Data Visualization

Author: Mai Trần Phương Nhi

Contents

[Dataset overview 2](#_Toc154060039)

[Data visualization 3](#_Toc154060040)

[Distribution of Cutomer Satisfaction 3](#_Toc154060044)

[Distribution of 'Gender' and its influence on satisfaction levels 5](#_Toc154060045)

[Distribution of 'Age' and its influence on satisfaction levels 7](#_Toc154060046)

[Sastifaction by Age and Gender 10](#_Toc154060047)

[Distribution of ‘Customer Type’ and its influence on satisfaction levels 11](#_Toc154060048)

[Distribution of ‘Type of Travel’ and ‘Travel Class’ 12](#_Toc154060049)

[Distribution of Flight Distance 14](#_Toc154060050)

[Service quality analysis 17](#_Toc154060051)

[Correlation between delays (departure and arrival) and satisfaction levels 21](#_Toc154060052)

# Dataset overview

This dataset contains a survey of passenger satisfaction on flights. These survey factors are strongly correlated with passenger satisfaction (or dissatisfaction). The data set named “Airline Passenger Satisfaction” includes 2 files “test.csv” and “train.csv” including 25 attributes for each file as follows:

|  |  |
| --- | --- |
| Column | Explain |
| # | Numerical order |
| id | Flight ID code |
| Gender | Customer's gender (Male, Female) |
| Customer Type | Customer type (Loyal customer, disloyal customer) |
| Age | Customer's age |
| Type of Travel | Purpose of the customer's flight (Personal Travel, Business Travel) |
| Class | Customer's ticket class (Business, Eco, Eco Plus) |
| Flight distance | Distance of the flight journey |
| Inflight wifi service | Satisfaction level with in-flight wifi service (0:Not Applicable;1-5) |
| Departure/Arrival time convenient | Level of satisfaction with convenient departure/arrival time |
| Ease of Online booking | Level of satisfaction when booking tickets online |
| Gate location | Level of satisfaction with Gate location |
| Food and drink | Level of satisfaction with food and drinks |
| Online boarding | Satisfaction level with online check-in |
| Seat comfort | Level of satisfaction with seat comfort |
| Inflight entertainment | Level of satisfaction with in-flight entertainment |
| On-board service | Level of satisfaction with on-board service |
| Leg room service | Level of satisfaction with seats with wide legroom |
| Baggage handling | Level of satisfaction with baggage handling |
| Check-in service | Level of satisfaction with Check-in service |
| Inflight service | Level of satisfaction with in-flight service |
| Cleanliness | Level of satisfaction with cleanliness |
| Departure Delay in Minutes | Departure minutes |
| Arrival Delay in Minutes | Number of minutes delayed upon arrival |
| Satisfaction | Customer satisfaction level with the airline (Satisfaction, neutral or dissatisfaction) |

# Data visualization



## Distribution of Cutomer Satisfaction

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, hàng

Mô tả được tạo tự động

FigureX: Distribution of Passenger Satisfaction countplot code

Creating a countplot using Seaborn. It visualizes the distribution of a categorical variable ('satisfaction'). The 'data' parameter specifies the dataframe, and x specifies the column to be plotted.

Ảnh có chứa văn bản, ảnh chụp màn hình, Hình chữ nhật, hình vuông

Mô tả được tạo tự động

FigureX: Distribution of Passenger Satisfaction countplot

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, hàng

Mô tả được tạo tự động

FigureX: Distribution of Passenger Satisfaction pie chart code

First, we count the occurrences of each unique value in the 'satisfaction' column and stores the counts in the satisfaction\_counts variable.

Next, createing a pie chart using the counts from satisfaction\_counts. The labels parameter assigns labels to the pie chart segments based on the unique values in the 'satisfaction' column. The autopct parameter specifies that the percentage of each segment should be displayed with one decimal place.

Ảnh có chứa văn bản, ảnh chụp màn hình, biểu đồ, vòng tròn

Mô tả được tạo tự động

FigureX: Distribution of Passenger Satisfaction pie chart

This analysis reveals that the number of dissatisfied passengers exceeds that of satisfied passengers by a noticeable margin, indicating an imbalance in our train data.

## Distribution of 'Gender' and its influence on satisfaction levels



FigureX: Distribution of Gender code

Line 1: Calculating the counts of each unique value in the 'Gender' column and stores the result in the gender\_counts variable.

Line 2: Calculating the percentage distribution of each gender category by dividing the counts by the total number of rows in the training dataset (len(train)) and multiplying by 100. The results are stored in the gender\_percentage variable.

Generating a Waffle chart to illustrate the distribution of 'Gender' in the training data. The chart represents the percentage of females and males using icons (female and male symbols).

Ảnh có chứa Phông chữ, thiết kế

Mô tả được tạo tự động

FigureX: Distribution of Gender

The distribution of gender among passengers is nearly balanced, with a relatively equal representation of male and female travelers.

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, hàng

Mô tả được tạo tự động

Figure: Create a contingency table (cross-tabulation)

Using the pd.crosstab function from the Pandas library to create a contingency table (cross-tabulation) that shows the relationship between the 'Gender' and 'satisfaction' columns in the DataFrame train.

The resulting table, named cross\_tab, displays the count of occurrences for each combination of gender and satisfaction level. Each row represents a gender category ('Female' or 'Male'), and each column represents a satisfaction level ('neutral or dissatisfied' or 'satisfied'). The values in the table indicate how many passengers fall into each category based on their gender and satisfaction level.

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, biên lai

Mô tả được tạo tự động

Figure: Cross-tabulation between the 'Gender' and 'satisfaction'

There are 22534 satisfied and 30193 dissatisfied female customer.

There are 22491 satisfied and 28868 dissatisfied male customer.

Ảnh có chứa văn bản, Phông chữ, hàng, ảnh chụp màn hình

Mô tả được tạo tự động

Figure: Distribution of satisfaction between genders code

Ảnh có chứa văn bản, ảnh chụp màn hình, biểu đồ, Phông chữ

Mô tả được tạo tự động

Figure: Distribution of satisfaction between genders pie chart

In general, the data indicates an even distribution of satisfaction between genders, with a slightly higher percentage of satisfied individuals in both male and female groups.

## Distribution of 'Age' and its influence on satisfaction levels

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, hàng

Mô tả được tạo tự động

Figure: List all the unique value in ‘Age’

List all the unique value in ‘Age’

Ảnh có chứa văn bản, Phông chữ, ảnh chụp màn hình, chữ viết tay

Mô tả được tạo tự động

Figure: Number of Passengers by Age plot code

Ảnh có chứa ảnh chụp màn hình, Sơ đồ, biểu đồ, văn bản

Mô tả được tạo tự động

Figure: Number of Passengers by Age plot

The individuals in the dataset span a wide range of ages, from 7 to 85 years old. This broad age range reflects the diversity of generations.

**Predominant age group among the customers:**

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, hàng

Mô tả được tạo tự động

Figure: Define age group bins and labels and count the number of them

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, số

Mô tả được tạo tự động

Figure: Number of customer age groups

Ảnh có chứa văn bản, ảnh chụp màn hình, biểu đồ, số

Mô tả được tạo tự động

Figure: Number of customer age groups countplot

Overall, the data exhibits a diverse distribution across various age groups, with a notable concentration in the '41-50,' '21-30,' and '31-40' categories. Meanwhile, the '0-20' group. Although they are the smallest, still represents a significant portion of the dataset."

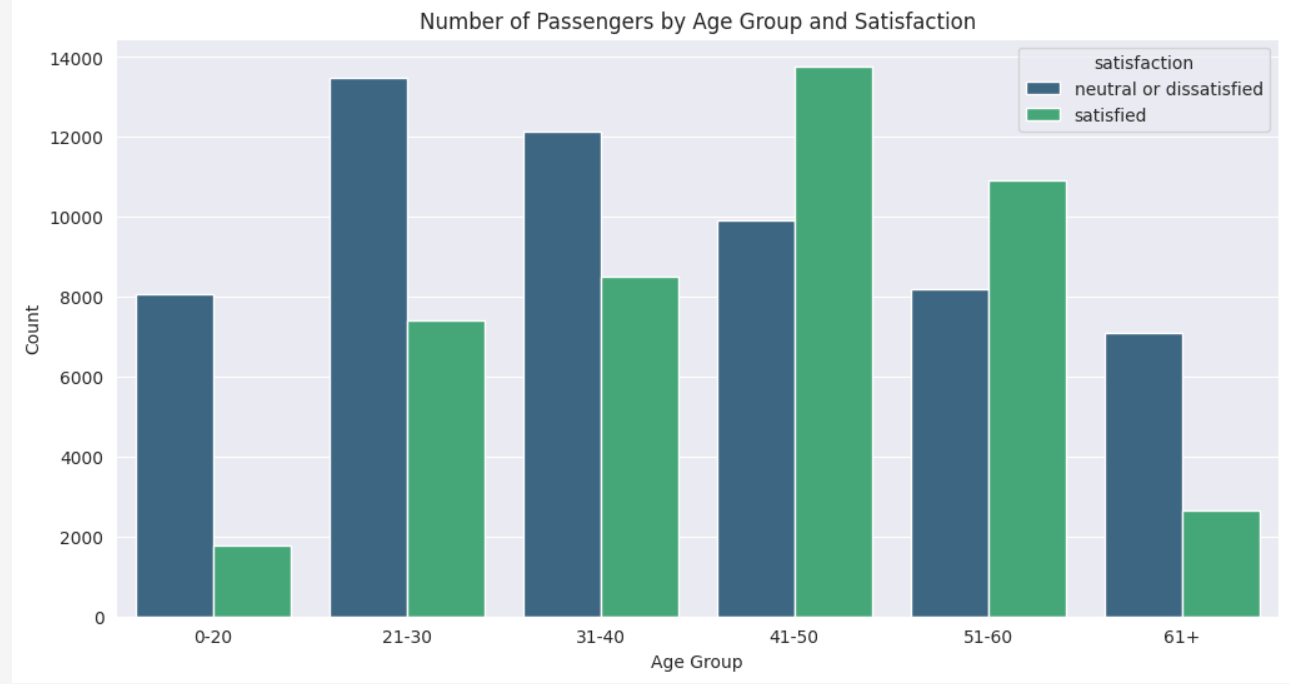


Figure: Number of Passengers by Age Group and Satisfaction countplot

To sum up, the data uncovers intriguing satisfaction patterns among various age groups. Higher satisfaction levels are observed in the 41-50 and 51-60 age brackets, whereas the 0-20 and 61+ age groups demonstrate larger proportions of individuals expressing neutrality or dissatisfaction. The 31-40 age group falls in the middle, suggesting a relatively balanced distribution of sentiments.

## Sastifaction by Age and Gender

Ảnh có chứa văn bản, ảnh chụp màn hình, phần mềm, màn hình

Mô tả được tạo tự động

Figure: 3.4.4. Sastifaction by Age and Gender ratio

Ảnh có chứa văn bản, Sơ đồ, hàng, biểu đồ

Mô tả được tạo tự động

Figure: Relationship Between Age, Gender, and Satisfaction

In a broader sense, the data suggests consistent satisfaction patterns between genders throughout various age groups, with only slight variations. Generally, younger age groups tend to show higher proportions of neutrality or dissatisfaction, whereas older age groups demonstrate elevated levels of satisfaction.

## Distribution of ‘Customer Type’ and its influence on satisfaction levels

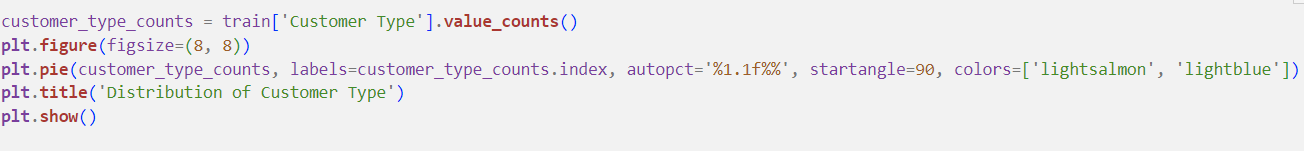


Figure: Distribution of Customer Type pie chart code

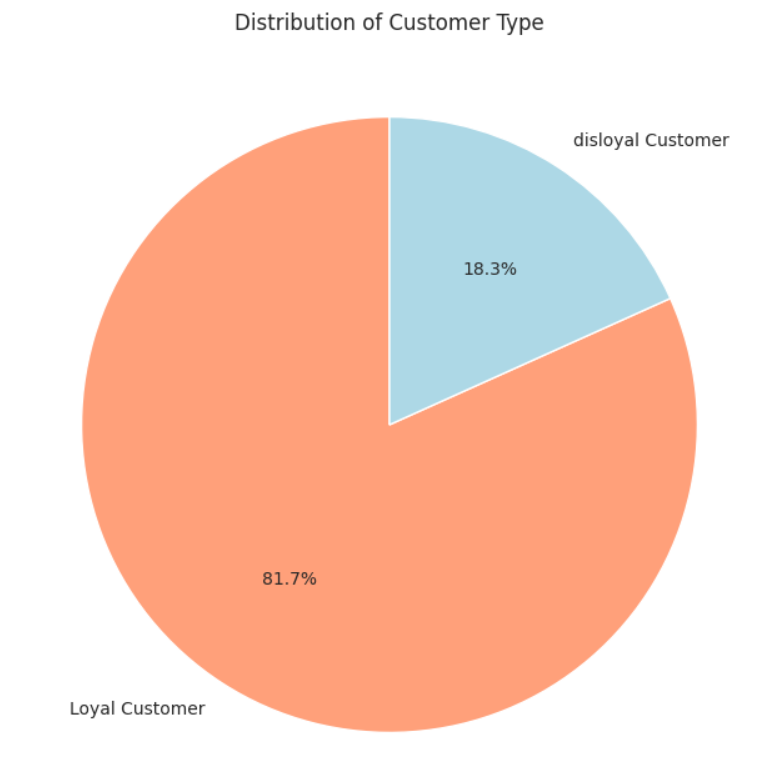


Figure: Distribution of Customer Type pie chart

Most customers (over 80%) belong to the loyal customer category, underscoring a robust rate of customer retention.

**Relationship between satisfaction, age and loyalty:**

Ảnh có chứa văn bản, Phông chữ, hàng, ảnh chụp màn hình

Mô tả được tạo tự động

Figure: Crosstab between satisfaction and customer type

Loyal customers typically exhibit a well-balanced distribution between neutral/dissatisfied and satisfied sentiments, while a significant majority of disloyal customers express either neutrality or dissatisfaction.

Ảnh có chứa văn bản, Phông chữ, hàng, ảnh chụp màn hình

Mô tả được tạo tự động

Figure: Average age of customer type

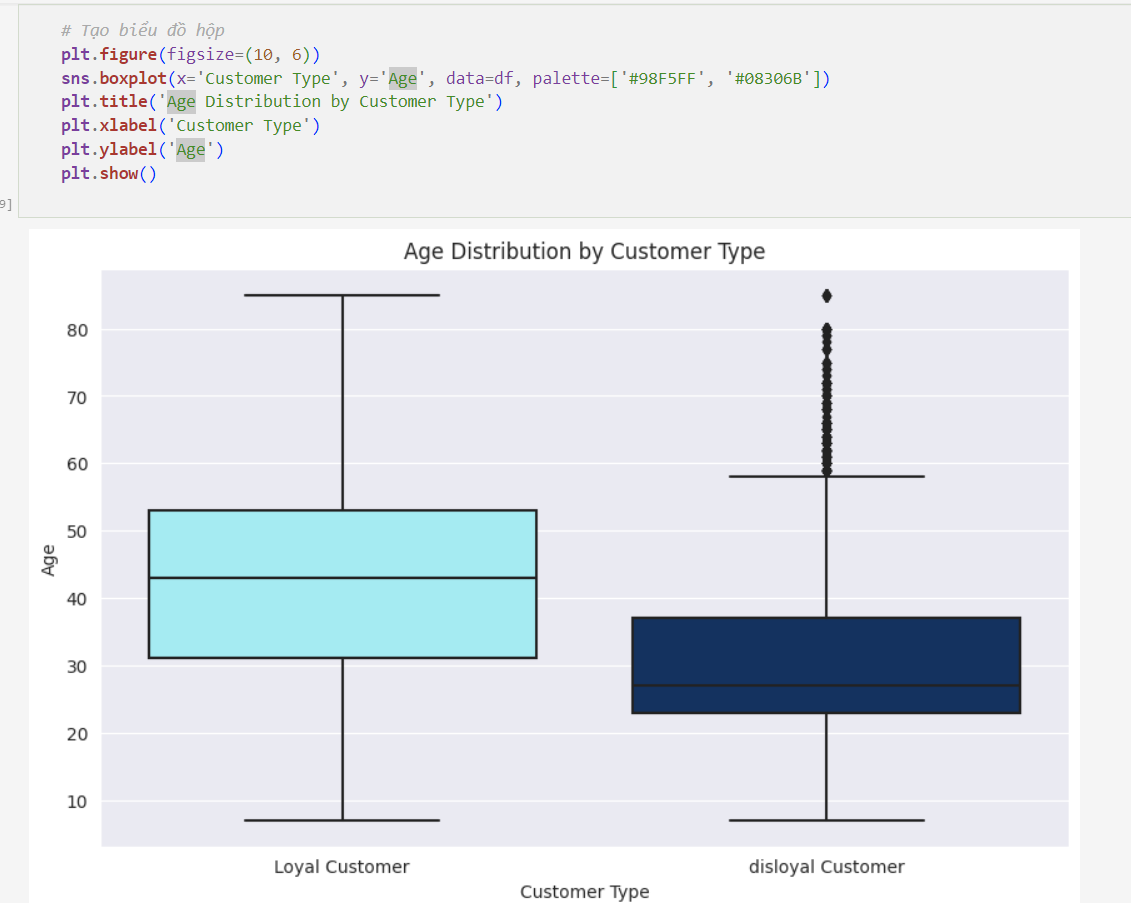


Figure: Age Distribution by Customer Type boxplot

On average, loyal customers are older in comparison to their disloyal counterparts. Age appears to be a potential contributing factor to customer loyalty, as older individuals demonstrate a higher likelihood of being classified as loyal customers.

## Distribution of ‘Type of Travel’ and ‘Travel Class’

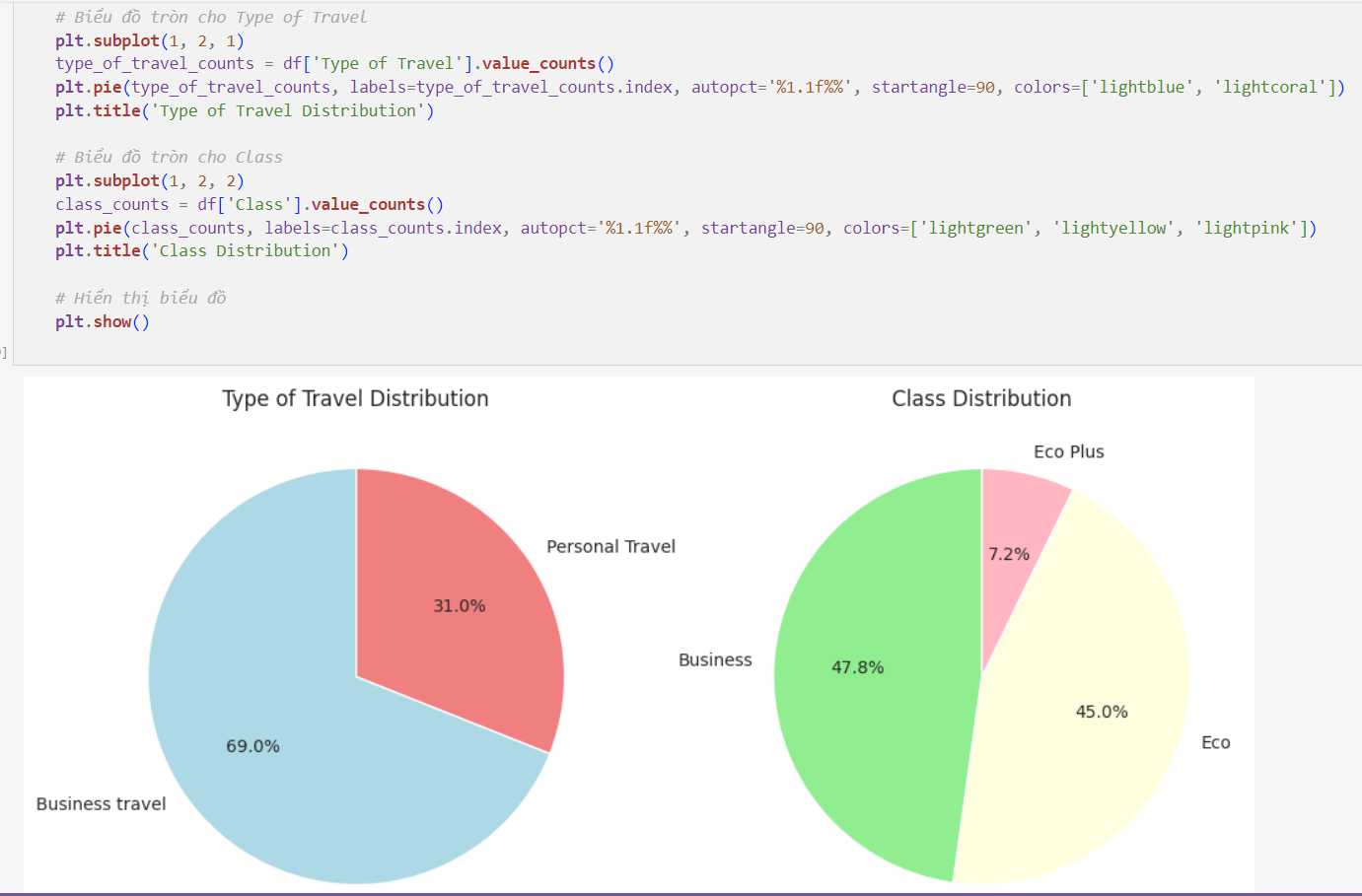


Figure: 3.4.6. Distribution of ‘Type of Travel’ and ‘Travel Class’ pie chart

In general, ‘Business Travel’ take a higher count compared to ‘Personal Travel’. Furthermore, the Economy class exhibits the highest ticket count, followed by Business class and Economy Plus class in descending order.

Ảnh có chứa văn bản, ảnh chụp màn hình, biểu đồ, số

Mô tả được tạo tự động

Figure: Satisfaction Percentage by Type of Travel chart

"Business Travelers show a significantly higher satisfaction rate compared to Personal Travelers, with the majority of Business Travelers expressing satisfaction, whereas the majority of Personal Travelers tend to express neutrality or dissatisfaction."

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, số

Mô tả được tạo tự động

Figure: Satisfaction Percentage by Class chart

The data reveals that 'Business Class' has the highest proportion of satisfied customers, accounting for 69.43% of total satisfaction across all classes. Conversely, 'Eco Plus Class' exhibits the highest percentage of dissatisfied or neutral customers, comprising 75.39% of such sentiments across all classes.

## Distribution of Flight Distance

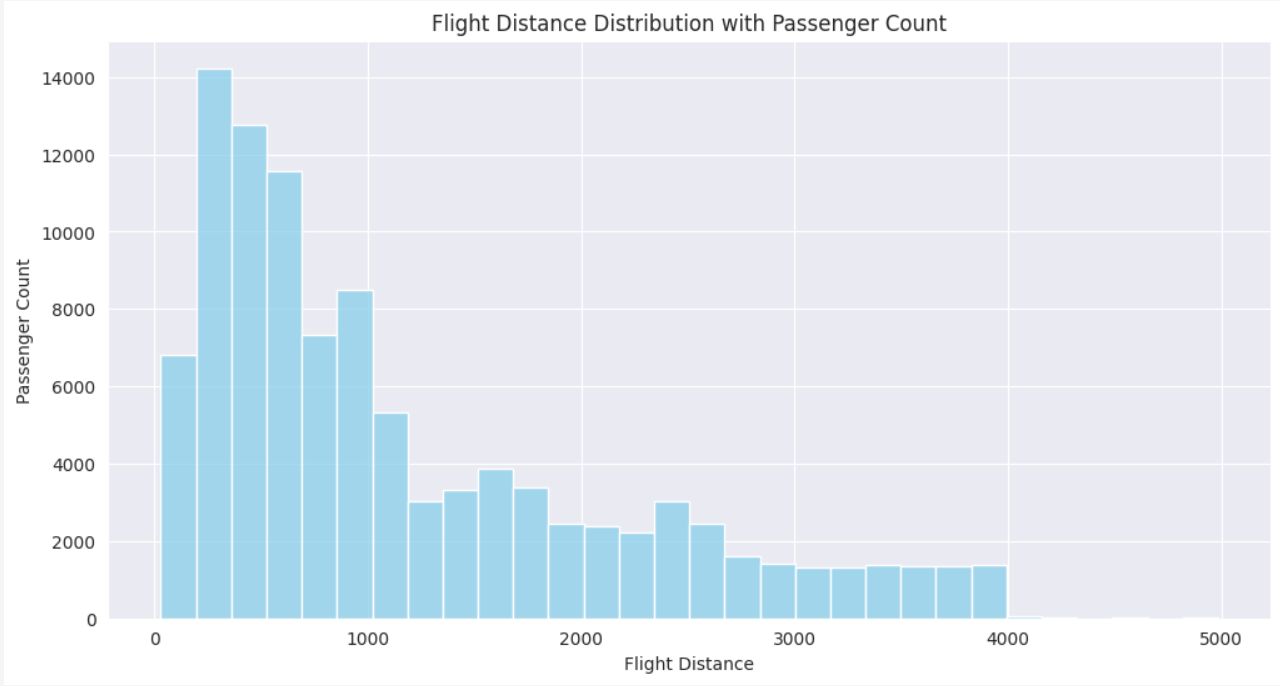


Figure: Flight Distance Distribution with Passenger Count

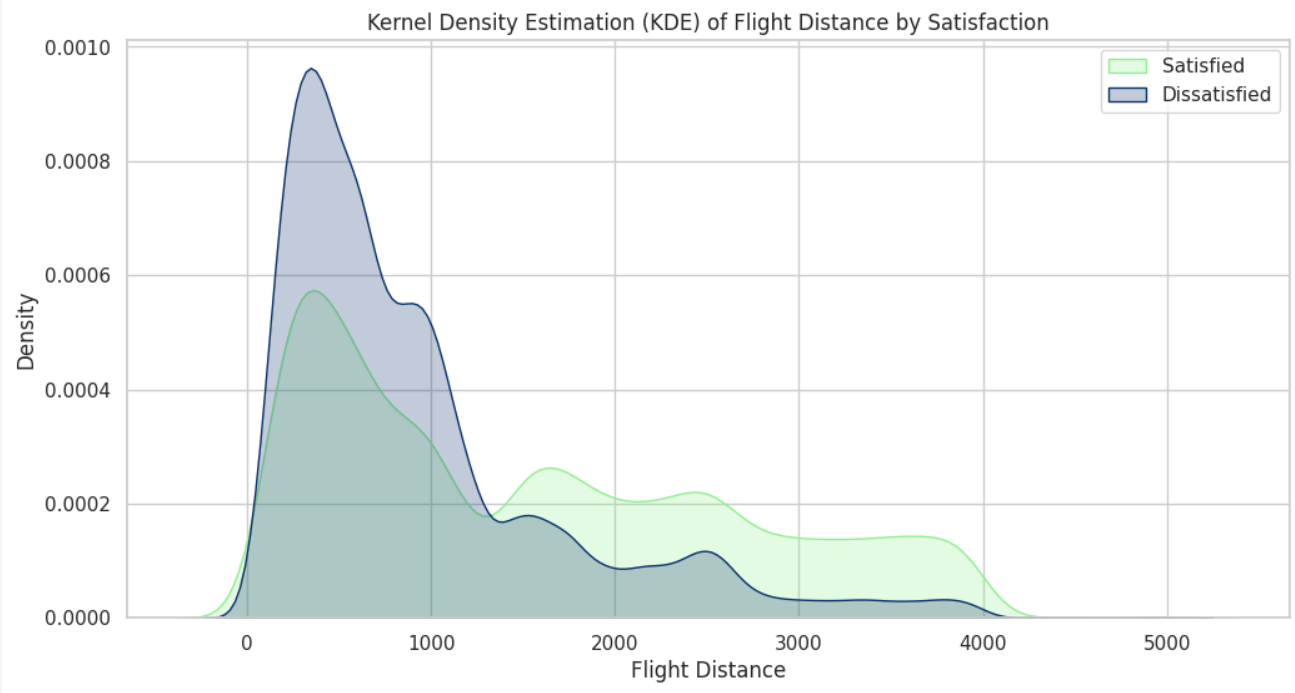


Figure: Kernel Density Estimation (KDE) of Flight Distance by Satisfaction

There appears to be a positive correlation between travel distance and satisfaction, suggesting that longer journeys are more likely to result in higher satisfaction. This correlation may be attributed to the observation that services on longer flights tend to be of higher quality compared to shorter ones.

**Average flight distance for each class of travel:**

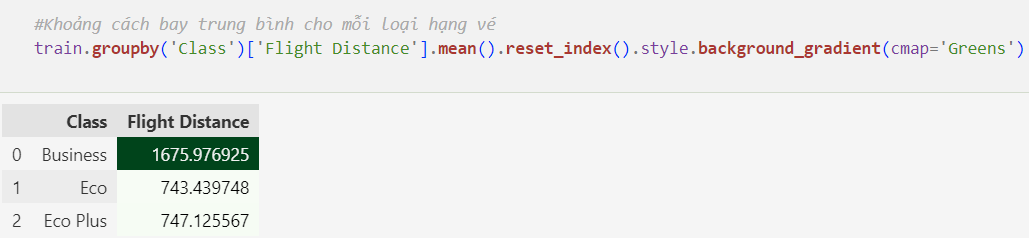
****

Figure: Average flight distance for each class of travel

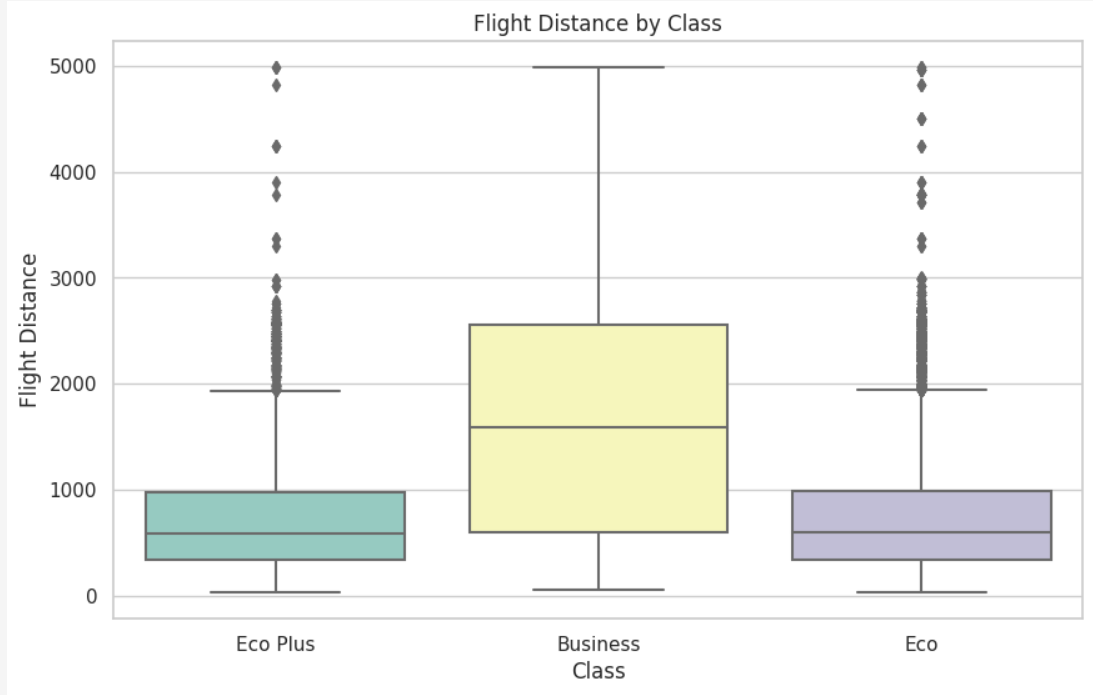


Figure: Flight Distance by Class boxplot

Business Class travelers, on average, cover longer flight distances compared to passengers in Economy Class and Economy Plus Class. In contrast, the flight distances for Economy Class and Economy Plus Class are notably shorter and exhibit a similar range.



Figure: Define bins and labels for flight distance groups

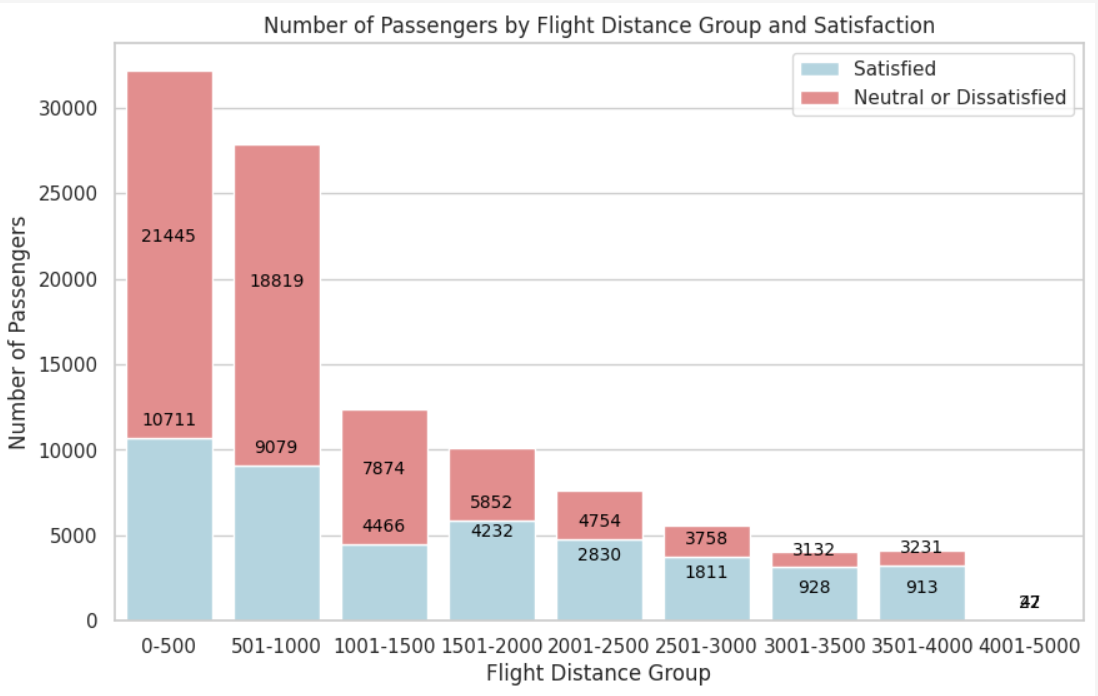


Figure: Number of Passengers by Flight Distance Group and Satisfaction

The majority of flights are concentrated within shorter distance ranges (0-500 miles and 501-1000 miles), with a gradual decrease in count as the flight distance range expands. The data offers a comprehensive overview of flight distribution across different distance groups.

**Correlation between travel class and satisfaction levels**

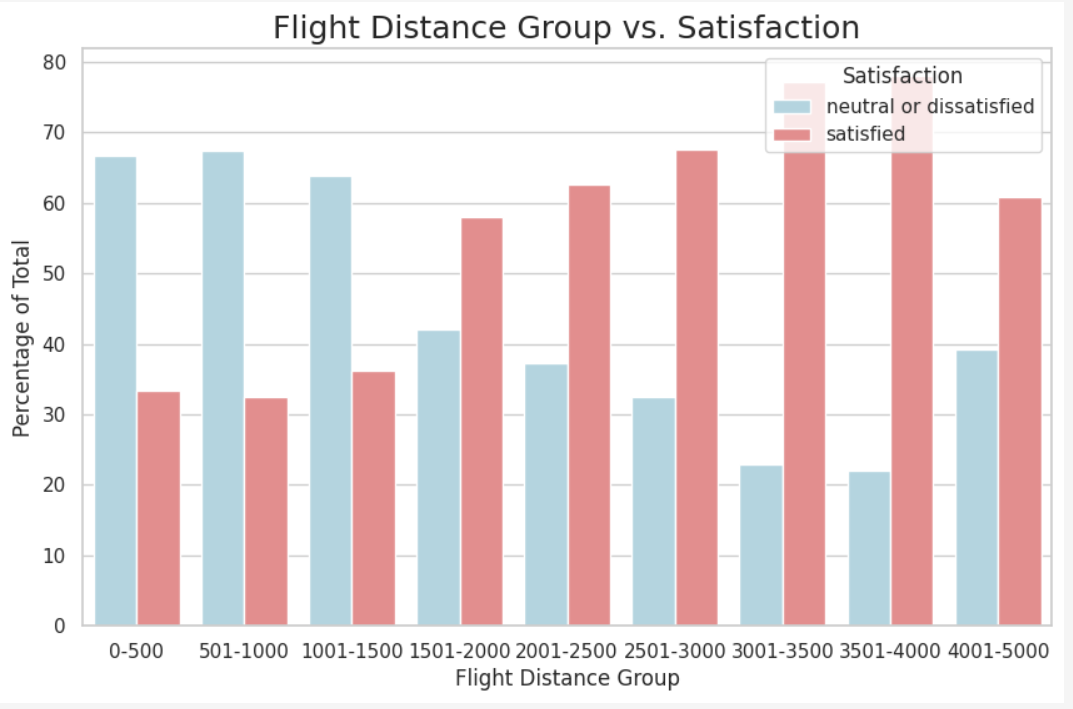


Figure: Flight Distance Group vs. Satisfaction

Generally, shorter-distance flights tend to have lower satisfaction levels, whereas long-distance flights exhibit a mix of satisfaction and neutral/dissatisfied sentiments. Some distance groups show a majority of passengers expressing satisfaction.

## Service quality analysis

Let see what service factors have the highest and lowest satisfaction levels:

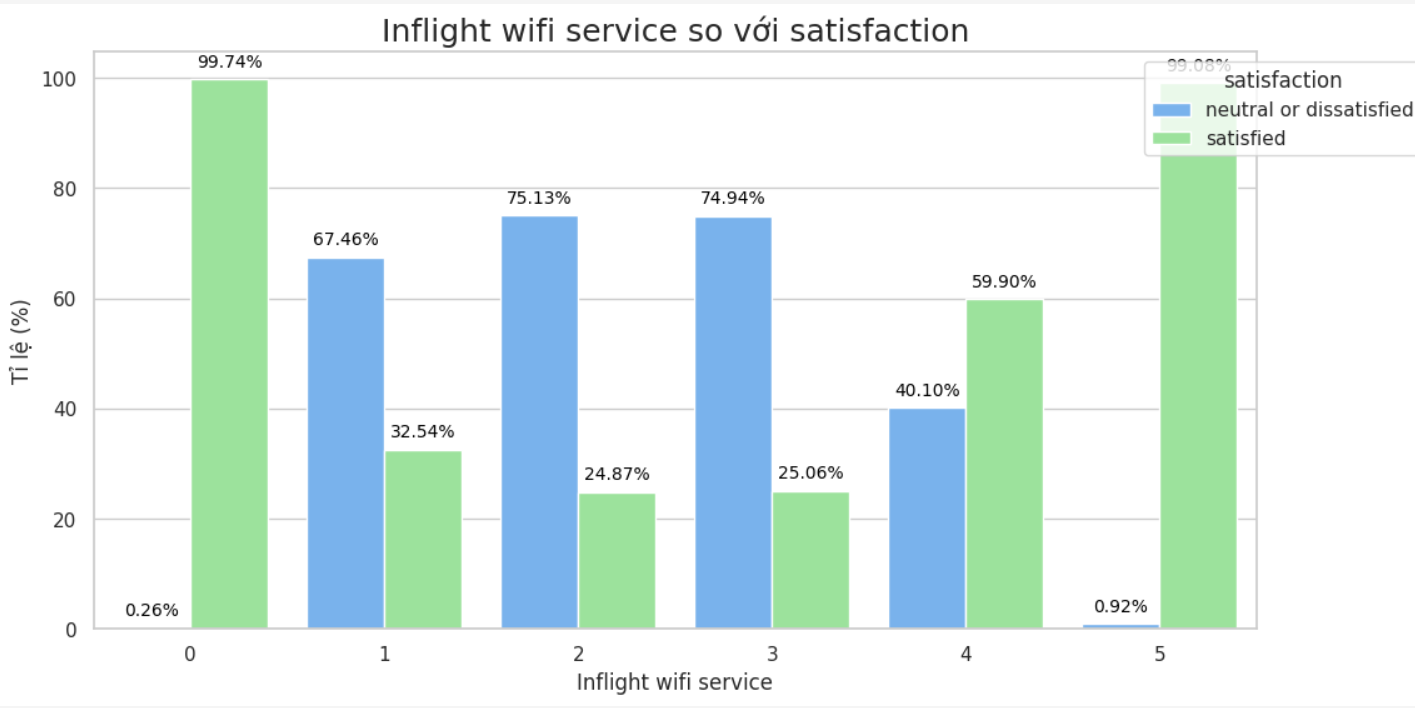


Figure: Inflight wifi service vs satisfaction

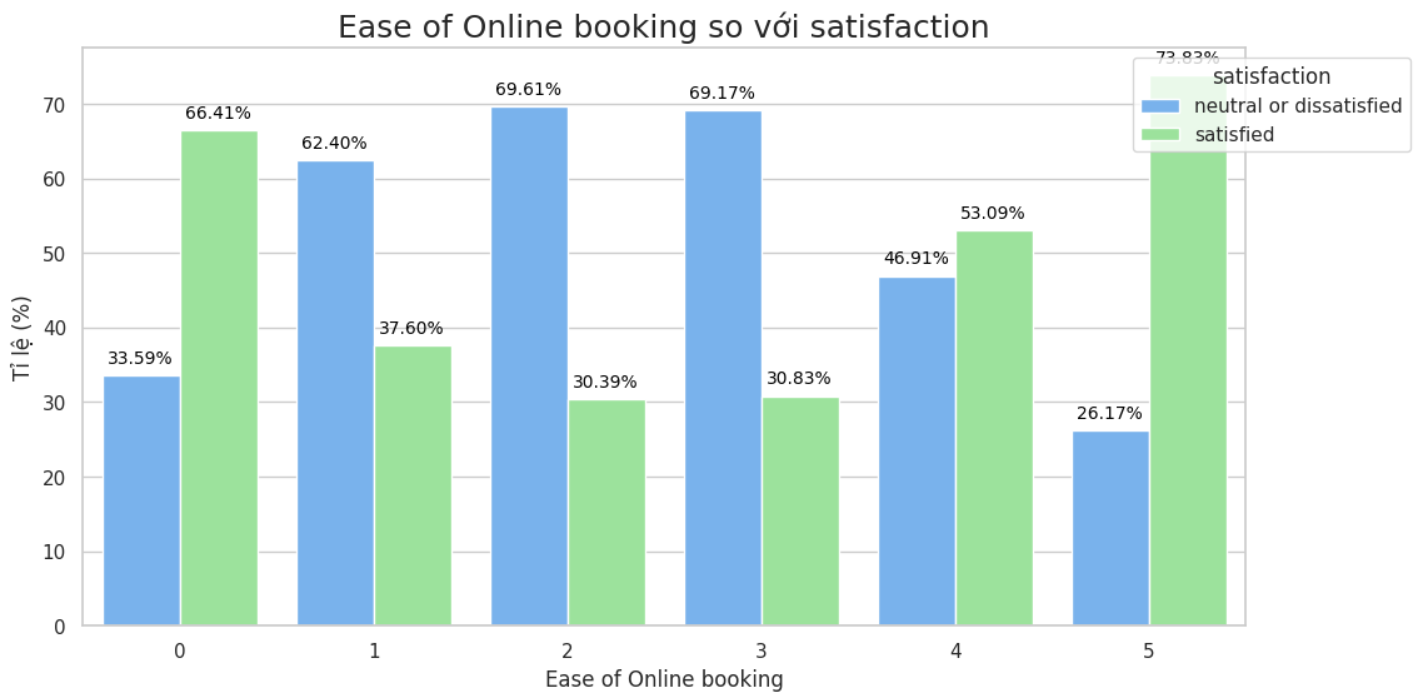


Figure: Ease of Online booking vs satisfaction

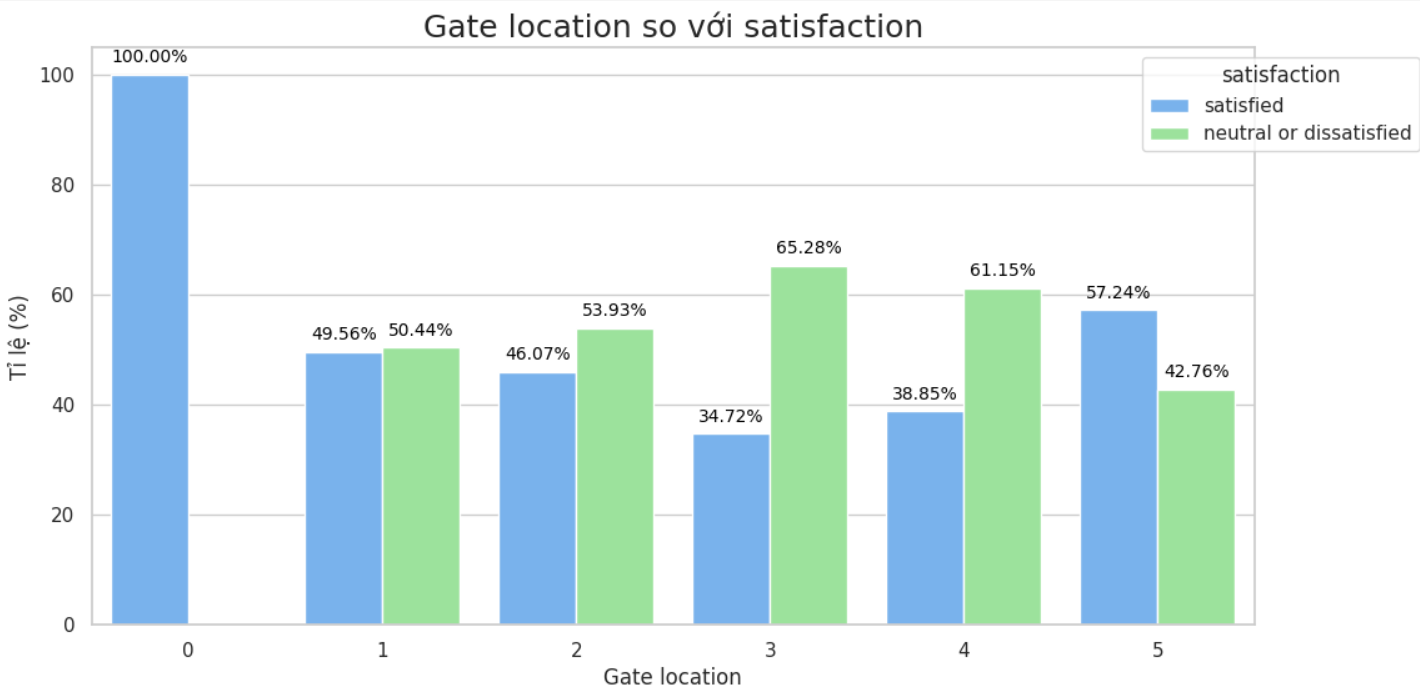


Figure: Gate location vs satisfaction

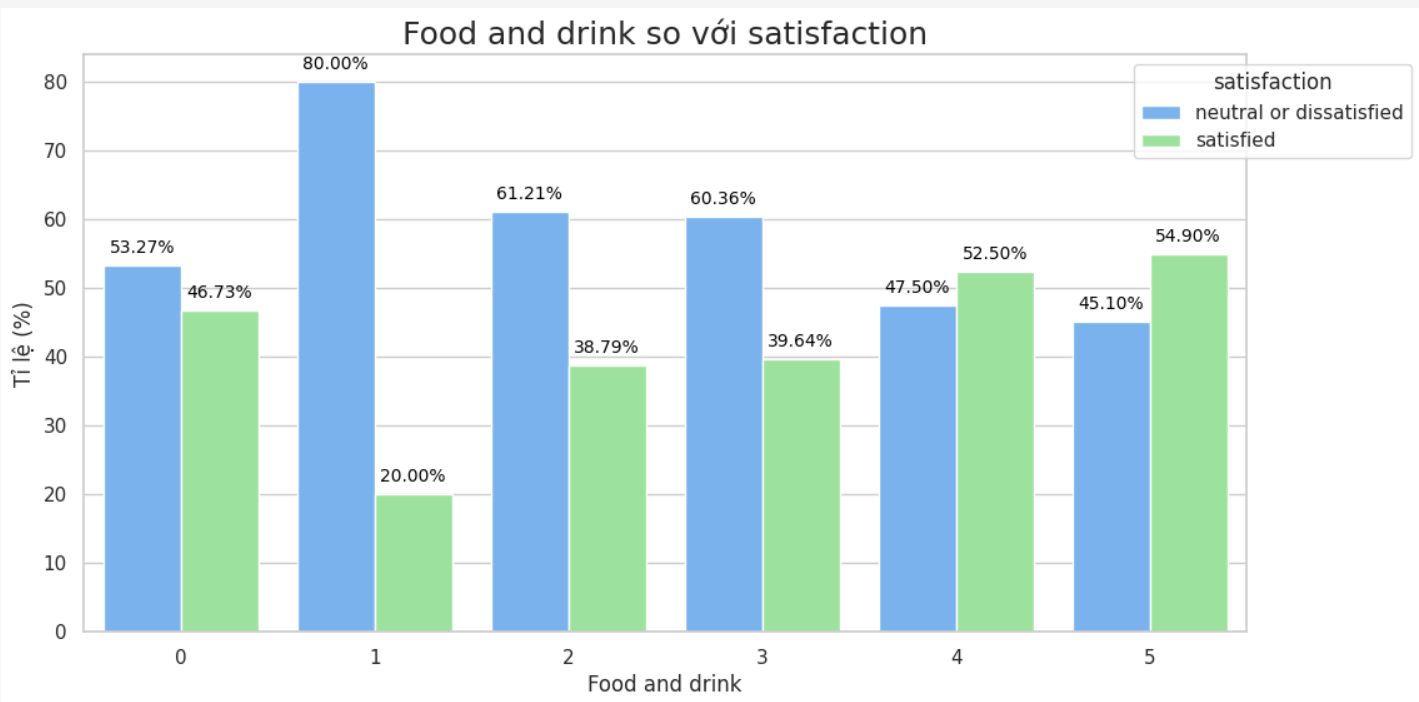


Figure: Food and drink vs satisfaction

Ảnh có chứa văn bản, ảnh chụp màn hình, biểu đồ, số

Mô tả được tạo tự động

Figure: Online boarding vs satisfaction

Ảnh có chứa văn bản, ảnh chụp màn hình, biểu đồ, số

Mô tả được tạo tự động

Figure: Seat comfort vs satisfaction

Ảnh có chứa văn bản, ảnh chụp màn hình, biểu đồ, số

Mô tả được tạo tự động

Figure: Inflight entertainment vs satisfaction

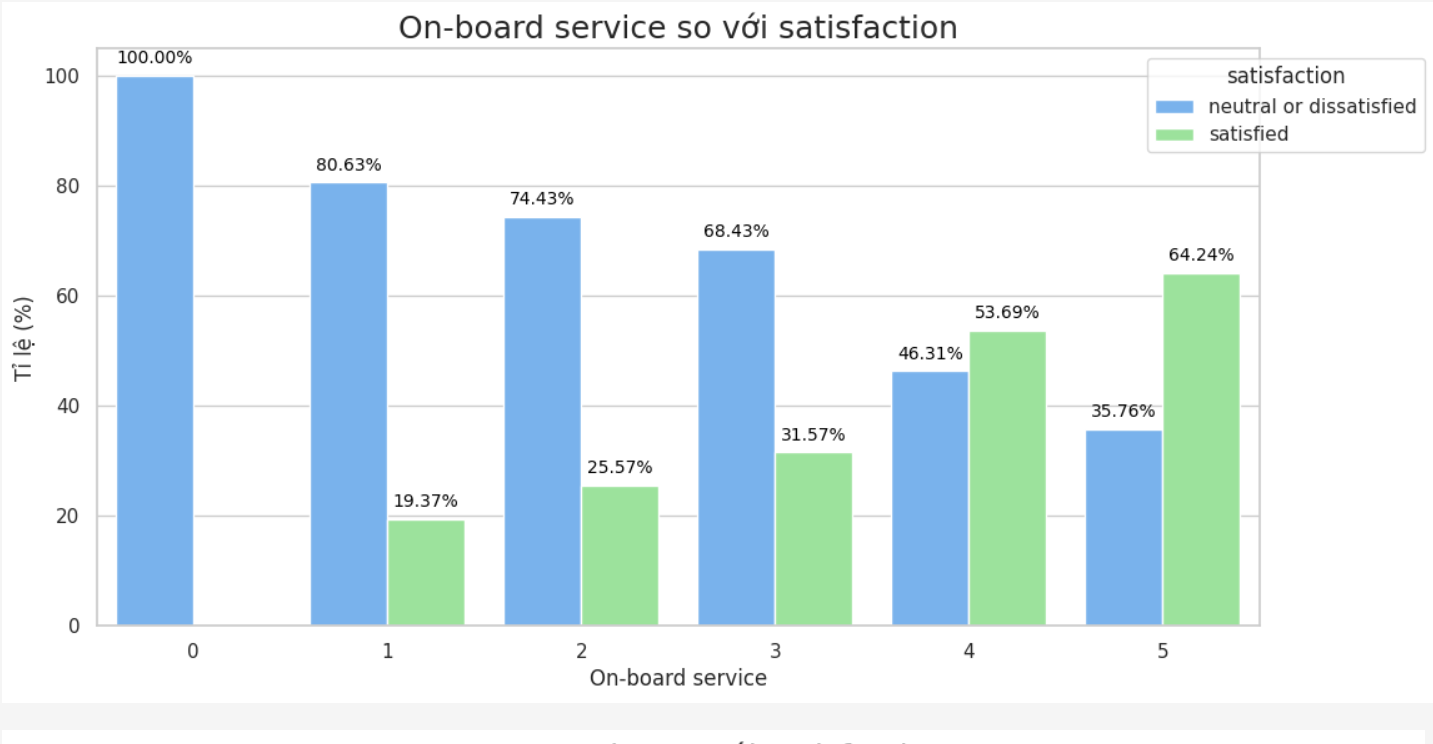


Figure: On-board service vs satisfaction

Ảnh có chứa văn bản, ảnh chụp màn hình, biểu đồ, số

Mô tả được tạo tự động

Figure: Leg room service vs satisfaction

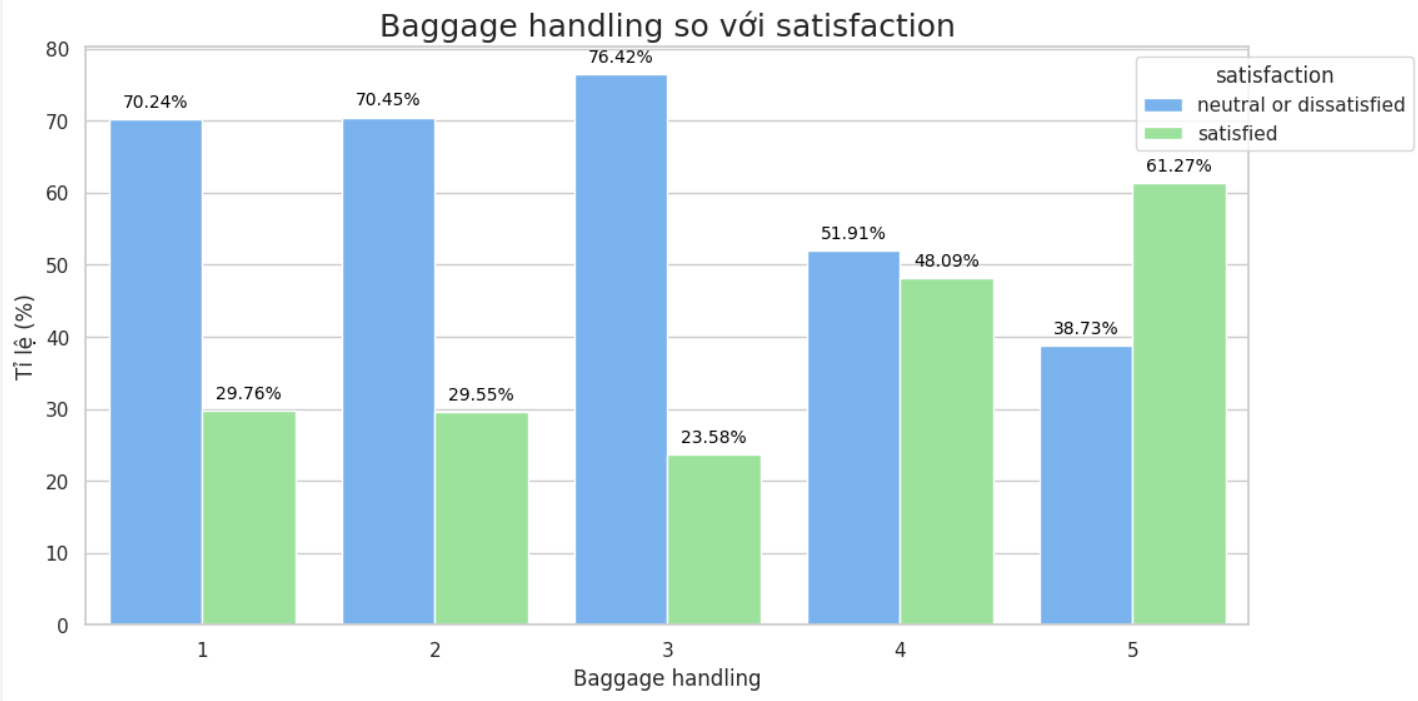


Figure: Baggage handling vs satisfactionẢnh có chứa văn bản, ảnh chụp màn hình, biểu đồ, số

Mô tả được tạo tự động

Figure: Checkin service vs satisfaction

Ảnh có chứa văn bản, ảnh chụp màn hình, biểu đồ, số

Mô tả được tạo tự động

Figure: Inflight service vs satisfaction

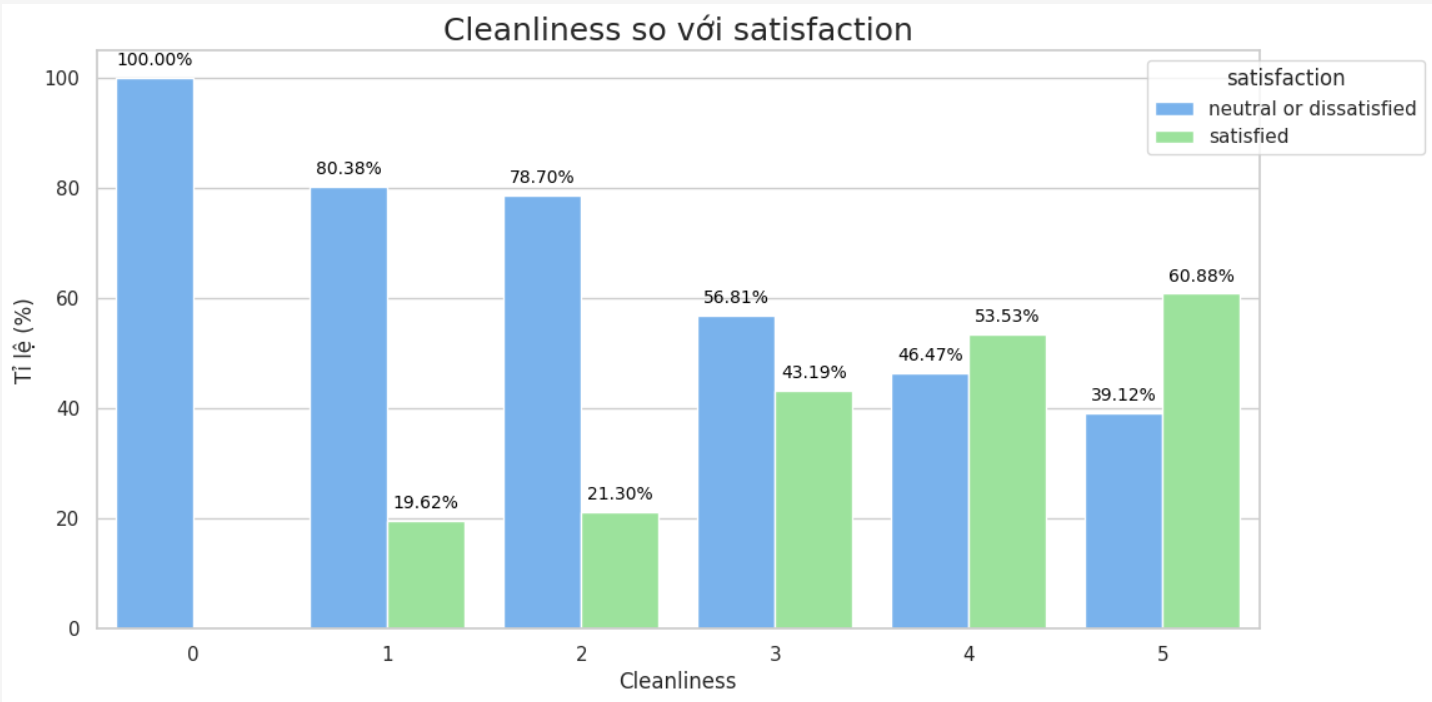


Figure: Cleanliness vs satisfaction

## Correlation between delays (departure and arrival) and satisfaction levels

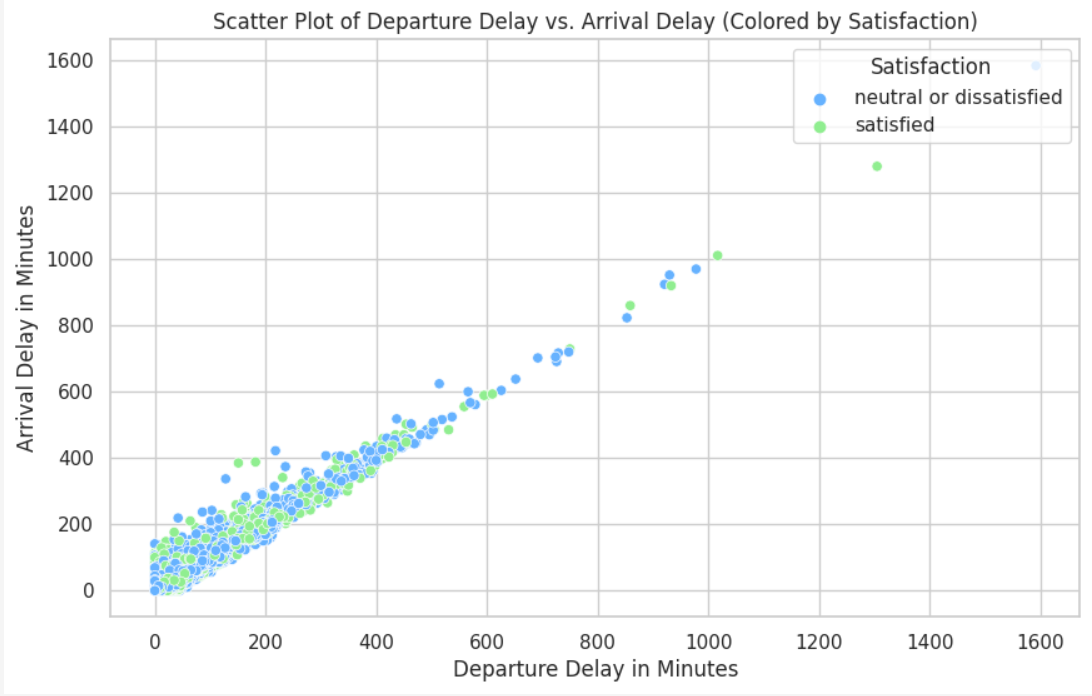


Figure: Scatter Plot of Departure Delay vs. Arrival Delay

The data emphasizes a notable pattern: individuals expressing neutral or dissatisfied sentiments frequently encounter more significant delays in their journeys compared to satisfied passengers. This underscores the pivotal role that punctuality and minimal delays play in shaping overall customer satisfaction.