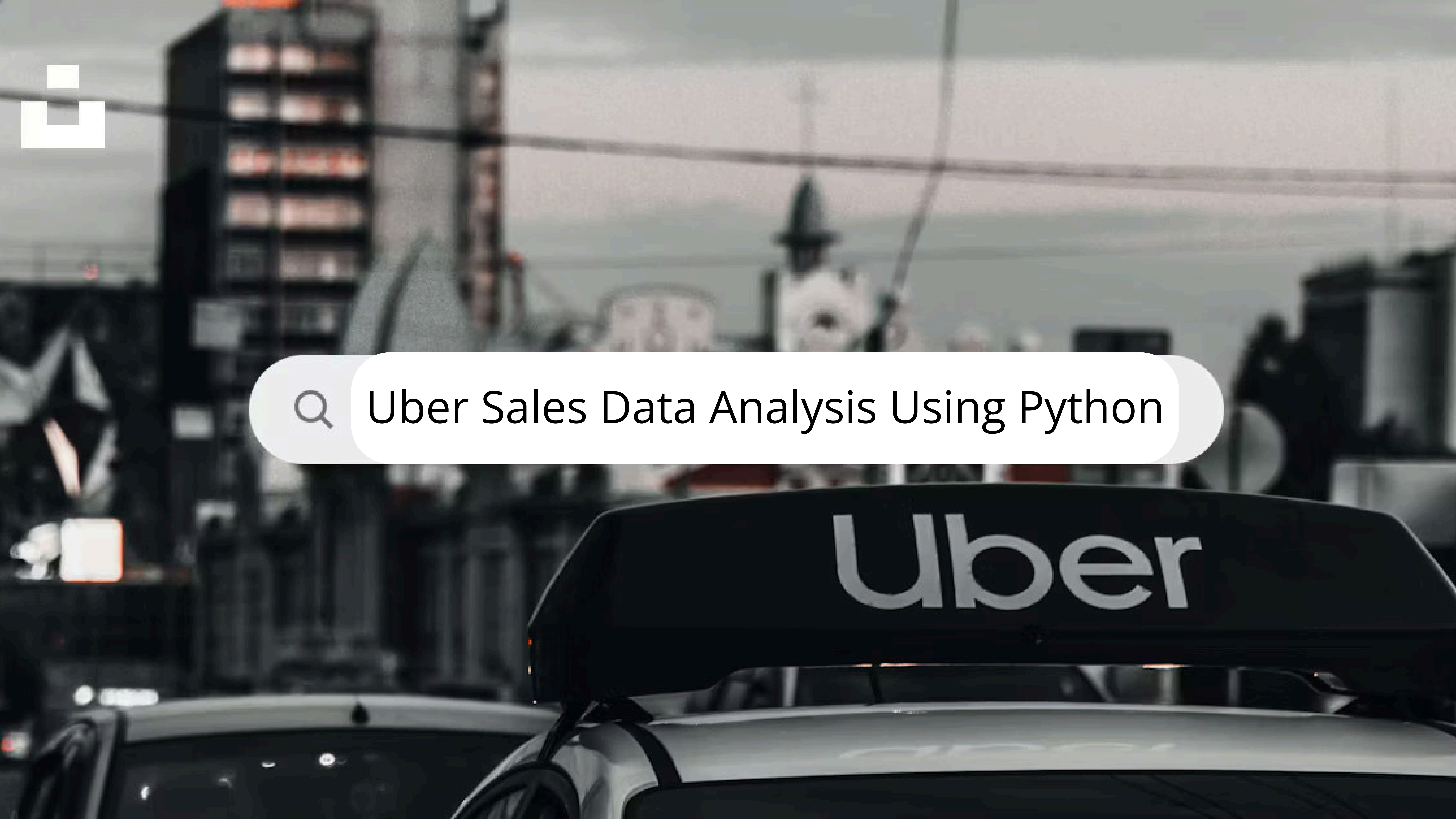
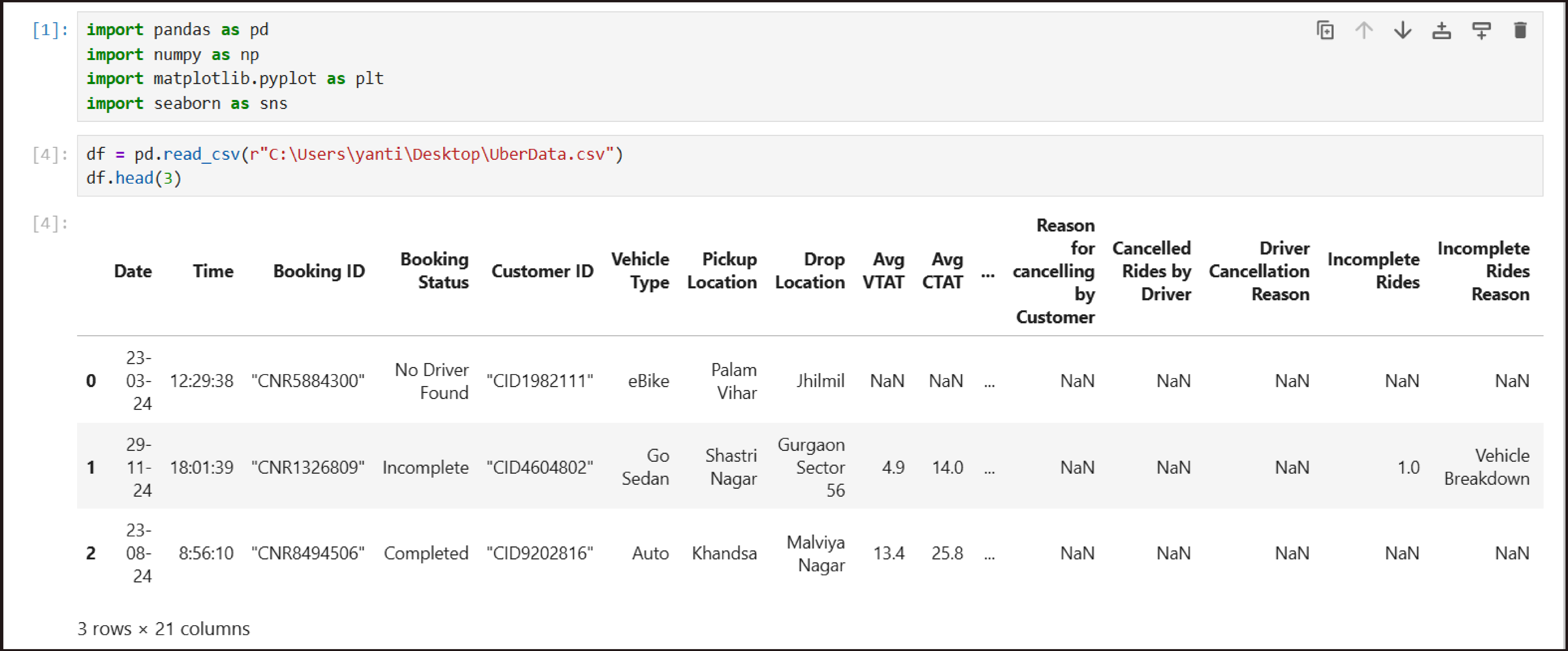


Uber Sales Data Analysis Using Python

Overview



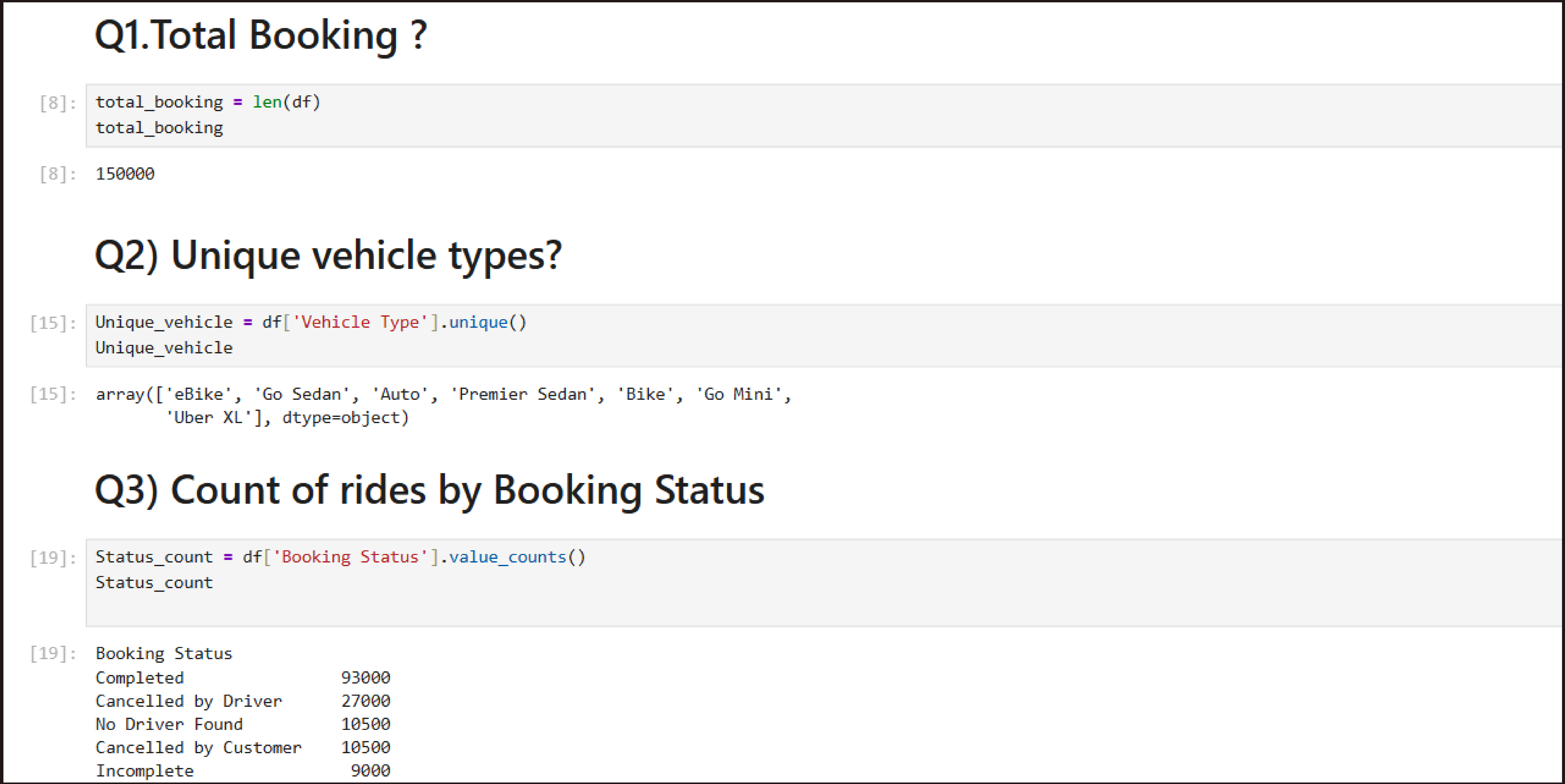
[1]:
```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

[4]:
```python
df = pd.read_csv(r"C:\Users\yanti\Desktop\UberData.csv")
df.head(3)
```

[4]:

	Date	Time	Booking ID	Booking Status	Customer ID	Vehicle Type	Pickup Location	Drop Location	Avg VTAT	Avg CTAT	...	Reason for cancelling by Customer	Cancelled Rides by Driver	Driver Cancellation Reason	Incomplete Rides	Incomplete Rides Reason
0	23-03-24	12:29:38	"CNR5884300"	No Driver Found	"CID1982111"	eBike	Palam Vihar	Jhilmil	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN
1	29-11-24	18:01:39	"CNR1326809"	Incomplete	"CID4604802"	Go Sedan	Shastri Nagar	Gurgaon Sector 56	4.9	14.0	...	NaN	NaN	NaN	1.0	Vehicle Breakdown
2	23-08-24	8:56:10	"CNR8494506"	Completed	"CID9202816"	Auto	Khandsa	Malviya Nagar	13.4	25.8	...	NaN	NaN	NaN	NaN	NaN

3 rows × 21 columns



Q1.Total Booking ?

```
[8]: total_booking = len(df)
     total_booking
```

```
[8]: 150000
```

Q2) Unique vehicle types?

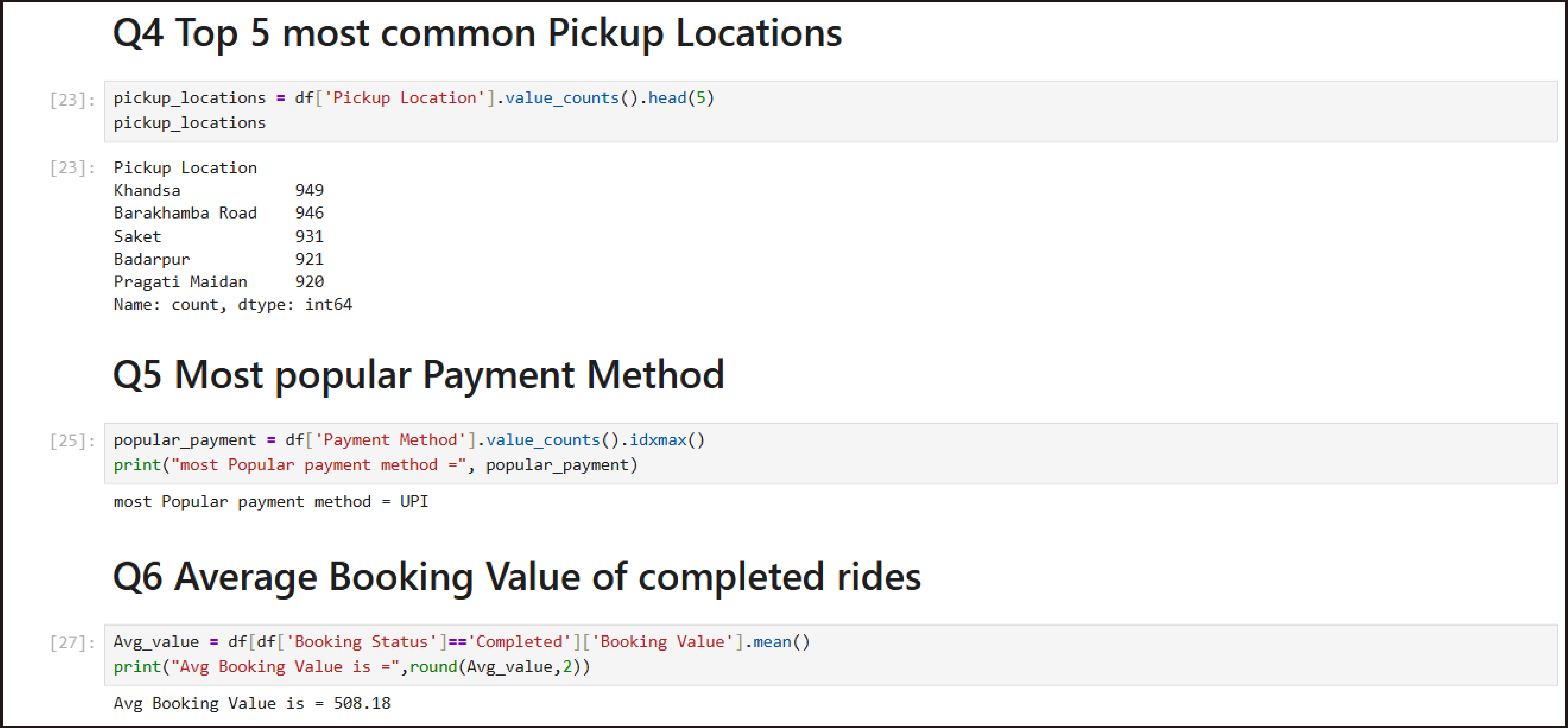
```
[15]: Unique_vehicle = df['Vehicle Type'].unique()
      Unique_vehicle
```

```
[15]: array(['eBike', 'Go Sedan', 'Auto', 'Premier Sedan', 'Bike', 'Go Mini',
             'Uber XL'], dtype=object)
```

Q3) Count of rides by Booking Status

```
[19]: Status_count = df['Booking Status'].value_counts()
      Status_count
```

```
[19]: Booking Status
      Completed                93000
      Cancelled by Driver      27000
      No Driver Found          10500
      Cancelled by Customer    10500
      Incomplete                9000
```



Q4 Top 5 most common Pickup Locations

```
[23]: pickup_locations = df['Pickup Location'].value_counts().head(5)
      pickup_locations
```

```
[23]: Pickup Location
      Khandsa            949
      Barakhamba Road    946
      Saket              931
      Badarpur           921
      Pragati Maidan     920
      Name: count, dtype: int64
```

Q5 Most popular Payment Method

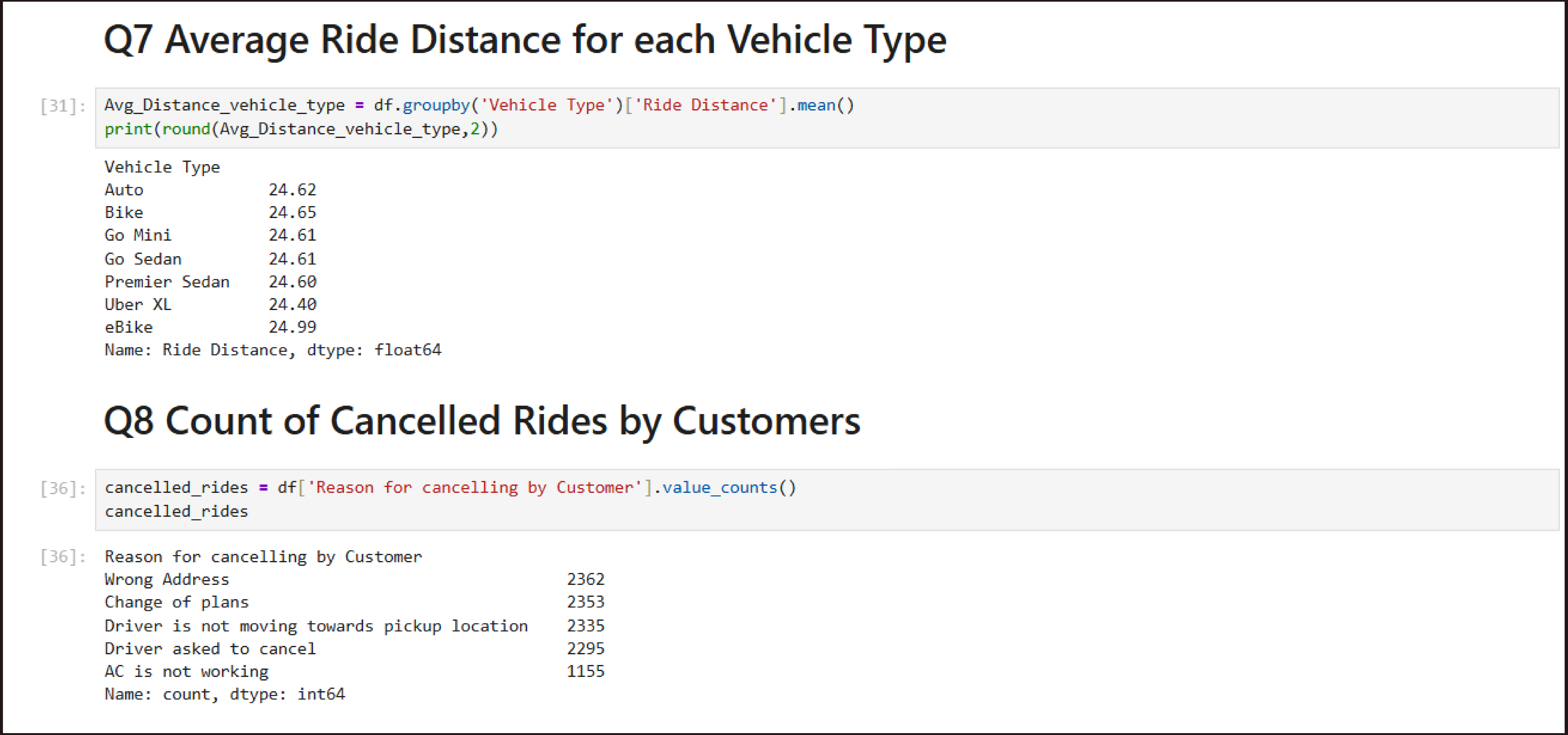
```
[25]: popular_payment = df['Payment Method'].value_counts().idxmax()
      print("most Popular payment method =", popular_payment)
```

```
most Popular payment method = UPI
```

Q6 Average Booking Value of completed rides

```
[27]: Avg_value = df[df['Booking Status']=='Completed']['Booking Value'].mean()
      print("Avg Booking Value is =",round(Avg_value,2))
```

```
Avg Booking Value is = 508.18
```



Q7 Average Ride Distance for each Vehicle Type

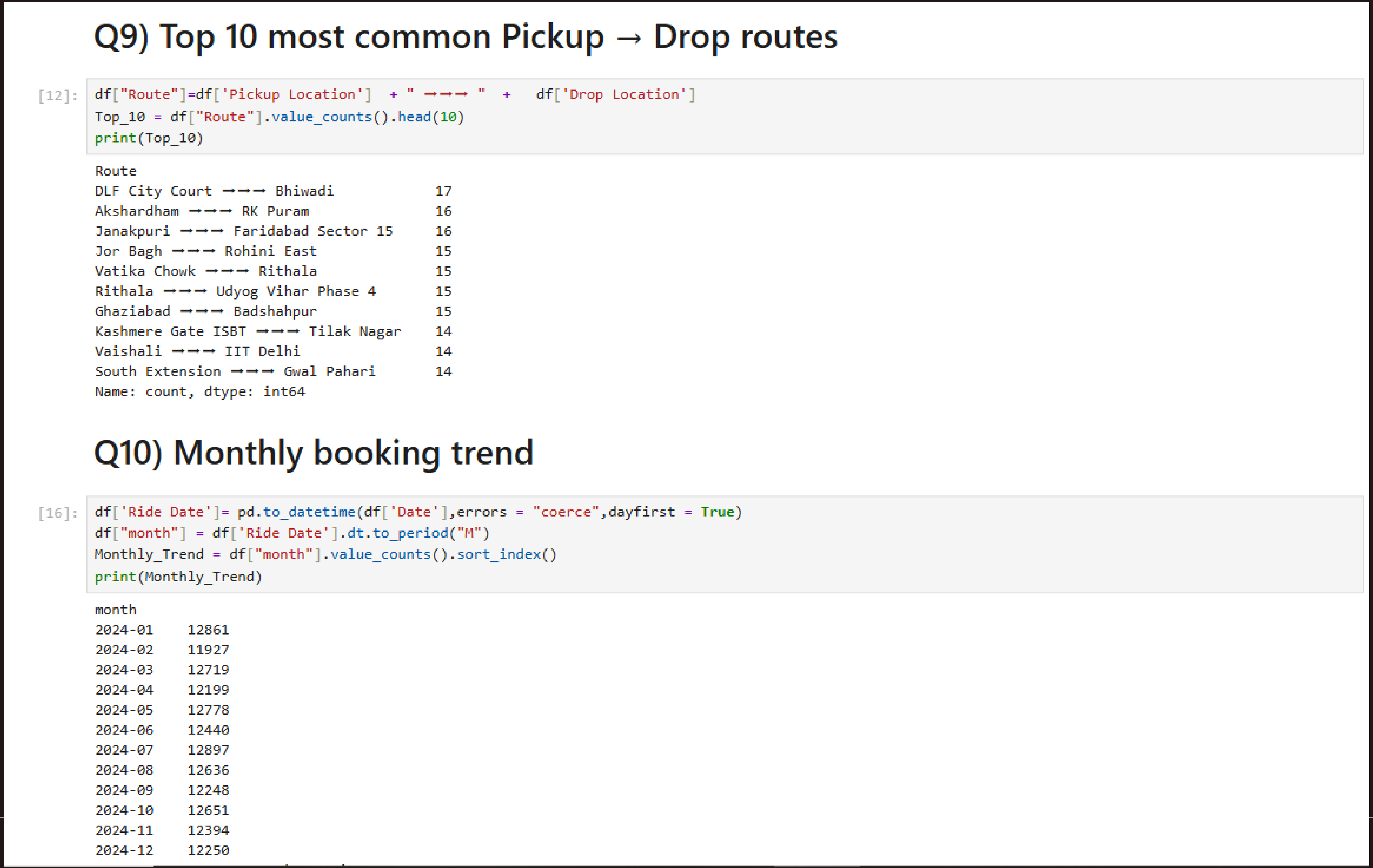
```
[31]: Avg_Distance_vehicle_type = df.groupby('Vehicle Type')['Ride Distance'].mean()
      print(round(Avg_Distance_vehicle_type,2))
```

```
Vehicle Type
Auto             24.62
Bike             24.65
Go Mini          24.61
Go Sedan         24.61
Premier Sedan    24.60
Uber XL          24.40
eBike            24.99
Name: Ride Distance, dtype: float64
```

Q8 Count of Cancelled Rides by Customers

```
[36]: cancelled_rides = df['Reason for cancelling by Customer'].value_counts()
      cancelled_rides
```

```
[36]: Reason for cancelling by Customer
      Wrong Address                               2362
      Change of plans                             2353
      Driver is not moving towards pickup location    2335
      Driver asked to cancel                      2295
      AC is not working                           1155
      Name: count, dtype: int64
```



Q9) Top 10 most common Pickup → Drop routes

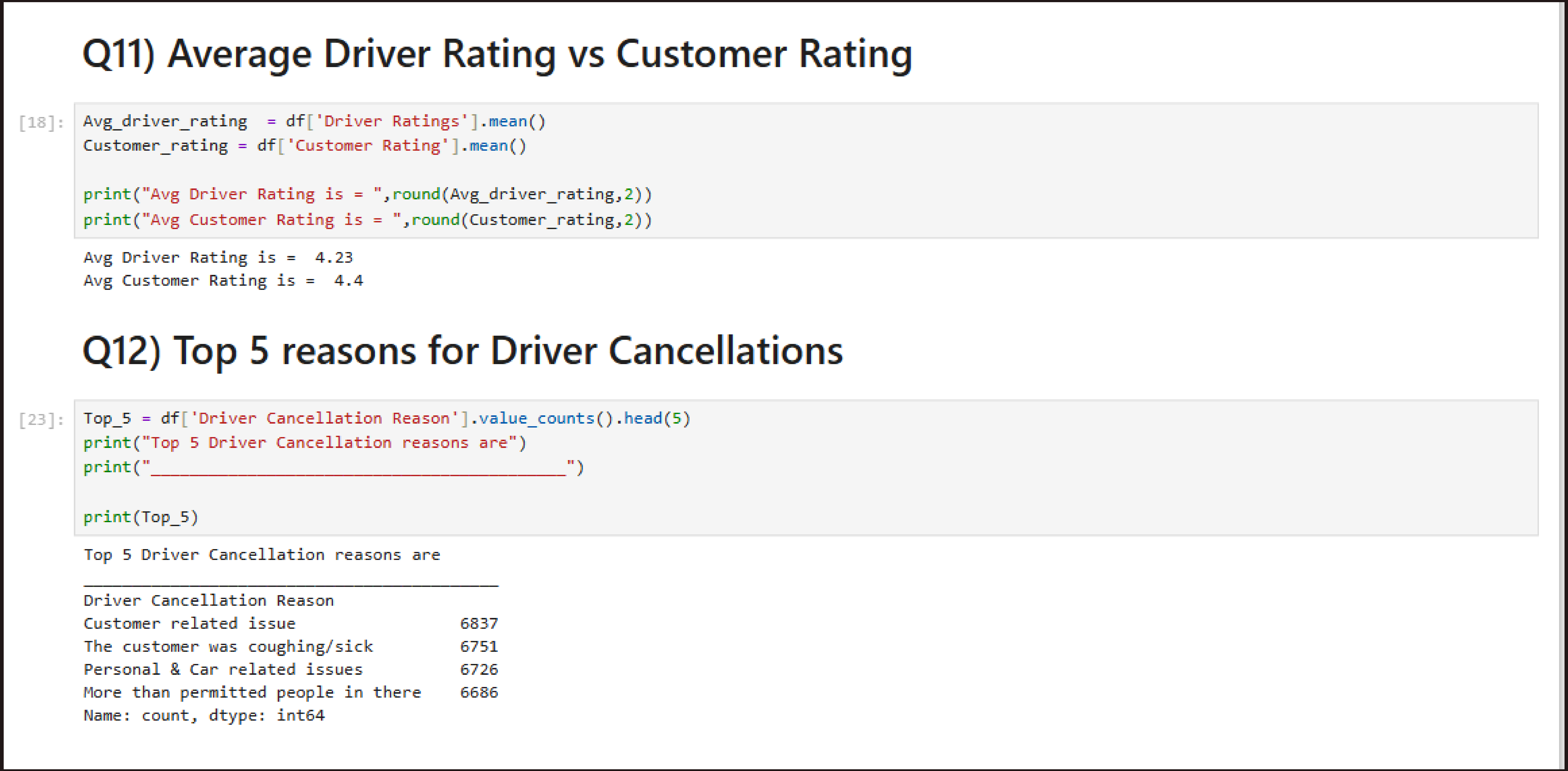
```
[12]: df["Route"]=df['Pickup Location']  + " →→→ "  +   df['Drop Location']
      Top_10 = df["Route"].value_counts().head(10)
      print(Top_10)
```

```
Route
DLF City Court →→→ Bhiwadi            17
Akshardham →→→ RK Puram               16
Janakpuri →→→ Faridabad Sector 15     16
Jor Bagh →→→ Rohini East              15
Vatika Chowk →→→ Rithala              15
Rithala →→→ Udyog Vihar Phase 4       15
Ghaziabad →→→ Badshahpur              15
Kashmere Gate ISBT →→→ Tilak Nagar    14
Vaishali →→→ IIT Delhi                14
South Extension →→→ Gwal Pahari       14
Name: count, dtype: int64
```

Q10) Monthly booking trend

```
[16]: df['Ride Date']= pd.to_datetime(df['Date'],errors = "coerce",dayfirst = True)
      df["month"] = df['Ride Date'].dt.to_period("M")
      Monthly_Trend = df["month"].value_counts().sort_index()
      print(Monthly_Trend)
```

```
month
2024-01    12861
2024-02    11927
2024-03    12719
2024-04    12199
2024-05    12778
2024-06    12440
2024-07    12897
2024-08    12636
2024-09    12248
2024-10    12651
2024-11    12394
2024-12    12250
```



Q11) Average Driver Rating vs Customer Rating

[18]:

```
Avg_driver_rating  = df['Driver Ratings'].mean()
Customer_rating = df['Customer Rating'].mean()

print("Avg Driver Rating is = ",round(Avg_driver_rating,2))
print("Avg Customer Rating is = ",round(Customer_rating,2))
```

```
Avg Driver Rating is =  4.23
Avg Customer Rating is =  4.4
```

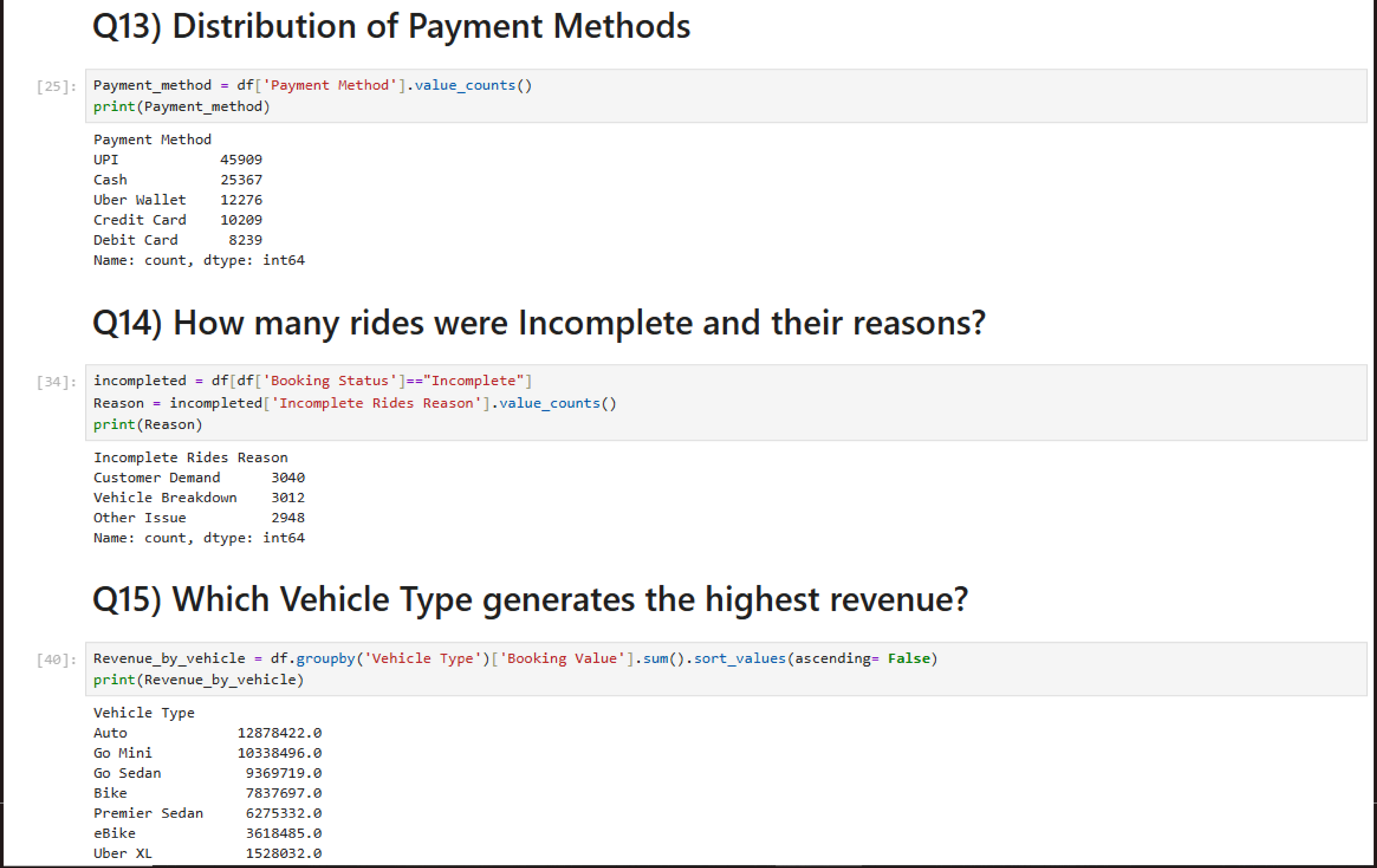
Q12) Top 5 reasons for Driver Cancellations

[23]:

```
Top_5 = df['Driver Cancellation Reason'].value_counts().head(5)
print("Top 5 Driver Cancellation reasons are")
print("___________________________________________")

print(Top_5)
```

```
Top 5 Driver Cancellation reasons are
___________________________________________
Driver Cancellation Reason
Customer related issue                 6837
The customer was coughing/sick         6751
Personal & Car related issues          6726
More than permitted people in there    6686
Name: count, dtype: int64
```



Q13) Distribution of Payment Methods

[25]:

```
Payment_method = df['Payment Method'].value_counts()
print(Payment_method)
```

```
Payment Method
UPI            45909
Cash           25367
Uber Wallet    12276
Credit Card    10209
Debit Card      8239
Name: count, dtype: int64
```

Q14) How many rides were Incomplete and their reasons?

[34]:

```
incompleted = df[df['Booking Status']=="Incomplete"]
Reason = incompleted['Incomplete Rides Reason'].value_counts()
print(Reason)
```

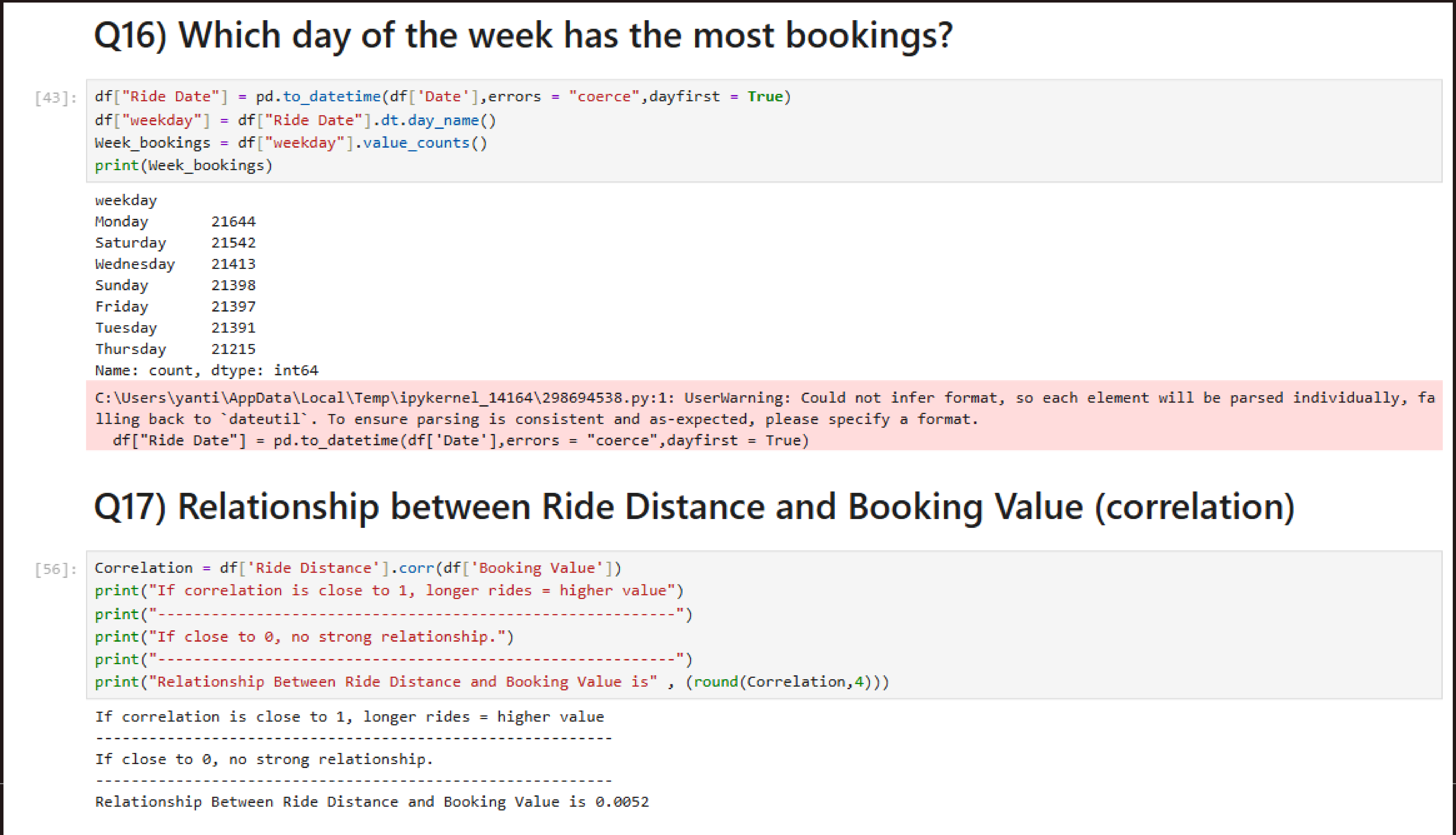
```
Incomplete Rides Reason
Customer Demand      3040
Vehicle Breakdown    3012
Other Issue          2948
Name: count, dtype: int64
```

Q15) Which Vehicle Type generates the highest revenue?

[40]:

```
Revenue_by_vehicle = df.groupby('Vehicle Type')['Booking Value'].sum().sort_values(ascending= False)
print(Revenue_by_vehicle)
```

```
Vehicle Type
Auto             12878422.0
Go Mini          10338496.0
Go Sedan          9369719.0
Bike              7837697.0
Premier Sedan     6275332.0
eBike             3618485.0
Uber XL           1528032.0
```



Q16) Which day of the week has the most bookings?

[43]:

```python
df["Ride Date"] = pd.to_datetime(df['Date'],errors = "coerce",dayfirst = True)
df["weekday"] = df["Ride Date"].dt.day_name()
Week_bookings = df["weekday"].value_counts()
print(Week_bookings)
```

```
weekday
Monday       21644
Saturday     21542
Wednesday    21413
Sunday       21398
Friday       21397
Tuesday      21391
Thursday     21215
Name: count, dtype: int64
```

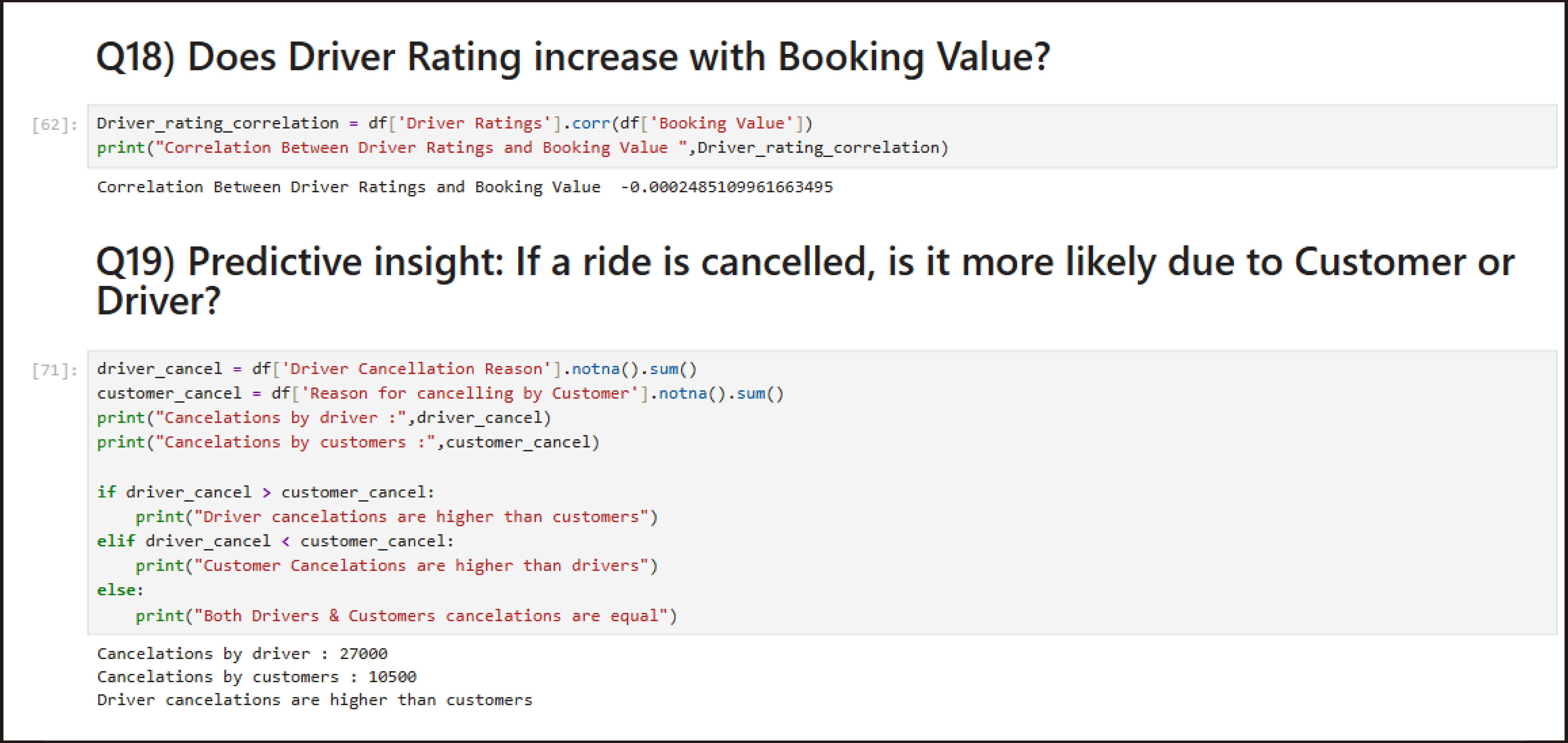
```
C:\Users\yanti\AppData\Local\Temp\ipykernel_14164\298694538.py:1: UserWarning: Could not infer format, so each element will be parsed individually, fa
lling back to `dateutil`. To ensure parsing is consistent and as-expected, please specify a format.
  df["Ride Date"] = pd.to_datetime(df['Date'],errors = "coerce",dayfirst = True)
```

Q17) Relationship between Ride Distance and Booking Value (correlation)

[56]:

```python
Correlation = df['Ride Distance'].corr(df['Booking Value'])
print("If correlation is close to 1, longer rides = higher value")
print("----------------------------------------------------------")
print("If close to 0, no strong relationship.")
print("----------------------------------------------------------")
print("Relationship Between Ride Distance and Booking Value is" , (round(Correlation,4)))
```

```
If correlation is close to 1, longer rides = higher value
----------------------------------------------------------
If close to 0, no strong relationship.
----------------------------------------------------------
Relationship Between Ride Distance and Booking Value is 0.0052
```



Q18) Does Driver Rating increase with Booking Value?

[62]:

```
Driver_rating_correlation = df['Driver Ratings'].corr(df['Booking Value'])
print("Correlation Between Driver Ratings and Booking Value ",Driver_rating_correlation)
```

```
Correlation Between Driver Ratings and Booking Value  -0.0002485109961663495
```

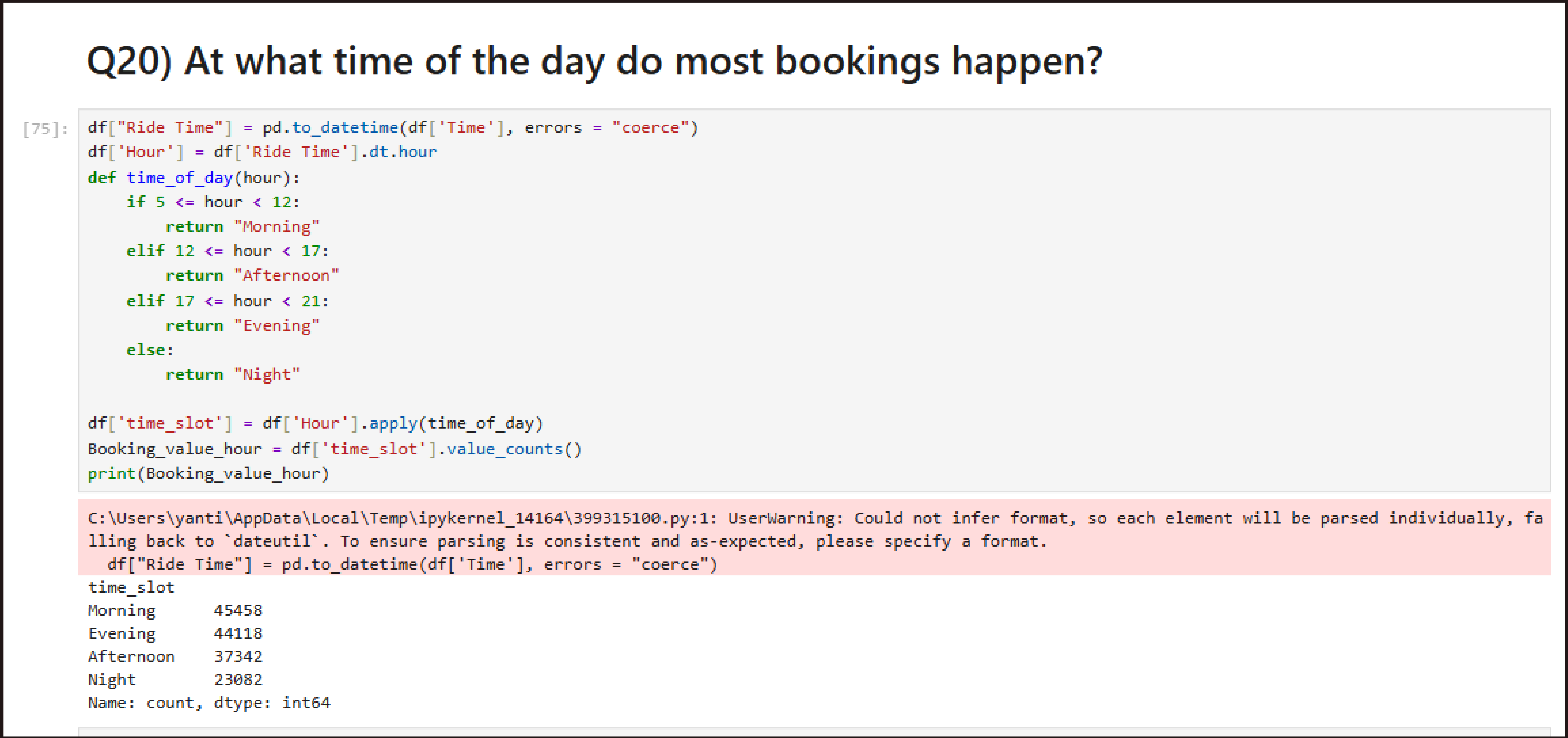
Q19) Predictive insight: If a ride is cancelled, is it more likely due to Customer or Driver?

[71]:

```
driver_cancel = df['Driver Cancellation Reason'].notna().sum()
customer_cancel = df['Reason for cancelling by Customer'].notna().sum()
print("Cancelations by driver :",driver_cancel)
print("Cancelations by customers :",customer_cancel)

if driver_cancel > customer_cancel:
    print("Driver cancelations are higher than customers")
elif driver_cancel < customer_cancel:
    print("Customer Cancelations are higher than drivers")
else:
    print("Both Drivers & Customers cancelations are equal")
```

```
Cancelations by driver : 27000
Cancelations by customers : 10500
Driver cancelations are higher than customers
```



Q20) At what time of the day do most bookings happen?

```
[75]: df["Ride Time"] = pd.to_datetime(df['Time'], errors = "coerce")
      df['Hour'] = df['Ride Time'].dt.hour
      def time_of_day(hour):
          if 5 <= hour < 12:
              return "Morning"
          elif 12 <= hour < 17:
              return "Afternoon"
          elif 17 <= hour < 21:
              return "Evening"
          else:
              return "Night"

      df['time_slot'] = df['Hour'].apply(time_of_day)
      Booking_value_hour = df['time_slot'].value_counts()
      print(Booking_value_hour)
```

```
C:\Users\yanti\AppData\Local\Temp\ipykernel_14164\399315100.py:1: UserWarning: Could not infer format, so each element will be parsed individually, fa
lling back to `dateutil`. To ensure parsing is consistent and as-expected, please specify a format.
  df["Ride Time"] = pd.to_datetime(df['Time'], errors = "coerce")
time_slot
Morning      45458
Evening      44118
Afternoon    37342
Night        23082
Name: count, dtype: int64
```

THANK YOU

Follow For More

Presented By:

Jatin Sharma