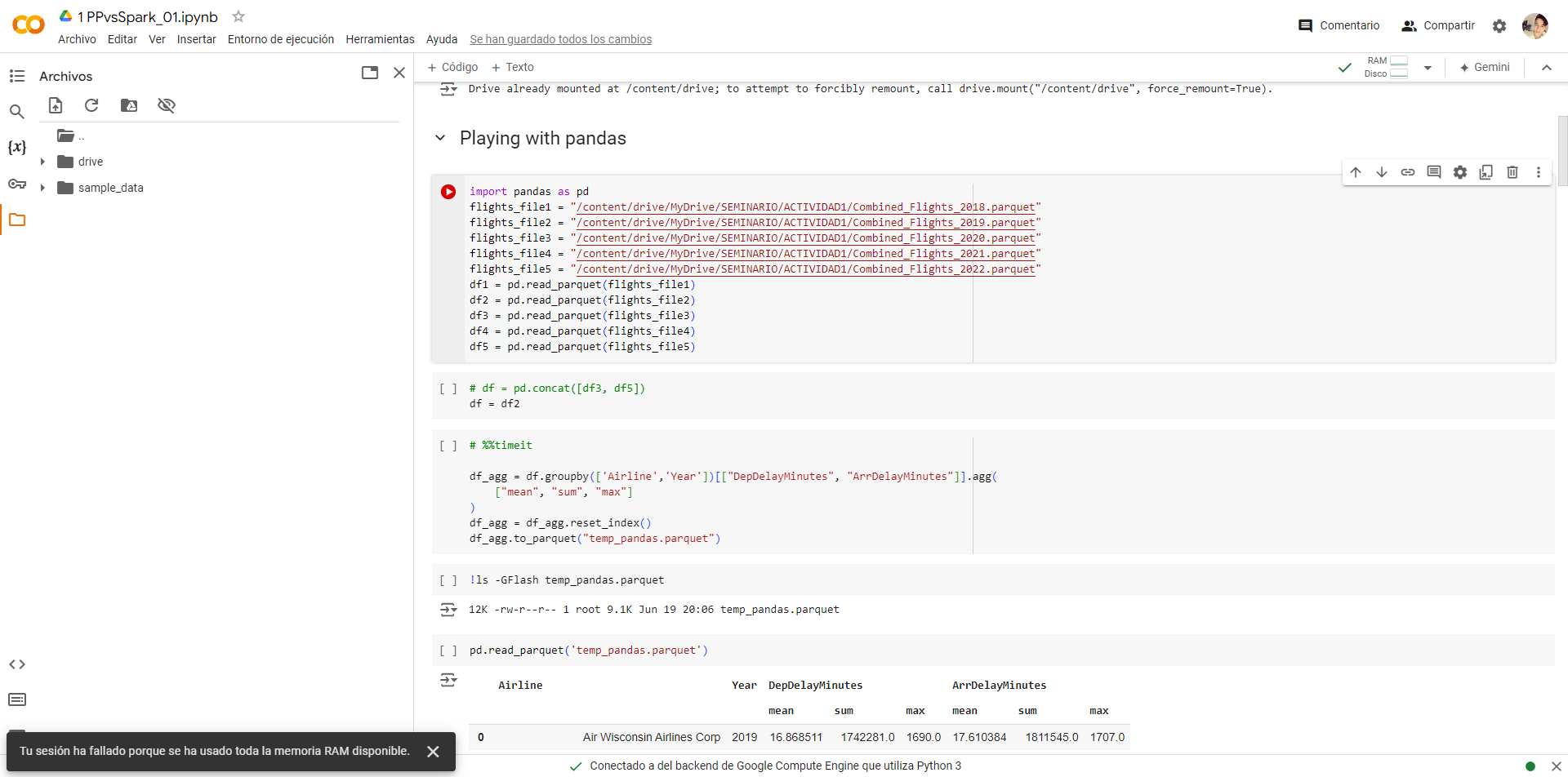
Informe Pandas vs Polars vs Spark vs Dask

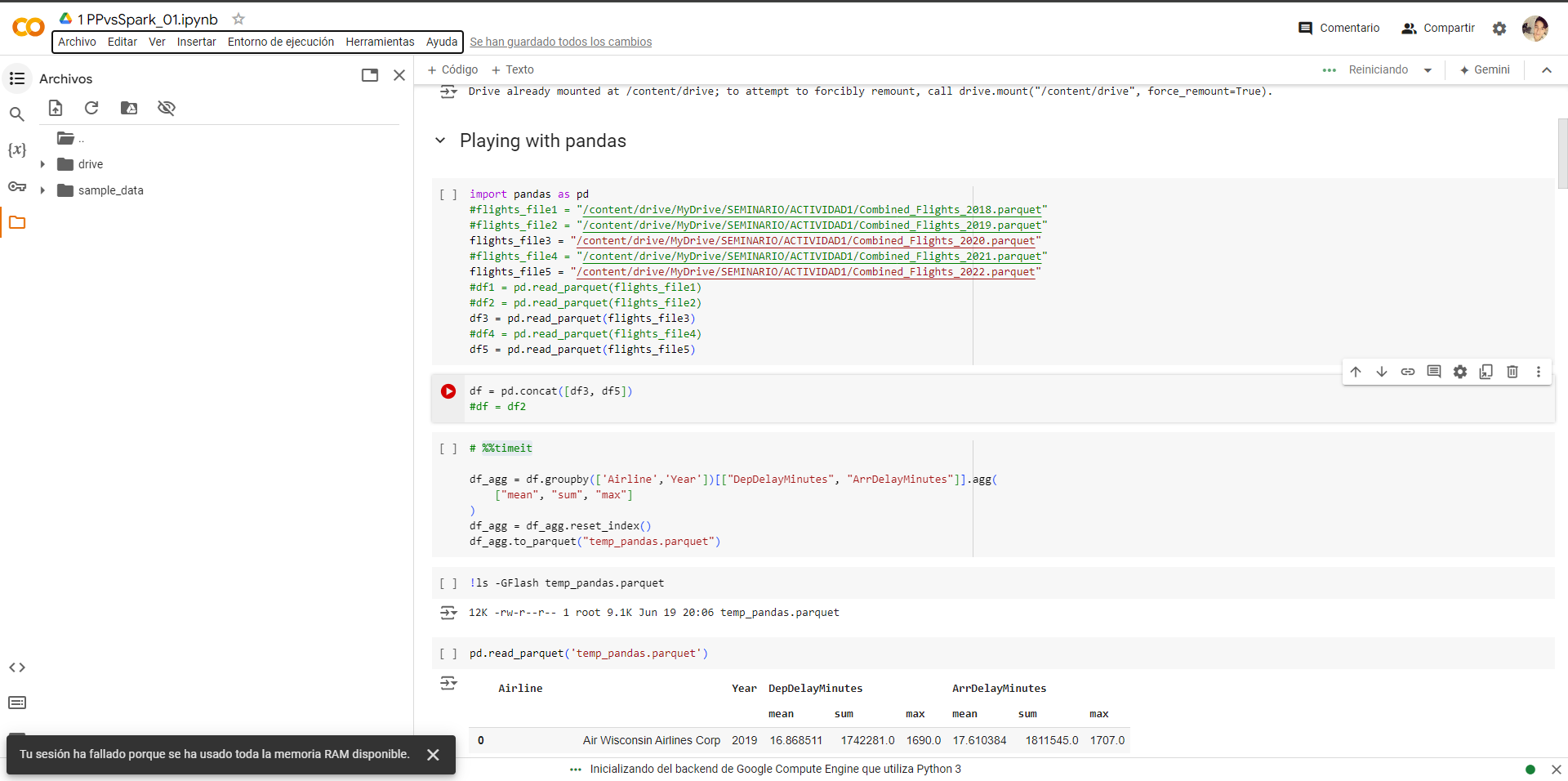
Integrantes:

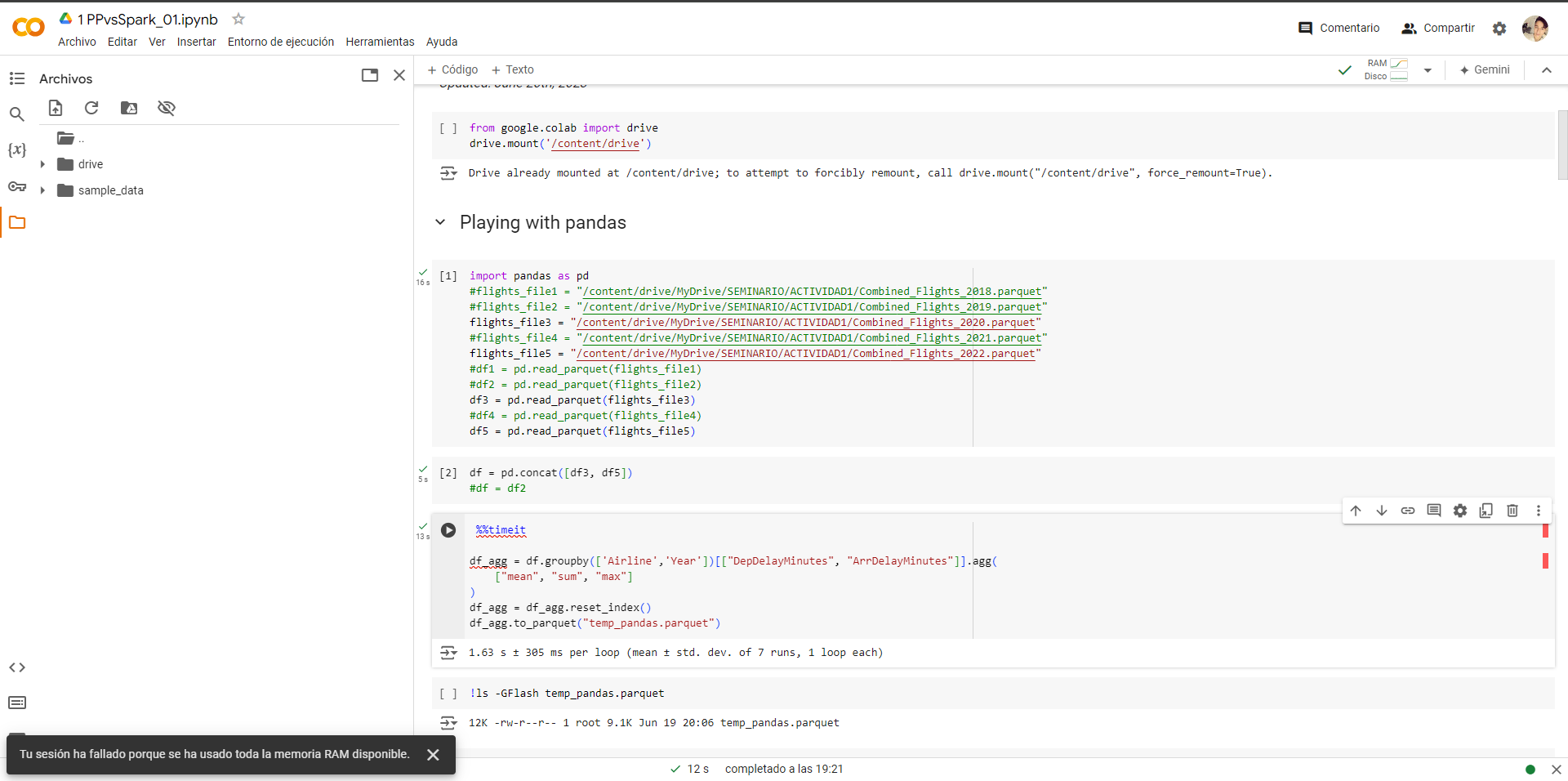
Juan Sebastian Losada Gonzalez 83760

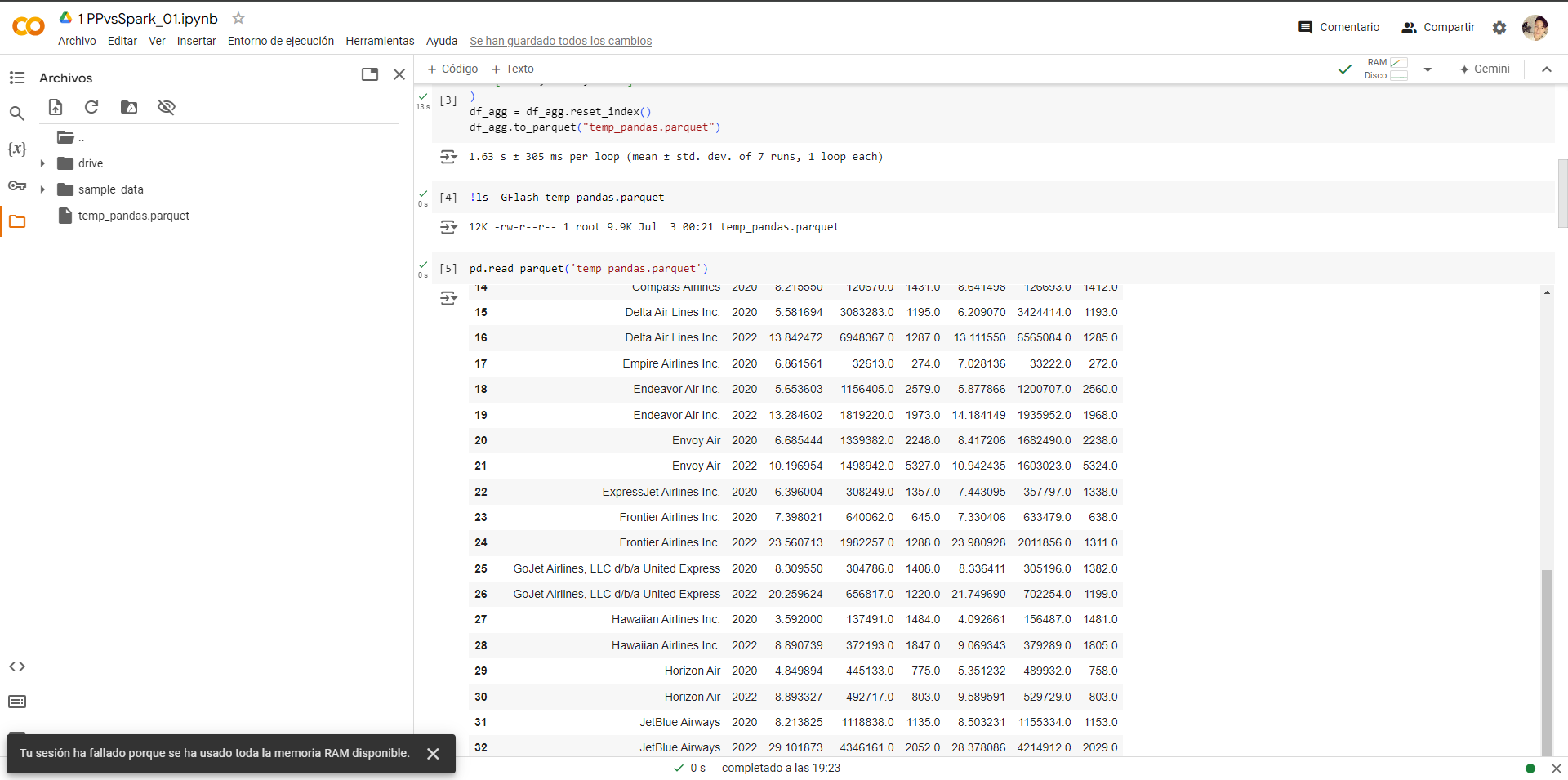
Cristhian David Cuellar Florez 85736

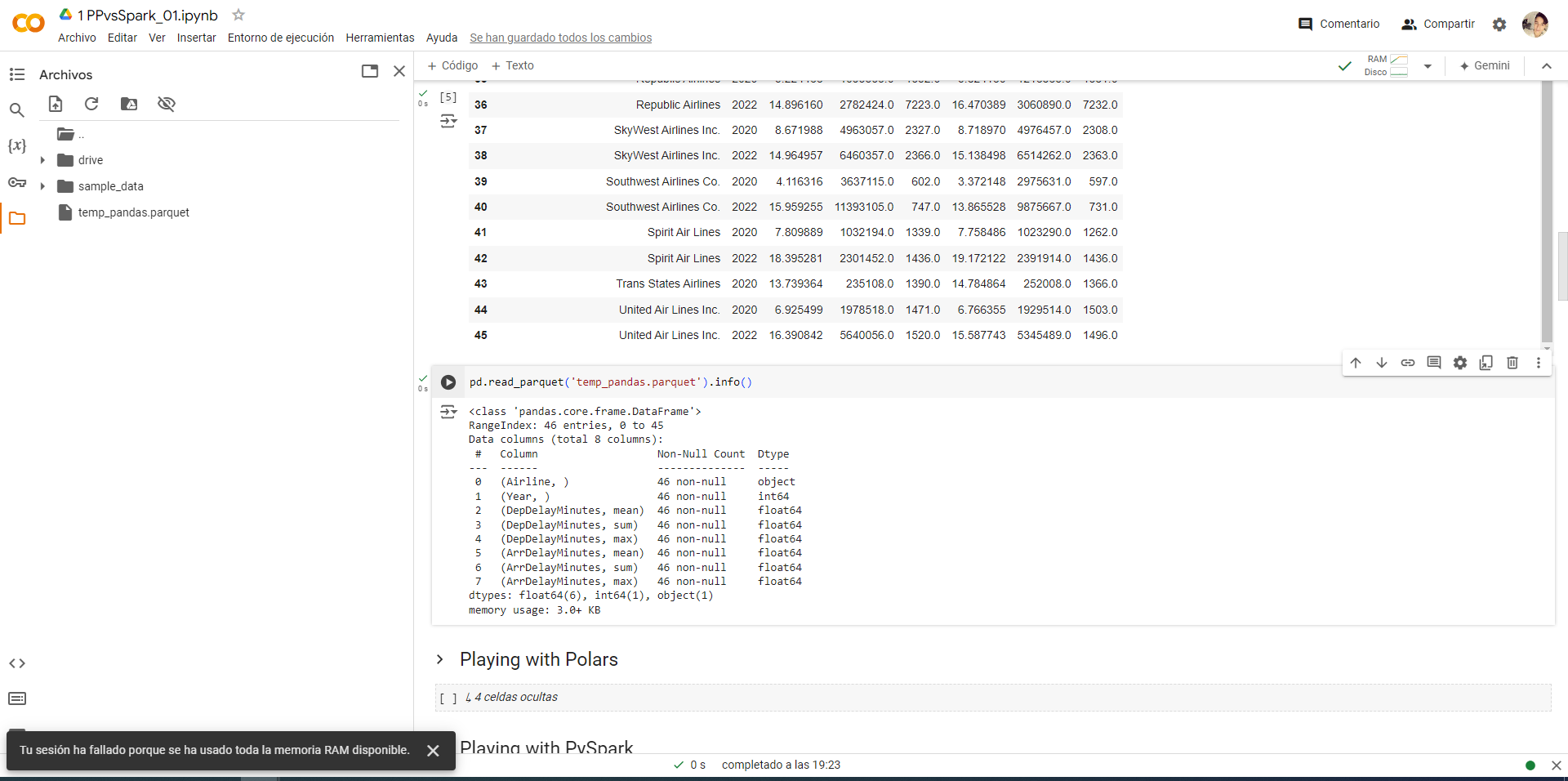
Resultados de Pandas



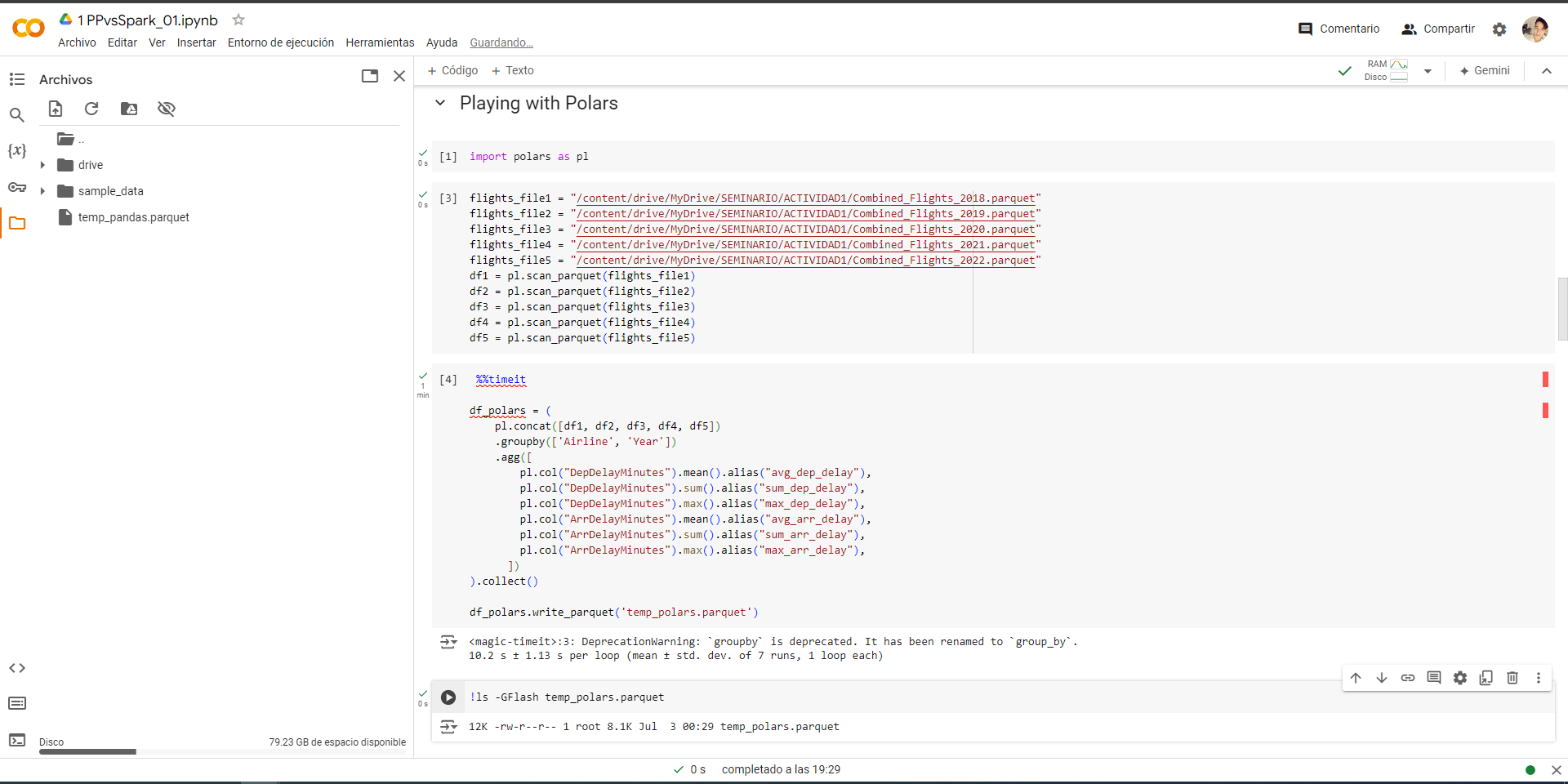




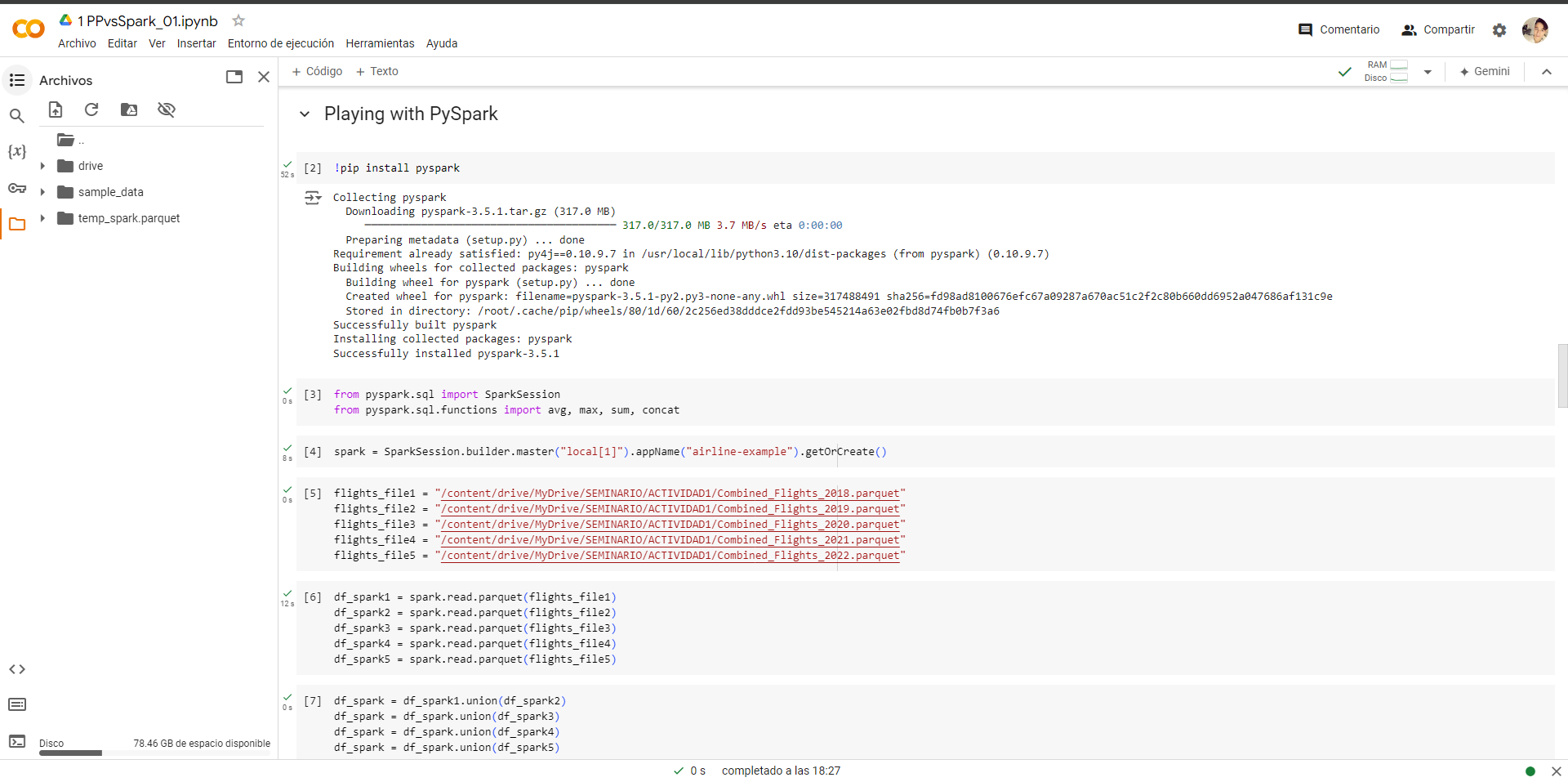


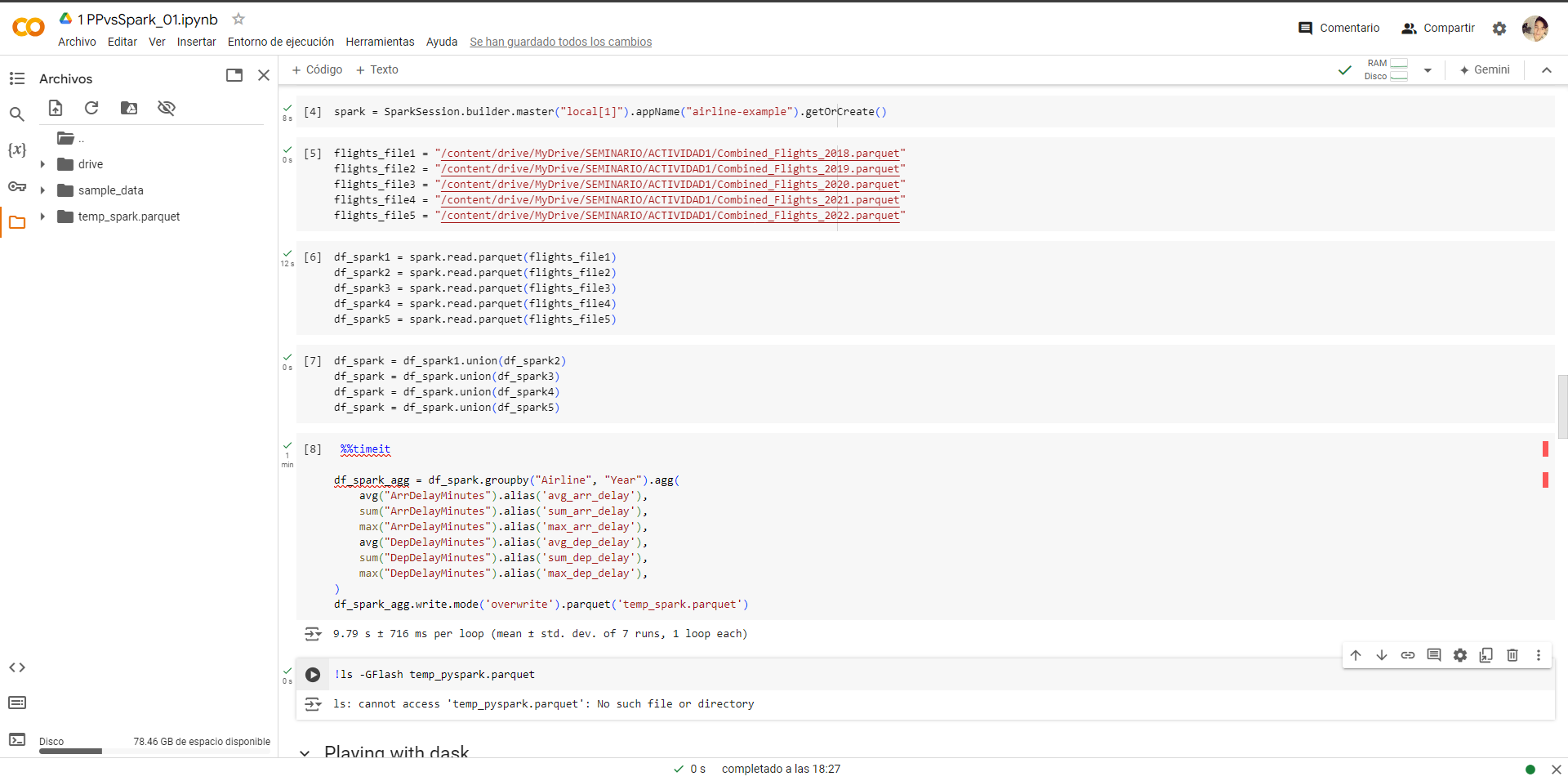


Resultados de Polars

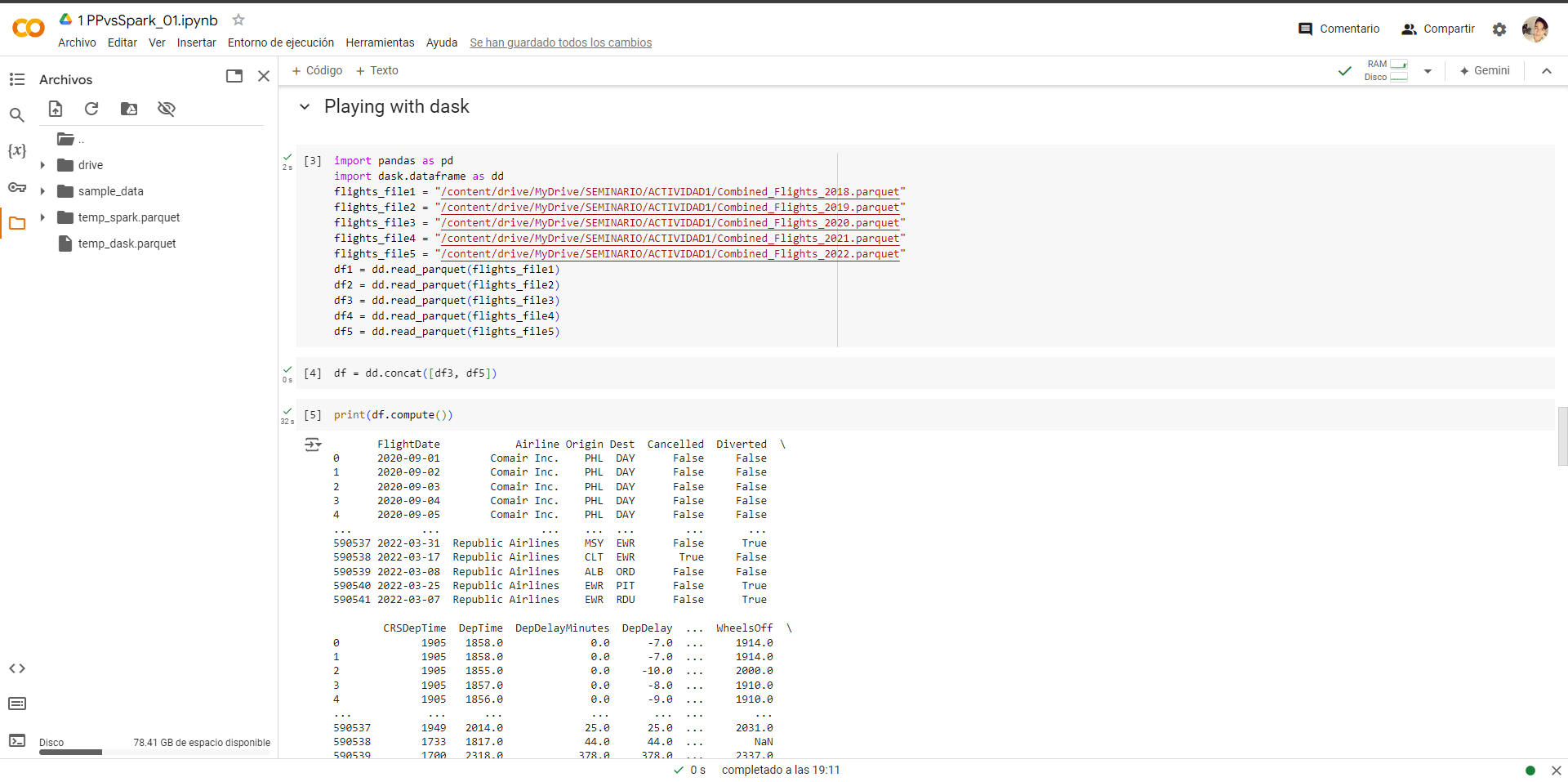


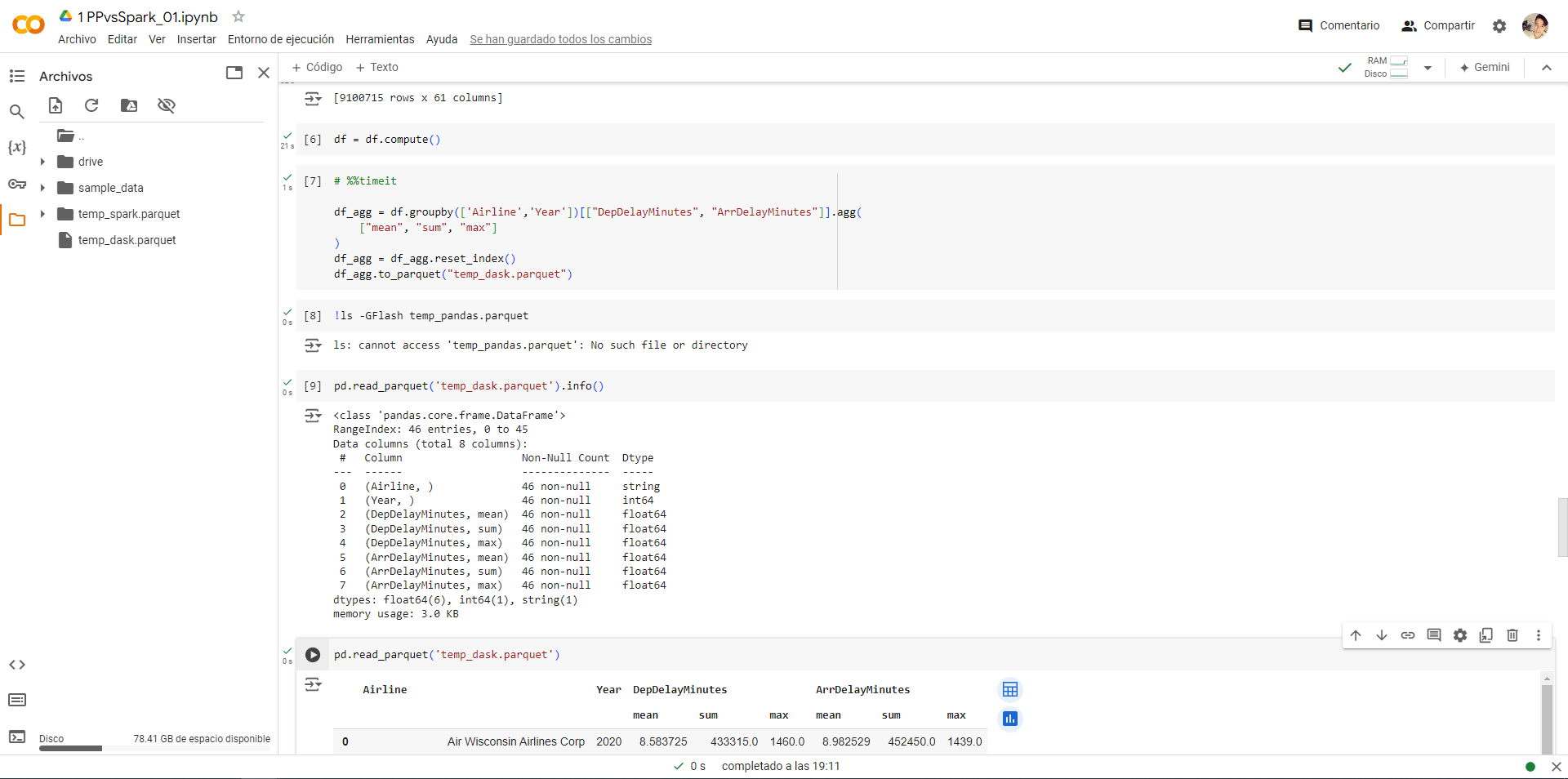
Resultados de Pyspark





Resultados de Dask





Comparación

* Pandas no puede analizar 5 datos a la vez, solo puede procesar 2 y solo permite agrupar de a 2 por lo cual no se demora mucho
* Polaris sí puede analizar 5 datos a la vez, y puede agrupar 5 datos 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 datos a la vez, pero no puede agrupar datos, sólo puede analizarlos en paralelo.