



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

ANÁLISIS DE REINCIDENCIA DEL PROGRAMA VOLVER A EMPEZAR
UTILIZANDO REGRESIÓN LOGÍSTICA, RANDOM FOREST Y XGBOOST

TESIS PARA OPTAR AL GRADO DE MAGISTER EN GESTIÓN Y POLÍTICAS
PÚBLICAS
MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

MATÍAS ROBERTO BUSTAMANTE ALARCÓN

PROFESOR GUÍA:
FERNANDO ORDOÑEZ PIZARRO

MIEMBROS DE LA COMISIÓN:
GUILMA ESPINOZA MAVILA
ALEJANDRO AREVALO SARCE

SANTIAGO DE CHILE
2023

ANÁLISIS DE REINCIDENCIA DEL PROGRAMA VOLVER A EMPEZAR UTILIZANDO REGRESIÓN LOGÍSTICA, RANDOM FOREST Y XGBOOST

La presente investigación se centra en analizar y evaluar los resultados del programa Volver a Empezar (VAE) implementado a partir del año 2017 por la Seremi de Justicia de la Región Metropolitana, el cual buscó introducir modelo de gestión de casos de reinserción social en las diez comunas de la región metropolitana con los mayores índices de delincuencia. Este programa presenta un sistema de intervención basado en el modelo de RNR (Riesgo, Necesidad y capacidad de Respuesta) formulado por Andrews y Bonta en el año 1990. El programa Volver a Empezar posee una evaluación previa, realizada a fines del año 2018 por la Universidad Católica y el Centro de Estudios Justicia y Sociedad, que si bien, analiza aspectos fundamentales de la implementación, posee poca profundidad en aspectos cuantitativos vinculados a la medición del efecto de este programa sobre las tasas de reincidencia, esto debido a la cercanía de esta evaluación con la fecha de término del programa.

La pregunta de investigación establecida para esta tesis es *¿Qué características de los usuarios y de la implementación del Programa Volver a Empezar fueron claves para prevenir la reincidencia?* Por lo que se busca principalmente establecer a través de metodologías cuantitativas cuales son las principales variables para tener en consideración en el diseño o rediseño de programas de reinserción social de características similares a las del VAE o que estén basados en el modelo RNR o en su defecto en el Manual de procedimientos del VAE. Las metodologías que fueron utilizadas fueron seleccionadas a partir de su adaptabilidad a los datos obtenidos, tales metodologías fueron la regresión logística de Cox o también llamada regresión logística multivariada, más dos metodologías basadas en arboles de decisión, las que son Random Forest y XGBost, estas metodologías permiten obtener estimaciones de las importancias relativas de las variables, utilizando modelos aleatorios y modelos iterativos.

Los resultados indican que, por una parte, el cumplimiento de los planes de reinserción establecidos en el programa Volver a Empezar poseen muy poca relevancia ante el resto de las variables estudiadas, lo cual resulta sumamente alarmante, debido a que es esperable que sea esta una variable clave sobre la reincidencia de los usuarios, lo que a su vez revela serios problemas de gestión de datos dentro del programa. Por otra parte, los resultados muestran cuatro variables claves, las que son la edad de los usuarios, el tamaño de sus redes de apoyo, su nivel educacional y a el acceso al trabajo durante su paso por el programa, la relevancia de estas variables se encuentra respaldada en la literatura, pero se evidencia la necesidad de complementarlas con información que permita una mejor interpretación, como por ejemplo la calidad de los vínculos de sus redes de apoyo.

Tanto que agradecer y tan poco espacio...

Agradezco a mi madre, Carmen, por formarme como persona y entregarme las herramientas y los valores que me han permitido afrontar todos los desafíos con los que me he encontrado.

Agradezco a Savka por ser mi soporte en este proceso, por siempre alentarme y nunca permitirme bajar los brazos.

Agradezco a Tomás y Rodrigo, mis hermanos, por inspirarme y motivarme todos los días a ser una mejor persona. También agradezco a mi Nina que siempre nos ha entregado su amor incondicional.

Agradezco a todos mis amigos y a esas personas importantes que han cruzado por mi vida. A mis bachis y a mis PRs, a Camila, Ignacio, Bryan, Valentina y Rodrigo, que sin ellos recorrer este largo camino no hubiese sido posible. Agradezco especialmente a Mirko por su disposición y cariño al apoyarme en el análisis de datos.

Agradezco a mi padre, Ricardo, por apoyar mi ingreso al magister y darme el respaldo que necesité y agradezco a su padre, Rogelio, quien para mi y mis hermanos siempre ha sido un modelo a seguir.

Agradezco a mi profesor guía, Fernando Ordoñez, por su apoyo, paciencia y comprensión a lo largo de este proceso y a Alejandro Arévalo de Gendarmería por la ayuda entregada en el momento más necesario.

Agradezco a Colo Colo, que para los colocolinos representa más que un club de fútbol, es una manera de ver la vida, es una postura política. También agradezco a Marcelo Barticciotto por ser un ejemplo como persona, como colocolino y por darnos la Libertadores.

Finalmente quiero dedicar este trabajo realizado con mucho esfuerzo y mucho cariño a todas las personas, hombres, mujeres y jóvenes que se encuentran siendo procesados, cumpliendo una condena o en postpenitenciario, a los niños y jóvenes de poblaciones, con infancias vulneradas que probablemente emprendan carreras delictivas, a las personas víctimas de la delincuencia y a quienes viven con temor a ella, a todos ellos les dedico este trabajo, porque la reinserción social es un proceso necesario para permitirnos sanar como sociedad.

TABLA DE CONTENIDO

1. Introducción	1
2. Pregunta de investigación, objetivos y resultados esperados.....	4
2.1. Pregunta de investigación	4
2.2. Objetivo general	4
2.3. Objetivos específicos	4
2.4. Resultados esperados.....	4
3. Marco Conceptual.....	5
3.1. Concepto de Reinserción Social	5
3.2. Concepto de Reincidencia	7
3.3. ¿Cómo medir la reincidencia?.....	9
3.4. Experiencia internacional	11
3.4.1. Inglaterra	12
3.4.2. Canadá	12
3.4.3. Estados Unidos.....	13
3.5. Factores de riesgo asociados a la medición de reincidencia	14
3.6. Modelos considerados	15
3.6.1. Análisis prospectivo	15
3.6.2. Análisis Retrospectivo	17
3.7. Fundamentos para la elección del modelo a utilizar	20
3.8. Programa Volver a Empezar	22
3.8.1. Modelo Riesgo Necesidad y capacidad de Respuesta (RNR).....	22
3.8.2. Implementación	24
3.8.3. Análisis del nivel de cumplimiento de los objetivos del Programa Volver a Empezar	28
4. Metodología.....	33
4.1. Tratamiento de datos	33
4.1.1. Variables dependientes:	35
4.2. Regresión logística.....	36
4.3. Random Forest	40
4.4. XGBoost.....	41
4.5. Random Forest vs XGBoost.....	41
5. Análisis exploratorio de datos	43

5.1.	Valores perdidos	43
5.1.1.	Edad	43
5.1.2.	Género.....	45
5.1.3.	Profesión	46
5.1.4.	Nivel educacional.....	48
5.1.5.	Comuna	49
5.1.6.	Reincidente.....	50
5.1.7.	Grupo de referencia.....	50
5.1.8.	Cumplimiento del plan de reinserción	51
5.1.9.	Colocado laboralmente si lo requiere	52
5.2.	Distribución condicionada de la variable dependiente	53
5.2.1.	Reincidencia según edad	54
5.2.2.	Reincidencia según tamaño del grupo de referencia.....	55
5.2.3.	Reincidencia según colocación laboral.....	56
5.2.4.	Nivel educacional.....	57
5.2.5.	Reincidencia según comuna.....	58
5.2.6.	Reincidencia según género	59
5.2.7.	Reincidencia según variable Profesión.....	60
5.2.8.	Reincidencia según cumplimiento del programa de reinserción social 61	
6.	Resultados.....	62
6.1.	Regresión logística multivariada	62
6.2.	Importancias con Random Forest y XGBoost	64
7.	Análisis	68
7.1.	Regresión logística.....	68
7.2.	Análisis de importancias con Random Forest y XGBoost	69
7.2.1.	La variable más importante	69
7.2.2.	Variables de alta importancia	70
7.2.3.	Variables de mediana importancia.....	71
7.2.4.	Variables de menor importancia	71
8.	Conclusiones	72
9.	Bibliografía.....	76
	Anexos	80

Anexo A: Listado de variables.....	80
Anexo B: Resultados regresión logística multivariada	84
Anexo C: Importancias para XGBoost	85
Anexo D: Importancias para Random Forest	86

INDICE DE TABLAS

Tabla 1 - Ciclos de Capacitación VAE (elaboración externa)	25
Tabla 2 - Reincidencia según criterios (elaboración externa).....	29
Tabla 3 - Continuidad del programa VAE (elaboración externa)	31
Tabla 4 - Valores perdidos por variable (elaboración propia).....	35
Tabla 5 - Medidas de tendencia central, variable Edad (elaboración propia)	43
Tabla 6 - Distribución de la variable edad (elaboración propia).....	45
Tabla 7 - Distribución de la variable comuna, antes y después de la limpieza de datos (elaboración propia).....	49
Tabla 8 - Medidas de tendencia central para la variable comuna, antes y después de la limpieza de datos (elaboración propia)	50
Tabla 9 - Reincidencia según tramo etario (elaboración propia)	54
Tabla 10 - Reincidencia según tamaño del grupo de referencia (elaboración propia)	56
Tabla 11 - Reincidencia según variable colocado laboralmente (elaboración propia)	57
Tabla 12 - Reincidencia según nivel educacional (elaboración propia)	57
Tabla 13 - Reincidencia según comuna (elaboración propia)	59
Tabla 14 - Reincidencia según cumplimiento del programa de reinserción (elaboración propia)	61
Tabla 15 - Resultados regresión logística (elaboración propia).....	63
Tabla 16 - Estadísticos Regresión logística.....	64
Tabla 17 - Importancias RF con función de pérdidas MSE (elaboración propia) ..	64
Tabla 18 - Importancias RF con función de pérdidas NodePurity (elaboración propia)	65
Tabla 19 - Importancias XGB (elaboración propia)	65
Tabla 20 - Importancias finales normalizadas (elaboración propia)	66
Tabla 21 - Resumen variables significativas (elaboración propia).....	69

INDICE DE FIGURAS

Figura 1 - Flujograma de atenciones (Elaboración externa).....	27
--	----

INDICE DE GRÁFICOS

Gráfico 1 - Encuesta utilidad del programa (elaboración externa).....	30
---	----

Gráfico 2 - Encuesta evaluación del programa (elaboración externa).....	31
Gráfico 3 - Matriz de correlación (elaboración propia).....	35
Gráfico 4 - Comparación Random Forest vs XGBoost (elaboración propia)	42
Gráfico 5 - Histograma variable edad, previo a la limpieza de datos (elaboración propia)	44
Gráfico 6 - Histograma variable edad, posterior a la limpieza de datos (elaboración propia)	44
Gráfico 7 - Distribución de la variable género, previo a la limpieza de datos (elaboración propia)	45
Gráfico 8 - Distribución de la variable género, posterior a la limpieza de datos (elaboración propia)	46
Gráfico 9 - Distribución de la variable profesión, previo a la limpieza de datos (elaboración propia)	47
Gráfico 10 - Distribución de la variable profesión, posterior a la limpieza de datos (elaboración propia)	47
Gráfico 11 - Distribución de la variable nivel educacional, antes y después de la limpieza de datos (elaboración propia).....	48
Gráfico 12 - Distribución de la variable grupo de referencia, antes y después de la limpieza de datos (elaboración propia).....	51
Gráfico 13 - Distribución de la variable cumplimiento del plan de reinserción, previo a la limpieza de datos (elaboración propia).....	52
Gráfico 14 - Distribución de la variable cumplimiento del plan de reinserción, posterior a la limpieza de datos (elaboración propia)	52
Gráfico 15 - Distribución de la variable colocado laboralmente, antes y después de la limpieza de datos (elaboración propia).....	53
Gráfico 16 - Distribución de la variable dependiente (reincidencia) (elaboración propia)	54
Gráfico 17 - Reincidencia según tramo etario (elaboración propia).....	55
Gráfico 18 - Reincidencia según tamaño del grupo de referencia (elaboración propia)	56
Gráfico 19 - Reincidencia según nivel educacional (elaboración propia)	58
Gráfico 20 - - Reincidencia según comuna (elaboración propia).....	59
Gráfico 21 - Reincidencia según género (elaboración propia).....	60
Gráfico 22 - Reincidencia según profesión (elaboración propia)	60

INDICE DE CÓDIGOS

Código 1 - Regresión logística	63
Código 2 - Random Forest y XGBoost	67

1. Introducción

La delincuencia es un fenómeno social que constituye un problema de interés público a nivel mundial. Dos de los indicadores más comunes para medir el nivel de seguridad pública por parte de los estados son la tasa de victimización y la percepción de exposición al delito. Según el INE durante 2019 la tasa de *victimización personal* fue de un 8.9%, lo cual es menor al 9.3% del año anterior, pese a esto la *percepción de aumento de la delincuencia y la exposición al delito* en el país fue de un 82%, mayor al 76.8% de 2018, cerca del 50% de las personas encuestadas señalaron que la principal fuente de información con la sustentan su opinión es la televisión (Instituto Nacional de Estadística, 2020).

En cuanto al fenómeno del delito, este se ve reflejado en la cantidad de detenciones que realizan las policías, la *Policía de Investigaciones (PDI)* detuvo un total de 49.801 personas en 2019, mientras que *Carabineros* detuvo 471.535 personas, lo que representa una disminución de un 14% y un alza de un 1,6%. Mientras por un lado las detenciones por parte de la PDI han presentado una disminución sostenida durante los últimos 5 años, en Carabineros la cifra de personas detenidas se ha mantenido sin cambios significativos (Instituto Nacional de Estadísticas, 2021).

La cárcel es sin duda el tratamiento más común para el problema de la delincuencia. En Chile, la institución encargada de la custodia de las personas que cumplen condena es Gendarmería, la cual divide sus funciones en tres subsistemas, los cuales son *abierto, cerrado y postpenitenciario*. El subsistema cerrado es el que concentra a la mayor parte de la población penitenciaria, 45.422 personas correspondientes al 40.7%, seguido por el subsistema abierto compuesto por 44.252 personas, igual al 39.7% del total, por último 21.845 personas, equivalentes al 19.6% forman parte del subsistema postpenitenciario (Gendarmería de Chile, 2020). Con respecto al subsistema cerrado no hay mayores variaciones en cuanto a la población, de hecho, en el mes de enero hubo 3.428 ingresos y 3.500 egresos, generando una disminución de 72 personas a nivel nacional.

Pese a que la idea de combatir delitos con cárcel puede parecer completamente lógica y correcta, no es una verdadera solución del problema. La estrategia de aislar a las personas que cometen delitos no elimina el fenómeno de la delincuencia, pese a que Chile es el segundo país de la OCDE con la *mayor tasa de encarcelamiento* (266 personas privadas de libertad por cada 100.000 habitantes), no logra eliminar la delincuencia, llegando a una *tasa de reincidencia* del 52.9%.

Dentro del *Decreto Ley 2859* que fija la Ley Orgánica de Gendarmería, se señala que uno de los principales fines de la institución es contribuir a la reinserción social de las personas que fueron privadas de libertad (Ministerio de Justicia de Chile, 2020). Pese a esto es conocido que existen factores que obstaculizan los procesos de reinserción social, como las condiciones de hacinamiento, el acceso a educación, las oportunidades laborales, el apoyo de las comunidades donde los internos se dirigen posterior a su egreso, entre algunas otras variables.

Pese a los múltiples factores que dificultan la reinserción social de infractores de ley, existen distintos programas de intervención que son aplicados a nivel nacional y local que apuntan a mejorar las probabilidades de reinserción y disminuir las de reincidencia. En el estudio “Una propuesta de modelo integral de reinserción social para infractores de ley” realizado por el Centro de Estudios de Seguridad Ciudadana (CESC) y la Fundación Paz Ciudadana se identificaron diez programas vinculados a la reinserción social de los cuales nueve se encuentran en operación, si bien estos programas poseen características bastante disímiles, se encuentran conectados por varios aspectos transversales, dentro de los cuales se encuentra (i) reducidos niveles de cobertura, (ii) limitados recursos económicos destinados, (iii) oferta programática escasamente basada en la evidencia, (iv) actividades desarrolladas con escasos niveles de calidad, (v) poca especialización por perfiles delictuales específicos, (vi) poca intervención coordinada entre el mundo intra y extra muros y por último (vii) pocos trabajos que permitan estimar rigurosamente los resultados de atribuibles a las intervenciones (Morales Peillard, Pantoja Vera, Piñol Arriagada, & Sanchez Cea, 2018).

Además de la desarticulada oferta programática respecto a reinserción social que presentan distintas instituciones como Gendarmería, el Ministerio de Justicia, el ministerio del interior o SENDA, durante la última década en nuestro país se pueden identificar dos políticas públicas que han operado a partir de planes piloto, la primera de estas, el *Programa Volver a Empezar (VAE)* que inició su operación en 2018 y que según su propio manual de procedimientos declara que “se pretende instalar un modelo de gestión municipal integral y estandarizado para abordar el proceso de transición a la comunidad de personas actualmente sujetas al control de Gendarmería, siendo fundamental para aquello capacitar en reinserción social y dotar a los equipos municipales de las competencias técnicas necesarias para su apropiada ejecución” (Fundación Tierra Esperanza, 2018), este programa se implementó en las diez comunas con mayor población en el sistema postpenitenciario hasta junio del 2019.

La segunda política pública implementada aún se encuentra en vigencia, es el *Proyecto +R* que fue implementado desde 2019 y que tiene como objetivo “Implementar una política pública que, a través de la colaboración público – privada, permita capacitar e insertar laboralmente a quienes han estado privados de libertad, disminuyendo de esta manera la reincidencia delictual”, este programa posee una componente de capacitación acompañada de otra de intermediación laboral, pero no existen publicaciones que describan el modelo de gestión u operación del programa.

Durante los últimos años se han creado distintos programas de reinserción a nivel municipal que buscan disminuir la probabilidad de reincidencia de vecinos dentro de sus propias comunas. Pese a los registros en redes sociales o breves artículos noticiosos en los propios portales comunales, existe muy poca información disponible que pueda describir el funcionamiento de este tipo de programas o su grado de efectividad.

En primera instancia esta tesis buscó conocer distintas iniciativas municipales con la finalidad de caracterizarlas y descubrir cuales son las variables claves que inciden en sus resultados. Debido a las barreras burocráticas ligadas al acceso de datos desagregados y a que estas complejidades se multiplican por la cantidad de programas de distintos

municipios que se desee estudiar, es que se optó por realizar un análisis de la implementación del Programa Volver a Empezar (previamente mencionado), con el fin de distinguir sus principales fortalezas y debilidades y como estas pueden ser utilizadas en futuras aplicaciones de políticas públicas.

Se realizaron dos análisis estadísticos, el primero fue una regresión logística que buscó relacionar a la reincidencia como variable dependiente con un set de variables dependientes obtenidas de la implementación del programa VAE con el fin de estimar los efectos provocados en términos de probabilidad de ocurrencia de una reincidencia por cada una de las variables. El segundo análisis consistió en estudio de las importancias relativas de cada una de las variables dependientes, este se realizó a través de metodologías basadas en árboles de decisión, estas metodologías fueron Random Forest y XGBoost las que al ser normalizadas se pudieron combinar para obtener un orden de importancias relativas para las variables.

2. Pregunta de investigación, objetivos y resultados esperados

2.1. Pregunta de investigación

¿Qué características de los usuarios y de la implementación del Programa Volver a Empezar fueron claves para prevenir la reincidencia?

2.2. Objetivo general

Estimar la importancia, respecto a la reincidencia delictiva, de variables demográficas y de implementación, propias de los usuarios del Programa Volver a Empezar.

2.3. Objetivos específicos

- Conocer implementaciones de programas de reinserción social o prevención de la reincidencia delictiva a nivel internacional junto con sus metodologías de análisis y evaluación.
- Comprender el contexto político en el que se gestó y el mecanismo de funcionamiento impartido por el Programa Volver a Empezar durante su implementación.
- Establecer cuáles son las principales condicionantes sociales a tener en cuenta para el diseño e implementación de políticas públicas respecto a reinserción social

2.4. Resultados esperados

- Evidenciar el nivel de eficacia de la implementación del Programa Volver a Empezar.
- Reconocer a las variables edad, nivel educacional y profesión u oficio, como variables facilitadoras de los procesos de reinserción.
- Distinguir, mediante Random Forest y XGBoost, mínimas diferencias entre la importancia que posee la comuna de procedencia del usuario del programa.
- Reconocer la relevancia de las variables de inserción y capacitación laboral sobre el resultado final.

3. Marco Conceptual

3.1. Concepto de Reinserción Social

Ciertamente la Reinserción Social como concepto, tiene un origen asociado a otros conceptos usados en las ciencias sociales, tales como la reeducación, resocialización, rehabilitación y la readaptación. Muchos de los conceptos a su vez, apunta a la reinserción como un proceso, de hecho, la RAE define a la reinserción como la acción de reinsertar, la cual a su vez la define como *“Hacer que se adapte a la vida social una persona que ha vivido separada de ella, proporcionándole los medios necesarios para que no le resulte difícil”*.

Si bien, en Chile no existe una definición jurídica para la reinserción, es un concepto habitualmente utilizado en la legislación que, por ejemplo, se encuentra presente en los objetivos de Gendarmería. A pesar de esto no se encuentra tratada de forma expresa en nuestra constitución a diferencia de países como México, Italia y España, donde en la última se precisa que *“Toda pena privativa de libertad o medida de seguridad dispuesta hacia la población penal, debe estar siempre orientada hacia la reeducación y reinserción social, prohibiéndose cualquier clase de trabajos forzados”* (Jarufer Bader, 2019).

Así mismo, es posible encontrar convenciones internacionales que si bien, no entregan una definición precisa, hacen un recurrente uso del concepto. A nivel internacional, las *“Reglas de las Naciones Unidas para el Tratamiento de las Reclusas y Medidas no Privativas de Libertad para las Mujeres Delincuentes”* o más conocidas como *“Reglas Bangkok”* estipulan un set de medidas que buscan el resguardo y la promoción de los derechos humanos junto con impulsar la reinserción social de mujeres privadas de libertad, decretando, por ejemplo, que *“Tras su puesta en libertad, se prestará apoyo suplementario a las mujeres que requieran ayuda psicológica, médica, jurídica y práctica, en cooperación con los servicios comunitarios, a fin de asegurar que su reinserción social tenga éxito”* (UNODC, Oficina de las Naciones Unidas Contra las Drogas y el Delito, 2011). Otro antecedente internacional es el de las *“Reglas Mínimas de las Naciones Unidas para el Tratamiento de los Reclusos”*, más conocidas como *“Reglas Mandela”*, donde se consigna que *“Los objetivos de las penas y medidas privativas de libertad son principalmente proteger a la sociedad contra el delito y reducir la reincidencia. Esos objetivos solo pueden alcanzarse si se aprovecha el período de privación de libertad para lograr, en lo posible, la reinserción de los exreclusos en la sociedad tras su puesta en libertad, de modo que puedan vivir conforme a la ley y mantenerse con el producto de su trabajo”* (UNODC, Oficina de las Naciones Unidas Contra las Drogas y el Delito, 2016), presentando a la reinserción como uno de los objetivos de las penas privativas y como una forma de disminuir.

Por otro lado, para comprender completamente este concepto, es necesario conocer la interpretación que le entregan quienes lo experimentan. En una tesis en donde se estudió la perspectiva de los asistentes al Centro de Apoyo para la Reintegración Social de Santiago, se logró asociar el concepto de reinserción social a once códigos lingüísticos, entre los cuales se encuentran “ser normal”, “ser persona”, “trabajar”, “cambio de vida”, “adaptación”, entre otros. Pese a que cada uno de estos códigos responde a enfoques

diferentes de la percepción de lo que debería ser el proceso postpenitenciario, es posible notar que en gran parte estos códigos se encuentran vinculados al desistimiento delictual mediante el trabajo remunerado (Vargas Subiabre, 2014).

Dado todo lo anterior es posible entender a la reinserción como un derecho y un proceso de cambio desde la perspectiva de la persona privada de libertad, pero a su vez como el principal objetivo de las penas privativas de libertad, asumiendo un compromiso por parte del estado con el individuo, tanto durante la sanción, como una vez puesto en libertad. Entendiendo la importancia del rol del estado respecto a la reinserción social como un asunto de interés público, es que resulta trascendental comprender su percepción del concepto. Para esto se presenta a continuación la definición entregada por el Ministerio de Justicia y Derechos Humanos y que se encuentra publicada en su página web:

“La reinserción social es entendida como un proceso sistemático de acciones orientado a favorecer la integración a la sociedad de una persona que ha sido condenada por infringir la ley penal. Estas acciones buscan abordar la mayor cantidad de factores que han contribuido al involucramiento de una persona en la actividad delictiva, con el objetivo de disminuir sus probabilidades de reincidencia y promover el cambio hacia conductas prosociales.

La reinserción social como proceso se inicia durante el periodo de cumplimiento de una condena, y continúa cuando la persona retorna a su vida en la comunidad. Se caracteriza por el desarrollo de competencias en el ámbito individual, social y laboral; y por el fortalecimiento de los aspectos protectores que facilitan la integración a la sociedad” (Ministerio de Justicia y Derechos Humanos, 2021).

De esta definición se pueden extraer dos principales apreciaciones, la primera es que la reinserción es una problemática multifactorial que engloba tres ámbitos, individual, social y laboral. La segunda apreciación tiene que ver con el objetivo del proceso que empieza durante la condena y que continúa en el retorno a la libertad, este objetivo es la disminución de las probabilidades de reincidencia y la modificación de la conducta antisocial.

Para obtener una perspectiva completa del concepto es necesario complementar las definiciones anteriores con otras dos definiciones halladas en la literatura nacional. La primera se encuentra en un trabajo de la Fundación Paz Ciudadana denominado “Reinserción social y laboral de infractores de ley, estudio comparado de la evidencia”, en donde se plantea que:

“Resulta pertinente la definición propuesta por Gallizo (2007), quien plantea que la reinserción social es exitosa en la medida que se cumplen dos grandes criterios. Primero, restaurar la participación de una persona en la sociedad; y segundo, que ésta logre tener autonomía para decidir sobre su vida. Para alcanzar ambos objetivos son importantes las competencias que posea para desenvolverse adecuadamente y los recursos que le permitan mantener un nivel de vida apropiado en salud o educación, entre otros” (Morales Peillard, Welsch Chauán, Cárcamo Cáceres, Aguilar Chamorro, & Sosa , 2015).

La segunda definición es la presentada por Carolina Villagra en el documento “Hacia una política postpenitenciaria en Chile” publicado por el Centro de Estudios de Seguridad Ciudadana de la Universidad de Chile. En esta definición la autora propone que la reinserción es:

“Un proceso sistemático de acciones que se inician desde el ingreso de una persona a la cárcel y continúan con posterioridad a su retorno a la vida libre. Este proceso busca incidir en la mayor cantidad de factores individuales y sociales que puedan haber colaborado con el involucramiento de una persona en actividades delictivas. Abarca la totalidad de actividades en que participan voluntariamente, los reclusos, sus familias y organizaciones públicas, privadas y voluntarias, tanto a nivel central como local. La reinserción cumple con los objetivos de favorecer la integración del exrecluso a la sociedad y de mejorar la seguridad pública” (Villagra Pincheira, Hacia una Política Postpenitenciaria en Chile, 2008).

Las últimas definiciones presentadas nos permiten interpretar que la totalidad de ámbitos involucrados en la reinserción no son fácilmente delimitables, agregando en este caso el desarrollo familiar del individuo, su educación y su salud, así mismo se podrían sumar aspectos como el acceso a la vivienda o a la participación ciudadana. Por otra parte, ambas definiciones nos proponen la bidireccionalidad de los objetivos entre el individuo y la sociedad. Por un lado, es el exrecluso quien logra reintegrarse a la sociedad de manera autónoma. Por el otro, es la sociedad quien se beneficia al mejorar la seguridad pública.

Comprendiendo que el objetivo de esta tesis es conocer las variables claves para el éxito o fracaso de un programa de reinserción con las características del VAE, es que se opta por definir el concepto de reinserción de manera acotada como:

Un proceso que comienza durante el cumplimiento de una sanción, pero que se extiende más allá del término de una condena y que consiste en la modificación de múltiples comportamientos y factores que inciden en la conducta delictiva, con el fin de mejorar la seguridad pública disminuyendo la probabilidad de reincidencia de quienes han cumplido condena.

En los siguientes capítulos se profundiza en el concepto de reincidencia y en como su probabilidad puede ser medida. Se revisan casos de implementación de modelos de análisis en diferentes países para posteriormente plantear dos posibles metodologías a utilizar en esta tesis. Además, se realizará una caracterización del Programa Volver a Empezar, que permita comprender la aplicación del modelo de análisis escogido.

3.2. Concepto de Reincidencia

Así como la reinserción, el concepto de reincidencia tampoco se encuentra definido con precisión como lo señala la Fundación Paz Ciudadana en conjunto con la Universidad Adolfo Ibáñez en su estudio “La Reincidencia en el Sistema Penitenciario Chileno”, en donde explican que “El concepto de reincidencia no es unívoco. Es utilizado haciendo alusión a distintos fenómenos por las instituciones del sistema penal y por la legislación” (Morales Peillard, Muñoz Correa, Welsch Chauán, & Fábrega Lacoa, 2012).

En cuanto a su raíz etimológica, es posible distinguir que se encuentra vinculada a la palabra “*recidere*” en latín que significa caer de nuevo o recaer. En términos generales, la literatura internacional ha entendido a la reincidencia como “*la repetición de una acción delictual*” (Payne, 2007).

Desde la perspectiva jurídica de la reincidencia en el marco de la política criminal, suele entenderse como un agravante de responsabilidad, el cual se encuentra en la legislación penal general o en disposiciones especiales. De la misma manera en la legislación chilena existe el decreto de ley 64 y el 409 regulan el proceso de eliminación de antecedentes penales como una medida de no obstaculizar la reintegración a la sociedad por partes de quienes cumplieron condenas penales, estableciendo que el penado no quedará “*marcado para toda su vida con el estigma de haber sido presidiario*” y que a éste “*debe dársele la seguridad de que, una vez cumplida su condena y después de haber llenado ciertos requisitos, pasará a formar parte de la sociedad en las mismas condiciones que los demás miembros de ella y de que no quedará el menor recuerdo de su paso por la prisión*”. Desde la literatura también se desprende que esta lógica jurídica de agravante está vinculada a la justificación y finalidad de la pena de retribución o prevención. Lo anterior cobra relevancia atendiendo a que el cumplimiento del proceso de eliminación de antecedentes puede ser una variable relevante a la hora de explicar el desistimiento delictual.

Al abordar el ámbito de la criminología, es posible distinguir distintas variables de la reincidencia, que dependen de como el sistema judicial y penal se vinculan al infractor, generando o no un registro formal de las acciones de este. Es así como lo explica el Centro de Estudios de la Fundación Paz Ciudadana:

“(…) se ha hablado de una “*triple dimensión de la reincidencia*” (Mir, 1974: 7-13), distinguiendo entre *reincidencia legal* (“*comisión de una infracción penal por parte de quien, con anterioridad a la misma, ha sido condenado por otra infracción penal*”), *penitenciaria* (“*quien se encuentra en prisión tras haber sido ya condenado anteriormente por la comisión de un delito*”) y *criminológica* (“*quien, habiendo sido sujeto de una condena precedente, comete a continuación un delito, sea o no descubierto*”)” (Villagra Pincheira, Espinoza Mavila, & Martínez Mercado, La Medición de la Reincidencia y sus Implicancias en la Política Penal, 2014).

Además de la distinción anterior, es útil señalar que desde la doctrina del derecho penal en Chile existe consenso al diferenciar la reincidencia de la reiteración, entendiendo que el principal factor diferenciador entre ambos conceptos es la existencia del cumplimiento de una condena que media entre la comisión de dos actos delictuales. En este caso el concepto de “*repetición de una acción delictual*” está más asociado a la reiteración que a la reincidencia.

Entendiendo a la reincidencia también como un indicador que permite comprender los efectos o los resultados del sistema penal respecto a la reinserción social, cobra vital relevancia realizar una exacta definición de esta previo a medirla, dado a que como se ha revisado en este trabajo, diferentes interpretaciones pueden conducir a distintos resultados, subestimando el efecto que se desee estudiar. Por otra parte, desde una

perspectiva criminológica se entiende la existencia de un sesgo en la medición de la reincidencia penal o legal, dado que existe una dependencia directa de la capacidad de persecución y sanción del sistema judicial y las policías. Por el contrario, la reincidencia criminológica debiese ser el fenómeno a observar si se quisiera explicar la probabilidad real de reincidencia condicionada a factores individuales y sociales, el problema metodológico se encuentra en la capacidad de generar esta medida de manera precisa, produciéndose lo que algunos llaman como el “*número negro*”.

En la literatura, también se hace mención del momento en el que ocurre el contacto con las instancias de control penal, reconociendo dos distinciones. Por un lado, se encuentra la “*Reincidencia Policial*”, que se produce cuando el individuo enfrenta una nueva detención policial y por otra parte se distingue la “*Reincidencia Penal*”, que hace alusión a un nuevo procesamiento. Esta última definición, agrupa tanto a un nuevo procesamiento, requerimiento o acusación, no siendo exacta en la interpretación del momento y el fenómeno exacto. En relación con lo anterior, se distinguen otros dos tipos de reincidencia, la “*Judicial*”, vinculada a la resolución de una nueva pena en virtud de una sentencia condenatoria, sin hacer distinción del tipo de condena o de si esta pueda ser cumplida en el medio libre o en privación de libertad. Finalmente se distingue también la “*Reincidencia Penitenciaria*” o también llamada “*re-encarcelamiento*”, que supone un nuevo ingreso a la cárcel y que puede aplicar para medidas cautelares o el cumplimiento de una pena privativa de libertad.

Finalmente, es necesario incorporar otras dos distinciones vinculadas al tipo de delito efectuado en el acto reincidente. Bajo la anterior premisa la literatura nos indica que la reincidencia puede ser “*Genérica*” o “*Específica*”, esto dependiendo si el nuevo delito cometido es de la misma especie o no que el delito anterior.

Para el objetivo de esta tesis y dado los datos que se poseen, la reincidencia será estudiada desde una perspectiva genérica y desde la perspectiva penitenciaria o del re-encarcelamiento. Es importante destacar que este abordaje puede ser causal de errores, dado que por un lado la reincidencia penitencia podría generar una sobreestimación de los efectos preventivos del VAE, mientras que no considerar el delito específico podría inducir una subestimación de los efectos del VAE, por ejemplo, en casos en que la carrera delictiva se haya redirigido a delitos de menos violencia o menor connotación.

3.3. ¿Cómo medir la reincidencia?

El estudio del CESC (2014) se distingue que a la hora de estimar la probabilidad de reincidencia existen distintos factores que pueden ser agrupados en dos tipos, los factores estáticos que apelan a características demográficas, como la edad, la residencia y el historial delictivo, entre otros. Mientras que por otra parte se encuentran los factores dinámicos, asociados principalmente a la conducta de los infractores como la motivación, la autonomía, la proactividad y sus redes de apoyo, entre otros. Ante la existencia de estos diferentes tipos de factores, los estudios predictivos adquieren distintos enfoques y a su vez pueden ser agrupados en: actuariales, clínicas y mixtas.

Respecto a las herramientas actuariales, este es el término que se usa en la literatura para referirse a modelos probabilísticos que usan la evidencia de grupos de personas con características estáticas definidas, mediante los cuales se generan parámetros estimadores que permiten predecir la probabilidad de reincidencia de otros a partir de sus propias características. Algunas metodologías usadas son la regresión de logística de Cox y el análisis de Kaplan-Meier, ambos con un enfoque de supervivencia proveniente de las ciencias médicas y validados en el marco de las ciencias sociales.

Por otra parte, las herramientas clínicas se caracterizan por considerar tanto variables estáticas como dinámicas, teniendo enfoque en el individuo particular más en aspectos generalizados y por poseer un grado mayor de subjetividad del evaluador. Es así como por un lado estas herramientas logran capturar variables que las herramientas actuariales tienden a excluir, pero por otra parte disminuyen la precisión de la medida al incorporar la perspectiva de un tercero.

Además de lo anterior, la literatura señala posibilidad de combinar ambos tipos de herramientas para generar una medida más confiable.

La Fundación Paz Ciudadana (2012) afirma que, a la hora de estudiar la reincidencia, existen tres factores principales que son necesarios tener en consideración, estos factores son la población objeto del estudio, los eventos a analizar y el periodo temporal que se pretende observar.

El grupo de individuos a observar se encuentra completamente ligado a los objetivos de la investigación, existen tanto investigaciones que se refieren al escenario general de un país que buscan, por ejemplo, medir el impacto de políticas generales, como también existen investigaciones específicas o focalizadas que pretendan observar los efectos de una intervención sobre una parte específica de la población. Así es como para lograr una precisa identificación de la población objeto del estudio es necesario establecer de manera clara los criterios de corte, que pueden ser demográficos, por tipos de delitos, por participación de programas específicos, entre otros, en donde los últimos pueden poseer sus propios criterios de exclusión. Un ejemplo de esto es mencionado en el estudio de la Fundación Paz Ciudadana en donde se explica que *“en la evaluación de los Tribunales de Tratamiento de Drogas, en que la incorporación al programa supone la aplicación de filtros asociados a la existencia de consumo problemático de drogas, que el infractor no enfrente condena por un delito violento, que se trate de personas mayores de 18 años, que el delito haya sido cometido en la jurisdicción de la Corte, que los imputados reconozcan los hechos y que enfrente penas privativas de libertad”* (Morales Peillard, Muñoz Correa, Welsch Chauán, & Fábrega Lacoa, 2012).

En cuanto a los eventos a estudiar, autores como Maltz sostienen que el arresto es mucho mejor indicador criminológico que el dictamen de una nueva condena, pero en la literatura en general existen ciertos reparos a esta idea. Mientras que por un lado se establece la existencia de un sesgo en la entrada al sistema penitenciario dado que existen diversos factores que pueden incidir en el veredicto (como, por ejemplo, la falta de pruebas) que van más allá de la culpabilidad o no del sujeto, generando finalmente una subestimación de la reincidencia a nivel global. Por otra parte, el arresto en sí, considerado como un

indicador de reincidencia es insuficiente, dado que al interpretar directamente al sujeto como autor material de un delito sin tener en consideración el debido proceso, pasa a llevar el principio de presunción de inocencia, produciendo un sesgo inverso al del caso anterior, generando una sobrestimación del indicador. También se señala el posible sesgo ocasionado por la “*pseudo-reincidencia*”, la cual se ocasiona cuando una persona que cumple sanción en libertad comete un nuevo delito sin culminar el proceso de sanción o que cursando una pena privativa es acusada de otro delito previo, este fenómeno contribuye a sobreestimar las tasas de reincidencia, pero es posible de corregirlo mediante el correcto tratamiento de los datos.

Respecto al periodo de estudio, es necesario delimitar los eventos de inicio y termino del fenómeno medido. En la mayoría de los casos el evento inicial es el egreso del sujeto desde el centro penitenciario, pero cuando se trata de la evaluación de un determinado programa suele utilizarse el egreso de este mismo como evento inicial. Por otra parte, el evento final esta vinculado al acto reincidente y como se ha decidido medir este, cabe decir que no es lo mismo considerar el arresto o el encarcelamiento, sobre todo considerando el desfase de tiempo que existe entre estos y que en algunos sistemas puede llegar a ser de años. En cuanto a la relevancia que cobra el periodo a evaluar es destacable el periodo de un año, dado que según distintos estudios de Australia, Inglaterra y Nueva Zelanda indican que más de la mitad de los reincidentes son detenidos durante el primer año de libertad.

3.4. Experiencia internacional

En cuanto al escenario internacional, la Fundación Paz Ciudadana presenta un análisis comparativo de doce estudios diferentes de reincidencia en distintas ciudades y países, entre los que se encuentran Australia, Estados Unidos, Inglaterra y Gales, Escocia, Suiza, Nueva Zelanda, Cataluña, Irlanda, Finlandia, Canadá e Islandia. Los estudios fueron publicados entre el año 1993 y 2009, teniendo principalmente como objetivo la retroalimentación de políticas públicas existentes y la obtención de evidencia para futuras políticas. Entre las características y resultados más relevantes de esta comparación se encuentran los eventos de inicio y termino, que en la gran mayoría de los casos fueron, el egreso de la cárcel por un lado, y el dictamen de una nueva condena privativa de libertad por otro; otra característica apreciable es el tiempo de estudio, los cuales fueron de un mínimo de doce meses y un máximo de sesenta meses; en cuanto a sus resultados, en los periodos de medición más cortos los países obtuvieron resultados entre el veintiséis por ciento y el treinta y nueve por ciento, mientras que en los periodos más largos de medición se obtuvieron resultados de reincidencia superiores al cincuenta por ciento.

La literatura destaca la relevancia de las capacidades institucionales a la hora de enfrentar un proceso de producción de un sistema de medición que permita recopilar la información necesaria para realizar un completo análisis de los factores que inciden en este fenómeno. Pese a la escasa existencia de literatura que aborde la evaluación de programas específicos, se pueden observar los casos de países que han desarrollado un programa de medición global que les permita identificar distintas intervenciones por las

cuales los usuarios han transitado y poder generar un diagnóstico a nivel general de estas.

En el estudio del CESC (2014) se describe el proceso de producción de un sistema de medición para tres países (Inglaterra, Canadá y Estados Unidos), pese a que estos son procesos institucionales a nivel de estado, en esta tesis se analizarán de manera breve los casos con el fin de comprender las principales determinantes asociadas a la obtención y la gestión de los datos.

3.4.1. Inglaterra

En cuanto a la estructura organizacional de este país, el Ministerio de Justicia es quien se encuentra a cargo del Servicio de Prisiones y del Servicio Nacional de Manejo de Infractores, ambas instituciones son claves en la administración de las cárceles y del tratamiento penitenciario y uno de sus principales objetivos son la reducción del riesgo de reincidencia.

El año 2006 se marca un hito en materia de medición de reincidencia, dado que previo a este año solo existía una medición de si el infractor reincidía o no en un periodo de dos años posterior a su egreso del sistema penal. En la actualidad, los datos gestionados por el sistema inglés les permite conocer además la frecuencia individual de reincidencias, el tipo de delitos, la versatilidad delictiva o las variaciones en la gravedad de los delitos.

A partir de un acuerdo llamado “Hacer Comunidades Más Seguras”, el Ministerio de Justicia se propuso, en un plazo de 6 años, reducir en un diez por ciento las cifras de reincidencia. Para lograr sus objetivos fue necesario unificar bases de datos institucionales que le permitieran ampliar la gama de información de reincidencia, permitiéndoles identificar a infractores serios y prolíficos para mejorar la focalización de sus programas de tratamiento y reinserción.

En cuanto a los aspectos metodológicos, es relevante notar que el caso inglés selecciona una cohorte con todos los infractores sancionados entre enero y marzo de cada año, esta muestra se compone de cuarenta mil a sesenta mil personas. De esta manera también se evitan casos de doble conteo, lo que podría suceder tomando distintas cohortes al año o ampliando el periodo. Además, se asumen diferencias estacionarias, las cuales podrían impactar los resultados de otros trimestres.

La finalidad de esta medición no es generar una representación de la reincidencia anual, si no que generar un punto de comparación a lo largo del tiempo.

Para facilitar las comparaciones sin depender del tamaño final de la cohorte, se normalizan los datos de frecuencia por cada cien infractores, lo que implica en términos prácticos dividir la cantidad de delitos totales cometidos por el número total de infractores y multiplicar esto por cien. Esta normalización se realiza tanto para el conteo general, como para el conteo frecuencia vinculada a delitos graves.

3.4.2. Canadá

El caso canadiense presenta una división en la administración del sistema penitenciario. Por un lado, las penas privativas de libertad menores a dos años son administradas por los Estados Provinciales o Territorios, mientras que las penas mayores son administradas por el Estado Federal, a través del Servicio Correccional de Canadá, que se encuentra bajo el cargo del Ministerio de Seguridad Pública. Sin embargo, el organismo encargado de las definiciones metodológicas sobre la temática y, por ende, de la construcción de las cifras de reincidencia en el país es el Centro Canadiense para las Estadísticas Judiciales.

Como hito en su proceso de medición se destaca que durante 2008 la información aportada por el Centro Canadiense para las Estadísticas Judiciales pasó a ser un módulo de la Agencia Nacional de Estadísticas de Canadá, lo que abrió el acceso a los datos y generar mediciones más precisas y completas.

La estructura con la que Canadá logra recaudar datos penitenciarios les permite generar distintos análisis de valor tanto en ámbito académico como en políticas públicas, así lo indica el CESC que apunta que *“los datos indicarían que un alto porcentaje de los sujetos reincidentes son multi-reincidentes (con más de 6 o 8 condenas previas), que los hombres tienen mayor probabilidad de reincidir que las mujeres, que mientras menor es la edad de inicio en el delito mayor es la probabilidad de reincidencia, que a mayor tiempo de encarcelamiento mayor es la probabilidad de reincidir, y que un número importante de los delitos violentos serían cometidos por sujetos multi-reincidentes”*.

3.4.3. Estados Unidos

En cuanto a su organización, Estados Unidos posee un sistema bastante complejo, no solo en cuanto a sus distribuciones territoriales que son similares a las canadienses, si no que posee sutiles matices administrativos en los distintos niveles de organización de cada estado.

La temática penitenciaria es administrada a partir del Departamento de Justicia, entidad con rango de Ministerio y que es dirigida por el Fiscal General de la Nación y de la cual depende el Buró Federal de Prisiones, quienes son responsables de la custodia y los cuidados de doscientos diez mil personas condenadas a nivel federal. En particular, quien genera, organiza y administra la información proveniente de los sistemas penitenciarios es el Buró de Estadísticas Judiciales y los datos que estudian son utilizados para la toma de decisiones en la generación de políticas públicas.

Una complejidad extra enfrentada por el sistema estadounidense es la diseminación de la información, ya que la mayor parte de la población se encuentra bajo supervisión estatal en cárceles administrada por los gobiernos locales, lo cual genera que la información muchas veces no se encuentre estandarizada provocando sesgos en las mediciones. Sin embargo, existe la Asociación de Correccionales Estatales, que funciona como una instancia de intercambio y organización para las distintas prisiones estatales y que ha generado conocimiento y aprendizajes conjuntos que ha favorecido al Buró para generar mediciones de reincidencia a nivel nacional.

Los estudios publicados por el Departamento Nacional de Justicia Criminal entregan distintos resultados que son relevantes de considerar en la ejecución de este trabajo y que son presentados a continuación.

En una observación de tres años posteriores a su liberación se concluyó que un 68% de quienes cumplieron condena por delitos en contra de la propiedad fueron nuevamente condenados, en comparación a un 56,6% de quienes fueron condenados por delitos violentos y un 54,6% de quienes fueron apresados por infracciones al orden público. Además, quienes no tenían condenas anteriores y cumplieron condena en cárceles federares tuvieron un 44% de reincidencia, mientras que, para el mismo grupo en cárceles estatales, su reincidencia fue del 22%.

Finalmente, la información publicada presenta otros dos puntos importantes. En primer lugar, se encuentra una invariabilidad en el promedio nacional durante la primera década del siglo XXI, en donde paso de un 45% a un 43%. El otro punto relevante son las distintas medidas obtenidas en los diferentes estados que oscilan entre el 24% y el 65% y que sus diferencias se deben tanto a fenómenos contextuales y culturales de los propios estados, pero también al uso de distintas metodologías de medición.

Si bien, las experiencias revisadas en este capítulo apuntan a sistemas nacionales de medición de reincidencia, son importantes para relevar factores claves en la implementación de las políticas públicas y en el uso práctico del concepto de reincidencia.

3.5. Factores de riesgo asociados a la medición de reincidencia

En el análisis hecho por el CESC en 2014 se precisan seis factores de riesgo identificados por la Oficina Nacional de Auditoría del Ministerio de Justicia Británico, lo cuales son explicados a continuación:

- 1) **La complejidad del proceso de recolección de datos:** Los datos necesarios en el análisis son variados, referentes a las carreras delictuales, a datos demográficos y a datos socioeconómicos y de conducta, es por esto por lo que las fuentes de las que provienen los datos son múltiples y de diversas instituciones, por ejemplo, en el caso chileno los datos podrían provenir del Ministerio de Justicia, del Ministerio de Desarrollo Social, del Poder Judicial, de Gendarmería, etc. Además, el proceso de recolección a nivel nacional implicaría un gran volumen de datos, sumado todo lo anterior esto siempre puede decantar en pérdidas de información y errores en el procesamiento, tanto como problemas en la integración de las bases de datos.
- 2) **La complejidad en el procesamiento y el análisis de datos:** En este punto existen por lo menos dos factores relevantes. El primero tiene que ver con el *“matching”* que se debe realizar entre los datos que provienen de diferentes fuentes, dado que es probable que existan casos que no tengan registro en alguna base de datos, mientras que en las otras sí. El otro factor relevante que se establece son las *“reglas de contéo”*, las que responden a cuáles deben ser los ingresos considerados para efecto del análisis, estas reglas son múltiples y responden

tanto a los casos para los que se posee la información mínima necesaria, como a casos que posean ciertas discrepancias.

- 3) **Nivel de subjetividad:** Tiene relación con la discretización de ciertas variables y el grado de libertad que posee el operador para determinar las categorías a las que pertenece cada caso. Se indica que es relevante asegurar que el sistema requiera mínimos grados de subjetividad y que los que existen no alteren los resultados.
- 4) **Madurez del sistema de datos:** Se señala que un sistema de datos maduro es aquel que ha sido usado varias veces y que los datos que entrega son confiables, sin embargo, también se indica que es necesario garantizar la constante supervisión de la calidad de los métodos y procedimientos. De todas maneras, esto no pretende que el sistema de recolección de datos sea centralizado, pero si que en su conjunto responda a un cierto estándar.
- 5) **Experticia de los operadores del sistema:** El monitoreo a las fuentes de información y el continuo trabajo de mejoras es relevante en la fidelidad de la medición, así como las continuas pruebas al modelo estadístico que aseguren la correcta medición del fenómeno.
- 6) **Uso de datos para el manejo de recompensas:** Finalmente se indica que el manejo de datos no debe usarse para la entrega de recompensas y/o sanciones para los equipos que administran los datos o a quienes ejecutan los programas, dado que estas recompensas pueden distorsionar el uso de datos, así como el desempeño de las personas o equipos.

Estos factores son de suma importancia de tener en consideración a la hora de proponer un modelo de evaluación permanente, que les permita a los distintos programas comprender cuales son las variables o factores más relevantes para mejorar o potenciar dentro de sus gestiones.

3.6. Modelos considerados

En el marco de la presente tesis se estudiaron diversas implementaciones de modelos que apuntaran a la medición de la probabilidad de reincidencia, pero se escogió como referencia el estudio presentado por la Fundación Paz Ciudadana en 2010, enfocado en población de jóvenes infractores, condenados por la ley de Responsabilidad Penal Adolescente (RPA), el cual se denomina "*Construcción de indicadores de reinserción social de adolescentes infractores de la ley penal*", este estudio presenta dos metodologías distintas que serán descritas en esta sección y que se diferencian por su nivel de subjetividad. La elección se hizo en base a la cercanía del caso de estudio y la accesibilidad a sus autores para poder tener una comprensión más acabada del enfoque de estudio.

3.6.1. Análisis prospectivo

En este análisis se estudia el fenómeno desde lo que la literatura denomina como un "*modelo actuarial*" que considera primordialmente datos de tipo estáticos relacionados a características demográficas y al historial delictivo de los sujetos.

En base a lo anterior, el estudio define al análisis prospectivo de la siguiente manera:

“El análisis prospectivo refiere a la estimación de tasas de reincidencia para la población que ha estado cumpliendo condena bajo la ley de responsabilidad penal adolescente. Se analiza a la población que ingreso a cumplir condena durante el primer año de la entrada en vigencia de la ley de responsabilidad penal adolescente. Se utilizan datos del Sename y del Ministerio Publico. Se analizan las tasas para un seguimiento de 1 año, aunque se presenta la probabilidad condicional de reincidir para un periodo de hasta 947 días”.

Además, se declara lo siguiente como objetivo general del análisis:

“Evaluar la reincidencia en jóvenes infractores de ley en el nivel nacional según tipo de medida y sanción, por centros y programas”.

En cuanto a otras consideraciones generales respecto al análisis, en primer lugar, el análisis es realizado únicamente sobre la población condena, por lo que se tienen en consideración cinco tipos de cumplimiento de condena, que van desde el medio cerrado, hasta los diferentes tipos de cumplimiento en el sistema abierto. Dado lo anterior, es que las fuentes de datos seleccionadas para este análisis fueron SENAME, el Ministerio Público, Gendarmería, Carabineros y la Policía de Investigaciones. Pese a esto, los datos de Gendarmería, Carabineros y la Policía de Investigaciones no fueron utilizados, dado a distintos que factores que dificultaban su incorporación, pero, además, dado a que se consideran suficientes los datos provenientes de SENAME y el Ministerio Público.

El inicio del periodo de seguimiento es desde el ingreso del condenado a las medidas particulares, con excepción de los Centro de Régimen Cerrado (CRC) para los cuales se tomó como inicio el egreso, considerando solo “egresos reales” y dejando afuera a los casos en que el egreso se asocia a traslados.

En cuanto al concepto de reincidencia, se define de dos formas: la existencia de una nueva causa y la existencia de una nueva condena. En base a lo anterior se conecta la fecha de ingreso o egreso obtenida desde el SENAME con la fecha de comisión de un nuevo delito y no hasta la fecha de la nueva condena. Lo anterior es dado que se busca intervenir el acto delictual.

El periodo de mayor seguimiento posible va desde la fecha del primer ingreso registrado por SENAME, hasta la del último registro de comisión de delito del Ministerio Público, lo que completa un periodo total de 953 días. Pese a lo anterior se escogió utilizar un periodo de observación de un año para cada caso, dado que se estimo un periodo deseable de observar de manera recurrente en el tiempo, con el que se pudiera evaluar mejoras anuales en los respectivos programas.

Por último, es relevante destacar que la variable de reincidencia es dicotómica, es decir de SI o NO, pese a esto también se cuantifican las condenas posteriores a la referencia para poder distinguir casos multireincidentes.

Luego de realizar los respectivos cruces en las bases de datos se pudo determinar un total de 5.001 registros, para los cuales se realizaron dos análisis paralelos. El primero fue un análisis de supervivencia de Kaplan-Meier, que es una metodología aplicada en

las ciencias médicas para estudiar los efectos de ciertos tratamientos, pero que ha sido incorporada en los estudios criminalísticos a través de la psicología jurídica, presentando resultados positivos en sus predicciones de conductas violentas y de reincidencia. El otro análisis que se realizó fue el de predicción de reincidencia al año, el cual consistió en realizar una regresión logística de Cox que además fuese capaz de incorporar ciertas variables que explicaran la reincidencia.

Para el caso del análisis de supervivencia fueron utilizados todos los registros disponibles, mientras que, en el caso de reincidencia al año, fueron excluidos los registros con ventanas temporales entre los sucesos de referencia que fuesen inferiores a 365 días y los casos que cumpliendo condena en un CRC no presentaran un egreso real.

3.6.2. Análisis Retrospectivo

En esta metodología, el estudio plantea un análisis del tipo “clínico” por lo que se basa en un modelo internacional denominado “*Youth Level of Service/Case Management Inventory*” o YLS/CMI y que se encuentra basado en el modelo de intervención de Riesgo, Necesidad y Responsividad o Capacidad de Respuesta (RNR) de Andrews y Bonta, modelo que es comúnmente utilizado por los distintos programas de intervención en materia penitenciaria. Este análisis, como fue señalado anteriormente posee un mayor nivel de subjetividad y presenta dificultades a la hora de entregar una lectura agregada, ya que es más eficaz en la implementación caso a caso. En concreto, la metodología consiste en una entrevista realizada por un profesional, quien se carga de entregar una interpretación clínica de las respuestas del usuario respecto a una serie de preguntas divididas en los tres ítems principales (Riesgo, Necesidad y Capacidad de Respuesta), en base a la evaluación del profesional genera una ponderación de sus respuestas lo que entrega un puntaje que es clasificado en categorías que determinan su probabilidad de reincidencia.

Cabe señalar que, si bien la validación se realiza sobre un instrumento que aborda la situación específica de jóvenes infractores, este proviene de otro instrumento llamado Level of Service Inventory-Revised que posee foco en población adulta y que puede ser aplicable para efectos de esta tesis.

En la descripción del análisis, el estudio señala que:

“El presente estudio se propuso en primer lugar, la validación de un instrumento de estimación del nivel de riesgo de reincidencia (...) La presente validación del instrumento YLS/CMI requiere, en cuanto a su capacidad real para predecir reincidencia, la generalización de su uso y contrastar los resultados obtenidos con población que a futuro reingresa. Lo cual implica un amplio espacio de tiempo entre sujetos cuantificados con el instrumento y conducta posterior (...) Sin embargo, una validación rigurosa requiere de una muestra mayor, de un análisis prospectivo, de una capacitación de mayor tiempo para los encuestadores, de acceso a fuentes de datos históricas y de acceso a diversas fuentes de información. Por lo tanto, este es solo un primer paso en el camino de la validación de este instrumento”.

Para efecto de esta tesis es importante considerar el desarrollo metodológico utilizado en este análisis con el fin de comprender el uso conjunto de técnicas que permitan obtener tanto una interpretación agregada del fenómeno, como también una que logre incorporar ciertos factores dinámicos asociados a características de personalidad de los usuarios que permiten realizar interpretaciones y predicciones más precisas.

En este escenario, el análisis plantea el siguiente objetivo general:

“Validar un instrumento capaz de cuantificar el nivel riesgo de reincidencia en jóvenes infractores de ley”

Para este caso se utilizan como eventos de referencia para medir la reincidencia, el dictamen de una primera condena y algún nuevo contacto con el sistema judicial ya sea como condenado, imputado o detenido. Se entiende como reincidencia el presentar condenas que derivaron en un ingreso a SENAME anteriores al ingreso por el cual se entrevista al sujeto.

El instrumento se divide en 6 partes, las cuales se presentan a continuación:

Evaluación de riesgos y necesidades: Se divide en 42 ítems que reciben una puntuación de 0 o 1, los ítems se encuentran agrupados en las siguientes 8 escalas:

- Delitos y medidas judiciales pasadas y actuales: 5 ítems (0 a 5 puntos)
- Circunstancias de las familias y de los padres: 6 ítems (0 a 6 puntos)
- Educación formal/empleo: 7 ítems (0 a 7 puntos)
- Relación con sus grupos de pares: 4 ítems (0 a 4 puntos)
- Consumo de sustancias: 5 ítems (0 a 5 puntos)
- Tiempo libre/diversión: 3 ítems (0 a 3 puntos)
- Personalidad/comportamiento: 7 ítems (0 a 7 puntos)
- Actitudes/valores y creencias: 5 ítems (0 a 5 puntos)

Resumen de factores de riesgo/necesidad: Se agregan los resultados de las escalas, generando cuatro niveles de riesgo:

- Riesgo bajo: entre 0 y 8 puntos
- Riesgo medio: entre 9 y 22 puntos
- Riesgo alto: entre 23 y 34 puntos
- Riesgo muy alto: entre 35 y 42 puntos

Evaluación de otras necesidades/consideraciones especiales: En esta parte el evaluador tiene la posibilidad de considerar otros factores de riesgo o necesidades que pueda identificar fuera de los descritos en la pauta.

Evaluación del nivel general de riesgo / necesidad: En este paso el profesional a cargo debe reconsiderar los resultados consignados en la rúbrica, teniendo en consideración la evaluación en general más allá de cada ítem en particular.

Nivel de contacto: El evaluador debe definir el nivel de contacto y seguimiento que se aplicará al plan de intervención del usuario.

Plan de gestión de caso: En esta etapa se debe establecer el plan de intervención en consideración de la evaluación realizada, fijando metas que sean coherentes a las necesidades y los riesgos detectados.

Posteriormente el profesional también queda a cargo de realizar el seguimiento al cumplimiento del plan de intervención.

Aunque, el instrumento comprende todos los pasos mencionados, la validación planteada en el estudio se enfoca únicamente en el primer y segundo paso, dado que estos en conjunto proporcionan los cuatro niveles de riesgo y la información necesaria de las ocho escalas.

En cuanto a las validaciones del modelo a nivel internacional, el estudio indica que este ha sido validado en múltiples ocasiones en Estados Unidos y Canadá, además de un caso particular en España, en donde se han utilizado distintos métodos estadísticos como ANOVA o regresiones logísticas. En las validaciones mencionadas se ha concluido que el instrumento posee la capacidad de discriminar entre jóvenes reincidentes y no reincidentes, pero no se asegura su efectividad ante el condicionamiento a poblaciones específicas, por ejemplo, género o etnias.

El instrumento fue validado en tres planos fundamentales entre los que se encuentra la consistencia, es decir la suficiencia del instrumento para la medición y la coherencia entre las partes del instrumento. Capacidad discriminante, es decir, evaluar si el instrumento es capaz de detectar diferencias entre las poblaciones objetivos. Por último, su confiabilidad, es decir, cuanto se condice el nivel de riesgo detectado con la reincidencia real medida.

Además, el estudio propuso dos resultados, por un lado, una descripción de las poblaciones de estudio a través de los niveles de riesgo y sus características sociodemográficas. Por otra parte, la predicción de la reincidencia y la identificación de los factores de riesgo que más inciden en esta.

Resultados

Análisis prospectivo:

En este caso se entregan dos resultados, el primero consiste en lo arrojado por el análisis de Supervivencia de Kaplan-Meier, el que arroja una tasa de reincidencia para un periodo de 899 días, mientras que al considerar solo casos de reincidencia una vez cumplida la condena efectiva el porcentaje llega al 69,1% en un periodo de 883 días. En estas mediciones también se presentan los resultados condicionados a variables demográficas y variables relacionadas con las carreras delictivas, como tipo de delito, cantidad de condenas previas, tipo de sanción, entre otras.

Posteriormente se presentan los resultados obtenidos para la regresión logística de Cox, en donde se obtienen valores de coeficientes para cada variable evaluada, como región, edad, sexo, tipo de delito y tipo de medida. Posteriormente se entregan los resultados agregados de las tasas promedios por tipo de sanción y total considerando un año de

seguimiento posterior a la condena. Las tasas de reincidencia obtenidas son de 70,1% para cualquier nueva causa y de un 55,2% para una nueva condena.

Análisis retrospectivo:

Luego de la validación se realiza una regresión logística de Cox teniendo como variables explicativas los puntajes obtenidos en cada escala de evaluación con el fin de determinar la probabilidad de reincidencia. Se constata que el instrumento logra generar una correcta interpretación del nivel de riesgo de reincidencia en un 79% de las ocasiones, pese a estos resultados el modelo determina que solo dos de las variables utilizadas son significativas, las condenas previas o actuales y los rasgos de personalidad o conducta.

En cuanto a las conclusiones de este análisis, en el estudio se manifiesta que existen a lo menos tres factores determinantes y necesarios de mejorar:

- 1) Profundidad de la aplicación del instrumento: Se recomienda aumentar las duraciones de las entrevistas y complementarlas con otras fuentes de datos como informes de sus colegios, informes de consumo o entrevistas con los padres. Además, se recomienda reconceptualizar los indicadores para evitar problemas de comprensión.
- 2) Preparación de los evaluadores: Se constata la deseabilidad de que los evaluadores sean personas del área de la salud mental, pero con experiencia en criminología, que cursen un proceso de capacitación extenso y que puedan ser evaluados previamente a la implementación.
- 3) Aumento del tamaño muestral: Se atribuye la no significancia de las variables a lo acotado del espacio muestral, por lo que se insta a ampliar este para mejores resultados en futuras implementaciones.

3.7. Fundamentos para la elección del modelo a utilizar

Como ha sido explicado en este marco conceptual, la reincidencia no es un concepto que posea una definición univoca, y además se encuentra determinado por múltiples factores como los eventos de referencia, los tipos de delitos e incluso los modelos utilizados para estimarla. A la hora de escoger un modelo de análisis es necesario tener en claro las implicancias de este. Como ha sido revisado, los modelos actuariales, vinculado a mediciones econométricas nos permiten obtener una perspectiva más clara de cifras agregadas, pero a partir principalmente de datos estáticos que tienden a obviar características de personalidad y comportamiento. Por otro lado, los modelos clínicos, permiten incorporar variables dinámicas, pero agregando un mayor grado de subjetividad que dificulta a su vez obtener correctas estimaciones a nivel agregado.

En vista de lo anterior, una buena alternativa es optar por metodologías mixtas, que permitan tener estimaciones más precisas, pero incorporando a su vez variables dinámicas. Para efecto de esta tesis se ha decidido trabajar utilizando los modelos implementados en el análisis prospectivo de la Fundación Paz Ciudadana, tanto el

análisis de supervivencia de Kaplan-Meier como la regresión logística de Cox. Hubiese sido ideal combinar ambos modelos revisados, pero existen una serie de argumentos que apuntan a la aplicabilidad de ambas opciones que son detallados a continuación:

- Imposibilidad de llevar a cabo una validación de un instrumento clínico:
 - 1) Tiempo de implementación: Como se establece en el estudio de Paz Ciudadana, llevar a cabo una validación de un instrumento requiere de tiempo de capacitación, implementación y seguimiento, periodos que podrían exceder los 18 meses, lo cual excede por mucho el tiempo disponible en el marco de esta tesis.
 - 2) Dependencia de Profesionales de la salud: Para la correcta implementación del modelo es necesario contar con evaluadores que sean profesionales de la salud mental, recurso con el que no se cuenta y que podría ser complejo de conseguir.
 - 3) Restricciones por contexto de pandemia: Aunque los factores ya mencionados pudiesen ser sorteados, la implementación de esta metodología requiere un significativo nivel de presencialidad, lo cual es opuesto a las restricciones presentes hoy en día.
- Aplicabilidad de un instrumento econométrico:
 - 1) Disponibilidad de fuentes de información: Dado que para este caso las fuentes de datos debiesen ser principalmente los programas de reinserción municipales, Gendarmería y el Ministerio Público, se han realizado gestiones iniciales para evaluar la disponibilidad de los datos, obteniéndose resultados positivos.
 - 2) Capacidad de análisis: Los modelos a utilizar son conocidos y presentan una mayor facilidad para el análisis.
 - 3) Adaptabilidad con el objetivo de la tesis: Seguramente este es el punto más relevante, dado que el tipo de análisis permite trabajar con datos antiguos, es decir registros de años anteriores, lo que aumenta el volumen de datos disponibles. Dado lo anterior se vuelve innecesario realizar un levantamiento de nueva información.

Habiendo escogido los modelos de análisis es importante delimitar las fuentes de las que provienen los datos. En primer lugar, del Poder Judicial deben provenir los datos referentes a la carrera delictiva, como el historial de causas y de condenas, tipos de condenas, recintos de cumplimiento de penas y las fechas de los eventos de referencia. En segundo lugar, de Gendarmería es necesario obtener los registros de egreso del cumplimiento de la pena y posiblemente, datos referentes a los tipos de intervenciones efectuados sobre los sujetos. Por último, desde los programas municipales es necesario incorporar en principio la información que defina la muestra a estudiar, y los objetivos cumplidos por los usuarios dentro de sus intervenciones.

Si bien, en un caso ideal, las fuentes mencionadas anteriormente son las requeridas, para esta tesis debido a la complejidad de la obtención de los datos, por el nivel de privacidad de estos y por las burocracias internas de las instituciones que los resguardan, es que finalmente se trabajó con datos pertenecientes a la implementación del Programa Volver

a Empezar, más la variable de reincidencia efectiva otorgada directamente por el área de estudios de Gendarmería.

3.8. Programa Volver a Empezar

Como fue comentado previamente, este estudio originalmente pretendió analizar a diversos programas municipales que se encuentran operativos en la actualidad, pero por la dificultad de acceder a data estandarizable y analizable dentro de los plazos de este estudio se optó por analizar los resultados obtenidos por el Programa Volver a Empezar. Afortunadamente para este análisis, las directrices metodológicas bajo las que opera el mencionado programa y los diferentes programas municipales previamente considerados son prácticamente las mismas o muy similares.

3.8.1. Modelo Riesgo Necesidad y capacidad de Respuesta (RNR)

Si bien este modelo, creado por Andrews y Bonta en la década de los 80, ha sido nombrado previamente, no ha sido explicado en detalle y no se ha aclarado cual es el nivel de implicancia que tiene dentro del funcionamiento de los programas municipales. El modelo es en si una respuesta a las perspectivas punitivistas, planteando la intervención de quienes delinquen desde una perspectiva individual y atendiendo a sus necesidades. Señala además que la sanción por si sola no puede eliminar el riesgo de que una persona que ha delinquido vuelva a hacerlo.

El modelo se descompone en tres principios que dan origen a su nombre y que se describen a continuación:

- 1) Principio de riesgo: Comprende dos aspectos primordiales, en primer lugar, postula que el riesgo se puede predecir. Por otra parte, apunta a que el nivel de necesidad de servicios debe ser similar al nivel de riesgo, esto quiere decir que la parte de la población que presente mayores riesgos de reincidencia, son quienes deben ser intervenidos más intensivamente.
- 2) Principio de necesidades: Son los ámbitos que se deben intervenir para que un programa sea exitoso y se reduzca la probabilidad de reincidencia. Es necesario delimitar que estas necesidades son solamente del tipo criminógenas, entiendo que el sujeto puede tener más necesidades, pero que estas no son relevantes a la hora de referirse a reincidencia. Los autores determinaron ocho factores de riesgo centrales:
 - Historial de conducta antisocial
 - Patrón antisocial de personalidad
 - Pensamiento/actitud antisocial
 - Asociación con pares antisociales
 - Circunstancias familiares/materiales
 - Trabajo/escuela
 - Tiempo libre/recreación
 - Abuso de sustancias

- 3) Principio de responsividad/capacidad de responder: La capacidad de respuesta se refiere a la manera en que debe ser abordada la intervención, atendiendo a las características de los sujetos y preocupándose de los factores biopsicosociales.

Además, el modelo posee otros cuatro tipos de principios entre los que se encuentran los principios globales, que apuntan a cosas generales, como por ejemplo que los programas deben realizarse de manera ética, legítima, moral y humana, otro principio global hace referencia a una concepción cognitivo-social de la personalidad y los planteamientos de la conducta criminal. Otros, son los principios de evaluación, se resalta lo importante que es usar instrumentos validados mediante la suficiente evidencia, también aboga a que durante las evaluaciones y las mismas intervenciones debe existir una flexibilidad que sea parte de una mirada humanizadora y le permita al profesional tomar las decisiones adecuadas, más allá que se encuentren fuera de las pautas de evaluación. En cuanto a los principios de organización se plantean dos puntos centrales, la organización con la base comunitaria, sin redes de apoyo las implementaciones no son posibles, y la continuidad del servicio, la integración debe ser más allá de la intervención durante la condena, por lo que se deben tener capacidades de seguimiento. Finalmente se encuentran los principios de entrega, vinculado a como los profesionales entregan el servicio, aquí encontramos el principio de dosificación, el cual es clave para quienes poseen altos índices de riesgo, con el fin de evitar el abandono a los programas, por otra parte, se vincula a la preparación de quienes imparten el programa, lo que se encuentra orientado a generar un trato cálido, respetuoso y colaborativo.

Junto con los principios declarados, el Modelo RNR utilizado por el VAE, categoriza ocho factores de riesgo:

- 1- **Historia delictual:** considera todos los actos delictivos que la persona ha realizado en su historia de vida, estos actos pueden ser caracterizados de distintas maneras, ya sea por su tipo o forma. Es en estos casos muy relevante comprender tanto el nivel de violencia como contra quien se encuentra dirigido (propiedad o personas). Este es un factor estático claramente, dada la imposibilidad de cambiar el pasado de las personas.
- 2- **Educación y empleo:** también se considerado un factor estático, y los planteamientos de trabajo van orientados a mejor o reforzar su valoración, nivelando la educación o generando apresto laboral, enfocándose también en la brecha personal a la hora de acceder a empleos y no solo en la brecha social.
- 3- **Familia y pareja:** este factor si bien, por un lado, posee variables estáticas como el estado civil o el número de hijos, también se encuentra la dinámica familiar (pareja, roles parentales y sociales) que debe ser considera por sus componentes conductuales, como una variable dinámica.
- 4- **Uso del tiempo libre:** el ocio es un factor dinámico y relevante que, en los casos negativos, se conservan patrones de conducta de la vida carcelaria, mientras que como objetivo se busca utilizarlo en beneficio de la vida familiar.
- 5- **Grupo de pares o amigos:** tiene relación con redes que en los peores casos pertenecen a círculos delictuales, es responsabilidad de la comunidad generar una

alternativa para el usuario sobre cualquier grupo de pares negativos y evitar que este tenga que mudarse o buscar empleo en otra parte en donde no lo conozcan.

- 6- **Consumo de Alcohol y drogas:** tiene relación con una condición individual, que puede o no ser incontrolable para el sujeto. Este factor adquiere una relevancia clave, ya que se vincula directamente a más factores de riesgo de índole social.
- 7- **Actitudes y orientación pro-criminal:** referente a recursos personales y valores éticos-morales que el sujeto a acumulado a lo largo de su historia, aún así es un factor dinámico, ya que pertenece a su conjunto de creencias que pueden ser cambiadas.
- 8- **Patrón antisocial:** tiene que ver con variables, atributos o características individuales de personalidad y salud mental, este factor no representa un riesgo de reincidencia en sí, pero si se reconoce como riesgo para la adaptación.

3.8.2. Implementación

Si bien durante su implementación el programa VAE atendió a más de cuatrocientos usuarios, aquel no fue el principal objetivo de este proceso, el programa tenía como objetivo la replicabilidad de sí mismo, es decir, pretendían generar las redes comunales para que el municipio pudiese dar continuidad al programa una vez cerrado el piloto. Para conseguir tal meta, el programa se ejecutó en cuatro pasos:

Paso 1:

El objetivo en esta etapa fue conformar una red intersectorial de reinserción y aprender a gestionarla, en primer lugar, se pretendía vincular a distintos servicios comunales con enfoques en ámbitos familiar, social, laboral y jurídico. En términos prácticos, durante el paso uno, se conformaron las Mesas Ejecutivas Comunales (MEC) que fueron presididas por los alcaldes de cada comuna y que contaron con participación privada y de la sociedad civil, se conformaron capacitaron los equipos comunales y estos a su vez convocaron y capacitaron a los funcionarios municipales respecto al modelo de gestión de casos de reinserción.

En la siguiente tabla es posible apreciar el detalle de los ciclos de capacitación llevados a cabo a lo largo del programa y principalmente en este paso.

Objetivo	Capacitar a los Equipos Comunales y a los Funcionarios Municipales de la oferta programática en los ámbitos familiar vincular, social, laboral/capacitacional y jurídico, que sean seleccionados para formar parte de la Red Intersectorial de Reinserción.	
Resultado final	Equipos Comunales y Funcionarios Municipales capacitados en Modelo de Gestión de Casos para la Reinserción.	
INDUCCIÓN INICIAL	Resultado intermedio 1	Inducción inicial Equipos Comunales Equipos Comunales habilitados teórica y prácticamente en las temáticas sistematizadas en el Manual de Procedimientos, con énfasis en el Modelo de Gestión de Casos para la Reinserción y la aplicación de sus instrumentos.
	Resultado intermedio 2	Inducción inicial Funcionarios Municipales Funcionarios Municipales sensibilizados y capacitados en reinserción, integrando algunas de las temáticas del Manual de Procedimientos, con énfasis en el Modelo de Gestión de Casos para la Reinserción, y los procedimientos e instrumentos requeridos para la atención directa de usuarios.
	Resultado intermedio 3	Entrenamiento/atención de casos Equipos Comunales Los Equipos Comunales gestionan casos de usuarios de primera línea aplicando los conocimientos contenidos en el Manual de Procedimientos.
ENTRENAMIENTO/ ATENCIÓN DE CASOS	Resultado intermedio 4	Entrenamiento/atención de casos Funcionarios Municipales Los Funcionarios Municipales –según su cargo y funciones– gestionan casos de usuarios que solicitan atención por demanda espontánea y aquellos casos referidos por el Equipo Comunal, cuya supervisión y seguimiento debe ser realizada por este último. La gestión de casos implica que el Equipo Comunal refiere a las unidades municipales correspondientes los casos que requieren de sus servicios en los ámbitos Familiar vincular, Social, Laboral/capacitacional y Jurídico.
CAPACITACIONES PERIÓDICAS	Resultado intermedio 5	Se entrega a los Equipos Comunales y a los Funcionarios Municipales –en dos momentos secuenciales del proceso de capacitación- herramientas y conocimientos respecto de las necesidades levantadas por los Equipos Comunales y los Funcionarios Municipales, relevantes en la aplicación del Modelo de Gestión de Casos para la Reinserción.

Tabla 1 - Ciclos de Capacitación VAE (elaboración externa)

Paso 2:

En esta etapa se busca iniciar las intervenciones con usuarios y realizar las iteraciones necesarias en los mecanismos de derivación interna entre los diversos servicios

municipales. Durante este proceso se hicieron efectivos los convenios adquiridos con gendarmería y se realizaron atenciones focalizadas para usuarios de primera línea (derivación directa de gendarmería) y usuarios de segunda línea (demanda espontánea).

Con precisión, el programa define ambas categorías de usuarios de la siguiente manera:

“Primera línea: población penal y su grupo de referencia que acceden al programa a través de la **derivación por medio de los registros de Gendarmería**. En cuanto al grupo de referencia, se estima el trabajo incluyendo un promedio de 3 personas en dicho grupo. La atención a beneficiarios de primera línea será entregada de manera directa por los Equipos Comunales. La derivación permite identificar a los usuarios que en el corto plazo terminarán de cumplir su condena o que se encuentran en proceso de eliminación de antecedentes, periodo identificado por la bibliografía especializada como el más crítico para la reincidencia y que, por tanto, requiere de intervenciones oportunas en su proceso de transición a la comunidad. La atención de estos usuarios debiera estar vinculada en complementariedad con el trabajo de Gendarmería y con ello favorecer un proceso de continuidad entre el cambio personal que pudiera estar desarrollándose y la vinculación con oportunidades en la red comunal para proyectar una vida sin reincidencia del delito.”

“Segunda línea: corresponde a los usuarios, hombres y mujeres mayores de 18 años, que se encuentran sujetos a Gendarmería de Chile y que soliciten, por demanda espontánea la atención de algún servicio municipal. Las personas-usuarias de segunda línea encuentran su justificación en el mandato del municipio de acoger e identificar las demandas de reinserción que pueden aparecer en distintos programas y áreas de trabajo de los municipios, esto con el fin de planificar y gestionar la información y atención a dichos usuarios, optimizando la intervención y evitando la burocratización de la prestación de servicios. A la acción habitual que el funcionario municipal realizaría, se espera que amplíe su acción a una vinculación más amplia de la persona-usuaria de segunda línea en atención a favorecer el proceso de reinserción social. En este caso, la continuidad entre las acciones de Gendarmería y las del municipio pueden haber sido todas en momentos difíciles, pero pueden ser igualmente complementarias y aportativas al proceso de reinserción.”

A su vez establece intervenciones y etapas diferentes para cada tipo de usuario.

En el caso de los usuarios de primera línea establece seis tipos de intervenciones: Entrevista de riesgo de reincidencia, Entrevista motivacional, Acompañamiento personalizado, Gestión de redes focales, Análisis de necesidades de vinculación para el apoyo en reinserción e intervención en crisis. Mientras que las fases en las que operó el programa fueron: 1) preingreso; 2) Selección; 3) Diagnóstico; 4) Formulación de plan de reinserción; 5) Ejecución del plan de reinserción; 6) Seguimiento y evaluación del plan; 7) Egreso y satisfacción usuaria.

En el caso de los usuarios de segunda línea, las intervenciones consideradas fueron: Análisis de detección de necesidades de vinculación y apoyo en reinserción, consejería y gestión de red focal. Mientras que las fases establecidas fueron: 1) Detección del usuario por la red municipal y derivación al equipo comunal; 2) Coordinación y complementariedad; 3) Plan de reinserción; 4) Monitoreo del plan; 5) Egreso.

Las etapas descritas anteriormente se presentan en la siguiente imagen.

FLUJOGRAMA DE ATENCIONES PARA PERSONAS-USUARIAS

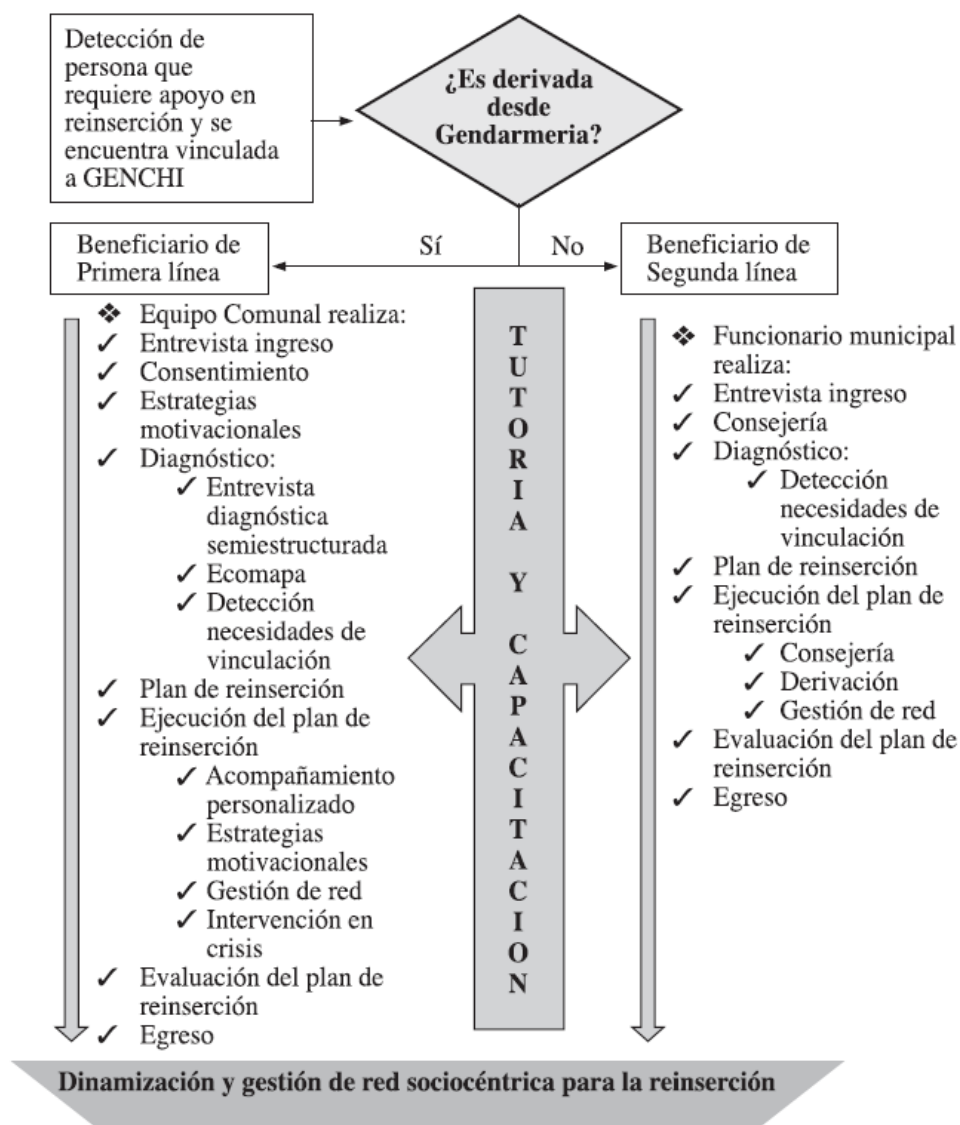


Figura 1 - Flujoograma de atenciones (Elaboración externa)

Para efectos de esta tesis es relevante puntualizar en la etapa de egreso, dado que esta medida se vincula a la aplicación completa del programa y podría ser una variable clave en la evaluación de la eficacia de este. A continuación, se presenta una breve revisión de las condiciones de egreso que establece el Manual de Procedimientos del VAE.

Se establece que el egreso de los usuarios se realizará posterior a una reunión de evaluación conjunta en donde se medirá el cumplimiento del plan de reinserción, estrategias de prevención de conductas reincidentes y la evolución del comportamiento prosocial del usuario. Una vez realizado un balance positivo de esta evaluación, el equipo comunal redactará un informe de cierre del proceso otorgando el egreso al usuario. Si

bien el egreso está vinculado en primera instancia al cumplimiento del plan de reinserción, este se encuentra también supeditado a otras causales como:

- Traslado de comuna, en caso de traslado a otra comuna donde operase el VAE se realizarían gestiones para continuar con el plan de reinserción.
- Cambio de sistema penal o nuevas sanciones.
- Deserción o pérdida de contacto prolongada.
- Renuncia voluntaria.
- Fallecimiento.

Paso 3:

Este paso se encontró marcado por actividades orientadas a fortalecer las redes comunitarias incorporando nuevos actores. En específico los objetivos fueron establecer vínculos colaborativos con organizaciones comunitarias, realizar actividades de sensibilización en conjunto con las organizaciones comunitarias y generar articulación entre los actores comunitarios y las redes institucionales.

Paso 4:

Por último, se buscó la consolidación de la red sociocéntrica de reinserción, en palabras simples, consistió en una serie de reuniones diseñadas para coordinar la ejecución y el seguimiento de los planes de reinserción de casos, reforzar los vínculos con organizaciones comunitarias y realizar reuniones de evaluación de estructura y funcionamiento de la red, revisión de procedimientos y proyección del trabajo en reinserción social y su rol en la seguridad ciudadana.

3.8.3. Análisis del nivel de cumplimiento de los objetivos del Programa Volver a Empezar

El 27 de junio de 2019, poco tiempo después de haber finalizado la implementación del programa, el Centro de Estudios Justicia y Sociedad en conjunto con el Instituto de Sociología de la Universidad Católica presentaron un informe en donde analizaron el cumplimiento de los objetivos del VAE.

En el mencionado informe se midió cumplimiento de diversos objetivos relacionados a la creación e implementación del sistema de gestión de casos, a la articulación de redes y a la atención de usuarios. Un aspecto específico dentro de la atención de usuarios fue la reincidencia de estos posterior al programa. Debido al poco tiempo transcurrido el análisis fue analizado con datos de reincidencia entre diciembre de 2018 y mayo de 2019, es decir alrededor de seis meses.

En el respectivo informe, previo a la entrega de los resultados, se describen las siguientes consideraciones:

“En esta evaluación, se solicita desde las bases medir reincidencia judicial y penitenciaria, las que representan solo el dato administrativo del comportamiento delictivo, dejando fuera otras cifras negras, tales como reiteraciones, reingresos carcelarios,

comportamientos de desistimientos del delito, entre otros. Por otro lado, hay que considerar que:

1. *El periodo de observación de reincidencia penitenciaria para esta evaluación es bastante breve en comparación a otros estudios de reincidencia.*
2. *Solo se pudo contar con datos de Gendarmería de Chile del Sistema cerrado y abierto, y no del Poder Judicial.*
3. *EL no poder atribuir que las bajas tasas de reincidencia sean por el Programa.”*

En el informe se puede encontrar la siguiente tabla que permite distinguir la población reincidente y no reincidente.

Tipo de reincidencia	N° de usuarios beneficiarios de Primera línea	% de usuarios de Primera línea (n=403)
No Reincidentes	377	94%
Reincidencia Legal / reingreso al sistema abierto GENCHI post- Programa ⁵¹	9	2%
Reincidencia Legal / reingreso al sistema cerrado GENCHI post- Programa	9	2%
Reincidencia penitenciaria	8	2%

Tabla 2 - Reincidencia según criterios (elaboración externa)

Cabe destacar que el análisis realizado fue únicamente sobre usuarios primera línea y con casos cerrados para diciembre de 2018, con data de reincidencia entre diciembre de 2018 y mayo de 2019, aproximadamente cinco meses, lo cual no refleja necesariamente cinco meses luego del egreso de cada usuario, es relevante señalar esto, dado que la cifra que reincidencia obtenida no es uniformemente comparable, ya que puede haber usuarios que hayan egresado varios meses antes. Con las consideraciones presentadas, el informe establece que las cifras obtenidas son “*alentadoras*”, ya que “*se destaca positivamente que existe una baja tasa de reincidencia de los usuarios del Programa, aunque solo se esté presentando la cifra oficial de Gendarmería*”. En concreto, los resultados arrojaron que un 94% de los usuarios, correspondiente a 377 personas eran no reincidentes a esa altura y que solo 9 personas se encontraban cumpliendo condena en el sistema cerrado, lo que corresponde a un 2%.

De forma paralela, además de los resultados reincidentes, el informe se pronuncia en cuanto a los usuarios reinsertados. Para el levantamiento de este dato utilizan tres métricas distintas, en primer lugar la cantidad de usuarios reinsertados según cumplimiento de los planes de reinsertión, en donde 175 (43%) de un total de 403 personas fueron reinsertadas. En segundo lugar, se obtuvo la cantidad de reinsertados a opinión de la dupla técnica de cada comuna, donde la cifra fue de 160 personas (40%) sobre el mismo total. Finalmente se estimó la cantidad de reinsertados según la definición del programa VAE, donde el total fue de 158 personas (39%).

Respecto a la evaluación general del programa, cabe destacar que uno de los principales elementos a la hora de evaluar el programa por parte del mencionado informe, fue la satisfacción usuaria. A partir del siguiente gráfico se puede desprender la utilidad que tuvo el programa para los usuarios encuestados, es relevante indicar, que la encuesta de satisfacción usuaria de este estudio fue aplicada a un 20% aproximado de los usuarios egresados.

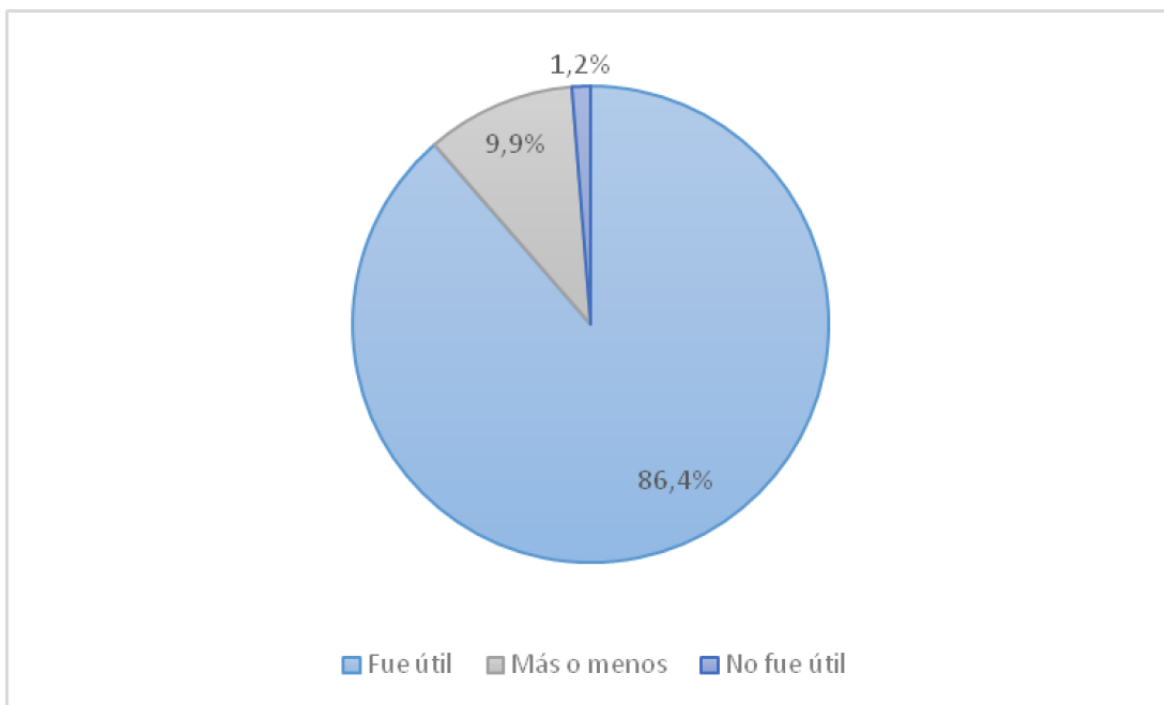


Gráfico 1 - Encuesta utilidad del programa (elaboración externa)

En cuanto a la satisfacción se puede observar que un gran porcentaje consideró que el programa les fue útil en facilitar alguna materia de sus procesos de reinserción, las temáticas más valoradas fueron lo laboral, lo educacional y la eliminación de antecedentes, otro aspecto muy valorado fue la relación forjada con los gestores y la ayuda psicosocial, en el sentido del desarrollo de la autoestima, el autocuidado y el ánimo de enfrentarse a nuevos desafíos. Finalmente, entre los aspectos menos valorados se encontró la vinculación a redes municipales y el apoyo en temáticas familiares.

Además, a los usuarios se les otorgó la posibilidad de evaluar el programa con una nota de 1 a 7, en donde el 83,5% calificó el programa con un 7.

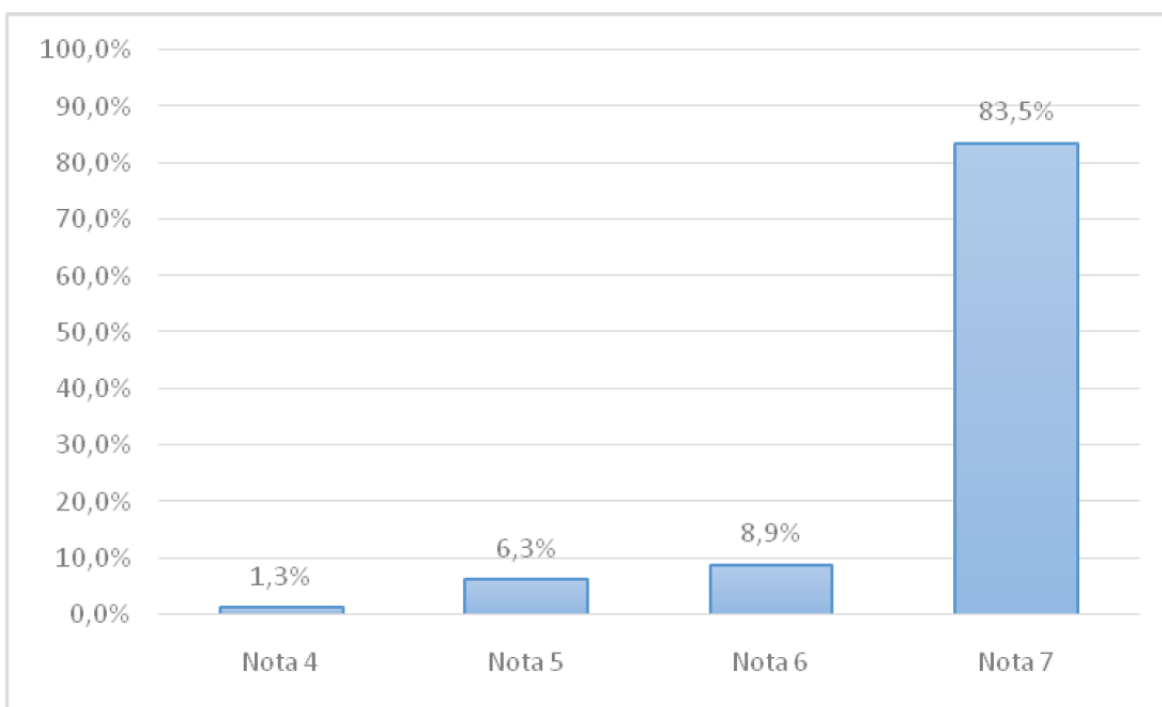


Gráfico 2 - Encuesta evaluación del programa (elaboración externa)

Un gran punto positivo valorado por el reporte es el interés de seis de las diez comunas por dar continuidad al programa, además se recalca la intención de replicar el modelo por parte de comunas que no participaron en el programa como San Joaquín, Las Condes, Estación Central y Pedro Aguirre Cerda.

Comuna	Municipio demuestra interés en continuar programa VAE	Municipio continuará con implementación del programa VAE	En caso de que continúe con Programa: ¿De qué manera continuará con el Programa? (ej: contratación de dupla, modelo de gestión de caso, modelo de reinserción, otro)
Pudahuel	SI	SI	DUPLA
Quilicura	NO	NO	NO
Maipú	SI	SI	DUPLA
Renca	NO	NO	NO
Santiago	SI	NO ESTA CLARO	DUPLA
Puente Alto	SI	SI	DUPLA
El Bosque	NO	NO	NO
La Pintana	NO	SI	PPROGRAMA PROPIO
San Bernardo	SI	SI	MODELO
Peñalolén	SI	SI	DUPLA

Tabla 3 - Continuidad del programa VAE (elaboración externa)

En cuanto a las conclusiones del informe, entre las principales críticas se menciona que “no hubo una correcta comprensión respecto a si el objetivo central del programa era la transferencia de competencias, la capacitación, la atención a los usuarios, o desarrollar un modelo de gestión de casos con conocimiento técnico”, lo mencionado presenta una alerta al plantear el análisis de los datos recopilados, dado que al no encontrarse claro el

objetivo, puede que la recolección de datos presente cierto vicios provenientes de estas inexactitudes.

Otra crítica relevante para efecto de esta tesis fue *“Se considera innecesaria la diferenciación entre beneficiarios de primera y segunda línea. Junto con ello, faltó una estrategia clara para contabilizar a aquellos pertenecientes a la segunda línea, así como corroborar si los atendidos como segunda línea cumplían con los requisitos de ingreso establecidos en el diseño del programa. Del mismo modo, tampoco quedó claro como atender específicamente a los usuarios de segunda línea”*, dicho punto alerta respecto de la utilidad de variable que diferencia el tipo de usuario, información que será relevante posteriormente en el tratamiento de la data.

Otro punto crítico resaltado por el informe fue la baja adherencia de los usuarios del programa, contando con tan solo un 42%.

Por último, un aspecto preocupante es la alta rotación en el personal en las duplas comunales, solo un tercio de las duplas se mantuvieron de inicio a fin del programa, considerando aún el breve periodo de implementación y un caso puntualmente preocupante es el de San Bernardo en donde estuvieron sin dupla durante varias semanas. Este antecedente genera preocupación respecto a la estandarización de los criterios aplicados a la hora de recolectar la data.

4. Metodología

Inicialmente se planteó realizar dos análisis cuantitativos que consistían en una regresión logística y un análisis de supervivencia.

Para la regresión logística se esperaba tener una matriz de datos individuales que permitiera describir a los usuarios demográficamente y a la vez registrar sus trayectorias dentro del programa, para posteriormente vincular estos a una variable que indicara de manera binaria la reincidencia de cada usuario en el periodo desde el cierre del programa hasta la actualidad. Para lograr tener la mejor representación posible de la reincidencia, se planteo conseguir datos provenientes del ministerio público que pudieran representar la reincidencia jurídica, es decir si la persona había sido o no procesada.

Para el análisis de supervivencia se necesitaba adicionalmente contar con la fecha de la reincidencia y poder configurar un grupo de control aleatorio y de características similares a la muestra.

Dadas diversas dificultades a la hora de conseguir los datos para efectos de este análisis, los cuales pueden ser atribuidos al nivel de resguardo por tema de privacidad de estos y dado a los altos grados de burocracia de las instituciones que los custodian, es que únicamente pudo ser realizada la regresión logística, teniendo que dejar de lado el estudio de supervivencias.

Debido a inconsistencias en la data y al bajo volumen de datos disponibles, combinado con deficiencias en el registro de esta es que los resultados obtenidos mediante la regresión logística han sido complementados con un análisis de importancias obtenido a través de dos metodologías basadas en arboles de decisión, en las cuales se profundizará durante este capítulo.

4.1. Tratamiento de datos

Inicialmente la data fue entregada por Karina Pino, profesional de Gendarmería, quien estuvo a cargo de la coordinación del Programa Volver a Empezar, la data otorgada constó del registro de cuatrocientos sesenta y ocho casos que pasaron por el programa, con ochenta y ocho columnas, que se encuentran detalladas en el Anexo 1, donde se registraron datos demográficos (edad, género, comuna, último año de educación, etc), como propios del registro de la gestión de casos (plan de reinserción, tipo de usuario, cumplimiento de los planes, etc). Cabe agregar que los datos otorgados se encontraban anonimizados por medidas de respeto a la privacidad de los usuarios.

Con el primer panel de datos, se realizaron gestiones tanto con Gendarmería como con el Ministerio Público, para obtener la reincidencia de los usuarios, sus fechas y generar un grupo de control que permitiera realizar los análisis inicialmente propuestos. Como se comentó previamente, las gestiones con el Ministerio Público fueron inviables respecto los plazos de la tesis y finalmente se obtuvo la data a través de la unidad de estudios de Gendarmería, quienes por temas de dificultades técnicas y plazos facilitaron únicamente el dato de reincidencia, de forma de variable binaria que indica si el usuario posee o no

una condena dentro de una ventana de tiempo entre el 1 de enero de 2018 y el 13 de septiembre de 2022.

Una vez realizado el cruce el siguiente paso fue reducir el número de columnas de nuestra matriz, inicialmente fueron retiradas las columnas de comentarios y las columnas de fechas al no ser relevantes para el modelo, además de otras columnas de información, como el nombre del autor, el ID y la edad en formato texto. Habiendo aislado las columnas que representaban información relevante, se conservaron las columnas de resultados por sobre las de proceso, por ejemplo, se eliminaron las columnas `aplicacion_diagnostico_area_judicial` y `cuenta_con_plan_de_reinsercion_en_el_area_judicial` para dar lugar a `reinsertado_en_ambito_judicial` y así evitar la alta correlación entre estos tipos de variables.

Se realizaron ajustes menores en la data con fin de estandarizar su contenido, como la verificación de los nombres de las comunas con o sin letras mayúsculas y corrigiendo errores de tipeo. También se realizó una adaptación de la variable `ultimo_anio_cursado_en_educacion` estandarizando su contenido y transformándolo en una variable que indica el nivel educacional con las categorías “Sin educacion”, “Básica incompleta”, “Básica completa”, “Media incompleta”, “Media completa”, “Tecnica incompleta”, “Tecnica completa”, “Superior incompleta”, “Superior completa”. Posteriormente esta variable fue discretizada con números correlativos desde el cero al ocho.

Otra variable modificada fue “Profesión”, esta variable inicialmente entregaba la profesión u oficio que el usuario declaró ejercer, para simplificar el análisis esta variable recibió un valor de 1 si es que el usuario declaró alguna profesión y un 0 en el caso contrario.

Con las variables consideradas se realizó un conteo de los valores perdidos por variables obteniéndose la siguiente tabla:

Edad	27
Genero	0
Profesion	0
Nivel educacional	65
Comuna	20
Reincidente	0
Grupo de referencia	0
Cumplimiento del Plan de Reinserción	0
Capacitado laboralmente si requiere	0
Colocado laboralmente si requiere	0
<code>reinsertado_en_ambito_familiar</code>	348
<code>reinsertado_en_ambito_social</code>	350
<code>reinsertado_en_ambito_laboral</code>	344
<code>reinsertado_en_ambito_judicial</code>	349

egresado	282
Sujeto Reinsertado FINAL	297

Tabla 4 - Valores perdidos por variable (elaboración propia)

Cabe destacar en primer lugar que las variables “Cumplimiento del Plan de Reinserción”, “Capacitado laboralmente si requiere” y “Colocado laboralmente si requiere”, en su momento fueron campos de entrada obligatorio. Dado el alto nivel de pérdida de datos en el resto de las variables propias del programa, estas fueron las únicas finalmente consideradas.

El siguiente paso fue convertir en dummies las variables categóricas (en lo que se profundizará más adelante), para así obtener la matriz de correlación, a partir de la cual se optó por descartar la variable “Capacitado laboralmente si requiere” dado a su alta correlación (0,821618) con la variable “Colocado laboralmente si requiere” (variables 2 y 3 para efecto de la gráfica).

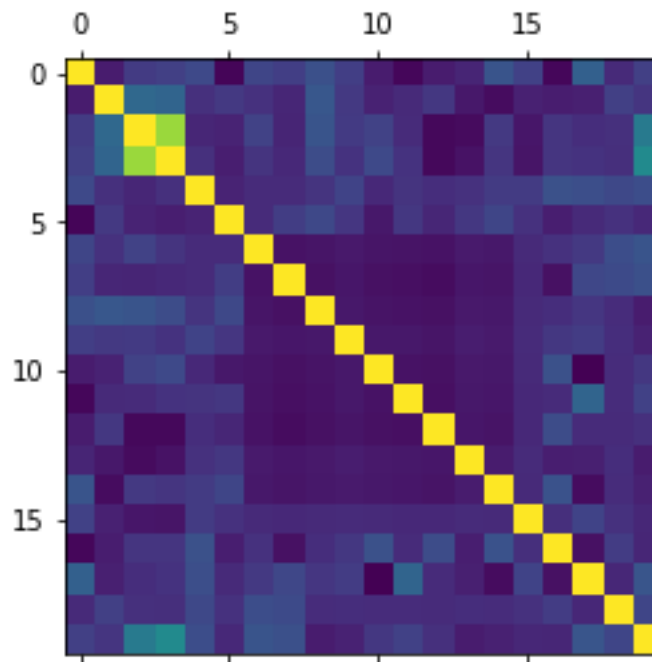


Gráfico 3 - Matriz de correlación (elaboración propia)

Una vez definido el set de variables a utilizar fueron eliminadas las filas con valores vacíos, obteniendo un dataframe con treientos setenta y siete registros, equivalente a un 80,77% de la data inicial. Con el paso de variables categóricas a dummies, la matriz de datos contó con un total de 19 variables independientes que pasan a ser descritas a continuación:

4.1.1. Variables dependientes:

- Edad: edad de los usuarios en formato de número entero
- Grupo de referencia: cantidad de personas pertenecientes al grupo de referencia o redes de apoyo, declarado por el usuario, en formato de número entero

- Colocado laboralmente si requiere: variable categórica, entrega un -1 si el usuario declaró no requerir ser colocado, un 1 si declaro requerirlo y fue colocado y un 0 si no
- Nivel educacional numérica: posee valores discretos entre 0 y 8 en orden respectivo a “Sin educacion”, “Básica incompleta”, “Básica completa”, “Media incompleta”, Media completa”, “Tecnica incompleta”, “Tecnica completa”, “Superior incompleta” y “Superior completa”.
- Comuna_La Pintana: Variable dummy, entrega 1 si el usuario pertenece a la comuna de La Pintana y 0 si no.
- Comuna_Maipu: Variable dummy, entrega 1 si el usuario pertenece a la comuna de Maipu y 0 si no.
- Comuna_Peñalolen: Variable dummy, entrega 1 si el usuario pertenece a la comuna de Peñalolen y 0 si no.
- Comuna_Pudahuel: Variable dummy, entrega 1 si el usuario pertenece a la comuna de Pudahuel y 0 si no.
- Comuna_Puente Alto: Variable dummy, entrega 1 si el usuario pertenece a la comuna de Puente Alto y 0 si no.
- Comuna_Quilicura: Variable dummy, entrega 1 si el usuario pertenece a la comuna de Quilicura y 0 si no.
- Comuna_Renca: Variable dummy, entrega 1 si el usuario pertenece a la comuna de Renca y 0 si no.
- Comuna_San Bernardo: Variable dummy, entrega 1 si el usuario pertenece a la comuna de San Bernardo y 0 si no.
- Comuna_Santiago: Variable dummy, entrega 1 si el usuario pertenece a la comuna de Santiago y 0 si no.
- Genero_Masculino: Variable dummy, entrega un 1 si el usuario es varón y un 0 si es mujer.
- Profesion_Si: Variable dummy, entrega un 1 si el usuario declaró contar con alguna profesión u oficio y un 0 si no.
- Reincidente_Sí: Variable dummy, entrega 1 si el usuario era reincidente al momento de cursar el programa y un 0 si no.
- Cumplimiento del Plan de Reinserción_Si: Variable Dummy, entrega 1 si el usuario completo su plan de reinserción y 0 si no.

4.2. Regresión logística

Una regresión logística es un método estadístico multivariante que al igual que cualquier tipo de regresión, busca relacionar y explicar una variable a través de otro grupo de variables. A diferencia de la regresión lineal, la variable dependiente no es una variable métrica, sino que particularmente es una variable dicotómica, generalmente que adopta valores de cero o uno, mientras que el conjunto de variables independientes puede ser de tipo métrica o no.

Las variables explicativas pueden ser cuantitativas o cualitativas y la ecuación que representa el modelo no es lineal sino más bien exponencial, que puede ser simplificada mediante transformación logarítmica.

La regresión logística tiene por objetivo estimar la probabilidad de ocurrencia de un suceso a partir de variables regresoras y a su vez identificar como estas variables regresoras afectan la posibilidad de ocurrencia del suceso.

Si se quisiese realizar este análisis utilizando una regresión lineal múltiple se producirían dos problemas, en primer lugar, se violaría el supuesto de la distribución normal de los errores aleatorios y además los valores predictados no podrían ser considerados como probabilidades, ya que no toman valores entre cero y uno.

Por otro lado, si se quisiese realizar un Análisis Discriminante, se encontraría un problema al abordar el supuesto de multinormalidad de las variables regresoras, es por esto que el método de regresión logística multivariada nos entrega la flexibilidad necesaria para trabajar con variables de distintos tipos con menores restricciones.

En cuanto al modelo, llamaremos $X^T = (x_1, x_2, \dots, x_k)$ al vector de variables regresoras y denotaremos a la probabilidad condicionada de que la variable dependiente sea 1 como:

$$P(y = 1|X) = \pi(X) \tag{1}$$

El logaritmo del modelo de regresión logística está dado por:

$$g(X) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n \tag{2}$$

Entonces el modelo de regresión logística múltiple es:

$$\pi(X) = \frac{e^{g(X)}}{1 + e^{g(X)}} \tag{3}$$

Así como para la regresión lineal múltiple, los coeficientes son calculados mediante el método de mínimos cuadrados, para la regresión logística múltiple es necesario usar el método de máxima verosimilitud. Si bien los cálculos serán realizados mediante el software R Studio, es necesario describir el procedimiento en este capítulo en función de su posterior análisis.

La función de verosimilitud expresa la probabilidad de los datos observados como una función de parámetros desconocidos. Los Estimadores de Máxima Verosimilitud de esos parámetros son aquellos que están en concordancia con los datos observados.

Inicialmente debemos considerar n observaciones de nuestra variable dependiente tal que esta variable pueda tomar valores de $y = 0$ o $y = 1$.

Sea $y^T = (y_1, y_2, y_3, \dots, y_n)$ donde $y_i \sim B(1, \pi_i)$ y sea $x_i^T = (1, x_{i1}, x_{i2}, x_{i3}, \dots, x_{in})$ la i-esima observación para las k variables explicativas.

Así el modelo de regresión logística se encuentra dado por:

$$P[y_i = 1|X_i] = \pi(X_i) = \frac{e^{g(X_i)}}{1 + e^{g(X_i)}} \tag{4}$$

De manera equivalente:

$$P[y_i = 1|X_i] = \frac{\text{Exp}(\beta_0 + \sum_{j=1}^k \beta_j x_{ij})}{1 + \text{Exp}(\beta_0 + \sum_{j=1}^k \beta_j x_{ij})} \quad (5)$$

Además, la probabilidad de que $y=0$ es:

$$P[y_i = 0|X_i] = 1 - P[y_i = 1|X_i] \quad (6)$$

Por lo tanto:

$$P[y_i = 0|X_i] = \frac{1}{1 + \text{Exp}(\beta_0 + \sum_{j=1}^k \beta_j x_{ij})} \quad (7)$$

Para facilitar la notación definiremos $x_{i0} = 1$, entonces:

$$P[y_i = 1|X_i] = \pi(X_i) = \frac{e^{\beta X_i^T}}{1 + e^{\beta X_i^T}} \quad (8)$$

Donde β corresponde al vector de coeficientes y X_i^T corresponde al vector transpuesto de variables explicativas. En cuanto a la transformación logística de la probabilidad π_i , la podemos definir como:

$$\lambda_i = \text{Ln}\left(\frac{\pi_i}{1 - \pi_i}\right) = \sum_{j=0}^k \beta_j x_{ij} \quad (9)$$

Para obtener la estimación máximo verosímil para el vector β se debe escribir la función de densidad de probabilidad del vector y , el cual es equivalente a n funciones $B(1, \pi_i)$.

$$f(y_i; \pi_i) = \prod_{i=1}^n \pi_i^{y_i} (1 - \pi_i)^{1-y_i} \quad (10)$$

$$f(y_i; \pi_i) = \prod_{i=1}^n \left(\frac{\pi_i}{1 - \pi_i}\right)^{y_i} (1 - \pi_i) \quad (11)$$

$$f(y_i; \pi_i) = \left\{ \prod_{i=1}^n (1 - \pi_i) \right\} \left\{ \prod_{i=1}^n \text{Exp}\left[\text{Ln}\left(\frac{\pi_i}{1 - \pi_i}\right)^{y_i}\right] \right\} \quad (12)$$

$$f(y_i; \pi_i) = \left\{ \prod_{i=1}^n (1 - \pi_i) \right\} \text{Exp}\left[\sum_{i=1}^n y_i \text{Ln}\left(\frac{\pi_i}{1 - \pi_i}\right)\right] \quad (13)$$

Reemplazando (9) en (13) se obtiene:

$$f(y_i; \pi_i) = \left\{ \prod_{i=1}^n (1 - \pi_i) \right\} \text{Exp}\left\{\sum_{i=1}^n y_i \sum_{j=0}^k \beta_j x_{ij}\right\} \quad (14)$$

$$f(y_i; \pi_i) = \left\{ \prod_{i=1}^n (1 - \pi_i) \right\} \text{Exp}\left\{\sum_{j=0}^k \left(\sum_{i=1}^n y_i x_{ij}\right) \beta_j\right\} \quad (15)$$

Calculando el logaritmo natural de (15), también llamado función soporte.

$$l(\pi_i, y_i) = \sum_{j=0}^k \left(\sum_{i=1}^n y_i x_{ij} \right) \beta_j + \sum_{i=1}^n \text{Ln} (1 - \pi_i) \quad (16)$$

Pero $1 - \pi_i$ es igual a (7), por lo tanto:

$$\text{Ln}(1 - \pi_i) = -\text{Ln} \left[1 + \text{Exp} \left(\sum_{j=0}^k \beta_j x_{ij} \right) \right] \quad (17)$$

Reemplazando (17) en (16):

$$l(\pi_i, y_i) = \sum_{j=0}^k \left(\sum_{i=1}^n y_i x_{ij} \right) \beta_j - \sum_{i=1}^n \text{Ln} \left[1 + \text{Exp} \left(\sum_{j=0}^k \beta_j x_{ij} \right) \right] \quad (18)$$

Como se puede observar, la función ya no depende de π_i , es entonces que puede ser denotada como:

$$l(\beta) = \sum_{j=0}^k \left(\sum_{i=1}^n y_i x_{ij} \right) \beta_j - \sum_{i=1}^n \text{Ln} \left[1 + \text{Exp} \left(\sum_{j=0}^k \beta_j x_{ij} \right) \right] \quad (19)$$

Sea:

$$t_j = \sum_{i=1}^n y_i x_{ij} \quad (20)$$

Entonces:

$$l(\beta) = \sum_{j=0}^k t_j \beta_j - \sum_{i=1}^n \text{Ln} \left[1 + \text{Exp} \left(\sum_{j=0}^k \beta_j x_{ij} \right) \right] \quad (21)$$

Las ecuaciones de verosimilitud se obtienen derivando con respecto a β_j e igualando a cero.

$$\frac{\partial L}{\partial \beta_j} = \sum_{i=1}^n y_i x_{ij} - \sum_{i=1}^n x_{ij} \left[\frac{\text{Exp}(\sum_{j=0}^k \beta_j x_{ij})}{1 + \text{Exp}(\sum_{j=0}^k \beta_j x_{ij})} \right] \quad (22)$$

Por lo tanto, las ecuaciones de verosimilitud son:

$$\sum_{i=0}^n x_{ij} y_i - \sum_{i=0}^n x_{ij} \pi_i = 0 \quad \text{para todo } j = 0, 1, 2, 3, \dots, k \quad (23)$$

$$\sum_{i=0}^n x_{ij} (y_i - \pi_i) = 0 \quad (24)$$

Finalmente, el cometido es encontrar el vector β tal que el estimador π_i cumpla con las ecuaciones de verosimilitud.

4.3. Random Forest

Random Forest es un algoritmo de machine learning introducido por Leo Breiman en 2001 perteneciente a los métodos de ensamblaje, dado que el modelo a su vez depende de submodelos, de modo que los indicadores que este entrega dependen de los modelos más pequeños. Los submodelos utilizados por esta metodología son Árboles de Clasificación y Regresión (ACR) y son claves en la comprensión de la metodología.

Los ACR son modelos que se basan en la subdivisión o partición de datos para estimar la distribución condicional de la variable dependiente, la cual se representa como un vector:

$$Y = (y_1, y_2, y_3, \dots, y_n) \quad (25)$$

Mientras que las variables explicativas se encuentran representadas a través de una matriz:

$$X = (X_1, X_2, X_3, \dots, X_p) \quad (26)$$

$$\text{con } X_j = (x_{1j}, x_{2j}, x_{3j}, \dots, x_{nj}) \text{ para } j \in \{1, \dots, p\} \quad (27)$$

El objetivo de las subdivisiones es agrupar Y en función de X de forma que cada subgrupo sea lo más homogéneo posible.

Con respecto a las variables dependientes dicotómicas, el modelo interpreta cada valor individual de las variables explicativas como un posible punto de corte binario del espacio, de modo que las áreas que se formen a partir de este punto sean lo más homogéneas posibles respecto a la variable dependiente.

Para definir si se debe generar una nueva segmentación es necesario analizar si el error disminuye o aumenta con un nuevo nodo. Siendo los datos específicos en un nodo (m) del ACR:

$$Y^{(m)} = (y_1^{(m)}, \dots, y_{n^{(m)}}^{(m)}) \quad y \quad X^{(m)} = (X_1^{(m)}, \dots, X_p^{(m)}) \quad (28)$$

Además, sea $X_s^{(m)}$ la variable explicativa considerada para agregar el corte en el nodo m, sea $C^{(m)} = \{x_i^{(m)}\}_{i \in \{1, \dots, n^{(m)}\}}$ el conjunto de los posibles valores únicos para la variable s en el nodo m y se $c \in C$ la observación considerada para realizar el corte. A raíz de lo anterior, el nodo inicial $Y^{(m)}$, queda dividido en dos, $Y^{(mi)}$ y $Y^{(md)}$, $Y^{(mi)}$ corresponde al grupo de elementos tales que $X_s^{(m)} \leq c$, mientras que $Y^{(md)}$ corresponde al grupo de elementos tales que $X_s^{(m)} > c$. Así, el incremento o disminución del error estaría definido por la siguiente formula:

$$\Delta(Y^{(m)}) = L(Y^{(m)}) - \left[\frac{n^{(mi)}}{n^m} L(Y^{(mi)}) + \frac{n^{(md)}}{n^m} L(Y^{(md)}) \right] \quad (29)$$

En este caso $n^{(mi)}$ y $n^{(md)}$ representan la cantidad de elementos que se encuentran a la izquierda y a la derecha del nodo respectivamente, mientras que L(.) es una función de perdida a escoger a partir de la variable dependiente.

Según Breiman el bagging o también conocido como bootstrap aggregating, es una técnica que, al combinar múltiples árboles, busca reducir la varianza de los valores ajustados de los árboles de regresión de manera que se reduzca significativamente el riesgo de sobreajuste. Lo anterior se logra al combinar los resultados de múltiples modelos y promediar los valores predichos para obtener un estimador estandarizado.

Una fortaleza de estos modelos es que cada submodelo es entrenado con una remuestra, la cual es obtenida mediante bootstrapping con reemplazo o con una rotación, de modo que cada muestra deja fuera una porción de los datos.

Otro punto fuerte es la posibilidad que entregan estos modelos de poder calcular la importancia de las variables, dado que emplea la misma lógica para los predictores, es decir, por cada submodelo, genera una selección aleatoria de las variables a considerar, lo que permite identificar submodelos que poseen mayor error al no contar con variables más importantes.

4.4. XGBoost

El modelo Extreme Gradient Boosting propuesto por Chen y Gestrin en 2016 es de aprendizaje supervisado y al igual que el anterior consiste en ensamblaje de árboles. Es entrenado de forma iterativa y aditiva, y su predicción tiene la siguiente forma:

$$\hat{y}_l = \sum_{k=1}^n f_k(x_i), \quad f_k \in F \quad (30)$$

Donde \hat{y}_l es el valor predicho, f_k representa un árbol de regresión, x_i es un valor de entrada, n es el número de funciones regresoras y F es el espacio de todo f_k posible. La siguiente es la función objetivo minimizada en el modelo:

$$L(\phi) = \sum_{i=1}^m l(y_i, \hat{y}_l) + \sum_{k=1}^n \Omega(f_k) \quad (31)$$

Donde:

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|\omega\|^2 \quad (32)$$

En este caso l es la función de pérdida entre el valor observado y_i y el valor predicho \hat{y}_l . El segundo término, Ω es el término de regularización, que penaliza la complejidad del modelo para evitar el sobreajuste. γ es la complejidad de cada hoja. T es el número de hojas en cada árbol de regresión. λ es el parámetro de compensación que escala la penalización y ω es el vector de puntajes en las hojas.

A diferencia de Random Forest, XGBoost es un boosting de árboles de decisión, lo que implica que cada árbol aprende del árbol anterior.

4.5. Random Forest vs XGBoost

A continuación, se presenta un cuadro comparativo entre ambas metodologías:

Random Forest	XGBoost
---------------	---------

En Random Forest los árboles de decisión son construidos de manera independiente y de forma simultánea, de modo que si se construyen n árboles en paralelo, estos poseerán distintos sets de datos y de variables.	XGBoost construye un árbol a la vez, ya que la data usada para un árbol es tomada en cuenta en el siguiente, buscando corregir los errores en cada iteración y sumando ponderadamente al resultado final.
Una vez generados todos los árboles de decisión el resultado es calculado como el promedio de los valores de todos los árboles.	El resultado va siendo agregado en cada uno de los pasos, lo que permite obtener una estimación del resultado antes de que el algoritmo termine sus iteraciones.
Si bien, Random Forest es menos usado que XGBoost por sus tiempos de cálculo, posee valor en la simplicidad de estos, por lo que es frecuentemente usado por principiantes.	Sus tiempos de cálculo son óptimos, pero su nivel de complejidad es de los más altos entre los algoritmos basados en arboles de decisión.
Random Forest es óptimo de usar en muestras donde existen grandes cantidades de valores perdidos o mucho ruido ya que la técnica se centra en evitar el sobreajuste del modelo.	El análisis de XGBoost es mucho más sensible al ruido y los valores perdidos de la muestra, presenta un óptimo rendimiento en al tratar con datos de alta calidad.
Random Forest tiene muchos árboles con hojas de igual peso, por lo que se puede obtener fácilmente una gran precisión con los datos disponibles.	XGBoost no tiene en cuenta el número de hojas presentes en el algoritmo. Si la predictibilidad del modelo no es buena, el algoritmo funciona mejor con más hojas en el árbol de decisión. Esto mejora el sesgo y los resultados dependen completamente de los datos presentes en el algoritmo.

Gráfico 4 - Comparación Random Forest vs XGBoost (elaboración propia)

Con el fin de conseguir resultados que otorguen una mayor confiabilidad es que se realizó análisis combinado entre Random Forest y XGBoost. Dado que XGBoost posee debilidades al trabajar con data incompleta, es que se decidió trabajar con el dataframe obtenido al eliminar los valores perdidos (el mismo con el que se realiza el análisis de regresión logística multivariable).

Para combinar los resultados obtenidos con ambas metodologías y entendiendo que los resultados de importancia de las variables explicativas se encuentran en diferentes escalas, es que se realizó una normalización de los resultados, ordenando las importancias con valores entre 0 y 1, en donde la variable menos importante vale 0 y la más importante vale 1. Una vez realizada la normalización, los resultados por variables fueron promediados obteniéndose los resultados finales.

5. Análisis exploratorio de datos

En el presente capítulo nos enfocaremos en analizar el estado inicial de la data con las variables consideradas, es decir, previo a eliminar los datos con valores perdidos, para poder comparar ambas muestras e interpretar como el tratamiento de datos afecta a la distribución de las variables.

Además, se busca comprender la composición de la muestra estudiando como la distribución de las variables se relaciona a priori con la variable dependiente, buscando visibilizar para que valores de cada variable se encuentran las mayores concentraciones en la tasa de reincidencia.

5.1. Valores perdidos

Inicialmente el dataset se componía de cuatrocientos sesenta y ocho registros y nueve variables explicativas, entre las cuales se registraron variables categóricas, numéricas y binarias. Posteriormente, una vez realizada la limpieza de valores perdidos, el dataset utilizado en los análisis estuvo compuesto por trescientos setenta y siete registros y dieciocho variables, de las cuales una es numérica (Edad) y dos son categóricas (Nivel Educativo y Grupo de Referencia), el resto son solo variables dummies obtenidas a partir de las variables iniciales.

5.1.1. Edad

En un inicio se encontraron cuatrocientos cuarenta registros respecto a esta variable. Los usuarios del programa promediaron 34,90 años con una desviación estándar igual a 10,20. Mientras que, de los trescientos setenta y siete registros empleados finalmente, equivalentes al 85,7% de la muestra inicial, la variable promedió 34,56 años, con una desviación igual a 10,02. A continuación se presentan los datos resumidos en una tabla en conjunto con los histogramas de la variable.

Edad	Cantidad	Promedio	Desviación estándar
Antes de limpieza	440	34,90	10,20
Después de limpieza	377	34,56	10,02

Tabla 5 - Medidas de tendencia central, variable Edad (elaboración propia)

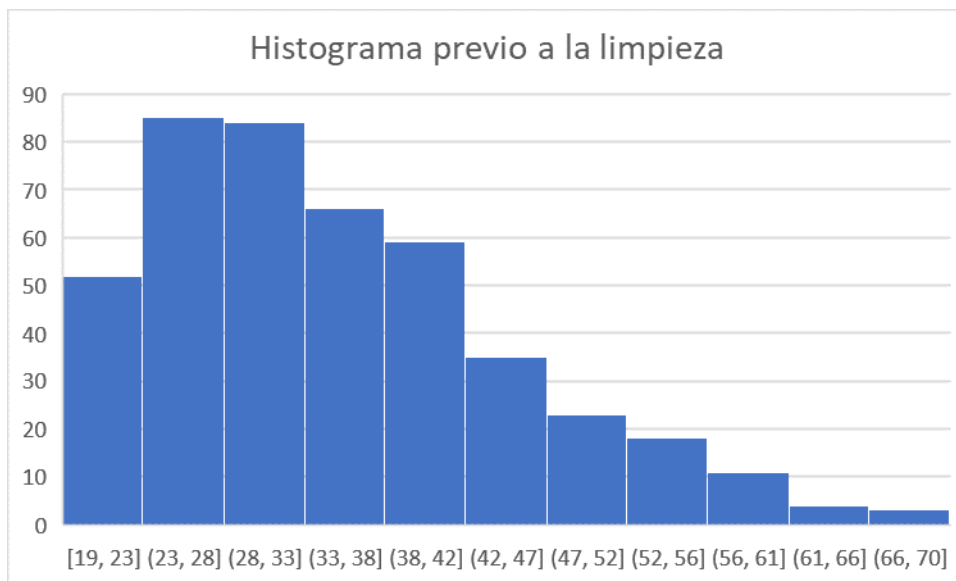


Gráfico 5 - Histograma variable edad, previo a la limpieza de datos (elaboración propia)

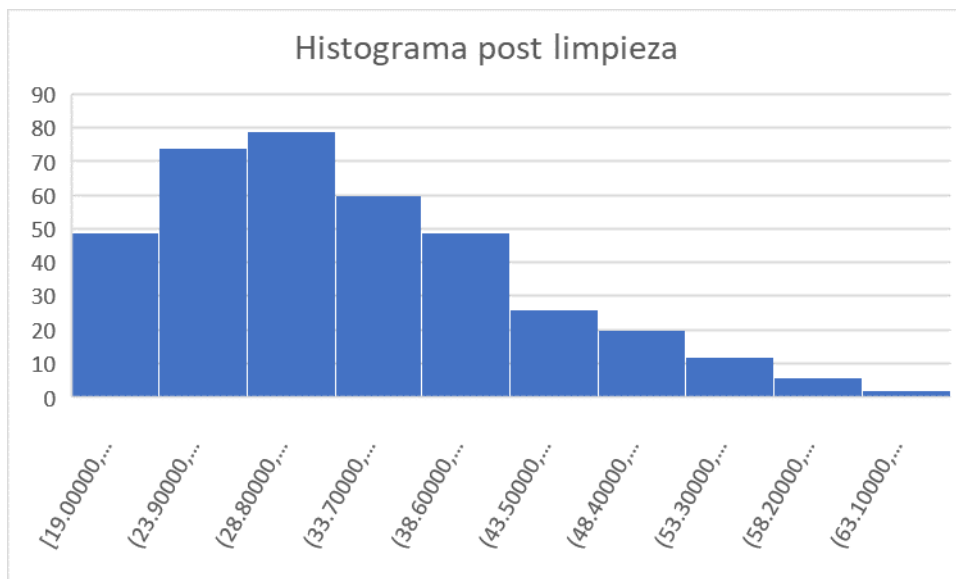


Gráfico 6 - Histograma variable edad, posterior a la limpieza de datos (elaboración propia)

A partir de los histogramas se puede observar que luego de la limpieza de datos ocurren variaciones mínimas, manteniéndose la distribución de datos con una máxima frecuencia entre los 28 y 33 años.

Para completar el análisis a continuación se presenta la distribución de los quintiles antes y después de la limpieza de datos:

Quintil	Antes de la limpieza	Después de la limpieza
---------	----------------------	------------------------

1	26.0	25.2
2	30.6	30.0
3	36.0	36.0
4	42.6	42.8
5	68.0	68.0

Tabla 6 - Distribución de la variable edad (elaboración propia)

Es posible observar mínimas variaciones respecto a la data original, más aún, los quintiles tres y cinco se mantienen exactamente iguales, lo que permite suponer que la muestra final respecto a la variable edad posee representatividad respecto a la población del programa.

5.1.2. Género

Inicialmente se encontró registro para todos los elementos de la muestra, es decir, se encontraron cuatrocientos sesenta y ocho registros de la variable género, la cual es una variable binaria que toma valores de masculino y femenino. En la data original trescientos ochenta registros correspondían al género masculino, lo que es equivalente al 81,2% de la muestra, siendo complementariamente ochenta y ocho registros pertenecientes al género femenino, lo que representa el restante 18,8% de la muestra.

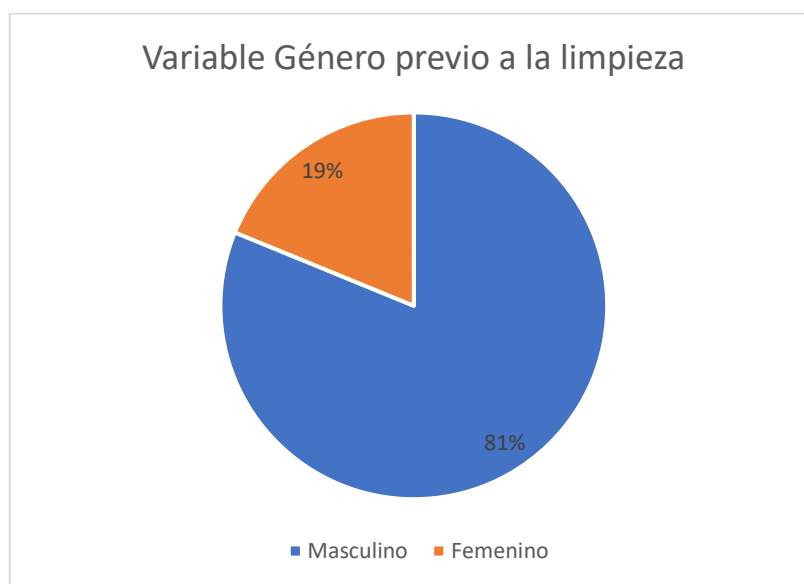


Gráfico 7 - Distribución de la variable género, previo a la limpieza de datos (elaboración propia)

Luego de la limpieza de datos, los trescientos sesenta y siete registros se distribuyeron de la siguiente manera:

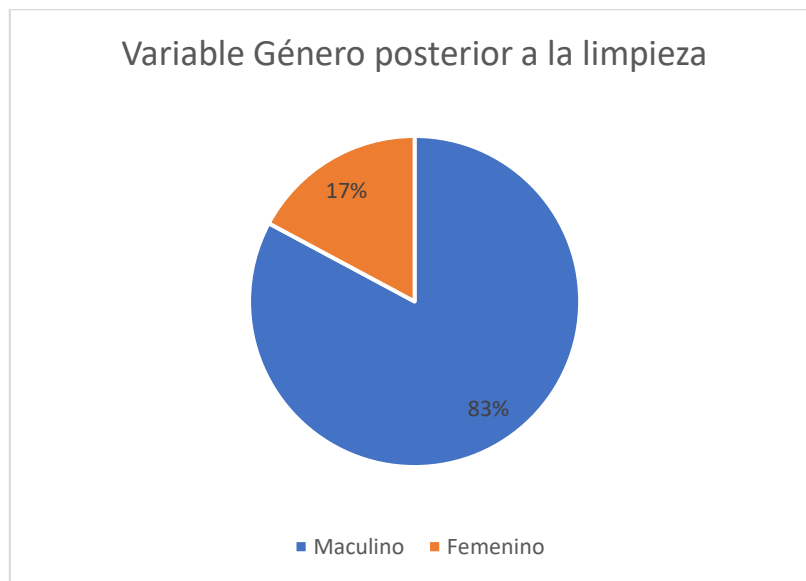


Gráfico 8 - Distribución de la variable género, posterior a la limpieza de datos (elaboración propia)

Se contabilizaron trescientos cuatro registros de género masculino y sesenta y tres registros de género femenino, lo que equivale a un 82,8% y un 17,2% respectivamente. Si bien se observan mínimas diferencias entre la distribución de la variable en uno y en otro dataset, se estima que estas diferencias no generan cambios significativos y que la variable se encuentra bien representada.

Cabe acotar que, durante el tratamiento de datos, la variable pasó de ser una variable categórica binaria que entregaba los valores “masculino” o “femenino” a una variable dummy que entregaba un 1 si el género era masculino y un 0 si no.

5.1.3. Profesión

Inicialmente la variable poseía registro para todos los usuarios, es decir cuatrocientos sesenta y ocho registros y podía representar dos valores “SI” y “NO”, donde “SI” indicaba que el usuario declaró haber ejercido una profesión u oficio previo a su encarcelamiento y “NO” si declaraba no haber trabajado o ejercido profesión u oficio alguno.

De la data inicial, trescientos treinta y dos personas declararon haber tenido un oficio u profesión y ciento treinta y seis personas declararon nunca haber ejercido oficio alguno. Lo que es equivalente al 70,9% y 29,1% de la muestra respectivamente.

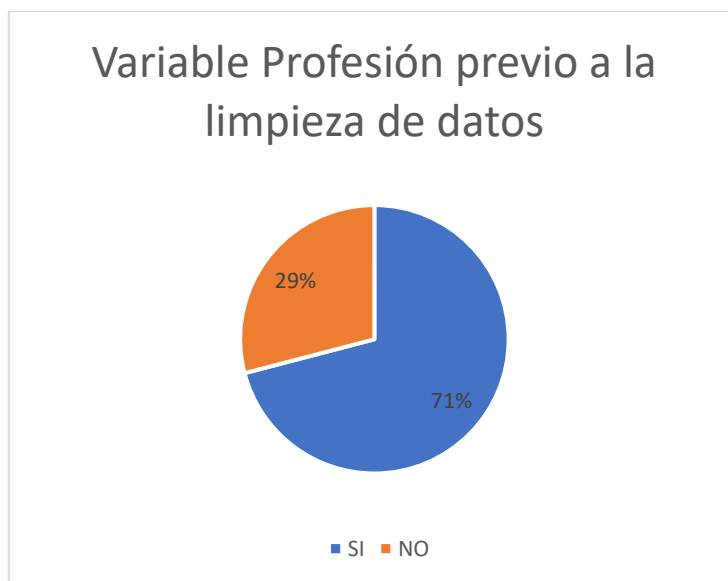


Gráfico 9 - Distribución de la variable profesión, previo a la limpieza de datos (elaboración propia)

A continuación, se presenta la distribución para la variable “Profesión” una vez realizada la limpieza de datos

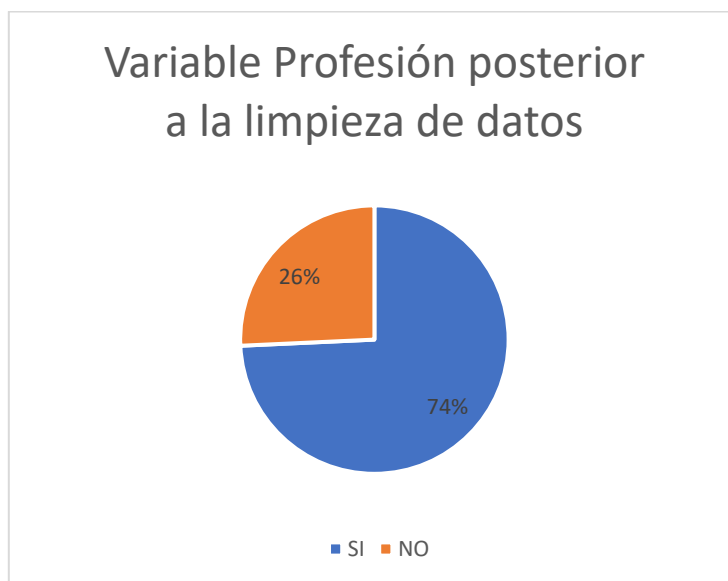


Gráfico 10 - Distribución de la variable profesión, posterior a la limpieza de datos (elaboración propia)

De los treientos setenta y siete registros finales, un 74,3% declaró haber poseído una profesión u oficio, lo que equivale a doscientos ochenta usuarios, mientras que el 25,7% de los usuarios declaró no haber tenido ninguna profesión previa a su condena, los que representa a noventa y siete usuarios.

Si bien, esta es una de las variables que presentó una de las mayores variaciones porcentuales posterior a la limpieza de datos, se considera que la muestra continúa

siendo representativa. Cabe destacar que la variable pasó de ser categórica a dummy, en donde la variable entrega un 1 cuando el usuario declaró haber tenido una profesión u oficio y un 0 si no.

5.1.4. Nivel educacional

Inicialmente esta variable poseía cuatrocientos un registro, equivalente al 85,7% de los cuatrocientos sesenta y ocho registros iniciales. Esta variable comprende nueve valores posibles ordenados correlativamente desde “Sin educación” Pasando por “Educación básica”, “Educación media”, “Educación técnica”, “Educación superior”, todas en sus variaciones “Completa” e “Incompleta”. A continuación, se presentan las distribuciones de las variables previo y posterior a la limpieza de datos.

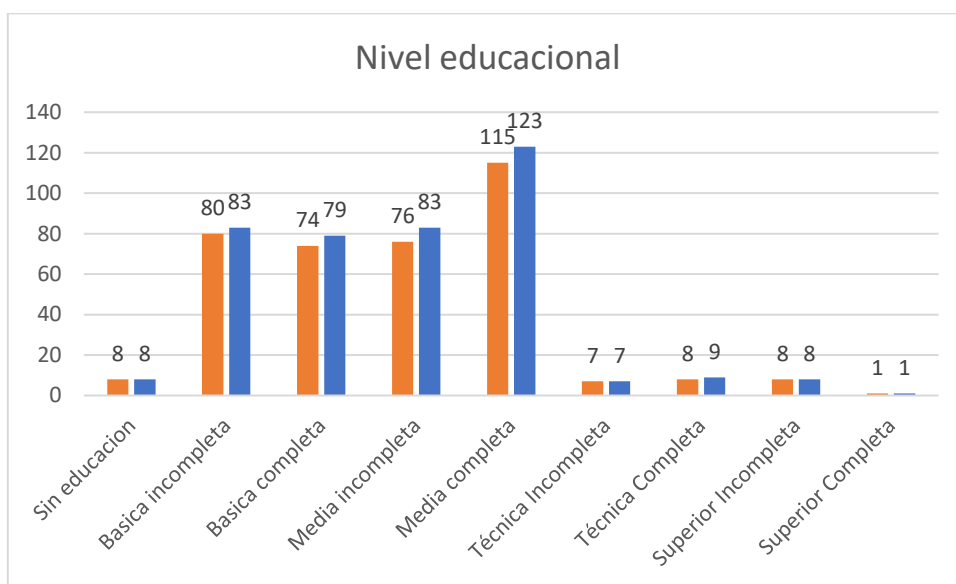


Gráfico 11 - Distribución de la variable nivel educacional, antes y después de la limpieza de datos (elaboración propia)

El gráfico presenta en color azul los datos previos a la limpieza de datos y en naranja los datos posteriores a la limpieza de datos. Se puede apreciar que la distribución en ambos casos es similar concentrándose el grueso de los datos entre la educación básica incompleta y la educación media completa. Es posible apreciar que en ambos casos la alternativa con mayor frecuencia fue la educación media completa con ciento veintitrés y ciento quince registros, lo que equivale a un 30,7% y un 30,5% respectivamente.

Se aprecia que la variable conserva representatividad de la muestra original. Finalmente es necesario mencionar que, con objeto de realizar los posteriores análisis, la variable fue convertida en una numérica discreta, asociándose a los niveles valores desde cero a ocho de forma correlativa.

5.1.5. Comuna

Esta variable posee inicialmente cuatrocientos cuarenta y ocho registros, los que corresponden al 95,7% de los cuatrocientos sesenta y ocho registros originales. La entrega los nombres de las diez comunas que participaron del programa, las cuales son: “El Bosque”, “La Pintana”, “Maipú”, “Peñalolén”, “Pudahuel”, “Puente Alto”, “Quilicura”, “Renca”, “San Bernardo” y “Santiago”. Luego de la limpieza, la cantidad de registros descendió a trecientos setenta y siete, lo que implica un cambio porcentual de 19,4%, lo que apunta a que los cambios porcentuales de los registros de cada comuna se deberían mantener en ese rango.

Esta es una variable interesante de evaluar, ya que las variaciones netas de cada comuna se deben a sus prácticas de registros de datos, es decir, las comunas que fueron más rigurosas en registrar los datos de los usuarios son aquellas que menos perdidas presentarán por valores vacíos. A continuación, se presenta la tabla resumen con las variaciones porcentuales respectivas a cada comuna antes y luego de la limpieza de datos.

Comuna	Registros iniciales	Registros Finales	Variación porcentual
El Bosque	50	49	2.0%
La Pintana	43	35	18.6%
Maipú	43	43	0.0%
Peñalolén	41	37	9.8%
Pudahuel	43	29	32.6%
Puente Alto	41	37	9.8%
Quilicura	51	40	21.6%
Renca	52	49	5.8%
San Bernardo	42	27	35.7%
Santiago	42	31	26.2%

Tabla 7 - Distribución de la variable comuna, antes y después de la limpieza de datos (elaboración propia)

Cabe destacar a Maipú (0,0%), El Bosque (2,0%) y Renca (5,8%) las cuales fueron las comunas con menor variación dado a la baja cantidad de valores perdidos en sus registros. Mientras que San Bernardo (35,7%), Pudahuel (32,6%) y Santiago (26,2%) fueron las comunas con el peor rendimiento y la mayor cantidad de valores perdidos, el caso de Pudahuel es realmente alarmante al conservar menos de treinta registros.

Es posible observar que inicialmente los registros poseen una distribución bastante homogénea, lo que posteriormente se pierde al realizar la limpieza de datos. Lo anterior se encuentra refrendado al comparar sus promedios y desviaciones a partir de la siguiente tabla:

	Registros iniciales	Registros Finales
Promedio	44.8	37.7
Desviación	4.36	7.68

*Tabla 8 - Medidas de tendencia central para la variable comuna, antes y después de la limpieza de datos
(elaboración propia)*

Es necesario aclarar que esta variable categórica, se transformó para propósitos del modelo en nueve variables dummies, dejando fuera a la comuna de El Bosque. Si bien es posible distinguir que la variable final no representa bien la población original, la relevancia de esta variable es muy alta para el análisis dado el fuerte componente territorial del Programa Volver a Empezar. Por lo tanto, esta variable será considerada en el análisis, teniendo los resguardos correspondientes a la hora de interpretar los resultados.

5.1.6. Reincidente

Esta variable adopta como propósito indicar si el usuario del programa era reincidente al momento de pasar por este. Inicialmente de los cuatrocientos sesenta y ocho registros de la variable, solo dos registros indicaban ser reincidentes. Si bien, posterior a la limpieza de datos estos registros se mantuvieron, el bajo número de usuarios reincidentes no se condice con la proporción de la tasa de reincidencia nacional (52,9%) y es insuficiente para poder realizar un análisis estadístico, es por esto a raíz de este análisis se decide desestimar el uso de esta variable de cara a los posteriores análisis.

Se sugieren dos posibles explicaciones a la distribución de esta variable, en primer lugar, se apunta a un posible error en los registros, donde posiblemente la alternativa “NO” se encontrase como predefinida y que al no ingresar información el usuario quedase registrado como no reincidente. Mientras que por otro lado se apunta a la selección de los usuarios realizada por Gendarmería para el ingreso al programa, generalmente los usuarios que optan a este tipo de beneficios son usuarios con buena conducta, quienes probablemente posean una menor tasa de reincidencia.

5.1.7. Grupo de referencia

Esta variable apunta a cuantificar las redes de apoyo que el usuario declara al momento de ingresar al programa, lo cual se encuentra registrado con una variable numérica que adopta el valor de 0 cuando el usuario no declara grupo de referencia.

Inicialmente esta variable contó con cuatrocientos sesenta y ocho registros, equivalentes al total de la muestra inicial. A continuación, se presentan los gráficos de distribución de la variable antes y después de la limpieza de datos.

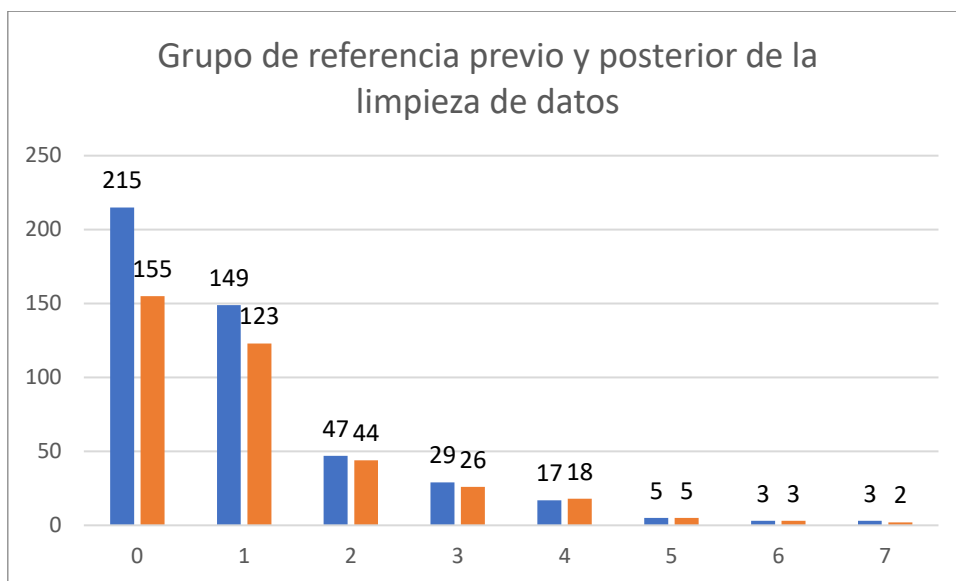


Gráfico 12 - Distribución de la variable grupo de referencia, antes y después de la limpieza de datos (elaboración propia)

Si bien la pérdida de datos es mayor en proporción para los registros de mayor frecuencia, esto se puede deber a un mal registro de datos, en donde haya una confusión entre usuarios que declararon no tener grupo de referencia, con quienes no declararon nada. Pese a lo anterior, es posible observar dos distribuciones con tendencias similares, en ambos casos se observa que los registros con grupos de referencias más numerosos son siempre menos probables.

En consideración de lo representado, es que se propone que la variable si se encuentra bien representada en la muestra final obtenida luego de la limpieza de datos.

5.1.8. Cumplimiento del plan de re inserción

Esta variable define si el usuario cumplió con el plan de re inserción a cabalidad o si no lo hizo. Esta variable posee todos los registros iniciales a diferencia de la mayoría de las variables propias del programa, las cuales presentaban una muy baja cantidad de registros. A continuación, se presentan los gráficos para la distribución de la variable antes y después de la limpieza de datos.



Gráfico 13 - Distribución de la variable cumplimiento del plan de reinserción, previo a la limpieza de datos (elaboración propia)



Gráfico 14 - Distribución de la variable cumplimiento del plan de reinserción, posterior a la limpieza de datos (elaboración propia)

Es posible ver que las distribuciones en ambos casos son muy similares, por lo que se entiende que la variable si representa de forma correcta la muestra inicial. Cabe destacar que como se realizó con otras variables, esta fue transformada en una dummy que entrega un 1 si el plan fue cumplido y un 0 si no.

5.1.1.9. Colocado laboralmente si lo requiere

Inicialmente esta variable posee tres distinciones, en primer lugar, registra un -1 si el usuario no requirió colocación laboral, de haber requerido colocación laboral y no haber sido colocado, la variable entrega un 0 y un 1 en caso de haber sido colocado laboralmente. Esta variable al igual que la anterior cuenta con todos los registros de la muestra inicial.

El siguiente gráfico presenta una comparación para la distribución de la variable antes y después de la limpieza de datos.

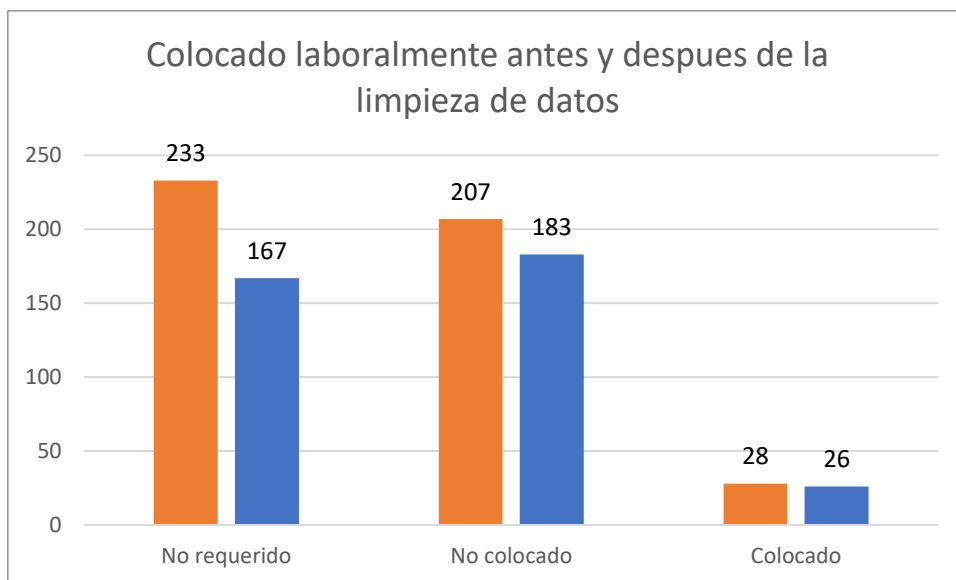


Gráfico 15 - Distribución de la variable colocado laboralmente, antes y después de la limpieza de datos (elaboración propia)

Si bien se aprecia que en las opciones “No colocado” y “Colocado” existen pérdidas proporcionalmente similares, es evidente que la opción “No requerido” posee pérdidas significativamente mayores, lo cual puede ser atribuible a malas prácticas a la hora de registrar los datos, por ejemplo, que “No requerido” fuese la opción predefinida y que la consulta no se hiciera de manera correcta al usuario. Pese a lo anterior se reconoce que las variables se encuentran en magnitudes similares para ambos casos, por lo que se decide usar esta variable para el análisis de todas formas.

5.2. Distribución condicionada de la variable dependiente

Para esta sección se analizará como distribuye la variable “Reincidencia” respecto a las otras variables regresaras con la finalidad de interpretar a priori como se relacionan estas variables y poder contrastar con los posteriores análisis. Cabe señalar que para este análisis exploratorio se utilizarán los datos obtenidos posterior a la limpieza de datos.

En cuanto a la variable Reincidencia, para el set de datos a analizar, podemos decir que el 52,3% de los usuarios volvió a recibir una condena administrada por gendarmería, lo que equivale a ciento noventa y nueve personas. Mientras que el restante 47,7% no volvió a ser encarcelado, lo que corresponde a 178 personas.

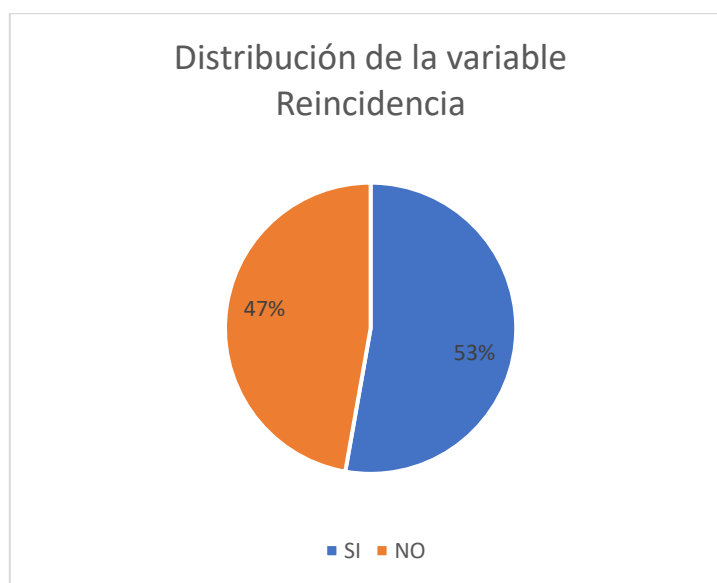


Gráfico 16 - Distribución de la variable dependiente (reincidencia) (elaboración propia)

A priori es posible observar que la tasa de reincidencia del programa es muy similar a la tasa de reincidencia a nivel nacional (52,9%), lo cual podría hacer inferir que el programa no consiguió ningún tipo de resultado significativo en lo que respecta a la reducción del delito, no obstante, esta tesis pretende realizar un análisis más profundo sobre las variables de proceso y demográficas que mejor responden ante este tipo de programas.

5.2.1. Reincidencia según edad

Para este análisis la variable edad fue segmentada en 5 conjuntos para poder calcular la tasa de reincidencia perteneciente a cada conjunto. Según la literatura, el resultado a esperar es que las tasas de reincidencia decaigan a medida que la edad aumenta.

Rango de edad	Total usuarios	Tasa de reincidencia
[18,30]	154	61.0%
(30,40]	127	49.6%
(40,50]	65	40.0%
(50,60]	25	56.0%
(60,68]	6	33.3%

Tabla 9 - Reincidencia según tramo etario (elaboración propia)

Es posible apreciar que a medida que la edad aumenta las tasas tienden a disminuir, salvo en el caso del rango entre 50 y 60 años, donde se produce un aumento, el cual puede estar relacionado con la escasa cantidad de usuarios. Pese a lo anterior se puede

afirmar que sobre los treinta años la tasa de reincidencia es menor a la tasa nacional, lo cual pretende inferir que esta variable es un factor relevante ante el desistimiento del delito y a la hora de diseñar programas de intervención.

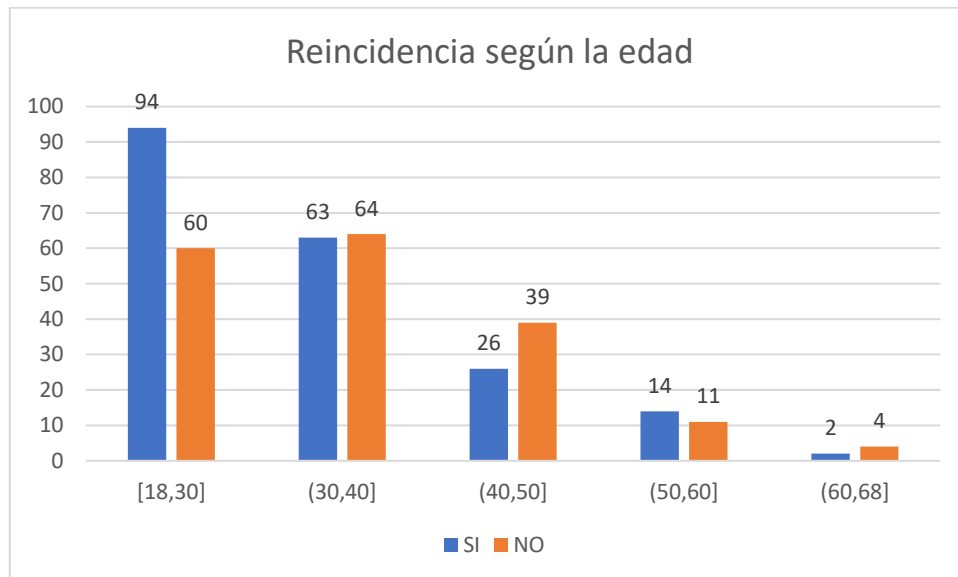


Gráfico 17 - Reincidencia según tramo etario (elaboración propia)

El gráfico nos permite observar las cantidades asociadas a cada rango de edad en donde se puede ver la diferencia aplastante en el rango entre 18 y 30 años con respecto al resto, en donde la tasa se eleva al punto en que es mucho más probable que un joven vuelva a ser encarcelado a que logre reinserirse. También podemos apreciar que el valor obtenido para el rango entre 50 y 60 es explicado por la baja cantidad de datos existentes, ya que la diferencia real es solo de dos usuarios.

5.2.2. Reincidencia según tamaño del grupo de referencia

Como se mencionó, la variable grupo de referencia define el tamaño declarado por los usuarios de su grupo de referencia o sus redes de apoyo más cercanas. A priori se esperaría que mientras mayor sea el grupo de apoyo las posibilidades de reincidencia disminuyan, lamentablemente los registros entregados por el programa no permiten generar ningún tipo de medida que indique la calidad de estas redes de apoyos, ya que sabemos gracias a la literatura que los círculos cercanos pueden ser factores claves en el desistimiento del delito, como también pueden ser factores de riesgo frente a la reincidencia.

A continuación, se presenta la distribución de la tasa de reincidencia sobre la variable de grupo de referencia.

Tamaño grupo de referencia	Total usuarios	Tasa de reincidencia
0	155	52.3%
1	123	48.8%

2	45	60.0%
3	26	65.4%
4	18	50.0%
5	5	80.0%
6	3	33.3%
7	2	0.0%

Tabla 10 - Reincidencia según tamaño del grupo de referencia (elaboración propia)

Para efectos de este análisis preliminar se ha decidido no contemplar las tasas relacionadas a los usuarios que declararon grupos de cinco personas o mayores, dado el bajo número de usuarios y el peso estadístico que cobran los casos particulares. Dicho lo anterior es que no se puede observar una tendencia clara en cuanto a la tasa de reincidencia, obteniéndose incluso resultados en donde con grupos de referencia más pequeños se obtienen menores tasas de reincidencia, lo a priori puede parecer contra intuitivo.

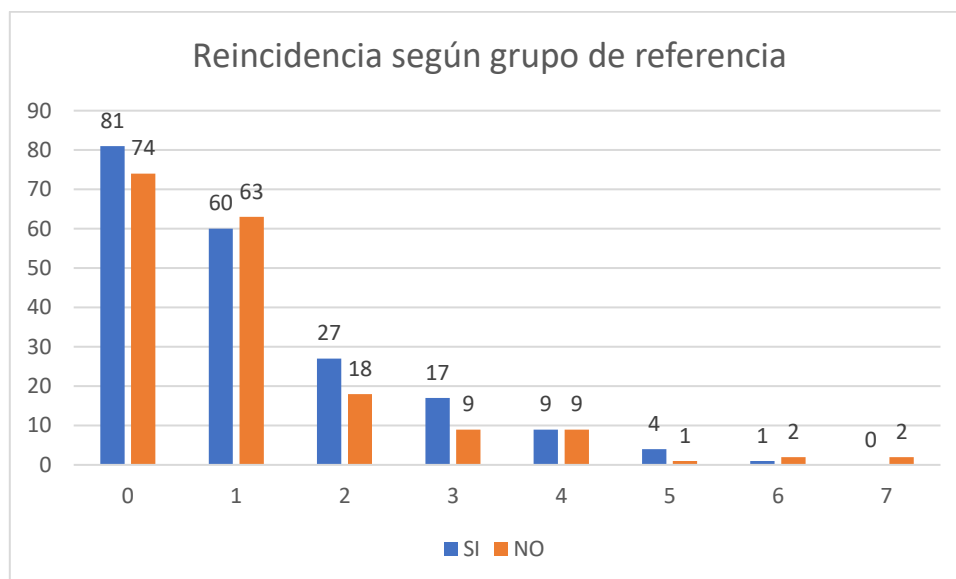


Gráfico 18 - Reincidencia según tamaño del grupo de referencia (elaboración propia)

El gráfico permite apreciar más claramente la falta de una tendencia en cuanto a la proporción de los datos. Todo lo presentado genera dudas respecto a relevancia que pueda cobrar esta variable respecto al modelo, dudas que serán disipadas posteriormente en el análisis de los resultados.

5.2.3. Reincidencia según colocación laboral

En base a la literatura es sabido que el factor laboral resulta ser preponderante para la reinserción social, es por esto que es de esperar que quienes lograron ser colocados laboralmente durante el programa posean una tasa de reincidencia menor.

En la siguiente tabla se desglosan las tasas de reincidencia para los distintos registros de esta variable.

Colocado laboralmente si requiere	Total usuarios	Tasa de reincidencia
No requiere	167	56.3%
No colocado	184	51.1%
Colocado	26	42.3%

Tabla 11 - Reincidencia según variable colocado laboralmente (elaboración propia)

Si bien la tabla muestra resultados esperables respecto a las tasas de reincidencia, es debido ser cautos, ya que, si bien, la tasa de reincidencia es menor para quienes consiguieron empleo, estos usuarios representan una porción muy baja de la muestra total (menos del 10%).

5.2.4. Nivel educacional

Se espera que esta sea una variable relevante al abordar el desistimiento delictivo, dado que un mayor nivel de educación promueve la caída de barreras en temas laborales, de acceso a la justicia y mejora la autogestión del sujeto, por lo que se espera que, a mayor nivel educacional, la probabilidad de reincidencia disminuya.

La siguiente tabla presenta las tasas de reincidencia para los distintos niveles educacionales.

Nivel educacional	Total usuarios	Tasa de reincidencia
Sin educación	8	37.5%
Básica incompleta	80	52.5%
Básica completa	74	58.1%
Media incompleta	76	63.2%
Media completa	115	45.2%
Técnica incompleta	7	42.9%
Técnica completa	8	62.5%
Superior incompleta	8	37.5%
Superior completa	1	0.0%

Tabla 12 - Reincidencia según nivel educacional (elaboración propia)

En primer lugar, es necesario separar del análisis las variables que poseen pocos registros, dado a la variabilidad estadísticas que estas pueden arrastrar. Por lo tanto, solo se consideran las variables “Básica incompleta”, “Básica completa”, “Media incompleta” y “Media completa”. Luego de lo anterior, se destaca que los registros de “Media completa” generan un quiebre en la tendencia de la tasa de reincidencia, este factor puede ser relevante de considerar al momento de analizar los modelos.

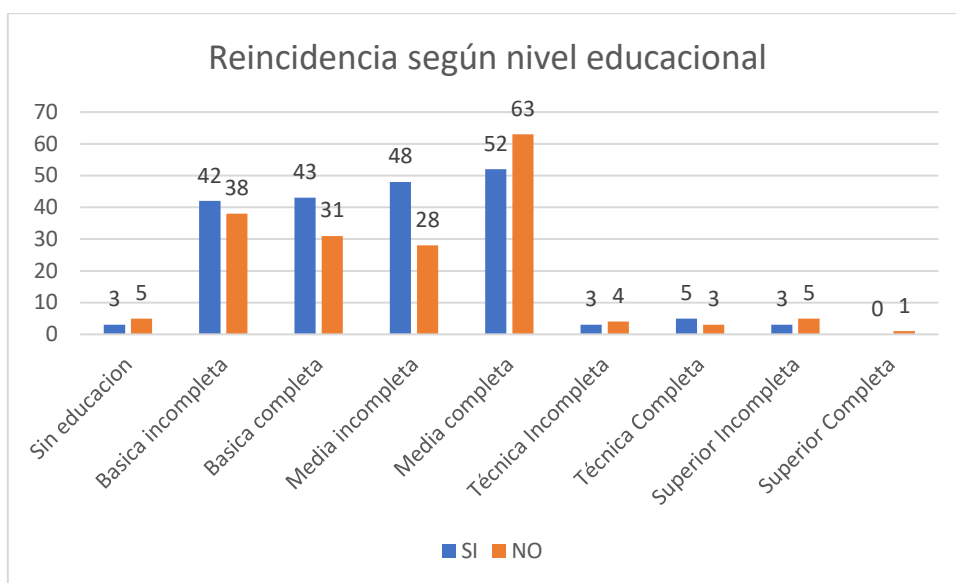


Gráfico 19 - Reincidencia según nivel educacional (elaboración propia)

El gráfico permite ilustrar las diferencias netas que se producen para los distintos valores, ya que, si bien en la tabla se pueden apreciar las diferencias para las tasas, al ver el gráfico, se puede distinguir que entre “Media incompleta” y “Media completa” existe un aumento de cuatro usuarios reincidentes, mientras hay treinta y cinco usuarios más no reincidentes, lo que es más del doble. Se espera que esta sea una variable que cobre relevancia dentro de los análisis de los modelos.

5.2.5. Reincidencia según comuna

El caso de esta variable es particular, ya que posee un componente demográfico y también de rendimiento del programa, ya que por un lado representa las variables situacionales a las que el usuario accede en su comuna, mientras que también posee un componente de la calidad de la gestión de los profesionales del programa de su comuna.

A continuación, se presenta la tabla de las tasas de reincidencia según la distribución de la variable.

Comuna	Total usuarios	Tasa de reincidencia
El Bosque	49	46.9%
La Pintana	35	48.6%
Maipú	43	58.1%
Peñalolén	37	64.9%
Pudahuel	29	55.2%
Puente Alto	37	37.8%
Quilicura	40	55.0%
Renca	49	46.9%
San Bernardo	27	55.6%

Santiago	31	64.5%
----------	----	-------

Tabla 13 - Reincidencia según comuna (elaboración propia)

De esta tabla cabe destacar el rendimiento de Puente Alto, como la comuna con la menor tasa de reincidencia (37,8%), junto con el de Santiago, quienes además de ser la comuna con la peor gestión de datos, es la comuna con la segunda más alta tasa de reincidencia (64,5%) después de Peñalolén (64,9%). También cabe mencionar a El Bosque y Renca, comunas que, habiendo tenido una buena gestión de datos, además poseen una tasa de reincidencia (46,9%) menor al promedio nacional.

El gráfico a continuación muestra los valores netos de reincidencia y reinserción por comuna para ilustrar de manera gráfica lo analizado.

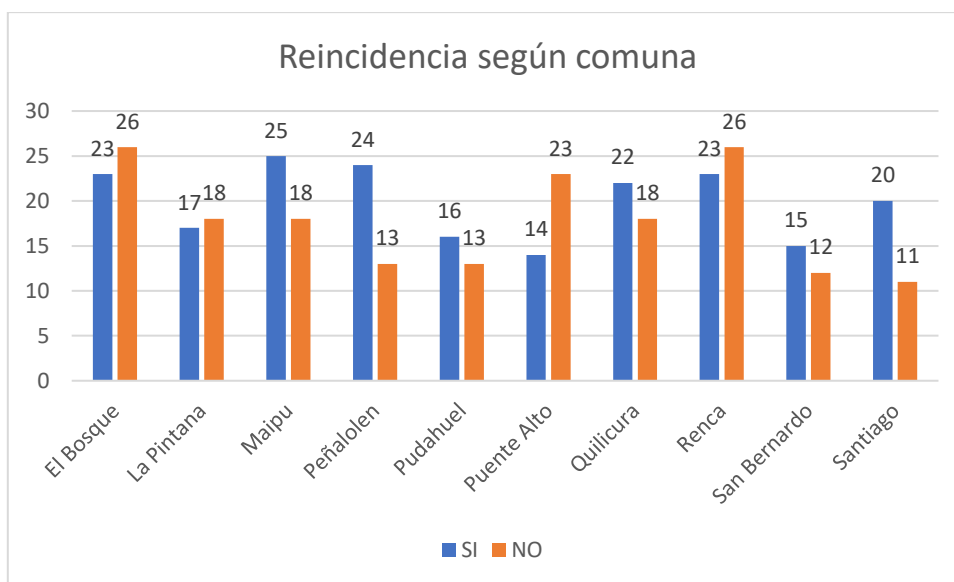


Gráfico 20 - - Reincidencia según comuna (elaboración propia)

Al ser analizadas estas variables como dummies, posiblemente las variables relacionadas a Puente Alto, Peñalolén y Santiago posean una mayor importancia para los análisis futuros.

5.2.6. Reincidencia según género

El siguiente gráfico ilustra la distribución de la variable reincidencia según género.

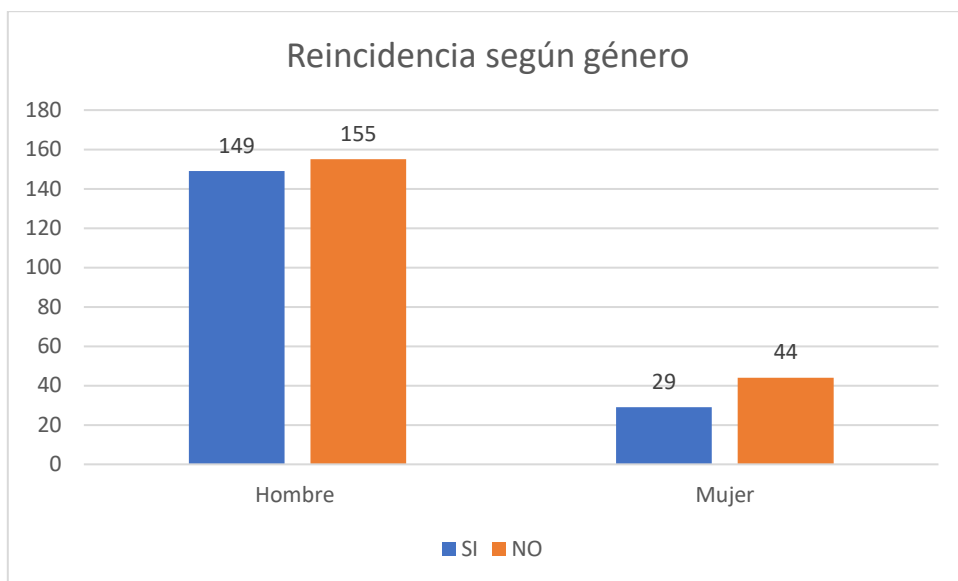


Gráfico 21 - Reincidencia según género (elaboración propia)

Porcentualmente la reincidencia en hombres alcanza un 51,0%, mientras que en mujeres un 60,3%, si bien la diferencia proporcional entre géneros es enorme, existe una mayor tendencia a la reincidencia por parte de las mujeres, lo cual puede cobrar cierta relevancia para los análisis posteriores.

5.2.7. Reincidencia según variable Profesión

El gráfico a continuación muestra la distribución de la reincidencia según si los usuarios declararon previamente haber tenido una profesión.

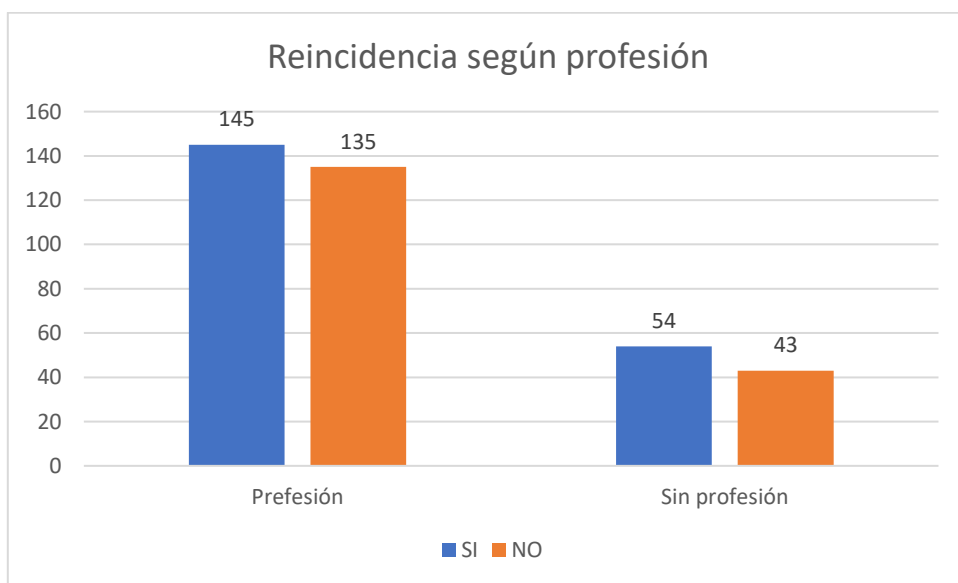


Gráfico 22 - Reincidencia según profesión (elaboración propia)

Porcentualmente la tasa de reincidencia para las personas que declararon haber tenido una profesión es de 51.8%, mientras que para quienes declararon no haber tenido una, es de 55,7%. Dado que las diferencias porcentuales son bajas y ambas se encuentran cercanas a la reincidencia a nivel nacional, además de las grandes diferencias proporcionales en sus dos alternativas, no permiten generar una inferencia clara del efecto de tener una profesión previamente, sobre el desistimiento delictivo.

5.2.8. Reincidencia según cumplimiento del programa de reinserción social

A continuación, se presenta las tasas de reincidencia según la distribución de la variable “Cumplimiento de plan de reinserción”.

Cumplimiento	Total usuarios	Tasa de reincidencia
SI	54	48.1%
NO	323	53.6%

Tabla 14 - Reincidencia según cumplimiento del programa de reinserción (elaboración propia)

Al igual que con la variable anteriormente analizada, se observan diferencias mínimas en las tasas de reincidencia, lo que sumado a la diferencia proporcional entre ambas alternativas, impide generar alguna inferencia clara respecto a la importancia de esta variable.

6. Resultados

6.1. Regresión logística multivariada

Para realizar la regresión logística con el set de datos ya tratados se diseñó un script de código en RStudio, donde se utilizó cuatro paqueterías diferentes y se realizaron los siguientes pasos:

1. En primer lugar, se importó el paquete “mise” el que se utilizó para limpiar el espacio de trabajo con el fin de evitar errores relacionados a variables y funciones previamente declaradas.
2. Se importó la librería “readxl”, la que permitió importar la matriz de variables explicativas “X” y la variable dependiente “y”.
3. Posteriormente con la función `as.factor` se convirtieron todas las variables en factores, exceptuando la Edad, Grupo de referencia y el Nivel educacional. Para R una variable de tipo factor, es una variable categórica.
4. Se combinó la data de las variables en un solo dataframe, para facilitar la ejecución del código.
5. Se importó la librería “caret”, librería que permite a través de múltiples funciones, compilar distintos modelos de regresiones, en particular, se utilizó la función `glm()` para compilar el modelo.
6. Se importó la librería “rms” que permite añadir al modelo ciertos estimadores estadísticos para medir la calidad de este.

A continuación, se presenta el código usado para la implementación del modelo de regresión logística.

```
library(mise)
mise()

library(readxl)
X = read_excel("C:/Users/matia/Desktop/Mati/Beauchef/2021/Otoño/Tesis 1/R/data2.xlsx",
              sheet = "X")

y = read_excel("C:/Users/matia/Desktop/Mati/Beauchef/2021/Otoño/Tesis 1/R/data2.xlsx",
              sheet = "Y")

X[-1] = apply(X[-1], 2, function(x){as.factor(x)})
X$`Grupo de referencia` = as.numeric(X$`Grupo de referencia`)
X$`Nivel educacional numerica` = as.numeric(X$`Nivel educacional numerica`)

library(caret)

BBDD = cbind(X, y)
BBDD = as.data.frame(BBDD)
colnames(BBDD)[17] = "Y"
model = glm(Y ~., data = BBDD, family = binomial)

summary(model)

library(rms)
```

```
colnames(BBDD) = gsub(" ", "_", colnames(BBDD))
model_sum = lrm(formula = as.formula(paste("Y ~", paste(names(BBDD[,-c(ncol(BBDD))]), collapse='+')), collapse = "")), data = BBDD)

print(model_sum)
```

Código 1 - Regresión logística

En seguida se presentan los resultados para los coeficientes del modelo y sus significancias:

Coeficientes	Estimador	Error estándar	z valor	Pr(> z)	Significancia
(Intercepto)	2.26459	0.63529	3.565	0.000364	***
Edad	-0.04415	0.01207	-3.657	0.000255	***
`Grupo de referencia`	0.02045	0.08616	0.237	0.812399	
`Colocado laboralmente si requiere`0	-0.28808	0.24191	-1.191	0.233718	
`Colocado laboralmente si requiere`1	-0.57422	0.52559	-1.093	0.274602	
`Nivel educacional numerica`	-0.05323	0.07544	-0.706	0.480427	
Genero_Masculino1	-0.52965	0.29379	-1.803	0.071419	.
Profesion_Si1	-0.0348	0.26889	-0.129	0.897025	
`Cumplimiento del Plan de Reinserción_Si`1	0.1308	0.37444	0.349	0.726855	
Comuna_La_Pintana1	0.01798	0.49143	0.037	0.970815	
Comuna_Maipu1	0.30411	0.47126	0.645	0.518726	
Comuna_Peñalolen1	0.84499	0.51281	1.648	0.099402	.
Comuna_Pudahuel1	0.32051	0.52076	0.615	0.538244	
Comuna_Puente_Alto1	-0.51560	0.49931	-1.033	0.301780	
Comuna_Renca1	-0.23115	0.45108	-0.512	0.608339	
Comuna_San_Bernardo1	0.03930	0.52971	0.074	0.940851	
Comuna_Santiago1	0.95331	0.55449	1.719	0.085569	.
Comuna_El_Bosque1	-0.15609	0.45629	-0.342	0.732287	

Tabla 15 - Resultados regresión logística (elaboración propia)

Código de significancias:

‘***’: variable significativa al 99,9% de confianza

‘**’: variable significativa al 99% de confianza

‘*’: variable significativa al 95% de confianza

‘.’: variable significativa al 90% de confianza

‘ ’: variable no significativa

Junto con lo anterior se calculó el R^2 y el estadístico para el Model Likelihood Ratio Test (MLRT)

R^2	0.103
-------	-------

MLRT P(>chi ²)	0.0253
----------------------------	--------

Tabla 16 - Estadísticos Regresión logística

6.2. Importancias con Random Forest y XGBoost

Para este análisis se utilizó el mismo set de datos y se diseñó un nuevo script en RStudio, el que incorporó el uso de cuatro librerías dentro de los siguientes pasos:

1. En primer lugar, se importó el paquete “mise” el que se utilizó para limpiar el espacio de trabajo con el fin de evitar errores relacionados a variables y funciones previamente declaradas.
2. Se importó la librería “readxl”, la que permitió importar la matriz de variables explicativas “X” y la variable dependiente “y”.
3. Posteriormente con la función as.factor se convirtieron todas las variables en factores, exceptuando la Edad, Grupo de referencia y el Nivel educacional. Para R una variable de tipo factor, es una variable categórica.
4. Se combinó la data de las variables en un solo dataframe, para facilitar la ejecución del código.
5. Se define y compila el modelo de Random Forest obteniéndose como resultado las listas de importancias para dos funciones de pérdida, Error Cuadrático Medio (MSE) y Pureza del Nudo (NodePurity), las cuales se presentan a continuación:

Variable	MSE
Edad	0.02399
Colocado_laboralmente_si_requiere	0.00866
Comuna_El_Bosque	0.00604
Nivel_educacional_numerica	0.00576
Genero_Masculino	0.00541
Grupo_de_referencia	0.00484
Comuna_Puente_Alto	0.00474
Comuna_Renca	0.00425
Comuna_Pudahuel	0.00329
Comuna_Santiago	0.00263
Comuna_Peñalolen	0.00198
Profesion_Si	0.00097
Comuna_La_Pintana	0.00066
Comuna_San_Bernardo	0.00061
Comuna_Maipu	-0.00034
Cumplimiento_del_Plan_de_Reinserción_Si	-0.00150

Tabla 17 - Importancias RF con función de pérdidas MSE (elaboración propia)

Variable	NodePurity
Edad	20.686
Nivel_educacional_numerica	9.614

Grupo_de_referencia	7.770
Colocado_laboralmente_si_requiere	5.143
Genero_Masculino	2.790
Profesion_Si	2.702
Comuna_Renca	2.439
Comuna_Puente_Alto	2.372
Comuna_El_Bosque	2.317
Comuna_Maipu	2.204
Comuna_Santiago	2.185
Comuna_Pudahuel	2.002
Comuna_Peñalolen	1.936
Cumplimiento_del_Plan_de_Reinserción_Si	1.842
Comuna_San_Bernardo	1.663
Comuna_La_Pintana	1.566

Tabla 18 - Importancias RF con función de pérdidas NodePurity (elaboración propia)

6. Posteriormente se define y compila el modelo de XGBoost, obteniéndose la lista de importancias que se presenta a continuación:

Variable	Importance_XGB
Edad	0.37703
Nivel_educacional_numerica	0.10677
Grupo_de_referencia	0.08555
Colocado_laboralmente_si_requiere	0.07028
Comuna_Puente_Alto	0.05297
Profesion_Si	0.04207
Comuna_El_Bosque	0.03713
Comuna_La_Pintana	0.03567
Genero_Masculino	0.03444
Comuna_Renca	0.03311
Comuna_Peñalolen	0.02724
Comuna_Pudahuel	0.02567
Cumplimiento_del_Plan_de_Reinserción_Si	0.02337
Comuna_Santiago	0.02296
Comuna_Maipu	0.02037
Comuna_San_Bernardo	0.00537

Tabla 19 - Importancias XGB (elaboración propia)

7. Entendiendo que los resultados se encuentran en diferentes escalas y se pretende obtener una escala de importancias final, es que se decide normalizar los resultados, ordenando cada tabla en una escala de cero a uno en donde 1 es el valor más alto y cero el más bajo, mientras que todos los otros valores se ordenan

de manera proporcional. Para lo anterior en este paso se define una función que normaliza cada una de las listas.

8. Se normalizan los valores obtenidos para Random Forest y posteriormente se suman, para volver a normalizar el resultado de la suma.
9. Se normaliza el resultado de XGBoost.
10. Se suman los resultados de Random Forest con los de XGBoost y se normaliza la suma, para finalmente obtener la importancia final. A continuación, se presentan los resultados de la importancia final y las importancias de cada método, ordenados por la importancia final.

Variable	importancia_f inal	importancia_ RF	importancia_ XGB
Edad	1.00000	1.00000	1.00000
Nivel_educacional_numerica	0.29751	0.35457	0.27284
Grupo_de_referencia	0.23446	0.25206	0.21574
Colocado_laboralmente_si_requiere	0.21667	0.35132	0.17466
Comuna_Puente_Alto	0.11631	0.14988	0.12807
Comuna_El_Bosque	0.10702	0.17060	0.08547
Genero_Masculino	0.10326	0.13890	0.07822
Comuna_Renca	0.08513	0.16362	0.07464
Profesion_Si	0.06787	0.10488	0.09874
Comuna_Pudahuel	0.05937	0.13858	0.05463
Comuna_Santiago	0.05146	0.11572	0.04734
Comuna_Peñalolen	0.04744	0.11078	0.05886
Comuna_La_Pintana	0.04078	0.02069	0.08153
Comuna_Maipu	0.01826	0.05504	0.04036
Cumplimiento_del_Plan_de_Reinserción_Si	0.00586	0.00000	0.04845
Comuna_San_Bernardo	0.00000	0.04679	0.00000

Tabla 20 - Importancias finales normalizadas (elaboración propia)

Para complementar los resultados presentados, se muestra a continuación el código utilizado:

```
library(mise)
library(readxl)
library(randomForest)

library(xgboost)

mise()

X = read_excel("C:/Users/matia/Desktop/Mati/Beauchef/2021/Otoño/Tesis 1/R/data2.xlsx",
              sheet = "X")

y = read_excel("C:/Users/matia/Desktop/Mati/Beauchef/2021/Otoño/Tesis 1/R/data2.xlsx",
              sheet = "Y")

X[-1] = apply(X[-1], 2, function(x){as.factor(x)})
```

```

X$`Grupo de referencia` = as.numeric(X$`Grupo de referencia`)

BBDD = cbind(X, y)
BBDD = as.data.frame(BBDD)

colnames(BBDD) = gsub(" ", "_", colnames(BBDD))

modelo_RF = randomForest(Reincidencia_Sí ~., data = BBDD, importance = TRUE)

varImpPlot(modelo_RF)

train_x = data.matrix(BBDD[, -17])
train_y = data.matrix(BBDD[, 17])
xgb_train = xgb.DMatrix(data = train_x, label = train_y)
modelo_XGB = xgb.train(data = xgb_train, max.depth = 3, nrounds = 100)
importance_matrix = xgb.importance(colnames(xgb_train), model = modelo_XGB)

importance_matrix

matriz_auxiliar = data.frame("variable" = rownames(modelo_RF$importance), "importancia_MSE" = m
odelo_RF$importance[,1], "importancia_NodePurity" = modelo_RF$importance[,2])

rownames(matriz_auxiliar) = NULL

normalizacion = function(x, na.rm = TRUE) {
  return((x- min(x)) / (max(x)-min(x)))
}

matriz_auxiliar$norm_MSE = normalizacion(matriz_auxiliar$importancia_MSE)
matriz_auxiliar$norm_NodePurity = normalizacion(matriz_auxiliar$importancia_NodePurity)
matriz_auxiliar$norm_general = matriz_auxiliar$norm_MSE + matriz_auxiliar$norm_NodePurity

matriz_importancias = data.frame("variable" = matriz_auxiliar$variable, "importancia_RF" = norm
alizacion(matriz_auxiliar$norm_general))
matriz_importancias = matriz_importancias[order(matriz_importancias$variable),]

matriz_auxiliar2 = data.frame("variable" = importance_matrix$Feature, "importancia_XGB" = norma
lizacion(importance_matrix$Gain))
matriz_auxiliar2 = matriz_auxiliar2[order(matriz_auxiliar2$variable), ]

matriz_auxiliar2$variable == matriz_importancias$variable[-length(matriz_importancias$variable)
]

matriz_importancias$importancia_XGB = matriz_auxiliar2$importancia_XGB

matriz_importancias$sum_importancias = matriz_importancias$importancia_RF + matriz_importancia
s$importancia_XGB
matriz_importancias$importancia_fianl = normalizacion(matriz_importancias$sum_importancias)

matriz_importancias = matriz_importancias[order(matriz_importancias$importancia_fianl, decreasi
ng = TRUE),]
matriz_importancias = matriz_importancias[,c(1,5)]

```

Código 2 - Random Forest y XGBoost

7. Análisis

7.1. Regresión logística

Este análisis comprende tres factores decisivos que son los estadísticos generales del modelo, en este caso el R^2 y el Model Likelihood Ratio Test, la significancia de las variables y los coeficientes del modelo.

En cuanto al Model Likelihood Ratio Test, este es una prueba sobre la bondad del ajuste del modelo, que posee como hipótesis nula la existencia de un modelo anidado que posee un ajuste igual o mejor. Se dice que un modelo B es anidado de un modelo A cuando el modelo B es construido a partir de un subconjunto de las variables que componen el modelo A y, además, B es distinto de A. El estadístico que se usa para construir este test es un χ^2 de un grado de libertad. En el caso del modelo estudiado, la probabilidad con la que se acepta la hipótesis nula es de 0,025, lo que nos asegura que se puede rechazar esta hipótesis a un 95% de confianza, lo cual implica que, a partir del set de variables utilizados, no se puede construir un modelo con mejor ajuste a este nivel de confianza.

En cuanto al R^2 , este es un estadístico que nos permite estimar que porción de la varianza de una variable es posible explicar a partir de un modelo. En el caso del modelo estudiado, el R^2 es de 0,103, lo cual indica que este modelo con estas variables logra explicar alrededor de un 10% de la varianza de la reincidencia. En términos generales este no es un buen resultado, ya que, para estudios de este tipo, valores desde 0,40 pueden ser considerados buenos resultados. En síntesis, este modelo permite explicar con poca precisión cuales son las variables fundamentales y cuales la probabilidad de que una persona reincida en el delito.

En cuanto a significancia los resultados son lamentables. Solo una variable posee significancia al 99,9%, la cual fue la variable edad, mientras que otras tres variables fueron significativas al 90%, ellas son la variable “Género”, “Comuna de Peñalolén” y “Comuna de Santiago”.

En cuanto a los coeficientes de las variables significativas podemos decir que:

1. El intercepto posee un valor alto (2,26), otorgando una mayor probabilidad basal de reincidencia, es decir que, pese a que un usuario posea atributos a favor de la reinserción, sigue existiendo una probabilidad significativa de que reincida, esto puede deberse a la falta de variables que expliquen el comportamiento de la variable.
2. Si bien la Edad al tener un coeficiente negativo de -0,04 posee un efecto reductor ante la posibilidad de reincidencia, lo cual es respaldado por la literatura y por análisis exploratorio realizado, este efecto es bajo, ya que en módulo el coeficiente es menor en comparación a otros.
3. La variable Género, si bien no tiene el más alto nivel de confianza, indica tener un impacto reductor en la probabilidad de reincidencia para los varones.
4. En cuanto a las variables de comunas, se mencionó durante el análisis exploratorio la posibilidad de que estas variables cobraran relevancia al ser Santiago y Peñalolén las comunas con los mayores índices de reincidencia marcadamente.

Sus coeficientes son de los más altos y al ser positivos nos indican que los usuarios de esas comunas poseen una mayor probabilidad de reincidencia.

A continuación, se presenta en la siguiente tabla las probabilidades asociadas a los coeficientes, obtenidas a partir de la transformación logística mencionada en la metodología:

Variable	Coefficiente	Probabilidad
(Intercepto)	2.26459	90.6%
Edad	-0.04415	48.9%
Genero_Masculino1	-0.52965	37.1%
Comuna_Peñalolen1	0.84499	70.0%
Comuna_Santiago1	0.95331	72.2%

Tabla 21 - Resumen variables significativas (elaboración propia)

7.2. Análisis de importancias con Random Forest y XGBoost

Estos análisis buscan determinar qué tan importante es cada variable a la hora de predecir la reincidencia, no así determinar de qué forma afecta cada variable, si positiva o negativamente, esto último debe ser complementado con el análisis exploratorio, la regresión logística o, en el caso de las variables no significativas, con cualquier otro análisis de tendencia más profundo que se desee realizar.

A continuación, el análisis será segmentado en cuatro partes según el nivel de importancia de las variables determinado a través de los modelos utilizados.

7.2.1. La variable más importante

Los resultados arrojaron que siempre, para cada uno de los algoritmos y para cada una de las funciones de pérdida, y además por gran diferencia, la variable más importante es la edad de los usuarios.

Laub y Sampson (2001), explican como el ámbito de los orígenes de las carreras delictivas y la reincidencia se encuentra mucho más profundizado que los estudios de desistimiento del delito y que se desconoce cómo se formulan las mecánicas sociales que posibilitan a las personas poner término a sus carreras delictivas.

Los autores, sin pesar de lo anteriormente mencionado, afirman que una de las principales variables asociadas al desistimiento es la edad, lo que quiere decir que a mayor edad es menor la probabilidad de reincidencia, pese a esto se admite que la preponderancia otorgada por la variable edad reside en su vinculación con variables “ocultas” como la creación o recomposición de vínculos familiares, la madurez del usuario, la eliminación de vínculos delictivos, la pérdida de habilidades delictivas o la mejoría en la autopercepción, entre otras.

Bajo la perspectiva de Laub y Sampson, los resultados obtenidos respecto a la importancia de la variable edad son nada más que un indicio del poco conocimiento e información que se posee sobre la muestra. Lo anterior apunta en dos sentidos, por un

lado, es necesario articular programas que permitan conocer e interpretar de mejor manera la situación personal de los usuarios sobre todo en lo que respecta a variables dinámicas que sean sensibles al paso del tiempo, por otra parte, si bien la edad es una variable que incorpora el efecto de otras variables relevantes, esto no quiere decir que debe ser despreciada la relevancia de la edad sobre la aplicación de programas de reinserción, por el contrario, esta se presenta como una variable clave para el diseño de estos.

7.2.2. Variables de alta importancia

De cara a los resultados obtenidos, luego de analizar la importancia de la variable edad en el desistimiento delictivo, es posible distinguir un siguiente grupo de variables que presentan una relevancia significativamente mayor al resto. Las variables mencionadas son Nivel educacional, Grupo de referencia y Colocación laboral.

En el estudio publicado por la Fundación Paternitas en 2021, en donde se estudia y caracteriza a los usuarios del Programa +R (programa piloto a nivel nacional que aborda la reinserción desde el punto de vista del acceso al trabajo) establece la relevancia del nivel educacional no solo sobre el desistimiento delictivo, sino que también sobre el desistimiento en el consumo de drogas, mostrando que quienes poseen mayor nivel educacional, también poseen mayor adherencia en los programas de drogas.

Si bien la literatura vincula el efecto del nivel educacional a distintas causas, es posible inferir que existen dos motivos principales, el primero habla de la influencia que esta variable tiene en el ámbito laboral debido a un nivel de estudios mínimo exigido para la inserción al mundo laboral, lo cual sucede en Chile, en donde generalmente se exige cuarto medio terminado para el acceso a cualquier trabajo. El otro motivo al que se asocia el efecto de esta variable es la autonomía de los usuarios a la hora de afrontar sus propios procesos de reinserción, esto tiene principalmente que ver con la autopercepción de sus capacidades para enfrentarse a realizar trámites personales, emprender nuevos proyectos y conseguir empleo.

Un aspecto interesante de los resultados es el contraste entre la relevancia del grupo de referencia y la inexistencia de una tendencia en el análisis exploratorio, a priori se podría aseverar que esta variable es relevante, pero para poseer una correcta interpretación sería necesario generar una variable que pueda indicar la calidad de las redes existentes dentro del grupo de referencia.

Finalmente, la última variable considerada de gran importancia es la variable de colocación laboral, lo cual es un resultado completamente esperable según la literatura donde típicamente se le atribuye un papel preponderante a lo laboral dentro de los procesos de reinserción, pero de la misma manera la literatura indica que no basta solamente con la colocación laboral, hay dos variables claves relacionadas a la colocación laboral, una es la mantención del puesto de trabajo por seis meses y la otra es la conformidad con el empleo. Por otra parte, Ana Alcázar Belmonte (2022), en su estudio *“Reencalcelamiento, reincidencia y empleo”* analiza cómo se relacionan las variables macroeconómicas del desempleo y el delito llegando a la conclusión de que no

existe evidencia que permita relacionarlas, lo que implica que no existiría una relación bidireccional de estas variables.

7.2.3. Variables de mediana importancia

En este grupo se encuentran dos variables de comuna y la variable género, las cuales, si bien poseen una importancia menor en comparación a la edad, no es posible descartar a priori su relevancia. En cuanto a la variable de género, es posible realizar dos consideraciones. Por un lado, el análisis exploratorio mostró una clara diferencia entre las tasas de reincidencia para hombres y para mujeres. Por otra parte, en la muestra estudiada la cantidad de mujeres es demasiado baja como para ser realmente representativa, por lo que no se considera completamente confiables las conclusiones que se puedan obtener respecto a esta variable.

En cuanto a las variables de comuna, estas son las relacionadas a la comuna de Puente Alto y El Bosque que a su vez son dos de las tres comunas con menor índice de reincidencia. Este resultado es verdaderamente relevante ya que, lo esperable sería que no existieran diferencias en las comunas de los programas dado al criterio socioeconómico con que estas fueron seleccionadas. Los resultados posiblemente apuntan a dos razones, una son las propias características de las comunas, razón que pierde fuerza al considerar los criterios aplicados en la selección y que las comunas poseen índices de delincuencia similares. La otra razón alude a la propia gestión del programa, a la prolijidad en la recolección de datos y a la rigurosidad en la aplicación de las actividades del programa.

7.2.4. Variables de menor importancia

En este grupo se encuentra el resto de las variables relacionadas a comunas, lo cual es esperable, mientras que también se encuentra la variable asociada a la tenencia previa de una profesión u oficio, y la variable vinculada al cumplimiento del plan de reinserción.

Con respecto a la tenencia previa de una profesión u oficio se infiere que la baja relevancia de esta puede estar vinculada a las barreras comunes que poseen todos los usuarios con respecto al acceso al trabajo, una de las barreras mas relevantes que hoy se presentan es el papel de antecedentes para fines laborales, el cual es pedido comúnmente en los empleos de baja calificación a los que postulan principalmente los usuarios del programa dado a sus bajos niveles de estudios.

Finalmente, es alarmante señalar la importancia que se le otorga al cumplimiento del plan de reinserción que para algunos de los análisis resulta ser la variable menos importante. Esto puede ser explicado de dos maneras, por un lado, una muy mala calidad en el registro de la variable y en la gestión de datos en general, por otro lado, mas preocupante, existe la posibilidad de que la aplicación del programa posea una bajísima efectividad y que ciertamente no haya obtenido ningún resultado relevante.

8. Conclusiones

Como se ha definido en este trabajo, la reinserción social es *“Un proceso que comienza durante el cumplimiento de una sanción, pero que se extiende más allá del término de una condena y que consiste en la modificación de múltiples comportamientos y factores que inciden en la conducta delictiva, con el fin de mejorar la seguridad pública disminuyendo la probabilidad de reincidencia de quienes han cumplido condena”*.

Al ser la delincuencia un problema público es que el estado tiene la obligación de dar respuesta a este fenómeno, por un lado, el estado lo hace a través de las fuerzas de orden público, invirtiendo en seguridad, represión e investigación de los delitos. También se invierte en medidas preventivas vinculadas a los principales factores de riesgo, como programas extracurriculares para jóvenes, programas para la prevención del consumo de drogas, programas de empleo para jóvenes, entre otros. Finalmente, un tipo de intervención estatal respecto de la delincuencia tiene que ver con la reinserción social, lo cual se realiza principalmente mediante programas que en su gran mayoría son impartidos en el interior de la cárcel, estos programas pueden tener enfoques en educación, apresto laboral y desarrollo de oficios, colocación laboral, apoyo psicológico, tratamiento de drogas, intervenciones familiares, entre otros.

Generalmente los esfuerzos estatales en intervenciones pro-reinserción social han sido llevados a cabo por gendarmería mediante programas intrapenitenciarios, con una marcada debilidad en la gestión del proceso postpenitenciario, en donde sus intervenciones se han visto limitadas a los Centros de Estudio y Trabajo (CET) en donde Gendarmería ofrece cupos limitados. Durante los últimos años los gobiernos han implementado iniciativas de programas de reinserción implementadas en el medio libre con distintos enfoques. Recientemente se implementó el programa +R con un enfoque laboral, el cual tenía como principal objetivo gestionar la inserción laboral de usuarios del sistema postpenitenciario, administrado por las delegaciones regionales.

Anterior al programa +R se implementó el programa Volver a Empezar (VAE) en donde el enfoque fue distinto, se utilizó un abordaje psicosocial con foco en la gestión territorial, en donde se plantearon objetivos multidimensionales entre los que destacan la formalización de una metodología de trabajo, la implantación de un sistema de gestión de casos en los municipios y la atención de usuarios. En relación con lo señalado en la evaluación del programa realizada en 2019, es verdaderamente preocupante la dimensión política en la implementación de este programa dado que existían distintos niveles de compromiso por parte de las distintas autoridades comunales, ministeriales y de Gendarmería. Otro aspecto débil en la evaluación fueron temas administrativos, como la rotación de las duplas psicosociales, el no pago a los trabajadores del programa, problemas de gestión con otros departamentos municipales y la baja adherencia lograda por los usuarios del programa.

Pese a los análisis entregados por la evaluación del programa, el componente de gestión de casos presentó dificultades para ser evaluado cuantitativamente, dadas las fechas de solicitud de la evaluación en donde el tiempo transcurrido no permitía realizar análisis de reincidencia que entregasen resultados de calidad. Dado lo anterior es que los análisis

presentados en este documento cobran especial importancia para la evaluación del proyecto, en consecuencia con lo anterior hay dos resultados muy destacables, el primero es la relevancia de la edad como una variable preponderante en el análisis de resultados de este tipo de programas, mientras que la baja relevancia del cumplimiento del plan de reinserción entrega un indicio de la eficacia de la aplicación del programa respecto de la reducción de la reincidencia.

Durante la ejecución de esta tesis, la obtención y el tratamiento de datos fueron los principales desafíos, dadas las características de las organizaciones propietarias de las fuentes de datos (Gendarmería y el Ministerio Público) y sus procesos burocráticos, dada la sensibilidad de la data como registros personales, dada la dispersión de criterios a la hora de realizar registros debido a los diversos equipos comunales y a la alta rotación de estos. En relación con los factores de riesgo asociados a la medición de reincidencia mencionados previamente en el marco teórico, se puede reconocer la presencia de cuatro de los cinco factores descritos, estos factores son la complejidad del proceso de recolección de datos, la complejidad en el procesamiento y el análisis de datos, el nivel de subjetividad y la experticia de los operadores del sistema.

El presente trabajo a permitido distinguir las variables más importantes para el éxito del programa Volver a Empezar o para programas que posean características similares. Es así que se pudo distinguir cuatro variables relevantes a la hora de considerar un rediseño, estas variables son la edad de los usuarios, el tamaño de los grupos de apoyo, el nivel educacional de los usuarios y su situación laboral.

Dicho lo anterior es que se requiere que el programa atienda de manera distinta a los diferentes grupos etarios, considerando las diferentes probabilidades de éxito que poseen, se recomienda que en cualquier caso el tratamiento destinado a los usuarios de menor edad sea mucho más intensivo para garantizar mayores tasas de adherencia y mejores resultados en cuanto al desistimiento delictivo. Junto con lo anterior se recomienda realizar un exhaustivo monitorio de las variables dinámicas que se asocian a la edad para conseguir una caracterización más completa de los usuarios.

El programa VAE pone sus enfoques metodológicos en cuatro dimensiones, las cuales son lo social-territorial, lo jurídico, lo familiar y lo laboral. Para el caso de las variables relevantes es que se puede señalar que la colocación laboral y el nivel educacional, están fuertemente ligado con la dimensión laboral, mientras que el tamaño del grupo de referencia se encuentra vinculado tanto a la dimensión familiar, como también a lo social-territorial, sin desmedro de lo anterior es que se señala que los datos estudiados no permiten realizar aseveraciones confiables de la relevancia de los aspectos jurídicos en la intervención del programa.

En cuanto al tamaño del grupo de referencia es que se sugiere complementar esta medida y nutrir el análisis, estudiando variables relacionadas que permitan medir la calidad de los vínculos presentes en las redes de apoyo, estas variables pueden ser construidas como indicadores, pero es necesarios que permitan distinguir factores de riesgo como el comportamiento delictivo, patrones de violencia y de consumo o dinámicas prosociales.

Respecto a la colocación laboral y el nivel educacional es que se sugiere otorgar gran relevancia a la gestión de redes laborales, que permitan a los usuarios acceder de manera simple a un empleo, por otra parte es necesario complementar esta información con variables que permitan conocer el estado y evaluación de los usuarios con respecto al apresto laboral, habilidades laborales, actitud frente a la búsqueda de empleo, además de considerar un seguimiento extendido durante un año posterior a la obtención de un empleo. También es necesario fortalecer el programa en cuanto a opciones de continuidad de estudios o acceso a becas de estudios superiores.

Se establece a lo largo de esta tesis las debilidades presentes en el programa en cuanto a la gestión de datos y la falta de orientación al análisis y evaluación de este, lo cual se ve reflejado en los distintos programas municipales que persisten a la fecha. Más allá de la identificación de las variables claves, es necesario que los programas posean dicha orientación como parte de su diseño de gestión estratégica, de modo que les permita establecer métricas de rendimiento en cuanto a la medición de reincidencia dado que esta representa un indicador clave en la evaluación de desempeño del programa. Como fue presentado en el marco teórico, estos análisis pueden y deben ser tanto actuariales como clínicos, es decir, que aborden modelos probabilísticos que permitan percibir de manera numérica la relevancia de que poseen las variables para el diseño de los programas, como también herramientas que permitan incluir de manera efectiva la evaluación de variables dinámicas complejas de cuantificar.

Un factor de vital importancia para el análisis y evaluación es el aspecto de coordinación político-administrativo, el cual presentó evidentes dificultades en la implementación del programa. Por un lado, es crítico reducir el nivel de rotación de los profesionales que llevan a cabo el programa, porque su gestión en materia de coordinación es fundamental no tan solo en el desarrollo de los planes de reinserción de cada usuario, sino que también en el seguimiento de las derivaciones tanto dentro como fuera del municipio. Por otra parte, la obtención de datos de fuentes externas es vital para la evaluación del desempeño, por lo que se recomienda la gestión política de las autoridades para lograr acuerdos de colaboración en materia de acceso a datos de reincidencia y de registros de intervenciones previas por parte de Gendarmería como del Ministerio público. Estos acuerdos deberían permitir realizar evaluaciones con mayor periodicidad y generar análisis de mayor calidad por medio de distintas medidas de reincidencia (legal, procesal y/o penal) e incluso optar a la formulación de grupos de control que permitan realizar mejores análisis de “supervivencia”.

Es preciso indicar que en la implementación del programa VAE, tanto como de los programas municipales existentes se observa una interrupción del ciclo de políticas públicas en donde no existen procesos de evaluación y rediseño que puedan nutrir a los programas. Se sugiere considerar como crítica la implementación de evaluaciones que permitan conocer los rendimientos de los programas respecto a métricas de reincidencia, ya que como se pudo apreciar en el caso del VAE, podría tener efectos considerables en la implementación de los programas el llegar a conocer que exista una baja eficacia en estos, luego de años de implementación.

Finalmente, se valora de manera positiva la iniciativa, dado que permitió la generación de conocimiento en un ámbito muy poco investigado a nivel internacional y que previamente al año 2010 en nuestro país fue prácticamente inexistente. También se sugiere considerar las implementaciones en Puente Alto y El Bosque como casos de estudio particulares dado su nivel de éxito. Si bien este trabajo rebela resultados preocupantes, en cuanto a la efectividad del programa y al manejo de datos, se señala que estas situaciones tienen que ver con la poca madurez del campo, pero que a su vez tanto esta iniciativa como el programa +R han generado aportes significativos en materia de gestión del conocimiento, como también en cuanto a la sensibilización y a la promoción de otras iniciativas pertenecientes al ámbito.

9. Bibliografía

- Alto, P. R. (22 de Junio de 2021). Obtenido de <https://prezi.com/p/0pfoastdqs0c/programa-renaciendo/?frame=379ad964edd186757c9c54eec714533afe40a22b>
- Canadian Centre for Justice, Statistics Canada. (2005). *Criminal Justice Indicators*. Ottawa, Ontario. Obtenido de <https://www150.statcan.gc.ca/n1/en/pub/85-227-x/85-227-x2002000-eng.pdf?st=S7z7AxPO>
- Centro de Políticas Públicas UC. (2017). Sistema carcelario en Chile: propuestas para avanzar hacia una mayor efectividad y reinserción. *Temas de Agenda Pública*. Obtenido de <https://politicaspublicas.uc.cl/wp-content/uploads/2017/05/Art%C3%ADculo-Sistema-carcelario-en-Chile.pdf>
- Fundación Paz Ciudadana. (2010). *Construcción de Indicadores de Reinserción Social para Adolescentes Infractores de Ley Penal*. Santiago. Obtenido de <https://pazciudadana.cl/biblioteca/documentos/construccion-de-indicadores-de-reinsercion-social-de-adolescentes-infractores-de-la-ley-penal-informe-final/>
- Fundación Tierra Esperanza. (2018). *Manual de Procedimientos, Modelo de Gestión de Casos para la Reinserción*. Obtenido de <http://www.reinsercionsocial.gob.cl/media/2019/02/MANUAL-DE-PROCEDIMIENTO-VAE.pdf>
- Gendarmería de Chile. (2020). *Compendio Estadístico Penitenciario, 2019*. Obtenido de https://html.gendarmeria.gob.cl/doc/estadisticas/Compendio_Estadistico_Peniten-ciario_2019.pdf
- Instituto Nacional de Estadística. (2020). *Encuesta Nacional Urbana de Seguridad Ciudadana, 2019*. Obtenido de https://ine.cl/docs/default-source/seguridad-ciudadana/publicaciones-y-anuarios/2019/bolet%C3%ADn---xvi-enusc-2019.pdf?sfvrsn=6a96af3c_2
- Instituto Nacional de Estadísticas. (2021). *Informe Anual de Estadísticas Judiciales, 2019*. Obtenido de https://ine.cl/docs/default-source/justicia/publicaciones-y-anuarios/difusi%C3%B3n/informe-anual-estad%C3%ADsticas-judiciales-2019.pdf?sfvrsn=25a4f678_2
- Instituto Nacional de Estadísticas. (2021). *Informe Anual de Estadísticas Policiales, 2019*. Obtenido de https://ine.cl/docs/default-source/estadisticas-policiales/publicaciones-y-anuarios/estadisticas-policiales/2019-informeanual-estad%C3%ADsticas-policiales.pdf?sfvrsn=b82b340a_2
- Jarufer Bader, J. (2019). *Reconocimiento Constitucional a la Reinserción Social de Reos y Condenados, Experiencia Internacional*. Biblioteca del Congreso Nacional de Chile. Obtenido de <https://obtienearchivo.bcn.cl/obtienearchivo?id=repositorio/10221/27460/1/Recon>

ocimiento_constitucional_a_la_reinsercion_social_de_reos_y_condenados_Experiencia_internacional.pdf

- Ministerio de Justicia de Chile. (2020). *Decreto ley 2.859 | Fija Ley Orgánica de Gendarmaría de Chile*. Biblioteca del Congreso Nacional de Chile. Obtenido de <https://www.bcn.cl/leychile/navegar?idNorma=7015>
- Ministerio de Justicia y Derechos Humanos. (16 de Abril de 2021). *Reinserción Social, Construyamos Oportunidades*. Obtenido de <https://www.reinsercionsocial.gob.cl/que-es-la-reinsercion/>
- Ministerio de Justicia y Derechos Humanos. (22 de Junio de 2021). *Reinserción Social, Construyamos Oportunidades*. Obtenido de <https://www.reinsercionsocial.gob.cl/dupla-programa-volver-a-empezar-de-maipu-la-reinsercion-laboral-es-una-puerta-de-entrada-para-las-personas-que-son-usuarios-de-este-programa/>
- Ministry of Justice of United Kingdom. (2009). *National Reoffending Measures – a Guide, An Explanation of the Headline National Reoffending Measure*. Obtenido de https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/162859/national-reoffending-measures-guide-210509.pdf
- Ministry of Justice of United Kingdom. (2009). *OGRS 3: The Revised Offender Group*. Obtenido de <https://eprints.lancs.ac.uk/id/eprint/49988/1/ogrs3.pdf>
- Morales Peillard, A., Muñoz Correa , N., Welsch Chauán, G., & Fábrega Lacoa, J. (2012). *La Reincidencia en el Sistema Penitenciario Chileno*. Obtenido de <https://pazciudadana.cl/biblioteca/documentos/la-reincidencia-en-el-sistema-penitenciario-chileno/>
- Morales Peillard, A., Pantoja Vera, R., Piñol Arriagada, D., & Sanchez Cea, M. (2018). *Una Propuesta de Modelo Integral de Reinserción Social para Infractores de Ley*. Obtenido de https://www.cesc.uchile.cl/Modelo_ReinsercionSocial_CESC_FPC.pdf
- Morales Peillard, A., Welsch Chauán, G., Cárcamo Cáceres, J., Aguilar Chamorro, L., & Sosa , M. (2015). *Reinserción Social y Laboral de Infractores de Ley, Estudio Comparado de la Evidencia*. Santiago. Obtenido de [file:///C:/Users/matia/Downloads/estudio-comparado-de-la-evidencia-en-reinsercion-social-y-laboral%20\(1\).pdf](file:///C:/Users/matia/Downloads/estudio-comparado-de-la-evidencia-en-reinsercion-social-y-laboral%20(1).pdf)
- Municipalidad de Pudahuel. (22 de Junio de 2021). Obtenido de <https://slidetodoc.com/programa-de-reinsercin-social-caractersticas-usuariosas-de-pudahuel/>
- Novello Tapia, V. L. (2019). *Naturaleza Jurídica de la Reinserción Social en el Sistema Penitenciario de Adulto: ¿Un Derecho o un Beneficio?* Santiago. Obtenido de <http://repositorio.uchile.cl/bitstream/handle/2250/170567/Naturaleza-juridica-de->

la-reinsercion-social-en-el-sistema-penitenciario-de-adultos.pdf?sequence=1&isAllowed=y

- Payne, J. (2007). *Recidivism in Australia: Findings and Future Research*. Australian Institute of Criminology. Obtenido de https://www.researchgate.net/profile/Jason-Payne/publication/267703058_Recidivism_in_Australia_findings_and_future_research/links/550a66380cf26198a63aa9ae/Recidivism-in-Australia-findings-and-future-research.pdf
- Sandoval Verges, M. A. (2014). *Facilitadores y Obstaculizadores de las Comunidades de Egreso de Adultos y Adultas de un Programa de Reinserción de la Comuna de la Pintana para sus Procesos de Reinserción*. Obtenido de <http://repositorio.uchile.cl/bitstream/handle/2250/130139/Tesis%20MASV%202014.pdf?sequence=1&isAllowed=y>
- UNODC, Oficina de las Naciones Unidas Contra las Drogas y el Delito. (2011). *Reglas de las Naciones Unidas para el Tratamiento de las Reclusas y Medidas no Privativas de Libertad para las Mujeres Delincuentes (Reglas de Bangkok)*. Obtenido de https://www.unodc.org/documents/justice-and-prison-reform/Bangkok_Rules_ESP_24032015.pdf
- UNODC, Oficina de las Naciones Unidas Contra las Drogas y el Delito. (2016). *Reglas Mínimas de las Naciones Unidas para el Tratamiento de Reclusos (Reglas Mandela)*. Obtenido de https://www.unodc.org/documents/justice-and-prison-reform/Nelson_Mandela_Rules-S-ebook.pdf
- Vargas Subiabre, J. (2014). *LOS SIGNIFICADOS DE LA REINSERCIÓN, SEGÚN LA MIRADA DE QUIENES RETORNAN A LA LIBERTAD. EL CASO DEL PROGRAMA DE REINSERCIÓN LABORAL DEL CENTRO DE APOYO A LA INTEGRACIÓN SOCIAL DE SANTIAGO*. Santiago. Obtenido de <https://repositorio.uahurtado.cl/bitstream/handle/11242/7547/M-ISOCVargasJ.pdf?sequence=1&isAllowed=y>
- Villagra Pincheira, C. (2008). *Hacia una Política Postpenitenciaria en Chile*. Santiago: Ril Editores. Obtenido de <https://www.cesc.uchile.cl/9789562846530%20-%20Villagra,%20Carolina%20-%202009%20-%20Hacia%20una%20pol%C3%ADtica%20postpenitenciaria%20en%20Chile%20Bajar.pdf>
- Villagra Pincheira, C., Espinoza Mavila, O., & Martínez Mercado, F. (2014). *La Medición de la Reincidencia y sus Implicancias en la Política Penal*. Santiago. Obtenido de https://www.cesc.uchile.cl/Publicacion_CESC_web_creditos.pdf
- Ramakers, A., Nieuwebeerta, P., Van Wilsem, J. & Dirkzwager, A. (2017, noviembre). Not Just Any Job Will Do: A Study on Employment Characteristics and Recidivism Risks After Release. *journals.sagepub.com*. Recuperado 22 de diciembre de 2022, de <https://journals.sagepub.com/doi/pdf/10.1177/0306624X16636141>

- Alcázar, A. (2022). ENCARCELAMIENTO, REINCIDENCIA Y EMPLEO [Investigación]. Universidad Pontificia Comillas.
- Laub, J. H. & Sampson, R. J. (2019). Entender el desistimiento del delito [Libro Online]. En Seguridad Ciudadana. Lecturas fundamentales (p. 521). CAF.
<https://scioteca.caf.com/bitstream/handle/123456789/1433/Seguridad%20ciudadana-lecciones%20fundamentales.pdf?sequence=1&isAllowed=y#page=523>
- Bustos, C. & Croquevielle, C. (2018). Caracterización de la población privada de libertad beneficiaria del programa +R, implementada por la fundación Paternitas. [Informe]. PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE, FACULTAD DE CIENCIAS SOCIALES, INSTITUTO DE SOCIOLOGÍA.

Anexos

Anexo A: Listado de variables

Edad

genero

Profesion

Fecha de ingreso al programa

ultimo_anio_cursado_en_educacion

Comuna de residencia actual

tipo_de_beneficiario

Centro Reinsercion Social

estado condena

reincidente

Fecha Egreso del Sistema Penitenciario

Origen_demanda_espontanea

N_tutoria_2dilinea

N_horas_tutoria_2dilinea

última_fecha_tutoria

familia_o_grupo_ref_contactada

n_grupo_referencia

n_grupo_referencia_mayor

n_grupo_referencia_menor

ref_parentesco

aplicacion_diagnostico_area_familiar

aplicacion_diagnostico_area_familiar_referencias_consideradas

aplicacion_diagnostico_area_familiar_referencias_consideradas_total

fecha_diagnostico_en_area_familiar

aplicacion_diagnostico_area_familiar_completo

aplicacion_diagnostico_area_familiar_incompleto

aplicacion_diagnostico_area_social

aplicacion_diagnostico_area_social_referencias_consideradas
aplicacion_diagnostico_area_social_referencias_consideradas_total
aplicacion_diagnostico_area_social_completo
aplicacion_diagnostico_area_social_incompleto
fecha_diagnostico_en_area_social
aplicacion_diagnostico_area_laboral
aplicacion_diagnostico_area_laboral_referencias_consideradas
aplicacion_diagnostico_area_laboral_referencias_consideradas_total
fecha_diagnostico_en_area_laboral
aplicacion_diagnostico_area_laboral_completo
aplicacion_diagnostico_area_laboral_incompleto
aplicacion_diagnostico_area_judicial
fecha_diagnostico_en_area_judicial
aplicacion_diagnostico_area_judicial_referencias_consideradas
aplicacion_diagnostico_area_judicial_referencias_consideradas_total
aplicacion_diagnostico_area_judicial_completo
aplicacion_diagnostico_area_judicial_incompleto
aplicacion_diagnostico_iniciales_completo
aplicacion_diagnostico_iniciales_incompleto
cuenta_con_plan_de_reinsercion_en_el_area_familiar
fecha_plan_de_reinsercion_en_el_area_familiar
cuenta_con_plan_de_reinsercion_en_el_area_social
fecha_plan_de_reinsercion_en_el_area_social
cuenta_con_plan_de_reinsercion_en_el_area_laboral
fecha_plan_de_reinsercion_en_el_area_laboral
cuenta_con_plan_de_colocacion_laboral
cuenta_con_plan_de_capacitacion_laboral
Cuenta con plan de apresto laboral
cuenta_con_plan_de_reinsercion_en_el_area_judicial

fecha_plan_de_reinsercion_en_el_area_judicial
reinsertado_en_ambito_familiar
fecha_reinsertado_en_ambito_familiar
reinsertado_en_ambito_social
fecha_reinsertado_en_ambito_social
reinsertado_en_ambito_laboral
fecha_reinsertado_en_ambito_laboral
Colocado laboralmente
Capacitado laboralmente
Aprestado laboralmente
reinsertado_en_ambito_judicial
fecha_reinsertado_en_ambito_judicial
egresado
fecha_egreso
Razon de egreso
Sujeto Reinsertado FINAL
fecha_reinsercion_final
observaciones_reinsercion
diagnosticos_a_dos_meses_de_ingreso
cuentan_con_plan_de_reinsercion
requieren_plan_de_reinsercion
planes_a_dos_meses_de_ingreso
capacitado_laboralmente_si_requiere
colocado_laboralmente_si_requiere
autor_mayus estado
Id
Autor
Fecha Ingreso de Ficha
fecha_ultima_modificacion_de_ficha

Autor Completo

Edad Num

Anexo B: Resultados regresión logística multivariada

```

call:
glm(formula = Y ~ ., family = binomial, data = BBDD)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.7476 -1.1400  0.7028  1.0620  1.6578

Coefficients:
                    Estimate Std. Error z value Pr(>|z|)
(Intercept)          2.26459    0.63529   3.565 0.000364 ***
Edad                 -0.04415    0.01207  -3.657 0.000255 ***
`Grupo de referencia`  0.02045    0.08616   0.237 0.812399
`Colocado laboralmente si requiere`0 -0.28808    0.24191  -1.191 0.233718
`Colocado laboralmente si requiere`1 -0.57422    0.52559  -1.093 0.274602
`Nivel educacional numerica` -0.05323    0.07544  -0.706 0.480427
Genero_Masculino1    -0.52965    0.29379  -1.803 0.071419 .
Profesion_Si1        -0.03480    0.26889  -0.129 0.897025
`Cumplimiento del Plan de Reinserción_Si`1 0.13080    0.37444   0.349 0.726855
`Comuna_La Pintana`1    0.01798    0.49143   0.037 0.970815
Comuna_Maipu1         0.30411    0.47126   0.645 0.518726
Comuna_Peñalolen1     0.84499    0.51281   1.648 0.099402 .
Comuna_Pudahuel1      0.32051    0.52076   0.615 0.538244
`Comuna_Puente Alto`1  -0.51560    0.49931  -1.033 0.301780
Comuna_Renca1        -0.23115    0.45108  -0.512 0.608339
`Comuna_San Bernardo`1 0.03930    0.52971   0.074 0.940851
Comuna_Santiago1      0.95331    0.55449   1.719 0.085569 .
Comuna_El_Bosque1    -0.15609    0.45629  -0.342 0.732287
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Coeficientes y medidas de tendencia

```

lrm(formula = as.formula(paste("Y ~", paste(names(BBDD[, -c(ncol(BBDD))])),
collapse = "+"), collapse = "")), data = BBDD)

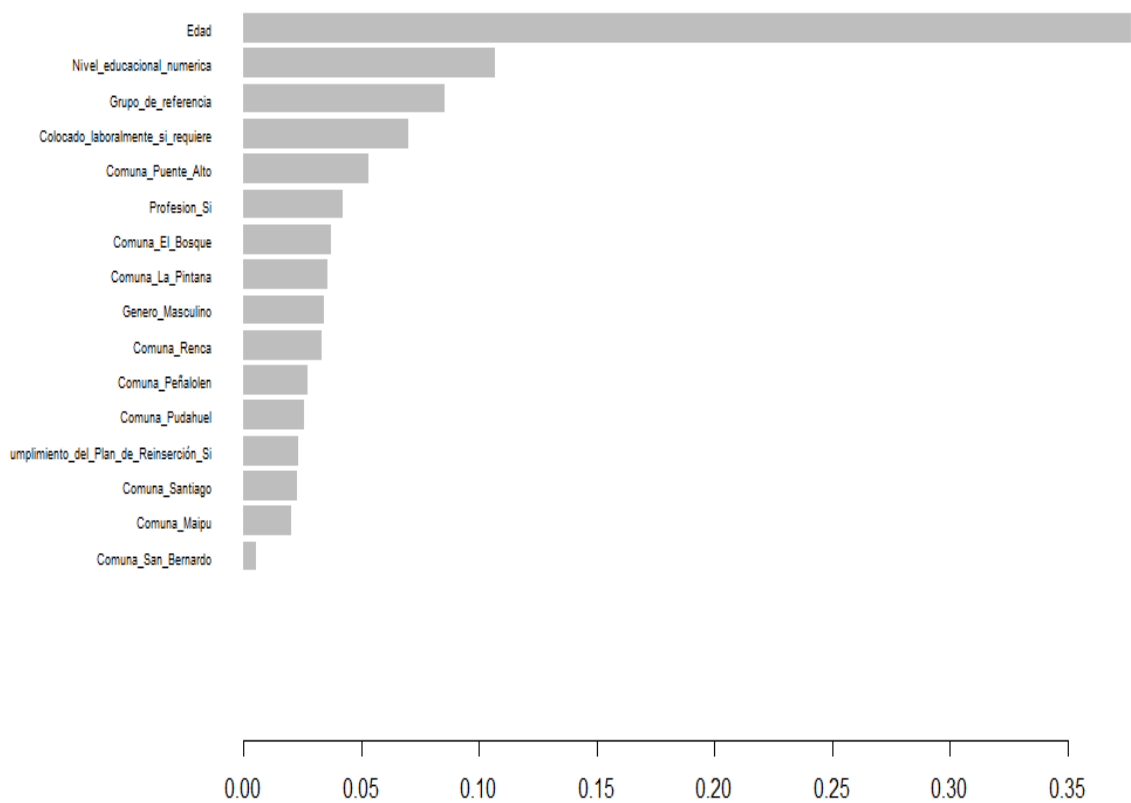
```

		Model Likelihood Ratio Test	Discrimination Indexes	Rank	Discrim. Indexes
Obs	377	LR chi2 30.15	R2 0.103	C	0.659
0	178	d.f. 17	R2(17,377)0.034	Dxy	0.318
1	199	Pr(> chi2) 0.0253	R2(17,281.9)0.046	gamma	0.318
max deriv	8e-12		Brier 0.230	tau-a	0.159

	Coef	S.E.	wald Z	Pr(> Z)
Intercept	2.2646	0.6353	3.56	0.0004
Edad	-0.0441	0.0121	-3.66	0.0003
Grupo_de_referencia	0.0204	0.0862	0.24	0.8124
Colocado_laboralmente_si_requiere=0	-0.2881	0.2419	-1.19	0.2337
Colocado_laboralmente_si_requiere=1	-0.5742	0.5256	-1.09	0.2746
Nivel_educacional_numerica	-0.0532	0.0754	-0.71	0.4804
Genero_Masculino=1	-0.5297	0.2938	-1.80	0.0714
Profesion_Si=1	-0.0348	0.2689	-0.13	0.8970
Cumplimiento_del_Plan_de_Reinserción_Si=1	0.1308	0.3744	0.35	0.7269
Comuna_La_Pintana=1	0.0180	0.4914	0.04	0.9708
Comuna_Maipu=1	0.3041	0.4713	0.65	0.5187
Comuna_Peñalolen=1	0.8450	0.5128	1.65	0.0994
Comuna_Pudahuel=1	0.3205	0.5208	0.62	0.5382
Comuna_Puente_Alto=1	-0.5156	0.4993	-1.03	0.3018
Comuna_Renca=1	-0.2312	0.4511	-0.51	0.6083
Comuna_San_Bernardo=1	0.0393	0.5297	0.07	0.9409
Comuna_Santiago=1	0.9533	0.5545	1.72	0.0856
Comuna_El_Bosque=1	-0.1561	0.4563	-0.34	0.7323

Estimadores de estadísticos relevantes

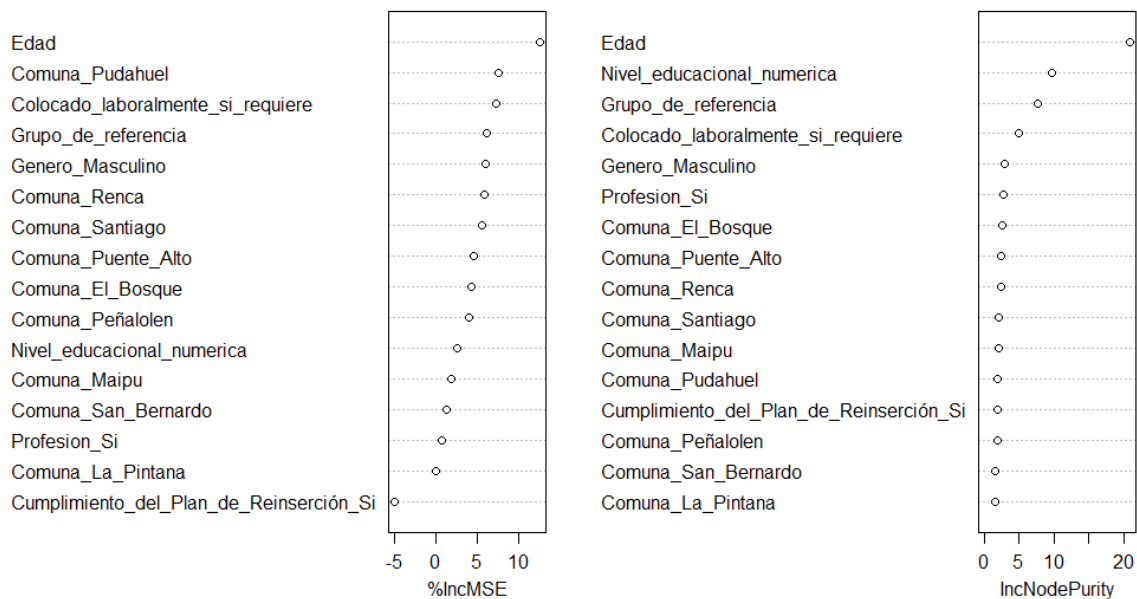
Anexo C: Importancias para XGBoost



Resultados de XGBoost obtenidos con RStudio.

Anexo D: Importancias para Random Forest

modelo_RF



Resultados con función MSE y Node Purity