MGT 6203 Group Project Proposal Template

**Please edit the following template to record your responses and provide details on your project plan.**

**TEAM INFORMATION (1 point)**

**Team #:**

**Team Members:**

1. Team Member 1 Name; GT Id (OMSA) or EdX username (MM)

[Insert background information: Name, professional background, education background, previous analytics related projects you have worked on]

1. Ryan Chandler, EdX username Zujin87
2. Team Member 3 Name; GT Id or EdX username
3. Team Member 4 Name; GT Id or EdX username
4. Team Member 5 Name; GT Id or EdX username

**OBJECTIVE/PROBLEM (5 points)**

**Project Title:** Finding the most valuable predictors for airline customer satisfaction

**Background Information on chosen project topic:**

Customer satisfaction, loyalty, and returns are of the utmost importance to the airline industry. Maximizing customer satisfaction for the lowest cost is one way to help maximize profits. The chosen data set represents one airline company and its customer’s satisfaction results along a series of predictors.

**Problem Statement (clear and concise statement explaining purpose of your analysis and investigation):**

What actions can the airline take to raise the likelihood of customer satisfaction at the most cost-efficient manner?

**State your Primary Research Question (RQ):**

What are the most important and cost-efficient predictors for making customers satisfied with their care?

**Add some possible Supporting Research Questions (2-4 RQs that support problem statement):**

1. Are there any interactions among predictors that affect the probability of a satisfied customer?

**Business Justification:** **(Why is this problem interesting to solve from a business viewpoint? Try to quantify the financial, marketing or operational aspects and implications of this problem, as if you were running a company, non-profit organization, city or government that is encountering this problem.)**  
Based on analysis results, the airline will be able to focus on actions and items that lead to the highest customer satisfaction results. The airline could be able to predict the results of budget cuts to predictors and the resulting affects on customer satisfaction. This result could be used to optimize expenditures versus customer satisfaction in order to maximize overall net profit.

Operationally, the airline has limited resources, both human and material. The model could be able to assist with prioritizing resources to maximize customer satisfaction. Depending on the type of resources being used, the company could potentially reduce “wasteful” or carbon emitting predictors that have a low impact on satisfaction scores and market those reductions as green initiatives.

**DATASET/PLAN FOR DATA (4 points)**

**Data Sources (links, attachments, etc.):**

Primary data set: <https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction>

Secondary data set: <https://www.kaggle.com/datasets/open-flights/flight-route-database>

**Data Description (describe each of your data sources, include screenshots of a few rows of data):**

The primary data set contains anonymized customer data describing their satisfaction level with various interactions and services with the airline. The data set also includes information about the customer, distance traveled, and whether it was a personal or business trip.

The secondary data set is a list of airlines, routes, and type of airplane used for that route.

**Key Variables: (which ones will be considered independent and dependent? Are you going to create new variables?** **What variables do you hypothesize beforehand to be most important?)**

The dependent variable will be “satisfaction” with all other predictors being independent. There are several variables that may lend themselves well to being factored out such as “Customer Type”, “Type of Travel”, and “Class”. One hypothesis is that factors for shorter flights may be more or less important than other factors for longer flights. An example is that inflight WIFI service and Leg Room may be more important for customer satisfaction on longer flights while these factors may not be as strongly correlated for shorter flights.

**APPROACH/METHODOLOGY (8 points)**

**Planned Approach (In paragraph(s), describe the approach you will take and what are the models you will try to use? Mention any data transformations that would need to happen. How do you plan to compare your models? How do you plan to train and optimize your model hyper-parameters?))**

While this is overall a classification problem, there are a few different ways to model the data. A support vector machine is an obvious choice for classification problems but may lack some flexibility required to properly model the data. Using a logistic model could potentially work as well but may take some tuning to find the correct threshold value. A forest model combined with logistic regression could be used in the case that a higher flexibility is required.

At this time, there is no known need for data transformation other than the creation of dummy variables.

The primary plan to compare the modes is to use a confusion matrix to compare each model against the test set. One limiting factor of this data set is that the neutral and dissatisfied customers are grouped together. This makes it difficult to understand which factors most correlate to an unsatisfied customer. Based on the belief that a dissatisfied customer will cause more harm to the brand overall than a satisfied customer can help the brand, the confusion matrix analysis will focus on minimizing “neutral or dissatisfied” errors.

The data are broken into a training and test data sets. The data will be modeled using different modeling techniques as to find the optimal technique. Once a technique is chosen, fine tuning of hyper-parameters will be based on the classification and my brain has stopped working.

**Anticipated Conclusions/Hypothesis (what results do you expect, how will you approach lead you to determining the final conclusion of your analysis) Note: At the end of the project, you do not have to be correct or have acceptable accuracy, the purpose is to walk us through an analysis that gives the reader insight into the conclusion regarding your objective/problem statement**

**What business decisions will be impacted by the results of your analysis? What could be some benefits?**

**PROJECT TIMELINE/PLANNING (2 points)**

**Project Timeline/Mention key dates you hope to achieve certain milestones by:**

**Appendix (any preliminary figures or charts that you would like to include):**