

## Introducción a la Ciencia de Datos - 2C 2022

### Guía para la entrega parcial

Trabajar de manera individual. Elegir uno de los tres datasets que se proponen (vuelos, árboles o seguros). **No usar el que se analizó para la entrega grupal.**

Explorar el dataset según algún aspecto específico que se busca analizar (hay una lista de sugerencias abajo) y realizar una infografía que permita extraer la mayor cantidad de información posible sobre este aspecto a partir de gráficos y textos.

El formato de entrega es una carilla **hoja A4**, con la cantidad de gráficos y texto que consideren necesaria, que sea legible con un zoom de no más del 125%. Entregar preferentemente en PDF.

Recuerden que siempre vale mostrar versiones preliminares a familiares o amigos.

**Fecha límite para entrega:** 17 de octubre de 2022 a las 00h00.

---

Ideas de aspectos para explorar. Es opcional elegir uno de esta lista; pueden hacerlo con algún aspecto que les interese particularmente, o que encuentren interesante.

#### Vuelos

- Distribución de retrasos por aerolíneas. ¿Qué aerolíneas son más confiables? ¿Por qué? ¿Para qué rutas?
- Funcionamiento de los distintos aeropuertos de salida. ¿Hay alguno que sea mejor? Tener en cuenta tantos factores como sea posible. Recuerdan que está la tabla de clima.
- Funcionamiento de los distintos aeropuertos de llegada. ¿Hay alguno que sea mejor? Tener en cuenta tantos factores como sea posible.
- Distribución de retrasos a lo largo del año (pueden usar `lubridate`; capítulo 16 del libro) para una aerolínea dada
- Distribución de retrasos a lo largo del año según origen o destino.
- Analizar la performance de aviones individuales.
- Analizar cuántos pasajeros son afectados por retrasos teniendo en cuenta el tamaño de los aviones.
- Analizar la dependencia de los retrasos con eventos climáticos puntuales.

#### Árboles

- Distribución geográfica de árboles en la ciudad de Buenos Aires. ¿Existe algún patrón? ¿Cómo cambia según la especie?
- Alturas de los árboles según su ubicación en la ciudad. ¿Existe alguna relación entre los árboles que se encuentran en avenidas más anchas y la altura? ¿Y según la orientación de la calle y la vereda en la que

se encuentren?

- Estados de las planteras según comuna o ubicación en la ciudad (avenidas comerciales, avenidas de mucha circulación, etc.)
- Altura de los árboles en relación con la ocupación de las planteras.
- Analizar la dependencia del ancho de los árboles en función de su altura y cómo cambia según diferentes factores.
- Analizar si se verifican las restricciones dadas en [este sitio](#).

## Seguros

- ¿Se encuentra una relación entre la cantidad de hijos y el estado de fumador / no fumador?
- ¿Cuánto del costo del seguro de personas no fumadoras puede ser explicado por la cantidad de hijos? Segmentar para BMI mayor o menor que  $30 \text{ kg/m}^2$ .
- Analizar si hay alguna relación entre el sexo y el promedio del costo del seguro, si la hay explique por qué ocurre. Piensen qué variables pueden ser confundidoras.
- Analizar la relación entre el costo del seguro y la ubicación geográfica. De nuevo, no olviden qué variables pueden ser las confundidoras.
- Analizar la dependencia de los costos con la ubicación geográfica para un dado estado de salud (bmi y si es fumador o no; y tal vez número de hijos).