Airline Passenger Satisfaction Analysis

Supervised By

Eng.Tarek Ziad

Presented by: Abdelmoneim Moustafa

AGENDA

Problem Definition → **Data Description** → **EDA (Exploratory Data Analysis) and Business Solution**

Data Pre-Processing → **Data Modeling** → **Conclusion**
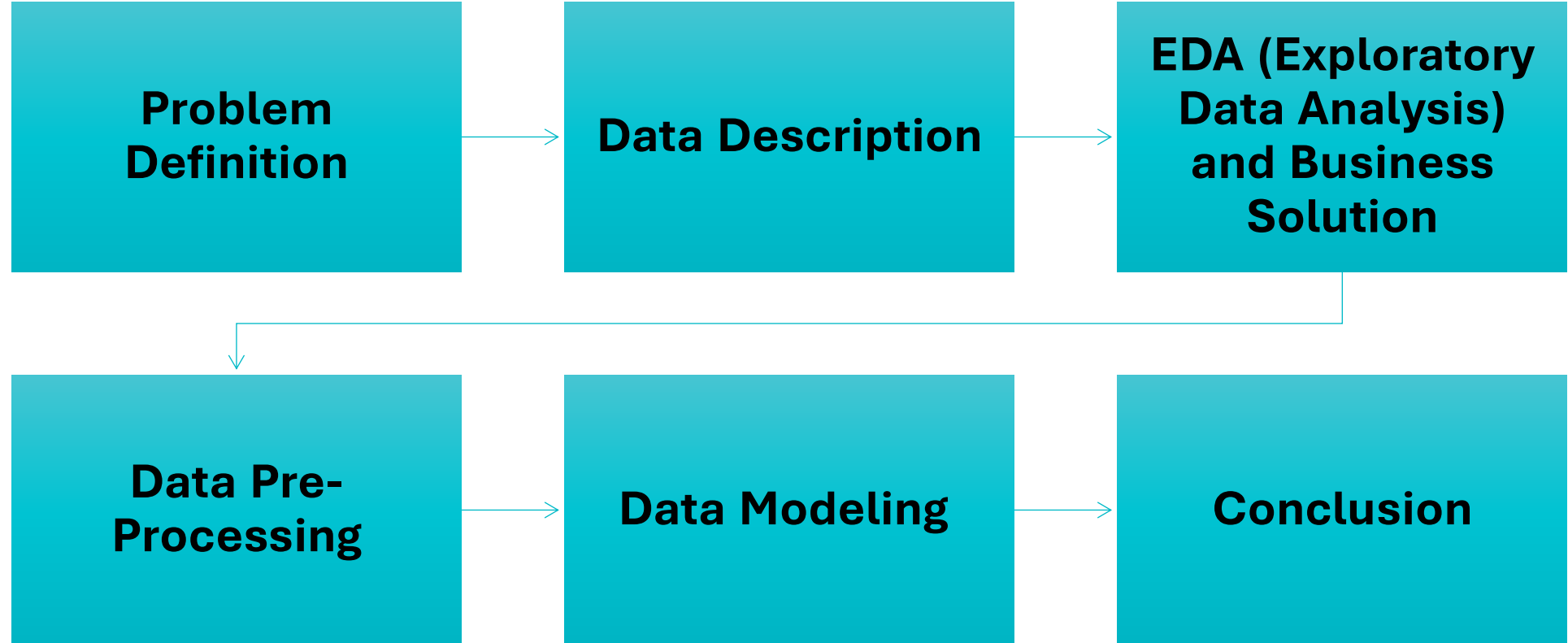
Problem Definition

Airlines operate in a competitive market where customer satisfaction is crucial.

Identifying key factors influencing satisfaction is essential for service improvement.

A data-driven approach can replace slow traditional feedback methods.

The goal is to analyze trends and predict passenger satisfaction to enhance airline services.

Problem Definition

The purpose of data is to uncover insights, identify patterns, and drive informed decision-making for improved passenger satisfaction.

Data Description

Overview of data

Dataset Summary:
- **Source:** Kaggle – *Airline Passenger Satisfaction Dataset*
- **Total Records:** 129,880
- **Total Features:** 24
- **Target Variable:** Passenger Satisfaction (Satisfied / Unsatisfied)

Dataset Statistics

Number of Variables	24
Number of Rows	129880
Missing Cells	393
Missing Cells (%)	0.0%
Duplicate Rows	0
Duplicate Rows (%)	0.0%
Total Size in Memory	59.9 MB
Average Row Size in Memory	484.0 B
Variable Types	Numerical: 5 Categorical: 19

Dataset Insights

`ID` is uniformly distributed	Uniform
`Departure Delay` and `Arrival Delay` have similar distributions	Similar Distribution
`Departure Delay` is skewed	Skewed
`Arrival Delay` is skewed	Skewed
`Type of Travel` has constant length 8	Constant Length
`Departure and Arrival Time Conv…` has constant length 1	Constant Length
`Ease of Online Booking` has constant length 1	Constant Length
`Check-in Service` has constant length 1	Constant Length
`Online Boarding` has constant length 1	Constant Length
`Gate Location` has constant length 1	Constant Length

1 2 3

Data Description

## Overview of data

Key Features:
- **Passenger Information:** Age, Travel Type, Class
- **Flight Experience:** Seat Comfort, Food & Drink, In-flight Entertainment
- **Service Ratings:** Check-in Service, Online Booking, Customer Support
- **Flight Details:** Departure Delay, Arrival Delay, Flight Distance

Data Insights:
- **Dataset Size:** 129,880 rows and 24 variables.
- **Missing Values:** 393 missing cells.
- **Duplicate Rows:** 0 duplicate rows.
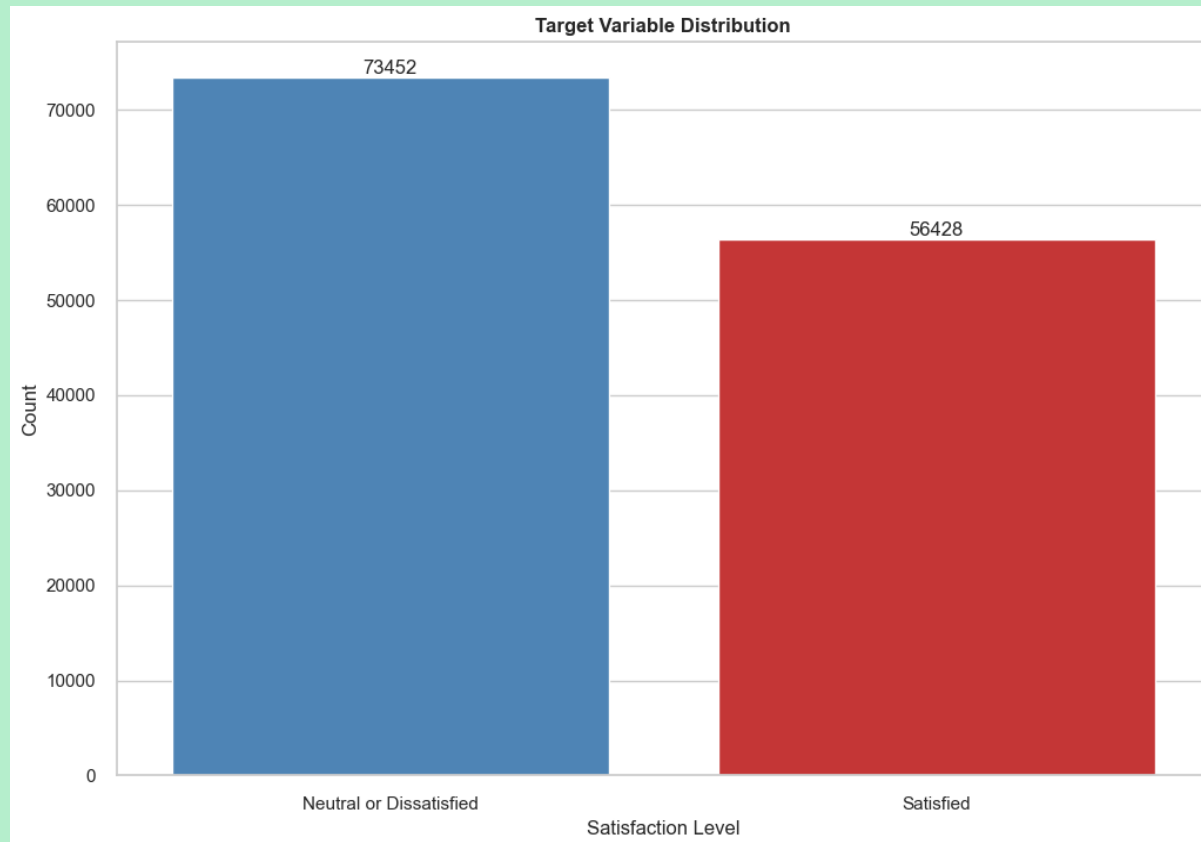- **Outliers:** Present, and will be handled later.

EDA

How does customer type relate to satisfaction and age?
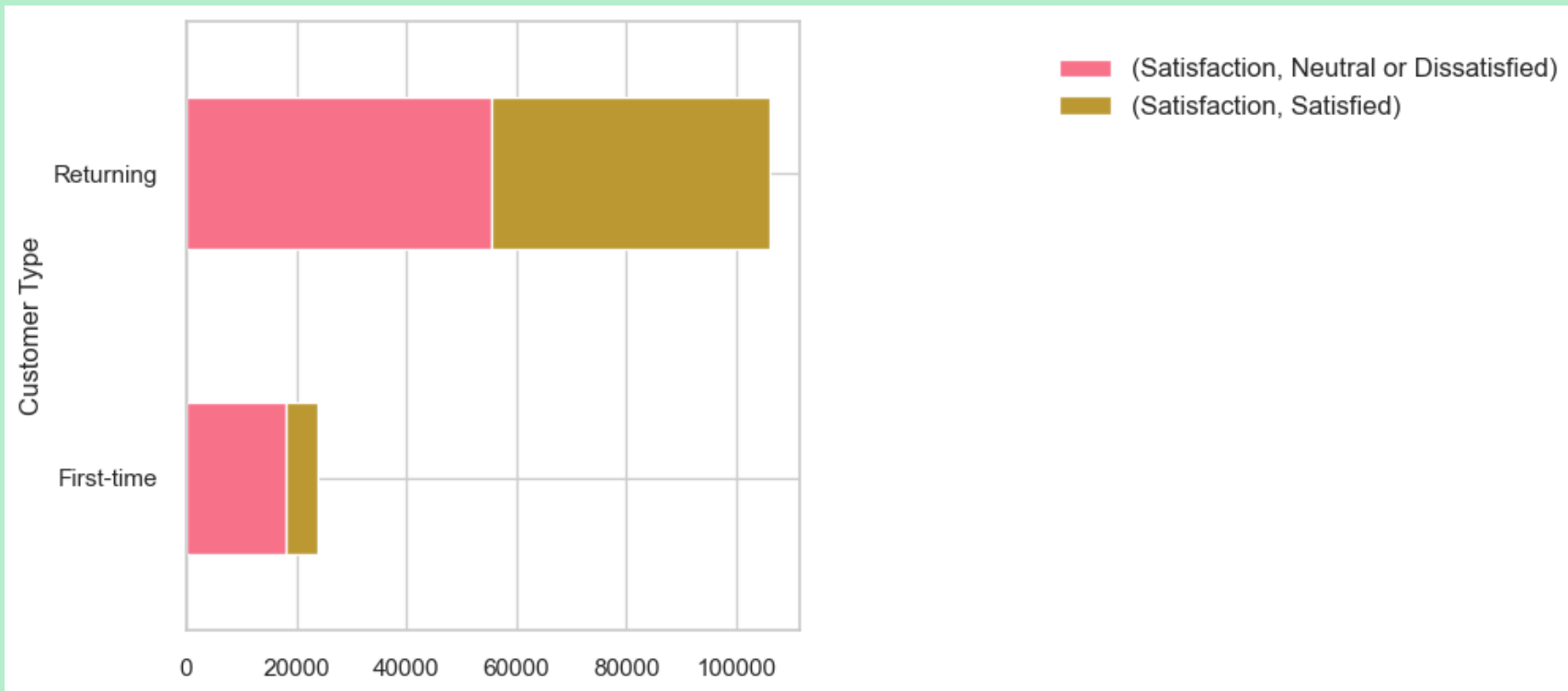
- **Most customers are Returning Passengers, so they have experienced the services before.**
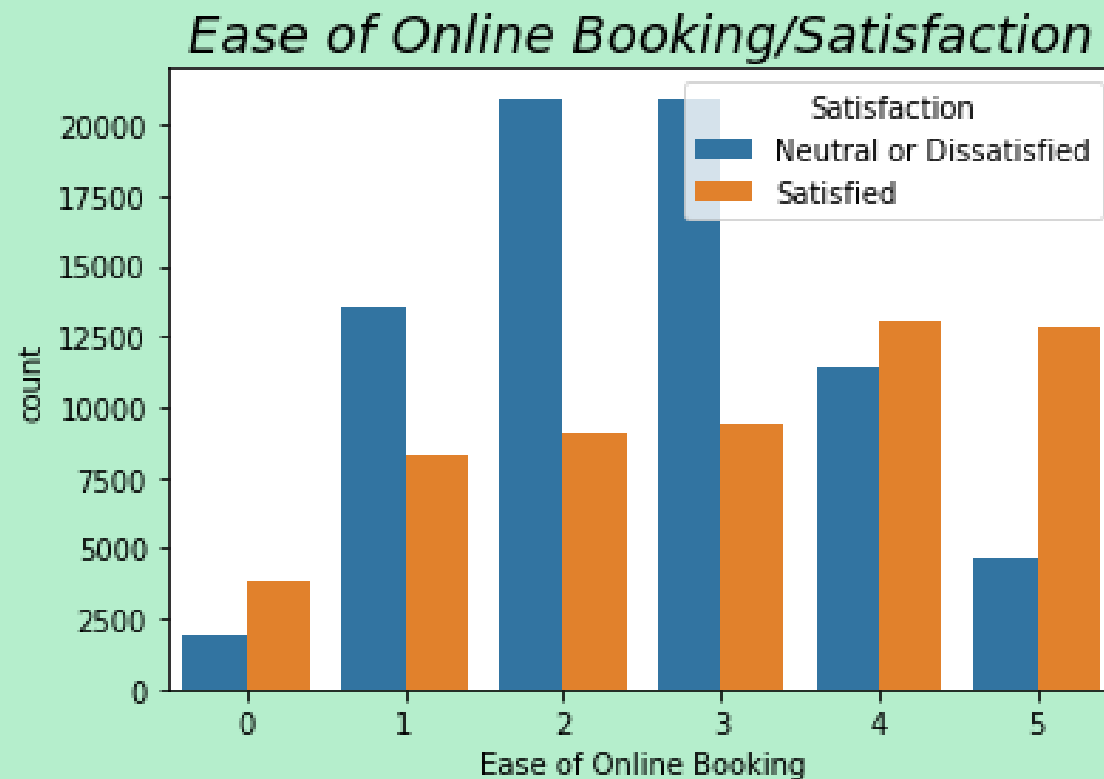
EDA

How does customer type relate to satisfaction and age?

- **Over 60% of travelers are either dissatisfied or neutral.**

EDA

What is the relationship between satisfaction and ease of online booking?

Ease of Online Booking/Satisfaction

- It was found that the most satisfied people are happy with the online reservation and the largest percentage of the people are satisfied with the reservation average
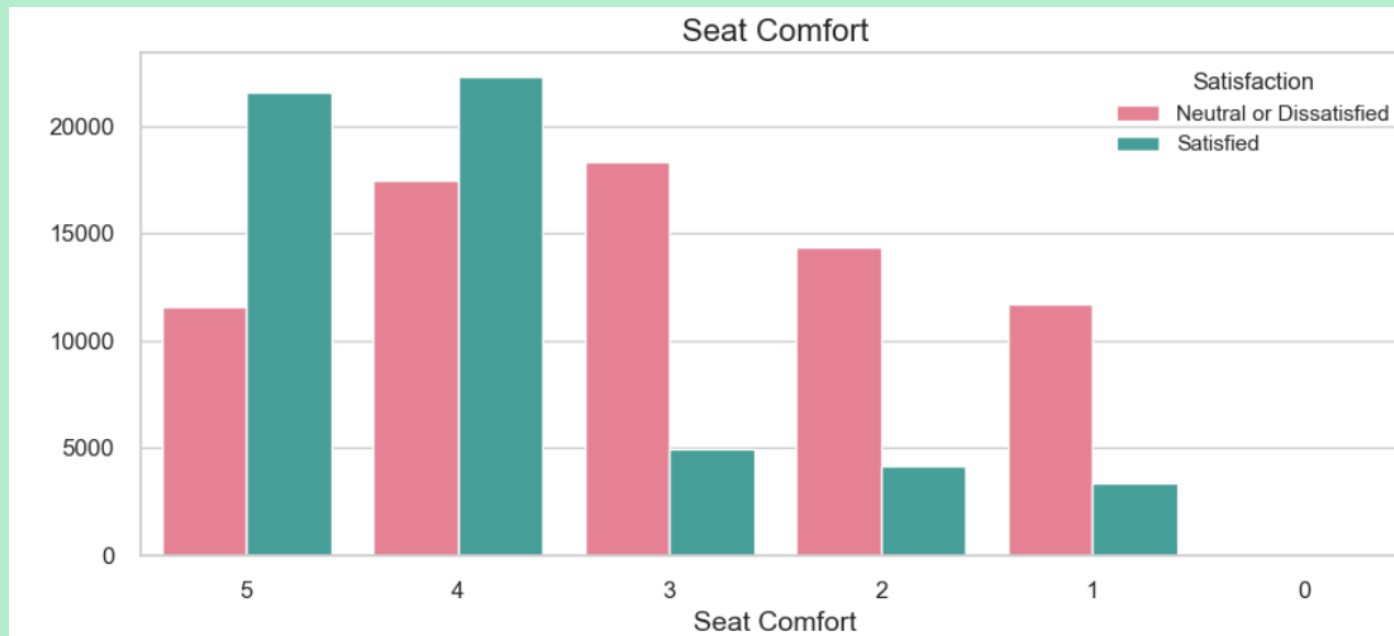
EDA

Satisfaction relationship with on-board service?
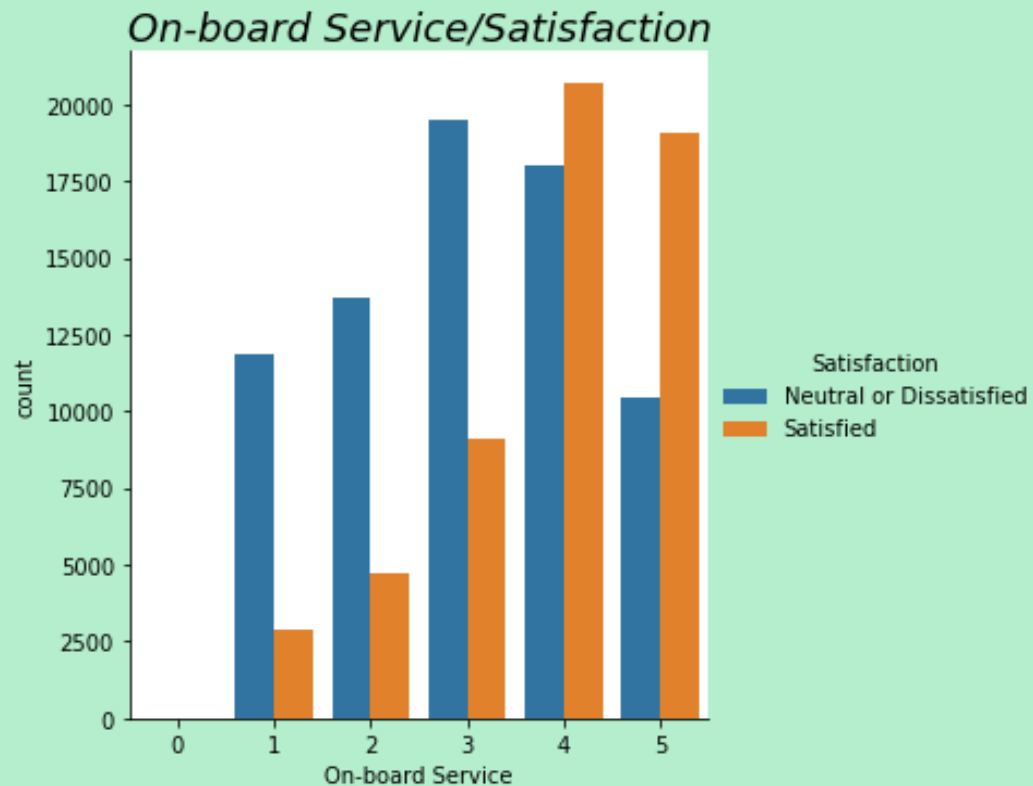
- A significant percentage of people are dissatisfied with the in-flight service, while the highest percentage of satisfied customers express happiness with the service.
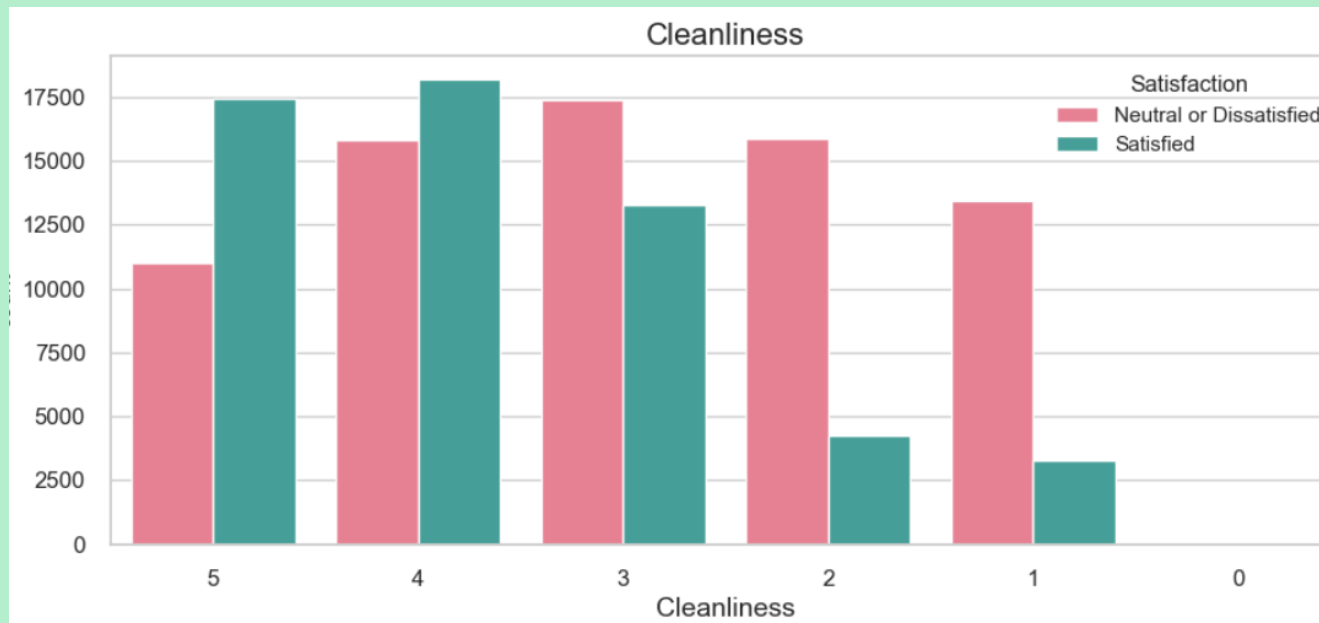
EDA

Satisfaction relationship with seat comfort?

- **Like the rest of the previous comparisons, a good number of dissatisfied people are happy with the seats, so the problem is not here**

EDA

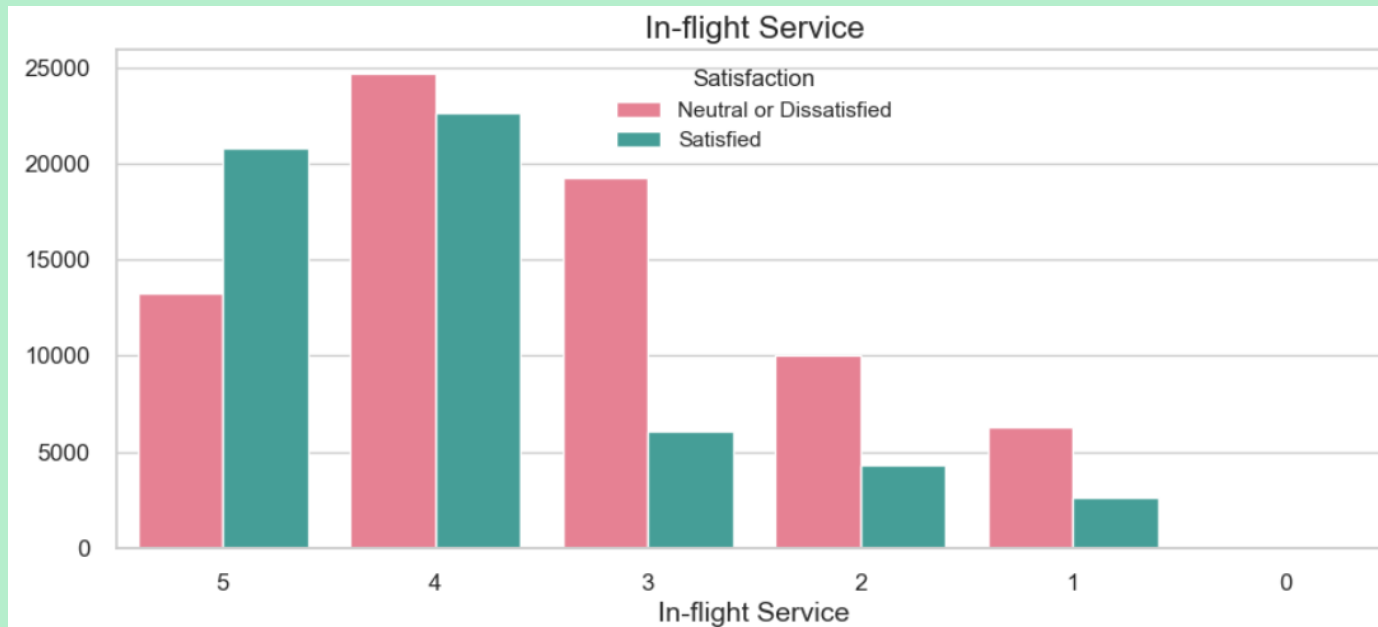## The relationship of the satisfaction with cleanliness?

- %22of dissatisfied individuals are unhappy with hygiene, whereas 27% of satisfied individuals are very pleased with cleanliness.
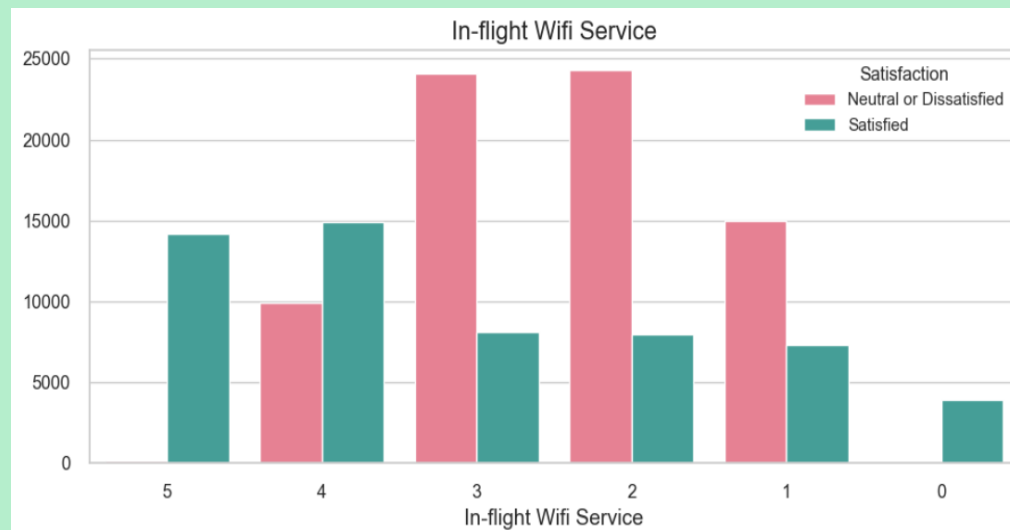
EDA

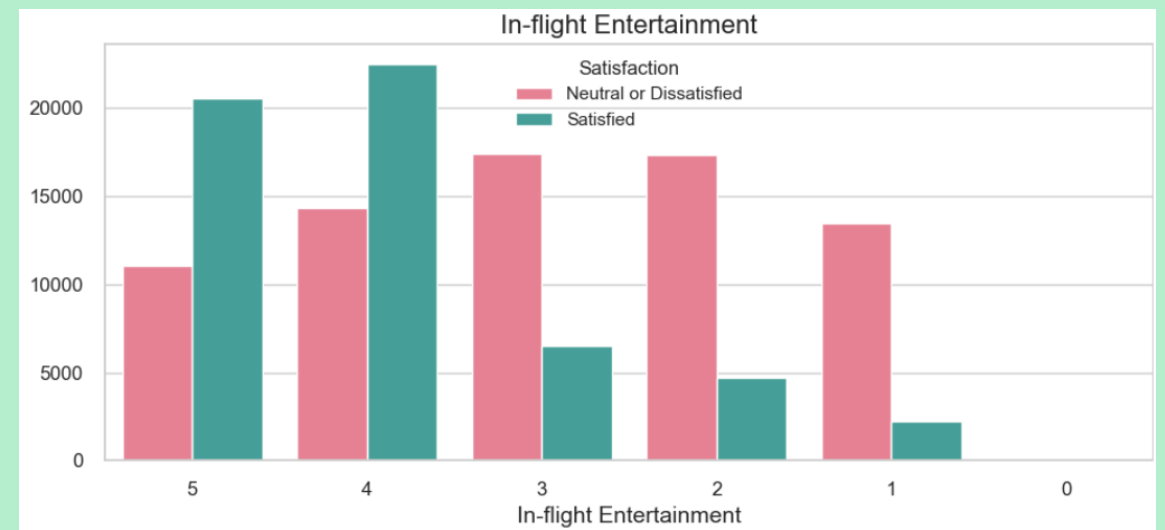**The relationship of the satisfaction with in-flight service?**

- The majority of the reviews are between ,5-3which is very positive. There are very few low ratings so there is not much to analyze there. The overall satisfaction is mostly neutral or dissatisfied at the low ratings but it gets better on ratings 4 and .5

**The relationship of the satisfaction with in-flight wifi service?
And with in-flight entertainment?**

Most ratings fall between 2 and egareva na gnitacidni ,4

iF-iW roop htiw sregnessap taht swohs ataD .erocs llarevo

(3-1)tend to be neutral or dissatisfied, which is

understandable given Wi-Fi's importance today.

Most reviews are positive, mainly between 4 and .5

tnemniatretne thgilf-ni taht swohs tolp ehT

.noitcafsitas regnessap llarevo stceffa yltnacifingis

14

EDA

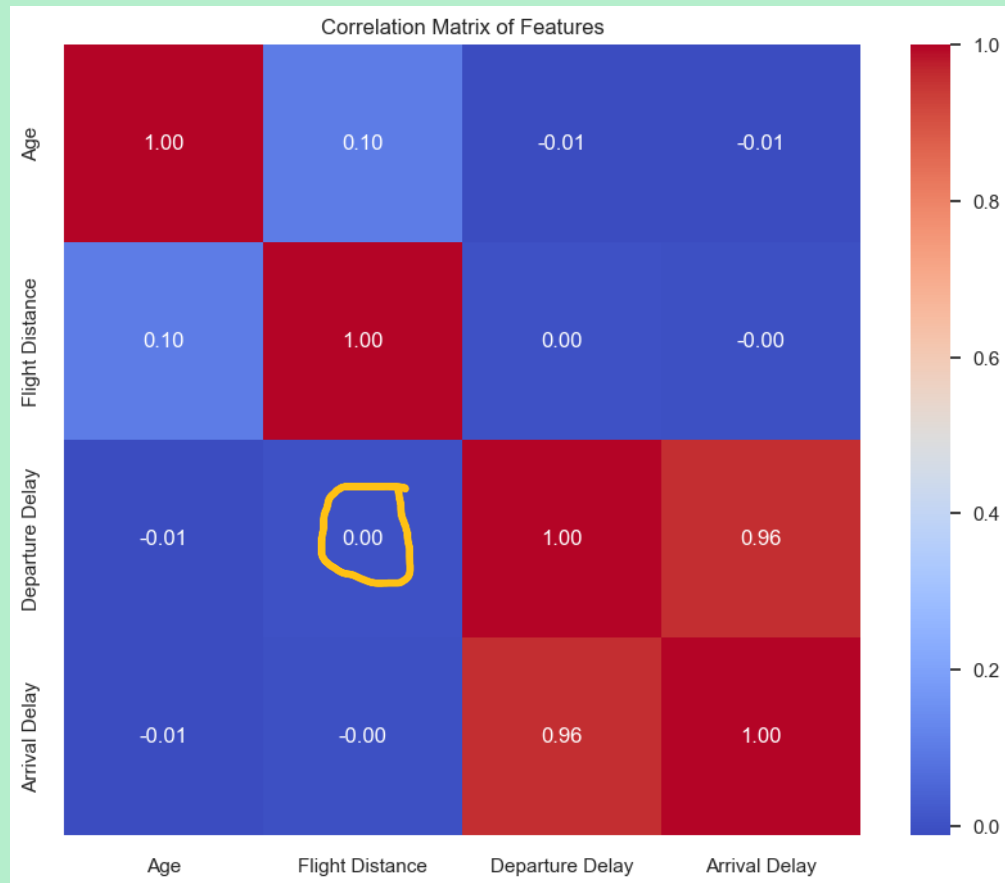**What is the relationship between the gender of the customer and the class Travel?**
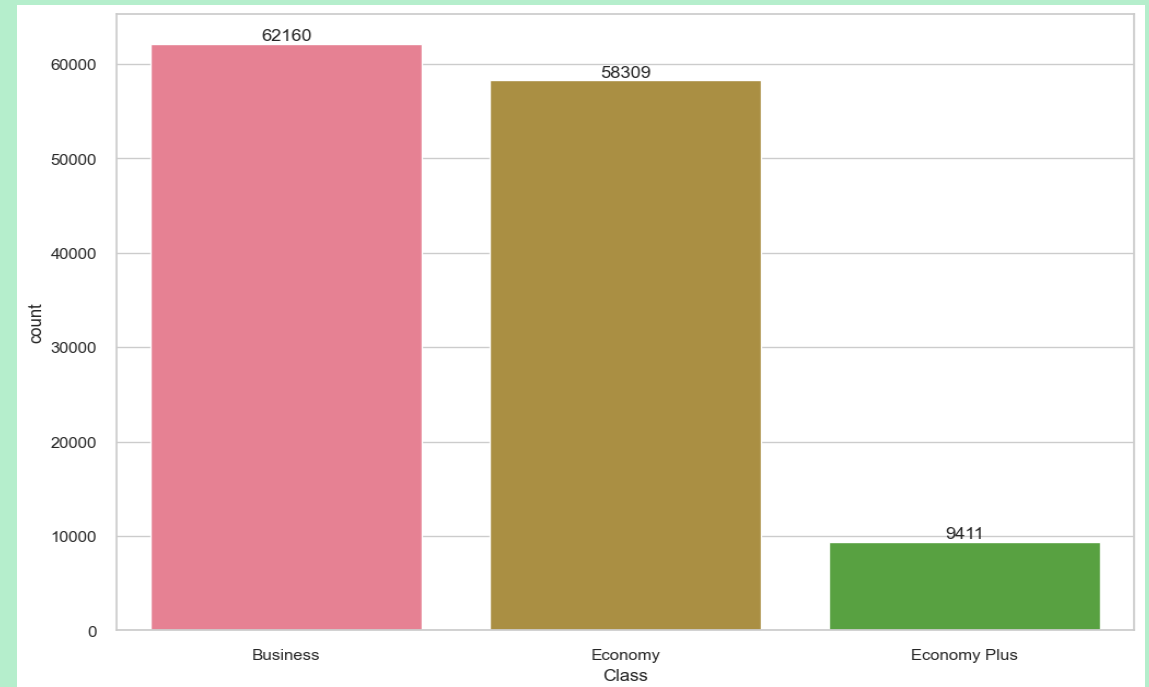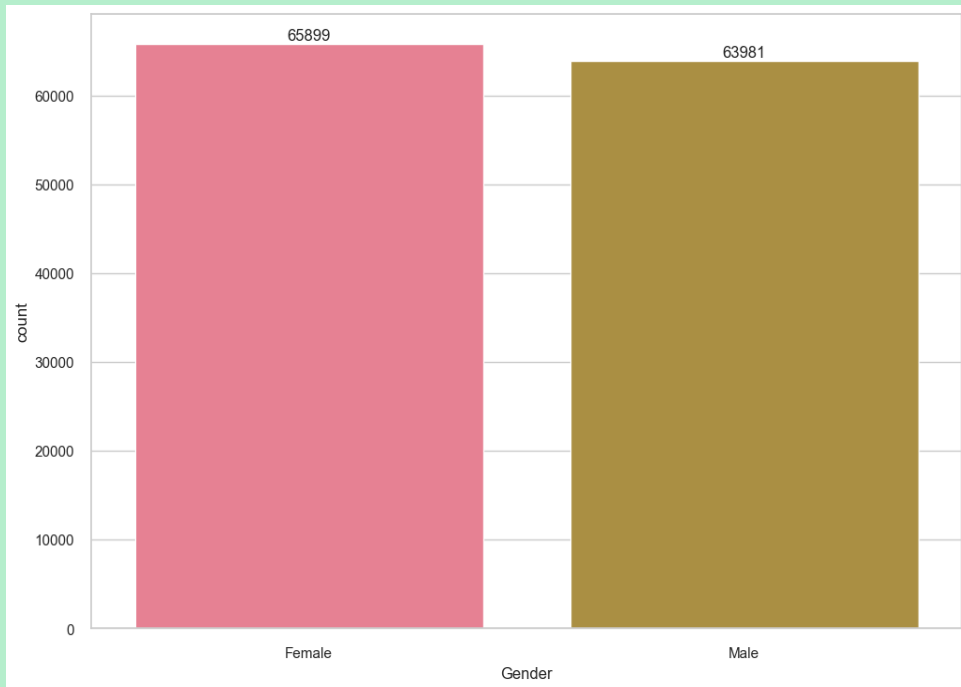
- **Most Passengers were in Business Class and the percentage of passengers in Economy Plus is relatively small.**
- **Gender seemed to be Equal in data.**

EDA

What is the relationship between the flight distance and departure delay?

- **We see that there is no relationship between flight distance and departure delay**
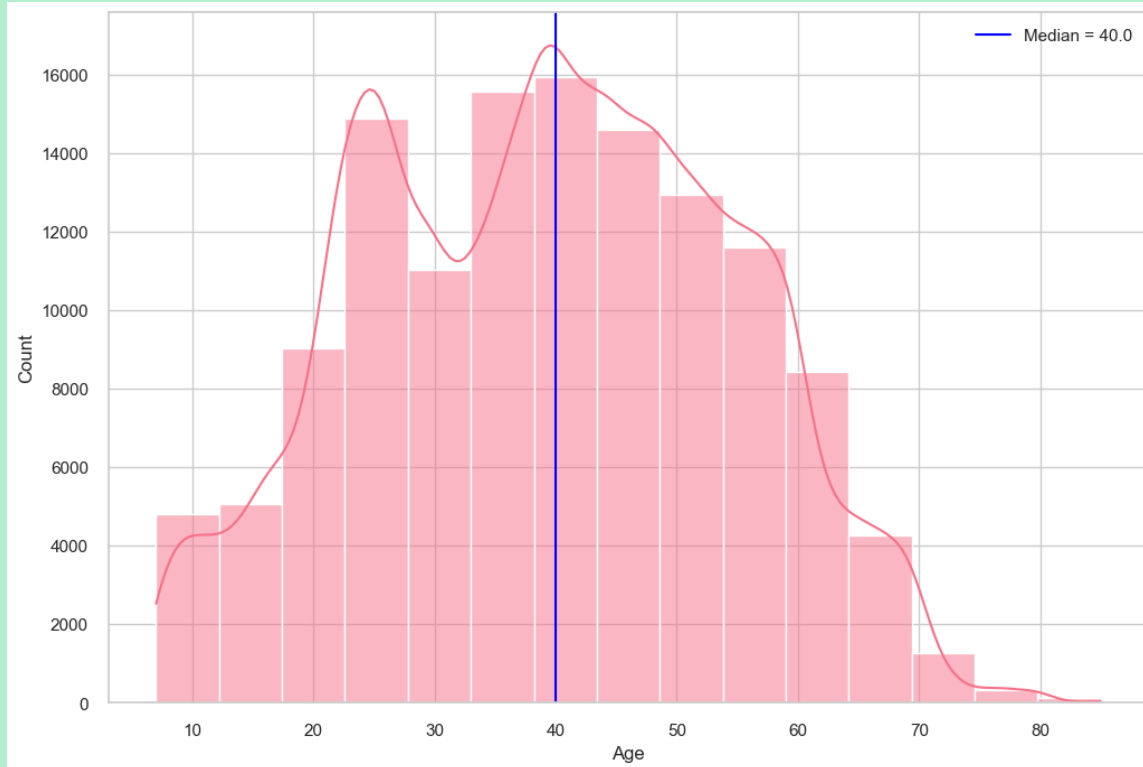
EDA

What is the relationship between the age and type Travel?
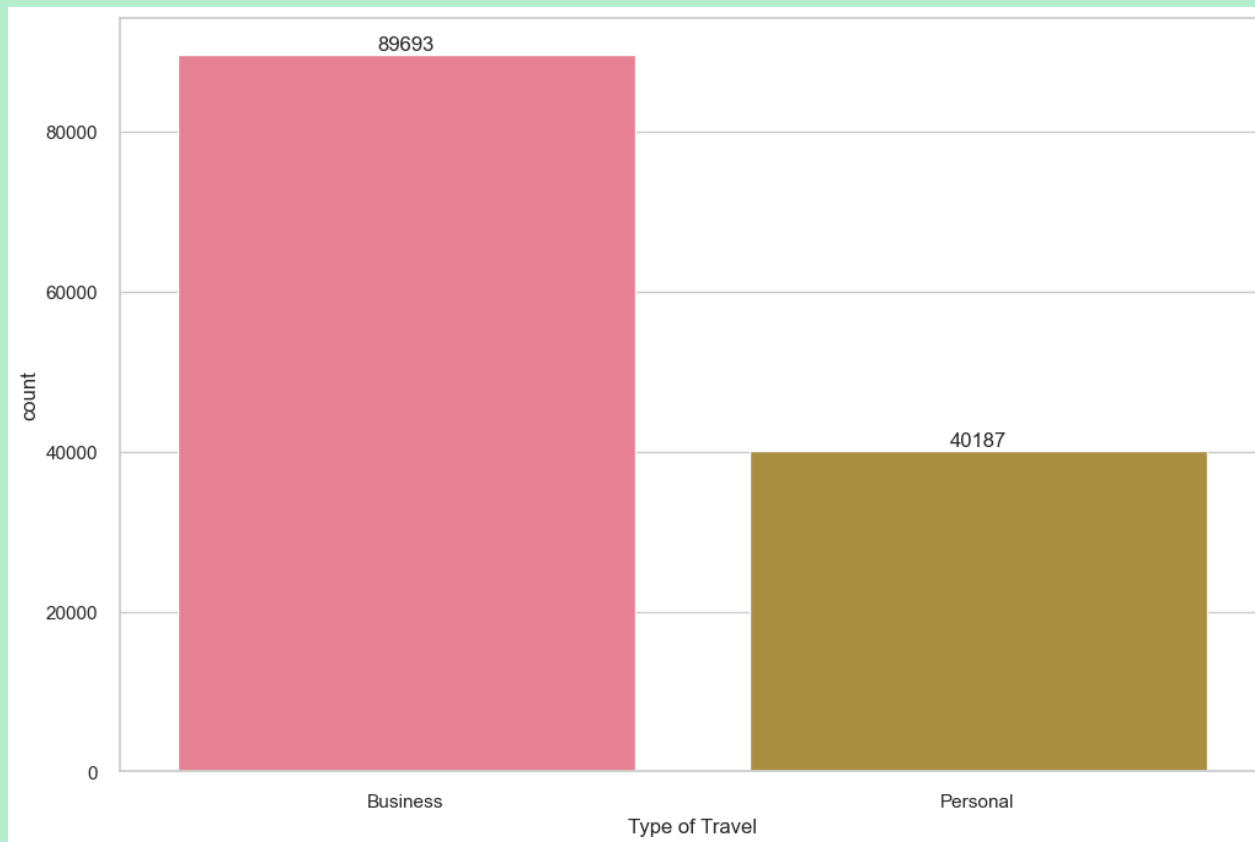
- **The Most common Type of Travel is Business**

EDA

How is the Age affect ?

- we can see from Age Distribution that:

- Ages has Normal ditribution

- Average Age in Passengers is dlo sraey 40

- Age .oot tola tneserp si 25

Data Pre-Processing

So, Let's handle Missing Values

Approximate Distinct Count		472
Approximate Unique (%)		0.4%
Missing		393
Missing (%)		0.3%
Infinite		0
Infinite (%)		0.0%
Memory Size		2.0 MB

Arrival Delay
numerical

Hide Details

- There is one column with missing values we identify it from his type (float) which indicates to either float answers or nan Values.
- The 'arrival delay in minutes' has some null values

Data Pre-Processing

So, Let's handle outliers
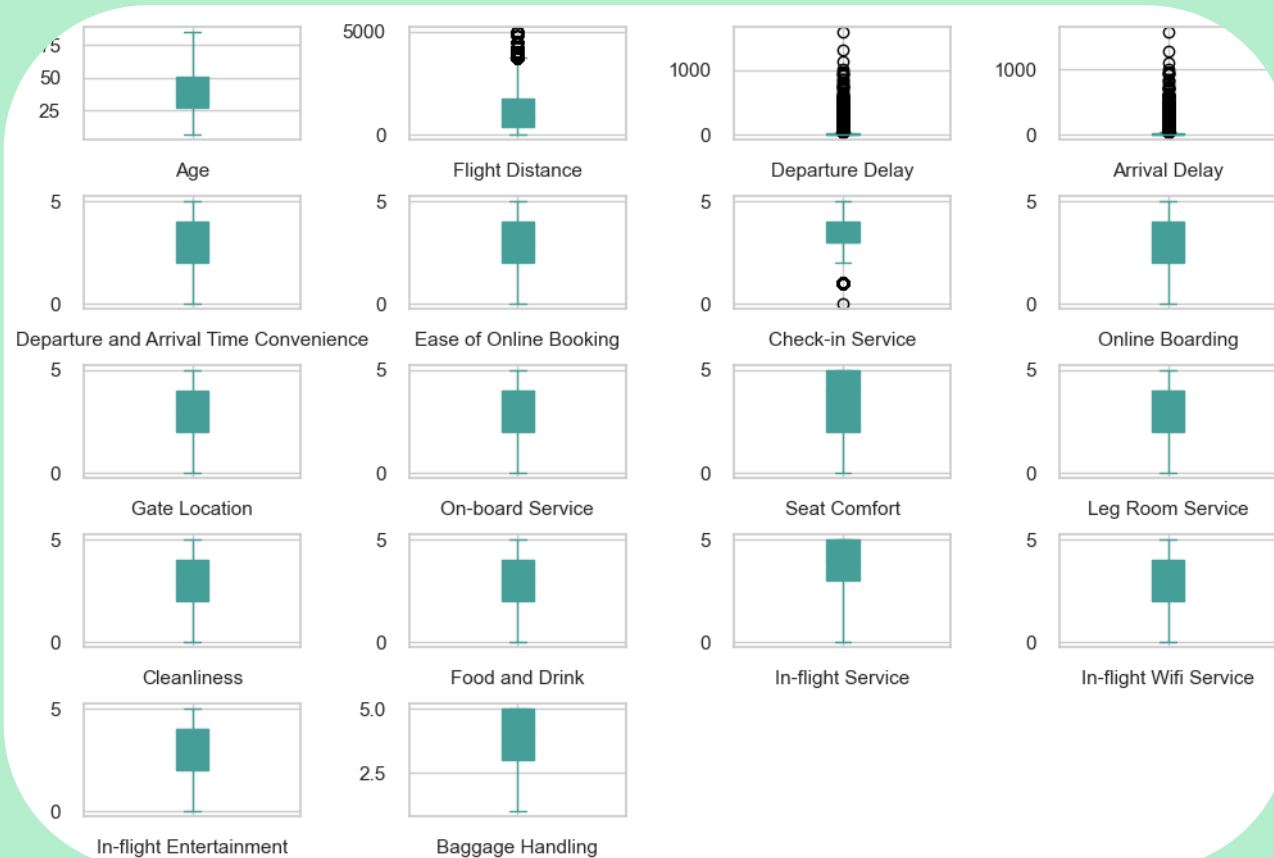
- After processing the reason of outliers , We found that the best solution is to remove them

Data Modeling

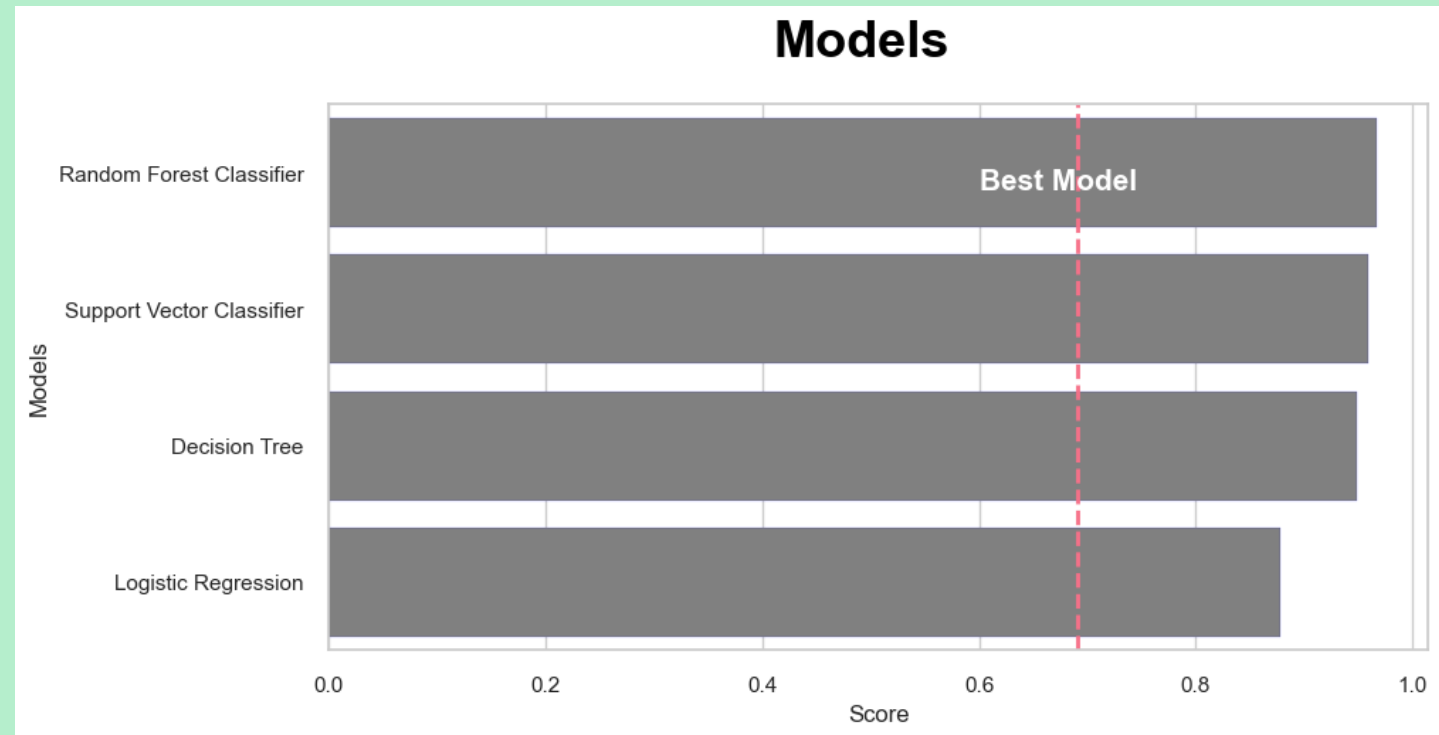As we see that we have many categorical variables so the optimal solution to use the ensemble model

- Decision Tree Regression
- Random Forest Regression
- Logistic Regression
- Support Vector Classifier
  - The best model is : Random Forest Regression

Conclusion

Age distributions for passengers vary by travel type, satisfaction, customer type, and class:

Older passengers are generally satisfied, returning customers in business class, while younger passengers are often dissatisfied, first-time customers in economy class. Age distribution is similar across genders.

Flight distance varies by travel type, satisfaction, customer type, and class:

Business travelers, satisfied passengers, and returning customers tend to travel longer distances. Gender and total time delays show similar distributions for kilometers traveled.

Categories:

Returning customers are nearly equally satisfied and dissatisfied, whereas first-time customers are mostly dissatisfied. Business class passengers are more satisfied than those in economy and eco plus classes. Business travelers report higher satisfaction than personal travelers. Satisfaction levels are similar for men and women.

Conclusion

Services

In-flight service, baggage handling, and seat comfort receive the highest ratings. Wi-Fi, ease of online booking, and gate location receive poor ratings. Business class gives higher service ratings, but gate location and time delays are rated poorly. Economy class rates Wi-Fi and online booking the lowest.

Returning customers rate online boarding, seat comfort, and in-flight entertainment highly, while first-time customers rate these services poorly. Satisfied passengers give high service ratings, while gate location receives moderate ratings from both satisfied and dissatisfied passengers.

Service ratings are similar for both genders. Business travelers give higher service ratings, while personal travelers rate delays excellently. Check-in service ratings are moderate and equal for personal and business passengers.

Delays

There is a strong relationship between departure and arrival delay times. Total time delays show no correlation with travel type, satisfaction, gender, class, or customer type and are distributed equally.

THANKS!

Do you have any questions?

Data Link :

https://www.kaggle.com/datasets/mysarahmadbhat/airline-passenger-satisfaction