Dataset: Airline Passenger Satisfaction.

This dataset contains information of an airline passenger satisfaction collected during the survey for the satisfaction of passenger in the airline. I downloaded the dataset from Kaggle, upon downloading it, it was a zipped file containing both test and training data. The test data was 20% while the training data is 80% of the whole dataset. The purpose of splitting this data into test and train was to build the model to predict customer satisfaction based on different features above. However, during this assignment I was asked to explore the data with univariate, bivariate, and then multivariate exploration using Power BI or Table which means that there is no need to split the data to build the predictive model. Therefore, I combined both test and train data into one final dataset.

But before combining, the **test** data has **25976 rows** and **24 columns** with **83 missing** values in the "**Arrival Delay in Minutes**" column. While the **train** data has **103904 rows** and **24 columns** with **310 missing** values in the "**Arrival Delay in Minutes**" column. Therefore, before concatenating both datasets, I **imputed the missing values**. Then I combined them into a single dataset. The **final dataset** has **129880 rows** and **24 columns** with **no missing** value.

Some columns have clear meanings like Gender, Customer Type, Type of Travel, Class, and Satisfaction which are categorical columns indicating the kind or type of passenger based on those categories.

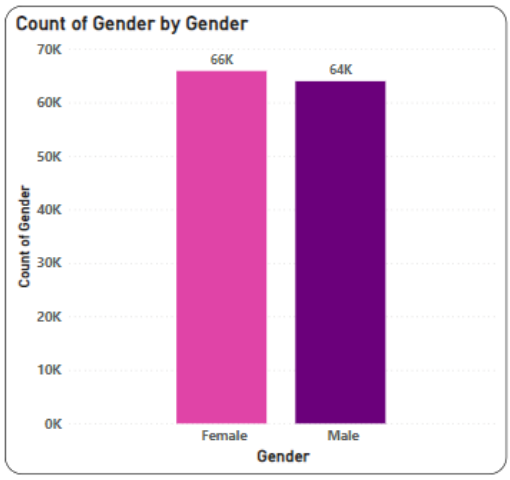
The rest columns are numerical and some of them need clear explanation. However, it is clear for Id, Age, Flight Distance, Departure Delay in Minutes and Arrival Delay in Minutes which are represented by the numerical values indicating their values.

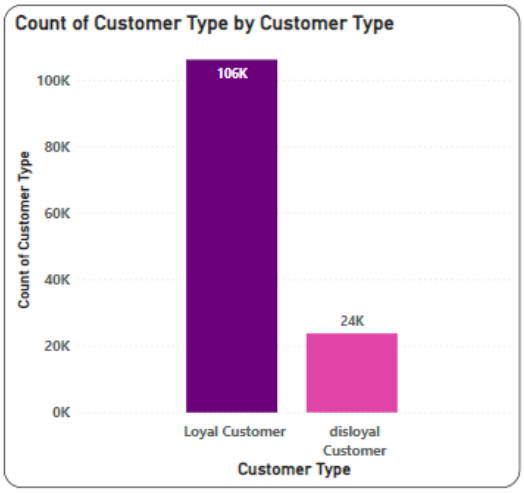
Other remaining columns such as inflight wifi service, Departure/Arrival time convenient, Ease of Online booking, Gate location, Food and drink, online boarding, Seat comfort, inflight entertainment, On-board service, Leg room service, Baggage handling, Checkin service, inflict service, and Cleanliness are **classified in numbers between 1-5** where **0** indicates **Not Applicable**, **1-5 range** indicates the **satisfactory level** of the service. Means that **1** indicates **weakly satisfied** while **5** indicates **strongly satisfied**.

I split the Analysis into **6 stages** which are:

1. Demographics Analysis
2. Travel preferences Analysis
3. Service satisfaction Analysis
4. Delay Analysis
5. Correlation Analysis
6. Satisfaction by Type of Travel, Class, Customer Type and Gender Analysis

1. **Demographics**: Let me analyze the data by gender, age, and customer type to understand the demographics of the airline customers. This can help managers understand who is most likely to fly with the airline and tailor their marketing and customer service strategies accordingly.

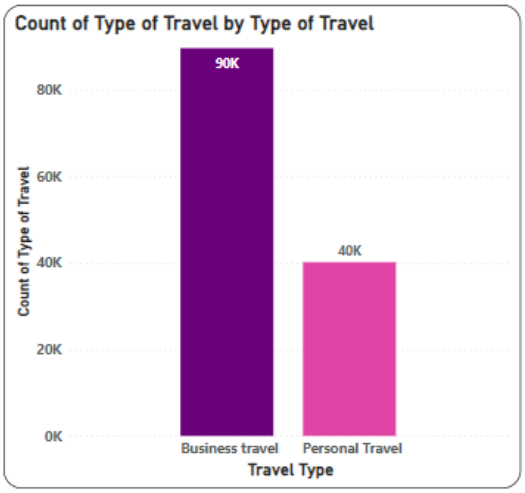


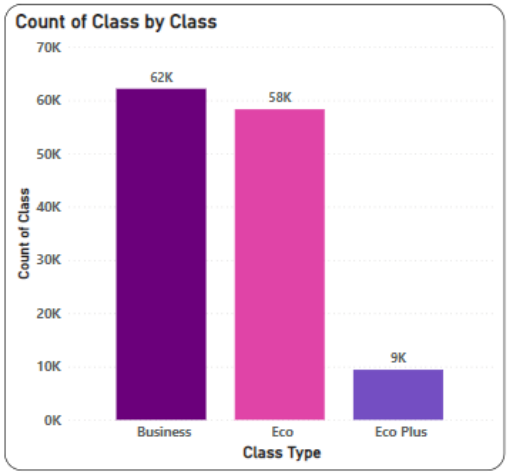


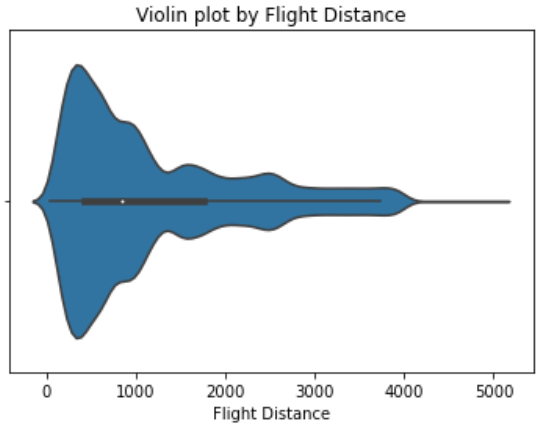
From these visuals, we see that **Female customers** are more than **Male customers** though there is little difference. Then from the data we have, we see that there are **more customers** from **Loyal type** Customer than the ones from **Disloyal Customers**. And on the **Age**, from this violin plot, we see that the **mean age** of all the customers is around **40** and more customers ranges from **22** to **55** years.

Therefore, while setting the policies on who and how to provide airline satisfactions' services, managers should put into consideration the **Female** customers than **Male** customers, they should also prioritize the **Loyal customers** than **Disloyal customers** and moreover, they should consider providing those services to customers whose years range from **22-55** years than the other ones.

2. **Travel preferences**: Then, lets analyze the data by type of travel, class, and flight distance to understand the travel preferences of the airline customers. This can help managers understand the types of flights and routes that are most popular and make decisions about route planning and fleet deployment.

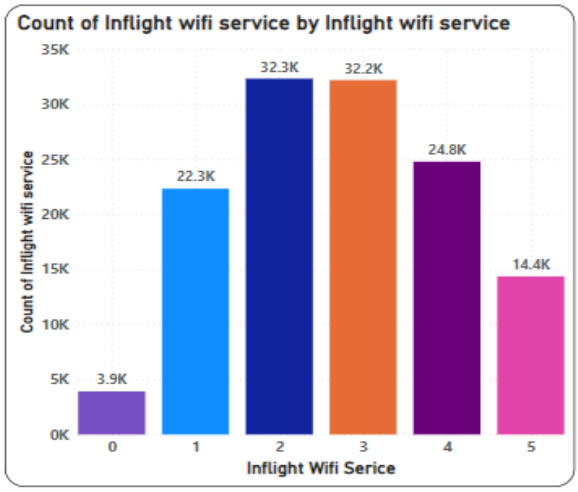


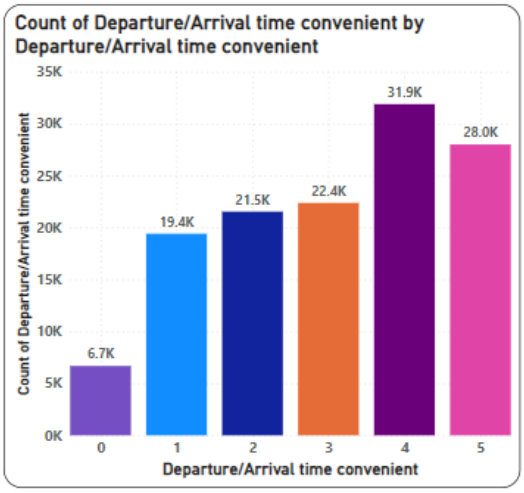


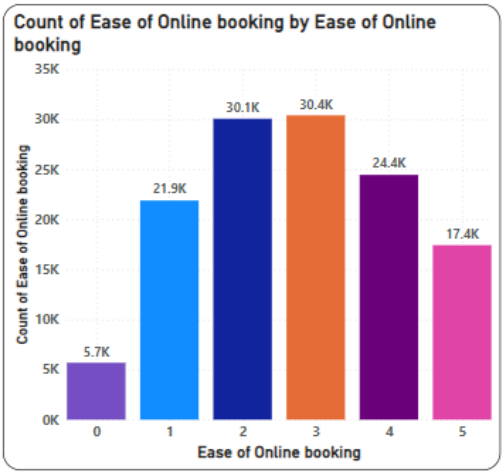


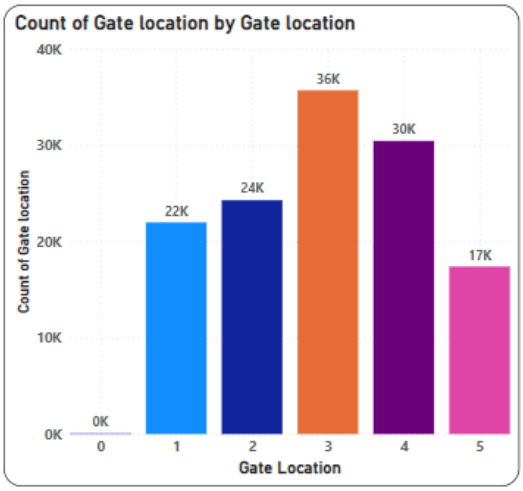
Then here we see that more customers come from **Business Travel** than the ones from **Personal Travel**. Therefore, priority should be given to **Business Travel**. In addition, **Eco** and **Business** class have more customers than **Eco plus** which has very few customers compared to those. Therefore, priority should be given to **customers from Eco and Business class**. Moreover, the violin plot shows us that the **average flight distance is around 1200** and more customers took the flights with distance between **1000 and 2000**. Then more satisfactions services should be given to this group of customers with the flight distance ranging **between 1000 and 2000**.

3. **Service satisfaction**: Then, lets analyze the data by 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, check-in service, inflight service, cleanliness, and satisfaction to understand **how customers rate the different aspects of the service**. This can help managers identify areas where **service needs improvement** and make changes to enhance the customer experience.

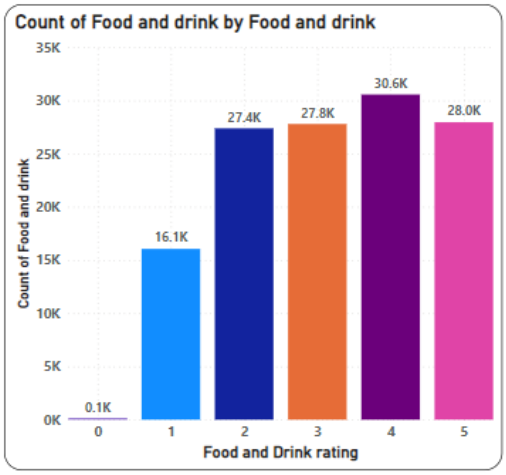


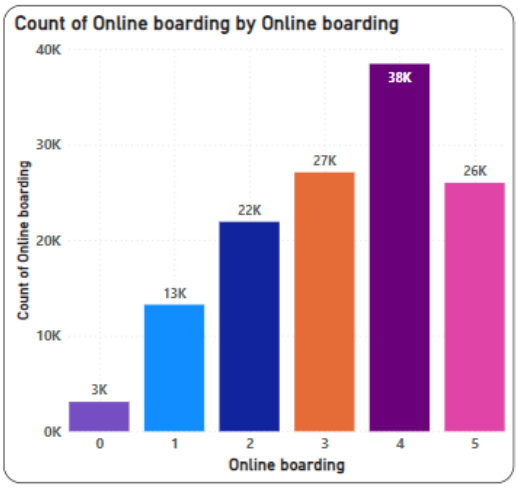


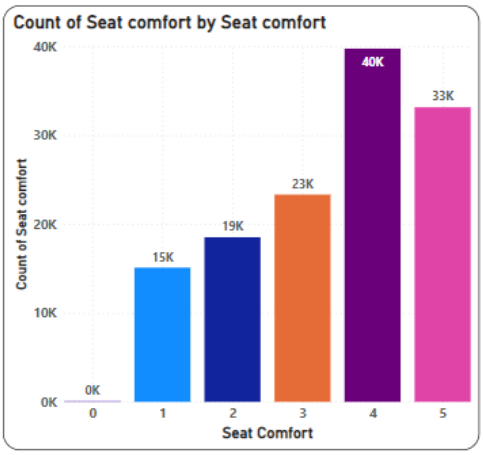


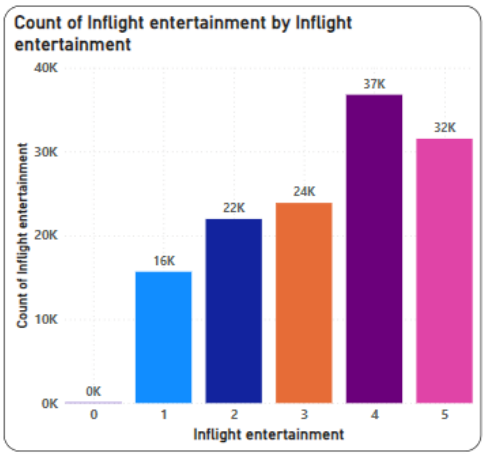


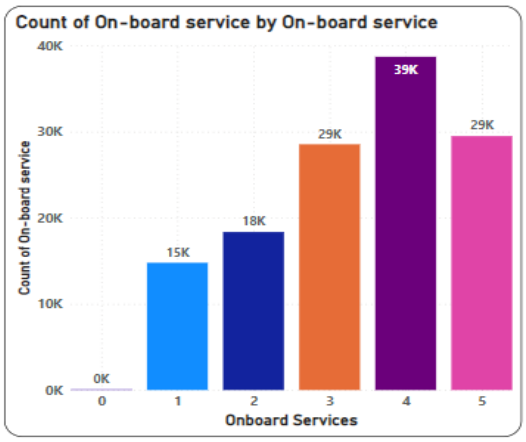
We can see that for the **Departure/Arrival time convenient**, more customers rated this to its **maximum i.e 4&5** which means that this service **need little improvements**. However, we can see that for **inflight wifi services, ease of online booking and gate location**, only few customers rated these services to their maximum i.e **4&5** instead **most of them rated these services as moderate** i.e most of them rated these services to be **somewhere between 2&3**, and **3&4** which can be considered as **moderate**. Therefore, these services need **strong improvement** to **enhance customers satisfactions**.

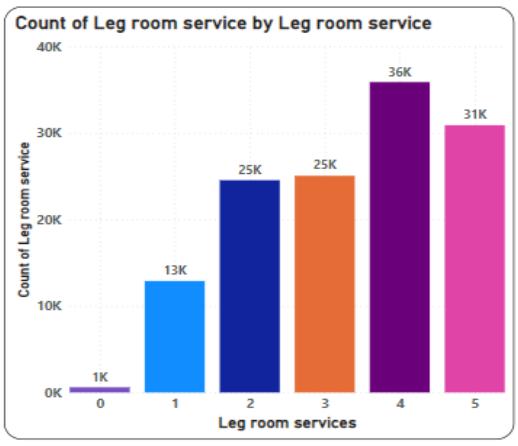

Count of Food and drink by Food and drink
Count of Food and drink
35K
30K
25K
20K
15K
10K
5K
0K
0.1K
16.1K
27.4K
27.8K
30.6K
28.0K
0
1
2
3
4
5
Food and Drink rating

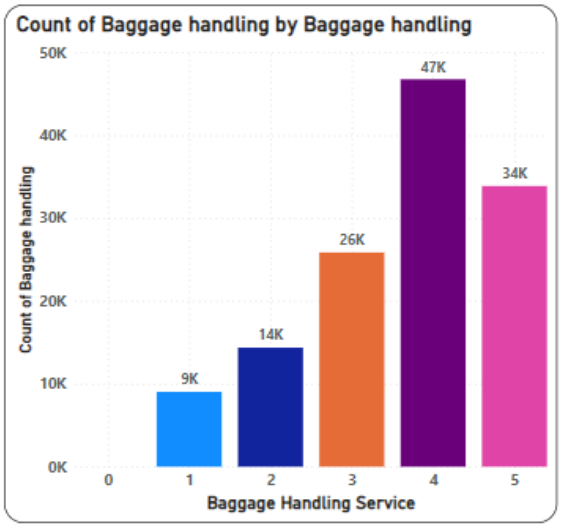

Count of Online boarding by Online boarding
Count of Online boarding
40K
30K
20K
10K
0K
3K
13K
22K
27K
38K
26K
0
1
2
3
4
5
Online boarding

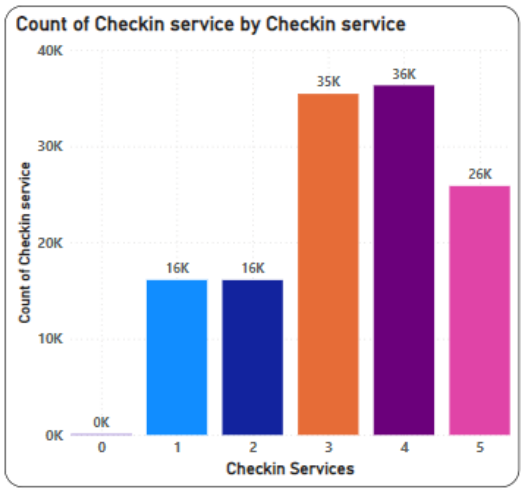

Count of Seat comfort by Seat comfort
Count of Seat comfort
40K
30K
20K
10K
0K
0K
15K
19K
23K
40K
33K
0
1
2
3
4
5
Seat Comfort

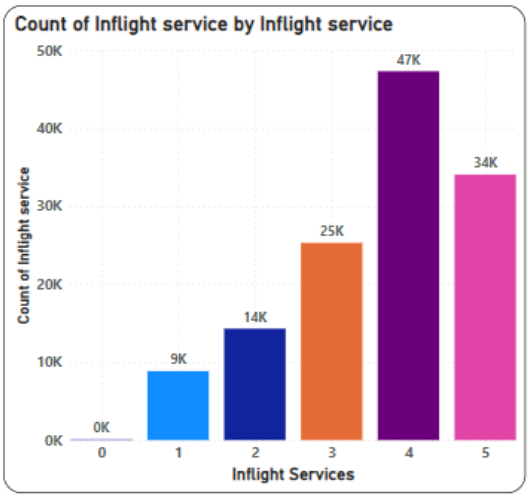

Count of Inflight entertainment by Inflight entertainment
Count of Inflight entertainment
40K
30K
20K
10K
0K
0K
16K
22K
24K
37K
32K
0
1
2
3
4
5
Inflight entertainment

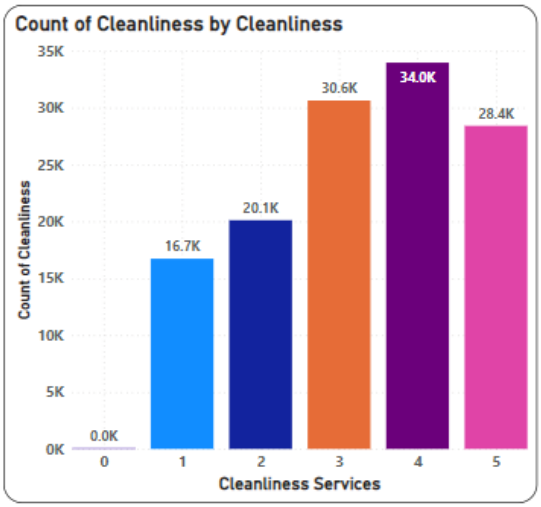

Count of On-board service by On-board service
Count of On-board service
40K
30K
20K
10K
0K
0K
15K
18K
29K
39K
29K
0
1
2
3
4
5
Onboard Services


Count of Leg room service by Leg room service
Count of Leg room service
40K
30K
20K
10K
0K
1K
13K
25K
25K
36K
31K
0
1
2
3
4
5
Leg room services

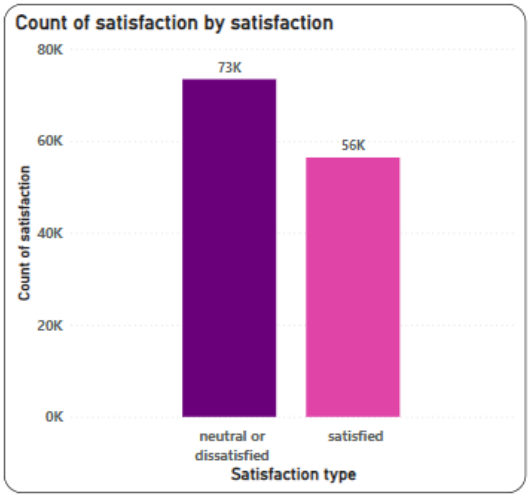






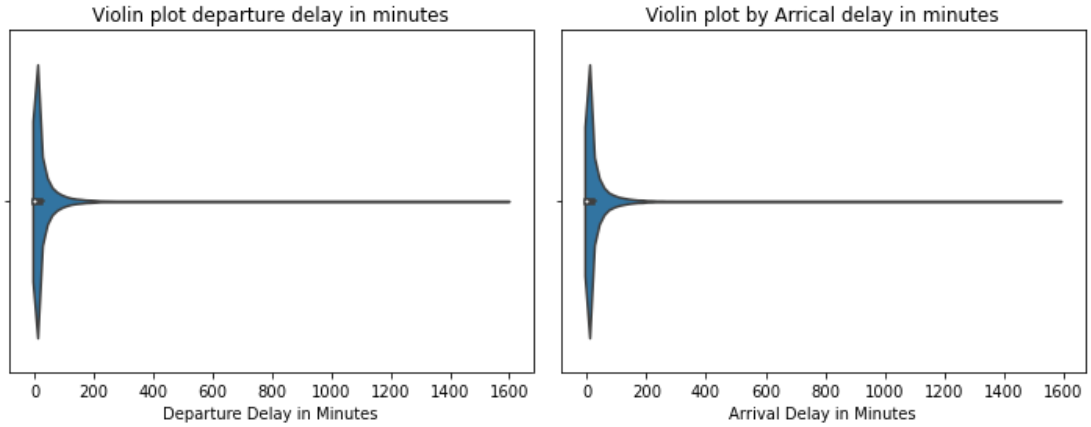


Then also for these services, **food and drink, online boarding, seat comfort, inflight entertainment, onboard service, leg room service**, **Baggage handling, check-in service, inflight service and cleanliness**, most of the customers rated these services as **maximum** i.e somewhere **between 4&5** and also as **medium** i.e somewhere between **3&4**. Therefore, these services need **moderate improvement** to enhance the customers satisfactions.



Finally, from the data given, we can see the customers that were **dissatisfied or neutral are more than the customers that were satisfied**. Therefore, **proper improvement and measures** need to be taken (with respective to the satisfactions services mentioned above) to **enhance the overall satisfaction of the customers.**

4. **Delay**: lets analyze the data by departure delay in minutes and arrival delay in minutes to understand how often flights are delayed and how long the delays typically last. This can help managers understand the causes of delay and take steps to reduce them.

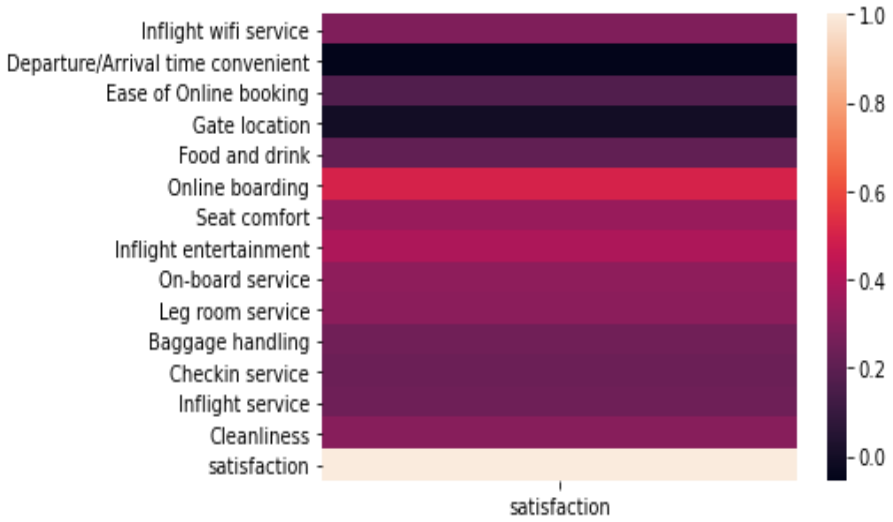


From the violin plot above, we can see that there are outliers for each variable i.e departure delay in minutes and arrival delay in minutes. Also, by using this violin plot, we can see that the **mean delay in minutes for each variable is around 15 mins.**

5. **Correlation Analysis**: let use the correlation between each feature and the target variable which is satisfaction and identify the relationship between them.

By using correlation method, here is the heatmap of the correlation between each feature.

	satisfaction
Inflight wifi service	0.283460
Departure/Arrival time convenient	-0.054270
Ease of Online booking	0.168877
Gate location	-0.002793
Food and drink	0.211340
Online boarding	0.501749
Seat comfort	0.348829
Inflight entertainment	0.398234
On-board service	0.322205
Leg room service	0.312424
Baggage handling	0.248680
Checkin service	0.237252
Inflight service	0.244918
Cleanliness	0.307035
satisfaction	1.000000

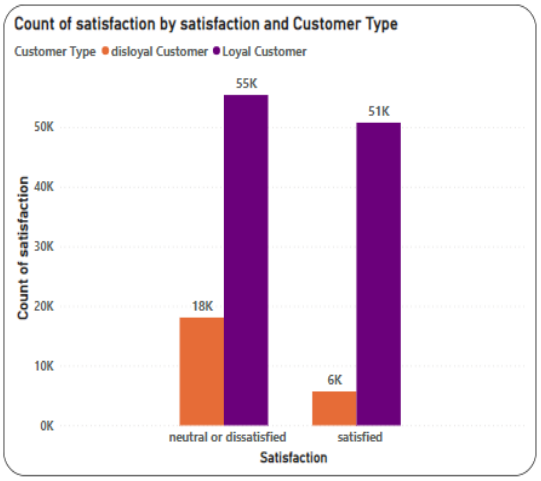


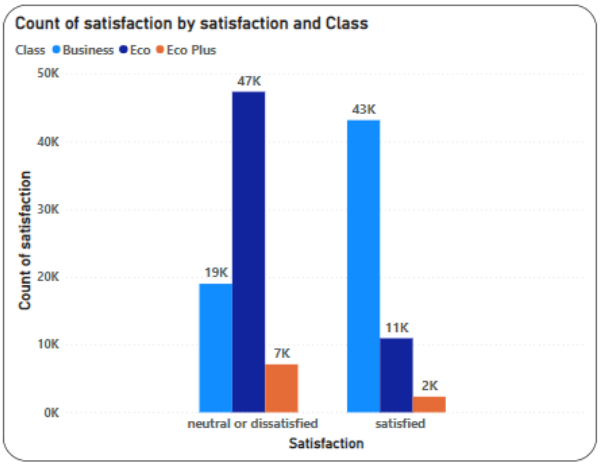
After performing the correlation coefficient between variables, I found that **online boarding** is more correlated to **satisfaction** (which is the target variable) than any other independent variable. It has a correlation coefficient of **0.5**. this mean that increasing the level of satisfactory for this service will automatically increase the number of passengers that will be satisfied by the Airline services. Other variables that are highly correlated with the target output (satisfaction) are:

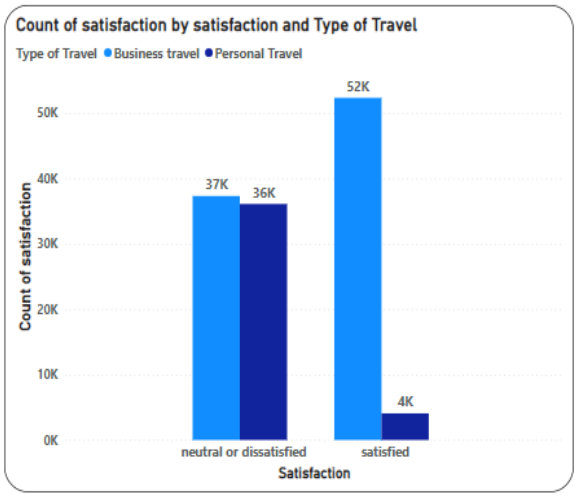
- Inflight wifi service with a correlation of **0.28**
- Seat comfort with a correlation of around **0.35**
- Inflight entertainment with a correlation of around **0.4**
- Onboard service with a correlation of around **0.322**
- Leg room service with a correlation of around **0.31**
- Cleanliness with a correlation of **0.3**

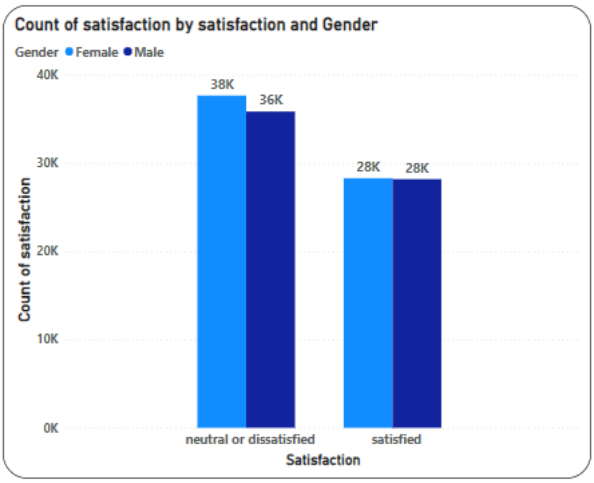
These variables are positively correlated to the target variables and have good correlation. This means that increasing their level of satisfactory to the passengers will automatically increase the number of passengers that will be satisfied the Airline services.

6. **Let's then analyze the satisfaction by Type of Travel, Class, Customer Type and Gender**





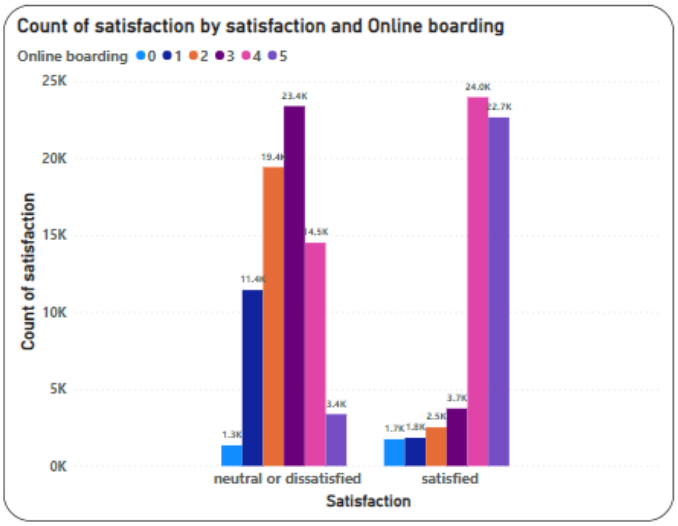




As obtained from bar plots above in Power BI, based on the **type of travel**, **Business travel** passengers are more **satisfied** while **personal travel** passengers are more **dissatisfied** by the Airline services. In addition, based on the **Customer type**, **Loyal Customer** passengers are mostly **satisfied** than **Disloyal Customer** passengers. Measures shows that a high number of **Business Travel** and high number of **Loyal Customers** were **satisfied** while a very low number of **Personal Travel** and **Disloyal Customer** were satisfied i.e most of them were **dissatisfied** by the Airline services.

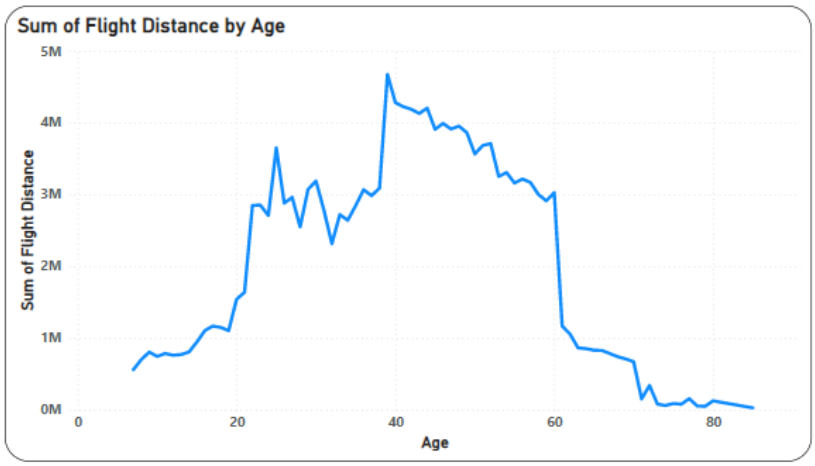
Moreover, based on the **Class**, **Business passenger** are more **satisfied**. While figures shows that passengers from **Eco and Eco plus** Class are mostly **dissatisfied** by the Airline. Therefore, great emphasize should be carried out on these two classes which are **Eco and Eco plus** to see what can be done to improve their satisfaction of the Airline. And **Female** passengers tends to be more **dissatisfied** than **Male** passengers.

Let's use **online boarding** as most correlated service to the target variable (satisfaction) and see how the ratings for these services affect the **final satisfaction** for the passengers.



According to the visuals, the graphs show that as the **level of satisfaction increase** from **1 to 5** in these columns 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, inflict service, and Cleanliness, the **number of satisfied passengers in satisfaction column (target)** increases too. This means that the **satisfactory level of 4 and 5** for each of the above columns (services) have more passengers who have **satisfied** by the Airline while the **satisfactory level between 1-3** have few passengers who have been satisfied by the Airline Services i.e most of them were **dissatisfied**.

Now, let also get some insights from Flight Distance and the Age of passenger.



Visuals shows that passengers between **20- and 60-years** old travel for **long Flight Distance** while people under **20 and those over 60 years** old travel for short Flight Distance. It also shows that the more people that travel in Airline have between **20-60 years old**.