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 predict flight delays and cancellations

The project analyzes flight data from The U.S. Department of Transportation's (DOT) Bureau of Transportation Statistics and predicts flight delays greater than 60 minutes or cancellations based on origin airport, destination airport, day, month of year, time of depature and airline.

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.

2.1 Analysis and Decision Prediction

The analysis of delay and cancellations would be made based on the following predictors

2.2 Descriptions of Variables

```
----------------------------------------------------------------
| Variable Description  |  Type      |  Variable Category   |
----------------------------------------------------------------
|  Airline              | Predictor  |  Categorical         |
|  Departure Time       | Predictor  |  Continuous          |
|  Dep Day of Week      | Predictor  |  Discrete            |
|  Departure Month      | Predictor  |  Discrete            |
|  Origin Airport       | Predictor  |  Categorial          |
|  Destination Airport  | Predictor  |  Categorial          |
|  Cancellations        | Target     |  Boolean             |
|  Airline Delay > 60min| Target     |  Boolean             |
----------------------------------------------------------------
```

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)

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 canceellations and their reasons in scheduling and flight route planning.
4. Regulatory boards: this analysis will assist in identifying consistenly 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 an application running in Rscript that predicts Airline delay or cancellation based on input predictors

Specific analysis to be done and questions to be explored

```
Chart of all airlines and number of flights cancelled.

Chart of all airlines and number of flights average delays.

Based on predictor variables mentioned in table above,

Application to predict airline delay greater than 60 min 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 70% 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 70% 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-application to predict delays and cancellations

Deployment to Airline Passengers Impact: The final application would be deployed to all airline passengers on a yearly basis.

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

10 Timeline of Project

1. Data Availablity: Yearly data made available by The U.S. Department of Transportation's (DOT) Bureau of Transport Statistics

2. Prediction Application Availability : One month after yearly data is made available

3. Effort Required for Intial Model Development : One man-month Effort Required for Yearly model retraining and application deployment : One man-month

11 Costs

To be estimated