# AIRLINE CUSTOMER SATISFACTION

## Dashboard Construction Notebook

SFSU
MSBA Spring 2024
ISYS 850

Simran Suri, Aditya Kochle, Thanapol Singjaroen & Lina Perez

## Table of Contents

## Introduction

In today's airline industry, customer satisfaction is key. Airlines must continuously improve the passenger experience to retain loyalty and attract new customers. Understanding satisfaction levels is vital for identifying areas for enhancement.

This dataset, sourced from an airline, provides insights into passenger satisfaction, covering factors like gender, customer type, travel purpose, class, flight distance, and satisfaction ratings across diverse services.

Assessing customer satisfaction helps airlines gauge service effectiveness, retain customers, and avoid negative feedback. By analyzing this dataset in Tableau, we aim to uncover insights to guide airlines in service optimization and customer retention, thus increasing premium passengers. This report documents our exploratory analysis process.

## Data Preparation

### Data Source

The dataset utilized for this analysis is sourced from an airline and can be accessed on Kaggle at https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction/data. It includes insights from an airline passenger satisfaction survey, featuring a broad range of demographic and travel-related attributes. Additionally, the dataset includes satisfaction ratings across various service dimensions, making it a comprehensive resource for exploring passenger preferences and satisfaction levels within the airline industry.

### Data Specification and Data Cleaning

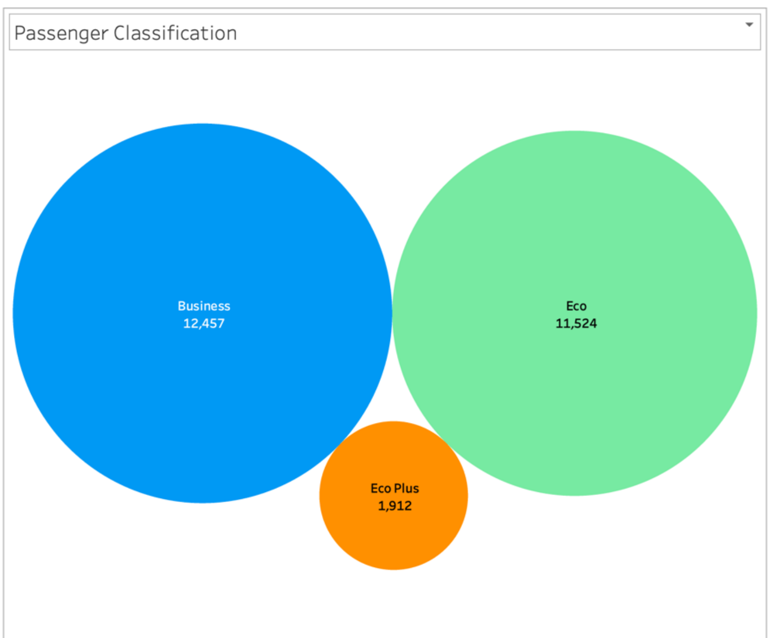
This dataset contains information about an airline passenger satisfaction survey. It has 25 columns and 26,000 rows. Before using the dataset for exploratory data analysis, we need to ensure that the dataset is clean. This means checking for missing data, duplicate rows, outliers in numerical variables, and typos in categorical variables. Once we clean the data, we will have a dataset that is conducive for analysis. We used Python to clean the dataset by identifying and removing 83 rows with missing values. This process ensured that the data was consistent and ready for analysis. The final dataset is the CSV file "Clean_PassengerSatisfaction.csv" and contains 25,893 rows without any duplicates, outliers, or typos. The screenshots of the data cleaning process can be found in appendix 1.

## Question & Graphs

### Initial Question

As we explore the construction of the dashboard, our primary focus centers on optimizing customer service and retention, particularly for premium passengers. The fundamental goal is to enhance the flying experience for premium clientele, ensuring their satisfaction and fostering long-term loyalty. To achieve this objective, we aim to address the next question: What are the key factors influencing customer satisfaction and loyalty among passengers? To answer the previous question, we created 5 graphs.

### Graph 1: Passengers Distribution




Caption

Passenger by class, where color and circle size indicate travel class and relative passenger volume, respectively.

Based on the dataset from the Airline in the "Clean_PassengerSatisfaction.csv" file, we conducted an analysis using a bubble chart to understand the demographics of our passengers and their satisfaction levels. Our initial focus was on identifying the majority passenger segment and assessing whether premium passengers significantly contribute to our revenue.

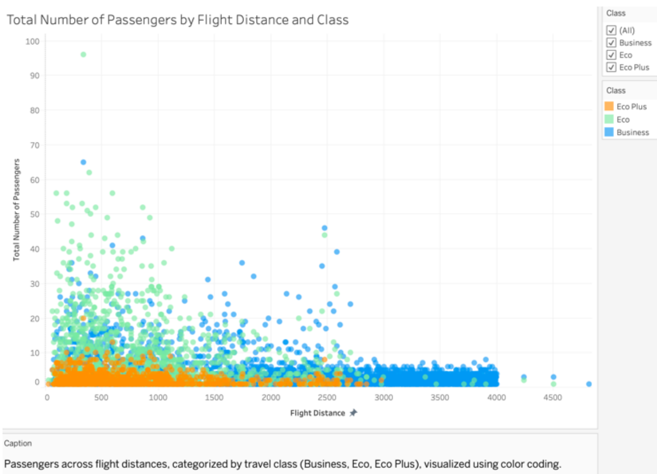
The bubble chart provides a clear visualization of the passenger categories: "Business," "Eco Plus," and "Eco." It is evident that the "Business" category dominates, representing approximately 55% of our surveyed passengers. This suggests that a sizable portion of our passengers are premium travelers. The "Eco Plus" category accounts for around 35%, indicating a smaller but notable segment, while the "Eco" category makes up the remaining 10%.

Given that the Business class comprises the largest share, we delved deeper into understanding areas where Business class passengers might be less satisfied. This analysis aims to pinpoint specific aspects within the Business class experience that require improvement, leveraging insights from the survey data.

**Construction Process:**

Based on the Passenger Satisfaction dataset we aimed to classify passengers into three categories: Business, Economy, and Economy Plus. Our primary goal was to identify the predominant passenger segment. Initially, we considered using a bar chart for this purpose. However, upon further reflection, we realized that the bar chart did not effectively convey the relationships between the categories. Undeterred, we experimented with a Venn diagram. Unfortunately, since there were no commonalities among the three types of passengers, the Venn diagram proved to be ineffective. We then conducted an analysis using a bubble chart to understand the demographics of our passengers and their satisfaction levels. We believed that the varying sizes of bubbles could indicate the relative importance or presence of each category.

### Graph 2: Total Number of Passengers by Flight Distance and Class

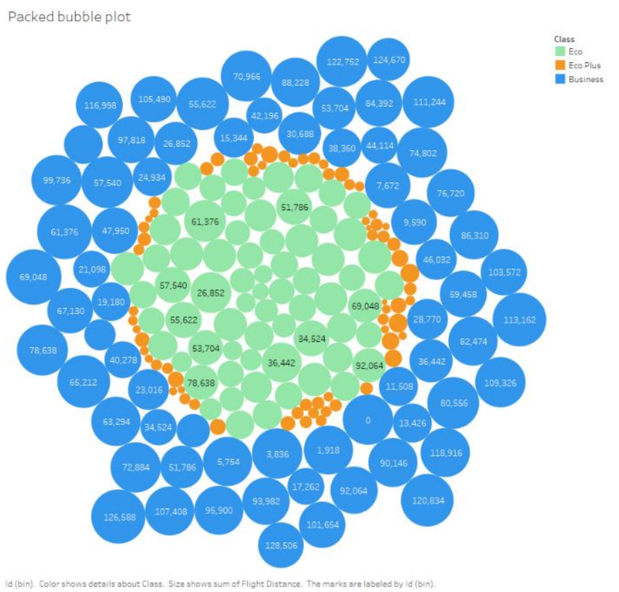


Caption
Passengers across flight distances, categorized by travel class (Business, Eco, Eco Plus), visualized using color coding.

Business class dominates distance vs number of passengers

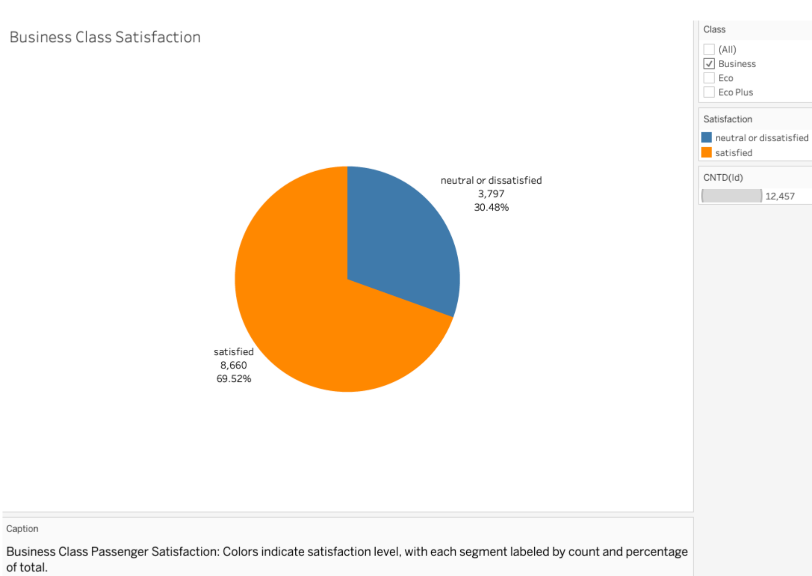
Airlines promote long-haul and ultra long-haul flights (2000+ mi) with premium services to generate more revenue. Options like extra legroom, location of seat, diverse food and beverage options is considered a pre-requisite within premium class. We expected to find a relation between the furthest distance travelled and the class of travel of all passengers. For this purpose, we chose the scatter plot to perform our analysis. We decided to proceed with the same color combination as above to distinguish the passengers in different classes. First impressions of the visualization portray passengers from the business class traveling the furthest. Out of the total passengers, 21% of passengers travelled more than 2000+ mi. Off these passengers 89% passengers constituted the business class passengers. Based on this observation, we decided to explore areas influencing passenger behavior in business class of travel.

**Construction Process:**




We tried a different approach. This time we decided to use the bubble plot. The intent was to have the inter-bubble distance as further for longer distances, colors to distinguish class of travel and labels on each bubble to show the passenger count. For this, the 'flight distances' measure and the 'Id' dimension were chosen. The 'Id' dimension was converted to count distinct. The class was used as a differentiation factor to convert to show distinct differences in class of travel. The circle's positions were linked to the distances they travelled while the size of the circles denoted the number of passengers from that class. Figure shows the final plot. We were not satisfied with the final product. That is how we ended up with the scatter plot.

## Graph 3: Business Class Satisfaction




Caption
Business Class Passenger Satisfaction: Colors indicate satisfaction level, with each segment labeled by count and percentage of total.

Once we identified that most of our passengers were premium travelers, the next critical question was assessing their satisfaction with our airline services. To delve into this, we created a detailed pie chart showing two main proportions: satisfied passengers and those who are neutral or dissatisfied. The chart revealed that approximately one-third of premium passengers fall into the neutral or dissatisfied category regarding our services. This insight is invaluable as it directs our attention towards improving the experience for this significant segment of passengers. By focusing on addressing the concerns of these neutral or dissatisfied passengers, we aim to enhance overall satisfaction levels and strengthen our reputation for premium service quality.

**Construction Process:**

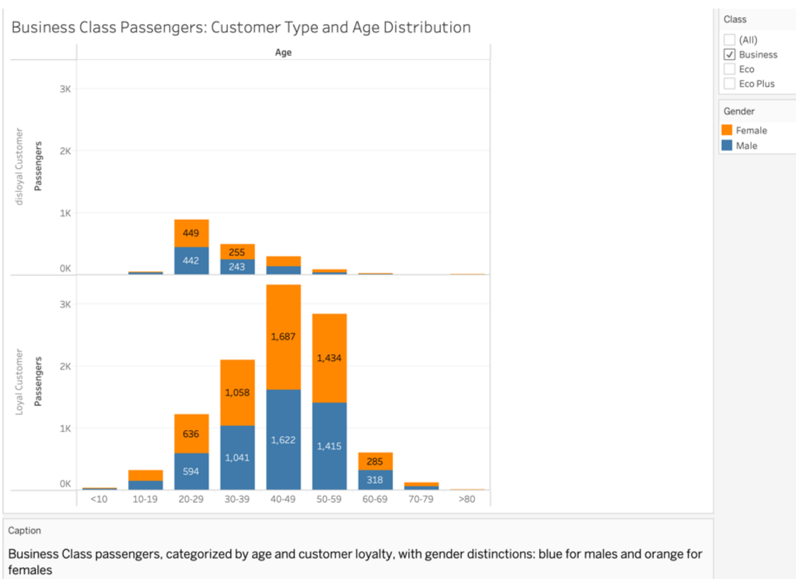
Once we identified that most of our passengers were premium travelers, the next critical question was assessing their satisfaction with the airline services. To delve into this, we decided to create a detailed pie chart highlighting two main proportions: satisfied passengers and those who are neutral or dissatisfied. Initially, while dragging and dropping columns and rows, we encountered a challenge. The chart options area, known as "Show me," did not offer a direct pie chart option. It became apparent that specific dimensions and measures were necessary to create the desired visualization.

So, we worked on defining the required dimensions and measures. Through careful data organization and manipulation, we successfully overcame this limitation and generated a pie chart to represent the satisfaction levels of business class passengers.

However, it did not end with the creation of the pie chart. We sought to enhance its effectiveness by incorporating both the percentage signs and numerical values. Our intention was to provide viewers with a clear understanding of the proportion of satisfied and dissatisfied business class passengers. Leveraging Tableau's formatting options, we skillfully added these elements to the pie chart, effectively communicating the percentages and elucidating the satisfaction levels.

In conclusion, we got a detailed pie chart clearly displaying the number as well as the percentage of satisfied and dissatisfied Business class passengers.

## Graph 4: Business Class Passengers: Customer Type and Age Distribution



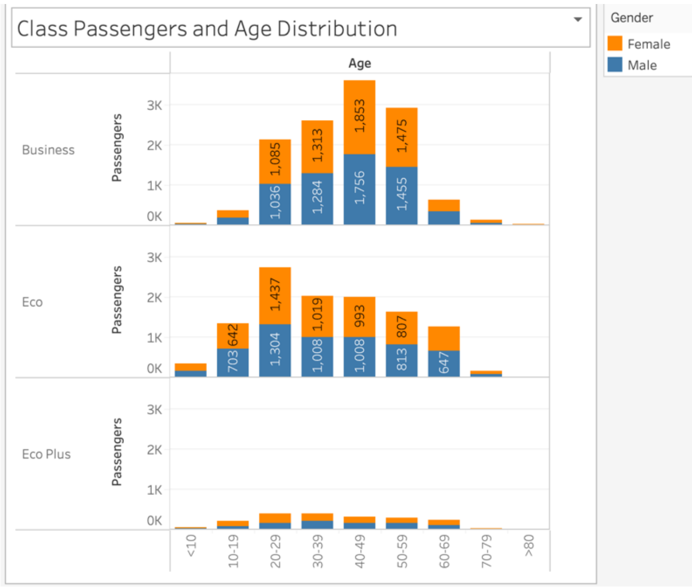

The graph illustrates the distribution of business class of 'Loyal' and 'Disloyal' customers across various age groups, segmented by gender (orange for females, blue for males).

The age groups 30-39,40-49, and 50-59 are the most populated, suggesting these middle-aged passengers are prime targets for targeted marketing and service enhancements due to their significant numbers.

Gender analysis reveals a balanced representation overall, but some age groups, like 40-49, display a slightly higher loyalty among males. This suggests potential gender-based preferences or satisfaction levels that could be explored further. The graph also highlights a higher proportion of disloyal customers in the youngest (<10 and 10-19). Strategies could include developing age-specific services or loyalty programs tailored to these groups.

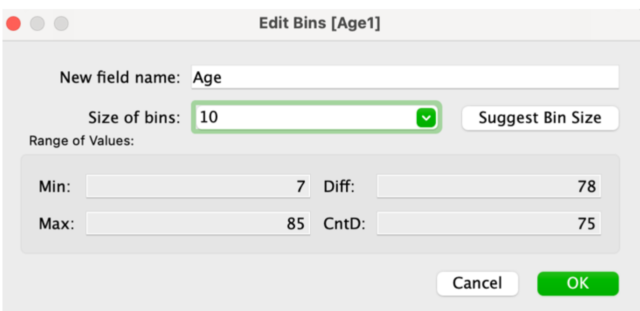
In conclusion, while middle-aged passengers show strong loyalty, opportunities exist to enhance service for younger to further boost overall passenger satisfaction and loyalty.

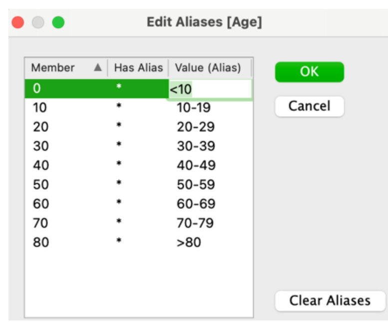
**Construction Process:**




In our analysis of the airline's customer satisfaction dataset, we initially focused on the age distribution across different passenger classes—Eco, Eco Plus, and Business. Recognizing that Business class passengers, who often travel longer distances, could significantly influence the airline's revenue, we decided to narrow our focus. We opted to filter specifically for Business class to delve deeper into this demographic. Additionally, we introduced a passenger loyalty dimension to our analysis. This allowed us to explore the relationship between age and loyalty within the Business class segment, aiming to identify strategic targets based on these attributes.

To create the previous graph, we segmented the age data into ten bins derived from the 'Age' column, as illustrated in figure (). Additionally, to enhance the graph's readability, we modified the aliases for each age group as shown in the figure with tittle "Edit Aliases."




## Graph 5: Business Passenger Satisfaction Analysis

After identifying the majority of age brackets of business-class passengers, we would like to see how well the airline provided services to this passenger group. We want to identify services that underperformed so that the airline can take action to improve the services, thus attracting more premium passengers. To get a general idea of how business-class passengers think positively or negatively about each airline service, we adopted a Likert scale, which shows the number of negative and positive responses. In other words, the further right that a bar leans, the more positive the passengers' responses are. The Likert scale is usually defined as the amount to which someone agrees or disagrees with something. It allows us to measure attitudes and opinions with a greater degree of nuance than just a yes or no question.

In our dataset, the survey questions are in the format that goes from left to right column **(Figure 1)**; however, this format is not conducive to creating a Likert scale. After researching how to tackle this issue, we found a solution that required pivoting the table. We needed to pivot each question column into one column and each response value into the next adjacent column **(Figure 2)**.

| # Clean_PassengerSatisfaction.csv Inflight wifi service | # Clean_PassengerSatisfaction.csv Departure/Arrival time ... | # Clean_PassengerSatisfaction.csv Ease of Online booking |
|---|---|---|
| 5 | 4 | 3 |
| 1 | 1 | 3 |
| 2 | 0 | 2 |
| 0 | 0 | 0 |
| 2 | 3 | 4 |
| 3 | 3 | 3 |
| 5 | 5 | 5 |

**Figure 1:** An example of the format of survey questions and their values before pivoting

| Abc Pivot Questions | # Pivot Pivot Field Values |
|---|---|
| Baggage handling | 5 |
| Checkin service | 2 |
| Cleanliness | 5 |
| Departure/Arrival time conve... | 4 |
| Ease of Online booking | 3 |

**Figure 2:** The format of survey questions and their values after pivoting

To start creating the visualization, we first tried to put the "questions" variable on the row shelf and their response values column on the column shelf. As a result, we got the total response value. However, we did not want to add these values up. We actually wanted our response values of questions to be a dimension, so we converted "Pivot Field Values" from a measure to a dimension. Then, we got a zero-to-five scale in the column. Then, we dragged the "ID" variable into the text shelf and set the "ID" to a count distinct. This showed us how many responses there are in each of these categories **(Figure 3).**

| | Pivot Field Values | | | | | |
|---|---|---|---|---|---|---|
| Questions | 0 | 1 | 2 | 3 | 4 | 5 |
| Baggage handling | | 1,785 | 2,833 | 5,204 | 9,357 | 6,714 |
| Checkin service | | 3,206 | 3,202 | 6,987 | 7,254 | 5,244 |
| Cleanliness | 2 | 3,404 | 3,968 | 6,046 | 6,771 | 5,702 |
| Departure/Arrival time co.. | 1,374 | 3,899 | 4,336 | 4,399 | 6,312 | 5,573 |
| Ease of Online booking | 1,193 | 4,342 | 6,021 | 5,927 | 4,854 | 3,556 |
| Food and drink | 25 | 3,210 | 5,375 | 5,474 | 6,183 | 5,626 |
| Gate location | | 4,415 | 4,823 | 7,122 | 6,023 | 3,510 |
| Inflight entertainment | 4 | 3,193 | 4,318 | 4,725 | 7,347 | 6,306 |
| Inflight service | 2 | 1,775 | 2,838 | 5,005 | 9,352 | 6,921 |
| Inflight wifi service | 812 | 4,469 | 6,481 | 6,298 | 4,965 | 2,868 |
| Leg room service | 126 | 2,536 | 5,000 | 4,940 | 7,075 | 6,216 |
| On-board service | 2 | 2,906 | 3,658 | 5,690 | 7,814 | 5,823 |
| Online boarding | 651 | 2,558 | 4,417 | 5,296 | 7,682 | 5,289 |
| Seat comfort | | 3,028 | 3,616 | 4,617 | 7,969 | 6,663 |

**Figure 3:** The survey questions table with the number of responses to each question and the Likert score

In the next step, we must create a divergence bar chart to draw a Likert scale. We also need to create six new separate calculations. Each calculation will represent a scale from zero to five. We used a count distinct of passengers' "ID" with an IF condition to give us the number of responses of each Likert score **(Figure 4)**. For the Likert score "3", we divided it into two since it represented the middle of the bar. Then, we dragged these six calculations into the table to get a stacked bar. We created another calculation field to calculate a net promoter score **(Figure 5)**, then changed the color of each score and rearranged the sequence (very dissatisfied or "0" to the left and very satisfied or "5" to the right). As a result, we had the Likert scale, as shown in **Figure 6**.

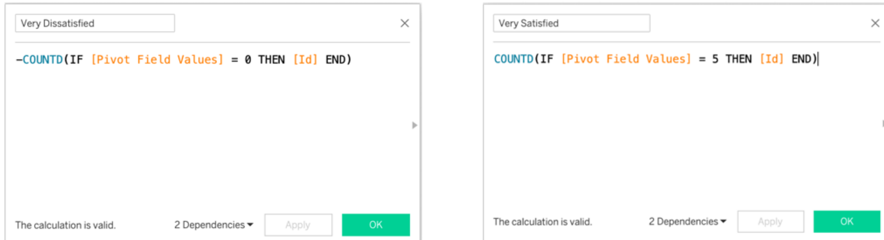
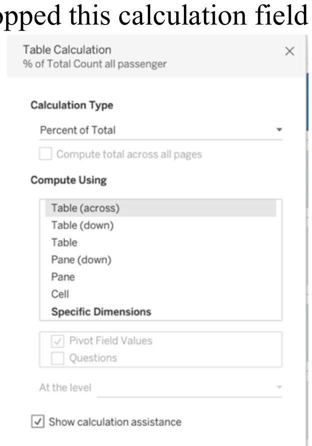


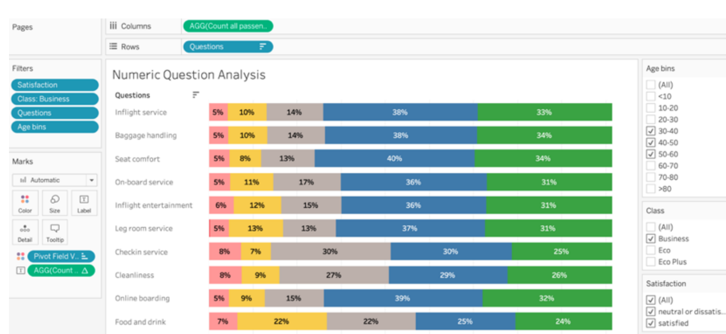

**Figure 4:** Examples of creating a calculation field for each Likert score


**Figure 5:** The calculation field of Net Promoter Score


**Figure 6:** A Likert scale graph

However, this graph did not give us rich information. We decided to create another stacked bar graph showing the percentage of each like score. First, we made a calculation field named "Count all passengers," which used a count distinct function to give us the total number of passengers. Then, we dropped this calculation field in the column shelf and saw 25,893 responses in each question. We



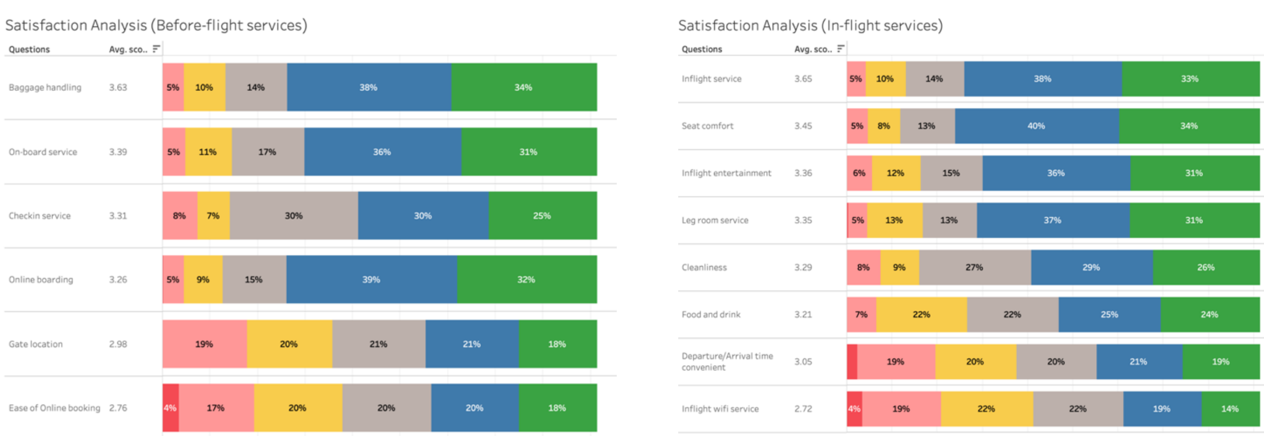
dragged "Pivot Field Values" to color and changed it to a dimension, giving us distinct colors for each like score. To create a percentage of each like score, we made a copy of "Count all passengers" by using a control dragging "Count all passengers" from the column shelf to Label. After that, we used the "add calculation table" to calculate the percentage by rows **(Figure 7)**. We also formatted the percentage that showed on the bar since the bar looked visually busy. In the format, we selected a percentage format and then reduced the number of decimals to zero. Now, we had a bar chart that showed us the percentage of each like score in each question **(Figure 8)**.

We then divided the services into before-flight and in-flight services, added an average score for each service, and put it next to the question to get a general idea of sentiment and quickly identify services that needed immediate improvement. To set this up, we needed another calculation field that used a "FIXED" function and an average ("AVG") function **(Figure 9)**. Then, we added a filter by questions to create a separate graph for before-flight and in-flight services **(5th Visualization)**. We visually identified that the ease of online booking and in-flight Wi-Fi had the lowest satisfaction scores, with approximately 40 percent of dissatisfied business-class passengers. Therefore, the airlines must improve the quality of the ease of online booking and in-flight Wi-Fi services had it wants to increase the number of premium passengers.

**Figure 7:** The setting in calculation table

**Figure 8:** A stacked bar with the percentage of each Likert score by question


**Figure 9:** a calculation field for an average satisfaction score by question




**5th Visualization:** Percentage breakdown of business passenger Likert scores by Question

## Recommendation Summarize the areas for further analysis?

From the different charts within the dashboards, it is evident that 'Business' class of travel dominates with its high passenger count, satisfaction rating and number of loyalty members. Some of the areas uncovered for exploratory analysis are as follows which we are addressed point by point in Appendix 2: Areas for further analysis:

1) Class preference analysis
2) Operational Optimization
3) Customer satisfaction enhancement
4) Revenue segmentation
5) Targeted marketing

## Conclusion

Why are premium passengers important? What are areas that need to improve? What action could be taken to improve the experience of this passenger segment?

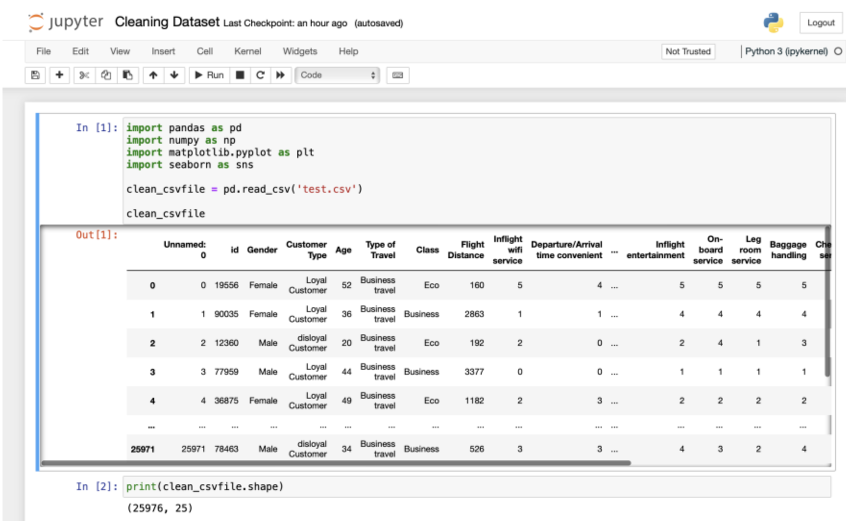
Premium passengers play a crucial role in the airline industry for several reasons. Firstly, they contribute significantly to revenue generation, as they often purchase higher-priced tickets and avail themselves of premium services. Secondly, premium passengers tend to be more loyal and have a higher lifetime value, making them an essential customer segment for airlines to retain. Lastly, premium passengers can influence the perception and reputation of an airline, as their experiences and satisfaction levels have the potential to attract new customers.

The analysis of the Airline Passenger Satisfaction dataset revealed areas that need improvement to enhance the experience of passengers. While the Business class dominated in terms of passenger count, there is an opportunity to identify specific aspects within the Business class experience that require attention. By understanding the factors that influence customer satisfaction and loyalty among premium passengers, airlines can focus their efforts on addressing these areas such as ease of online booking and in-flight Wi-Fi services that have the lowest average score in customer satisfaction.

In conclusion, premium passengers are vital for airlines due to their significant contribution to revenue, loyalty, and brand reputation. To improve the experience of this passenger segment, airlines should focus on enhancing premium services, prioritizing customer service, leveraging technology, and actively seeking feedback. By consistently delivering exceptional experiences, airlines can foster long-term loyalty and attract new premium passengers, ensuring sustainable growth in a competitive industry.

# Appendix

## Appendix 1: Data Cleaning in python



jupyter Cleaning Dataset Last Checkpoint: an hour ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Not Trusted

Python 3 (ipykernel)

Run

Code

In [1]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

clean_csvfile = pd.read_csv('test.csv')

clean_csvfile
```

Out[1]:

| | Unnamed: 0 | id | Gender | Customer Type | Age | Type of Travel | Class | Flight Distance | Inflight wifi service | Departure/Arrival time convenient | ... | Inflight entertainment | On-board service | Leg room service | Baggage handling | Che ser |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 19556 | Female | Loyal Customer | 52 | Business travel | Eco | 160 | 5 | 4 | ... | 5 | 5 | 5 | 5 | |
| 1 | 1 | 90035 | Female | Loyal Customer | 36 | Business travel | Business | 2863 | 1 | 1 | ... | 4 | 4 | 4 | 4 | |
| 2 | 2 | 12360 | Male | disloyal Customer | 20 | Business travel | Eco | 192 | 2 | 0 | ... | 2 | 4 | 1 | 3 | |
| 3 | 3 | 77959 | Male | Loyal Customer | 44 | Business travel | Business | 3377 | 0 | 0 | ... | 1 | 1 | 1 | 1 | |
| 4 | 4 | 36875 | Female | Loyal Customer | 49 | Business travel | Eco | 1182 | 2 | 3 | ... | 2 | 2 | 2 | 2 | |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| 25971 | 25971 | 78463 | Male | disloyal Customer | 34 | Business travel | Business | 526 | 3 | 3 | ... | 4 | 3 | 2 | 4 | |

In [2]:

```
print(clean_csvfile.shape)
```

```
(25976, 25)
```

jupyter Cleaning Dataset Last Checkpoint: an hour ago (autosaved)

File Edit View Insert Cell Kernel Widgets Help | Not Trusted | Python 3 (ipykernel)

In [3]:
```python
clean_csvfile.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 25976 entries, 0 to 25975
Data columns (total 25 columns):
 #   Column                             Non-Null Count  Dtype
---  ------                             --------------  -----
 0   Unnamed: 0                         25976 non-null  int64
 1   id                                 25976 non-null  int64
 2   Gender                             25976 non-null  object
 3   Customer Type                      25976 non-null  object
 4   Age                                25976 non-null  int64
 5   Type of Travel                     25976 non-null  object
 6   Class                              25976 non-null  object
 7   Flight Distance                    25976 non-null  int64
 8   Inflight wifi service              25976 non-null  int64
 9   Departure/Arrival time convenient  25976 non-null  int64
 10  Ease of Online booking             25976 non-null  int64
 11  Gate location                      25976 non-null  int64
 12  Food and drink                     25976 non-null  int64
 13  Online boarding                    25976 non-null  int64
```

In [4]:
```python
# check NaN values

check_nan_in_data = clean_csvfile.isnull().values.any()
print(check_nan_in_data)
```

```
True
```

In [5]:
```python
# Count NaN values

count_nan_values_in_data = clean_csvfile.isnull().sum().sum()
print(count_nan_values_in_data)
```

```
83
```

jupyter Cleaning Dataset Last Checkpoint: an hour ago (autosaved)

File Edit View Insert Cell Kernel Widgets Help | Not Trusted | Python 3 (ipykernel)

In [6]:
```python
# Count NaN values by column

clean_csvfile.isnull().sum()
```

```
Flight Distance                       0
Inflight wifi service                 0
Departure/Arrival time convenient     0
Ease of Online booking                0
Gate location                         0
Food and drink                        0
Online boarding                       0
Seat comfort                          0
Inflight entertainment                0
On-board service                      0
Leg room service                      0
Baggage handling                      0
Checkin service                       0
Inflight service                      0
Cleanliness                           0
Departure Delay in Minutes            0
Arrival Delay in Minutes             83
satisfaction                          0
dtype: int64
```

In [7]:
```python
# Eliminate rows of missed values

clean_csvfile.dropna(inplace=True)
```

In [9]:
```python
clean_csvfile.shape

#83 rows are deleted
```

```
Out[9]: (25893, 25)
```

jupyter Cleaning Dataset Last Checkpoint: an hour ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

Out[9]: (25893, 25)

In [10]:
```
# Check if there is any repeated row

print(f'Size of the set before eliminate the repeated rows: {clean_csvfile.shape}')
clean_csvfile.drop_duplicates(inplace=True)

print(f'Size of the set after eliminate the repeated rows: {clean_csvfile.shape}')

# No repeated row
```

```
Size of the set before eliminate the repeated rows: (25893, 25)
Size of the set after eliminate the repeated rows: (25893, 25)
```

In [11]:
```
# check numerical columns if there is any outstanding value

clean_csvfile.describe()

# all min and max values in each numerical column look reasonable
```

Out[11]:

| | Unnamed: 0 | id | Age | Flight Distance | Inflight wifi service | Departure/Arrival time convenient | Ease of Online booking | Gate location | Food and drink | Online boarding | S |
|---|---|---|---|---|---|---|---|---|---|---|---|
| count | 25893.000000 | 25893.000000 | 25893.000000 | 25893.000000 | 25893.000000 | 25893.000000 | 25893.000000 | 25893.000000 | 25893.000000 | 25893.000000 | 25 |
| mean | 12987.838566 | 65021.974858 | 39.621983 | 1193.753254 | 2.723709 | 3.046422 | 2.755996 | 2.976442 | 3.214923 | 3.261615 | |
| std | 7499.175165 | 37606.098635 | 15.134224 | 998.626779 | 1.334711 | 1.532971 | 1.412552 | 1.281661 | 1.331895 | 1.355505 | |
| min | 0.000000 | 17.000000 | 7.000000 | 31.000000 | 0.000000 | 0.000000 | 0.000000 | 1.000000 | 0.000000 | 0.000000 | |
| 25% | 6496.000000 | 32209.000000 | 27.000000 | 414.000000 | 2.000000 | 2.000000 | 2.000000 | 2.000000 | 2.000000 | 2.000000 | |
| 50% | 12984.000000 | 65344.000000 | 40.000000 | 849.000000 | 3.000000 | 3.000000 | 3.000000 | 3.000000 | 3.000000 | 4.000000 | |
| 75% | 19482.000000 | 97623.000000 | 51.000000 | 1744.000000 | 4.000000 | 4.000000 | 4.000000 | 4.000000 | 4.000000 | 4.000000 | |
| max | 25975.000000 | 129877.000000 | 85.000000 | 4983.000000 | 5.000000 | 5.000000 | 5.000000 | 5.000000 | 5.000000 | 5.000000 | |


```
Size of the set after eliminate the repeated rows: (25893, 25)
```

]:
```
# check numerical columns if there is any outstanding value

clean_csvfile.describe()

# all min and max values in each numerical column look reasonable
```

]:

| | Unnamed: 0 | id | Age | Flight Distance | Inflight wifi service | Departure/Arrival time convenient | Ease of Online booking | Gate location | Food and drink | Online boarding | S |
|---|---|---|---|---|---|---|---|---|---|---|---|
| count | 25893.000000 | 25893.000000 | 25893.000000 | 25893.000000 | 25893.000000 | 25893.000000 | 25893.000000 | 25893.000000 | 25893.000000 | 25893.000000 | 25 |
| mean | 12987.838566 | 65021.974858 | 39.621983 | 1193.753254 | 2.723709 | 3.046422 | 2.755996 | 2.976442 | 3.214923 | 3.261615 | |
| std | 7499.175165 | 37606.098635 | 15.134224 | 998.626779 | 1.334711 | 1.532971 | 1.412552 | 1.281661 | 1.331895 | 1.355505 | |
| min | 0.000000 | 17.000000 | 7.000000 | 31.000000 | 0.000000 | 0.000000 | 0.000000 | 1.000000 | 0.000000 | 0.000000 | |
| 25% | 6496.000000 | 32209.000000 | 27.000000 | 414.000000 | 2.000000 | 2.000000 | 2.000000 | 2.000000 | 2.000000 | 2.000000 | |
| 50% | 12984.000000 | 65344.000000 | 40.000000 | 849.000000 | 3.000000 | 3.000000 | 3.000000 | 3.000000 | 3.000000 | 4.000000 | |
| 75% | 19482.000000 | 97623.000000 | 51.000000 | 1744.000000 | 4.000000 | 4.000000 | 4.000000 | 4.000000 | 4.000000 | 4.000000 | |
| max | 25975.000000 | 129877.000000 | 85.000000 | 4983.000000 | 5.000000 | 5.000000 | 5.000000 | 5.000000 | 5.000000 | 5.000000 | |


]:
```
# save clean data to "Clean_PassengerSatisfaction"

clean_csvfile.to_csv('Clean_PassengerSatisfaction.csv')
```

## Appendix 2: Areas for Further Analysis

It is an unwritten law; people tend to spend more during travel. If the distances are longer, the spending goes up proportionally. Expectations for amenities and in-flight services are maximized. We tried to find this relation in this dataset. We discovered an interrelation between the flight classes, satisfaction, loyalty, and distances. The three most important insights developed may help optimize resource allocation, flight operations and revenue generation.

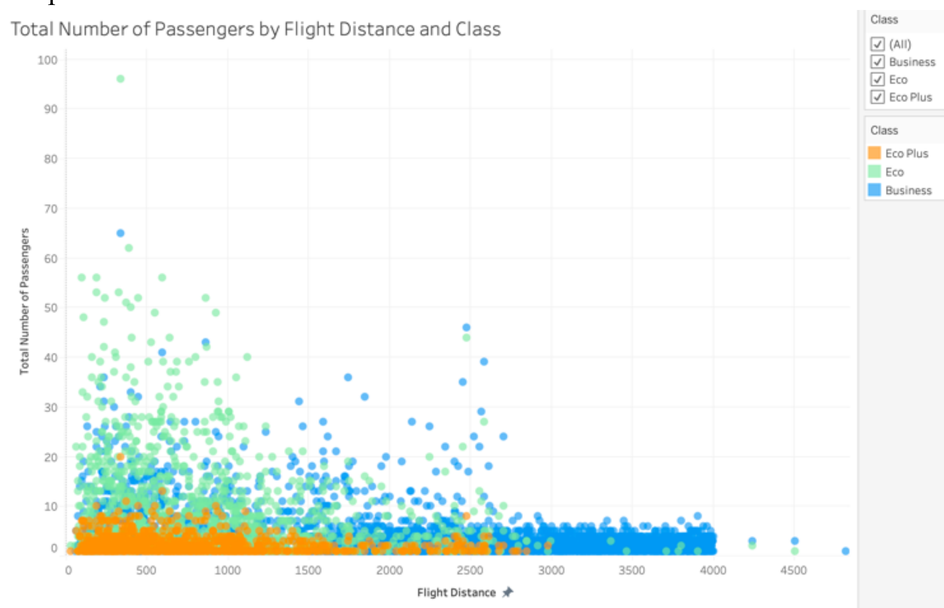
From the different charts within the dashboards, it can be ascertained that ‘Business’ Class of travel dominates with its high passenger count, satisfaction rating and number of loyalty members. Some of the areas uncovered for exploratory analysis are as follows which we will address point by point:

1) Class preference analysis
2) Operational Optimization
3) Customer satisfaction enhancement
4) Revenue segmentation

5) Targeted marketing

## Class preference analysis

The figure below paints a better picture of the flight distances, class of travel and passenger count. Shorter flights being dominated by Economy class while longer flights see a higher proportion of Business class travelers. This can be used to optimize seating arrangements like position of seat, & type of seat offered whereas in the case of service offerings, special exotic meal options or preferences, which can be ordered before the flight. If the customer is a frequent flyer, the airline could keep track of the meal preferences of the passenger. Pricing strategies can be based on the various offerings mentioned in relation to the competition.




Caption

Passengers across flight distances, categorized by travel class (Business, Eco, Eco Plus), visualized using color coding.

## Operational Optimization

Flight distances have an influence on operations and demand optimum resource allocation. Factors such as type of aircraft, passenger load, route planning, fuel consumption & crew scheduling come into play. Some aircraft are purpose built for short haul and ultra long-haul range. Metro cities and holiday destinations have peak passenger demands during weekdays, weekends, or festivities. Routes need to be chartered accordingly if the flights are connecting or direct. Fuel consumption based on passenger load and distance affects revenues. All these factors demand higher operational efficiency.

## Customer satisfaction enhancement

Overall customer satisfaction is summation of multiple factors like seat comfort, inflight amenities, and quality of service. Business class passengers are inclined to have higher expectations in this area. Passenger feedback must be given the utmost importance to find the areas for improvement. This can also be used as an indicator to channelize investments based on findings.

**Targeted marketing**

The figure could also be used to facilitate analysis on travel behavior of the passengers. In the case of business travelers' preference of premium classes for long-haul flights while leisure travelers may opt for economy class on shorter routes. This can be used to tailor marketing campaigns for destinations, class of travel and service expectations, loyalty programs for points accumulation leading to free sector travels and service offerings approportionate to the class.

**Revenue Segmentation**

Through the chart above (Figure), further analysis can be done to analyze revenue patterns and identify high-yield routes to maximize profits.