**Ejercicio 1**

**Aviación Civil**

La Administración Nacional de Aviación Civil necesita una serie de informes para elevar al ministerio de transporte acerca de los aterrizajes y despegues en todo el territorio Argentino, como puede ser: cuales aviones son los que más volaron, cuántos pasajeros volaron, ciudades de partidas y aterrizajes entre fechas determinadas, etc. Usted como data engineer deberá realizar un pipeline con esta información, automatizarlo y realizar los análisis de datos solicitados que permita responder las preguntas de negocio, y hacer sus recomendaciones con respecto al estado actual.

Listado de vuelos realizados:

<https://datos.transporte.gob.ar/dataset/aterrizajes-y-despegues-procesados-por-la-administracion-nacional-de-aviacion-civil-anac/archivo/1b4f569a-cab0-4560-993c-5bf96c3e7cf0>

Listado de detalles de aeropuertos de Argentina: <https://datos.transporte.gob.ar/dataset/lista-aeropuertos>

**TAREAS**

1. Hacer ingest de los siguientes files relacionados con transporte aéreo de Argentina:

2021:

<https://edvaibucket.blob.core.windows.net/data-engineer-edvai/2021-informe-ministerio.csv?sp=r&st=2023-11-06T12:59:46Z&se=2025-11-06T20:59:46Z&sv=2022-11-02&sr=b&sig=%2BSs5xIW3qcwmRh5TTmheIY9ZBa9BJC8XQDcI%2FPLRe9Y%3D>

2022:

<https://edvaibucket.blob.core.windows.net/data-engineer-edvai/202206-informe-ministerio.csv?sp=r&st=2023-11-06T12:52:39Z&se=2025-11-06T20:52:39Z&sv=2022-11-02&sr=c&sig=J4Ddi2c7Ep23OhQLPisbYaerlH472iigPwc1%2FkG80EM%3D>

Aeropuertos\_detalles:

<https://edvaibucket.blob.core.windows.net/data-engineer-edvai/aeropuertos_detalle.csv?sp=r&st=2023-11-06T12:52:39Z&se=2025-11-06T20:52:39Z&sv=2022-11-02&sr=c&sig=J4Ddi2c7Ep23OhQLPisbYaerlH472iigPwc1%2FkG80EM%3D>

**Solución:**

<https://github.com/luxfo/edvai-data-engineering/blob/main/Ejercicios/TPFinal/Ejercicio1/tp1_ingest.sh>

2. Crear 2 tablas en el datawarehouse, una para los vuelos realizados en 2021 y 2022 (2021-informe-ministerio.csv y 202206-informe-ministerio) y otra tabla para el detalle de los aeropuertos (aeropuertos\_detalle.csv)

**Solución:**

<https://github.com/luxfo/edvai-data-engineering/blob/main/Ejercicios/TPFinal/Ejercicio1/hive_scripts.sh>

3. Realizar un proceso automático orquestado por airflow que ingeste los archivos previamente mencionados entre las fechas 01/01/2021 y 30/06/2022 en las dos columnas creadas.

Los archivos 202206-informe-ministerio.csv y 202206-informe-ministerio.csv → en la tabla aeropuerto\_tabla

El archivo aeropuertos\_detalle.csv → en la tabla aeropuerto\_detalles\_tabla

**Solución:**

<https://github.com/luxfo/edvai-data-engineering/blob/main/Ejercicios/TPFinal/Ejercicio1/dag_tp1.py>

4. Realizar las siguiente transformaciones en los pipelines de datos:

● Eliminar la columna inhab ya que no se utilizará para el análisis

● Eliminar la columna fir ya que no se utilizará para el análisis

● Eliminar la columna “calidad del dato” ya que no se utilizará para el análisis

● Filtrar los vuelos internacionales ya que solamente se analizarán los vuelos domésticos

● En el campo pasajeros si se encuentran campos en Null convertirlos en 0 (cero)

● En el campo distancia\_ref si se encuentran campos en Null convertirlos en 0 (cero)

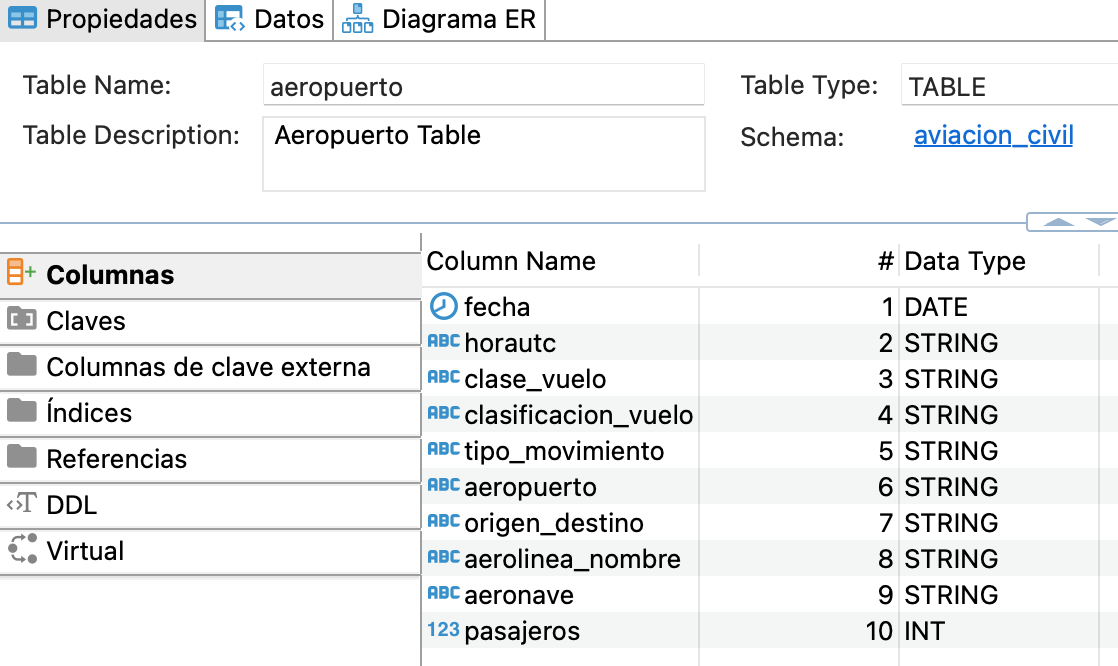
**Solución:**

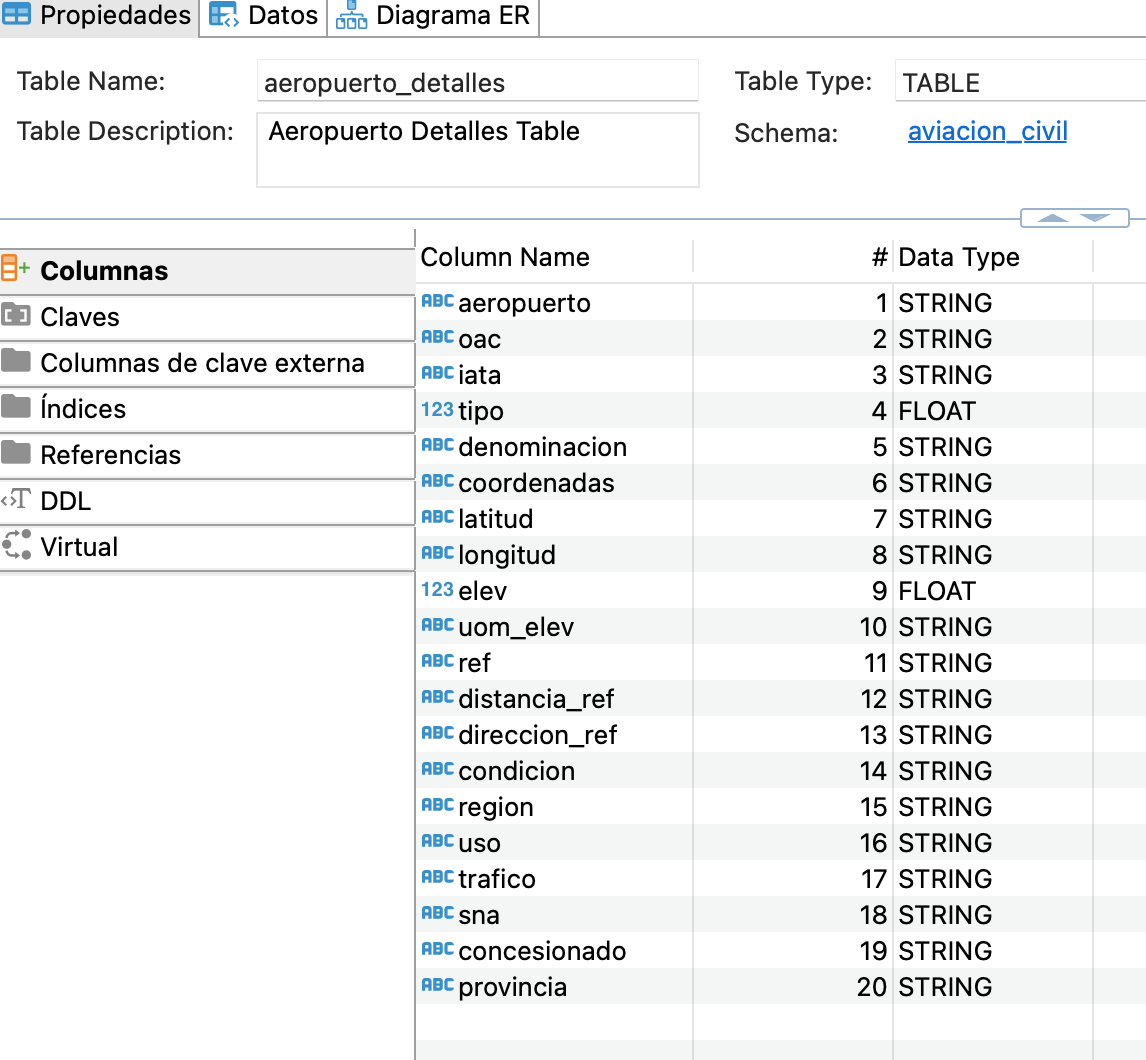
<https://github.com/luxfo/edvai-data-engineering/blob/main/Ejercicios/TPFinal/Ejercicio1/tp1_transform_1.py>

<https://github.com/luxfo/edvai-data-engineering/blob/main/Ejercicios/TPFinal/Ejercicio1/tp1_transform_2.py>

5. Mostrar mediante una impresión de pantalla, que los tipos de campos de las tablas sean los solicitados en el datawarehouse (ej: fecha date, aeronave string, pasajeros integer, etc.)

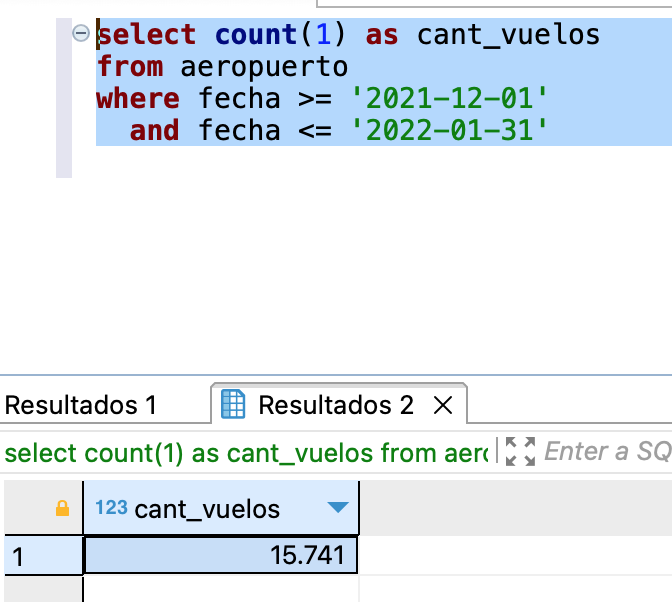
**Solución:**

****



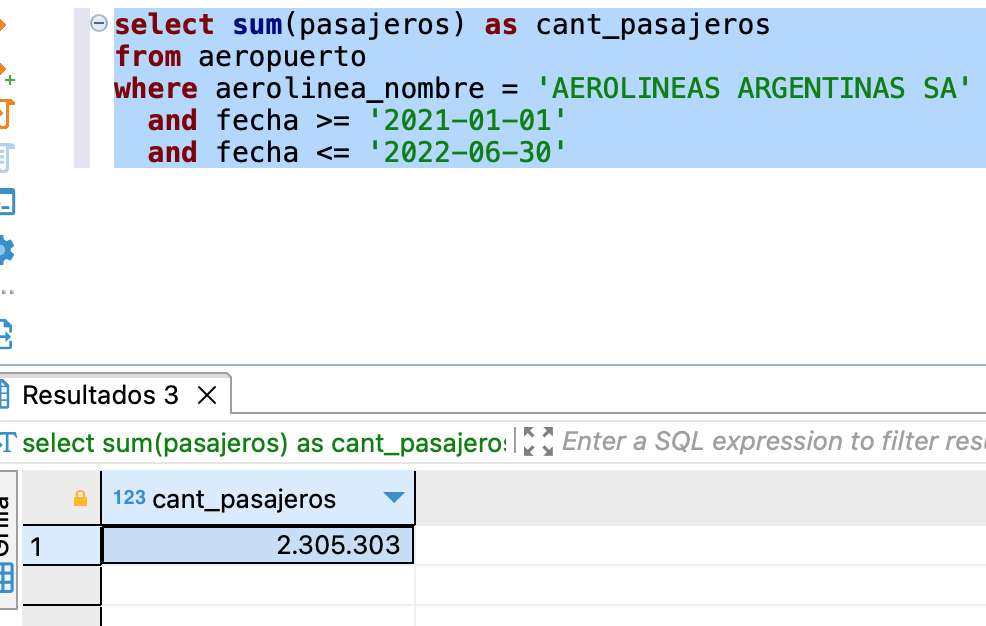
6. Determinar la cantidad de vuelos entre las fechas 01/12/2021 y 31/01/2022. Mostrar consulta y Resultado de la query

**Solución:**

****

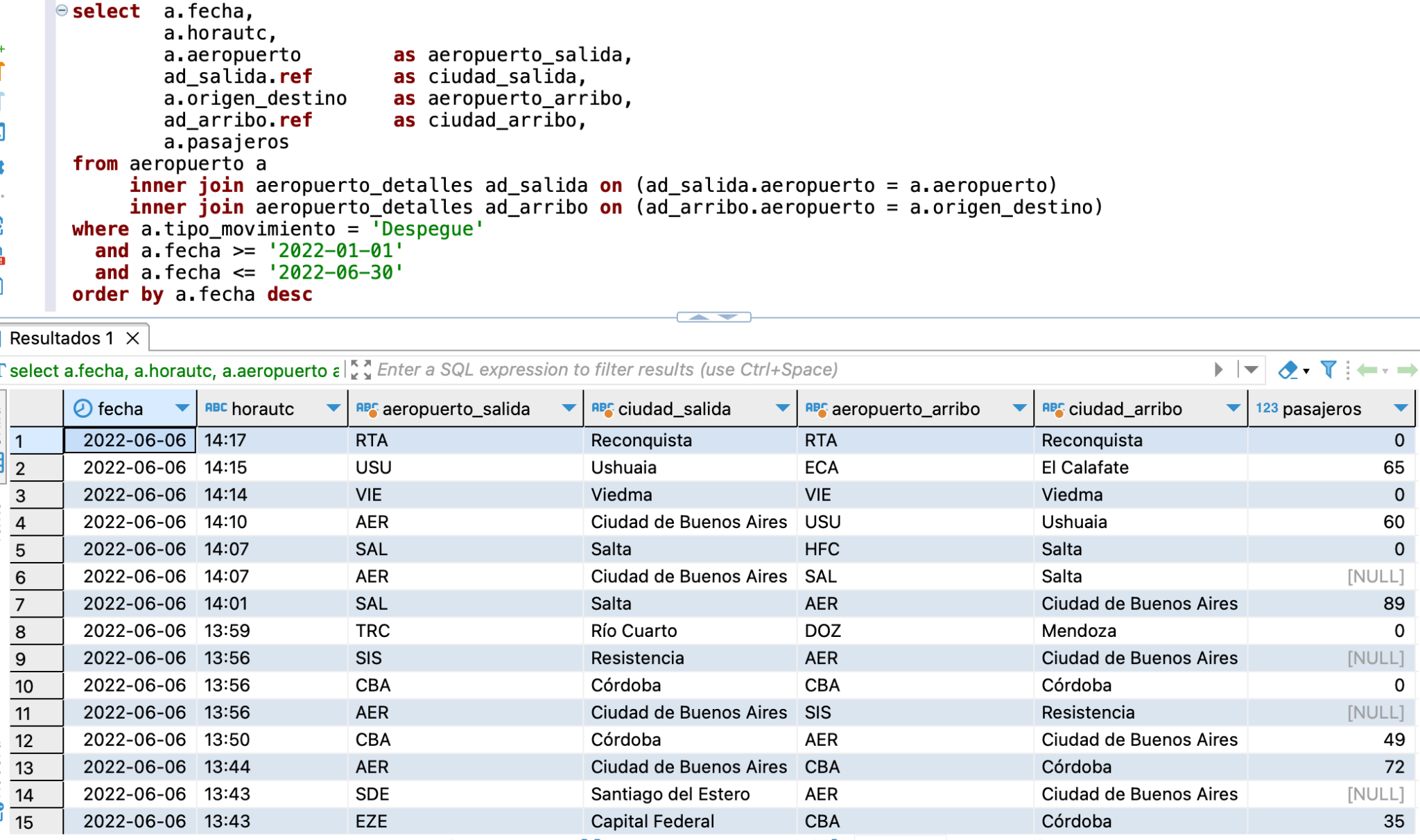
7. Cantidad de pasajeros que viajaron en Aerolíneas Argentinas entre el 01/01/2021 y 30/06/2022. Mostrar consulta y Resultado de la query

**Solución:**

****

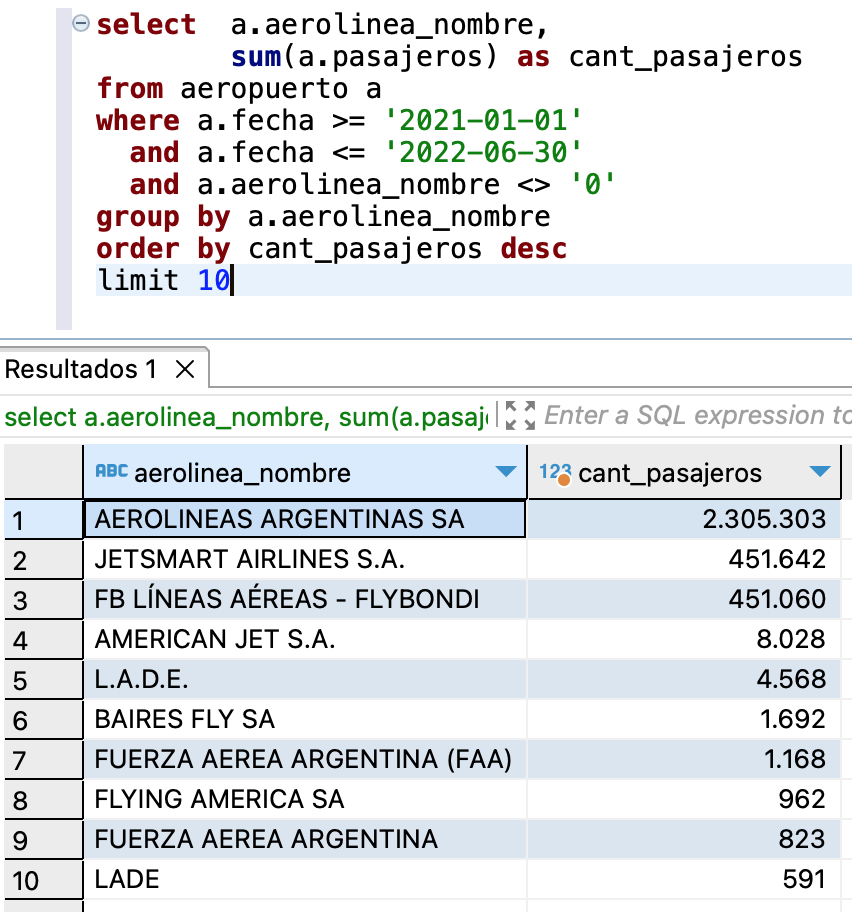
8. Mostrar fecha, hora, código aeropuerto salida, ciudad de salida, código de aeropuerto de arribo, ciudad de arribo, y cantidad de pasajeros de cada vuelo, entre el 01/01/2022 y el 30/06/2022 ordenados por fecha de manera descendiente. Mostrar consulta y Resultado de la query

**Solución:**

****

9. Cuales son las 10 aerolíneas que más pasajeros llevaron entre el 01/01/2021 y el 30/06/2022 exceptuando aquellas aerolíneas que no tengan nombre. Mostrar consulta y Visualización

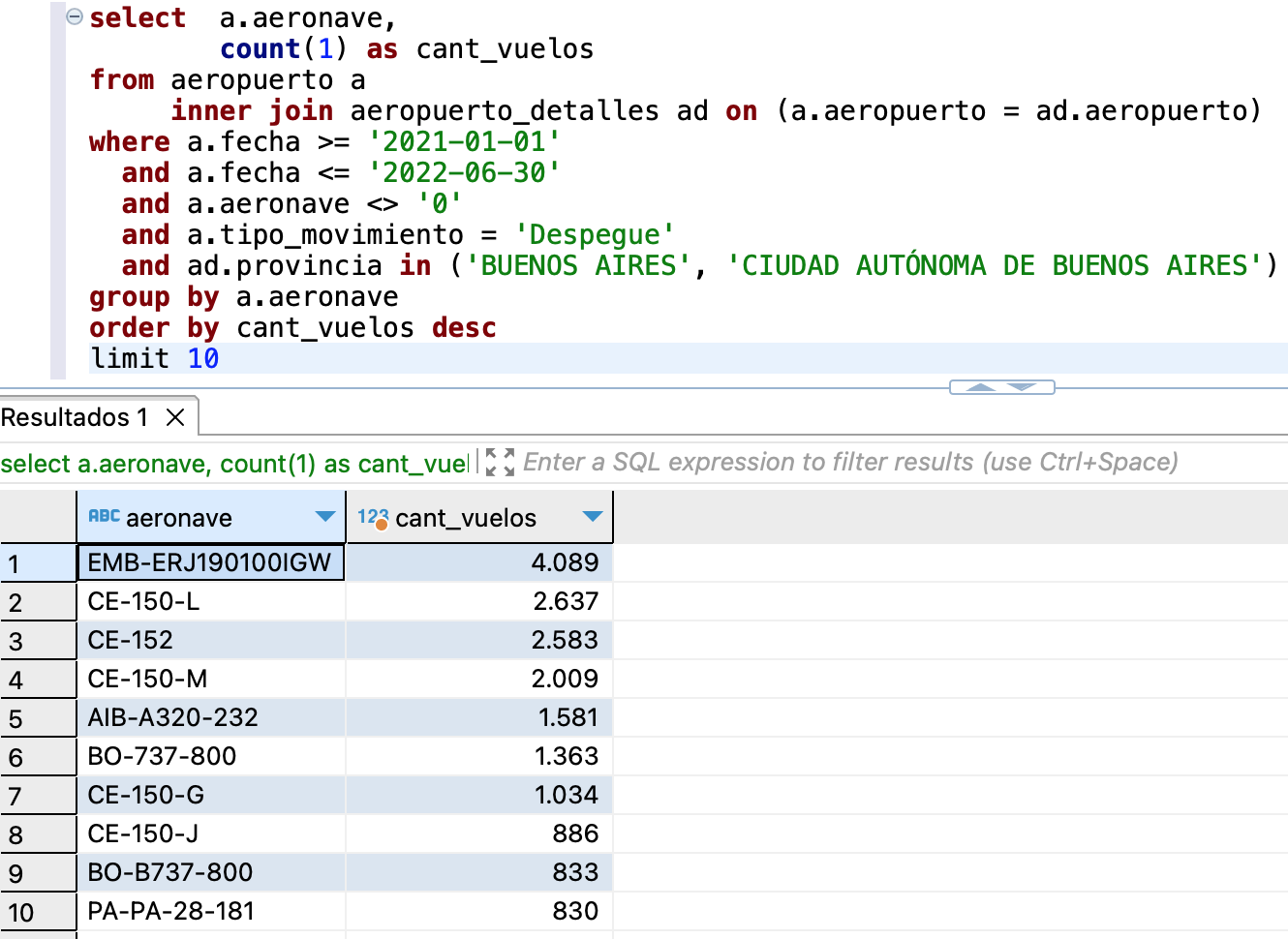
**Solución:**

****

<https://github.com/luxfo/edvai-data-engineering/blob/main/Ejercicios/TPFinal/Ejercicio1/TP1_Informe.pdf>

10. Cuales son las 10 aeronaves más utilizadas entre el 01/01/2021 y el 30/06/22 que despegaron desde la Ciudad autónoma de Buenos Aires o de Buenos Aires, exceptuando aquellas aeronaves que no cuentan con nombre. Mostrar consulta y Visualización.

**Solución:**

****

<https://github.com/luxfo/edvai-data-engineering/blob/main/Ejercicios/TPFinal/Ejercicio1/TP1_Informe.pdf>

11. Qué datos externos agregaría en este dataset que mejoraría el análisis de los datos.

**Solución:**

Nro. de Vuelo: Para poder identificar mejor la relación de los despegues y aterrizajes.

Duración: Duración del vuelo (en caso de tener la relación con el nro de vuelo se puede calcular el mismo con las fechas de despegue y aterrizaje).

Estado: Estado del vuelo (Demorado, En tiempo, etc)

Motivo: Motivo en caso de demora.

Cantidad asientos: Asientos disponibles en el avión (para ver si estuvo lleno o vacío el vuelo).

12. Elabore sus conclusiones y recomendaciones sobre este proyecto.

**Solución:**

Considerando que los datasets a utilizar se encuentran en la nube se recomienda una arquitectura cloud para mantener el proceso dentro de la misma, y de esta forma tener asegurado un buen mantenimiento, seguridad, escalabilidad y abaratar costos.

También considerar agregar los datos del punto 11 para mejorar el análisis y poder dar más herramientas al negocio para la toma de decisiones.

13. Proponer una arquitectura alternativa para este proceso ya sea con herramientas on premise o cloud (Sí aplica)

**Solución:**

