Zamato dataset

```
import mysql.connector
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings("ignore")

#pip install PyMySQL
df=pd.read_csv(r"C:\Users\Viren Rajendra Gamre\OneDrive\Desktop\zomato
delivery operations dataset\archive\Zomato Dataset.csv")

#import pymysql
#import pandas as pd
#
## Create a connection using PyMySQL
#connection = pymysql.connect(
#    user='root',
#    password='12345',
#    host='localhost',
#    database='zomato_operations'
#)
#
## Query your data
#query = 'SELECT * FROM z'
#df = pd.read_sql(query, connection)
#
## Close the connection
#connection.close()
#
#print(df.head())

df
```

```
           ID Delivery_person_ID  Delivery_person_Age  \
0      0xcdcd      DEHRES17DEL01                 36.0
1      0xd987      KOCRES16DEL01                 21.0
2      0x2784     PUNERES13DEL03                 23.0
3      0xc8b6     LUDHRES15DEL02                 34.0
4      0xdb64      KNPRES14DEL02                 24.0
...       ...                ...                  ...
45579  0x7c09      JAPRES04DEL01                 30.0
45580  0xd641      AGRRES16DEL01                 21.0
45581  0x4f8d     CHENRES08DEL03                 30.0
45582  0x5eee    COIMBRES11DEL01                 20.0
45583  0x5fb2   RANCHIRES09DEL02                 23.0
```

```
       Delivery_person_Ratings  Restaurant_latitude
Restaurant_longitude  \
0                          4.2            30.327968
78.046106
1                          4.7            10.003064
76.307589
2                          4.7            18.562450
73.916619
3                          4.3            30.899584
75.809346
4                          4.7            26.463504
80.372929
...                        ...                  ...
...
45579                      4.8            26.902328
75.794257
45580                      4.6             0.000000
0.000000
45581                      4.9            13.022394
80.242439
45582                      4.7            11.001753
76.986241
45583                      4.9            23.351058
85.325731

       Delivery_location_latitude  Delivery_location_longitude
Order_Date  \
0                       30.397968                    78.116106  12-02-
2022
1                       10.043064                    76.347589  13-02-
2022
2                       18.652450                    74.006619  04-03-
2022
3                       30.919584                    75.829346  13-02-
2022
4                       26.593504                    80.502929  14-02-
2022
...                           ...                          ...
...
45579                   26.912328                    75.804257  24-03-
2022
45580                    0.070000                     0.070000  16-02-
2022
45581                   13.052394                    80.272439  11-03-
2022
45582                   11.041753                    77.026241  07-03-
2022
45583                   23.431058                    85.405731  02-03-
```

```
2022

      Time_Orderd Time_Order_picked Weather_conditions
Road_traffic_density  \
0           21:55              22:10                Fog
Jam
1           14:55              15:05             Stormy
High
2           17:30              17:40         Sandstorms
Medium
3           09:20              09:30         Sandstorms
Low
4           19:50              20:05                Fog
Jam
...           ...                ...                ...
...
45579       11:35              11:45              Windy
High
45580       19:55              20:10              Windy
Jam
45581       23:50           24:05:00             Cloudy
Low
45582       13:35              13:40             Cloudy
High
45583       17:10              17:15                Fog
Medium

       Vehicle_condition Type_of_order Type_of_vehicle
multiple_deliveries  \
0                      2         Snack      motorcycle
3.0
1                      1          Meal      motorcycle
1.0
2                      1        Drinks         scooter
1.0
3                      0        Buffet      motorcycle
0.0
4                      1         Snack         scooter
1.0
...                  ...           ...             ...
...
45579                  1          Meal      motorcycle
0.0
45580                  0        Buffet      motorcycle
1.0
45581                  1        Drinks         scooter
0.0
45582                  0         Snack      motorcycle
1.0
```

```
45583                  2         Snack         scooter
1.0

      Festival           City  Time_taken (min)
0           No  Metropolitian                46
1           No  Metropolitian                23
2           No  Metropolitian                21
3           No  Metropolitian                20
4           No  Metropolitian                41
...        ...            ...               ...
45579       No  Metropolitian                32
45580       No  Metropolitian                36
45581       No  Metropolitian                16
45582       No  Metropolitian                26
45583       No  Metropolitian                36

[45584 rows x 20 columns]
```

```
df.isnull().sum()
```

```
ID                                0
Delivery_person_ID                0
Delivery_person_Age            1854
Delivery_person_Ratings        1908
Restaurant_latitude               0
Restaurant_longitude              0
Delivery_location_latitude        0
Delivery_location_longitude       0
Order_Date                        0
Time_Orderd                    1731
Time_Order_picked                 0
Weather_conditions              616
Road_traffic_density            601
Vehicle_condition                 0
Type_of_order                     0
Type_of_vehicle                   0
multiple_deliveries             993
Festival                        228
City                           1200
Time_taken (min)                  0
dtype: int64
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 45584 entries, 0 to 45583
Data columns (total 20 columns):
 #   Column                       Non-Null Count  Dtype
---  ------                       --------------  -----
 0   ID                           45584 non-null  object
```

```
 1   Delivery_person_ID           45584 non-null  object
 2   Delivery_person_Age          43730 non-null  float64
 3   Delivery_person_Ratings      43676 non-null  float64
 4   Restaurant_latitude          45584 non-null  float64
 5   Restaurant_longitude         45584 non-null  float64
 6   Delivery_location_latitude   45584 non-null  float64
 7   Delivery_location_longitude  45584 non-null  float64
 8   Order_Date                   45584 non-null  object
 9   Time_Orderd                  43853 non-null  object
 10  Time_Order_picked            45584 non-null  object
 11  Weather_conditions           44968 non-null  object
 12  Road_traffic_density         44983 non-null  object
 13  Vehicle_condition            45584 non-null  int64
 14  Type_of_order                45584 non-null  object
 15  Type_of_vehicle              45584 non-null  object
 16  multiple_deliveries          44591 non-null  float64
 17  Festival                     45356 non-null  object
 18  City                         44384 non-null  object
 19  Time_taken (min)             45584 non-null  int64
dtypes: float64(7), int64(2), object(11)
memory usage: 7.0+ MB
```

```
df["Order_Date"] = pd.to_datetime(df["Order_Date"],format="mixed")
```

Renaming the columns

```
df=df.rename(columns={"Delivery_person_Age":"del_boy_age","Delivery_pe
rson_Ratings":"Del_boy_Ratings","Road_traffic_density":"traffic_densit
y",

"Delivery_location_latitude":"del_loc_lat","Delivery_location_longitud
e":"del_loc_log"})

df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 45584 entries, 0 to 45583
Data columns (total 20 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   ID                    45584 non-null  object
 1   Delivery_person_ID    45584 non-null  object
 2   del_boy_age           43730 non-null  float64
 3   Del_boy_Ratings       43676 non-null  float64
 4   Restaurant_latitude   45584 non-null  float64
 5   Restaurant_longitude  45584 non-null  float64
 6   del_loc_lat           45584 non-null  float64
 7   del_loc_log           45584 non-null  float64
```

```
 8   Order_Date             45584 non-null  datetime64[ns]
 9   Time_Orderd            43853 non-null  object
 10  Time_Order_picked      45584 non-null  object
 11  Weather_conditions     44968 non-null  object
 12  traffic_density        44983 non-null  object
 13  Vehicle_condition      45584 non-null  int64
 14  Type_of_order          45584 non-null  object
 15  Type_of_vehicle        45584 non-null  object
 16  multiple_deliveries    44591 non-null  float64
 17  Festival               45356 non-null  object
 18  City                   44384 non-null  object
 19  Time_taken (min)       45584 non-null  int64
dtypes: datetime64[ns](1), float64(7), int64(2), object(10)
memory usage: 7.0+ MB
```

Removing dublicate

```
#removing duplicates if any.
df.drop_duplicates()
```

```
           ID Delivery_person_ID  del_boy_age  Del_boy_Ratings  \
0      0xcdcd      DEHRES17DEL01         36.0              4.2
1      0xd987      KOCRES16DEL01         21.0              4.7
2      0x2784     PUNERES13DEL03         23.0              4.7
3      0xc8b6     LUDHRES15DEL02         34.0              4.3
4      0xdb64      KNPRES14DEL02         24.0              4.7
...       ...                ...          ...              ...
45579  0x7c09      JAPRES04DEL01         30.0              4.8
45580  0xd641      AGRRES16DEL01         21.0              4.6
45581  0x4f8d     CHENRES08DEL03         30.0              4.9
45582  0x5eee    COIMBRES11DEL01         20.0              4.7
45583  0x5fb2   RANCHIRES09DEL02         23.0              4.9

       Restaurant_latitude  Restaurant_longitude  del_loc_lat
del_loc_log  \
0                30.327968             78.046106    30.397968
78.116106
1                10.003064             76.307589    10.043064
76.347589
2                18.562450             73.916619    18.652450
74.006619
3                30.899584             75.809346    30.919584
75.829346
4                26.463504             80.372929    26.593504
80.502929
...                    ...                   ...          ...
...
45579            26.902328             75.794257    26.912328
```

```
75.804257
45580              0.000000              0.000000     0.070000
0.070000
45581             13.022394             80.242439    13.052394
80.272439
45582             11.001753             76.986241    11.041753
77.026241
45583             23.351058             85.325731    23.431058
85.405731

      Order_Date Time_Orderd Time_Order_picked Weather_conditions  \
0     2022-12-02       21:55              22:10                Fog
1     2022-02-13       14:55              15:05             Stormy
2     2022-04-03       17:30              17:40         Sandstorms
3     2022-02-13       09:20              09:30         Sandstorms
4     2022-02-14       19:50              20:05                Fog
...          ...         ...                ...                ...
45579 2022-03-24       11:35              11:45              Windy
45580 2022-02-16       19:55              20:10              Windy
45581 2022-11-03       23:50           24:05:00             Cloudy
45582 2022-07-03       13:35              13:40             Cloudy
45583 2022-02-03       17:10              17:15                Fog

      traffic_density  Vehicle_condition Type_of_order Type_of_vehicle
\
0                 Jam                  2         Snack      motorcycle

1                High                  1          Meal      motorcycle

2              Medium                  1        Drinks         scooter

3                 Low                  0        Buffet      motorcycle

4                 Jam                  1         Snack         scooter

...               ...                ...           ...             ...

45579            High                  1          Meal      motorcycle

45580             Jam                  0        Buffet      motorcycle

45581             Low                  1        Drinks         scooter

45582            High                  0         Snack      motorcycle

45583          Medium                  2         Snack         scooter


       multiple_deliveries Festival           City  Time_taken (min)
0                      3.0       No  Metropolitian                46
1                      1.0       No  Metropolitian                23
```

```
2                       1.0       No  Metropolitian                 21
3                       0.0       No  Metropolitian                 20
4                       1.0       No  Metropolitian                 41
...                     ...      ...            ...                ...
45579                   0.0       No  Metropolitian                 32
45580                   1.0       No  Metropolitian                 36
45581                   0.0       No  Metropolitian                 16
45582                   1.0       No  Metropolitian                 26
45583                   1.0       No  Metropolitian                 36

[45584 rows x 20 columns]
```

```
df.describe()
```

```
        del_boy_age  Del_boy_Ratings  Restaurant_latitude  \
count  43730.000000     43676.000000         45584.000000
mean      29.566911         4.633774            17.017948
min       15.000000         1.000000           -30.905562
25%       25.000000         4.500000            12.933284
50%       30.000000         4.700000            18.551440
75%       35.000000         4.900000            22.728163
max       50.000000         6.000000            30.914057
std        5.815064         0.334744             8.185674

       Restaurant_longitude   del_loc_lat   del_loc_log  \
count          45584.000000  45584.000000  45584.000000
mean              70.229684     17.465480     70.844161
min              -88.366217      0.010000      0.010000
25%               73.170000     12.988453     73.280000
50%               75.897963     18.633934     76.002574
75%               78.044095     22.785049     78.107044
max               88.433452     31.054057     88.563452
std               22.885575      7.335562     21.120578

                          Order_Date  Vehicle_condition
multiple_deliveries  \
count                          45584       45584.000000
44591.000000
mean   2022-04-20 10:57:49.849069824           1.023385
0.744635
min              2022-01-03 00:00:00           0.000000
0.000000
25%              2022-03-04 00:00:00           0.000000
0.000000
50%              2022-03-24 00:00:00           1.000000
1.000000
75%              2022-05-03 00:00:00           2.000000
1.000000
max              2022-12-03 00:00:00           3.000000
3.000000
```

```
std                                  NaN           0.839055
0.572510

       Time_taken (min)
count      45584.000000
mean          26.293963
min           10.000000
25%           19.000000
50%           26.000000
75%           32.000000
max           54.000000
std            9.384298
```

```
df["Type_of_vehicle"].unique()
```

```
array(['motorcycle', 'scooter', 'electric_scooter', 'bicycle'],
      dtype=object)
```

```
df[df["del_boy_age"].isnull()==True]
```

```
           ID Delivery_person_ID  del_boy_age  Del_boy_Ratings  \
33       0x4f0      MUMRES17DEL02          NaN              NaN
65       0xa9f     BANGRES13DEL01          NaN              NaN
111      0x6e2      VADRES06DEL01          NaN              NaN
131     0x1ec7     PUNERES05DEL03          NaN              NaN
170     0x1b58      JAPRES03DEL01          NaN              NaN
...        ...                ...          ...              ...
45499   0x426b      HYDRES06DEL01          NaN              NaN
45509    0x863     INDORES05DEL01          NaN              NaN
45510   0xd0be      AGRRES07DEL01          NaN              NaN
45517   0x1ed1      JAPRES18DEL01          NaN              NaN
45557   0x1163      SURRES06DEL02          NaN              NaN

       Restaurant_latitude  Restaurant_longitude  del_loc_lat
del_loc_log  \
33               19.121999             72.908493    19.201999
72.988493
65               12.935662             77.614130    12.975662
77.654130
111              22.312790             73.170283    22.422790
73.280283
131              18.543626             73.905101    18.573626
73.935101
170              26.913483             75.803139    26.983483
75.873139
...                    ...                   ...          ...
...
45499            17.455894             78.375467    17.525894
78.445467
45509            22.727021             75.884167    22.737021
```

```
75.894167
45510            27.160934              78.044095    27.170934
78.054095
45517            26.913987              75.752891    27.023987
75.862891
45557            21.185047              72.808590    21.205047
72.828590

      Order_Date Time_Orderd Time_Order_picked Weather_conditions  \
33    2022-06-04         NaN              18:35             Cloudy
65    2022-11-03         NaN              15:30                NaN
111   2022-02-04         NaN              18:25                NaN
131   2022-09-03         NaN              23:40             Stormy
170   2022-12-03         NaN              17:50         Sandstorms
...          ...         ...                ...                ...
45499 2022-03-23       17:55               0.75                Fog
45509 2022-03-19         NaN              11:40                NaN
45510 2022-02-13         NaN              09:05                NaN
45517 2022-03-27       20:35              20:45             Cloudy
45557 2022-03-04         NaN              10:50              Windy

      traffic_density  Vehicle_condition Type_of_order
Type_of_vehicle  \
33             Medium                  1        Drinks
scooter
65                NaN                  1        Drinks
scooter
111               NaN                  3         Snack
electric_scooter
131               Low                  2         Snack
scooter
170            Medium                  0        Drinks
motorcycle
...               ...                ...           ...                 .
..
45499          Medium                  0          Meal
motorcycle
45509             NaN                  3         Snack
electric_scooter
45510             NaN                  3          Meal
electric_scooter
45517             Jam                  2         Snack
scooter
45557             Low                  2          Meal
scooter

       multiple_deliveries Festival           City  Time_taken (min)
33                     1.0       No  Metropolitian                33
65                     0.0       No  Metropolitian                19
111                    0.0       No  Metropolitian                25
```

```
131                    0.0       No            NaN                 19
170                    0.0       No  Metropolitian                 43
...                    ...      ...            ...                ...
45499                  1.0       No          Urban                 36
45509                  0.0       No  Metropolitian                 25
45510                  1.0       No          Urban                 17
45517                  0.0       No  Metropolitian                 48
45557                  0.0       No          Urban                 12

[1854 rows x 20 columns]
```

```
df.groupby(["Delivery_person_ID"])["Time_taken (min)"].count()
```

```
Delivery_person_ID
AGRRES010DEL01    14
AGRRES010DEL02    15
AGRRES010DEL03    13
AGRRES01DEL01     12
AGRRES01DEL02     14
                  ..
VADRES19DEL02     59
VADRES19DEL03     41
VADRES20DEL01     59
VADRES20DEL02     50
VADRES20DEL03     37
Name: Time_taken (min), Length: 1320, dtype: int64
```

```
# the delivery id has unique pattern of city embaded in the id lets
extract the city code from their id
```

```
df["Delivery_person_ID"].str[:3]
```

```
0        DEH
1        KOC
2        PUN
3        LUD
4        KNP
        ...
45579    JAP
45580    AGR
45581    CHE
45582    COI
45583    RAN
Name: Delivery_person_ID, Length: 45584, dtype: object
```

```
df["city_name"] = df["Delivery_person_ID"].str[:3]
```

```
df
```

```
           ID Delivery_person_ID  del_boy_age  Del_boy_Ratings  \
0      0xcdcd      DEHRES17DEL01         36.0              4.2
```

```
1      0xd987      KOCRES16DEL01         21.0              4.7
2      0x2784     PUNERES13DEL03         23.0              4.7
3      0xc8b6     LUDHRES15DEL02         34.0              4.3
4      0xdb64      KNPRES14DEL02         24.0              4.7
...       ...                ...          ...              ...
45579  0x7c09      JAPRES04DEL01         30.0              4.8
45580  0xd641      AGRRES16DEL01         21.0              4.6
45581  0x4f8d     CHENRES08DEL03         30.0              4.9
45582  0x5eee    COIMBRES11DEL01         20.0              4.7
45583  0x5fb2   RANCHIRES09DEL02         23.0              4.9

       Restaurant_latitude  Restaurant_longitude  del_loc_lat
del_loc_log  \
0                30.327968             78.046106    30.397968
78.116106
1                10.003064             76.307589    10.043064
76.347589
2                18.562450             73.916619    18.652450
74.006619
3                30.899584             75.809346    30.919584
75.829346
4                26.463504             80.372929    26.593504
80.502929
...                    ...                   ...          ...
...
45579            26.902328             75.794257    26.912328
75.804257
45580             0.000000              0.000000     0.070000
0.070000
45581            13.022394             80.242439    13.052394
80.272439
45582            11.001753             76.986241    11.041753
77.026241
45583            23.351058             85.325731    23.431058
85.405731

      Order_Date Time_Orderd  ... Weather_conditions
traffic_density  \
0     2022-12-02       21:55  ...               Fog             Jam

1     2022-02-13       14:55  ...             Stormy            High

2     2022-04-03       17:30  ...         Sandstorms          Medium

3     2022-02-13       09:20  ...         Sandstorms             Low

4     2022-02-14       19:50  ...               Fog             Jam

...          ...         ...  ...               ...             ...
```

```
45579 2022-03-24       11:35  ...               Windy             High
45580 2022-02-16       19:55  ...               Windy              Jam
45581 2022-11-03       23:50  ...              Cloudy              Low
45582 2022-07-03       13:35  ...              Cloudy             High
45583 2022-02-03       17:10  ...                 Fog           Medium

      Vehicle_condition  Type_of_order Type_of_vehicle
multiple_deliveries  \
0                     2          Snack      motorcycle
3.0
1                     1           Meal      motorcycle
1.0
2                     1         Drinks         scooter
1.0
3                     0         Buffet      motorcycle
0.0
4                     1          Snack         scooter
1.0
...                 ...            ...             ...
...
45579                 1           Meal      motorcycle
0.0
45580                 0         Buffet      motorcycle
1.0
45581                 1         Drinks         scooter
0.0
45582                 0          Snack      motorcycle
1.0
45583                 2          Snack         scooter
1.0

       Festival           City Time_taken (min)  city_name
0            No  Metropolitian                46        DEH
1            No  Metropolitian                23        KOC
2            No  Metropolitian                21        PUN
3            No  Metropolitian                20        LUD
4            No  Metropolitian                41        KNP
...         ...            ...               ...        ...
45579        No  Metropolitian                32        JAP
45580        No  Metropolitian                36        AGR
45581        No  Metropolitian                16        CHE
45582        No  Metropolitian                26        COI
45583        No  Metropolitian                36        RAN

[45584 rows x 21 columns]
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 45584 entries, 0 to 45583
Data columns (total 21 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   ID                    45584 non-null  object
 1   Delivery_person_ID    45584 non-null  object
 2   del_boy_age           43730 non-null  float64
 3   Del_boy_Ratings       43676 non-null  float64
 4   Restaurant_latitude   45584 non-null  float64
 5   Restaurant_longitude  45584 non-null  float64
 6   del_loc_lat           45584 non-null  float64
 7   del_loc_log           45584 non-null  float64
 8   Order_Date            45584 non-null  datetime64[ns]
 9   Time_Orderd           43853 non-null  object
 10  Time_Order_picked     45584 non-null  object
 11  Weather_conditions    44968 non-null  object
 12  traffic_density       44983 non-null  object
 13  Vehicle_condition     45584 non-null  int64
 14  Type_of_order         45584 non-null  object
 15  Type_of_vehicle       45584 non-null  object
 16  multiple_deliveries   44591 non-null  float64
 17  Festival              45356 non-null  object
 18  City                  44384 non-null  object
 19  Time_taken (min)      45584 non-null  int64
 20  city_name             45584 non-null  object
dtypes: datetime64[ns](1), float64(7), int64(2), object(11)
memory usage: 7.3+ MB
```

```
df["Time_Orderd"]=
df["Time_Orderd"].str.replace(":",".").astype("float")
df
```

```
           ID Delivery_person_ID  del_boy_age  Del_boy_Ratings  \
0      0xcdcd      DEHRES17DEL01         36.0              4.2
1      0xd987      KOCRES16DEL01         21.0              4.7
2      0x2784     PUNERES13DEL03         23.0              4.7
3      0xc8b6     LUDHRES15DEL02         34.0              4.3
4      0xdb64      KNPRES14DEL02         24.0              4.7
...       ...                ...          ...              ...
45579  0x7c09      JAPRES04DEL01         30.0              4.8
45580  0xd641      AGRRES16DEL01         21.0              4.6
45581  0x4f8d     CHENRES08DEL03         30.0              4.9
45582  0x5eee    COIMBRES11DEL01         20.0              4.7
45583  0x5fb2   RANCHIRES09DEL02         23.0              4.9

       Restaurant_latitude  Restaurant_longitude  del_loc_lat
del_loc_log  \
```

```
0                30.327968              78.046106    30.397968
78.116106
1                10.003064              76.307589    10.043064
76.347589
2                18.562450              73.916619    18.652450
74.006619
3                30.899584              75.809346    30.919584
75.829346
4                26.463504              80.372929    26.593504
80.502929
...                    ...                    ...          ...
...
45579            26.902328              75.794257    26.912328
75.804257
45580             0.000000               0.000000     0.070000
0.070000
45581            13.022394              80.242439    13.052394
80.272439
45582            11.001753              76.986241    11.041753
77.026241
45583            23.351058              85.325731    23.431058
85.405731

      Order_Date  Time_Orderd  ... Weather_conditions traffic_density
\
0     2022-12-02        21.55  ...                Fog             Jam

1     2022-02-13        14.55  ...             Stormy            High

2     2022-04-03        17.30  ...         Sandstorms          Medium

3     2022-02-13         9.20  ...         Sandstorms             Low

4     2022-02-14        19.50  ...                Fog             Jam

...          ...          ...  ...                ...             ...

45579 2022-03-24        11.35  ...              Windy            High

45580 2022-02-16        19.55  ...              Windy             Jam

45581 2022-11-03        23.50  ...             Cloudy             Low

45582 2022-07-03        13.35  ...             Cloudy            High

45583 2022-02-03        17.10  ...                Fog          Medium


      Vehicle_condition  Type_of_order Type_of_vehicle
multiple_deliveries  \
```

```
0                     2          Snack      motorcycle
3.0
1                     1           Meal      motorcycle
1.0
2                     1         Drinks         scooter
1.0
3                     0         Buffet      motorcycle
0.0
4                     1          Snack         scooter
1.0
...                 ...            ...             ...
...
45579                 1           Meal      motorcycle
0.0
45580                 0         Buffet      motorcycle
1.0
45581                 1         Drinks         scooter
0.0
45582                 0          Snack      motorcycle
1.0
45583                 2          Snack         scooter
1.0
```

```
       Festival           City Time_taken (min)  city_name
0            No  Metropolitian                46        DEH
1            No  Metropolitian                23        KOC
2            No  Metropolitian                21        PUN
3            No  Metropolitian                20        LUD
4            No  Metropolitian                41        KNP
...         ...            ...               ...        ...
45579        No  Metropolitian                32        JAP
45580        No  Metropolitian                36        AGR
45581        No  Metropolitian                16        CHE
45582        No  Metropolitian                26        COI
45583        No  Metropolitian                36        RAN

[45584 rows x 21 columns]
```

```
df["Time_Order_picked"]
```

```
0           22:10
1           15:05
2           17:40
3           09:30
4           20:05
           ...
45579       11:45
45580       20:10
45581    24:05:00
45582       13:40
```

```
45583       17:15
Name: Time_Order_picked, Length: 45584, dtype: object
```

```
# function to remove seconds from the column
def x(i):

    i = str(i)
    if len(i)>5:

        i=i.replace(":",".")
        i = i[:-3]
        return i
    else:
        i=i.replace(":",".")
        return i



df["Time_Order_picked"] = df["Time_Order_picked"].apply(x)
```

```
df["Time_Order_picked"] = df["Time_Order_picked"].astype('float')
```

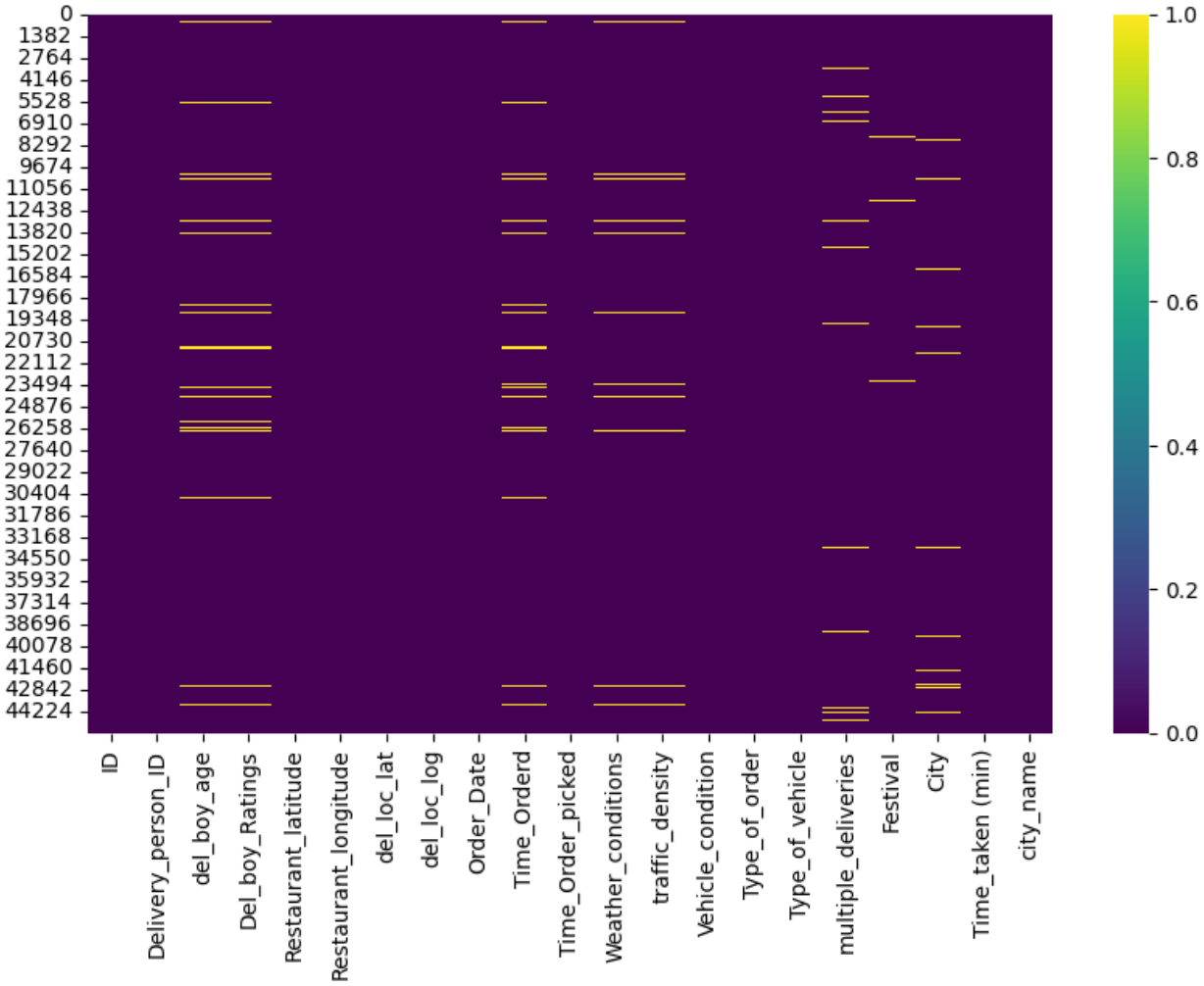
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 45584 entries, 0 to 45583
Data columns (total 21 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   ID                    45584 non-null  object
 1   Delivery_person_ID    45584 non-null  object
 2   del_boy_age           43730 non-null  float64
 3   Del_boy_Ratings       43676 non-null  float64
 4   Restaurant_latitude   45584 non-null  float64
 5   Restaurant_longitude  45584 non-null  float64
 6   del_loc_lat           45584 non-null  float64
 7   del_loc_log           45584 non-null  float64
 8   Order_Date            45584 non-null  datetime64[ns]
 9   Time_Orderd           43853 non-null  float64
 10  Time_Order_picked     45584 non-null  float64
 11  Weather_conditions    44968 non-null  object
 12  traffic_density       44983 non-null  object
 13  Vehicle_condition     45584 non-null  int64
 14  Type_of_order         45584 non-null  object
 15  Type_of_vehicle       45584 non-null  object
 16  multiple_deliveries   44591 non-null  float64
 17  Festival              45356 non-null  object
 18  City                  44384 non-null  object
 19  Time_taken (min)      45584 non-null  int64
 20  city_name             45584 non-null  object
```

```
dtypes: datetime64[ns](1), float64(9), int64(2), object(9)
memory usage: 7.3+ MB
```

finding any pattern in null values

```
import seaborn as sns
import matplotlib.pyplot as plt

plt.figure(figsize=(10, 6))
sns.heatmap(df.isnull(), cbar=True, cmap='viridis')
plt.show()
```




```
 null_1 = df[df["del_boy_age"].isnull() &
df["Del_boy_Ratings"].isnull()]
```

```
null_2 = df[df["traffic_density"].isnull() &
df["Weather_conditions"].isnull()]

null_1
```

```
           ID Delivery_person_ID  del_boy_age  Del_boy_Ratings  \
33      0x4f0      MUMRES17DEL02          NaN              NaN
65      0xa9f     BANGRES13DEL01          NaN              NaN
111     0x6e2      VADRES06DEL01          NaN              NaN
131    0x1ec7     PUNERES05DEL03          NaN              NaN
170    0x1b58      JAPRES03DEL01          NaN              NaN
...       ...                ...          ...              ...
45499  0x426b      HYDRES06DEL01          NaN              NaN
45509   0x863     INDORES05DEL01          NaN              NaN
45510  0xd0be      AGRRES07DEL01          NaN              NaN
45517  0x1ed1      JAPRES18DEL01          NaN              NaN
45557  0x1163      SURRES06DEL02          NaN              NaN

       Restaurant_latitude  Restaurant_longitude  del_loc_lat
del_loc_log  \
33               19.121999             72.908493    19.201999
72.988493
65               12.935662             77.614130    12.975662
77.654130
111              22.312790             73.170283    22.422790
73.280283
131              18.543626             73.905101    18.573626
73.935101
170              26.913483             75.803139    26.983483
75.873139
...                    ...                   ...          ...
...
45499            17.455894             78.375467    17.525894
78.445467
45509            22.727021             75.884167    22.737021
75.894167
45510            27.160934             78.044095    27.170934
78.054095
45517            26.913987             75.752891    27.023987
75.862891
45557            21.185047             72.808590    21.205047
72.828590

      Order_Date  Time_Orderd  ...  Weather_conditions traffic_density
\
33    2022-06-04          NaN  ...              Cloudy          Medium

65    2022-11-03          NaN  ...                 NaN             NaN

111   2022-02-04          NaN  ...                 NaN             NaN
```

```
131   2022-09-03          NaN  ...              Stormy              Low
170   2022-12-03          NaN  ...          Sandstorms           Medium
...          ...          ...  ...                 ...              ...
45499 2022-03-23        17.55  ...                 Fog           Medium
45509 2022-03-19          NaN  ...                 NaN              NaN
45510 2022-02-13          NaN  ...                 NaN              NaN
45517 2022-03-27        20.35  ...              Cloudy              Jam
45557 2022-03-04          NaN  ...               Windy              Low

      Vehicle_condition  Type_of_order   Type_of_vehicle
multiple_deliveries  \
33                    1         Drinks           scooter
1.0
65                    1         Drinks           scooter
0.0
111                   3          Snack  electric_scooter
0.0
131                   2          Snack           scooter
0.0
170                   0         Drinks        motorcycle
0.0
...                 ...            ...               ...
...
45499                 0           Meal        motorcycle
1.0
45509                 3          Snack  electric_scooter
0.0
45510                 3           Meal  electric_scooter
1.0
45517                 2          Snack           scooter
0.0
45557                 2           Meal           scooter
0.0

       Festival           City Time_taken (min)  city_name
33           No  Metropolitian               33        MUM
65           No  Metropolitian               19        BAN
111          No  Metropolitian               25        VAD
131          No            NaN               19        PUN
170          No  Metropolitian               43        JAP
...         ...            ...              ...        ...
```

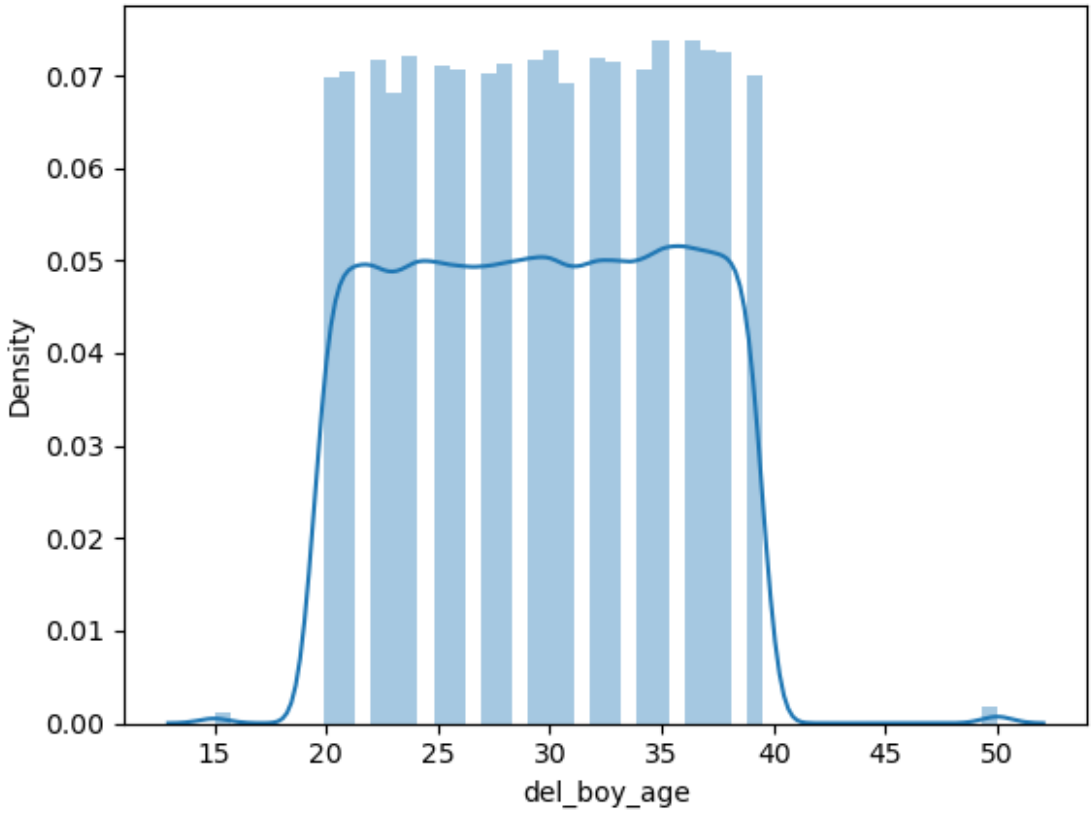
```
45499         No           Urban                36         HYD
45509         No  Metropolitian                 25         IND
45510         No           Urban                17         AGR
45517         No  Metropolitian                 48         JAP
45557         No           Urban                12         SUR

[1854 rows x 21 columns]
```

```
df['del_boy_age'].corr(df['Del_boy_Ratings'])
```

```
-0.0679594840339842
```

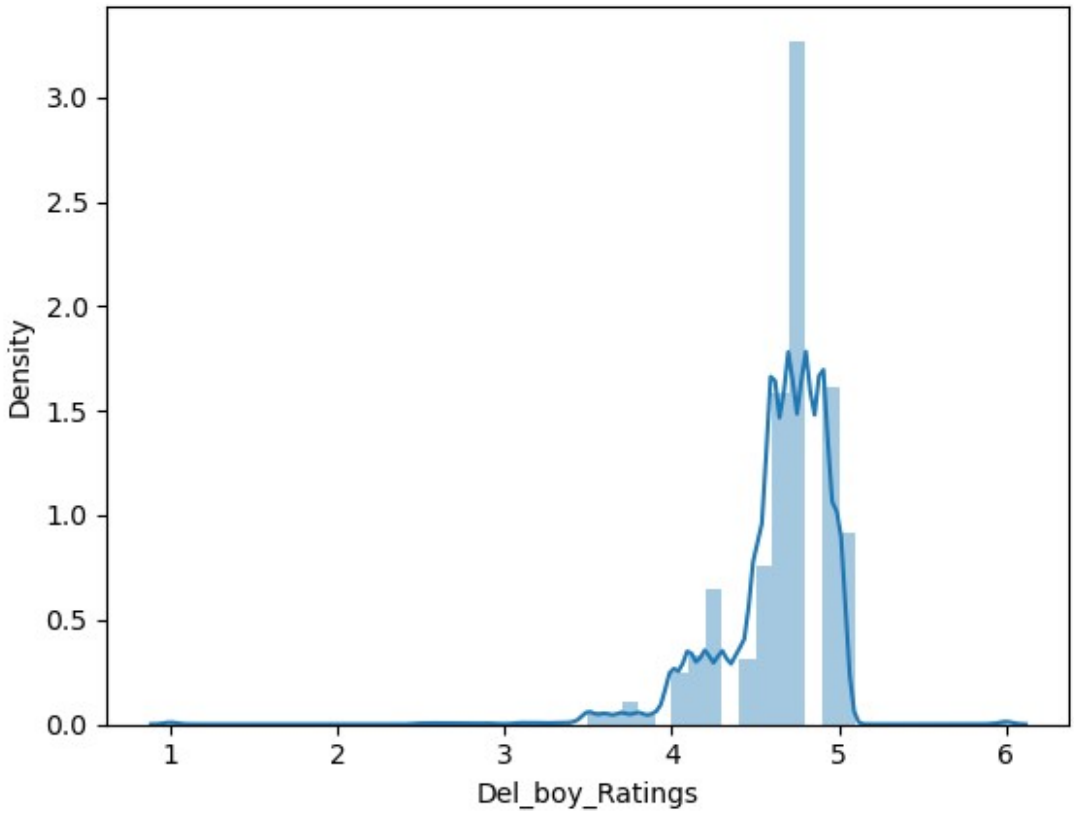
```
sns.distplot(df['del_boy_age'])
```

```
<Axes: xlabel='del_boy_age', ylabel='Density'>
```




```
sns.distplot(df['Del_boy_Ratings'])
```

```
<Axes: xlabel='Del_boy_Ratings', ylabel='Density'>
```



The null values in del_boy_age can be filled with median and del_boy_rating can be filled with mode values

```
df['del_boy_age'].fillna(df['del_boy_age'].median(), inplace=True)
```

```
df['del_boy_age'].isnull().sum()
```

```
0
```

```
df['Del_boy_Ratings'].fillna(df['Del_boy_Ratings'].mode()[0],
inplace=True)
```

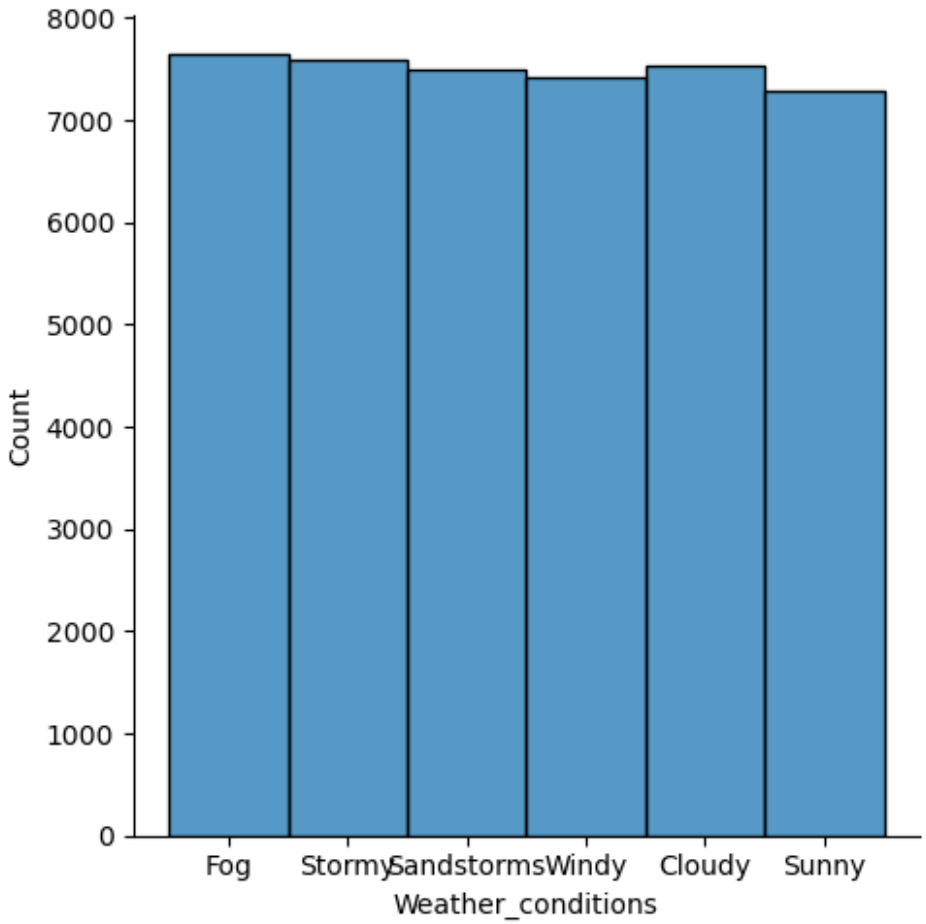
```
df["Del_boy_Ratings"].isnull().sum()
```

```
0
```

The data shows some relation between weather and traffic density ut the data is too small to put any efforts for the anaylis so we drop the the null values

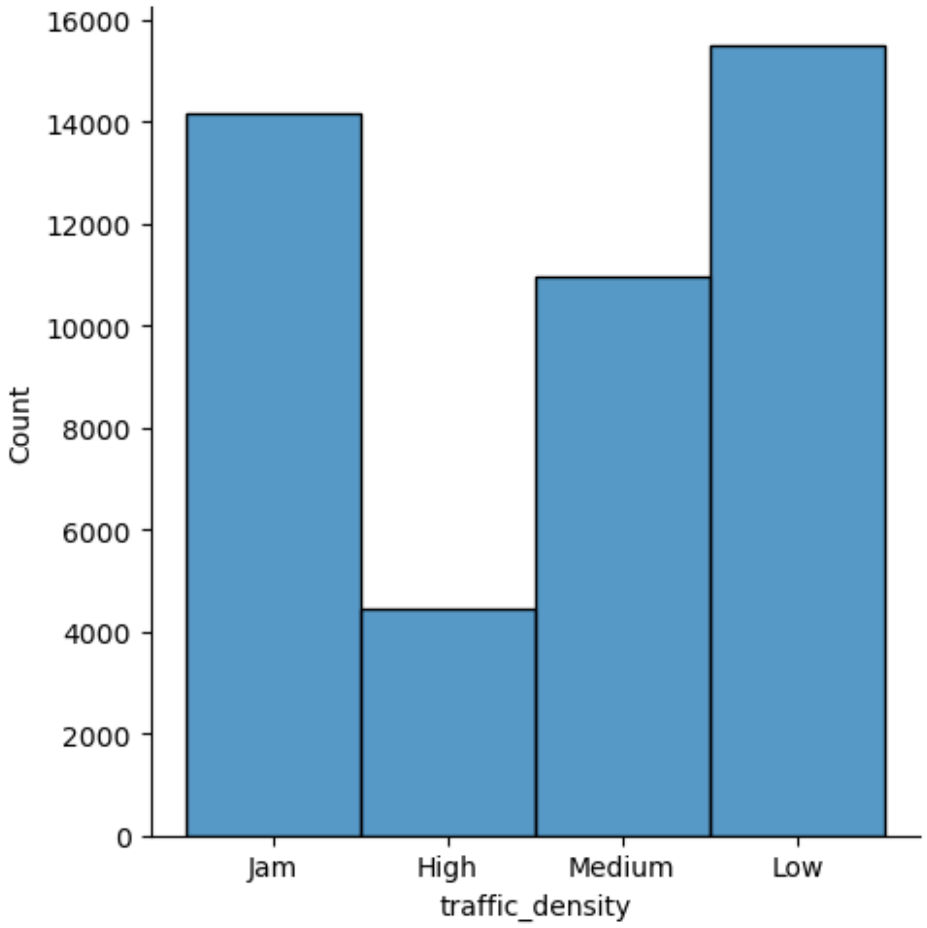
```
df['Weather_conditions'].value_counts()
sns.displot(df['Weather_conditions'])
```

```
<seaborn.axisgrid.FacetGrid at 0x20acaae15b0>
```




```
df['traffic_density'].value_counts()
sns.displot(df['traffic_density'])
```

```
<seaborn.axisgrid.FacetGrid at 0x20acab5de20>
```




```
df['Weather_conditions'].fillna(df['Weather_conditions'].mode()[0],
inplace=True)
df['traffic_density'].fillna(df['traffic_density'].mode()[0],
inplace=True)
```

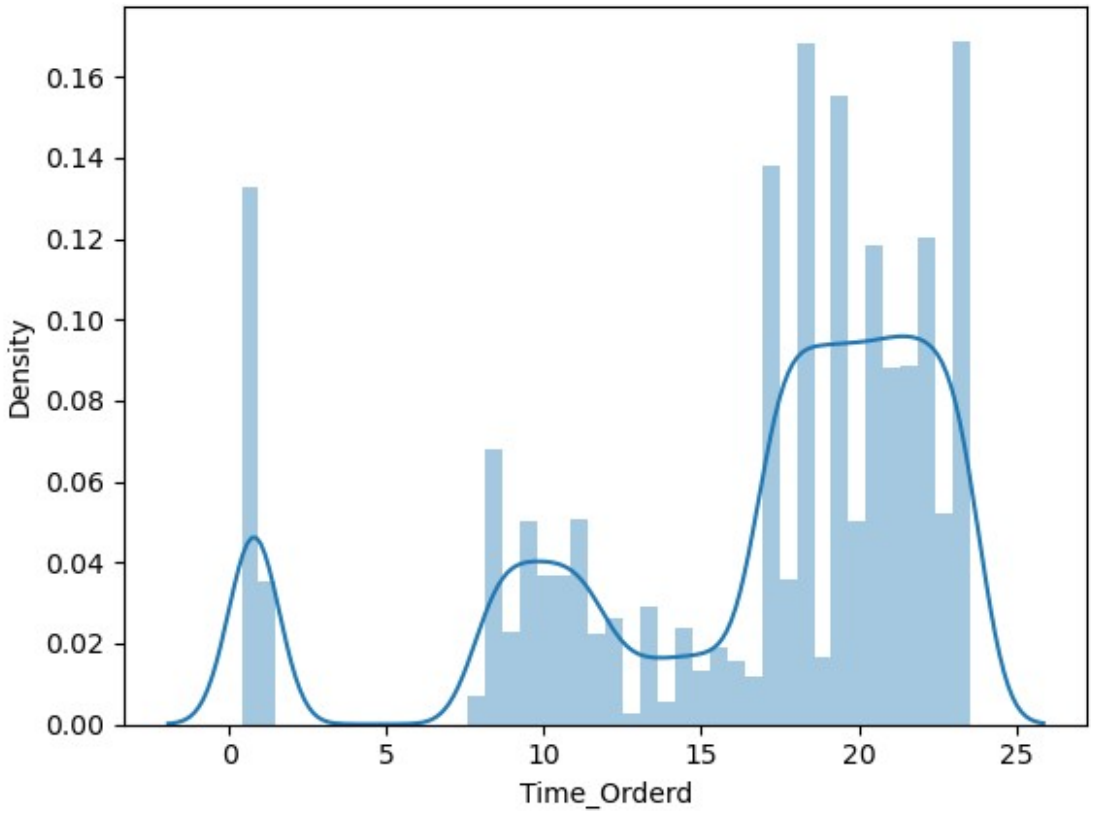
Finding skewed to fill the data accordingly

```
df["Time_Orderd"].isnull().sum()
```

```
1731
```

```
sns.distplot(df["Time_Orderd"])
```

```
<Axes: xlabel='Time_Orderd', ylabel='Density'>
```

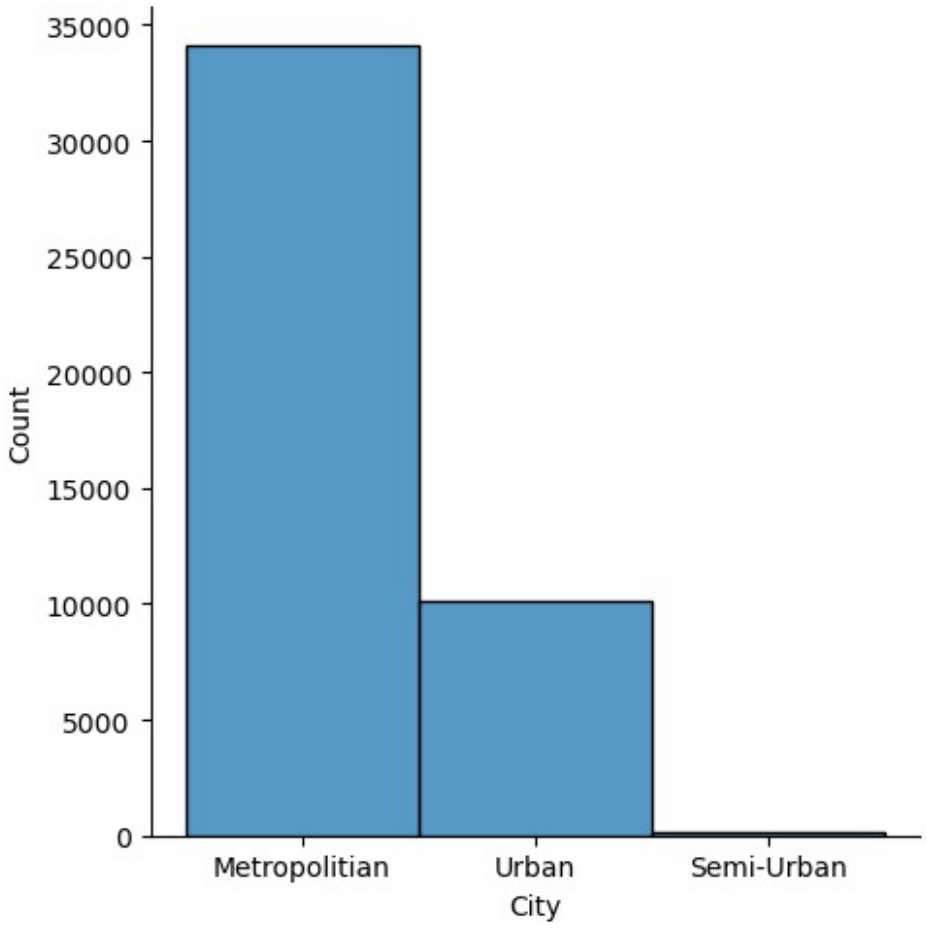



```
df['Time_Orderd'].fillna(df['Time_Orderd'].median(), inplace=True)
```

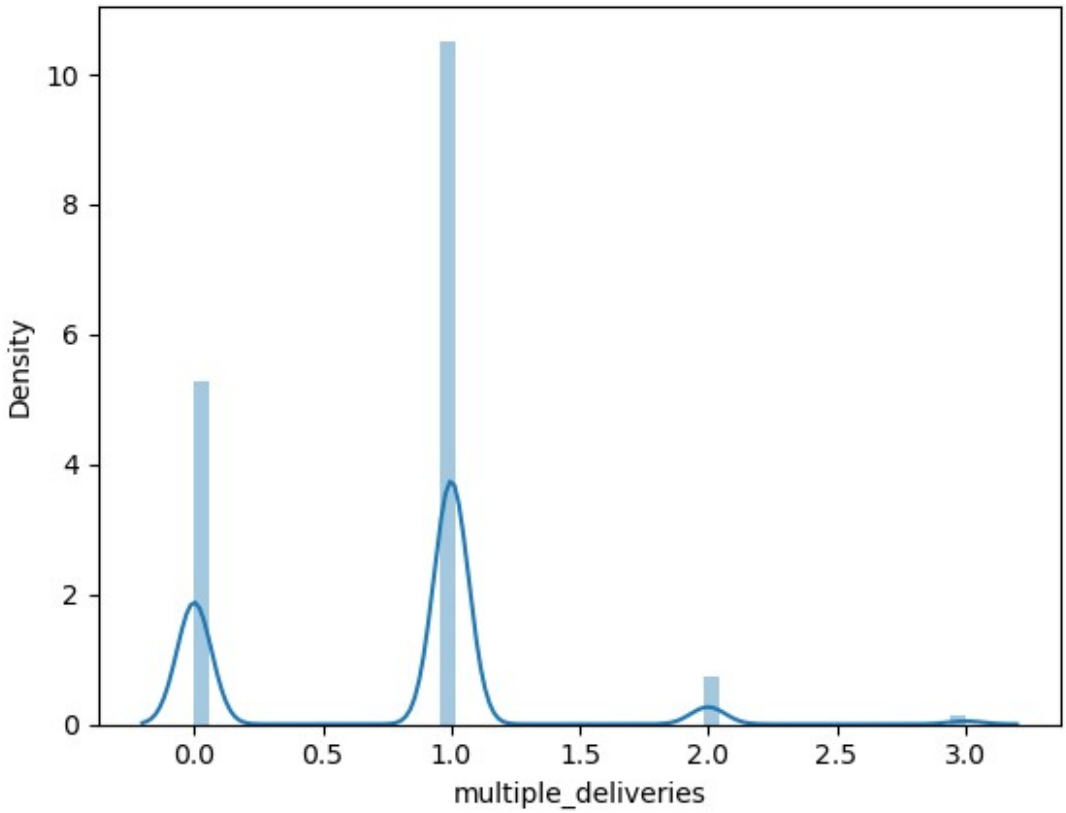
filling futhur null values

```
sns.displot(df["City"])
```

```
<seaborn.axisgrid.FacetGrid at 0x20acd107a40>
```

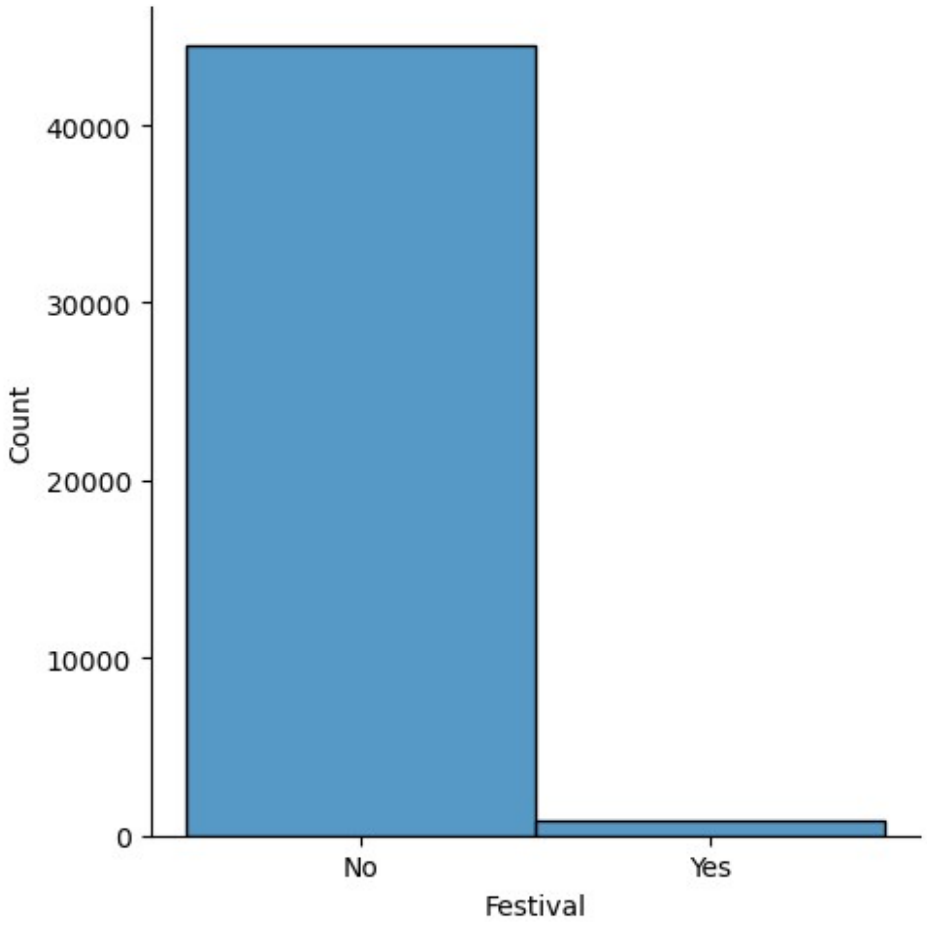



```
df['City'].fillna(df['City'].mode()[0], inplace=True)
sns.distplot(df["multiple_deliveries"])
<Axes: xlabel='multiple_deliveries', ylabel='Density'>
```




```
df['multiple_deliveries'].fillna(df['multiple_deliveries'].median(),
inplace=True)

sns.displot(df["Festival"])
```

```
<seaborn.axisgrid.FacetGrid at 0x20ac98a81a0>
```




```
df['Festival'].fillna(df['Festival'].mode()[0], inplace=True)

df.isnull().sum()
```

```
ID                          0
Delivery_person_ID          0
del_boy_age                 0
Del_boy_Ratings             0
Restaurant_latitude         0
Restaurant_longitude        0
del_loc_lat                 0
del_loc_log                 0
Order_Date                  0
Time_Orderd                 0
Time_Order_picked           0
Weather_conditions          0
traffic_density             0
Vehicle_condition           0
Type_of_order               0
Type_of_vehicle             0
multiple_deliveries         0
```

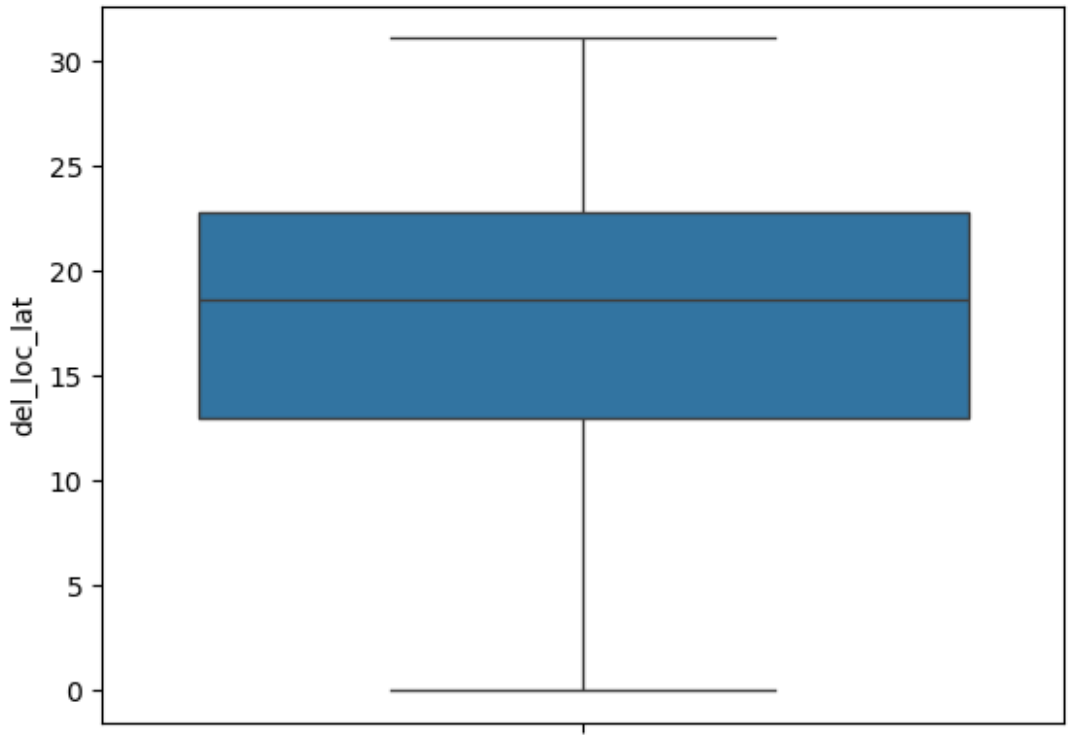
```
Festival                0
City                    0
Time_taken (min)        0
city_name               0
dtype: int64
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 45584 entries, 0 to 45583
Data columns (total 21 columns):
 #   Column                Non-Null Count  Dtype         
---  ------                --------------  -----         
 0   ID                    45584 non-null  object        
 1   Delivery_person_ID    45584 non-null  object        
 2   del_boy_age           45584 non-null  float64       
 3   Del_boy_Ratings       45584 non-null  float64       
 4   Restaurant_latitude   45584 non-null  float64       
 5   Restaurant_longitude  45584 non-null  float64       
 6   del_loc_lat           45584 non-null  float64       
 7   del_loc_log           45584 non-null  float64       
 8   Order_Date            45584 non-null  datetime64[ns]
 9   Time_Orderd           45584 non-null  float64       
 10  Time_Order_picked     45584 non-null  float64       
 11  Weather_conditions    45584 non-null  object        
 12  traffic_density       45584 non-null  object        
 13  Vehicle_condition     45584 non-null  int64         
 14  Type_of_order         45584 non-null  object        
 15  Type_of_vehicle       45584 non-null  object        
 16  multiple_deliveries   45584 non-null  float64       
 17  Festival              45584 non-null  object        
 18  City                  45584 non-null  object        
 19  Time_taken (min)      45584 non-null  int64         
 20  city_name             45584 non-null  object        
dtypes: datetime64[ns](1), float64(9), int64(2), object(9)
memory usage: 7.3+ MB
```

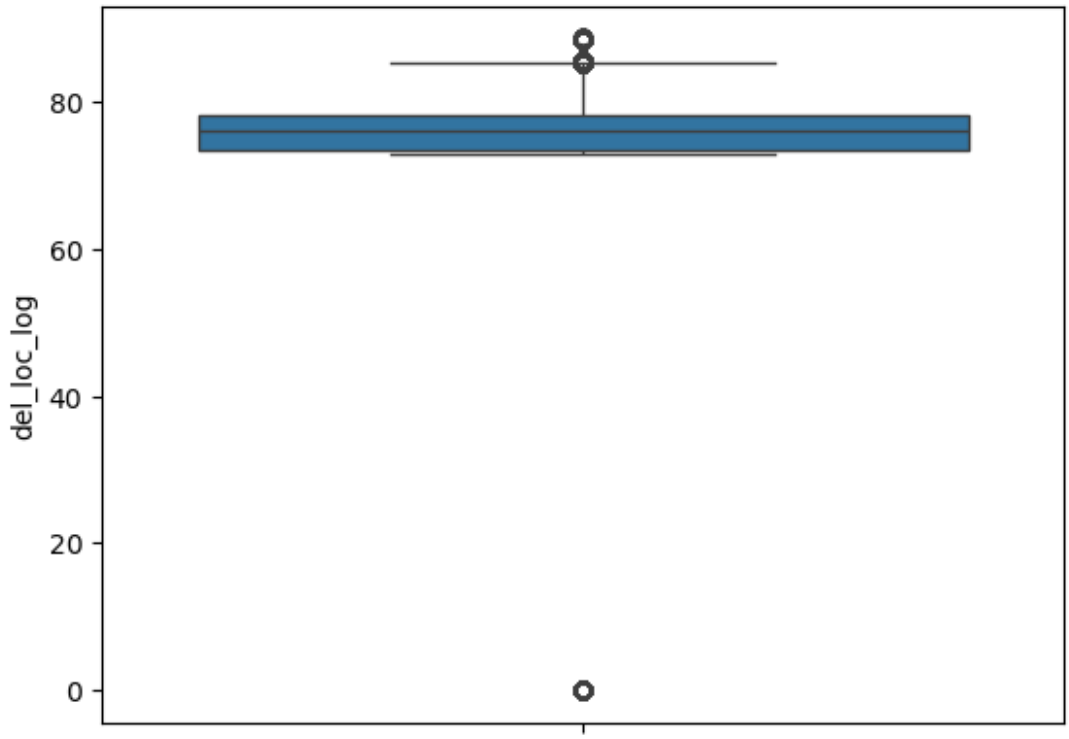
```
sns.boxplot(df["del_loc_lat"])
```

```
<Axes: ylabel='del_loc_lat'>
```




```
sns.boxplot(df["del_loc_log"])
```

```
<Axes: ylabel='del_loc_log'>
```



```
df.head()
```

```
       ID Delivery_person_ID  del_boy_age  Del_boy_Ratings  \
0  0xcdcd      DEHRES17DEL01         36.0              4.2
1  0xd987      KOCRES16DEL01         21.0              4.7
2  0x2784     PUNERES13DEL03         23.0              4.7
3  0xc8b6     LUDHRES15DEL02         34.0              4.3
4  0xdb64      KNPRES14DEL02         24.0              4.7

   Restaurant_latitude  Restaurant_longitude  del_loc_lat  del_loc_log
\
0            30.327968             78.046106    30.397968    78.116106

1            10.003064             76.307589    10.043064    76.347589

2            18.562450             73.916619    18.652450    74.006619

3            30.899584             75.809346    30.919584    75.829346

4            26.463504             80.372929    26.593504    80.502929


  Order_Date  Time_Orderd  ...  Weather_conditions traffic_density  \
0 2022-12-02        21.55  ...                 Fog             Jam
1 2022-02-13        14.55  ...              Stormy            High
2 2022-04-03        17.30  ...          Sandstorms          Medium
3 2022-02-13         9.20  ...          Sandstorms             Low
4 2022-02-14        19.50  ...                 Fog             Jam

  Vehicle_condition  Type_of_order Type_of_vehicle multiple_deliveries
\
0                 2          Snack      motorcycle                 3.0

1                 1           Meal      motorcycle                 1.0

2                 1         Drinks         scooter                 1.0

3                 0         Buffet      motorcycle                 0.0

4                 1          Snack         scooter                 1.0


   Festival           City Time_taken (min)  city_name
0        No  Metropolitian               46        DEH
1        No  Metropolitian               23        KOC
2        No  Metropolitian               21        PUN
3        No  Metropolitian               20        LUD
4        No  Metropolitian               41        KNP

[5 rows x 21 columns]
```

```
import pandas as pd
import pymysql

# Sample DataFrame
# df = pd.read_csv('your_data.csv')  # Load your DataFrame here

# Database connection parameters
db_config = {
    'host': 'localhost',          # e.g., 'localhost'
    'user': 'root',
    'password': '12345',
    'database': 'zomato_operations'
}

# Create a connection
connection = pymysql.connect(**db_config)

# Create a cursor object
cursor = connection.cursor()

# Create the table if it doesn't exist
create_table_query = """
CREATE TABLE IF NOT EXISTS delivery_data (
    ID VARCHAR(255),
    Delivery_person_ID VARCHAR(255),
    del_boy_age FLOAT,
    Del_boy_Ratings FLOAT,
    Restaurant_latitude FLOAT,
    Restaurant_longitude FLOAT,
    del_loc_lat FLOAT,
    del_loc_log FLOAT,
    Order_Date DATETIME,
    Time_Orderd FLOAT,
    Time_Order_picked FLOAT,
    Weather_conditions VARCHAR(255),
    traffic_density VARCHAR(255),
    Vehicle_condition INT,
    Type_of_order VARCHAR(255),
    Type_of_vehicle VARCHAR(255),
    multiple_deliveries FLOAT,
    Festival VARCHAR(255),
    City VARCHAR(255),
    Time_taken_min INT,
    city_name VARCHAR(255)
)
"""
cursor.execute(create_table_query)

# Insert DataFrame into MySQL
for index, row in df.iterrows():
```

```
    insert_query = """
    INSERT INTO delivery_data (ID, Delivery_person_ID, del_boy_age,
Del_boy_Ratings,
    Restaurant_latitude, Restaurant_longitude, del_loc_lat,
del_loc_log,
    Order_Date, Time_Orderd, Time_Order_picked, Weather_conditions,
    traffic_density, Vehicle_condition, Type_of_order,
Type_of_vehicle,
    multiple_deliveries, Festival, City, Time_taken_min, city_name)
    VALUES (%s, %s, %s, %s, %s, %s, %s, %s, %s, %s, %s, %s, %s, %s,
%s, %s, %s, %s, %s, %s, %s)
    """
    cursor.execute(insert_query, tuple(row))

# Commit changes
connection.commit()

# Close the cursor and connection
cursor.close()
connection.close()
```