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 PvMvSOL
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 PyMySOL
#connection = pymysql.connect(
     user='root',
     password='12345',
#
#
     host='localhost',
     database='zomato operations'
#
#)
#
## Query your data
#querv = '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 \
       0xcdcd
                   DEHRES17DEL01
                                                  36.0
0
       0xd987
1
                                                  21.0
                   KOCRES16DEL01
2
       0x2784
                  PUNERES13DEL03
                                                  23.0
3
       0xc8b6
                  LUDHRES15DEL02
                                                  34.0
4
       0xdb64
                                                  24.0
                   KNPRES14DEL02
45579 0x7c09
                   JAPRES04DEL01
                                                  30.0
45580
      0xd641
                   AGRRES16DEL01
                                                  21.0
45581 0x4f8d
                  CHENRES08DEL03
                                                  30.0
45582 0x5eee
                 COIMBRES11DEL01
                                                  20.0
45583 0x5fb2
                                                  23.0
                RANCHIRES09DEL02
```

Delivery_person		taurant_latitude		
Restaurant_longitude 0	4.2	30.327968		
78.046106		10.000004		
1 76.307589	4.7	10.003064		
2	4.7	18.562450		
73.916619	4.2	20 000504		
3 75.809346	4.3	30.899584		
4	4.7	26.463504		
80.372929				
		• • •		
45579	4.8	26.902328		
75.794257 45580	4.6	0.000000		
0.000000	4.0	0.00000		
45581	4.9	13.022394		
80.242439 45582	4.7	11.001753		
76.986241	4.7	11.001/33		
45583	4.9	23.351058		
85.325731				
Delivery_locat:	ion_latitude	Delivery_location_l	ongitude	
Order_Date \	20 207060	<u>-</u>	70 116106	12 02
0 2022	30.397968	,	78.116106	12-02-
1	10.043064	7	76.347589	13-02-
2022	18.652450	_	74 006610	04 02
2 2022	18.052450	,	74.006619	04-03-
3	30.919584	7	75.829346	13-02-
2022 4	26.593504	C	30.502929	14-02-
2022	20.393304	C	10.302929	14-02-
 45579	26.912328	7	75.804257	24-03-
2022	20.912320	,	3.004237	24-03-
45580	0.070000		0.070000	16-02-
2022 45581	13.052394	Ç	30.272439	11-03-
2022	13.032334		101212433	11-00-
45582	11.041753	7	77.026241	07-03-
2022 45583	23.431058	۶	35.405731	02-03-
13303	231431030		,51105/51	02 05

7	0	7	7
_	v	_	_

			r_picked Weat	her_conditions
_	ffic_density	\	22:10	Гол
0 Jam	21:55		22:10	Fog
Jaili 1	14:55		15:05	Stormy
High	14.33		13.03	3 COT III y
2	17:30		17:40	Sandstorms
Medium	17130		17110	Sanasconiis
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	22.50		24-05-00	C1 d
45581	23:50		24:05:00	Cloudy
Low 45582	13:35		13:40	Claudy
45562 High	15:55		13:40	Cloudy
45583	17:10		17:15	Fog
Medium	17.10		17.13	1 09
ricaram				
V				
V	ehicle condit	ion Ty	pe of order T	ype of vehicle
	ehicle_condit _deliveries	ion Ty	pe_of_order T	ype_of_vehicle
multiple 0		ion Ty \ 2	pe_of_order T Snack	ype_of_vehicle motorcycle
multiple 0 3.0		2	Snack	motorcycle
multiple 0 3.0 1		\		
multiple 0 3.0 1		2	Snack Meal	motorcycle motorcycle
multiple 0 3.0 1 1.0 2		2	Snack	motorcycle
multiple 0 3.0 1 1.0 2		2 1 1	Snack Meal Drinks	motorcycle motorcycle scooter
multiple 0 3.0 1 1.0 2 1.0 3		2	Snack Meal	motorcycle motorcycle
multiple 0 3.0 1 1.0 2 1.0 3		2 1 1 0	Snack Meal Drinks Buffet	motorcycle motorcycle scooter motorcycle
multiple 0 3.0 1 1.0 2 1.0 3 0.0 4		2 1 1	Snack Meal Drinks	motorcycle motorcycle scooter
multiple 0 3.0 1 1.0 2 1.0 3		2 1 1 0	Snack Meal Drinks Buffet	motorcycle motorcycle scooter motorcycle
multiple 0 3.0 1 1.0 2 1.0 3 0.0 4		2 1 1 0	Snack Meal Drinks Buffet	motorcycle motorcycle scooter motorcycle
multiple 0 3.0 1 1.0 2 1.0 3 0.0 4 1.0		2 1 1 0 1	Snack Meal Drinks Buffet Snack	motorcycle motorcycle scooter motorcycle scooter
multiple 0 3.0 1 1.0 2 1.0 3 0.0 4 1.0 45579		2 1 1 0	Snack Meal Drinks Buffet	motorcycle motorcycle scooter motorcycle
multiple 0 3.0 1 1.0 2 1.0 3 0.0 4 1.0		2 1 1 0 1	Snack Meal Drinks Buffet Snack	motorcycle motorcycle scooter motorcycle scooter motorcycle
multiple 0 3.0 1 1.0 2 1.0 3 0.0 4 1.0 45579 0.0		2 1 1 0 1	Snack Meal Drinks Buffet Snack Meal	motorcycle motorcycle scooter motorcycle scooter
multiple 0 3.0 1 1.0 2 1.0 3 0.0 4 1.0 45579 0.0 45580		2 1 1 0 1	Snack Meal Drinks Buffet Snack Meal	motorcycle motorcycle scooter motorcycle scooter motorcycle
multiple 0 3.0 1 1.0 2 1.0 3 0.0 4 1.0 45579 0.0 45580 1.0 45581		2 1 1 0 1 1 0	Snack Meal Drinks Buffet Snack Meal Buffet Drinks	motorcycle motorcycle scooter motorcycle scooter motorcycle motorcycle scooter
multiple 0 3.0 1 1.0 2 1.0 3 0.0 4 1.0 45579 0.0 45580 1.0 45581		2 1 1 0 1 	Snack Meal Drinks Buffet Snack Meal Buffet	motorcycle motorcycle scooter motorcycle scooter motorcycle motorcycle

```
45583
                        2
                                   Snack
                                                  scooter
1.0
      Festival
                                 Time taken (min)
                           City
0
                 Metropolitian
            No
                                                46
1
                 Metropolitian
                                                23
            No
2
                 Metropolitian
                                                21
            No
3
            No
                 Metropolitian
                                                20
4
                 Metropolitian
                                                41
            No
                                               . . .
45579
            No
                 Metropolitian
                                                32
                 Metropolitian
                                                36
45580
            No
45581
            No
                 Metropolitian
                                                16
45582
                 Metropolitian
                                                26
            No
                 Metropolitian
45583
            No
                                                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 \overline{0}rderd
                                 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
                                  228
Festival
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
    Delivery_location_longitude 45584 non-null float64
 7
 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
                                 44384 non-null
    City
                                                 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
у",
"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
                           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
```

```
Order Date
                                         datetime64[ns]
                          45584 non-null
    Time Orderd
9
                          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
    City
                          44384 non-null
18
                                         object
19 Time taken (min)
                          45584 non-null
                                         int64
dtypes: \overline{datetime64[ns](1)}, float64(7), int64(2), object(10)
memory usage: 7.0+ MB
```

Removing dublicate

```
#removing duplicates if any.
df.drop duplicates()
                                   del boy_age
                                                Del boy Ratings \
           ID Delivery person ID
                   DEHRES17DEL01
       0xcdcd
                                          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
                   KNPRES14DEL02
                                          24.0
       0xdb64
                                                             4.7
                   JAPRES04DEL01
45579
       0x7c09
                                          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 \
                 30.327968
                                        78.046106
                                                      30.397968
78.116106
                 10.003064
                                        76.307589
                                                      10.043064
76.347589
                 18.562450
                                        73.916619
                                                     18.652450
74.006619
                 30.899584
                                        75.809346
                                                      30.919584
75.829346
                 26.463504
                                        80.372929
                                                     26.593504
80.502929
. . .
                 26.902328
                                        75.794257
                                                     26.912328
45579
```

75.804	1257								
45580	1237	0.000	0000		6	0.0000	0	0.0700	00
0.0700 45581	000	13.022	20/		26).24243	20	13.0523	0/1
80.272	2439								
45582 77.026	5241	11.001	.753		76	5.98624	1	11.0417	53
45583		23.351	.058		85	.32573	31	23.4310	58
85.405	0/31								
0	Order_Date 2022-12-02	Time_C	orderd 21:55	Time_Or	der_p	icked 22:10	Weath	er_cond	itions \ Fog
1	2022-02-13		14:55			15:05			Stormy
2	2022-04-03 2022-02-13		17:30 09:20			17:40 09:30			storms storms
4	2022-02-14		19:50			20:05		30	Fog
45579	2022-03-24		11:35			11:45			Windy
	2022-02-16 2022-11-03		19:55 23:50		24.	20:10 05:00			Windy Cloudy
45582	2022-07-03		13:35		24.	13:40			Cloudy
45583	2022-02-03		17:10			17:15			Fog
\	traffic_der	nsity	Vehic	le_condi	tion	Type_c	of_ord	er Type	_of_vehicle
0		Jam			2		Sna	ck	motorcycle
1		High			1		Me	al	motorcycle
2	Me	edium			1		Drin	ks	scooter
3		Low			0		Buff	et	motorcycle
4		Jam			1		Sna	ck	scooter
45579		High			1		Me	al	motorcycle
45580		Jam			0		Buff	et	motorcycle
45581		Low			1		Drin	ks	scooter
45582		High			0		Sna	ck	motorcycle
45583	Me	edium			2		Sna	ck	scooter
0	multiple_d	deliver	3.0	No		ropolit	ian	Time_ta	ken (min) 46
1			1.0	No	Metr	ropolit	ian		23

2 3 4		1.0 0.0 1.0	No No No	Metropo Metropo Metropo	olitian	21 20 41
45579 45580 45581 45582 45583		0.0 1.0 0.0 1.0	No No No No No	Metropo Metropo Metropo Metropo	olitian olitian olitian	32 36 16 26 36
[45584	rows x 20 colu	ımns]				
df.desc	ribe()					
count mean min 25% 50% 75% max std	del_boy_age 43730.000000 29.566911 15.000000 25.000000 30.000000 35.000000 50.000000 5.815064		/_Ratings 76.000000 4.633774 1.000000 4.500000 4.700000 4.900000 6.000000 0.334744	Restau	urant_latitud 45584.00000 17.01794 -30.90556 12.93328 18.55144 22.72816 30.91405 8.18567	0 8 2 4 0 3 7
count mean min 25% 50% 75% max std	70. -88. 73. 75. 78. 88.	ngitude 000000 229684 366217 170000 897963 044095 433452 885575	0.03 12.98 18.63 22.78 31.09	_	del_loc_log 45584.000000 70.844161 0.010000 73.280000 76.002574 78.107044 88.563452 21.120578	
mul+inl	.e_deliveries		rder_Date	Vehicl	le_condition	
count	_	\	45584	4	45584.000000	
44591.0 mean 0.74463	2022-04-20 10:	57:49.8	349069824		1.023385	
min 0.00000		2-01-03	00:00:00		0.000000	
25% 0.00000	2022	2-03-04	00:00:00		0.000000	
50%	2022	2-03-24	00:00:00		1.000000	
1.00000 75%	2022	2-05-03	00:00:00		2.000000	
1.00000 max 3.00000	2022	?-12-03	00:00:00		3.000000	

```
NaN
                                                  0.839055
std
0.572510
       Time taken (min)
           45584.000000
count
              26.293963
mean
               10.000000
min
25%
              19.000000
50%
              26.000000
              32.000000
75%
              54.000000
max
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 \
                  19.121999
                                         72.908493
                                                       19.201999
33
72.988493
65
                  12.935662
                                         77.614130
                                                       12.975662
77.654130
                  22.312790
                                         73.170283
                                                       22,422790
111
73.280283
                  18.543626
                                         73.905101
                                                       18.573626
131
73.935101
170
                  26.913483
                                         75.803139
                                                       26.983483
75.873139
. . .
. . .
45499
                  17.455894
                                         78.375467
                                                       17.525894
78.445467
                                                       22.737021
45509
                  22.727021
                                         75.884167
```

75.89	4167	27.164	2024			70.0440	0.5	27 170	2024	
45510 78.054	4095	27.166	9934			78.0440	95	27.176	9934	
45517	2001	26.913	3987			75.7528	91	27.023	3987	
75.862 45557	2891	21.185	5047			72.8085	90	21.205	5047	
72.828	8590									
33 65 111 131 170	Order_Date 2022-06-04 2022-11-03 2022-02-04 2022-09-03 2022-12-03	_	Orderd NaN NaN NaN NaN NaN		_Orde	r_picked 18:35 15:30 18:25 23:40 17:50	Weath	_	ndition: Cloud Nal Nal Storm	y N N y
45499 45509 45510 45517	2022-03-23 2022-03-19 2022-02-13 2022-03-27 2022-03-04		17:55 NaN NaN 20:35			0.75 11:40 09:05 20:45 10:50		341	Fo Fo Na Na Cloud Wind	g N N
	traffic_de	nsity	Vehic	le_cor	nditi	on Type_	of_ord	er		
Type_0 33 scoote	M	\ edium				1	Drin	ks		
65		NaN				1	Drin	ks		
scoote 111		NaN				3	Sna	ck		
elect	ric_scooter	Low				2	Sna	ck		
scoote		edium				0	Drin			
motor		eatum				U	ודוט	K5		
					•		•			•
45499		edium				0	Me	al		
motor 45509		NaN				3	Sna	ck		
elect 45510	ric_scooter	NaN				3	Me	al		
elect	ric_scooter	Jam				2	Sna			
scoote	er									
45557 scoote	er	Low				2	Me	al		
	multiple	بمينة إما	cias F	ectiva	a l		City	Time +	aken (ı	min)
33 65 111	шасстрсе_	ue (IVE)	1.0 0.0 0.0	N N	No Mo	etropoli etropoli etropoli	tian tian	1 TIII 6 _ (Lanell (I	33 19 25

```
131
                        0.0
                                   No
                                                                       19
                                                  NaN
170
                        0.0
                                   No
                                       Metropolitian
                                                                      43
                                  . . .
. . .
45499
                        1.0
                                                Urban
                                                                      36
                                   No
45509
                        0.0
                                   No
                                       Metropolitian
                                                                      25
                                                                      17
45510
                        1.0
                                                Urban
                                   No
45517
                        0.0
                                   No
                                       Metropolitian
                                                                      48
45557
                        0.0
                                                                      12
                                   No
                                                Urban
[1854 rows x 20 columns]
df.groupby(["Delivery person ID"])["Time taken (min)"].count()
Delivery person ID
AGRRES010DEL01
                   14
AGRRES010DEL02
                   15
AGRRES010DEL03
                   13
                   12
AGRRES01DEL01
AGRRES01DEL02
                   14
VADRES19DEL02
                   59
VADRES19DEL03
                   41
                   59
VADRES20DEL01
VADRES20DEL02
                   50
                   37
VADRES20DEL03
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]
         DEH
1
         K<sub>0</sub>C
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 \
                    DEHRES17DEL01
0
       0xcdcd
                                            36.0
                                                               4.2
```

1 2 3 4	0xd987 0x2784 0xc8b6 0xdb64	KOCRES16 PUNERES13 LUDHRES15 KNPRES14	DEL03 DEL02	21.0 23.0 34.0 24.0	4. 4. 4. 4.	7 3
45579 45580 45581 45582 45583	0xd641 0x4f8d 0x5eee	JAPRES04 AGRRES16 CHENRES08 COIMBRES11 RANCHIRES09	DEL01 DEL03 DEL01	30.0 21.0 30.0 20.0 23.0	4.8 4.0 4.9 4.9	3 5 9 7
		nt_latitude	Restaui	rant_longitude	del_loc_lat	
del_l 0	oc_log \	30.327968		78.046106	30.397968	
78.11 1	6106	10.003064		76.307589	10.043064	
76.34 2	7589	18.562450		73.916619	18.652450	
74.00	6619					
3 75.82	9346	30.899584		75.809346	30.919584	
4 80.50	2929	26.463504		80.372929	26.593504	
45579		26.902328		75.794257	26.912328	
75.80 45580		0.000000		0.00000	0.070000	
0.070 45581		13.022394		80.242439	13.052394	
80.27 45582		11.001753		76.986241	11.041753	
77.02	6241					
45583 85.40		23.351058		85.325731	23.431058	
		e Time_Order	d V	Weather_condition	ons	
traff 0	ic_density 2022-12-02		5	F	og	Jam
1	2022-02-13	3 14:5	5	Stor	^my	High
2	2022-04-03	3 17:3	9	Sandsto	^ms	Medium
3	2022-02-13	3 09:2	9	Sandsto	^ms	Low
4	2022-02-14	19:5	0	F	og	Jam
		2313			- 3	
		•••				

45579 2022-03-24	11:35		Windy	High
45580 2022-02-16	19:55		Windy	Jam
45581 2022-11-03	23:50		Cloudy	Low
45501 2022-11-05	23.30		Cloudy	LOW
45582 2022-07-03	13:35		Cloudy	High
45583 2022-02-03	17:10		Fog	Medium
Vehicle_cor		of_order Typ	oe_of_vehicle	
0	2	Snack	motorcycle	
3.0 1	1	Meal	motorcycle	
1.0	1	Drinks	scooter	
1.0				
3 0.0	0	Buffet	motorcycle	
4	1	Snack	scooter	
1.0				
45570	1	Maa 1	ma+amaala	
45579 0.0	1	Meal	motorcycle	
45580	Θ	Buffet	motorcycle	
1.0 45581	1	Drinks	scooter	
0.0	0	Canal	ma+amaala	
45582 1.0	Θ	Snack	motorcycle	
45583	2	Snack	scooter	
1.0	_		500010.	
Festival		Time_taken		
0 No	Metropolitian		46 23	DEH KOC
1 No 2 No 3 No	Metropolitian Metropolitian		23	PUN
3 No	Metropolitian		20	LUD
4 No	Metropolitian		41	KNP
45579 No	${\tt Metropolitian}$		32	JAP
45580 No	Metropolitian		36	AGR
45581 No	Metropolitian		16	CHE
45582 No	Metropolitian Metropolitian		26 36	COI RAN
45583 No			36	NAIV
[45584 rows x 21	co cumins J			

```
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
                                           obiect
 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
     city name
                           45584 non-null object
 20
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
                                  del boy age
           ID Delivery person ID
                                               Del boy Ratings \
       0xcdcd
0
                   DEHRES17DEL01
                                         36.0
                                                           4.2
                                                           4.7
1
       0xd987
                   KOCRES16DEL01
                                         21.0
2
                                         23.0
                                                           4.7
       0x2784
                  PUNERES13DEL03
3
       0xc8b6
                                         34.0
                                                           4.3
                  LUDHRES15DEL02
4
       0xdb64
                   KNPRES14DEL02
                                         24.0
                                                           4.7
                                                            . . .
                   JAPRES04DEL01
45579
      0x7c09
                                         30.0
                                                           4.8
                                                           4.6
45580
      0xd641
                   AGRRES16DEL01
                                         21.0
                                                           4.9
45581
       0x4f8d
                  CHENRES08DEL03
                                         30.0
                                                           4.7
45582
      0x5eee
                 COIMBRES11DEL01
                                         20.0
                RANCHIRES09DEL02
                                         23.0
                                                           4.9
45583
      0x5fb2
       Restaurant latitude Restaurant longitude del loc lat
del loc log \
```

0 78.116	5106	30.327968		78.046106	30.397968	
1		10.003064		76.307589	10.043064	
76.347 2		18.562450		73.916619	18.652450	
74.006 3	5619	30.899584		75.809346	30.919584	
75.829 4	9346	26.463504		80.372929	26.593504	
80.502	2929					
				•••		
45579 75.804	1257	26.902328		75.794257	26.912328	
45580		0.000000		0.00000	0.070000	
0.0700 45581		13.022394		80.242439	13.052394	
80.272 45582		11.001753		76.986241	11.041753	
77.026 45583 85.405		23.351058		85.325731	23.431058	
	Order_Date	Time_Orderd	V	Veather_condition	ons traffic_	_density
0	2022-12-02	21.55		F	og	Jam
1	2022-02-13	14.55		Stor	my	High
2	2022-04-03	17.30		Sandstor	ms	Medium
3	2022-02-13	9.20		Sandstor	ms	Low
4	2022-02-14	19.50		F	og	Jam
45579	2022-03-24	11.35		Wir	ndy	High
45580	2022-02-16	19.55		Wir	ndy	Jam
45581	2022-11-03	23.50		Clou	ıdy	Low
45582	2022-07-03	13.35		Clou	ıdy	High
45583	2022-02-03	17.10		F	og	Medium
multip	Vehicle_co ole_deliver:		of_ord	der Type_of_vehi	.cle	

0			2		Snack	mot	corcycle	
3.0 1			1		Meal	mot	orcycle	
1.0							-	
2 1.0			1		Drinks		scooter	
3			0		Buffet	mot	orcycle	
0.0 4			1		Snack		scooter	
1.0			1		Silack		30000	
45579			1		Meal	mot	orcycle	
0.0							_	
45580 1.0			0		Buffet	mot	orcycle	
45581			1		Drinks		scooter	
0.0 45582			0		Snack	mot	orcycle	
1.0						iiio (orcyctc	
45583 1.0			2		Snack		scooter	
1.0								
0	Festi	val. No	Metropoli		Time_taker	n (min) 46		name DEH
1		No	Metropoli	tian		23		K0C
2		No No	Metropoli Metropoli			21 20		PUN LUD
1		No	Metropoti			41		KNP
		 No	Motropoli					
45579 45580		No No	Metropoli Metropoli			32 36		JAP AGR
45581		No	Metropoli	tian		16	5	CHE
15582 15583		No No	Metropoli Metropoli			26 36		COI RAN
			•					
45584	rows	X 21	columns]					
lf["Tin	ne_0rd	ler_p	icked"]					
)		22:1						
1		15:05 17:40						
2 3 4		09:3	-					
1		20:0	5					
45579		11:4						
45580 45581	24.	20:10						
45581 45582	24:	05:00 13:40						

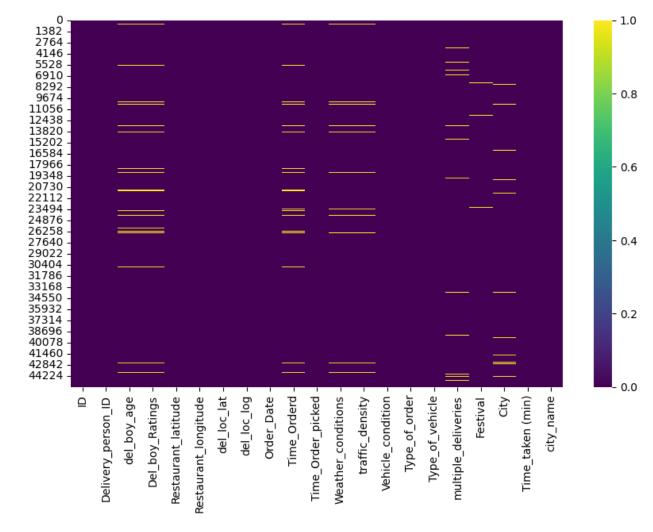
```
45583
           17:15
Name: Time Order picked, Length: 45584, dtype: object
# function to remove seconds from the column
def x(i):
   i = str(i)
   if len(i)>5:
       i=i.replace(":",".")
       i = i[:-3]
        return i
   else:
        i=i.replace(":",".")
        return i
df["Time Order picked"] = 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
                          45584 non-null
    ID
                                          object
 1
    Delivery person ID
                          45584 non-null object
 2
    del boy age
                          43730 non-null float64
 3
                          43676 non-null float64
    Del_boy_Ratings
 4
                          45584 non-null float64
    Restaurant_latitude
 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
                                          obiect
 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
                          44591 non-null
 16 multiple deliveries
                                          float64
 17 Festival
                          45356 non-null
                                          object
                          44384 non-null
 18 City
                                          object
 19 Time taken (min)
                          45584 non-null
                                          int64
                          45584 non-null object
 20 city name
```

```
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()
```

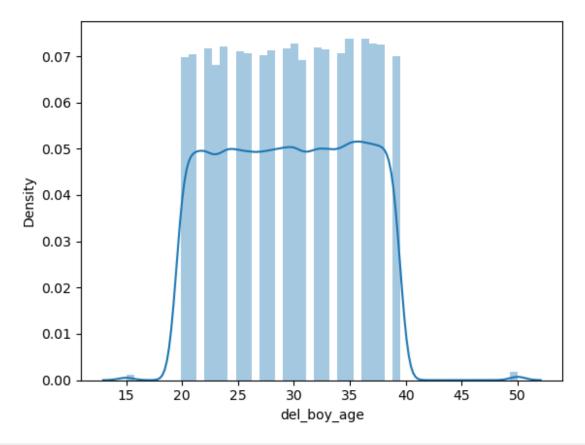


```
null_1 = df[df["del_boy_age"].isnull() &
df["Del_boy_Ratings"].isnull()]
```

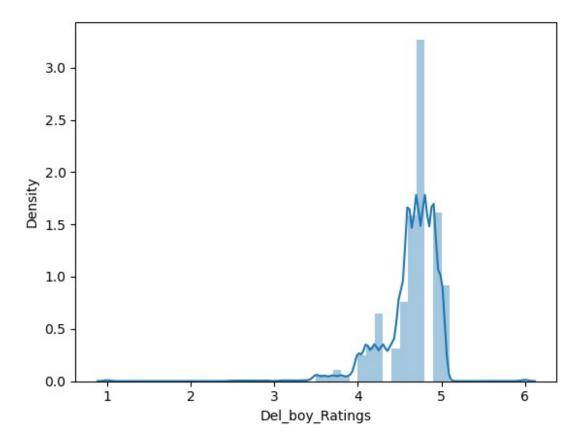
```
null 2 = df[df["traffic density"].isnull() &
df["Weather conditions"].isnull()]
null 1
           ID Delivery_person_ID
                                    del_boy_age
                                                 Del_boy_Ratings \
33
        0x4f0
                    MUMRES17DEL02
                                            NaN
                                                              NaN
        0xa9f
65
                   BANGRES13DEL01
                                            NaN
                                                              NaN
111
        0x6e2
                    VADRES06DEL01
                                            NaN
                                                              NaN
131
       0x1ec7
                   PUNERES05DEL03
                                            NaN
                                                              NaN
170
       0x1b58
                    JAPRES03DEL01
                                            NaN
                                                              NaN
                    HYDRES06DEL01
45499
       0x426b
                                            NaN
                                                              NaN
45509
        0x863
                   INDORES05DEL01
                                            NaN
                                                              NaN
       0xd0be
45510
                    AGRRES07DEL01
                                            NaN
                                                              NaN
45517
       0x1ed1
                    JAPRES18DEL01
                                            NaN
                                                              NaN
                    SURRES06DEL02
                                            NaN
45557
       0x1163
                                                              NaN
       Restaurant latitude Restaurant longitude del loc lat
del loc log
                  19.121999
                                         72.908493
                                                       19.201999
33
72.988493
                  12.935662
                                         77.614130
                                                       12.975662
77.654130
                  22.312790
                                         73.170283
111
                                                       22.422790
73.280283
131
                  18.543626
                                         73.905101
                                                       18.573626
73.935101
170
                  26.913483
                                         75.803139
                                                       26.983483
75.873139
. . .
. . .
                                         78.375467
45499
                  17.455894
                                                       17.525894
78.445467
                  22.727021
                                         75.884167
                                                       22.737021
45509
75.894167
                  27.160934
                                                       27.170934
                                         78.044095
45510
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		9	Stormy	Low
170	2022-12-03		NaN		Sand	storms	Medium
45499	2022-03-23	1	.7.55			Fog	Medium
45509	2022-03-19		NaN			NaN	NaN
45510	2022-02-13		NaN			NaN	NaN
45517	2022-03-27	2	0.35		(Cloudy	Jam
45557	2022-03-04		NaN			Windy	Low
multi	Vehicle_com ple deliver:		Type_	of_order	Type_of	_vehicle	
33 1.0	p to_uc t1vc1.	1		Drinks		scooter	
65		1		Drinks		scooter	
0.0 111		3		Snack	electric	_scooter	
0.0 131		2		Snack		scooter	
0.0 170		0		Drinks	mo ⁻	torcycle	
0.0							
 45499		0		Meal	mo ⁻	torcycle	
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							
33 65 111 131 170	Festival No No No No No	Metropol Metropol Metropol	itian itian itian NaN	Time_tak	en (min) 33 19 25 19 43	city_name MUM BAN VAD PUN JAP	

```
45499
             No
                          Urban
                                               36
                                                          HYD
45509
             No
                  Metropolitian
                                               25
                                                          IND
45510
             No
                          Urban
                                               17
                                                          AGR
45517
                  Metropolitian
                                               48
             No
                                                          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'>
```

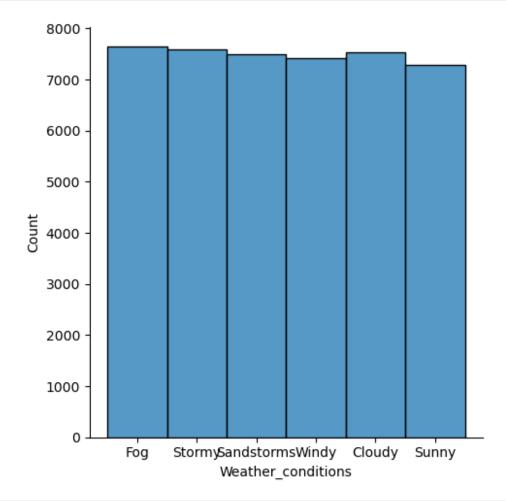


The null values in del_boy_age can be filled with median and del_boy_rating can be filled with mode values

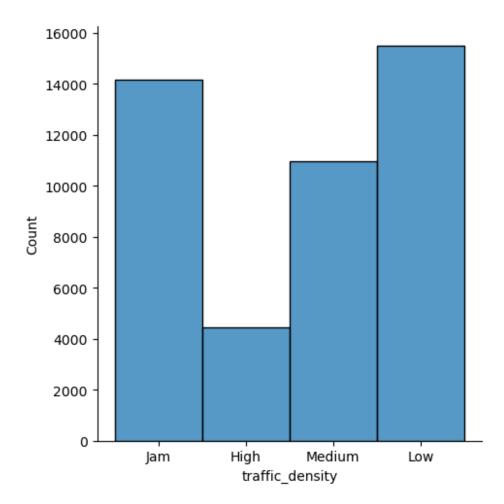
```
df['del_boy_age'].fillna(df['del_boy_age'].median(), inplace=True)
df['del_boy_age'].isnull().sum()
0
df['Del_boy_Ratings'].fillna(df['Del_boy_Ratings'].mode()[0],
inplace=True)
df["Del_boy_Ratings"].isnull().sum()
0
```

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>
```



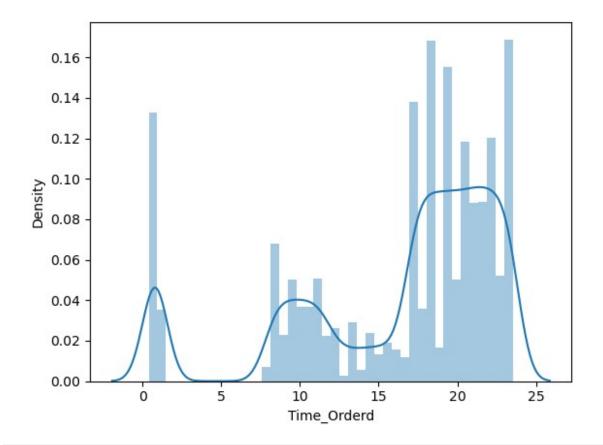
```
df['traffic_density'].value_counts()
sns.displot(df['traffic_density'])
<seaborn.axisgrid.FacetGrid at 0x20acab5de20>
```



```
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'>
```

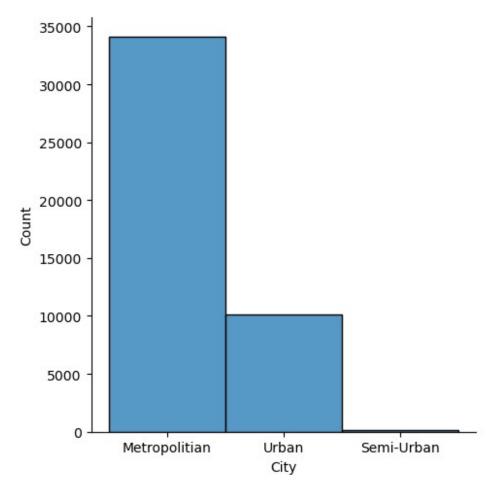


df['Time_Orderd'].fillna(df['Time_Orderd'].median(), inplace=True)

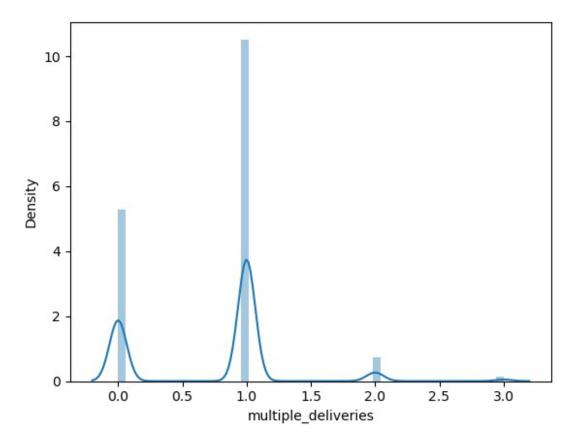
filling futhur null values

sns.displot(df["City"])

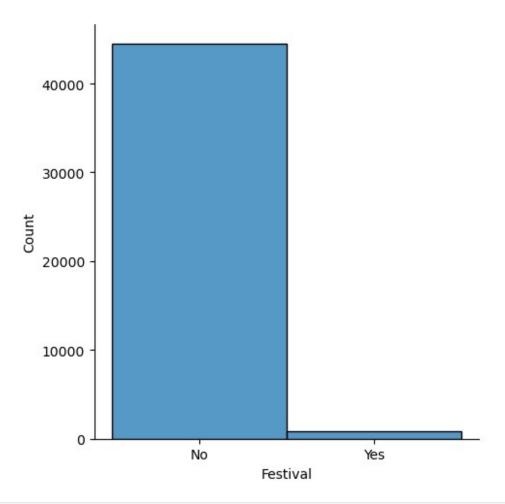
<seaborn.axisgrid.FacetGrid at 0x20acd107a40>



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

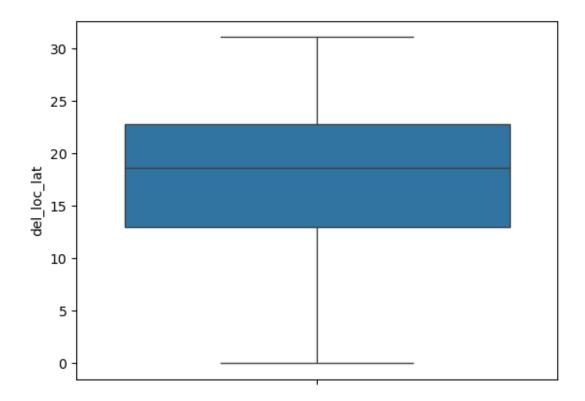


```
df['multiple_deliveries'].fillna(df['multiple_deliveries'].median(),
inplace=True)
sns.displot(df["Festival"])
<seaborn.axisgrid.FacetGrid at 0x20ac98a81a0>
```



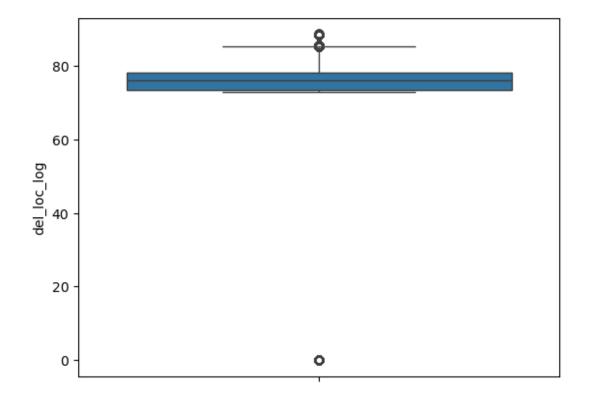
```
df['Festival'].fillna(df['Festival'].mode()[0], inplace=True)
df.isnull().sum()
                           0
ID
Delivery_person_ID
                           0
del_boy_age
                           0
Del_boy_Ratings
                           0
Restaurant latitude
                           0
Restaurant_longitude
                           0
del_loc_lat
                           0
del_loc_log
                           0
Order_Date
                           0
Time_{\overline{0}}rderd
                           0
Time_Order_picked Weather_conditions
                           0
                           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
                        0
city name
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
                           45584 non-null
     Restaurant_latitude
                                            float64
 5
     Restaurant longitude
                           45584 non-null
                                            float64
 6
     del_loc_lat
                           45584 non-null
                                            float64
 7
                           45584 non-null
                                            float64
     del loc log
 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
    Citv
                           45584 non-null
                                            obiect
19
    Time taken (min)
                           45584 non-null
                                            int64
    city name
                           45584 non-null object
 20
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 \
               DEHRES17DEL01
   0xcdcd
                                       36.0
                                                          4.2
1
  0xd987
               KOCRES16DEL01
                                       21.0
                                                          4.7
2
                                       23.0
  0x2784
              PUNERES13DEL03
                                                          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
/
             30.327968
                                     78.046106
                                                  30.397968
                                                                78.116106
0
             10.003064
                                     76.307589
                                                  10.043064
                                                                76.347589
1
2
             18.562450
                                     73.916619
                                                  18.652450
                                                                74.006619
3
             30.899584
                                     75.809346
                                                  30.919584
                                                                75.829346
             26.463504
                                     80.372929
                                                  26.593504
                                                                80.502929
                                 Weather conditions traffic density \
  Order Date
              Time Orderd
0 2022-12-02
                     21.55
                                                 Fog
                                                                  Jam
1 2022-02-13
                     14.55
                                              Stormy
                                                                 High
                     17.30
2 2022-04-03
                                          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
                               Meal
                                          motorcycle
                                                                      1.0
2
                                                                      1.0
                             Drinks
                                             scooter
3
                             Buffet
                                                                      0.0
                                          motorcycle
                              Snack
                                             scooter
                                                                      1.0
   Festival
                       City Time taken (min)
                                               city name
             Metropolitian
0
         No
                                           46
                                                      DEH
                                           23
1
             Metropolitian
                                                      K0C
         No
2
         No
             Metropolitian
                                           21
                                                      PUN
3
         No
             Metropolitian
                                           20
                                                      LUD
4
             Metropolitian
                                           41
                                                      KNP
         No
[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)
   %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()
```