

## **PROJECT MVP**

### **MODULE 3: DATA ENGINEERING**

#### **THE DATA:**

A dataset of 5.8m rows has been secured from Kaggle. It is a static dataset, but large enough to meet the project requirement.

The dataset contains delayed flight data of the year 2015, collected on 322 airports and 14 US airlines.

The data has been cleaned for analysis, revealing some key insights:

#### **1. Types of Delay**

The National Bureau of Transportation Statistics (or the BTS) is the branch of the National Transportation & Safety Bureau (NTSB) charged with the collection, analysis and delivery of commercial aviation statistics to the public domain.

The BTS has categorized commercial aviation delay data into 5 following broad categories:

#### **1. Air-system Delay**

Delays caused by non-severe weather conditions like head-winds, cross-winds, rain, ice and sleet, heavy traffic and national air safety regulations...these are completely beyond the carrier's control. For instance, aircraft may be put into a holding pattern due to congestion, or diverted to go around a weather system in the air etc..

#### **2. Weather Delay**

Delays caused by severe weather systems like hurricanes, tornadoes, blizzards, fog etc..The carrier has absolutely no control over this.

#### **3. Security Delay**

Delays caused due to slow nonexistent or malfunctioning security screening equipment, resulting in much longer boarding times etc... This does impact the airline delay score even though it is not counted against the airline since a carrier does not secure its own passengers. You can't have your cake and eat it.

#### **4. Airline Delay**

Covers various delays caused due to off-loading aircraft, loading aircraft, cleaning the aircraft, misplaced baggage, passenger boarding (post security check), delays caused due to connecting passengers, deplaning of passengers due to security violations on-board, ill patients etc..

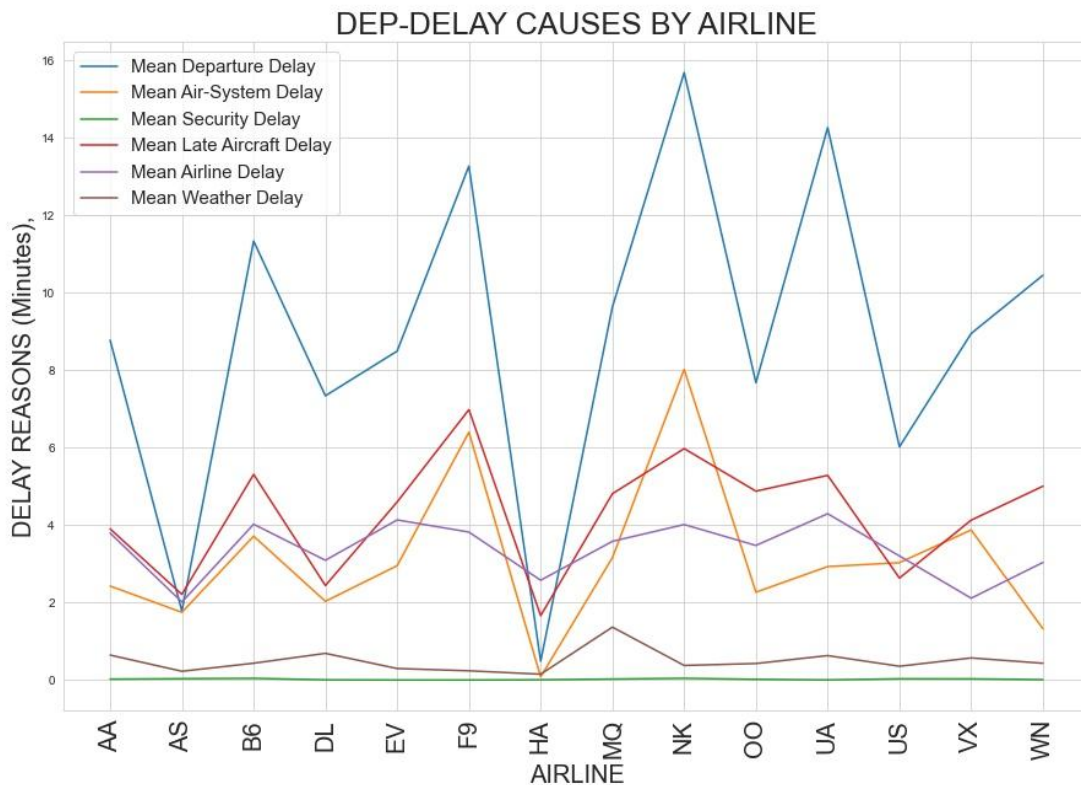
#### **5. Late Aircraft Delay**

This type of delay is caused by a delayed flight whose equipment (aircraft) is going to be used for a subsequent flight.

The first 3 are considered to be beyond the carrier's control while the last 2 are deemed to be firmly within the airline/carrier's ability to minimize.

## EDA Insights:

The most tardy airline of the year on Mean Departure Delay was Spirit Airlines(NK) with a mean delay of 15.8min while the most punctual (or least tardy) was Hawaiian Airlines (HA), coming in just over 0.5min) :

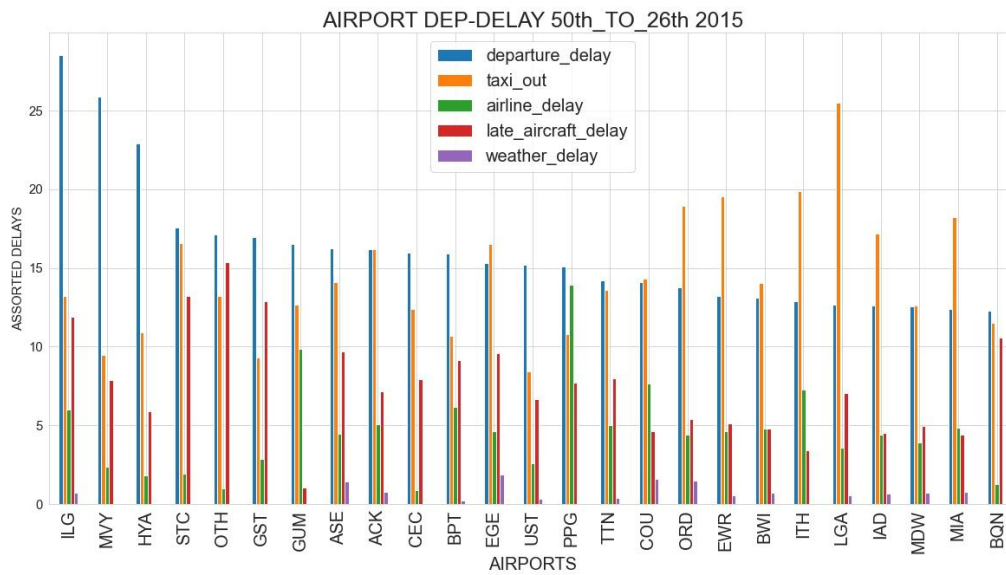


The major causes of departure delay were late aircraft delay and airline delay and air\_system delays.

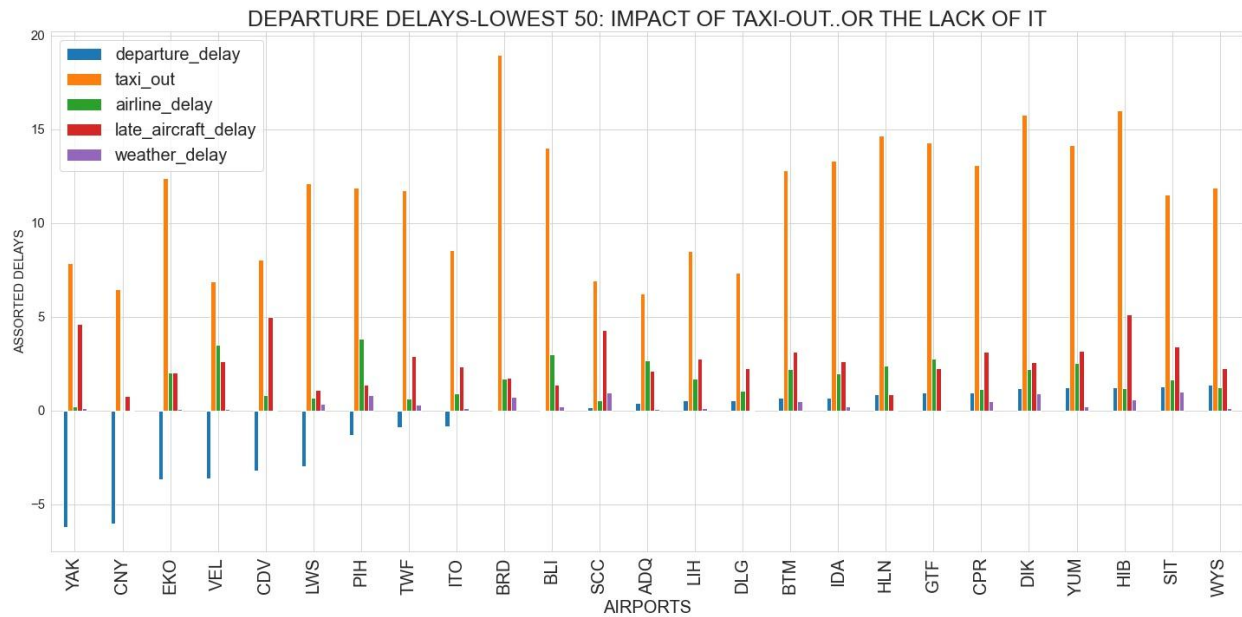
The worst performing airport of the year on Mean Departure Delay was ILG- Wilmington International, Wilmington DE with a mean departure delay of 28min. (see plot below)

The above-factors have been plotted for the graph below as well as Taxi-Out, since taxi-ing is something every aircraft must perform to access the run-way before becoming airborne.

And yes, taxi-out does seem to be a significant factor to departure delay. But is it??



Taking a look at the airports with the best departure ranking (smallest departure delays);



As it turns out, taxi-out in the best performing is just as high as in the worst performers.

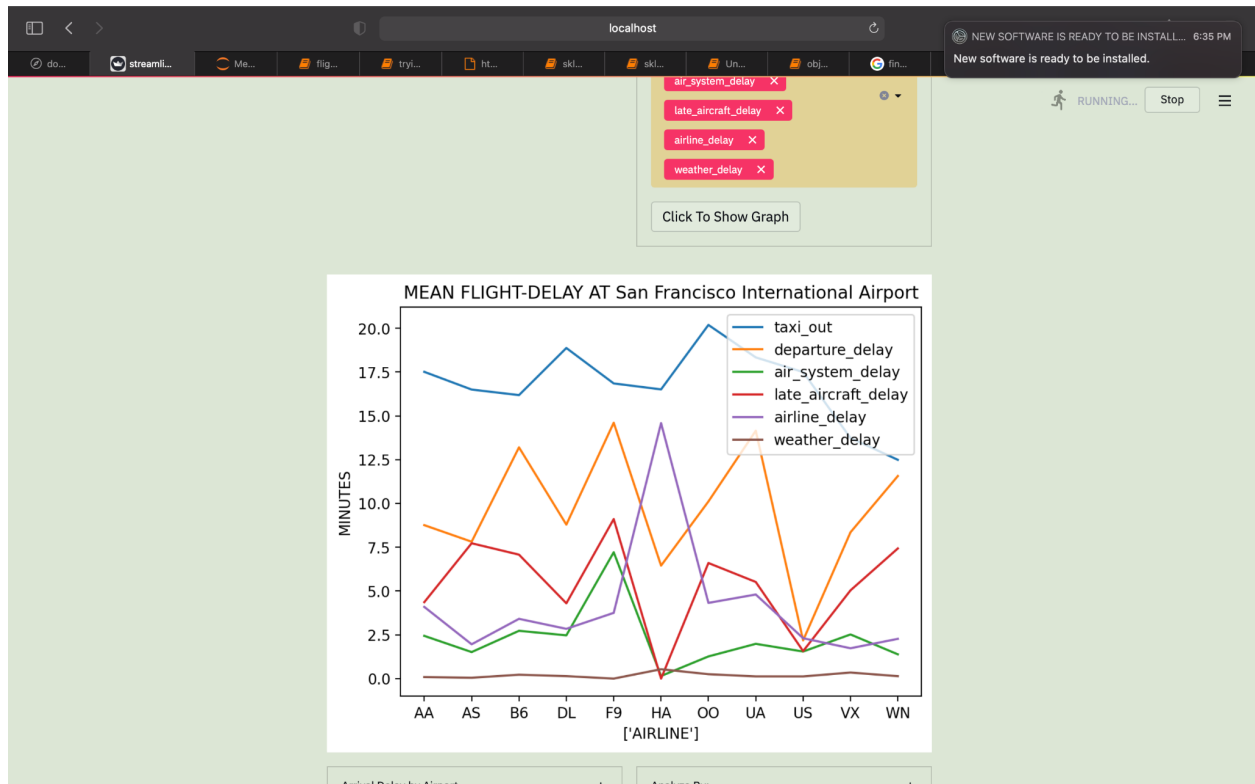
**WEB-APP**

A Streamlit app has been built to pass the above and other queries to the database in real-time. Here, we just took a look at airline performance at Ted Stevens International, Anchorage Alaska, with metrics selected from a multi-select drop-down menu.



United Airlines encountered 26min delays at Anchorage, mainly due to late-aircraft delay.

A similar performance metric can be extracted for SFO - San Francisco International, with more metrics though:



At SFO, it is United(unfortunately) and Spirit Airlines are high on the delay rankings with a mean 15min delay.

Major causes are Airline Delay, Late-Aircraft delay, probably due to mid-air delays, since Air\_System Delay is also quite high.

### Conclusion:

We have the data, as well as a properly working template to meet the objectives of our project. Our project is on track.

Dennis Ssekamaanya

Metis Engineering - May17th to May28th 2021.