



ANÁLISIS SIMCE EN EL RENDIMIENTO ESCOLAR EN CHILE

Grupo 12

Integrantes:

Agustin Córdova

Martín Muñoz

Vicente de Gavardo

Juan Levipil

Sebastian Mena

Índice

Contexto y Motivación.....	3
Objetivos.....	4
Datos.....	5
Preguntas De Investigación.....	7
Diseño Tentativo.....	8
BIBLIOGRAFÍA.....	9

Contexto y Motivación

¿Qué es el SIMCE?

La definición del SIMCE es Sistema de Medición de la Calidad de la Educación. Creado en 1968, es una prueba estandarizada administrada por el Ministerio de Educación para evaluar a los estudiantes y su rendimiento académico, con el fin de determinar el nivel de educación en Chile.

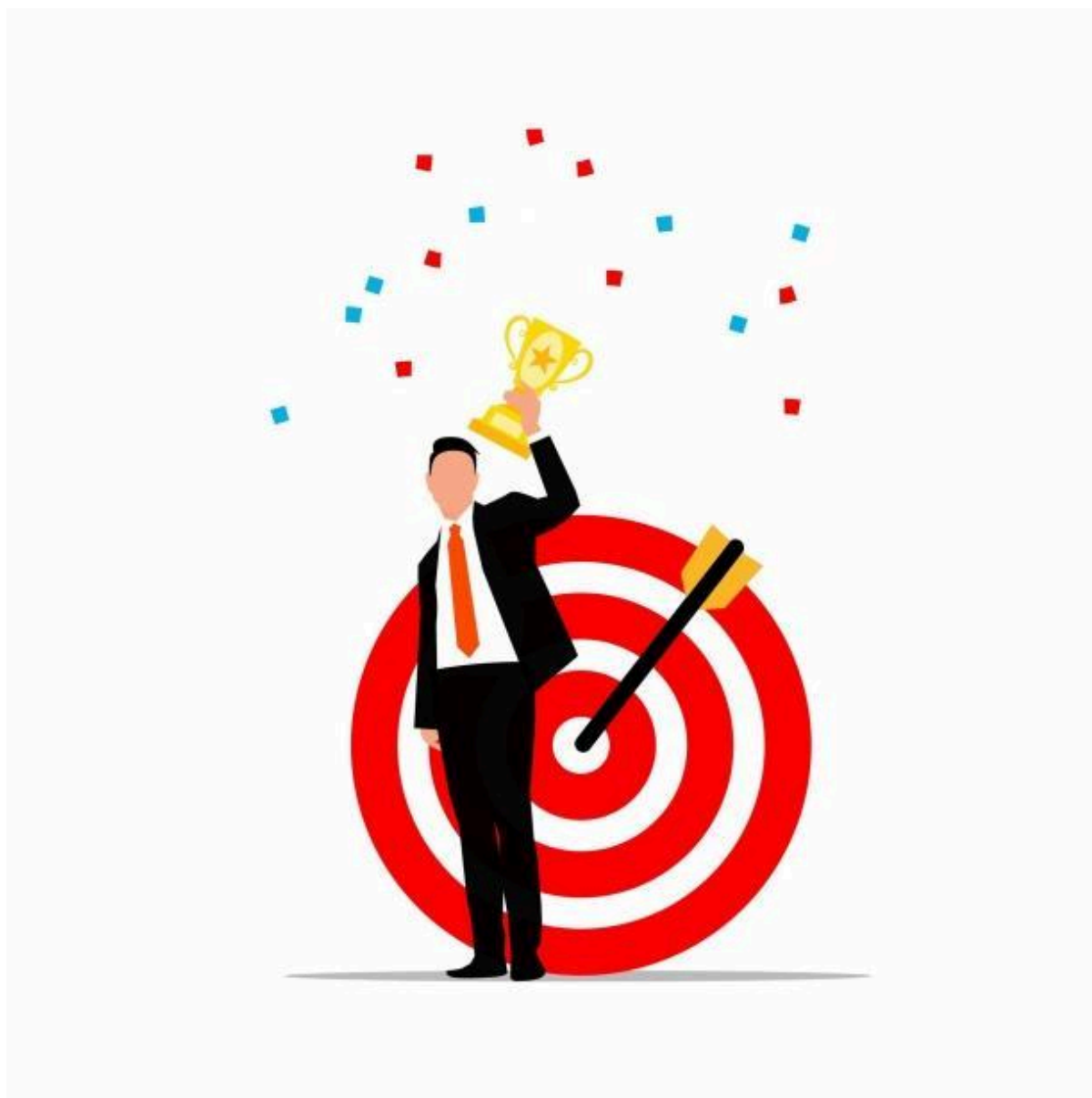
Nuestra motivación es investigar si existe una diferencia significativa entre los puntajes de matemáticas y lenguaje en nuestro país. Esto nos permitirá identificar áreas clave en las que se deben enfocar los esfuerzos para mejorar la educación en Chile.



Objetivos

Con nuestro proyecto buscamos comparar no solo la calidad de educación a través de los años, si no también ver si existe alguna correlación con el nivel de formación según qué tan vulnerable o privilegiado sea un sector.

Con nuestros datos extraídos directamente de la base de datos del ministerio de educación, organizaremos y limpiaremos los archivos para convertirlos en un *dataFrame* legible y fácil de comprender utilizando el lenguaje de programación *Python* para posteriormente comparar estos datos con mayor facilidad.



Datos

Los data frames consisten en las calificaciones de las pruebas SIMCE evaluadas en distintos años y grados, en base a los colegios de las distintas comunas y regiones a nivel nacional. Estos datos fueron extraídos directamente de la agencia de calidad de la educación. Los cursos a analizar son los siguientes:

2014 - 4º básico

2016 - 6º básico

2019 - 8º básico

2022 - 2º medio

Nuestros datos son proporcionados públicamente por el ministerio de educación. Dentro de estos archivos se encuentra información sobre las 346 comunas de Chile y más de 7 mil colegios existentes ***aunque cabe destacar que mientras más antiguos sean los datos, menos colegios y comunas estarán disponibles para comparar.***



LAS VARIABLES PRESENTADAS SON LAS SIGUIENTES

nom_rbd: Nombre del establecimiento

agno: Año de evaluación

cod_grupo: Grade socioeconómico (1 = bajo, 5 = alto)

cod_rural_rbd: Código de ruralidad (1 = urbano, 2 = rural)

nom_reg_rbd: Nombre de la región del establecimiento

nom_com_rbd: Nombre de la comuna del establecimiento

nom_pro_rbd: Nombre de la provincia del establecimiento

prom_lec2m_rbd: Puntaje promedio del establecimiento en lectura

prom_mate2m_rbd: Puntaje promedio del establecimiento en matematicas

dif_lec2m_rbd: Diferencia en lectura respecto al año anterior

dif_mate2m_rbd: Diferencia en matematicas respecto al año anterior

cod_depe2: Codigo de dependencia administrativa (1 = municipal, 2 = subvencionado, 3 = pagado, 4 = servicio local)

Preguntas De Investigación

- ¿Los resultados del SIMCE difieren significativamente dependiendo de la comuna?
- ¿Cómo afectó la pandemia a los resultados en la prueba?
- ¿Existe una mejoría o un deterioro en los resultados del SIMCE a través de los años?
- ¿Existe una diferencia significativa en los resultados del SIMCE entre estudiantes de colegios públicos y privados?
- ¿Las generaciones mantienen sus puntajes al pasar de los años?
- ¿Afecta el grupo socioeconómico del establecimiento en el puntaje del SIMCE?
- ¿Existe mucha diferencia entre los resultados de institutos rurales y los urbanos?

Diseño Tentativo

Para el desarrollo de este proyecto, utilizaremos Python como nuestro lenguaje de programación, junto con diversas librerías que facilitan la limpieza y organización de datos.

Una vez obtenidos los datos desde la página del gobierno y descargados los archivos Excel, procederemos a convertirlos en *dataframes* utilizando el módulo Pandas. Con los *dataframes* cargados como variables, realizaremos la limpieza de campos innecesarios, fechas incorrectas y valores atípicos o absurdos.

Después de limpiar los *dataframes*, responderemos las preguntas de investigación mediante la manipulación de estos datos. Por último, crearemos gráficos para ilustrar nuestras conclusiones de manera efectiva, utilizando Matplotlib u otras librerías similares.

LIBRERIAS A UTILIZAR:

pandas

numpy

matPlotLib

geopandas

BIBLIOGRAFÍA

Portal de Estudios. (s. f.). <https://informacionestadistica.agenciaeducacion.cl/#/bases>

O'Neil, C. (2016). Weapons of Math Destruction. Obtenido de

https://edisciplinas.usp.br/pluginfile.php/7574239/mod_resource/content/1/%28FFLCH%29%20LIVRO%20Weapons%20of%20Math%20Destruction%20-%20Cathy%20ONeal.pdf