



UNIVERSIDAD DE CHILE  
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS  
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

REDISEÑO DEL PROCESO DE CONTROL DE CALIDAD PARA UNA START-UP  
BASADA EN LA ECONOMÍA COLABORATIVA

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

SEBASTIÁN NICOLÁS VERA PÁEZ

PROFESOR GUÍA:  
OMAR CERDA INOSTROZA

MIEMBROS DE LA COMISIÓN:  
RENÉ ESQUIVEL CABRERA  
MARÍA JOSÉ CONTRERAS

SANTIAGO DE CHILE  
2020

## **REDISEÑO DEL PROCESO DE CONTROL DE CALIDAD PARA UNA START-UP BASADA EN LA ECONOMÍA COLABORATIVA**

En la actualidad, el futuro de las organizaciones ha estado migrando a modelos de producción y consumo en los cuales se privilegia la relación entre pares, los cuales brindan el servicio subyacente por medio de una plataforma digital (denominado como economía colaborativa). El mercado de estas empresas ha crecido exponencialmente y se espera que en 2025 sus facturaciones aumenten en un 2.000% en comparación con el año 2015. En este contexto cada vez más competitivo, es donde se desenvuelve Rocketpin, empresa especializada en el servicio de levantamiento de información en terreno, la cual es realizada por usuarios (shoppers) que descargan su aplicación móvil. El principal servicio provisto por la empresa consiste en la verificación de domicilio de personas que solicitan créditos bancarios y/o licencias médicas, el que es realizado por los shoppers a través de un formulario que deben llenar con toda la información del domicilio y la persona en cuestión (en adelante, "misiones").

Rocketpin se debe encargar de verificar la validez de la información entregada por los shoppers. Para cumplir con los estándares de calidad exigidos por su principal cliente (EQUIFAX), Rocketpin realiza una revisión del 100% de las misiones. Si el porcentaje de misiones correctas es inferior al 95% Rocketpin incurre en cuantiosas multas. El área de control de calidad es la encargada de revisar el contenido de las misiones y no puede cumplir con el flujo diario correspondiente a 800 misiones, revisando tan solo 450. Con un tiempo de revisión promedio de 100 segundos, los digitadores deberían trabajar 22,2 horas diarias para cumplir la meta. Dado esto, se propone un rediseño de este proceso, en base al diagnóstico de la situación actual, análisis de bases de datos y control estadístico del proceso.

Este análisis permitió identificar que el proceso de revisión costó un 23,4% de los ingresos en 2019 y que las variables más influyentes en la aprobación de misiones corresponden a: distancia entre los puntos de inicio y término, remuneración, tiempo de duración, confiabilidad y cantidad de misiones realizadas por el shopper. Estos resultados se validaron por medio de una regresión logística, modelo que posee una precisión del 98,9% para distinguir entre rechazo y aprobación. El rediseño propuesto se basó en la generación de un sistema de control que apruebe aquellas misiones con alta probabilidad de estar aprobadas y genere un conglomerado de las que deben revisarse más exhaustivamente. La implementación de este rediseño permitiría reducir el tiempo de revisión de las 800 misiones de 22,2 a 2,4 horas, lo cual implicaría que la capacidad diaria de revisión aumentaría de 1.350 a 9.600 misiones. Finalmente, se estimó que la diferencia entre el VAN de la situación actual y el rediseño propuesto a un horizonte de 5 años es igual a \$316.673.306, permitiendo concluir que el rediseño propuesto es factible y provechoso para Rocketpin.

## **DEDICATORIA**

*Para Fuki, espero donde quiera que estés, sepas que te amo con todo mi corazón y agradezco todos los hermosos momentos que vivimos juntos.*

## **AGRADECIMIENTOS**

En primer lugar, agradecer a mis padres, Soledad y Fernando por apoyarme en cada parte de este trayecto, brindándome su amor y cariño. Gracias madre por ser tan dedicada, esforzada y preocupada por nuestras necesidades, creo que con mi hermano no pudimos haber tenido más suerte. Estoy seguro que me van a faltar años de vida para pagarte todo lo que has hecho por mí, espero al menos poder regresarte la mínima parte. Gracias padre por ser un hombre amoroso, cariñoso y enfocado en sacar esta familia adelante. Gracias por tu entereza y tu fuerza para salir de las dificultades siempre victorioso. Gracias por tu alma tan noble y caritativa que siempre ha dado hasta que duela. Gracias a mi hermano por ser mi compañero de risas y aventuras, por las tardes que nos divertimos jugando videojuegos y por ser un niño lleno de bondad (te admiro profundamente). Gracias a mi primo Matías, que siempre ha sabido sacarme una sonrisa con sus payasadas y por amenizar el ambiente con su calidez (te considero un hermano más). Gracias a mi abuela Norma, que siempre quiso verme como profesional, espero donde estés, celebres conmigo. Gracias abuelo José por los trucos de magia y por aguantar a tus dos revoltosos nietos.

Quiero agradecer a mi principal pilar en todo este proceso universitario, mi polola Manuela, que ha estado estos 6 años conmigo. Gracias por tus múltiples y diversos regalos, que siempre supieron sorprenderme y llenarme el corazón de alegría. Gracias por compartir tantos momentos hermosos y que espero sigamos teniendo en el futuro próximo. Gracias por nunca dejarme desistir, a pesar de la infinidad de veces en que eso paso por mi mente. Gracias por ser mi compañera de batallas y porque juntos hemos podido avanzar y crecer como personas. Espero las risas y hamburguesas juntos no se terminen nunca y ten por seguro que vamos a cumplir todos nuestros sueños juntos.

Gracias a Felipe, fiel compañero de estudios, fiestas, festines de comida y diversión. Eres por lejos una de las mejores personas que he conocido y espero nuestra amistad se mantenga siempre. Gracias a Ricardo por sus consejos y su gran compañía. Gracias a Alan, María José, Stefan y Jorge por los buenos momentos.

Finalmente, quiero agradecer a Fuki porque cuando nos vimos por primera vez me robaste el corazón. Gracias por todo el tiempo que pasamos juntos, aunque fuera poco, no lo cambio por nada. Renovaste mi vida y me trajiste la mejor de las fortunas. Espero que te hayas ido tan llena de amor que no tengas ningún conflicto en esta tierra. Donde estés, ojalá sepas que no hay día que no te recuerde y te prenda una vela. ¡Muchas gracias!

## **TABLA DE CONTENIDOS**

<b>1. INTRODUCCIÓN</b> .....	<b>9</b>
<b>1.1. Antecedentes generales de la economía colaborativa</b> .....	<b>9</b>
<b>2. CARACTERIZACIÓN DE LA ORGANIZACIÓN</b> .....	<b>13</b>
<b>2.1. Funcionamiento de Rocketpin</b> .....	<b>14</b>
<b>2.2. Relación con el sector industrial</b> .....	<b>15</b>
<b>2.3. Estructura Organizacional</b> .....	<b>15</b>
<b>3. PLANTEAMIENTO DEL PROBLEMA Y JUSTIFICACIÓN</b> .....	<b>16</b>
<b>4. OBJETIVOS</b> .....	<b>22</b>
<b>4.1. Objetivo General</b> .....	<b>22</b>
<b>4.2. Objetivos Específicos</b> .....	<b>22</b>
<b>5. MARCO CONCEPTUAL</b> .....	<b>23</b>
<b>5.1. Rediseño de procesos</b> .....	<b>23</b>
5.1.1. Métodos de rediseño .....	25
<b>5.2. Procedimientos de control de procesos</b> .....	<b>28</b>
5.2.1. Control de procesos por atributos: uso de gráficas P .....	29
5.2.2. Control de procesos por variables: uso de gráficas X y R.....	30
5.2.3. Modelos de clasificación .....	31
5.2.4. Regresión logística .....	33
<b>5.3. CALIDAD COMO CONCEPTO ESTADÍSTICO</b> .....	<b>35</b>
5.3.1. Costos de calidad .....	36
<b>6. METODOLOGÍA</b> .....	<b>36</b>
<b>6.1. ANÁLISIS DE LA SITUACIÓN ACTUAL</b> .....	<b>37</b>
<b>6.2. ANÁLISIS EXPLORATORIO DE LAS BASES DE DATOS</b> .....	<b>38</b>
<b>6.3. PROCEDIMIENTOS DE CONTROL</b> .....	<b>38</b>
<b>6.4. REDISEÑO DEL PROCESO ACTUAL</b> .....	<b>39</b>
<b>6.5. EVALUACIÓN DE LOS CAMBIOS PROPUESTOS</b> .....	<b>40</b>
<b>7. DIAGNÓSTICO DE LA SITUACIÓN ACTUAL</b> .....	<b>40</b>
<b>7.1. Aplicación de Rocketpin</b> .....	<b>40</b>
<b>7.2. Macroprocesos del negocio</b> .....	<b>43</b>
7.2.1. Capacitación de shoppers.....	44
7.2.2. Realización de misiones .....	45
7.2.3. Revisión de misiones .....	46

7.3.	Descripción de la situación actual del negocio .....	49
7.4.	Análisis del proceso de revisión .....	52
7.5.	Selección de la variable crítica.....	55
<b>8.</b>	<b>ANÁLISIS DE LAS BASES DE DATOS .....</b>	<b>55</b>
8.1.	Clasificación de motivos de rechazo .....	56
8.2.	Análisis exploratorio de la base de misiones .....	57
8.3.	Análisis exploratorio de la base de shoppers .....	64
8.4.	Análisis conjunto de las bases de datos.....	69
<b>9.</b>	<b>CONTROL DEL PROCESO DE REVISIÓN.....</b>	<b>73</b>
9.1.	Gráficos de control .....	74
9.1.1.	Uso de gráficas por atributo: tipo P .....	74
9.1.2.	Uso de gráficas por variables: tipo X y R.....	76
9.1.3.	Regresión Logística .....	83
9.1.4.	Validación del modelo.....	90
9.1.5.	Interpretación de los resultados.....	94
<b>10.</b>	<b>REDISEÑO DEL PROCESO DE REVISIÓN .....</b>	<b>96</b>
10.1.	Elección de la potencial solución .....	96
10.2.	Diseño de la solución escogida .....	99
<b>11.</b>	<b>EVALUACIÓN TEÓRICA DE LOS CAMBIOS PROPUESTOS .....</b>	<b>103</b>
11.1.	Mejoras potenciales de la variable crítica .....	103
11.2.	Factibilidad financiera de la solución propuesta .....	104
11.2.1.	Costos del proceso de revisión actual.....	104
11.2.2.	Costos de la solución propuesta .....	107
11.2.3.	Estimación de los indicadores financieros .....	108
<b>12.</b>	<b>CONCLUSIONES.....</b>	<b>111</b>
12.1.	Conclusiones .....	111
12.2.	Recomendaciones .....	114
12.3.	Comentarios finales .....	115
<b>13.</b>	<b>BIBLIOGRAFÍA .....</b>	<b>117</b>
<b>14.</b>	<b>ANEXO.....</b>	<b>123</b>

## ÍNDICE DE ILUSTRACIONES

Ilustración 1: Ingresos por ámbitos de economía colaborativa en Europa [4]. ...	9
Ilustración 2: Distribución porcentual de la inversión en plataformas de economía colaborativa a nivel mundial en el período de 2000 a 2015, por sectores [4]. .....	10
Ilustración 3: Distribución porcentual de los usuarios participantes en economía colaborativa en 2014 a nivel mundial, por tipo de industria [4]. .....	11
Ilustración 4: Distribución porcentual de usuarios por sector de aplicaciones de economía colaborativa en Chile. Fuente: Pontificia Universidad Católica de Chile. ....	12
Ilustración 5: Cronología de los hitos más importantes en la trayectoria de Rocketpin. ....	14
Ilustración 6: Organigrama de Rocketpin. ....	15
Ilustración 7: Cantidad de misiones completadas versus revisadas por los digitadores. ....	19
Ilustración 8: Cantidad de misiones revisadas por los digitadores. ....	19
Ilustración 9: Desglose del problema estudiando mediante un árbol de problemas. ....	21
Ilustración 10: Visión de conjunto del método de rediseño de procesos. ....	26
Ilustración 11: Perspectivas del costo de la variabilidad para la sociedad. ....	29
Ilustración 12: Ejemplo de elaboración de un gráfico de control P en base a 10 muestras de 300 formas de cobro. ....	30
Ilustración 13: Representación de la función logística. ....	34
Ilustración 14: Concepto estadístico de la calidad. ....	35
Ilustración 15: Funcionalidades de la aplicación Rocketpin. ....	41
Ilustración 16: Perfil del shopper en la aplicación Rocketpin. ....	41
Ilustración 17: Vistas de las misiones en la aplicación Rocketpin. ....	42
Ilustración 18: Misiones de entrenamiento para los shoppers. ....	44
Ilustración 19: Diagrama BPMN de la situación actual del proceso de revisión. ....	49
Ilustración 20: Gráfico de barra de la cantidad de misiones realizadas por mes. ....	50
Ilustración 21: Gráfico de barra del gasto mensual en pago a shoppers. ....	51
Ilustración 22: Gráfico de barra de los ingresos mensuales de Rocketpin. ....	52
Ilustración 23: Gráfico de distribución de los tiempos de revisión de misiones. ....	53
Ilustración 24: Descripción de las variables en la base de datos de motivos de rechazo. ....	56
Ilustración 25: Categorización de los motivos de rechazo de las misiones. ....	57
Ilustración 26: Descripción de las variables de la base de misiones. ....	58
Ilustración 27: Descripción de los tipos de datos en Python [23]. ....	59
Ilustración 28: Variables creadas a partir de la base de misiones. ....	60
Ilustración 29: Comunas donde más se realizan misiones de verificación domiciliaria. ....	61
Ilustración 30: Frecuencia de las verificaciones realizadas por cliente. ....	62
Ilustración 31: Boxplot del pago a shoppers. ....	62

Ilustración 32: Correlación entre las variables de la base de datos de misiones. .....	64
Ilustración 33: Descripción de las variables en la base de datos de shoppers. .	65
Ilustración 34: Variables creadas a partir de la base de shoppers. ....	66
Ilustración 35: Distribución de la edad de los shoppers. ....	66
Ilustración 36: Distribución de la experiencia de los shoppers. ....	67
Ilustración 37: Distribución del género de los shoppers.....	68
Ilustración 38: Distribución de las regiones de origen de los shoppers.....	68
Ilustración 39: Distribución de la tasa de cumplimiento de los shoppers. ....	69
Ilustración 40: Ejemplificación de un left join mediante la teoría de conjuntos.	70
Ilustración 41: Análisis de las nuevas variables agregadas a la base de misiones. .....	70
Ilustración 42: Distribución de las misiones por edad y género del shopper. ...	71
Ilustración 43: Distribución de las misiones por edad y estado de la misión. ...	72
Ilustración 44: Correlograma de las variables de shoppers incluidas en la base de misiones. ....	73
Ilustración 45: Gráfico del tipo P para la tasa de desaprobación de misiones... 76	76
Ilustración 46: Gráficos del tipo X y R para la distancia entre inicio y fin de una misión.....	78
Ilustración 47: Gráficos del tipo X y R para las calificaciones promedio de los shoppers. ....	79
Ilustración 48: Gráficos del tipo X y R para el Score de los shoppers.....	80
Ilustración 49: Gráficos del tipo X y R para la duración de una misión. ....	81
Ilustración 50: Gráficos del tipo X y R para el tiempo entre que se tome e inicia una misión. ....	82
Ilustración 51: Gráficos del tipo X y R para la tasa de aprobación en los últimos 3 meses. ....	83
Ilustración 52: Correlograma de las variables más influyentes en el rechazo de misiones.....	85
Ilustración 53: Particiones realizadas a la base de datos para la regresión logística. ....	87
Ilustración 54: Resultados de la regresión logística en Python.....	88
Ilustración 55: Ejemplificación de una matriz de confusión. ....	90
Ilustración 56: Matriz de confusión realizada para la predicción de la regresión logística. ....	93
Ilustración 57: Métricas resultantes de la evaluación de la regresión logística. 93	93
Ilustración 58: Cross Validation de 8 secciones para la base de entrenamiento original. ....	94
Ilustración 59: Calificación ponderada de las soluciones usando el enfoque de factores.....	98
Ilustración 60: Gráfico del tipo X para la probabilidad de aprobación.....	100
Ilustración 61: Distribución de las misiones según cuadrante.....	101
Ilustración 62: Ejemplo de un sistema de control de lazo cerrado. ....	102
Ilustración 63: Diagrama BPMN del rediseño aplicado al proceso de revisión. 103	103



Ilustración 64: Simulación del proceso de revisión para estimar costos de falla externa. ....	106
Ilustración 65: Ítems necesarios para la estimación del flujo de caja. ....	109
Ilustración 66: Flujo de caja de Rocketpin considerando la situación actual... ..	110
Ilustración 67: Flujo de caja de Rocketpin considerando la solución propuesta. ....	110
Ilustración 68: Marco del contrato de prestación de servicios entre EQUIFAX y Rocketpin. ....	123
Ilustración 69: Proporción de las misiones realizadas por estado en el año 2019. ....	124
Ilustración 70: Gráficos de control de evidencias para una investigación.....	125
Ilustración 71: Factores para determinar los límites de control de las gráficas X y R.....	126
Ilustración 72: Ejemplo de un informe de costos de calidad por categoría.....	127
Ilustración 73: Preguntas realizadas en la misión de Entrenamiento 1.....	128
Ilustración 74: Preguntas realizadas en la misión de Entrenamiento 2.....	129
Ilustración 75: Ejemplo del formulario de una verificación domiciliaria. ....	131
Ilustración 76: Medición de los tiempos de revisión para 150 misiones.....	132
Ilustración 77: Distribución de la edad de los shoppers mediante un boxplot.	133
Ilustración 78: Ejemplo de una cross validation. ....	133

## **1. INTRODUCCIÓN**

### **1.1. Antecedentes generales de la economía colaborativa**

La cantidad de empresas que utilizan un modelo de negocios basado en la economía colaborativa ha ido en aumento, tanto en Chile como en la mayoría de los países del mundo, las cuales están presentes en distintas industrias como el transporte, hotelería, retail, entre otros. La economía colaborativa conforma aquellos modelos de producción, consumo o financiación que se basan en la intermediación entre la oferta y la demanda generada en relaciones entre iguales (P2P o B2B), o bien de particular a profesional, a través de plataformas digitales que no prestan el servicio subyacente, generando un mayor aprovechamiento de los bienes y recursos existentes, los cuales son infrautilizados. De esta manera, la economía colaborativa permite utilizar, compartir, intercambiar o invertir recursos o bienes, pudiendo existir o no una contraprestación monetaria entre los usuarios y generalmente la interacción entre los shoppers se ve mediada por plataformas digitales [1].

Por ello, el principal valor agregado de esta nueva tendencia a la colaboración, además de ganancias económicas, es la producción y el desarrollo de conocimiento, ya que al ser compartido puede tomarse como punto de partida para iniciar nuevos modelos de negocio. En este sentido, el modelo de la economía colaborativa puede traer beneficios personales, tanto a nivel económico como de crecimiento personal e intelectual, pero está sujeto a una limitación: el deseo del individuo. El verdadero potencial de este modelo radica en la integración de varias personas en el proceso de generación de valor, tanto económico como social. Al establecer que cada individuo puede trabajar en un área de especialización diferente en esta economía, se descubre que existe un potencial infinito de posibilidades de creación, innovación e incluso empoderamiento financiero en pequeños, medianos y grandes proyectos [2].

Se estima que la facturación de las organizaciones que utilizan este modelo de negocio aumente en un 2.000% para el año 2025 en comparación al año 2015, pasando de 15.000 a 335.000 millones de dólares. En España estas compañías suponen entre el 1% y el 1,4% del PIB total, pero se espera que para 2025 contemplen aproximadamente el 2,9% de éste [3]. Si se segmentan los ingresos de estas organizaciones por sectores industriales, hasta el año 2015, el área de transporte es el que más ingresos ha registrado en Europa con un total de 1.650 millones de euros; seguido por el sector de alojamiento con un total de 1.150 millones de euros [4], lo cual puede apreciarse en el siguiente gráfico:

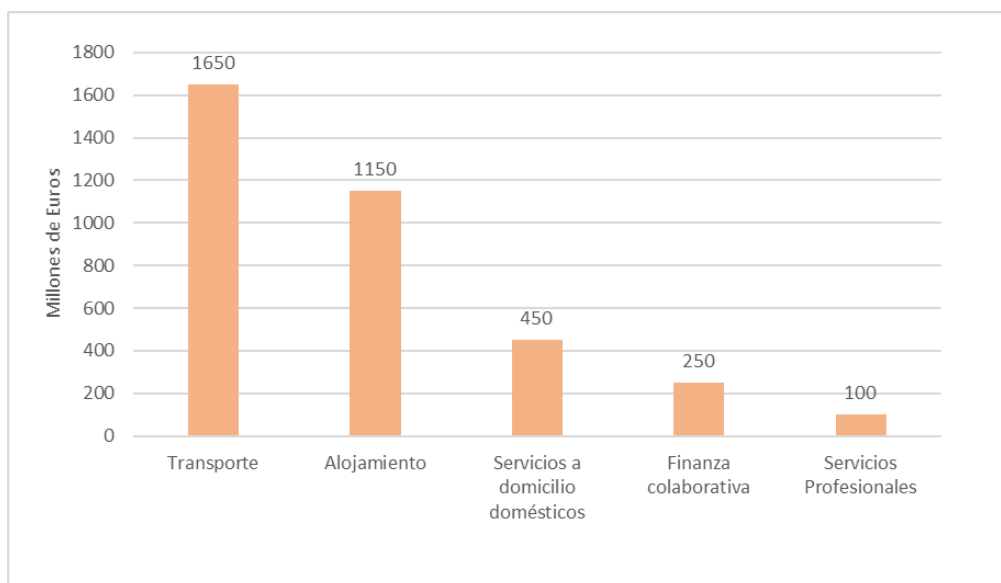


Ilustración 1: Ingresos por ámbitos de economía colaborativa en Europa  
Fuente: The Ostelea School of The Tourism & Hospitality.

Que el sector de transporte sea el que más ingresos ha obtenido va muy ligado con el nivel de inversión que ha habido en ese rubro, motivado principalmente por la gran cantidad de usuarios de plataformas y empresas de economía colaborativa ligadas presentes en el mercado, lo cual puede apreciarse en los gráficos presentados a continuación:

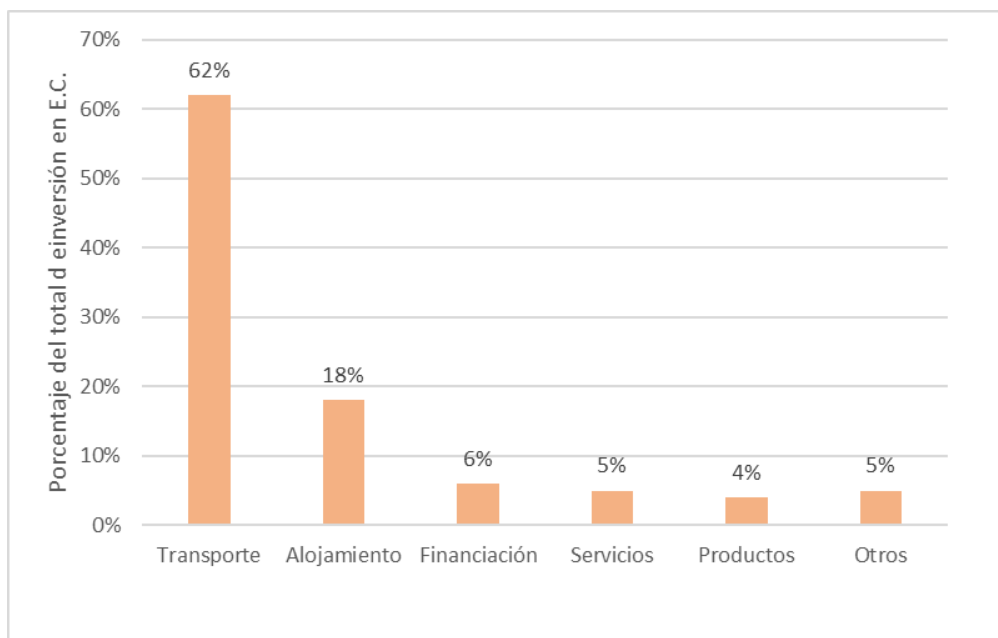


Ilustración 2: Distribución porcentual de la inversión en plataformas de economía colaborativa a nivel mundial en el período de 2000 a 2015, por sectores [4].  
Fuente: The Ostelea School of The Tourism & Hospitality.

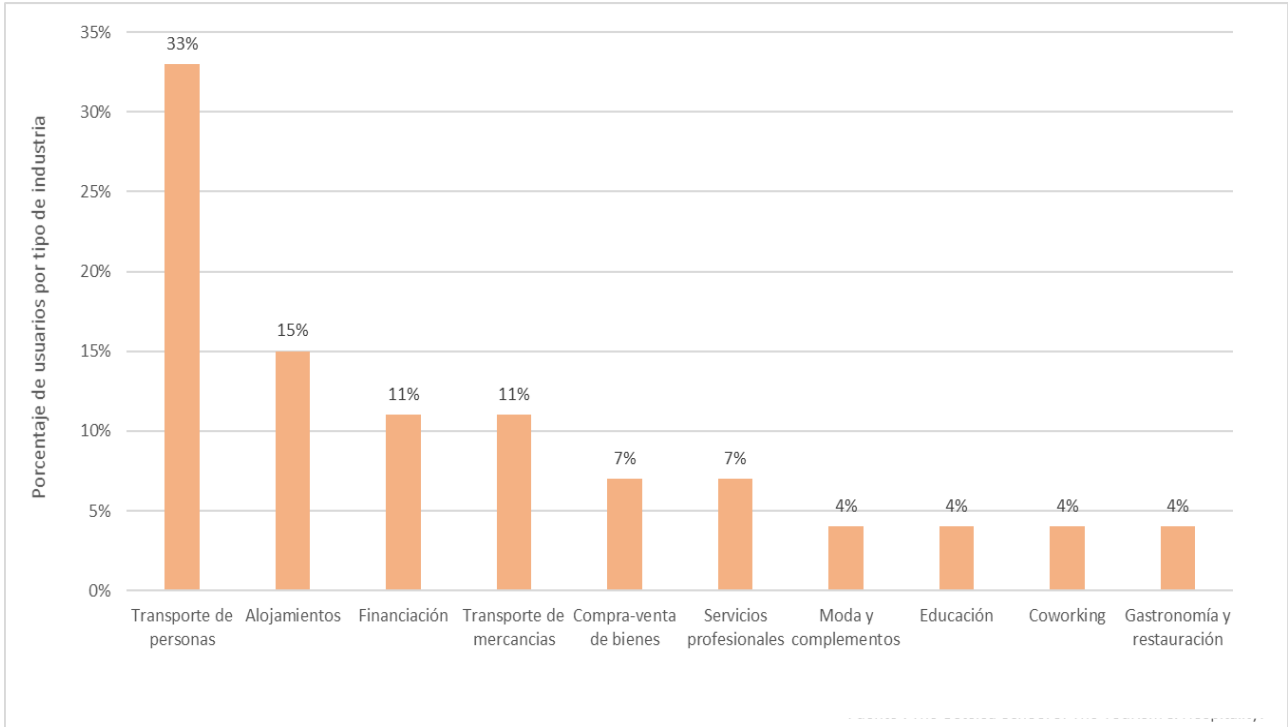


Ilustración 3: Distribución porcentual de los usuarios participantes en economía colaborativa en 2014 a nivel mundial, por tipo de industria [4].

Fuente: The Ostelea School of The Tourism & Hospitality.

En Chile, el crecimiento de este tipo de aplicaciones también ha seguido la tendencia mundial y el más claro ejemplo de esto es Uber, aplicación móvil en la que una persona puede utilizar su propio vehículo para volverse taxi y llevar a otras personas a su destino. Desde que llegaron a Chile en enero de 2014 con 50 choferes, se han expandido para estar presentes en 26 ciudades del país, contar con 85.000 conductores y transportar a cerca de 2,2 millones de usuarios cada mes [5].

La distribución porcentual de usuarios participantes en aplicaciones de economía compartida por sector en Chile, también sigue una tendencia similar a lo ocurrido en Europa, donde se destaca la utilización de aplicaciones ligadas al transporte de pasajeros. Esto puede apreciarse en el siguiente gráfico, el cual fue realizado por la Universidad Católica de Chile y TrenDigital [6] usando la información obtenida de una encuesta realizada a 1.141 personas con acceso a internet:

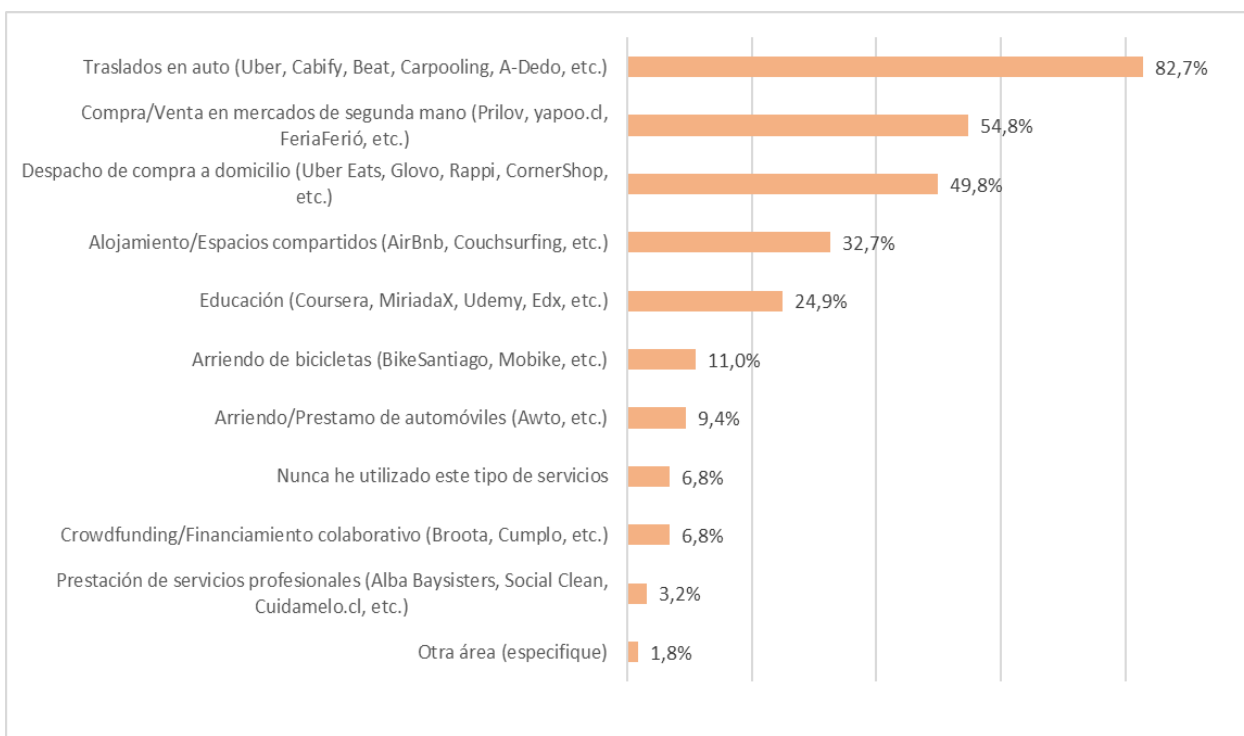


Ilustración 4: Distribución porcentual de usuarios por sector de aplicaciones de economía colaborativa en Chile.

Fuente: Pontificia Universidad Católica de Chile.

Este tipo de organizaciones representan, según varias fuentes, el futuro de la economía global, de hecho, para el año 2011 la revista Times señaló que la economía colaborativa es una de las 10 ideas que cambiaran el mundo. Dado esto, es que se hace necesario tener un marco regulatorio y legislativo en el que basarse para mediar las relaciones entre los actores pertenecientes a estas organizaciones, ya que actualmente no existe ninguna ley y/o ente encargado de supervisar el correcto funcionamiento de estas empresas ni que asegure el cumplimiento de los derechos laborales de los trabajadores de este tipo de negocios.

En este entorno versátil, creciente y competitivo se está abriendo paso la *Start-Up* chilena Rocketpin, la cual basa su modelo de negocios en el concepto de la economía colaborativa. Su principal servicio corresponde al levantamiento de información *on-demand* por medio de usuarios que descargan su aplicación móvil, aprovechando que estos tienen tiempos ociosos en sus rutinas diarias (y, por ende, subutilizan ese recurso). El presente informe se centrará en los principales dolores a los que se está viendo enfrentada la organización, con el fin de definir un proyecto que pueda hacerse cargo de estos.

## 2. CARACTERIZACIÓN DE LA ORGANIZACIÓN

Rocketpin es una empresa conformada como una sociedad por acciones (SPA) con principal foco en levantamiento de información en terreno *on-demand*. Fue fundada el año 2015 en Santiago de Chile y actualmente también cuenta con presencia en Argentina y Uruguay. El negocio de Rocketpin consiste en entregar a las empresas que los contratan, el levantamiento de información en terreno para campañas donde les sería muy costoso recabar la información por sí mismos. Para llevar a cabo esta obtención de información, Rocketpin cuenta con una aplicación móvil, la cual debe ser descargada por los usuarios (en adelante shoppers) para poder brindar el servicio. La recolección de información por parte de los shoppers (en adelante misiones), se realiza llenando un formulario presente en la aplicación, el cual es diferente dependiente el tipo de servicio de levantamiento de información. Una vez recolectada toda la información relativa a una campaña en específico, Rocketpin genera el entregable final que consiste en un conglomerado de los datos recabados.

La empresa lleva como misión y visión (no hacen distinción entre ambas) “ser la comunidad de trabajo colaborativo más grande de Latinoamérica, generando ingresos extras para nuestros usuarios e información ágil, confiable y económica a nuestros clientes”, especificando como su propósito el mejorar la calidad de vida de sus shoppers.

La idea de fundar Rocketpin surgió después de que sus principales accionistas, Diego y Raúl Reeves notasen el poder de las plataformas colaborativas que hasta el 2015 estaban creciendo con mayor presencia en el mercado. Ejemplos como el de Uber, Waze y AirBnB donde los usuarios eran quienes proveían la información que sustentaba a la aplicación, les dieron la idea de extrapolar esta plataforma colaborativa al sector de verificaciones de domicilio, con la promesa de incentivar a sus usuarios mediante remuneraciones por su labor.

Desde entonces, Rocketpin ha ido en ascenso; trascendiendo de un volumen de 200 millones en ventas en el 2016 a contar con 1.000 millones en ventas hacia el 2018, aumentando el flujo de misiones realizadas mensualmente de 2.500 a 40.000 e incrementando la cantidad total de trabajadores en planta de 5 a 25, dentro de ese mismo período.

Este crecimiento se puede explicar gracias a la gran cantidad de fondos concursables obtenidos por la empresa, los cuales se especifican dentro de la siguiente línea temporal de hitos relevantes de Rocketpin:



Ilustración 5: Cronología de los hitos más importantes en la trayectoria de Rocketpin.  
Fuente: Rocketpin.

Hacia el año 2018, Rocketpin inicia operaciones fuera del país instaurándose primero en Argentina y posteriormente en Uruguay. Esta expansión se debió en parte a que la gerencia de la empresa conocía personas capaces de hacerse cargo de la operación fuera del país, además de ver en aquellos lugares oportunidades de poder extrapolar el modelo de negocios chileno y asegurar una gran presencia de mercado. De momento, las operaciones y escalas en aquellos países son muy inferiores a las obtenidas en Chile, y todavía es muy pronto para poder delimitar hasta qué punto va a escalar la empresa en el extranjero.

## 2.1. Funcionamiento de Rocketpin

En líneas generales, el funcionamiento de Rocketpin consiste en ser contratada por empresas que requieren obtener información sobre la operación de su negocio o, principalmente en el caso de instituciones financieras, obtener información sobre sus clientes que normalmente les tomaría mucho tiempo recolectar (como una verificación laboral o de domicilio).

Una vez se define la información que se busca recopilar, el conjunto de misiones con un fin similar se denomina campaña y la información de las sucursales dónde se deben obtener estos levantamientos de información se habilitan dentro de la aplicación móvil de Rocketpin. Estando la información allí, los shoppers pueden ir a realizar la misión, y la información es enviada a

la base de datos interna de Rocketpin, donde es verificada por un trabajador de la empresa para comprobar que la información sea coherente.

Si todo está en orden, se almacena la información en la base de datos y se le entrega una remuneración al shopper por su trabajo. Además de lo anterior, también se realizan las denominadas Campañas Play, que son encuestas que los shoppers pueden responder desde la comodidad de su celular pero que en su mayoría no son remuneradas. Estas encuestas solían utilizarse como método de encuesta masivo en etapas iniciales del desarrollo del negocio.

## 2.2. Relación con el sector industrial

Se puede señalar que Rocketpin tiene presencia tanto en los sectores industriales de Estudios de Mercado, contando con una participación de mercado de un 1% en este rubro y con especialidad en levantamiento de información en terreno; y el sector de Servicios de la Información, con una presencia de mercado del 30% y con principal foco en las instituciones financieras<sup>1</sup>.

## 2.3. Estructura Organizacional

La estructura organizacional de Rocketpin queda delimitada por el siguiente organigrama, cabe mencionar que en este no se especifican los trabajadores de Argentina (2) ni Uruguay (2):

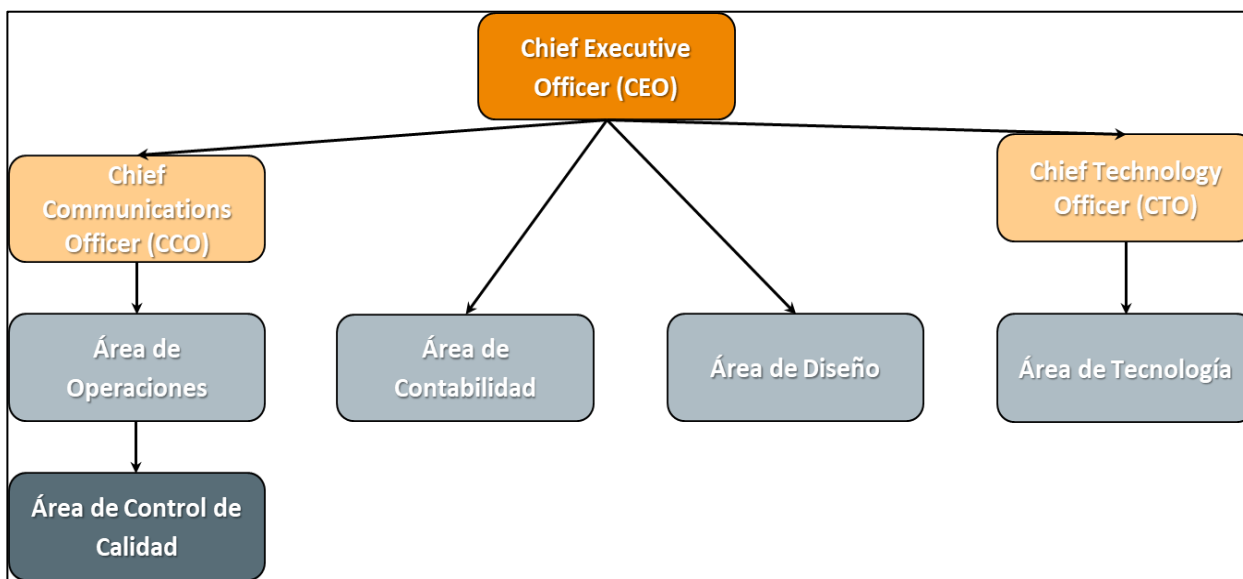


Ilustración 6: Organigrama de Rocketpin.

Fuente: Elaboración Propia.

<sup>1</sup> Cifra obtenida por elaboración propia de Rocketpin. Esta métrica se tomó comparando las presencias en redes sociales frente a sus principales competidores (Roamer y Snuuper).



De la ilustración anterior, es posible apreciar que las áreas de contabilidad y diseño son consideradas como las de mayor independencia dentro de la empresa, ya que solo deben reportar su trabajo directamente al CEO de Rocketpin, lo cual limita su interacción con el resto de las áreas en proyectos puntuales y basándose principalmente en los datos que proveen estas otras, más que en una relación constante sobre lo realizado día a día. El área de contabilidad se encarga principalmente de realizar los pagos a los shoppers y a los trabajadores de la empresa, mientras que el área de diseño se encarga de confeccionar las campañas publicitarias y de la estética visual del sitio web.

Distinto es el caso del área de tecnología y operaciones, las cuales están fuertemente relacionadas. El área de operaciones tiene como principal labor supervisar que las misiones para cada campaña se realicen en tiempo y forma, implicando hacer una revisión constante de las misiones que los shoppers envían a la empresa (tarea que se le asignó a la nueva área de control de calidad) y también contactando shoppers para que realicen misiones que están atrasándose con respecto a los plazos pactados con el cliente. En el caso del área de tecnología, ellos velan principalmente por mantener tanto la aplicación como el sitio web y bases de datos operacionales en todo momento, además de esto, se encargan de proveer soporte tecnológico a los trabajadores del área de operaciones, ya sea confeccionando un archivo entregable a los clientes o aprovechando su experiencia técnica para realizar tareas de forma más expedita.

Finalmente, el área donde se realizará el trabajo de memoria explicitado en este informe corresponde a la de control de calidad. Esta área fue creada recientemente como una división del área de operaciones, para que esta última se encargara principalmente de gestionar el cumplimiento del servicio solicitado por los clientes de la empresa. Dado esto, el personal del área de control de calidad está compuesto principalmente por personal antiguo del área de operaciones, con Jorge Díaz (*Key Account Manager*) encargado de supervisar a 3 digitadores, los cuales se encargan del proceso de revisión de misiones enviadas por los shoppers a Rocketpin. Esta labor es fundamental, ya que sin ella Rocketpin no podría cumplir con las altas exigencias en cuando a rapidez y calidad exigida por sus principales clientes.

### **3. PLANTEAMIENTO DEL PROBLEMA Y JUSTIFICACIÓN**

Rocketpin es una *Start-Up* cuyo principal servicio corresponde al levantamiento de información en terreno por medio de shoppers *on-demand*. El servicio se sustenta en que hay empresas que requieren obtener información sobre la operación de su negocio o de sus clientes, pero la recopilación de estos datos les tomaría mucho tiempo y recursos.

El levantamiento de información se realiza a través de una aplicación móvil, donde los shoppers corresponden a usuarios activos de la aplicación, los cuales por medio de misiones provistas por Rocketpin, levantan información a cambio de una remuneración. Para cumplir las misiones, los shoppers deben llenar un formulario con la información del servicio, por ejemplo, para una misión de verificación domiciliaria, deben ingresar los datos de la persona que habita en el lugar y tomar múltiples fotografías a la fachada de la vivienda. Rocketpin debe comprobar que las misiones se realizan de forma adecuada, para lo cual es necesario validar que la información recopilada por los shoppers sea correcta y verídica. Finalmente, cuando se aprueban todas las misiones relativas a un cliente, Rocketpin consolida la información recopilada en una base de datos y procede a realizar la entrega de ésta al cliente.

El principal cliente de Rocketpin es EQUIFAX, la cual es una agencia de informes de crédito al consumo. EQUIFAX recopila y agrega información de más de 800 millones de consumidores individuales y más de 88 millones de empresas en todo el mundo (principalmente bancos). Dada la necesidad de recopilar información, EQUIFAX subcontrata a Rocketpin para realizar verificaciones de domicilio cuyo objetivo es validar la información de clientes que piden créditos bancarios. El volumen de verificaciones solicitadas por EQUIFAX corresponde al 80% del total de misiones que Rocketpin provee en su aplicación (alrededor de 32 mil mensuales) y representa el 50% del total de ingresos del negocio (aproximadamente 500 millones en 2018).

Para realizar las misiones de verificación de domicilio provistas por EQUIFAX, los shoppers deben llenar un formulario específico que es diferente al que se debe llenar para verificaciones de domicilio solicitadas por otras empresas. En este formulario el agente debe ingresar fotografías de la fachada, reja y numeración del domicilio, ingresar información acerca de la peligrosidad de la zona y una serie de datos acerca de la persona que lo atendió. Dada la importancia de este cliente para Rocketpin, se generó un área (control de calidad) que se dedicara exclusivamente a la revisión de las misiones de verificación de domicilio provistas por EQUIFAX. A pesar de lo anterior, el proceso de revisión de misiones es uno de los principales desafíos de Rocketpin, ya que tan solo un error encontrado en la base de datos enviada al cliente se podría traducir en cuantiosas multas (se descuenta la cantidad de misiones desaprobadas por el cliente, sumado a un cargo extra por incumplimiento de los plazos pactados) y además hace que la empresa pierda prestigio.

Si bien se estudiará el problema que genera el proceso de revisión de misiones para el caso particular de EQUIFAX, es importante señalar que los demás servicios de levantamiento de información brindados por Rocketpin siguen

procedimientos similares, por lo cual se puede decir que las deficiencias encontradas en este proceso son generalizables para la gran mayoría de servicios y clientes de la empresa. Para comprender los estándares de cumplimiento que impone EQUIFAX a Rocketpin, es útil analizar el marco del contrato de prestación de servicios entre ambas empresas, el cual puede apreciarse en la ilustración 68 de anexo de este informe. Los puntos evidenciados en el contrato permiten entender las altas exigencias que tiene Rocketpin en cuanto al volumen de verificaciones aprobadas (75% del total) y los estándares de calidad que debe cumplir (95% de aprobación). Es posible apreciar que las multas son bastante elevadas dependiendo del nivel de retraso de las verificaciones (se cobra dos veces el precio de la verificación dependiendo de la cantidad de días atrasados y de la región) y la tasa de aprobación. Sin embargo, Rocketpin también puede obtener beneficios si es que el cumplimiento supera el 95%, por lo cual existen incentivos a realizar un control exhaustivo de las misiones con el fin de obtener el multiplicador a la facturación neta.

Actualmente, Rocketpin puede cumplir con estas altas exigencias en la mayoría de los casos y, por ende, con las expectativas de sus clientes. De hecho, la propuesta de valor de la organización se sustenta en realizar el levantamiento de información 100% correcta en el menor tiempo posible, siendo estos los principales factores que le permiten a Rocketpin diferenciarse de sus competidores. Sin embargo, para cumplir esta propuesta de valor, el área de control de calidad debe revisar el 100% de las misiones de verificación domiciliaria (aproximadamente 2.400 diarias). Esta tarea se divide en los 3 digitadores que posee el área de control de calidad, por ende, cada uno tiene como meta revisar 800 misiones diarias. Para cumplir esta meta, los digitadores trabajan en jornadas de 9:30 a 19:00 horas (47,5 horas semanales) y además deben trabajar un sábado al mes (horas extra).

El proceso de revisión de misiones es bastante engorroso, ya que los digitadores deben comprobar que cada campo del formulario fue llenado correctamente y dada la complejidad de este, la detección de errores es una tarea difícil. Cada revisión toma aproximadamente 40 segundos cuando poseen confianza en el shopper que lo envía (este criterio se basa principalmente en la experiencia anterior que hayan tenido los digitadores con el shopper) y toma entre 2-3 minutos cuando posee algún error. Si se considera en promedio 1 minuto por revisión, los digitadores deberían tener jornadas de 13 horas diarias para cumplir con su meta, lo cual es excesivo y produce mucha presión. Para ejemplificar la situación anterior, se graficó para un período de 17 días, la cantidad de misiones revisadas por los digitadores del área de control de calidad en contraste con la cantidad de misiones realizadas por los shoppers, lo cual puede notarse a continuación:

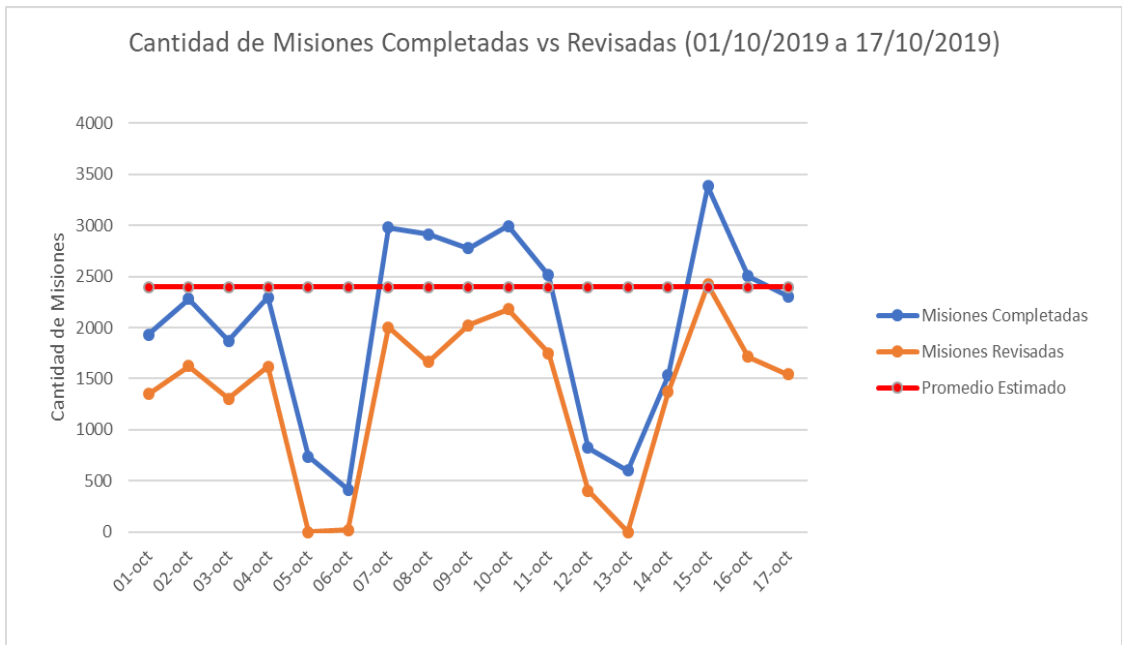


Ilustración 7: Cantidad de misiones completadas versus revisadas por los digitadores.  
Fuente: Elaboración Propia.

La ilustración 7 indica que, en ninguno de los días estudiados, los digitadores pudieron revisar el total de misiones realizadas y la brecha es aproximadamente de 500 misiones en promedio. Al analizar la cantidad de misiones revisadas por digitador se obtiene lo siguiente:

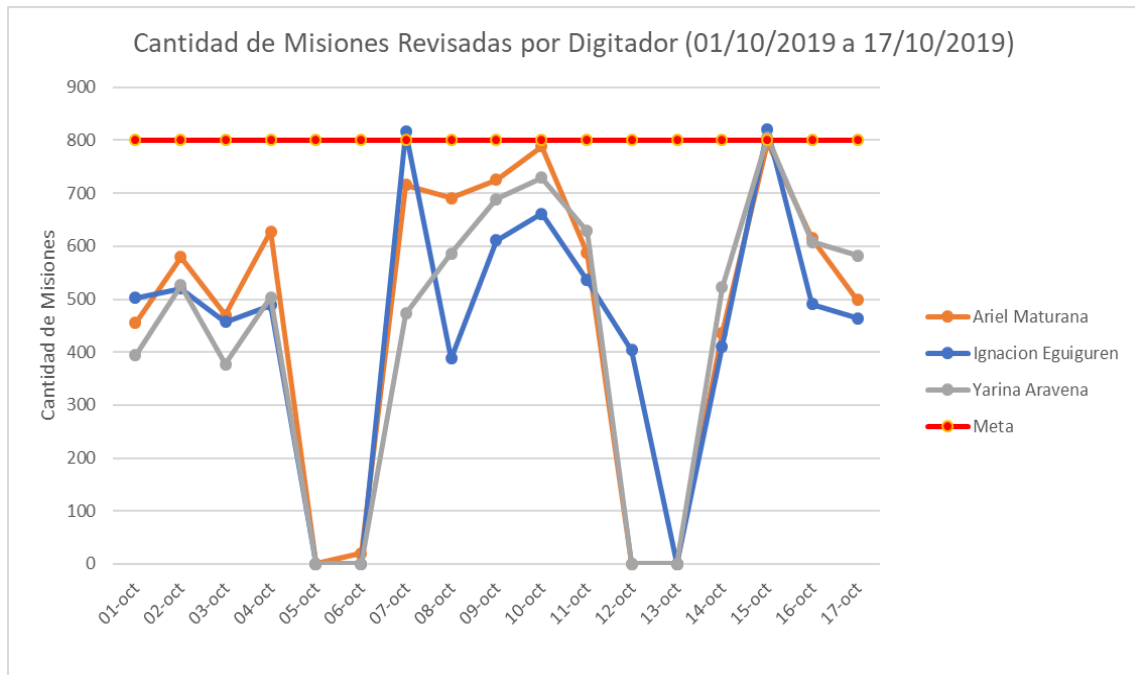


Ilustración 8: Cantidad de misiones revisadas por los digitadores.  
Fuente: Elaboración Propia.

De la ilustración 8 se puede extraer que, de los 17 días estudiados, tan solo en 2 de estos los digitadores pudieron cumplir con la meta de 800 misiones revisadas, teniendo un promedio de 450 misiones revisadas por día. Además, si bien existen normas que les permiten a los digitadores evaluar el formulario, este proceso no está estandarizado y, por ende, tampoco se puede evaluar que cada digitador está realizando bien su labor.

El porcentaje de misiones rechazadas por los digitadores es de aproximadamente un 9,6% (esto puede apreciarse en la ilustración 69 del anexo de este informe) y la cantidad de falsos positivos (entiéndase por falsos positivos, la cantidad de misiones que fueron aprobadas y posteriormente, rechazadas por la empresa contratante cuando se revisó el consolidado de las misiones) está entre un 2,5% a 10% del total de misiones realizadas<sup>2</sup>, lo que sugiere que en algunos casos se incumple con los estándares de calidad estipulados en el contrato, provocando el cobro de multas a la organización.

Rocketpin no posee ninguna métrica y/o indicador sobre el desempeño de los usuarios, lo que impide facilitar la tarea de revisión de misiones. En base a lo anterior y a modo de ejemplo, si a shoppers con mayor experiencia y tasa de cumplimiento se les controlara menos que a usuarios nuevos, se podría reducir la carga de trabajo para el área de control de calidad, ya que no tendrían que revisar el 100% de las misiones. Además, es importante destacar que la estrategia de Rocketpin contempla crecer en volumen de servicios brindados (y por ende, también en cantidad de misiones provistas a sus shoppers), para lo cual no es sostenible seguir teniendo el mismo proceso de control de misiones, ya que con el volumen actual, el personal ya se encuentra sobrecargado.

Dado todo lo anterior, es que un rediseño de este proceso es de suma importancia y urgencia para Rocketpin, ya que no es sostenible para el cumplimiento de su estrategia ni para el área de control de calidad seguir con el proceso de revisión actual. A modo de recapitulación y para entender el problema encontrado en el proceso de revisión de manera más general, se utilizará la herramienta gráfica conocida como árbol de problemas, ya que permite representar qué es lo que está ocurriendo (problema principal), por qué está ocurriendo (causas) y que es lo que esto está ocasionando (los efectos o consecuencias):

---

<sup>2</sup> Cifra otorgada por el jefe del área de control de calidad, en base a su experiencia.

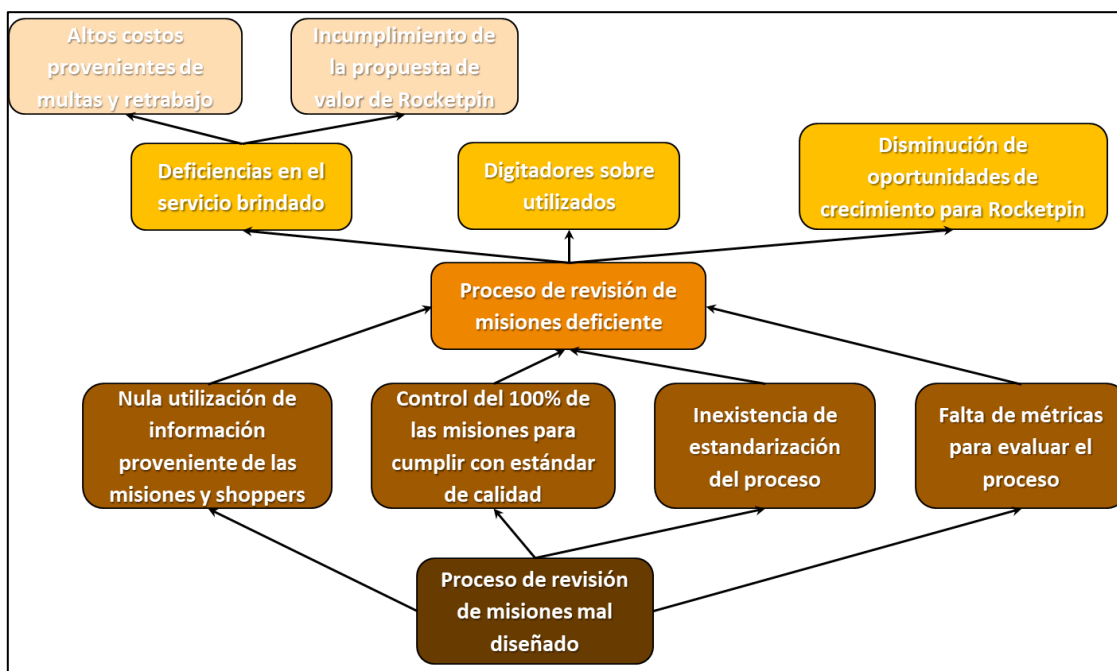


Ilustración 9: Desglose del problema estudiando mediante un árbol de problemas.  
Fuente: Elaboración Propia.

Una de las primeras soluciones que se pueden pensar para corregir las deficiencias en este proceso al leer esta memoria, correspondería a cambiar el formulario que deben llenar los shoppers para cumplir con las misiones de verificación de domicilio de EQUIFAX, sin embargo, esta solución de momento es infactible, puesto que por ordenanza de EQUIFAX ningún campo de este formulario puede ser modificado.

La solución más factible que surge por medio de la investigación y conversaciones con el área de control de calidad corresponde al rediseño del proceso de revisión de misiones, mediante la implementación de un modelo que genere como output un indicador o métrica que permita obtener información del posible grado de cumplimiento que tendrá la misión enviada por los shoppers, lo cual podría hacer más eficiente el control al no tener que revisar el 100% de las misiones. El input que se le entregaría al modelo podría basarse en los principales factores que influyen en la aprobación y desaprobación de las misiones, como, por ejemplo, la experiencia del shopper en Rocketpin. Este último factor es importante, ya que se esperaría que aquellos shoppers que llevan mayor tiempo en la aplicación, tengan una menor probabilidad de equivocarse, puesto que deberían tener internalizado el procedimiento estándar del servicio.

Otro factor importante que podría influir en la desaprobación de las misiones corresponde al tiempo que le toma al shopper realizarla. Por ejemplo, en promedio completar una misión de verificación de domicilio toma entre 5 a 10

minutos, dado esto es que cualquier misión que se realice en un tiempo mayor al señalado debería ser revisada de manera más minuciosa, ya que existen algunos usuarios que llenan la información de los formularios por medio de aplicaciones de geolocalización como *Google Maps Street View* (lo cual podría indicar por qué se demoran más en completar estas misiones). Existe otro factor que va muy ligado con el señalado anteriormente y corresponde a la distancia de donde el usuario inicia la misión versus donde la termina. Dado que las misiones están geolocalizadas y en algunos casos la dirección indicada en la aplicación no corresponde a la realidad, los shoppers tienen la facultad de mover el "pin" que indica su posición en el mapa para señalar de forma "correcta" donde en realidad debería estar localizada la misión. Sin embargo, han existido tanto casos de shoppers que cambian verídicamente la localización de la misión debido a que la dirección del domicilio no correspondía, como también existen casos de shoppers que modifican la ubicación para su propio beneficio, como, por ejemplo, haciendo que la misión este cerca de su casa y realizarla desde ahí utilizando *Google Maps Street View* para sacar las fotos de los domicilios.

Para finalizar, el rediseño del proceso de revisión supondría el principal cambio a realizar, y generaría alto valor en Rocketpin, ya que podría disminuir el tiempo que los digitadores utilizan para revisar las misiones, permitiría sentar las bases para la automatización de este proceso, haría factible un aumento del volumen de misiones de las cuales podría hacerse cargo Rocketpin y, por ende, es un factor crucial para la implementación de su estrategia.

## **4. OBJETIVOS**

### **4.1. Objetivo General**

Rediseñar el proceso de revisión de la veracidad y pertinencia de los datos obtenidos en las misiones que posee Rocketpin para asegurar la calidad de su servicio de forma menos costosa.

### **4.2. Objetivos Específicos**

1. Estimar los costos en los que incurre Rocketpin debido a las deficiencias del proceso.
2. Identificar las características de los shoppers y misiones que más influyen en la aprobación y rechazo de estas, mediante el uso de gráficas de control.

3. Diseñar indicadores que permitan controlar de forma más eficiente el proceso de revisión de misiones y faciliten asegurar la calidad del servicio.
4. Elaborar una propuesta de rediseño basado en la evaluación y estudio del proceso de revisión de misiones.
5. Evaluar el impacto potencial que tendría la implementación de este rediseño.

## **5. MARCO CONCEPTUAL**

En esta sección se presentan los tres principales enfoques desde los que se abordará la problemática estudiada, lo cual permitirá diseñar una solución que se ajuste a las necesidades de la organización. Los dominios expuestos a continuación van desde la conceptualización, esquematización y rediseño de procesos, pasando por los procedimientos necesarios para controlar procesos desde una mirada estadística, hasta el concepto de calidad, haciendo énfasis en los costos generados por un control deficiente de ésta.

### **5.1. Rediseño de procesos**

El problema que posee Rocketpin se basa principalmente en que no se están cumpliendo las metas de revisión de verificaciones domiciliarias, lo cual se origina debido a que el proceso actual es ineficiente y altamente costoso para la organización, ya que se está controlando el 100% de las misiones.

Una teoría ampliamente utilizada para hacer los procesos más eficientes y efectivos corresponde al rediseño de procesos. Juan Bravo expone que el rediseño de procesos surge de la necesidad de generar cambios en los procesos, por el simple hecho de que estos se van degradando posterior a su implementación y con el paso del tiempo van surgiendo nuevas metodologías, herramientas y/o tecnologías que se ajustan mejor a los requerimientos del mercado [7]. Es por esto que el rediseño de procesos está ligado a la innovación, ya que por medio de la mejora continua pretende revertir el deterioro de los procesos y lograr mantener los estándares de servicio, lo cual ayuda a las organizaciones a sobrevivir en un mercado cada vez más competitivo.

El rediseño permite identificar, modelar, diseñar, controlar, mejorar y hacer más productivos los procesos de la organización con el fin de ganar la confianza del cliente. El aumento de la productividad incluye la optimización de recursos y la eficacia para lograr los objetivos propuestos por la empresa.



Además, el rediseño de procesos debe estar alineado con la estrategia de la organización, teniendo como principal foco cumplir con las necesidades del cliente y satisfacer a las partes interesadas (comunidad, colaboraciones, dueños y otras). Lo anterior, proviene directamente de la definición de procesos expuesta por Juan Bravo [7, p. 17], la cual se plantea a continuación:

Un proceso es una competencia de la organización que le agrega valor al cliente y a las demás partes interesadas a través del trabajo en equipo de personas, en una secuencia organizada de actividades, interacciones, estructura y recursos que trasciende áreas. Esta definición cuenta de tres partes:

- La primera: "Un proceso es una competencia de la organización que le agrega valor al cliente y a las demás partes interesadas", esto se refiere a que el proceso se apalanca en las fortalezas de la organización, lo cual permite generar valor tanto al cliente como a las partes interesadas. Entender esto es sustancial, ya que cuando un proceso deja de generar valor al cliente y de ser sostenible para la organización, se hace necesario modificarlo, en beneficio de todos los actores involucrados.
- La segunda: "a través del trabajo en equipo de personas, en una secuencia organizada de actividades, interacciones, estructura y recursos", hace referencia a la esquematización de un proceso, enfatizando en que la serie de tareas que lo componen siguen un flujo estructurado y son realizadas por personas organizadas bajo cierta estructura y que manejan diversos recursos: infraestructura, tecnologías de apoyo, productos e información, entre otros. Además, es importante señalar que se entiende por actividad, el quehacer de una persona en un bloque de tiempo, pudiendo diferentes actividades interactuar entre sí (como, por ejemplo, actividades que requieren un documento hecho por otra persona para iniciarse).
- La tercera: "que trasciende a las áreas", sugiere que el proceso debe ser considerado como parte de un todo y no como una secuencia de pasos dependiente de una parte de la organización en específico. Los procesos realizados por una organización generalmente se cruzan horizontalmente, por lo que deben estar bien integrados entre ellos para asegurar un correcto funcionamiento.

Esta definición es clave para entender el funcionamiento de una organización, ya que la operación del negocio puede traducirse a una serie de procesos interrelacionados entre sí que tienen como objetivo generar valor al cliente. El entendimiento general de los procesos de Rocketpin es fundamental para

comprender el contexto en el que se enmarca el proceso deficiente, ya que permite saber su posición en la cadena de valor y visualizar los problemas que genera desde un punto de vista macro. Es por esto, que se hace necesaria la utilización de una notación estándar para modelar y visualizar procesos, ya que teniendo una visión gráfica es posible alinear la visión de diseño y ejecución, logrando que todas las partes involucradas tengan un mayor entendimiento de los flujos de información, las actividades realizadas e interacciones, permitiendo mejorar la productividad y calidad de los procesos en cuestión.

La notación escogida para modelar procesos corresponde a *Business Process Model Notation* (BPMN), la cual cuenta con elementos gráficos estándares que se definen a continuación [8]:

- **Objetos de flujos:** eventos, actividades, rombos de control de flujo (*gateways*).
- **Objetos de conexión:** flujo de secuencia, flujo de mensaje, asociación.
- **Carriles de nado (*swimlanes*):** piscina, carril, permiten diferenciar entre los actores encargados de realizar las actividades.
- **Artefactos:** objetos de datos, grupo y anotación.

La combinación de dichos elementos permite modelar los procesos poniéndose en todas las situaciones hipotéticas que puedan existir. El proceso se modela utilizando un *swimlane* que comprenda a todos los participantes del proceso y las tareas a ejecutar. Por otro lado, existen diferentes carriles para representar a los participantes del proceso e identificar así a los responsables de ejecutar cada una de las tareas en cuestión. Dentro de cada carril existen eventos (algo que sucede), tareas (algo que se ejecuta) o subprocesos (conjunto de tareas que son agrupadas para representar una tarea particular), que se conectan temporalmente mediante flechas de flujo y compuertas lógicas, que permiten modelar bifurcaciones de diferentes rutas posibles del proceso.

### 5.1.1. Métodos de rediseño

La contextualización del proceso deficiente es parte de la primera fase (investigar) del método expuesto por Juan Bravo [7, p. 22]. Aplicando el enfoque de proyectos, todo comienza por conocer bien el problema y así evitar soluciones que solo atacan el síntoma. Se trata de conocer el problema de fondo y luego buscar y evaluar soluciones en el espectro más amplio posible, para evitar la dependencia de una única solución. Una vez seleccionada la

solución, se propone una arquitectura que se refiera a la estrategia, procesos, estructura, tecnología y personas. Esta fase de investigación concluye con estudios económicos y un plan de proyecto.

Posteriormente, se profundiza en la fase de desarrollo del proyecto: diseñar, implementar y desplegar. Durante la vida útil de la solución se la mantiene y al concluir este plazo, nuevamente es necesario rediseñar, tres, cinco o diez años después, según la naturaleza del proceso rediseñado y de acuerdo con la estrategia actualizada. Las etapas del método de rediseño se pueden apreciar de mejor manera en la siguiente ilustración:

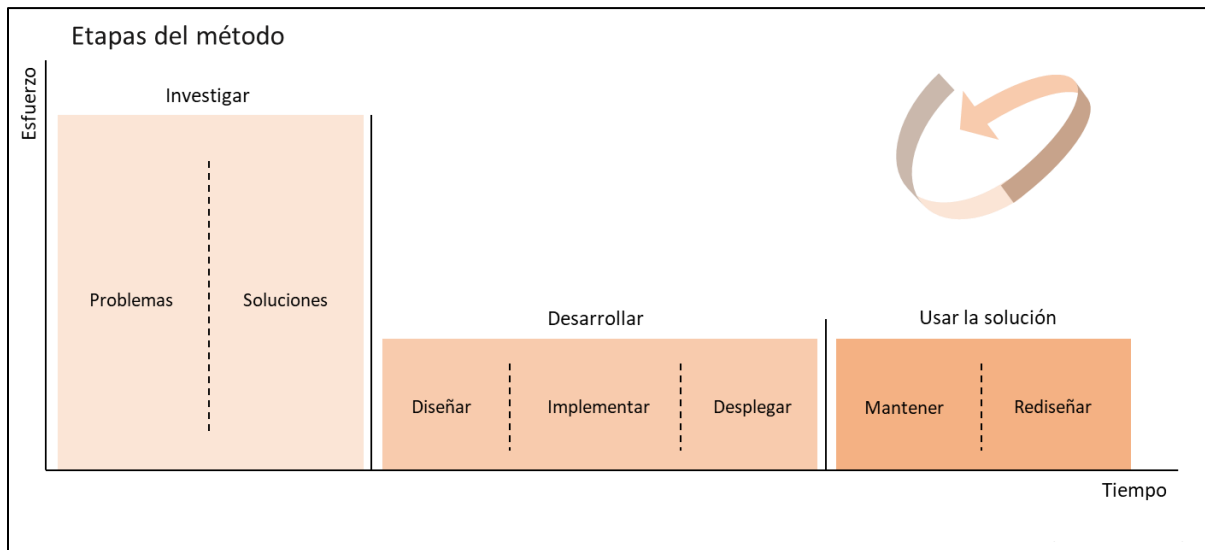


Ilustración 10: Visión de conjunto del método de rediseño de procesos.  
Fuente: Rediseño de Procesos (Juan Bravo, 2016).

La espiral presentada a la derecha de la figura indica que se debe trabajar de forma iterativa para favorecer la mejora continua de los procesos, ya que de esta forma se puede resolver en cada vuelta un porcentaje mayor de los requerimientos. Es importante notar que la primera fase del método corresponde a la que requiere mayor esfuerzo. En Latinoamérica se pierden anualmente cientos de millones de dólares debido a fallas evitables en la gestión de proyectos. Estas fallas generalmente se deben a que las organizaciones plantean una solución improvisada o que proviene de una investigación poco rigurosa.

Cada proyecto de rediseño consta de las mismas etapas (las cuales representan la subdivisión de las fases), siendo el alcance de cada una la que puede diferir según condiciones particulares. La descripción de cada etapa se presenta a continuación:

1. **Investigar problemas:** consiste en concebir el problema, el cual puede tomar diferentes formas: una necesidad o una dificultad específica, entre otras. Se entrega en la forma de un enunciado validado, con una variable crítica objetivamente evaluada y determinando sus causas. Se usa la palabra problema, aun cuando este se refiera a situaciones positivas, tal como expandir una buena práctica. Esta etapa requiere de amplios estudios y puede ser la más extensa.
2. **Investigar soluciones:** consiste en obtener una solución después de un barrido creativo de muchas alternativas y de un estudio detallado de algunas de ellas. La solución seleccionada tiene un grado de elaboración que llega hasta plantear la arquitectura para cuantificar y elaborar la carta Gantt del proyecto. Incluye un análisis de los requerimientos críticos para desarrollar la solución.
3. **Diseñar:** consiste en obtener el detalle de la solución completa, identificando la estrategia, procesos, estructura, tecnología y personas. El diseño contempla también como se mantiene y perfecciona la solución anterior, la que está vigente, hasta asegurar que la nueva propuesta es efectivamente mejor.
4. **Implementar:** consiste en llevar a la práctica la solución completa. Es fundamental para esta etapa evaluar que efectivamente la solución resuelve la causa del problema que origino su creación. Se concluye con una aplicación de carácter piloto de la solución, la cual se adecua a la realidad de la organización. La verificación puede incluso llevar a concluir que es mejor no desplegar la solución.
5. **Desplegar:** consiste en replicar o expandir la solución generada hasta ser bien utilizada por todos los usuarios previstos en el plan de proyecto. El cierre del proyecto concluye con un análisis de resultados. Un aspecto central de esta etapa es capitalizar las lecciones aprendidas llevando ese aprendizaje a los documentos y la práctica del método de gestión de proyectos de rediseño.
6. **Mantener:** consiste en mantener la solución en mejora continua y en buen funcionamiento hasta que cumpla su vida útil o sea reemplazada por otra solución.
7. **Rediseñar:** Concluye con el uso de la vida útil de la solución anterior con el inicio de un rediseño del proceso que proveerá otra solución. Es una espiral donde se van realizando sucesivos ciclos de rediseño hasta que la necesidad este superada o el problema completamente resuelto.

## 5.2. Procedimientos de control de procesos

Para las organizaciones es fundamental vigilar la calidad del proceso mientras se produce el producto o servicio, ya que este factor es primordial para satisfacer las necesidades del cliente y cumplir con la estrategia de la organización. En el caso de Rocketpin, se necesita utilizar indicadores que permitan asegurar la calidad del servicio brindado de forma menos costosa, lo cual puede hacerse al centrar la investigación en el control de proceso deficiente, en los problemas de línea, y no solo en el servicio final otorgado.

Es importante definir que si bien Rocketpin entrega un servicio a sus clientes, el cual consiste en el levantamiento de información *on-demand*, este se puede considerar como un conjunto de verificaciones domiciliarias. Esta distinción permite controlar el proceso a nivel "*producto*", ya que cada verificación domiciliaria posee características intrínsecas que pueden vigilarse, como, por ejemplo, el tiempo que le toma al shopper realizarla o ratio de calificación de este individuo.

Para vigilar el correcto funcionamiento del proceso, se plantea utilizar uno de los diez enfoques de la gestión de calidad (GC) que expone César Camisón en su libro *Gestión de la calidad: conceptos, enfoques, modelos y sistemas*, llamado control estadístico de procesos (CEP). La particularidad de este enfoque es que reconoce a cada proceso como variable, percatándose que existe variabilidad en productos y/o servicios que deberían ser "*idénticos*", la cual es consustancial a la producción, debido a causas aleatorias como diferencias en materias primas, problemas con los equipos o grados de destreza diferentes entre los trabajadores.

Dado esto, se entiende que no es posible estandarizar en un 100% los procesos, sin embargo, si se puede garantizar la estabilidad de este definiendo ciertos límites de tolerancia [9]. Para estimar y fijar estos límites de variación aceptable, el CEP propone dos principales enfoques estándar para controlar procesos, los cuales tienen la particularidad de dar como resultado final, la elaboración de gráficas de control. Los gráficos de control fueron diseñados como parte de técnicas estadísticas sencillas que permiten fijar los límites de variación aceptable y permiten distinguir entre la variación normal, producto de causas aleatorias, y la variación excesiva, cuyas causas pueden determinarse. Eliminar las causas que generan esta variabilidad excesiva permite asegurar la calidad del producto o servicio final, ya que es generalmente aceptado que a medida que disminuye la variabilidad, la calidad mejora y también se incurre en menores costos, ya que los productos o servicios que no cumplen con las especificaciones probablemente no podrán ser vendidos. Esto se ilustra en la siguiente ilustración:

A continuación, se describen los dos enfoques de control de procesos que propone el CEP, donde el primero se basa en la construcción de gráficas de control mediante medición por atributos y el segundo sobre la medición por variables.

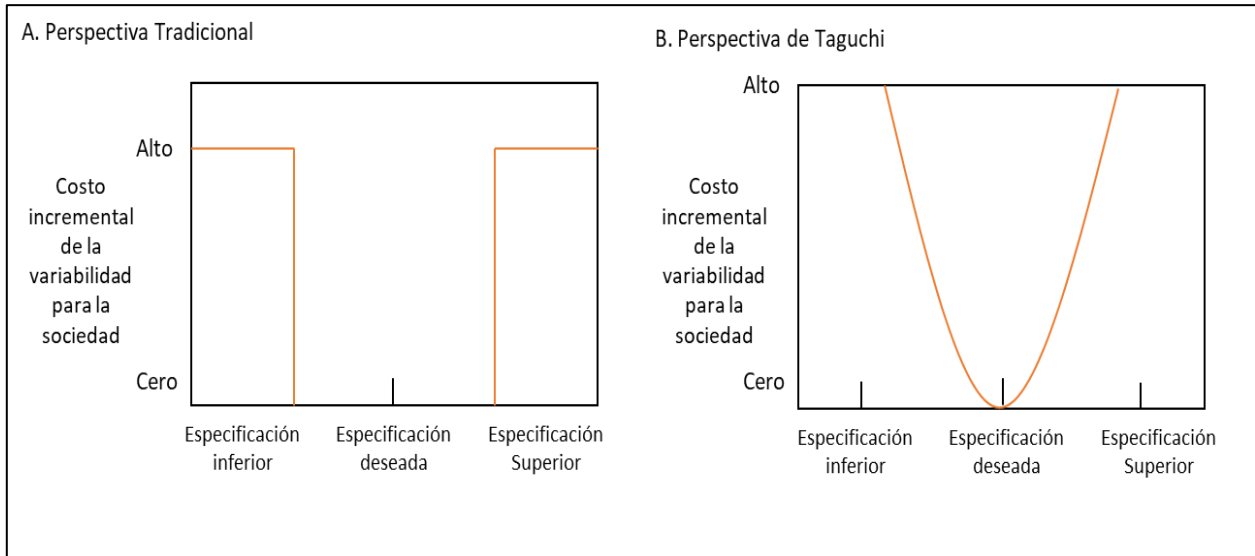


Ilustración 11: Perspectivas del costo de la variabilidad para la sociedad.  
Fuente: Gestión de la calidad (C. CAMISÓN, S. CRUZ y T. GONZÁLEZ, 2006).

### 5.2.1. Control de procesos por atributos: uso de gráficas P

Los atributos son características de calidad que se clasifican como que cumplen o no con la especificación deseada, distinguiendo entre productos o servicios bien realizados y los que son defectuosos. La medición por atributos significa tomar muestras y tomar una sola decisión: el artículo es bueno o malo. Como se trata de una decisión de sí o no, se utiliza la estadística simple para crear una gráfica de control P, con un límite de control superior (LCS) y un límite de control inferior (LCI). Se pueden trazar estos límites de control en una gráfica y luego representar la fracción de defectos de cada una de las muestras probadas. Se supone que el proceso funciona de manera correcta cuando las muestras, permanecen entre los límites de control.

$$p = \frac{\text{Número total de defectos de todas las muestras}}{\text{Número de muestras} \times \text{Tamaño de la muestra}}$$

$$\theta_p = \sqrt{\frac{(1-p)p}{n}}$$

$$LCS = p + z * \theta_p$$

$$LCI = p - z * \theta_p$$

Donde p es la fracción defectuosa,  $\theta_p$  es la desviación estándar, n es el tamaño de la muestra y z es el número de desviaciones estándar para una confianza específica. Por lo general, se usa z=3 (99,7% de confianza) o z=2.58 (99% de confianza). Es posible apreciar un ejemplo de un gráfico de control realizado con este enfoque a continuación:

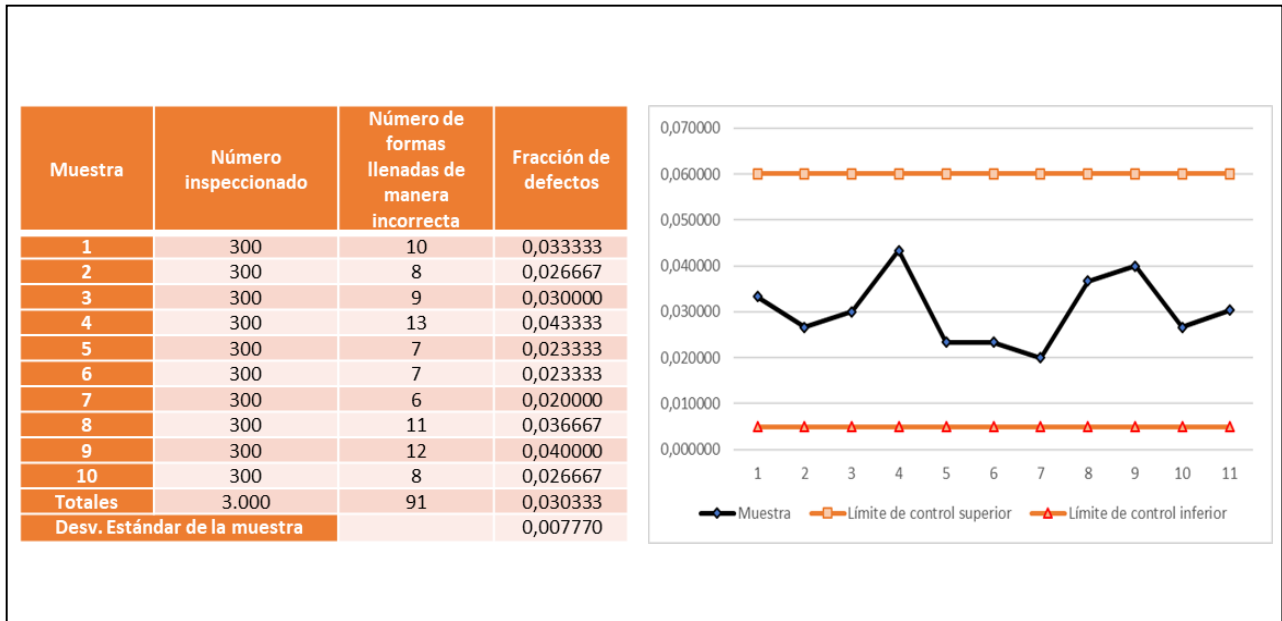


Ilustración 12: Ejemplo de elaboración de un gráfico de control P en base a 10 muestras de 300 formas de cobro.

Fuente: Administración de Operaciones (CHASE, R., JACOBS, R. y AQUILANO, N. 2009).

El tamaño de la muestra debe ser suficientemente grande para permitir el conteo del atributo. Por ejemplo, si una máquina produce 1% de defectos, una muestra de cinco unidades pocas veces capturaría un defecto. Una regla práctica al crear una gráfica p es hacer que la muestra tenga el tamaño suficiente para esperar contar el atributo dos veces en cada muestra. De modo que, si el índice aproximado de defectos es 1%, el tamaño de la muestra apropiado sería de 200 unidades. En la ilustración 70 del anexo de este informe, se pueden apreciar los casos en que es necesario ahondar más en el proceso según el comportamiento de los gráficos de control [10].

### 5.2.2. Control de procesos por variables: uso de gráficas X y R

Las gráficas X y R (de rango) se utilizan con frecuencia en el control estadístico del proceso. En la medición de variables, se mide el peso, volumen, número de pulgadas o cualquier otra variable real, y se desarrollan gráficas de control

para determinar el grado de aceptación o rechazo del proceso, con base en esas mediciones. Para crear estas gráficas de control se debe considerar 4 principales aspectos: tamaño de las muestras, el número de muestras, la frecuencia de las muestras y los límites de control.

E. L. Grant y R. Leavenworth calcularon una tabla (ilustración 71 del anexo de este informe) que permite calcular con facilidad los límites de control superior e inferior tanto para la gráfica X como R. Las ecuaciones son las siguientes [10]:

$$LCS \text{ para gráfica } X = \bar{\bar{X}} + A_2 * \bar{R}$$

$$LCI \text{ para gráfica } X = \bar{\bar{X}} - A_2 * \bar{R}$$

$$LCS \text{ para gráfica } R = D_4 * \bar{R}$$

$$LCI \text{ para gráfica } R = D_3 * \bar{R}$$

Donde:

$$\bar{\bar{X}} = \textit{Promedio de las medias de la muestra}$$

$$\bar{R} = \textit{Promedio de las resta entre el mínimo y máximo valor de cada muestra}$$

### 5.2.3. Modelos de clasificación

Dentro del mundo del aprendizaje de máquinas (*machine learning*), existen distintos tipos de modelos y/o algoritmos que se utilizan para encontrar patrones en los datos, segmentar clientes en base a sus características o predecir la fuga de usuarios en un determinado período de tiempo [11].

Mediante el empleo de técnicas de aprendizaje automático, se puede capturar la presencia de patrones de comportamiento en base a los factores que mayor relación tiene con la aprobación o desaprobación de las misiones de verificación domiciliaria, usando todos los hallazgos obtenidos en las secciones de diagnóstico de la situación actual, análisis de bases de datos y gráficas de control. Antes de iniciar el análisis es recomendable hacer una distinción entre los tipos de aprendizaje automático que existen en el mercado actualmente:

- a) Aprendizaje supervisado:** la primera modalidad de aprendizaje que tiene el *machine learning* es la de aprendizaje supervisado. Corresponde a un conjunto de técnicas que permite realizar predicciones futuras basadas en comportamientos o características analizadas en datos



históricos etiquetados [12]. La etiqueta corresponde a la clasificación entregada a la observación, la cual está previamente definida. Por ejemplo, la etiqueta para un paciente podría ser que este tuviera o no cáncer. Existen variables que permiten diferenciar si es que el paciente tiene o no cáncer, con cierto grado de certeza como, por ejemplo: niveles de glóbulos blancos en la sangre, signos vitales, porcentaje de masa corporal, entre otras. Para este caso, un algoritmo de aprendizaje supervisado podría etiquetar a un paciente en base a sus características, utilizando como input, información histórica de personas internadas que ya fueron previamente clasificadas como pacientes con cáncer o no. Esto implicaría que el algoritmo "aprende" a clasificar utilizando la información pasada y puede predecir en base a eso, cual es la etiqueta que con mayor probabilidad tendrá el paciente.

Existen dos tipos de aprendizaje supervisado:

- **Regresión:** tiene como resultado un número específico. Si las etiquetas suelen ser un valor numérico, mediante las variables de las características, se pueden obtener dígitos como dato resultante.
- **Clasificación:** en este tipo, el algoritmo encuentra diferentes patrones y tiene por objetivo clasificar los elementos en diferentes grupos [13]. Dentro de la subcategoría del aprendizaje supervisado que es la clasificación, existen dos diferentes tipos:
  - **Clasificación Binaria:** es un tipo de clasificación en el que tan solo se pueden asignar dos clases diferentes (0 o 1). El ejemplo sobre el paciente con cáncer o no corresponde a este tipo de clasificación.
  - **Clasificación Multiclase:** se pueden asignar múltiples categorías a las observaciones. Como el reconocimiento de caracteres de escritura manual de números (en el que las clases van de 0 a 9) [14].

**b) Aprendizaje no supervisado:** a diferencia del aprendizaje supervisado, en el no supervisado solo se le otorgan las características, sin proporcionarle al algoritmo ninguna etiqueta. Su función principal es la agrupación, el cual es el proceso de particionar un conjunto de datos (u objetos) en un conjunto de subclases significativas llamadas grupos (*clusters*). Un grupo es una colección de objetos de datos que son similares a otros y así pueden ser tratados colectivamente como un

grupo. El agrupamiento es una forma de clasificación no supervisada en la que no se conocen las etiquetas de las clases (no hay clases predefinidas) y puede que tampoco se conozca el número de grupos. Un ejemplo claro de este tipo de algoritmos sería la segmentación de clientes en base a su comportamiento de compra, ya que no se tiene información previa acerca de la etiqueta o clase del cliente, pero se pueden establecer similitudes entre ellos en base a sus características y patrones. Es importante destacar que en estos casos no existen métricas ni formas de analizar si la segmentación realizada es la mejor o la correcta, ya que existen infinitos diferentes resultados que podrían ajustarse a las necesidades del negocio que solicita el análisis [15].

El caso presentado en esta memoria correspondería a un algoritmo de aprendizaje supervisado, donde las observaciones serían las misiones y la etiqueta correspondería a si está rechazada o no (0 si está aprobada, 1 sino). La clasificación sería del tipo binaria, ya que solo existen dos clases diferentes. El algoritmo permitirá determinar que variables son las más influyentes a la hora de definir si una misión está rechazada o no y dependiendo de su desempeño, se podría predecir en base a las características de la misión, a que categoría debería pertenecer ésta. El rendimiento del algoritmo dependerá netamente de su calibración, programación y si es que las variables utilizadas como input permiten visualizar diferencias entre las categorías.

#### **5.2.4. Regresión logística**

En estadística, la regresión logística es un tipo de análisis de regresión utilizado para predecir el resultado de una variable categórica (una variable que puede adoptar un número limitado de categorías) en función de las variables independientes o predictoras. Es útil para modelar la probabilidad de un evento ocurriendo como función de otros factores. El análisis de regresión logística se enmarca en el conjunto de Modelos Lineales Generalizados (GLM por sus siglas en inglés) que usa como función de enlace la función *logit*. Las probabilidades que describen el posible resultado de un único ensayo se modelan, como una función de variables explicativas, utilizando una función logística.

La regresión logística unidimensional puede usarse para tratar de correlacionar la probabilidad de una variable cualitativa binaria (asumiremos que puede tomar los valores reales "0" y "1") con una variable escalar  $x$ . La idea es que la regresión logística aproxime la probabilidad de obtener "0" (no ocurre cierto suceso) o "1" (ocurre el suceso) con el valor de la variable explicativa  $x$ . En esas condiciones, la probabilidad aproximada del suceso se aproximará mediante una función logística del tipo [16]:

$$\pi(x) = \frac{e^{(\beta_0 + \beta_1 x)}}{e^{(\beta_0 + \beta_1 x)} + 1} = \frac{1}{e^{-(\beta_0 + \beta_1 x)} + 1}$$

que puede reducirse al cálculo de una regresión lineal para la función logit de la probabilidad:

$$g(x) = \ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \beta_0 + \beta_1 x$$

El gráfico de la función logística se muestra a continuación, donde la variable independiente es la combinación lineal  $\beta_0 + \beta_1 x$  y la variable dependiente es la probabilidad estimada  $\pi(x)$ :

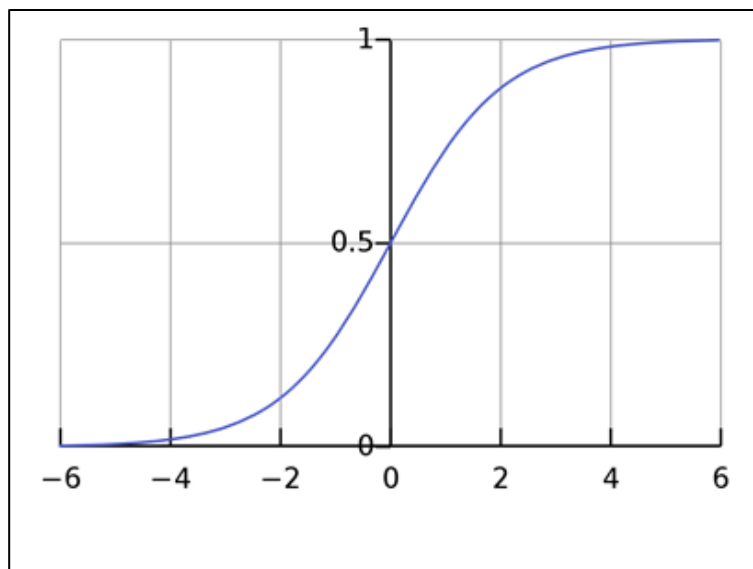


Ilustración 13: Representación de la función logística.

Fuente: Elaboración propia.

Para el problema abordado en esta memoria, la utilización de este algoritmo será de particular ayuda para asignar una probabilidad de aprobación a las misiones, dependiendo de sus características y del perfil del shopper que la realiza. Esta probabilidad puede ser usada como un atributo o variable de las misiones a validar por el proceso de revisión, por lo que se puede definir un rango de aceptación de su variabilidad, utilizando gráficas P, X y R. Además, la particularidad del gráfico de control generado a partir del output de la regresión logística permitirá sentar las bases para modificar la forma en que se controla, ya que se pueden determinar niveles de exhaustividad en la revisión dependiendo de la probabilidad de aprobación de la misión.

### 5.3.CALIDAD COMO CONCEPTO ESTADÍSTICO

El control estadístico de procesos indica que un producto o servicio tiene calidad cuando sus características están dentro de los límites de tolerancia superior e inferior de las especificaciones. Dado lo anterior, es necesario considerar una definición del concepto calidad que sea coherente con la visión de procesos que se tiene del problema estudiando, ya que considerar la calidad desde todas sus posibles aristas, entorpecería el diseño de una solución adecuada para la organización. Es por esto que, se considera la definición aportada por Wickoff (1984): "*calidad es el grado de excelencia prometido y el control de la variabilidad en el alcance de esa excelencia en respuesta a los requerimientos de los consumidores*".

Esta filosofía de calidad permite comprender que la estadística es importante, pero lo crucial es un pensamiento estadístico, que oponga el concepto de variedad al de calidad, lo cual puede apreciarse de mejor manera en la siguiente ilustración:

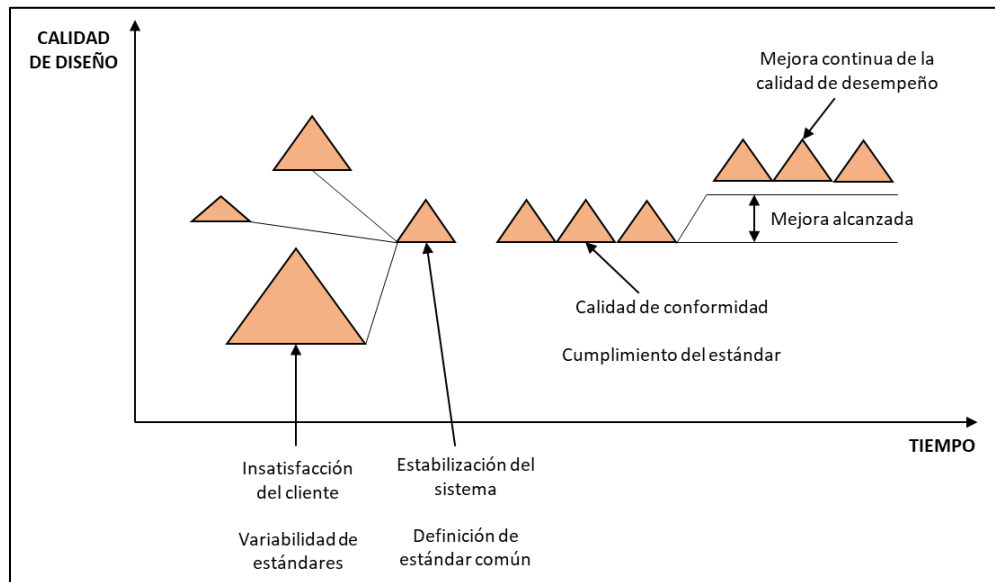


Ilustración 14: Concepto estadístico de la calidad.

Fuente: Gestión de la calidad (C. CAMISÓN, S. CRUZ y T. GONZÁLEZ, 2006).

El concepto estadístico de la calidad como uniformidad se centra en los aspectos internos de la producción, enfocándose en lograr productos libres de errores que satisfagan con precisión las metas de las especificaciones de diseño. El concepto estadístico de la calidad gira entonces alrededor de los procesos. La calidad es consecuencia de reducir la variabilidad de los comportamientos en los procesos: no se puede garantizar una calidad uniforme y su mejora continua sin disminuir la variabilidad de las características del producto y el servicio [9, p. 157].

### 5.3.1. Costos de calidad

Una de las formas en las que se puede entender el impacto económico que genera un proceso deficiente corresponde a la estimación de los costos de calidad. Esto es fundamental, ya que aun cuando pocos pueden discrepar sobre la noción de prevención, la gerencia a menudo necesita cifras sólidas para determinar el costo de las actividades de prevención. Desde el punto de vista purista, significa todos los costos atribuibles a la producción de calidad que no es 100% perfecta. Una definición menos estricta sólo considera los costos que son la diferencia entre lo que se espera de un desempeño excelente y los costos actuales. Se estima que entre el 15 y 20% de cada dólar de venta, corresponde a costos de retrabajo, desperdicio, servicio repetido, pruebas, garantías y otros elementos relacionados a la calidad. Los expertos establecen que el costo correcto de un programa de gestión de calidad bien dirigido debe ser inferior a 2,5%. Hay 3 suposiciones básicas que justifican un análisis de costos de calidad: 1) las fallas son provocadas, 2) la prevención es más barata y 3) se puede medir el desempeño. Generalmente, los costos de calidad son de 4 tipos (en la ilustración 73 del anexo de este informe puede apreciarse un informe de costos de calidad por categoría):

- 1) Costos de evaluación:** Los costos de la inspección, pruebas y demás tareas que garantizan que el producto o proceso sea aceptable.
- 2) Costos de prevención:** La suma de todos los costos para prevenir defectos, como los costos de identificar la causa del defecto, poner en práctica la medida correctiva para eliminar la causa, capacitar al personal, rediseñar el producto o sistema y comprar equipo nuevo o realizar modificaciones.
- 3) Costos de falla interna:** costos por defecto en los que se incurrió dentro del sistema: desperdicio, retrabajo y/o reparación.
- 4) Costos de falla externa:** Costos por defectos que pasan el sistema: reemplazos por garantía al cliente, pérdida de los clientes, manejo de quejas y reparación del producto [10, p. 311].

## 6. METODOLOGÍA

Para abordar la problemática señalada en la sección de planteamiento del problema y justificación del presente informe, se dividirá el trabajo a realizar en 5 etapas que se basan principalmente en la metodología propuesta por Juan Bravo en su libro "*Rediseño de Procesos*", poniendo énfasis en el control de los procesos y en el aseguramiento de la calidad del servicio. Las tres primeras

etapas de la metodología presentada se engloban en la fase de investigación del método propuesto por Juan Bravo, y las últimas 2 en la fase de desarrollo. Las etapas en las que se abordará el trabajo de memoria son definidas a continuación:

## **6.1. ANÁLISIS DE LA SITUACIÓN ACTUAL**

Una vez se tenga claridad de los conceptos y marcos conceptuales que definen el tema propuesto por el estudiante, será necesario realizar un análisis de la situación actual de Rocketpin respecto al proceso de revisión de misiones. Para realizar esto, en primer lugar, será necesario entender el funcionamiento de la organización, comprender la estrategia (misión, visión y objetivos a largo plazo), conocer todos los macroprocesos involucrados y analizar cómo se sitúa la revisión de misiones en este contexto general, para de esta forma medir cuantitativa y cualitativamente cómo aporta este proceso a la creación de valor de cara al cliente.

Será necesario levantar información respecto al proceso de revisión de misiones, para conocer los subprocesos que lo componen, saber cuáles son los actores involucrados en la realización de cada tarea. Para esto será fundamental entrevistar a los actores involucrados en este proceso, utilizando particularmente preguntas abiertas, ya que estas permiten simplificar la comunicación con el entrevistado, proporciona una gran riqueza de detalles y revela nuevas alternativas sobre preguntas no consideradas. Los principales objetivos de las entrevistas al personal del área de revisión de misiones será saber con qué métricas se está analizando su productividad, conocer sus testimonios respecto al ambiente de trabajo y condiciones laborales, además de generar un entendimiento general de su participación en el proceso de revisión de misiones. Se espera que las entrevistas permitan comprender los principales indicadores cualitativos en los que se basa el personal del área de control de calidad para desaprobar misiones. Esto será fundamental, puesto que, si se quiere generar algún KPI o métrica que permita facilitar el proceso de revisión basado en las características de los shoppers que más influyen en la aprobación de las misiones, será clave que este indicador sea intuitivo y fácil de interpretar, ya que de no serlo podría generarse aversión al cambio y hacer que el trabajo de memoria realizado no sea efectivo.

Se modelará la situación actual del proceso mediante la herramienta BPMN (*Business Process Model and Notation*), la cual corresponde a una notación gráfica estandarizada que permite la descripción de procesos de negocios, en un formato de flujo de trabajo (*workflow*). La modelación se realizará considerando las reglas estandarizadas sobre el diseño de flujos de trabajo y se realizará utilizando el software Bizagi.

Además, en esta etapa se debe definir la variable crítica o un conjunto de ellas que se utilizará para medir el desempeño del proceso, la cual además servirá para evaluar el éxito o fracaso de la solución propuesta. Un ejemplo que surge en base a la investigación actual podría sugerir como variables críticas, las siguientes: la tasa de misiones aprobadas por el área de control de calidad que son rechazadas por las otras empresas y la cantidad promedio de misiones revisadas por día. La primera variable permite supervisar el estándar de calidad fijado para el servicio (95% misiones correctamente aprobadas) y la segunda permite controlar el desempeño de los digitadores en base a las metas propuestas. Finalmente, se espera poder identificar los principales dolores que genera este proceso a la organización y a los actores que lo componen, para de esta forma medir de forma cualitativa y cuantitativa los efectos producidos por las ineficiencias encontradas.

## **6.2. ANÁLISIS EXPLORATORIO DE LAS BASES DE DATOS**

Debido a que el proceso de revisión en estudio se sustenta, en las misiones que los shoppers realizan para Rocketpin, será fundamental tener una idea clara de su comportamiento, caracterizarlos y conocer más a fondo su participación en este proceso. Para esto, será clave utilizar la base de datos de shoppers y misiones que Rocketpin posee (la cual se definió en la sección de planteamiento de la problemática y justificación). Con el análisis exploratorio se espera identificar las características que poseen los shoppers y las misiones que son más influyentes en la aprobación y desaprobación de estas últimas. Se espera que los resultados obtenidos estén altamente correlacionados y sean consecuentes con las ideas generales que el área de control de calidad tiene respecto a las características de los usuarios y misiones que más influyen en la desaprobación.

Además, esta etapa abarca el estudio de estadísticos descriptivos y gráficos de dispersión de las variables relativas a shoppers y misiones, con el fin de generar una mayor comprensión de su comportamiento. La información obtenida será fundamental para la etapa de rediseño, ya que permitirá presentar los lineamientos generales que se deberán utilizar para la implementación de un modelo y/o generación de indicadores que permitan facilitar el proceso de revisión al personal del área (en otras palabras, el posible cambio a implementar).

## **6.3. PROCEDIMIENTOS DE CONTROL**

Se deben estudiar las fuentes de variabilidad provenientes del proceso de revisión de misiones, ya que como se ha mencionado anteriormente, una alta

variabilidad induce a una calidad deficiente. Para esto, se utilizarán herramientas provenientes del control estadístico de procesos, correspondientes a gráficas de control.

Se realizarán gráficas de control tipo X y R para las variables más influyentes en la aprobación de misiones, como podrían ser, por ejemplo, el tiempo total y la distancia de inicio versus la de término. Esta herramienta permitirá establecer los límites de variación aceptadas para cada variable, los que representan información valiosa para el diseño de la solución al problema encontrado en la revisión de misiones, ya que entrega las bases para estandarizar este proceso y hacer más eficiente supervisar la calidad del servicio.

Además, se utilizarán algoritmos de clasificación para modelar la probabilidad de aprobación de las misiones en base a sus características que posean mayor variabilidad y tengan mayor influencia en este resultado. A la probabilidad resultante se le puede aplicar las gráficas de control tipo X y R, lo cual permite conocer el límite inferior y superior de la probabilidad de aprobación o rechazo.

#### **6.4. REDISEÑO DEL PROCESO ACTUAL**

Se definirán las direcciones de cambio que establecerán las diferencias entre el proceso de revisión de misiones que existe actualmente y el rediseño propuesto. Será necesario buscar cambios a implementar en el proceso de revisión que se sustenten en los estudios obtenidos de las etapas anteriores, priorizando aquellos que sean más factibles de diseñar e implementar en la organización. Para el diseño de los cambios se debe establecer la estrategia para su implementación, la tecnología y estructura a utilizar, las personas que implementarán la solución y como esta última, podría ajustarse a otros servicios provistos por Rocketpin.

Posteriormente se debe modelar el rediseño, lo cual consiste en realizar una representación de los nuevos procesos que implementarán el cambio establecido anteriormente.

Finalmente, se deberán utilizar las variables críticas escogidas o diseñar indicadores que permitan medir el desempeño en caso de que se implemente el cambio en la organización, definiendo los límites en los cuales el rendimiento del proceso será aceptable.



## **6.5.EVALUACIÓN DE LOS CAMBIOS PROPUESTOS**

Se realizará una evaluación del impacto potencial que tendría la implementación de los cambios propuestos en la etapa de rediseño, mediante una evaluación económica. Identificando los costos iniciales o de inversión, los costos recurrentes y los beneficios que traería consigo la implementación del proyecto. Para medir el beneficio potencial, se estimarán los flujos de caja de la empresa con y sin proyecto, lo cual permitirá estimar indicadores financieros que den indicios de su factibilidad (VAN o TIR).

Finalmente, se deberán medir las potenciales disminuciones en los tiempos y costos de la revisión utilizando simulaciones o mediante la implementación de un piloto de la solución, y además se podrían utilizar métodos cualitativos para validar que el rediseño propuesto sería beneficioso para todos los actores involucrados.

## **7. DIAGNÓSTICO DE LA SITUACIÓN ACTUAL**

### **7.1.Aplicación de Rocketpin**

Rocketpin es una *Start-Up* tecnológica cuyo principal servicio consiste en el levantamiento de información en terreno por medio de shoppers, personas que subutilizan su tiempo y quieren rentabilizarlo. Para formar parte de la comunidad Rocketpin, los interesados deben descargar su aplicación móvil, la cual está disponible para dispositivos con sistemas operativos Android e iOS. Posterior a la descarga de la aplicación móvil, el potencial shopper debe indicar algunos datos personales, como fecha de nacimiento, género, domicilio, entre otras. Las funcionalidades más importantes de la aplicación pueden apreciarse en las ilustraciones a continuación:

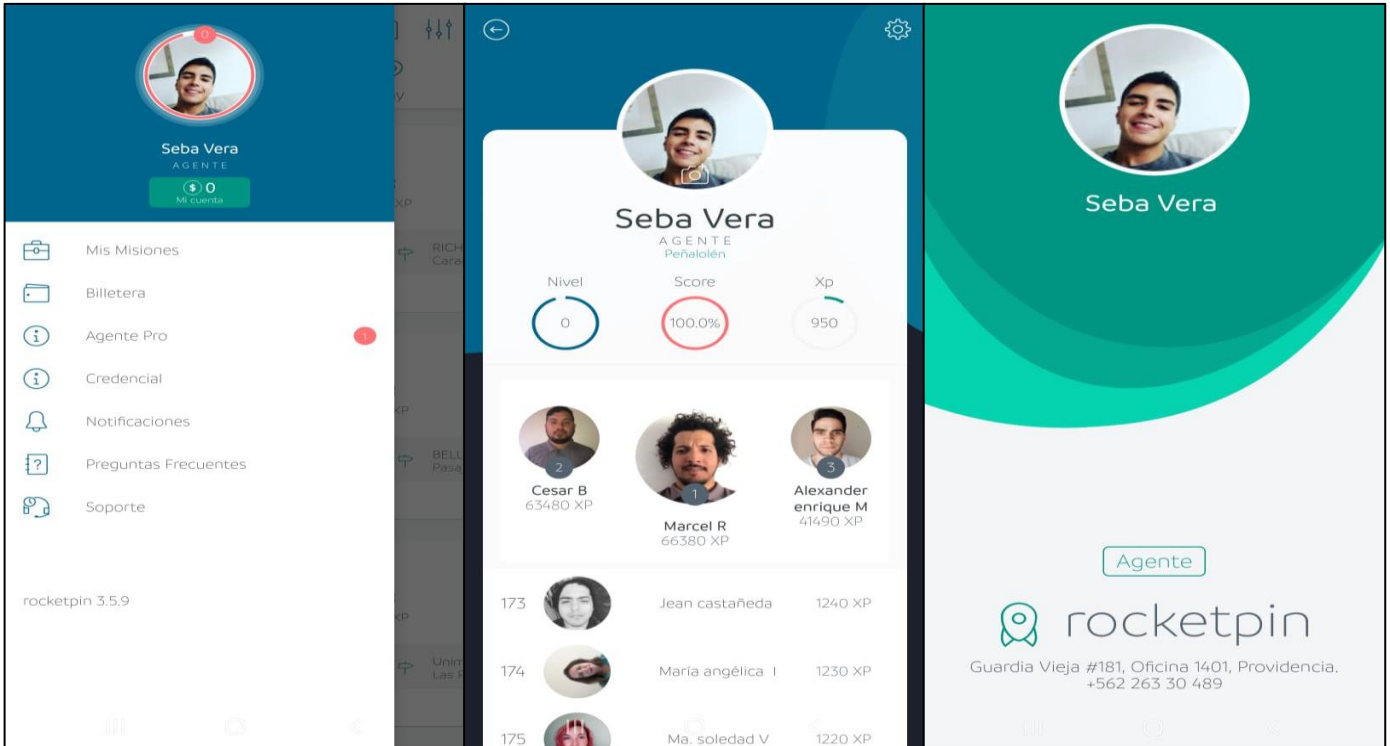


Ilustración 15: Perfil del shopper en la aplicación Rocketpin.  
Fuente: Aplicación móvil de Rocketpin.

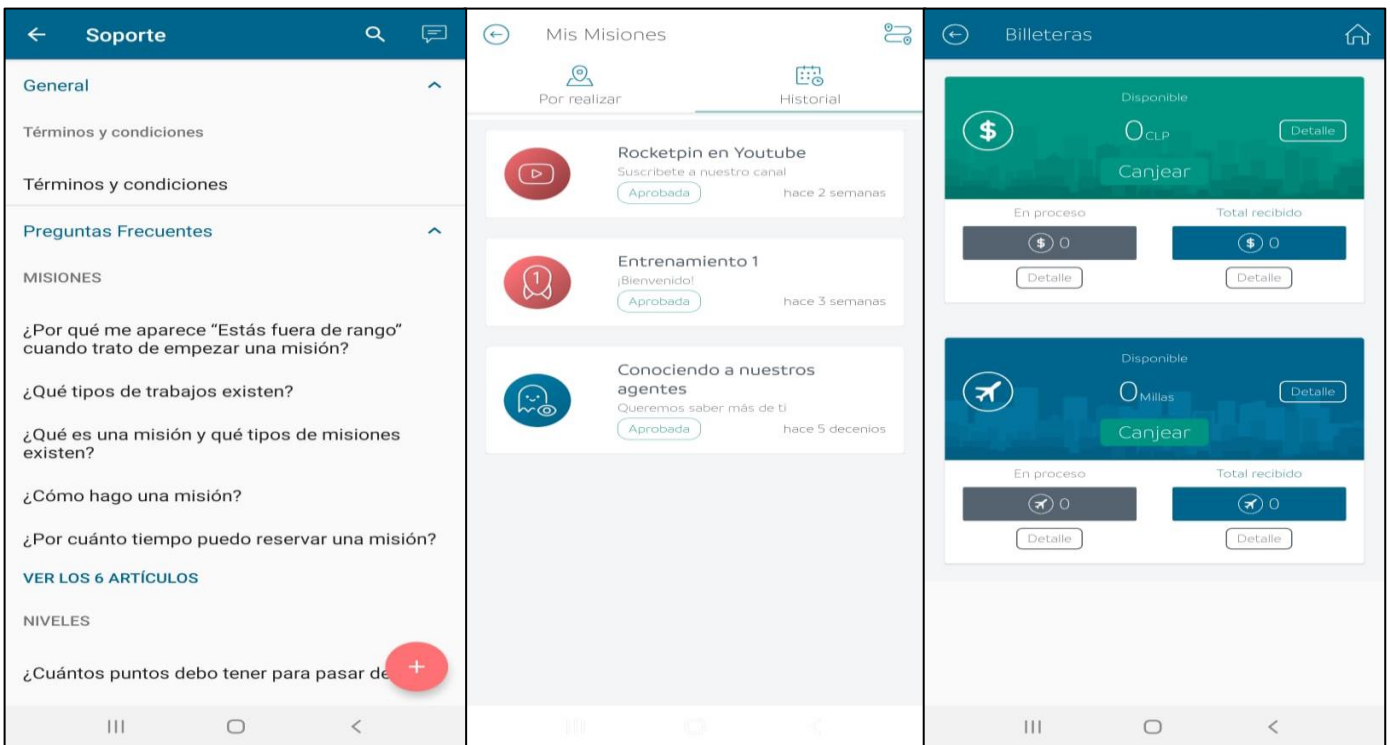


Ilustración 16: Funcionalidades de la aplicación Rocketpin.  
Fuente: Aplicación móvil de Rocketpin.

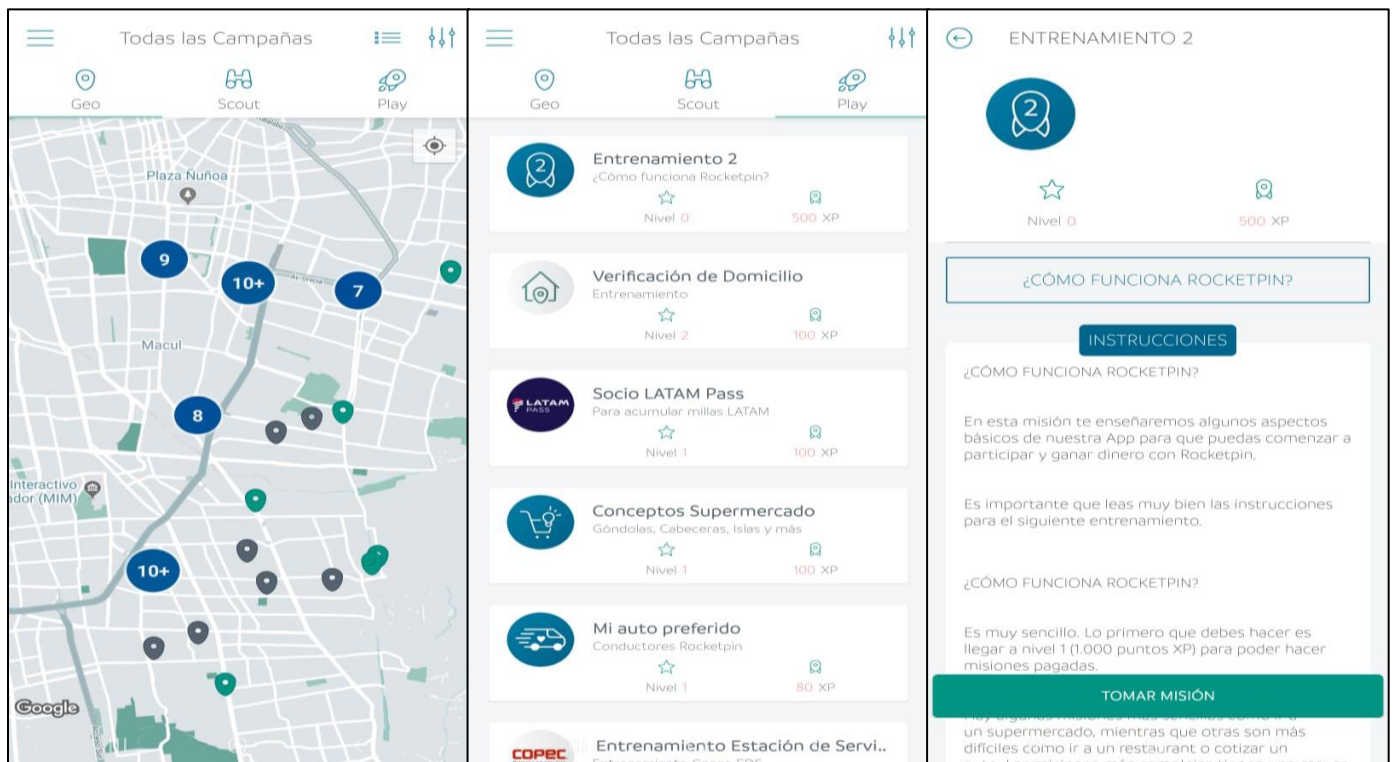


Ilustración 17: Vistas de las misiones en la aplicación Rocketpin.  
Fuente: Aplicación móvil de Rocketpin.

En la primera ilustración, se muestra el perfil que genera la aplicación una vez que el shopper ha ingresado sus datos personales. Rocketpin es una aplicación en la que a medida que se realicen más misiones se pueden acumular más puntos de experiencia, lo cual permite a los shoppers acceder a misiones que otorgan mayor remuneración y/o beneficios. Los indicadores presentes en el perfil permiten identificar el nivel del shopper (y, por ende, su compromiso con la empresa), la experiencia que ha ganado en la semana y su porcentaje de cumplimiento o *Score*. Esta última métrica corresponde a un indicador de "confiabilidad", es decir, si es que el shopper cumple a cabalidad con los trabajos que se compromete. Para mejorar este indicador, los shoppers deben ser responsables y cumplir con todas las misiones que reservan en los plazos y estándares óptimos.

Las variables que la empresa analiza para la generación de ese *Score* son las siguientes:

- a) **Protocolo de presentación:** *cumples a cabalidad con el protocolo de salud al momento de realizar tus misiones, independiente de la marca, adecuas la comunicación respetando el speech que te entrega Rocketpin, saludas amablemente, utilizas un vocabulario respetuoso y formal.*

- b) **Calidad de fotografías:** *las fotografías de tus misiones son de excelente calidad no solo en cuanto a la resolución de éstas, sino que también te preocupas de tomarlas desde los ángulos que te solicitamos, respetando enfoques y planos abiertos.*
- c) **Utilidad del comentario:** *las observaciones generales que registras en cada misión son coherentes con las respuestas que seleccionas en los formularios, tienen relación con el levantamiento de información que estás realizando y son comentarios útiles que aportan valor a la marca.*
- d) **Ortografía y redacción:** *te preocupas de enviar tus misiones a proceso de revisión velando por la buena ortografía, además redactas tus comentarios de manera clara y elocuente.*

En la ilustración 16, se indican algunas funcionalidades que tiene la aplicación, como la sección de preguntas frecuentes, la “billetera”, donde los shopper pueden ver la remuneración que han acumulado y el historial de las misiones que han realizado con su correspondiente estado de revisión, que puede ser *Aprobado, Rechazado o En Espera.*

La ilustración 17 presenta las vistas de cómo se realizan las misiones en la aplicación, la primera muestra las misiones geolocalizadas en un mapa tomando como referencia la ubicación del shopper (por ende, son las misiones más cercanas), en la segunda vista se puede apreciar la descripción de las misiones cercanas, donde se señala el nivel necesario para tomar la misión, el pago y los puntos de experiencia que se otorgarán una vez aprobada.

## **7.2. Macroprocesos del negocio**

Existen 3 procesos clave que debe llevar a cabo la empresa para que el negocio opere de forma correcta y entregue su servicio con el estándar que lo diferencia de la competencia. Éstos son:

- a) Capacitación de los shoppers
- b) Realización de la misión por parte de los shoppers
- c) Revisión de la misión por el área de control de calidad

Para comprender la situación actual de la empresa y situar el problema encontrado en el proceso de revisión de misiones, es necesario tener claridad respecto a la importancia del proceso en la operación del negocio y saber como

este se conecta con otros, por lo que se describirá de manera general el proceso de capacitación y realización de misiones, para posteriormente poner un mayor énfasis en el proceso a estudiar.

### 7.2.1. Capacitación de shoppers

Para la mayoría de los servicios que provee Rocketpin, la información levantada en terreno será utilizada por un tercero, como, por ejemplo, ejecutivos bancarios, por lo que es bastante probable que estos últimos se percaten si es que la información provista por Rocketpin es errónea. Dado esto es que, para mantener su imagen y estándar de servicio, Rocketpin debe asegurar que sus shoppers estén correctamente capacitados y tengan los conocimientos necesarios para cumplir con la calidad del servicio comprometida.

En primer lugar, cuando un shopper descarga la aplicación inicia en nivel 0, con 0 puntos de experiencia. Para comenzar a realizar misiones remuneradas, todos los nuevos usuarios de la aplicación de Rocketpin deben aprobar las misiones de entrenamiento y subir de nivel. Lo anterior puede apreciarse en las siguientes vistas de la aplicación:

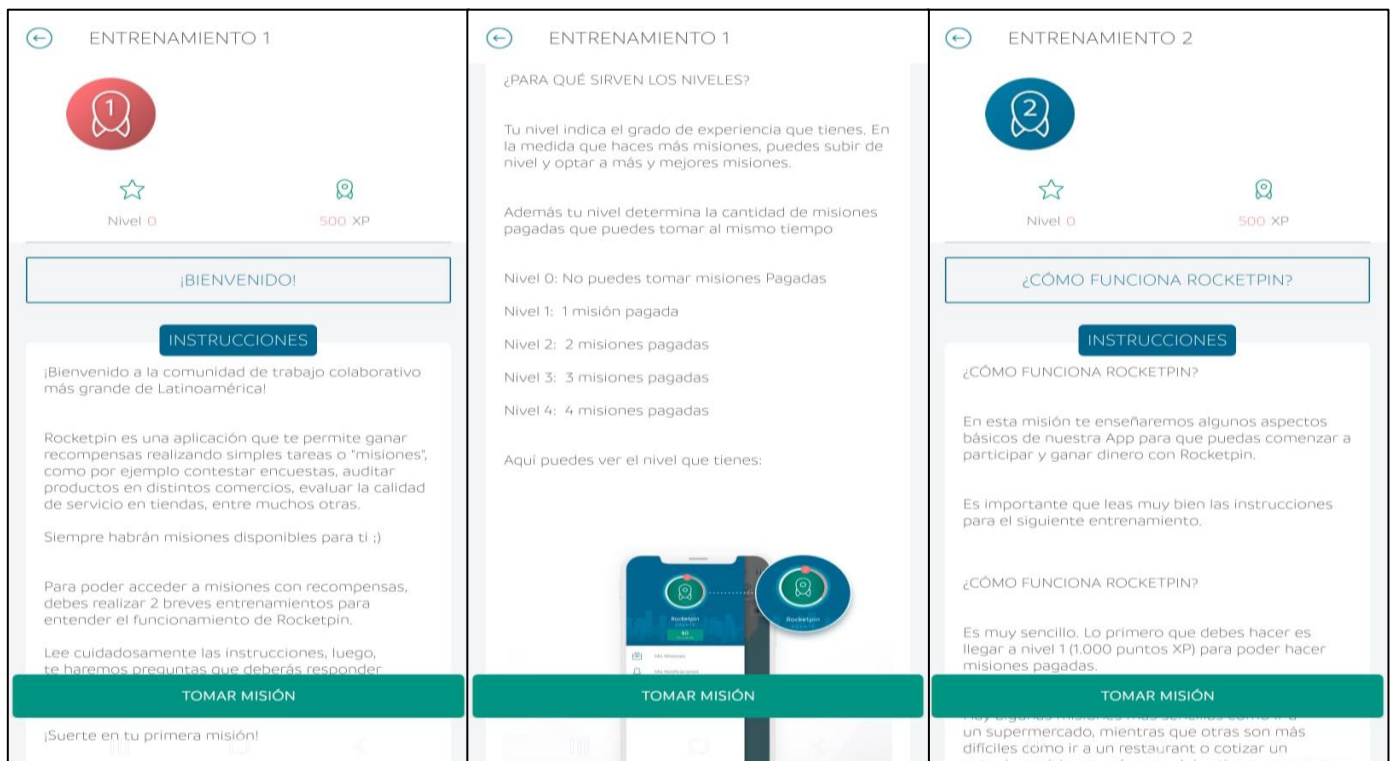


Ilustración 18: Misiones de entrenamiento para los shoppers.  
Fuente: Aplicación móvil de Rocketpin.

La primera misión que el shopper puede realizar cuando se une a la comunidad corresponde al *Entrenamiento 1*, en la cual se explica el funcionamiento general de la aplicación, a que corresponden las misiones, el sistema de niveles necesario para poder optar a más misiones remuneradas y los estándares de calidad que deben cumplir para que la misión éste aprobada. Para verificar que el shopper ha internalizado toda esta información, se realiza un cuestionario de opción múltiple de 4 preguntas, en las que el shopper debe conseguir que todas estén correctas para su aprobación. Si la misión es rechazada, el shopper puede optar a realizarla de nuevo hasta que consiga el puntaje perfecto. Las preguntas realizadas en esta misión pueden apreciarse en la ilustración 73 del anexo de este informe.

Posterior a la aprobación del *Entrenamiento 1*, el shopper deberá realizar la misión de *Entrenamiento 2*, para la cual deberá leer una serie de instrucciones que detallan los requisitos necesarios para pertenecer a la comunidad, como: ser mayor de 18 años, tener teléfono con datos móviles y contar con una cuenta bancaria propia. Además, se detalla el pago de las remuneraciones obtenidas por cada misión, el que será depositado en la cuenta del shopper en las próximas 48 horas posteriores al cumplimiento de éstas. Para completar esta misión, el shopper debe responder 4 preguntas de selección múltiple, siguiendo la misma modalidad del *Entrenamiento 1*. Las preguntas realizadas en esta misión pueden apreciarse en la ilustración 74 del anexo de este informe. Finalmente, los shopper deberán realizar la misión de entrenamiento especializada para cada servicio de Rocketpin, en las que se detallan los pasos que deben seguir para la aprobación de la misión y los estándares mínimos de calidad que deben cumplir.

### **7.2.2. Realización de misiones**

Las misiones remuneradas, las cuales corresponden a brindar el servicio de levantamiento de información en terreno, pueden empezar a realizarse a partir del nivel 1 (al cual se accede con 1.000 puntos de experiencia), habiendo completado previamente las misiones de capacitación. Para realizar una misión de verificación domiciliaria (servicio en el que se enfocará esta memoria), en primer lugar, debe buscarse en el mapa de la aplicación la ubicación de esta o acceder a misiones que no están georreferenciadas.

Luego el shopper debe reservar la misión, lo cual le da un plazo de 3 horas para completarla. Si este no completa la misión dentro del plazo, esta será rechazada y se liberará para otros usuarios. Para que el shopper pueda comenzar a realizar la misión, deberá estar a menos de 200 metros para aquellas que están geolocalizadas, para las otras el shopper deberá indicar el punto de inicio.

La verificación de domicilio consiste básicamente en una visita donde el shopper debe presentarse cordialmente indicando que viene de parte de Rocketpin, donde posteriormente deberá corroborar que cierta persona vive en el domicilio expuesto en la misión. Por lo tanto, la única información con la que cuenta el shopper corresponde al nombre de la persona a verificar y la dirección donde ésta vive. Es importante que el shopper tenga claridad de que la visita es algo que la persona a verificar espera, ya que significa que está solicitando una apertura de cuenta, un producto bancario o algo similar. Es por esto que, el shopper no debería tener ningún inconveniente al momento de realizar la visita domiciliaria.

Una vez que el shopper este a menos de 200 metros de la misión o que indique su inicio en la aplicación, se empezará a grabar un audio automáticamente, el cual permite a Rocketpin verificar el correcto desempeño del shopper. Posteriormente, la primera tarea del usuario será presentarse con su nombre y mostrar la credencial digital presente en la aplicación. Es posible que el shopper se enfrente a cualquiera de las siguientes situaciones:

1. Que lo reciba la misma persona que se requiere verificar.
2. Que lo reciba algún familiar, empleada o conserje del edificio.
3. Que nadie conteste al llegar al domicilio.
4. Que haya una persona en el domicilio, pero que no quiera entregar información.

En los primeros dos casos, el shopper deberá indicar en sus respuestas la persona que lo recibió y corroborar si la persona a verificar vive o no en el domicilio, sacar dos fotografías de la fachada de la casa, edificio o sector de conserjería. Además, el shopper deberá indicar en sus respuestas acerca de la vivienda, el color de la fachada, cantidad de pisos y deslinde. En los últimos dos casos, el shopper deberá indicar la situación sucedida en los comentarios y finalizar la misión.

### **7.2.3.Revisión de misiones**

Este proceso comienza una vez que el formulario de la misión es llenado por el shopper esperando su aprobación o rechazo. Un ejemplo del formulario a revisar por el digitador puede apreciarse en la ilustración 75 del anexo de este informe.

La revisión de misiones es realizada por 3 digitadores, bajo la supervisión del jefe de área. Esta labor es primordial para la operación del negocio, puesto que se verifica que la información entregada por los shopper es fidedigna, ya que de no serlo la empresa pierde prestigio y está sujeta a multas por incumplimiento de contrato.

Existen 6 puntos que el digitador debe revisar a la hora de aprobar o rechazar una misión, los cuales se mencionan a continuación:

**a) Audio:** el audio deberá ser revisado obligatoriamente en los siguientes casos:

- No hay nadie en el domicilio ni alrededores.
- El nombre de la persona a verificar es inválido.
- El shopper es nuevo (ha realizado pocas misiones).
- El audio tiene poca duración (menor a 1 minuto).

La misión será rechazada por el audio dado los siguientes motivos:

- No se evidencia interacción con el informante (persona que atiende al shopper en el domicilio).
- No se evidencia que el shopper tenga intención de encontrar al cliente (no se escucha un "Aló/Hola", no se pregunta a los vecinos, etc.).
- No sigue el protocolo de presentación.

**b) Fotos:** las fotos deberán revisarse en todos los casos. El digitador deberá rechazar la misión por fotos dado los siguientes motivos:

- Foto borrosa.
- Sin numeración (o sin la explicación pertinente en el comentario).
- Foto con mal ángulo.

**c) Comentario:** los comentarios deberán revisarse en todos los casos. El digitador debe corregir el comentario solo si es que existen faltas ortográficas leves. El digitador deberá rechazar la misión por el comentario dado los siguientes motivos:



- Falta de ortografía grosera.
- Comentario no atinente a la misión (ejemplo: se quedó congelada la aplicación, la misión queda lejos del hogar del shopper, etc.).
- Comentario inconsistente con las respuestas.

**d) Ubicación PIN:** la ubicación deberá ser revisada obligatoriamente en los siguientes casos:

- No existe la dirección: El digitador deberá revisar en el mapa si es que existe realmente.
- Fotos sin numeración: El digitador deberá revisar la numeración más cercana en el mapa.
- Si es que la distancia entre los pines de inicio y fin de la misión están muy alejados entre sí.

La misión deberá rechazarse dado los siguientes motivos:

- Si el domicilio no posee numeración o algún otro error de dirección.

**e) Tiempo de duración:** existe una alta correlación entre misiones de baja duración con fraudes de información. Dado esto es que el digitador deberá rechazar la misión si es que ésta es excesivamente corta (dura menos de un minuto).

**f) Consistencia de respuestas:** la consistencia debe existir a nivel de respuestas-fotos (color de la casa, cantidad de pisos, etc.) y combinación de respuestas. Estas últimas son las más importantes, ya que no puede haber inconsistencias en la información.

El digitador deberá rechazar la misión por errores en las respuestas, dado los siguientes motivos:

- Nombre del informante: si la persona que atiende al shopper en el domicilio no se identifica con su nombre y apellido, la verificación no es válida. En ese caso la pregunta "¿Había alguien en el domicilio?" debe responderse con la alternativa "No, ni en el domicilio ni en alrededores" y dejar la observación en el comentario.

- Error de dirección: Si se logra dar con el cliente, pero la dirección entregada es incorrecta (por ejemplo, Monjitas 550 pero el cliente vive en Monjitas 551), la verificación de domicilio no es válida y se procede igual que en el caso anterior.

El detalle del proceso puede apreciarse de manera gráfica mediante la metodología BPMN presentada a continuación:

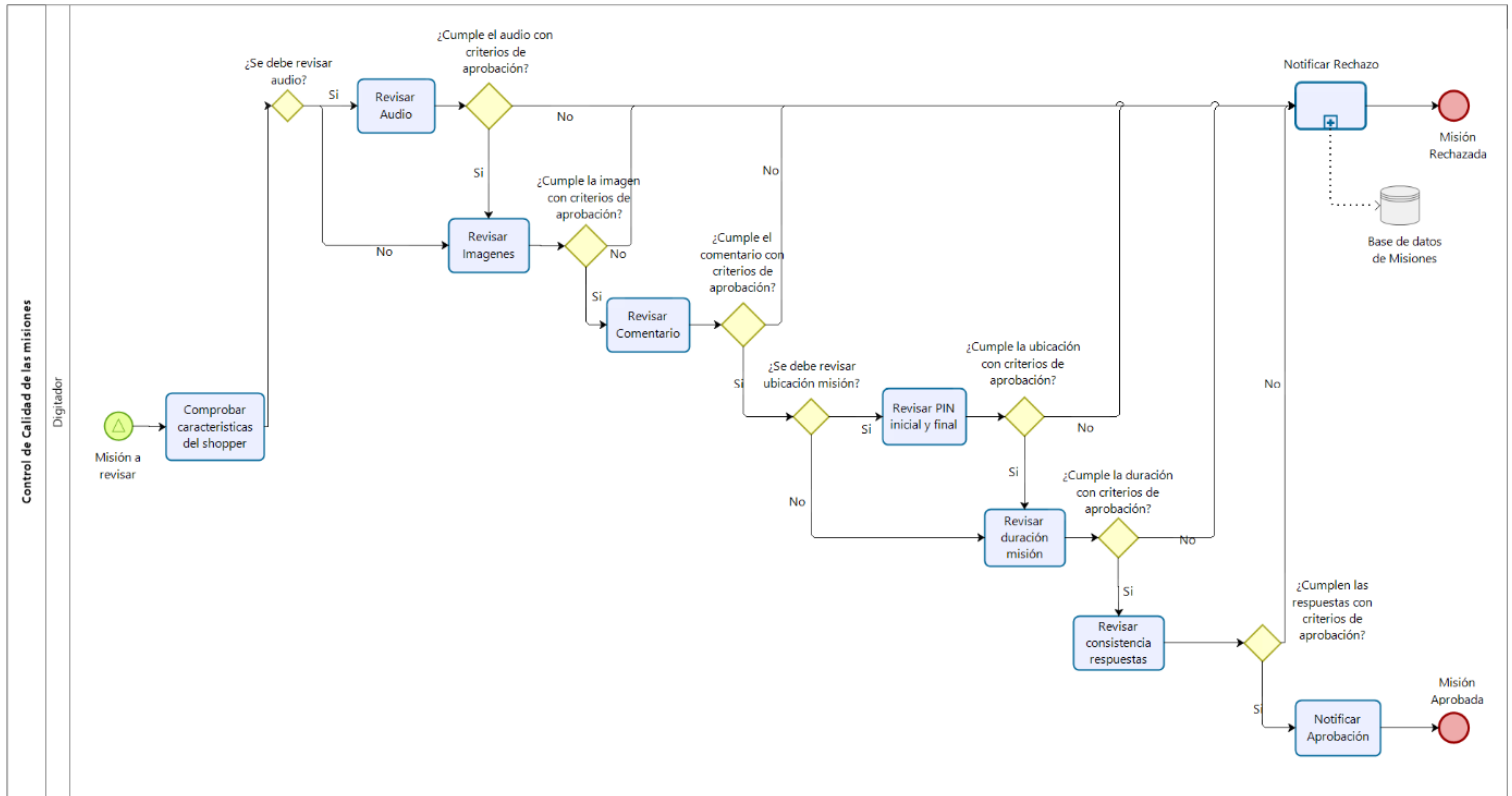


Ilustración 19: Diagrama BPMN de la situación actual del proceso de revisión.  
Fuente: Elaboración Propia.

### 7.3. Descripción de la situación actual del negocio

Para describir la situación en la que se encuentra Rocketpin, se estudiará la operación del negocio en la ventana de tiempo comprendida entre enero y diciembre de 2019.

En ese período de tiempo, se realizaron un total de 536.444 verificaciones domiciliarias, donde 485.020 fueron aprobadas (90,4% del total) y 51.424 fueron rechazadas (9,6% del total). Estas misiones fueron realizadas por 1.297 shoppers diferentes, donde en promedio cada uno realizó 413,60 misiones en el año. En el gráfico a continuación se puede apreciar la cantidad de misiones realizadas mensualmente en el período de análisis:

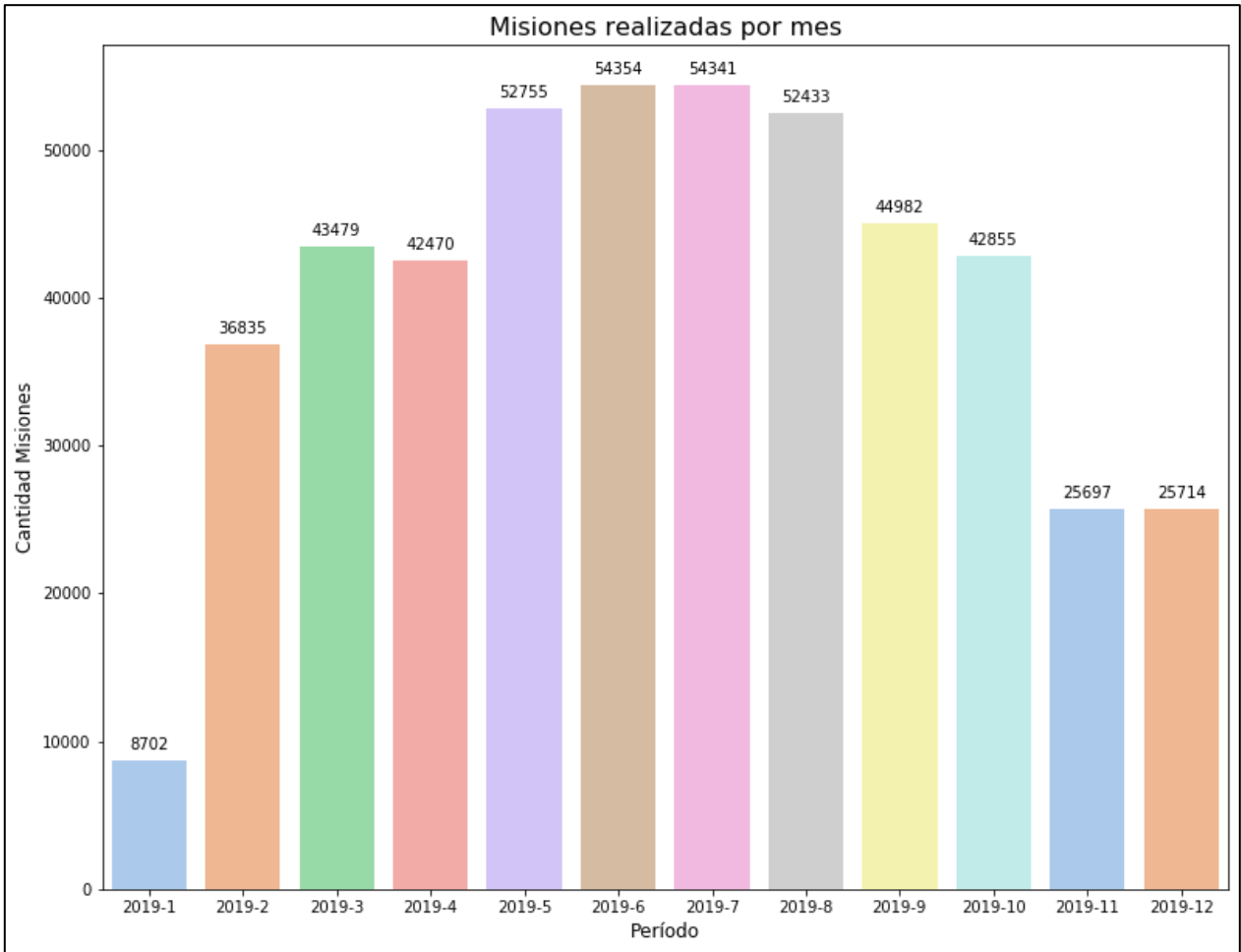


Ilustración 20: Gráfico de barra de la cantidad de misiones realizadas por mes.  
Fuente: Elaboración Propia.

Los *peaks* alcanzados se encuentran en los meses de junio y julio, con un total aproximado de 54.350 misiones, siendo el promedio de misiones realizadas por mes igual a 40.385.

El promedio de la remuneración por misión para el período de análisis corresponde a \$1.101,79, por lo cual se puede extrapolar esta cifra para determinar el gasto mensual en el que ha incurrido Rocketpin en términos de pago a shoppers, lo cual puede apreciarse en el siguiente gráfico:

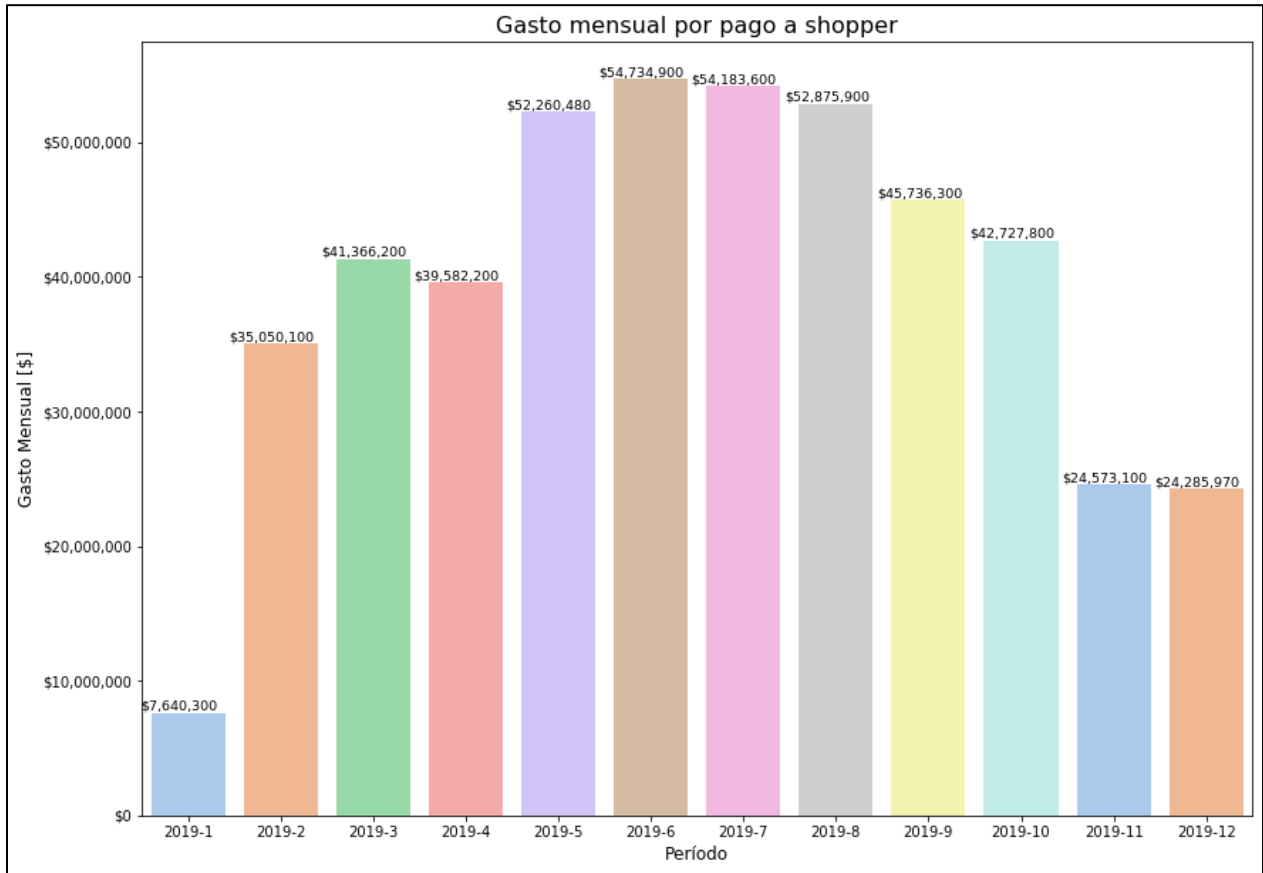


Ilustración 21: Gráfico de barra del gasto mensual en pago a shoppers.  
Fuente: Elaboración Propia.

De este último se desprende que los meses de mayor gasto corresponden a junio y julio. El gasto total en pago a shoppers para el período de análisis asciende a \$ 475.016.850. Rocketpin margina aproximadamente el 30% de la remuneración realizada a los shoppers, por lo que en promedio obtienen un total de \$1.573,99 por misión.

Los ingresos mensuales que obtuvo Rocketpin, solo considerando el servicio de verificaciones domiciliarias, puede apreciarse en el siguiente gráfico:

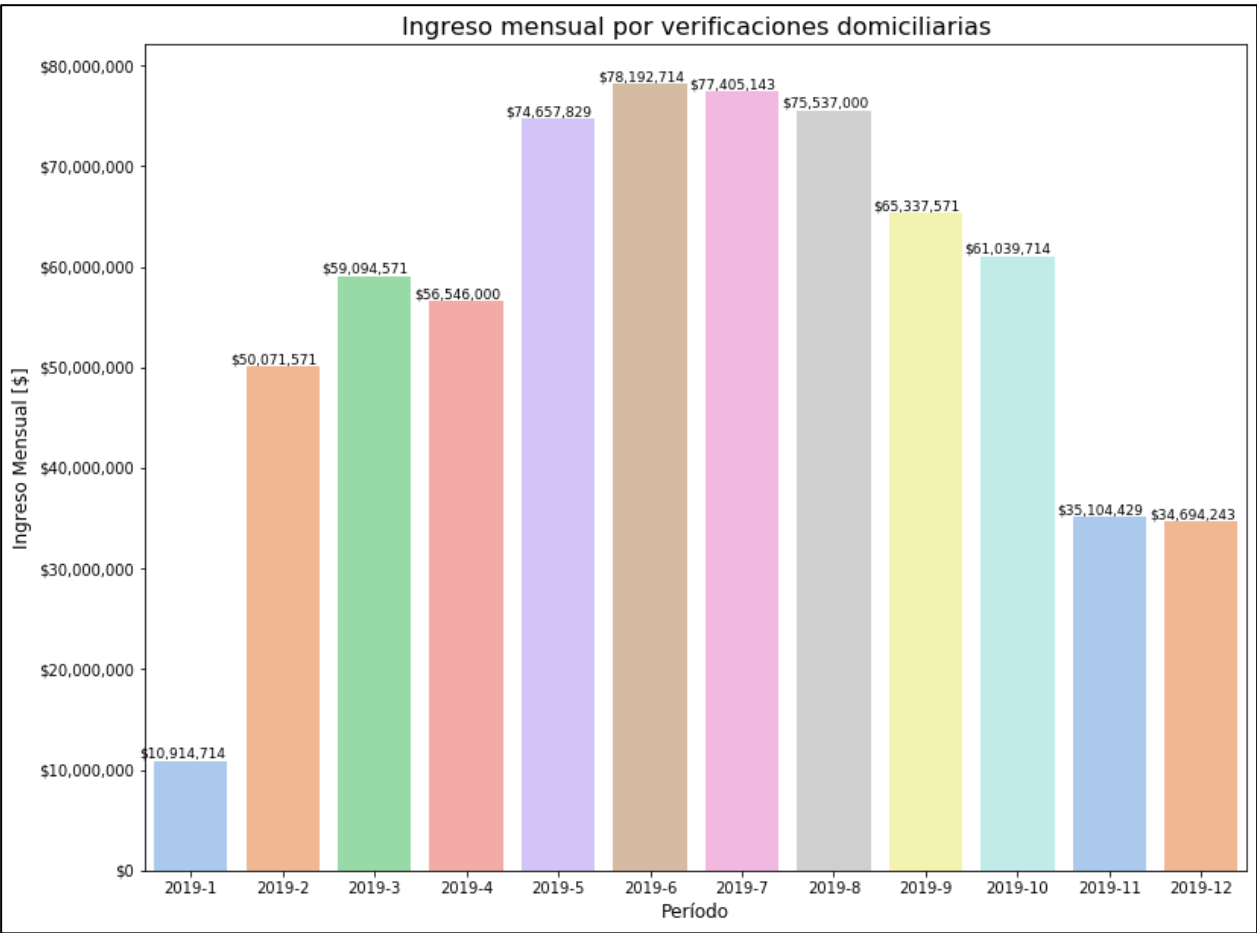


Ilustración 22: Gráfico de barra de los ingresos mensuales de Rocketpin.  
Fuente: Elaboración Propia.

Los ingresos de Rocketpin para el servicio de verificaciones domiciliarias en el año 2019 fueron de \$678.596.500. Esto indicaría que el beneficio de la organización, considerando solamente los costos variables provenientes del pago a shoppers, sería de \$203.579.650 (aproximadamente un 30% del total).

Dado el nivel de facturación de la empresa y que la cantidad de empleados según la planilla es igual a 25, Rocketpin sería clasificada como una empresa de pequeño tamaño, ya que estaría compuesta por entre 6 a 50 trabajadores, y/o su volumen de ventas oscila entre 2.401 a 25.000 U.F anuales [17].

**7.4. Análisis del proceso de revisión**

Como se mencionó en la sección 7.2, el proceso de revisión de misiones para el servicio de verificación domiciliaria sigue una serie de exhaustivos pasos que deben realizarse a cabalidad para aprobar o rechazar una misión. Para ello

Rocketpin cuenta con 3 digitadores y 1 jefe de área, encargado de supervisar el correcto funcionamiento del proceso.

De la ilustración 8 en la sección 3, puede apreciarse que en promedio los digitadores revisan 450 misiones diarias, no pudiendo alcanzar la meta observada de 800 misiones. Para comprender el porqué de esta situación, se procedió a medir los tiempos que demoraba un digitador en revisar sus misiones. Se realizó la medición del tiempo para un total de 150 misiones, lo cual puede verificarse en la ilustración 76 del anexo de este informe.

La distribución de los tiempos de revisión puede apreciarse en la siguiente ilustración:

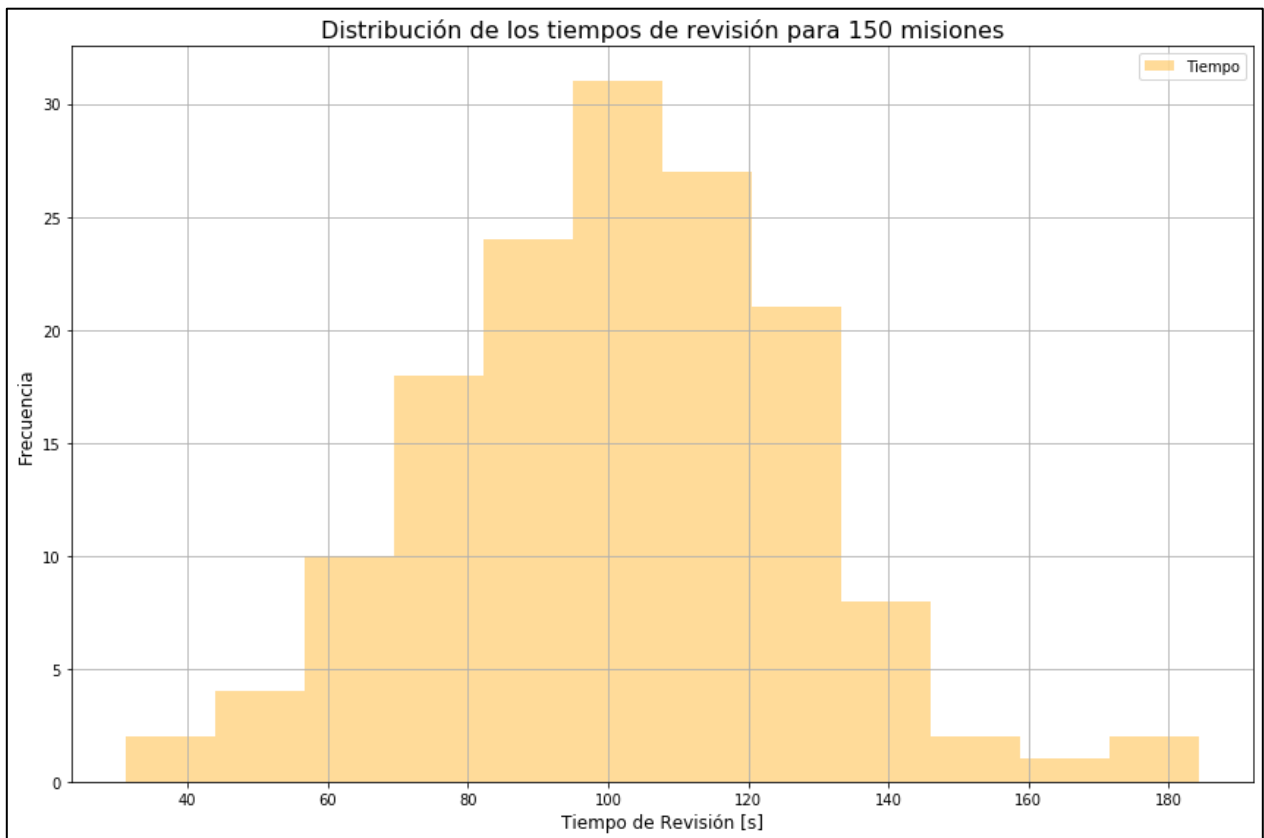


Ilustración 23: Gráfico de distribución de los tiempos de revisión de misiones.  
Fuente: Elaboración Propia.

De la ilustración anterior, puede apreciarse que los tiempos de revisión se asemejan a una campana de Gauss, por lo que podría decirse que distribuyen como una normal. El tiempo promedio de revisión es de aproximadamente 100,92 [s] (o 1 minuto y 40 segundos) con una desviación estándar de 80 [s]. Que la desviación estándar sea casi idéntica al promedio de la distribución, sugiere que existe alta variabilidad en este proceso, lo cual explica que existan misiones que se revisan en 31 segundos y otras en 3 minutos.

Tomando el promedio del tiempo de revisión, para cumplir con la meta de 800 misiones, los digitadores deberían trabajar aproximadamente 22,2 horas diarias, lo cual excede con creces su jornada laboral de 9:30 a 19:00 horas. De hecho, en base a estos resultados, a los digitadores les tomaría aproximadamente 12,5 horas revisar las 450 misiones que corresponden a su rendimiento promedio diario. Dada esta información, se procedió a entrevistar a los digitadores del área, con el fin de consultarles acerca de la situación, sobre la estandarización del proceso y sus condiciones laborales.

Los digitadores señalan que, si bien existe una rúbrica de pasos que deben seguir para revisar correctamente una misión, en muchos casos utilizan su intuición para acelerar el proceso de revisión, ya que de hacerlo de forma tan rigurosa no podrían siquiera cumplir con el mínimo diario solicitado. Además, señalan que en muchos casos no alcanzan a terminar con la cuota de misiones a revisar en su jornada laboral, por lo que tienen que completar este trabajo en sus hogares y en tiempos que deberían destinarse al ocio/descanso.

La intuición en la que se basan para revisar las misiones de forma más expedita consiste en ser menos rigurosos con aquellos shoppers que tienen buen desempeño. Esto pueden realizarlo dado que recuerdan a los shoppers que cumplen con los estándares fijados, sin embargo, esta información no se presenta en el formulario que los digitadores revisan. Otro factor que utilizan corresponde a la distancia entre la ubicación inicial y final de la misión, ya que esto podría indicar que el shopper falsificó su ubicación o que no llegó a la dirección geolocalizada en la aplicación. Lo anterior también se realiza de forma visual e intuitiva, ya que el formulario no incluye esta distancia. Además, para facilitar su labor, los digitadores analizan el tiempo que le toma al shopper realizar la misión y la duración del audio, sin embargo, no tienen una idea clara del rango normal en el que estos indicadores deberían moverse.

Respecto a su remuneración, los digitadores declaran que ganan un sueldo de \$400.000 brutos y el jefe de área un total de \$550.000, por lo que anualmente el proceso de revisión en términos de planilla cuesta a Rocketpin \$21.000.000 (aproximadamente un 3,09% del total de ingresos en verificaciones domiciliarias). Este monto podría reducirse si es que se rediseña y/o automatiza el proceso, puesto que Rocketpin no debería revisar el 100% de las misiones, lo cual permitiría reducir al personal y/o incluirlos en otras áreas.

Dado todo lo anterior, se extrae que el proceso de revisión de misiones no está estandarizado, no aprovecha la información generada por los shoppers ni las misiones y además genera complicaciones a los digitadores. Es por esto que, se hace evidente la necesidad de rediseñar el proceso para mantener los

estándares de calidad y poder aumentar el volumen de misiones realizadas y revisadas, buscando optimizar los costos y tiempos en los que se incurre.

### **7.5. Selección de la variable crítica**

Para todo proyecto de rediseño de procesos es necesario enfocar los esfuerzos de cambio, por lo que según la metodología expuesta por Juan Bravo [9, p. 56], se debe priorizar la selección de solo un indicador para efectos de evaluación. En el caso del proceso de revisión de misiones, el indicador que se medirá para evaluar el rediseño será el tiempo de revisión para el flujo diario de 800 misiones.

Este indicador en la actualidad es igual a 22,2 horas, considerando que en promedio un digitador se demora 1 minuto y 40 segundos en revisar una misión. Posterior a la definición de la variable crítica, se hace necesario definir la meta o valor deseado al cual se quiere llegar, lo cual puede realizarse desde distintos enfoques.

La meta propuesta para el rediseño consiste en que el tiempo de revisión disminuya por lo menos para que los digitadores logren cumplir la meta de 800 misiones en su jornada laboral diaria, lo cual quiere decir que se espera que el valor de la variable crítica disminuya de 22,2 a 9,5 horas, lo que representa una reducción de un 53%.

## **8. ANÁLISIS DE LAS BASES DE DATOS**

El análisis exploratorio de datos es una forma de analizar información definido por John W. Tukey, el cual corresponde al tratamiento estadístico al que se someten las muestras recogidas durante un proceso de investigación en cualquier campo científico. Para mayor rapidez y precisión, generalmente todo este proceso suele realizarse por medios informáticos [18]. En este caso, todo el análisis exploratorio de datos se realizó utilizando el lenguaje de programación Python y el visualizador Jupyter Notebook.

Rocketpin cuenta con múltiples bases de datos con información relativa a las misiones y shoppers, con lo cual se puede generar un entendimiento más riguroso del proceso de revisión, caracterizar a los usuarios de la aplicación y su comportamiento. Además, el análisis será útil para identificar qué características poseen los shoppers y misiones que son más influyentes en la aprobación y rechazo de estas últimas.



## 8.1. Clasificación de motivos de rechazo

Para entender cuáles son las principales observaciones que realizan los digitadores al momento de rechazar una misión, se procedió a analizar una base de datos entregada por Rocketpin, la cual poseía el motivo de rechazo de 1.092 misiones desaprobadas. Las variables existentes en la base de datos se pueden apreciar a continuación:

Variables	Significado
id_mision	Llave única que identifica la misión
mail_shopper	Email del shopper
motivo	Observación del porqué del rechazo entregada por el digitador

Ilustración 24: Descripción de las variables en la base de datos de motivos de rechazo.  
Fuente: Elaboración Propia.

El problema encontrado en esta base de datos es que no existe una estandarización de los motivos de rechazo indicados por los digitadores. De hecho, al analizar las distintas observaciones realizadas, existen 701 motivos diferentes para las 1.092 misiones, lo cual no permite visualizar de forma clara por qué se rechazan las misiones. Además, esta situación no se condice con los motivos de rechazo que se les indican a los digitadores (que corresponden a: fotos, audios, ubicación PIN, comentarios, respuestas y duración de la misión), lo que puede deberse a una desorganización o a la falta de supervisión en esta labor.

El problema anterior puede ejemplificarse al considerar observaciones de rechazo distintas, que podrían ser agrupadas en una misma categoría: *"formulario mal llenado no se obtiene apellido de informante"*, *"formulario mal llenado no se obtiene nombre de informante"*, *"no se obtiene datos del informante"*, *"no se indica nombre de cliente a verificar"*, entre otras. Dado esto, se procedió a agrupar en categorías generales, todas las observaciones entregadas por los digitadores, para de esta forma identificar los principales motivos de rechazo.

De este proceso, se generaron 13 categorías de rechazo distintas, las que se listan a continuación:

Categoría	Frecuencia	Porcentaje
SIN NUMERACIÓN	152	14%
NO CORRESPONDE A LA COMUNA	149	14%
FORMULARIO MAL COMPLETADO	141	13%
PIN FINAL NO CORRESPONDE	138	13%
SI EXISTE DIRECCION	126	12%
AUDIO NO DEMUESTRA INTERACCIÓN	84	8%
FALTAN DATOS INFORMANTE	83	8%
SECTOR PELIGROSO	68	6%
SECTOR FUERA DE COBERTURA	62	6%
FOTOGRAFÍA NO CORRESPONDE	27	2%
HORARIO RESTRINGIDO	21	2%
COMENTARIO NO CONSISTENTE	21	2%
CONDOMINIO CERRADO	20	2%

Ilustración 25: Categorización de los motivos de rechazo de las misiones.  
Fuente: Elaboración Propia.

Dentro de los principales motivos de rechazo se encuentran las misiones en las que los shoppers no envían la foto de la numeración del domicilio y aquellas en las que la dirección indicada no correspondía a la geolocalización. Dentro de los motivos que indican una falta de capacitación y/o experiencia del shopper se destacan: formulario mal completado, comentario no consistente, faltan datos del informante, fotografía no corresponde y formulario mal completado. Otro de los motivos de rechazo más frecuentes corresponde a que el PIN final no se adecua a la geolocalización, lo cual indica que la ubicación final está a más de 200 metros del lugar donde debería.

Este análisis permite generar *insights* iniciales respecto a las variables que podrían tener una mayor relación con la aprobación y/o rechazo de las misiones como, por ejemplo: experiencia del shopper, misiones realizadas, porcentaje de misiones aprobadas, distancia entre los pines, entre otras.

## 8.2. Análisis exploratorio de la base de misiones

Para este análisis se considerará una ventana de análisis entre enero y diciembre de 2019. La base de datos de misiones posee las siguientes variables, cuya descripción puede apreciarse a continuación:

Variables	Descripción	Inválidos	Validos	%Inválidos	Clase
<b>id</b>	Llave única que identifica la misión	0	536.444	0,00%	object
<b>location</b>	Coordenadas de la ubicación donde el shopper inicia la misión	0	536.444	0,00%	object
<b>end_coords</b>	Coordenadas de la ubicación donde el shopper termina la misión	47.227	489.217	8,80%	object
<b>headquarter_street</b>	Dirección de la misión	1	536.444	0,00%	object
<b>headquarter_location</b>	Coordenadas de la geolocalización de la misión	1	536.443	0,00%	object
<b>shopper_email</b>	Email del shopper	44.422	492.022	8,30%	object
<b>approver_id</b>	Llave que identifica al digitador	0	536.444	0,00%	int64
<b>approver_email</b>	Email del digitador	50.958	485.486	9,50%	object
<b>commune</b>	Comuna donde se realiza la misión	1	536.443	0,00%	object
<b>country</b>	País donde se realiza la misión	1	536.443	0,00%	object
<b>rut</b>	Rut del shopper	3	536.441	0,00%	object
<b>campain_id</b>	Llave que identifica la campaña de verificación domiciliaria	0	536.444	0,00%	int64
<b>created_at</b>	Fecha y hora en la que se creo la misión	0	536.444	0,00%	object
<b>completed_date_time</b>	Fecha y hora en la que se completó la misión	45.179	491.265	8,40%	object
<b>started_date_time</b>	Fecha y hora en la que comenzó la misión	45.657	490.787	8,50%	object
<b>taken_at</b>	Fecha y hora en la que el shopper tomo la misión	20.651	515.793	3,80%	object
<b>approved_at</b>	Fecha y hora en la que se aprobó la misión	50.958	485.486	9,50%	object
<b>disapproved_at</b>	Fecha y hora en la que se rechazó la misión	485.682	50.762	90,50%	object
<b>state</b>	Estado de la misión	0	536.444	0,00%	object
<b>amount</b>	Pago realizado al shopper por la misión	0	536.444	0,00%	float64
<b>company_id</b>	Llave que identifica a la compañía que contrato a Rocketpin	28.291	508.153	5,30%	float64
<b>company_name</b>	Nombre de la compañía	28.291	508.153	5,30%	object

Ilustración 26: Descripción de las variables de la base de misiones.  
Fuente: Elaboración propia.

De la tabla, puede apreciarse que la base de datos cuenta con un total de 536.444 observaciones y 22 variables. Respecto a la calidad de la información expuesta, se observa que la mayoría de las variables tiene menos de un 10% de valores nulos, a excepción de "disapproved\_at", lo cual se debe a que solo el 9,6% de las misiones totales fueron desaprobadas.

El tratamiento de los valores nulos depende del origen de la falta de información, la cual puede ser *Missing Completely at Random* (MCAR), *Missing at Random* (MAR) o *Missing Not at Random* (MNAR) [19]. El primer caso indica que el hecho de que falte una observación no está relacionado con el o los valores faltantes ni con los existentes (o que cualquier valor tiene la misma probabilidad de faltar que algún otro). Por ejemplo, una falla ocasional que impide hacer una medición o el olvido ocasional en registrar un dato. El segundo caso corresponde a que una o varias características registradas pueden explicar la distribución de datos faltantes (también llamado falta de datos manejable). Por ejemplo: el nivel de respuestas faltantes en una encuesta está relacionado con el nivel socioeconómico o el número de pacientes con un electrocardiograma preoperatorio está relacionado con la edad de los pacientes. El último caso y el más difícil de tratar corresponde a

cuando los datos faltantes probablemente dependen o estén relacionados con datos no observados. Por ejemplo, que un hombre no conteste una encuesta sobre depresión debido a su nivel de depresión [20]. Para el caso estudiado se podría decir que la falta de datos corresponde al tipo MAR, ya que los valores nulos pueden explicarse completamente por medio de otras variables registradas (como el caso de "dissaproved\_at" y "approved\_at"). Para los casos MCAR y MAR se dice que la falta de datos es ignorable y no es necesario un tratamiento específico, dependiendo de la cantidad de estos valores [21].

La tolerancia aceptable para el porcentaje de valores nulos que puede poseer una variable, según se indica en la literatura, debería encontrarse en un rango de entre 1-10% del total [22]. De no ser así, existen múltiples métodos para el tratamiento de la data faltante, como reemplazar estas observaciones con el promedio de las demás, por la observación más frecuente y/o métodos más complejos, como KNN (*k-nearest neighbors*) e imputación multivariada. Dado todo lo anterior, se procedió simplemente a eliminar la variable "dissaproved\_at" y trabajar con los datos en bruto. Respecto a la clase a la que pertenece cada variable, se señala la descripción de cada una en la ilustración a continuación:

Pandas Type	Native Python Type	Descripción
object	string	El tipo de dato más general. Se les asigna a columnas que tengan texto y valores numéricos.
int64	int	Caracteres numéricos. El 64 corresponde a la cantidad de números dígitos máximo que puede tener.
float64	float	Caracteres numéricos con decimales
datetime64	N/A	Datos del tipo fecha y hora

Ilustración 27: Descripción de los tipos de datos en Python [23].

Fuente: Elaboración propia.

Para el caso a estudiar, se cargó la base de datos en Jupyter Notebook usando la librería Pandas (una de las más utilizadas para la manipulación y análisis de datos), lo cual explica el tipo de las variables en la ilustración 27. El primer tratamiento realizado, correspondió a transformar las variables que poseían la clase incorrecta, como el caso de aquellas terminadas en "\_at" y "datetime", las cuales deberían ser del tipo datetime64 en vez de object, ya que, de no hacerlo no se podría trabajar con las fechas y horas. Esto se realizó utilizando la librería "datetime" de Python, la que permite transformar datos del tipo texto a fecha.

En base a este trabajo, se pudieron generar nuevas variables a partir de las modificadas, las cuales corresponden a tiempos entre las acciones que debe realizar un shopper para completar una misión. Para trabajar las variables que

contienen coordenadas ("location", "end\_coords" y "headquarter\_location"), fue necesario separarlas en dos para obtener la latitud y longitud, ya que estaban organizadas de la forma: (LAT, LONG). Para determinar la distancia entre 2 puntos geográficos, es necesario utilizar la formula del Haversine [24]:

$$R = \text{Radio de la Tierra} = 6.371 \text{ [km]}$$

$$a = \sin^2\left(\frac{\Delta Lat}{2}\right) + \cos(Latitud_1) * \cos(Latitud_2) * \sin^2\left(\frac{\Delta Long}{2}\right)$$

$$c = 2 * a * \tan2(\sqrt{a}, \sqrt{1-a})$$

$$\text{Distancia} = R * c \text{ [km]}$$

Es importante destacar que la base de datos de misiones no posee ninguna variable que permita determinar la experiencia de los shoppers, lo cual según se vio en la sección 8.1, es fundamental a la hora de explicar los rechazos de las misiones. Es por esto, que se generaron dos variables que incorporarán esta información, las que corresponden a la cantidad de misiones realizadas por los shoppers en los últimos 3 meses y su tasa de aprobación en ese período. La intuición del porque se consideran los últimos 3 meses y no toda la ventana de análisis, se debe a que de esta forma se da mayor peso a la actividad reciente, ya que seguramente un shopper que realizó una misión hace poco tenga menor probabilidad de equivocarse que otro que realizó la misma misión hace mucho tiempo atrás (el primero estará más actualizado sobre la realización del servicio). El objetivo de trabajar los datos consiste en facilitar su manipulación y en este caso, permitió generar nuevas variables a partir de las modificadas, las cuales pueden apreciarse a continuación:

Variables	Descripción	Inválidos	Validos	%Inválidos	Clase
LAT_ini	Latitud del punto donde se inicia la misión	0	536.444	0%	float64
LONG_ini	Longitud del punto donde se inicia la misión	0	536.444	0%	float64
LAT_fin	Latitud del punto donde se finaliza la misión	47.227	489.217	9%	float64
LONG_fin	Longitud del punto donde se finaliza la misión	47.227	489.217	9%	float64
LAT_headquarter	Latitud de la geolocalización de la misión	1	536.443	0%	float64
LONG_headquarter	Longitud de la geolocalización de la misión	1	536.443	0%	float64
distancia	Distancia entre el punto final e inicial de la misión	47.227	489.217	9%	float64
distancia_headquarter	Distancia entre el punto final y su geolocalización	47.228	489.216	9%	float64
t_started_vs_completed	Tiempo entre que comienza la misión y se completa	45.749	490.695	9%	float64
t_taken_vs_completed	Tiempo entre que comienza la misión y se toma	47.194	489.250	9%	float64
t_taken_vs_started	Tiempo entre que se toma la misión y se comienza	47.755	488.689	9%	float64
approved_mean	Tasa de aprobación de misiones en los últimos 3 meses	129.043	407.401	24%	float64
approved_count	Cantidad de misiones realizadas en los últimos 3 meses	129.043	407.401	24%	float64
Approved	Dummy que toma el valor 1 si la misión está aprobada, 0 sino	0	536.444	0%	Int64
Dissaproved	Dummy que toma el valor 1 si la misión está desaprobada, 0 sino	0	536.444	0%	Int64

Ilustración 28: Variables creadas a partir de la base de misiones.

Fuente: Elaboración Propia.

Una vez generadas las variables necesarias para caracterizar el comportamiento que poseen las misiones, se procedió a generar gráficos descriptivos, los cuales permiten representar gráficamente los datos y visualizar tendencias y anomalías. En el siguiente gráfico se pueden apreciar las 10 principales comunas donde se realizan las misiones de verificación domiciliaria:

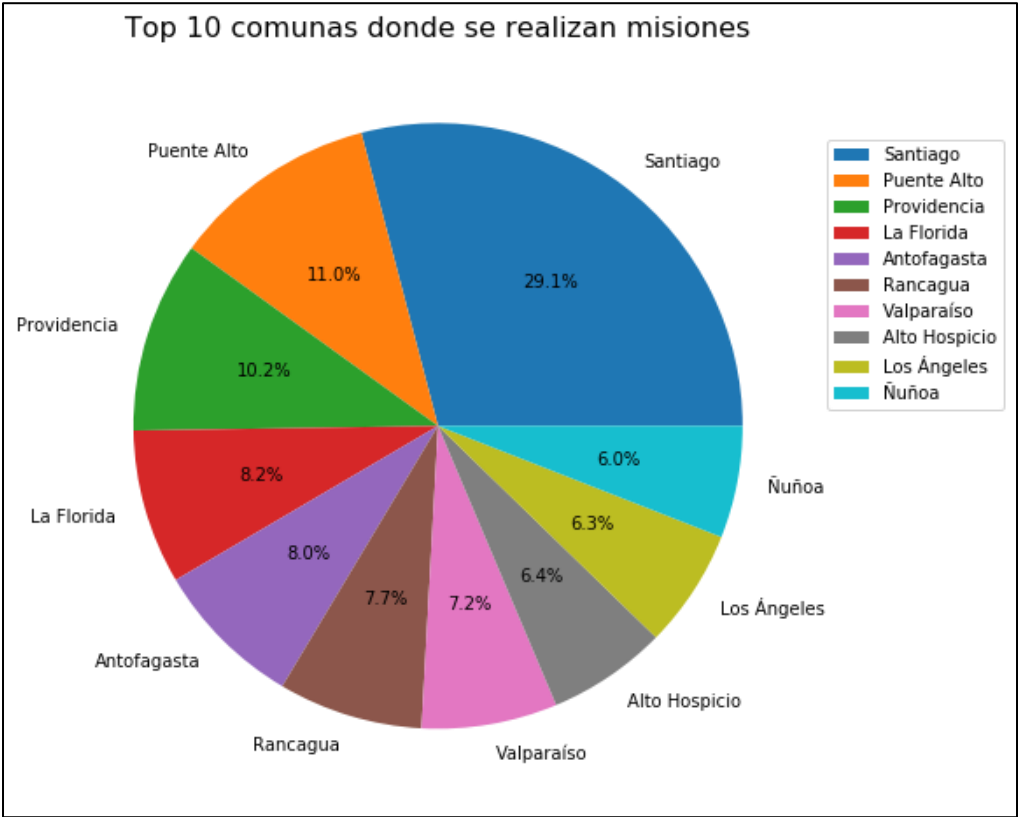


Ilustración 29: Comunas donde más se realizan misiones de verificación domiciliaria.  
Fuente: Elaboración Propia.

De este, se destaca que las principales comunas corresponden a la región metropolitana, destacándose Santiago, Puente Alto y Providencia. Dentro de los principales clientes de Rocketpin que solicitan verificaciones domiciliarias se encuentran: Hites, Banco Santander, Scotiabank, Cencosud, La Polar y Banco Security. Se destaca que la mayoría de estos clientes tienen presencia en el sector bancario, lo cual es esperable dado que el servicio de verificación domiciliaria generalmente se solicita para los préstamos crediticios.

Lo anterior, puede apreciarse en la ilustración 30 y es importante destacar que el intermediario de estas empresas frente a Rocketpin corresponde a EQUIFAX.

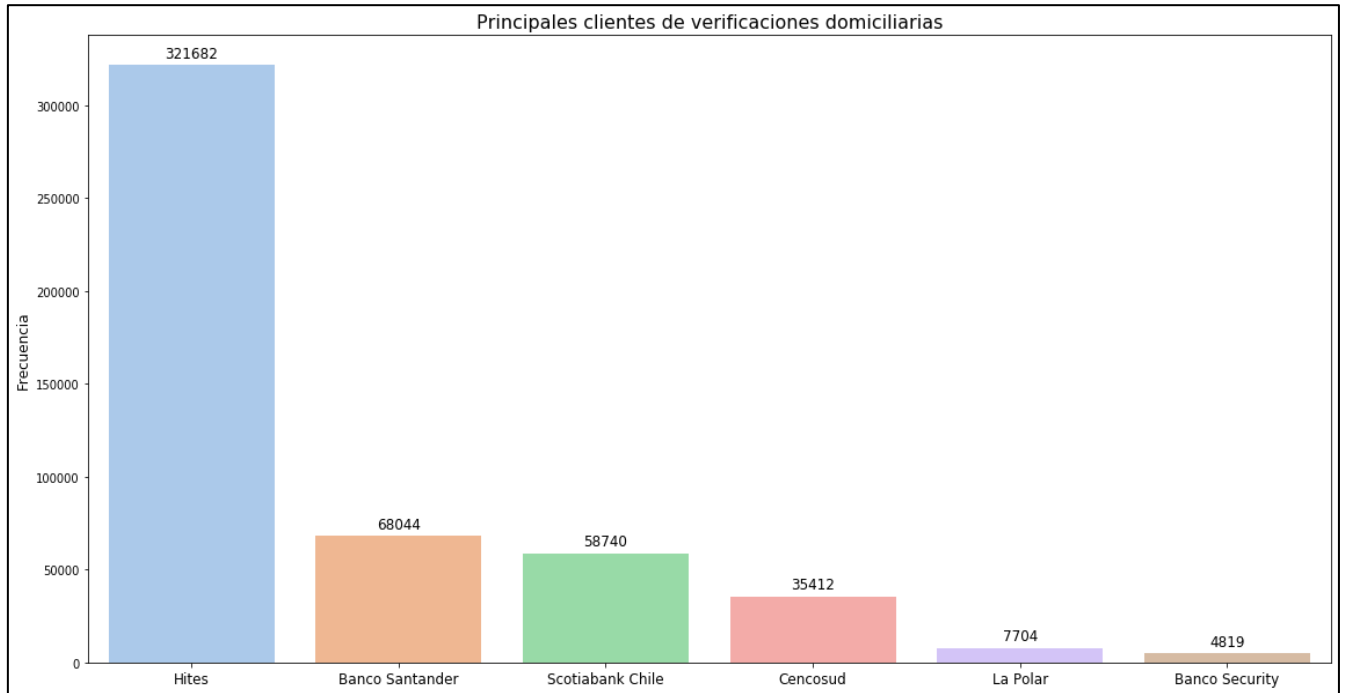


Ilustración 30: Frecuencia de las verificaciones realizadas por cliente.  
Fuente: Elaboración Propia.

Al analizar la distribución del pago por misión realizado a los shoppers, se puede notar que es bastante homogéneo, encontrándose el 75% de los valores entre \$800 y \$1.500, lo cual puede apreciarse en la ilustración 31:

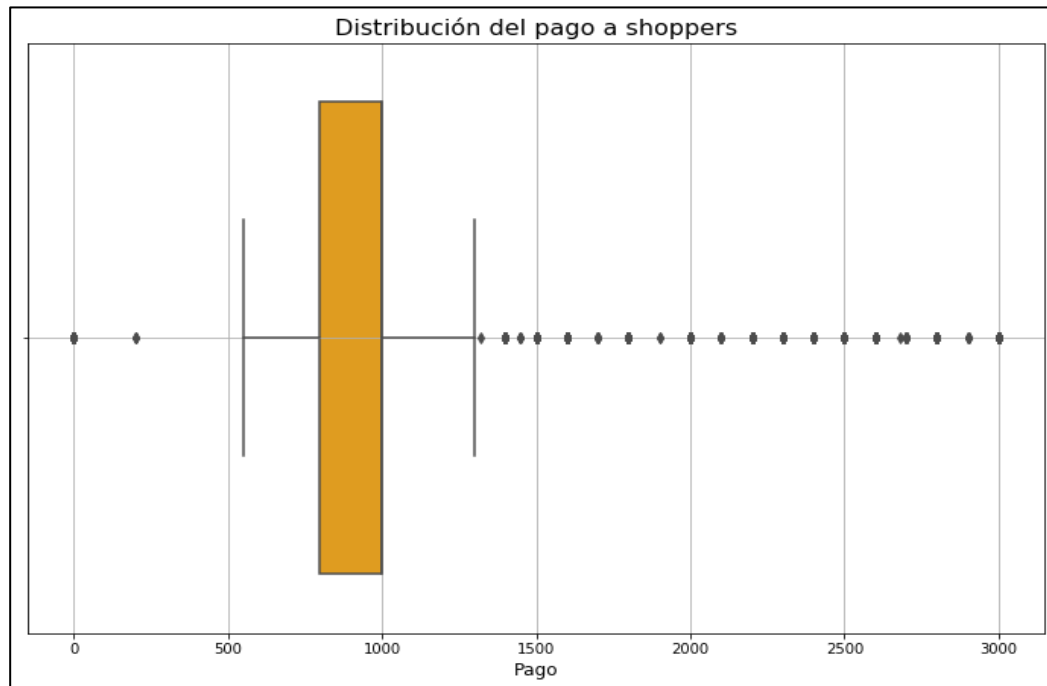


Ilustración 31: *Boxplot* del pago a shoppers.  
Fuente: Elaboración Propia.

Respecto a las nuevas variables generadas, se estudió como éstas se relacionaban con la aprobación de las misiones, para lo cual se usará el coeficiente de correlación de Pearson. La ecuación que define este indicador se presenta a continuación:

$$\rho_{x,y} = \frac{\theta_{x,y}}{\theta_x \theta_y}$$

$\theta_{x,y}$  = *Covarianza entre la variable X e Y*

$\theta_x$  = *Desviación estándar de la variable X*

$\theta_y$  = *Desviación estándar de la variable Y*

La correlación es en esencia una medida normalizada de asociación o covariación lineal entre dos variables. Sin embargo, es importante hacer notar que esta covariación no necesariamente implica causalidad, ya que la correlación puede ser fortuita. Un ejemplo típico de este último caso corresponde a la alta correlación encontrada entre el aumento en la temperatura global y la disminución del número de piratas. Estas variables no tienen ninguna relación causal y su alta correlación se explica por medio de una tercera variable, que corresponde simplemente al paso del tiempo. Es por esto que, el análisis de correlación debe ser complementado por otros procedimientos para validar sus resultados.

El coeficiente de correlación de Pearson puede variar entre -1 y 1, indicando correlaciones perfectas, negativa y positiva respectivamente. Una correlación con un valor igual a 0, indica que no existe relación lineal entre las variables [25].

El tamaño de la correlación puede interpretarse de la siguiente forma:

- Correlación baja:  $\rho \leq |0,3|$
- Correlación mediana:  $|0,3| < \rho \leq |0,5|$
- Correlación alta o fuerte:  $\rho > |0,5|$

En base a lo anterior, se procedió a generar un correlograma que permitiera cuantificar el grado de relación existente entre las variables, el cual puede apreciarse a continuación:



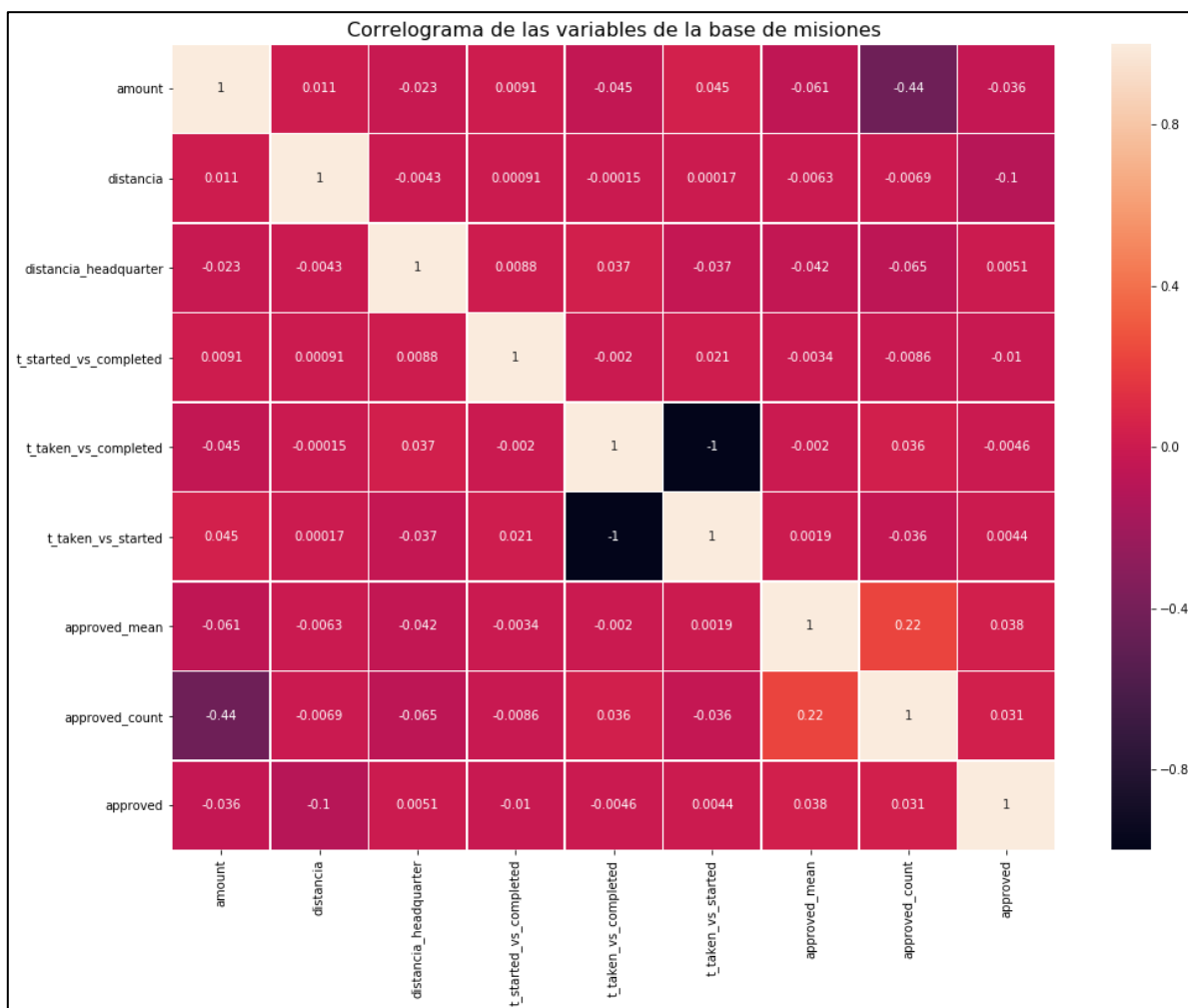


Ilustración 32: Correlación entre las variables de la base de datos de misiones.  
Fuente: Elaboración Propia.

Del correlograma, se aprecia que todas las variables estudiadas poseen una baja correlación con respecto a "approved", la cual indica si la misión fue aprobada o no. Es por esto que, para análisis posteriores solo se considerarán aquellas variables cuya correlación sea comparativamente más alta. Estas variables corresponden a: "amount", "distancia", "t\_started\_vs\_completed", "t\_taken\_vs\_completed", "approved\_mean" y "approved\_count".

### 8.3. Análisis exploratorio de la base de shoppers

Para este análisis se consideró a 1.293 shoppers que habían realizado parte de las 536.444 misiones expuestas anteriormente, utilizando la misma ventana de análisis (enero a diciembre 2019). Si bien la totalidad de las misiones fueron realizadas por 1.297 shoppers, los 4 restantes poseen poca información debido a gestiones deficientes de la información por parte de Rocketpin (no se guardó su información inicial, problemas en el servidor de la

página, entre otros). La falta de información puede considerarse totalmente aleatoria (MCAR), ya que en ningún caso se debe a comportamientos o características particulares de los shoppers, por lo cual el análisis no se vería sesgado. La base de datos de shoppers posee las siguientes variables, cuya descripción se presenta a continuación:

Variable	Descripción	Inválidos	Válidos	%Inválidos	Clase
shopper_email	Email del shopper	0	1.293	0%	object
id shopper	Llave única que identifica al shopper	0	1.293	0%	int64
Calificaciones	Promedio de calificaciones del shopper en la ventana de análisis	899	394	70%	float64
Score	Score del shopper	0	1.293	0%	float64
Cantidad_Misiones	Cantidad de misiones realizadas en la ventana de análisis	0	1.293	0%	int64
nombre	Nombre del shopper	1	1.292	0%	object
apellido	Apellido del shopper	168	1.125	13%	object
sexo	Género del shopper	17	1.276	1%	object
fecha nacimiento	Fecha de nacimiento del shopper	100	1.193	8%	object
fecha ingreso	Fecha en la que se registró el shopper en Rocketpin	0	1.293	0%	object
fecha confirmacinnn	Fecha de confirmación del registro del shopper	3	1.290	0%	object
Comuna	Comuna donde vive el shopper	0	1.293	0%	object

Ilustración 33: Descripción de las variables en la base de datos de shoppers.

Fuente: Elaboración Propia.

De la tabla anterior, se aprecia que la cantidad de valores nulos es despreciable para la mayoría de las variables, a excepción de "Calificaciones". Esta variable corresponde al promedio de la puntuación que recibe el shopper por sus misiones, la cual puede variar entre 1 a 5. Generalmente, las misiones son aprobadas cuando la calificación otorgada por el digitador es mayor a 3, y rechazadas en el caso contrario. La falta de información se debe a que esta política de asignar puntuaciones a las misiones es relativamente nueva, y se creó como una forma de calificar a los shoppers para dar mayor importancia a aquellos con buen rendimiento. Otra variable importante que explicar corresponde al "Score", la que indica el porcentaje de cumplimiento o confiabilidad del shopper desde que forma parte de la comunidad Rocketpin.

Teniendo un conocimiento general de la base, se estudia la creación de nuevas variables a partir de las existentes. En primer lugar, se modificaron aquellas variables relativas a fechas para que tuvieran la clase "datetime64", las que corresponden a: "fecha nacimiento", "fecha ingreso" y "fecha confirmacinnn". Posteriormente, utilizando como fecha de referencia el 31 de diciembre de 2019, se generaron variables que según la sección 8.2, podrían estar correlacionadas con el rechazo o aprobación de misiones como, por ejemplo: edad y experiencia del shopper. Estas se crearon calculando la diferencia entre la fecha de referencia y la fecha de nacimiento e ingreso del shopper a Rocketpin. Todas las variables nuevas pueden apreciarse en la siguiente ilustración:

Variable	Descripción	Inválidos	Válidos	%Inválidos	Clase
edad	Edad del shopper	100	1.193	8%	float64
experiencia	Experiencia en días del shopper en Rocketpin	3	1.290	0%	int64
female	Dummy que toma el valor 1 si el shopper es hombre, 0 sino	17	1.276	1%	float64
male	Dummy que toma el valor 1 si el shopper es mujer, 0 sino	17	1.276	1%	float64

Ilustración 34: Variables creadas a partir de la base de shoppers.

Fuente: Elaboración Propia.

Con respecto a la edad de los shoppers, cerca del 50% tiene entre 20 y 35 años, lo cual puede apreciarse en la siguiente ilustración:

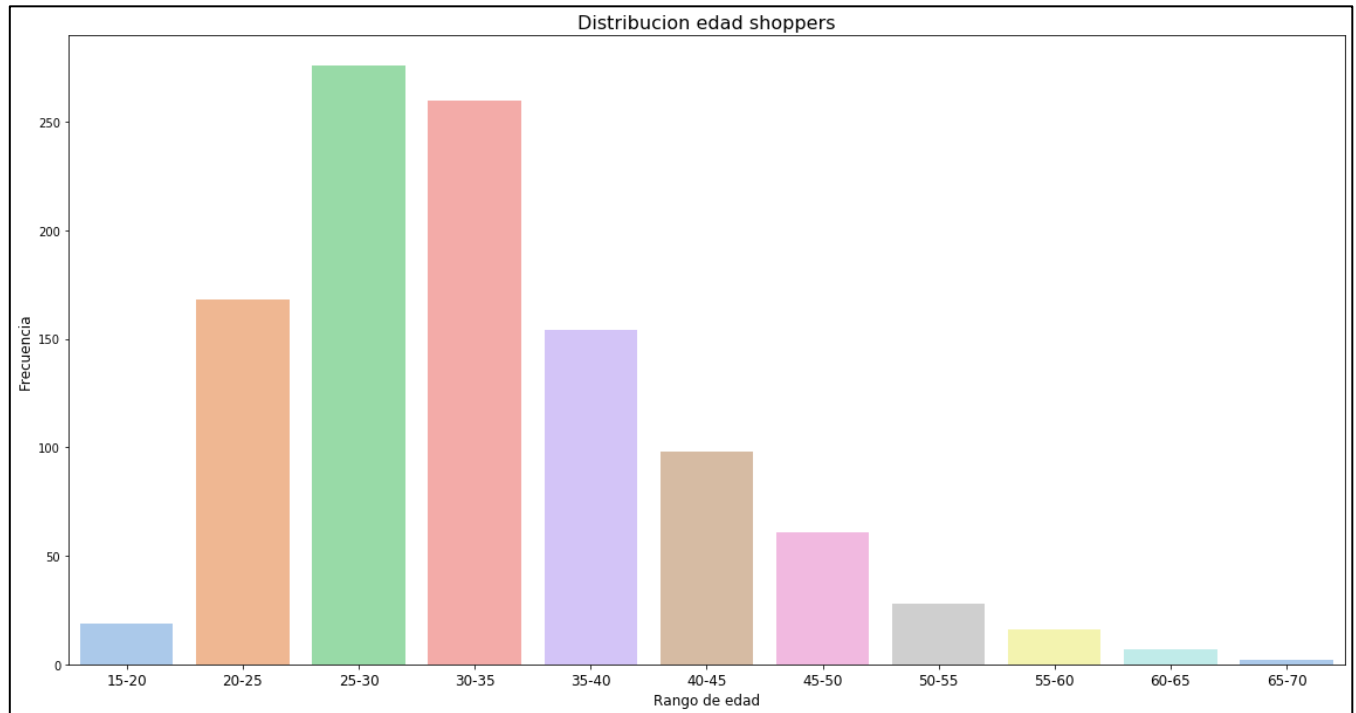


Ilustración 35: Distribución de la edad de los shoppers.

Fuente: Elaboración Propia.

Esto sugiere que la aplicación de Rocketpin es mayoritariamente ocupada por un público joven. El promedio de edad de los shoppers es de 32,65 años y la desviación estándar de la distribución es de 8,8 años, lo cual indica baja dispersión en esta variable (el *boxplot* de la edad sin considerar rangos puede apreciarse en la ilustración 77 del anexo de este informe). En cambio, la experiencia de los shoppers es bastante heterogénea, ya que el promedio de la distribución es de 413,77 días y su desviación estándar es de 277, por lo que el coeficiente de variación es igual a 67% ( $C_v = \frac{\theta}{\mu}$ , con  $\theta$  igual a la desviación estándar y  $\mu$  igual al promedio de la distribución). Esto puede apreciarse en el gráfico a continuación:

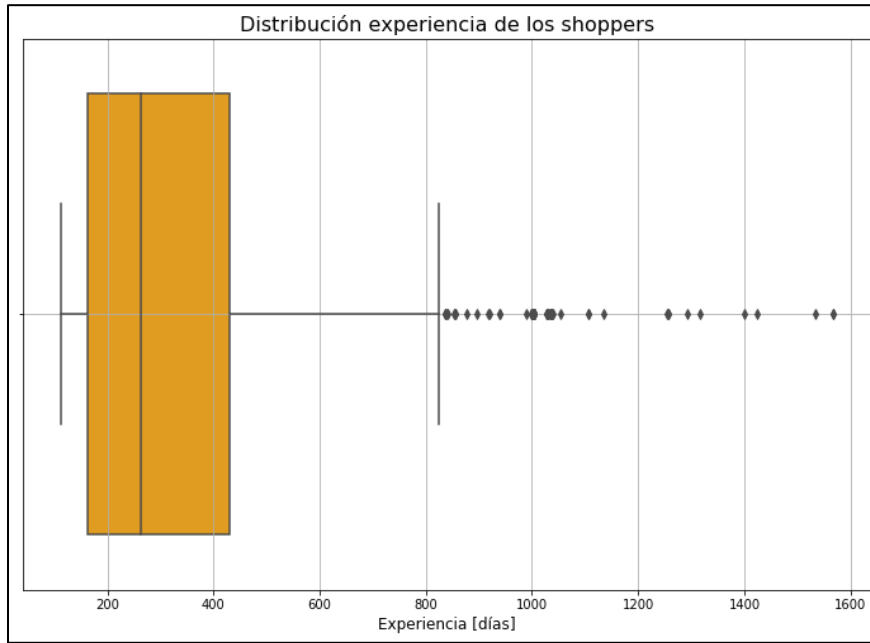


Ilustración 36: Distribución de la experiencia de los shoppers.  
Fuente: Elaboración Propia.

El 75% de las observaciones se encuentran entre 113 y 431 días de experiencia. Que la distribución sea heterogénea indica que existe alta variabilidad, por lo que los datos son muy dispersos y se alejan de la media. Esto permitiría sugerir que, solo considerando la experiencia como variable, existen grupos o segmentos de shoppers con comportamientos diferentes. Con respecto al género, se puede apreciar una clara predominancia de los hombres, ya que corresponden al 64,1% de los shoppers analizados. Esto puede apreciarse en la ilustración expuesta a continuación:

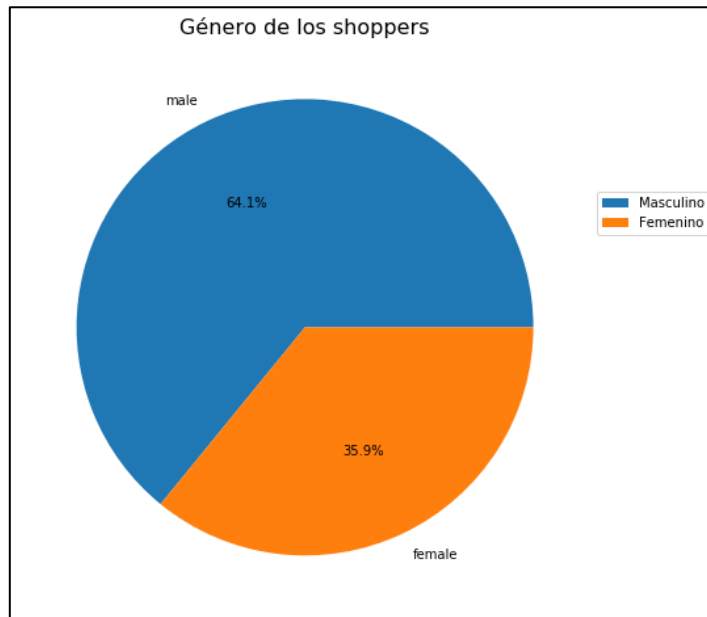


Ilustración 37: Distribución del género de los shoppers.  
Fuente: Elaboración Propia.

Para analizar la región de procedencia de los shoppers fue necesario utilizar una base de datos de la encuesta de caracterización socioeconómica nacional (CASEN), la cual es levantada por el ministerio de desarrollo social (MDS) [26]. Utilizando la variable "Comuna" proveniente de la base de shoppers se pudo unir esta con la CASEN, para obtener la región de Chile a la que pertenece cada comuna. Dado esto, se obtuvo que los shoppers provienen en su mayoría de la región Metropolitana (36.1%), seguido por las regiones de Bío Bío y Los Lagos (9,4% y 9,1% respectivamente). Lo anterior, se muestra gráficamente en la siguiente ilustración:

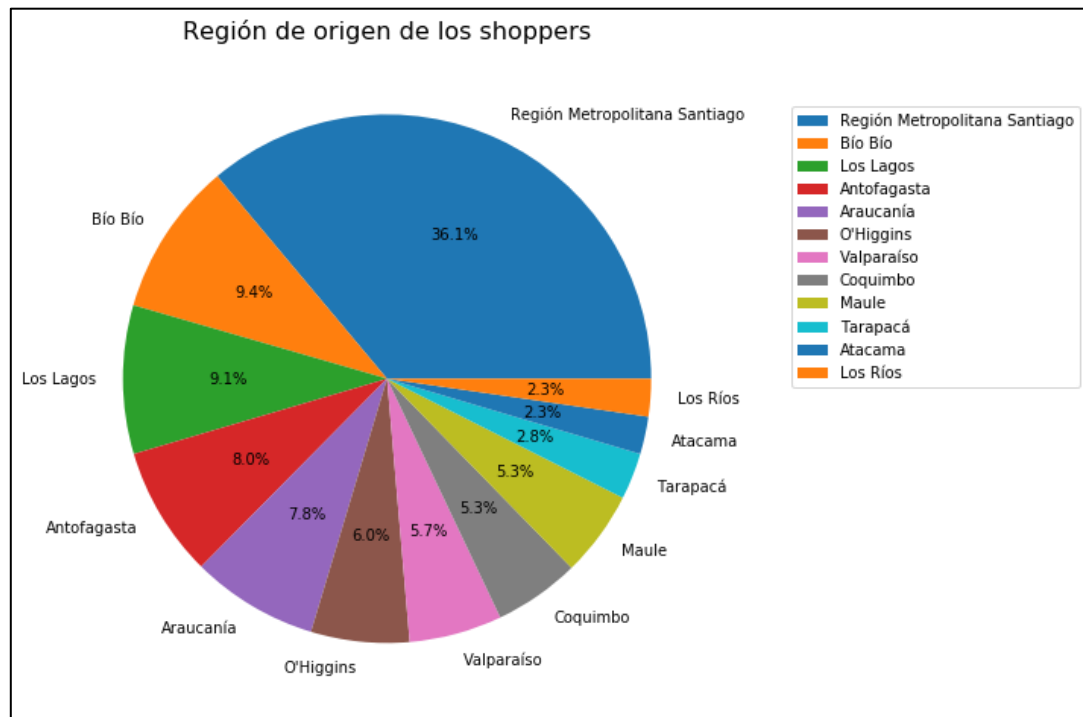


Ilustración 38: Distribución de las regiones de origen de los shoppers.  
Fuente: Elaboración Propia.

Esta grafica indica que, a pesar del claro predominio de la región Metropolitana, Rocketpin tiene presencia a lo largo de todo Chile. Finalmente, con relación a la confiabilidad o "Score" de los shoppers, es posible apreciar que la mayoría posee entre un 90 a 100%, seguido por el rango de 80 a 90%. Para Rocketpin es primordial tener shoppers con alta tasa de cumplimiento, ya que sus estándares de calidad son muy elevados (95% de misiones con información verídica). Dado esto, es que se hace primordial revisar aquellos casos con tasas inferiores al rango de entre 90-100% y en medida de lo posible que aquellos pertenecientes a ese rango, se acerquen o superen un cumplimiento de 95%.

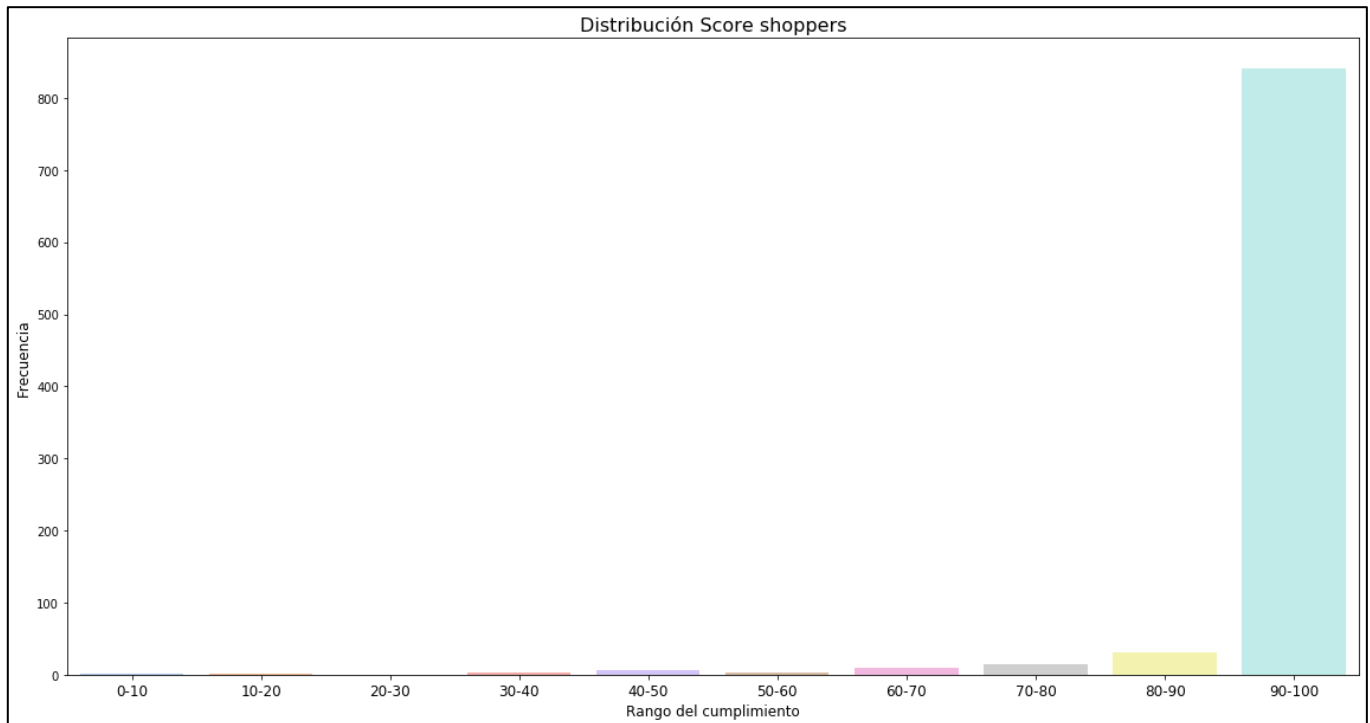


Ilustración 39: Distribución de la tasa de cumplimiento de los shoppers.  
Fuente: Elaboración Propia.

#### 8.4. Análisis conjunto de las bases de datos

Esta etapa consiste en unir las bases de datos previamente analizadas y generar un entendimiento de las misiones a nivel global, considerando sus características intrínsecas y las de los shoppers que las realizan. Además, este análisis permitirá determinar qué características de los shoppers influyen en mayor medida en la aprobación de las misiones.

Para generar esta base conjunta fue necesario unir ambas bases de datos utilizando aquella variable que poseen en común, la cual corresponde a "shopper\_email". La base resultante posee 536.444 observaciones y 65 variables. El hecho de que la cantidad de observaciones de la nueva base sea igual a la de misiones, se debe a que se realizó un "left join". Esto puede entenderse gráficamente por medio de la siguiente ilustración:

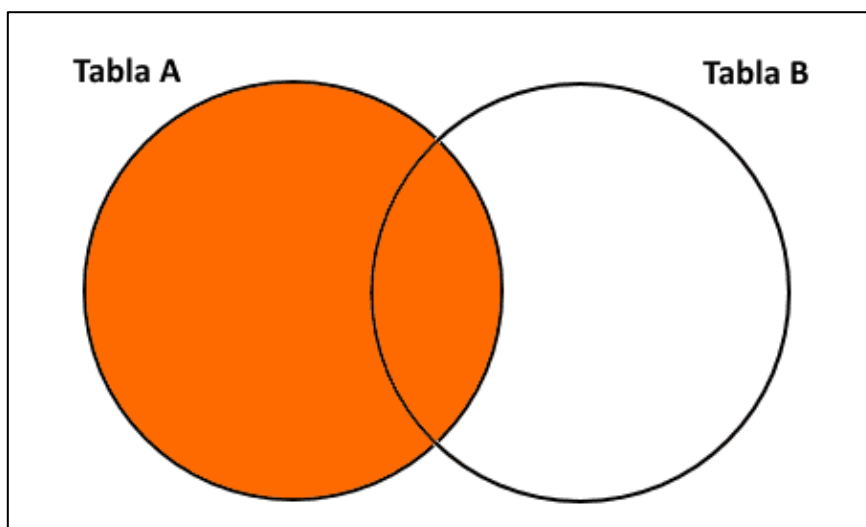


Ilustración 40: Ejemplificación de un *left join* mediante la teoría de conjuntos.  
Fuente: Elaboración Propia.

Al unir estas bases de datos, las variables de la base de la derecha se incorporan a la de la izquierda, en la medida que coincida la variable en común que poseen (en este caso, "shopper\_email"). Es decir, si es que los emails de los agentes en la base de misiones y de shoppers coinciden, las variables de esta última se incorporarán a la primera y de no ser así las nuevas variables se agregarán con valores nulos. Dado esto, se puede estudiar la cantidad de valores inválidos de las nuevas variables agregadas a la base de misiones, lo cual se detalla a continuación:

Variable	Inválidos	Válidos	%Inválidos
Calificaciones	151.984	384.460	28%
Score	44.444	492.000	8%
Cantidad_Misiones	44.444	492.000	8%
id shopper	44.444	492.000	8%
email	44.444	492.000	8%
nombre	44.445	491.999	8%
apellido	60.728	475.716	11%
sexo	45.642	490.802	9%
fecha nacimiento	64.648	471.796	12%
fecha ingreso	44.444	492.000	8%
fecha confirmacinnn	46.870	489.574	9%
Comuna	44.444	492.000	8%
edad	64.648	471.796	12%
experiencia	46.870	489.574	9%
female	45.642	490.802	9%
male	45.642	490.802	9%

Ilustración 41: Análisis de las nuevas variables agregadas a la base de misiones.  
Fuente: Elaboración Propia.

La mayoría de las nuevas variables tiene una tasa cercana al 10% de valores nulos, a excepción de la variable "Calificaciones", lo cual se debe a que esta medida se implementó hace poco tiempo.

Al analizar las variables de la base de shoppers a nivel de misiones, se pueden estudiar comportamientos que antes no. Por ejemplo, la siguiente ilustración muestra la distribución de cantidad de misiones realizadas por edad y género del shopper:

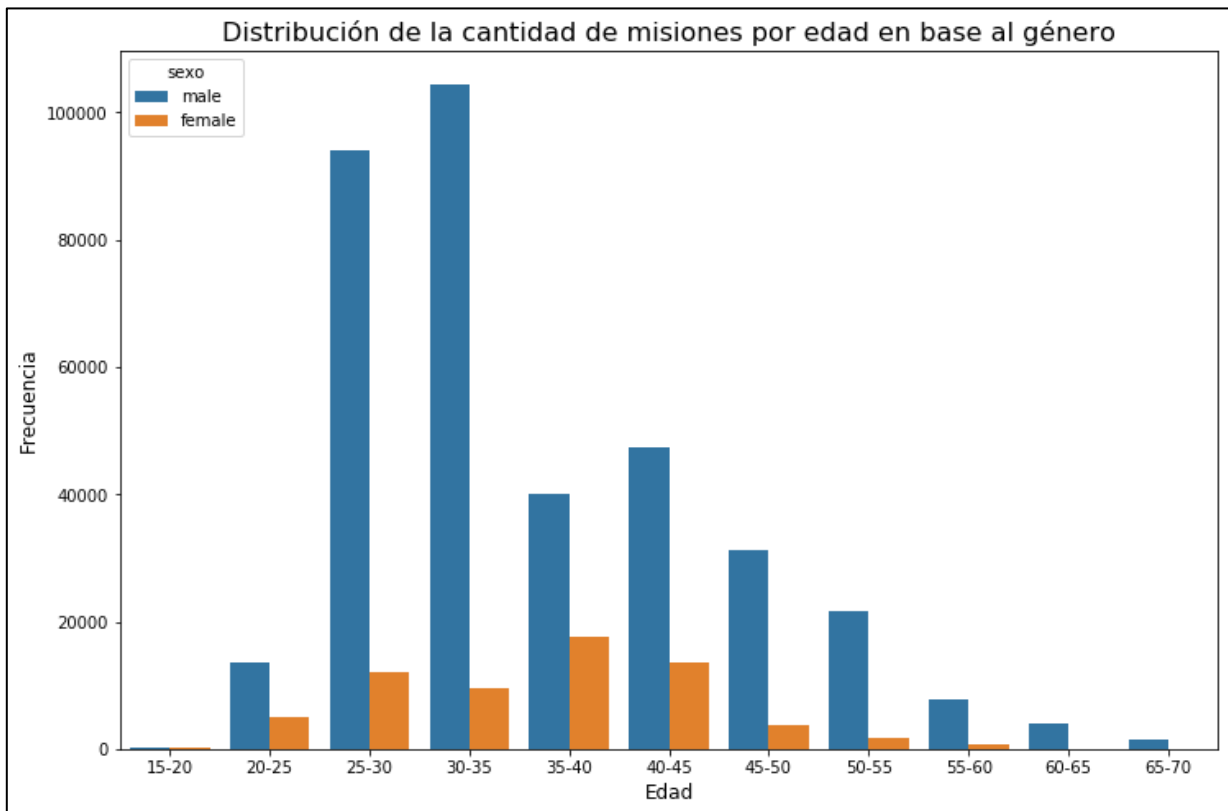


Ilustración 42: Distribución de las misiones por edad y género del shopper.  
Fuente: Elaboración Propia.

De ésta se extrae que, la mayoría de las misiones son realizadas por hombres. Los rangos de edad en los hombres que más realizan misiones son 25-30 y 30-35 años. En cambio, para las mujeres los rangos son de entre 35-40 y 40-45 años. Este hallazgo es interesante ya que, si se encontrará una relación positiva entre la edad y el rechazo de una misión, se podría establecer que algún género tiene predominancia a cometer más errores. Sin embargo, al analizar la cantidad de misiones aprobadas y rechazadas por rango de edad, no se aprecia ninguna relación entre el rechazo de la misión y la edad del shopper, lo cual puede notarse en la siguiente ilustración:



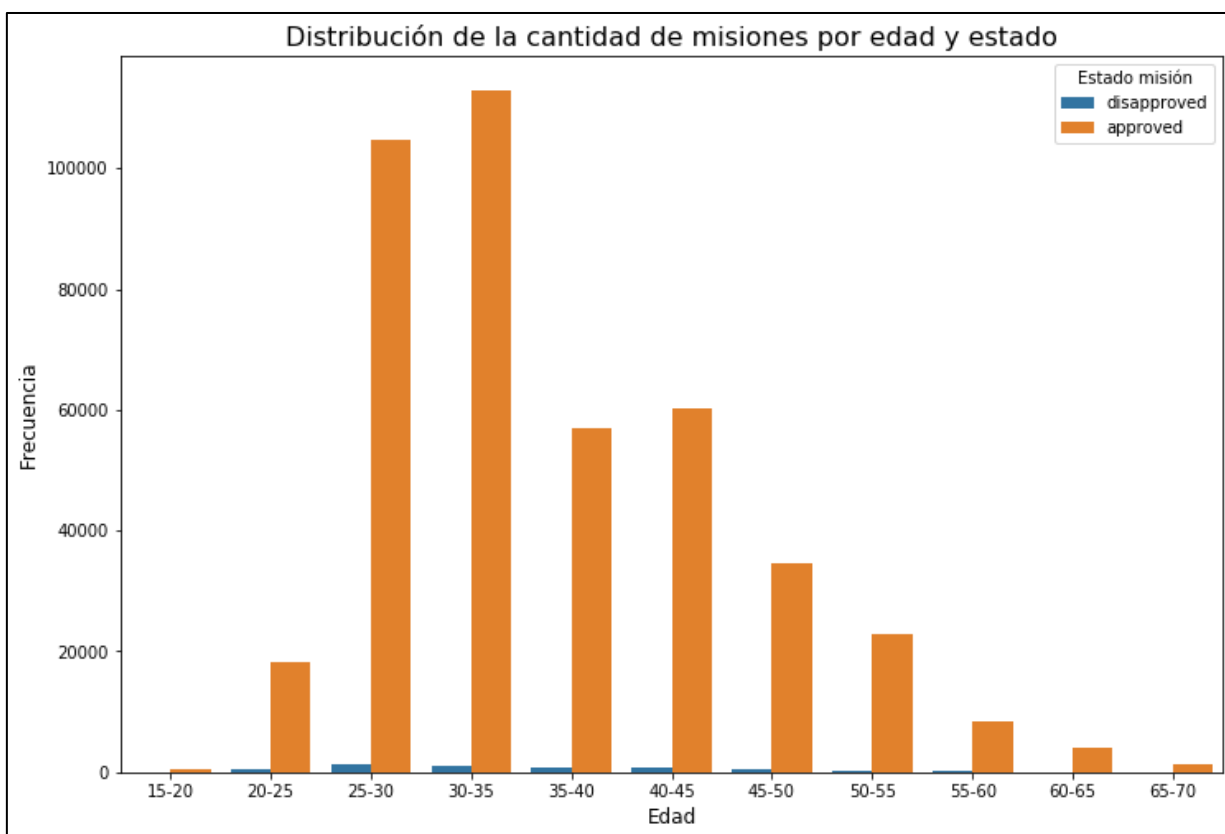


Ilustración 43: Distribución de las misiones por edad y estado de la misión.  
Fuente: Elaboración Propia.

Al igual que en la sección 8.2, se estudiará la relación entre estas nuevas variables y la aprobación de misiones, mediante la correlación de Pearson, lo cual puede apreciarse en la ilustración a continuación:

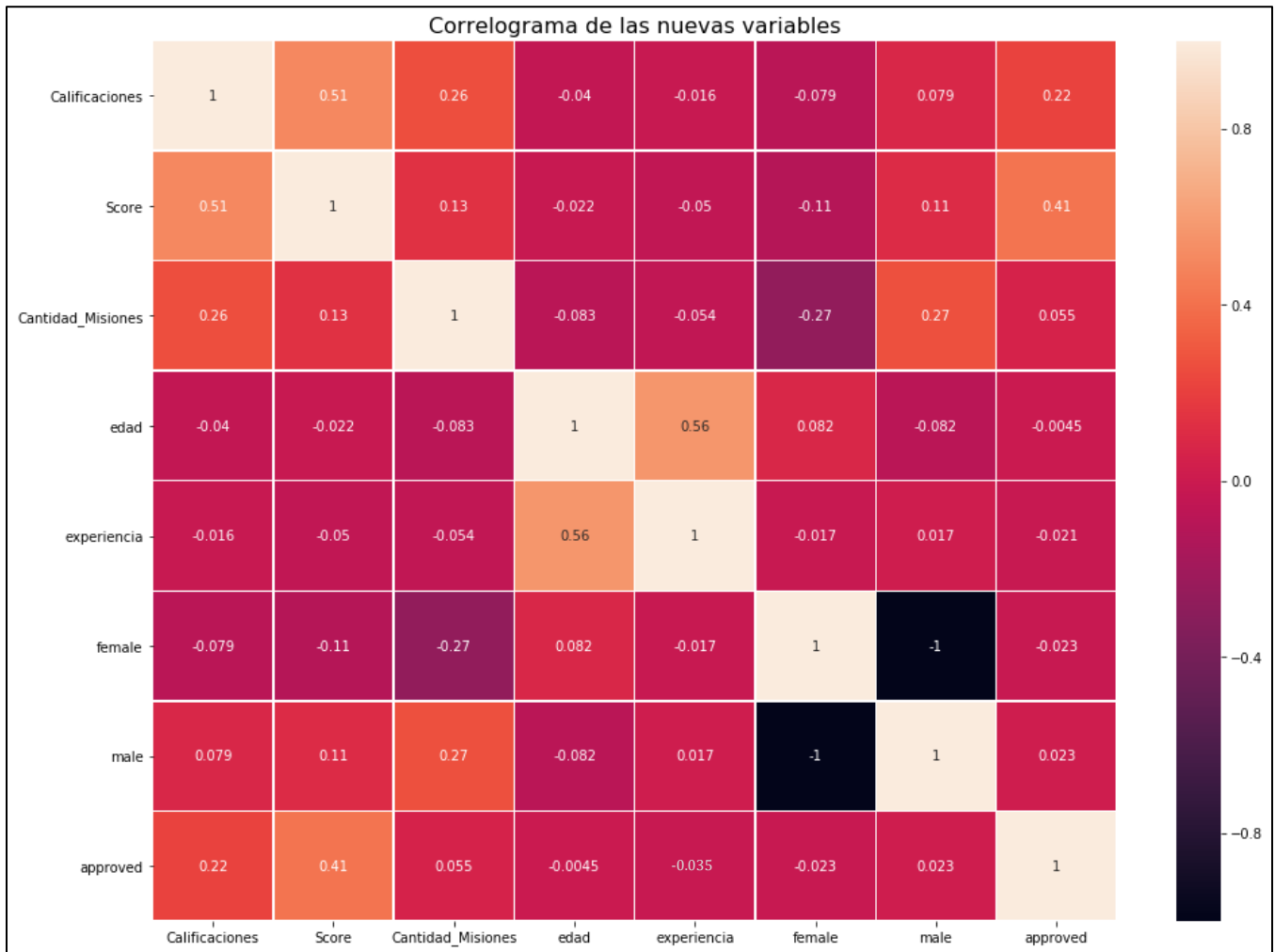


Ilustración 44: Correlograma de las variables de shoppers incluidas en la base de misiones.  
Fuente: Elaboración Propia.

Las variables que con un coeficiente de correlación más alto con respecto a "approved" corresponden a: "Calificaciones", "Score", "Cantidad\_Misiones" y "experiencia". Además, tanto la edad como el género parecen no influir en la aprobación de misiones. Dado esto, solo se seguirán estudiando aquellas variables que posean mayor correlación con la aprobación, las cuales son: "amount", "distancia", "t\_started\_vs\_completed", "t\_taken\_vs\_completed", "approved\_mean", "approved\_count", "Calificaciones", "Score", "Cantidad\_Misiones" y "experiencia"

## 9. CONTROL DEL PROCESO DE REVISIÓN

El control de procesos se ocupa de vigilar la calidad mientras se produce el producto o servicio. Los objetivos típicos de los planes de control de procesos son proporcionar información oportuna sobre si la producción en ese momento cumple con las especificaciones de diseño y detectar cambios en el proceso

que indiquen que es probable que productos futuros no cumplan con esas especificaciones [27].

La principal causa que podría hacer que un producto no cumpla con las especificaciones en el futuro, consiste en la variabilidad de los atributos de este. A pesar de que la calidad se relacione directamente con el concepto de uniformidad y que mientras menor varianza tenga la producción mayor es el aseguramiento de ésta, se entiende que no es posible que esto ocurra en el 100% de los casos y, por ende, se dice que un producto tiene calidad y que un proceso está controlado, cuando sus características están dentro de los límites de tolerancia superior e inferior de la especificación [9, p. 156]. Para entender la variabilidad del proceso de revisión de misiones, se estudiarán las variables más influyentes en la aprobación y/o rechazo de éstas. Esto ayudará a entender los límites aceptables en los que deben estar estas variables y a verificar si el proceso está controlado o no.

## **9.1. Gráficos de control**

### **9.1.1. Uso de gráficas por atributo: tipo P**

Este tipo de gráfica es útil cuando no se desea controlar el valor de una magnitud medible, sino simplemente si el producto es adecuado o no lo es; o, en general, si posee o no cierto atributo. Las ventajas que otorga este tipo de medición versus el control por variables son las siguientes:

- Suele ser más sencillo y rápido. Por tanto, es más económico.
- Permite resumir las características de varias variables. Un artículo o servicio puede ser defectuoso o no dependiendo de un conjunto de variables y no de una sola. No se controla una característica medible sino la ausencia o presencia de un atributo (rechazo/ no rechazo).

La monitorización de un proceso a través de este tipo de mediciones se denomina control por atributos. Existen varios gráficos que permiten estudiar la evolución de este tipo de información, sin embargo, en el caso de esta memoria, se utilizará aquel que observa la variación de la proporción de productos defectuosos en sucesivas muestras de tamaño  $n$  [28].

El caso que se estudiará con este tipo de gráfica corresponde al rechazo o aprobación de las misiones. Se podría considerar el estado de la misión, como el atributo a controlar, ya que permite estudiar si el producto (misión) está defectuoso (rechazada) o no (aprobada) y la variación de esta proporción en las muestras escogidas. Para elegir la muestra adecuada, el tamaño de ésta

debe ser lo suficientemente grande para permitir el conteo del atributo. Una regla práctica al crear una gráfica P es hacer que la muestra tenga el tamaño suficiente para esperar contar el atributo dos veces en cada muestra. En este caso, como la tasa de rechazos es del 9,6%, la muestra debería tener un tamaño 20,83 misiones. Dado que se tiene una base cuenta con un total de 536.444 misiones, se aumentará la muestra para que el efecto a estudiar pueda contarse al menos 20 veces, lo cual indicaría que el tamaño de la muestra debe ser de aproximadamente 200 misiones.

Para cumplir más adecuadamente con la función de predecir cuantos productos del proceso van a satisfacer las especificaciones fue creado  $C_p$ , llamado índice de capacidad potencial del proceso, que mide la dispersión permitida por la medida de la real dispersión del proceso. Esta queda definida de la siguiente forma:

$$C_p = \frac{LCS - LCI}{6\theta_p}$$

Donde  $\theta_p$ , corresponde a la desviación estándar del proceso. Se asume una dispersión de  $6\theta_p$ , lo que representa en la teoría normal, la anchura del intervalo que contiene al 99,73% de la población (lo cual también es conocido como *six sigma*) [29]. Las variables  $LCS$  y  $LCI$  corresponden al límite superior e inferior de la gráfica P respectivamente. Dado lo anterior y como se señaló en el marco conceptual de esta memoria, se calcularían asumiendo un  $z = 3$  según la distribución normal, puesto que permite abarcar entre los límites de control, el 99,73% del total de casos. La evaluación del proceso a través del uso del índice de capacidad se hace de la siguiente manera:

$C_p \geq 2,0 \Rightarrow$  *proceso excelente, altamente confiable.*

$1,33 \leq C_p < 2,0 \Rightarrow$  *proceso capaz, relativamente confiable.*

$1,0 \leq C_p < 1,33 \Rightarrow$  *proceso relativamente incapaz, poco confiable.*

$0 \leq C_p < 1,0 \Rightarrow$  *proceso incapaz, pudiendo tener producción defectuosa.*

$C_p < 0 \Rightarrow$  *proceso totalmente incapaz.*

Utilizando todos los conceptos anteriores, se procedió a generar el gráfico de control del tipo P para la proporción de misiones aceptadas y/o rechazadas, utilizando un total de 80 muestras con 200 observaciones cada una, lo cual puede apreciarse en la ilustración presentada a continuación:

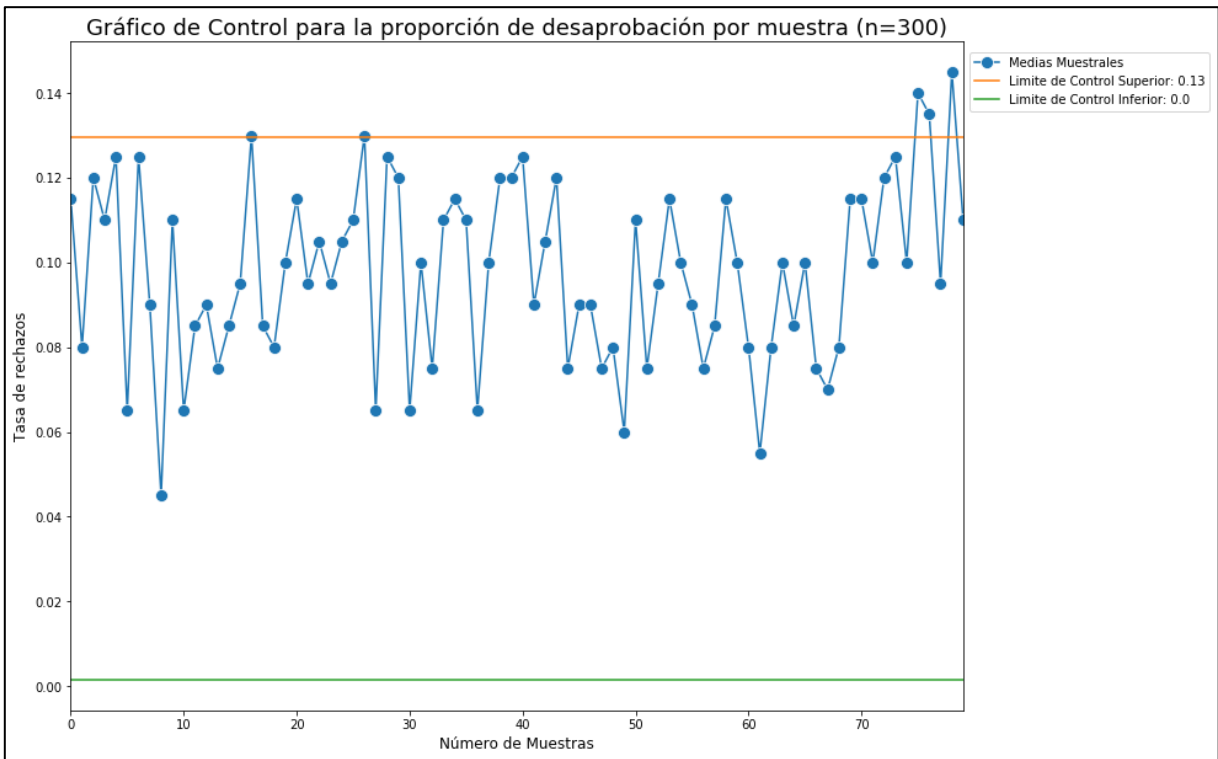


Ilustración 45: Gráfico del tipo P para la tasa de desaprobación de misiones.  
Fuente: Elaboración Propia.

De este puede notarse que la tasa de rechazos se encuentra con un 99,73% de confianza entre un 1% y 13%, siendo superado el límite superior en múltiples ocasiones. La desviación estándar del proceso es de 2,13%. Si se calcula el índice de capacidad del proceso éste da un total de  $C_p = \frac{13\% - 0\%}{6 * 2,13\%} = \frac{13\%}{12,78\%} = 1,017$ . En base a este resultado, puede decirse que el proceso de revisión de misiones es relativamente incapaz y poco confiable. Puesto que se determinó que el proceso no es apto para producir misiones aptas bajo las especificaciones establecidas, se reafirma la necesidad de adoptar la medida de mejorar y/o rediseñar el proceso. Finalmente, los límites obtenidos podrían servir como indicador para supervisar las labores de los digitadores, ya que se conoce el rango de la tasa de aprobación (o rechazo) que debería tener el lote que diariamente revisa cada uno.

### 9.1.2. Uso de gráficas por variables: tipo X y R

Las gráficas de control del tipo X y R se utilizan generalmente cuando se quiere realizar una medición de variables continuas. En el muestreo por atributos, se determina si algo es bueno o malo, si queda bien o no; se trata de una situación de seguir o no. Sin embargo, en la medición de variables, se mide cualquier variable real, y se desarrollan gráficas de control para determinar el

grado de aceptación o rechazo del proceso, con base en esas mediciones. Por ejemplo, en el muestreo por atributos, podría decirse que se rechaza una misión si es que la duración de ésta es mayor a 10 minutos. En el muestreo por variables, se mide una muestra y se registran duraciones de 9 y 11 minutos. Estos valores se usan para crear o modificar las gráficas de control y saber si se encuentran dentro de los límites aceptables. Los 4 aspectos más importantes a la hora de generar gráficos de control por variables corresponden a:

- I. Tamaño de las muestras:** El tamaño de las muestras preferido para aplicaciones industriales es de entre 4 a 5 unidades. Las mayores de 5 dan límites de control más estrechos y, por lo tanto, mayor sensibilidad. De hecho, para detectar variaciones más finas de un proceso, quizá es necesario utilizar muestras más extensas. En el caso a estudiar, se utilizarán muestras de 15 unidades, ya que es el límite superior en el que se hace mejor ocupar gráficas del tipo X con rango R versus gráficas del tipo X con desviación estándar  $\alpha$ . Dado un  $n = 15$ , según la ilustración 71 del anexo de este informe, se debe considerar para la construcción de las gráficas del tipo X y R, un  $A_2 = 0,22$ ,  $D_3 = 0,35$  y  $D_4 = 1,65$ .
- II. Número de muestras:** Una vez creada la gráfica, es posible comprar cada muestra con los límites de control y tomar una decisión sobre si el proceso es aceptable. Sin embargo, para elaborar las gráficas, la prudencia y las estadísticas sugieren que se tomen alrededor de 25 muestras. Para el caso actual y dado que la base de datos de misiones es bastante extensa, se tomarán cerca de 50 muestras por variable.
- III. Frecuencia de las muestras:** La frecuencia con la que es necesario tomar las muestras depende del costo del muestreo y el beneficio de ajustar el sistema. En este caso, dado que se analizará una base estática para una ventana de análisis dada, este aspecto no aplica.
- IV. Límites de control:** Una práctica estándar, en el control estadístico de proceso para las variables, es establecer límites de control tres desviaciones estándar sobre la media y tres desviaciones estándar debajo de esta. Esto significa que se espera un 99,73% de las medias de la muestra caigan dentro de los límites de control (es decir, un intervalo de confianza de 99,73%).

Contando con un entendimiento de los conceptos anteriores, se procede a generar gráficas de control del tipo X y R para cada una de las variables con mayor correlación con la aprobación y/o rechazo de misiones, lo cual se estudió en la sección 8.4 de esta memoria. Para la variable que indica la

distancia entre el lugar donde el shopper inicia la misión versus donde la termina, se generaron los siguientes gráficos de control:

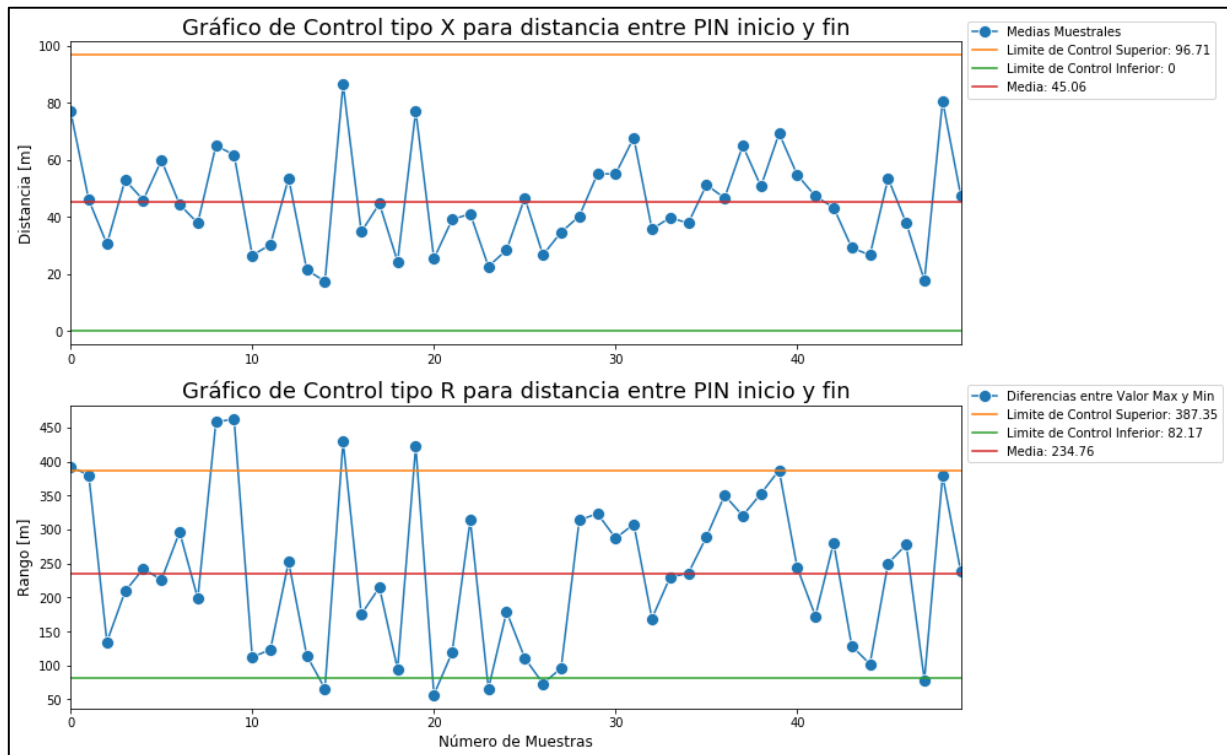


Ilustración 46: Gráficos del tipo X y R para la distancia entre inicio y fin de una misión.  
Fuente: Elaboración Propia.

El gráfico de control del tipo X, señala que el proceso está controlado, ya que se encuentra bajo los límites de control. El límite de control superior de las muestras es igual a 96,71 metros y el inferior igual a 0, donde la media corresponde a 45,06 metros. Esto indica que la distancia entre el lugar donde el shopper inicia la misión y donde la termina no debería ser superior a aproximadamente 100 metros, lo cual es bastante acertado si se considera que la restricción impuesta por Rocketpin es de un máximo de 200 metros.

Al analizar el gráfico de control del tipo R, se puede apreciar que existen varios puntos que sobrepasan tanto el límite de control superior como inferior. Esta gráfica representa la diferencia entre el punto máximo y el mínimo de las muestras (también llamado rango), por lo que representa una medida de la variabilidad del proceso. En muchas ocasiones se analiza la variabilidad a través del rango, en lugar de la desviación típica normal, lo cual se debe a la simplicidad de su cálculo [30]. Se puede apreciar que los límites de control para este gráfico son superiores que para el del tipo X, lo cual sugiere que las muestras presentan una alta dispersión y como bien se ha señalado anteriormente, la alta variabilidad es un síntoma de deficiencias en la calidad.

Dado esto, se sugiere que la distancia entre los puntos de inicio y fin de una misión corresponde a una variable controversial, a la cual debería ponerse un mayor énfasis para vigilar el correcto funcionamiento del negocio de Rocketpin.

Respecto a las calificaciones promedio que tienen los shoppers en las misiones, se generaron los siguientes gráficos de control:

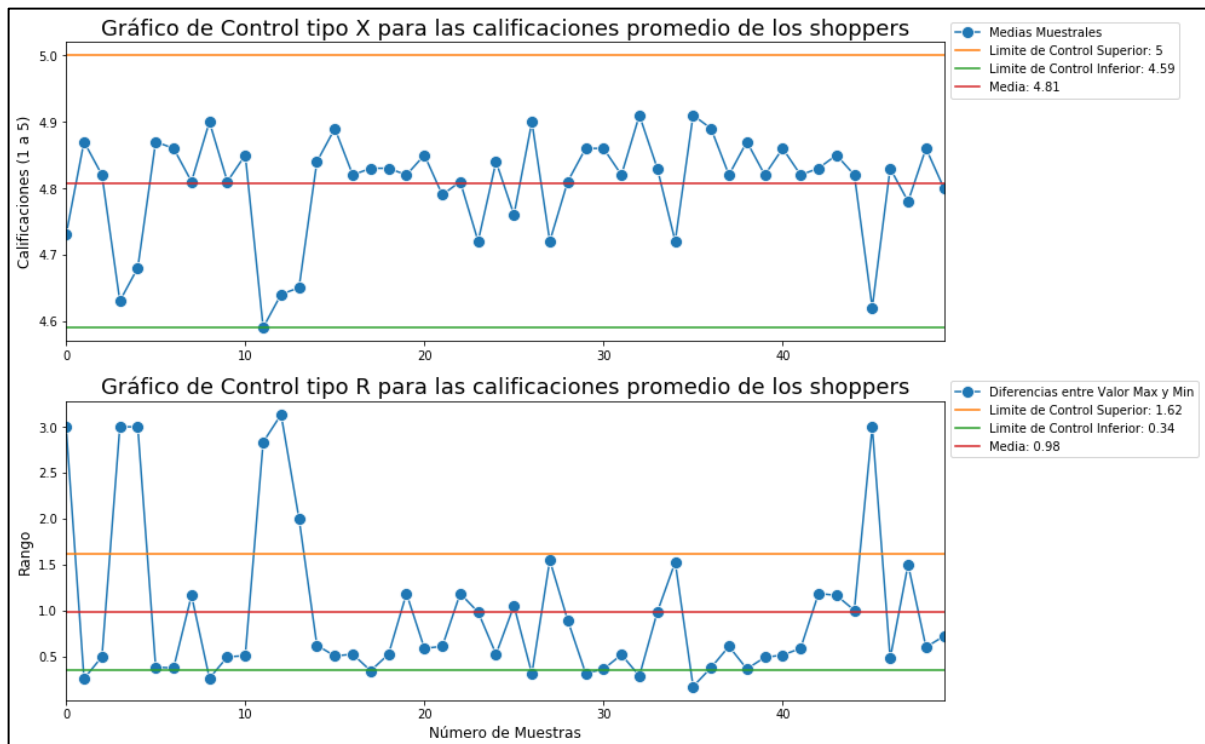


Ilustración 47: Gráficos del tipo X y R para las calificaciones promedio de los shoppers.  
Fuente: Elaboración Propia.

El límite superior para la media corresponde a la máxima calificación que puede tener una misión, es decir, una nota igual a 5 y que el límite inferior es igual a 4,6. Esto sugiere que la mayoría de los shoppers poseen calificaciones bastante elevadas en sus misiones, lo que valida los estrictos estándares de calidad que debe tener Rocketpin con sus clientes. Sin embargo, de la gráfica de tipo R, se puede notar que esta variable tiene una alta dispersión aparente, ya que varios puntos se alejan de los límites de control. Esto quiere decir que las muestras representan grupos heterogéneos, donde hay shoppers con muy buenas calificaciones y otros con un desempeño inferior, lo cual explicaría los puntos extremos. De lo anterior, se extrae que esta variable debe vigilarse para controlar la revisión de misiones y que además se debe buscar homogeneizar el desempeño de los shoppers para disminuir la variabilidad del proceso (quizá capacitándolos de mejor manera y/o prefiriendo perfiles específicos que se adapten a las necesidades de la empresa).



El comportamiento de las calificaciones es bastante similar al encontrado en los gráficos de control para el *Score* o índice de confiabilidad de los shoppers:

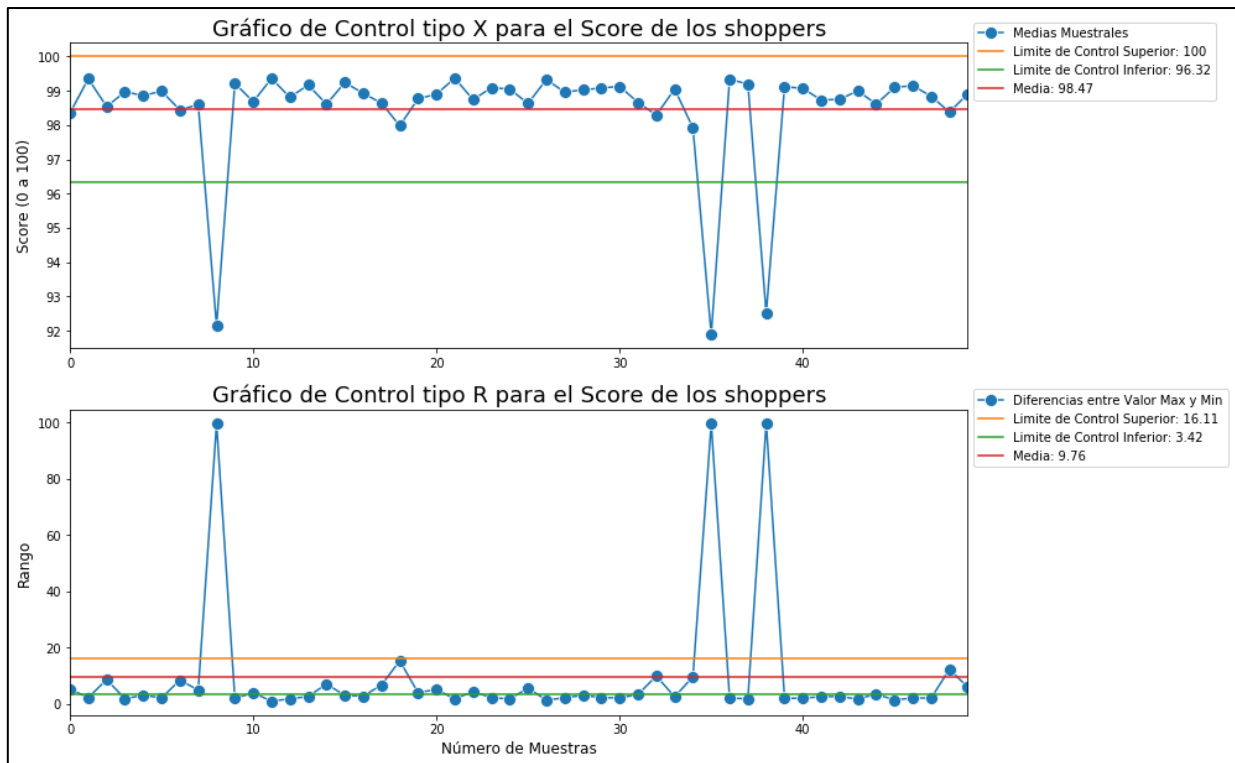


Ilustración 48: Gráficos del tipo X y R para el Score de los shoppers.  
Fuente: Elaboración Propia.

De la ilustración se extrae que la mayoría de las muestras señalan que los shoppers tienen altos índices de confiabilidad, ya que tanto los límites de control como la media tienen valores bastantes cercanos (*LCS* igual a 100, *LCI* igual a 96,32 y media igual a 98,47). Sin embargo, en ambas gráficas existen puntos que sobrepasan los límites de control, lo cual señala que existen shoppers de baja confiabilidad que desestabilizan el proceso. Esto puede apreciarse en el gráfico del tipo R, ya que los puntos que se alejan de los límites tienen un valor igual a 100, lo cual señala que en esas muestras existen shoppers con un índice de confiabilidad igual a 0.

Para la duración de la misión, los gráficos de control son los siguientes:

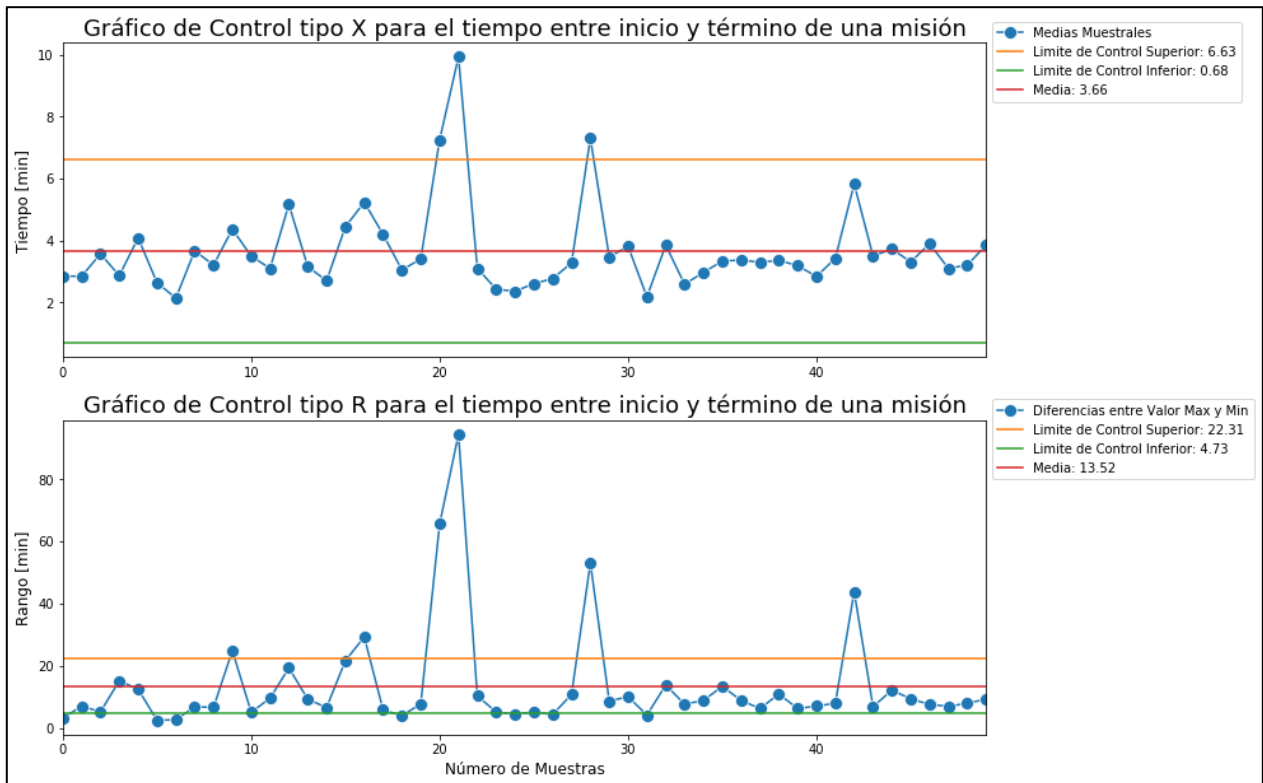


Ilustración 49: Gráficos del tipo X y R para la duración de una misión.  
Fuente: Elaboración Propia.

De la gráfica del tipo X anterior, se desprende que en promedio una misión de verificación domiciliaria dura aproximadamente 4 minutos. El límite de control inferior corresponde a 0 minutos (lo cual es improbable, pero ninguna de las misiones alcanza ese punto). La máxima duración que debería tener una misión es cercana a 7 minutos, sin embargo, existen puntos que sobrepasan este valor. Se extrae de este gráfico, que las verificaciones domiciliares son relativamente cortas, lo cual explica por qué en promedio los shoppers realizan 413 misiones en un año (1.297 shoppers realizan 536.444 misiones). En la gráfica de rangos se vislumbra que, existen muestras que poseen misiones con diferencias en su duración de aproximadamente 100 minutos, lo cual es preocupante si se considera que la duración promedio es de 4 minutos. Esto podría deberse a complicaciones particulares o aleatorias, ya que no se aprecia un patrón claro en los puntos que sobrepasan el límite de control superior. A pesar de esto, se sugiere que la duración de las misiones es una variable que debe controlarse en mayor medida, ya que su dispersión es bastante alta.

Para analizar el tiempo entre que se toma e inicia una misión, se crearon los siguientes gráficos de control:

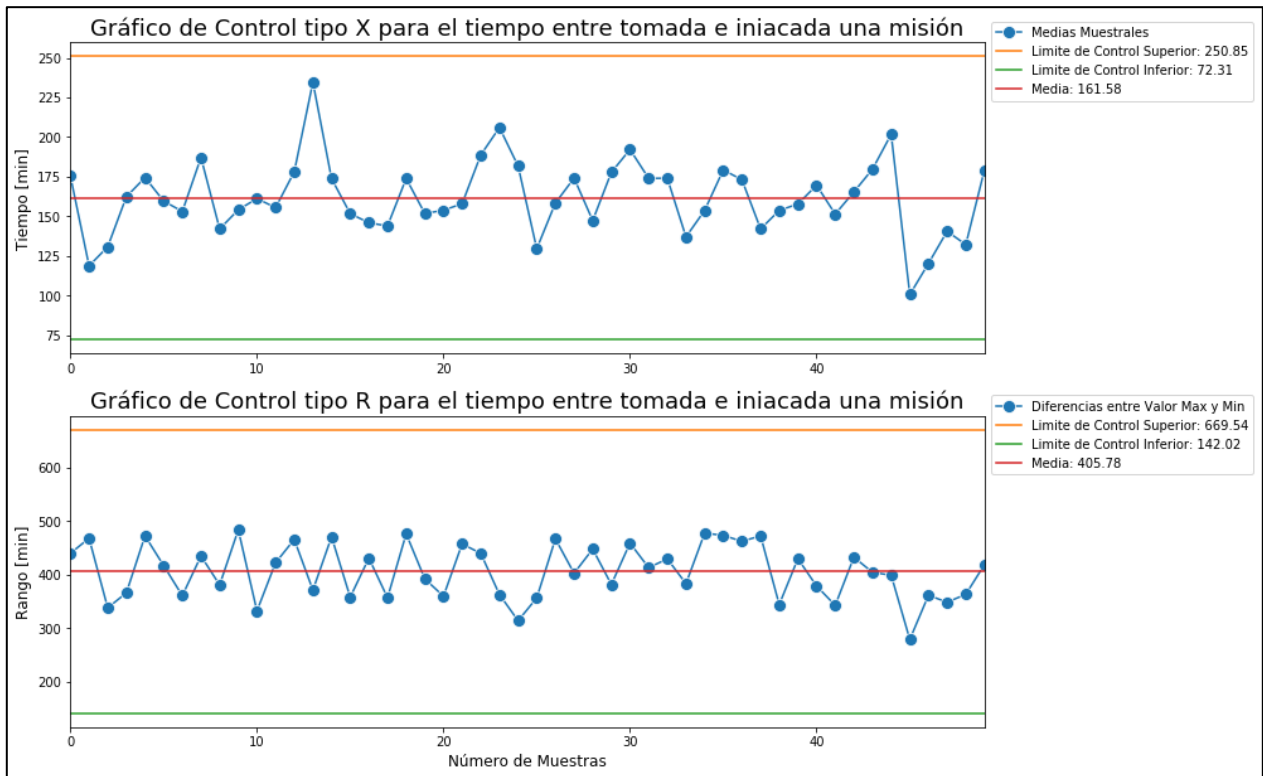


Ilustración 50: Gráficos del tipo X y R para el tiempo entre que se tome e inicia una misión.  
Fuente: Elaboración Propia.

De ambas gráficas se evidencia que bajo esta variable el proceso estaría bajo control, ya que en ninguna se sobrepasan los límites de control. El tiempo promedio entre que se toma e inicia una misión corresponde a 161,58 minutos, el límite de control inferior es de 72,31 minutos y el superior de 250,85 minutos. Este último valor es el que genera complicaciones, puesto que según lo estipulado por Rocketpin, las misiones que fueron tomadas deberían ser iniciadas antes de 3 horas, lo cual es bastante inferior al LCS (aproximadamente 4 horas y 15 minutos). Además, se estipula que sobre las 3 horas las misiones son rechazadas, por lo cual se aprecia claramente un problema en la realización de las misiones, supervisión de los shoppers o digitadores. Esto se comprueba al mirar la gráfica del tipo R, ya que en promedio la diferencia entre el mínimo y máximo valor de las muestras es de 405,78 minutos, valor muy cercano a 7 horas.

Con respecto al desempeño de los shoppers en sus misiones realizadas en los últimos 3 meses, los gráficos de control se muestran a continuación:

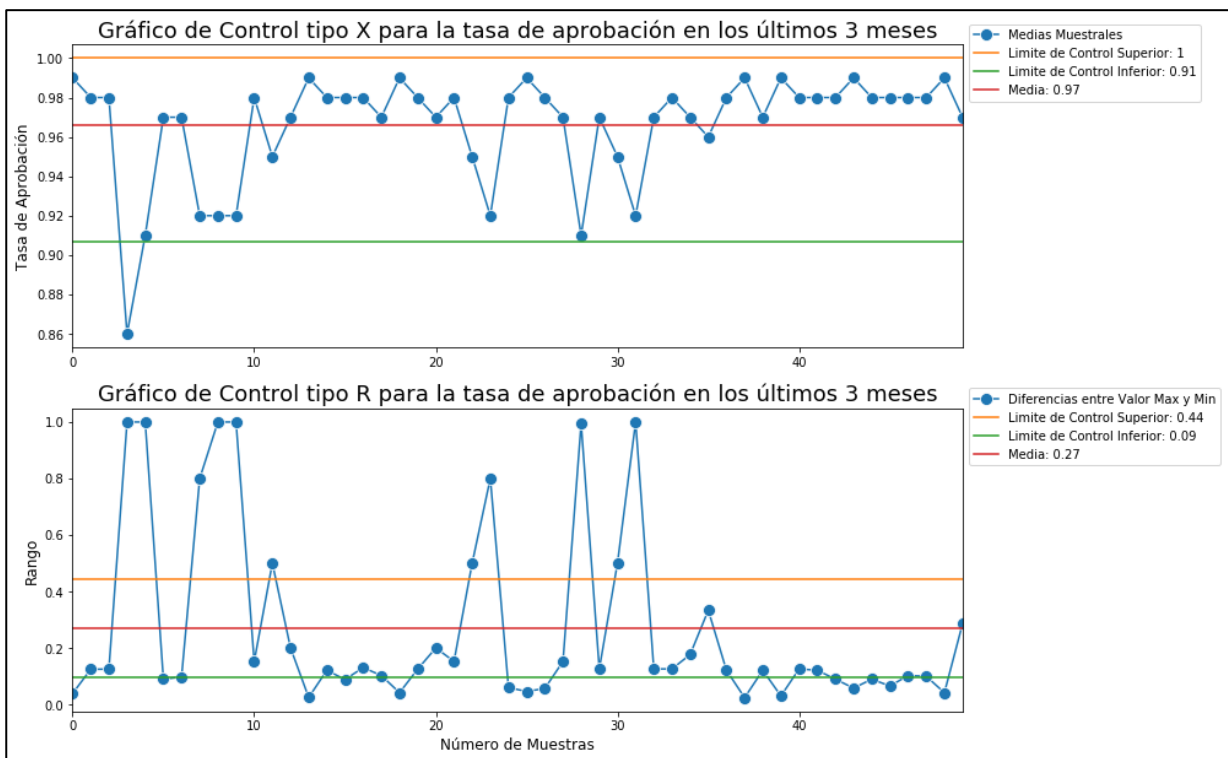


Ilustración 51: Gráficos del tipo X y R para la tasa de aprobación en los últimos 3 meses. Fuente: Elaboración Propia.

Estos últimos gráficos reafirman que en promedio el desempeño de los shoppers es sumamente elevado, ya que el límite de control inferior es de un 91% de aprobación. Sin embargo, también potencian lo dicho anteriormente sobre la existencia de shoppers que desequilibran el proceso, lo cual se observa en los puntos que se alejan de los límites de control en ambas gráficas. El desempeño en los últimos 3 meses tiende a ser inferior al *Score*, lo cual podría explicarse por la ventana de análisis escogida. Finalmente, no se realizaron gráficos de control para las variables "amount", "approved\_count" y "experiencia", ya que su análisis no aportaría hallazgos relevantes al estudio del proceso de revisión.

### 9.1.3. Regresión Logística

La regresión logística es un procedimiento cuantitativo sumamente útil para problemas en los que la variable dependiente toma valores de un conjunto finito. Este modelo permite predecir la variable categórica a partir de datos de otras variables conocidas, bien cuantitativas o cualitativas que actúan como variables explicativas.

La característica que se quiere explicar puede ser: a) una cualidad que puede únicamente tomar dos modalidades (modelos binomiales), b) una cualidad que

puede tomar más de dos modalidades diferentes, exhaustivas y mutuamente excluyentes (modelos multinomiales), c) una característica con varias modalidades que presentan un orden natural (modelos ordenados) y d) la característica a explicar corresponde a una decisión que puede suponer decisiones encadenadas (modelos anidados).

El caso a estudiar en este informe corresponde a un modelo binomial, donde la característica a estudiar es el rechazo o aprobación de una misión, lo cual fue modelado utilizando una variable *dummy* que toma el valor 1 si la misión fue aprobada y 0 si no ("approved"). Mediante la regresión logística se pretende modelar la probabilidad de que la misión este aprobada como función de ciertas variables que se presumen relevantes o influyentes. Por lo tanto, este algoritmo consiste principalmente en obtener una función logística de las variables independientes que permita clasificar las misiones en una de las dos subpoblaciones o grupos establecidos por la variable dependiente. La función logística es aquella que halla, para cada individuo según los valores de una serie de variables ( $X_i$ ), la probabilidad ( $P$ ) de que presente el efecto estudiado (esta puede apreciarse en la sección 5.2.3 de este informe) [31]. Dentro de los beneficios que otorga la utilización de este algoritmo, se destaca que permite determinar la influencia de cada variable explicativa en la dependiente y además otorga la funcionalidad de predecir la categoría de un individuo u objeto en base a sus características.

Sin embargo, a pesar de todos los beneficios potenciales que posee este algoritmo, existen algunos riesgos que deben sortearse previo a su utilización, ya que en la actualidad la generalización del uso de paquetes estadísticos aumenta el riesgo de que usuarios lo utilicen sin criterio y muchas veces sin comprender lo que hacen. La primera labor que debe realizarse antes de implementar el modelo consiste en eliminar el riesgo de multicolinealidad en las variables explicativas.

La multicolinealidad en una regresión corresponde a una relación de dependencia lineal fuerte (correlación alta o fuerte:  $\rho > |0,5|$ ) entre más de dos variables explicativas, es decir, una correlación alta entre éstas. Este fenómeno implica un aumento en la varianza de los estimadores y, por ende, los coeficientes no pueden estimarse con precisión. Dentro de las complicaciones que trae consigo este problema, se destacan los siguientes:

- Los intervalos de confianza para los coeficientes son muy amplios.
- Los errores estándares son muy altos.

- Los estimadores son poco robustos (pequeños cambios en la base de datos generan cambios importantes en los coeficientes estimados).

En otras palabras, no solucionar este problema podría implicar dar importancia predictiva a variables que la tienen o desestimar variables influyentes debido a su alta varianza [32]. La solución existente para la multicolinealidad consiste en estudiar la correlación de las variables explicativas y eliminar del modelo solo una de aquellas que tengan un alto grado de relación lineal, privilegiando a la que tenga mayor correlación con la variable dependiente. Para esto se realizó un correlograma con todas aquellas variables indicadas como influyentes en las secciones 8.4 y 9.1 de esta memoria:

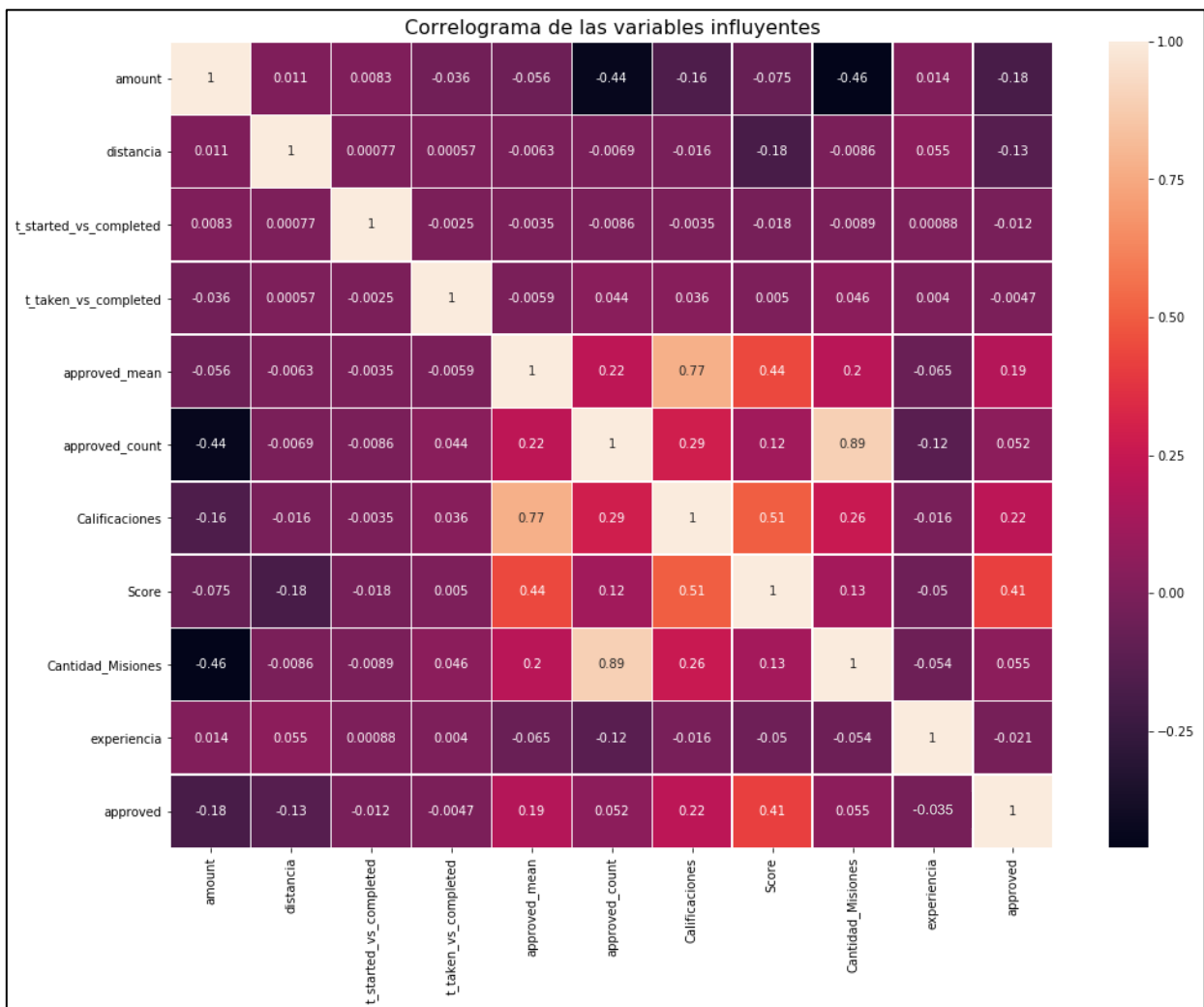


Ilustración 52: Correlograma de las variables más influyentes en el rechazo de misiones. Fuente: Elaboración Propia.

De este gráfico, se puede apreciar que existen múltiples variables que poseen una alta correlación entre ellas, destacándose las siguientes:

- i. La variable "Calificaciones" tiene una correlación de 0,77 con "approved\_mean".
- ii. La variable "Cantidad\_Misiones" tiene una correlación de 0,89 con "approved\_count".
- iii. La variable "Score" tiene una correlación de 0,51 con "Calificaciones".

Dado el método de selección de variables mencionado anteriormente, se privilegiará a aquellas variables con mayor correlación con "approved". Del punto i) la variable escogida correspondería a "Calificaciones", ya que tiene una correlación de 0,22 con "approved" versus 0,19 de "approved\_mean". Para el caso de ii) la variable escogida correspondería a "Cantidad\_Misiones", ya que tiene una correlación de 0,055 con "approved" versus 0,052 de "approved\_count". Finalmente, para el punto iii) la variable escogida corresponde a "Score", puesto que posee una correlación igual a 0,41 con "approved" versus 0,22 de "Calificaciones".

Luego de esta selección, se tiene que las variables escogidas para utilizar el algoritmo serían: "amount", "distancia", "t\_started\_vs\_completed", "t\_taken\_vs\_completed", "Cantidad\_Misiones", "Score" y "experiencia".

Teniendo contempladas las variables para implementar el modelo, se separa la base de datos en un set de entrenamiento y otro para testear el algoritmo. Esta práctica es ampliamente aceptada y utilizada en el aprendizaje de máquinas, puesto que la primera base de datos se utiliza para que el modelo generalice los datos, lo cual se realiza estimando los parámetros que mejor se ajustan a éstos. Para el caso de la regresión logística, la base de entrenamiento permitirá estimar los coeficientes que acompañan a las variables predictoras. El objetivo de la segunda partición de los datos es evaluar la capacidad predictiva del modelo, ya que se puede comparar la clasificación hecha por este versus la realidad observada [33].

Por ejemplo, las misiones expuestas en la base de datos ya están previamente clasificadas como aprobadas o rechazadas, por lo que con la base de *testing* se puede comprobar qué tan bien clasifica el algoritmo, ya que previamente se conoce la categoría real a la que pertenece la misión.

Esta metodología puede entenderse de forma gráfica mediante la siguiente ilustración:

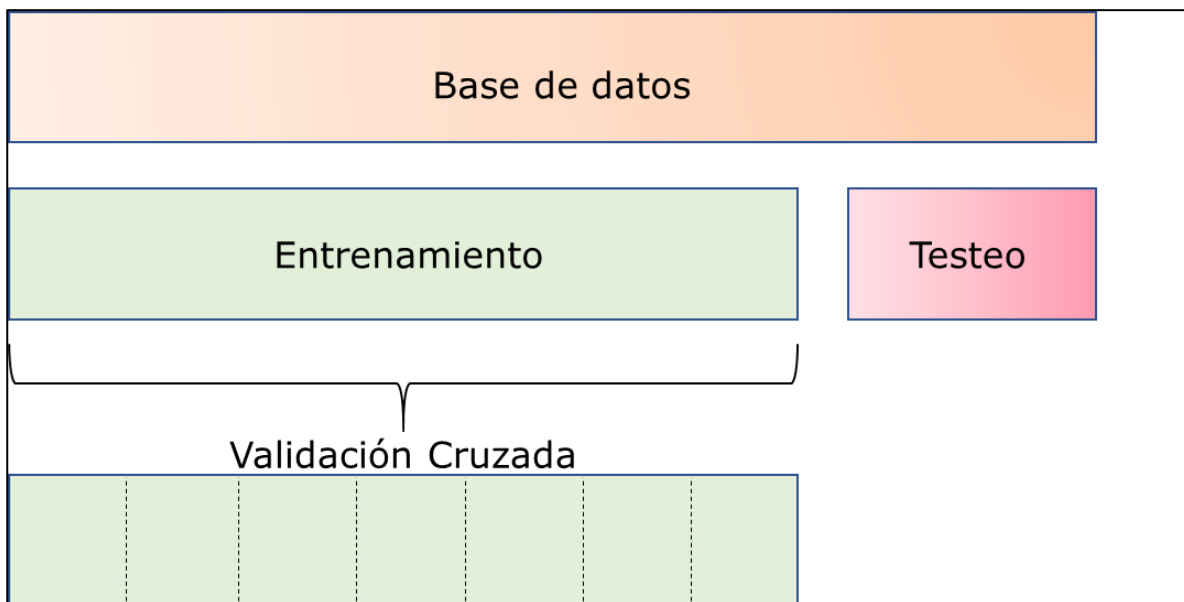


Ilustración 53: Particiones realizadas a la base de datos para la regresión logística.  
Fuente: Elaboración Propia.

Las principales controversias existentes a la hora de particionar la base de datos se refieren a qué porcentaje del total de datos debería tener cada división. Actualmente no existe un consenso sobre esta temática, pero el rango utilizado generalmente por expertos para la base de entrenamiento va entre un 70% a 80% del total de datos y para la de *testing* de entre 20% a 30% [34].

Para el problema a estudiar, se considerará que la base de entrenamiento tendrá el 70% de los datos y la de *testing* el 30% restante. Otra complicación que surge tras la división de la base de datos corresponde a que cada participación generada sea representativa de la población total. Por ejemplo, puede que para la base de *testing* no aparezca el evento a estudiar en la variable dependiente, es decir, puede que no haya ninguna misión rechazada, puesto que este evento ocurre tan solo en el 9,6% del total de misiones aproximadamente. Para solucionar esta problemática se utilizó la librería especializada en *machine learning* de Python llamada "Scikit-learn", la cual incluye el módulo "model\_selection" que a su vez posee la funcionalidad "train\_test\_split". La función "train\_test\_split", recibe como input la base que contiene solo las variables explicativas (x), la variable explicativa (y), el tamaño de la base de testeo (test\_size=0,3) y el parámetro "stratify" [35].

Este último parámetro proviene del término "Stratification", el cual es una forma de representación estadística que muestra cómo se comporta una característica o variable de una población a través de hacer evidente el cambio de dicha variable en subpoblaciones o estratos en los que se ha dividido.



Consiste en la división previa de la población de estudio en grupos o clases que se suponen homogéneos respecto a la característica a estudiar y que no se solapen [36]. Para el caso de las misiones, el parámetro que se le entrega corresponde a la variable dependiente (y), ya que la función utiliza el porcentaje que representa cada clase y genera la partición considerando que esta proporción de categorías debe cumplirse en cada subgrupo generado. Dado esto, tanto la base de entrenamiento como la de testeo, deberán tener un 90,4% de misiones aprobadas y un 9,6% de misiones rechazadas.

La cantidad de misiones que posee la base de entrenamiento corresponde a 375.510 (70% del total) y la de testeo 160.934 (30%). La regresión logística se programó utilizando Python, con la librería "Scikit-learn" y su módulo "LogisticRegresion". El modelo se aplicó a las variables explicativas presentes en la base de entrenamiento, utilizando la variable "approved" como dependiente. Los resultados del algoritmo pueden apreciarse en la ilustración presentada a continuación:

Resultados	
N° Observaciones	375.510
Variable dependiente	approved
Pseudo R cuadrado	0,844

Variables	Coefficientes	Error estándar	Estimador	P-value	[0.025	0.975]
amount	-0,0002	0,0000	10,5691	0,0000	0,0002	0,0002
distancia	-0,0026	0,0000	128,0441	0,0000	0,0026	0,0026
t_started_vs_completed	-0,0002	0,0000	3,8828	0,0001	0,0001	0,0003
t_taken_vs_completed	-0,0000	0,0000	3,0607	0,0022	0,0000	0,0000
Cantidad_Misiones	0,0001	0,0000	-16,7560	0,0000	0,0001	0,0001
Score	0,0486	0,0004	-132,6377	0,0000	0,0493	0,0478
experiencia	-0,0000	0,0000	2,9117	0,0036	0,0000	0,0000

Ilustración 54: Resultados de la regresión logística en Python.  
Fuente: Elaboración Propia.

Para el caso de la regresión logística, los coeficientes de las variables explicativas permiten determinar el impacto en la probabilidad de que ocurra el suceso de estudio (aprobación o rechazo). Esto se debe a que la variable dependiente, se transforma mediante la función logística a una probabilidad, con la finalidad de que el output del modelo siempre éste entre el rango 0 y 1. Lo anterior queda definido por medio de las siguientes ecuaciones:

$$P(Y|X) = \text{Probabilidad de que ocurra el evento } Y \text{ sujeto a } X.$$

$$\text{Se define } P(Y|X) = \frac{1}{1 + e^{-\sum_{i=1}^n x_i * \beta_i}}$$

Donde  $x_i$  corresponde a las variables explicativas y  $\beta_i$  a los coeficientes que las acompañan. La probabilidad de que ocurra el evento se modela bajo la función logística, sin embargo, su interpretación es poco intuitiva bajo la ecuación anterior, por lo que se utilizará la siguiente transformación:

$$\ln\left(\frac{P(Y|X)}{1 - P(Y|X)}\right) = \sum_{i=1}^n x_i * \beta_i$$

**donde para el caso binomial:**  $P(Y = 1|X) = 1 - P(Y = 0|X)$

$$\Rightarrow \ln\left(\frac{P(Y = 1|X)}{P(Y = 0|X)}\right) = \sum_{i=1}^n x_i * \beta_i \Rightarrow \frac{P(Y = 1|X)}{P(Y = 0|X)} = \sum_{i=1}^n e^{x_i * \beta_i}$$

El primer término a la izquierda es llamado *logit*, esto es, el logaritmo natural del cociente entre la probabilidad de pertenecer a cada una de las categorías [37]. Esta ecuación permite deducir que, si un coeficiente de una variable explicativa es mayor a 0, influye positivamente en la probabilidad de que ocurra el evento  $P(Y = 1|X)$ , ya que aumenta la razón de este término versus su complemento. La situación es inversa en el caso de que el coeficiente sea negativo. Si la variable posee un coeficiente igual a 0, esto indica tiene importancia a la hora de describir la probabilidad, puesto que  $e^{0 * x_i} = 1$ . Dada esta interpretación de los coeficientes, se tiene la siguiente descripción de los resultados obtenidos:

- i. Las variables que influyen positivamente en la probabilidad de aprobación corresponden a "Score" y "Cantidad\_Misiones", lo cual responde a la intuición de los digitadores y es un resultado esperable, ya que, si aumenta la confiabilidad y la cantidad de misiones que realiza un shopper, es más probable que las misiones que realice estén aprobadas.
- ii. Las variables que afectan negativamente corresponden a: "distancia", "t\_started\_vs\_completed" y "amount". Que el efecto de las primeras dos variables sea negativo responde a la esperado, ya que si aumenta la distancia entre el punto de inicio y el final de una misión es más probable que ésta sea rechazada y pasa exactamente igual con el tiempo de duración. Sin embargo, que el coeficiente asociado al pago de la misión sea negativo genera controversia. Podría esperarse que los shoppers tuvieran incentivos a realizar de mejor manera misiones con un mayor

pago. A pesar de esto, una posible explicación a este fenómeno podría ser que exista una alta correlación entre misiones de mayor pago y sectores fuera de cobertura o peligrosos, lo cual dificultaría la labor del shopper.

- iii. Las variables que no afectan la probabilidad de aprobación corresponden a: "experiencia" y "t\_taken\_vs\_completed". Esto puede deberse a que el efecto de estas variables es explicado por otras. Por ejemplo, el efecto de la variable experiencia podría estar viéndose comprometido por "Cantidad\_Misiones" o "Score" y para "t\_taken\_vs\_completed" podría verse afectada por "t\_started\_vs\_completed". La otra explicación posible, es que en realidad estas variables no sean influyentes en la probabilidad de aprobación de una misión.

#### 9.1.4. Validación del modelo

El modelo puede validarse al analizar su capacidad predictiva utilizando la base de *testing*, explicada en la sección 9.2.1. Para esto, se debe aplicar la regresión logística (previamente ajustada a la base de entrenamiento) a las variables explicativas de la base de *testing*. Esta implementación generará como output, la categoría a la que pertenece cada misión de la base utilizada, que puede ser aprobada (output igual a 1) o rechazada (output igual a 0). Puesto que la base de *testing* posee la clase a la que pertenece cada misión (variable "approved"), se puede obtener la bondad de ajuste del modelo comparando la categoría predicha por éste versus la realidad observada.

Esta comparación puede realizarse utilizando la herramienta llamada matriz de confusión, la cual puede apreciarse a continuación:

		Predicción	
		Positivos	Negativos
Observación	Positivos	Verdaderos Positivos (VP)	Falsos Negativos (FN)
	Negativos	Falsos Positivos (FP)	Verdaderos Negativos (VN)

Ilustración 55: Ejemplificación de una matriz de confusión.  
Fuente: Elaboración Propia.

La matriz de confusión permite visualizar el desempeño de un algoritmo que se emplea en el aprendizaje supervisado. Cada columna representa el número de predicciones hechas para cada clase, mientras que cada fila representa a las instancias reales de cada clase [38]. Existen múltiples métricas de bondad de ajuste que se derivan de esta herramienta, sin embargo, para el caso a estudiar se debe privilegiar aquellas que penalicen por la cantidad de falsos positivos (FP). Esto se debe a que el evento negativo en cuestión corresponde al rechazo de una misión, el cual tiene una ocurrencia promedio del 9,6% en el total de casos, y, por ende, un modelo que prediga que el 100% de las misiones están aprobadas, tendría una efectividad del 90,4%. Claramente ese modelo no cumpliría con los estándares buscados, puesto que es necesario privilegiar predecir con exactitud aquellos eventos que implican un mayor costo a la empresa.

Para el caso de Rocketpin, las misiones rechazadas deben ser reprocesadas y realizadas por otro shopper si es que son identificadas a tiempo. Si las misiones rechazadas no son detectadas, Rocketpin se enfrentaría a multas por incumplimientos de contrato que superan considerablemente el monto pagado por la misión, ya que disminuyen el estándar de calidad pactado y su retrabajo probablemente excedería el plazo de revisión solicitado. Para el caso en el que equivocadamente se clasifica una misión como rechazada, esta debe ser realizada por otro shopper dentro del plazo, por lo que su costo equivaldría a pagar dos veces la remuneración por la misión. Esto sugiere que el costo de rechazar una misión que debería estar aprobada (falso negativo) es menor al costo de aprobar una misión que debería estar rechazada (falso positivo). Es por esto, que las métricas escogidas para validar el modelo en base a la matriz de correlación son las siguientes:

$$\text{True Negative Ratio} = \frac{VN}{VN + FP}$$

$$\text{Accuracy} = \frac{VP + VN}{VP + VN + FP + FN}$$

$$\text{Matthews Correlation Coeficients} = \frac{VP * VN - FP * FN}{\sqrt{(VP + FP)(VP + FN)(VN + FP)(VN + FN)}}$$

$$\text{AUC} = \frac{TPR + TNR}{2} = \frac{FP * VP + VN * FN + 2 * VN * VP}{2(VN + FP)(VP + FN)}$$

El *True Negative Ratio* o *Specificity* permite determinar la capacidad que posee el algoritmo para predecir correctamente los valores negativos (misiones rechazadas en este caso), ya que se calcula como el cociente entre los valores

negativos en los que acierta el modelo versus la totalidad de casos negativos (considerando que los valores falsos positivos en realidad son negativos) [39].

La métrica *accuracy* es útil para evaluar la capacidad del modelo para predecir correctamente las diferentes categorías existentes, sin embargo, debe tenerse cuidado con conjuntos de datos donde hay una disparidad significativa entre el número de etiquetas positivas y negativas. Al igual que el ejemplo dado anteriormente, un modelo que predijera que todas las misiones están aprobadas tendría un *accuracy* de 90,4%, lo cual es bastante alto, sin embargo, tendría un *True Negative Ratio* de 0%, ya que no predice correctamente ningún caso negativo. Dado esto, es importante que esta métrica se utilice en conjunto con otras para evaluar de forma integral la capacidad predictiva de un algoritmo [40].

El *Matthews Correlation Coeficients* es una medida de la calidad de un modelo de clasificación binario o multiclase. Es una métrica bastante utilizada para bases de datos desbalanceadas, ya que tiene en cuenta las relaciones de equilibrio de las 4 categorías de la matriz de confusión (verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos). El MCC toma valores entre -1 y 1. El coeficiente igual 1 indica una predicción perfecta, el 0 representa una predicción igual a una aleatoria y el -1 corresponde a un desacuerdo total entre predicción y observación [41].

El AUC (*Area Under the Curve*) corresponde al área bajo la curva ROC. La curva Roc se construye al variar la probabilidad de corte para la clasificación y graficar el *True Positive Ratio* versus el *False Positive Ratio* (los cuales se construyen de manera similar al *True Negative Ratio*). Esta métrica permite evaluar la capacidad que posee un algoritmo para distinguir entre las diferentes clases y es ampliamente aceptada en distintos campos científicos, ya que es independiente del punto de corte de la probabilidad. El AUC toma valores entre 0 y 1, donde un AUC cercano a 1 indica que el modelo es capaz de distinguir correctamente entre las categorías [42].

Una vez definidas las métricas a utilizar, se procedió a crear la matriz de confusión para el modelo de regresión logística utilizando la base de *testing* y a estimar cada una de las medidas de bondad de ajuste señaladas anteriormente:

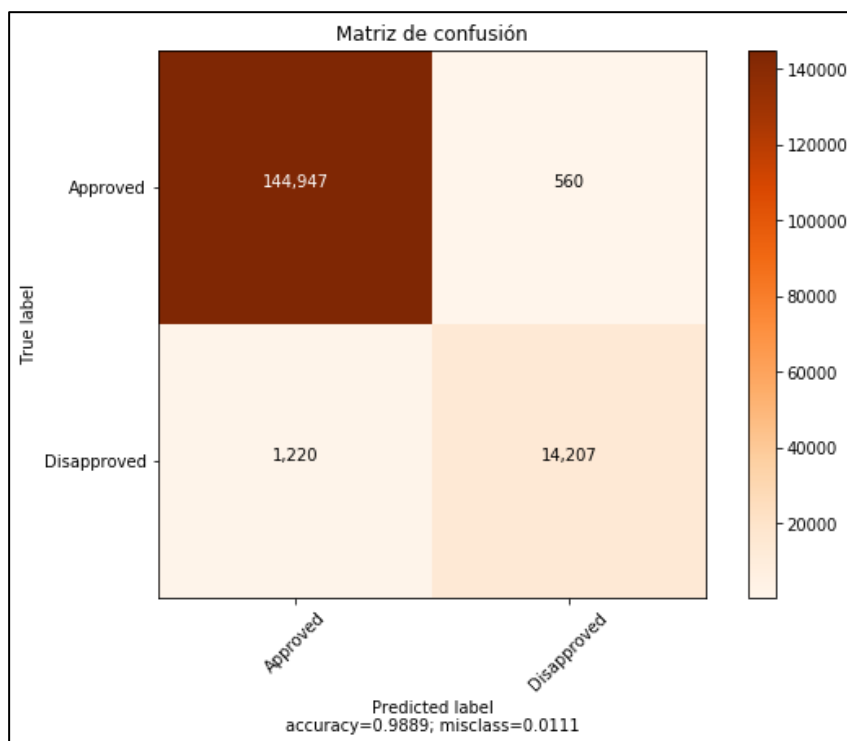


Ilustración 56: Matriz de confusión realizada para la predicción de la regresión logística.  
Fuente: Elaboración Propia.

Métricas	Valor
True Negative Rate	0,921
Accuracy	0,989
AUC	0,959
Matthews Correlation Coefficient	0,935

Ilustración 57: Métricas resultantes de la evaluación de la regresión  
Fuente: Elaboración Propia.

De la matriz de confusión puede apreciarse que el modelo distingue correctamente 144.947 de las 145.507 misiones realmente aprobadas (99,6%) y 14.207 de las 15.427 misiones realmente rechazadas (92,1%). A simple vista puede apreciarse que el rendimiento del modelo es destacable, puesto que distingue con bastante exactitud entre las misiones aprobadas y rechazadas. Esto puede comprobarse al revisar los valores resultantes de las métricas señaladas anteriormente, donde todas son superiores a un 90%.

Finalmente, la última forma a utilizar para validar la capacidad predictiva del modelo es mediante la técnica llamada *cross validation*, la cual consiste en dividir la base de entrenamiento en  $k$  secciones diferentes, utilizando  $k - 1$  divisiones de la base original como base de entrenamiento y la división restante como base de *testing*. Posteriormente, se entrena el modelo

utilizando la nueva base de entrenamiento y se valida su capacidad predictiva con la nueva base de *testing*. Este proceso se repite tantas veces como la cantidad de particiones realizada al conjunto de datos (*k* veces) y para cada iteración se estiman las métricas de ajuste, con el fin de calcular el promedio y error de los *k* valores obtenidos para cada métrica. La utilidad de este método consiste en validar que los resultados del modelo no se deben a la elección de los conjuntos de entrenamiento y *testing* iniciales, y que además el modelo es aplicable a la totalidad de la base de datos [43]. Un ejemplo gráfico de este proceso puede apreciarse en la ilustración 78 del anexo de este informe.

Los resultados obtenidos de la *cross validation* utilizando 8 divisiones de la base de entrenamiento original pueden verse a continuación:

Iteración	AUC	Accuracy	MCC	TNR
1	0,959035	0,98913	0,935	0,921892
2	0,958679	0,98889	0,935	0,921157
3	0,958575	0,98917	0,935	0,920902
4	0,955731	0,98871	0,935	0,914973
5	0,955295	0,98814	0,935	0,914743
6	0,959192	0,9888	0,935	0,922495
7	0,956152	0,98804	0,935	0,916628
8	0,958805	0,98896	0,935	0,921376
<b>Promedio</b>	0,957683	0,988731	0,935	0,919271
<b>Error</b>	0,16%	0,04%	0,00%	0,32%

Ilustración 58: *Cross Validation* de 8 secciones para la base de entrenamiento original.  
Fuente: Elaboración Propia.

De esta ilustración puede apreciarse que la dispersión entre las métricas para cada iteración es mínima, ya que el error estándar no supera un 0,5%. En cuanto a los promedios, es posible notar que son casi idénticos a los valores obtenidos para las métricas de la ilustración 57. Esto indica que el modelo de regresión logística utilizado para clasificar misiones de verificación domiciliaria es válido y posee una capacidad predictiva notable.

### 9.1.5. Interpretación de los resultados

Los resultados de la regresión logística aplicada al problema en el proceso de revisión de misiones, permitió descubrir que las variables más influyentes en la probabilidad de aprobación corresponden a: la distancia entre el punto de inicio y final de la misión, duración y remuneración de ésta, tasa de confiabilidad y cantidad de misiones realizadas por el shopper en la ventana de análisis. Sin embargo, a pesar de los coeficientes asociados a cada variable, no es posible indicar que alguna posea más importancia que otra, ya que las

magnitudes pueden no ser comparables. Por ejemplo, la variable "Score" tiene asociado un coeficiente igual a 0,0486 y "distancia" un coeficiente de -0,0026. Podría pensarse que la confiabilidad del shopper es más importante, pero esto no es correcto, puesto que ambas pueden tomar distintos valores. El "Score" puede variar entre 0 y 100%, en cambio la "distancia" puede tomar cualquier valor positivo en metros, lo cual podría indicar que el efecto de ambas variables puede contrarrestarse a pesar de que una tenga mayor coeficiente que la otra.

Esto puede apreciarse al considerar la ecuación proveniente de la regresión logística para determinar la probabilidad de aprobación y los coeficientes obtenidos para cada variable, lo cual es presentado a continuación:

$$P(Y = 1 | \beta_{pago}, \beta_{distancia}, \beta_{duración}, \beta_{cantidad\ misiones}, \beta_{Score}) = \frac{1}{1 + e^{-\sum_{i=1}^5 x_i * \beta_i}}$$

$$x_{distancia} = 230,85 [m]$$

$$x_{Score} = 95\%$$

$$x_{distancia} * \beta_{distancia} + x_{Score} * \beta_{Score} = 230,85 * 0,0486 - 0,95 * 0,0002 = 0$$

Lo anterior indica que el efecto en la probabilidad de aprobación de una misión realizada por un shopper con una confiabilidad del 95% puede verse contrarrestada por una distancia entre PIN inicial y final de 230 metros. A modo de ejemplificación, se estimará la probabilidad de aprobación para una misión que contempla los valores de las medias muestrales de cada variable influyente, los cuales fueron obtenidos en la sección 9.1.2:

$$x_{pago} = \$1.101,79$$

$$x_{distancia} = 96,71 [m]$$

$$x_{duración} = 3,68 [min]$$

$$x_{cantidad\ misiones} = 413$$

$$x_{Score} = 98,47\%$$

$$P(Y = 1 | \beta_{pago}, \beta_{distancia}, \beta_{duración}, \beta_{cantidad\ misiones}, \beta_{Score}) = 59,46\%$$

Esto sugiere que una misión "promedio" posee una probabilidad de ser aprobada de aproximadamente un 60%, lo cual es bastante acertado con la realidad. De hecho, este resultado permite validar la efectividad del algoritmo,



puesto que podría pensarse que un shopper con una confiabilidad alta (98,47%) necesariamente realizará misiones aprobables, sin embargo, el modelo indica que hasta shoppers con una confiabilidad del 100% pueden tener una alta probabilidad de rechazo, dependiendo de la distancia, duración y remuneración de la misión, lo cual demuestra la robustez de los resultados.

Finalmente, se pueden extraer múltiples soluciones aplicables al problema encontrado en el proceso de revisión basado en los resultados expuestos, los cuales pueden basarse principalmente en la probabilidad de aprobación entregada por la regresión logística y/o en los límites de control encontrados para las variables más influyentes.

## **10. REDISEÑO DEL PROCESO DE REVISIÓN**

### **10.1. Elección de la potencial solución**

De la investigación y hallazgos realizados en las secciones 7, 8 y 9 de esta memoria, surgen 4 potenciales soluciones aplicables al proceso de revisión actual, las cuales se definen a continuación:

- A. Automatizar el proceso utilizando el modelo de regresión logística, ya que como se apreció anteriormente, el algoritmo puede clasificar las misiones usando las bases de datos recopiladas por Rocketpin.
- B. Crear un sistema de control para el proceso usando el modelo de regresión logística, el cual contemple distintas exhaustividades de revisión dependiendo de la probabilidad de aprobación asociada a la misión.
- C. Crear un sistema de control dependiendo de los niveles de confiabilidad de los shoppers.
- D. Crear un sistema de control que contemple diferentes grados de revisión dependiendo de los límites de control que poseen las variables más influyentes en la aprobación de misiones.

Estas 4 soluciones tienen como objetivos: asegurar la calidad del servicio de Rocketpin de forma más eficiente y menos costosa, reducir el tiempo de revisión para los digitadores y facilitarles su labor diaria.

Para elegir la solución que se ajuste mejor a los requerimientos y necesidades de Rocketpin, se utilizará el enfoque propuesto por Juan Bravo, el cual consiste en generar una lista preliminar ponderada, de acuerdo con factores de decisión

expresados en forma de valoración cualitativa, normalmente con notas de 1 a 5.

Los factores de decisión provienen de aspectos que son relevantes desde la estrategia de la organización y sirven para evaluar proyectos. También surgen desde restricciones, ya que se trata de aspectos que acotan el universo de posibilidades como, por ejemplo, el costo de la solución no puede superar el costo del problema o un determinado monto, ya sea por restricciones presupuestarios o porque aparece un costo de oportunidad, es decir, pueden existir otros problemas de costo alto con soluciones de bajo costo para la organización que darían más rentabilidad interna con el mismo nivel de inversión. También existirán restricciones, técnicas, de estandarización, legales y de otra índole.

Existe un grupo de interés, llamado destinatarios de la misión (se les dice clientes para simplificar), a cuyos intereses no se les llama restricciones, sino que se habla de requerimientos. Es necesario preguntar, ¿Qué desean los clientes? Las respuestas pueden ser a nivel de procesos individuales o grupos de procesos, por ejemplo, en el otorgamiento de créditos bancarios, el cliente requiere rapidez y baja tasa. Estos aspectos también pueden ser considerados como factores de decisión [7, p. 68].

Dado lo anterior, se evaluarán las soluciones propuestas en base a los siguientes factores de decisión, los cuales surgen de las necesidades de la organización, clientes y personal del área de control de calidad:

- i. Nivel de inversión inicial.
- ii. Beneficios a largo plazo.
- iii. Tiempo de implementación.
- iv. Calidad de cara al cliente
- v. Alineamiento con la estrategia de la organización
- vi. Dificultad de diseño e implementación.
- vii. Riesgo de adversidad al cambio.

La calificación y ponderación otorgada a cada factor debe realizarse por la directiva de la organización, un panel de expertos y el personal del área a la que el proyecto afectará directamente. Para el caso de Rocketpin, se consideró al personal del área de tecnología como el panel de expertos, puesto que tiene un alto grado de conocimiento en proyectos que involucran inteligencia de negocios, análisis de datos, inteligencia artificial y aprendizaje de máquinas. Además, tienen manejos en diferentes lenguajes de programación como: Python, R, C++ y Java, entre otros.

En la siguiente ilustración puede apreciarse la calificación y ponderación otorgada para cada factor considerando las soluciones propuestas:

Factores	Ponderación	Automatizar revisión	Sistema de control - probabilidad de aprobación	Sistema de control - shoppers	Sistema de control - Límites de control
Nivel de inversión	10%	2	4	4	5
Beneficios largo plazo	25%	5	4	3	2
Tiempo de implementación	10%	1	3	3	5
Calidad de cara al cliente	15%	5	5	3	3
Alineamiento estrategia	15%	5	5	4	4
Dificultad diseño e implementación	15%	1	3	3	5
Riesgos asociados	10%	2	5	3	3
Calificación Ponderada	100%	3,4	4,15	3,25	3,6

Ilustración 59: Calificación ponderada de las soluciones usando el enfoque de factores.  
Fuente: Elaboración Propia.

Se procederá a explicar el porqué de las calificaciones otorgadas para cada factor:

- El nivel de inversión a realizar para automatizar el proceso de revisión es bastante superior en comparación a la implementación de un sistema de control ya que, según el área de tecnología, se deberá hacer un gasto intensivo en contratar personal calificado e invertir en licencias y softwares.
- Las soluciones basadas en sistemas de control contemplan mantener al personal del área de revisión y aumentar el flujo de misiones por medio de la diferenciación en los niveles de control. En cambio, automatizar el proceso permitirá aumentar sustancialmente la cantidad de misiones revisadas y no se necesitaría tanto personal para su supervisión.
- El tiempo de diseño, programación e implementación es mayor para la automatización, en comparación a las demás soluciones. De hecho, el sistema de control basado en los límites de control puede implementarse fácilmente aplicándose restricciones a las variables más influyentes encontradas anteriormente.
- Tanto el proyecto de automatización como el sistema de control basado en probabilidad de aprobación permitirán aumentar el flujo de misiones a revisar, asegurando la calidad del servicio, ya que como se apreció anteriormente, el modelo de regresión logística tiene métricas de predicción excepcionales. En cambio, un sistema de control basado en la confiabilidad de los shoppers podría generar mayor cantidad de

errores, ya que Rocketpin asigna un 100% de confiabilidad a los agentes nuevos y, además, no gestionan información sustancial que permitiría clasificar a los shoppers de mejor manera como, por ejemplo: nivel de escolaridad, uso de otras aplicaciones basadas en la economía colaborativa, situación laboral, entre otras. Para el caso del sistema de control basado en los límites de control, puede suceder que, dada la aleatoriedad de las muestras escogidas para su estimación, se generen límites poco realistas (como, por ejemplo, misiones con una duración de 0 segundos dentro del rango aceptable).

- Se puede apreciar que tanto la automatización del proceso y el sistema de control basado en probabilidad tienen la máxima calificación, porque permiten aumentar el flujo de misiones revisadas, lo cual es sustancial para el crecimiento de Rocketpin. Los sistemas de control basado en confiabilidad y límites de control también permiten lo anterior, pero tienen deficiencias en el aseguramiento de la calidad del servicio.
- Los sistemas de control basados en probabilidad y confiabilidad tienen una dificultad media en su diseño e implementación, ya que pueden realizarse utilizando una combinación de Python y SQL. La automatización del proceso es intensiva en el uso de nuevas tecnologías (o librerías especializadas de Python). Para el caso del sistema de control basado en límites de control, su implementación puede realizarse fácilmente, ya que solo contemplaría una serie de restricciones lineales que pueden implementarse en Python.
- Para la dirección de Rocketpin es importante que exista una intervención humana en el proceso de revisión, ya que existe un cierto temor a los resultados de un software automático. Además, este proyecto generaría reticencia al cambio, ya que el personal del área de control de calidad no estaría dispuesto a implementarlo, no así con los sistemas de control.

En base al análisis realizado, se escoge la solución que presenta el mayor promedio ponderado de las calificaciones en cada factor, siendo elegida la creación de un sistema de control basado en la regresión logística, con un promedio de 4,15.

## **10.2. Diseño de la solución escogida**

En la sección 9.2.2. se pudo notar que el modelo de regresión logística es bastante bueno distinguiendo entre las misiones aprobadas y rechazadas. Esto es posible, ya que el modelo estima la probabilidad de una misión de pertenecer a cada una de las categorías.

La probabilidad de aprobación del modelo puede modelarse mediante un gráfico de control del tipo X, ya que es una variable continua, lo cual permitirá entender el comportamiento de ésta:

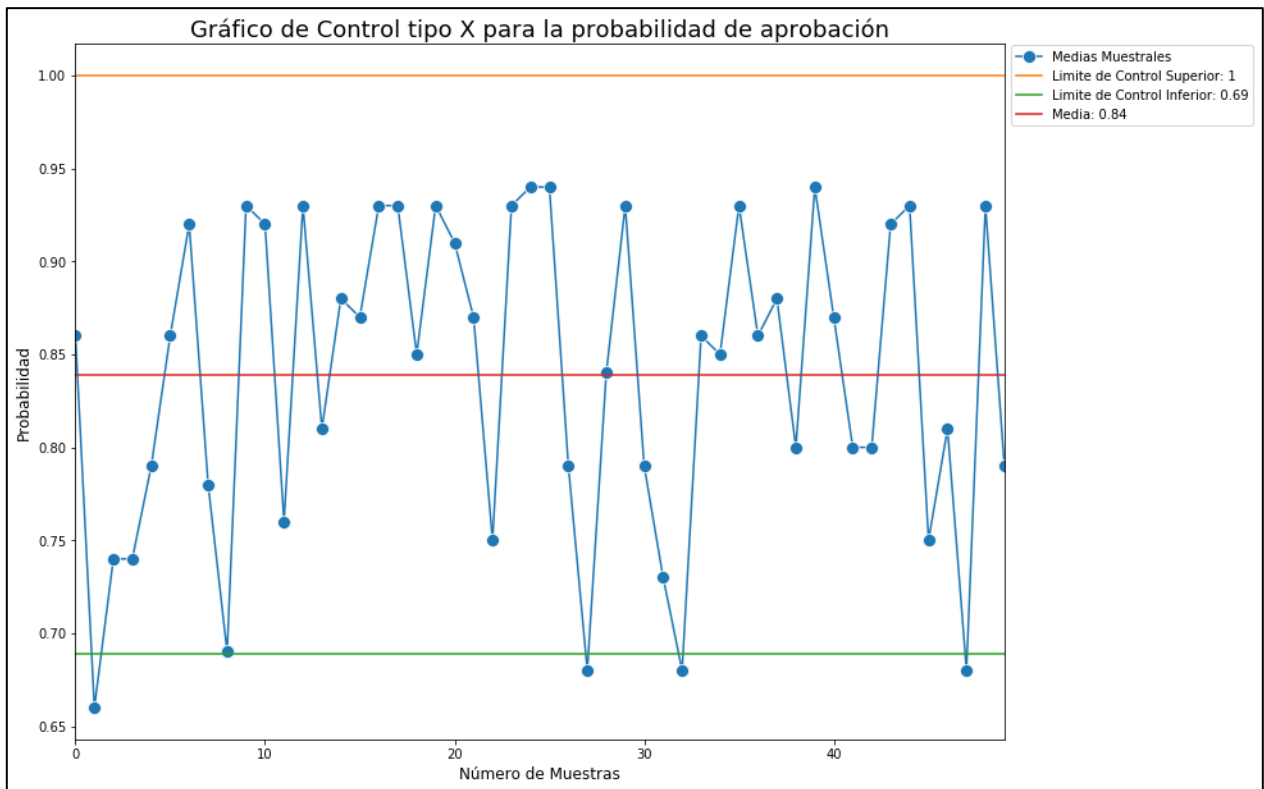


Ilustración 60: Gráfico del tipo X para la probabilidad de aprobación.  
Fuente: Elaboración Propia.

De este, puede apreciarse que la media de la probabilidad de aprobación es bastante alta, cercana a un 84%, lo cual es esperable, ya que la cantidad de misiones aprobadas supera con creces al total de desaprobadas. Además, del gráfico se extrae que existen 3 diferentes cuadrantes para la probabilidad de aprobación:

- I. El primero corresponde a aquellas misiones cuya probabilidad se encuentra entre el límite superior y la media.
- II. El segundo corresponde a aquellas misiones cuya probabilidad se encuentra entre la media y el límite inferior.
- III. El tercer y último cuadrante, corresponde a aquellas misiones cuya probabilidad es inferior al límite inferior.

La solución propuesta propone utilizar la información extraída de este gráfico para generar el sistema de control. Las misiones ubicadas en el primer cuadrante no deberán ser revisada por los digitadores, debido a su alta probabilidad de aprobación. Para el segundo cuadrante se realizará una revisión aleatoria, priorizando las misiones realizadas por shoppers con menor confiabilidad y experiencia. Finalmente, se propone una revisión completa de las misiones ubicadas en el último cuadrante, debido a su baja probabilidad de aprobación.

Basándose en la predicción del modelo de regresión logística para la base de entrenamiento (160.934 observaciones), se estima que la cantidad de misiones perteneciente a cada cuadrante sería la siguiente:

Segmento	Cantidad Misiones	Porcentaje
<b>Cuadrante I</b>	141.624	88,0%
<b>Cuadrante II</b>	3.871	2,4%
<b>Cuadrante III</b>	15.439	9,6%
<b>Total</b>	160.934	100%

Ilustración 61: Distribución de las misiones según cuadrante.

Fuente: Elaboración Propia.

Se plantea para el cuadrante II, revisar de forma aleatoria el 50% del total de misiones, lo cual implicaría que, con esta solución, los digitadores pasarían de revisar el 100% de las misiones, a tan solo un 10,8% del total.

La solución deberá ser implementada por el área de tecnología, utilizando como base el código generado para la programación de la regresión logística. El sistema de control diseñado deberá ser capaz de procesar la información de la base de misiones y shoppers en conjunto, generando las variables necesarias para la implementación del algoritmo. Posteriormente, deberá estimar las probabilidades de aprobación de cada misión, aprobar las misiones del cuadrante I, filtrar de forma aleatoria las misiones a revisar del cuadrante II y aprobar las restantes y finalmente, generar un conglomerado de misiones del cuadrante II y III para que sean revisadas por los digitadores.

Junto con esto, es importante que el sistema de control tenga un sensor que permita retroalimentar el proceso, lo cual puede hacerse al considerar el porcentaje del total de misiones de cada cuadrante como variable de referencia. Si existe un desbalance importante entre los porcentajes de referencia y salida que tiene cada cuadrante, el sensor debe ser capaz de retroalimentar al controlador para que este corrija las deficiencias o busque alguna explicación del cambio generado. Dado esto, el sistema de control a

implementar correspondería a uno de lazo cerrado. Estos sistemas son aquellos en los que la acción de control está en función de la señal de salida; es decir, en los sistemas de control de lazo cerrado o sistemas de control con realimentación, la salida que se desea controlar se realimenta para compararla con la entrada (valor deseado) y así generar un error que recibe el controlador para decidir la acción a tomar sobre el proceso, con el fin de disminuir dicho error y por tanto, llevar la salida del sistema al valor deseado [44]. Esto puede entenderse de mejor forma en la siguiente ilustración:

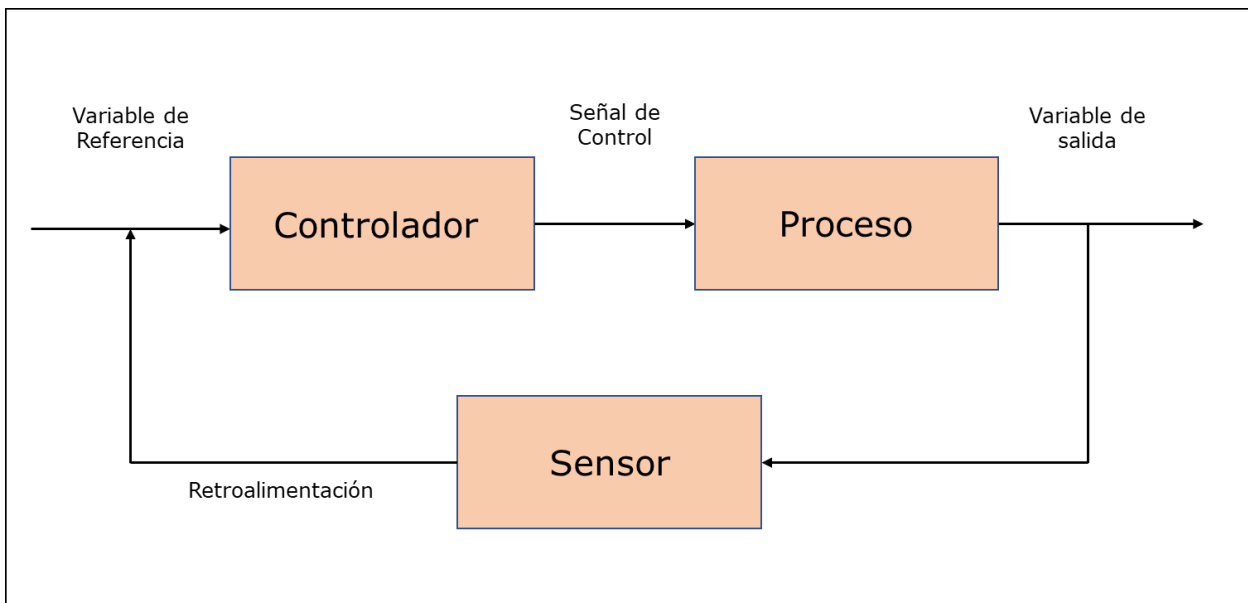


Ilustración 62: Ejemplo de un sistema de control de lazo cerrado.  
Fuente: Introducción a los sistemas de control (M. PÉREZ, 2017).

La variable de entrada para el sistema de control propuesto correspondería a los porcentajes del total de misiones relativos a cada cuadrante. El controlador deberá ser un ingeniero civil informático y/o carrera afín que se encargue de supervisar el correcto funcionamiento del algoritmo y que sea capaz de actualizarlo en la medida de que cambien las condiciones externas del problema como, por ejemplo: cambios en las políticas de verificación, problemas en la capacitación de shoppers, entre otros. El proceso corresponderá a la estimación de la probabilidad de aprobación para cada misión. La variable de salida será el total de misiones a revisar por el área de control de calidad (que debería coincidir con el 10,8% del total de misiones). El sensor deberá ser algún dispositivo o programa que permita medir en tiempo real la variable de salida y entregue de forma constante el error en la medición al controlador.

Finalmente, en base a la solución propuesta, el rediseño del proceso de revisión de misiones puede visualizarse mediante la metodología BPMN a través de la siguiente ilustración:

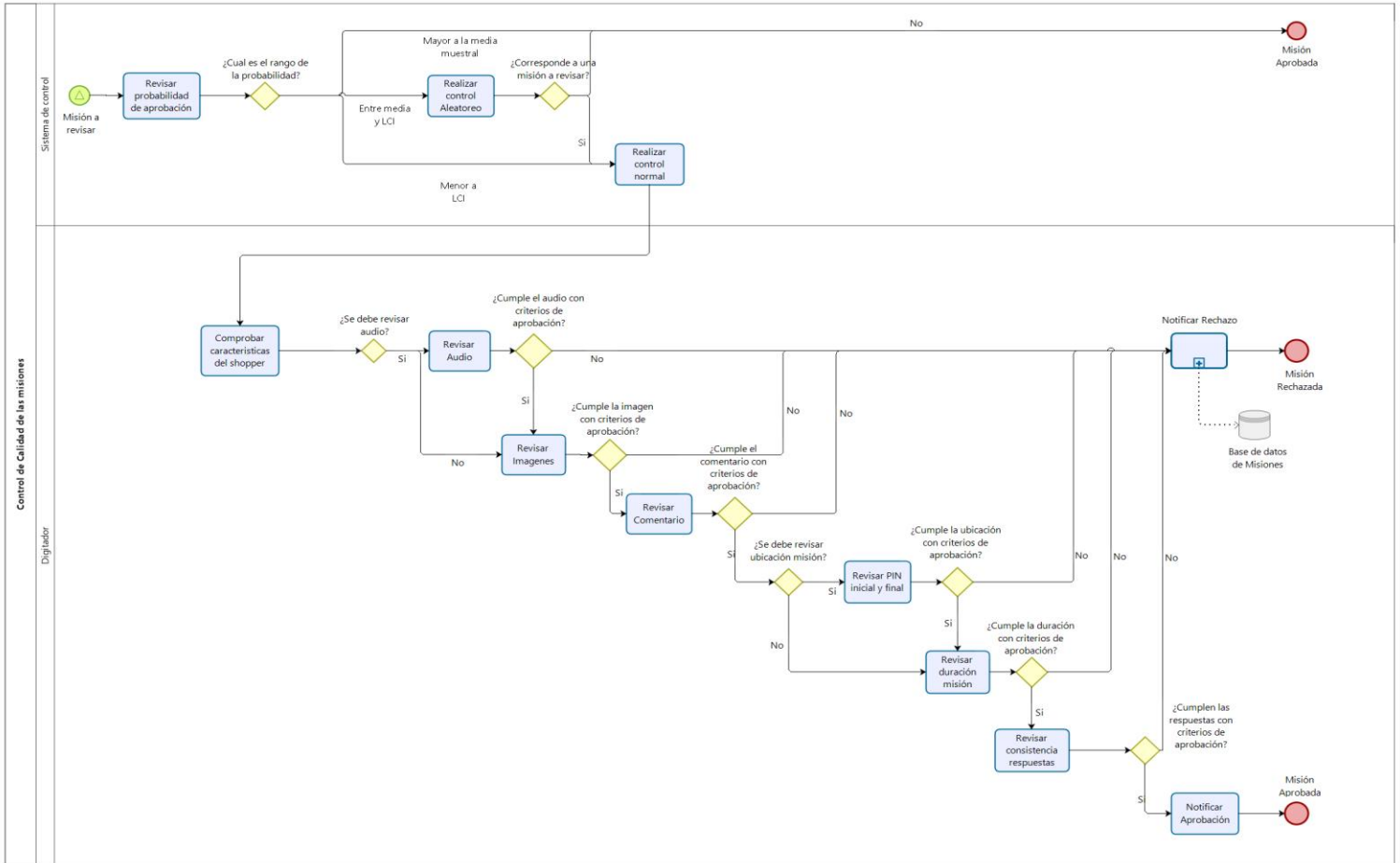


Ilustración 63: Diagrama BPMN del rediseño aplicado al proceso de revisión.  
Fuente: Elaboración Propia.

## 11.EVALUACIÓN TEÓRICA DE LOS CAMBIOS PROPUESTOS

### 11.1.Mejoras potenciales de la variable crítica

En primer lugar, se analizarán los cambios en la variable crítica escogida que generaría la implementación de la solución propuesta. Actualmente, considerando que el tiempo promedio de revisión es de 1 minuto y 40 segundos, a los digitadores les tomaría aproximadamente 22,2 horas revisar la meta diaria de 800 misiones. En cambio, gracias al sistema de control los digitadores solo deberían revisar el 10,8% del total de misiones (suponiendo que las probabilidades de aprobación de las 800 misiones diarias poseen la misma distribución que la presentada en la ilustración 61).



Esto se traduciría en que los digitadores solo deberían revisar un total aproximado de 86 misiones diarias y aun así se cumpliría la meta diaria. Dado esto, la variable crítica pasaría de 22,2 a 2,4 horas, lo cual corresponde a una reducción del 91%. Extrapolando este resultado, se podrían revisar aproximadamente 9.600 misiones por día, considerando que los 3 digitadores tienen una jornada de 9,5 horas y, por ende, cada uno debería revisar un total de 344 misiones (sin tener que trabajar en jornadas extralaborales). Esto significaría que Rocketpin podría aumentar el volumen diario de verificaciones realizadas en un 400% (de 2.400 a 9.600). De hecho, con tan solo 1 digitador, se podrían revisar en día un total de 3.200 misiones, lo cual supera con creces a las 1.350 misiones que revisaban en promedio los 3 digitadores.

## **11.2.Factibilidad financiera de la solución propuesta**

### **11.2.1.Costos del proceso de revisión actual**

Para analizar la factibilidad de la solución versus el proceso actual, se compararán los indicadores financieros (VAN y TIR) provenientes del flujo de caja del negocio. Dentro de los costos de calidad que posee el proceso actual de revisión se encuentran:

- **Costos de evaluación:** corresponden a las pruebas realizadas para asegurar la calidad del producto, que para el proceso de revisión serían realizadas por los digitadores. Dado esto, se considerará en este ítem el salario del personal del área de control de calidad, lo cual corresponde a un total de \$21.000.000 anuales.
- **Costos de prevención:** el proceso actual no cuenta con ningún método para prevenir errores, por lo cual no se considerará este costo.
- **Costo de falla interna:** el reprocesamiento de misiones debido a que son aprobadas por Rocketpin y rechazadas por la empresa contratista se incluirá en este ítem. No existen estimaciones basadas en datos, pero se aceptará la aproximación entregada por el área de control de calidad, la cual indica que la tasa de falsos positivos está comprendida entre un 2,5 a 10% del total de misiones aprobadas diariamente. Esto quiere decir que, de las 2.400 misiones que deberían ser revisadas diariamente, 2.170 son aprobadas (90,4%) y 109 corresponden a falsos positivos (considerando una tasa de falsos positivos igual a 5%). Dado que se debe pagar nuevamente a los shoppers para que realicen estas misiones, el costo diario por reprocesamiento ascendería a un total de \$119.534 y anualmente a \$32.988.121 (considerando 23 días hábiles por mes).

- **Costos de falla externa:** corresponden a las multas por incumplimiento de los estándares de calidad pactados en el contrato con EQUIFAX.

Para estimar los costos provenientes de las multas, se simula el proceso de revisión, para lo cual se utilizan las siguientes consideraciones y supuestos:

- i. El total de misiones que realizarán los shoppers por mes, será de 40.200, lo cual es aproximadamente el promedio mensual de 2019.
- ii. La revisión de estas misiones se repartirá de forma equitativa entre los digitadores, por lo cual cada uno revisará un total de 13.600 misiones mensuales.
- iii. El plazo de revisión será de un mes, lo cual para Rocketpin comprende un total de 23 días hábiles de trabajo (generalmente se utilizan 22, pero los digitadores trabajan un sábado al mes).
- iv. Para facilitar la estimación del costo, se supondrá que todas las misiones se realizan en la región metropolitana, ya que existe una diferenciación para las multas dependiendo de su origen.
- v. Los shoppers realizan un total de 800 misiones diarias, sin excepciones.
- vi. Como se apreció en la sección 3 de este informe, se considerará que los digitadores revisan en promedio un total de 450 misiones diarias.
- vii. Se consideran aquellas misiones que se revisen posteriores al plazo de un mes como retrasadas, para lo cual se contará cada día hábil pasada la fecha de término, con el fin de estimar la multa correspondiente por incumplimiento del plazo pactado.

La simulación del proceso de revisión realizando todas las consideraciones y supuestos, puede apreciarse en la siguiente ilustración:

Días de trabajo	Misiones diarias	Revisiones diarias	Misiones pendientes	Retraso [Días]
1	800	450	350	0
2	800	450	700	0
3	800	450	1.050	0
4	800	450	1.400	0
5	800	450	1.750	0
6	800	450	2.100	0
7	800	450	2.450	0
8	800	450	2.800	0
9	800	450	3.150	0
10	800	450	3.500	0
11	800	450	3.850	0
12	800	450	4.200	0
13	800	450	4.550	0
14	800	450	4.900	0
15	800	450	5.250	0
16	800	450	5.600	0
17	800	450	5.950	0
18	0	450	5.500	0
19	0	450	5.050	0
20	0	450	4.600	0
21	0	450	4.150	0
22	0	450	3.700	0
23	0	450	3.250	0
24	0	450	2.800	1
25	0	450	2.350	2
26	0	450	1.900	3
27	0	450	1.450	4
28	0	450	1.000	5
29	0	450	550	6
30	0	450	100	7
31	0	100	0	0

Ilustración 64: Simulación del proceso de revisión para estimar costos de falla  
Fuente: Elaboración Propia.

De la tabla, puede notarse que en un plazo de 17 días los shoppers serían capaces de cumplir con las 13.600 misiones que debe revisar 1 digitador al mes, ya que realizan 800 diarias. Para esa misma fecha, los digitadores tendrían un déficit de 5.950 misiones, lo cual se debe a que el proceso de revisión actual no permite sopesar el flujo diario de misiones. En el día 23, los digitadores tendrían un total de 3.250 misiones pendientes, por lo que necesariamente Rocketpin incurriría en multas por incumplimiento de contrato. El primer rango de multa corresponde a entre 2 a 6 días hábiles de retraso, por lo que desde el día 25, Rocketpin no recibirá el pago de 2.350 misiones. Posterior al sexto día de retraso, Rocketpin deberá pagar 2 veces el valor de la misión en concepto de multas, lo que según la tabla correspondería a 100 misiones.

Recapitulando, Rocketpin deberá pagar a los shoppers por 2.350 misiones de las cuales no recibirá ganancia alguna, lo que corresponde a un total de \$2.589.207. Además, tiene que pagar el doble del precio pagado por la empresa contratista por las 100 misiones con un atraso de 7 días, lo cual

asciende a un monto de \$314.798. El costo de falla externa mensual para 1 digitador sería de \$2.904.005. El costo anual que considera a los 3 digitadores ascendería a un total de \$104.544.162, lo cual representa un 15,4% de los ingresos anuales en términos de verificaciones domiciliarias (\$678.596.500 en 2019).

En total, los costos provenientes de la gestión de calidad para el servicio de verificación domiciliaria en el año 2019 ascienden a \$158.542.283, lo que corresponde a un 23,4% de los ingresos anuales, 9 veces superior al 2,5% que recomiendan los expertos en la temática [10, p. 311].

### **11.2.2. Costos de la solución propuesta**

Los costos de la calidad para la solución propuesta serían los siguientes:

- **Costos de evaluación:** manteniendo al personal actual del área de revisión, se deberá agregar el salario del ingeniero civil informativo y/o carrera afín que se encargará de supervisar la implementación de la solución. Esta labor puede externalizarse al subcontratar el servicio por hora de ingenieros presentes en diversos países, los cuales brindan tarifas más económicas y horarios de trabajo atractivos para las empresas contratistas.

El factor principal para considerar esta opción por sobre mantener a un ingeniero civil en computación de manera *full time*, es que hacerlo implicaría el riesgo de depender de este trabajador, ya que se le estaría confiando solamente a él la implementación total y regulación del proceso de revisión de misiones, el cual es fundamental en la operación del negocio de Rocketpin. En cambio, para el caso externalizado, la solución deberá estar debidamente documentada y explicada, y ser lo suficientemente genérica para ser fácilmente implementada y/o modificada por cualquier profesional del área, lo que otorgaría mayor flexibilidad a la organización.

El costo por hora de un ingeniero *freelance* proveniente de India es de aproximadamente 15 dólares [45] (un total de \$11.700, considerando el precio del dólar en \$780). Se estima que, trabajando aproximadamente 45 horas cada 3 meses, el ingeniero informático podría ser capaz de programar la solución a la medida de la organización, actualizarlo en base a las condiciones del sistema y monitorear la variable de referencia. En base a lo anterior, el costo anual de mantener a un controlador en el sistema es de \$2.106.000. Finalmente, este ítem contemplaría un total de \$23.106.000 anuales.

- **Costos de prevención:** se considerará la inversión a realizar en la implementación del sistema de control en este ítem. La solución propuesta no necesita el pago de softwares y licencias, ya que puede programarse en Python y SQL, los cuales son lenguajes de código abierto. Sin embargo, dada la necesidad de procesar constantemente grandes volúmenes de datos, será necesario que Rocketpin pague por un servicio de *Cloud Computing*. Este servicio corresponde a una tecnología que permite acceso remoto a softwares, almacenamientos de archivos y procesamiento de datos por medio de internet, siendo así, una alternativa a la ejecución en una computadora personal o servidor local [46]. Esto permitirá a Rocketpin procesar los datos de forma remota y no tener que invertir en computadores de alto rendimiento (activo fijo que se depreciaría en el corto plazo). El costo mensual de este servicio, para un almacenamiento de 50GB, uso de red de 10GB, 10.000 operaciones básicas de procesamiento de datos y 50.000 operaciones de mayor complejidad, corresponde a 2,57 dólares (un total de \$2.000, considerando el precio del dólar en \$780) en la plataforma Google Cloud [47]. Anualmente el costo ascendería a \$24.000.
- **Costos de falla interna:** se puede apreciar de la ilustración 56 que la cantidad de falsos positivos corresponde a 1.220 misiones, lo cual representa un 0,75% del total perteneciente a la base de *testing* (160.934 misiones). Extrapolando este resultado al flujo diario de 2.400 misiones para el 2019, se tiene que aproximadamente 18 misiones corresponderían a falsos positivos, lo que implicaría un costo de \$19.818 diarios. Anualmente este costo ascendería a \$5.469.768.
- **Costos de falla externa:** Rocketpin no incurriría en este tipo de costos al implementar la solución actual, ya que puede revisar un total de 9.600 misiones diarias manteniendo al personal actual del área de control de calidad.

Se tiene que los costos de calidad provenientes de la implementación del rediseño propuesto, tomando como base el año 2019, corresponden a \$28.599.768, lo cual representa un 4,21% del total de ingresos obtenidos para ese año.

### **11.2.3. Estimación de los indicadores financieros**

Para estimar los indicadores financieros, se realizará un flujo de caja a un horizonte de 5 años. Se hace necesario inferir cómo será el crecimiento del volumen de verificaciones domiciliarias que se realizarán anualmente. Tomando como referencias las facturaciones relativas a las verificaciones

domiciliarias de los años 2018 y 2019, las cuales fueron de \$500.000.000 y \$678.596.500 respectivamente, se puede estimar la tasa de crecimiento de los ingresos en un 35,7% anual. Esta cifra es bastante creíble si se considera la constante entrada al mercado chileno de aplicaciones que basan su modelo de negocios en la economía colaborativa y que múltiples organizaciones internacionales prevén que las facturaciones de este tipo de negocios aumentaran en un 2.000% para el año 2025 en comparación al año 2015 [3].

En la ilustración a continuación se presentan las facturaciones correspondientes a un horizonte de 5 años, la estimación del costo por pago a shoppers y la cantidad de misiones realizadas a nivel anual, mensual y diario:

Año	Cantidad Facturada	Pago a shoppers	Misiones Anuales	Misiones mensuales	Misiones diarias
1	\$ 678.596.500	\$ 475.016.850	536.444	40.385	2.356
2	\$ 920.986.420	\$ 644.226.487	585.129	48.761	2.720
3	\$ 1.249.956.322	\$ 874.339.678	794.132	66.178	3.477
4	\$ 1.696.431.970	\$1.186.647.691	1.077.791	89.816	4.505
5	\$ 2.302.385.595	\$1.610.509.940	1.462.770	121.898	5.900

Ilustración 65: Ítems necesarios para la estimación del flujo de caja.

Fuente: Elaboración Propia.

Dada la tasa de crecimiento estimada, es posible apreciar que, para el cuarto año Rocketpin podría ver triplicada su facturación en comparación al primer año. Otro factor por considerar corresponde a que la cantidad de misiones diarias a revisar no superan, siquiera en el quinto año de operación estimado, las 9.600 misiones que puede revisar el equipo de control de calidad gracias a la potencial implementación del sistema de control diseñado. En cambio, para la situación actual, Rocketpin debería contratar a 3 digitadores el primer año para aumentar la capacidad del proceso, ya que contando con 6 digitadores en el área de control de calidad se podrían revisar 2.700 misiones diarias aproximadamente, considerando que cada uno en promedio revisa 450 misiones al día.

Siguiendo esta lógica, Rocketpin deberá contratar a 2 digitadores más para el tercer año, en el cuarto a otros 2 y finalmente, en el quinto año se tendrá que contar con un total de 13 digitadores para disminuir al mínimo los costos producidos por las multas. Se supondrá que esta contratación de personal generará una disminución de este costo de falla externa de un 15,4% de los ingresos en 2019 (\$104.544.162) a un 5% sostenible en el tiempo. Tomando todas estas consideraciones, el flujo de caja para la situación actual y la solución propuesta pueden verse a continuación:

Años	1	2	3	4	5
Ingresos por verificaciones	\$678.596.500	\$920.986.420	\$1.249.956.322	\$1.696.431.970	\$2.302.385.595
<b>Costos Fijos</b>					
Sueldos Digitadores	\$-28.800.000	\$-28.800.000	\$-38.400.000	\$-48.000.000	\$-62.400.000
Sueldo Jefe Área	\$-6.600.000	\$-6.600.000	\$-6.600.000	\$-6.600.000	\$-6.600.000
<b>Costos Variables</b>					
Reprocesamiento	\$-32.988.121	\$-37.386.961	\$-47.795.541	\$-61.921.992	\$-81.094.312
Multas	\$-33.929.825	\$-46.049.321	\$-62.497.816	\$-84.821.599	\$-115.119.280
Pago Shoppers	\$-475.016.850	\$-644.688.738	\$-874.967.043	\$-1.187.499.146	\$-1.611.665.528
Utilidad antes de impuesto	\$101.261.704	\$157.461.400	\$219.695.922	\$307.589.234	\$425.506.475
Impuestos (27%)	\$-27.340.660	\$-42.514.578	\$-59.317.899	\$-83.049.093	\$-114.886.748
Flujo total	\$73.921.044	\$114.946.822	\$160.378.023	\$224.540.140	\$310.619.727

Ilustración 66: Flujo de caja de Rocketpin considerando la situación actual.

Fuente: Elaboración Propia.

Años	1	2	3	4	5
Ingresos por verificaciones	\$678.596.500	\$920.986.420	\$1.249.956.322	\$1.696.431.970	\$2.302.385.595
<b>Costos Fijos</b>					
Sueldos Digitadores	\$-14.400.000	\$-14.400.000	\$-14.400.000	\$-14.400.000	\$-14.400.000
Sueldo ingeniero informático	\$-2.106.000	\$-2.106.000	\$-2.106.000	\$-2.106.000	\$-2.106.000
Sueldo Jefe Área	\$-6.600.000	\$-6.600.000	\$-6.600.000	\$-6.600.000	\$-6.600.000
<b>Costos Variables</b>					
Reprocesamiento	\$-5.469.768	\$-6.199.070	\$-7.924.326	\$-10.267.210	\$-13.446.513
Pago Google Cloud	\$-24.000	\$-37.680	\$-59.158	\$-92.877	\$-145.818
Pago Shoppers	\$-475.016.850	\$-644.688.738	\$-874.967.043	\$-1.187.499.146	\$-1.611.665.528
Utilidad antes de impuesto	\$174.979.882	\$246.954.932	\$343.899.795	\$475.466.737	\$654.021.736
Impuestos (27%)	\$-47.244.568	\$-66.677.832	\$-92.852.945	\$-128.376.019	\$-176.585.869
Flujo total	\$127.735.314	\$180.277.100	\$251.046.850	\$347.090.718	\$477.435.867

Ilustración 67: Flujo de caja de Rocketpin considerando la solución propuesta.

Fuente: Elaboración Propia.

La implementación de la solución aumentaría los beneficios netos de Rocketpin a más del doble que utilizando el proceso de revisión actual. Para calcular el VAN y la TIR, se debe encontrar la tasa de descuento asociada al negocio. Basándose en el mercado estadounidense, la tasa de descuento otorgada a *Start-Ups* basadas en el uso intensivo de tecnología es de 14,13% [48]. El VAN del negocio relativo al proceso de revisión actual corresponde a \$553.647.282 y el que considera la implementación del sistema de control es igual a \$870.320.588. Dado esto, se extrae que el valor actual neto de Rocketpin en base al sistema de control es superior en \$316.673.306 al del proceso actual. Con respecto a la TIR, no es una métrica aconsejable a utilizar

para evaluar el proyecto, puesto que éste no contempla inversión ni posee flujos negativos, y dado que esta métrica corresponde a la tasa de descuento con la cual el VAN se hace igual a 0, su estimación correspondería a un porcentaje demasiado alto (mayor a un 100%) para los 2 casos a estudiar, y, por ende, no aportaría ningún valor añadido a la evaluación del negocio.

## **12.CONCLUSIONES**

### **12.1.Conclusiones**

Del trabajo realizado a lo largo de esta memoria, se pueden obtener las siguientes conclusiones:

En primer lugar, se destaca que el mercado de los negocios basados en la economía colaborativa está en constante crecimiento, ya que permiten generar un mayor aprovechamiento de los bienes y recursos existentes, lo cual genera una intermediación entre la oferta y la demanda de forma rápida, eficiente y menos costosa. Esta tendencia a la colaboración entre iguales se ve potenciada por el fenómeno de globalización, donde los mercados internacionales están cada vez más correlacionados y existe una mayor facilidad en la comunicación entre las personas gracias a los medios digitales. Por lo cual, el ambiente en el que se desarrollan este tipo de empresas es altamente competitivo, existiendo múltiples incentivos a mejorar la calidad del servicio brindado, reducir tiempos de espera y costos. En este mercado es donde se desenvuelve Rocketpin y, por ende, se hace fundamental optimizar sus procesos para aumentar sus probabilidades de supervivencia.

Rocketpin es una organización dedicada a realizar servicios de levantamiento de información en terreno, por medio de usuarios que descargan su aplicación móvil. Dentro de todos los servicios de levantamiento de información que provee Rocketpin, el que representa el mayor volumen y porcentaje del total de ventas corresponde a las verificaciones domiciliarias. Para brindar este servicio, los shoppers deben básicamente comprobar que la persona a verificar vive efectivamente en el domicilio señalado (lo que Rocketpin denomina como "misiones") y posteriormente, llenar un formulario con toda la información recabada. Debido a los altos estándares de calidad exigidos a Rocketpin por parte de la empresa contratista (EQUIFAX), fue necesario generar un área que se especializara en comprobar la validez de las misiones enviadas por los shoppers. Esta nueva área, bautizada como control de calidad, está compuesta por 3 digitadores, los cuales revisan el contenido de cada misión y por un jefe de área, quien se encarga de supervisar este proceso. El problema que surge de este proceso se debe a que, para cumplir con los estándares de calidad, se exige una revisión del 100% de las misiones realizadas por los shoppers, lo



cual produce una sobreutilización de los digitadores, altos costos debido al retraso en la provisión del servicio y no permite a la organización mantener su crecimiento. Sumado a lo anterior y a pesar de guardar información relativa a la operación del negocio y tener suposiciones acerca de los factores que más influyen en el rechazo o desaprobación de las misiones, Rocketpin no posee ninguna metodología, modelo y/o métrica que les permita validar estas creencias. Dado todo lo anterior, se hace urgente rediseñar el proceso de revisión, buscando mantener los estándares de calidad solicitados, pero disminuyendo los tiempos y costos asociados.

En la etapa de investigación de esta memoria, se pudo comprobar formalmente los problemas generados por el proceso de revisión actual. Se estimó que el flujo de misiones diarias es de aproximadamente 2.400, lo que equitativamente representaría un total de 800 misiones a revisar por cada digitador. Sin embargo, éstos solo pueden revisar en promedio un total de 450 misiones diarias, lo que representa un 56,25% de la meta. Además, mediante la medición del tiempo de revisión para un total de 150 misiones, se pudo apreciar que en promedio les toma 1 minuto y 40 segundos a los digitadores aprobar o rechazar una misión, lo cual indica que para completar el flujo diario de misiones deberían trabajar aproximadamente 22,2 horas. De hecho, extrapolando este resultado, a los digitadores les tomaría 12,5 horas revisar las 450 misiones que revisan a diario. Esto explica por qué señalan que trabajan en jornadas extralaborales, ya que su jornada diaria solo es de 9,5 horas. En términos económicos, se demostró que el proceso de revisión actual generó costos de retrabajo de \$32.988.121 anuales y costos por infracciones de contrato por \$104.544.162. Estos costos representan un 20,2% de los ingresos generados por verificaciones domiciliarias el año 2019. Todos los factores señalados anteriormente, comprobaron la necesidad de hacer más eficiente el proceso actual.

Para realizar esta labor, se estudió la revisión de misiones mediante un enfoque de procesos, con el objetivo de comprender la serie de pasos que seguían los digitadores para verificar la veracidad de la información. Además, este análisis permitió entender las suposiciones realizadas por los directivos y digitadores de Rocketpin acerca de los factores más influyentes en la aprobación y rechazo de misiones. En base a estas creencias, se procedió a comprobar formalmente el efecto de estos factores en la clasificación de las misiones como aprobadas o rechazadas. Lo anterior, se realizó en 3 etapas diferentes. La primera consistió en un análisis exploratorio de las bases de misiones y shoppers, donde mediante elementos gráficos y estudios de correlación, se pudieron encontrar indicios formales sobre las variables más influyentes en la aprobación de misiones, las cuales calzaban casi en su totalidad con las suposiciones realizadas por directivos y digitadores de

Rocketpin. La segunda etapa, basada en el control del proceso mediante el estudio de la variabilidad de los factores más influyentes, permitió descubrir los límites aceptables de estas variables. La última etapa de este análisis tuvo como finalidad estimar formalmente el efecto de estas variables en la probabilidad de aprobación de las misiones, lo cual se realizó al aplicar un modelo de regresión logística. De este análisis, se concluyó que tanto shoppers con mayor confiabilidad y cantidad de misiones realizadas, aumentaban la probabilidad de aprobación, en cambio, la distancia entre el PIN inicial y final, duración y remuneración de la misión, incidían negativamente en la aprobación. También, se determinó que variables que se suponían influyentes no tenían ningún efecto en la probabilidad de aprobación, las cuales corresponden a la experiencia del shopper y el tiempo entre tomada e iniciada la misión.

Para validar los resultados obtenidos del algoritmo, se estudió su capacidad predictiva, lo cual se realizó entrenando al modelo con un 70% de los datos originales y probando su efectividad al predecir la clasificación de las misiones en el 30% restante de los datos. Además, era primordial evaluar que el modelo fuera capaz de predecir correctamente las misiones rechazadas, ya que su errónea clasificación es la que trae mayores complicaciones y costos a Rocketpin. La comprobación de la eficacia del algoritmo se realizó al estudiar métricas de evaluación de sus predicciones, las cuales miden el error entre la clasificación real y la predicha. De este análisis, se concluyó que la regresión logística tenía una capacidad predictiva notable para predecir el estado de las misiones, ya que pudo distinguir el 92,1% de las misiones rechazadas, con una exactitud de 98,9% para predecir el total de misiones y una capacidad de distinguir entre clases de un 95,6%, lo cual permitió validar que las variables escogidas eran las más influyentes a la hora de clasificar una misión.

En base a toda la investigación realizada, se determinó que la mejor solución para rediseñar el proceso de revisión consistía en elaborar un sistema de control basado en la capacidad de la regresión logística para predecir la probabilidad de aprobación. El diseño de esta solución se basó en la creación de niveles de exhaustividad de revisión. Estos niveles se generaron en base a los límites de control provenientes de la gráfica de control del tipo X de la probabilidad de aprobación. Se estimaron 3 niveles de revisión: el cuadrante I, II y III. Las misiones ubicadas en el primer cuadrante serán aprobadas automáticamente debido a su alta probabilidad de aprobación. Para el cuadrante II se realizará un control aleatorio del 50% de las misiones priorizando aquellas hechas por shoppers con baja confiabilidad y, por último, el cuadrante III contempla aquellas misiones que deben ser revisadas en su totalidad. De esta parte del trabajo de memoria, se concluye que para implementar este sistema de control será necesario contar con un controlador

capaz de actualizar y actuar en base a cambios en el porcentaje total de misiones que debería poseer cada cuadrante (variable de referencia). Este controlador deberá ser un profesional capaz de manejar múltiples lenguajes de programación, tener un alto conocimiento de estadística y algoritmos de *machine learning*, por lo cual se consideró que la subcontratación de un ingeniero civil en computación y/o carrera afín calzaría con el perfil del cargo. Dentro de los diferentes tipos de sistemas de control, se prefirió la utilización de uno de lazo cerrado, ya que puede retroalimentarse mediante un sensor, el cual permite mantener informado constantemente al controlador sobre los errores generados entre la variable de referencia y salida del proceso.

La implementación de este sistema de control permitiría aumentar el flujo de misiones que pueden revisar 1 digitador de 1.350 a 3.200, lo que significa una capacidad diaria de 9.600 misiones diarias considerando a los 3 digitadores del área de control de calidad (4 veces más que el flujo diario actual de 2.400 misiones). Esto se debe principalmente a la reducción de la variable crítica escogida, ya que el tiempo necesario para revisar las 800 misiones diarias se reduciría de 22,2 a 2,4 horas, puesto que los digitadores solo deberán revisar el 10,8% del flujo total. Se concluye de este análisis, que la potencial implementación del sistema de control propuesto permitiría asegurar la calidad del servicio de Rocketpin de forma más eficiente y rápida. Finalmente, respecto al análisis financiero de la situación actual versus el rediseño propuesto, se estimó que la diferencia entre el valor actual neto (VAN) de ambos proyectos es de \$316.673.306 a favor del sistema de control, con lo cual se puede concluir que éste permite aumentar los flujos de capital de Rocketpin mediante una reducción de sus costos operacionales, lo cual logra validar el cumplimiento del objetivo general de esta memoria.

## **12.2.Recomendaciones**

Con respecto al funcionamiento de Rocketpin y a su manejo de la información, se tienen las siguientes recomendaciones:

La organización deberá preocuparse en mayor medida de asegurar la confiabilidad de los shoppers que ingresan a la aplicación, ya que esto permitiría facilitar el proceso de control de calidad. Para esto, se podría implementar un modelo de clasificación que permita predecir qué shoppers serán confiables en base a sus características. Sin embargo, Rocketpin no posee la información necesaria de sus agentes para desarrollar este tipo de análisis, por lo que se le recomienda generar instancias para aumentar la cantidad de información y variables que tienen almacenada acerca de sus agentes.

Otra recomendación que surge de la deficiente gestión del manejo de los datos consiste en recolectar información relativa al largo de los audios y calidad de las fotografías tomadas por los shoppers en las misiones de verificación domiciliaria. Como se pudo apreciar en la etapa de investigación, estos factores se suponían importantes a la hora de determinar la clasificación de las misiones, pero no se pudo comprobar su validez debido a que Rocketpin no gestiona esta información.

Con respecto al diseño e implementación de la solución propuesta se tienen las siguientes recomendaciones:

La regresión logística se escogió dado que permite determinar el efecto que poseen las variables en el fenómeno a estudiar, sin embargo, en el amplio mundo del aprendizaje de máquinas y *Deep learning*, existen modelos más complejos que podrían tener aún mejores resultados en la predicción de la clasificación de misiones. Algunos ejemplos de estos modelos, que se basan en la unión de múltiples árboles de decisión, corresponden a: *Light GBM*, *XGBoost* y *CatBoost*. Estos 3 modelos han ganado diversas competencias en torneos de clasificación y actualmente, son considerados como los mejores para este tipo de problemas [49].

Finalmente, se recomendaría a Rocketpin en un mediano o largo plazo, automatizar el proceso de revisión de misiones, puesto que cuando el flujo diario de misiones supere el umbral de 9.600 obtenido mediante la potencial implementación del rediseño propuesto, la empresa incurrirá nuevamente en costos por incumplimientos de contrato y/o contratación de nuevos digitadores. Estos costos pueden evitarse automatizando el proceso, el cual además puede verse potenciado en términos de capacidad predictiva, si Rocketpin recaba más información de la operación del negocio e implementa un mejor modelo de clasificación que la regresión logística.

### **12.3.Comentarios finales**

El diseño de la solución escogida fue presentado a Rocketpin en una reunión con la dirigencia y el personal del área de control de calidad. Además, se les explicaron las diferentes métricas utilizadas para evaluar el modelo y los resultados potenciales que podría conllevar su implementación. Las reacciones fueron variadas. La dirigencia estuvo bastante de acuerdo con el sistema de control de gestión basado en la probabilidad de aprobación, ya que la reducción tanto de costos como de la ocupación del personal sería bastante destacable, pudiendo aumentar el flujo diario de misiones y mantener los estándares de calidad. Respecto al nuevo puesto que generaría la implementación de este sistema (controlador), la dirigencia estuvo de acuerdo

con que era una buena alternativa, puesto que han contratado de manera remota a ingenieros venezolanos y peruanos, los cuales han hecho trabajos impecables y tienen muy buena reputación. Sin embargo, el área de control de calidad, indicó que podría generar complicaciones no tener contacto directo con la persona encargada de entregarles el compilado de misiones a revisar, puesto que dada la modalidad de trabajo propuesta para el controlador (45 horas cada 3 meses), reduciría la capacidad de entregarle *feedback* sobre su labor y haría menos ágil la revisión de misiones que son falsos positivos (aprobadas por Rocketpin y rechazadas por la empresa contratista), puesto que el sistema de control solo entregaría como *output* el compilado de misiones a revisar.

Con respecto al porcentaje del total de misiones a revisar (10,8%), existió un consenso respecto a que podría ser una medida muy agresiva en el corto plazo, ya que existe cierto temor a que cambie el comportamiento tanto de los shoppers como de las misiones, lo cual podría generar una disminución en la capacidad predictiva de la regresión logística, produciendo el aumento de multas y/o pérdida de la imagen de Rocketpin. Lo anterior, también se debe a que la organización tiene arraigado el pensamiento acerca de que revisar el 100% de las misiones es un factor decisivo para el aseguramiento de la calidad de su servicio, por lo que existe cierta reticencia al cambio en este sentido.

Es por esto que, en conjunto con el memorista, se decidió que la solución propuesta debería lanzarse de manera piloto, para lo cual se sugirió que se probase con un número reducido de misiones de verificación domiciliaria, las cuales podrían filtrarse por el cliente que las solicita (para más detalles ver ilustración 30) y donde el controlador podría ser, en primera instancia, algún informático del área de tecnología. Esto sería de suma utilidad para Rocketpin, ya que les permitiría comprobar la eficacia de la solución sin contraer tantos riesgos y esta, podría ser escalable a todas las misiones en el corto-mediano plazo. Además, daría tiempo al área de control de calidad para adaptarse al nuevo proceso de revisión, ya que deberán ser más minuciosos para revisar el compilado de misiones y seguir rigurosamente la rúbrica de trabajos señaladas en el diagrama BPMN del rediseño propuesto.

Finalmente, se extraen múltiples aprendizajes de la presentación realizada a Rocketpin. Se infiere que para un ingeniero civil industrial es sustancial tener la capacidad de comunicar las ideas con claridad, escuchar y entender las necesidades del cliente e identificar la cultura organizacional de las empresas en las que trabaja, ya que en muchos casos existe la tendencia a preferir soluciones complejas (o estrictas) basadas en herramientas y/o tecnologías emergentes, a pesar de que una solución simple puede ser la que mejor se adapte a la cultura, personas y requerimientos de la organización.

### **13. BIBLIOGRAFÍA**

[1] CIPPEC. ¿De qué hablamos cuando hablamos de economía de plataformas? [En línea] <<https://www.revista-uno.com/número-20/economia-colaborativa-la-revolucion-del-consumo-mundial/>> [Consulta: 03 octubre 2019]

[2] LLYC. 2015. Economía colaborativa: la revolución del consumo mundial. [En línea] <<https://www.revista-uno.com/número-20/economia-colaborativa-la-revolucion-del-consumo-mundial/>> [Consulta: 03 octubre 2019]

[3] Las plataformas de economía colaborativa facturarán 335.000 millones en 2025. [En línea] Decoop Chile. 8 de abril, 2018. <<http://www.decoopchile.cl/las-plataformas-de-economia-colaborativa-facturaran-335-000-millones-en-2025/>> [Consulta: 03 octubre 2019]

[4] THE OSTELEA SCHOOL OF TOURISM & HOSPITALITY. PLATAFORMAS DE ECONOMÍA COLABORATIVA: UNA MIRADA GLOBAL. [En línea] <[http://www.aept.org/archivos/documentos/ostelea\\_informe\\_economia\\_colaborativa.pdf](http://www.aept.org/archivos/documentos/ostelea_informe_economia_colaborativa.pdf)> [Consulta: 03 octubre 2019]

[5] Cinco años de Uber en Chile: app traslada a más de dos millones de pasajeros al mes. [En línea] El dinamo. 1 de febrero, 2019. <<https://www.eldinamo.cl/nacional/2019/02/01/cinco-anos-de-uber-en-chile-app-traslada-a-mas-de-dos-millones-de-pasajeros-al-mes/>> [Consulta: 03 octubre 2019]

[6] MediaInteractive, Pontificia Universidad Católica de Chile y TrenDigital. LA ECONOMÍA DE LAS PLATAFORMAS COLABORATIVAS EN CHILE: USO Y PERCEPCIÓN DE LOS CHILENOS. [En línea] <<http://www.mediainteractive.cl/informe-plataformas-colaborativas.pdf>> [Consulta: 03 octubre 2019]

[7] BRAVO CARRASCO, J. 2018. Rediseño de Procesos. Santiago, Chile. Editorial Evolución S.A. 257 p.

[8] "[http://resources.bizagi.com/docs/BPMN\\_Guia\\_de\\_Referencia\\_ESP.pdf](http://resources.bizagi.com/docs/BPMN_Guia_de_Referencia_ESP.pdf)"

[9] CAMISÓN, C., CRUZ, S. y GONZÁLEZ, T. 2006. Gestión de la calidad: conceptos, enfoques, modelos y sistemas. Madrid. PEARSON EDUCATION S.A. 1.464 p.

- [10] CHASE, R., JACOBS, R. y AQUILANO, N. 2009. Administración de Operaciones: Producción y Cadena de Suministros. 12° ed. McGraw-Hill. 800 p.
- [11] SANCHO, F. 2019. Aprendizaje Supervisado y No Supervisado. [En línea] <<http://www.cs.us.es/~fsancho/?e=77>> [Consulta: 10 enero 2020]
- [12] CALVO, D. 2019. Aprendizaje supervisado. [En línea] <<http://www.diegocalvo.es/aprendizaje-supervisado/>> [Consulta: 10 enero 2020]
- [13] ZAMBRANO, J. 2018. ¿Aprendizaje supervisado o no supervisado? Conoce sus diferencias dentro del machine learning y la automatización inteligente. [En línea] <<https://medium.com/@juanzambrano/aprendizaje-supervisado-o-no-supervisado-39ccf1fd6e7b>> [Consulta: 10 enero 2020]
- [14] ROMAN, V. Aprendizaje Supervisado: Introducción a la Clasificación y Principales Algoritmos. [En línea] <<https://medium.com/datos-y-ciencia/aprendizaje-supervisado-introducci%C3%B3n-a-la-clasificaci%C3%B3n-y-principales-algoritmos-dadee99c9407>> [Consulta: 10 enero 2020]
- [15] CÁCERES, J. 2019. Reconocimiento de patrones y aprendizaje no supervisado. [En línea] <[https://www.researchgate.net/profile/Jesus\\_Tello/publication/228857048\\_Reconocimiento\\_de\\_patrones\\_y\\_el\\_aprendizaje\\_no\\_supervisado/links/0c960517e7e677b522000000.pdf](https://www.researchgate.net/profile/Jesus_Tello/publication/228857048_Reconocimiento_de_patrones_y_el_aprendizaje_no_supervisado/links/0c960517e7e677b522000000.pdf)> [Consulta: 13 enero 2020]
- [16] AJTUN, J. Regresión Logística. [En línea] <<https://www.sutori.com/story/regresion-logistica--uQJxeraWVgZEcvhugaCNTwt4>> [Consulta: 17 noviembre 2019]
- [17] LEÓN, F. 2019. Clasificación de empresas en Chile. [En línea] <<https://www.rankia.cl/blog/ideas-emprendimiento-chile/3644120-clasificacion-empresas-chile>> [Consulta: 20 diciembre 2019]
- [18] C. BATANERO, A. ESTEPA y J. GODINO. 1991. ANÁLISIS EXPLORATORIO DE DATOS: SUS POSIBILIDADES EN LA ENSEÑANZA SECUNDARIA. [En línea] <[https://www.researchgate.net/publication/255765681\\_ANALISIS\\_EXPLORATORIO\\_DE\\_DATOS\\_SUS\\_POSIBILIDADES\\_EN\\_LA\\_ENSEÑANZA\\_SECUNDARIA](https://www.researchgate.net/publication/255765681_ANALISIS_EXPLORATORIO_DE_DATOS_SUS_POSIBILIDADES_EN_LA_ENSEÑANZA_SECUNDARIA)> [Consulta: 21 diciembre 2019]

[19] W. OUTWAITE y S. P. TURNER. 2007. The SAGE Handbook of Social Science Methodology. SAGE Publications Ltd. 623 p.

[20] DAGNINO, J. 2015. DATOS FALTANTES (MISSING VALUES). [En línea] <<https://revistachilenadeanestesia.cl/datos-faltantes-missing-values/>> [Consulta: 26 diciembre 2019]

[21] Scheffer, J. 2002. Dealing with Missing Data. [En línea] <[https://mro.massey.ac.nz/bitstream/handle/10179/4355/Dealing\\_with\\_Missing\\_Data.pdf?sequence=1&isAllowed=y](https://mro.massey.ac.nz/bitstream/handle/10179/4355/Dealing_with_Missing_Data.pdf?sequence=1&isAllowed=y)> [Consulta: 26 diciembre 2019]

[22] E. ACUÑA y C. RODRIGUEZ. 2004. The Treatment of Missing Values and its Effect on Classifier Accuracy. [En línea] <[https://link.springer.com/chapter/10.1007/978-3-642-17103-1\\_60](https://link.springer.com/chapter/10.1007/978-3-642-17103-1_60)> [Consulta: 26 diciembre 2019]

[23] datacarpentry.org. Data Types and Formats. [En línea] <<https://datacarpentry.org/python-ecology-lesson/04-data-types-and-format/>> [Consulta: 30 diciembre 2019]

[24] MIGUEL, J. 2011. ¿Cómo calcular la distancia entre dos puntos geográficos en C#? (Fórmula de Haversine). [En línea] <<https://www.genbeta.com/desarrollo/como-calcular-la-distancia-entre-dos-puntos-geograficos-en-c-formula-de-haversine>> [Consulta: 30 diciembre 2019].

[25] VINUESA, P. 2016. Tema 8 - Correlación: teoría y práctica. [En línea] <[https://www.ccg.unam.mx/~vinuesa/R4biosciences/docs/Tema8\\_correlacion.html](https://www.ccg.unam.mx/~vinuesa/R4biosciences/docs/Tema8_correlacion.html)> [Consulta: 3 enero 2020]

[26] Ministerio de Desarrollo Social y Familia. Archivo histórico de Encuesta CASEN. [En línea] <[http://observatorio.ministeriodesarrollosocial.gob.cl/casen/casen\\_obj.php](http://observatorio.ministeriodesarrollosocial.gob.cl/casen/casen_obj.php)> [Consulta: 4 enero 2020]

[27] R. CARRO y D. GONZÁLEZ. CONTROL ESTADÍSTICO DE PROCESOS. [En línea] <[http://nulan.mdp.edu.ar/1617/1/12\\_control\\_estadistico.pdf](http://nulan.mdp.edu.ar/1617/1/12_control_estadistico.pdf)> [Consulta: 4 enero 2020]

[28] SÁNCHEZ, I. Control de procesos por atributos. [En línea] <[http://www.est.uc3m.es/esp/nueva\\_docencia/leganes/ing\\_telecomunicacion/metodos\\_mejora\\_calidad/MEMC/doc\\_generica/Temario/CapCAtributos/CapCAtributos.pdf](http://www.est.uc3m.es/esp/nueva_docencia/leganes/ing_telecomunicacion/metodos_mejora_calidad/MEMC/doc_generica/Temario/CapCAtributos/CapCAtributos.pdf)> [Consulta: 5 enero 2020]



- [29] HERNÁNDEZ, C. 2016. Aplicación del control estadístico de procesos (CEP) en el control de su calidad. [En línea] <[http://scielo.sld.cu/scielo.php?pid=S2224-61852016000100010&script=sci\\_arttext&lng=pt](http://scielo.sld.cu/scielo.php?pid=S2224-61852016000100010&script=sci_arttext&lng=pt)> [Consulta: 5 enero 2020]
- [30] SÁNCHEZ, I. Control de procesos por variables. [En línea] <[http://www.est.uc3m.es/esp/nueva\\_docencia/leganes/ing\\_telecomunicacion/metodos\\_mejora\\_calidad/MEMC/doc\\_generica/Temario/CapCVar/CapCVariables.pdf](http://www.est.uc3m.es/esp/nueva_docencia/leganes/ing_telecomunicacion/metodos_mejora_calidad/MEMC/doc_generica/Temario/CapCVar/CapCVariables.pdf)> [Consulta: 5 enero 2020]
- [31] M. DOLORES y J. C. RODRÍGUEZ. 2000. La regresión logística: una herramienta versátil. Revista Nefrología: 477-565. Vol. 20. Num. 6. [En línea] <<https://www.revistanefrologia.com/es-relacionados-la-regresion-logistica-una-herramienta-versatil-articulo-X0211699500035664>> [Consulta: 13 enero 2020]
- [32] PALACIOS, L. 2018. El Concepto de la Multicolinealidad. [En línea] <<https://medium.com/high-data/el-concepto-de-multicolinealidad-51713b18a0b1>> [Consulta: 14 enero 2020]
- [33] G. JAMES, D. WITTEN y T. HASTIE. 2013. An Introduction to Statistical Learning with Applications in R. Springer. 440 p.
- [34] BROWNLEE, J. 2019. How Much Training Data is Required for Machine Learning?. [En línea] <<https://machinelearningmastery.com/much-training-data-required-machine-learning/>> [Consulta: 17 enero 2020]
- [35] "[https://scikit-learn.org/stable/modules/generated/sklearn.model\\_selection.train\\_test\\_split.html](https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html)"
- [36] T. OTZEN y C. MANTEROLA. 2017. Técnicas de Muestreo sobre una Población a Estudio. [En línea] <[https://scielo.conicyt.cl/scielo.php?script=sci\\_arttext&pid=S0717-95022017000100037](https://scielo.conicyt.cl/scielo.php?script=sci_arttext&pid=S0717-95022017000100037)> [Consulta: 19 enero 2020]
- [37] FERRE, M. 2019. FEIR: Regresión Logística. [En línea] <<http://gauss.inf.um.es/feir/45/>> [Consulta: 20 enero 2020]
- [38] BARRIOS, J. I. 2019. La matriz de confusión y sus métricas. [En línea] <<https://www.juanbarrios.com/matriz-de-confusion-y-sus-metricas/>> [Consulta: 23 enero 2020]

- [39] GHONEIM, S. 2019. Accuracy, Recall, Precision, F-Score & Specificity, which to optimize on?. [En línea] <<https://towardsdatascience.com/accuracy-recall-precision-f-score-specificity-which-to-optimize-on-867d3f11124>> [Consulta: 24 enero 2020]
- [40] GOOGLE DEVELOPERS. Clasificación: Exactitud. [En línea] <<https://developers.google.com/machine-learning/crash-course/classification/accuracy>> [Consulta: 25 enero 2020]
- [41] KAGGLE. 2019. Demythifying Matthew Correlation Coefficients (MCC) [En línea] <<https://www.kaggle.com/ratthachat/demythifying-matthew-correlation-coefficients-mcc>> [Consulta: 27 enero 2020]
- [42] NARKHEDE, S. 2018. Understanding AUC - ROC Curve. [En línea] <<https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5>> [Consulta: 28 enero 2020]
- [43] SANJAY, M. 2018. Why and how to Cross Validate a Model?. [En línea] <<https://towardsdatascience.com/why-and-how-to-cross-validate-a-model-d6424b45261f>> [Consulta: 28 enero 2020]
- [44] M. PÉREZ, A. PÉREZ Y E. PÉREZ. 2017. INTRODUCCIÓN A LOS SISTEMAS DE CONTROL Y MODELO MATEMÁTICO PARA SISTEMAS LINEALES INVARIANTES EN EL TIEMPO. [En línea] <<http://dea.unsj.edu.ar/control1/apuntes/unidad1y2.pdf>> [Consulta: 30 enero 2020]
- [45] WORKANA. 2019. Freelancers – Programación y Tecnología [En línea]. <<https://www.workana.com/es/freelancers/india?category=it-programming>> [Consulta: 7 marzo 2020]
- [46] SALESFORCE. ¿Qué es Cloud Computing?. [En línea] <<https://www.salesforce.com/mx/cloud-computing/>> [Consulta: 4 febrero 2020]
- [47] GOOGLE CLOUD. Precios de Cloud Storage. [En línea] <<https://cloud.google.com/storage/pricing?hl=es>> [Consulta: 4 febrero 2020]
- [48] DAMODARAN, A. 2019. Costs of Capital by Industry Sector. [En línea] <[http://pages.stern.nyu.edu/~adamodar/New\\_Home\\_Page/datacurrent.html](http://pages.stern.nyu.edu/~adamodar/New_Home_Page/datacurrent.html)> [Consulta: 8 febrero 2020]

[49] SWALIN, A. 2018. CatBoost vs. Light GBM vs. XGBoost. [En línea] <<https://towardsdatascience.com/catboost-vs-light-gbm-vs-xgboost-5f93620723db>> [Consulta: 9 febrero 2020]

## 14.ANEXO

### 4. Estándares de Cumplimiento por Volumen:

- a) El volumen de verificaciones realizadas con resultado positivo, debe ser mayor o igual al 75%. En caso de que se obtenga un resultado de verificaciones positivas menor, el Contratista tiene la obligación de reportar a EQUIFAX el motivo de esa situación y coordinar un plan de trabajo para mejorar los resultados.

$$\text{Nivel de Verificaciones Positivas} = \frac{\text{Verificaciones Positivas}}{\text{Total de Verificaciones Realizadas}} \geq 75\%$$

- b) El volumen de verificaciones rechazadas no puede superar el 10% del volumen total de verificaciones.

Entiéndase por verificaciones positivas, aquellas que realizan los shoppers y son aprobadas por Rocketpin, para luego pasar el filtro de EQUIFAX.

### 5. Estándares del Servicio

El Contratista, deberá asegurar el cumplimiento de los siguientes estándares del Servicio para evitar las multas y optar a los incentivos que se detallan en la cláusula siguiente.

- 95% en la medición general agregada sobre el total de las verificaciones realizadas.
- 95% del total de cada tipo de verificación realizada (Diurna, Diurna Laboral, Vespertina, Premium y Premium Express).

### SEGUNDO: Multas e incentivos:

#### 1. Multas.

Equifax aplicará multas al Contratista por incumplimiento de los estándares de Servicio, especificados en la cláusula precedente, utilizando los siguientes criterios:

- a) Para el servicio se aplicarán las siguientes multas:
- a. Multas por cada verificación informada en sistema atrasada.
- i. Si una verificación de Región Metropolitana tiene un retraso:
1.  $\geq 2$  días de atraso y  $\leq 4$  días de atraso, no se pagará el valor de la verificación
  2.  $> 5$  días de atraso, el valor de la multa será 2 veces su precio según el que corresponda en la tabla contenida en la cláusula siguiente.
- ii. Si una verificación de Regiones tiene un retraso:
1.  $\geq 3$  días de atraso y  $\leq 5$  días de atraso, no se pagará el valor de la verificación.
  2.  $> 6$  días de atraso, el valor de la multa será 2 veces su precio, según el que corresponda en la tabla contenida en la cláusula siguiente.
- b. Multas por incumplimiento del servicio:
- i. Si el cumplimiento general del servicio es inferior al 95%, se descontará el 5% del valor neto de la facturación del servicio.

#### 2. Incentivos.

Equifax pagará un incentivo monetario al Contratista si el desempeño del Servicio excede los estándares que exige el presente anexo aplicando los siguientes criterios:

- a) Si el cumplimiento general del Servicio es mayor al 95%, se entregará una bonificación sobre la facturación neta de 0,5% por cada punto porcentual de cumplimiento sobre el 95%.

Ejemplo:

- Si el cumplimiento del Servicio General=96%  $\Rightarrow$  (Facturación Neta $\times$ 1,005).
- Si el cumplimiento del Servicio General=97%  $\Rightarrow$  (Facturación Neta $\times$ 1,010).
- Si el cumplimiento del Servicio General=98%  $\Rightarrow$  (Facturación Neta $\times$ 1,015).
- Si el cumplimiento del Servicio General=99%  $\Rightarrow$  (Facturación Neta $\times$ 1,020).
- Si el cumplimiento del Servicio General=100%  $\Rightarrow$  (Facturación Neta $\times$ 1,025).

Ilustración 68: Marco del contrato de prestación de servicios entre EQUIFAX y Rocketpin.

Fuente: Elaboración Propia.

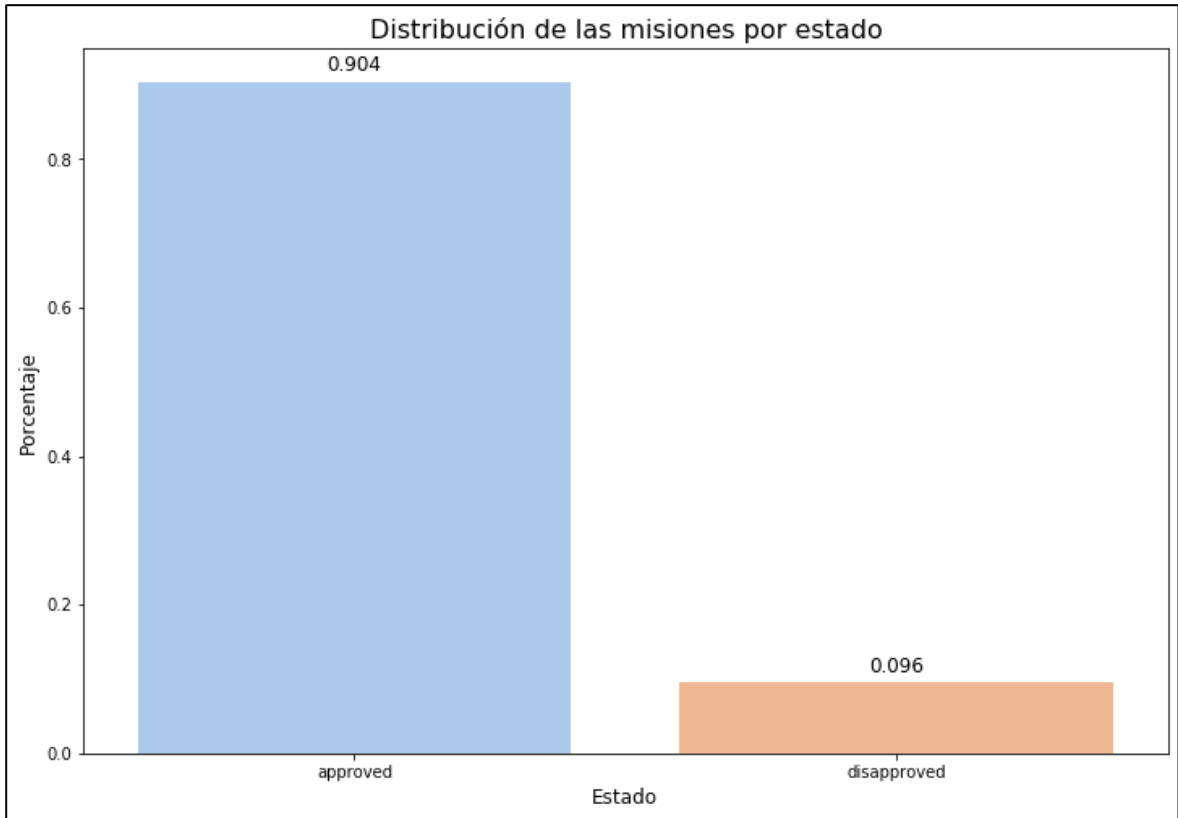


Ilustración 69: Proporción de las misiones realizadas por estado en el año 2019.  
Fuente: Elaboración Propia.

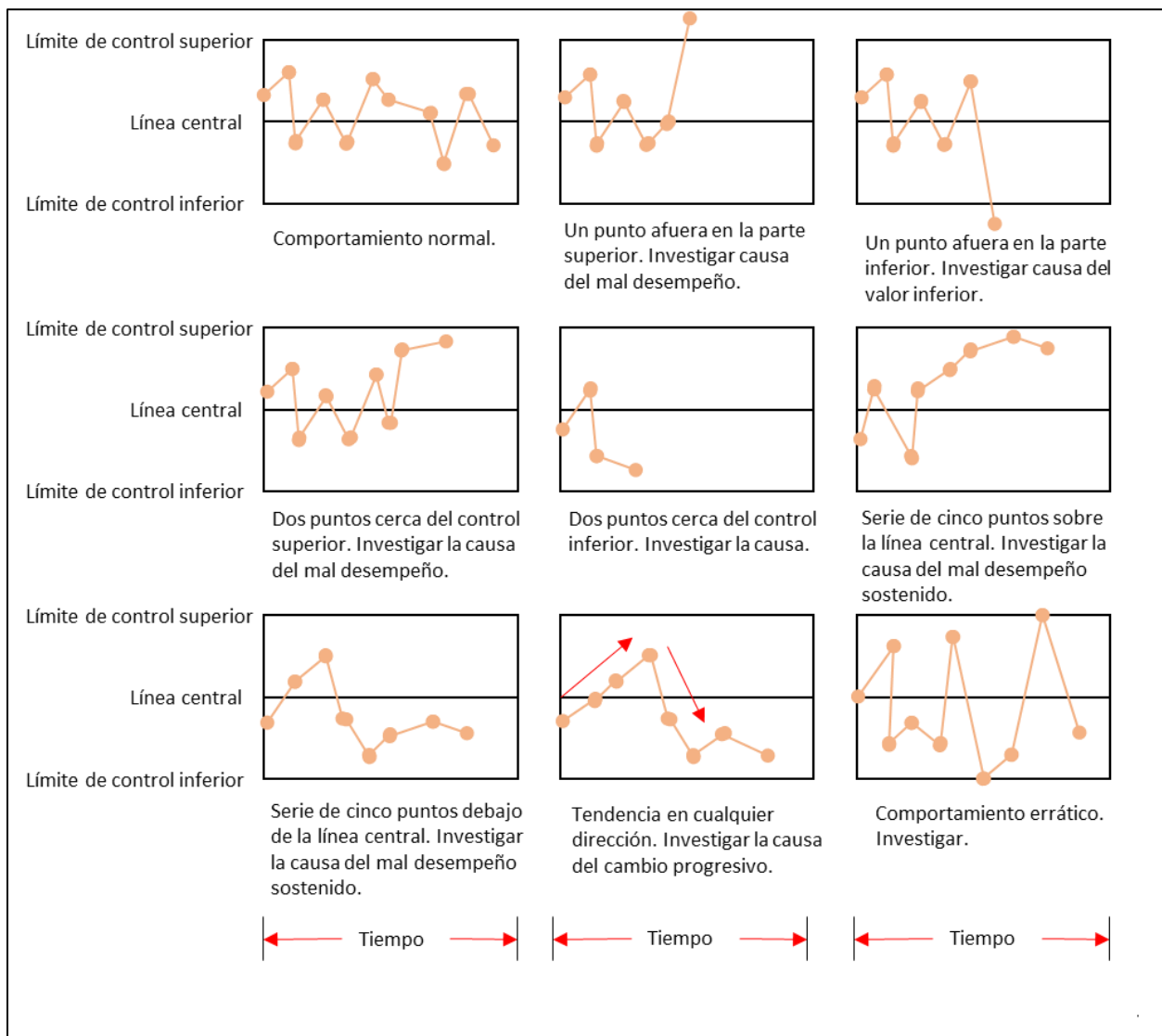


Ilustración 70: Gráficos de control de evidencias para una investigación.  
Fuente: *Administración de Operaciones (CHASE, R., JACOBS, R. y AQUILANO, N. 2009).*

Número de observaciones en el subgrupo	Factor para la gráfica $\bar{X}$	Factores para la gráfica R	
		Límite de control inferior	Límite de control superior
$n$	$A_2$	$D_3$	$D_4$
2	1,88	0	3,27
3	1,02	0	2,57
4	0,73	0	2,28
5	0,58	0	2,11
6	0,48	0	2
7	0,42	0,08	1,92
8	0,37	0,14	1,86
9	0,34	0,18	1,82
10	0,31	0,22	1,78
11	0,29	0,26	1,74
12	0,27	0,28	1,72
13	0,25	0,31	1,69
14	0,24	0,33	1,67
15	0,22	0,35	1,65
16	0,21	0,36	1,64
17	0,2	0,38	1,62
18	0,19	0,39	1,61
19	0,19	0,4	1,6
20	0,18	0,41	1,59

Ilustración 71: Factores para determinar los límites de control de las gráficas X y R.  
Fuente: Administración de Operaciones (CHASE, R., JACOBS, R. y AQUILANO, N. 2009).

Detalle	Costos del mes actual	Porcentaje del total
<b>Costos de Prevención</b>		
Capacitación de calidad	\$ 2.000	1,3%
Asesoría en confiabilidad	\$ 10.000	6,5%
Corridas de producción piloto	\$ 5.000	3,3%
Desarrollo de sistemas	\$ 8.000	5,2%
<b>Total de prevención</b>	<b>\$ 25.000</b>	<b>16,3%</b>
<b>Costos de Evaluación</b>		
Inspección de materiales	\$ 6.000	3,9%
Inspección de suministros	\$ 3.000	2,0%
Pruebas de confiabilidad	\$ 5.000	3,3%
Pruebas de laboratorio	\$ 25.000	16,3%
<b>Total de evaluación</b>	<b>\$ 39.000</b>	<b>25,5%</b>
<b>Costos de falla interna</b>		
Desperdicio	\$ 15.000	9,8%
Reparación	\$ 18.000	11,8%
Retrabajo	\$ 12.000	7,8%
Tiempo muerto	\$ 6.000	3,9%
<b>Total de falla interna</b>	<b>\$ 51.000</b>	<b>33,3%</b>
<b>Costos de falla externa</b>		
Costos de Garantía	\$ 14.000	9,2%
Reparaciones y reemplazo fuera de garantía	\$ 6.000	3,9%
Quejas del cliente	\$ 3.000	2,0%
Responsabilidad del producto	\$ 10.000	6,5%
Pérdidas de transporte	\$ 5.000	3,3%
<b>Total falla externa</b>	<b>\$ 38.000</b>	<b>24,8%</b>
<b>Total costos de calidad</b>	<b>\$ 153.000</b>	<b>100,0%</b>

Ilustración 72: Ejemplo de un informe de costos de calidad por categoría.  
Fuente: Administración de Operaciones (CHASE, R., JACOBS, R. y AQUILANO, N. 2009).



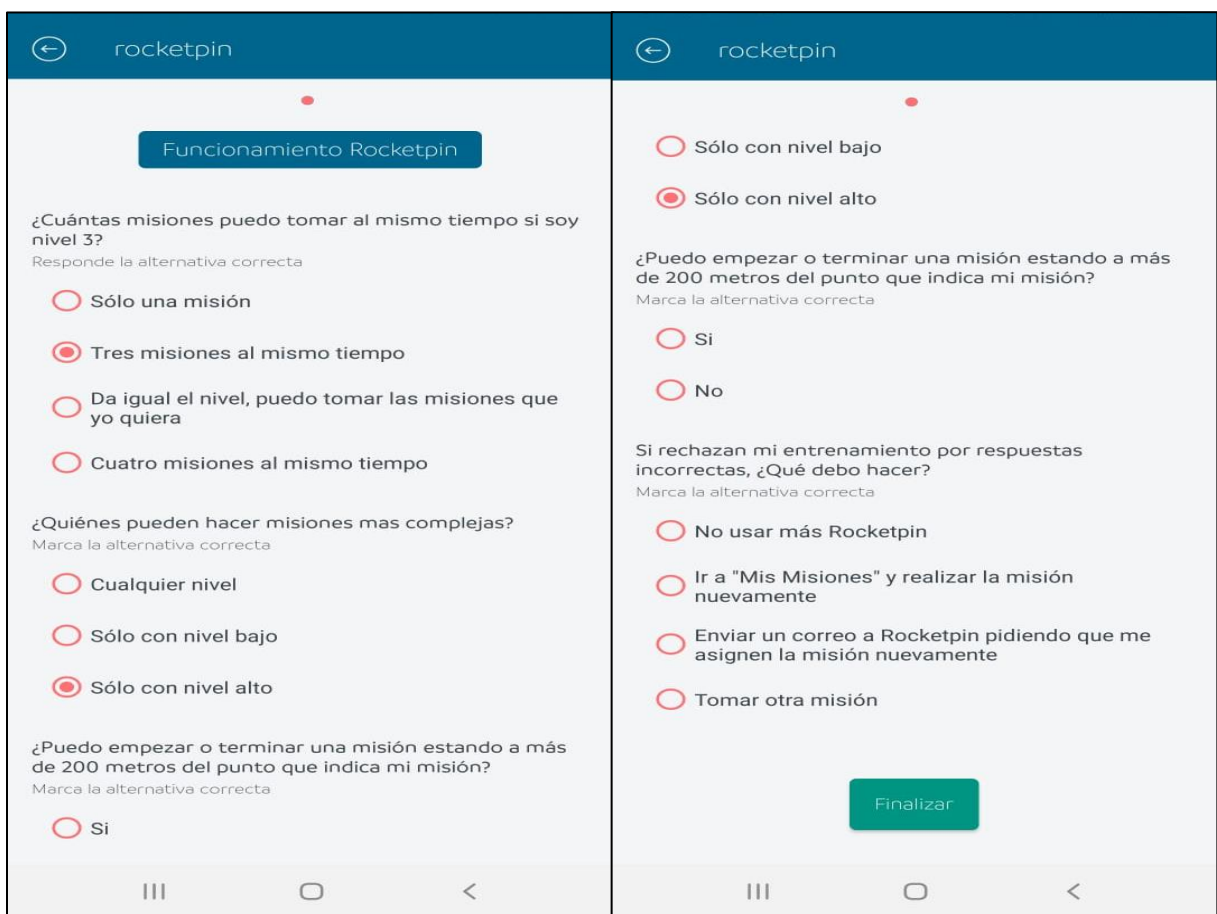


Ilustración 73: Preguntas realizadas en la misión de Entrenamiento 1.  
Fuente: Aplicación móvil de Rocketpin.

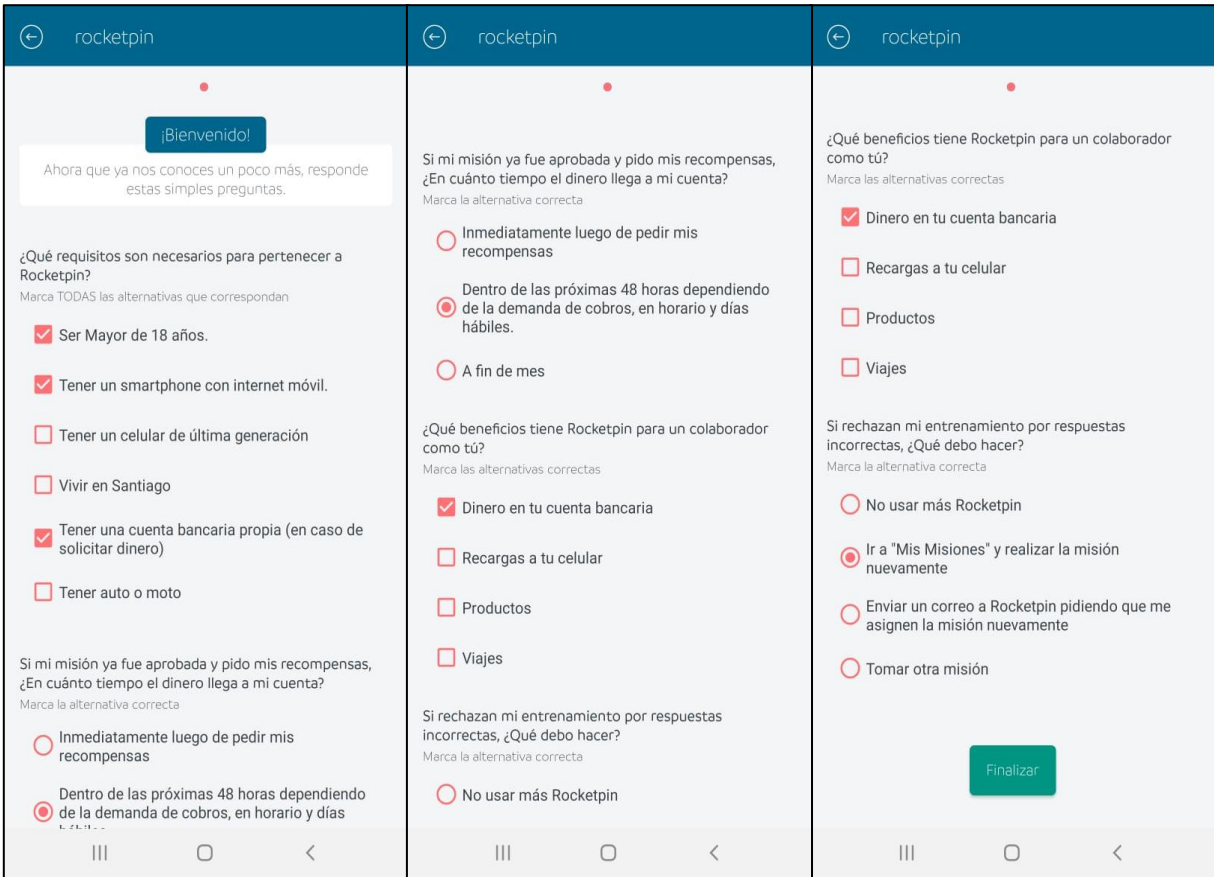


Ilustración 74: Preguntas realizadas en la misión de Entrenamiento 2.  
Fuente: Aplicación móvil de Rocketpin.

rocketpin.com/admin/missions#/show\_mission/5e2616a0ffbbe9212cf42c6b

Inicio Admin Camp Play Clientes Formularios 2.0 Misiones Orders Sucursales Usuarios Cuenta Dashboard Salir

# Missions

**Detalles de mision: Verificación domicilio EQUIFAX** Editar detalles Clonar

ESTADO	completed	PAGO	1000
PAGO	1000		
DURACIÓN	2m 17s	SHOPPER	Mario Henriquez
SUCURSAL	[HIT20] CALLE 7 Y MEDIO ORIENTE A 401, TALCA (3878088)	DIRECCIÓN	CALLE 7 Y MEDIO ORIENTE A 401 BLOCK 2940 DEPTO 401, TALCA NI LAS AMERICAS 11 0212 PASAJE 18 1 2 NORTE
VERIFICACIÓN DOMICILIO	SI	RESULTADO VERIFICACIÓN	

**Acciones**

Esta mision debe ser aprobada o rechazada para continuar con el workflow

Aprobar Rechazar

**Evaluación de mision** (Mario Henriquez)

☆☆☆☆

**Inicio** 21/01/2020, 3:09 pm / **Fin:** 21/01/2020, 3:12 pm

**Datos Cliente a Verificar**

NOMBRE	JULIA SALAS OLIVARES	EMAIL	
TELÉFONO		SUB INFO	
SUB EMPRESA	Administradora Plaza (Hites)	LOGO	rocketpin

**Archivos Adjuntos**

Audio 0 0:00

Descargar 0

**Respuestas**

Verificación de domicilio


Si tienes consultas o preguntas sobre la verificación, comunícate al +56957137554

**Existencia de domicilio**

Sí  
 No  
*La dirección solicitada corresponde a un domicilio*


---

**Foto fachada del domicilio \***

Ningún archivo seleccionado  Si el domicilio es un departamento, toma la fotografía a la fachada del edificio. Si es una casa, toma una buena fotografía donde ésta se vea claramente

---

**Foto Numeración Domicilio \***

Ningún archivo seleccionado  Toma una fotografía a la numeración del domicilio, si el domicilio es un departamento, tómale una foto a la numeración del edificio

---

**Tipo de domicilio**

Casa  
 Departamento  
*Indica a qué tipo de edificación corresponde el domicilio*

---

**¿Había alguien para contestar?**

Sí  
 No  
*Indica si habla o no una persona en el domicilio que pudiera contestar tus consultas (Incluye al conserje en edificios y condominios)*

---

Es importante obtener el nombre y apellido o al menos el nombre de quien responde ya que de no ser así, la verificación quedara como no válida para el cliente.

---

**Observación adicional**

Toqué timbre y llamé tres veces pero nadie salió

Observación: Al no encontrarse nadie en el domicilio se procede a verificación SMS vía llamado telefónico y la persona a verificar [Paula Salvador] responde de manera correcta a las preguntas de seguridad para la verificación. Se procede a su aprobación.

*Indica cualquier observación adicional acerca de tu visita en esta sección*

---

●  
VERIFICACIÓN DE DOMICILIO

---

**Acciones**

El estado actual de la misión no requiere de una acción para continuar con el workflow

Ilustración 75: Ejemplo del formulario de una verificación domiciliaria.  
Fuente: Rocketpin.

Misión	Tiempo [s]	Misión	Tiempo [s]	Misión	Tiempo [s]	Misión	Tiempo [s]	Misión	Tiempo [s]
0	77,13	30	109,54	60	116,18	90	83,93	120	88,50
1	96,56	31	58,55	61	95,72	91	67,24	121	137,50
2	87,24	32	112,28	62	45,73	92	184,32	122	96,01
3	102,63	33	91,08	63	122,23	93	66,70	123	116,73
4	126,99	34	100,82	64	71,11	94	74,79	124	138,26
5	123,89	35	116,89	65	96,78	95	86,40	125	76,33
6	119,63	36	124,59	66	68,60	96	49,81	126	130,60
7	125,90	37	96,00	67	85,35	97	141,85	127	72,68
8	82,44	38	72,91	68	113,94	98	109,27	128	118,17
9	77,14	39	114,00	69	67,42	99	31,10	129	88,39
10	69,94	40	83,88	70	106,08	100	103,83	130	79,82
11	170,51	41	97,29	71	103,03	101	72,57	131	78,37
12	121,85	42	102,24	72	97,71	102	121,06	132	121,21
13	92,91	43	127,97	73	91,73	103	98,69	133	88,75
14	61,02	44	101,35	74	105,53	104	98,35	134	74,45
15	144,46	45	111,34	75	101,19	105	79,08	135	101,38
16	69,28	46	133,68	76	36,77	106	147,62	136	118,58
17	100,47	47	49,65	77	80,05	107	120,84	137	87,07
18	44,89	48	100,45	78	126,75	108	100,24	138	92,60
19	93,77	49	74,00	79	137,11	109	149,29	139	83,61
20	109,49	50	114,63	80	101,21	110	118,16	140	106,90
21	77,44	51	127,99	81	109,24	111	110,50	141	130,28
22	107,25	52	125,94	82	99,96	112	174,56	142	68,11
23	129,75	53	120,99	83	140,12	113	132,39	143	82,46
24	124,18	54	91,52	84	93,41	114	110,21	144	101,93
25	84,30	55	137,18	85	108,54	115	132,72	145	107,72
26	91,19	56	89,74	86	112,10	116	109,64	146	58,51
27	104,31	57	111,27	87	105,18	117	81,68	147	110,44
28	84,06	58	75,79	88	84,82	118	114,41	148	60,01
29	101,48	59	110,61	89	104,29	119	113,00	149	125,64

Ilustración 76: Medición de los tiempos de revisión para 150 misiones.  
Fuente: Elaboración Propia.

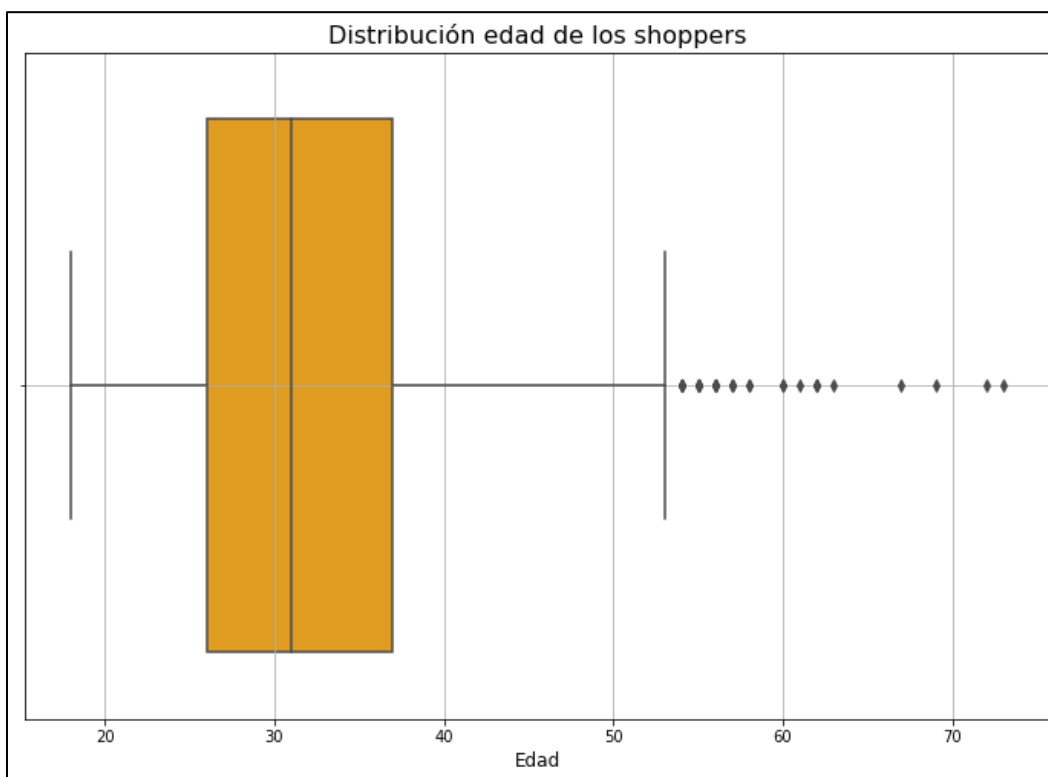


Ilustración 77: Distribución de la edad de los shoppers mediante un *boxplot*.  
Fuente: Elaboración Propia.

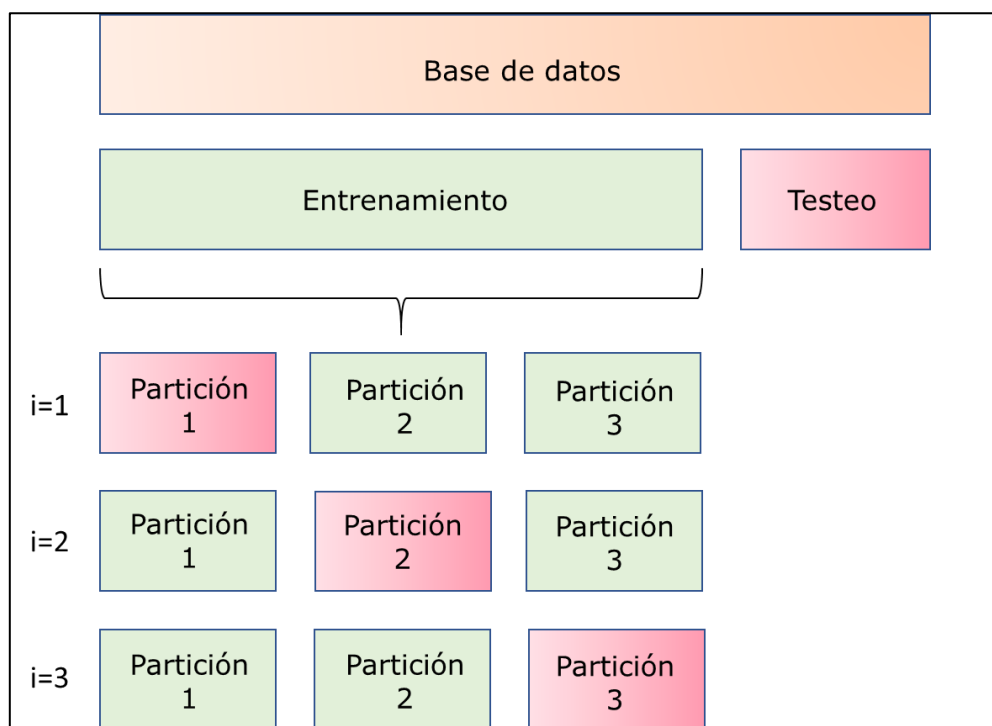


Ilustración 78: Ejemplo de una *Cross Validation*.  
Fuente: Elaboración Propia.