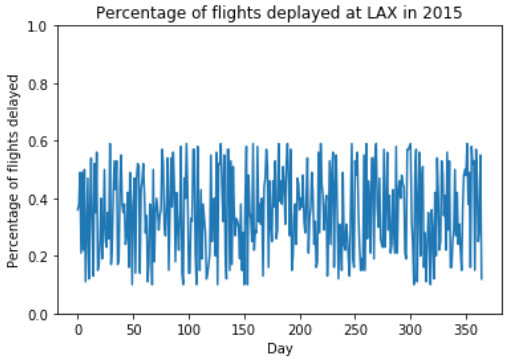
Project Proposal: Flight Delay Analysis and Prediction

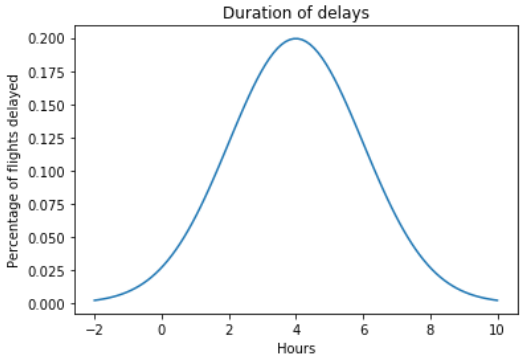
ECE 180 Team 4
Jinhao Tang, Sam Vineyard, Yuhan Chen, Siyi Wang, Zexi Chen

This project aims to predict the possibilities of flight delays, which may help passengers in the US choose their flights to minimize delays. People suffer from flight delays especially when they have urgent need to arrive at the destination on time. This project can serve as a guide for them to avoid such problems. We obtained the data from Department of Transportation(DOT). As all the flights in US need to be registered in DOT, the data is very reliable.

To address our purpose using the data, we first need to download the data from DOT website. We intend to use 7 related features in the dataset: 1) Date of the flight 2) Airline of the flight 3) Original and Destination of the Airport 4) Cause of the delay 5) Distance of the flight 6) Scheduled Departure and Arrival time 7) Delay Duration. We choose those features with respective reasons as following: a) The flight traffic can vary on different days. Traffic can be extremely dense on holidays, thus leading to the delay. b) Different airlines have different target customers and value differently on flight delays. c) Large cities may have more flights, which often indicates higher probability of delay. d) The cause of delay can give user an idea of how to avoid delay. e) According to the lengths of flight, the delay distribution can vary. f) Different flight time during the day also leads to different flight delays based on flight traffic. h) The delay duration can help customers decide to what extent they can bear the delay. The Python libraries we will be using are: Numpy, Pandas, Matplotlib, basemap and Scikit-learn. Matplotlib can help with the visualization. basemap deals with associating geometric data with flights. Numpy and Pandas help to organize data and extract the statistics information of the data in a fast fashion. Scikit-learn will provide machine learning algorithms that we might use.

Some examples of graphs that we expect to see are shown below:





Timeline for this project:

Time	Jinhao	Sam	Yuhan	Siyi	Zexi
11/14 - 11/20	Download Data	Organize Data	Make Tables	Examine Data	Examine Data
11/21 - 11/27	Write code	extract features	Write code	Extract Features	run functions
11/28 - 12/2	Get results	make visual graphs	make visual graphs	make visual graphs	Get Results
12/3 - 12/4	Prepare for presentation	Prepare for presentation	Prepare for presentation	Prepare for presentation	Prepare for presentation