

Project #2 Luther Proposal for Po-Yan Tsang

Goal

How much delay is your flight likely going to have? The goal of this project is to predict flight delay time for a given flight (flight #, date).

Data Sources

- 2015 Flight delay and cancellation data from <https://www.kaggle.com/usdot/flight-delays#flights.csv>
- 2015 airport volume data scraped from Bureau of Transportation Statistics website: <https://www.transtats.bts.gov/airports.asp?pn=1>
- Historic airport weather data scraped from Wunderground website: <https://www.wunderground.com/history/daily/us/wa/seattle/KABE>

Methodology

1. Filter out 2015 flights data to only departing flights that are NOT cancelled and with departure time data
2. Merge airport volume data and weather data with flight data
3. Build linear regression model with the merged flights and airport data to predict flight delay time for a given flight

Features

1. Airport - # departure passengers for airport where flight departs
2. Airport - # scheduled flights per day for airport where flight departs
3. Airport - # passengers handled for the airline at given airport where flight departs
4. Flight - DAY
5. Flight - DAY_OF_WEEK
6. Flight - SCHEDULED_DEPARTURE
7. Flight - AIR_TIME
8. Flight - DISTANCE
9. Flight - Airline
10. Airport weather- wind speed
11. Airport weather- precipitation
12. Airport weather- temperature