ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ
SCHOOL OF BUSINESS

ΤΜΗΜΑ ΔΙΟΙΚΗΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ & ΤΕΧΝΟΛΟΓΙΑΣ
DEPARTMENT OF MANAGEMENT SCIENCE & TECHNOLOGY

ΟΠΑ
AUEB

Flight Delays
USA - 2015

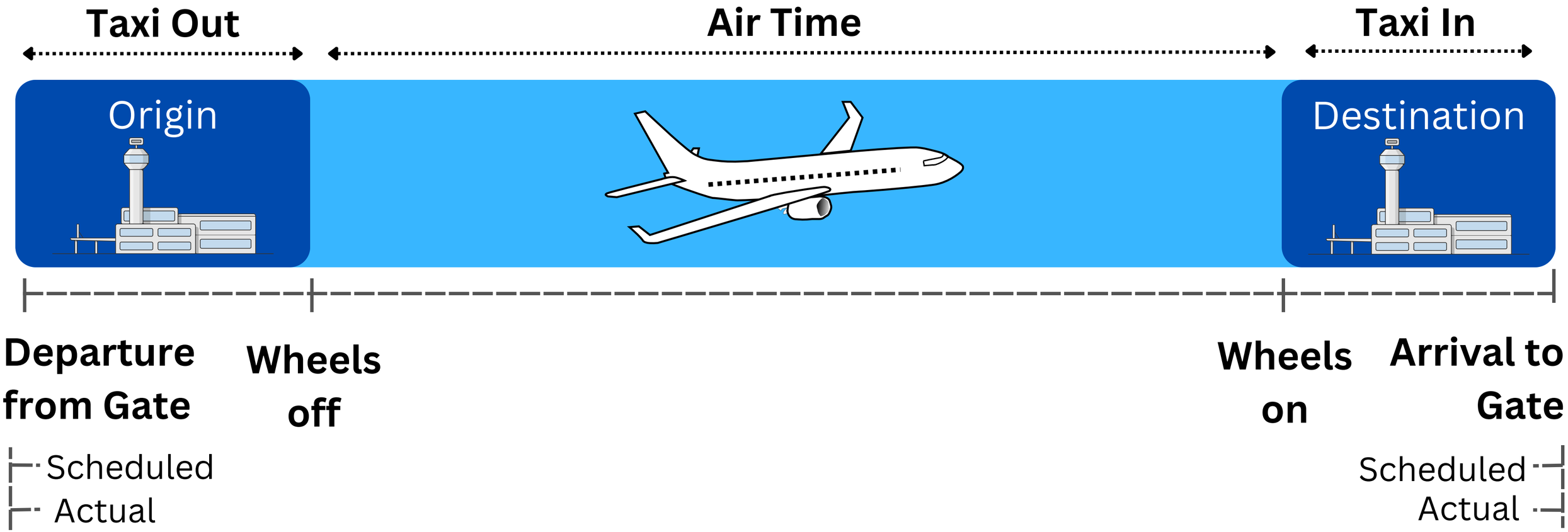
Dimitris Bouris 8190119
Philippos Priovolos 8190147

2015 Flight Delays and Cancellations

Main Info

- Flight
  - Date - Time
  - Scheduled & actual departure time
  - Scheduled & actual arrival time
  - Arrival delay
- Airline
- Airport
  - Origin and Destination
  - State
  - Latitude and Longitude

Flight Procedure



Delay Metrics

- **Departure Delay** = Scheduled departure - Actual departure
- **Arrival Delay** = Scheduled arrival - Actual Arrival

Early, On-Time & Delayed

- According to The United States Federal Aviation Administration (FAA), a flight is considered to be **delayed** when it arrives **10 minutes later** than its scheduled time.
- We consider a flight to be **early** when it arrives at least **10 minutes earlier** that the scheduled time

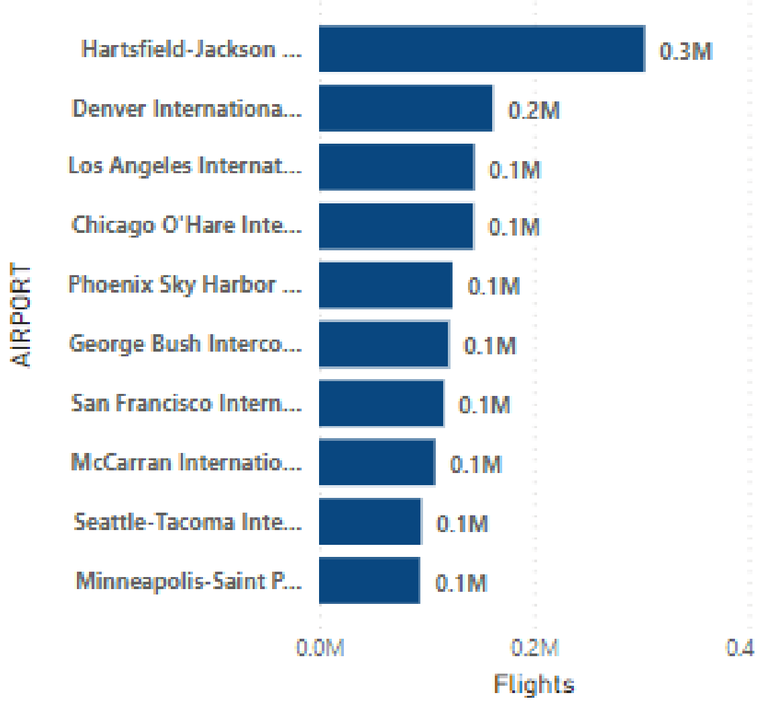
Additional Aircraft Data

Federal Aviation Administration

Overview

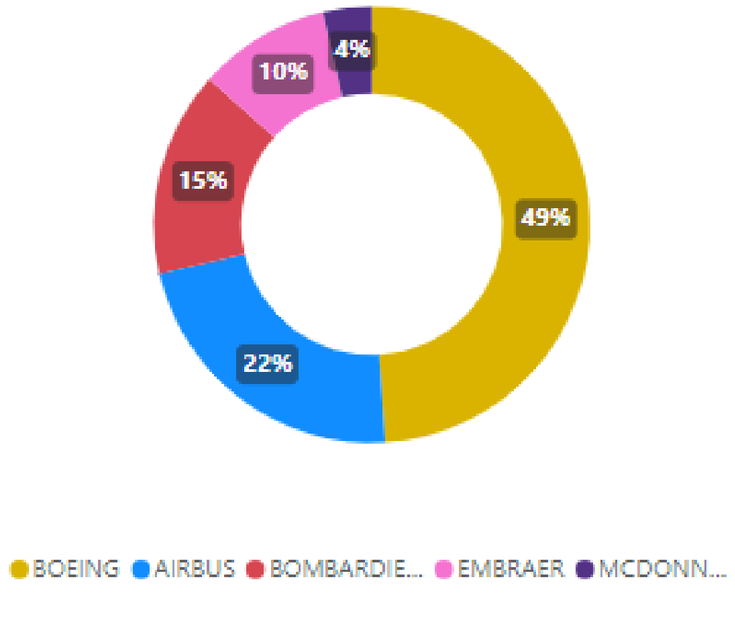
FLIGHTS	EARLY	ON-TIME	DELAYED
4M	35%	50%	15%

Busiest Airports



Busiest Airlines

Top Manufacturers

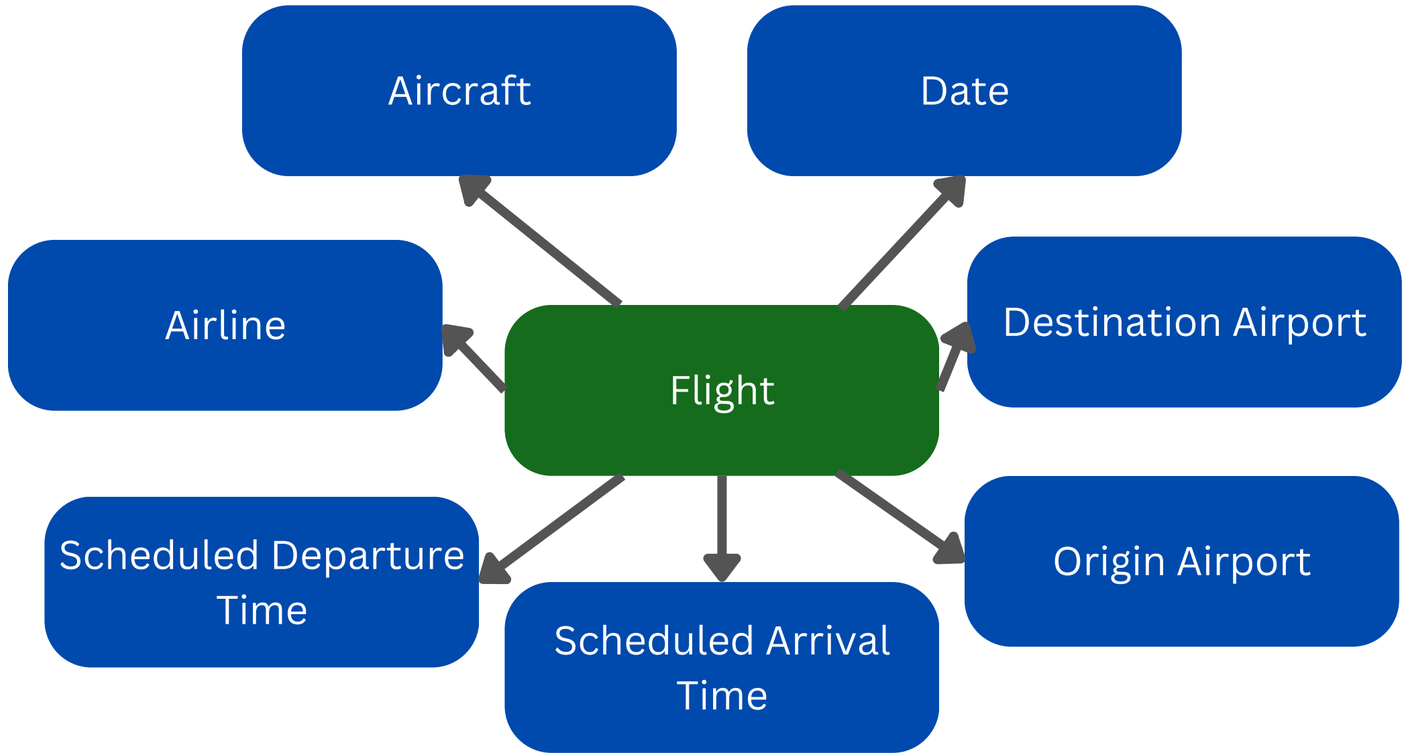


Cube Design

Star Schema

1 Fact table & 7 Dimensions

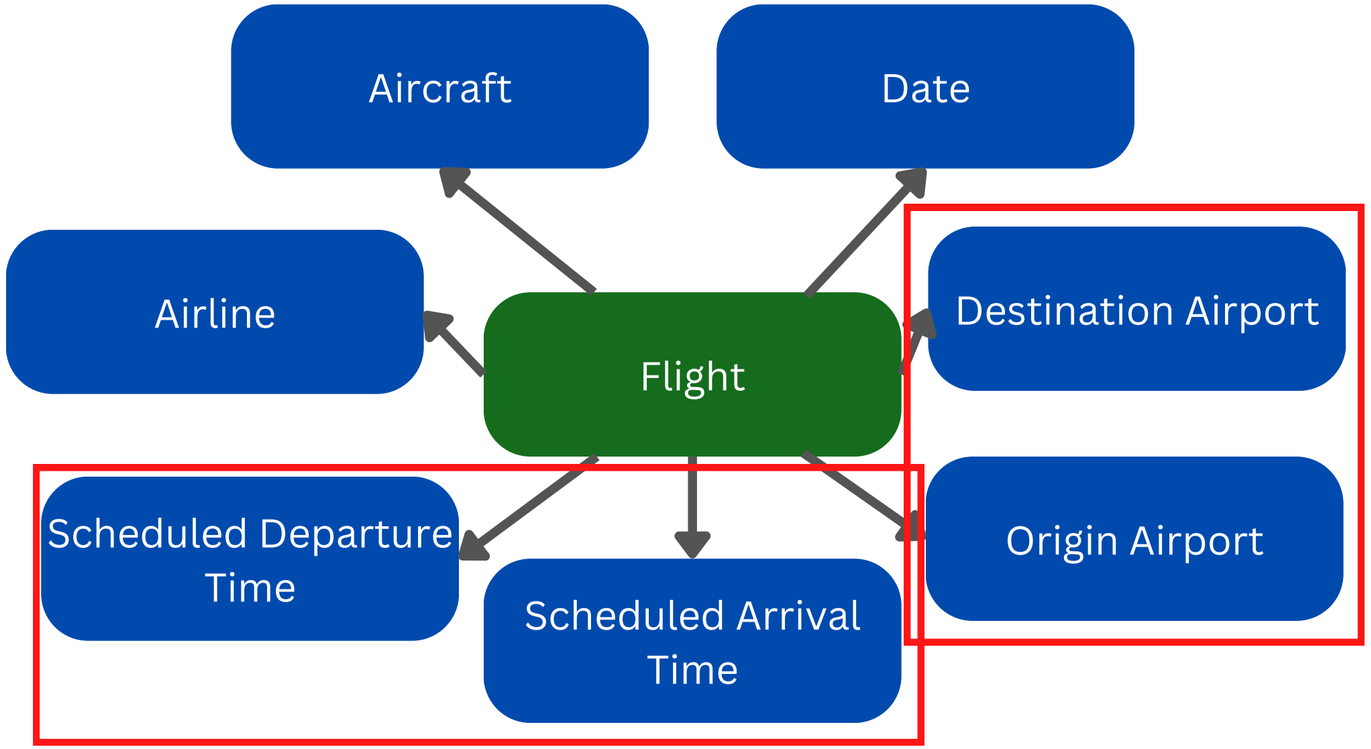
- Fact table
  - Flight Data and Measures
- Dimensions
  - Date
  - Departure and Arrival Time
  - Origin and Destination Airport
  - Airline
  - Aircraft



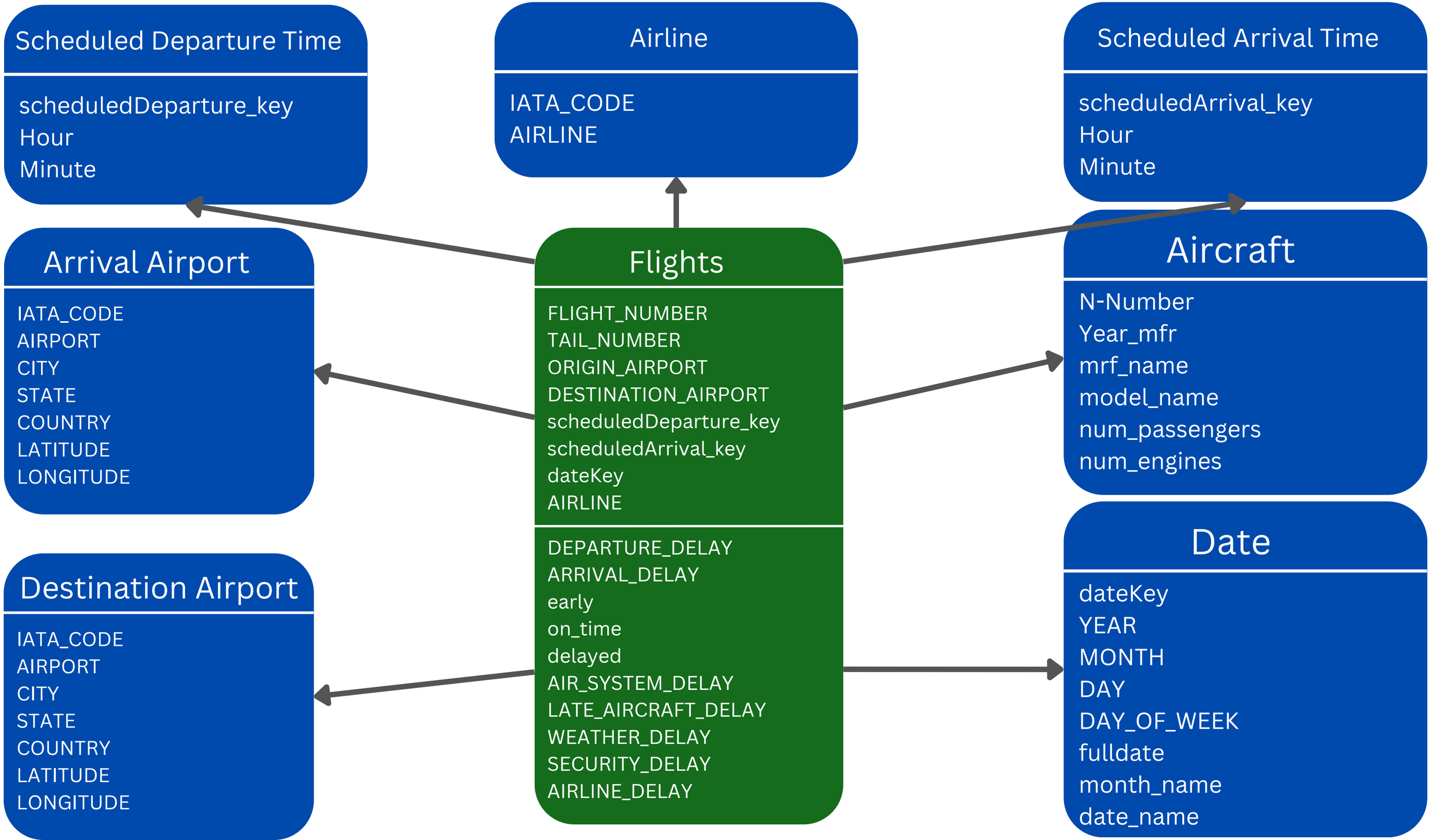
Airport and Time Dimenssion

Airport and Time are modeled in two separate dimensions

- **Origin - Destination** Airport
- Scheduled **Departure and Arrival** Time



This separation will help us analyze the **individual effect** of the Origin and Destination airport on the delay. Same for the Time dimension.


Scheduled Departure Time
scheduledDeparture_key
Hour
Minute
Airline
IATA_CODE
AIRLINE
Scheduled Arrival Time
scheduledArrival_key
Hour
Minute
Arrival Airport
IATA_CODE
AIRPORT
CITY
STATE
COUNTRY
LATITUDE
LONGITUDE
Flights
FLIGHT_NUMBER
TAIL_NUMBER
ORIGIN_AIRPORT
DESTINATION_AIRPORT
scheduledDeparture_key
scheduledArrival_key
dateKey
AIRLINE
DEPARTURE_DELAY
ARRIVAL_DELAY
early
on_time
delayed
AIR_SYSTEM_DELAY
LATE_AIRCRAFT_DELAY
WEATHER_DELAY
SECURITY_DELAY
AIRLINE_DELAY
Aircraft
N-Number
Year_mfr
mrf_name
model_name
num_passengers
num_engines
Destination Airport
IATA_CODE
AIRPORT
CITY
STATE
COUNTRY
LATITUDE
LONGITUDE
Date
dateKey
YEAR
MONTH
DAY
DAY_OF_WEEK
fulldate
month_name
date_name

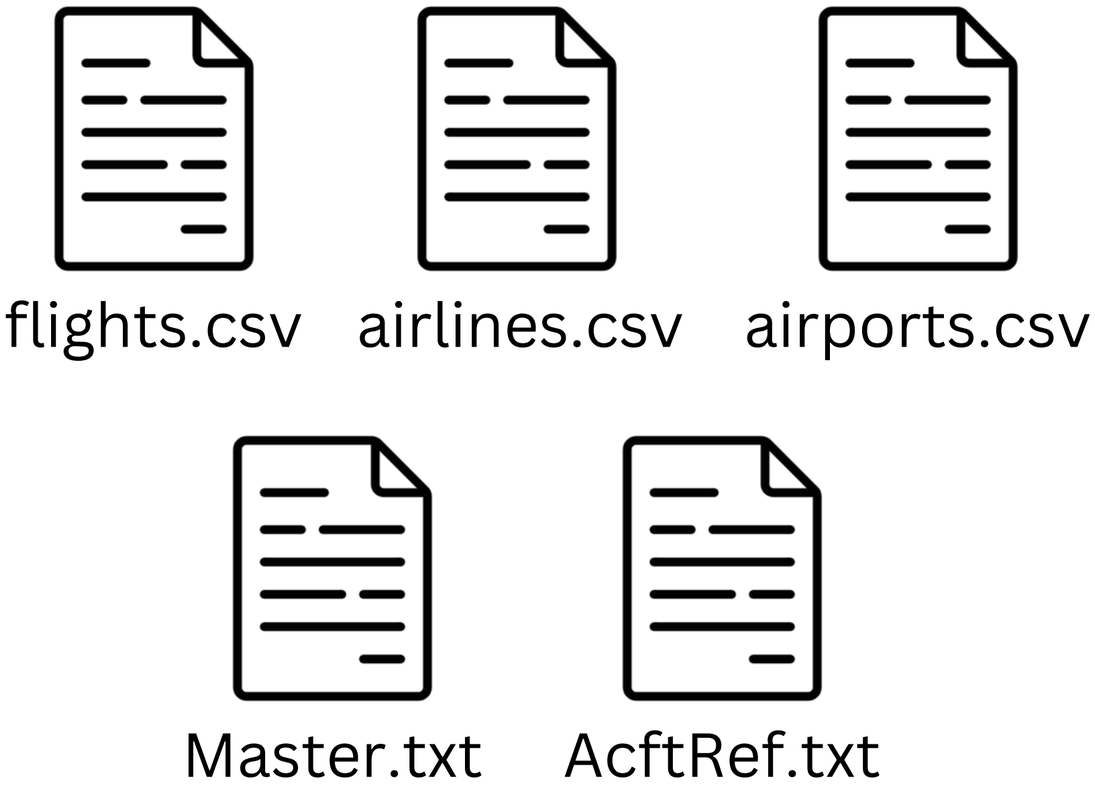
Extract - Transform - Load

ETL

Extract

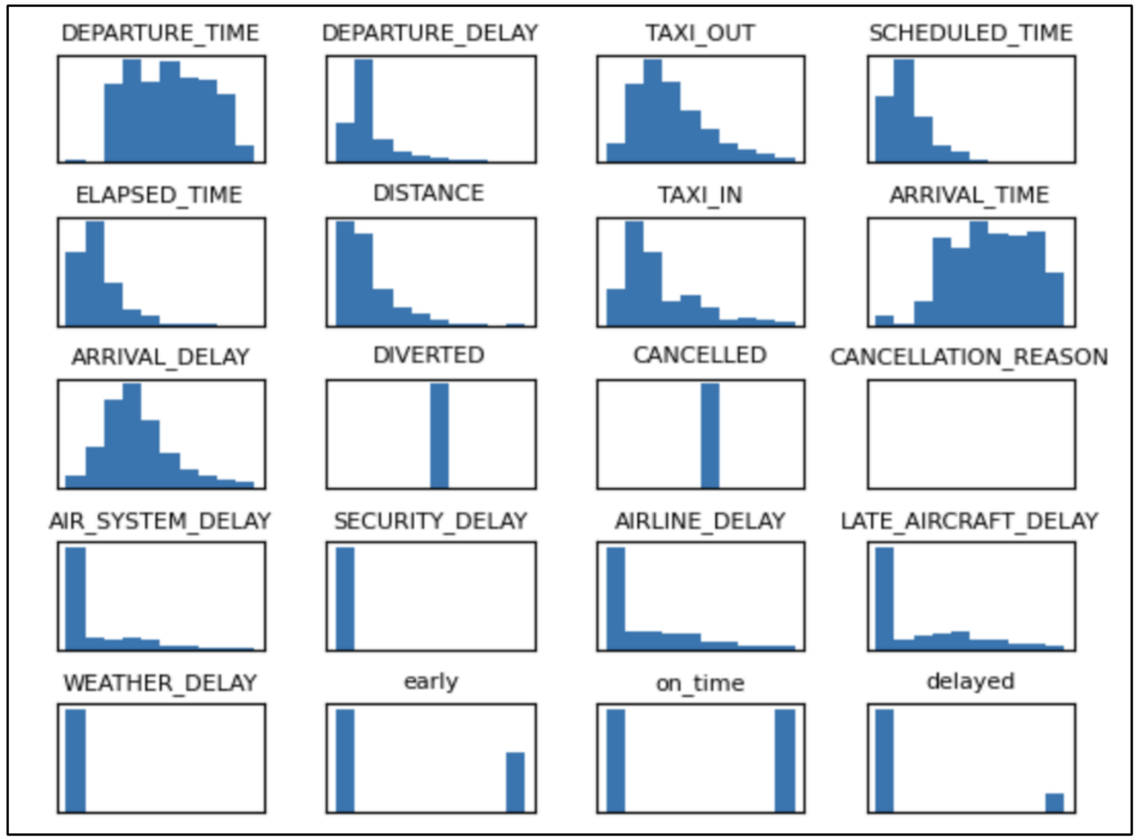
5 files downloaded

Federal Aviation Administration



Transform (1)

Data Cleaning



From the variables

- DEPARTURE_DELAY,
- ARRIVAL_DELAY
- TAXI_IN
- TAXI_OUT

top 95% and bottom 1% of the data were removed

- Flights without reference to the aircraft Master file were removed

4M Flights
14 Airlines
322 Airports
4099 Aircrafts

Transform (2)

3 new boolean attributes created

Early

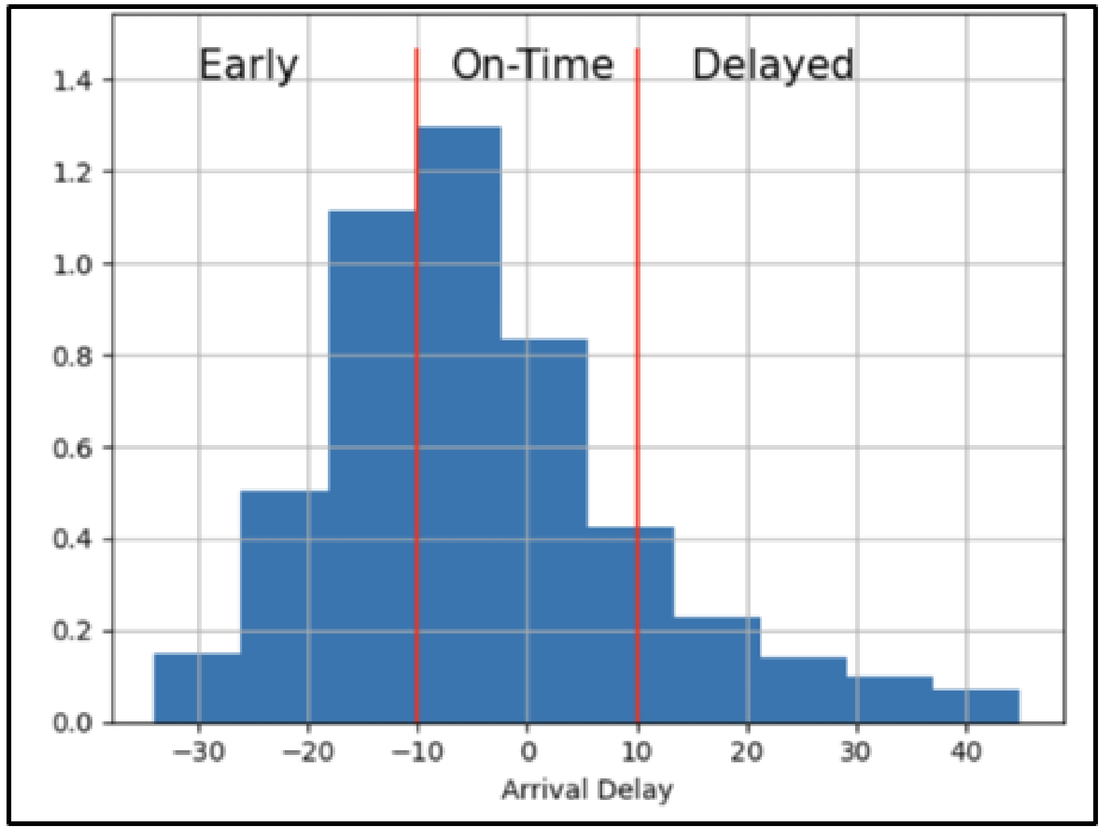
ARRIVAL_DELAY< -10

On-Time

-10 <= ARRIVAL_DELAY<= -10

Delayed

ARRIVAL_DELAY > 10



Transform (2)

8 files created (Fact + dimensions)

flights.csv

- date_dim.csv
- departureTime_dim.csv
- arrivalTime_dim.csv
- flights_fact.csv

airlines.csv

- airline_dim.csv

airports.csv

- departureAirport_dim.csv
- arrivalAirport_dim.csv

Master.txt & AcftRef.txt

- aircraft_dim

Load

Step 1

Create Tables

create the fact and dimension tables and add their foreign key relationships

Step 2

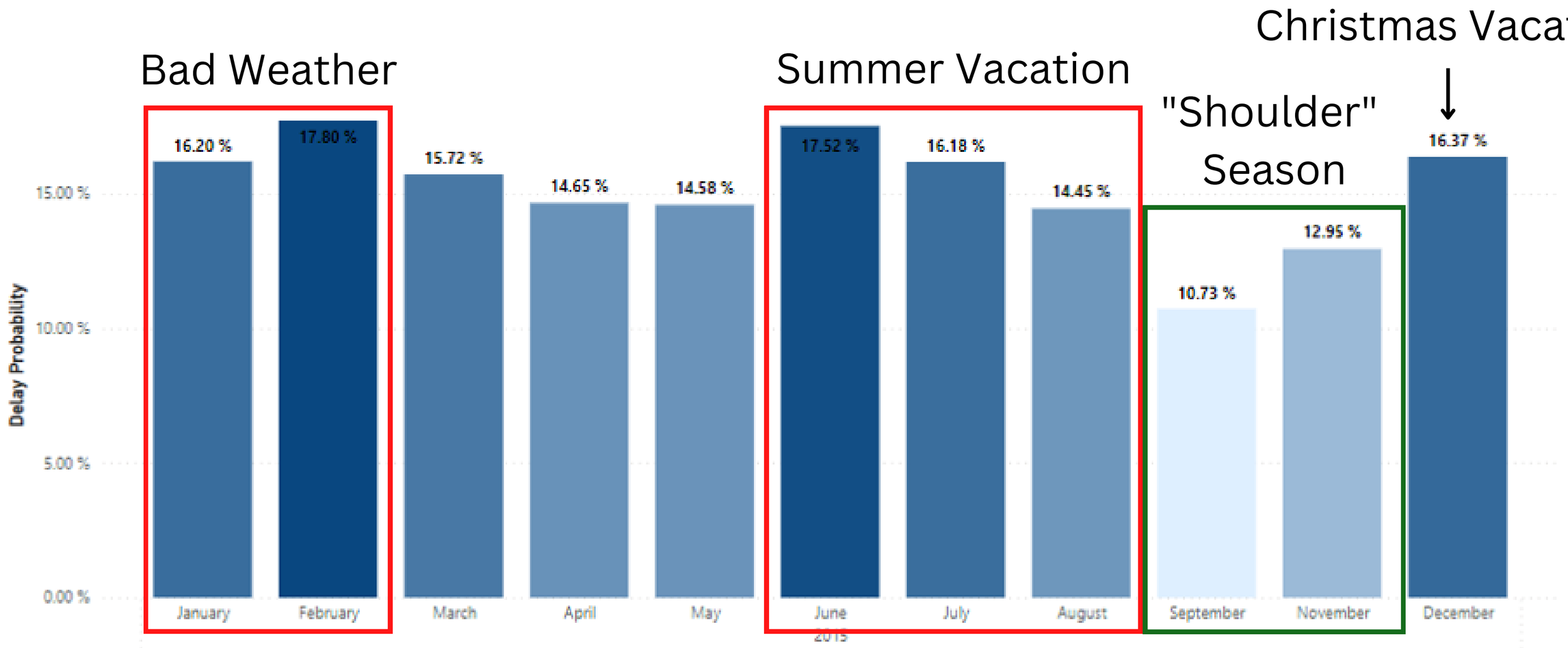
Truncate → **Import CSV**

empty the tables

read the CSV's, manipulate data types and import them in the DB

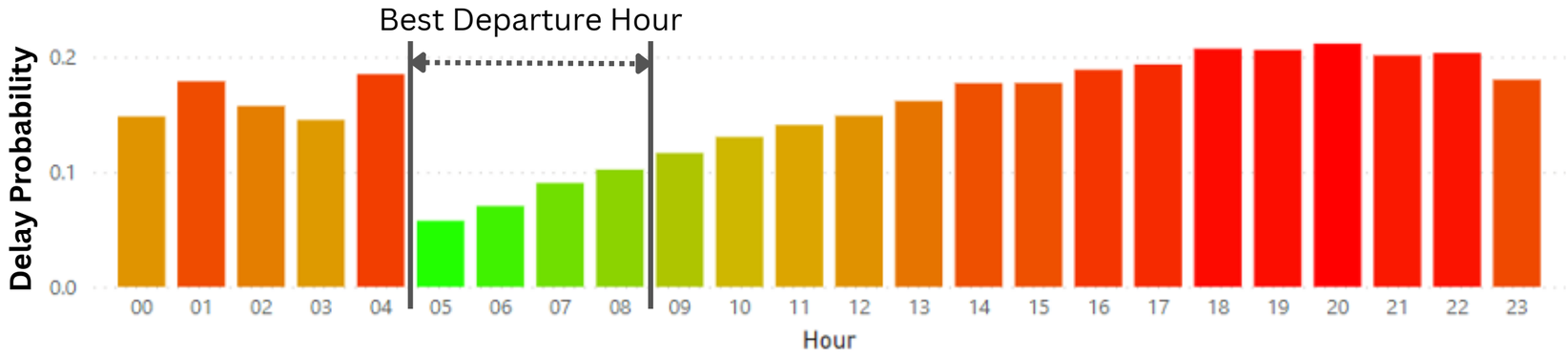
Visualizations

Best and Worse Months to fly



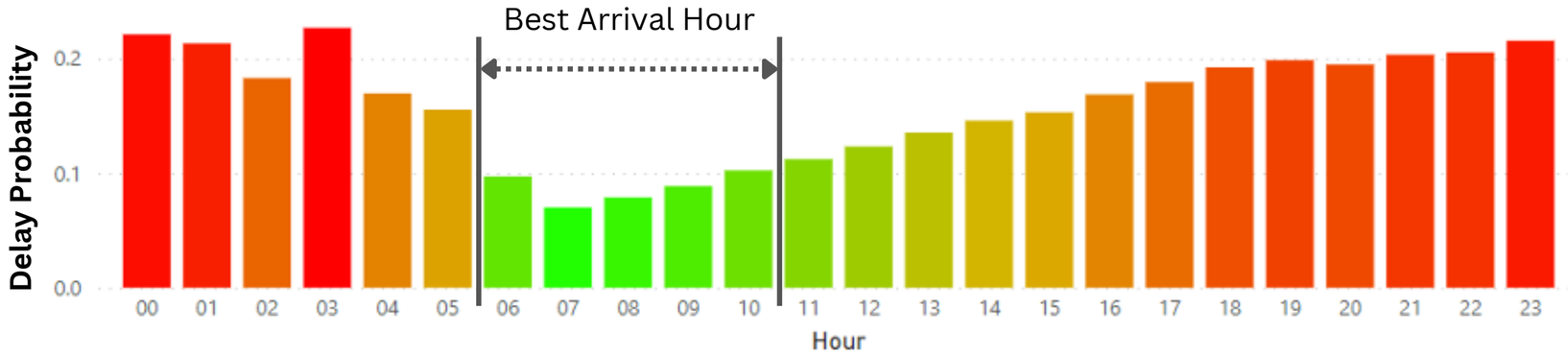
Best Take-off and Landing Time

Departure Hour



05-08 AM

Arrival Hour



06-10 AM

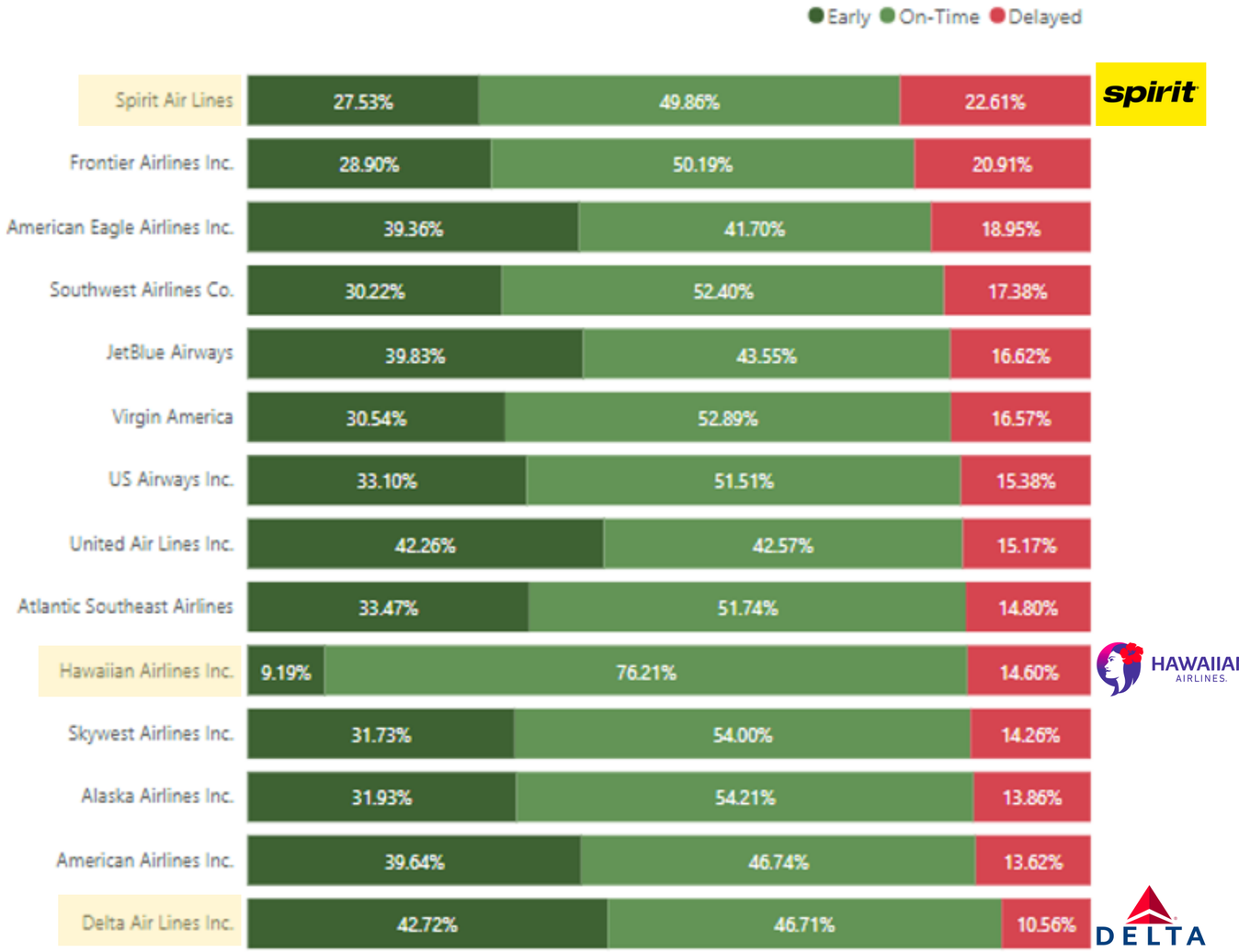
Delays within overnight hours (00-04 AM) are caused by less air traffic controllers on duty, crew availability and bad weather.

Which Airline is the best?

Delta Airlines ranks 1st in Early and On-Time performance

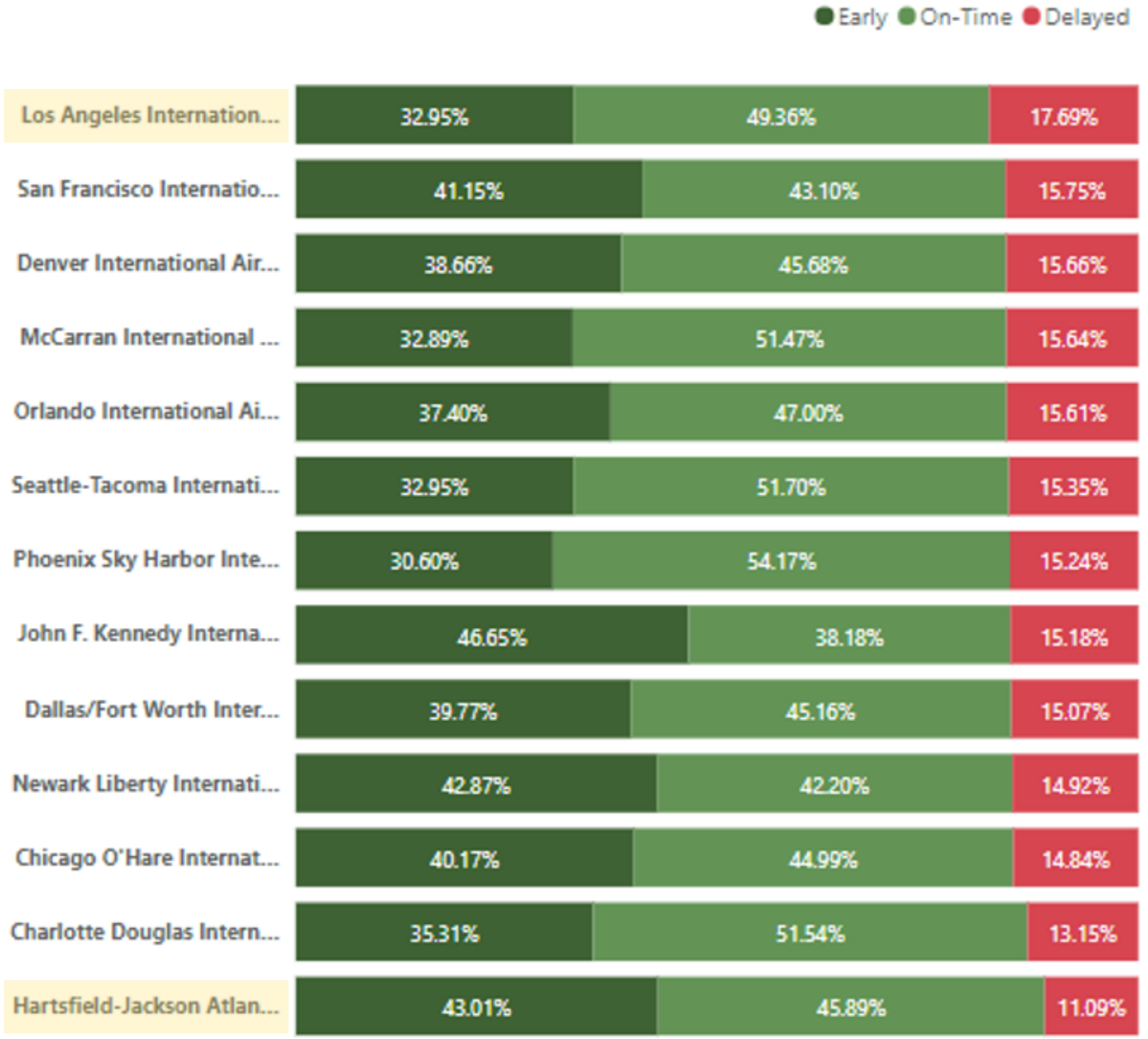
- **Spirit Air** is late with a 22% probability
- **Hawaiian Airlines** flights are mostly on time (76%)
- **Delta Airlines** flights arrive early or on time with a 89% probability

Delta Airlines ranked 2nd at the overall Customer satisfaction study in 2015 [J.D. Power]



Which is the best Arrival Airport?

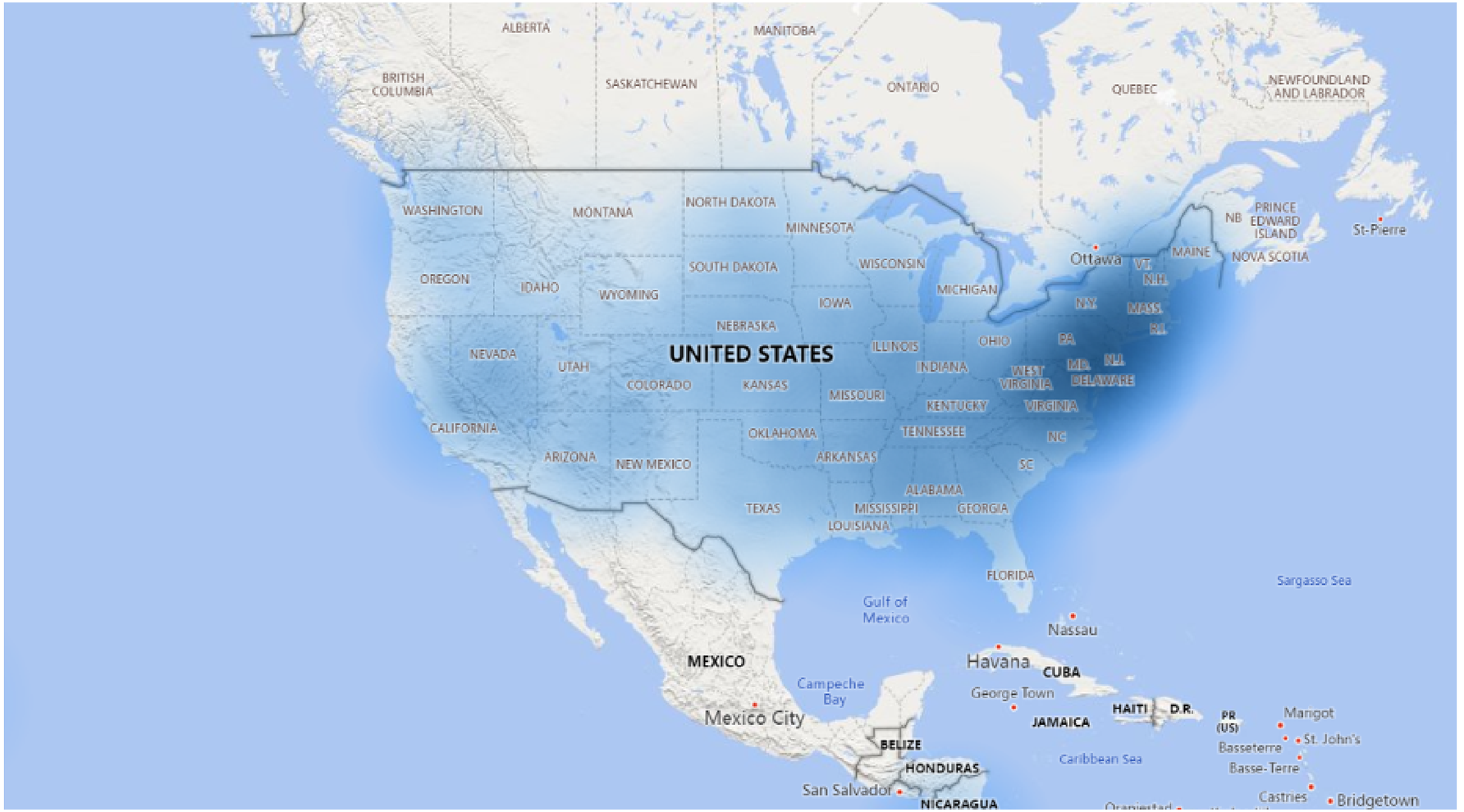
- Flights landing at the **Los Angeles International Airport (LAX)** arrive delayed with **18%** probability.
- Flights landing in **Hartsfield Jackson Atlanta International Airport** arrive mostly early or on time (**89%**)



** Presented are the Top 13 airports by passenger traffic [Forbes]*

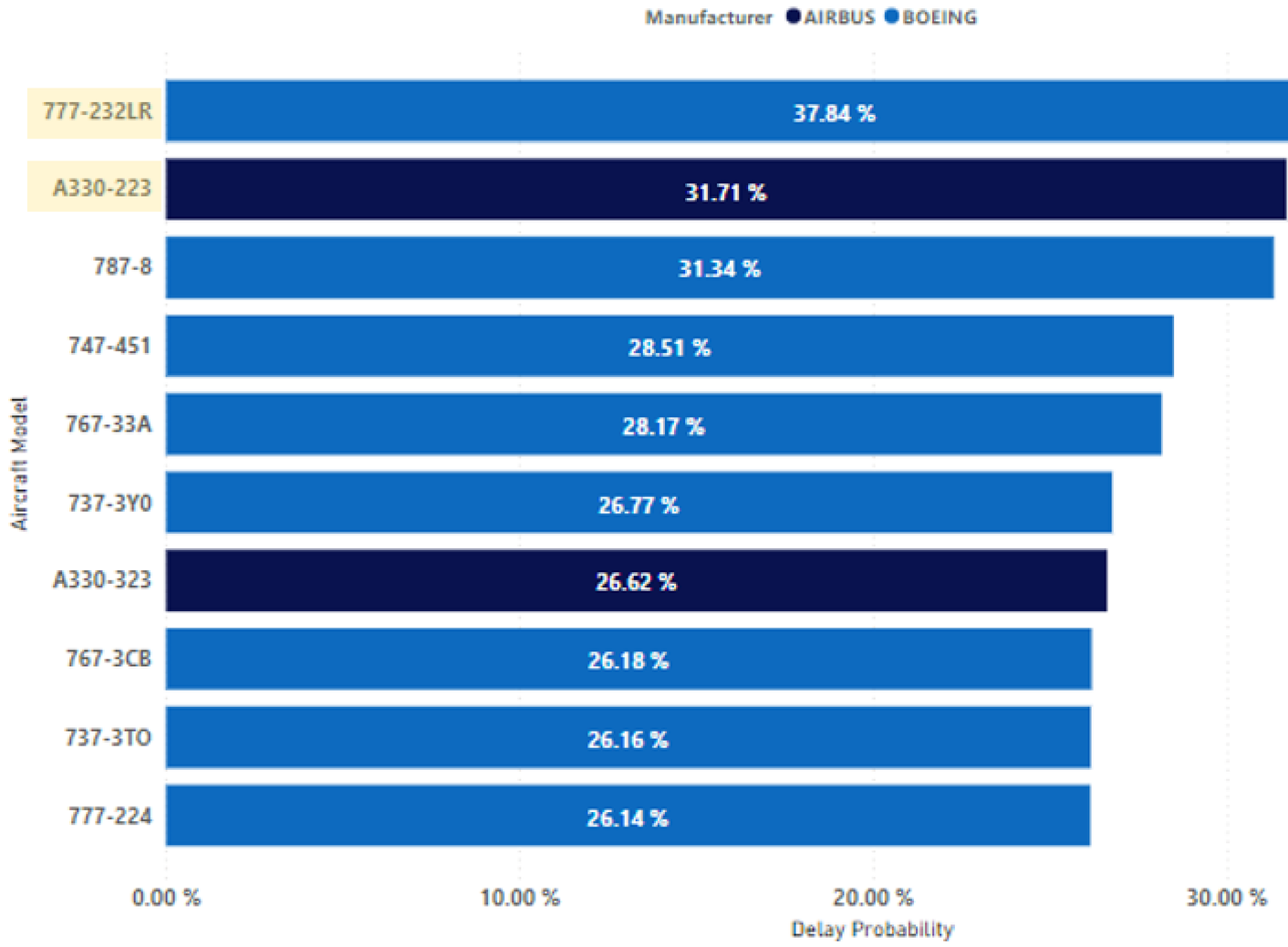
Delay Heat Map

Delay Probability rises for flights between Northeast states



Delay Probability by Aircraft

The Boeing 777-232LR has the highest delay probability



Aircraft factors that cause delays

- mechanical problems
- refueling
- maintenance

[National Business Aviation Administration]

Data Mining Models

3 Models in total

Clustering Model	→	Grouping flights and reasons of delay
Regression Model	→	Predicting minutes of delay
Classification Model	→	Predicting if a flight will be delayed or not

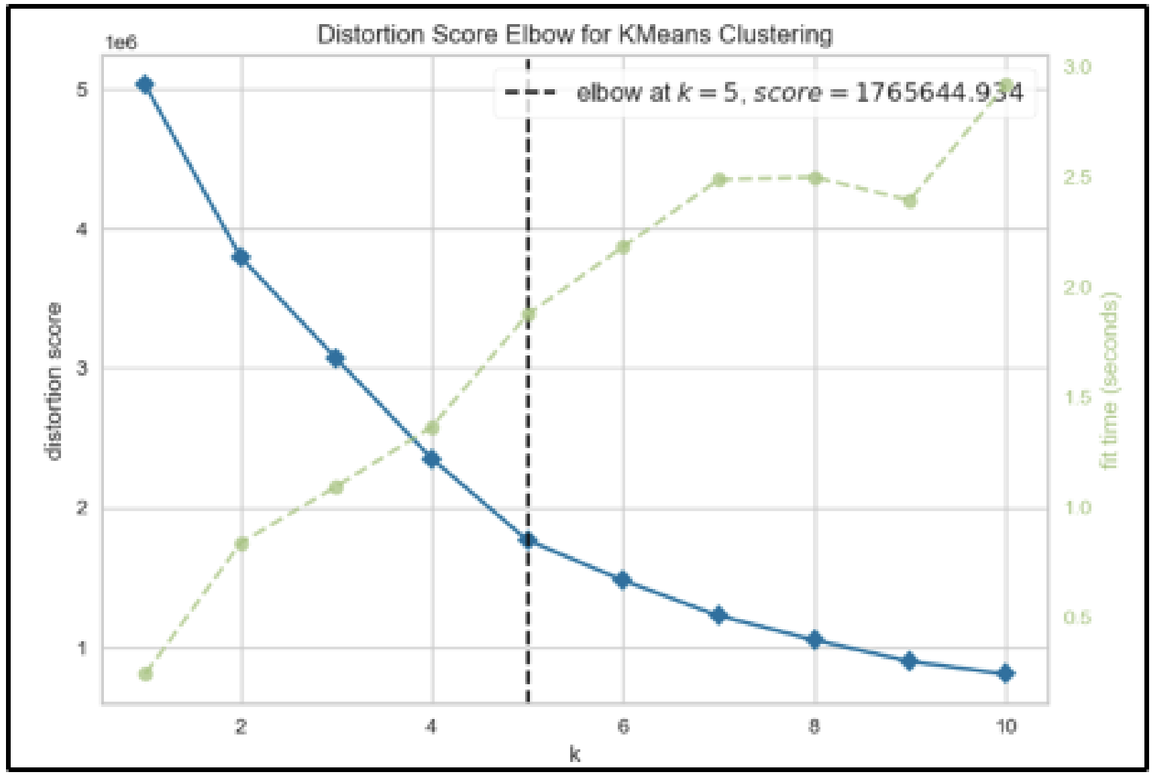
Flights Clustering (1)

Target:

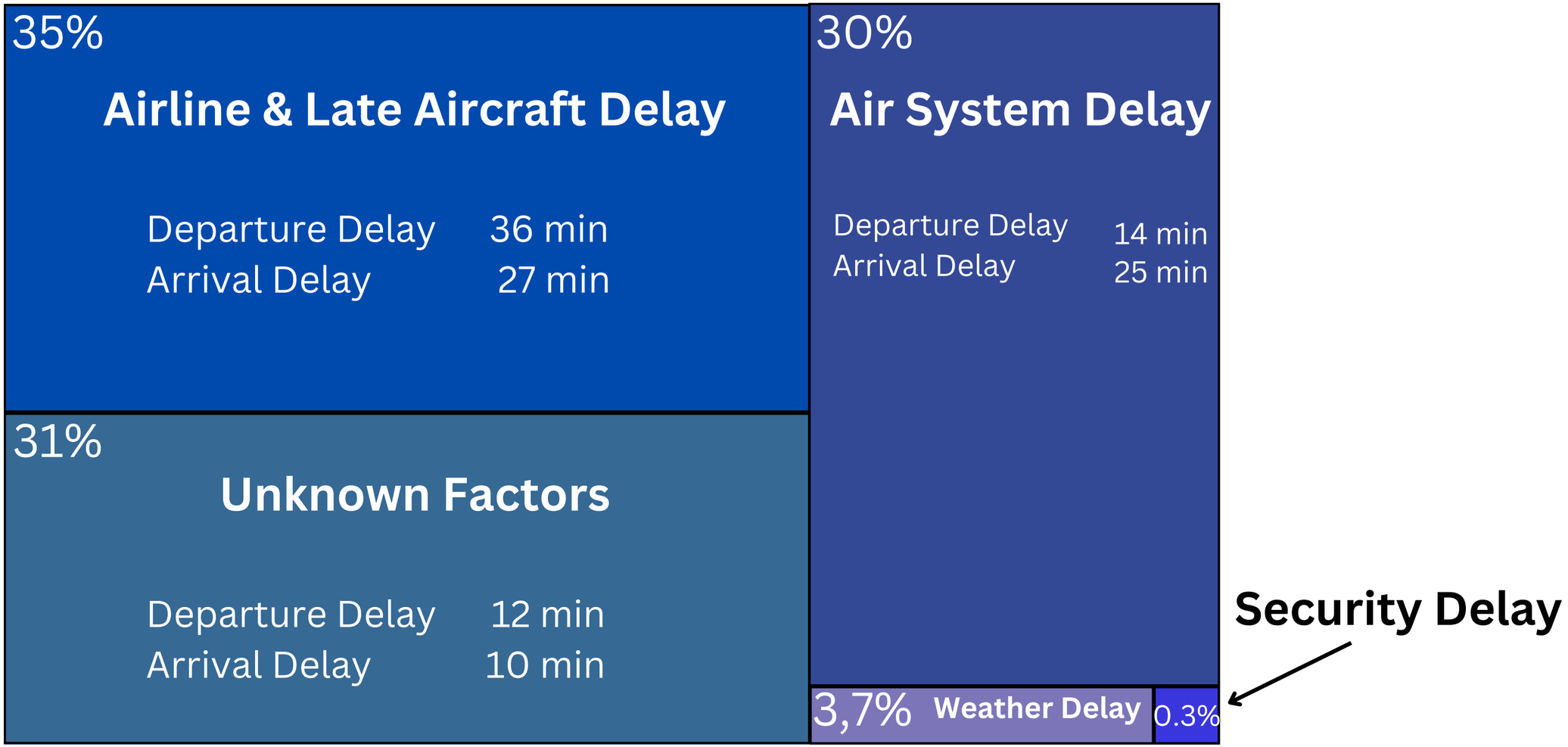
Cluster the flights and Investigate the delay causes

Attributes Used:

- DEPARTURE_DELAY
- ARRIVAL_DELAY
- AIR_SYSTEM_DELAY
- WEATHER_DELAY
- LATE_AIRCRAFT_DELAY
- AIRLINE_DELAY



Flights Clustering (2)



Regression Model

Use:

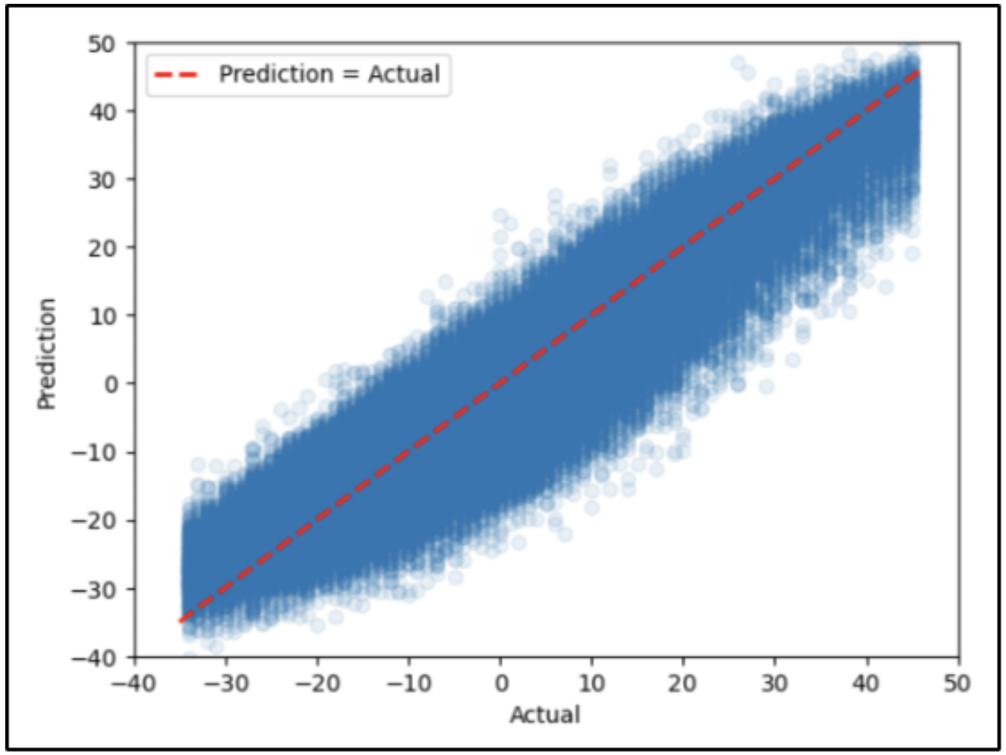
Predict the Arrival Delay based on the Departure Delay

Useful in the Arrival Airport's planning.

Model: XGBRegressor

Mean Absolute Error: 4.5 min

R: 85%



Classification Model

Goal: Predict whether a flight will arrive delayed or not

Use: More accurate planning by Airlines and Passengers

Model: XGBClassifier

Accuracy: 89%

	precision	recall	f1-score	support
0	0.94	0.93	0.93	829197
1	0.62	0.65	0.64	148620
accuracy			0.89	977817
macro avg	0.78	0.79	0.79	977817
weighted avg	0.89	0.89	0.89	977817

F1 score: 0.8873633819007033

Traveling Tips

- The least delayed airline is **Delta.**
- The best airport to avoid delays is the **Hartsfield Jackson Atlanta International Airport.**
- The best month to fly is **September**
- The best time to depart and arrive is **early morning** (04-09 AM).
- The aircraft to fly in order to avoid delays is the **Boeing 737-732**