Part_I_exploration_template

February 13, 2023

1 Part I - Airline Passenger Satisfaction Dataset Exploration

1.1 by Haifaa Mohamad Alzahrani

1.2 Introduction

The goal of this project is to analyze the **Airline Passenger Satisfaction** dataset. It is published by Kaggle and contains an airline passenger satisfaction survey. It provides 129880 observations with 24 different factors to help analyze the passenger's satisfaction.

> The dataset is in csv format, but it is divided into Train and Test sections. However, since I am not interested in prediction, I would combine them into one dataframe.

Below more details about the factors as documented by the source: * 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:Not Applicable;1-5) * 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)

1.3 Preliminary Wrangling

In [1]:
```
# import all packages and set plots to be embedded inline
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sb

%matplotlib inline
```

In [2]:
```python
# To show all columns
pd.set_option("display.max_columns", 40)
```

In [3]:
```python
dataframe_train = pd.read_csv("train.csv")
```

In [4]:
```python
dataframe_test = pd.read_csv("test.csv")
```

In [5]:
```python
print(dataframe_test.shape)
print(dataframe_train.shape)
```

```
(25976, 25)
(103904, 25)
```

In [6]:
```python
dataframe = pd.concat([dataframe_train, dataframe_test])
```

In [7]:
```python
dataframe = dataframe.drop_duplicates()
```

In [8]:
```python
dataframe.to_csv('satisfaction.csv', encoding='utf-8')
```

In [9]:
```python
dataframe.head()
```

Out[9]:

```
   Unnamed: 0      id  Gender      Customer Type  Age   Type of Travel  \
0           0   70172    Male     Loyal Customer   13  Personal Travel
1           1    5047    Male  disloyal Customer   25  Business travel
2           2  110028  Female     Loyal Customer   26  Business travel
3           3   24026  Female     Loyal Customer   25  Business travel
4           4  119299    Male     Loyal Customer   61  Business travel

      Class  Flight Distance  Inflight wifi service  \
0  Eco Plus              460                      3
1  Business              235                      3
2  Business             1142                      2
3  Business              562                      2
4  Business              214                      3

   Departure/Arrival time convenient  Ease of Online booking  Gate location  \
0                                  4                       3              1
1                                  2                       3              3
2                                  2                       2              2
3                                  5                       5              5
4                                  3                       3              3

   Food and drink  Online boarding  Seat comfort  Inflight entertainment  \
0               5                3             5                       5
1               1                3             1                       1
2               5                5             5                       5
3               2                2             2                       2
4               4                5             5                       3
```

```
   On-board service  Leg room service  Baggage handling  Checkin service  \
0                 4                 3                 4                4
1                 1                 5                 3                1
2                 4                 3                 4                4
3                 2                 5                 3                1
4                 3                 4                 4                3

   Inflight service  Cleanliness  Departure Delay in Minutes  \
0                 5            5                          25
1                 4            1                           1
2                 4            5                           0
3                 4            2                          11
4                 3            3                           0

   Arrival Delay in Minutes             satisfaction
0                      18.0  neutral or dissatisfied
1                       6.0  neutral or dissatisfied
2                       0.0                satisfied
3                       9.0  neutral or dissatisfied
4                       0.0                satisfied
```

```
In [10]: dataframe = dataframe.drop(["Unnamed: 0"], axis=1)
```

```
In [11]: dataframe.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 129880 entries, 0 to 25975
Data columns (total 24 columns):
 #   Column                             Non-Null Count   Dtype
---  ------                             --------------   -----
 0   id                                 129880 non-null  int64
 1   Gender                             129880 non-null  object
 2   Customer Type                      129880 non-null  object
 3   Age                                129880 non-null  int64
 4   Type of Travel                     129880 non-null  object
 5   Class                              129880 non-null  object
 6   Flight Distance                    129880 non-null  int64
 7   Inflight wifi service              129880 non-null  int64
 8   Departure/Arrival time convenient  129880 non-null  int64
 9   Ease of Online booking             129880 non-null  int64
 10  Gate location                      129880 non-null  int64
 11  Food and drink                     129880 non-null  int64
 12  Online boarding                    129880 non-null  int64
 13  Seat comfort                       129880 non-null  int64
 14  Inflight entertainment             129880 non-null  int64
 15  On-board service                   129880 non-null  int64
 16  Leg room service                   129880 non-null  int64
```

```
 17  Baggage handling                   129880 non-null  int64
 18  Checkin service                    129880 non-null  int64
 19  Inflight service                   129880 non-null  int64
 20  Cleanliness                        129880 non-null  int64
 21  Departure Delay in Minutes         129880 non-null  int64
 22  Arrival Delay in Minutes           129487 non-null  float64
 23  satisfaction                       129880 non-null  object
dtypes: float64(1), int64(18), object(5)
memory usage: 24.8+ MB
```

In [12]:
```python
# Data type of Arrival Delay in Minutes should be int not float
# 1) Fill na with 0
dataframe["Arrival Delay in Minutes"] = dataframe["Arrival Delay in Minutes"].fillna(0)
# 2) Convert
dataframe["Arrival Delay in Minutes"] = dataframe["Arrival Delay in Minutes"].astype(in
```

In [13]:
```python
dataframe.rename(columns={"satisfaction": "Satisfaction"}, inplace=True)
```

In [14]:
```python
dataframe.describe()
```

Out[14]:
```
                  id            Age  Flight Distance  Inflight wifi service  \
count  129880.000000  129880.000000    129880.000000          129880.000000
mean    64940.500000      39.427957      1190.316392               2.728696
std     37493.270818      15.119360       997.452477               1.329340
min         1.000000       7.000000        31.000000               0.000000
25%     32470.750000      27.000000       414.000000               2.000000
50%     64940.500000      40.000000       844.000000               3.000000
75%     97410.250000      51.000000      1744.000000               4.000000
max    129880.000000      85.000000      4983.000000               5.000000

       Departure/Arrival time convenient  Ease of Online booking  \
count                      129880.000000           129880.000000
mean                            3.057599                2.756876
std                             1.526741                1.401740
min                             0.000000                0.000000
25%                             2.000000                2.000000
50%                             3.000000                3.000000
75%                             4.000000                4.000000
max                             5.000000                5.000000

       Gate location  Food and drink  Online boarding   Seat comfort  \
count  129880.000000   129880.000000    129880.000000  129880.000000
mean        2.976925        3.204774         3.252633       3.441361
std         1.278520        1.329933         1.350719       1.319289
min         0.000000        0.000000         0.000000       0.000000
25%         2.000000        2.000000         2.000000       2.000000
50%         3.000000        3.000000         3.000000       4.000000
75%         4.000000        4.000000         4.000000       5.000000
```

```
max          5.000000          5.000000          5.000000         5.000000

       Inflight entertainment  On-board service  Leg room service  \
count           129880.000000     129880.000000     129880.000000
mean                 3.358077          3.383023          3.350878
std                  1.334049          1.287099          1.316252
min                  0.000000          0.000000          0.000000
25%                  2.000000          2.000000          2.000000
50%                  4.000000          4.000000          4.000000
75%                  4.000000          4.000000          4.000000
max                  5.000000          5.000000          5.000000

       Baggage handling  Checkin service  Inflight service    Cleanliness  \
count     129880.000000    129880.000000     129880.000000  129880.000000
mean           3.632114         3.306267          3.642193       3.286326
std            1.180025         1.266185          1.176669       1.313682
min            1.000000         0.000000          0.000000       0.000000
25%            3.000000         3.000000          3.000000       2.000000
50%            4.000000         3.000000          4.000000       3.000000
75%            5.000000         4.000000          5.000000       4.000000
max            5.000000         5.000000          5.000000       5.000000

       Departure Delay in Minutes  Arrival Delay in Minutes
count               129880.000000             129880.000000
mean                    14.713713                 15.045465
std                     38.071126                 38.416353
min                      0.000000                  0.000000
25%                      0.000000                  0.000000
50%                      0.000000                  0.000000
75%                     12.000000                 13.000000
max                   1592.000000               1584.000000
```

In [15]:
```python
print("Gender values: ", dataframe["Gender"].unique())
print("Customer Type values: ", dataframe["Customer Type"].unique())
print("Age values: ", dataframe["Age"].unique())
print("Type of Travel values: ", dataframe["Type of Travel"].unique())
print("Class: ", dataframe["Class"].unique())
print("Satisfaction values: ", dataframe["Satisfaction"].unique())
```

```
Gender values:  ['Male' 'Female']
Customer Type values:  ['Loyal Customer' 'disloyal Customer']
Age values:  [13 25 26 61 47 52 41 20 24 12 53 33 45 38  9 17 43 58 23 57 49 36 22 31
 15 35 67 37 40 34 39 50 29 54 21 28 27 69 60 48 59 46 30 66 64 44 51 32
 19 42 16 11 62  8 56 68 55 18 65 72 70 63 10  7 14 80 74 71 85 73 76 77
 75 79 78]
Type of Travel values:  ['Personal Travel' 'Business travel']
Class:  ['Eco Plus' 'Business' 'Eco']
Satisfaction values:  ['neutral or dissatisfied' 'satisfied']
```

1.3.1 What is the structure of your dataset?

There are 129880 survey observations in the dataset with 24 features described above. Most variables are numeric (int) in nature, but there are categorical/qualitative variables as following: | Feature | Type | Examples | | ----------- | ----------- | ----------- | | Gender | Nominal | 'Male' 'Female' | Customer Type | Nominal | 'Loyal Customer' 'Disloyal Customer' | Age | Ordinal | 13 25 26 61 47 52 41 20 24 12 53 33 45 38 9 17 43 58 23 57 49 36 22 31 15 35 67 37 40 34 39 50 29 54 21 28 27 69 60 48 59 46 30 66 64 44 51 32 19 42 16 11 62 8 56 68 55 18 65 72 70 63 10 7 14 80 74 71 85 73 76 77 75 79 78 | Type of Travel | Nominal | 'Personal Travel' 'Business travel' | Class | Ordinal | 'Eco' 'Eco Plus' 'Business' | Satisfaction | Nominal | 'Neutral or Dissatisfied' 'Satisfied'

1.3.2 What is/are the main feature(s) of interest in your dataset?

I'm most interested in figuring out what features are best for predicting the passenger's satisfaction in the dataset.

1.3.3 What features in the dataset do you think will help support your investigation into your feature(s) of interest?

I expect that carat will have the strongest effect on each diamond's price: the larger the diamond, the higher the price. I also think that the other big "C"s of diamonds: cut, color, and clarity, will have effects on the price, though to a much smaller degree than the main effect of carat.

1.4 Univariate Exploration

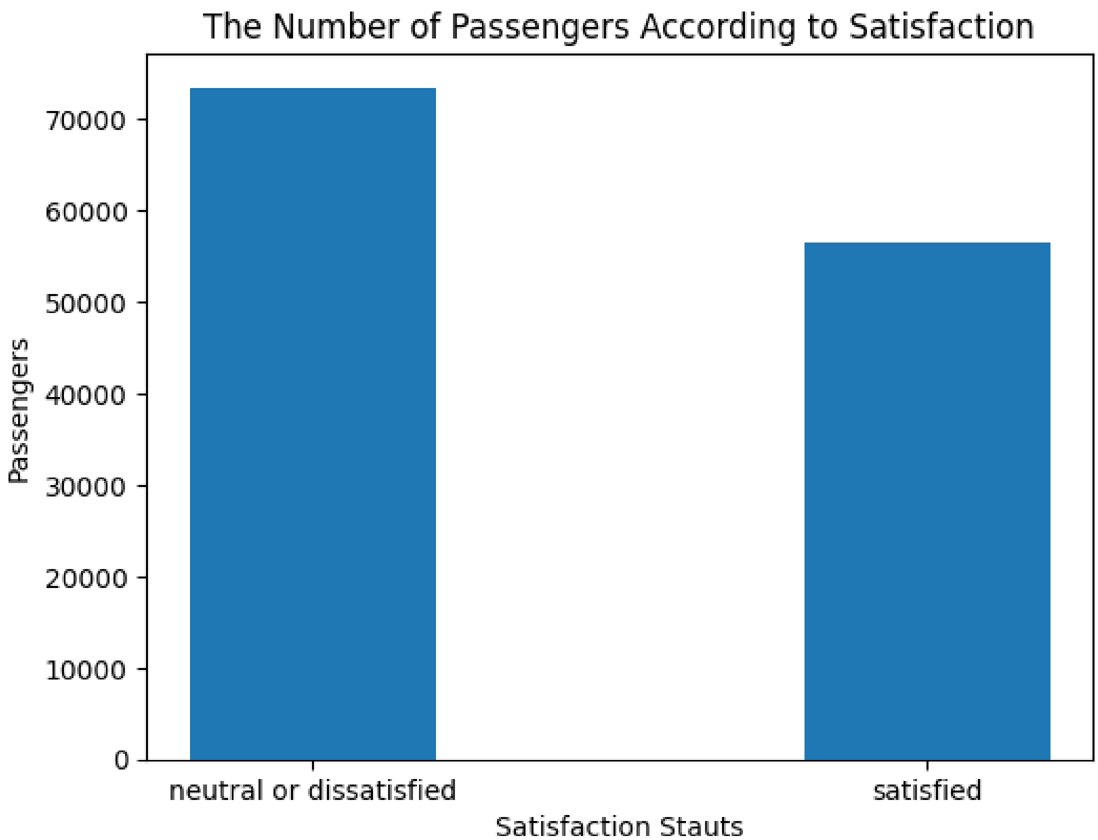
I'll start by looking at the distribution of the main variable of interest: satisfaction.

```
In [16]: # This function is used to setup all charts' details such as title and labels
         def drawChart(title, xlabelText, ylabelText, xticksRotate ):
             plt.title(title)
             # plt.grid(zorder=0)
             plt.xticks(rotation=xticksRotate)
             plt.xlabel(xlabelText)
             plt.ylabel(ylabelText)
             plt.show()
```

1.4.1 1 Distribution of Satisfaction

1.1 Question How many passengers are satisfied, and how many are not(neutral or dissatisfied)? #### 1.2 Visualization

```
In [17]: satisfaction = dataframe['Satisfaction'].value_counts()
         plt.bar(pd.Series(satisfaction.index),satisfaction.values,
                 width = 0.4, zorder=3)
         drawChart( "The Number of Passengers According to Satisfaction", "Satisfaction Stauts",
         satisfaction
```




```
Out[17]: neutral or dissatisfied    73452
         satisfied                  56428
         Name: Satisfaction, dtype: int64
```

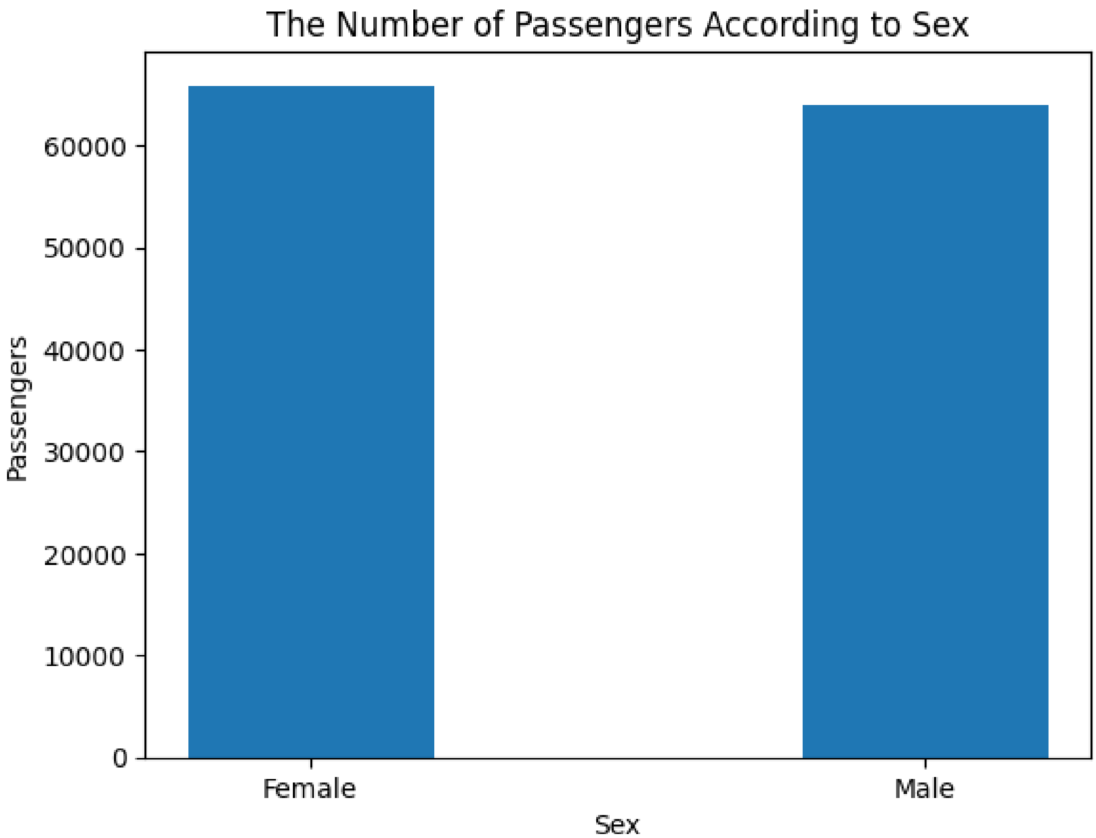
1.4.2 1.3 Observation

According to our previous analysis, the number of passengers with neutral or dissatisfied opinions is 17024 more compared to satisfied passengers, which seems reasonable since both neutral and dissatisfied opinions are reflected.

1.4.3 2 Distribution of sex

2.1 Question How many males and females in the dataset? #### 2.2 Visualization

```
In [18]: sex = dataframe['Gender'].value_counts()
         plt.bar(pd.Series(sex.index),sex.values, width = 0.4)
         drawChart( "The Number of Passengers According to Sex", "Sex" , "Passengers", 0)
         sex
```




```
Out[18]: Female    65899
         Male      63981
         Name: Gender, dtype: int64
```

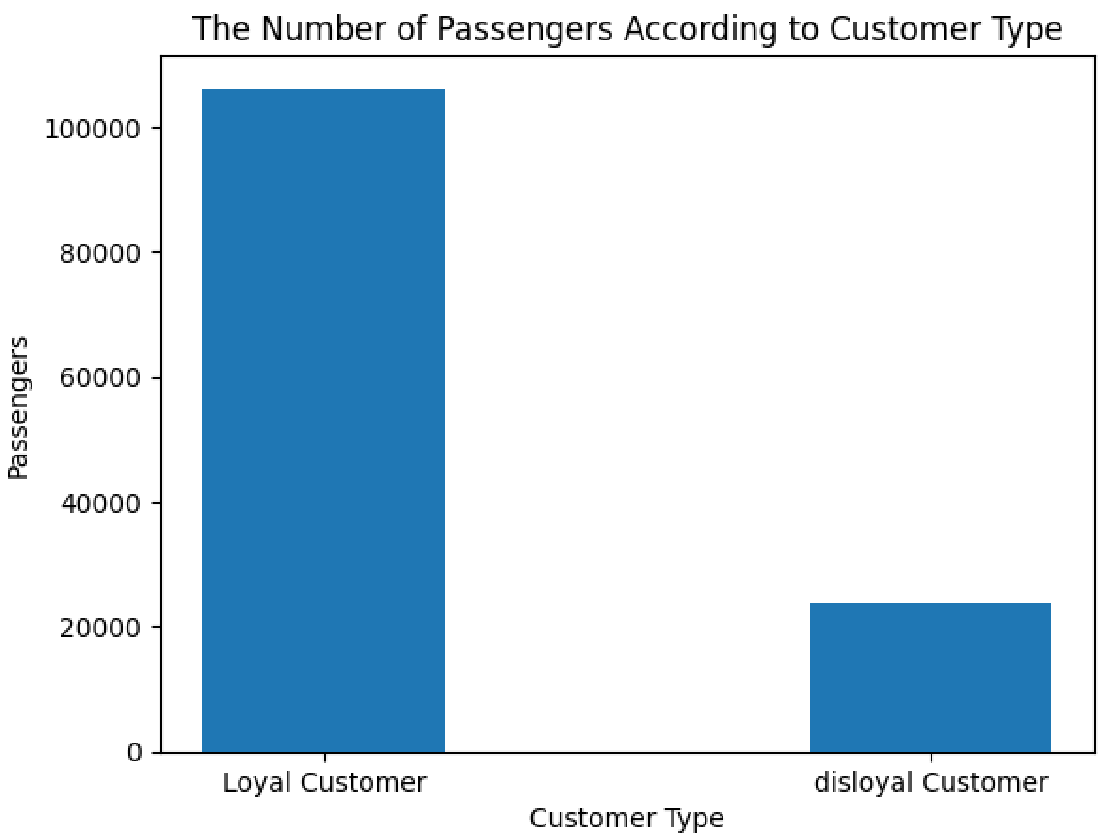
1.4.4 2.3 Observation

The number of females and males is close, but the females are about more with 1918 passengers.

1.4.5 3 Distribution of Customer Type

3.1 Question How many loyal and disloyal passengers in the dataset? #### 3.2 Visualization

```
In [19]: customer_Type = dataframe['Customer Type'].value_counts()
         plt.bar(pd.Series(customer_Type.index),customer_Type.values,
                 width = 0.4, zorder=3)
         drawChart( "The Number of Passengers According to Customer Type", "Customer Type" , "Pa
         customer_Type
```




```
Out[19]: Loyal Customer       106100
         disloyal Customer     23780
         Name: Customer Type, dtype: int64
```

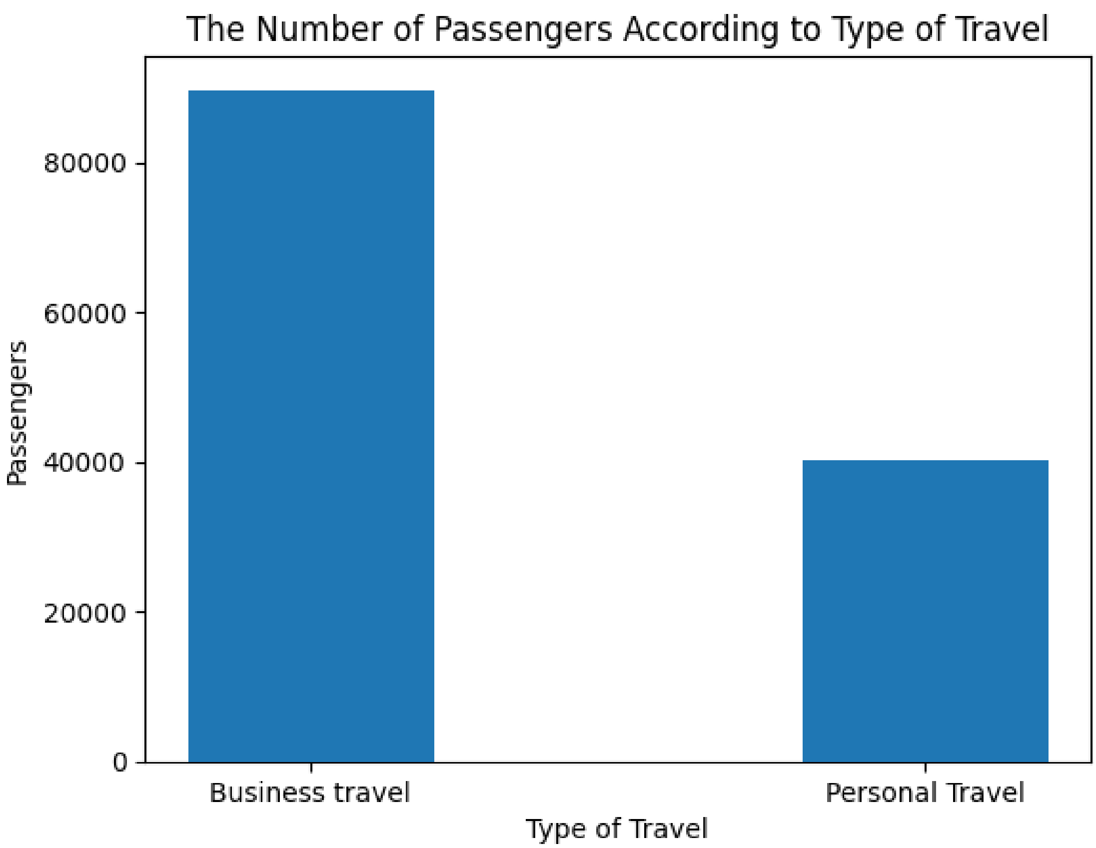
1.4.6 3.3 Observation

The number of loyal customers is more than disloyal by about 4 times.

1.4.7 4 Distribution of Type of Travel

4.1 Question How many business and personal travels do we have? #### 4.2 Visualization

```
In [20]: travel_type = dataframe['Type of Travel'].value_counts()
         plt.bar(pd.Series(travel_type.index), travel_type.values,
                 width = 0.4, zorder=3)
         drawChart( "The Number of Passengers According to Type of Travel", "Type of Travel" , "
         travel_type
```



Out[20]:
```
Business travel    89693
Personal Travel    40187
Name: Type of Travel, dtype: int64
```

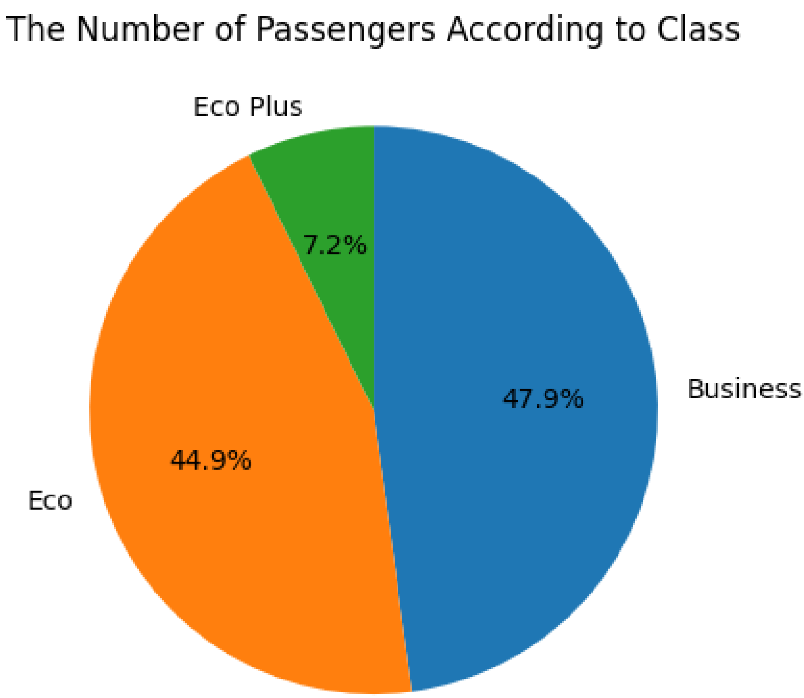
1.4.8 4.3 Observation

Number of business travels is more than double of personal travels.

1.4.9 5 Distribution of Class

5.1 Question How many passengers we have in busniss, eco plus, and ceo classes in the dataset? #### 5.2 Visualization

In [21]:
```
travel_class = dataframe['Class'].value_counts()
plt.pie( travel_class.values, labels=pd.Series(travel_class.index),
        startangle = 90, counterclock = False, autopct   = '%.1f%%')
drawChart("The Number of Passengers According to Class", "" , "", 0)
travel_class
```




```
Out[21]: Business    62160
         Eco         58309
         Eco Plus     9411
         Name: Class, dtype: int64
```

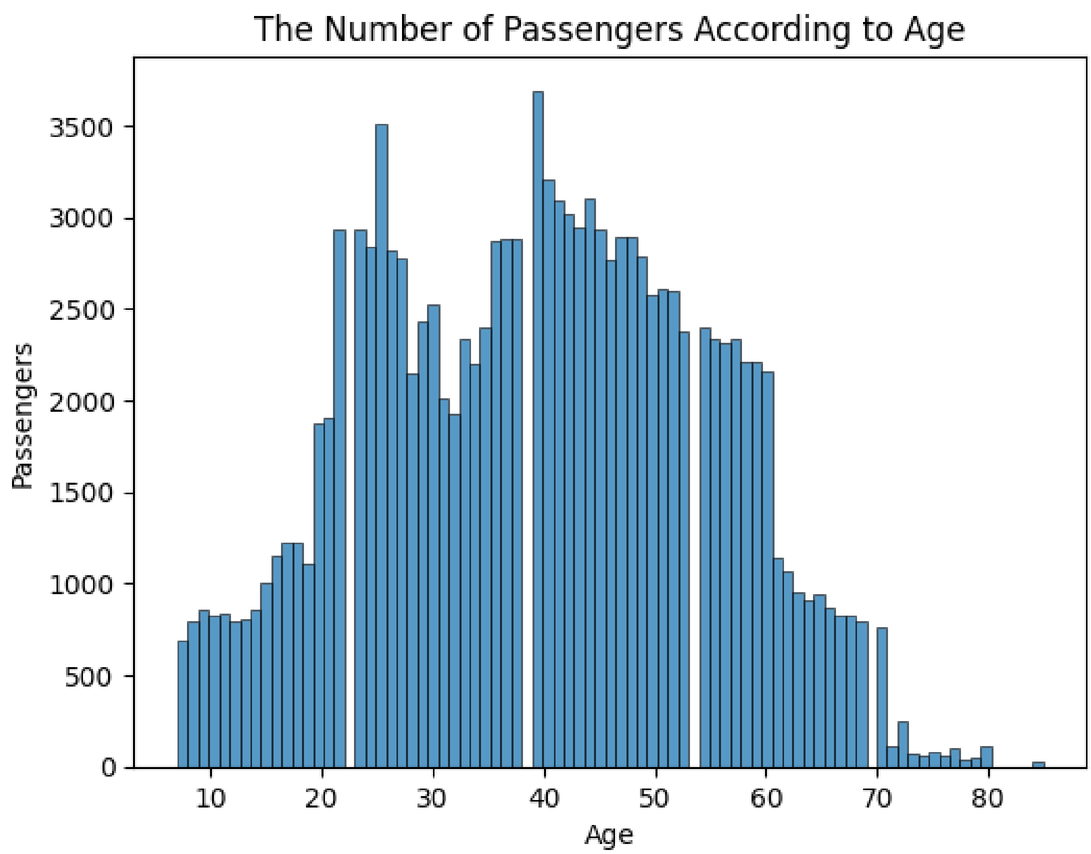
1.4.10 5.3 Observation

The number of Business flights is too close to the number of Eco class, but the number of Eco Plus is very small.

1.4.11 6 Distribution of Age

6.1 Question What are the ages of passengers? #### 6.2 Visualization

```
In [22]: # dataframe.hist(column=["Age"], grid=False, range=[1, 100])
         # sb.catplot(x="Age", data=dataframe, aspect=3.0, kind='count', hue='Satisfaction', ord
         sb.histplot( x= "Age", data = dataframe)
         drawChart("The Number of Passengers According to Age", "Age" , "Passengers", 0)
```



1.4.12 6.3 Observation

Age has a very short-tailed distribution on the high age end, with 2 peraks, so it can be concidered as Bimodal. Most of data is between 20-60, with one peak between 20 and 30, and a second peak a little below 40. Interestingly, there're 4 gaps at 23, 38, 54, 69, and over 80.

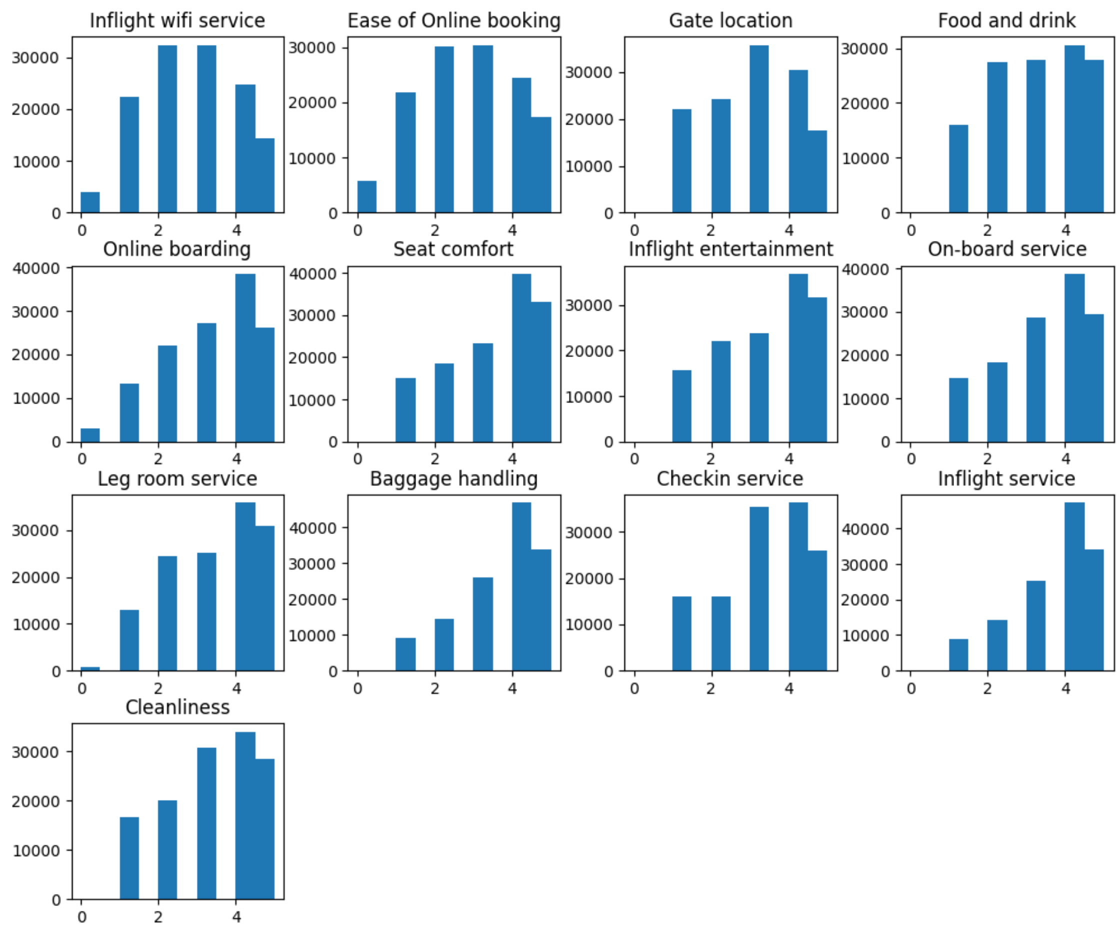
1.4.13 7 Distribution of Passengers' Opinions Against Survey's Questions

7.1 Question What are the distributions of passengers' opinions against survey's questions? #### 7.2 Visualization

```
In [23]: dataframe.hist(column=["Inflight wifi service",
                                "Ease of Online booking", "Gate location", "Food and drink", "On
                                "Seat comfort", "Inflight entertainment", "On-board service",
                                "Leg room service", "Baggage handling", "Checkin service", "Infl
                                "Cleanliness"], range=[0, 5], figsize=(12,10), grid=False)
```

```
Out[23]: array([[<AxesSubplot: title={'center': 'Inflight wifi service'}>,
                 <AxesSubplot: title={'center': 'Ease of Online booking'}>,
                 <AxesSubplot: title={'center': 'Gate location'}>,
```

```
 <AxesSubplot: title={'center': 'Food and drink'}>],
[<AxesSubplot: title={'center': 'Online boarding'}>,
 <AxesSubplot: title={'center': 'Seat comfort'}>,
 <AxesSubplot: title={'center': 'Inflight entertainment'}>,
 <AxesSubplot: title={'center': 'On-board service'}>],
[<AxesSubplot: title={'center': 'Leg room service'}>,
 <AxesSubplot: title={'center': 'Baggage handling'}>,
 <AxesSubplot: title={'center': 'Checkin service'}>,
 <AxesSubplot: title={'center': 'Inflight service'}>],
[<AxesSubplot: title={'center': 'Cleanliness'}>, <AxesSubplot: >,
 <AxesSubplot: >, <AxesSubplot: >]], dtype=object)
```



7.3 Observation The distributions seem acceptable for the survey questions. Later on, I'll investigate 3 questions in more details.

1.4.14 Discuss the distribution(s) of your variable(s) of interest. Were there any unusual points? Did you need to perform any transformations?

As shown above, the satisfaction variable took only 2 values (binary classification). There is no need for any transformation. Also for the other variables we dont need to apply transformation.

1.4.15 Of the features you investigated, were there any unusual distributions? Did you perform any operations on the data to tidy, adjust, or change the form of the data? If so, why did you do this?

No, there is no unusual distributions and I did not make any operations to resolve the issue.

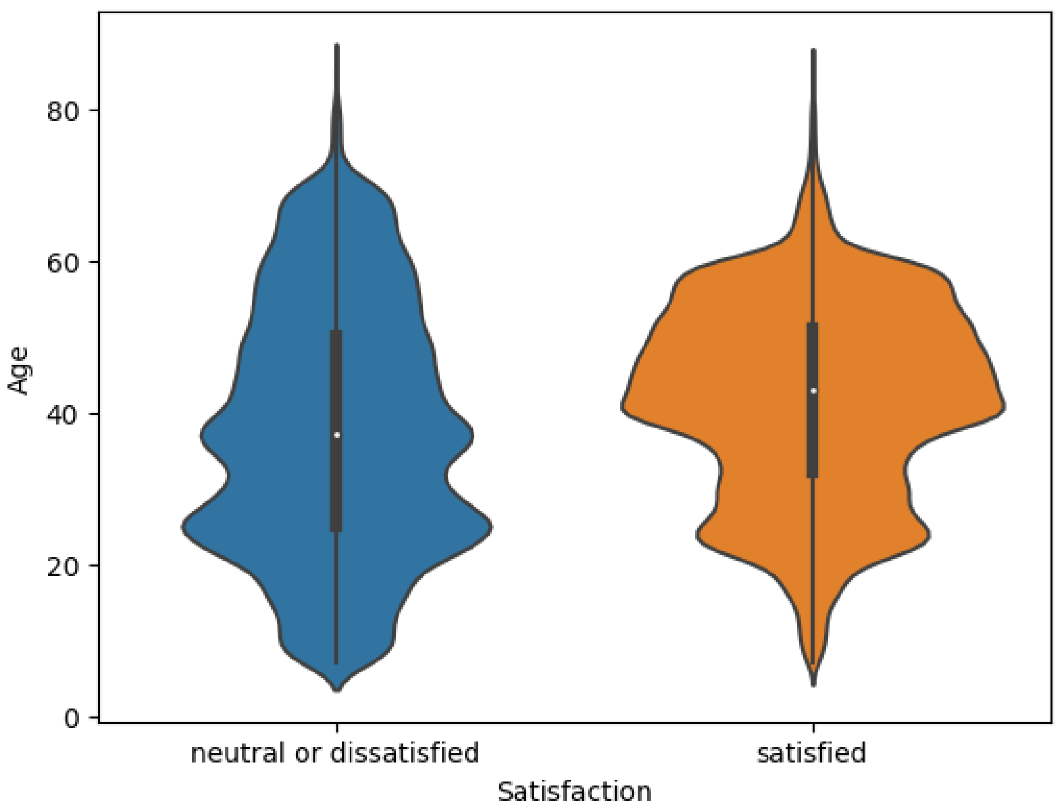
1.5 Bivariate Exploration

1.5.1 8 Correlation of Satisfaction and Age

8.1 Question Which ages are more like to be satisfied or neutral or dissatisfied? #### 8.2 Visualization

```
In [24]: sb.violinplot(data = dataframe, x = 'Satisfaction', y = 'Age')
```

```
Out[24]: <AxesSubplot: xlabel='Satisfaction', ylabel='Age'>
```



1.5.2 8.3 Observation

Most of correlation between age and neutral or dissatisfied is beteewn 20 and 30, then between 30 and 40. We found less correlation over 40s for neutral or dissatisfied. While the satisfied opinions are more between 40 and below 60, then between 20 and 25. Also, we have outliers over 70.

1.5.3 9 Correlation of Satisfaction and Sex

9.1 Question Which gender is more satisfied, neutral, or dissatisfied? #### 9.2 Visualization

```
In [25]: g = sb.FacetGrid(dataframe, col="Satisfaction", hue="Gender")
         g.map(sb.countplot, "Gender", order=["Male", "Female"])
         g.add_legend()
```

```
Out[25]: <seaborn.axisgrid.FacetGrid at 0x1fe25bf0a30>
```

1.5.4 9.3 Observation

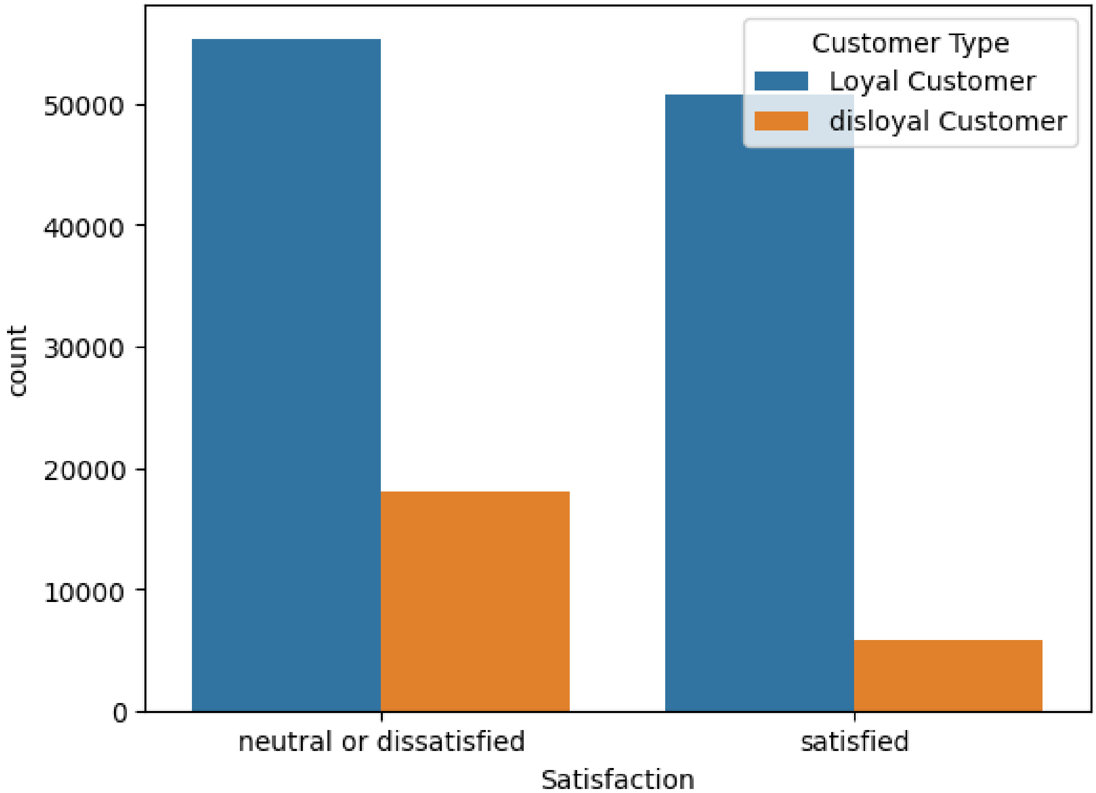
The number of neutral or dissatisfied opinions for females is larger than for males. while in satisfied, we see a similar number of observations for both males and females.

1.5.5 10 Correlation of Satisfaction and Customer Type

10.1 Question What is the relation between loyality and satisfaction? #### 10.2 Visualization

```
In [26]: sb.countplot(data = dataframe, x = 'Satisfaction', hue = 'Customer Type')
```

```
Out[26]: <AxesSubplot: xlabel='Satisfaction', ylabel='count'>
```



1.5.6 10.3 Observation

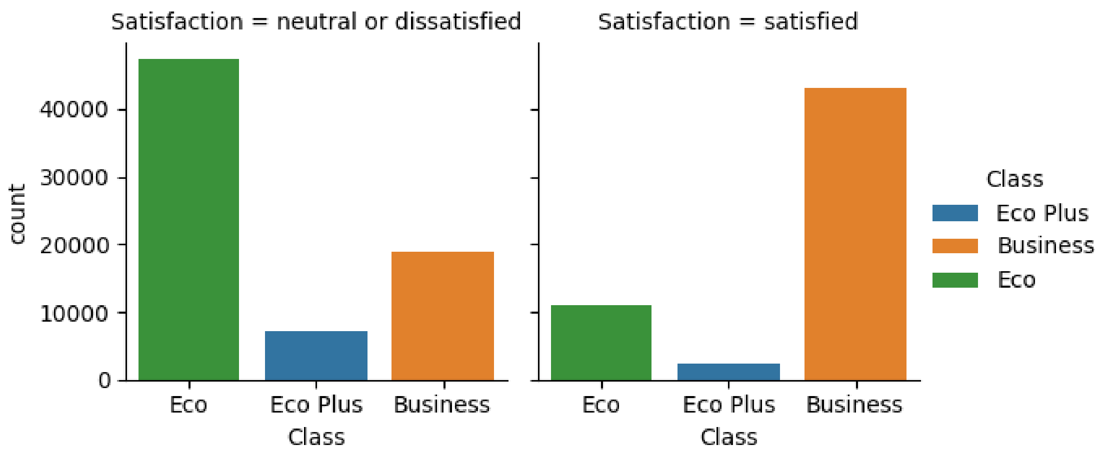
As shown before, the number of loyal customers is larger than disloyal. So, here we found that the number of neutral or dissatisfied opinions for the loyal customers is much larger than disloyal. Also, we see a larger difference between satisfied loyal customers and disloyal.

1.5.7 11 Correlation of Satisfaction and Class

11.1 Question Is there a coorelation between Satisfaction and Class? #### 11.2 Visualization

```
In [27]: g = sb.FacetGrid(dataframe, col="Satisfaction", hue="Class")
         g.map(sb.countplot, "Class", order=["Eco", "Eco Plus", "Business"])
         g.add_legend()

Out[27]: <seaborn.axisgrid.FacetGrid at 0x1fe2ab43fa0>
```



1.5.8 11.3 Observation

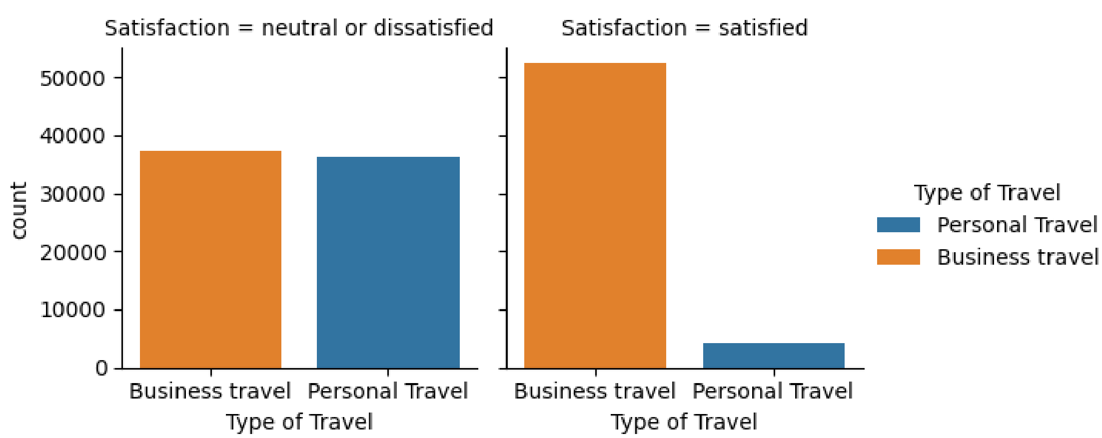
Interestingly, a huge number (40000+) of passengers from Eco class were neutral or dissatisfied against only about 10000 satisfied. On the contrary, a huge number (40000+) of passengers from the Business class were satisfied against only about 20000 neutral or dissatisfied. Finally, most of the passengers from Eco Plus were neutral or dissatisfied.

1.5.9 12 Correlation of Satisfaction and Type of Travel

12.1 Question What is the relationship between Type of Travel and Satisfaction? #### 12.2 Visualization

```
In [28]: g = sb.FacetGrid(dataframe, col="Satisfaction", hue="Type of Travel")
         g.map(sb.countplot, "Type of Travel", order=["Business travel", "Personal Travel"])
         g.add_legend()
```

```
Out[28]: <seaborn.axisgrid.FacetGrid at 0x1fe24b50880>
```



1.5.10 12.3 Observation

Here, we have more than 35000 of Business travelers were neutral or dissatisfied, and more than 50000 were satisfied. On the other hand, most of personal travelers passengers were neutral or dissatisfied against less than 50000 satisfied.

1.5.11 Talk about some of the relationships you observed in this part of the investigation. How did the feature(s) of interest vary with other features in the dataset?

As displayed in previous charts, the strongest correlation was between Travel Type and Satisfaction. Furthermore, we found a huge number of passengers from Eco class tend to be neutral or dissatisfied against only about less than a quarter of them were satisfied. Contrarily, a significant portion of business class passengers reported being satisfied, as compared to only approximately half of them who reported being neutral or dissatisfied. Finally, the majority of Eco Plus passengers expressed neutral or negative satisfaction.

1.5.12 Did you observe any interesting relationships between the other features (not the main feature(s) of interest)?

No, I was focused on relationships with satisfaction.

1.6 Multivariate Exploration

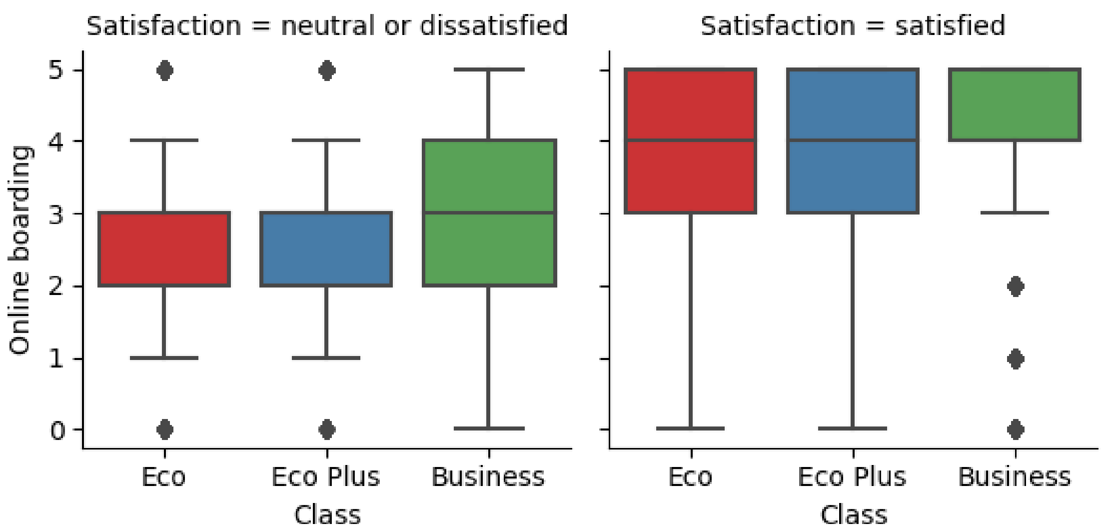
Next, I'll investigate the relationships between 2 randomly selected factors from survey with the class and satisfaction. Beside, 1 another factor from survey with the type of travel and satisfaction

1.6.1 13 Relationship of Online boarding, Class, and Satisfaction

13.1 Question What is the relationship between Online boarding, Class, and Satisfaction? #### 13.2 Visualization

```
In [29]: g = sb.FacetGrid(data = dataframe, col = 'Satisfaction')
         g.map(sb.boxplot, 'Class', 'Online boarding', order=["Eco", "Eco Plus", "Business"], pa
```

```
Out[29]: <seaborn.axisgrid.FacetGrid at 0x1fe27c4e1a0>
```



1.6.2 13.3 Observation

For neutral or dissatisfied & Online Boarding: * The box plots for Eco and Eco Plus classes suggest that overall passengers have a high level of agreement with each other on online boarding (2-3) with small number of outliers. * For Business class, the observations have a more diverse (2-4) with no outliers.

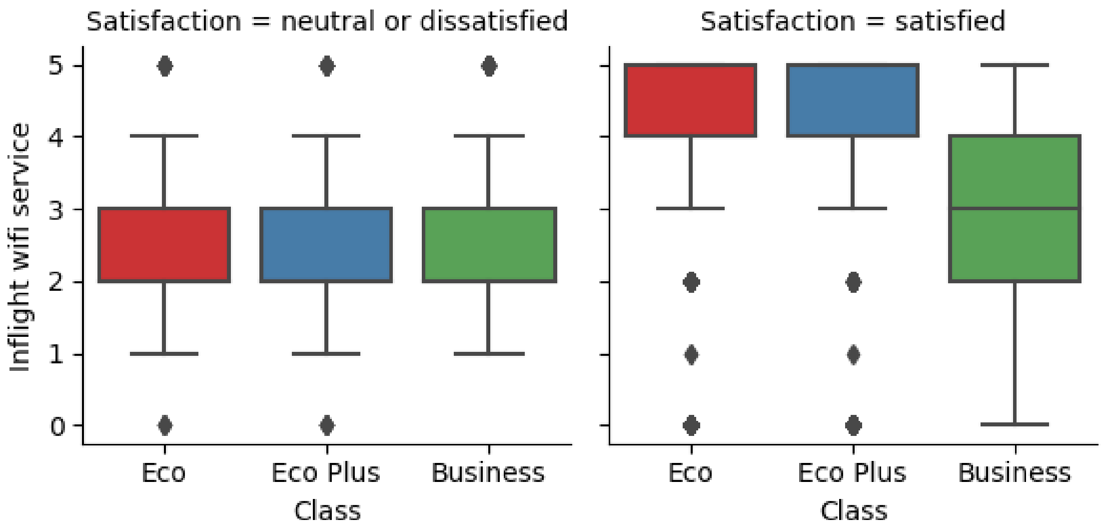
For satisfied & Online Boarding: * The box plots for Eco and Eco Plus classes suggest that observations have a wider range of opinions (3-5) with no outliers. * For Business class, it shows high level of agreement (4-5) with some outliers.

1.6.3 14 Relationship of Inflight wifi service, Class, and Satisfaction

14.1 Question What is the relationship between Inflight wifi service, Class, and Satisfaction? #### 14.2 Visualization

```
In [30]: g = sb.FacetGrid(data = dataframe, col = 'Satisfaction')
         g.map(sb.boxplot, 'Class', 'Inflight wifi service', order=["Eco", "Eco Plus", "Business
```

```
Out[30]: <seaborn.axisgrid.FacetGrid at 0x1fe27c72e60>
```



1.6.4 14.3 Observation

For neutral or dissatisfied about Inflight wifi service: * The three classes: Eco, Eco Plus, and Business show that overall passengers have a high level of agreement with each other regarding the inflight wifi service (2-3) with small number of outliers.

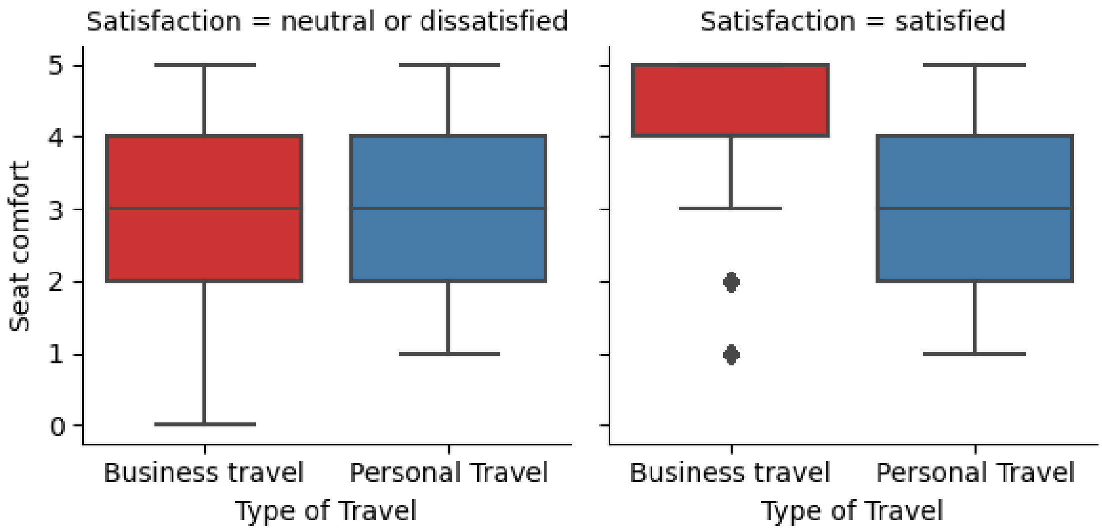
For satisfied about Inflight wifi service: * The box plots for Eco and Eco Plus classes suggest that overall passengers have a high level of agreement on high satisfaction of the inflight wifi service with some outliers. * For Business class, it shows more diversity of opionins and median of 3, with no outliers.

1.6.5 15 Relationship of Inflight wifi service, Type of Travel, and Satisfaction

15.1 Question What is the relationship between Inflight wifi service, Type of Travel, and Satisfaction? #### 15.2 Visualization

```
In [31]: g = sb.FacetGrid(data = dataframe, col = 'Satisfaction')
         g.map(sb.boxplot, 'Type of Travel', 'Seat comfort', order = ["Business travel", "Person
```

```
Out[31]: <seaborn.axisgrid.FacetGrid at 0x1fe280e6d40>
```



1.6.6 15.3 Observation

For neutral or dissatisfied & Seat comfort: * Both types of Business and Personal travels show that passengers' opinions mostly range from 2-4.

For satisfied & Seat comfort: * For Business travel, it shows more agreement on opionins (4-5) with some outliers. * For Personal travel, it shows that passengers' opinions mostly range from 2-4 with median of 3.

1.6.7 Talk about some of the relationships you observed in this part of the investigation. Were there features that strengthened each other in terms of looking at your feature(s) of interest?

- Referring to 13.2, most of passengers from Business class tend to like the online boarding (4-5) if they were satisfied. On the contrary, their opinins range from 2-4 if they were neutral or dissatisfied.

1.6.8 Were there any interesting or surprising interactions between features?

- Referring to 15.2, the outliers in satisfied Business passengers who tend to gave the Seat comfort 1 & 2 scores were surprising for me.

1.7 Conclusions

Here is a summary of the main findings: * The number of passengers with neutral or dissatisfied opinions is 17024 more compared to satisfied passengers, which seems reasonable since both neutral and dissatisfied opinions are reflected. * The number of females and males is close, but the females are about more with 1918 passengers. * The number of loyal customers is more than disloyal by about 4 times. * Number of business travels is more than double of personal travels. * The number of Business flights is too close to the number of Eco class, but the number of Eco Plus is very small. * A huge number (40000+) of passengers from Eco class were neutral or dissatisfied against only about 10000 satisfied. * A huge number (40000+) of passengers from the Business class were satisfied against only about 20000 neutral or dissatisfied. * Most of the passengers from Eco Plus were neutral or dissatisfied. * More than 35000 of Business travelers were neutral or dissatisfied, and more than 50000 were satisfied. * Most of personal travelers passengers were neutral or dissatisfied against less than 50000 satisfied.