

# Indicaciones Trabajo Práctico

## Fundamentos de Ciencia de Datos - Recursado 2024

### Descripción General

El objetivo de este Trabajo Práctico es que los estudiantes adquieran habilidades en la manipulación, limpieza y análisis exploratorio de datos utilizando Python.

En grupos de a 2 (dos) estudiantes, deberán seleccionar un dataset de su interés (*ver posibles fuentes más abajo*) y, luego de realizar cualquier tarea de limpieza, transformación y adecuación que consideren necesaria, llevar adelante un análisis exploratorio, incluyendo el cálculo e interpretación de medidas descriptivas y la elaboración de herramientas gráficas de visualización.

### Detalles de trabajo

#### Selección del dataset

Cada grupo deberá seleccionar un dataset de una temática que resulte de su interés. Para sacar el máximo provecho de las herramientas del análisis exploratorio abordadas durante el curso, se valorará que el dataset seleccionado este compuesto por al menos 6 variables con datos de diferente naturaleza (numéricos, categóricos, geográficos, etc.).

**IMPORTANTE.** El dataset elegido no deberá provenir de un repositorio en el que se expongan, además, herramientas para su análisis (por ejemplo, Kaggle).

Algunas fuentes sugeridas:

- Datos Abiertos de Argentina (<https://www.datos.gob.ar>)
- Datos Abiertos del Ministerio de Salud (<http://datos.salud.gob.ar/>)
- Datos Abiertos Ciudad de Buenos Aires (<https://data.buenosaires.gob.ar/dataset/>)
- Datos de la NASA sobre la Tierra (<https://www.earthdata.nasa.gov/>)
- Datos de Agencias Federales de los EE. UU., incluyendo tópicos como atención médica, energía, transporte y educación, entre otros (<https://catalog.data.gov/dataset>)
- FiveThirtyEight (<https://data.fivethirtyeight.com/>)
- Datos Abiertos del Banco Mundial (<https://datos.bancomundial.org/>)
- Datos Abiertos de UNICEF (<https://data.unicef.org/>)
- Datos de Investigación de CONICET (<https://datosdeinvestigacion.conicet.gov.ar/>)

#### Descripción del dataset elegido

Se deberá proporcionar una breve descripción del dataset, incluyendo su origen, tipo de datos que contiene y un diccionario de variables que especifique la información que se encuentra en cada columna.

## Manipulación y transformación del dataset

Realizar cualquier tarea de limpieza, transformación y adecuación del dataset que se considere necesaria para su posterior análisis, incluyendo, si corresponde, manejo de datos faltantes, duplicados y/o inconsistentes y modificación de nombres de variables y/o estructura de los datos. Cada una de las acciones que se realicen sobre el dataset deberá estar documentada y debidamente justificada.

## Análisis exploratorio de los datos

Realizar un análisis exploratorio de los datos, extrayendo cualquier tipo de información que el grupo juzgue de interés según la temática elegida. El mismo deberá incluir el cálculo de las medidas descriptivas que se consideren adecuadas de acuerdo a las características de las variables, así como la construcción de diferentes herramientas de visualización, **las cuales deberán estar debidamente interpretadas y comentadas.**

Se valorará positivamente la presentación de diferentes tipos de gráficos, así como el correcto uso de etiquetas de ejes, paletas de colores, leyendas, etc.

## Sobre la entrega

### Detalles de la entrega:

Cada grupo deberá entregar un **archivo .zip o .rar** con nombre **Apellido1\_Apellido2.zip** que contenga:

- Archivo/s de datos utilizado/s.
- Notebook (.ipynb) con el informe del trabajo realizado.

### Sobre el informe a entregar:

- El informe debe contener una descripción clara y concisa de cada una de las etapas mencionadas en la sección Detalles de trabajo, así como el código de Python utilizado, el cual deberá correr sin errores y estar correctamente documentado y estructurado.
- Las distintas secciones deberán estar organizadas de una manera prolija, utilizando títulos y subtítulos, y todos los comentarios, interpretaciones y elaboraciones que se realicen a partir de los resultados del análisis deberán incluirse como celdas de texto en lenguaje *Markdown*.
- La entrega se realizará vía Campus en un espacio destinado para tal fin. Sólo uno de los integrantes del grupo deberá hacer la carga del archivo.

**Fecha de entrega:** viernes 8 de noviembre (*tentativa*).