Ola

June 18, 2025

[1]:
```
import pandas as pd
import numpy as nb
import seaborn as sb
import matplotlib.pyplot as plt
```

[2]:
```
df=pd.read_csv("Bengaluru Ola.csv")
```

[3]:
```
df
```

[3]:

	Date	Time	Booking ID	Booking Status	Customer ID	\
0	28/01/2024	6:00:00	CNR1721175	Success	329258	
1	26/01/2024	3:00:00	CNR2871422	Cancelled by Driver	201414	
2	15/01/2024	16:00:00	CNR6875935	Cancelled by Driver	301629	
3	02/01/2024	22:00:00	CNR6798834	Cancelled by Driver	319684	
4	30/01/2024	22:00:00	CNR9661713	Incomplete	330283	
…	…	…	…	…	…	
49994	02/01/2024	0:00:00	CNR7299340	Success	454202	
49995	27/01/2024	2:00:00	CNR0378586	Success	349873	
49996	17/01/2024	23:00:00	CNR2461856	Cancelled by Driver	880053	
49997	18/01/2024	22:00:00	CNR9524579	Success	752806	
49998	28/01/2024	14:00:00	CNR3698787	Success	159552	

	Vehicle Type	Pickup Location	Drop Location	Avg VTAT	Avg CTAT	…	\
0	Auto	Area-3	Area-2	5.42	18.46	…	
1	Mini	Area-7	Area-6	NaN	NaN	…	
2	Bike	Area-40	Area-24	NaN	NaN	…	
3	Prime Sedan	Area-11	Area-24	NaN	NaN	…	
4	Bike	Area-41	Area-45	NaN	NaN	…	
…	…	…	…	…	…	…	
49994	eBike	Area-39	Area-34	11.53	2.42	…	
49995	Auto	Area-17	Area-21	5.56	9.05	…	
49996	Auto	Area-37	Area-43	NaN	NaN	…	
49997	eBike	Area-48	Area-29	5.15	13.72	…	
49998	Bike	Area-1	Area-37	14.89	28.07	…	

	Reason for Cancelling by Customer	Cancelled Rides by Driver	\
0	NaN	0	
1	NaN	1	

```
2                                    NaN                          1
3                                    NaN                          1
4                                    NaN                          0
…                                    …                            …
49994                                NaN                          0
49995                                NaN                          0
49996                                NaN                          1
49997                                NaN                          0
49998                                NaN                          0

           Reason for Cancelling by Driver Incomplete Rides  \
0                                      NaN                 0
1            The customer was coughing/sick                0
2             Personal & Car related issues                0
3            The customer was coughing/sick                0
4                                      NaN                 1
…                                      …                 …
49994                                  NaN                 0
49995                                  NaN                 0
49996  More than permitted people in there                 0
49997                                  NaN                 0
49998                                  NaN                 0

       Incomplete Rides Reason Booking Value  Payment Method Ride Distance  \
0                          NaN        868.06          Wallet         28.50
1                          NaN           NaN             NaN           NaN
2                          NaN           NaN             NaN           NaN
3                          NaN           NaN             NaN           NaN
4                  Other Issue           NaN             NaN           NaN
…                          …             …               …             …
49994                      NaN       1217.43            Card         10.52
49995                      NaN       1369.51             UPI         11.04
49996                      NaN           NaN             NaN           NaN
49997                      NaN       1930.49             UPI         14.49
49998                      NaN       1534.66            Cash          4.60

       Driver Ratings  Customer Rating
0                 4.4              4.4
1                 NaN              NaN
2                 NaN              NaN
3                 NaN              NaN
4                 NaN              NaN
…                 …                …
49994             4.1              3.2
49995             5.0              4.3
49996             NaN              NaN
49997             3.1              4.5
```

```
49998            3.3             4.1

[49999 rows x 21 columns]
```

[4]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 49999 entries, 0 to 49998
Data columns (total 21 columns):
 #   Column                              Non-Null Count  Dtype
---  ------                              --------------  -----
 0   Date                                49999 non-null  object
 1   Time                                49999 non-null  object
 2   Booking ID                          49999 non-null  object
 3   Booking Status                      49999 non-null  object
 4   Customer ID                         49999 non-null  int64
 5   Vehicle Type                        49999 non-null  object
 6   Pickup Location                     49999 non-null  object
 7   Drop Location                       49999 non-null  object
 8   Avg VTAT                            33484 non-null  float64
 9   Avg CTAT                            33484 non-null  float64
 10  Cancelled  by Customer              49999 non-null  int64
 11  Reason for Cancelling by Customer   3799 non-null   object
 12  Cancelled Rides by Driver           49999 non-null  int64
 13  Reason for Cancelling by Driver     9610 non-null   object
 14  Incomplete Rides                    49999 non-null  int64
 15  Incomplete Rides Reason             3106 non-null   object
 16  Booking Value                       33484 non-null  float64
 17  Payment Method                      33484 non-null  object
 18  Ride Distance                       33484 non-null  float64
 19  Driver Ratings                      33484 non-null  float64
 20  Customer Rating                     33484 non-null  float64
dtypes: float64(6), int64(4), object(11)
memory usage: 8.0+ MB
```

[5]:
```
df.shape
```

[5]:
```
(49999, 21)
```

[6]:
```
df.head()
```

[6]:

	Date	Time	Booking ID	Booking Status	Customer ID	\
0	28/01/2024	6:00:00	CNR1721175	Success	329258	
1	26/01/2024	3:00:00	CNR2871422	Cancelled by Driver	201414	
2	15/01/2024	16:00:00	CNR6875935	Cancelled by Driver	301629	
3	02/01/2024	22:00:00	CNR6798834	Cancelled by Driver	319684	
4	30/01/2024	22:00:00	CNR9661713	Incomplete	330283	

```
  Vehicle Type Pickup Location Drop Location  Avg VTAT  Avg CTAT  …  \
0         Auto          Area-3        Area-2      5.42     18.46  …
1         Mini          Area-7        Area-6       NaN       NaN  …
2         Bike         Area-40       Area-24       NaN       NaN  …
3  Prime Sedan         Area-11       Area-24       NaN       NaN  …
4         Bike         Area-41       Area-45       NaN       NaN  …

   Reason for Cancelling by Customer Cancelled Rides by Driver  \
0                                NaN                         0
1                                NaN                         1
2                                NaN                         1
3                                NaN                         1
4                                NaN                         0

   Reason for Cancelling by Driver Incomplete Rides  Incomplete Rides Reason  \
0                              NaN                0                      NaN
1   The customer was coughing/sick                0                      NaN
2    Personal & Car related issues                0                      NaN
3   The customer was coughing/sick                0                      NaN
4                              NaN                1              Other Issue

  Booking Value  Payment Method Ride Distance  Driver Ratings  Customer Rating
0        868.06          Wallet          28.5             4.4              4.4
1           NaN             NaN           NaN             NaN              NaN
2           NaN             NaN           NaN             NaN              NaN
3           NaN             NaN           NaN             NaN              NaN
4           NaN             NaN           NaN             NaN              NaN

[5 rows x 21 columns]
```

[7]:
```
df.tail()
```

[7]:
```
             Date      Time  Booking ID       Booking Status  Customer ID  \
49994  02/01/2024   0:00:00  CNR7299340              Success       454202
49995  27/01/2024   2:00:00  CNR0378586              Success       349873
49996  17/01/2024  23:00:00  CNR2461856  Cancelled by Driver       880053
49997  18/01/2024  22:00:00  CNR9524579              Success       752806
49998  28/01/2024  14:00:00  CNR3698787              Success       159552

      Vehicle Type Pickup Location Drop Location  Avg VTAT  Avg CTAT  …  \
49994        eBike         Area-39       Area-34     11.53      2.42  …
49995         Auto         Area-17       Area-21      5.56      9.05  …
49996         Auto         Area-37       Area-43       NaN       NaN  …
49997        eBike         Area-48       Area-29      5.15     13.72  …
49998         Bike          Area-1       Area-37     14.89     28.07  …

       Reason for Cancelling by Customer Cancelled Rides by Driver  \
```

```
49994                                NaN                          0
49995                                NaN                          0
49996                                NaN                          1
49997                                NaN                          0
49998                                NaN                          0

          Reason for Cancelling by Driver Incomplete Rides  \
49994                                NaN                0
49995                                NaN                0
49996  More than permitted people in there                0
49997                                NaN                0
49998                                NaN                0

       Incomplete Rides Reason Booking Value  Payment Method Ride Distance  \
49994                      NaN       1217.43            Card         10.52
49995                      NaN       1369.51             UPI         11.04
49996                      NaN           NaN             NaN           NaN
49997                      NaN       1930.49             UPI         14.49
49998                      NaN       1534.66            Cash          4.60

       Driver Ratings  Customer Rating
49994             4.1              3.2
49995             5.0              4.3
49996             NaN              NaN
49997             3.1              4.5
49998             3.3              4.1

[5 rows x 21 columns]
```

[8]:
```
df.describe()
```

[8]:
```
         Customer ID      Avg VTAT      Avg CTAT  Cancelled  by Customer  \
count   49999.000000  33484.000000  33484.000000            49999.000000
mean   550979.937539     10.481654     15.564320                0.075982
std    258855.930945      5.490853      8.365313                0.264971
min    100001.000000      1.000000      1.000000                0.000000
25%    328740.500000      5.730000      8.310000                0.000000
50%    552598.000000     10.460000     15.610000                0.000000
75%    773970.000000     15.240000     22.790000                0.000000
max    999991.000000     20.000000     30.000000                1.000000

       Cancelled Rides by Driver  Incomplete Rides  Booking Value  \
count               49999.000000      49999.000000   33484.000000
mean                    0.192204          0.062121    1023.374286
std                     0.394036          0.241378     563.595072
min                     0.000000          0.000000      50.100000
25%                     0.000000          0.000000     532.942500
```

```
50%                    0.000000          0.000000    1023.350000
75%                    0.000000          0.000000    1514.677500
max                    1.000000          1.000000    2000.000000

       Ride Distance  Driver Ratings  Customer Rating
count   33484.000000     33484.00000     33484.000000
mean       25.447143         4.00304         4.001726
std        14.175953         0.57968         0.579177
min         1.000000         3.00000         3.000000
25%        13.100000         3.50000         3.500000
50%        25.460000         4.00000         4.000000
75%        37.670000         4.50000         4.500000
max        50.000000         5.00000         5.000000
```

[9]:
```
df.isnull().sum()
```

[9]:
```
Date                                    0
Time                                    0
Booking ID                              0
Booking Status                          0
Customer ID                             0
Vehicle Type                            0
Pickup Location                         0
Drop Location                           0
Avg VTAT                            16515
Avg CTAT                            16515
Cancelled  by Customer                  0
Reason for Cancelling by Customer   46200
Cancelled Rides by Driver               0
Reason for Cancelling by Driver     40389
Incomplete Rides                        0
Incomplete Rides Reason             46893
Booking Value                       16515
Payment Method                      16515
Ride Distance                       16515
Driver Ratings                      16515
Customer Rating                     16515
dtype: int64
```

[10]:
```
df['Avg VTAT']=df['Avg VTAT'].fillna(0)
```

[11]:
```
df.isnull().sum()
```

[11]:
```
Date                    0
Time                    0
Booking ID              0
Booking Status          0
```

```
Customer ID                            0
Vehicle Type                           0
Pickup Location                        0
Drop Location                          0
Avg VTAT                               0
Avg CTAT                           16515
Cancelled  by Customer                 0
Reason for Cancelling by Customer  46200
Cancelled Rides by Driver              0
Reason for Cancelling by Driver    40389
Incomplete Rides                       0
Incomplete Rides Reason            46893
Booking Value                      16515
Payment Method                     16515
Ride Distance                      16515
Driver Ratings                     16515
Customer Rating                    16515
dtype: int64
```

[12]:
```
df['Avg CTAT'].fillna(1,inplace=True)
```

```
C:\Users\admin\AppData\Local\Temp\ipykernel_13040\1426927941.py:1:
FutureWarning: A value is trying to be set on a copy of a DataFrame or Series
through chained assignment using an inplace method.
The behavior will change in pandas 3.0. This inplace method will never work
because the intermediate object on which we are setting values always behaves as
a copy.

For example, when doing 'df[col].method(value, inplace=True)', try using
'df.method({col: value}, inplace=True)' or df[col] = df[col].method(value)
instead, to perform the operation inplace on the original object.


  df['Avg CTAT'].fillna(1,inplace=True)
```

[13]:
```
df.isnull().sum()
```

[13]:
```
Date                               0
Time                               0
Booking ID                         0
Booking Status                     0
Customer ID                        0
Vehicle Type                       0
Pickup Location                    0
Drop Location                      0
Avg VTAT                           0
Avg CTAT                           0
Cancelled  by Customer             0
```

```
Reason for Cancelling by Customer    46200
Cancelled Rides by Driver                0
Reason for Cancelling by Driver      40389
Incomplete Rides                         0
Incomplete Rides Reason              46893
Booking Value                        16515
Payment Method                       16515
Ride Distance                        16515
Driver Ratings                       16515
Customer Rating                      16515
dtype: int64
```

[14]:
```
df['Reason for Cancelling by Driver']=df['Reason for Cancelling by Driver'].
 ↪fillna('Health Issue')
```

[15]:
```
df.isnull().sum()
```

[15]:
```
Date                                 0
Time                                 0
Booking ID                           0
Booking Status                       0
Customer ID                          0
Vehicle Type                         0
Pickup Location                      0
Drop Location                        0
Avg VTAT                             0
Avg CTAT                             0
Cancelled  by Customer               0
Reason for Cancelling by Customer    46200
Cancelled Rides by Driver                0
Reason for Cancelling by Driver          0
Incomplete Rides                         0
Incomplete Rides Reason              46893
Booking Value                        16515
Payment Method                       16515
Ride Distance                        16515
Driver Ratings                       16515
Customer Rating                      16515
dtype: int64
```

[16]:
```
df['Reason for Cancelling by Customer'].fillna('Traffic',inplace=True)
```

```
C:\Users\admin\AppData\Local\Temp\ipykernel_13040\252031365.py:1: FutureWarning:
A value is trying to be set on a copy of a DataFrame or Series through chained
assignment using an inplace method.
The behavior will change in pandas 3.0. This inplace method will never work
because the intermediate object on which we are setting values always behaves as
a copy.
```

```
For example, when doing 'df[col].method(value, inplace=True)', try using
'df.method({col: value}, inplace=True)' or df[col] = df[col].method(value)
instead, to perform the operation inplace on the original object.
```

```
df['Reason for Cancelling by Customer'].fillna('Traffic',inplace=True)
```

[17]:
```
df.isnull().sum()
```

[17]:
```
Date                                    0
Time                                    0
Booking ID                              0
Booking Status                          0
Customer ID                             0
Vehicle Type                            0
Pickup Location                         0
Drop Location                           0
Avg VTAT                                0
Avg CTAT                                0
Cancelled  by Customer                  0
Reason for Cancelling by Customer       0
Cancelled Rides by Driver               0
Reason for Cancelling by Driver         0
Incomplete Rides                        0
Incomplete Rides Reason             46893
Booking Value                       16515
Payment Method                      16515
Ride Distance                       16515
Driver Ratings                      16515
Customer Rating                     16515
dtype: int64
```

[18]:
```
df
```

[18]:

	Date	Time	Booking ID	Booking Status	Customer ID	\
0	28/01/2024	6:00:00	CNR1721175	Success	329258	
1	26/01/2024	3:00:00	CNR2871422	Cancelled by Driver	201414	
2	15/01/2024	16:00:00	CNR6875935	Cancelled by Driver	301629	
3	02/01/2024	22:00:00	CNR6798834	Cancelled by Driver	319684	
4	30/01/2024	22:00:00	CNR9661713	Incomplete	330283	
…	…	…	…	…	…	
49994	02/01/2024	0:00:00	CNR7299340	Success	454202	
49995	27/01/2024	2:00:00	CNR0378586	Success	349873	
49996	17/01/2024	23:00:00	CNR2461856	Cancelled by Driver	880053	
49997	18/01/2024	22:00:00	CNR9524579	Success	752806	
49998	28/01/2024	14:00:00	CNR3698787	Success	159552	

```
      Vehicle Type Pickup Location Drop Location  Avg VTAT  Avg CTAT  …  \
0             Auto          Area-3        Area-2      5.42     18.46  …
1             Mini          Area-7        Area-6      0.00      1.00  …
2             Bike         Area-40       Area-24      0.00      1.00  …
3      Prime Sedan         Area-11       Area-24      0.00      1.00  …
4             Bike         Area-41       Area-45      0.00      1.00  …
…              …              …            …         …       …  …
49994        eBike         Area-39       Area-34     11.53      2.42  …
49995         Auto         Area-17       Area-21      5.56      9.05  …
49996         Auto         Area-37       Area-43      0.00      1.00  …
49997        eBike         Area-48       Area-29      5.15     13.72  …
49998         Bike          Area-1       Area-37     14.89     28.07  …

       Reason for Cancelling by Customer Cancelled Rides by Driver  \
0                               Traffic                          0
1                               Traffic                          1
2                               Traffic                          1
3                               Traffic                          1
4                               Traffic                          0
…                                 …                            …
49994                           Traffic                          0
49995                           Traffic                          0
49996                           Traffic                          1
49997                           Traffic                          0
49998                           Traffic                          0

           Reason for Cancelling by Driver Incomplete Rides  \
0                             Health Issue                0
1           The customer was coughing/sick                0
2            Personal & Car related issues                0
3           The customer was coughing/sick                0
4                             Health Issue                1
…                                    …                  …
49994                         Health Issue                0
49995                         Health Issue                0
49996  More than permitted people in there                0
49997                         Health Issue                0
49998                         Health Issue                0

       Incomplete Rides Reason Booking Value  Payment Method Ride Distance  \
0                         NaN        868.06          Wallet         28.50
1                         NaN           NaN             NaN           NaN
2                         NaN           NaN             NaN           NaN
3                         NaN           NaN             NaN           NaN
4                Other Issue            NaN             NaN           NaN
…                        …              …               …             …
49994                     NaN       1217.43            Card         10.52
```

```
49995                      NaN       1369.51             UPI         11.04
49996                      NaN           NaN             NaN          NaN
49997                      NaN       1930.49             UPI         14.49
49998                      NaN       1534.66            Cash          4.60

       Driver Ratings  Customer Rating
0                 4.4              4.4
1                 NaN              NaN
2                 NaN              NaN
3                 NaN              NaN
4                 NaN              NaN
…                   …                …
49994             4.1              3.2
49995             5.0              4.3
49996             NaN              NaN
49997             3.1              4.5
49998             3.3              4.1

[49999 rows x 21 columns]
```

[19]:
```
df.isnull().sum()
```

[19]:
```
Date                                    0
Time                                    0
Booking ID                              0
Booking Status                          0
Customer ID                             0
Vehicle Type                            0
Pickup Location                         0
Drop Location                           0
Avg VTAT                                0
Avg CTAT                                0
Cancelled  by Customer                  0
Reason for Cancelling by Customer       0
Cancelled Rides by Driver               0
Reason for Cancelling by Driver         0
Incomplete Rides                        0
Incomplete Rides Reason             46893
Booking Value                       16515
Payment Method                      16515
Ride Distance                       16515
Driver Ratings                      16515
Customer Rating                     16515
dtype: int64
```

[20]:
```
df['Incomplete Rides Reason'].fillna('Critical Situation',inplace=True)
```

```
C:\Users\admin\AppData\Local\Temp\ipykernel_13040\578893116.py:1: FutureWarning:
```

```
A value is trying to be set on a copy of a DataFrame or Series through chained
assignment using an inplace method.
The behavior will change in pandas 3.0. This inplace method will never work
because the intermediate object on which we are setting values always behaves as
a copy.

For example, when doing 'df[col].method(value, inplace=True)', try using
'df.method({col: value}, inplace=True)' or df[col] = df[col].method(value)
instead, to perform the operation inplace on the original object.


  df['Incomplete Rides Reason'].fillna('Critical Situation',inplace=True)
```

[21]:
```
df.isnull().sum()
```

[21]:
```
Date                                    0
Time                                    0
Booking ID                              0
Booking Status                          0
Customer ID                             0
Vehicle Type                            0
Pickup Location                         0
Drop Location                           0
Avg VTAT                                0
Avg CTAT                                0
Cancelled  by Customer                  0
Reason for Cancelling by Customer       0
Cancelled Rides by Driver               0
Reason for Cancelling by Driver         0
Incomplete Rides                        0
Incomplete Rides Reason                 0
Booking Value                       16515
Payment Method                      16515
Ride Distance                       16515
Driver Ratings                      16515
Customer Rating                     16515
dtype: int64
```

[22]:
```
df['Booking Value'].fillna(1000.00,inplace=True)
```

```
C:\Users\admin\AppData\Local\Temp\ipykernel_13040\339474169.py:1: FutureWarning:
A value is trying to be set on a copy of a DataFrame or Series through chained
assignment using an inplace method.
The behavior will change in pandas 3.0. This inplace method will never work
because the intermediate object on which we are setting values always behaves as
a copy.

For example, when doing 'df[col].method(value, inplace=True)', try using
```

```
'df.method({col: value}, inplace=True)' or df[col] = df[col].method(value)
instead, to perform the operation inplace on the original object.


  df['Booking Value'].fillna(1000.00,inplace=True)
```

[23]:
```
df['Payment Method'].fillna('online',inplace=True)
```

```
C:\Users\admin\AppData\Local\Temp\ipykernel_13040\1307734987.py:1:
FutureWarning: A value is trying to be set on a copy of a DataFrame or Series
through chained assignment using an inplace method.
The behavior will change in pandas 3.0. This inplace method will never work
because the intermediate object on which we are setting values always behaves as
a copy.

For example, when doing 'df[col].method(value, inplace=True)', try using
'df.method({col: value}, inplace=True)' or df[col] = df[col].method(value)
instead, to perform the operation inplace on the original object.


  df['Payment Method'].fillna('online',inplace=True)
```

[24]:
```
df.isnull().sum()
```

[24]:
```
Date                                0
Time                                0
Booking ID                          0
Booking Status                      0
Customer ID                         0
Vehicle Type                        0
Pickup Location                     0
Drop Location                       0
Avg VTAT                            0
Avg CTAT                            0
Cancelled  by Customer              0
Reason for Cancelling by Customer   0
Cancelled Rides by Driver           0
Reason for Cancelling by Driver     0
Incomplete Rides                    0
Incomplete Rides Reason             0
Booking Value                       0
Payment Method                      0
Ride Distance                   16515
Driver Ratings                  16515
Customer Rating                 16515
dtype: int64
```

[ ]:

[25]:
```
df['Ride Distance'].fillna(30,inplace=True)
```

```
C:\Users\admin\AppData\Local\Temp\ipykernel_13040\3087966065.py:1:
FutureWarning: A value is trying to be set on a copy of a DataFrame or Series
through chained assignment using an inplace method.
The behavior will change in pandas 3.0. This inplace method will never work
because the intermediate object on which we are setting values always behaves as
a copy.

For example, when doing 'df[col].method(value, inplace=True)', try using
'df.method({col: value}, inplace=True)' or df[col] = df[col].method(value)
instead, to perform the operation inplace on the original object.


  df['Ride Distance'].fillna(30,inplace=True)
```

[26]:
```
df['Customer Rating']=df['Customer Rating'].fillna(4.0)
```

[27]:
```
df.isnull().sum()
```

[27]:
```
Date                                    0
Time                                    0
Booking ID                              0
Booking Status                          0
Customer ID                             0
Vehicle Type                            0
Pickup Location                         0
Drop Location                           0
Avg VTAT                                0
Avg CTAT                                0
Cancelled  by Customer                  0
Reason for Cancelling by Customer       0
Cancelled Rides by Driver               0
Reason for Cancelling by Driver         0
Incomplete Rides                        0
Incomplete Rides Reason                 0
Booking Value                           0
Payment Method                          0
Ride Distance                           0
Driver Ratings                      16515
Customer Rating                         0
dtype: int64
```

[28]:
```
df["Driver Ratings"]=df["Driver Ratings"].fillna(4.5)
```

[29]:
```
df
```

[29]:

	Date	Time	Booking ID	Booking Status	Customer ID	\
0	28/01/2024	6:00:00	CNR1721175	Success	329258	
1	26/01/2024	3:00:00	CNR2871422	Cancelled by Driver	201414	
2	15/01/2024	16:00:00	CNR6875935	Cancelled by Driver	301629	
3	02/01/2024	22:00:00	CNR6798834	Cancelled by Driver	319684	
4	30/01/2024	22:00:00	CNR9661713	Incomplete	330283	
…	…	…	…	…	…	
49994	02/01/2024	0:00:00	CNR7299340	Success	454202	
49995	27/01/2024	2:00:00	CNR0378586	Success	349873	
49996	17/01/2024	23:00:00	CNR2461856	Cancelled by Driver	880053	
49997	18/01/2024	22:00:00	CNR9524579	Success	752806	
49998	28/01/2024	14:00:00	CNR3698787	Success	159552	

	Vehicle Type	Pickup Location	Drop Location	Avg VTAT	Avg CTAT	…	\
0	Auto	Area-3	Area-2	5.42	18.46	…	
1	Mini	Area-7	Area-6	0.00	1.00	…	
2	Bike	Area-40	Area-24	0.00	1.00	…	
3	Prime Sedan	Area-11	Area-24	0.00	1.00	…	
4	Bike	Area-41	Area-45	0.00	1.00	…	
…	…	…	…	…	…	…	
49994	eBike	Area-39	Area-34	11.53	2.42	…	
49995	Auto	Area-17	Area-21	5.56	9.05	…	
49996	Auto	Area-37	Area-43	0.00	1.00	…	
49997	eBike	Area-48	Area-29	5.15	13.72	…	
49998	Bike	Area-1	Area-37	14.89	28.07	…	

	Reason for Cancelling by Customer	Cancelled Rides by Driver	\
0	Traffic	0	
1	Traffic	1	
2	Traffic	1	
3	Traffic	1	
4	Traffic	0	
…	…	…	
49994	Traffic	0	
49995	Traffic	0	
49996	Traffic	1	
49997	Traffic	0	
49998	Traffic	0	

	Reason for Cancelling by Driver	Incomplete Rides	\
0	Health Issue	0	
1	The customer was coughing/sick	0	
2	Personal & Car related issues	0	
3	The customer was coughing/sick	0	
4	Health Issue	1	
…	…	…	
49994	Health Issue	0	

```
49995                          Health Issue                0
49996  More than permitted people in there                0
49997                          Health Issue                0
49998                          Health Issue                0

       Incomplete Rides Reason Booking Value  Payment Method Ride Distance  \
0           Critical Situation        868.06          Wallet         28.50
1           Critical Situation       1000.00          online         30.00
2           Critical Situation       1000.00          online         30.00
3           Critical Situation       1000.00          online         30.00
4                 Other Issue       1000.00          online         30.00
…                          …             …             …             …
49994       Critical Situation       1217.43            Card         10.52
49995       Critical Situation       1369.51             UPI         11.04
49996       Critical Situation       1000.00          online         30.00
49997       Critical Situation       1930.49             UPI         14.49
49998       Critical Situation       1534.66            Cash          4.60

       Driver Ratings  Customer Rating
0                 4.4              4.4
1                 4.5              4.0
2                 4.5              4.0
3                 4.5              4.0
4                 4.5              4.0
…                 …                …
49994             4.1              3.2
49995             5.0              4.3
49996             4.5              4.0
49997             3.1              4.5
49998             3.3              4.1

[49999 rows x 21 columns]
```

[30]:
```
df.head(5)
```

[30]:
```
         Date      Time  Booking ID       Booking Status  Customer ID  \
0  28/01/2024   6:00:00  CNR1721175              Success       329258
1  26/01/2024   3:00:00  CNR2871422  Cancelled by Driver       201414
2  15/01/2024  16:00:00  CNR6875935  Cancelled by Driver       301629
3  02/01/2024  22:00:00  CNR6798834  Cancelled by Driver       319684
4  30/01/2024  22:00:00  CNR9661713           Incomplete       330283

  Vehicle Type Pickup Location Drop Location  Avg VTAT  Avg CTAT  …  \
0         Auto          Area-3        Area-2      5.42     18.46  …
1         Mini          Area-7        Area-6      0.00      1.00  …
2         Bike         Area-40       Area-24      0.00      1.00  …
3  Prime Sedan         Area-11       Area-24      0.00      1.00  …
```

```
4         Bike         Area-41       Area-45      0.00      1.00  …

   Reason for Cancelling by Customer Cancelled Rides by Driver  \
0                            Traffic                          0
1                            Traffic                          1
2                            Traffic                          1
3                            Traffic                          1
4                            Traffic                          0

   Reason for Cancelling by Driver Incomplete Rides  Incomplete Rides Reason  \
0                     Health Issue                0       Critical Situation
1   The customer was coughing/sick                0       Critical Situation
2    Personal & Car related issues                0       Critical Situation
3   The customer was coughing/sick                0       Critical Situation
4                     Health Issue                1              Other Issue

  Booking Value  Payment Method Ride Distance  Driver Ratings  Customer Rating
0        868.06          Wallet          28.5             4.4              4.4
1       1000.00          online          30.0             4.5              4.0
2       1000.00          online          30.0             4.5              4.0
3       1000.00          online          30.0             4.5              4.0
4       1000.00          online          30.0             4.5              4.0

[5 rows x 21 columns]
```

[31]:
```
df[df['Vehicle Type']=='Auto'].sum()
```

[31]:
```
Date
28/01/202409/01/202410/01/202422/01/202422/01/…
Time
6:00:002:00:0012:00:0022:00:0010:00:0012:00:00…
Booking ID
CNR1721175CNR2955705CNR0138926CNR8084056CNR931…
Booking Status
SuccessSuccessSuccessSuccessSuccessSuccessCanc…
Customer ID
3922891981
Vehicle Type
AutoAutoAutoAutoAutoAutoAutoAutoAutoAutoAu…
Pickup Location
Area-3Area-34Area-35Area-29Area-42Area-15Area-…
Drop Location
Area-2Area-11Area-22Area-49Area-10Area-48Area-…
Avg VTAT
49938.22
Avg CTAT
76346.03
```

```
Cancelled  by Customer
518
Reason for Cancelling by Customer
TrafficTrafficTrafficTrafficTrafficTrafficTraf…
Cancelled Rides by Driver
1338
Reason for Cancelling by Driver      Health IssueHealth IssueHealth IssueHealth
Iss…
Incomplete Rides
454
Incomplete Rides Reason              Critical SituationCritical
SituationCritical S…
Booking Value
7180612.58
Payment Method
WalletCardUPICardWalletCashonlineUPICardonline…
Ride Distance
192543.2
Driver Ratings
29574.0
Customer Rating
28419.4
dtype: object
```

[32]:
```
df[df['Vehicle Type']=='Bike'].value_counts()
```

[32]:
```
Date        Time     Booking ID  Booking Status         Customer ID  Vehicle
Type  Pickup Location  Drop Location  Avg VTAT  Avg CTAT  Cancelled  by Customer
Reason for Cancelling by Customer             Cancelled Rides by Driver  Reason
for Cancelling by Driver  Incomplete Rides  Incomplete Rides Reason  Booking
Value  Payment Method  Ride Distance  Driver Ratings  Customer Rating
31/01/2024  0:00:00  CNR9106501  Success                713692       Bike
Area-21          Area-15        10.76     16.80     0
Traffic                                    0                         Health
Issue                     0                Critical Situation       221.45
Cash            1.26           4.8             4.1                1
01/01/2024  0:00:00  CNR0182483  Cancelled by Customer  321326       Bike
Area-30          Area-26        0.00      1.00      1
Change of plans                            0                         Health
Issue                     0                Critical Situation       1000.00
online          30.00          4.5             4.0                1
                     CNR0736513  Success                944975       Bike
Area-27          Area-38        19.38     3.29      0
Traffic                                    0                         Health
Issue                     0                Critical Situation       438.73
Cash            41.45          4.5             4.9                1
                     CNR1428423  Success                403869       Bike
```

```
Area-11          Area-8         1.16      27.34     0
Traffic                                       0                          Health
Issue                     0                 Critical Situation        101.52
Wallet          1.19           3.9              4.5                  1
                     CNR2283026  Success                 296651       Bike
Area-4           Area-40        13.40     25.69     0
Traffic                                       0                          Health
Issue                     0                 Critical Situation        1967.61
Cash            35.52          4.0              4.4                  1
                                                                        ..
                     CNR8924669  Success                 280936       Bike
Area-12          Area-10        4.91      4.70      0
Traffic                                       0                          Health
Issue                     0                 Critical Situation        1230.87
Card            44.92          3.4              3.7                  1
                     CNR7676214  Success                 664766       Bike
Area-25          Area-30        8.47      24.40     0
Traffic                                       0                          Health
Issue                     0                 Critical Situation        1344.06
Card            21.59          4.6              4.8                  1
                     CNR6897816  Success                 480196       Bike
Area-22          Area-5         3.65      10.58     0
Traffic                                       0                          Health
Issue                     0                 Critical Situation        826.79
Wallet          49.50          4.2              4.4                  1
                     CNR4718636  Cancelled by Customer  587007       Bike
Area-14          Area-6         0.00      1.00      1
Driver is not moving towards pickup location  0                          Health
Issue                     0                 Critical Situation        1000.00
online          30.00          4.5              4.0                  1
                     CNR4620663  Success                 466768       Bike
Area-50          Area-5         8.83      5.90      0
Traffic                                       0                          Health
Issue                     0                 Critical Situation        1013.42
Wallet          23.63          5.0              4.2                  1
Name: count, Length: 7223, dtype: int64
```

[33]:
```
(df['Vehicle Type']=='Auto').sum()
```

[33]:
```
np.int64(7098)
```

[34]:
```
(df['Vehicle Type']=='Mini').sum()
```

[34]:
```
np.int64(7010)
```

[35]:
```
(df['Vehicle Type']=='Bike').sum()
```

[35]:
```
np.int64(7223)
```

[36]:
```
(df['Vehicle Type']=='Prime Sedan').sum()
```

[36]:
```
np.int64(7179)
```

[37]:
```
(df['Vehicle Type']=='eBike').sum()
```

[37]:
```
np.int64(7097)
```

[38]:
```
df['Vehicle Type'].nunique()
```

[38]:
```
7
```

[39]:
```
df.duplicated()
```

[39]:
```
0        False
1        False
2        False
3        False
4        False
         …
49994    False
49995    False
49996    False
49997    False
49998    False
Length: 49999, dtype: bool
```

[40]:
```
df.drop_duplicates(inplace=True)
```

[41]:
```
df.head()
```

[41]:

	Date	Time	Booking ID	Booking Status	Customer ID	\
0	28/01/2024	6:00:00	CNR1721175	Success	329258	
1	26/01/2024	3:00:00	CNR2871422	Cancelled by Driver	201414	
2	15/01/2024	16:00:00	CNR6875935	Cancelled by Driver	301629	
3	02/01/2024	22:00:00	CNR6798834	Cancelled by Driver	319684	
4	30/01/2024	22:00:00	CNR9661713	Incomplete	330283	

	Vehicle Type	Pickup Location	Drop Location	Avg VTAT	Avg CTAT	…	\
0	Auto	Area-3	Area-2	5.42	18.46	…	
1	Mini	Area-7	Area-6	0.00	1.00	…	
2	Bike	Area-40	Area-24	0.00	1.00	…	
3	Prime Sedan	Area-11	Area-24	0.00	1.00	…	
4	Bike	Area-41	Area-45	0.00	1.00	…	

	Reason for Cancelling by Customer	Cancelled Rides by Driver	\

```
0                              Traffic                          0
1                              Traffic                          1
2                              Traffic                          1
3                              Traffic                          1
4                              Traffic                          0

   Reason for Cancelling by Driver Incomplete Rides  Incomplete Rides Reason  \
0                      Health Issue                0       Critical Situation
1   The customer was coughing/sick                 0       Critical Situation
2    Personal & Car related issues                 0       Critical Situation
3   The customer was coughing/sick                 0       Critical Situation
4                      Health Issue                1              Other Issue

  Booking Value  Payment Method Ride Distance  Driver Ratings  Customer Rating
0        868.06          Wallet          28.5             4.4              4.4
1       1000.00          online          30.0             4.5              4.0
2       1000.00          online          30.0             4.5              4.0
3       1000.00          online          30.0             4.5              4.0
4       1000.00          online          30.0             4.5              4.0

[5 rows x 21 columns]
```

[42]:
```
df.rename(columns={'Booking Status' : 'booking_status'},inplace=True)
```

[43]:
```
df.head()
```

[43]:
```
         Date      Time  Booking ID       booking_status  Customer ID  \
0  28/01/2024   6:00:00  CNR1721175              Success       329258
1  26/01/2024   3:00:00  CNR2871422  Cancelled by Driver       201414
2  15/01/2024  16:00:00  CNR6875935  Cancelled by Driver       301629
3  02/01/2024  22:00:00  CNR6798834  Cancelled by Driver       319684
4  30/01/2024  22:00:00  CNR9661713           Incomplete       330283

  Vehicle Type Pickup Location Drop Location  Avg VTAT  Avg CTAT  …  \
0         Auto          Area-3        Area-2      5.42     18.46  …
1         Mini          Area-7        Area-6      0.00      1.00  …
2         Bike         Area-40       Area-24      0.00      1.00  …
3  Prime Sedan         Area-11       Area-24      0.00      1.00  …
4         Bike         Area-41       Area-45      0.00      1.00  …

   Reason for Cancelling by Customer Cancelled Rides by Driver  \
0                              Traffic                          0
1                              Traffic                          1
2                              Traffic                          1
3                              Traffic                          1
4                              Traffic                          0
```

```
   Reason for Cancelling by Driver Incomplete Rides  Incomplete Rides Reason  \
0                     Health Issue                0       Critical Situation
1   The customer was coughing/sick                0       Critical Situation
2    Personal & Car related issues                0       Critical Situation
3   The customer was coughing/sick                0       Critical Situation
4                     Health Issue                1              Other Issue

  Booking Value  Payment Method Ride Distance  Driver Ratings  Customer Rating
0        868.06          Wallet          28.5             4.4              4.4
1       1000.00          online          30.0             4.5              4.0
2       1000.00          online          30.0             4.5              4.0
3       1000.00          online          30.0             4.5              4.0
4       1000.00          online          30.0             4.5              4.0

[5 rows x 21 columns]
```

[44]:
```
df.rename(columns={'Customer ID':'customer_id'},inplace=True)
```

[45]:
```
df.head()
```

[45]:
```
         Date      Time  Booking ID       booking_status  customer_id  \
0  28/01/2024   6:00:00  CNR1721175              Success       329258
1  26/01/2024   3:00:00  CNR2871422  Cancelled by Driver       201414
2  15/01/2024  16:00:00  CNR6875935  Cancelled by Driver       301629
3  02/01/2024  22:00:00  CNR6798834  Cancelled by Driver       319684
4  30/01/2024  22:00:00  CNR9661713           Incomplete       330283

  Vehicle Type Pickup Location Drop Location  Avg VTAT  Avg CTAT  …  \
0         Auto          Area-3        Area-2      5.42     18.46  …
1         Mini          Area-7        Area-6      0.00      1.00  …
2         Bike         Area-40       Area-24      0.00      1.00  …
3  Prime Sedan         Area-11       Area-24      0.00      1.00  …
4         Bike         Area-41       Area-45      0.00      1.00  …

   Reason for Cancelling by Customer Cancelled Rides by Driver  \
0                           Traffic                          0
1                           Traffic                          1
2                           Traffic                          1
3                           Traffic                          1
4                           Traffic                          0

   Reason for Cancelling by Driver Incomplete Rides  Incomplete Rides Reason  \
0                     Health Issue                0       Critical Situation
1   The customer was coughing/sick                0       Critical Situation
2    Personal & Car related issues                0       Critical Situation
3   The customer was coughing/sick                0       Critical Situation
4                     Health Issue                1              Other Issue
```

```
  Booking Value  Payment Method Ride Distance  Driver Ratings  Customer Rating
0        868.06          Wallet          28.5             4.4              4.4
1       1000.00          online          30.0             4.5              4.0
2       1000.00          online          30.0             4.5              4.0
3       1000.00          online          30.0             4.5              4.0
4       1000.00          online          30.0             4.5              4.0

[5 rows x 21 columns]
```

[46]:
```
df.rename(columns={'Vehicle Type':'vehicle_type'},inplace=True)
```

[47]:
```
df.head()
```

[47]:
```
         Date      Time  Booking ID       booking_status  customer_id  \
0  28/01/2024   6:00:00  CNR1721175              Success       329258
1  26/01/2024   3:00:00  CNR2871422  Cancelled by Driver       201414
2  15/01/2024  16:00:00  CNR6875935  Cancelled by Driver       301629
3  02/01/2024  22:00:00  CNR6798834  Cancelled by Driver       319684
4  30/01/2024  22:00:00  CNR9661713           Incomplete       330283

  vehicle_type Pickup Location Drop Location  Avg VTAT  Avg CTAT  …  \
0         Auto          Area-3        Area-2      5.42     18.46  …
1         Mini          Area-7        Area-6      0.00      1.00  …
2         Bike         Area-40       Area-24      0.00      1.00  …
3  Prime Sedan         Area-11       Area-24      0.00      1.00  …
4         Bike         Area-41       Area-45      0.00      1.00  …

   Reason for Cancelling by Customer Cancelled Rides by Driver  \
0                           Traffic                          0
1                           Traffic                          1
2                           Traffic                          1
3                           Traffic                          1
4                           Traffic                          0

   Reason for Cancelling by Driver Incomplete Rides  Incomplete Rides Reason  \
0                     Health Issue                0       Critical Situation
1   The customer was coughing/sick                0       Critical Situation
2    Personal & Car related issues                0       Critical Situation
3   The customer was coughing/sick                0       Critical Situation
4                     Health Issue                1              Other Issue

  Booking Value  Payment Method Ride Distance  Driver Ratings  Customer Rating
0        868.06          Wallet          28.5             4.4              4.4
1       1000.00          online          30.0             4.5              4.0
2       1000.00          online          30.0             4.5              4.0
3       1000.00          online          30.0             4.5              4.0
```

```
4       1000.00          online          30.0             4.5              4.0

[5 rows x 21 columns]
```

[48]:
```
df.rename(columns={'Pickup Location':'pickup_location'},inplace=True)
```

[49]:
```
df.head()
```

[49]:
```
         Date      Time  Booking ID       booking_status  customer_id  \
0  28/01/2024   6:00:00  CNR1721175              Success       329258
1  26/01/2024   3:00:00  CNR2871422  Cancelled by Driver       201414
2  15/01/2024  16:00:00  CNR6875935  Cancelled by Driver       301629
3  02/01/2024  22:00:00  CNR6798834  Cancelled by Driver       319684
4  30/01/2024  22:00:00  CNR9661713           Incomplete       330283

  vehicle_type pickup_location Drop Location  Avg VTAT  Avg CTAT  …  \
0         Auto          Area-3        Area-2      5.42     18.46  …
1         Mini          Area-7        Area-6      0.00      1.00  …
2         Bike         Area-40       Area-24      0.00      1.00  …
3  Prime Sedan         Area-11       Area-24      0.00      1.00  …
4         Bike         Area-41       Area-45      0.00      1.00  …

   Reason for Cancelling by Customer Cancelled Rides by Driver  \
0                           Traffic                          0
1                           Traffic                          1
2                           Traffic                          1
3                           Traffic                          1
4                           Traffic                          0

   Reason for Cancelling by Driver Incomplete Rides  Incomplete Rides Reason  \
0                     Health Issue                0       Critical Situation
1   The customer was coughing/sick                0       Critical Situation
2    Personal & Car related issues                0       Critical Situation
3   The customer was coughing/sick                0       Critical Situation
4                     Health Issue                1              Other Issue

  Booking Value  Payment Method Ride Distance  Driver Ratings  Customer Rating
0        868.06          Wallet          28.5             4.4              4.4
1       1000.00          online          30.0             4.5              4.0
2       1000.00          online          30.0             4.5              4.0
3       1000.00          online          30.0             4.5              4.0
4       1000.00          online          30.0             4.5              4.0

[5 rows x 21 columns]
```

[50]:
```
df.rename(columns={'Drop Location':'drop_location'},inplace=True)
```

[51]:
```
df.rename(columns={'Booking ID':'booking_id'},inplace=True)
```

[52]:
```
df.head()
```

[52]:

	Date	Time	booking_id	booking_status	customer_id	\
0	28/01/2024	6:00:00	CNR1721175	Success	329258	
1	26/01/2024	3:00:00	CNR2871422	Cancelled by Driver	201414	
2	15/01/2024	16:00:00	CNR6875935	Cancelled by Driver	301629	
3	02/01/2024	22:00:00	CNR6798834	Cancelled by Driver	319684	
4	30/01/2024	22:00:00	CNR9661713	Incomplete	330283	

	vehicle_type	pickup_location	drop_location	Avg VTAT	Avg CTAT	…	\
0	Auto	Area-3	Area-2	5.42	18.46	…	
1	Mini	Area-7	Area-6	0.00	1.00	…	
2	Bike	Area-40	Area-24	0.00	1.00	…	
3	Prime Sedan	Area-11	Area-24	0.00	1.00	…	
4	Bike	Area-41	Area-45	0.00	1.00	…	

	Reason for Cancelling by Customer	Cancelled Rides by Driver	\
0	Traffic	0	
1	Traffic	1	
2	Traffic	1	
3	Traffic	1	
4	Traffic	0	

	Reason for Cancelling by Driver	Incomplete Rides	Incomplete Rides Reason	\
0	Health Issue	0	Critical Situation	
1	The customer was coughing/sick	0	Critical Situation	
2	Personal & Car related issues	0	Critical Situation	
3	The customer was coughing/sick	0	Critical Situation	
4	Health Issue	1	Other Issue	

	Booking Value	Payment Method	Ride Distance	Driver Ratings	Customer Rating
0	868.06	Wallet	28.5	4.4	4.4
1	1000.00	online	30.0	4.5	4.0
2	1000.00	online	30.0	4.5	4.0
3	1000.00	online	30.0	4.5	4.0
4	1000.00	online	30.0	4.5	4.0

[5 rows x 21 columns]

[53]:
```
df[['Date','Time','booking_id','booking_status','customer_id','vehicle_type','pickup_location'
```

[53]:

	Date	Time	booking_id	booking_status	customer_id	\
0	28/01/2024	6:00:00	CNR1721175	Success	329258	
1	26/01/2024	3:00:00	CNR2871422	Cancelled by Driver	201414	
2	15/01/2024	16:00:00	CNR6875935	Cancelled by Driver	301629	

```
3      02/01/2024  22:00:00  CNR6798834  Cancelled by Driver       319684
4      30/01/2024  22:00:00  CNR9661713           Incomplete       330283
…           …        …          …                    …              …
49994  02/01/2024   0:00:00  CNR7299340              Success       454202
49995  27/01/2024   2:00:00  CNR0378586              Success       349873
49996  17/01/2024  23:00:00  CNR2461856  Cancelled by Driver       880053
49997  18/01/2024  22:00:00  CNR9524579              Success       752806
49998  28/01/2024  14:00:00  CNR3698787              Success       159552

      vehicle_type pickup_location drop_location
0             Auto          Area-3        Area-2
1             Mini          Area-7        Area-6
2             Bike         Area-40       Area-24
3      Prime Sedan         Area-11       Area-24
4             Bike         Area-41       Area-45
…              …              …            …
49994        eBike         Area-39       Area-34
49995         Auto         Area-17       Area-21
49996         Auto         Area-37       Area-43
49997        eBike         Area-48       Area-29
49998         Bike          Area-1       Area-37

[49999 rows x 8 columns]
```

[54]:
```
df[df['Time']=='22:00:00'].count()
```

[54]:
```
Date                                2087
Time                                2087
booking_id                          2087
booking_status                      2087
customer_id                         2087
vehicle_type                        2087
pickup_location                     2087
drop_location                       2087
Avg VTAT                            2087
Avg CTAT                            2087
Cancelled  by Customer              2087
Reason for Cancelling by Customer   2087
Cancelled Rides by Driver           2087
Reason for Cancelling by Driver     2087
Incomplete Rides                    2087
Incomplete Rides Reason             2087
Booking Value                       2087
Payment Method                      2087
Ride Distance                       2087
Driver Ratings                      2087
Customer Rating                     2087
```

```
dtype: int64
```

[55]:
```
df[df['Time']=='23:00:00'].count()
```

[55]:
```
Date                                2006
Time                                2006
booking_id                          2006
booking_status                      2006
customer_id                         2006
vehicle_type                        2006
pickup_location                     2006
drop_location                       2006
Avg VTAT                            2006
Avg CTAT                            2006
Cancelled  by Customer              2006
Reason for Cancelling by Customer   2006
Cancelled Rides by Driver           2006
Reason for Cancelling by Driver     2006
Incomplete Rides                    2006
Incomplete Rides Reason             2006
Booking Value                       2006
Payment Method                      2006
Ride Distance                       2006
Driver Ratings                      2006
Customer Rating                     2006
dtype: int64
```

[56]:
```
df.head()
```

[56]:

	Date	Time	booking_id	booking_status	customer_id	\
0	28/01/2024	6:00:00	CNR1721175	Success	329258	
1	26/01/2024	3:00:00	CNR2871422	Cancelled by Driver	201414	
2	15/01/2024	16:00:00	CNR6875935	Cancelled by Driver	301629	
3	02/01/2024	22:00:00	CNR6798834	Cancelled by Driver	319684	
4	30/01/2024	22:00:00	CNR9661713	Incomplete	330283	

	vehicle_type	pickup_location	drop_location	Avg VTAT	Avg CTAT	…	\
0	Auto	Area-3	Area-2	5.42	18.46	…	
1	Mini	Area-7	Area-6	0.00	1.00	…	
2	Bike	Area-40	Area-24	0.00	1.00	…	
3	Prime Sedan	Area-11	Area-24	0.00	1.00	…	
4	Bike	Area-41	Area-45	0.00	1.00	…	

	Reason for Cancelling by Customer	Cancelled Rides by Driver	\
0	Traffic	0	
1	Traffic	1	
2	Traffic	1	

```
3                              Traffic                          1
4                              Traffic                          0

   Reason for Cancelling by Driver Incomplete Rides  Incomplete Rides Reason  \
0                      Health Issue                0       Critical Situation
1   The customer was coughing/sick                 0       Critical Situation
2    Personal & Car related issues                 0       Critical Situation
3   The customer was coughing/sick                 0       Critical Situation
4                      Health Issue                1              Other Issue

  Booking Value  Payment Method Ride Distance  Driver Ratings  Customer Rating
0        868.06          Wallet          28.5             4.4              4.4
1       1000.00          online          30.0             4.5              4.0
2       1000.00          online          30.0             4.5              4.0
3       1000.00          online          30.0             4.5              4.0
4       1000.00          online          30.0             4.5              4.0

[5 rows x 21 columns]
```

[57]:
```
df[df['Date']=='28/01/2024'].count()
```

[57]:
```
Date                                1664
Time                                1664
booking_id                          1664
booking_status                      1664
customer_id                         1664
vehicle_type                        1664
pickup_location                     1664
drop_location                       1664
Avg VTAT                            1664
Avg CTAT                            1664
Cancelled  by Customer              1664
Reason for Cancelling by Customer   1664
Cancelled Rides by Driver           1664
Reason for Cancelling by Driver     1664
Incomplete Rides                    1664
Incomplete Rides Reason             1664
Booking Value                       1664
Payment Method                      1664
Ride Distance                       1664
Driver Ratings                      1664
Customer Rating                     1664
dtype: int64
```

[58]:
```
df[(df['Date']>='01/01/2024') & (df['Date']<='31/01/2024')].head(5)
```

[58]:
```
         Date      Time  booking_id       booking_status  customer_id  \
0  28/01/2024   6:00:00  CNR1721175              Success       329258
1  26/01/2024   3:00:00  CNR2871422  Cancelled by Driver       201414
2  15/01/2024  16:00:00  CNR6875935  Cancelled by Driver       301629
3  02/01/2024  22:00:00  CNR6798834  Cancelled by Driver       319684
4  30/01/2024  22:00:00  CNR9661713           Incomplete       330283

  vehicle_type pickup_location drop_location  Avg VTAT  Avg CTAT  …  \
0         Auto          Area-3        Area-2      5.42     18.46  …
1         Mini          Area-7        Area-6      0.00      1.00  …
2         Bike         Area-40       Area-24      0.00      1.00  …
3  Prime Sedan         Area-11       Area-24      0.00      1.00  …
4         Bike         Area-41       Area-45      0.00      1.00  …

   Reason for Cancelling by Customer Cancelled Rides by Driver  \
0                            Traffic                         0
1                            Traffic                         1
2                            Traffic                         1
3                            Traffic                         1
4                            Traffic                         0

   Reason for Cancelling by Driver Incomplete Rides  Incomplete Rides Reason  \
0                     Health Issue                0       Critical Situation
1   The customer was coughing/sick                0       Critical Situation
2    Personal & Car related issues                0       Critical Situation
3   The customer was coughing/sick                0       Critical Situation
4                     Health Issue                1              Other Issue

  Booking Value  Payment Method Ride Distance  Driver Ratings  Customer Rating
0        868.06          Wallet          28.5             4.4              4.4
1       1000.00          online          30.0             4.5              4.0
2       1000.00          online          30.0             4.5              4.0
3       1000.00          online          30.0             4.5              4.0
4       1000.00          online          30.0             4.5              4.0

[5 rows x 21 columns]
```

[59]:
```
df[(df['Time']>='16:00:00') & (df['Time']<='21:00:00')].sort_index().head()
```

[59]:
```
         Date      Time  booking_id         booking_status  customer_id  \
2  15/01/2024  16:00:00  CNR6875935    Cancelled by Driver       301629
5  14/01/2024  20:00:00  CNR8720322             Incomplete       787277
6  17/01/2024  21:00:00  CNR8611757                Success       329974
13 23/01/2024  21:00:00  CNR4422948                Success       976638
16 25/01/2024  16:00:00  CNR5645592  Cancelled by Customer       237235

   vehicle_type pickup_location drop_location  Avg VTAT  Avg CTAT  …  \
```

```
2          Bike         Area-40       Area-24      0.00      1.00  …
5    Prime Plus         Area-46       Area-21      0.00      1.00  …
6   Prime Sedan         Area-38       Area-26     19.95      5.01  …
13         Bike          Area-4       Area-42      1.28      3.70  …
16         Mini         Area-16       Area-17      0.00      1.00  …

    Reason for Cancelling by Customer Cancelled Rides by Driver  \
2                             Traffic                         1
5                             Traffic                         0
6                             Traffic                         0
13                            Traffic                         0
16                 AC is not working                          0

    Reason for Cancelling by Driver Incomplete Rides  Incomplete Rides Reason  \
2     Personal & Car related issues                0       Critical Situation
5                      Health Issue                1        Vehicle Breakdown
6                      Health Issue                0       Critical Situation
13                     Health Issue                0       Critical Situation
16                     Health Issue                0       Critical Situation

   Booking Value  Payment Method Ride Distance  Driver Ratings  \
2        1000.00          online         30.00             4.5
5        1000.00          online         30.00             4.5
6         348.04            Card         25.18             4.5
13        988.65          Wallet         37.92             3.5
16       1000.00          online         30.00             4.5

    Customer Rating
2               4.0
5               4.0
6               4.7
13              3.8
16              4.0

[5 rows x 21 columns]
```

[60]:
```
result=df[(df['Time']>='15:00:00') & (df['Time']<='18:00:00')].count()
print(result)
```

```
Date                                 8476
Time                                 8476
booking_id                           8476
booking_status                       8476
customer_id                          8476
vehicle_type                         8476
pickup_location                      8476
drop_location                        8476
Avg VTAT                             8476
```

```
Avg CTAT                              8476
Cancelled  by Customer                8476
Reason for Cancelling by Customer     8476
Cancelled Rides by Driver             8476
Reason for Cancelling by Driver       8476
Incomplete Rides                      8476
Incomplete Rides Reason               8476
Booking Value                         8476
Payment Method                        8476
Ride Distance                         8476
Driver Ratings                        8476
Customer Rating                       8476
dtype: int64
```

[61]:
```
df[df['Date'].between('01/01/2024','10/01/2024')].count()
```

[61]:
```
Date                                  16475
Time                                  16475
booking_id                            16475
booking_status                        16475
customer_id                           16475
vehicle_type                          16475
pickup_location                       16475
drop_location                         16475
Avg VTAT                              16475
Avg CTAT                              16475
Cancelled  by Customer                16475
Reason for Cancelling by Customer     16475
Cancelled Rides by Driver             16475
Reason for Cancelling by Driver       16475
Incomplete Rides                      16475
Incomplete Rides Reason               16475
Booking Value                         16475
Payment Method                        16475
Ride Distance                         16475
Driver Ratings                        16475
Customer Rating                       16475
dtype: int64
```

[62]:
```
df[df['Date'].between('11/01/2024','20/01/2024')].count()
```

[62]:
```
Date                                  16817
Time                                  16817
booking_id                            16817
booking_status                        16817
customer_id                           16817
vehicle_type                          16817
pickup_location                       16817
```

```
drop_location                        16817
Avg VTAT                             16817
Avg CTAT                             16817
Cancelled  by Customer               16817
Reason for Cancelling by Customer    16817
Cancelled Rides by Driver            16817
Reason for Cancelling by Driver      16817
Incomplete Rides                     16817
Incomplete Rides Reason              16817
Booking Value                        16817
Payment Method                       16817
Ride Distance                        16817
Driver Ratings                       16817
Customer Rating                      16817
dtype: int64
```

[63]:
```
df[df['Date'].between('21/01/2024','31/01/2024')].count()
```

[63]:
```
Date                                 16707
Time                                 16707
booking_id                           16707
booking_status                       16707
customer_id                          16707
vehicle_type                         16707
pickup_location                      16707
drop_location                        16707
Avg VTAT                             16707
Avg CTAT                             16707
Cancelled  by Customer               16707
Reason for Cancelling by Customer    16707
Cancelled Rides by Driver            16707
Reason for Cancelling by Driver      16707
Incomplete Rides                     16707
Incomplete Rides Reason              16707
Booking Value                        16707
Payment Method                       16707
Ride Distance                        16707
Driver Ratings                       16707
Customer Rating                      16707
dtype: int64
```

[64]:
```
df[df['Date'].between('01/01/2024','31/01/2024')].count()
```

[64]:
```
Date                                 49999
Time                                 49999
booking_id                           49999
booking_status                       49999
```

```
customer_id                          49999
vehicle_type                         49999
pickup_location                      49999
drop_location                        49999
Avg VTAT                             49999
Avg CTAT                             49999
Cancelled  by Customer               49999
Reason for Cancelling by Customer    49999
Cancelled Rides by Driver            49999
Reason for Cancelling by Driver      49999
Incomplete Rides                     49999
Incomplete Rides Reason              49999
Booking Value                        49999
Payment Method                       49999
Ride Distance                        49999
Driver Ratings                       49999
Customer Rating                      49999
dtype: int64
```

[65]:
```
df['DateTime']=pd.to_datetime(df['Date']+' '+df['Time'])
```

```
C:\Users\admin\AppData\Local\Temp\ipykernel_13040\1743183297.py:1: UserWarning:
Parsing dates in %d/%m/%Y %H:%M:%S format when dayfirst=False (the default) was
specified. Pass `dayfirst=True` or specify a format to silence this warning.
  df['DateTime']=pd.to_datetime(df['Date']+' '+df['Time'])
```

[66]:
```
df
```

[66]:

	Date	Time	booking_id	booking_status	customer_id	\
0	28/01/2024	6:00:00	CNR1721175	Success	329258	
1	26/01/2024	3:00:00	CNR2871422	Cancelled by Driver	201414	
2	15/01/2024	16:00:00	CNR6875935	Cancelled by Driver	301629	
3	02/01/2024	22:00:00	CNR6798834	Cancelled by Driver	319684	
4	30/01/2024	22:00:00	CNR9661713	Incomplete	330283	
…	…	…	…	…	…	
49994	02/01/2024	0:00:00	CNR7299340	Success	454202	
49995	27/01/2024	2:00:00	CNR0378586	Success	349873	
49996	17/01/2024	23:00:00	CNR2461856	Cancelled by Driver	880053	
49997	18/01/2024	22:00:00	CNR9524579	Success	752806	
49998	28/01/2024	14:00:00	CNR3698787	Success	159552	

	vehicle_type	pickup_location	drop_location	Avg VTAT	Avg CTAT	…	\
0	Auto	Area-3	Area-2	5.42	18.46	…	
1	Mini	Area-7	Area-6	0.00	1.00	…	
2	Bike	Area-40	Area-24	0.00	1.00	…	
3	Prime Sedan	Area-11	Area-24	0.00	1.00	…	
4	Bike	Area-41	Area-45	0.00	1.00	…	
…	…	…	…	…	…	…	…

```
49994        eBike         Area-39       Area-34     11.53      2.42  …
49995         Auto         Area-17       Area-21      5.56      9.05  …
49996         Auto         Area-37       Area-43      0.00      1.00  …
49997        eBike         Area-48       Area-29      5.15     13.72  …
49998         Bike          Area-1       Area-37     14.89     28.07  …

       Cancelled Rides by Driver      Reason for Cancelling by Driver  \
0                              0                          Health Issue
1                              1        The customer was coughing/sick
2                              1         Personal & Car related issues
3                              1        The customer was coughing/sick
4                              0                          Health Issue
…                            …                                    …
49994                          0                          Health Issue
49995                          0                          Health Issue
49996                          1  More than permitted people in there
49997                          0                          Health Issue
49998                          0                          Health Issue

       Incomplete Rides Incomplete Rides Reason  Booking Value Payment Method  \
0                     0      Critical Situation         868.06         Wallet
1                     0      Critical Situation        1000.00         online
2                     0      Critical Situation        1000.00         online
3                     0      Critical Situation        1000.00         online
4                     1             Other Issue        1000.00         online
…                   …                       …             …            …
49994                 0      Critical Situation        1217.43           Card
49995                 0      Critical Situation        1369.51            UPI
49996                 0      Critical Situation        1000.00         online
49997                 0      Critical Situation        1930.49            UPI
49998                 0      Critical Situation        1534.66           Cash

       Ride Distance Driver Ratings  Customer Rating            DateTime
0             28.50            4.4              4.4 2024-01-28 06:00:00
1             30.00            4.5              4.0 2024-01-26 03:00:00
2             30.00            4.5              4.0 2024-01-15 16:00:00
3             30.00            4.5              4.0 2024-01-02 22:00:00
4             30.00            4.5              4.0 2024-01-30 22:00:00
…             …              …                …                   …
49994         10.52            4.1              3.2 2024-01-02 00:00:00
49995         11.04            5.0              4.3 2024-01-27 02:00:00
49996         30.00            4.5              4.0 2024-01-17 23:00:00
49997         14.49            3.1              4.5 2024-01-18 22:00:00
49998          4.60            3.3              4.1 2024-01-28 14:00:00

[49999 rows x 22 columns]
```

[67]:
```
df[df['DateTime'].between('01/01/2024 12:00:00','31/01/2024 11:59:59')]
```

[67]:

	Date	Time	booking_id	booking_status	customer_id	\
0	28/01/2024	6:00:00	CNR1721175	Success	329258	
1	26/01/2024	3:00:00	CNR2871422	Cancelled by Driver	201414	
2	15/01/2024	16:00:00	CNR6875935	Cancelled by Driver	301629	
3	02/01/2024	22:00:00	CNR6798834	Cancelled by Driver	319684	
4	30/01/2024	22:00:00	CNR9661713	Incomplete	330283	
…	…	…	…	…	…	
49994	02/01/2024	0:00:00	CNR7299340	Success	454202	
49995	27/01/2024	2:00:00	CNR0378586	Success	349873	
49996	17/01/2024	23:00:00	CNR2461856	Cancelled by Driver	880053	
49997	18/01/2024	22:00:00	CNR9524579	Success	752806	
49998	28/01/2024	14:00:00	CNR3698787	Success	159552	

	vehicle_type	pickup_location	drop_location	Avg VTAT	Avg CTAT	…	\
0	Auto	Area-3	Area-2	5.42	18.46	…	
1	Mini	Area-7	Area-6	0.00	1.00	…	
2	Bike	Area-40	Area-24	0.00	1.00	…	
3	Prime Sedan	Area-11	Area-24	0.00	1.00	…	
4	Bike	Area-41	Area-45	0.00	1.00	…	
…	…	…	…	…	…	…	
49994	eBike	Area-39	Area-34	11.53	2.42	…	
49995	Auto	Area-17	Area-21	5.56	9.05	…	
49996	Auto	Area-37	Area-43	0.00	1.00	…	
49997	eBike	Area-48	Area-29	5.15	13.72	…	
49998	Bike	Area-1	Area-37	14.89	28.07	…	

	Cancelled Rides by Driver	Reason for Cancelling by Driver	\
0	0	Health Issue	
1	1	The customer was coughing/sick	
2	1	Personal & Car related issues	
3	1	The customer was coughing/sick	
4	0	Health Issue	
…	…	…	
49994	0	Health Issue	
49995	0	Health Issue	
49996	1	More than permitted people in there	
49997	0	Health Issue	
49998	0	Health Issue	

	Incomplete Rides	Incomplete Rides Reason	Booking Value	Payment Method	\
0	0	Critical Situation	868.06	Wallet	
1	0	Critical Situation	1000.00	online	
2	0	Critical Situation	1000.00	online	
3	0	Critical Situation	1000.00	online	
4	1	Other Issue	1000.00	online	

```
…                  …                      …              …              …
49994                0      Critical Situation        1217.43           Card
49995                0      Critical Situation        1369.51            UPI
49996                0      Critical Situation        1000.00         online
49997                0      Critical Situation        1930.49            UPI
49998                0      Critical Situation        1534.66           Cash

       Ride Distance Driver Ratings  Customer Rating            DateTime
0              28.50             4.4              4.4 2024-01-28 06:00:00
1              30.00             4.5              4.0 2024-01-26 03:00:00
2              30.00             4.5              4.0 2024-01-15 16:00:00
3              30.00             4.5              4.0 2024-01-02 22:00:00
4              30.00             4.5              4.0 2024-01-30 22:00:00
…                  …               …                …                  …
49994          10.52             4.1              3.2 2024-01-02 00:00:00
49995          11.04             5.0              4.3 2024-01-27 02:00:00
49996          30.00             4.5              4.0 2024-01-17 23:00:00
49997          14.49             3.1              4.5 2024-01-18 22:00:00
49998           4.60             3.3              4.1 2024-01-28 14:00:00

[49162 rows x 22 columns]
```

[68]:
```
df['Date'] = pd.to_datetime(df['Date'])
df['Time'] = pd.to_datetime(df['Time'], format='%H:%M:%S').dt.time
```

```
C:\Users\admin\AppData\Local\Temp\ipykernel_13040\850320488.py:1: UserWarning:
Parsing dates in %d/%m/%Y format when dayfirst=False (the default) was
specified. Pass `dayfirst=True` or specify a format to silence this warning.
  df['Date'] = pd.to_datetime(df['Date'])
```

[69]:
```
df['Date']
```

[69]:
```
0       2024-01-28
1       2024-01-26
2       2024-01-15
3       2024-01-02
4       2024-01-30
           …
49994   2024-01-02
49995   2024-01-27
49996   2024-01-17
49997   2024-01-18
49998   2024-01-28
Name: Date, Length: 49999, dtype: datetime64[ns]
```

[70]:
```
df.groupby('booking_status').sum(numeric_only=True)
```

[70]:

booking_status	customer_id	Avg VTAT	Avg CTAT	Cancelled by Customer	Cancelled Rides by Driver	Incomplete Rides	Booking Value	Ride Distance	Driver Ratings	Customer Rating
Cancelled by Customer	2101886126	0.00	3799.00	3799	0	0	3799000.00	113970.00	17095.5	15196.0
Cancelled by Driver	5345534305	0.00	9610.00	0	9610	0	9610000.00	288300.00	43245.0	38440.0
Incomplete	1712090228	0.00	3106.00	0	0	3106	3106000.00	93180.00	13977.0	12424.0
Success	18388935238	350967.71	521155.68	0	0	0	34266664.59	852072.14	134037.8	133993.8

[71]:
```
df.groupby('vehicle_type').sum(numeric_only=True)
```

[71]:

vehicle_type	customer_id	Avg VTAT	Avg CTAT	Cancelled by Customer	Cancelled Rides by Driver	Incomplete Rides	Booking Value
Auto	3922891981	49938.22	76346.03	518	1338	454	7180612.58
Bike	4011373510	50629.92	77945.56	561	1394	415	7327790.69
Mini	3870396355	49094.98	75470.81	550	1318	416	7124216.80
Prime Plus	3997533568	50200.57	76939.49	571	1415	462	7354974.66
Prime SUV	3921213939	50193.53	76452.05	529	1395	449	7257966.98
Prime Sedan	3936127072	50992.07	77288.44	531			
eBike	3888909472	49918.42	77228.30	539			

```
Prime Sedan                          1406              452     7308408.86
eBike                                1344              458     7227694.02

              Ride Distance  Driver Ratings  Customer Rating
vehicle_type
Auto             192543.20         29574.0          28419.4
Bike             194421.88         30079.1          28873.5
Mini             189429.16         29202.2          28002.9
Prime Plus       194278.72         30220.6          29035.1
Prime SUV        192253.74         29752.9          28623.8
Prime Sedan      193859.23         29958.9          28710.3
eBike            190736.21         29567.6          28388.8
```

[72]:
```
df.groupby('pickup_location').sum(numeric_only=True)
```

[72]:
```
                 customer_id  Avg VTAT  Avg CTAT  Cancelled  by Customer  \
pickup_location
Area-1             576556157   7634.29  11192.36                     74
Area-10            536267697   6982.81  10761.43                     83
Area-11            561837661   7465.75  11876.31                     71
Area-12            555083480   7479.84  11408.06                     77
Area-13            548748994   6941.66  10524.06                     62
Area-14            524543711   6334.85   9727.79                     73
Area-15            556881458   7336.15  10681.47                     66
Area-16            549195669   7039.57  10630.02                     74
Area-17            550926249   6951.12  10680.48                     69
Area-18            538295609   7104.89  10878.30                     57
Area-19            549338763   7292.00  10417.84                     83
Area-2             529384662   7106.82  10965.54                     70
Area-20            559262201   6948.62  10830.50                     78
Area-21            535773123   7117.27  11133.83                     67
Area-22            581800737   7027.02  10996.36                     82
Area-23            544101191   6992.99  10610.24                     83
Area-24            574484470   7191.56  11284.83                     87
Area-25            529481162   6898.39  10308.93                     84
Area-26            538326669   6492.51  10283.98                     77
Area-27            536073476   6303.10   9777.51                     82
Area-28            525072972   6898.40  10547.83                     67
Area-29            579362165   7395.40  11123.16                     72
Area-3             530539971   6794.49  10253.34                     73
Area-30            548804619   6848.47  10732.33                     76
Area-31            529046442   6541.02  10139.15                     66
Area-32            535047434   6634.71  10119.55                     62
Area-33            543914161   7312.99  10778.08                     71
Area-34            547531464   7021.51  10218.83                     85
Area-35            527902014   6759.34  10479.30                     74
Area-36            534788095   6543.95   9755.47                     80
```

Area-37	521755708	6441.98	9722.03	73
Area-38	552314748	6863.41	11085.74	70
Area-39	595158618	8000.02	12055.33	97
Area-4	587342594	7035.60	11345.31	83
Area-40	536322612	6774.84	10916.88	76
Area-41	568333268	6948.95	10626.21	92
Area-42	536175168	6756.49	10138.41	78
Area-43	548668079	7018.17	10918.74	64
Area-44	577806183	7290.40	11172.22	79
Area-45	551506229	6971.82	11121.59	81
Area-46	553635533	6992.88	10616.84	84
Area-47	556496826	7334.73	11663.37	80
Area-48	573981712	7032.71	10829.89	94
Area-49	540385009	7131.32	10644.32	70
Area-5	561543761	7392.90	11005.53	77
Area-50	554566646	7548.85	11226.99	67
Area-6	542638581	6974.93	10725.13	81
Area-7	531712952	6510.48	10195.68	79
Area-8	604127056	7351.04	11239.07	76
Area-9	575602138	7204.70	11304.52	73

pickup_location	Cancelled Rides by Driver	Incomplete Rides	Booking Value	\
Area-1	204	53	1057576.27	
Area-10	197	53	984966.43	
Area-11	172	72	1052214.56	
Area-12	191	53	1046744.62	
Area-13	210	64	1000400.29	
Area-14	192	70	985641.11	
Area-15	188	71	1038609.76	
Area-16	192	58	1005856.09	
Area-17	193	53	1014104.37	
Area-18	200	47	1034095.72	
Area-19	188	58	1004876.95	
Area-2	157	56	979988.72	
Area-20	183	64	1017428.00	
Area-21	164	63	978645.31	
Area-22	197	71	1063727.38	
Area-23	183	48	1015895.79	
Area-24	180	64	1039861.41	
Area-25	168	77	990236.58	
Area-26	194	74	1009275.26	
Area-27	216	59	962997.76	
Area-28	191	59	975952.37	
Area-29	215	55	1057825.53	
Area-3	201	60	985054.54	
Area-30	203	78	1030106.58	

Area-31	210	57	983784.00
Area-32	208	58	972488.62
Area-33	184	51	1011436.72
Area-34	208	66	1007915.29
Area-35	167	63	978320.56
Area-36	198	77	990750.66
Area-37	206	61	944904.32
Area-38	201	73	1013283.71
Area-39	184	62	1115256.62
Area-4	217	73	1057315.16
Area-40	197	71	999522.86
Area-41	212	50	1026835.55
Area-42	208	52	997781.74
Area-43	188	65	1006019.61
Area-44	193	71	1062822.34
Area-45	172	67	1012792.09
Area-46	176	61	1006816.25
Area-47	163	52	996493.07
Area-48	201	64	1031671.96
Area-49	179	56	1022374.07
Area-5	189	65	1034912.27
Area-50	186	62	1036473.87
Area-6	188	61	1028309.78
Area-7	192	63	985751.91
Area-8	201	67	1060354.00
Area-9	203	58	1065196.16

pickup_location	Ride Distance	Driver Ratings	Customer Rating
Area-1	27909.42	4263.4	4100.5
Area-10	26597.67	4133.1	3953.8
Area-11	27607.02	4315.5	4152.3
Area-12	26986.45	4244.6	4067.6
Area-13	26268.05	4097.2	3938.0
Area-14	25945.51	3956.3	3799.8
Area-15	27390.26	4228.4	4038.0
Area-16	27006.67	4150.0	3983.5
Area-17	26174.01	4090.9	3906.8
Area-18	26533.84	4101.5	3943.2
Area-19	26984.82	4185.1	4015.1
Area-2	25774.61	3998.2	3846.0
Area-20	27105.61	4191.6	4044.4
Area-21	26186.96	3994.0	3847.0
Area-22	27963.77	4309.2	4142.0
Area-23	26806.33	4093.9	3935.1
Area-24	27923.92	4266.0	4084.1
Area-25	26107.69	4075.2	3922.5

```
Area-26                25974.79          4109.5           3949.1
Area-27                26011.50          4023.1           3837.2
Area-28                26320.84          4043.7           3874.4
Area-29                27785.13          4345.4           4171.5
Area-3                 26215.36          4081.4           3920.5
Area-30                26795.73          4230.0           4027.0
Area-31                26619.14          4035.5           3868.5
Area-32                26457.50          4034.3           3888.1
Area-33                26445.88          4136.7           3988.2
Area-34                27447.03          4256.1           4043.8
Area-35                25978.37          4002.2           3851.3
Area-36                26073.24          4076.4           3897.0
Area-37                25696.43          3943.7           3802.7
Area-38                27727.66          4201.2           4039.1
Area-39                29226.64          4585.4           4410.1
Area-4                 28483.78          4409.8           4220.4
Area-40                26847.85          4155.1           3997.0
Area-41                27248.18          4243.6           4062.7
Area-42                26428.91          4087.5           3941.4
Area-43                27053.18          4092.1           3962.3
Area-44                27972.41          4351.4           4165.5
Area-45                26952.74          4160.7           3996.1
Area-46                26980.77          4123.6           4040.4
Area-47                26732.85          4123.2           3960.8
Area-48                28408.68          4321.4           4136.4
Area-49                26541.14          4107.7           3931.9
Area-5                 27380.52          4261.2           4082.1
Area-50                26914.72          4253.8           4065.4
Area-6                 27162.85          4159.9           4023.4
Area-7                 26109.63          3980.6           3840.8
Area-8                 28059.61          4370.9           4188.6
Area-9                 28196.47          4354.1           4150.4
```

[73]:
```python
result = df.groupby('booking_status').agg({
    'Date': 'last',    # or 'min'/'max'
    'Time': 'last',
    'Driver Ratings': 'sum'     # example numerical column
})
print(result)
```

```
                              Date      Time  Driver Ratings
booking_status
Cancelled by Customer 2024-01-13  20:00:00         17095.5
Cancelled by Driver   2024-01-17  23:00:00         43245.0
Incomplete            2024-01-11  15:00:00         13977.0
Success               2024-01-28  14:00:00        134037.8
```

[74]:
```
result=df.groupby('vehicle_type').agg({'Date':'first',
                                       'Time':'last',
                                       'Customer Rating':'sum'})
print(result)
```

```
                  Date      Time  Customer Rating
vehicle_type
Auto        2024-01-28  23:00:00          28419.4
Bike        2024-01-15  14:00:00          28873.5
Mini        2024-01-26  03:00:00          28002.9
Prime Plus  2024-01-14  18:00:00          29035.1
Prime SUV   2024-01-22  20:00:00          28623.8
Prime Sedan 2024-01-02  08:00:00          28710.3
eBike       2024-01-30  22:00:00          28388.8
```

[75]:
```
df.drop('DateTime',axis=1,inplace=True)
```

[76]:
```
df
```

[76]:

	Date	Time	booking_id	booking_status	customer_id	\
0	2024-01-28	06:00:00	CNR1721175	Success	329258	
1	2024-01-26	03:00:00	CNR2871422	Cancelled by Driver	201414	
2	2024-01-15	16:00:00	CNR6875935	Cancelled by Driver	301629	
3	2024-01-02	22:00:00	CNR6798834	Cancelled by Driver	319684	
4	2024-01-30	22:00:00	CNR9661713	Incomplete	330283	
…	…	…	…	…	…	
49994	2024-01-02	00:00:00	CNR7299340	Success	454202	
49995	2024-01-27	02:00:00	CNR0378586	Success	349873	
49996	2024-01-17	23:00:00	CNR2461856	Cancelled by Driver	880053	
49997	2024-01-18	22:00:00	CNR9524579	Success	752806	
49998	2024-01-28	14:00:00	CNR3698787	Success	159552	

	vehicle_type	pickup_location	drop_location	Avg VTAT	Avg CTAT	…	\
0	Auto	Area-3	Area-2	5.42	18.46	…	
1	Mini	Area-7	Area-6	0.00	1.00	…	
2	Bike	Area-40	Area-24	0.00	1.00	…	
3	Prime Sedan	Area-11	Area-24	0.00	1.00	…	
4	Bike	Area-41	Area-45	0.00	1.00	…	
…	…	…	…	…	…	…	
49994	eBike	Area-39	Area-34	11.53	2.42	…	
49995	Auto	Area-17	Area-21	5.56	9.05	…	
49996	Auto	Area-37	Area-43	0.00	1.00	…	
49997	eBike	Area-48	Area-29	5.15	13.72	…	
49998	Bike	Area-1	Area-37	14.89	28.07	…	

	Reason for Cancelling by Customer	Cancelled Rides by Driver	\
0	Traffic	0	
1	Traffic	1	

```
2                                Traffic                          1
3                                Traffic                          1
4                                Traffic                          0
…                                  …                            …
49994                            Traffic                          0
49995                            Traffic                          0
49996                            Traffic                          1
49997                            Traffic                          0
49998                            Traffic                          0

           Reason for Cancelling by Driver Incomplete Rides  \
0                            Health Issue                 0
1             The customer was coughing/sick              0
2              Personal & Car related issues              0
3             The customer was coughing/sick              0
4                            Health Issue                 1
…                                    …                  …
49994                        Health Issue                 0
49995                        Health Issue                 0
49996  More than permitted people in there                0
49997                        Health Issue                 0
49998                        Health Issue                 0

       Incomplete Rides Reason Booking Value  Payment Method Ride Distance  \
0           Critical Situation        868.06          Wallet         28.50
1           Critical Situation       1000.00          online         30.00
2           Critical Situation       1000.00          online         30.00
3           Critical Situation       1000.00          online         30.00
4                  Other Issue       1000.00          online         30.00
…                          …            …               …            …
49994       Critical Situation       1217.43            Card         10.52
49995       Critical Situation       1369.51             UPI         11.04
49996       Critical Situation       1000.00          online         30.00
49997       Critical Situation       1930.49             UPI         14.49
49998       Critical Situation       1534.66            Cash          4.60

       Driver Ratings  Customer Rating
0                 4.4              4.4
1                 4.5              4.0
2                 4.5              4.0
3                 4.5              4.0
4                 4.5              4.0
…                   …                …
49994             4.1              3.2
49995             5.0              4.3
49996             4.5              4.0
49997             3.1              4.5
```

```
49998             3.3              4.1

[49999 rows x 21 columns]
```

[77]:
```
result=df.groupby('booking_status').agg({'Date':'first',
                                         'Time':'last',
                                         'vehicle_type':'sum'})
print(result)
```

```
                              Date      Time  \
booking_status
Cancelled by Customer 2024-01-01  20:00:00
Cancelled by Driver   2024-01-26  23:00:00
Incomplete            2024-01-30  15:00:00
Success               2024-01-28  14:00:00

                                                                vehicle_type
booking_status
Cancelled by Customer  Prime PluseBikeMiniMiniPrime SUVPrime SUVBikeB…
Cancelled by Driver    MiniBikePrime SedanMiniPrime SedanPrime PlusPr…
Incomplete             BikePrime PluseBikeMiniPrime SUVPrime SedaneBi…
Success                AutoPrime SedaneBikePrime SUVPrime SedaneBikeB…
```

[78]:
```
df['vehicle_type'].value_counts()
```

[78]:
```
vehicle_type
Prime Plus     7252
Bike           7223
Prime Sedan    7179
Prime SUV      7140
Auto           7098
eBike          7097
Mini           7010
Name: count, dtype: int64
```

[79]:
```
df['booking_status'].value_counts()
```

[79]:
```
booking_status
Success                 33484
Cancelled by Driver      9610
Cancelled by Customer    3799
Incomplete               3106
Name: count, dtype: int64
```

[80]:
```
df['booking_status']=df['booking_status'].str.upper()
```

[81]:
```
df.head()
```

[81]:
```
        Date      Time  booking_id       booking_status  customer_id  \
0 2024-01-28  06:00:00  CNR1721175              SUCCESS       329258
1 2024-01-26  03:00:00  CNR2871422  CANCELLED BY DRIVER       201414
2 2024-01-15  16:00:00  CNR6875935  CANCELLED BY DRIVER       301629
3 2024-01-02  22:00:00  CNR6798834  CANCELLED BY DRIVER       319684
4 2024-01-30  22:00:00  CNR9661713           INCOMPLETE       330283

  vehicle_type pickup_location drop_location  Avg VTAT  Avg CTAT  …  \
0         Auto          Area-3        Area-2      5.42     18.46  …
1         Mini          Area-7        Area-6      0.00      1.00  …
2         Bike         Area-40       Area-24      0.00      1.00  …
3  Prime Sedan         Area-11       Area-24      0.00      1.00  …
4         Bike         Area-41       Area-45      0.00      1.00  …

   Reason for Cancelling by Customer Cancelled Rides by Driver  \
0                           Traffic                          0
1                           Traffic                          1
2                           Traffic                          1
3                           Traffic                          1
4                           Traffic                          0

   Reason for Cancelling by Driver Incomplete Rides  Incomplete Rides Reason  \
0                     Health Issue                0       Critical Situation
1   The customer was coughing/sick                0       Critical Situation
2    Personal & Car related issues                0       Critical Situation
3   The customer was coughing/sick                0       Critical Situation
4                     Health Issue                1              Other Issue

  Booking Value  Payment Method Ride Distance  Driver Ratings  Customer Rating
0        868.06          Wallet          28.5             4.4              4.4
1       1000.00          online          30.0             4.5              4.0
2       1000.00          online          30.0             4.5              4.0
3       1000.00          online          30.0             4.5              4.0
4       1000.00          online          30.0             4.5              4.0

[5 rows x 21 columns]
```

[82]:
```
df['vehicle_type']=df['vehicle_type'].str.upper()
```

[83]:
```
df.head()
```

[83]:
```
        Date      Time  booking_id       booking_status  customer_id  \
0 2024-01-28  06:00:00  CNR1721175              SUCCESS       329258
1 2024-01-26  03:00:00  CNR2871422  CANCELLED BY DRIVER       201414
2 2024-01-15  16:00:00  CNR6875935  CANCELLED BY DRIVER       301629
3 2024-01-02  22:00:00  CNR6798834  CANCELLED BY DRIVER       319684
4 2024-01-30  22:00:00  CNR9661713           INCOMPLETE       330283
```

```
  vehicle_type pickup_location drop_location  Avg VTAT  Avg CTAT  …  \
0         AUTO          Area-3        Area-2      5.42     18.46  …
1         MINI          Area-7        Area-6      0.00      1.00  …
2         BIKE         Area-40       Area-24      0.00      1.00  …
3  PRIME SEDAN         Area-11       Area-24      0.00      1.00  …
4         BIKE         Area-41       Area-45      0.00      1.00  …

   Reason for Cancelling by Customer Cancelled Rides by Driver  \
0                           Traffic                          0
1                           Traffic                          1
2                           Traffic                          1
3                           Traffic                          1
4                           Traffic                          0

   Reason for Cancelling by Driver Incomplete Rides  Incomplete Rides Reason  \
0                     Health Issue                0       Critical Situation
1   The customer was coughing/sick                0       Critical Situation
2    Personal & Car related issues                0       Critical Situation
3   The customer was coughing/sick                0       Critical Situation
4                     Health Issue                1              Other Issue

  Booking Value  Payment Method Ride Distance  Driver Ratings  Customer Rating
0        868.06          Wallet          28.5             4.4              4.4
1       1000.00          online          30.0             4.5              4.0
2       1000.00          online          30.0             4.5              4.0
3       1000.00          online          30.0             4.5              4.0
4       1000.00          online          30.0             4.5              4.0

[5 rows x 21 columns]
```

[84]:
```
df['driver_ratings']=df['Driver Ratings']*0.1
```

[85]:
```
df.head()
```

[85]:
```
        Date      Time  booking_id       booking_status  customer_id  \
0 2024-01-28  06:00:00  CNR1721175              SUCCESS       329258
1 2024-01-26  03:00:00  CNR2871422  CANCELLED BY DRIVER       201414
2 2024-01-15  16:00:00  CNR6875935  CANCELLED BY DRIVER       301629
3 2024-01-02  22:00:00  CNR6798834  CANCELLED BY DRIVER       319684
4 2024-01-30  22:00:00  CNR9661713           INCOMPLETE       330283

  vehicle_type pickup_location drop_location  Avg VTAT  Avg CTAT  …  \
0         AUTO          Area-3        Area-2      5.42     18.46  …
1         MINI          Area-7        Area-6      0.00      1.00  …
2         BIKE         Area-40       Area-24      0.00      1.00  …
3  PRIME SEDAN         Area-11       Area-24      0.00      1.00  …
```

```
4         BIKE         Area-41       Area-45      0.00      1.00  …

   Cancelled Rides by Driver Reason for Cancelling by Driver  \
0                         0                    Health Issue
1                         1  The customer was coughing/sick
2                         1   Personal & Car related issues
3                         1  The customer was coughing/sick
4                         0                    Health Issue

   Incomplete Rides Incomplete Rides Reason  Booking Value Payment Method  \
0                 0      Critical Situation         868.06         Wallet
1                 0      Critical Situation        1000.00         online
2                 0      Critical Situation        1000.00         online
3                 0      Critical Situation        1000.00         online
4                 1             Other Issue        1000.00         online

   Ride Distance Driver Ratings  Customer Rating  driver_ratings
0           28.5            4.4              4.4            0.44
1           30.0            4.5              4.0            0.45
2           30.0            4.5              4.0            0.45
3           30.0            4.5              4.0            0.45
4           30.0            4.5              4.0            0.45

[5 rows x 22 columns]
```

[86]:
```
df['customer_rating']=df['Customer Rating']*20
```

[87]:
```
df.head()
```

[87]:
```
        Date      Time  booking_id       booking_status  customer_id  \
0 2024-01-28  06:00:00  CNR1721175              SUCCESS       329258
1 2024-01-26  03:00:00  CNR2871422  CANCELLED BY DRIVER       201414
2 2024-01-15  16:00:00  CNR6875935  CANCELLED BY DRIVER       301629
3 2024-01-02  22:00:00  CNR6798834  CANCELLED BY DRIVER       319684
4 2024-01-30  22:00:00  CNR9661713           INCOMPLETE       330283

  vehicle_type pickup_location drop_location  Avg VTAT  Avg CTAT  …  \
0         AUTO          Area-3        Area-2      5.42     18.46  …
1         MINI          Area-7        Area-6      0.00      1.00  …
2         BIKE         Area-40       Area-24      0.00      1.00  …
3  PRIME SEDAN         Area-11       Area-24      0.00      1.00  …
4         BIKE         Area-41       Area-45      0.00      1.00  …

   Reason for Cancelling by Driver Incomplete Rides  Incomplete Rides Reason  \
0                     Health Issue                0       Critical Situation
1   The customer was coughing/sick                0       Critical Situation
2    Personal & Car related issues                0       Critical Situation
```

```
3   The customer was coughing/sick                0       Critical Situation
4                     Health Issue                1              Other Issue

  Booking Value  Payment Method Ride Distance  Driver Ratings Customer Rating  \
0        868.06          Wallet          28.5             4.4             4.4
1       1000.00          online          30.0             4.5             4.0
2       1000.00          online          30.0             4.5             4.0
3       1000.00          online          30.0             4.5             4.0
4       1000.00          online          30.0             4.5             4.0

   driver_ratings  customer_rating
0            0.44             88.0
1            0.45             80.0
2            0.45             80.0
3            0.45             80.0
4            0.45             80.0

[5 rows x 23 columns]
```

[88]:
```
df.drop('driver_ratings',axis=1,inplace=True)
```

[89]:
```
df.drop('customer_rating',axis=1,inplace=True)
```

[90]:
```
df
```

[90]:
```
            Date      Time  booking_id       booking_status  customer_id  \
0     2024-01-28  06:00:00  CNR1721175              SUCCESS       329258
1     2024-01-26  03:00:00  CNR2871422  CANCELLED BY DRIVER       201414
2     2024-01-15  16:00:00  CNR6875935  CANCELLED BY DRIVER       301629
3     2024-01-02  22:00:00  CNR6798834  CANCELLED BY DRIVER       319684
4     2024-01-30  22:00:00  CNR9661713           INCOMPLETE       330283
…            …       …          …                    …           …
49994 2024-01-02  00:00:00  CNR7299340              SUCCESS       454202
49995 2024-01-27  02:00:00  CNR0378586              SUCCESS       349873
49996 2024-01-17  23:00:00  CNR2461856  CANCELLED BY DRIVER       880053
49997 2024-01-18  22:00:00  CNR9524579              SUCCESS       752806
49998 2024-01-28  14:00:00  CNR3698787              SUCCESS       159552

      vehicle_type pickup_location drop_location  Avg VTAT  Avg CTAT  …  \
0             AUTO          Area-3        Area-2      5.42     18.46  …
1             MINI          Area-7        Area-6      0.00      1.00  …
2             BIKE         Area-40       Area-24      0.00      1.00  …
3      PRIME SEDAN         Area-11       Area-24      0.00      1.00  …
4             BIKE         Area-41       Area-45      0.00      1.00  …
…             …            …          …        …       …  …
49994        EBIKE         Area-39       Area-34     11.53      2.42  …
49995         AUTO         Area-17       Area-21      5.56      9.05  …
```

```
49996         AUTO         Area-37       Area-43      0.00      1.00  …
49997        EBIKE         Area-48       Area-29      5.15     13.72  …
49998         BIKE          Area-1       Area-37     14.89     28.07  …

       Reason for Cancelling by Customer Cancelled Rides by Driver  \
0                                Traffic                          0
1                                Traffic                          1
2                                Traffic                          1
3                                Traffic                          1
4                                Traffic                          0
…                                  …                            …
49994                            Traffic                          0
49995                            Traffic                          0
49996                            Traffic                          1
49997                            Traffic                          0
49998                            Traffic                          0

           Reason for Cancelling by Driver Incomplete Rides  \
0                             Health Issue                 0
1           The customer was coughing/sick                 0
2            Personal & Car related issues                 0
3           The customer was coughing/sick                 0
4                             Health Issue                 1
…                                    …                   …
49994                         Health Issue                 0
49995                         Health Issue                 0
49996  More than permitted people in there                 0
49997                         Health Issue                 0
49998                         Health Issue                 0

       Incomplete Rides Reason Booking Value  Payment Method Ride Distance  \
0           Critical Situation        868.06          Wallet         28.50
1           Critical Situation       1000.00          online         30.00
2           Critical Situation       1000.00          online         30.00
3           Critical Situation       1000.00          online         30.00
4                 Other Issue       1000.00          online         30.00
…                         …             …               …             …
49994       Critical Situation       1217.43            Card         10.52
49995       Critical Situation       1369.51             UPI         11.04
49996       Critical Situation       1000.00          online         30.00
49997       Critical Situation       1930.49             UPI         14.49
49998       Critical Situation       1534.66            Cash          4.60

       Driver Ratings  Customer Rating
0                 4.4              4.4
1                 4.5              4.0
2                 4.5              4.0
```

```
3                 4.5              4.0
4                 4.5              4.0
…                 …                …
49994             4.1              3.2
49995             5.0              4.3
49996             4.5              4.0
49997             3.1              4.5
49998             3.3              4.1

[49999 rows x 21 columns]
```

[91]:
```
df['Payment Method']=df['Payment Method'].str.upper()
```

[92]:
```
df
```

[92]:
```
            Date      Time  booking_id       booking_status  customer_id  \
0     2024-01-28  06:00:00  CNR1721175              SUCCESS       329258
1     2024-01-26  03:00:00  CNR2871422  CANCELLED BY DRIVER       201414
2     2024-01-15  16:00:00  CNR6875935  CANCELLED BY DRIVER       301629
3     2024-01-02  22:00:00  CNR6798834  CANCELLED BY DRIVER       319684
4     2024-01-30  22:00:00  CNR9661713           INCOMPLETE       330283
…            …         …           …                    …            …
49994 2024-01-02  00:00:00  CNR7299340              SUCCESS       454202
49995 2024-01-27  02:00:00  CNR0378586              SUCCESS       349873
49996 2024-01-17  23:00:00  CNR2461856  CANCELLED BY DRIVER       880053
49997 2024-01-18  22:00:00  CNR9524579              SUCCESS       752806
49998 2024-01-28  14:00:00  CNR3698787              SUCCESS       159552

      vehicle_type pickup_location drop_location  Avg VTAT  Avg CTAT  …  \
0             AUTO          Area-3        Area-2      5.42     18.46  …
1             MINI          Area-7        Area-6      0.00      1.00  …
2             BIKE         Area-40       Area-24      0.00      1.00  …
3      PRIME SEDAN         Area-11       Area-24      0.00      1.00  …
4             BIKE         Area-41       Area-45      0.00      1.00  …
…                …               …             …         …         …  …
49994        EBIKE         Area-39       Area-34     11.53      2.42  …
49995         AUTO         Area-17       Area-21      5.56      9.05  …
49996         AUTO         Area-37       Area-43      0.00      1.00  …
49997        EBIKE         Area-48       Area-29      5.15     13.72  …
49998         BIKE          Area-1       Area-37     14.89     28.07  …

       Reason for Cancelling by Customer Cancelled Rides by Driver  \
0                                Traffic                         0
1                                Traffic                         1
2                                Traffic                         1
3                                Traffic                         1
4                                Traffic                         0
```

```
…                                 …                          …
49994                          Traffic                          0
49995                          Traffic                          0
49996                          Traffic                          1
49997                          Traffic                          0
49998                          Traffic                          0

           Reason for Cancelling by Driver Incomplete Rides  \
0                            Health Issue                 0
1             The customer was coughing/sick                0
2              Personal & Car related issues                0
3             The customer was coughing/sick                0
4                            Health Issue                 1
…                                      …               …
49994                        Health Issue                 0
49995                        Health Issue                 0
49996  More than permitted people in there                0
49997                        Health Issue                 0
49998                        Health Issue                 0

       Incomplete Rides Reason Booking Value  Payment Method Ride Distance  \
0           Critical Situation        868.06           WALLET         28.50
1           Critical Situation       1000.00           ONLINE         30.00
2           Critical Situation       1000.00           ONLINE         30.00
3           Critical Situation       1000.00           ONLINE         30.00
4                 Other Issue        1000.00           ONLINE         30.00
…                         …               …                …           …
49994       Critical Situation       1217.43             CARD         10.52
49995       Critical Situation       1369.51              UPI         11.04
49996       Critical Situation       1000.00           ONLINE         30.00
49997       Critical Situation       1930.49              UPI         14.49
49998       Critical Situation       1534.66             CASH          4.60

       Driver Ratings  Customer Rating
0                 4.4              4.4
1                 4.5              4.0
2                 4.5              4.0
3                 4.5              4.0
4                 4.5              4.0
…                   …                …
49994             4.1              3.2
49995             5.0              4.3
49996             4.5              4.0
49997             3.1              4.5
49998             3.3              4.1

[49999 rows x 21 columns]
```

[93]:
```
df.drop('Payment Method',axis=1,inplace=True)
```

[94]:
```
df.head()
```

[94]:

	Date	Time	booking_id	booking_status	customer_id	vehicle_type	pickup_location	drop_location	Avg VTAT	Avg CTAT	Cancelled by Customer	Reason for Cancelling by Customer	Cancelled Rides by Driver	Reason for Cancelling by Driver	Incomplete Rides	Incomplete Rides Reason	Booking Value	Ride Distance	Driver Ratings	Customer Rating
0	2024-01-28	06:00:00	CNR1721175	SUCCESS	329258	AUTO	Area-3	Area-2	5.42	18.46	0	Traffic	0	Health Issue	0	Critical Situation	868.06	28.5	4.4	4.4
1	2024-01-26	03:00:00	CNR2871422	CANCELLED BY DRIVER	201414	MINI	Area-7	Area-6	0.00	1.00	0	Traffic	1	The customer was coughing/sick	0	Critical Situation	1000.00	30.0	4.5	4.0
2	2024-01-15	16:00:00	CNR6875935	CANCELLED BY DRIVER	301629	BIKE	Area-40	Area-24	0.00	1.00	0	Traffic	1	Personal & Car related issues	0	Critical Situation	1000.00	30.0	4.5	4.0
3	2024-01-02	22:00:00	CNR6798834	CANCELLED BY DRIVER	319684	PRIME SEDAN	Area-11	Area-24	0.00	1.00	0	Traffic	1	The customer was coughing/sick	0	Critical Situation	1000.00	30.0	4.5	4.0
4	2024-01-30	22:00:00	CNR9661713	INCOMPLETE	330283	BIKE	Area-41	Area-45	0.00	1.00	0	Traffic	0	Health Issue	1	Other Issue	1000.00	30.0	4.5	4.0

[ ]:

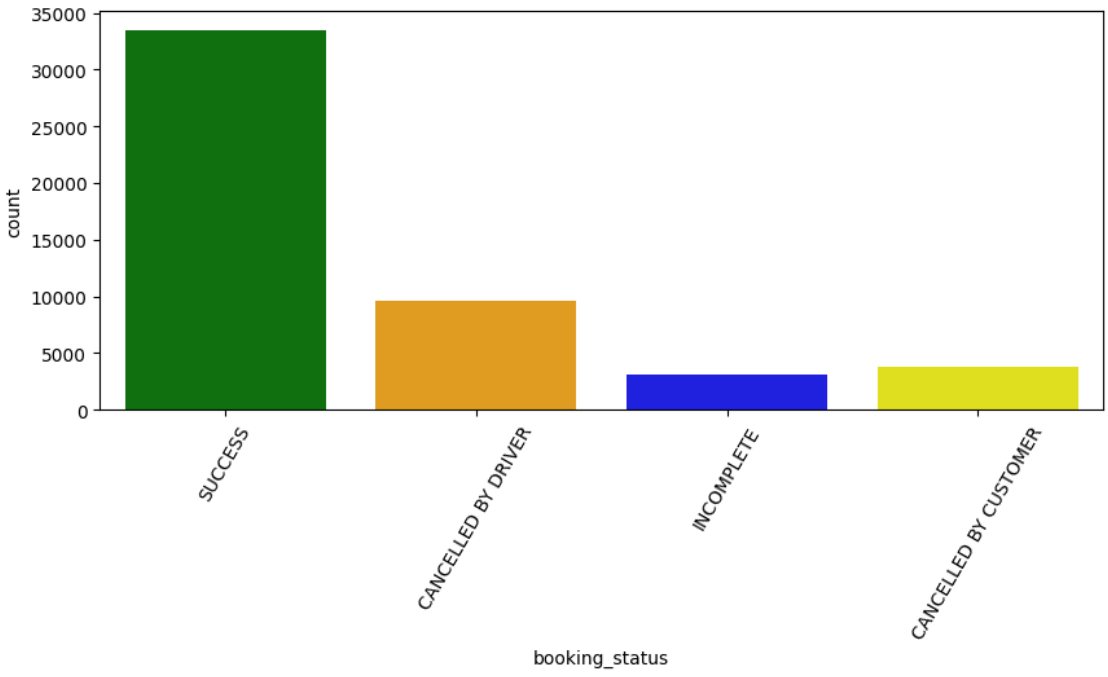
[95]:
```
### CHART␣
↪Work=============================================###=======================================
```

[96]:
```
df.columns
```

[96]:
```
Index(['Date', 'Time', 'booking_id', 'booking_status', 'customer_id',
       'vehicle_type', 'pickup_location', 'drop_location', 'Avg VTAT',
       'Avg CTAT', 'Cancelled  by Customer',
       'Reason for Cancelling by Customer', 'Cancelled Rides by Driver',
       'Reason for Cancelling by Driver', 'Incomplete Rides',
       'Incomplete Rides Reason', 'Booking Value', 'Ride Distance',
       'Driver Ratings', 'Customer Rating'],
      dtype='object')
```

[97]:
```
plt.figure(figsize=(10,4))
colors=('green','orange','blue','yellow')
ax=sb.countplot(data=df,x='booking_status',palette=colors)
plt.xticks(rotation=60)
plt.show()
```

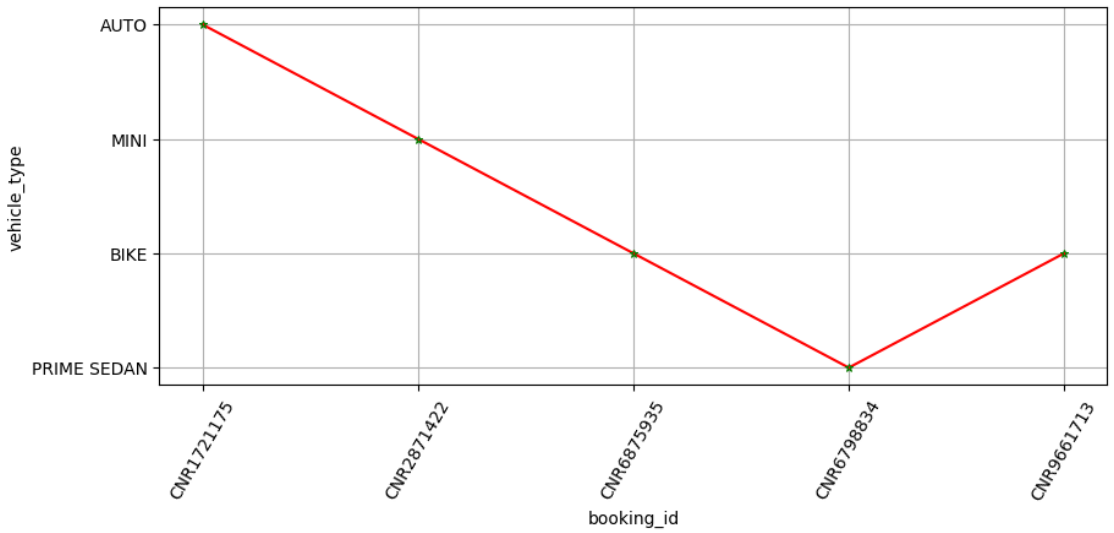
```
C:\Users\admin\AppData\Local\Temp\ipykernel_13040\3976460694.py:3:
FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in
v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same
effect.

  ax=sb.countplot(data=df,x='booking_status',palette=colors)
```



[98]:
```
df.columns
```

[98]:
```
Index(['Date', 'Time', 'booking_id', 'booking_status', 'customer_id',
       'vehicle_type', 'pickup_location', 'drop_location', 'Avg VTAT',
       'Avg CTAT', 'Cancelled  by Customer',
       'Reason for Cancelling by Customer', 'Cancelled Rides by Driver',
       'Reason for Cancelling by Driver', 'Incomplete Rides',
       'Incomplete Rides Reason', 'Booking Value', 'Ride Distance',
       'Driver Ratings', 'Customer Rating'],
      dtype='object')
```

[99]:
```
plt.figure(figsize=(10,4))
sb.lineplot(data=df.
 ↪head(5),x='booking_id',y='vehicle_type',color='red',marker='*',␣
 ↪mec='green',ms='5')
plt.xticks(rotation=60)
plt.grid()
plt.show()
```



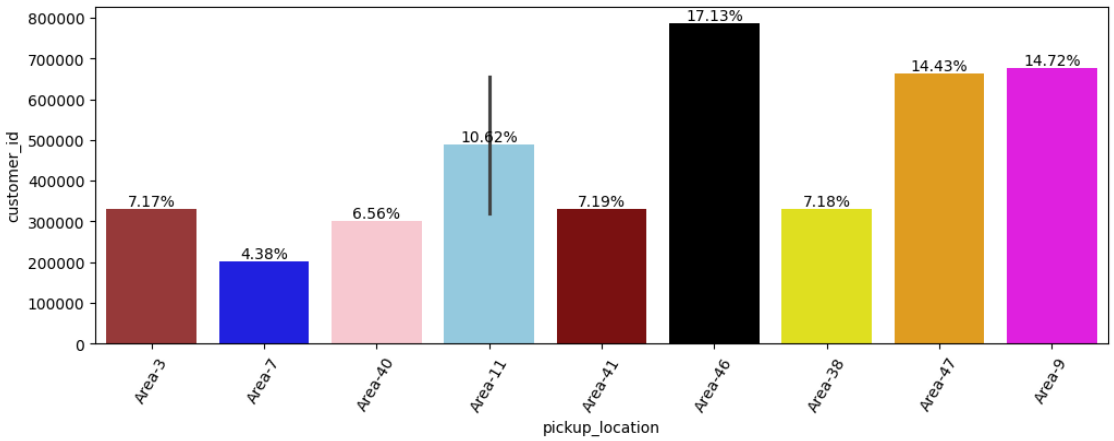
[100]:
```
df.columns
```

[100]:
```
Index(['Date', 'Time', 'booking_id', 'booking_status', 'customer_id',
       'vehicle_type', 'pickup_location', 'drop_location', 'Avg VTAT',
       'Avg CTAT', 'Cancelled  by Customer',
       'Reason for Cancelling by Customer', 'Cancelled Rides by Driver',
       'Reason for Cancelling by Driver', 'Incomplete Rides',
```

```
       'Incomplete Rides Reason', 'Booking Value', 'Ride Distance',
       'Driver Ratings', 'Customer Rating'],
      dtype='object')
```

[101]:
```
plt.figure(figsize=(12,4))
colors=('brown','blue','pink','skyblue','darkred','black','yellow','orange','magenta')
ax=sb.barplot(data=df.
 ↪head(10),x='pickup_location',y='customer_id',palette=colors)
total = sum(df.head(10)['customer_id'])
for bar in ax.patches:
    height = bar.get_height()
    percentage = f'{100 * height / total:.2f}%'  # Calculate % of total
    ax.annotate(percentage,                 # Text to display
                (bar.get_x() + bar.get_width() / 2, height),  # Position␣
 ↪(center of bar)
                ha='center', va='bottom', fontsize=10, color='black')

plt.xticks(rotation=60)
plt.show()
```

```
C:\Users\admin\AppData\Local\Temp\ipykernel_13040\1705995479.py:3:
FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in
v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same
effect.

  ax=sb.barplot(data=df.head(10),x='pickup_location',y='customer_id',palette=col
ors)
```

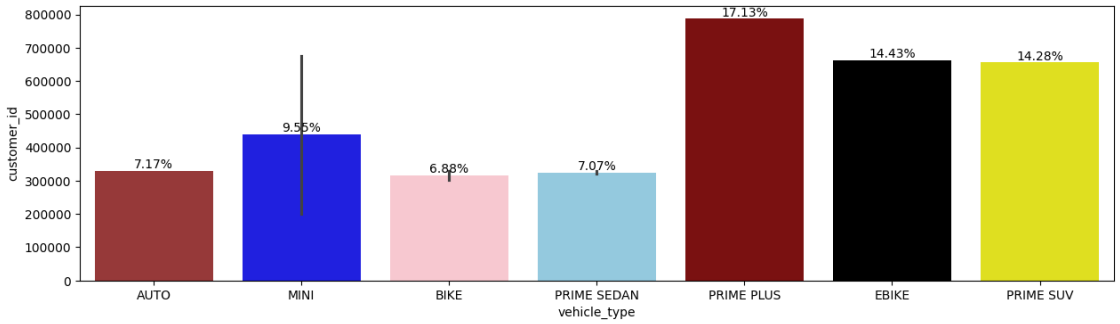


[102]:
```
df.columns
```

[102]:
```
Index(['Date', 'Time', 'booking_id', 'booking_status', 'customer_id',
       'vehicle_type', 'pickup_location', 'drop_location', 'Avg VTAT',
       'Avg CTAT', 'Cancelled  by Customer',
       'Reason for Cancelling by Customer', 'Cancelled Rides by Driver',
       'Reason for Cancelling by Driver', 'Incomplete Rides',
       'Incomplete Rides Reason', 'Booking Value', 'Ride Distance',
       'Driver Ratings', 'Customer Rating'],
      dtype='object')
```

[103]:
```
plt.figure(figsize=(15,4))
colors=('brown','blue','pink','skyblue','darkred','black','yellow','orange','magenta')
ax=sb.barplot(data=df.head(10),x='vehicle_type',y='customer_id',palette=colors)
total=sum(df.head(10)['customer_id'])
for bar in ax.patches:
    height=bar.get_height()
    percentage=f'{100*height/total:.2f}%'
    ax.annotate(percentage,(bar.get_x()+bar.get_width()/2,height),
    ha='center',va='bottom',fontsize=10,color='black')
```

```
C:\Users\admin\AppData\Local\Temp\ipykernel_13040\3263018117.py:3:
FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in
v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same
effect.

ax=sb.barplot(data=df.head(10),x='vehicle_type',y='customer_id',palette=colors)
```



[104]:
```
df.columns
```

[104]:
```
Index(['Date', 'Time', 'booking_id', 'booking_status', 'customer_id',
       'vehicle_type', 'pickup_location', 'drop_location', 'Avg VTAT',
       'Avg CTAT', 'Cancelled  by Customer',
       'Reason for Cancelling by Customer', 'Cancelled Rides by Driver',
       'Reason for Cancelling by Driver', 'Incomplete Rides',
```

```
       'Incomplete Rides Reason', 'Booking Value', 'Ride Distance',
       'Driver Ratings', 'Customer Rating'],
      dtype='object')
```

[105]:
```
df
```

[105]:
```
            Date      Time  booking_id       booking_status  customer_id  \
0     2024-01-28  06:00:00  CNR1721175              SUCCESS       329258
1     2024-01-26  03:00:00  CNR2871422  CANCELLED BY DRIVER       201414
2     2024-01-15  16:00:00  CNR6875935  CANCELLED BY DRIVER       301629
3     2024-01-02  22:00:00  CNR6798834  CANCELLED BY DRIVER       319684
4     2024-01-30  22:00:00  CNR9661713           INCOMPLETE       330283
…            …         …           …                    …            …
49994 2024-01-02  00:00:00  CNR7299340              SUCCESS       454202
49995 2024-01-27  02:00:00  CNR0378586              SUCCESS       349873
49996 2024-01-17  23:00:00  CNR2461856  CANCELLED BY DRIVER       880053
49997 2024-01-18  22:00:00  CNR9524579              SUCCESS       752806
49998 2024-01-28  14:00:00  CNR3698787              SUCCESS       159552

      vehicle_type pickup_location drop_location  Avg VTAT  Avg CTAT  \
0             AUTO          Area-3        Area-2      5.42     18.46
1             MINI          Area-7        Area-6      0.00      1.00
2             BIKE         Area-40       Area-24      0.00      1.00
3      PRIME SEDAN         Area-11       Area-24      0.00      1.00
4             BIKE         Area-41       Area-45      0.00      1.00
…              …              …             …         …         …
49994        EBIKE         Area-39       Area-34     11.53      2.42
49995         AUTO         Area-17       Area-21      5.56      9.05
49996         AUTO         Area-37       Area-43      0.00      1.00
49997        EBIKE         Area-48       Area-29      5.15     13.72
49998         BIKE          Area-1       Area-37     14.89     28.07

       Cancelled  by Customer Reason for Cancelling by Customer  \
0                          0                          Traffic
1                          0                          Traffic
2                          0                          Traffic
3                          0                          Traffic
4                          0                          Traffic
…                        …                                …
49994                      0                          Traffic
49995                      0                          Traffic
49996                      0                          Traffic
49997                      0                          Traffic
49998                      0                          Traffic

       Cancelled Rides by Driver      Reason for Cancelling by Driver  \
0                              0                          Health Issue
```

```
1                            1       The customer was coughing/sick
2                            1        Personal & Car related issues
3                            1       The customer was coughing/sick
4                            0                        Health Issue
…                          …                                 …
49994                        0                        Health Issue
49995                        0                        Health Issue
49996                        1  More than permitted people in there
49997                        0                        Health Issue
49998                        0                        Health Issue

       Incomplete Rides Incomplete Rides Reason  Booking Value  Ride Distance  \
0                     0      Critical Situation         868.06          28.50
1                     0      Critical Situation        1000.00          30.00
2                     0      Critical Situation        1000.00          30.00
3                     0      Critical Situation        1000.00          30.00
4                     1             Other Issue        1000.00          30.00
…                   …                       …             …             …
49994                 0      Critical Situation        1217.43          10.52
49995                 0      Critical Situation        1369.51          11.04
49996                 0      Critical Situation        1000.00          30.00
49997                 0      Critical Situation        1930.49          14.49
49998                 0      Critical Situation        1534.66           4.60

       Driver Ratings  Customer Rating
0                 4.4              4.4
1                 4.5              4.0
2                 4.5              4.0
3                 4.5              4.0
4                 4.5              4.0
…                 …                …
49994             4.1              3.2
49995             5.0              4.3
49996             4.5              4.0
49997             3.1              4.5
49998             3.3              4.1

[49999 rows x 20 columns]
```

[106]:
```
df.columns
```

[106]:
```
Index(['Date', 'Time', 'booking_id', 'booking_status', 'customer_id',
       'vehicle_type', 'pickup_location', 'drop_location', 'Avg VTAT',
       'Avg CTAT', 'Cancelled  by Customer',
       'Reason for Cancelling by Customer', 'Cancelled Rides by Driver',
       'Reason for Cancelling by Driver', 'Incomplete Rides',
       'Incomplete Rides Reason', 'Booking Value', 'Ride Distance',
```

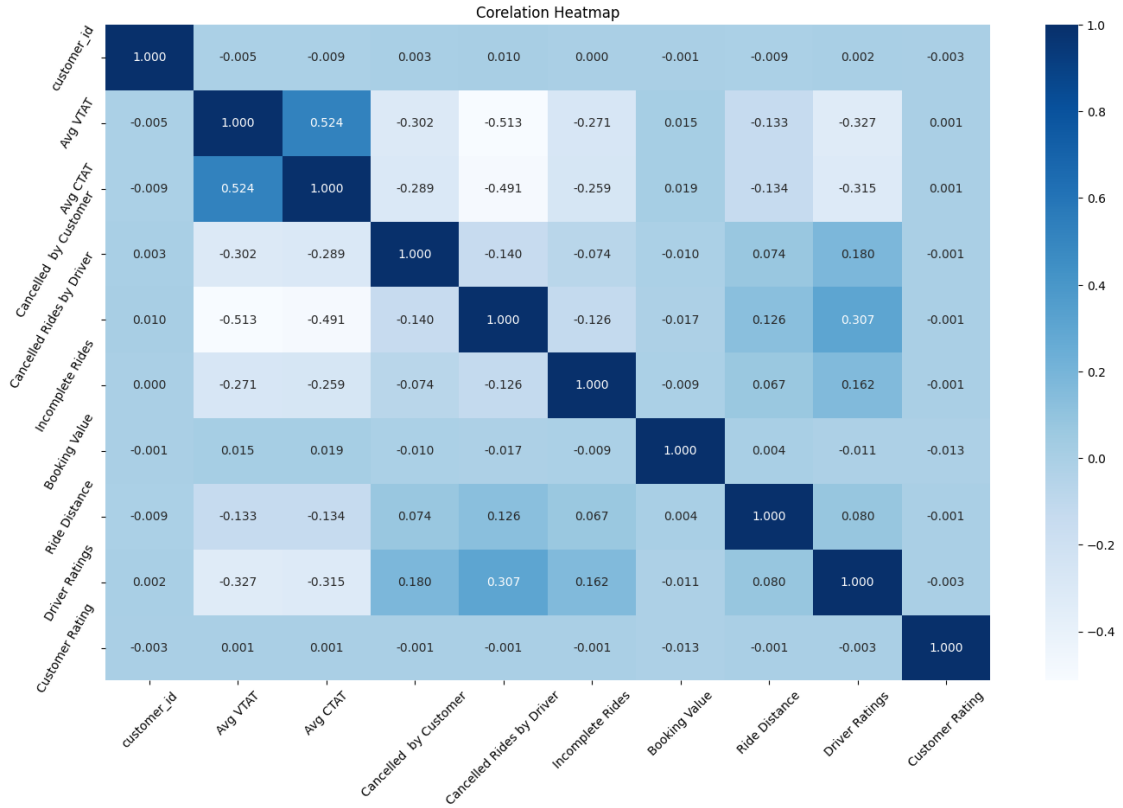
```
       'Driver Ratings', 'Customer Rating'],
      dtype='object')
```

[107]:
```
df['Avg CTAT'].count()
```

[107]:
```
np.int64(49999)
```

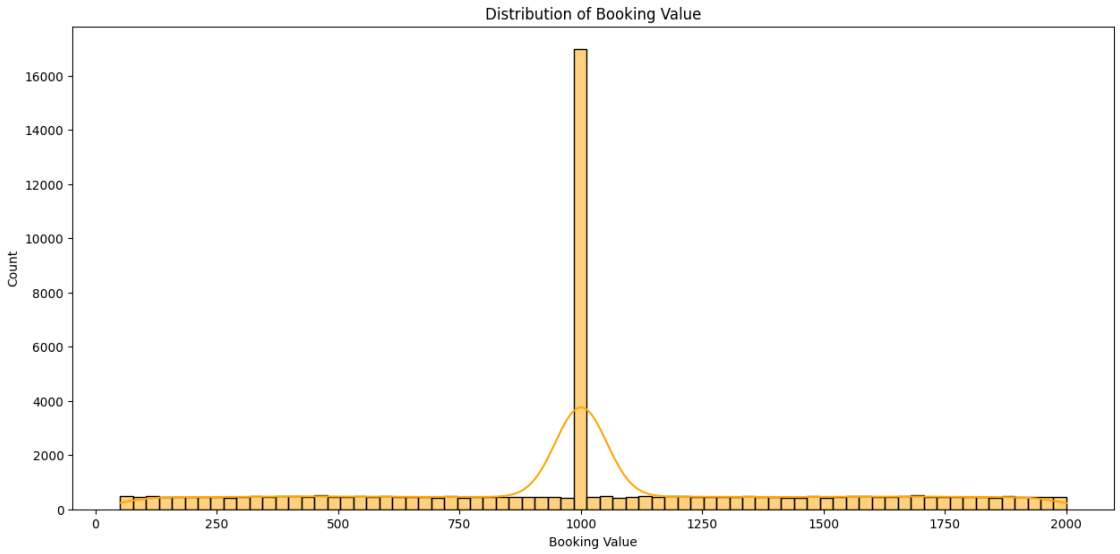
[ ]:
```
##Corelation␣
↪Heatmap==================##############################==========================############
```

[134]:
```
plt.figure(figsize=(15,10))
sb.heatmap(df.corr(numeric_only=True),annot=True,cmap='Blues',fmt=".3f")
plt.title('Corelation Heatmap')
plt.tight_layout()
plt.xticks(rotation=45)
plt.yticks(rotation=60)
plt.show()
```



[ ]:
```
####################=====================Distribuation plot for Numerical␣
↪Column=======================########################
```

[115]:
```
if 'Booking Value' in df.columns:
    plt.figure(figsize=(15,7))
    sb.histplot(df["Booking Value"],kde=True,color='orange')
    plt.title("Distribution of Booking Value")
    plt.show()
```



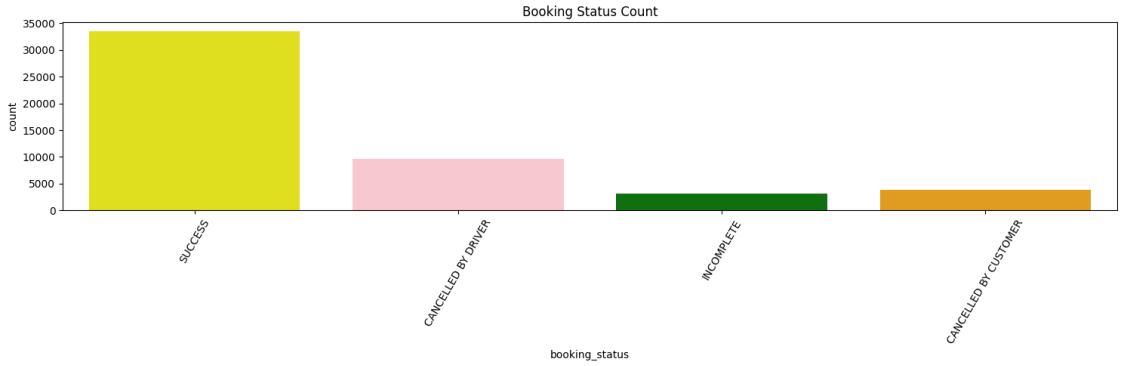
[ ]:
```
# Countplot for Categorical␣
↪Columns########################################==========================================
```

[116]:
```
df.columns
```

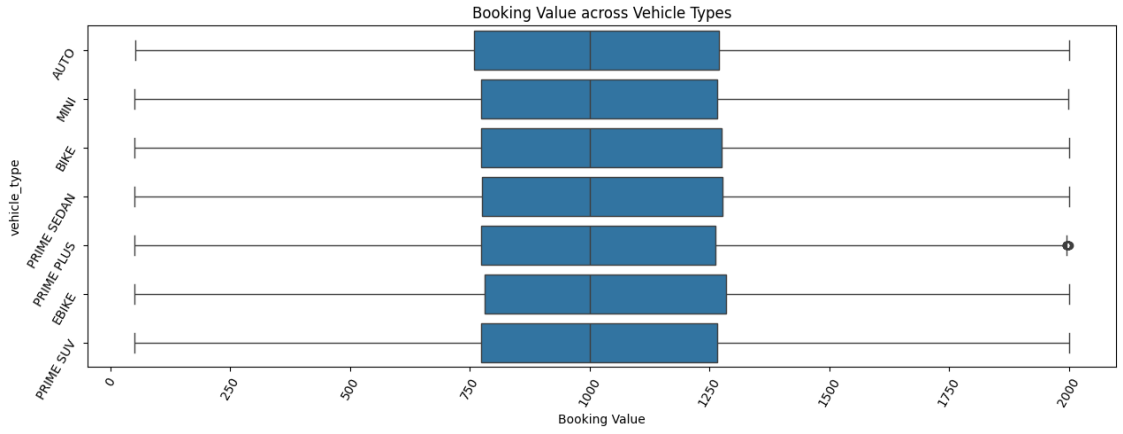
[116]:
```
Index(['Date', 'Time', 'booking_id', 'booking_status', 'customer_id',
       'vehicle_type', 'pickup_location', 'drop_location', 'Avg VTAT',
       'Avg CTAT', 'Cancelled  by Customer',
       'Reason for Cancelling by Customer', 'Cancelled Rides by Driver',
       'Reason for Cancelling by Driver', 'Incomplete Rides',
       'Incomplete Rides Reason', 'Booking Value', 'Ride Distance',
       'Driver Ratings', 'Customer Rating'],
      dtype='object')
```

[121]:
```
if 'booking_status' in df.columns:
    plt.figure(figsize=(15,5))
    colors=('yellow','pink','green','orange')
    sb.countplot(data=df,x='booking_status',palette=colors)
    plt.title("Booking Status Count")
    plt.xticks(rotation=60)
    plt.tight_layout()
    plt.show()
```

```
C:\Users\admin\AppData\Local\Temp\ipykernel_13040\1021343052.py:4:
FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in
v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same
effect.

  sb.countplot(data=df,x='booking_status',palette=colors)
```



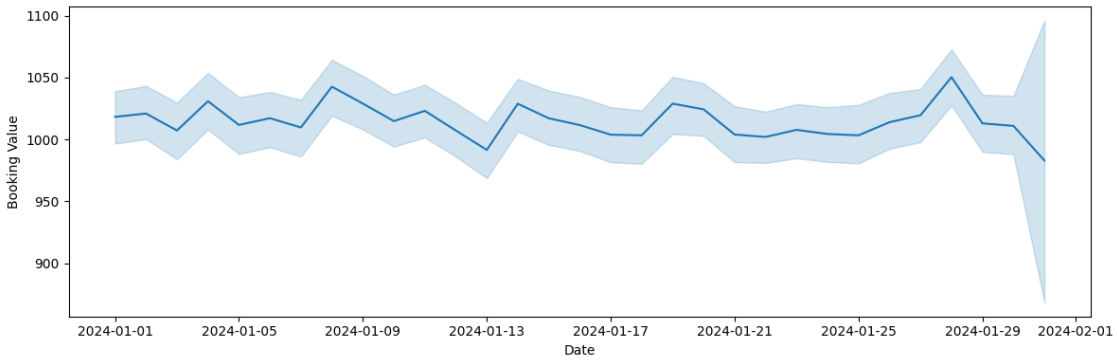
[124]:
```
if 'Booking Value' in df.columns and 'vehicle_type' in df.columns:
    plt.figure(figsize=(15,5))
    sb.boxplot(data=df,x='Booking Value',y='vehicle_type')
    plt.title("Booking Value across Vehicle Types")
    plt.xticks(rotation=60)
    plt.yticks(rotation=60)
    plt.show()
```



[126]:
```
df.dtypes
```

[126]:
```
Date                                  datetime64[ns]
Time                                          object
booking_id                                    object
booking_status                                object
customer_id                                    int64
vehicle_type                                  object
pickup_location                               object
drop_location                                 object
Avg VTAT                                     float64
Avg CTAT                                     float64
Cancelled  by Customer                         int64
Reason for Cancelling by Customer             object
Cancelled Rides by Driver                      int64
Reason for Cancelling by Driver               object
Incomplete Rides                               int64
Incomplete Rides Reason                       object
Booking Value                                float64
Ride Distance                                float64
Driver Ratings                               float64
Customer Rating                              float64
dtype: object
```

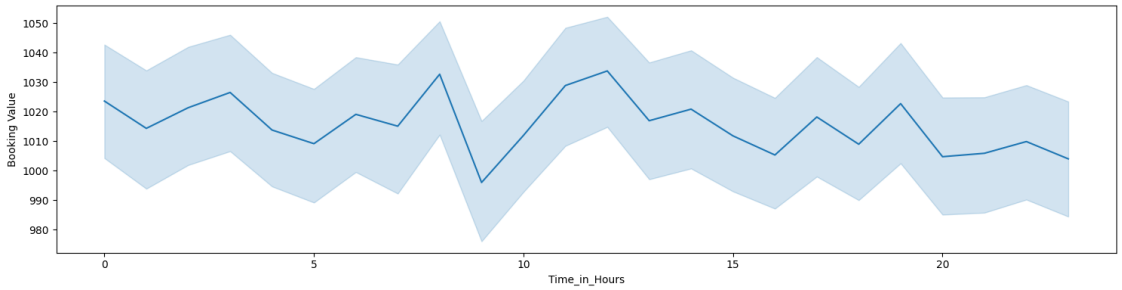
[135]:
```
if 'Date' in df.columns and 'Booking Value' in df.columns:
    plt.figure(figsize=(12,4))
    sb.lineplot(data=df,x='Date',y='Booking Value')
    plt.tight_layout()
    plt.show()
```



[144]:
```
df['Time'] = pd.to_datetime(df['Time'], format='%H:%M:%S').dt.time
```

[146]:
```
df['Time_in_Hours'] = pd.to_datetime(df['Time'], format='%H:%M:%S',␣
 ↪errors='coerce').dt.hour + \
                      pd.to_datetime(df['Time'], format='%H:%M:%S',␣
 ↪errors='coerce').dt.minute / 60 + \
                      pd.to_datetime(df['Time'], format='%H:%M:%S',␣
 ↪errors='coerce').dt.second / 3600
```

[148]:
```
if 'Time_in_Hours' in df.columns and 'Booking Value' in df.columns:
    plt.figure(figsize=(15,4))
    sb.lineplot(data=df,x='Time_in_Hours',y='Booking Value')
    plt.tight_layout()
    plt.show()
```



[ ]:
```
###Revenue␣
 ↪analysis==========================================================================================
```

[151]:
```
revenue=df['Booking Value'].sum()
print("Total Revenue:",round(revenue,0))
```

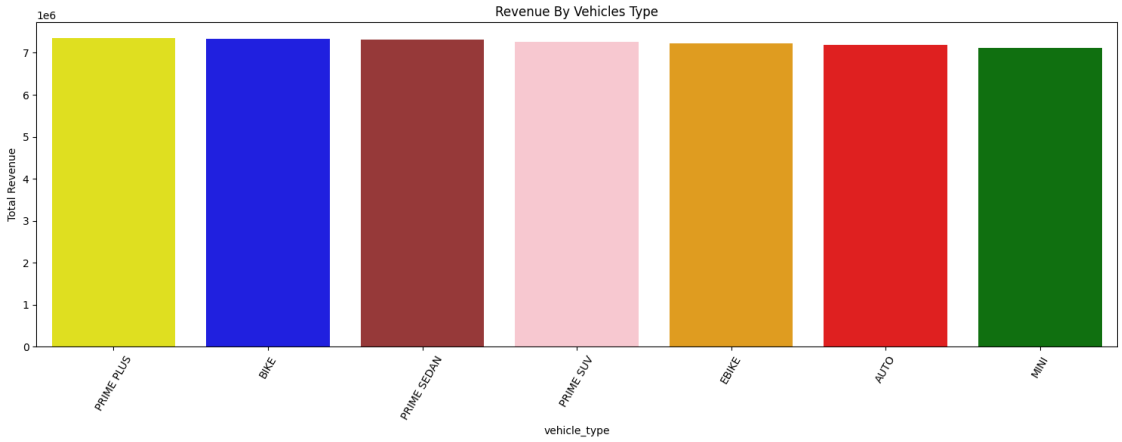
```
Total Revenue: 50781665.0
```

[152]:
```
revenue_by_vehicles=df.groupby('vehicle_type')['Booking Value'].sum().
 ↪sort_values(ascending=False)
print("Revenue By Vehicles:",round(revenue_by_vehicles,0))
```

```
Revenue By Vehicles: vehicle_type
PRIME PLUS     7354975.0
BIKE           7327791.0
PRIME SEDAN    7308409.0
PRIME SUV      7257967.0
EBIKE          7227694.0
AUTO           7180613.0
MINI           7124217.0
Name: Booking Value, dtype: float64
```

[163]:
```
plt.figure(figsize=(15,6))
colors=('yellow','blue','brown','pink','orange','red','green')
sb.barplot(x=revenue_by_vehicles.index,y=revenue_by_vehicles.
 ↪values,palette=colors)
plt.title("Revenue By Vehicles Type")
plt.ylabel("Total Revenue")
plt.xticks(rotation=60)
plt.tight_layout()
plt.show()
```

```
C:\Users\admin\AppData\Local\Temp\ipykernel_13040\2079756544.py:3:
FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in
v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same
effect.

  sb.barplot(x=revenue_by_vehicles.index,y=revenue_by_vehicles.values,palette=co
lors)
```



[164]:
```
revenue_by_status=df.groupby('booking_status')['Booking Value'].sum().
 ↪sort_values(ascending=False)
```

[168]:
```
print("Revenue by Booking Status:",round(revenue_by_status,0))
```

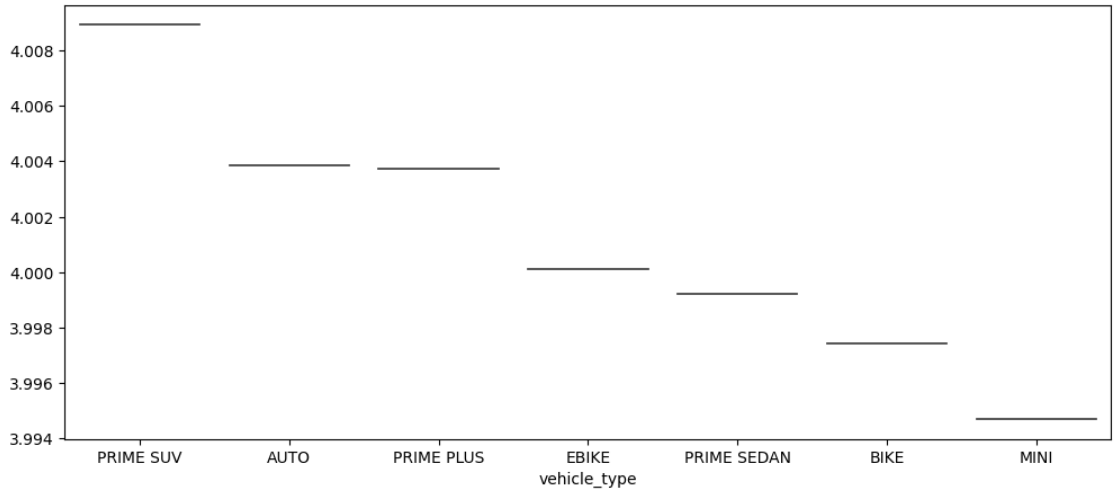
```
Revenue by Booking Status: booking_status
SUCCESS                   34266665.0
CANCELLED BY DRIVER        9610000.0
CANCELLED BY CUSTOMER      3799000.0
INCOMPLETE                 3106000.0
Name: Booking Value, dtype: float64
```

[171]:
```
rating_by_vehicle=df.groupby('vehicle_type')['Customer Rating'].mean().
  ↪sort_values(ascending=False)
print('Average rating by vehicle type:',round(rating_by_vehicle,0))
```

```
Average rating by vehicle type: vehicle_type
PRIME SUV      4.0
AUTO           4.0
PRIME PLUS     4.0
EBIKE          4.0
PRIME SEDAN    4.0
BIKE           4.0
MINI           4.0
Name: Customer Rating, dtype: float64
```

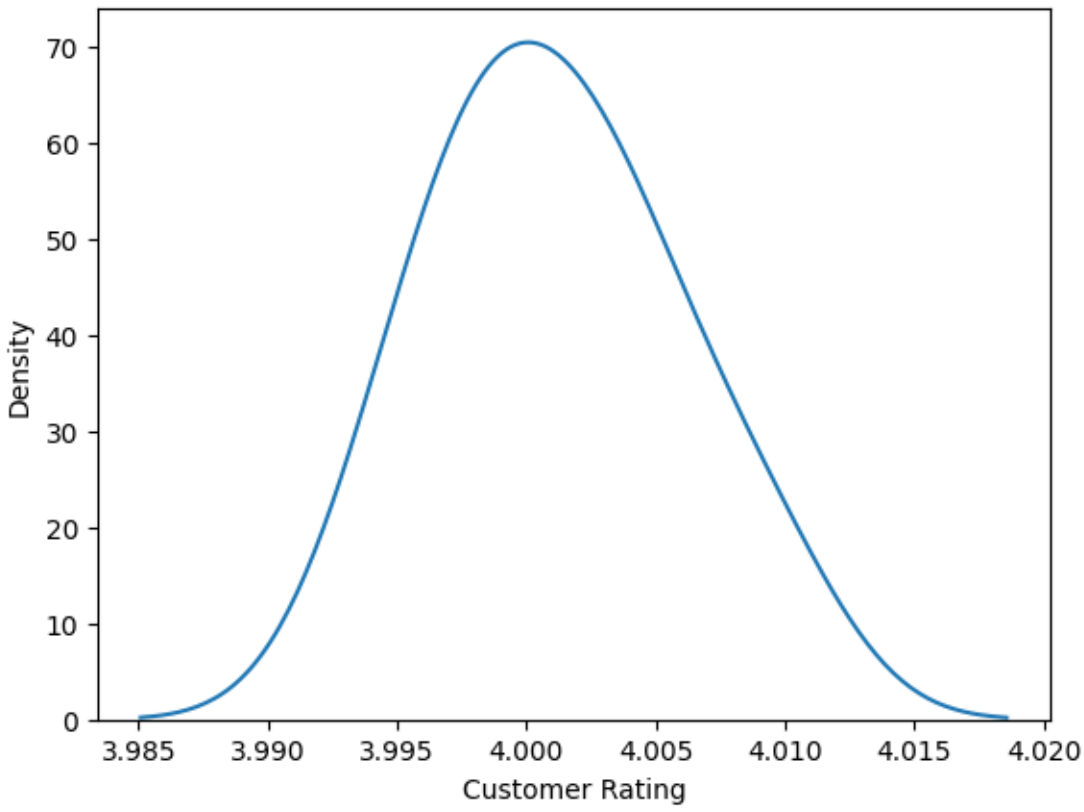
[175]:
```
plt.figure(figsize=(12,5))
sb.violinplot(x=rating_by_vehicle.index,y=rating_by_vehicle.values)
```

[175]: <Axes: xlabel='vehicle_type'>



[176]:
```
sb.kdeplot(rating_by_vehicle)
```

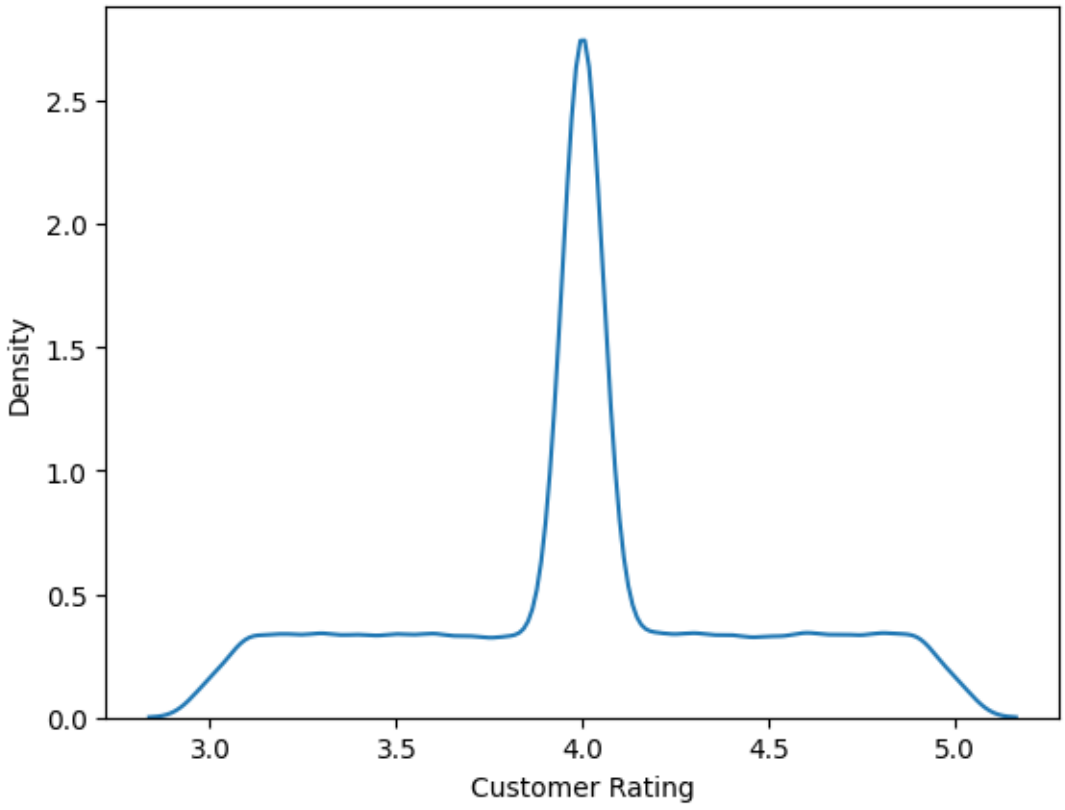
[176]: <Axes: xlabel='Customer Rating', ylabel='Density'>



[ ]:

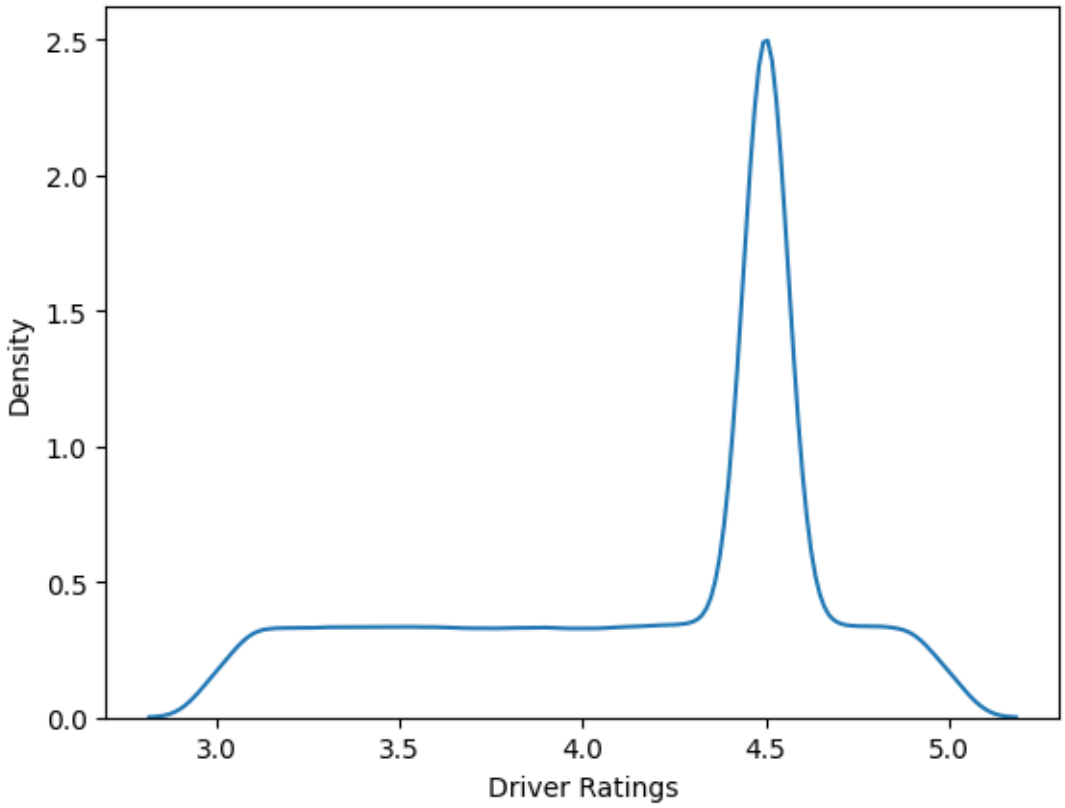
[177]:
```
sb.kdeplot(df['Customer Rating'])
```

[177]:
```
<Axes: xlabel='Customer Rating', ylabel='Density'>
```



[178]:
```
sb.kdeplot(df['Driver Ratings'])
```

[178]: <Axes: xlabel='Driver Ratings', ylabel='Density'>


2.5
2.0
1.5
1.0
0.5
0.0
Density
3.0
3.5
4.0
4.5
5.0
Driver Ratings

[ ]:

[ ]: