Airlines Operations Scenario

Airline CFO or SVP Strategy/Operations:
"Departure delays are causing us a lot of financial loss. I want our operations-technology team to look into potential solutions for decreasing losses caused by delays."

Director Operations:
" Departure delays are causing us a lot of financial loss. Big boss wants us to look deeper and come up with potential solutions."

Manager Operations Technology & Team:
"Let's breakdown the problem and look at what's causing the losses. If we do, we can figure how to mitigate or manage delays, as well as improve customer satisfaction."

The team comes up with the following:

Top Reasons for Delays	*Contribution to Financial Loss*	*Controllable? Yes/No*
Carrier Delays	x %	yes – Mitigation Plan
Weather Delays	y%	no - Management Plan
NAS_Delay	z%	no – Management Plan
Security Delay	a%	no – Mitigation Plan
Late Aircraft Delay	b%	no - Management plan

Of these reasons for delay, some can be controlled, such as carrier delays (perhaps due to unforeseen repair/maintenance and such)

Flight delays create problems in scheduling for airlines and airports, leading to passenger inconvenience, and huge economic losses. As a result there is growing interest in predicting flight delays beforehand in order to optimize operations and improve customer satisfaction. In this project, you will be predicting flight delays using the datasets provided. For now, the problem to be tackled in this project is framed as follows:

"Predict a departure delay or not where a delay is defined as 15-minute delay (or greater) with respect to the planned time of departure. This prediction should be done two hours ahead of departure (thereby giving airlines and airports time to regroup and passengers a heads up on a delay)."

Airlines Data

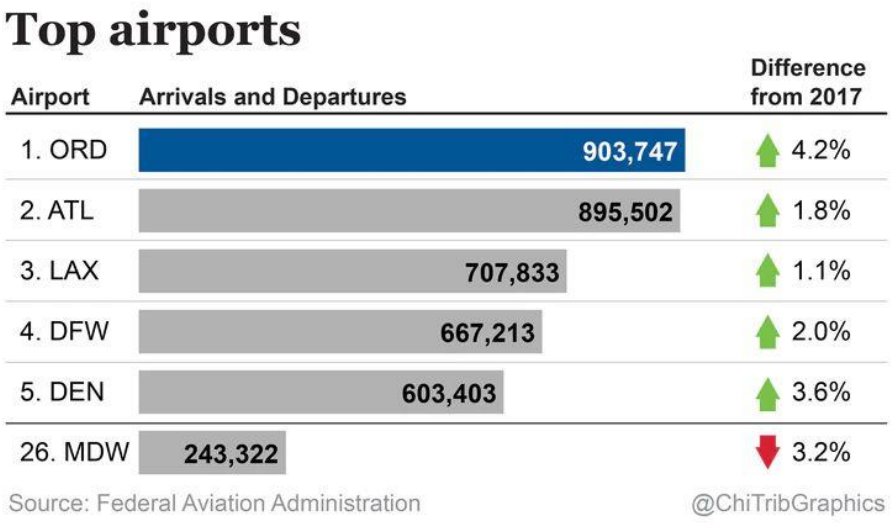
The data for this project comes in the form of two (BIG) tables:

1. **Flights table**: This is a subset of the passenger flight's on-time performance data taken from the TranStats data collection available from the U.S. Department of Transportation (DOT) from 2015-2017. http://www.transtats.bts.gov/DL_SelectFields.asp?Table_ID=236&DB_Short_Na me=On-Time

   A Data Dictionary for this dataset is located here: https://www.transtats.bts.gov/Glossary.asp?index=C

2. **Weather table**: Flight departures (and arrivals) often get affected by weather conditions, so it makes sense to collect and process weather data corresponding to the origin and destination airports at the time of departure and arrival respectively, and build features based upon this data. A weather table has been pre-downloaded from the National Oceanic and Atmospheric Administration repository to S3 in the form of parquet files (thereby enabling pushdown querying and efficient joins).

The following table summarizes the busiest airports in the US in terms of flight arrivals and departures. On average some of the busier airports handle 1,500 arrivals and 1,500 departures on average per day, leading to daily transaction files of 3-5MB in size.



In 2018, O'Hare also handled more than 83.4 million passengers and on average 1,500 flight arrivals and 1,500 flight departures per day.

[Source: https://www.chicagotribune.com/news/breaking/ct-biz-ohare-flight-numbers-20190204-st ory.html]