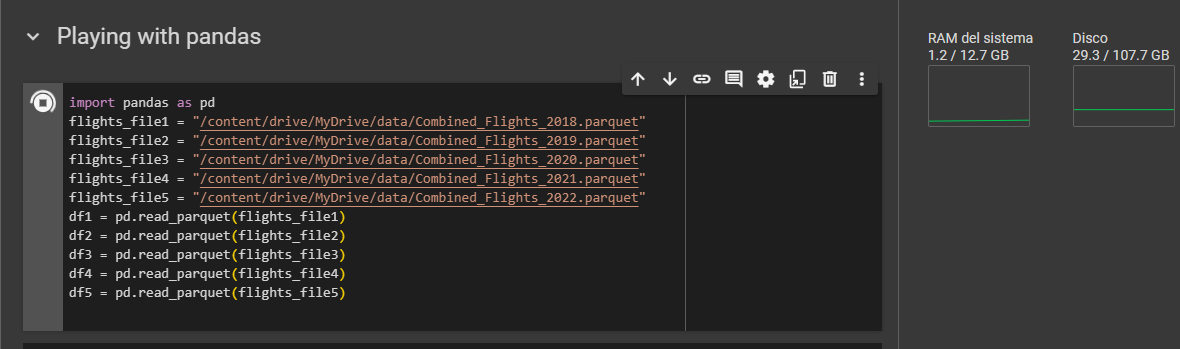
Laboratorio 2

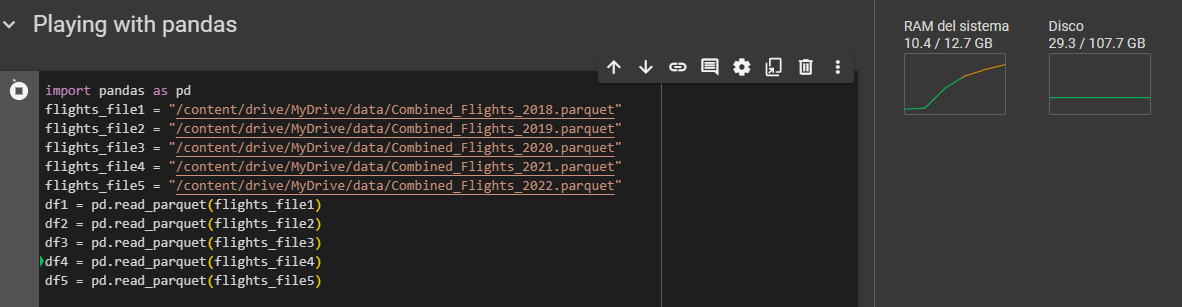
Presentado por: William Steven Moreno Vargas - 95629, Dayana Melissa Martínez Urrego- 109969

# **Pruebas pandas**

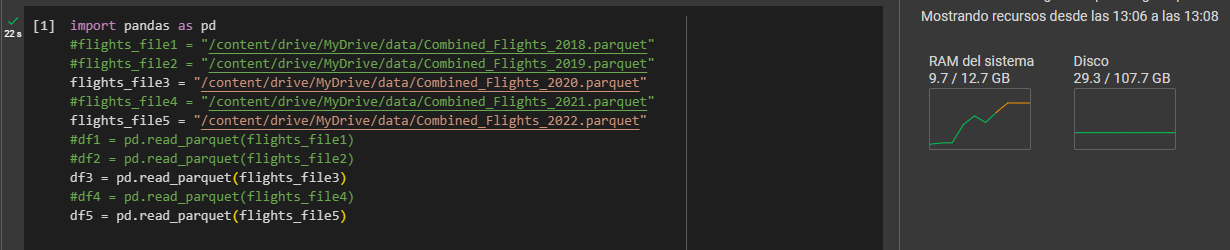
Antes de ejecutar pruebas en pandas



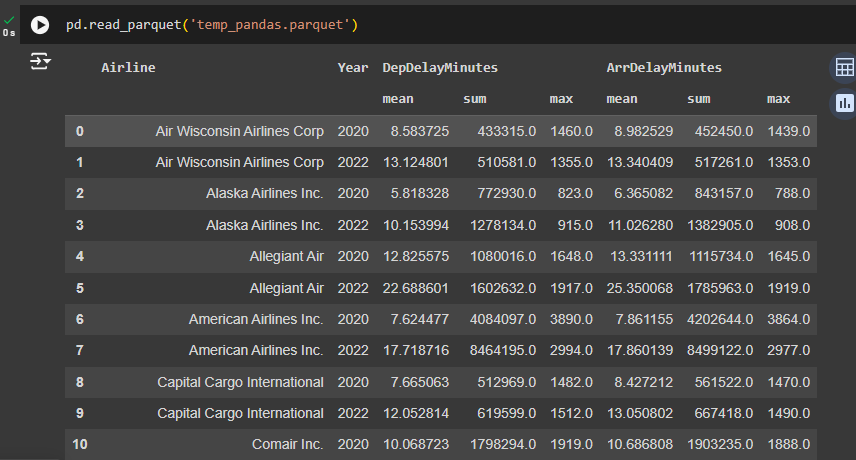
Falla por capacidad de memoria



Se intenta con 2 registros y pasa correctamente

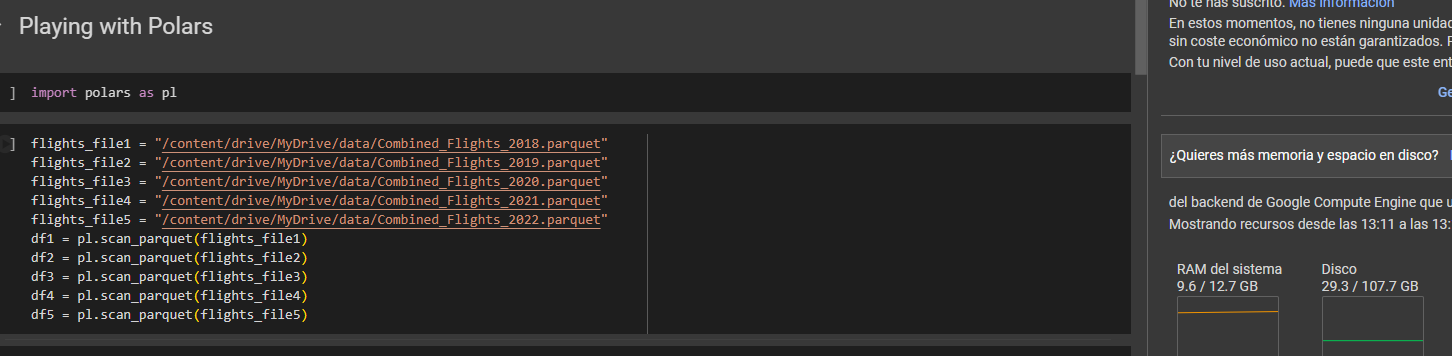


Resultados

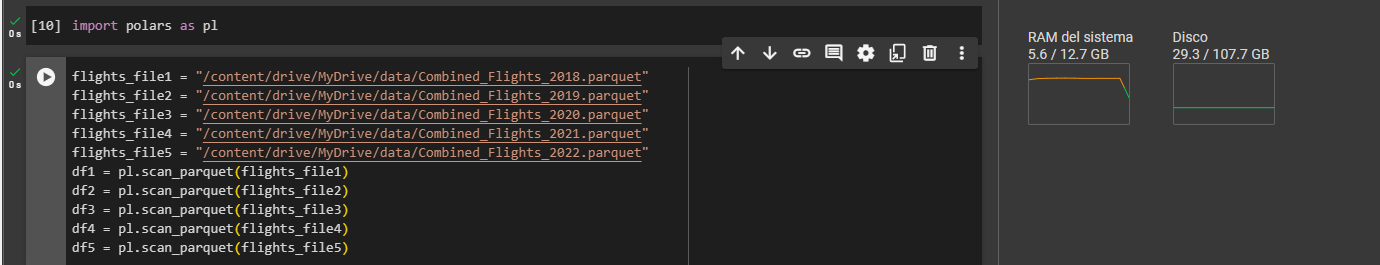


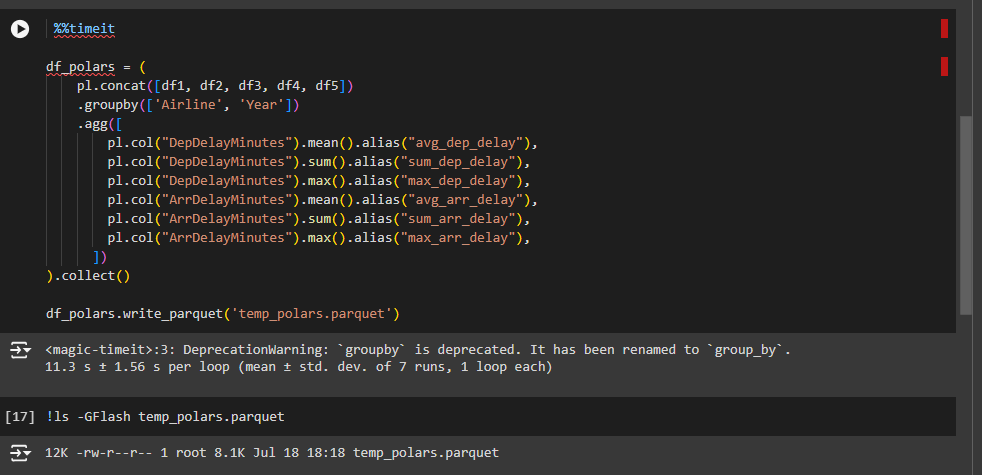
# **Pruebas Polars**

Antes de ejecutar pruebas en Polars



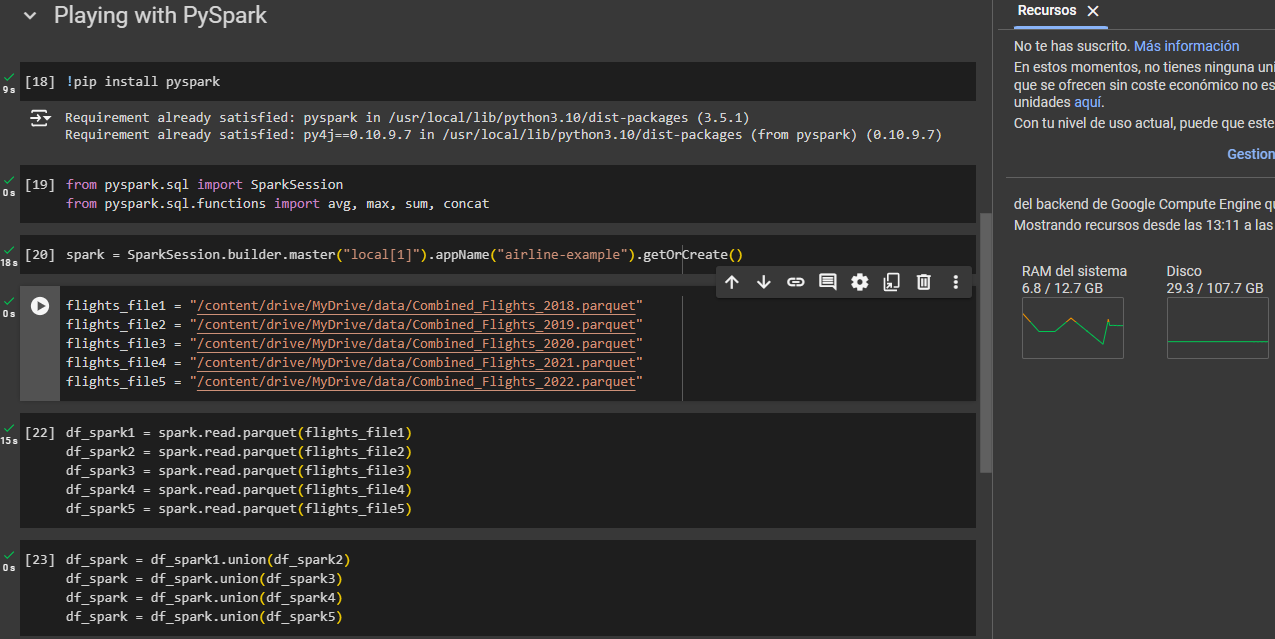
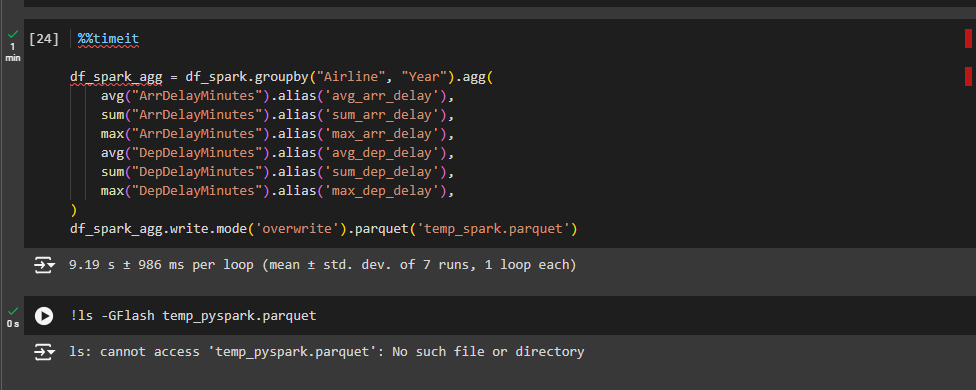
Se intenta con 5 registros y pasa correctamente



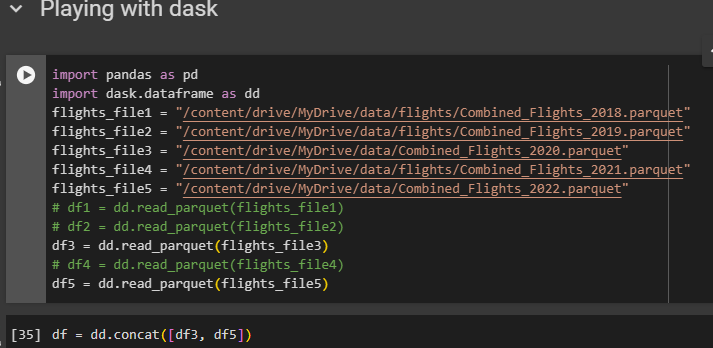


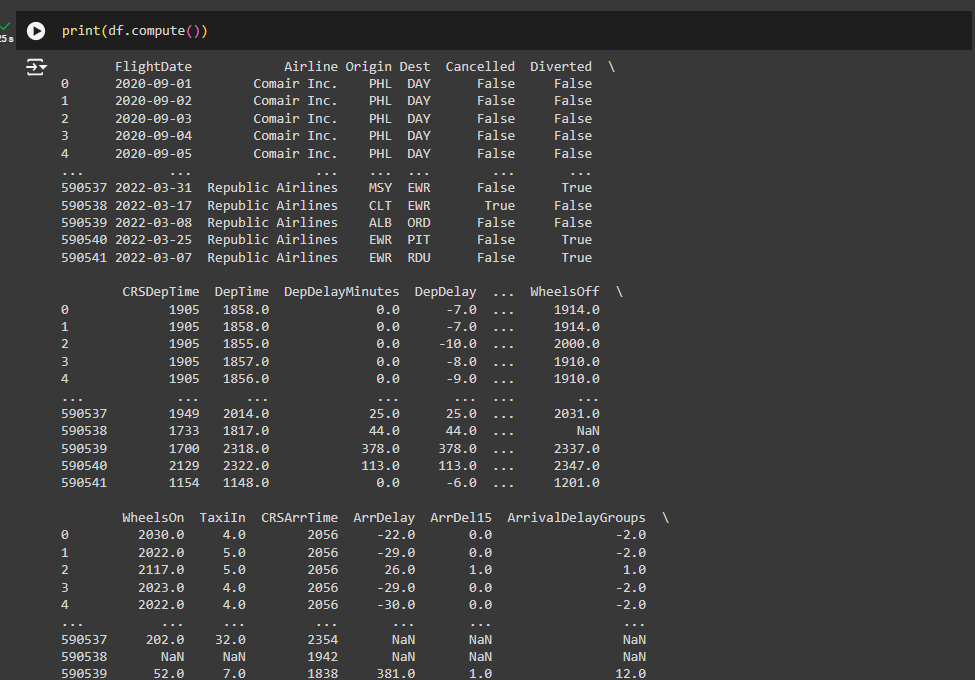
# **Pruebas PySpark**

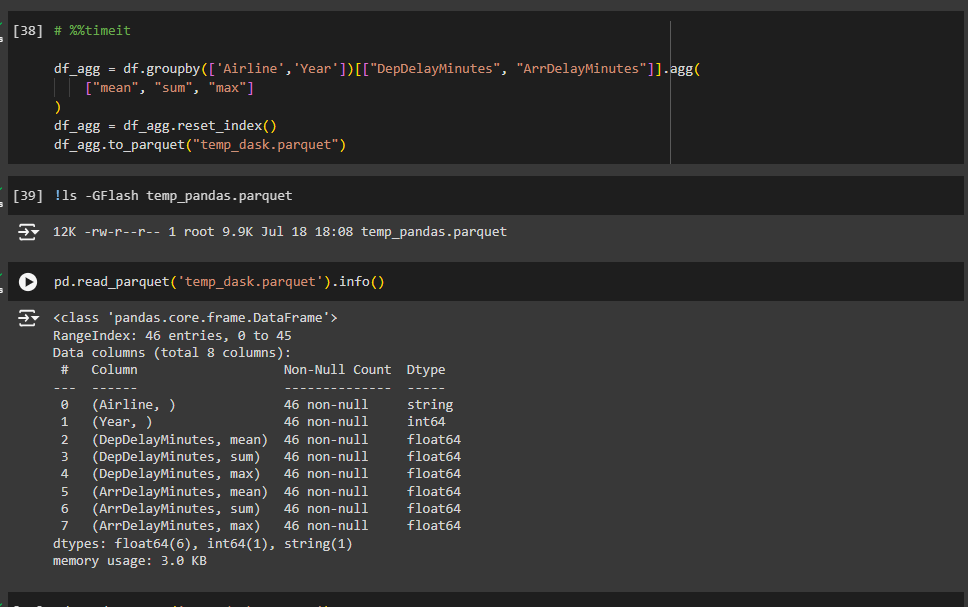
Se intenta con 5 registros y pasa correctamente

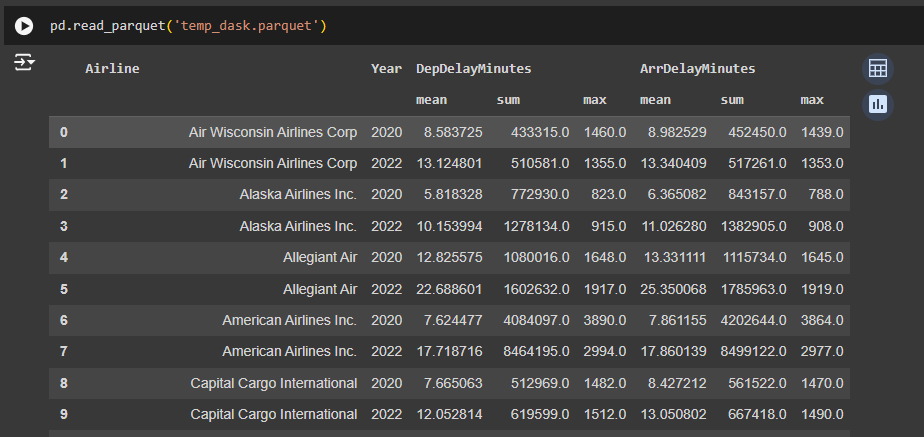
 

# **Pruebas dask**



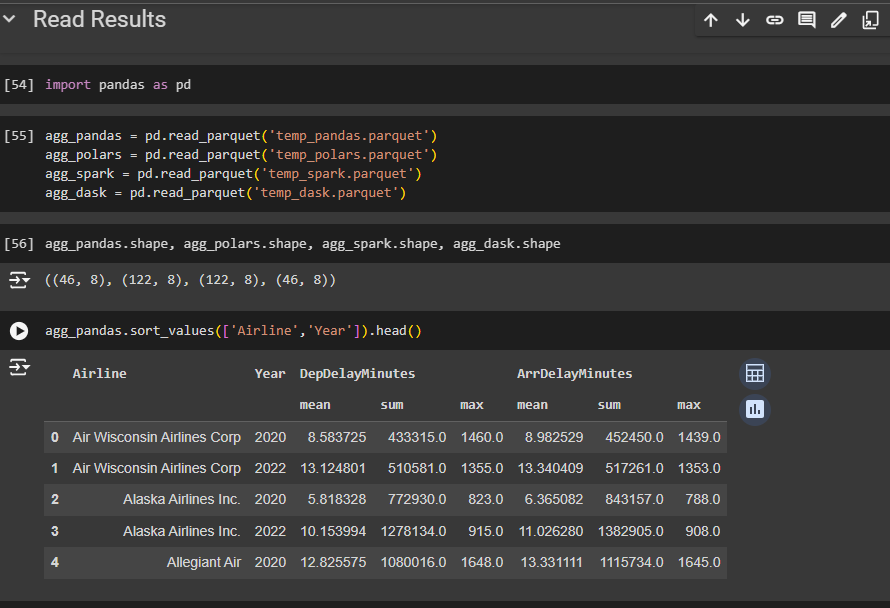




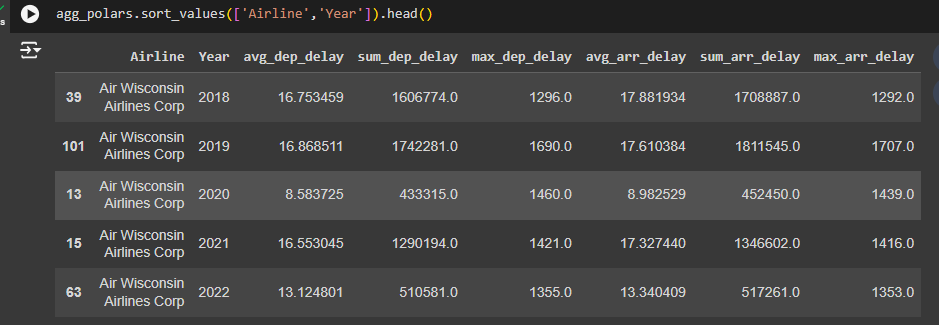


# **Resultados**

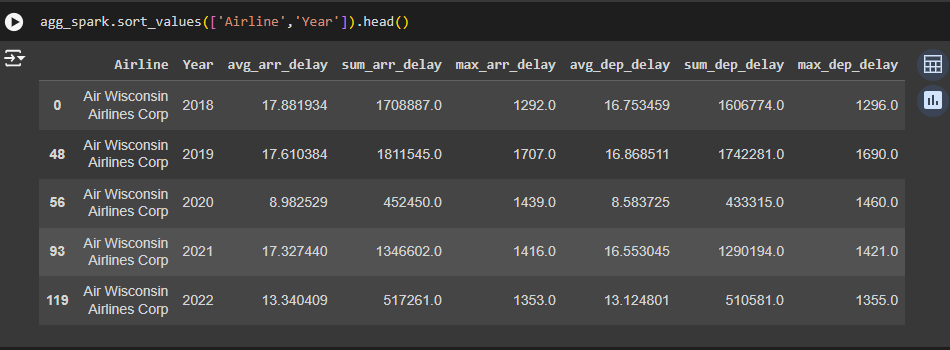
**Pandas**



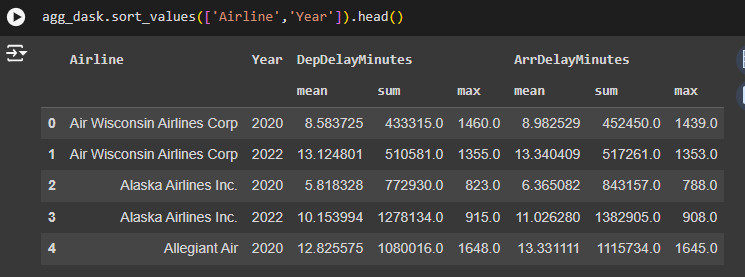
**Polars**



**Spark**



**Dask**



# **Conclusiones**

* La librería panda consume mucha memoria RAM lo cual hace imposible ejecutar los 5 paquetes.
* Polaris sí puede analizar 5 paquetes a la vez, y puede agruparlos, pero demora mas
* PySpark si puede analizar 5 datos a la vez, pero no puede agrupar datos, sólo puede analizarlos en paralelo y tiene un rendimiento similar a Polaris
* Dask si puede analizar 5 paquetes a la vez, pero no puede agruparlos, sólo puede analizarlos en paralelo.