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 #:** 6

**Team Members:**

1. Henri SALOMON; henri\_salomon

Henri Salomon is a data analyst at the United Nations in New York where he specializes in data-driven insights about internal operations. With a background in management and consulting, Henri has worked with the World Food Programme (WFP) and PwC's Strategy&, demonstrating interest and expertise in innovation, efficiency, and strategic analysis.

1. Team Member 2 Name; GT Id or EdX username
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:** Analyzing factors influencing airline passenger satisfaction

**Background Information on chosen project topic:**

While demand for air travel has been growing back and beyond pre-pandemic levels, the airline industry has been facing structural trends and needs more than ever to generate value for both passengers and investors. With the growth of low-cost actors, ancillaries have become the norm for the industry as passengers enjoy greater control of their travel experience. Price and schedule are no more the only factors for customers, and reputations along with service quality help companies differentiate themselves. Loyalty schemes, online sales, and mobile apps have for example become trusted channels to engage directly with customers. In this context, identifying which factors best influence airline passenger satisfaction can help empower airline companies, and adjust their offer as the traditional travel classes/ travel types no longer represent the only customer segmentation.

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

The purpose of this analysis is to investigate factors influencing passenger satisfaction while being able to predict passenger satisfaction.

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

Which factors (most) influence airline passenger satisfaction?

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

1. Can we predict airline passenger satisfaction based on a set of factors?
2. Beyond traditional characteristics for segmenting customers (e.g., travel class, type of travel), can we further define customer segments?
3. Do airline passenger satisfaction levels accurately reflect the true quality of airline services (e.g., leg room for Business class passengers?

**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.)**

The airline industry has evolved around different trends, marked by the hybridization of business models – the traditional distinction between full-service and low-cost carriers has been blurring – and by the frequency of financial distress due to internal (mismanagement) or exogenous (e.g., oil prices). In this competitive context, improving air passenger satisfaction can have a positive impact on both companies’ revenues and costs. First, air companies can focus on operational efficiency while addressing customers’ real and personalized needs to increase satisfaction levels and reduce efforts on unnecessary needs (“value for money”). Second, improving customer satisfaction will reduce customer churn and improve customer loyalty, which has become a barrier to entry with higher customer acquisition costs, consequently increasing their revenues.

Finally, on the marketing side, improving customer experience will contribute to building a strong brand image and reputation. For example, passengers would be more inclined to share their experience (e.g., word of mouth, social media), which could attract new passengers.

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

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

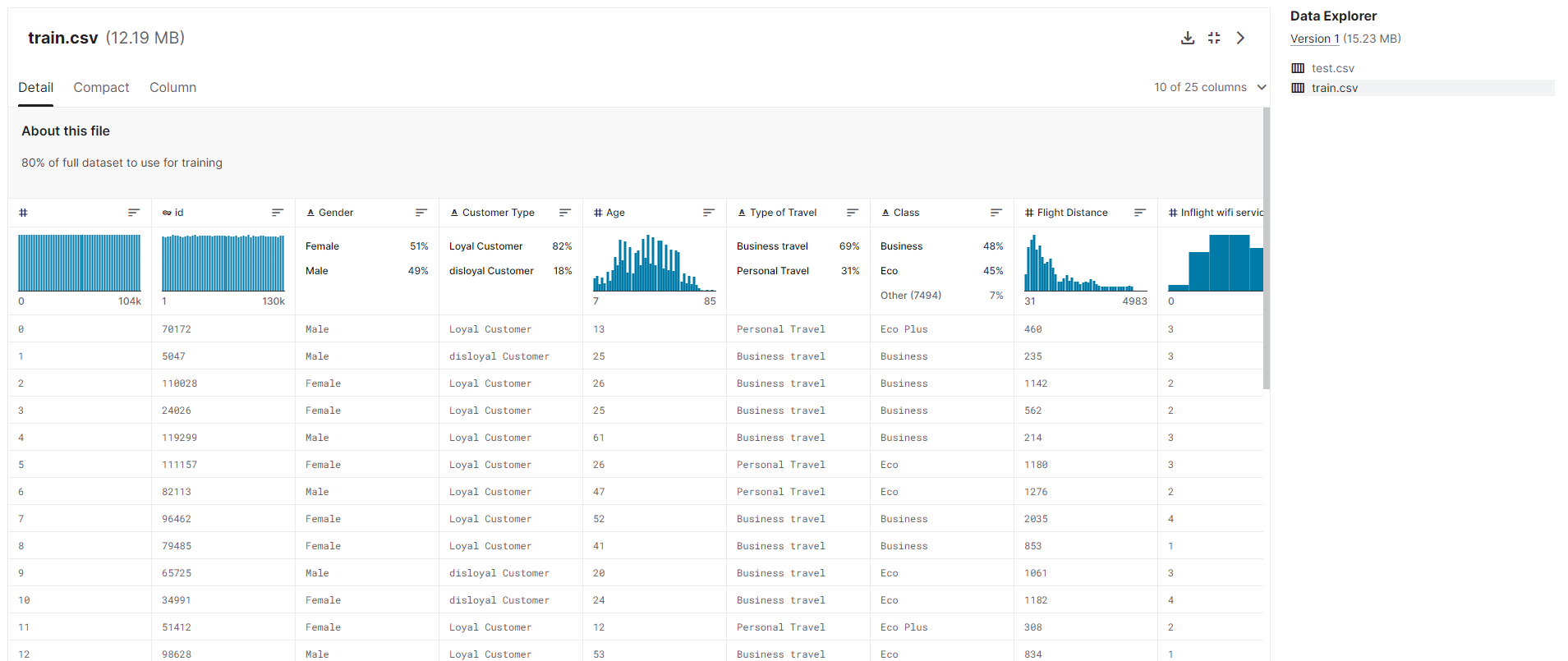
The primary data source identified is from a US Airline passenger satisfaction survey from 2015 and is available on Kaggle[[1]](#footnote-1).

An alternate data source is from the American Customer Satisfaction Index (ACSI) and is available on Mendeley[[2]](#footnote-2). It contains customer satisfaction across 50 industries, including airlines in 2015.

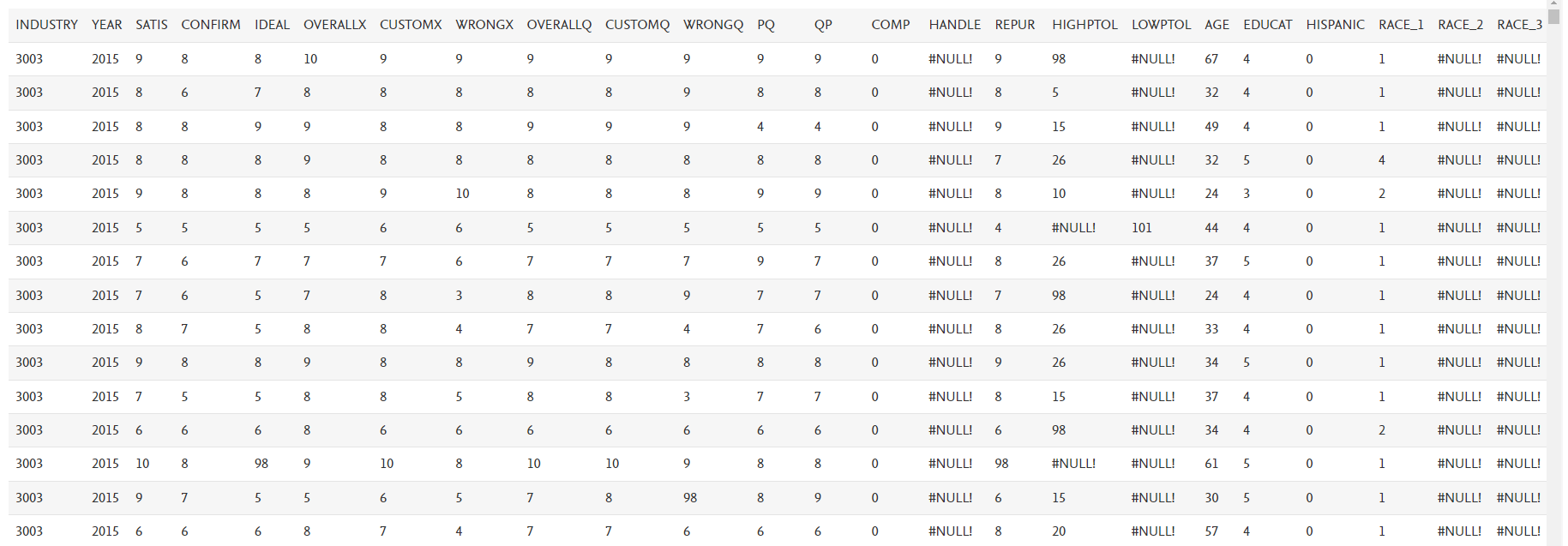
A secondary data source is from Opengflights.org and is available on Kaggle[[3]](#footnote-3). It contains information about flight routes between airports on different airlines worldwide.

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

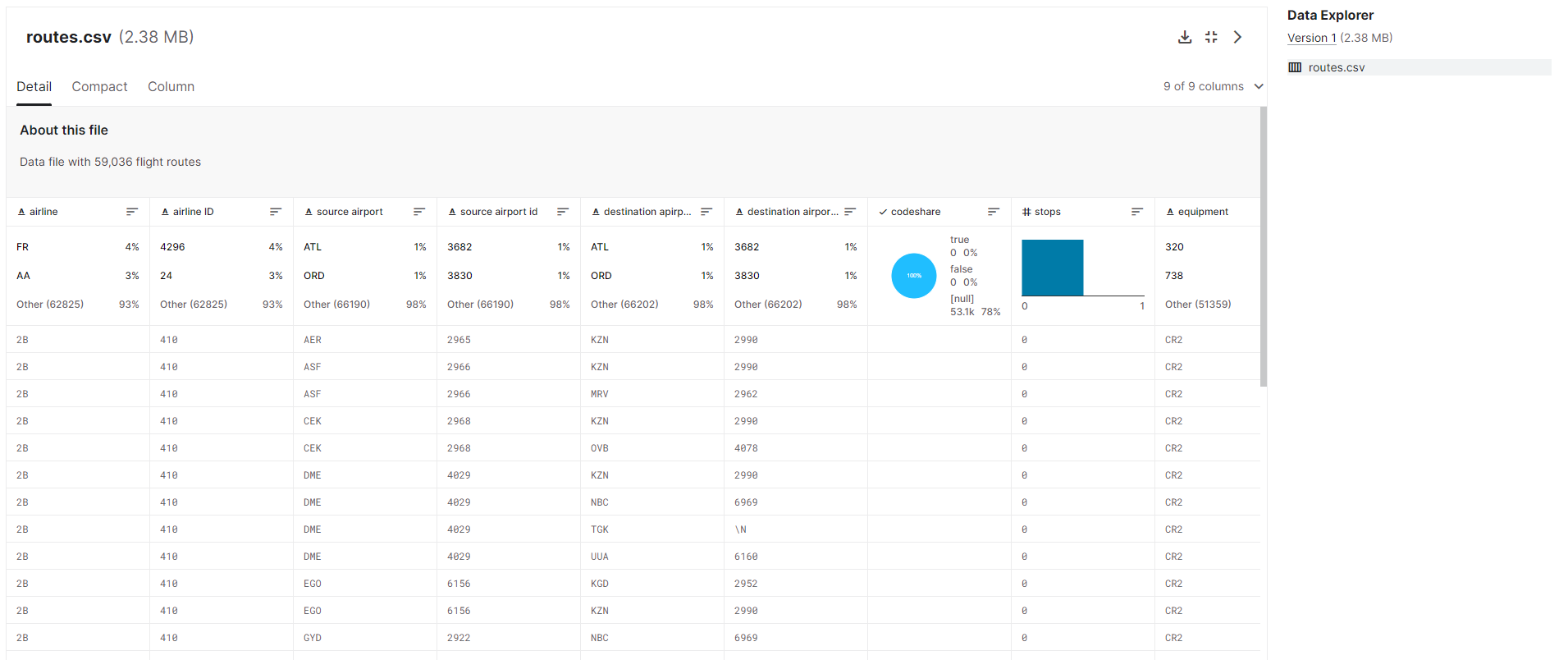
**The primary data source** presents results from a satisfaction survey (satisfied, neutral or dissatisfied) conducted by a US airline passenger and includes passenger demographics information (e.g., Gender, Age, type of travel), flight information (e.g., Type of travel, Flight delay) and customer experience surveyed at a granular level (e.g., Inflight wifi service, Food and drink, Seat comfort). The data set contains 25 columns and 130k rows.



**Alternate data source:** The alternate data set provides a sample of survey data collected by the American Customer Satisfaction Index (ACSI) about customer satisfaction in 50 consumer industries, including airline companies in 2015.



**Secondary data source:** The flight route data set contains data about 59036 routes between 3209 airports on 531 airlines around the globe.



**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 in this analysis would be "Satisfaction," representing the overall satisfaction level of passengers. The independent variables would include all the other features in the dataset, such as gender, age, customer type, type of travel, class, flight distance, and satisfaction levels of various services.

Based on our personal experience, we hypothesized beforehand that the following would be most important: inflight service, delay, seat comfort, inflight entertainment, food and drinks, cleanliness and baggage handling.

**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?))**

* **Data collection, cleaning and transformation:** Afterloading the data sets, missing values, outliers and data quality issues will be identified. Categorical variables would be encoded, and any necessary data transformations or scaling would be applied.
* **Exploratory Data Analysis:** Preliminary and descriptive analyses will be conducted on the data sets, providing already useful insights. For example, we will look at correlation matrix and distributions but also at average satisfaction across different groups (e.g., class type, flight type, loyalty program, gender, age,…).
* **Modelling:** Different models will be tested and assessed against performance. To compare models, the data set will be split into training, validation and testing with a 60:20:20 ratio, and the best model chosen best on performance on test data. The analysis of coefficients will help us reply to the research question 1. These models would include:
* Regression: Linear, Log-linear, Log-log
* Random forest
* Feature selection including Lasso
* PCA

**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**

Based on the analysis, we expect to identify key factors that significantly influence passenger satisfaction. For example, we expect to find that certain services such as seat comfort, and inflight entertainments, have a strong correlation with overall satisfaction.

Focusing on certain customer segments based on demographics, travel class, flight type, etc., we expect to find different key factors impacting passenger satisfaction. For example, we hypothesize that seat comfort will have a greater impact on passenger satisfaction for long-distance flights than for short-distance flights. In addition, we expect to identify clusters of customers based on characteristics and passenger satisfaction.

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

Factors with a high impact on passenger satisfaction will have to be prioritized as an improvement of the factor could yield a significant increase in customer satisfaction. Overall, it will help identify actionable recommendations and prioritize activities for operational efficiency improvements at the global level and ‘local’ levels (per class type, per travel type, …). For example, companies might want to invest or dedicate more resources to some activities, while reducing activities for others. As data comes from a survey, perception might be misled by other realities, and the company might need to better engage and communicate with passengers to change their perception rather than services.

Finally, identifying clusters of customers will help rethink the marketing strategy and better understand customer expectations and needs in the medium and long term.

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

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

1. **Team formation, Data & Topic Selection**

* Activity 1: Identity team, topic and data sets
* Milestone 1: Selected and framed topic

1. **Project Proposal (by 21 June 2023) and Proposal Video presentation (by 2 July 2023)**

* Activity 1: Prepare project proposal including problem statement and research questions
* Activity 2: Meet with TA and reflect feedback by updating the project proposal
* Activity 3: Load data set and conduct first Exploratory Data Analysis (EDA)
* Activity 4: Prepare Proposal Video
* Milestone 1: Proposal video prepared
* Milestone 2: Cleaned and ready-to-use data sets

1. **Progress report (by 9 July 2023)**

* Activity 1: Identify academic research or past results
* Activity 2: Identify and perform models on data sets
* Activity 3: Brainstorm and draw initial business implications from models
* Milestone 1: Summary of models/ variables used and one/several model(s) identified with good prediction rate/ reduced error

1. **Final Project Submission (by 20 July 2023, and for Final video by 23 July 2023)**

* Activity 1: Finalize report and project submission
* Activity 2: Realize the final video for reports
* Milestone 1: Final results and reports validated

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

1. <https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction> [↑](#footnote-ref-1)
2. <https://data.mendeley.com/datasets/64xkbj2ry5/1/files/e4e4b8b0-0d7d-41a9-a2be-c1e586897d7e> [↑](#footnote-ref-2)
3. <https://www.kaggle.com/datasets/open-flights/flight-route-database> [↑](#footnote-ref-3)