



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

IDENTIFICACIÓN DE CLIENTES CON ALTA PROBABILIDAD DE LLAMAR AL  
CONTACT CENTER DE UNA AEROLÍNEA

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

EMILIO ALONSO VARAS ROMÁN

PROFESOR GUÍA:  
ALEJANDRA PUENTE CHANDÍA

MIEMBROS DE LA COMISIÓN  
PABLO MARÍN VICUÑA  
MARCEL GOIC FIGUEROA

SANTIAGO DE CHILE  
2016

## IDENTIFICACIÓN DE CLIENTES CON ALTA PROBABILIDAD DE LLAMAR AL CONTACT CENTER DE UNA AEROLÍNEA

El programa de fidelización LANPASS, tiene como uno de sus principales servicios de consultas el contact center, el cual ha incurrido en costos más allá de lo presupuestado en el último tiempo. Por esto, la compañía se ha visto en la necesidad de estudiar a los clientes que utilizan este servicio, de modo de poder reducir tales costos.

El presente trabajo tiene como objetivo identificar cuáles son los clientes más propensos a llamar al contact center por motivo de “Check-in Upgrade”, el cual representa el 28% de los llamados de la atención preferente. La elección de unir estos dos motivos “Check-in” y “Upgrade” se basa en dos razones. Primero por la baja participación del motivo upgrade (7%) y segundo debido a la experiencia del negocio, la cual dice que ambos procesos se realizan en la misma llamada. Para el desarrollo del trabajo, se utilizó la información disponible de las llamadas del año 2014, de los socios de categoría Premium Silver, Comodoro y Black, que representan el 24% éstas.

Se realizaron dos modelos de propensión, regresión logística y árbol de decisión, para poder identificar cuáles eran las variables que aumentaban la probabilidad de llamada al contact center. En términos de la curva de ganancia, el modelo de árbol de decisión obtuvo los mejores resultados capturando el 30% de los casos en el primer decil, en comparación al 20% del modelo de regresión. Debido a su simpleza en la interpretación de los resultados y a distintas métricas de desempeño, se escogió el modelo de árboles. De éste último, se obtuvo que las principales variables que aumentan la probabilidad son el canje del pasaje, si el vuelo es internacional o doméstico, si el viaje presenta escalas, la nacionalidad y categoría del socio. Luego se escogieron los 4 perfiles más relevantes en base a éstas variables. A éstos se les diseñó una acción ya sea, un plan comunicacional, un diagnóstico o un proyecto, para la modificación de los procesos regulares.

De forma complementaria, se realizó una evaluación económica, la cual mostró que en un escenario conservador (30% de respuestas positivas), se generaría un ahorro del 20% de los costos generados por dichos perfiles, correspondiente a más de 15 millones de pesos anuales. Además, con estas acciones se estaría reduciendo aproximadamente el 50% del exceso de presupuesto producido por la atención preferente.

Finalmente, como lineamiento de trabajo futuro se espera la realización de los experimentos de cada una de las acciones diseñadas, la incorporación de nuevas fuentes de información, tales como la duración de las llamadas, que se cree que serán un gran aporte para realizar mejores estimaciones de los beneficios a obtener.

## **AGRADECIMIENTOS**

Habiendo finalizado mi etapa universitaria, me gustaría agradecer a algunas personas que hicieron que ésta fuera una inolvidable y grandiosa experiencia.

Primero que todo, agradezco profundamente a mi familia, mi madre, mi hermano y a mi padrastro, por el amor que siempre me han entregado, por su apoyo incondicional y por su fe intacta de que podría llegar a este día de la mejor manera. Por acompañarme en todo momento, celebrando los buenos y aconsejándome en los más difíciles. Espero que se sientan muy orgullosos de mí, porque yo estoy muy orgulloso de todos ustedes.

Quiero agradecer a mis profesores Alejandra Puente y Pablo Marín, y también a Juan Pablo Valencia, quienes me brindaron toda la ayuda posible en este trabajo. Muchas gracias por estar siempre disponibles ante cualquier duda o problema que pudiese surgir. Aprendí mucho de ustedes y eso lo agradeceré y valoraré por siempre.

No cabe duda que por mi paso por la universidad rescaté muchas cosas, pero lo más importante son las personas que conocí en este camino. Primero, quería mencionar a Alejandro Encina, que si bien no estudiamos en la misma institución, lo conozco hace mucho tiempo y fue fundamental en toda esta etapa, me hace demasiado feliz el saber que siempre estarás ahí, te quiero hermano mío. También agradecer a Felipe Sánchez, amigo mío un gustazo haberte conocido desde primer año, si bien no seguimos las mismas especialidades, siempre estuviste ahí, en los carretes, en las dificultades, en la selección de fútbol, en todo, mi 10!, te quiero un montón! A la bandida de industrias: Carlos, Imanol, Chanta y Sebastián, como no mencionarlos, si estuvimos juntos estos últimos 4 años, cuantas bienvenidas, anécdotas, cuantos momentos lindos vivimos y nos quedan por vivir, muchas gracias muchachos por hacer de las tareas, los trabajos y las tardes de estudios las más entretenidas! Espero sigamos siempre en contacto. Como no mencionar a todas las personas de la Feria Empresarial, que fueron fundamental en mí proceso por la universidad. Conocí personas bellísimas, sequísimas y que me hicieron crecer y formar la persona que soy hoy en día. Sepan que a cada uno de los que estuvo aquellos 3 años los llevaré siempre conmigo. No puedo dejar de mencionar a ustedes Matías, Pelao y Ricardo, grandes amigos que siempre de distintas formas se han hecho presente, gracias por la paciencia, apoyo y cariño. Por último, a todos los chiquillos de puro bacilon por todos los lindos recuerdos en nuestros tranquilos carretes.

Gracias a todos los chiquillos de LANPASS, en especial al área de operaciones, por haberme acogido y hacerme sentir a gusto. Les deseo lo mejor a cada uno de ustedes.

Gracias a cada uno de los que estuvo presente de una u otra forma todos estos años.

Finalmente, quería agradecer de manera muy especial a mi padre, que desde arriba siempre se ha hecho presente y me ha dado las fuerzas necesarias para seguir adelante.

# TABLA DE CONTENIDO

<b>TABLA DE CONTENIDO .....</b>	<b>III</b>
<b>INDICE DE TABLAS .....</b>	<b>V</b>
<b>INDICE DE GRÁFICOS .....</b>	<b>VI</b>
<b>1 INTRODUCCIÓN .....</b>	<b>1</b>
<b>2 JUSTIFICACIÓN.....</b>	<b>2</b>
<b>3 OBJETIVOS.....</b>	<b>3</b>
3.1 OBJETIVO GENERAL.....	3
3.2 OBJETIVOS ESPECÍFICOS.....	3
<b>4 ALCANCES .....</b>	<b>4</b>
<b>5 MARCO CONCEPTUAL .....</b>	<b>4</b>
5.1 PROGRAMAS DE FIDELIZACIÓN .....	4
5.2 CONTACT CENTER .....	5
<b>6 MARCO TEORICO.....</b>	<b>5</b>
6.1 MODELOS DE PROPENSIÓN .....	5
6.2 REGRESIÓN LOGÍSTICA (LOGIT) .....	6
6.3 ÁRBOLES DE DECISIÓN .....	7
6.4 MEDIDAS DE ERROR.....	8
6.4.1 Raíz del error cuadrático medio .....	8
6.4.2 Error absoluto medio.....	9
6.4.3 Máximo error absoluto porcentual .....	9
6.4.4 Medidas de evaluación de desempeño .....	9
6.5 TEST DE HIPÓTESIS.....	14
6.5.1 Test de significancia individual .....	14
6.5.2 Test de significancia global.....	15
<b>7 METODOLOGÍA Y DESARROLLO METODOLÓGICO .....</b>	<b>15</b>
7.1 OBTENCIÓN DE DATOS .....	16
7.2 ANÁLISIS EXPLORATORIO: LLAMADOS 2014 .....	17
7.3 LIMPIEZA DE DATOS .....	19
7.4 MOTIVO DE ESTUDIO: CHECK-IN UPGRADE .....	22
7.5 CARACTERIZACIÓN DE CLIENTES E IDENTIFICACIÓN DE VARIABLES RELEVANTES .....	24
7.5.1 Análisis exploratorio: “Check-in Upgrade”.....	25
7.6 DESARROLLO MODELOS DE PROPENSIÓN .....	33
7.6.1 Modelo Logit .....	33
7.6.2 Modelo Logit 1.....	36
7.6.3 Modelo Logit 2 .....	44
7.6.4 Comparación modelo balanceado y no balanceado .....	46
7.7 MODELO ÁRBOL DE DECISIÓN.....	47
7.8 COMPARACIÓN DE MODELOS.....	57
7.9 DISEÑO DE ACCIONES .....	59
7.9.1 Elección perfiles de clientes.....	60
7.9.2 Acciones .....	61
7.9.3 Evaluación económica.....	63
<b>8 CONCLUSIONES .....</b>	<b>68</b>
8.1 CONCLUSIONES DEL TRABAJO REALIZADO.....	68
8.2 LIMITACIONES DEL TRABAJO .....	69

8.3	RECOMENDACIONES Y TRABAJOS FUTUROS .....	69
<b>9</b>	<b>BIBLIOGRAFÍA .....</b>	<b>71</b>
<b>10</b>	<b>ANEXOS Y APENDICES .....</b>	<b>73</b>
10.1	ANEXO 1: TRÁFICO DE PASAJEROS NACIONAL .....	73
10.2	ANEXO 2: TRÁFICO DE PASAJEROS INTERNACIONAL .....	73
10.3	ANEXO 3: TABLA DE PRECISIÓN SEGÚN AUC .....	74
10.4	ANEXO 4: VARIABLES HORARIO Y DÍA .....	74
10.5	ANEXO 5: EXTRACTO DE UN FLUJO DE TRANSFORMACIÓN DE LA HERRAMIENTA PENTAHO .....	77
10.6	ANEXO 6: DESARROLLO TEST DE PROPORCIONES DE LA VARIABLE SALDO .....	77
10.7	ANEXO 7: RESULTADO MODELO 2 LOGIT .....	78
10.8	ANEXO 8: MATRIZ DE CONFUSIÓN MODELO 2 LOGIT PARTICIÓN TESTEO .....	79
10.9	ANEXO 9: MÉTRICAS DE DESEMPEÑO MODELO 2 LOGIT .....	79
10.10	ANEXO 10: TABLA DE GANANCIA DE INFORMACIÓN MODELO 2 LOGIT PARTICIÓN TESTEO .....	79
10.11	ANEXO 11: RESULTADOS MODELO 1 LOGIT NO BALANCEADO .....	80
10.12	ANEXO 12: RESULTADOS MODELO 2 LOGIT NO BALANCEADO .....	82
10.13	ANEXO 13: ANÁLISIS DEL ÁRBOL DESDE EL NODO INTERNACIONAL .....	84
10.14	ANEXO 14: ANÁLISIS DEL ÁRBOL DESDE EL NODO DOMÉSTICO .....	88
10.15	ANEXO 15: RESUMEN ÁRBOL DE DECISIÓN .....	91
10.16	ANEXO 16: RESUMEN PERFILES DEL ÁRBOL DE DECISIÓN EN BASE TOTAL .....	94

## INDICE DE TABLAS

TABLA 1: MATRIZ DE CONFUSIÓN .....	9
TABLA 2: METODOLOGÍA MEMORIA Y SU RELACIÓN CON EL PROCESO KDD. ....	16
TABLA 3: TIPIFICACIÓN FINAL A TRABAJAR. ....	19
TABLA 4: PARTICIPACIÓN DE TIPIFICACIONES FINALES EN LA DEMANDA DEL CONTACT CENTER.....	20
TABLA 5: MUESTRA DE LA CANTIDAD DE LLAMADAS POR SEMANA DE SOCIOS PREFERENTES POR CHECK-IN UPGRADE. ....	22
TABLA 6: VARIABLES DE ESTUDIO.....	24
TABLA 7: HIPÓTESIS GENERADAS EN BASE AL ANÁLISIS EXPLORATORIO. ....	33
TABLA 8: VARIABLES EXPLICATIVAS PARA MODELO LOGIT.....	35
TABLA 9: RESULTADO MODELO 1 LOGIT.....	37
TABLA 10: PRECISIÓN PARA LOS DISTINTOS PUNTOS DE CORTE PARA EL MODELO 1 EN LA PARTICIÓN DE TESTEO. ...	41
TABLA 11: MATRIZ DE CONFUSIÓN MODELO 1 EN PARTICIÓN DE TESTEO CON DATOS BALANCEADOS. ....	41
TABLA 12: GANANCIA DE INFORMACIÓN MODELO 1 LOGIT .....	42
TABLA 13: INTERACCIÓN DE VARIABLES INCLUIDAS EN MODELO 2 LOGIT. ....	44
TABLA 14: DIFERENCIAS DE MÉTRICAS ENTRE MODELO 1 BALANCEADO Y NO BALANCEADO.....	47
TABLA 15: MATRIZ DE CONFUSIÓN ÁRBOL DE DECISIÓN PARA PARTICIÓN DE VALIDACIÓN.....	54
TABLA 16: MÉTRICAS DE EVALUACIÓN DE DESEMPEÑO DE LOS DISTINTOS MODELOS.....	58
TABLA 17: PERFILES SELECCIONADOS PARA TOMAR ACCIONES. ....	60
TABLA 18: DESCRIPCIÓN CUANTITATIVA DE LOS PERFILES ESCOGIDOS. ....	61
TABLA 19: ACCIÓN PARA CADA PERFIL DE SOCIO. ....	62
TABLA 20: DEFINICIÓN DEL PORCENTAJE DE RESPUESTA POSITIVA PARA CADA UNO DE LOS TRES ESCENARIOS.....	63
TABLA 21: DISPONIBILIDAD DE CONTACTO Y OPEN RATE DE LOS DISTINTOS TIPOS DE COMUNICACIÓN.....	64
TABLA 22: DESCRIPCIÓN DE COSTOS FIJOS ASOCIADOS A CADA ACCIÓN. ....	64
TABLA 23: DESCRIPCIÓN DE COSTOS VARIABLES ASOCIADOS A CADA ACCIÓN. ....	65
TABLA 24: COSTOS TOTALES POR ACCIÓN PARA LOS DISTINTOS PERFILES SEGÚN EL NÚMERO DE SOCIOS AFECTADOS. .....	65
TABLA 25: AHORRO GENERADO POR LAS DIFERENTES ACCIONES PARAR LOS PERFILES CORRESPONDIENTES.....	66
TABLA 26: TIEMPO PARA QUE LA ACCIÓN OBTENGA AHORROS PARA LA EMPRESA EN MESES. ....	66
TABLA 27: BREAK-EVEN PARA LAS DISTINTAS ACCIONES.....	67

## INDICE DE GRÁFICOS

GRÁFICO 1: CURVA ROC .....	13
GRÁFICO 2: LLAMADOS TIPIFICADOS 2014. ....	17
GRÁFICO 3: DISTRIBUCIÓN TIPIFICACIONES POR LÍNEA DE COMUNICACIÓN.....	18
GRÁFICO 4: DISTRIBUCIÓN TIPIFICACIONES ATENCIÓN GENERAL VS PREFERENTE.....	18
GRÁFICO 5: DISTRIBUCIÓN TIPIFICACIONES FINALES ATENCIÓN PREFERENTE. ....	20
GRÁFICO 6: DISTRIBUCIÓN DE LA CANTIDAD DE LLAMADAS A TRAVÉS DE LAS SEMANAS DEL 2014 EN LA ATENCIÓN PREFERENTE. ....	21
GRÁFICO 7: DISTRIBUCIÓN DE LA CANTIDAD DE LLAMADAS POR MOTIVO A TRAVÉS DE LAS SEMANAS DEL 2014. ....	21
GRÁFICO 8: ASA DE LLAMADA VS CANTIDAD DE SOCIOS EN ANTICIPACIÓN DE COMPRA EN DÍAS. ....	25
GRÁFICO 9.- TASA DE LLAMADA VS CANTIDAD DE SOCIOS EN EDAD DE LOS SOCIOS. ....	26
GRÁFICO 10.- TASA DE LLAMADA VS CANTIDAD DE SOCIOS RESPECTO AL SALDO ACUMULADO (MILES DE KMS) ....	26
GRÁFICO 11.- TASA DE LLAMADA VS CANTIDAD DE SOCIOS SEGÚN EL TIPO DE VUELO.....	27
GRÁFICO 12.- TASA DE LLAMADA VS CANTIDAD DE SOCIOS EN VUELOS CON O SIN ESCALAS. ....	27
GRÁFICO 13.- TASA DE LLAMADA VS CANTIDAD DE SOCIOS RESPECTO A LA ANTIGÜEDAD EN EL PROGRAMA DEL CLIENTE. ....	28
GRÁFICO 14.- TASA DE LLAMADA VS CANTIDAD DE SOCIOS EN LA COMPRA DE PASAJE VÍA INTERNET. ....	28
GRÁFICO 15.- TASA DE LLAMADA VS CANTIDAD DE SOCIOS SEGÚN LA COMPRA DE PASAJE VÍA CALL CENTER.....	29
GRÁFICO 16.- TASA DE LLAMADA VS CANTIDAD DE SOCIOS RESPECTOS LOS CLIENTES QUE TENÍAN OPCIÓN DE POSTULAR A UPGRADE. ....	29
GRÁFICO 17.- TASA DE LLAMADA VS CANTIDAD DE SOCIOS SEGÚN EL GÉNERO. ....	30
GRÁFICO 18.- TASA DE LLAMADA VS CANTIDAD DE SOCIOS RESPECTO DE LOS CLIENTES QUE CANJEARON SU PASAJE COMO FORMA DE PAGO. ....	30
GRÁFICO 19.- TASA DE LLAMADA VS CANTIDAD DE SOCIOS, PARA LA NACIONALIDAD DE LOS SOCIOS.....	31
GRÁFICO 20.- TASA DE LLAMADA VS CANTIDAD DE SOCIOS, RESPECTO A LOS INTENTOS DE CHEQUEO EN LA WEB...31	31
GRÁFICO 21.- TASA DE LLAMADA VS CANTIDAD DE SOCIOS SEGÚN LA CATEGORÍA DE LOS SOCIOS.....	32
GRÁFICO 22: CURVA ROC DE LA REGRESIÓN LOGÍSTICA PARA EL MODELO DE LA TABLA 6 EN LA PARTICIÓN DE CALIBRACIÓN. ....	39
GRÁFICO 23: CURVA ROC PARA REGRESIÓN LOGÍSTICA PARA LA PARTICIÓN DE TESTEO. ....	40
GRÁFICO 24: CURVA DE GANANCIA MODELO 1 LOGIT BALANCEADO PARA PARTICIÓN DE CALIBRACIÓN (IZQUIERDA) Y TESTEO (DERECHA). ....	43
GRÁFICO 25: CURVA DE ELEVACIÓN PARA MODELO 1 LOGIT BALANCEADO EN PARTICIÓN DE CALIBRACIÓN (IZQUIERDA) Y TESTEO (DERECHA). ....	43
GRÁFICO 26: CURVA ROC MODELO 2 LOGIT BALANCEADO PARA PARTICIÓN DE CALIBRACIÓN (IZQUIERDA) Y TESTEO (DERECHA). ....	45
GRÁFICO 27: CURVA DE GANANCIA MODELO 2 LOGIT BALANCEADO EN PARTICIÓN DE CALIBRACIÓN (IZQUIERDA) Y TESTEO (DERECHA). ....	45
GRÁFICO 28: CURVA DE ELEVACIÓN PARA MODELO 2 LOGIT BALANCEADO EN PARTICIÓN DE CALIBRACIÓN (IZQUIERDA) Y TESTEO (DERECHA). ....	46
GRÁFICO 29: IMPORTANCIA DE VARIABLES PARA EL MODELO DE ÁRBOLES DE DECISIÓN. ....	49
GRÁFICO 30: CURVA ROC MODELO ÁRBOL DE DECISIÓN PARA LAS DISTINTAS PARTICIONES DE LA BASE DE DATOS.56	56
GRÁFICO 31: GANANCIA DE INFORMACIÓN PARA MODELO ÁRBOL DE DECISIÓN PARA LAS DISTINTAS PARTICIONES DE LA BASE DE DATOS.....	56
GRÁFICO 32: ELEVACIÓN PARA MODELO DE ÁRBOL DE DECISIÓN PARA LAS DISTINTAS PARTICIONES DE LA BASE DE DATOS.....	57
GRÁFICO 33: DISTRIBUCIÓN DE LLAMADOS POR BLOQUE HORARIO EN 2014.....	74
GRÁFICO 34: DISTRIBUCIÓN DE LAS TIPIFICACIONES 2014 POR BLOQUE HORARIO. ....	75
GRÁFICO 35: DISTRIBUCIÓN DE LLAMADOS POR TIPO DE DÍA EN 2014. ....	76
GRÁFICO 36: DISTRIBUCIÓN TIPIFICACIONES POR TIPO DE DÍA EN 2014. ....	76

## INDICE DE ILUSTRACIONES

ILUSTRACIÓN 1: PROCESO KDD.....	15
ILUSTRACIÓN 2: PROCESAMIENTO DE BASE DE DATOS PARA TRABAJAR EL MODELO LOGIT. ....	36
ILUSTRACIÓN 3: PROCESAMIENTO DE BASE DE DATOS PARA TRABAJAR EL MODELO LOGIT. ....	48
ILUSTRACIÓN 4: NODO 0 MODELO ÁRBOL CHAID: VARIABLE OBJETIVO.....	49
ILUSTRACIÓN 5: PRIMER NIVEL MODELO ÁRBOL CHAID EXHAUSTIVO: ABERTURA POR TIPO DE VUELO. ....	50
ILUSTRACIÓN 6: NIVEL 2 MODELO ÁRBOL CHAID EXHAUSTIVO: ABERTURA POR POSIBILIDAD DE POSTULAR A UPGRADE.....	50
ILUSTRACIÓN 7: TERCER NIVEL ÁRBOL: APERTURA POR SI VIAJE PRESENTA ESCALAS O NO. ....	51
ILUSTRACIÓN 8 RAMIFICACIÓN DEL ÁRBOL A PARTIR DEL NODO 5: VUELO INTERNACIONAL SIN OPCIONES DE POSTULAR A UPGRADE. ....	52
ILUSTRACIÓN 9: RAMIFICACIÓN SEGUNDO NIVEL PARA VARIABLE: VUELO DOMÉSTICO. ....	52
ILUSTRACIÓN 10: RAMIFICACIÓN DEL NODO 7 SOCIO NO INTENTA CHEQUEARSE VÍA WEB (NODO PADRE: SOCIO NO ES COMODORO). ....	53
ILUSTRACIÓN 11: RAMIFICACIÓN DEL NODO 8 SOCIO INTENTA CHEQUEARSE VÍA WEB (NODO PADRE: SOCIO NO ES COMODORO).....	53
ILUSTRACIÓN 12: SITUACIÓN ACTUAL DE COBRO DE CUPONES UPG PARA LA POSTULACIÓN A UPGRADE. ....	62
ILUSTRACIÓN 13: RAMIFICACIÓN DESDE EL NODO 18: VUELO CON ESCALAS.....	85
ILUSTRACIÓN 14: RAMIFICACIÓN DESDE EL NODO 17: VUELO SIN ESCALAS. ....	86
ILUSTRACIÓN 15: RAMIFICACIÓN DEL ÁRBOL A PARTIR DEL NODO 5: VUELO INTERNACIONAL SIN OPCIONES DE POSTULAR A UPGRADE. ....	87
ILUSTRACIÓN 16: RAMIFICACIÓN DEL NODO 7 SOCIO NO INTENTA CHEQUEARSE VÍA WEB (NODO PADRE: SOCIO NO ES COMODORO).....	88
ILUSTRACIÓN 17: RAMIFICACIÓN NODO 4: SOCIO ES DE CATEGORÍA COMODORO (NODO PADRE: VUELO DOMÉSTICO). ....	89
ILUSTRACIÓN 18: CONTINUACIÓN RAMIFICACIÓN NODO 4: SOCIO ES DE CATEGORÍA COMODORO (NODO PADRE: VUELO DOMÉSTICO). ....	90
ILUSTRACIÓN 19: CONTINUACIÓN RAMIFICACIÓN NODO 4: SOCIO ES DE CATEGORÍA COMODORO (NODO PADRE: VUELO DOMÉSTICO). ....	90

## 1 INTRODUCCIÓN

La industria aérea en Chile viene sufriendo cambios hace ya varios años. Según la Junta Aeronáutica Civil desde el año 2009 se ha presenciado un aumento en el tráfico de pasajeros (sección 10 anexo 1 y 2). Sumado a esto, está la entrada de nuevas firmas, el aumento de los costos operacionales, la baja en los precios de pasajes, etc. lo que da cabida para la continua búsqueda de mejoras por parte del programa de fidelización de clientes de LAN, denominado LANPASS.

El programa de fidelización de LAN, LANPASS, consiste en que la aerolínea premia a sus socios con “kilómetros LANPASS”, los cuales pueden ser obtenidos de diversas maneras, tales como: viajando en las líneas aéreas LAN, TAM, alianza Oneworld<sup>1</sup> o asociadas; usando la tarjeta Santander, comprando en comercios asociados y finalmente prefiriendo servicios asociados a la hora de viajar<sup>2</sup>. Además, los kilómetros acumulados pueden ser canjeados por dos tipos de beneficios: vuelos de las mismas aerolíneas mencionadas anteriormente o por productos del catálogo LANPASS<sup>3</sup>.

Dentro del programa podremos distinguir distintas categorías de clasificación para los clientes, dependiendo de la suma de kilómetros, segmentos o combinación de ambos<sup>4</sup>. Específicamente LANPASS tiene una clasificación de 5 categorías: LANPASS, Premium, Premium Silver, Comodoro y Black, siendo LANPASS la más baja y Black la más alta. Además, el contact center se divide en dos grandes equipos de atención: general y preferente, los que se diferencian en las categorías de socios que atienden, siendo el segundo quien atiende a los socios tipo premium silver, comodoro y black.

En general, en la mayoría de las empresas se busca la forma de satisfacer a los clientes, y mayor aún en un programa de fidelización que busca reafirmar esta relación. Es por esto mismo que nacen servicios como lo es el contact center, en el cual el cliente tiene la posibilidad de llamar para poder realizar diversos tipos de acciones, desde una consulta por un vuelo hasta canjear un producto del catálogo.

Ahora bien, todo servicio que se pretende brindar para la satisfacción del cliente conlleva costos, y es aquí donde en los últimos años se ha producido una dificultad para LANPASS, ya que no se ha logrado cumplir con el presupuesto que se tenía contabilizado para el contact center. Debido a esto, es donde se enmarca el trabajo a realizar en la presente memoria.

---

1 Oneworld es una alianza global que reúne a 15 de las mejores y mayores líneas aéreas del mundo, entre las cuales cuenta con: LAN, American Airlines, Cathay Pacific, Iberia, Malaysia Airlines, Qatar Airways, S7 Airlines, Airberlin, British Airways, Finnair, Japan Airlines, Qantas, Royal Jordanian y SriLankan Airlines.

2 Sobre LANPASS en: [http://www.lan.com/es\\_cl/sitio\\_personas/lanpass/sobre-lanpass/](http://www.lan.com/es_cl/sitio_personas/lanpass/sobre-lanpass/)

3 Catálogo LANPASS en: <http://www.lan.com/catalogolanpass/>

4 Como calificar LANPASS en: [http://www.lan.com/es\\_cl/sitio\\_personas/lanpass/sobre-lanpass/como-calificar/](http://www.lan.com/es_cl/sitio_personas/lanpass/sobre-lanpass/como-calificar/)

El presente trabajo se encargara de poder identificar clientes que tengan una alta probabilidad de llamar al contact center del programa, a raíz de su contexto, a través de un modelo de propensión, de tal forma de poder tomar alguna acción preventiva sobre este tipo de clientes, específicamente para los clientes que llaman por el motivo de “Check-in Upgrade”.

Finalmente, el trabajo realizado se enmarca dentro del área de operaciones del programa de fidelización de LAN, la cual es la que está a cargo de la funcionalidad del contact center del programa.

## **2 JUSTIFICACIÓN**

Los distintos beneficios que ofrece el programa sumado a las eventualidades que pueden pasar a la hora de realizar un vuelo, hacen que los socios llamen al contact center por varias razones, lo que se vio reflejado en que, en el año 2014 se realizaron un total de 848.000 llamados, y tomando en cuenta que tal servicio cobra por minuto hablado, esto se tradujo en un costo de varios millones de dólares, lo cual como se mencionó anteriormente, estuvo por sobre lo presupuestado para el año respectivo.

Es por esto que es de vital importancia conocer los principales motivos por los cuales lo socios están llamando, y más aún identificar las distintas variables que permitan caracterizar a los clientes que provocan este aumento en la demanda del contact center.

El presente proyecto surge de la necesidad de caracterizar a los clientes y poder saber cuáles son las principales diferencias entre los socios que llaman y los que no, con el fin de poder tomar acciones que eviten en algún porcentaje las respectivas llamadas. Esto último toma importancia ya que, por ejemplo si se evita sólo un 1% de las llamadas de la atención preferente significaría un ahorro de USD\$ 12.000 para la compañía, es decir, que pequeños cambios pueden llegar a transformarse en grandes ahorros. Además una reducción en los llamados también puede traducirse como una mayor satisfacción de los clientes más importantes del programa con los servicios ofrecidos por LANPASS.

Para esto, dentro de los motivos más demandantes del año 2014, como lo son el “Check-in Upgrade”, cambios y devoluciones, acreditaciones, venta, canje e información, se identificarán las variables más significativas que describen el contexto de los socios que están ocupando el servicio del contact center. En particular, se estudiará el motivo de “Check-in Upgrade” debido a su gran impacto en el total de las llamadas, siendo el principal motivo por el cual los socios al día de hoy están llamando y también por ser un proceso que cuenta con un proceso regular, el cual puede ser más fácil de corregir desde el área operativa del programa.

Una vez identificadas estas variables, se procederá a estimar un modelo de propensión en el motivo a estudiar “Check-in Upgrade”, de tal forma que se pueda saber cuáles variables del contexto del socio son las que están provocando un aumento considerable en la probabilidad de llamar al contact center.

Finalmente, esto constituirá una nueva herramienta para la empresa, ya que actualmente para el servicio del “Check-in Upgrade” de los pasajeros sólo existe el envío de un email 72 horas antes del vuelo, y aún con dicha herramienta se están produciendo 48.820 llamados anuales por los clientes preferentes relacionados a este motivo. Por esto mismo, se ve reflejado que no se tiene una apreciación de cuáles son las características principales que hacen que los clientes ocupen el servicios del contact center. Por lo mismo, con los resultados de este trabajo se tendrá el poder de realizar acciones de manera más acertada y argumentada.

### **3 OBJETIVOS**

#### **3.1 Objetivo General**

- Identificar a los clientes que tienen una alta probabilidad de llamar al contact center, con el fin de poder realizar recomendaciones de acciones a cada uno de ellos que puedan evitar dicho llamado.

#### **3.2 Objetivos Específicos**

- Identificar los motivos<sup>5</sup> más demandados del contact center según el año 2014.
- Describir las variables que definen el perfil de cliente que llaman al contact center según el motivo estudiado.
- Desarrollar un modelo de propensión, que permita identificar cuáles son las variables que hacen aumentar la probabilidad de llamada de clientes, probar las hipótesis planteadas y construir los perfiles más propensos a llamar.
- Describir los lineamientos y la metodología de las acciones posibles a tomar, en base a los resultados encontrados, junto con una propuesta de evaluación económica respectiva a cada acción.

---

<sup>5</sup> Motivo: se refiere a la razón general por que el socio esta llamado al contact center de la aerolínea, como puede ser: check-in, devoluciones, canje, información, etc.

## **4 ALCANCES**

- El presente trabajo solo analizará la atención preferente, debido a la información disponible, ya que de la forma en que los ejecutivos del contact center tipifican las llamadas, sólo es posible conectar los motivos de llamada con el socio que llama en el grupo de atención antes mencionado.
- La información se limitara a los datos del 2014.
- Se trabajará sólo con el motivo “Check-in Upgrade”, por la utilidad que presentan a la hora de poder tomar alguna acción futura.
- La presente memoria, no busca hacer un cambio en la estructura del contact center, como por ejemplo cambiar la capacidad, definir las habilidades de los ejecutivos, etc., sino que pretende dar información argumentada para que el área operativa del programa pueda tomar acciones sobre los clientes o perfiles de clientes encontrados
- En el presente trabajo se definirán las acciones que se tomaran con los distintos perfiles de clientes, su metodología y sus costos respectivos, pero no se harán los experimentos de éstas.

## **5 MARCO CONCEPTUAL**

### **5.1 Programas de fidelización**

Un programa de fidelización es un programa de marketing que es diseñado para construir una lealtad por parte del consumidor, proporcionando incentivos a los clientes rentables. Este último se basa a menudo en varias proposiciones, como las siguientes [2]:

- Los clientes puede que quieran una mayor relación con los productos que compran.
- Una proporción de estos consumidores muestra una tendencia de ser leal.
- Existe un grupo rentable (ley del 80/20).
- Es posible reforzar la lealtad de estos consumidores a través de un programa de fidelización.

Pueden existir distintas definiciones para lo que es un programa de fidelización, pero todas tienen como factor común el establecer una relación con clientes premiándolos por

la preferencia a la marca. Aunque el problema reside en como cultivar dicha preferencia en presencia de un mercado con una alta competencia.

El presente trabajo estará desarrollado en el programa de fidelización de LAN, denominado LANPASS.

## **5.2 Contact Center**

Un contact center es una oficina centralizada usada con el propósito de recibir y transmitir una amplia cantidad de llamados y pedidos a través del teléfono, los cuales se pueden realizar por canales adicionales al teléfono, tales como fax, email, mensajes de texto, mensajes multimedia, entre otros<sup>6</sup>. Es esto último lo que diferencia un contact center de un call center, ya que este último sólo ocupa el medio telefónico para la prestación de servicios.

La presente memoria aborda un problema de negocio en el contexto de la reducción de las cantidades de llamada a un contact center realizadas por las categorías más altas de un programa de fidelización.

## **6 MARCO TEORICO**

### **6.1 Modelos de propensión**

Los modelos que pueden ser ocupados en marketing, tienen una variada gama de usos debido al comportamiento de los consumidores. Las principales razones se mencionan a continuación [3]:

- Consumidores son diferentes.
- Alternativas de decisión son distintas.
- El contexto de compra es distinto.
- Necesidades gerenciales difieren.

Por esto mismo, la generación de un modelo es uno de los problemas más interesante a resolver, debido a la dificultad que presenta estimar la respuesta de los consumidores que se espera que las herramientas de marketing generen sobre ellos.

---

<sup>6</sup> Centro de llamadas Wikipedia: [https://es.wikipedia.org/wiki/Centro\\_de\\_llamadas](https://es.wikipedia.org/wiki/Centro_de_llamadas)

A continuación se describirán conceptualmente los modelos a utilizar en el trabajo de memoria.

## 6.2 Regresión Logística (Logit)

El modelo de regresión logística se basa en la obtención de la probabilidad de que una observación pertenezca a un conjunto determinado, en función del comportamiento de las variables independientes [4].

Por lo tanto, buscamos describir la probabilidad de que un individuo  $i$  elija la alternativa  $j$ . Asumiendo que el individuo decide entre las alternativas maximizando la utilidad, esta última se descompondrá en una componente determinística y otra estocástica, de la siguiente forma:

$$u_{ij} = v_{ij} + \varepsilon_{ij}$$

Lo que hace que la probabilidad de que el individuo  $i$  escoja la alternativa  $j$ , es definida como:

$$\begin{aligned} P_{ij} &= \Pr(u_{ij} > u_{ik}, \forall j \neq k) \\ &= \Pr(v_{ij} + \varepsilon_{ij} > v_{ik} + \varepsilon_{ik}, \forall j \neq k) \\ &= \int \mathbf{1}_{[v_{ij} + \varepsilon_{ij} > v_{ik} + \varepsilon_{ik}, \forall j \neq k]} f(\varepsilon_i) d\varepsilon_i \end{aligned} \quad (1)$$

En este modelo, la distribución de cada  $\varepsilon_{ij}$  asume que se distribuye como valor extremo:

$$f(\varepsilon_{ij}) = e^{-\varepsilon_{ij}} e^{-e^{-\varepsilon_{ij}}}$$

Desarrollando matemáticamente la expresión (1), llegamos a que la probabilidad de elección queda descrita como:

$$P_{ij} = \frac{e^{\beta' x_{ij}}}{\sum_k e^{\beta' x_{ik}}} \quad (2)$$

Donde se estimarán los  $\beta$  de (2) a través del método de máxima verosimilitud y se calculará la probabilidad asociada a cada individuo.

### 6.3 Árboles de decisión

Los árboles de decisión es una técnica de minería de datos muy potente de clasificación y una de las más usadas debido tanto a las posibilidades que brinda, como la facilidad para interpretar sus resultados.

Los árboles de decisión consisten en un modelo estadístico en los que interesa explicar una variable dependiente, en nuestro caso variable binaria donde 1 es si el socio llama y 0 si el cliente no llama, en función de una serie de variables explicativas. Además, esta técnica permite segmentar, clasificar, predecir, reducir la dimensión de los datos, identificar y recodificar.

El árbol está formado por nodos y ramas, donde cada rama del árbol se define como nivel, donde las primeras ramificaciones se llevan a cabo con la variable que mejor discrimina. La metodología que usan los árboles para ramificar es clasificada como Greedy, es por esto que se da la posibilidad de que un atributo aparezca en diferentes niveles de las ramificaciones. Además se debe cumplir que haya un solo nodo raíz, es decir, un nodo que no tiene padre y que cada nodo distinto de la raíz tenga un solo padre.

Existen cuatro métodos de clasificación vía árboles:

- **CHAID (Chi-square automatic interaction detector):** Consiste en un rápido algoritmo de árbol estadístico y multidireccional que explora datos de forma rápida y eficaz, y crea segmentos y perfiles con respecto al resultado deseado. Permite la detección automática de interacciones mediante Chi-cuadrado. En cada paso, CHAID elige la variable independiente (predictora) que presenta la interacción más fuerte con la variable dependiente. Las categorías de cada predictor se funden si no son significativamente distintas respecto a la variable dependiente [5].
- **CHAID exhaustivo:** Supone una modificación de CHAID que examina todas las divisiones posibles para cada predictor y trata todas las variables por igual, independientemente del tipo y el número de categorías [5].
- **Árboles de clasificación y regresión (CRT-Classification and regression trees):** Consiste en un algoritmo de árbol binario completo que hace particiones de los datos y genera subconjuntos precisos y homogéneos. CRT divide los datos en segmentos para que sean lo más homogéneos posible respecto a la variable dependiente [5]. El grado en que un nodo no representa homogeneidad es una indicación de impureza. Para la medida de la impureza para variables categóricas como es el caso que se presenta, se pueden seleccionar distintas medidas:

- **Gini:** es una medida de impureza general que aplica a la rama probabilidades de pertenencia a categorías. Este indicador se calcula de la siguiente manera:

$$Gini = 1 - \sum_i Q(t)_i^2$$

Donde  $Q(t)_i$  es la proporción de casos en el nodo  $t$  que pertenecen a la categoría de  $i$  del campo resultado.

- **Binaria:** las categorías de la variable dependiente se agrupan en dos subclases. Se obtienen las divisiones que mejor separan los dos grupos.
- **QUEST (Quick, unbiased, efficient, statistical tree):** Consiste en un algoritmo estadístico que selecciona variables sin sesgo y crea árboles binarios precisos de forma rápida y eficaz. Con cuatro algoritmos tenemos la posibilidad de probar métodos diferentes de crecimiento de los árboles y encontrar el que mejor se adapte a nuestros datos. Es un método rápido y que evita el sesgo que presentan otros métodos al favorecer los predictores con muchas categorías. Sólo puede especificarse QUEST si la variable dependiente es nominal [5].

## 6.4 Medidas de error

Para evaluar las predicciones se utilizan distintas medidas del error dependiendo del modelo. Para estimaciones del tipo regresión lineal (simple o múltiple), se utiliza el  $R^2$  y el  $R^2 - \text{ajustado}$ . El coeficiente de determinación o  $R^2$ , varía entre 0 y 1, y mide la proporción del total de la varianza de la variable dependiente que es explicada por las variables independientes. El  $R^2 - \text{ajustado}$  mide lo mismo que  $R^2$ , pero ajustado por el grado de libertad del modelo [6].

Para poder medir como ha sido el desempeño del pronóstico, se utilizaran medidas para comparar el valor predicho con el real. También estas medidas pueden ser utilizadas para la calibración de datos. Para esto se tienen las siguientes medidas de error:

### 6.4.1 Raíz del error cuadrático medio

$$RMSE = \sqrt{\frac{1}{n} \sum_i (y_i - \hat{y}_i)^2}$$

Este indicador nos entrega el promedio de los errores cuadráticos obtenidos en el pronóstico.

#### 6.4.2 Error absoluto medio

$$MAE = \frac{1}{n} \sum_i |y_i - \hat{y}_i|$$

Este indicador en palabras simples nos dice cuanto se equivoca el modelo.

#### 6.4.3 Máximo error absoluto porcentual

$$MAPE = \frac{1}{n} \sum_i \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

Este indicador entrega en promedio el error absoluto obtenido en el pronóstico de toda la serie, como porcentaje de la serie original.

#### 6.4.4 Medidas de evaluación de desempeño

Otras medidas que son frecuentemente utilizadas en estadísticas para cuantificar cuan bueno o confiable es el modelo son la sensibilidad, especificidad y la precisión. Por un lado la sensibilidad evalúa la proporción de verdaderos positivos que son correctamente identificados por el modelo, por otro lado la especificidad evalúa la proporción de verdaderos negativos que son clasificados correctamente. Para ver de mejor manera cual es la relación que existe entre estas dos medidas, está lo que se denomina curva ROC, la cual ayuda gráficamente en la decisión de si un modelo es bueno o no.

En la tabla 1, se puede observar cada uno de los casos obtenido por el modelo correspondiente y el valor real, de donde además nacen las correspondientes medidas de evaluación.

Real	Pronosticado	
	Positivo	Negativo
Verdadero	TP	FN
Negativo	FP	TN

*Tabla 1: Matriz de confusión*

De esta tabla se desprende por ejemplo, que los “Verdaderos Positivos” (TP), corresponden a cuando el modelo clasifica a un cliente como que llamó y en la realidad efectivamente el cliente llamó. Del mismo modo los “Verdaderos Negativos” (TN), corresponde a cuando el modelo clasifica a un cliente como que no llamó y lo observado el cliente no llamó. Ahora bien, por otro lado “Falsos Positivos” (FP) quiere decir cuando el modelo dice que un cliente llamó, pero en la realidad dicho cliente no llamó y “Falsos Negativos” (FN) corresponde a decir que el modelo dice que un cliente no llamó, pero en la realidad dicho cliente llamó. Estas dos últimas clasificaciones corresponden al error tipo I y error tipo II respectivamente.

A raíz de estas definiciones podemos dar paso a la descripción de las siguientes medidas de evaluación de desempeño:

#### **6.4.4.1 Sensibilidad**

La sensibilidad como se mencionó anteriormente nos proporciona una medida de cuantos casos positivos son bien calificados sobre el total de casos positivos. Este suele representarse en forma de porcentaje, tal como se muestra a continuación:

$$\text{Sensibilidad} = \frac{TP}{TP + FN}$$

Este indicador es cuan bueno es el modelo para identificar que el cliente llamó, es decir, una sensibilidad de un 99% significará que cuando realizamos el diagnóstico para un cliente que llamo, con un 99% de probabilidad el modelo lo clasificara como que llamó.

#### **6.4.4.2 Especificidad**

Al igual como se mencionó en un principio, la especificidad muestra la capacidad del modelo de poder clasificar a los clientes que no llaman efectivamente como clientes que no llaman. Esta suele calcularse de la siguiente manera:

$$\text{Especificidad} = \frac{TN}{TN + FP}$$

En otras palabras, lo que quiere decir este indicador es la probabilidad de clasificar bien a un cliente que no llamó, es decir, si la especificidad es de un 99%, quiere decir que cuando hacemos un diagnostico en un cliente que no llamó, hay un 99% de probabilidad que se le califique como un cliente que no llamó.

### 6.4.4.3 Accuracy

Otro indicador importante a la hora de evaluar un modelo es la “accuracy”, la cual corresponde a la proporción de verdaderos sobre toda la población. Se suele representar de la siguiente forma:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Este indicador lo que busca mostrar es cuan bueno es el modelo para identificar los casos verdaderos respecto a todos los casos, es decir, la efectividad de clasificar a los clientes que llamaron como que llamaron y a los que no llamaron como que no llamaron. Al igual que los indicadores anteriores también se mide generalmente en porcentaje y entre más cercano a un 100% mejor será el modelo, ya que esto significaría que con un 100% el modelo clasifica bien cada caso.

De forma alternativa, la accuracy se puede calcular de la siguiente forma:

$$Accuracy = sensibilidad * prevalencia + especificidad(1 - prevalencia)$$

Donde la prevalencia corresponde para este caso, a la probabilidad de llamar en la población en un momento dado del tiempo.

Finalmente cabe recalcar que un modelo puede ser bueno en especificidad sin ser bueno en sensibilidad o viceversa, ambos indicadores son igual de importantes, aunque obviamente un modelo es mejor si ambos indicadores tienen un valor alto.

### 6.4.4.4 Precisión

La precisión mide la proporción de verdaderos positivos frente a todos los casos positivos. Generalmente se expresa como un porcentaje y se calcula de la siguiente manera:

$$Precisión = \frac{TP}{TP + FP}$$

Este indicador muestra como son nuestras predicciones en relación a lo que realmente ocurrió.

#### 6.4.4.5 LIFT

El lift es otra medida que permite comparar diferentes resultados producidos por distintos algoritmos. Está orientado a la evaluación de problemas de clasificación, ya que principalmente mide la diferencia de eficacia entre dos modelos.

Este indicador generalmente se calcula como el cociente entre el porcentaje de concentración de elementos de una determinada clase, sobre la concentración que presenta la población en su conjunto, tal como se muestra en la siguiente ecuación:

$$LIFT = \frac{\textit{Porcentaje objetivo en clase}}{\textit{Porcentaje objetivo en población}}$$

Esta medida no es un porcentaje, sino que es un indicador de cuantas veces es mejor el modelo en la captación del hecho objetivo que la aleatoriedad, es decir, que a mayor magnitud del lift mejor candidato será el modelo, debido a que éste estará evaluado como un modelo de buen desempeño.

#### 6.4.4.6 Curva ROC

Para formar el espacio ROC, se necesitan calcular los ratios de verdaderos positivos (TPR) y falsos positivos (FPR) de la siguiente manera:

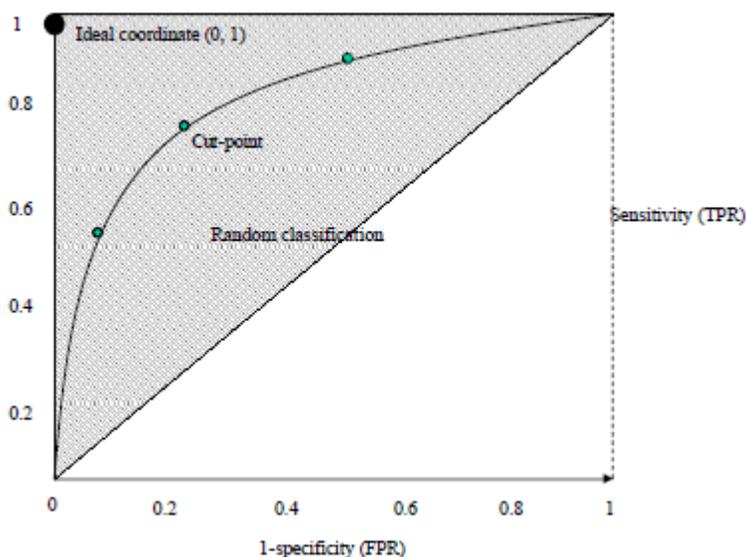
$$TPR = \frac{TP}{TP + FN} = \textit{Sensibilidad}$$

$$FPR = \frac{FP}{FP + TN} = (1 - \textit{Especificidad})$$

Una vez calculados estos ratios, el espacio ROC se define como todas las combinaciones posibles entre TPR y FPR, donde una combinación particular entre ellos representa un simple punto en dicho espacio y la posición de este punto representa el trade-off entre la sensibilidad y la especificidad, es decir, un incremento en la sensibilidad del modelo es acompañada en una disminución en la especificidad. De tal forma, la ubicación de un punto en el espacio ROC corresponde a si la clasificación es buena o mala.

En una situación ideal, un punto determinado por TPR y FPR con coordenada (0,1) representa una sensibilidad del 100% y una especificidad del 100%, lo que se conoce como clasificación perfecta. Si se tiene una sensibilidad del 50% y además una especificidad del 50%, esto se podrá observar por la recta generada por los puntos (0,0) y (1,0), como se

muestra en el gráfico 1. Teóricamente una clasificación al azar dará un punto perteneciente a esta recta. Al realizar la predicción en base un modelo, si un punto da sobre la recta antes mencionada, se puede considerar como un buen modelo.



**Gráfico 1: Curva ROC**

Los distintos puntos de corte definen la curva ROC, donde cada punto de la curva entre más cercano a la diagonal peor es el modelo, y entre más cercano al punto de clasificación perfecta mejor es el modelo. Otra interpretación interesante, es que entre más rápido la curva se acerca al punto ideal (0,1), los resultados del modelo serán mejores.

#### 6.4.4.7 AUC

Un indicador relevante a la hora de clasificar las curvas ROC, es el área bajo la curva o AUC (“Area Under Curve”), la cual provee una medida de la precisión del modelo. A mayor el área bajo la curva, mejor será la predicción del modelo. Sus valores varían entre [0,5; 1] y se calcula de la siguiente manera:

$$AUC = \int_0^1 ROC(t)dt$$

Donde  $t = (1 - \text{especificidad})$  y  $ROC(t)$  es la sensibilidad.

Otra manera de interpretar este indicador es que el valor de AUC en presencia de dos individuos, representa la probabilidad de clasificarlos de manera correcta a cada uno.

Con estos criterios, se puede concluir que la curva ROC es un buen método para seleccionar un posible punto de corte óptimo para determinar la precisión del modelo. Una clasificación de la precisión usando AUC puede observarse en la sección 10 anexos 3.

## 6.5 Test de Hipótesis

En estadística, en muchas ocasiones el objetivo es poder hacer inferencia con respecto a ciertos parámetros poblacionales desconocidos, que están hechos en base a la información obtenida mediante datos muestrales. Estas inferencias pueden expresarse de dos maneras, como estimaciones de los parámetros o como test de hipótesis referentes a sus valores.

Respecto a la segunda forma, test de hipótesis, tenemos los siguientes tipos de contrastes [7]:

### 6.5.1 Test de significancia individual

Se plantea como:

$$\begin{aligned} H_0: \beta_k &= 0 \\ H_A: \beta_k &\neq 0 \end{aligned}$$

El estadístico correspondiente se define como:

$$\frac{\widehat{\beta}_k - \beta_k}{DS(\widehat{\beta}_k)} \mid X \sim T_{n-k}$$

Por lo tanto asumiendo que  $H_0$  es correcto, se cumpliría que:

$$\Pr(-T_{1-\frac{\alpha}{2}, n-k} < \frac{\widehat{\beta}_k}{DS(\widehat{\beta}_k)} < T_{1-\frac{\alpha}{2}, n-k} \mid X) = 1 - \alpha$$

Por ello rechazamos  $H_0$ , si  $\left| \frac{\widehat{\beta}_k}{DS(\widehat{\beta}_k)} \right| > T_{1-\frac{\alpha}{2}, n-k}$ , lo que determina la región de rechazo.

### 6.5.2 Test de significancia global

Se define el test de la siguiente forma:

$$\begin{aligned} H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0 \\ H_A: \beta_1 \neq \beta_2 \neq \dots \neq \beta_k \neq 0 \end{aligned}$$

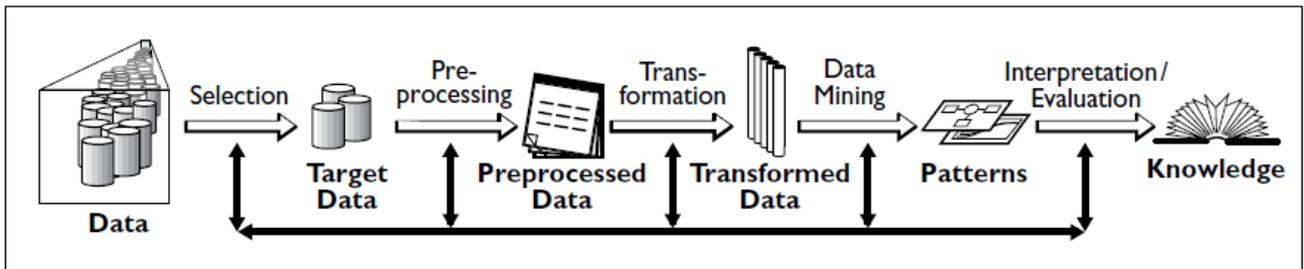
Si escribimos el modelo de desvíos respecto a la media  $y = x\beta^* + U$ , entonces  $C\beta^* = 0$ , con  $C = I_K$ . Reemplazando en el test  $C\beta$  obtenemos que:

$$F = \frac{\widehat{\beta}^{*'} x' x \widehat{\beta}^*}{K \widehat{\sigma}^2} = \frac{\frac{SCE}{K}}{\frac{SCR}{N-K}} = \frac{R^2}{1-R^2} \left( \frac{n-k}{k} \right) \sim f_{k, n-k}$$

Hay que tener en cuenta que, aún con covarianza cero, un test conjunto de R coeficientes no es equivalente a R tests individuales, ya que el test conjunto construye la región de mínimo tamaño (área, volumen, etc.) donde  $H_0$  sea rechazado incorrectamente sólo con probabilidad  $\alpha$  considerando las covarianzas. Estas últimas entregan información importante para evaluar la plausibilidad de un valor de un coeficiente dado los valores de otros.

## 7 METODOLOGÍA Y DESARROLLO METODOLÓGICO

La metodología del presente trabajo está basado en el proceso KDD (Knowledge Discovery in Databases), la cual puede ser resumida en la ilustración 1 [8].



*Ilustración 1: Proceso KDD*

Los pasos de este proceso se pueden resumir en los siguientes 9 puntos:

1. **Aprender el dominio de la aplicación:** aprender el conocimiento y objetivos relevantes de la aplicación.

2. **Crear un conjunto de datos objetivos:** selección del conjunto de datos, muestra o sub muestra.
3. **Limpieza y pre procesamiento de los datos:** incluye las operaciones básicas para reunir la información necesaria para el modelo.
4. **Reducción y proyección de los datos:** incluye encontrar interesantes características para la representación de los datos.
5. **Elección de la función de data mining:** decisión del propósito del modelo derivado del algoritmo de data mining.
6. **Escoger el algoritmo de data mining:** selección del método.
7. **Data mining:** búsqueda de patrones de interés que representen el conjunto.
8. **Interpretación:** en esta etapa se interpretan los patrones encontrados en la etapa anterior.
9. **Descubrimiento de conocimiento:** esta etapa incluye el tomar acciones, reportar los resultados, etc.

En base a los 9 puntos anteriores, para esta memoria la metodología será desarrollada como se describe en la tabla 2:

Proceso Memoria	Proceso KDD
Obtención de datos.	Selección de datos
Selección y procesamiento de motivo escogido.	Pre procesamiento y transformación
Caracterización de clientes e identificación de variables relevantes.	Pre procesamiento y transformación
Análisis exploratorio.	Reducción y proyección de los datos.
Desarrollo, análisis y evaluación de modelos de propensión.	Data Mining e interpretación
Diseño de acciones.	Descubrimiento de conocimiento.

*Tabla 2: Metodología memoria y su relación con el proceso KDD.*

*Fuente: Elaboración Propia.*

## 7.1 Obtención de datos

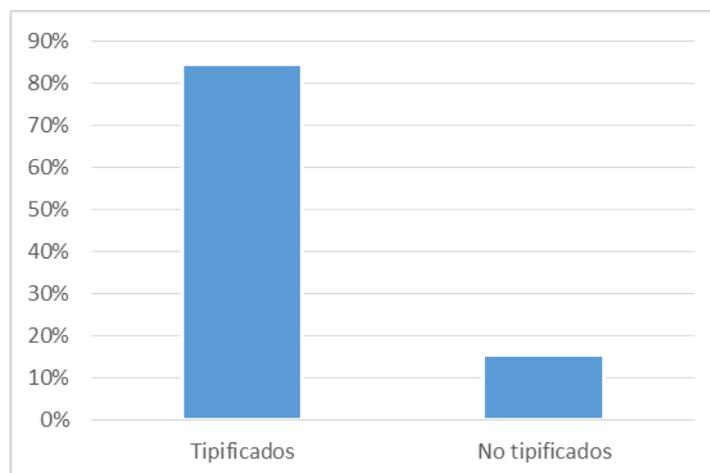
Para el presente trabajo se ocuparan cuatro fuentes principales de datos proporcionados por el área de operaciones de LANPASS, los cuales son:

- **Tipificaciones:** Principalmente de esta base de datos se obtienen los motivos por el cual los clientes están llamando al contact center y también se puede encontrar el ID que vincula a las tipificaciones con la siguiente fuente llamada IVR. Además se pueden encontrar datos de interés como el mercado, día, hora, especialidad, etc.
- **IVR:** Esta fuente de datos principalmente nos da acceso al número de socio que realizo la llamada, con lo cual podremos acceder a todo lo que son las variables socio-demográficas, actividad, eventos, etc. del socio presente en la fuente "Customer".

- **Customer:** de esta base de datos, lograremos obtener la información que describirá los vuelos correspondientes asociado a cada tipificación, es decir, será nuestra principal fuente de información.
- **LAN.COM:** de esta base podremos obtener información sobre los clientes y su navegación en la página.

## 7.2 Análisis exploratorio: Llamados 2014

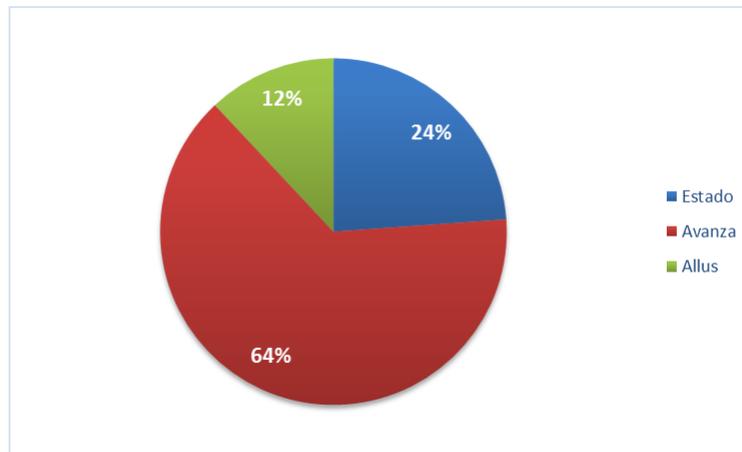
En el año 2014 hubo aproximadamente 848.000 llamados al contact center, de los cuales sólo el 85% de estos fueron tipificados (gráfico 2), es decir, los ejecutivos pudieron describir a que motivo pertenecía tal llamado.



**Gráfico 2: Llamados Tipificados 2014.**

**Fuente: Datos Empresa, Elaboración Propia.**

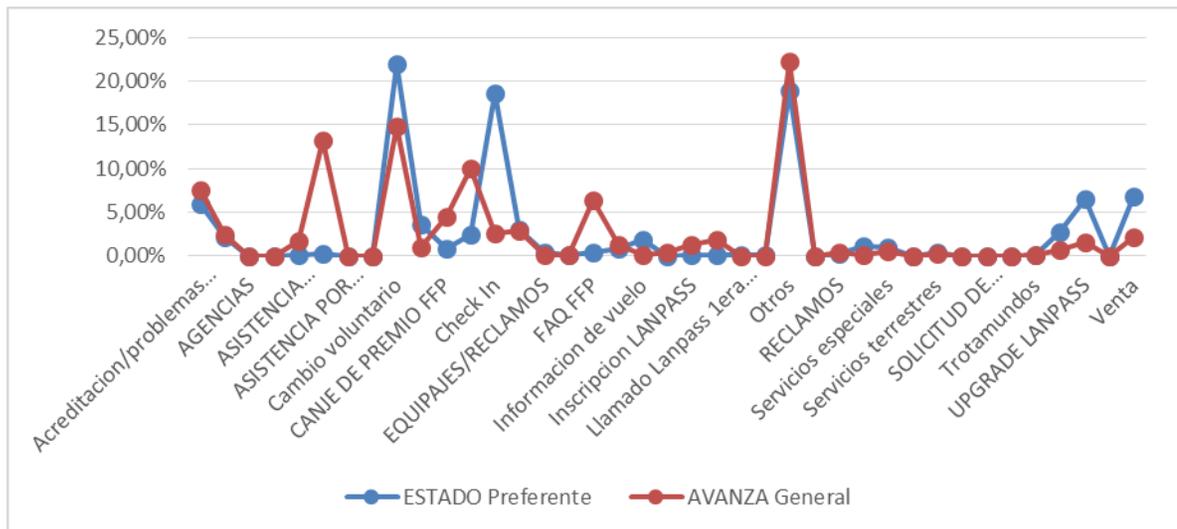
Por otro lado, existen distintas líneas de comunicación entre los clientes y el programa, esto por medidas de seguridad en caso que falle alguno de los servicios de contact center. Dentro de estos se encuentran los tres principales al día de hoy: Allus (servidor peruano), Avanza (servidor colombiano) y Estado (servidor interno del programa). De los tres mencionados los dos primeros se hacen cargo de la atención general, en cambio el tercero se hace cargo de la atención preferente. La distribución de llamadas tipificadas por línea de comunicación se puede observar en el gráfico 3.



**Gráfico 3: Distribución Tipificaciones por Línea de Comunicación.**

**Fuente: Datos Empresa, Elaboración Propia.**

Debido a la forma de tipificar los llamados por parte de los ejecutivos del contact center, el presente trabajo se realizará sólo analizando a la atención preferente. Aunque esto se realiza por la disponibilidad de datos, cabe mencionar que por un lado ésta atención se hace cargo de los socios más valiosos del programa, es decir, son los socios a los que hay que tener más satisfechos y por otro lado, según la distribución de llamados por motivo entre ambas atenciones se puede decir, que ambos grupos se comportan de manera similar en la mayoría de los motivos a llamar, como se puede observar en el gráfico 4.



**Gráfico 4: Distribución Tipificaciones Atención General vs Preferente.**

**Fuente: Datos Empresa, Elaboración Propia.**

Dado lo anterior, podemos empezar a trabajar en el análisis de nuestras variables candidatas para el modelo, sólo con la información de la atención preferente y además como se puede observar en el mismo grafico anterior, primero se deberá limpiar la base de datos, ya que hay tipificaciones que significan lo mismo y también hay otras que presentan una demanda casi nula. Por esto mismo, el trabajo final a realizar será con las tipificaciones que presenten un mayor efecto en la demanda al contact center.

### 7.3 Limpieza de datos

Como vimos anteriormente, en las tipificaciones 2014 existían por un lado algunos motivos a lo largo del año que tuvieron una demanda casi nula y por otro, tenemos motivos que representaban el mismo tipo de llamado, por esto se procedió a realizar una consolidación de estos datos, ya sea en base a juicio de expertos del negocio o sólo por pequeñas diferencias a la hora de tipificar, lo que dio como resultado lo que se muestra en la tabla 3.

Tipificación Final	Tipificaciones Originales
Acreditación (*)	Acreditacion/problemas en datos de cartola LANPASS
Asistencia web e información (*)	ASISTENCIA EN LAN.COM ASISTENCIA POR CAMBIO DE ITINERARIO SOLICITUD DE INFORMACION
Cambios y devoluciones	CAMBIO / REEMISION Cambio voluntario Cambios involuntarios Devoluciones
Check-in Upgrade (*)	Check in UPGRADE FFP
Inscripción (*)	INSCRIPCION FFP
Reclamos (*)	EQUIPAJES/RECLAMOS
Canje (*)	CANJE DE PREMIO FFP
Información (*)	INFORMACION DE VIAJE
Llamado 1ª línea (*)	LLAMADO 1ERA LINEA
Otros	Todo los demás

**Tabla 3: Tipificación Final a Trabajar.**

**Fuente: Datos Empresa, Elaboración Propia.**

(\*): Son las tipificaciones finales, en las que su tipificación original se repetía debido a detalles mínimos como acentos, mayúsculas, etc.

Luego de limpiar las tipificaciones como se muestran en la tabla 3, se volvió a calcular la participación que tuvo cada tipificación en el año 2014, la cual se puede ver reflejado en la tabla 4. Cabe mencionar que el motivo “Otros” si bien tiene un 18%, no se considerará, ya que debido a la forma de tipificar de los ejecutivos del contact center no es posible asociarle un motivo específico.

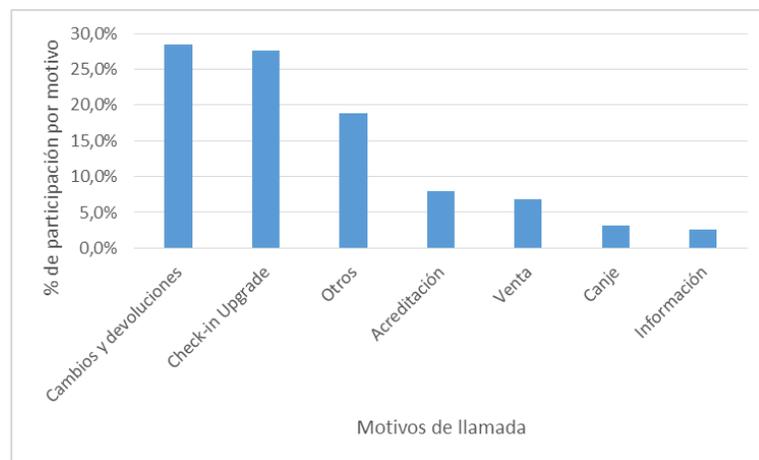
Tipificación	%
Cambios y devoluciones	28,53%
Check-in Upgrade	27,55%
Otros	18,92%
Acreditación	8,00%
Venta	6,80%
Canje	3,13%
Información	2,58%
Reclamos	1,71%
Servicios especiales	0,94%

Servicios terrestres	0,84%
Asistencia web e información	0,38%
FAQ FFP	0,33%
Inscripción	0,12%
Llamado 1ª línea	0,10%
EXCEPCIONES	0,05%
ONE WORLD	0,02%
Trotamundos	0,01%
AGENCIAS	0,00%
SERVICIOS LANPASS	0,00%
SIN TIPIFICACION	0,00%
STAFF TRAVEL (FUNCIONARIOS AEROLINEAS)	0,00%

**Tabla 4: Participación de Tipificaciones finales en la demanda del contact center.**

**Fuente: Datos Empresa, Elaboración Propia.**

De la tabla 4, también se puede observar que existen motivos que llamamos “Otros” que tienen una alta participación (18,92%), esto puede ser importante a trabajar más adelante para lograr una tipificación del 100% y tener resultados más certeros. Finalmente para poder trabajar de una manera más sencilla, se agruparán todos estos motivos de baja participación en el motivo llamado “Otros”, quedando la distribución final lo que se puede observar en el gráfico 5.

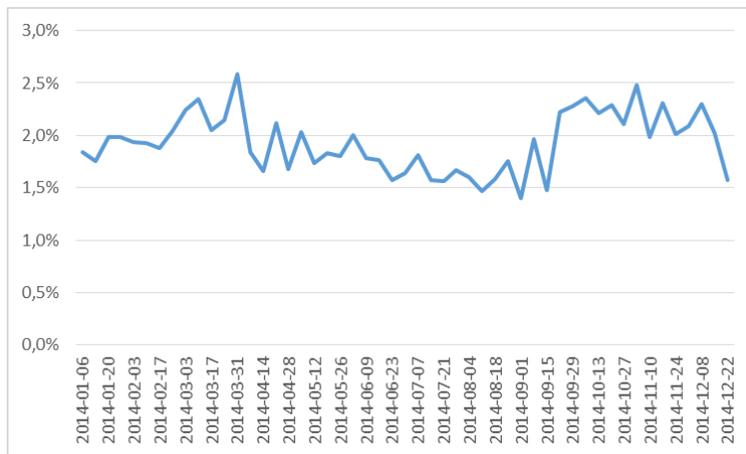


**Gráfico 5: Distribución Tipificaciones Finales Atención Preferente.**

**Fuente: Datos Empresa, Elaboración Propia.**

Una vez realizada toda la consolidación de los datos anteriores, podemos empezar a trabajar con los motivos que más impactan la demanda del contact center.

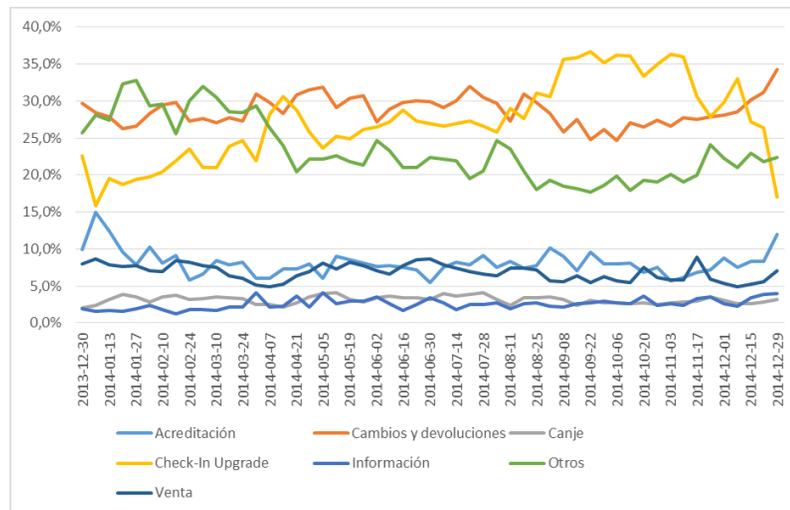
Como punto de partida se analizó la evolución de la cantidad de llamados a nivel agregado de los motivos antes encontrados a través de las semanas del 2014. Esto con el propósito de ver cómo se comporta la demanda del servicio ofrecido por el programa de fidelización y ver en qué espacio de tiempo se producen los mayores aumentos de demanda. Esto se puede apreciar en el gráfico 6.



**Gráfico 6: Distribución de la cantidad de llamadas a través de las semanas del 2014 en la atención preferente.**

**Fuente: Datos Empresa, Elaboración Propia.**

Como se puede observar en el gráfico 6, hay semanas en que la demanda del contact center tiene aumentos significativos, por lo mismo nace la necesidad de observar la misma evolución a través del tiempo, pero a nivel desagregado por motivo. Ésto para poder ver como es el comportamiento de cada uno, y poder tomar una mejor decisión sobre qué motivos trabajar más adelante. Esto se puede apreciar en el gráfico 7.



**Gráfico 7: Distribución de la cantidad de llamadas por Motivo a través de las semanas del 2014.**

**Fuente: Datos Empresa, Elaboración Propia.**

Del gráfico 7 se puede observar que, en general, salvo por el motivo de llamado “Check-in Upgrade”, los motivos principales de la demanda del contact center se mantienen relativamente “constantes” a través del tiempo. Lo que se puede traducir en que, si bien pueden no ser las mismas personas las que llaman (aunque hay evidencia que dentro de la categoría preferente si hay que gente que se repite, ver ejemplo de la tabla 5), el programa de fidelización no se ha hecho cargo de conocer cuál es el contexto que caracteriza a las personas que están haciendo uso de este servicio.

N° Socio	30-12-2013	06-01-2014	13-01-2014	20-01-2014	27-01-2014	03-02-2014	10-02-2014	17-02-2014	24-02-2014	03-03-2014	10-03-2014
1149XXX	1										
4418XXX											
5420XXX										3	
5899XXX											
6525XXX											
6713XXX					1						
6799XXX											
9028XXX											
10842XXX											1
16092XXX											3
16376XXX										1	2
17373XXX											
18831XXX		1									
21271XXX			1		2		2	3			
21508XXX										1	
22808XXX											
22869XXX						1		1		2	
23000XXX											
23051XXX				1	2						

**Tabla 5: Muestra de la cantidad de llamadas por semana de socios preferentes por Check-in Upgrade.**

**Fuente: Datos Empresa, Elaboración Propia.**

Antes de comenzar con la caracterización de clientes, es necesario mencionar que las variables horario y día de llamada, en un principio se pensaron como relevantes para el modelo de propensión, pero esta idea fue descartada debido a que no existía la posibilidad de comparación entre los clientes que no llamaron y los que si llamaron al contact center. Ahora bien la descripción de ambas variables se encuentra en la sección 10 anexo 4, debido a que puede crear una oportunidad de mejora para la organización del contact center.

A continuación se describirá el motivo a trabajar y el objetivo principal del presente trabajo de memoria.

#### **7.4 Motivo de estudio: Check-in Upgrade**

Hasta el momento en todos los análisis anteriores se ha podido observar el motivo “Check-in Upgrade”, el cual está conformado por la unión de los motivos “Check-In” y “Upgrade”, pero no se ha hecho una definición formal del motivo propiamente tal y la razón de porque se optó por unir estos dos.

Primero se justificará a través de 2 razones la unión estos dos motivos antes de su definición final.

La primera y principal razón tiene relación con el negocio, ya que la condición para poder postular a un upgrade es haber realizado el check-in, y además dicha postulación se otorga según la hora a la que se hizo el check-in, es decir, se va concediendo esta postulación según orden de anticipación de chequeo. Por lo mismo, con la empresa se optó por unir ambos motivos debido a que a juicio de los expertos del negocio, la mayoría de los que llaman y se notifican como upgrade, son debido a la urgencia de realizar el

procedimiento de check-in lo antes posible y ser los primeros candidatos a quedarse con la postulación, es decir, ambos procesos se realizan en una sola llamada.

La segunda razón para la unión de estos dos motivos se debe a la baja participación que tiene el motivo “Upgrade” en general (ver gráfico 4), ya que sólo aporta con aproximadamente un 7%, versus el motivo “Check-in” que por si solo se queda con un quinto de los motivos totales (20%). Por esto mismo, debido a su bajo impacto, se decidió unir estos dos motivos con el fin de poder atacar aproximadamente el 28% de las llamadas de este tipo de socios.

Si bien existe una tercera razón, no se pudo comprobar que entre los socios que llaman por “Check-in” o por “Upgrade” no exista una gran diferencia identificable para lograr estudiar ambos procesos por separado.

Con las dos primeras razones anteriores damos pie para unir ambos motivos y que la definición propiamente tal sea la siguiente:

**“Check-in Upgrade”:** *se define este motivo, como todas las llamadas relacionadas con el check-in para un vuelo o los llamados para las postulaciones a up-grade<sup>7</sup> correspondientes a un vuelo.”*

Es importante mencionar que el motivo “Cambios y devoluciones” siendo que tiene cerca de un 29% de participación en la demanda, no se consideró como parte del trabajo de memoria, debido a que el único canal para abordar estos problemas en el programa de fidelización es a través del contact center. Aunque no se descarta que en un futuro, se pueda estudiar este motivo y tomar acciones para reducir los llamados asociados a éste.

Ahora bien, aunque estamos abordando 1 de los 7 motivos principales (incluyendo “Otros”), el seleccionado para el trabajo representa cerca de un 28% de los llamados del 2014 para la atención preferente, lo que un cambio en su reducción puede significar un gran ahorro en los costos de la aerolínea.

Por otro lado, actualmente frente a este proceso, el área operativa del programa LANPASS, como se mencionó en un principio, sólo toma la acción de 72 horas antes del vuelo mandar un e-mail recordando al socio que se debe realizar su check-in.

Como se ha mencionado anteriormente, la idea será evaluar las variables que se describirán a continuación, para ver cuáles son las que más aportan para el aumento de la probabilidad de llamada de un cliente preferente.

---

<sup>7</sup> UP-GRADE LAN en: [http://www.lan.com/es\\_cl/sitio\\_personas/lanpass/upgrade-de-cabina/index.html](http://www.lan.com/es_cl/sitio_personas/lanpass/upgrade-de-cabina/index.html)

## 7.5 Caracterización de clientes e identificación de variables relevantes

Esta etapa consta de dos procesos principales, en primer lugar se caracterizará a los clientes preferentes según su información socio-demográfica y en segundo lugar, con el motivo escogido, “Check-in Upgrade”, se evaluarán cuáles son las variables más relevantes para usar en el modelo de propensión a desarrollar.

Para poder realizar el cruce de información de las bases de datos mencionadas en la sección 7.1, se ocupará la herramienta de business intelligence llamada PENTAHO<sup>8</sup> (sección 10 anexos 5). A través de este programa se podrá recolectar y armar la base de datos para el modelo de propensión de una manera más rápida y eficiente.

Una vez claro el motivo a trabajar y el tamaño de la muestra, lo siguiente será identificar las variables relevantes para poder armar la base de datos que logre caracterizar el contexto de los clientes que llamaron y los que no, con el fin de poder identificar cuáles son las que causan una mayor diferenciación entre ellos y un mayor impacto en la demanda del contact center. Es importante mencionar que las variables a estudiar dependerán del motivo en el cual se esté trabajando.

Las variables identificadas a estudiar por este motivo son las siguientes:

Variable	Definición
Anticipación de compra	Número de días de anticipación de compra del vuelo.
Clase	Indica si el pasajero puede postular a Upgrade.
Tipo de Vuelo	Doméstico o internacional.
Forma de Pago	Compra o canje.
Canal de pago	Internet, call center y otros.
Intento de Check-In	Indica si el pasajero intentó hacer check-in vía web (no implica que falló el chequeo web)
Escalas	Indica si el vuelo tiene escalas.
Edad	Edad del socio en años.
Genero	Genero del socio.
Antigüedad en el programa	Años que lleva el socio unido al programa.
Saldo	Kilómetros LANPASS acumulados hasta la fecha de extracción de los datos.
País	Nacionalidad del socio.
Categoría	Categoría del cliente preferente.

*Tabla 6: Variables de estudio.*

*Fuente: Elaboración Propia.*

---

<sup>8</sup> PENTAHO, programa de inteligencia de negocios, en: <http://www.pentaho.com/>

Para la realización del modelo de propensión, será necesario poder extraer la información de las variables antes mencionadas tanto de los clientes preferentes que llamaron al contact center como los que no.

Por último, cabe mencionar que luego de haber hecho el primer cruce de información, el total de llamados identificables para este motivo son de 48.820 registros, y 886.242 registros de clientes correspondientes a los mismos vuelos que no llamaron, es decir, aproximadamente sólo un 5,2% de todos los posibles socios llamaron. Es por esta razón que más adelante, al desarrollar los modelos de propensión se analizará la opción de trabajar con una muestra balanceada, para luego verificar que dicho resultado no está sesgado y poder aplicarlo a toda la base de datos.

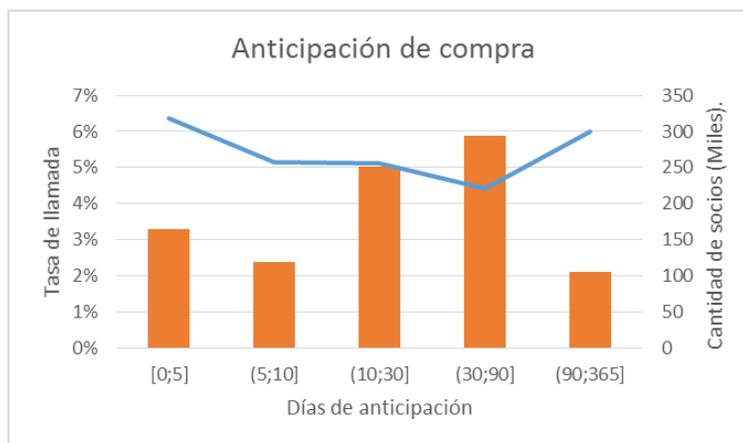
### 7.5.1 Análisis exploratorio: “Check-in Upgrade”

Antes de comenzar con dicho análisis, se aclarará cómo se calculó la tasa de llamada presentada en el siguiente análisis, para todas las variables en cuestión. Tenemos que en cada variable existen distintos segmentos o casos, como por ejemplo, para la anticipación de compra tenemos 5 segmentos o tramos de anticipación, para el tipo de vuelo tenemos si es internacional o doméstico, etc. y para cada uno de ellos tenemos una cantidad de socios que llamó y otra que no. Por lo tanto, la tasa de llamada se calcula como los que llamaron sobre el total para cada segmento, esto se puede apreciar de mejor forma en la siguiente ecuación:

$$Tasa\ de\ llamada_{anticipación_t} = \frac{(Socio\ llama)_{anticipación_t}}{(Socio\ llama)_{anticipación_t} + (Socio\ no\ llama)_{anticipación_t}}, \text{ con } t \in [0, 5]$$

Una vez aclarado esto, se muestra el análisis de las variables antes mencionadas donde el gráfico de barras corresponde a la cantidad de socios y el de líneas a la tasa de llamada.

#### 7.5.1.1 Anticipación de compra

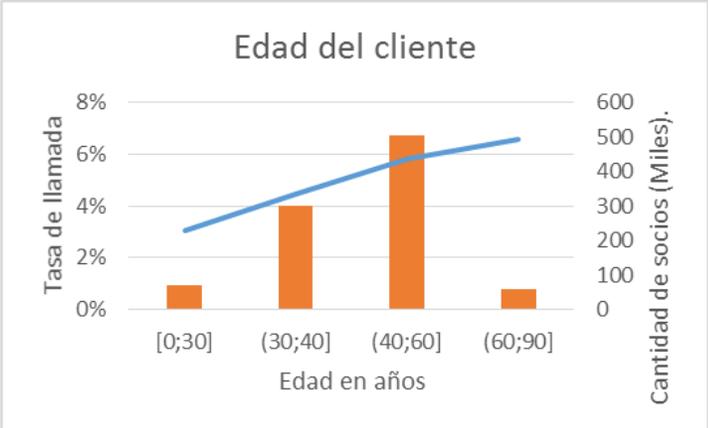


**Gráfico 8:** Tasa de llamada vs Cantidad de socios en Anticipación de compra en días.

*Fuente: Datos Empresa, Elaboración Propia.*

Del gráfico 8 se puede apreciar que tanto en el tramo de poca anticipación de compra (0; 5] como en el de más de 3 meses (90; 365], la probabilidad de que una persona haya llamado es alta, siendo inclusive una menor cantidad de socios con respecto a los otros tramos. Por eso mismo, la variable anticipación de compra en tramos puede ser interesante de estudiar, debido al comportamiento que presentan los socios tanto en el primer como en el último tramo.

**7.5.1.2 Edad del socio**

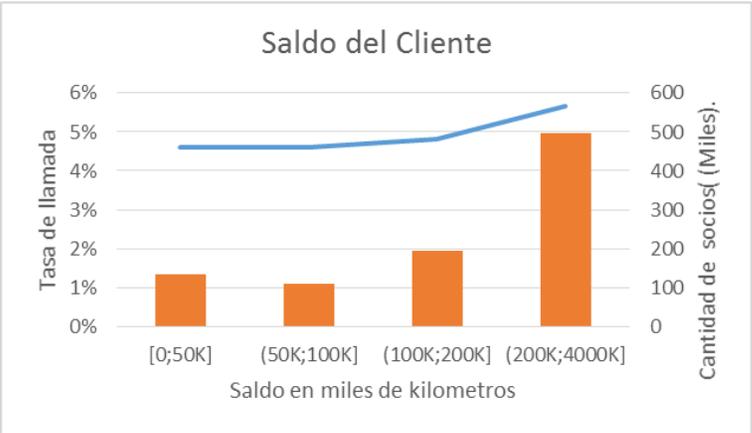


**Gráfico 9.- Tasa de llamada vs Cantidad de socios en Edad de los socios.**

*Fuente: Datos Empresa, Elaboración Propia.*

Según el gráfico 9, podemos notar que si bien a mayor edad del socio, la probabilidad de que llame es mayor, es de interés estudiar a los clientes jóvenes (menores de 30 años) y a los adultos mayores a 60 años, debido a que en cuanto a cantidad de socios no representan una mayor participación, pero su tasa de llamada es alta, en especial en el último tramo mencionado anteriormente.

**7.5.1.3 Saldo**

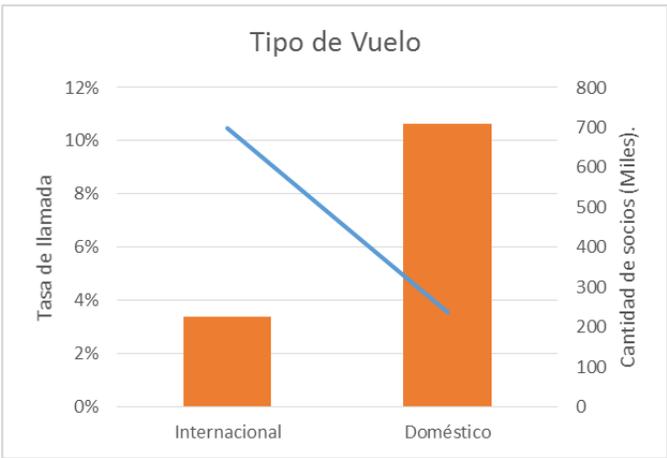


**Gráfico 10.- Tasa de llamada vs Cantidad de socios respecto al saldo acumulado (miles de KMS)**

*Fuente: Datos Empresa, Elaboración Propia*

Del gráfico anterior podemos observar que la probabilidad de que un socio llame según su saldo acumulado se comporta de manera similar, salvo en los últimos dos rangos de saldo. Para esto se realizó un test de proporciones para los últimos dos rangos de esta variable, dando como resultado que su diferencia sí es significativa, por lo cual podría ser de interés ver cómo influye el tener un saldo mayor a 200 mil kilómetros a la hora de realizar el llamado al contact center. El desarrollo del test se puede observar en la sección 10 anexo 6.

**7.5.1.4 Vuelo internacional**



**Gráfico 11.- Tasa de llamada vs Cantidad de socios según el tipo de vuelo.**

*Fuente: Datos Empresa, Elaboración Propia*

El gráfico 11, nos muestra un comportamiento interesante, debido a que nos presenta que si bien la cantidad de socios que tienen un viaje internacional son pocos, ellos son muy propensos a llamar para realizar su “Check-in Upgrade”.

**7.5.1.5 Vuelos con escalas**



**Gráfico 12.- Tasa de llamada vs Cantidad de socios en Vuelos con o sin escalas.**

*Fuente: Datos Empresa, Elaboración Empresa.*

Al igual que en el gráfico anterior, que un socio tenga un viaje con escalas, se comportará de manera similar a cuando tiene un viaje internacional, es decir, que son pocos los casos con escalas, pero su tasa de llamada es alta.

### 7.5.1.6 Antigüedad en el programa LANPASS

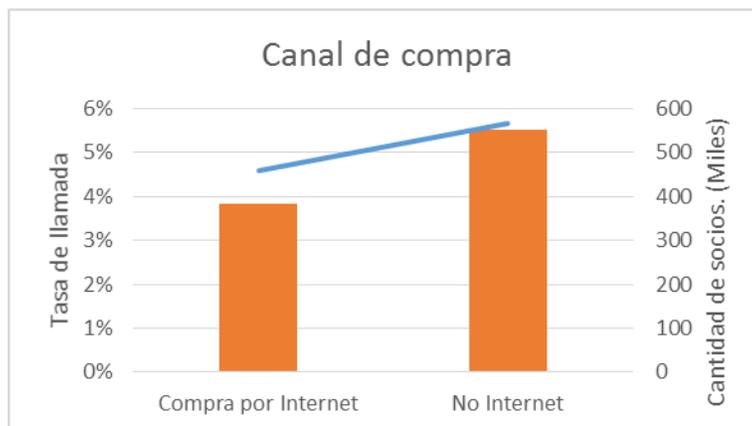


**Gráfico 13.- Tasa de llamada vs Cantidad de socios respecto a la antigüedad en el programa del cliente.**

*Fuente: Datos Empresa, Elaboración Empresa.*

El gráfico 13 nos muestra que la tasa de llamada con respecto a la antigüedad del socio en el programa tiene sentido, es decir, a mayor cantidad de socios existe una mayor probabilidad de llamar, pero además nos muestra claramente que a mayor antigüedad del socio la probabilidad de llamar es mayor, por lo cual es una variable de interés para estudiar en el modelo a desarrollar.

### 7.5.1.7 Canal de compra pasaje internet

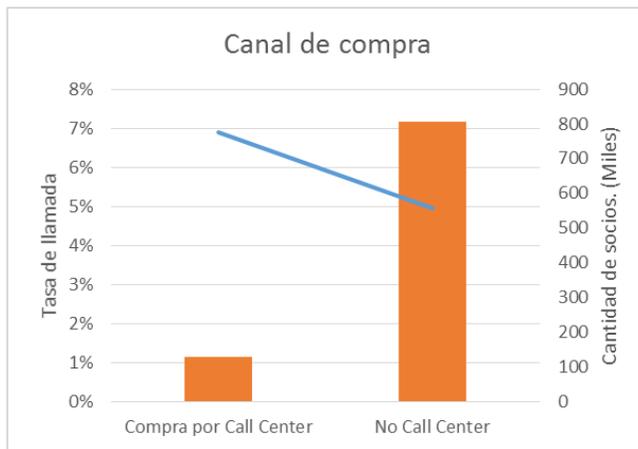


**Gráfico 14.- Tasa de llamada vs Cantidad de socios en la compra de pasaje vía internet.**

*Fuente: Datos Empresa, Elaboración Empresa.*

Tal como la antigüedad en el programa, el comportamiento de los socios que compran el pasaje vía internet tiene sentido, es decir, a mayor casos mayor probabilidad.

### 7.5.1.8 Canal de compra pasaje call center

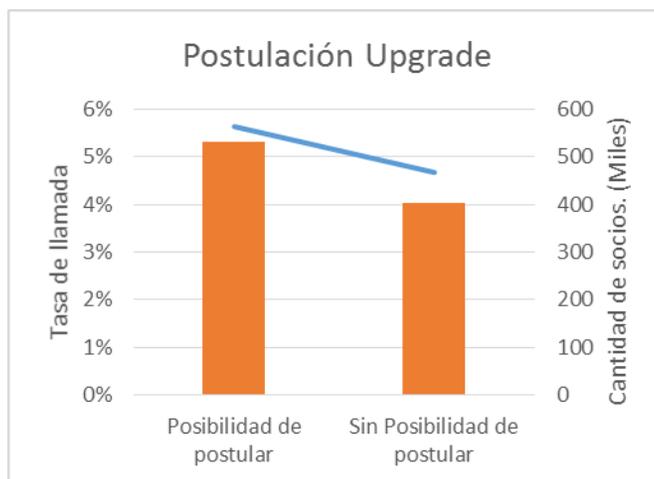


**Gráfico 15.- Tasa de llamada vs Cantidad de socios según la compra de pasaje vía call center.**

**Fuente: Datos Empresa, Elaboración Empresa.**

En el gráfico 15, nos encontramos con un comportamiento de interés, debido a que los socios que compran su pasaje vía el call center, son más propensos a llamar para realizar su “Check-in Upgrade”. Esto se puede deber a la confianza y/o costumbre al canal utilizado para realizar los diferentes tipos de trámites asociados a la línea aérea.

### 7.5.1.9 Opción de postular a Upgrade

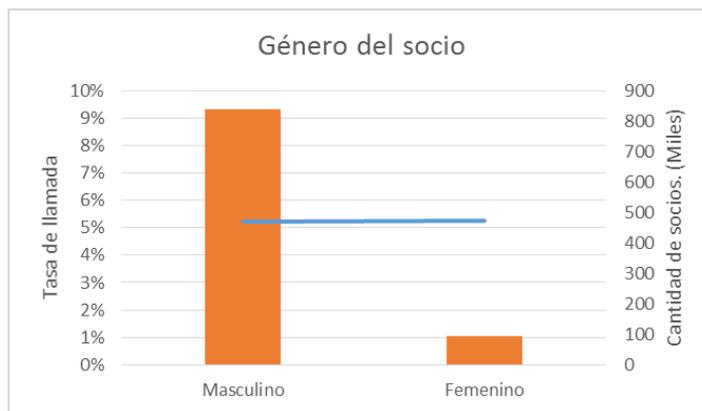


**Gráfico 16.- Tasa de llamada vs Cantidad de socios respecto los clientes que tenían opción de postular a upgrade.**

**Fuente: Datos Empresa, Elaboración Empresa.**

De este gráfico podemos observar que la tasa de llamada hace sentido con la cantidad de socios, pero también es interesante preguntarse porque la gente que tiene opción a upgrade llama al contact center para realizar su respectivo “Check-in Upgrade”.

### 7.5.1.10 Género del socio

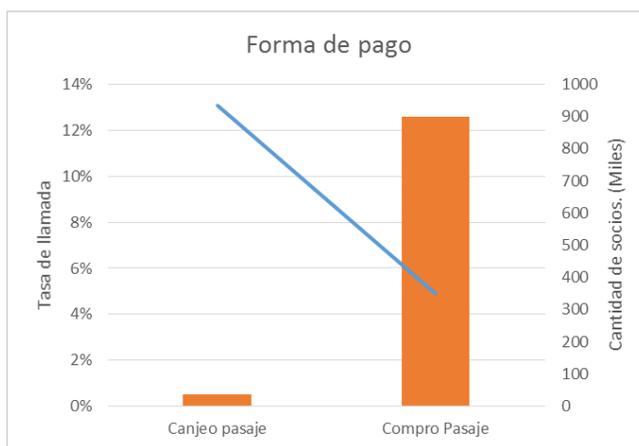


**Gráfico 17.- Tasa de llamada vs Cantidad de socios según el género.**

**Fuente: Datos Empresa, Elaboración Empresa.**

Del gráfico anterior se puede desprender que las tasas de llamadas tanto para las mujeres como para los hombres es la misma, por lo que no genera un gran interés por estudiar dicha variable.

### 7.5.1.11 Socio canjea pasaje



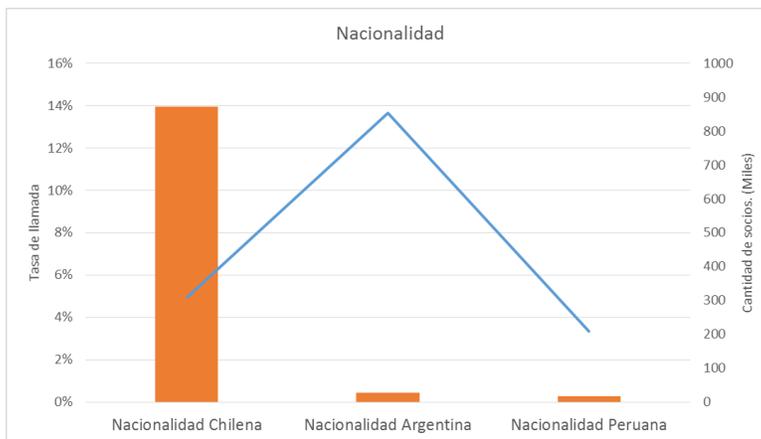
**Gráfico 18.- Tasa de llamada vs Cantidad de socios respecto de los clientes que canjearon su pasaje como forma de pago.**

**Fuente: Datos Empresa, Elaboración Empresa.**

Esta es una de las variables más interesantes a estudiar, debido a que es importante preguntarse ¿porque los socios que canjean se comportan tan distinto a los socios que

pagan por su pasaje? Por lo mismo será una variable relevante a mirar en el desarrollo del modelo y sus futuras conclusiones.

### 7.5.1.12 País de nacimiento del socio

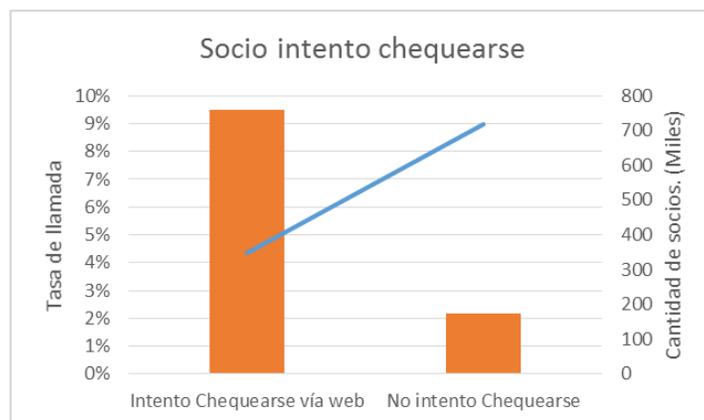


**Gráfico 19.- Tasa de llamada vs Cantidad de socios, para la nacionalidad de los socios.**

**Fuente: Datos Empresa, Elaboración Empresa.**

Del gráfico 19, podemos observar que los socios que son de Argentina, son más propensos a llamar para realizar su “Check-in Upgrade”, lo que da luces de que el país de origen podría llegar a ser una variable relevante a observar. Además, nos muestra que los socios de origen chileno no son tan propensos a llamar al contact center respecto a los demás países. Finalmente, los socios de nacionalidad peruana tienen un análisis similar al país argentino, pero en menor magnitud.

### 7.5.1.13 Socio intento realizar check-in por la web

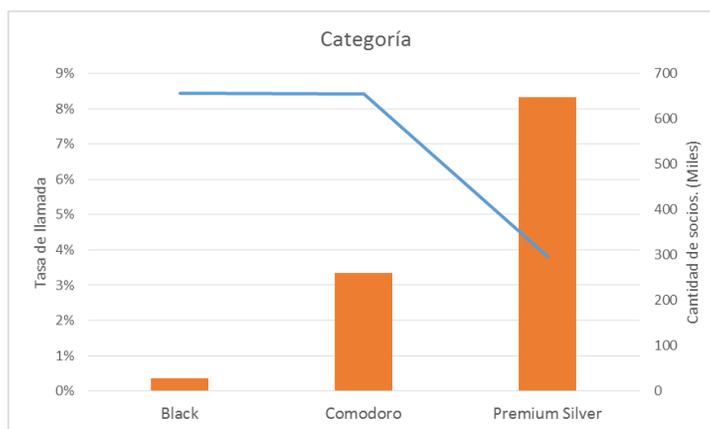


**Gráfico 20.- Tasa de llamada vs Cantidad de socios, respecto a los intentos de chequeo en la web.**

**Fuente: Datos Empresa, Elaboración Empresa.**

Esta variable nos muestra la cantidad de clientes que intentaron realizar el check-in vía web, no distinguiendo entre sí tuvo éxito o no. Por lo mismo nos sirve sólo como un indicador de que, los que no lo intentan tienen una alta probabilidad de llamar.

### 7.5.1.14 Categoría del socio



**Gráfico 21.- Tasa de llamada vs Cantidad de socios según la categoría de los socios.**

*Fuente: Datos Empresa, Elaboración Empresa.*

El siguiente gráfico nos describe que la tasa de llamada es alta para los socios que son de categoría black, lo cual es interesante de estudiar, debido a que ellos son el tipo de cliente más valioso para el programa, por lo cual se deberían satisfacer en lo más posible sus necesidades. Además, los socios comodoros se comportan de manera similar, y siendo los segundos más importantes. Finalmente, los socios de categoría premium silver tienen una probabilidad menor a llamar al contact center.

Debido al análisis exploratorio podemos notar que las principales variables que harán que la probabilidad de llamada aumente en un socio son:

- Tipo de viaje.
- Categoría del socio.
- Forma de compra.
- País de nacionalidad.
- Canje.

Con el análisis descrito podemos generar las primeras hipótesis de porque los socios del programa están llamados al contact center, lo que se podrá avalar o rechazar con los

modelos que se desarrollen en la siguiente sección. Las principales hipótesis planteadas según el análisis exploratorio son:

H1	Los socios que canjeen su pasaje en vez de comprarlos, tendrán una mayor probabilidad de llamar al contact center del programa por motivos de “Check-in Upgrade”.
H2	Los socios que tengan viajes internacionales, tendrán una mayor probabilidad de llamar que los socios que sólo realicen viajes domésticos.
H3	Los socios que tengan escalas en sus vuelos, tendrán una mayor probabilidad de llamar que los socios que los socios que su vuelo es directo.
H4	Los socios que correspondan a categoría black y comodoro, tendrán una mayor probabilidad de llamar al contact center del programa.
H5	Los socios de nacionalidad argentina o chilena, tendrán una mayor probabilidad de llamar al contact center, que los socios de otros países para realizar su “Check-in Upgrade”.

*Tabla 7: Hipótesis generadas en base al análisis exploratorio.*

*Fuente: Elaboración Propia.*

En la siguiente sección, se desarrollarán los modelos pertinentes para corroborar las hipótesis mencionadas y poder identificar los perfiles de clientes más problemáticos que darán a lugar al diseño de acciones correspondientes.

## **7.6 DESARROLLO MODELOS DE PROPENSIÓN**

La siguiente etapa del presente trabajo, constará de realizar un modelo que permita calcular la probabilidad de la posible llamada de un cliente al contact center del programa, basado en las características y contexto del cliente que fue descrito anteriormente. Esto con el fin de poder encontrar cuales de las variables utilizadas son las más relevantes a la hora de formar o crear un perfil de cliente, el cual nos permitirá diseñar acciones acorde a ellos y permitan tener una alta posibilidad de evitar dicho llamado.

### **7.6.1 Modelo Logit**

Para esto se ocupará un modelo de regresión logística binario, debido al comportamiento de la variable dependiente, la cual puede describirse como:

$$y_n = \begin{cases} 1, & \text{si cliente } n \text{ llama} \\ 0, & \text{si cliente } n \text{ no llama} \end{cases}$$

Las variables dependientes corresponderán a las mencionadas en la sección 7.5.

Teniendo el resultado del modelo y comprobando su bondad de ajuste basado en distintas métricas, podremos a través de los parámetros estimados, ver cuáles son las variables que están causando un mayor impacto en la probabilidad de llamar al contact center, y es de aquí de donde se podrán concluir qué medidas tomar según el tipo de socio desde el área de operaciones.

Para poder validar lo anteriormente dicho, necesitamos verificar que nuestro modelo posee un buen poder predictivo, que las variables son relevantes tanto individual como grupalmente, que tienen un buen ajuste, etc., es decir, necesitamos evaluar una serie de métricas de ajustes para poder asegurar que el uso del modelo es correcto.

### **7.6.1.1 Tratamiento previo de variables para modelo logit**

A continuación se describirá como se llegó a las variables incluidas en el modelo. Para esto lo primero que se hizo fue estudiar el comportamiento de las variables continuas dentro del modelo, es decir, se quería evaluar si convenía introducirlas tal cual o había que hacer alguna transformación lineal, categorizarlas o simplemente no incluirlas por no ser significantes, todo esto antes de decidir qué modelo trabajar finalmente.

Como resultado de esto se obtuvo lo siguiente:

- Si introducimos todas las variables continuas sin ninguna transformación, resulta en que todas son significativas, pero el signo del  $\beta_{edad}$  no hace sentido según el análisis previamente hecho.
- Si hacemos la transformación del logaritmo de las variables, resulta que tanto la anticipación y la edad no hacen sentido con lo previamente estudiado.
- Si categorizamos las variables continuas en distintos rangos, se gana en interpretación del modelo, en el sentido de la anticipación y en significancia.

Además realizando una prueba previa de la regresión logística, nos pudimos dar cuenta de lo siguiente:

- El saldo no es significativo.
- El género no es significativo.
- Los niveles base para cada variable son: Anticipación entre 30 y 90 días, Edad mayor de 60 años, Categoría Premium Silver, Canal Otros y País Otros.

Por estas razones para el modelo final no se introducirán el saldo ni el género y además la variable antigüedad en el programa se transformará con el logaritmo de la misma

variable, debido a que de esta forma si hace sentido según el análisis previo. Las variables que se incluirán en el modelo se muestran en la tabla 8.

<b>Variables</b>	<b>Tipo</b>
Anticipación compra 5 o menos días	Categoría
Anticipación compra entre 6 y 10 días	Categoría
Anticipación compra entre 10 y 30 días	Categoría
Anticipación compra entre 30 y 90 días	Categoría
Anticipación compra más de 90 días	Categoría
Categoría Premium Silver	Categoría
Categoría Comodoro	Categoría
Categoría Black	Categoría
Edad de 30 o menos años	Categoría
Edad entre 30 y 40 años	Categoría
Edad entre 40 y 60 años	Categoría
Edad mayor a 60 años	Categoría
Canal de compra: Internet	Categoría
Canal de compra: Call Center	Categoría
Vuelo con escalas	Categoría
Cliente intento chequearse por web	Categoría
Vuelo internacional	Categoría
Nacionalidad Argentina	Categoría
Nacionalidad Chilena	Categoría
Nacionalidad Peruana	Categoría
Posibilidad de postular a upgrade	Categoría
Canjeó pasaje	Categoría
Ln(Antigüedad en el programa)	Continua

*Tabla 8: Variables explicativas para modelo logit*

*Fuente: Elaboración Propia*

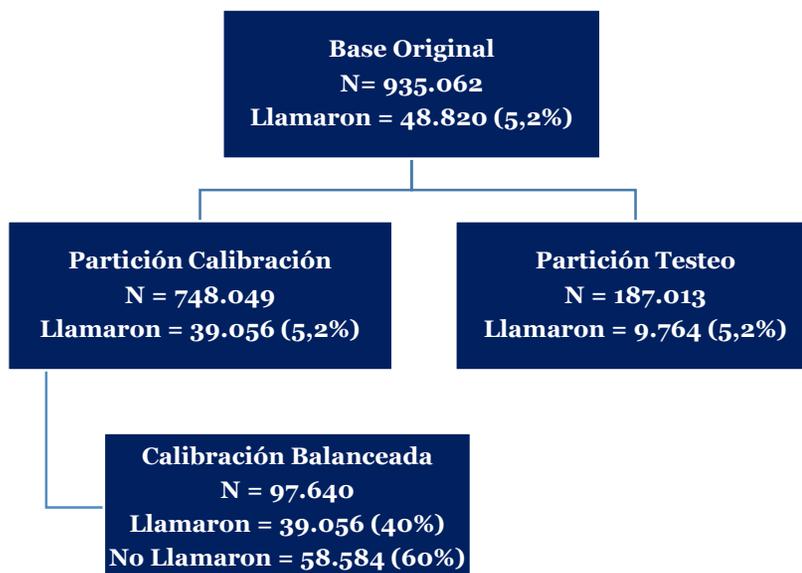
### **7.6.1.2 Balance de la muestra**

Antes de realizar el modelo definitivo, como se mencionó anteriormente, la base a utilizar sólo tiene un 5,2% de socios que llamaron (“1”), de lo cual se observa claramente que los datos están muy desbalanceados. Por esto mismo, una vez realizada la división total de la muestra en un 80% para la calibración y un 20% para el testeo, en la muestra de calibración se equilibrará la proporción entre “1” y “0”, de tal forma que el aprendizaje que tenga el modelo para los socios que llaman sea mejor y más rápido. Luego se comprobarán los resultados tanto en la muestra balanceada como en la que no, para ver si hay algún grado de sesgo al realizar este balanceo previo.

Como se mencionó, lo primero que se hizo fue dividir la muestra en un 80% para la calibración del modelo y un 20% para el testeo de éste. Cabe recalcar que ambas bases tendrán la misma proporción de clientes que llamaron respecto los que no. Esto se traduce en que en la muestra de calibración se tendrá un total de 748.049 datos, de los cuales sólo 39.056 corresponden a socios que llamaron al contact center. De igual forma la base de testeo tendrá 187.013, de los cuales sólo 9.764 corresponde a socios que llamaron. Como

se puede apreciar en ambas bases la proporción de clientes que llama al contact center es del 5,2%.

Luego se realizó el balanceo de la muestra de calibración, para lo cual hay dos formas posibles de realizarlo. Por un lado se pueden multiplicar los “1”, es decir, aumentar la cantidad de socios que llamaron ya sea, duplicándolos, triplicándolos, etc. y por otro, se pueden eliminar los “0”, es decir, eliminar socios que no llamaron. Para este caso, se optó por eliminar de forma aleatoria registros de socios que no llamaron. De esta forma la muestra quedó compuesta de un 40% de socios que llamaron y 60% de socios que no, lo que equivalió a 39.056 y 58.584 registros respectivamente. Todo este proceso se puede visualizar de manera más sencilla en la ilustración 2.



*Ilustración 2: Procesamiento de base de datos para trabajar el modelo logit.*

*Fuente: Elaboración Propia.*

Una vez con la base de datos resultante y con el análisis exploratorio antes descrito, se dispuso a realizar la regresión logística antes mencionada.

### 7.6.2 Modelo Logit 1

Antes de analizar cada variable, tenemos que ver si el modelo como conjunto es significativo, ya que si no lo fuera quedaría inmediatamente descartado para su uso como predictor. Para poder saber esto necesitamos fijarnos en el estadístico Chi-cuadrado y ver su significancia, lo que para este caso tenemos:

$$X^2_{20} = 16021,696$$

$$p - \text{valor} = 0,000$$

Por el “p-valor”, resulta que se rechaza la hipótesis nula  $H_0: \beta_1 = \dots = \beta_k = 0$ , por lo que nuestro modelo es significativo en su conjunto. Esto significa que en una primera instancia, antes de analizar las métricas de ajuste, podemos usar este modelo para predecir la llamada de un socio.

Variable		Beta	Significancia	Exp(Beta)
Anticipación de Compra	5 o menos días	0,235	0	1,265
	entre 6 y 10 días	0,066	0,008	1,068
	Entre 10 y 30 días	-0,007	0,734	0,993
	Más de 90 días	0,252	0	1,287
Categoría	Categoría Black	0,47	0	1,6
	Categoría Comodoro	0,775	0	2,107
Edad	30 o menos años	0,1	0	1,105
	Entre 30 y 40 años	0,177	0	1,194
	Entre 40 y 60 años	0,071	0,012	1,074
Canal de Compra	Internet	0,113	0	1,12
	Call center	0,044	0,035	1,045
Nacionalidad	Argentina	0,694	0	2,001
	Chile	0,137	0	1,146
	Perú	-0,663	0	0,515
Posibilidad de postular a Upgrade		0,201	0,003	1,223
Canjeo pasaje		0,646	0	1,908
Ln(Antigüedad en el programa)		0,53	0	1,699
Vuelo con escalas		0,488	0	1,63
Cliente intento chequearse por web		-0,532	0	0,587
Vuelo Internacional		0,662	0	1,939
Constante		-2,495	0	0,083

*Tabla 9: Resultado Modelo 1 logit*

*Fuente: Elaboración Propia.*

Para analizar las variables y su influencia en el modelo debemos observar la tabla 9, de donde a primera vista, si el socio pertenece a la categoría comodoro, su viaje presenta escalas, es de procedencia argentina, en vez de comprar el pasaje lo canjeo, su viaje es internacional y a medida que lleva más años en el programa LANPASS, su probabilidad

de llamar al contact center será más alta. Ahora bien, debido a que estos  $\beta$ 's no están estandarizados no podemos asegurar que entre más alto el beta mayor es el efecto que tiene sobre la probabilidad.

Para esto último, nos tenemos que fijar en la columna “Exp(Beta)” de la tabla 9, la cual se interpreta de la siguiente manera, por ejemplo para la variable de si el viaje es internacional, la cual tiene un valor de  $Exp(Beta) = 1,939$ , quiere decir que el ser un viaje internacional aumenta la razón de llamado al contact center 1,939 veces. Recordar lo que significa razón: es la frecuencia que representa la ocurrencia de un suceso sobre la frecuencia de su no ocurrencia (más conocida como “Odds Ratio”). Ahora, cuando el valor “Exp(Beta)” es negativa significa que dicha razón disminuye. También con esta columna podemos ver, que variables son las que más influyen en la probabilidad de llamado de un socio tanto positiva como negativamente. Para poder realizar esto necesitamos dejar los valores menores que 1 en la misma escala que los demás por lo que es necesario realizar la siguiente transformación:

$$Si \text{Exp}(Beta) < 1, \quad \text{Exp}(Beta)' = \frac{1}{\text{Exp}(Beta)}$$

Como se puede observar sólo las variables “Si el cliente intento chequearse por web y Nacionalidad peruana” son menores a uno, por lo que sus valores quedan como 1,704 y 1,942 respectivamente. Por lo tanto si realizamos un top 10 de variables más importantes, donde (+) es que la variable aumenta la probabilidad y (-) la disminuye, tenemos lo siguiente:

1. Categoría Comodoro (+).
2. Nacionalidad Argentina (+).
3. Nacionalidad Peruana (-).
4. Vuelo Internacional (+).
5. Canjeo pasaje (+).
6. Cliente intento chequearse por web (-).
7. Antigüedad en el programa (+).
8. Vuelo con escalas (+).
9. Categoría Black (+).
10. Anticipación compra más de 90 días (+).

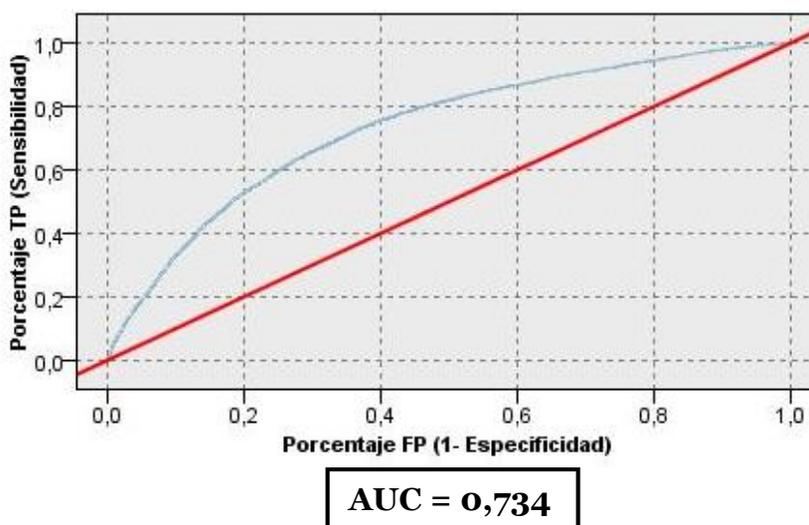
Por otro lado, en el modelo presentado en la tabla 9, podemos notar que el signo de los  $\beta$ 's correspondientes hacen sentido según lo visto en la sección 7.5.1. También, que la única variable no significativa es si la persona compra su pasaje con una anticipación entre 10 y 30 días. Por otro lado, se confirma que las variables analizadas son las que más influyen en la probabilidad de que un socio llame al contact center de la aerolínea, como lo son el tipo de viaje, la forma de pago, la categoría y la nacionalidad del socio.

Para evaluar el desempeño del modelo, se evaluarán las métricas de ajustes mencionadas en la sección 6.4 tanto en la base balanceada como en la base de testeo. Luego para ver si existe algún tipo de sesgo, se compararan ambas métricas en las distintas bases tanto balanceadas como no.

En primer lugar se muestra el gráfico 22, en el cual se observa la curva ROC para la regresión logística del modelo analizado anteriormente, para la partición de calibración, en donde la recta de color rojo representa la clasificación al azar y la azul corresponde a la curva ROC propiamente tal. Además se puede ver que la detección de verdaderos positivos es mayor a la de falsos positivos, es decir, la capacidad del modelo de clasificar a un cliente que llamo cómo llamo, es mayor que clasificar a un cliente como no llamo, que en verdad haya llamado.

También se puede apreciar el valor AUC (área bajo la curva), el cual asciende hasta 0,734, lo que significa que en la presencia de dos individuos, el modelo tiene un 73,4% de clasificarlos a cada uno correctamente.

Finalmente, con el análisis de este gráfico, podemos decir que el modelo es un buen predictor y es mucho mejor que realizar una clasificación totalmente al azar, es decir, se gana en la predicción de socios que llamaran al contact center. Sin embargo, falta corroborar el modelo en la base de testeo para poder confirmar la validez del modelo.

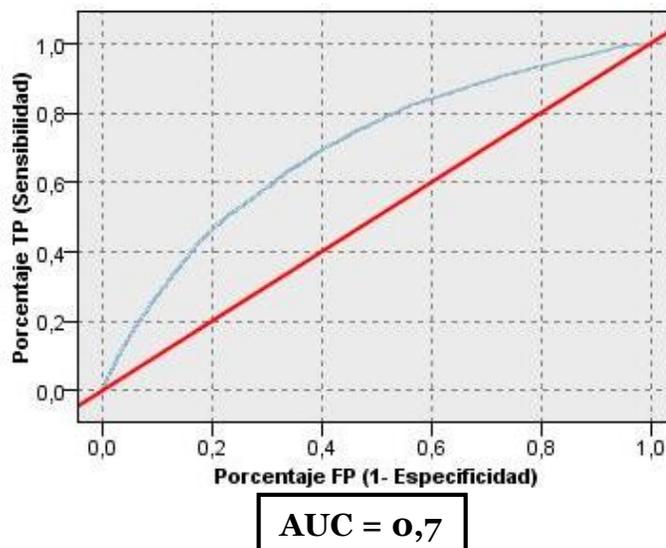


*Gráfico 22: Curva ROC de la regresión logística para el modelo de la tabla 6 en la partición de calibración.*

*Fuente: Elaboración Propia.*

A diferencia del gráfico 22, en el gráfico 23 se muestra la curva ROC del mismo modelo pero sobre la base de testeo. De aquí se puede observar que el área bajo la curva (AUC) en

comparación a la base de calibración baja en cuanto a magnitud (0,734 a 0,7), es decir, sólo un 3% menos de probabilidad para identificar correctamente si el socio llamó o no. Además podemos ver que si vemos dos puntos cualesquiera en ambos gráficos la diferencia no es significativa, por ejemplo, si vemos cuando  $1 - \text{Especificidad}$  es 0,2, vemos que en la partición de calibración la sensibilidad es aproximadamente 0,56, en cambio en la de testeo es 0,46. Con este resultado podemos ratificar que el modelo es aceptable, ya que no hay muchas diferencias entre los resultados de la partición de calibración y testeo.



**Gráfico 23: Curva ROC para regresión logística para la partición de testeo.**

*Fuente: Elaboración propia.*

Si siguiendo con la evaluación del desempeño del modelo, se pueden observar en la tabla 10, para los distintos puntos de corte de probabilidad (columna “PROB”), la cantidad de socios tanto que llaman como que no, el total de socios, la precisión (calculada como la cantidad de socios que llaman sobre el total) y la precisión acumulada (calculada como la suma acumulada del total de socios sobre el total), todo en la base de testeo.

PROB	Y=0	Y=1	Total	%PRECISION (Y=1/TOTAL)	%PRECISION ACUM
80,0%	1800	390	2190	17,81%	17,81%
70,0%	6962	1135	8097	14,02%	14,82%
60,0%	11535	1453	12988	11,19%	12,79%
50,0%	16219	1635	17854	9,16%	11,22%
40,0%	24928	1646	26574	6,19%	9,24%
30,0%	30265	1459	31724	4,60%	7,76%
20,0%	52390	1468	53858	2,73%	5,99%
10,0%	31411	567	31978	1,77%	5,26%
9,0%	1046	10	1056	0,95%	5,24%
8,0%	369	1	370	0,27%	5,23%
7,0%	190	0	190	0,00%	5,22%

6,0%	99	0	99	0,00%	5,22%
5,0%	26	0	26	0,00%	5,22%
4,0%	3	0	3	0,00%	5,22%
3,0%	0	0	0	0,00%	5,22%
2,0%	1	0	1	0,00%	5,22%
1,0%	3	0	3	0,00%	5,22%
0,0%	2	0	2	0,00%	5,22%
Total	177249	9764	187013	5,22%	

**Tabla 10: Precisión para los distintos puntos de corte para el modelo 1 en la partición de testeo.**

**Fuente: Elaboración Propia.**

De esta tabla se pueden rescatar dos resultados importantes. El primero es que podemos observar la cantidad de personas que son clasificadas que llaman y que no dependiendo del corte de probabilidad. Esto es muy importante porque a la hora de diseñar las acciones y ver a que socios se afectara, se podrá estudiar económicamente hasta qué punto será conveniente para la empresa realizar dicha acción. En segundo lugar, el resultado tanto de la precisión como el de la precisión acumulada, nos dan una intuición de que tan bueno es el modelo, lo que al parecer según esta métrica no pareciera ser del todo bueno, ya que sólo un 18% es clasificado de forma correcta si lo observado fue llamar, es decir, bajo esta métrica el modelo no sería muy recomendable para su uso.

Por otro lado, podemos observar la matriz de confusión en la tabla 11, donde lo más importante, es el porcentaje correcto que pronostica nuestro modelo. De aquí el dato que más importa es la correcta clasificación de los socios que llaman, es decir, con un criterio de que el socio llamará si la probabilidad es superior al 50%, el 47% de los socios son clasificados correctamente.

Modelo 1 Logit		Observado	
		0	1
Pronostico	0	140733	5151
	1	36516	4613
%Correcto		79,40%	47,24%

**Tabla 11: Matriz de confusión modelo 1 en partición de testeo con datos balanceados.**

**Fuente: Elaboración Propia.**

De la matriz de confusión anterior, pueden calcularse las distintas métricas como: **Accuracy**, **Especificidad**, **Sensibilidad** y **Precisión**, y además se muestran el **LIFT** y  $R^2_{Nalgerkerke}$  obteniendo lo siguiente:

$$\begin{aligned}
 Accuracy_{Modelo1} &= 77,72\% \\
 Especificidad_{Modelo1} &= 79,40\% \\
 Sensibilidad_{Modelo1} &= 47,24\% \\
 LIFT_{Modelo1} &= 1,52 \\
 Precisión_{Modelo1} &= 11,22\% \\
 R^2_{Nalgerkerke} &= 0,205
 \end{aligned}$$

Esto nos quiere decir que aproximadamente un 78% de los casos verdaderos son clasificados correctamente por el modelo, lo que es bastante aceptable para un modelo de propensión. Además lo ideal es tener tanto la sensibilidad como la especificidad alta, y como se ve en los resultados la especificidad es de un 79,4% y la sensibilidad 47,24%, lo que hace bastante aceptable al modelo respecto a estas medidas.

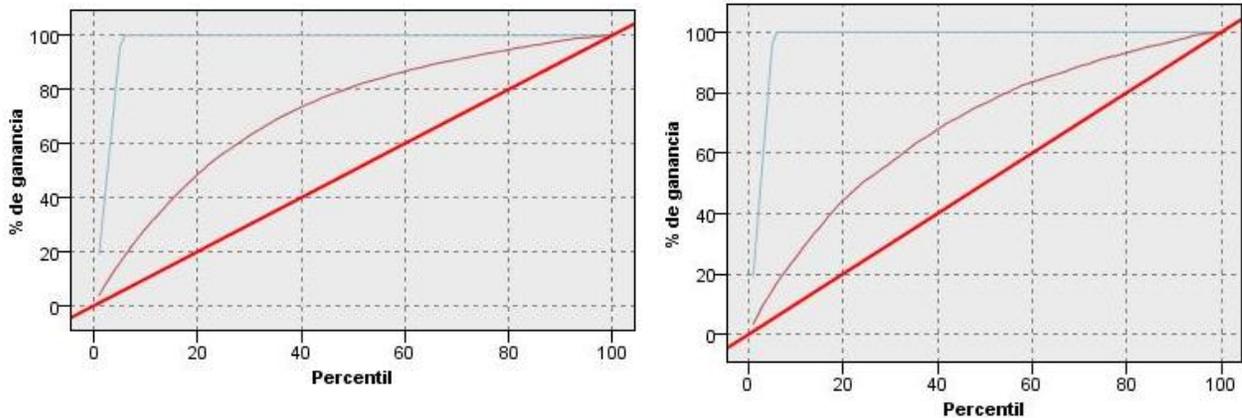
En la tabla 12, se puede observar al igual que en la tabla 10 los distintos puntos de corte de probabilidad, pero ahora con su respectiva sensibilidad asociada y el porcentaje de casos acumulados. De ésta se pueden interpretar dos resultados importantes. El primero nos muestra cual es la sensibilidad del modelo dependiendo del punto de corte de probabilidad, por ejemplo a un nivel de corte del 40% tenemos un sensibilidad del 64,1%, esto quiere decir que el modelo clasifica correctamente un 64,1% de los socios que en lo observado llaman, con un criterio de que sobre un 40% de probabilidad el cliente llamará. En segundo lugar y muy importante es lo que llamamos *ganancia de información*, esto quiere decir con cuanta información el modelo es capaz de recuperar a los socios que llaman. Esto último se puede observar relacionando las columnas de “SENSIBILIDAD” y “%CASOS ACUM”, donde por ejemplo si nos fijamos con un 36,2% de los socios más propensos a llamar, estamos obteniendo una sensibilidad cerca del 64,1%, es decir, que con un tercio de los datos estamos obteniendo un alto porcentaje de clasificación correcta. Por esto mismo, podemos decir que la capacidad de ganancia de información del modelo es alta, ya que con poca información es capaz de recuperar la mayoría de los casos importantes (socios que llaman).

PROB	SENSIBILIDAD	% CASOS ACUM
80,0%	4,0%	1,17%
70,0%	15,6%	5,50%
60,0%	30,5%	12,45%
50,0%	47,2%	21,99%
40,0%	64,1%	36,20%
30,0%	79,0%	53,17%
20,0%	94,1%	81,96%
10,0%	99,9%	99,06%
9,0%	100,0%	99,63%
8,0%	100,0%	99,83%
7,0%	100,0%	99,93%
6,0%	100,0%	99,98%
5,0%	100,0%	100,00%
4,0%	100,0%	100,00%
3,0%	100,0%	100,00%
2,0%	100,0%	100,00%
1,0%	100,0%	100,00%
0,0%	100,0%	100,00%

**Tabla 12: Ganancia de información modelo 1 logit**

**Fuente: Elaboración Propia.**

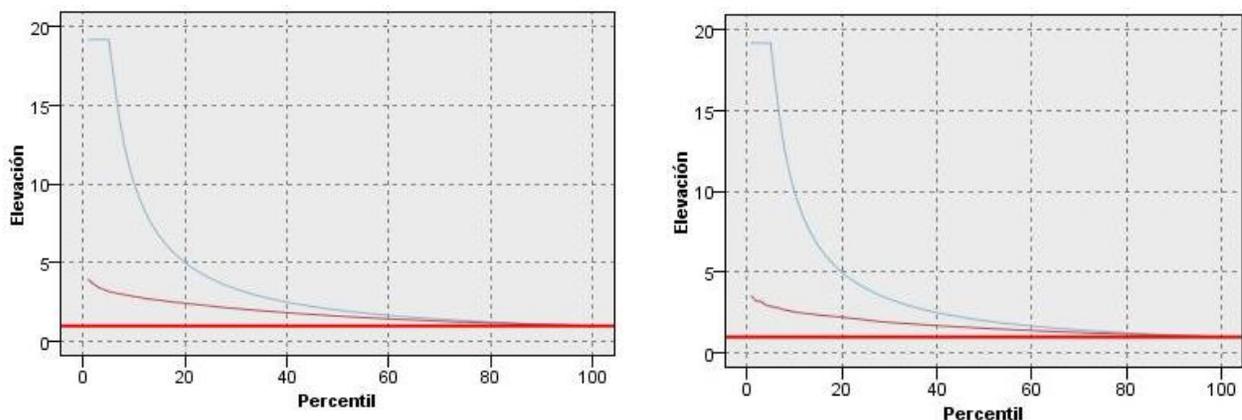
La tabla 12 se puede observar gráficamente en el gráfico 24, donde se ratifica que el ajuste del modelo es bueno, ya que logra detectar el 60% de los socios que llaman al contact center con el 30% de los socios más probable que llamen. Además la línea roja representa la selección aleatoria y la azul la ganancia perfecta.



**Gráfico 24:** Curva de ganancia modelo 1 logit balanceado para partición de calibración (izquierda) y testeo (derecha).

*Fuente: Elaboración Propia.*

Otra métrica interesante de evaluar en el desempeño del modelo es lo que llamamos curva de elevación o lift, la cual se puede observar en el gráfico 25 tanto para la partición de calibración como para la de testeo. Este gráfico explica cuanto más efectivo es tomar una acción sobre un segmento de socios específico que sobre un segmento al azar. Por ejemplo si vemos el gráfico de la derecha, la curva de color rojo claro nos dice que para el 30% de socios con mayor probabilidad de llamar al contact center, el efecto de tomar una acción será 2,1 veces mayor que al no usar el modelo (aleatoriedad), lo que nos dice que el uso del modelo es bueno y recomendable.



**Gráfico 25:** Curva de elevación para modelo 1 logit balanceado en partición de calibración (izquierda) y testeo (derecha).

*Fuente: Elaboración Propia.*

Finalmente, a modo de resumen para este primer modelo, podemos observar que es bastante aceptable para la predicción de los socios que llaman según distintas métricas, como lo son la curva ROC, accuracy, ganancia de información y curva de elevación. Por lo mismo podemos postularlo como un candidato a la hora de seleccionar un modelo que nos presente los distintos perfiles de clientes más propensos a llamar al contact center de la aerolínea.

A continuación se presenta el segundo modelo de regresión logística, donde la única diferencia es que se incluyen ciertos cruces de variables interesantes para el negocio. Cabe recalcar que no se quiso incluir todos los cruces posibles debido a un posible sobreajuste del modelo.

### 7.6.3 Modelo Logit 2

Como se mencionó anteriormente la única diferencia del modelo 2 con el modelo 1 es que se agregan ciertos cruces de variables interesantes para la empresa, pero se evaluará si su inclusión agrega valor al modelo. Los cruces que se introdujeron para este modelo son:

- Anticipación de compra x Canje (4).
- Anticipación de compra x Internacional (4).

Primero que todo cabe mencionar que el estadístico Chi-cuadrado para este modelo tiene el siguiente valor:  $X^2_{28} = 16113,745$ , lo cual hace que el modelo en su conjunto sea significativo. Con esto podemos realizar el análisis de cada una de las variables, donde lo principal a resaltar son la significancia y sentido de las variables que se incluyeron respecto al primer modelo, lo que puede observarse en la tabla 13. El resto de las variables se pueden encontrar en la sección 10 en el anexo 7.

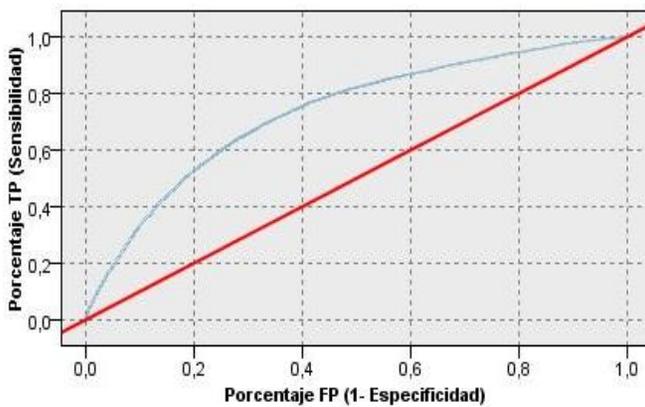
Variable	Beta	Significancia	Exp(Beta)
Anticipación entre 6 y 10 días x Canje	-0,296	0,039	0,744
Anticipación entre 10 y 30 días x Canje	0,164	0,092	1,178
Anticipación entre 30 y 90 días x Canje	0,108	0,193	1,114
Anticipación de 5 o menos días x Inter	-0,154	0,007	0,857
Anticipación entre 6 y 10 días x Inter	-0,053	0,394	0,949
Anticipación entre 10 y 30 días x Inter	0,051	0,334	1,053
Anticipación entre 30 y 90 días x Inter	0,141	0,009	1,151
Constante	-2,537	0	0,079

*Tabla 13: Interacción de variables incluidas en modelo 2 logit.*

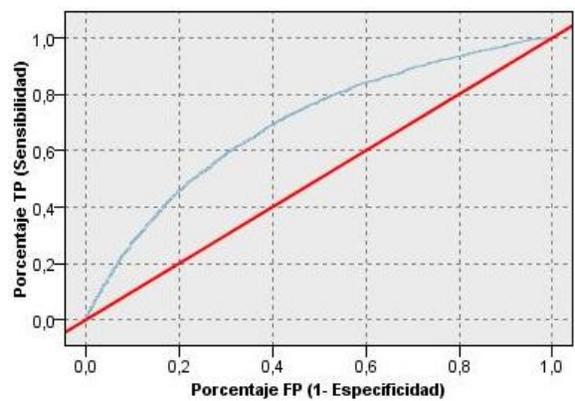
*Fuente: Elaboración Propia.*

De aquí se pueden concluir dos cosas importantes, la primera es que los socios que canjean su pasaje con menor anticipación tienen mayor probabilidad de llamar al contact center y en segundo lugar, es que los socios que anticipan su compra y su viaje sea internacional, su probabilidad de llamar aumenta. Ahora bien, aunque se pueda concluir lo recién mencionado, hay que notar que 4 de las 8 variables incluidas son no significativas, por lo que en general su inclusión no será de mucho valor comparado al modelo 1.

Una vez dicho esto, si se realiza la misma evaluación que para el modelo 1, se puede observar según los gráficos 26, 27 y 28, los indicadores tales como curva ROC, ganancia de información y curva de elevación no cambian de manera significativa, es más hay algunos de ellos como la ganancia y la elevación que empeoran con respecto a los resultados del modelo 1. El resto de los indicadores del modelo 2 se encuentran en la sección 10 en anexo 8, 9 y 10.



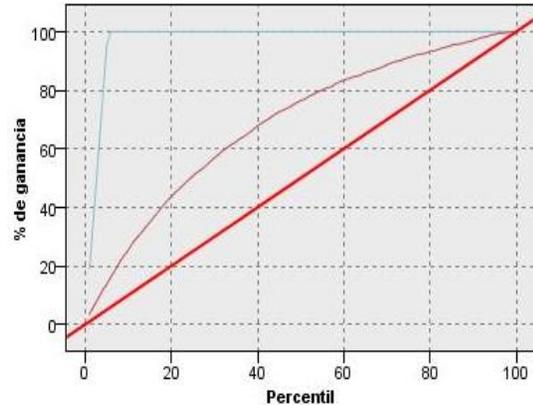
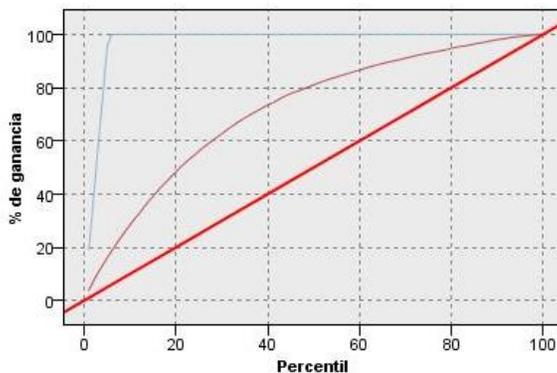
**AUC = 0,734**



**AUC = 0,695**

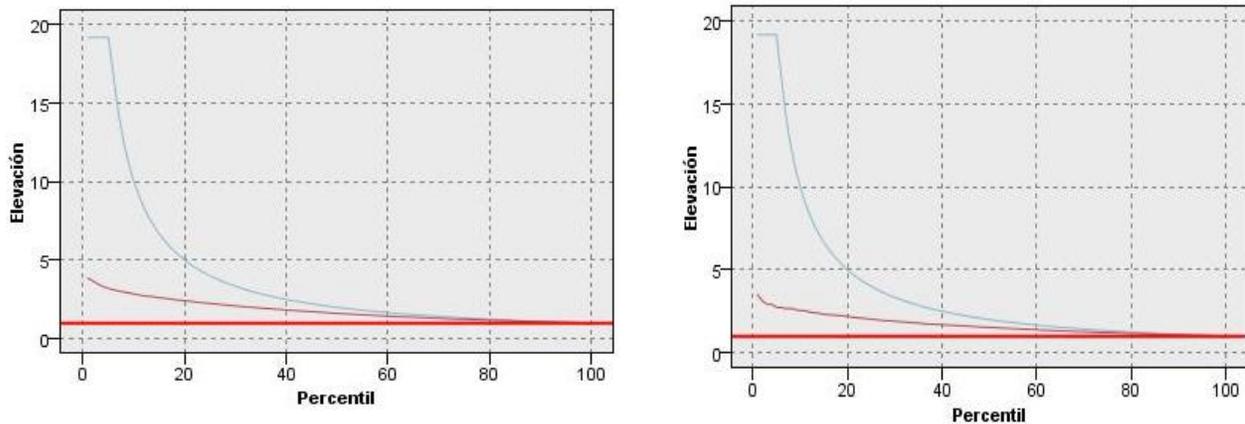
**Gráfico 26: Curva ROC modelo 2 logit balanceado para partición de calibración (izquierda) y testeo (derecha).**

*Fuente: Elaboración Propia.*



**Gráfico 27: Curva de ganancia modelo 2 logit balanceado en partición de calibración (izquierda) y testeo (derecha).**

*Fuente: Elaboración Propia.*



**Gráfico 28:** Curva de elevación para modelo 2 logit balanceado en partición de calibración (izquierda) y testeo (derecha).

*Fuente: Elaboración Propia.*

Luego de ver los dos modelos descritos anteriormente y antes de realizar el análisis en base a árboles de decisión, se optará por usar el modelo 1 para identificar a los clientes que tienen alta probabilidad de llamar al contact center de una aerolínea, debido a la comparación de todas las métricas, que en general son mejores en el modelo 1 y además claramente la inclusión del cruce de variables no aporta gran valor, sino más bien aumenta la posibilidad de un sobreajuste.

Como se mencionó en un principio, al balancear la base de datos, una vez con los resultados de los modelos había que verificar que éstos no estuvieran sesgados, es decir, que las métricas de desempeño no se sobre ajustaran dando un modelo casi perfecto en comparación a los resultados con los datos no balanceado. Por esto mismo surge la necesidad de comparar los resultados recién descritos con los resultados no balanceados, los cuales se pueden observar en la sección 10 en los anexos 11 y 12.

#### 7.6.4 Comparación modelo balanceado y no balanceado

A continuación, se compararán los resultados entre modelos balanceados y no balanceados, donde se puede observar que las curvas de ganancia, ROC y de elevación no presentan una gran diferencia, sino que las mejoras más importantes viene reflejada en la sensibilidad del modelo y el  $R^2_{nalgerkerke}$ , donde el primero es uno de los puntos más importantes para este caso, ya que nos interesa poder identificar de mejor forma a los socios que llaman al contact center. Si bien, como se puede observar en la tabla 14, indicadores como la accuracy o especificidad bajan en un porcentaje, esta baja no hace que el modelo sea malo para poder predecir el comportamiento estudiado. Además esto nos dice que, por un lado los resultados no están sesgados, ya que si fuera así nos encontraríamos con todas las métricas completamente distintas y los resultados del modelo balanceado serían cercanos a lo perfecto, y por otro que el balancear los datos hizo

que el modelo tuviera un mejor aprendizaje de los socios que llaman al contact center (“1”), lo cual era lo que se buscaba con esta acción.

Por otro lado una métrica importante para argumentar el uso del balanceo del modelo es el análisis de la precisión de éste, ya que se debe mostrar que tan bueno es el pronóstico con respecto a la realidad, es decir, cuántos de los socios pronosticados en verdad llaman. De la tabla 14, se puede observar que la precisión del modelo no balanceado es mayor al balanceado, lo que se debe a la cantidad de datos en cada una de las muestras. Ahora bien, esta diferencia de 2,5%, no presenta ser tan significativa para descartar el modelo balanceado tanto para ocuparlo para la elección de perfiles de clientes, como para definirlo como un buen modelo. Además esta diferencia no es comparable con el gran aumento que se da en la sensibilidad del modelo balanceado (30%).

Finalmente, debido a los argumentos anteriores, se optara por utilizar el modelo balanceado, principalmente por su gran aumento en la sensibilidad, sacrificando sólo una pequeña baja de 2,5% en la precisión.

Las principales diferencias del modelo 1, medidas en la base de testeo se ven reflejadas en la tabla 14.

Métrica	No Balanceado	Balanceado
Accuracy	89,72%	77,72%
Sensibilidad	18,47%	47,24%
Especificidad	93,64%	79,4%
Precisión	13,79%	11,22%
$R^2_{\text{naïve}}/gerkerke$	0,99	0,2

*Tabla 14: Diferencias de métricas entre modelo 1 balanceado y no balanceado.*

*Fuente: Elaboración Propia.*

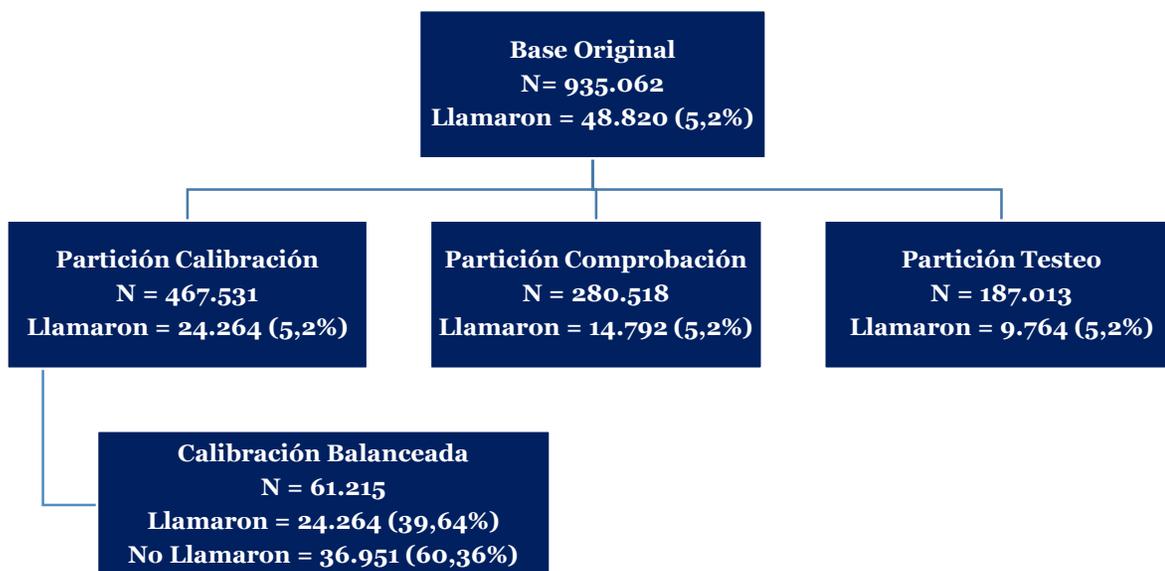
## 7.7 Modelo árbol de decisión

El modelo de árboles de decisión se realiza por tres motivos principales:

1. **Comparación de modelos:** No basta sólo con realizar un modelo, logit en este caso, ya que puede existir otro que nos dé un mejor resultado, y además es bueno verificar que ambos entregan sentidos similares del problema.
2. **Facilidad de comprensión:** A diferencia del modelo logit, los árboles de decisión son muy fáciles de interpretar y leer.
3. **Construcción perfiles de clientes:** El resultado del modelo de árboles de decisión forma automáticamente los perfiles más propensos a llamar al contact

center, junto con su probabilidad y total de llamados posibles, lo que es uno de los objetivos principales de este proyecto.

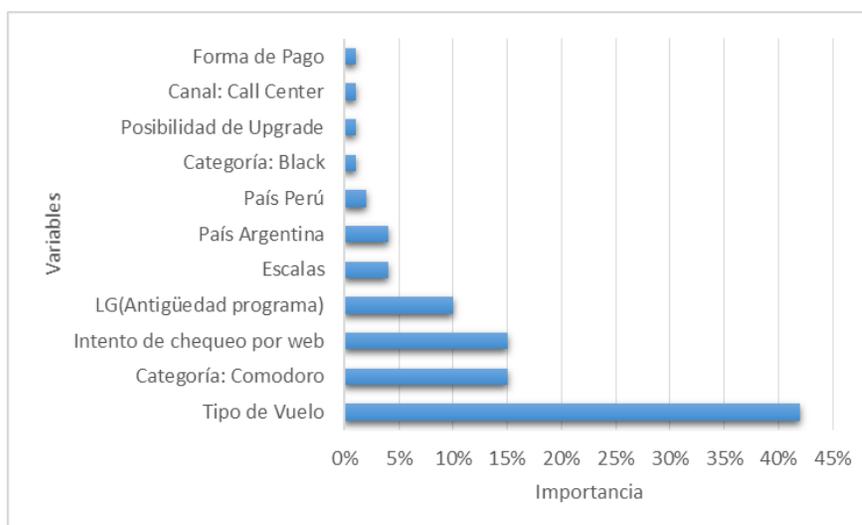
Una vez dicho esto, el modelo que se realizó fue un árbol de tipo CHAID exhaustivo, de una profundidad 5, con las mismas variables utilizadas para el modelo 1 logit y con variable objetivo si el cliente llama. El árbol total contó con 86 nodos y se pudo para que éstos no tuvieran muy pocos casos. Además, para este modelo se dividió la muestra en un 50% de calibración, un 30% de comprobación y un 20% de testeo. Al igual que en el modelo logit, dentro de la partición de calibración se hará el balanceo de la muestra de tal forma que los que llaman representen un 40% y los que no un 60%, tal como se puede observar en la ilustración 3.



*Ilustración 3: Procesamiento de base de datos para trabajar el modelo logit.*

*Fuente: Elaboración Propia.*

El primer resultado, se puede observar en el gráfico 29, donde se muestra la importancia de las variables para predecir la variable objetivo, es decir, el tipo de vuelo (internacional o no) es la variable más importante para predecir si el cliente llamará o no al contact center con un 42%. Le sigue en segundo lugar si la categoría del socio es comodoro, en tercer lugar si el socio intento chequearse por la web y en cuarto el logaritmo de la antigüedad en el programa. En el mismo gráfico se pueden observar el resto de las variables.



**Gráfico 29: Importancia de variables para el modelo de árboles de decisión.**

**Fuente: Elaboración Propia.**

La importancia de las variables independientes se calcula sobre la muestra de comprobación y en su conjunto suman 1 en total (100%). Además, la importancia de las variables no se relaciona con las métricas de ajustes del modelo, sino que es una medida de cuanto afecta dicha variable en la predicción de la variable objetivo.

Primero que todo, hay que mencionar que el análisis del árbol se realizará sólo en los nodos más significativos, debido a su impacto en lo que busca este trabajo. Además, cada nodo está compuesto por 3 columnas, donde la primera identifica la variable objetivo (No llama y Llama), la segunda indica la probabilidad de llamar (o no llamar) del tipo de perfil generado y la tercera corresponde al total de individuos con ese perfil correspondiente a cada elección. También se podrá observar la participación total del tipo de perfil respecto a la muestra. Por ejemplo en la ilustración 4 se observa el nodo Padre, donde un 39,64% llama, lo cual corresponde a un total de 24.264 socios, los que no llaman son un 60,30% (36.951 socios) y el total de socios son 61.215 el cual representa un 100% de la base de calibración.

Nodo 0: Variable Objetivo		
No Llama	60,36%	36951
Llama	39,64%	24264
Total	100,00%	61215

**Ilustración 4: Nodo 0 modelo árbol CHAID: variable objetivo.**

**Fuente: Elaboración Propia.**

Una vez dicho esto, en la ilustración 5, se puede observar que el primer nivel del modelo a través del tipo árbol CHAID exhaustivo es el tipo de vuelo (doméstico o

internacional), la cual como se vio en el grafico anterior, es la variable más importante para la predicción.

Nodo 0: Variable Objetivo		
No Llama	60,36%	36951
Llama	39,64%	24264
Total	100,00%	61215

Nodo 1: Vuelo Domestico		
No Llama	69,47%	28700
Llama	30,53%	12611
Total	67,48%	41311

Nodo 2: Vuelo Internacional		
No Llama	41,45%	8251
Llama	58,55%	11653
Total	32,52%	19904

**Ilustración 5: Primer nivel modelo árbol CHAID exhaustivo: Abertura por tipo de vuelo.**

*Fuente: Elaboración Propia.*

De este primer nivel podemos observar que el vuelo domestico explica el 67,48% de los socios que vuelan. Ahora bien, que el vuelo sea internacional y represente un 32,52% del total no deja de ser importante, ya que este último hace que los socios tengan un 28,02% más de probabilidad de llamar al contact center para realizar su “Check-in Upgrade” con respecto a los de tipo doméstico, es más, aunque represente una participación menor, los 11.653 llamados que contemplan los viajes internacionales, deja desde ya, un lugar para focalizar esfuerzos en este tipo de socios, pudiendo generar un gran impacto en los costos asociados al servicios del contact center.

Si continuamos con el segundo nivel a través del nodo 2, tipo de vuelo internacional, se puede observar en la ilustración 6, que la siguiente abertura es debido a la variable de si el socio tiene posibilidad de postular a upgrade.

Nodo 2: Vuelo Internacional		
No Llama	41,45%	8251
Llama	58,55%	11653
Total	32,52%	19904

Nodo 5: No UPG		
No Llama	49,94%	3967
Llama	50,06%	3976
Total	12,98%	7943

Nodo 6: UPG		
No Llama	35,86%	4284
Llama	64,18%	7677
Total	19,54%	11961

**Ilustración 6: Nivel 2 modelo árbol CHAID exhaustivo: Abertura por posibilidad de postular a upgrade.**

*Fuente: Elaboración Propia.*

De la opción de poder postular a un upgrade, podemos observar que el socio tendrá un 64,18% de probabilidad de usar el contact center. Esto se condice con lo que se mencionó en un principio, ya que para poder ganarse esta postulación se requiere que el socio este en los primeros lugares de la realización de su check-in, por lo mismo, éstos prefieren optar por asegurar la realización de este proceso a través de una llamada (que en un 100% saldrá correcto), versus la posibilidad de realizarlo a primera hora por el sitio web, pero con la posibilidad de que exista algún error.

Cabe mencionar el cómo se van armando las correspondientes cifras de los distintos nodos, por ejemplo al generar la división del nodo 2 en el nodo 5 y 6 de la ilustración 5, nos damos cuenta que la suma de llamadas, 3.976 y 7.677 respectivamente, hacen el total de 11.653 llamadas que se tenían en el nodo 2, pero no así el porcentaje total, ya que esto se calcula cómo:

$$\text{Participación total del perfil} = \frac{\text{Socios}_{\text{nodo}(x)}}{\text{Total Socios}}$$

También podemos ir observando que las cantidades de llamada cada vez van siendo menores, debido a que el árbol por sí sólo nos va armando los perfiles de clientes, y por ende, cada nivel que avanzamos nos entrega un socio más específico, lo que hace que sea una herramienta muy útil para lo que se busca en este trabajo.

Continuando con el análisis del árbol, en la ilustración 7 se puede observar que la siguiente variable evaluada para los socios que tienen la posibilidad de realizar un upgrade, es ver si su vuelo presenta escalas. De ésta se puede ver que la presencia de escalas en el vuelo, hace que los socios tengan una alta de probabilidad de llamar, más específicamente, los socios que tienen viajes internacionales, con posibilidades de postular a upgrade y vuelo con escalas, tienen un 71,32% de usar el servicio del contact center, lo que equivale a 3.959 llamados (nodo 18).

Ahora bien, no se debe dejar de lado el grupo de socios que no tienen escalas en sus vuelos, debido a que su probabilidad asciende a un 58%, lo que representa 3.718 llamados, es decir, si bien los socios con escalas presentan un 13% más de probabilidad de llamar, no hay una gran diferencia entre las cantidades de llamadas de ambos tipos de socios, por lo que a la hora de tomar acciones hasta este nivel, se podría pensar en ambos grupos para la posible definición de acciones.

Nodo 6: UPG		
No Llama	35,86%	4284
Llama	64,18%	7677
Total	19,54%	11961

Nodo 17: No Escalas		
No Llama	41,99%	2692
Llama	58,01%	3718
Total	10,47%	6410

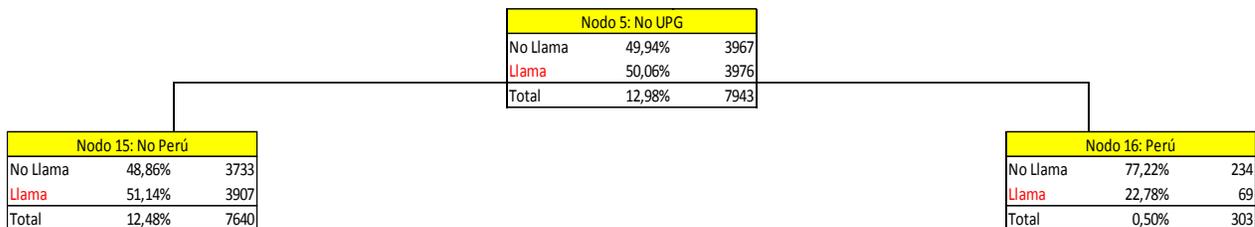
Nodo 18: Escalas		
No Llama	28,68%	1592
Llama	71,32%	3959
Total	9,01%	5551

**Ilustración 7: Tercer nivel árbol: Apertura por si viaje presenta escalas o no.**

**Fuente: Elaboración Propia.**

En la ilustración 8, se puede observar la ramificación que viene desde la ilustración 5, correspondiente a que si el viaje es de tipo internacional y el socio no tiene posibilidad de postula a upgrade (ramificación nodo 5).

De esta rama del árbol se puede observar en primer lugar que la variable más significativa es si el socio es de nacionalidad peruana o no, donde lo más relevante es la probabilidad de llamada que tienen los socios que no son de nacionalidad peruana con un 51,14%.



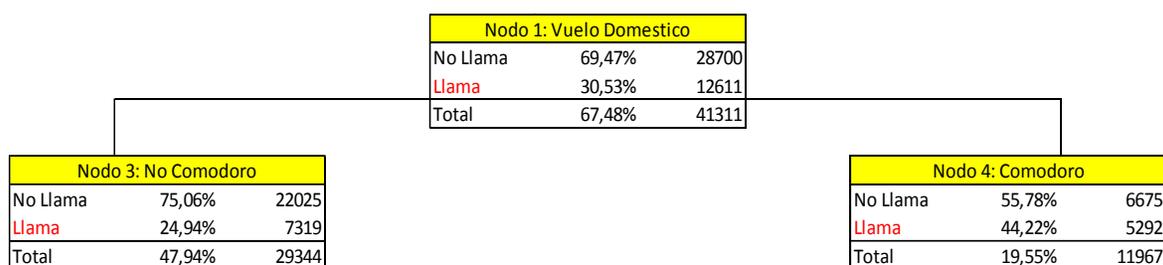
**Ilustración 8 Ramificación del árbol a partir del nodo 5: Vuelo Internacional sin opciones de postular a upgrade.**

*Fuente Elaboración Propia.*

Todo el análisis anterior fue hecho sobre el nodo 2: vuelo internacional, el resto del análisis de este tipo de clientes se encuentre en la sección 10 anexo 13.

A continuación se analizará rama del árbol (nodo 1), en la cual el vuelo del socio es de tipo doméstico.

Como se puede observar en la ilustración 9, la variable más significativa para el tipo de vuelo doméstico, es si el socio pertenece o no a la categoría comodoro. De aquí se puede observar que siendo un vuelo doméstico la probabilidad de llamar para todas las categorías es baja, por lo que un perfil de clientes tan general con estos dos niveles no nos ayuda a poder focalizar las acciones que se vayan a diseñar.



**Ilustración 9: Ramificación segundo nivel para variable: Vuelo doméstico.**

*Fuente: Elaboración Propia.*

A continuación se analizará la rama que genera el nodo 3, donde el socio no es comodoro y sus ramificaciones más relevantes para el objetivo del trabajo. El análisis de los demás nodos se encuentra en la sección 10 anexos 14.

En primer lugar en la ilustración 10, se puede observar que la variable más significativa para ramificar este nodo es la variable si el socio intento chequearse vía web. Para esto analizaremos primero la rama en que el socio no lo intenta, donde la probabilidad de llamar es de un 39,83%. Ahora lo más rescatable de este tipo, es la cantidad de socios que

llaman que son 1991, lo cual es una cifra relevante para la posible reducción de llamados. También es importante recalcar la siguiente ramificación, donde encontramos a los socios que canjean su pasaje (nodo 20), los cuales tienen una alta probabilidad de llamar al contact center. Además este último resultado se condice con lo encontrado en el modelo anterior donde la variable “canjea” era una de las más importante a la hora de aumentar la probabilidad de llamada, es decir, una vez más vemos que ambos modelos van en la misma dirección de los resultados.

Nodo 3: No Comodoro		
No Llama	75,06%	22025
Llama	24,94%	7319
Total	47,94%	29344

Nodo 7: No Intento		
No Llama	60,17%	3008
Llama	39,83%	1991
Total	8,17%	4999

Nodo 20: Canjea		
No Llama	28,50%	61
Llama	71,50%	153
Total	0,35%	214

**Ilustración 10: Ramificación del nodo 7 socio no intenta chequearse vía web (nodo padre: socio no es comodoro).**

**Fuente: Elaboración Propia.**

Continuando con el análisis, la ramificación donde el socio sigue siendo no comodoro, pero ahora si intento chequearse por el sitio web, ilustración 11, podemos resaltar dos casos principalmente.

Nodo 8: Intento		
No Llama	78,12%	19017
Llama	21,88%	5328
Total	39,77%	24345

Nodo 22: Antigüedad entre 3 y 12 años		
No Llama	80,25%	9095
Llama	19,75%	2237
Total	18,52%	11332

Nodo 23: Antigüedad entre 12 y 20 años		
No Llama	73,41%	6270
Llama	26,59%	2270
Total	13,95%	8540

Nodo 51: No Argentino		
No Llama	80,92%	90,36
Llama	19,08%	2130
Total	18,24%	11166

Nodo 52: Argentino		
No Llama	35,54%	59
Llama	64,46%	107
Total	0,27%	166

Nodo 53: No Black		
No Llama	74,25%	6185
Llama	25,75%	2144
Total	13,61%	8329

Nodo 54: Black		
No Llama	40,28%	85
Llama	59,72%	126
Total	0,35%	211

**Ilustración 11: Ramificación del nodo 8 socio intenta chequearse vía web (nodo Padre: socio no es comodoro).**

**Fuente: Elaboración Propia.**

Primero nos encontramos con el nodo 52, el cual representa a un socio donde su viaje es doméstico, no comodoro, intento chequearse por el sitio web, tiene entre 3 y 12 años en el programa LANPASS y es de nacionalidad argentina. Este perfil tiene un 64,46% de probabilidad de llamar al contact center. Al igual que en casos anteriores nos encontramos con un perfil de personas que tienen una alta probabilidad de llamada, pero que su número de llamadas es comparativamente bajo con respecto a otros tipos de socios. Lo mismo sucede con el nodo 54. Por esto mismo, hace que a la hora de elegir posibles candidatos para tomar acciones no estén en las primeras preferencias, aunque no se descartan debido a su posible impacto económico que podrían tener estas aproximadamente 230 llamadas.

No hay que dejar de notar que en el nodo 53, si bien la probabilidad de llamar de este tipo de socios es de un 25,75%, la cantidad de llamadas es un número considerable (2.144), por lo cual podría existir algún tipo de interés en analizar de porque se están realizando.

Finalmente, con todo el análisis descrito podemos observar que se vuelven a ratificar las hipótesis planteadas y el análisis exploratorio realizado anteriormente, principalmente las que mencionan el tipo de viaje, la presencia de escalas, la categoría del socio, etc.

De esta manera, se da por finalizado el análisis de los respectivos nodos y ramificaciones del árbol de decisión, junto con sus probabilidades y cantidades de llamada. El resumen del árbol completo se puede observar en la sección 10 anexos 15 y además en la sección 10 anexos 16, se encuentra el resumen de los perfiles para la base total.

A continuación se presentarán las mismas evaluaciones de desempeño que se hicieron para el modelo logit, con el objetivo de poder medir cuan bueno es el modelo mediante árboles de decisión y de ésta forma poder comparar ambos modelos en la sección 8.3.

Primero que todo analizaremos la tabla 15, la cual muestra la matriz de confusión de la partición de validación del modelo de árbol de decisión. De ella se puede observar la columna “%Correcto” la que muestra tanto la especificidad y sensibilidad del modelo, con un 80% y 52% respectivamente, es decir, es un modelo que logra clasificar cada tipo de socio (no llama como no llama y llama como llama), con un alto porcentaje.

Modelo Árbol de decisión		Observado	
		0	1
Pronostico	0	141800	4686
	1	35449	5078
%Correcto		80%	52%

**Tabla 15: Matriz de confusión árbol de decisión para partición de validación.**

**Fuente: Elaboración Propia.**

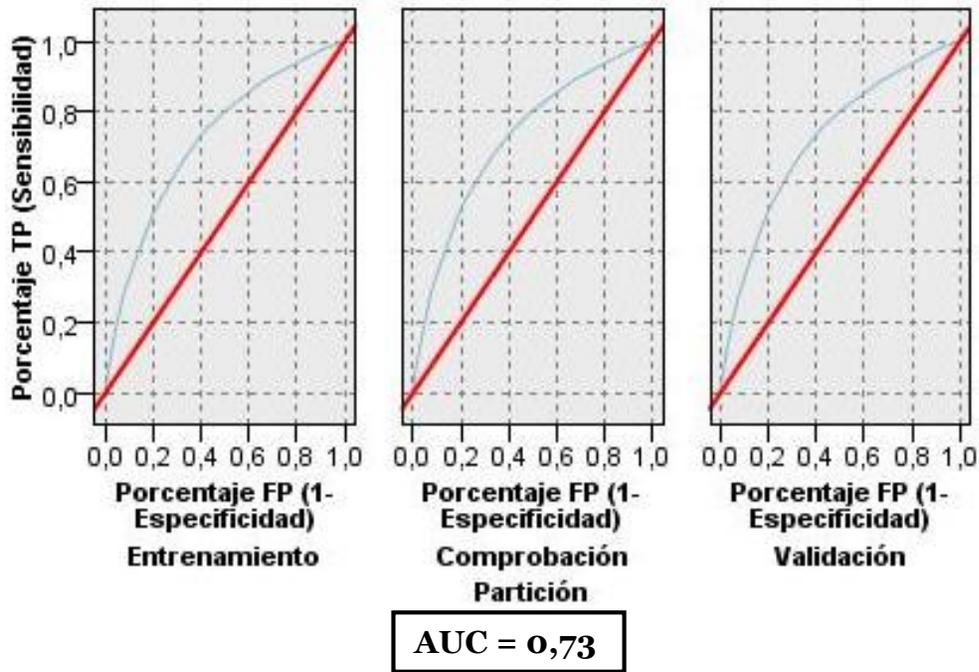
También se presentan las siguientes métricas de desempeño: accuracy, especificidad, sensibilidad, lift, precisión y  $R^2_{Nalgerkerke}$ :

$$\begin{aligned} \mathbf{Accuracy}_{Modelo1} &= \mathbf{78,54\%} \\ \mathbf{Especificidad}_{Modelo1} &= \mathbf{80\%} \\ \mathbf{Sensibilidad}_{Modelo1} &= \mathbf{52\%} \\ \mathbf{LIFT}_{Modelo1} &= \mathbf{1,54} \\ \mathbf{Precisión}_{Modelo1} &= \mathbf{12,5\%} \\ \mathbf{R^2}_{Nalgerkerke} &= \mathbf{0,21} \end{aligned}$$

Con esta información se puede decir que el modelo con aproximadamente un 79% clasifica los valores verdaderos correctamente. Además como ya mencionamos anteriormente, el modelo clasifica a los que no llaman correctamente con un 80% y a los socios que si llaman con un 52%. Además presenta un lift de 1,54, lo que significa que el modelo es 1,54 veces mejor detectando a los socios que llamaran al contact center que la aleatoriedad. Por último es importante mencionar la precisión del modelo la cual es de un 12,5%.

Siguiendo con la evaluación del desempeño del modelo, podemos observar el gráfico 30, el cual nos muestra la curva ROC para cada una de las distintas particiones (entrenamiento, comprobación y validación). Ahora bien al igual que en los análisis anteriores, nos fijaremos en la partición de validación del modelo.

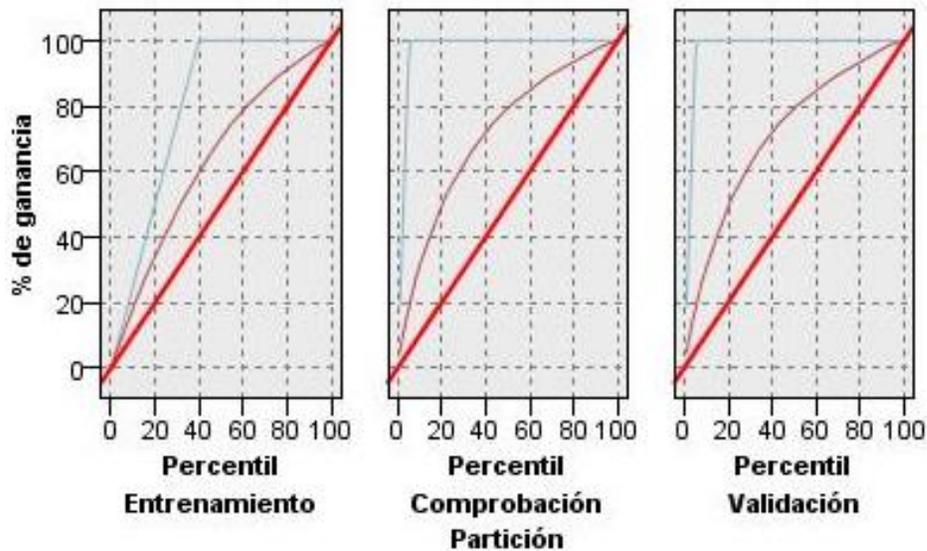
Se puede observar que la línea azul, que representa la curva ROC, muestra que el modelo a través de árboles clasifica de mejor manera a los distintos socios, frente a la aleatoriedad (línea roja). Además se adjunta en el gráfico el valor del AUC para la partición de validación, el cual corresponde a un 0,73, lo que significa que frente a dos socios, el modelo tiene un 73% de probabilidad de clasificarlos correctamente a cada uno.



*Gráfico 30: Curva ROC modelo árbol de decisión para las distintas particiones de la base de datos.*

*Fuente: Elaboración Propia.*

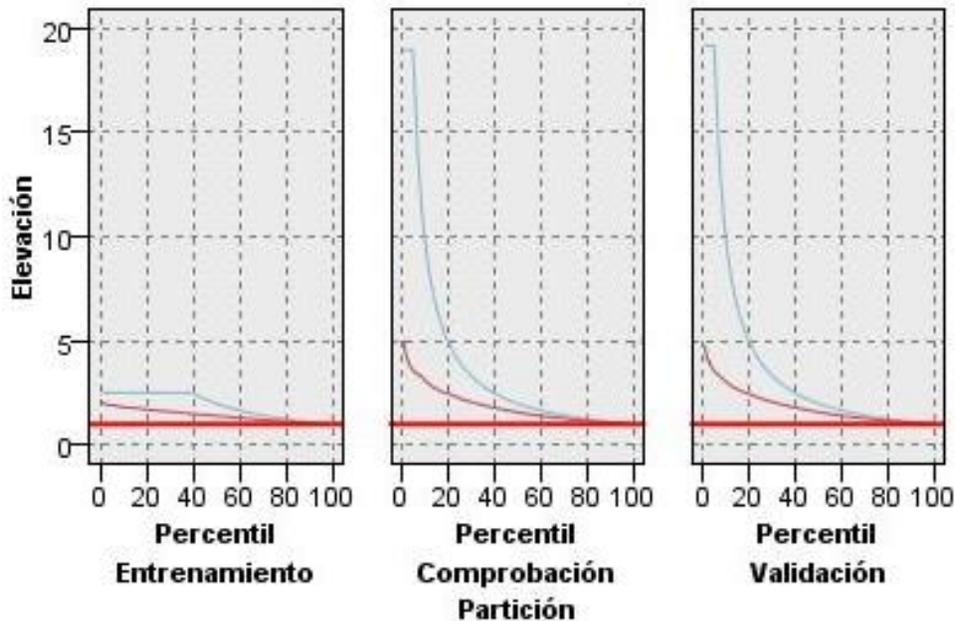
Otra métrica importante es la que se representa en el gráfico 31, la cual muestra la ganancia de información que presenta el modelo. De este se puede observar que para la partición de validación con menos del 30% de los socios más propensos a llamar al contact center, se obtienen más del 60% de los casos, es decir, el modelo es capaz de captar la mayoría de los casos importantes (socios que llaman) con poca información. Este resultado hace que el modelo tenga un gran poder de identificación para el problema en cuestión.



*Gráfico 31: Ganancia de información para modelo árbol de decisión para las distintas particiones de la base de datos.*

*Fuente: Elaboración Propia.*

Finalmente, se encuentra el gráfico 32, el cual muestra la evaluación de la elevación (lift) del modelo. Al igual que en el gráfico anterior, si tomamos el 30% de los socios más propensos a llamar, podemos ver que el modelo entrega una elevación de 2,1, lo que significa que se captura 2,1 veces más el hecho objetivo (socios que llaman) que la aleatoriedad.



**Gráfico 32:** Elevación para modelo de árbol de decisión para las distintas particiones de la base de datos.

*Fuente: Elaboración Propia.*

Con todas las métricas anteriormente descritas, se puede concluir que el modelo a través de árboles de decisión, es sumamente atractivo para pronosticar quienes serán los clientes propensos a llamar al contact center de la línea aérea, debido a su alto poder de detección, su buen comportamiento y principalmente por el output entregado, el cual es de vital importancia para lo que se busca en este proyecto, el cual es poder identificar a los clientes y más aún poder formar perfiles de clientes para poder tomar una acción sobre ellos.

## 7.8 Comparación de Modelos

En esta sección se compararán los distintos modelos analizados anteriormente para poder tomar la decisión de cuál de ellos utilizar para la obtención de los distintos perfiles de clientes y la respectiva elección de ellos.

Antes de mencionar las diferencias de las distintas métricas entre ambos modelos, es importante recalcar que los resultados tanto del logit como del árbol de decisión, entregan casi las mismas variables que afectan más en la probabilidad de llamada, es decir, ambos

van en el mismo sentido de cuál es el contexto que está haciendo que los socios ocupen el servicio del contact center.

A continuación se comparan las distintas métricas de evaluación de desempeño para los distintos modelos. Esto se muestra en la tabla 16.

Métrica	Modelo Logit	Modelo Árbol de decisión
Accuracy	77,72%	78,54%
Sensibilidad	47,24%	52%
Especificidad	79,40%	80%
Precisión	11,22%	12,50%
LIFT (*)	1,52	1,54
Ganancia de información (*)	76%	80%
AUC	0,7	0,73
$R^2_{nalgelkerke}$	0,2	0,21

**Tabla 16: Métricas de evaluación de desempeño de los distintos modelos.**

**(\*): Métricas con el 50% más propenso a llamar.**

**Fuente: Elaboración Propia.**

De la tabla anterior se puede observar que el modelo a través de árboles de decisión presenta en todas las métricas de evaluación de desempeño, una mejoría con respecto al modelo logit. Por otra parte también es cierto que las diferencias no son tan significativas, siendo la que presenta una mayor diferencia es la sensibilidad del modelo con un 4,76% más en el modelo de árbol. Ahora bien un dato interesante a observar es la precisión del modelo de árboles, ya que vuelve a aumentar con respecto al logit, a un 12,5%, lo que hace recordar que este balanceo del modelo fue una opción donde definitivamente se ganó en poder predictivo, ya que con este último sólo se perdió en un 1% de precisión pero se ganó en 35% de sensibilidad.

Para seguir argumentando el uso del modelo de árboles sobre el logit para este trabajo, nos encontramos con cuatro criterios en los cuáles también podemos notar ciertas diferencias [9]:

- **Velocidad:** esta medida se refiere al costo computacional que requiere llevar a cabo el modelo.
- **Robustez:** esta se refiere a la capacidad predictiva del modelo dado datos con errores o con valores perdidos.
- **Escalabilidad:** esta medida muestra la habilidad de construir un predictor eficiente dada las grandes cantidades de datos.

- **Interpretabilidad:** esta muestra el nivel de entendimiento y de las percepciones que puede generar el modelo predictor.

En las 3 primeras medidas, ambos modelos para este trabajo, no se diferencian mucho, tanto en su velocidad, robustez y escalabilidad, se aprecia que pueden ir a la par, pero para este problema en particular, la interpretabilidad del modelo de árboles de decisiones es mejor y más fácil que la del logit, debido a que automáticamente nos entrega los perfiles de clientes más propenso a llamar, especificando su probabilidad y número de llamadas, lo cual es uno de los objetivos principales para este trabajo.

También hay que destacar que ambos modelos, tanto logit como árboles de decisión, entregan resultados similares y apuntan hacia la misma dirección, es decir, que las variables que hacen que la probabilidad aumente son casi las mismas para ambos.

Finalmente, con todos los resultados anteriormente descritos, se decidió que el modelo con que se trabajará en las siguientes secciones para el diseño de acciones, será el modelo de árboles de decisión, debido a su mejor desempeño frente a las distintas métricas y a su destacada interpretabilidad, que hacen de él, un modelo sumamente atractivo para el objetivo buscado.

## 7.9 DISEÑO DE ACCIONES

En esta sección se describirán los perfiles de clientes a los cuales la empresa enfocará sus actividades. A su vez se definirán las acciones correspondientes a cada tipo de perfil, para la posible reducción de las cantidades de llamadas y finalmente se hará una evaluación económica de cada una de éstas, para ver cuán conveniente es tomar estas medidas por parte del área operativa de LANPASS.

Antes de comenzar con la descripción de la elección de los perfiles a trabajar, es importante mencionar, que una vez construidos los perfiles, se pudieron obtener sus datos cuantitativos (probabilidad, lift, cantidad de llamados, etc.) para la base total (sección 10 anexo 16). Ahora bien, como esta base no está balanceada y sólo hay un 5,2% de socios que llaman, es lógico que la probabilidad de llamada disminuya considerablemente respecto a las probabilidades mostradas por el árbol en el análisis anterior, ya que aquel se desarrolla con la partición de calibración que se encontraba balanceada. Por lo tanto, para la elección de los perfiles se tomó en cuenta tanto la probabilidad entregada por el árbol como el lift entregado en la base total, ya que esta última nos muestra cuantas veces más probable es perfil escogido en el problema en cuestión que el promedio.

### 7.9.1 Elección perfiles de clientes

Cómo se mencionó anteriormente, para la elección de los perfiles a trabajar, se siguieron dos criterios. Primero se observaron los distintos nodos finales del árbol de decisión. Con esta información se eligieron los tipos de socios que, por un lado tenían una alta probabilidad de llamada y representaban un alto número de socios, y por otro lado los que hacían sentido para el negocio según la experiencia del área operativa del programa.

En segundo lugar, para los perfiles candidatos se pensaron las posibles acciones a tomar, y con estas en mente, dentro del área operativa, se discutió si tales acciones podían afectar a perfiles de clientes de un nivel superior, es decir, elegir perfiles más generales que específicos, como lo son los de 5 características (nodos finales del árbol). Luego de esta discusión se llegó a la conclusión que los perfiles elegidos por la empresa son los perfiles 1, 2, 3 y 4 descritos en la tabla 17.

Perfil	C. 1	C.2	C.3	C.4	C.5
1	Vuelo Doméstico	No Comodoro	No intento chequearse por página web		
2	Vuelo Doméstico	No Comodoro	Intento chequearse por página web		
3	Vuelo Internacional	No permite Upgrade	Nacionalidad que no sea Peruana		
4	Vuelo Internacional	Permite Upgrade	Vuelo con escalas		
5	Vuelo Doméstico	No Comodoro	No intento chequearse por página web	Canjea pasaje	
6	Vuelo Doméstico	No Comodoro	Intento chequearse por página web	Antigüedad entre 12 y 20 años en el programa	Black

**Tabla 17: Perfiles seleccionados para tomar acciones.**

**Fuente: Elaboración Propia.**

Por otro lado, se puede observar en la misma tabla la existencia de dos perfiles (5 y 6) los cuales no corresponden a los que la empresa eligió atacar, sino que son dos que nacen de los perfiles 1 y 2 respectivamente, es decir, son los perfiles 1 y 2 pero más específicos. Esto debido a que según lo que puede observarse en la tabla 18, el lift de los perfiles 1 y 2 no era mayor a 1, lo que significa que éstos no capturan en una mayor proporción el hecho

de llamar al contact center, con respecto al promedio (esto también se ve reflejado en que, en la base de calibración su probabilidad no era mayor a un 50%), más bien, estos se destacaban principalmente en la cantidad de socios afectados. Por esto mismo, como recomendación propia se decidió por acotar dichos perfiles, para poder tener un tipo de socio que tenga tanto una probabilidad más alta de llamar como un mayor lift, aunque el número de socios afectados disminuya significativamente. Por lo tanto, se evaluarán económicamente los 6 perfiles anteriormente mencionados, para ver si los dos propuestos hacen sentido especificarlos.

Perfil	#Llama	#No Llama	Probabilidad	Participación Llamam	Participación Total	LIFT
1	3.862	70.796	5,17%	7,91%	7,98%	0,99
2	10.673	454.764	2,29%	21,86%	49,78%	0,43
3	7.915	91.129	7,99%	16,21%	10,59%	1,53
4	8.104	38.873	17,25%	16,6%	5,02%	3,3
5	277	1.315	17,4%	0,57%	0,17%	3,33
6	73	665	9,89%	0,15%	0,08%	1,89

**Tabla 18: Descripción cuantitativa de los perfiles escogidos.**

*Fuente: Elaboración Propia.*

### 7.9.2 Acciones

A continuación, en la tabla 19 se definirán las acciones para cada perfil mencionado en la sección 9.1.

Perfil	Acción	Argumento
Socios: - Vuelos domésticos. - black y premium silver. - Sin intento de chequearse por la página web. * Canjea	Plan comunicacional a través de SMS y e-mail invitando a realizar el Check-in.	Debido a que no intentan chequearse por la página web, son un grupo que se identifica como ajeno a ocupar la página para dicho proceso, por lo mismo se personalizará la acción de invitar a realizar el Check-in a través de un SMS y un e-mail.
Socios: - Vuelos domésticos. - black y premium silver. - Con intento de chequeo por la página web. *Antigüedad entre 12 y 20 años en el programa.	Se realizará un diagnóstico sobre los vuelos del cliente, para ver si tuvo reemisiones, cambios de ruta o problemas con la selección de asientos y se dará la solución correspondiente.	Se observa que estos clientes sí ocupan la página web para realizar sus procesos, pero por alguna de estas razones no lo están llevando a cabo por completo, lo que hace que en definitiva ocupen el servicio del contact center. Por lo mismo, una vez identificado el problema, se realizarán las mejoras respectivas.

<b>Socios:</b> <ul style="list-style-type: none"> <li>- Vuelos Internacionales.</li> <li>- No permite upgrade.</li> <li>- Nacionalidad chilena y argentina.</li> </ul>	Se modificará el correo de confirmación, personalizándolo y destacando que no existe la posibilidad de postular a un upgrade.	Esto se debe a que el e-mail actual, no indica si el ticket permite o no permite postular a un upgrade, por lo que el socio tiende a llamar para preguntar e informarse si tiene la posibilidad de éste.
<b>Socios:</b> <ul style="list-style-type: none"> <li>- Vuelos Internacionales.</li> <li>- Permite upgrade.</li> <li>- Vuelos con escalas.</li> </ul>	Proyecto UPG LATAM: Rediseño de la aplicación del sitio UPG.	Este proyecto específicamente ataca dos problemas. El primero es, porque al existir escalas en los vuelos y querer postular a un upgrade, el cálculo de los “cupones upgrade” es distinto si se hace con el mismo origen y destino, pero directo (ver ilustración 16), lo que hace que este tipo de socios hagan uso del contact center para ratificar la cantidad de cupones ocupados. El segundo problema, es que al tratar de hacer Check-in a través de la página cuando existen escalas, se puede dar el caso de que el Check-in de donde se hace la respectiva escala, no esté abierto por la diferencia horaria, lo cual hace que la gente llame al contact center por no poder realizar el proceso completo.

**Tabla 19: Acción para cada perfil de socio.**

**(\*):** *Corresponde a la(s) características(s) específicas, pero que tienen la misma acción que los perfiles más generales.*

*Fuente: Elaboración Propia.*



**Ilustración 12: Situación actual de cobro de cupones UPG para la postulación a upgrade.**

*Se puede observar que si se viaja de Buenos Aires a Miami a través de Santiago, se ocupan 4 cupones, siendo que sólo debería ser 3, ya que el tener escalas o no, no debe influir en la cantidad de cupones ocupados.*

*Fuente: Elaboración Propia.*

Cabe recalcar que para las distintas acciones, se realizarán diferentes experimentos con sus respectivos grupos de tratamiento y control, donde al primero se le aplicará la

acción y al segundo no, con el fin de poder tener un punto de comparación a la hora de ver el porcentaje de llamados al contact center. De esta manera se podrá visualizar concretamente si la acción está teniendo una respuesta positiva por parte de los socios identificados o no.

### 7.9.3 Evaluación económica

Primero que todo, como antecedente para esta parte, cabe recordar que, como se mencionó en un principio, el costo que involucra el contact center es por minuto hablado. Ahora bien, a lo largo de este proyecto se ha trabajado con la cantidad de llamadas, ya que por un lado no se tenía disponible la información de la duración de éstas, y por otro que si se logra reducir una llamada entrante, por si solo se estará reduciendo dichos minutos asociados.

Para realizar la evaluación económica de las acciones descritas anteriormente, se tendrán tres escenarios denominados como: optimista, conservador y pesimista. Estos se diferenciarán en el porcentaje de socios que tendrán una respuesta favorable frente a cada acción, es decir, una vez ejecutada ésta última, se evitara que un porcentaje de los socios afectados llamen al contact center del programa. Primero que todo definimos el porcentaje de aceptación del escenario conservador en un 30%. Este porcentaje nace en base a la experiencia del área operativa, debido a la respuesta en acciones realizadas anteriormente por ellos. De esta última definición se fijan los otros dos escenarios, donde el optimista toma el valor de una respuesta positiva por parte del 50% de los socios accionados y el pesimista con un valor del 10%. Es importante notar que un 50%, no pareciera ser tan atractivo como respuesta del desempeño de una acción, pero para el área operativa, es más bien un exitazo, debido a la respuesta normal que han tenido en varias de las acciones pasadas. Esto se puede observar en la tabla 20.

Escenario	Porcentaje de respuesta favorable
Optimista	50%
Conservador	30%
Pesimista	10%

*Tabla 20: Definición del porcentaje de respuesta positiva para cada uno de los tres escenarios.*

*Fuente: Elaboración Propia.*

Por otra parte, antes de pasar al cálculo de los respectivos beneficios y costos, se tiene que definir a cuantos clientes se podrá atacar para las acciones correspondientes, tanto de SMS como de e-mail, ya que por una parte no se tiene el contacto de todos los socios y por otro la apertura de los correos electrónicos no es siempre del 100%. Esto queda definido por la información disponible desde el área operativa del programa en la tabla 21.

Tipo de comunicación	Disponibilidad de contacto	Open Rate
SMS	90%	100%
E-MAIL	90%	35%

*Tabla 21: Disponibilidad de contacto y open rate de los distintos tipos de comunicación.*

*Fuente: Datos Empresa, Elaboración Propia.*

Finalmente, para tener una visión de ahorros y costos en cuanto a las llamadas propiamente tal, se debe definir el valor de una llamada del tipo “Check-in Upgrade”. Para esto, según la experiencia del área de operaciones, éstas tienen una duración en promedio de 10 minutos, lo que tiene un valor de USD\$ 4,5.

### 7.9.3.1 Cálculo de beneficios

El beneficio que traerán las acciones anteriormente descritas, será principalmente un ahorro en los costos que se están produciendo el contact center, ya que como se evitarán las llamadas de ciertos tipos de socios, se podrán ahorrar los minutos hablados que generarían éstos mismo, si es que no existiese dicha acción. Por esta misma razón, los beneficios se calcularán de la siguiente forma:

$$\text{Beneficio}_i = (\text{socios con respuesta positiva})_i * \text{valor promedio llamada}$$

Donde  $i$  representa cada uno de los distintos escenarios, ya sea optimista, conservador o pesimista. En el caso de las acciones correspondientes a SMS y e-mail también se deberá multiplicar tanto por el porcentaje de información del contacto de los socios, como por el open rate de éstos mismos medios.

### 7.9.3.2 Cálculo de costos

En las tablas 22 y 23 se definen el costos fijos (one-time) y variables para cada acción descrita anteriormente, lo que involucra la definición del valor de las horas hombre involucradas y del envío de cada SMS.

Acción	Descripción Costo	Valor
SMS	Horas hombre dedicadas a la creación del mensaje de texto tipo.	- 25 horas hombre
E-MAIL	Horas hombre dedicadas a la creación del correo tipo.	- 25 horas hombre
Diagnóstico y solución: reemisión, cambios de ruta o selección de asientos	Horas hombre dedicadas por parte del área operativa para el diagnóstico de éstos problemas.	40 horas hombre
Proyecto UPG LATAM	Porcentaje de las horas hombre dedicadas a la solución de los problemas mencionados en la tabla 20.	55 horas hombre

*Tabla 22: Descripción de costos fijos asociados a cada acción.*

*Fuente: Elaboración Propia.*

Acción	Descripción Costo	Valor
SMS	Se hará el cobro por cada envío de los mensajes de texto, más las horas de dedicadas a éste.	- 2.5 horas hombre semanal (envío) - USD\$ 0,049 por envío de mensaje
E-MAIL	Horas hombres dedicadas al envío.	- 2.5 horas hombre semanal (envío)

**Tabla 23: Descripción de costos variables asociados a cada acción.**

*Fuente: Elaboración Propia.*

Finalmente antes de pasar al resultado de costos/beneficios de todas las acciones, se define, el valor de la hora hombre en 0,9 UF (1 UF = CLP\$ 25.629,09) y el valor del dólar en USD\$ 1 = CLP\$708,5.

### 7.9.3.3 Resultado Costo/Beneficio

En la tabla 24, se puede observar el costo total anual que conlleva realizar las distintas acciones para cada uno de los perfiles descritos en la tabla 17.

Perfil	Acción	#Llaman	#NoLlaman	Costo Fijo	Costo Variable	Costo Total
1	SMS	8104	38873	\$ 904	\$ 6.791	\$ 7.696
	E-MAIL	8104	38873	\$ 904	\$ 4.703	\$ 5.607
2	Diagnóstico y solución	7915	91129	\$ 1.447	\$ 0	\$ 1.447
3	E-MAIL	3862	70796	\$ 904	\$ 4.703	\$ 5.607
4	UPG LATAM	10673	454764	\$ 1.990	\$ 0	\$ 1.990
5	SMS	277	1315	\$ 904	\$ 4.781	\$ 5.686
	E-MAIL	277	1315	\$ 904	\$ 4.703	\$ 5.607
6	Diagnóstico y solución	73	665	\$ 1.447	\$ 0	\$ 1.447

**Tabla 24: Costos totales por acción para los distintos perfiles según el número de socios afectados.**

*Fuente: Elaboración Propia.*

Con este resultado, podemos realizar el cálculo del ahorro (beneficio – costo), que traerá el aplicar las acciones correspondientes. De la tabla 25 podemos observar que en dos acciones, independientemente del escenario en cual nos encontremos, traerá resultados positivos para la empresa, es decir, se aprecia su conveniencia de realizar dichas acciones para los resultados económicos de la empresa.

Por otra parte, las acciones para el perfil 1, también se muestran esperanzadoras, debido a que son dos acciones diferentes para el mismo tipo de socios, el total ahorrado será la suma de éstas dos, por lo que tanto en el escenario optimista como en el

conservador traen consigo un ahorro considerable para la empresa. Además se puede observar que la acción ejecutada para el perfil 3 no entrega resultados positivos en cuanto ahorro, por lo que se podría descartar por un lado o bien tratar de optimizar los costos asociados a dicha acción para obtener resultados positivos.

Por último, de la tabla 25, se concluye que económicamente el especificar los perfiles 1 y 2, en el 5 y 6 respectivamente, no trae buenos resultados, por lo que se decidió continuar con el tratamiento a los primeros cuatro perfiles seleccionados por la empresa en un principio.

Perfil	Acción	Ahorro (USD\$)		
		Escenario Optimista	Escenario Conservador	Escenario Pesimista
1	SMS	\$ 8.715	\$ 2.151	\$ -6.547
	E-MAIL	\$ 137	\$ -2.161	\$ -2.325
2	Diagnóstico y solución	\$ 15.819	\$ 8.696	\$ 1.572
3	E-MAIL	\$ -2.870	\$ -3.965	\$ -5.059
4	UPG LATAM	\$ 22.025	\$ 12.419	\$ 2.813
5	SMS	\$ -5.125	\$ -5.349	\$ -5.573
	E-MAIL	\$ -5.411	\$ -5.489	\$ -5.568
6	Diagnóstico y solución	\$ -1.283	\$ -1.348	\$ -1.414

**Tabla 25: Ahorro generado por las diferentes acciones para los perfiles correspondientes.**

*Fuente: Elaboración Propia.*

En la tabla 26 se puede observar el tiempo necesario, en meses, para que la acción empiece a tener ahorros para la empresa.

Perfil	Acción	Tiempo de Recuperación (Mensual)
1	SMS	2
	E-MAIL	11
2	Diagnóstico y solución	2
3	E-MAIL	-
4	UPG LATAM	1

**Tabla 26: Tiempo para que la acción obtenga ahorros para la empresa en meses.**

*Fuente: Elaboración Propia.*

De esta tabla se puede observar que las acciones para el perfil 1, 2 y 4 empiezan a generar ahorros antes de que finalice el primer año. Por otro lado, la acción para el perfil 3, como se vio en la tabla 25 no genera ahorro, por lo cual no genera ningún ahorro en el corto ni largo plazo. Con esta información se puede dar un primer indicio de cuales acciones se deberían priorizar en realizar. Para este caso debería ser UPG LATAM en primer lugar, luego SMS e Diagnóstico y solución y finalmente e-mail.

Finalmente, en la tabla 27 se hizo el cálculo del break-even para los 4 perfiles y sus distintas acciones, el cual corresponde al porcentaje necesario de respuestas positivas para que la acción realizada tenga un ahorro igual a cero, es decir, representa el mínimo porcentaje de aceptación para que la realización de éstas sea rentable.

De aquí se puede observar que según el break-even, las más rentables para realizar son UPG LATAM, SMS y Diagnóstico. Ahora bien, también se nos muestra el lift de cada perfil, para poder observar cómo conversan estas dos medidas, ya que por ejemplo, para la acción de diagnóstico, si bien tenemos un break-even bastante bajo (4%), los socios de este tipo tienen un 0,43 veces más probabilidad de llamar que el promedio, por lo que se debe tener en cuenta la relación existente entre el break-even y lift (probabilidad) de los perfiles, a la hora de priorizar el orden de la realización de cada una de las acciones. Además se observa que el orden para realizar las acciones según el break-even, coinciden con la priorización según el tiempo de recuperación mostrado anteriormente.

Perfil	Acción	Break-Even	LIFT
1	SMS	23%	0,99
	E-MAIL	49%	0,99
2	Diagnostico	4%	0,43
3	E-MAIL	102%	1,53
4	UPG LATAM	4%	3,3

**Tabla 27: Break-Even para las distintas acciones.**

*Fuente: Elaboración Propia.*

De los análisis realizados, se concluye que con tres de las cuatro acciones diseñadas se obtendrá un ahorro positivo para la empresa. Además en conjunto, este tipo de perfiles representan aproximadamente el 60% de las llamadas, los cuales en un escenario conservador, la activación de dichas acciones generaría un ahorro de la quinta parte (US\$24.820 = CLP\$17.584.898) de lo que están costando en promedio este tipo de llamadas. Además, como es de esperar, un mayor ahorro está asociado a mayores tasas de respuestas positivas ante las diferentes acciones y también a un mayor porcentaje tanto de la información disponible del socio como de los esfuerzos por mejorar el open rate.

Finalmente, es preferible realizar las acciones en el orden donde el break-even sea bajo, el lift alto y el tiempo recuperación sea bajo, debido a que de esta manera se espera tener una tasa de éxito mayor que en el caso contrario. Concretamente tenemos el caso del proyecto “UPG LATAM”, el cual tiene un break-even de 4% y un lift de 3,3, por lo mismo es el candidato a tener mayor éxito para reducir la cantidad de llamadas realizadas al contact center, es más, éste es el que presenta un mayor ahorro en la evaluación hecha anteriormente.

## 8 CONCLUSIONES

### 8.1 Conclusiones del trabajo realizado

Las primeras conclusiones generales del trabajo se enmarcan en base a los resultados obtenidos por los modelos predictivos realizados. Primero que todo se decidió trabajar con el motivo “Check-in Upgrade” debido a que representa aproximadamente el 28% de las llamadas y también porque la puesta en marcha de las acciones a realizar es completamente factible. En segundo lugar, el modelo a través de árboles de decisión logro capturar aproximadamente un 30% de los llamados al contact center en el primer decil, lo cual es muy satisfactorio comparado con el modelo aleatorio que captura solo un 10%. Con la interpretación de los resultados del árbol, se pudo observar que éste era el más adecuado en cuanto a los objetivos planteados en un principio del trabajo, debido a la confección rápida y entendible de los distintos perfiles de clientes. Además se pudo verificar que las variables más relevantes que afectan en el aumento de la probabilidad de llamado fueron que el tipo de viaje, la posibilidad de postular a upgrade, viajes con escalas y las distintas categorías de los socios.

Por otra parte, el modelo elegido fue comparado con un modelo logit en el cual con el primer decil se capturo aproximadamente un 20% de los llamados, el cual sigue siendo mejor que la aleatoriedad, pero no mejor que el modelo de árboles de decisión. Ahora bien, lo interesante de la comparación de este modelo, es que ambos compartían las variables relevantes para el hecho objetivo, aunque éste último le da más importancia como variable relevante a la nacionalidad del socio, es decir, éste da cuenta de que puede ser interesante un trato diferenciado a los socios de distintos países enfocando los esfuerzos en los que muestran una mayor probabilidad de llamada, como lo es el caso de los socios argentinos.

Respecto al diseño de acciones, se escogieron en total 4 perfiles de clientes que, por un lado hacían sentido al negocio y por otro tenían una alta probabilidad de llamar al contact center con respecto al promedio. Éstos se caracterizaban principalmente por el tipo de viaje del socio (internacional o doméstico), la categoría, la nacionalidad, la presencia de escalas y el intento previa de haber tratado de realizar el respectivo “Check-in Upgrade”. Éstos representan el 73,68% de los socios de la base total y el 62,58% de los llamados. Para cada uno de estos se confeccionó una acción con tres distintos propósitos, ya sea educar a los socios, rediseñar el conducto regular de los procesos que no estaban cumpliendo su función a cabalidad y finalmente, pero más importante, el poder reducir el mayor porcentaje posible de las llamadas hacia el contact center.

Desde la evaluación económica realizada para cada acción, se desprende que con la puesta en marcha de estas últimas, en un escenario conservador, se podrá lograr un ahorro aproximadamente del 20% de lo que actualmente estos perfiles están costando. Además este 20% viene a representar una cantidad superior a los 15 millones de pesos y podría llegar a ser superior a los 35 millones, si se logra tener una respuesta positiva del 50%. De esto se concluyó que son resultados satisfactorios debido a que con este ahorro

producido, se está reduciendo en aproximadamente el 50% del exceso del presupuesto del año 2014 correspondiente a la atención preferente, y más aún, este resultado sólo está siendo producido por el 70% de los socios preferentes, los cuales representan el 24% de las llamadas al contact center, es decir, que como se vio en un principio del trabajo, al ser la distribución de las llamadas parecidas entre preferentes y generales, el aplicar estas acciones a todas las categorías de socios, podría terminar en un completa reducción del exceso de presupuesto, o más aún, en una reducción del presupuesto propiamente tal. Por otro lado, se calculó el break-even de dichas acciones, el cual en 3 de 4 de ellas se observó que con un pequeño porcentaje (4%-6%) de respuestas positivas la acción se hacía rentable.

A modo de resumen, se concluyen que todos los objetivos planteados al comienzo del trabajo fueron realizados en su totalidad, obteniendo modelos con resultados satisfactorios que validaron las hipótesis planteadas en un principio, junto con la identificación de las variable relevantes para identificación de un llamado al contact center y además se obtuvo la evaluación económica correspondiente a las distintas acciones confeccionadas, dando como resultado un ahorro superior a los 15 millones de pesos en su escenario conservador.

## **8.2 Limitaciones del trabajo**

Dentro de las limitaciones del trabajo se destaca que, debido a la forma de tipificación de los ejecutivos del contact center, no fue posible poder vincular a los socios de atención general (LANPASS y Premium) con el motivo de la llamada correspondiente. Por lo mismo el trabajo se basó sólo en el estudio de los socios de atención preferente.

Una segunda limitación del trabajo, fue que en primera instancia, el objetivo de la información requerida era el obtener la duración de las llamadas de los socios preferentes, debido a que el cobro de este servicio es por minuto hablado. Lo que no se pudo concluir por la disponibilidad de información al momento de realizar este trabajo, lo que termino en la decisión de trabajar con la cantidad de llamadas.

Una tercera limitación del trabajo, fue la dificultad que se tuvo para la extracción de los datos por parte de la base de CUSTOMER. Debido a que por problemas de restricciones de uso en los servidores, se tuvo que sacar los datos requeridos día por día para todo el año 2014, lo que significó una gran cantidad de tiempo.

## **8.3 Recomendaciones y trabajos futuros**

Uno de los objetivos de futuros trabajos debiese apuntar en realizar los correspondientes experimentos de las acciones diseñadas. Para las acciones de SMS y e-mail se debiese tomar un grupo de control, al cual no se le debe aplicar nada, y un grupo

testeo al cual se le aplica lo propuesto en este trabajo. Para la parte de diagnóstico y proyecto UPG LATAM se puede aplicar el mismo procedimiento, pero del tipo A/B testing<sup>9</sup>, para luego comparar los resultados de ambos grupos y ver la efectividad de las acciones realizadas.

Además, en línea con lo anterior se recomienda que la realización de acciones se ejecuten por las que pueden tener mayor probabilidad de éxito, es decir, por las cuales el break-even sea bajo y el lift sea alto, ya que esto significará que, por un lado la probabilidad llamar del socio es alta y por otro, no se necesita que un gran porcentaje de socios tenga una respuesta positiva ante tal acción para hacer de ésta, una acción rentable para la empresa. En concreto con los resultados, se recomienda realizar en este orden las acciones: Proyecto UPG LATAM, SMS, Diagnóstico y E-mail (1). Es importante recalcar que se debe evaluar una posible modificación de la acción de envío de e-mail para el socio que tiene viaje internacional, no puede postular a upgrade y no es de nacionalidad peruana, para que ésta pueda ser rentable en un escenario conservador.

También, como se mencionó en las limitaciones del trabajo, no fue posible estudiar el comportamiento de las duraciones de las llamadas. Por lo que se plantea como futura línea de investigación, el análisis de las duraciones de las llamadas de los socios preferentes, para ver qué motivo es el que se está llevando en términos de minutos, el mayor costo para el programa, de tal forma de ver si se ratifica o cambia la importancia de los motivos estudiados.

Otra información que podría haber sido valiosa para el trabajo, hubiese sido tener una información más cualitativa respecto de porque los socios están llamando, a modo de ejemplo si se tuviese la información de si los clientes no tienen internet, o información sobre los procesos que hacen que abandonen la página web, sería posible tomar acciones aún más específicas, de tal forma que el break-even para que éstas sean rentables, sea aún más bajo.

Finalmente, respecto a la evaluación económica, puede haber enriquecido el trabajo, el haber aplicado las características de los perfiles encontrados, a los socios de categorías LANPASS y Premium, es decir, al 76% de las llamadas restantes, con el objetivo de ver el ahorro total que pudiesen generar las acciones diseñadas. Además otro aporte valioso para el trabajo, sería perfeccionar la estimación de los costos correspondientes a las horas hombre y al promedio de minutos hablados para el motivo estudiado.

---

<sup>9</sup> A/B Testing, test para hacer pruebas de acciones sobre páginas web, en: <https://vwo.com/ab-testing/>

## 9 BIBLIOGRAFÍA

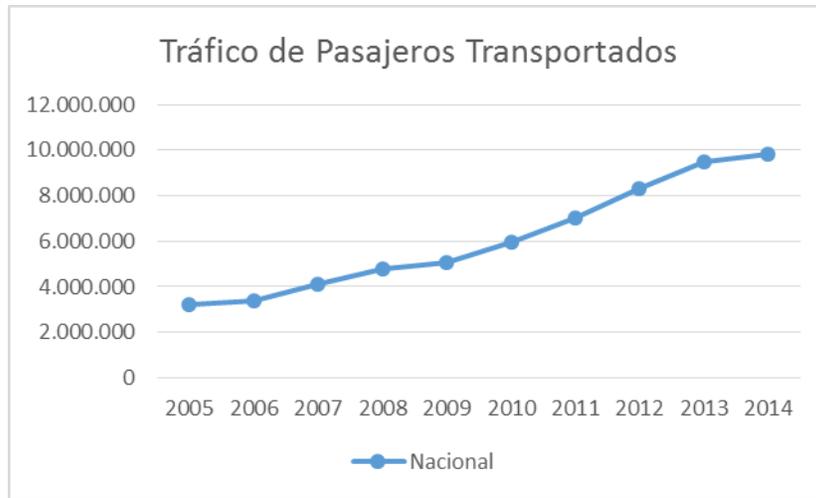
- [1] O. Colmenares and J. Saavedra (2007), “Theoretical review of the brand loyalty: approaches and valuations”. Universidad de Zulia.
- [2] Y. Yi and H. Jeon, “Effects of loyalty programs on value perception, program loyalty, and brand loyalty”. *Journal of the Academy of Marketing Science*. Junio 2003.
- [3] ROBERTS, JOHN y LILIEN, GARY. 1993. Chapter 2 Explanatory and predictive models of consumer behavior. *Handbooks in Operations Research and Management Science*, Vol.5, 27-82.
- [4] A. Vieytes y C. González (2004), “Predicción del rendimiento académico final a partir de pruebas previas en asignaturas cuantitativas”. Universidad Pontificia Comillas de Madrid.
- [5] Berlanga Vanesa, Rubio María y Vilà Ruth (2013), “Cómo aplicar árboles de decisión en SPSS”. Universidad de Barcelona, España.
- [6] Greene, W.H., (2003) “Econometric Analysis”, tercera edición, Prentice-Hall.
- [7] B. Villena, “Curso: Aplicaciones de Probabilidades y Estadística en Gestión”. Universidad de Chile, Chile.
- [8] U. Fayyad, G. Piatetsky-Shapiro and P. Smith (1996) “The KDD process for extracting useful knowledge from volumes of data”.
- [9] J. Han and M. Kamber, “*Data Mining: Concepts and Techniques 2nd. Ed.*”
- [10] Ares Felipe (2008), “*Valorización marginal del programa de pasajero frecuente en una línea aérea*”. Santiago, Chile.
- [11] Fuenzalida Bárbara (2012), “Pronostico de efectividad de promociones sobre clientes, dadas sus características y respuestas pasadas”. Santiago, Chile.
- [12] D. Kleinbaum and M. Klein (2002), “Logistic Regression A Self-Learning Text”.
- [13] M. Goic, “Curso: Marketing II”. Universidad de Chile, Chile.
- [14] Rodríguez, G. (2007), “Chapter 3: Logit Models for Binary Data. Lecture Notes on Generalized Linear Models”. Princeton University, Estados Unidos.
- [15] Wen Zhu, Nancy Zeng and Ning Wang (2010), “Sensitivity, Specificity, Accuracy, Associated Confidence Interval and ROC Analysis with Practical SAS® Implementations”.

[16] Rojas Constanza (2015), “Modelo predictivo de desafiliación de empresas para una caja de compensación de asignación familiar”. Santiago, Chile.

[17] Vargas María (2009), “Rediseño de un modelo de incentivos para vendedores de una tienda por departamentos”. Santiago, Chile.

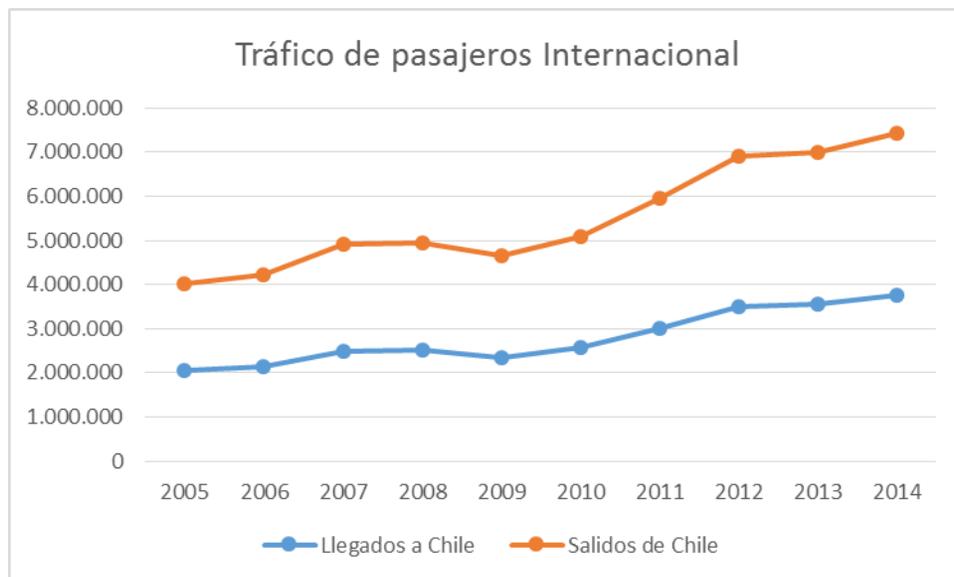
## 10 ANEXOS Y APENDICES

### 10.1 Anexo 1: Tráfico de pasajeros Nacional



**FUENTE: JUNTA DE AERONÁUTICA CIVIL**

### 10.2 Anexo 2: Tráfico de pasajeros Internacional



**FUENTE: JUNTA DE AERONÁUTICA CIVIL**

### 10.3 Anexo 3: Tabla de precisión según AUC

AUC Range	Classification
$0.9 < AUC < 1.0$	Excellent
$0.8 < AUC < 0.9$	Good
$0.7 < AUC < 0.8$	Worthless
$0.6 < AUC < 0.7$	Not good

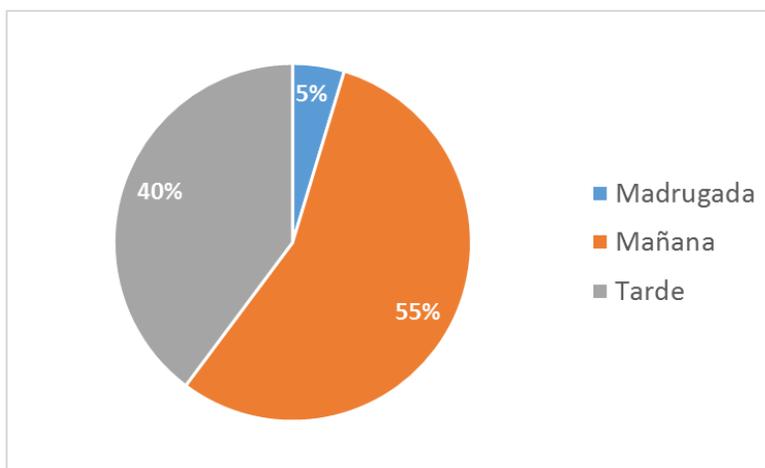
### 10.4 Anexo 4: Variables horario y día

La primera de éstas, es el horario en que se producen las llamadas, para esto utilizamos la tipificación final y describimos las 24 horas del día en tres bloques de 8 horas, tal como se describe a continuación:

- **Madrugada:** bloque de horario correspondiente entre las 00:00 a 07:59 A.M.
- **Mañana:** bloque de horario correspondiente entre las 08:00 A.M a 15:59 P.M.
- **Tarde:** bloque de horario correspondiente entre las 16:00 P.M. a 23:59 P.M.

Primero que todo debemos mencionar que la cantidad de llamados en los distintos bloques difiere de gran manera (gráfico 33), ya que para cada bloque tenemos las siguientes cantidades de llamados:

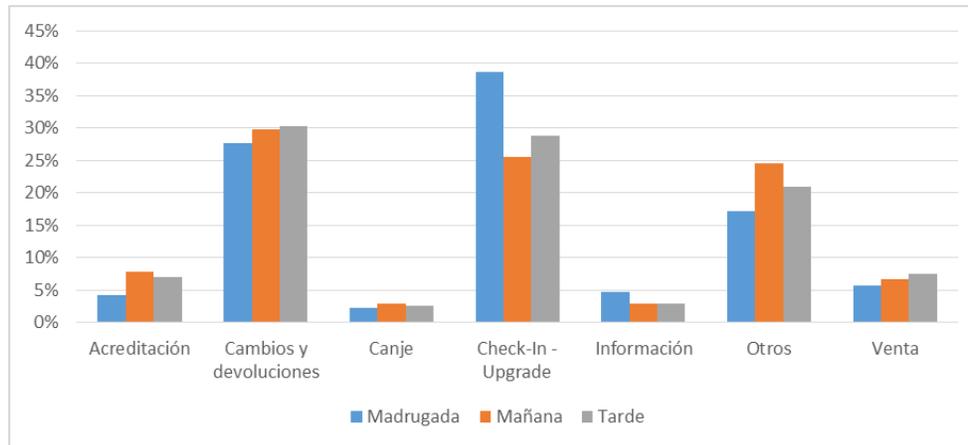
- **Madrugada:** 4.428 llamados, correspondiente al 4,68%.
- **Mañana:** 52.606 llamados, correspondiente al 55,55%.
- **Tarde:** 37.668 llamados, correspondiente al 39,78%.



**Gráfico 33: Distribución de llamados por bloque horario en 2014.**

**Fuente: Datos Empresa, Elaboración Propia.**

Ahora bien, en el gráfico 34 se puede observar la distribución de las principales tipificaciones en los distintos bloques de horarios en todo el año 2014.



**Gráfico 34: Distribución de las tipificaciones 2014 por bloque horario.**

**Fuente: Datos Empresa, Elaboración Propia.**

En el gráfico anterior se puede observar que la participación que tienen las distintas tipificaciones del 2014 en los distintos bloques de horarios es muy parecida a excepción de los llamados por motivos de “Check-in Upgrade” en el horario de la madrugada, es decir, que para las horas entre 00:00 y las 7:59, el 40% de los llamados se debe al motivo recién mencionado.

Por lo tanto, con esta información, lo que se podría llegar a recomendar es que para el bloque de la madrugada, las habilidades que presenten los ejecutivos a cargo del contact center, estén enfocados en este tipo de problemas, para poder tener una mayor eficiencia a la hora de responder este tipo de llamados, y así poder ayudar a la disminución de minutos hablados, manteniendo el nivel de servicio del contact center de LANPASS. También es importante recalcar, que otra habilidad a reforzar, en todo horario, es la capacidad de atender llamados del tipo “Cambios y devoluciones”, que representan cerca del 30% en los distintos bloques.

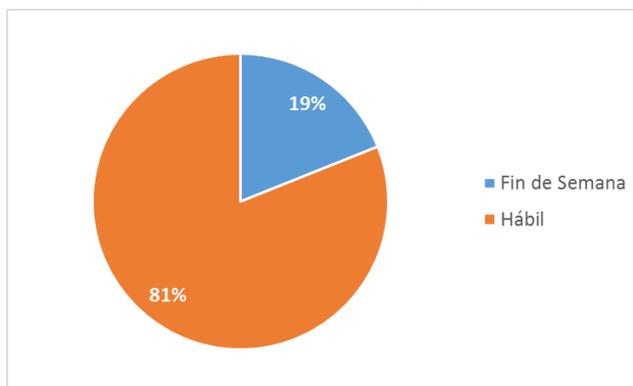
En segundo lugar, la otra variable interesante a describir es el tipo de día que llaman los clientes preferentes, es decir, como se distribuyen los llamados entre los días hábiles y los fines de semanas. Para efectos de esta variable se definen los siguientes tipos de días:

- **Hábil:** días entre lunes y viernes.
- **Fin de semana:** día sábado y domingo.

Debido a la división de los días, la cantidad de llamados en ambas también difiere en aproximadamente 59.000 llamados, ya que por un lado los días hábiles presentan una

cantidad de cinco días y los fines de semana sólo dos. Las cantidades correspondientes para cada tipo de día son (gráfico 35):

- **Hábil:** 76.757 llamados, correspondiente al 81%.
- **Fin de semana:** 17.945 llamados, correspondiente al 19%.

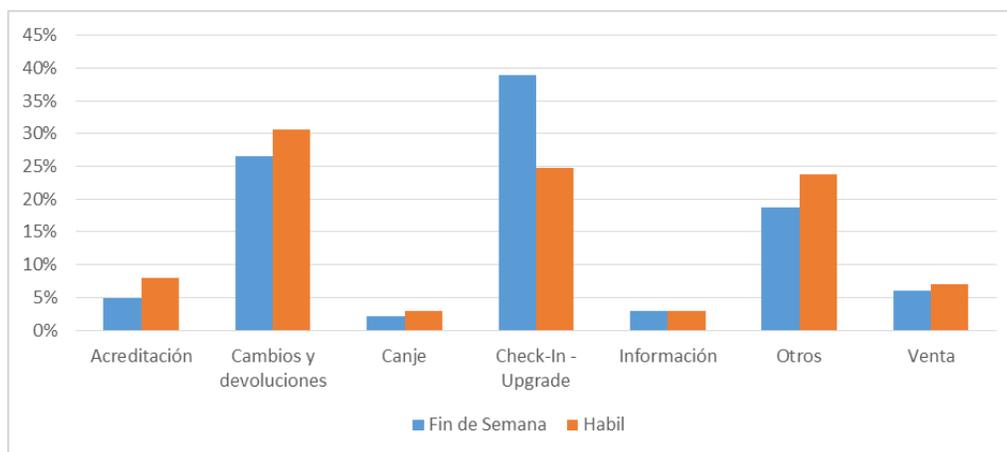


**Gráfico 35: Distribución de llamados por tipo de día en 2014.**

*Fuente: Datos Empresa, Elaboración Propia.*

Al igual que en la variable anterior, las cantidades de llamada ayudan a decidir la cantidad de ejecutivos a asignar por tipo de día, donde claramente los fines de semana se necesitan menos que en el transcurso de los días hábiles.

Ahora bien, la distribución de las tipificaciones en estos dos tipos de días, es la que se presenta en el gráfico 36.



**Gráfico 36: Distribución tipificaciones por tipo de día en 2014.**

*Fuente: Datos Empresa, Elaboración Propia.*

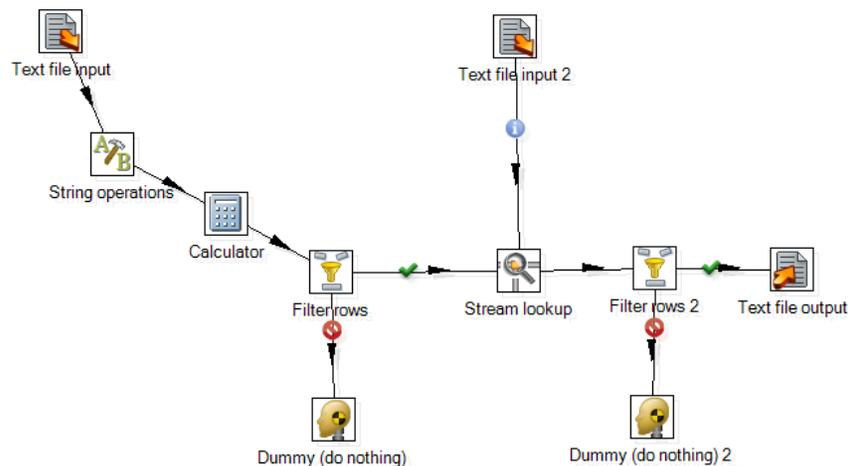
Del gráfico anterior se puede observar claramente que los días correspondientes a los fines de semana, el motivo de llamados “Check-in Upgrade” tiene el mayor de los pesos, lo que puede deberse principalmente a que la gente acostumbra a viajar en este tipo de días.

Al igual que para la variable bloque de horario, esta descripción sirve más para una organización del contact center, y poder decidir las habilidades que se necesitan principalmente los diferentes tipos de días.

Finalmente, para estas dos variables y sus recomendaciones correspondientes, hay que tener en cuenta que todo este análisis es para la atención preferente del contact center de LANPASS, por lo que tanto las habilidades como la capacidad que se pueda organizar están enfocadas en este tipo de atención, aunque cómo vimos anteriormente por su similar distribución, esto también puede ser útil para la atención general.

Además cabe mencionar que, como se puede observar, el análisis presentado está de forma agregada y se dan directrices de lo que se puede hacer de forma general, y aunque éste no es el principal objetivo de la presente memoria, esto puede servir más como un apoyo a las decisiones que se están tomando con respecto a este servicio del programa.

### 10.5 Anexo 5: Extracto de un flujo de transformación de la herramienta PENTAHO.



FUENTE: ELABORACIÓN PROPIA

### 10.6 Anexo 6: Desarrollo test de proporciones de la variable saldo

$$H_0: p_{(100K;200K]} = p_{(200K;4000k]}$$

$$H_A: p_{(100K;200K]} \neq p_{(200K;4000k]}$$

$$p_{(100K;200K]} = 5,68\% ; p_{(200K;4000k]} = 4,81\%$$

$$n = 193.747 ; m = 497.909$$

$$\hat{p} = \frac{9.319 + 28.275}{193.747 + 497.909} = 0,054$$

$$Z = \frac{5,68\% - 4,81\%}{\sqrt{(0,054 * 0,945) * \left(\frac{1}{193.747} + \frac{1}{497.909}\right)}} = 14,312 > 1,96$$

### 10.7 Anexo 7: Resultado modelo 2 Logit

Variable		Beta	Significancia	Exp(Beta)
Anticipación de Compra	Anticipación de 5 o menos días	0,304	0	1,356
	Anticipación entre 6 y 10 días	0,139	0	1,149
	Anticipación entre 10 y 30 días	0,019	0,436	1,019
	Anticipación más de 90 días	0,317	0	1,373
Categoría	Categoría Black	0,482	0	1,62
	Categoría Comodoro	0,777	0	2,175
Edad	Edad de 30 o menos años	0,102	0,023	1,107
	Edad entre 30 y 40 años	0,183	0	1,201
	Edad entre 40 y 60 años	0,076	0,007	1,079
Canal de Compra	Canal de compra: Internet	0,121	0	1,129
	Canal de compra: Call center.	0,034	0,15	1,035
Nacionalidad	Nacionalidad Argentina	0,699	0	2,011
	Nacionalidad Chilena	0,142	0,004	1,153
	Nacionalidad Peruana	-0,655	0	0,519
Vuelo con escalas		0,489	0	1,631
Cliente intento chequearse por web		-0,532	0	0,587
Vuelo Internacional		0,643	0	1,903
Posibilidad de postular a Upgrade		0,203	0	1,225
Canjeo pasaje		0,495	0	1,641
Ln(Antigüedad en el programa)		0,526	0	1,693
Anticipación de 5 o menos días x Canje		0,628	0	1,873
Anticipación entre 6 y 10 días x Canje		-0,296	0,039	0,744
Anticipación entre 10 y 30 días x Canje		0,164	0,092	1,178
Anticipación entre 30 y 90 días x Canje		0,108	0,193	1,114
Anticipación de 5 o menos días x Inter		-0,154	0,007	0,857
Anticipación entre 6 y 10 días x Inter		-0,053	0,394	0,949
Anticipación entre 10 y 30 días x Inter		0,051	0,334	1,053
Anticipación entre 30 y 90 días x Inter		0,141	0,009	1,151
Constante		-2,537	0	0,079

FUENTE: ELABORACIÓN PROPIA

### 10.8 Anexo 8: Matriz de confusión modelo 2 Logit partición testeo

Modelo 2 Logit		Observado	
		0	1
Pronostico	0	141297	5264
	1	35952	4500
%Correcto		79,72%	46,09%

FUENTE: ELABORACIÓN PROPIA

### 10.9 Anexo 9: Métricas de desempeño modelo 2 Logit

Métrica	Valor
Accuracy	77,96%
Sensibilidad	46,09%
Especificidad	79,72%
LIFT	1,4
$R^2_{nalgelkerke}$	0,2

FUENTE: ELABORACIÓN PROPIA

### 10.10 Anexo 10: Tabla de ganancia de información modelo 2 Logit partición testeo

PROB	SENSIBILIDAD	% CASOS ACUM
80,0%	3,5%	1,00%
70,0%	13,7%	4,98%
60,0%	29,1%	11,76%
50,0%	46,1%	21,63%
40,0%	62,7%	35,27%
30,0%	78,6%	52,82%
20,0%	93,6%	81,11%
10,0%	99,9%	99,03%
9,0%	100,0%	99,49%
8,0%	100,0%	99,80%
7,0%	100,0%	99,91%
6,0%	100,0%	99,98%
5,0%	100,0%	99,99%
4,0%	100,0%	100,00%
3,0%	100,0%	100,00%
2,0%	100,0%	100,00%
1,0%	100,0%	100,00%
0,0%	100,0%	100,00%

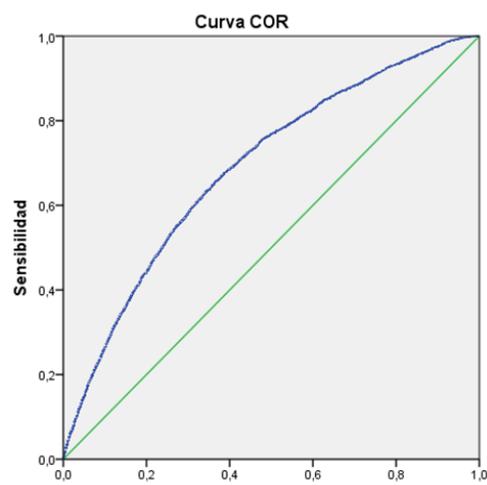
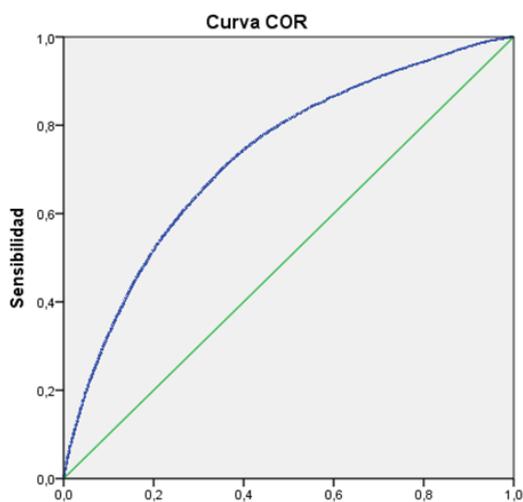
FUENTE ELABORACIÓN PROPIA

### 10.11 Anexo 11: Resultados modelo 1 logit no balanceado

Variable		Beta	Significancia	Exp(Beta)
Anticipación de Compra	Anticipación de 5 o menos días	0,342	0	1,407
	Anticipación entre 6 y 10 días	0,13	0	1,139
	Anticipación entre 10 y 30 días	0,102	0	1,107
	Anticipación más de 90 días	0,126	0	1,134
Categoría	Categoría Black	0,277	0	1,319
	Categoría Comodoro	0,669	0	1,953
Edad	Edad de 30 o menos años	0,184	0	1,202
	Edad entre 30 y 40 años	0,235	0	1,265
	Edad entre 40 y 60 años	0,12	0	1,128
Canal de Compra	Canal de compra: Internet	0,185	0	1,203
	Canal de compra: Call center.	0,203	0	1,225
Nacionalidad	Nacionalidad Argentina	0,739	0	2,094
	Nacionalidad Chilena	0,24	0	1,271
	Nacionalidad Peruana	-0,592	0	0,553
Vuelo con escalas		0,574	0	1,776
Cliente intento chequearse por web		-0,519	0	0,595
Vuelo Internacional		0,606	0	1,833
Posibilidad de postular a Upgrade		0,188	0	1,207
Canjeo pasaje		0,588	0	1,801
Ln(Antigüedad en el programa)		1,142	0	3,134
Constante		-5,106	0	0,006

#### Resultado coeficientes modelo 1 logit.

Fuente: Elaboración Propia.



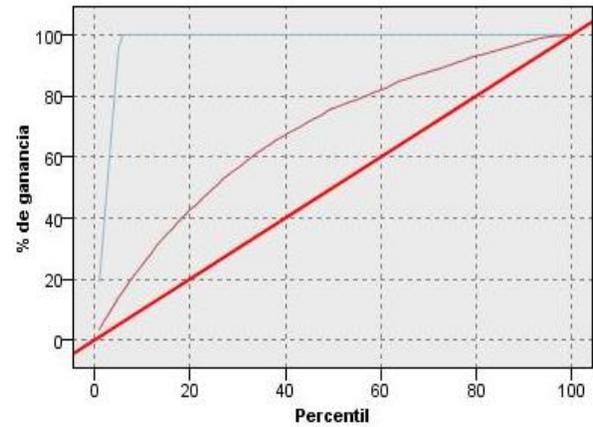
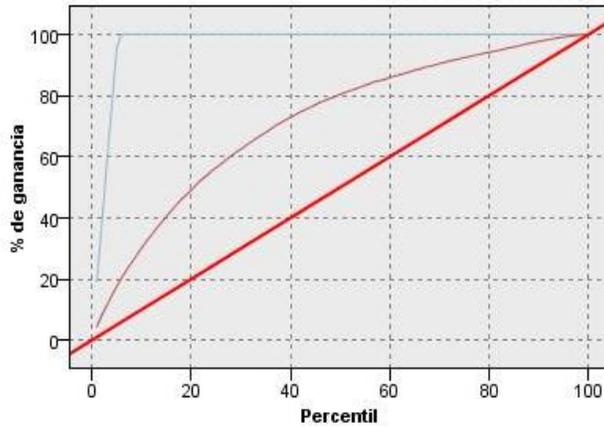
**Curva ROC modelo 1 logit (Calibración y Testeo)**  
**Fuente: Elaboración Propia.**

Modelo 1 Logit No Balanceado		Observado	
		0	1
Pronostico	0	165979	7961
	1	11269	1803
%Correcto		94%	18%

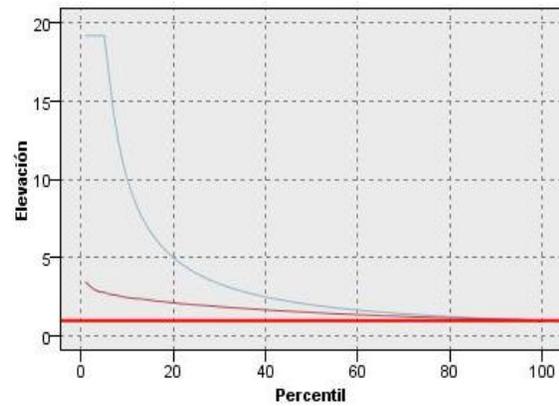
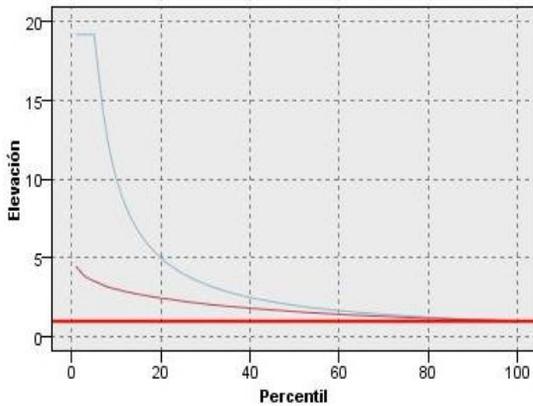
**Matriz de confusión modelo 1 logit**  
**Fuente: Elaboración Propia.**

Métrica	Valor
Accuracy	89,72%
Sensibilidad	18,47%
Especificidad	93,64%
LIFT	1,5
$R^2_{nalgelkerke}$	0,99

**Métricas de desempeño modelo 1 logit**  
**Fuente: Elaboración Propia**



**Curva de elevación modelo 1 logit (Calibración y Testeo)**  
**Fuente: Elaboración Propia.**

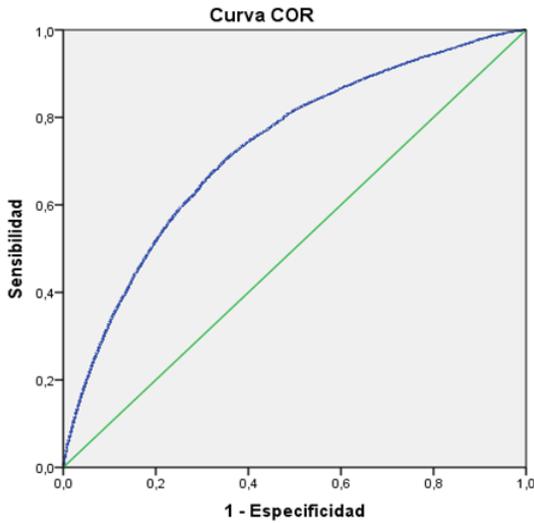


**Curva de ganancia modelo 1 logit (Calibración y Testeo)**  
**Fuente: Elaboración Propia.**

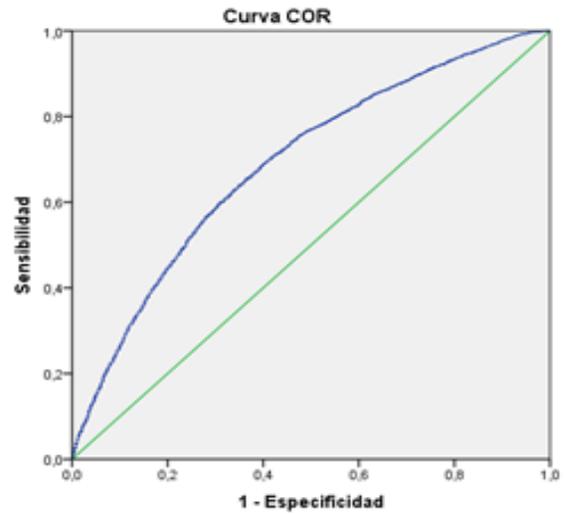
## 10.12 Anexo 12: Resultados modelo 2 logit no balanceado

Variable		Beta	Significancia	Exp(Beta)
Anticipación de Compra	Anticipación de 5 o menos días	0,463	0	1,589
	Anticipación entre 6 y 10 días	0,257	0	1,293
	Anticipación entre 10 y 30 días	0,178	0	1,194
	Anticipación más de 90 días	0,27	0	1,31
Categoría	Categoría Black	0,29	0	1,337
	Categoría Comodoro	0,67	0	1,955
Edad	Edad de 30 o menos años	0,185	0	1,203
	Edad entre 30 y 40 años	0,242	0	1,273
	Edad entre 40 y 60 años	0,128	0	1,136
Canal de Compra	Canal de compra: Internet	0,189	0	1,209
	Canal de compra: Call center.	0,187	0	1,206
Nacionalidad	Nacionalidad Argentina	0,746	0	2,108
	Nacionalidad Chilena	0,249	0	1,283
	Nacionalidad Peruana	-0,585	0	0,557
Vuelo con escalas		0,575	0	1,777
Cliente intento chequearse por web		-0,518	0	0,596
Vuelo Internacional		0,61	0	1,84
Posibilidad de postular a Upgrade		0,194	0	1,214
Canjeo pasaje		0,304	0	1,356
Ln(Antigüedad en el programa)		1,124	0	3,078
Anticipación de 5 o menos días x Canje		0,706	0	2,026
Anticipación entre 6 y 10 días x Canje		0,222	0,028	1,248
Anticipación entre 10 y 30 días x Canje		0,314	0	1,369
Anticipación entre 30 y 90 días x Canje		0,287	0	1,333
Anticipación de 5 o menos días x Inter		-0,172	0	0,842
Anticipación entre 6 y 10 días x Inter		-0,106	0,017	0,899
Anticipación entre 10 y 30 días x Inter		0,003	0,927	1,003
Anticipación entre 30 y 90 días x Inter		0,182	0	1,199
Constante		-5,187	0	0,006

**Resultado coeficientes modelo 2 logit.  
Fuente: Elaboración Propia.**



**AUC = 0,72**



**AUC = 0,69**

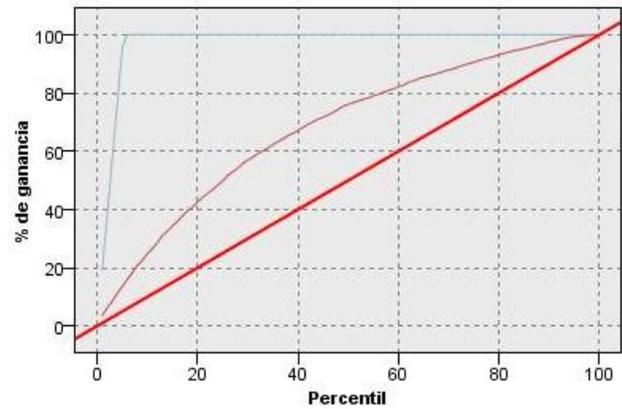
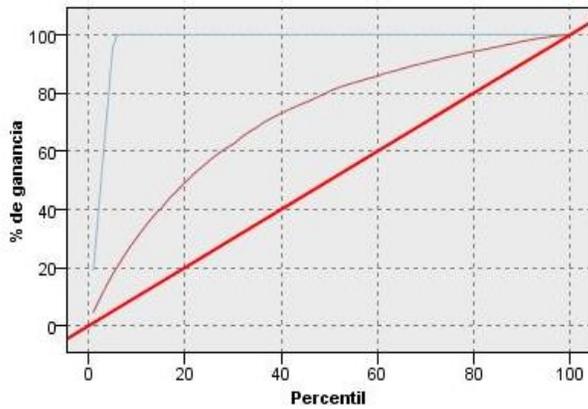
**Curva ROC modelo 2 logit (Calibración y Testeo)  
Fuente: Elaboración Propia.**

Modelo 2 Logit No Balanceado		Observado	
		0	1
Pronostico	0	166446	8054
	1	10803	1710
%Correcto		93,91%	17,51%

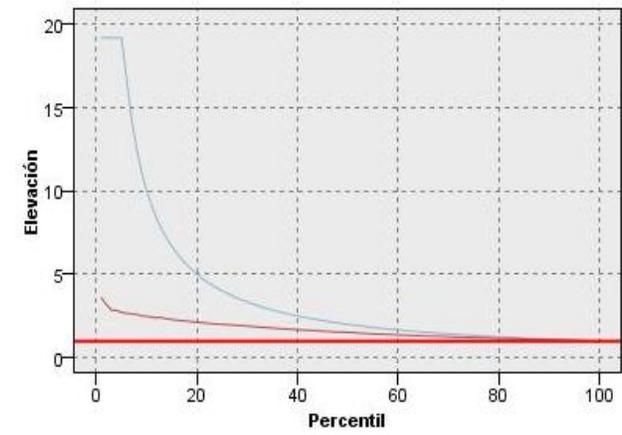
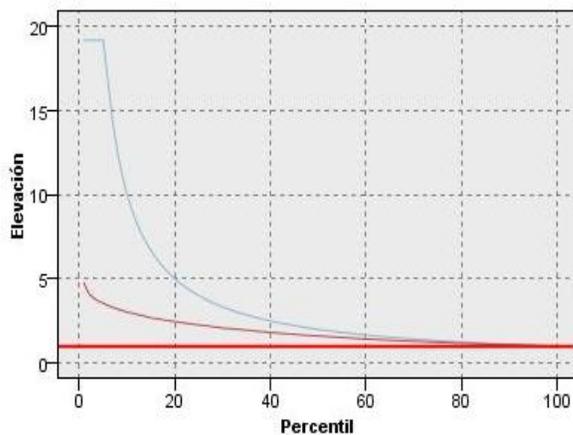
**Matriz de confusión modelo 2 logit  
Fuente: Elaboración Propia.**

Métrica	Valor
Accuracy	89,9%
Sensibilidad	17,5%
Especificidad	93,9%
LIFT	1,5
$R^2_{nalgerkerke}$	0,1

**Métricas de desempeño modelo 2 logit  
Fuente: Elaboración Propia**



**Curva de ganancia modelo 2 logit (Calibración y Testeo)**  
**Fuente: Elaboración Propia.**



**Curva de elevación modelo 2 logit (Calibración y Testeo)**  
**Fuente: Elaboración Propia.**

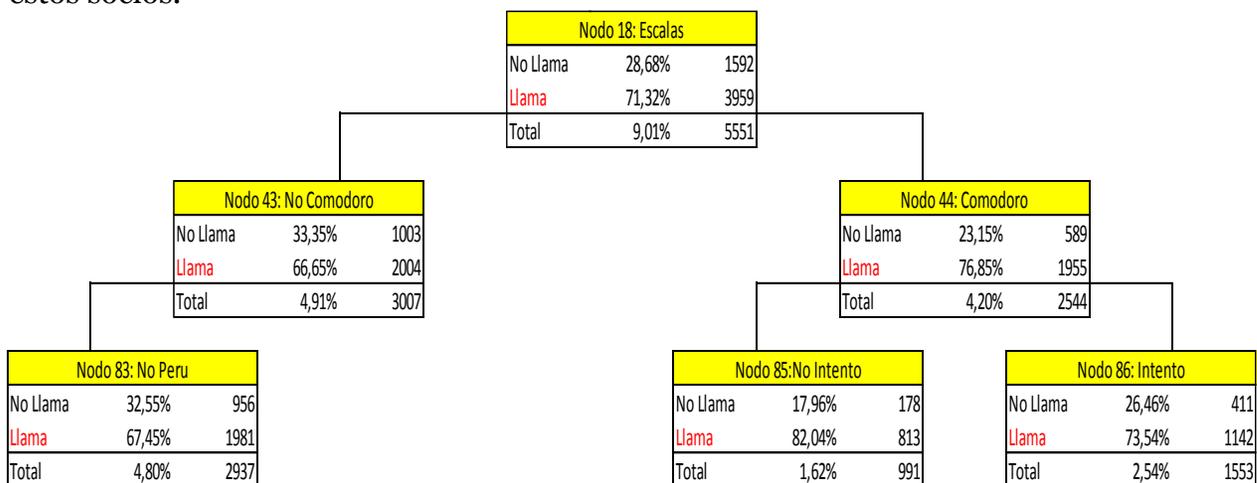
### 10.13 Anexo 13: Análisis del árbol desde el nodo internacional

En la ilustración 13, se muestra el resto del árbol desde el nodo 18, es decir, podemos observar el cuarto y quinto nivel a partir de los vuelos internacionales, con posibilidad de postular a upgrade y vuelos con escalas. En primer lugar, vemos que la primera ramificación, es decir, la variable más significativa para los vuelos con escalas es si el socio pertenece (nodo 44) o no (nodo 43) a la categoría comodoro. De esta se puede apreciar que las llamadas aumentan para ambos grupos, aunque para los socios comodoro, su probabilidad es aproximadamente un 10% mayor. Ahora bien, como ambos grupos

presentan una alta probabilidad de llamada, se hace necesario estudiar el próximo nivel, para poder identificar y focalizar las acciones de una mejor forma.

De esta forma para el nodo 43 la variable más significativa es, si el socio es de nacionalidad peruana. Se puede observar que sólo se presenta el nodo 83, ya que como se mencionó en un principio, los nodos que no tenían un aporte significativo fueron podados del árbol. Por esto mismo, se puede concluir que un socio que tiene un viaje internacional, tiene la posibilidad de postular a upgrade, vuelo con escalas, es categoría black o premium silver y no es de nacionalidad peruana, tienen un 67,45% de probabilidad de llamar al contact center, lo que equivale a 1.981 llamadas, haciendo de este perfil, un candidato a tratar en el diseño de acciones.

Por otro lado, si el socio es comodoro (nodo 44), la variable escogida debido a su significancia, es si el socio intento chequearse por el sitio web. De donde se puede observar que si el socio no lo intento, tiene una probabilidad de un 82,04%, lo que hace mucho sentido, debido a que si no se usa el medio regular para realizar el proceso del “Check-in Upgrade”, la otra opción será usar el servicio del contact center. Ahora si el socio intento chequearse también existe una alta probabilidad de llamar, 73,54%, lo que puede deberse por un lado, a dudas por el viaje, ya que recordemos que este perfil de clientes cuenta con un viaje internacional y con escalas, lo que se hace más complicado en el sentido de los transbordos y distintas medidas para los diferentes países, y por otro, debido a algún error que presentó el sitio web. Además, es importante recalcar que si bien las llamadas que representa este último tipo de socios son menores a las conformadas por el tipo del nodo 83 (1.981 vs 813 vs 1.142), sus probabilidades hacen que sean muy atractivos a la hora de la elección de perfiles a tratar, por la mayor seguridad de reducir los costos que producen estos socios.



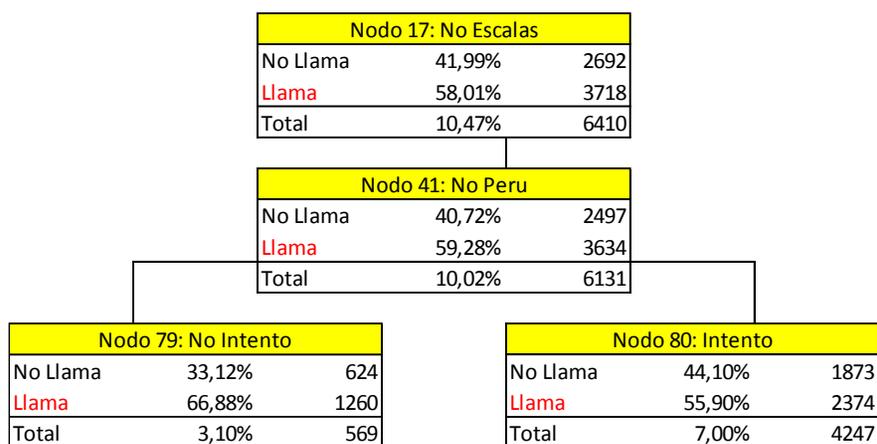
**Ilustración 13: Ramificación desde el nodo 18: Vuelo con escalas.**

**Fuente Elaboración Propia.**

En la ilustración 14, podemos ver la ramificación de los vuelos sin escalas, desde el nodo 6, donde la variable más significativa para realizar la siguiente ramificación es, si el socio es de nacionalidad peruana. Al igual que en la ilustración anterior se pudo la rama que correspondía al nodo de ser peruano. Luego, el nodo 41, que responde a que el socio

no es de nacionalidad peruana se vuelve a dividir según la variable intento (si el socio intento o no chequearse por el sitio web).

De esta parte del árbol se puede concluir que si el cliente no intento chequearse vía web su probabilidad de llamar al contact center es un 10% mayor que si lo intento. Esto puede reflejar que el vuelo al no tener escalas, disminuya la posibilidad de errores por el conducto regular o haya menos dudas del vuelo por parte de los socios. Ahora bien, a la hora de la elección del perfil para tomar acciones, no bastara sólo con mirar la probabilidad de llamada, ya que cómo se ha hecho a lo largo del trabajo con los modelos, que un socio tenga una probabilidad mayor al 50% ya se puede ser considerado como un socio candidato a llamar, por lo mismo, el perfil de clientes que crea el nodo 80, se vuelve relevante, ya que su probabilidad es del 55%, más aún, en cuanto a magnitud de número de llamadas, este nodo es casi el doble que el otro (2.373 vs 1.260), y además este tipo de socio concentra un 7% de la muestra total. Por lo que si se decide enfocar los esfuerzos en este tipo de clientes “Socio con viaje internacional, con posibilidades de postular a upgrade, vuelo sin escalas, que no sea de nacionalidad peruana y que intento chequearse por el sitio web”, podemos lograr un alto impacto en la reducción de costos, el cual es el objetivo final del diseño de acciones para los diferentes tipos de clientes.



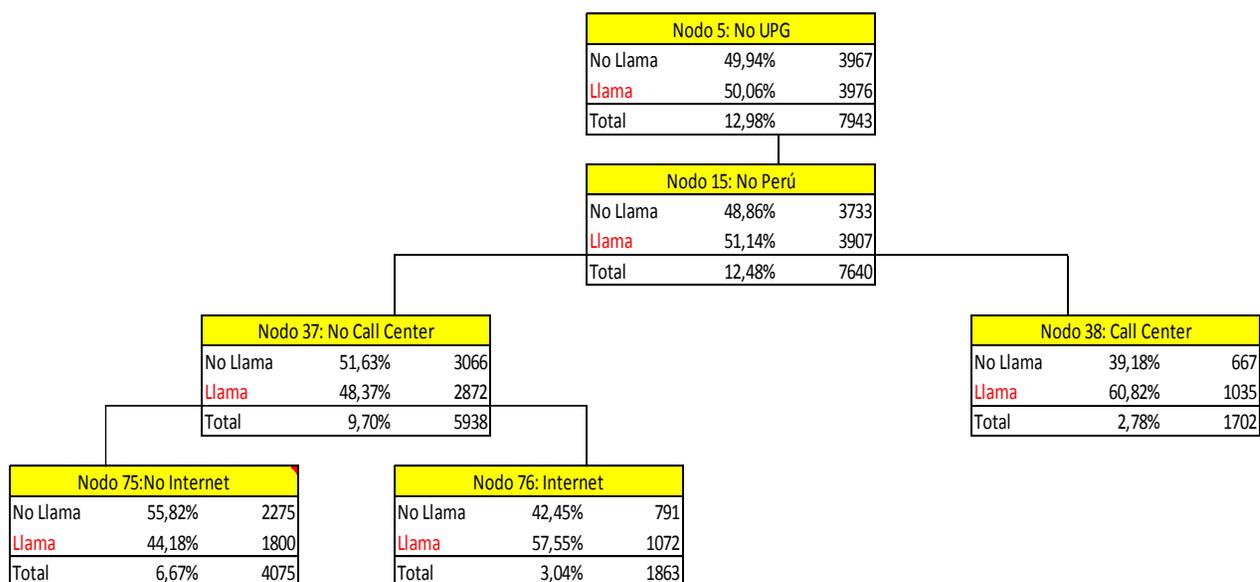
**Ilustración 14: Ramificación desde el nodo 17: Vuelo sin escalas.**

**Fuente Elaboración Propia.**

En la ilustración 15, se puede observar que el nodo 15 (socio no es de nacionalidad peruana), se ramifica según la variable de canal de compra call center, donde el nodo 37 hace referencia a que el socio no ocupó este medio para comprar su pasaje y el nodo 38 muestra a los que sí ocupan dicho canal. Ahora bien, podemos observar que el nodo 38 no tiene un siguiente nivel en comparación al nodo 37, debido a la poda del modelo. Por su parte, el nodo 37, se ramifica según la variable de canal de compra internet. Esto último permite que con los nodos 37, 75 y 76, podemos especificar el canal de compra de este tipo de socios.

Si se observa el nodo 76, podemos notar que los socios que tienen un viaje internacional, sin posibilidad de postular a upgrade, vuelo sin escalas, no son de nacionalidad peruana y compran su pasaje a través de internet, tienen una probabilidad de llamar de un 57,55%, correspondiente a 1.072 llamados. Este perfil de clientes nos da una clara señal de que en el conducto regular (sitio web) para realizar el trámite del “Check-in Upgrade”, tiene algo que no está funcionando, ya que al no tener posibilidades de postular al upgrade y que el vuelo no tenga escalas, hace que el proceso que debe llevar a cabo sea más simple, pero la variable de compra vía internet, significa que el socio tiene manejo y acceso a este servicio, por lo cual la posibilidad de que al momento de realizar el proceso antes mencionado existan errores o dudas por parte socio, hace que deba recurrir al contact center. Por esto mismo estos socios quedan como un candidato potencial a tomar acciones por parte del área operativa del programa.

Finalmente, es importante notar que el nodo 38 es de suma importancia, ya que si un socio compra su pasaje a través del call center tiene una alta probabilidad, 60,82%, de llamar al contact center para realizar su respectivo “Check-in Upgrade”. Esto se puede deber principalmente a que si un socio acostumbra a utilizar este canal para comprar, es porque tiene confianza en él, por lo cual aplicará el mismo medio para realizar todo el proceso. Por lo mismo es un segmento de clientes interesante, ya que existe un desafío preventivo importante desde el área de operaciones, de hacer que estos clientes puedan hacer uso del sitio web, de tal forma de evitar un porcentaje de las llamadas de este perfil de socios.

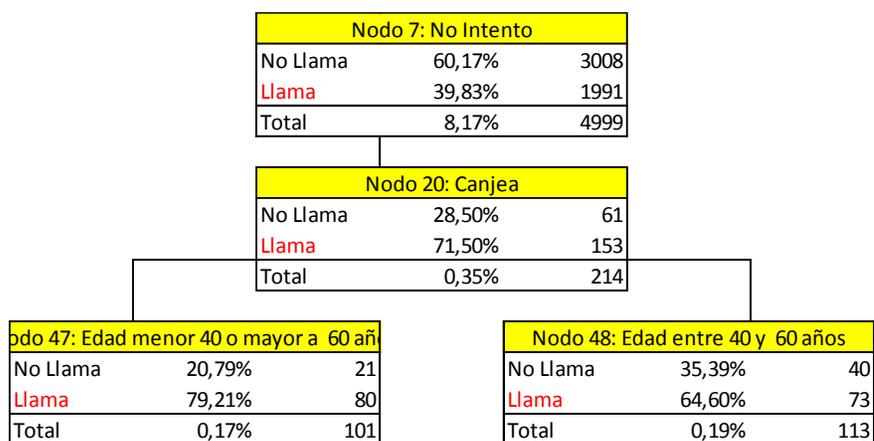


**Ilustración 15: Ramificación del árbol a partir del nodo 5: Vuelo Internacional sin opciones de postular a upgrade.**

**Fuente Elaboración Propia.**

## 10.14 Anexo 14: Análisis del árbol desde el nodo doméstico

La ilustración 16 se quiso mostrar debido a la alta probabilidad de llamar del tipo de socios para el nodo 47 (79,21%) y nodo 48 (64,6%), los cuales se definen como, nodo 47: viaje doméstico, no comodoro, no intenta chequearse por el sitio web, canjea su pasaje y edad menor a 40 o mayor a 60 años. Por otro lado el nodo 48: viaje doméstico, no comodoro, no intenta chequearse por el sitio web, canjea su pasaje y edad entre 40 y 60 años. Ahora bien, hay que notar que la totalidad de estos tipos de socios solo representan un 0,17% y 0,19% respectivamente de la muestra. Además, la cantidad de llamadas son solo 80 y 73 respectivamente, lo cual a la hora de la elección de un perfil de clientes, no tendrá el mismo impacto sobre un perfil donde la cantidad de llamadas sea sobre 1.000 (Notar que tampoco se muestra el nodo 19: No Canjea, debido a su poca participación y probabilidad de llamada).



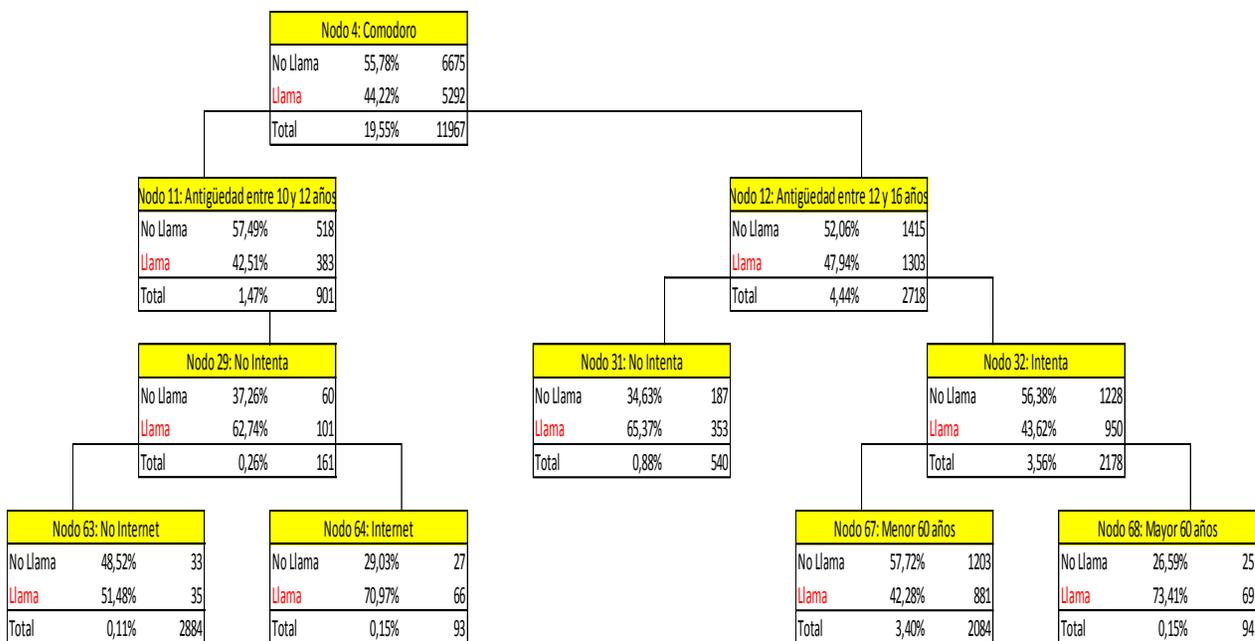
**Ilustración 16: Ramificación del nodo 7 socio no intenta chequearse vía web (nodo padre: socio no es comodoro).**

**Fuente: Elaboración Propia.**

La última ramificación a analizar es la que nace desde el nodo 4, donde el socio tiene viaje doméstico, pero su categoría es comodoro, la cual representa un 19,55% de la base y tienen un 55,78% de no llamar al contact center. Ahora la variable más significativa para seguir clasificando es la antigüedad en el programa, que se analizará en las ilustraciones 17, 18 y 19.

De la ilustración 17 se puede observar que los nodos con mayor probabilidad de llamar al contact center son el nodo 29 con un 62,74% de probabilidad de llamar, el nodo 64 con un 70,97% de probabilidad que a diferencia del nodo 29 sólo agrega que la compra del pasaje fue por internet. También el nodo 31 con un 65,37% de probabilidad y por último el nodo 58 que tiene un 73,41% de probabilidad de llamar.

Sin embargo, estos nodos por un lado presentan una alta probabilidad, pero a su vez representan una baja cantidad de llamadas, lo que nuevamente hace que a la hora de elegir los perfiles respectivos, se tengan en cuenta tanto su evaluación económica como las distintas métricas (lift, ganancia, etc.) para tomar una decisión más acertada, ya que estos podrían significar una reducción mucho más segura que otros casos.



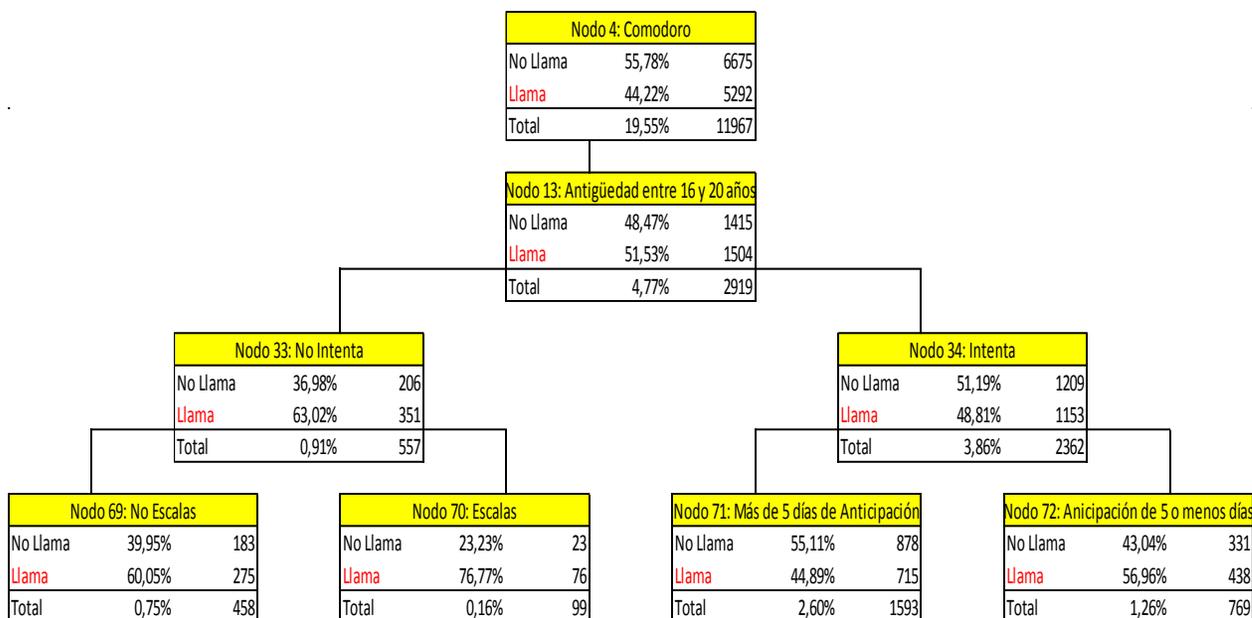
**Ilustración 17: Ramificación nodo 4: socio es de categoría comodoro (nodo padre: Vuelo Doméstico).**

*Fuente: Elaboración Propia.*

En la ilustración 18 podemos observar la variable antigüedad entre 16 y 20 años en el programa, la cual representa una participación del 4,77%, en particular el nodo 13, tiene una cantidad considerable de llamadas (1.505), el cual tiene un 51,53% de probabilidades de llamar al contact center por motivos de “Check-in Upgrade”.

Si seguimos por las demás ramificaciones podemos encontrar perfiles de harto interés respecto a su probabilidad de llamada al contact center. Los siguientes nodos tienen en común que son viajes domésticos, los socios son comodoros y tienen entre 16 y 20 años en el programa, y se diferencian en lo siguiente, el nodo 69 no intenta chequearse vía web y su viaje no contiene escalas, el cual tiene una probabilidad de llamar igual al 60,05%. El nodo 70 no intenta chequearse vía web y su viaje si contiene escalas y su probabilidad asciende a 76,77%. Por último el nodo 72 intenta chequearse vía web y compra su pasaje entre 0 y 5 días de anticipación y la probabilidad que llame por tema de check-in upgrade es igual a 56,96%.

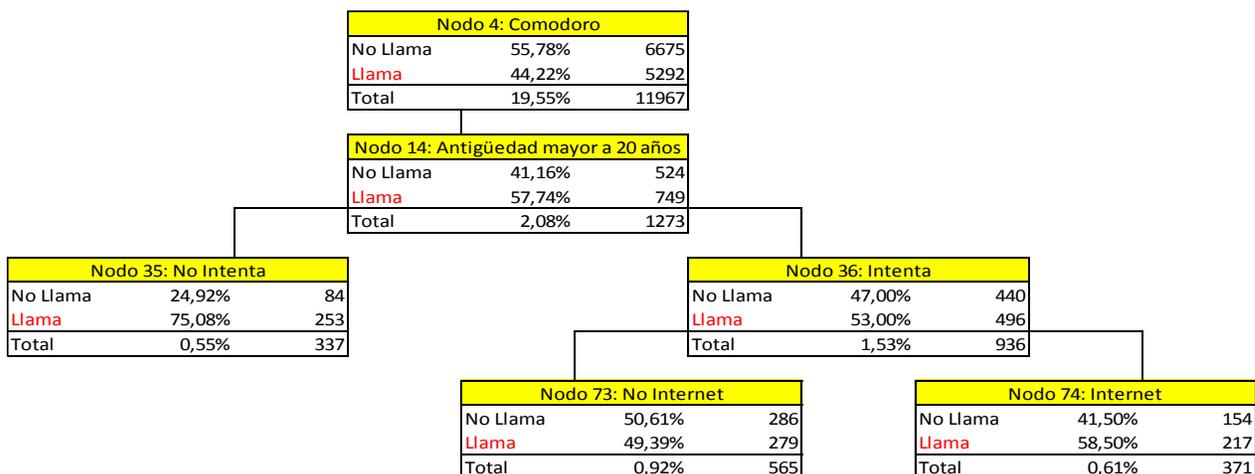
Ahora bien, al igual que en los casos anteriores a la hora de ver los candidatos posibles existirá el trade-off entre la probabilidad de llamada y los casos pertinentes, que habrá que evaluar junto con otras métricas para poder tomar la mejor decisión.



**Ilustración 18: Continuación ramificación nodo 4: socio es de categoría comodoro (nodo padre: Vuelo Doméstico).**

**Fuente: Elaboración Propia.**

Finalmente, en la ilustración 19 podemos ver la última ramificación del nodo 4 respecto a la antigüedad del socio en el programa. De aquí podemos observar, que para los socios que llevan más de 20 años en el programa, el no intentar chequearse por la web hace que tengan un 75,08% de probabilidad de llamar al contact center, representando un 0.55% de la muestra total. Por otro lado para los socios que si intentan chequearse, tenemos que para los que compran su pasaje vía internet su probabilidad de llamada asciende a un 58.5%.



**Ilustración 19: Continuación ramificación nodo 4: socio es de categoría comodoro (nodo padre: Vuelo Doméstico).**

**Fuente: Elaboración Propia.**

## 10.15 Anexo 15: Resumen árbol de decisión

No do	C. 1	C. 2	C. 3	C. 4	C. 5	#Llamada	#NoLlamada	Prob.	%Total	LIFT
85	Internacional	UPG	Escalas	Comodoro	No intento	813	178	82,04%	1,62%	2,07
46	Doméstico	No Comodoro	No intento	Compra Pasaje	Argentino	61	14	81,33%	0,12%	2,05
47	Doméstico	No Comodoro	No intento	Canjeo Pasaje	NO 4160años	80	21	79,21%	0,16%	2,00
44	Internacional	UPG	Escalas	Comodoro		1955	589	76,85%	4,16%	1,94
70	Doméstico	Comodoro	16<=Antigüedad<20	No intento	Escalas	76	23	76,77%	0,16%	1,94
35	Doméstico	Comodoro	Antigüedad>20	No Intento		253	84	75,07%	0,55%	1,89
78	Internacional	No Upg	No Perú	Call Center	Anticipa entre 0y5 días	425	144	74,69%	0,93%	1,88
86	Internacional	UPG	Escalas	Comodoro	Intento	1142	411	73,54%	2,54%	1,86
68	Doméstico	Comodoro	12<=Antigüedad<16	Intento	Mayor 60 años	69	25	73,40%	0,15%	1,85
20	Doméstico	No Comodoro	No intento	Canjeo Pasaje		153	61	71,50%	0,35%	1,80
18	Internacional	UPG	Escalas			3959	1592	71,32%	9,07%	1,80
64	Doméstico	Comodoro	10<=Antigüedad<12	No Intento	Internet	66	27	70,97%	0,15%	1,79
83	Internacional	UPG	Escalas	No Comodoro	No Perú	1981	956	67,45%	4,80%	1,70
79	Internacional	UPG	No Escalas	No Perú	No intento	1260	624	66,88%	3,08%	1,69
43	Internacional	UPG	Escalas	No Comodoro		2004	1003	66,64%	4,91%	1,68
31	Doméstico	Comodoro	12<=Antigüedad<16	No Intento		353	187	65,37%	0,88%	1,65
48	Doméstico	No Comodoro	No intento	Canjeo Pasaje	4160años	73	40	64,60%	0,18%	1,63
52	Doméstico	No Comodoro	Intento	3<Antigüedad<=12	Argentino	107	59	64,46%	0,27%	1,63
6	Internacional	UPG				7677	4284	64,18%	19,54%	1,62
33	Doméstico	Comodoro	16<=Antigüedad<20	No intento		351	206	63,02%	0,91%	1,59
82	Internacional	UPG	No Escalas	Perú	Antigüedad >15	39	23	62,90%	0,10%	1,59
29	Doméstico	Comodoro	10<=Antigüedad<12	No Intento		101	60	62,73%	0,26%	1,58
61	Doméstico	Comodoro	3<=Antigüedad<10	Intento	No Chileno	70	43	61,95%	0,18%	1,56
56	Doméstico	No Comodoro	Intento	Antigüedad >20	Black	52	32	61,90%	0,14%	1,56
38	Internacional	No Upg	No Perú	Call Center		1035	667	60,81%	2,78%	1,53

69	Doméstico	Comodoro	16<=Antigüedad<20	No intento	No escalas	275	183	60,04%	0,75%	1,51
54	Doméstico	No Comodoro	Intento	12<Antigüedad<=20	Black	126	85	59,72%	0,34%	1,51
41	Internacional	UPG	No Escalas	No Perú		3634	2497	59,27%	10,02%	1,50
59	Doméstico	Comodoro	3<=Antigüedad<10	No intento	No internet	186	130	58,86%	0,52%	1,48
14	Doméstico	Comodoro	Antigüedad>20			749	524	58,84%	2,08%	1,48
2	Internacional					11653	8251	58,55%	32,51%	1,48
74	Doméstico	Comodoro	Antigüedad>20	Intento	Internet	217	154	58,49%	0,61%	1,48
17	Internacional	UPG	No Escalas			3718	2692	58,00%	10,47%	1,46
76	Internacional	No Upg	No Perú	No Call Center	Internet	1072	791	57,54%	3,04%	1,45
72	Doméstico	Comodoro	16<=Antigüedad<20	Intento	Anticipa entre 0 y 5 días	438	331	56,96%	1,26%	1,44
80	Internacional	UPG	No Escalas	No Perú	Intento	2374	1873	55,90%	6,94%	1,41
77	Internacional	No Upg	No Perú	Call Center	No Anticipa entre 0 y 5 días	610	523	53,84%	1,85%	1,36
36	Doméstico	Comodoro	Antigüedad>20	Intento		496	440	52,99%	1,53%	1,34
27	Doméstico	Comodoro	3<=Antigüedad<10	No intento		267	241	52,56%	0,83%	1,33
13	Doméstico	Comodoro	16<=Antigüedad<20			1504	1415	51,52%	4,77%	1,30
63	Doméstico	Comodoro	10<=Antigüedad<12	No Intento	No internet	35	33	51,47%	0,11%	1,30
15	Internacional	No Upg	No Perú			3907	3733	51,14%	12,48%	1,29
5	Internacional	No Upg				3976	3967	50,06%	12,98%	1,26
73	Doméstico	Comodoro	Antigüedad>20	Intento	No Internet	279	286	49,38%	0,92%	1,25
34	Doméstico	Comodoro	16<=Antigüedad<20	Intento		1153	1209	48,81%	3,86%	1,23
37	Internacional	No Upg	No Perú	No Call Center		2872	3066	48,37%	9,70%	1,22
12	Doméstico	Comodoro	12<=Antigüedad<16			1303	1415	47,94%	4,44%	1,21
71	Doméstico	Comodoro	16<=Antigüedad<20	Intento	No Anticipa entre 0 y 5 días	715	878	44,88%	2,60%	1,13
4	Doméstico	Comodoro				5292	6675	44,22%	19,55%	1,12
75	Internacional	No Upg	No Perú	No Call Center	No internet	1800	2275	44,17%	6,66%	1,11
32	Doméstico	Comodoro	12<=Antigüedad<16	Intento		950	1228	43,62%	3,56%	1,10
11	Doméstico	Comodoro	10<=Antigüedad<12			383	518	42,51%	1,47%	1,07

67	Doméstico	Comodoro	12<=Antigüedad<16	Intento	NO Mayor 60 años	881	1203	42,27%	3,40%	1,07
60	Doméstico	Comodoro	3<=Antigüedad<10	No intento	Internet	81	111	42,19%	0,31%	1,06
65	Doméstico	Comodoro	10<=Antigüedad<12	Intento	No Escalas	254	383	39,87%	1,04%	1,01
7	Doméstico	No Comodoro	No intento			1991	3008	39,83%	8,17%	1,00
19	Doméstico	No Comodoro	No intento	Compro Pasaje		1838	2947	38,41%	7,82%	0,97
30	Doméstico	Comodoro	10<=Antigüedad<12	Intento		282	458	38,11%	1,21%	0,96
45	Doméstico	No Comodoro	No intento	Compro Pasaje	No Argentino	1777	2933	37,73%	7,69%	0,95
10	Doméstico	Comodoro	3<=Antigüedad<10			1206	2299	34,41%	5,73%	0,87
40	Internacional	No Upg	Perú	Call Center		27	52	34,18%	0,13%	0,86
24	Doméstico	No Comodoro	Intento	Antigüedad >20		410	808	33,66%	1,99%	0,85
25	Doméstico	Comodoro	Antigüedad<3	No intento		30	61	32,97%	0,15%	0,83
84	Internacional	UPG	Escalas	No Comodoro	Perú	23	47	32,86%	0,11%	0,83
50	Doméstico	No Comodoro	Intento	Antigüedad <3	Canjeo Pasaje	18	38	32,14%	0,09%	0,81
55	Doméstico	No Comodoro	Intento	Antigüedad >20	No Black	358	776	31,57%	1,85%	0,80
28	Doméstico	Comodoro	3<=Antigüedad<10	Intento		939	2058	31,33%	4,90%	0,79
1	Doméstico					12611	28700	30,53%	67,49%	0,77
62	Doméstico	Comodoro	3<=Antigüedad<10	Intento	Chileno	869	2015	30,13%	4,71%	0,76
42	Internacional	UPG	No Escalas	Perú		84	195	30,11%	0,46%	0,76
66	Doméstico	Comodoro	10<=Antigüedad<12	Intento	Escalas	28	75	27,18%	0,17%	0,69
23	Doméstico	No Comodoro	Intento	12<Antigüedad<=20		2270	6270	26,58%	13,95%	0,67
58	Doméstico	Comodoro	Antigüedad<3	Intento	Internet	57	164	25,79%	0,36%	0,65
53	Doméstico	No Comodoro	Intento	12<Antigüedad<=20	NO Black	2144	6185	25,74%	13,61%	0,65
3	Doméstico	No Comodoro				7319	22025	24,94%	47,94%	0,63
16	Internacional	No Upg	Perú			69	234	22,77%	0,49%	0,57
9	Doméstico	Comodoro	Antigüedad<3			147	504	22,58%	1,06%	0,57
8	Doméstico	No Comodoro	Intento			5328	19017	21,89%	39,77%	0,55

26	Doméstico	Comodoro	Antigüedad<3	Intento		117	443	20,89%	0,91%	0,53
81	Internacional	UPG	No Escalas	Perú	Antigüedad <=15	45	172	20,74%	0,35%	0,52
22	Doméstico	No Comodoro	Intento	3<Antigüedad<=12		2237	9095	19,74%	18,51%	0,50
51	Doméstico	No Comodoro	Intento	3<Antigüedad<=12	No Argentino	2130	9036	19,08%	18,24%	0,48
39	Internacional	No Upg	Perú	No Call Center		42	182	18,75%	0,37%	0,47
57	Doméstico	Comodoro	Antigüedad<3	Intento	No internet	60	279	17,70%	0,55%	0,45
21	Doméstico	No Comodoro	Intento	Antigüedad <3		411	2844	12,63%	5,32%	0,32
49	Doméstico	No Comodoro	Intento	Antigüedad <3	Compro Pasaje	393	2806	12,29%	5,23%	0,31

**FUENTE: ELABORACIÓN PROPIA**

### 10.16 Anexo 16: Resumen perfiles del árbol de decisión en base total

Nodo	C.1	C.2	C.3	C.4	C.5	#Llamada	#NoLlamada	Prob.	%Total	LIFT
85	Internacional	UPG	Escalas	Comodoro	No intento	1700	4493	27,45%	0,66%	5,26
70	Doméstico	Comodoro	16<=Antigüedad<20	No intento	Escalas	159	513	23,66%	0,07%	4,53
44	Internacional	UPG	Escalas	Comodoro		4007	14770	21,34%	2,01%	4,09
68	Doméstico	Comodoro	12<=Antigüedad<16	Intento	Mayor 60 años	145	545	21,01%	0,07%	4,02
46	Doméstico	No Comodoro	No intento	Compro Pasaje	Argentino	132	546	19,47%	0,07%	3,73
47	Doméstico	No Comodoro	No intento	Canjeo Pasaje	NO 4160años	148	625	19,15%	0,08%	3,67
78	Internacional	No Upg	No Perú	Call Center	Anticipa entre 0y5 días	883	3767	18,99%	0,50%	3,64
35	Doméstico	Comodoro	Antigüedad>20	No Intento		541	2372	18,57%	0,31%	3,56

86	Internacional	UPG	Escalas	Comodoro	Intento	2307	10277	18,33%	1,35%	3,51
20	Domestico	No Comodoro	No intento	Canjeo Pasaje		277	1315	17,40%	0,17%	3,33
18	Internacional	UPG	Escalas			8104	38873	17,25%	5,02%	3,30
48	Domestico	No Comodoro	No intento	Canjeo Pasaje	4160años	129	690	15,75%	0,09%	3,02
83	Internacional	UPG	Escalas	No Comodoro	No Perú	4058	22954	15,02%	2,89%	2,88
79	Internacional	UPG	No Escalas	No Perú	No intento	2548	14718	14,76%	1,85%	2,83
43	Internacional	UPG	Escalas	No Comodoro		4097	24103	14,53%	3,02%	2,78
64	Domestico	Comodoro	10<=Antigüedad<12	No Intento	Internet	127	786	13,91%	0,10%	2,66
31	Domestico	Comodoro	12<=Antigüedad<16	No Intento		687	4315	13,73%	0,53%	2,63
33	Domestico	Comodoro	16<=Antigüedad<20	No intento		690	4349	13,69%	0,54%	2,62
6	Internacional	UPG				15592	104656	12,97%	12,86%	2,48
52	Domestico	No Comodoro	Intento	3<Antigüedad<=12	Argentino	219	1486	12,84%	0,18%	2,46
29	Domestico	Comodoro	10<=Antigüedad<12	No Intento		207	1419	12,73%	0,17%	2,44
69	Domestico	Comodoro	16<=Antigüedad<20	No intento	No escalas	531	3836	12,16%	0,47%	2,33
38	Internacional	No Upg	No Perú	Call Center		2109	16171	11,54%	1,95%	2,21
63	Domestico	Comodoro	10<=Antigüedad<12	No Intento	No internet	80	633	11,22%	0,08%	2,15
41	Internacional	UPG	No Escalas	No Perú		7318	60695	10,76%	7,27%	2,06
59	Domestico	Comodoro	3<=Antigüedad<10	No intento	No internet	362	3008	10,74%	0,36%	2,06
2	Internacional					23648	201770	10,49%	24,11%	2,01
61	Domestico	Comodoro	3<=Antigüedad<10	Intento	No Chileno	141	1207	10,46%	0,14%	2,00
17	Internacional	UPG	No Escalas			7488	65783	10,22%	7,84%	1,96
56	Domestico	No Comodoro	Intento	Antigüedad>20	Black	92	825	10,03%	0,10%	1,92
76	Internacional	No Upg	No Perú	No Call Center	Internet	2195	19877	9,94%	2,36%	1,90
14	Domestico	Comodoro	Antigüedad>20			1731	15739	9,91%	1,87%	1,90
54	Domestico	No Comodoro	Intento	12<Antigüedad<=20	Black	73	665	9,89%	0,08%	1,89
72	Domestico	Comodoro	16<=Antigüedad<20	Intento	Anticipa entre 0 y 5 días	812	7402	9,89%	0,88%	1,89
80	Internacional	UPG	No Escalas	No Perú	Intento	4770	45977	9,40%	5,43%	1,80
77	Internacional	No Upg	No Perú	Call Center	No Anticipa entre 0y5 días	1226	12404	8,99%	1,46%	1,72
27	Domestico	Comodoro	3<=Antigüedad<10	No intento		554	5744	8,80%	0,67%	1,68

74	Domestic	Comodoro	Antigüedad >20	Intento	Internet	502	5279	8,68%	0,62%	1,66
13	Domestic	Comodoro	16<=Antigüedad<20			2862	31653	8,29%	3,69%	1,59
36	Domestic	Comodoro	Antigüedad >20	Intento		1190	13367	8,17%	1,56%	1,57
15	Internacional	No Upg	No Perú			7915	91129	7,99%	10,59%	1,53
73	Domestic	Comodoro	Antigüedad >20	Intento	No Internet	688	8088	7,84%	0,94%	1,50
5	Internacional	No Upg				8056	97114	7,66%	11,25%	1,47
34	Domestic	Comodoro	16<=Antigüedad<20	Intento		2172	27304	7,37%	3,15%	1,41
37	Internacional	No Upg	No Perú	No Call Center		5806	74958	7,19%	8,64%	1,38
12	Domestic	Comodoro	12<=Antigüedad<16			2543	33034	7,15%	3,80%	1,37
60	Domestic	Comodoro	3<=Antigüedad<10	No intento	Internet	192	2736	6,56%	0,31%	1,26
71	Domestic	Comodoro	16<=Antigüedad<20	Intento	No Anticipa entre 0y5 días	1360	19902	6,40%	2,27%	1,23
4	Domestic	Comodoro				10637	158912	6,27%	18,13%	1,20
82	Internacional	UPG	No Escalas	Perú	Antigüedad >15	77	1165	6,20%	0,13%	1,19
75	Internacional	No Upg	No Perú	No Call Center	No internet	3611	55081	6,15%	6,28%	1,18
32	Domestic	Comodoro	12<=Antigüedad<16	Intento		1856	28719	6,07%	3,27%	1,16
11	Domestic	Comodoro	10<=Antigüedad<12			764	12091	5,94%	1,37%	1,14
67	Domestic	Comodoro	12<=Antigüedad<16	Intento	NO Mayor 60años	1711	28174	5,73%	3,20%	1,10
25	Domestic	Comodoro	Antigüedad <3	No intento		75	1294	5,48%	0,15%	1,05
7	Domestic	No Comodoro	No intento			3862	70796	5,17%	7,98%	0,99
65	Domestic	Comodoro	10<=Antigüedad<12	Intento	No Escalas	475	8956	5,04%	1,01%	0,96
30	Domestic	Comodoro	10<=Antigüedad<12	Intento		557	10672	4,96%	1,20%	0,95
19	Domestic	No Comodoro	No intento	Compro Pasaje		3585	69481	4,91%	7,81%	0,94
45	Domestic	No Comodoro	No intento	Compro Pasaje	No Argentino	3453	68935	4,77%	7,74%	0,91
66	Domestic	Comodoro	10<=Antigüedad<12	Intento	Escalas	82	1716	4,56%	0,19%	0,87
40	Internacional	No Upg	Perú	Call Center		47	1062	4,24%	0,12%	0,81
10	Domestic	Comodoro	3<=Antigüedad<10			2413	54709	4,22%	6,11%	0,81
24	Domestic	No Comodoro	Intento	Antigüedad >20		945	24538	3,71%	2,73%	0,71
28	Domestic	Comodoro	3<=Antigüedad<10	Intento		1859	48965	3,66%	5,44%	0,70
1	Domestic					25172	684472	3,55%	75,89%	0,68
62	Domestic	Comodoro	3<=Antigüedad<10	Intento	Chileno	1718	47758	3,47%	5,29%	0,67
84	Internacional	UPG	Escalas	No Comodoro	Perú	39	1149	3,28%	0,13%	0,63
50	Domestic	No Comodoro	Intento	Antigüedad <3	Canjeo Pasaje	36	1076	3,24%	0,12%	0,62

42	Internacional	UPG	No Escalas	Perú		170	5088	3,23%	0,56%	0,62
58	Doméstico	Comodoro	Antigüedad <3	Intento	Internet	115	3572	3,12%	0,39%	0,60
23	Doméstico	No Comodoro	Intento	12 < Antigüedad <= 20		4363	147275	2,88%	16,22%	0,55
9	Doméstico	Comodoro	Antigüedad <3			324	11686	2,70%	1,28%	0,52
3	Doméstico	No Comodoro				14535	525560	2,69%	57,76%	0,52
26	Doméstico	Comodoro	Antigüedad <3	Intento		249	10392	2,34%	1,14%	0,45
81	Internacional	UPG	No Escalas	Perú	Antigüedad >15	93	3923	2,32%	0,43%	0,44
16	Internacional	No Upg	Perú			141	5985	2,30%	0,66%	0,44
8	Doméstico	No Comodoro	Intento			10673	454764	2,29%	49,78%	0,44
22	Doméstico	No Comodoro	Intento	3 < Antigüedad <= 12		4530	214916	2,06%	23,47%	0,40
53	Doméstico	No Comodoro	Intento	12 < Antigüedad <= 20	NO Black	4457	214251	2,04%	23,39%	0,39
51	Doméstico	No Comodoro	Intento	3 < Antigüedad <= 12	No Argentino	4311	213430	1,98%	23,29%	0,38
57	Doméstico	Comodoro	Antigüedad <3	Intento	No internet	134	6820	1,93%	0,74%	0,37
39	Internacional	No Upg	Perú	No Call Center		94	4923	1,87%	0,54%	0,36
21	Doméstico	No Comodoro	Intento	Antigüedad <3		835	68035	1,21%	7,37%	0,23
55	Doméstico	No Comodoro	Intento	Antigüedad >20	No Black	831	67975	1,21%	7,36%	0,23
49	Doméstico	No Comodoro	Intento	Antigüedad <3	Compro Pasaje	799	66959	1,18%	7,25%	0,23

**FUENTE: ELABORACIÓN PROPIA**