Ciclo de formación en Python

Introducción a Python con orientación al análisis de datos

# Trabajo Final

## Consigna 1

A su dirección llega un pedido externo pidiendo información sobre las fuentes de recaudación del estado. Creen que la mejor forma de comunicar estos números es en relación con al producto interno bruto.

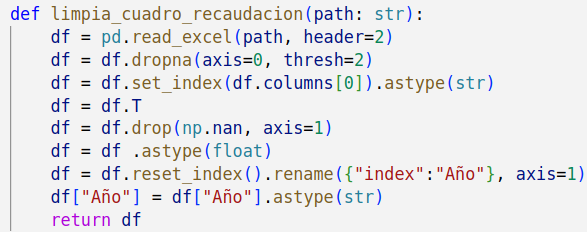
Para poder generar estas series será necesario:

### A. Descargar y limpiar los datos

Las tablas las brinda el INDEC en los siguientes vínculos:

* Recaudación: https://www.indec.gob.ar/ftp/cuadros/economia/070301\_2021.xlsx
* PIB: https://www.indec.gob.ar/ftp/cuadros/economia/sh\_oferta\_demanda\_09\_23.xls

Unx compañerx que ya había trabajado con estos datos le facilita unas funciones para limpiar esos datos y que se acoplen al formato “tidy data”. Analice las dos funciones y explique lo mejor que pueda el proceso línea por línea.





### B. Exportar trabajo

Entender el código que le envío su compañero le ha llevado bastante trabajo, por lo que muy razonablemente cree que es conveniente guardar el progreso hecho y exportar los dos cuadros limpios en un excel. Además a otrx compañerx le sería muy beneficioso contar con los cuadros limpios que usted ha trabajado. Por esa razón decide exportar su trabajo en formato excel.

Exporte los dos cuadros en un excel que contenga dos hojas, una por cada cuadro.

### C. Explorar y seleccionar información

* Recaudación: determine cuales son las 5 principales fuentes de recaudación además de la recaudación bruta total. Genere un nuevo dataframe que contenga las variables de año, recaudación total y las cinco fuentes de recaudación que usted encontró.
* PIB: Genere un dataframe que solo contenga la variable de año y la de producto bruto interno

### D. Consolidar información

Genere un nuevo dataframe producto de la cruza de los datos de los cuadros previamente trabajados según la variable año.

## Consigna 2

### A. En función del PIB

Transforme los datos de los ingresos en valores porcentuales relativos al PIB. Es decir que el total del PIB = 100.

### C. Análisis temporal

Quiere ver el progreso de las series en el tiempo. Cree un gráfico de líneas en el que pueda comparar el progreso de las cinco series seleccionadas simultáneamente.

### E. Exportación de los datos

Parece que van a incluir su análisis en el informe de su dirección pero le piden poder publicar los datos en formato excel. Consideran conveniente tener las cinco series seleccionadas con su valor original y su valor porcentual relativo al PIB.

Exporte en un excel cinco hojas por cada serie, con dos colmnas (monto y monto relativo al PIB) con los datos de los últimos 10 años.

### ¡BONUS!: Análisis estructural

Si bien las series tienen tienen más observaciones, el período de análisis tiene un recorte en la última década. Genere dos gráficos de torta con *Plotly*, que contengan a las cinco fuentes de ingresos seleccionadas y una sexta llamada “resto”. Las seis categorías deben sumar 100. El primer gráfico debe ser del promedio de la década y el segundo del último año observado.

¿Ha cambiado esa distribución estructuralmente comparando las dos tortas?

**!**

**Ya que su gráfico será publicado, el equipo de difusión le solicita que utilice coma para separar los decimales y punto como separador de miles.**