**Independent Project – Overview**

**Flights Delay and Cancellation Analysis**

**Dataset Context and Project Aim**

The dataset provides information about the on-time performance (cancellations and delays) of domestic flights operated by large air carriers in the US. The main aim of this Shiny app would be to provide government agencies and airline companies with information and insights to find their areas of weakness and help them identify what factors are causing these problems to the general public.

**App Structure:**

The app would be divided into number of pages, as per this flow –

1. **Landing page** – This page would be the face of the app and would first, briefly describe the problem and the what insights the user could expect from the app. It would also have a Map of US, showing all airports and basic information about them like number of daily flights, average delay, etc. Then I plan to put some graphs/charts showing the major carriers in US, mean delay and cancellation rates for different cities.

Answers to questions from this page -

* 1. What is this app about?
  2. Which geographical region does this app analyzes (cities)?
  3. Which airline companies analyzed and how do they fare overall in terms of delays / cancellations (w.r.t others)?

1. **Main analysis/ Interactive plots page –** In this page, I would be focusing on the main aim of this analysis, which is to assess the factors affecting the delays and cancellations. I am planning to have some grid-heat plot kind of a plot between airport and airline companies.

Answers to questions from this page -

* 1. What is the relationship between an airport/city and a particular airline company operating form there?
  2. Why does an airline perform bad at a particular airport? Does it have a history of delayed departures/ arrivals?
  3. Which airport is always congested and does this result in delay for all airlines flying through that port ? Does it give preference to a particular airline?

1. **Interesting Findings page** – Here I plan to do analysis which could provide critical insights using other features in the dataset like Day of week, Taxi time, etc. More to be included, after further analyzing the data

Answers to questions from this page -

* 1. Which day of the week a particular airport is most busy? What steps could be taken for that (policy analysis)?
  2. How do taxi IN and taxi OUT times affect arrival/ departure delays?

1. **Machine Learning page -**  I plan to include some type of machine learning in my analysis to find patterns in the data not clearly visible through data analysis and visualization. Currently, I am thinking of training a Random Forrest Ensemble model to identify the most crucial matters contributing to delays and cancellation of flights across the nation. I may modify/ change this model as I get to know more about the data.