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

**Deliverables**:

**Deliverable #1**

One slide presentation (10-15 slides) 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 show R code or quote terms like “R-squared” or “p-value”, but rather describe your statistical results in plain language. Your group will present this slide deck on the last day of classes at the normal class time.

**Deliverable #2**

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.

**Deliverable #3**

All your R Code

**Rules of Engagement**: This is an honor system assignment: You **may**consult with your other group members, IST687 professors and faculty assistants, the textbook, and publications on the Internet at any time. Your attribution statement, at the top of your R-code file, must reflect these constraints. As a group, you may not share your results or work in progress with any other human besides your professors and faculty assistants.

**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: Clean your Data*. There are several columns in the dataset that may contain missing data, need cleaning, or transformation to a different data type, etc. 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. Write code to clean your data, as well as any needed code to transform your data.
* *Phase 2: Getting to know your Data and prepare business questions -* For all of your significant numeric variables, create visualizations and comments that describes what value your gained from these. Describe in detail how this drove your business question development(include a list of those business questions(~10), as well as the data segments you are using in each questions.

**Exploratory Analysis Phase**

* *Phase 3: Predictive Modeling.*Many columns contain data relating to the characteristics of each customer’s trip. To get further insights into your business questions, use the modeling techniques we learned in the class (Linear Modeling, Assoc Rules, SVM) or others you are familiar with in R, to develop 3-5 different predictive models that analyze the data. This is 3-5 models per modeling technique. This means you need to change parameters and/or model inputs to get the best model.

**Business Recommendations Development Phase**

* ***Phase 4****: Make sense of your models and Develop Marketing Plan*. Reduce your business questions to your top 5 (e.g. Low Likelihood to recommend for middle age, female, business traveler, in the northeast). Finally recommend 5 actionable insights for each segment that are **actionable** and you believe would increase the NPS for the segment.