

Caso IATA

Vladimir Castillo Pérez ¹, Maria Belarcazar ², Cesar Fuentes ³

¹ Estudiante Maestría en Ciencia de Datos, Pontificia Universidad Javeriana, Calí, Colombia (email: vladimir2022@javerianacali.edu.co) ² Estudiante Maestría en Ciencia de Datos, Pontificia Universidad Javeriana, Calí, Colombia (email: "mariabp@javerianacali.edu.co) ³ Estudiante Maestría en Ciencia de Datos, Pontificia Universidad Javeriana, Calí, Colombia (email: cesarfu94@javerianacali.edu.co)

Fecha: 1 de junio de 2024

Resumen:

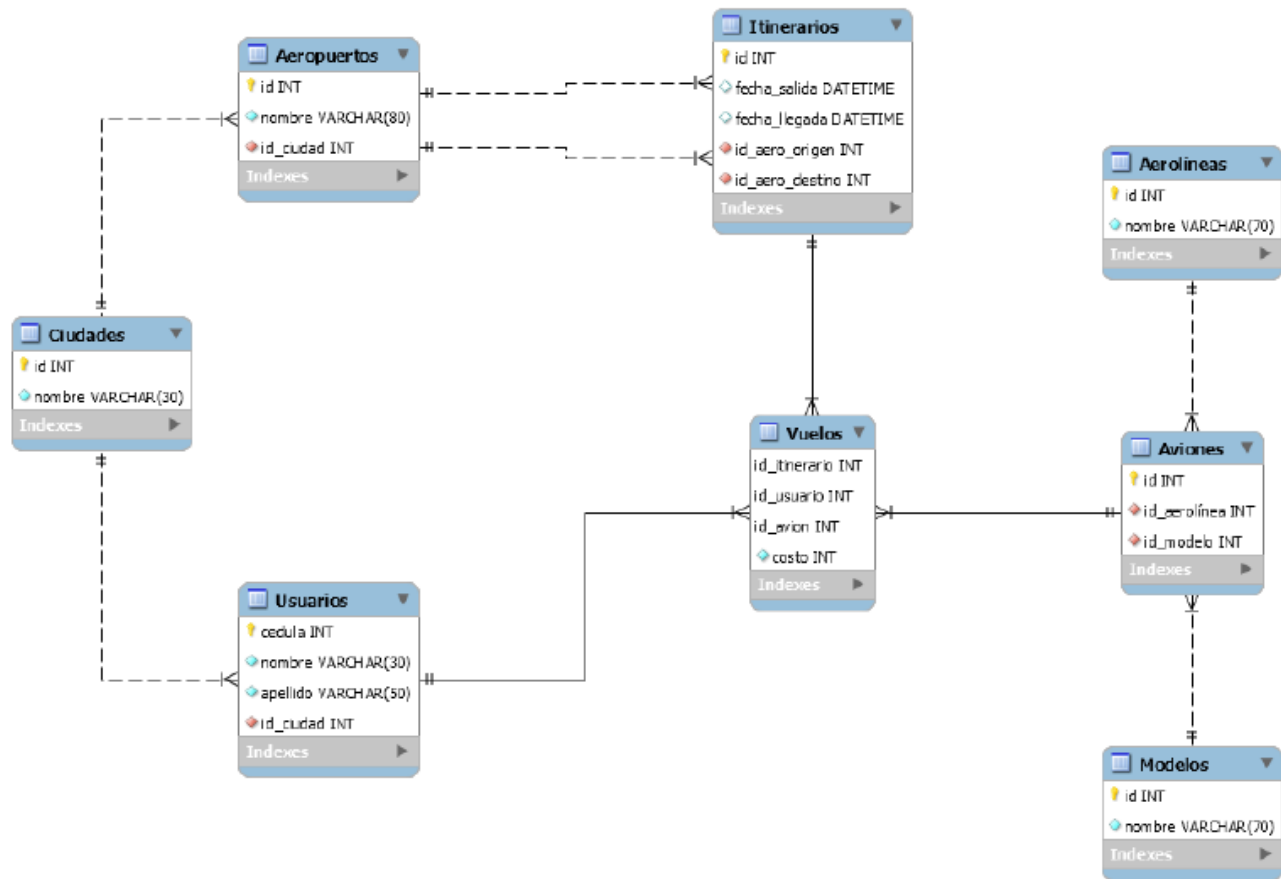
Palabras clave: COVID-19, Pandemia, Datos de contagios, Visualización de datos, Análisis estadístico

Contexto

Uno de los métodos de transporte más utilizados por los seres humanos es el transporte aéreo. Aunque el número de vuelos diarios se vio afectado por la pandemia de la COVID-19, es bien sabido que son muchos los aviones que diariamente se movilizan de un sitio a otro llevando pasajeros. Es por ello que, para tener un control de estos movimientos, se han creado diversas organizaciones internacionales que llevan un registro de aerolíneas, aviones, itinerarios, equipajes, entre otras cosas. Una de ellas es la Asociación Internacional de Transporte Aéreo, IATA por sus siglas en inglés. Esta asociación surge en el año 1919 en los Países Bajos y estaba conformada por 57 miembros de 31 países pertenecientes, principalmente a Europa y Norteamérica. Años más tarde, en 1945, fue relanzada en Cuba. Actualmente, incluye 290 aerolíneas de 120 países. Sus misiones principales comprenden la promoción de la seguridad, la protección y la fiabilidad en el transporte aéreo y el cuidado del medio ambiente. Todo esto en pro del beneficio económico de sus accionistas. Adicional a esto la IATA representa a las 290 aerolíneas que la conforman, simplifica los procesos y reduce los costos que aumentan su flujo financiero, y actúa como un puente para asegurar el movimiento de las personas alrededor del mundo.

Caso

La Asociación Internacional de Transporte Aéreo, IATA, posee una base de datos donde tiene almacenada la información recolectada de todos los vuelos hechos a nivel global a partir del año 1945. El Modelo Relacional que la IATA utilizó para la construcción de su Base de Datos Relacional fue el siguiente:



Alcance

En este Modelo Relacional se encuentran almacenados los siguientes datos:

1. Los usuarios y su respectiva ciudad de origen.
2. Los aeropuertos y la ciudad donde se encuentran ubicados.
3. Los aviones, sus modelos y la aerolínea a la que pertenece cada uno.
4. Los itinerarios, incluyendo fecha de salida, fecha de llegada, ciudad de origen del vuelo y ciudad de destino del vuelo.
5. Los vuelos, incluyendo itinerarios de cada vuelo, aviones implicados en cada vuelo, usuarios que tomaron determinado vuelo y costo de los diferentes vuelos registrados.

Con los datos almacenados en su base de datos relacional, la IATA quiere responder algunos requerimientos de análisis relacionados con los vuelos.

La IATA quiere analizar en qué medida la pandemia de COVID-19 afectó el transporte aéreo de pasajeros durante el año 2020. Los requerimientos de análisis son los siguientes:

1. ¿Cuál aerolínea realizó el mayor número de vuelos a la ciudad de Roma en el año 2019 y cuál en el año 2020?
2. Total de dinero recaudado por vuelos de cada aerolínea en el primer semestre del año 2019 y en el primer semestre del año 2020.
3. ¿Cuál modelo de avión realizó el mayor número de vuelos en el año 2019 y cuál en el año 2020?
4. ¿Cuál fue la ciudad cuyos habitantes viajaron más en el año 2019 y cuál en el año 2020?

Para este fin, la IATA contrata a un grupo de científicos de datos, expertos en la gestión de datos en bases de datos analíticas, para que implemente, a partir de su base de datos relacional, un Modelo Estrella que permita dar respuesta a los requerimientos de análisis. Para responder estos requerimientos y dar completa solución al caso deben tener en cuenta los siguientes puntos:

Análisis Crítico

El análisis crítico involucra evaluar la estructura de datos existente y la relevancia de los datos almacenados para responder a los requerimientos planteados. La base de datos relacional de la IATA incluye datos de usuarios, aeropuertos, aviones, itinerarios, modelos, ciudades, aerolíneas y vuelos.

Ventajas del Enfoque Relacional

- **Integridad de Datos:** La estructura relacional garantiza la consistencia y precisión de los datos a través de claves primarias y foráneas.
- **Facilidad de Acceso:** Permite la realización de consultas complejas mediante SQL, facilitando el acceso a información específica y detallada.

Desventajas del Enfoque Relacional

- **Rendimiento:** Consultas sobre grandes volúmenes de datos pueden ser lentas, especialmente al realizar análisis complejos que requieren unir múltiples tablas.
- **Escalabilidad:** Manejar y escalar bases de datos relacionales puede ser un desafío, particularmente con el creciente volumen de datos en la industria aérea.

Proceso de Construcción del Modelo Multidimensional

Para responder a los requerimientos de análisis de la IATA, se implementará un Modelo Estrella. Este modelo es adecuado para análisis OLAP (Procesamiento Analítico en Línea), proporcionando una estructura eficiente para consultas analíticas.

Pasos para la Construcción del Modelo Estrella:

Identificación de Hechos y Dimensiones:

Tabla de Hechos: Registro de vuelos, incluyendo métricas como costos.

Dimensiones:

- Dimensión Avión (modelo, aerolínea)
- Dimensión Itinerario (año, semestre, ciudad)

Diseño del Esquema Estrella:

- **Tabla de Hechos:** Contendrá claves foráneas de las dimensiones y medidas como total costos.
- **Tablas de Dimensiones:** Cada tabla de dimensión contendrá atributos detallados que describen las características específicas de cada dimensión.

Esquema Estrella Propuesto

Tabla de Hechos (Vuelos):

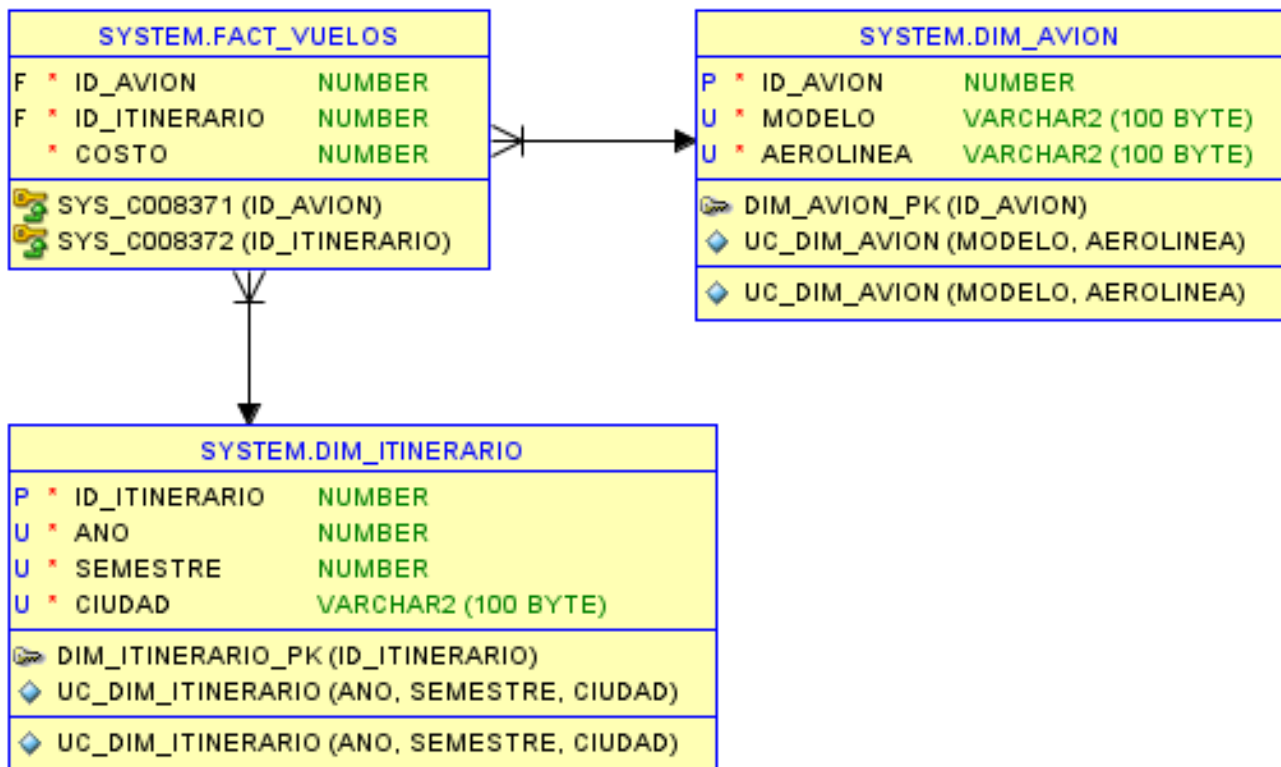
- ID_Avion
- ID_Itinerario
- Costos

Dimensión Avion:

- ID_Avion
- Modelo
- Aerolinea

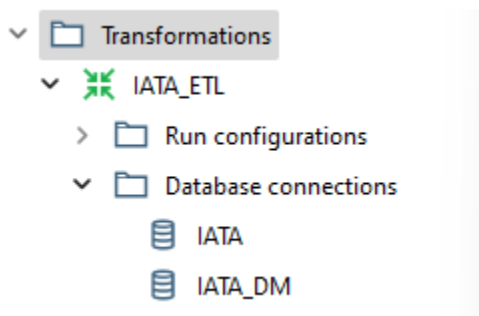
Dimensión Itinerario:

- ID_Itinerario
- Año
- Semestre
- Ciudad



Procesos ETL con Data-Integration

Se ingresa a la herramienta Spoon para realizar los procesos de carga al Data Mart, realizando inicialmente la conexión a las bases de datos



A continuación, se realizan los procesos de transformación, iniciando por “Input - Table Input”

Table input

Step name: Avion_EXT

Connection: IATA Edit... New... Wizard...

SQL Get SQL select statement...

```
SELECT AVIONES.ID_AVION, MODELOS.NOMBRE AS MODELO, AEROLINEAS.NOMBRE AS AEROLINEA
FROM AVIONES
JOIN MODELOS
ON MODELOS.ID_MODELO = AVIONES.ID_MODELO
JOIN AEROLINEAS
ON AEROLINEAS.ID_AEROLINEA = AVIONES.ID_AEROLINEA
```

Table input

Step name: Itinerario_EXT

Connection: IATA Edit... New... Wizard...

SQL Get SQL select statement...

```
SELECT
  I.ID_ITINERARIO,
  CAST(EXTRACT(YEAR FROM I.FECHA_SALIDA) AS INT) AS ANO,
  CAST(CASE
    WHEN EXTRACT(MONTH FROM I.FECHA_SALIDA) BETWEEN 1 AND 6 THEN 1
    WHEN EXTRACT(MONTH FROM I.FECHA_SALIDA) BETWEEN 7 AND 12 THEN 2
  END AS INT) AS SEMESTRE,
  CI.NOMBRE AS CIUDAD
FROM ITINERARIOS I
JOIN AEROPUERTOS A ON I.ID_AEROPUERTO_ORIGEN = A.ID_AEROPUERTO
JOIN CIUDADES CI ON A.ID_CIUDAD = CI.ID_CIUDAD
```

Table input

Step name: Vuelos_EXT

Connection: IATA Edit... New... Wizard...

SQL Get SQL select statement...

```
SELECT ID_ITINERARIO, ID_AVION, COSTO FROM VUELOS
```

Ahora se realizan los procesos “Output - Insert/Update”

Insert / update

Step name: Avion_LOAD

Connection: IATA_DM [Edit...] [New...] [Wizard...]

Target schema: C#VLADIMIR [Browse...]

Target table: DIM_AVION [Browse...]

Commit size: 100

Don't perform any updates: ☐

The key(s) to look up the value(s):

#	Table field	Comparator	Stream field1	Stream field2
1	ID_AVION	=	ID_AVION	
2	MODELO	=	MODELO	
3	AEROLINEA	=	AEROLINEA	

[Get fields]

Update fields:

#	Table field	Stream field	Update
1	ID_AVION	ID_AVION	Y
2	MODELO	MODELO	Y
3	AEROLINEA	AEROLINEA	Y

[Get update fields] [Edit mapping]

[?] Help [OK] [Cancel] [SQL]

Insert / update

Step name: Itinerario_LOAD

Connection: IATA_DM [Edit...] [New...] [Wizard...]

Target schema: C#VLADIMIR [Browse...]

Target table: DIM_ITINERARIO [Browse...]

Commit size: 100

Don't perform any updates: ☐

The key(s) to look up the value(s):

#	Table field	Comparator	Stream field1	Stream field2
1	ID_ITINERARIO	=	ID_ITINERARIO	
2	ANO	=	ANO	
3	SEMESTRE	=	SEMESTRE	
4	CIUDAD	=	CIUDAD	

[Get fields]

Update fields:

#	Table field	Stream field	Update
1	ID_ITINERARIO	ID_ITINERARIO	Y
2	ANO	ANO	Y
3	SEMESTRE	SEMESTRE	Y
4	CIUDAD	CIUDAD	Y

[Get update fields] [Edit mapping]

[?] Help [OK] [Cancel] [SQL]

Insert / update

Step name: Vuelos_LOAD

Connection: IATA_DM [Edit...] [New...] [Wizard...]

Target schema: C##VLADIMIR [Browse...]

Target table: FACT_VUELOS [Browse...]

Commit size: 100

Don't perform any updates: ☐

The key(s) to look up the value(s):

#	Table field	Comparator	Stream field1	Stream field2
1	ID_ITINERARIO	=	ID_ITINERARIO	
2	ID_AVION	=	ID_AVION	
3	COSTO	=	COSTO	

[Get fields]

Update fields:

#	Table field	Stream field	Update
1	ID_ITINERARIO	ID_ITINERARIO	Y
2	ID_AVION	ID_AVION	Y
3	COSTO	COSTO	Y

[Get update fields] [Edit mapping]

[Help] [OK] [Cancel] [SQL]

Se realiza el proceso ETL para cargar los datos al data mart

View Design

Search

- Transformations
 - IATA_ETL
 - Run configurations
 - Database connections
 - IATA
 - IATA_DM
 - Steps
 - Hops
 - Partition schemas
 - Slave server
 - Kettle cluster schemas
 - VFS Connections

Welcome! IATA_ETL

Avion_EXT → Avion_LOAD

Itinerario_EXT → Itinerario_LOAD

Vuelos_EXT → Vuelos_LOAD

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Pre

2024/06/02 21:15:01 - Spoon - Transformation opened.

2024/06/02 21:15:01 - Spoon - Launching transformation [IATA_ETL]...

2024/06/02 21:15:01 - Spoon - Started the transformation execution.

2024/06/02 21:15:01 - IATA_ETL - Dispatching started for transformation [IATA_ETL]

Verificación de la carga de los datos

The screenshot shows the SQL Developer interface. On the left, the 'Conexiones' pane displays the database schema hierarchy: 'Oracle conexiones' > 'IATA' > 'IATA_DM' > 'Tablas (Filtrado)' > 'DIM_AVION'. The 'DIM_AVION' table is selected. On the right, the 'Datos' tab shows the table's contents. The table has three columns: 'ID_AVION', 'MODELO', and 'AEROLINEA'. It contains 10 rows of data.

ID_AVION	MODELO	AEROLINEA
1	1 Airbus 320	Avianca
2	2 Airbus 320	Latam
3	3 Boeing 747	Avianca
4	4 Airbus 320	Wingo
5	5 Boeing 747	Avianca
6	6 Airbus 320	Avianca
7	7 Airbus 320	Latam
8	8 Boeing 747	Latam
9	9 Boeing 747	Latam
10	10 Boeing 747	Wingo

The screenshot shows the SQL Developer interface. On the left, the 'Conexiones' pane displays the database schema hierarchy: 'Oracle conexiones' > 'IATA' > 'IATA_DM' > 'Tablas (Filtrado)' > 'DIM_ITINERARIO'. The 'DIM_ITINERARIO' table is selected. On the right, the 'Datos' tab shows the table's contents. The table has four columns: 'ID_ITINERARIO', 'ANO', 'SEMESTRE', and 'CIUDAD'. It contains 15 rows of data.

ID_ITINERARIO	ANO	SEMESTRE	CIUDAD
1	1	2019	1 Santiago de Cali
2	2	2019	1 Bogotá
3	3	2019	1 Bogotá
4	4	2019	1 Medellín
5	5	2019	1 Medellín
6	6	2019	1 Medellín
7	7	2019	1 Medellín
8	8	2019	1 Bogotá
9	9	2019	1 Bogotá
10	10	2019	1 Bogotá
11	11	2019	1 Santiago de Cali
12	12	2019	1 Barcelona
13	13	2019	1 Santiago de Cali
14	14	2019	1 Santiago de Cali
15	15	2019	1 París

The screenshot displays the SQL Developer interface. On the left, the 'Conexiones' (Connections) pane shows the 'IATA_DM' database selected under 'Oracle conexiones'. The 'Tablas (Filtrado)' (Filtered Tables) folder is expanded, showing the 'FACT_VUELOS' table. Below this, the 'Buscar Objeto de Base de Datos' (Search Database Object) pane is visible. On the right, the 'FACT_VUELOS' table is open, showing its columns: ID_AVION, ID_ITINERARIO, and COSTO. The table contains 15 rows of data.

	ID_AVION	ID_ITINERARIO	COSTO
1	1	1	600000
2	1	1	650000
3	1	1	750000
4	7	2	3500500
5	7	2	3000550
6	2	5	4115450
7	9	6	4115450
8	9	6	600000
9	9	6	750000
10	6	7	3225000
11	10	8	4115450
12	5	9	2789000
13	5	9	3000000
14	3	10	3500500
15	2	11	750000

Visualizando el Data Mart y el Cubo OLAP con PowerBI

Primero se debe instalar el cliente Oracle para las herramientas de Microsoft disponible en el siguiente link: https://download.oracle.com/otn_software/odac/Oracle-Client-for-Microsoft-Tools-64-bit.exe

A continuación, se realiza la conexión con la base de datos de Oracle IATA_DM

Oracle Database

Servidor

localhost/XEPDB1

Modo Conectividad de datos ⓘ

☒ Importar

☐ DirectQuery

▶ Opciones avanzadas

Aceptar

Cancelar

Navegador

Opciones de presentación ▼

localhost/XEPDB1 [28]

▶ ANONYMOUS

▶ APPQOSSYS

▶ AUDSYS

▶ c##vladimir

▶ C##VLADIMIR [6]

☐ BIN\$fsJnopilRXmVKhZ25xcDrA==\$0

☐ BIN\$hclEZRxeTt+50cHUpeQKyw==\$0

☐ BIN\$Hi3Q/e/VTYi9pwwwCSL6aQ==\$0

☒ DIM_AVION

☒ DIM_ITINERARIO

☒ FACT_VUELOS

▶ DBSFUSER

▶ DBSNMP

▶ DGPDB_INT

▶ DIP

▶ DVF

▶ DVSYS

▶ GGSYS

▶ GSMADMIN_INTERNAL

FACT_VUELOS

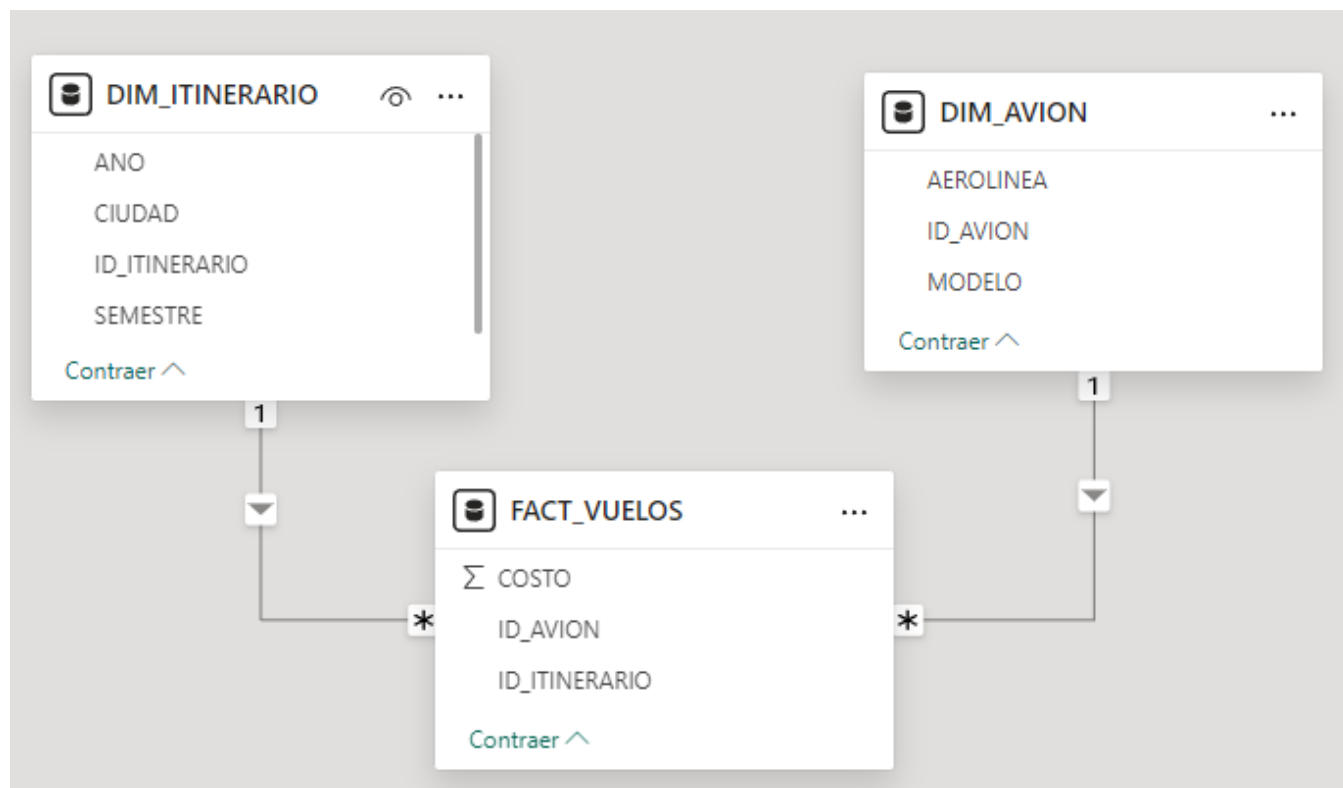
ID_AVION	ID_ITINERARIO	COSTO
1	1	600000
1	1	650000
1	1	750000
7	2	3500500
7	2	3000550
2	5	4115450
9	6	4115450
9	6	600000
9	6	750000
6	7	3225000
10	8	4115450
5	9	2789000
5	9	3000000
3	10	3500500
2	11	750000
1	12	1200000
5	13	3500500
2	14	3500500
4	15	1545000
10	16	3500500
7	17	3500500
8	18	1200000
9	19	4115450
1	20	2789000

Seleccionar tablas relacionadas

Cargar

Transformar datos

Cancelar



1. ¿Cuál aerolínea realizó el mayor número de vuelos a la ciudad de Roma en el año 2019 y cuál en el año 2020?

¿Cuál aerolínea realizó el mayor número de vuelos a la ciudad de Roma en el año 2019 y cuál en el año 2020?

ANO

2.019

2.020

CIUDAD

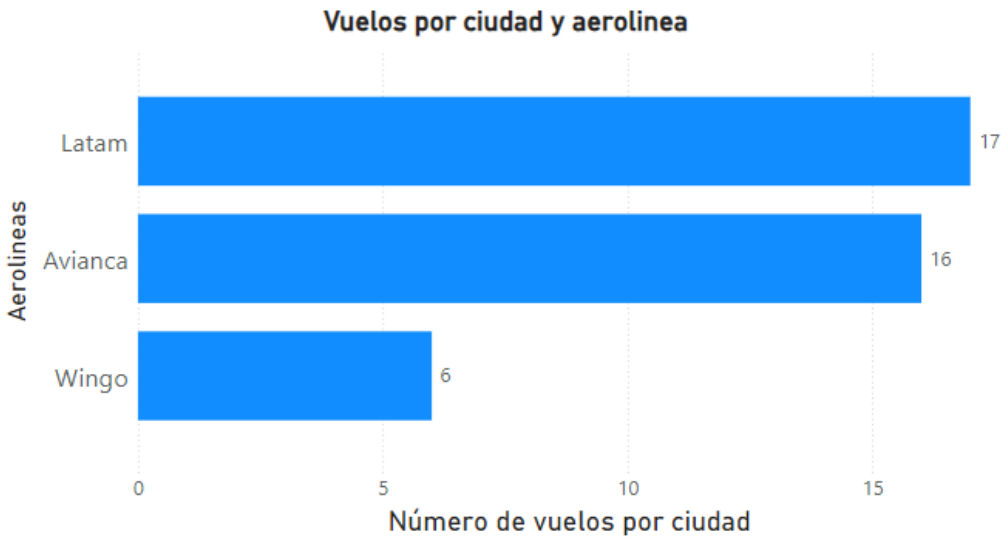
Barcelona

Bogotá

Medellín

París

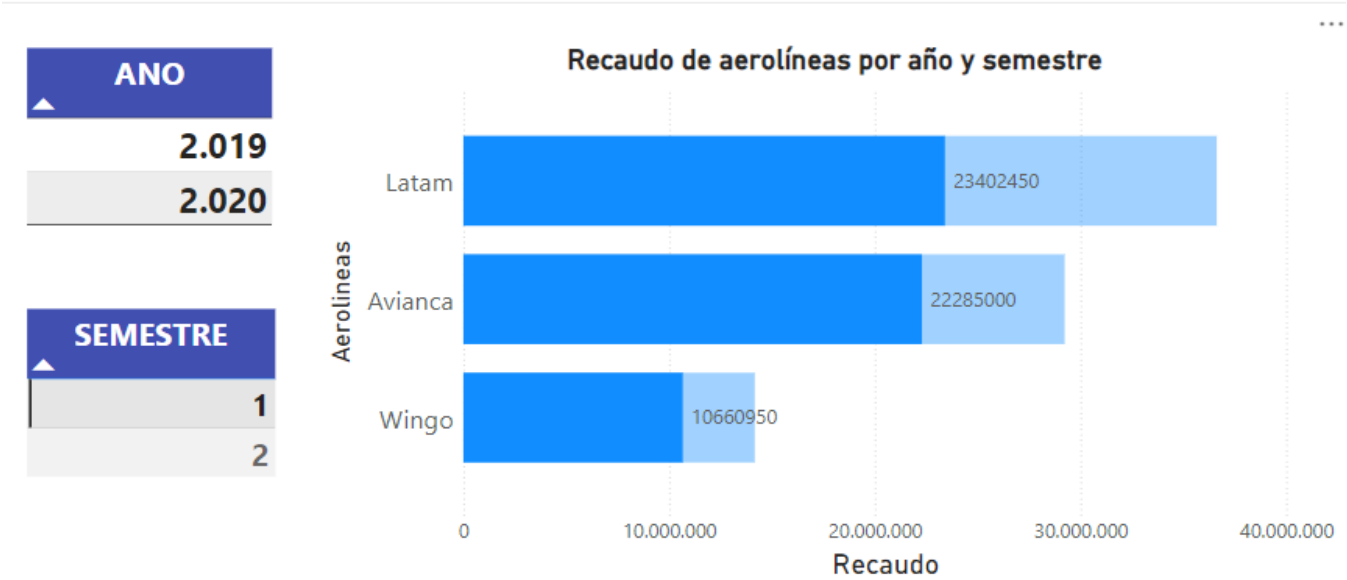
Santiago de Cali



La base de datos no presenta vuelos a la ciudad de roma durante los periodos 2019 y 2020

2. Total de dinero recaudado por vuelos de cada aerolínea en el primer semestre del año 2019 y en el primer semestre del año 2020.

Total de dinero recaudado por vuelos de cada aerolínea en el primer semestre del año 2019 y en el primer semestre del año 2020

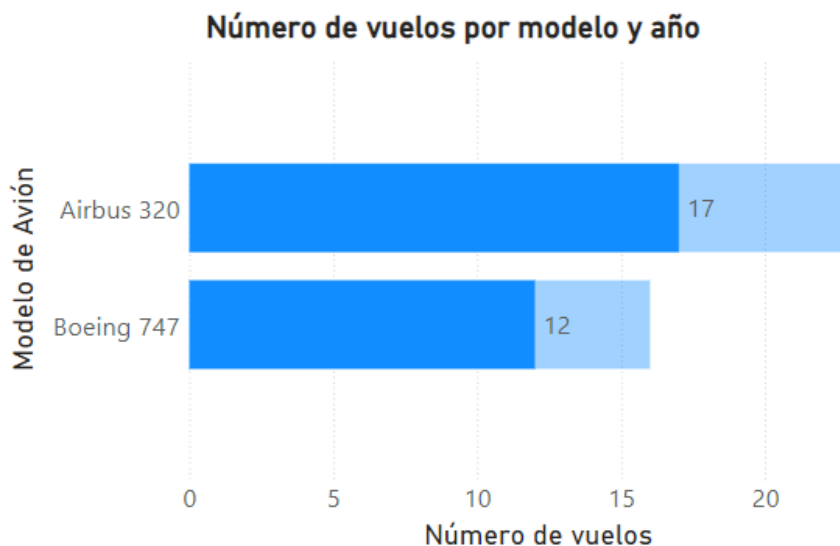


Durante el primer semestre del año 2019 y primer semestre del año 2020; Latam recaudo \$23.402.450, Avianca \$22.285.000 y Wingo \$10.660.950

3. ¿Cuál modelo de avión realizó el mayor número de vuelos en el año 2019 y cuál en el año 2020?

¿Cuál modelo de avión realizó el mayor número de vuelos en el año 2019 y cuál en el año 2020?

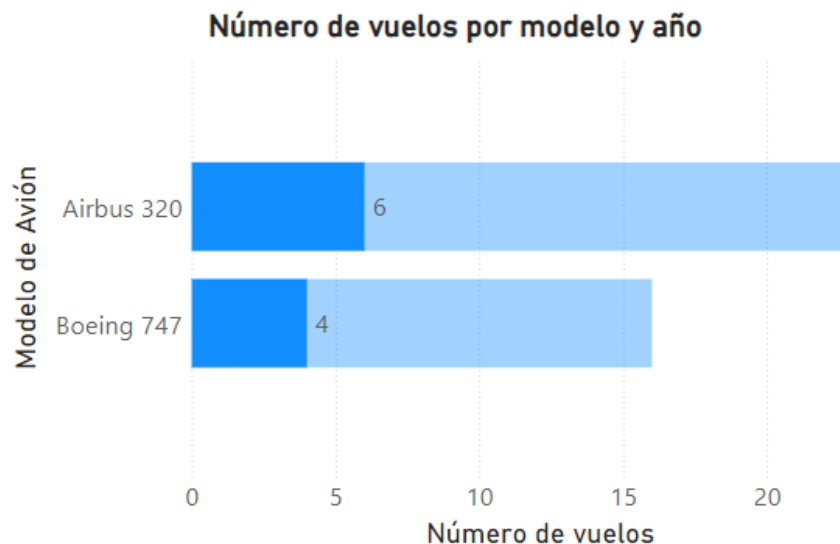
ANO
2.019
2.020



En el año 2019 los aviones de modelo Airbus 320 realizaron 17 vuelos

¿Cuál modelo de avión realizó el mayor número de vuelos en el año 2019 y cuál en el año 2020?

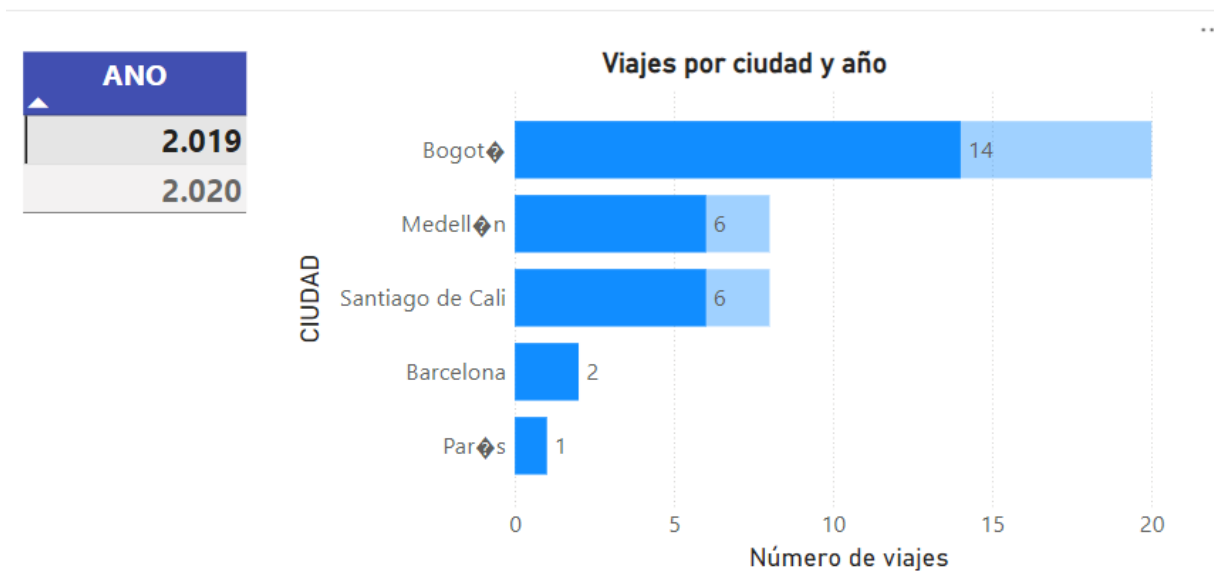
ANO
2.019
2.020



En el año 2020 los aviones de modelo Airbus 320 realizaron 6 vuelos

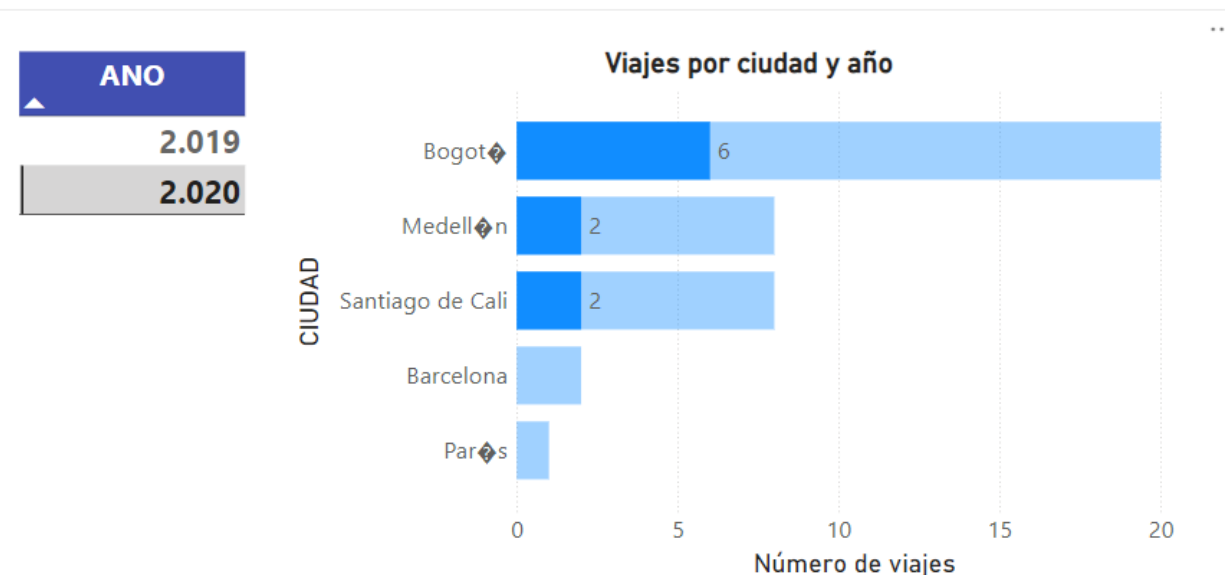
4. ¿Cuál fue la ciudad cuyos habitantes viajaron más en el año 2019 y cuál en el año 2020?

¿Cuál fue la ciudad cuyos habitantes viajaron más en el año 2019 y cuál en el año 2020?



Durante el año 2019 la ciudad cuyos habitantes viajaron mas durante este año fue Bogotá con un total de 14 viajes

¿Cuál fue la ciudad cuyos habitantes viajaron más en el año 2019 y cuál en el año 2020?



Durante el año 2020 la ciudad cuyos habitantes viajaron más durante este año fue Bogotá con un total de 6 viajes

Conclusión

El análisis crítico-contextual y la construcción del Modelo Multidimensional propuesto permiten a la IATA obtener una visión clara del impacto de la pandemia en el transporte aéreo. La implementación del Modelo Estrella facilita la realización de consultas complejas y la generación de informes detallados, contribuyendo a una toma de decisiones informada y estratégica.