

CSX415 Project : Airline Delays and Cancellation Analysis

Ajey Patil, ajeypatil@gmail.com (mailto:ajeypatil@gmail.com)

04/15/2018, 04/28/2018

1 Objective

The Objective of this project is to analyze Flight Delay and Cancellations for US Airlines and determine best airline for least flight delays and cancellations on the 100 busiest routes in country

The project analyzes flight data from The U.S. Department of Transportation's (DOT) Bureau of Transportation Statistics and predicts Airline Performance based on target variables of cancelled number of flights and average flight delays (which is not due to airport, weather, security or air system delays) and when delays are only due to airline delays.

2 Overview

When multiple airlines are available in metro area, then residents have choice of airline. An analysis of airlines performance in terms of flight delay and cancellation is required for passengers to choose an airline with least cancelations and delays on the 100 busiest routes in country.

Decision Prediction The predictions of best airline of choice in US would be made based on how busiest airline is in terms of number of flights and relative delays and cancellations per flight and 100 busiest routes would be selected based on sorted (unique airport of source and destination) list of flights between 2 airports.

2.1 Descriptions of Variables

Variable Description	Type	Variable Category
Airline	Target	Categorical
Dep Date	Predictor	Ordinal
Dep Time	Predictor	Continuous
Air Time	Predictor	Continuous
Distance	Predictor	Continuous
Arrival Delay	Predictor	Continuous
Weather Delay	Predictor	Continuous
Weather Condition	Predictor	Ordinal
Security Delay	Predictor	Continuous
Air System Delay	Predictor	Continuous
Late Aircraft Delay	Predictor	Continuous
Cancellations	Target	Continuous
Airline Delay	Target	Continuous

3 Data Sources

3.1 Flight Data

The U.S. Department of Transportation’s (DOT) Bureau of Transportation Statistics tracks the on-time performance of domestic flights operated by large air carriers. Summary information on the number of on-time, delayed, canceled, and diverted flights is published in DOT’s monthly Air Travel Consumer Report and in this dataset of 2015 flight delays and cancellations. <https://www.kaggle.com/usdot/flight-delays/data> (<https://www.kaggle.com/usdot/flight-delays/data>)

3.2 Weather Data

The weather data is obtained from Hourly land-based weather observations from NOAA at <https://www.ncdc.noaa.gov/orders/qclcd/> (<https://www.ncdc.noaa.gov/orders/qclcd/>), This source contains hourly weather data at various airports.

4 Key Roles

4.1 Model Creation

1. Model Creator : Ajey Patil

4.2 Model Experts

1. The U.S. Department of Transportation’s (DOT) Bureau of Transportation Statistics

4.3 Model Users

1. Airline Passengers: this analysis will help them choose airline depending on airport performance of flight delays and cancellations
2. Airport management: this report will help them determine systemic and process related issues affecting flight delays and cancellations
3. Airline management: this report will help them determine best performing airports in terms of flight delays and cancellations and their reasons in scheduling and flight route planning.
4. Regulatory boards: this analysis will assist in identifying consistently low performing airlines in terms of flight cancellations and delays and their reasons

5 Current State and Reasons for Project Consideration:

This project was considered for analysis due to increasing number of complaints from airline passengers about the delays and cancellations of certain airlines and make predictions available to passengers of an airline based on past performance.

6 Deliverables

The deliverables for this project will include a report of all the findings, r code, all of which will be published on GitHub.

Specific analysis to be done and questions to be explored

Chart of all airlines and number of flights cancelled per day of week.

Chart of all airlines and number of flights average delays per day of week.

Chart of average delay time and all airlines and day of week.

Chart of Category the delay reason for flights for airline and numbers of delays per delay reason

Based on predictor variables mentioned in table above,
predict best airline for minimum flight delay and cancellations.

7 Success Criteria

Success criteria of project:

Benefits of the Project: Airline Passengers will be able to choose best airline and Airline Management will be made aware of quality of their service

Project Success Metric: When the analysis is done year-or-year, the improvement in average delays and cancellations per airline is measurable and is expected to reduce 10% year-over-year

Model Analysis Success Metric: The model analysis and prediction should predict airline delays and cancellations with accuracy greater than 75% and ROC(AUC) greater than 0.65

8 Risks

1. Availability of data in required format for different years of observations of data
2. Project Risks: Airlines cannot improve improve delays and cancellations year-over-year
3. Prediction Risks: Model Analysis and Prediction cannot predict with accuracy greater than 75% and ROC(AUC) greater than 0.65 then more data is needed

9 Deployment Phases

Initial Analysis Phase: R-code for preliminary analysis of data

Initial Prediction Phase: R-code for preliminary prediction of deliverables

Final Model Deployment Phase: R-code to generate reports for yearly data

Deployment to Airline Passengers Impact: This model would be deployed to all airline passengers as yearly report and they can make choice of airline for up-coming year based on report.

Risks of Deployment: The deployment of model as yearly report may be rejected by airline and hence analysis performed must be made public to airlines which can then verify the scientific analysis of the data.

10 Timeline of Project

1. Data Availability: Yearly data made available The U.S. Department of Transportation's (DOT) Bureau of Transport Statistics
2. Prediction Report Availability : One month after yearly data is made available
3. Effort Required for Predictions : One man-month

11 Costs

To be estimated