[**Final Project Submission**](https://blackboard.syracuse.edu/webapps/assignment/uploadAssignment?content_id=_5322499_1&course_id=_393757_1&group_id=&mode=view)

**Deliverables**:

One 10-15 slide presentation that summarizes your analysis. The audience for this will be Executives level leadership of an Airline company. Please assume these executives do not know too much about statistics, so you probably should not quote terms like “R-squared” or “p-value” but rather describe your statistical results in plain language.

One MS-Word file, containing a detailed report all your work. This report should include sections for all the phases of data science discussed in this course. A suggested template of the deliverables will be provided. Your intended audience for this report will be your Data Science professors, who understand R code and Data Science. Please make sure to include all assumptions made and any analysis completed, whether you found it significant or not.

**Rules of Engagement**: This is an honor system assignment: You **may**consult with IST687 professors and faculty assistants, the textbook, and publications on the Internet at any time. You **may not**consult, collaborate, or seek assistance from any other human besides this. Your attribution statement, at the top of your R-code file, must reflect these constraints. You may not share your results or work in progress with any other human besides your professors and faculty assistants. Note that your data file is unique to you: The results that other students in IST687 obtain will be different from yours. Project updates from you will be due at the dates provided to ensure you are on track and there are not any outstanding questions.

**Project Goal**: The goal of this term project is for you to use all of the skills you have developed in the IST687 class/labs/homeworks to make sense of a novel dataset; to perform some essential analyses on the dataset; and to explain/document what you have done. The dataset contains survey data of air travel within the U.S, one row per customer, per trip.

**Accessing Your Data File**: The data will be available to you in Blackboard. The file contains about 32 columns/variables.

**Recommended Project Phases**

**Data Pre-processing / Data Preparation Phase**

* *Phase 1: Mitigate Missing Data*. There are several columns in the dataset that may contain missing data. Write code that examines each column to see if it contains missing (or NA) data. To mitigate missing data, use any of the methods learned in class and make sure to note your approach in your final documentation. Use comments in your code to document how many missing data values you had to repair and your approach to do that.
* *Phase 2: Summarize variables.* For each numeric variable, create a histogram, add a comment that describes the shape of the histogram as symmetric, positively skewed (long right tail), or negatively skewed (long left tail). For each factor variable (e.g., Gender), use the table() command to summarize how many observations are in each category.

**Exploratory Analysis Phase**

* *Phase 3: Predictive Modeling.*Many columns contain data relating to the characteristics of each customer’s trip. Using the modeling techniques, we learned in the class (Liner Modeling, Assoc Rules, SVM), develop 3-5 different predictive models that analyze the data.
* *Phase 4: Map Low Satisfaction Routes.*Subset your data to create a smaller data set containing only the trips where customers reported the lowest levels of satisfaction. The latitude and longitude of each origin and destination is shown in the data set. Use ggplot to place route curves onto an outline map of the U.S. states. The geom\_curve() geometry supports this kind of plotting.

**Business Recommendations Development Phase**

* *Phase 5: Make Sense of Low Satisfaction Segments.* The client wants to know why customers become dissatisfied with their air travel. Use insights from Phase 3 and Phase 4 to explain why certain trips have low satisfaction. Conduct any appropriate follow-up analyses to provide evidence for your ideas. Make sure to document any additional code with appropriate comments.
* *Phase 6: Develop Marketing Plan*. Identify three interesting Market Segments. Define the demographic characteristics associated with each Market Segment. Finally recommend three actionable insights for each segment that you believe would increase the NPS for the segment.

Your presentation should provide the client (presumably the Executive/leaders of the airline) with an explanation of your results in language that is suitable for an Executive to understand. Your report should contain the data and visualizations that supports your actionable insights you are trying to communicate to the client.