

CoderHouse

Curso Data Science

Informe del Proyecto Final

**Análisis socioeducativo de los habitantes de la Ciudad de
Buenos Aires**

Profesor: Damian Dapuetto

Tutor: Héctor Alonso

Grupo de Trabajo:

Lucia Buzzeo, Lucia Hukovsky,
Jose Saint German, Juan Martín Carini

3 de septiembre de 2022



Índice

1. Presentación del problema y fuente de información	3
1.1. Presentación del problema	3
1.2. Definición de la fuente de información	3
2. Pregunta y objetivos de investigación	3
3. Orden de trabajo	3
4. Introducción a las variables: Análisis exploratorio de los datos	3
4.1. Análisis univariado	3
4.2. Análisis bivariado	3
4.3. Análisis multivariado	3
5. Modelos analíticos	3

1. Presentación del problema y fuente de información

1.1. Presentación del problema

Nos es de gran de interés vivir en una comunidad con políticas públicas eficaces que mejoren las condiciones de vida de las personas. En este sentido, hemos decidido analizar los diferentes ejes que en nuestro país se rigen por políticas publicas. Al respecto, encontramos una gran limitación en el eje de educación al reconocer que su acceso dista de ser equitativo. Este aspecto no nos resultó una novedad, sin embargo, nos dio el pie para comenzar una investigación que permita dar una explicación teórica a la problemática. En concreto, nos ha permitido conocer mejor la situación educativa actual de CABA y descubrir las principales variables que afectan el nivel educativo.

El análisis realizado en el marco del presente proyecto podría establecer una base de requerimientos que permitan generar políticas públicas efectivas, no solo en el ámbito educativo, sino en el económico, cultural, social y geográfico, entre otros.

1.2. Definición de la fuente de información

Para trabajar esta problemática, hemos decidido recurrir a la [datos abiertos](#) del Gobierno de la Ciudad de Buenos Aires para el año 2019. El mismo está disponible en la base de [Encuesta Anual de Hogares](#) del GCBA.

Esta encuesta contiene información demográfica, social, económica, educativa y de salud de 14319 habitantes de la Ciudad, la cual es una muestra representativa que permite obtener un vistazo de la población de la Ciudad.

2. Pregunta y objetivos de investigación

Nuestro objetivo principal es descubrir las principales variables intervinientes en el nivel máximo educativo alcanzado por la población de la Ciudad Autónoma de Buenos Aires (CABA).

De este objetivo principal se desprenden los siguientes sub-objetivos:

- Determinar si la ubicación geográfica del encuestado es determinante para alcanzar ciertos niveles educativos. De este objetivo se desprende determinar la relación entre el nivel educativo y la comuna del encuestado, así como la relación entre la misma variable y el hecho de que el encuestado habite en una villa de emergencia.
- Establecer la fuerza con la que el nivel socio-económico afecta la variable target.
- Explorar la relación del target con otras variables, como el sexo del encuestado, la cantidad de hijos, la afiliación de salud o la edad.

3. Orden de trabajo

Este trabajo estará dividido en 3 partes:

1. **Introducción a las variables del problema:** Se hará un análisis de las variables en donde se buscará conocer su performance dentro del dataset y su potencial significanos para la pregunta que buscamos responder. A la vez, queremos ver cómo las variables interactúan entre si. Esta parte es lo que se conoce como análisis univariado, bivariado y multivariado,
2. **Modelos analíticos:** En esta sección se llevarán a cabo diversos modelos analíticos y algoritmos que nos servirán para acercarnos a la respuesta a nuestra pregunta de investigación,
3. **Conclusión:** Haremos conclusiones finales sobre nuestros hallazgos. Además, discutiremos posibles limitaciones que tuviera y plantearemos futuras líneas de análisis a partir del análisis presente.

4. Introducción a las variables: Análisis exploratorio de los datos

4.1. Análisis univariado

4.2. Análisis bivariado

4.3. Análisis multivariado

5. Modelos analíticos