

# Trabajo Práctico 1

## Introducción

El presente trabajo práctico tiene como objetivo aplicar los conocimientos adquiridos en las unidades 2, 3 y 4 del curso. En particular, nos enfocaremos en el análisis descriptivo y exploratorio de datos en diferentes contextos, haciendo uso de herramientas y técnicas de manipulación, limpieza y visualización de datos. Los ejercicios propuestos abordarán problemáticas y conjuntos de datos relacionados con la demografía y economía de las provincias de Argentina, así como con el tipo de cambio y el turismo en el país.

Este trabajo consta de dos ejercicios principales. En el primero, se realizará un análisis exploratorio de un conjunto de datos que contiene información demográfica y económica de las provincias de Argentina. En el segundo, se llevará a cabo un análisis descriptivo y exploratorio de dos dataframes relacionados con los tipos de cambio oficial y paralelo en Argentina, y la cantidad de turistas que llegaron desde países limítrofes.

Cada ejercicio incluirá la manipulación, limpieza y curación de datos, así como la aplicación de técnicas de visualización y estadística descriptiva para identificar patrones, tendencias y relaciones entre las variables analizadas. A lo largo de este trabajo, se pondrán en práctica las habilidades y conocimientos adquiridos en las unidades mencionadas, para garantizar un análisis riguroso y fundamentado en datos de calidad.

## Ejercicio 1

### Objetivo:

El propósito de este ejercicio es llevar a cabo un análisis exploratorio de un conjunto de datos que contiene información demográfica y económica de las provincias de Argentina (provincias\_datos\_2022.xlsx). Este análisis tiene como objetivo realizar una exploración de las variables e identificar características relevantes o asociaciones que permitan obtener conocimientos relevantes para la toma de decisiones informadas en diversos ámbitos, como el económico, social y político.

### Instrucciones:

1. Realizar las tareas de limpieza, transformación y normalización de los datos contenidos en el archivo "provincias\_datos\_2022.xlsx" a fin de asegurar la calidad y homogeneidad de la información.

2. Una vez que los datos hayan sido curados, proceder a explorar cada variable y las posibles relaciones entre las variables del conjunto de datos, tales como la población, la cantidad de hogares, los ingresos laborales, la tasa de empleo y la superficie en km<sup>2</sup>. Para ello, se deben emplear técnicas de visualización y estadística descriptiva.
3. Al concluir el análisis, se espera que el estudiante presente sus hallazgos y conclusiones, destacando cualquier patrón o tendencia que resulte relevante y contribuya a una mayor comprensión de las relaciones entre las variables estudiadas.
4. Escribir el informe incluyendo código Python utilizado, las visualizaciones y las conclusiones en un cuaderno Google Colab. El dataframe final curado debe ser exportado a formato .parquet.

## Ejercicio 2

Análisis Descriptivo y Exploratorio de Dataframes sobre Tipos de Cambio y Turismo en Argentina

### Objetivo:

El propósito de este ejercicio es realizar un análisis descriptivo y exploratorio de dos dataframes, relacionados con los tipos de cambio oficial y paralelo en Argentina y la cantidad de turistas que llegaron desde países limítrofes. El análisis permitirá conocer mejor la distribución de las variables y las posibles relaciones entre ellas, asegurando que los datos sean precisos y confiables antes de continuar con el análisis.

### Instrucciones:

1. Importar los dos dataframes desde el archivo "tc\_turistas.xlsx": un dataframe con datos diarios de tipos de cambio oficial y paralelo en Argentina durante el período de junio de 2017 a febrero de 2020, y otro con la cantidad de turistas que llegaron desde países limítrofes en frecuencia mensual.
2. Ajustar la frecuencia del dataframe de tipos de cambio a mensual, para que sea compatible con el dataframe de turistas.
3. Explorar ambos dataframes mediante gráficos y medidas de resumen. Verificar la distribución de las variables, la presencia de valores faltantes o extremos, y las correlaciones entre las variables.
4. Realizar un análisis exploratorio de los datos para conocer mejor la distribución de los mismos y las posibles relaciones entre las variables.
5. Si se encuentra algún problema en los datos, determine la mejor manera de solucionarlo. Verificar que los datos sean precisos y confiables antes de continuar con el análisis.
6. Realizar cualquier tarea de limpieza o curación de los datos que sea necesaria para asegurar la calidad de la información.

7. Presentar los hallazgos y conclusiones del análisis, destacando cualquier patrón o tendencia que resulte relevante y contribuya a una mayor comprensión de las relaciones entre las variables estudiadas.
8. Escribir el informe incluyendo código Python utilizado, las visualizaciones y las conclusiones en un cuaderno Google Colab. El dataframe final curado debe ser exportado a formato .parquet.

## Entrega del TP:

- Fecha de entrega (por el campus):
  - C1: Martes 23 de Mayo 2023
  - C2: Lunes 22 de Mayo 2023
- Debe realizarse en **grupos de no más de 2 personas**
- La entrega del TP se realizará en un archivo .zip que contendrá dos carpetas: "ejercicio\_1" y "ejercicio\_2", con la resolución de los ejercicios 1 y 2, respectivamente.
- La resolución de cada ejercicio deberá ser presentada en un cuaderno de Google Colab, con un formato tipo informe con títulos, texto explicativo y el código de python utilizado. Además, la carpeta del archivo .zip de cada ejercicio deberá contener los dataframes finales (curados) en formato .parquet (no los datasets originales).
- El nombre del archivo zip deberá contener la comisión y los apellidos de los alumnos en orden alfabético, por ejemplo: C1\_DellaCeca\_Tadeo.zip  
C2\_Geary\_Manson.zip

## Consideraciones:

- El informe con la resolución de cada ejercicio:
  - debe ser claro y conciso (para ello deberán agregar los títulos correspondientes y el texto aclaratorio/descriptivo/conclusiones que consideren necesario)
  - contener el código Python utilizado
  - contener al menos 4 tipos de visualizaciones diferentes (boxplot, scatterplots, etc) en la resolución de cada ejercicio, que sean adecuadas para lo que desean mostrar