ml-model

May 9, 2023

0.1 Step01- Import Libraries

[1]:
```
import pandas as pd # Used to dataframe operations purpose
import numpy as np # for numeric operation
import seaborn as sns # Making Statistical graphics
import matplotlib.pyplot as plt# Making Statistical graphics
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import␣
 ↪confusion_matrix,multilabel_confusion_matrix,classification_report,accuracy_score
import warnings
warnings.filterwarnings('ignore')
import joblib
import os
from flask import jsonify, Flask, request
from flask_restful import Api, Resource
```

0.2 Step02- Data Gathering

[ ]:
```
# First thing and main is understand the problem statement and relative data␣
 ↪information.
# After understand the problem, to start the ML process, collect the data in␣
 ↪required format.
```

[2]:
```
df = pd.read_csv('data (1).csv')  # Using pandas we read the data file.
df
```

[2]:

	id	Gender	Customer Type	Age	Type of Travel	Class	\
0	19556	Female	Loyal Customer	52	Business travel	Eco	
1	90035	Female	Loyal Customer	36	Business travel	Business	
2	12360	Male	disloyal Customer	20	Business travel	Eco	
3	77959	Male	Loyal Customer	44	Business travel	Business	
4	36875	Female	Loyal Customer	49	Business travel	Eco	
…	…	…	…	…	…	…	
25971	78463	Male	disloyal Customer	34	Business travel	Business	
25972	71167	Male	Loyal Customer	23	Business travel	Business	
25973	37675	Female	Loyal Customer	17	Personal Travel	Eco	

```
25974  90086    Male     Loyal Customer   14  Business travel  Business
25975  34799  Female     Loyal Customer   42  Personal Travel       Eco

       Flight Distance  Inflight wifi service  \
0                  160                      5
1                 2863                      1
2                  192                      2
3                 3377                      0
4                 1182                      2
…                    …                      …
25971              526                      3
25972              646                      4
25973              828                      2
25974             1127                      3
25975              264                      2

       Departure/Arrival time convenient  Ease of Online booking  …  \
0                                      4                       3  …
1                                      1                       3  …
2                                      0                       2  …
3                                      0                       0  …
4                                      3                       4  …
…                                    …                     …  …
25971                                  3                       3  …
25972                                  4                       4  …
25973                                  5                       1  …
25974                                  3                       3  …
25975                                  5                       2  …

       Inflight entertainment  On-board service  Leg room service  \
0                           5                 5                 5
1                           4                 4                 4
2                           2                 4                 1
3                           1                 1                 1
4                           2                 2                 2
…                         …                 …                 …
25971                       4                 3                 2
25972                       4                 4                 5
25973                       2                 4                 3
25974                       4                 3                 2
25975                       1                 1                 2

       Baggage handling  Checkin service  Inflight service  Cleanliness  \
0                     5                2                 5            5
1                     4                3                 4            5
2                     3                2                 2            2
3                     1                3                 1            4
```

```
4                    2                4                2            4
…                  …                …                …          …
25971                4                4                5            4
25972                5                5                5            4
25973                4                5                4            2
25974                5                4                5            4
25975                1                1                1            1

       Departure Delay in Minutes  Arrival Delay in Minutes  \
0                              50                      44.0
1                               0                       0.0
2                               0                       0.0
3                               0                       6.0
4                               0                      20.0
…                             …                        …
25971                           0                       0.0
25972                           0                       0.0
25973                           0                       0.0
25974                           0                       0.0
25975                           0                       0.0

                  satisfaction
0                    satisfied
1                    satisfied
2      neutral or dissatisfied
3                    satisfied
4                    satisfied
…                          …
25971  neutral or dissatisfied
25972                satisfied
25973  neutral or dissatisfied
25974                satisfied
25975  neutral or dissatisfied

[25976 rows x 24 columns]
```

[3]:
```
df.head()
```

[3]:
```
      id  Gender      Customer Type  Age   Type of Travel     Class  \
0  19556  Female     Loyal Customer   52  Business travel       Eco
1  90035  Female     Loyal Customer   36  Business travel  Business
2  12360    Male  disloyal Customer   20  Business travel       Eco
3  77959    Male     Loyal Customer   44  Business travel  Business
4  36875  Female     Loyal Customer   49  Business travel       Eco

   Flight Distance  Inflight wifi service  Departure/Arrival time convenient  \
0              160                      5                                  4
```

```
1             2863                     1                                  1
2              192                     2                                  0
3             3377                     0                                  0
4             1182                     2                                  3

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

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

   Cleanliness  Departure Delay in Minutes  Arrival Delay in Minutes  \
0            5                          50                      44.0
1            5                           0                       0.0
2            2                           0                       0.0
3            4                           0                       6.0
4            4                           0                      20.0

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

[5 rows x 24 columns]
```

[4]:
```
df.info() # Here we can understand the structure of data, name and number of Featuers,Data type of Feature, Data size so on.
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 25976 entries, 0 to 25975
Data columns (total 24 columns):
 #   Column                             Non-Null Count  Dtype
---  ------                             --------------  -----
 0   id                                 25976 non-null  int64
 1   Gender                             25976 non-null  object
 2   Customer Type                      25966 non-null  object
 3   Age                                25976 non-null  int64
```

```
 4   Type of Travel                     25976 non-null  object
 5   Class                              25976 non-null  object
 6   Flight Distance                    25976 non-null  int64
 7   Inflight wifi service              25976 non-null  int64
 8   Departure/Arrival time convenient  25976 non-null  int64
 9   Ease of Online booking             25976 non-null  int64
 10  Gate location                      25976 non-null  int64
 11  Food and drink                     25976 non-null  int64
 12  Online boarding                    25976 non-null  int64
 13  Seat comfort                       25976 non-null  int64
 14  Inflight entertainment             25976 non-null  int64
 15  On-board service                   25976 non-null  int64
 16  Leg room service                   25976 non-null  int64
 17  Baggage handling                   25976 non-null  int64
 18  Checkin service                    25976 non-null  int64
 19  Inflight service                   25976 non-null  int64
 20  Cleanliness                        25976 non-null  int64
 21  Departure Delay in Minutes         25976 non-null  int64
 22  Arrival Delay in Minutes           25893 non-null  float64
 23  satisfaction                       25976 non-null  object
dtypes: float64(1), int64(18), object(5)
memory usage: 4.8+ MB
```

[5]:
```
df.shape # column and row size
```

[5]:
```
(25976, 24)
```

[6]:
```
df.describe(include='all')  # to understand main information we use this␣
 ↪Describe method
```

[6]:

	id	Gender	Customer Type	Age	Type of Travel	\
count	25976.000000	25976	25966	25976.000000	25976	
unique	NaN	2	3	NaN	2	
top	NaN	Female	Loyal Customer	NaN	Business travel	
freq	NaN	13172	21170	NaN	18038	
mean	65005.657992	NaN	NaN	39.620958	NaN	
std	37611.526647	NaN	NaN	15.135685	NaN	
min	17.000000	NaN	NaN	7.000000	NaN	
25%	32170.500000	NaN	NaN	27.000000	NaN	
50%	65319.500000	NaN	NaN	40.000000	NaN	
75%	97584.250000	NaN	NaN	51.000000	NaN	
max	129877.000000	NaN	NaN	85.000000	NaN	

	Class	Flight Distance	Inflight wifi service	\
count	25976	25976.000000	25976.000000	
unique	5	NaN	NaN	
top	Business	NaN	NaN	

freq	12490	NaN	NaN
mean	NaN	1193.788459	2.724746
std	NaN	998.683999	1.335384
min	NaN	31.000000	0.000000
25%	NaN	414.000000	2.000000
50%	NaN	849.000000	3.000000
75%	NaN	1744.000000	4.000000
max	NaN	4983.000000	5.000000

	Departure/Arrival time convenient	Ease of Online booking	…	\
count	25976.000000	25976.000000	…	
unique	NaN	NaN	…	
top	NaN	NaN	…	
freq	NaN	NaN	…	
mean	3.046812	2.756775	…	
std	1.533371	1.412951	…	
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	…	

	Inflight entertainment	On-board service	Leg room service	\
count	25976.000000	25976.000000	25976.000000	
unique	NaN	NaN	NaN	
top	NaN	NaN	NaN	
freq	NaN	NaN	NaN	
mean	3.357753	3.385664	3.350169	
std	1.338299	1.282088	1.318862	
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	25976.000000	25976.000000	25976.000000	25976.000000	
unique	NaN	NaN	NaN	NaN	
top	NaN	NaN	NaN	NaN	
freq	NaN	NaN	NaN	NaN	
mean	3.633238	3.314175	3.649253	3.286226	
std	1.176525	1.269332	1.180681	1.319330	
min	1.000000	1.000000	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                    25976.00000               25893.000000
unique                       NaN                        NaN
top                          NaN                        NaN
freq                         NaN                        NaN
mean                     14.30609                  14.740857
std                      37.42316                  37.517539
min                       0.00000                   0.000000
25%                       0.00000                   0.000000
50%                       0.00000                   0.000000
75%                      12.00000                  13.000000
max                    1128.00000                1115.000000

                   satisfaction
count                     25976
unique                        2
top     neutral or dissatisfied
freq                      14573
mean                        NaN
std                         NaN
min                         NaN
25%                         NaN
50%                         NaN
75%                         NaN
max                         NaN

[11 rows x 24 columns]
```

[7]:
```
df.drop(['id'],axis=1,inplace=True) # ID column doesent tell any info and not␣
 ↪useful for model, Due to all Unique no, so we drop that.
```

[8]:
```
df
```

[8]:
```
       Gender      Customer Type  Age   Type of Travel     Class  \
0      Female     Loyal Customer   52  Business travel       Eco
1      Female     Loyal Customer   36  Business travel  Business
2        Male  disloyal Customer   20  Business travel       Eco
3        Male     Loyal Customer   44  Business travel  Business
4      Female     Loyal Customer   49  Business travel       Eco
…       …                 …  …            …       …
25971    Male  disloyal Customer   34  Business travel  Business
25972    Male     Loyal Customer   23  Business travel  Business
25973  Female     Loyal Customer   17  Personal Travel       Eco
25974    Male     Loyal Customer   14  Business travel  Business
25975  Female     Loyal Customer   42  Personal Travel       Eco
```

```
       Flight Distance  Inflight wifi service  \
0                  160                      5
1                 2863                      1
2                  192                      2
3                 3377                      0
4                 1182                      2
…                    …                      …
25971              526                      3
25972              646                      4
25973              828                      2
25974             1127                      3
25975              264                      2

       Departure/Arrival time convenient  Ease of Online booking  \
0                                      4                       3
1                                      1                       3
2                                      0                       2
3                                      0                       0
4                                      3                       4
…                                      …                       …
25971                                  3                       3
25972                                  4                       4
25973                                  5                       1
25974                                  3                       3
25975                                  5                       2

       Gate location  …  Inflight entertainment  On-board service  \
0                  4  …                       5                 5
1                  1  …                       4                 4
2                  4  …                       2                 4
3                  2  …                       1                 1
4                  3  …                       2                 2
…                  …  …                       …                 …
25971              1  …                       4                 3
25972              4  …                       4                 4
25973              5  …                       2                 4
25974              3  …                       4                 3
25975              5  …                       1                 1

       Leg room service  Baggage handling  Checkin service  Inflight service  \
0                     5                 5                2                 5
1                     4                 4                3                 4
2                     1                 3                2                 2
3                     1                 1                3                 1
4                     2                 2                4                 2
…                   …                 …                …                 …
25971                 2                 4                4                 5
```

```
25972                5                  5                  5                  5
25973                3                  4                  5                  4
25974                2                  5                  4                  5
25975                2                  1                  1                  1

       Cleanliness  Departure Delay in Minutes  Arrival Delay in Minutes  \
0                5                          50                      44.0
1                5                           0                       0.0
2                2                           0                       0.0
3                4                           0                       6.0
4                4                           0                      20.0
…              …                           …                         …
25971            4                           0                       0.0
25972            4                           0                       0.0
25973            2                           0                       0.0
25974            4                           0                       0.0
25975            1                           0                       0.0

                  satisfaction
0                    satisfied
1                    satisfied
2      neutral or dissatisfied
3                    satisfied
4                    satisfied
…                          …
25971  neutral or dissatisfied
25972                satisfied
25973  neutral or dissatisfied
25974                satisfied
25975  neutral or dissatisfied

[25976 rows x 23 columns]
```

0.3 Step03- EDA & Feature Engg

[ ]:
```
# here we study and analyse the each feature and if Chategorical data is there␣
 ↪then we convert that data. There is so many steps
# to work on that for each if required, also we work on mainly Missing value,␣
 ↪Outlires, Skewed data, Encoding, Distribution of data.
```

[9]:
```
# Here I created Outliers function for to use directly it help to maintain code␣
 ↪simple and also save time once created
def Outliers(i):
    q1=i.quantile(0.25)
    q2=i.quantile(0.50)
    q3=i.quantile(0.75)
```

```
    print(q1,q2,q3)
    print()
    iqr=q3-q1
    iqr
    print()
    uppr_lmt=q3+(1.5*iqr)
    lwr_lmt=q1-(1.5*iqr)
    print(uppr_lmt,lwr_lmt)
    print()
    outliers=df[(i>uppr_lmt)|(i<lwr_lmt)].index
    outliers
    print()
    import numpy as np
    i.mean()
    print()
    import numpy as np
    i=np.where(i>uppr_lmt,uppr_lmt,i)
    print()
    i.mean()
    print()
    sns.boxplot(i)
    return i
```

0.3.1 Feature 1:- Gender

[10]:
```
# count of unique value

df['Gender'].value_counts()
```

[10]:
```
Female    13172
Male      12804
Name: Gender, dtype: int64
```

[11]:
```
# Total unique values

df['Gender'].nunique()
```

[11]:
```
2
```

[12]:
```
df['Gender'].replace({'Male':1,'Female':2},inplace=True)
df.head()
```

[12]:

	Gender	Customer Type	Age	Type of Travel	Class	Flight Distance	\
0	2	Loyal Customer	52	Business travel	Eco	160	
1	2	Loyal Customer	36	Business travel	Business	2863	
2	1	disloyal Customer	20	Business travel	Eco	192	
3	1	Loyal Customer	44	Business travel	Business	3377	

```
4       2     Loyal Customer   49  Business travel       Eco             1182

   Inflight wifi service  Departure/Arrival time convenient  \
0                      5                                  4
1                      1                                  1
2                      2                                  0
3                      0                                  0
4                      2                                  3

   Ease of Online booking  Gate location  …  Inflight entertainment  \
0                       3             4  …                       5
1                       3             1  …                       4
2                       2             4  …                       2
3                       0             2  …                       1
4                       4             3  …                       2

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

   Inflight service  Cleanliness  Departure Delay in Minutes  \
0                 5            5                          50
1                 4            5                           0
2                 2            2                           0
3                 1            4                           0
4                 2            4                           0

   Arrival Delay in Minutes             satisfaction
0                      44.0                satisfied
1                       0.0                satisfied
2                       0.0  neutral or dissatisfied
3                       6.0                satisfied
4                      20.0                satisfied

[5 rows x 23 columns]
```

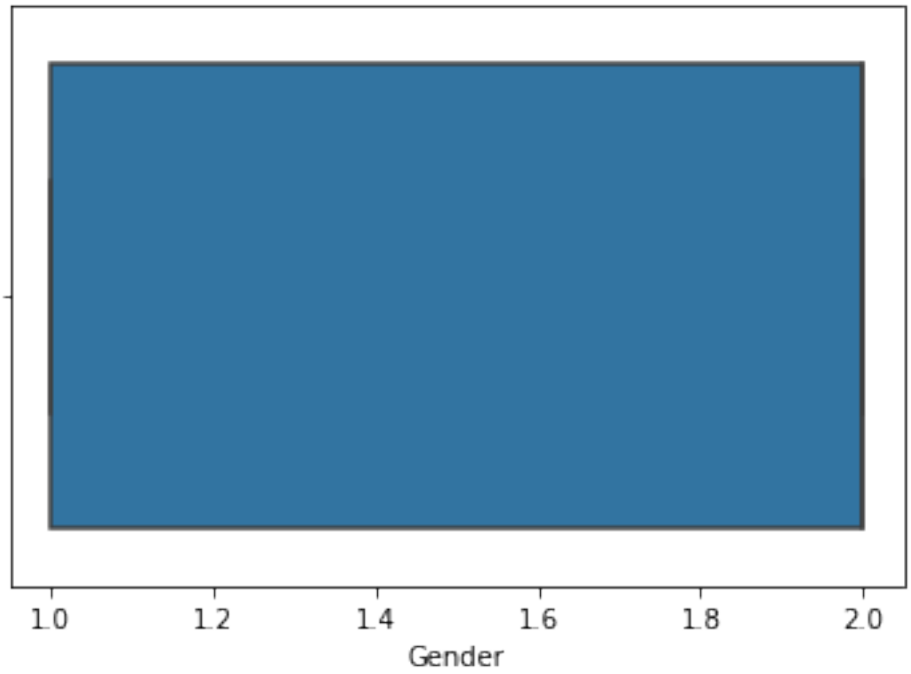
[13]:
```
df['Gender'].value_counts(normalize=True)
```

[13]:
```
2    0.507083
1    0.492917
Name: Gender, dtype: float64
```

[14]:
```
# Boxplot used to idntify Outlires from feature
```

```
sns.boxplot(df['Gender'])
```

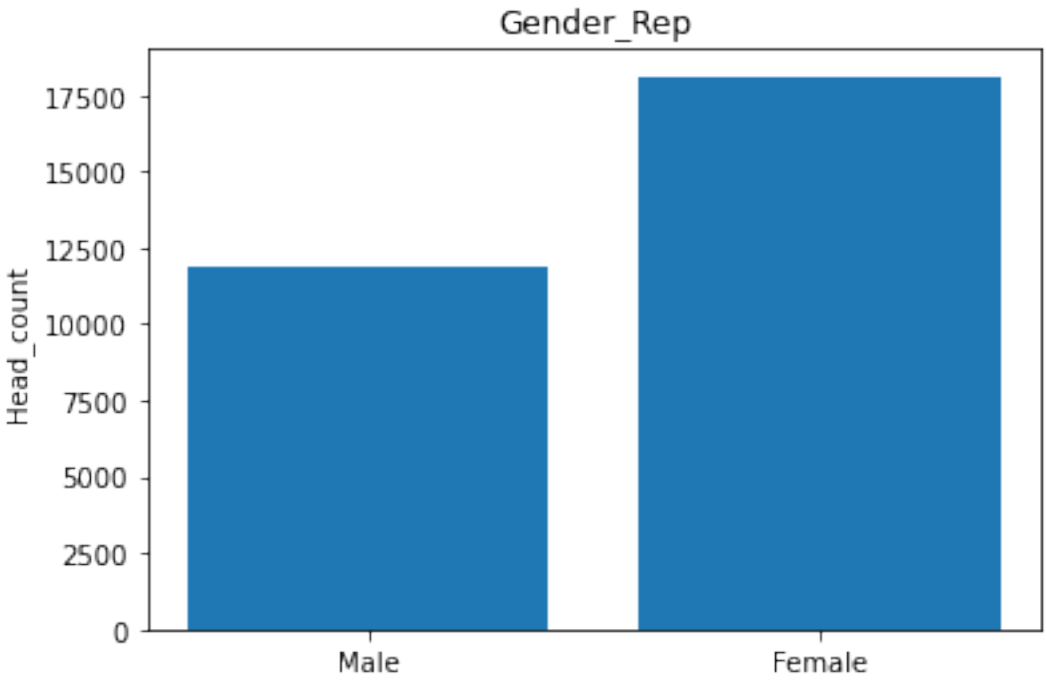
[14]: <AxesSubplot:xlabel='Gender'>



[15]:
```
df['Gender'] = df['Gender'].fillna(df['Gender'].mode())
```

[16]:
```
# To ubderstand Data Distribution

Gender = {'Male':11888,
      'Female':18112}
fig,ax= plt.subplots()
ax.bar(Gender.keys(),Gender.values())
ax.set(title='Gender_Rep',ylabel='Head_count')
```

[16]: [Text(0.5, 1.0, 'Gender_Rep'), Text(0, 0.5, 'Head_count')]



0.3.2 Feature 2:- Customer Type

[17]:
```
df['Customer Type'].value_counts()
```

[17]:
```
Loyal Customer       21170
disloyal Customer     4794
Disloyal Customer        2
Name: Customer Type, dtype: int64
```

[18]:
```
df['Customer Type'].nunique()
```

[18]:
```
3
```

[19]:
```
df['Customer Type'].isnull().mean()*100
```

[19]:
```
0.0384970742223591
```

[20]:
```
df['Customer Type'].isnull().mean()*100
```

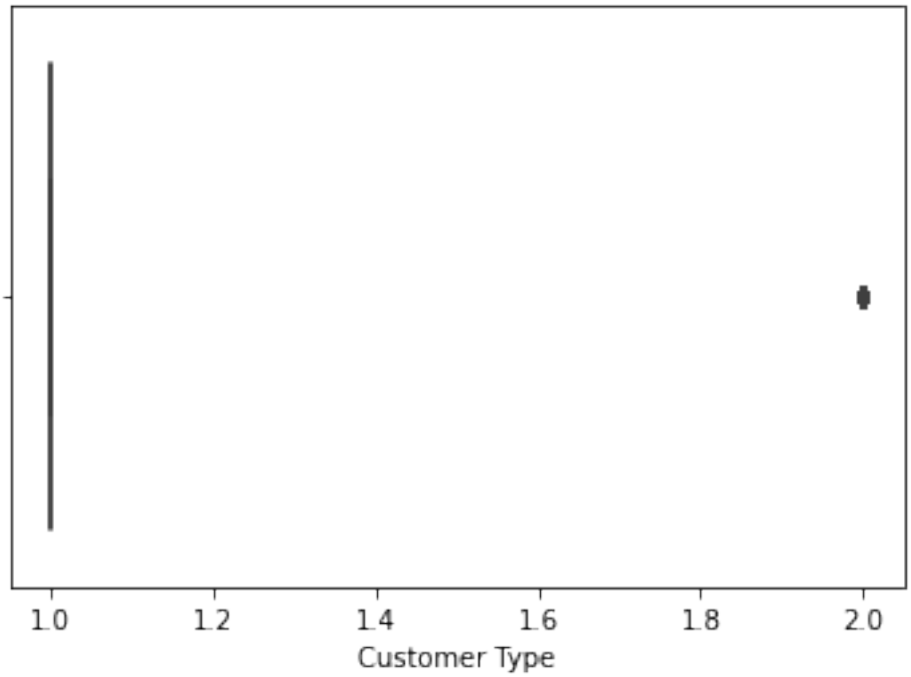
[20]:
```
0.0384970742223591
```

[21]:
```
df['Customer Type'].replace({'Loyal Customer':1,'disloyal Customer':2,'Disloyal␣
↪Customer':2},inplace=True)
df.head()
```

[21]:
```
   Gender  Customer Type  Age   Type of Travel     Class  Flight Distance  \
0       2            1.0   52  Business travel       Eco              160
1       2            1.0   36  Business travel  Business             2863
2       1            2.0   20  Business travel       Eco              192
3       1            1.0   44  Business travel  Business             3377
4       2            1.0   49  Business travel       Eco             1182

   Inflight wifi service  Departure/Arrival time convenient  \
0                      5                                  4
1                      1                                  1
2                      2                                  0
3                      0                                  0
4                      2                                  3

   Ease of Online booking  Gate location  …  Inflight entertainment  \
0                       3              4  …                       5
1                       3              1  …                       4
2                       2              4  …                       2
3                       0              2  …                       1
4                       4              3  …                       2

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

   Inflight service  Cleanliness  Departure Delay in Minutes  \
0                 5            5                          50
1                 4            5                           0
2                 2            2                           0
3                 1            4                           0
4                 2            4                           0

   Arrival Delay in Minutes             satisfaction
0                      44.0                satisfied
1                       0.0                satisfied
2                       0.0  neutral or dissatisfied
3                       6.0                satisfied
4                      20.0                satisfied

[5 rows x 23 columns]
```

[22]:
```
sns.boxplot(df['Customer Type'])
```

[22]:
```
<AxesSubplot:xlabel='Customer Type'>
```



[23]:
```
df['Customer Type'].value_counts()
```

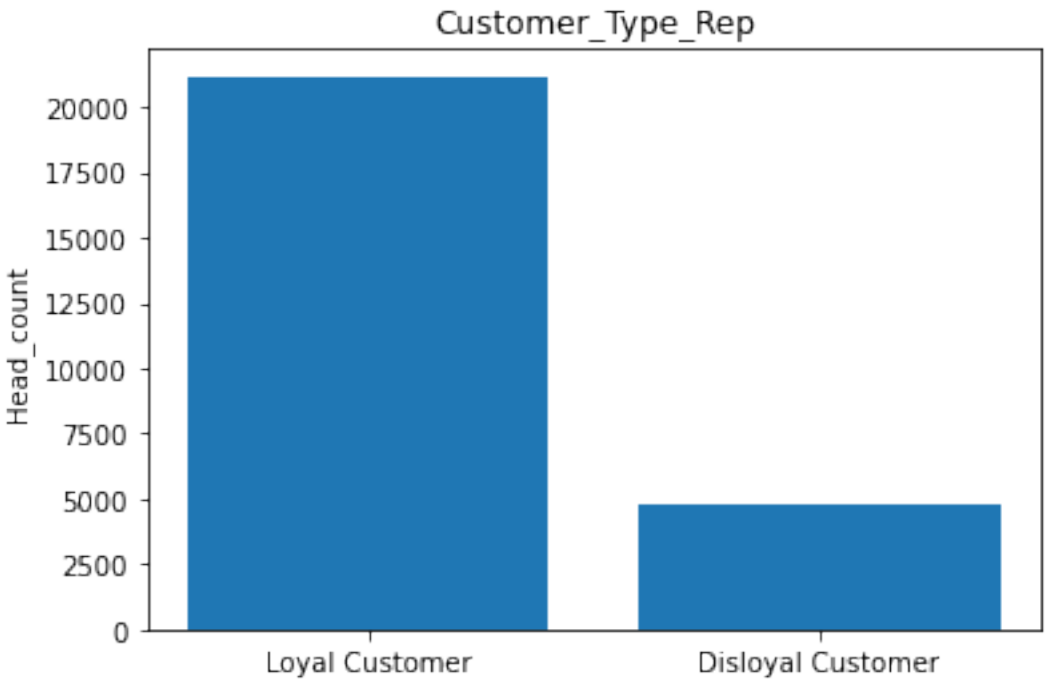
[23]:
```
1.0    21170
2.0     4796
Name: Customer Type, dtype: int64
```

[24]:
```
df['Customer Type'].value_counts(normalize=True)
```

[24]:
```
1.0    0.815297
2.0    0.184703
Name: Customer Type, dtype: float64
```

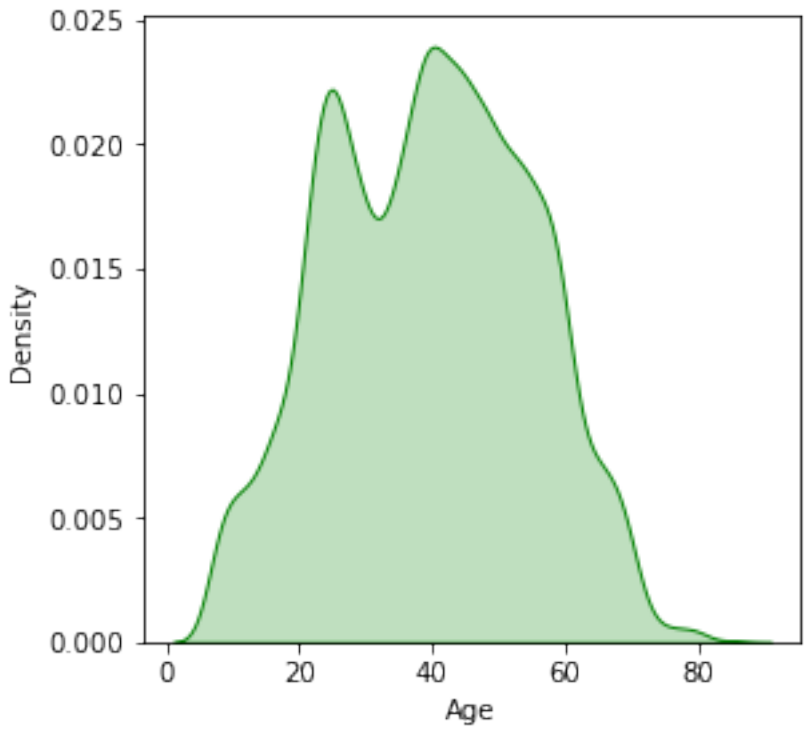
[25]:
```
Customer_Type = {'Loyal Customer':21170,
      'Disloyal Customer':4796}
fig,ax= plt.subplots()
ax.bar(Customer_Type.keys(),Customer_Type.values())
ax.set(title='Customer_Type_Rep',ylabel='Head_count')
```

[25]:
```
[Text(0.5, 1.0, 'Customer_Type_Rep'), Text(0, 0.5, 'Head_count')]
```



0.3.3 Feature 3:-Age

[26]:
```
# Understand Data distribution usning  graph

plt.figure(figsize=(15,25))
a = 5
b = 3
c = 1

colors = ['green']
plt.subplot(a,b,c)
ax = sns.kdeplot(x =df['Age'], fill=True, color=colors)

plt.show()
```



0.3.4 Feature 3:-Type of Travel

[27]:
```
df['Type of Travel'].value_counts()
```

[27]:
```
Business travel    18038
Personal Travel     7938
Name: Type of Travel, dtype: int64
```

[28]:
```
df['Type of Travel'].nunique()
```

[28]:
```
2
```

[29]:
```
# We Check Here Null Persentage

df['Type of Travel'].isnull().mean()*100
```

[29]:
```
0.0
```

[30]:
```
# Replace Characterized data into Numeric form

df['Type of Travel'].replace({'Business travel':1,'Personal Travel':
 ↪2},inplace=True)
df.head()
```

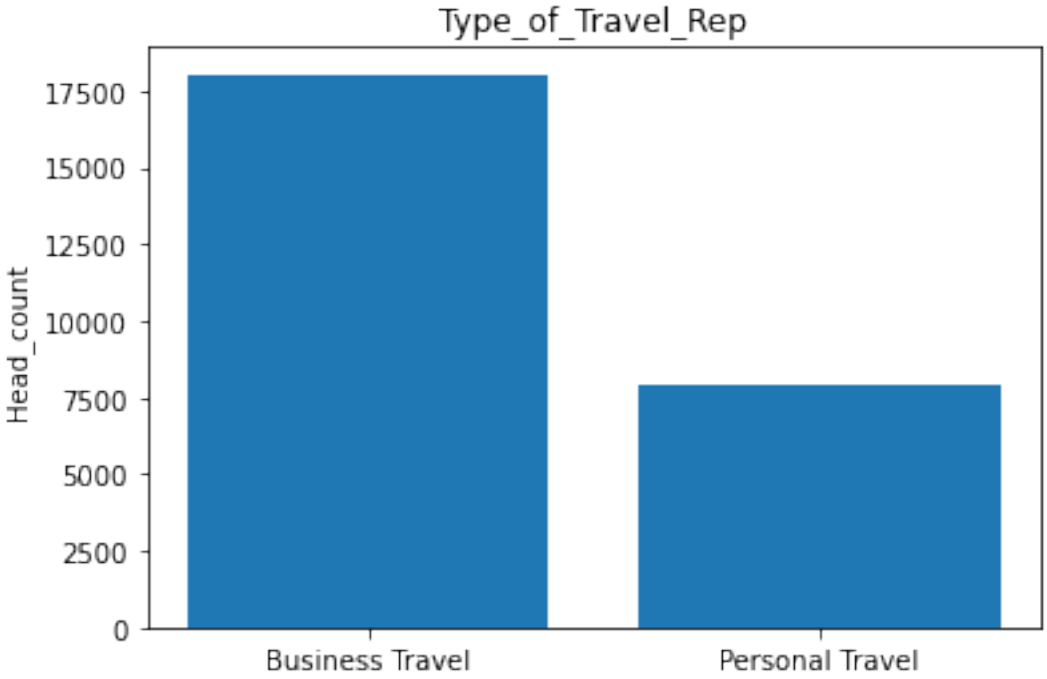
[30]:
```
   Gender  Customer Type  Age  Type of Travel     Class  Flight Distance  \
0       2            1.0   52               1       Eco              160
1       2            1.0   36               1  Business             2863
2       1            2.0   20               1       Eco              192
3       1            1.0   44               1  Business             3377
4       2            1.0   49               1       Eco             1182

   Inflight wifi service  Departure/Arrival time convenient  \
0                      5                                  4
1                      1                                  1
2                      2                                  0
3                      0                                  0
4                      2                                  3

   Ease of Online booking  Gate location  …  Inflight entertainment  \
0                       3              4  …                       5
1                       3              1  …                       4
2                       2              4  …                       2
3                       0              2  …                       1
4                       4              3  …                       2

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

   Inflight service  Cleanliness  Departure Delay in Minutes  \
0                 5            5                          50
1                 4            5                           0
2                 2            2                           0
3                 1            4                           0
4                 2            4                           0

   Arrival Delay in Minutes             satisfaction
0                      44.0                satisfied
1                       0.0                satisfied
2                       0.0  neutral or dissatisfied
3                       6.0                satisfied
4                      20.0                satisfied

[5 rows x 23 columns]
```

[31]:
```
df['Type of Travel'].value_counts(normalize=True)
```

[31]:
```
1    0.69441
2    0.30559
Name: Type of Travel, dtype: float64
```

[32]:
```
Type_of_Travel = {'Business Travel':18038,
      'Personal Travel':7938}
fig,ax= plt.subplots()
ax.bar(Type_of_Travel.keys(),Type_of_Travel.values())
ax.set(title='Type_of_Travel_Rep',ylabel='Head_count')
```

[32]:
```
[Text(0.5, 1.0, 'Type_of_Travel_Rep'), Text(0, 0.5, 'Head_count')]
```



0.3.5 Feature 4:- Class

[33]:
```
df['Class'].value_counts()
```

[33]:
```
Business    12490
Eco         11563
Eco Plus     1916
Eco plus        5
eco             2
Name: Class, dtype: int64
```

[34]:
```
df['Class'].nunique()
```

[34]:
```
5
```

[35]:
```
df['Class'].isnull().mean()*100
```

[35]:
```
0.0
```

[36]:
```
df['Class'].replace({'Business':1,'Eco Plus':2,'Eco plus':2,'eco':3,'Eco':
  ↪3},inplace=True)
df.head()
```

[36]:

	Gender	Customer Type	Age	Type of Travel	Class	Flight Distance	\
0	2	1.0	52	1	3	160	
1	2	1.0	36	1	1	2863	
2	1	2.0	20	1	3	192	
3	1	1.0	44	1	1	3377	
4	2	1.0	49	1	3	1182	

	Inflight wifi service	Departure/Arrival time convenient	\
0	5	4	
1	1	1	
2	2	0	
3	0	0	
4	2	3	

	Ease of Online booking	Gate location	…	Inflight entertainment	\
0	3	4	…	5	
1	3	1	…	4	
2	2	4	…	2	
3	0	2	…	1	
4	4	3	…	2	

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

	Inflight service	Cleanliness	Departure Delay in Minutes	\
0	5	5	50	
1	4	5	0	
2	2	2	0	
3	1	4	0	
4	2	4	0	

	Arrival Delay in Minutes	satisfaction
0	44.0	satisfied

```
1                      0.0                satisfied
2                      0.0  neutral or dissatisfied
3                      6.0                satisfied
4                     20.0                satisfied

[5 rows x 23 columns]
```

[37]:
```
df['Class'].value_counts()
```

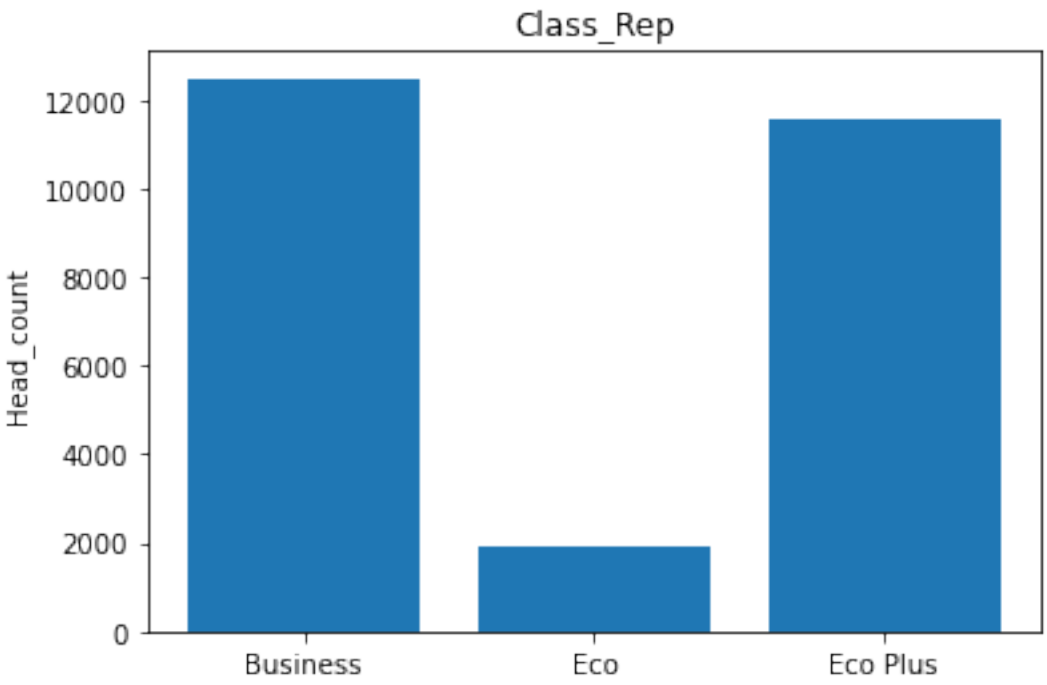
[37]:
```
1    12490
3    11565
2     1921
Name: Class, dtype: int64
```

[38]:
```
df['Class'].value_counts(normalize=True)
```

[38]:
```
1    0.480828
3    0.445219
2    0.073953
Name: Class, dtype: float64
```

[39]:
```
Class = {'Business':12490,
      'Eco':1921,'Eco Plus':11565}
fig,ax= plt.subplots()
ax.bar(Class.keys(),Class.values())
ax.set(title='Class_Rep',ylabel='Head_count')
```

[39]:
```
[Text(0.5, 1.0, 'Class_Rep'), Text(0, 0.5, 'Head_count')]
```



0.3.6 Feature 5:- Flight Distance

[40]:
```
df['Flight Distance'].value_counts()
```

[40]:
```
337     181
594     110
862     104
2475    101
447      95
       ...
1168      1
1381      1
2314      1
3883      1
3518      1
Name: Flight Distance, Length: 3281, dtype: int64
```

[41]:
```
df['Flight Distance'].nunique()
```

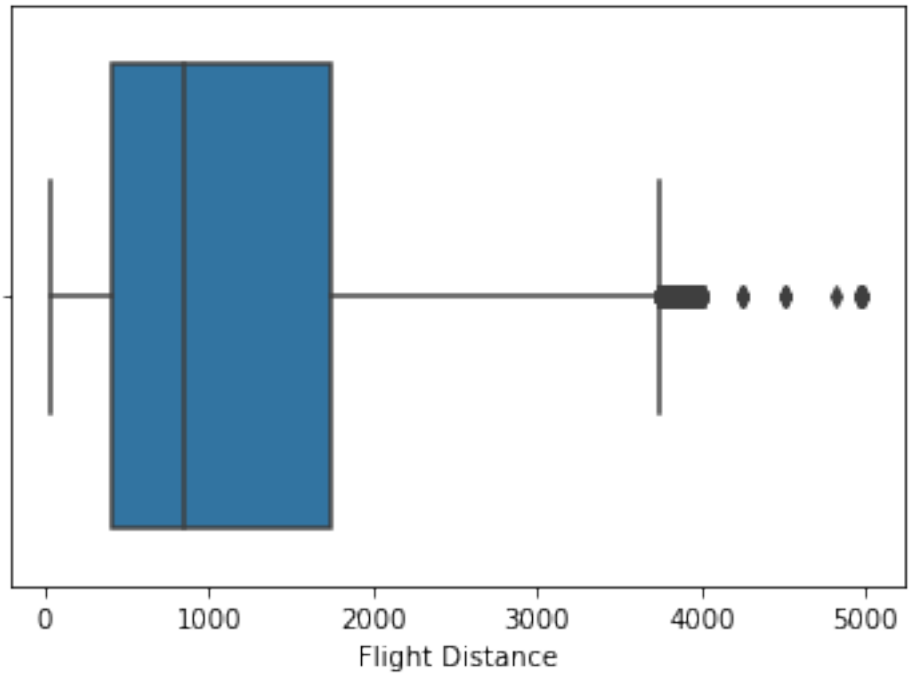
[41]:
```
3281
```

[42]:
```
df['Flight Distance'].isnull().mean()*100
```

[42]:
```
0.0
```

[43]:
```
sns.boxplot(df['Flight Distance'])
```

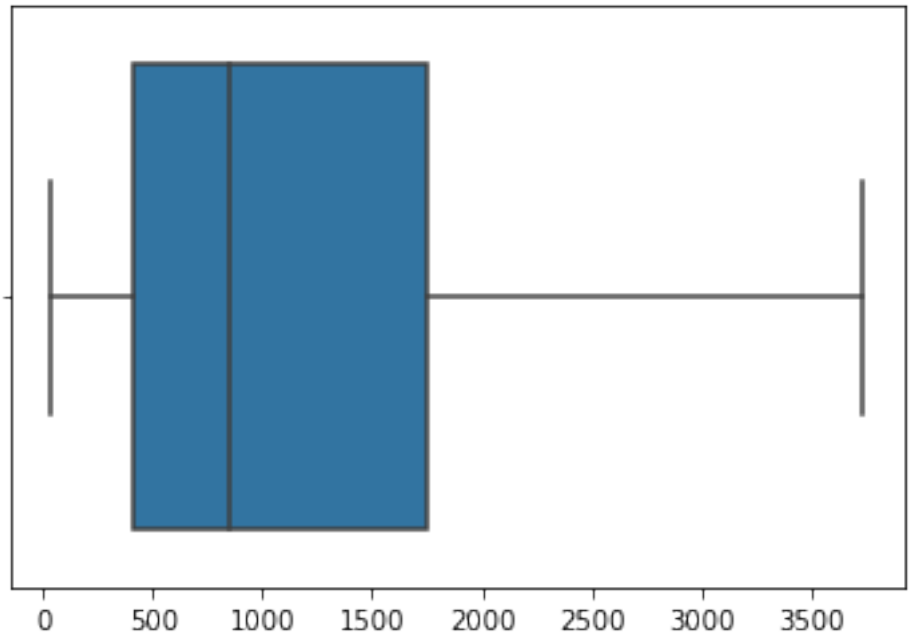
[43]: <AxesSubplot:xlabel='Flight Distance'>



[44]:
```
Outliers(df['Flight Distance'])
```

```
414.0 849.0 1744.0


3739.0 -1581.0
```

[44]:
```
array([ 160., 2863.,  192., …,  828., 1127.,  264.])
```



0.3.7 Feature 6:- Inflight wifi service

[45]:
```
df['Inflight wifi service'].value_counts()
```

[45]:
```
2    6490
3    6317
4    4981
1    4488
5    2887
0     813
Name: Inflight wifi service, dtype: int64
```

[46]:
```
df['Inflight wifi service'].nunique()
```

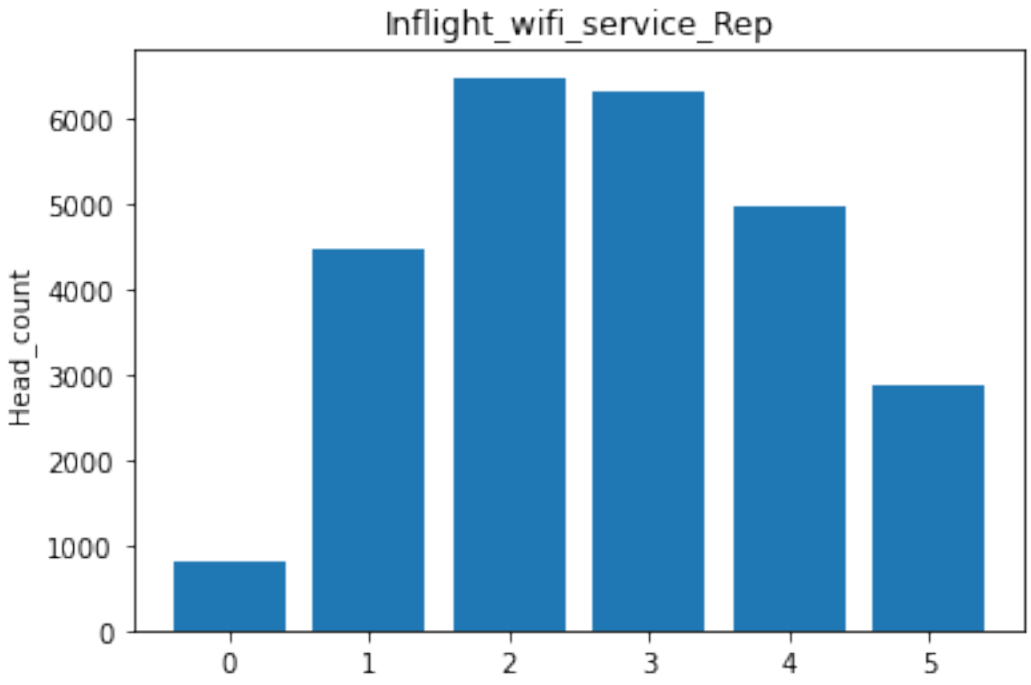
[46]:
```
6
```

[47]:
```
df['Inflight wifi service'].isnull().mean()*100
```

[47]:
```
0.0
```

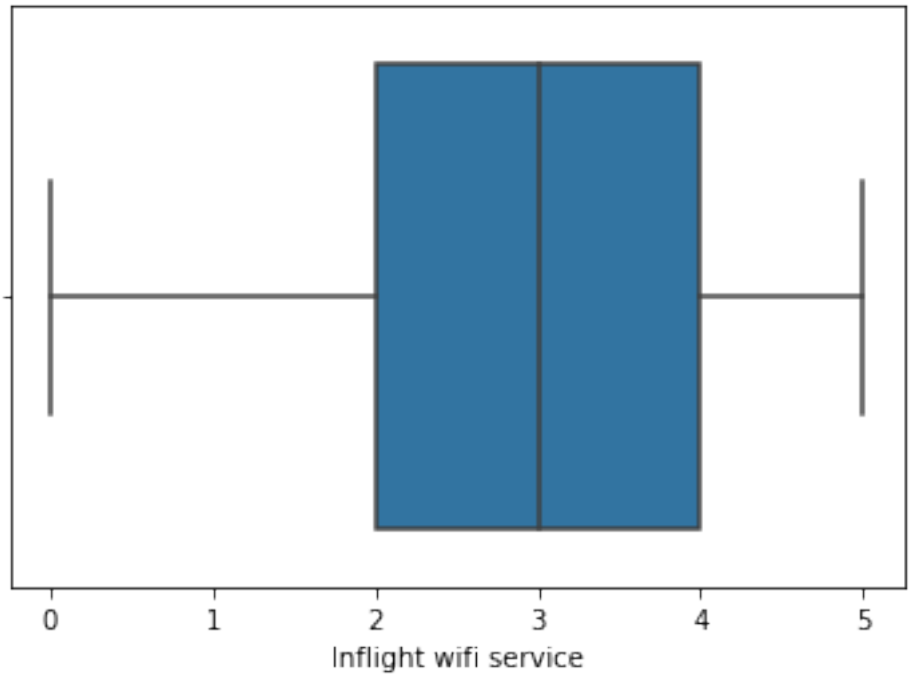
[48]:
```
Inflight_wifi_service = {'0':813,
      '1':4488,'2':6490,'3':6317,'4':4981,'5':2887}
fig,ax= plt.subplots()
ax.bar(Inflight_wifi_service.keys(),Inflight_wifi_service.values())
ax.set(title='Inflight_wifi_service_Rep',ylabel='Head_count')
```

[48]: [Text(0.5, 1.0, 'Inflight_wifi_service_Rep'), Text(0, 0.5, 'Head_count')]



[49]:
```
sns.boxplot(df['Inflight wifi service'])
```

[49]: <AxesSubplot:xlabel='Inflight wifi service'>



0.3.8 Feature 7:- Departure/Arrival time convenient

[50]:
```
df['Departure/Arrival time convenient'].value_counts()
```

[50]:
```
4    6334
5    5595
3    4412
2    4343
1    3911
0    1381
Name: Departure/Arrival time convenient, dtype: int64
```

[51]:
```
df['Departure/Arrival time convenient'].nunique()
```

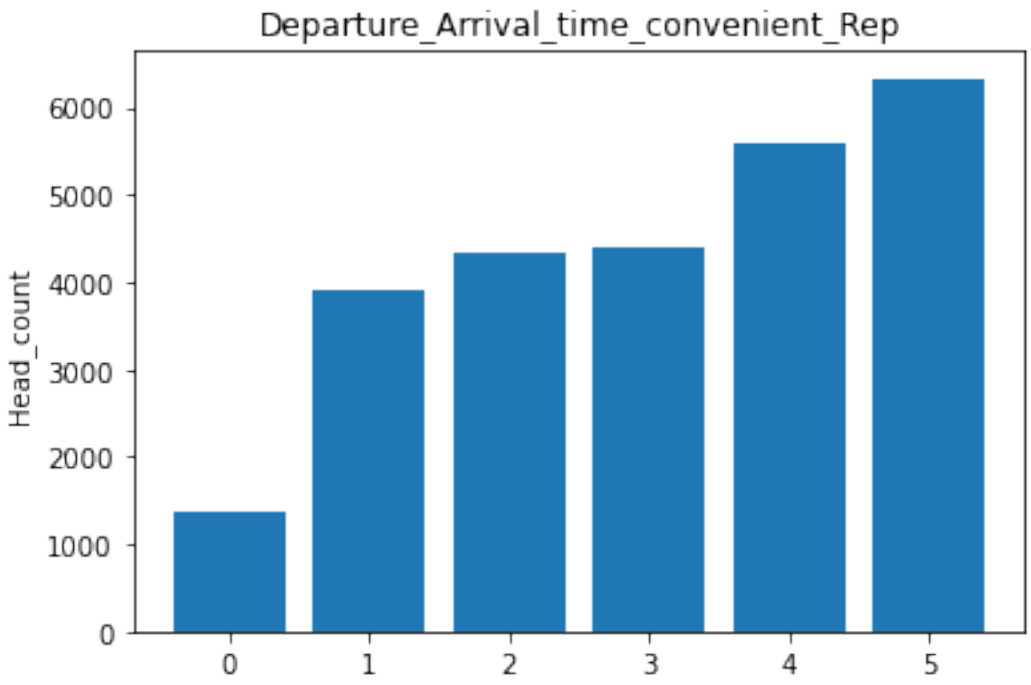
[51]:
```
6
```

[52]:
```
df['Departure/Arrival time convenient'].isnull().mean()*100
```

[52]:
```
0.0
```

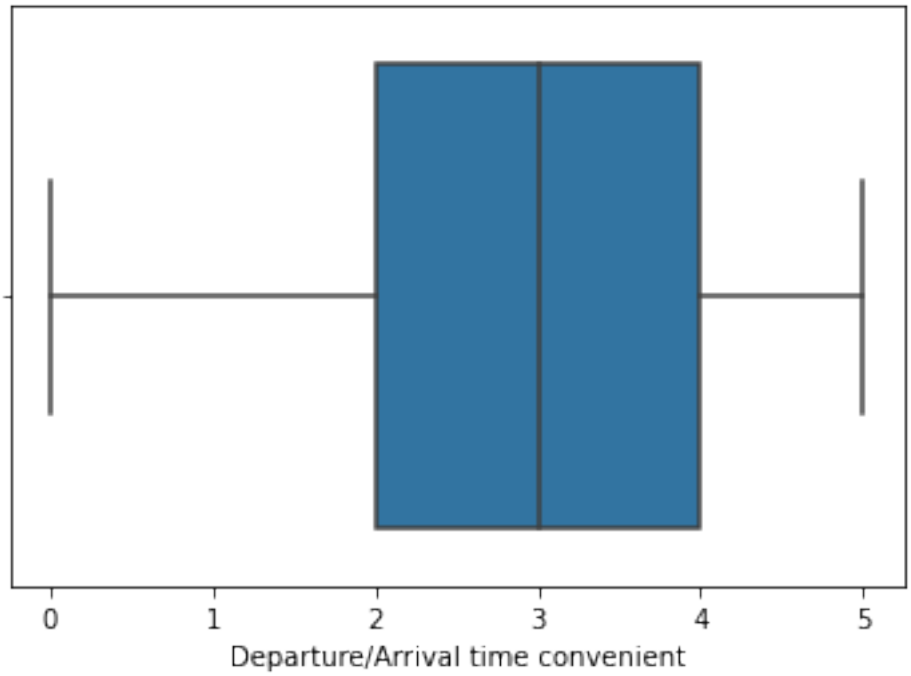
[53]:
```
Departure_Arrival_time_convenient = {'0':1381,
      '1':3911,'2':4343,'3':4412,'4':5595,'5':6334}
fig,ax= plt.subplots()
ax.bar(Departure_Arrival_time_convenient.
 ↪keys(),Departure_Arrival_time_convenient.values())
ax.set(title='Departure_Arrival_time_convenient_Rep',ylabel='Head_count')
```

[53]:
```
[Text(0.5, 1.0, 'Departure_Arrival_time_convenient_Rep'),
 Text(0, 0.5, 'Head_count')]
```



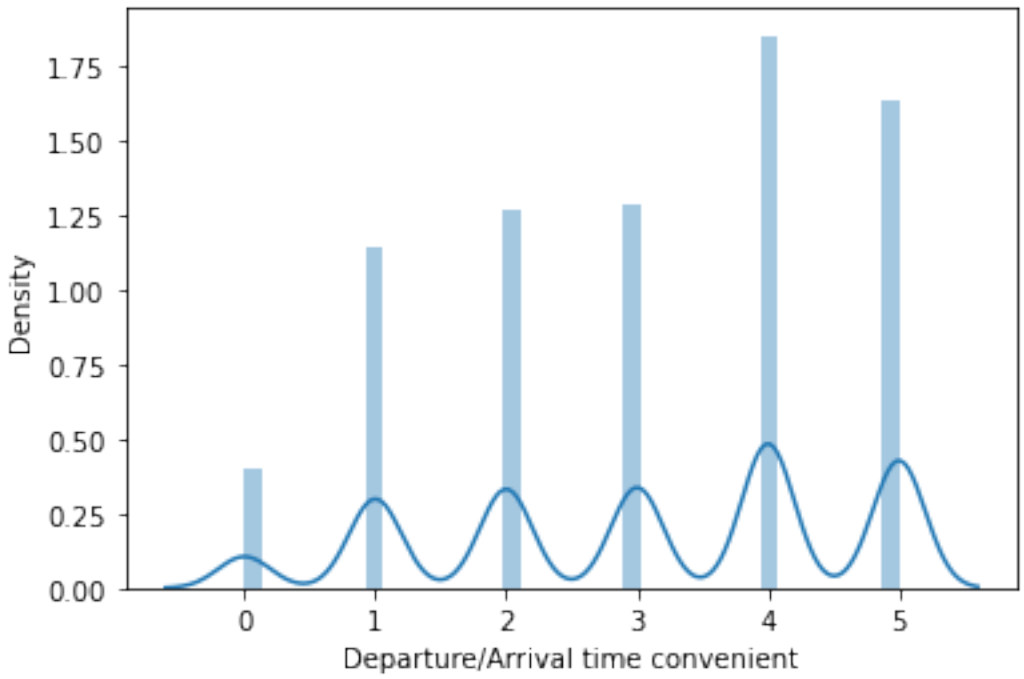
[54]:
```
sns.boxplot(df['Departure/Arrival time convenient'])
```

[54]: <AxesSubplot:xlabel='Departure/Arrival time convenient'>



[55]:
```
sns.distplot(df['Departure/Arrival time convenient'])
```

[55]: <AxesSubplot:xlabel='Departure/Arrival time convenient', ylabel='Density'>



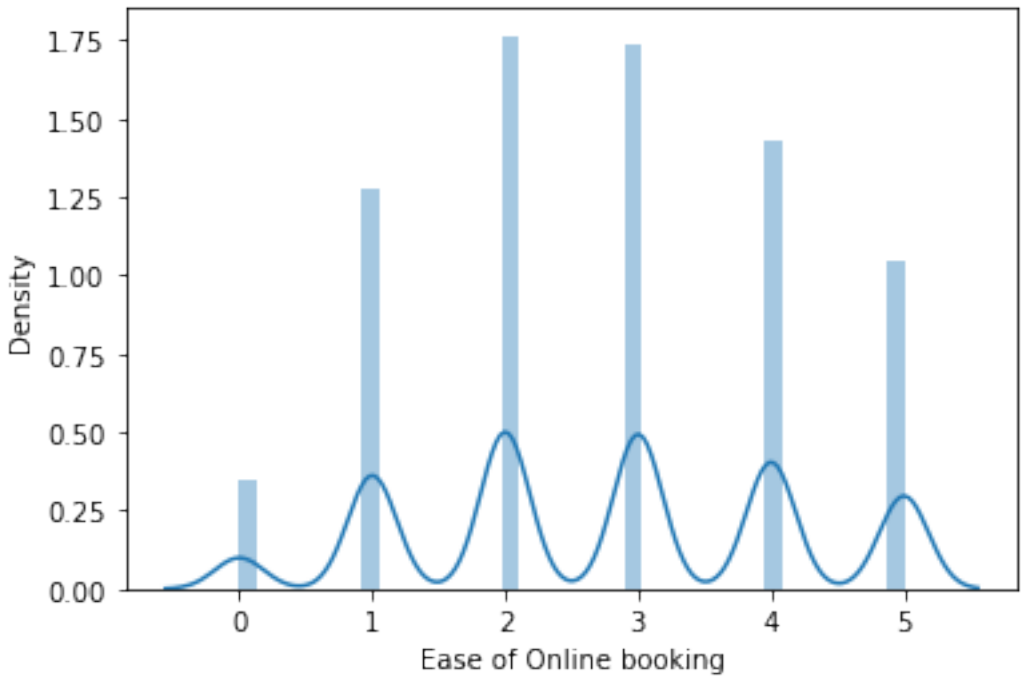
0.3.9 Feature 8- Ease of Online booking

[56]:
```
df['Ease of Online booking'].value_counts()
```

[56]:
```
2    6030
3    5944
4    4873
1    4361
5    3573
0    1195
Name: Ease of Online booking, dtype: int64
```

[57]:
```
sns.distplot(df['Ease of Online booking'])
```

[57]: <AxesSubplot:xlabel='Ease of Online booking', ylabel='Density'>



[58]:
```
df['Ease of Online booking'].nunique()
```

[58]: 6

[59]:
```
df['Ease of Online booking'].isnull().mean()*100
```

[59]: 0.0

0.3.10 Feature 9- Gate location

[60]:
```
df['Gate location'].value_counts()
```

[60]:
```
3    7140
4    6040
2    4837
1    4429
5    3530
Name: Gate location, dtype: int64
```

[61]:
```
df['Gate location'].nunique()
```

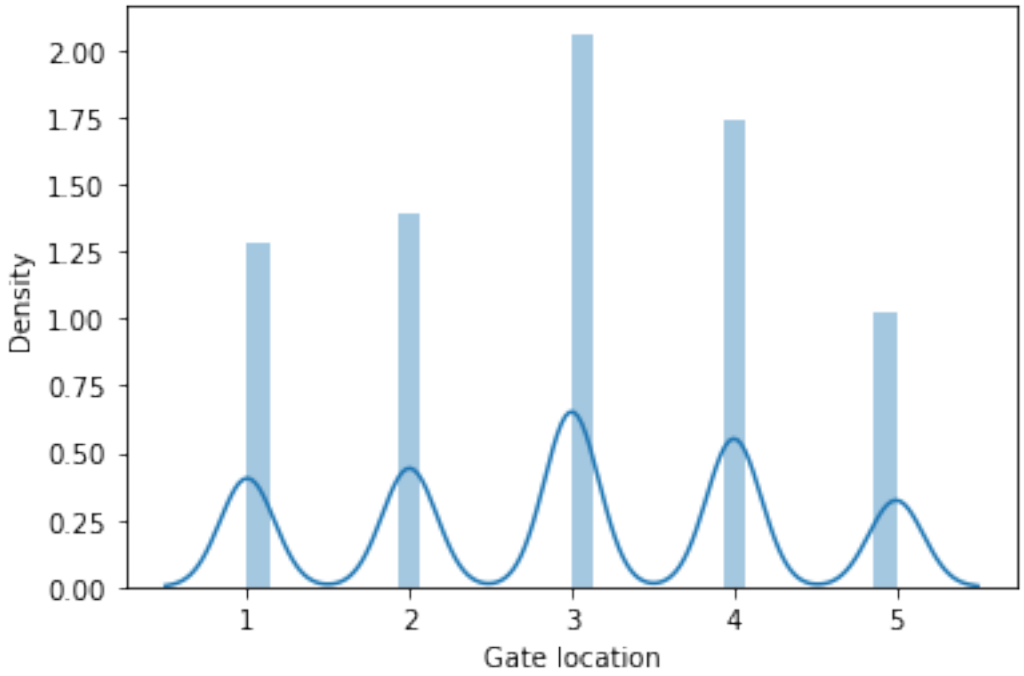
[61]: 5

[62]:
```
df['Gate location'].isnull().mean()*100
```

[62]: 0.0

[63]:
```
sns.distplot(df['Gate location'])
```

[63]: <AxesSubplot:xlabel='Gate location', ylabel='Density'>



0.3.11 Feature 10- Food and drink

[64]:
```
df['Food and drink'].value_counts()
```

[64]:
```
4    6204
5    5644
3    5494
2    5395
1    3214
0      25
Name: Food and drink, dtype: int64
```

[65]:
```
df['Food and drink'].nunique()
```

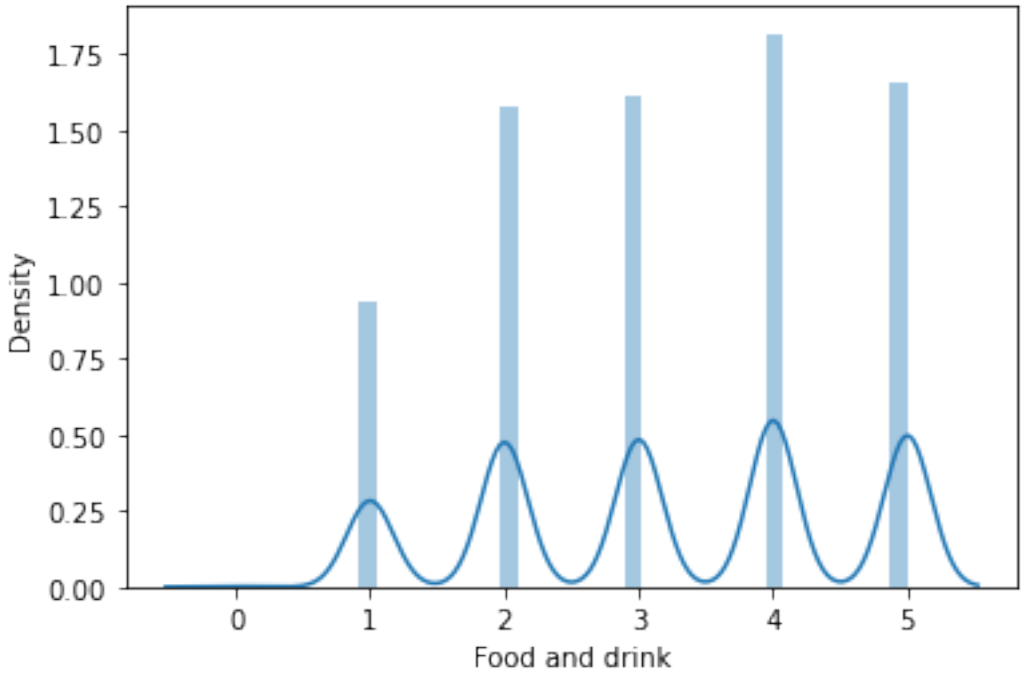
[65]: 6

[66]:
```
df['Food and drink'].isnull().mean()*100
```

[66]: 0.0

[67]:
```
sns.distplot(df['Food and drink'])
```

[67]: <AxesSubplot:xlabel='Food and drink', ylabel='Density'>



0.3.12 Feature 11- Online boarding

[68]:
```
df['Online boarding'].value_counts()
```

[68]:
```
4    7706
3    5313
5    5307
2    4429
1    2569
0     652
Name: Online boarding, dtype: int64
```

[69]:
```
df['Online boarding'].nunique()
```

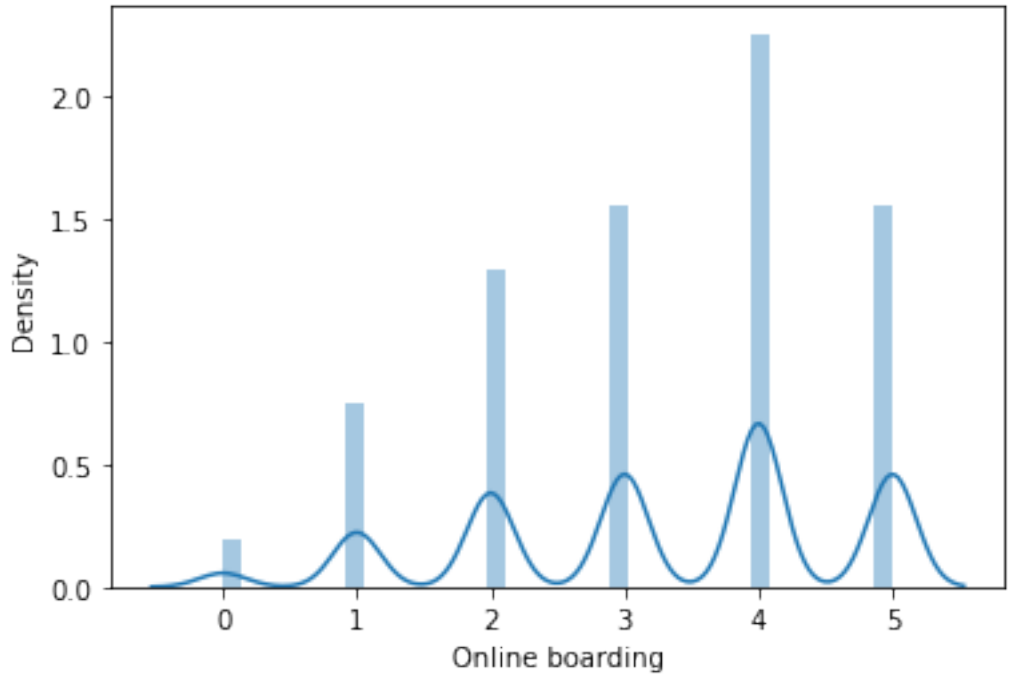
[69]: 6

[70]:
```
df['Online boarding'].isnull().mean()*100
```

[70]: 0.0

[71]:
```
sns.distplot(df['Online boarding'])
```

[71]: <AxesSubplot:xlabel='Online boarding', ylabel='Density'>



0.3.13 Feature 12- Seat comfort

[72]:
```
df['Seat comfort'].value_counts()
```

[72]:
```
4    7991
5    6688
3    4632
2    3632
1    3033
Name: Seat comfort, dtype: int64
```

[73]:
```
df['Seat comfort'].nunique()
```

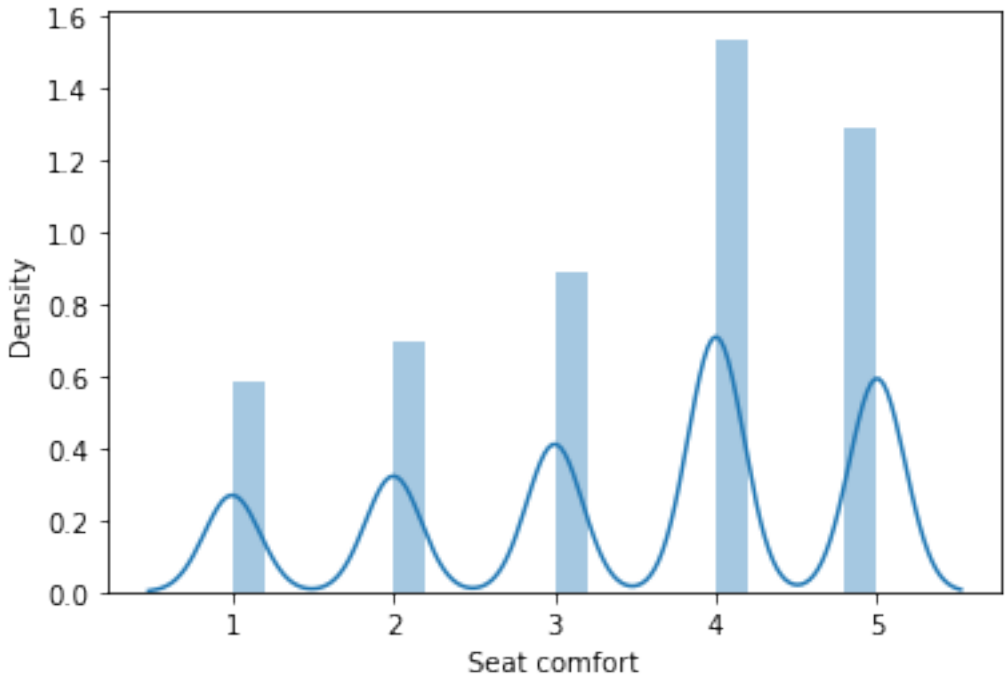
[73]: 5

[74]:
```
df['Seat comfort'].isnull().mean()*100
```

[74]: 0.0

[75]:
```
sns.distplot(df['Seat comfort'])
```

[75]:
```
<AxesSubplot:xlabel='Seat comfort', ylabel='Density'>
```



0.3.14 Feature 13- Inflight entertainment

[76]:
```
df['Inflight entertainment'].value_counts()
```

[76]:
```
4    7368
5    6331
3    4745
2    4331
1    3197
0       4
Name: Inflight entertainment, dtype: int64
```

[77]:
```
df['Inflight entertainment'].nunique()
```

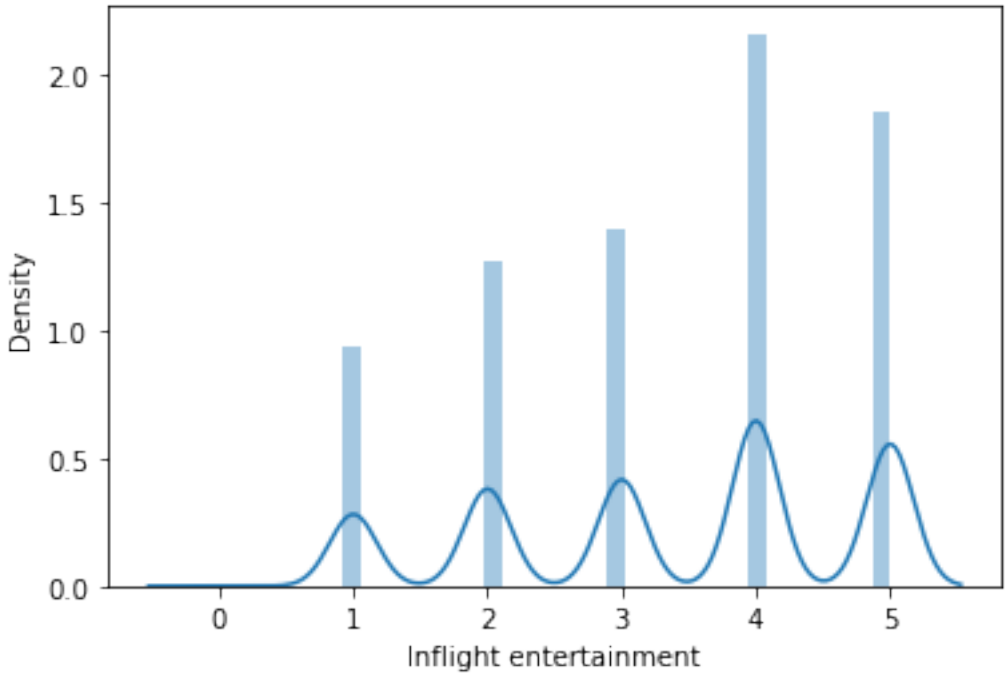
[77]:
```
6
```

[78]:
```
df['Inflight entertainment'].isnull().mean()*100
```

[78]:
```
0.0
```

[79]:
```
sns.distplot(df['Inflight entertainment'])
```

[79]:
```
<AxesSubplot:xlabel='Inflight entertainment', ylabel='Density'>
```



0.3.15 Feature 14- On-board service

[80]:
```
df['On-board service'].value_counts()
```

[80]:
```
4    7836
5    5844
3    5709
2    3670
1    2915
0       2
Name: On-board service, dtype: int64
```

[81]:
```
df['On-board service'].nunique()
```

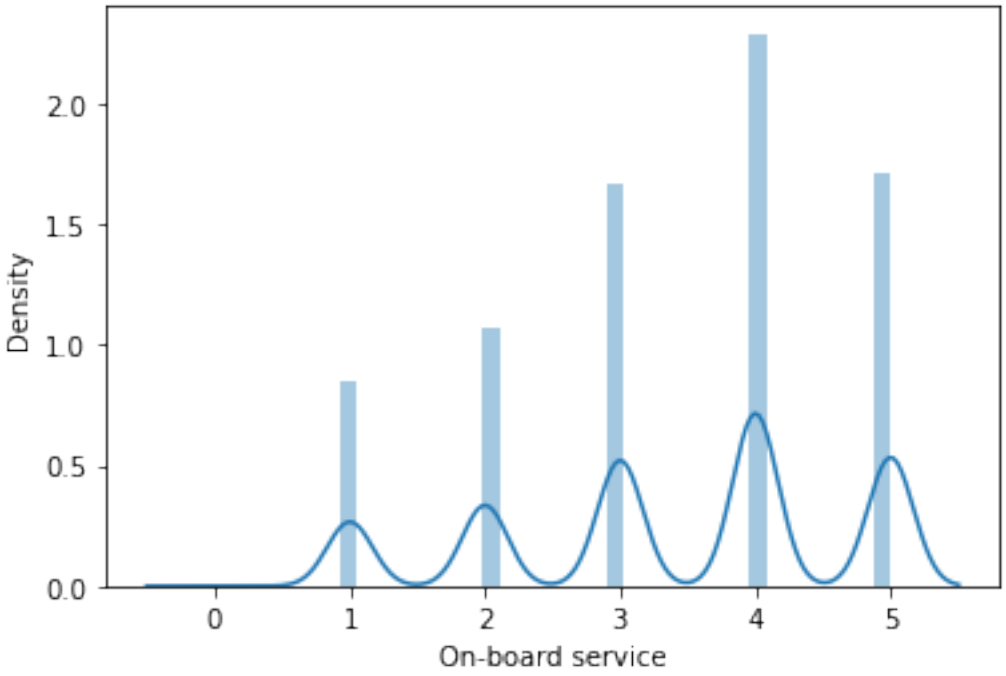
[81]:
```
6
```

[82]:
```
df['On-board service'].isnull().mean()*100
```

[82]:
```
0.0
```

[83]:
```
sns.distplot(df['On-board service'])
```

[83]:
```
<AxesSubplot:xlabel='On-board service', ylabel='Density'>
```



0.3.16 Feature 15- Leg room service

[84]:
```
df['Leg room service'].value_counts()
```

[84]:
```
4    7097
5    6238
2    5015
3    4958
1    2542
0     126
Name: Leg room service, dtype: int64
```

[85]:
```
df['Leg room service'].nunique()
```

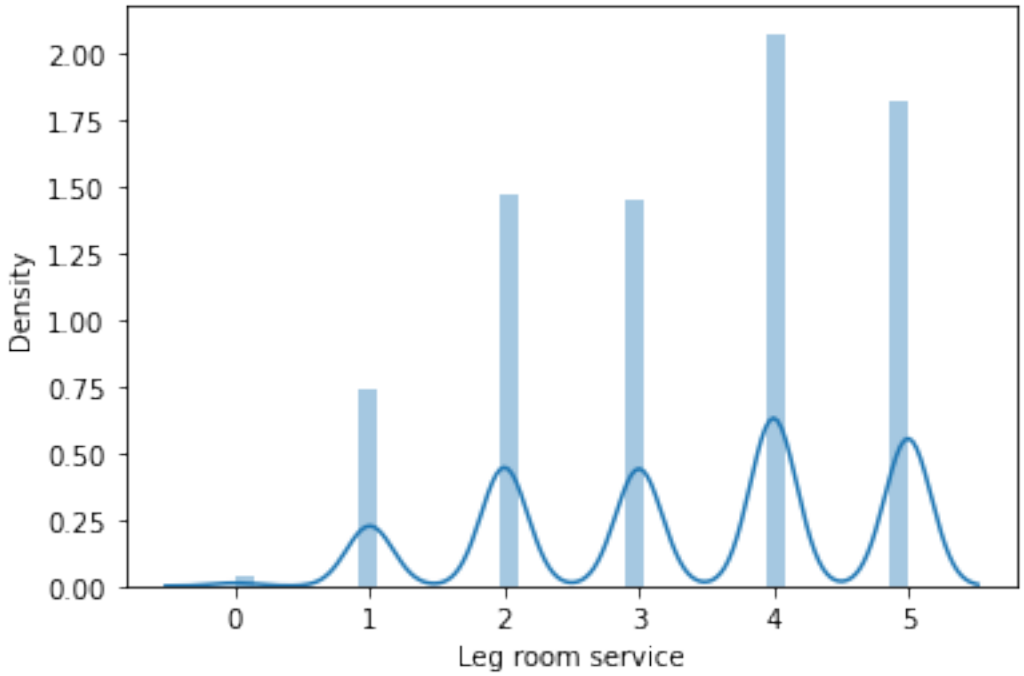
[85]:
```
6
```

[86]:
```
df['Leg room service'].isnull().mean()*100
```

[86]:
```
0.0
```

[87]:
```
sns.distplot(df['Leg room service'])
```

[87]:
```
<AxesSubplot:xlabel='Leg room service', ylabel='Density'>
```



0.3.17 Feature 16- Baggage handling

[88]:
```
df['Baggage handling'].value_counts()
```

[88]:
```
4    9378
5    6747
3    5219
2    2841
1    1791
Name: Baggage handling, dtype: int64
```

[89]:
```
df['Baggage handling'].nunique()
```

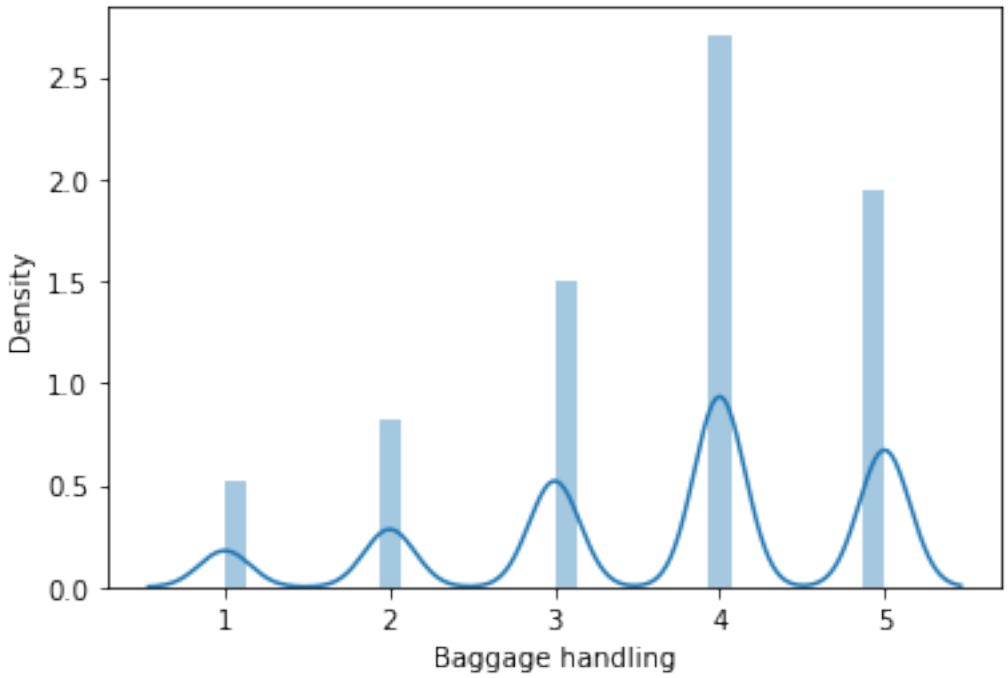
[89]:
```
5
```

[90]:
```
df['Baggage handling'].isnull().mean()*100
```

[90]:
```
0.0
```

[91]:
```
sns.distplot(df['Baggage handling'])
```

[91]: <AxesSubplot:xlabel='Baggage handling', ylabel='Density'>



0.3.18 Feature 17- Checkin service

[92]:
```
df['Checkin service'].value_counts()
```

[92]:
```
4    7278
3    7007
5    5264
1    3218
2    3209
Name: Checkin service, dtype: int64
```

[93]:
```
df['Checkin service'].nunique()
```

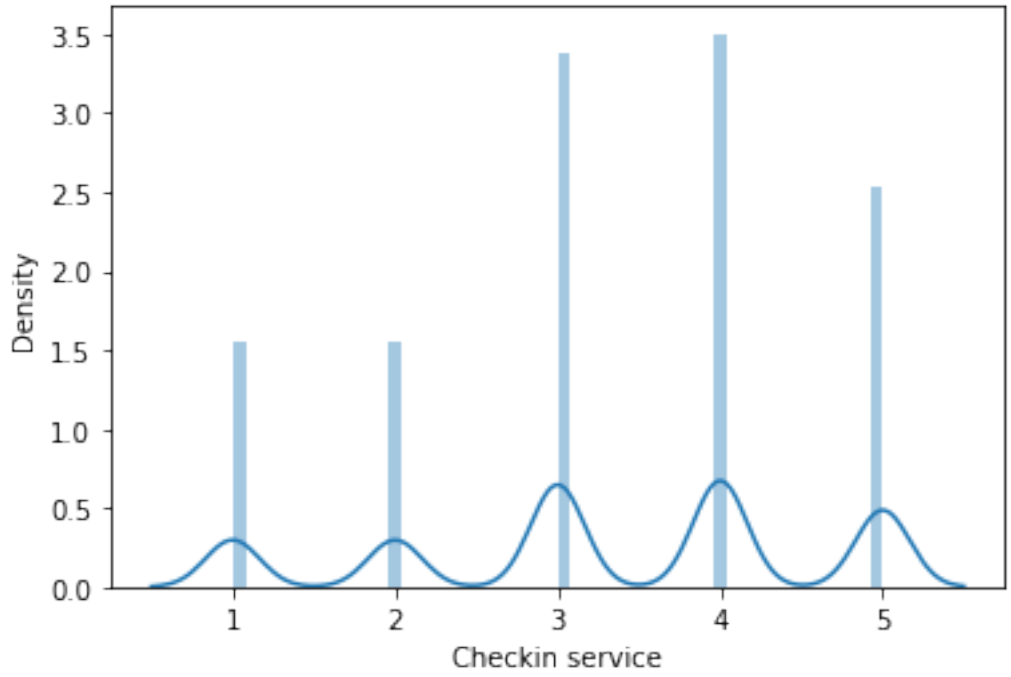
[93]: 5

[94]:
```
df['Checkin service'].isnull().mean()*100
```

[94]: 0.0

[95]:
```
sns.distplot(df['Checkin service'])
```

[95]: <AxesSubplot:xlabel='Checkin service', ylabel='Density'>



0.3.19 Feature 18- Inflight service

[96]:
```
df['Inflight service'].value_counts()
```

[96]:
```
4    9378
5    6950
3    5017
2    2851
1    1778
0       2
Name: Inflight service, dtype: int64
```

[97]:
```
df['Inflight service'].nunique()
```

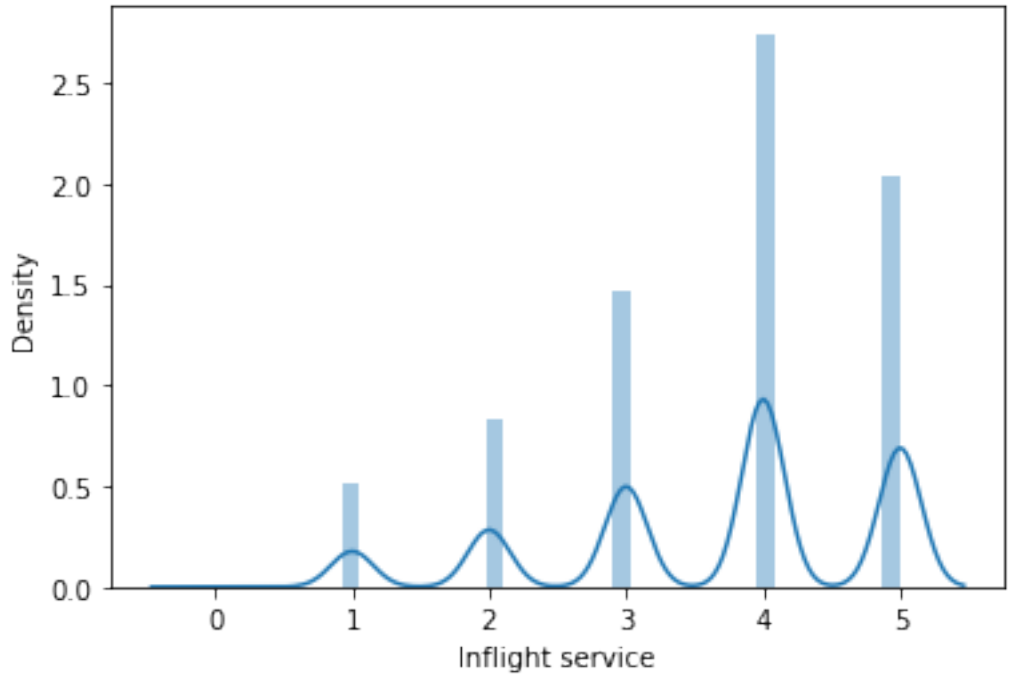
[97]:
```
6
```

[98]:
```
df['Inflight service'].isnull().mean()*100
```

[98]:
```
0.0
```

[99]:
```
sns.distplot(df['Inflight service'])
```

[99]:
```
<AxesSubplot:xlabel='Inflight service', ylabel='Density'>
```



0.3.20 Feature 19- Cleanliness

[100]:
```
df['Cleanliness'].value_counts()
```

[100]:
```
4    6790
3    6065
5    5727
2    3981
1    3411
0       2
Name: Cleanliness, dtype: int64
```

[101]:
```
df['Cleanliness'].nunique()
```

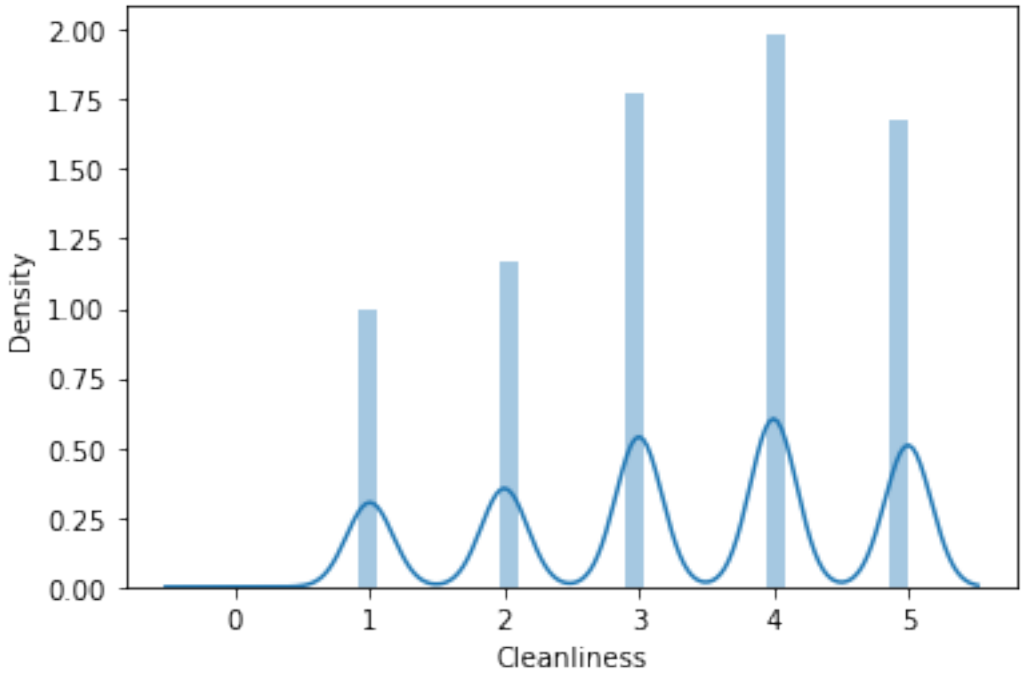
[101]:
```
6
```

[102]:
```
df['Cleanliness'].isnull().mean()*100
```

[102]:
```
0.0
```

[103]:
```
sns.distplot(df['Cleanliness'])
```

[103]:
```
<AxesSubplot:xlabel='Cleanliness', ylabel='Density'>
```



0.3.21 Feature 20- Departure Delay in Minutes

[104]:
```
df['Departure Delay in Minutes'].value_counts()
```

[104]:
```
0      14688
1        734
2        581
3        526
4        455
       ...
333        1
252        1
147        1
360        1
295        1
Name: Departure Delay in Minutes, Length: 313, dtype: int64
```

[105]:
```
df['Departure Delay in Minutes'].nunique()
```

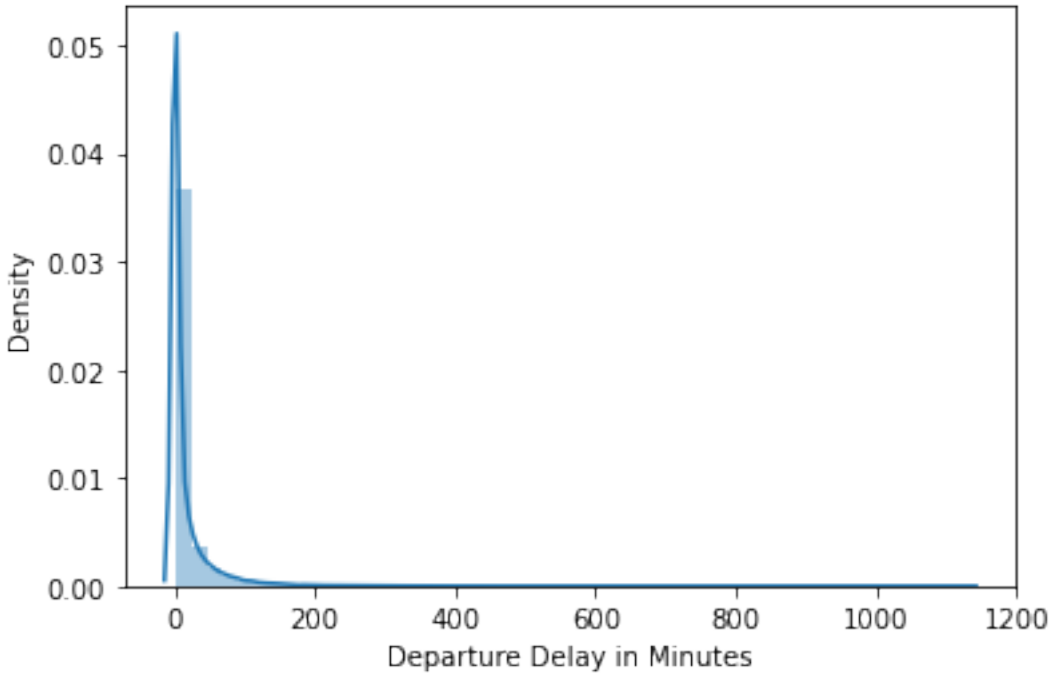
[105]:
```
313
```

[106]:
```
df['Departure Delay in Minutes'].isnull().mean()*100
```

[106]:
```
0.0
```

[107]:
```
sns.distplot(df['Departure Delay in Minutes'])
```

[107]:
```
<AxesSubplot:xlabel='Departure Delay in Minutes', ylabel='Density'>
```



0.3.22 Feature 21- Arrival Delay in Minutes

[108]:
```
df['Arrival Delay in Minutes'].value_counts()
```

[108]:
```
0.0      14594
1.0        536
2.0        523
3.0        490
4.0        466
          ...
307.0        1
285.0        1
347.0        1
795.0        1
288.0        1
Name: Arrival Delay in Minutes, Length: 320, dtype: int64
```

[109]:
```
df['Arrival Delay in Minutes'].nunique()
```

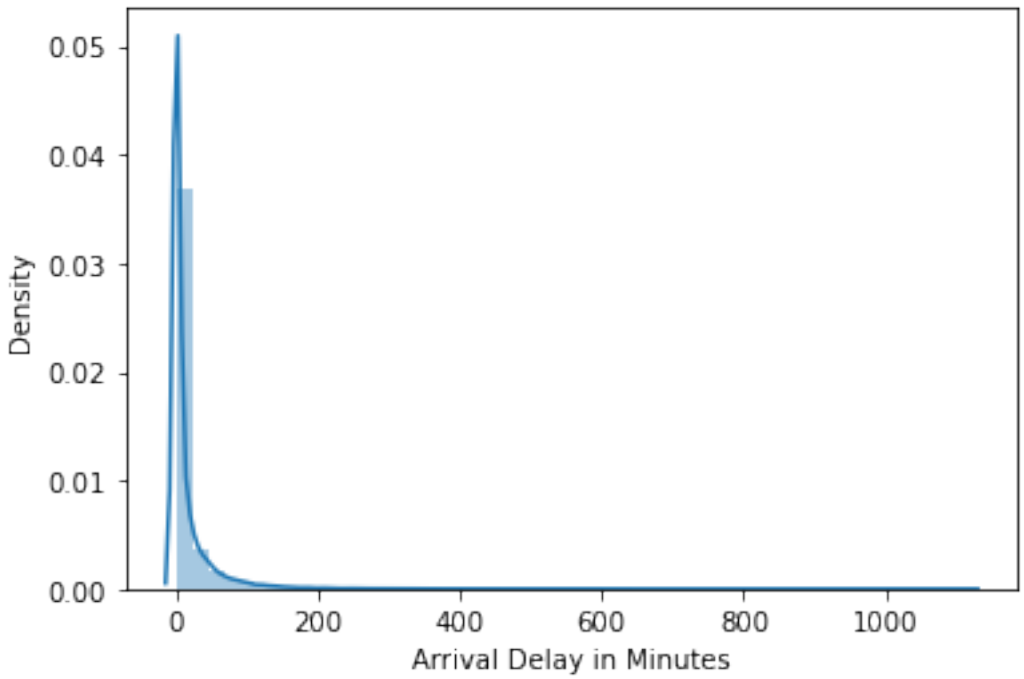
[109]:
```
320
```

[110]:
```
df['Arrival Delay in Minutes'].isnull().mean()*100
```

[110]:
```
0.31952571604558055
```

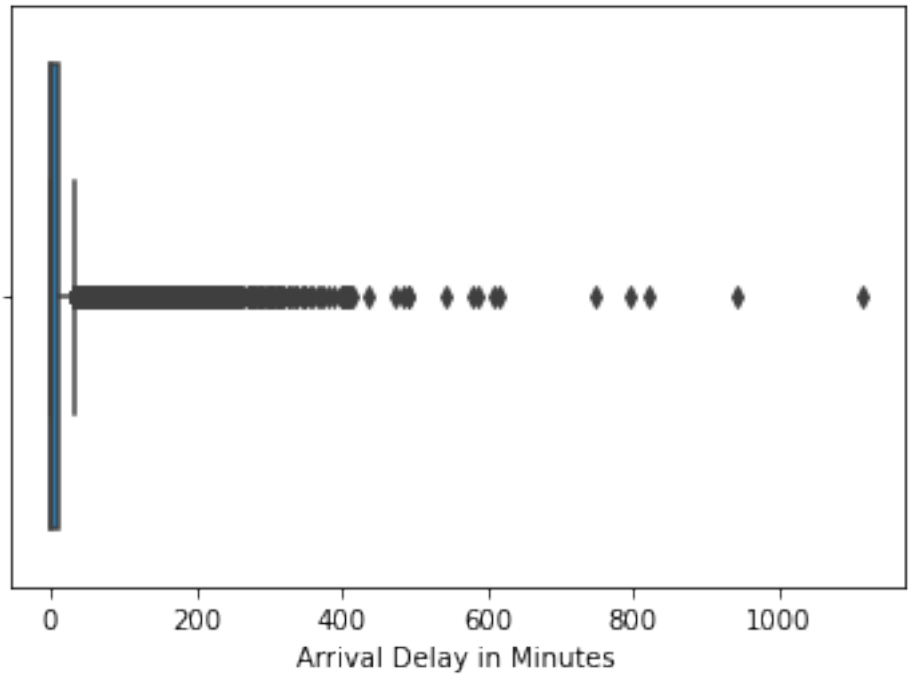
[111]:
```
sns.distplot(df['Arrival Delay in Minutes'])
```

[111]:
```
<AxesSubplot:xlabel='Arrival Delay in Minutes', ylabel='Density'>
```



[112]:
```
sns.boxplot(df['Arrival Delay in Minutes'])
```

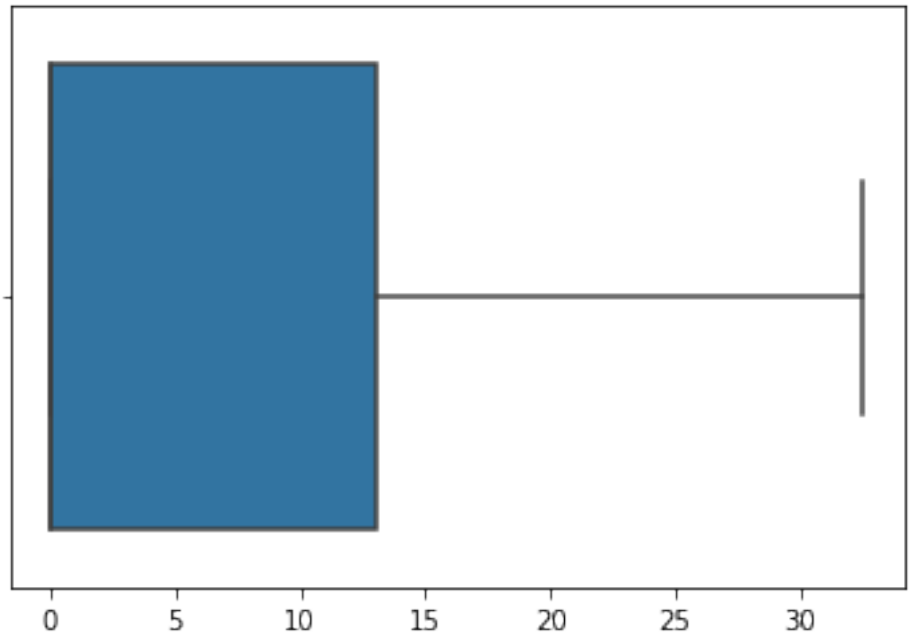
[112]:
```
<AxesSubplot:xlabel='Arrival Delay in Minutes'>
```



[113]:
```
Outliers(df['Arrival Delay in Minutes'])
```

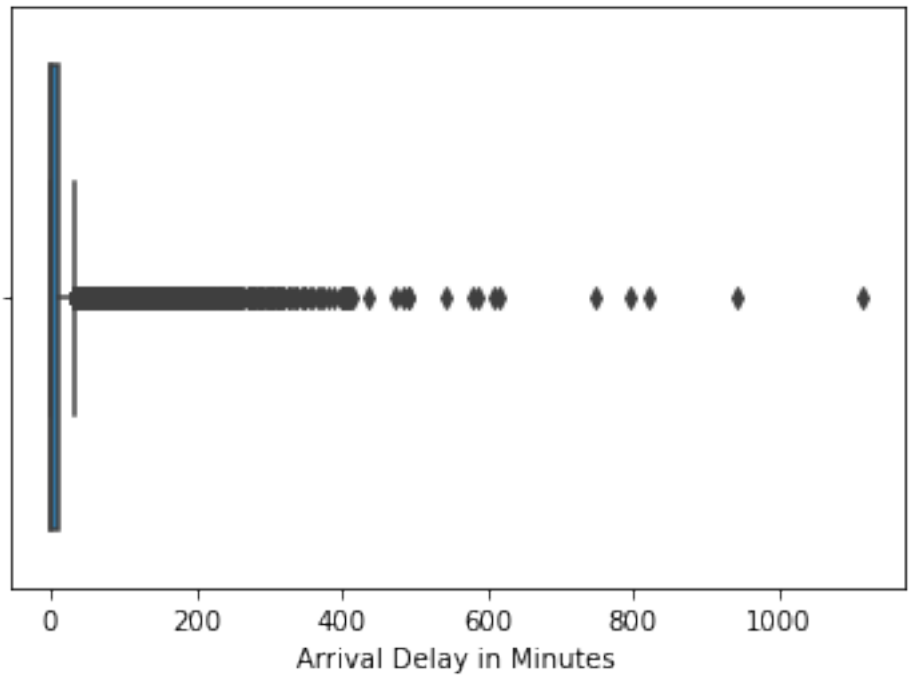
```
0.0 0.0 13.0


32.5 -19.5
```

[113]:
```
array([32.5,  0. ,  0. , …,  0. ,  0. ,  0. ])
```



[114]:
```
sns.boxplot(df['Arrival Delay in Minutes'])
```

[114]: <AxesSubplot:xlabel='Arrival Delay in Minutes'>



0.3.23 Target Variable- satisfaction

[115]:
```
df['satisfaction'].value_counts()
```

[115]:
```
neutral or dissatisfied    14573
satisfied                  11403
Name: satisfaction, dtype: int64
```

[116]:
```
df['satisfaction'].nunique()
```

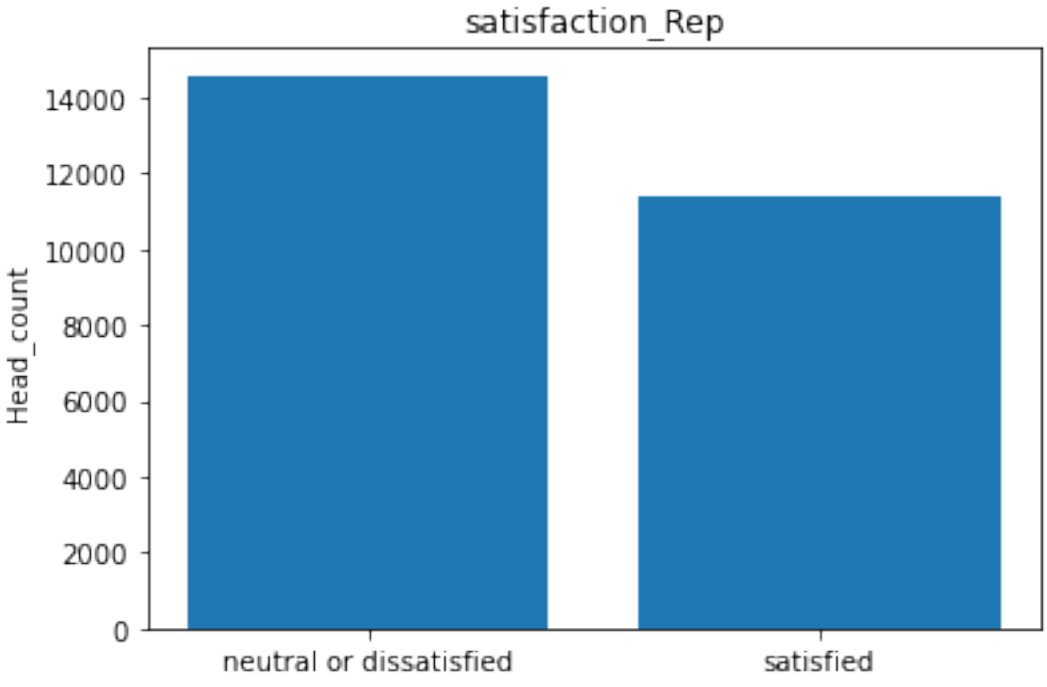
[116]:
```
2
```

[117]:
```
df['satisfaction'].isnull().mean()*100
```

[117]:
```
0.0
```

[118]:
```
satisfaction = {'neutral or dissatisfied':14573,
      'satisfied':11403}
fig,ax= plt.subplots()
ax.bar(satisfaction.keys(),satisfaction.values())
ax.set(title='satisfaction_Rep',ylabel='Head_count')
```

[118]:
```
[Text(0.5, 1.0, 'satisfaction_Rep'), Text(0, 0.5, 'Head_count')]
```



[119]:
```
df['satisfaction'].replace({'neutral or dissatisfied':0,'satisfied':
  ↪1},inplace=True)
df.head()
```

[119]:

	Gender	Customer Type	Age	Type of Travel	Class	Flight Distance
0	2	1.0	52	1	3	160
1	2	1.0	36	1	1	2863
2	1	2.0	20	1	3	192
3	1	1.0	44	1	1	3377
4	2	1.0	49	1	3	1182

	Inflight wifi service	Departure/Arrival time convenient
0	5	4
1	1	1
2	2	0
3	0	0
4	2	3

	Ease of Online booking	Gate location	…	Inflight entertainment
0	3	4	…	5
1	3	1	…	4
2	2	4	…	2
3	0	2	…	1
4	4	3	…	2

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

	Inflight service	Cleanliness	Departure Delay in Minutes
0	5	5	50
1	4	5	0
2	2	2	0
3	1	4	0
4	2	4	0

	Arrival Delay in Minutes	satisfaction
0	44.0	1
1	0.0	1
2	0.0	0
3	6.0	1
4	20.0	1

[5 rows x 23 columns]

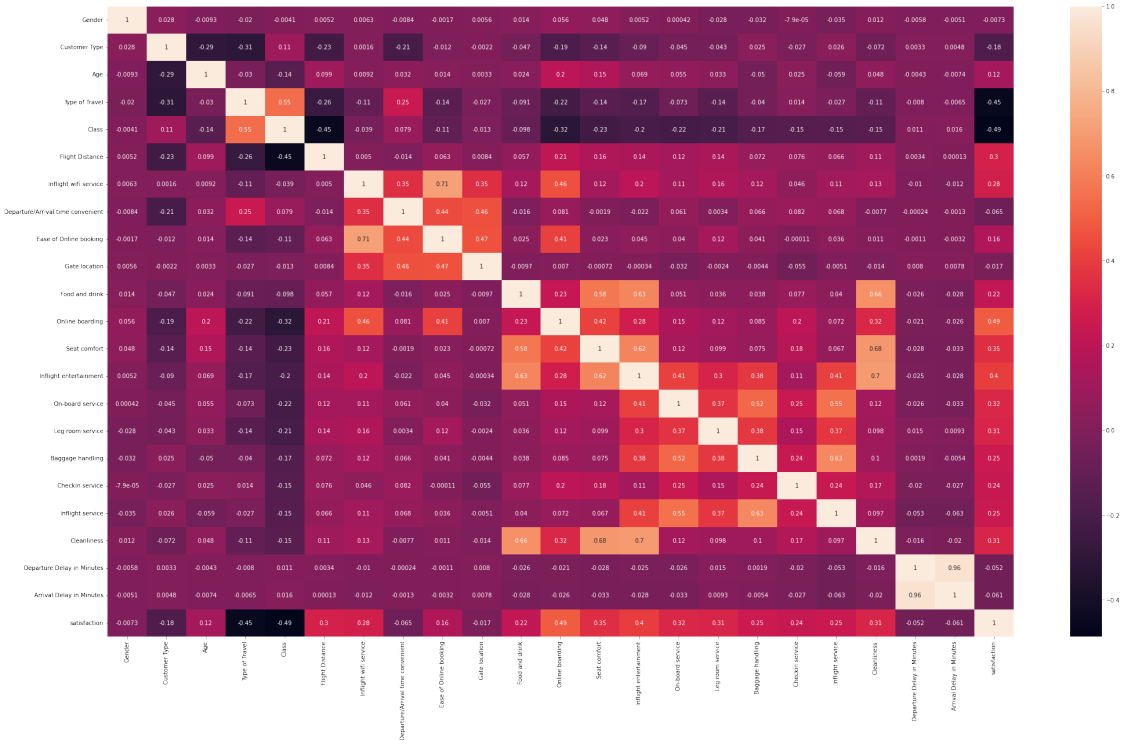
[120]:
```
df['satisfaction'].value_counts(normalize=True)
```

[120]:
```
0    0.561018
1    0.438982
Name: satisfaction, dtype: float64
```

[121]:
```
corr = df.corr() # Here we check the Co-relation between features using Heatmap␣
↪and Pairplot
```

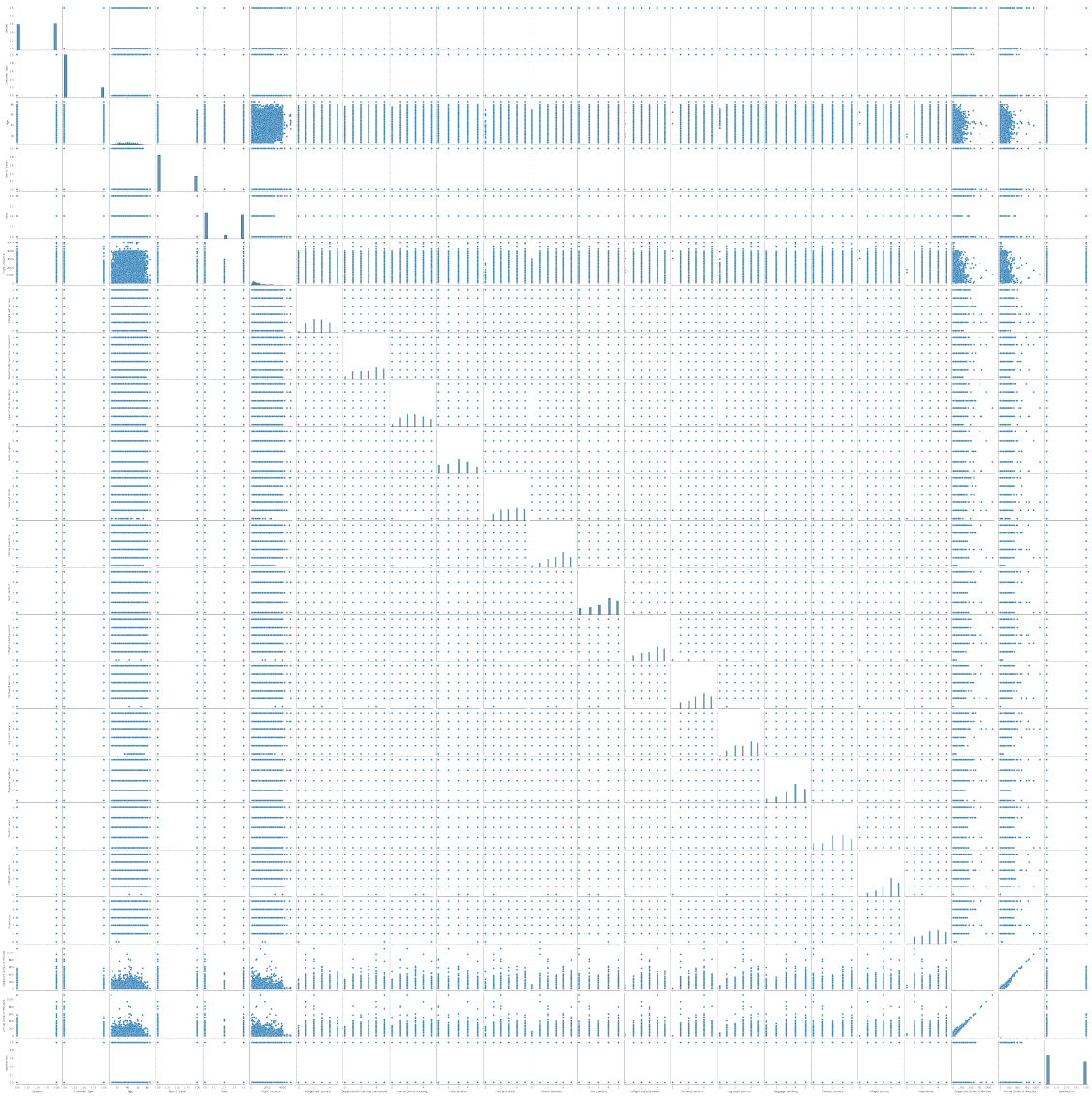
[122]:
```
plt.figure(figsize=(35,20))
sns.heatmap(corr,annot=True)
```

[122]:
```
<AxesSubplot:>
```



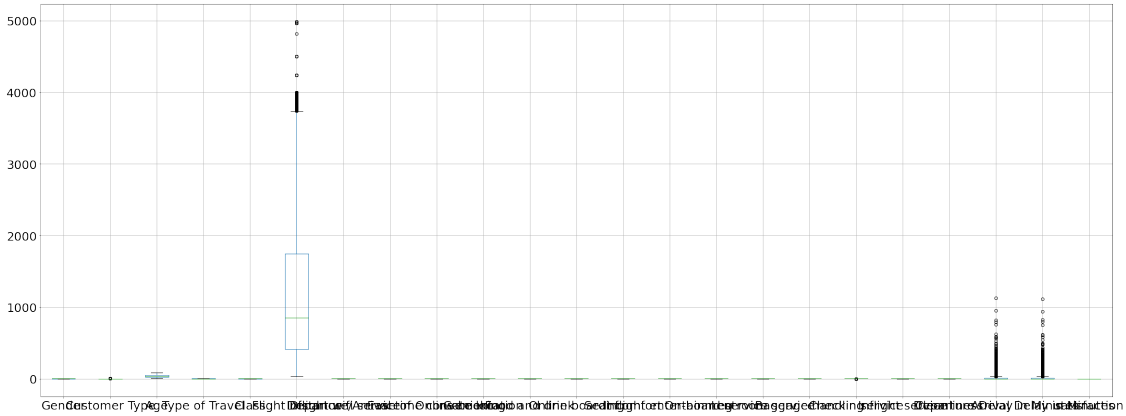
[123]:
```
sns.pairplot(df)
```

[123]:
```
<seaborn.axisgrid.PairGrid at 0x242a66f68b0>
```



[124]:
```
df.boxplot(figsize='40,15',fontsize=25,) # Overall we check outlires
```

[124]: <AxesSubplot:>



[125]:
```
df= df.replace(np.nan,0)
```

[126]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 25976 entries, 0 to 25975
Data columns (total 23 columns):
 #   Column                             Non-Null Count  Dtype
---  ------                             --------------  -----
 0   Gender                             25976 non-null  int64
 1   Customer Type                      25976 non-null  float64
 2   Age                                25976 non-null  int64
 3   Type of Travel                     25976 non-null  int64
 4   Class                              25976 non-null  int64
 5   Flight Distance                    25976 non-null  int64
 6   Inflight wifi service              25976 non-null  int64
 7   Departure/Arrival time convenient  25976 non-null  int64
 8   Ease of Online booking             25976 non-null  int64
 9   Gate location                      25976 non-null  int64
 10  Food and drink                     25976 non-null  int64
 11  Online boarding                    25976 non-null  int64
 12  Seat comfort                       25976 non-null  int64
 13  Inflight entertainment             25976 non-null  int64
 14  On-board service                   25976 non-null  int64
 15  Leg room service                   25976 non-null  int64
 16  Baggage handling                   25976 non-null  int64
 17  Checkin service                    25976 non-null  int64
 18  Inflight service                   25976 non-null  int64
 19  Cleanliness                        25976 non-null  int64
 20  Departure Delay in Minutes         25976 non-null  int64
 21  Arrival Delay in Minutes           25976 non-null  float64
 22  satisfaction                       25976 non-null  int64
dtypes: float64(2), int64(21)
```

```
memory usage: 4.6 MB
```

[127]:
```
df.describe(include='all')
```

[127]:

	Gender	Customer Type	Age	Type of Travel	\
count	25976.000000	25976.000000	25976.000000	25976.000000	
mean	1.507083	1.184247	39.620958	1.305590	
std	0.499959	0.388685	15.135685	0.460666	
min	1.000000	0.000000	7.000000	1.000000	
25%	1.000000	1.000000	27.000000	1.000000	
50%	2.000000	1.000000	40.000000	1.000000	
75%	2.000000	1.000000	51.000000	2.000000	
max	2.000000	2.000000	85.000000	2.000000	

	Class	Flight Distance	Inflight wifi service	\
count	25976.000000	25976.000000	25976.000000	
mean	1.964390	1193.788459	2.724746	
std	0.961673	998.683999	1.335384	
min	1.000000	31.000000	0.000000	
25%	1.000000	414.000000	2.000000	
50%	2.000000	849.000000	3.000000	
75%	3.000000	1744.000000	4.000000	
max	3.000000	4983.000000	5.000000	

	Departure/Arrival time convenient	Ease of Online booking	\
count	25976.000000	25976.000000	
mean	3.046812	2.756775	
std	1.533371	1.412951	
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	…	Inflight entertainment	On-board service	\
count	25976.000000	…	25976.000000	25976.000000	
mean	2.977094	…	3.357753	3.385664	
std	1.282133	…	1.338299	1.282088	
min	1.000000	…	0.000000	0.000000	
25%	2.000000	…	2.000000	2.000000	
50%	3.000000	…	4.000000	4.000000	
75%	4.000000	…	4.000000	4.000000	
max	5.000000	…	5.000000	5.000000	

	Leg room service	Baggage handling	Checkin service	Inflight service	\
count	25976.000000	25976.000000	25976.000000	25976.000000	
mean	3.350169	3.633238	3.314175	3.649253	

```
std           1.318862          1.176525          1.269332          1.180681
min           0.000000          1.000000          1.000000          0.000000
25%           2.000000          3.000000          3.000000          3.000000
50%           4.000000          4.000000          3.000000          4.000000
75%           4.000000          5.000000          4.000000          5.000000
max           5.000000          5.000000          5.000000          5.000000

        Cleanliness  Departure Delay in Minutes  Arrival Delay in Minutes  \
count  25976.000000                 25976.00000              25976.000000
mean       3.286226                    14.30609                 14.693756
std        1.319330                    37.42316                 37.466787
min        0.000000                     0.00000                  0.000000
25%        2.000000                     0.00000                  0.000000
50%        3.000000                     0.00000                  0.000000
75%        4.000000                    12.00000                 13.000000
max        5.000000                  1128.00000               1115.000000

       satisfaction
count  25976.000000
mean       0.438982
std        0.496272
min        0.000000
25%        0.000000
50%        0.000000
75%        1.000000
max        1.000000

[8 rows x 23 columns]
```

[128]:
```
df
```

[128]:
```
       Gender  Customer Type  Age  Type of Travel  Class  Flight Distance  \
0           2            1.0   52               1      3              160
1           2            1.0   36               1      1             2863
2           1            2.0   20               1      3              192
3           1            1.0   44               1      1             3377
4           2            1.0   49               1      3             1182
…       …             …  …              …    …              …
25971       1            2.0   34               1      1              526
25972       1            1.0   23               1      1              646
25973       2            1.0   17               2      3              828
25974       1            1.0   14               1      1             1127
25975       2            1.0   42               2      3              264

       Inflight wifi service  Departure/Arrival time convenient  \
0                          5                                 4
1                          1                                 1
```

```
2                          2                                  0
3                          0                                  0
4                          2                                  3
…                        …                                  …
25971                      3                                  3
25972                      4                                  4
25973                      2                                  5
25974                      3                                  3
25975                      2                                  5

       Ease of Online booking  Gate location  …  Inflight entertainment  \
0                           3              4  …                       5
1                           3              1  …                       4
2                           2              4  …                       2
3                           0              2  …                       1
4                           4              3  …                       2
…                         …              …  …                     …
25971                       3              1  …                       4
25972                       4              4  …                       4
25973                       1              5  …                       2
25974                       3              3  …                       4
25975                       2              5  …                       1

       On-board service  Leg room service  Baggage handling  Checkin service  \
0                     5                 5                 5                2
1                     4                 4                 4                3
2                     4                 1                 3                2
3                     1                 1                 1                3
4                     2                 2                 2                4
…                   …                 …                 …              …
25971                 3                 2                 4                4
25972                 4                 5                 5                5
25973                 4                 3                 4                5
25974                 3                 2                 5                4
25975                 1                 2                 1                1

       Inflight service  Cleanliness  Departure Delay in Minutes  \
0                     5            5                          50
1                     4            5                           0
2                     2            2                           0
3                     1            4                           0
4                     2            4                           0
…                   …            …                         …
25971                 5            4                           0
25972                 5            4                           0
25973                 4            2                           0
25974                 5            4                           0
```

```
25975                 1            1                           0

       Arrival Delay in Minutes  satisfaction
0                          44.0             1
1                           0.0             1
2                           0.0             0
3                           6.0             1
4                          20.0             1
…                           …             …
25971                       0.0             0
25972                       0.0             1
25973                       0.0             0
25974                       0.0             1
25975                       0.0             0

[25976 rows x 23 columns]
```

0.4 Train Test Split

[ ]:
```
# Due to training purpose we saperate independant varible and Target variable
```

[129]:
```
y = df['satisfaction']
x = df.drop('satisfaction',axis=1)
```

[130]:
```
x
```

[130]:
```
       Gender  Customer Type  Age  Type of Travel  Class  Flight Distance  \
0           2            1.0   52               1      3              160
1           2            1.0   36               1      1             2863
2           1            2.0   20               1      3              192
3           1            1.0   44               1      1             3377
4           2            1.0   49               1      3             1182
…         …              …   …                …    …               …
25971       1            2.0   34               1      1              526
25972       1            1.0   23               1      1              646
25973       2            1.0   17               2      3              828
25974       1            1.0   14               1      1             1127
25975       2            1.0   42               2      3              264

       Inflight wifi service  Departure/Arrival time convenient  \
0                          5                                 4
1                          1                                 1
2                          2                                 0
3                          0                                 0
4                          2                                 3
…                        …                                 …
25971                      3                                 3
```

```
25972                      4                                  4
25973                      2                                  5
25974                      3                                  3
25975                      2                                  5

       Ease of Online booking  Gate location  …  Seat comfort  \
0                           3              4  …             3
1                           3              1  …             5
2                           2              4  …             2
3                           0              2  …             4
4                           4              3  …             2
…                         …              …  …           …
25971                       3              1  …             4
25972                       4              4  …             4
25973                       1              5  …             2
25974                       3              3  …             4
25975                       2              5  …             2

       Inflight entertainment  On-board service  Leg room service  \
0                           5                 5                 5
1                           4                 4                 4
2                           2                 4                 1
3                           1                 1                 1
4                           2                 2                 2
…                         …                 …               …
25971                       4                 3                 2
25972                       4                 4                 5
25973                       2                 4                 3
25974                       4                 3                 2
25975                       1                 1                 2

       Baggage handling  Checkin service  Inflight service  Cleanliness  \
0                     5                2                 5            5
1                     4                3                 4            5
2                     3                2                 2            2
3                     1                3                 1            4
4                     2                4                 2            4
…                   …                …                 …          …
25971                 4                4                 5            4
25972                 5                5                 5            4
25973                 4                5                 4            2
25974                 5                4                 5            4
25975                 1                1                 1            1

       Departure Delay in Minutes  Arrival Delay in Minutes
0                              50                      44.0
1                               0                       0.0
```

```
2                                0                      0.0
3                                0                      6.0
4                                0                     20.0
…                              …                        …
25971                            0                      0.0
25972                            0                      0.0
25973                            0                      0.0
25974                            0                      0.0
25975                            0                      0.0

[25976 rows x 22 columns]
```

[131]:
```
y
```

[131]:
```
0        1
1        1
2        0
3        1
4        1
        ..
25971    0
25972    1
25973    0
25974    1
25975    0
Name: satisfaction, Length: 25976, dtype: int64
```

0.5 Model Building

[ ]:
```
# We can do apply first multiple algorithm on the basis of problem,
#from which algorithm give us max accuracy we can go with them.
# But Here lR algorithm is giving good accuracy on train and test both, so we␣
 ↪applying LR algo
```

[132]:
```
x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.
 ↪2,stratify=y,random_state=1)
```

[133]:
```
print(x_train.shape,x_test.shape)
```

```
(20780, 22) (5196, 22)
```

[134]:
```
lr_clf = LogisticRegression()
```

[ ]:
```

```

[135]:
```
lr_clf.fit(x_train,y_train)
```

[135]:
```
LogisticRegression()
```

[136]:
```
y_pred = lr_clf.predict(x_test)
```

[137]:
```
y_pred_train = lr_clf.predict(x_train)
```

[138]:
```
y_pred[:5]
```

[138]:
```
array([0, 0, 1, 0, 0], dtype=int64)
```

[139]:
```
y_test[:5]
```

[139]:
```
16693    1
9266     1
14233    1
14814    0
23002    0
Name: satisfaction, dtype: int64
```

0.6 Step 6 -> Model Evaluation

[ ]:
```
# On the basis of prob stat regression, Classification we use evaluation matrix.
# In Evaluation we can understand insights of accuracy
```

[140]:
```
residual = y_test - y_pred
residual
```

[140]:
```
16693    1
9266     1
14233    0
14814    0
23002    0
        ..
24409    0
19857    0
793      0
12028    0
22631    0
Name: satisfaction, Length: 5196, dtype: int64
```

[141]:
```
# Accuracy of Training Data
# We check here accuracy, Precision, Recall, F1-Score

cnf_matrix = confusion_matrix(y_train,y_pred_train)
print('Confusion Matrix is:\n',cnf_matrix)
clf_report = classification_report(y_train,y_pred_train)
print('Classification Report is:\n',clf_report)
```

```
accuracy = accuracy_score(y_train,y_pred_train)
print('Accuracy of Model is:\n',accuracy)
```

```
Confusion Matrix is:
 [[10001  1657]
 [ 1667  7455]]
Classification Report is:
              precision    recall  f1-score   support

           0       0.86      0.86      0.86     11658
           1       0.82      0.82      0.82      9122

    accuracy                           0.84     20780
   macro avg       0.84      0.84      0.84     20780
weighted avg       0.84      0.84      0.84     20780

Accuracy of Model is:
 0.8400384985563042
```

[142]:
```
# Accuracy of Testing Data

cnf_matrix = confusion_matrix(y_test,y_pred)
print('Confusion Matrix is:\n',cnf_matrix)
clf_report = classification_report(y_test,y_pred)
print('Classification Report is:\n',clf_report)
accuracy = accuracy_score(y_test,y_pred)
print('Accuracy of Model is:\n',accuracy)
```

```
Confusion Matrix is:
 [[2482  433]
 [ 383 1898]]
Classification Report is:
              precision    recall  f1-score   support

           0       0.87      0.85      0.86      2915
           1       0.81      0.83      0.82      2281

    accuracy                           0.84      5196
   macro avg       0.84      0.84      0.84      5196
weighted avg       0.84      0.84      0.84      5196

Accuracy of Model is:
 0.8429561200923787
```

[143]:
```
joblib.dump(lr_clf,"LR_model.model") # Here we Created model file using JOBLIB
```

[143]:
```
['LR_model.model']
```

[144]:
```
app = Flask(__name__) # Flask Access
api = Api(app)
```

[ ]:
```

```