DSC540 - Milestone 1

Airlines On-Time Performance, Delays, Cancellations and Diversions

Introduction:

> Airline cancellations or delays are one of the major causes for passenger inconvenience. With the publicly available dataset, using datascience I am hoping to gain meaningful insights into the best performing airlines and understanding the causes for delays, diversions and cancellations across different airline carriers. For the final project I would like to analyze airline data to identify different factors and their effects on a carrier's performance. As a performance measure, I would like to explore on-time arrivals, number of cancellations by carrier and also explore different reasons for delays and diversions. Based on the outcome, carriers can take necessary actions to focus on the problem areas.

Data Source:

- Flat File: Excel files from BTS. The excel data has airline performance factors such as cancelled, diverted, delayed and on-time data. The downloaded raw data has upto 34 columns. https://www.transtats.bts.gov/OT_Delay/OT_DelayCause1.asp?20=E (Download Raw Data link for data).
- API: API provides historical weather information. https://visual-crossing-weather.p.rapidapi.com/history?startDateTime={}&aggregateHours=24&location={}&endDateTime={}&unitGroup=us
- Website: Website consists a list of diverted flights. https://www.diverted.eu/

Relationships:

> Flat file is the main data source with scheduled flight information. Flat File - API Data from the flat file has cancellations and delays due to weather. I would like to lookup the weather information for the flight date at the origin/destination of flights cancelled or delayed due to bad weather. The Bureau data has upto January 2023 data. In order to lookup weather for a past date, I would need the historic weather data. The API gets the historic weather data for a location (origin or destination city name). This will enable to validate if there truly was a bad weather situation for a flight to be delayed or cancelled. With this, we can also identify the cause of bad weather like storm,snow,wind,etc. Flat file has many to many relation with the API. We will need to pass flight date and the origin or destination city to the API to get weather information for a particular date and place.
> Flat File - Website Flat file has a column for diverted flights but does not have any information on the cause for diversion. I would like to lookup the reason for a flight being diverted. The website and flat file can be matched on flight date, origin and destination to lookup diverted flight information. Flat file has many to many relation with the Website. We will need to pass flight date and the origin and destination city to the website to get flight diversion details for a particular date and route.

Project Subject Area:

> The project aims on idenifying various performance measures in airline operations. Using statiscal analysis we can gain insights on best and least performing airline carriers and most common reasons for delays and cancellations.

Ethical implications:

> Data source for the flat file is a genuine and reliable (Bureau of Transportation). However, the API and website may not hold accurate and reliable information because it is not from a government or FAA authorized source.
>
> **Challenges:**
>
> The flight performance data size is huge (flat file). I would have to find ways to reduce data to a reasonable size without losing meaningful information.

Conclusion:

> For the first project milestone, I have identified data from different sources in different formats. I will be applying different data cleansing and visualization techniques on this dataset to gain meaningful insights in the upcoming project milestones.