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

UNA METODOLOGÍA PARA PREDECIR LAS PREFERENCIAS POR ESTABLECIMIENTOS ESCOLARES DE ESTUDIANTES Y APODERADOS A PARTIR DEL SISTEMA DE ADMISIÓN ESCOLAR CHILENO

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN GESTIÓN Y POLÍTICAS PÚBLICAS
MEMORIA PARA OPTAR AL TÍTULO DE INGENIERA CIVIL INDUSTRIAL

FRANCISCA CONSTANZA GUZMÁN ALISTE

PROFESOR GUÍA:
PATRICIO RODRÍGUEZ VALDÉS

MIEMBROS DE LA COMISIÓN:
FELIPE TOBAR HENRÍQUEZ
JUAN PABLO VALENZUELA BARROS

SANTIAGO DE CHILE
2023

RESUMEN DE LA TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN GESTIÓN Y POLÍTICAS PÚBLICAS
RESUMEN DE LA MEMORIA PARA OPTAR AL TÍTULO DE INGENIERA CIVIL INDUSTRIAL
POR: FRANCISCA CONSTANZA GUZMÁN ALISTE
FECHA: 2023
PROFESOR GUÍA: PATRICIO RODRÍGUEZ VALDÉS

UNA METODOLOGÍA PARA PREDECIR LAS PREFERENCIAS POR ESTABLECIMIENTOS ESCOLARES DE ESTUDIANTES Y APODERADOS A PARTIR DEL SISTEMA DE ADMISIÓN ESCOLAR CHILENO

Con la promulgación de la Ley de Inclusión Escolar (2015) que prohíbe la selección, elimina el copago y prohíbe el lucro en establecimientos educacionales que reciben aportes del Estado, se crea el Sistema de Admisión Escolar (SAE). El SAE es un mecanismo de postulación centralizado del Ministerio de Educación, a través del cual las familias postulan a los establecimientos educacionales públicos y particulares subvencionados de su preferencia. El SAE utiliza un algoritmo matemático que asigna la demanda a la oferta de forma justa basado en la solución de Gale-Shapley.

La aplicación del SAE comenzó el año 2017, y ha ido acumulando valiosos datos de los postulantes y sus preferencias. En particular, es posible conocer la real demanda de los establecimientos públicos y particulares subvencionados, que antes no era explícita por la autogestión de los establecimientos en materia de admisión.

El propósito de esta tesis es generar evidencia estadística y proponer una metodología para estudiar la demanda por educación pública a partir de los datos que otorga el SAE desde su creación, para habilitar la toma de decisiones informada en la formulación de políticas públicas para la planificación de la oferta educativa, especialmente la pública.

La presente investigación identifica para cada territorio, la cantidad de oferta y demanda en los establecimientos educacionales, y estima la capacidad ociosa de establecimientos en un territorio dado. Además, propone 2 modelos para predecir demanda a partir de las preferencias de las familias y de las vacantes ofrecidas y utilizadas por cada establecimiento educacional según los resultados del SAE.

Los hallazgos sugieren porcentajes críticos de desocupación en las regiones de Valparaíso, Maule, Ñuble, La Araucanía y Biobío. Por ello, el estudio se focaliza en esta última región. Existe una desocupación mayor en establecimientos públicos versus particulares subvencionados. Además, solo un 66% de los estudiantes se matriculó efectivamente en el EE asignado por el SAE en el proceso de admisión 2020.

Sobre los modelos propuestos, se releva la importancia del nivel socioeconómico en la preferencia de las familias, así como la distancia del estudiante al establecimiento, lo que se condice con las investigaciones previas. En términos de desempeño, los modelos tienen amplias oportunidades de mejora y perfeccionamiento, pero constituyen un primer acercamiento a la predicción de demanda a partir de los datos del SAE y se espera una continuación en futuras investigaciones.

A mis padres, Erika y Juan, por regalarme la oportunidad de soñar sin límites

Agradecimientos

Terminando esta intensa pero feliz etapa, quisiera comenzar agradeciendo a mi madre Erika Aliste y a mi padre Juan Guzmán, quienes son el pilar fundamental de mi vida y han sido mi mayor apoyo en esta carrera, animándome siempre a ir más lejos y confiando en cada uno de los pasos y decisiones que he tomado. Gracias por su amor incondicional, por creer en mí y permitirme soñar sin límites, ambos son un tremendo ejemplo de esfuerzo, resiliencia y bondad que me inspira a crecer como persona y profesional día a día.

A todas aquellas personas con las que he tenido la oportunidad de compartir en mi proceso universitario, por enseñarme y nutrirme de nuevas experiencias y conocimientos. Especialmente a mis amigos/as de la universidad: Pau, Dani, Vale, Coni, Fer, Chino, Yuyin, Takechi, Chelo, Rafa, Rodri Z., Mati C., Pancho, y tantos otros que tengo el placer de conocer y no alcanzo a nombrar en este breve apartado. Gracias por las caminatas al metro, los almuerzos, los cafecitos, las tutorías, los carretes y tantos momentos vividos en estos largos años.

Un reconocimiento especial a mis queridas Pau y Dani, "las industriales", por todos estos años de pánico y estrés, pero también de risas y momentos inolvidables. La universidad no hubiera sido lo mismo sin ustedes, y hoy puedo decir que cada trasnochada valió la pena. Gracias por sus consejos y por ser mis partners de trabajo y de vida.

A mi compañero Mati, quien me ha alentado a seguir y ha creído en mí incondicionalmente. Gracias por apoyarme con tu amor, entrega y paciencia durante todo este proceso.

A mi mejor amigo Vicho, por ser mi hermano, mi mejor consejero y escucha. Gracias por todo. A mis amigas del colegio, por su apoyo constante y sus palabras de aliento.

A mi profesor guía Patricio, a quien agradezco su gran apoyo, dedicación y trabajo puesto en esta investigación.

Por último, a todos quienes sueñan con un país más justo y menos desigual, y trabajan día a día para lograrlo, para que nunca perdamos la esperanza.

Tabla de Contenido

Índice de Tablas

Índice de Ilustraciones

Capítulo 1: Introducción

En Chile, el sistema educativo funciona respondiendo a lógicas de mercado, en función de las reformas y políticas impuestas durante la dictadura militar en el año 1980, que se sustentan en la implementación del sistema neoliberal. Así, los establecimientos educacionales han tenido que competir entre sí por la demanda de familias que desean acceder a este servicio. Como resultado, se generó un sistema educacional altamente segregado y estratificado, pues en la práctica las familias de nivel socioeconómico alto acceden a la educación privada, las familias de nivel socioeconómico medio acceden a la educación subvencionada y las familias de nivel socioeconómico medio-bajo y bajo asisten a la educación pública (Elacqua, 2012; Valenzuela, Bellei, & Ríos, 2014; Bellei, Contreras, Canales, & Orellana, 2019).

Además, por las características de este sistema, el estudio de los criterios que sustentan las preferencias de las familias al elegir establecimientos educacionales para sus hijos se ha convertido en un aspecto importante de análisis, y si bien ya existe la hipótesis instalada de que la elección no sería completamente racional, aún hay un campo inexplorado a la hora de entender qué variables influyen exactamente en esta decisión. Variados autores postulan que, si bien la calidad educacional es un factor considerado, también se incluye la ponderación de otros factores tales como los sociodemográficos, los económicos y los culturales (Orellana, Caviedes, Bellei, & Contreras, 2018). Por ejemplo, se afirma que las familias de nivel socioeconómico bajo operan guiadas por consideraciones prácticas como la distancia, la gratuidad o el conocimiento previo del establecimiento escolar (Córdoba, 2014; Chumacero, Gomez, & Paredes, 2008; Raczynski, Salinas, de la Fuente, Hernandez, & Lattz, 2010).

En definitiva, está claro que debido a la forma en la que se configura el sistema escolar y las dinámicas de segregación existentes, la educación pública ha sido relegada a un rol de sustentar la demanda que no es satisfecha por los establecimientos privados o subvencionados, los que son percibidos de mejor calidad o mejor ambiente social. Actualmente, un 63% de la oferta educativa es particular (54% particular subvencionado y 9% particular pagado) versus un 36% del sector público (CIPER Académico, 2020). En esa línea, es relevante cuestionarse cuál debería ser el rol del Estado en la provisión de educación para las familias, y además cuestionarse qué le depara a la educación pública chilena bajo el escenario actual, en que prácticamente sólo los estudiantes que no logran entrar a la educación privada entran a la pública.

Con la implementación de la Ley de inclusión escolar en el año 2015, que prohíbe la selección, elimina gradualmente el financiamiento compartido (copago) y prohíbe el lucro

en la educación pública, se busca revertir paulatinamente la segregación escolar. En esa línea es que en el año 2017 se comienza a implementar el Sistema de Admisión Escolar (SAE), con el objetivo de tener un sistema objetivo, transparente, y en que todos los estudiantes estén en igualdad de condiciones sin importar sus notas o nivel socioeconómico. La implementación del SAE ha servido de insumo para estudiar el proceso de elección de las familias, pero hay otro valor poco explotado en el sistema, que está en el poder estudiar la ocupación de los establecimientos y su cantidad precisa de oferta y demanda, pudiendo visibilizar territorios donde hay sub o sobre oferta para una mejor utilización de los recursos y provisión del servicio de educación por parte del Estado.

Así, esta tesis pretende aportar en ese ámbito, particularmente en el estudio de la ocupación escolar, y de cómo se podría abordar una predicción de demanda a partir del entendimiento y pronóstico matemático de las preferencias de los estudiantes y sus familias y/o de los resultados finales de oferta y demanda por establecimiento escolar y nivel. Además, también pretende ser exploratoria, en cuánto aún no se estudia profundamente si el uso del SAE se traduce efectivamente en estudiantes matriculados a los establecimientos asignados por él. Así, se propone como objetivo general el estudiar los niveles de desocupación escolar, así como el nivel de uso del sistema de admisión escolar. Asimismo, generar un modelo para predecir las preferencias de los padres como primer acercamiento para lograr predecir la demanda de educación pública en los próximos años, para la toma de decisiones informada respecto a la configuración de la oferta en los distintos territorios.

El aporte de conocimiento respecto a la ocupación escolar y al nivel de uso del SAE que generará el presente trabajo de tesis, en conjunto con el facilitar una metodología y un instrumento para predecir la demanda por educación pública en el corto plazo, es sumamente valioso para la toma de decisiones y generación de políticas públicas eficientes y de impacto, basadas en evidencia y uso de la tecnología. El uso de analítica avanzada en educación ha tomado gran relevancia y la aplicación de modelos predictivos presenta una oportunidad para desarrollar una herramienta que determine la preferencia de las familias en base a las variables que la literatura ha apuntado que la definen.

El presente documento está estructurado como sigue. En el capítulo 2, se plantean los objetivos de investigación. Luego, en el capítulo 3 se revisa la literatura, caracterizando el sistema educativo chileno e identificando trabajos previos de las preferencias de las familias. El capítulo 4 se justifica la importancia de predecir la demanda, así como definir los modelos a utilizar para ello en el contexto de *machine learning*. El capítulo 5 explica la metodología, y se enfoca particularmente en la elección de la muestra, variables a utilizar y métricas para medir el desempeño. Luego, el capítulo 6 ilustra los resultados, tanto del análisis estadístico como de la predicción que arrojan los modelos, haciendo un análisis preliminar de su eficiencia y precisión. Finalmente, se presentan las conclusiones obtenidas a partir del trabajo desarrollado.

Capítulo 2: Objetivos

2.1. Objetivo general

El objetivo general de esta investigación es estudiar los niveles de desocupación escolar, así como el nivel de uso del sistema de admisión escolar. Asimismo, generar un modelo para predecir las preferencias de los padres como primer acercamiento para lograr predecir la demanda de educación pública en los próximos años, para la toma de decisiones informada respecto a la configuración de la oferta en los distintos territorios.

2.2. Objetivos específicos

Los objetivos específicos de esta investigación son los siguientes:

1. Diagnosticar el estado actual de la ocupación de establecimientos para identificar territorios con niveles críticos de subocupación de establecimientos.
2. Diagnosticar el estado actual del nivel de uso del sistema de admisión escolar.
3. Explorar la posibilidad de predecir la demanda general por educación, y en particular por la pública, a través de algoritmos de *machine learning*.

Capítulo 3: Marco teórico

El sistema educacional chileno se caracteriza por operar como un cuasi mercado en el contexto de los procesos de privatización y expansión de políticas neoliberales impuestas en los últimos 40 años en el país. En esa línea, se cuestiona el rol que debiese tener el Estado respecto de la provisión, aseguramiento y regulación del sistema educativo y se destaca la llamada "libertad de elección" de las familias por escoger un establecimiento (Valenzuela, Bellei, & Ríos, 2014).

El sistema de *voucher* o subvención fue pensado en una lógica de mercado, bajo la idea de que el financiamiento estatal sería eficientemente distribuido hacia los mejores establecimientos educacionales, pues las familias en el modelo de *school choice* preferirían los mejores establecimientos y, por ende, estos querrían mejorar en todo ámbito para ser "atractivos" dentro del mercado. Sin embargo, la consecuencia de este sistema es una alta segregación social y estratificación socioeconómica entre establecimientos educacionales y alumnos que acceden a ellos (Elacqua, 2012). En concreto, las familias de nivel socioeconómico alto acceden a la educación privada, las familias de nivel socioeconómico medio acceden a la educación subvencionada (en donde la mayoría actualmente son gratuitos, pero una parte posee la modalidad de copago, es decir, reciben recursos del Estado y además cobran un arancel obligatorio), y las familias de nivel socioeconómico medio-bajo y bajo asisten a la educación pública (Elacqua, 2012; Valenzuela, Bellei, & Ríos, 2014; Bellei, Contreras, Canales, & Orellana, 2019). Lo anterior se debe a que "ni todos los padres valoran y priorizan los mismos elementos o factores al elegir escuelas para sus hijos, ni todos están en igualdad de condiciones para elegir en función de sus reales preferencias" (Román & Murillo, 2014). Así, son los sectores más privilegiados y con mayor capital cultural quienes acceden a mayor y mejor información, y por ello son quienes han podido maximizar los beneficios del modelo de *school choice*. (Carrasco, Gutiérrez, & Flores, 2017; Cucchiara, 2013; Lareau & Goyette, 2014; Orellana, Caviedes, Bellei, & Contreras, 2018; Orfield & Frankenberg, 2013; Reardon & Owens, 2014).

Aun así, existe muy poca evidencia en la literatura que explique concretamente las variables que se juegan en las preferencias y elecciones de las familias al elegir un establecimiento educacional. Variados autores afirman que la elección va más allá de la teoría de la acción racional, incluyendo la ponderación de otros factores tales como los sociodemográficos, los económicos y los culturales (Orellana, Caviedes, Bellei, & Contreras, 2018). Por ejemplo, se afirma que las familias de nivel socioeconómico bajo operan guiadas por consideraciones prácticas como la distancia, la gratuidad o el conocimiento previo del establecimiento escolar (Córdoba, 2014; Chumacero, Gomez, & Paredes, 2008; Raczynski,

Salinas, de la Fuente, Hernandez, & Lattz, 2010). En esa línea, un estudio sobre la experiencia subjetiva del proceso de elección de establecimientos educacionales en apoderamos de escuelas municipales de la RM (2013), concluyó que todos los entrevistados priorizan razones prácticas frente a las académicas a la hora de elegir un establecimiento, siendo la distancia el criterio determinante en esta elección (Gubbins, 2013). Otro estudio realizado el año 2021, buscó predecir los atributos de los establecimientos educacionales que pueden predecir la preferencia de las familias, observó que al controlar las restricciones del copago, se demostró estadísticamente que las familias en general escogen "lo mejor" dentro de lo disponible, considerando su radio de decisión y restricción presupuestaria, que particularmente las familias de bajo nivel educacional (educación básica completa o menor) escogían establecimientos de menor rendimiento y mayor proporción de estudiantes vulnerables, aludiendo al fenómeno de la auto segregación (Garay & Sillard, 2021). En conclusión, parece ser que la elección de las familias por establecimientos educacionales no es completamente racional ni basada en razones de calidad o excelencia académica, y además involucra información asimétrica entre las familias, lo que genera una auto segregación.

Respecto a las familias de clase media, algunos autores plantean que están en búsqueda de las mejores opciones posibles según sus restricciones presupuestarias y de información, apuntando idealmente a establecimientos subvencionados o a los "mejores" establecimientos públicos (Hernández & Raczynski, 2015; Corvalán & Román, 2016). Este grupo social actúa bajo expectativas de aislamiento social (Elacqua, Schneider, & Buckley, 2006; Chumacero & Paredes, 2012), en donde siempre preferirán establecimientos que se perciban mejor tanto en su calidad como en estándares de seguridad, respeto y organización (Duk & Murillo, 2019), buscando formas de diferenciarse y alejarse de los sectores populares, de aquellos alumnos catalogados como indeseados (Canales, Bellei, & Orellana, 2016) y de establecimientos que se perciben con problemas de seguridad, convivencia deteriorada y donde los aprendizajes parecen no mejorar con el tiempo (Dirección de Educación Pública, 2018).

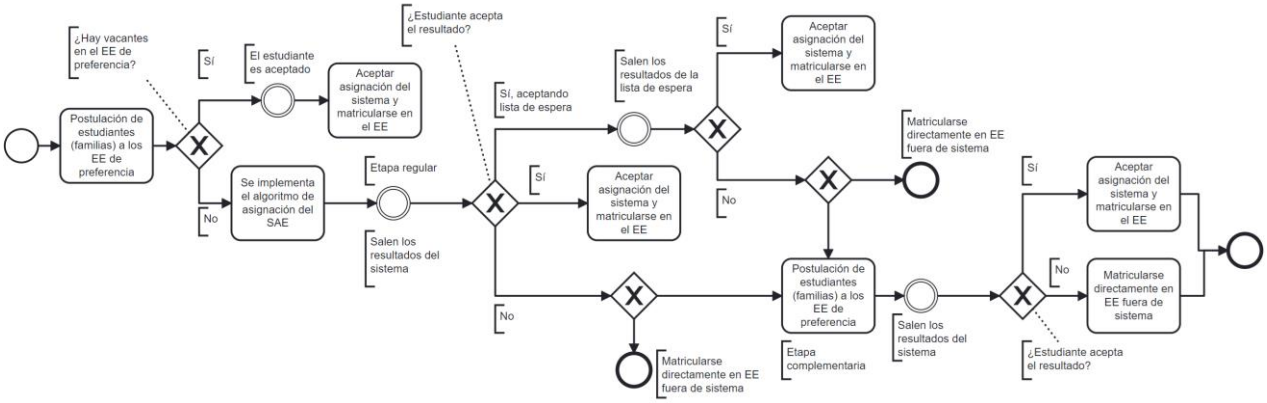
Con todo lo anterior, se puede concluir que en Chile la educación particular subvencionada no ha crecido junto a la educación pública, sino que ha crecido a costa de ella, ya que la existencia del copago restringe la capacidad de los sectores más vulnerables al no poder solventar los costos de arancel y matrícula. Actualmente, un 63% de la oferta educativa es particular (54% particular subvencionado y 9% particular pagado) versus un 36% del sector público (CIPER Académico, 2020). Es más, la encuesta CEP 2011 mostró que la mayor proporción de las personas encuestadas, y con hijos en edad escolar, preferiría un colegio particular subvencionado a una escuela o liceo municipal (69%) (CEP, 2011).

La regulación más relevante para el sector privado subvencionado en el último tiempo es la Ley de Inclusión Escolar N°20.845 (2015), que busca hacerse cargo de la fuerte segregación social y desregulación del modelo de mercado educativo. Esta ley regula la admisión

de los y las estudiantes (prohíbe la selección), elimina el financiamiento compartido (co-pago) y prohíbe el lucro en establecimientos educacionales que reciben aportes del Estado (Biblioteca del Congreso Nacional, 2015). Posteriormente, se promulga la Ley de Nueva Educación Pública (2017), que involucra el traspaso de la gestión y administración de la educación pública desde los municipios a nuevos entes autónomos, llamados Servicios Locales de Educación Pública (SLEP) (Biblioteca del Congreso Nacional, 2017).

Así, a partir de estas nuevas regulaciones, se crea el Sistema de Admisión Escolar (SAE). El SAE es un mecanismo de postulación del Ministerio de Educación (MINEDUC), para que padres y apoderados puedan, a través de una plataforma centralizada en Internet, postular en orden de preferencia a los establecimientos educacionales públicos y particulares subvencionados que deseen para sus hijos/as, según el nivel en que estén, según el flujo señalado en la ilustración 1[1]. Importante es señalar que posterior a las postulaciones, la asignación de cupos se realiza a través de un algoritmo matemático (Correa, y otros, 2021) que considera la preferencia de los padres, un set de reglas de asignación definidas por la ley y los cupos disponibles en cada establecimiento educativo (Ministerio de Educación, s.f). Tal y como se puede observar en la ilustración, existe una etapa regular que considera el avance de una lista de espera: y una etapa complementaria (donde se suman estudiantes que no quedaron en su preferencia o no fueron asignados en la etapa regular, y algunos estudiantes nuevos que no se inscribieron a tiempo en la regular). Además, posterior a la asignación de ambas etapas, y finalizando el proceso del SAE, corresponde un periodo de regularización general, donde a discreción de cada familia y cada EE, se pueden abrir cupos y se genera un proceso de selección directa en los EE.

Ilustración 1: Funcionamiento del SAE



[1] La ilustración 1 pretende dar un lineamiento general al lector sobre cómo funciona el SAE, más no está construido como una herramienta técnica ni abarca todos los casos posibles. Para más información ver en: https://www.mineduc.cl/wp-content/uploads/sites/19/2022/10/17102022_Protocolo-resultados-y-listas-de-espera_EEEE.pdf

Fuente: Elaboración propia

Cabe destacar que el SAE lleva seis años funcionando. El primer año (2017) se aplicó como marcha blanca solo en la región de Magallanes; el segundo (2018) se implementó además en las regiones de Tarapacá, Coquimbo, O'Higgins y Los Lagos. Luego en 2019 el SAE funcionó en todas las regiones, excepto en la Región Metropolitana, que se sumó finalmente en 2020.

La incorporación del SAE es relevante, pues no sólo busca centralizar y facilitar el proceso de postulación a las familias, sino también velar por que el proceso sea objetivo, transparente y no discriminatorio. A pesar de ser un aspecto poco analizado, es evidente que el rol que juega la selección escolar en la segregación es crucial, pues la selección no sólo se opone a la libertad de elección de las familias, sino que también desde la perspectiva del derecho a la educación y a la inclusión es discriminatoria al limitar las oportunidades de los estudiantes y afectar su dignidad (Duk & Murillo, 2019). Así, con la implementación del SAE se busca que todos los estudiantes estén en igualdad de condiciones sin importar sus notas o nivel socioeconómico y con ello se disminuya la segregación escolar (Sillard, Garay, & Troncoso, 2018). A pesar de esto, no se debe subestimar que la implementación del SAE posee una base de legitimidad moral aún frágil (Carrasco & Honey, 2019), debido a las contrarias posturas en torno al principio de acceso igualitario, sustentadas en la defensa de la meritocracia como valor que hace justicia al mérito y esfuerzo personal (Duk & Murillo, 2019). Un estudio realizó entrevistas a familiares que pasaron por el SAE en sus primeros años de implementación, y los resultados indican como elementos positivos la reducción de tiempos y costos al postular y una sensación de equidad y transparencia. Además, como elementos negativos menciona la ansiedad por los resultados y la despersonalización del proceso (Carrasco & Honey, 2019).

Hasta el momento los estudios en torno al SAE han sido principalmente para analizar el comportamiento de las familias y su proceso de elección (Duk & Murillo, 2019). Sin embargo, no se ha profundizado en torno al valor que entrega el SAE como insumo para transparentar la real demanda de los establecimientos públicos y particulares subvencionados, lo que antes no era explícito por la autogestión de los establecimientos en materia de admisión. Se ha mencionado que "esta dimensión de bien público informativo que genera el sistema a partir de las postulaciones no ha sido suficientemente destacada ni explotada" (Eyzaguirre, Hernando, & Blanco, 2018). Así, la implementación del SAE permite visibilizar para cada territorio, la cantidad de oferta y demanda (cupos totales y postulantes) y la sub o sobre oferta en los establecimientos educacionales (Garay & Sillard, 2021), con lo que se puede obtener una estimación de la "capacidad ociosa" de establecimientos en un territorio dado, entendiendo que esto es un insumo relevante para la toma de decisiones en política pública debido a los grandes recursos monetarios que invierte el Estado en educación. En esa línea, se torna relevante el profundizar en los estudios de las variables

del SAE, pues su análisis permitiría dilucidar aspectos relevantes del fenómeno de elección escolar y, a su vez este nuevo conocimiento generaría políticas públicas más efectivas en base a datos e información palpable (Garay & Sillard, 2021).

El mayor desafío de esta tesis, además de diagnosticar la desocupación y el uso del SAE en la realidad, está en lograr modelar la demanda por educación pública, en un contexto de alta segmentación social y con atributos poco claros para entender el proceso de toma de decisiones sobre las preferencias. Además de esto, se debe considerar que la elección de establecimiento escolar también depende del territorio y del nivel en que se estudie, pues no será lo mismo ingresar por primera vez al sistema escolar en primero básico, que cambiarse de establecimiento ya dentro del sistema.

Capítulo 4: Marco conceptual

4.1. Valor público en el uso de analítica avanzada

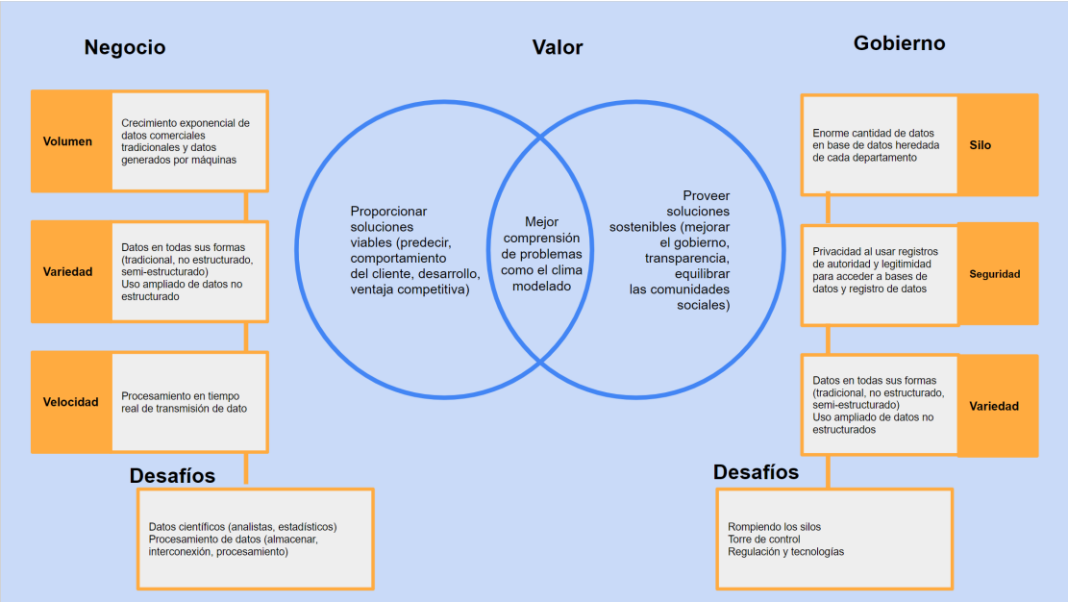
El uso de analítica avanzada en el sector público es importante tanto en el diseño, implementación y evaluación de políticas públicas basadas en evidencia pertinente, de calidad y oportuna (Rodríguez, Palomino, & Mondaca, 2017).

La toma de decisiones guiadas por datos o *data-driven decision making* en el contexto de lo público, produce valor para la economía mundial, mejorando la productividad y competitividad privada y públicas, y creando excedentes económicos para los consumidores. En esa línea, el uso de analítica avanzada provee beneficios en el ejercicio de la administración pública, al proveer de más y mejores soluciones que satisfagan necesidades de salud, educación, transporte, vivienda, atención e inclusión de grupos minoritarios, entre otras, a partir de contextos sociales, demográficos y territoriales particulares (Manyika & otros, 2011).

Rodríguez et. al (2017) postulan que el acceso a datos masivos y el uso de técnicas analíticas adecuadas permiten el desarrollo de una "inteligencia de valor público", concepto asociado a la inteligencia de negocios que se trabaja en el mundo privado, y que en este caso habilitaría una dimensión de valor estratégico para la toma de decisiones y el diseño, implementación y evaluación de políticas públicas dentro de los gobiernos de América Latina y el Caribe.

Los gobiernos, en el proceso de implementación de leyes y reglamentos y la prestación de servicios públicos y financieros, acumulan una enorme cantidad de transacciones de datos con atributos, valores, y desafíos que difieren de los que se manejan en el sector privado, tal y como se puede observar en la Ilustración 2.

Ilustración 2: Caracterización de datos y atributos en el sector público y privado



Fuente: Adaptado y traducido de Gang-Hoon, Silvana, & Ji-Hyong (2014)

En esa línea, no se puede olvidar que las transacciones difieren entre los sectores, debido a que los proyectos y la finalidad que tiene cada uno es muy diferente: el sector privado busca dar la mayor utilidad posible a su negocio, mientras que el sector público busca el desarrollo sustentable de la sociedad (ver Tabla 1). Además, existe una dificultad mayor en el sector público por su característica rígida y poco ágil, de lograr desarrollar nuevas capacidades y adoptar nuevas tecnologías para transformarla en información a través de la analítica de datos. En esa línea, si bien esta temática se presenta como un desafío, se espera que el *Big data* mejore la capacidad de los gobiernos para servir a los ciudadanos y abordar los principales desafíos nacionales que involucran la economía, la atención de la salud, la creación de empleo, los desastres naturales y el terrorismo (Gang-Hoon, Silvana, & Ji-Hyong, 2014).

Tabla 1: Caracterización de proyectos en el sector público y privado

Atributo	Sector privado	Sector público
Meta	Ganancia de los accionistas	Tranquilidad doméstica y desarrollo sustentable
Misión	Desarrollo de una ventaja competitiva y satisfacción del cliente	Aseguramiento de los derechos básicos (equidad, libertad y justicia), promoción del bienestar general y crecimiento económico
Toma de decisiones	Procesos de toma de decisiones de corto plazo para maximizar el interés personal y minimizar los costos	Procesos de toma de decisiones de largo plazo para maximizar el interés personal y promover el interés público
Actores en las decisiones	Número limitado de actores en las decisiones	Diversos actores en las decisiones
Estructura organizacional	Jerárquico	Gobernanza
Recursos financieros	Ingresos	Impuestos

Atributo	Sector privado	Sector público
Naturaleza de la actividad colectiva	Competencia y fidelización	Cooperación y control

Fuente: Adaptado y traducido de Gang-Hoon, Silvana, & Ji-Hyong (2014)

4.2. ¿Por qué calcular la demanda?

Predecir la demanda en cualquier contexto revela información valiosa como insumo para las decisiones de operaciones relacionadas con el diseño de procesos, la planeación de la capacidad, la asignación de las personas, las decisiones de equipamiento, entre otros (Schroeder, Meyer, & Rungtusanatham, 2011). Particularmente en educación pública, la pregunta de por qué es importante calcular la demanda se responde fácilmente al analizar la cantidad de recursos de todo tipo que invierten los Estados en proveer este servicio público, y con ello la importancia de optimizar los recursos teniendo en cuenta las dimensiones de público que recibirá el servicio y sus características. El tema de esta tesis se justifica al generar una metodología que aportará en la predicción de la demanda por educación pública en el país, y considerando que es un tema poco estudiado, viene a ser una invitación para continuar este análisis a futuro en otros proyectos de investigación, en aporte a políticas públicas basadas en evidencia.

En particular, la presente tesis, en conjunto con la existencia de los datos que provee el SAE, permite informar al sistema y alimentar las políticas públicas respecto de la relación entre la oferta y la demanda, aportando en comprender la sobreoferta de escuelas o la sobredemanda de postulantes en distintas zonas de Chile. Además, permite visibilizar la capacidad ociosa del sistema, es decir, la capacidad de los establecimientos que no son demandados por las familias. Las dificultades que evidencian los establecimientos públicos para ocupar las vacantes disponibles hacen presente la tendencia a maximizar matrícula según permitan infraestructura y regulaciones vigentes. La hipótesis es que existen territorios con niveles críticos de subocupación escolar, lo que sería una manifestación del sobre *stock* de establecimientos y vacantes, lo que es imperante de abordar a nivel de política pública en un contexto donde el número de hijos por familia se reduce y las tasas de cobertura son prácticamente universales (Rodríguez, Espinosa, & Padilla, 2020).

Un buen sistema de planificación educacional debería estructurar la oferta en términos de satisfacer requerimientos de calidad, cobertura, localización y accesibilidad. Esto se debe hacer usando instrumentos de planificación basados en evidencia e incorporando la dimensión espacial, geográfica y demográfica en la construcción y consolidación de la oferta pública (Amaya, y otros, 2021). Un buen sistema de planificación deberá adaptarse a los requerimientos de la demanda, lo que implica diversos desafíos, entre ellos:

1. Lidiar con territorios de baja matrícula que generan establecimientos no sustentables en términos económicos, y con infraestructura y dotación de personal mayor a las necesidades actuales.

2. Proveer nueva oferta en zonas de expansión urbana.

3. Propender a la equidad territorial, velando por reducir los tiempos de viajes de los estudiantes para asistir a establecimientos públicos de calidad.[2]

4.3. *Machine learning*

Teniendo en cuenta la situación actual de la educación pública en relación con la privada, es que se hace necesario entender la futura ocupación de los establecimientos educacionales públicos en el país para una toma de decisiones efectiva. En esa línea, se propone investigar para obtener una metodología que logre este propósito en cuánto al cálculo e identificación de establecimientos con subocupación de matrículas, a través del uso de *machine learning*.

El *machine learning* es una forma de inteligencia artificial, que permite a un sistema aprender de los datos en lugar de aprender mediante la programación explícita. Así, el algoritmo ingiere datos de entrenamiento, produciendo modelos más precisos basados en datos (IBM, s.f). En particular, mencionar que las técnicas de aprendizaje automático poseen múltiples aplicaciones, y una de ellas es su utilización como método para la previsión de la demanda, demostrando a través de los años su gran potencial en mejorar la eficiencia de la cadena de suministro y la planificación territorial (Mohamed, Putu, & Made, 2021).

Para esta tesis se probaron distintos algoritmos, y a continuación se describirá el que finalmente se utiliza para reportar resultados: *Random Forest.*

Los árboles de decisión son una clase de modelos de *machine learning* que se pueden considerar como una secuencia de declaraciones "si" que se aplican a una entrada para determinar la predicción. Un modelo *Random Forest* está formado por un conjunto de árboles de decisión individuales, cada uno entrenado con una muestra ligeramente distinta de los datos de entrenamiento generada mediante *bootstrapping*. La predicción de una nueva observación se obtiene agregando las predicciones de todos los árboles individuales que forman el modelo (Cutler, Cutler, & Stevens, 2011). Cabe destacar que este modelo puede usarse tanto para respuestas categóricas (discretas), es decir, como un clasificador;

[2] Desafíos obtenidos del estudio de Rodríguez, P.; Valenzuela, J.; Trufello, R.; Ulloa, J.; Matas, M.; Quintana, D.; Hernández, C.; Muñoz, C.; Requena, B. (2019)

o para respuestas continuas, es decir, como un regresor. Lógicamente en el caso de esta investigación la variable a predecir es un regresor dado que se busca predecir la demanda por educación.

Los modelos *Random Forest* son atractivos de utilizar pues, entre otras cosas, son relativamente rápidos para entrenar y predecir, dependen de uno o dos parámetros de sintonización, puede usarse directamente para problemas de alta dimensión y estadísticamente logran medir la importancia priorizada de las variables involucradas (Cutler, Cutler, & Stevens, 2011). En particular, el algoritmo de *Random Forest* funciona de la siguiente forma:

Sea $D = \{(x_1, y_1),...,(x_N, y_N)\}$ los datos de entrenamiento, con $x_i = (x_{i,1},..., x_{i,p})^T$. Para $j = 1$ a J:

1. Tomar una muestra de arranque D_j de tamaño N de D.
2. Usando la muestra de partida D_j como datos de entrenamiento, ajustar un árbol usando partición binaria recursiva:
    a) Comenzar con todas las observaciones en un solo nodo.
    b) Repetir los siguientes pasos recursivamente para cada nodo no dividido hasta que se cumpla el criterio de parada:
        i. Seleccionar m predictores al azar de los p predictores disponibles.
        ii. Encontrar la mejor división binaria entre todas las divisiones binarias en los m predictores del paso i.
        iii. Dividir el nodo en dos nodos descendientes utilizando el paso dividir desde ii.

Para hacer una predicción en un nuevo punto x,

$$\hat{f}(x) = \frac{1}{J}\sum_{j=1}^{J} \widehat{h_j}(x) \qquad \textit{Ecuación 1}$$

$$\hat{f}(x) = argmax_y \sum_{j=1}^{J} I(\widehat{h_j}(x) = y) \qquad \textit{Ecuación 2}$$

donde $\widehat{h_j}(x)$ es la predicción de la variable de respuesta en x utilizando el j-ésimo árbol (Cutler, Cutler, & Stevens, 2011).

Los métodos de *machine learning* han demostrado tener éxito en la industria, la academia y el aprendizaje automático competitivo (Wang & Hastie, 2014). En esta ocasión, para implementar los modelos se utilizó la librería *scikit-learn* de Python.

4.4. Métricas de evaluación

4.4.1. Ranked Biased Overlap (RBO)[3]

Considerando las características particulares del problema que se está abordando, en que parte de los resultados se reportan en listas priorizadas de preferencias por establecimientos de cada estudiante, es que se buscó una métrica apta para listas. RBO es una medida de similitud entre listas, que otorga pesos para cada posición de rango en la lista. Los pesos se derivan de una serie convergente, y la siguiente ecuación describe el RBO de dos listas clasificadas infinitas S y T:

$$RBO(S,T,p) = (1-p)\sum_{d=1}^{\infty} p^{d-1} A_d \qquad \textit{Ecuación 3}$$

donde,

$X_d = |\ S_{:d}\ \cap\ T_{:d}|$ (tamaño de la sobreposición de S y T hasta la profundidad 'd')

$A_d = X_d\ /\ d$ (acuerdo entre S y T dado por la proporción del tamaño de la superposición hasta la profundidad 'd')

El parámetro p es un parámetro ajustable en el rango (0, 1) que se puede usar para determinar la contribución de los rangos d superiores al valor final de la medida de similitud de RBO. Se debe considerar que el término de la sumatoria es convergente al ser una serie geométrica. Su suma viene dada por $\frac{1}{(1-p)}$ y como $0 < p < 1$, la suma es finita.

Para obtener un valor único de la métrica RBO, es posible realizar una extrapolación a partir de la información disponible, asumiendo que la concordancia vista hasta una profundidad k se mantiene indefinidamente entre las dos listas. Así, RBO se puede calcular usando:

[3] La descripción de esta métrica fue traducida desde: https://towardsdatascience.com/rbo-v-s-kendall-tau-to-compare-ranked-lists-of-items-8776c5182899

$$RBO(S,T,p,k) = \left(\frac{X_k}{k}\right) p^k + (\frac{(1-p)}{p} \sum_{d=1}^{n} p^d A_d) \qquad \textit{Ecuación 4}$$

En la métrica RBO, la elección del valor de p determina el grado de ponderación superior que representa el valor de RBO resultante. Se puede calcular el peso total que los rangos d superiores contribuyen al cálculo de RBO de la siguiente forma:

$$W_{RBO[1:d]} = 1 - p^{d-1} + \left(\frac{(1-p)}{p}\right) * d * \left(ln\left(\frac{1}{(1-p)}\right)\right) - \sum_{i=1}^{d-1} \frac{p^i}{i} \qquad \textit{Ecuación 5}$$

donde d es el largo del ranking evaluado.

Para el valor de $p = 0{,}9$ y $d = 10$ (se examinan los 10 primeros rangos), $W_{RBO}[1{:}d]$ es 0,8556, es decir, los 10 primeros rangos contribuyen en un 85,56 % a la medida final de RBO. Por lo tanto, dependiendo de la cantidad de contribución que le gustaría de los mejores resultados de d, puede elegir el valor de p en consecuencia usando la ecuación 5 (Webber, Moffar, & Zobel, 2010).

4.4.2. Métricas complementarias

Considerando las características particulares del problema que se está abordando, en que parte de los resultados se reportan en listas priorizadas de preferencias por establecimientos de cada estudiante, es que en esta investigación se crean otras métricas complementarias al RBO que podrían ser útiles para ver directamente cuán certero es el modelo. A continuación, se muestra el nombre de cada métrica creada, junto a su descripción y fórmula:

- **"Contador_presentes":** cuenta cuántos establecimientos educacionales (EE) de la lista predicha están presentes realmente en la lista real a la que postula el estudiante.

- **"Contador_presentes_porcentual"**: cuenta porcentualmente cuántos establecimientos educacionales (EE) de la lista predicha están presentes realmente en la lista real a la que postula el estudiante.

$$Contado\_presentes\_porcentual = \frac{(N°\ de\ EE\ de\ la\ lista\ predicha\ presentes\ en\ la\ lista\ real)}{N°\ total\ de\ EE\ de\ la\ lista\ real} * 100 \qquad \textit{Ecuación 6}$$

- **"*Match*":** cuenta cuántos EE de la lista predicha están presentes y además están en el orden correcto de la lista real a la que postula el estudiante.

- **"*Match*_porcentual":** cuenta porcentualmente cuántos EE de la lista predicha están presentes y además están en el orden correcto de la lista real a la que postula el estudiante.

$$Match\_porcentual = \frac{N° \text{ de EE de la lista predicha presentes en la lista real en el orden correcto}}{N° \text{ total de EE de la lista real}} * 100 \qquad \textit{Ecuación 7}$$

- **"Ratio_correctos":** es la división entre "*Match*" y "Contador_presentes", lo que permite entender cuál es la proporción de EE que están presentes y además se encuentran en el orden acertado.

$$Ratio\_correctos = \frac{Match}{Contador\_presentes} * 100 \qquad \textit{Ecuación 8}$$

De cada una de estas métricas se obtiene el promedio agregado, en conjunto con un histograma que permite ver la distribución de cada una.

4.4.3. Errores medios: MSE, RMSE y MAE[4]

Los errores medios son métricas ampliamente utilizadas para medir el desempeño de modelos de *machine learning*. Se subentiende que, para todo error, entre menor sea el valor, mejor es el desempeño del modelo. En particular, en esta tesis se utilizarán 3 errores para regresores, que se describen a continuación junto a su fórmula:

- **Error cuadrático medio (MSE):** mide el error cuadrado promedio de las predicciones, y para cada punto calcula la diferencia cuadrada entre las predicciones y los valores reales, sacando su promedio. Es una métrica que puede subestimar o sobre estimar el desempeño del modelo si es que existen *outliers*. Su fórmula es:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \qquad \textit{Ecuación 9}$$

Donde y_i es el resultado real esperado y $\hat{y}_i$ es la predicción del modelo.

- **Raíz del error cuadrático medio (RMSE):** es la raíz cuadrada del MSE. El efecto que tiene la raíz cuadrada es hacer que la escala de los errores sea igual a la escala de los valores reales. Su fórmula es:

[4] Descripciones y fórmulas obtenidas de: https://towardsdatascience.com/comparing-robustness-of-mae-mse-and-rmse-6d69da870828

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \qquad \textit{Ecuación 10}$$

- **Error absoluto medio (MAE):** es el promedio de diferencias absolutas entre los valores reales y las predicciones. Esta métrica tiene como ventaja que penaliza errores enormes, por lo que no es tan sensible a *outliers* como las métricas anteriores.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \qquad \textit{Ecuación 11}$$

Capítulo 5: Metodología

El estudio de caso se desarrollará con un enfoque cuantitativo a través de la aplicación de estadística descriptiva y modelos de *machine learning* para predicción de demanda, asociados a las preferencias de los padres y apoderados explicitadas en el SAE y a la oferta y demanda escolar respectiva. En esta investigación se busca justificar el problema a través de análisis estadísticos que sostengan la necesidad de conocer la ocupación de los establecimientos escolares públicos en los distintos territorios, para así aplicar los modelos de *machine learning* que permitan predecir la demanda a partir de las preferencias por establecimientos de los padres y apoderados y a la oferta y demanda escolar respectiva. El pilar en el que se sustenta el trabajo es la herramienta de Google, *Colaboratory*[5], que ejecuta código en lenguaje Python.

A grandes rasgos, lo que se hará en esta tesis es un primer análisis estadístico para conocer las principales cifras sobre SAE admisión 2020 (escogiendo este año al ser prepandemia de COVID-19, en que el escenario podría haber sido muy dispar a la regularidad). Luego, un análisis extenso sobre la ocupación escolar en los distintos territorios del país, mostrando las cifras más críticas e ilustraciones tipo mapa. Además, en el análisis también se estudió el nivel de uso real del SAE traducido en matrículas efectivas post asignación. Posterior al análisis estadístico, se pasará al modelo de predicción de demanda, que se ha estudiado de 2 formas. Primeramente, la metodología consiste en tomar una muestra de estudiantes y establecimientos escolares (y sus respectivas características sociodemográficas) a partir de los datos del SAE, y conociendo el orden de preferencia que asignó el estudiante (o familia) a cada establecimiento en la postulación, se probarán los datos en un modelo de *machine learning* que luego logre predecir a qué establecimientos escolares postulará un estudiante y el orden de preferencia que le asigna. El objetivo es que, con esta predicción de preferencias, se tendrán todas las listas de postulación de cada estudiante y se podrá ingresar esta información al algoritmo actual que utiliza el SAE y junto con ello se podrá hacer un análisis de oferta y demanda para identificar territorios con subocupación crítica, o capacidad ociosa relevante. Por otro lado, también se aplicó un modelo de *machine learning* basado directamente en la demanda a nivel de establecimiento escolar y nivel de enseñanza (curso), considerando las vacantes ofrecidas por el sistema, y las efectivamente utilizadas post asignación de este. Así, existirían 2 posibles formas de predecir la demanda por educación pública, como primer acercamiento a una metodología consolidada para ello. Cabe destacar que la mayor parte del trabajo de esta tesis se enfocó

[5] Para más información: https://colab.research.google.com/

en el primer modelo de predicción mencionado, ya que el segundo surge como una alternativa a explorar dado los resultados que entregó el primer modelo.

5.1. Elección de la muestra

Para obtener mejores resultados en el modelo, se debe considerar la hipótesis de que la elección de establecimiento educacional es un fenómeno social heterogéneo. No es lo mismo vivir ni estudiar en zona urbana que en zona rural, tampoco es lo mismo estudiar en las distintas regiones del país debido a las formas de desplazamiento que existen y otras dinámicas territoriales. Así también, el nivel al que se postula cambia los factores de elección, al menos al considerar una postulación a primero básico que es la entrada al sistema escolar, respecto de otro nivel en que ya se está dentro del sistema.

Así, a continuación, se describe las decisiones estratégicas que se tomaron para la elección de la muestra testeada en el modelo, junto a su debida justificación.

5.1.1. Ruralidad

Corresponde mencionar en esta sección, la lógica para elegir entre seleccionar las zonas urbanas o rurales para este estudio. Principalmente, es vital pensar la política pública como una herramienta posibilitadora de oportunidades a los ciudadanos y ciudadanas. Por ende, de tomar decisiones de política pública en torno a la desocupación crítica, siempre tendrá que hacerse con el debido resguardo de no mermar el derecho a educación de los estudiantes.

Así, es coherente pensar que independiente de qué tan desocupados puedan estar los establecimientos educacionales rurales debido a su lejanía geográfica, cantidad de habitantes en edad escolar, u otra variable; no será razonable para el Estado querer optimizar estos establecimientos en número y tamaño, pues el servicio que entregan debe ser impartido independiente de cuántas personas accedan a él. Esto no quiere decir que no se pueda mejorar la calidad, infraestructura y la planificación territorial, pero no sería tan valioso centrarse en este territorio debido a lo recién mencionado, y especialmente, no sería parte de los alcances de esta tesis.

En definitiva, debido a lo recién expuesto, es que se escogerá la **zona urbana del país** como primer filtro de estudio, considerando que existe un mayor rango de acción y mayor oferta y demanda tanto de establecimientos escolares como de estudiantes.

5.1.2. Territorio

En Chile hay 319 comunas que presentan población en zona urbana. La población de estas comunas posee distintos rasgos estadísticos ilustrados en el anexo A.1. En particular, hay 31 comunas con más de 150.000 habitantes (la distribución se puede ver en el anexo A.2), las que se asumen que al ser las con mayor población, son también las que más establecimientos educacionales tienen y por ende donde más inversión hay. Por ende, es que se eligen estas comunas para hacer el análisis de desocupación de establecimientos educacionales públicos, según el primer objetivo específico formulado.

En línea con lo anterior, se hace un análisis del nivel de ocupación escolar de cada una de las 31 comunas (ver en anexo A.3), identificando primeramente que algunas de ellas son de la Región Metropolitana, y considerando que esta región tiene características particulares que difieren de las demás en su esquema de desplazamiento escolar, es que se obvió esta región en el análisis. Lo anterior, sumado a que por interés propio de esta investigación no se desea caer en el centralismo y se priorizará escoger comunas de otras regiones. Dado lo anterior, quedan 18 comunas candidatas a ser estudiadas, y se procede a analizar la cantidad total de establecimientos educacionales (EE) públicos urbanos en cada una, buscando cuántos de estos tienen una desocupación de **30% o más como porcentaje crítico de desocupación**, quedándose con aquellas que presenten situación crítica, lo que se puede observar en la tabla 2. El porcentaje de criticidad fue definido al observar la distribución de desocupación en las comunas, siendo una decisión estratégica.

Tabla 2: Comunas seleccionadas con subocupación crítica.

Comuna	Región	Población total (Censo 2017, zona urbana)6	Población en edad estudiantil (5 a 19 años) *	Total de EE públicos	EE públicos con desocupación de 30% o más.
Viña del Mar	Valparaíso	295.918	61.687	45	27 (60%)
Valparaíso	Valparaíso	334.248	56.837	45	22 (48,9%)
Talca	Maule	210.916	43.293	33	17 (51,5%)
Chillán	Ñuble	168.647	35.435	25	14 (56%)
Los Ángeles	Biobío	151.087	34.029	25	10 (40%)
Talcahuano	Biobío	150.320	30.099	28	14 (50%)

[6] Fuente: https://www.ine.cl/estadisticas/sociales/censos-de-poblacion-y-vivienda/informacion-historica-censo-de-poblacion-y-vivienda

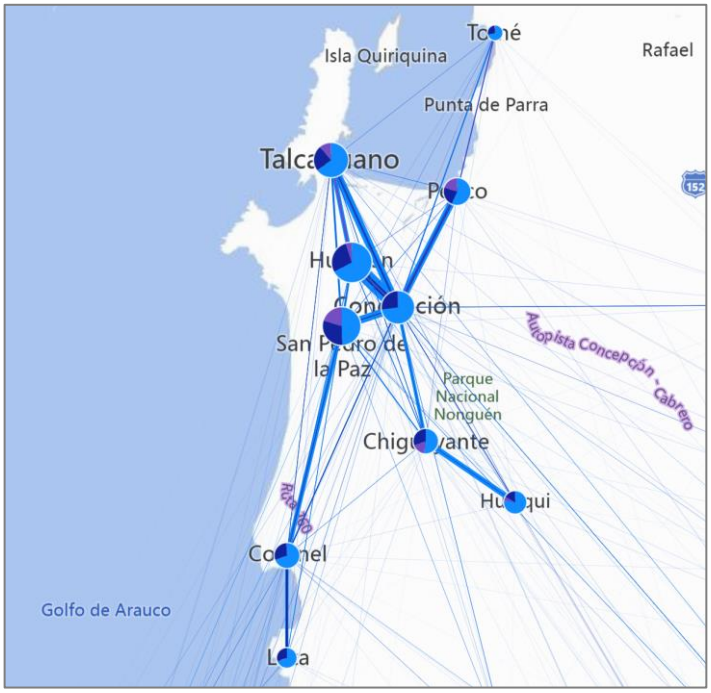
Comuna	Región	Población total (Censo 2017, zona urbana)6	Población en edad estudiantil (5 a 19 años) *	Total de EE públicos	EE públicos con desocupación de 30% o más.
Concepción	**Biobío**	**219.057**	**39.729**	**30**	**18 (60%)**
Temuco	La Araucanía	263.165	54.893	25	10 (40%)

Fuente: Elaboración propia

Así, como decisión estratégica y por ser una de las regiones con mayor porcentaje de EE con desocupación crítica, se decide trabajar con la **Región del Biobío**, en particular con aquellas comunas que poseen mayor población y desocupación. Dada esta decisión, se debe considerar que los datos de esta comuna están disponibles desde el año 2018 y la aplicación del SAE comienza de forma parcelada, sólo con 1° básico, 7° básico y I° medio.

Cabe destacar que, a pesar de hacer una elección acotada de comunas para trabajar, hay que considerar que existen desplazamientos de estudiantes entre comunas para ir a estudiar, por lo que no se puede analizar las comunas como sistemas cerrados, sino que es imprescindible estudiar la movilidad y tomar todas las comunas que interactúen dentro del sistema. Es por ello, que, dentro de la región del Biobío, se toma la comuna de Concepción como la localidad con alto nivel de desocupación a estudiar, y luego, se procede a estudiar el mapa de desplazamiento escolar elaborado por el académico del CIAE Patricio Rodríguez, el cual toma los datos de las comunas de origen y destino del viaje de los estudiantes desde las bases de datos de matrícula del MINEDUC, para establecer flujos que relacionan comunas. Así, se concluye que existe un circuito de desplazamiento especialmente marcado entre las comunas de Concepción, Talcahuano, Hualpén, San Pedro de la Paz, Penco, Chiguayante, Coronel, Tomé, Hualqui y Lota, tal y como se puede observar en la Ilustración 3. Dado esto, es que se decide **aplicar el modelo de esta tesis en esas 10 comunas.**

Ilustración 3: Mapa de flujos de desplazamiento estudiantil



Fuente: Estudio de desplazamientos de académico Patricio Rodríguez, Instituto de Estudios Avanzados en Educación[7]

5.1.1. Niveles educativos

Es relevante destacar que, como decisión estratégica, sólo se estudiaron los datos provenientes de postulaciones de educación básica y media de niños/as y jóvenes, excluyendo del análisis a la educación parvularia y la educación para adultos.

Hay algunos niveles que poseen mayor desocupación, así como mayor número de postulaciones. Particularmente se debe pensar en aquellos cursos de entrada, es decir, aquellos en que los alumnos ingresan por primera vez al EE por ser el nivel mínimo que este ofrece: 1° básico, 7° básico y I° medio son los niveles de mayor interés.

Se espera que el modelo logre diferenciar entre niveles, y para esta tesis se utilizarán todos los niveles disponibles en la región a estudiar, que este caso (para la postulación 2018 admisión 2019) son 1° básico, 7° básico y I° medio, que son precisamente los niveles con mayor movilidad, como fue recién mencionado.

Así, finalmente la muestra será de aquellos estudiantes que postulan en los 3 niveles mencionados a EE de zona urbana, en las comunas de Concepción, Talcahuano, Hualpén,

[7] Mapa en Power BI: https://www.ie.uchile.cl/index.php?page=view_vinculacion&id=1885&langSite=es

San Pedro de la Paz, Penco, Chiguayante, Coronel, Tomé, Hualqui, Lota, en Región del Biobío.

5.2. Datos y variables del modelo

Las bases de datos que serán utilizadas para realizar el análisis y la predicción son las proporcionadas por el MINEDUC en su plataforma web[8] respecto al Sistema de Admisión Escolar (bases de datos con información de los establecimientos, postulantes, postulaciones y resultados). Además, se utilizarán las bases de datos de Directorio de Establecimientos Educacionales, que posee la información de todos los establecimientos del país (nombre, tipo, ubicación, entre otros); la de Matrícula por alumno, que posee información propia de cada estudiante que asiste a un establecimiento educacional chileno. Finalmente, también se utilizan bases de datos para algunos índices agregados al modelo, como el índice de vulnerabilidad multidimensional (IVM) de JUNAEB y el SNED, también del MINEDUC. A continuación, las bases de datos utilizadas y su fuente:

Tabla 3: Bases de datos utilizadas

Base de datos	Fuente
Oferta establecimientos etapa regular y complementaria 2019 y 2020	Centro de Estudios del MINEDUC: Sistema de Admisión Escolar (SAE)
Postulantes establecimientos etapa regular y complementaria 2019 y 2020	Centro de Estudios del MINEDUC: Sistema de Admisión Escolar (SAE)
Postulaciones etapa regular y complementaria 2019 y 2020	Centro de Estudios del MINEDUC: Sistema de Admisión Escolar (SAE)
Resultados etapa regular y complementaria 2019 y 2020	Centro de Estudios del MINEDUC: Sistema de Admisión Escolar (SAE)
Matrícula única 2018, 2019 y 2020	Centro de Estudios del MINEDUC: Matrícula por estudiante
Directorio oficial EE 2018, 2019 y 2020	Centro de Estudios del MINEDUC: Directorio de Establecimientos Educacionales
IVM Establecimientos 2018 y 2019	JUNAEB
SNED 2018 y 2019	Centro de Estudios del MINEDUC: Sistema Nacional Evaluación del Desempeño (SNED)

Fuente: Elaboración propia

A continuación, se describen las variables que se utilizarán en el modelo. Estas variables son características del proceso de postulación, tanto de los postulantes como de los EE en oferta.

[8] Página web del Centro de Estudios del MINEDUC: https://datosabiertos.mineduc.cl/

Tabla 4: Variables del modelo

Variable	Descripción	Tipo	Fuente
MRUN	Enmascaramiento del RUN del postulante	Numérica	SAE C1
PREFERENCIA_POSTULANTE	Lugar de preferencia dentro del ranking declarado en la plataforma de postulación	Numérica	SAE C1
LAT_CON_ERROR/LON_CON_ERROR	Coordenadas geográficas con error aleatorio	Numérica	SAE B1
COD_COM_ALU	Código oficial comuna de residencia (auto declarado y voluntario)	Categórica	Matrícula por alumno
ES_MUJER	Indicador de si el postulante es mujer	Categórica (binaria)	SAE B1
EDAD_ALU	Edad al 30 de junio del correspondiente año escolar	Numérica	Matrícula por alumno
COD_NIVEL	Nivel al que postula el estudiante	Categórica	SAE C1
COD_ENSE	Código de enseñanza	Categórica	SAE C1
COD_ESPE	Código de especialidad	Categórica	SAE C1
ALTO_RENDIMIENTO	Indicador si el postulante proviene del 20% superior del ranking notas	Categórica (binaria)	SAE B1
PRIORIDAD_MATRICULADO	Prioridad por estar matriculado en el establecimiento	Categórica (binaria)	SAE C1
PRIORIDAD_HIJO_FUNCIONARIO	Prioridad por tener padre y/o madre trabajando en el establecimiento	Categórica (binaria)	SAE C1
PRIORIDAD_EXALUMNO	Prioridad por ser exalumno del establecimiento	Categórica (binaria)	SAE C1
PRIORIDAD_HERMANO	Prioridad por tener un hermano matriculado en el establecimiento	Categórica (binaria)	SAE C1
NSE	Nivel socioeconómico	Categórica	-
RBD	Código del establecimiento al que postula	Numérica	SAE C1
LAT_RBD/LON_RBD	Coordenadas geográficas del establecimiento educacional	Numérica	SAE A1
COD_COM_RBD	Código oficial comuna en que se ubica el establecimiento	Categórica	Matrícula por alumno
DISTANCIA	Distancia real en kilómetros entre el mrun y el EE	Numérica	Elaboración propia
COD_DEPE2	Código que indica tipo de dependencia (Municipal, Particular Subvencionado, SLEP, etc.)	Categórica	Directorio oficial EE
TIPO_EE	Indica si el EE es sólo de básica, sólo de media HC/TP, o completo HC/TP9	Categórica	Elaboración propia
CON_COPAGO	Indica si el EE es con copago	Categórica (binaria)	SAE A1
COD_JOR	Indica si el EE tiene jornada de mañana, tarde o completa	Categórica	SAE C1
COD_SEDE	N° de sede	Numérica	SAE C1

[9] HC: humanista científico y TP: técnico profesional

ORI_RELIGIOSA	Orientación religiosa del EE	Categórica	Directorio oficial EE
RURAL_RBD	Indica si el EE está en zona rural o urbana	Categórica (binaria)	Directorio oficial EE
PAGO_MATRICULA	Valor de la matrícula del EE, por tramos	Categórica	Directorio oficial EE
PAGO_MENSUAL	Valor de la mensualidad del EE, por tramos	Categórica	Directorio oficial EE
CUPOS_TOTALES	Cupos totales disponibles para el nivel y RBD	Numérica	SAE A1
VACANTES	Número de vacantes para el sistema de admisión	Numérica	SAE A1
VACANTES_USADAS	Número de vacantes asignadas post resultados del SAE	Numérica	Elaboración propia
SNED	Índice de calidad educativa	Categórica	SNED
IVM	Índice de vulnerabilidad escolar del establecimiento educacional	Numérica	IVM JUNAEB
NSE	Nivel socioeconómico del estudiante	Numérica	CIAE[10]

Fuente: Elaboración propia en base a bases de datos del MINEDUC

Cabe destacar que hubo un trabajo de selección de variables de distintas bases de datos, a partir de la revisión bibliográfica que sustenta este trabajo de tesis. Además, el procesamiento de los datos y el estudio de las variables permitió hacer los ajustes necesarios de cada una, para poder ser utilizadas en el modelo en cuestión.

El modelo de preferencias individuales posee como variable dependiente la "`PREFERENCIA_POSTULANTE`", y como variables independientes las siguientes: `LAT_MRUN, LON_MRUN, ES_MUJER, EDAD_ALU, NSE, COD_NIVEL, ALTO_RENDIMIENTO, PRIORIDAD_MATRICULADO, PRIORIDAD_HIJO_FUNCIONARIO, PRIORIDAD_EXALUMNO, PRIORIDAD_HERMANO, LAT_RBD, LON_RBD, COD_COM_RBD, COD_DEPE2, TIPO_EE, CON_COPAGO, ORI_RELIGIOSA, PAGO_MATRICULA, PAGO_MENSUAL, SNED, IVM, DISTANCIA, CUPOS_TOTALES`. En este caso, se implementaron 3 modelos de preferencias individuales:

1. Modelo de orden: se configura como una base de datos que por cada fila indica para un estudiante el EE al que postula y la preferencia en que ubica al EE en el ranking. Por ende, hay tal cantidad de filas como estudiantes y EE postulados por cada uno. En este modelo se le entrega previamente los EE a los que postula cada estudiante, por ende, el modelo al ser entrenado y validado, sólo debe ordenar el set de EE postulados por cada estudiante. Esta es

[10] Esta base de datos fue otorgada por el profesor guía Patricio Rodríguez desde el CIAE. Posee mrun de estudiantes desde el 2005 hasta el 2019, por ende para el año 2020 en el nivel 1° básico se generan la mayor cantidad de valores nulos. Para no eliminar tantas filas, se decide inputar el promedio del NSE del EE al que asiste el estudiante antes de cambiarse como su NSE si es que el valor era nulo.

una primera prueba que no representa la realidad, pues no muestra el set total de EE a los que podría postular cada estudiante.

2. Modelo de selección y orden: se configura como una base de datos que, por cada estudiante, se simula que postulara a todos los EE factibles (es decir, que ofrezcan cupos en el nivel correspondiente, ver en anexo C.1) del territorio, lo que implica que hay tal cantidad de filas como estudiantes y EE factibles para cada uno en el territorio. En este caso, dado que se introducen filas nuevas indicando postulaciones a EE para las que el estudiante en la realidad no postuló, se indica un orden de preferencia consecutivo al último EE postulado, de acuerdo a la distancia más cercana del estudiante al EE (dado que esta es una variable relevante en el comportamiento de elección). Este modelo no solo debe ordenar la preferencia de los EE postulados por el estudiante, sino que ahora previamente debe seleccionar a qué EE postularía efectivamente un estudiante, y luego además ordenarlos según cuáles serían sus preferencias, lo cual es la situación realista.

3. Modelo de orden de cercanos: se configura como una base de datos que, por cada estudiante, se simula que postulara a los 4 EE más cercanos a su georreferenciación reportada, lo que implica que hay un n° de filas equivalente a 4 veces la cantidad de estudiantes. Este modelo se probó considerando que el modelo de selección y orden tenía un set de EE demasiado grande para seleccionar, y esta era una prueba dado que la variable distancia es relevante en el comportamiento de elección de acuerdo a la bibliografía. Así, este modelo tendría que ordenar las preferencias de esos 4 EE para cada estudiante.

Lo que se busca en todos estos modelos de preferencias individuales es predecir la preferencia con que padres y apoderados escogerán un EE determinado, sólo a partir de las características del estudiante, del EE y de la potencial postulación. Así, posterior a la predicción, se podrá armar una lista de preferencias por cada estudiante que esté en el proceso de postulación. A continuación, la lista de preferencias será ingresada al algoritmo actual del SAE, pudiendo asignar a cada estudiante a un EE, para finalmente calcular la ocupación potencial de cada EE y con ello cumplir con el objetivo de identificar aquellos territorios con desocupación crítica a futuro.

El segundo modelo probado posee como variable dependiente las "`VACANTES_USADAS`", y como variables independientes las siguientes: `COD_NIVEL`, `COD_GRADO`, `COD_ENSE`, `COD_JOR`, `COD_ESPE`, `COD_SEDE`, `LAT_RBD`, `LON_RBD`, `COD_COM_RBD`, `COD_DEPE2`, `TIPO_EE`, `CON_COPAGO`, `ORI_RELIGIOSA`, `PAGO_MATRICULA`, `PAGO_MENSUAL`, `SNED`, `IVM`, `VACANTES`. Se tiene una base de datos que por cada fila indica para una determinada combinación de EE, nivel y código de curso (con sus respectivas

características sociodemográficas), cuántas vacantes se ofrecen por el sistema SAE, y cuántas efectivamente son ocupadas según la asignación. Así, objetivo es que el modelo logre predecir esta ocupación de vacantes, lo que conlleva a entender la demanda.

Por último, es relevante mencionar que para la realización de esta tesis se probaron distintos algoritmos de machine learning: árbol de decisión (LGBM Regressor), CatBoost, XGBoost y Random Forest. Luego de validar su desempeño, se decidió implementar los modelos utilizando Random Forest, y abordando los resultados sólo desde su aplicación.

Capítulo 6: Análisis y resultados

6.1. Análisis exploratorio de datos (EDA)

Como ya se mencionó previamente, la implementación del SAE ha permitido acumular valiosos datos en torno a las características de los postulantes, las postulaciones y las vacantes. A continuación, se hacen distintos análisis exploratorios, que buscan dar valor a los datos transformándolos en información y evidencia estadística. Primeramente, se hace un análisis diagnóstico del proceso de admisión 2020. No se utilizan procesos posteriores, debido a que aún es un enigma el cómo se comportaron las mismas durante la pandemia de COVID-19 que afectó al país desde el 2020 en adelante. Luego, se hace un análisis regional respecto a la ocupación escolar, lo que permite dar resolución al primer objetivo específico de esta investigación, Por último, se hace un estudio respecto al nivel de uso del SAE en relación con cuántas de las asignaciones se convierten efectivamente en matrículas, así como a tener el dato de cuántos estudiantes se matriculan en EE que son parte del SAE, sin pasar por el sistema de admisión. Este último estudio permite dar resolución al segundo objetivo específico de esta investigación.

6.1.1. Análisis SAE admisión 2020

En esta sección se realiza un diagnóstico respecto a las cifras principales de postulantes, oferentes y postulaciones a nivel nacional. Cabe destacar que en este análisis se consideró al mismo tipo de perfil descrito en la sección anterior, es decir, sólo se estudiaron los datos provenientes de postulaciones de educación básica y media de niños/as y jóvenes, excluyendo del análisis a la educación parvularia y la educación para adultos. Además, sólo se consideraron postulaciones a EE urbanos. Lo anterior, simplemente para acotar el alcance de la investigación.

Para el año señalado, y sin hacer otros filtros adicionales a los ya mencionados, se observa que hay 1.034.992 preferencias[11], correspondientes a 290.216 postulantes y 4.751 EE participantes. Además, los EE ofrecen en total 560.957 vacantes a través del sistema, distribuidas en los distintos niveles de la siguiente forma:

[11] Recordar que cada postulante (estudiante) ingresa una postulación, que se compone de distintas "preferencias", que son una asociación estudiante-EE con un orden de preferencia.

Tabla 5: Distribución de vacantes por nivel

Nivel	Vacantes	Porcentaje
1° básico	91.667	16,34%
2° básico	32.398	5,78%
3° básico	29.842	5,32%
4° básico	27.579	4,92%
5° básico	26.584	4,74%
6° básico	25.267	4,50%
7° básico	64.502	11,50%
8° básico	32.393	5,77%
I° medio	139.310	24,83%
II° medio	29.382	5,24%
III° medio	32.234	5,75%
IV° medio	29.799	5,31%
Total	**560.957**	**100%**

Fuente: Elaboración propia

Se observa que un 24,83% de las vacantes ofrecidas son para I° medio. Le siguen 1° básico con un 16,34% y 7° básico con un 11,5% de las vacantes. El nivel donde menos se ofrecen es 6° básico con un 4,5%.

A continuación, se estudió la distribución de vacantes por región:

Tabla 6: Distribución de vacantes por región

Región	Vacantes	Porcentaje
Tarapacá	12.914	2,30%
Antofagasta	22.926	4,09%
Atacama	14.662	2,61%
Coquimbo	31.191	5,56%
Valparaíso	80.686	14,38%
Libertador Gral. Bernardo O'Higgins	37.908	6,76%
Maule	41.581	7,41%
Biobío	67.291	12,00%
La Araucanía	49.605	8,84%
Los Lagos	35.983	6,41%
Aysén	5.832	1,04%
Magallanes	6.146	1,10%
Región Metropolitana	104.386	18,61%
Los Ríos	17.531	3,13%
Arica	8.352	1,49%
Ñuble	23.963	4,27%
Total	**560.957**	**100%**

Fuente: Elaboración propia

De esta tabla se observa que las regiones con mayor cantidad de cupos son la Región Metropolitana y la Región de Valparaíso, con un 18,61% y 14,38% respectivamente. Le sigue de cerca la Región del Biobío con un 12% de vacantes. La región con menos vacantes ofrecidas fue la Región de Aysén, con 5.832 vacantes (1,04%).

Respecto a la distribución de vacantes por tipo de dependencia, el porcentaje se distribuye de la siguiente forma:

Tabla 7: Distribución de vacantes por tipo de dependencia

Tipo de dependencia	Vacantes	Porcentaje
Corporación Municipal	83.174	14,83%
Municipal DAEM	208.822	37,23%
Particular Subvencionado	237.078	42,26%
Corporación de Administración Delegada (DL 3166)	18.983	3,38%
Servicio Local de Educación	12.900	2,30%
Total	**560.957**	**100%**

REcibidoFuente: Elaboración propia

Se observa un predominio de vacantes en los EE particulares subvencionados (42,26%) y los municipales DAEM (37,23%). Cabe destacar que **no todos** los establecimientos subvencionados son parte del SAE, en particular, en promedio un 25% de los EE registrados en el MINEDUC en 2019 no participó del SAE, porcentaje que fue mayor en las regiones del norte del país (34% promedio) y la Región Metropolitana (35%).

Adicionalmente, se calculó el promedio de vacantes por EE, separando por tipos de dependencia:

Tabla 8: Promedio de vacantes por tipo de dependencia

Tipo de dependencia	Promedio de vacantes por EE
Corporación Municipal	136,35
Municipal DAEM	148,63
Particular Subvencionado	93,01
Corporación de Administración Delegada (DL 3.166)	275,12
Servicio Local de Educación	109,32

Fuente: Elaboración propia

De ello, se observa que el mayor promedio de vacantes reportada es en las Corporaciones de Administración Delegada, seguido por los Municipales DAEM.

Para simplificar el análisis en torno a los tipos de dependencia, se agrupan los tipos de establecimientos en "Público" (considerando los tipos de dependencia: Corporación Municipal, Municipal DAEM, Corporación de Administración Delegada (DL 3166) y Servicio Local de Educación), "Subvencionado con copago" (considera tipo de dependencia

particular subvencionado y con pago mensual) y "Subvencionado gratuito" (considera tipo de dependencia particular subvencionado y sin pago mensual) a través de una nueva variable. Así, la distribución de vacantes para esta variable queda:

Tabla 9: Promedio de vacantes por tipo de dependencia agrupada

Tipo de dependencia agrupada	Cantidad
Público	15.372
Subvencionado gratuito	11.614
Subvencionado con copago	6.959

Fuente: Elaboración propia

En relación con ello, se observa una distribución relativamente similar entre EE públicos y subvencionados, pero se hace la distinción de que los subvencionados con copago son quienes presentan menos vacantes a través del sistema. Para ahondar en el detalle, se puede revisar una tabla de distribución de vacantes por región y tipo de dependencia en el anexo B.1. Con esta información, y considerando que un 36% de la matrícula es pública, se infiere que existe sobreoferta en este sector, porque ofrece casi la misma cantidad de vacantes, teniendo mucha menos matrícula capturada.

Como caracterización de los EE particulares subvencionados, se observa la siguiente distribución de pago mensual:

Tabla 10: Distribución de EE según pago mensual

Pago mensual	N° de EE	Porcentaje
Sin información	17	0,36%
Gratuito	3.841	80,85%
$1.000 a $10.000	12	0,25%
$10.001 a $25.000	117	2,46%
$25.001 a $50.000	386	8,12%
$50.001 a $100.000	343	7,22%
Más de $100.000	35	0,74%
Total	**4.751**	**100%**

Fuente: Elaboración propia

Tal como fue recién visto, la mayor parte de establecimientos subvencionados son gratuitos (80,85%), y los que no lo son se concentran en el rango de precio $25.001 a $100.000 en un 15,34% de la muestra restante.

Como ya se mencionó, hay 1.034.992 de preferencias en el año estudiado, lo que, dado que hay 290.216 postulantes, indica un promedio de 3,56 preferencias por estudiante. El mínimo de preferencias que tiene un estudiante es a 1 EE, y el máximo es a 94 EE (sólo un caso). De todas formas, se realiza un análisis desagregado por nivel, para analizar el promedio de preferencias por postulación en cada uno:

Tabla 11: Promedio de preferencias por postulación

Nivel	Promedio de preferencias por postulación
1° básico	3,65
2° básico	3,61
3° básico	3,51
4° básico	3,51
5° básico	3,44
6° básico	3,46
7° básico	3,46
8° básico	3,37
I° medio	3,60
II° medio	3,41
III° medio	3,46
IV° medio	3,38

Fuente: Elaboración propia

Se observa que a medida que aumenta el nivel postulado, disminuye la cantidad de EE en las preferencias de cada estudiante, aunque no hay grandes diferencias. Así mismo, no hay grandes diferencias entre regiones (ver en anexo B.2).

Continuando con el análisis de preferencias, estas se desagregan por región, obteniendo los siguientes resultados:

Tabla 12: Preferencias por EE por región

Región	Preferencias	Porcentaje del total
Tarapacá	32.273	3,12%
Antofagasta	63.912	6,18%
Atacama	28.957	2,80%
Coquimbo	73.032	7,06%
Valparaíso	118.855	11,48%
Libertador Gral. Bernardo O'Higgins	62.068	6,00%
Maule	70.988	6,86%
Biobío	98.196	9,49%
La Araucanía	54.446	5,26%
Los Lagos	51.050	4,93%
Aysén	6.121	0,59%
Magallanes	11.209	1,08%
Región Metropolitana	301.575	29,14%
Los Ríos	20.769	2,01%
Arica	17.498	1,69%
Ñuble	24.043	2,32%
Total	**1.034.992**	**100,00%**

Fuente: Elaboración propia

De ellos, se observa que la mayor cantidad de postulaciones se encuentra en aquellas regiones con mayor población, como es de esperar. En una mayor medida, la Región Metropolitana se lleva el 29,14% de las postulaciones, seguida por la Región de Valparaíso y Biobío.

Por último, se estudia la distribución de postulaciones de primera preferencia de cada estudiante:

Tabla 13: Distribución de postulaciones de primera preferencia por tipo de dependencia

Tipo de dependencia	Cantidad postulaciones 1ª preferencia	Porcentaje
Público	99.533	35,4%
Subvencionado gratuito	94.046	33,5%
Subvencionado con copago	87.428	31,1%

Fuente: Elaboración propia

Se observa que, de las primeras preferencias de las postulaciones de cada estudiante, la mayor parte de ellas son de EE subvencionados (181.474 en total). Sólo si se desagrega entre EE subvencionados, los EE públicos serían aquellos con mayor número de postulaciones.

6.1.2. Análisis regional de ocupación escolar

A continuación, se muestra un resumen general de lo que se pudo observar a partir del estudio de ocupación escolar pública y subvencionada que se realizó para todas las regiones del país. Si bien cada región presenta características particulares y en algunas de ellas se observa una desocupación más acentuada, hay aspectos comunes que se ven a lo largo de todo el territorio.

En la mayor parte de las regiones, predomina la educación pública versus la subvencionada en términos de número de EE. Además, si se hace la diferenciación según copago, le seguiría la subvencionada gratuita sobre la de copago. Por otro lado, entendiendo que los establecimientos en Chile pueden ser sólo de básica, sólo de media (humanista científico (HC) o técnico profesional (TP)) o de ambos (básica + media HC o TP), lo que se observa en la generalidad es una fuerte presencia de establecimientos de básica, especialmente en las zonas rurales o rezagadas de las distintas regiones, pero también en los centros urbanos. Seguido a los establecimientos de básica, se encuentran los establecimientos con oferta completa (básica y media) humanista científica (HC). En menor medida se encuentra la oferta sólo de media o completa TP.

Haciendo una intersección entre el tipo de dependencia (pública o subvencionada) y el tipo de establecimiento (sólo de básica, media o completa HC/TP), en general se

identifica que la mayor parte de los establecimientos educacionales en las regiones son públicos y de educación básica, seguidos de los públicos de educación completa HC. Además, para los subvencionados gratuitos se repite la misma situación ya mencionada, y en los subvencionados con copago hay un predominio de establecimientos completos HC. Respecto a la ruralidad, en la mayor parte de las regiones hay más establecimientos urbanos que rurales, pero es variable y como es de esperar, la situación cambia a medida que se está en regiones más extremas del país, especialmente en el sur, en donde predomina la ruralidad.

Tabla 14: Ocupación por región

Región	**Tipo de dependencia**	**Q1 % de ocupación**	**Mediana % de ocupación**	**Mínimo % de ocupación**
Tarapacá	Público	52	81,4	20,0
	Subvencionado con copago	94,9	97,6	81,5
	Subvencionado gratuito	85,6	92,7	56,7
Antofagasta	Público	82,6	92,7	25,0
	Subvencionado con copago	96,7	97,6	88,3
	Subvencionado gratuito	88,8	92,7	48,2
Atacama	Público	60	76,7	30,0
	Subvencionado con copago	94	96,7	62,4
	Subvencionado gratuito	80,1	91,1	49,3
Coquimbo	Público	40	59,4	10,0
	Subvencionado con copago	95,9	97,8	20,0
	Subvencionado gratuito	41	85,9	15,8
Valparaíso	Público	56,2	72,8	16,7
	Subvencionado con copago	93,2	96,7	30,0
	Subvencionado gratuito	77,9	89,6	14,3
Libertador Gral. Bernardo O'Higgins	Público	50	66,3	10,0
	Subvencionado con copago	93,1	96,6	34,6
	Subvencionado gratuito	78,2	93,3	38,1
Maule	Público	47,5	62,8	8,6
	Subvencionado con copago	95,9	98,3	67,2
	Subvencionado gratuito	74	88,9	19,2
Biobío	Público	49	65,4	11,7
	Subvencionado con copago	95,3	97,3	70,7
	Subvencionado gratuito	66,2	85,2	16,7
La Araucanía	Público	43,8	58,8	13,3
	Subvencionado con copago	89	94,7	27,5
	Subvencionado gratuito	41,7	58,6	10,6
Los Lagos	Público	40	56,2	10,0
	Subvencionado con copago	95	97,4	79,2
	Subvencionado gratuito	48	73,4	9,3

Región	Tipo de dependencia	Q1 % de ocupación	Mediana % de ocupación	Mínimo % de ocupación
Aysén	Público	47	62,6	20,0
	Subvencionado con copago	97	97,5	96,5
	Subvencionado gratuito	80,8	87,0	53,4
Magallanes	Público	49,2	74,5	20,0
	Subvencionado con copago	94	97,5	89,5
	Subvencionado gratuito	88,6	94,3	85,7
Región Metropolitana	Público	70	85,5	20,0
	Subvencionado con copago	91	96,0	23,5
	Subvencionado gratuito	80	91,9	13,8
Los Ríos	Público	42	57,2	14,2
	Subvencionado con copago	92,6	97,9	67,5
	Subvencionado gratuito	47,7	62,6	3,0
Arica	Público	59,8	80,8	5,7
	Subvencionado con copago	97,4	97,5	97,3
	Subvencionado gratuito	85,8	91,6	59,2
Ñuble	Público	36	53,9	10,0
	Subvencionado con copago	94	98,8	78,5
	Subvencionado gratuito	70,8	85,0	33,3

Fuente: Elaboración propia

Respecto a la ocupación, y centrándose en los establecimientos urbanos, lo que se observa a grandes rasgos es que la mayor ocupación (midiéndose a través de la mediana, mínimo y primer cuartil) se da en los establecimientos subvencionados con copago, seguido de los subvencionados gratuitos, y finalmente los públicos, tal como se puede observar en la tabla 14. Lo anterior se condice con lo que señala la evidencia bibliográfica, que sugiere que padres y apoderados prefieren un establecimiento subvencionado a uno público, por lo que la educación pública sigue la lógica del "chorreo" (Canales, Bellei, & Orellana, 2016). Por otra parte, en términos de tipo de establecimiento, se observa que son los establecimientos de básica los que presentan la mayor desocupación, lo que tiene sentido al recordar que casi todos estos establecimientos son públicos. También se observa un mejor desempeño en términos de desocupación para los establecimientos de educación completa HC. Ahora, al observar con más detalle y analizar solamente los establecimientos públicos, lo que se observa al mirar por curso, es que en básica en general la tendencia es a tener mayor ocupación a medida que se aumenta de curso, y en media todo lo contrario. A pesar de lo recién mencionado, pareciera ser que los aumentos o disminuciones en ocupación no son significativos.

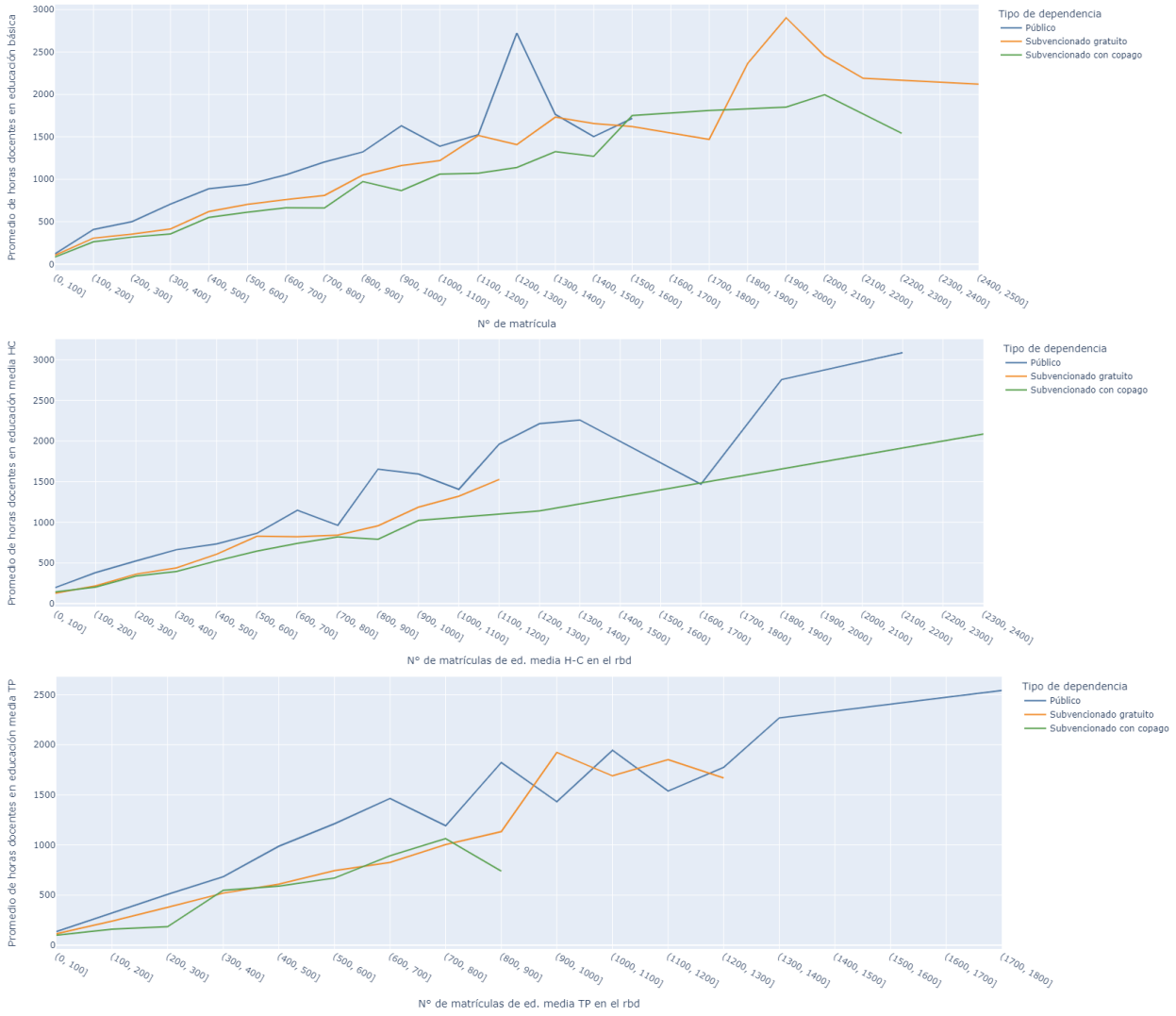
Al observar la desocupación de los establecimientos en un mapa interactivo (ver anexo C.2), lo que se concluye principalmente es que los establecimientos subvencionados con copago siempre son menos, y se encuentran principalmente en los centros urbanos de las regiones, con una desocupación muy baja, en general menor a 5%. Luego, al observar

los establecimientos subvencionados gratuitos, se ve que estos amplían su alcance ya no sólo estando en los centros urbanos, sino también en localidades urbanas más pequeñas y rezagadas, aunque manteniéndose al margen de la ruralidad. Además, se observa para este tipo de dependencia una mayor desocupación (esto observado en los centros urbanos, bajo la premisa de que en las localidades pequeñas es normal que exista desocupación alta), aunque aun así un desempeño bastante bueno con casos aislados de desocupación crítica. Ahora, en lo que respecta a la educación pública, se evidencia que esta es la que se hace cargo de las zonas rurales y rezagadas, y que, tanto en lo rural como en lo urbano, es este el tipo de dependencia con mayor desocupación, incluso en los centros urbanos.

En general parece ser que la desocupación es crítica en establecimientos públicos de básica, lo que es consecuente con lo mencionado anteriormente respecto a la baja ocupación en este tipo de establecimiento. Por otro lado, se concluye al mirar las cifras, que un 30% de desocupación es un porcentaje importante en casi todas las regiones, pues una buena parte de establecimientos públicos urbanos presenta esta cifra o más de desocupación en sus aulas. En términos de tamaño de los establecimientos, se estudió la relación entre el promedio de desocupación versus el número de matrículas disponibles en cada establecimiento (vale decir, su tamaño), y lo que se concluyó en general es que existe una tendencia a la baja en la desocupación a medida que el establecimiento es más grande.

Comentando un poco acerca de la docencia en los distintos tipos de dependencia, la evidencia del estudio realizado muestra que, a nivel país, la educación pública es la que más horas promedio de docentes en aula presenta, seguido de los establecimientos subvencionados gratuitos y por últimos los con copago. Lógicamente, una mayor cantidad de horas docentes implica una mayor inversión en este ítem, lo que no se está retribuyendo en una mayor ocupación de los establecimientos públicos, o, dicho de otra forma, en una mayor preferencia de padres y apoderados por este tipo de establecimiento. Cabe preguntarse la sostenibilidad económica de establecimientos que son los menos preferidos por padres y apoderados, 100% financiados por el Estado, y que además sobre invierten en horas docentes.

Ilustración 4: Promedio de horas docentes por intervalos de matrícula en básica y media HC y TP para todo Chile



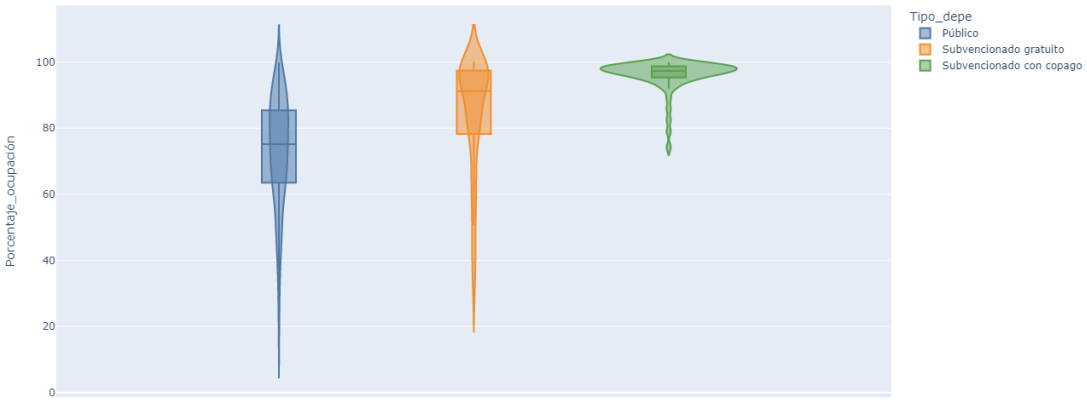
Fuente: Elaboración propia

Según lo establecido en la subsección 5.1.2, en lo que sigue, se describirán los resultados principales del análisis de ocupación escolar en la **Región del Biobío**, al ser el territorio escogido para estudiar en esta investigación.

Se estudió el porcentaje de ocupación en los distintos tipos de dependencia. Cabe destacar que se consideraron **EE urbanos**, dado que la mayor parte de EE rurales son públicos y algunos poseen una alta desocupación, lo que podría distorsionar los resultados a nivel agregado.

Se generaron gráficos de violín, que indican por un lado la dispersión de los datos como histograma, pero también las principales medidas: mínimo, máximo, mediana y cuartiles.

Ilustración 5: Porcentajes de ocupación por tipo de dependencia, Región del Biobío

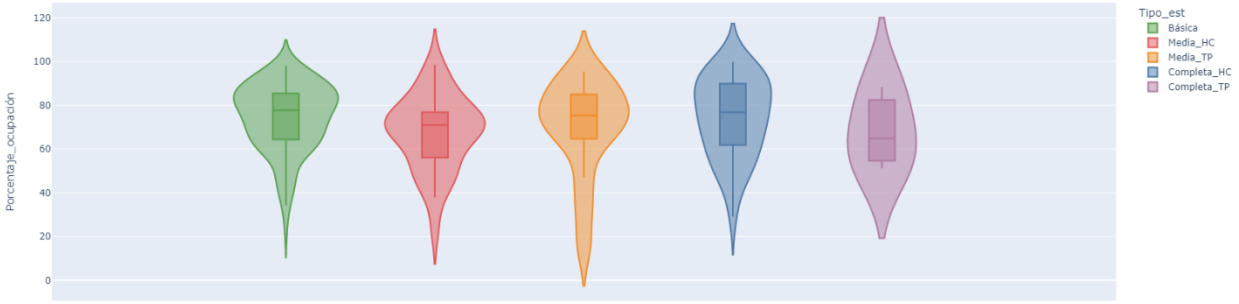


Fuente: Elaboración propia

De la ilustración 5, se observa que el tipo de dependencia con mayor porcentaje de ocupación son los subvencionados con copago, luego el subvencionado gratuito y finalmente los públicos. Además, estos últimos tienen una mayor dispersión, menor mediana (75,15%) y un mínimo de 15,9% (a diferencia del subvencionado gratuito con un mínimo de 29,7% y mediana 91,2%).

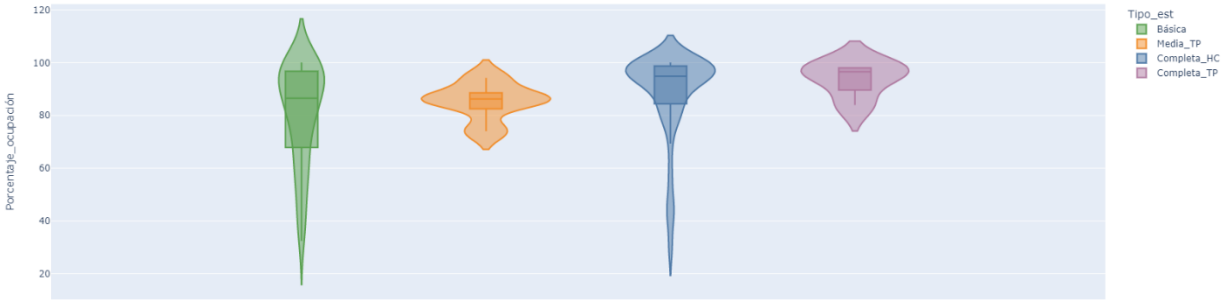
A continuación, se muestra el gráfico de porcentaje de ocupación por tipo de establecimiento.

Ilustración 6: Porcentajes de ocupación por tipo de establecimiento para EE públicos, Región del Biobío



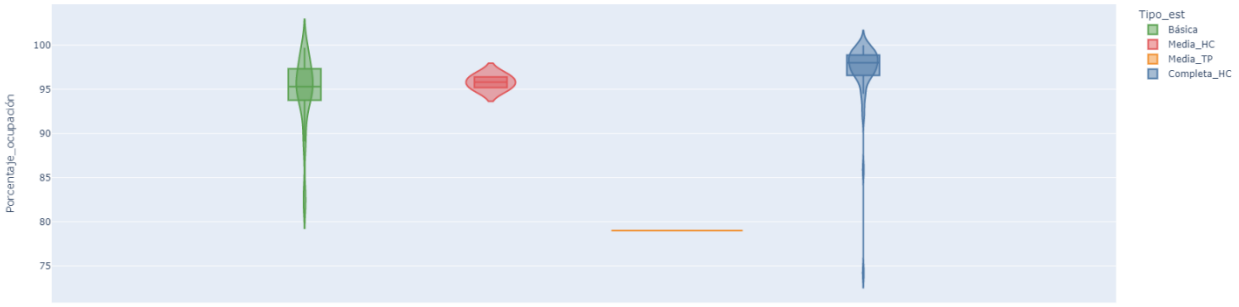
Fuente: Elaboración propia

Ilustración 7: Porcentajes de ocupación por tipo de establecimiento para EE subvencionados gratuitos, Región del Biobío



Fuente: Elaboración propia

Ilustración 8: Porcentajes de ocupación por tipo de establecimiento para EE subvencionados con copago, Región del Biobío

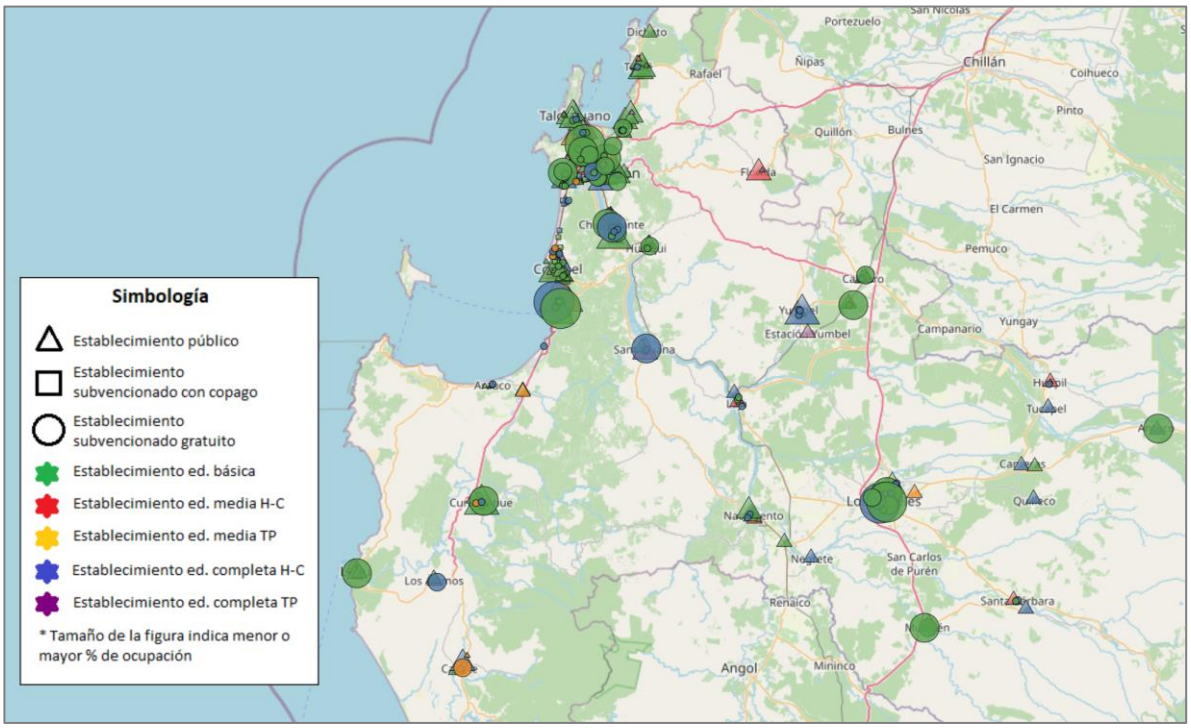


Fuente: Elaboración propia

De las ilustraciones 6, 7 y 8 se observa que el tipo de establecimiento con mejor desempeño a nivel general son los completos HC, que poseen mayor mediana y máximo. Respecto a los EE públicos, que son donde se concentra la mayor cantidad de establecimientos, se observa bastante similitud entre los tipos de establecimiento en relación con la mediana, presentándose todos los tipos establecimiento. Para los EE subvencionados gratuitos, no se observan EE de media HC y se observa un buen desempeño de los EE técnico profesionales, lo que está en realidad ligado a la baja cantidad de EE de este tipo. Sobre los EE subvencionados con copago, son los que menor dispersión y mayor ocupación presentan, pero también cabe destacar que son los que menos cantidad representan. Por último, en general para todos los tipos de dependencia, se observa que los EE de básica poseen la mayor dispersión.

Se realizó además un mapa interactivo que ilustrara la ubicación de los EE de la región presentes en el SAE, así como su tipo de dependencia, tipo de establecimiento y porcentaje de desocupación según tamaño.

Ilustración 9: Mapa interactivo de la distribución de desocupación en EE urbanos de la Región del Biobío

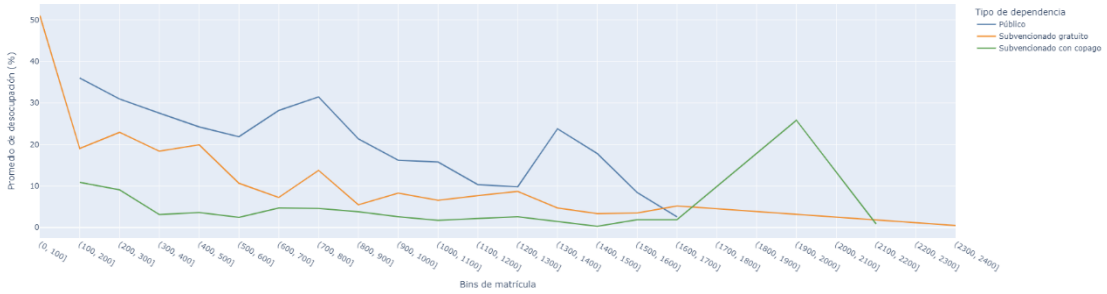


Fuente: Elaboración propia

De este mapa se observa que los EE subvencionados con copago están ubicados principalmente en Concepción y Los Ángeles, y poseen una desocupación muy baja, excepto por 2 EE con desocupación de 80% aproximadamente, ubicados en la periferia de Concepción. Sobre los EE subvencionados gratuitos, se ve que están presentes en más localidades pequeñas urbanas. Además, se nota una mayor desocupación. Hay 8 EE de Los Ángeles con desocupación mayor a 30%, la mayoría de básica y completos HC. Hay 2 EE en la misma situación en Lota, con desocupación entre 60-70%. Hay 4 EE en Chiguayante con la misma situación. Así también, hay 4 en Hualpén y 5 en Concepción con alta desocupación, todos de básica. Por último, respecto a los EE públicos, estos son los que más se encuentran en las localidades urbanas más pequeñas. A simple vista hay una mayor desocupación, y en las localidades pequeñas hay entre 2 y 3 EE públicos generalmente de básica y con una desocupación de entre 25-40%. En Los Ángeles, hay 10 EE con una desocupación mayor a 30%, de los aprox. 20 que hay (todos de básica). En Lota y Coronel hay 4 EE en la misma situación. En la zona de Talcahuano, Hualpén y Concepción hay una alta desocupación para casi todos los tipos de EE.

A continuación, se estudió el promedio de desocupación por tipo de dependencia y por número de estudiantes matriculados en el EE, para observar si existe alguna relación entre el tamaño del EE y su desocupación.

Ilustración 10: Promedio de desocupación por tipo de dependencia y por intervalos de matrícula en EE urbanos, Región del Biobío

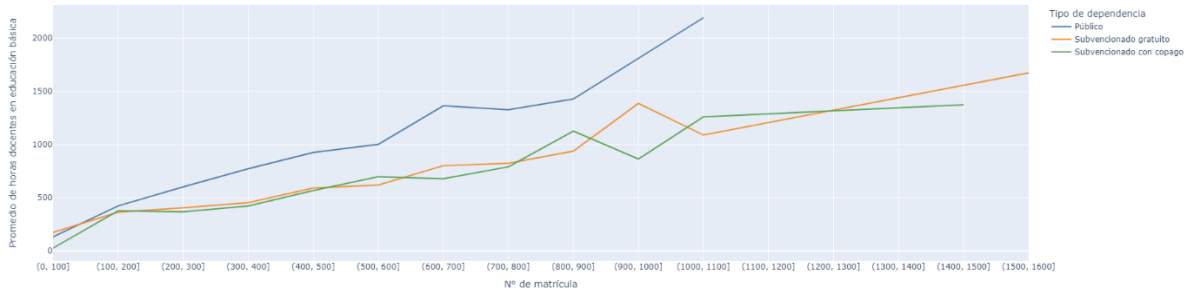


Fuente: Elaboración propia

De este gráfico de promedio de desocupación versus intervalos de tamaño de matrícula por tipo de dependencia, se observa que los EE públicos poseen un promedio de desocupación muy superior a los demás tipos de dependencia, seguido de los subvencionados gratuitos. Los subvencionados con copago tienen una desocupación muy baja en general. Además, hay que destacar que para todos los tipos de dependencia a medida que crece el EE, disminuye la desocupación promedio.

Por último, como insumo para entender las dimensiones de inversión de cada tipo de establecimiento y cómo afecta la desocupación escolar dado ello, es que se estudió a modo general el promedio de horas docentes en aula según tipo de dependencia, para educación básica, educación media HC y educación media TP.

Ilustración 11: Promedio de horas docentes por n° de matrículas en EE educación básica urbanos, Región del Biobío



Fuente: Elaboración propia

Lo que se observa de este gráfico de cantidad promedio de horas docentes en aula para educación básica (ilustración 11), es que los EE subvencionados gratuitos llegan a tener mayor número de estudiantes (públicos llegan hasta el intervalo de matrícula 1.000-1.100 mientras que los subvencionados gratuitos llegan a 1.500-1.600. Por su lado, los con copago llegan a 1.400-1.500). Sobre los EE públicos, estos demuestran una cantidad de horas promedio superior que los demás tipos de dependencia para todos los intervalos de tamaño de matrícula de matrícula, mientras que los subvencionados tanto gratuitos como

con copago poseen una cantidad de horas casi idéntica para todos los intervalos de matrícula.

Ilustración 12: Promedio de horas docentes por n° de matrículas en EE educación media HC urbanos, Región del Biobío

Fuente: Elaboración propia

Lo que se observa de este gráfico de cantidad promedio de horas docentes en aula para educación media HC (ilustración 12), es la misma situación que para básica, a excepción de que acá los EE públicos presentan los intervalos de tamaño de matrícula más altos (1.300-1.400), mientras que ambos subvencionados llegan a 700-800.

Ilustración 13: Promedio de horas docentes por n° de matrículas en educación media TP

Fuente: Elaboración propia

La ilustración 13 muestra que para educación media TP, hay pocos EE en la región, especialmente subvencionados con copago. Para todos los intervalos de matrícula donde hay presencia de todos los tipos de dependencia, los EE públicos superan levemente la cantidad promedio de horas docentes independiente del número de estudiantes, mientras que los EE subvencionados poseen un promedio de horas casi idéntico.

6.1.3. Análisis del nivel de uso del Sistema de Admisión Escolar (SAE)

En esta sección, se verifican cuántas asignaciones del SAE se traducen efectivamente en matrículas, en relación con lo propuesto en el segundo objetivo específico de esta tesis. Lo anterior, pues si bien el SAE genera una asignación a través del sistema que puede ser aceptada por las familias, esto no necesariamente se traduce en una matrícula, pues de hecho el sistema no está integrado con los sistemas de admisión de los establecimientos, sino que sólo otorga un código que luego debe ser presentado por cada estudiante en el establecimiento asignado. Por lo tanto, no necesariamente las familias siguen estos pasos, y por ende la verdadera distribución de la demanda no se puede establecer sólo con las asignaciones que entrega el SAE. Cabe destacar que este es un estudio a nivel general de las postulaciones y asignaciones, es decir, se estudia todo el país en todos los niveles educativos.

Se comienza replicando el sistema de asignación del SAE a partir de los resultados de este, generando una base de datos con cada estudiante, el EE asignado y su respectivo código de curso[12]. Para ello, se siguieron los siguientes pasos:

1. Se utilizaron las bases D1 y D2 del SAE (ver Tabla 3), correspondientes a los resultados del proceso en su etapa regular y complementaria, respectivamente.

2. Se toma primeramente los resultados de la base de datos D1 de la etapa regular (ver Ilustración 1), en que hay una columna que indica la respuesta del postulante luego de la asignación de etapa regular. Las respuestas pueden ser: 1: Acepta asignación; 2: Acepta asignación y espera por si corre la lista de espera; 3: Rechaza asignación; 4: Rechaza asignación y espera por si corre la lista de espera; 5: Sin respuesta; 6: Obligado a esperar lista de espera, pues no tiene asignación; 7: Sale del proceso. Así, en base a ello, se dejan aquellas filas en que el postulante indicó 1 o 5, pues en este último caso el sistema automáticamente acepta la asignación si es que el postulante no responde.

3. El sistema hace los ajustes de aceptación de vacantes por parte de las familias, y según las restricciones de capacidad, corre la lista de espera y se genera una reasignación para todos los alumnos que no seleccionaron las 2 alternativas previamente mencionadas (ver Ilustración 1). En esta oportunidad, los postulantes

[12] El código de curso es una tupla de variables características del establecimiento: `COD_GRADO`, `COD_ENSE`, `COD_ESPE`, `COD_SEDE` (número de sede), `COD_GENERO` y `COD_JOR`, todas mencionadas en la sección de variables del modelo.

pueden responder a la asignación de la siguiente forma[13]: 1: Acepta asignación; 3: Rechaza asignación; 5: Sin respuesta; 6: Sin asignación. Nuevamente, se deja aquellas filas en que el postulante indicó 1 o 5.

4. Habiendo trabajado ya con las opciones factibles de asignación en la etapa regular, se pasa a las asignaciones de la etapa complementaria, cuyos resultados se expresan en la base de datos D2. Esta base de datos no posee una respuesta por parte del postulante, pues simplemente genera una asignación para cada postulante en esta etapa. Recordar que en esta etapa hay estudiantes provenientes de la etapa complementaria y también hay estudiantes nuevos.

5. Finalmente, se consolidan las bases de datos resultantes de los pasos anteriores, generando la base de datos final con la asignación del SAE para cada postulante según todas sus etapas.

6.1.3.1. Admisión 2019

La base de postulantes del SAE admisión 2019 indica que hay 274.990 postulantes en la etapa regular y 46.698 postulantes en la etapa complementaria. En estas bases hay estudiantes repetidos, y al unirlas y eliminar duplicados quedan **294.768 postulantes**. Además, la base de datos recién construida en los 5 pasos previamente descritos muestra que en ese año hay **279.487 estudiantes asignados** a EE. La diferencia entre estos números es porque hay estudiantes que rechazan la asignación y no continúan luego en la etapa complementaria.

Para estudiar cuántos de los estudiantes asignados por el SAE realmente se matriculan en los respectivos EE, se toma la base de matrículas del año 2019, y se hace una comparación a nivel de estudiante - EE con la base de datos creada en los 5 pasos anteriores que indica la asignación del SAE (en este caso SAE admisión 2019), observando si hay coincidencia entre la asignación del estudiante a un EE para el año 2019 y el EE en el que está matriculado en el ese año. Al hacer esta comparación, se observa que hay **216.318 estudiantes que efectivamente se matricularon en el EE que les asignó el SAE** (según la base de datos creada). Dado ello, implicaría que un **77%** de los estudiantes se matricula en el EE asignado por el SAE (216.318/279.487). Lo anterior es relevante de estudiar, pues hay un supuesto muy grande relacionado a que los datos que aporta el SAE en términos de asignación de estudiantes a EE permitirían entender la movilidad estudiantil y es necesario entender la representatividad de ello respecto al total de elecciones, lo que impactará en el error de las predicciones. Acá se está observando que hay personas que no siguen la asignación.

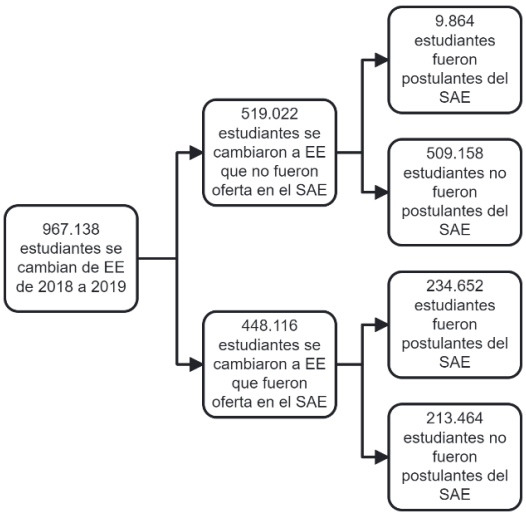
[13] Las opciones 2 y 4 de la etapa regular no existen porque ya corrió la lista de espera.

Se estudiaron cambios de EE de estudiantes entre el año 2018 y 2019, ilustrando cifras en torno a si los estudiantes fueron o no postulantes del SAE y si se cambiaron o no a un EE que fue oferta del SAE en ese año. Para ello, se compararon las bases de matrícula 2018 y 2019, identificando que **967.138** estudiantes se matricularon por primera vez o aparecieron en un EE distinto al que estaban en 2018 (ver Ilustración 14)[14]. Esta cifra incluye EE y estudiantes que no son parte del SAE, pues las bases de datos de matrícula son a nivel nacional, de todo tipo de establecimientos, incluso privados. Así, y para efectos de acotar a aquellos estudiantes que sí debieron usar el SAE, se identificó y eliminó a aquellos estudiantes que se cambiaron o ingresaron por primera vez a EE que no son parte de la oferta del SAE, resultando **448.116** estudiantes, por lo que habría **519.022** estudiantes que se cambiaron a otros EE que no fueron del sistema. Sobre los cambios a EE del SAE, se observó que **234.652** estudiantes efectivamente fueron postulantes del SAE[15], mientras que **213.464** fueron estudiantes que a pesar de matricularse en EE que fueron parte de la oferta del SAE, jamás fueron postulantes del sistema. Sobre los cambios a EE que no fueron del SAE, se observa que hay **9.864** estudiantes que fueron postulantes del SAE, es decir, desertaron del sistema en algún momento o quedaron sin asignación y optaron por EE fuera del SAE. Así mismo, hay **509.158** estudiantes que no fueron postulantes del SAE, y se matricularon en un EE por fuera del sistema (aquí se encuentran los EE privados, por ejemplo).

[14] La idea de calcular este número en bruto a nivel nacional, es tener una noción de cuántos estudiantes en general se cambian de EE, para luego ir acotando a lo que ocurre vía SAE, complementado además con lo estudiado en la sección 6.1.1 sobre las principales cifras del SAE.

[15] La diferencia de este número con los postulantes totales, que son 294.768, es porque el estudiante o se quedó en el mismo EE en el que estaba antes, o porque se salió del sistema SAE al final del proceso y se fue a un EE que no era parte de la oferta del SAE.

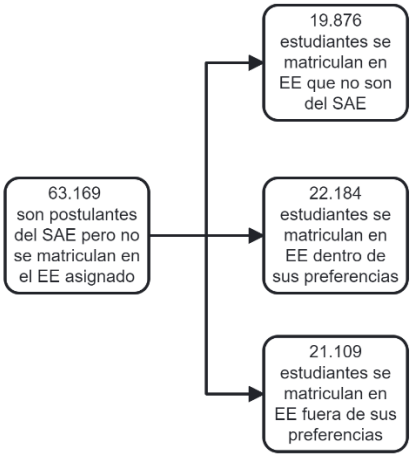
Ilustración 14: Esquema de análisis uso del SAE admisión 2019



Fuente: Elaboración propia

Se consideró particularmente interesante entender qué ocurre con aquellos estudiantes que, siendo postulantes del SAE, no se matriculan en el EE asignado por el sistema (ver Ilustración 15). Así, para el año 2019 se observa que hay 63.169 estudiantes en estas circunstancias, de los cuales 19.876 (31,46%) se matricularon finalmente en EE que no fueron parte de la oferta del SAE; 22.184 (35,12%) estudiantes se matricularon en EE que fueron parte de la oferta del SAE y también de su lista de preferencias al postular; y 21.109 (33,42%) estudiantes se matricularon en EE que fueron parte de la oferta del SAE, más no de su lista de preferencias al postular.

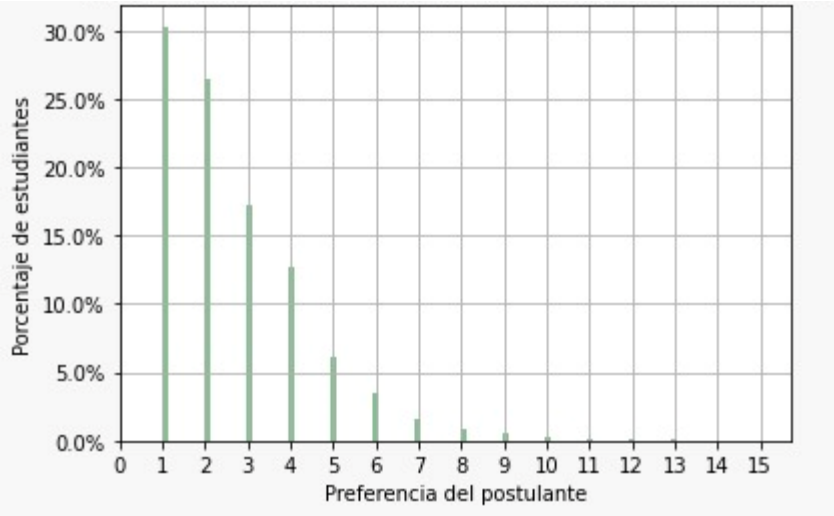
Ilustración 15: Distribución de postulantes del SAE admisión 2019 que no se matriculan en EE asignado



Fuente: Elaboración propia

Por último, a modo de estudiar con mayor profundidad el tema, se analiza en cuál de las preferencias de la lista está el EE al que se matriculan aquellos estudiantes que no siguen la asignación del SAE, lo que se puede observar en la Ilustración 16.

Ilustración 16: Distribución de preferencias para estudiantes que no siguen la asignación del SAE admisión 2019



Fuente: Elaboración propia

De este gráfico, se puede observar que aproximadamente un 30% de los estudiantes que desobedecen la asignación del SAE, finalmente logran matricularse en su primera preferencia por fuera del sistema. En general, se observa que 73% de estos estudiantes logra matricularse en sus primeras 3 preferencias, a pesar de probablemente haber quedado por sistema en una posterior, que no deseaban.

6.1.3.1. Admisión 2020

La base de postulantes del SAE admisión 2020 indica que hay **483.070** postulantes en la etapa regular y **87.604** postulantes en la etapa complementaria. En estas bases hay estudiantes repetidos, y al unirlas y eliminar duplicados quedan **526.686 postulantes**. Además, la base de datos recién construida en los 5 pasos previamente descritos muestra que en ese año hay **492.317 estudiantes asignados** a EE. La diferencia entre estos números es porque hay estudiantes que rechazan la asignación y no continúan luego en la etapa complementaria.

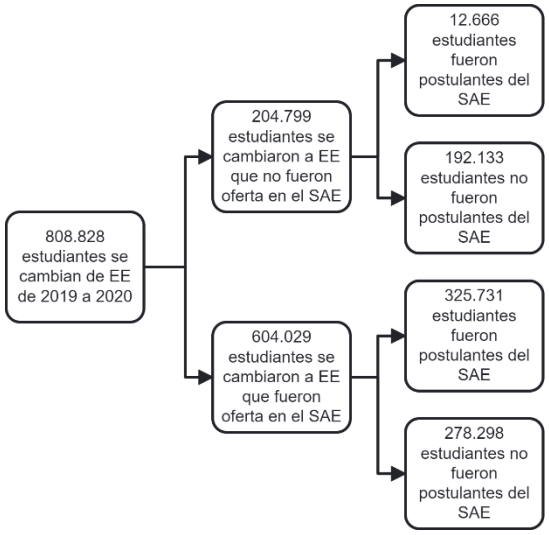
Para estudiar cuántos de los estudiantes asignados por el SAE realmente se matriculan en los respectivos EE, se toma la base de matrículas del año 2020, y se hace una comparación a nivel de estudiante - EE con la base de datos creada en los 5 pasos anteriores que indica la asignación del SAE (en este caso SAE admisión 2020), observando si hay coincidencia entre la asignación del estudiante a un EE para el año 2020 y el EE en

el que está matriculado en el ese año. Al hacer esta comparación, se observa que hay **325.731 estudiantes que efectivamente se matricularon en el EE que les asignó el SAE** (según la base de datos creada). Dado ello, implicaría que un **66%** de los estudiantes se matricula en el EE asignado por el SAE (325.731/492.317), lo que significa una disminución del 11% respecto al año anterior. Nuevamente, lo anterior es relevante de estudiar, por las razones previamente mencionadas en que al haber personas que desobedecen la asignación, no se puede estudiar la misma como una vacante utilizada al momento de calcular demanda.

Se estudiaron cambios de EE de estudiantes entre el año 2019 y 2020, ilustrando cifras en torno a si los estudiantes fueron o no postulantes del SAE y si se cambiaron o no a un EE que fue oferta del SAE en ese año. Para ello, se compararon las bases de matrícula 2019 y 2020, identificando que **808.828** estudiantes se matricularon por primera vez o aparecieron en un EE distinto al que estaban en 2019 (ver Ilustración 17). Esta cifra incluye EE y estudiantes que no son parte del SAE, pues las bases de datos de matrícula son a nivel nacional, de todo tipo de establecimientos, incluso privados. Así, y para efectos de acotar a aquellos estudiantes que sí debieron usar el SAE, se identificó y eliminó a aquellos estudiantes que se cambiaron o ingresaron por primera vez a EE que no son parte de la oferta del SAE, resultando **604.029** estudiantes, por lo que habría **204.799** estudiantes que se cambiaron a otros EE que no fueron del sistema. Sobre los cambios a EE del SAE, se observó que **325.731** estudiantes efectivamente fueron postulantes del SAE[16], mientras que **278.298** fueron estudiantes que a pesar de matricularse en EE que fueron parte de la oferta del SAE, jamás fueron postulantes del sistema. Sobre los cambios a EE que no fueron del SAE, se observa que hay **12.666** estudiantes que fueron postulantes del SAE, es decir, desertaron del sistema en algún momento o quedaron sin asignación y optaron por EE fuera del SAE. Así mismo, hay **192.133** estudiantes que no fueron postulantes del SAE, y se matricularon en un EE por fuera del sistema (aquí se encuentran los EE privados, por ejemplo). Respecto al año anterior, se observa un aumento de las postulaciones vía SAE, lo que se condice con que en este cambio de año hay más EE disponibles en oferta del SAE versus los que poseen sistemas de selección propios.

[16] La diferencia de este número con los postulantes totales, que son 526.686, es porque el estudiante o se quedó en el mismo EE en el que estaba antes, o porque se salió del sistema SAE al final del proceso y se fue a un EE que no era parte de la oferta del SAE.

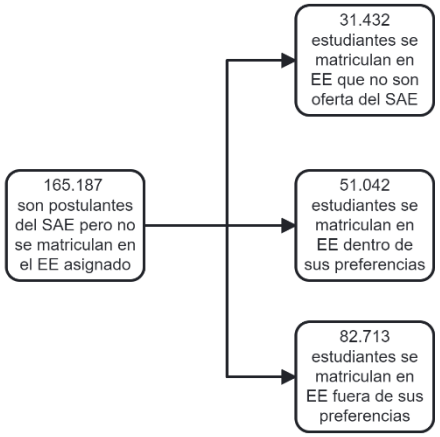
Ilustración 17: Esquema de análisis uso del SAE admisión 2020



Fuente: Elaboración propia

Se consideró particularmente interesante entender qué ocurre con aquellos estudiantes que, siendo postulantes del SAE, no se matriculan en el EE asignado por el sistema (ver Ilustración 18). Así, para el año 2020 se observa que hay 165.187 estudiantes en estas circunstancias, de los cuales 31.432 (19,03%) se matricularon finalmente en EE que no fueron parte de la oferta del SAE; 51.042 (30,9%) estudiantes se matricularon en EE que fueron parte de la oferta del SAE y también de su lista de preferencias al postular; y 82.713 (50,07%) estudiantes se matricularon en EE que fueron parte de la oferta del SAE, más no de su lista de preferencias al postular. Respecto al año anterior, se observa que hay un aumento de estudiantes que se matriculan en EE del SAE que están fuera de sus preferencias en la postulación, mientras que disminuyen los estudiantes que se matriculan en EE que no son oferta del SAE. Lo anterior se condice con el aumento de EE que son oferta del SAE para este año.

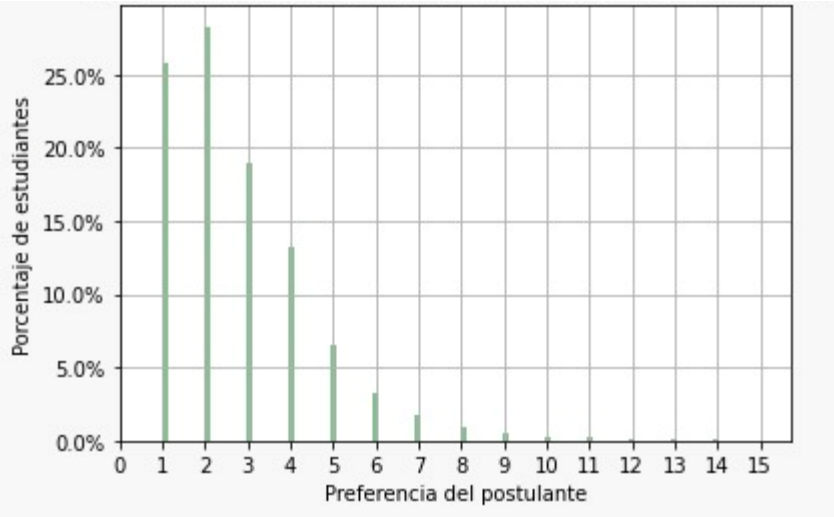
Ilustración 18: Distribución de postulantes del SAE admisión 2020 que no se matriculan en EE asignado



Fuente: Elaboración propia

Por último, a modo de estudiar con mayor profundidad el tema, se analiza en cuál de las preferencias de la lista está el EE al que se matriculan aquellos estudiantes que no siguen la asignación del SAE, lo que se puede observar en la Ilustración 19.

Ilustración 19: Distribución de preferencias para estudiantes que no siguen la asignación del SAE admisión 2020



Fuente: Elaboración propia

De este gráfico, se puede observar que aproximadamente un 26% de los estudiantes que desobedecen la asignación del SAE, finalmente logran matricularse en su primera preferencia por fuera del sistema. En general, se observa que 74% de estos estudiantes logra matricularse en sus primeras 3 preferencias, a pesar de probablemente haber quedado por sistema en una posterior, que no deseaban.

Por lo tanto, estos estudios generales sobre el uso del SAE para efectos de cambios de establecimiento o primer ingreso, permiten concluir que la adherencia al sistema aún es baja, pues incluso a pesar de que un estudiante pase por el sistema, esto no necesariamente se traduce en una matrícula efectiva en el EE asignado, y por ende el utilizar las preferencias declaradas en el SAE para caracterizar los atributos de la oferta que busca la demanda, conducirá necesariamente una subestimación que es necesario considerar en los análisis.

6.2. Exploración de predicción de la demanda por establecimiento educacional usando preferencias del SAE

6.2.1. Limpieza de datos

Dado que el estudio se centra en la Región del Biobío, se filtran las bases de datos utilizadas (la de admisión 2019 y 2020). A continuación, se describirá el **proceso para la base de datos de admisión 2020**, a modo de ilustrar los cambios realizados, que son análogos para el año anterior.

Existen 1.034.992 preferencias a EE de básica y media para niños, niñas y jóvenes urbanos. Se excluyen de este análisis la educación parvularia y los establecimientos educacionales de adultos, según las decisiones estratégicas definidas en el capítulo 5.

Acotando a las comunas escogidas de la Región del Biobío, existen 67.092 preferencias. Luego, se observa que hay 805 preferencias de 406 estudiantes en que la distancia del estudiante al EE postulado es mayor a 50 km. Se eliminan estas filas porque estas elecciones se consideran valores atípicos. Así, quedan 66.287 preferencias.

Es importante destacar que en la postulación admisión 2019, la Región del Biobío implementó el SAE solamente para los niveles 1° básico, 7° básico y I° medio, y luego en la admisión 2020 se implementa el sistema para todos los niveles. Así, a modo de estudiar bajo las mismas condiciones, es que se consideran solamente estos 3 niveles en la base de datos, para tener un mismo objeto de estudio. Al hacer este filtro, quedan 39.136 preferencias para admisión 2020.

Por otro lado, teniendo lo anterior, se observa el porcentaje de valores nulos en las variables escogidas (ver sección 5.2). Un 3,68% de los datos poseen valores nulos en la variable Nivel Socioeconómico, un 2,06% de los datos poseen valores nulos en la columna Índice de Vulnerabilidad Multidimensional y un 1,81% de los datos poseen valores nulos en la edad del alumno. Así, se decide eliminar estas filas, quedando 36.634 preferencias.

Por último, se utiliza la georreferenciación que aporta la base de postulantes del SAE. La calidad de dicha georreferenciación se reporta en una variable llamada

CALIDAD_GEOREF, que indica cuán certeros son los datos de latitud y longitud[17] de los estudiantes, de acuerdo con las siguientes categorías:

1. Respuesta única de la Google Geocoding API, calidad "rooftop" o "range_interpolated" en la variable "location_type".
2. Respuesta única de la Google Geocoding API, calidad "geometric_center" en la variable "location_type".
3. Múltiples respuestas coherentes de la Google Geocoding API, se imputó el centro de las respuestas.
4. Se imputaron las coordenadas de la municipalidad.
5. El usuario compartió su ubicación.

Dado que se necesita calcular la distancia entre la dirección del estudiante y el EE (variable que según la literatura es clave para las familias, y por ende para la predicción), se decidió dejar sólo aquellas preferencias con `CALIDAD_GEOREF` de valores 1, 2 o 5, reduciendo la muestra a 30.284 preferencias.

Por ende, la caracterización de la muestra con la que se validarán los modelos[18] corresponde a 30.284 preferencias de 10 comunas de la Región del Biobío, que involucran a 8.116 estudiantes y 286 establecimientos educacionales participantes. Lo anterior, indica que en promedio hay 3,73 postulaciones por estudiante. De forma más precisa, al analizar los datos se observa que para 1° básico hay 3,56 preferencias en promedio, para 7° básico hay 3,57 preferencias en promedio y en I° medio hay 3,84 preferencias en promedio. Es relevante además destacar, que con esta muestra quedan 8.700 preferencias para 1° básico, 3.620 preferencias en 7° básico y 18.760 preferencias para I° medio.

A continuación, se describen los 2 modelos utilizados para entrenar y validar los datos, lo que permite luego predecir las preferencias de los estudiantes y la cantidad de vacantes utilizadas, respectivamente. Cabe destacar que la mayor parte de la tesis se trabajó en el modelo de preferencias individuales, que era la idea inicial que luego daría paso a una simulación del SAE con los datos predichos, pero dado que el modelo no dio los resultados esperados, se generó otro modelo como posible camino para resolver el problema de predicción de demanda, en concordancia con el tercer objetivo específico de esta investigación.

[17] Recordar que la georreferenciación del estudiante además posee un error aleatorio de hasta 200 metros para resguardar su ubicación real.

[18] Recordando que el modelo se entrena con los datos de la admisión 2019 y se validan con la admisión 2020

6.2.2. Modelo de predicción a nivel de preferencias individuales

Se utiliza el algoritmo *Random Forest* para predecir las preferencias de los estudiantes de manera individual, es decir, para lograr predecir a qué EE postulará un estudiante y en qué orden (tal como funciona el SAE, en que se prioriza una lista de EE a postular, como ya ha sido explicado previamente), lo cual de funcionar correctamente, se utilizaría como insumo para el algoritmo de asignación del SAE, permitiendo simular las postulaciones y con ello calcular la demanda, identificando así territorios con capacidad ociosa en sus EE, que es el fin último de este trabajo.

Primeramente, es clave destacar que la base de datos está construida de tal forma que se presentan los estudiantes (con sus vectores de características) asociados a los EE a postular (con su vector de características), y que la variable a predecir es la preferencia del postulante, es decir, en qué lugar de la lista posiciona al establecimiento el estudiante (recordar que los modelos fueron descritos ampliamente en la sección 5.2).

6.2.2.1. Resultado del modelo de orden

Se realiza un primer estudio para probar la efectividad del modelo propuesto para ordenar los establecimientos según la preferencia que el postulante le da a los EE en su lista de preferencias, es decir, dándole previamente al modelo los EE a los que postula cada estudiante, y que el modelo los ordene según sus preferencias. Este es un primer acercamiento, que lógicamente no se asemeja a la realidad de lo que debería llegar a hacer el modelo, que no sólo es priorizar un EE por sobre otro, sino previo a ello escoger a cuáles EE postularía el estudiante, para luego ordenarlos.

Para los fines de este estudio, se utilizó la base de datos de postulaciones del SAE de admisión 2019 (es decir, en *t*) para entrenar el modelo, y la base de datos de postulaciones del SAE de admisión 2020 (es decir, en *t+1*), para validar el modelo. Como variables dependientes, están todas las mencionadas en la sección 5.2.1, y como variable independiente está la preferencia del estudiante (el orden asignado según la preferencia).

Primeramente, se entrena un regresor *Random Forest*, con 10 árboles de decisión e hiperparámetros por defecto de la biblioteca. En este modelo, se observa que las variables más relevantes son la distancia, el Nivel Socioeconómico (NSE), el Índice de Vulnerabilidad Multidimensional (IVM) y el SNED (ver anexo D.1). Para mejorar el desempeño del modelo, se hace un *grid search* (o búsqueda de subconjunto de hiperparámetros) para optimizar los hiperparámetros del mismo. Así, se buscan combinaciones de n_estimators,

max_features y max_depth[19] que generen el mayor R^2, siendo estos valores 150, 9, y 10, respectivamente. A continuación, se genera un nuevo modelo usando los hiperparámetros. La importancia de cada variable en este nuevo modelo es bastante similar al previo, pero cambiando el SNED por la variable cupos totales y subiendo en su importancia (ver anexo D.2).

Al aplicar este modelo entrenándolo en un año y validando en el siguiente (calculando la diferencia entre orden de la preferencia en la lista real vs la predicha[20]), se observa un error cuadrático medio (MSE) de 3,34, un RMSE de 1,83 y un error absoluto medio (MAE) de 1,37. Para tener en cuenta la dimensión de estos errores, recordar que en promedio cada estudiante postula a 4 EE.

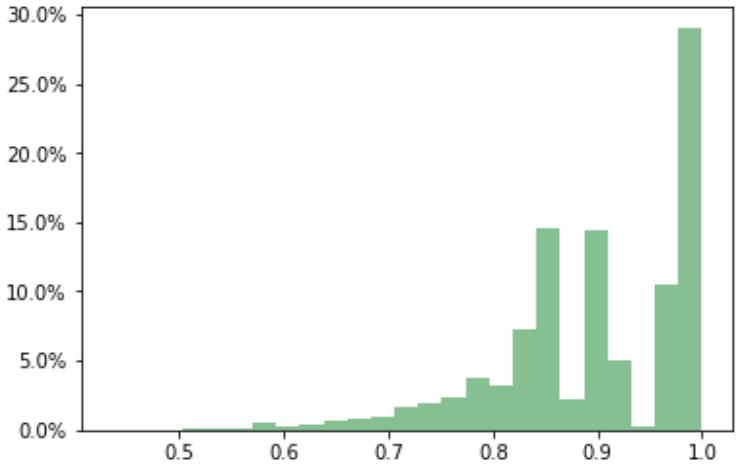
Luego, con las predicciones de este modelo para el año *t+1* se reconstruyó la lista de preferencias priorizada de cada estudiante, y se midió la capacidad de ordenar cada lista en comparación a la original (desconocida para el modelo, puesto que se entrenó con preferencias del año *t*). Nuevamente, es importante recordar que este modelo se le está entregando previamente la lista de EE a la que postula cada estudiante, y por ende la tarea que tiene es netamente ordenar la lista, más no hacer una selección de EE entre todos los disponibles, que es la forma realista y que se espera desarrollar a continuación.

El RBO (ver sección 4.4.1) promedio de 0,89, lo cual a priori es muy bueno considerando que esta métrica fluctúa entre 0 y 1, donde 1 es correspondencia total. Sólo un 14% de los datos poseen un RBO menor a 0,8 y no se observaron diferencias significativas de esta métrica entre niveles. A continuación, se muestra cómo se distribuye la frecuencia del RBO en las postulaciones.

[19] Para más información sobre los parámetros: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html

[20] Recordar que el modelo no arroja valores enteros dado que es un regresor continuo, por ende no necesariamente arroja en primera preferencia un valor de 1, sino un valor superior que al ordenarlo quedaría en primera preferencia, por ejemplo.

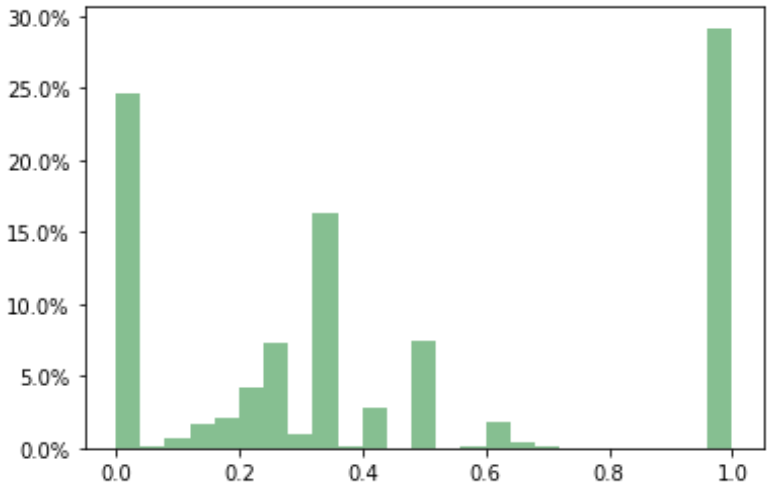
Ilustración 20: Distribución de la métrica RBO en modelo de orden



Fuente: Elaboración propia

Además de esta métrica, en este modelo también se usa el *Match_porcentual*, métrica construida para esta investigación a partir del cálculo entre cuántos EE estaban en el orden correcto versus la cantidad total de EE en la lista. Esta métrica permite tener una medida más concreta y de interpretación más sencilla de los resultados de este modelo, teniendo resultados que van entre 0 y 1. Su cálculo arrojó un promedio de 0,45, sin diferencias significativas entre niveles. En específico, se observa la siguiente distribución de resultados:

Ilustración 21: Distribución de la métrica Match promedio en modelo de orden



Fuente: Elaboración propia

De este último resultado, se concluye que el modelo tiene un bajo desempeño, porque en aproximadamente un 25% de las preferencias el modelo no es capaz de priorizar bien ninguno de los EE, siendo exitoso solo en aproximadamente un 30% de las preferencias.

El resto de los datos se distribuye en valores de *Match porcentual* entre 0,2 y 0,5 en su mayoría. Con estos resultados, también se podría decir que la métrica RBO está sobrevalorando los resultados pues se pensaría que con un valor promedio de 0,89 el modelo está siendo muy preciso en el ordenamiento, pero este valor refleja principalmente la ponderación acentuada en que el modelo le acierte a los primeros valores de la lista, lo cual es valioso pues demuestras una buena predicción para las primeras prioridades del estudiante, donde idealmente el sistema intentará asignarlo.

6.2.2.1. Resultados de modelo de selección y orden

Considerando que el modelo anterior fue sólo un primer acercamiento a ordenar los EE a los que cada estudiante postula, el siguiente paso es generar un modelo que seleccione los EE a postular dentro de todos los EE factibles[21] de postular ofrecidos por el sistema en la región y luego ordenarlos según la prioridad del estudiante. Sólo de esta forma, el modelo estará preparado para predecir en un escenario realista.

Al igual que en la sección anterior, se utilizó la base de datos de postulaciones del SAE admisión 2019 (es decir, en *t*) para entrenar el modelo, y la base de datos de postulaciones del SAE admisión 2020 (es decir, en *t+1*), para validar el modelo. Como variables dependientes, están todas las mencionadas en la sección 5.2.1, y como variable independiente está la preferencia del estudiante.

En particular, para ambas admisiones se construye una muestra en que cada estudiante postula a todos los establecimientos factibles del territorio, nominando la variable `PREFERENCIA_POSTULANTE` de forma sucesiva según cercanía de distancia del EE con el postulante para aquellos EE donde los estudiantes en la realidad no postularon (por ejemplo, si un estudiante postuló a 3 EE, la variable `PREFERENCIA_POSTULANTE` tendrá valor 1, 2 y 3 respectivamente según su prioridad, pero para todos los otros EE disponibles, a esta variable se le asignará los valores 4, 5, 6, y así sucesivamente; del EE más cercano al más lejano en distancia del postulante). Lo anterior, permite que el modelo sea entrenado no sólo con los EE reales a los que postula cada estudiante, sino con todo el universo de EE factibles (ver en anexo C.1). Cabe destacar que luego de predecir, se trunca la lista de preferencias de cada estudiante para que queden los primeros 4 establecimientos, haciendo referencia al promedio de postulaciones por estudiante y abogando por un análisis de listas que no considerara todo el universo, sino cuántos establecimientos de forma realista estaría postulando una familia o estudiante.

[21] Factible se refiere a que el EE imparta el nivel al que el estudiante postula. Por ejemplo, los liceos no son EE factibles para un estudiante que postula a 1° básico.

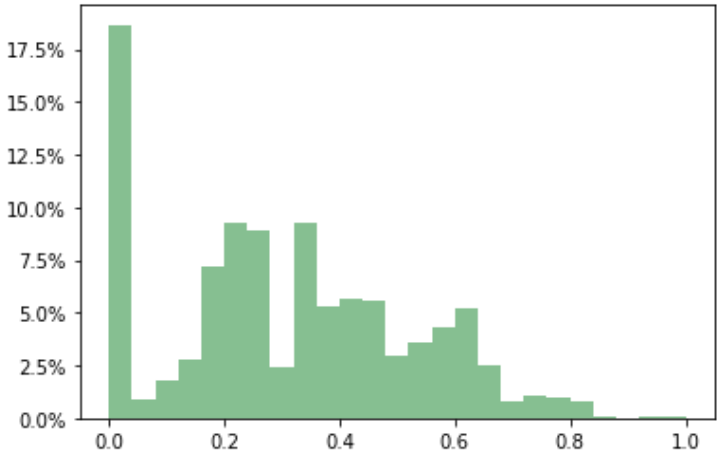
Nuevamente, se implementa un regresor *Random Forest*, con 10 árboles de decisión e hiperparámetros por defecto de la biblioteca. Para este modelo, las variables más importantes son la distancia, el nivel, la comuna y el nivel socioeconómico (ver anexo D.3). Con el objetivo de mejorar el desempeño del modelo, se hace un *grid search* (o búsqueda de subconjunto de hiperparámetros) para optimizar los hiperparámetros del mismo. Así, se buscan combinaciones de *n_estimators, max_features y max_depth*[22] que generen el mayor R^2, siendo estos valores 101, 9, y 20, respectivamente. En el modelo generado con los hiperparámetros recién mencionados, la importancia de cada variable tiene bastante similitud con el previo, pero intercambiando la importancia del nivel socioeconómico por la edad (ver anexo D.4).

Al aplicar este modelo entrenándolo en el año *t* y validándolo en el *t+1*, se observa un error cuadrático medio (MSE) de 422,61, un RMSE de 20,56 y un error absoluto medio (MAE) de 13,16.

Luego, con las predicciones de este modelo se reconstruyó la lista de preferencias priorizada de cada estudiante, y se midió la capacidad de ordenar cada lista en comparación a la original.

Para esta iteración, se obtuvo un RBO promedio de 0,3, donde esta métrica fluctúa entre 0 y 1, con 1 correspondencia total. En este modelo ya se comienza a observar un rendimiento menor, tal como se muestra en la ilustración 22 que indica cómo se distribuye la frecuencia del RBO en las postulaciones. Se observa que el grueso de los datos se distribuye entre 0,2 y 0,6, habiendo aproximadamente un 18% de los datos con RBO igual a 0.

Ilustración 22: Distribución de la métrica RBO en modelo de selección y orden

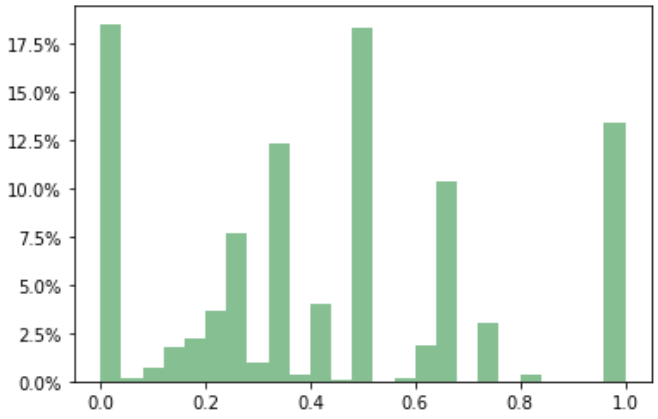


Fuente: Elaboración propia

[22] Para más información sobre los parámetros: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html

Se calculó la métrica *Contador_presentes_porcentual*, que indica cuántos de los EE a los que efectivamente postula un estudiante se encuentran en la lista cortada de 4 EE predicha de forma porcentual en relación con el largo de la lista real. Así, al sacar el promedio global de esta métrica se obtiene un resultado de 0,4, lo que indica que hay en promedio 0,4 EE en la lista de predicciones, que se condice con lo realmente postulado, es decir, probablemente en la mayor parte de las postulaciones no haya ninguna coincidencia. A continuación, se puede ver la distribución porcentual de datos, que muestra que todos los datos poseen una coincidencia en la lista:

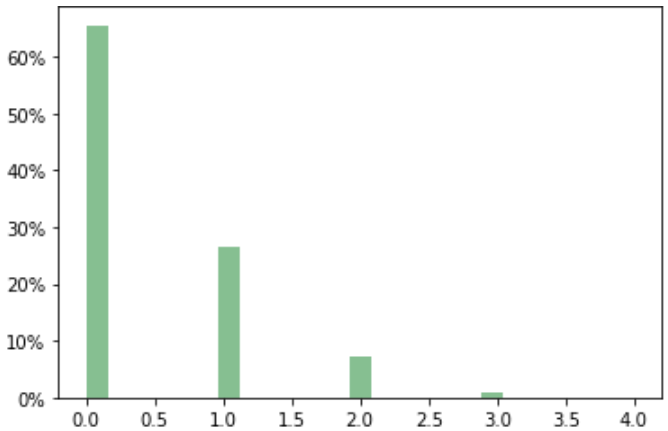
Ilustración 23: Distribución de la métrica Contador presentes porcentual en modelo de selección y orden



Fuente: Elaboración propia

Así mismo, se calculó la métrica "Match_porcentual", obtenida también en la sección anterior. El promedio de esta métrica da 0,15, indicando que en casi ninguno de los casos existe un "*match*" entre la correcta elección de EE postulado y su orden en la lista, lo que es análogo a la métrica anterior, por la baja cantidad de aciertos. En específico, se observa la siguiente distribución de resultados:

Ilustración 24: Distribución de la métrica Match promedio en modelo de orden y selección



Fuente: Elaboración propia

De este gráfico y de los resultados generales de este modelo, se observa un mal desempeño, ya que en ninguna de las postulaciones el modelo ordena perfectamente. Con estos resultados, nuevamente se corrobora que la métrica RBO está sobrevalorando los resultados pues se pensaría que con un valor promedio de 0,61 el modelo está siendo muy preciso en el ordenamiento y selección, pero este valor refleja principalmente la ponderación acentuada en que el modelo le acierte a los primeros valores de la lista, lo cual es valioso pues demuestras una buena predicción para las primeras prioridades del estudiante, donde idealmente el sistema intentará asignarlo.

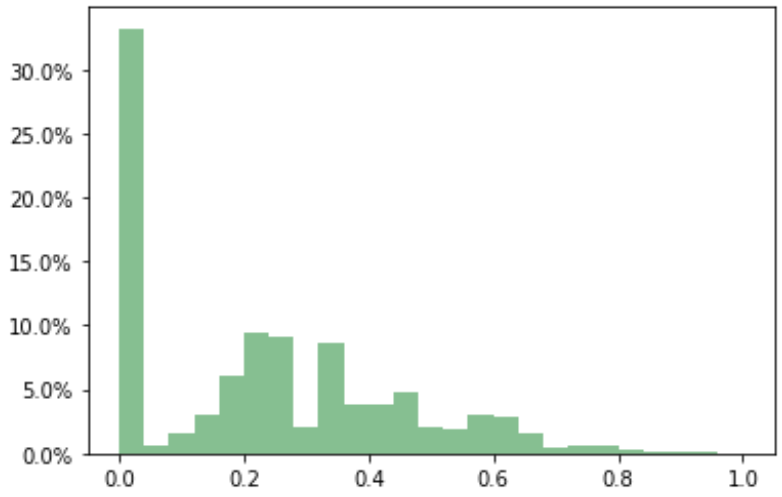
Finalmente, se utilizó la métrica *Ratio_correctos*, descrita previamente. El promedio de esta métrica es de 0,3 y se observa que aproximadamente un 18% de los datos posee un ratio de 1, indicando una baja efectividad del modelo.

6.2.2.2. Modelo de orden de cercanos

Bajo el supuesto confirmado (por la literatura y por la importancia que da el modelo a la variable) de que la distancia entre el estudiante y el EE es relevante dentro de las preferencias de las familias al escoger EE, es que se hace la prueba de, en vez de predecir simulando que un estudiante postula a todos los EE del territorio, predecir solamente el orden que el modelo daría si es que el estudiante postulara a los 4 EE más cercanos a su domicilio. Al aplicar lo anteriormente mencionado, se observa un error cuadrático medio (MSE) de 11,8, un RMSE de 10,57 y un error absoluto medio (MAE) de 1,36.

Además, respecto al desempeño anterior, empeora levemente el RBO, arrojando un promedio de 0,23 y una distribución porcentual de los datos de la siguiente forma:

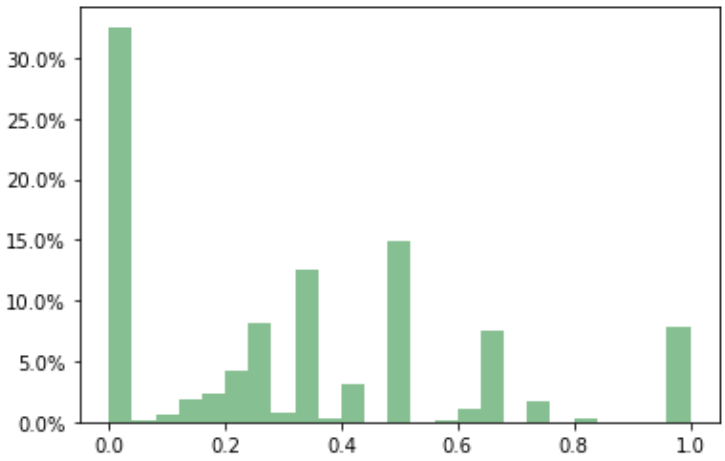
Ilustración 25: Distribución de la métrica RBO en modelo de orden de cercanos



Fuente: Elaboración propia

Respecto a la métrica *Contador_presente_porcentual*, se observa un promedio de 0,32, con una distribución que muestra más del 30% de los datos en cero y aproximadamente un 10% de los datos en el valor máximo 1, como se puede ver a continuación:

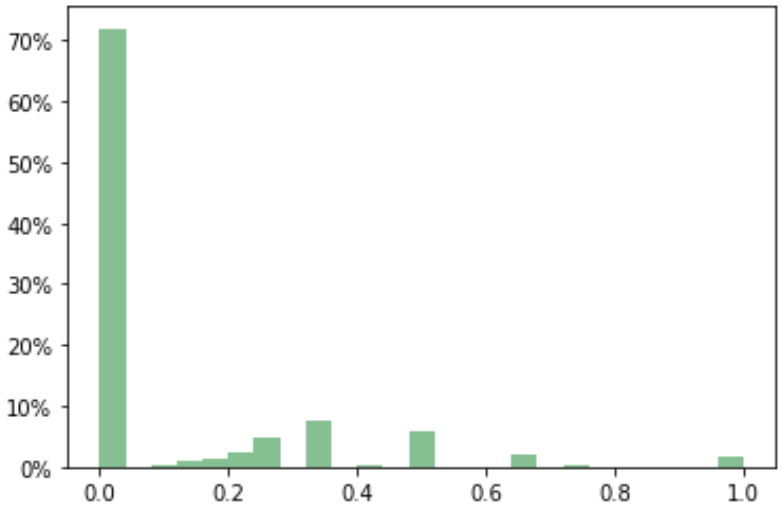
Ilustración 26: Distribución de la métrica Contador presente porcentual en modelo de orden de cercanos



Fuente: Elaboración propia

Así mismo, la métrica *Match_porcentual*, arroja un promedio de 0,1 con la siguiente distribución:

Ilustración 27: Distribución de la métrica Match porcentual en modelo de orden de cercanos



Fuente: Elaboración propia

Por último, la métrica *Ratio_correctos* indica un promedio de 0,33, concluyendo que nuevamente la efectividad del modelo es baja.

Finalmente, y a pesar de los intentos de usar distintos algoritmos y de optimizar el modelo, este parece no ser capaz de seleccionar y ordenar los EE a postular por cada estudiante, mostrando un bajo desempeño. El modelo escogido no se logra adaptar a la problemática, y en particular al formato de lista que ésta presenta, con lo que la predicción de la preferencia del postulante como variable dependiente no lleva a buenos resultados. Se cree que los resultados podrían mejorar de haber un modelo que trabaje específicamente con listas priorizadas por individuo. Además, el modelo tiene dificultades para seleccionar entre todo el universo de establecimientos que se ofrecen a través del sistema en la región, por ser un número muy grande, por lo que se podría mejorar con un modelo intermedio que se base sólo en la selección de EE candidatos por postulante, según variables relevantes. Lo anterior se podría hacer mediante algún método de clusterización o similares.

En definitiva, y dada la evidencia tanto del nivel de uso del SAE como de los modelos de predicción de la preferencia para construir postulaciones, se decide no continuar con la metodología definida inicialmente. Por lo tanto, conociendo las limitaciones de los datos del SAE en relación con capturar las preferencias del total de estudiantes que ingresan o se cambian al sistema escolar, se decidió a modo exploratorio, generar un modelo a nivel agregado. Eso significa que ya no se predicen las preferencias, sino que se busca determinar directamente la demanda por cada nivel y EE en el territorio.

6.2.3. Modelo de predicción a nivel agregado

A modo de explorar una forma alternativa a la inicialmente planteada para tener una aproximación sobre la demanda escolar, es que se plantea un nuevo modelo, ya no basado en las preferencias individuales de cada estudiante o familia postulante a través del SAE, sino ahora considerando las vacantes ofrecidas a través del sistema, y luego cuántas de ellas son ocupadas.

Cabe destacar que este modelo posee un supuesto base, que fue refutado parcialmente en la sección 6.1.3 del presente documento, que es asumir que todas las asignaciones del SAE que son aceptadas por sistema luego se convierten en una matrícula efectiva, lo cual se comprobó que no siempre es así, con lo cual se genera de base una incongruencia de cálculo. De todas maneras, esto no impide hacer pruebas con el modelo, pues si este demuestra un buen desempeño, los datos pueden ir refinándose a futuro y los resultados finales de utilizarlo podrían ser muy positivos.

El modelo utilizado en esta sección se basa en una predicción de demanda a nivel de EE, nivel ofrecido y código de curso asociado (que es una tupla de variables características del establecimiento: `COD_GRADO`, `COD_ENSE`, `COD_ESPE`, `COD_SEDE`, `COD_GENERO` y `COD_JOR`, todas mencionadas en la sección 5.2). El modelo se compone de variables

independientes, que corresponden a un vector de variables sociodemográficas del EE en cuestión junto con las vacantes ofrecidas (`COD_NIVEL`, `COD_GRADO`, `COD_ENSE`, `COD_JOR`, `COD_ESPE`, `COD_SEDE`, `LAT_RBD`, `LON_RBD`, `COD_COM_RBD`, `COD_DEPE2`, `TIPO_EE`, `CON_COPAGO`, `ORI_RELIGIOSA`, `PAGO_MATRICULA`, `PAGO_MENSUAL`, `SNED`, `IVM`, `VACANTES`), y una variable dependiente o a predecir, que serían las vacantes utilizadas (que deben ser menores o iguales a las vacantes ofrecidas). Además, cabe destacar que la predicción se hace entrenando los datos de resultados del SAE 2018 y testeando y prediciendo en el año 2019.

En términos de la construcción de la muestra, lo que se hizo fue reproducir la asignación del SAE, tanto de su etapa regular como complementaria, utilizando las bases de datos de resultados del SAE (D1 y D2), que contienen las respuestas de los postulantes. Para lo anterior, se siguen los mismos pasos descritos en la sección 6.1.3 de este documento, y luego se hace una concatenación con la base de oferta de EE del SAE, que contiene las características sociodemográficas de cada establecimiento. Con ello, se obtiene una base de datos consolidada según lo descrito en el párrafo anterior. Cabe destacar que teniendo esta base de datos, se hace el filtro para quedarse con los niveles 1° básico, 7° básico y I° medio de las comunas de Concepción, Talcahuano, Hualpén, San Pedro de la Paz, Penco, Chiguayante, Coronel, Tomé, Hualqui y Lota de la Región del Biobío, tal y como se ha trabajado durante todo esta investigación (esto se hizo por un tema de coherencia con la línea de investigación de la tesis, pero se podría extender a otros territorios y niveles con el mismo código propuesto).

Así, se implementa el algoritmo *Random Forest* a través de la librería *scikit-learn* de Python, utilizando 10 árboles de decisión y parámetros estándar, con las variables recién mencionadas. Al hacer esto, la priorización de variables en el modelo indica una gran relevancia de las vacantes, el SNED y el IVM, así como las coordenadas geográficas del EE (ver anexo D.5). Lo anterior tiene mucho sentido, pues hay una alta correlación entre las vacantes ofrecidas y las utilizadas. Además, los índices de desempeño y de vulnerabilidad escolar son variables explicativas de cómo es un EE en una gran medida.

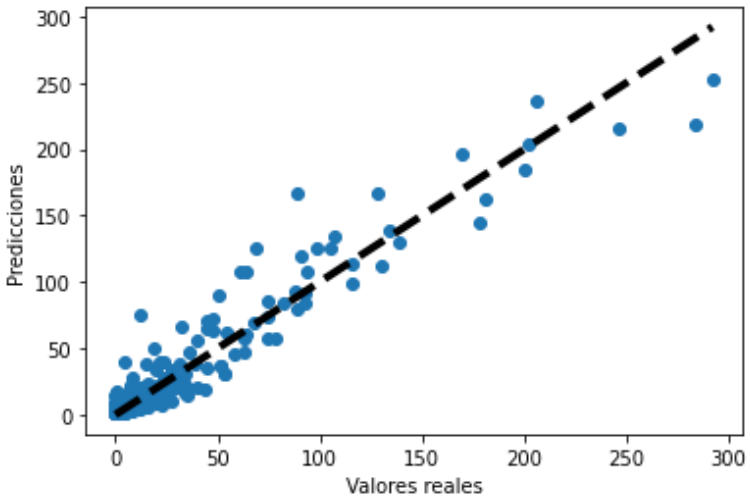
Con el objetivo de mejorar el desempeño del modelo, se hace un *grid search* (o búsqueda de subconjunto de hiperparámetros) para optimizar los hiperparámetros del mismo. Así, se buscan combinaciones de *n_estimators, max_features y max_depth*[23] que generen el mayor R^2, siendo estos valores 150, 9, y 10, respectivamente. En el modelo generado con los hiperparámetros recién mencionados, la importancia de cada variable tiene bastante similitud con el previo, pero priorizando esta vez el "tipo de establecimiento" (si es de

[23] Para más información sobre los parámetros: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html

básica, media HC, media TP o completo) posterior a las vacantes, seguido del IVM y SNED (ver anexo D.6).

Al aplicar este modelo entrenándolo en un año y testeándolo en el siguiente, se observa un error cuadrático medio (MSE) de 91,74, un RMSE de 9,58 y un error absoluto medio (MAE) de 14,9. Para tener en cuenta la dimensión de estos errores, recordar que las vacantes van entre 1 y 300, con una concentración mayor entre 1 y 30.

Ilustración 28: Valores reales versus predichos de vacantes usadas



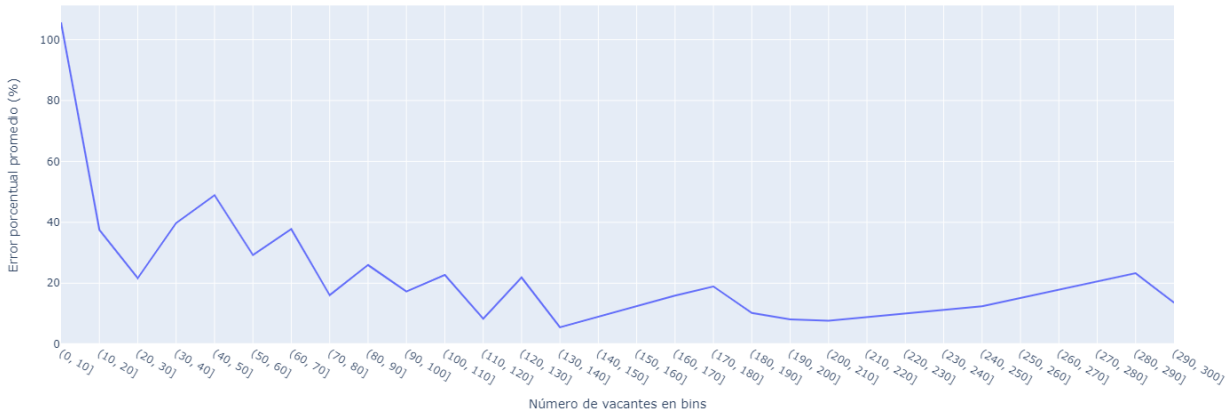
Fuente: Elaboración propia

Como se puede observar en la ilustración 28, si bien el modelo presenta un error de base que ya fue mencionado, y además tiene un error medio considerable, a pesar de ello

parece seguir una tendencia lineal, logrando predicciones bastante cercanas a la realidad dentro del rango en que transita la variable.

Para indagar más en lo anterior, se generó un gráfico segmentando las vacantes reales por intervalos, y obteniendo el promedio del error porcentual absoluto entre lo predicho y lo real.

Ilustración 29: Promedio del error porcentual absoluto para intervalos de vacantes



Fuente: Elaboración propia

De este gráfico, se observa que el error porcentual absoluto promedio va disminuyendo conforme aumentan las vacantes ofrecidas, lo cual no es tan sorprendente al recordar que la mayor cantidad de datos se concentra entre las 0 y 40 vacantes. De todas maneras, en la generalidad se observa que el error promedio para cada intervalo fluctúa entre 20% y 50%.

Este modelo puede seguir siendo perfeccionado, y no pretende ser el centro de estudio de esta investigación, pero dado su desempeño parece tener un buen potencial de desarrollo y se considera un descubrimiento pertinente al estudio.

Capítulo 7: Discusión y Conclusiones

En primer lugar, es relevante destacar que los resultados de esta tesis se condicen con las investigaciones y la literatura principal existente en la materia, que argumenta que una de las variables fundamentales en la elección de las familias de establecimientos escolares, tiene que ver con su nivel socioeconómico, lo que se asocia además con variables como el nivel de educación y la información de la que dispone cada familia para escoger, que es dispar. Además, también se visibiliza que la variable distancia (entre el domicilio del estudiante y el EE a postular) tiene una gran relevancia, concluyendo que entre menor es el nivel o curso al que se postula y/o menor es el nivel socioeconómico, menor es la distancia que prefieren las familias. Por último, la investigación también arroja coincidencias con lo que demuestra la literatura en torno a la prioridad de las familias de estar en un establecimiento educacional particular subvencionado por sobre uno público, lo que se observa de forma explícita en los niveles de desocupación críticos de algunos establecimientos públicos, donde los particulares subvencionados en casi todos los casos mostraron una ocupación casi completa.

En términos de cumplimiento de objetivos, la tesis logra lo propuesto. En particular, y de acuerdo con el primer objetivo específico, aporta un diagnóstico respecto a la ocupación escolar para las distintas regiones del país, estudiando tanto por tipo de dependencia como por niveles, y haciendo un análisis exhaustivo mediante la creación de mapas interactivos, logrando definir un 30% de desocupación escolar como un porcentaje crítico que estaba presente en diversos territorios. En resumen, se observa una alta desocupación en algunas localidades de las regiones de Valparaíso, Maule, Ñuble, Biobío y La Araucanía, y como decisión estratégica se decide profundizar el estudio en 10 comunas de la zona de Concepción, de la Región del Biobío, donde se observa una desocupación crítica. Respecto al segundo objetivo específico, este también se cumple, en cuánto se identifica que el Sistema de Admisión Escolar no está siendo necesariamente una herramienta que se convierta en matrículas efectivas a pesar de la asignación. Los resultados arrojan que un 77% y 66% de los estudiantes se matriculó efectivamente en el EE asignado en los procesos de admisión 2019 y 2020, respectivamente. Además, también se encuentra que hay estudiantes que postulan de forma directa a EE que son parte de la oferta del SAE, jamás habiendo pasado por él, lo que puede significar que existe un bajo conocimiento del sistema, falta de herramientas digitales y/o físicas para utilizarlo, o también desinterés. El no utilizar el sistema, provoca distorsión en los datos al momento de hacer cálculos de demanda sin utilizar este porcentaje de personas, y además resta evidencia respecto al conocimiento de las preferencias de estas familias que constituyen demanda de educación pública o particular subvencionada. Por último, respecto al tercer objetivo específico, se logra generar 2 modelos que

podrían conducir a una buena predicción de demanda. Actualmente, ninguno de los modelos reportó un buen desempeño, pero en términos metodológicos se ve potencial para seguir desarrollando la investigación, y de todas formas se logró lo propuesto, que era analizar la eficacia de los métodos.

Respecto a limitaciones de la investigación, corresponde mencionar la cantidad de años que tiene el SAE y lo progresiva que ha sido su incorporación tanto en términos de regiones, como de EE y niveles, lo que implica que existan datos dispares entre años, y aún la muestra sea pequeña. Además, para la región escogida, los datos estaban disponibles recién desde el año 2018, y se decidió no analizar los años de pandemia, pues se considera que el análisis sobre el comportamiento de las personas en todo ámbito de cosas en este tiempo podría ser considerado irregular y los resultados no serían generalizables. Otra limitación, tuvo que ver con la cantidad de datos perdidos en limpieza de datos, los cuales debieron ser removidos para tener una muestra adecuada para el entrenamiento del modelo. En particular, era relevante poder incluir los datos del nivel socioeconómico de los estudiantes, lo cual se pudo conseguir sólo parcialmente. Lo mismo con las coordenadas geográficas del estudiante, variable fundamental para medir distancias, la cual era entregada con errores y valores nulos o imputados por el sistema, a pesar de haber hecho esfuerzos sin éxito por conseguir los valores reales. Por último, es destacable como una limitación, que el modelo utilizado para la predicción a nivel de preferencias individuales no captura de forma orgánica listas ordenadas y priorizadas de elementos, y para obtener este formato se realizó una predicción adaptada a través de la "preferencia del postulante" como variable dependiente, asumiendo órdenes impuestos según distancia para los EE no postulados, lo que implicó una distorsión de los datos y además una limitación no superada del modelo, pues no sólo tenía que priorizar los EE a postular, sino también debía ser capaz de elegir los EE factibles a los que efectivamente postularía un estudiante o su familia entre todo el universo de EE disponibles en el sistema, y esto no fue capaz de hacerlo. En relación con el modelo de predicción a nivel agregado, la mayor limitación tenía que ver con que las vacantes usadas no eran un dato completamente correcto, lo cual se descubrió posterior al estudio de utilización del SAE, que concluyó que la asignación no necesariamente se refleja en matrículas efectivas, por lo que este modelo tenía un error de base que generó que sus resultados no sean verídicos de antemano.

Este trabajo de tesis da pie para continuar futuras investigaciones, y junto con sus limitaciones, tiene grandes oportunidades de mejora o de expansión del campo de estudio. En particular, el modelo de preferencias individuales permite profundización en la investigación de nuevas formas de modelar o de predecir datos capturados en lógica de listas priorizadas. Además, sería relevante a futuro poder realizar investigación de modelos de panel, que logren capturar el contexto y evolución temporal del fenómeno, y además permitan estudiar con un mayor número de datos. Otra oportunidad de estudio teniendo más

datos, sería generar evidencia sobre "*clusters*" de estudiantes y/o "*clusters*" de establecimientos educacionales, según sus características, logrando hacer un match entre ambos para caracterizar las preferencias. Esta y otras ideas pueden surgir, sólo gracias a la existencia e implementación del Sistema de Admisión Escolar y otros datos sobre las matrículas, que permite dar cuenta de la preferencia explícita de los estudiantes y sus familias, y que acumulando años de información permitirán tener suficientes datos para generar evidencia relevante que habilite la toma de decisiones informada en la formulación de políticas públicas en materia de educación.

Por último, como recomendaciones generales de política pública, es relevante que se fortalezca el uso del SAE, levantando un diagnóstico y evaluación sobre su uso y generando una estrategia para su sostenibilidad futura. Es importante recordar que aún existe una resistencia a su uso, ya sea por desinformación o por su base de legitimidad moral aún frágil (Carrasco & Honey, 2019), debido a las contrarias posturas en torno al principio de acceso igualitario, sustentadas en la defensa de la meritocracia como valor que hace justicia al mérito y esfuerzo personal (Duk & Murillo, 2019). En ese sentido, aún queda trabajo por hacer para lograr instalar el sistema culturalmente y de forma masiva. Además, según los resultados obtenidos en esta investigación, es fundamental que el Estado tome medidas en relación con la capacidad ociosa de los establecimientos educacionales urbanos, considerando la inversión pública que hay detrás de cada uno y la poca eficiencia que significa tener desocupación, considerando además que los EE reciben un financiamiento bajo, lo que les impide ser sustentable creando una oferta precarizada que no ayuda a la mejora del sistema escolar. En esa línea, es fundamental implementar una buena planificación territorial, y sistemas de monitoreo de la ocupación en los distintos territorios, que permita tener siempre en cuenta la desocupación crítica (también, por el contrario, cuando hay sobredemanda) para minimizar el impacto en las comunidades escolares, también ligado a la oportunidad tanto para la mejora de infraestructura y/o relocalización, como también la mejora de las condiciones de profesores y asistentes en educación. Por último, y a modo de cierre, es importante que el Estado se haga cargo del fortalecimiento de la educación pública en todo ámbito, tal que deje de existir la lógica del "chorreo" desde la educación particular subvencionada, y todos los estudiantes puedan acceder a una educación de calidad, pensada desde el enfoque de derecho y de la equidad, y no con la extrema segregación con la que se desenvuelve hoy el sistema escolar chileno.

Bibliografía

Amaya, J., Canals, C., Mizala, A., Rodríguez, P., Uribe, P., & Valenzuela, J. (2021). *Planificación Territorial de la Oferta Escolar Pública: Avanzando en sustentabilidad y equidad.*

Bellei, C., Contreras, M., Canales, M., & Orellana, V. (2019). The Production of Socio-economic Segregation in Chilean Education: School Choice, Social Class and Market Dynamics. *Bloomsbury Publishing*, 221.

Biblioteca del Congreso Nacional. (2015). *Ley 20.845 de inclusión escolar que regula la admisión de los y las estudiantes, elimina el financiamiento compartido y prohíbe el lucro en establecimientos educacionales que reciben aportes del Estado.* Obtenido de https://www.bcn.cl/leychile/navegar?idNorma=1078172

Biblioteca del Congreso Nacional. (2017). *Ley 21.040 crea el sistema de educación pública.* Obtenido de https://www.bcn.cl/leychile/navegar?idNorma=1111237

Canales, M., Bellei, C., & Orellana, V. (2016). *¿Por qué elegir una escuela privada subvencionada? Sectores medios emergentes y elección de escuela en un sistema de mercado.* Obtenido de Estudios pedagógicos de Valdivia, 42(3), 89-109: http://dx.doi.org/10.4067/S0718-07052016000400005

Carrasco, A., & Honey, N. (2019). Nuevo Sistema de Admisión Escolar y su capacidad de atenuar la desigualdad de acceso a colegios de calidad; al inicio de un largo camino. *Estudios en Justicia Educacional.*

Carrasco, A., Gutiérrez, G., & Flores, C. (2017). Failed regulations and school composition: selective admission practices in Chilean primary schools. *Journal of Education Policy*, 642-672.

CEP. (2011). Estudio Nacional de Opinión Pública.

Chumacero, R., & Paredes, R. (2012). Vouchers, choice, and public policy: An overview. *Estudios de economía*, 115-122.

Chumacero, R., Gomez, D., & Paredes, R. (2008). I World walk 500 miles (if paid): Vouchers and school choice in Chile. *Latin-American Econometrica Society Meeting.* Rio de Janeiro, Brazil.

CIPER Académico. (2020). *Cómo terminar con el lugar privilegiado de la educación privada en Chile*. Obtenido de https://www.ciperchile.cl/2020/12/26/como-terminar-con-el-lugar-privilegiado-de-la-educacion-privada-en-chile/

Córdoba, C. (2014). La elección de escuela en sectores pobres: Resultados de un estudiocualitativo. *Psicoperspectivas*, 56-67.

Correa, J., Epstein, N., Epstein, R., Escobar, J., Rios, I., Aramayo, N., . . . Subiabre, F. (2021). School Choice in Chile. *Operations Research*, 70(2):1066-1087.

Corvalán, J., & Román, M. (2016). Dicen que esta escuela es mala, pero nosotros la encontramos buena. Elección de escuela en familias pobres en Chile. *En J. Corvalán, A. Carrasco y J. E. García Huidobro, Mercado escolar y oportunidad educacional. Libertad, diversidad y desigualdad*, 209-231.

Cucchiara, M. (2013). Marketing schools, Marketing cities: who win and who loses when schools become urban amenities. *Chicago, IL: University of Chicago Press.*

Cutler, A., Cutler, D., & Stevens, J. (2011). Random Forests.

Derpanis, K. (2010). *Overview of the RANSAC Algorithm.*

Dirección de Educación Pública. (2018). Relatos sobre la creación del Sistema Nacional de Educación Pública.

Duk, C., & Murillo, J. (2019). Segregación Escolar y Meritocracia. *Revista latinoamericana de educación inclusiva*, 11-13.

Elacqua, G. (2012). The impact of school choice and public policy on segregation. Evidence from Chile. *International Journal of Educational Development*, 444-453. DOI; https://doi.org/10.1016/j.ijedudev.2011.08.003.

Elacqua, G., Schneider, M., & Buckley, J. (2006). School Choice in Chile: Is it Class or the Classroom? *Journal of Policy Analysis and Management*, 577-601.

Eyzaguirre, S., Hernando, A., & Blanco, N. (2018). Cargando con la mochila ajena. Resultados y desafíos del nuevo Sistema de Admisión Escolar. *Centro de Estudios Públicos*, 20.

Fischler, M., & Bolles, R. (1981). *Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography.*

Gang-Hoon, K., Silvana, T., & Ji-Hyong, C. (2014). Big-Data aaplications in the government sector. *Communications of the ACM.*

Garay, M., & Sillard, M. (2021). Los atributos de los establecimientos educacionales que pueden predecir la preferencia por parte de los apoderados. *Calidad en la educación. DOI: https://dx.doi.org/10.31619/caledu.n54.909*, 173-211.

Gubbins, V. (2013). La Experiencia Subjetiva del proceso de Elección de Establecimiento Educacional en Apoderados de Escuelas Municipales de la Región Metropolitana. *Estudios pedagógicos (Valdivia). DOI: https://dx.doi.org/10.4067/S0718-07052013000200011*, 165-178.

Hernández, M., & Raczynski, D. (2015). Elección de escuela en Chile: De las dinámicas de distinción y exclusión a la segregación socioeconómica del sistema escolar. *Estudios Pedagógicos*, 127-141.

IBM. (s.f). *¿Qué es el Machine Learning?* Obtenido de https://www.ibm.com/cl-es/analytics/machine-learning

Kenley, R., & Seppänen, O. (2010). *Location-Based Management for Construction: Planning, Scheduling and Control.* Abingdon, Inglaterra: Spon Press.

Lara, B., Mizala, A., & Repetto, A. (2011). The Effectiveness of Private Voucher Education. Educational Evaluation and Policy Analysis, 33(2). doi:10.3102/0162373711402990.

Lareau, A., & Goyette, K. (2014). Choosing homes, choosing schools.

Manyika, J., & otros. (2011). *Big Data: The next frontier for innovation, competition, and productivity.* McKinsey Global Institute.

Ministerio de Educación. (s.f). *¿Qué es SAE?* Obtenido de https://parvularia.mineduc.cl/que-es-2/

Mohamed, A., Putu, L., & Made, I. (2021). *Data Analytics in the Supply Chain Management: Review of Machine Learning Applications in Demand Forecasting .*

Orellana, V., Caviedes, S., Bellei, C., & Contreras, M. (2018). La elección de escuela como fenómeno sociológico. Una revisión de literatura. *Revista Brasileira de Educação*, 1-19.

Orfield, G., & Frankenberg, E. (2013). Educational Delusions? Why Choice Can Deepen Inequality and How to Make Schools Fair. *Berkeley: University of California Press.*

Raczynski, D., Salinas, D., de la Fuente, L., Hernandez, M., & Lattz, M. (2010). Hacia una estrategia de validación de la educación pública-municipal: imaginarios, valoraciones y demandas de las familias. *Proyecto FONIDE N°: F310827.*

Reardon, S., & Owens, A. (2014). 60 Years after Brown: Trends and Consequences of School Segregation. *Annual Review of Sociology*, 199-218.

Rodríguez, C., Espinosa, D., & Padilla, G. (2020). Dónde quiero que estudien mis hijos/as: caracterización de la oferta educativa y sus niveles de demanda en Chile. *Revista de estudios y experiencias en educación. DOI: https://dx.doi.org/10.21703/rexe.20201941rodriguez4*, 19(41), 57-70.

Rodríguez, P., Palomino, N., & Mondaca, J. (2017). *El uso de datos masivos y sus técnicas analíticas para el diseño e implementación de políticas públicas.* BID.

Román, M., & Murillo, F. J. (2014). Uso de los resultados de las evaluaciones estandarizadas como criterio de elección y selección de escuelas. *Revista Iberoamericana de Evaluación Educativa*, 5-7.

Schroeder, R., Meyer, S., & Rungtusanatham, J. (2011). *Administración de operaciones 5a edición.* Mcgraw Hill.

Sillard, M., Garay, M., & Troncoso, I. (2018). Analysis of the new school admission system in chile: the region of magallanes as a pilot experience. *Calidad en la educación. DOI: https://dx.doi.org/10.31619/caledu.n49.578*, 112-136.

Valenzuela, J., Bellei, C., & Ríos, D. (2014). Socioeconomic school segregation in a market-oriented educational system. The case of Chile. *Journal of Education Policy. DOI: https://doi.org/10.1080/0268 0939.2013.806995*, 217-241.

Wang, J., & Hastie, T. (2014). Boosted varying-coefficient regression models. *Journal of Computational and Graphical Statistics, vol. 23, no. 2*, 361–382.

Webber, W., Moffar, A., & Zobel, J. (2010). *A similarity measure for indefinite rankings.* Obtenido de https://doi.org/10.1145/1852102.1852106

Wooldridge, J. (2009). *Introducción a la econometría: Un enfoque moderno.* CENGAGE Learning.

Anexos

Anexo A

A.1. Caracterización de la población urbana

Tabla 15: Caracterización de la población urbana

Medida de tendencia central	N° de habitantes
Promedio	48.352
Mediana	13.253
Mínimo	1.075
Máximo	568.094

Fuente: Elaboración propia

A.2. Cantidad de comunas por rango de n° de habitantes

Tabla 16: Cantidad de comunas por rango de n° de habitantes

Rango de n° de habitantes	Cantidad de comunas
(0-150.000]	288
(150.000-300.000]	25
(300.000-450.000]	4
(450.000-600.000]	2

Fuente: Elaboración propia

A.3. Características de comunas urbanas con más de 150.000 habitantes

NOMBRE REGIÓN	NOMBRE COMUNA	TOTAL POBLACIÓN CENSADA	TOTAL ÁREA URBANA	TOTAL EE PÚBLICO URBANO	EE CON +30% DESOCUPACIÓN
ARICA Y PARINACOTA	ARICA	221.364	205.079	26	2 (7,7%)
TARAPACÁ	IQUIQUE	191.468	189.065	19	3 (15,8%)
ANTOFAGASTA	ANTOFAGASTA	361.873	354.104	41	1 (2,4%)
	CALAMA	165.731	158.487	22	1 (5,3%)
ATACAMA	COPIAPÓ	153.937	150.962	23	3 (13%)
COQUIMBO	LA SERENA	221.054	200.640	22	6 (27,3%)
	COQUIMBO	227.730	214.550	22	5 (22,7%)
VALPARAÍSO	VALPARAÍSO	296.655	295.918	45	27 (60%)
	VIÑA DEL MAR	334.248	334.248	45	22 (48,9%)
LIBERTADOR GENERAL BERNARDO O'HIGGINS	RANCAGUA	241.774	234.183	28	4 (14,3%)
MAULE	TALCA	220.357	210.916	33	17 (51,5%)
ÑUBLE	CHILLÁN	184.739	168.647	25	14 (56%)
BIOBÍO	CONCEPCIÓN	223.574	219.057	30	18 (60%)
	TALCAHUANO	151.749	150.320	28	14 (50%)
	LOS ÁNGELES	202.331	151.087	23	10 (43,5%)
LA ARAUCANÍA	TEMUCO	282.415	263.165	25	10 (40%)
LOS RÍOS	VALDIVIA	166.080	154.716	23	5 (21,7%)
LOS LAGOS	PUERTO MONTT	245.902	220.143	29	8 (27,6%)

Tabla 17: Comunas urbanas con más de 150.000 habitantes

Fuente: Elaboración propia

Anexo B

B.1. Distribución de vacantes por región y tipo de dependencia

Tabla 18: Distribución de vacantes por región y tipo de dependencia

Región	Público	Subvencionado con copago	Subvencionado gratuito	N° de vacantes	% total de vacantes
Tarapacá	4.345	1.873	6.696	12.914	2,30%
Antofagasta	15.646	4.237	3.043	22.926	4,09%
Atacama	11.770	1.856	1.036	14.662	2,61%
Coquimbo	18.905	3.855	8.431	31.191	5,56%
Valparaíso	43.263	11.839	25.584	80.686	14,38%
Libertador Gral. Bernardo O'Higgins	26.986	4.074	6.848	37.908	6,76%
Maule	29.131	2.540	9.910	41.581	7,41%
Biobío	45.464	6.150	15.677	67.291	12,00%
La Araucanía	24.000	3.398	22.207	49.605	8,84%
Los Lagos	23.480	1.781	10.722	35.983	6,41%
Aysén	3.614	237	1.981	5.832	1,04%
Magallanes	5.015	862	269	6.146	1,10%
Región Metropolitana	43.408	24.083	36.895	104.386	18,61%
Los Ríos	9.929	809	6.793	17.531	3,13%
Arica	3.753	152	4.447	8.352	1,49%
Ñuble	15.170	1.319	7.474	23.963	4,27%
Total	323.879	69.065	168.013	560.957	100,00%

Fuente: Elaboración propia

B.2. Promedio de preferencias por curso y región

Tabla 19: Promedio de preferencias por curso y región

Región	Básica								Media			
	1º	2°	3°	4°	5°	6°	7°	8°	Iº	IIº	IIIº	IVº
Tarapacá	3,13	3,41	3,23	3,29	3,00	3,29	3,02	3,32	3,03	3,27	3,32	2,96
Antofagasta	3,75	3,82	3,82	3,64	3,71	3,73	3,39	3,52	3,93	3,49	3,89	3,82
Atacama	3,87	3,78	3,65	3,80	3,52	3,70	3,07	3,37	3,61	3,35	3,90	4,00
Coquimbo	3,95	4,15	3,95	3,96	4,02	3,99	3,52	3,67	3,61	3,83	3,71	3,55
Valparaíso	**3,48**	**3,61**	**3,54**	**3,48**	**3,50**	**3,54**	**3,36**	**3,35**	**3,54**	**3,49**	**3,43**	**3,38**

Región	Básica								Media			
	1º	2°	3°	4°	5°	6°	7°	8°	Iº	IIº	IIIº	IVº
Libertador Gral. Bernardo O'Higgins	3,12	3,34	3,25	3,32	3,14	2,99	2,76	3,07	3,43	3,38	3,50	3,17
Maule	3,37	3,49	3,27	3,41	3,44	3,27	3,06	3,25	3,78	3,38	3,72	3,44
Biobío	3,40	3,77	3,73	3,63	3,56	3,48	3,28	3,59	3,54	3,62	3,40	3,01
La Araucanía	3,07	3,29	3,30	3,14	3,10	3,09	2,91	3,18	2,89	3,15	3,08	3,31
Los Lagos	3,07	3,30	3,13	3,31	3,18	3,21	2,76	3,10	3,18	3,28	3,16	3,24
Aysén	2,99	3,24	3,31	3,12	3,04	3,43	2,88	3,02	2,78	3,05	2,96	4,13
Magallanes	3,61	3,59	3,40	4,26	3,39	3,61	3,59	3,88	3,32	3,39	3,21	3,53
Región Metropolitana	3,91	3,52	3,39	3,25	3,38	3,38	3,94	3,32	3,98	3,31	3,63	3,23
Los Ríos	3,36	3,38	3,25	3,27	3,19	3,51	2,71	3,03	2,71	2,75	2,65	2,82
Arica	3,63	3,37	3,53	3,44	3,45	3,45	3,25	3,57	3,66	3,46	3,58	3,75
Ñuble	2,96	3,48	3,19	3,43	3,18	3,18	2,91	3,42	2,80	3,08	3,10	3,96

Fuente: Elaboración propia

Anexo C

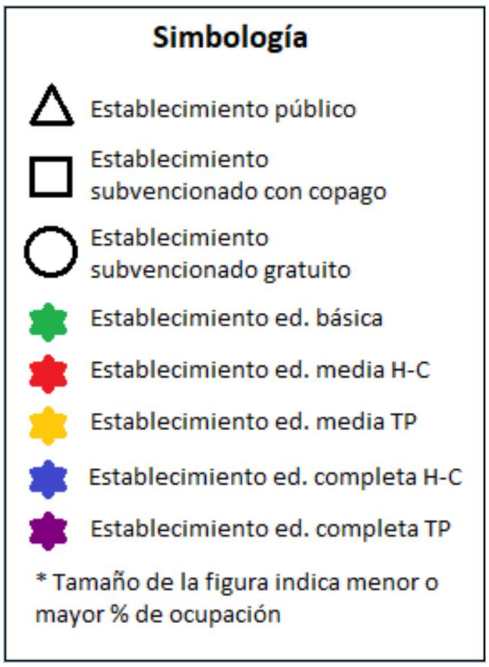
C.1. Cantidad de EE disponibles para cada nivel en la Región del Biobío

Tabla 20: Cantidad de EE disponibles para cada nivel en la Región del Biobío

Nivel	Cantidad de EE disponibles
1°básico	225
7° básico	108
1° medio	124

C.2. Mapas interactivos

Ilustración 30: Simbología de mapas interactivos



Fuente: Elaboración propia

Ilustración 31: Mapa Región Arica y Parinacota

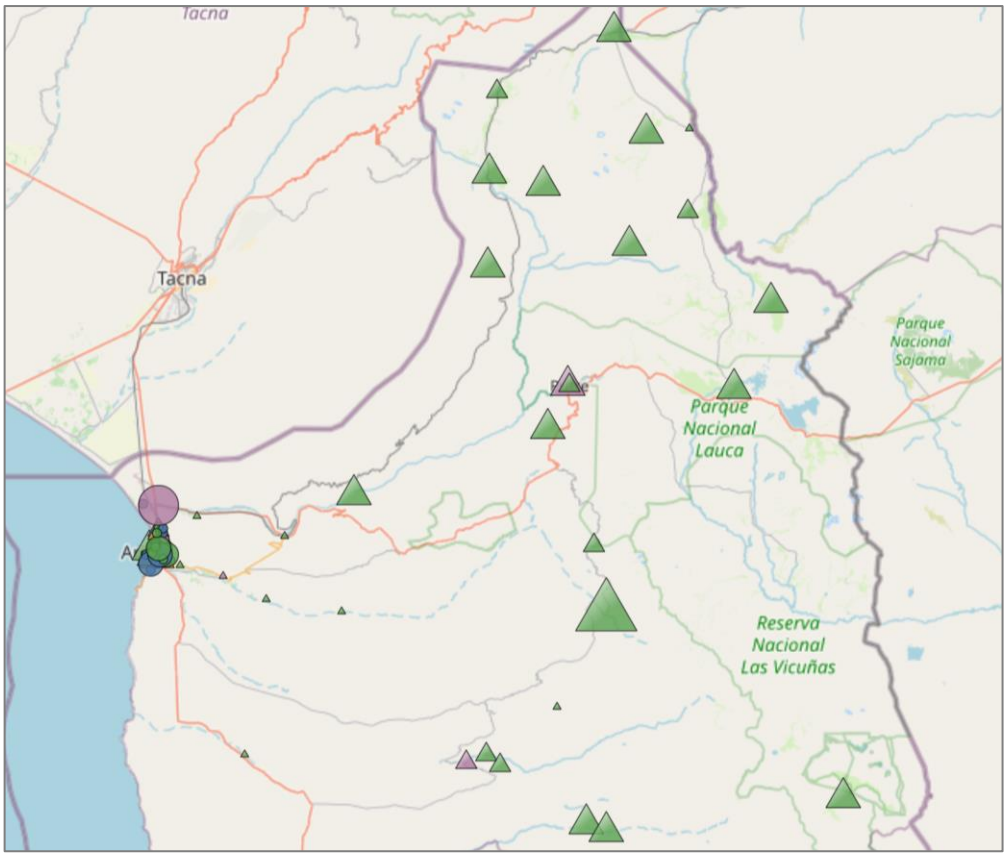


Ilustración 32: Mapa Región de Antofagasta

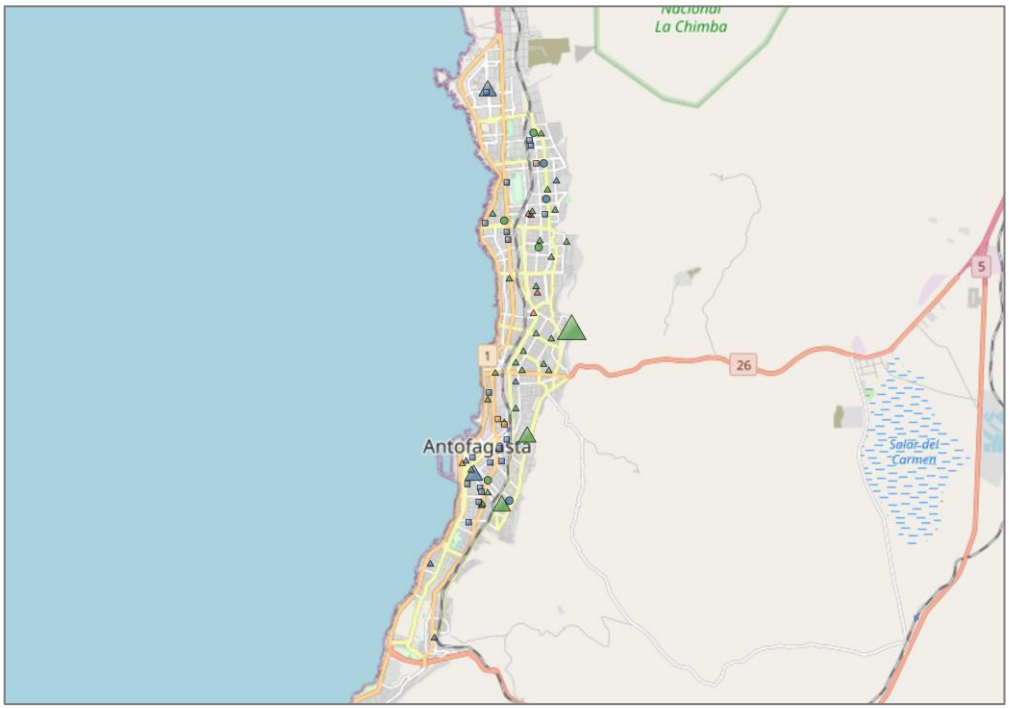


Ilustración 33: Mapa Región de Tarapacá

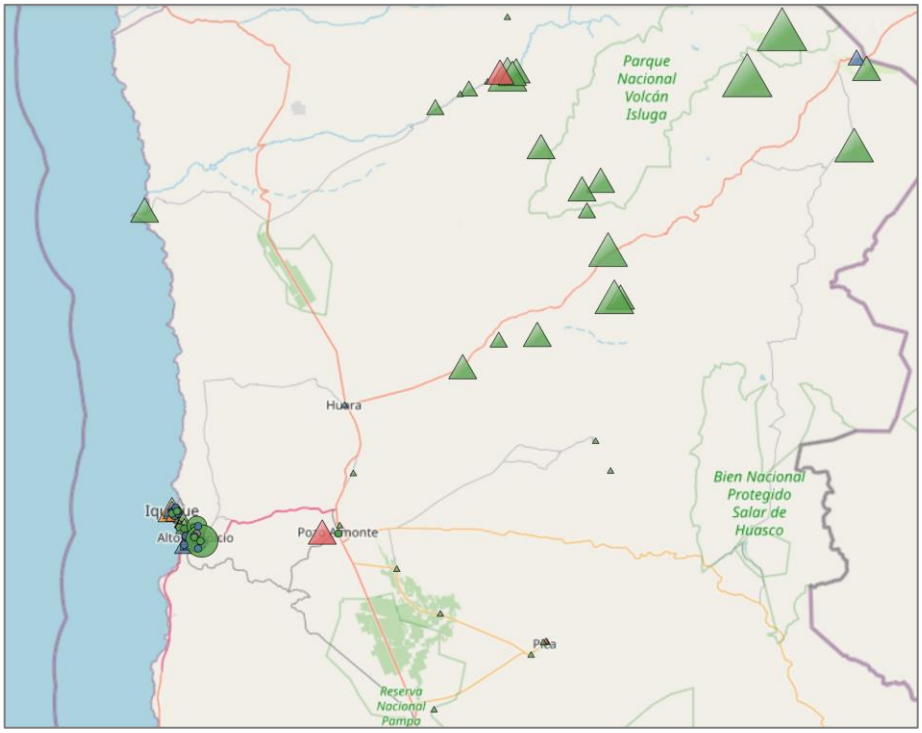


Ilustración 34: Mapa Región de Coquimbo

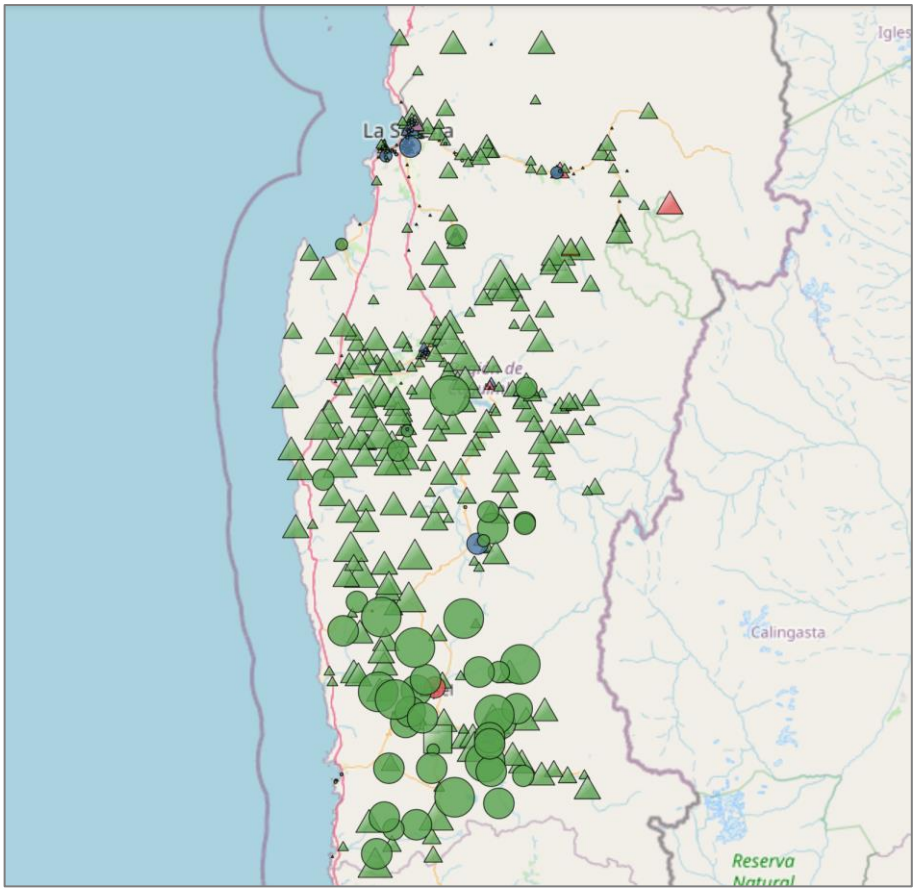


Ilustración 35: Mapa Región de Valparaíso

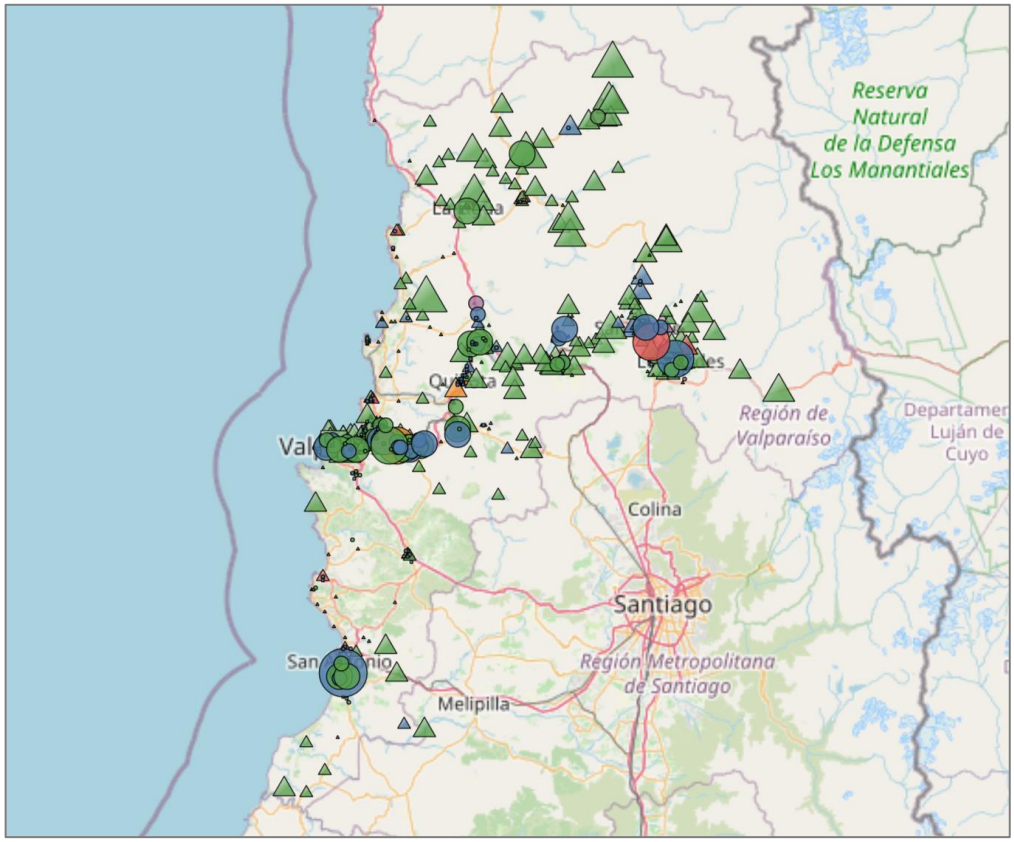


Ilustración 36: Mapa Región Metropolitana

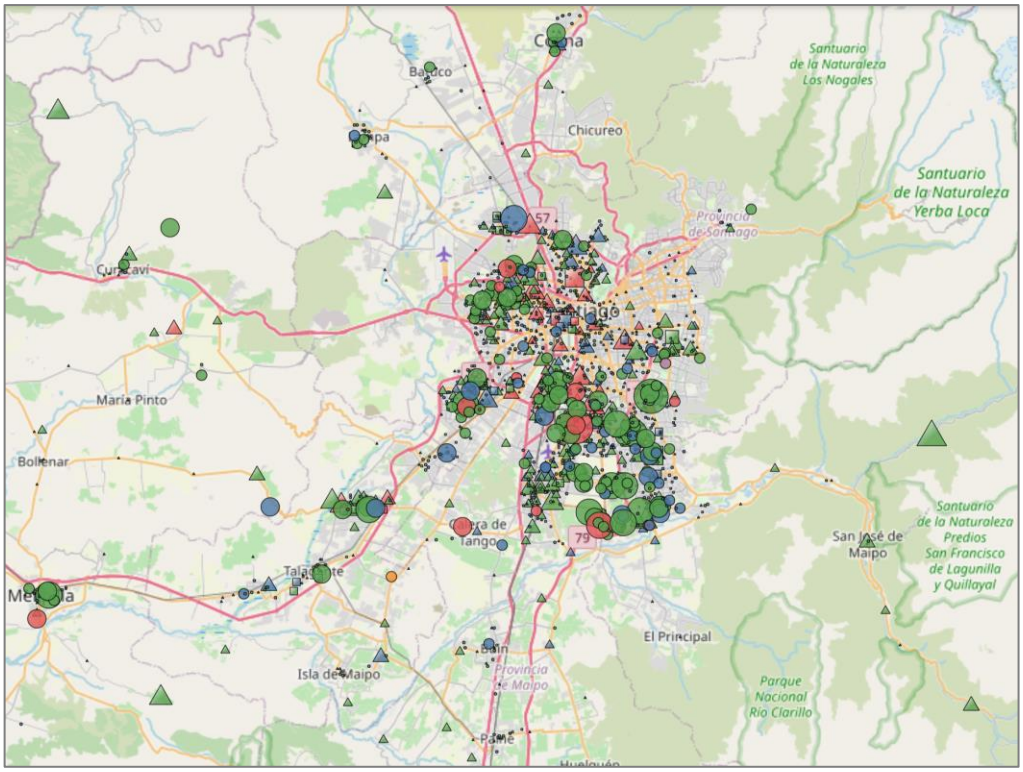


Ilustración 37: Mapa Región Libertador Bernardo O'Higgins

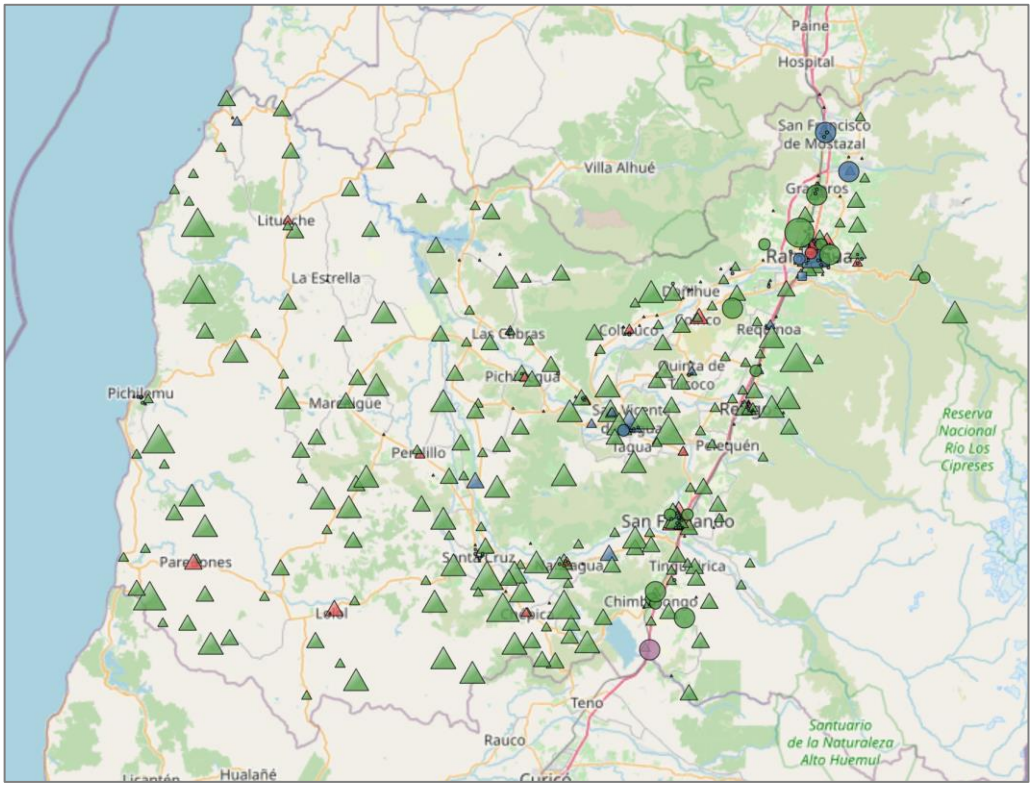


Ilustración 38: Mapa Región del Maule

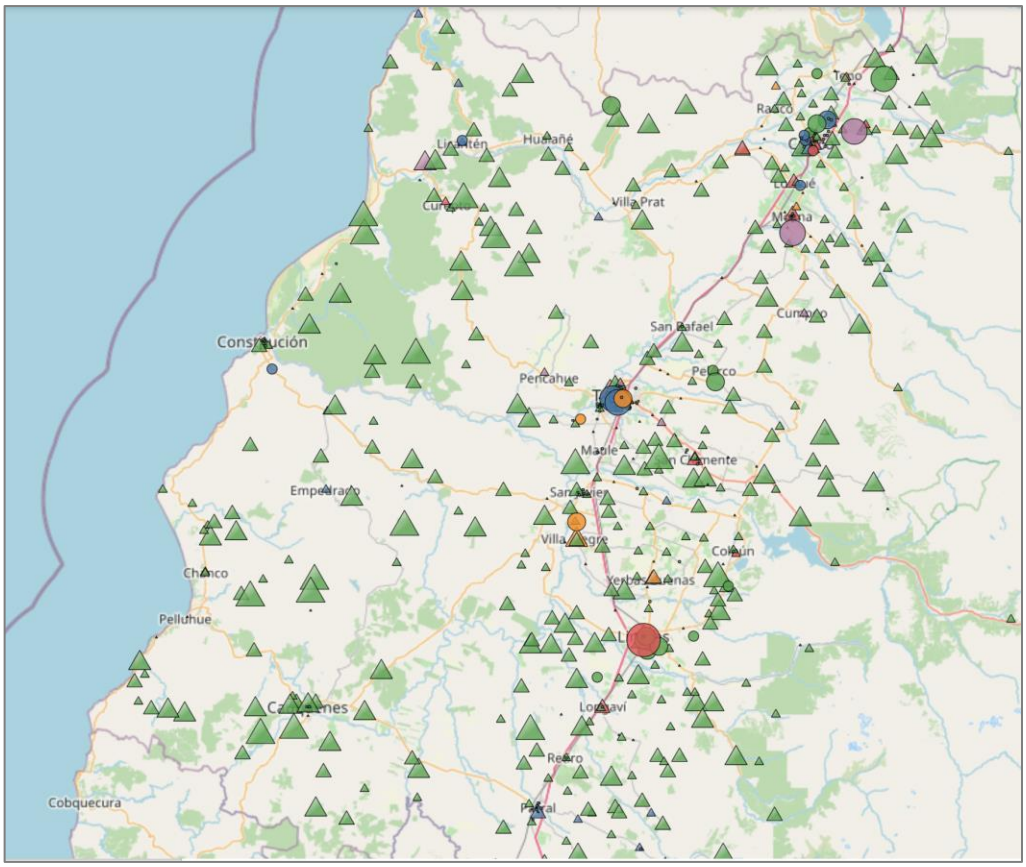


Ilustración 39: Mapa Región de Ñuble

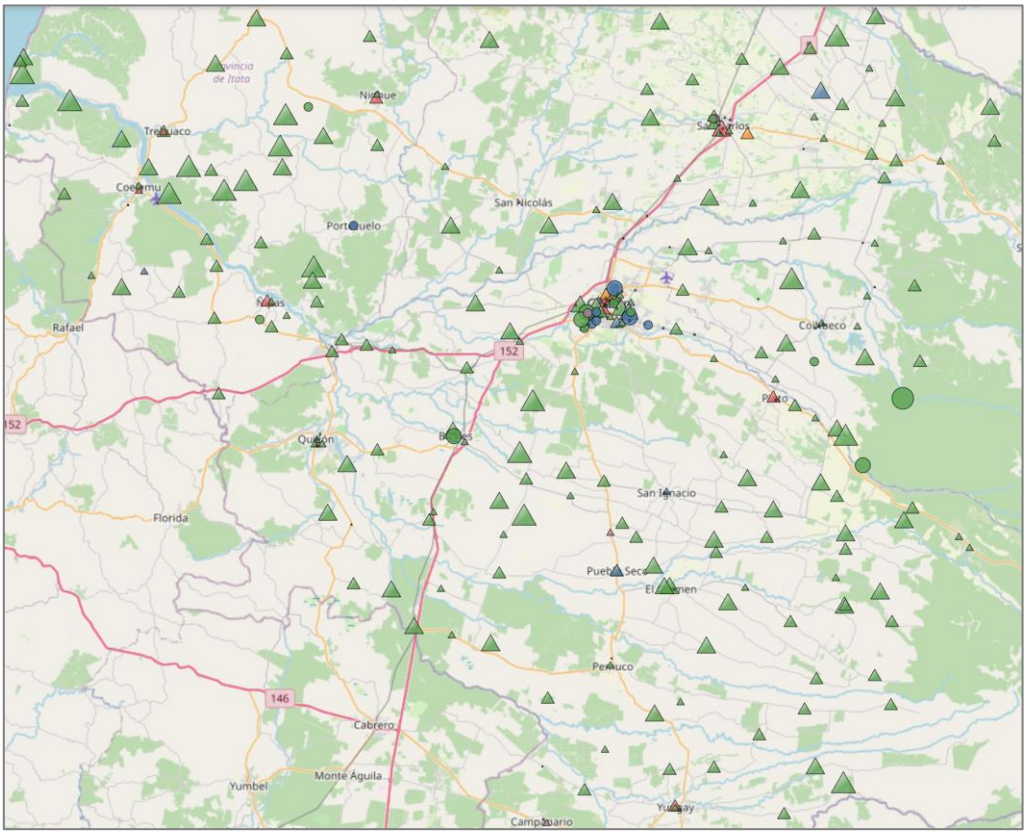


Ilustración 40: Mapa Región del Biobío

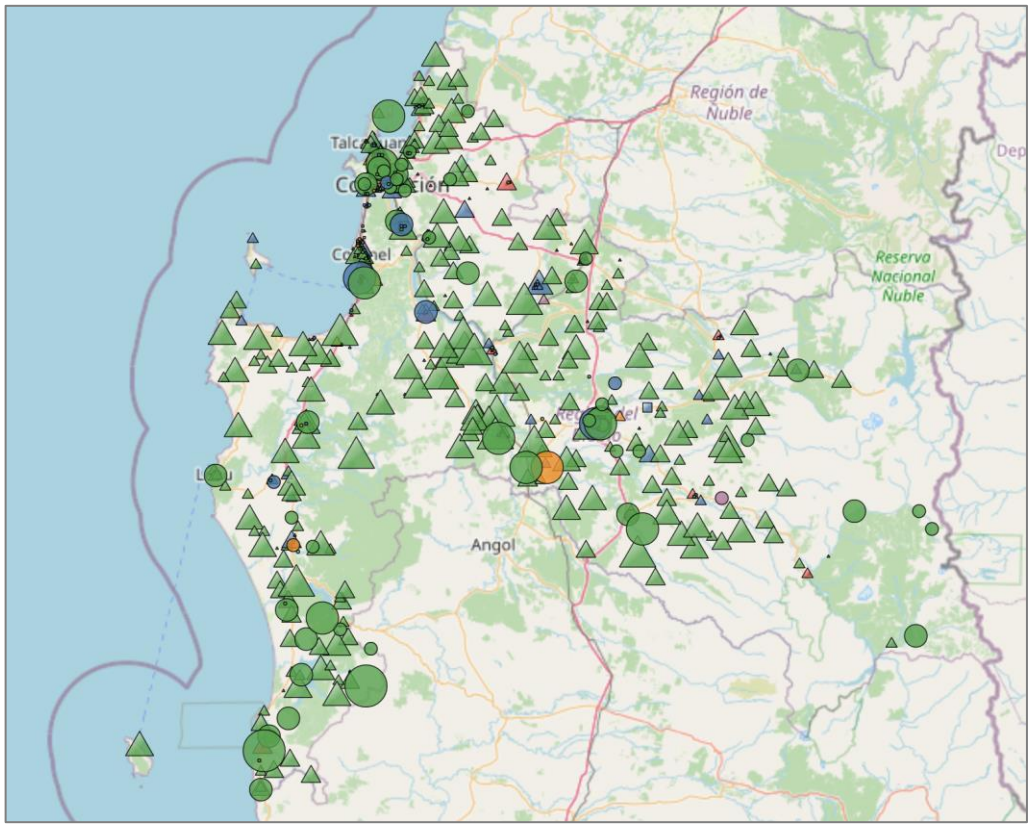


Ilustración 41: Mapa Región de La Araucanía

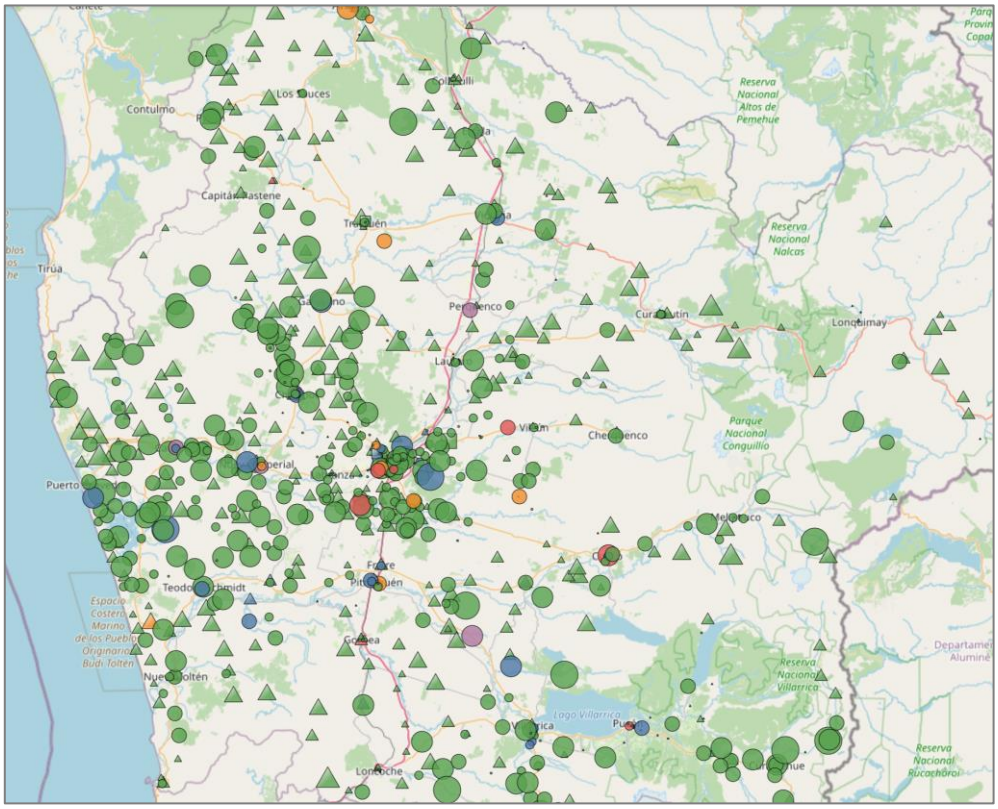


Ilustración 42: Mapa Región de Los Ríos

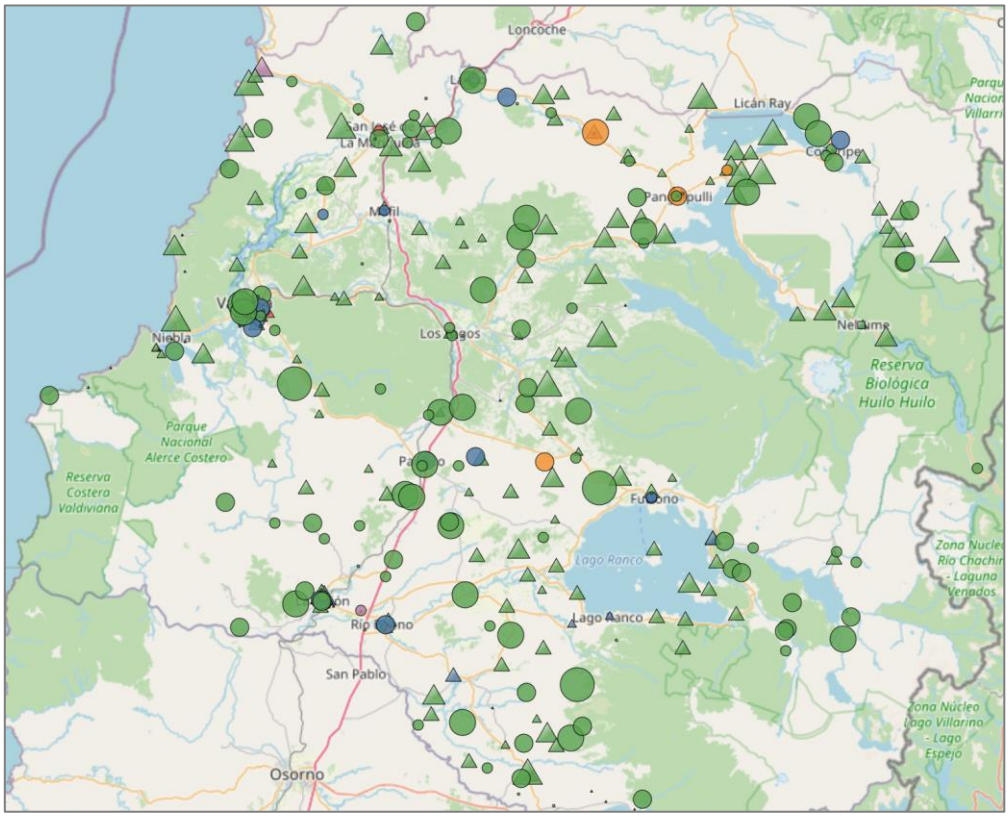


Ilustración 43: Mapa Región de Los Lagos

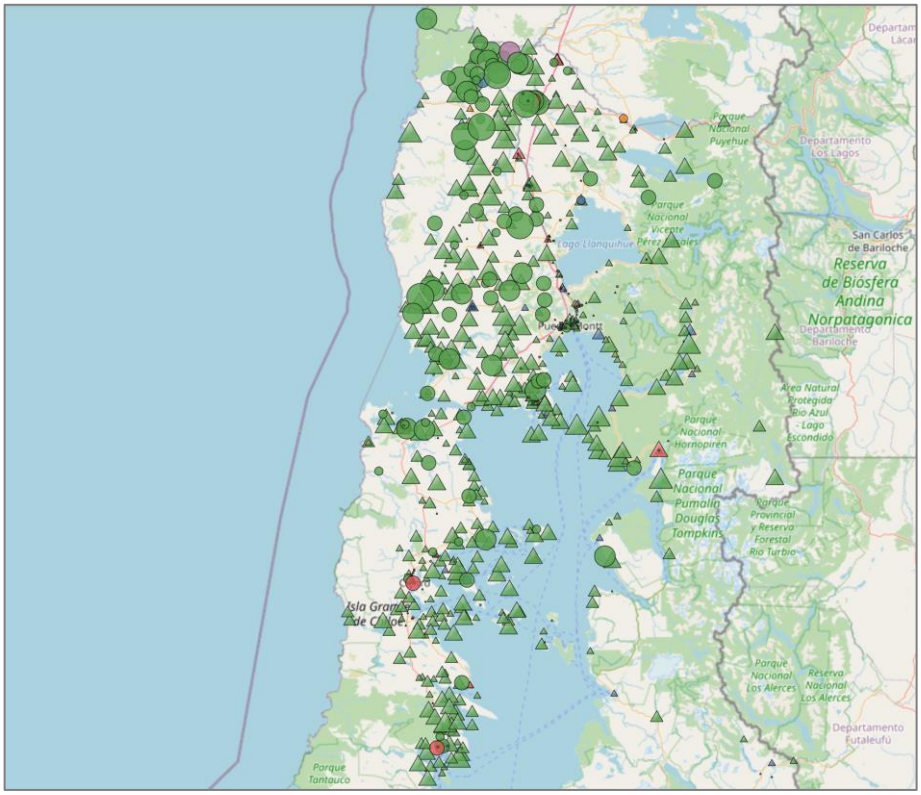


Ilustración 44: Mapa Región de Aysén

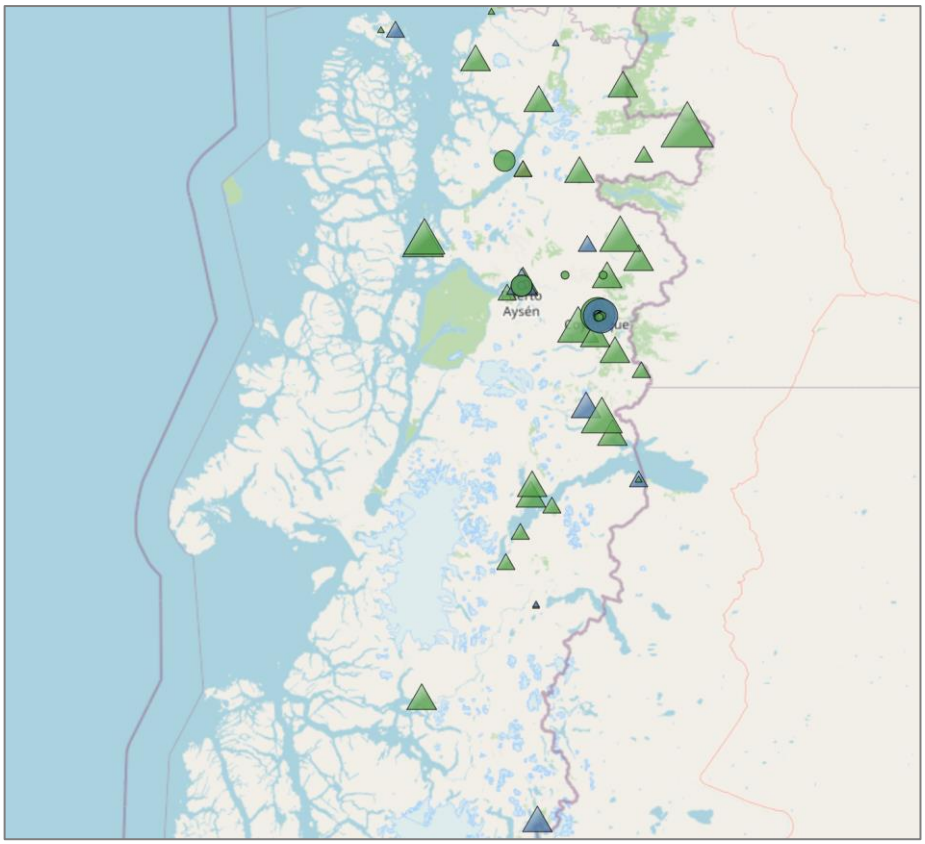
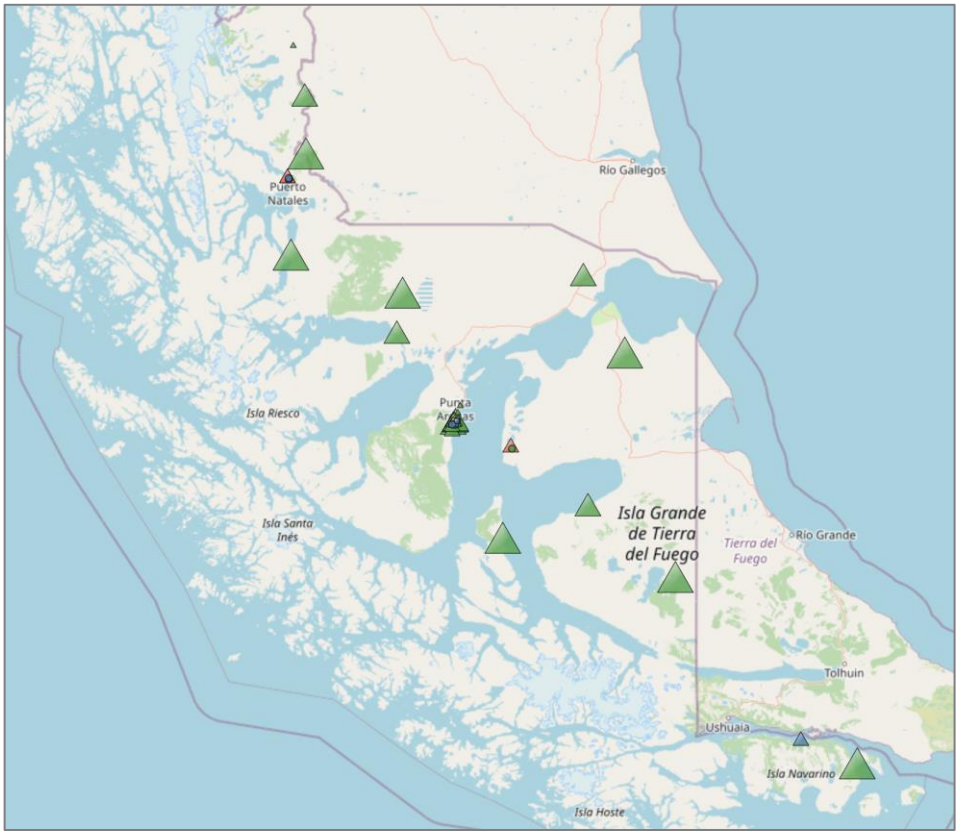


Ilustración 45: Mapa Región de Magallanes



Fuente: Elaboración propia

Anexo D

D.1. Importancia de variables en modelo de ordenamiento estándar

Variable	Importancia
DISTANCIA	0.360650
NSE	0.179069
IVM	0.078381
SNED	0.064125

CUPOS_TOTALES	0.060723
EDAD_ALU	0.050413
ES_MUJER	0.035750
COD_COM_RBD	0.032069
TIPO_EE	0.025054
ALTO_RENDIMIENTO	0.019800
PRIORIDAD_HERMANO	0.019629
PAGO_MENSUAL	0.015736
ORI_RELIGIOSA	0.015596
PRIORIDAD_MATRICULADO	0.011054
CON_COPAGO	0.008328
PAGO_MATRICULA	0.007031
COD_DEPE2	0.006555
COD_NIVEL	0.005913
PRIORIDAD_EXALUMNO	0.003935
PRIORIDAD_HIJO_FUNCIONARIO	0.000190

Fuente: Elaboración propia

D.2. Importancia de variables en modelo de ordenamiento optimizado

Variable	**Importancia**
DISTANCIA	0.220032
NSE	0.144229
CUPOS_TOTALES	0.097349
IVM	0.087282
PRIORIDAD_HERMANO	0.073035

Variable	Importancia
COD_COM_RBD	0.068802
SNED	0.061931
TIPO_EE	0.056878
EDAD_ALU	0.046646
PRIORIDAD_MATRICULADO	0.036036
ES_MUJER	0.019940
COD_NIVEL	0.016477
PAGO_MENSUAL	0.014835
ORI_RELIGIOSA	0.014132
CON_COPAGO	0.011575
ALTO_RENDIMIENTO	0.011314
PAGO_MATRICULA	0.008154
COD_DEPE2	0.007555
PRIORIDAD_EXALUMNO	0.003675
PRIORIDAD_HIJO_FUNCIONARIO	0.000123

Fuente: Elaboración propia

D.3. Importancia de variables en modelo de ordenamiento y selección estándar

Variable	Importancia
DISTANCIA	0.566330
COD_NIVEL	0.283166
COD_COM_RBD	0.080919
NSE	0.022668
IVM	0.010643
SNED	0.007950

Variable	Importancia
CUPOS_TOTALES	0.007015
EDAD_ALU	0.005980
ES_MUJER	0.004184
ORI_RELIGIOSA	0.002829
PAGO_MENSUAL	0.002069
ALTO_RENDIMIENTO	0.001892
TIPO_EE	0.001357
COD_DEPE2	0.001011
CON_COPAGO	0.000738
PAGO_MATRICULA	0.000579
PRIORIDAD_MATRICULADO	0.000419
PRIORIDAD_HERMANO	0.000184
PRIORIDAD_EXALUMNO	0.000060
PRIORIDAD_HIJO_FUNCIONARIO	0.000007

Fuente: Elaboración propia

D.4. Importancia de variables en modelo de ordenamiento y selección optimizado

Variable	Importancia
DISTANCIA	0.590454
COD_NIVEL	0.134185
COD_COM_RBD	0.083701
EDAD_ALU	0.063209
IVM	0.060329
NSE	0.015315
TIPO_EE	0.013607

Variable	Importancia
SNED	0.011232
CUPOS_TOTALES	0.009529
COD_DEPE2	0.005377
ORI_RELIGIOSA	0.003637
PAGO_MENSUAL	0.002374
CON_COPAGO	0.002355
ES_MUJER	0.001656
PAGO_MATRICULA	0.001174
ALTO_RENDIMIENTO	0.001078
PRIORIDAD_MATRICULADO	0.000540
PRIORIDAD_HERMANO	0.000183
PRIORIDAD_EXALUMNO	0.000058
PRIORIDAD_HIJO_FUNCIONARIO	0.000008

Fuente: Elaboración propia

D.5. Importancia de variables en modelo de predicción a nivel agregado estándar

Variable	Importancia
VACANTES	0.866337
SNED	0.043065
IVM	0.037819
LON_RBD	0.024689
LAT_RBD	0.007285
COD_COM_RBD	0.005764
COD_DEPE2	0.005034

Variable	Importancia
COD_JOR	0.002615
COD_GRADO	0.002032
ORI_RELIGIOSA	0.001845
TIPO_EE	0.001362
CON_COPAGO	0.001017
PAGO_MENSUAL	0.000489
COD_NIVEL	0.000311
COD_ENSE	0.000185
COD_SEDE	0.000116
PAGO_MATRICULA	0.000037
COD_ESPE	0.000000

Fuente: Elaboración propia

D.6. Importancia de variables en modelo de predicción a nivel agregado optimizado

Variable	Importancia
VACANTES	0.590624
TIPO_EE	0.168860
IVM	0.076310
SNED	0.039586
LON_RBD	0.037992
COD_ENSE	0.018644
LAT_RBD	0.018625
COD_DEPE2	0.010795
COD_NIVEL	0.010104

Variable	Importancia
COD_COM_RBD	0.010058
COD_JOR	0.004457
COD_GRADO	0.003968
ORI_RELIGIOSA	0.003281
CON_COPAGO	0.002609
PAGO_MENSUAL	0.002019
PAGO_MATRICULA	0.001868
COD_SEDE	0.000198
COD_ESPE	0.000000

Fuente: Elaboración propia