

Desafío 1

Especialización en Políticas Públicas | Clínica de Datos | Jueves 7 de Abril 2022

Sobre los datos

En este desafío, se trabajará con un dataset de delitos ocurridos en la Ciudad Autónoma de Buenos Aires en 2019 y 2020. El dataset cuenta con un preprocesamiento previo de estandarización de variables, limpieza de datos ubicados geográficamente fuera de la ciudad e incorporación de nuevas variables de trabajo. Los datos originales pueden encontrarse en el portal BADATA

Sobre la Ciudad Autónoma de Buenos Aires La Ciudad Autónoma de Buenos Aires se encuentra organizada administrativa en Comunas (Fuente: Wikipedia - GCBA). A su vez, las comunas se componen de 1 o más barrios.

Abrir el dataset

Se recomienda cargar el dataset incluyendo el parámetro `stringsAsFactors = TRUE`

Análisis exploratorio

¿Cuántas filas y columnas tiene el dataset?

¿Cómo es la estructura del dataset? Comentar sobre los nombres de las variables y sobre los tipos de datos de las variables. ¿Alguna variable es de texto libre? Analizar particularidades de las variables “fecha” y “comuna”.

Realizar un resumen de los datos y comentar ¿Qué tipo de delito es el más frecuente en el dataset? ¿Cuántas observaciones no tienen información sobre la comuna?

Preprocesamiento de los datos

Parte A

Del resumen anterior, se observa que la variable comuna funciona como una variable continua numérica. Si bien la comuna es un número, la variable es categórica: el número representa el nombre de la comuna. Se busca obtener un resumen estadístico de delitos por comuna. Para ello se sugieren los siguientes pasos:

- seleccionar unicamente la variable de interés (comuna)
- filtrar las observaciones sin comuna (NA). *Tip: la función `is.na()` permite obtener las observaciones que son NA o sin información*
- transformar la variable numérica a una variable categórica *Tip: la función `as_factor()` permite transformar una variable a categórica*
- generar el resumen estadístico

¿Cuál es la comuna con más delitos?

Parte B

Habiendo detectado la comuna con mayor delitos del dataset, se busca discriminar la cantidad de delitos por año y por comuna. Para ello, se sugieren los siguientes pasos:

- generar un dataset agrupando por año y comuna, y calculando la cantidad de delitos
- ordenar de forma descendiente
- mostrar los primeros 10 datos

Comentar los resultados.