

Ejercicio DataMart

José Julián Barbosa Ayala
Michael Hernández Vera
Camilo Vega Ramírez

Contenido

| | |
|---|----------|
| Logica creación modelo estrella. | 2 |
| Modelo relacional original | 2 |
| Requerimientos de analisis. | 3 |
| Análisis de los hechos. | 3 |
| Análisis de las dimensiones. | 3 |
| Análisis de los indicadores. | 3 |
| Análisis de las jerarquías. | 3 |
| Diagrama del Modelo. | 3 |
| Construcción Data Mart. | 5 |
| Construcción Cubo OLAP | 5 |
| Proceso de ETL | 6 |
| Solución Requerimientos de analisis. | 6 |
| ¿Cuál aerolínea realizó el mayor número de vuelos a la ciudad de Roma en el año 2019 y cuál en el año 2020? | 9 |
| Total de dinero recaudado por vuelos de cada aerolínea en el primer semestre del año 2019 y en el primer semestre del año 2020. | 9 |
| ¿Cuál modelo de avión realizó el mayor número de vuelos en el año 2019 y cuál en el año 2020? . . | 10 |
| ¿Cuál fue la ciudad cuyos habitantes viajaron más en el año 2019 y cuál en el año 2020? | 10 |

Logica creación modelo estrella.

A continuación se explica de forma breve el proceso para llegar al modelo estrella seleccionado.

Modelo relacional original

En la figura 1 podemos ver el modelo relacional original.

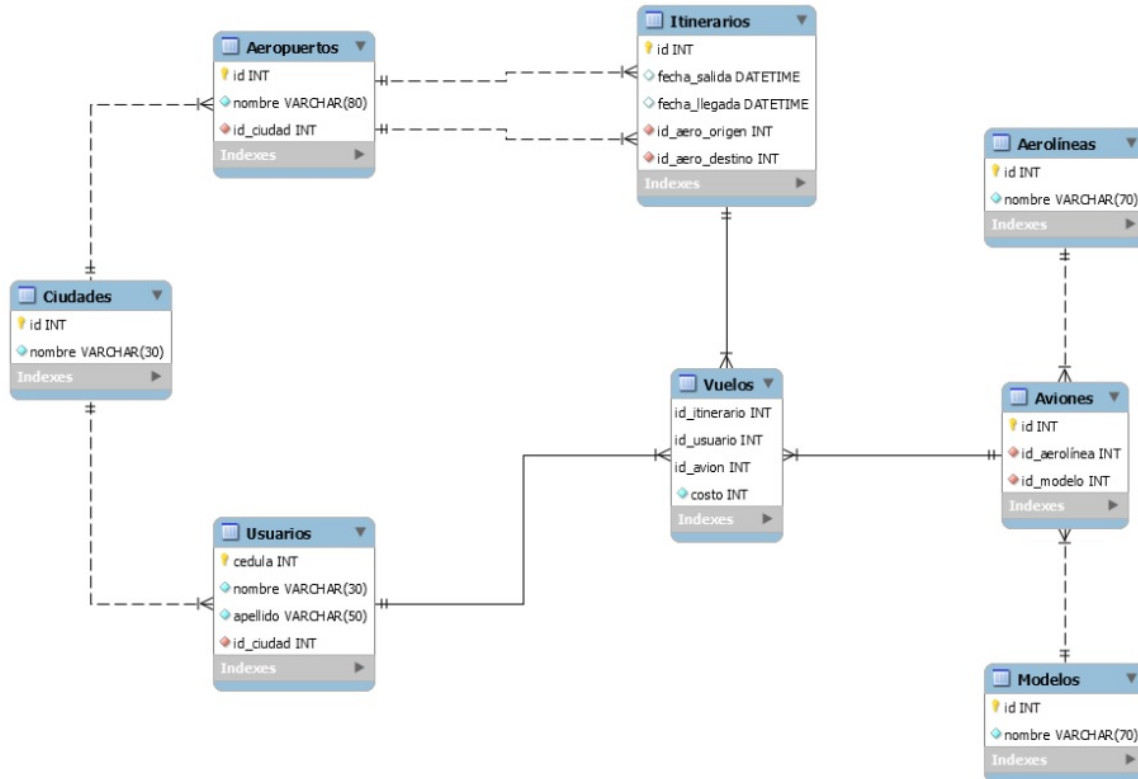


Figure 1: Modelo relacional original

Del mismo vemos:

- Los vuelos se relacionan con los itinerarios donde cada itinerario puede contar con varios vuelos.
- Los vuelos se relacionan con los aviones donde cada avión puede contar con varios vuelos.
- Los vuelos se relacionan con los usuarios donde cada usuario puede realizar varios vuelos.
- Los aviones se relacionan con los modelos donde cada modelo puede tener varios aviones.
- Los aviones se relacionan con las aerolíneas donde cada aerolínea puede tener varios aviones.
- Los Usuarios se relacionan con las ciudades donde cada ciudad puede tener varios usuarios.
- Los itinerarios se relacionan con los aeropuertos donde cada aeropuerto puede tener varios itinerarios de origen o varios itinerarios de destino.
- Los aeropuertos se relacionan con las ciudades donde cada ciudad puede tener varios aeropuertos.

Se entiende que los vuelos hacen referencia a cada pasaje vendido a un usuario.

Requerimientos de analisis.

Las siguientes son las preguntas a las cuales se quiere dar respuesta con este analisis:

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

Análisis de los hechos.

De los requerimientos de analisis podemos ver que los hechos buscados estan relacionados con el conteo de vuelos y con el costo total de los vuelos, por lo cual la tabla principal de donde nos podemos basar es la de VUELOS.

Análisis de las dimensiones.

De las pregunta de negocio se idenfician las dimensiones *aerolinea*, *ciudad de destino*, *fecha*, *modelo de avión*, *ciudad usuario*. Sobre estas cinco dimensiones se debe modelar para poder sacar de los hechos las cantidades y total costo.

Análisis de los indicadores.

Las dos medidas solicitadas son *cantidades* entendidas como el condeo de vuelos, y *total costo* entendido como la sumatorio del campo *costo* de cada vuelo.

Análisis de las jerarquías.

Teniendo en cuenta las preguntas y las dimensiones, con la excepción de la *fecha* que cuenta con los niveles fecha, año y semestre, nuestras demas jerarquias solo cuentan con un nivel de la siguiente forma.

- AEROLINEAS
- CIUDAD DESTINO
- CIUDAD USUARIO
- FECHAS -> AÑO -> SEMESTRE
- MODELOS

Diagrama del Modelo.

Teniendo en cuenta todo el analisis previo, nuestro diagrama de estrella cuenta en el centro con la tabla de hechos *VUELOS* con las medidas *COSTO* y *CANTIDAD* y los ids para relacionarlos con las dimensiones *AEROLINEAS*, *CIUDAD_DESTINO*, *CIUDAD_USUARIO*, *FECHA* y *MODELOS* como se ve en la figura 2.

Cabe notar que este modelo es creado a la medida para contestar a las preguntas de interes del presente caso.

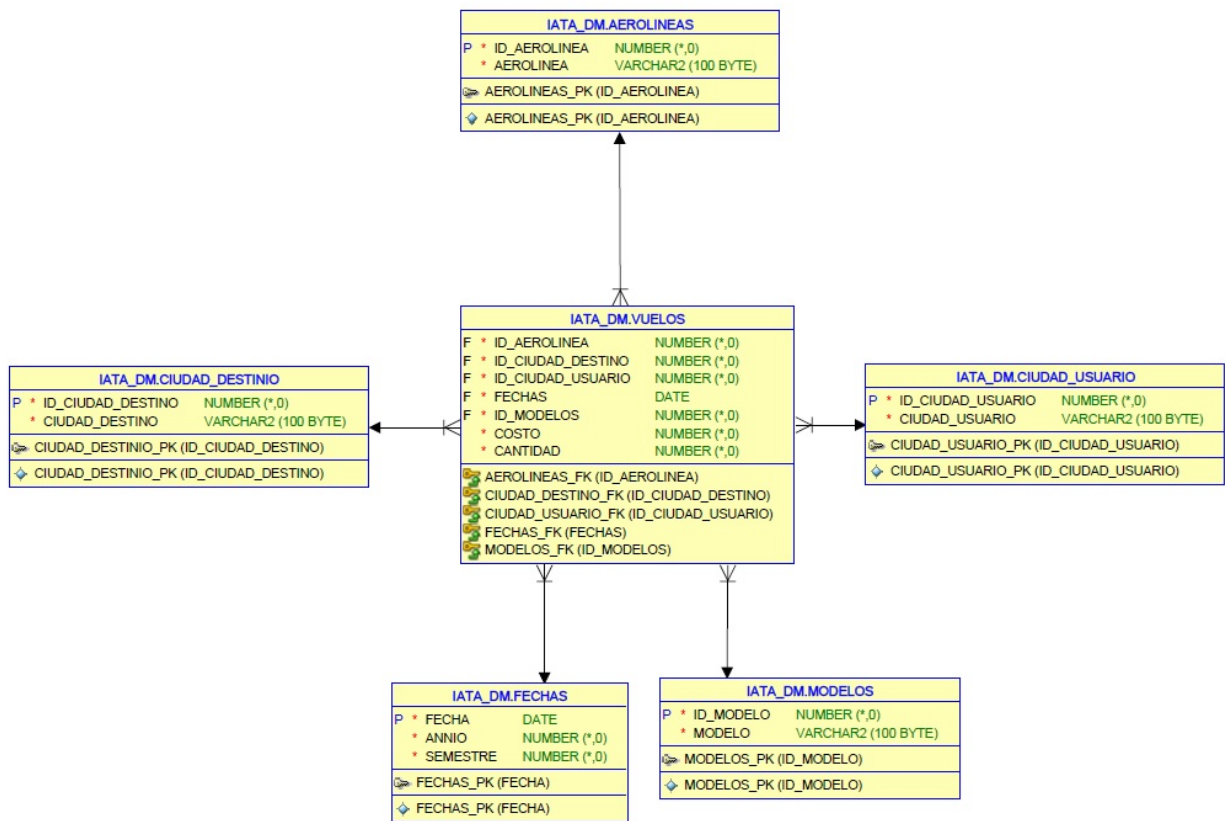


Figure 2: Modelo Estrella Datamart

Construcción Data Mart.

Para la construcción del Data Mart se usó Sql Developer construyendo el esquema IATA_DM y las tablas y columnas como se aprecian en la figura 3

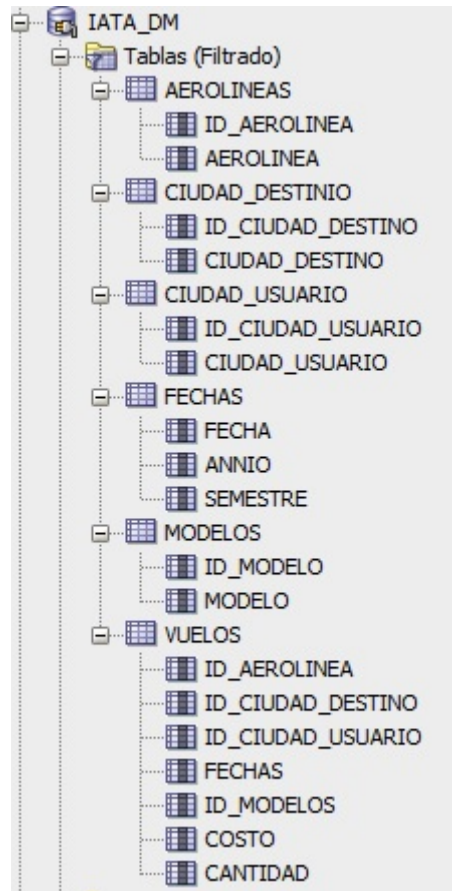


Figure 3: Data Mart SQL Developer

Se crean las llaves principales * AEROLINEAS_PK * CIUDAD_DESTINO_PK * CIUDAD_USUARIO_PK * FECHAS_PK * MODELOS_PK

Igualmente se crean las llaves foraneas * AEROLINEAS_FK * CIUDAD_DESTINO_FK * CIUDAD_USUARIO_FK * FECHAS_FK * MODELOS_FK

Las relaciones entre las tablas las llaves foraneas y llaves principales, así como el tipo de datos se puede apreciar en la gráfica relacional del modelo estrella expuesto con anterioridad en la figura 2.

Construcción Cubo OLAP

Para la construcción del cubo olap se usó la herramienta Pentaho Workbench. Se creó el cubo IATA_CUBE_GRUPO4A, a la cual se adiciona la tabla de hechos VUELOS, las dimensiones AEROLINEAS con la jerarquía AEROLINEA y la tabla AEROLINEAS, CIUDAD_DESTINO con la jerarquía CIUDAD_DESTINO y la tabla CIUDAD_DESTINO, CIUDAD_USUARIO con la jerarquía CIUDAD_USUARIO y la tabla CIUDAD_USUARIO, FECHAS con la jerarquías ANNIO, SEMESTRE y la tabla FECHAS, MODELOS con la jerarquía MODELOS y la tabla MODELOS, así mismo las medidas

COSTO que es la sumatoria de la columna COSTO de la tabla de hechos y la medida CANTIDA que es el conteo de la tabla CANTIDAD de la tabla de hechos. La estructura del cubo OLAP la podemos apreciar en la figura 4.

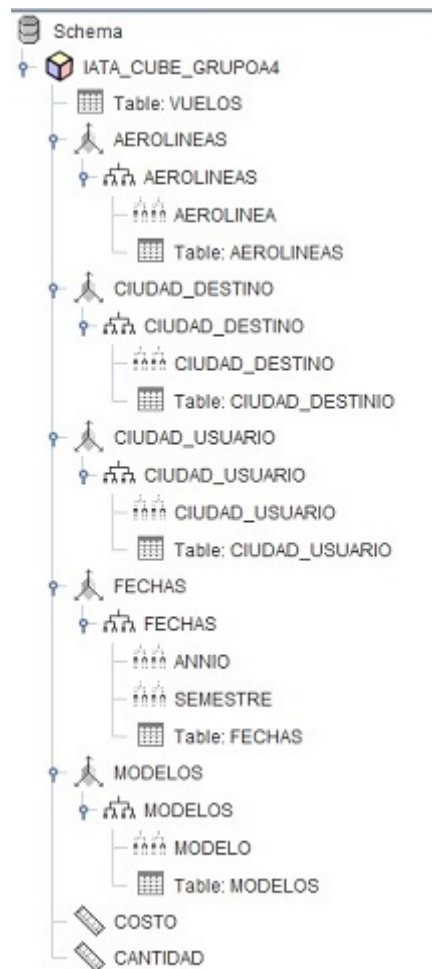


Figure 4: Cubo Olap IATA_CUBE_GRUPOA4

Proceso de ETL

El ETL se realizó utilizando la herramienta Pentaho Integration Services, para eso se realizó un flujo por cada uno de las tablas de dimensiones y tabla de hechos del Data Mart donde se creó un Input correspondiente a un query de SQL que extrae los datos de la base IATA y los transforma y los deposita en el Data Mart IATA_DM como se ve en la figura 5.

Las queries utilizadas se pueden ver en las figuras 6 a 10.

Solución Requerimientos de analisis.

Para dar respuesta a los requerimientos de analisis se usó la herramienta LinceBI a la cual se conectó el Data Mart IATA_DM y el cubo IATA_CUBE_GRUPOA4 para visualizar los resultados.

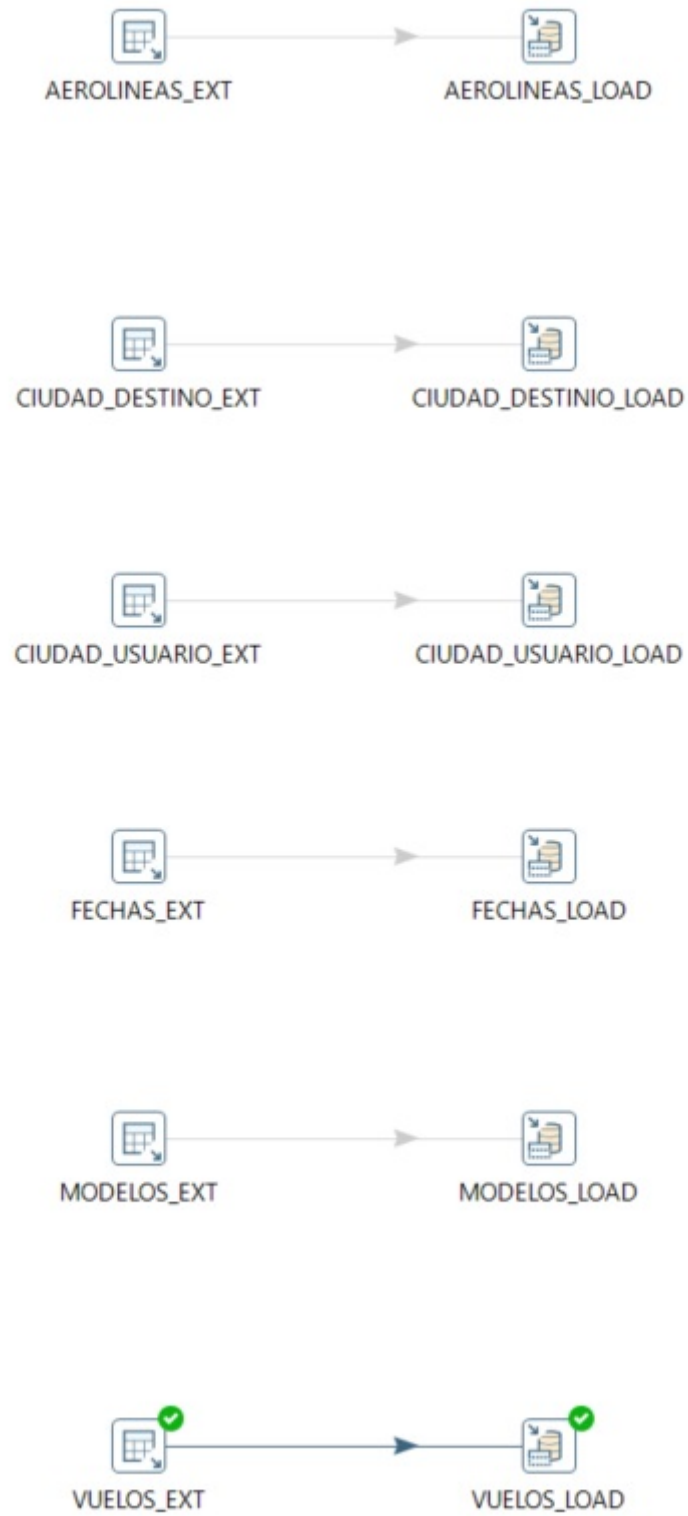


Figure 5: Flujo ETL IATA a IATA_dm

```

SELECT
    ID_AEROLINEA,
    NOMBRE AS AEROLINEA
FROM AEROLINEAS

```

Figure 6: Query AEROLINEAS_EXT

```

SELECT DISTINCT
    AEROPUERTOS.ID_CIUADAD AS ID_CIUADAD_DESTINO,
    CIUDADES.NOMBRE AS CIUADAD_DESTINO
FROM AEROPUERTOS
LEFT JOIN CIUDADES
    ON AEROPUERTOS.ID_CIUADAD = CIUDADES.ID_CIUADAD

```

Figure 7: Query CIUADAD_DESTINO_EXT

```

SELECT DISTINCT
    AEROPUERTOS.ID_CIUADAD AS ID_CIUADAD_DESTINO,
    CIUDADES.NOMBRE AS CIUADAD_DESTINO
FROM AEROPUERTOS
LEFT JOIN CIUDADES
    ON AEROPUERTOS.ID_CIUADAD = CIUDADES.ID_CIUADAD

```

Figure 8: Query CIUADAD_USUARIO_EXT

```

SELECT DISTINCT
    TRUNC(FECHA_SALIDA) AS FECHA,
    CAST(EXTRACT(YEAR FROM FECHA_SALIDA) AS INT) AS ANNIO,
    TO_CHAR(FECHA_SALIDA, 'YYYY') || ' SEM ' ||
        CASE
            WHEN TO_NUMBER(TO_CHAR(FECHA_SALIDA, 'MM')) <= 6 THEN '1'
            ELSE '2'
        END AS SEMESTRE
FROM ITINERARIOS

```

Figure 9: Query FECHAS_EXT

```

SELECT
    AVIONES.ID_AEROLINEA,
    AEROPUERTOS.ID_CIUADAD AS ID_CIUADAD_DESTINO,
    USUARIOS.ID_CIUADAD AS ID_CIUADAD_USUARIO,
    TRUNC(ITINERARIOS.FECHA_SALIDA) AS FECHAS,
    AVIONES.ID_MODELO AS ID_MODELOS,
    CAST(VUELOS.COSTO AS INT) AS COSTO,
    CAST(1 AS INT) AS CANTIDAD
FROM VUELOS
LEFT JOIN AVIONES
    ON VUELOS.ID_AVION = AVIONES.ID_AVION
LEFT JOIN ITINERARIOS
    ON VUELOS.ID_ITINERARIO = ITINERARIOS.ID_ITINERARIO
LEFT JOIN AEROPUERTOS
    ON ITINERARIOS.ID_AEROPUERTO_DESTINO = AEROPUERTOS.ID_AEROPUERTO
LEFT JOIN USUARIOS
    ON VUELOS.ID_USUARIO = USUARIOS.CEDULA

```

Figure 10: VUELOS_EXT

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

En la figura 11 podemos ver la visualización creada para dar respuesta al requerimiento.

STOLap View x

OLAP MDX CHART TABLE MISC

| | CIUDAD_DESTINO | FECHAS | AEROLINEAS | Measures |
|--|----------------|--------|------------|----------|
| | Roma | + 2019 | Avianca | 11 |
| | | | Wingo | 4 |

Figure 11: Punto 1

La aerolínea que realizó el mayor número de vuelos a la ciudad de Roma en el año 2019 fue Avianca, para 2020 los datos suministrados no contienen vuelos a la ciudad de Roma.

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

En la figura 12 podemos ver la visualización creada para dar respuesta al requerimiento.

OLAP MDX CHART TABLE MISC

| | FECHAS | AEROLINEAS | Measures |
|--|------------|------------|------------|
| | 2019 SEM 1 | Avianca | 90,089,500 |
| | | Latam | 81,830,300 |
| | | Wingo | 46,104,600 |
| | 2020 SEM 1 | Avianca | 9,890,000 |
| | | Latam | 14,530,000 |
| | | Wingo | 6,000,000 |

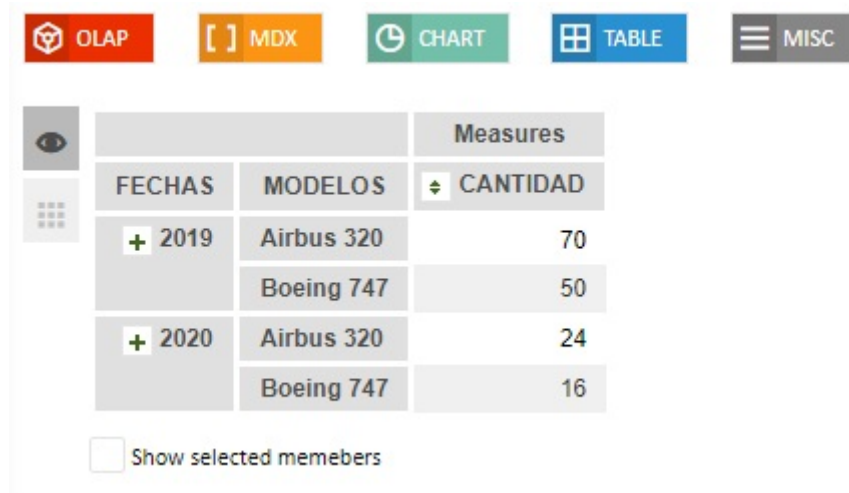
☐ Show selected members

Figure 12: Punto 2

De la misma tabla de la figura 12 podemos ver los ingresos por recaudos de cada una de las aerolíneas para el primer semestre del año 2019 y 2020.

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

En la figura 13 podemos ver la visualización creada para dar respuesta al requerimiento.



The screenshot shows a BI tool interface with a top navigation bar containing buttons for OLAP, MDX, CHART, TABLE, and MISC. The main area displays a table with the following data:

| | | Measures |
|--------|------------|----------|
| FECHAS | MODELOS | CANTIDAD |
| + 2019 | Airbus 320 | 70 |
| | Boeing 747 | 50 |
| + 2020 | Airbus 320 | 24 |
| | Boeing 747 | 16 |

Below the table, there is a checkbox labeled "Show selected members" which is currently unchecked.

Figure 13: Punto 3

En tanto en el año 2019 como en el 2020, el modelo de avión que realizó el mayor número de vuelos fue el Airbus 320.

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

En la figura 14 podemos ver la visualización creada para dar respuesta al requerimiento.



The screenshot shows a BI tool interface with a top navigation bar containing buttons for OLAP, MDX, CHART, TABLE, and MISC. The main area displays a table with the following data:

| | | Measures |
|--------|------------------|----------|
| FECHAS | CIUDAD_USUARIO | CANTIDAD |
| + 2019 | Medellín | 36 |
| | Santiago de Cali | 30 |
| | Bogotá | 30 |
| | Roma | 12 |
| | Nueva York | 6 |
| | Londres | 6 |
| | | |
| + 2020 | Medellín | 12 |
| | Santiago de Cali | 10 |
| | Bogotá | 10 |
| | Roma | 4 |
| | Nueva York | 2 |
| | Londres | 2 |

Below the table, there is a checkbox labeled "Show selected members" which is currently unchecked.

Figure 14: Punto 4

En tanto en el año 2019 como en el 2020 los habitantes de la ciudad de Medellin fueron los que realizaron el mayor número de vuelos.