Contents

[Data Warehouse SQL 1](#_Toc118152637)

[Transformations 3](#_Toc118152638)

[dim\_airline 3](#_Toc118152639)

[dim\_airplane 4](#_Toc118152640)

[dim\_arrival 5](#_Toc118152641)

[dim\_departure 6](#_Toc118152642)

[dim\_airport 7](#_Toc118152643)

[fact\_flight 8](#_Toc118152644)

[Jobs 10](#_Toc118152645)

[load\_dw 10](#_Toc118152646)

[load\_fact 10](#_Toc118152647)

[Cube Definition 11](#_Toc118152648)

[Saiku Analysis 14](#_Toc118152649)

[airports.sql 14](#_Toc118152650)

[passengers and revenue by airline and month 14](#_Toc118152651)

[airplane type and month of departure by month of arrival 15](#_Toc118152652)

[how much money spent/gained on flights between countries 16](#_Toc118152653)

[airports-large.sql 17](#_Toc118152654)

[passengers and revenue by airline and month 17](#_Toc118152655)

[airplane type and month of departure by month of arrival 18](#_Toc118152656)

[how much money spent/gained on flights between countries 19](#_Toc118152657)

[airports-large-extra.sql 20](#_Toc118152658)

[passengers and revenue by airline and month 20](#_Toc118152659)

[airplane type and month of departure by month of arrival 21](#_Toc118152660)

[how much money spent/gained on flights between countries 22](#_Toc118152661)

[Comparisons 23](#_Toc118152662)

# 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),

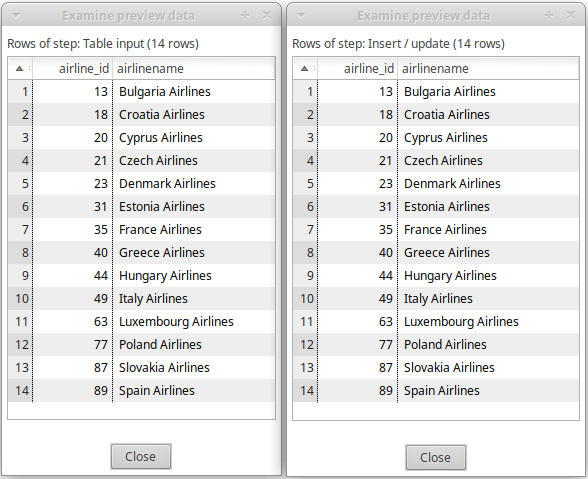
*FOREIGN KEY* (DEPARTURE) *REFERENCES* dim\_departure (TIME\_ID),

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

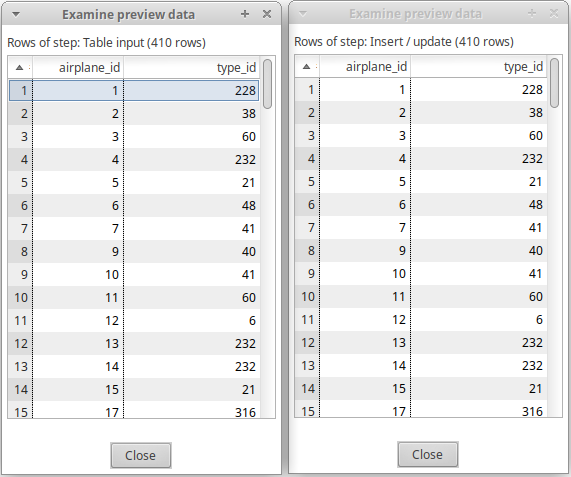
)

# Transformations

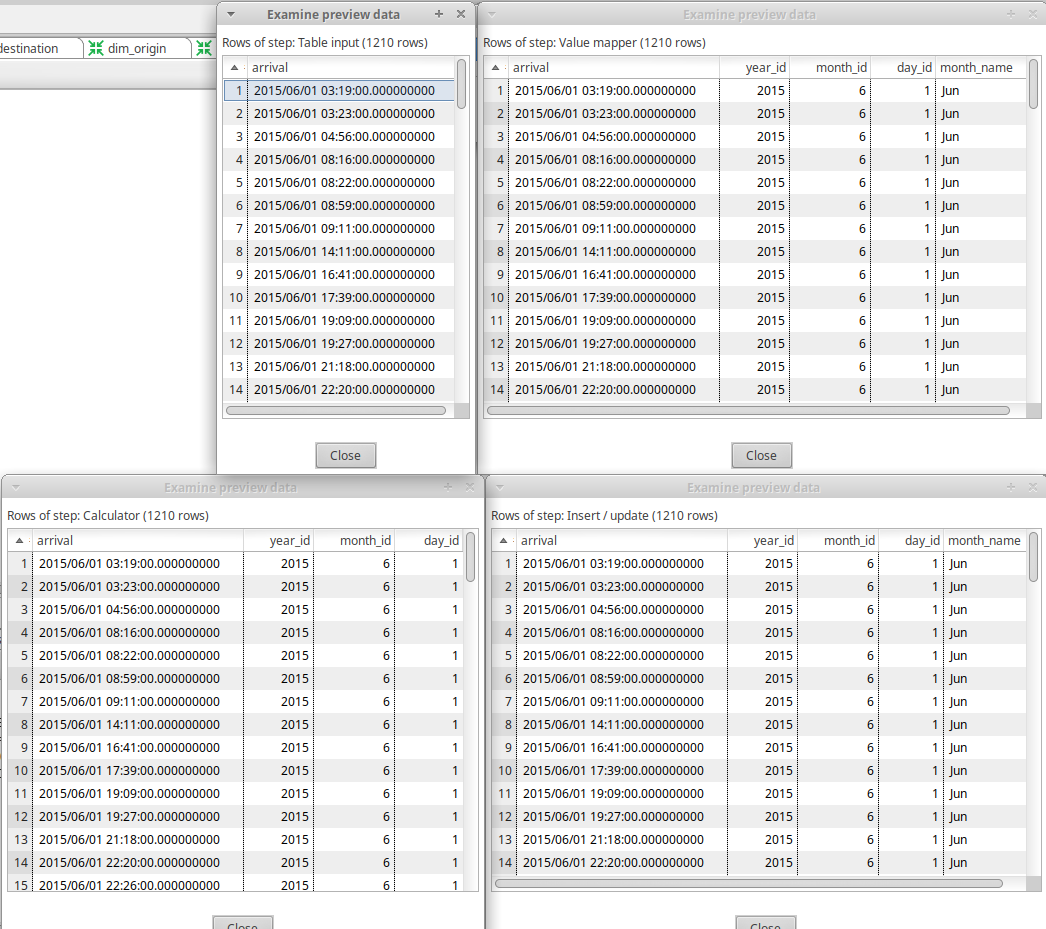
## Graphical user interface, text, application, email Description automatically generateddim\_airline



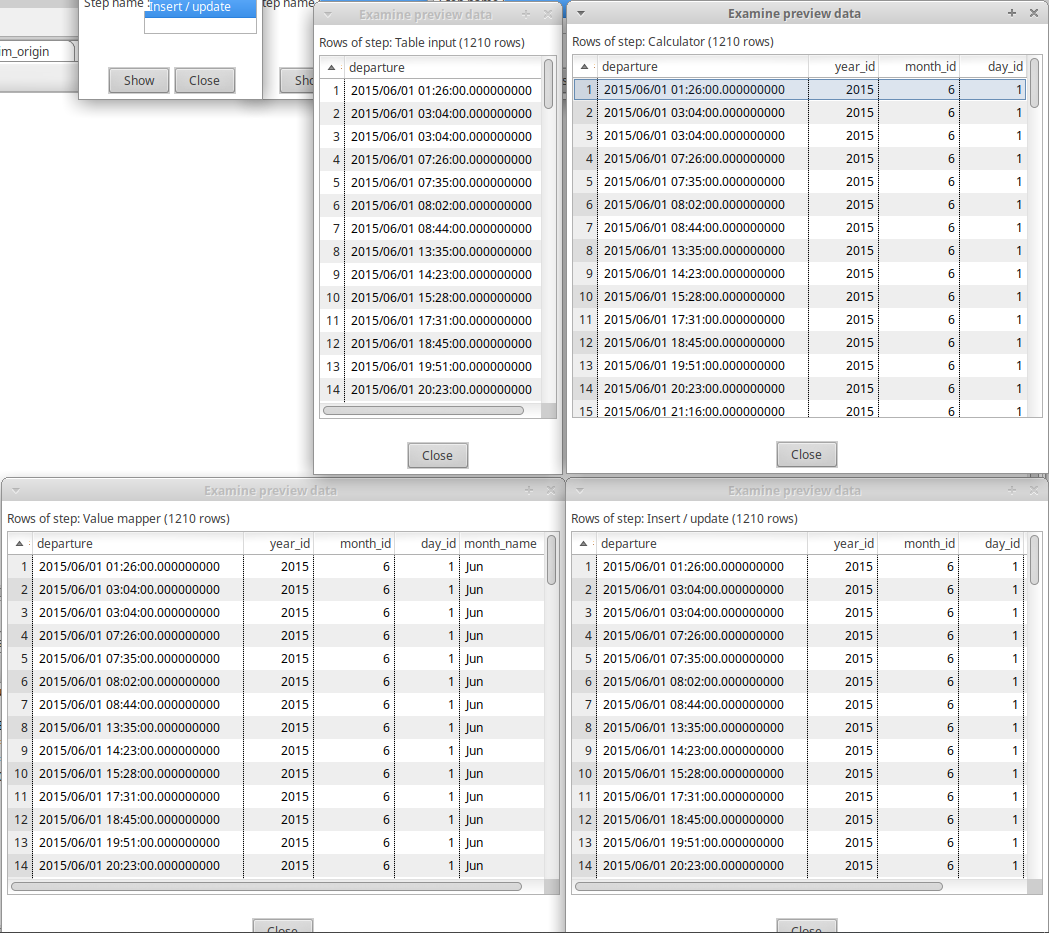
## dim\_airplane



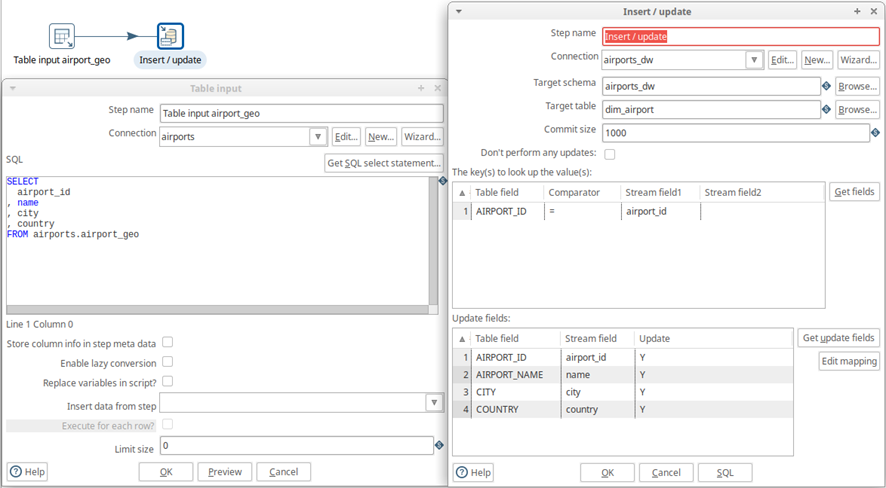
## dim\_arrival

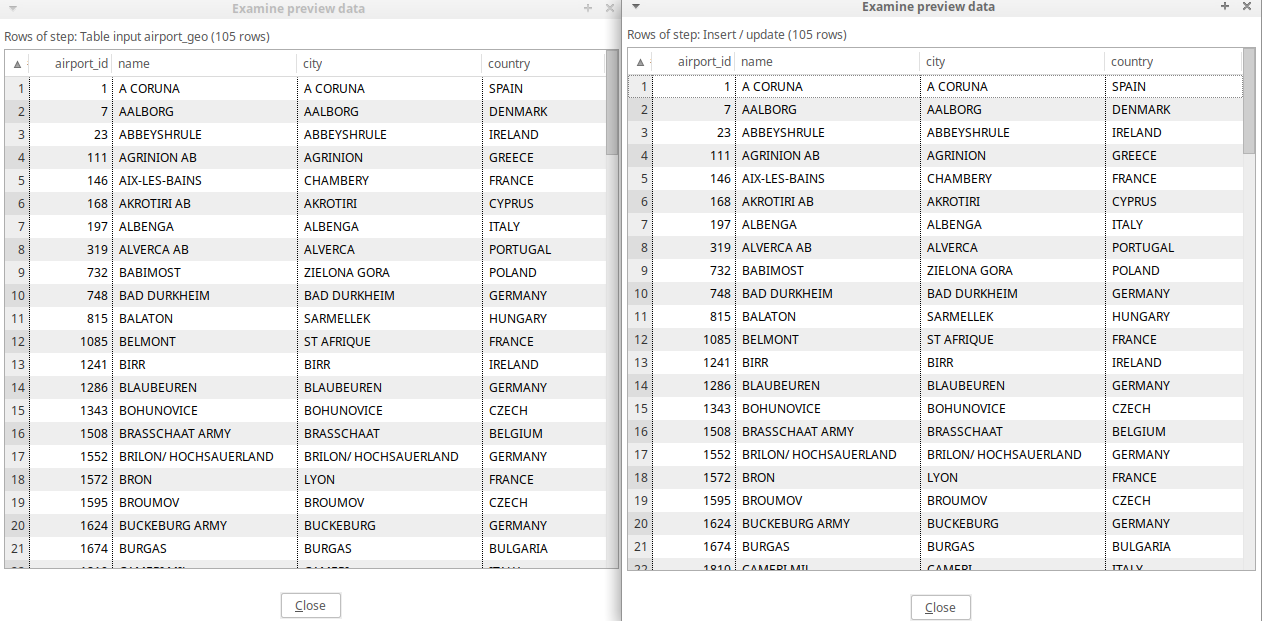


## dim\_departure

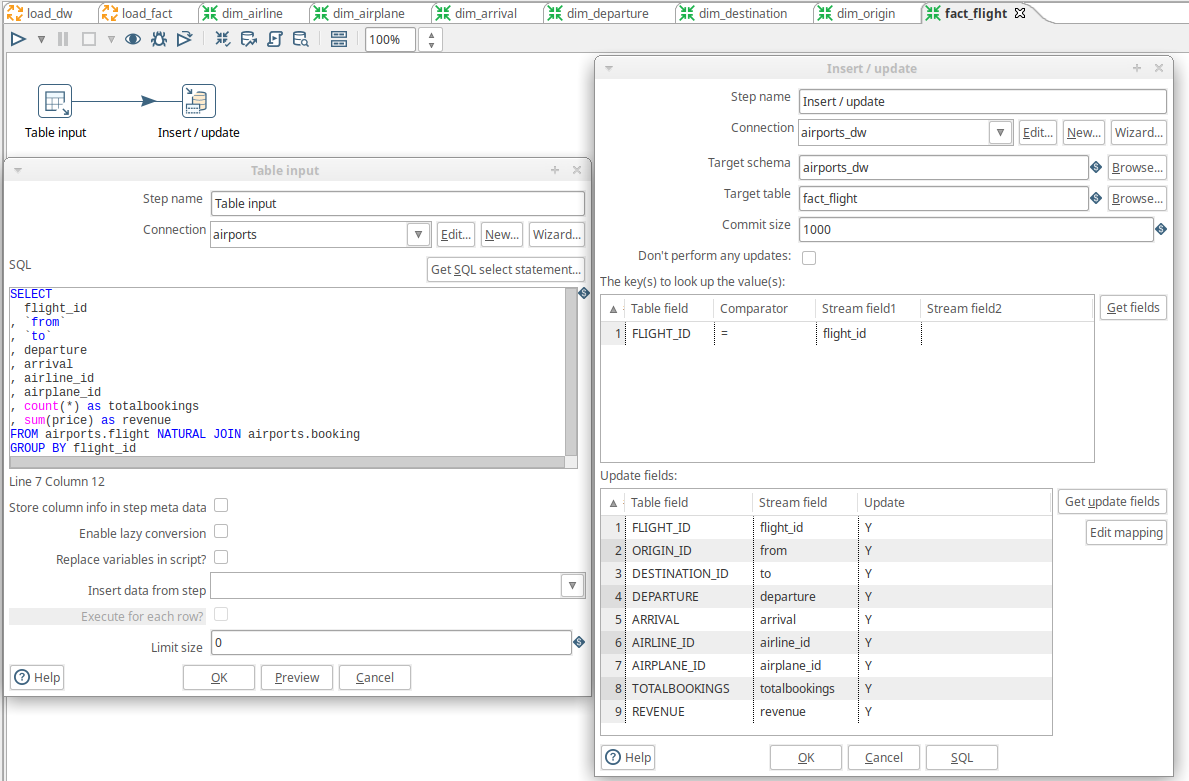


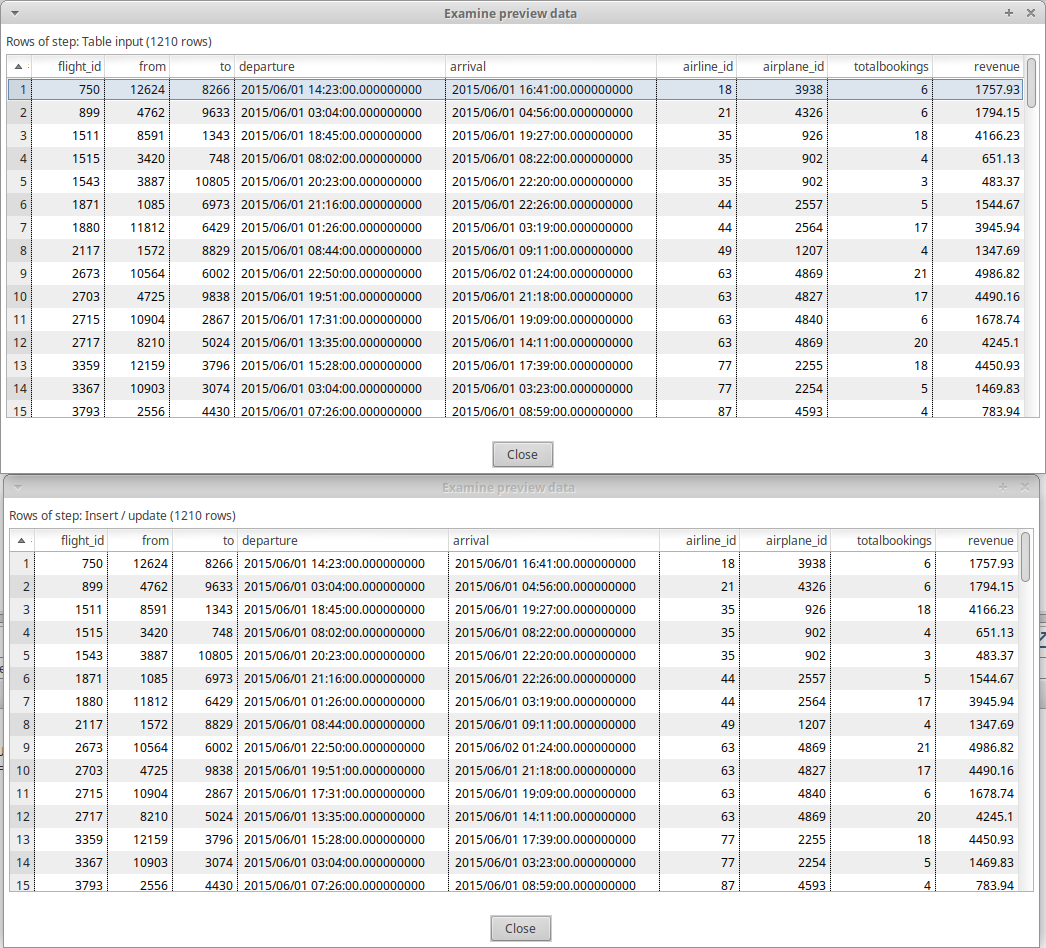
## dim\_airport

(note: dim\_origin and dim\_destination were replaced by dim\_airport)



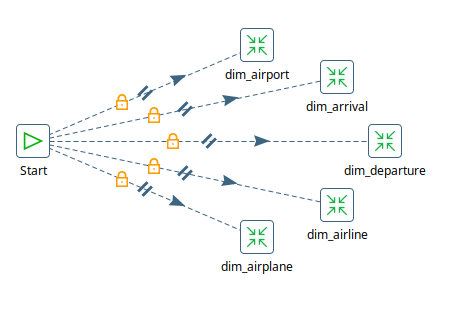
## fact\_flight



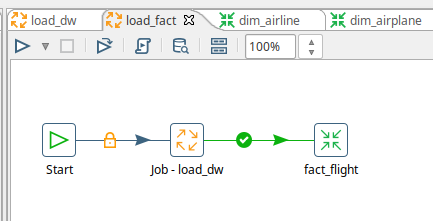


# Jobs

load\_dw – loads dimension in parallel



load\_fact – loads fact\_flight table after all dimensions 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>

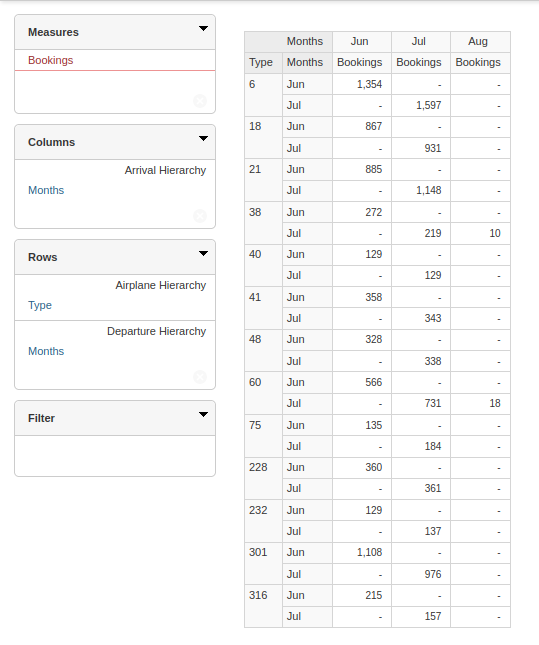
# Saiku Analysis

## airports.sql

### passengers and revenue by airline and month

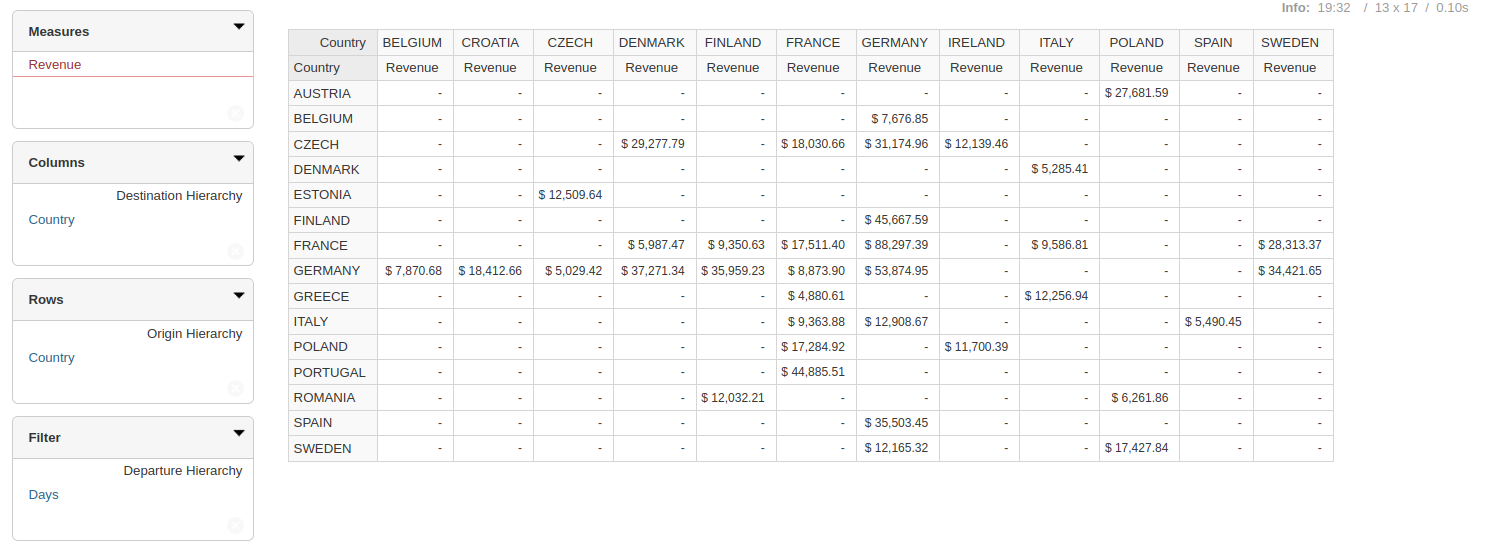
### airplane type and month of departure by month of arrival

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



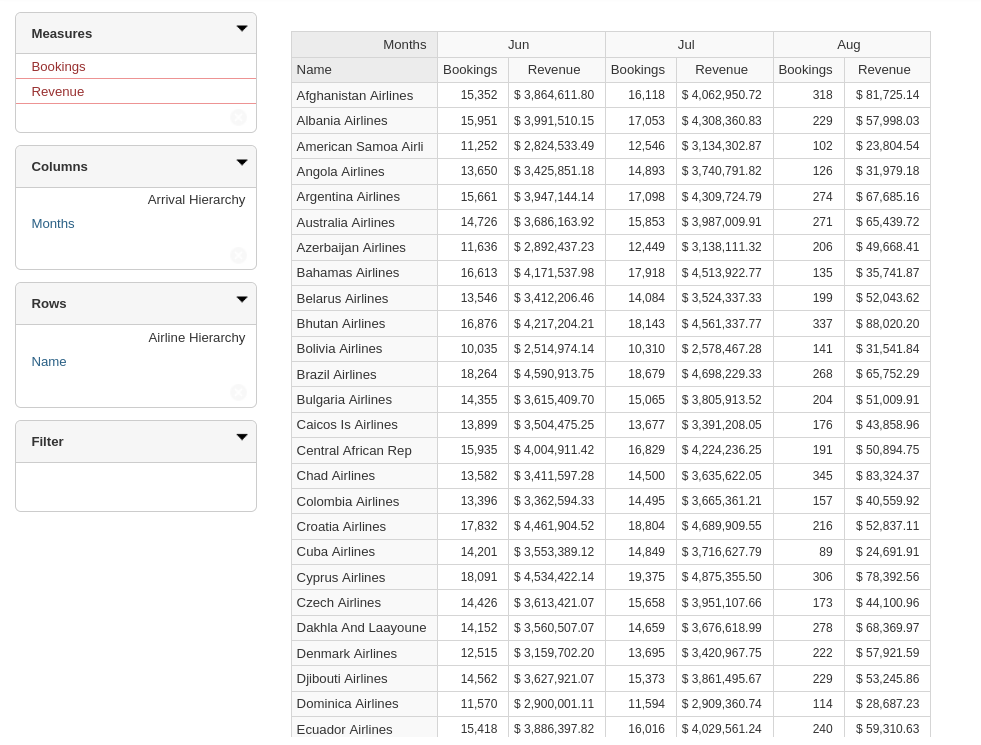
### how much money spent/gained on flights between countries

(on the 10 days with highest revenue)



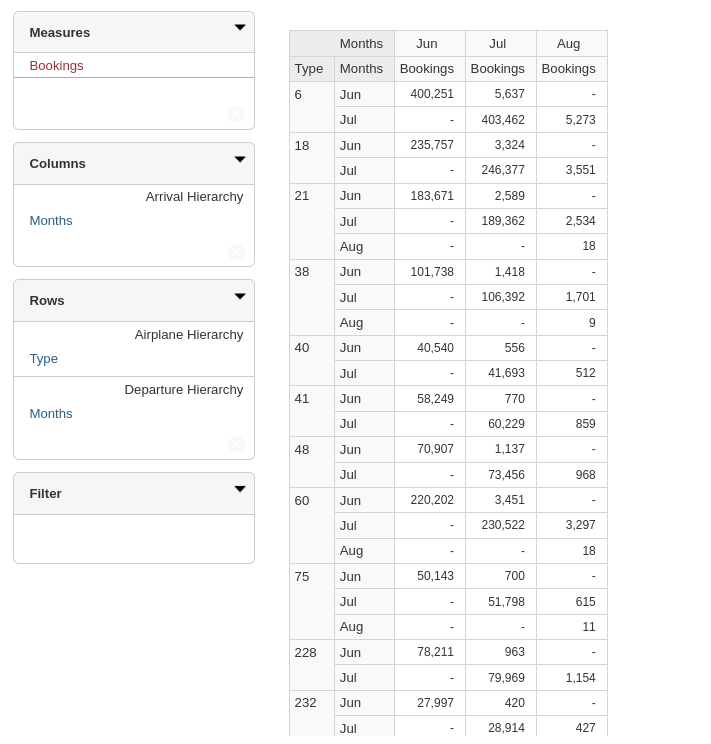
## airports-large.sql

### passengers and revenue by airline and month



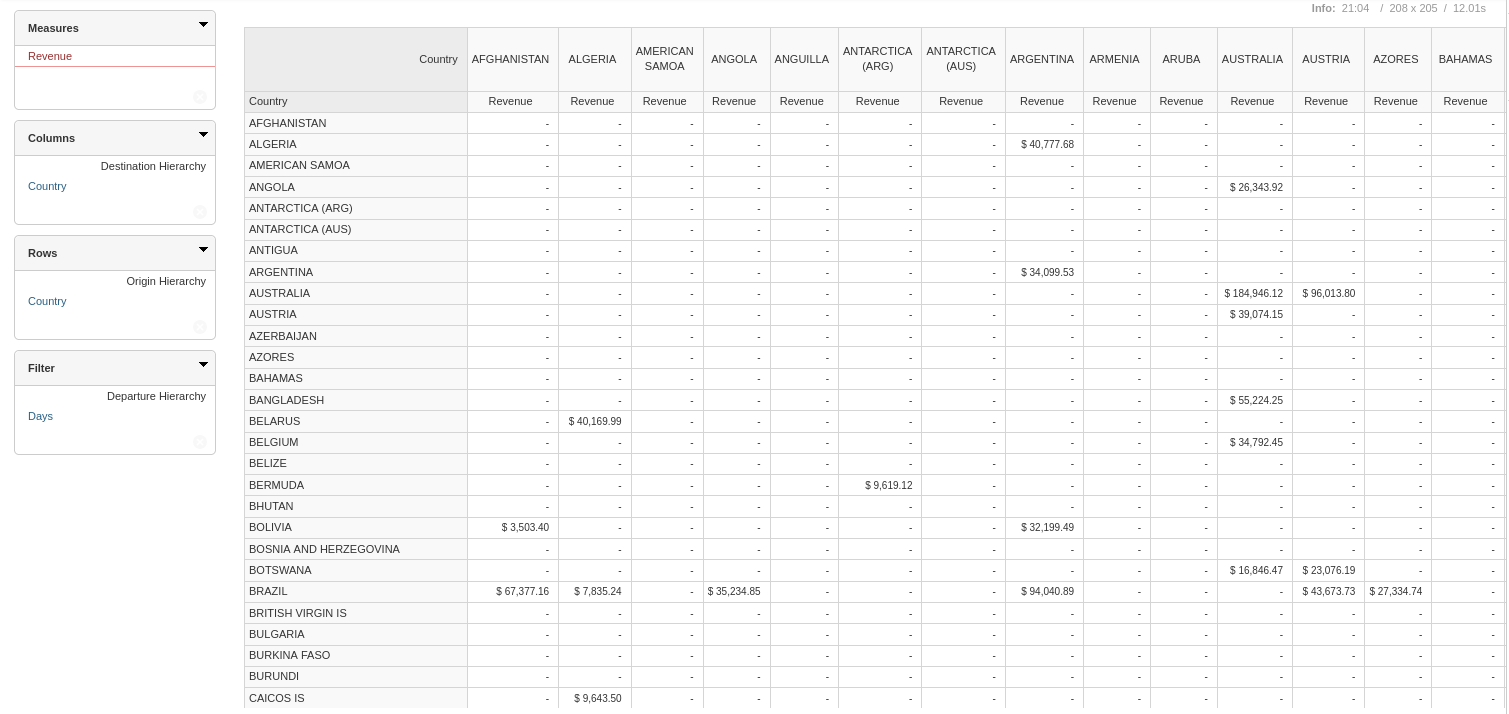
### airplane type and month of departure by month of arrival

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



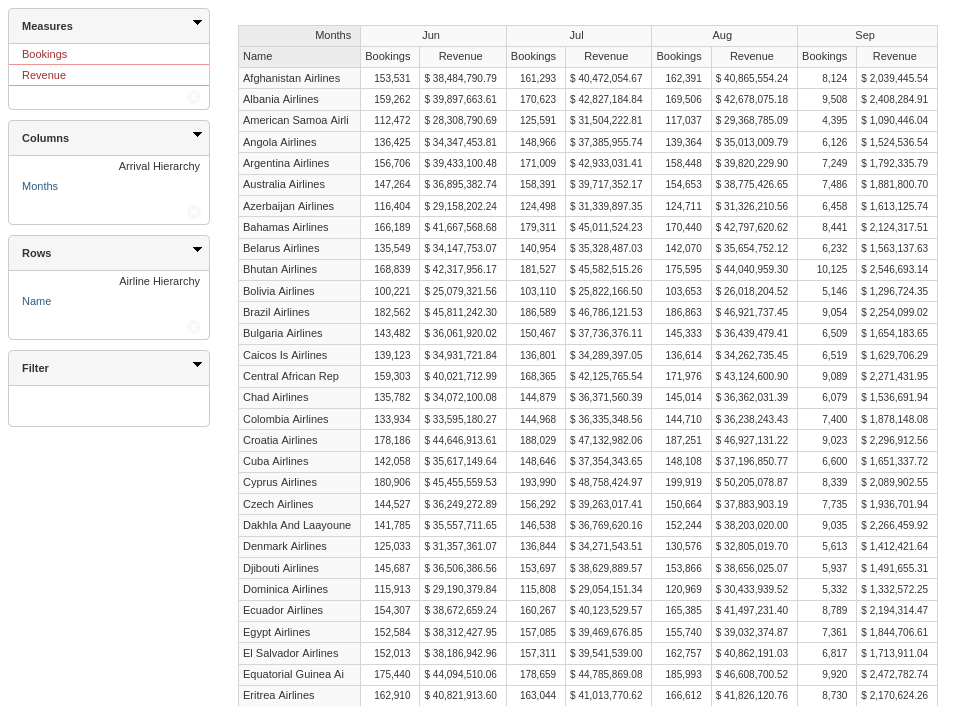
### how much money spent/gained on flights between countries

(on the 10 days with highest revenue)



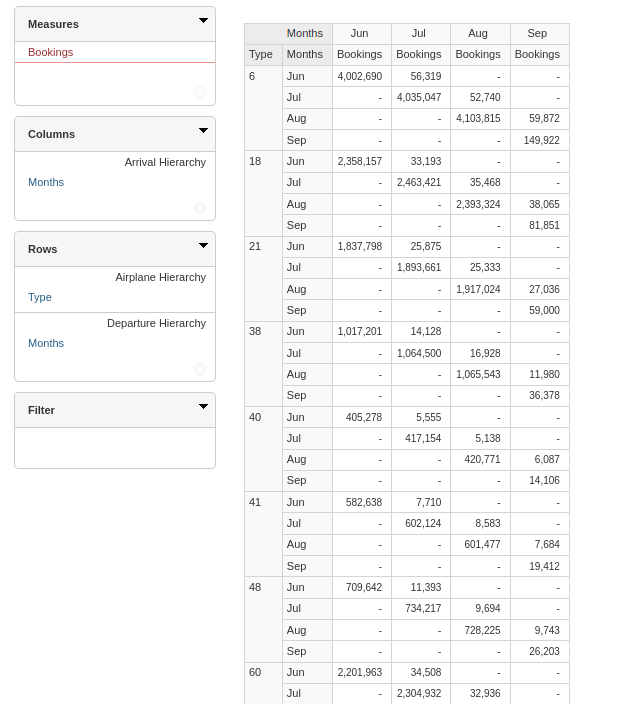
## airports-large-extra.sql

### passengers and revenue by airline and month



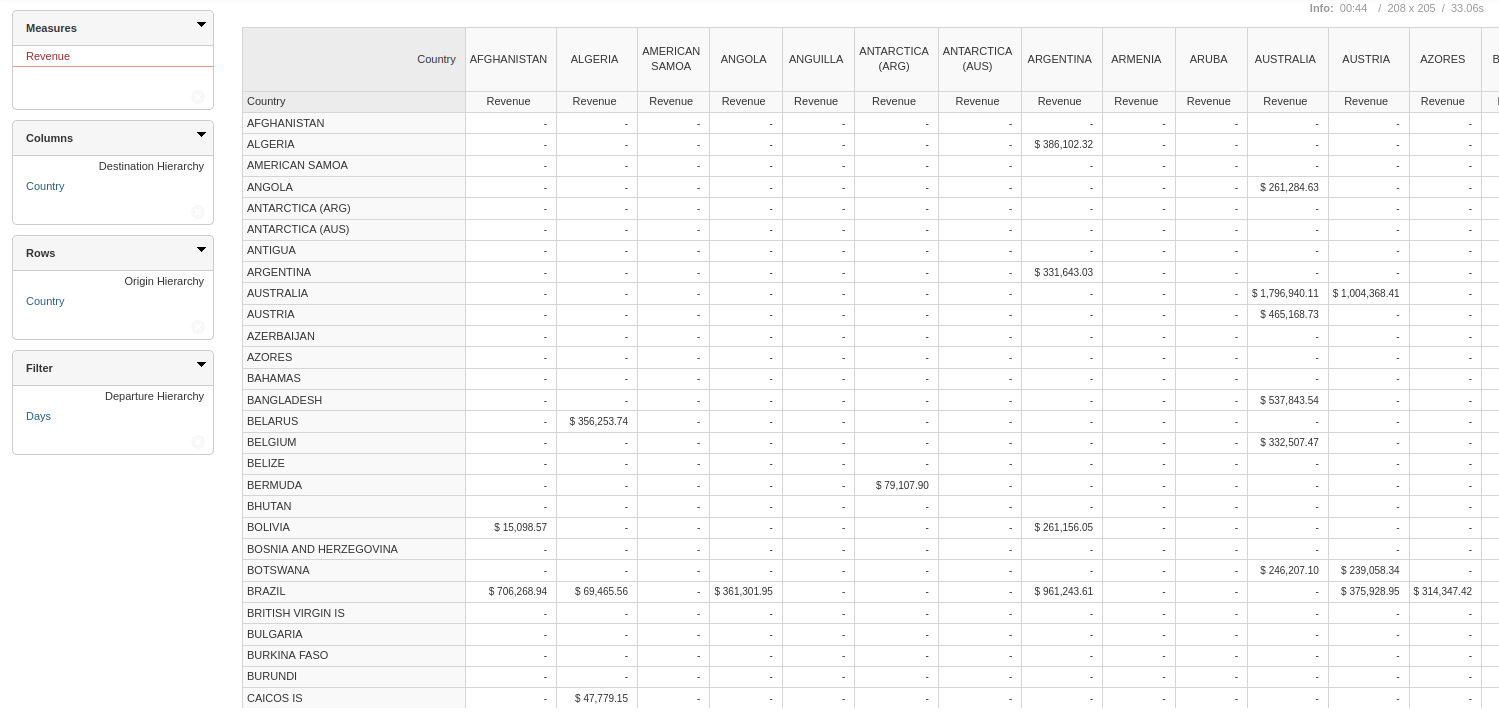
### airplane type and month of departure by month of arrival

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

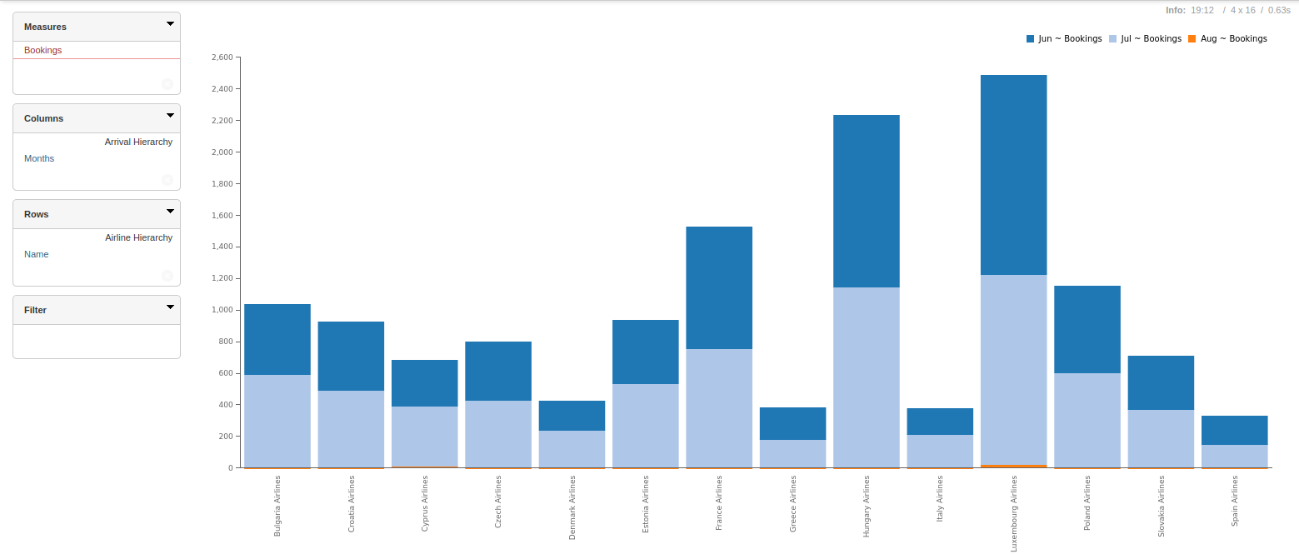


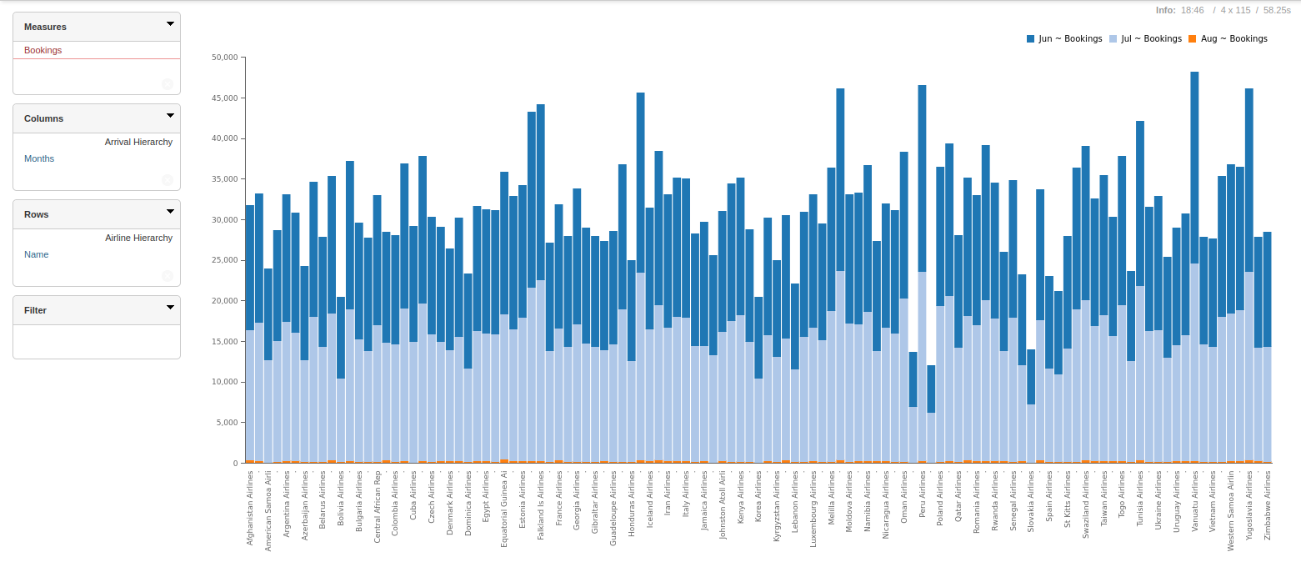
### how much money spent/gained on flights between countries

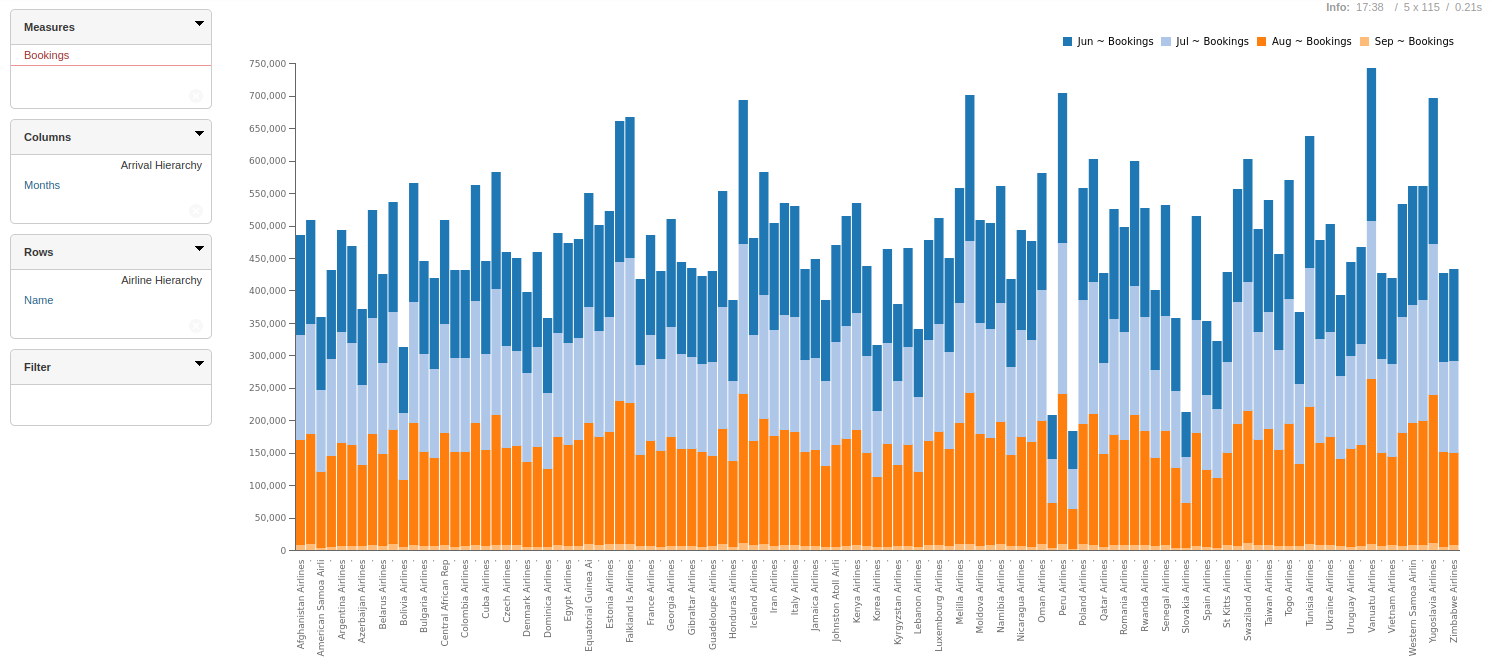
(on the 10 days with highest revenue)



# Comparisons

airports.sql

airports-large.sql

airports-large-extra.sql

We decided to compare the 3 datasets using the first query from exercise 5, just using the bookings measure.

From the graphs, we can see that scale increases 20-fold between the first and second graphs, and 15-fold between the second and third graphs.

From the first to the second graphs, we see more airlines. From the second to the third graphs, there is around the same number of airlines, but a new month is added, September.

Thus, we can conclude that we are dealing with increasingly denser datasets.