Project Goals

Team Members:
- Panav Munshi: **pmunshi2**
- Rohan Batra: **rohanb4**
- Anuraag Agarwal: **anuraag3**
- Abhay Narayanan: **abhayn2**

Dataset Link:
- http://stat-computing.org/dataexpo/2009/the-data.html
- https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/HG7NV7

Note: The aforementioned links point to the same dataset.

Description:
We will be using the above dataset and extracting flight data from them in the form of CSV files. These files contain data on the airline, the flight code, its routes, its departure and arrival timings, the elapsed time, the origin, and the destination. We will be applying this data to solve a problem based on the travelling salesman problem. The salesman has a given set of areas it could make sales at, and a given time-period within which he needs to make the maximum number of sales possible. The data provided will help us code a program that will optimize the traveling salesman's route by deciding which areas (and therefore, airports) he needs to visit in his journey, taking into account the buffer-time in between flights and all other factors, overall resulting in having visited the maximum number of areas.

The scope of this project is quite large, and as a group, we still need to finalize some of the specifics about the problem statement and the kind of solution our program will provide; however, we do have a general idea.

Architecture:
- As the dataset has many fields, which includes the name of the airports, the arrival and departure times, etc., we believe that arranging the data in a **graph structure** will enable us to look for shortest paths between the connected components.
- As the data will be constructed into a graph, we plan on using algorithms, such as **Dijkstra's Algorithm** to find the shortest path between two nodes in the graph.
- We are also looking into providing a **graphical output of the entire graph** and highlight the edges we traverse finally while travelling from Point A to Point B.