



Universidad del Desarrollo
Facultad de Ingeniería

RETORNO SEGURO

Como evaluar el riesgo de los establecimientos educacionales en Chile,
ante una posible reapertura, durante la pandemia de Covid-19

POR: ROY ALBERTO BARRERA RICHARDS

Proyecto de grado presentado a la Facultad de Ingeniería de la Universidad
del Desarrollo para optar al grado académico de Magíster en Data Science

PROFESOR GUÍA:

Dra. LORETO BRAVO

Diciembre 2020

SANTIAGO

para Zuni, Rocío y María Esperanza.

Ustedes son mi motivo para seguir adelante.

Las Amo.

AGRADECIMIENTOS

Agradezco a Dios, a mi familia y a la universidad por haber podido tomar esta oportunidad de desarrollo profesional.

Además, va mi agradecimiento a profesores y compañeros de magister, por su amistad, constante apoyo y motivación. Destaco entre todos a la Doctora Loreto Bravo, por su paciencia y dedicación; y a aquellos compañeros que generosamente me ayudaron alguna vez con una investigación o con las revisiones de los preliminares de este documento: Israel Díaz, Gorki González, Alex Huerta, Cristóbal Eulufi y en especial Sergio Arancibia.

Por último a José Pérez Torres, por ser mi amigo, por acompañarme desde el principio y durante todo el proceso de magister; por no rendirse, ni haber permitido que yo lo hiciera.

Gracias Totales.

TABLA DE CONTENIDO

RESUMEN	5
1. INTRODUCCIÓN	6
2. HIPÓTESIS Y OBJETIVOS	10
3. DATOS Y METODOLOGÍA	11
3.1. DATOS.....	11
3.2. METODOLOGÍA.....	14
4. EXPLORACIÓN DE LOS DATOS Y APLICACIÓN DE LA METODOLOGÍA	17
4.1. LIMPIEZA PRELIMINAR.....	17
4.2. ANÁLISIS EXPLORATORIO.....	20
4.3. ALGORITMOS Y MODELOS.....	28
5. RESULTADOS	37
6. CONCLUSIONES	43
6.1. DISCUSIÓN.....	43
6.2. LIMITACIONES.....	43
6.3. TRABAJO FUTURO.....	44
7. BIBLIOGRAFÍA	45
ANEXOS	52
ANEXO A: EJES DEL PLAN RETORNO SEGURO DEL MINEDUC.....	52
ANEXO B: PELIGROSIDAD EN BASE A COMUNAS.....	54
ANEXO C: LA ENTROPÍA COMO MEDIDA DE MEZCLA SOCIAL.....	62
ANEXO D: ESTUDIO DE UN CASO DE EJEMPLO.....	69
ANEXO E: RESULTADOS OBTENIDOS.....	87

Resumen

El objetivo de este proyecto es determinar una forma para evaluar el riesgo de los establecimientos educacionales, actualmente cerrados en Chile, ante una eventual reapertura de estos, en el contexto de la epidemia de Covid-19

Para ello, la metodología propuesta consiste en definir un modelo que considere el número de alumnos matriculados y la entropía de Shannon como medida de la mezcla social que se da al interior del establecimiento, y así cuantificar el riesgo potencial de cada plantel.

Los datos provienen de la base de matrícula de estudiantes por establecimiento educacional 2019, de libre acceso, publicado en el sitio del ministerio de educación (Mineduc), complementándose con información de otras fuentes oficiales.

Como resultado, se logra establecer el indicador de riesgo potencial asociado a cada establecimiento, el cual puede utilizarse en conjunto con otras medidas, para extremar las precauciones y diseñar un plan de reapertura gradual que minimice la probabilidad de que nuevos contagios pudieran propagarse explosivamente en nuestras ciudades.

Los establecimientos educacionales con mayores riesgos potenciales tienden a estar en zonas urbanas, densamente pobladas donde la conectividad y el sistema de transporte urbano favorecen los desplazamientos entre comunas, siendo el emblemático Instituto Nacional (Santiago Centro), el más riesgoso de todos.

1. Introducción

La pandemia de Covid-19 ha originado profundos cambios en nuestra sociedad y en muchos aspectos de nuestra vida cotidiana. En todos los países del mundo, siguiendo las recomendaciones de la Organización mundial de la Salud, OMS, se han implementado una serie de protocolos sanitarios destinados a resguardar la salud de la población; muchas de estos protocolos provienen del aprendizaje obtenido tras la anterior gripe pandémica del virus Influenza A (H1N1) en 2009.

Las recomendaciones realizadas por la OMS se basan en experiencias de varios países, así como en estudios de las consecuencias sanitarias, económicas y sociales del cierre de escuelas. Otras medidas que se pueden adaptar en función de las situaciones epidemiológicas locales, los recursos disponibles y el rol social que cumplen las escuelas.

“Las escuelas deberían promover la higiene de las manos y una conducta cívica al toser y estornudar, y hacer acopio de los suministros apropiados. Asimismo, se aconseja limpiar y ventilar adecuadamente los locales, para disminuir los efectos de la aglomeración de personas.” (OMS / Medidas para los establecimientos escolares, 2009)

Respecto de decretar el cierre de escuelas, la OMS recomienda evaluar la conveniencia respecto de la disminución de nuevos casos, versus los costos sociales y económicos de esta medida ya que la conveniencia de la medida depende en gran medida del contexto.

Si bien el cierre proactivo de escuelas puede servir para frenar la transmisión dentro del establecimiento y disminuir la propagación del brote al resto de la comunidad de una zona determinada y en consecuencia el número de infecciones en el peor momento de la crisis, las autoridades sanitarias y escolares deben tener presentes que los costos económicos y sociales, derivados principalmente del ausentismo laboral de los padres que deben quedarse en casa para ocuparse de los niños, pueden ser muy elevados respecto de los beneficios potenciales de la medida.

“De las estimaciones de los estudios se desprende que los cierres escolares pueden provocar el ausentismo del 16% de la fuerza de trabajo, que habrá que sumar a los niveles ordinarios de ausentismo y a los provocados por la enfermedad. De todos modos, esas estimaciones varían considerablemente según los países en función de varios factores, en particular de la estructura de la fuerza laboral.” (OMS | Medidas para los establecimientos escolares, 2009)

Por otra parte, mantener esta medida de forma indefinida resulta especialmente preocupante en los sectores más vulnerables de la población, donde -por ejemplo- la escuela cumple un rol importante para la alimentación y para permitir trabajar a los padres. (La OMS y UNICEF instan a la pronta reapertura de las escuelas africanas, 2020)

Varios autores han estudiado los efectos que el cierre de escuelas supone no solo en la educación, sino en toda la sociedad.

“Desde la creación de los sistemas educativos modernos, éstos han tenido dos grandes funciones complementarias junto con la de custodia: la formación de las nuevas generaciones para ejercer una profesión y su preparación para desarrollar la vida en sociedad.”(Tarabini, 2020).

Además de los factores económicos, el cierre de colegios tiene fuertes impactos en cuanto a equidad social

“La paralización de la actividad presencial de las escuelas ha puesto en evidencia las amplias brechas que existen entre familias y escuelas para desarrollar su función educativa y ha evidenciado las múltiples formas de exclusión que limitan la función igualadora de debería cumplir la institución escolar” (Tarabini, 2020)

Todo esto sin mencionar una serie de efectos negativos sobre la salud física y mental de los estudiantes, principalmente niños, niñas y adolescentes, que no solo requieren instrucción y contenidos curriculares, sino que también interacciones con sus pares para el desarrollo de otros aspectos importantes de su personalidad. (Pietrobelli et al., 2020)

En vista de los antecedentes mencionado, es clara la necesidad de reabrir los colegios, el problema es entonces ¿Cómo hacer que el retorno a las clases presenciales sea seguro? En varios países ya se discute el tema (Chopra, 2020).

A la fecha de este reporte algunos países ya han decretado la reapertura parcial extremando las medidas de seguridad y adoptando estrictas medidas de seguridad.

En Chile, la medida de cerrar escuelas, fue decretada el 15 de marzo de 2020, (Gob.cl, 2020) tras conocerse los primeros casos de Covid-19, inicialmente por 2 semanas. Posteriormente se extendió (MINEDUC, 2020) y a la fecha, la medida aún se encuentra vigente en la mayor parte del territorio.

Se ha estimado que a Agosto de 2020, 3.6 millones de estudiantes del país ya habían perdido más el 44% de las clases presenciales del primer semestre (Said, 2020) y ya se han recibido varias solicitudes de escuelas para su reapertura (*Retorno a clases*, 2020)

A este respecto, el Ministerio de Educación, ya trabaja en un plan de retorno gradual bajo el nombre “Orientaciones para el año escolar 2020, preparando el regreso” (Rehbein, 2020) y cuyos ejes principales se presentan en el anexo A de este informe.

Debido a estos antecedentes, la relevancia, así como la urgencia de planificar una reapertura gradual de escuelas y un retorno seguro a las clases presenciales, son claros.

Surgiendo entonces, como objetivo principal de este proyecto, la necesidad poder cuantificar el riesgo potencial o “peligrosidad” asociado a cada escuela para ayudar a las autoridades a definir estrategia para reabrir los establecimientos educacionales de manera gradual y segura.

2. Hipótesis y Objetivos

En una pandemia existen muchos vectores que inciden en la propagación del virus, entre ellos la movilidad de los infectados entre una y otra zona geográfica; por esto las cuarentenas son de las medidas más efectivas que se pueden adoptar; y en este sentido, el cierre de escuelas surte el mismo efecto. En recientes investigaciones realizadas en nuestro país, se llegó a determinar que *“una reducción del 10% en la movilidad de las personas incide en una disminución entre un 15% y hasta 25% de nuevos casos”* de COVID-19. (Olivares et al., 2020).

La hipótesis en que se basa el presente trabajo es que si consideramos por una parte los desplazamientos entre la comuna de residencia de los alumnos y la comuna donde se ubica el establecimiento educacional; y por otra, como se mezclan los alumnos pertenecientes a cada comuna que se reúnen al interior de este (representatividad de cada comuna dada por la entropía), podríamos establecer un indicador de riesgo o peligrosidad para cada establecimiento.

De esta manera el objetivo principal de este proyecto es llegar a establecer el indicador de riesgo o peligrosidad para cada uno de los establecimientos educacionales del país.

En este sentido, se definen dos objetivos específicos:

1. Establecer un mecanismo de cálculo válido para el indicador de riesgo y
2. Obtener el valor del indicador de riesgo para cada establecimiento

3. Datos y Metodología

3.1. Datos

Como en todo proyecto de ciencia de datos, el primer gran desafío consiste en escoger las fuentes de datos adecuadas, completas, correctas y confiables. Para este estudio se trabajará con la información oficial, de libre acceso, disponible en los sitios web del Ministerio de Educación de Chile (Mineduc); complementándose con información de otras fuentes oficiales, como son Subsecretaría de Telecomunicaciones (SUBTEL), Subsecretaria de Desarrollo Regional y Administrativo (SUBDERE), así como el Instituto Nacional de Estadísticas (INE)

De esta forma los principales insumos serán los siguientes archivos de datos:

ARCHIVO 1	MATRÍCULA 2019.RAR
ORIGEN	MINEDUC
CONTENIDO DEL ZIP	<ul style="list-style-type: none">• 20191028_Matricula_unica_2019_20190430_PUBL.CSV• ER_Matricula_por_alumno_PUBL_MRUN_WEB.pdf
UBICACION	http://datos.mineduc.cl/dashboards/19776/descarga-bases-de-datos-de-matricula-por-estudiante/
ULTIMA CONSULTA	05/SEP/2020
DESCRIPCIÓN	Detalle de la matrícula de estudiantes por Establecimiento Educativo correspondiente al año 2019 (información más reciente del tema)
DATASET	<ul style="list-style-type: none">• 3.624.343 Filas (Estudiantes)• 35 columnas (features o características)

Tabla 1: Describe estructura de archivo de datos Detalle Matrícula.

ARCHIVO 2	Resumen_Matricula_EE_Oficial_2019.csv
ORIGEN	MINEDUC
UBICACION	http://datos.mineduc.cl/dashboards/20022/descarga-bases-de-datos-de-resumen-matricula-por-establecimiento/
ULTIMA CONSULTA	05/SEP/2020
DESCRIPCIÓN	Resumen de la matrícula de estudiantes por Establecimiento Educativo correspondiente al año 2019 (información más reciente del tema)
DATASET	<ul style="list-style-type: none"> • 16.236 Filas (Establecimientos Educativos) • 55 columnas (features o características)
DESCRIPCIÓN DE CAMPOS	ER_Resumen_Matricula_Oficial_EE_WEB_2019.pdf (misma ubicación y fecha de última consulta)

Tabla 2: Describe estructura de archivo de datos Resumen Matrícula

ARCHIVO 3	20191121_Directorio_Oficial_EE_2019_20190430_PUBL.csv
ORIGEN	MINEDUC
UBICACION	http://datos.mineduc.cl/dashboards/20015/descarga-bases-de-datos-directorio-de-establecimientos-educacionales/
ULTIMA CONSULTA	05/SEP/2020
DESCRIPCIÓN	Directorio de Establecimientos Educativos correspondiente al año 2019 (información más reciente del tema)
DATASET	<ul style="list-style-type: none"> • 16.236 Filas (Establecimientos Educativos) • 38 columnas (features o características)
DESCRIPCIÓN DE CAMPOS	ER_Directorio_Oficial_EE_WEB.pdf (misma ubicación y fecha de última consulta)

Tabla 3: Describe estructura de archivo de datos Directorio EE.

ARCHIVO 4	CUT_2018_v04.xls
ORIGEN	SUBDERE
UBICACION	http://www.subdere.gov.cl/documentacion/c%C3%B3digos-%C3%BAnicos-territoriales-actualizados-al-06-de-septiembre-2018
ULTIMA CONSULTA	05/SEP/2020
DESCRIPCIÓN	Códigos únicos territoriales para todas las regiones, provincias y comunas de Chile, Actualizado a septiembre de 2018 (Creación de la Región de Ñuble y aprobado el Decreto 1115, que define las abreviaturas oficiales para el nombre de las regiones)
DATASET	<ul style="list-style-type: none"> • 346 Filas (Comunas) • 7 columnas (features o características)
DESCRIPCIÓN DE CAMPOS	DTO-1115-SEPT_2018.pdf (misma ubicación y fecha de última consulta)

Tabla 4: Describe estructura de archivo de datos Códigos Territoriales.

ARCHIVO 5	1_1_POBLACION.xls
ORIGEN	INE
UBICACION	https://www.ine.cl/estadisticas/sociales/censos-de-poblacion-y-vivienda/poblacion-y-vivienda
ULTIMA CONSULTA	05/SEP/2020
DESCRIPCIÓN	Número de habitantes por Edad, Sexo y Comuna, en base al último Censo de Población y Vivienda realizado en 2017 (información más reciente del tema)
DATASET	<ul style="list-style-type: none"> • 35.939 Filas (Comunas/Edad) • 17 columnas (features o características)

Tabla 5: Describe estructura de archivo de datos Población según Censo 2017.

3.2. Metodología

Para establecer el indicador de riesgo potencial, la primera tarea es establecer un mecanismo de cálculo válido. La primera aproximación consiste en considerar el número de comunas distintas “que se mezclan” al interior del establecimiento; de esta forma, de manera intuitiva, se establece que: **a mayor número de comunas presente en la unidad educativa, existe un mayor riesgo potencial**, como se grafica a continuación y se explica con mayor profundidad en el anexo B.

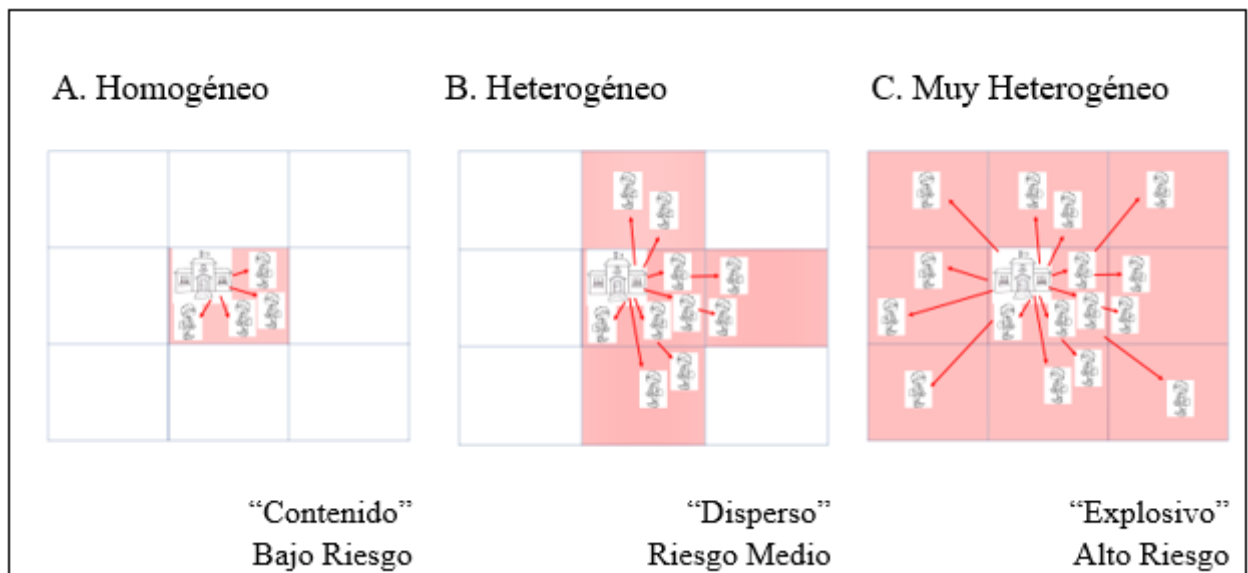


Figura 1: Riesgo en base al número de comunas distintas presentes en el establecimiento.

El problema de esta primera aproximación es que no considera el número de alumnos por comuna, es decir su “representatividad” o “peso” a la hora de calcular el indicador, por lo cual la validez del indicador podría cuestionarse.

En este contexto se revisó el trabajo “**Shopping mall attraction and social mixing at a city scale**” (Beiró et al., 2018), en el cual sus autores estudian la mezcla social que se produce al interior de los centros comerciales. Si pensamos que los centros comerciales y las escuelas son lugares que: provocan el desplazamiento de las personas (hacia y desde ellos); y donde se produce mezcla social (por comunas de residencia); estas similitudes nos permiten homologar los modelos; pero, dado que el presente trabajo no se enfoca en la segregación de los estudiantes o en analizar sus desplazamientos, se determinó que la medida más adecuada para lograr el objetivo propuesto es utilizar la entropía de Shannon como una medida de la diversidad social (Beiró et al., 2018); que nos permite obtener un indicador más confiable, que ya ha sido validado y que podemos usar para comparar escuelas, independiente de la cantidad de estudiantes y de comunas.

La fórmula de la entropía de Shannon está dada por
donde:

$$S_m = - \sum_{q=1}^Q p_q \cdot \log p_q$$

Fórmula 1: Entropía de Shannon

- m = Colegio que quiero estudiar
- Q = Total de comunas presentes al interior del colegio
- p_q = Probabilidad de la ocurrencia de la comuna q-esima (para todo q entre 1 y Q), representada por una fracción entre el número de estudiantes de esa comuna (n) sobre el total de estudiantes del colegio (N)
- S_m = Entropía de Shannon para el Colegio m

En resumen: se calculan las probabilidades para cada comuna, a continuación, se calculan los logaritmos para cada probabilidad, luego se multiplican ambos términos y finalmente, se realiza la sumatoria de todos los parciales para obtener el valor final. (Ver anexo C)

Pero ninguno de estos elementos por si solo es suficiente para establecer el indicador de riesgo potencial. Si vamos a la definición, encontramos que riesgo es la posibilidad que una amenaza determinada se haga realidad y la medida de la magnitud de los daños o el impacto que esta podría tener o producir (ISO9001:2015, ISO31000:2018)

Podemos operacionalizar esta definición estableciendo la siguiente ecuación:

$$\text{Riesgo} = \text{Impacto} * \text{Probabilidad}$$

(si se concreta la amenaza) (que ocurra realmente)

Fórmula 2: Definición de Riesgo en base a Impacto y Probabilidad.

De esta forma, ante la amenaza de que un nuevo contagio al interior del Establecimiento Educativo pueda propagarse a toda la comunidad; el impacto puede medirse a través del número de comunas que potencialmente se verían afectadas, mientras que la probabilidad podría medirse como la ocurrencia de un caso, es decir que un alumno, de una comuna particular se contagie, por la exposición a otros contagiados. Considerando la entropía como medida de esta mezcla social y reemplazando en la ecuación anterior, obtenemos:

$$\text{Riesgo} = \text{N}^\circ \text{ de Comunas} * \text{Entropía}$$

(donde residen los estudiantes) (que considera la probabilidad en base al n° de alumnos por comuna)

Fórmula 3: Definición de Riesgo en base al número de comunas y la entropía

Siendo esta la fórmula que aplicaremos para calcular el indicador de riesgo potencial para todos los establecimientos educacionales.

4. Exploración de los Datos y Aplicación de la Metodología

4.1. Limpieza preliminar

Para realizar el proceso de carga, limpieza y exploración de la data se utilizó un enfoque mixto que combinó el uso de herramientas web (Google Colab y Jupyter Notebooks) y herramientas locales (Oracle 11g Express Edition, Microsoft Excel). En primer término, se realizó la carga de los archivos y la depuración de estos.

En cuanto a la matrícula de los Establecimientos Educativos (EE):

1. Se logró cargar, comparar y validar los datos del detalle y el resumen de la matrícula de 2019, así como el directorio de establecimientos educativos, tras lo cual se trabajó en depurar la información del detalle de matrícula.
2. Se implementaron los siguientes criterios para filtrar la información :
 - Se incluye sólo **11.451** establecimientos que están en funcionamiento
 - Se excluyen **3.100** alumnos 'sin información' de su región de residencia
 - Se excluyen **21** alumnos que residen en la comuna 12202 : Antártica , porque en esta no existen establecimientos educativos (del Mineduc)
 - Se excluye **4** alumnos de los cuales no hay información de género.

Con estas correcciones el dataset se redujo de 3.624.343 a **3.620.758** registros, es decir, en menos de un **0,1%** respecto de la data original

3. Respecto del campo EDAD, se detectaron datos no numéricos y nulos, los cuales se corrigieron a partir del campo FECHA_DE_NACIMIENTO.
4. Se eliminaron las columnas consideradas como irrelevantes para el alcance del proyecto; reduciéndose a las 15 columnas que se muestran en la tabla 6.

#	CAMPO	TIPO	DESCRIPCION
1	NOMBRE_EST	Texto	Nombre del Establecimiento Educacional
2	COD_REGION_EST	Numérico	Código de Región donde se ubica el EE
3	COD_PROVINCIA_EST	Numérico	Código de Provincia del EE
4	COD_COM_EST	Numérico	Código de Comuna del EE
5	COMUNA_EST	Texto	Nombre de la Comuna del EE
6	DEPENDENCIA	Numérico	Tipo de Dependencia del EE
7	URBANO_RURAL	Numérico	Indicador Urbano o Rural para el EE
8	COD_ENSEÑANZA	Numérico	Código de Enseñanza que imparte el EE
9	CURSO_O_GRADO	Numérico	Curso o Grado del Alumno
10	COD_JORNADA	Numérico	Código de la Jornada del Alumno
11	GENERO_ALU	Numérico	Género del Alumno
12	EDAD_ALU	Numérico	Edad (en Años) del Alumno
13	COD_REGION_ALU	Numérico	Código de región de residencia del Alumno
14	COD_COM_ALU	Numérico	Código de Comuna de residencia del Alumno
15	COMUNA_ALU	Texto	Nombre de la Comuna de residencia del Alumno

Tabla 6: Describe estructura de tabla MAESTRA posterior a la limpieza y depuración de la data base.

En cuanto a las comunas:

Inicialmente se investigaba una aproximación basada en los desplazamientos entre comunas, en este contexto

1. Se logró combinar la información territorial de las comunas, con la información de superficie (Km²), la información de población (Censo2017) y las coordenadas de Latitud y Longitud para consolidar una base única.
2. Se realizó un análisis manual en base a información actualizada del Instituto Geográfico Militar para generar una matriz de adyacencia de las comunas y regiones del país.

Aunque finalmente no se utilizaron, estos “resultados intermedios”: la base territorial con coordenadas y ambas matrices de adyacencia, constituyen los primeros productos de datos derivados a partir de este proyecto.

4.2. Análisis Exploratorio

El análisis exploratorio se realizó contrastando la información de los archivos 1 y 2 de la fuente oficial (Mineduc), mencionados en la Sección 3.1, a saber:

- **20191028_Matricula_unica_2019_20190430_PUBL.CSV** (MAESTRO) y
- **Resumen_Matricula_EE_Oficial_2019.csv** (RESUMEN).

Se logró determinar algunos insights, los cuales se presentan a continuación.

1.- Establecimientos Educativos según su estado :

- ✓ Del total de EE, sólo el 74.3% (12.069) está en funcionamiento (abierto).
- ✓ De estos, sólo el 94,8% (11.451), registra alumnos matriculados.
- ✓ El universo de alumnos (individuos) matriculados en estos EE es 3.623.883.-

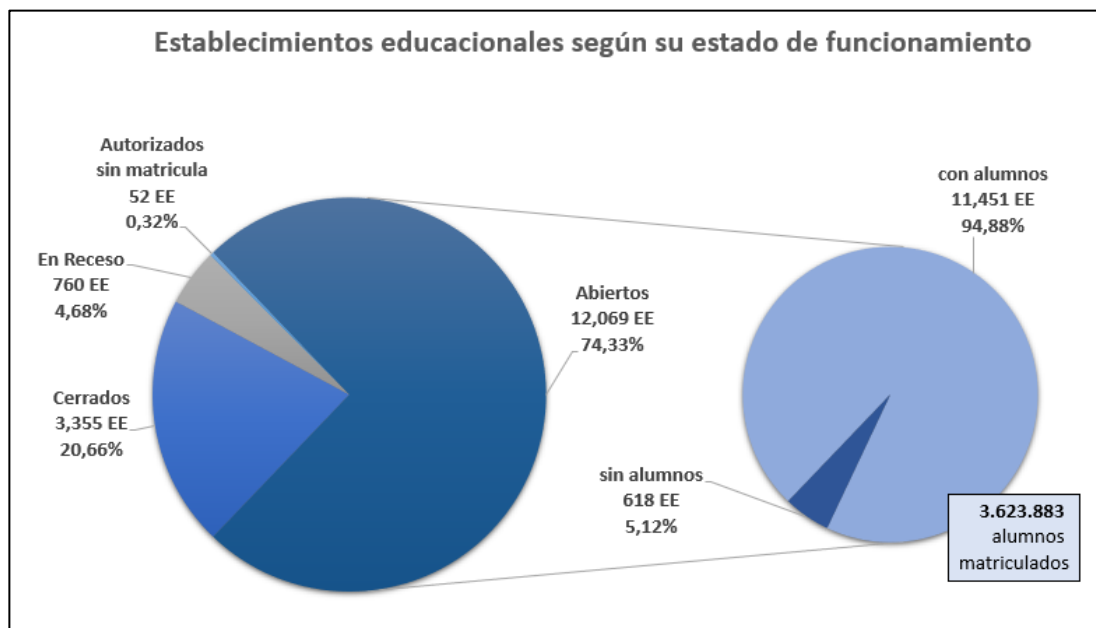


Figura 2: Distribución de Establecimientos Educativos según su estado de funcionamiento

2.- Alumnos en Establecimientos Educativos Abiertos, por tipo de Enseñanza

En total, entre los 11.451 EE Abiertos, con matrícula activa, sus 3.623.883 Alumnos y Alumnas; en su mayoría corresponden a alumnos niños de enseñanza básica (56%) y jóvenes de enseñanza media (25 %). Les siguen a estos infantes de educación parvularia (15%) y en mucho menor porcentaje, adultos que se están escolarizando (4%).

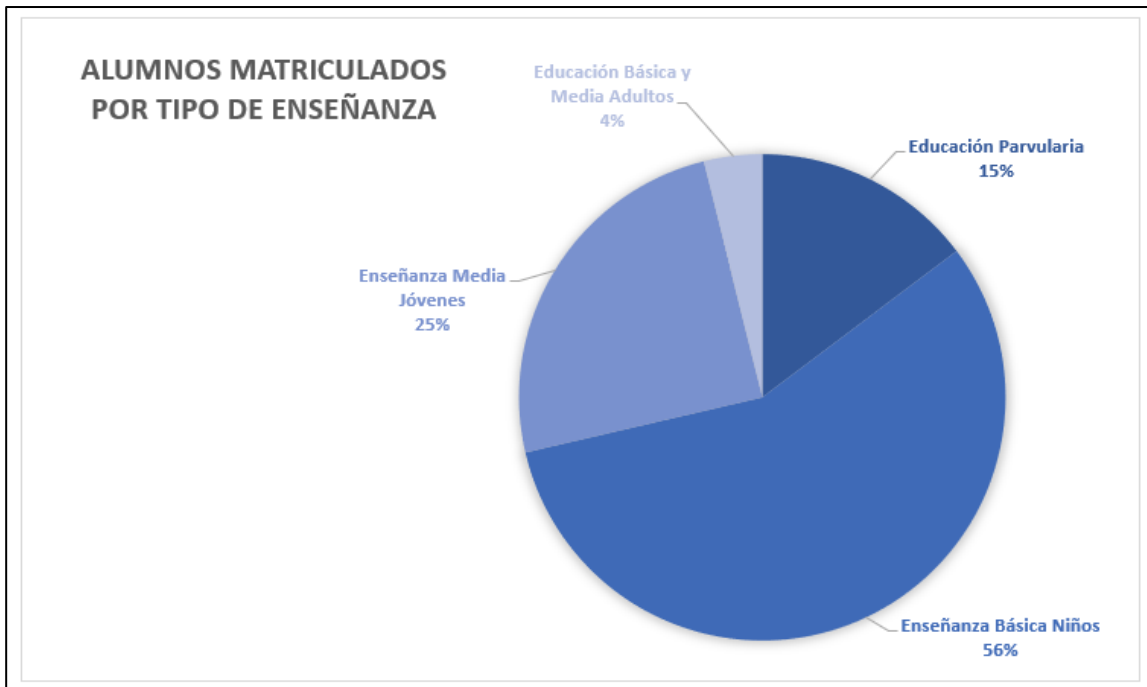


Figura 3: Distribución de Alumnos matriculados en EE abiertos, por tipo de Enseñanza

Vemos que la información del Mineduc no incluye alumnos de instituciones e educación superior (universidades, institutos profesionales, centros de formación técnica y escuelas matrices de las fuerzas armadas), razón por la cual, estos estudiantes quedan fuera del presente estudio.

3.- Alumnos Matriculados por área geográfica: Región y Comuna

La distribución de EE abiertos y alumnos matriculados, corresponde a las regiones con mayor densidad poblacional, según las cifras del Censo 2017.

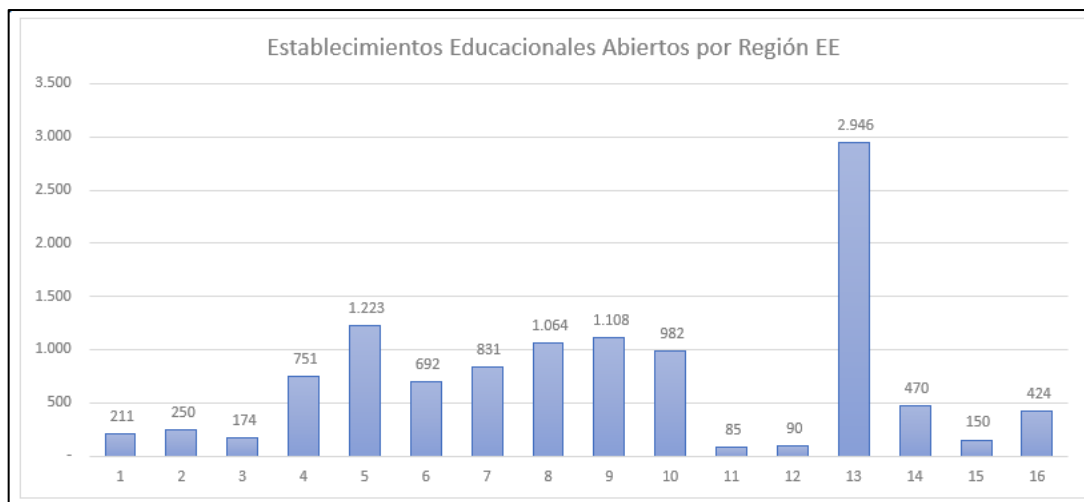


Figura 4: Establecimientos Educacionales Abiertos por Región

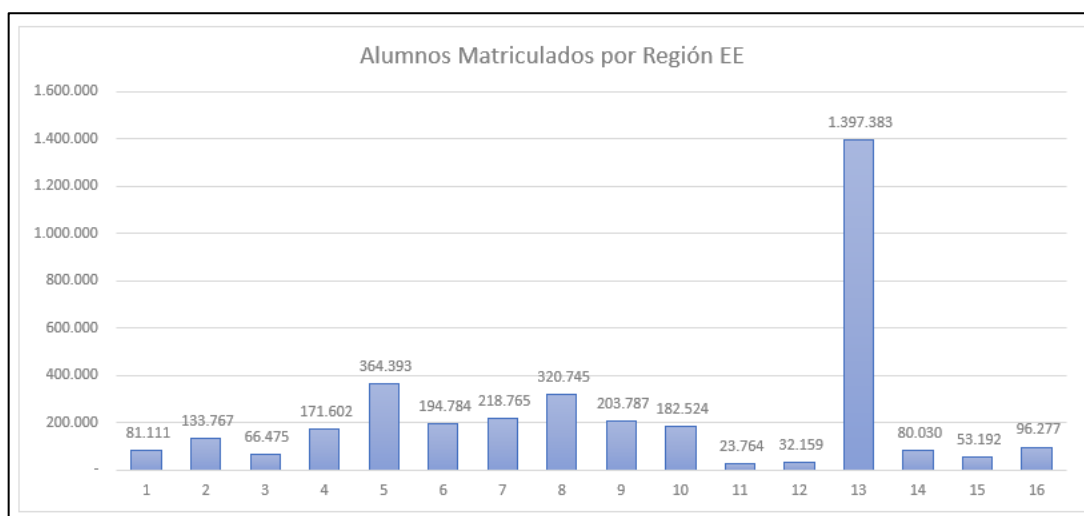


Figura 5: Alumnos Matriculados por Región

La mayor cantidad de estudiantes se registra en la Región Metropolitana (38.6%), seguida por las regiones de Valparaíso (10.1%), Bio Bío (8,9%), El Maule (6.0%), Araucanía (5.6%), Del Libertador Bernardo O'Higgins (5.4%), De los Lagos (5.0%) y Coquimbo (4.7%). Cada una de las restantes tiene menos del 4% de los estudiantes.

4.- Comunas con EE Abiertos por Región

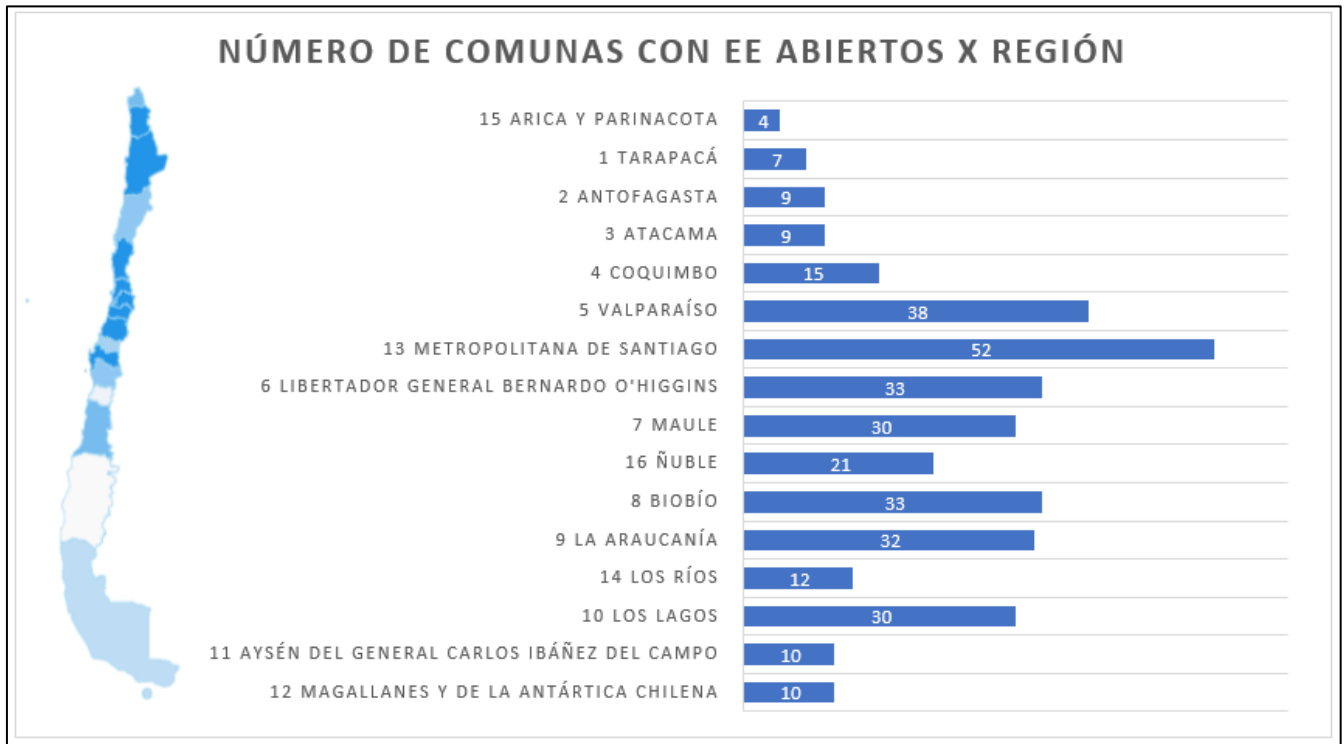


Figura 6: Comunas por Región con EE abiertos y con matrícula.

Chile se divide en 16 regiones y 346 comunas, no obstante, en la Antártica no hay EE abiertos dependientes del Mineduc, por lo cual se reducen a 345 comunas.

En la Región Metropolitana se encuentra la mayor cantidad de: comunas = 52, Establecimientos Educativos = 2.946 y alumnos matriculados = 1.397.383, por esta razón, el análisis se centra en primer lugar en esta región.

La Región Metropolitana comprende 52 comunas agrupadas en las siguientes 6 provincias: Santiago (32), Cordillera (3), Chacabuco (3), Maipo (4), Melipilla (5) y Talagante (5).

5.- Comunas de residencia de Alumnos para la región metropolitana

Aunque la gran mayoría de estudiantes matriculados (68%) declara tener residencia en la misma comuna donde se ubica su establecimiento, muchos otros (30%) declaran tener residencia en otras comunas de la misma región y sólo algunos (2%) declara que reside en una región distinta.

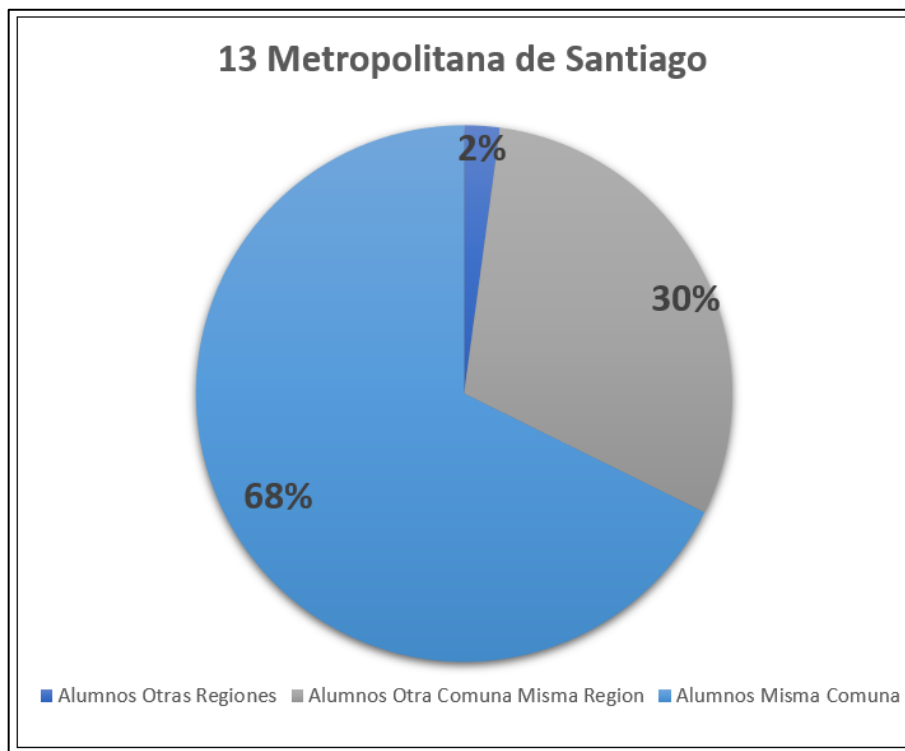


Figura 7: Composición de Alumnos por Comuna, Caso puntual Región Metropolitana.

6.- Comunas de residencia de Alumnos por región

Los estudiantes asisten mayoritariamente a establecimientos educacionales que se ubican en las mismas comunas de su residencia o en otras comunas dentro de la misma región.

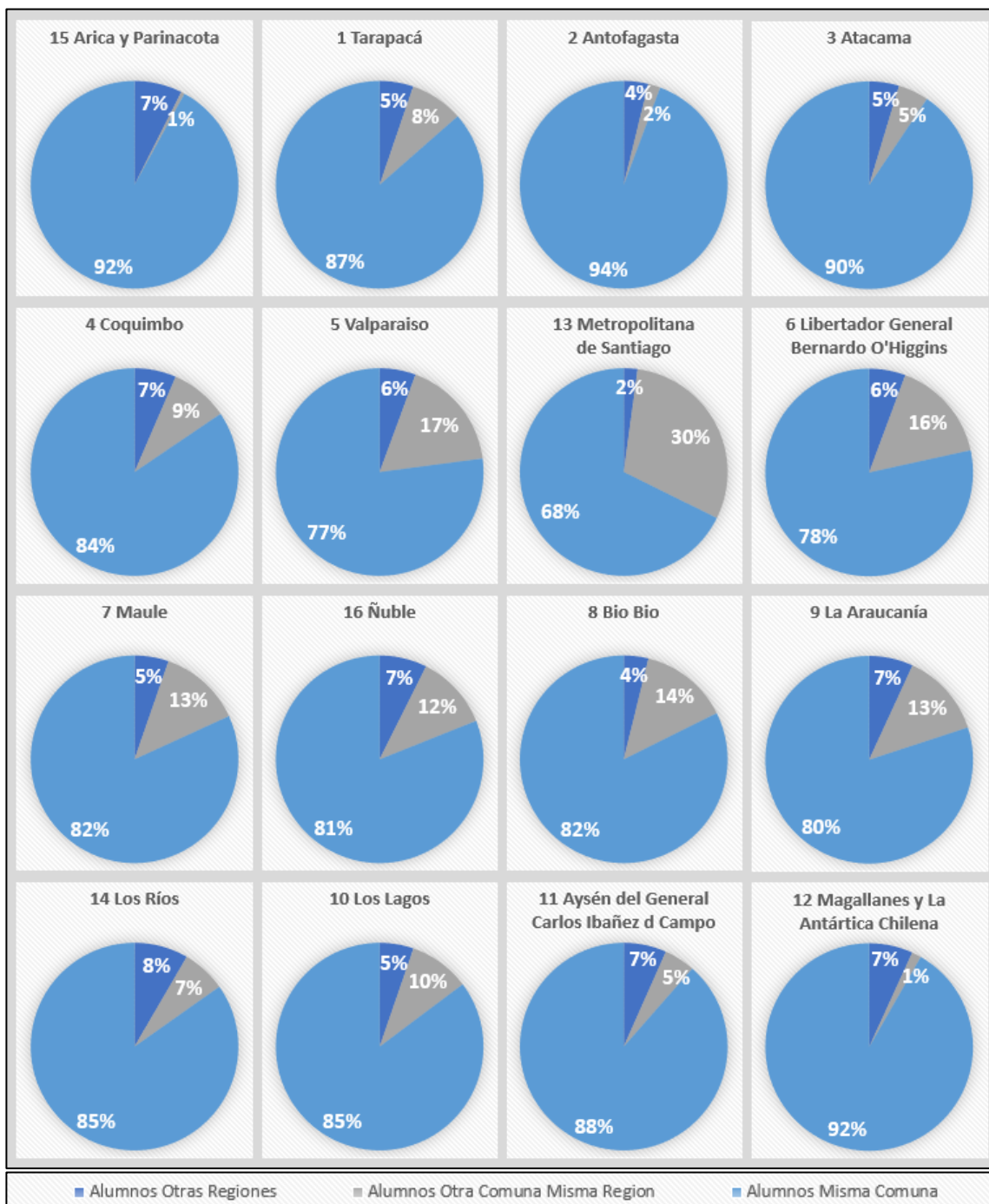


Figura 8: Composición de Alumnos por Comuna, todas las regiones del país.

Si damos una mirada más de cerca, por comuna, ordenando por cantidad de establecimientos educacionales abiertos, se observa la misma tendencia: 78% misma comuna, 19% otra comuna misma región y 3% otras regiones.

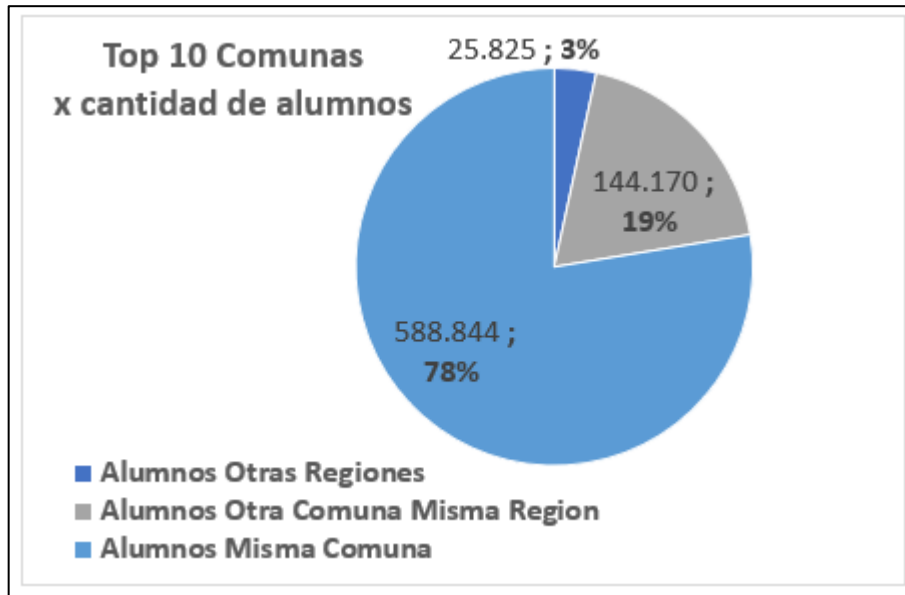


Figura 9: Comunas Top 10 x Cantidad de Alumnos (porcentajes).

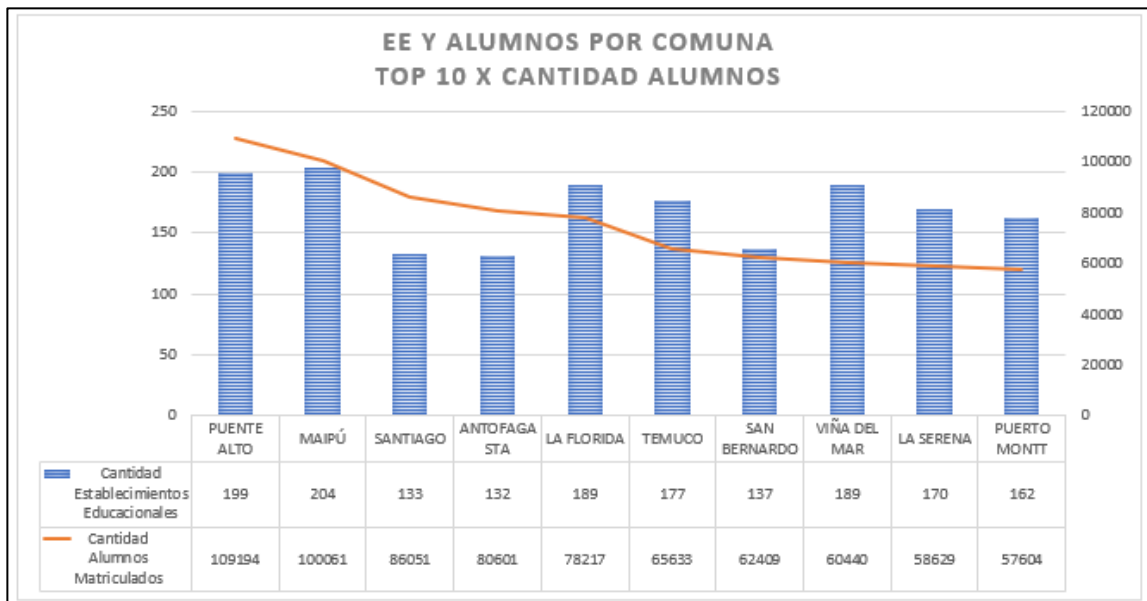


Figura 10: Comunas Top 10 x Cantidad de Alumnos (valores)

Tres de las diez comunas con más EE se ubican en la región metropolitana.

Lo mismo ocurre si damos una mirada por comuna, ordenando por cantidad de Alumnos Matriculados, se observa la misma tendencia: 81% misma comuna, 15% otra comuna misma región y 4% otras regiones.

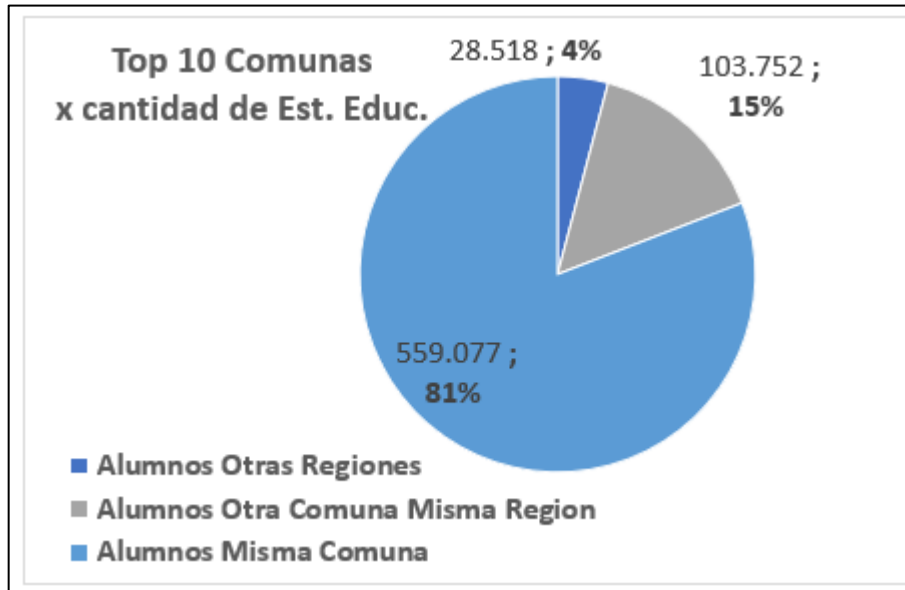


Figura 11: Comunas Top 10 x Cantidad de Establecimientos Educativos (porcentajes).

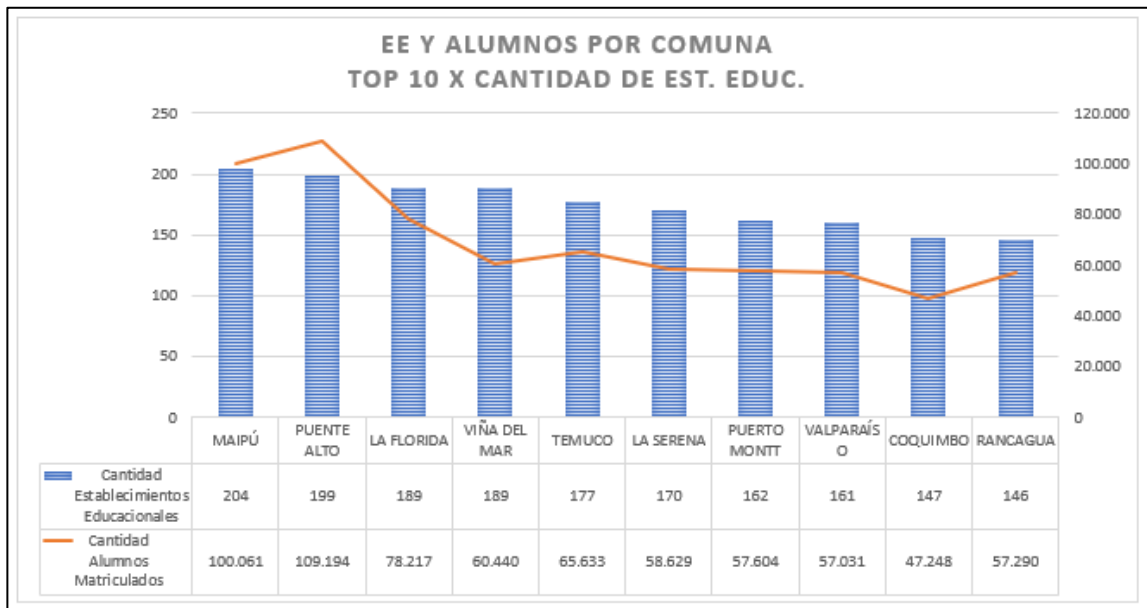


Figura 12: Comunas Top 10 x Cantidad de Establecimientos Educativos (valores).

Cinco de las diez comunas con más matriculados se ubican en la región metropolitana.

4.3. Algoritmos y Modelos

Para seguir adelante con el análisis exploratorio de los datos y alcanzar el objetivo principal del presente trabajo, se procede a comparar las relaciones entre la comuna donde se ubica este y las comunas de residencia de sus alumnos, mediante el desarrollo de un caso de ejemplo. Con este propósito, escojo la unidad educativa: *Colegio Compañía de María-Seminario*, Providencia, Santiago; código de establecimiento: 8950.

Este establecimiento cuenta con 853 alumnos matriculados, pertenecientes a 70 comunas diferentes; que supera las 32 comunas de la provincia de Santiago incluso las 52 comunas de la Región Metropolitana.

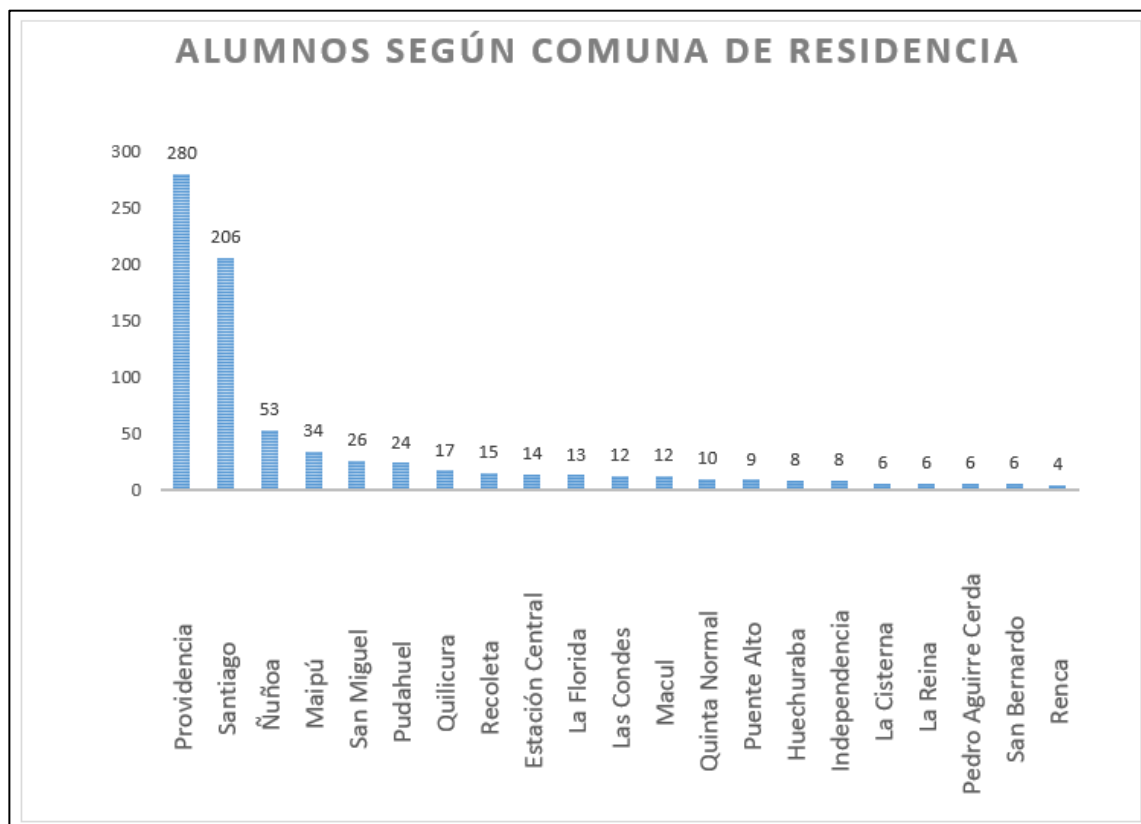


Figura 13: Alumnos por Comuna de Residencia, Colegio Compañía de María-Seminario, Providencia.

El detalle del análisis, incluyendo el código, se encuentra en el Anexo D.

La figura 16 muestra la distribución de los alumnos matriculados según su comuna de residencia, acotado a aquellas 20 comunas donde el porcentaje es igual o superior al 0.7%, que equivale a decir, al menos 5 alumnos.

Se aprecia como la mayoría de los alumnos matriculados (63%) en este establecimiento, se concentra en las siguientes comunas:

- 33% : Providencia, misma comuna donde se ubica el EE
- 24% : Santiago y
- 6% : Ñuñoa, ambas comunas vecinas, con grandes facilidades para el desplazamiento entre ellas (locomoción pública, avenidas)

En menor porcentaje (27%) los alumnos residen en las siguientes comunas:

- Maipú con un 4%,
- San Miguel y Pudahuel , ambas con un 3% cada una,
- Quilicura, Recoleta, Estación Central y la Florida con un 2% cada una y
- Las Condes, Macul, Quinta Normal, Puente Alto, Independencia, Huechuraba, La Reina, La Cisterna, Pedro Aguirre Cerda y San Bernardo con 1% cada una

El 10% restante se reparte en 50 comunas distintas, de las cuales:

- 17 son de la región metropolitana con 36 alumnos equivalentes al 4% ,
- 14 son de las regiones 5 y 6, adyacentes a la región metropolitana, con 18 alumnos equivalentes al 2% y

- 19 son de regiones más lejanas que no tienen comunicación terrestre directa con la región metropolitana, con 34 alumnos equivalentes al 4% restante

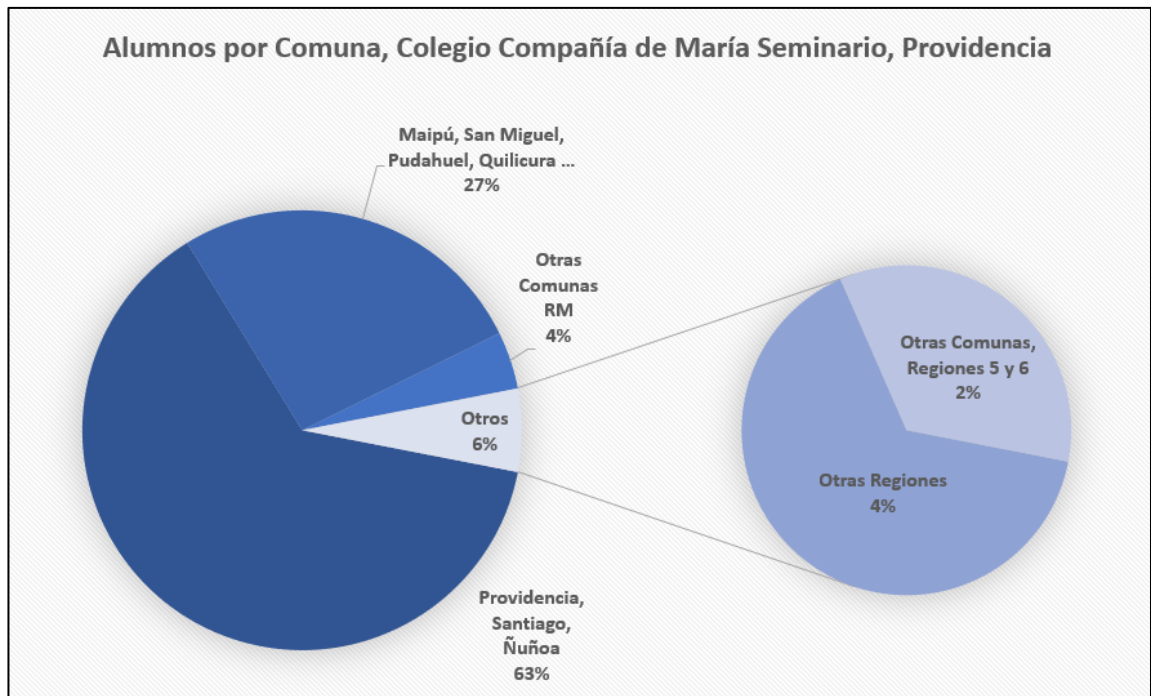


Figura 14: Alumnos por Comuna de Residencia, Colegio Compañía de María-Seminario, Providencia.

Tras validar las comunas y las regiones de residencia de los alumnos, se intenta interpretar los resultados obtenidos que se muestran en la figura 17, sobre todo en lo referente a alumnos que residen en otras regiones, a gran distancia de la RM.

Revisando la documentación encontramos la región del alumno se determina en función de la comuna y que este es un dato auto declarado y voluntario; por lo cual podría contener información imprecisa, ya que podría haber sido interpretado como PROCEDENCIA (donde nació el alumno) en lugar de RESIDENCIA, por lo que constituye una fuente de distorsión para el análisis.

La figura 15 a continuación intenta demostrar gráficamente la imposibilidad práctica del desplazamiento diario desde otras regiones para asistir a clases en el establecimiento.

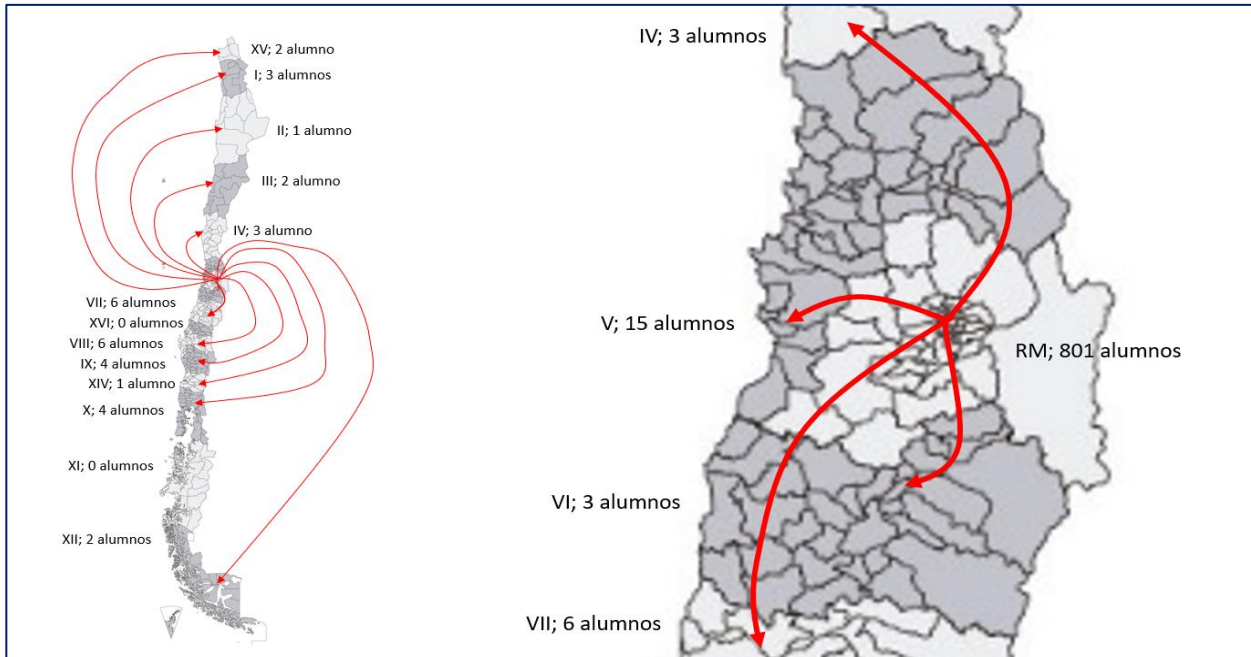


Figura 15: Desplazamiento teórico de los estudiantes de regiones para asistir al establecimiento del ejemplo

Se decide, por tanto, tratar estos datos como outliers y eliminarlos del dataset, antes de continuar. Es necesario definir un criterio adecuado para realizar esta tarea. Se analiza la factibilidad de hacerlo calculando las distancias entre la comuna de establecimiento y la del alumno, para luego eliminar los que superen un valor dado.

Esta aproximación además ineficiente es costosa de implementar y no considera la accidentada geografía de nuestro país o la infraestructura y los sistemas de transporte públicos existentes en Santiago y las otras grandes capitales regionales que permitirían desplazarse al alumno distancias considerables con relativa facilidad.

Se opta entonces por un criterio diferente, más sencillo de interpretar e implementar, que además aprovecha el conocimiento de la regionalización del país, al establecer un grafo no dirigido entre las regiones que luego se operacionaliza como una matriz de adyacencia. De esta forma, si la región del establecimiento y la región del alumno no son la misma, ni adyacentes, entonces se considerará como dato inválido.

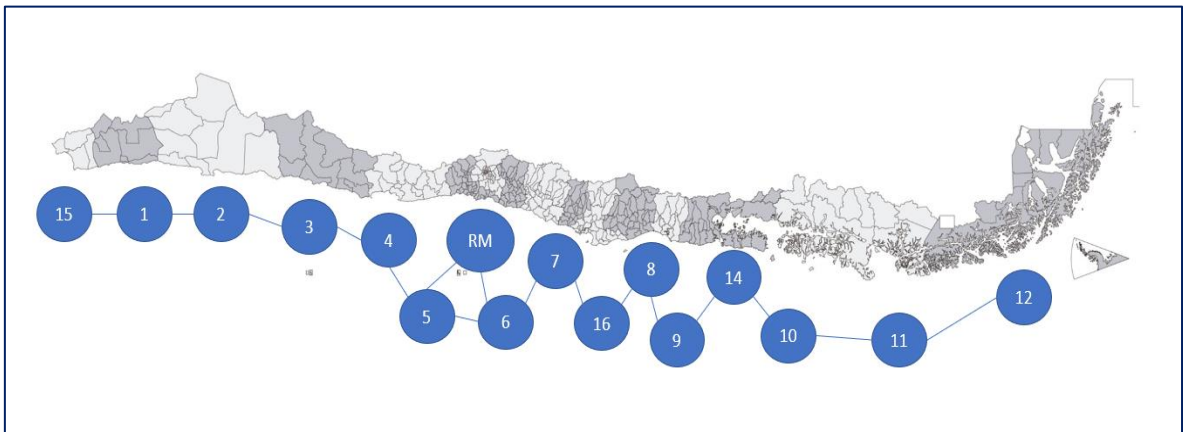


Figura 16: Grafo no Dirigido para las regiones de Chile. Comunas ordenadas de norte a sur.

Región	R15	R01	R02	R03	R04	R05	R13	R06	R07	R16	R08	R09	R14	R10	R11	R12
R15	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R01	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
R02	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
R03	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0
R04	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0
R05	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0
R13	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0
R06	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0
R07	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0
R16	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0
R08	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0
R09	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0
R14	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0
R10	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0
R11	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
R12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1

Tabla 7: Matriz de Adyacencia de Comunas (1:Son Adyacentes; 0: no lo son).

Tras la aplicación del filtro por regiones adyacentes al caso de estudio del ejemplo, la cantidad de alumnos se redujo de 853 a 819, es decir, el 4% que se esperaba, al tiempo que la cantidad de comunas se redujo de 70 a 51 y las regiones de 14 a 3 (RM, 5 y 6)

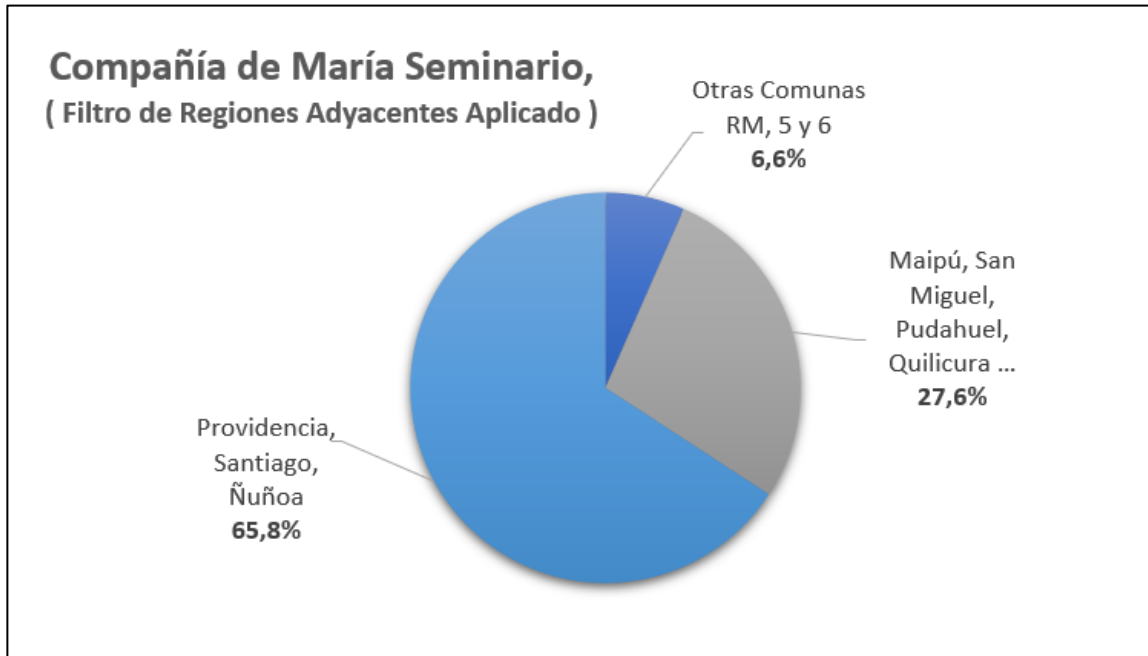


Figura 17: Alumnos Colegio Compañía de María-Seminario, tras aplicar filtro de comunas

Si bien este criterio no es perfecto, el resultado que se obtiene hace mucho más sentido, considerando que para un establecimiento de la Región Metropolitana podría ser posible tener alumnos de las regiones 5 y 6. Se ha logrado descubrir y solucionar un problema con los datos que pudo haber pasado desapercibido, generando “distorsión” o “ruido” en la determinación del indicador de riesgo potencial, pudiendo incluso llegar a invalidarlo.

De esta forma se concluye la etapa de análisis exploratorio de los datos y se procede a la etapa de algoritmos y modelos donde realizaremos el cálculo del indicador de riesgo o peligrosidad en primer lugar para el establecimiento del ejemplo y posteriormente para el resto de los Establecimientos Educativos de la muestra.

En la Sección 3.2 se comentaba que una primera aproximación natural o intuitiva para determinar el indicador de riesgo potencial se podría establecer en función del número de comunas que se mezclan al interior de este (ver Anexo B). Empleando esta aproximación, el indicador de riesgo potencial estaría en directa relación con el “alcance de comunas” del establecimiento.

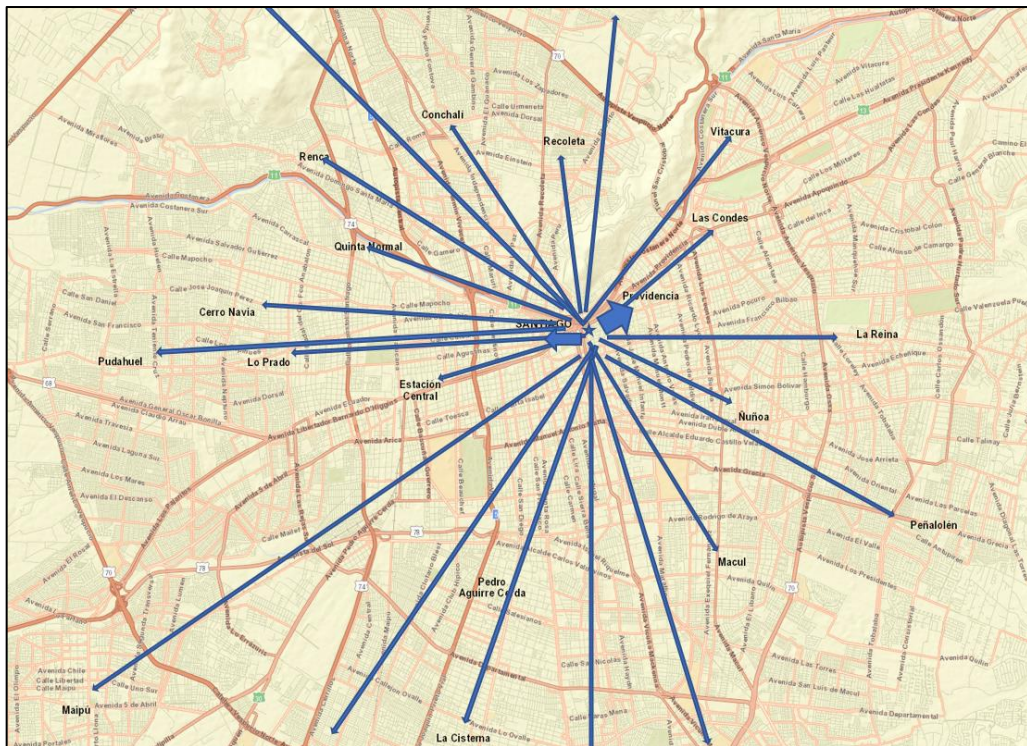


Figura 18: “Alcance de Comunas” que tiene el establecimiento del ejemplo, ubicado en Providencia

Para el caso del ejemplo, al estar ubicado en una comuna céntrica, en una zona densamente poblada, el colegio tendría una composición “**muy heterogénea**”, por lo que un nuevo contagio podría tener una propagación “**explosiva**”, catalogándose por tanto al establecimiento como de “**Alto Riesgo**”

Llevado a números el primer indicador sería: **51**, que equivale al número de comunas distintas que es posible identificar entre los alumnos matriculados en 2019, (después de aplicar el filtro de regiones adyacentes).

Luego se calcula el segundo indicador de mezcla social, en base a la entropía, como se muestra en la figura a continuación, obteniéndose el valor: **1,035**.

$$S_m = - \sum_{q=1}^Q p_q \cdot \log p_q$$

Código EE	Cód Comuna EE	Cód Comuna Alumnos	Cant de Alumnos	Total Alumnos EE	Probabilidad	Logaritmo	Prob x Log P	Sumatoria	Orden
8950	13123	13123	280	819	0,341880342	-0,46612587	-0,15935927	-0,15935927	1
8950	13123	13101	206	819	0,251526252	-0,599416681	-0,15076903	-0,3101283	2
8950	13123	13120	53	819	0,064713065	-1,189008032	-0,07694435	-0,38707266	3
8950	13123	13119	34	819	0,041514042	-1,381804985	-0,05736431	-0,44443697	4
8950	13123	13130	26	819	0,031746032	-1,498310554	-0,04756541	-0,49200238	5
8950	13123	13111	1	819	0,001221001	-2,913283902	-0,00355712	-1,02091228	47
8950	13123	13115	1	819	0,001221001	-2,913283902	-0,00355712	-1,0244694	48
8950	13123	13132	1	819	0,001221001	-2,913283902	-0,00355712	-1,02802652	49
8950	13123	13601	1	819	0,001221001	-2,913283902	-0,00355712	-1,03158365	50
8950	13123	13605	1	819	0,001221001	-2,913283902	-0,00355712	-1,03514077	51

Tabla 8: Cálculo la entropía que tiene el Colegio del ejemplo

Finalmente, se aplica la fórmula definida en la sección anterior, multiplicando ambos valores parciales para obtener el Indicador de Riesgo para el establecimiento del ejemplo:

$$\text{Riesgo} = \text{N}^\circ \text{ de Comunas} * \text{Entropía}$$

(donde residen los estudiantes) (que considera la probabilidad en base al n° de alumnos por comuna)

$$R_{\text{Cía. de María Seminario}} = 51 * 1,035 = 52.785$$

$$R_{\text{Cía. de María Seminario}} \approx 52.8$$

De esta forma hemos podido establecer un mecanismo de cálculo válido para el indicador de riesgo, cumpliéndose el primero de los objetivos específicos propuestos.

A continuación, se aplica el mismo cálculo para el resto de los Establecimientos del país:

1.- Se cruza la tabla de Matrículas (11.451 EE) con la matriz de adyacencia de Regiones, dejando fuera 6 establecimientos, donde todos sus alumnos declaran residencia en comunas y regiones distintas a aquellas donde se ubica el establecimiento.

Código EE	Nombre EE	Region EE	Comuna EE	Region Alumno	Comuna Alumno	Cant Alumnos
4479	ESCUELA CHEQUENAL	8	NACIMIENTO	10	PUERTO MONTT	2
5364	ESCUELA MANCHURIA	9	CURACAUTÍN	13	EL MONTE	1
6578	ESCUELA PARTICULAR LA MONTANA	9	CARAHUE	5	QUILPUÉ	1
7545	ESCUELA RURAL PUTRIHUE	10	RÍO NEGRO	13	TALAGANTE	1
33327	TERNURA	4	SALAMANCA	13	SANTIAGO	17
36141	SALA CUNA ALTUE	14	VALDIVIA	13	SANTIAGO	19

Tabla 9: EE descartados por el criterio de regiones adyacentes.

2.- Para los 11.445 EE restantes se realiza el cálculo de la entropía de Shannon, de la forma antes descrita obteniéndose valores como los que se muestran a continuación:

#	Código EE	Código Comuna EE	Alumnos Matriculados	a) Comunas Alumnos	b) Entropía	Indicador de Riesgo a x b
1	1	15101	627	8	0,116	0,928
2	2	15101	211	1	0	0
3	3	15101	407	1	0	0
4	4	15101	918	9	0,097	0,873
5	5	15101	919	4	0,062	0,248
6	7	15101	1051	8	0,107	0,856
5341	8945	13123	581	31	1,092	33,852
5342	8950	13123	819	51	1,035	52,785
5343	8953	13132	634	30	0,674	20,22
11440	41135	13601	137	11	0,549	6,039
11441	41173	4105	121	7	0,245	1,715
11442	41179	4301	70	3	0,089	0,267
11443	41239	6101	7	1	0	0
11444	41264	13120	48	13	0,793	10,309
11445	41280	8301	77	1	0	0

Caso ejemplo

Tabla 10: Calculando el indicador de riesgo para todos los EE no descartados

De esta forma hemos podido obtener el valor del indicador de riesgo para cada establecimiento, cumpliéndose el segundo de los objetivos específicos propuestos.

5. Resultados

Habiéndose completado ambos objetivos específicos propuestos:

1. Establecer un mecanismo de cálculo válido para el indicador de riesgo y
2. Obtener el valor del indicador de riesgo para cada establecimiento

se ha logrado cumplir también el objetivo principal definido en el principio de este documento: “establecer el indicador de riesgo o peligrosidad para cada uno de los establecimientos educacionales del país”.

De esta forma el producto de datos final, derivado de este análisis, es un dataset generado a partir del Directorio de Establecimientos Educacionales y el indicador de riesgo potencial o peligrosidad asociado a cada uno de ellos.

No obstante, en el Anexo E se incluyen otros resultados; productos de datos adicionales, derivados del proceso y de la investigación realizada:

- ✓ Consolidación de las comunas en una única base con información territorial, superficie (KM2), población por género y edad (Censo) y coordenadas GPS.
- ✓ Matriz de adyacencia de comunas de Chile
- ✓ Matriz de adyacencia de regiones de Chile

A continuación se muestran y comentan los resultados más relevantes obtenidos, respecto del objetivo principal; Indicador de riesgo calculado para todos los EE

1.- Distribución de Establecimientos Educativos con indicador de riesgo > 0:

Los valores obtenidos para del indicador de riesgo van entre 0 y 111.15. Considerando solamente a aquellos 10.080 Establecimientos con valores > 0, obtenemos que la gran mayoría de estos, 9779 (97%) se clasificarían como de bajo riesgo (valores inferiores a 30), mientras que 273 (2,7%) se clasificación como de riesgo medio y tan solo 28 (0,3%) de clasificarían como de Alto Riesgo (valores superiores a 60)

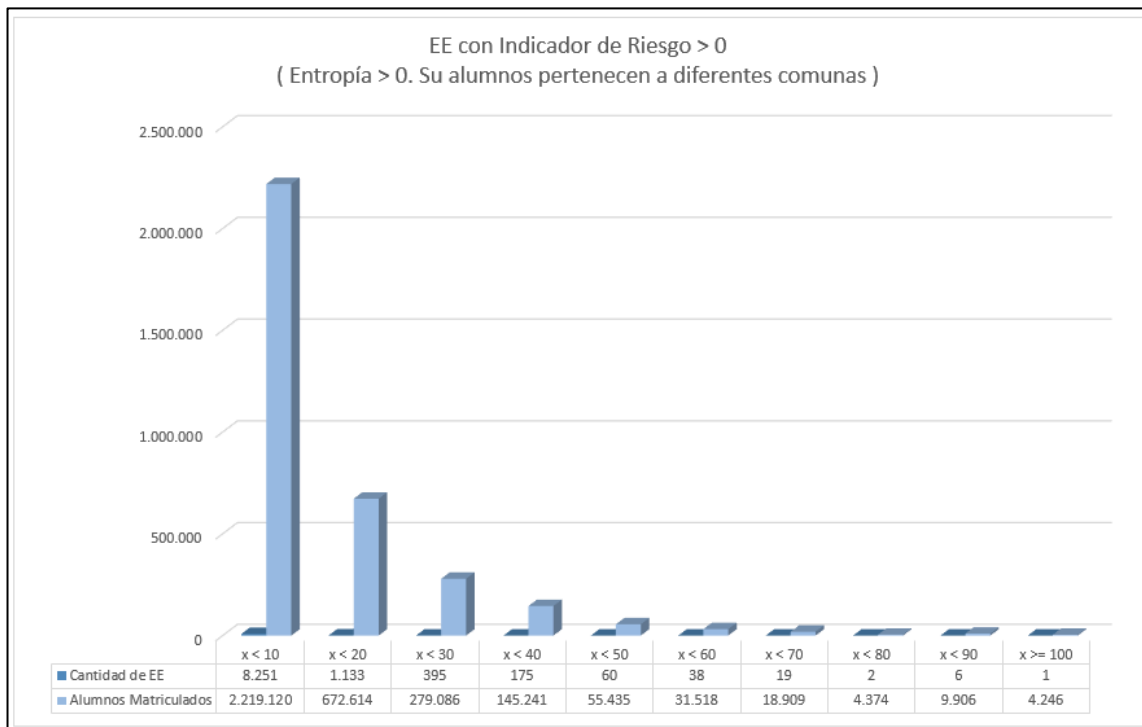


Figura 19: Distribución de EE con indicador de Riesgo > 0.

2.- Establecimientos Educativos con indicador de riesgo más alto (>60)

Revisando el listado de los colegios TOP respecto de su peligrosidad, obtenemos que 1 de ellos se encuentra en la Región del Bío Bío, 2 en la Región de O'Higgins y los 25 restantes en la Región Metropolitana.

#	Codigo Est. Educ.	Nombre de Establecimiento Educativo	Codigo Comuna	Nombre Comuna	Alumnos Matriculados	Comunas Distintas	Entropía Shannon	Indicador de Riesgo
1	8485	LICEO INSTITUTO NACIONAL	13101	SANTIAGO	4246	75	1,482	111,15
2	25818	COLEGIO PART. ADULTOS INSTITUTO ICEL	13101	SANTIAGO	1935	64	1,395	89,28
3	8499	INTERNADO NACIONAL BARROS ARANA	13101	SANTIAGO	1260	63	1,38	86,94
4	8928	LICEO JOSE VICTORINO LASTARRIA	13123	PROVIDENCIA	2573	61	1,408	85,888
5	10715	INSTITUTO SUBTTE.LUIS CRUZ MARTINEZ	13601	TALAGANTE	618	57	1,468	83,676
6	31030	ESCUELA PRE MILITAR HEROES DE LA CONCEPCION	13109	LA CISTERNA	1666	58	1,415	82,07
7	8491	LICEO DE APLICACION RECTOR JORGE E SCHNE	13101	SANTIAGO	1854	58	1,387	80,446
8	8487	LICEO JAVIERA CARRERA	13101	SANTIAGO	2612	54	1,411	76,194
9	8927	LICEO CARMELA CARVAJAL DE PRAT	13123	PROVIDENCIA	1762	51	1,426	72,726
10	8603	CENTRO EDUCATIVO SALESIANOS ALAMEDA	13101	SANTIAGO	1693	52	1,308	68,016
11	8748	COLEGIO INSTITUTO ALONSO DE ERCILLA	13101	SANTIAGO	1939	54	1,24	66,96
12	2552	INSTITUTO POLITECNICO	6310	SANTA CRUZ	1195	65	1,006	65,39
13	8930	LICEO B 42 TAJAMAR	13123	PROVIDENCIA	978	45	1,436	64,62
14	15777	COLEGIO JUAN LUIS VIVES	6101	RANCAGUA	558	63	1,022	64,386
15	17607	COLEGIO INGLÉS BRITISH ROYAL SCHOOL	8101	CONCEPCIÓN	1041	59	1,089	64,251
16	24473	COLEGIO TEC.HOTELERIA Y GASTRONOMIA ACHIGA CO	13114	LAS CONDES	528	44	1,446	63,624
17	8643	INSTITUTO COMERCIAL BLAS CANAS	13101	SANTIAGO	702	49	1,282	62,818
18	31167	CENTRO EDUCACIONAL INTEGRAL DE ADULT ACUARIO	13101	SANTIAGO	207	42	1,486	62,412
19	9472	LICEO POLITECNICO SAN LUIS	13130	SAN MIGUEL	1393	51	1,222	62,322
20	24840	COLEGIO DE ADULTOS INSTITUTO TABANCURA SEDE LONDRES	13101	SANTIAGO	235	44	1,405	61,82
21	8926	LICEO N°7 DE PROVIDENCIA LUISA SAAVEDRA	13123	PROVIDENCIA	1257	43	1,435	61,705
22	8492	LICEO MANUEL BARROS BORGONO	13101	SANTIAGO	831	45	1,37	61,65
23	20310	COLEGIO DE ADULTOS PULMAHUE	13101	SANTIAGO	296	44	1,401	61,644
24	8627	COLEGIO SANTA MARIA DE SANTIAGO	13101	SANTIAGO	2059	65	0,945	61,425
25	8925	LICEO POLIVALENTE ARTURO ALESSANDRI P.	13123	PROVIDENCIA	1024	43	1,414	60,802
26	8756	COLEGIO EXCELSIOR	13101	SANTIAGO	1395	57	1,066	60,762
27	9073	LICEO CARMELA SILVA DONOSO	13120	ÑUÑO A	1103	55	1,1	60,5
28	9077	COMPLEJO EDUC.MUN. BRIGIDA WALKER.	13120	ÑUÑO A	475	45	1,334	60,03

Tabla 11: Establecimientos Educativos con los valores más altos para el indicador de Riesgo.

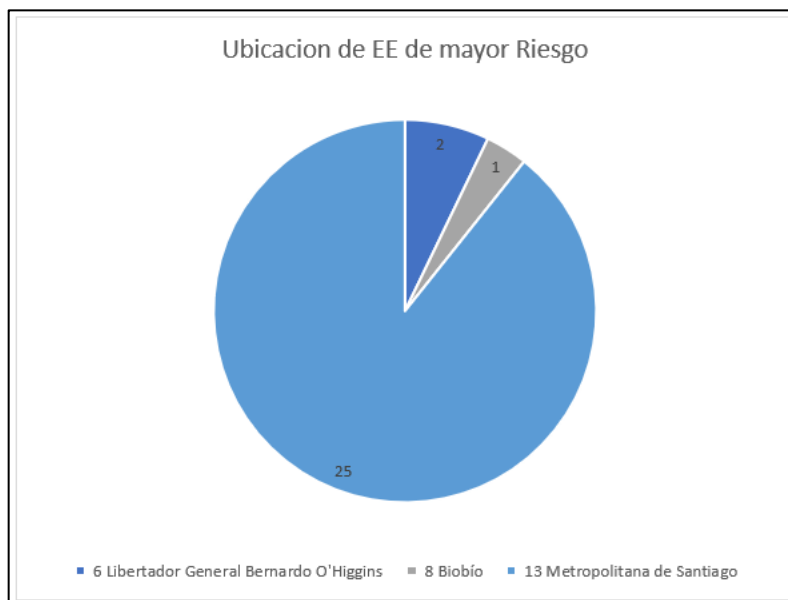


Figura 20: Establecimientos Educativos con los valores más altos para el indicador de Riesgo.

3.- Establecimientos Educacionales con indicador de riesgo = 0:

En el extremo opuesto de los datos, tenemos a aquellos establecimientos con indicador de Riesgo = 0. Al respecto 2 comportamientos que destacar: En primer lugar, establecimientos con muy escasa matrícula: solo 1 alumno.

#	Codigo Est. Educ.	Nombre de Establecimiento Educacional	Codigo Comuna	Nombre Comuna	Alumnos Matriculados	Comunas Distintas	Entropía Shannon	Indicador de Riesgo
11408	5362	ESCUELA RIO BLANCO	9203	CURACAUTÍN	1	1	0	0
11409	5365	ESCUELA JOSE CHANIN SILHY	9203	CURACAUTÍN	1	1	0	0
11410	6867	ESC.RURAL JORGE ROBERTO FONTANNAZ BLUAS	14106	MARIQUINA	1	1	0	0
11411	8075	ESCUELA RURAL QUILAR	10202	ANCUD	1	1	0	0
11412	8080	ESCUELA RURAL COGOMO	10202	ANCUD	1	1	0	0
11413	8141	ESCUELA RURAL SAN JOSE	10209	QUEMCHI	1	1	0	0
11414	8155	ESCUELA RURAL DALLICO	10205	DALCAHUE	1	1	0	0
11415	5984	ESC. BASICA LO SCHMITH	9103	CUNCO	1	1	0	0
11416	7502	ESCUELA PARTICULAR N.181 LAS GAVIOTAS	10302	PUERTO OCTAY	1	1	0	0
11417	6522	ESCUELA BASICA PUYANGUE	9102	CARAHUE	1	1	0	0
11418	6527	ESCUELA BASICA ALTO YUPEHUE	9102	CARAHUE	1	1	0	0
11419	6547	ESCUELA PARTICULAR CENTRO LONCOYAMO	9102	CARAHUE	1	1	0	0
11420	7318	ESCUELA PARTICULAR HUEIMEN	14203	LAGO RANCO	1	1	0	0
11421	6567	ESCUELA PARTICULAR JOSE ABELARDO NUNEZ	9111	NEUEVA IMPERIAL	1	1	0	0
11422	11425	ESCUELA CIFUENTES BAJO	8205	CURANILAHUE	1	1	0	0
11423	11543	ESCUELA PARTICULAR TRIPAYANTE	14106	MARIQUINA	1	1	0	0
11424	8277	ESCUELA RURAL SAN JUAN	10208	QUELLÓN	1	1	0	0
11425	8304	ESCUELA RURAL RIO AMARILLO	10401	CHAITÉN	1	1	0	0
11426	8322	ESCUELA RURAL HUINAY	10403	HUALAIHUÉ	1	1	0	0
11427	8330	ESCUELA PARTICULAR N.252 AULEN	10403	HUALAIHUÉ	1	1	0	0
11428	4859	ESCUELA BASICA CHUPALLAR	8111	TOMÉ	1	1	0	0
11429	4907	ESCUELA GOMERO	8105	HUALQUI	1	1	0	0
11430	4941	ESCUELA PODUCO BAJO	8109	SANTA JUANA	1	1	0	0
11431	5078	ESCUELA CERRO COLORADO	8202	ARAUCO	1	1	0	0
11432	5099	ESCUELA TRONGOL ALTO	8205	CURANILAHUE	1	1	0	0
11433	5163	ESCUELA GRANO DE TRIGO	8204	CONTULMO	1	1	0	0
11434	5168	ESCUELA VILLA RIVAS	8204	CONTULMO	1	1	0	0
11435	5186	ESCUELA PARTICULAR VILLA HERMOSA	8204	CONTULMO	1	1	0	0
11436	11206	ESCUELA MUNICIPAL LAS PUERTAS	5402	CABILDO	1	1	0	0
11437	4314	ESCUELA LAS HIJUELAS	8312	TUCAPEL	1	1	0	0
11438	4323	ESCUELA LA COLONIA	8312	TUCAPEL	1	1	0	0
11439	4478	ESCUELA LOS CASTANOS	8306	NACIMIENTO	1	1	0	0
11440	3384	ESCUELA RAMBERTO CARRERA MARTINEZ	7403	LONGAVÍ	1	1	0	0
11441	3444	ESCUELA AMANDA BENAVENTE D.	7405	RETIRO	1	1	0	0
11442	3989	ESCUELA RINCONADA DE COLTON	16102	BULNES	1	1	0	0
11443	3189	ESCUELA TERESA CONSUELO	7102	CONSTITUCIÓN	1	1	0	0
11444	3196	ESCUELA LEONTINA LETELIER LETELIER	7102	CONSTITUCIÓN	1	1	0	0
11445	3229	ESCUELA MANANTIALES	7103	CUREPTO	1	1	0	0

Tabla 12: Establecimientos con riesgo 0: 1 alumno matriculado, 1 comuna, Entropía = 0

Por las comunas donde se ubican estos y dado el conocimiento de la geografía del país, estas situaciones se corresponden a zonas rurales (Lago Rango, Hualaihué, Nueva Imperial, etc.) e insulares (Ancud, Dalcahue, Quellón, etc.). En todo caso sería conveniente verificar estos datos, dado que podrían tratarse como outliers y eliminarse del reporte. Esto queda abierto para un eventual trabajo futuro.

El segundo comportamiento para analizar, lo componen 1.365 establecimientos educacionales del país con más de un alumno, pero donde todos ellos residen en la misma comuna, ocasionando que la entropía sea 0 (no hay incertidumbre) y por lo tanto el riesgo también es 0

#	Codigo Est. Educ.	Nombre de Establecimiento Educacional	Codigo Comuna	Nombre Comuna	Alumnos Matriculados	Comunas Distintas	Entropía Shannon	Indicador de Riesgo
10081	12870	C.E.I.A. ANTONIO RENDIC.	2101	ANTOFAGASTA	1737	1	0	0
10082	10915	COLEGIO CARDENAL ANTONIO SAMORE	15101	ARICA	1457	1	0	0
10083	288	ESCUELA JAPON	2101	ANTOFAGASTA	1192	1	0	0
10084	24900	LICEO HUMBERTO DIAZ CASANUEVA	13301	COLINA	1131	1	0	0
10085	648	COLEGIO SANTA MARTA	4102	COQUIMBO	1094	1	0	0
10086	287	LICEO DOMINGO HERRERA RIVERA	2101	ANTOFAGASTA	1087	1	0	0
10087	4830	LICEO REPÚBLICA DEL ECUADOR	8111	TOMÉ	1044	1	0	0
10088	366	HRVATSKA SKOLA S. ESTEBAN (EX-J. SKOLA)	2101	ANTOFAGASTA	1027	1	0	0
10089	5026	ESCUELA FRESIA GRACIELA MULLER RUIZ	8201	LEBU	1020	1	0	0
10090	11681	COLEGIO MIGUEL DE CERVANTES	12101	PUNTA ARENAS	993	1	0	0
10091	319	ESCUELA LA BANDERA	2101	ANTOFAGASTA	933	1	0	0
10092	286	LICEO LA PORTADA	2101	ANTOFAGASTA	832	1	0	0
10093	191	LIC. ALCDE.SERGIO GONZALEZ GUTIERREZ	1401	POZO ALMONTE	767	1	0	0
10094	8424	INSTIT.SUPERIOR DE COMERCIO JOSE MENENDEZ	12101	PUNTA ARENAS	703	1	0	0
10095	25790	COLEGIO POLIV. DIEGO PORTALES DE MAIPU	13119	MAIPÚ	683	1	0	0
10096	24318	ESCUELA PADRE ALBERTO HURTADO CRUCHAGA	12101	PUNTA ARENAS	630	1	0	0
10097	2734	COLEGIO ALESSANDRI	7301	CURICÓ	626	1	0	0
10098	296	ESCUELA REPUBLICA DE LOS ESTADOS UNIDOS	2101	ANTOFAGASTA	597	1	0	0
10099	5019	ESCUELA GENARO RIOS CAMPOS	8102	CORONEL	574	1	0	0
10100	31010	LICEO BICENTENARIO CIUDAD DE LOS RIOS VALDIVI	14101	VALDIVIA	562	1	0	0
10101	7131	ESCUELA PRESIDENTE JORGE ALESSANDRI R.	14201	LA UNIÓN	501	1	0	0

Tabla 13: Muestra de Establecimientos con riesgo 0: >500 alumno matriculado, 1 comuna, Entropía = 0

Se observa en la imagen que la comuna de Antofagasta aparece 7 veces en la muestra. Siendo esta una de las comunas con mayor población del norte del país, me cuestiono si el indicador será 100% válido para todos los casos, planteando un trabajo futuro que pueda hacerse cargo de dar una mirada más profunda al indicador para casos de borde como este.

Por ahora, para efectos del análisis, se agrupa estos establecimientos por rangos según la cantidad de alumnos matriculados, en las siguientes categorías : menos de 50 alumnos (<50), mas de 50 y menos de 100 alumnos (>=50), entre 100 y 200 (>= 100), entre 200 y 300 (>= 200), entre 300 y 400 (>=300), entre 400 y 500 (>=400), entre 500 y 1000 (>=500) y finalmente sobre 1000 alumnos (>= 1000).

Los siguientes son los resultados:

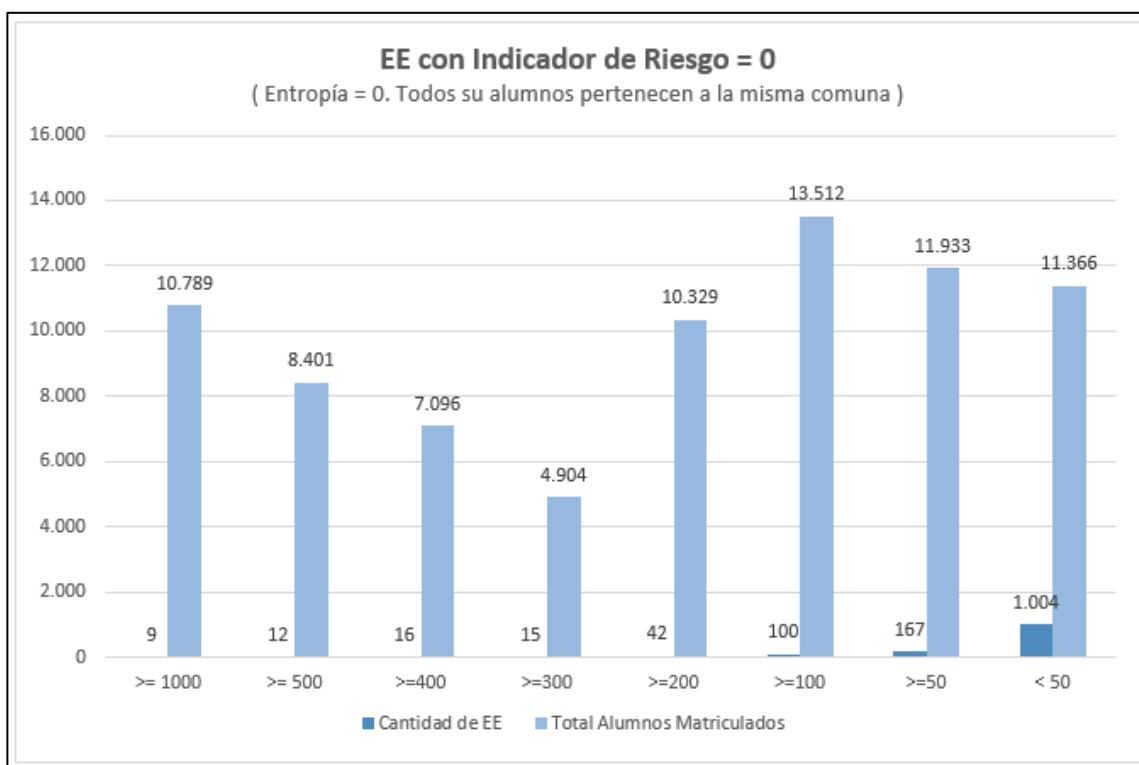


Figura 21: Establecimientos con riesgo 0, distribuidos por rangos, según la cantidad de alumnos

Para mayor detalle de los números ver Anexo E.

6. Conclusiones

6.1. Discusión

Se logró establecer un marco metodológico que permitió alcanzar los objetivos propuestos dentro de los plazos definidos, con un alto nivel de confianza en la calidad de los productos de datos elaborados a partir del análisis.

Se logró desarrollar el indicador de riesgo o peligrosidad y a la luz de los resultados analizados, se considera factible que este pueda ser usado para una orientación a la hora de elaborar una estrategia que permita reabrir de forma segura y gradual todos los establecimientos educacionales del país.

6.2. Limitaciones

El indicador de riesgo que se obtuvo puede considerarse confiable en cuanto a fórmula y la matemática involucrada, pero debido a que la data utilizada (disponible) en estudio es de 2019, es necesario actualizar la data y los valores calculados, para lograr un resultado que sea verdaderamente utilizable.

De modo similar, el filtro de alumnos, basado en la adyacencia de las regiones, debería cambiarse por un criterio más preciso, por ejemplo, basado en la distancia geográfica entre el lugar de residencia del alumno y la ubicación del establecimiento. Lamentablemente esta información no estaba disponible a la hora del presente trabajo.

6.3. Trabajo Futuro

Sería interesante poder dar una segunda revisión al tema, profundizando en aquellos aspectos que por las condiciones actuales no es posible desarrollar, por ejemplo, verificar con el Mineduc u otra fuente la matrícula de establecimientos con 1 solo alumnos o revisar como se obtiene los valores para el campo comuna del alumno, para reducir al máximo la distorsión.

Del mismo modo, si pudiéramos contar con información acerca del número de salas y/o superficie de estas, podríamos estimar el número óptimo de alumnos por sala, para mantener el distanciamiento físico y poder justificar otras alternativas como la implementación de jornadas alternadas por nivel.

Por último, si tuviéramos acceso a información georreferenciada se podrían elaborar mapas de establecimientos con el mismo riesgo potencial, que podrían ser de utilidad para ayudar a enfrentar con éxito un posible nuevo escenario de pandemia, ya sea por nuevos brotes de COVID u otros tipos de Influenza que pudieran golpearnos a futuro.

Pero lo más interesante podría ser ayudar a elaborar un plan de reapertura que, sin dejar de lado las medidas de prevención ya establecidas, permita organizar la vuelta a presencialidad gradual y segura de todos los establecimientos educacionales del país.

7. Bibliografía

1. Chopra, Ritika. «Coronavirus Lockdown: Back to School in Europe». The Indian Express (blog), 22 de mayo de 2020. <https://indianexpress.com/article/explained/coronavirus-lockdown-back-to-school-in-europe-6421568/>.
2. Beiró, Mariano G., Loreto Bravo, Diego Caro, Ciro Cattuto, Leo Ferres, y Eduardo Graells-Garrido. «Shopping Mall Attraction and Social Mixing at a City Scale». EPJ Data Science 7, n.o 1 (diciembre de 2018): 1-21. <https://doi.org/10.1140/epjds/s13688-018-0157-5>.
3. Cooperativa.cl. «Plan de reapertura escolar: Mineduc informó que 935 colegios han emitido solicitudes». Cooperativa.cl. Accedido 6 de noviembre de 2020. <https://cooperativa.cl/noticias/pais/educacion/plan-de-reapertura-escolar-mineduc-informo-que-935-colegios-han-emitido/2020-10-30/115616.html>.
4. Correa Villa, Mauricio. Fundamentos de la teoría de la información. ITM, 2008.
5. Di Domenico, Laura, Giulia Pullano, Chiara E. Sabbatini, Pierre-Yves Boëlle, y Vittoria Colizza. «Expected Impact of Reopening Schools after Lockdown on COVID-19 Epidemic in Île-de-France». Preprint. Infectious Diseases (except HIV/AIDS), 12 de mayo de 2020. <https://doi.org/10.1101/2020.05.08.20095521>.
6. Diario U de Chile. «Presidente Piñera anuncia suspensión de clases y nuevas medidas para enfrentar COVID-19 «Diario y Radio U Chile». DiarioUCHile, 15 de marzo de 2020. <https://radio.uchile.cl/2020/03/15/presidente-pinera-anuncia-suspension-de-clases-y-nuevas-medidas-para-enfrentar-covid-19/>.

7. El Confidencial. «Las claves de la vuelta al cole: clases con 20 niños y distancia para mayores de 10 años». El Confidencial, 11 de junio de 2020.
https://www.elconfidencial.com/espana/2020-06-11/propuestas-educacion-curso-2020-2021-vuelta-al-cole_2633972/.
8. El Desconcierto. «Retorno a clases: Mineduc ha recibido 45 solicitudes de reapertura y ya aprobó un tercio de ellas». El Desconcierto, 12 de agosto de 2020.
<https://www.eldesconcierto.cl/2020/08/12/retorno-a-clases-mineduc-aprobo-un-tercio-de-las-solicitudes/>.
9. El Mercurio. «Reapertura de colegios: 43 municipios del país han pedido volver y en la RM se sumó Renca | Emol.com». Emol, 29 de octubre de 2020.
<https://www.emol.com/noticias/Nacional/2020/10/29/1002157/Renca-pidio-reabrir-colegios-municipales.html>.
10. El País. «Expertos de la Universidad de Granada calculan que meter a 20 niños en un aula supone 808 contactos cruzados en dos días | Educación | EL PAÍS». Accedido 8 de diciembre de 2020. <https://elpais.com/educacion/2020-06-17/expertos-de-la-universidad-de-granada-calculan-que-meter-a-20-ninos-en-un-aula-supone-808-contactos-cruzados-en-dos-dias.html>.
11. Gob.cl, Anuncios Presidente. «Gob.cl - Artículo: Presidente Piñera Anuncia Nuevas Medidas Para Frenar El Avance Del Coronavirus». Gobierno de Chile, 15 de marzo de 2020. <https://www.gob.cl/noticias/presidente-pinera-anuncia-nuevas-medidas-para-enfrentar-el-coronavirus/>.

12. Gob.cl. «Gob.cl - Cifras Oficiales». Gobierno de Chile. Accedido 22 de mayo de 2020.
<https://www.gob.cl/coronavirus/cifrasoficiales/>.
13. Infobae. «Cómo hizo Dinamarca para reabrir rápido sus escuelas y que no aumenten los casos de coronavirus». infobae. Accedido 6 de septiembre de 2020.
[/america/mundo/2020/06/02/las-claves-del-exito-con-que-dinamarca-logro-reabrir-las-escuelas-cerradas-por-el-covid-19/](https://www.infobae.com/america/mundo/2020/06/02/las-claves-del-exito-con-que-dinamarca-logro-reabrir-las-escuelas-cerradas-por-el-covid-19/).
14. Johansen, Oscar. Introducción a la teoría general de sistemas. Editorial Limusa, 1982.
15. La Tercera. «Mineduc dice que colegios deben prepararse para un regreso gradual a clases». La Tercera, 5 de junio de 2020.
<https://www.latercera.com/nacional/noticia/mineduc-dice-que-colegios-deben-prepararse-para-un-regreso-gradual-a-clases/DLSGZHRBYZA67FB7LMVGGXMXU/>.
16. La Tercera. «Regreso a clases con testeos en colegios: Salud aplicará examen de detección de anticuerpos a escolares». La Tercera, 20 de abril de 2020.
<https://www.latercera.com/la-tercera-pm/noticia/regreso-a-clases-con-testeos-en-colegios-salud-aplicara-examen-de-deteccion-de-anticuerpos-a-escolares/7VDAET4RO5BODOCO3B5FFAU7JQ/>.
17. Lau, Hien, Veria Khosrawipour, Piotr Kocbach, Agata Mikolajczyk, Justyna Schubert, Jacek Bania, y Tanja Khosrawipour. «The Positive Impact of Lockdown in Wuhan on Containing the COVID-19 Outbreak in China». *Journal of Travel Medicine* 27, n.o 3 (18 de mayo de 2020). <https://doi.org/10.1093/jtm/taaa037>.
18. Luzmore, Ruth. «Reopening My School after Lockdown Is a Huge Logistical Challenge | Ruth Luzmore». *The Guardian*, 31 de mayo de 2020, sec. Education.

- <https://www.theguardian.com/education/commentisfree/2020/may/31/reopening-my-school-after-coronavirus-lockdown-is-a-huge-logistical-challenge>.
19. McKinsey. «How to Safely Reopen Schools after COVID-19 Closures | McKinsey». Accedido 24 de mayo de 2020. <https://www.mckinsey.com/industries/social-sector/our-insights/safely-back-to-school-after-coronavirus-closures>.
 20. Mental Health Foundation. «Returning to School after the Coronavirus Lockdown». Mental Health Foundation, 29 de mayo de 2020. <https://www.mentalhealth.org.uk/coronavirus/returning-school-after-lockdown>.
 21. Ministerio de Ciencias. «Base de Datos COVID-19 | Ministerio de Ciencia». Accedido 22 de mayo de 2020. <http://www.minciencia.gob.cl/covid19>.
 22. Ministerio de Educación. «Apoyos del Mineduc durante la pandemia del Covid-19». Ministerio de educación, 14 de mayo de 2020. <https://www.mineduc.cl/apoyos-del-mineduc-durante-la-pandemia-del-covid-19/>.
 23. Ministerio de Educación. «Descarga Bases de datos de matrícula por estudiante · MINEDUC». Ministerio de Educación. Accedido 22 de mayo de 2020. <http://datos.mineduc.cl/dashboards/19776/descarga-bases-de-datos-de-matricula-por-estudiante/>.
 24. Ministerio de Educación. «Descarga Bases de datos de Resumen Matrícula por Establecimiento · MINEDUC». Ministerio de Educación. Accedido 22 de mayo de 2020. <http://datos.mineduc.cl/dashboards/20022/descarga-bases-de-datos-de-resumen-matricula-por-establecimiento/>.

25. Ministerio de Educación. «MINEDUC: Se extiende período de suspensión de clases». Ministerio de educación, 25 de marzo de 2020. <https://www.mineduc.cl/se-extiende-periodo-de-suspension-de-clases/>.
26. Ministerio de Educacion. «Plan Retorno A Clases». Accedido 12 de octubre de 2020. <https://www.colegiodeprofesores.cl/wp-content/uploads/2020/07/PlanRetornoAClases-08.06.pdf>.
27. Noticias ONU. «La OMS y UNICEF instan a la pronta reapertura de las escuelas africanas». Noticias ONU, 20 de agosto de 2020. <https://news.un.org/es/story/2020/08/1479252>.
28. Olivares, M, M Goic, G Weintraub, A Carranza, J Covarrubia, C Escobedo, y L Basso. «Informe-Movilidad-y-contagios-RM-28-de-julio.pdf». Accedido 12 de octubre de 2020. <https://isci.cl/wp-content/uploads/2020/07/Informe-Movilidad-y-contagios-RM-28-de-julio.pdf>.
29. OMS. «Medidas para los establecimientos escolares». WHO. World Health Organization, 11 de septiembre de 2009. https://www.who.int/csr/disease/swineflu/notes/h1n1_school_measures_20090911/es/.
30. Our World in Data. «Daily confirmed COVID-19 cases per million people». Our World in Data. Accedido 23 de mayo de 2020. <https://ourworldindata.org/grapher/new-covid-cases-per-million>.
31. Pancevski, Bojan, y Naja Dandanell. «¿Es seguro reabrir las escuelas? Estos países dicen que sí». La Tercera, 2 de junio de 2020. <https://www.latercera.com/que-pasa/noticia/es->

seguro-reabrir-las-escuelas-estos-paises-dicen-que-
si/5ANEBA5APNB75CFDIQN2XX3Y7Y/.

32. Pietrobelli, Angelo, Luca Pecoraro, Alessandro Ferruzzi, Moonseong Heo, Myles Faith, Thomas Zoller, Franco Antoniazzi, Giorgio Piacentini, S. Nicole Fearnbach, y Steven B. Heymsfield. «Effects of COVID-19 Lockdown on Lifestyle Behaviors in Children with Obesity Living in Verona, Italy: A Longitudinal Study». *Obesity* n/a, n.o n/a (2020).
<https://doi.org/10.1002/oby.22861>.
33. Radio Bio Bio. «Mineduc reporta 261 colegios abiertos: otros 674 solicitaron apertura y se encuentran en revisión». *BioBioChile* - 30 octubre de 2020.
<https://www.biobiochile.cl/especial/educacion/noticias/2020/10/30/mineduc-reporta-261-colegios-abiertos-otros-674-solicitaron-apertura-y-se-encuentran-en-revision.shtml>.
34. Radio Duna. «Presidente Piñera posterga el retorno a clases y defiende el regreso gradual de los trabajadores públicos». *RadioDuna-Noticias*, 19 de abril de 2020.
<https://www.duna.cl/noticias/2020/04/19/presidente-pinera-posterga-el-retorno-a-clases-y-defiende-el-regreso-gradual-de-los-trabajadores-publicos/>.
35. Radio Duna. «Retorno a clases: ¿Cómo se está haciendo en los lugares donde ya no hay nuevos contagios?» <https://www.facebook.com/RadioDuna>. Consultado 24 de mayo de 2020. <https://www.duna.cl/noticias/2020/04/21/retorno-a-clases-como-se-esta-haciendo-en-los-lugares-donde-ya-no-hay-nuevos-contagios/>.
36. Radio Regional. «Expertos asesorarán reapertura de colegios para 2020 y 2021». *Radio Regional*, 27 de octubre de 2020. <https://www.radiosregionales.cl/expertos-asesoraran-reapertura-de-colegios-para-2020-y-2021/>.

37. Rehbein, Consuelo. «Documento sobre “retorno seguro” a clases del Mineduc se filtró: incluye protocolo para casos positivos de covid-19 y transporte escolar». *Publmetro Chile*, 2020. <https://www.publmetro.cl/cl/noticias/2020/07/30/retorno-seguro-clases-mineduc.html>.
38. Robertson, Emma, Karen Hershenfield, Sherry Lynn Grace, y Donna Eileen Stewart. «The Psychosocial Effects of Being Quarantined Following Exposure to SARS: A Qualitative Study of Toronto Health Care Workers». *The Canadian Journal of Psychiatry* 49, n 6 (1 de junio de 2004): 403-7. <https://doi.org/10.1177/070674370404900612>.
39. Tarabini, Aina. «¿Para qué sirve la escuela? Reflexiones sociológicas en tiempos de pandemia global». *Revista de Sociología de la Educación-RASE* 13, n. 2 (29 de abril de 2020): 145-55. <https://doi.org/10.7203/RASE.13.2.17135>.
40. Universidad de Chile, ISCI. «Investigación de las universidades de Chile y Stanford revela impacto de reducción de movilidad en la contención de contagios - Universidad de Chile». Consultado 12 de octubre de 2020. <https://uchile.cl/noticias/165677/el-impacto-de-reducir-la-movilidad-para-contener-contagios-covid-19>.

ANEXOS

Anexo A: Ejes del plan Retorno Seguro del Mineduc

Proponiendo a la escuela como lugar de reencuentro, el Mineduc elabora el plan “Retorno Seguro” que incluye medidas sanitarias, pedagógicas y administrativas, considerando tanto el bienestar físico como emocional no solo de los alumnos, sino de toda la comunidad educativa. A continuación, se detallan los principios rectores o ejes de este plan:

Seguridad y protección

- *Se retornará a clases cuando las condiciones sanitarias lo permitan*
- *Generar condiciones ambientales básicas de higiene, limpieza, cuidados personales y de distanciamiento social en los establecimientos para evitar contagios.*
- *Reforzar la instalación de una cultura preventiva en todos los establecimientos.*

Contención socioemocional

- *Abordar el aprendizaje y contención socioemocional como eje prioritario.*
 - *Promover y desarrollar una convivencia escolar respetuosa e inclusiva como aprendizaje transversal.*
-

Flexibilidad y gradualidad

- *Implementar medidas adecuadas y pertinentes al contexto de cada comunidad educativa, considerando sus desafíos particulares.*

Centralidad del proceso pedagógico

- *Planificación del retorno a clases presenciales bajo criterios pedagógicos, a fin de resguardar las trayectorias educativas de todos los estudiantes.*

Equidad

- *Poner al centro de las preocupaciones la protección, el cuidado y el resguardo de las trayectorias educativas de todos los estudiantes, identificando a tiempo los apoyos diferenciados que requieren y gestionando los instrumentos e instancias disponibles para ello.*

Proponiendo énfasis en la gradualidad y en la protección, esta guía ayuda a planificar el retorno seguro de estudiantes y profesores a la presencialidad.

Para mayor información, el texto completo de este plan se encuentra disponible en

<https://www.colegiodeprofesores.cl/wp-content/uploads/2020/07/PlanRetornoAClases-08.06.pdf>

Anexo B: Peligrosidad en base a comunas

Intentando establecer un indicador de riesgo potencial para un establecimiento educacional, la primera aproximación que surge es estimarlo en función del número de comunas que podrían verse afectadas en caso un nuevo contagio.

Este enfoque se basa en que siempre existe una posibilidad de contagio por transmisión directa asociado al número de interacciones sociales que se producen al interior de un recinto cerrado, por ejemplo la sala de clases.

De manera intuitiva entendemos que a mayor número de personas en un recinto, mayores las posibilidades de contagio. ¿Cómo demostrarlo?, simplemente graficando el número de interacciones posibles, según el número de personas, como se muestra a continuación:

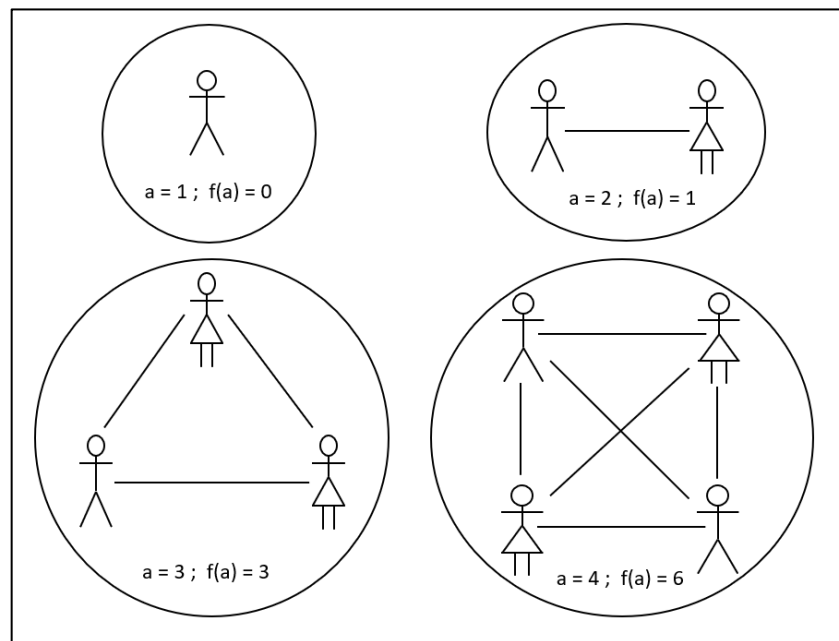


Figura B1: Interacciones posibles entre 1, 2, 3 y 4 alumnos en una sala de clases

También podemos generar una aplicación que determine y muestre estas interacciones:

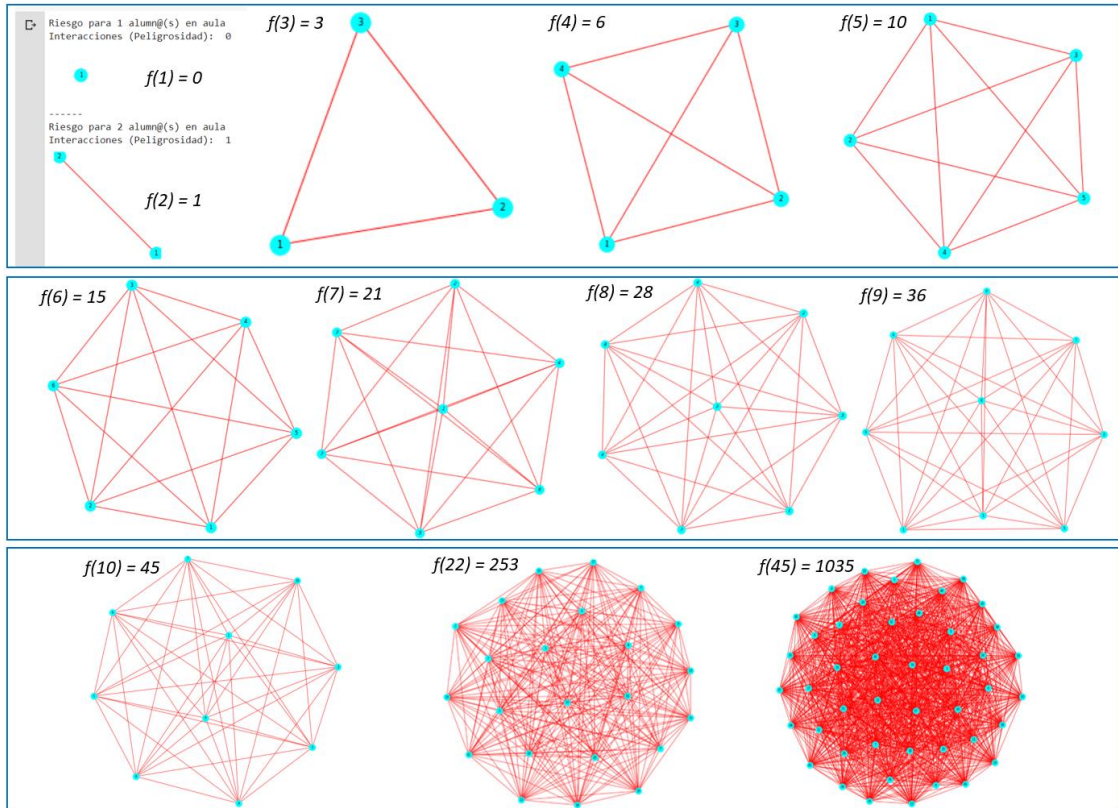


Figura B2: Interacciones posibles entre distintos números de alumnos en una sala de clases

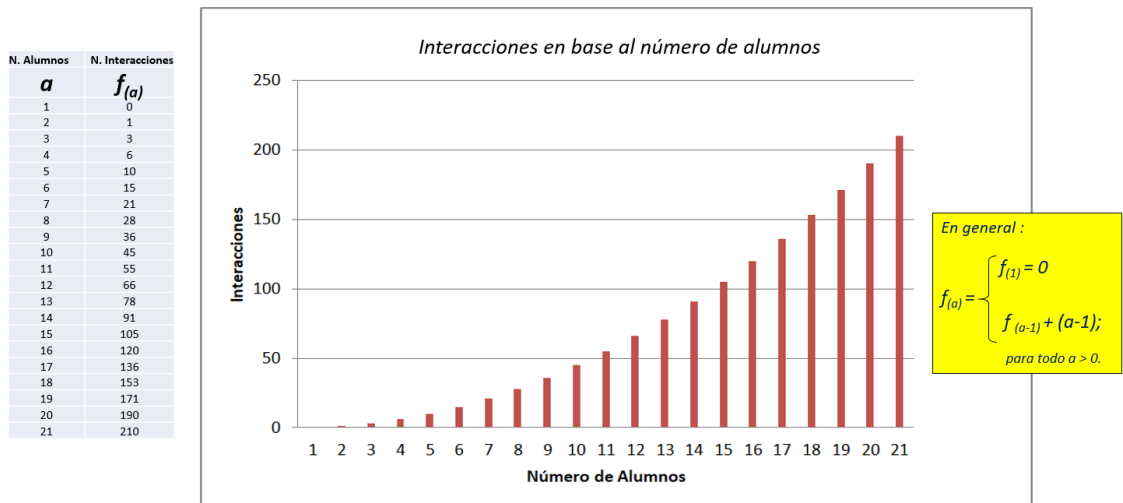


Figura B3: Análisis de las Interacciones posibles entre distintos números de alumnos en una sala de clases

Extrapolando esta idea desde las aulas a las comunas donde se ubican los establecimientos educacionales, al congregarse alumnos de distintas comunas en un recinto, existe la posibilidad que en los desplazamientos desde y hacia sus hogares, un estudiante enfermo (incluso si es asintomático) podría entrar en contacto con otras personas y propagar un contagio, el cual, podría potencialmente afectar a toda la comuna donde reside.

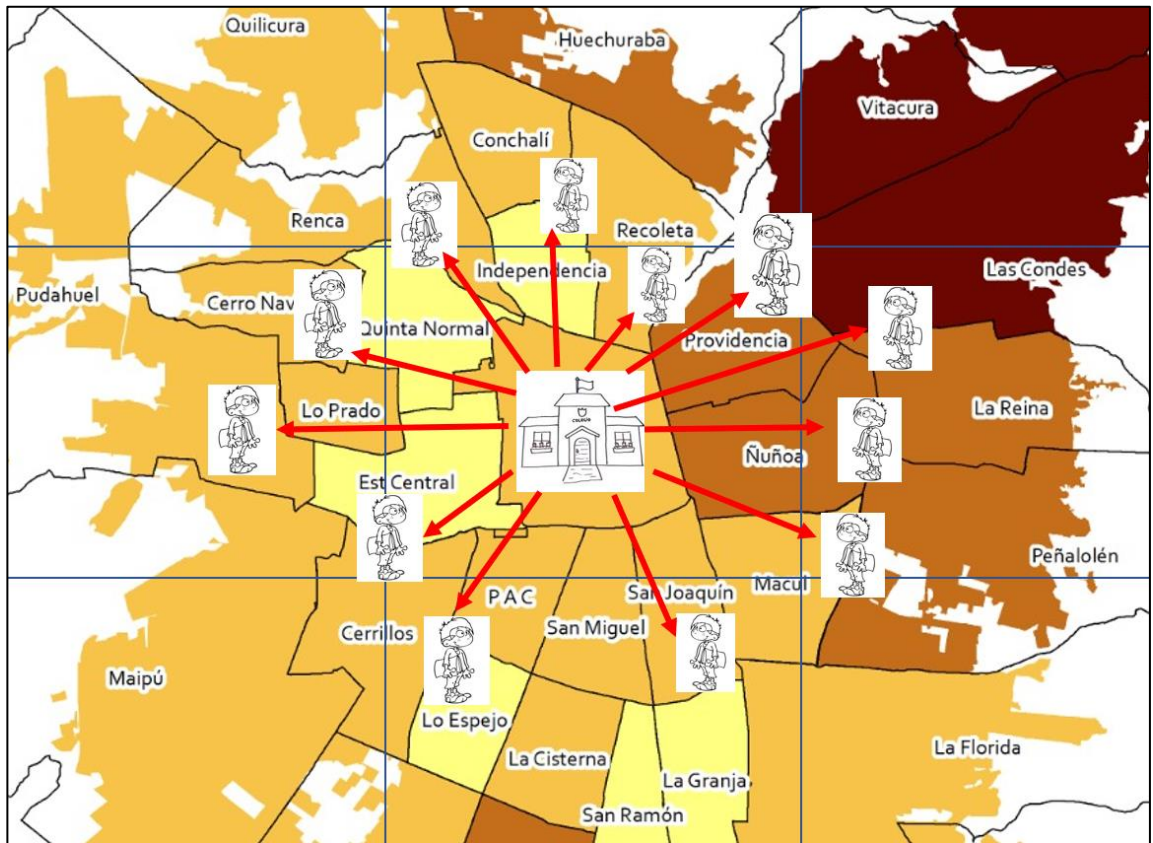


Figura B4: Impacto teórico en comunas

A continuación, se presentan tres escenarios hipotéticos A.B y C donde se intenta teorizar sobre el riesgo potencial de los establecimientos, respecto de la composición de su plantel es decir, en función de las interacciones entre alumnos de distintas comunas.

El caso A presenta una composición HOMOGENEA, donde todos sus alumnos viven en la misma comuna en la cual se ubica el establecimiento, por lo que no hay desplazamientos fuera de ella. En este escenario el riesgo de propagación está CONTENIDO solo a esa comuna, por lo que podría catalogarse como de BAJO RIESGO.

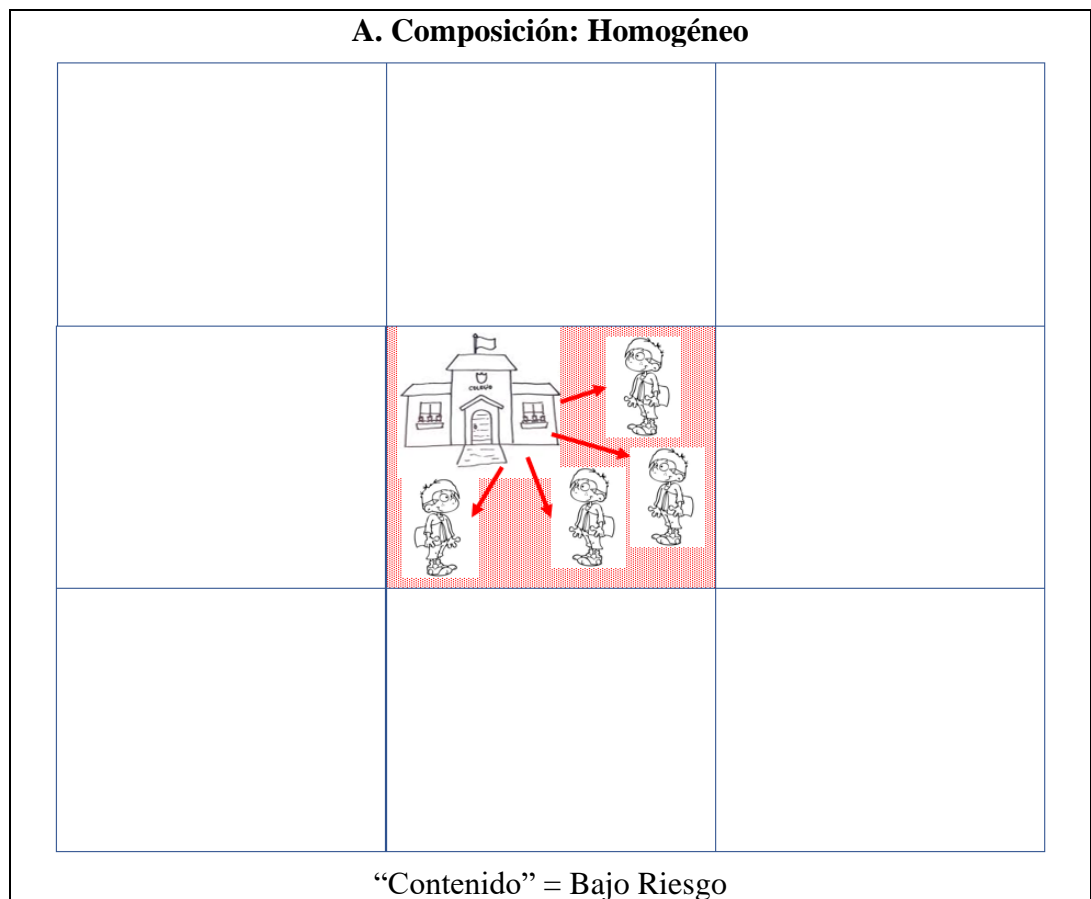
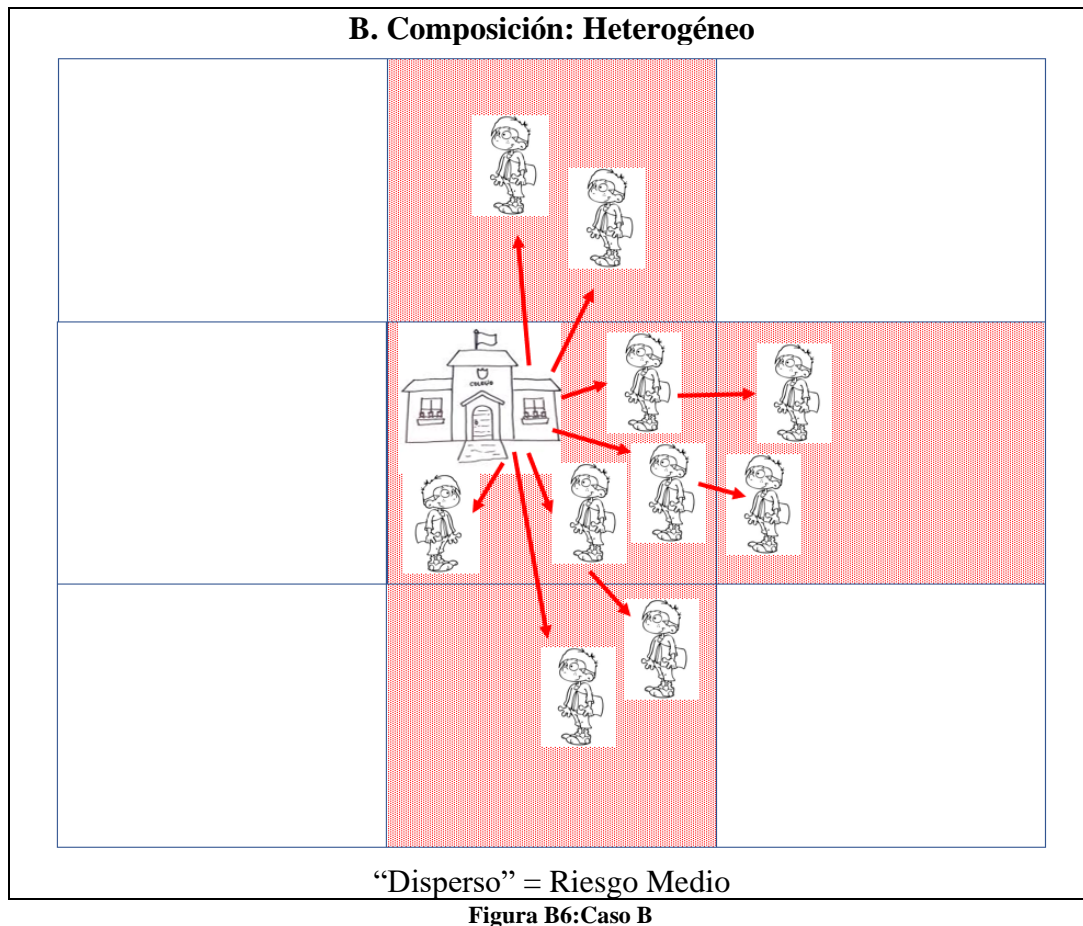


Figura B5: Caso A

Teóricamente A podría ubicarse en los territorios insulares del país, donde la movilidad y los desplazamientos son limitadas desde y hacia otros territorios.

El caso B presenta una composición HETEROGÉNEA, donde algunos alumnos viven en la misma comuna o bien en comunas cercanas a la que se ubica el establecimiento, por lo cual existen algunos desplazamientos fuera de la comuna. En este escenario el riesgo de propagación es DISPERSO ya que podría afectarse esa comuna y sus vecinas, por lo que podría catalogarse como de RIESGO MEDIO.



Teóricamente B podría ubicarse en los territorios extremos del país, donde las distancias sean muy grandes, su geografía muy accidentada o existan notables diferencias de altitud que dificulten la comunicación y los desplazamientos entre ellos.

El caso C presenta una composición MUY HETEROGÉNEA, donde algunos alumnos viven en la misma comuna o bien en comunas cercanas a la que se ubica el establecimiento mientras que otros lo hacen en comunas más alejadas, por lo que pueden existir un gran número de desplazamientos dentro y fuera de la comuna.

En este escenario el riesgo de propagación es EXPLOSIVO ya que podrían resultar afectadas varias comunas, por lo que podría catalogarse como de ALTO RIESGO.

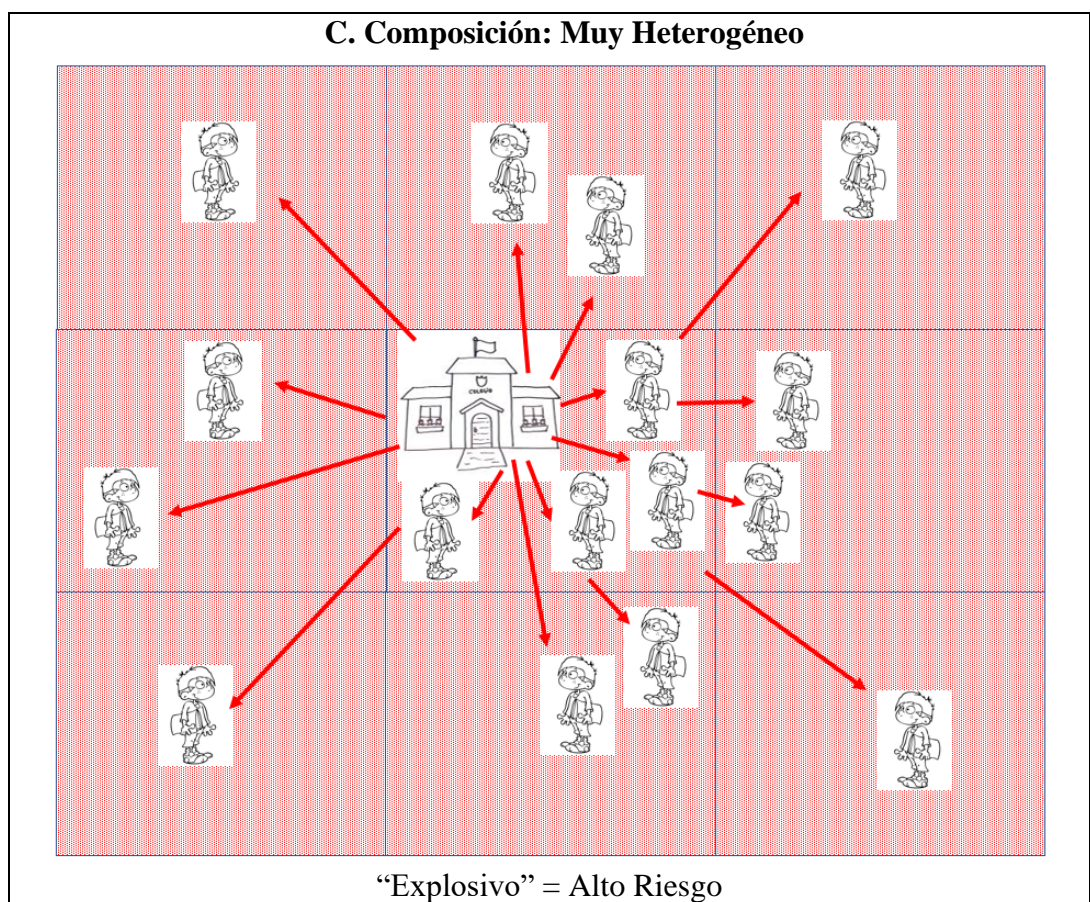


Figura B7:Caso C

Teóricamente C podría ubicarse en zonas densamente pobladas del país, donde existe mejor conectividad y la infraestructura así como los sistemas de transporte facilitan la movilidad y el desplazamiento de las personas, aun cuando se trate de grandes distancias.

De esta forma podríamos resumir :

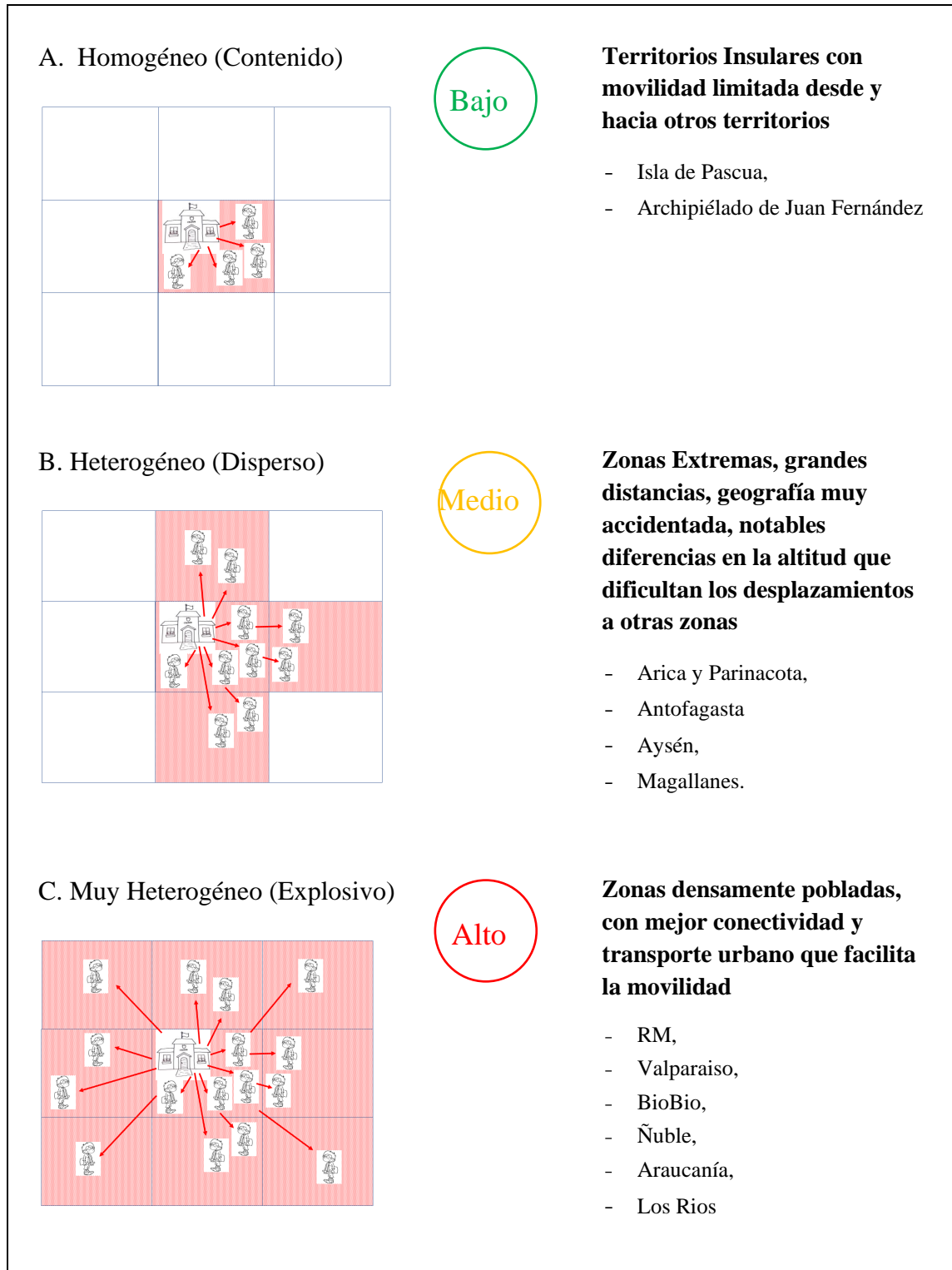


Figura B8:Resumen

El problema de esta aproximación es cuantificar en términos absolutos (y medibles) que se considera como ALTO, MEDIO o BAJO riesgo; ya que considerar solamente el número de comunas da el mismo peso o importancia a cada una de ellas sin considerar la distribución de alumnos en ella.

Surge entonces la necesidad de implementar un criterio adicional que nos permita dar cuenta del grado de dispersión de los alumnos en las distintas comunas, o dicho de otro modo, la mezcla social que se produce al interior del establecimiento; por lo cual se decide complementar este criterio de peligrosidad en base a comunas, incluyendo el criterio de entropía, el cual se explica en detalle en el anexo C.

Anexo C: la Entropía como medida de mezcla social.

Esta aproximación intenta conceder un peso o importancia a cada comuna en base a la distribución o cantidad de alumnos que pertenece a ella.

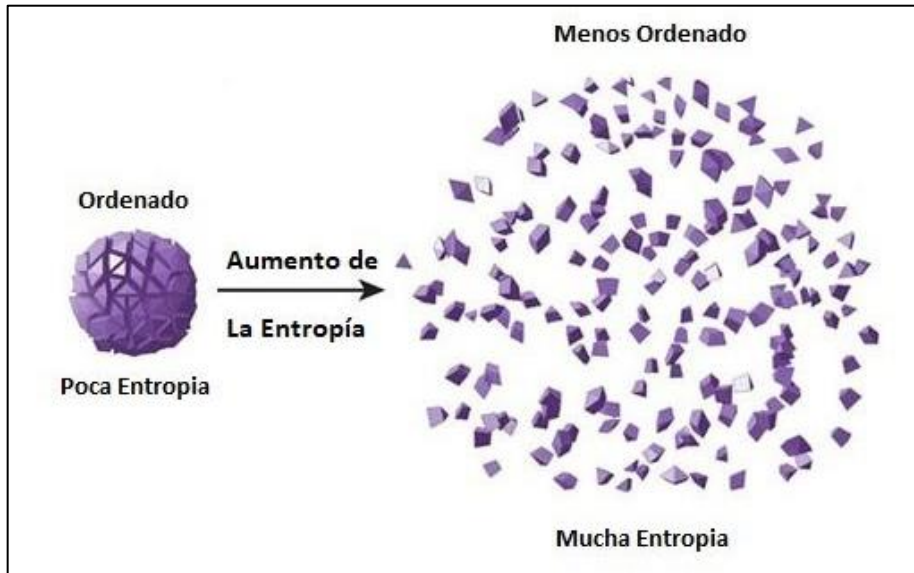


Figura C1:Concepto gráfico de Entropía

En términos simples, para la teoría de la información, la entropía se concibe como una “medida del desorden” o la “peculiaridad de ciertas combinaciones”, por lo tanto, se puede considerar la entropía como una medida de la incertidumbre y de la información que se necesita para poder acotar, reducir o eliminar esta incertidumbre, en cualquier proceso.


Si el caso fuera calcular las probabilidades de obtener un resultado lanzando un dado convencional de 6 caras, donde cada lado tiene la misma posibilidad de salir en un lanzamiento, la probabilidad sería $1/6$ ya que la formula nos dice: un resultado favorable dividido entre el seis resultados posibles.

Pero ¿Cuál sería la probabilidad de obtener un resultado si se lanzan dos dados? O dicho de otra manera ¿todos los resultados posibles tienen la misma probabilidad de salir?

Entropía

Para un dado común de 6 caras (no cargado); la probabilidad de obtener un resultado es la misma para cada cara del dado, es decir, $1/6$

Para un dado convencional de 6 caras :




Casos favorables : 1 ; Casos totales : 6; Probabilidad : $1/6$


Probabilidades : $P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = 1/6$

Combinaciones : (Valores Posibles ^{numero de dados}); $6^1 = 6$

Pero ¿cuál es la probabilidad, si se lanzan dos dados a la vez?

Para dos dados convencionales de 6 caras :






Casos favorables = 1

Casos totales = 11 (sumas posibles entre 2 y 12)

Probabilidades = $1/11$ para obtener 2 o 12
 $2/11$ para obtener 3 u 11
 $3/11$ para obtener 4 o 10
 $4/11$ para obtener 5 o 9
 $5/11$ para obtener 6 u 8
y $6/11$ para obtener 7 ...

Combinaciones : $6^2 = 36$



Estado Inespecífico	Estados Específicos	Probabilidad
P(2)	1+1	1/11
P(3)	1+2, 2+1	2/11
P(4)	1+3, 2+2, 3+1	3/11
P(5)	1+4, 2+3, 3+2, 4+1	4/11
P(6)	1+5, 2+4, 3+3, 4+2, 5+1	5/11
P(7)	1+6, 2+5, 3+4, 4+3, 5+2, 6+1	6/11
P(8)	2+6, 3+5, 4+4, 5+3, 6+2	5/11
P(9)	3+6, 4+5, 5+4, 6+3	4/11
P(10)	4+6, 5+5, 6+4	3/11
P(11)	5+6, 6+5	2/11
P(12)	6+6	1/11

En general :

Para N dados convencionales de 6 caras :

Casos favorables : 1

Casos totales : $(n \cdot 6) - (n-1)$

Probabilidades : Distribución Normal
(las mayores probabilidades tienden a estar en el centro)

Combinaciones : 6^n

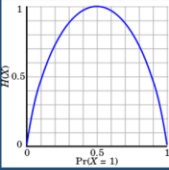


Figura C2:Explicando el concepto de Entropía

Como se observa en la figura, hay números que tienen mayores probabilidades que otros; por ejemplo: existen seis combinaciones posibles para obtener un siete, pero solo una combinación posible para obtener un dos o un doce.

Aplicando este concepto al tema que estamos estudiando, si consideramos los alumnos de un establecimiento, según su comuna de residencia, notaremos que las concentraciones son distintas, es decir, que hay comunas que tienen una mayor representatividad que otras, por lo tanto, siendo la entropía una medida de la distribución de los elementos del sistema, (a mayor distribución, mayor la entropía y a menor entropía menor la incertidumbre), esta nos sirve como un indicador de la mezcla al interior del establecimiento.

Para explicarlo de manera sencilla se utilizará un caso hipotético de una escuela con 200 alumnos matriculados que viven en 5 comunas distintas, ubicándose el colegio en una de estas, (aunque no necesariamente)

En primer lugar, se cuenta la cantidad de alumnos por cada comuna.

Colegio Ejemplo	Número de Alumnos
Comuna 1	100
Comuna 2	80
Comuna 3	10
Comuna 4	6
Comuna 5	4
Total Alumnos	200

Tabla C1: Ejemplo hipotético

A continuación, se establece el porcentaje asociado a cada comuna, como una medida de la representatividad de esta comuna en la escuela

Colegio Ejemplo	Número de Alumnos	Representatividad (%)
Comuna 1	100	50%
Comuna 2	80	40%
Comuna 3	10	5%
Comuna 4	6	3%
Comuna 5	4	2%
Total Alumnos	200	100%

Tabla C2: Ejemplo hipotético

En este ejemplo se observa que las comunas 1 y 2 concentran la mayoría de los alumnos. Las posibilidades son claramente distintas, la comuna 1 al ser más grande (mayor número de alumnos) tendría una probabilidad mayor de ocurrencia que la comuna 5. Por lo tanto, si escogiéramos un alumno cualquiera, al azar, es mayormente probable que pertenezca a las comunas 1 o 2, que a las comunas 3, 4 o 5.

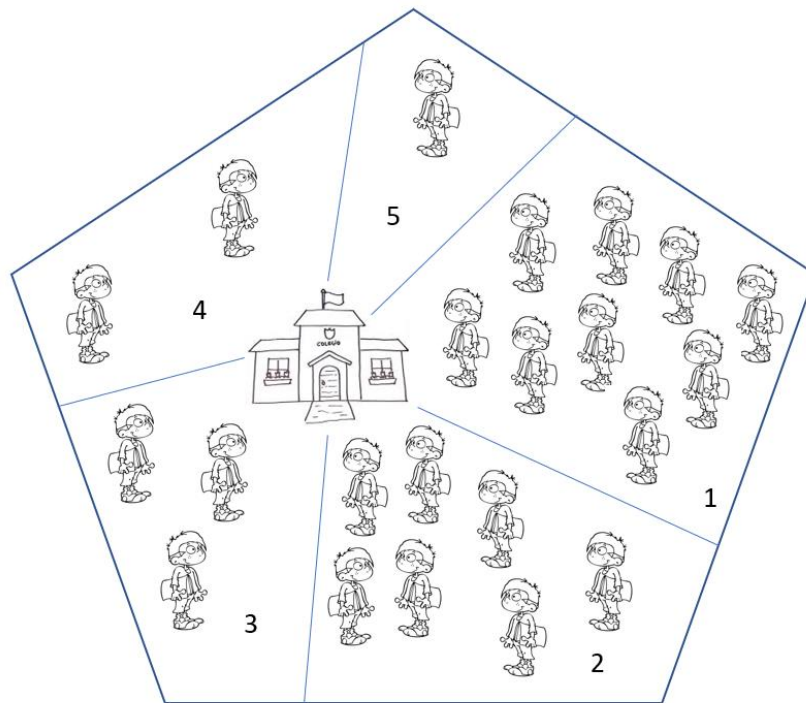


Figura C3: Entropía como medida de mezcla de comunas al interior de un establecimiento

La idea de usar la entropía de Shannon, como medida útil de la diversidad social al interior de un recinto, surge de la revisión del trabajo “Shopping mall attraction and social mixing at a city scale” (Beiró et al., 2018), en donde se encuentra una forma sencilla de operacionalizarlo para obtener un valor numérico que podamos cuantificar.

De esta forma para determinar el valor de la Entropía se requiere calcular la sumatoria de las probabilidades por el logaritmo de la probabilidad, para los Q términos de la serie. La fórmula se expresa a continuación y se demuestra su aplicabilidad en este contexto.

$$S_m = - \sum_{q=1}^Q p_q \cdot \log p_q$$

De esta forma, siendo:

- m = Colegio que quiero estudiar
- Q = Total de comunas presentes al interior del colegio (en el ejemplo: 5)
- p_q = Probabilidad de la ocurrencia de la comuna q-esima : representada por una fracción entre el número de estudiantes de esa comuna (n) sobre el total de estudiantes del colegio ($N:200$)
- S_m = Entropía de Shannon para el Colegio m

Desarrollamos la fórmula con los datos del ejemplo:

$$S_m = (\text{Prob C1} * \text{Log Prob C1}) + (\text{Prob C2} * \text{Log Prob C2}) + (\text{Prob C3} * \text{Log Prob C3}) + (\text{Prob C4} * \text{Log Prob C4}) + (\text{Prob C5} * \text{Log Prob C5})$$

$$S_m = (100/200 * \text{Log } 100/200) + (80/200 * \text{Log } 80/200) + (10/200 * \text{Log } 10/200) + (6/200 * \text{Log } 6/200) + (4/200 * \text{Log } 4/200)$$

$$S_m = (0.5 * \text{Log } 0.5) + (0.4 * \text{Log } 0.4) + (0.05 * \text{Log } 0.05) + (0.03 * \text{Log } 0.03) + (0.02 * \text{Log } 0.02)$$

$$S_m = (0.5 * -0,30103) + (0.4 * -0.39794) + (0.05 * -1.30103) + (0.03 * -1.52288) + (0.02 * -1.69897)$$

$$S_m = (-0.150515) + (-0.15918) + (-0.06505) + (-0.045686) + (-0.03398)$$

$$S_m = (-0.454411)$$

Colegio Ejemplo	Alumnos (n)	Pq : (n/N)	Logaritmos LogPq	Pq * LogPq
Comuna 1	100	50%	-0,30103	-0,15051
Comuna 2	80	40%	-0,39794	-0,15918
Comuna 3	10	5%	-1,30103	-0,06505
Comuna 4	6	3%	-1,52288	-0,04569
Comuna 5	4	2%	-1,69897	-0,03398
Total Alumnos (N)	200	100%	Sumatoria:	-0,45441
			Entropía:	0,45441

Tabla C3: Ejemplo hipotético

VALIDEZ DE SUPUESTO

Para validar que la entropía es una medida adecuada para los propósitos del presente estudio, se realizó una comparación entre otros colegios hipotéticos con la misma cantidad de alumnos (200) y el mismo número de comunas (5); pero con diferentes distribuciones de alumnos por comunas, entre cada una de ellas.

Id. Colegio Ejemplo	Caso 1	Caso 2	Caso 3	Caso 4	Caso 5	Caso 6	Caso 7	Caso 8
Comuna 1	180	160	140	120	100	80	60	40
Comuna 2	5	10	15	20	25	30	35	40
Comuna 3	5	10	15	20	25	30	35	40
Comuna 4	5	10	15	20	25	30	35	40
Comuna 5	5	10	15	20	25	30	35	40
Total Alumnos (N)	200	200	200	200	200	200	200	200
Probabilidades Pq	0,9	0,8	0,7	0,6	0,5	0,4	0,3	0,2
	0,025	0,05	0,075	0,1	0,125	0,15	0,175	0,2
	0,025	0,05	0,075	0,1	0,125	0,15	0,175	0,2
	0,025	0,05	0,075	0,1	0,125	0,15	0,175	0,2
	0,025	0,05	0,075	0,1	0,125	0,15	0,175	0,2
Logaritmos Log Pq	-0,04576	-0,09691	-0,1549	-0,22185	-0,30103	-0,39794	-0,52288	-0,69897
	-1,60206	-1,30103	-1,12494	-1	-0,90309	-0,82391	-0,75696	-0,69897
	-1,60206	-1,30103	-1,12494	-1	-0,90309	-0,82391	-0,75696	-0,69897
	-1,60206	-1,30103	-1,12494	-1	-0,90309	-0,82391	-0,75696	-0,69897
	-1,60206	-1,30103	-1,12494	-1	-0,90309	-0,82391	-0,75696	-0,69897
Términos de la Sumatoria (Pq * LogPq)	-0,04118	-0,07753	-0,10843	-0,13311	-0,15051	-0,15918	-0,15686	-0,13979
	-0,04005	-0,06505	-0,08437	-0,1	-0,11289	-0,12359	-0,13247	-0,13979
	-0,04005	-0,06505	-0,08437	-0,1	-0,11289	-0,12359	-0,13247	-0,13979
	-0,04005	-0,06505	-0,08437	-0,1	-0,11289	-0,12359	-0,13247	-0,13979
	-0,04005	-0,06505	-0,08437	-0,1	-0,11289	-0,12359	-0,13247	-0,13979
Entropía	0,20139	0,33773	0,44591	0,53311	0,60206	0,65352	0,68674	0,69897

Tabla C4: Comparación entre 8 ejemplos hipotéticos

En la tabla se observa como en el caso 8, donde los alumnos están distribuidos de manera más uniforme en varias comunas, tiene un valor mayor que el caso 1, donde la mayoría de los alumnos se concentran en una sola, demostrándose que es un valor adecuado para los propósitos de este trabajo.

Anexo D: Estudio de un caso de Ejemplo

Para elaborar un ejemplo escogí la unidad educativa: *Colegio Compañía de María-Seminario*, en la comuna de Providencia, Santiago.

El ejemplo se desarrolló con una base de datos Oracle 11g XE, se adjuntan las imágenes como evidencia de la consulta SQL el resultado obtenido en cada caso

PREPARACION

Se crea ambiente de trabajo en Oracle

```
-- 1. CREA TABLESPACE Y DATAFILE
CREATE TABLESPACE TESIS
DATAFILE 'C:\oraclexe\app\oracle\oradata\XE\mitesis.dbf'
SIZE 1024m
REUSE
AUTOEXTEND ON NEXT 10M MAXSIZE 2048M;

-- 2. CREA USUARIO Y ASOCIA A TABLESPACE
CREATE USER TESIS IDENTIFIED BY tesis
DEFAULT TABLESPACE "TESIS"
TEMPORARY TABLESPACE "TEMP";

-- 3. ASIGNA QUOTAS ILIMITADA
ALTER USER TESIS QUOTA UNLIMITED ON TESIS;

-- 4. ASIGNA ROL QUE PERMITE CREAR TABLAS
GRANT "RESOURCE" TO TESIS ;

-- 5. OTORGA PRIVILEGIOS
GRANT CREATE SESSION TO TESIS ;
GRANT CREATE SYNONYM TO TESIS;
GRANT CREATE VIEW TO TESIS;
```

Figura D1: Preparación del Ambiente

Se crean las tablas y se cargan estas a partir de los datos del Mineduc

```
-- 2. CARGANDO DATOS BASES DESDE EL ORIGEN
-----
-- ARCHIVO ZIP : MATRICULA 2019.RAR
-- ARCHIVO DE DATOS : 20191028_Matricula_unica_2019_20190430_PUBL.CSV
-- DESCRIPCION DE LOS CAMPOS : ER_Matricula_por_alumno_PUBL_MRUN_WEB.pdf
-- RECUPERADO DE : http://datos.mineduc.cl/dashboards/19776/descarga-bases-de-datos-de-matricula-por-estudiante/
-- FECHA CONSULTA : 05/SEP/2020
-----
DROP TABLE T1_DETALLE_MATRICULA_EE_2019;
-----
CREATE TABLE T1_DETALLE_MATRICULA_EE_2019
(
  AGNO          NUMBER(4),
  RBD           NUMBER(10),
  DGV RBD      CHAR(2),
)
```

Figura D2: Preparación del Ambiente, Carga de Datos (muestra)

Se realizan los primeros análisis exploratorios, que no viene al caso detallar ahora

```
-----
-- En este punto se produce la importación de datos
-- usando el utilitario SQL-LOADER que incluye por defecto
-- SQL-DEVELOPER en la acción de "IMPORTAR".
-----
CREATE SYNONYM MAESTRO
FOR T1_DETALLE_MATRICULA_EE_2019 ;
-----
SELECT COUNT(1) FROM MAESTRO;
-- 3.624.343 : Mismo número declarado en el documento
-----
SELECT ESTADO_ESTAB, count(1)
FROM MAESTRO
group by ESTADO_ESTAB;
-- "ESTADO_ESTAB" "COUNT(1)"
-- 1 3.623.883 : Mismo número declarado en el documento
-- 4 385
-- 3 75
-----
```

Figura D3: Preparación del Ambiente, Carga de Datos (muestra)

Se cargan los códigos territoriales y otras acciones de preparación del ambiente

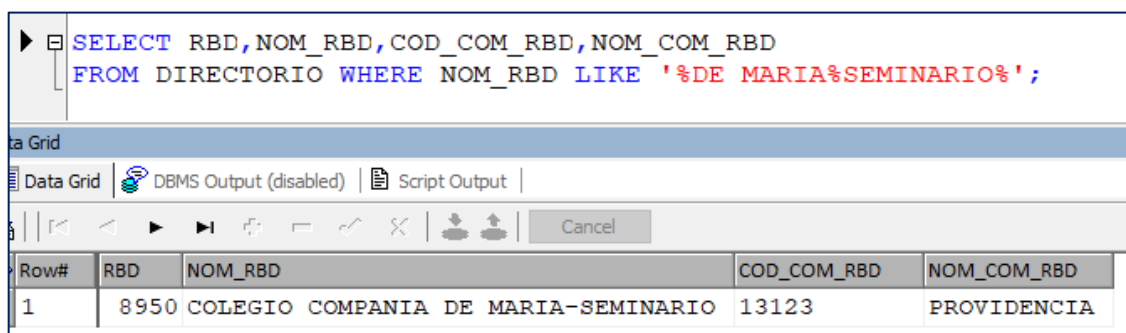
```
-- =====
----- REQUIERO NOMBRES DE COMUNAS
-- http://www.subdere.gov.cl/documentacion/c%C3%B3digos-%C3%BAnico
DROP TABLE T5_CODIGOS_TERRITORIALES_2018;
--
CREATE TABLE T5_CODIGOS_TERRITORIALES_2018
(
  Cod_Region          NUMBER(3),
  Nombre_Region       VARCHAR2(100),
  Abrev_Region        VARCHAR2(10),
  Cod_Provincia        NUMBER(5),
  Nombre_Provincia    VARCHAR2(100),
  Cod_Comuna           VARCHAR2(6),
  Nombre_Comuna       VARCHAR2(100)
);
--
CREATE SYNONYM TERRITORIO
FOR T5_CODIGOS_TERRITORIALES_2018;
--
```

Figura D4: Preparación del Ambiente

Una vez validado el ambiente, analizados y depurados los datos con los que se va a trabajar, se procede a realizar el análisis mediante un caso de ejemplo.

DESARROLLO

1.- Determinando el código del colegio y la comuna donde se ubica :



The screenshot shows a database query window with the following SQL statement:

```
SELECT RBD, NOM_RBD, COD_COM_RBD, NOM_COM_RBD
FROM DIRECTORIO WHERE NOM_RBD LIKE '%DE MARIA%SEMINARIO%';
```

Below the query window is a data grid with the following data:

Row#	RBD	NOM_RBD	COD_COM_RBD	NOM_COM_RBD
1	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA

Figura D5: Ubicación del Caso de pruebas para el ejemplo

2.- Determinando la cantidad de alumnos matriculados :

```

SELECT COUNT(1) AS TOTAL_ALUMNOS
FROM MATRICULA
WHERE COD_EST = 8950;
    
```

Row#	TOTAL_ALUMNOS
1	853

Figura D6: Cantidad de Alumnos

3.- Verificando las comunas de residencia de los alumnos, versus la del colegio.

```

SELECT COD_EST, NOMBRE_EST,
       COD_COM_EST, COMUNA_EST, COD_REGION_EST,
       COD_COM_ALU, COMUNA_ALU, COD_REGION_ALU,
       COUNT(1) AS "CANTIDAD_ALUMNOS",
       ROUND(COUNT(1)/&TOTAL_ALUMNOS,4) AS "PORCENTAJE"
FROM MATRICULA
WHERE COD_EST = &CODIGO
GROUP BY COD_EST, NOMBRE_EST,
         COD_COM_EST, COMUNA_EST, COD_REGION_EST,
         COD_COM_ALU, COMUNA_ALU, COD_REGION_ALU
ORDER BY "CANTIDAD_ALUMNOS" DESC, COD_COM_ALU ASC;
    
```

Figura D7: Analizando las comunas de residencia de los alumnos

Row#	COD_EST	NOMBRE_EST	COD_C...	COMUNA_EST	CO...	COD_C...	COMUNA_ALU	CO...	CANTIDAD_ALUMNOS	PORCENTAJE
1	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	13123	PROVIDENCIA	13	280	0,3283
2	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	13101	SANTIAGO	13	206	0,2415
3	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	13120	?U?OA	13	53	0,0621
4	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	13119	MAIPU	13	34	0,0399
5	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	13130	SAN MIGUEL	13	26	0,0305
6	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	13124	PUDAHUEL	13	24	0,0281
7	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	13125	QUILICURA	13	17	0,0199
8	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	13127	RECOLETA	13	15	0,0176
9	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	13106	ESTACION CEN	13	14	0,0164
10	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	13110	LA FLORIDA	13	13	0,0152
11	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	13114	LAS CONDES	13	12	0,0141
12	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	13118	MACUL	13	12	0,0141
13	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	13126	QUINTA NORMA	13	10	0,0117
14	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	13201	PUENTE ALTO	13	9	0,0106
15	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	13107	HUECHURABA	13	8	0,0094
16	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	13108	INDEPENDENCI	13	8	0,0094

Figura D8: Analizando las comunas de residencia de los alumnos

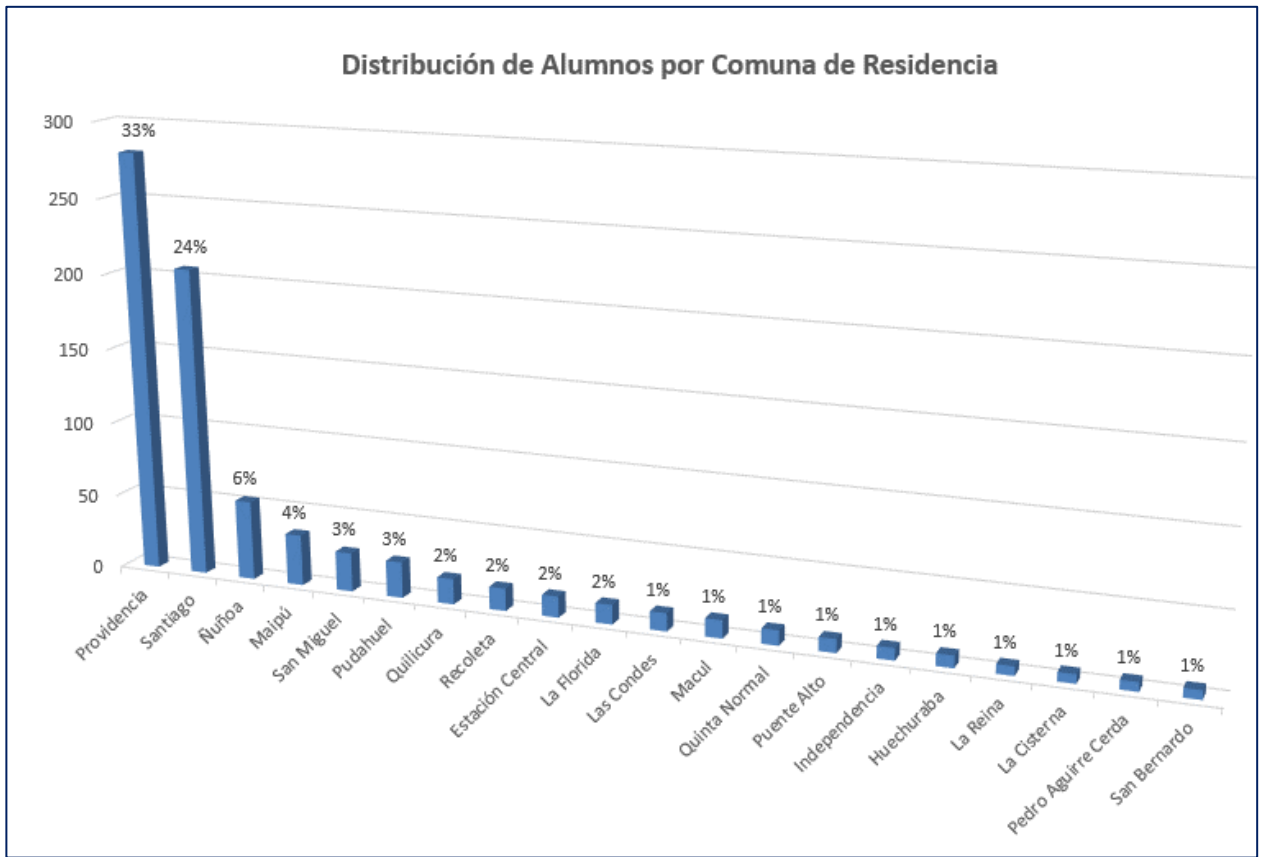


Figura D9: Alumnos según su comuna de residencia (para todo ≥ 5 alumnos)

Hasta aquí va bien, los primeros datos no acusan errores.

4.- Analizando el problema que surge por las comunas “muy distantes”

21	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	13128	RENCA	13	4	0,0047
22	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	10101	PUERTO MONTT	10	3	0,0035
23	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	1101	IQUIQUE	1	3	0,0035
24	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	13102	CERRILLOS	13	3	0,0035
30	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	13403	CALERA DE TA	13	3	0,0035
31	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	5101	VALPARAISO	5	3	0,0035
32	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	8101	CONCEPCION	8	3	0,0035
33	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	9101	TEMUCO	9	3	0,0035
34	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	12101	PUNTA ARENAS	12	2	0,0023
35	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	13103	CERRO NAVIA	13	2	0,0023
36	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	13129	SAN JOAQUIN	13	2	0,0023
37	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	15101	ARICA	15	2	0,0023
38	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	3101	COPIAPO	3	2	0,0023
39	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	4101	LA SERENA	4	2	0,0023
40	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	5109	VIPA DEL MAR	5	2	0,0023
41	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	5502	CALERA	5	2	0,0023
42	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	7101	TALCA	7	2	0,0023
43	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	7301	CURICO	7	2	0,0023
44	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	8108	SAN PEDRO DE	8	2	0,0023
45	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	10102	CALBUCCO	10	1	0,0012
46	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	13104	CONCHALI	13	1	0,0012
52	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	13605	PEÑAFLOR	13	1	0,0012
53	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	14101	VALDIVIA	14	1	0,0012
54	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	2101	ANTOFAGASTA	2	1	0,0012
55	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	4102	COQUIMBO	4	1	0,0012
56	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	5401	LA LIGUA	5	1	0,0012
57	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	5402	CABILDO	5	1	0,0012
58	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	5501	QUILLOTA	5	1	0,0012
59	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	5601	SAN ANTONIO	5	1	0,0012
60	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	5604	EL QUISCO	5	1	0,0012
61	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	5801	QUILPUE	5	1	0,0012
62	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	5803	OLMUE	5	1	0,0012
63	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	5804	VILLA ALEMÁN	5	1	0,0012
64	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	6108	MACHALI	6	1	0,0012
65	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	6115	RENGO	6	1	0,0012
66	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	6301	SAN FERNANDO	6	1	0,0012
67	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	7102	CONSTITUCION	7	1	0,0012
68	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	7401	LINARES	7	1	0,0012
69	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	8205	CURANILAHUE	8	1	0,0012
70	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13123	PROVIDENCIA	13	9114	PITRUFQUEN	9	1	0,0012

Figura D10: Alumnos según su comuna de residencia (para comunas con menos de 5 alumnos)

Se observa que aparecen comunas geográficamente muy distantes, de donde sería **imposible** el desplazamiento diario de los estudiantes para asistir a clases, se decide tratarlos como outliers

5.- Recuento de comunas distintas al interior del colegio

```

SELECT COD_EST, NOMBRE_EST,
       COUNT(DISTINCT COD_COM_EST) AS "COMUNA ESTABLECIMIENTO",
       COUNT(DISTINCT COD_COM_ALU) AS "COMUNAS DE LOS ALUMNOS"
FROM MATRICULA
WHERE COD_EST = &CODIGO
GROUP BY COD_EST, NOMBRE_EST;

```

Row#	COD_EST	NOMBRE_EST	COMUNA ESTABLECIMIENTO	COMUNAS DE LOS ALUMNOS
1	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	1	70

Figura D11: Recuento de comunas distintas

6.- Se validan las comunas y las regiones de residencia de los estudiantes.

```

SELECT COD_EST, NOMBRE_EST, COD_REGION_ALU AS "REGION",
       COUNT(1) AS "ALUMNOS X REGION"
FROM MATRICULA
WHERE COD_EST = &CODIGO
GROUP BY COD_EST, NOMBRE_EST, COD_REGION_ALU
ORDER BY "ALUMNOS X REGION" DESC;

```

Figura D12: Recuento de comunas distintas

Row#	COD_EST	NOMBRE_EST	REGION	ALUMNOS X REGION
1	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13	801
2	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	5	15
3	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	7	6
4	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	8	6
5	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	10	4
6	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	9	4
7	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	6	3
8	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	1	3
9	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	4	3
10	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	12	2
11	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	3	2
12	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	15	2
13	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	14	1
14	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	2	1

16 Row 1 of 14 total rows TESIS@127.0.0.1:1521/xe Modified

Figura D13: Recuento de comunas distintas

7.- Se intenta interpretar los resultados obtenidos.

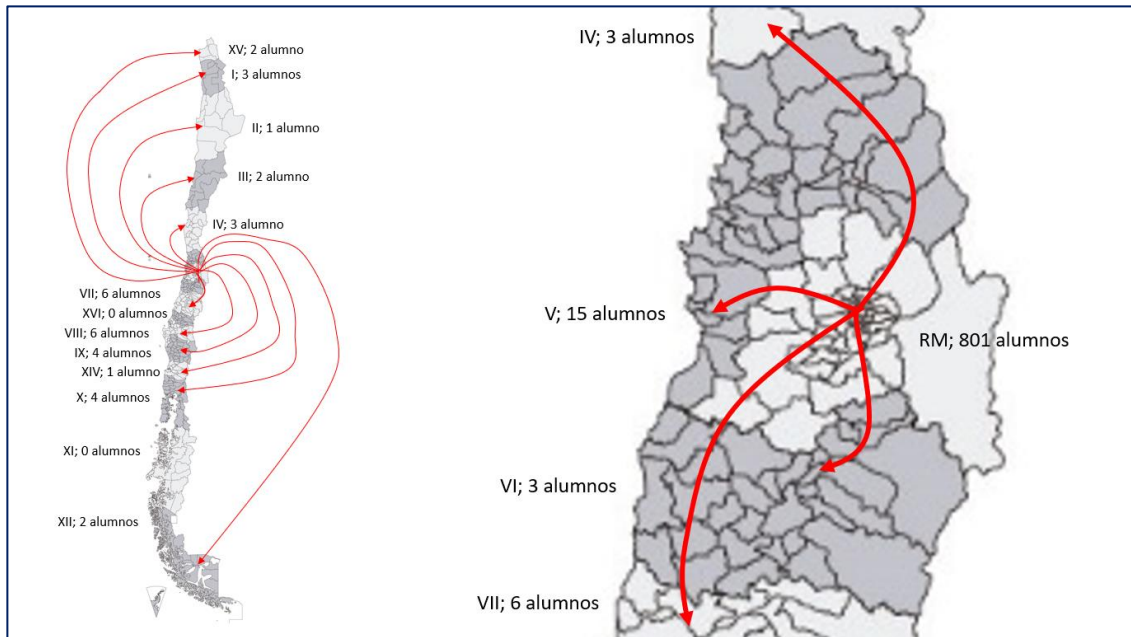


Figura D14: Interpretando los viajes interregionales

Revisando la documentación de los datos para intentar encontrar una explicación a estos valores se encuentra que la región se determina en función de la comuna del alumno y esta aparece con el siguiente comentario:(auto declarado y voluntario).

Como se observa en la Figura, este dato podría contener información imprecisa, ya que podría ser interpretado como PROCEDENCIA (donde nació el alumno) en lugar de RESIDENCIA, por lo que constituyen una fuente de distorsión para el análisis. Se decide por tanto tratar estos datos como OUTLIERS y eliminarlos del dataset, antes de continuar.

Es necesario definir un criterio adecuado para realizar esta tarea. Se analiza la factibilidad de hacerlo calculando las distancias entre la comuna de establecimiento y la del alumno, para luego eliminar los que superen un valor KM determinado.

Esta aproximación además ineficiente es costosa de implementar y no considera la accidentada geografía de nuestro país o la infraestructura y los sistemas de transporte públicos existentes en Santiago y las otras grandes capitales regionales que permitirían desplazarse al alumno distancias considerables con relativa facilidad.

Se opta entonces por un criterio diferente, más sencillo de interpretar e implementar, que además aprovecha el conocimiento de la regionalización del país, al establecer un **grafo no dirigido** entre las regiones que luego se operacionaliza como una matriz de adyacencia.

De esta forma, **si la región del establecimiento y la región del alumno no son la misma, ni adyacentes**, entonces se considerará como dato inválido.

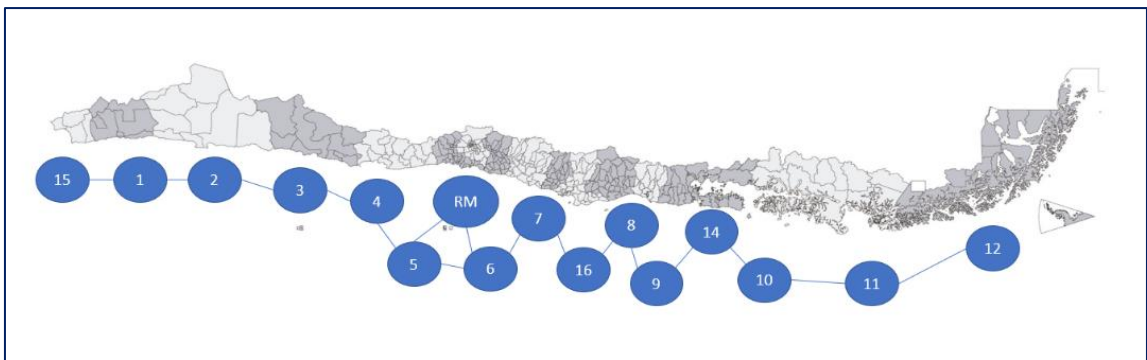


Figura D15: Grafo no dirigido para las regiones del país

7.- Operacionalizando el concepto de matriz de adyacencia, mediante una función:

Row#	COD_REGION	R15	R01	R02	R03	R04	R05	R13	R06	R07	R16	R08	R09	R14	R10	R11	R12
1	R15	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	R01	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
3	R02	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
4	R03	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0
5	R04	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0
6	R05	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0
7	R13	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0
8	R06	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0
9	R07	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0
10	R16	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0
11	R08	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0
12	R09	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0
13	R14	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0
14	R10	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0
15	R11	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
16	R12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1

Figura D16: Matriz de Adyacencia para las regiones del país

```

CREATE OR REPLACE
FUNCTION ADYACENCIA_REGION (REGION_1 IN NUMBER, REGION_2 IN NUMBER)
RETURN NUMBER IS
  COD_REGION_1 VARCHAR2 (3);
  COD_REGION_2 VARCHAR2 (3);
  TEXTO_SQL VARCHAR2 (100);
  ADYACENCIA NUMBER;
BEGIN
  COD_REGION_1 := 'R' || LPAD (REGION_1, 2, '0');
  COD_REGION_2 := 'R' || LPAD (REGION_2, 2, '0');
  TEXTO_SQL := 'SELECT ' || COD_REGION_2 ||
    ' FROM T82_ADYACENCIA_REGION WHERE COD_REGION=' ||
    COD_REGION_1 || '''';
  EXECUTE IMMEDIATE TEXTO_SQL INTO ADYACENCIA;
  RETURN ADYACENCIA;
END ADYACENCIA_REGION;
/

```

Figura D17: Procedimiento almacenado que aprovecha la matriz de adyacencia

8.- Aplicando la función de adyacencia para volver a analizar el colegio de ejemplo:

```

SELECT COD_EST, NOMBRE_EST,
--COD_COM_EST,
COMUNA_EST, COD_REGION_EST,
--COD_COM_ALU,
COMUNA_ALU, COD_REGION_ALU,
ADYACENCIA_REGION(COD_REGION_EST,COD_REGION_ALU) AS ADYACENCIA,
COUNT(1) AS "CANT_ALUMNOS",
ROUND(COUNT(1)/853,4) AS "PORCENTAJE"
FROM MATRICULA
WHERE COD_EST = &CODIGO
GROUP BY COD_EST, NOMBRE_EST,
COD_COM_EST, COMUNA_EST, COD_REGION_EST,

```

Row#	COD_EST	NOMBRE_EST	COMUNA_EST	COD...	COMUNA_ALU	COD...	ADYACENCIA	CANT_ALUMNOS	PORCENTAJE
31	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	PROVIDENCIA	13	VALPARAISO	5	1	3	0,0035
32	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	PROVIDENCIA	13	CONCEPCION	8	0	3	0,0035
33	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	PROVIDENCIA	13	TEMUCO	9	0	3	0,0035
34	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	PROVIDENCIA	13	PUNTA ARENAS	12	0	2	0,0023
35	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	PROVIDENCIA	13	CERRO NAVIA	13	1	2	0,0023
36	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	PROVIDENCIA	13	SAN JOAQUIN	13	1	2	0,0023
37	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	PROVIDENCIA	13	ARICA	15	0	2	0,0023
38	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	PROVIDENCIA	13	COPIAPO	3	0	2	0,0023
39	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	PROVIDENCIA	13	LA SERENA	4	0	2	0,0023
40	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	PROVIDENCIA	13	VIÑA DEL MAR	5	1	2	0,0023

Figura D18: Aplicación de función de adyacencia

9.- Aplicando la función de adyacencia para eliminar outliers y acotar el universo:

```

CREATE TABLE MATRICULA_REDUCIDA
AS SELECT * FROM MATRICULA
WHERE ADYACENCIA_REGION(COD_REGION_EST,COD_REGION_ALU) = 1;

```

Figura D19: Aplicación de función de adyacencia para depurar la base

```

SELECT COUNT(1) AS TOTAL_ALUMNOS
FROM MATRICULA_REDUCIDA
WHERE COD_EST = &CODIGO;

```

Row#	TOTAL_ALUMNOS
1	819

Figura D20: base depurada después del criterio de adyacencia

10.- Volviendo a revisar el colegio, pero sobre la base reducida (sin outliers):

```

SELECT COD_EST, NOMBRE_EST,
       COMUNA_EST, COD_REGION_EST,
       COMUNA_ALU, COD_REGION_ALU,
       ADYACENCIA_REGION(COD_REGION_EST,COD_REGION_ALU) AS ADYACENCIA,
       COUNT(1) AS "CANT_ALUMNOS",
       ROUND(COUNT(1)/&TOTAL_ALUMNOS,4) AS "PORCENTAJE"
FROM MATRICULA_REDUcida
WHERE COD_EST = &CODIGO
GROUP BY COD_EST, NOMBRE_EST,
         COMUNA_EST, COD_REGION_EST,
         COMUNA_ALU, COD_REGION_ALU
ORDER BY "CANT_ALUMNOS" DESC, COMUNA_ALU ASC;

```

Row#	COD_EST	NOMBRE_EST	COMUNA_EST	CO...	COMUNA_ALU	COD...	ADYACENCIA	CANT_ALUMNOS	PORCENTAJE
43	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	PROVIDENCIA	13	PE?AFLO	13	1	1	0,0012
44	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	PROVIDENCIA	13	QUILLOTA	5	1	1	0,0012
45	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	PROVIDENCIA	13	QUILPUE	5	1	1	0,0012
46	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	PROVIDENCIA	13	RENGO	6	1	1	0,0012
47	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	PROVIDENCIA	13	SAN ANTONIO	5	1	1	0,0012
48	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	PROVIDENCIA	13	SAN FERNANDO	6	1	1	0,0012
49	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	PROVIDENCIA	13	TALAGANTE	13	1	1	0,0012
50	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	PROVIDENCIA	13	VILLA ALEMAN	5	1	1	0,0012
51	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	PROVIDENCIA	13	VITACURA	13	1	1	0,0012

Figura D21: revisión del colegio sobre la base depurada

```

SELECT COD_EST, NOMBRE_EST, COD_REGION_ALU AS "REGION",
       COUNT(1) AS "ALUMNOS X REGION"
FROM MATRICULA_REDUcida
WHERE COD_EST = &CODIGO
GROUP BY COD_EST, NOMBRE_EST, COD_REGION_ALU
ORDER BY "ALUMNOS X REGION" DESC;

```

Row#	COD_EST	NOMBRE_EST	REGION	ALUMNOS X REGION
1	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	13	801
2	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	5	15
3	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	6	3

Figura D22: revisión del colegio sobre la base depurada

Se observa que si bien el criterio de regiones adyacentes aplicado no es perfecto, el resultado que se obtiene con él hace mucho más sentido, considerando que para un establecimiento de la región metropolitana podría ser posible tener alumnos de las regiones 5 y 6.

11., Se establece el valor actualizado del número de comunas del colegio ejemplo:

The screenshot shows a SQL query in a text editor and its results in a data grid. The query is as follows:

```

2 SELECT COD_EST, NOMBRE_EST,
3       COUNT(DISTINCT COD_COM_EST) AS "COMUNA ESTABLECIMIENTO",
4       COUNT(DISTINCT COD_COM_ALU) AS "COMUNAS DE LOS ALUMNOS"
5 FROM MATRICULA_REDUcida
6 WHERE COD_EST = &CODIGO
7 GROUP BY COD_EST, NOMBRE_EST;
8

```

The data grid below the query shows the following results:

Row#	COD_EST	NOMBRE_EST	COMUNA ESTABLECIMIENTO	COMUNAS DE LOS ALUMNOS
1	8950	COLEGIO COMPANIA DE MARIA-SEMINARIO	1	51

Figura D23: revisión del colegio sobre la base depurada

12.- Se crea un procedimiento que permita calcular el valor de los n términos de la sumatoria , para los N establecimientos educacionales que están en la base.

En primer lugar se crea nueva tabla para realizar los cálculos parciales

```

DROP TABLE INDICADOR_RIESGO;

CREATE TABLE INDICADOR_RIESGO AS
select cod_est, cod_com_est, cod_com_alu, count(1) as num_alu
from matricula_reducida
group by cod_est, cod_com_est, cod_com_alu
order by cod_est, cod_com_alu ;

update INDICADOR_RIESGO A
set A.cod_com_est = ( select B.cod_comuna from COMUNA B
                    where B.codigo_comuna = A.cod_com_est ) ;

update INDICADOR_RIESGO A
set A.cod_com_alu = ( select B.cod_comuna from COMUNA B
                    where B.codigo_comuna = A.cod_com_alu ) ;

alter table INDICADOR_RIESGO add matriculados number (5) ;

update INDICADOR_RIESGO A
set A.matriculados = (select count(1) from matricula_reducida B
                    where B.cod_est = A.cod_est) ;

commit;

alter table INDICADOR_RIESGO
add constraint pk_indicador_riesgo
primary key (cod_est, cod_com_alu);

```

Figura D24: calculando paso a paso el indicador de riesgo

y#	COD_EST	COD_COM_EST	COD_COM_ALU	NUM_ALU	MATRICULADOS
37	440	03102	03101	2	1157
38	440	03102	03102	1153	1157
39	440	03102	04102	1	1157
40	441	03103	02101	1	234
41	441	03103	03101	29	234
42	441	03103	03103	191	234
43	441	03103	03301	4	234
44	441	03103	03302	2	234

Figura D25: calculando paso a paso el indicador de riesgo

13. Luego se agregan a la tabla los n valores necesarios para el cálculo de la entropía

```

alter table INDICADOR_RIESGO add probabilidad number;

update INDICADOR_RIESGO set probabilidad = (NUM_ALU/MATRICULADOS);

alter table INDICADOR_RIESGO add logaritmo number;

update INDICADOR_RIESGO set logaritmo = log(10,probabilidad);

alter table INDICADOR_RIESGO add prob_x_log_prob number;

update INDICADOR_RIESGO set prob_x_log_prob = (logaritmo * probabilidad);

commit;

```

Figura D26: agregando los n términos de la sumatoria para determinar la entropía

COD_EST	COD_COM_EST	COD_COM_ALU	NUM_ALU	MATRICULADOS	PROBABILIDAD	LOGARITMO	PROB_X_LOG_PROB
4	15101	01107	7	918	0,00762527233115468	-2,11774464118699	-0,0161483796168942
4	15101	01402	2	918	0,00217864923747277	-2,66181268553726	-0,00579915617764109
4	15101	15202	1	918	0,00108932461873638	-2,96284268120124	-0,00322749747407543
7	15101	01401	5	1051	0,00475737392959087	-2,32263271169222	-0,0110496323106195
7	15101	15202	5	1051	0,00475737392959087	-2,32263271169222	-0,0110496323106195
8	15101	01107	5	594	0,00841750841750842	-2,07481644064517	-0,0174647848539156
9	15101	15101	1359	1393	0,975592246949031	-0,0107316596914692	-0,0104697239918928
10	15101	15101	159	168	0,946428571428571	-0,0239121574054114	-0,0226311489729786
12	15101	01107	3	303	0,0099009900990099	-2,00432137378264	-0,0198447660770559
13	15101	15101	274	274	1	0	0
14	15101	01101	2	320	0,00625	-2,20411998265592	-0,0137757498915995

Figura D27: agregando los n términos de la sumatoria para determinar la entropía

14. A continuación se agregan columnas para el total de la sumatoria y una auxiliar para ordenar y poder calcular el valor para el total

```
alter table INDICADOR_RIESGO add sumatoria number default 0;
alter table INDICADOR_RIESGO add orden number default 0;
commit;

DECLARE
    SUMA NUMBER;
    i NUMBER;
    --
    cursor ee is
    select distinct cod_est
    from INDICADOR_RIESGO
    ORDER BY COD_EST ASC;
    --
    CURSOR COLE (COLEGIO IN NUMBER) IS
    select *
    from INDICADOR_RIESGO
    where cod_est=COLEGIO
    order by num_alu desc, cod_com_alu ASC
    FOR UPDATE OF SUMATORIA, ORDEN;
    --
BEGIN
    FOR c IN ee LOOP
        SUMA := 0;
        I := 0;
        FOR R IN COLE( C.COD_EST ) LOOP
            suma := suma + r.prob_x_log_prob ;
            I := I + 1;
            update INDICADOR_RIESGO
            SET SUMATORIA = suma,
                ORDEN = I
            WHERE CURRENT OF COLE;
        END LOOP;
    END LOOP;
    COMMIT;
END;
/
```

Figura D28: calculando la suma acumulativa para cada término

De esta forma es posible, no solo realizar el cálculo, sino que también revisar cualquier valor individual en caso de ser necesario. Se verifican los valores calculados por este procedimiento, para el establecimiento de ejemplo.

15. Se establece una segunda tabla, resumen de la anterior, donde además se aplica la fórmula para el cálculo del riesgo en base a la entropía y la cantidad de comunas

```

01
02 create table PELIGROSIDAD as
03 select INDICADOR_RIESGO.cod_est,
04         cod_com_est,
05         INDICADOR_RIESGO.matriculados,
06         orden as comunas_distintas,
07         to_char(0-sumatoria,'0d999') as Shannon
08 from INDICADOR_RIESGO join ( select cod_est, max(orden) as maxim
09                             from INDICADOR_RIESGO
10                             group by cod_est ) x
11     on  INDICADOR_RIESGO.cod_est = x.cod_est
12     and INDICADOR_RIESGO.orden = x.maxim
13 order by 1 asc ;
14
15 alter table PELIGROSIDAD add Riesgo number;
16
17 update PELIGROSIDAD set riesgo = shannon * comunas_distintas;
18
19 commit;
20
21 select * from PELIGROSIDAD;
22

```

Data Grid

Data Grid | DBMS Output (disabled) | Script Output

Cancel

Row#	COD_EST	COD_COM_EST	MATRICULADOS	COMUNAS_DISTINTAS	SHANNON	RIESGO
5163	8942	13401	1065	23	0.122	2,806
5164	8944	13123	843	49	1.045	51,205
5165	8945	13123	581	31	1.092	33,852
5166	8950	13123	819	51	1.035	52,785
5167	8953	13132	634	30	0.674	20,22
5168	8954	13123	957	40	1.159	46,36
5169	8956	13123	181	31	0.884	27,404

Figura D30: calculando la suma acumulativa para cada término

16. Por último, se realiza un cruce de información con el directorio de nombres de establecimientos para producir la base resultado y producto final de este estudio.

```

22
23 create table RESULTADO_FINAL as
24 select  c.cod_est,
25         d.nom_rbd,
26         c.cod_com_est,
27         d.nom_com_rbd,
28         c.matriculados,
29         c.comunas_distintas,
30         c.shannon as entropia,
31         c.riesgo
32 from PELIGROSIDAD c join directorio d on c.COD_EST = d.RBD
33 order by riesgo desc, comunas_distintas desc, matriculados desc;
34

```

Row#	COD_EST	NOM_RBD	COD_CO...	NOM_COM_RBD	MATRIC...	COMUNAS...	ENTROPIA	RIESGO
1	8485	LICEO INSTITUTO NACIONAL	13101	SANTIAGO	4246	75	1.482	111,15
2	25818	COLEGIO PART. ADULTOS INSTITUTO ICEL	13101	SANTIAGO	1935	64	1.395	89,28
3	8499	INTERNADO NACIONAL BARROS ARANA	13101	SANTIAGO	1260	63	1.380	86,94
4	8928	LICEO JOSE VICTORINO LASTARRIA	13123	PROVIDENCIA	2573	61	1.408	85,888
5	10715	INSTITUTO SUBTTE.LUIS CRUZ MARTINEZ	13601	TALAGANTE	618	57	1.468	83,676
6	31030	ESCUELA PRE MILITAR HEROES DE LA CONCEPCION	13109	LA CISTERNA	1666	58	1.415	82,07
7	8491	LICEO DE APLICACION RECTOR JORGE E SCHNE	13101	SANTIAGO	1854	58	1.387	80,446
8	8487	LICEO JAVIERA CARRERA	13101	SANTIAGO	2612	54	1.411	76,194
9	8927	LICEO CARMELA CARVAJAL DE PRAT	13123	PROVIDENCIA	1762	51	1.426	72,726
10	8603	CENTRO EDUCATIVO SALESIANOS ALAMEDA	13101	SANTIAGO	1693	52	1.308	68,016
11	8748	COLEGIO INSTITUTO ALONSO DE ERCILLA	13101	SANTIAGO	1939	54	1.240	66,96
12	2552	INSTITUTO POLITECNICO	06310	SANTA CRUZ	1195	65	1.006	65,39
13	8930	LICEO B 42 TAJAMAR	13123	PROVIDENCIA	978	45	1.436	64,62
14	15777	COLEGIO JUAN LUIS VIVES	06101	RANCAGUA	558	63	1.022	64,386
15	17607	COLEGIO INGLES BRITISH ROYAL SCHOOL	08101	CONCEPCION	1041	59	1.089	64,251
16	24473	COLEGIO TEC.HOTELERIA Y GASTRONOMIA ACHIGA CO	13114	LAS CONDES	528	44	1.446	63,624
17	8643	INSTITUTO COMERCIAL BLAS CANAS	13101	SANTIAGO	702	49	1.282	62,818
18	31167	CENTRO EDUCACIONAL INTEGRAL DE ADULT ACUARIO	13101	SANTIAGO	207	42	1.486	62,412
19	9472	LICEO POLITECNICO SAN LUIS	13130	SAN MIGUEL	1393	51	1.222	62,322
20	24840	COLEGIO DE ADULTOS INSTITUTO TABANCURA SEDE LONDRES	13101	SANTIAGO	235	44	1.405	61,82
21	8926	LICEO N°7 DE PROVIDENCIA LUISA SAAVEDRA	13123	PROVIDENCIA	1257	43	1.435	61,705
22	8492	LICEO MANUEL BARROS BORGONO	13101	SANTIAGO	831	45	1.370	61,65
23	20310	COLEGIO DE ADULTOS PULMAHUE	13101	SANTIAGO	296	44	1.401	61,644

Figura D31: calculando la suma acumulativa para cada término

De esta forma se logra obtener el valor para el indicador de riesgo, según la fórmula planteada, no solo para el colegio del ejemplo, sino también, para todos los establecimientos educacionales de la base; cumpliéndose el objetivo propuesto al inicio de este documento.

Anexo E: Resultados Obtenidos

En el siguiente vínculo https://github.com/roybarrera/Indicador_de_Riesgo_EE se pueden encontrar las siguientes bases, resultados directos e indirectos del presente trabajo :

- 1.- Base con el indicador de riesgo potencial o peligrosidad asociado a cada uno de los establecimientos educacionales del país.
- 2.- Consolidación de las comunas en una única base con información territorial, superficie (KM2), población por género y edad (Censo) y coordenadas GPS.
- 3.- Matriz de adyacencia de comunas de Chile
- 4.- Matriz de adyacencia de regiones de Chile

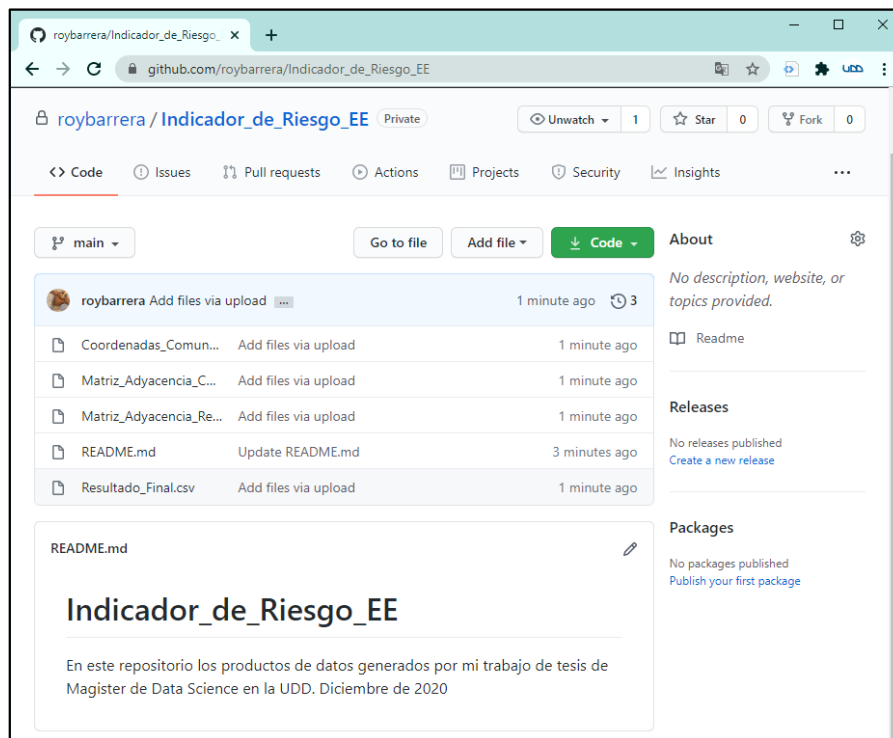


Figura E1: Archivos publicados en Github
