Group1_ButCharLo_MiniProject1_EDA_Report

Customer Satisfaction with United Airlines (UA)

by But But, Charles Wong, Hazel Lo

06/12/2024

Executive Summary

To improve customer satisfaction at United Airlines (UA), this report focuses on analyzing 10,356 valid survey records from Economy Class passengers. The analysis highlights three critical areas for improvement: Inflight Wi-Fi service, Ease of Online Booking, and Online Boarding. For business travelers in Economy Class, additional attention is needed on Food and Drink, Seat Comfort, Inflight Entertainment, and Cleanliness. Further analysis is recommended to prioritize these areas to maximize their impact on customer satisfaction.

Project Details and Objectives

This exploratory data analysis (EDA) aims to identify key factors correlated with passenger dissatisfaction based on UA's customer satisfaction survey. Insights derived will guide UA in targeted service improvements to enhance overall customer experience.

Dataset Description

Introduction to the dataset

The dataset was sourced from Kaggle (*https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction*) and includes survey responses from 25,976 passengers across 25 features.

Dataset Characteristics

Key attributes include:

Index and ID: Removed for redundancy during preprocessing.
Demographic details: Gender, Age, and Customer Type.
Travel details: Class, Type of Travel, and Flight Distance.
Service ratings: 14 features (e.g., Inflight Wi-Fi, Seat Comfort) scored on a 5-point scale.
Outcome: Satisfaction level categorized as "Satisfied" or "Neutral/Dissatisfied."

Detail attributes

#	Column name	Data Type	Remark/ Description
1	Index	int64	Index number start from zero
2	id	int64	Customer id number
3	Gender	string	Male/Female
4	Customer Type	string	Loyal/Disloyal Customer
5	Age	int64	Customer age
6	Type of Travel	string	Business/Personal travel
7	Class	string	Eco/ Eco Plus/ Business Class
8	Flight Distance	int64	Distance in nautical miles
9	Inflight wifi service	int64	5-point scale from 1 - 5 (1=lowest, 5=highest)
10	Departure/Arrival time convenient	int64	5-point scale from 1 - 5 (1=lowest, 5=highest)
11	Ease of Online booking	int64	5-point scale from 1 - 5 (1=lowest, 5=highest)
12	Gate location	int64	5-point scale from 1 - 5 (1=lowest, 5=highest)
13	Food and drink	int64	5-point scale from 1 - 5 (1=lowest, 5=highest)
14	Online boarding	int64	5-point scale from 1 - 5 (1=lowest, 5=highest)
15	Seat comfort	int64	5-point scale from 1 - 5 (1=lowest, 5=highest)
16	Inflight entertainment	int64	5-point scale from 1 - 5 (1=lowest, 5=highest)
17	On-board service	int64	5-point scale from 1 - 5 (1=lowest, 5=highest)
18	Leg room service	int64	5-point scale from 1 - 5 (1=lowest, 5=highest)
19	Baggage handling	int64	5-point scale from 1 - 5 (1=lowest, 5=highest)
20	Checkin service	int64	5-point scale from 1 - 5 (1=lowest, 5=highest)
21	Inflight service	int64	5-point scale from 1 - 5 (1=lowest, 5=highest)
22	Cleanliness	int64	5-point scale from 1 - 5 (1=lowest, 5=highest)
23	Departure Delay in Minutes	int64	Departure Delay Time
24	Arrival Delay in Minutes	float64	Arrival Delay Time
25	satisfaction	string	Satisfied/ Neutral or dissatisfied

Methodology

Data preprocessing steps

The following data preprocessing steps were being taken before analysis:

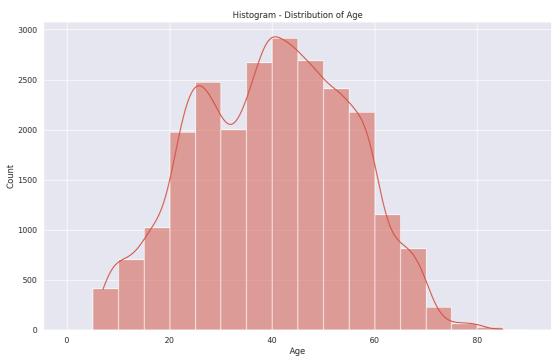
#	Steps included:
1	Dropping redundant columns (Index and ID).
2	Removing duplicate and invalid records (e.g., rows with zero service ratings).
3	Handling missing values by row deletion.
4	Validating data consistency (e.g., ensuring string-based columns contain appropriate values)

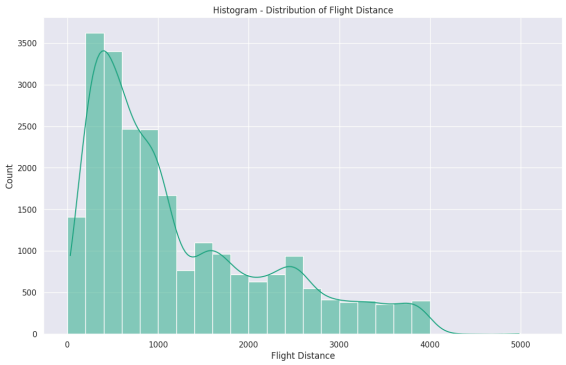
Rationale(s) behind

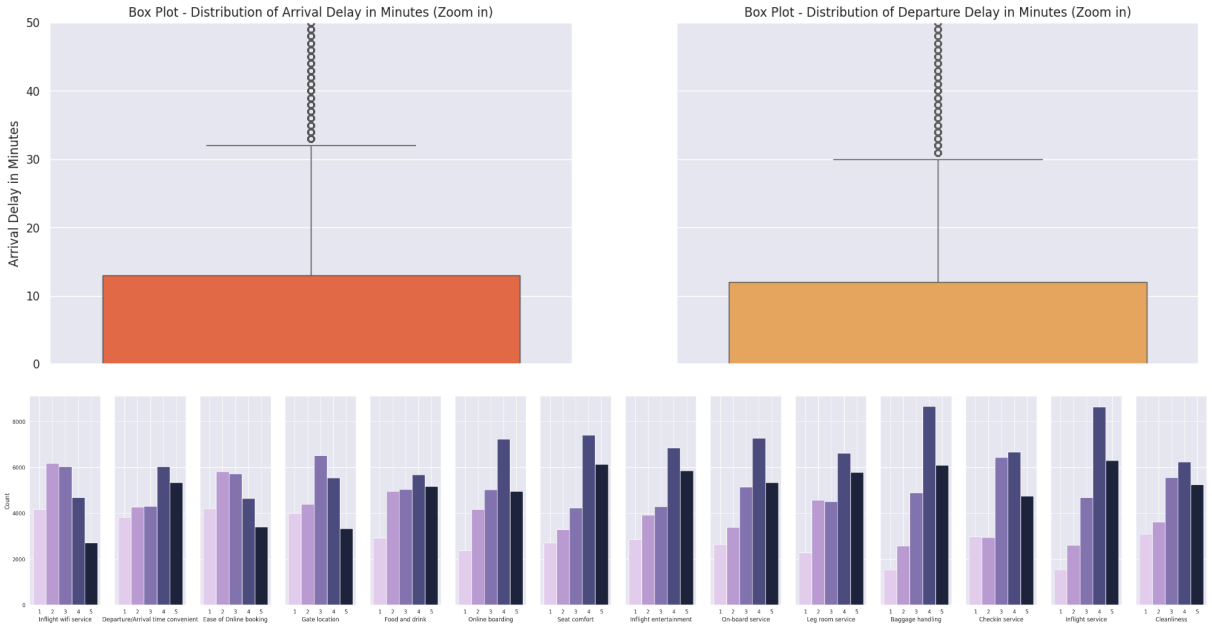
High-quality data is essential for reliable analysis; hence, preprocessing eliminates potential biases and inaccuracies.

Analysis

Univariate Analysis







After preprocessing the dataset, 23,789 valid records were analyzed to explore the general distribution of variables. Key insights include:

- **Demographics:** Gender distribution is nearly balanced between male and female passengers. The majority of customers are categorized as "Loyal," indicating frequent participation in United Airlines' MileagePlus program. Most passengers are in the age group of 40–50 years old, with the distribution resembling a normal curve.
- **Travel Characteristics:** Business travel dominates over personal travel as the primary purpose, and Economy Class is the most commonly chosen class. Flights are typically short-haul, with a peak in flight distances around 400–600 nautical miles.
- **Service Ratings:** Service features such as **Baggage Handling** and **Inflight Service** consistently receive high ratings (4–5), while **Inflight Wi-Fi Service** stands out with predominantly low ratings (1–2), highlighting a significant area for improvement.
- **Satisfaction Levels:** A slightly higher proportion of passengers report being "Neutral or Dissatisfied" compared to those who are "Satisfied." This indicates room for enhancing customer satisfaction through targeted service improvements.

Bivariate Analysis

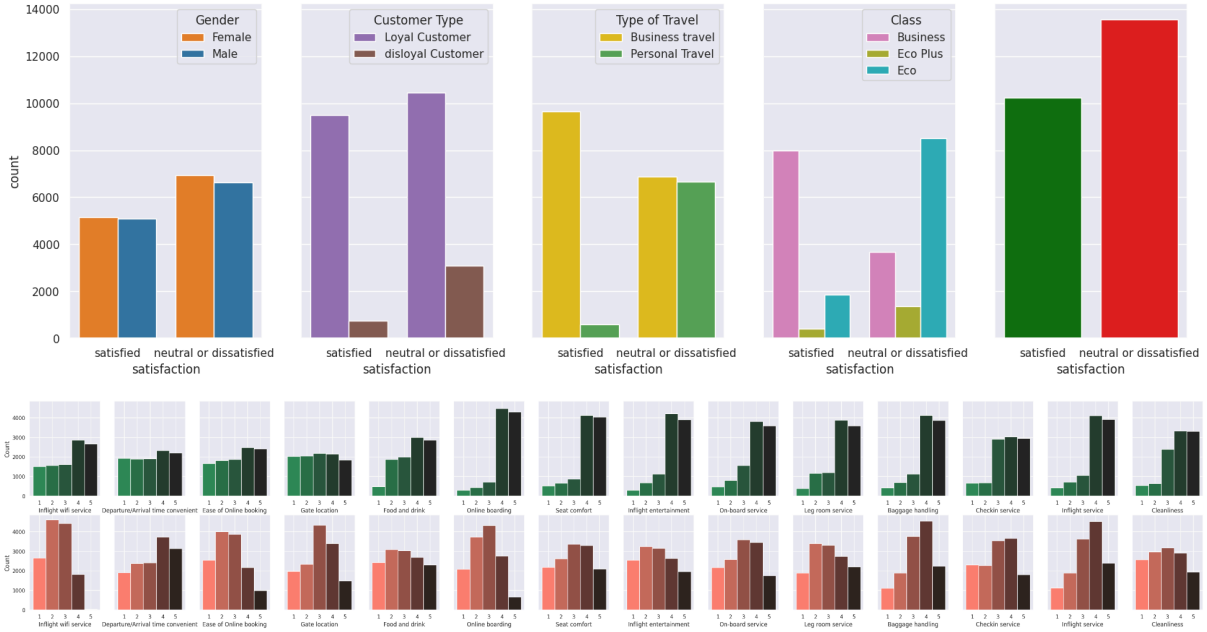
Analysis Logic

To identify key factors contributing to dissatisfaction, we focused on how service ratings correlate with customer satisfaction levels. The analysis logic is as follows:

- If the ratings for a service are concentrated in the **4–5 range**, but the customer feedback is "Neutral or Dissatisfied," this suggests the service is **not a primary determinant** of dissatisfaction.
- Conversely, ratings concentrated in the **1–2 range** are indicative of **critical pain points** that likely contribute to dissatisfaction.

Using this approach, we evaluated the distribution of service ratings against satisfaction levels to uncover meaningful patterns.

Bivariate Analysis: Whole Dataset

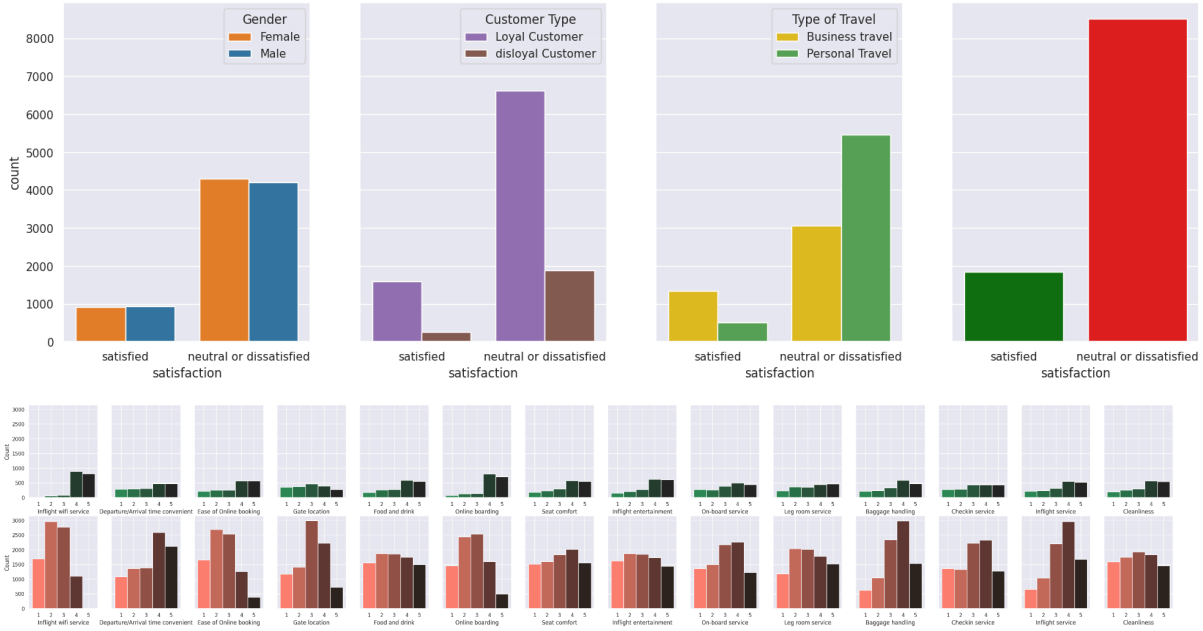


1. **Satisfaction vs. Class:**
    - Economy Class passengers account for the largest share of "Neutral or Dissatisfied" customers, while Business Class passengers show a significantly higher proportion of "Satisfied" ratings.
    - Economy Plus shows mixed satisfaction levels, likely influenced by its intermediate service offerings.
2. **Satisfaction vs. Service Ratings:**
    - Service features with low ratings (1–2) are strongly correlated with dissatisfaction. The key contributors include:
        - **Inflight Wi-Fi Service:** Displays consistently poor ratings, with a high concentration of 1s and 2s and no 5-point ratings, indicating a severe pain point.
        - **Ease of Online Booking** and **Online Boarding:** Both features show a significant proportion of low ratings among dissatisfied customers, highlighting areas for improvement.
3. **Non-Determinative Factors:**
    - Features such as **Gate Location** and **Departure/Arrival Time Convenience** have high ratings (4–5) even among dissatisfied customers, indicating these factors do not heavily influence satisfaction.
4. **Visual Insights:**
    - The graphical representation of satisfaction versus features demonstrates clear differences in rating distributions between "Satisfied" and "Neutral or Dissatisfied" groups, particularly for the top contributors like **Inflight Wi-Fi Service** and **Online Boarding.**

Insights from Satisfaction(Binary) vs Class:

Key observation: Among the customers who voted 'neutral or dissatisfied' for the overall service to UA, Eco class has the most 'neutral or dissatisfied' customers.

Bivariate Analysis: Focus on Economy Class

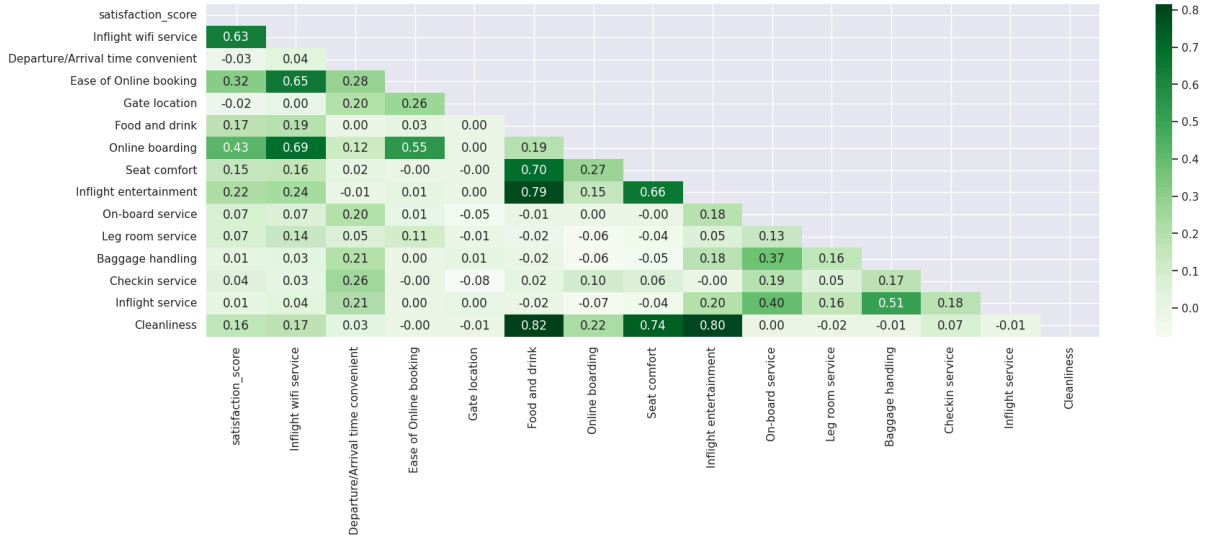


Given the high concentration of dissatisfaction in Economy Class, a focused analysis was conducted to uncover specific drivers within this segment:

1. **Satisfaction vs. Service Ratings in Economy Class:**
    - The trends observed in the whole dataset are magnified within Economy Class:
        - **Inflight Wi-Fi Service** has the highest proportion of 1–2 ratings, reaffirming its critical impact on dissatisfaction.
        - **Ease of Online Booking** and **Online Boarding** also exhibit low ratings, consistent with the broader dataset.

2. **Insights on Additional Features:**
    - Economy Class passengers rate features like **Food and Drink**, **Seat Comfort**, and **Inflight Entertainment** lower compared to passengers in other classes. These features are likely secondary contributors to dissatisfaction but warrant attention for improvement.
3. **Key Observations:**
    - The proportion of "Neutral or Dissatisfied" customers remains significantly higher than "Satisfied" within Economy Class, reflecting a pressing need for targeted interventions in this segment.
    - The focus on low-rated features within Economy Class aligns with overall dissatisfaction trends but also introduces class-specific pain points like **Seat Comfort** and **Food and Drink.**
4. **Actionable Insights for Economy Class:**
    - Investments should prioritize **Inflight Wi-Fi Service**, **Ease of Online Booking**, and **Online Boarding**, as these features are critical to reducing dissatisfaction in this class.
    - Enhancing secondary aspects, such as **Food and Drink** and **Seat Comfort**, may further improve satisfaction for Economy Class passengers.

Multivariate Analysis



To uncover deeper relationships, a multivariate analysis was conducted by examining correlations between satisfaction and service ratings. Key insights include:

- **Correlation Matrix:** The analysis revealed that **Inflight Wi-Fi Service** has the highest correlation with satisfaction (0.63), followed by **Online Boarding** (0.43) and **Ease of Online Booking** (0.32). These findings align with observations from bivariate analysis, confirming their critical impact on customer satisfaction.
- **Segmented Analysis:** For Economy Class passengers:
    - The same three features (**Inflight Wi-Fi Service**, **Online Boarding**, and **Ease of Online Booking**) remain the top contributors to satisfaction.

    - **Inflight Entertainment** also emerges as a notable factor, particularly for business travelers in Economy Class.
- **Limitations:** While correlations indicate relationships between variables, they do not imply causation. Further analysis, potentially involving predictive modeling, is required to establish causal links.
- **Actionable Insights:** The multivariate findings highlight the need to prioritize investments in improving inflight Wi-Fi, streamlining the online booking process, and enhancing the boarding experience to address dissatisfaction effectively.

Challenges and Limitations

Challenges Faced

This project presented several challenges, particularly due to its scale and time constraints:

1. **Large Dataset Handling:** Being the first experience with a relatively large-scale dataset, understanding and managing the multi-column data structure required additional effort.
2. **Insight Generation:** At the initial stages, identifying meaningful patterns or insights was difficult, which delayed the progression to deeper analysis.
3. **Visualization Complexity:** Generating and interpreting effective visualizations took considerable time, limiting the team's ability to explore more advanced analysis techniques.

Limitations

1. **Lack of Pricing Data:** The absence of ticket pricing information limited our ability to analyze the relationship between price and perceived satisfaction. Pricing data could have provided valuable insights into whether higher-paying customers receive better service quality or are more satisfied.
2. **Binary Satisfaction Data:** The dataset only classifies satisfaction as either "Satisfied" or "Neutral/Dissatisfied." This binary categorization lacks the granularity needed for more precise analysis, such as identifying levels of dissatisfaction or degrees of satisfaction.
3. **Data Insensitivity:** Service ratings are recorded on a 5-point interval scale without considering variations within the intervals. This lack of sensitivity could have reduced the depth of insights obtained from the analysis.

Future Work

To further build upon this analysis, several areas of improvement and expansion are suggested:

1. **Incorporate Additional Data:**
    - Include pricing information and more granular satisfaction scores (e.g., a 10-point scale) to better understand customer preferences and expectations.
    - Collect data on specific customer demographics, such as income level or travel frequency, for improved segmentation analysis.

2. **Advanced Analytical Techniques:**
    - Implement machine learning models to predict customer satisfaction based on service ratings and other features.
    - Use clustering algorithms to identify distinct customer segments with similar satisfaction patterns and needs.
3. **Narrow Focus Areas:**
    - Conduct class-specific analyses (e.g., Economy, Economy Plus, Business) to tailor recommendations to each customer segment.
    - Evaluate other contributing factors such as customer loyalty programs and their impact on satisfaction.
4. **Time Management Enhancements:**
    - Establish a clear schedule and prioritize tasks to allow for deeper exploration of selected areas, particularly under tight deadlines.

Description of any additional ideas or approaches that were not implemented

During the course of this project, several additional ideas and approaches were identified but not implemented due to resource constraints. These include:

1. **Correlation Analysis by Travel Purpose:**
    - We explored the idea of segmenting the dataset further by travel purpose (e.g., business or personal) to identify how satisfaction factors differ between these groups. While initial results indicated similar trends, deeper segmentation analysis was not conducted.
2. **Advanced Predictive Modeling:**
    - Machine learning techniques, such as logistic regression or decision trees, could have been used to predict satisfaction levels and identify the most critical features with greater accuracy.
3. **Temporal Analysis:**
    - Examining trends over time, such as the impact of flight delays or service changes, could provide additional insights into fluctuating customer satisfaction levels.

Reasons for not implementing these ideas (e.g. time constraints, complexity)

1. **Time Constraints:**

- The limited timeframe of this project made it impractical to delve into advanced analyses or implement new methodologies. A more focused approach was necessary to meet deadlines.

2. **Complexity:**

- The team had limited exposure to advanced statistical and machine learning methods. Learning and applying these techniques within the project scope would have required additional resources and expertise.

3. **Data Availability:**

- Certain analyses, such as temporal trends, required data that was not available in the provided dataset, such as historical customer feedback or operational data.

Summary of the project's objectives and achievements

The primary objective of this project was to analyze customer satisfaction survey data to identify key factors contributing to dissatisfaction and provide actionable recommendations for United Airlines. The project successfully achieved the following:

1. **Key Insights Identified:**
    - Highlighted critical dissatisfaction factors, including **Inflight Wi-Fi Service**, **Ease of Online Booking**, and **Online Boarding**.
    - Discovered additional concerns in Economy Class, such as **Seat Comfort**, **Food and Drink**, and **Inflight Entertainment**.
2. **Data Preprocessing and Visualization:**
    - Successfully prepared the dataset for analysis by handling duplicates, null values, and invalid data.
    - Created clear visualizations to support findings and provide actionable insights.
3. **Actionable Recommendations:**
    - Provided targeted suggestions for service improvements based on the most impactful features affecting satisfaction.

Key takeaways and insights gained

1. **Critical Areas for Improvement:**
    - **Inflight Wi-Fi Service** emerged as the most significant pain point for dissatisfied passengers.
    - Improving **Ease of Online Booking** and **Online Boarding** could have a substantial impact on satisfaction, particularly for Economy Class travelers.
2. **Segment-Specific Findings:**
    - Economy Class accounts for the majority of dissatisfied customers, necessitating focused improvements in this segment.
3. **Team Collaboration:**
    - The project demonstrated the importance of efficient teamwork, particularly in handling large datasets and managing time effectively.
4. **Future Opportunities:**
    - The project highlighted areas for future analysis, such as advanced modeling and segmentation, to enhance the depth of insights.

Reference

1. Kaggle. *Airline Passenger Satisfaction Dataset.* Retrieved from https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction.
2. United Airlines. *MileagePlus Program.* Retrieved from https://www.united.com/en/us/fly/mileageplus.html.
3. Wikipedia. *United Airlines.* Retrieved from https://en.wikipedia.org/wiki/United_Airlines.

Distribution of Work

All team members contributed equally across various aspects of the project:

- **Code Implementation:** But But, Charles Wong, Hazel Lo
- **PowerPoint Presentation:** But But, Charles Wong, Hazel Lo
- **Report Writing:** But But, Charles Wong, Hazel Lo

The collaborative effort ensured a comprehensive approach to the project deliverables, with each team member playing a crucial role in its successful completion.

This is the end of the report.