Big Data y Machine Learning (UBA) Grupo 4 - 2025

Trabajo Práctico 1: Jugando con APIs y WebScraping

Reglas de formato y presentación

* El trabajo debe estar debidamente documentado comentado (utilizando #) para que tanto los docentes como sus compañeros puedan comprender el código fácilmente.
* El mismo debe ser completado en este Jupyter Notebook y entregado como tal, es decir en un archivo .ipynb

Fecha de entrega:

Viernes 5 de Septiembre a las 13:00 hs

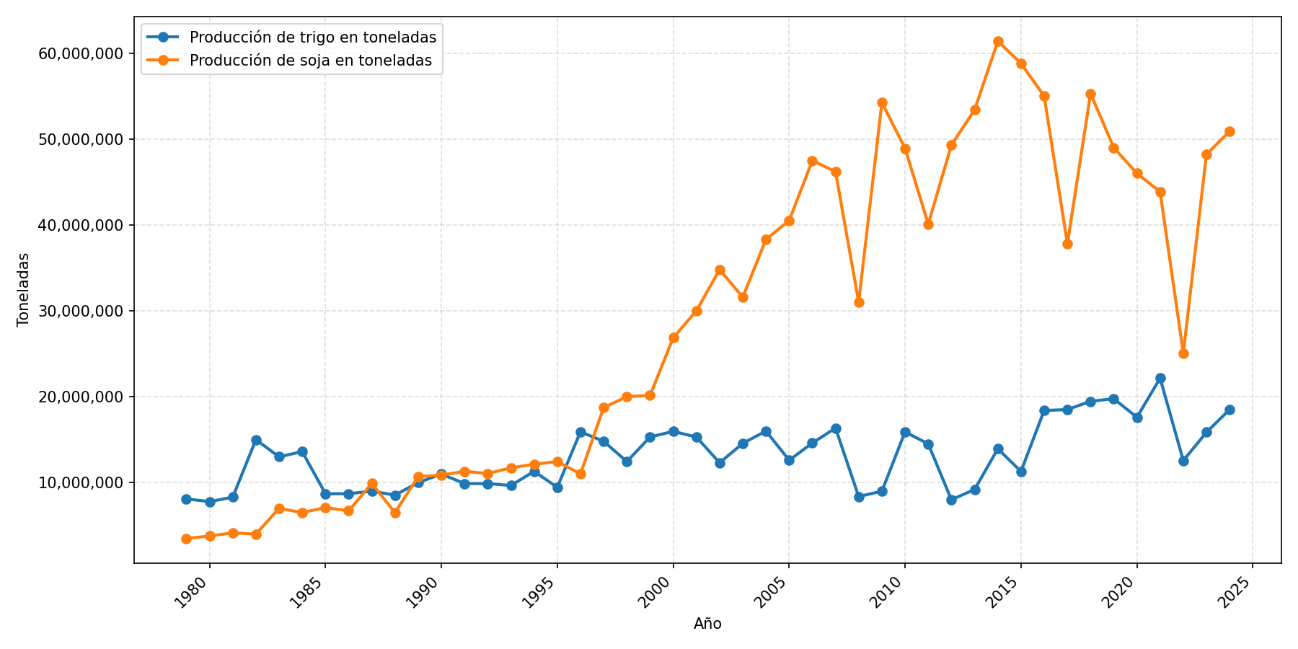
Modalidad de entrega

* Al finalizar el trabajo práctico deben hacer un último *commit* en su repositorio de GitHub llamado “Entrega final del tp”.
* Asegurense de haber creado una carpeta llamada TP1. Este Jupyter Notebook y el correspondiente al TP1 deben estar dentro de esa carpeta.
* También deben enviar el link de su repositorio -para que pueda ser clonado y corregido- a mi correo 25RO35480961@campus.economicas.uba.ar. Usar de asunto de email *"Big Data - TP 1 - Grupo #"* y nombrar el archivo *"TP1\_Grupo #"* donde # es el número de grupo que le fue asignado.
* La última versión en el repositorio es la que será evaluada. Por lo que es importante que:
  + No envien el correo hasta no haber terminado y estar seguros de que han hecho el *commit y push* a la versión final que quieren entregar.
  + No hagan nuevos *push* despues de haber entregado su versión final. Esto generaría confusión acerca de que versión es la que quieren que se les corrija.
* En resumen, la carpeta del repositorio debe incluir:
  + El codigo
  + Un documento Word (Parte A) donde esten las figuras y una breve descripción de las mismas.
  + El excel con los links webscrappeados (Parte B)

### Parte A

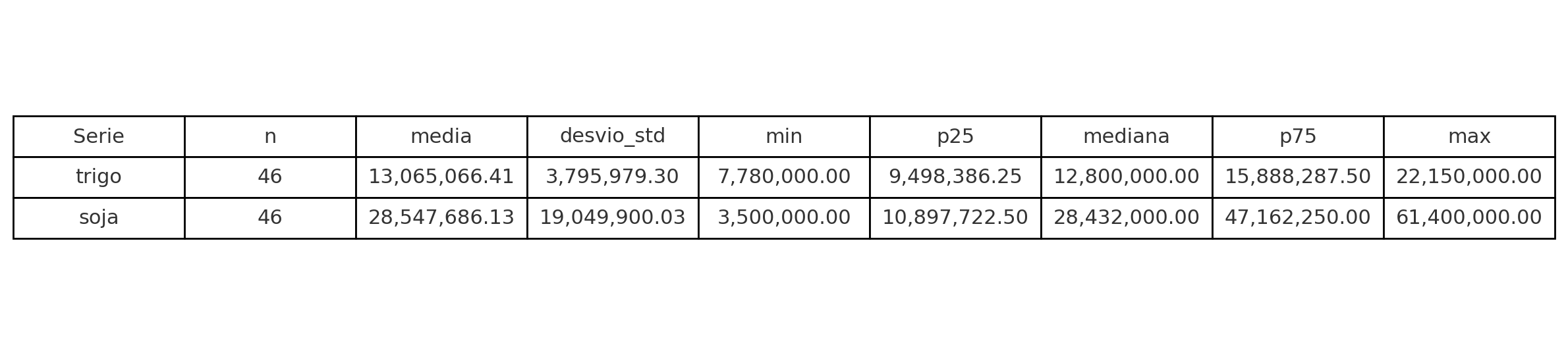
#### **Ejercicio 1 - Jugando con APIs**

Usando la API del Series de Tiempo de la Republica Argentina [link](https://datosgobar.github.io/series-tiempo-ar-api/applications/) , obtener dos series de indicadores del Ministerio de Agricultura, Ganadería y Pesca. En total, dicho Ministerio tiene 72 series anuales y pueden buscar dos serie de indicadores de su interés.



El gráfico muestra la evolución anual (1979–2025) de la producción de trigo (azul) y soja (naranja) en toneladas. La soja crece con fuerza desde fines de los ’90, alcanza picos >60 Mt a mediados de los 2010, tiene caídas fuertes (2009, 2018 y 2023) y rebota luego hacia ~50 Mt. El trigo se mantiene en niveles menores (8–22 Mt), con tendencia suave al alza y menor volatilidad, con picos cerca de 20–22 Mt alrededor de 2019–2021. En síntesis: la soja creció fuertemente y se consolidó como el principal producto exportador del país, en cambio, el trigo, aunque mantuvo una producción relativamente estable, perdió participación y dejó de ser protagonista desde mediados de los ’90.

#### **Ejercicio 2 - Repaso de Pandas**

Realicen una estadistica descriptiva ambas series de indicadores del Ministerio de Ganaderia y presente en una tabla en su reporte. Dicha tabla debe mostrar: numero de observaciones, media, desvío standard, min, p25, mediana, p75 y max. Comente la tabla en **el reporte**.

La tabla resume 46 años de datos y muestra que la soja produce, en promedio, unas 28,5 Mt, más del doble que el trigo (13,1 Mt). Además, la soja es mucho más volátil: su desvío estándar ronda 19,0 Mt frente a 3,8 Mt del trigo , y su rango intercuartílico es mucho más amplio (36,3 Mt vs 6,4 Mt). En un año típico, la mediana de soja (28,4 Mt) también supera con claridad a la del trigo (12,8 Mt). Los extremos refuerzan esta idea: la soja alcanza picos de 61,4 Mt y pisos de 3,5 Mt, mientras que el trigo se mueve entre 7,8 y 22,2 Mt, evidenciando un comportamiento más estable. En conjunto, la soja creció fuertemente y domina en volumen, pero con marcada variabilidad interanual; el trigo, en cambio, muestra niveles más bajos y un recorrido más parejo.