Prediction and Analysis of Commercial Airline Delays Based on Atmospheric Conditions

ECE 180 Final Project Proposal

Team 15: Zi Deng, Ruixian Liu, Junhao Su, Hannah-Marie S. Valenzuela, Hsiao-Chen Huang

I. Introduction

Commercial flights are in integral part of modern society, over 600 millions commercial airline flights happen every year in America. Thus, any relevant insights regarding air travel can result in a big impact in our lives. The main purpose of our project is to provide accurate predictions of commercial aircraft flight delay percentages based on future weather forecasts. Overall, the prediction model would allow passengers to schedule plans with a better expectation of flight times as well as to make more informed decisions regarding air travel purchases.

II. Datasets and Quantitative Analysis

The datasets we will be using are the "Airline On-Time Statistics and Delay Causes" public database from the Bureau of Transportation Statistics as well as the "Historical Weather" public data from Weather Underground. Furthermore we will be doing data scraping on Weather Underground for both past weather reports and future weather forecasts.

Data of airlines - US Department of Transportation, Bureau of Transportation Statistics:
https://www.transtats.bts.gov/DL_SelectFields.asp?Table_ID=236&DB_Short_Name=On-Time

Data of weather - Weather Underground:
https://www.wunderground.com/

III. Proposed Framework

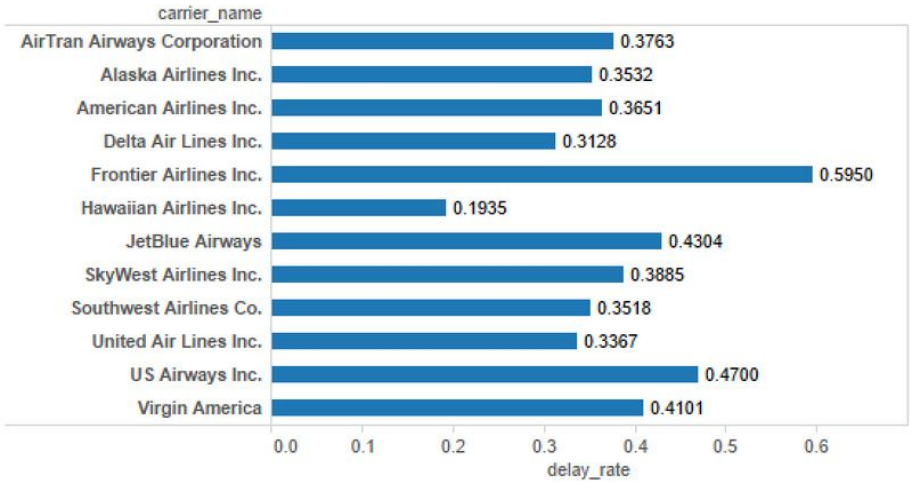
We will analyze the delay of flights versus the various conditions of weather, airline companies, dates, airport size and so on. Using relative libraries of Python, we can model the relationship between weather condition and flight delays and then use this model to predict future delays based on new weather forecast data. We will be specifically looking at data from 5 large cities, 5 medium sized cities, and 5 small cities to create a robust model that accounts for size and location variations.

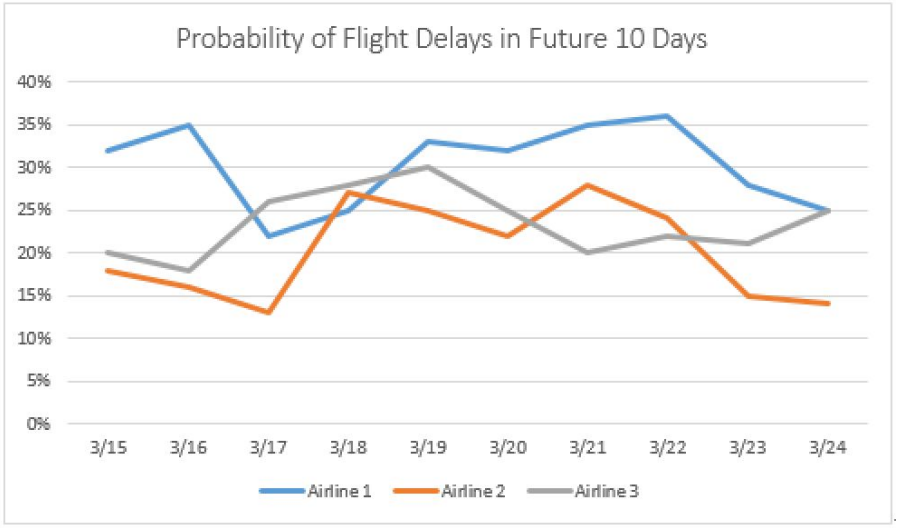
IV. Implementation and Expected Result

For data scraping, cleaning and formatting, we are going to use Python libraries such as Urllib2, BeautifulSoup, Numpy, Pandas which can easily help to collect and integrate "big data". Then, using Matplotlib, basemap to help present those data in the form of statistical charts. Finally, implement the machine learning algorithms to predict the probability of flight delays base on the conditions of weather.

The main goal in this project is to predict the flight delays in the future 10 days.

Example of Data Visualization we expect to achieve:





V. Timeline

Task	Expected Time	Assigned Member
Collecting Data and Data Scraping	2/07/18 - 2/14/18	Zi, Junhao
Analysing Delays vs Multiple Factors	2/14/18 - 2/21/18	Ruixian, Hsiao-Chen
Making Models for Prediction	2/21/18 - 2/28/18	Ruixian, Junhao, Hsiao-Chen
Validating the Robustness of Models and Results & Preparing Presentation	2/28/18 - 3/07/18	Zi, Ruixian, Hsiao-Chen, Junhao, Hannah-Marie