

Untitled

Jorge Melendez B.

2024-06-07

Informe de Análisis de Datos

Master en Data Science - Universidad de las Americas

Alumno: Jorge Melendez Bastías

1. Definición del problema / Objetivo de investigación

La matriculación de alumnos en la enseñanza obligatoria (Preescolar, Básica y Media) en Chile es un conjunto de datos de gran relevancia y tamaño para analizar en el contexto del análisis de datos. Con aproximadamente 3.6 millones de observaciones por año y un total de 36 variables a analizar, este conjunto de datos proporciona una amplia gama de información para explorar. ### Objetivo General Analizar la matriculación de alumnos en la enseñanza obligatoria en Chile durante los años 2019-2023, utilizando técnicas de análisis de datos en R Studio y MySQL, con el objetivo de identificar tendencias y patrones que puedan ser útiles para la toma de decisiones en el ámbito educativo.

Objetivos Especificos

- Analizar la evolución de la matriculación de alumnos en la enseñanza obligatoria en Chile durante los años 2019-2023.
- Identificar tendencias y patrones en la matriculación de alumnos por nivel educativo y región.
- Evaluar el impacto de políticas educativas y otros factores en la matriculación de alumnos.
- Utilizar técnicas de programación en R Studio, especialmente las librerías del mundo tidiverse, para el análisis de datos.
- Utilizar MySQL para almacenar y gestionar eficientemente la gran cantidad de registros, aplicando buenas prácticas para agilizar el análisis.

Fuente de Datos *Los datos utilizados en este análisis son proporcionados por el Ministerio de Educación de Chile (MINEDUC) a través de la plataforma de datos abiertos (<https://datosabiertos.mineduc.cl/>). Se cumple con la protección de datos personales según lo establecido en la Ley N° 19.628.*

2. Introducción

Descripción del conjunto de datos seleccionado y los objetivos del análisis.

3. Preprocesamiento de Datos

Descripción detallada de los pasos tomados para limpiar y preparar los datos para el análisis utilizando Tidyverse.

4. Análisis Exploratorio de Datos

Implementación de códigos en R para explorar los datos. Esto puede incluir visualizaciones de datos, medidas de tendencia central, etc.

5. Modelado de Datos

Aplicación de técnicas de modelado de datos y algoritmos de aprendizaje automático, si corresponde.

6. Interpretación de Resultados

Discusión de los resultados del análisis y su relevancia para los objetivos establecidos en la introducción.

7. Código

Incluir todo el código R utilizado en el análisis como un apéndice o integrado en el informe. El código debe ser legible y propiamente comentado.