Airline Passenger Satisfaction

Author: Ishan Kuchroo

5/3/2022

App Link - https://dashapp-ujgxf7gouq-uk.a.run.app/

Table of Contents

Acknowledgment

I thank Prof. Reza Jafari for guiding us and providing us with techniques to analyze and visualize data. I thank my classmates for keeping the Spring session interactive.

Abstract

Data visualization is the graphical representation of information and data. It helps to tell stories by curating data into a form easier to understand, highlighting the trends and removing the noise. By using visual elements like charts, and graphs, it tools provide an accessible way to see and understand trends, outliers, and patterns in data.

The final project details the implementation of tools, libraries and techniques acquired during the course.

Introduction

Final project will talk about application of data visualization techniques to analyze “customer satisfaction” survey data for the airlines. We will talk about factors that are highly correlated to a satisfied (or dissatisfied) passenger giving the airline an opportunity to improve their services.

Methods – Theory and Procedure

Data Extraction/Preparation

1. **Survey Data**:

Data Source – https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction

The dataset contains an airline passenger satisfaction survey. We have following files:

- **Train.csv** – 103904 rows; 25 columns

Data Analysis - Statistical Methods

1. **Mean**:

Also known as "average", is the value we get when we add up all the terms and divide by the number of terms.

Formula

$$m = \frac{\text{sum of the terms}}{\text{number of terms}}$$

m = mean

Procedure: To implement this method, we use the function "**mean**" from Python library NumPy.

2. **Variance**:

Variance is the measurement of spread of numbers in a dataset, specifically how far is a number from the mean of the dataset.

Formula

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$

S^2 = sample variance
x_i = the value of the one observation
$\bar{x}$ = the mean value of all observations
n = the number of observations

Procedure: To implement this method, we use the function "**var**" from Python library NumPy.

3. **Pearson's Correlation Coefficient (r)**:

Pearson's coefficient measures the statistical relationship, or association, between two continuous variables

Formula

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

r = correlation coefficient
x_i = values of the x-variable in a sample
$\bar{x}$ = mean of the values of the x-variable
y_i = values of the y-variable in a sample
$\bar{y}$ = mean of the values of the y-variable

Data Overview & Pre-processing

In this section, we will talk about the RAW data and data preprocessing steps.

- Read **train.csv** file in a dataframe using "Pandas"
    - Original dataframe consists of **25 columns** and **103904 rows**
    - 6 continuous variables and 19 categorical variables

```
Columns in original dataset:
 Index(['Unnamed: 0', 'id', 'Gender', 'Customer Type', 'Age', 'Type of Travel',
       'Class', 'Flight Distance', 'Inflight wifi service',
       'Departure/Arrival time convenient', 'Ease of Online booking',
       'Gate location', 'Food and drink', 'Online boarding', 'Seat comfort',
       'Inflight entertainment', 'On-board service', 'Leg room service',
       'Baggage handling', 'Checkin service', 'Inflight service',
       'Cleanliness', 'Departure Delay in Minutes', 'Arrival Delay in Minutes',
       'satisfaction'],
      dtype='object')
```

    - **Dependent/Target Variable – "satisfaction"**

- Removing identifier columns ("**id**", " **Unnamed: 0**") from dataframe. This done because their value varies from one sample to another, they are irrelevant to any useful pattern and have no predictive power over the target.

- Check for NaN, Null or missing values in dataframe –
    - Only one column "**Arrival Delay in Minutes**"

```
Columns with NA values - Before removal:
Arrival Delay in Minutes : 310
```

    - I have removed the missing rows from the dataset.

```
Columns with NA values - After removal:
```

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Final Term Project – Data Visualization

Screengrab of Data Overview:

Snapshot

Snapshot

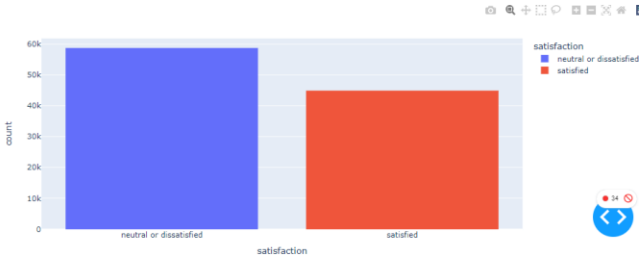
Gender	Customer Type	Age	Type of Travel	Class	Flight Distance	Inflight wifi service	Departure/Arrival time convenient	Ease of Online booking	Gate location	Food and drink	Online boarding	Seat comfort	Inflight entertainment	C... l... s...
Male	Loyal Customer	13	Personal Travel	Eco Plus	460	3	4	3	1	5	3	5	5	4
Male	disloyal Customer	25	Business travel	Business	235	3	2	3	3	1	3	1	1	1
Female	Loyal Customer	26	Business travel	Business	1142	2	2	2	2	5	5	5	5	4
Female	Loyal Customer	25	Business travel	Business	562	2	5	5	5	2	2	2	2	2
Male	Loyal Customer	61	Business travel	Business	214	3	3	3	3	4	5	5	3	3
Female	Loyal Customer	26	Personal Travel	Eco	1180	3	4	2	1	1	2	1	1	3
Male	Loyal Customer	47	Personal Travel	Eco	1276	2	4	2	3	2	2	2	2	3
Female	Loyal Customer	52	Business travel	Business	2035	4	3	4	4	5	5	5	5	5
Female	Loyal Customer	41	Business travel	Business	853	1	2	2	2	4	3	3	1	1
Male	disloyal Customer	20	Business travel	Eco	1061	3	3	3	4	2	3	3	2	2
Female	disloyal Customer	24	Business travel	Eco	1182	4	5	5	4	2	5	2	2	3
Female	Loyal Customer	12	Personal Travel	Eco Plus	308	2	4	2	2	1	2	1	1	1
Male	Loyal Customer	53	Business travel	Eco	834	1	4	4	4	1	1	1	1	1
Male	Loyal Customer	33	Personal Travel	Eco	946	4	2	4	3	4	4	4	4	4
Female	Loyal Customer	26	Personal Travel	Eco	453	3	2	3	2	2	3	2	2	4
Male	disloyal Customer	13	Business travel	Eco	486	2	1	2	3	4	2	1	4	2
Female	Loyal Customer	26	Business travel	Business	2123	3	3	3	3	4	4	4	4	5

Target Variable Analysis

Statistics

index	Age	Flight Distance	Inflight wifi service	Departure/Arrival time convenient	Ease of Online booking	Gate location	Food and drink	
mean	39.380466050157345	1186.0061055659594	2.7297526883796360	3.0600806996544203	2.7569839952120780	2.9770256964689072	3.2021256057300618	3
std	15.113125274715594	980.2646837167999	1.3278662408249535	1.5252328832652130	1.3989335816552615	1.2777250471261020	1.3294012031090188	1
min	7	31	0	0	0	0	0	0
25%	27	414	2	2	2	2	2	2
50%	40	842	3	3	3	3	3	3
75%	51	1743	4	4	4	4	4	4
max	85	3736.5	5	5	5	5	5	5

Customer Satisfaction - Target



Column Name	Column Description
Gender	Gender of the passengers (Female, Male)
Customer Type	The customer type (Loyal customer, disloyal customer)
Age	The actual age of the passengers
Type of Travel	Purpose of the flight of the passengers (Personal Travel, Business Travel)
Class	Travel class in the plane of the passengers (Business, Eco, Eco Plus)
Flight distance	The flight distance of this journey
Inflight wifi service	Satisfaction level of the inflight wifi service (0-5)
Departure/Arrival time convenient	Satisfaction level of Departure/Arrival time convenient
Ease of Online booking	Satisfaction level of online booking
Gate location	Satisfaction level of Gate location
Food and drink	Satisfaction level of Food and drink
Online boarding	Satisfaction level of online boarding
Seat comfort	Satisfaction level of Seat comfort
Inflight entertainment	Satisfaction level of inflight entertainment
On-board service	Satisfaction level of On-board service
Leg room service	Satisfaction level of Leg room service
Baggage handling	Satisfaction level of baggage handling
Check-in service	Satisfaction level of Check-in service
Inflight service	Satisfaction level of inflight service
Cleanliness	Satisfaction level of Cleanliness
Departure Delay in Minutes	Minutes delayed when departure
Arrival Delay in Minutes	Minutes delayed when Arrival
Satisfaction	Airline satisfaction level(Satisfaction, neutral or dissatisfaction)

Outlier Detection and Removal

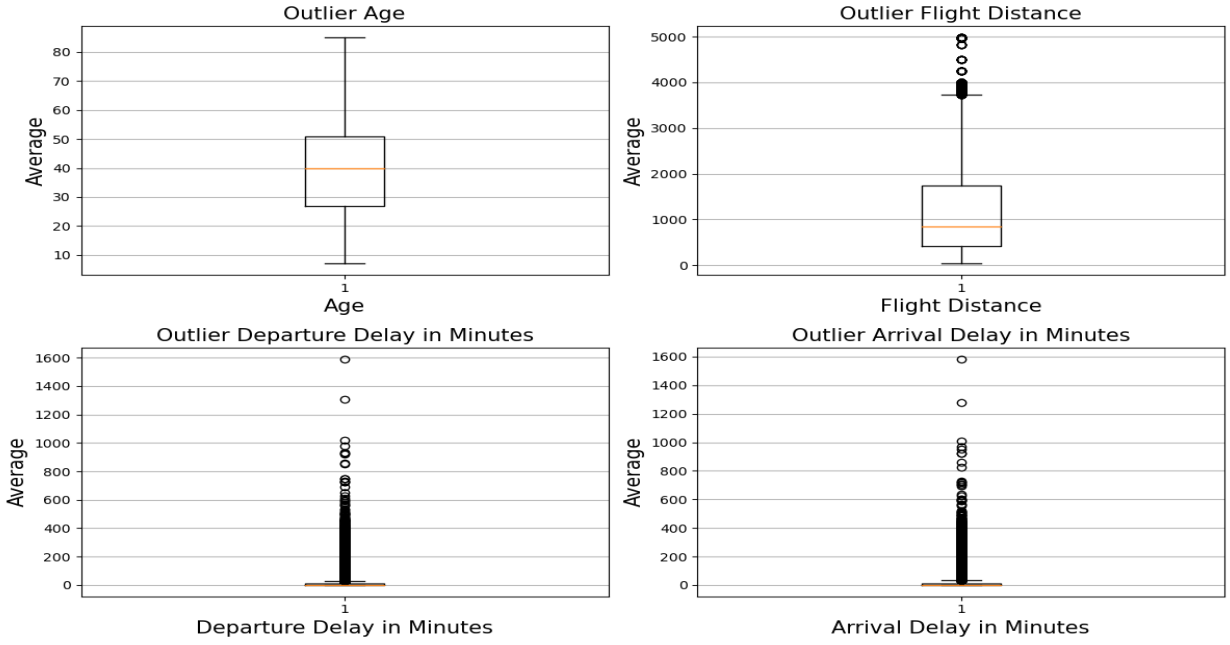
In this section, we will talk about outliers in our data and using the *Inter-Quartile Range* (IQR) method to detect and remove such outliers.

- **What are Outliers?**

As per *National Institute of Standards and Technology*, an outlier is an observation that lies an abnormal distance from other values in a random sample from a population. This definition leaves it up to the analyst (or a consensus process) to decide what is considered abnormal. Before abnormal observations can be singled out, it is necessary to characterize normal observations.

- **Detection – Box Plot and IQR**

A useful graphical display for describing the above behavior of the data is through Boxplot. The box plot uses the median and the lower and upper quartiles (defined as the 25th and 75th percentiles) so we can clearly see if we have outliers in the data. If the lower quartile is Q1 and the upper quartile is Q3, the difference (Q3 - Q1) called the interquartile range or IQR. IQR is for measuring variability by dividing a data set into quartiles.



- **Removal**

Once we know the IQR, we define lower and upper limits as follows:

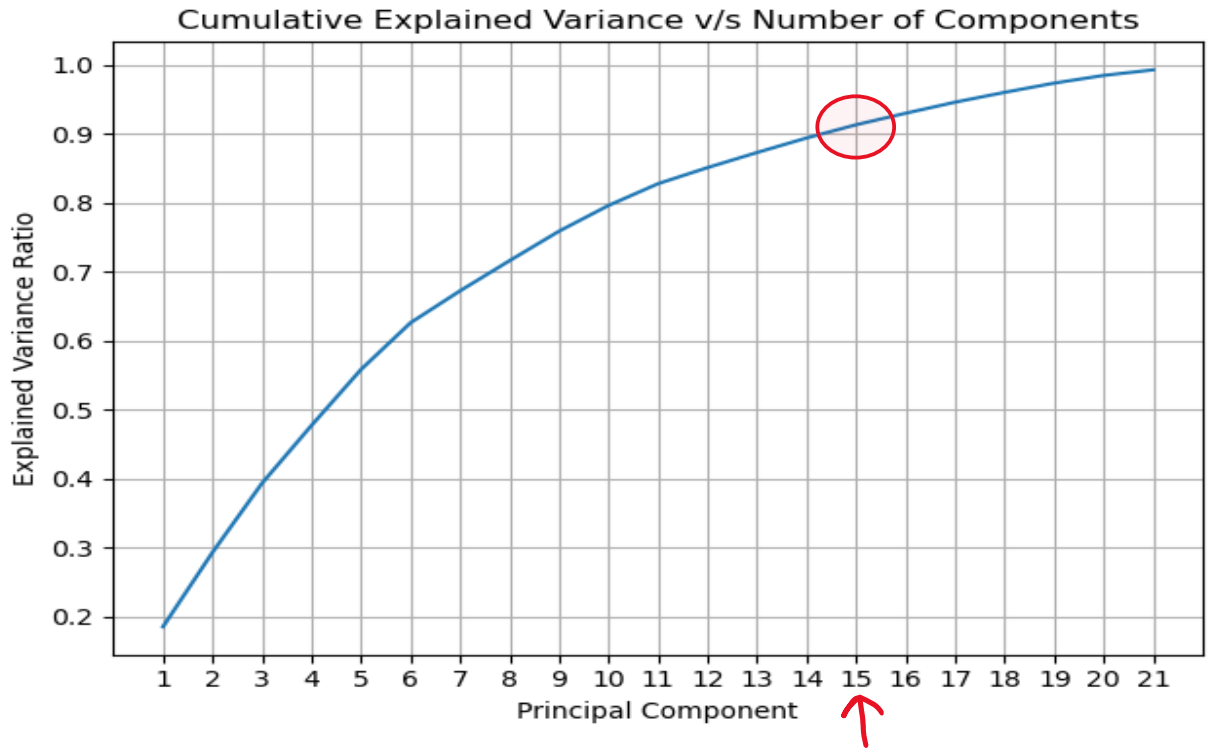
```
low_lim = Q1 - 1.5 * IQR
up_lim = Q3 + 1.5 * IQR
```

If a value is greater than up_lim, we impute the said value with up_lim. If it is less than low_lim, then we impute the value with low_lim.

Principal Component Analysis

Principal component analysis (PCA) is finding the principal components and using them to perform a change of basis on the data, sometimes using only the first few principal components and ignoring the rest. It is for dimensionality reduction wherein the goal is to find a smaller number of components to explain as much of the variance possible. It also helps us deal with multi-collinearity in our dataset.

For our dataset, as can be seen below, 15 components can explain more than 90% of the variance in our data.



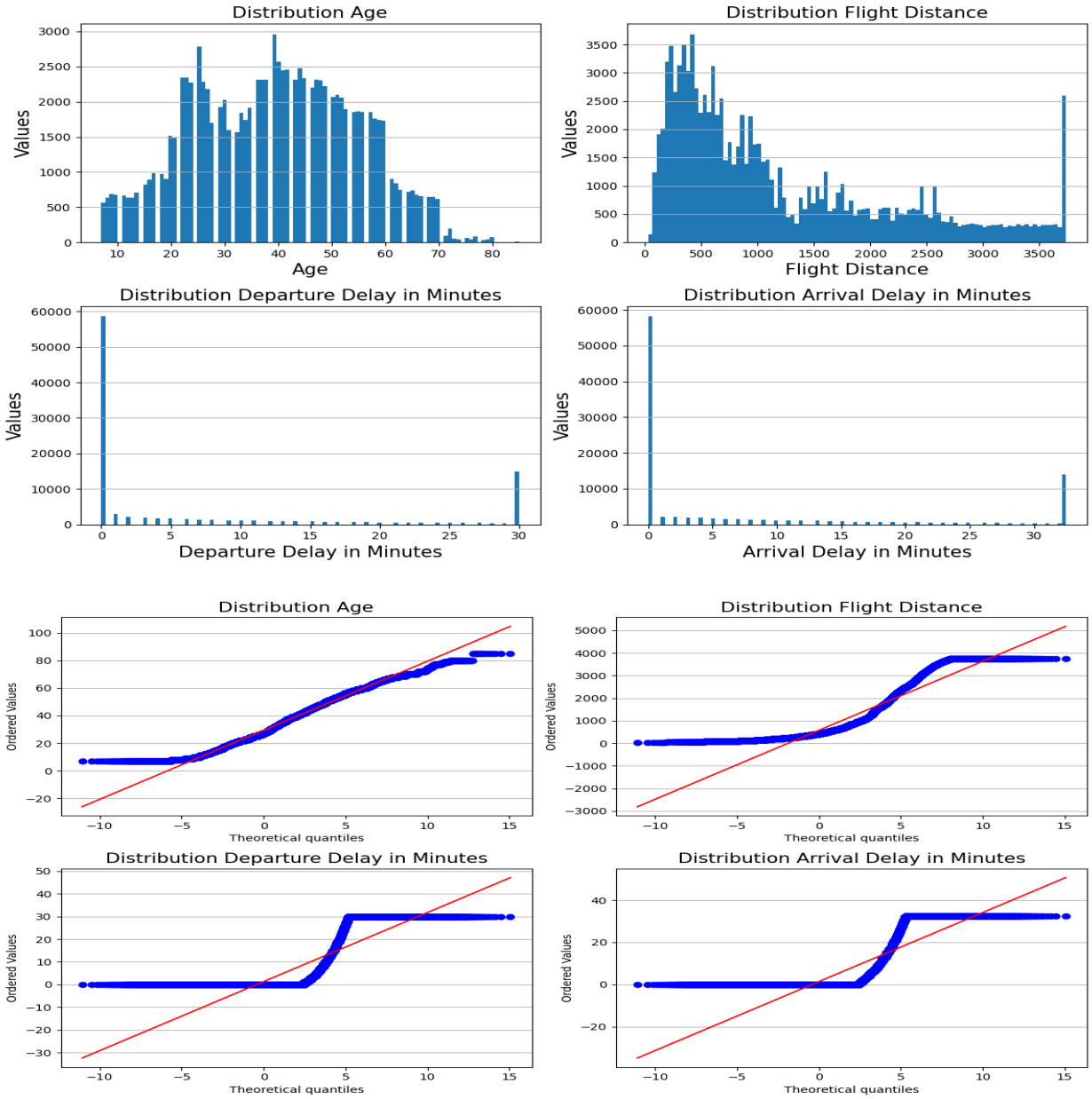

```
Singular value of dataframe are [421543.61352382 245923.21088318 228404.66521935 191381.74061017
 184898.35054934 154431.90574883 105549.34202756  99706.88780544
  97624.02732098  85045.90759272  71889.17976263  53219.75991447
  49914.74410206  48076.10139736  43642.6475836   38121.20059922
  36563.94958976  32490.49630834  30512.53093929  25419.07859791
  18390.27608836  16318.3838356 ]
Condition number for dataframe is 5.082561997417234
Explained Variance Ratio - Original v/s Reduced Feature Space:
 [0.18496316 0.10790517 0.10021845 0.08397369 0.08112893 0.06776099
 0.0463125  0.04374897 0.04283507 0.03731609 0.03154324 0.02335155
 0.02190138 0.02109463 0.01914934 0.01672666 0.01604338 0.01425605
 0.01338816 0.01115328 0.00806921]
##########################################################################################################
##########################################################################################################
Singular value of PCA dataframe are [421543.61352382 245923.21088318 228404.66521935 191381.74061017
 184898.35054934 154431.90574883 105549.34202756  99706.88780544
  97624.02732098  85045.90759272  71889.17976263  53219.75991447
  49914.74410206  48076.10139736  43642.6475836 ]
```

Normality Test and Transformation

Normality refers to a statistical distribution called a normal distribution, or sometimes the Gaussian distribution or bell-shaped curve is a symmetrical continuous distribution defined by the mean and standard deviation of the data. We will be testing for normality on the numeric columns in our data.

- **Histogram and QQ-plot**:

Both these plots help us understand if a data distribution is fitting the Gaussian distribution. As seen below, apart from Age none of the columns is even close to a Gaussian distribution.



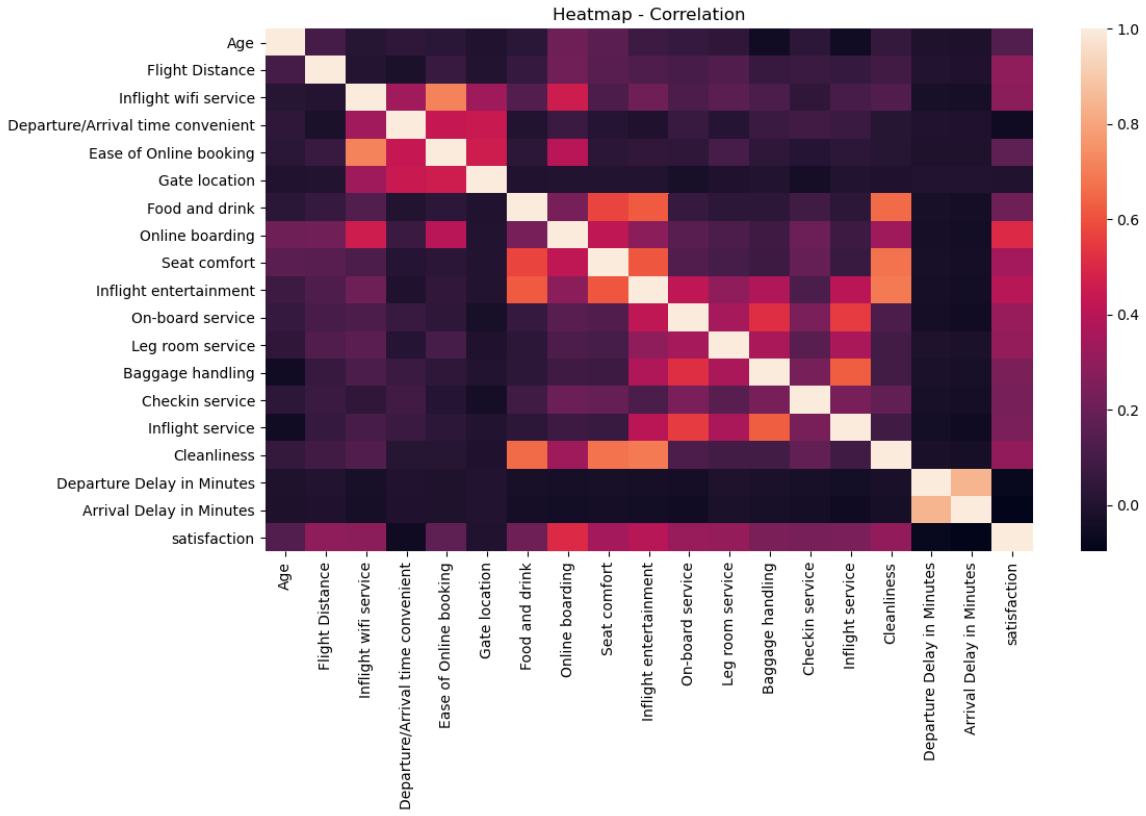
- **Normality Tests**:

We'll use the Shapiro-Wilk Test which will tell if a random sample comes from a normal distribution.

```
Shapiro-Wilk Test Result - Age:

Shapiro Test: statistics=0.9883827567100525; p-value=0.0
Shapiro Test: Age is not normal distributed
C:\Users\Ishan Kuchroo\Anaconda3\lib\site-packages\scipy\stats\morestats.py:1681: Use
  warnings.warn("p-value may not be accurate for N > 5000.")
C:/Users/Ishan Kuchroo/Desktop/GWU_MS_DataScience/Visualization_of_Complex_Data/Final
  tgt = np.log(tgt)
Shapiro Test: statistics=nan; p-value=1.0
Shapiro Test: Transformed Age is normally distributed


Shapiro-Wilk Test Result - Flight Distance:

Shapiro Test: statistics=0.8631260991096497; p-value=0.0
Shapiro Test: Flight Distance is not normal distributed
Shapiro Test: statistics=nan; p-value=1.0
Shapiro Test: Transformed Flight Distance is normally distributed


Shapiro-Wilk Test Result - Departure Delay in Minutes:

Shapiro Test: statistics=0.6642884016036987; p-value=0.0
Shapiro Test: Departure Delay in Minutes is not normal distributed
Shapiro Test: statistics=nan; p-value=1.0
Shapiro Test: Transformed Departure Delay in Minutes is normally distributed


Shapiro-Wilk Test Result - Arrival Delay in Minutes:

Shapiro Test: statistics=0.671425461769104; p-value=0.0
Shapiro Test: Arrival Delay in Minutes is not normal distributed
Shapiro Test: statistics=nan; p-value=1.0
Shapiro Test: Transformed Arrival Delay in Minutes is normally distributed
```

As seen from the results, the data was not normally distributed. We used **stats.norm.ppf** followed by a log transformation to make the data normally distributed.

Correlation Matrix



Observation:

The plot indicates a strong correlation between the “Departure Delay in Minutes” and “Arrival Delay in Minutes” features which is also true. We can also see high correlation between the “Cleanliness” and “Inflight entertainment,” but it cannot be explained with reason.

Target variable “satisfaction” is moderately correlated with all features, ranging from 20% to 35%, but is well correlated to “Online Booking” – 50%.

Visualization

I have divided my attributes into five categories:

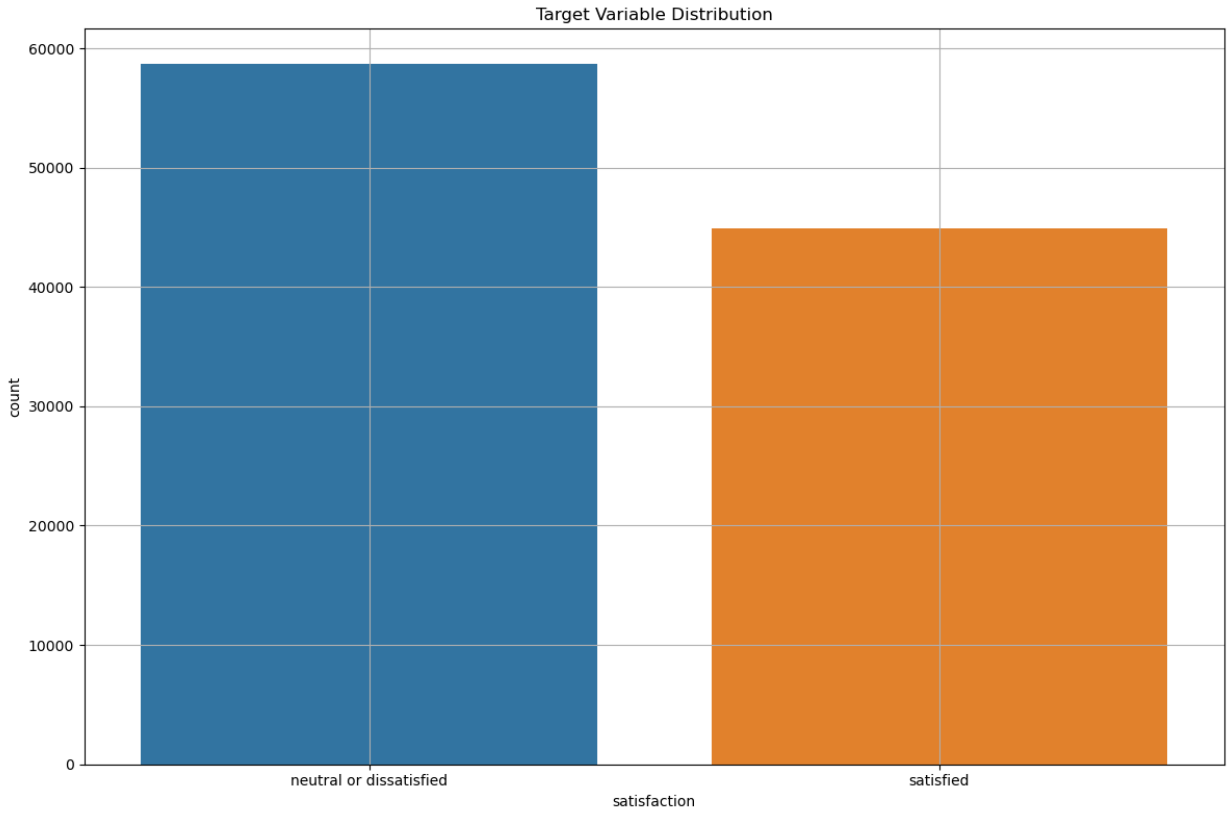
1. **Target Variable**
    a. Satisfaction (satisfied/ dissatisfied)
2. **Customer Demographics**
    a. Gender (Male/Female)
    b. Customer Type (Loyal or Disloyal)
    c. Type of Travel (Personal or Business)
    d. Class (Business, Economy, Economy Plus)
    e. Age
3. **Pre-Boarding Factors (Rating – 0 to 5)**
    a. Ease of Online booking
    b. Gate Location
    c. Online Boarding
    d. Check-in Service
    e. Departure Delay
    f. Departure/Arrival Time Convenient
4. **On-Board Factors (Rating – 0 to 5)**
    a. Inflight wi-fi service
    b. Food and drink
    c. Seat comfort
    d. Inflight entertainment
    e. On-board service
    f. Leg room
    g. Inflight service
    h. Cleanliness
5. **Post-flight Factors (Rating – 0 to 5)**
    a. Baggage handling
    b. Arrival Delay

I will be using customer demographics to understand how pre-boarding, on-board and post-flight factors have impacted the “satisfaction” of a customer.

As most of my variables are categorical, I’ll be using Pie chart to explain the distribution of data for the variable and Histogram to explain the interaction of the variable with customer demographics.

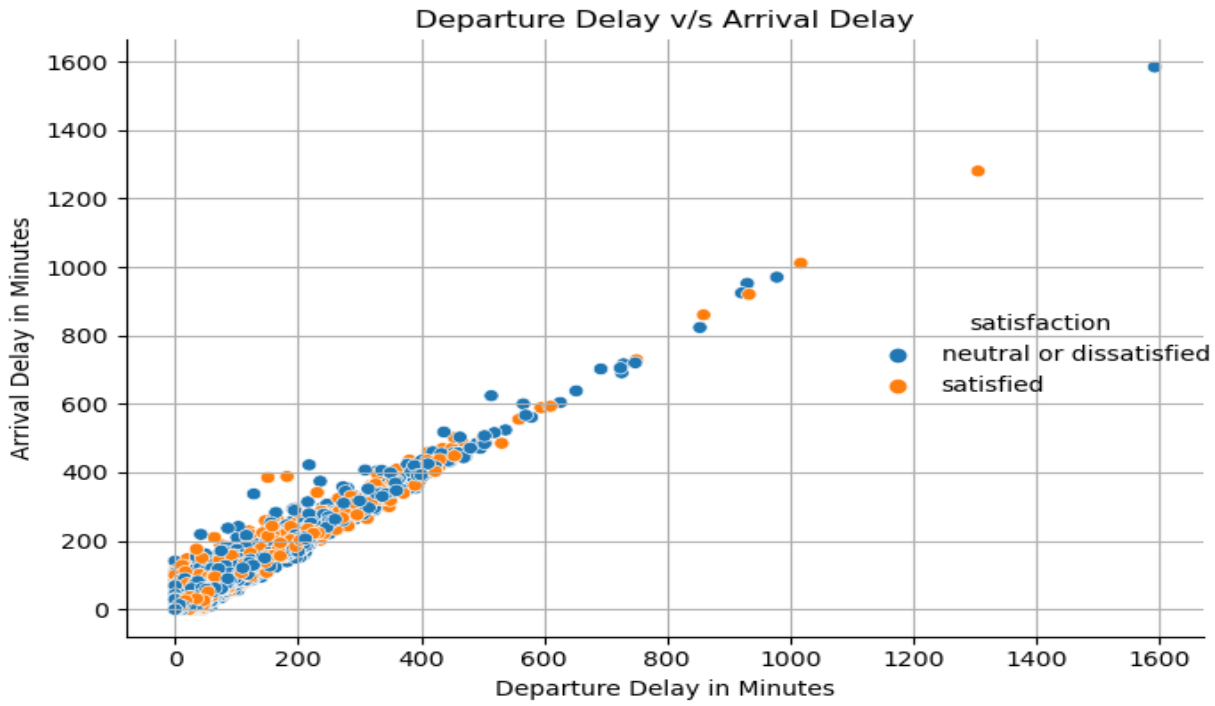
Customer Satisfaction Level (target variable) is used to change the distribution of data.

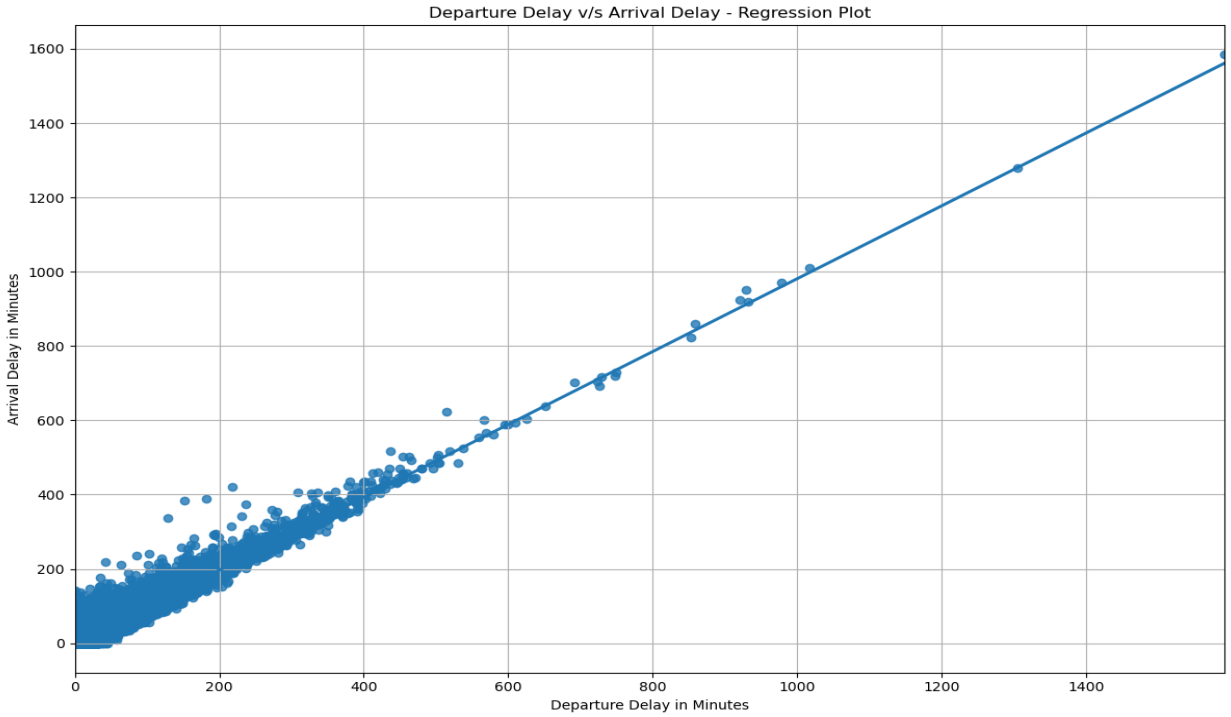
Target Variable



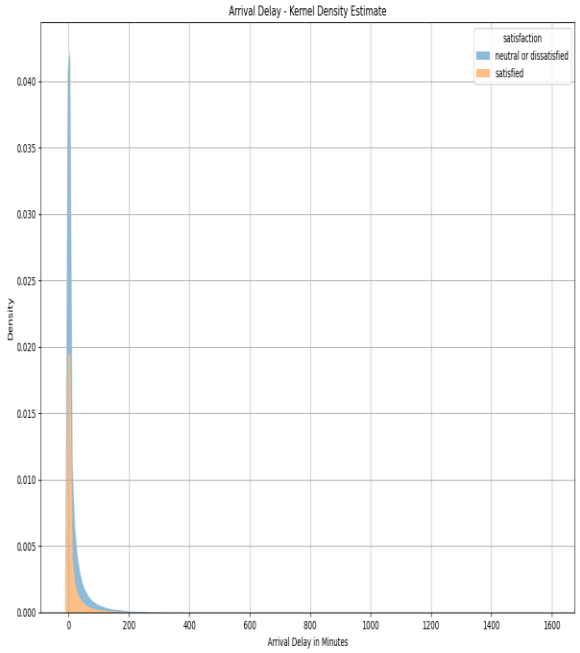
Observation: Mostly customers are not satisfied with the airline close to 57%.

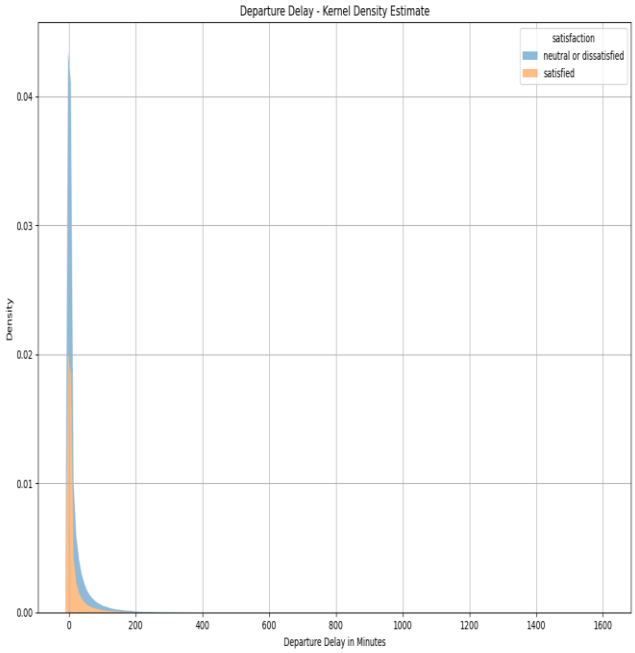
Departure and Arrival Delay



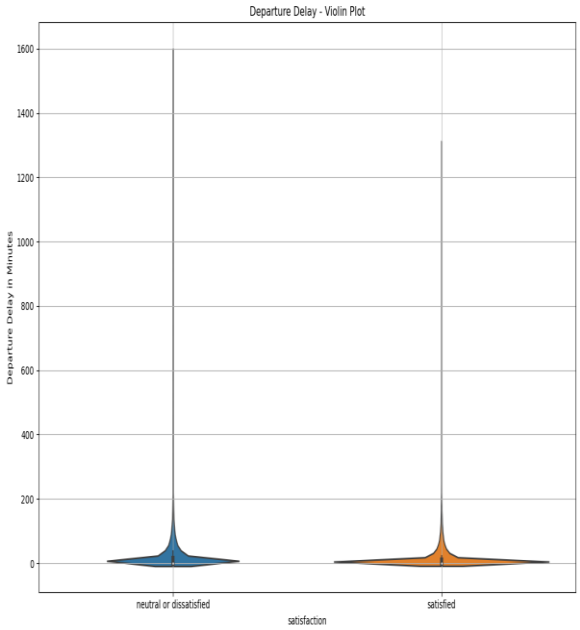


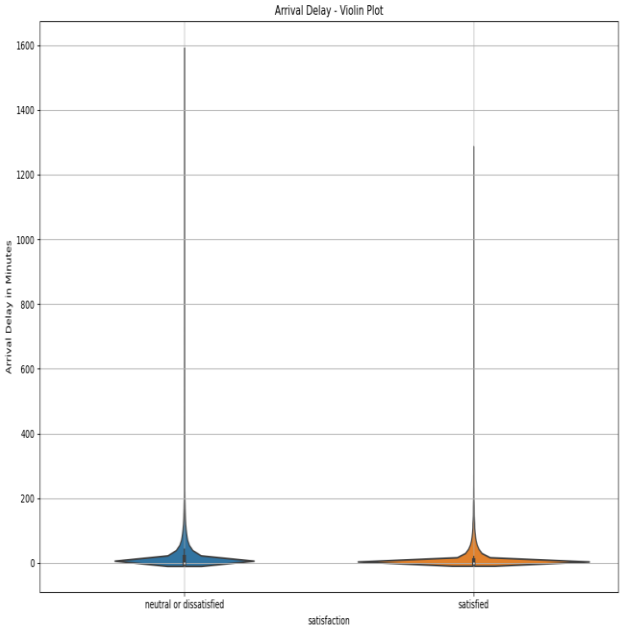
Observation: Departure and Arrival Delay are highly correlated with each other.



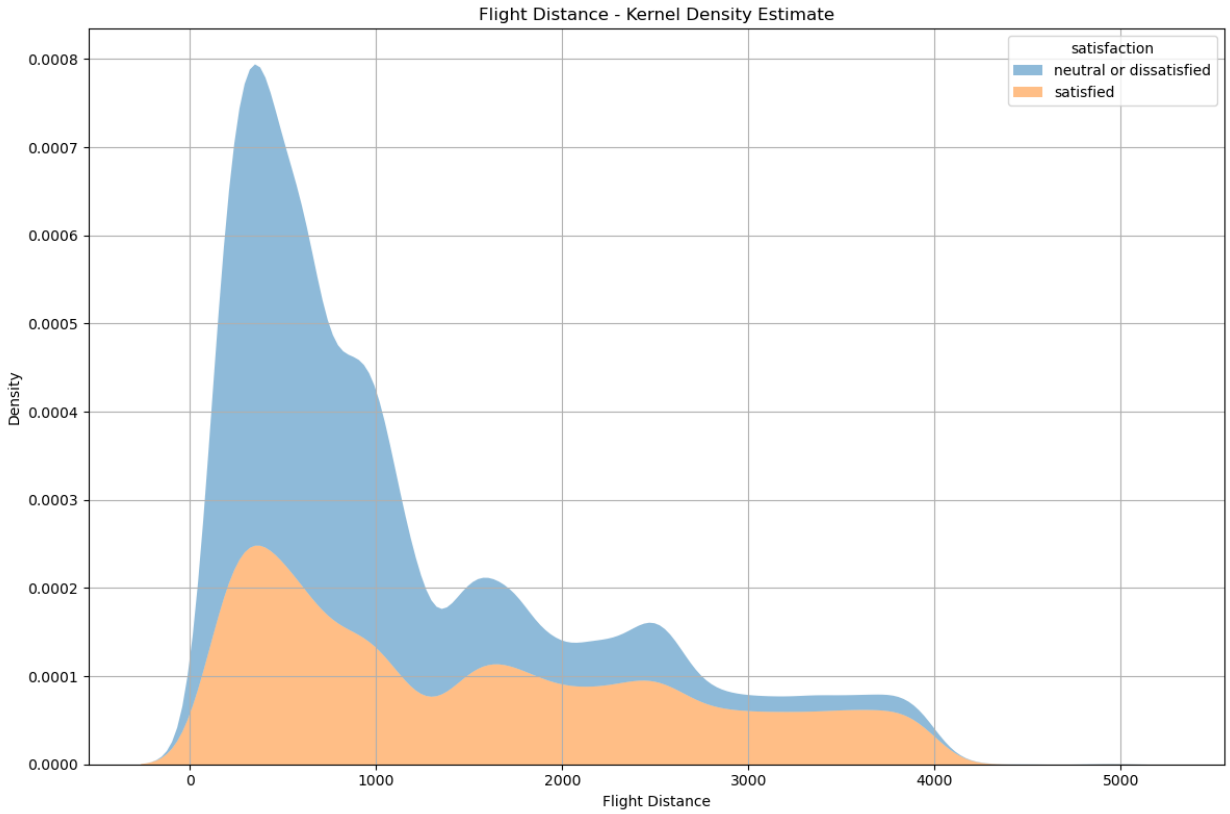


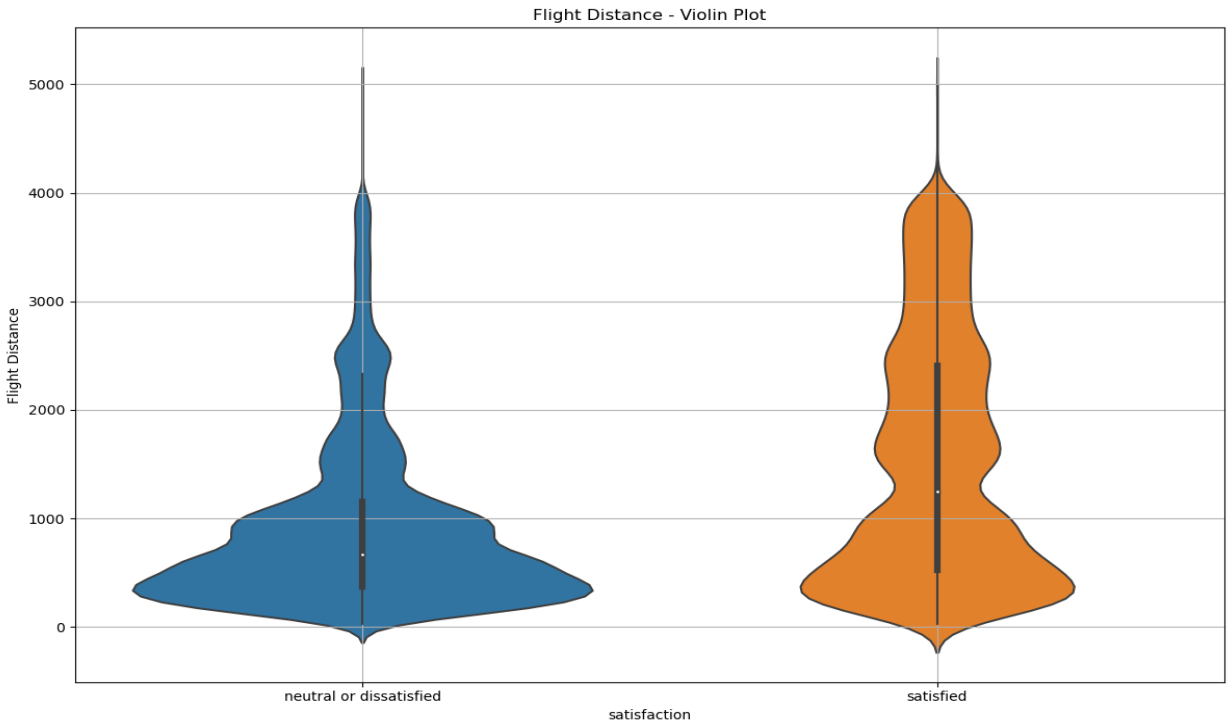
Both the variables are left-skewed and not normally distributed. We'll need to apply transformation.



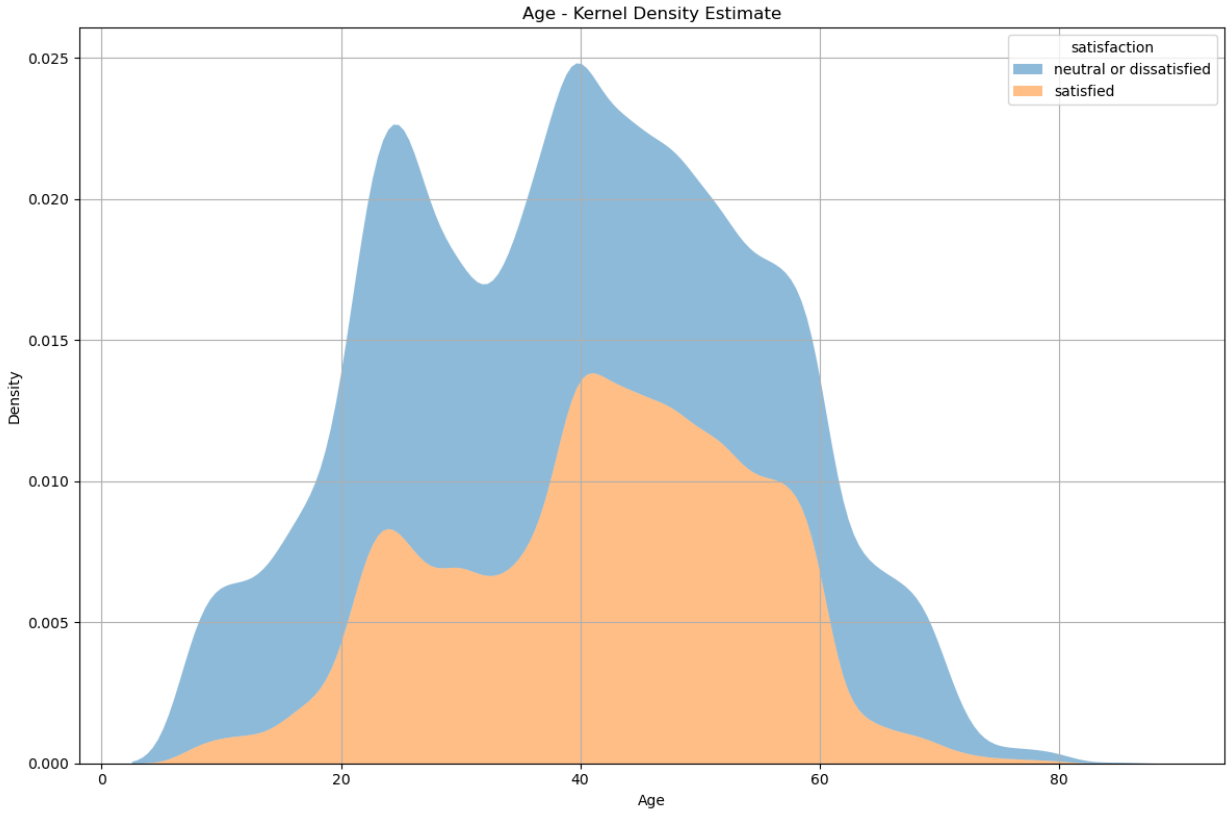


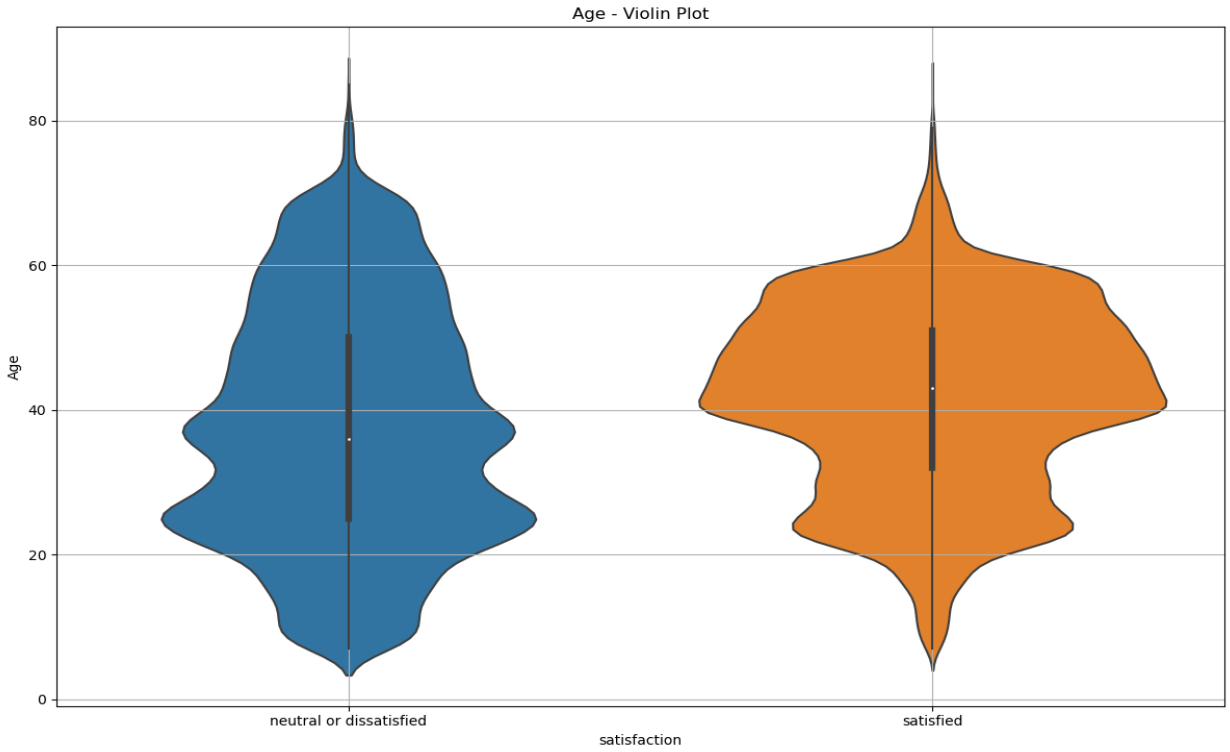
Flight Distance and Age





Final Term Project – Data Visualization





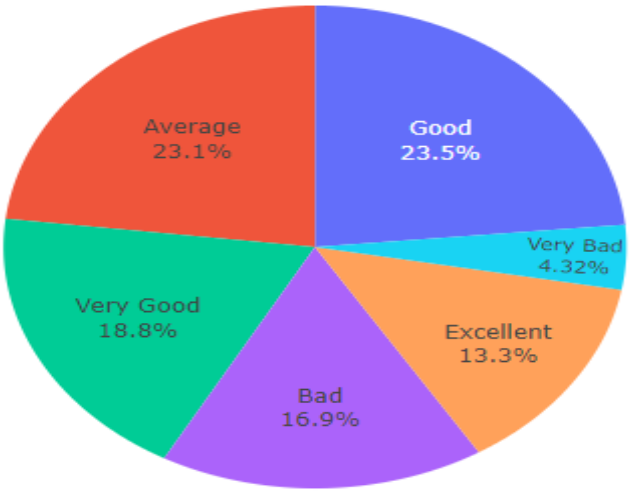
THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Final Term Project – Data Visualization

Pre-Boarding Factors

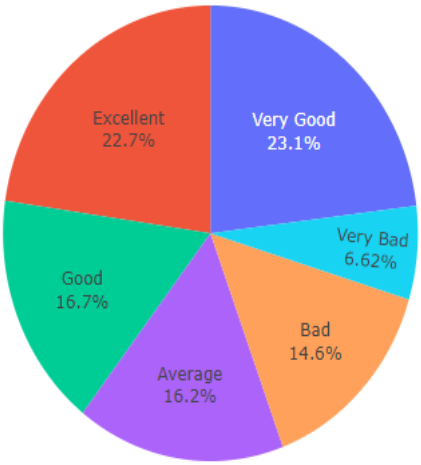
1. Ease of Online Booking:

Overall Distribution

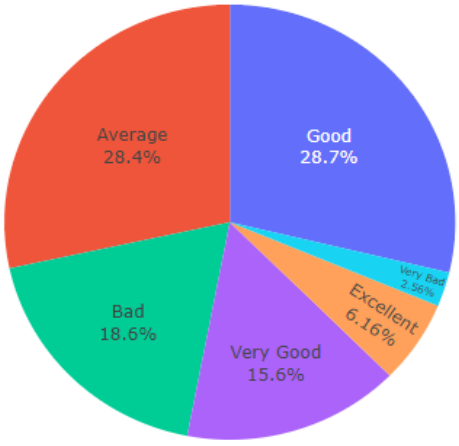


Distribution based on Customer Satisfaction

Satisfied Customers



Not Satisfied Customers



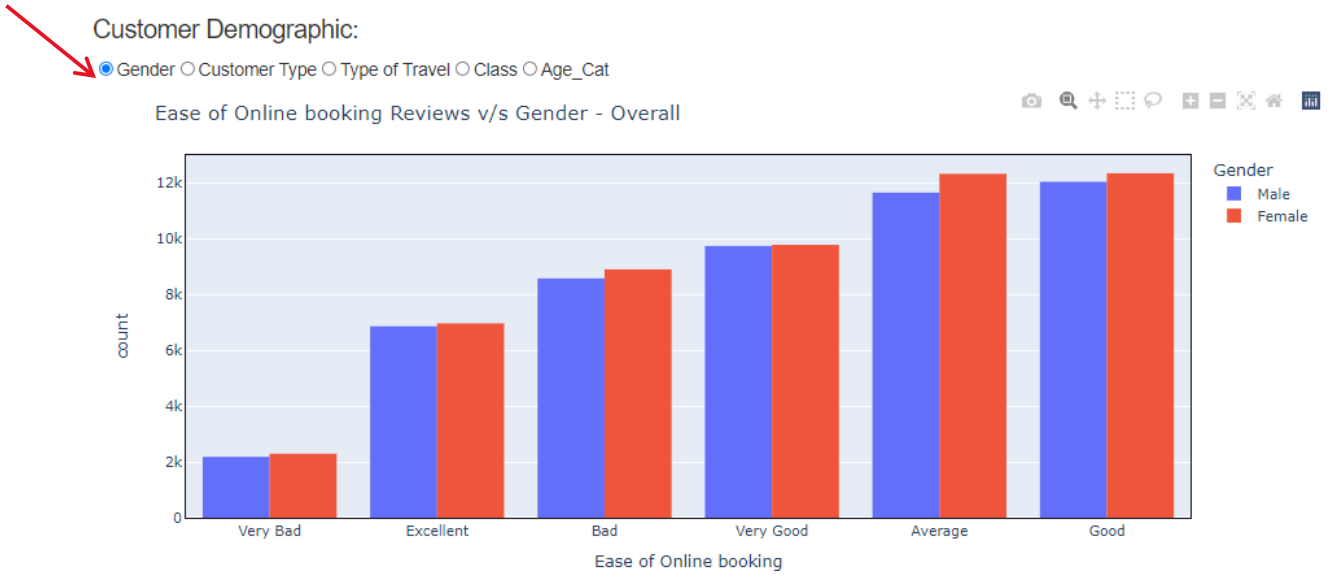
Observation:

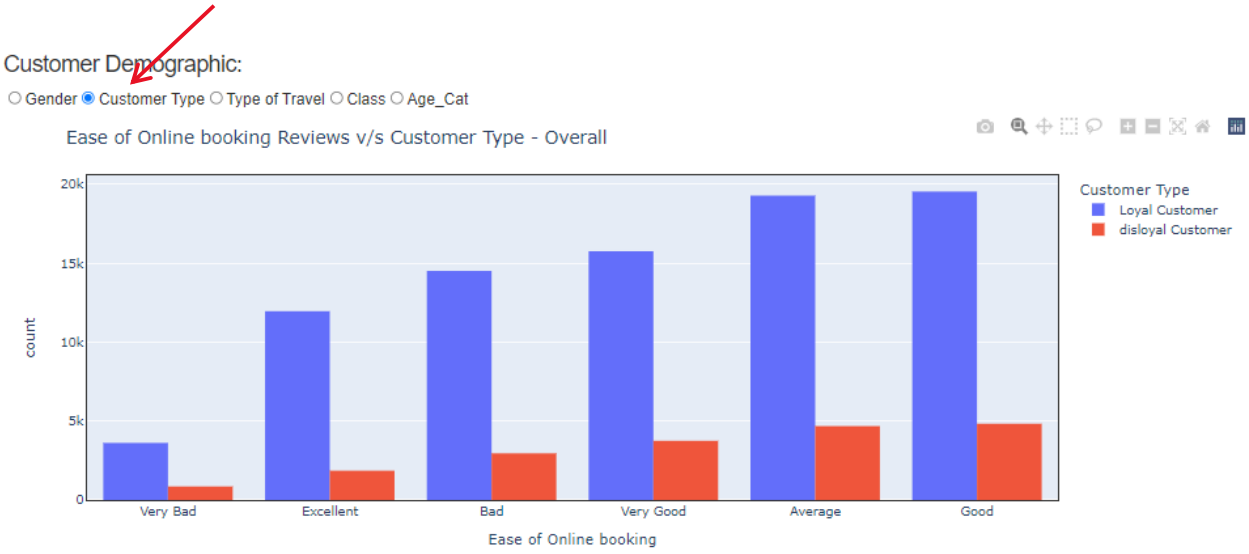
Overall customers have given a “Average” or a “Good” review for ease with which they were able to book a ticket online. However, satisfied customers have given more Excellent rating compared to customers not satisfied with service.

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Final Term Project – Data Visualization

Interaction with Customer Demographics (Gender, Customer Type, Type of Travel, Class, and Age)



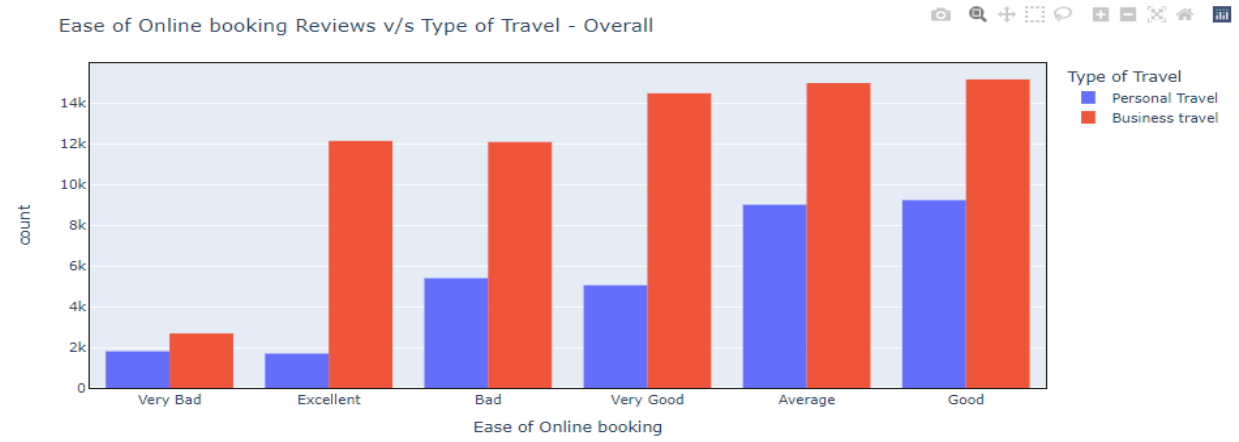


THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Final Term Project – Data Visualization

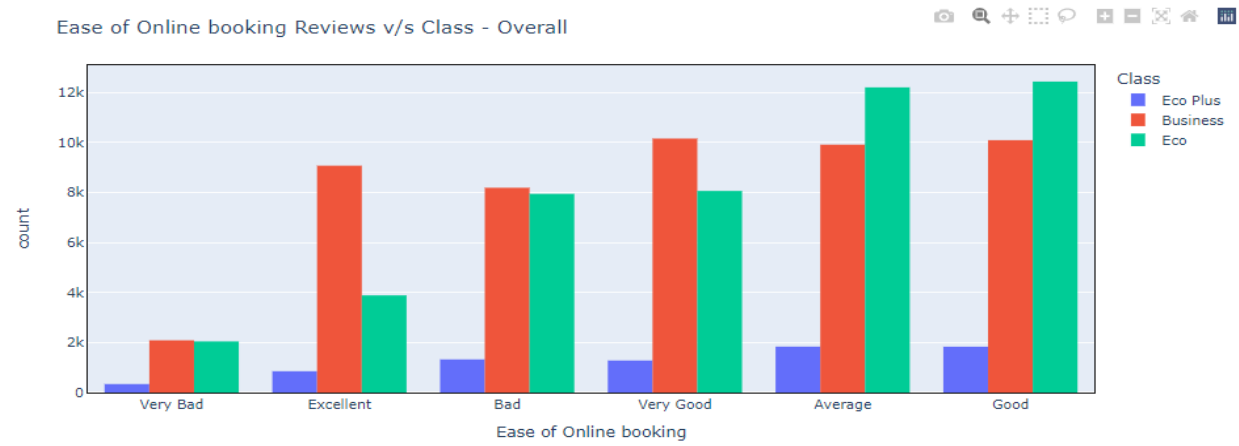
Customer Demographic:

○ Gender ○ Customer Type ◉ Type of Travel ○ Class ○ Age_Cat



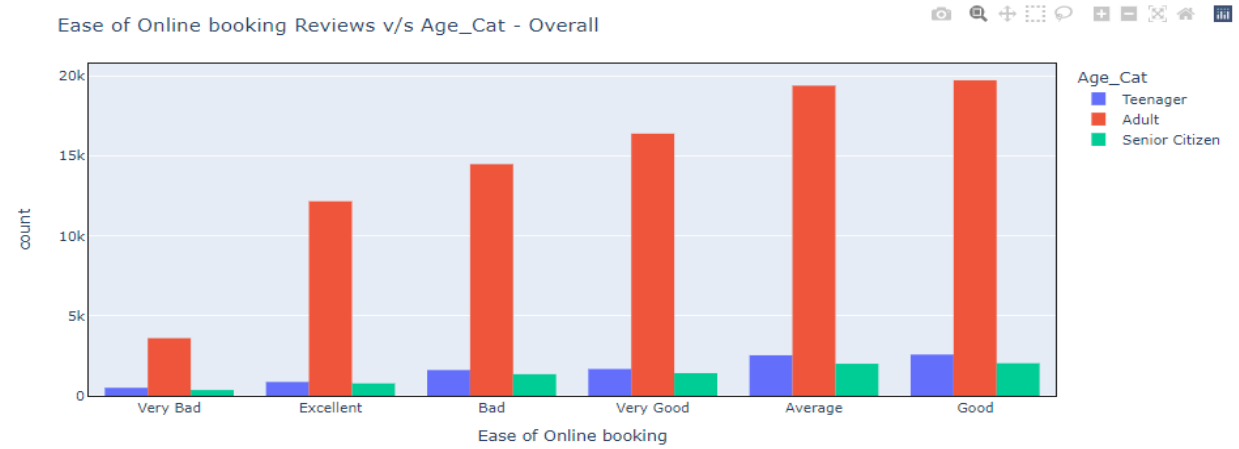
Customer Demographic:

○ Gender ○ Customer Type ○ Type of Travel ◉ Class ○ Age_Cat



Customer Demographic:

○ Gender ○ Customer Type ○ Type of Travel ○ Class ◉ Age_Cat

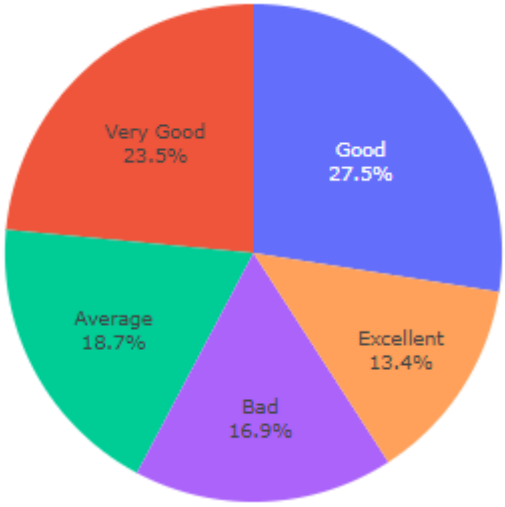


Observations:

- **Gender**: More reviews have been given by Females, but it appears that both Male and Females were happy with the "Ease of Online Booking"
- **Customer Type**: Mostly airlines loyal customers have given reviews and they are happy with the service. However, out of those close to 14.5k customers are not happy with this factor.
- **Type of Travel**: Customers travelling for business have given more reviews. Similar number of votes have come for "excellent" and "bad" rating.
- **Class**: Economy class passengers have given more reviews and are mostly good, like business or economy plus passengers. However, for economy class, the number of "bad" reviews are significantly more compared to "excellent" reviews. The same thing may not be true for business class, but number of "bad" reviews is very high.
- **Age**: As expected, adults have given more reviews and they are good with the service.

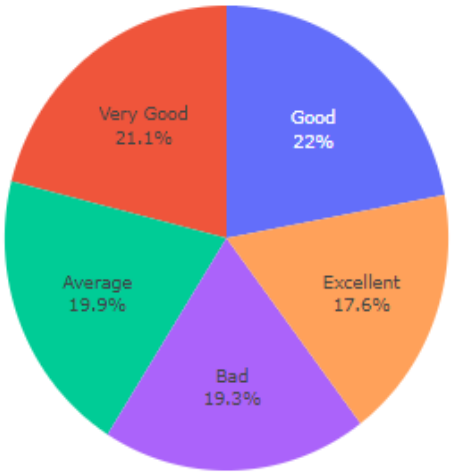
2. Gate Location:

Overall Distribution

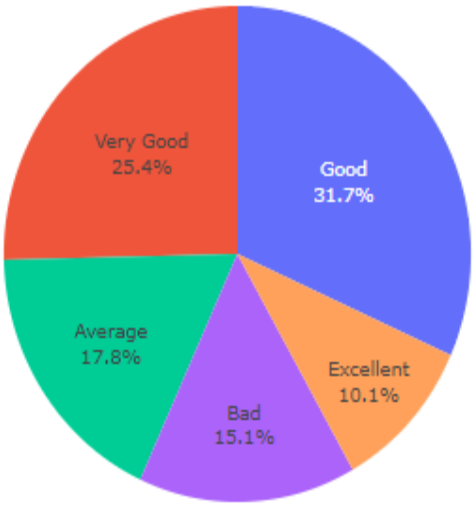


Distribution based on Customer Satisfaction

Satisfied Customers



Not Satisfied Customers



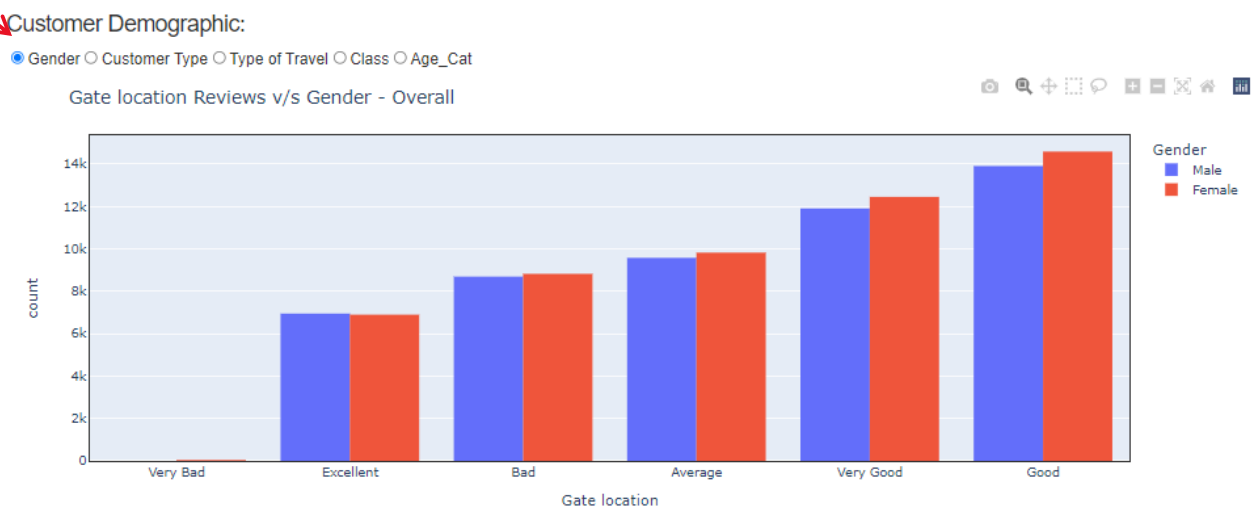
Observation:

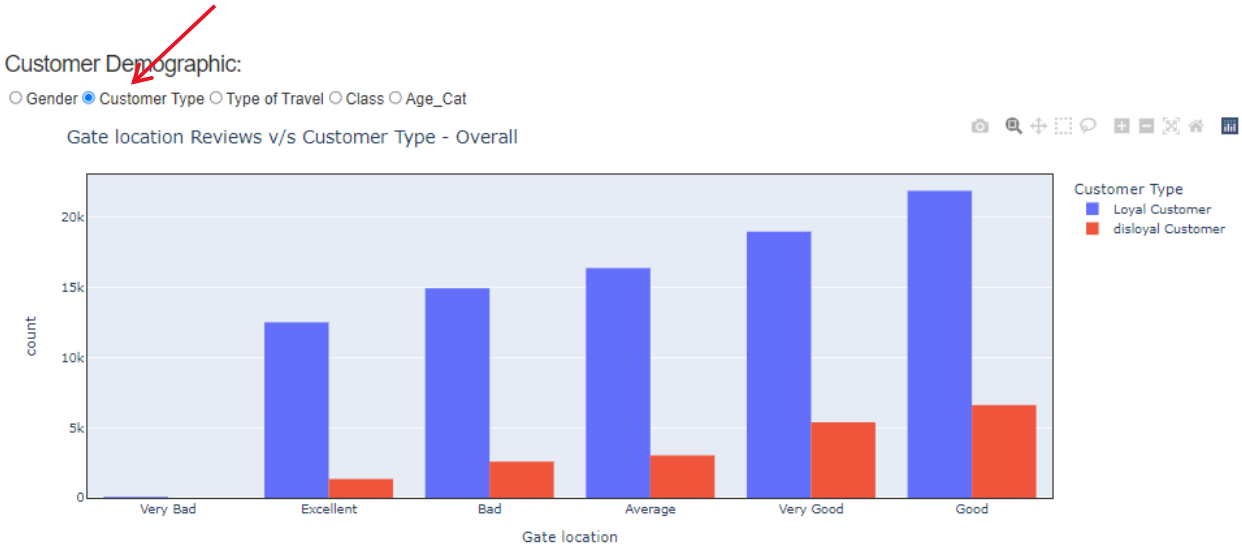
Overall customers have given a “Very Good” or a “Good” review for the gate location. However, not satisfied customers have given more **good** rating compared to customers satisfied with service.

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Final Term Project – Data Visualization

Interaction with Customer Demographics (Gender, Customer Type, Type of Travel, Class, and Age)



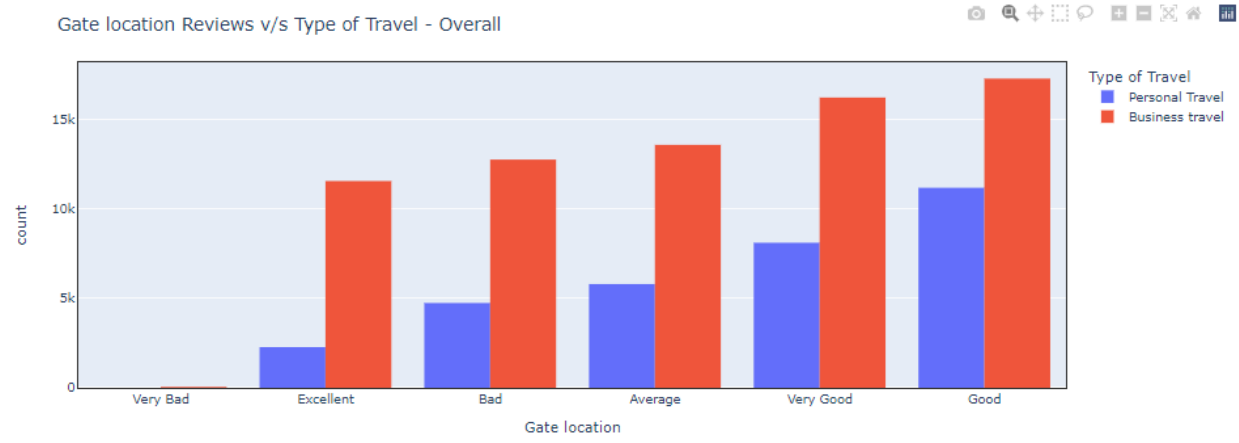


THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Final Term Project – Data Visualization

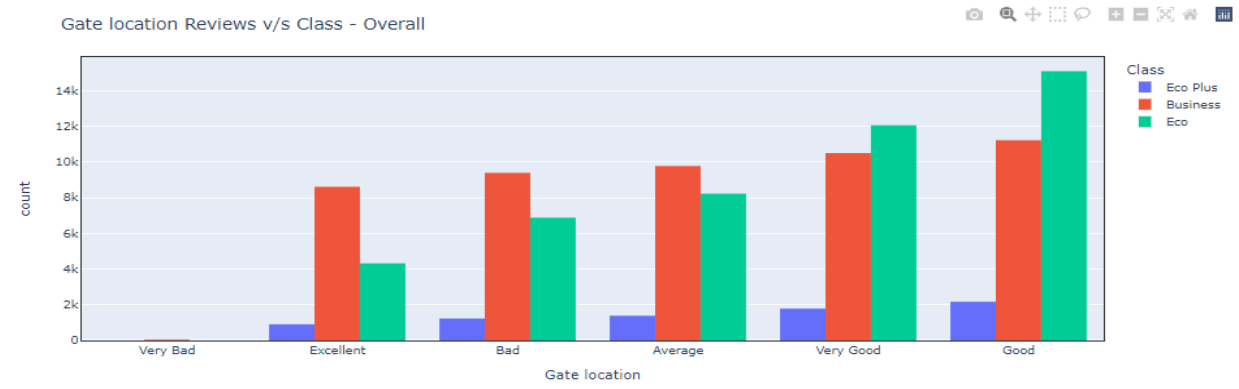
Customer Demographic:

○ Gender ○ Customer Type ◉ Type of Travel ○ Class ○ Age_Cat



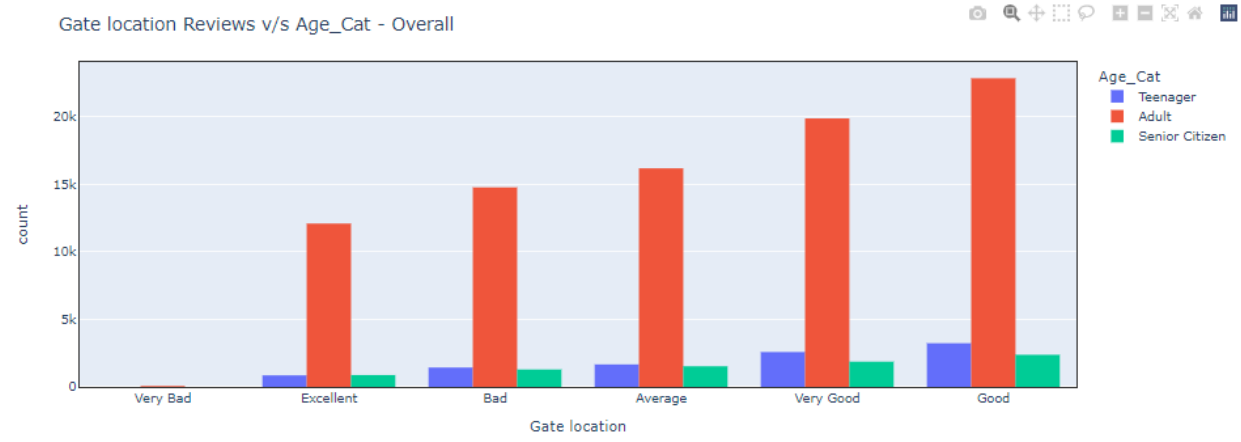
Customer Demographic:

○ Gender ○ Customer Type ○ Type of Travel ◉ Class ○ Age_Cat



Customer Demographic:

○ Gender ○ Customer Type ○ Type of Travel ○ Class ◉ Age_Cat

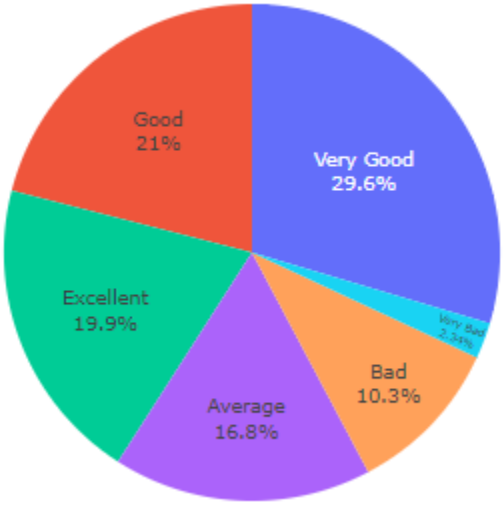


Observations:

- **Gender**: More reviews have been given by Females, but it appears that both Male and Females were happy with the "Gate Location".
- **Customer Type**: Mostly airlines loyal customers have given reviews and they are happy with the service.
- **Type of Travel**: Customers travelling for business have given more reviews and close to 13k have given a "bad" rating.
- **Class**: All the classes seem to have given largely a good review about this factor.
- **Age**: As expected, adults have given more reviews and they have given a lot of "Good" reviews but there are a lot of "bad" reviews given as well.

3. Online Boarding:

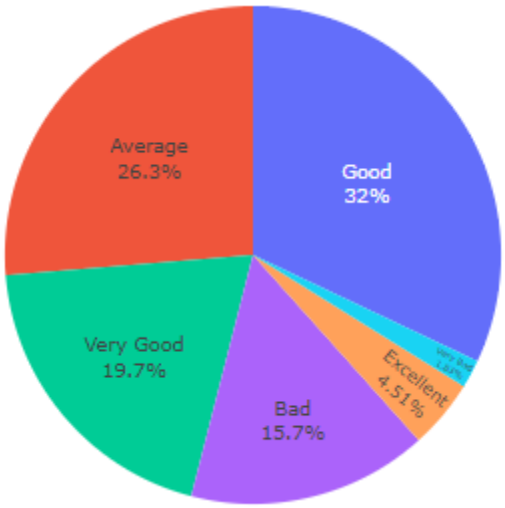
Overall Distribution



Distribution based on Customer Satisfaction

Satisfied Customers

Not Satisfied Customers



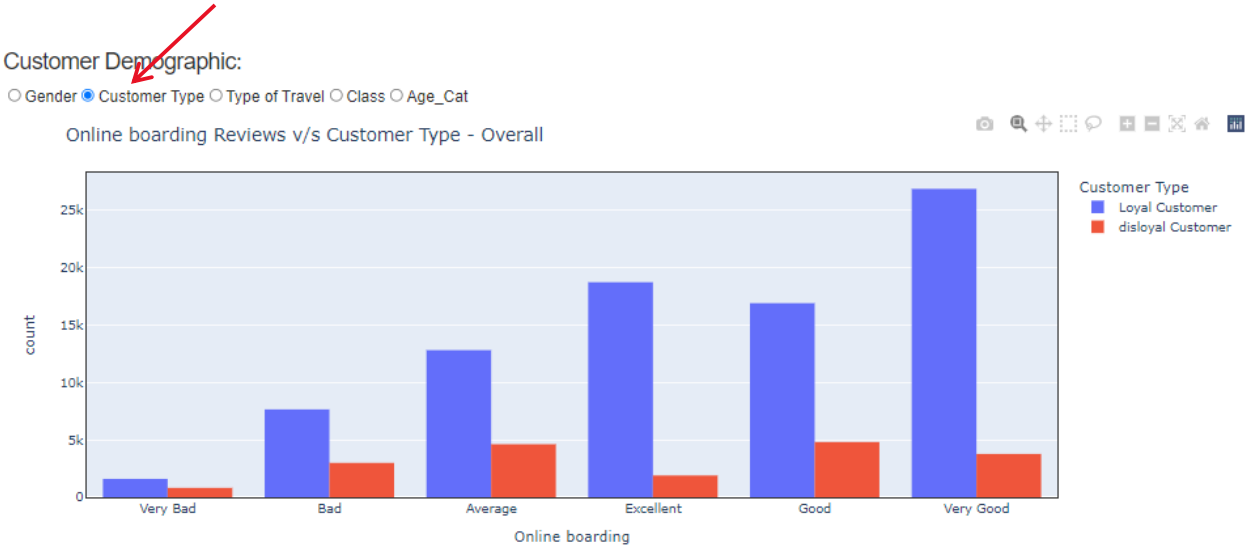
Observation:

Overall customers have mostly given a “Very Good” review for the online boarding. The satisfied customers have mainly given an “Excellent” or a “Very Good” review.

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Final Term Project – Data Visualization

Interaction with Customer Demographics (Gender, Customer Type, Type of Travel, Class, and Age)

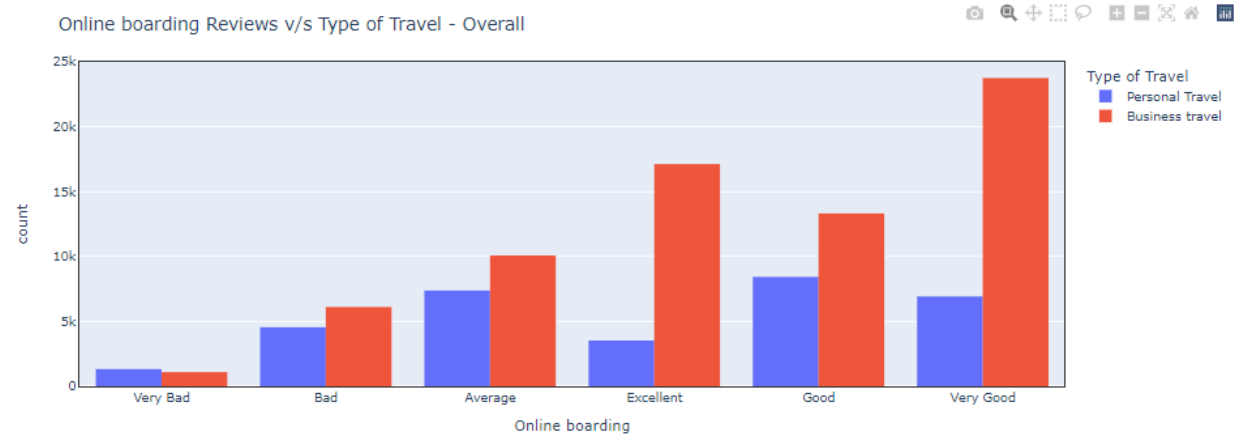


THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Final Term Project – Data Visualization

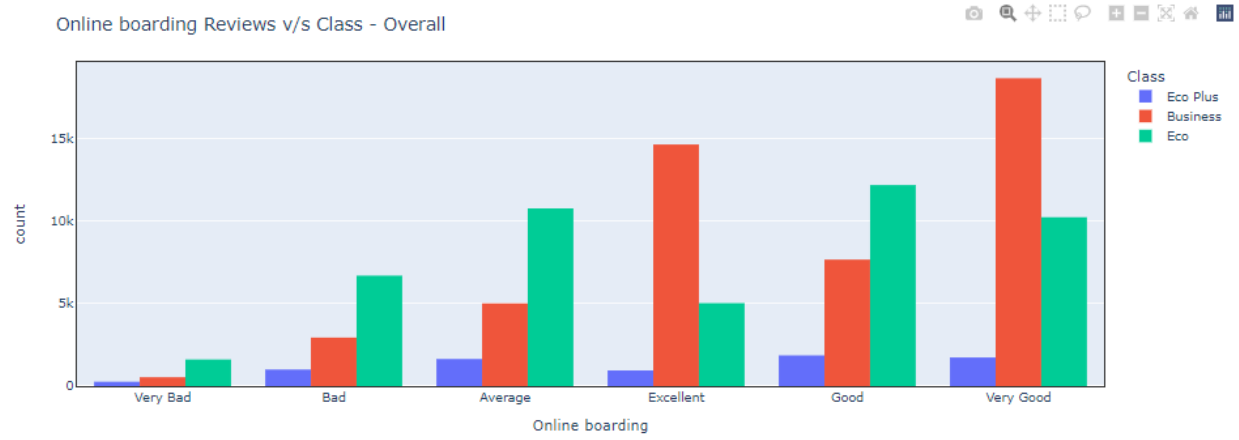
Customer Demographic:

○ Gender ○ Customer Type ◉ Type of Travel ○ Class ○ Age_Cat



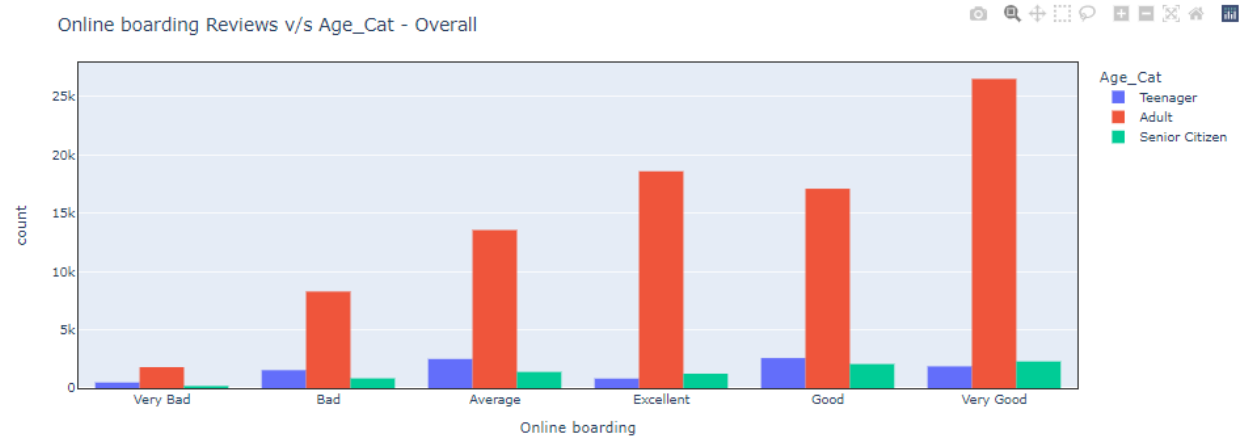
Customer Demographic:

○ Gender ○ Customer Type ○ Type of Travel ◉ Class ○ Age_Cat



Customer Demographic:

○ Gender ○ Customer Type ○ Type of Travel ○ Class ◉ Age_Cat



Observations:

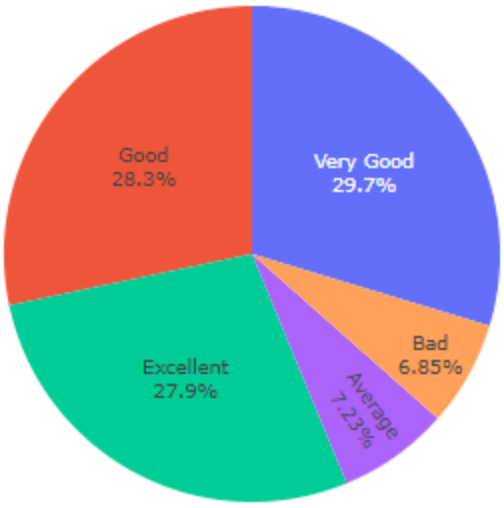
- **Gender**: Majority reviews are good. Number of males giving a "bad" rating is higher than females.
- **Customer Type**: Mostly airlines loyal customers have given reviews and they are happy with the service.
- **Type of Travel**: Customers travelling for business have given more reviews but both categories indicate they are happy with Online boarding process.
- **Class**: All the classes seem to have given largely a good review about this factor. However, economy class passengers have given more "bad" reviews compared to other class.
- **Age**: As expected, adults have given more reviews and they have given a lot of "Very Good" reviews.

4. **Check-in Service:**

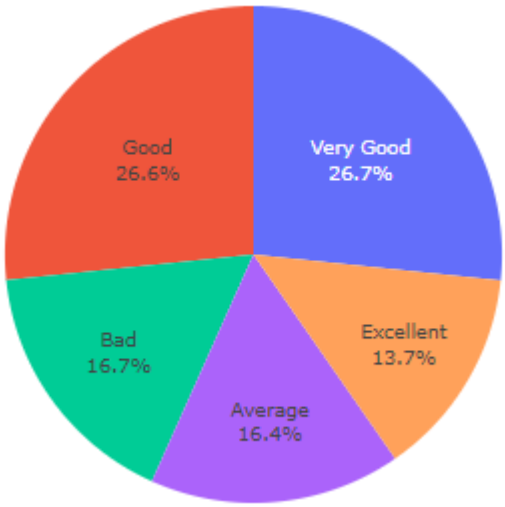
Overall Distribution

Distribution based on Customer Satisfaction

Satisfied Customers



Not Satisfied Customers



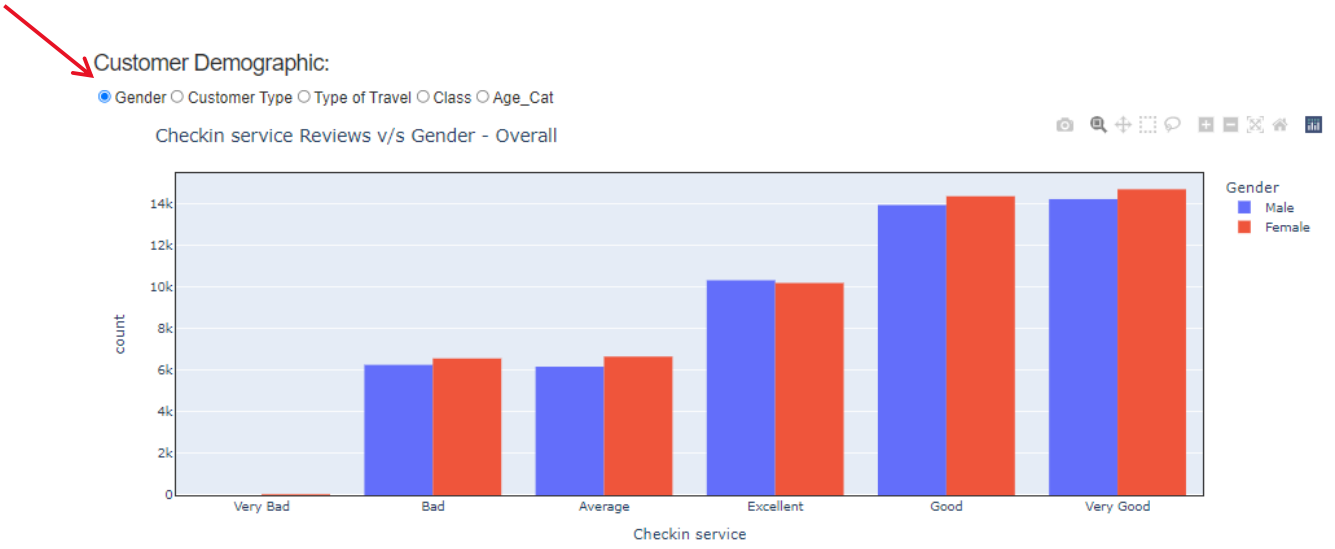
Observation:

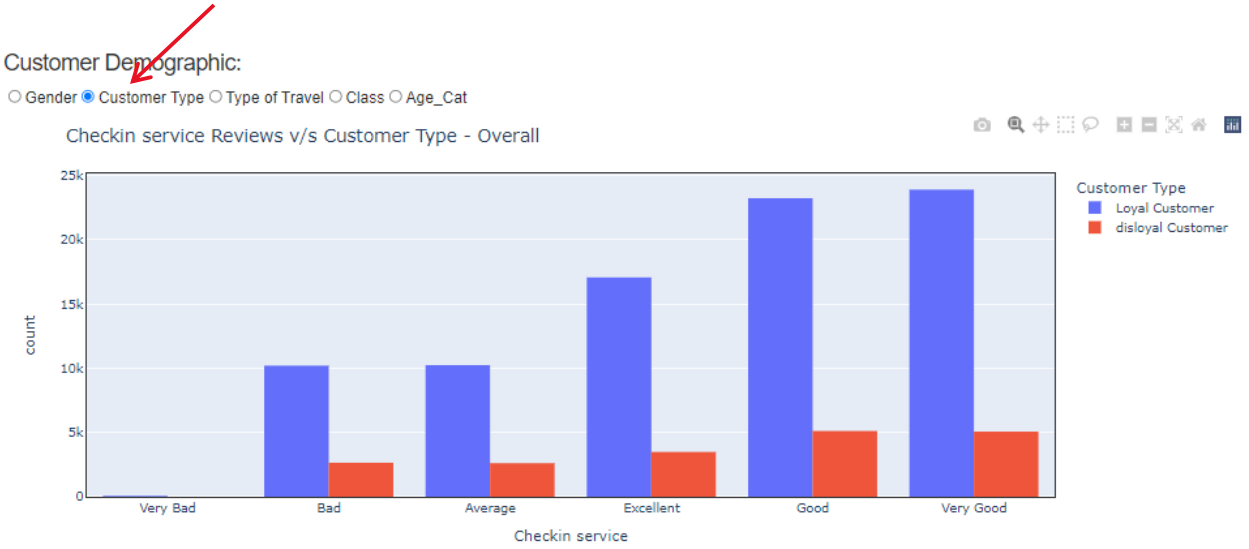
Overall customers have mostly given a “Very Good or Good” review. Also, there’s a good % of customers giving “bad” reviews to “Check-in Service”.

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Final Term Project – Data Visualization

Interaction with Customer Demographics (Gender, Customer Type, Type of Travel, Class, and Age)



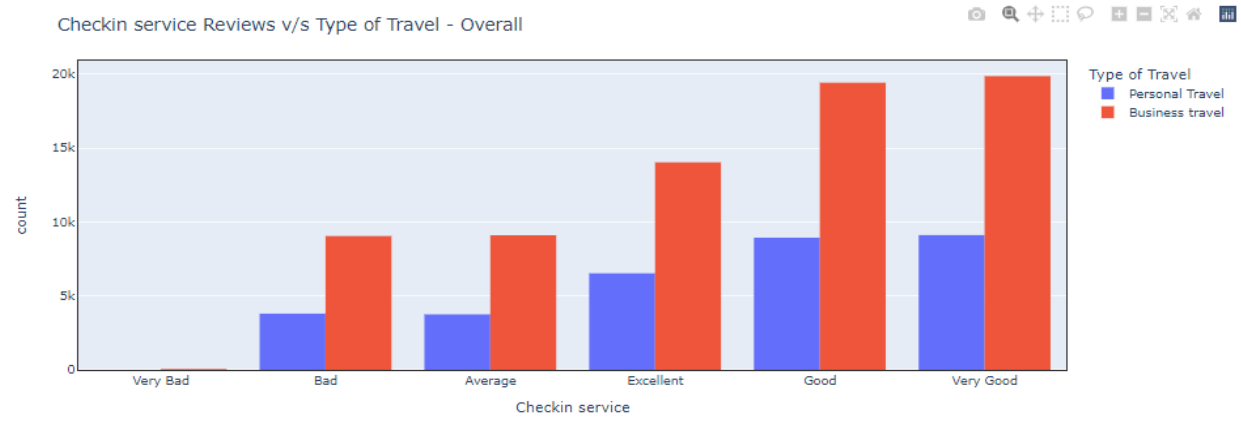


THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Final Term Project – Data Visualization

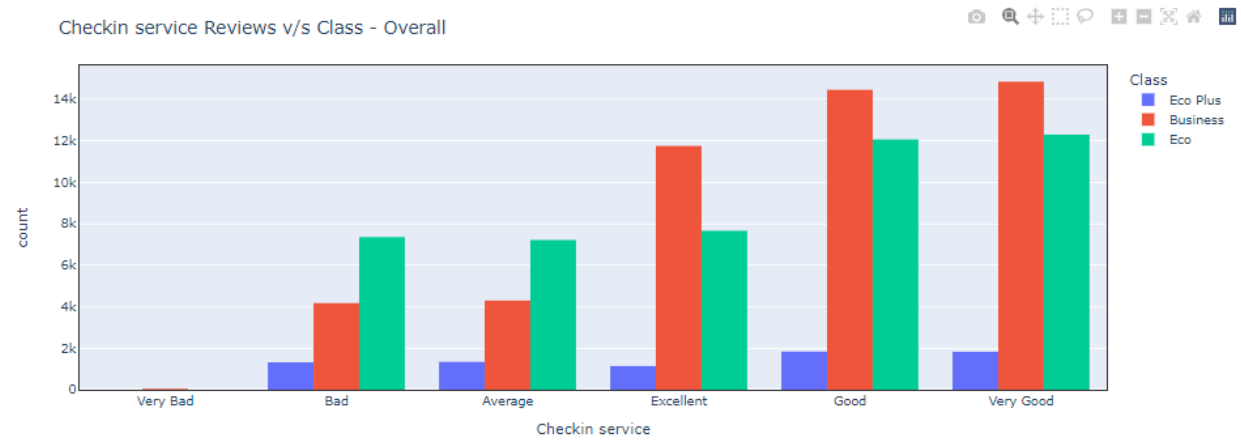
Customer Demographic:

○ Gender ○ Customer Type ◉ Type of Travel ○ Class ○ Age_Cat



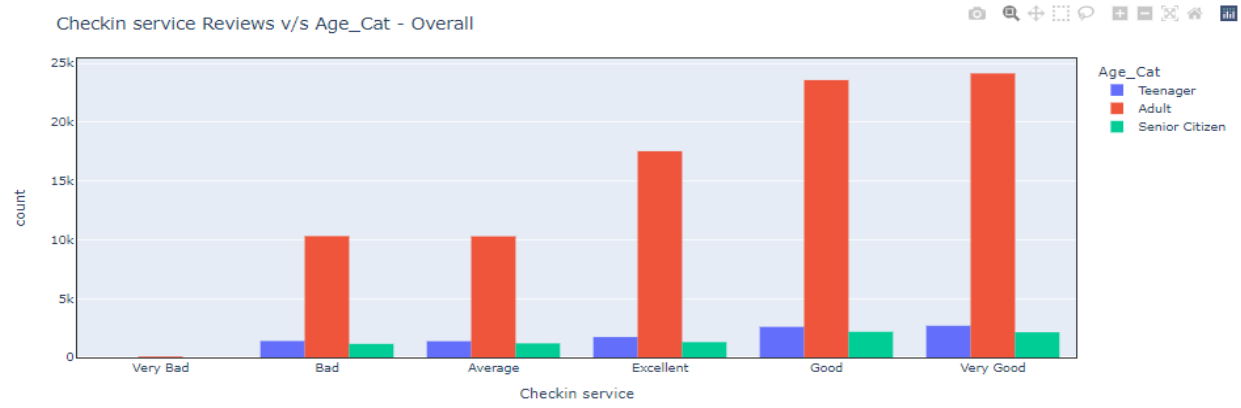
Customer Demographic:

○ Gender ○ Customer Type ○ Type of Travel ◉ Class ○ Age_Cat



Customer Demographic:

○ Gender ○ Customer Type ○ Type of Travel ○ Class ◉ Age_Cat

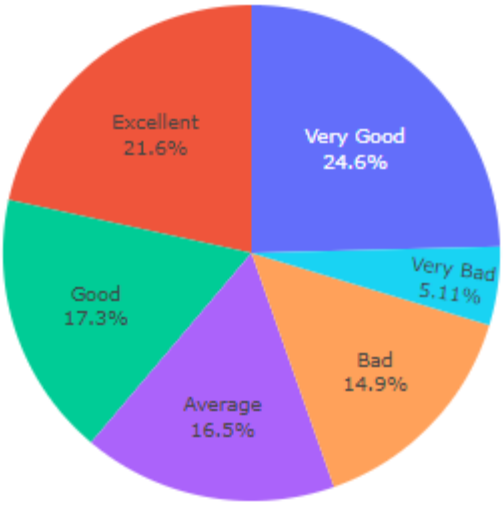


Observations:

- **Gender**: Both genders have given high rating to the "Check-in Service".
- **Customer Type**: Mostly airlines loyal customers have given reviews and they are happy with the service.
- **Type of Travel**: Customers travelling for business have given more reviews and have higher number of "bad" reviews compared to personal travel customers.
- **Class**: All the classes seem to have given largely a good review about this factor. However, economy class passengers have given more "bad" reviews compared to other class.
- **Age**: As expected, adults have given more reviews and they have given a lot of "Very Good" reviews.

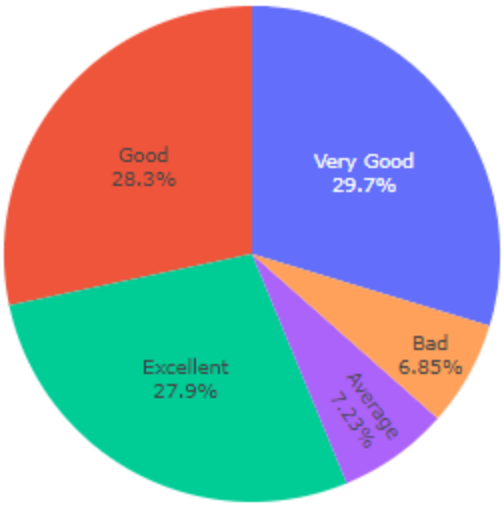
5. **Departure/Arrival Time Convenient:**

Overall Distribution

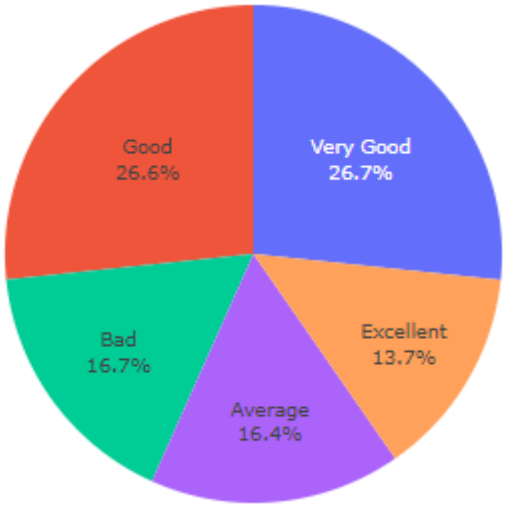


Distribution based on Customer Satisfaction

Satisfied Customers



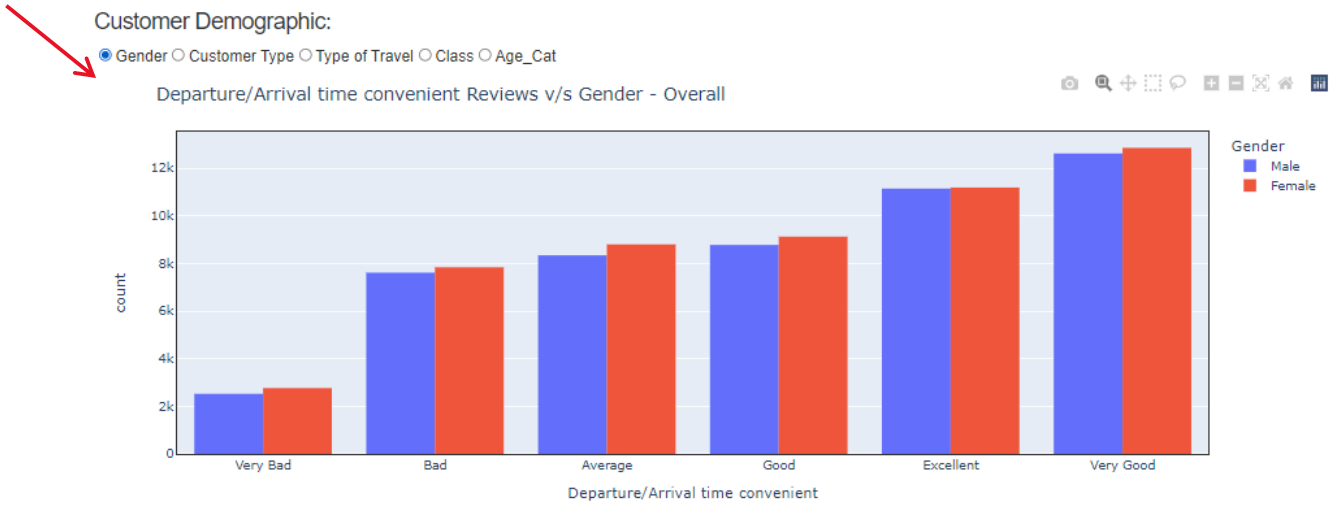
Not Satisfied Customers

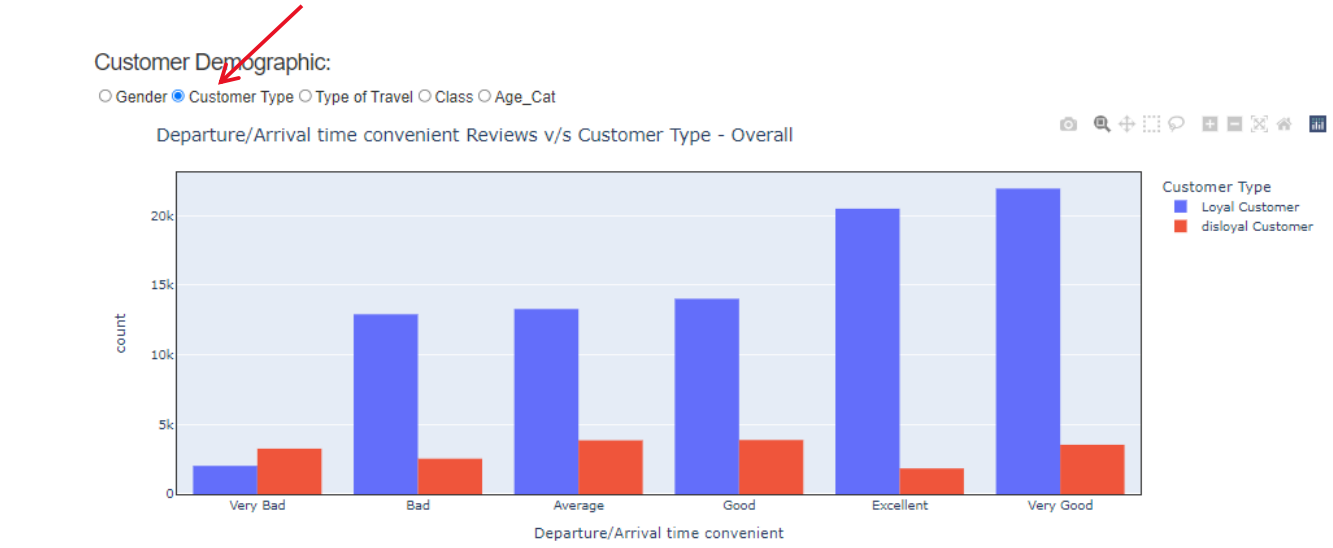


Observation:

Although customers, overall, are happy with the timings of the flight, close to 20% of customers have found the flight timings to be quite bad.

Interaction with Customer Demographics (Gender, Customer Type, Type of Travel, Class, and Age)



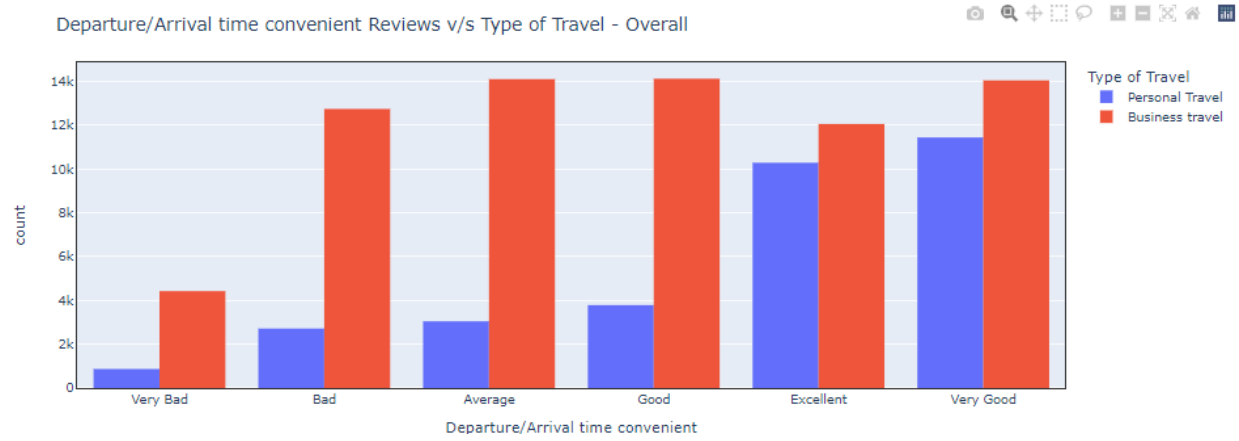


THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Final Term Project – Data Visualization

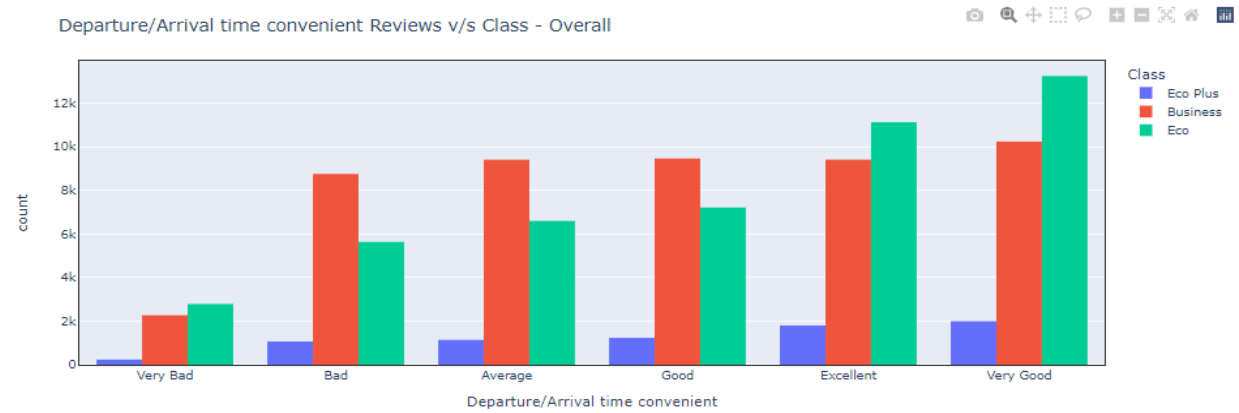
Customer Demographic:

○ Gender ○ Customer Type ◉ Type of Travel ○ Class ○ Age_Cat



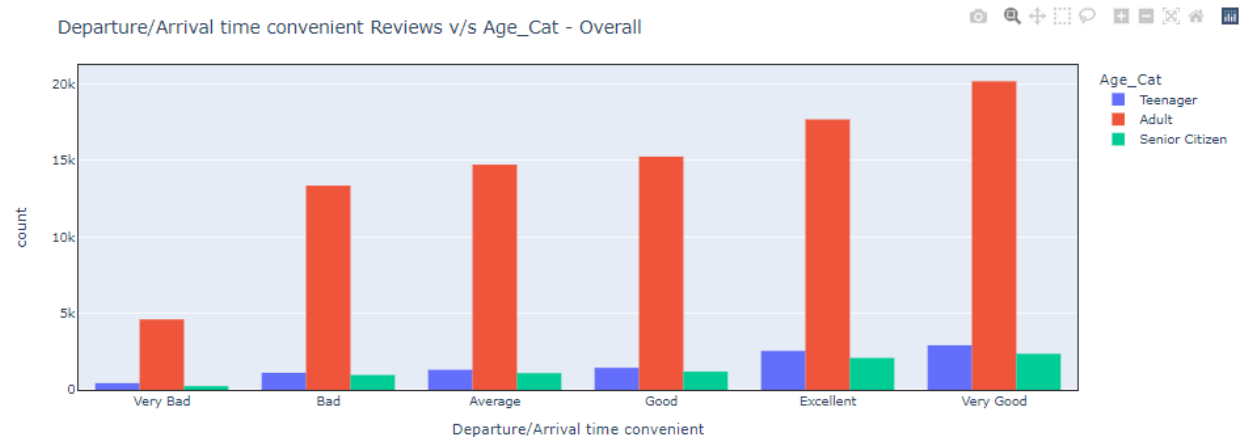
Customer Demographic:

○ Gender ○ Customer Type ○ Type of Travel ◉ Class ○ Age_Cat



Customer Demographic:

○ Gender ○ Customer Type ○ Type of Travel ○ Class ◉ Age_Cat

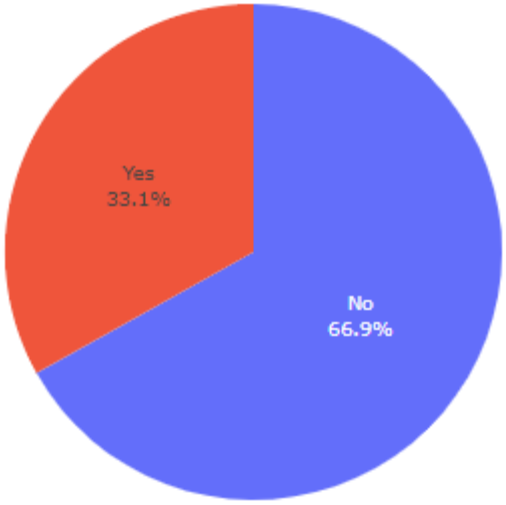


Observations:

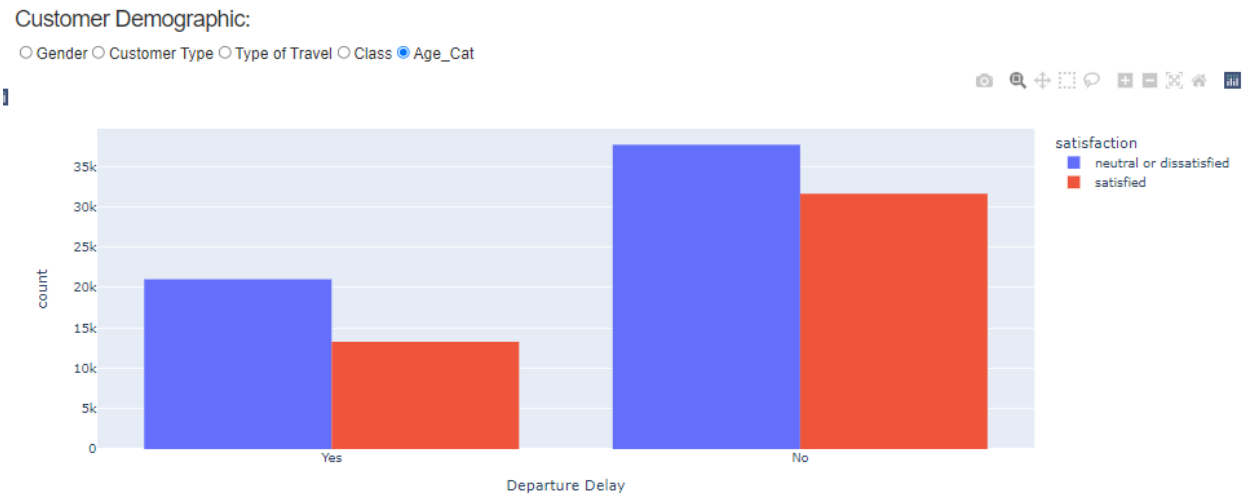
- **Gender**: Overall, both males and females are happy with the timing of the flights, but there are customers who have given "very bad" reviews about the timing.
- **Customer Type**: Mostly airlines loyal customers have given reviews and they are happy with the service. But there's a good count of customers giving a bad review.
- **Type of Travel**: Customers travelling for business have given more reviews and have higher number of "bad" reviews compared to personal travel customers.
- **Class**: All the classes seem to have given largely a good review about this factor. However, business class passengers have given more "bad" reviews compared to other class.
- **Age**: As expected, adults have given more reviews and they have given a lot of "bad" reviews.

6. Departure Delay:

Overall Distribution



Distribution based on Customer Satisfaction



NOTE: DEPARTURE DELAY WILL ONLY BE SHOWN w.r.t SATISFACTION

Observation:

One-third of the flights have had delayed departure. However, based on customer satisfaction, even if there was no delayed flight, many customers were still not satisfied.

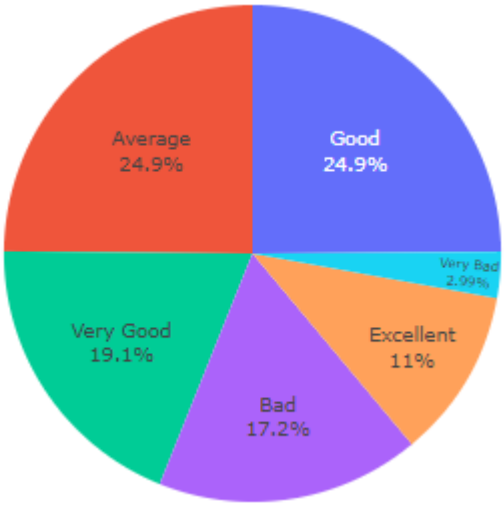
THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Final Term Project – Data Visualization

On-Board Factors

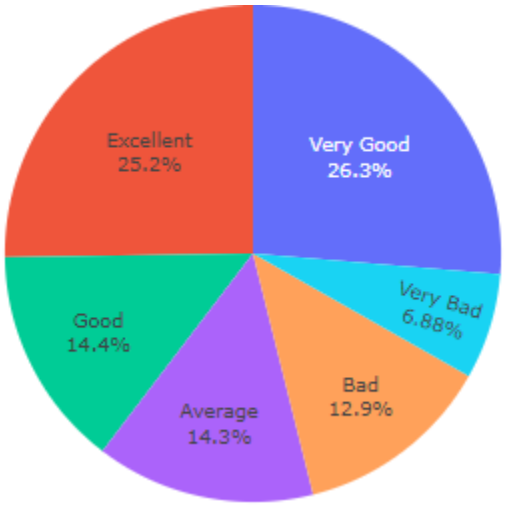
1. Inflight Wi-fi service:

Overall Distribution



Distribution based on Customer Satisfaction

Satisfied Customers

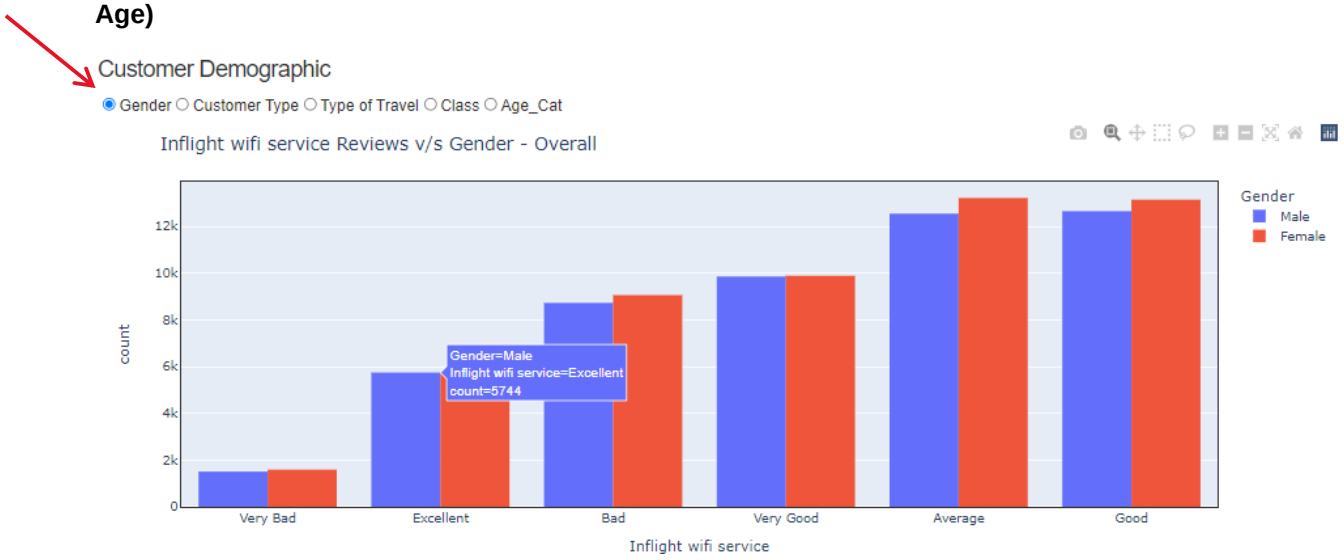


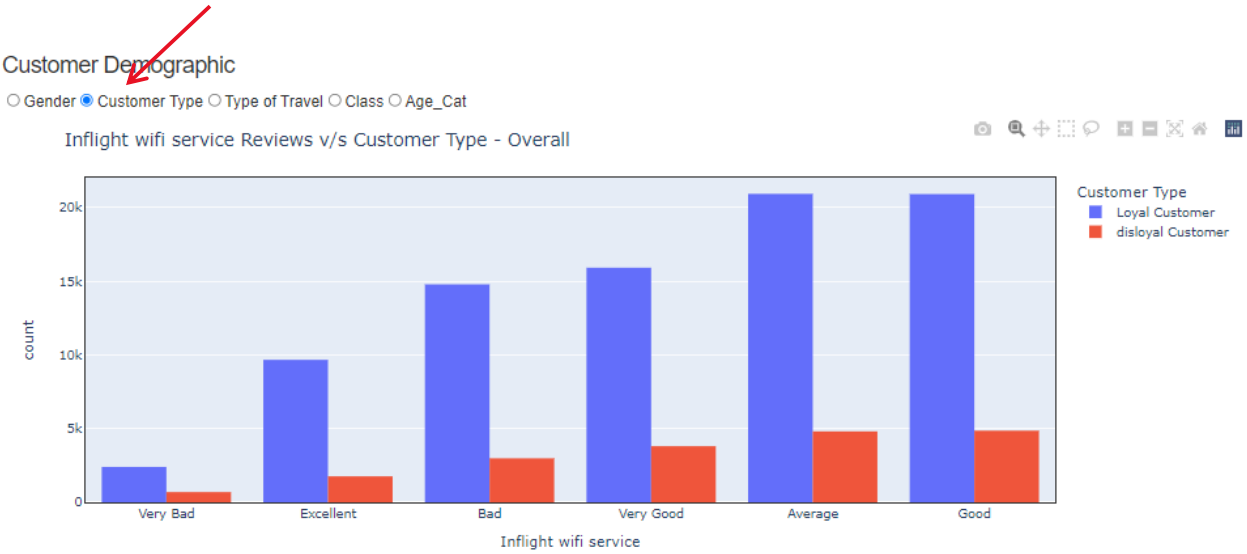
Not Satisfied Customers

Observation:

Overall customers have given a “Average” or a “Good” review. However, satisfied customers have given more Excellent rating compared to customers not satisfied with service.

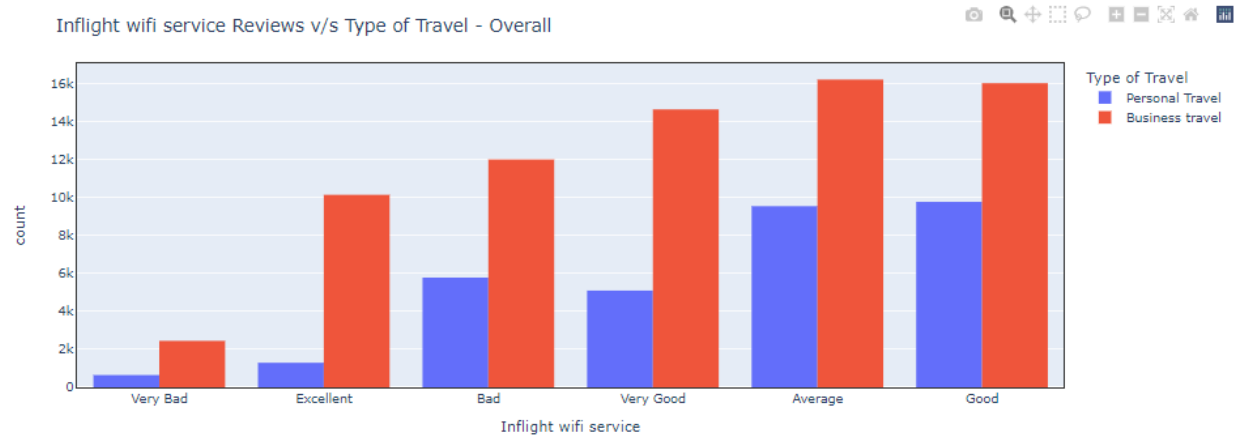
Interaction with Customer Demographics (Gender, Customer Type, Type of Travel, Class, and Age)





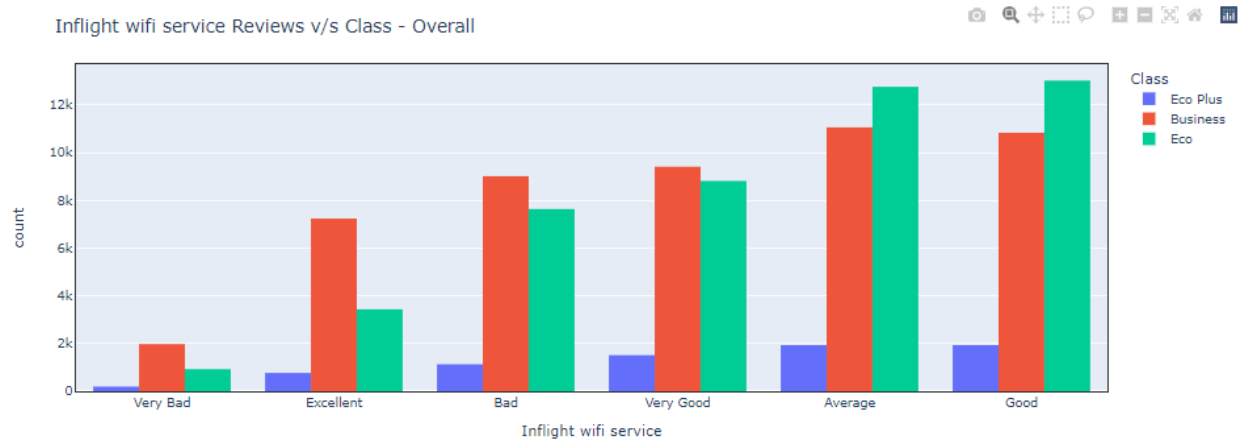
Customer Demographic

○ Gender ○ Customer Type ◉ Type of Travel ○ Class ○ Age_Cat



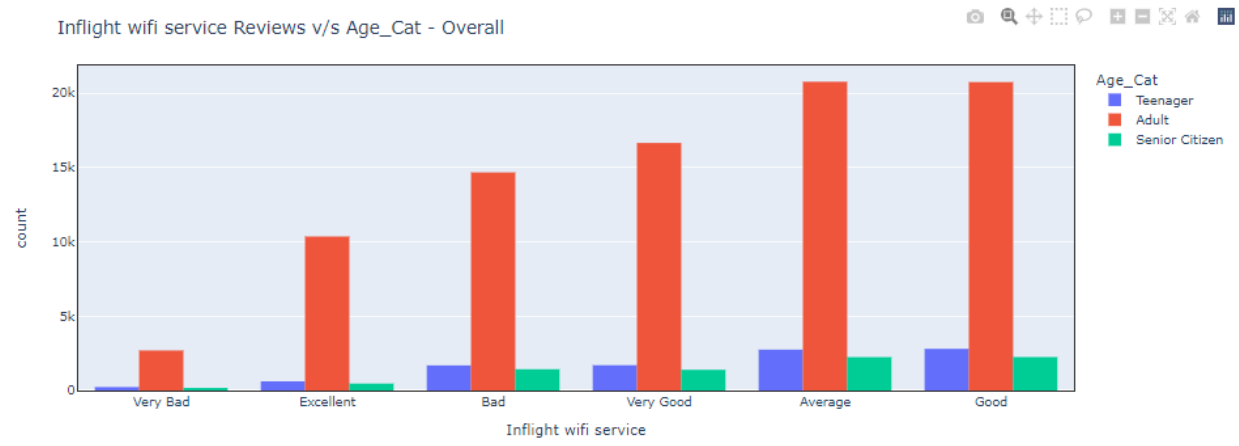
Customer Demographic

○ Gender ○ Customer Type ○ Type of Travel ◉ Class ○ Age_Cat



Customer Demographic

○ Gender ○ Customer Type ○ Type of Travel ○ Class ◉ Age_Cat

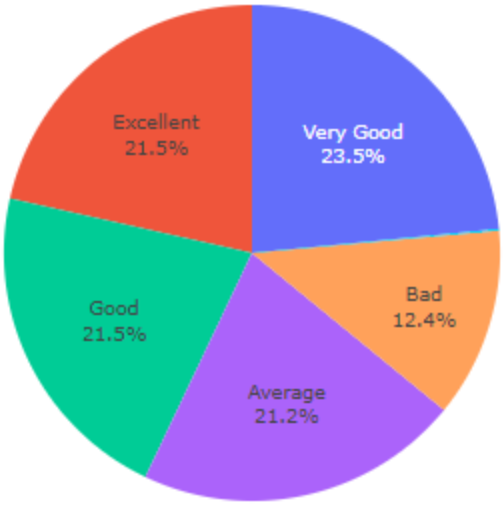


Observations:

- **Gender**: More reviews have been given by Females, but it appears that both Male and Females were happy with the service.
- **Customer Type**: Mostly airlines loyal customers have given reviews and they are happy with the service. However, out of those close to 14k customers are not happy with this factor.
- **Type of Travel**: Customers travelling for business have given more reviews. Wi-fi service has been rated as “bad” more than it was rated “excellent”.
- **Class**: Economy class passengers have given more reviews and have mostly given an “Average” review about the wi-fi. Business class passengers have given more “bad” reviews than any other class.
- **Age**: As expected, adults have given lot of review reviews and more adults have categorized Inflight wi-fi as “bad” than “excellent”.

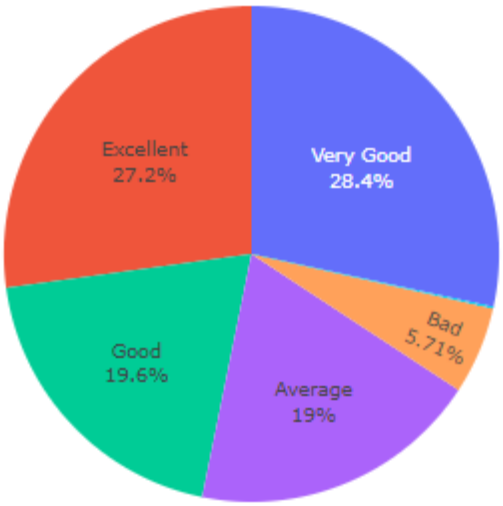
2. Food and Drink:

Overall Distribution

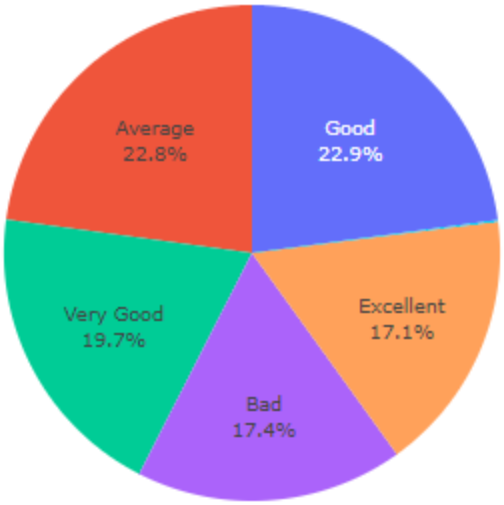


Distribution based on Customer Satisfaction

Satisfied Customers



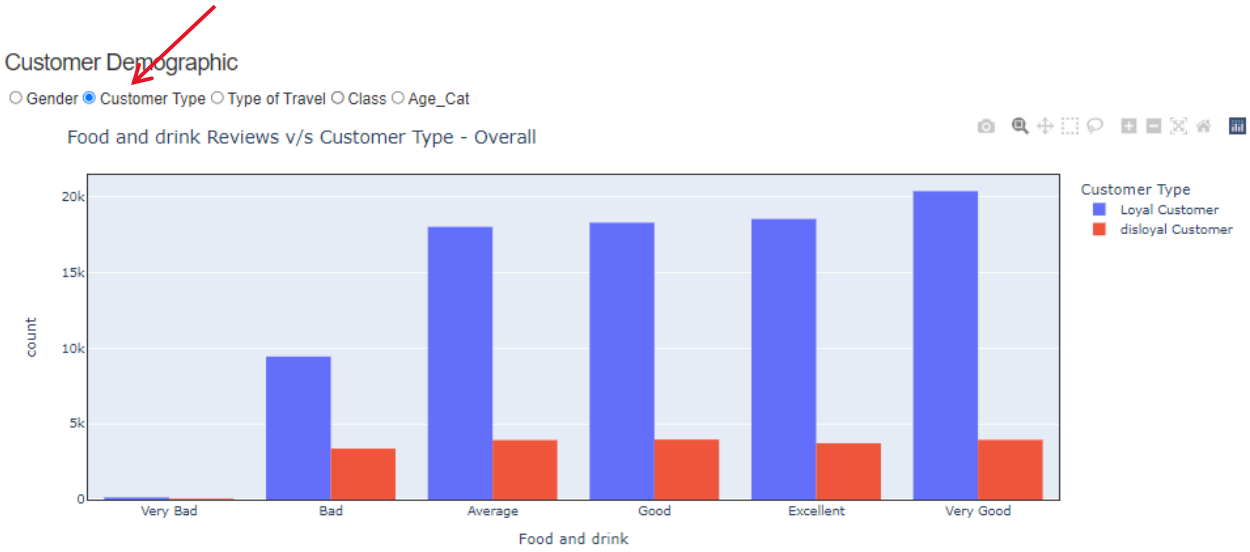
Not Satisfied Customers



Observation:

Overall customers have given extremely high rating for food and drinks. However, satisfied customers have given more bad reviews rating compared to customers not satisfied with service.

Interaction with Customer Demographics (Gender, Customer Type, Type of Travel, Class, and Age)

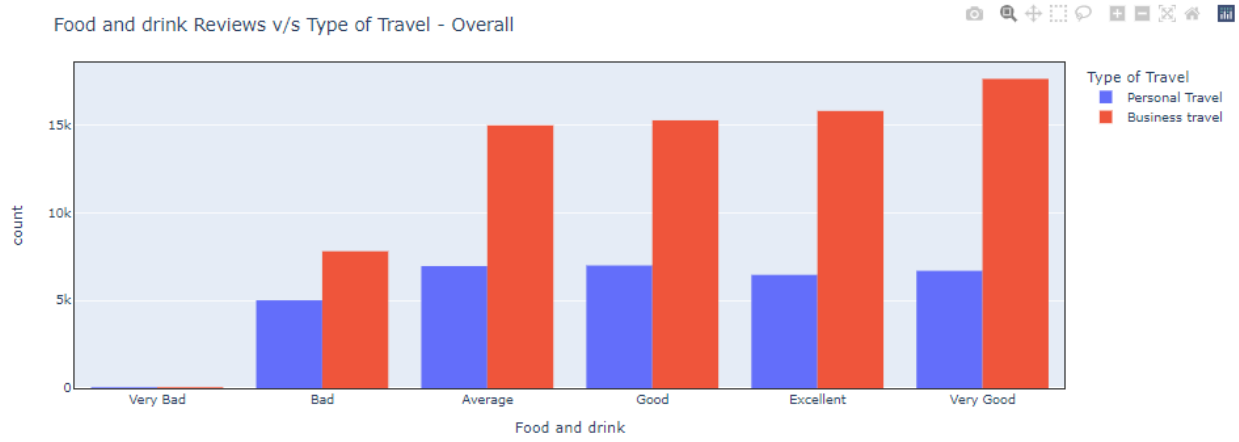


THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Final Term Project – Data Visualization

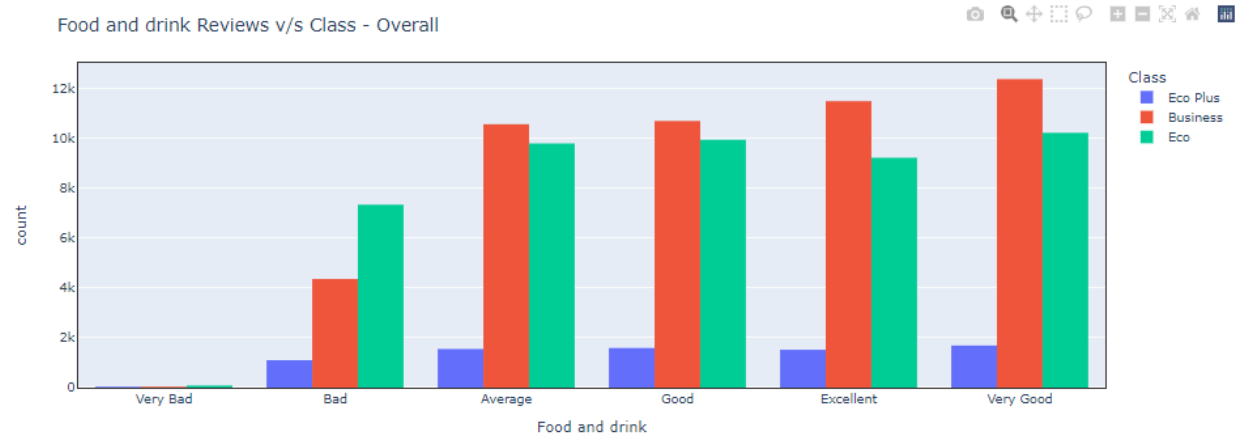
Customer Demographic

○ Gender ○ Customer Type ◉ Type of Travel ○ Class ○ Age_Cat



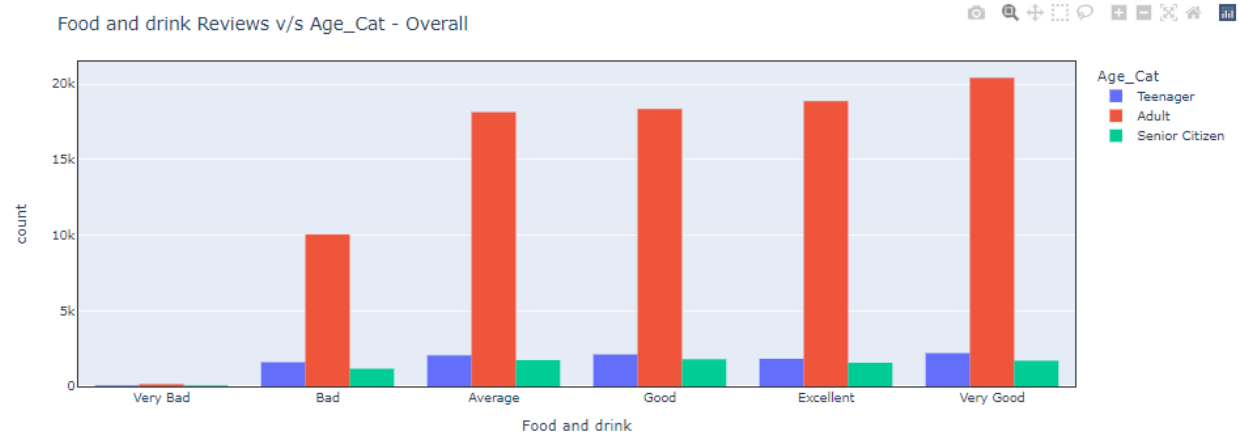
Customer Demographic

○ Gender ○ Customer Type ○ Type of Travel ◉ Class ○ Age_Cat



Customer Demographic

○ Gender ○ Customer Type ○ Type of Travel ○ Class ◉ Age_Cat

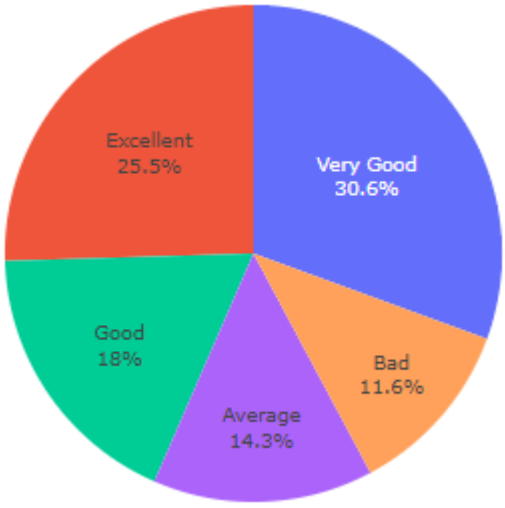


Observations:

- **Gender**: Both Male and Females were happy with the food and drink offered by airline.
- **Customer Type**: Mostly airlines loyal customers have given reviews and they are happy with the service.
- **Type of Travel**: Customers travelling for business have given more reviews. Wi-fi service has been rated as “very good” more than it was rated “bad”.
- **Class**: Business class passengers have given more reviews and have mostly given an “excellent” review about the food and drinks. Economy class passengers have given more “bad” reviews than any other class.
- **Age**: As expected, adults have given lot of review reviews and more adults have categorized the food and drinks as “excellent.”

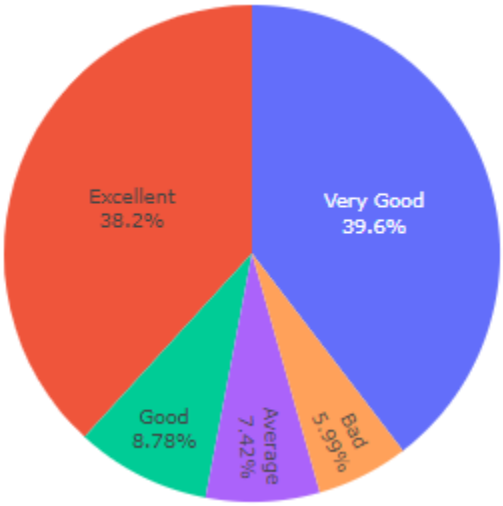
3. **Seat Comfort:**

Overall Distribution

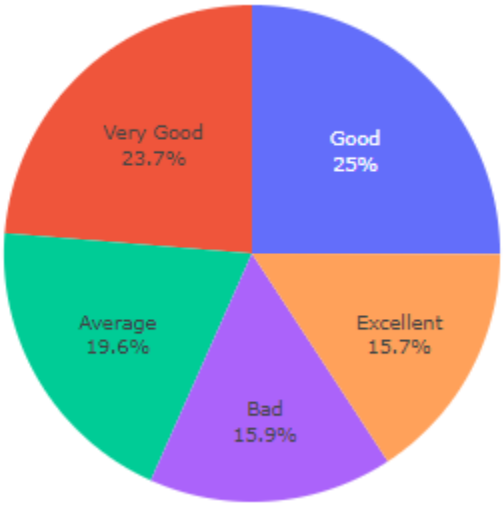


Distribution based on Customer Satisfaction

Satisfied Customers



Not Satisfied Customers



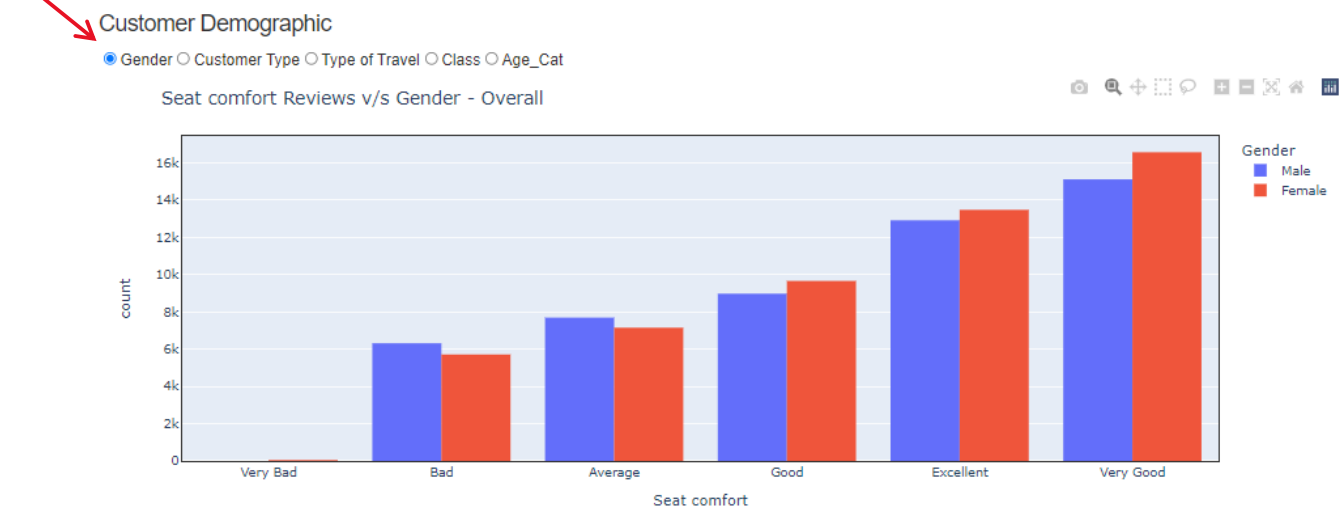
Observation:

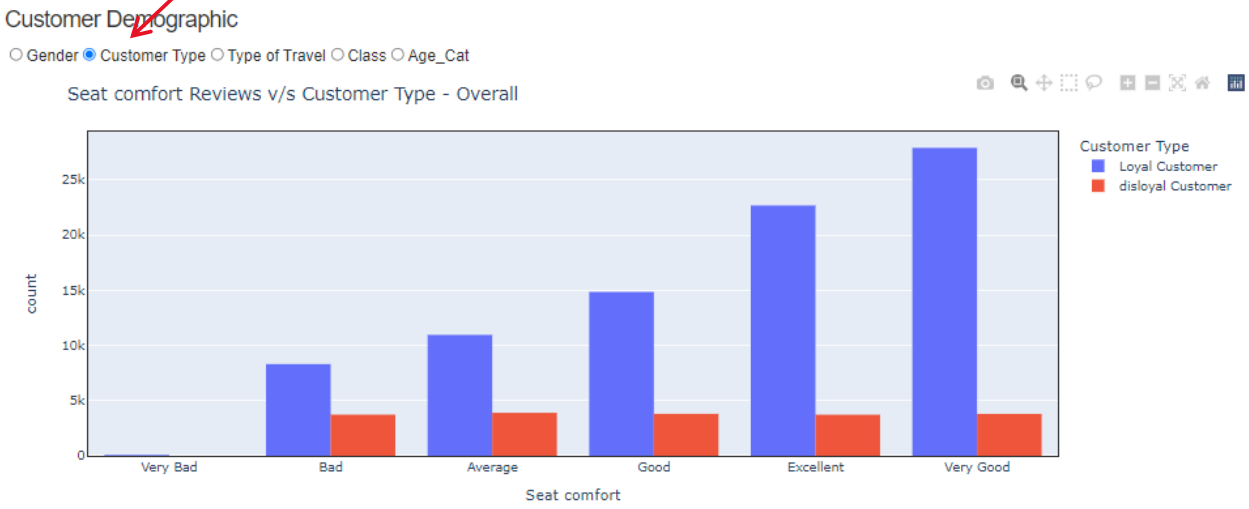
Overall customers have given extremely high rating for seat comfort.

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Final Term Project – Data Visualization

Interaction with Customer Demographics (Gender, Customer Type, Type of Travel, Class, and Age)



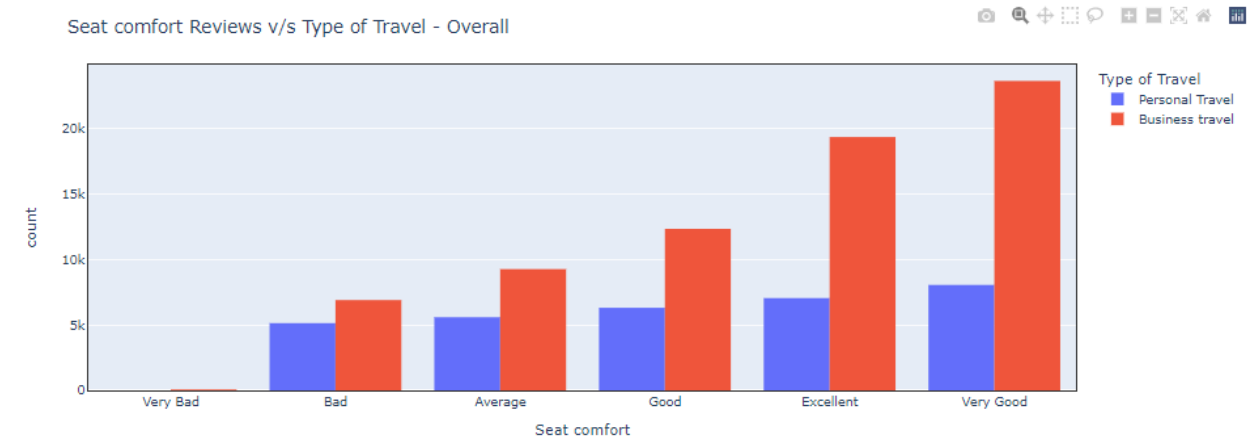


THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Final Term Project – Data Visualization

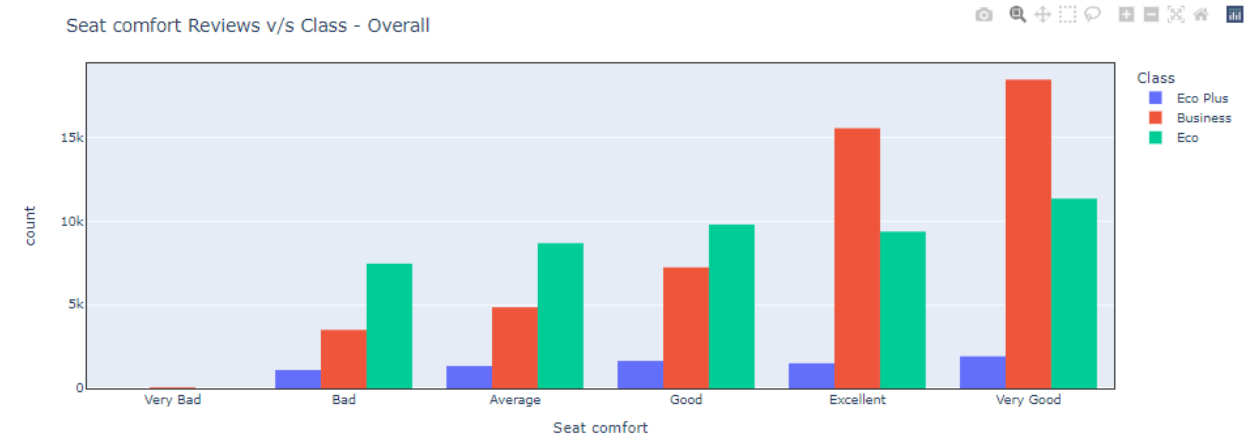
Customer Demographic

Gender | Customer Type | Type of Travel | Class | Age_Cat



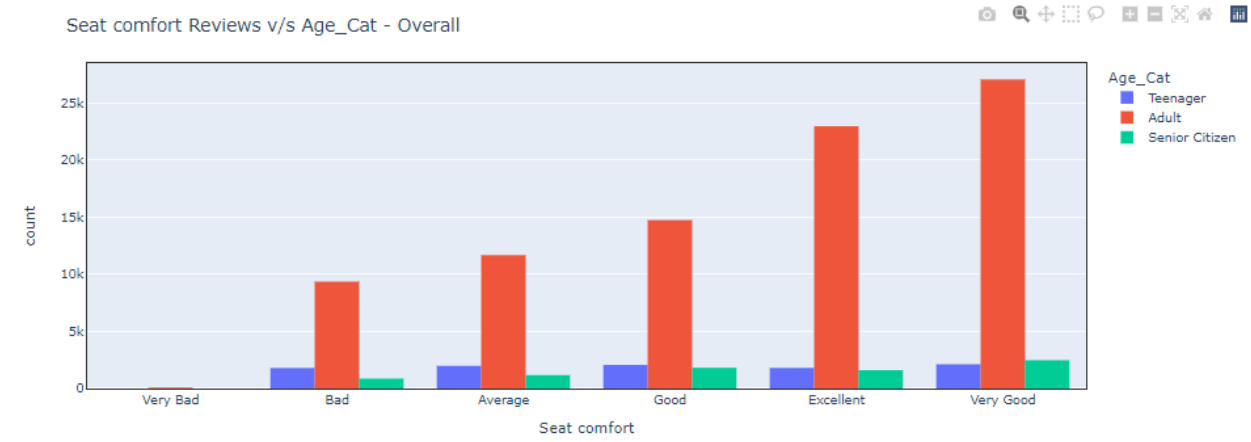
Customer Demographic

Gender | Customer Type | Type of Travel | Class | Age_Cat



Customer Demographic

Gender | Customer Type | Type of Travel | Class | Age_Cat



Observations:

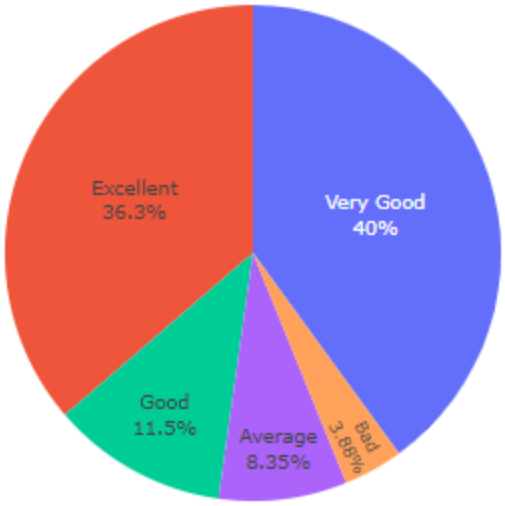
- **Gender**: Both Male and Females were happy with the seat comfort offered by airline. Males have given more bad reviews than females.
- **Customer Type**: Mostly airlines loyal customers have given reviews and they are happy with the seat comfort.
- **Type of Travel**: Customers travelling for business have given more reviews. Seat comfort has been rated as “very good” more than it was rated “bad”.
- **Class**: Seat comfort for all class passengers seems to be good.
- **Age**: As expected, adults have given lot of review reviews and more adults have categorized the seat comfort as “Very Good.”

4. **Inflight Entertainment:**

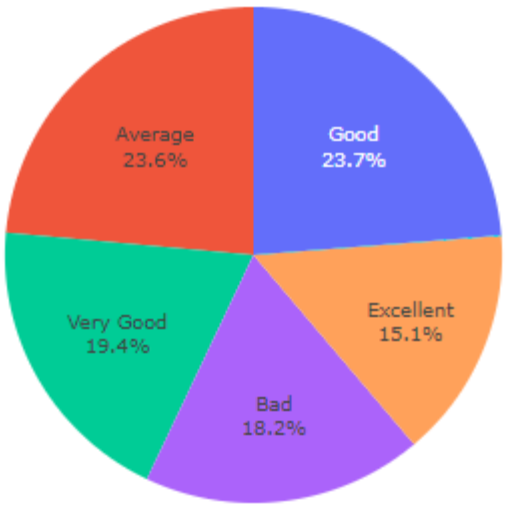
Overall Distribution

Distribution based on Customer Satisfaction

Satisfied Customers



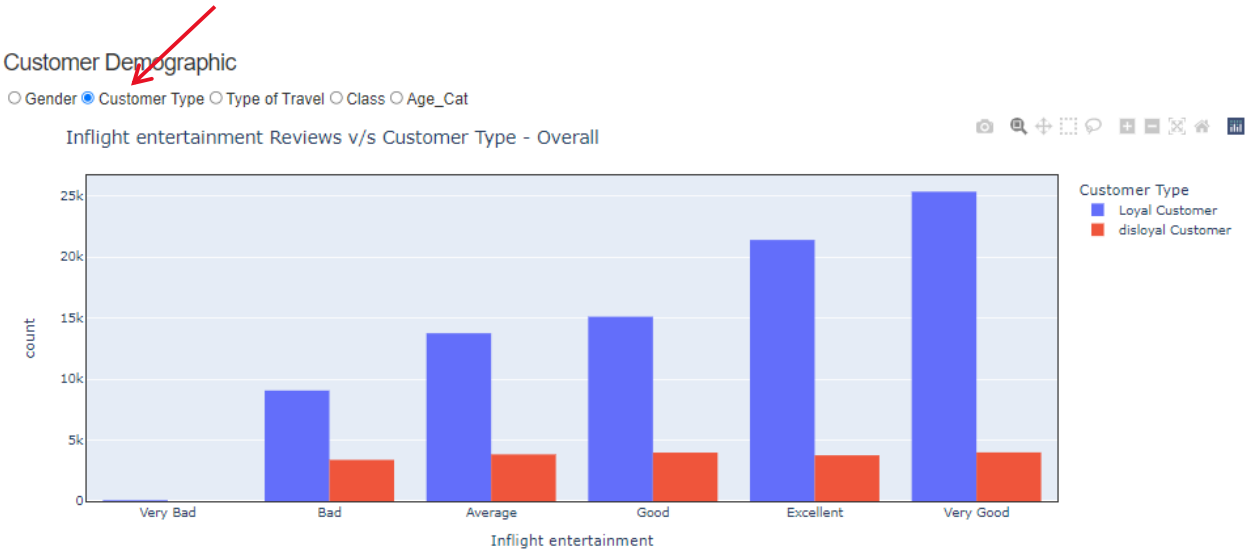
Not Satisfied Customers



Observation:

Overall customers have given extremely high rating for inflight entertainment. Lot of not-satisfied customers have rated the inflight entertainment as “bad”

Interaction with Customer Demographics (Gender, Customer Type, Type of Travel, Class, and Age)

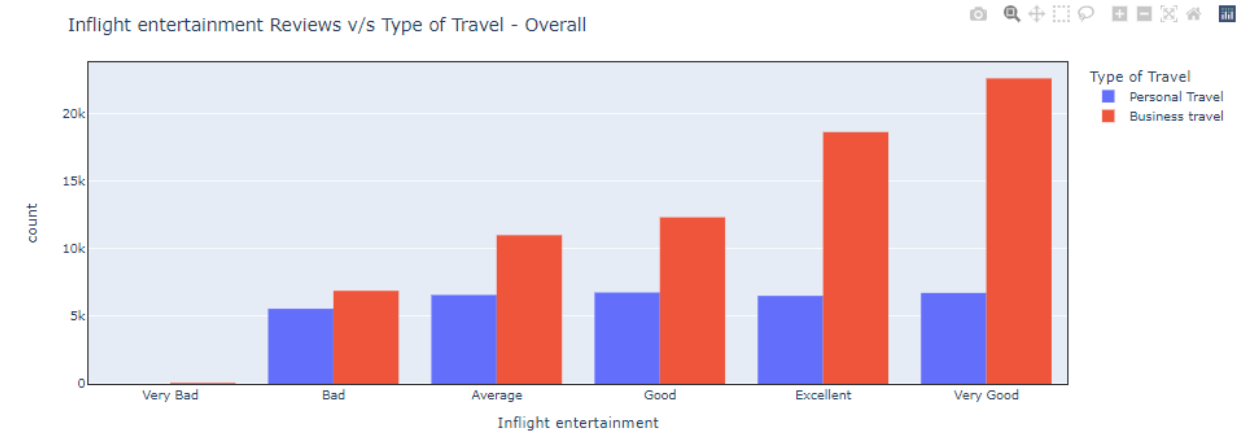


THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Final Term Project – Data Visualization

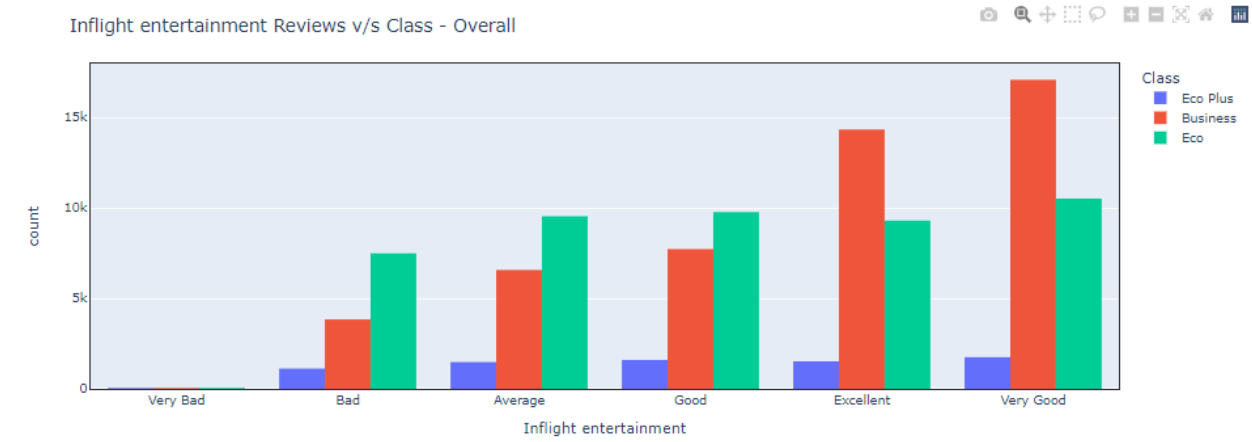
Customer Demographic

Gender Customer Type Type of Travel Class Age_Cat



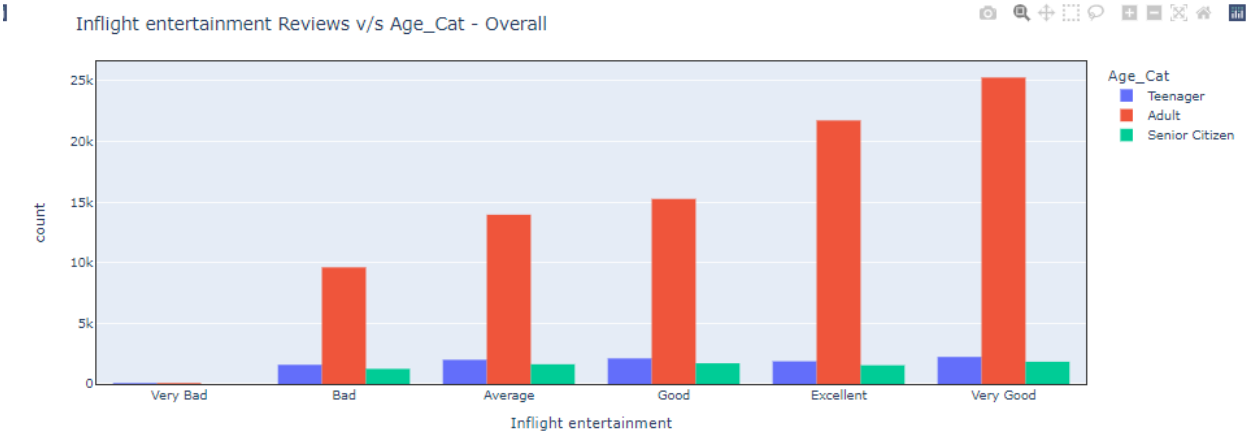
Customer Demographic

Gender Customer Type Type of Travel Class Age_Cat



Customer Demographic

Gender Customer Type Type of Travel Class Age_Cat

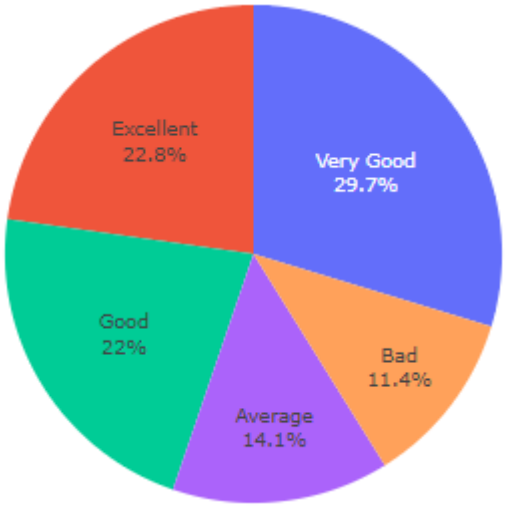


Observations:

- **Gender**: Both Male and Females were happy with the inflight entertainment offered by airline. Females have given more bad reviews than males.
- **Customer Type**: Mostly airlines loyal customers have given reviews and they are happy with the entertainment.
- **Type of Travel**: Customers travelling for business have given more reviews. Inflight entertainment has been rated as “very good” more than it was rated “bad”.
- **Class**: Passengers from all classes have rated the service highly.
- **Age**: As expected, adults have given lot of review reviews and more adults have categorized the entertainment as “Very Good.”

5. **On-board service:**

Overall Distribution



Distribution based on Customer Satisfaction

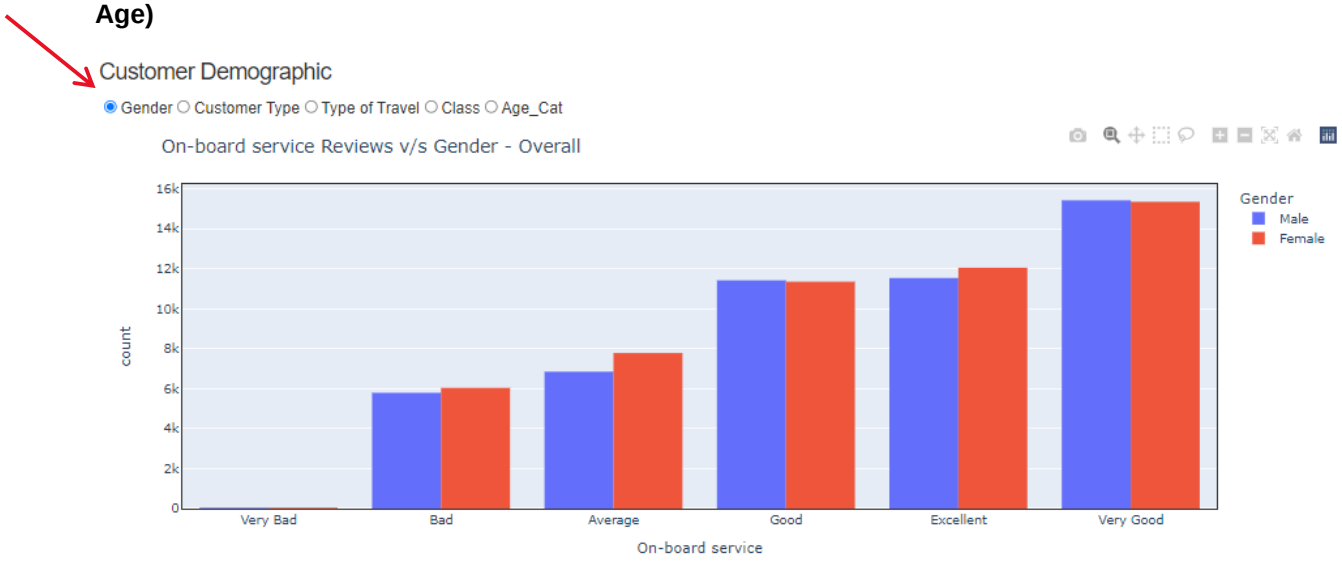
Satisfied Customers

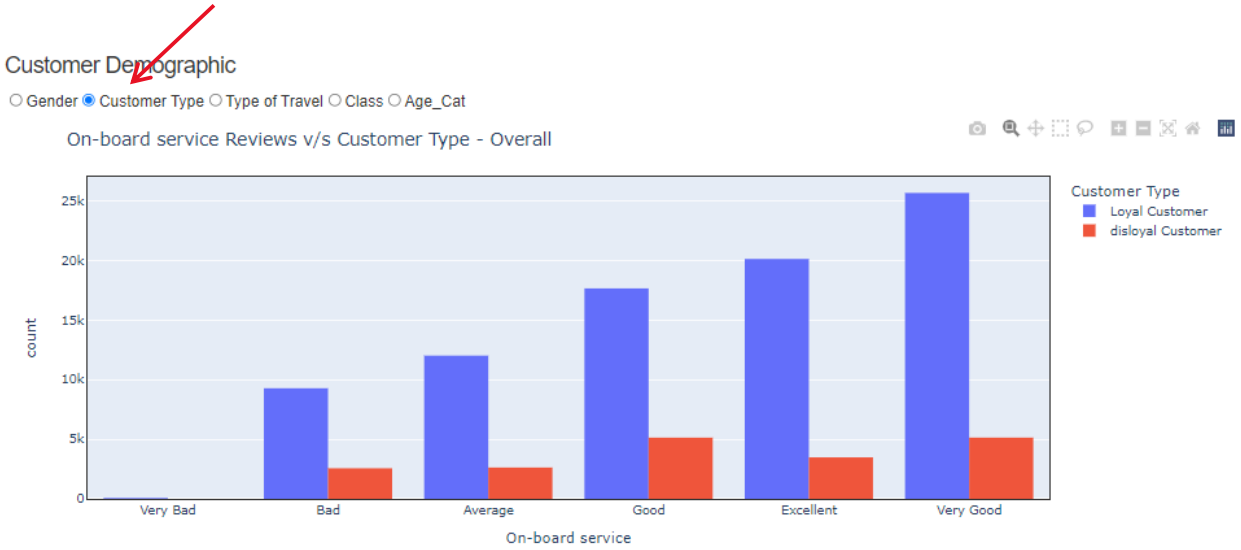
Not Satisfied Customers

Observation:

Overall customers have given extremely high rating for On-board service. More than 50% customers are happy with the service. Close to 1/6th of not-satisfied customers has rated the on-board service as “bad”

Interaction with Customer Demographics (Gender, Customer Type, Type of Travel, Class, and Age)



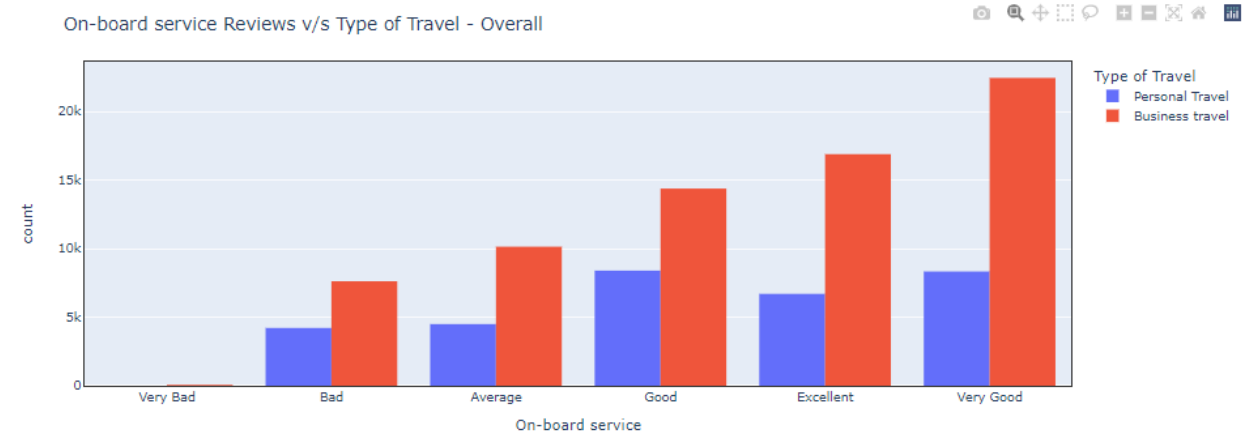


THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Final Term Project – Data Visualization

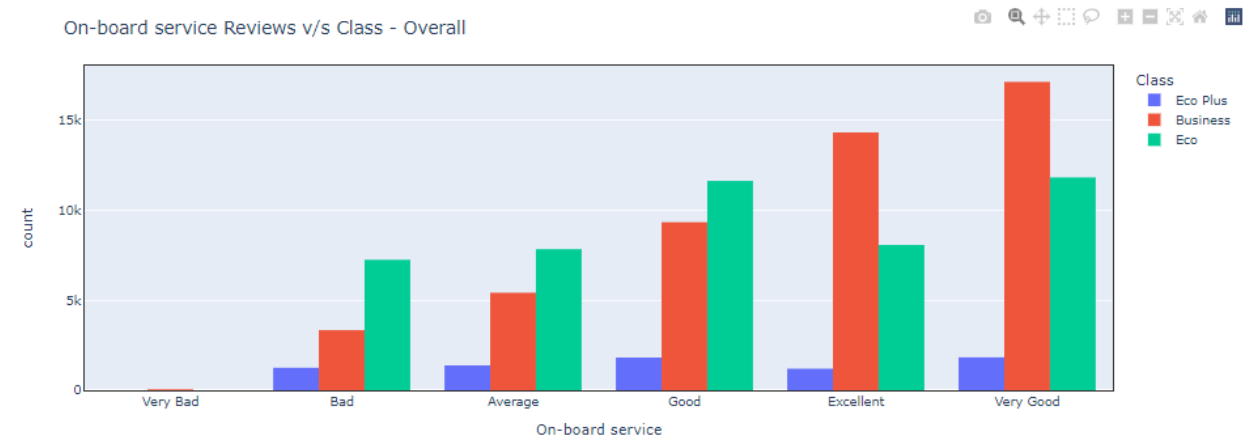
Customer Demographic

Gender | Customer Type | Type of Travel | Class | Age_Cat



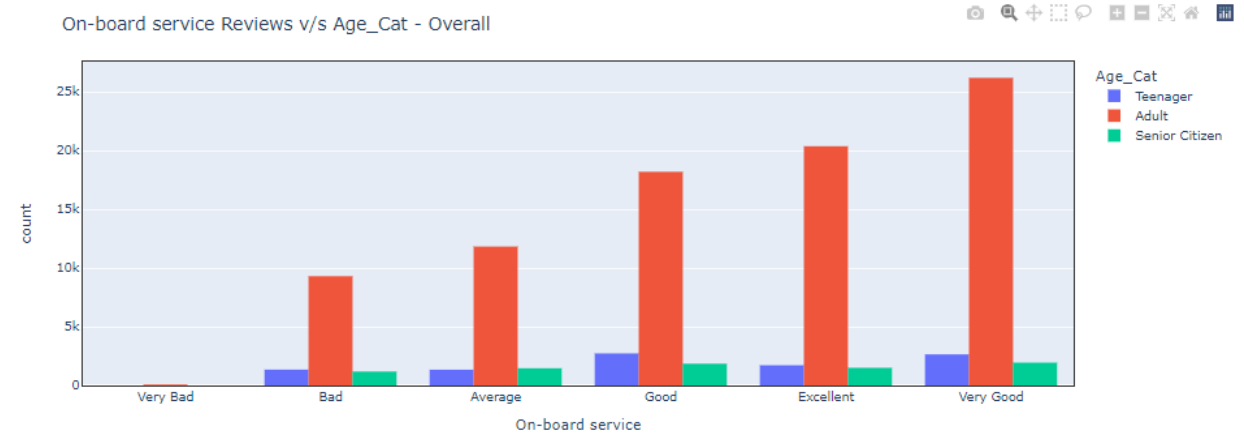
Customer Demographic

Gender | Customer Type | Type of Travel | Class | Age_Cat



Customer Demographic

Gender | Customer Type | Type of Travel | Class | Age_Cat

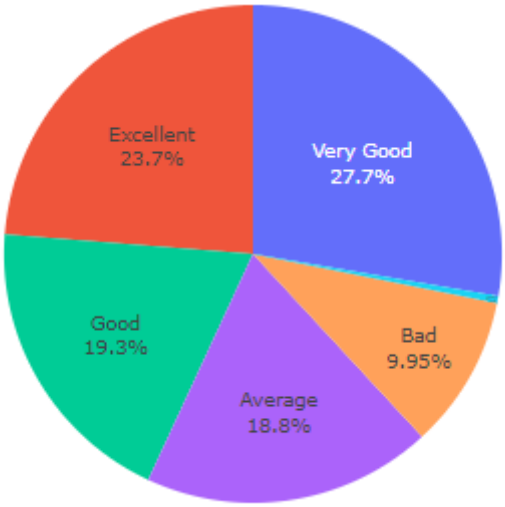


Observations:

- **Gender**: Both Male and Females were happy with the on-board offered by airline. Females have given more "average" reviews than males.
- **Customer Type**: Mostly airlines loyal customers have given reviews and they are happy with the on-board services.
- **Type of Travel**: Customers travelling for business have given more reviews. On-board services have been rated as "very good" more than it was rated "bad".
- **Class**: Passengers from all classes have rated the service highly. However, economy class passengers have given lot of "bad" reviews for on-board services.
- **Age**: As expected, adults have given lot of review reviews and more adults have categorized the on-board as "Very Good."

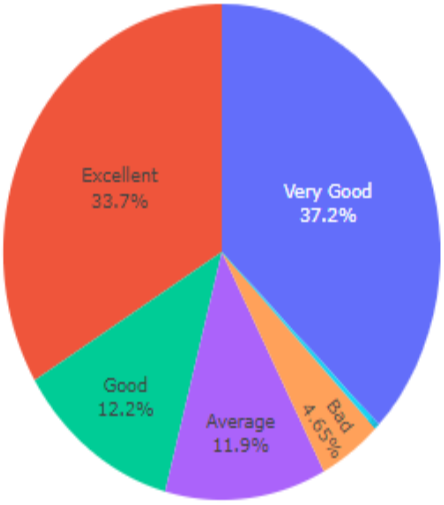
6. **Legroom:**

Overall Distribution

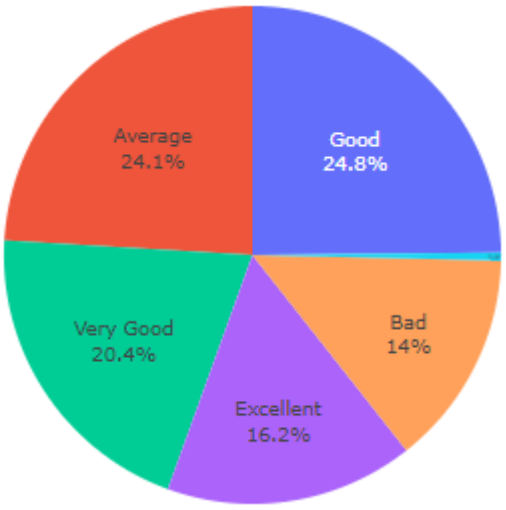


Distribution based on Customer Satisfaction

Satisfied Customers



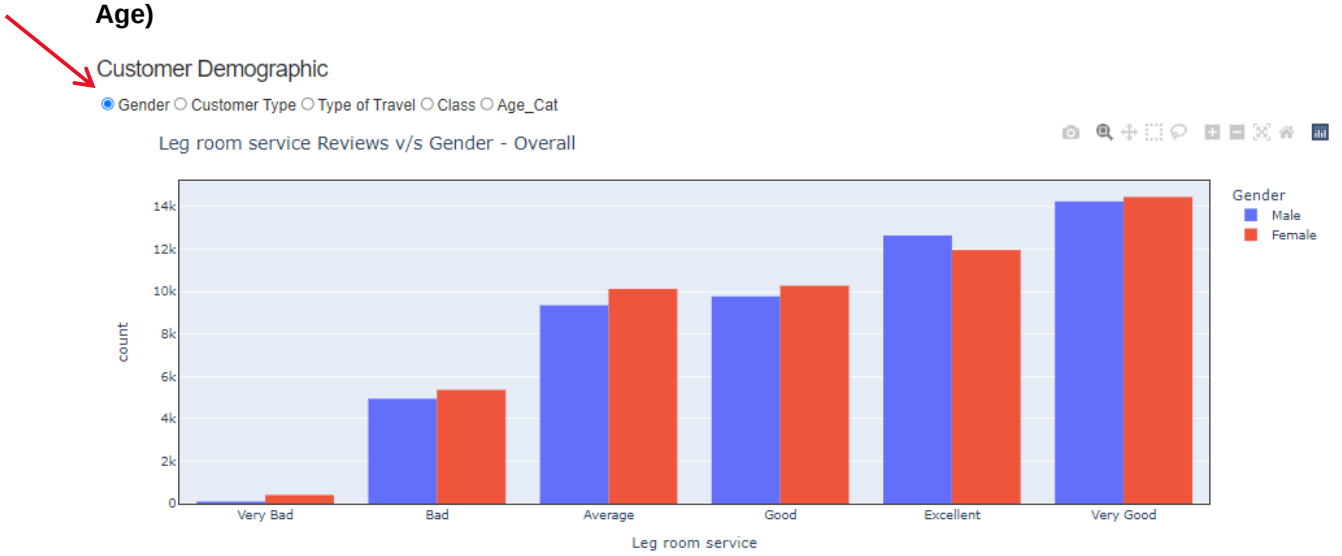
Not Satisfied Customers

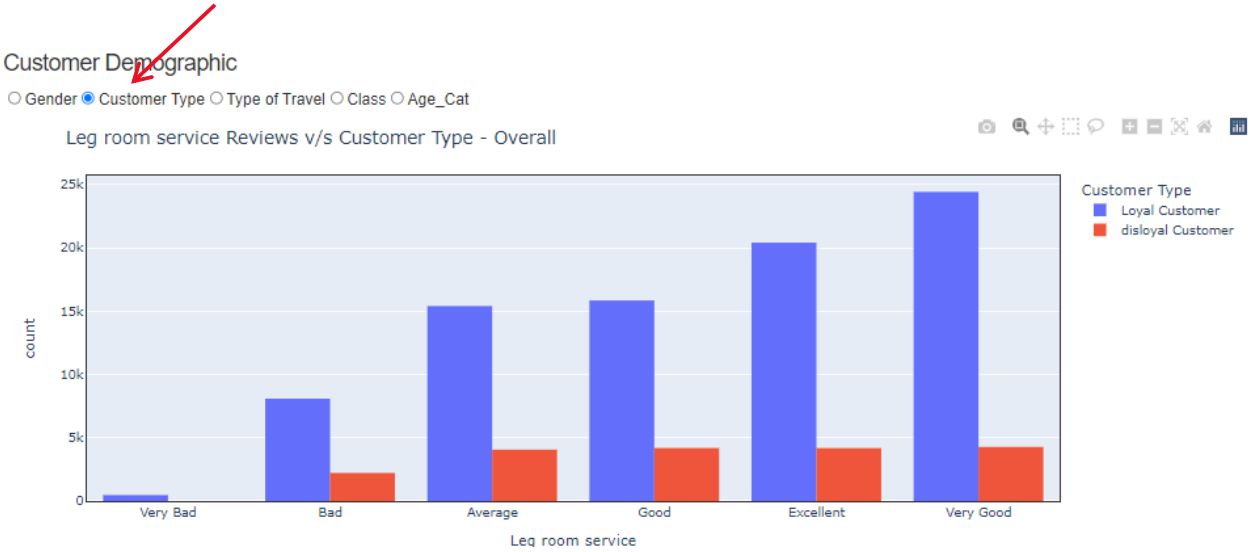


Observation:

Overall customers have given extremely high rating. More than 50% customers are happy with the legroom available.

Interaction with Customer Demographics (Gender, Customer Type, Type of Travel, Class, and Age)



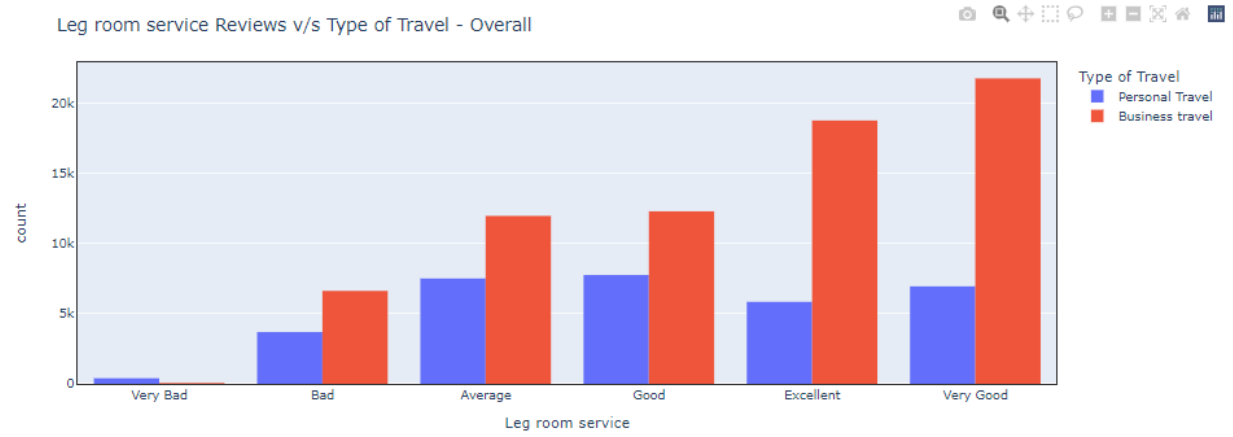


THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Final Term Project – Data Visualization

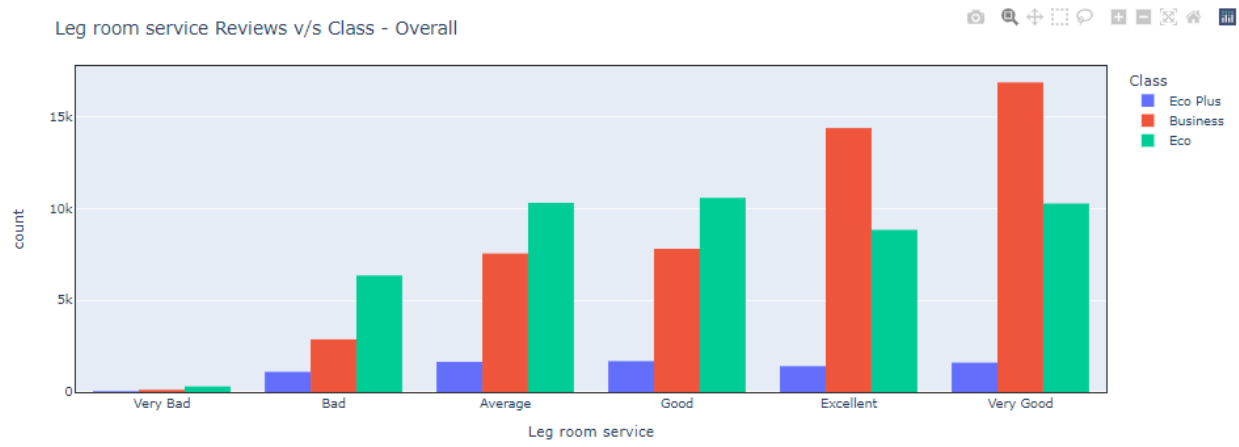
Customer Demographic

Gender | Customer Type | Type of Travel | Class | Age_Cat



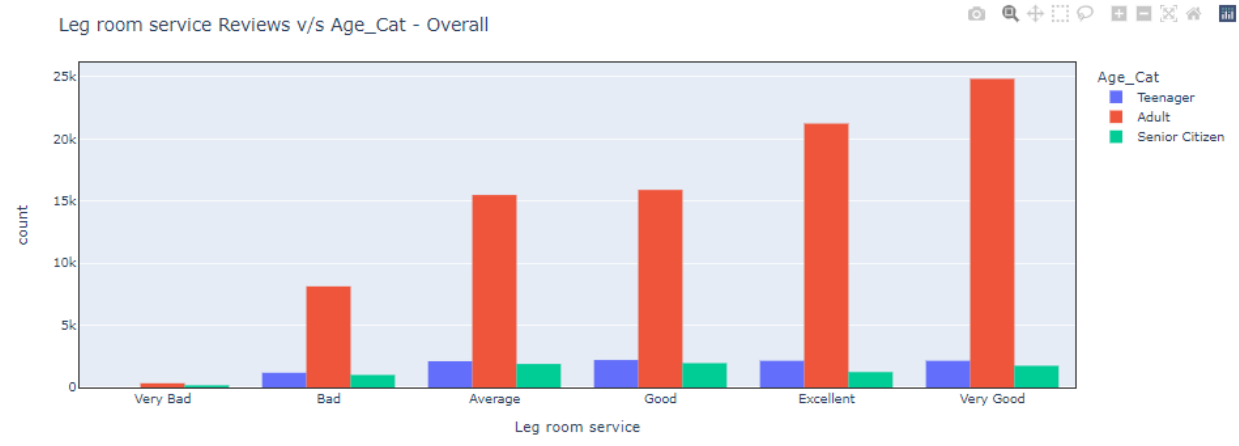
Customer Demographic

Gender | Customer Type | Type of Travel | Class | Age_Cat



Customer Demographic

Gender | Customer Type | Type of Travel | Class | Age_Cat

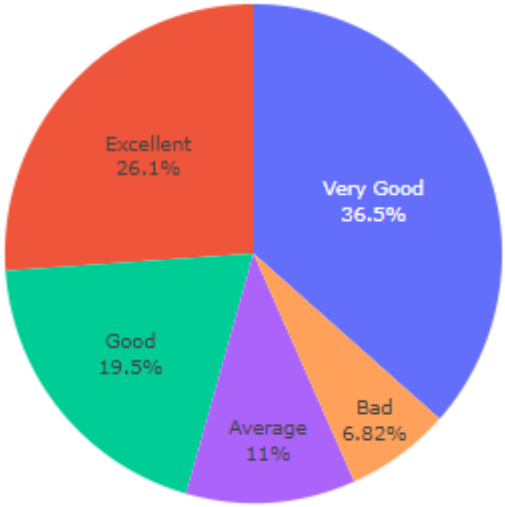


Observations:

- **Gender**: Both Male and Females were happy with the legroom offered by airline.
- **Customer Type**: Mostly airlines loyal customers have given reviews and they are happy with the legroom on-board.
- **Type of Travel**: Customers travelling for business have given more reviews. Legroom has been rated as “very good” more than it was rated “bad”.
- **Class**: Passengers from all classes have rated the service highly. However, economy class passengers have the highest “bad” reviews but overall, they appear satisfied with the legroom.
- **Age**: As expected, adults have given lot of review reviews and more adults have rated that they are happy with the legroom.

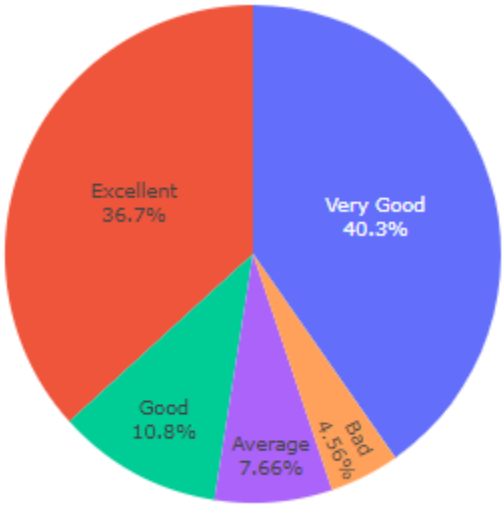
7. **In-flight service:**

Overall Distribution

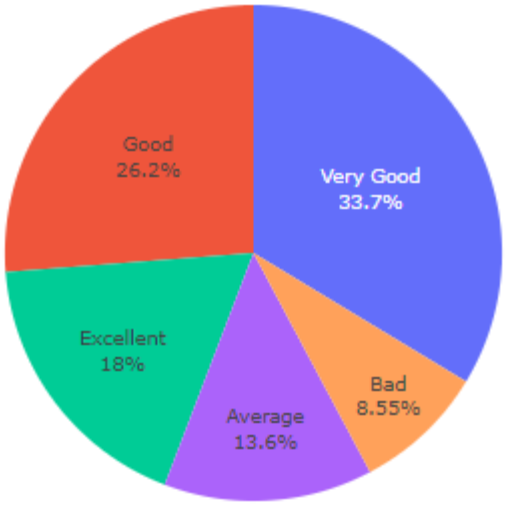


Distribution based on Customer Satisfaction

Satisfied Customers



Not Satisfied Customers



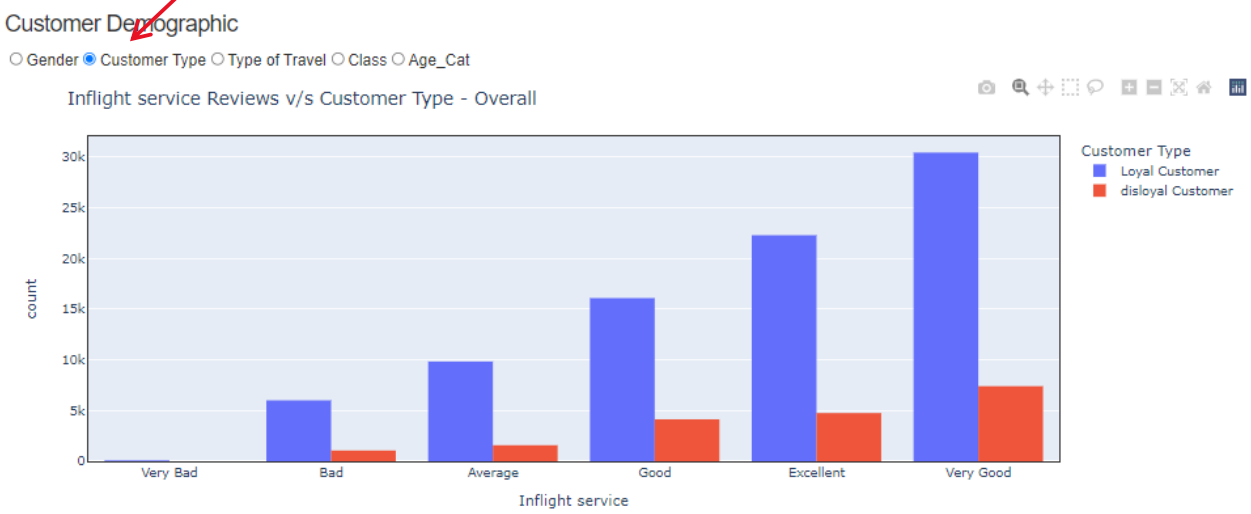
Observation:

Overall customers have given extremely high rating. More than 70% customers are happy with the in-flight services.

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Final Term Project – Data Visualization

Interaction with Customer Demographics (Gender, Customer Type, Type of Travel, Class, and Age)

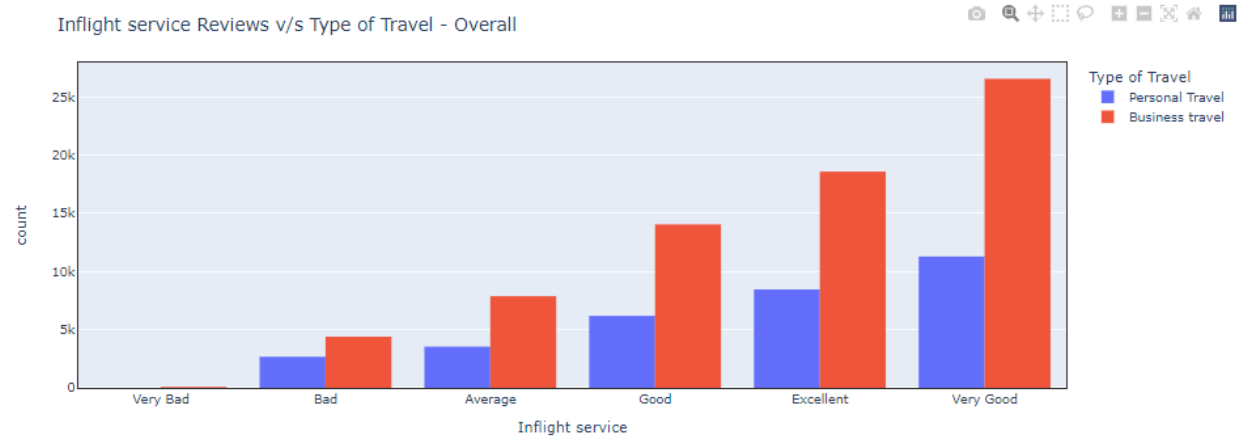


THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Final Term Project – Data Visualization

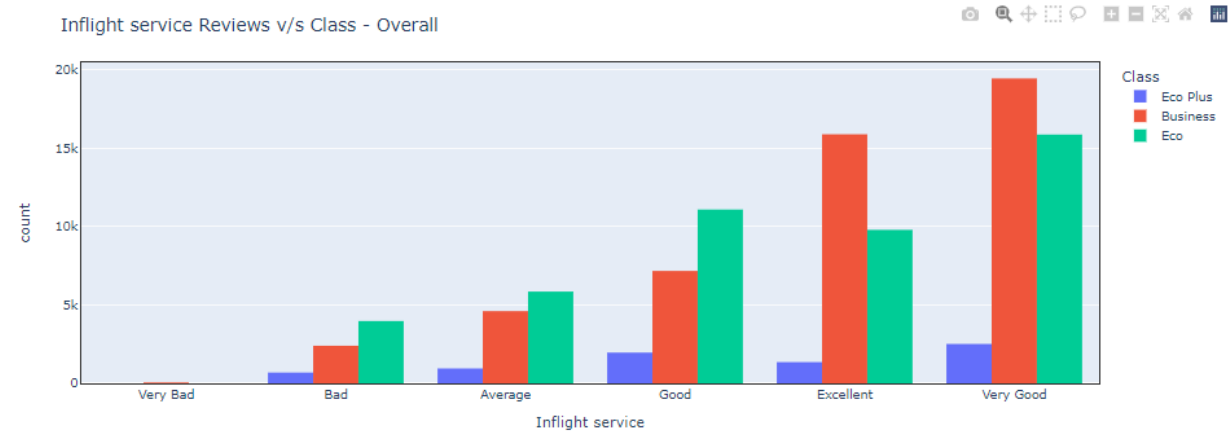
Customer Demographic

Gender | Customer Type | Type of Travel | Class | Age_Cat



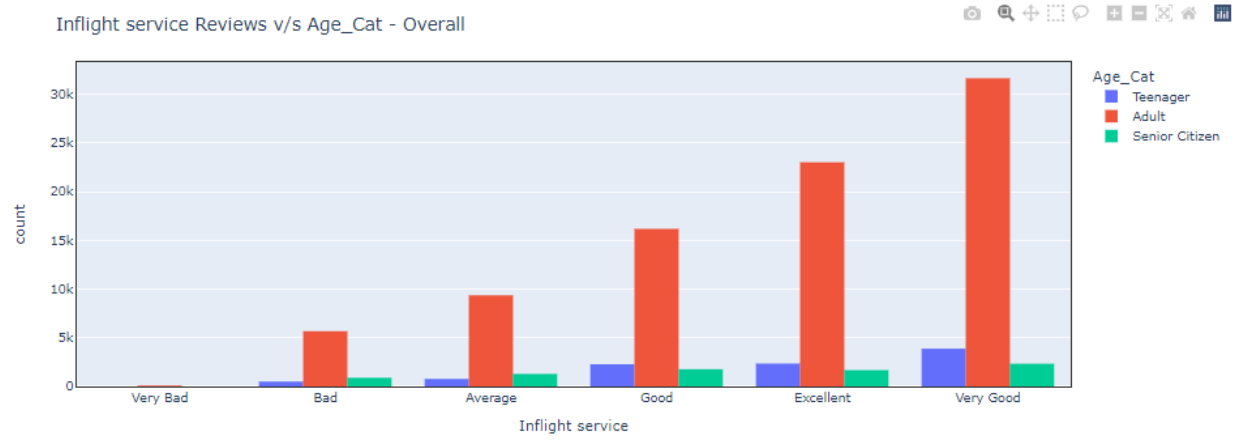
Customer Demographic

Gender | Customer Type | Type of Travel | Class | Age_Cat



Customer Demographic

Gender | Customer Type | Type of Travel | Class | Age_Cat

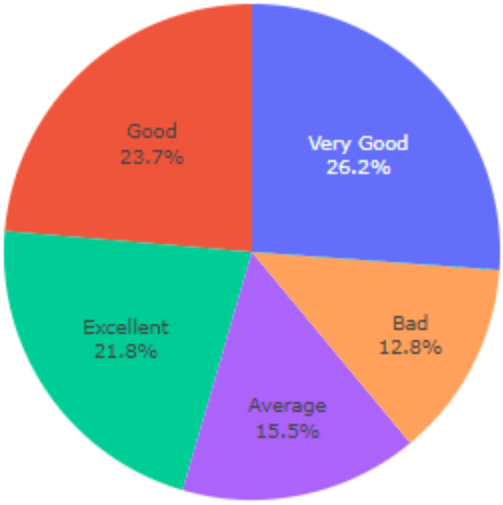


Observations:

- **Gender**: Both Male and Females have rated inflight service highly. However, females have given more "bad" reviews than males.
- **Customer Type**: Mostly airlines loyal customers have given reviews and they are happy with the service.
- **Type of Travel**: Customers travelling for business have given more reviews and have rated the service as "very good" more than it was rated "bad".
- **Class**: Passengers from all classes have rated the service highly.
- **Age**: As expected, adults have given lot of review reviews and more adults have rated that they are happy with the inflight service.

8. **Cleanliness:**

Overall Distribution



Distribution based on Customer Satisfaction

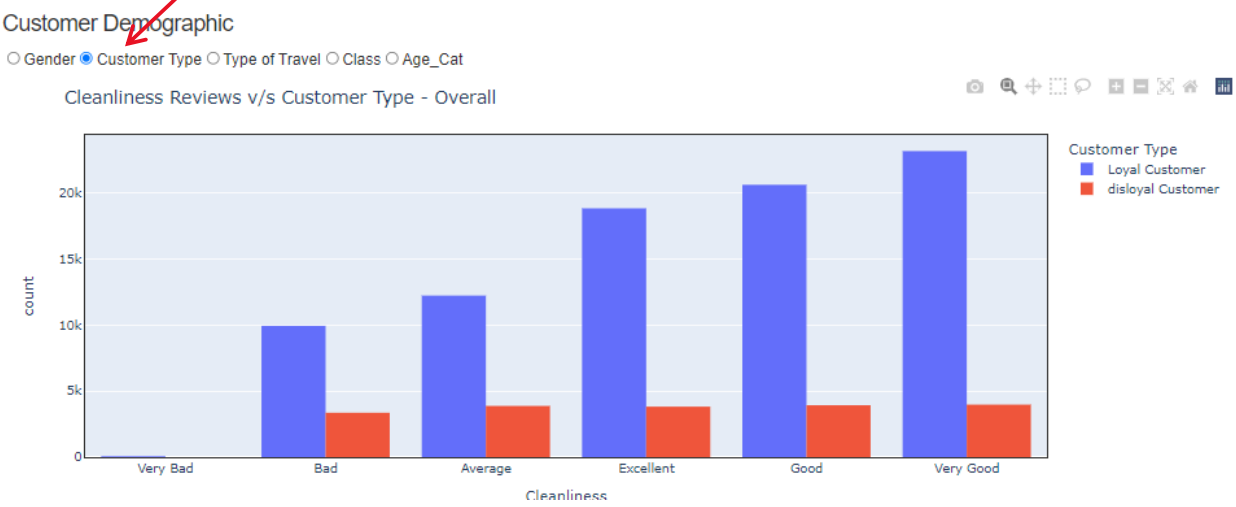
Satisfied Customers

Not Satisfied Customers

Observation:

Overall customers seem happy with the cleanliness on flight. But there are still 12.8% customers who think that the planes were not clean enough.

Interaction with Customer Demographics (Gender, Customer Type, Type of Travel, Class, and Age)



THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

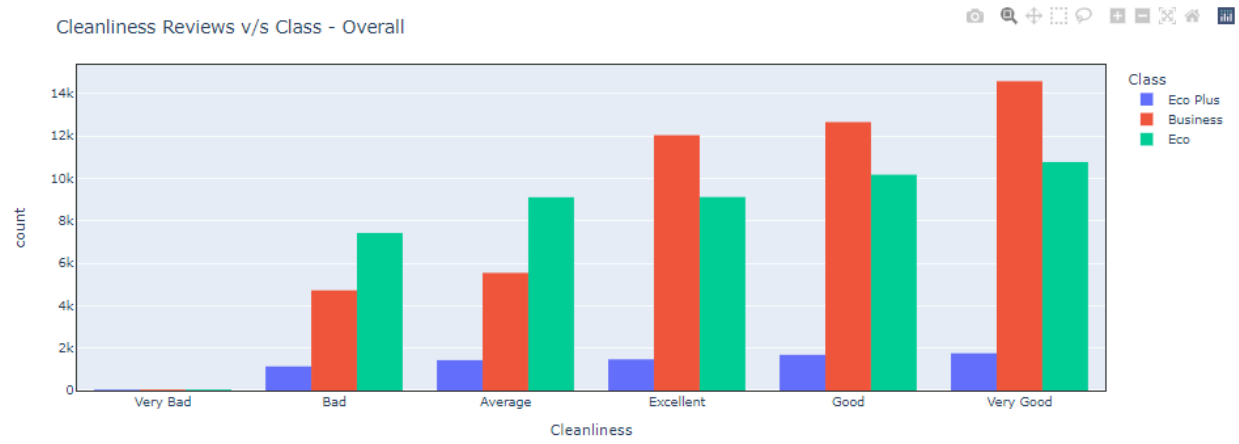
Final Term Project – Data Visualization

Customer Demographic

Gender | Customer Type | Type of Travel | Class | Age_Cat

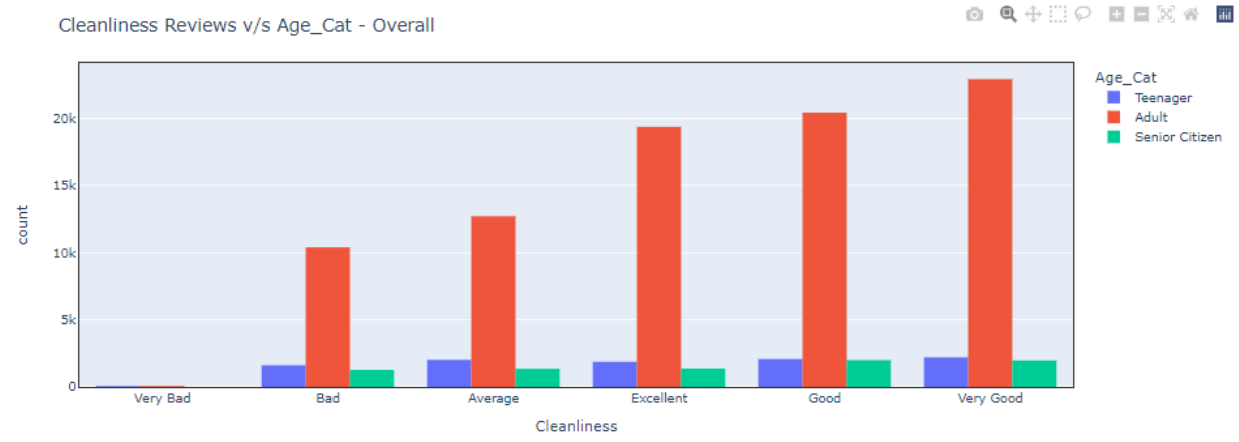
Customer Demographic

Gender | Customer Type | Type of Travel | Class | Age_Cat



Customer Demographic

Gender | Customer Type | Type of Travel | Class | Age_Cat



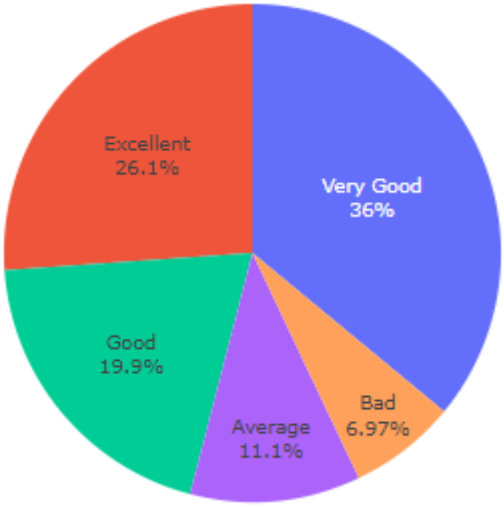
Observations:

- **Gender**: Both Male and Females have rated that the plane was well cleaned.
- **Customer Type**: Mostly airlines loyal customers have given reviews and they are happy with the cleanliness.
- **Type of Travel**: Customers travelling for business have given more reviews and have rated the service as “very good” and “excellent”.
- **Class**: Passengers from all classes have rated the cleanliness highly.
- **Age**: As expected, adults have given lot of review reviews and more adults have rated that they are happy with the cleanliness.

Post-flight Factors

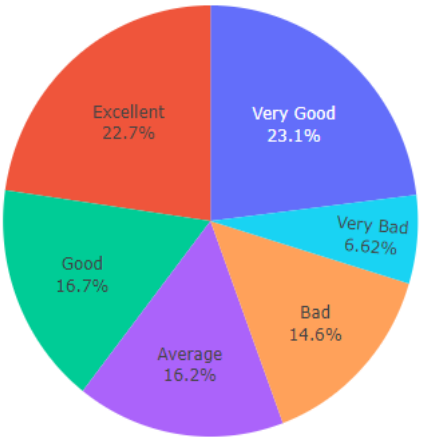
1. Baggage Handling:

Overall Distribution

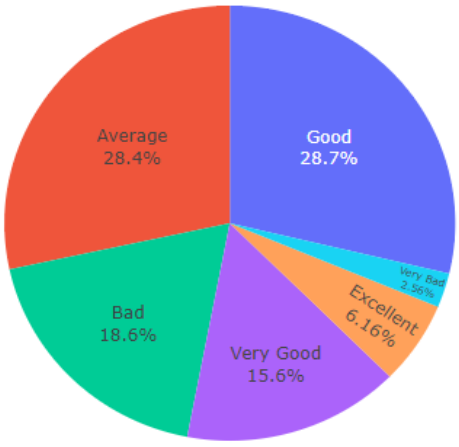


Distribution based on Customer Satisfaction

Satisfied Customers



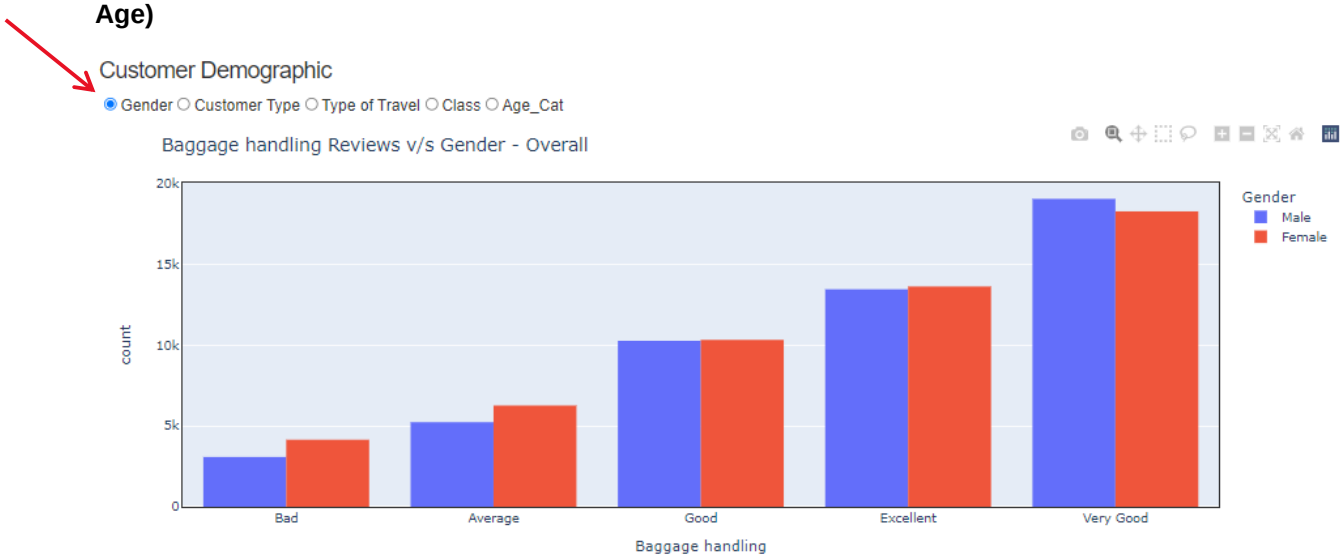
Not Satisfied Customers



Observation:

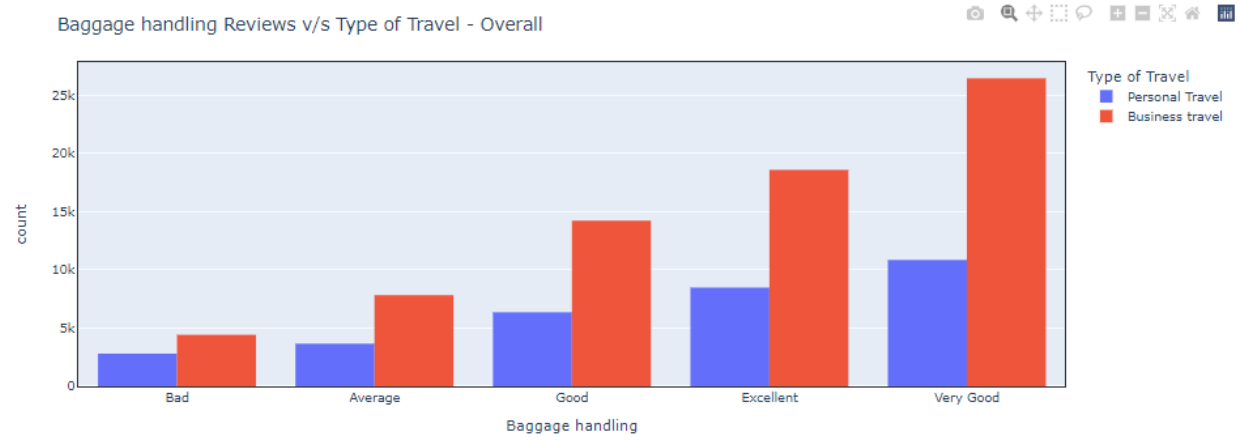
Overall customers are happy with how their baggage was handled.

Interaction with Customer Demographics (Gender, Customer Type, Type of Travel, Class, and Age)



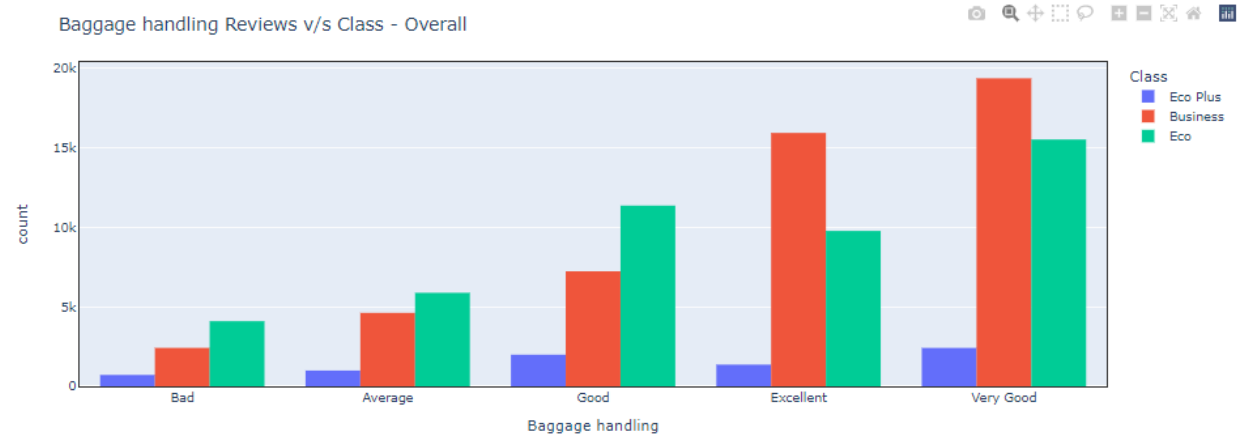
Customer Demographic

Gender | Customer Type | Type of Travel | Class | Age_Cat



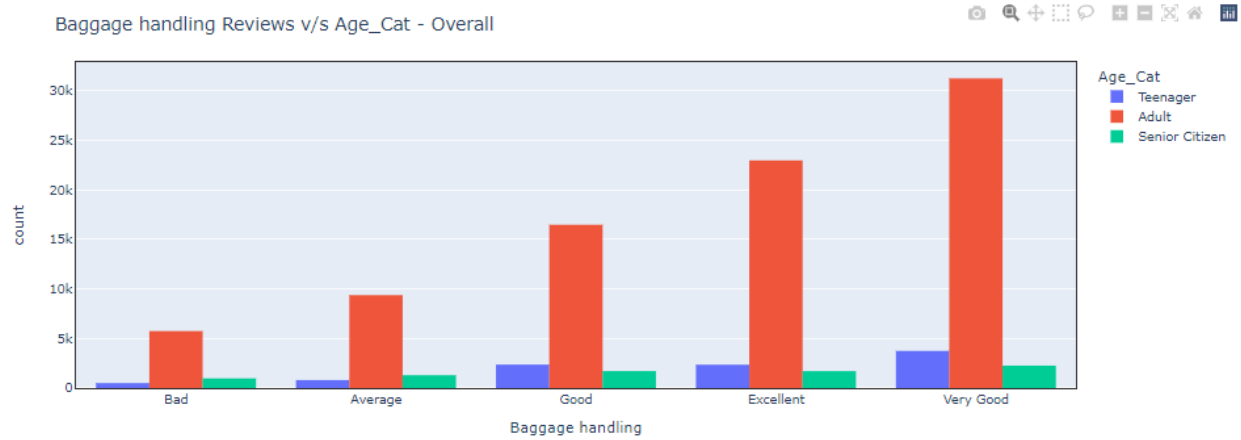
Customer Demographic

Gender | Customer Type | Type of Travel | Class | Age_Cat



Customer Demographic

Gender | Customer Type | Type of Travel | Class | Age_Cat

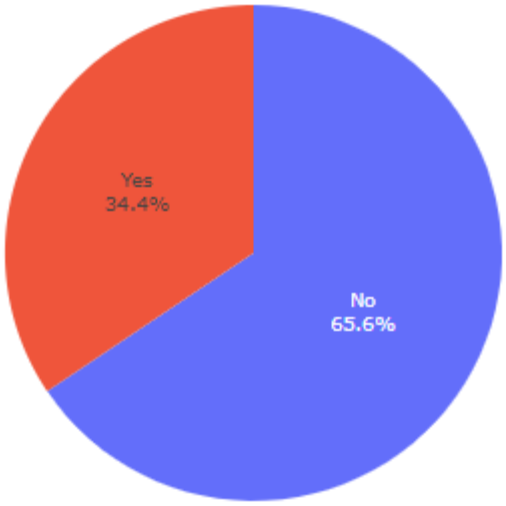


Observations:

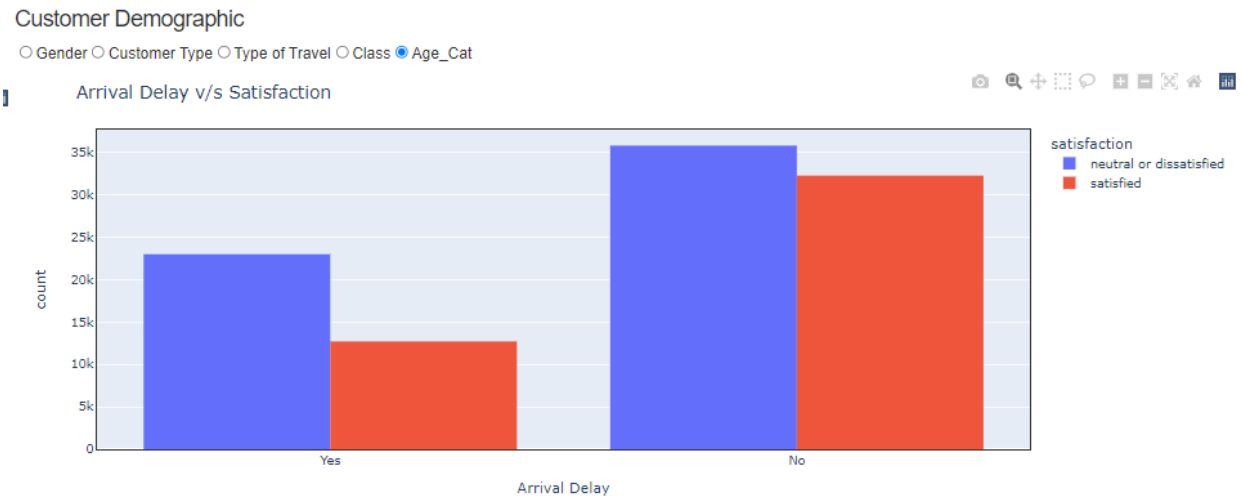
- **Gender**: More reviews have been given by Males, but it appears that both Male and Females were happy with the "Baggage Handling"
- **Customer Type**: Mostly airlines loyal customers have given reviews and they are happy with the service. A lot of customers have given "Excellent" reviews.
- **Type of Travel**: Customers travelling for business have given more reviews. Most number of reviews indicate that customers are happy with the service.
- **Class**: Business class passengers have given more reviews and are mostly good, like economy or economy plus passengers. However, for economy class, the number of "bad" reviews are significantly more compared to business class.
- **Age**: As expected, adults have given more reviews and they are good with the service.

2. Arrival Delay:

Overall Distribution



Distribution based on Customer Satisfaction



NOTE: ARRIVAL DELAY WILL ONLY BE SHOWN w.r.t SATISFACTION

Observation:

One-third of the flights have had arrived late. However, based on customer satisfaction, even if there was no delayed flight, many customers were still not satisfied.

Summary and Recommendations

Through various methods adopted to visualize the customer satisfaction data, we've able to drill down on each factor and establish its importance in the process. We've concluded that:

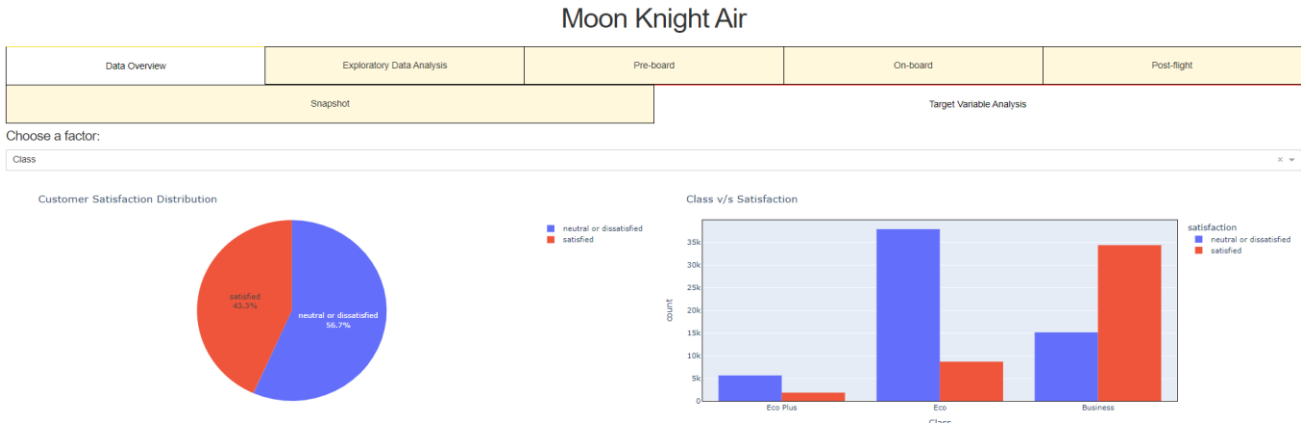
- Each factor is important in contributing towards a "satisfied" customer
- Any category with "Very Bad" reviews needs to be reviewed and improved
- Positive impact
    - Online boarding
    - Food and Drinks
    - In-flight entertainment
- Requires improvement
    - In-flight Wi-fi
    - Ease of Online Booking
    - Departure Delay

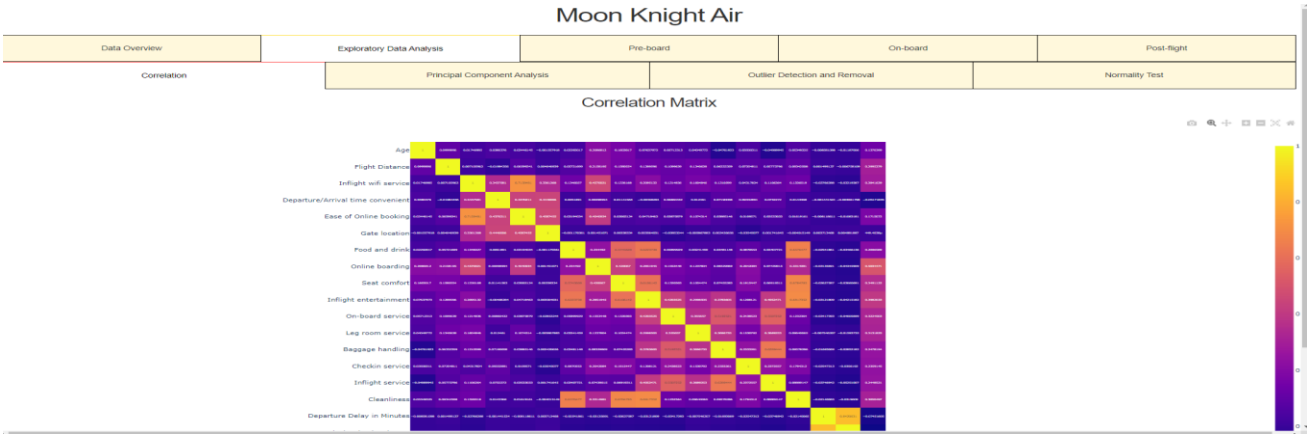
Furthermore, to strengthen our analysis, we can gather more factors like cost of ticket, flight duration, accessibility, routes offered, lounge access, etc. and observe their effect on customer satisfaction. We can also develop a classification model to better map features that impact the customer satisfaction

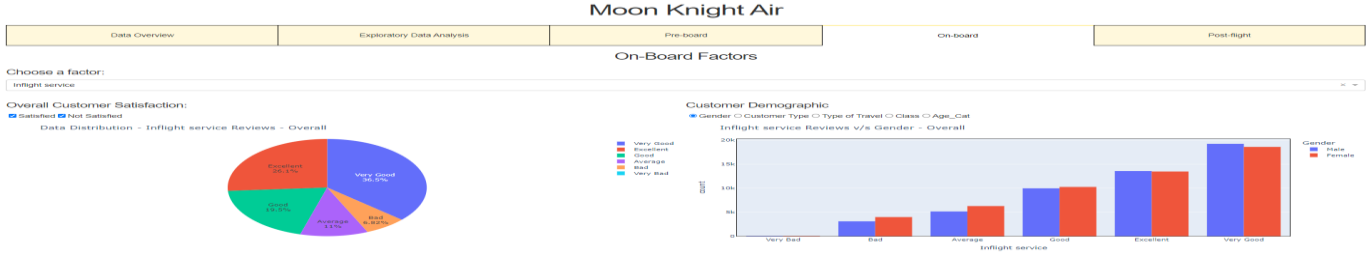
Appendix

Code GitHub Link – https://github.com/IshanKuchroo/Airline_Satisfaction_Survey

Google Cloud App Screenshots:







References

Plotly Graphs

https://plotly.com/

https://dash.plotly.com/dash-core-components

https://community.plotly.com/

Seaborn Graphs:

https://seaborn.pydata.org/tutorial.html

Miscellaneous:

https://www.jdpower.com/business/press-releases/2021-north-america-airline-satisfaction-study

https://www.theacsi.org/industries/travel/airlines/