Airline_Passenger_Satisfaction

April 29, 2024

1 Flight Passenger Satisfaction Predictor

Dataset: https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction

2 Import Libraries

[3]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import warnings

from sklearn.preprocessing import OneHotEncoder
from sklearn.preprocessing import OrdinalEncoder
from sklearn.preprocessing import StandardScaler
warnings.filterwarnings('ignore')
```

3 Import Data and Quick Snapshot

---

[4]:
```
#df_train = pd.read_csv('https://drive.google.com/uc?
↪export=download&id=1Vcl5dxH7xKarsGRMQMFxDzDsiOTlPwA5', index_col=[0])
#df_test = pd.read_csv('https://drive.google.com/uc?
↪export=download&id=1XqlCyshe7u0G29COgjLqa88w6FhvXdG8', index_col=[0])
df_train = pd.read_csv('./train.csv', index_col=[0])
df_test = pd.read_csv('./test.csv', index_col=[0])
```

[5]:
```
display(df_train)
```

[5]:
```
            id  Gender      Customer Type  Age   Type of Travel     Class  \
0        70172    Male     Loyal Customer   13  Personal Travel  Eco Plus
1         5047    Male  disloyal Customer   25  Business travel  Business
2       110028  Female     Loyal Customer   26  Business travel  Business
3        24026  Female     Loyal Customer   25  Business travel  Business
```

```
4       119299    Male     Loyal Customer   61  Business travel  Business
...        ...     ...                ...  ...              ...       ...
103899   94171  Female  disloyal Customer   23  Business travel       Eco
103900   73097    Male     Loyal Customer   49  Business travel  Business
103901   68825    Male  disloyal Customer   30  Business travel  Business
103902   54173  Female  disloyal Customer   22  Business travel       Eco
103903   62567    Male     Loyal Customer   27  Business travel  Business

        Flight Distance  Inflight wifi service  \
0                   460                      3
1                   235                      3
2                  1142                      2
3                   562                      2
4                   214                      3
...                 ...                    ...
103899              192                      2
103900             2347                      4
103901             1995                      1
103902             1000                      1
103903             1723                      1

        Departure/Arrival time convenient  Ease of Online booking  ...  \
0                                       4                       3  ...
1                                       2                       3  ...
2                                       2                       2  ...
3                                       5                       5  ...
4                                       3                       3  ...
...                                   ...                     ...  ...
103899                                  1                       2  ...
103900                                  4                       4  ...
103901                                  1                       1  ...
103902                                  1                       1  ...
103903                                  3                       3  ...

        Inflight entertainment  On-board service  Leg room service  \
0                            5                 4                 3
1                            1                 1                 5
2                            5                 4                 3
3                            2                 2                 5
4                            3                 3                 4
...                        ...               ...               ...
103899                       2                 3                 1
103900                       5                 5                 5
103901                       4                 3                 2
103902                       1                 4                 5
103903                       1                 1                 1
```

```
        Baggage handling  Checkin service  Inflight service  Cleanliness  \
0                      4                4                 5            5
1                      3                1                 4            1
2                      4                4                 4            5
3                      3                1                 4            2
4                      4                3                 3            3
...                  ...              ...               ...          ...
103899                 4                2                 3            2
103900                 5                5                 5            4
103901                 4                5                 5            4
103902                 1                5                 4            1
103903                 4                4                 3            1

        Departure Delay in Minutes  Arrival Delay in Minutes  \
0                               25                      18.0
1                                1                       6.0
2                                0                       0.0
3                               11                       9.0
4                                0                       0.0
...                            ...                       ...
103899                           3                       0.0
103900                           0                       0.0
103901                           7                      14.0
103902                           0                       0.0
103903                           0                       0.0

                   satisfaction
0       neutral or dissatisfied
1       neutral or dissatisfied
2                     satisfied
3       neutral or dissatisfied
4                     satisfied
...                         ...
103899  neutral or dissatisfied
103900                satisfied
103901  neutral or dissatisfied
103902  neutral or dissatisfied
103903  neutral or dissatisfied

[103904 rows x 24 columns]
```

[6]:
```
display(df_train.head())
```

[6]:
```
       id  Gender      Customer Type  Age   Type of Travel     Class  \
0   70172    Male     Loyal Customer   13  Personal Travel  Eco Plus
1    5047    Male  disloyal Customer   25  Business travel  Business
2  110028  Female     Loyal Customer   26  Business travel  Business
```

```
3   24026  Female     Loyal Customer   25  Business travel   Business
4  119299    Male     Loyal Customer   61  Business travel   Business

   Flight Distance  Inflight wifi service  Departure/Arrival time convenient  \
0              460                      3                                  4
1              235                      3                                  2
2             1142                      2                                  2
3              562                      2                                  5
4              214                      3                                  3

   Ease of Online booking  ...  Inflight entertainment  On-board service  \
0                       3  ...                       5                 4
1                       3  ...                       1                 1
2                       2  ...                       5                 4
3                       5  ...                       2                 2
4                       3  ...                       3                 3

   Leg room service  Baggage handling  Checkin service  Inflight service  \
0                 3                 4                4                 5
1                 5                 3                1                 4
2                 3                 4                4                 4
3                 5                 3                1                 4
4                 4                 4                3                 3

   Cleanliness  Departure Delay in Minutes  Arrival Delay in Minutes  \
0            5                          25                      18.0
1            1                           1                       6.0
2            5                           0                       0.0
3            2                          11                       9.0
4            3                           0                       0.0

              satisfaction
0  neutral or dissatisfied
1  neutral or dissatisfied
2                satisfied
3  neutral or dissatisfied
4                satisfied

[5 rows x 24 columns]
```

3.1 Info and Shape

[7]:
```python
display(df_train.info())

print("\n\nShape of Training Data: ", df_train.shape)
print("Shape of Testing Data: ", df_test.shape)
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 103904 entries, 0 to 103903
Data columns (total 24 columns):
 #   Column                             Non-Null Count   Dtype
---  ------                             --------------   -----
 0   id                                 103904 non-null  int64
 1   Gender                             103904 non-null  object
 2   Customer Type                      103904 non-null  object
 3   Age                                103904 non-null  int64
 4   Type of Travel                     103904 non-null  object
 5   Class                              103904 non-null  object
 6   Flight Distance                    103904 non-null  int64
 7   Inflight wifi service              103904 non-null  int64
 8   Departure/Arrival time convenient  103904 non-null  int64
 9   Ease of Online booking             103904 non-null  int64
 10  Gate location                      103904 non-null  int64
 11  Food and drink                     103904 non-null  int64
 12  Online boarding                    103904 non-null  int64
 13  Seat comfort                       103904 non-null  int64
 14  Inflight entertainment             103904 non-null  int64
 15  On-board service                   103904 non-null  int64
 16  Leg room service                   103904 non-null  int64
 17  Baggage handling                   103904 non-null  int64
 18  Checkin service                    103904 non-null  int64
 19  Inflight service                   103904 non-null  int64
 20  Cleanliness                        103904 non-null  int64
 21  Departure Delay in Minutes         103904 non-null  int64
 22  Arrival Delay in Minutes           103594 non-null  float64
 23  satisfaction                       103904 non-null  object
dtypes: float64(1), int64(18), object(5)
memory usage: 19.8+ MB
```

[7]: None

```
Shape of Training Data:  (103904, 24)
Shape of Testing Data:  (25976, 24)
```

3.2 DF Describe

[8]:
```
display(df_train.describe())
```

[8]:

	id	Age	Flight Distance	Inflight wifi service	\
count	103904.000000	103904.000000	103904.000000	103904.000000	
mean	64924.210502	39.379706	1189.448375	2.729683	
std	37463.812252	15.114964	997.147281	1.327829	
min	1.000000	7.000000	31.000000	0.000000	

25%	32533.750000	27.000000	414.000000	2.000000
50%	64856.500000	40.000000	843.000000	3.000000
75%	97368.250000	51.000000	1743.000000	4.000000
max	129880.000000	85.000000	4983.000000	5.000000

	Departure/Arrival time convenient	Ease of Online booking \
count	103904.000000	103904.000000
mean	3.060296	2.756901
std	1.525075	1.398929
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	103904.000000	103904.000000	103904.000000	103904.000000
mean	2.976883	3.202129	3.250375	3.439396
std	1.277621	1.329533	1.349509	1.319088
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	103904.000000	103904.000000	103904.000000
mean	3.358158	3.382363	3.351055
std	1.332991	1.288354	1.315605
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	103904.000000	103904.000000	103904.000000	103904.000000
mean	3.631833	3.304290	3.640428	3.286351
std	1.180903	1.265396	1.175663	1.312273
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	103904.000000	103594.000000

```
mean           14.815618      15.178678
std            38.230901      38.698682
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
```

4 Data Cleaning and Preprocessing

---

[9]:
```python
# Replace spaces in column names with underscore

df_train.columns = df_train.columns.str.replace(' ', '_')
df_test.columns = df_test.columns.str.replace(' ', '_')
```

[10]:
```python
# Drop the uneeded ID Column

df_train.drop(columns=['id'], inplace=True)
df_test.drop(columns=['id'], inplace=True)
```

4.1 Handling Null Values

4.1.1 Display Null Values

[11]:
```python
# Display the Null Values

display(df_train.isnull().sum())
display(df_test.isnull().sum())
display(df_train[pd.isnull(df_train).any(axis=1)])
```

[11]:
```
Gender                                0
Customer_Type                         0
Age                                   0
Type_of_Travel                        0
Class                                 0
Flight_Distance                       0
Inflight_wifi_service                 0
Departure/Arrival_time_convenient     0
Ease_of_Online_booking                0
Gate_location                         0
Food_and_drink                        0
Online_boarding                       0
Seat_comfort                          0
Inflight_entertainment                0
On-board_service                      0
```

```
Leg_room_service                      0
Baggage_handling                      0
Checkin_service                       0
Inflight_service                      0
Cleanliness                           0
Departure_Delay_in_Minutes            0
Arrival_Delay_in_Minutes            310
satisfaction                          0
dtype: int64
```

[11]:
```
Gender                               0
Customer_Type                        0
Age                                  0
Type_of_Travel                       0
Class                                0
Flight_Distance                      0
Inflight_wifi_service                0
Departure/Arrival_time_convenient    0
Ease_of_Online_booking               0
Gate_location                        0
Food_and_drink                       0
Online_boarding                      0
Seat_comfort                         0
Inflight_entertainment               0
On-board_service                     0
Leg_room_service                     0
Baggage_handling                     0
Checkin_service                      0
Inflight_service                     0
Cleanliness                          0
Departure_Delay_in_Minutes           0
Arrival_Delay_in_Minutes            83
satisfaction                         0
dtype: int64
```

[11]:
```
        Gender      Customer_Type  Age   Type_of_Travel     Class  \
213     Female     Loyal Customer   38  Business travel       Eco
1124      Male     Loyal Customer   53  Personal Travel       Eco
1529      Male     Loyal Customer   39  Business travel  Business
2004    Female  disloyal Customer   26  Business travel  Business
2108    Female     Loyal Customer   24  Personal Travel       Eco
...        ...                ...  ...              ...       ...
102067    Male     Loyal Customer   49  Personal Travel  Eco Plus
102384    Male     Loyal Customer   58  Business travel       Eco
102552  Female  disloyal Customer   29  Business travel       Eco
102960    Male     Loyal Customer   58  Business travel       Eco
103540  Female     Loyal Customer   33  Personal Travel       Eco
```

```
        Flight_Distance  Inflight_wifi_service  \
213                 109                      5
1124               1012                      3
1529                733                      2
2004               1035                      3
2108                417                      2
...                 ...                    ...
102067             1249                      2
102384              733                      3
102552             1107                      2
102960             1088                      4
103540              359                      4

        Departure/Arrival_time_convenient  Ease_of_Online_booking  \
213                                     3                       3
1124                                    2                       3
1529                                    5                       5
2004                                    3                       3
2108                                    1                       2
...                                   ...                     ...
102067                                  5                       2
102384                                  3                       3
102552                                  1                       1
102960                                  4                       1
103540                                  4                       4

        Gate_location  ...  Inflight_entertainment  On-board_service  \
213                 3  ...                       5                 5
1124                4  ...                       4                 4
1529                5  ...                       2                 2
2004                1  ...                       2                 3
2108                2  ...                       5                 1
...               ...  ...                     ...               ...
102067              3  ...                       3                 4
102384              3  ...                       3                 3
102552              1  ...                       5                 4
102960              1  ...                       5                 1
103540              3  ...                       4                 3

        Leg_room_service  Baggage_handling  Checkin_service  Inflight_service  \
213                    2                 4                1                 1
1124                   4                 4                3                 3
1529                   2                 2                2                 2
2004                   3                 4                5                 5
2108                   4                 2                1                 2
...                  ...               ...              ...               ...
```

```
102067                  5                 4                 3                 4
102384                  1                 2                 4                 2
102552                  1                 5                 5                 3
102960                  5                 5                 5                 3
103540                  2                 5                 3                 5

        Cleanliness  Departure_Delay_in_Minutes  Arrival_Delay_in_Minutes  \
213               5                          31                       NaN
1124              4                          38                       NaN
1529              3                          11                       NaN
2004              2                          41                       NaN
2108              5                           1                       NaN
...             ...                         ...                       ...
102067            3                         230                       NaN
102384            3                          55                       NaN
102552            5                           0                       NaN
102960            5                           0                       NaN
103540            4                          42                       NaN

                   satisfaction
213                   satisfied
1124    neutral or dissatisfied
1529    neutral or dissatisfied
2004    neutral or dissatisfied
2108    neutral or dissatisfied
...                         ...
102067  neutral or dissatisfied
102384  neutral or dissatisfied
102552  neutral or dissatisfied
102960                satisfied
103540  neutral or dissatisfied

[310 rows x 23 columns]
```

4.1.2 Filling Null Values

[12]:
```
# Set the null values of Arrival Delay to Departure Delay since if one departs a␣
↪certain time late, they will most likely arrive that certain time late.         ␣
↪
df_train['Arrival_Delay_in_Minutes'] = df_train['Arrival_Delay_in_Minutes'].
↪fillna(df_train['Departure_Delay_in_Minutes'])
df_test['Arrival_Delay_in_Minutes'] = df_test['Arrival_Delay_in_Minutes'].
↪fillna(df_test['Departure_Delay_in_Minutes'])
```

[13]:
```
# Null Values After fillna
display(df_train.isnull().sum())
```

```
display(df_test.isnull().sum())
```

[13]:
```
Gender                               0
Customer_Type                        0
Age                                  0
Type_of_Travel                       0
Class                                0
Flight_Distance                      0
Inflight_wifi_service                0
Departure/Arrival_time_convenient    0
Ease_of_Online_booking               0
Gate_location                        0
Food_and_drink                       0
Online_boarding                      0
Seat_comfort                         0
Inflight_entertainment               0
On-board_service                     0
Leg_room_service                     0
Baggage_handling                     0
Checkin_service                      0
Inflight_service                     0
Cleanliness                          0
Departure_Delay_in_Minutes           0
Arrival_Delay_in_Minutes             0
satisfaction                         0
dtype: int64
```

[13]:
```
Gender                               0
Customer_Type                        0
Age                                  0
Type_of_Travel                       0
Class                                0
Flight_Distance                      0
Inflight_wifi_service                0
Departure/Arrival_time_convenient    0
Ease_of_Online_booking               0
Gate_location                        0
Food_and_drink                       0
Online_boarding                      0
Seat_comfort                         0
Inflight_entertainment               0
On-board_service                     0
Leg_room_service                     0
Baggage_handling                     0
Checkin_service                      0
Inflight_service                     0
Cleanliness                          0
```

```
Departure_Delay_in_Minutes          0
Arrival_Delay_in_Minutes            0
satisfaction                        0
dtype: int64
```

4.2 Making Capitalization Consistent

[14]:
```
# Adjust capitalization of columns
df_train['Customer_Type'].replace('disloyal Customer', 'Disloyal_Customer',␣
 ↪inplace=True)
df_train['Customer_Type'].replace('Loyal Customer', 'Loyal_Customer',␣
 ↪inplace=True)
df_train['Type_of_Travel'].replace('Business travel', 'Business_Travel',␣
 ↪inplace=True)
df_train['Type_of_Travel'].replace('Personal travel', 'Personal_Travel',␣
 ↪inplace=True)
df_train = df_train.rename(columns={'satisfaction': 'Satisfaction',
                                    'Inflight_wifi_service':␣
 ↪'Inflight_Wifi_Service',
                                    'Departure/Arrival_time_convenient':␣
 ↪'Departure/Arrival_Time_Convenient',
                                    'Ease_of_Online_booking':␣
 ↪'Ease_of_Online_Booking',
                                    'Gate_location': 'Gate_Location',
                                    'Food_and_drink': 'Food_and_Drink',
                                    'Online_boarding': 'Online_Boarding',
                                    'Seat_comfort': 'Seat_Comfort',
                                    'Inflight_entertainment':␣
 ↪'Inflight_Entertainment',
                                    'On-board_service': 'On-board_Service',
                                    'Leg_room_service': 'Leg_Room_Service',
                                    'Baggage_handling': 'Baggage_Handling',
                                    'Checkin_service': 'Checkin_Service',
                                    'Inflight_service': 'Inflight_Service'})
df_train['Satisfaction'].replace('satisfied', 'Satisfied', inplace=True)
df_train['Satisfaction'].replace('neutral or dissatisfied',␣
 ↪'Neutral_Or_Dissatisfied', inplace=True)

df_test['Customer_Type'].replace('disloyal Customer', 'Disloyal_Customer',␣
 ↪inplace=True)
df_test['Customer_Type'].replace('Loyal Customer', 'Loyal_Customer',␣
 ↪inplace=True)
df_test['Type_of_Travel'].replace('Business travel', 'Business_Travel',␣
 ↪inplace=True)
df_test['Type_of_Travel'].replace('Personal travel', 'Personal_Travel',␣
 ↪inplace=True)
```

```
df_test = df_test.rename(columns={'satisfaction': 'Satisfaction',
                                   'Inflight_wifi_service':␣
 ↪'Inflight_Wifi_Service',
                                   'Departure/Arrival_time_convenient':␣
 ↪'Departure/Arrival_Time_Convenient',
                                   'Ease_of_Online_booking':␣
 ↪'Ease_of_Online_Booking',
                                   'Gate_location': 'Gate_Location',
                                   'Food_and_drink': 'Food_and_Drink',
                                   'Online_boarding': 'Online_Boarding',
                                   'Seat_comfort': 'Seat_Comfort',
                                   'Inflight_entertainment':␣
 ↪'Inflight_Entertainment',
                                   'On-board_service': 'On-board_Service',
                                   'Leg_room_service': 'Leg_Room_Service',
                                   'Baggage_handling': 'Baggage_Handling',
                                   'Checkin_service': 'Checkin_Service',
                                   'Inflight_service': 'Inflight_Service'})
df_test['Satisfaction'].replace('satisfied', 'Satisfied', inplace=True)
df_test['Satisfaction'].replace('neutral or dissatisfied',␣
 ↪'Neutral_Or_Dissatisfied', inplace=True)

display(df_train)
unstandardized_df_train = df_train.copy()
```

[14]:

```
        Gender      Customer_Type  Age   Type_of_Travel     Class  \
0         Male     Loyal_Customer   13  Personal Travel  Eco Plus
1         Male  Disloyal_Customer   25  Business_Travel  Business
2       Female     Loyal_Customer   26  Business_Travel  Business
3       Female     Loyal_Customer   25  Business_Travel  Business
4         Male     Loyal_Customer   61  Business_Travel  Business
...        ...                ...  ...              ...       ...
103899  Female  Disloyal_Customer   23  Business_Travel       Eco
103900    Male     Loyal_Customer   49  Business_Travel  Business
103901    Male  Disloyal_Customer   30  Business_Travel  Business
103902  Female  Disloyal_Customer   22  Business_Travel       Eco
103903    Male     Loyal_Customer   27  Business_Travel  Business

        Flight_Distance  Inflight_Wifi_Service  \
0                   460                      3
1                   235                      3
2                  1142                      2
3                   562                      2
4                   214                      3
...                 ...                    ...
103899              192                      2
103900             2347                      4
```

```
103901            1995                       1
103902            1000                       1
103903            1723                       1

        Departure/Arrival_Time_Convenient  Ease_of_Online_Booking  \
0                                       4                       3
1                                       2                       3
2                                       2                       2
3                                       5                       5
4                                       3                       3
...                                   ...                     ...
103899                                  1                       2
103900                                  4                       4
103901                                  1                       1
103902                                  1                       1
103903                                  3                       3

        Gate_Location  ...  Inflight_Entertainment  On-board_Service  \
0                   1  ...                       5                 4
1                   3  ...                       1                 1
2                   2  ...                       5                 4
3                   5  ...                       2                 2
4                   3  ...                       3                 3
...               ...  ...                     ...               ...
103899              3  ...                       2                 3
103900              4  ...                       5                 5
103901              3  ...                       4                 3
103902              5  ...                       1                 4
103903              3  ...                       1                 1

        Leg_Room_Service  Baggage_Handling  Checkin_Service  Inflight_Service  \
0                      3                 4                4                 5
1                      5                 3                1                 4
2                      3                 4                4                 4
3                      5                 3                1                 4
4                      4                 4                3                 3
...                  ...               ...              ...               ...
103899                 1                 4                2                 3
103900                 5                 5                5                 5
103901                 2                 4                5                 5
103902                 5                 1                5                 4
103903                 1                 4                4                 3

        Cleanliness  Departure_Delay_in_Minutes  Arrival_Delay_in_Minutes  \
0                 5                          25                      18.0
1                 1                           1                       6.0
2                 5                           0                       0.0
```

```
3                2                        11                         9.0
4                3                         0                         0.0
...            ...                       ...                         ...
103899           2                         3                         0.0
103900           4                         0                         0.0
103901           4                         7                        14.0
103902           1                         0                         0.0
103903           1                         0                         0.0

                   Satisfaction
0       Neutral_Or_Dissatisfied
1       Neutral_Or_Dissatisfied
2                     Satisfied
3       Neutral_Or_Dissatisfied
4                     Satisfied
...                         ...
103899  Neutral_Or_Dissatisfied
103900                Satisfied
103901  Neutral_Or_Dissatisfied
103902  Neutral_Or_Dissatisfied
103903  Neutral_Or_Dissatisfied

[103904 rows x 23 columns]
```

4.3 Z-Score Standardization

[15]:
```python
# Data Standardization

scaler = StandardScaler()
numCols = df_train.select_dtypes(include=np.number).columns
df_train[numCols] = pd.DataFrame(scaler.fit_transform(df_train[numCols]),␣
↪columns=df_train[numCols].columns)
display(df_train)

df_test[numCols] = pd.DataFrame(scaler.fit_transform(df_test[numCols]),␣
↪columns=df_test[numCols].columns)
```

[15]:
```
        Gender      Customer_Type       Age    Type_of_Travel     Class  \
0         Male     Loyal_Customer -1.745279  Personal Travel  Eco Plus
1         Male  Disloyal_Customer -0.951360  Business_Travel  Business
2       Female     Loyal_Customer -0.885200  Business_Travel  Business
3       Female     Loyal_Customer -0.951360  Business_Travel  Business
4         Male     Loyal_Customer  1.430397  Business_Travel  Business
...        ...                ...       ...              ...       ...
103899  Female  Disloyal_Customer -1.083680  Business_Travel       Eco
103900    Male     Loyal_Customer  0.636478  Business_Travel  Business
103901    Male  Disloyal_Customer -0.620561  Business_Travel  Business
```

```
103902  Female  Disloyal_Customer -1.149840  Business_Travel       Eco
103903    Male     Loyal_Customer -0.819040  Business_Travel  Business

        Flight_Distance  Inflight_Wifi_Service  \
0             -0.731539               0.203579
1             -0.957184               0.203579
2             -0.047584              -0.549533
3             -0.629246              -0.549533
4             -0.978244               0.203579
...                 ...                    ...
103899        -1.000307              -0.549533
103900         1.160869               0.956691
103901         0.807860              -1.302646
103902        -0.189991              -1.302646
103903         0.535081              -1.302646

        Departure/Arrival_Time_Convenient  Ease_of_Online_Booking  \
0                                0.616172                0.173776
1                               -0.695245                0.173776
2                               -0.695245               -0.541060
3                                1.271880                1.603448
4                               -0.039537                0.173776
...                                   ...                     ...
103899                          -1.350954               -0.541060
103900                           0.616172                0.888612
103901                          -1.350954               -1.255895
103902                          -1.350954               -1.255895
103903                          -0.039537                0.173776

        Gate_Location  ...  Inflight_Entertainment  On-board_Service  \
0           -1.547323  ...                1.231704          0.479403
1            0.018094  ...               -1.769081         -1.849161
2           -0.764614  ...                1.231704          0.479403
3            1.583511  ...               -1.018885         -1.072973
4            0.018094  ...               -0.268688         -0.296785
...               ...  ...                     ...               ...
103899       0.018094  ...               -1.018885         -0.296785
103900       0.800803  ...                1.231704          1.255590
103901       0.018094  ...                0.481508         -0.296785
103902       1.583511  ...               -1.769081          0.479403
103903       0.018094  ...               -1.769081         -1.849161

        Leg_Room_Service  Baggage_Handling  Checkin_Service  Inflight_Service  \
0              -0.266840          0.311769         0.549799          1.156436
1               1.253380         -0.535045        -1.821012          0.305848
2              -0.266840          0.311769         0.549799          0.305848
3               1.253380         -0.535045        -1.821012          0.305848
```

```
4               0.493270          0.311769         -0.240472          -0.544740
...                  ...               ...               ...                ...
103899         -1.787061          0.311769         -1.030742          -0.544740
103900          1.253380          1.158582          1.340069           1.156436
103901         -1.026951          0.311769          1.340069           1.156436
103902          1.253380         -2.228672          1.340069           0.305848
103903         -1.787061          0.311769          0.549799          -0.544740

        Cleanliness  Departure_Delay_in_Minutes  Arrival_Delay_in_Minutes  \
0          1.305870                    0.266393                  0.070988
1         -1.742292                   -0.361375                 -0.238223
2          1.305870                   -0.387532                 -0.392828
3         -0.980251                   -0.099805                 -0.160920
4         -0.218211                   -0.387532                 -0.392828
...             ...                         ...                       ...
103899    -0.980251                   -0.309061                 -0.392828
103900     0.543829                   -0.387532                 -0.392828
103901     0.543829                   -0.204433                 -0.032082
103902    -1.742292                   -0.387532                 -0.392828
103903    -1.742292                   -0.387532                 -0.392828

                   Satisfaction
0       Neutral_Or_Dissatisfied
1       Neutral_Or_Dissatisfied
2                     Satisfied
3       Neutral_Or_Dissatisfied
4                     Satisfied
...                         ...
103899  Neutral_Or_Dissatisfied
103900                Satisfied
103901  Neutral_Or_Dissatisfied
103902  Neutral_Or_Dissatisfied
103903  Neutral_Or_Dissatisfied

[103904 rows x 23 columns]
```

4.4 Encoding Categorical Data

4.4.1 Values in Categorical Columns

[16]:

```python
# All the different values for the Categorical Columns

print(df_train['Gender'].unique())
print(df_train['Customer_Type'].unique())
print(df_train['Type_of_Travel'].unique())
print(df_train['Class'].unique())
```

```
['Male' 'Female']
['Loyal_Customer' 'Disloyal_Customer']
['Personal Travel' 'Business_Travel']
['Eco Plus' 'Business' 'Eco']
```

4.4.2 One-Hot Encoding

[17]:
```
# One Hot Encoding for Categorical Features

categorical_columns = ['Gender', 'Customer_Type', 'Type_of_Travel']
ohe = OneHotEncoder(sparse_output=False)
one_hot_encoded = ohe.fit_transform(df_train[categorical_columns])
one_hot_df = pd.DataFrame(one_hot_encoded, columns=ohe.
 ↪get_feature_names_out(categorical_columns))

encoded_train = pd.concat([one_hot_df, df_train], axis=1)
encoded_train = encoded_train.drop(categorical_columns, axis=1)
display(encoded_train)

categorical_columns = ['Gender', 'Customer_Type', 'Type_of_Travel']
ohe = OneHotEncoder(sparse_output=False)
one_hot_encoded = ohe.fit_transform(df_test[categorical_columns])
one_hot_df = pd.DataFrame(one_hot_encoded, columns=ohe.
 ↪get_feature_names_out(categorical_columns))

encoded_test = pd.concat([one_hot_df, df_test], axis=1)
encoded_test = encoded_test.drop(categorical_columns, axis=1)
```

[17]:
```
        Gender_Female  Gender_Male  Customer_Type_Disloyal_Customer  \
0                 0.0          1.0                              0.0
1                 0.0          1.0                              1.0
2                 1.0          0.0                              0.0
3                 1.0          0.0                              0.0
4                 0.0          1.0                              0.0
...               ...          ...                              ...
103899            1.0          0.0                              1.0
103900            0.0          1.0                              0.0
103901            0.0          1.0                              1.0
103902            1.0          0.0                              1.0
103903            0.0          1.0                              0.0

        Customer_Type_Loyal_Customer  Type_of_Travel_Business_Travel  \
0                                1.0                             0.0
1                                0.0                             1.0
2                                1.0                             1.0
3                                1.0                             1.0
4                                1.0                             1.0
```

```
...                              ...                             ...
103899                           0.0                             1.0
103900                           1.0                             1.0
103901                           0.0                             1.0
103902                           0.0                             1.0
103903                           1.0                             1.0

        Type_of_Travel_Personal Travel        Age     Class  Flight_Distance  \
0                                  1.0 -1.745279  Eco Plus        -0.731539
1                                  0.0 -0.951360  Business        -0.957184
2                                  0.0 -0.885200  Business        -0.047584
3                                  0.0 -0.951360  Business        -0.629246
4                                  0.0  1.430397  Business        -0.978244
...                                ...       ...       ...              ...
103899                             0.0 -1.083680       Eco        -1.000307
103900                             0.0  0.636478  Business         1.160869
103901                             0.0 -0.620561  Business         0.807860
103902                             0.0 -1.149840       Eco        -0.189991
103903                             0.0 -0.819040  Business         0.535081

        Inflight_Wifi_Service  ...  Inflight_Entertainment  On-board_Service  \
0                    0.203579  ...                1.231704          0.479403
1                    0.203579  ...               -1.769081         -1.849161
2                   -0.549533  ...                1.231704          0.479403
3                   -0.549533  ...               -1.018885         -1.072973
4                    0.203579  ...               -0.268688         -0.296785
...                       ...  ...                     ...               ...
103899              -0.549533  ...               -1.018885         -0.296785
103900               0.956691  ...                1.231704          1.255590
103901              -1.302646  ...                0.481508         -0.296785
103902              -1.302646  ...               -1.769081          0.479403
103903              -1.302646  ...               -1.769081         -1.849161

        Leg_Room_Service  Baggage_Handling  Checkin_Service  Inflight_Service  \
0              -0.266840          0.311769         0.549799          1.156436
1               1.253380         -0.535045        -1.821012          0.305848
2              -0.266840          0.311769         0.549799          0.305848
3               1.253380         -0.535045        -1.821012          0.305848
4               0.493270          0.311769        -0.240472         -0.544740
...                  ...               ...              ...               ...
103899         -1.787061          0.311769        -1.030742         -0.544740
103900          1.253380          1.158582         1.340069          1.156436
103901         -1.026951          0.311769         1.340069          1.156436
103902          1.253380         -2.228672         1.340069          0.305848
103903         -1.787061          0.311769         0.549799         -0.544740

        Cleanliness  Departure_Delay_in_Minutes  Arrival_Delay_in_Minutes  \
```

```
0          1.305870                       0.266393                     0.070988
1         -1.742292                      -0.361375                    -0.238223
2          1.305870                      -0.387532                    -0.392828
3         -0.980251                      -0.099805                    -0.160920
4         -0.218211                      -0.387532                    -0.392828
...             ...                            ...                          ...
103899    -0.980251                      -0.309061                    -0.392828
103900     0.543829                      -0.387532                    -0.392828
103901     0.543829                      -0.204433                    -0.032082
103902    -1.742292                      -0.387532                    -0.392828
103903    -1.742292                      -0.387532                    -0.392828

                   Satisfaction
0       Neutral_Or_Dissatisfied
1       Neutral_Or_Dissatisfied
2                     Satisfied
3       Neutral_Or_Dissatisfied
4                     Satisfied
...                         ...
103899  Neutral_Or_Dissatisfied
103900                Satisfied
103901  Neutral_Or_Dissatisfied
103902  Neutral_Or_Dissatisfied
103903  Neutral_Or_Dissatisfied

[103904 rows x 26 columns]
```

4.4.3 Label Encoding (Ordinal Encoding)

[18]:
```
# Ordinal Encoding for the Class where the order is Eco, Eco Plus, and then␣
 ↪Business

enc = OrdinalEncoder(categories=[['Eco', 'Eco Plus','Business']])
encoded_train['Class'] = enc.fit_transform(encoded_train.loc[:,['Class']])
enc = OrdinalEncoder(categories=[['Neutral_Or_Dissatisfied', 'Satisfied']])
encoded_train['Satisfaction'] = enc.fit_transform(encoded_train.loc[:
 ↪,['Satisfaction']])

display(encoded_train)

enc = OrdinalEncoder(categories=[['Eco', 'Eco Plus','Business']])
encoded_test['Class'] = enc.fit_transform(encoded_test.loc[:,['Class']])
enc = OrdinalEncoder(categories=[['Neutral_Or_Dissatisfied', 'Satisfied']])
encoded_test['Satisfaction'] = enc.fit_transform(encoded_test.loc[:
 ↪,['Satisfaction']])
```

[18]:

```
        Gender_Female  Gender_Male  Customer_Type_Disloyal_Customer  \
0                 0.0          1.0                              0.0
1                 0.0          1.0                              1.0
2                 1.0          0.0                              0.0
3                 1.0          0.0                              0.0
4                 0.0          1.0                              0.0
...               ...          ...                              ...
103899            1.0          0.0                              1.0
103900            0.0          1.0                              0.0
103901            0.0          1.0                              1.0
103902            1.0          0.0                              1.0
103903            0.0          1.0                              0.0

        Customer_Type_Loyal_Customer  Type_of_Travel_Business_Travel  \
0                                1.0                             0.0
1                                0.0                             1.0
2                                1.0                             1.0
3                                1.0                             1.0
4                                1.0                             1.0
...                              ...                             ...
103899                           0.0                             1.0
103900                           1.0                             1.0
103901                           0.0                             1.0
103902                           0.0                             1.0
103903                           1.0                             1.0

        Type_of_Travel_Personal Travel       Age  Class  Flight_Distance  \
0                                  1.0 -1.745279    1.0        -0.731539
1                                  0.0 -0.951360    2.0        -0.957184
2                                  0.0 -0.885200    2.0        -0.047584
3                                  0.0 -0.951360    2.0        -0.629246
4                                  0.0  1.430397    2.0        -0.978244
...                                ...       ...    ...              ...
103899                             0.0 -1.083680    0.0        -1.000307
103900                             0.0  0.636478    2.0         1.160869
103901                             0.0 -0.620561    2.0         0.807860
103902                             0.0 -1.149840    0.0        -0.189991
103903                             0.0 -0.819040    2.0         0.535081

        Inflight_Wifi_Service  ...  Inflight_Entertainment  On-board_Service  \
0                    0.203579  ...                1.231704          0.479403
1                    0.203579  ...               -1.769081         -1.849161
2                   -0.549533  ...                1.231704          0.479403
3                   -0.549533  ...               -1.018885         -1.072973
4                    0.203579  ...               -0.268688         -0.296785
...                       ...  ...                     ...               ...
103899              -0.549533  ...               -1.018885         -0.296785
```

```
103900               0.956691  ...            1.231704          1.255590
103901              -1.302646  ...            0.481508         -0.296785
103902              -1.302646  ...           -1.769081          0.479403
103903              -1.302646  ...           -1.769081         -1.849161

        Leg_Room_Service  Baggage_Handling  Checkin_Service  Inflight_Service  \
0              -0.266840          0.311769         0.549799          1.156436
1               1.253380         -0.535045        -1.821012          0.305848
2              -0.266840          0.311769         0.549799          0.305848
3               1.253380         -0.535045        -1.821012          0.305848
4               0.493270          0.311769        -0.240472         -0.544740
...                  ...               ...              ...               ...
103899         -1.787061          0.311769        -1.030742         -0.544740
103900          1.253380          1.158582         1.340069          1.156436
103901         -1.026951          0.311769         1.340069          1.156436
103902          1.253380         -2.228672         1.340069          0.305848
103903         -1.787061          0.311769         0.549799         -0.544740

        Cleanliness  Departure_Delay_in_Minutes  Arrival_Delay_in_Minutes  \
0          1.305870                    0.266393                  0.070988
1         -1.742292                   -0.361375                 -0.238223
2          1.305870                   -0.387532                 -0.392828
3         -0.980251                   -0.099805                 -0.160920
4         -0.218211                   -0.387532                 -0.392828
...             ...                         ...                       ...
103899    -0.980251                   -0.309061                 -0.392828
103900     0.543829                   -0.387532                 -0.392828
103901     0.543829                   -0.204433                 -0.032082
103902    -1.742292                   -0.387532                 -0.392828
103903    -1.742292                   -0.387532                 -0.392828

        Satisfaction
0                0.0
1                0.0
2                1.0
3                0.0
4                1.0
...              ...
103899           0.0
103900           1.0
103901           0.0
103902           0.0
103903           0.0

[103904 rows x 26 columns]
```

4.4.4 Standardizing Encoded Data

[19]:
```
# Standardized Encoded Data
std_encoded_train = encoded_train
std_encoded_train['Class'] = pd.Series(scaler.
↪fit_transform(encoded_train[['Class']]).flatten())
display(std_encoded_train)

std_encoded_test = encoded_test
std_encoded_test['Class'] = pd.Series(scaler.
↪fit_transform(encoded_test[['Class']]).flatten())
```

[19]:
```
        Gender_Female  Gender_Male  Customer_Type_Disloyal_Customer  \
0                 0.0          1.0                              0.0
1                 0.0          1.0                              1.0
2                 1.0          0.0                              0.0
3                 1.0          0.0                              0.0
4                 0.0          1.0                              0.0
...               ...          ...                              ...
103899            1.0          0.0                              1.0
103900            0.0          1.0                              0.0
103901            0.0          1.0                              1.0
103902            1.0          0.0                              1.0
103903            0.0          1.0                              0.0

        Customer_Type_Loyal_Customer  Type_of_Travel_Business_Travel  \
0                                1.0                             0.0
1                                0.0                             1.0
2                                1.0                             1.0
3                                1.0                             1.0
4                                1.0                             1.0
...                              ...                             ...
103899                           0.0                             1.0
103900                           1.0                             1.0
103901                           0.0                             1.0
103902                           0.0                             1.0
103903                           1.0                             1.0

        Type_of_Travel_Personal Travel       Age     Class  Flight_Distance  \
0                                  1.0 -1.745279 -0.029187        -0.731539
1                                  0.0 -0.951360  1.009393        -0.957184
2                                  0.0 -0.885200  1.009393        -0.047584
3                                  0.0 -0.951360  1.009393        -0.629246
4                                  0.0  1.430397  1.009393        -0.978244
...                                ...       ...       ...              ...
103899                             0.0 -1.083680 -1.067767        -1.000307
103900                             0.0  0.636478  1.009393         1.160869
```

```
103901                            0.0 -0.620561  1.009393         0.807860
103902                            0.0 -1.149840 -1.067767        -0.189991
103903                            0.0 -0.819040  1.009393         0.535081

        Inflight_Wifi_Service  ...  Inflight_Entertainment  On-board_Service  \
0                    0.203579  ...                1.231704          0.479403
1                    0.203579  ...               -1.769081         -1.849161
2                   -0.549533  ...                1.231704          0.479403
3                   -0.549533  ...               -1.018885         -1.072973
4                    0.203579  ...               -0.268688         -0.296785
...                       ...  ...                     ...               ...
103899              -0.549533  ...               -1.018885         -0.296785
103900               0.956691  ...                1.231704          1.255590
103901              -1.302646  ...                0.481508         -0.296785
103902              -1.302646  ...               -1.769081          0.479403
103903              -1.302646  ...               -1.769081         -1.849161

        Leg_Room_Service  Baggage_Handling  Checkin_Service  Inflight_Service  \
0              -0.266840          0.311769         0.549799          1.156436
1               1.253380         -0.535045        -1.821012          0.305848
2              -0.266840          0.311769         0.549799          0.305848
3               1.253380         -0.535045        -1.821012          0.305848
4               0.493270          0.311769        -0.240472         -0.544740
...                  ...               ...              ...               ...
103899         -1.787061          0.311769        -1.030742         -0.544740
103900          1.253380          1.158582         1.340069          1.156436
103901         -1.026951          0.311769         1.340069          1.156436
103902          1.253380         -2.228672         1.340069          0.305848
103903         -1.787061          0.311769         0.549799         -0.544740

        Cleanliness  Departure_Delay_in_Minutes  Arrival_Delay_in_Minutes  \
0          1.305870                    0.266393                  0.070988
1         -1.742292                   -0.361375                 -0.238223
2          1.305870                   -0.387532                 -0.392828
3         -0.980251                   -0.099805                 -0.160920
4         -0.218211                   -0.387532                 -0.392828
...             ...                         ...                       ...
103899    -0.980251                   -0.309061                 -0.392828
103900     0.543829                   -0.387532                 -0.392828
103901     0.543829                   -0.204433                 -0.032082
103902    -1.742292                   -0.387532                 -0.392828
103903    -1.742292                   -0.387532                 -0.392828

        Satisfaction
0                0.0
1                0.0
2                1.0
```

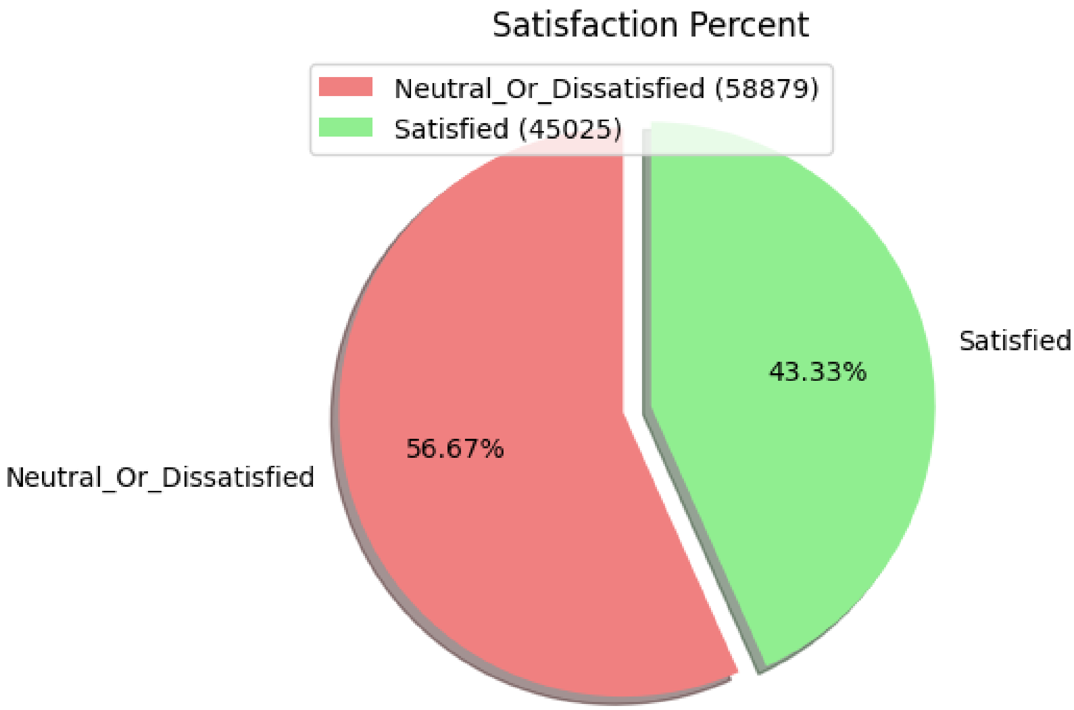
```
3                0.0
4                1.0
...              ...
103899           0.0
103900           1.0
103901           0.0
103902           0.0
103903           0.0

[103904 rows x 26 columns]
```

5 Exploratory Data Analysis

5.1 Satisfaction Percentage

[20]:
```
fig, axes = plt.subplots(nrows=1, ncols=1)
axes.pie(df_train['Satisfaction'].value_counts(),
         labels=df_train['Satisfaction'].value_counts().index,
         explode=[0.1, 0],
         autopct='%1.2f%%',
         shadow=True,
         startangle=90,
         colors=["lightcoral", "lightgreen"])
axes.set_title('Satisfaction Percent')
legend_labels = [f'{label} ({count})' for label, count in␣
 ↪zip(df_train['Satisfaction'].value_counts().index, df_train['Satisfaction'].
 ↪value_counts())]
axes.legend(legend_labels, loc='upper left')
plt.show()
```

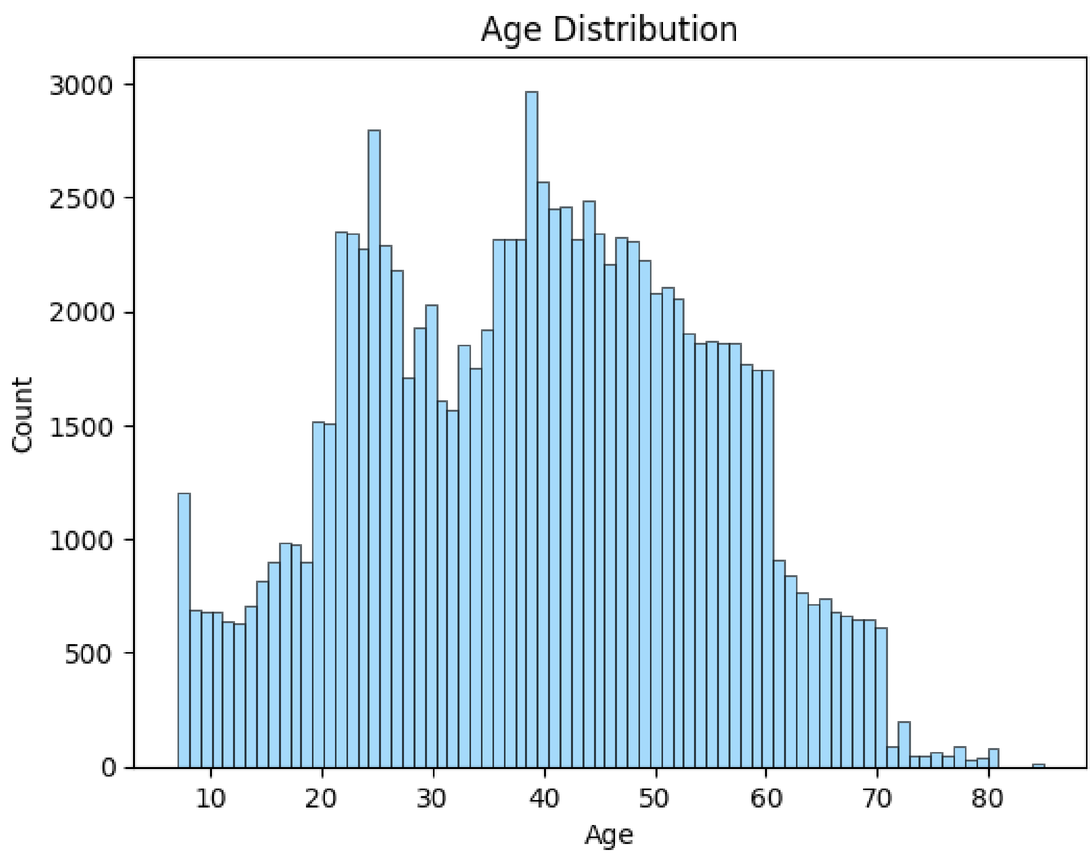
[20]:



5.2 Age Distribution

[21]:
```
sns.histplot(data=unstandardized_df_train,x='Age', color="lightskyblue")
plt.title("Age Distribution")
plt.show()
```

[21]:

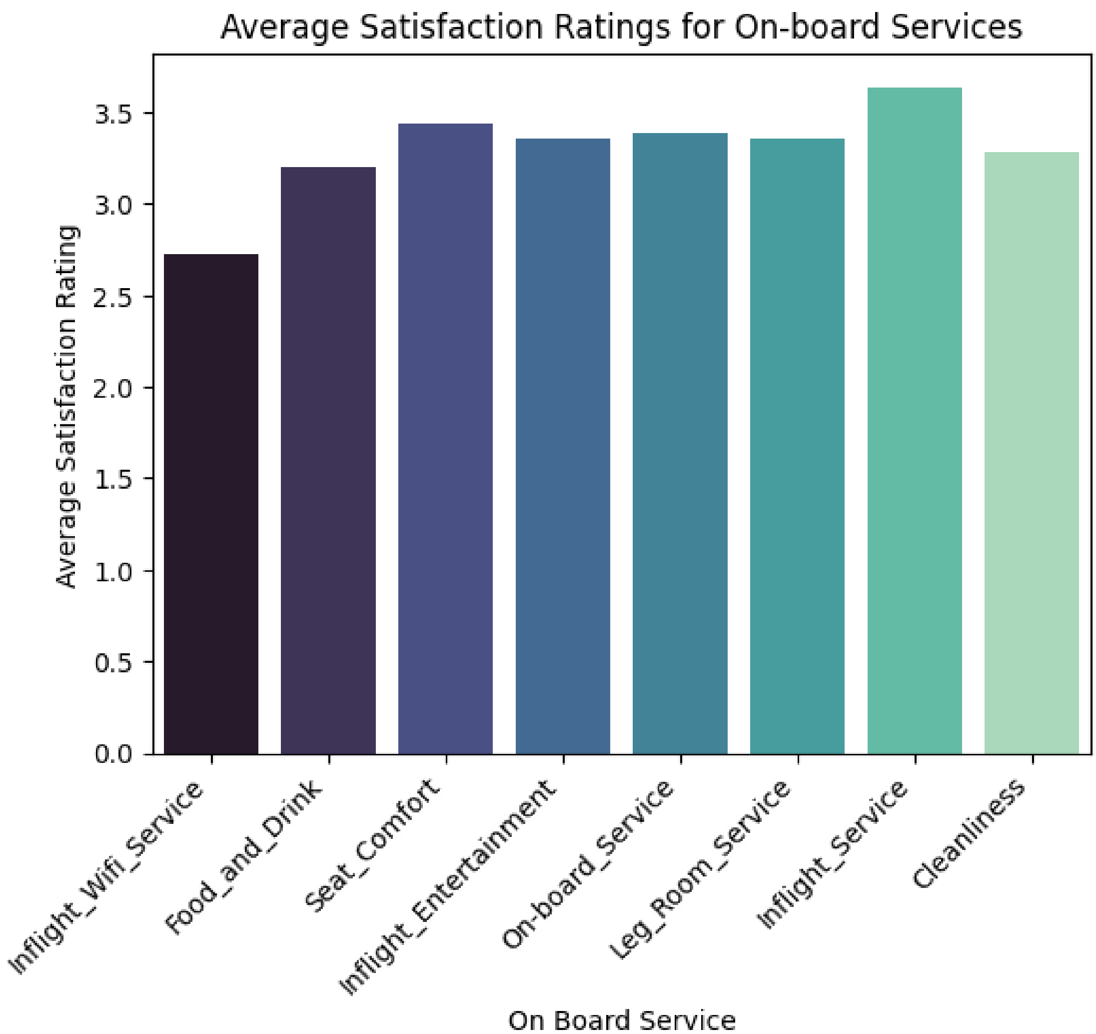


5.3 Average Satisfaction of Services

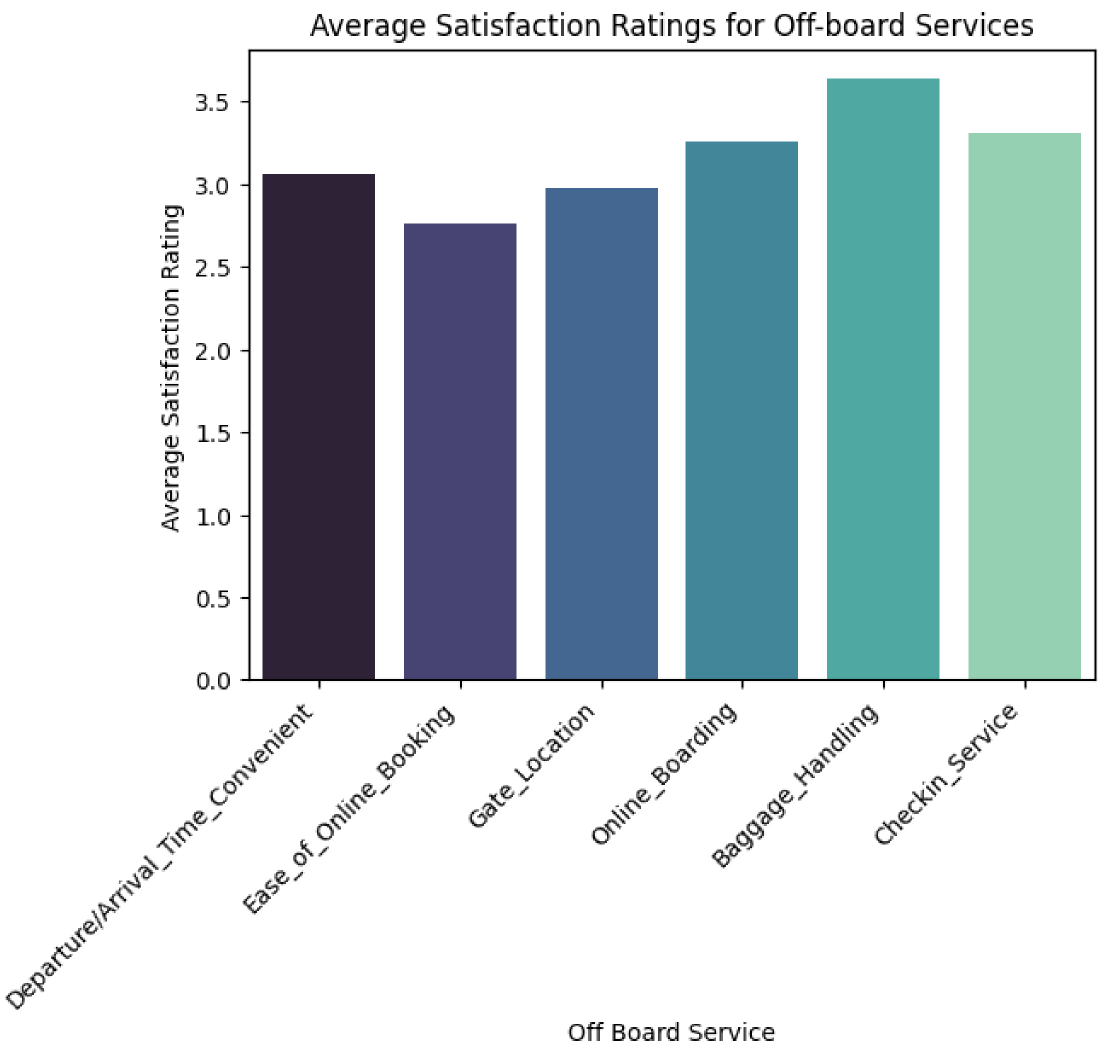
[22]:
```
on_board_services = ['Inflight_Wifi_Service', 'Food_and_Drink', 'Seat_Comfort',␣
↪'Inflight_Entertainment',␣
↪'On-board_Service','Leg_Room_Service','Inflight_Service', 'Cleanliness']
average_on_ratings = unstandardized_df_train[on_board_services].mean()
average_on_ratings_df = pd.DataFrame({'On Board Service': average_on_ratings.
↪index, 'Average Satisfaction Rating': average_on_ratings.values})
sns.barplot(x='On Board Service', y='Average Satisfaction Rating',␣
↪data=average_on_ratings_df, palette='mako')
plt.xticks(rotation=45, ha='right')
plt.title('Average Satisfaction Ratings for On-board Services')
plt.show()


off_board_services = ['Departure/Arrival_Time_Convenient',␣
↪'Ease_of_Online_Booking', 'Gate_Location', 'Online_Boarding',␣
↪'Baggage_Handling', 'Checkin_Service']
average_off_ratings = unstandardized_df_train[off_board_services].mean()
```

```
average_off_ratings_df = pd.DataFrame({'Off Board Service': average_off_ratings.
↪index, 'Average Satisfaction Rating': average_off_ratings.values})
sns.barplot(x='Off Board Service', y='Average Satisfaction Rating',␣
↪data=average_off_ratings_df, palette='mako')
plt.xticks(rotation=45, ha='right')
plt.title('Average Satisfaction Ratings for Off-board Services')
plt.show()
```

[22]:



[22]:



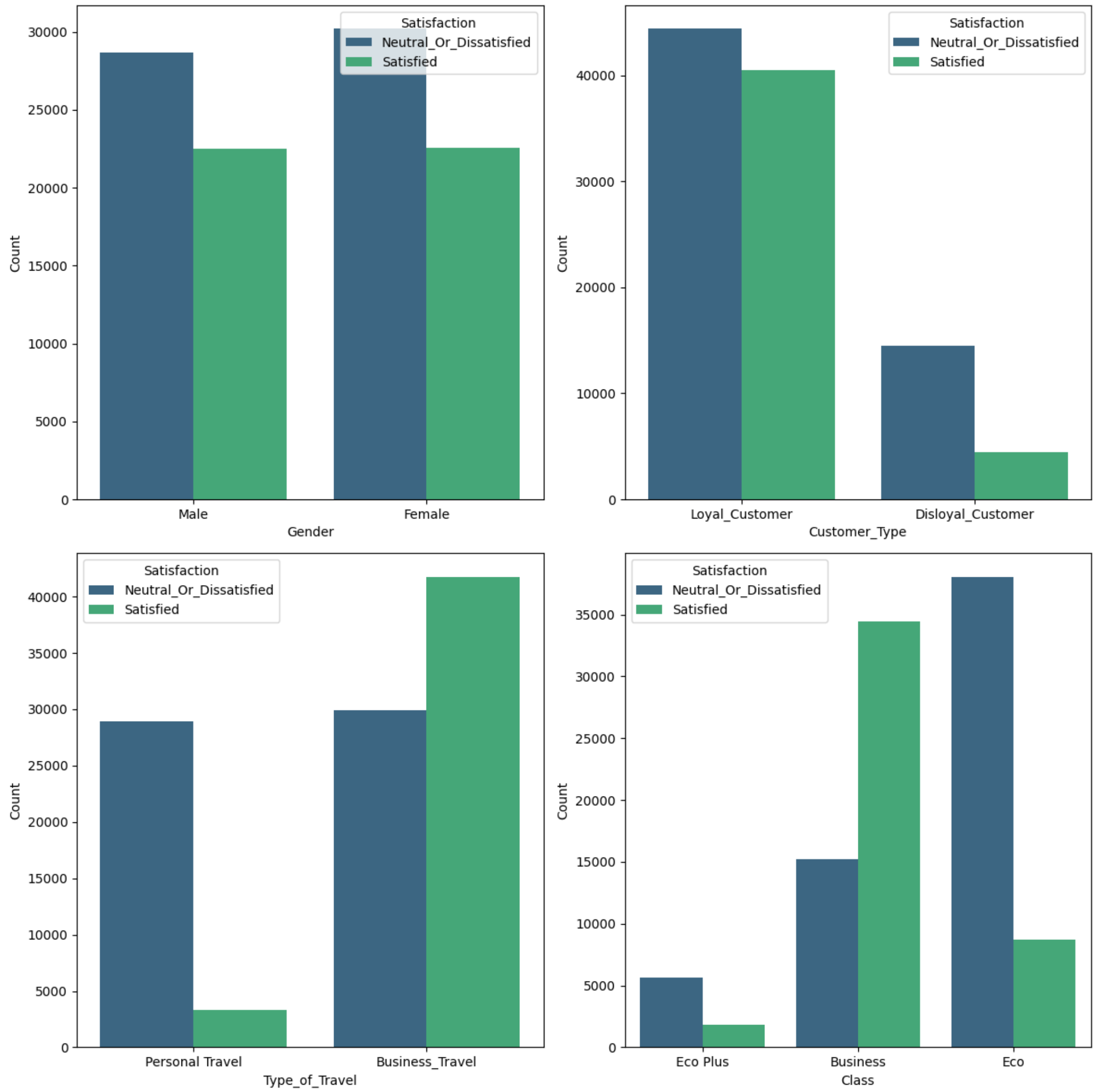
5.4 Satisfaction Distribution Based on Categorical Values

[23]:

```
fig, axes = plt.subplots(nrows=2, ncols=2, figsize=(12, 12))
data_cols = ['Gender', 'Customer_Type', 'Type_of_Travel', 'Class']

for i, col_name in enumerate(data_cols):
    row = i // 2
    col = i % 2
    sns.countplot(x=col_name, hue='Satisfaction', data=df_train, ax=axes[row,␣
↪col], palette='viridis')
    axes[row, col].set_xlabel(col_name)
    axes[row, col].set_ylabel('Count')

plt.tight_layout()
```

```
plt.show()
```

[23]:



5.5 Satisfaction Distribution Based on Numerical Values

[24]:
```
fig, axes = plt.subplots(nrows=4, ncols=4, figsize=(25, 25))
data_cols = ['Inflight_Wifi_Service',
             'Departure/Arrival_Time_Convenient',
             'Ease_of_Online_Booking',
             'Gate_Location',
             'Food_and_Drink',
             'Online_Boarding',
             'Seat_Comfort',
             'Inflight_Entertainment',
```

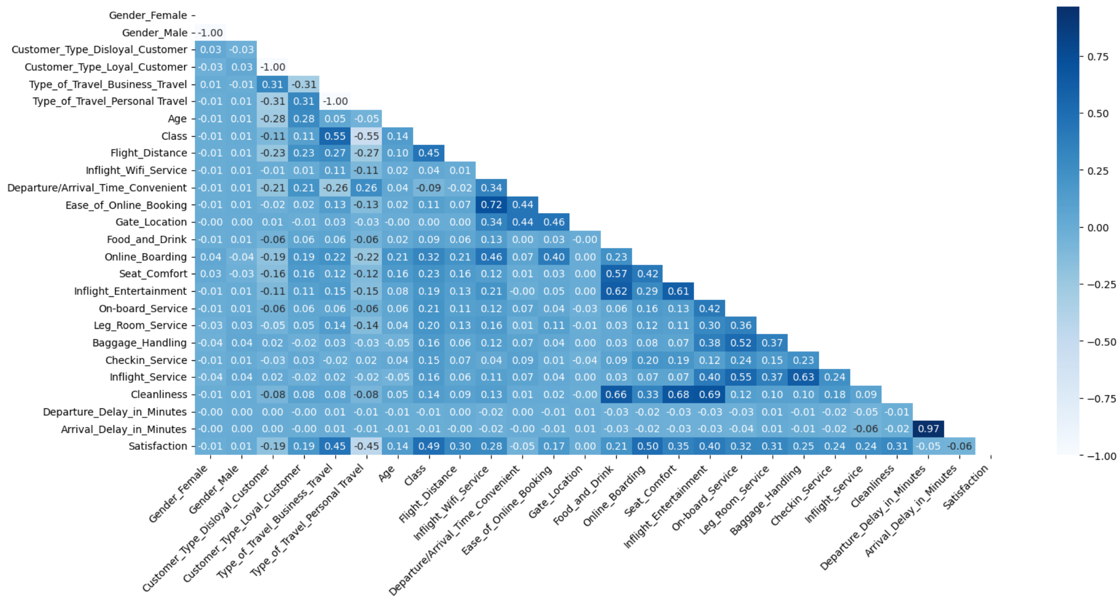
```
             'On-board_Service',
             'Leg_Room_Service',
             'Baggage_Handling',
             'Checkin_Service',
             'Inflight_Service',
             'Cleanliness']

for i, col_name in enumerate(data_cols):
    row = i // 4
    col = i % 4
    sns.countplot(x=col_name, hue='Satisfaction', data=unstandardized_df_train,␣
 ↪ax=axes[row, col], hue_order=["Satisfied", "Neutral_Or_Dissatisfied"],␣
 ↪palette='mako')
    axes[row, col].set_xlabel(col_name)
    axes[row, col].set_ylabel('Count')
plt.tight_layout()
plt.show()
```

[24]:

5.6 Correlation Matrix

[25]:
```
plt.figure(figsize=(18,8))

corr = std_encoded_train.corr()
mask = np.triu(corr)

sns.heatmap(corr, annot=True, fmt='.2f', mask=mask, cmap="Blues")
plt.xticks(rotation=45, ha='right')
plt.show()
```

[25]:



6 Model Building

6.1 Splitting Data

[26]:
```
X_train = std_encoded_train.drop('Satisfaction', axis=1)
X_test = std_encoded_test.drop('Satisfaction', axis=1)
Y_train = std_encoded_train['Satisfaction']
Y_test = std_encoded_test['Satisfaction']
```

6.2 Model Construction

[69]:
```
from sklearn.neighbors import KNeighborsClassifier
from sklearn.naive_bayes import GaussianNB
from sklearn.linear_model import Perceptron, LogisticRegression
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
classifiers = [("KNN", KNeighborsClassifier()),
               ("Gaussian Naive Bayes", GaussianNB()),
               ("Perceptron", Perceptron()),
               ("Logistic Regression", LogisticRegression()),
               ("Decision Trees", DecisionTreeClassifier()),
               ("Random Forest Classifier", RandomForestClassifier())]

for name, classifier in classifiers:
```

```
    classifier.fit(X_train, Y_train)
```

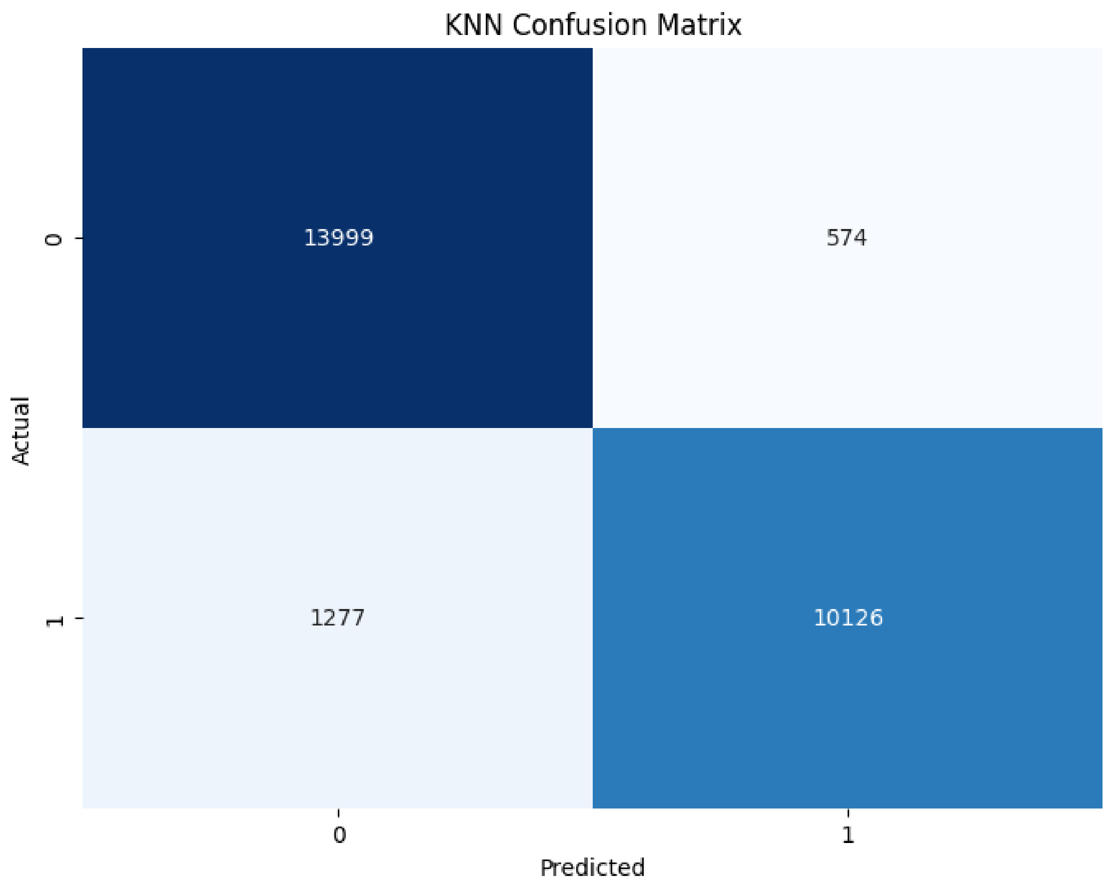
7 Model Testing

---

[67]:
```
from sklearn.metrics import accuracy_score, confusion_matrix

results = pd.DataFrame(columns=["Classifier", "Training Accuracy", "Testing␣
↪Accuracy"])

for name, classifier in classifiers:
    Y_pred_train = classifier.predict(X_train)
    Y_pred_test = classifier.predict(X_test)
    score = classifier.score(X_test, Y_test)
    train_accuracy_score = accuracy_score(Y_train, Y_pred_train)
    test_accuracy_score = accuracy_score(Y_test, Y_pred_test)
    print("Name: ", name)
    print("Training Accuracy Score: ", train_accuracy_score)
    print("Testing Accuracy Score: ", test_accuracy_score, "\n")
    data = pd.DataFrame([{"Classifier": name, "Training Accuracy":␣
↪train_accuracy_score, "Testing Accuracy": test_accuracy_score}])
    confusion = confusion_matrix(Y_test, Y_pred_test)
    plt.figure(figsize=(8, 6))
    sns.heatmap(confusion, annot=True, fmt="d", cmap="Blues", cbar=False)
    plt.xlabel('Predicted')
    plt.ylabel('Actual')
    plt.title(name + ' Confusion Matrix')
    plt.show()
    results = pd.concat([results, data], ignore_index=True)
```

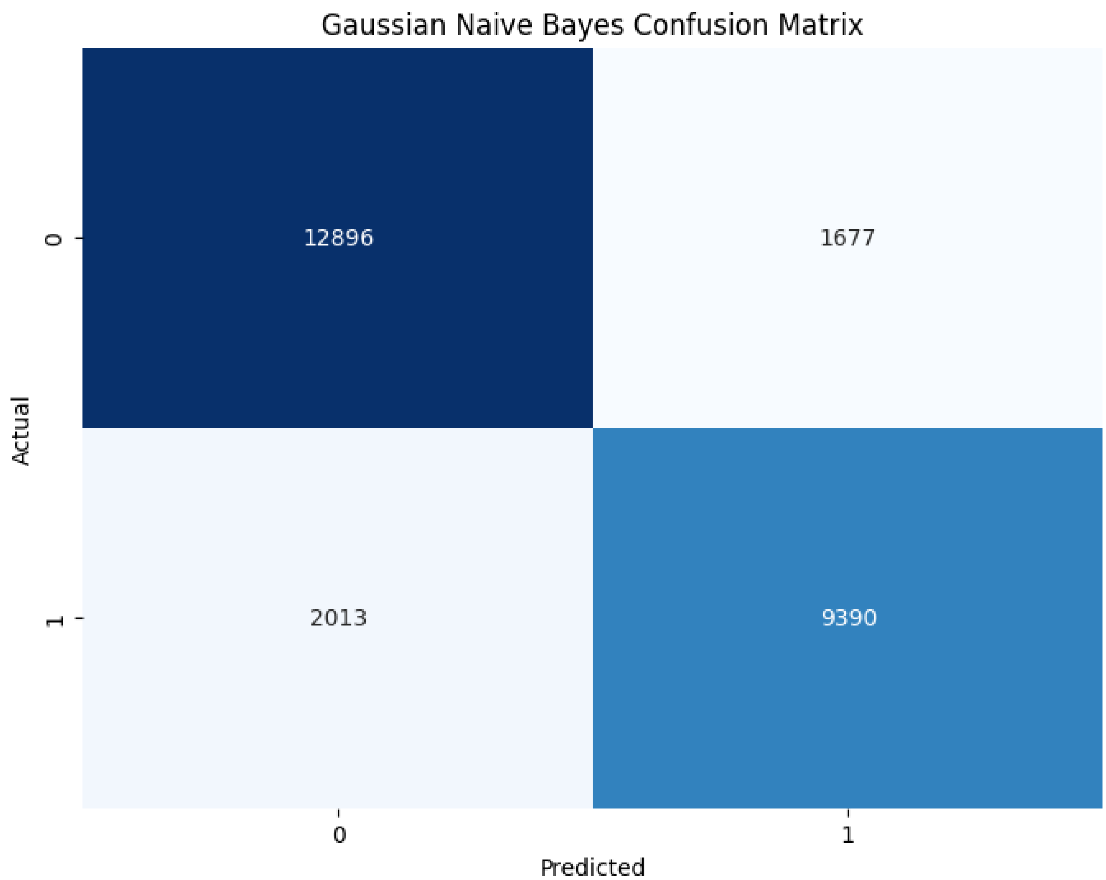
```
Name:  KNN
Training Accuracy Score:  0.9493763473975978
Testing Accuracy Score:  0.9287419156144133
```

[67]:



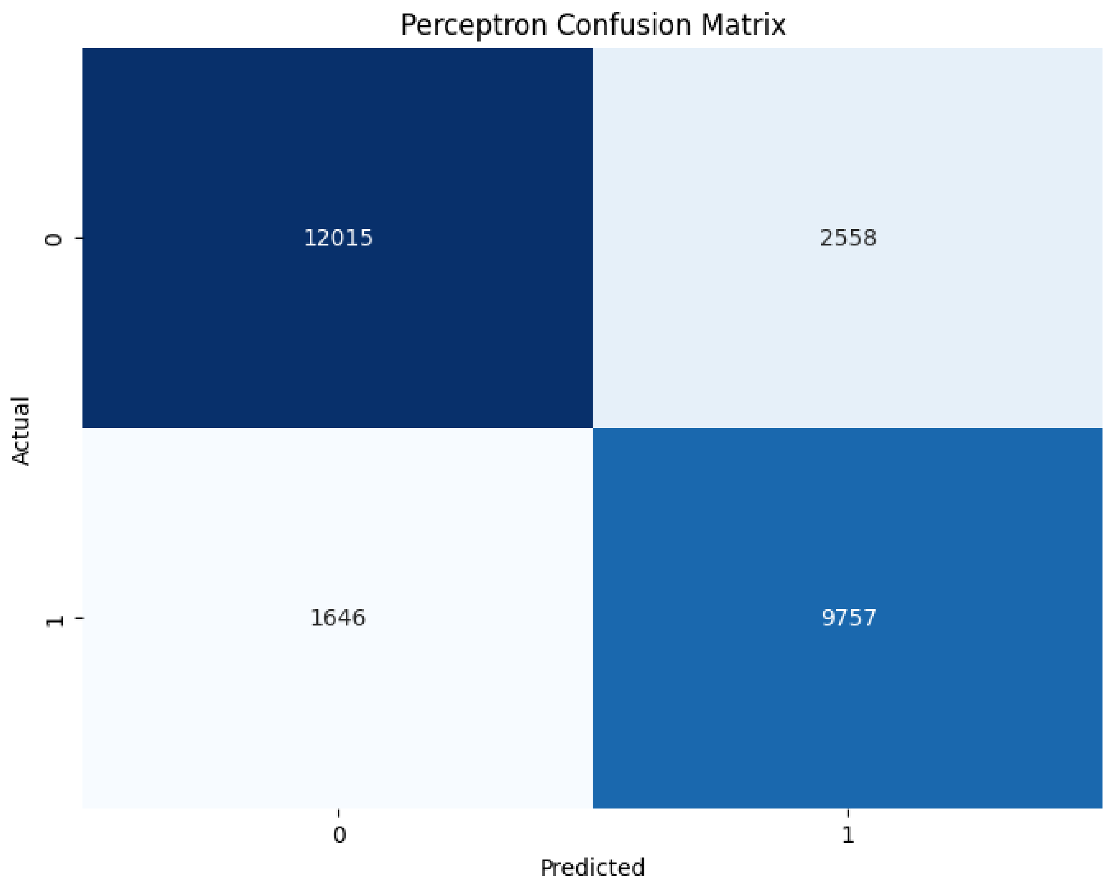

```
Name:  Gaussian Naive Bayes
Training Accuracy Score:  0.8620072374499538
Testing Accuracy Score:  0.857945796119495
```

[67]:



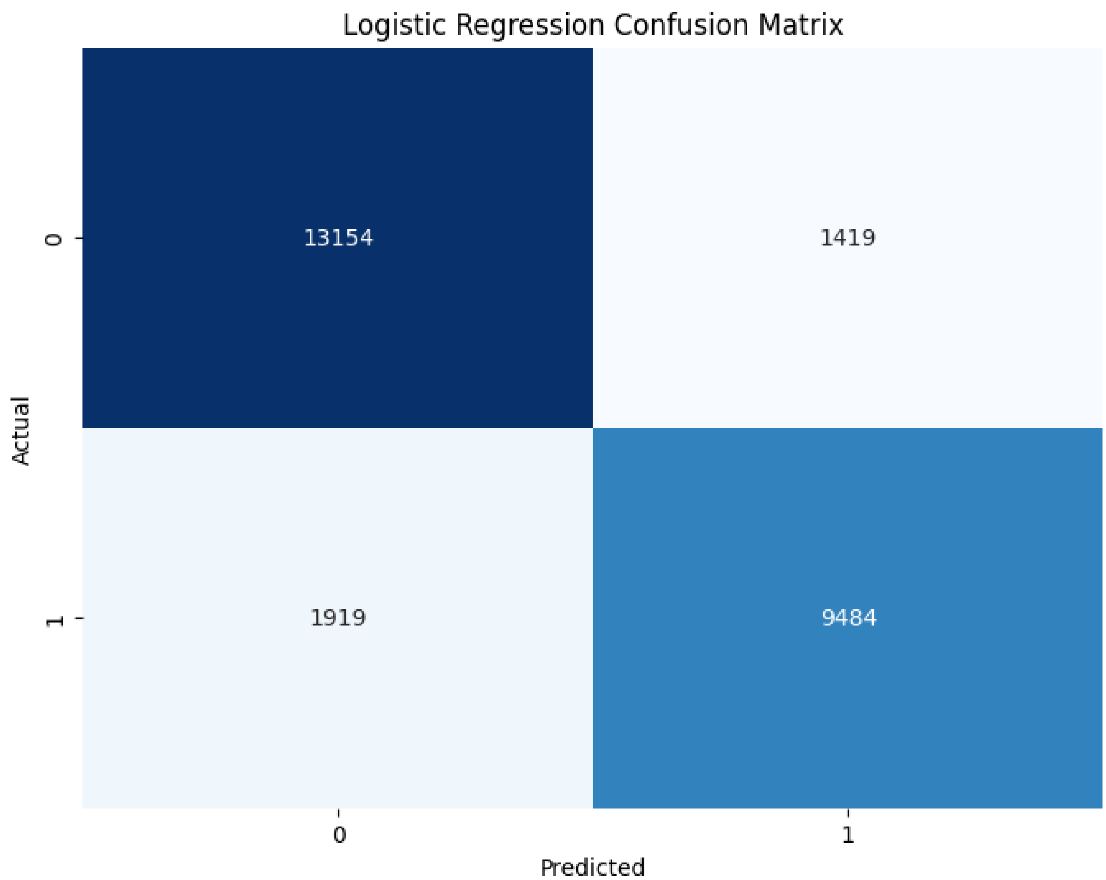

```
Name:  Perceptron
Training Accuracy Score:  0.8347128118263012
Testing Accuracy Score:  0.8381582999692023
```

[67]:



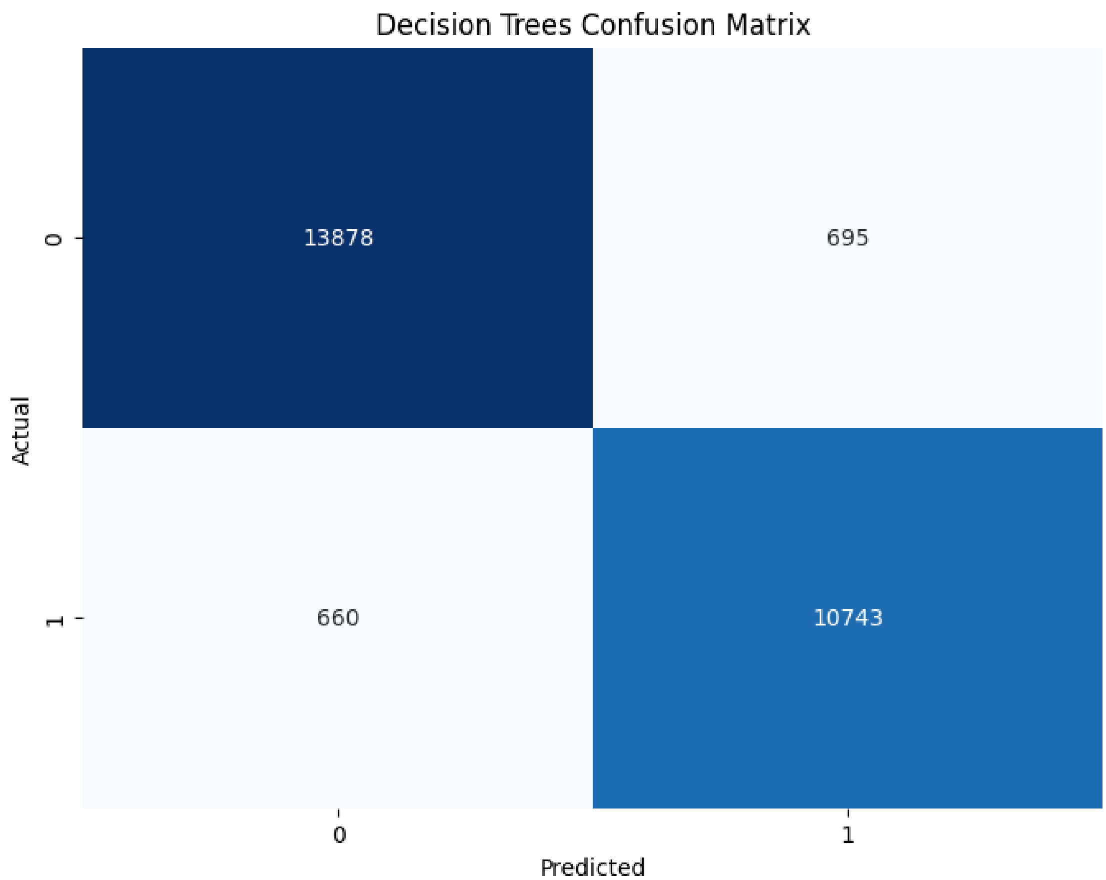

```
Name:  Logistic Regression
Training Accuracy Score:  0.8748748845087774
Testing Accuracy Score:  0.8714967662457653
```

[67]:



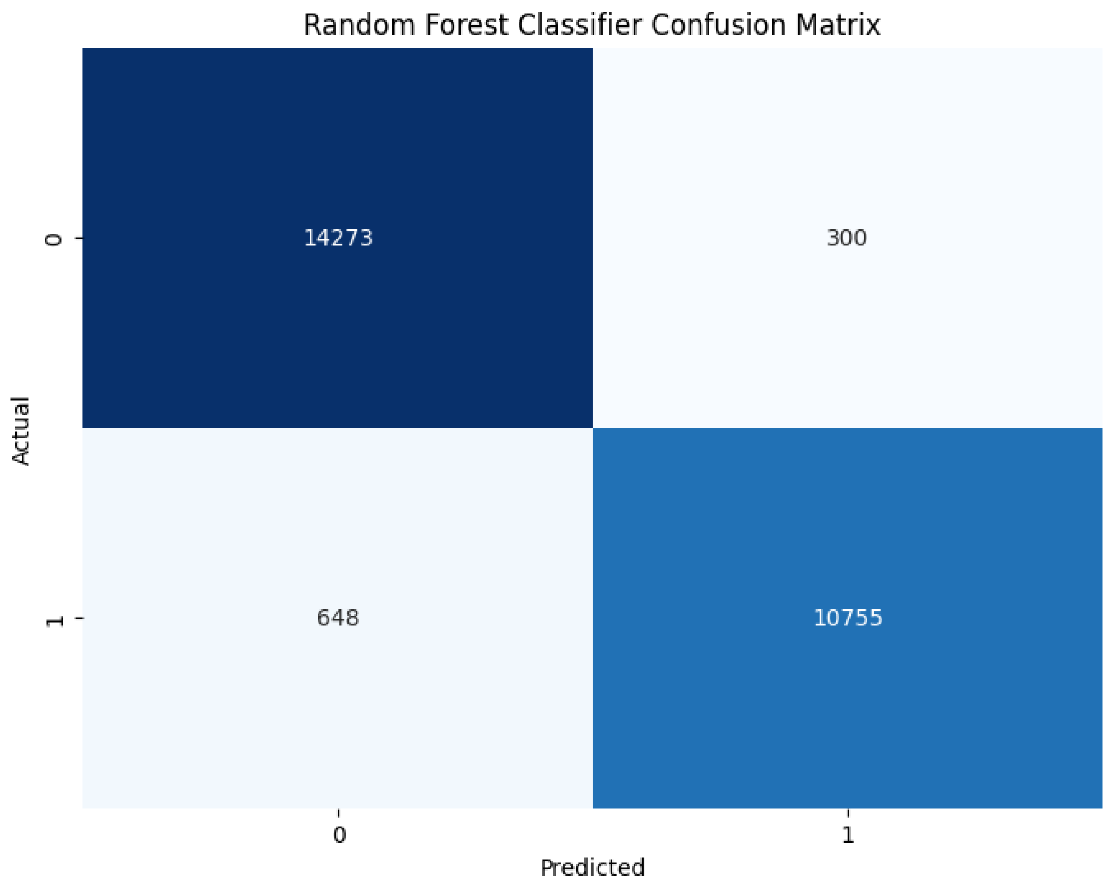

```
Name:  Decision Trees
Training Accuracy Score:  1.0
Testing Accuracy Score:  0.9478364644287034
```

[67]:




```
Name:  Random Forest Classifier
Training Accuracy Score:  1.0
Testing Accuracy Score:  0.9635047736372035
```

[67]:



[68]:
```
%%latex
$$Accuracy\ Score = \dfrac{TP + TN}{TP + TN + FP + FN}$$
```

[68]:

$$Accuracy\ Score = \dfrac{TP + TN}{TP + TN + FP + FN}$$

8 Model Comparison

---

[52]:
```
styled_results = results.style.background_gradient(cmap='Blues',␣
 ↪subset=["Training Accuracy", "Testing Accuracy"])
display(styled_results)
```

[52]: <pandas.io.formats.style.Styler at 0x7ff57d7a62f0>