

Trabajo Práctico

Laboratorio de Datos, 2do cuatrimestre de 2021

Último día para entregar: **Martes 02 de Noviembre**, 23:59hs

Grupos: <https://docs.google.com/spreadsheets/d/1q1hji2KH35anK9yIEXGwvF6MaLfQUMFykfvGbF-dgh8/edit?usp=sharing>

Link para entrega de Rmd y html: <https://forms.gle/Y1PZbNhqiTeKTS6aA>

Objetivo: Hacer un análisis descriptivo / exploratorio con un conjunto de datos.

Datasets:

2021: Aterrizajes y despegues procesados por la Administración Nacional de Aviación Civil (ANAC) Ministerio de Transporte.

<https://datos.gob.ar/dataset/transporte-aterrizajes-despegues-procesados-por-administracion-nacional-aviacion-civil-anac>

Descripción:

La idea **no** es que hagan un análisis exhaustivo utilizando todas las herramientas que aprendimos en la materia. Tienen que mostrarnos que pueden describir y explorar los datos, buscar patrones y relaciones entre variables, generar preguntas y representar los datos de forma eficiente para generar hipótesis que den posibles respuestas a las preguntas que plantearon. Por esta razón, es fundamental que la entrega sea una combinación de **código**, **visualizaciones** y **texto**. El **texto debe apoyarse en las visualizaciones y el código para formar una narrativa** de lo que ustedes hayan aprendido y explorado de los datasets.

El trabajo tiene dos partes: en la primera deberán **limpiar datasets**, e **incorporar información externa extraída de Wikipedia**. En la segunda, deberán **explorar los datos** y emplear lo aprendido en la materia para responder preguntas que les resulten interesantes. Piensen en que el destinatario es un compañero de ustedes. No den cosas por sabidas, expláyense en todo lo que consideren necesario.

Parte 1:

- Hagan una limpieza de los datos, revisando cada una de las variables a fin de ver que no contengan errores o casos extraños.
- Quedarse con vuelos nacionales.
- El dataset tiene el código de los aeropuertos. Agreguen nuevas variables con la ciudad, provincia y coordenadas que corresponde a cada aeropuerto. Pueden encontrar esta información en [este artículo de Wikipedia](#)
- Calculen la distancia recorrida por cada vuelo. La función **distHaversine** del paquete **geosphere** puede resultarles útil.
- Calculen la velocidad media de cada vuelo. Para esto deberán asociar distintas filas del dataset. El dataset contiene **Despegues** y **Aterrizajes**, que permiten encontrar la hora de partida y llegada de cada vuelo.

Parte 2:

- Usando las herramientas que aprendieron, exploren los datos. Piensen preguntas que les despierten interés y generen hipótesis que den posibles respuestas, en base a las magnitudes y visualizaciones que vimos en clase.
- Reporten las preguntas que se hicieron y las hipótesis que generaron, buscando construir una narrativa. Es necesario mezclar texto, código y visualizaciones.

Sugerencia para la Parte 2: Tomen cada variable y piensen qué información les brinda, cómo la representarían y resumirían. Luego, tomen cada par de variables posibles y piensen qué relación se obtiene de ellas. Tomen las que les resulten más interesantes, resumanlas y representenlas.

Entrega:

- Notebook en versión html y Rmd.