Index

[Data Warehouse SQL 1](#_Toc118211972)

[Transformations 3](#_Toc118211973)

[dim\_airline 3](#_Toc118211974)

[dim\_airplane 4](#_Toc118211975)

[dim\_arrival 5](#_Toc118211976)

[dim\_departure 6](#_Toc118211977)

[dim\_airport 7](#_Toc118211978)

[fact\_flight 8](#_Toc118211979)

[Jobs 10](#_Toc118211980)

[load\_dw 10](#_Toc118211981)

[load\_fact 10](#_Toc118211982)

[Cube Definition 11](#_Toc118211983)

[Saiku Analysis 14](#_Toc118211984)

[airports.sql 14](#_Toc118211985)

[airports-large.sql 17](#_Toc118211989)

[airports-large-extra.sql 20](#_Toc118211993)

[Comparisons 23](#_Toc118211997)

# 

# **Data Warehouse SQL**

***airports\_dw.sql***

DROP DATABASE IF EXISTS airports\_dw;

CREATE DATABASE airports\_dw;

USE airports\_dw;

CREATE TABLE dim\_airport (

    AIRPORT\_ID INT,

    AIRPORT\_NAME VARCHAR(255),

    CITY VARCHAR(255),

    COUNTRY VARCHAR(255),

    PRIMARY KEY (AIRPORT\_ID)

);

CREATE TABLE dim\_departure (

    TIME\_ID DATETIME,

    YEAR\_ID INT,

    MONTH\_ID INT,

    MONTH\_NAME VARCHAR(255),

    DAY\_ID INT,

    PRIMARY KEY (TIME\_ID)

);

CREATE TABLE dim\_arrival (

    TIME\_ID DATETIME,

    YEAR\_ID INT,

    MONTH\_ID INT,

    MONTH\_NAME VARCHAR(255),

    DAY\_ID INT,

    PRIMARY KEY (TIME\_ID)

);

CREATE TABLE dim\_airplane (

    AIRPLANE\_ID INT,

    AIRPLANE\_TYPE INT,

    PRIMARY KEY (AIRPLANE\_ID)

);

CREATE TABLE dim\_airline (

    AIRLINE\_ID INT,

    AIRLINE\_NAME VARCHAR(255),

    PRIMARY KEY (AIRLINE\_ID)

);

CREATE TABLE fact\_flight (

    FLIGHT\_ID INT,

    TOTALBOOKINGS INT,

    REVENUE DECIMAL(65,2),

    AIRLINE\_ID INT,

    AIRPLANE\_ID INT,

    ORIGIN\_ID INT,

    DESTINATION\_ID INT,

    DEPARTURE DATETIME,

    ARRIVAL DATETIME,

    PRIMARY KEY (FLIGHT\_ID),

*FOREIGN KEY* (AIRLINE\_ID) *REFERENCES* dim\_airline (AIRLINE\_ID),

*FOREIGN KEY* (AIRPLANE\_ID) *REFERENCES* dim\_airplane (AIRPLANE\_ID),

*FOREIGN KEY* (ORIGIN\_ID) *REFERENCES* dim\_airport (AIRPORT\_ID),

*FOREIGN KEY* (DESTINATION\_ID) *REFERENCES* dim\_airport (AIRPORT\_ID),

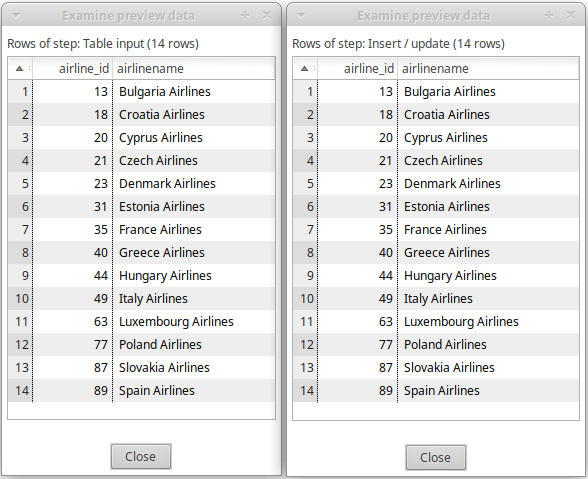
*FOREIGN KEY* (DEPARTURE) *REFERENCES* dim\_departure (TIME\_ID),

*FOREIGN KEY* (ARRIVAL) *REFERENCES* dim\_arrival (TIME\_ID)

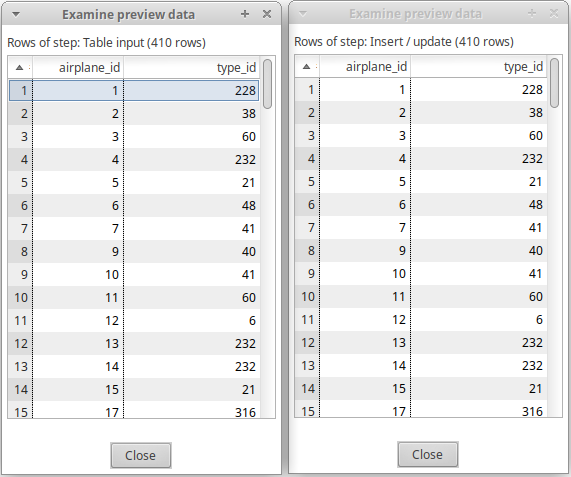
)

# **Transformations**

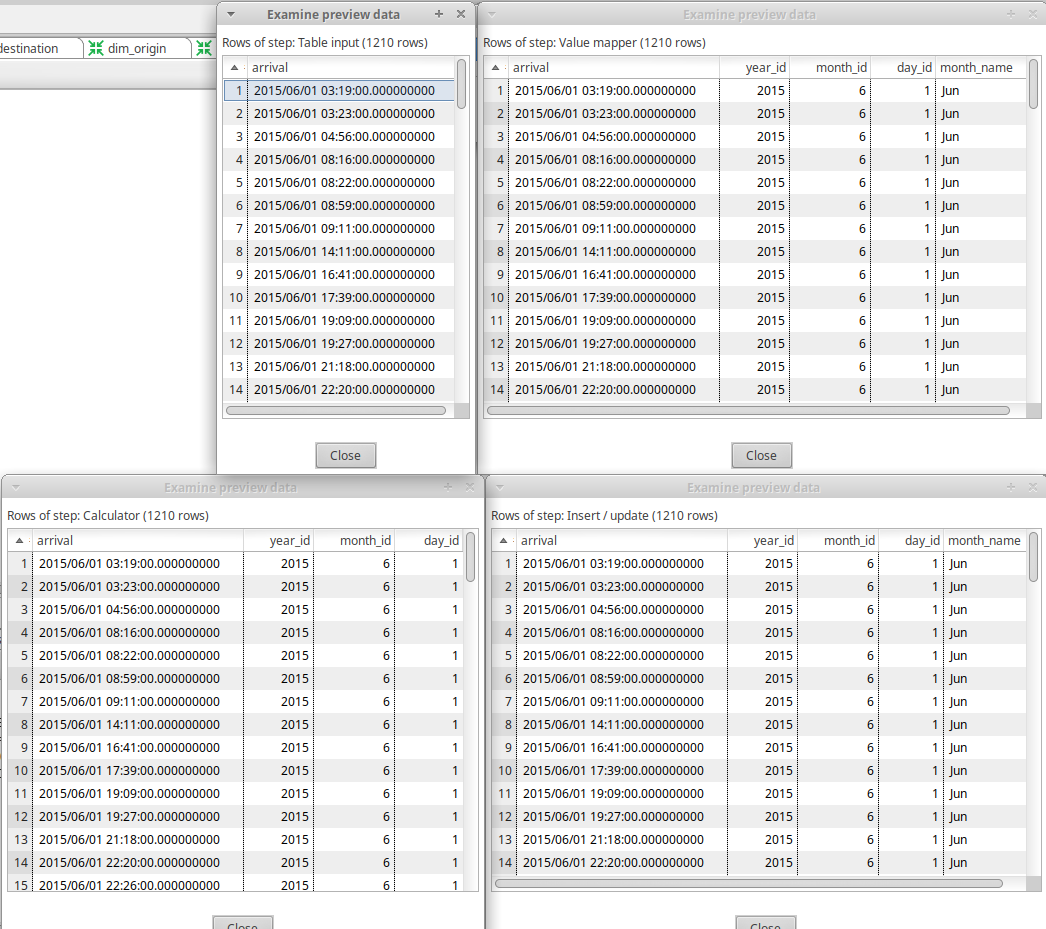
## **Graphical user interface, text, application, email Description automatically generateddim\_airline**



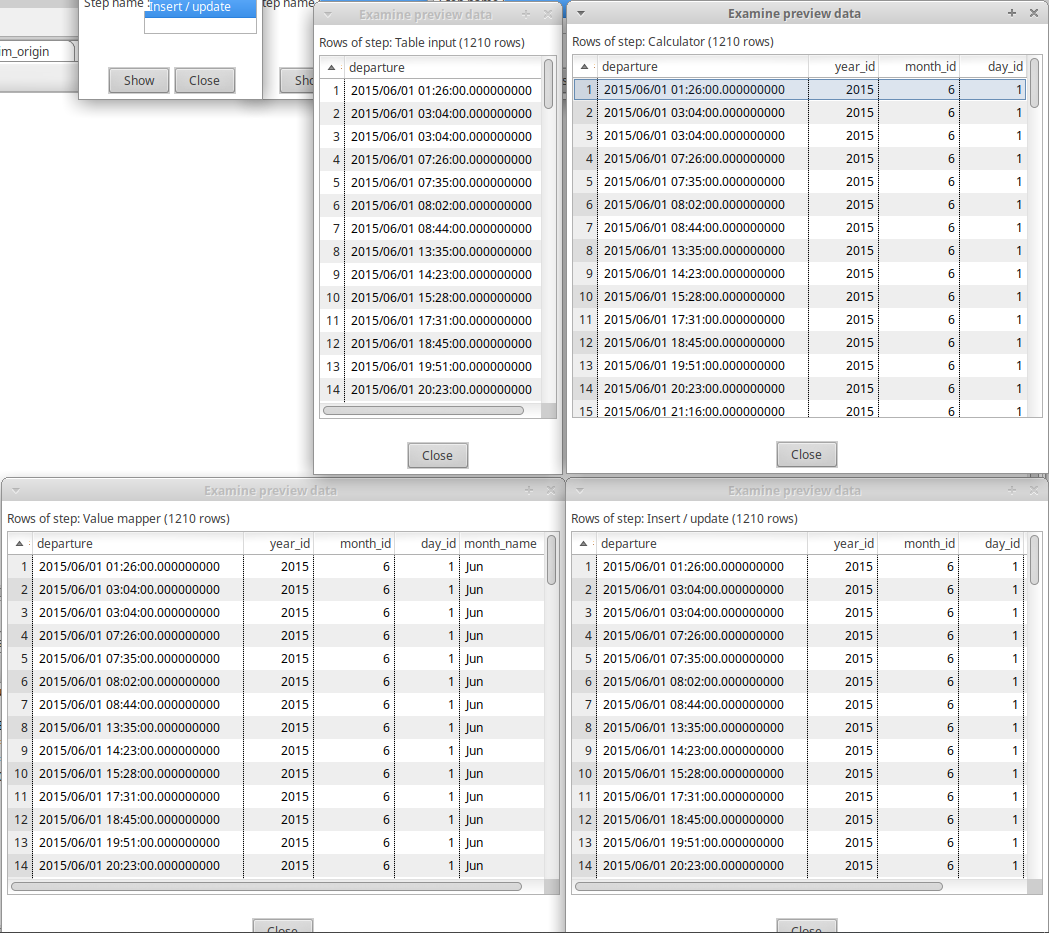
## **dim\_airplane**



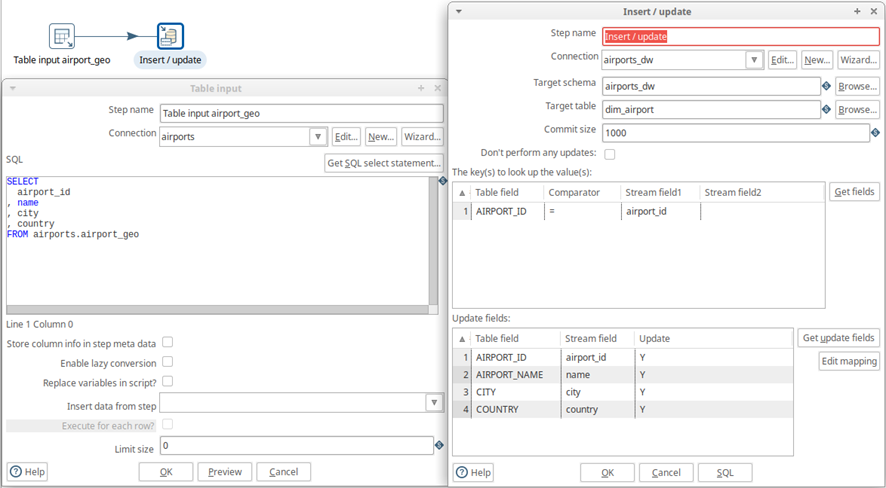
## **dim\_arrival**

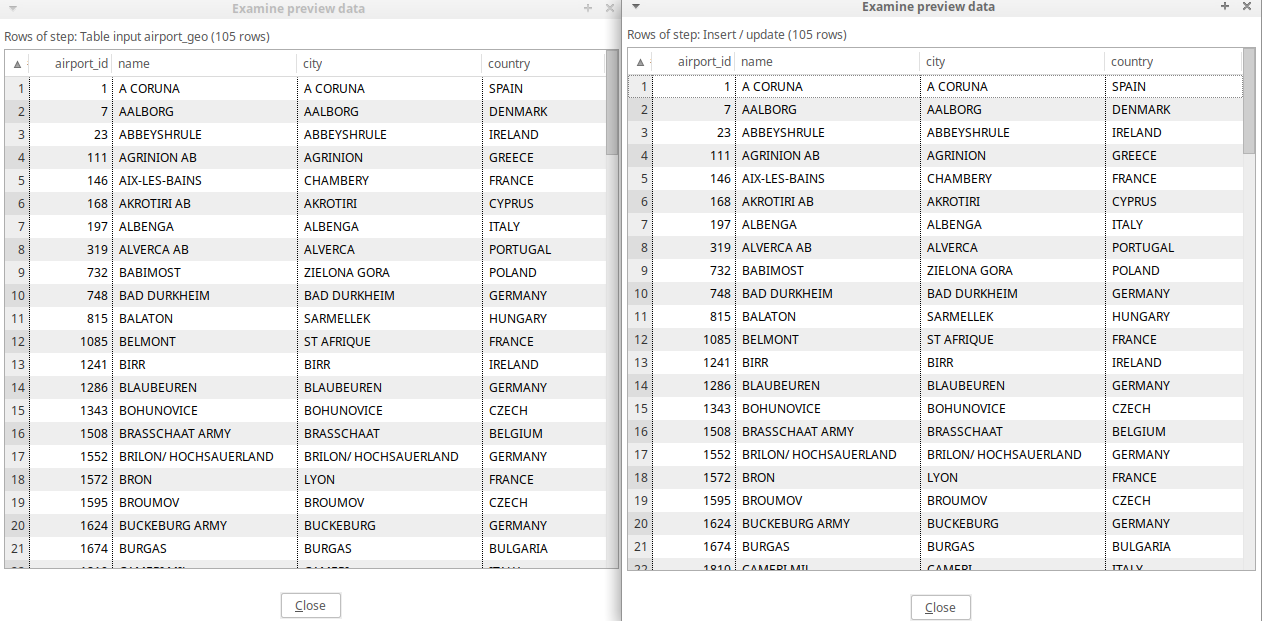


## **dim\_departure**

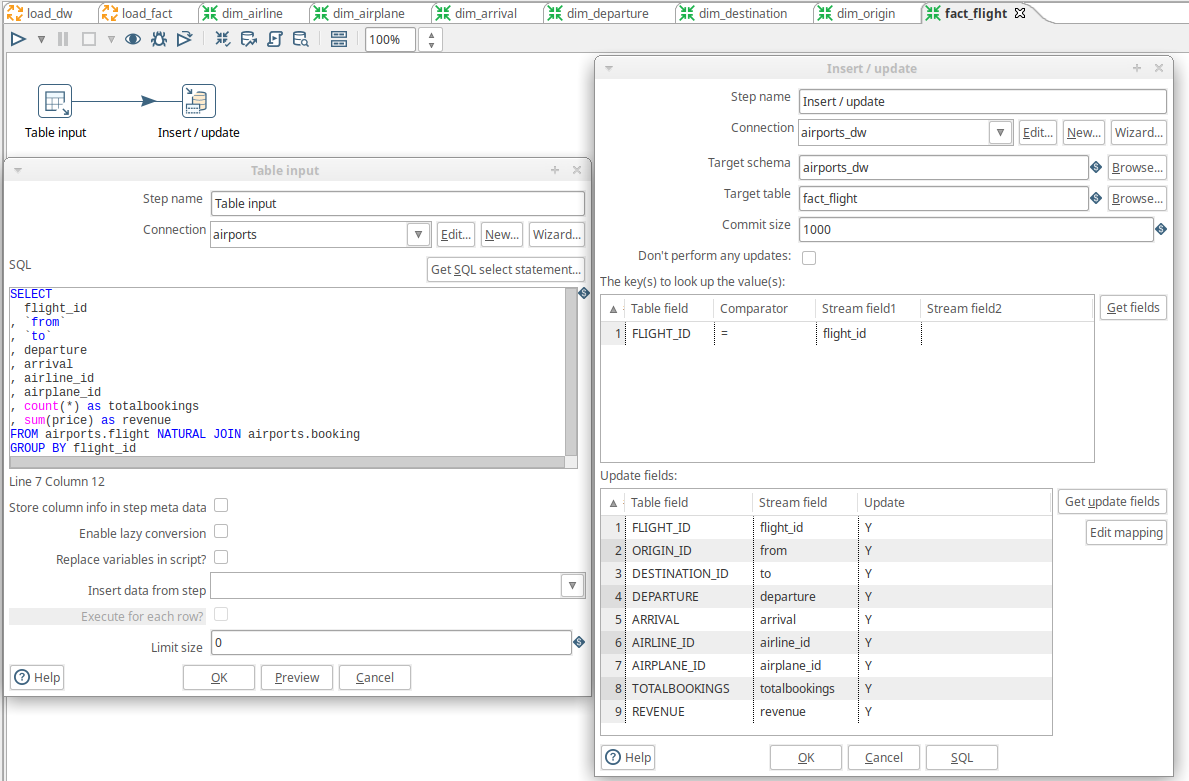


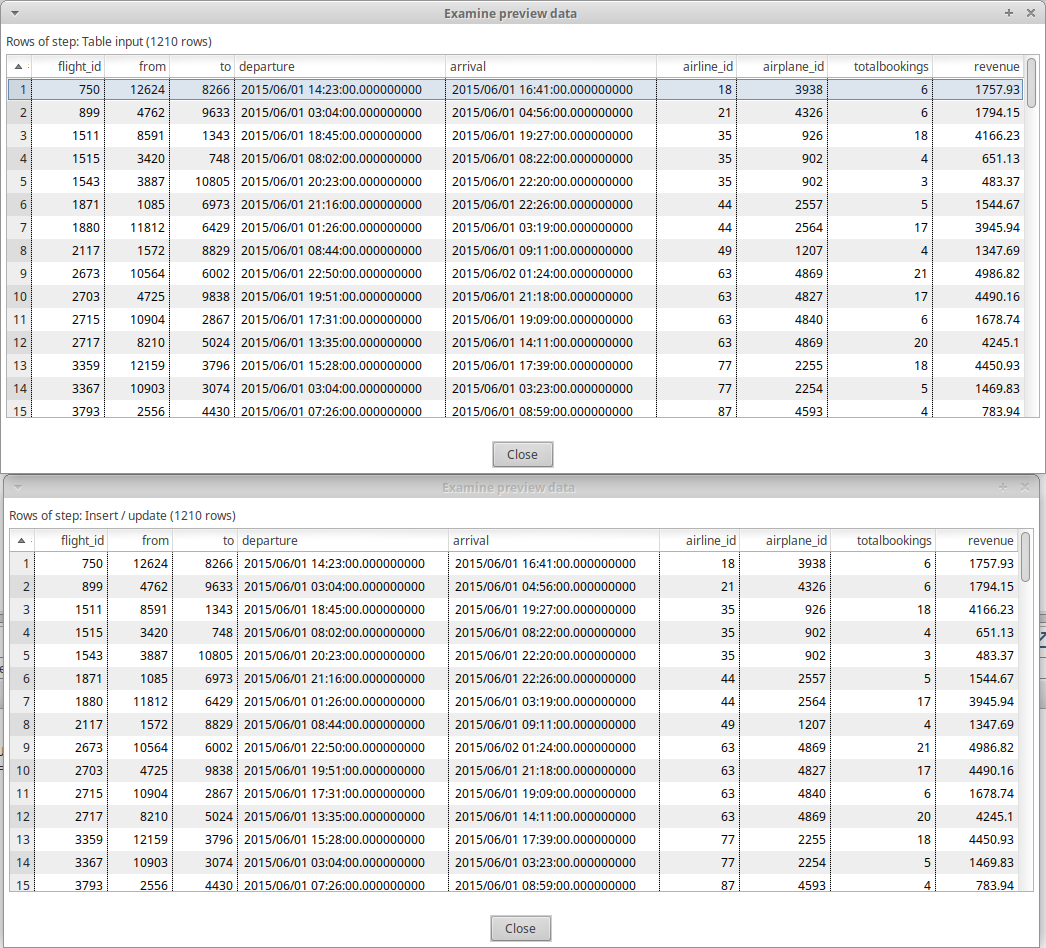
## **dim\_airport**

(Note: *dim\_origin* and *dim\_destination* were replaced by *dim\_airport.*)



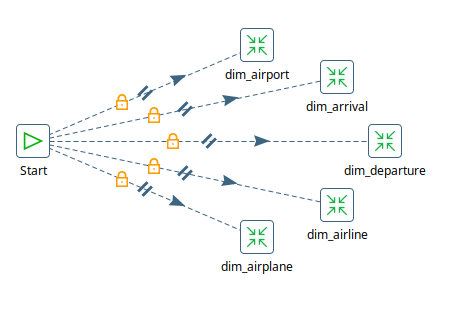
## **fact\_flight**



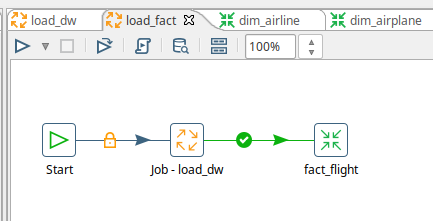


# **Jobs**

**load\_dw** – Loads dimensions in parallel.



**load\_fact** – Loads *fact\_flight* table after all dimensions are loaded.



# **Cube Definition**

<Schema name="airports\_dw">

  <Cube name="Flights" visible="true" cache="true" enabled="true">

    <Table name="fact\_flight">

    </Table>

    <Dimension type="StandardDimension" visible="true" foreignKey="ORIGIN\_ID" highCardinality="false" name="Origin">

      <Hierarchy name="Origin Hierarchy" visible="true" hasAll="true" allMemberName="AllOrigins" primaryKey="AIRPORT\_ID">

        <Table name="dim\_airport">

        </Table>

        <Level name="Country" visible="true" column="COUNTRY" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">

        </Level>

        <Level name="City" visible="true" column="CITY" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">

        </Level>

        <Level name="Airport Name" visible="true" column="AIRPORT\_NAME" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">

        </Level>

      </Hierarchy>

    </Dimension>

    <Dimension type="StandardDimension" visible="true" foreignKey="DESTINATION\_ID" highCardinality="false" name="Destination">

      <Hierarchy name="Destination Hierarchy" visible="true" hasAll="true" allMemberName="All Destinations" primaryKey="AIRPORT\_ID">

        <Table name="dim\_airport">

        </Table>

        <Level name="Country" visible="true" column="COUNTRY" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">

        </Level>

        <Level name="City" visible="true" column="CITY" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">

        </Level>

        <Level name="Airport Name" visible="true" column="AIRPORT\_NAME" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">

        </Level>

      </Hierarchy>

    </Dimension>

    <Dimension type="TimeDimension" visible="true" foreignKey="DEPARTURE" highCardinality="false" name="Departure">

      <Hierarchy name="Departure Hierarchy" visible="true" hasAll="true" allMemberName="All Departures" primaryKey="TIME\_ID">

        <Table name="dim\_departure">

        </Table>

        <Level name="Year" visible="true" column="YEAR\_ID" type="Integer" uniqueMembers="false" levelType="TimeYears" hideMemberIf="Never">

        </Level>

        <Level name="Months" visible="true" column="MONTH\_NAME" ordinalColumn="MONTH\_ID" type="String" uniqueMembers="false" levelType="TimeMonths" hideMemberIf="Never">

        </Level>

        <Level name="Days" visible="true" column="DAY\_ID" type="Integer" uniqueMembers="false" levelType="TimeDays" hideMemberIf="Never">

        </Level>

      </Hierarchy>

    </Dimension>

    <Dimension type="TimeDimension" visible="true" foreignKey="ARRIVAL" highCardinality="false" name="Arrival">

      <Hierarchy name="Arrival Hierarchy" visible="true" hasAll="true" allMemberName="All Arrivals" primaryKey="TIME\_ID">

        <Table name="dim\_arrival">

        </Table>

        <Level name="Year" visible="true" column="YEAR\_ID" type="Integer" uniqueMembers="false" levelType="TimeYears" hideMemberIf="Never">

        </Level>

        <Level name="Months" visible="true" column="MONTH\_NAME" ordinalColumn="MONTH\_ID" type="String" uniqueMembers="false" levelType="TimeMonths" hideMemberIf="Never">

        </Level>

        <Level name="Days" visible="true" column="DAY\_ID" type="Integer" uniqueMembers="false" levelType="TimeDays" hideMemberIf="Never">

        </Level>

      </Hierarchy>

    </Dimension>

    <Dimension type="StandardDimension" visible="true" foreignKey="AIRPLANE\_ID" highCardinality="false" name="Airplane">

      <Hierarchy name="Airplane Hierarchy" visible="true" hasAll="true" allMemberName="All Airplanes" primaryKey="AIRPLANE\_ID">

        <Table name="dim\_airplane">

        </Table>

        <Level name="Type" visible="true" column="AIRPLANE\_TYPE" type="Integer" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">

        </Level>

      </Hierarchy>

    </Dimension>

    <Dimension type="StandardDimension" visible="true" foreignKey="AIRLINE\_ID" highCardinality="false" name="Airlines">

      <Hierarchy name="Airline Hierarchy" visible="true" hasAll="true" allMemberName="All Airlines" primaryKey="AIRLINE\_ID">

        <Table name="dim\_airline">

        </Table>

        <Level name="Name" visible="true" column="AIRLINE\_NAME" type="String" uniqueMembers="false" levelType="Regular" hideMemberIf="Never">

        </Level>

      </Hierarchy>

    </Dimension>

    <Measure name="Bookings" column="TOTALBOOKINGS" datatype="Integer" formatString="#,###" aggregator="sum" visible="true">

    </Measure>

    <Measure name="Revenue" column="REVENUE" datatype="Integer" formatString="$ #,###.00" aggregator="sum" visible="true">

    </Measure>

  </Cube>

</Schema>

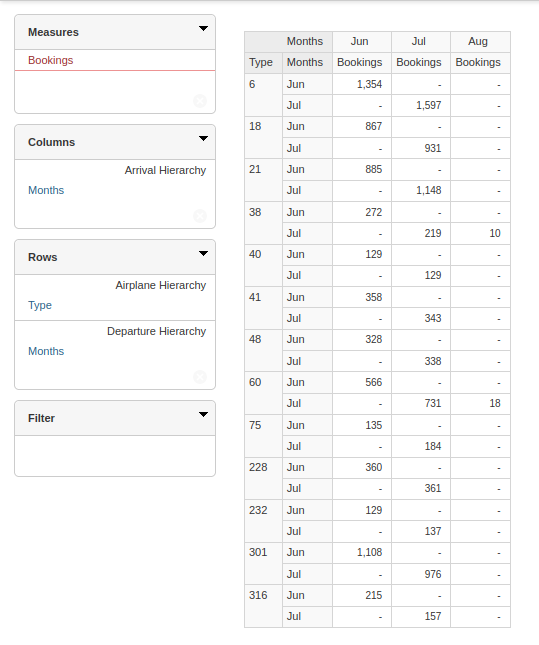
# **Saiku Analysis**

## **airports.sql**

### Passengers and revenue by airline and month.

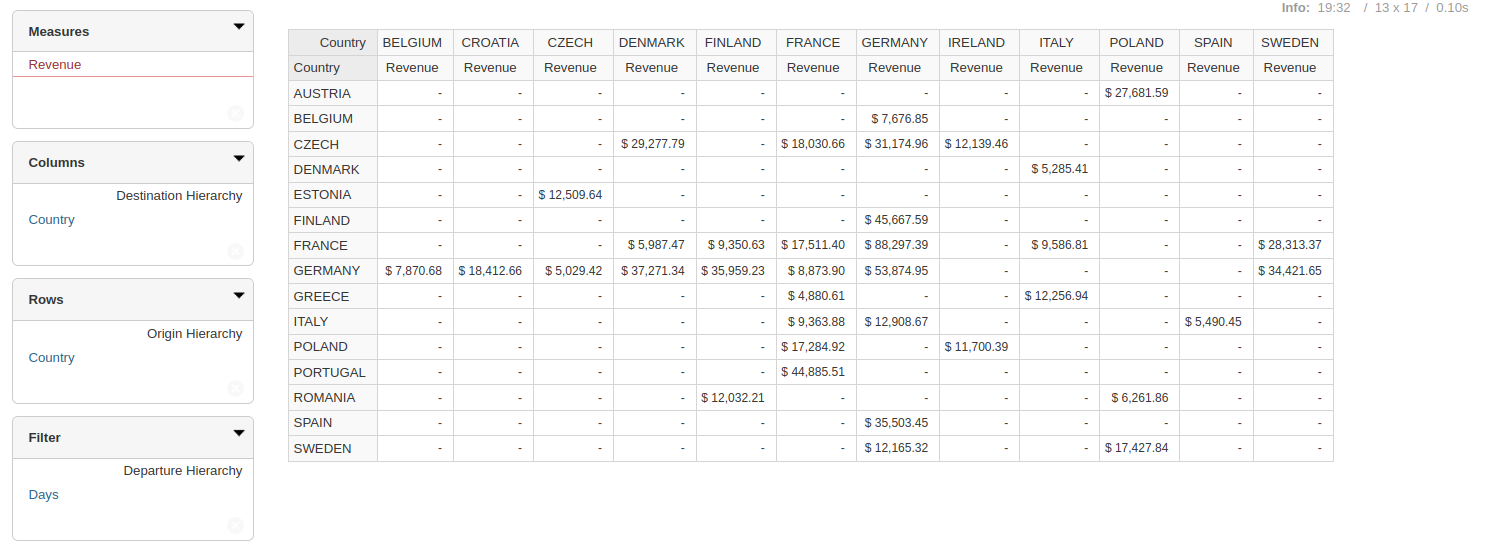
### Airplane type and month of departure by month of arrival.

(2 airplane types departed in July and arrived in August.)



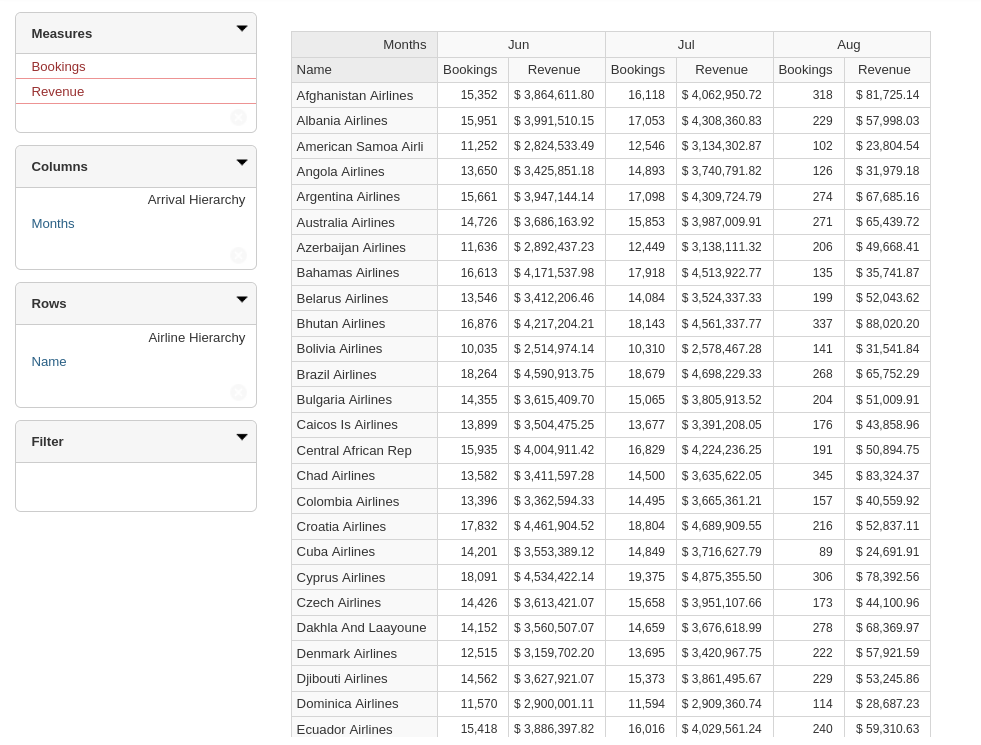
### Money spent/gained on flights between countries.

(On the 10 days with highest revenue – filter option.)



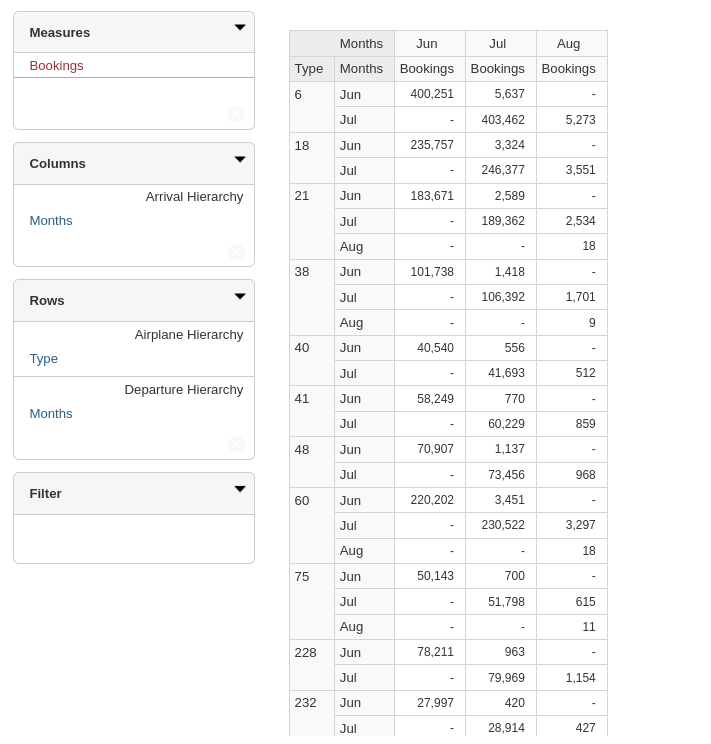
## **airports-large.sql**

### Passengers and revenue by airline and month.



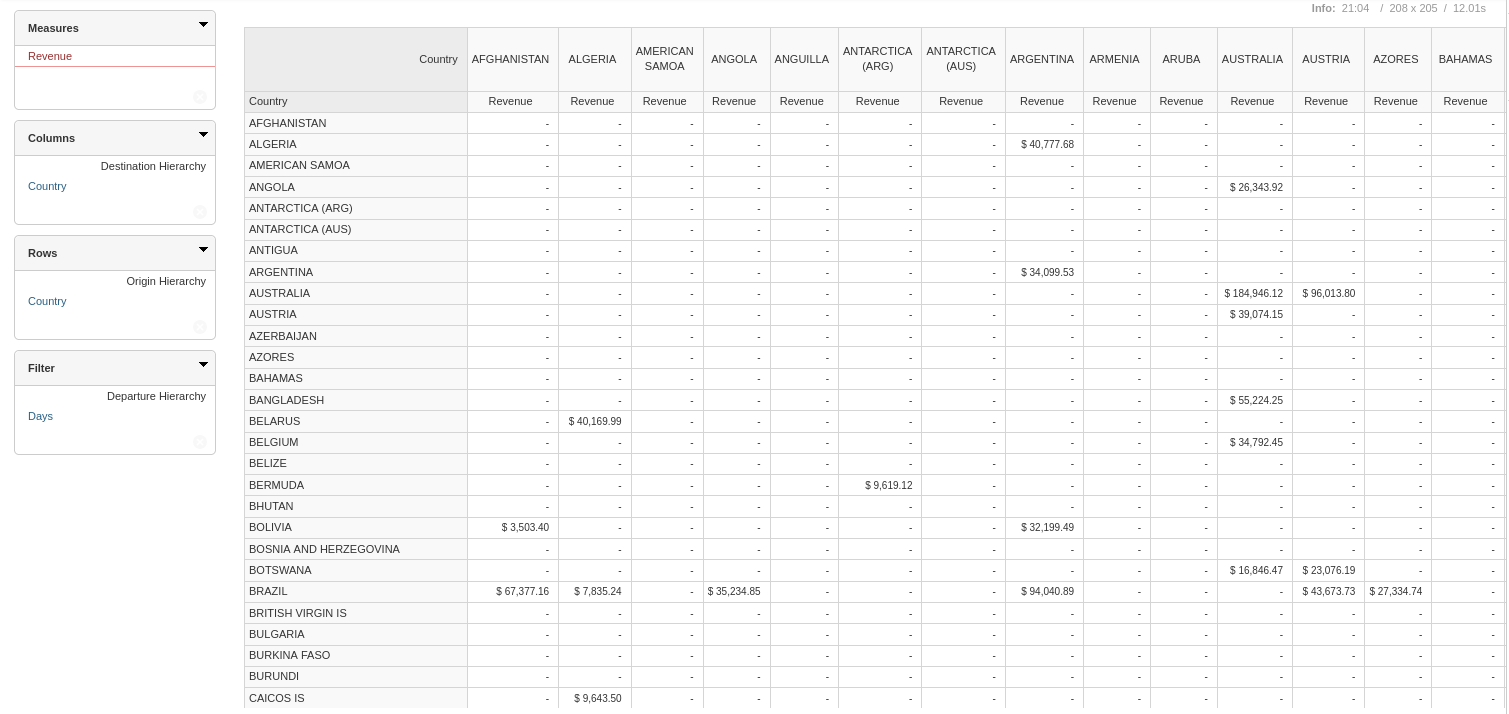
### Airplane type and month of departure by month of arrival.

(2 airplane types departed in July and arrived in August.)



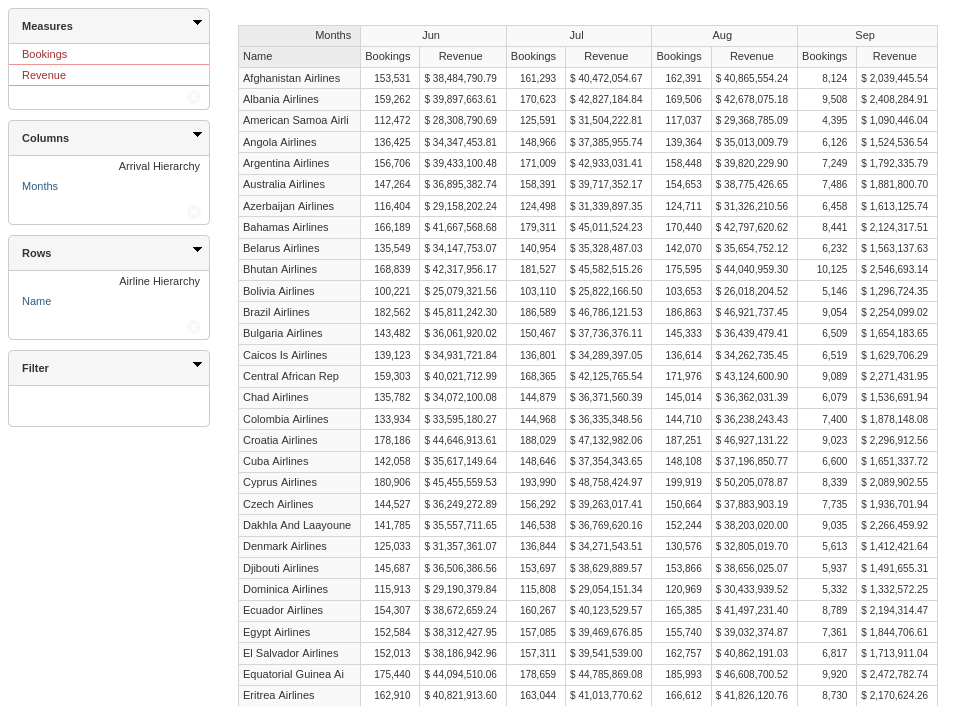
### Money spent/gained on flights between countries.

(On the 10 days with highest revenue – filter option.)



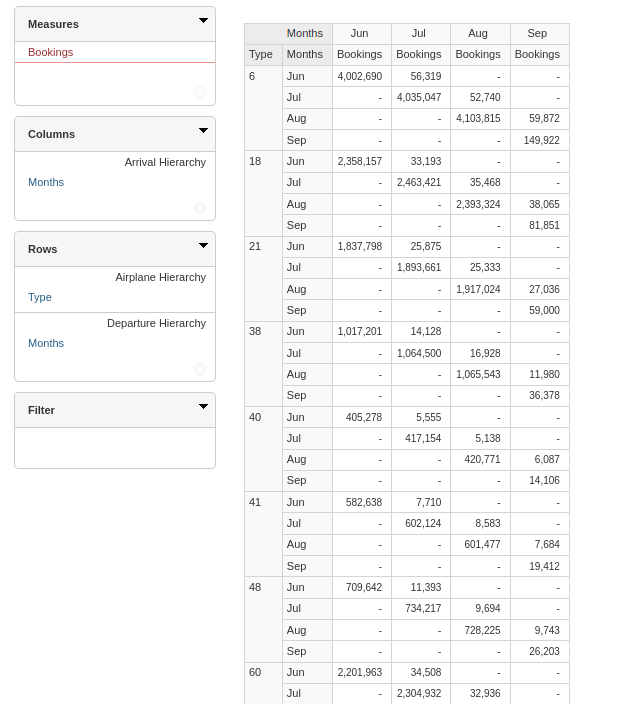
## **airports-large-extra.sql**

### Passengers and revenue by airline and month.



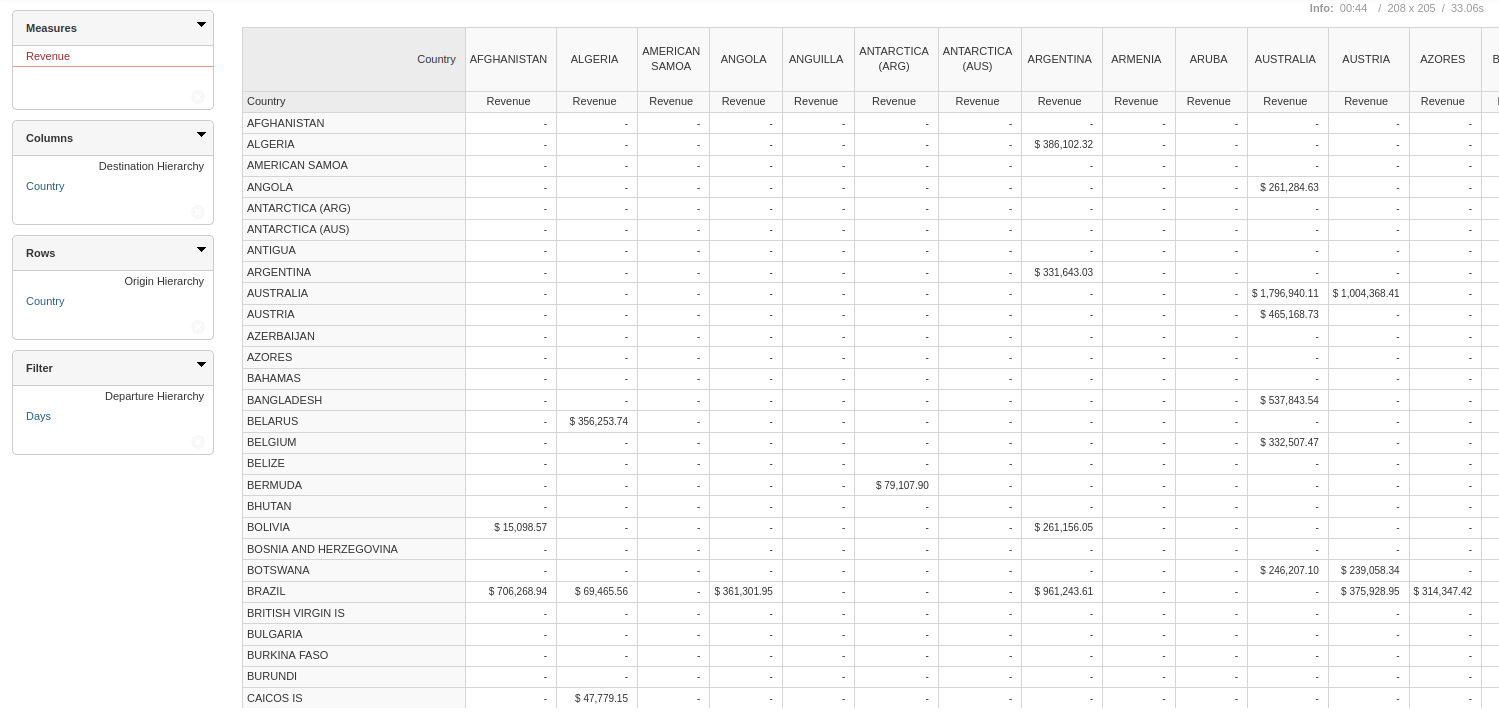
### Airplane type and month of departure by month of arrival.

(2 airplane types departed in July and arrived in August.)

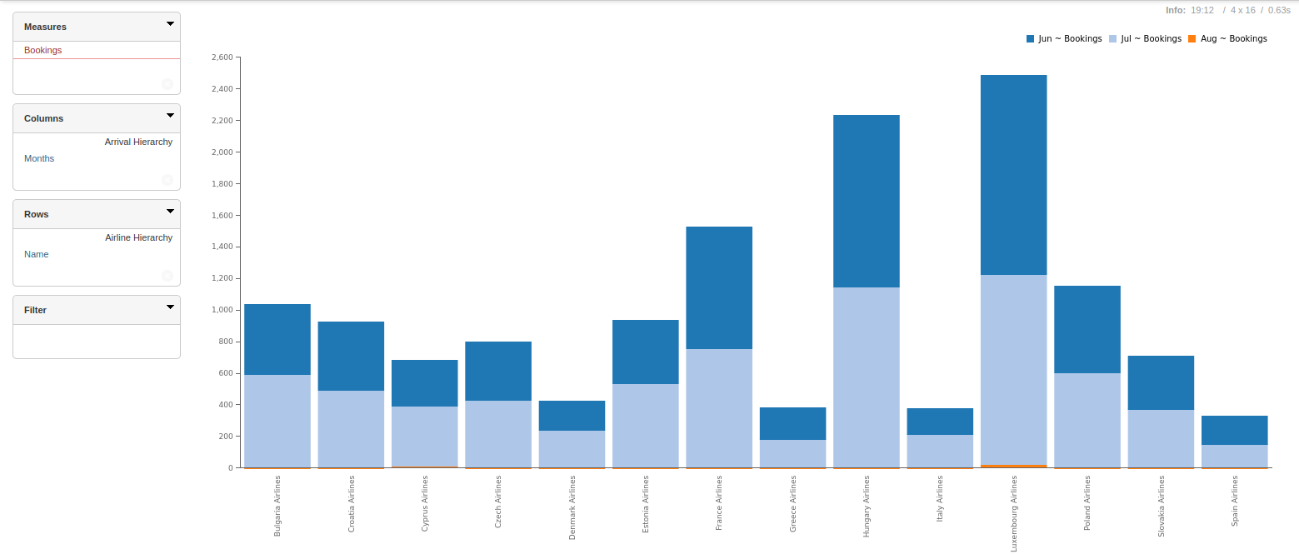


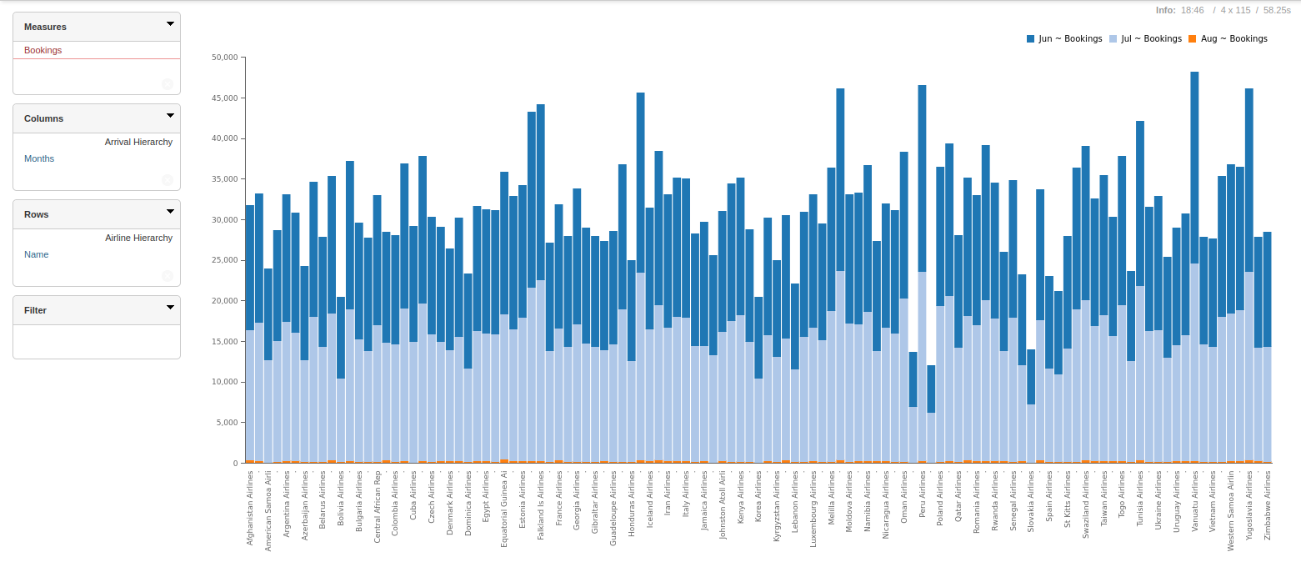
### Money spent/gained on flights between countries.

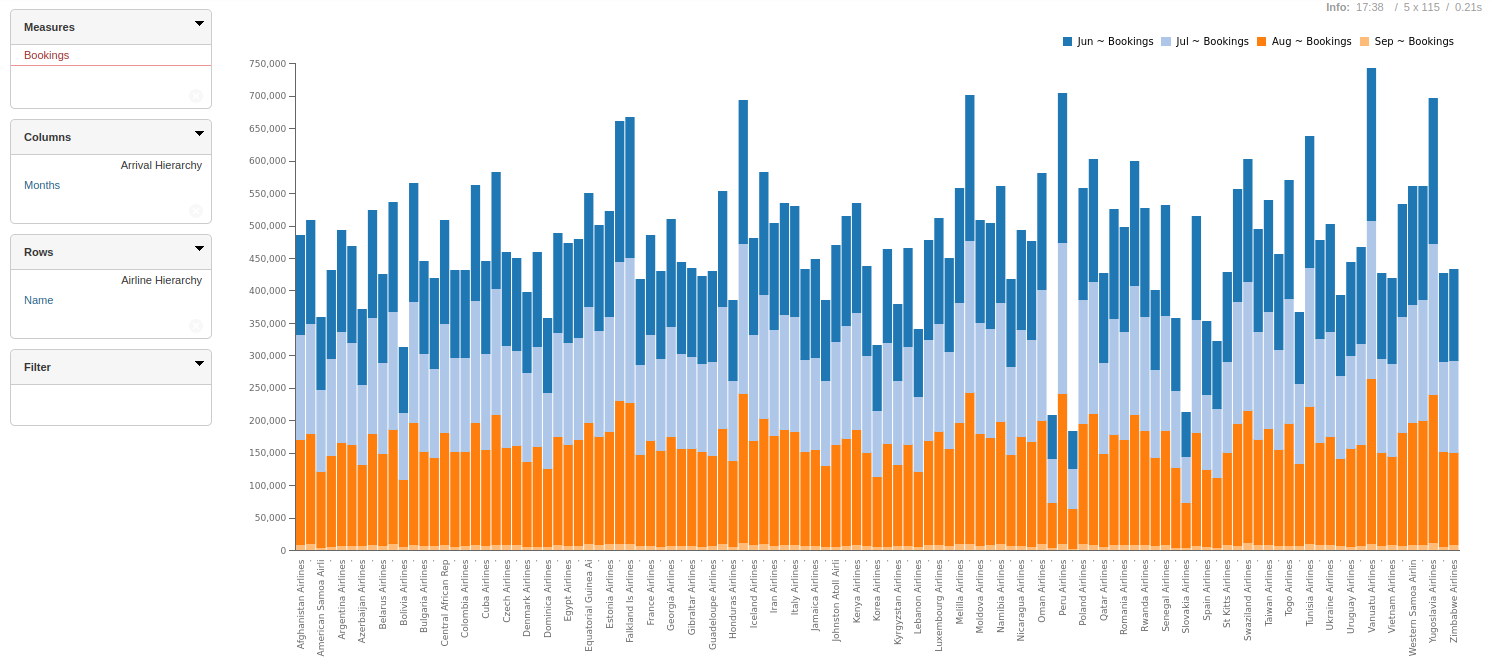
(On the 10 days with highest revenue – filter option.)



# **Comparisons**

***airports.sql***

***airports-large.sql***

***airports-large-extra.sql***

### 

### We decided to compare the three datasets using the first query from exercise 5 (*Passengers and revenue by airline and month*), just with the bookings measure.

### On an overall view, we can see the number of airlines growing considerably from the first to the second graphs, and then the appearance of a new month between the second and third ones.

Despite not comparing the number of airports and airplanes stored in the data warehouse, just by looking at the scales of the chosen graphs, we get a good idea of how the problem (manage information about flights) scales with the three *.sql*.

About the second query (*Airplane type and month of departure by month of arrival*), it only states the curiosity of the flights at the end of the month, which depart at one month and arrive at the other. The table only gets more lines as we increase the problem (can’t be seen by the picture).

In the third query (*Money spent/gained on flights between countries*), we are able to see the scaling of big datasets, because between the first and the second, a considerable number of airports were introduced and the table became more sparse (a higher density of null values). At *airports-large-extra*, the number of airports barely changes (at least in the picture), but there are more flights between the same countries, as the table continues to be sparse and the money involved is higher.