**LABORATORIO N°1**

**INTRO PYTHON FOR DS**

Andrés Felipe Esquivel Ruiz; Código 12967.

Nathaly Daniela Mejía Meléndez; Código 84588.

Universidad ECCI

Seminario Big Data

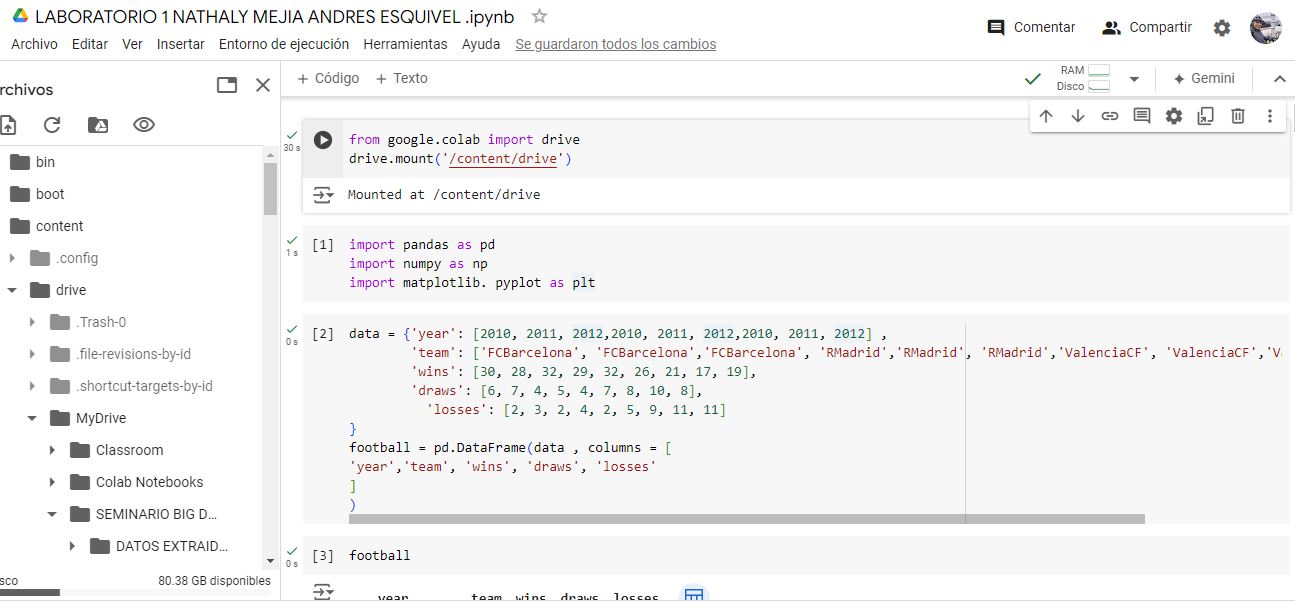
Elías Buitrago Bolívar

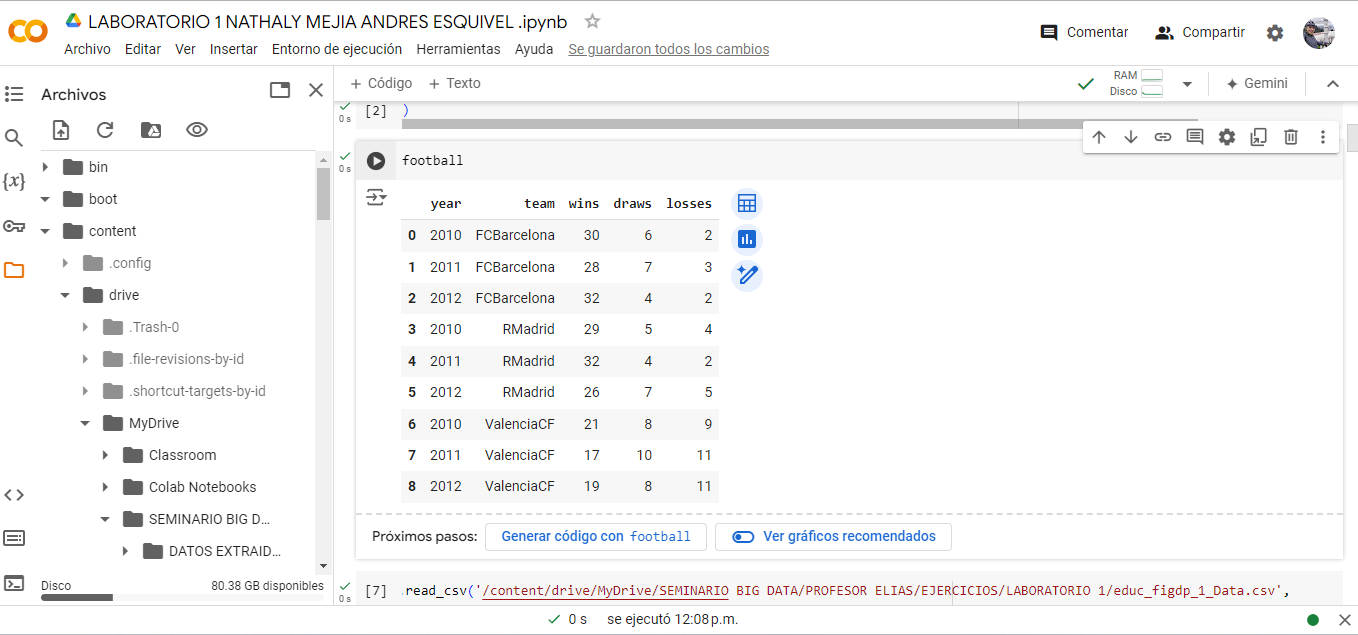
Julio de 2024

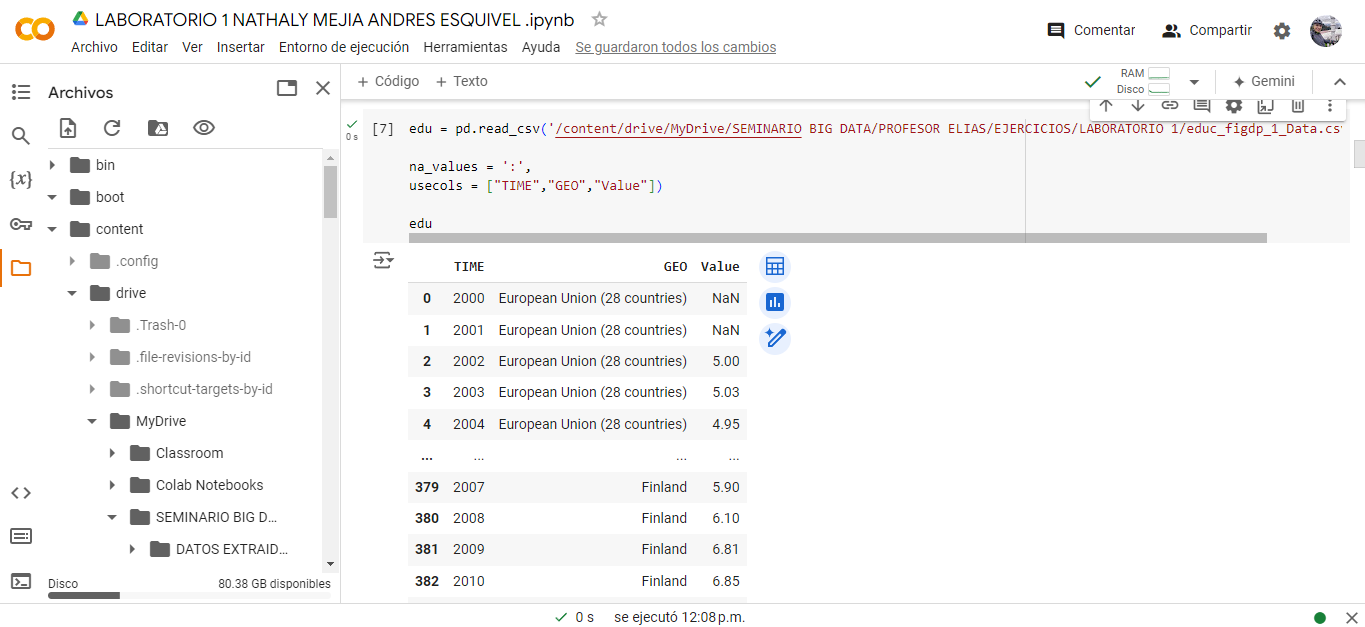
**Descripción de la Actividad:**

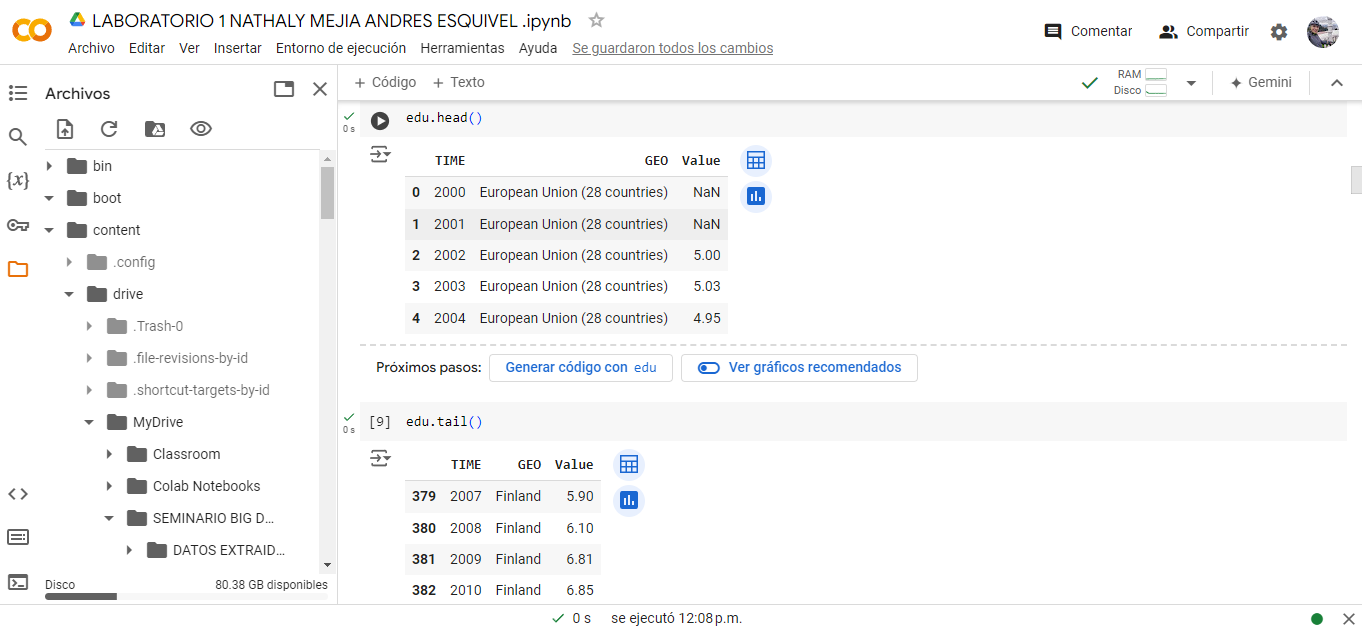
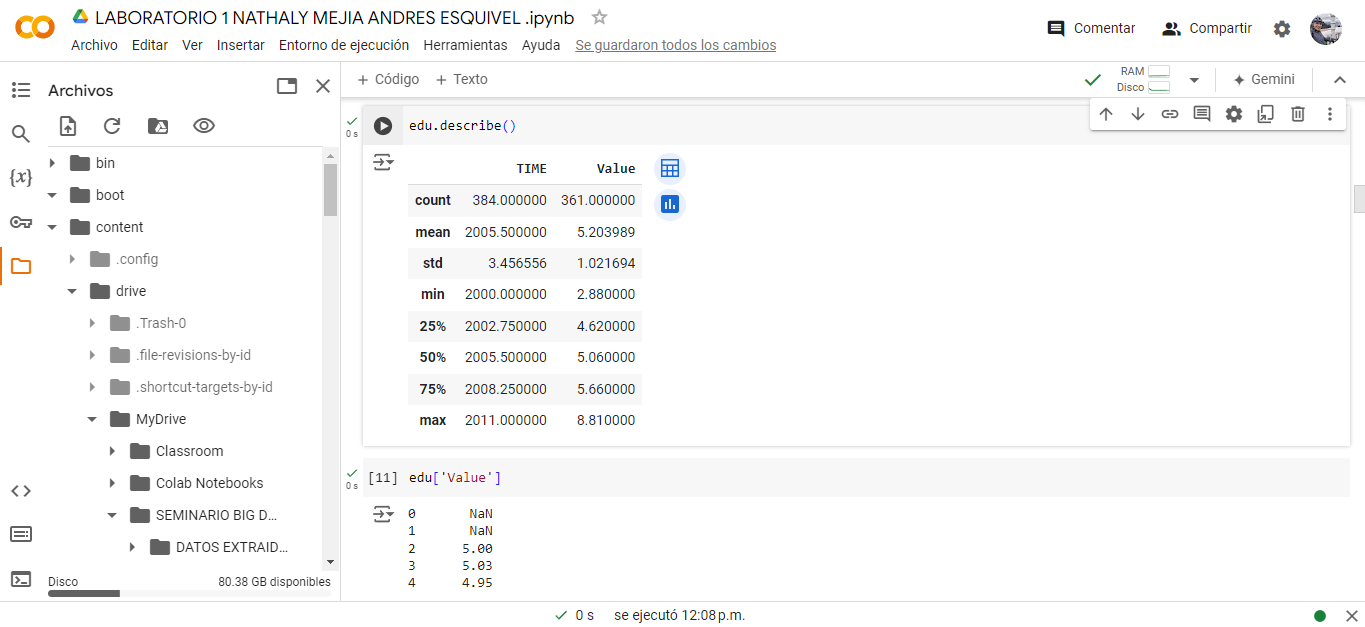
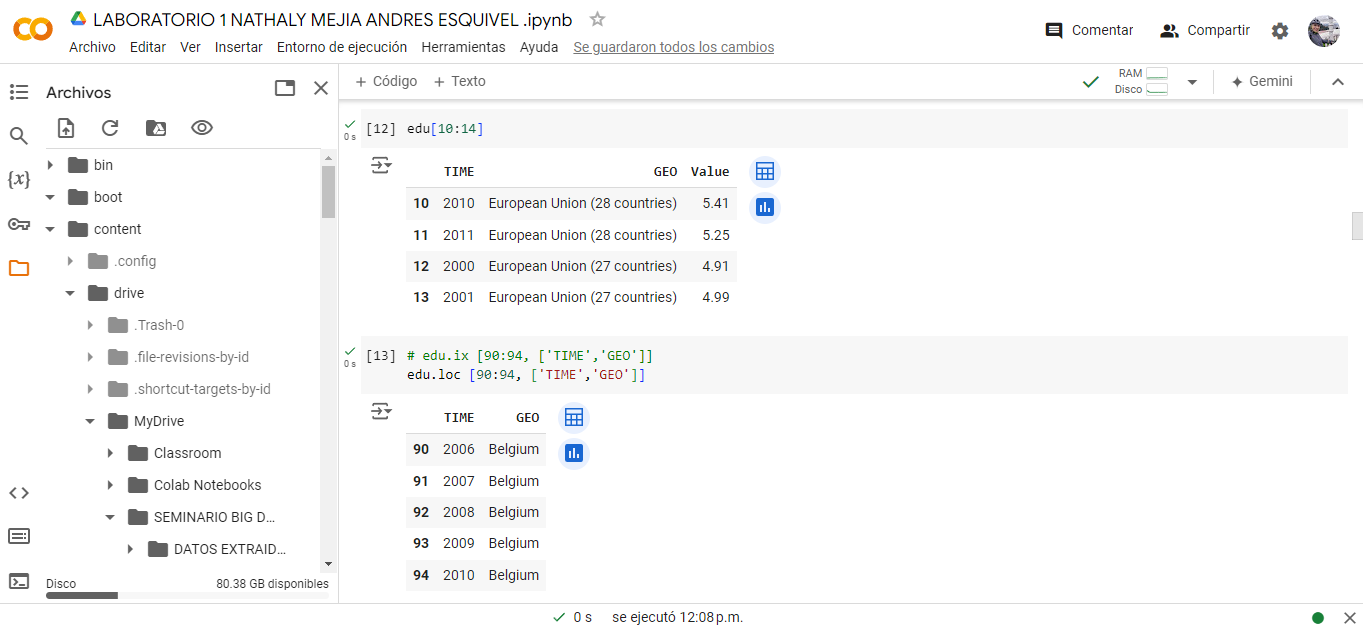
Comprender la introducción al lenguaje de programación en Python mediante el entorno Jupyter Notebook de acuerdo a la guía y la data compartida por el ingeniero Elías Buitrago Bolívar, A continuación decribimos los pasos realizados en Google Colab.

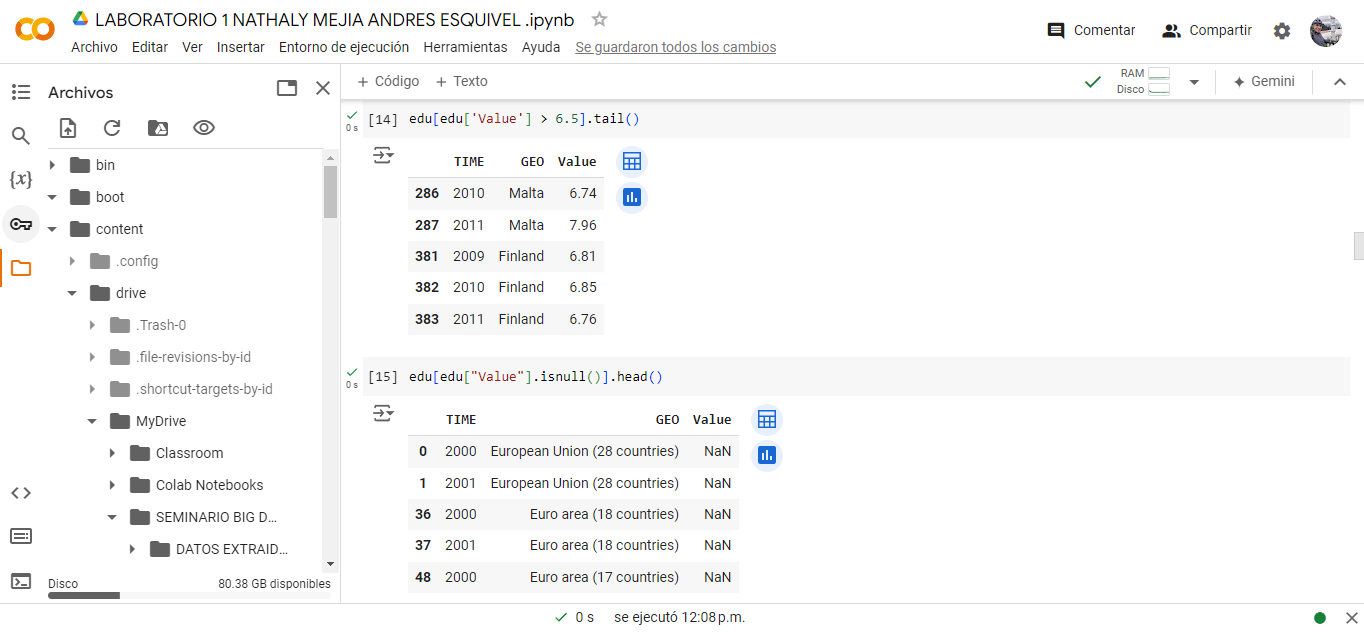
1.



2. 

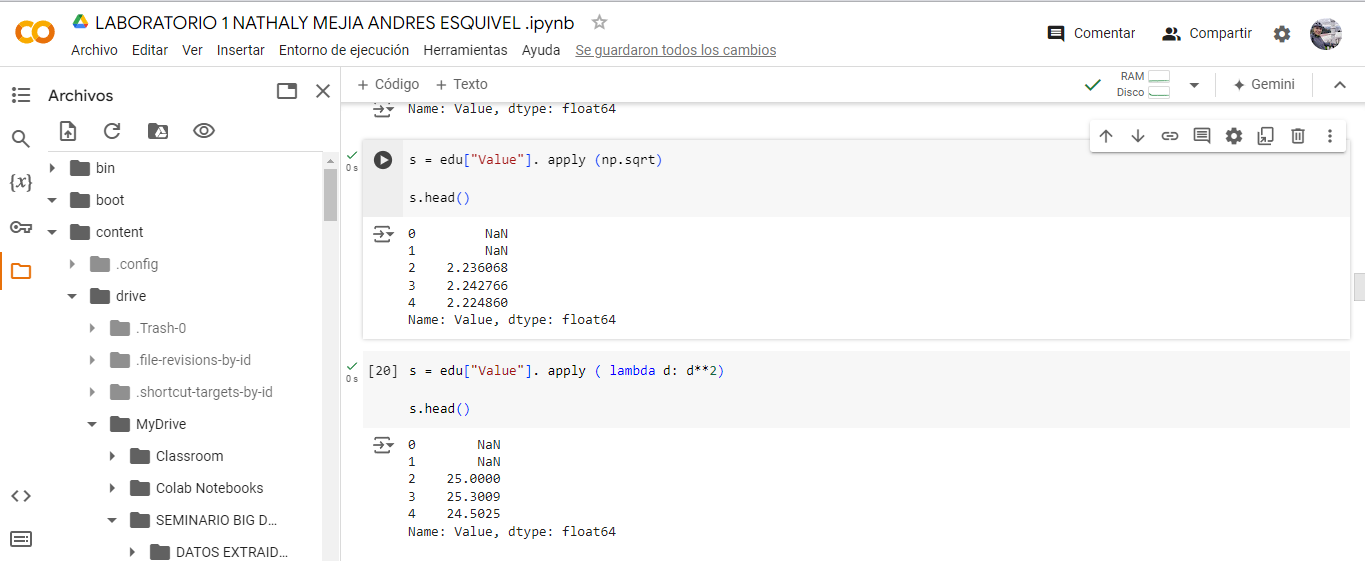
**3. **

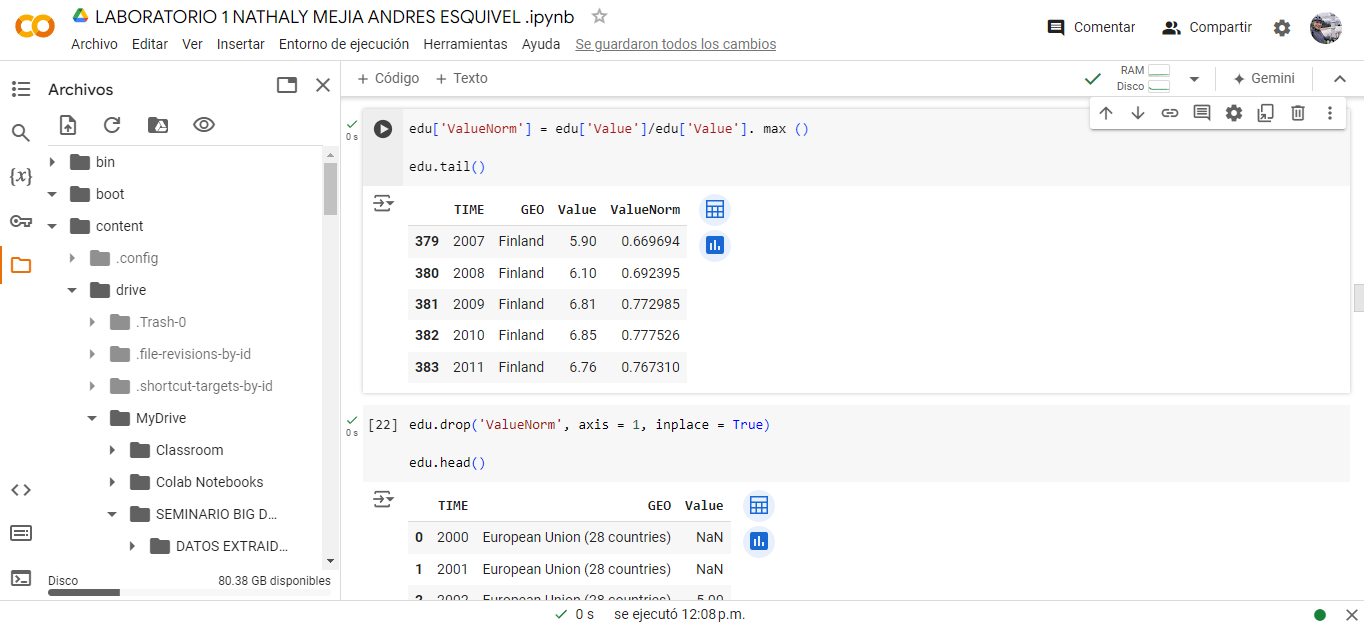
**4. 5. 6. **

**7. **

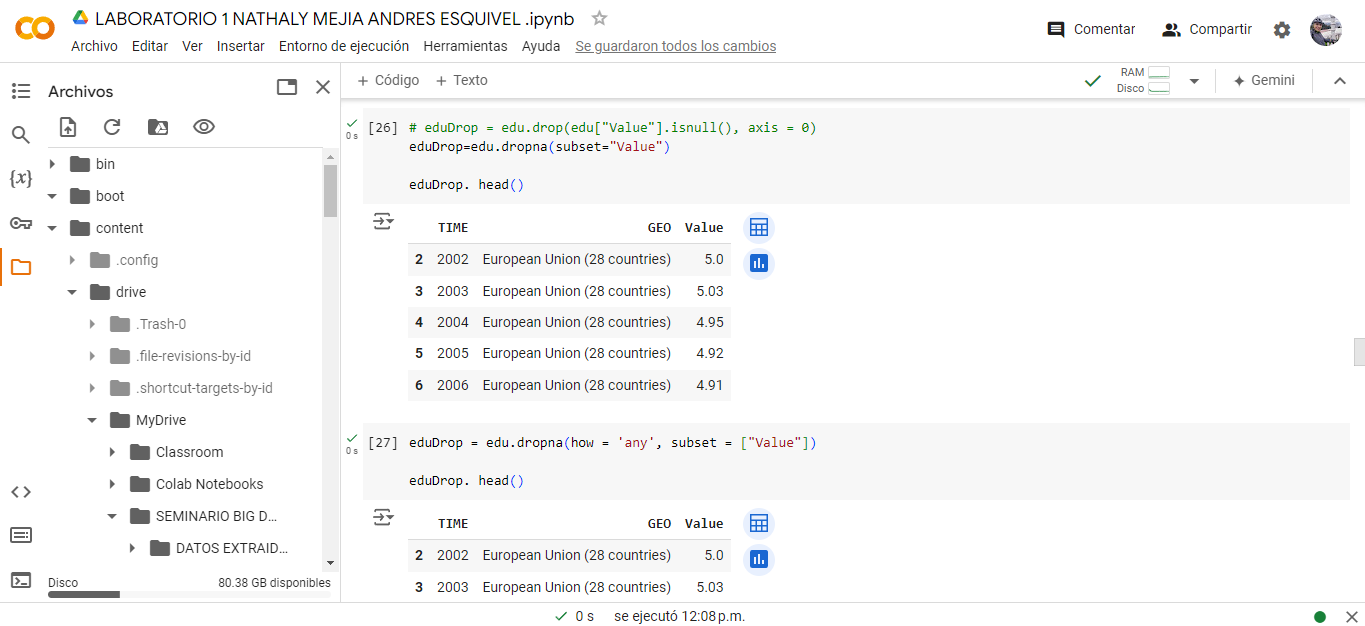
**8. **

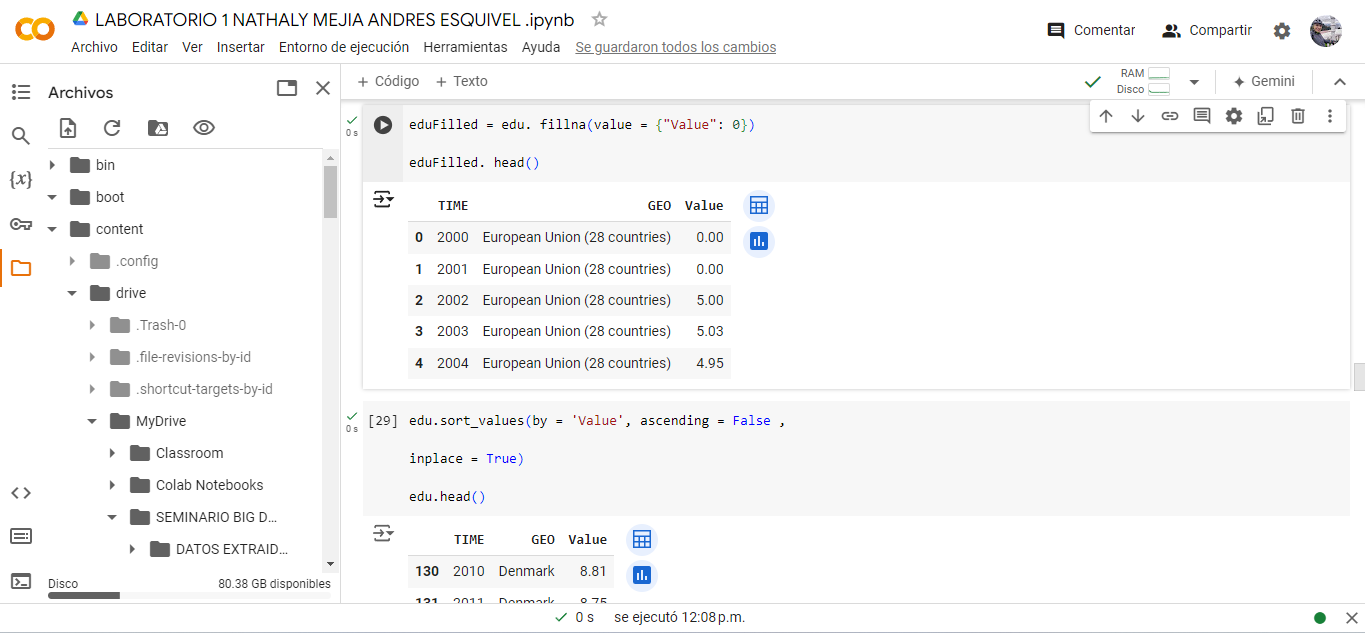
**9.**

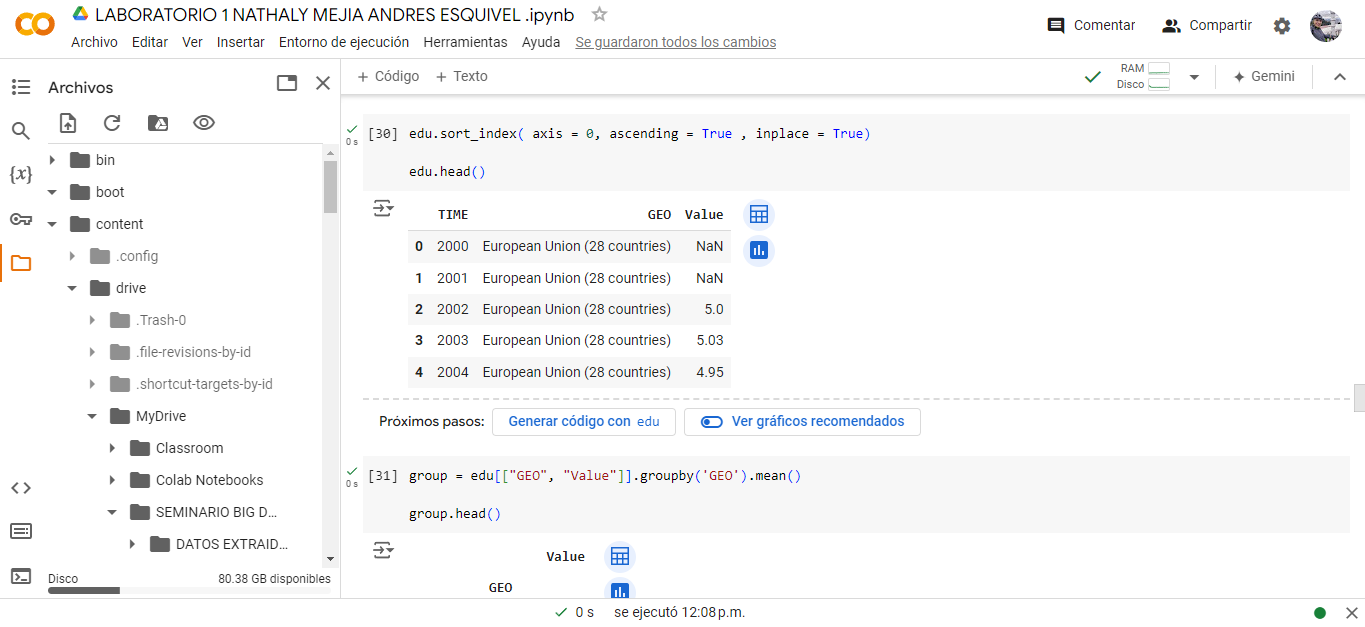
****

**10. **

**11. **

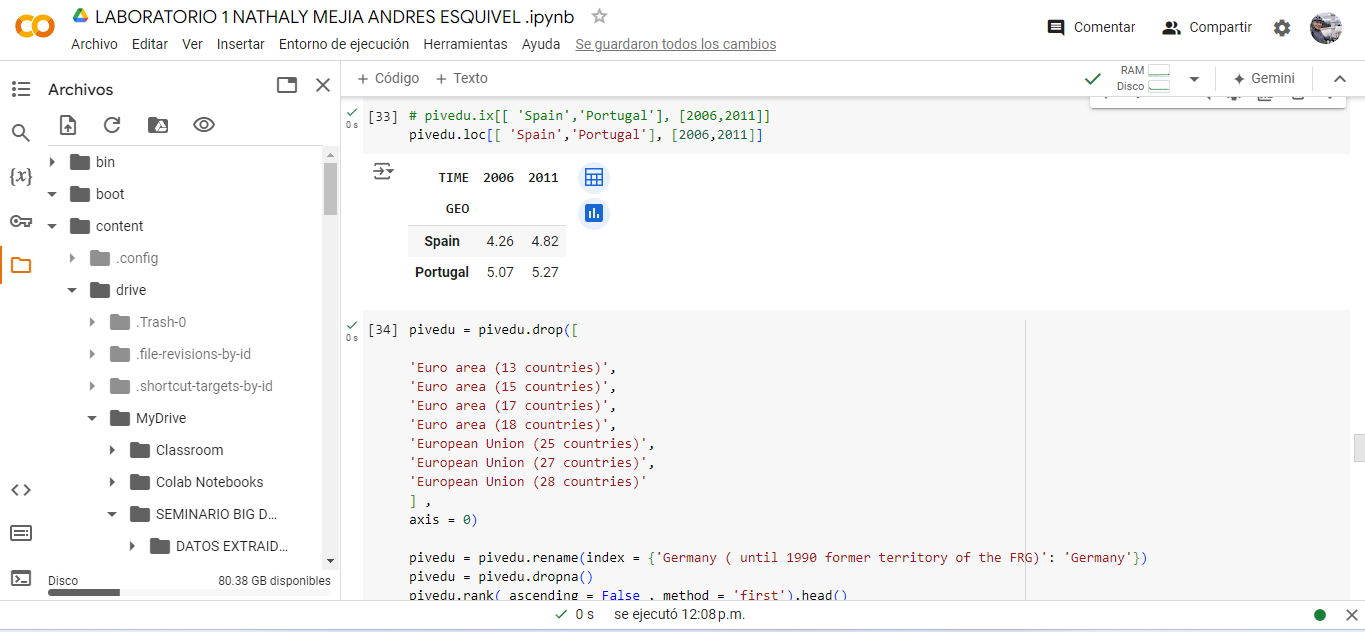
**12. **

**13. **

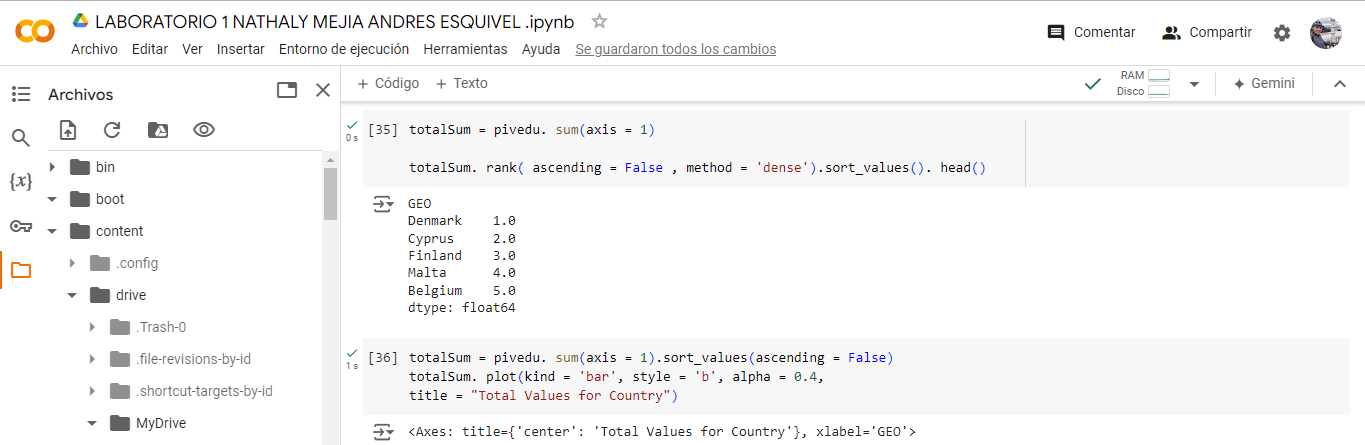
**14. **

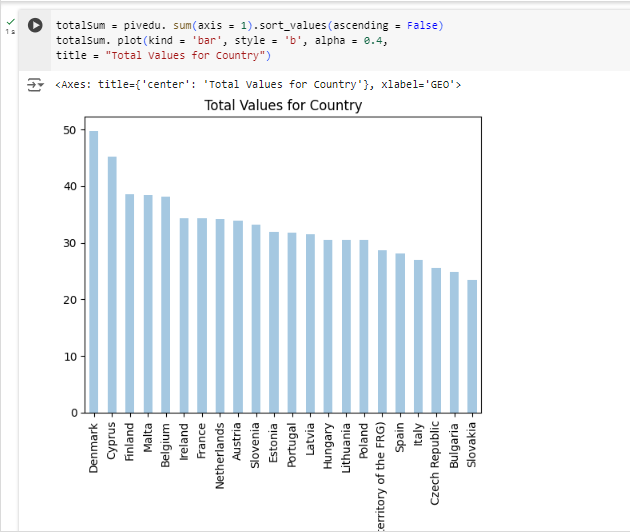
**15. **

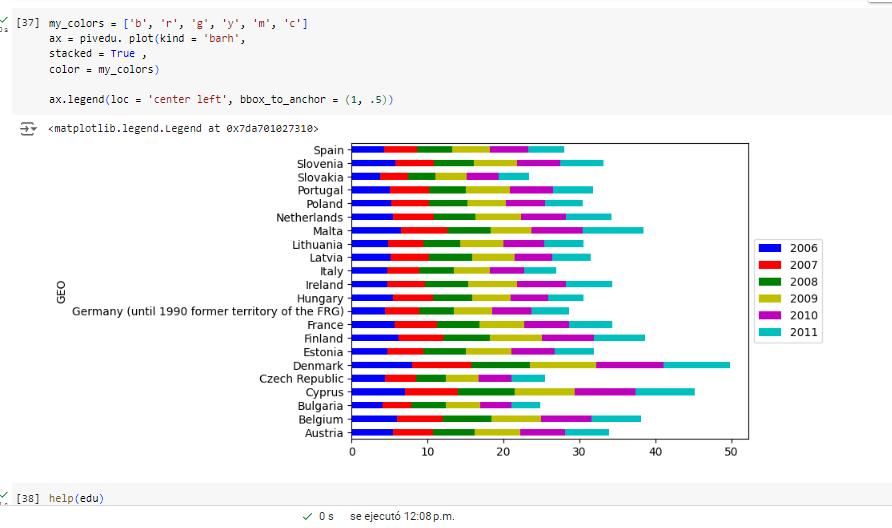
**16.**

****

**17.**

****

**18. **

**19. **

**CONCLUSION:**

Como conclusión podemos decir que la elección de la Biblioteca depende de las Necesidades especificas del proyecto, El tamaño de los datos y la capacidad de procesamiento disponible en el caso de Colab contamos con las 12Gb de Ram que ofrece en su versión Free, Cada una de las bibliotecas cuenta con sus fortalezas y debilidades y ya dependerá según los requisitos de análisis en cuestión, adicionalmente podemos concluir que PySpark y Dask están diseñados para escalar a grandes volúmenes de datos y múltiples nodos, lo cual es ideal para operaciones con grandes conjuntos de datos. Pandas y Polars están limitados a la capacidad de una sola máquina y su potencia de trabajo.