List of Steps Taken

1. Downloaded and loaded up data in R, viewed shape data and examined dimension of data and averages within columns.
2. Ran a few ggplots by airline\_sentiment and negative reason, facet\_wrapped by airline.
3. Cleaned existing text and produced new column fix\_test. Process removed special characters and emojis within each text cell. Used the link to obtain the code: <https://towardsdatascience.com/text-mining-with-r-gathering-and-cleaning-data-8f8b0d65e67c>
4. Moved to Excel after struggles to clean data with R. Used Excel to clean data by removing unnecessary columns
5. Created new columns in Excel: user\_region, Grade\_sentiment\_confidence, Grade\_negative\_reason. Used Vlookup to create new groupings for user\_region and used if statements to segment confidence by grade. (Perfect, Strong, Good, Ok, Weak)
6. Created new Pivot tables to analyze data relevant to airlines, user region, and confidence for sentiment and negative reasons at deeper level.
7. Created charts off Pivot tables to visualize distribution of confidences, tweets by region, proportion of tweets by negative reason, and more.
8. Constructed methodology to compare airlines through a competitive strength assessment. Researched online to see how strength assessments were formatted in consulting business.
9. Underwent calculations and set guidelines on how to score and grade each airline by the following factors: negative reason performance, user\_region performance, and airline sentiment performance. Used information available in Pivot table to base calculations off of.
10. Scored each airline based on these three factors, then listed takeaways by mentioning plans of actions for airlines, consulting companies, and possible limitations of assessment.