MGT 6203 Group Project Proposal Template

**Team 6:**

**Team Members:**

1. Name: Raajitha Middi; EdX username – Raajitha Middi

Holds engineering and MBA degrees. Worked as Data Analyst in various domains including IT, e-commerce, and healthcare. In my professional capacity, I worked on multiple projects like creating interactive dashboards on PowerBI, procurement data analysis etc.,

1. Team Member 3 Name; GT Id or EdX username
2. Team Member 4 Name; GT Id or EdX username
3. Team Member 5 Name; GT Id or EdX username

**OBJECTIVE/PROBLEM (5 points)**

**Project Title:** Airline Passenger Satisfaction Prediction

**Background Information on chosen project topic:**

Customer service plays an important role in any industry, particularly in airlines as it is a capital-intensive industry and hence customer satisfaction is of utmost importance. Ensuring customer satisfaction retains the customers which leads to more revenue. There might be different factors/customer touchpoints which influence customer satisfaction, and it is important for airlines to understand which factors play a key role in customer satisfaction.

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

Our objective is to identify the crucial factors and predict whether a customer is satisfied with the airline or not.

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

What are the important attributes in driving customer satisfaction in the airline industry?

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

1. Are these attributes influenced by each other – for example, departure time might be influenced by the arrival time? Some airlines might cover the delay time in air and arrive by the scheduled time but that might not be the case always.

**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.)**

Customer satisfaction is a key metric as it has a direct effect on the bottom line. The analysis of data helps us understand better the customer preferences and expectations. Airlines can further use the information in spend analysis for optimizing the operations, reallocating the costs to enhance the customer experience. This in turn not only increases the revenue by retaining the customers but also provides a competitive advantage.

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

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

Our dataset is found at: <https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction>

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

The dataset contains the information from an airline passenger satisfaction survey. It has a total of 24 variables.

Most of the data that has numbers, for example cleanliness, check in service etc., are not actually numeric as they are ratings and are to be considered as categorical variables.

The variables that are numeric are id, age, flight distance, departure delay and arrival delay.

A picture containing text, screenshot, font, number

Description automatically generated

A picture containing text, screenshot, line, font

Description automatically generated

**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?)**

We are considering the “satisfaction” variable as the dependent variable and the remaining variables except “id” to be dependent variables (id is irrelevant). The dependent variable has 2 classes – satisfied and neutral/dissatisfied.

Though logistic regression creates indicator variables in the model for us, we might want to add the significant ones to the data to predict for new data. In addition to this, we might classify the age into different groups, for ex: 20-35, 36-50 and so on. This makes us understand better if the satisfaction is varied across different age groups.

We think some of the important predictors to be:

* Check-in service
* Onboard service
* Seat comfort
* Cleanliness
* Departure delay
* Arrival delay

**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?))**

The first step is to perform explanatory data analysis (EDA) to identify the ineffective variables, outliers, missing values and understand the relationships between different variables.

Since the response variable is categorical, our main model would be logistic regression after including different indicators and interaction variables. We get the model output in terms of probability for which we should set a threshold value to classify them into satisfied or neutral/dissatisfied classes. We’ll check the accuracy using the confusion matrix and ROC curve.

In addition to logistic regression model, we would also like to build the model using classification trees, random forest, K-nearest neighbors (KNN) and perform cross-validation.

**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.**

We’re guessing that predictors like in-flight service, seat comfort, departure/arrival delays might play a major role in customer satisfaction.

* We can explore what kind of in-flight service will make an impact, costs associated with it and accordingly decisions can be made.
* Seat comfort can be difficult to achieve and involves a lot of money too.
* Sometimes, departure and arrival times are not in our control. They depend on various factors like the airport location, runway availability, air traffic and so on.

So further in-depth analysis can be performed by collecting data from various sources but that might be out of scope for now.

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

Based on the analysis, we identify which factors are the most important and of them, we can group take away the ones that are difficult to achieve (like seat comfort). And we can work on the factors that are achievable like in-flight service, online booking etc., We should make sure that the costs of the improvements should not outweigh the benefits/returns.

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

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

Proposal submission – 21st June

Data collection and EDA – 30th June

Fitting different models to the data – 10th July

Final checks and refinements – 15th July

Final report preparation and submission – 20th July

Final presentation – 23rd July

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