

## EXAMEN FINAL - RESPUESTAS

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-despegues-procesados-por-administracion-nacional-aviacion-civil-anac>

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://data-engineer-edvai.s3.amazonaws.com/2021-informe-ministerio.csv>

2022: <https://data-engineer-edvai.s3.amazonaws.com/202206-informe-ministerio.csv>

Aeropuertos\_detalle:

[https://data-engineer-edvai.s3.amazonaws.com/aeropuertos\\_detalle.csv](https://data-engineer-edvai.s3.amazonaws.com/aeropuertos_detalle.csv)

```
hadoop@0485e86b7577:~/scripts$ ./ingestion.sh
--2023-09-20 21:06:56-- https://data-engineer-edvai.s3.amazonaws.com/2021-informe-ministerio.csv
Resolving data-engineer-edvai.s3.amazonaws.com (data-engineer-edvai.s3.amazonaws.com)... 54.231.228.185, 52.216.207.163, 52.216.146.171, ...
Connecting to data-engineer-edvai.s3.amazonaws.com (data-engineer-edvai.s3.amazonaws.com)|54.231.228.185|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 32322556 (31M) [text/csv]
Saving to: '/home/hadoop/landing/2021-informe-ministerio.csv'

2021-informe-ministerio.csv      100%[=====]

2023-09-20 21:06:57 (69.0 MB/s) - '/home/hadoop/landing/2021-informe-ministerio.csv' saved [32322556/32322556]

--2023-09-20 21:06:57-- https://data-engineer-edvai.s3.amazonaws.com/202206-informe-ministerio.csv
Resolving data-engineer-edvai.s3.amazonaws.com (data-engineer-edvai.s3.amazonaws.com)... 52.217.117.105, 3.5.19.101, 16.182.107.73, ...
Connecting to data-engineer-edvai.s3.amazonaws.com (data-engineer-edvai.s3.amazonaws.com)|52.217.117.105|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 22833520 (22M) [text/csv]
Saving to: '/home/hadoop/landing/202206-informe-ministerio.csv'

202206-informe-ministerio.csv   100%[=====]

2023-09-20 21:06:57 (56.9 MB/s) - '/home/hadoop/landing/202206-informe-ministerio.csv' saved [22833520/22833520]

--2023-09-20 21:06:57-- https://data-engineer-edvai.s3.amazonaws.com/aeropuertos_detalle.csv
Resolving data-engineer-edvai.s3.amazonaws.com (data-engineer-edvai.s3.amazonaws.com)... 54.231.228.185, 52.216.207.163, 52.216.146.171, ...
Connecting to data-engineer-edvai.s3.amazonaws.com (data-engineer-edvai.s3.amazonaws.com)|54.231.228.185|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 136007 (133K) [text/csv]
Saving to: '/home/hadoop/landing/aeropuertos_detalle.csv'

aeropuertos_detalle.csv        100%[=====]

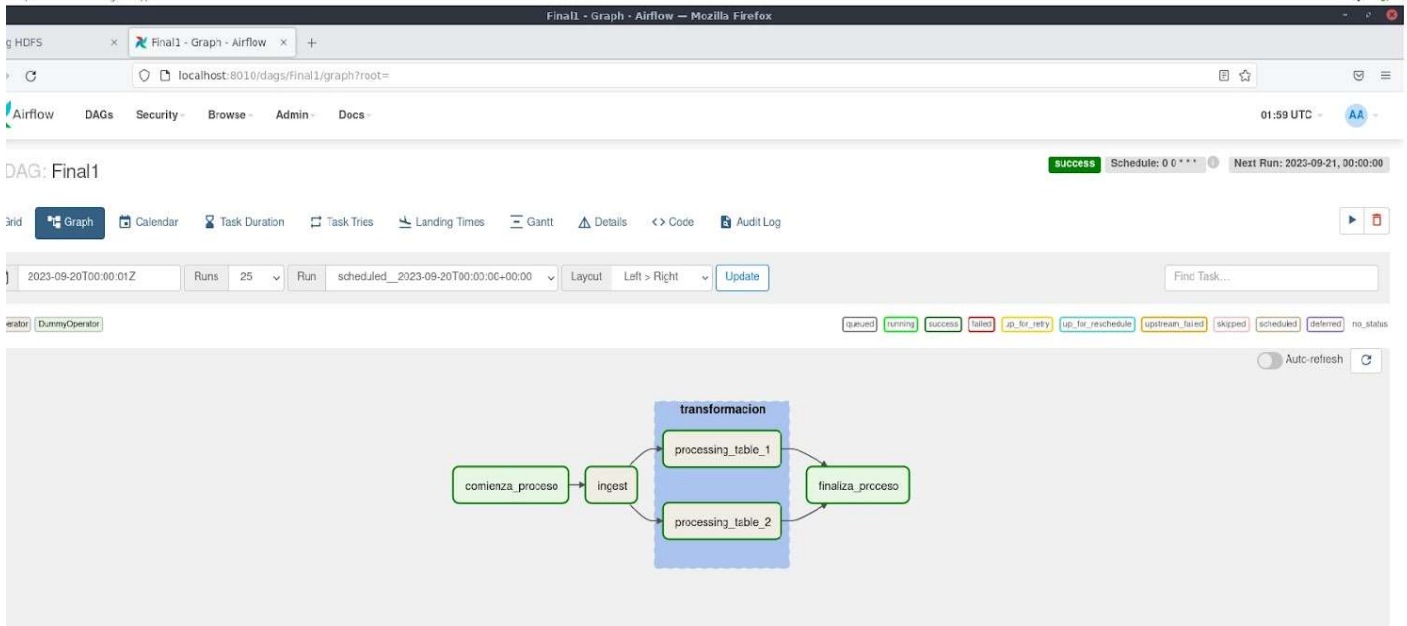
2023-09-20 21:06:57 (21.0 MB/s) - '/home/hadoop/landing/aeropuertos_detalle.csv' saved [136007/136007]

Deleted /ingest/2021-informe-ministerio.csv
Deleted /ingest/202206-informe-ministerio.csv
Deleted /ingest/aeropuertos_detalle.csv
hadoop@0485e86b7577:~/scripts$
```

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)

```
hive> show databases;
OK
default
f1
northwind_analytics
tripdata
viajes
Time taken: 1.198 seconds, Fetched: 5 row(s)
hive> use viajes
> ;
OK
Time taken: 0.104 seconds
hive> CREATE TABLE aeropuerto_tabla (fecha DATE, horaUTC STRING, clase_de_vuelo STRING, clasificacion_de_vuelo STRING, tipo_de_movimiento STRING, aeropuerto INT);
OK
Time taken: 1.901 seconds
hive> CREATE TABLE aeropuerto_detalle_tabla (aeropuerto STRING, oac STRING, iata STRING, tipo STRING, denominacion STRING, coordenadas STRING, latitud STRING, direccion_ref STRING, condicion STRING, control STRING, region STRING, uso STRING, trafico STRING, sna STRING, concesionado STRING, provincia STRING);
OK
Time taken: 0.189 seconds
hive> show tables;
OK
aeropuerto_detalle_tabla
aeropuerto_tabla
Time taken: 0.128 seconds, Fetched: 2 row(s)
hive>
```

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\_detalle\_tabla



4. Realizar las siguientes 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)

Amazon WorkSpaces

Amazon WorkSpaces

View Settings Support

DBeaver 22.3.5 - aeropuerto\_tabla

File Edit Navigate Search SQL Editor Database Window Help

SQL SQL Commit Rollback Auto tripdata tripdata

Da x Pr

Enter a part o...

Enter a SQL expression to filter results (use Ctrl+Space)

tripdata viajes Tables aeropuerto\_tabla

postgres - localhost:5432

tripdata - localhost:5432

default

f1

northwind\_analytics

tripdata

viages

Tables

aeropuerto\_det

aeropuerto\_tabla

Views

Indexes

Procedures

Data Types

Project - G x

Name

Bookmarks

Diagrams

Scripts

viages

Record

Refresh Save Cancel

Export data

200

200+

200 rows | fetched - 547ms (224ms fetch), on 2023-09-20 at 23:44:39

ART en\_US

|  | id | fecha      | msc_hora:utc | msc_clase_de_vuelo                   | msc_clasificacion_de_vuelo | msc_tipo_de_movimiento | msc_aeropuerto | msc_origen_destino | msc_aerolinea_nombre            | msc_aeronave   | msc_pasajeros |
|--|----|------------|--------------|--------------------------------------|----------------------------|------------------------|----------------|--------------------|---------------------------------|----------------|---------------|
|  | 1  | 2021-01-01 | 00:02        | Vuelo Privado con Matricula Nacional | Domestico                  | Despegue               | PAR            | ROS                | 0                               | PA-PA-28-181   | 0             |
|  | 2  | 2021-01-01 | 00:24        | Regular                              | Domestico                  | Aterrizaje             | EZE            | GRA                | AEROLINEAS ARGENTINAS           | BO-8737-88B    | 70            |
|  | 3  | 2021-01-01 | 00:26        | Regular                              | Domestico                  | Aterrizaje             | EZE            | ECA                | AEROLINEAS ARGENTINAS           | BO-737-800     | 70            |
|  | 4  | 2021-01-01 | 00:29        | Regular                              | Domestico                  | Aterrizaje             | EZE            | SAL                | AEROLINEAS ARGENTINAS           | BO-8-737-76N   | 12            |
|  | 5  | 2021-01-01 | 00:37        | Regular                              | Domestico                  | Aterrizaje             | EZE            | TUC                | AEROLINEAS ARGENTINAS           | EMB-ERJ190100G | 26            |
|  | 6  | 2021-01-01 | 01:00        | Vuelo Privado con Matricula Nacional | Domestico                  | Aterrizaje             | ROS            | PAR                | 0                               | PA-PA-28-181   | 0             |
|  | 7  | 2021-01-01 | 07:09        | Regular                              | Domestico                  | Despegue               | EZE            | ECA                | AEROLINEAS ARGENTINAS           | BO-8737-800    | 35            |
|  | 8  | 2021-01-01 | 08:10        | Regular                              | Domestico                  | Despegue               | EZE            | JUJ                | AEROLINEAS ARGENTINAS           | EMB-ERJ190100G | 46            |
|  | 9  | 2021-01-01 | 08:14        | Regular                              | Domestico                  | Despegue               | EZE            | BAR                | AEROLINEAS ARGENTINAS           | BO-737-800     | 81            |
|  | 10 | 2021-01-01 | 08:35        | Regular                              | Domestico                  | Despegue               | EZE            | USU                | AEROLINEAS ARGENTINAS           | BO-8737-88B    | 74            |
|  | 11 | 2021-01-01 | 09:04        | Regular                              | Domestico                  | Despegue               | EZE            | NEU                | AEROLINEAS ARGENTINAS           | BO-8737-800    | 67            |
|  | 12 | 2021-01-01 | 09:10        | Vuelo Oficial Nacional               | Domestico                  | Despegue               | CBA            | VRU                | 0                               | EUR-AS-350 B3  | 0             |
|  | 13 | 2021-01-01 | 09:11        | Regular                              | Domestico                  | Despegue               | EZE            | USU                | AEROLINEAS ARGENTINAS           | BO-737-85H     | 71            |
|  | 14 | 2021-01-01 | 09:50        | Regular                              | Domestico                  | Despegue               | EZE            | BAR                | FB LINEAS AEREAS - FLYBOI       | BO-737-88J     | 89            |
|  | 15 | 2021-01-01 | 10:00        | Regular                              | Domestico                  | Aterrizaje             | JUJ            | EZE                | AEROLINEAS ARGENTINAS           | EMB-ERJ190100G | 46            |
|  | 16 | 2021-01-01 | 10:11        | Regular                              | Domestico                  | Aterrizaje             | ECA            | EZE                | AEROLINEAS ARGENTINAS           | BO-8737-800    | 35            |
|  | 17 | 2021-01-01 | 10:14        | Regular                              | Domestico                  | Aterrizaje             | BAR            | EZE                | AEROLINEAS ARGENTINAS           | BO-737-800     | 80            |
|  | 18 | 2021-01-01 | 10:16        | Regular                              | Domestico                  | Despegue               | EZE            | BAR                | AEROLINEAS ARGENTINAS           | BO-8737-81D    | 80            |
|  | 19 | 2021-01-01 | 10:21        | Regular                              | Domestico                  | Despegue               | EZE            | CRV                | AEROLINEAS ARGENTINAS           | BO-737-88K     | 28            |
|  | 20 | 2021-01-01 | 10:34        | Regular                              | Domestico                  | Despegue               | EZE            | TUC                | AEROLINEAS ARGENTINAS           | EMB-ERJ190100G | 43            |
|  | 21 | 2021-01-01 | 10:37        | Regular                              | Domestico                  | Aterrizaje             | NEU            | EZE                | AEROLINEAS ARGENTINAS           | BO-8737-800    | 66            |
|  | 22 | 2021-01-01 | 10:39        | Regular                              | Domestico                  | Despegue               | EZE            | DOZ                | AEROLINEAS ARGENTINAS           | BO-737-800     | 67            |
|  | 23 | 2021-01-01 | 10:44        | Regular                              | Domestico                  | Despegue               | EZE            | SAL                | AEROLINEAS ARGENTINAS           | BO-8737-76N    | 59            |
|  | 24 | 2021-01-01 | 10:58        | Vuelo Oficial Nacional               | Domestico                  | Aterrizaje             | CBA            | VRU                | DIRECCION GENERAL DE AEROLINEAS | EUR-AS-350 B3  | 0             |
|  | 25 | 2021-01-01 | 11:02        | Regular                              | Domestico                  | Despegue               | ECA            | EZE                | AEROLINEAS ARGENTINAS           | BO-8737-800    | 3             |
|  | 26 | 2021-01-01 | 11:06        | Regular                              | Domestico                  | Aterrizaje             | SAL            | EZE                | AEROLINEAS ARGENTINAS           | BO-8-737-76N   | 59            |
|  | 27 | 2021-01-01 | 11:25        | Regular                              | Domestico                  | Despegue               | JUJ            | EZE                | AEROLINEAS ARGENTINAS           | EMB-ERJ190100G | 32            |
|  | 28 | 2021-01-01 | 11:29        | Regular                              | Domestico                  | Despegue               | EZE            | CBA                | AEROLINEAS ARGENTINAS           | BO-8737-81D    | 80            |
|  | 29 | 2021-01-01 | 11:30        | Regular                              | Domestico                  | Despegue               | BAR            | EZE                | AEROLINEAS ARGENTINAS           | BO-737-800     | 1             |
|  | 30 | 2021-01-01 | 11:43        | Regular                              | Domestico                  | Despegue               | NEU            | EZE                | AEROLINEAS ARGENTINAS           | BO-8737-800    | 46            |
|  | 31 | 2021-01-01 | 11:47        | Regular                              | Domestico                  | Despegue               | EZE            | ECA                | AEROLINEAS ARGENTINAS           | BO-737-81X     | 64            |
|  | 32 | 2021-01-01 | 11:48        | Regular                              | Domestico                  | Aterrizaje             | BAR            | EZE                | FB LINEAS AEREAS - FLYBOI       | BO-737-88J     | 87            |

Amazon WorkSpaces

Amazon WorkSpaces View Settings Support

DBeaver 22.3.5 - aeropuerto\_detalle\_tabla

File Edit Navigate Search SQL Editor Database Window Help

SQL Commit Rollback Auto tripdata tripdata

Da X Pr Properties Data ER Diagram

Enter a part o... Enter a SQL expression to filter results (use Ctrl+Space)

Grid

postgres - localhost:5432

tripdata - localhost:100

default

tripdata

northwind\_analytics

viajes

tables

aeropuerto\_detalle

aeropuerto\_detalle

Views

Indexes

Procedures

Data Types

Project - G x

Name

Bookmarks

Diagrams

Scripts

| id | idc aeropuerto | idc pac | idc iata | idc tipo  | idc denominacion                   | idc coordenadas        | idc latitud  | idc longitud | idc elev | idc uom_elev | idc ref                | idc distancia_ref |
|----|----------------|---------|----------|-----------|------------------------------------|------------------------|--------------|--------------|----------|--------------|------------------------|-------------------|
| 1  | ACH            | [NULL]  | [NULL]   | Aeródromo | CORONEL BOGADO/AGROSERVICIOS       | *33°16'20"S 60°34'14"W | -60.57066000 | -33.27226000 | 44       | Metros       | Coronel Bogado         | 6                 |
| 2  | ACH            | [NULL]  | [NULL]   | Aeródromo | GENERAL ACHA                       | *37°24' 6"S 64°36'49"W | -64.61351000 | -37.40164000 | 277      | Metros       | General Acha           | 3                 |
| 3  | ACM            | [NULL]  | [NULL]   | Aeródromo | ARRECIFES/LA CURA MALAL            | *34° 4'33"S 60° 8'30"W | -60.14170000 | -34.07574000 | 37       | Metros       | Arrecifes              | 4                 |
| 4  | ADO            | SAWD    | PUD      | Aeródromo | PUERTO DESEADO                     | *47°44' 6"S 65°54'15"W | -65.90410000 | -47.73511000 | 82       | Metros       | Puerto Deseado         | 2                 |
| 5  | ADT            | [NULL]  | [NULL]   | Aeródromo | BANDERA/AGROSERVICIOS DOÑA TERESA  | *28°51'19"S 62°15'53"W | -62.26462000 | -28.85541000 | 75       | Metros       | Bandera                | 4                 |
| 6  | ADU            | [NULL]  | [NULL]   | Aeródromo | BANDERA/DUTTO                      | *28°52' 1"S 62°14'17"W | -62.23812000 | -28.86691000 | 87       | Metros       | Bandera                | 3                 |
| 7  | AER            | SABE    | AEP      | Aeródromo | BUENOS AIRES/AEROPARQUE J. NEWBERY | *34°33'32"S 58°24'59"W | -58.41638889 | -34.55888889 | 5.6      | Metros       | Ciudad de Buenos Aires | 2                 |
| 8  | AGI            | SAWA    | [NULL]   | Aeródromo | PIEDRA DEL AGUILA                  | *40°11'32"S 70° 0'39"W | -70.01075000 | -40.19232000 | 649      | Metros       | Piedra del Águila      | 17                |
| 9  | AGR            | [NULL]  | [NULL]   | Aeródromo | AITA GRACIA                        | *31°39' 4"S 64°23'38"W | -64.39388889 | -31.65111111 | 533      | Metros       | Alta Gracia            | 2                 |
| 10 | AIB            | [NULL]  | [NULL]   | Aeródromo | CHACABUCCO/LAS DOS A               | *34°48'40"S 60°30'60"W | -60.51661000 | -34.81116000 | 60       | Metros       | Chacabuco              | 20                |
| 11 | AJA            | [NULL]  | [NULL]   | Aeródromo | LA LAJA                            | *31°21' 5"S 68°28'23"W | -68.47305556 | -31.35138889 | 650      | Metros       | La Laja                | 3                 |
| 12 | ALC            | [NULL]  | [NULL]   | Aeródromo | AMEGHINO/LA CHACRA                 | *34°48'34"S 62°21'23"W | -62.32293000 | -34.80942000 | 109      | Metros       | Florentino Ameghino    | 7                 |
| 13 | ALL            | [NULL]  | [NULL]   | Aeródromo | ALLEN                              | *38°57'30"S 67°48'10"W | -67.80277778 | -38.94833333 | 290      | Metros       | Allen                  | 3.5               |
| 14 | ALR            | [NULL]  | [NULL]   | Aeródromo | ALEJANDRO ROCA                     | *33°20'55"S 62°43'48"W | -62.73000000 | -33.34861111 | 210      | Metros       | Alejandro Roca         | 1.5               |
| 15 | AIT            | [NULL]  | [NULL]   | Aeródromo | CRUZ AITA                          | *33° 0'28"S 61°50'13"W | -61.83694444 | -33.00777778 | 97       | Metros       | Cruz Alta              | 2                 |
| 16 | ALV            | [NULL]  | [NULL]   | Aeródromo | ALVEAR/AGROALVEAR SRL              | *33° 4' 7"S 60°40' 1"W | -60.66682000 | -33.06853000 | 32       | Metros       | Alvear                 | 4.5               |
| 17 | AMA            | [NULL]  | [NULL]   | Aeródromo | ROBERTS/LA AMALIA                  | *35° 6' 2"S 62° 0'60"W | -62.01657000 | -35.10043000 | 100      | Metros       | Roberts                | 6.5               |
| 18 | AME            | [NULL]  | [NULL]   | Aeródromo | AMEGHINO/SIGFRIDO ROHR             | *34°49'50"S 62°28' 4"W | -62.46787000 | -34.83057000 | 110      | Metros       | Florentino Ameghino    | 2                 |
| 19 | AMG            | [NULL]  | [NULL]   | Aeródromo | AMEGHINO                           | *34°50'43"S 62°28'58"W | -62.48266000 | -34.84541000 | 110      | Metros       | Florentino Ameghino    | 2.5               |
| 20 | ANA            | [NULL]  | [NULL]   | Aeródromo | PARANÁ/ALROCLUB                    | *31°45'28"S 60°22'26"W | -60.37388889 | -31.75777778 | 60       | Metros       | La Picada              | 15                |
| 21 | AND            | [NULL]  | [NULL]   | Aeródromo | ANDALGALÁ                          | *27°37'41"S 66°20'34"W | -66.34277778 | -27.62805556 | 900      | Metros       | Andalgala              | 1                 |
| 22 | APD            | [NULL]  | [NULL]   | Aeródromo | APÓSTOLES                          | *27°54'12"S 55°45'56"W | -55.76555556 | -27.90333333 | 117      | Metros       | Apóstoles              | 1.6               |
| 23 | ARR            | [NULL]  | [NULL]   | Aeródromo | ARRIBEROS                          | *34°13'19"S 61°23'12"W | -61.38666667 | -34.22194444 | 82       | Metros       | Arriberos              | 3.3               |
| 24 | ARS            | SAWR    | ARR      | Aeródromo | ALTO RIO SENGUERRI/D. C. ZLAPELIS  | *45° 0'48"S 70°48'46"W | -70.81202000 | -45.01320000 | 697      | Metros       | Alto Rio Senguerr      | 3                 |
| 25 | ARY            | [NULL]  | [NULL]   | Aeródromo | ARROYITO/ARCOR                     | *31°25'20"S 63° 0'18"W | -63.00500000 | -31.42222222 | 148      | Metros       | Arroyito               | 4                 |
| 26 | AZE            | [NULL]  | [NULL]   | Aeródromo | ZARATE                             | *34° 7' 6"S 59° 4'58"W | -59.08266000 | -34.11829000 | 26       | Metros       | Zarate                 | 6                 |
| 27 | ATN            | [NULL]  | [NULL]   | Aeródromo | TRENQUE LAUQUEN/LA ARGENTINA       | *35°59'52"S 62°42'23"W | -62.70635000 | -35.99773000 | 95       | Metros       | Trenque Lauquen        | 2                 |
| 28 | AUL            | [NULL]  | [NULL]   | Aeródromo | AMEGHINO/AUGUSTO LEGUIZAMÓN        | *34°49' 5"S 62°28'38"W | -62.47709000 | -34.81797000 | 109      | Metros       | Florentino Ameghino    | 3                 |
| 29 | AVA            | [NULL]  | [NULL]   | Aeródromo | ALVEAR/AEROPARQUE ROSARIO          | *33° 2'49"S 60°35'51"W | -60.59756000 | -33.04691000 | 26       | Metros       | Alvear                 | 2.5               |
| 30 | AVL            | [NULL]  | [NULL]   | Aeródromo | AVELLANEDA/PRESIDENTE AVELLANEDA   | *29° 6'22"S 59°39'27"W | -59.65750000 | -29.10611111 | 45       | Metros       | Avellaneda             | 1                 |
| 31 | AYA            | [NULL]  | [NULL]   | Aeródromo | AYACUCHO                           | *37° 9'45"S 58°28'23"W | -58.47305556 | -37.16250000 | 77       | Metros       | Ayacucho               | 0.9               |
| 32 | BAD            | [NULL]  | [NULL]   | Aeródromo | BERÓN DE ASTRADA/DINALUCA S.A.     | *27°22'58"S 57°38'39"W | -57.64416667 | -27.38277778 | 67       | Metros       | Berón de Astrada       | 21                |

Refresh Save Cancel Export data 200 200+ 200 rows fetched - 418ms (135ms fetch), on 2023-09-20 at 23:45:29

ART en\_US

Applications Places system data-engineer Terminal data-engineer Final - Code - Airflow DBeaver 22.3.5 - aerop... Wed Sep 20, 23:49

11:49 PM 9/20/2023



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.)

The screenshot shows the DBeaver 22.3.5 interface with the 'aeropuerto\_detalle' table selected. The table is located in the 'viajes' schema of the 'tripdata' database. The table structure is as follows:

| Column Name   | #  | Data Type | Length | Scale | N |
|---------------|----|-----------|--------|-------|---|
| aeropuerto    | 1  | STRING    |        |       |   |
| oac           | 2  | STRING    |        |       |   |
| iata          | 3  | STRING    |        |       |   |
| tipo          | 4  | STRING    |        |       |   |
| denominacion  | 5  | STRING    |        |       |   |
| coordenadas   | 6  | STRING    |        |       |   |
| latitud       | 7  | STRING    |        |       |   |
| longitud      | 8  | STRING    |        |       |   |
| elev          | 9  | FLOAT     | 7      | 7     |   |
| uom_elev      | 10 | STRING    |        |       |   |
| ref           | 11 | STRING    |        |       |   |
| distancia_ref | 12 | FLOAT     | 7      | 7     |   |
| direccion_ref | 13 | STRING    |        |       |   |
| condicion     | 14 | STRING    |        |       |   |
| control       | 15 | STRING    |        |       |   |
| region        | 16 | STRING    |        |       |   |
| uso           | 17 | STRING    |        |       |   |
| trafico       | 18 | STRING    |        |       |   |
| sna           | 19 | STRING    |        |       |   |
| concesionado  | 20 | STRING    |        |       |   |
| provincia     | 21 | STRING    |        |       |   |

The screenshot shows the DBeaver 22.3.5 interface with the 'aeropuerto\_tabla' table selected. The table is located in the 'viajes' schema of the 'tripdata' database. The table structure is as follows:

| Column Name            | #  | Data Type | Length |
|------------------------|----|-----------|--------|
| fecha                  | 1  | DATE      | 10     |
| horautc                | 2  | STRING    |        |
| clase_de_vuelo         | 3  | STRING    |        |
| clasificacion_de_vuelo | 4  | STRING    |        |
| tipo_de_movimiento     | 5  | STRING    |        |
| aeropuerto             | 6  | STRING    |        |
| origen_destino         | 7  | STRING    |        |
| aerolinea_nombre       | 8  | STRING    |        |
| aeronave               | 9  | STRING    |        |
| pasajeros              | 10 | INT       | 10     |

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

The screenshot shows the DBeaver 22.3.5 interface. The SQL Editor contains the following query:

```
SELECT COUNT(*)
FROM aeropuerto_tabla AS cantidad_vuelos
WHERE Fecha >= '2021-12-01' AND Fecha <= '2022-01-31';
```

The query is executed, and the Results pane shows the following result:

| 1 | 57,984 |
|---|--------|
|   |        |

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

The screenshot shows the DBeaver 22.3.5 interface. The SQL Editor contains the following query:

```
SELECT tipo_de_movimiento, COUNT(*)
FROM aeropuerto_tabla
WHERE fecha >= '2021-01-01' AND fecha <= '2022-06-30'
AND aerolinea_nombre = 'AEROLINEAS ARGENTINAS SA'
GROUP BY tipo_de_movimiento;
```

The query is executed, and the Results pane shows the following result:

| 1 | Atterrissage | 73,173 |
|---|--------------|--------|
| 2 | Décollage    | 74,331 |

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

Amazon WorkSpaces

Amazon WorkSpaces View Settings Support

DBeaver 22.3.5 - <tripdata> Script-1

File Edit Navigate Search SQL Editor Database Window Help

Auto SQL tripdata viajes

Database Na Projects

Enter a part of object name h...

postgres - localhost:5432

tripdata - localhost:10000

car\_rental\_db

default

f1

northwind\_analytics

tripdata

viajes

Tables

Views

Indexes

Procedures

Data Types

```

SELECT
    vf.fecha,
    vf.horaUTC,
    vf.aeropuerto_salida,
    acs.ciudad AS ciudad_salida,
    vf.aeropuerto_arribo,
    aca.ciudad AS ciudad_arribo,
    vf.pasajeros
FROM (
    SELECT
        fecha,
        horaUTC,
        CASE
            WHEN tipo_de_movimiento = 'Despegue' THEN aeropuerto
            ELSE origen_destino
        END AS aeropuerto_salida,
        CASE
            WHEN tipo_de_movimiento = 'Aterrizaje' THEN aeropuerto
            ELSE origen_destino
        END AS aeropuerto_arribo,
        pasajeros
    FROM aeropuerto_detalle_tabla
    WHERE fecha BETWEEN '2022-01-01' AND '2022-06-30'
) vf
LEFT JOIN (
    SELECT aeropuerto, ref AS ciudad
    FROM aeropuerto_detalle_tabla
) acs ON vf.aeropuerto_salida = acs.aeropuerto
LEFT JOIN (
    SELECT aeropuerto, ref AS ciudad
    FROM aeropuerto_detalle_tabla
) aca ON vf.aeropuerto_arribo = aca.aeropuerto;
  
```

Results 1 X

SELECT vf.fecha, vf.horaUTC, vf.aeropuerto\_salida, acs.ciudad

| fecha      | horaUTC | aeropuerto_salida | ciudad_salida        | aeropuerto_arribo | ciudad_arribo          | pasajeros |
|------------|---------|-------------------|----------------------|-------------------|------------------------|-----------|
| 2022-01-01 | 00:01   | ECA               | El Calafate          | AER               | Ciudad de Buenos Aires | 69        |
| 2022-01-01 | 00:05   | SAL               | Salta                | AER               | Ciudad de Buenos Aires | 65        |
| 2022-01-01 | 00:05   | IGU               | Cataratas del Iguazú | AER               | Ciudad de Buenos Aires | 41        |
| 2022-01-01 | 00:09   | GAL               | Río Gallegos         | AER               | Ciudad de Buenos Aires | 73        |
| 2022-01-01 | 00:12   | DOZ               | Mendoza              | SRA               | San Rafael             | 0         |

Refresh Save Cancel Export data

ART en\_US Writeable Smart insert 33 : 1 : 898 Sel: 0 | 0

Applications Places System

data-engineer DBeaver 22.3.5 - <tripdata>

Tue Oct 10, 21:54

9:54 PM

10/10/2023

DBeaver 22.3.5 - <tripdata> Script-1

File Edit Navigate Search SQL Editor Database Window Help

Auto SQL tripdata viajes

Database Na Projects

Enter a part of object name h...

postgres - localhost:5432

tripdata - localhost:10000

car\_rental\_db

default

f1

northwind\_analytics

tripdata

viajes

Tables

Views

Indexes

Procedures

Data Types

```

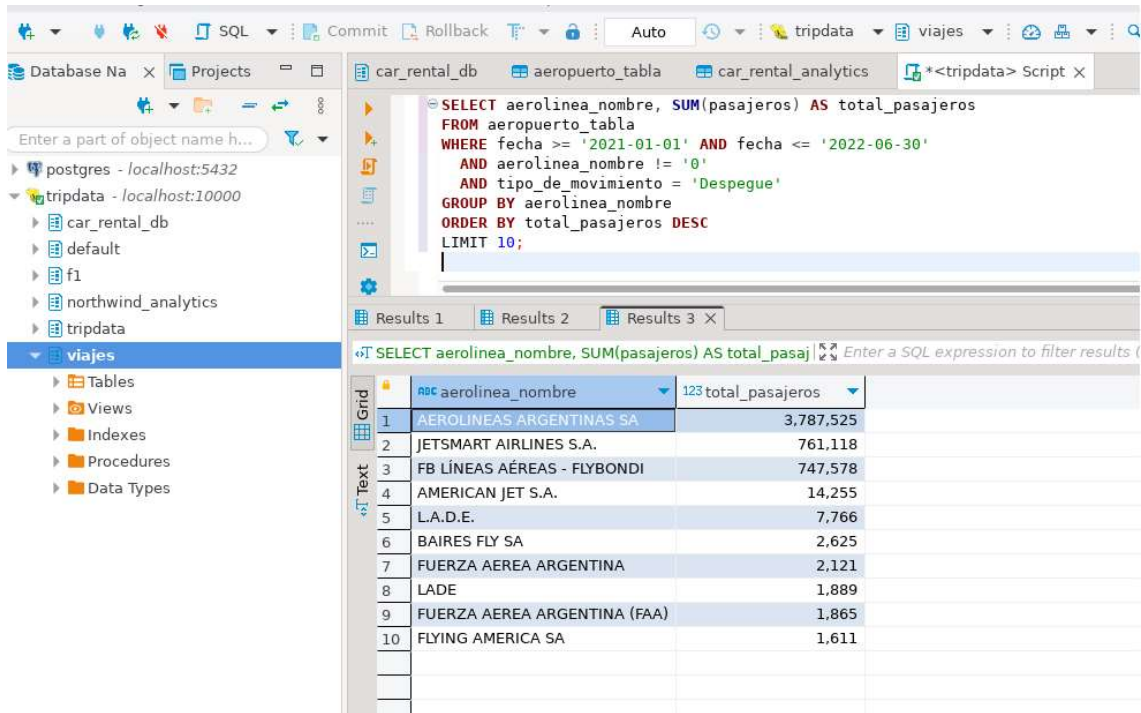
SELECT vf.fecha, vf.horaUTC, vf.aeropuerto_salida, acs.ciudad
  
```

Results 1 X

SELECT vf.fecha, vf.horaUTC, vf.aeropuerto\_salida, acs.ciudad

| fecha      | horaUTC | aeropuerto_salida | ciudad_salida           | aeropuerto_arribo | ciudad_arribo           | pasajeros |
|------------|---------|-------------------|-------------------------|-------------------|-------------------------|-----------|
| 2022-01-01 | 00:01   | ECA               | El Calafate             | AER               | Ciudad de Buenos Aires  | 69        |
| 2022-01-01 | 00:05   | SAL               | Salta                   | AER               | Ciudad de Buenos Aires  | 65        |
| 2022-01-01 | 00:05   | IGU               | Cataratas del Iguazú    | AER               | Ciudad de Buenos Aires  | 41        |
| 2022-01-01 | 00:09   | GAL               | Río Gallegos            | AER               | Ciudad de Buenos Aires  | 73        |
| 2022-01-01 | 00:12   | DOZ               | Mendoza                 | SRA               | San Rafael              | 0         |
| 2022-01-01 | 00:12   | AER               | Ciudad de Buenos Aires  | EZE               | Capital Federal         | 0         |
| 2022-01-01 | 00:16   | ROS               | Rosario                 | AER               | Ciudad de Buenos Aires  | 20        |
| 2022-01-01 | 00:21   | AER               | Ciudad de Buenos Aires  | EZE               | Capital Federal         | 0         |
| 2022-01-01 | 00:28   | ROS               | Rosario                 | AER               | Ciudad de Buenos Aires  | 5         |
| 2022-01-01 | 00:30   | CBA               | Córdoba                 | AER               | Ciudad de Buenos Aires  | 39        |
| 2022-01-01 | 00:42   | CRR               | Corrientes              | AER               | Ciudad de Buenos Aires  | 85        |
| 2022-01-01 | 00:42   | JUJ               | San Salvador de Jujuy   | EZE               | Capital Federal         | 88        |
| 2022-01-01 | 00:48   | AER               | Ciudad de Buenos Aires  | CBA               | Córdoba                 | 85        |
| 2022-01-01 | 01:00   | ROS               | Rosario                 | AER               | Ciudad de Buenos Aires  | 5         |
| 2022-01-01 | 01:02   | ECA               | El Calafate             | EZE               | Capital Federal         | 31        |
| 2022-01-01 | 01:33   | BAR               | San Carlos de Bariloche | AER               | Ciudad de Buenos Aires  | 11        |
| 2022-01-01 | 01:39   | CBA               | Córdoba                 | POS               | Posadas                 | 0         |
| 2022-01-01 | 02:25   | JUJ               | San Salvador de Jujuy   | EZE               | Capital Federal         | 91        |
| 2022-01-01 | 02:41   | IGU               | Cataratas del Iguazú    | AER               | Ciudad de Buenos Aires  | 86        |
| 2022-01-01 | 03:55   | AER               | Ciudad de Buenos Aires  | BAR               | San Carlos de Bariloche | 93        |
| 2022-01-01 | 04:44   | POS               | Posadas                 | AER               | Ciudad de Buenos Aires  | 57        |

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



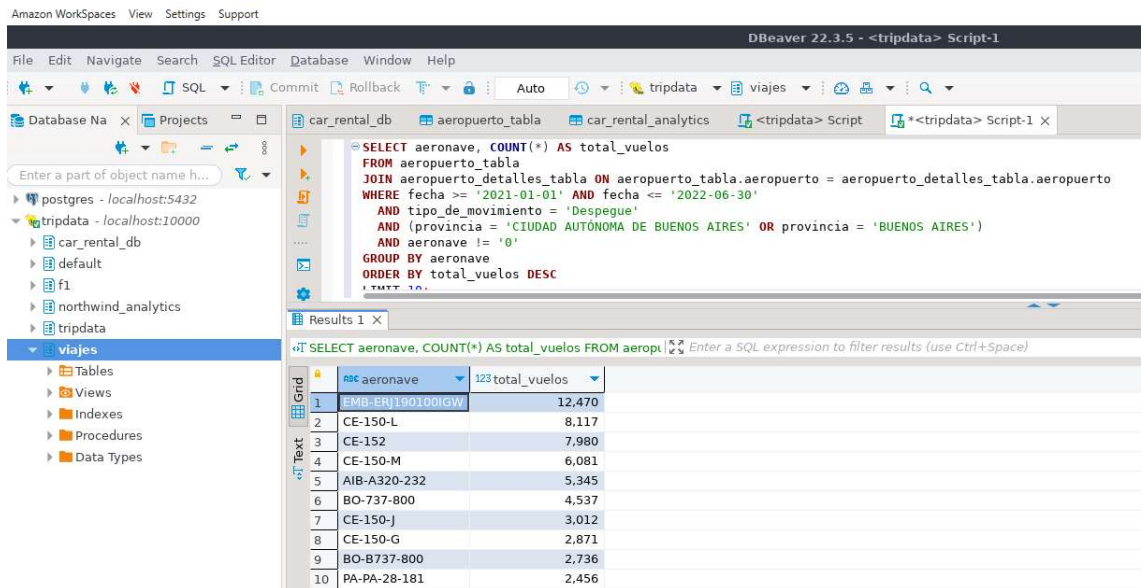
The screenshot shows the DBeaver SQL editor with a query to find the top 10 airlines by passenger count. The query is as follows:

```
SELECT aerolinea_nombre, SUM(pasajeros) AS total_pasajeros
FROM aeropuerto_tabla
WHERE fecha >= '2021-01-01' AND fecha <= '2022-06-30'
AND aerolinea_nombre != '0'
AND tipo_de_movimiento = 'Despegue'
GROUP BY aerolinea_nombre
ORDER BY total_pasajeros DESC
LIMIT 10;
```

The results are displayed in a grid view, showing the top 10 airlines and their total passenger counts:

| aerolinea_nombre             | total_pasajeros |
|------------------------------|-----------------|
| AEROLINEAS ARGENTINAS SA     | 3,787,525       |
| JETSMART AIRLINES S.A.       | 761,118         |
| FB LINEAS AEREAS - FLYBONDI  | 747,578         |
| AMERICAN JET S.A.            | 14,255          |
| L.A.D.E.                     | 7,766           |
| BAIRES FLY SA                | 2,625           |
| FUERZA AEREA ARGENTINA       | 2,121           |
| LADE                         | 1,889           |
| FUERZA AEREA ARGENTINA (FAA) | 1,865           |
| FLYING AMERICA SA            | 1,611           |

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



The screenshot shows the DBeaver SQL editor with a query to find the top 10 aircraft by flight count. The query is as follows:

```
SELECT aeronave, COUNT(*) AS total_vuelos
FROM aeropuerto_tabla
JOIN aeropuerto_detalle_tabla ON aeropuerto_tabla.aeropuerto = aeropuerto_detalle_tabla.aeropuerto
WHERE fecha >= '2021-01-01' AND fecha <= '2022-06-30'
AND tipo_de_movimiento = 'Despegue'
AND (provincia = 'CIUDAD AUTÓNOMA DE BUENOS AIRES' OR provincia = 'BUENOS AIRES')
AND aeronave != '0'
GROUP BY aeronave
ORDER BY total_vuelos DESC
LIMIT 10;
```

The results are displayed in a grid view, showing the top 10 aircraft and their total flight counts:

| aeronave         | total_vuelos |
|------------------|--------------|
| EMB-ER190-100IGW | 12,470       |
| CE-150-L         | 8,117        |
| CE-152           | 7,980        |
| CE-150-M         | 6,081        |
| AIB-A320-232     | 5,345        |
| BO-737-800       | 4,537        |
| CE-150-J         | 3,012        |
| CE-150-G         | 2,871        |
| BO-B737-800      | 2,736        |
| PA-PA-28-181     | 2,456        |



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

Información de Tráfico Aéreo: Entender el tráfico aéreo en diferentes horarios y aeropuertos puede ayudar a optimizar las rutas y horarios de vuelo.

Condiciones Meteorológicas: Agregaría datos sobre el clima en los despegues y aterrizajes de los horarios frecuentes, también en los aeropuertos específicos para ayudar a analizar la relación entre el clima y los retrasos, cancelaciones o incluso accidentes.

Datos de Mantenimiento de Aeronaves: para poder ver problemas o retrasos en vuelos y proporcionar información sobre la eficiencia y seguridad de las aerolíneas.

Información de Tráfico Aéreo: Entender el tráfico aéreo en diferentes horarios y aeropuertos puede ayudar a optimizar las rutas y horarios de vuelo.

12. Elabore sus conclusiones y recomendaciones sobre este proyecto.

Desde mi análisis puedo decir que he encontrado algunos faltantes de información, que dan incongruencias, por ejemplo la disminución o agregación de uno o más pasajeros en un viaje, puede que sea por una escala, pero este dato se encuentra sin contención, en caso de que se desee tomar los datos como reales, además el hecho de que el schema está de con terminología técnica y con caracteres, como tildes y o espacios, que intervienen en la fluida manipulación de los datos.

Recomiendo el uso de herramientas de visualización de datos para hacer que el análisis sea más comprensible y accesible.

Implementar una solución de almacenamiento de datos escalable y fácil de consultar, como un data warehouse o un lago de datos, para manejar múltiples fuentes de datos.

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

GCP

Cloud Storage >> DataProc