"Business Data Visualization Project Proposal – Group 3"

1. ***How you will collect and concatenate the monthly BTS datasets?***

a) Data will be retrieved from the Bureau of Transportation Statistics for different months starting from January 2018 to December 2018.
b) After retrieving the data, we will use R to concatenate all the individual datasets and filer for the unique airline code "OO" which represents SkyWest Airlines.
c) We will then delete the data for all other airline codes.

Below are some of the important attributes that we will include in our data:

- Time Period: Quarter, Month, DayOfWeek, FlightDate
- Origin: OriginAirportID, OriginAirportSeqID, OriginCityMarketID, Origin
- Departure Performance: DepDelay, DepartureDelayGroups
- Arrival Performance: ArrDelay, ArrDelayMinutes, ArrivalDelayGroups
- Cancellations and Diversions: Cancelled, CancellationCode, Diversion

2. ***How you will condition the BTS datasets for use in the group project?***

a) R will be used to condition the BTS Datasets.
b) After concatenating the datasets, we will then sub-divide our data into two files
    - Cancelled flights for Year 2018
    - Non – cancelled flights for Year 2018
c) We will then remove some of the highly correlated fields/variables to remove the redundancy, remove all the null values and clean the data for both the data sets.
d) Finally, we will load these two data sets in Tableau.

3. ***How you will explore your airline's data for meaningful patterns and trends?***

Below are some of the trends and patterns that we are planning to analyze in our project:

a) Cancellation analysis and cause - Number of flights which were cancelled throughout the year based on day of week, location source and destination and the reasons for cancellation on monthly basis
b) Analysis of Delays and cause – Monthly analysis filtered by location, months, day of week, origin of flight on the delayed flights and the flights which were on time or early, thereby analyzing the delay cause.
c) Diversion analysis and cause – Analyzing the frequency of diverted flights and its cause
d) Performance analysis – Analyze if SkyWest airline perform better at the start or end of the month
e) Peak hour performance - Average frequency of flights at different times of the day, to identify airline's performance in peak traveling hours
f) Quarterly performance – Comparing various quarters to stud various factors that influence performance
g) Metrics to be used in visualization: Avg Departure Performance over time, Avg Arrival Performance, Available Flying Time, Departure Delays by Origin, Departure Delays by Destination, Cancellation by Origin, Cancellation by Destination, %Cancellation Trend by Quarter.

Group Members : Marc Rodriguez, Nisarg Thakkar, Prajakta Iyer, Ritu Gangwal, Shazia Mariam