accident_severity_project

May 31, 2024

1 Importing libraries

[91]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
```

2 Loading datasets

[92]:
```
#import data from drive
from google.colab import drive
drive.mount('/content/drive')
```

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call
drive.mount("/content/drive", force_remount=True).
```

[93]:
```
#creating first dataframe
df_acc=pd.read_csv('/content/drive/MyDrive/accident_data.csv')
```

[94]:
```
#creating secong dataframe
df_veh=pd.read_csv('/content/drive/MyDrive/vehicle_data.csv',encoding='latin1')
```

[95]:
```
df_acc
```

[95]:
```
        Accident_Index 1st_Road_Class  1st_Road_Number 2nd_Road_Class  \
0        200501BS00001              A           3218.0            NaN
1        200501BS00002              B            450.0              C
2        200501BS00003              C              0.0            NaN
3        200501BS00004              A           3220.0            NaN
4        200501BS00005            NaN              0.0            NaN
…                  …             …                …              …
1048570  201091NM01760              A             95.0            NaN
1048571  201091NM01881              A             95.0            NaN
1048572  201091NM01935              A             96.0   Unclassified
```

```
1048573  201091NM01964              A              9.0   Unclassified
1048574  201091NM02142            NaN              0.0   Unclassified

         2nd_Road_Number Accident_Severity   Carriageway_Hazards        Date  \
0                    0.0           Serious                   NaN  04/01/2005   
1                    0.0            Slight                   NaN  05/01/2005   
2                    0.0            Slight                   NaN  06/01/2005   
3                    0.0            Slight                   NaN  07/01/2005   
4                    0.0            Slight                   NaN  10/01/2005   
…                  …               …                   …         …   
1048570              0.0            Slight                   NaN  18/02/2010   
1048571              0.0            Slight                   NaN  21/02/2010   
1048572              0.0            Slight                   NaN  23/02/2010   
1048573              0.0           Serious                   NaN  23/02/2010   
1048574              0.0           Serious  Other object on road  28/02/2010   

        Day_of_Week  Did_Police_Officer_Attend_Scene_of_Accident  …  \
0           Tuesday                                          1.0  …   
1         Wednesday                                          1.0  …   
2          Thursday                                          1.0  …   
3            Friday                                          1.0  …   
4            Monday                                          1.0  …   
…             …                                          …  …   
1048570    Thursday                                          1.0  …   
1048571      Sunday                                          1.0  …   
1048572     Tuesday                                          1.0  …   
1048573     Tuesday                                          1.0  …   
1048574      Sunday                                          1.0  …   

                Police_Force Road_Surface_Conditions           Road_Type  \
0        Metropolitan Police             Wet or damp  Single carriageway   
1        Metropolitan Police                     Dry    Dual carriageway   
2        Metropolitan Police                     Dry  Single carriageway   
3        Metropolitan Police                     Dry  Single carriageway   
4        Metropolitan Police             Wet or damp  Single carriageway   
…                      …                     …                 …   
1048570             Northern                     Dry  Single carriageway   
1048571             Northern            Frost or ice  Single carriageway   
1048572             Northern            Frost or ice  Single carriageway   
1048573             Northern             Wet or damp  Single carriageway   
1048574             Northern             Wet or damp    Dual carriageway   

        Special_Conditions_at_Site Speed_limit   Time  Urban_or_Rural_Area  \
0                              NaN          30  17:42                Urban   
1                              NaN          30  17:36                Urban   
2                              NaN          30  00:15                Urban   
3                              NaN          30  10:35                Urban   
```

```
4                              NaN         30  21:13               Urban
…                            …         …    …                 …
1048570                        NaN         60  07:00               Rural
1048571                        NaN         60  03:00               Rural
1048572                        NaN         30  09:38               Rural
1048573                        NaN         60  18:25               Rural
1048574                        NaN         60  15:45               Rural

            Weather_Conditions  Year InScotland
0        Raining no high winds  2005         No
1           Fine no high winds  2005         No
2           Fine no high winds  2005         No
3           Fine no high winds  2005         No
4           Fine no high winds  2005         No
…                         …   …         …
1048570     Fine no high winds  2010        Yes
1048571     Fine no high winds  2010        Yes
1048572     Fine no high winds  2010        Yes
1048573     Fine no high winds  2010        Yes
1048574  Snowing no high winds  2010        Yes

[1048575 rows x 34 columns]
```

[96]:
```
df_veh
```

[96]:
```
        Accident_Index Age_Band_of_Driver  Age_of_Vehicle  \
0         200401BS00001            26 - 35             3.0
1         200401BS00002            26 - 35             NaN
2         200401BS00003            26 - 35             4.0
3         200401BS00003            66 - 75             NaN
4         200401BS00004            26 - 35             1.0
…                    …                 …             …
2177200   2016984131116            21 - 25            14.0
2177201   2016984131116            56 - 65             NaN
2177202   2016984131216            56 - 65             NaN
2177203   2016984131316            16 - 20            13.0
2177204   2016984133416            46 - 55             5.0

                Driver_Home_Area_Type  Driver_IMD_Decile  \
0                          Urban area                4.0
1                          Urban area                3.0
2        Data missing or out of range                NaN
3        Data missing or out of range                NaN
4                          Urban area                4.0
…                                 …                …
2177200                    Urban area                NaN
2177201                    Small town                NaN
```

```
2177202                    Urban area                3.0
2177203                    Urban area                6.0
2177204                    Urban area                NaN

         Engine_Capacity_.CC. Hit_Object_in_Carriageway  \
0                      1588.0                       NaN
1                         NaN                       NaN
2                       998.0                       NaN
3                         NaN                       NaN
4                       124.0                       NaN
…                       …                       …
2177200                1598.0                       NaN
2177201                1598.0                       NaN
2177202                   NaN                       NaN
2177203                1796.0                       NaN
2177204                2184.0                       NaN

        Hit_Object_off_Carriageway     Journey_Purpose_of_Driver  \
0                              NaN  Data missing or out of range
1                              NaN  Data missing or out of range
2                              NaN  Data missing or out of range
3                              NaN  Data missing or out of range
4                              NaN  Data missing or out of range
…                            …                           …
2177200                        NaN                     Not known
2177201                        NaN        Commuting to/from work
2177202      Central crash barrier       Journey as part of work
2177203                       Tree                     Not known
2177204              Wall or fence                     Not known

                                    Junction_Location  …  \
0                        Data missing or out of range  …
1                        Data missing or out of range  …
2                        Data missing or out of range  …
3                        Data missing or out of range  …
4                        Data missing or out of range  …
…                                               …  …
2177200  Mid Junction - on roundabout or on main road  …
2177201  Mid Junction - on roundabout or on main road  …
2177202        Not at or within 20 metres of junction  …
2177203        Not at or within 20 metres of junction  …
2177204        Not at or within 20 metres of junction  …

        Skidding_and_Overturning Towing_and_Articulation  \
0                            NaN     No tow/articulation
1                            NaN     No tow/articulation
2                            NaN     No tow/articulation
```

```
3                            NaN     No tow/articulation
4                            NaN     No tow/articulation
…                          …                     …
2177200                      NaN     No tow/articulation
2177201                      NaN     No tow/articulation
2177202                      NaN     No tow/articulation
2177203                  Skidded     No tow/articulation
2177204   Skidded and overturned     No tow/articulation

               Vehicle_Leaving_Carriageway Vehicle_Location.Restricted_Lane  \
0                Did not leave carriageway                              0.0
1                Did not leave carriageway                              0.0
2                Did not leave carriageway                              0.0
3                Did not leave carriageway                              0.0
4                Did not leave carriageway                              0.0
…                                    …                              …
2177200          Did not leave carriageway                              0.0
2177201                            Offside                              0.0
2177202  Offside on to central reservation                              0.0
2177203                            Offside                              0.0
2177204                           Nearside                              0.0

         Vehicle_Manoeuvre Vehicle_Reference                 Vehicle_Type  \
0        Going ahead other                 2                          109
1        Going ahead other                 1                          109
2            Turning right                 1                          109
3        Going ahead other                 2                          109
4        Going ahead other                 1   Motorcycle 125cc and under
…                    …               …                             …
2177200  Going ahead other                 1                          Car
2177201  Going ahead other                 2                          Car
2177202  Going ahead other                 1  Goods 7.5 tonnes mgw and over
2177203  Going ahead other                 1                          Car
2177204  Going ahead other                 1                          Car

          Was_Vehicle_Left_Hand_Drive X1st_Point_of_Impact  Year
0        Data missing or out of range                Front  2004
1        Data missing or out of range                Front  2004
2        Data missing or out of range                Front  2004
3        Data missing or out of range                Front  2004
4        Data missing or out of range                Front  2004
…                               …                  …   …
2177200                            No                Front  2016
2177201                            No                Front  2016
2177202                            No              Offside  2016
2177203                            No                Front  2016
2177204                            No             Nearside  2016
```

```
[2177205 rows x 24 columns]
```

[97]:

```
#merging datas
df=pd.merge(df_veh, df_acc, on='Accident_Index')
df
```

[97]:

```
         Accident_Index            Age_Band_of_Driver  Age_of_Vehicle  \
0         200501BS00002                       36 - 45             3.0
1         200501BS00003                       26 - 35             5.0
2         200501BS00004                       46 - 55             4.0
3         200501BS00005                       46 - 55            10.0
4         200501BS00006                       46 - 55             1.0
…                   …                             …             …
528330    201034NK23180                       36 - 45             3.0
528331    201034NK23410  Data missing or out of range             NaN
528332    201034NK23410                       36 - 45             8.0
528333    201034NK23890                       56 - 65             8.0
528334    201034NK23890                       16 - 20             7.0

                Driver_Home_Area_Type  Driver_IMD_Decile  Engine_Capacity_.CC.  \
0        Data missing or out of range                NaN                8268.0
1                          Urban area                3.0                8300.0
2                          Urban area                1.0                1769.0
3        Data missing or out of range                NaN                  85.0
4                          Urban area                4.0                2976.0
…                                   …                 …                    …
528330                     Urban area                1.0                1248.0
528331   Data missing or out of range                NaN                   NaN
528332                     Urban area                3.0                1360.0
528333                     Urban area                8.0                1951.0
528334                     Small town                4.0                1587.0

         Hit_Object_in_Carriageway Hit_Object_off_Carriageway  \
0                              NaN                        NaN
1                   Parked vehicle                        NaN
2                              NaN                        NaN
3                             Kerb                        NaN
4                              NaN                        NaN
…                             …                         …
528330                         NaN                        NaN
528331                         NaN                        NaN
528332                         NaN                        NaN
528333                         NaN                        NaN
528334                         NaN                        NaN

          Journey_Purpose_of_Driver  \
0         Journey as part of work
1         Journey as part of work
2       Other/Not known (2005-10)
3       Other/Not known (2005-10)
4       Other/Not known (2005-10)
…                           …
528330    Journey as part of work
528331    Journey as part of work
528332  Other/Not known (2005-10)
528333  Other/Not known (2005-10)
528334  Other/Not known (2005-10)

                                        Junction_Location  …  \
0                                      Leaving roundabout  …
1                  Not at or within 20 metres of junction  …
2                  Not at or within 20 metres of junction  …
3                  Not at or within 20 metres of junction  …
4                  Not at or within 20 metres of junction  …
…                                                   …  …
528330  Approaching junction or waiting/parked at junc…  …
528331       Mid Junction - on roundabout or on main road  …
528332       Mid Junction - on roundabout or on main road  …
528333       Mid Junction - on roundabout or on main road  …
528334       Mid Junction - on roundabout or on main road  …

               Police_Force Road_Surface_Conditions          Road_Type  \
0       Metropolitan Police                    Dry   Dual carriageway
1       Metropolitan Police                    Dry  Single carriageway
2       Metropolitan Police                    Dry  Single carriageway
3       Metropolitan Police             Wet or damp  Single carriageway
4       Metropolitan Police             Wet or damp  Single carriageway
…                     …                     …                 …
528330     Northamptonshire                    Dry  Single carriageway
528331     Northamptonshire             Wet or damp  Single carriageway
528332     Northamptonshire             Wet or damp  Single carriageway
528333     Northamptonshire                    Dry  Single carriageway
528334     Northamptonshire                    Dry  Single carriageway

       Special_Conditions_at_Site Speed_limit   Time Urban_or_Rural_Area  \
0                             NaN          30  17:36               Urban
1                             NaN          30  00:15               Urban
2                             NaN          30  10:35               Urban
3                             NaN          30  21:13               Urban
4                   Oil or diesel          30  12:40               Urban
…                           …         …    …                 …
528330                        NaN          30  10:20               Urban
528331                        NaN          30  12:00               Rural
```

```
528332                        NaN          30  12:00               Rural
528333                        NaN          30  13:07               Urban
528334                        NaN          30  13:07               Urban

           Weather_Conditions Year_y  InScotland
0          Fine no high winds   2005          No
1          Fine no high winds   2005          No
2          Fine no high winds   2005          No
3          Fine no high winds   2005          No
4       Raining no high winds   2005          No
…                       …    …         …
528330     Fine no high winds   2010          No
528331  Raining no high winds   2010          No
528332  Raining no high winds   2010          No
528333  Raining no high winds   2010          No
528334  Raining no high winds   2010          No

[528335 rows x 57 columns]
```

3 Understanding the Dataset

[98]:

```
df.head()
```

[98]:

```
  Accident_Index Age_Band_of_Driver  Age_of_Vehicle  \
0  200501BS00002            36 - 45             3.0
1  200501BS00003            26 - 35             5.0
2  200501BS00004            46 - 55             4.0
3  200501BS00005            46 - 55            10.0
4  200501BS00006            46 - 55             1.0

          Driver_Home_Area_Type  Driver_IMD_Decile  Engine_Capacity_.CC.  \
0  Data missing or out of range                NaN                8268.0
1                    Urban area                3.0                8300.0
2                    Urban area                1.0                1769.0
3  Data missing or out of range                NaN                  85.0
4                    Urban area                4.0                2976.0

  Hit_Object_in_Carriageway Hit_Object_off_Carriageway  \
0                       NaN                        NaN
1            Parked vehicle                        NaN
2                       NaN                        NaN
3                      Kerb                        NaN
4                       NaN                        NaN

   Journey_Purpose_of_Driver                       Junction_Location  …  \
0    Journey as part of work                      Leaving roundabout  …
```

```
1    Journey as part of work  Not at or within 20 metres of junction  …
2  Other/Not known (2005-10)  Not at or within 20 metres of junction  …
3  Other/Not known (2005-10)  Not at or within 20 metres of junction  …
4  Other/Not known (2005-10)  Not at or within 20 metres of junction  …

          Police_Force Road_Surface_Conditions           Road_Type  \
0  Metropolitan Police                     Dry    Dual carriageway
1  Metropolitan Police                     Dry  Single carriageway
2  Metropolitan Police                     Dry  Single carriageway
3  Metropolitan Police             Wet or damp  Single carriageway
4  Metropolitan Police             Wet or damp  Single carriageway

  Special_Conditions_at_Site Speed_limit   Time Urban_or_Rural_Area  \
0                        NaN          30  17:36               Urban
1                        NaN          30  00:15               Urban
2                        NaN          30  10:35               Urban
3                        NaN          30  21:13               Urban
4              Oil or diesel          30  12:40               Urban

      Weather_Conditions Year_y  InScotland
0     Fine no high winds   2005          No
1     Fine no high winds   2005          No
2     Fine no high winds   2005          No
3     Fine no high winds   2005          No
4  Raining no high winds   2005          No

[5 rows x 57 columns]
```

[99]:
```python
df.tail()
```

[99]:
```
       Accident_Index            Age_Band_of_Driver  Age_of_Vehicle  \
528330  201034NK23180                       36 - 45             3.0
528331  201034NK23410  Data missing or out of range             NaN
528332  201034NK23410                       36 - 45             8.0
528333  201034NK23890                       56 - 65             8.0
528334  201034NK23890                       16 - 20             7.0

               Driver_Home_Area_Type  Driver_IMD_Decile  Engine_Capacity_.CC.  \
528330                    Urban area                1.0                1248.0
528331  Data missing or out of range                NaN                   NaN
528332                    Urban area                3.0                1360.0
528333                    Urban area                8.0                1951.0
528334                    Small town                4.0                1587.0

       Hit_Object_in_Carriageway Hit_Object_off_Carriageway  \
528330                       NaN                        NaN
528331                       NaN                        NaN
```

```
528332                       NaN                        NaN
528333                       NaN                        NaN
528334                       NaN                        NaN

        Journey_Purpose_of_Driver  \
528330    Journey as part of work
528331    Journey as part of work
528332  Other/Not known (2005-10)
528333  Other/Not known (2005-10)
528334  Other/Not known (2005-10)

                                        Junction_Location  …  \
528330  Approaching junction or waiting/parked at junc…  …
528331       Mid Junction - on roundabout or on main road  …
528332       Mid Junction - on roundabout or on main road  …
528333       Mid Junction - on roundabout or on main road  …
528334       Mid Junction - on roundabout or on main road  …

            Police_Force Road_Surface_Conditions           Road_Type  \
528330  Northamptonshire                     Dry  Single carriageway
528331  Northamptonshire             Wet or damp  Single carriageway
528332  Northamptonshire             Wet or damp  Single carriageway
528333  Northamptonshire                     Dry  Single carriageway
528334  Northamptonshire                     Dry  Single carriageway

       Special_Conditions_at_Site Speed_limit   Time Urban_or_Rural_Area  \
528330                        NaN          30  10:20               Urban
528331                        NaN          30  12:00               Rural
528332                        NaN          30  12:00               Rural
528333                        NaN          30  13:07               Urban
528334                        NaN          30  13:07               Urban

           Weather_Conditions Year_y  InScotland
528330     Fine no high winds   2010          No
528331  Raining no high winds   2010          No
528332  Raining no high winds   2010          No
528333  Raining no high winds   2010          No
528334  Raining no high winds   2010          No

[5 rows x 57 columns]
```

[100]:

```
df.isna().sum()
```

[100]:

```
Accident_Index                                      0
Age_Band_of_Driver                                  0
Age_of_Vehicle                                  69478
Driver_Home_Area_Type                               0
```

```
Driver_IMD_Decile                              115674
Engine_Capacity_.CC.                            45216
Hit_Object_in_Carriageway                      506767
Hit_Object_off_Carriageway                     482965
Journey_Purpose_of_Driver                           0
Junction_Location                                   0
make                                                0
model                                           75618
Propulsion_Code                                 37243
Sex_of_Driver                                       0
Skidding_and_Overturning                       456926
Towing_and_Articulation                             0
Vehicle_Leaving_Carriageway                         0
Vehicle_Location.Restricted_Lane                   11
Vehicle_Manoeuvre                                   0
Vehicle_Reference                                   0
Vehicle_Type                                        0
Was_Vehicle_Left_Hand_Drive                         0
X1st_Point_of_Impact                                0
Year_x                                              0
1st_Road_Class                                 130079
1st_Road_Number                                     1
2nd_Road_Class                                 204469
2nd_Road_Number                                  4490
Accident_Severity                                   0
Carriageway_Hazards                            519593
Date                                                0
Day_of_Week                                         0
Did_Police_Officer_Attend_Scene_of_Accident        66
Junction_Control                                    0
Junction_Detail                                     0
Latitude                                           23
Light_Conditions                                 1247
Local_Authority_(District)                          0
Local_Authority_(Highway)                           0
Location_Easting_OSGR                              23
Location_Northing_OSGR                             23
Longitude                                          23
LSOA_of_Accident_Location                       24306
Number_of_Casualties                                0
Number_of_Vehicles                                  0
Pedestrian_Crossing-Human_Control                  16
Pedestrian_Crossing-Physical_Facilities             8
Police_Force                                        0
Road_Surface_Conditions                           728
Road_Type                                        2994
Special_Conditions_at_Site                     514105
```

```
Speed_limit                                         0
Time                                                2
Urban_or_Rural_Area                                20
Weather_Conditions                               8555
Year_y                                              0
InScotland                                          3
dtype: int64
```

[101]:

```
df.dtypes
```

[101]:

```
Accident_Index                                  object
Age_Band_of_Driver                              object
Age_of_Vehicle                                 float64
Driver_Home_Area_Type                           object
Driver_IMD_Decile                              float64
Engine_Capacity_.CC.                           float64
Hit_Object_in_Carriageway                       object
Hit_Object_off_Carriageway                      object
Journey_Purpose_of_Driver                       object
Junction_Location                               object
make                                            object
model                                           object
Propulsion_Code                                 object
Sex_of_Driver                                   object
Skidding_and_Overturning                        object
Towing_and_Articulation                         object
Vehicle_Leaving_Carriageway                     object
Vehicle_Location.Restricted_Lane               float64
Vehicle_Manoeuvre                               object
Vehicle_Reference                                int64
Vehicle_Type                                    object
Was_Vehicle_Left_Hand_Drive                     object
X1st_Point_of_Impact                            object
Year_x                                           int64
1st_Road_Class                                  object
1st_Road_Number                                float64
2nd_Road_Class                                  object
2nd_Road_Number                                float64
Accident_Severity                               object
Carriageway_Hazards                             object
Date                                            object
Day_of_Week                                     object
Did_Police_Officer_Attend_Scene_of_Accident    float64
Junction_Control                                object
Junction_Detail                                 object
Latitude                                       float64
Light_Conditions                                object
```

```
Local_Authority_(District)                      object
Local_Authority_(Highway)                       object
Location_Easting_OSGR                          float64
Location_Northing_OSGR                         float64
Longitude                                      float64
LSOA_of_Accident_Location                       object
Number_of_Casualties                             int64
Number_of_Vehicles                               int64
Pedestrian_Crossing-Human_Control              float64
Pedestrian_Crossing-Physical_Facilities        float64
Police_Force                                    object
Road_Surface_Conditions                         object
Road_Type                                       object
Special_Conditions_at_Site                      object
Speed_limit                                      int64
Time                                            object
Urban_or_Rural_Area                             object
Weather_Conditions                              object
Year_y                                           int64
InScotland                                      object
dtype: object
```

[102]:

```python
#descriptive statistics
df.describe().T.style
```

[102]:

```
<pandas.io.formats.style.Styler at 0x790133984460>
```

[103]:

```python
#showing number of unique values
df.nunique()
```

[103]:

```
Accident_Index                                 403235
Age_Band_of_Driver                                 12
Age_of_Vehicle                                     72
Driver_Home_Area_Type                               4
Driver_IMD_Decile                                  10
Engine_Capacity_.CC.                             1754
Hit_Object_in_Carriageway                          12
Hit_Object_off_Carriageway                         11
Journey_Purpose_of_Driver                           6
Junction_Location                                  10
make                                              408
model                                           19749
Propulsion_Code                                     9
Sex_of_Driver                                       4
Skidding_and_Overturning                            6
Towing_and_Articulation                             7
Vehicle_Leaving_Carriageway                        10
```

```
Vehicle_Location.Restricted_Lane                   10
Vehicle_Manoeuvre                                  19
Vehicle_Reference                                  28
Vehicle_Type                                       17
Was_Vehicle_Left_Hand_Drive                         3
X1st_Point_of_Impact                                6
Year_x                                              6
1st_Road_Class                                      5
1st_Road_Number                                  4323
2nd_Road_Class                                      6
2nd_Road_Number                                  4674
Accident_Severity                                   3
Carriageway_Hazards                                 5
Date                                             2191
Day_of_Week                                         7
Did_Police_Officer_Attend_Scene_of_Accident         3
Junction_Control                                    7
Junction_Detail                                    10
Latitude                                       307189
Light_Conditions                                    5
Local_Authority_(District)                        298
Local_Authority_(Highway)                         153
Location_Easting_OSGR                           43751
Location_Northing_OSGR                          50481
Longitude                                      319250
LSOA_of_Accident_Location                       22839
Number_of_Casualties                               33
Number_of_Vehicles                                 20
Pedestrian_Crossing-Human_Control                   3
Pedestrian_Crossing-Physical_Facilities             6
Police_Force                                       35
Road_Surface_Conditions                             5
Road_Type                                           5
Special_Conditions_at_Site                          9
Speed_limit                                         8
Time                                             1439
Urban_or_Rural_Area                                 3
Weather_Conditions                                  8
Year_y                                              6
InScotland                                          2
dtype: int64
```

[104]:

```
#showing number of rows and columns
df.shape
```

[104]:

```
(528335, 57)
```

[105]:

```
#list of columns in the dataset
df.columns
```

[105]:

```
Index(['Accident_Index', 'Age_Band_of_Driver', 'Age_of_Vehicle',
       'Driver_Home_Area_Type', 'Driver_IMD_Decile', 'Engine_Capacity_.CC.',
       'Hit_Object_in_Carriageway', 'Hit_Object_off_Carriageway',
       'Journey_Purpose_of_Driver', 'Junction_Location', 'make', 'model',
       'Propulsion_Code', 'Sex_of_Driver', 'Skidding_and_Overturning',
       'Towing_and_Articulation', 'Vehicle_Leaving_Carriageway',
       'Vehicle_Location.Restricted_Lane', 'Vehicle_Manoeuvre',
       'Vehicle_Reference', 'Vehicle_Type', 'Was_Vehicle_Left_Hand_Drive',
       'X1st_Point_of_Impact', 'Year_x', '1st_Road_Class', '1st_Road_Number',
       '2nd_Road_Class', '2nd_Road_Number', 'Accident_Severity',
       'Carriageway_Hazards', 'Date', 'Day_of_Week',
       'Did_Police_Officer_Attend_Scene_of_Accident', 'Junction_Control',
       'Junction_Detail', 'Latitude', 'Light_Conditions',
       'Local_Authority_(District)', 'Local_Authority_(Highway)',
       'Location_Easting_OSGR', 'Location_Northing_OSGR', 'Longitude',
       'LSOA_of_Accident_Location', 'Number_of_Casualties',
       'Number_of_Vehicles', 'Pedestrian_Crossing-Human_Control',
       'Pedestrian_Crossing-Physical_Facilities', 'Police_Force',
       'Road_Surface_Conditions', 'Road_Type', 'Special_Conditions_at_Site',
       'Speed_limit', 'Time', 'Urban_or_Rural_Area', 'Weather_Conditions',
       'Year_y', 'InScotland'],
      dtype='object')
```

[106]:

```
#printing categorical and numerical columns separately
cats = []
nums = []

for i in df.columns:
  if df[i].dtype == 'object':
    cats.append(i)
  else:
    nums.append(i)

print('Categorical =', cats)
print('Numerical =',nums)
```

```
Categorical = ['Accident_Index', 'Age_Band_of_Driver', 'Driver_Home_Area_Type',
'Hit_Object_in_Carriageway', 'Hit_Object_off_Carriageway',
'Journey_Purpose_of_Driver', 'Junction_Location', 'make', 'model',
'Propulsion_Code', 'Sex_of_Driver', 'Skidding_and_Overturning',
'Towing_and_Articulation', 'Vehicle_Leaving_Carriageway', 'Vehicle_Manoeuvre',
'Vehicle_Type', 'Was_Vehicle_Left_Hand_Drive', 'X1st_Point_of_Impact',
'1st_Road_Class', '2nd_Road_Class', 'Accident_Severity', 'Carriageway_Hazards',
'Date', 'Day_of_Week', 'Junction_Control', 'Junction_Detail',
'Light_Conditions', 'Local_Authority_(District)', 'Local_Authority_(Highway)',
'LSOA_of_Accident_Location', 'Police_Force', 'Road_Surface_Conditions',
'Road_Type', 'Special_Conditions_at_Site', 'Time', 'Urban_or_Rural_Area',
'Weather_Conditions', 'InScotland']
Numerical = ['Age_of_Vehicle', 'Driver_IMD_Decile', 'Engine_Capacity_.CC.',
'Vehicle_Location.Restricted_Lane', 'Vehicle_Reference', 'Year_x',
'1st_Road_Number', '2nd_Road_Number',
'Did_Police_Officer_Attend_Scene_of_Accident', 'Latitude',
'Location_Easting_OSGR', 'Location_Northing_OSGR', 'Longitude',
'Number_of_Casualties', 'Number_of_Vehicles', 'Pedestrian_Crossing-
Human_Control', 'Pedestrian_Crossing-Physical_Facilities', 'Speed_limit',
'Year_y']
```

[107]:

```
#descriptive statistics of numerical columns
df[nums].describe().T
```

[107]:

```
                                                count           mean  \
Age_of_Vehicle                               458857.0       6.383588
Driver_IMD_Decile                            412661.0       5.260124
Engine_Capacity_.CC.                         483119.0    2151.738251
Vehicle_Location.Restricted_Lane             528324.0       0.097436
Vehicle_Reference                            528335.0       1.550823
Year_x                                       528335.0    2007.789590
1st_Road_Number                              528334.0     995.431594
2nd_Road_Number                              523845.0     414.535553
Did_Police_Officer_Attend_Scene_of_Accident  528269.0       1.172516
Latitude                                     528312.0      52.431487
Location_Easting_OSGR                        528312.0  443322.491407
Location_Northing_OSGR                       528312.0  282633.819202
Longitude                                    528312.0      -1.380425
Number_of_Casualties                         528335.0       1.453441
Number_of_Vehicles                           528335.0       2.117925
Pedestrian_Crossing-Human_Control            528319.0       0.006121
Pedestrian_Crossing-Physical_Facilities      528327.0       0.704429
Speed_limit                                  528335.0      39.308441
Year_y                                       528335.0    2007.789590

                                                     std          min  \
Age_of_Vehicle                                  4.468637     1.000000
Driver_IMD_Decile                               2.833934     1.000000
Engine_Capacity_.CC.                         2123.559860     1.000000
Vehicle_Location.Restricted_Lane                0.826387     0.000000
Vehicle_Reference                               0.750148     1.000000
Year_x                                          1.716688  2005.000000
1st_Road_Number                              1818.536437     0.000000
2nd_Road_Number                              1367.399055     0.000000
Did_Police_Officer_Attend_Scene_of_Accident     0.386767     1.000000
```

```
Latitude                                          1.252996     49.914488
Location_Easting_OSGR                         95953.389724  64950.000000
Location_Northing_OSGR                       139151.337560  10520.000000
Longitude                                         1.408710     -7.516225
Number_of_Casualties                              0.955524      1.000000
Number_of_Vehicles                                0.910613      1.000000
Pedestrian_Crossing-Human_Control                 0.101606      0.000000
Pedestrian_Crossing-Physical_Facilities           1.758357      0.000000
Speed_limit                                      14.290866     10.000000
Year_y                                            1.716688   2005.000000

                                                       25%            50%  \
Age_of_Vehicle                                    3.000000       6.000000
Driver_IMD_Decile                                 3.000000       5.000000
Engine_Capacity_.CC.                           1299.000000    1686.000000
Vehicle_Location.Restricted_Lane                  0.000000       0.000000
Vehicle_Reference                                 1.000000       1.000000
Year_x                                         2006.000000    2008.000000
1st_Road_Number                                   0.000000     124.000000
2nd_Road_Number                                   0.000000       0.000000
Did_Police_Officer_Attend_Scene_of_Accident       1.000000       1.000000
Latitude                                         51.513084      52.071845
Location_Easting_OSGR                        384300.000000  454410.000000
Location_Northing_OSGR                       180930.000000  244010.000000
Longitude                                        -2.236486      -1.194378
Number_of_Casualties                              1.000000       1.000000
Number_of_Vehicles                                2.000000       2.000000
Pedestrian_Crossing-Human_Control                 0.000000       0.000000
Pedestrian_Crossing-Physical_Facilities           0.000000       0.000000
Speed_limit                                      30.000000      30.000000
Year_y                                         2006.000000    2008.000000

                                                       75%           max
Age_of_Vehicle                                    9.000000  8.500000e+01
Driver_IMD_Decile                                 8.000000  1.000000e+01
Engine_Capacity_.CC.                           1998.000000  9.600000e+04
Vehicle_Location.Restricted_Lane                  0.000000  9.000000e+00
Vehicle_Reference                                 2.000000  3.200000e+01
Year_x                                         2009.000000  2.010000e+03
1st_Road_Number                                 666.000000  9.999000e+03
2nd_Road_Number                                   0.000000  9.999000e+03
Did_Police_Officer_Attend_Scene_of_Accident       1.000000  3.000000e+00
Latitude                                         53.397356  6.075754e+01
Location_Easting_OSGR                        524410.000000  6.552900e+05
Location_Northing_OSGR                       389232.500000  1.208800e+06
Longitude                                        -0.203736  1.758337e+00
Number_of_Casualties                              2.000000  6.800000e+01
```

```
Number_of_Vehicles                                2.000000  3.200000e+01
Pedestrian_Crossing-Human_Control                 0.000000  2.000000e+00
Pedestrian_Crossing-Physical_Facilities           0.000000  8.000000e+00
Speed_limit                                      50.000000  7.000000e+01
Year_y                                         2009.000000  2.010000e+03
```

[108]:

```
#descriptive statistics of categorical columns
df[cats].describe()
```

[108]:

```
       Accident_Index Age_Band_of_Driver Driver_Home_Area_Type  \
count          528335             528335                528335
unique         403235                 12                     4
top     2009559D05333            36 - 45            Urban area
freq               22             115777                361204

       Hit_Object_in_Carriageway Hit_Object_off_Carriageway  \
count                      21568                      45370
unique                        12                         11
top                         Kerb     Other permanent object
freq                        9291                      15862

        Journey_Purpose_of_Driver                       Junction_Location  \
count                      528335                                  528335
unique                          6                                      10
top     Other/Not known (2005-10)  Not at or within 20 metres of junction
freq                       348438                                  198077

            make               model Propulsion_Code  ...  \
count     528335              452717          491092  ...
unique       408               19749               9  ...
top     VAUXHALL  CLIO DYNAMIQUE 16V          Petrol  ...
freq       58829                2626          299655  ...

       Local_Authority_(Highway) LSOA_of_Accident_Location  \
count                     528335                    504029
unique                       153                     22839
top                       Surrey                 E01000004
freq                       22305                      1285

               Police_Force Road_Surface_Conditions           Road_Type  \
count                528335                  527607              525341
unique                   35                       5                   5
top     Metropolitan Police                     Dry  Single carriageway
freq                 110529                  360938              380113

       Special_Conditions_at_Site    Time Urban_or_Rural_Area  \
count                       14230  528333              528315
```

```
unique                          9    1439                   3
top                     Roadworks   17:00               Urban
freq                         7851    4638              349808

        Weather_Conditions InScotland
count               519780     528332
unique                   8          2
top     Fine no high winds         No
freq                421099     507008

[4 rows x 38 columns]
```

[109]:

```python
#calculating the percentage of missing values
percentage_na = df.isna().sum().sort_values(ascending=False)*100/len(df)
percentage_na
```

[109]:

```
Carriageway_Hazards                            98.345368
Special_Conditions_at_Site                     97.306633
Hit_Object_in_Carriageway                      95.917742
Hit_Object_off_Carriageway                     91.412645
Skidding_and_Overturning                       86.484144
2nd_Road_Class                                 38.700635
1st_Road_Class                                 24.620553
Driver_IMD_Decile                              21.894063
model                                          14.312510
Age_of_Vehicle                                 13.150369
Engine_Capacity_.CC.                            8.558206
Propulsion_Code                                 7.049126
LSOA_of_Accident_Location                       4.600490
Weather_Conditions                              1.619238
2nd_Road_Number                                 0.849840
Road_Type                                       0.566686
Light_Conditions                                0.236024
Road_Surface_Conditions                         0.137791
Did_Police_Officer_Attend_Scene_of_Accident     0.012492
Latitude                                        0.004353
Location_Easting_OSGR                           0.004353
Location_Northing_OSGR                          0.004353
Longitude                                       0.004353
Urban_or_Rural_Area                             0.003785
Pedestrian_Crossing-Human_Control               0.003028
Vehicle_Location.Restricted_Lane                0.002082
Pedestrian_Crossing-Physical_Facilities         0.001514
InScotland                                      0.000568
Time                                            0.000379
1st_Road_Number                                 0.000189
Year_y                                          0.000000
```

```
Police_Force                                    0.000000
Speed_limit                                     0.000000
Number_of_Casualties                            0.000000
Local_Authority_(Highway)                       0.000000
Number_of_Vehicles                              0.000000
Local_Authority_(District)                      0.000000
Accident_Index                                  0.000000
Junction_Detail                                 0.000000
Vehicle_Manoeuvre                               0.000000
Driver_Home_Area_Type                           0.000000
Journey_Purpose_of_Driver                       0.000000
Junction_Location                               0.000000
make                                            0.000000
Sex_of_Driver                                   0.000000
Towing_and_Articulation                         0.000000
Vehicle_Leaving_Carriageway                     0.000000
Vehicle_Reference                               0.000000
Junction_Control                                0.000000
Vehicle_Type                                    0.000000
Was_Vehicle_Left_Hand_Drive                     0.000000
X1st_Point_of_Impact                            0.000000
Year_x                                          0.000000
Age_Band_of_Driver                              0.000000
Date                                            0.000000
Day_of_Week                                     0.000000
Accident_Severity                               0.000000
dtype: float64
```

[110]:

```
#checking duplicate entries
print("Duplicate entries:", df.duplicated().sum())
```

```
Duplicate entries: 0
```

[111]:

```
#column names with their unique values
for column in df.columns:
    print(column + ":" + str(len(df[column].unique())))
    if(len(df[column].unique())<100):
        print(column + ':' + str(df[column].unique()))
    print('\n')
```

```
Accident_Index:403235



Age_Band_of_Driver:12
Age_Band_of_Driver:['36 - 45' '26 - 35' '46 - 55' '66 - 75' '16 - 20' '56 - 65'
 'Data missing or out of range' '21 - 25' 'Over 75' '11 - 15' '6 - 10'
 '0 - 5']
```

```
Age_of_Vehicle:73
Age_of_Vehicle:[ 3.  5.  4. 10.  1.  2. nan 16.  6.  7. 14.  9. 18. 39.  8. 15.
12. 13.
 19. 11. 27. 20. 21. 23. 17. 42. 40. 22. 24. 35. 25. 31. 37. 41. 26. 34.
 28. 38. 29. 48. 30. 36. 45. 33. 43. 47. 32. 52. 46. 44. 51. 49. 50. 76.
 85. 80. 61. 68. 59. 63. 53. 58. 55. 57. 65. 54. 66. 60. 56. 72. 71. 83.
 75.]


Driver_Home_Area_Type:4
Driver_Home_Area_Type:['Data missing or out of range' 'Urban area' 'Rural'
'Small town']


Driver_IMD_Decile:11
Driver_IMD_Decile:[nan  3.  1.  4.  2.  6.  7.  8. 10.  5.  9.]


Engine_Capacity_.CC.:1755


Hit_Object_in_Carriageway:13
Hit_Object_in_Carriageway:[nan 'Parked vehicle' 'Kerb' 'Bollard or refuge' 'Open
door of vehicle'
 'Other object' 'Previous accident' 'Road works'
 'Central island of roundabout' 'Bridge (side)' 'Bridge (roof)'
 'Data missing or out of range' 'Any animal (except ridden horse)']


Hit_Object_off_Carriageway:12
Hit_Object_off_Carriageway:[nan 'Road sign or traffic signal' 'Other permanent
object' 'Lamp post'
 'Near/Offside crash barrier' 'Central crash barrier' 'Tree'
 'Bus stop or bus shelter' 'Entered ditch' 'Telegraph or electricity pole'
 'Data missing or out of range' 'Submerged in water']


Journey_Purpose_of_Driver:6
Journey_Purpose_of_Driver:['Journey as part of work' 'Other/Not known (2005-10)'
 'Commuting to/from work' 'Pupil riding to/from school'
 'Taking pupil to/from school' 'Data missing or out of range']


Junction_Location:10
Junction_Location:['Leaving roundabout' 'Not at or within 20 metres of junction'
 'Approaching junction or waiting/parked at junction approach'
 'Mid Junction - on roundabout or on main road'
 'Cleared junction or waiting/parked at junction exit'
 'Entering main road' 'Entering from slip road' 'Leaving main road'
 'Entering roundabout' 'Data missing or out of range']


make:408


model:19750


Propulsion_Code:10
Propulsion_Code:['Heavy oil' 'Petrol' nan 'Hybrid electric' 'Gas/Bi-fuel' 'Gas'
 'Petrol/Gas (LPG)' 'Electric' 'New fuel technology' 'Gas Diesel']


Sex_of_Driver:4
Sex_of_Driver:['Male' 'Female' 'Not known' 'Data missing or out of range']


Skidding_and_Overturning:7
Skidding_and_Overturning:[nan 'Skidded' 'Skidded and overturned' 'Overturned'
'Jackknifed'
 'Jackknifed and overturned' 'Data missing or out of range']


Towing_and_Articulation:7
Towing_and_Articulation:['No tow/articulation' 'Articulated vehicle' 'Other tow'
'Single trailer'
 'Double or multiple trailer' 'Data missing or out of range' 'Caravan']


Vehicle_Leaving_Carriageway:10
Vehicle_Leaving_Carriageway:['Did not leave carriageway' 'Nearside' 'Offside'
 'Straight ahead at junction' 'Offside on to central reservation'
 'Offside - crossed central reservation' 'Nearside and rebounded'
 'Offside on to centrl res + rebounded' 'Offside and rebounded'
 'Data missing or out of range']


Vehicle_Location.Restricted_Lane:11
Vehicle_Location.Restricted_Lane:[ 0.  2.  8.  3.  4.  7.  9.  6.  1.  5. nan]


Vehicle_Manoeuvre:19
Vehicle_Manoeuvre:['Slowing or stopping' 'Going ahead right-hand bend' 'Going
ahead other'
 'Moving off' 'Parked' 'Turning left' 'Turning right'
 'Overtaking moving vehicle - offside' 'U-turn' 'Waiting to go - held up'
 'Overtaking static vehicle - offside' 'Waiting to turn right'
 'Going ahead left-hand bend' 'Changing lane to right'
 'Changing lane to left' 'Reversing' 'Waiting to turn left'
 'Overtaking - nearside' 'Data missing or out of range']


Vehicle_Reference:28
Vehicle_Reference:[ 1  2  3  4  5  6  8  7 13 11  9 10 16 18 12 14 15 17 19 21
23 24 25 26
 29 30 31 32]


Vehicle_Type:17
Vehicle_Type:['Bus or coach (17 or more pass seats)' 'Car' 'Motorcycle 125cc and
under'
 'Motorcycle over 500cc' 'Motorcycle over 125cc and up to 500cc'
 'Taxi/Private hire car' 'Goods 7.5 tonnes mgw and over'
 'Van / Goods 3.5 tonnes mgw or under' 'Motorcycle 50cc and under'
 'Goods over 3.5t. and under 7.5t' 'Minibus (8 - 16 passenger seats)'
 'Other vehicle' 'Agricultural vehicle' 'Pedal cycle' 'Tram'
 'Ridden horse' 'Data missing or out of range']


Was_Vehicle_Left_Hand_Drive:3
Was_Vehicle_Left_Hand_Drive:['No' 'Yes' 'Data missing or out of range']


X1st_Point_of_Impact:6
X1st_Point_of_Impact:['Nearside' 'Front' 'Did not impact' 'Back' 'Offside'
 'Data missing or out of range']


Year_x:6
Year_x:[2005 2006 2007 2008 2009 2010]


1st_Road_Class:6
1st_Road_Class:['B' 'C' 'A' nan 'Motorway' 'A(M)']


1st_Road_Number:4324


2nd_Road_Class:7
2nd_Road_Class:['C' nan 'Unclassified' 'B' 'A' 'Motorway' 'A(M)']
```

```
2nd_Road_Number:4675


Accident_Severity:3
Accident_Severity:['Slight' 'Serious' 'Fatal']


Carriageway_Hazards:6
Carriageway_Hazards:[nan 'Other object on road' 'Pedestrian in carriageway - not
injured'
 'Vehicle load on road' 'Previous accident'
 'Any animal in carriageway (except ridden horse)']


Date:2191


Day_of_Week:7
Day_of_Week:['Wednesday' 'Thursday' 'Friday' 'Monday' 'Tuesday' 'Sunday'
'Saturday']


Did_Police_Officer_Attend_Scene_of_Accident:4
Did_Police_Officer_Attend_Scene_of_Accident:[ 1.  2.  3. nan]


Junction_Control:7
Junction_Control:['Auto traffic signal' 'Data missing or out of range'
 'Give way or uncontrolled' 'Stop sign' 'Authorised person'
 'Auto traffic sigl' 'Not at junction or within 20 metres']


Junction_Detail:10
Junction_Detail:['Crossroads' 'Not at junction or within 20 metres'
 'T or staggered junction' 'Mini-roundabout' 'Private drive or entrance'
 'More than 4 arms (not roundabout)' 'Roundabout' 'Other junction'
 'Slip road' 'Data missing or out of range']


Latitude:307190


Light_Conditions:6
Light_Conditions:['Darkness - lights lit' nan 'Darkness - lighting unknown'
 'Darkness - lights unlit' 'Daylight' 'Darkness - no lighting']


Local_Authority_(District):298
Local_Authority_(Highway):153


Location_Easting_OSGR:43752


Location_Northing_OSGR:50482


Longitude:319251


LSOA_of_Accident_Location:22840


Number_of_Casualties:33
Number_of_Casualties:[ 1  2  4  3  5  6  7 10  9  8 17 13 19 11 15 12 14 23 18
21 22 26 16 40
 25 28 29 42 68 27 32 43 24]


Number_of_Vehicles:20
Number_of_Vehicles:[ 1  2  4  3  5  6  8  7 13 10 11 18  9 20 17 14 12 16 32 19]


Pedestrian_Crossing-Human_Control:4
Pedestrian_Crossing-Human_Control:[ 0.  2.  1. nan]


Pedestrian_Crossing-Physical_Facilities:7
Pedestrian_Crossing-Physical_Facilities:[ 5.  0.  1.  4.  7.  8. nan]


Police_Force:35
Police_Force:['Metropolitan Police' 'City of London' 'Lancashire' 'Merseyside'
 'Greater Manchester' 'Durham' 'South Yorkshire' 'Humberside' 'Cleveland'
 'West Midlands' 'Staffordshire' 'West Mercia' 'Derbyshire'
 'Nottinghamshire' 'Lincolnshire' 'Northamptonshire' 'Norfolk' 'Suffolk'
 'Bedfordshire' 'Hertfordshire' 'Essex' 'Thames Valley' 'Hampshire'
 'Surrey' 'Devon and Cornwall' 'Wiltshire' 'Dorset' 'North Wales'
 'Dyfed-Powys' 'Northern' 'Strathclyde' 'Avon and Somerset' 'Tayside'
 'North Yorkshire' 'West Yorkshire']


Road_Surface_Conditions:6
Road_Surface_Conditions:['Dry' 'Wet or damp' 'Snow' 'Frost or ice' 'Flood over
3cm. deep' nan]


Road_Type:6
Road_Type:['Dual carriageway' 'Single carriageway' 'One way street' 'Roundabout'
 'Slip road' nan]


Special_Conditions_at_Site:10
Special_Conditions_at_Site:[nan 'Oil or diesel' 'Roadworks' 'Auto signal part
defective'
 'Road surface defective' 'Auto traffic signal - out'
 'Road sign or marking defective or obscured' 'Mud'
 'Auto traffic sigl - out' 'Auto sigl part defective']


Speed_limit:8
Speed_limit:[30 40 50 20 60 70 10 15]


Time:1440


Urban_or_Rural_Area:4
Urban_or_Rural_Area:['Urban' 'Rural' 'Unallocated' nan]


Weather_Conditions:9
Weather_Conditions:['Fine no high winds' 'Raining no high winds' 'Snowing no
high winds'
 'Fine + high winds' 'Other' nan 'Raining + high winds' 'Fog or mist'
 'Snowing + high winds']


Year_y:6
Year_y:[2005 2006 2007 2008 2009 2010]


InScotland:3
InScotland:['No' 'Yes' nan]
```

4 Cleaning the Dataset

[112]:

```
#changing the value missing to nan
df['Age_Band_of_Driver'].replace('Data missing or out of range', np.NaN,␣
 ↪inplace = True)
df['Driver_Home_Area_Type'].replace('Data missing or out of range', np.NaN,␣
 ↪inplace = True)
df['Journey_Purpose_of_Driver'].replace('Data missing or out of range', np.NaN,␣
 ↪inplace = True)
df['Junction_Location'].replace('Data missing or out of range', np.NaN, inplace␣
 ↪= True)
df['Sex_of_Driver'].replace('Data missing or out of range', np.NaN, inplace =␣
 ↪True)
df['Towing_and_Articulation'].replace('Data missing or out of range', np.NaN,␣
 ↪inplace = True)
df['Vehicle_Leaving_Carriageway'].replace('Data missing or out of range', np.
 ↪NaN, inplace = True)
df['Vehicle_Manoeuvre'].replace('Data missing or out of range', np.NaN, inplace␣
 ↪= True)
df['Vehicle_Type'].replace('Data missing or out of range', np.NaN, inplace =␣
 ↪True)
df['Was_Vehicle_Left_Hand_Drive'].replace('Data missing or out of range', np.
 ↪NaN, inplace = True)
df['X1st_Point_of_Impact'].replace('Data missing or out of range', np.NaN,␣
 ↪inplace = True)
df['Junction_Detail'].replace('Data missing or out of range', np.NaN, inplace =␣
 ↪True)
```

[113]:

```
#changing the value of zero to nan
df['1st_Road_Number'].replace(0, np.NaN, inplace = True)
df['2nd_Road_Number'].replace(0, np.NaN, inplace = True)
```

[114]:

```
#calculating the percentage of missing values
percentage_na = df.isna().sum().sort_values(ascending=False)*100/len(df)
percentage_na
```

[114]:

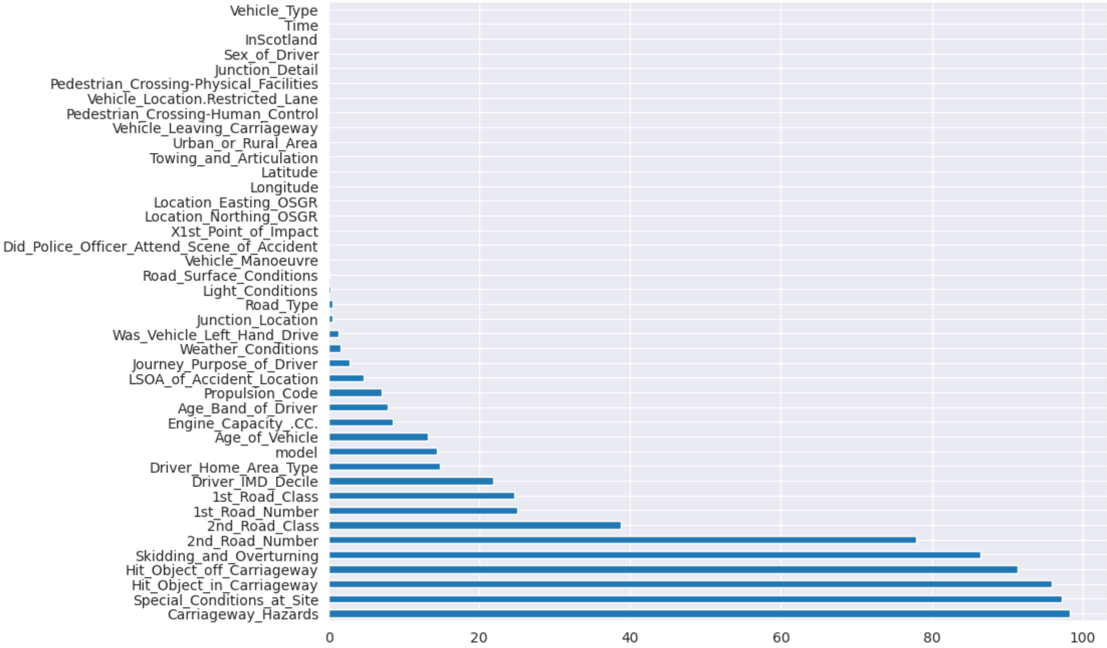
```
Carriageway_Hazards                               98.345368
Special_Conditions_at_Site                        97.306633
Hit_Object_in_Carriageway                         95.917742
Hit_Object_off_Carriageway                        91.412645
Skidding_and_Overturning                          86.484144
2nd_Road_Number                                   77.981205
2nd_Road_Class                                    38.700635
1st_Road_Number                                   25.096956
1st_Road_Class                                    24.620553
Driver_IMD_Decile                                 21.894063
```

```
Driver_Home_Area_Type                          14.814275
model                                          14.312510
Age_of_Vehicle                                 13.150369
Engine_Capacity_.CC.                            8.558206
Age_Band_of_Driver                              7.828366
Propulsion_Code                                 7.049126
LSOA_of_Accident_Location                       4.600490
Journey_Purpose_of_Driver                       2.761316
Weather_Conditions                              1.619238
Was_Vehicle_Left_Hand_Drive                     1.316589
Junction_Location                               0.575014
Road_Type                                       0.566686
Light_Conditions                                0.236024
Road_Surface_Conditions                         0.137791
Vehicle_Manoeuvre                               0.029148
Did_Police_Officer_Attend_Scene_of_Accident     0.012492
X1st_Point_of_Impact                            0.009653
Location_Northing_OSGR                          0.004353
Location_Easting_OSGR                           0.004353
Longitude                                       0.004353
Latitude                                        0.004353
Towing_and_Articulation                         0.003975
Urban_or_Rural_Area                             0.003785
Vehicle_Leaving_Carriageway                     0.003218
Pedestrian_Crossing-Human_Control               0.003028
Vehicle_Location.Restricted_Lane                0.002082
Pedestrian_Crossing-Physical_Facilities         0.001514
Junction_Detail                                 0.001136
Sex_of_Driver                                   0.001136
InScotland                                      0.000568
Time                                            0.000379
Vehicle_Type                                    0.000379
Speed_limit                                     0.000000
Year_y                                          0.000000
Police_Force                                    0.000000
Accident_Index                                  0.000000
Number_of_Vehicles                              0.000000
Number_of_Casualties                            0.000000
Local_Authority_(Highway)                       0.000000
Local_Authority_(District)                      0.000000
Junction_Control                                0.000000
Day_of_Week                                     0.000000
Date                                            0.000000
Year_x                                          0.000000
Vehicle_Reference                               0.000000
make                                            0.000000
Accident_Severity                               0.000000
```

```
dtype: float64
```

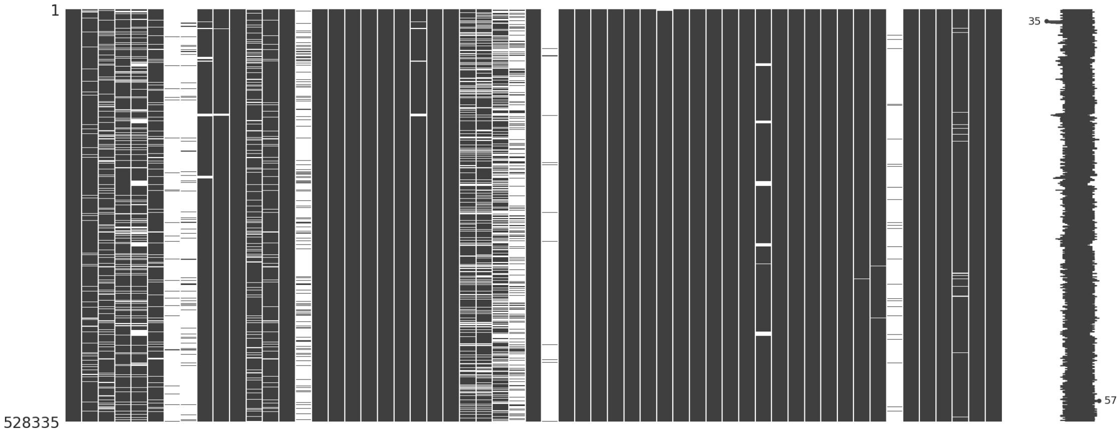
```
[115]: #Visualize percentage_na
import seaborn as sns
plt.figure(figsize=(10, 8))
sns.set_style("darkgrid")
percentage_na[percentage_na>0].plot(kind='barh')
```

```
[115]: <Axes: >
```



```
[116]: #visualizing missing values
import missingno as msno
msno.matrix(df)
```

```
[116]: <Axes: >
```



[117]:
```
#dropping columns with highest missing values
df.
 ↪drop(['Carriageway_Hazards','Special_Conditions_at_Site','Hit_Object_in_Carriageway','Hit_O
        'Skidding_and_Overturning','2nd_Road_Number'],axis=1,inplace=True)
df.head()
```

[117]:
```
  Accident_Index Age_Band_of_Driver  Age_of_Vehicle Driver_Home_Area_Type  \
0  200501BS00002            36 - 45             3.0                   NaN
1  200501BS00003            26 - 35             5.0            Urban area
2  200501BS00004            46 - 55             4.0            Urban area
3  200501BS00005            46 - 55            10.0                   NaN
4  200501BS00006            46 - 55             1.0            Urban area

   Driver_IMD_Decile  Engine_Capacity_.CC.  Journey_Purpose_of_Driver  \
0                NaN                8268.0    Journey as part of work
1                3.0                8300.0    Journey as part of work
2                1.0                1769.0  Other/Not known (2005-10)
3                NaN                  85.0  Other/Not known (2005-10)
4                4.0                2976.0  Other/Not known (2005-10)

                       Junction_Location    make                    model  \
0                    Leaving roundabout  DENNIS                      NaN
1  Not at or within 20 metres of junction  DENNIS                      NaN
2  Not at or within 20 metres of junction  NISSAN            ALMERA SE AUTO
3  Not at or within 20 metres of junction   HONDA                      NaN
4  Not at or within 20 metres of junction    AUDI  A4 SPORT CABRIOLET AUTO

   … Pedestrian_Crossing-Physical_Facilities         Police_Force  \
0  …                                     5.0  Metropolitan Police
1  …                                     0.0  Metropolitan Police
2  …                                     0.0  Metropolitan Police
3  …                                     0.0  Metropolitan Police
```

```
4  …                                      0.0  Metropolitan Police

  Road_Surface_Conditions           Road_Type  Speed_limit   Time  \
0                      Dry    Dual carriageway           30  17:36
1                      Dry  Single carriageway           30  00:15
2                      Dry  Single carriageway           30  10:35
3              Wet or damp  Single carriageway           30  21:13
4              Wet or damp  Single carriageway           30  12:40

   Urban_or_Rural_Area     Weather_Conditions Year_y InScotland
0                Urban     Fine no high winds   2005         No
1                Urban     Fine no high winds   2005         No
2                Urban     Fine no high winds   2005         No
3                Urban     Fine no high winds   2005         No
4                Urban  Raining no high winds   2005         No

[5 rows x 51 columns]
```

[118]:

```
df.isna().sum()
```

[118]:

```
Accident_Index                                      0
Age_Band_of_Driver                              41360
Age_of_Vehicle                                  69478
Driver_Home_Area_Type                           78269
Driver_IMD_Decile                              115674
Engine_Capacity_.CC.                            45216
Journey_Purpose_of_Driver                       14589
Junction_Location                                3038
make                                                0
model                                           75618
Propulsion_Code                                 37243
Sex_of_Driver                                       6
Towing_and_Articulation                            21
Vehicle_Leaving_Carriageway                        17
Vehicle_Location.Restricted_Lane                   11
Vehicle_Manoeuvre                                 154
Vehicle_Reference                                   0
Vehicle_Type                                        2
Was_Vehicle_Left_Hand_Drive                      6956
X1st_Point_of_Impact                               51
Year_x                                              0
1st_Road_Class                                 130079
1st_Road_Number                                132596
2nd_Road_Class                                 204469
Accident_Severity                                   0
Date                                                0
Day_of_Week                                         0
```

```
Did_Police_Officer_Attend_Scene_of_Accident        66
Junction_Control                                    0
Junction_Detail                                     6
Latitude                                           23
Light_Conditions                                 1247
Local_Authority_(District)                          0
Local_Authority_(Highway)                           0
Location_Easting_OSGR                              23
Location_Northing_OSGR                             23
Longitude                                          23
LSOA_of_Accident_Location                       24306
Number_of_Casualties                                0
Number_of_Vehicles                                  0
Pedestrian_Crossing-Human_Control                  16
Pedestrian_Crossing-Physical_Facilities             8
Police_Force                                        0
Road_Surface_Conditions                           728
Road_Type                                        2994
Speed_limit                                         0
Time                                                2
Urban_or_Rural_Area                                20
Weather_Conditions                               8555
Year_y                                              0
InScotland                                          3
dtype: int64
```

[119]:

```
#filling missing values
df['1st_Road_Class']=df['1st_Road_Class'].fillna(df['1st_Road_Class'].mode()[0])
df['1st_Road_Number']=df['1st_Road_Number'].fillna(df['1st_Road_Number'].mean())
df['2nd_Road_Class']=df['2nd_Road_Class'].fillna(df['2nd_Road_Class'].mode()[0])
df['Did_Police_Officer_Attend_Scene_of_Accident']=df['Did_Police_Officer_Attend_Scene_of_Accid
  ↪fillna(df['Did_Police_Officer_Attend_Scene_of_Accident'].mean())
df['Latitude']=df['Latitude'].fillna(df['Latitude'].mean())
df['Location_Easting_OSGR']=df['Location_Easting_OSGR'].
  ↪fillna(df['Location_Easting_OSGR'].mode()[0])
df['Location_Northing_OSGR']=df['Location_Northing_OSGR'].
  ↪fillna(df['Location_Northing_OSGR'].mode()[0])
df['Longitude']=df['Longitude'].fillna(df['Longitude'].mode()[0])
df['LSOA_of_Accident_Location']=df['LSOA_of_Accident_Location'].
  ↪fillna(df['LSOA_of_Accident_Location'].mode()[0])
df['Pedestrian_Crossing-Human_Control']=df['Pedestrian_Crossing-Human_Control'].
  ↪fillna(df['Pedestrian_Crossing-Human_Control'].mean())
df['Pedestrian_Crossing-Physical_Facilities']=df['Pedestrian_Crossing-Physical_Facilities'].
  ↪fillna(df['Pedestrian_Crossing-Physical_Facilities'].mean())
df['Road_Surface_Conditions']=df['Road_Surface_Conditions'].
  ↪fillna(df['Road_Surface_Conditions'].mode()[0])
```

```
df['Road_Type']=df['Road_Type'].fillna(df['Road_Type'].mode()[0])
df['Urban_or_Rural_Area']=df['Urban_or_Rural_Area'].
  ↪fillna(df['Urban_or_Rural_Area'].mode()[0])
df['Weather_Conditions']=df['Weather_Conditions'].
  ↪fillna(df['Weather_Conditions'].mode()[0])
df['Time']=df['Time'].fillna(df['Time'].mode()[0])
df['Light_Conditions']=df['Light_Conditions'].fillna(df['Light_Conditions'].
  ↪mode()[0])
df['Age_Band_of_Driver']=df['Age_Band_of_Driver'].
  ↪fillna(df['Age_Band_of_Driver'].mode()[0])
df['Age_of_Vehicle']=df['Age_of_Vehicle'].fillna(df['Age_of_Vehicle'].median())
df['Driver_Home_Area_Type']=df['Driver_Home_Area_Type'].
  ↪fillna(df['Driver_Home_Area_Type'].mode()[0])
df['Driver_IMD_Decile']=df['Driver_IMD_Decile'].fillna(df['Driver_IMD_Decile'].
  ↪median())
df['Engine_Capacity_.CC.']=df['Engine_Capacity_.CC.'].
  ↪fillna(df['Engine_Capacity_.CC.'].mode()[0])
df['Journey_Purpose_of_Driver']=df['Journey_Purpose_of_Driver'].
  ↪fillna(df['Journey_Purpose_of_Driver'].mode()[0])
df['Junction_Location']=df['Junction_Location'].fillna(df['Junction_Location'].
  ↪mode()[0])
df['model']=df['model'].fillna(df['model'].mode()[0])
df['Propulsion_Code']=df['Propulsion_Code'].fillna(df['Propulsion_Code'].
  ↪mode()[0])
df['Sex_of_Driver']=df['Sex_of_Driver'].fillna(df['Sex_of_Driver'].mode()[0])
df['Towing_and_Articulation']=df['Towing_and_Articulation'].
  ↪fillna(df['Towing_and_Articulation'].mode()[0])
df['Vehicle_Leaving_Carriageway']=df['Vehicle_Leaving_Carriageway'].
  ↪fillna(df['Vehicle_Leaving_Carriageway'].mode()[0])
df['Vehicle_Location.Restricted_Lane']=df['Vehicle_Location.Restricted_Lane'].
  ↪fillna(df['Vehicle_Location.Restricted_Lane'].mode()[0])
df['Vehicle_Manoeuvre']=df['Vehicle_Manoeuvre'].fillna(df['Vehicle_Manoeuvre'].
  ↪mode()[0])
df['Vehicle_Type']=df['Vehicle_Type'].fillna(df['Vehicle_Type'].mode()[0])
df['Was_Vehicle_Left_Hand_Drive']=df['Was_Vehicle_Left_Hand_Drive'].
  ↪fillna(df['Was_Vehicle_Left_Hand_Drive'].mode()[0])
df['X1st_Point_of_Impact']=df['X1st_Point_of_Impact'].
  ↪fillna(df['X1st_Point_of_Impact'].mode()[0])
df['Junction_Detail']=df['Junction_Detail'].fillna(df['Junction_Detail'].
  ↪mode()[0])
df['InScotland']=df['InScotland'].fillna(df['InScotland'].mode()[0])
df.isna().sum()
```

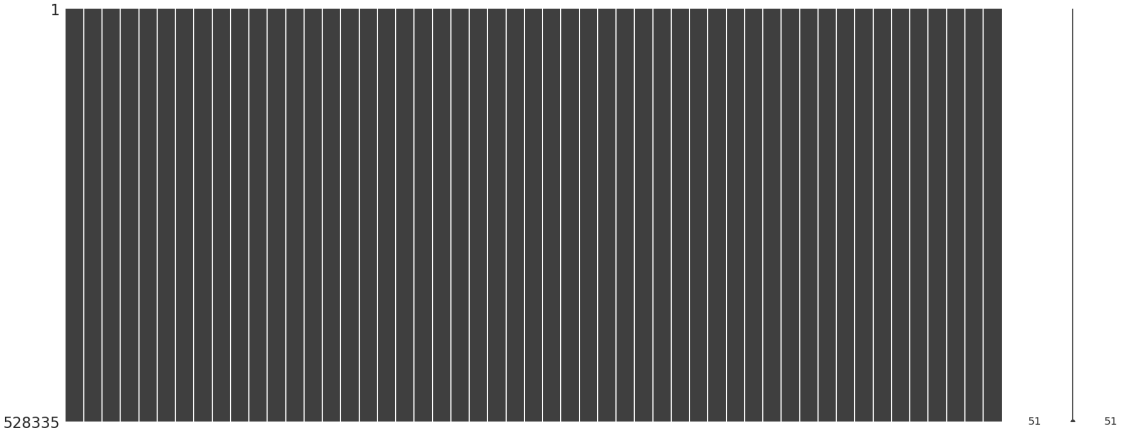
```
[119]: Accident_Index                                0
       Age_Band_of_Driver                            0
       Age_of_Vehicle                                0
```

```
Driver_Home_Area_Type                          0
Driver_IMD_Decile                              0
Engine_Capacity_.CC.                           0
Journey_Purpose_of_Driver                      0
Junction_Location                              0
make                                           0
model                                          0
Propulsion_Code                                0
Sex_of_Driver                                  0
Towing_and_Articulation                        0
Vehicle_Leaving_Carriageway                    0
Vehicle_Location.Restricted_Lane               0
Vehicle_Manoeuvre                              0
Vehicle_Reference                              0
Vehicle_Type                                   0
Was_Vehicle_Left_Hand_Drive                    0
X1st_Point_of_Impact                           0
Year_x                                         0
1st_Road_Class                                 0
1st_Road_Number                                0
2nd_Road_Class                                 0
Accident_Severity                              0
Date                                           0
Day_of_Week                                    0
Did_Police_Officer_Attend_Scene_of_Accident    0
Junction_Control                               0
Junction_Detail                                0
Latitude                                       0
Light_Conditions                               0
Local_Authority_(District)                     0
Local_Authority_(Highway)                      0
Location_Easting_OSGR                          0
Location_Northing_OSGR                         0
Longitude                                      0
LSOA_of_Accident_Location                      0
Number_of_Casualties                           0
Number_of_Vehicles                             0
Pedestrian_Crossing-Human_Control              0
Pedestrian_Crossing-Physical_Facilities        0
Police_Force                                   0
Road_Surface_Conditions                        0
Road_Type                                      0
Speed_limit                                    0
Time                                           0
Urban_or_Rural_Area                            0
Weather_Conditions                             0
Year_y                                         0
```

```
InScotland                                    0
dtype: int64
```

[120]:
```
#graphical representation of missing values
msno.matrix(df)
```

[120]:
```
<Axes: >
```



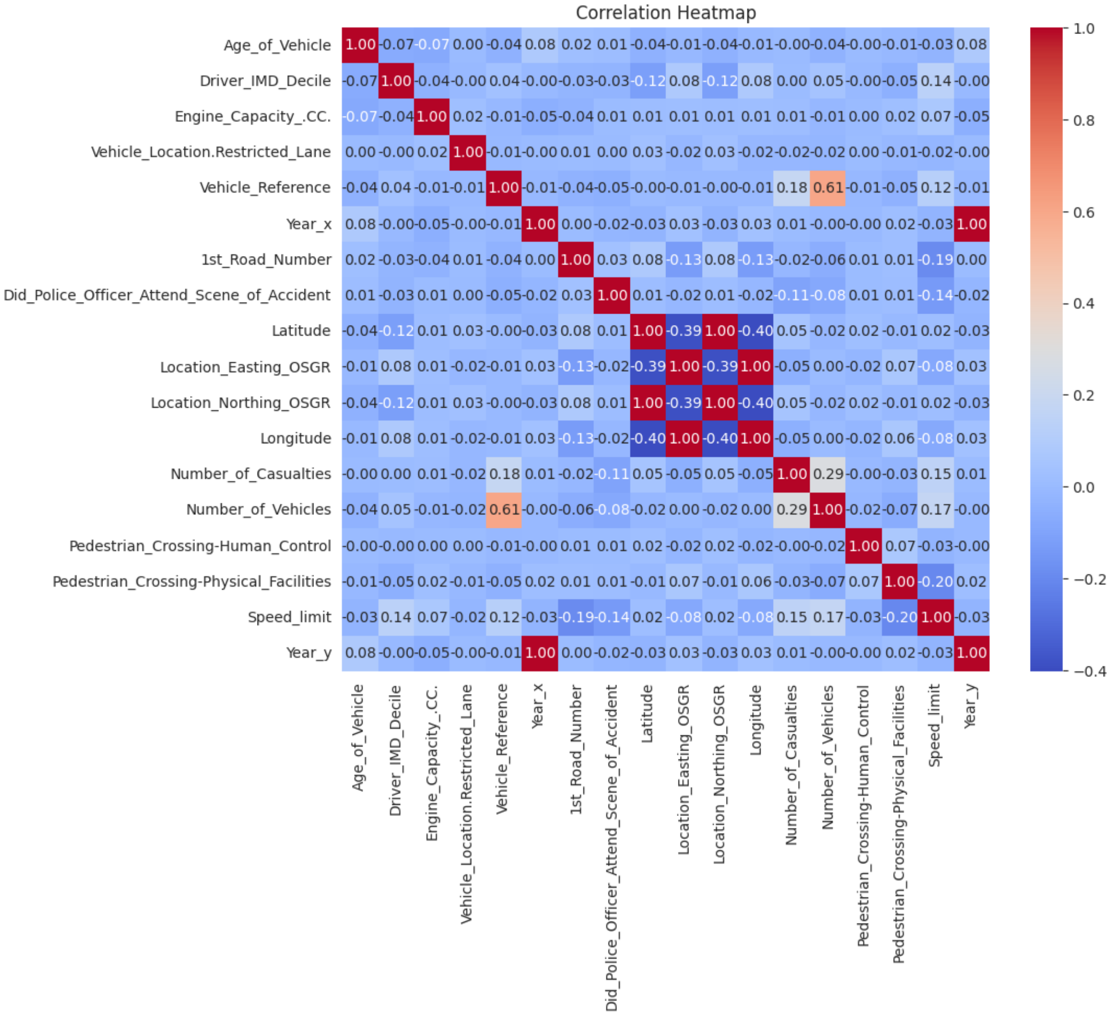
[121]:
```
#dropping unwanted column
df.drop(['Accident_Index'],axis=1,inplace=True)
df.head()
```

[121]:
```
  Age_Band_of_Driver  Age_of_Vehicle Driver_Home_Area_Type  Driver_IMD_Decile  \
0            36 - 45             3.0            Urban area                5.0
1            26 - 35             5.0            Urban area                3.0
2            46 - 55             4.0            Urban area                1.0
3            46 - 55            10.0            Urban area                5.0
4            46 - 55             1.0            Urban area                4.0

   Engine_Capacity_.CC.  Journey_Purpose_of_Driver  \
0                8268.0    Journey as part of work
1                8300.0    Journey as part of work
2                1769.0  Other/Not known (2005-10)
3                  85.0  Other/Not known (2005-10)
4                2976.0  Other/Not known (2005-10)

                        Junction_Location    make                   model  \
0                     Leaving roundabout  DENNIS       CLIO DYNAMIQUE 16V
1  Not at or within 20 metres of junction  DENNIS       CLIO DYNAMIQUE 16V
2  Not at or within 20 metres of junction  NISSAN           ALMERA SE AUTO
3  Not at or within 20 metres of junction   HONDA       CLIO DYNAMIQUE 16V
4  Not at or within 20 metres of junction    AUDI  A4 SPORT CABRIOLET AUTO
```

```
  Propulsion_Code  … Pedestrian_Crossing-Physical_Facilities  \
0       Heavy oil  …                                     5.0
1       Heavy oil  …                                     0.0
2          Petrol  …                                     0.0
3          Petrol  …                                     0.0
4          Petrol  …                                     0.0

          Police_Force Road_Surface_Conditions           Road_Type  \
0  Metropolitan Police                     Dry    Dual carriageway
1  Metropolitan Police                     Dry  Single carriageway
2  Metropolitan Police                     Dry  Single carriageway
3  Metropolitan Police             Wet or damp  Single carriageway
4  Metropolitan Police             Wet or damp  Single carriageway

  Speed_limit   Time Urban_or_Rural_Area     Weather_Conditions Year_y  \
0          30  17:36               Urban     Fine no high winds   2005
1          30  00:15               Urban     Fine no high winds   2005
2          30  10:35               Urban     Fine no high winds   2005
3          30  21:13               Urban     Fine no high winds   2005
4          30  12:40               Urban  Raining no high winds   2005

   InScotland
0          No
1          No
2          No
3          No
4          No

[5 rows x 50 columns]
```

[122]:

```
numerical_columns = df.select_dtypes(include=['int64', 'float64'])

# Compute the correlation matrix
correlation_matrix = numerical_columns.corr()

# Plot heatmap
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Correlation Heatmap')
plt.show()
```



```
[123]: df.drop(['Location_Easting_OSGR','Location_Northing_OSGR'],axis=1,inplace=True)
       df.head()
```

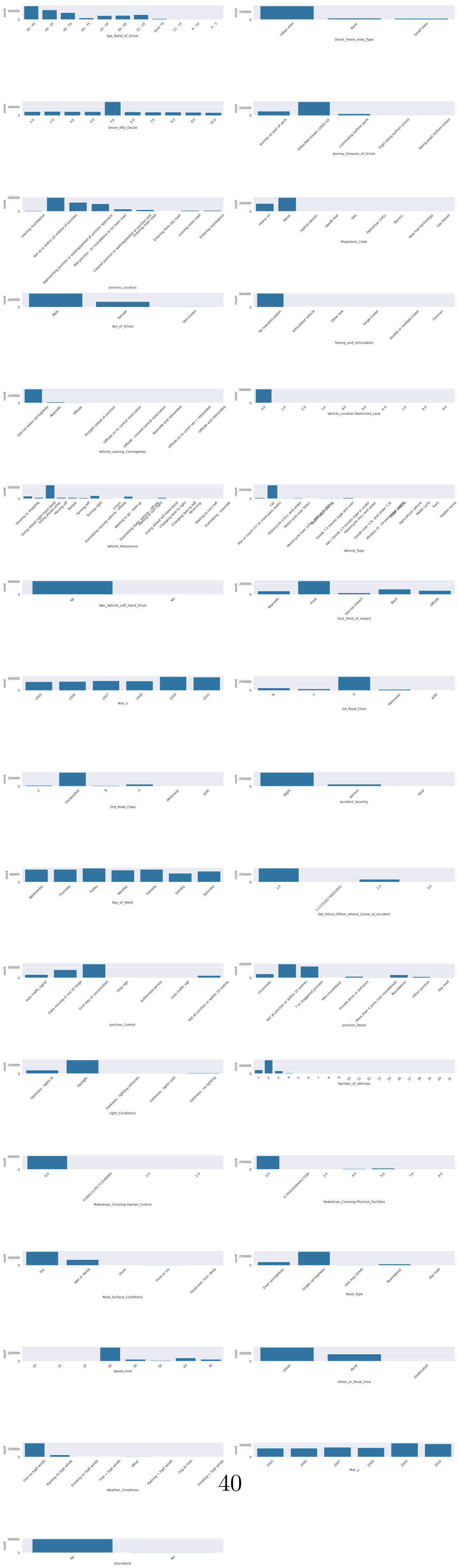
```
[123]:   Age_Band_of_Driver  Age_of_Vehicle Driver_Home_Area_Type  Driver_IMD_Decile  \
       0            36 - 45             3.0            Urban area                5.0
       1            26 - 35             5.0            Urban area                3.0
       2            46 - 55             4.0            Urban area                1.0
       3            46 - 55            10.0            Urban area                5.0
       4            46 - 55             1.0            Urban area                4.0

          Engine_Capacity_.CC.  Journey_Purpose_of_Driver  \
       0                8268.0    Journey as part of work
       1                8300.0    Journey as part of work
       2                1769.0  Other/Not known (2005-10)
       3                  85.0  Other/Not known (2005-10)
       4                2976.0  Other/Not known (2005-10)
```

```
                        Junction_Location    make                   model  \
0                      Leaving roundabout  DENNIS      CLIO DYNAMIQUE 16V
1  Not at or within 20 metres of junction  DENNIS      CLIO DYNAMIQUE 16V
2  Not at or within 20 metres of junction  NISSAN          ALMERA SE AUTO
3  Not at or within 20 metres of junction   HONDA      CLIO DYNAMIQUE 16V
4  Not at or within 20 metres of junction    AUDI  A4 SPORT CABRIOLET AUTO

  Propulsion_Code  … Pedestrian_Crossing-Physical_Facilities  \
0       Heavy oil  …                                     5.0
1       Heavy oil  …                                     0.0
2          Petrol  …                                     0.0
3          Petrol  …                                     0.0
4          Petrol  …                                     0.0

          Police_Force Road_Surface_Conditions           Road_Type  \
0  Metropolitan Police                     Dry    Dual carriageway
1  Metropolitan Police                     Dry  Single carriageway
2  Metropolitan Police                     Dry  Single carriageway
3  Metropolitan Police             Wet or damp  Single carriageway
4  Metropolitan Police             Wet or damp  Single carriageway

  Speed_limit   Time Urban_or_Rural_Area     Weather_Conditions Year_y  \
0          30  17:36               Urban     Fine no high winds   2005
1          30  00:15               Urban     Fine no high winds   2005
2          30  10:35               Urban     Fine no high winds   2005
3          30  21:13               Urban     Fine no high winds   2005
4          30  12:40               Urban  Raining no high winds   2005

   InScotland
0          No
1          No
2          No
3          No
4          No

[5 rows x 48 columns]
```

```
[124]: count_col = []
hist_col = []
for column in df.columns:
    unique_value = df[column].nunique()
    if unique_value <= 20:
        count_col.append(column)
    else:
        hist_col.append(column)
```

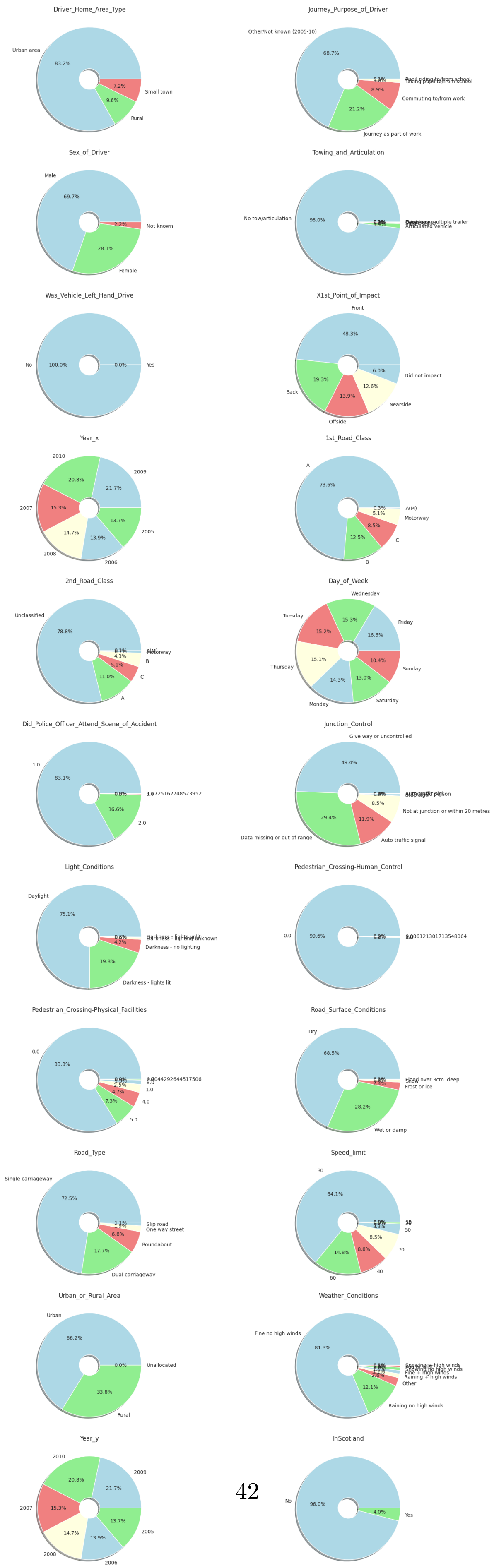
[125]:

```
#count plot
plt.figure(figsize=(20,80))
plot_num = 1
for col in count_col:
    plt.subplot(20,2,plot_num)
    sns.countplot(data=df, x=col)
    plot_num += 1
    plt.tight_layout()
    plt.xticks(rotation=45)
```


count
100000
0
36 - 45
26 - 35
46 - 55
66 - 75
16 - 20
56 - 65
21 - 25
Over 75
11 - 15
6 - 10
0 - 5
Age_Band_of_Driver
count
250000
0
Urban area
Rural
Small town
Driver_Home_Area_Type
count
100000
0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
9.0
10.0
Driver_IMD_Decile
count
200000
0
Journey as part of work
Other/Not known (2005-10)
Commuting to/from work
Pupil riding to/from school
Taking pupil to/from school
Journey_Purpose_of_Driver
count
200000
0
Leaving roundabout
Not at or within 20 metres of junction
Approaching junction or waiting/parked at junction approach
Mid Junction - on roundabout or on main road
Cleared junction or waiting/parked at junction exit
Entering main road
Entering from slip road
Leaving main road
Entering roundabout
Junction_Location
count
200000
0
Heavy oil
Petrol
Hybrid electric
Gas/Bi-fuel
Gas
Petrol/Gas (LPG)
Electric
New fuel technology
Gas Diesel
Propulsion_Code
count
200000
0
Male
Female
Not known
Sex_of_Driver
count
500000
0
No tow/articulation
Articulated vehicle
Other tow
Single trailer
Double or multiple trailer
Caravan
Towing_and_Articulation
count
250000
0
Did not leave carriageway
Nearside
Offside
Straight ahead at junction
Offside on to central reservation
Offside - crossed central reservation
Nearside and rebounded
Offside on to centrl res + rebounded
Offside and rebounded
Vehicle_Leaving_Carriageway
count
500000
0
0.0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
9.0
Vehicle_Location.Restricted_Lane
count
200000
0
Slowing or stopping
Going ahead right-hand bend
Going ahead other
Moving off
Parked
Turning left
Turning right
U-turn
Overtaking moving vehicle - offside
Waiting to go - held up
Overtaking static vehicle - offside
Waiting to turn right
Going ahead left-hand bend
Changing lane to right
Changing lane to left
Reversing
Waiting to turn left
Overtaking - nearside
Vehicle_Manoeuvre
count
250000
0
Bus or coach (17 or more pass seats)
Car
Motorcycle 125cc and under
Motorcycle over 500cc
Motorcycle over 125cc and up to 500cc
Taxi/Private hire car
Goods 7.5 tonnes mgw and over
Van / Goods 3.5 tonnes mgw or under
Motorcycle 50cc and under
Goods over 3.5t. and under 7.5t
Minibus (8 - 16 passenger seats)
Other vehicle
Agricultural vehicle
Pedal cycle
Tram
Ridden horse
Vehicle_Type
count
500000
0
No
Yes
Was_Vehicle_Left_Hand_Drive
count
200000
0
Nearside
Front
Did not impact
Back
Offside
X1st_Point_of_Impact
count
100000
0
2005
2006
2007
2008
2009
2010
Year_x
count
250000
0
B
C
A
Motorway
A(M)
1st_Road_Class
count
250000
0
C
Unclassified
B
A
Motorway
A(M)
2nd_Road_Class
count
250000
0
Slight
Serious
Fatal
Accident_Severity
count
50000
0
Wednesday
Thursday
Friday
Monday
Tuesday
Sunday
Saturday
Day_of_Week
count
250000
0
1.0
1.1725162748523952
2.0
3.0
Did_Police_Officer_Attend_Scene_of_Accident
count
200000
0
Auto traffic signal
Data missing or out of range
Give way or uncontrolled
Stop sign
Authorised person
Auto traffic sigl
Not at junction or within 20 metres
Junction_Control
count
200000
0
Crossroads
Not at junction or within 20 metres
T or staggered junction
Mini-roundabout
Private drive or entrance
More than 4 arms (not roundabout)
Roundabout
Other junction
Slip road
Junction_Detail
count
250000
0
Darkness - lights lit
Daylight
Darkness - lighting unknown
Darkness - lights unlit
Darkness - no lighting
Light_Conditions
count
200000
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
16
17
18
19
20
32
Number_of_Vehicles
count
500000
0
0.0
0.006121301713548064
1.0
2.0
Pedestrian_Crossing-Human_Control
count
250000
0
0.0
0.7044292644517506
1.0
4.0
5.0
7.0
8.0
Pedestrian_Crossing-Physical_Facilities
count
200000
0
Dry
Wet or damp
Snow
Frost or ice
Flood over 3cm. deep
Road_Surface_Conditions
count
250000
0
Dual carriageway
Single carriageway
One way street
Roundabout
Slip road
Road_Type
count
200000
0
10
15
20
30
40
50
60
70
Speed_limit
count
200000
0
Urban
Rural
Unallocated
Urban_or_Rural_Area
count
250000
0
Fine no high winds
Raining no high winds
Snowing no high winds
Fine + high winds
Other
Raining + high winds
Fog or mist
Snowing + high winds
Weather_Conditions
count
100000
0
2005
2006
2007
2008
2009
2010
Year_y
count
500000
0
No
Yes
InScotland

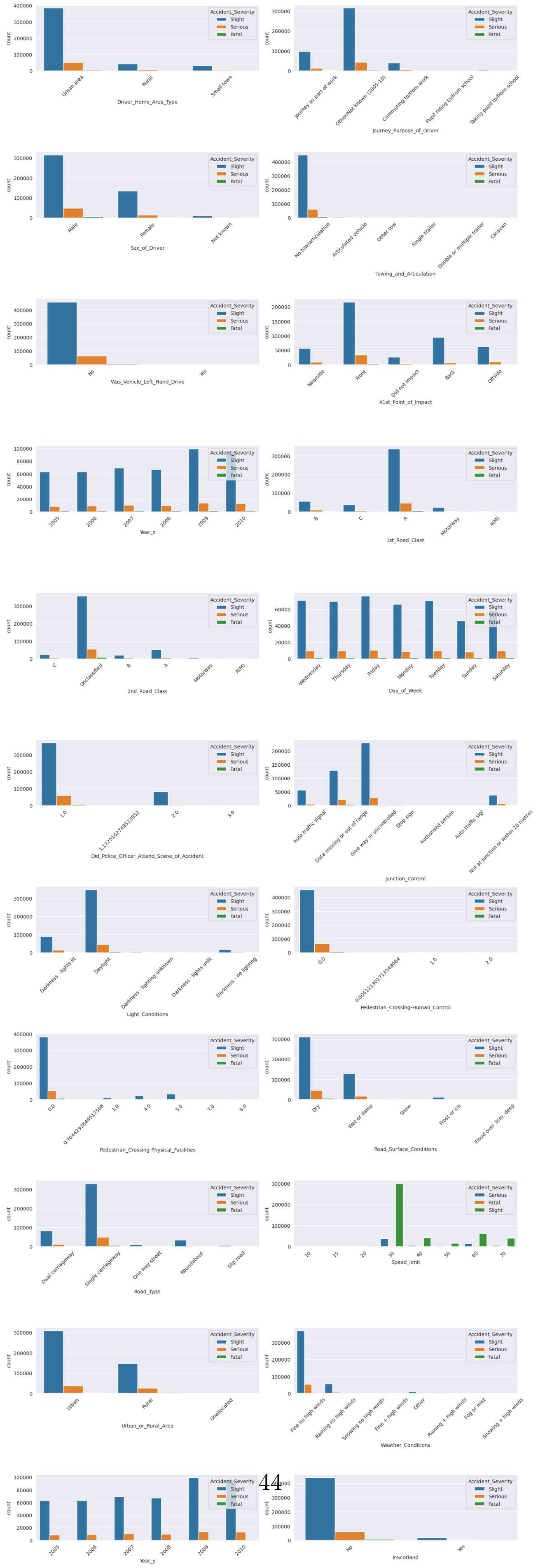
[126]:

```
#donut_chart
plt.figure(figsize=(15,80))
plot_num = 1
for col in count_col:
    if df[col].nunique() <= 8 and col != "Accident_Severity":
        plt.subplot(20,2,plot_num)
        plt.pie(data=df, x=df[col].value_counts().values,labels=df[col].
 ↪value_counts().index,shadow=True,autopct='%1.1f%%',
               colors=['lightblue', 'lightgreen', 'lightcoral',␣
 ↪'lightyellow'],wedgeprops=dict(width=0.8,edgecolor="w") )
        plt.title(col)
        plot_num += 1
        plt.tight_layout()
```


Driver_Home_Area_Type
Urban area
83.2%
7.2%
Small town
9.6%
Rural
Journey_Purpose_of_Driver
Other/Not known (2005-10)
68.7%
0.1%
1.0%
Pupil riding to/from school
Taking pupil to/from school
8.9%
Commuting to/from work
21.2%
Journey as part of work
Sex_of_Driver
Male
69.7%
2.2%
Not known
28.1%
Female
Towing_and_Articulation
No tow/articulation
98.0%
1.4%
Double or multiple trailer
Articulated vehicle
Was_Vehicle_Left_Hand_Drive
No
100.0%
0.0%
Yes
X1st_Point_of_Impact
Front
48.3%
6.0%
Did not impact
19.3%
Back
13.9%
Offside
12.6%
Nearside
Year_x
2010
20.8%
2009
21.7%
2007
15.3%
14.7%
2008
13.9%
2006
13.7%
2005
1st_Road_Class
A
73.6%
0.3%
A(M)
5.1%
Motorway
8.5%
C
12.5%
B
2nd_Road_Class
Unclassified
78.8%
0.1%
0.7%
A(M)
Motorway
4.3%
B
5.1%
C
11.0%
A
Day_of_Week
Wednesday
15.3%
Tuesday
15.2%
Friday
16.6%
Thursday
15.1%
10.4%
Sunday
14.3%
Monday
13.0%
Saturday
Did_Police_Officer_Attend_Scene_of_Accident
1.0
83.1%
0.0%
0.3%
3.0
1.0725162748523952
16.6%
2.0
Junction_Control
Give way or uncontrolled
49.4%
0.0%
0.6%
Auto traffic sigl
Authorised person
Stop sign
8.5%
Not at junction or within 20 metres
29.4%
Data missing or out of range
11.9%
Auto traffic signal
Light_Conditions
Daylight
75.1%
0.3%
0.6%
Darkness - lights unlit
Darkness - lighting unknown
4.2%
Darkness - no lighting
19.8%
Darkness - lights lit
Pedestrian_Crossing-Human_Control
0.0
99.6%
0.2%
0.0%
1.0
2.0
0.006121301713548064
Pedestrian_Crossing-Physical_Facilities
0.0
83.8%
0.0%
1.3%
7.0
2.044292644517506
8.0
2.5%
1.0
4.7%
4.0
7.3%
5.0
Road_Surface_Conditions
Dry
68.5%
0.1%
0.8%
Flood over 3cm. deep
Snow
2.4%
Frost or ice
28.2%
Wet or damp
Road_Type
Single carriageway
72.5%
1.1%
Slip road
1.9%
One way street
6.8%
Roundabout
17.7%
Dual carriageway
Speed_limit
30
64.1%
0.0%
0.5%
10
20
3.3%
50
8.5%
70
14.8%
60
8.8%
40
Urban_or_Rural_Area
Urban
66.2%
0.0%
Unallocated
33.8%
Rural
Weather_Conditions
Fine no high winds
81.3%
0.1%
1.0%
1.2%
2.6%
Snowing + high winds
Fog or mist
Snowing no high winds
Fine + high winds
Raining + high winds
Other
12.1%
Raining no high winds
Year_y
2010
20.8%
2009
21.7%
2007
15.3%
14.7%
2008
13.9%
2006
13.7%
2005
InScotland
No
96.0%
4.0%
Yes

[127]:

```
#count plots with hue accident severity
plt.figure(figsize=(15,80))
plot_num = 1
for col in count_col:
    if df[col].nunique() <= 8 and col != "Accident_Severity":
        plt.subplot(20,2,plot_num)
        sns.countplot(data=df, x=col, hue="Accident_Severity")
        plot_num += 1
        plt.tight_layout()
        plt.xticks(rotation=45)
```


Accident_Severity
Slight
Serious
Fatal
count
0
100000
200000
300000
400000
Urban area
Rural
Small town
Driver_Home_Area_Type
Journey as part of work
Other/Not known (2005-10)
Commuting to/from work
Pupil riding to/from school
Taking pupil to/from school
Journey_Purpose_of_Driver
Male
Female
Not known
Sex_of_Driver
No tow/articulation
Articulated vehicle
Other tow
Single trailer
Double or multiple trailer
Caravan
Towing_and_Articulation
No
Yes
Was_Vehicle_Left_Hand_Drive
50000
150000
Nearside
Front
Did not impact
Back
Offside
X1st_Point_of_Impact
20000
40000
60000
80000
2005
2006
2007
2008
2009
2010
Year_x
B
C
A
Motorway
A(M)
1st_Road_Class
C
Unclassified
B
A
Motorway
A(M)
2nd_Road_Class
Wednesday
Thursday
Friday
Monday
Tuesday
Sunday
Saturday
Day_of_Week
1.0
1.1725162748523952
2.0
3.0
Did_Police_Officer_Attend_Scene_of_Accident
Auto traffic signal
Data missing or out of range
Give way or uncontrolled
Stop sign
Authorised person
Auto traffic sigl
Not at junction or within 20 metres
Junction_Control
Darkness - lights lit
Daylight
Darkness - lighting unknown
Darkness - lights unlit
Darkness - no lighting
Light_Conditions
0.0
0.006121301713548064
1.0
2.0
Pedestrian_Crossing-Human_Control
0.0
0.7044292644517506
1.0
4.0
5.0
7.0
8.0
Pedestrian_Crossing-Physical_Facilities
Dry
Wet or damp
Snow
Frost or ice
Flood over 3cm. deep
Road_Surface_Conditions
Dual carriageway
Single carriageway
One way street
Roundabout
Slip road
Road_Type
Serious
Fatal
Slight
10
15
20
30
40
50
60
70
Speed_limit
Urban
Rural
Unallocated
Urban_or_Rural_Area
Fine no high winds
Raining no high winds
Snowing no high winds
Fine + high winds
Other
Raining + high winds
Fog or mist
Snowing + high winds
Weather_Conditions
Year_y
No
Yes
InScotland

[128]:

```
severities = df['Accident_Severity']
severity_levels = severities.unique()
severity_levels
```

[128]:

```
array(['Slight', 'Serious', 'Fatal'], dtype=object)
```

[129]:

```
severities.value_counts()
```

[129]:

```
Accident_Severity
Slight     457171
Serious     63913
Fatal        7251
Name: count, dtype: int64
```

[130]:

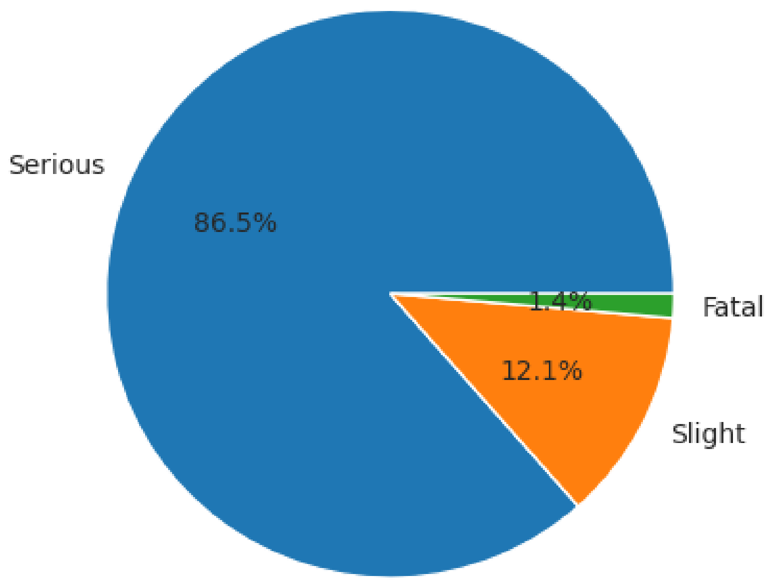
```
#Calculate Percentages
severities.value_counts()*100/len(severities)
```

[130]:

```
Accident_Severity
Slight     86.530516
Serious    12.097060
Fatal       1.372425
Name: count, dtype: float64
```

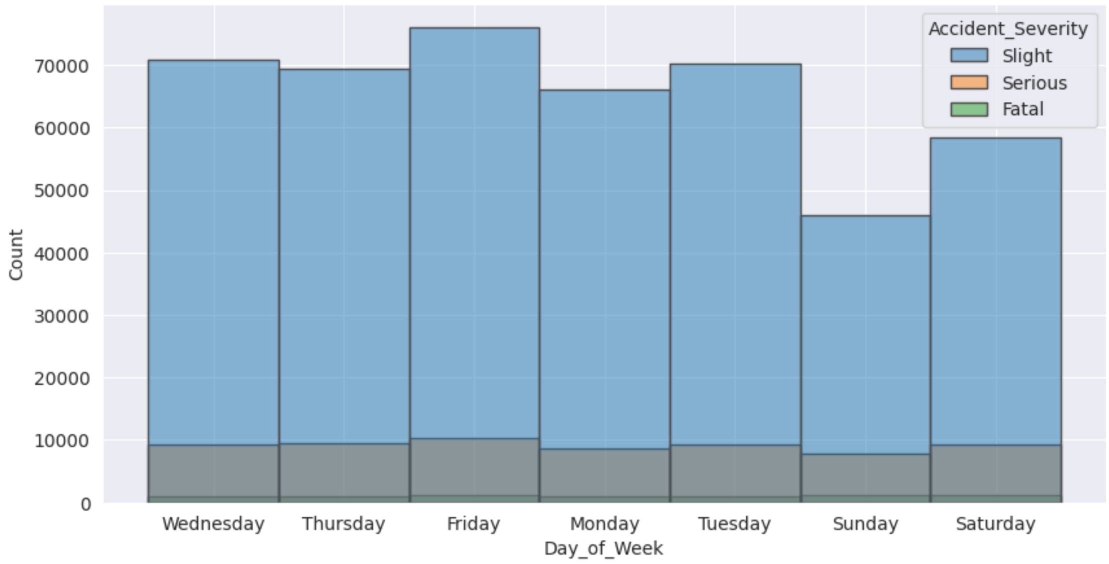
[131]:

```
#piechart
plt.pie(df.Accident_Severity.value_counts(), labels = ["Serious",␣
 ↪"Slight","Fatal"], autopct = '%1.1f%%')
pass
```



[132]:

```
ax = plt.subplots(1,figsize = (10,5))
sns.histplot( x = "Day_of_Week", hue = "Accident_Severity", edgecolor = ".3",␣
  ↪data=df)
pass
```



[133]:

```
df['Date']
```

[133]:

```
0         05/01/2005
1         06/01/2005
2         07/01/2005
3         10/01/2005
4         11/01/2005
             …
528330    09/06/2010
528331    08/06/2010
528332    08/06/2010
528333    10/06/2010
528334    10/06/2010
Name: Date, Length: 528335, dtype: object
```

[134]:

```
#Time:
times = df.Time
times
```

[134]:

```
0         17:36
1         00:15
2         10:35
3         21:13
4         12:40
          …
528330    10:20
528331    12:00
528332    12:00
528333    13:07
528334    13:07
Name: Time, Length: 528335, dtype: object
```

[135]:

```
# drop rows with time null:
df.dropna(subset=['Time'],inplace=True)
```

[136]:

```
# Create a new datetime columns based on the individual values given

df['DateTime'] = df['Date']+' '+ df['Time']
df.head()
```

[136]:

```
  Age_Band_of_Driver  Age_of_Vehicle Driver_Home_Area_Type  Driver_IMD_Decile  \
0            36 - 45             3.0            Urban area                5.0
1            26 - 35             5.0            Urban area                3.0
2            46 - 55             4.0            Urban area                1.0
3            46 - 55            10.0            Urban area                5.0
4            46 - 55             1.0            Urban area                4.0
```

```
   Engine_Capacity_.CC.  Journey_Purpose_of_Driver  \
0                8268.0    Journey as part of work
1                8300.0    Journey as part of work
2                1769.0  Other/Not known (2005-10)
3                  85.0  Other/Not known (2005-10)
4                2976.0  Other/Not known (2005-10)

                      Junction_Location    make                    model  \
0                     Leaving roundabout  DENNIS       CLIO DYNAMIQUE 16V
1  Not at or within 20 metres of junction  DENNIS       CLIO DYNAMIQUE 16V
2  Not at or within 20 metres of junction  NISSAN           ALMERA SE AUTO
3  Not at or within 20 metres of junction   HONDA       CLIO DYNAMIQUE 16V
4  Not at or within 20 metres of junction    AUDI  A4 SPORT CABRIOLET AUTO

  Propulsion_Code  …          Police_Force Road_Surface_Conditions  \
0       Heavy oil  …  Metropolitan Police                     Dry
1       Heavy oil  …  Metropolitan Police                     Dry
2          Petrol  …  Metropolitan Police                     Dry
3          Petrol  …  Metropolitan Police              Wet or damp
4          Petrol  …  Metropolitan Police              Wet or damp

            Road_Type  Speed_limit   Time  Urban_or_Rural_Area  \
0   Dual carriageway           30  17:36                Urban
1  Single carriageway          30  00:15                Urban
2  Single carriageway          30  10:35                Urban
3  Single carriageway          30  21:13                Urban
4  Single carriageway          30  12:40                Urban

      Weather_Conditions Year_y InScotland          DateTime
0     Fine no high winds   2005         No  05/01/2005 17:36
1     Fine no high winds   2005         No  06/01/2005 00:15
2     Fine no high winds   2005         No  07/01/2005 10:35
3     Fine no high winds   2005         No  10/01/2005 21:13
4  Raining no high winds   2005         No  11/01/2005 12:40

[5 rows x 49 columns]
```

[137]:

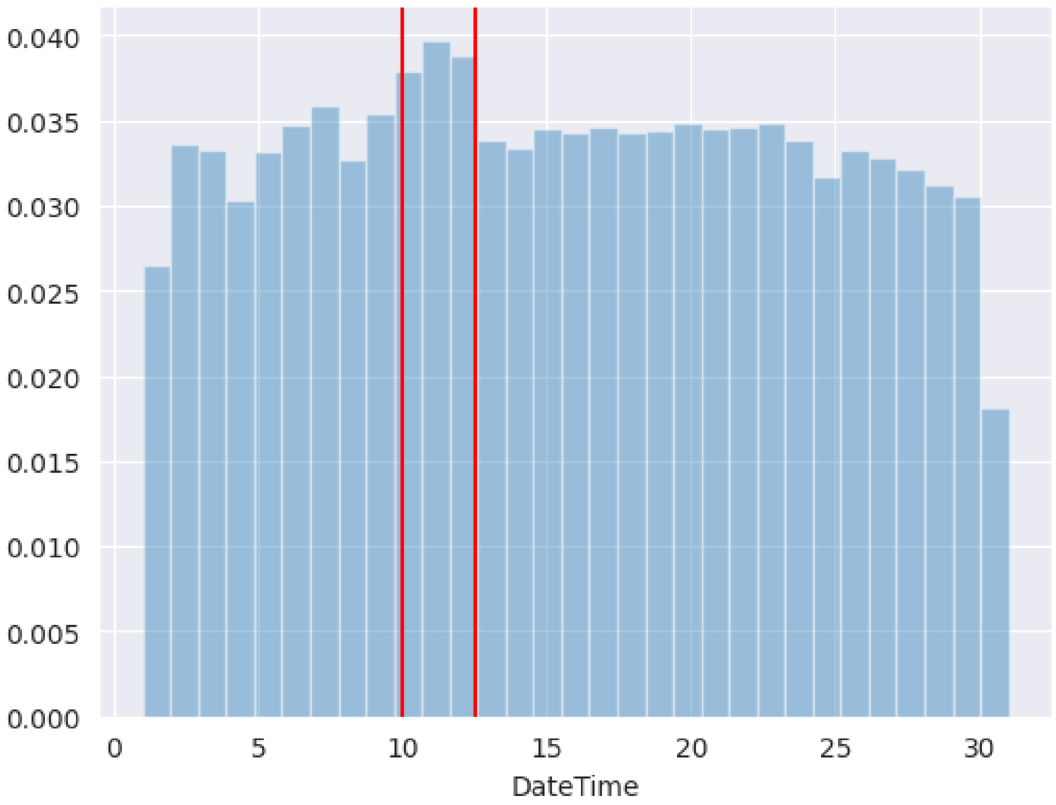
```
df['DateTime'] = pd.to_datetime(df['DateTime'],format='mixed') # Convert␣
 ↪Datetime to pandas datetime
```

[138]:

```
sns.distplot(df.DateTime.dt.day, bins=31, kde=False, norm_hist=True)
plt.axvline(10.0, 0,color='red')
plt.axvline(12.5, 0,color='red')
```

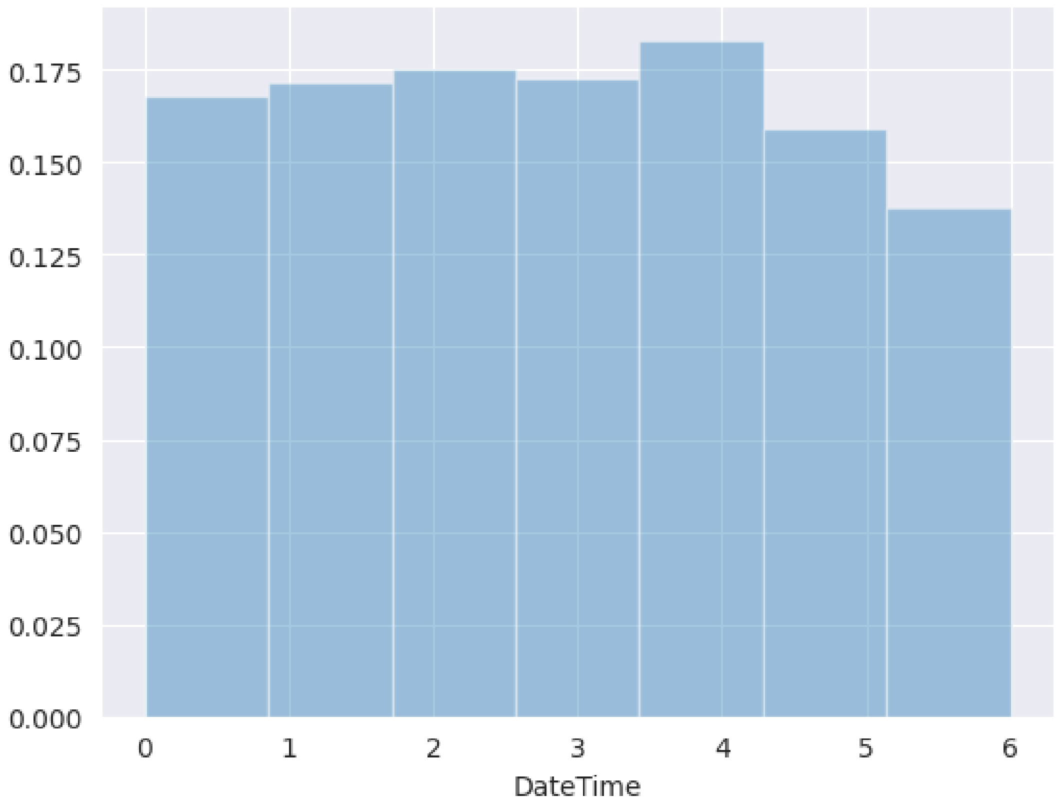
[138]:

```
<matplotlib.lines.Line2D at 0x790128cb62c0>
```



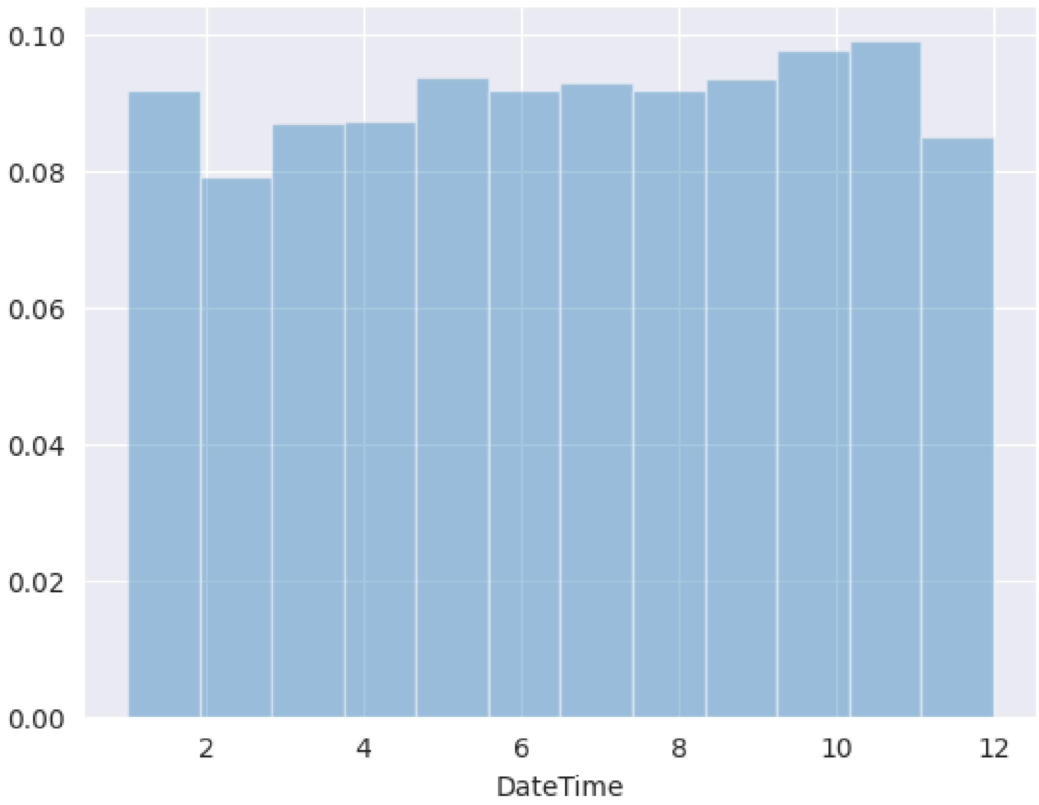
```
[139]: sns.distplot(df.DateTime.dt.day_of_week, bins=7, kde=False, norm_hist=True)
```

```
[139]: <Axes: xlabel='DateTime'>
```



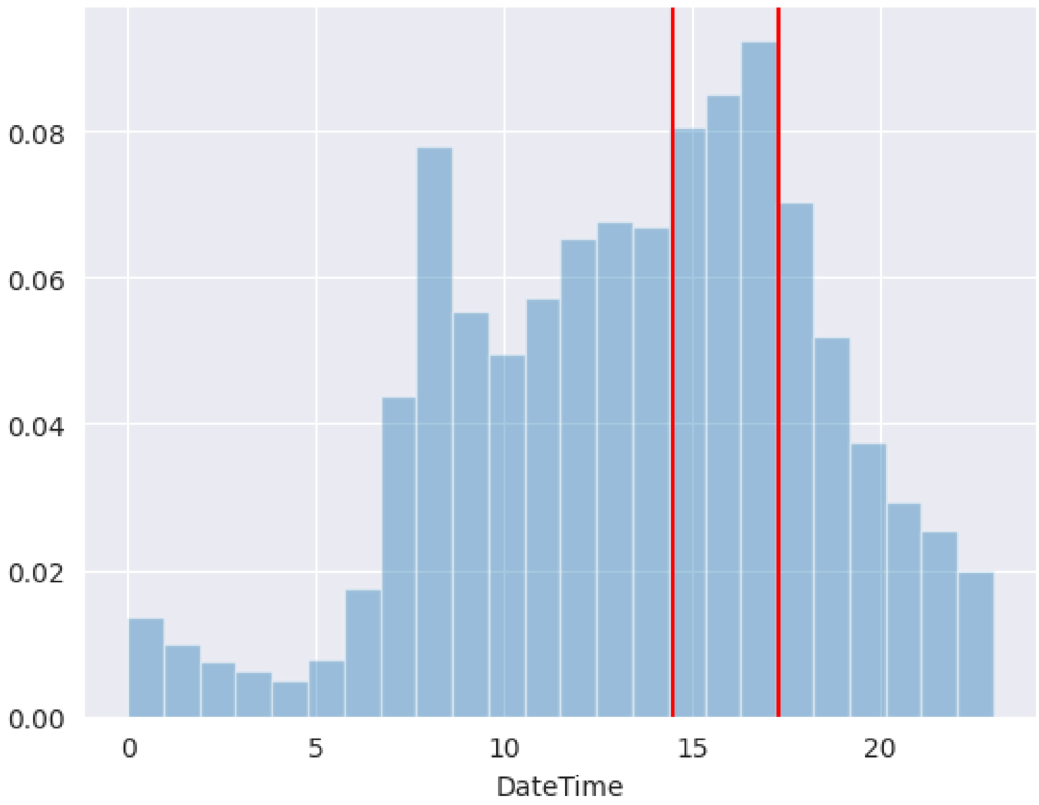
```
[140]: sns.distplot(df.DateTime.dt.month, bins=12, kde=False, norm_hist=True)
```

```
[140]: <Axes: xlabel='DateTime'>
```



[141]:
```
## Analyze time of occurance of collisions
sns.distplot(df.DateTime.dt.hour, bins=24, kde=False, norm_hist=True)

plt.axvline(14.5, 0,color='red')
plt.axvline(17.3, 0,color='red')
```

[141]:
```
<matplotlib.lines.Line2D at 0x79010555fe50>
```



```
[142]: police_responses = df.Did_Police_Officer_Attend_Scene_of_Accident
       police_responses.value_counts()
```

```
[142]: Did_Police_Officer_Attend_Scene_of_Accident
       1.000000    438939
       2.000000     87525
       3.000000      1805
       1.172516        66
       Name: count, dtype: int64
```

```
[143]: #calculate police response by percentages
       response_percentages = police_responses.value_counts()*100/len(police_responses)
       response_percentages
```

```
[143]: Did_Police_Officer_Attend_Scene_of_Accident
       1.000000    83.079675
       2.000000    16.566194
       3.000000     0.341639
       1.172516     0.012492
       Name: count, dtype: float64
```

[144]:

```
#get high casualty accidents:
high_casualty = df[df.Number_of_Casualties>5]
high_casualty.head()
```

[144]:

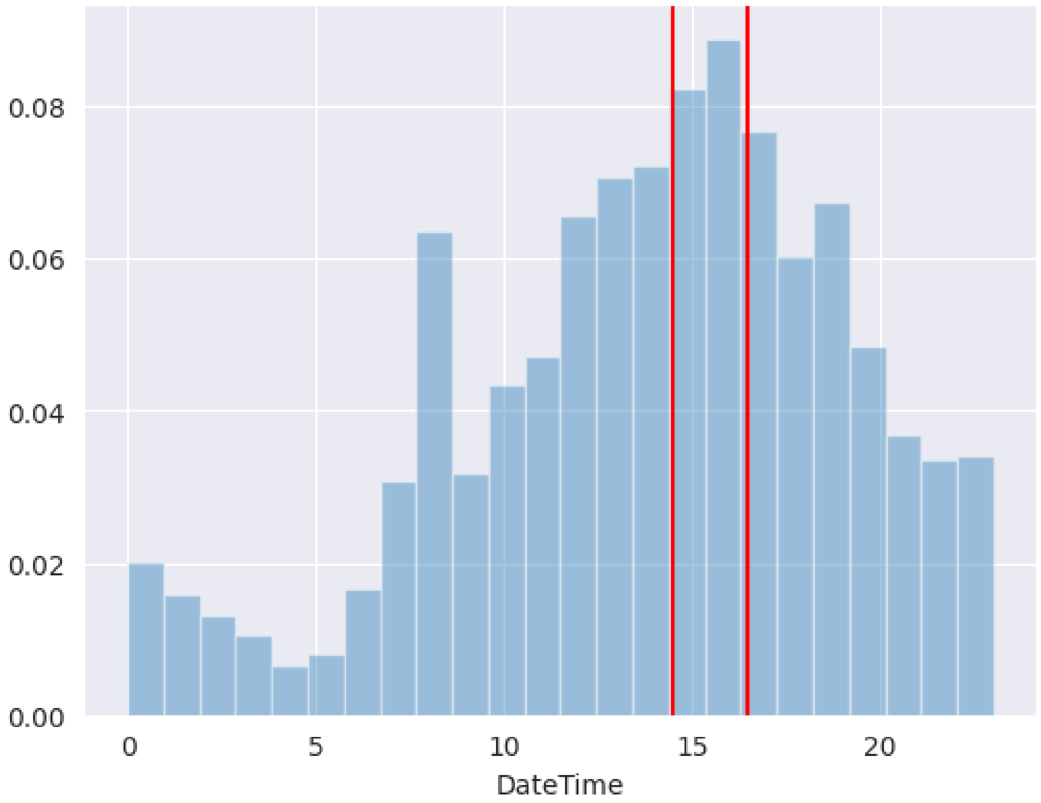
```
     Age_Band_of_Driver  Age_of_Vehicle Driver_Home_Area_Type  \
3837            21 - 25             7.0           Urban area
4253            36 - 45             1.0           Urban area
4602            56 - 65             5.0           Urban area
4603            36 - 45             2.0           Urban area
4604            26 - 35            10.0           Urban area

      Driver_IMD_Decile  Engine_Capacity_.CC.  Journey_Purpose_of_Driver  \
3837                3.0                1598.0  Other/Not known (2005-10)
4253                3.0                1124.0  Other/Not known (2005-10)
4602                5.0               11000.0    Journey as part of work
4603                9.0               12130.0    Journey as part of work
4604                2.0               12130.0    Journey as part of work

                           Junction_Location     make                 model  \
3837  Not at or within 20 metres of junction  RENAULT  MEGANE RT SPORT 1.6E
4253                      Leaving roundabout  PEUGEOT                 206 S
4602  Not at or within 20 metres of junction      ERF    CLIO DYNAMIQUE 16V
4603  Not at or within 20 metres of junction    VOLVO    CLIO DYNAMIQUE 16V
4604  Not at or within 20 metres of junction    VOLVO    CLIO DYNAMIQUE 16V

     Propulsion_Code  …         Police_Force Road_Surface_Conditions  \
3837          Petrol  …  Metropolitan Police             Wet or damp
4253          Petrol  …  Metropolitan Police                     Dry
4602       Heavy oil  …  Metropolitan Police                     Dry
4603       Heavy oil  …  Metropolitan Police                     Dry
4604       Heavy oil  …  Metropolitan Police                     Dry

               Road_Type  Speed_limit   Time  Urban_or_Rural_Area  \
3837  Single carriageway           30  21:38                Urban
4253  Single carriageway           30  12:35                Urban
4602    Dual carriageway           70  06:26                Rural
4603    Dual carriageway           70  06:26                Rural
4604    Dual carriageway           70  06:26                Rural

         Weather_Conditions Year_y InScotland            DateTime
3837  Raining no high winds   2005         No 2005-08-13 21:38:00
4253     Fine no high winds   2005         No 2005-08-15 12:35:00
4602     Fine no high winds   2005         No 2005-07-06 06:26:00
4603     Fine no high winds   2005         No 2005-07-06 06:26:00
4604     Fine no high winds   2005         No 2005-07-06 06:26:00

[5 rows x 49 columns]
```

[145]:

```
# when do these high casualty collisions occur happen?
sns.distplot(high_casualty.DateTime.dt.hour, bins=24, kde=False, norm_hist=True)
plt.axvline(14.5, 0,color='red')
plt.axvline(16.5, 0,color='red')
```

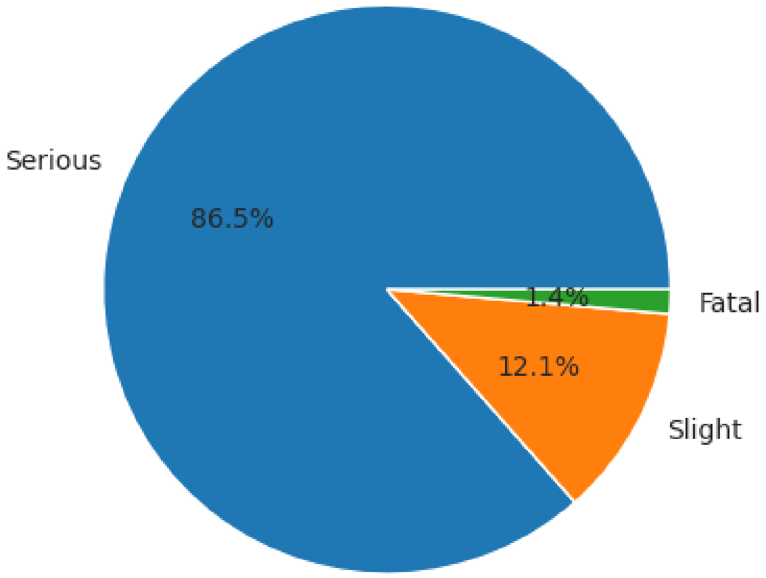
[145]:

```
<matplotlib.lines.Line2D at 0x79010555f880>
```



[146]:

```
#piechart
plt.pie(df.Accident_Severity.value_counts(), labels = ["Serious",␣
 ↪"Slight","Fatal"], autopct = '%1.1f%%')
pass
```



```
[147]: import folium
       from folium.plugins import FastMarkerCluster
```

```
[148]: df.Latitude
```

```
[148]: 0         51.520075
       1         51.525301
       2         51.482442
       3         51.495752
       4         51.515540
                   …
       528330    52.398088
       528331    52.418163
       528332    52.418163
       528333    52.390211
       528334    52.390211
       Name: Latitude, Length: 528335, dtype: float64
```

```
[149]: sample_data = df.sample(int(0.001 * len(df)))
       lat_lon_pairs = list(zip(list(sample_data.Latitude), list(sample_data.
        ↪Longitude)))

       map = folium.Map()
```

```
HeatMap(lat_lon_pairs).add_to(map)
FastMarkerCluster(lat_lon_pairs).add_to(map)

map
```

```
[149]: <folium.folium.Map at 0x790098210160>
```

```
[150]: df.dtypes
```

```
[150]: Age_Band_of_Driver                                     object
       Age_of_Vehicle                                        float64
       Driver_Home_Area_Type                                  object
       Driver_IMD_Decile                                     float64
       Engine_Capacity_.CC.                                  float64
       Journey_Purpose_of_Driver                              object
       Junction_Location                                      object
       make                                                   object
       model                                                  object
       Propulsion_Code                                        object
       Sex_of_Driver                                          object
       Towing_and_Articulation                                object
       Vehicle_Leaving_Carriageway                            object
       Vehicle_Location.Restricted_Lane                      float64
       Vehicle_Manoeuvre                                      object
       Vehicle_Reference                                       int64
       Vehicle_Type                                           object
       Was_Vehicle_Left_Hand_Drive                            object
       X1st_Point_of_Impact                                   object
       Year_x                                                  int64
       1st_Road_Class                                         object
       1st_Road_Number                                       float64
       2nd_Road_Class                                         object
       Accident_Severity                                      object
       Date                                                   object
       Day_of_Week                                            object
       Did_Police_Officer_Attend_Scene_of_Accident           float64
       Junction_Control                                       object
       Junction_Detail                                        object
       Latitude                                              float64
       Light_Conditions                                       object
       Local_Authority_(District)                             object
       Local_Authority_(Highway)                              object
       Longitude                                             float64
       LSOA_of_Accident_Location                              object
       Number_of_Casualties                                    int64
       Number_of_Vehicles                                      int64
       Pedestrian_Crossing-Human_Control                     float64
```

```
Pedestrian_Crossing-Physical_Facilities               float64
Police_Force                                           object
Road_Surface_Conditions                                object
Road_Type                                              object
Speed_limit                                             int64
Time                                                   object
Urban_or_Rural_Area                                    object
Weather_Conditions                                     object
Year_y                                                  int64
InScotland                                             object
DateTime                                       datetime64[ns]
dtype: object
```

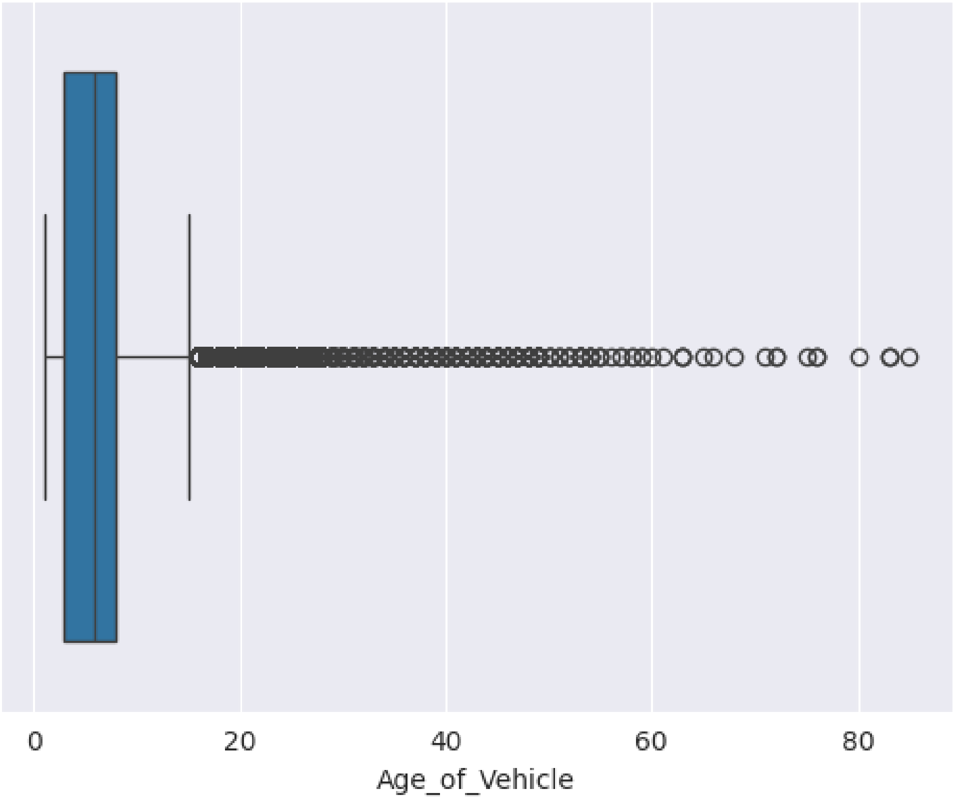
```
[151]: df.drop(['Date','Time','Year_y','Year_x','DateTime'],axis=1,inplace=True)
       df.head()
```

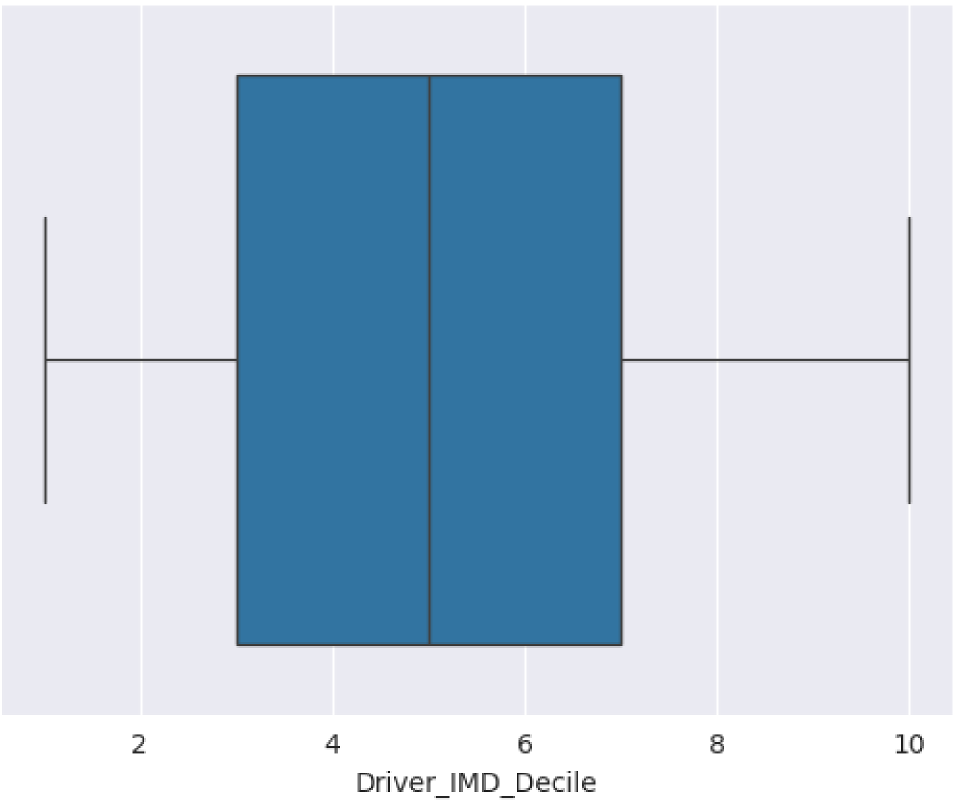
```
[151]:   Age_Band_of_Driver  Age_of_Vehicle Driver_Home_Area_Type  Driver_IMD_Decile  \
       0            36 - 45             3.0            Urban area                5.0
       1            26 - 35             5.0            Urban area                3.0
       2            46 - 55             4.0            Urban area                1.0
       3            46 - 55            10.0            Urban area                5.0
       4            46 - 55             1.0            Urban area                4.0

          Engine_Capacity_.CC.  Journey_Purpose_of_Driver  \
       0                8268.0    Journey as part of work
       1                8300.0    Journey as part of work
       2                1769.0  Other/Not known (2005-10)
       3                  85.0  Other/Not known (2005-10)
       4                2976.0  Other/Not known (2005-10)

                               Junction_Location    make                    model  \
       0                      Leaving roundabout  DENNIS       CLIO DYNAMIQUE 16V
       1  Not at or within 20 metres of junction  DENNIS       CLIO DYNAMIQUE 16V
       2  Not at or within 20 metres of junction  NISSAN           ALMERA SE AUTO
       3  Not at or within 20 metres of junction   HONDA       CLIO DYNAMIQUE 16V
       4  Not at or within 20 metres of junction    AUDI  A4 SPORT CABRIOLET AUTO

         Propulsion_Code  … Number_of_Vehicles Pedestrian_Crossing-Human_Control  \
       0       Heavy oil  …                  1                               0.0
       1       Heavy oil  …                  2                               0.0
       2          Petrol  …                  1                               0.0
       3          Petrol  …                  1                               0.0
       4          Petrol  …                  2                               0.0

         Pedestrian_Crossing-Physical_Facilities         Police_Force  \
       0                                     5.0  Metropolitan Police
       1                                     0.0  Metropolitan Police
```

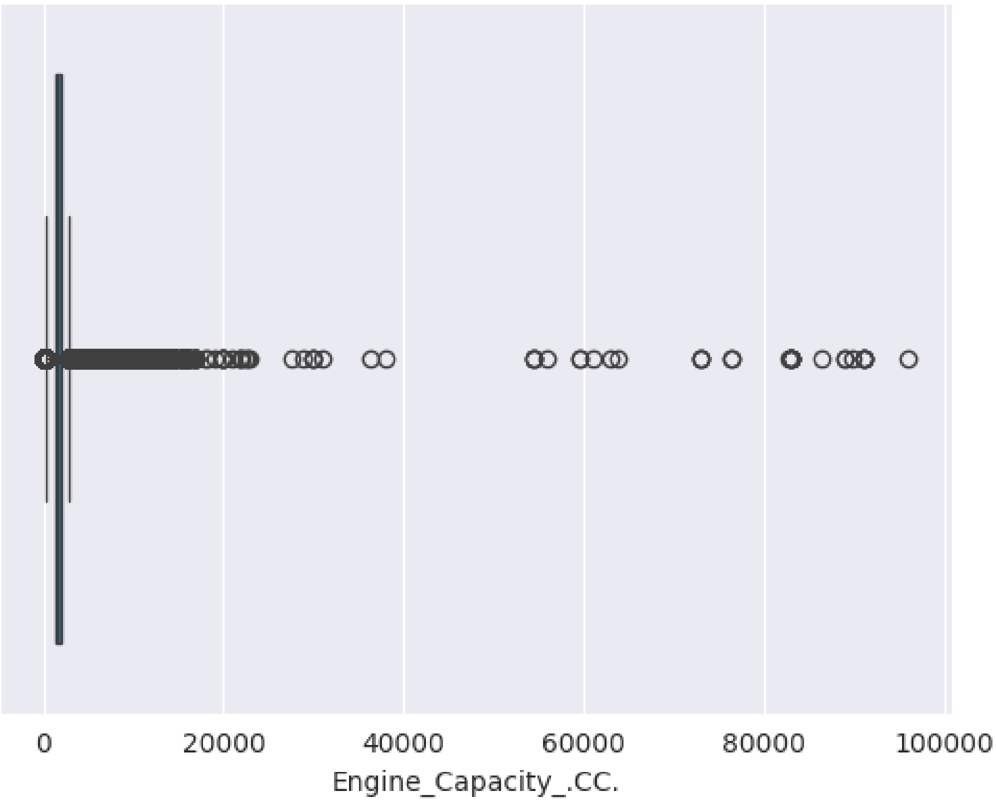
```
2                                     0.0  Metropolitan Police
3                                     0.0  Metropolitan Police
4                                     0.0  Metropolitan Police

  Road_Surface_Conditions           Road_Type Speed_limit Urban_or_Rural_Area  \
0                     Dry    Dual carriageway          30               Urban
1                     Dry  Single carriageway          30               Urban
2                     Dry  Single carriageway          30               Urban
3             Wet or damp  Single carriageway          30               Urban
4             Wet or damp  Single carriageway          30               Urban

      Weather_Conditions InScotland
0     Fine no high winds         No
1     Fine no high winds         No
2     Fine no high winds         No
3     Fine no high winds         No
4  Raining no high winds         No

[5 rows x 44 columns]
```

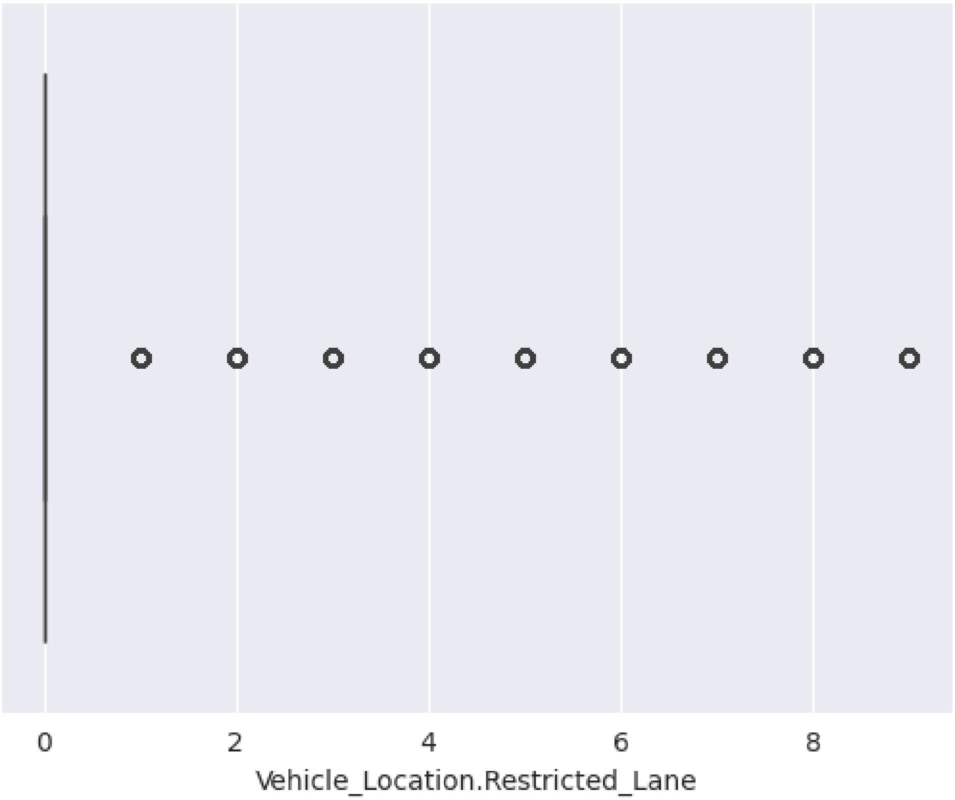
[152]:

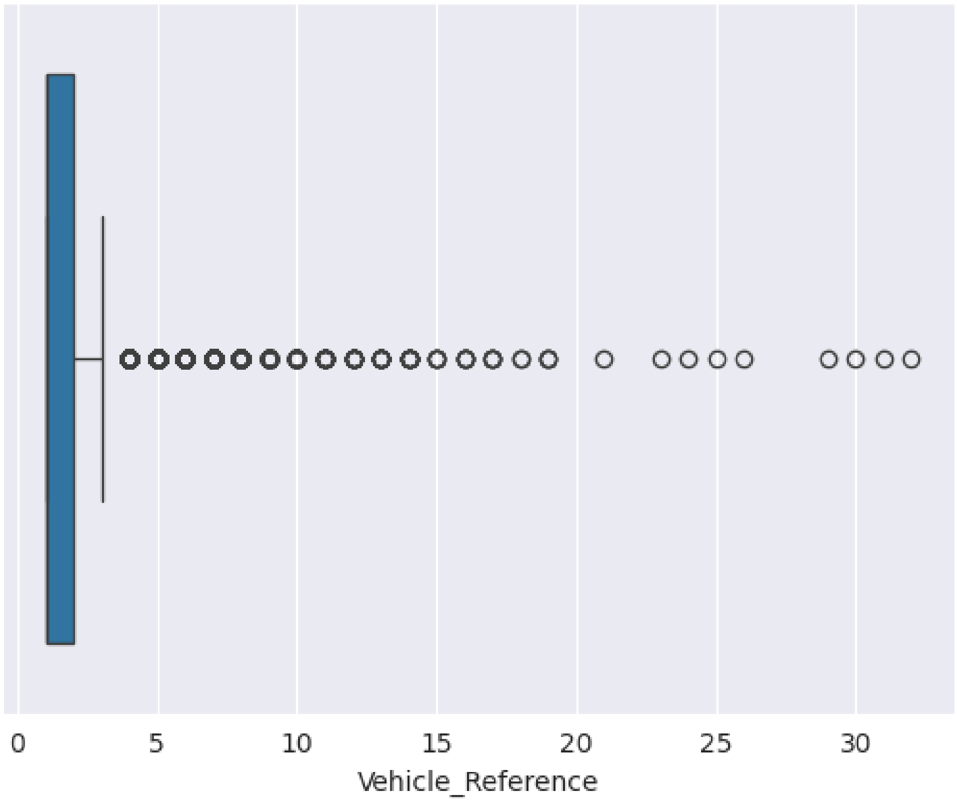
```
#boxplot
data1=df.select_dtypes(include=['int64','float64'])
for i in data1:
    sns.boxplot(x=i, data=df)
    plt.show()
```

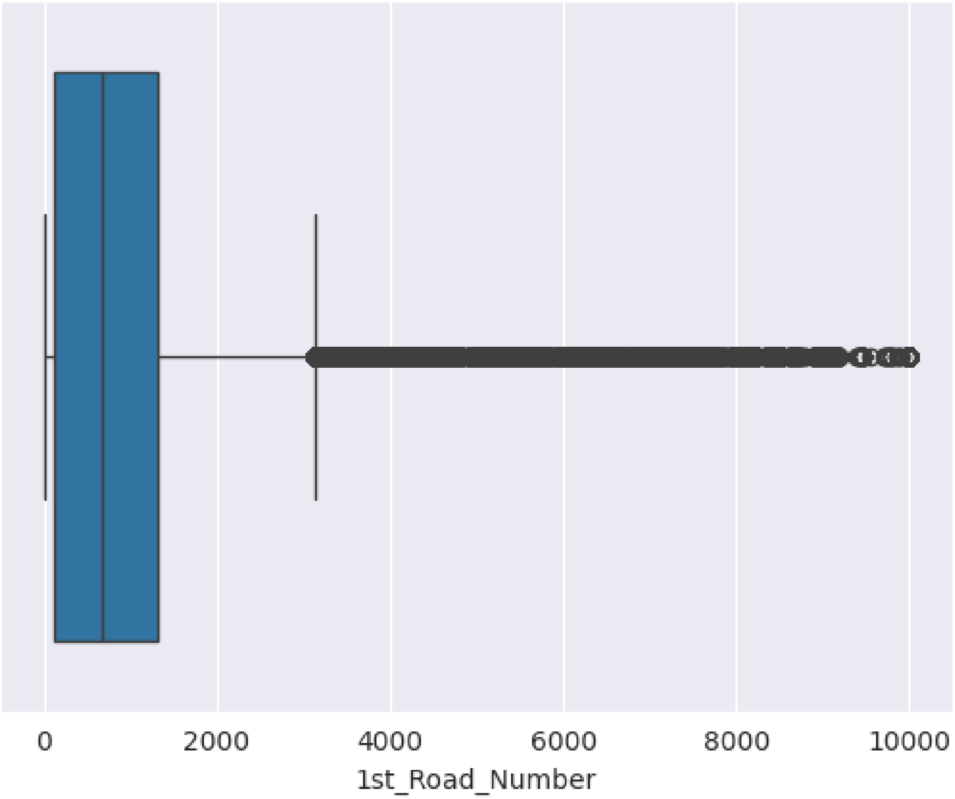

0
20
40
60
80
Age_of_Vehicle

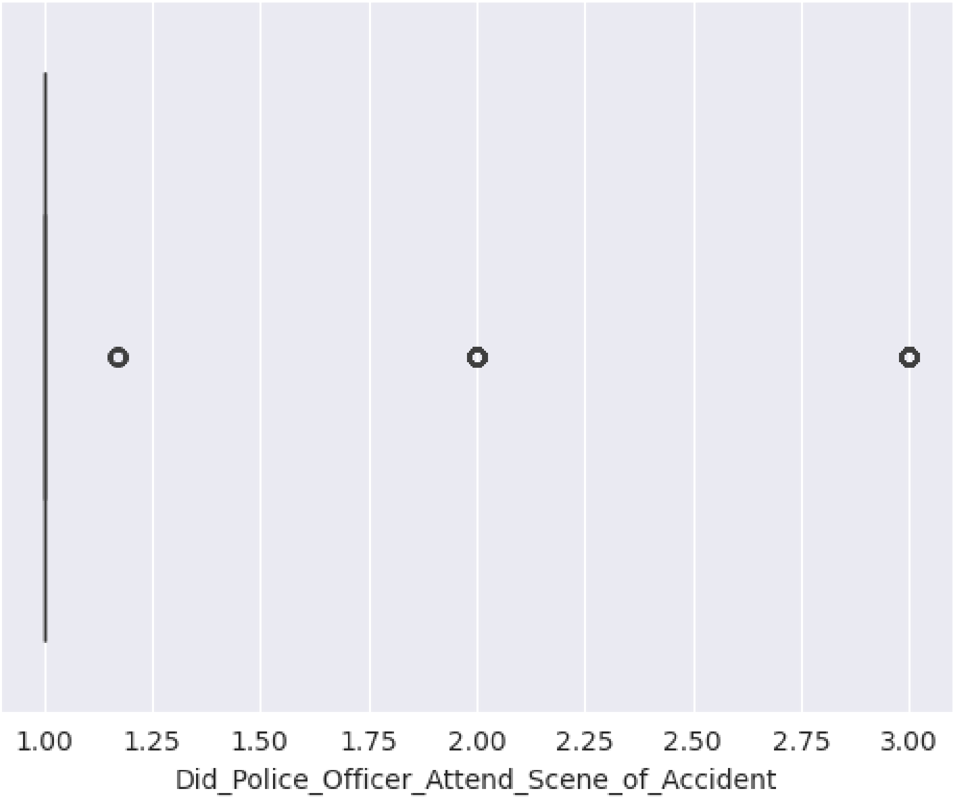

2
4
6
8
10
Driver_IMD_Decile

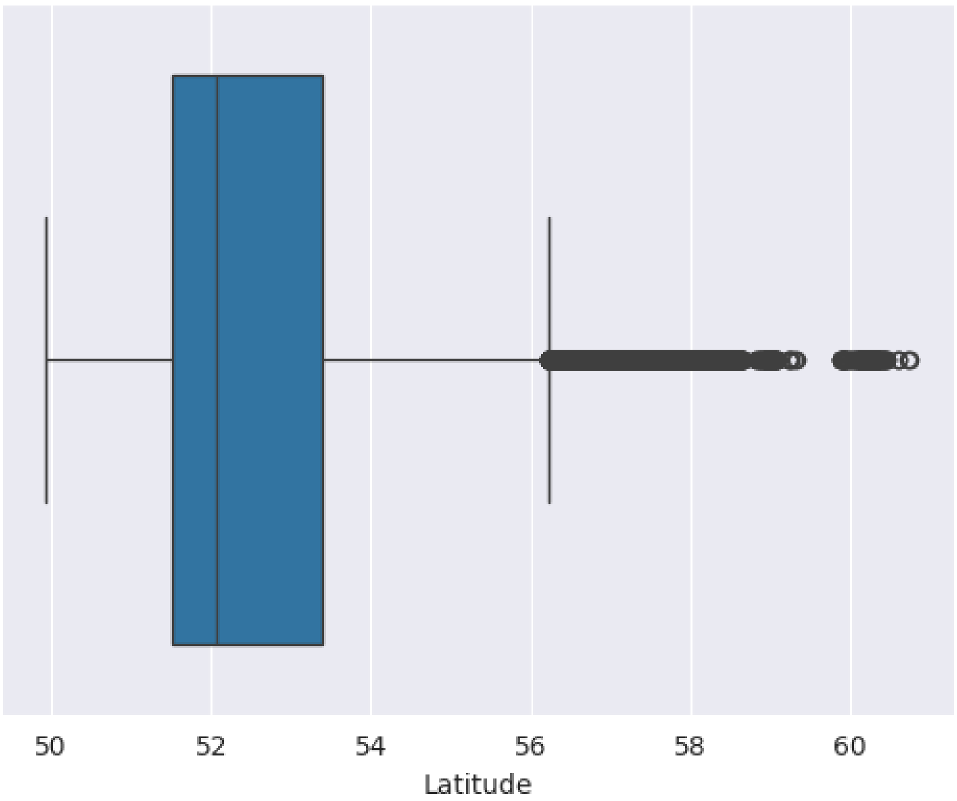

0
20000
40000
60000
80000
100000
Engine_Capacity_.CC.

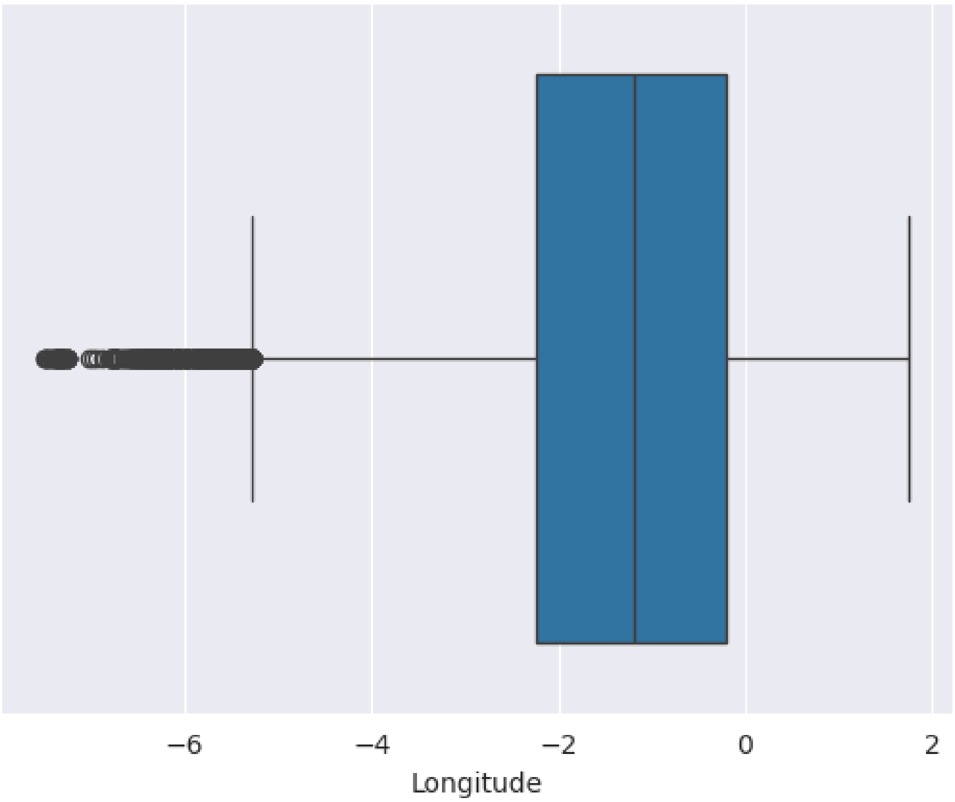

0
2
4
6
8
Vehicle_Location.Restricted_Lane

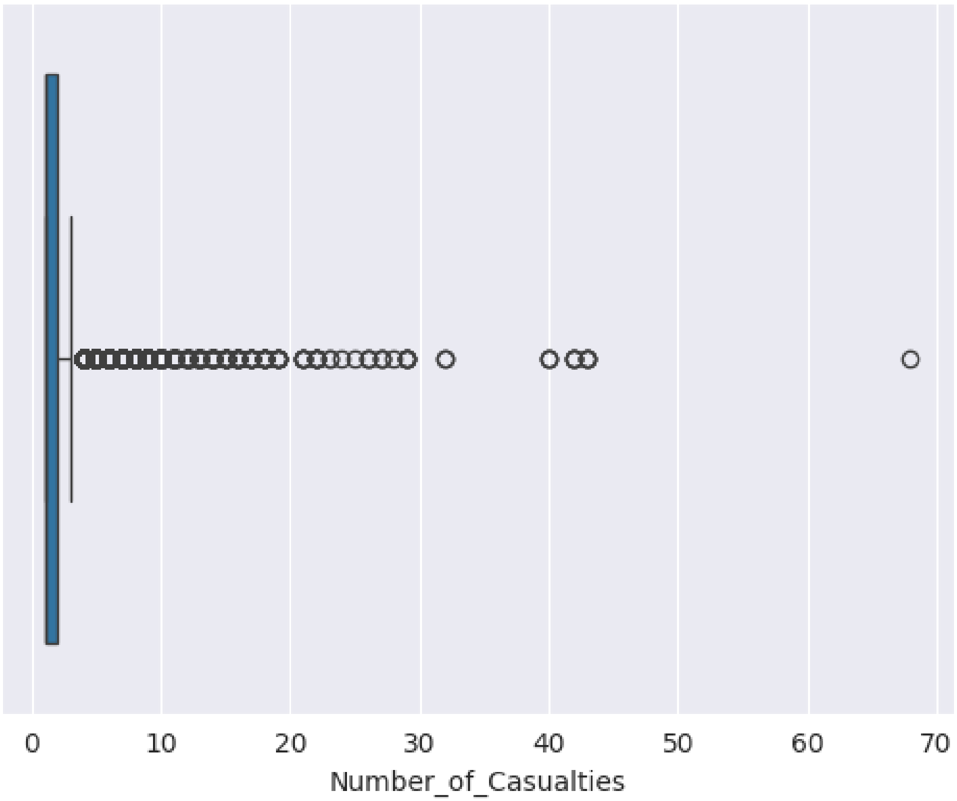

0
5
10
15
20
25
30
Vehicle_Reference

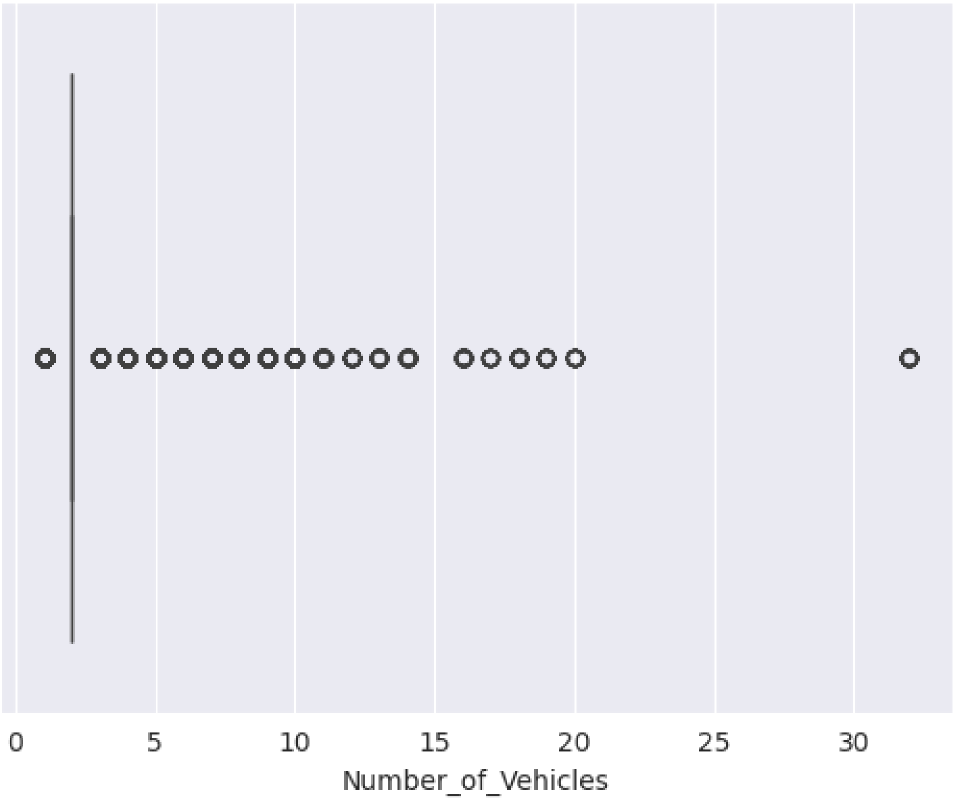

0
2000
4000
6000
8000
10000
1st_Road_Number

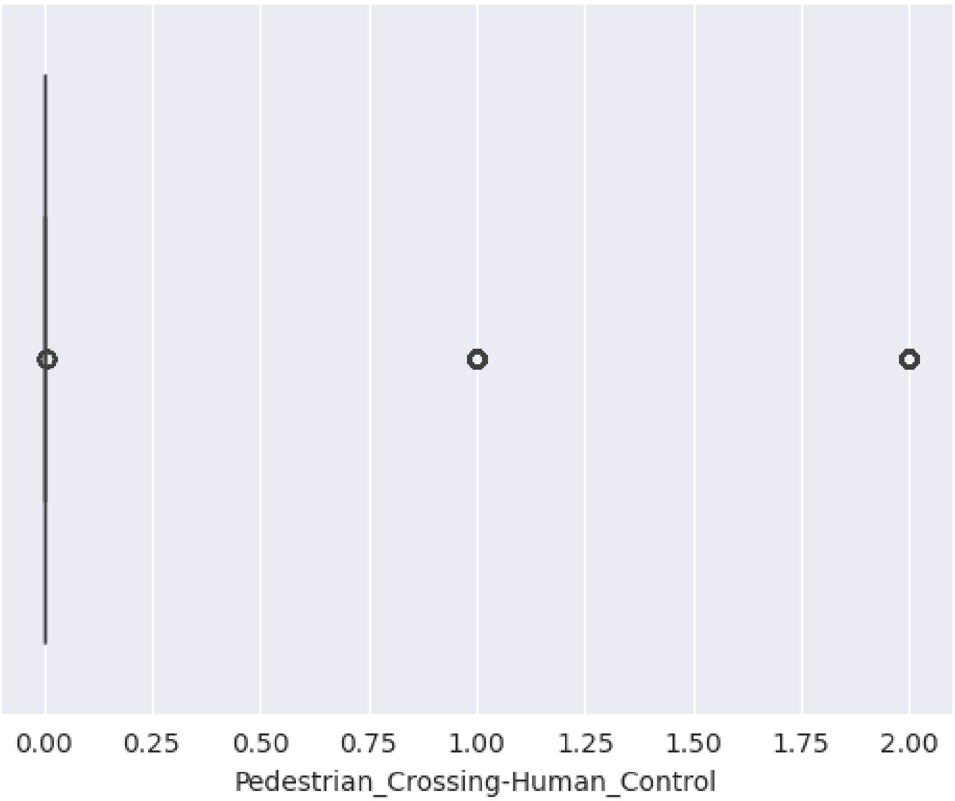

1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
Did_Police_Officer_Attend_Scene_of_Accident

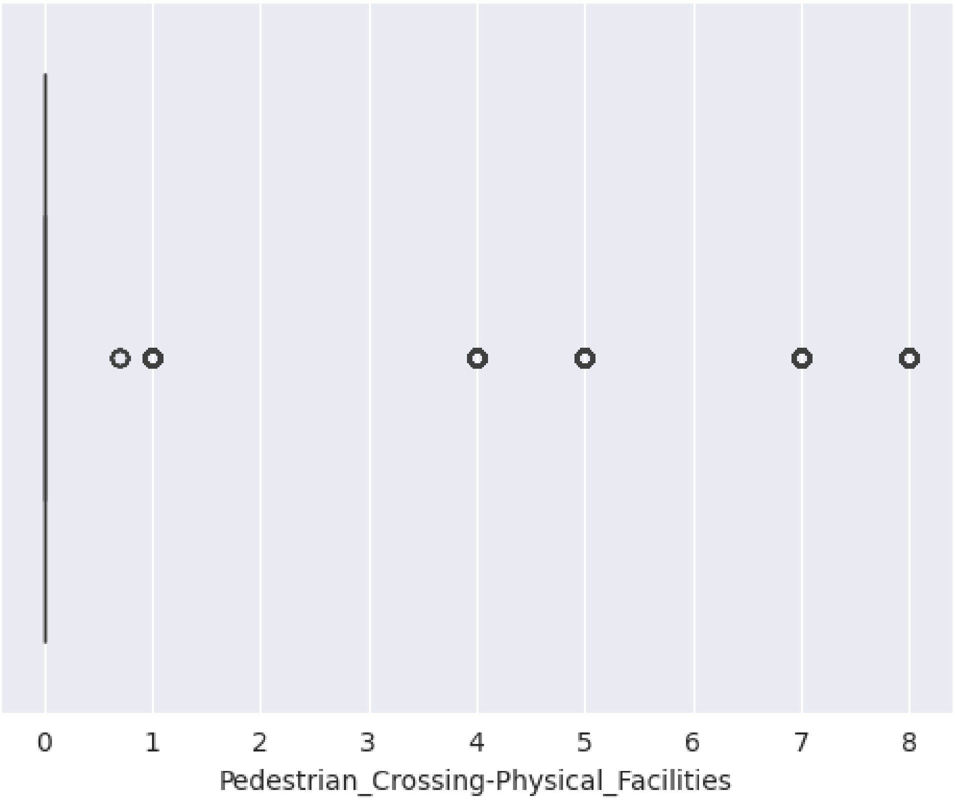

50
52
54
56
58
60
Latitude

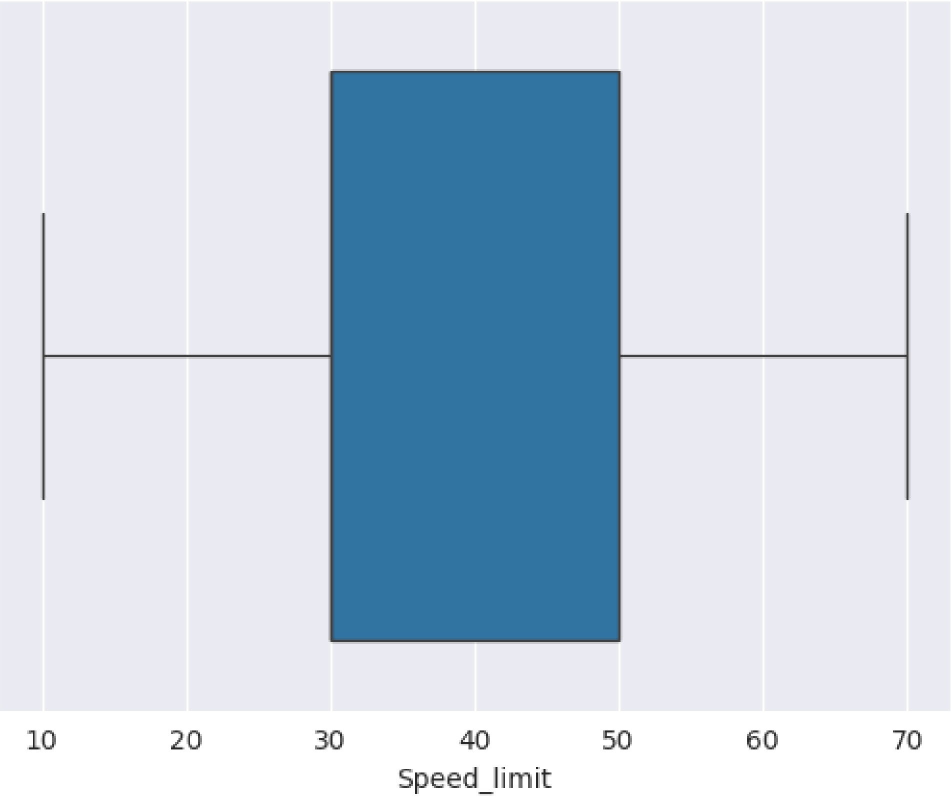

−6
−4
−2
0
2
Longitude


0
10
20
30
40
50
60
70
Number_of_Casualties


0
5
10
15
20
25
30
Number_of_Vehicles


0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
Pedestrian_Crossing-Human_Control


0
1
2
3
4
5
6
7
8
Pedestrian_Crossing-Physical_Facilities



```
[153]: lst=['Age_of_Vehicle','Vehicle_Reference','1st_Road_Number','Latitude','Longitude','Number_of_
         ↪CC.']
       for i in lst:
         print("Minimium value of ",i,"is",df[i].min())
         print("Mean value of ",i,"is",df[i].mean())
         print("Maximum value of ",i,"is",df[i].max())
         print("-"*75)
```

```
Minimium value of  Age_of_Vehicle is 1.0
Mean value of  Age_of_Vehicle is 6.333144690395298
Maximum value of  Age_of_Vehicle is 85.0
---------------------------------------------------------------------------
Minimium value of  Vehicle_Reference is 1
Mean value of  Vehicle_Reference is 1.5508228680666623
Maximum value of  Vehicle_Reference is 32
---------------------------------------------------------------------------
Minimium value of  1st_Road_Number is 1.0
Mean value of  1st_Road_Number is 1328.9576109506518
Maximum value of  1st_Road_Number is 9999.0
---------------------------------------------------------------------------
Minimium value of  Latitude is 49.914488
```

```
Mean value of  Latitude is 52.43148713785794
Maximum value of  Latitude is 60.757544
--------------------------------------------------------------------------
Minimium value of  Longitude is -7.516225
Mean value of  Longitude is -1.380407690620534
Maximum value of  Longitude is 1.758337
--------------------------------------------------------------------------
Minimium value of  Number_of_Casualties is 1
Mean value of  Number_of_Casualties is 1.4534414717934643
Maximum value of  Number_of_Casualties is 68
--------------------------------------------------------------------------
Minimium value of  Engine_Capacity_.CC. is 1.0
Mean value of  Engine_Capacity_.CC. is 2104.3481881760626
Maximum value of  Engine_Capacity_.CC. is 96000.0
--------------------------------------------------------------------------
```

[154]:

```python
df.shape
```

[154]:

```
(528335, 44)
```

[160]:

```python
def iqr_capping(df,variables):
    q1=df[variables].quantile(0.25)
    q3=df[variables].quantile(0.75)
    iqr=q3-q1
    upper_whisker=q3+(1.5*iqr)
    lower_whisker=q1-(1.5*iqr)
    return lower_whisker,upper_whisker
```

[161]:

```python
lower_lim,upper_lim=iqr_capping(df,'Age_of_Vehicle')
print('lower lim = ',lower_lim)
print('upper lim = ',upper_lim)
df['Age_of_Vehicle']=np.where(df['Age_of_Vehicle']>upper_lim,upper_lim,
                    np.
 ↪where(df['Age_of_Vehicle']<lower_lim,lower_lim,df['Age_of_Vehicle']))
```

```
lower lim =  -4.5
upper lim =  15.5
```

[162]:

```python
lower_lim,upper_lim=iqr_capping(df,'Engine_Capacity_.CC.')
print('lower lim = ',lower_lim)
print('upper lim = ',upper_lim)
df['Engine_Capacity_.CC.']=np.where(df['Engine_Capacity_.CC.
 ↪']>upper_lim,upper_lim,
                    np.where(df['Engine_Capacity_.CC.
 ↪']<lower_lim,lower_lim,df['Engine_Capacity_.CC.']))
```

```
lower lim =  404.5
upper lim =  2952.5
```

[163]:

```
lower_lim,upper_lim=iqr_capping(df,'Vehicle_Reference')
print('lower lim = ',lower_lim)
print('upper lim = ',upper_lim)
df['Vehicle_Reference']=np.where(df['Vehicle_Reference']>upper_lim,upper_lim,
                    np.
  ↪where(df['Vehicle_Reference']<lower_lim,lower_lim,df['Vehicle_Reference']))
```

```
lower lim =  -0.5
upper lim =  3.5
```

[164]:

```
lower_lim,upper_lim=iqr_capping(df,'Latitude')
print('lower lim = ',lower_lim)
print('upper lim = ',upper_lim)
df['Latitude']=np.where(df['Latitude']>upper_lim,upper_lim,
                    np.where(df['Latitude']<lower_lim,lower_lim,df['Latitude']))
```

```
lower lim =  48.68670175000001
upper lim =  56.22373175
```

[165]:

```
lower_lim,upper_lim=iqr_capping(df,'1st_Road_Number')
print('lower lim = ',lower_lim)
print('upper lim = ',upper_lim)
df['1st_Road_Number']=np.where(df['1st_Road_Number']>upper_lim,upper_lim,
                    np.
  ↪where(df['1st_Road_Number']<lower_lim,lower_lim,df['1st_Road_Number']))
```

```
lower lim =  -1678.4364164259778
upper lim =  3133.3940273766293
```

[166]:

```
lower_lim,upper_lim=iqr_capping(df,'Longitude')
print('lower lim = ',lower_lim)
print('upper lim = ',upper_lim)
df['Longitude']=np.where(df['Longitude']>upper_lim,upper_lim,
                    np.
  ↪where(df['Longitude']<lower_lim,lower_lim,df['Longitude']))
```

```
lower lim =  -5.2855834999999995
upper lim =  2.8453645
```

[167]:

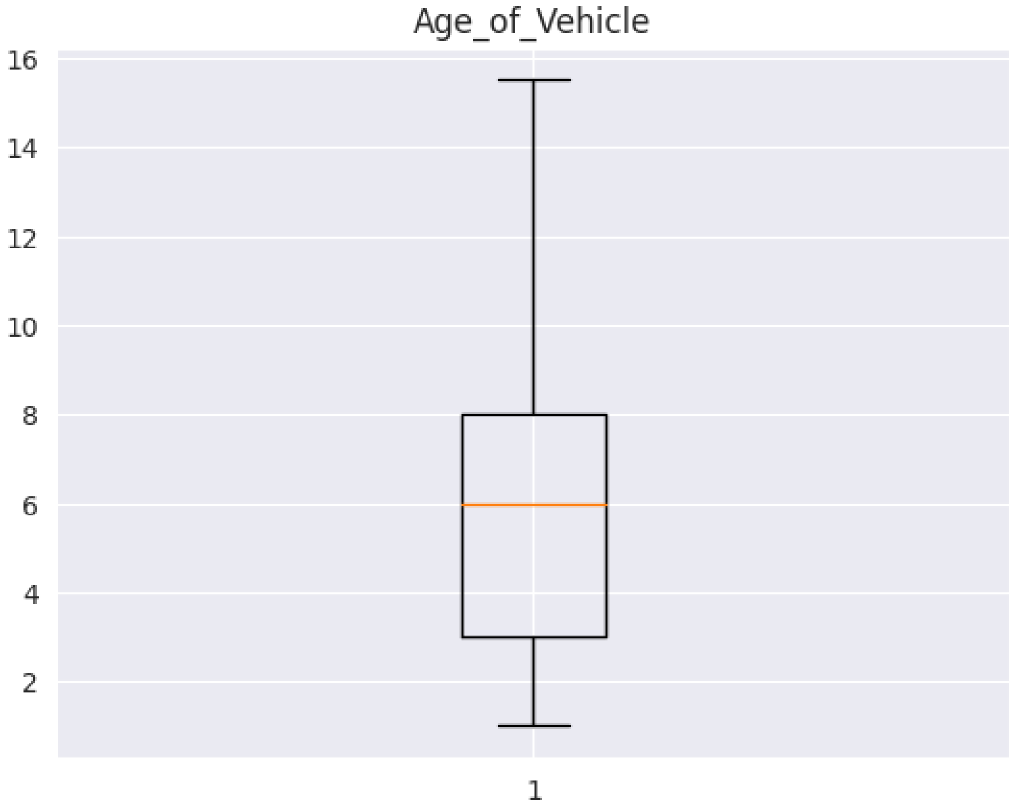
```
lower_lim,upper_lim=iqr_capping(df,'Number_of_Casualties')
print('lower lim = ',lower_lim)
print('upper lim = ',upper_lim)
df['Number_of_Casualties']=np.
  ↪where(df['Number_of_Casualties']>upper_lim,upper_lim,
                    np.
  ↪where(df['Number_of_Casualties']<lower_lim,lower_lim,df['Number_of_Casualties']))
```

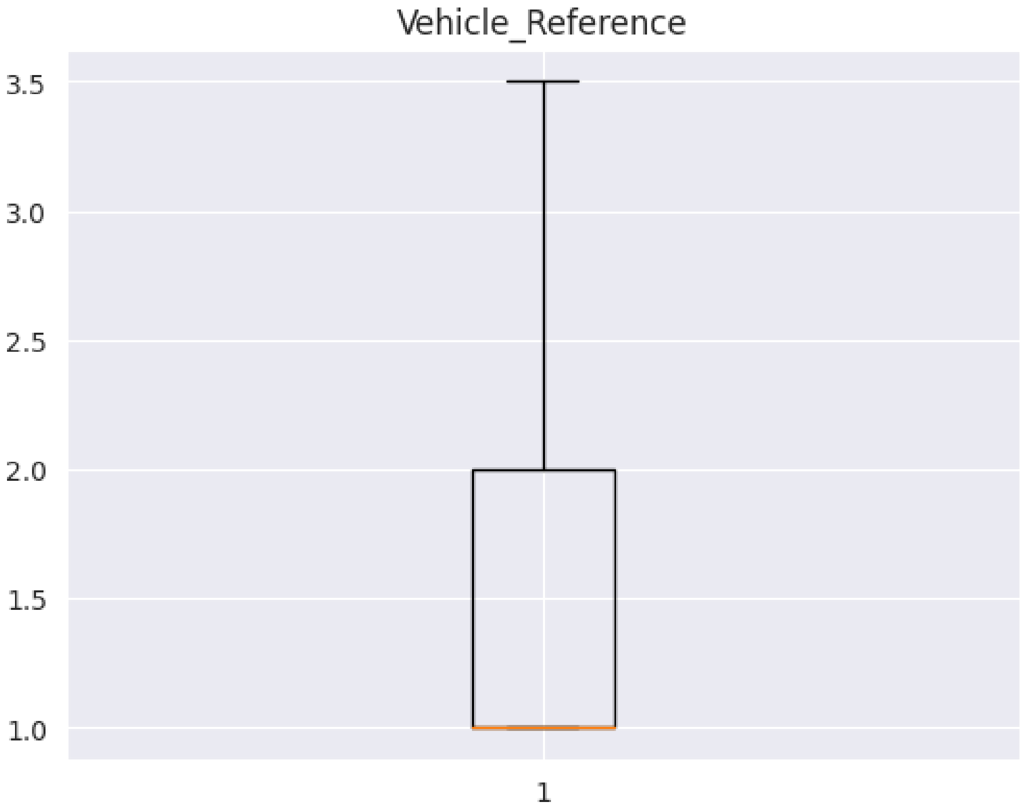
```
lower lim =  -0.5
```

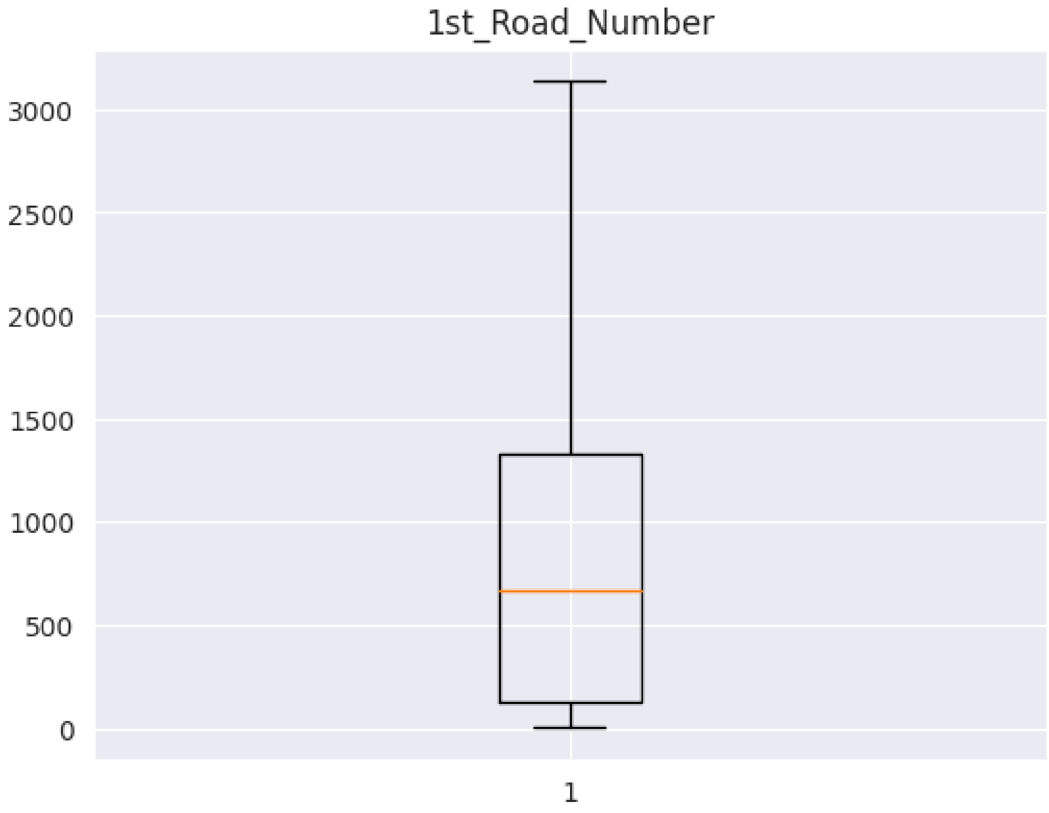
```
upper lim =  3.5
```

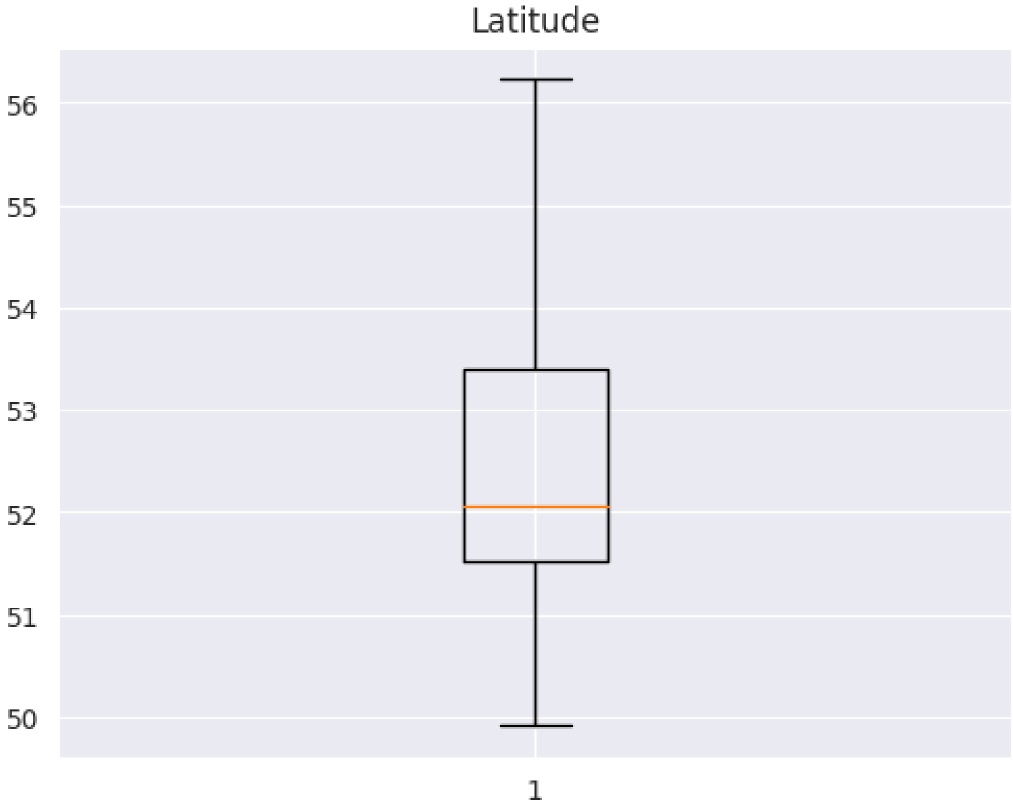
[168]:

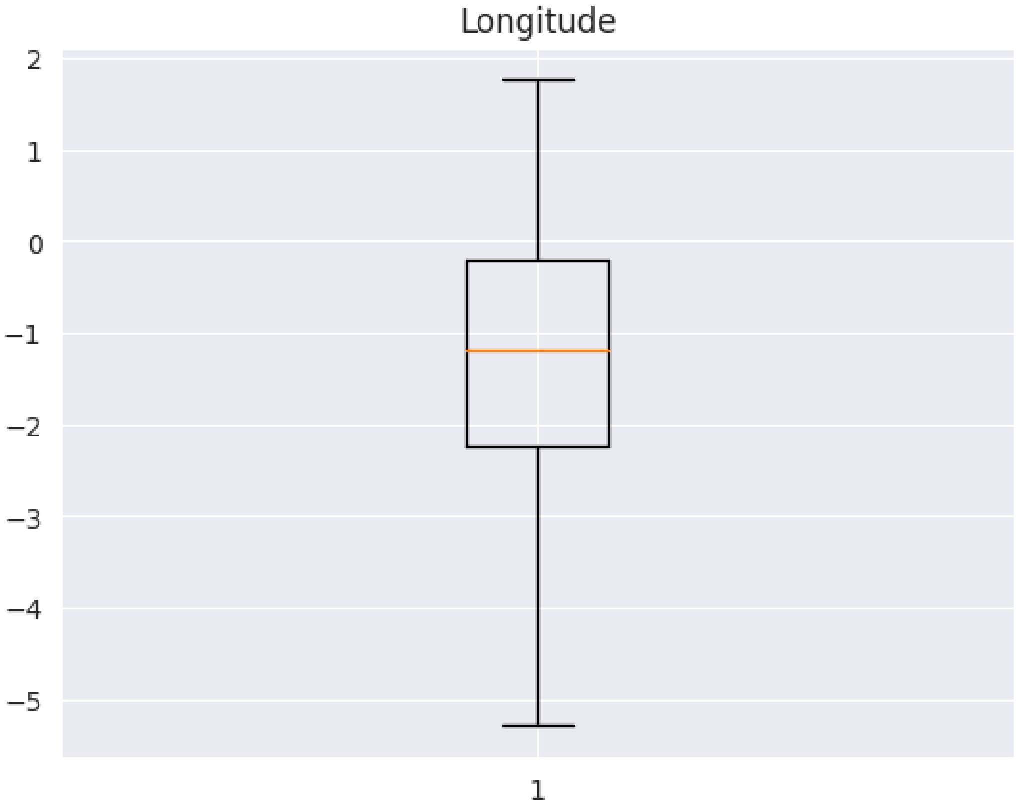
```
for i in␣
↪['Age_of_Vehicle','Vehicle_Reference','1st_Road_Number','Latitude','Longitude','Number_of_C
↪CC.']:
          plt.figure()
          plt.boxplot(df[i])
          plt.title(i)
```

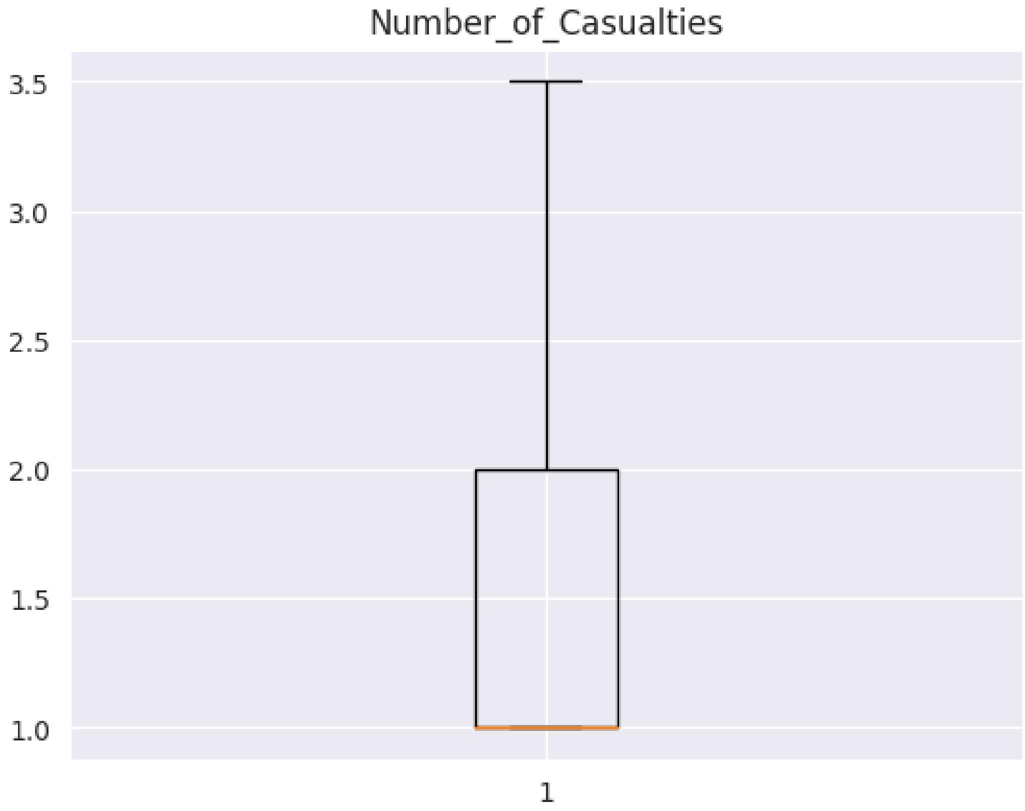


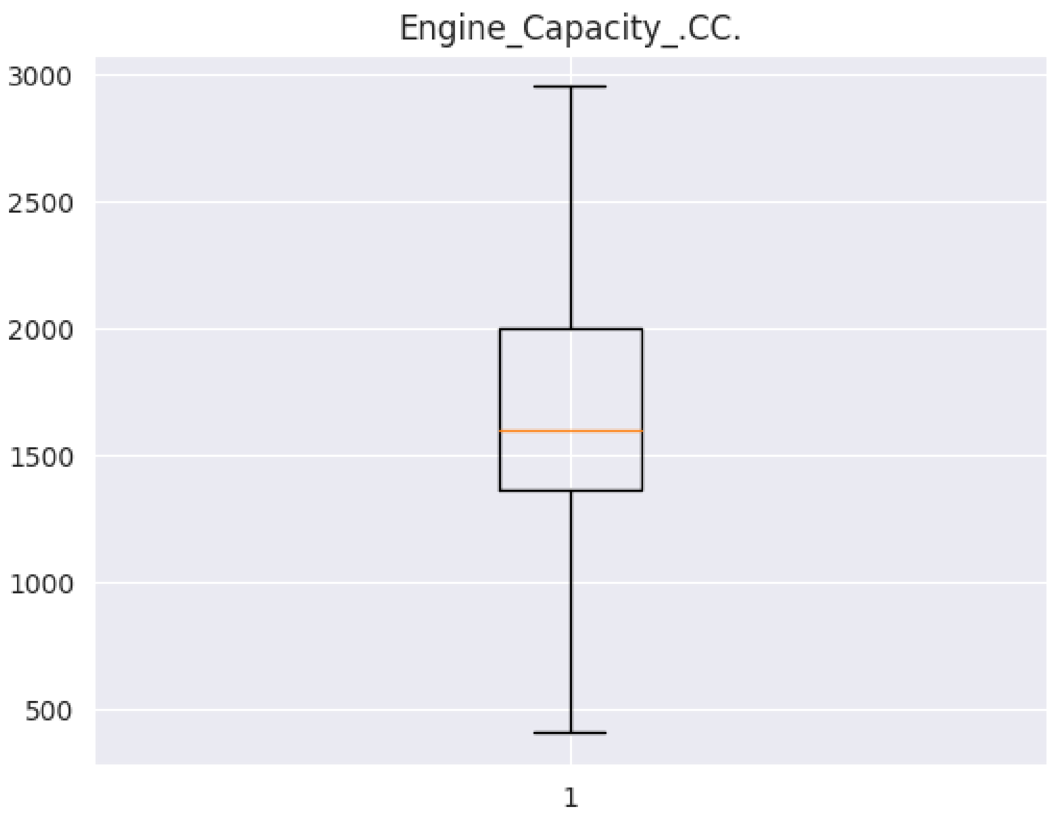

Vehicle_Reference
3.5
3.0
2.5
2.0
1.5
1.0
1


1st_Road_Number
3000
2500
2000
1500
1000
500
0
1


Latitude
56
55
54
53
52
51
50
1


Longitude
2
1
0
−1
−2
−3
−4
−5
1


Number_of_Casualties
3.5
3.0
2.5
2.0
1.5
1.0
1



[169]:
```
df.dtypes
```

[169]:
```
Age_Band_of_Driver                                object
Age_of_Vehicle                                   float64
Driver_Home_Area_Type                             object
Driver_IMD_Decile                                float64
Engine_Capacity_.CC.                             float64
Journey_Purpose_of_Driver                         object
Junction_Location                                 object
make                                              object
model                                             object
Propulsion_Code                                   object
Sex_of_Driver                                     object
Towing_and_Articulation                           object
Vehicle_Leaving_Carriageway                       object
Vehicle_Location.Restricted_Lane                 float64
Vehicle_Manoeuvre                                 object
Vehicle_Reference                                float64
Vehicle_Type                                      object
Was_Vehicle_Left_Hand_Drive                       object
X1st_Point_of_Impact                              object
```

```
1st_Road_Class                                  object
1st_Road_Number                                float64
2nd_Road_Class                                  object
Accident_Severity                               object
Day_of_Week                                     object
Did_Police_Officer_Attend_Scene_of_Accident    float64
Junction_Control                                object
Junction_Detail                                 object
Latitude                                       float64
Light_Conditions                                object
Local_Authority_(District)                      object
Local_Authority_(Highway)                       object
Longitude                                      float64
LSOA_of_Accident_Location                       object
Number_of_Casualties                           float64
Number_of_Vehicles                               int64
Pedestrian_Crossing-Human_Control              float64
Pedestrian_Crossing-Physical_Facilities        float64
Police_Force                                    object
Road_Surface_Conditions                         object
Road_Type                                       object
Speed_limit                                      int64
Urban_or_Rural_Area                             object
Weather_Conditions                              object
InScotland                                      object
dtype: object
```

[170]:

```
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
df['Age_Band_of_Driver']=le.fit_transform(df['Age_Band_of_Driver'])
df['Driver_Home_Area_Type']=le.fit_transform(df['Driver_Home_Area_Type'])
df['Journey_Purpose_of_Driver']=le.
 ↪fit_transform(df['Journey_Purpose_of_Driver'])
df['Junction_Location']=le.fit_transform(df['Junction_Location'])
df['make']=le.fit_transform(df['make'])
df['model']=le.fit_transform(df['model'])
df['Propulsion_Code']=le.fit_transform(df['Propulsion_Code'])
df['Sex_of_Driver']=le.fit_transform(df['Sex_of_Driver'])
df['Towing_and_Articulation']=le.fit_transform(df['Towing_and_Articulation'])
df['Vehicle_Leaving_Carriageway']=le.
 ↪fit_transform(df['Vehicle_Leaving_Carriageway'])
df['Vehicle_Manoeuvre']=le.fit_transform(df['Vehicle_Manoeuvre'])
df['Vehicle_Type']=le.fit_transform(df['Vehicle_Type'])
df['Was_Vehicle_Left_Hand_Drive']=le.
 ↪fit_transform(df['Was_Vehicle_Left_Hand_Drive'])
df['X1st_Point_of_Impact']=le.fit_transform(df['X1st_Point_of_Impact'])
df['1st_Road_Class']=le.fit_transform(df['1st_Road_Class'])
```

```
df['2nd_Road_Class']=le.fit_transform(df['2nd_Road_Class'])
df['Accident_Severity']=le.fit_transform(df['Accident_Severity'])
df['Day_of_Week']=le.fit_transform(df['Day_of_Week'])
df['Junction_Control']=le.fit_transform(df['Junction_Control'])
df['Junction_Detail']=le.fit_transform(df['Junction_Detail'])
df['Light_Conditions']=le.fit_transform(df['Light_Conditions'])
df['Local_Authority_(District)']=le.
 ↪fit_transform(df['Local_Authority_(District)'])
df['Local_Authority_(Highway)']=le.
 ↪fit_transform(df['Local_Authority_(Highway)'])
df['Police_Force']=le.fit_transform(df['Police_Force'])
df['Road_Surface_Conditions']=le.fit_transform(df['Road_Surface_Conditions'])
df['Road_Type']=le.fit_transform(df['Road_Type'])
df['Speed_limit']=le.fit_transform(df['Speed_limit'])
df['Urban_or_Rural_Area']=le.fit_transform(df['Urban_or_Rural_Area'])
df['Weather_Conditions']=le.fit_transform(df['Weather_Conditions'])
df['InScotland']=le.fit_transform(df['InScotland'])
df['LSOA_of_Accident_Location']=le.
 ↪fit_transform(df['LSOA_of_Accident_Location'])
df.dtypes
```

```
[170]: Age_Band_of_Driver                               int64
       Age_of_Vehicle                                 float64
       Driver_Home_Area_Type                            int64
       Driver_IMD_Decile                              float64
       Engine_Capacity_.CC.                           float64
       Journey_Purpose_of_Driver                        int64
       Junction_Location                                int64
       make                                             int64
       model                                            int64
       Propulsion_Code                                  int64
       Sex_of_Driver                                    int64
       Towing_and_Articulation                          int64
       Vehicle_Leaving_Carriageway                      int64
       Vehicle_Location.Restricted_Lane               float64
       Vehicle_Manoeuvre                                int64
       Vehicle_Reference                              float64
       Vehicle_Type                                     int64
       Was_Vehicle_Left_Hand_Drive                      int64
       X1st_Point_of_Impact                             int64
       1st_Road_Class                                   int64
       1st_Road_Number                                float64
       2nd_Road_Class                                   int64
       Accident_Severity                                int64
       Day_of_Week                                      int64
       Did_Police_Officer_Attend_Scene_of_Accident    float64
       Junction_Control                                 int64
```

```
Junction_Detail                                  int64
Latitude                                       float64
Light_Conditions                                 int64
Local_Authority_(District)                       int64
Local_Authority_(Highway)                        int64
Longitude                                      float64
LSOA_of_Accident_Location                        int64
Number_of_Casualties                           float64
Number_of_Vehicles                               int64
Pedestrian_Crossing-Human_Control              float64
Pedestrian_Crossing-Physical_Facilities        float64
Police_Force                                     int64
Road_Surface_Conditions                          int64
Road_Type                                        int64
Speed_limit                                      int64
Urban_or_Rural_Area                              int64
Weather_Conditions                               int64
InScotland                                       int64
dtype: object
```

[171]:

```
x=df.drop(['Accident_Severity'],axis=1).values
x
```

[171]:

```
array([[5., 3., 2., …, 2., 1., 0.],
       [4., 5., 2., …, 2., 1., 0.],
       [6., 4., 2., …, 2., 1., 0.],
       …,
       [5., 8., 2., …, 0., 5., 0.],
       [7., 8., 2., …, 2., 5., 0.],
       [2., 7., 1., …, 2., 5., 0.]])
```

[172]:

```
y=df['Accident_Severity'].values
y
```

[172]:

```
array([2, 2, 2, …, 2, 2, 2])
```

[173]:

```
from imblearn.over_sampling import SMOTE
oversample=SMOTE()
x1,y1=oversample.fit_resample(x,y)
```

[174]:

```
#checking the oversampling output
y2=pd.DataFrame(y1)
y2.value_counts()
```

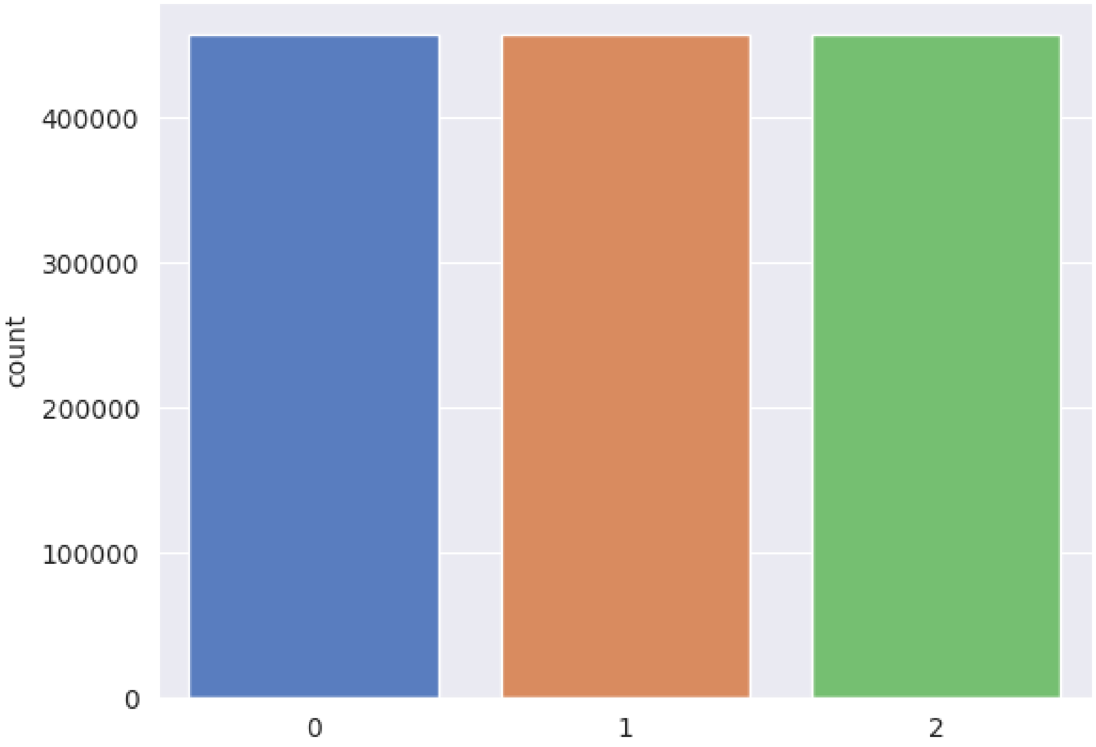
[174]:

```
0    457171
1    457171
2    457171
```

```
Name: count, dtype: int64
```

```
[175]: sns.countplot(x = y1, palette='muted')
```

```
[175]: <Axes: ylabel='count'>
```



```
[176]: from sklearn.model_selection import train_test_split
       x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.
        ↪30,random_state=1)
```

```
[177]: x_train
```

```
[177]: array([[ 2.,  8.,  2., …,  2.,  5.,  0.],
              [ 5.,  2.,  2., …,  2.,  1.,  0.],
              [ 5., 13.,  2., …,  2.,  1.,  1.],
              …,
              [ 3., 10.,  2., …,  2.,  1.,  0.],
              [ 4., 12.,  2., …,  2.,  5.,  0.],
              [ 4.,  1.,  2., …,  0.,  0.,  0.]])
```

```
[178]: x_test
```

```
[178]: array([[ 2. , 15.5,  2. , …,  2. ,  1. ,  0. ],
              [ 4. ,  7. ,  1. , …,  0. ,  1. ,  0. ],
              [ 2. , 15.5,  2. , …,  2. ,  1. ,  0. ],
              …,
              [ 4. ,  9. ,  2. , …,  2. ,  1. ,  0. ],
              [ 6. ,  7. ,  2. , …,  0. ,  1. ,  0. ],
              [ 5. ,  6. ,  2. , …,  2. ,  1. ,  0. ]])
```

```
[179]: y_train
```

```
[179]: array([2, 2, 2, …, 2, 2, 2])
```

```
[180]: y_test
```

```
[180]: array([2, 2, 2, …, 2, 2, 2])
```

```
[181]: from sklearn.preprocessing import StandardScaler
       scaler=StandardScaler()
       scaler.fit(x_train)
       x_train=scaler.transform(x_train)
       x_test=scaler.transform(x_test)
```

```
[182]: pip install catboost
```

```
Collecting catboost
  Downloading catboost-1.2.5-cp310-cp310-manylinux2014_x86_64.whl (98.2 MB)
                                 98.2/98.2 MB
6.0 MB/s eta 0:00:00
Requirement already satisfied: graphviz in /usr/local/lib/python3.10/dist-
packages (from catboost) (0.20.3)
Requirement already satisfied: matplotlib in /usr/local/lib/python3.10/dist-
packages (from catboost) (3.7.1)
Requirement already satisfied: numpy>=1.16.0 in /usr/local/lib/python3.10/dist-
packages (from catboost) (1.25.2)
Requirement already satisfied: pandas>=0.24 in /usr/local/lib/python3.10/dist-
packages (from catboost) (2.0.3)
Requirement already satisfied: scipy in /usr/local/lib/python3.10/dist-packages
(from catboost) (1.11.4)
Requirement already satisfied: plotly in /usr/local/lib/python3.10/dist-packages
(from catboost) (5.15.0)
Requirement already satisfied: six in /usr/local/lib/python3.10/dist-packages
(from catboost) (1.16.0)
Requirement already satisfied: python-dateutil>=2.8.2 in
/usr/local/lib/python3.10/dist-packages (from pandas>=0.24->catboost) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-
packages (from pandas>=0.24->catboost) (2023.4)
Requirement already satisfied: tzdata>=2022.1 in /usr/local/lib/python3.10/dist-
```

```
packages (from pandas>=0.24->catboost) (2024.1)
Requirement already satisfied: contourpy>=1.0.1 in
/usr/local/lib/python3.10/dist-packages (from matplotlib->catboost) (1.2.1)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.10/dist-
packages (from matplotlib->catboost) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in
/usr/local/lib/python3.10/dist-packages (from matplotlib->catboost) (4.51.0)
Requirement already satisfied: kiwisolver>=1.0.1 in
/usr/local/lib/python3.10/dist-packages (from matplotlib->catboost) (1.4.5)
Requirement already satisfied: packaging>=20.0 in
/usr/local/lib/python3.10/dist-packages (from matplotlib->catboost) (24.0)
Requirement already satisfied: pillow>=6.2.0 in /usr/local/lib/python3.10/dist-
packages (from matplotlib->catboost) (9.4.0)
Requirement already satisfied: pyparsing>=2.3.1 in
/usr/local/lib/python3.10/dist-packages (from matplotlib->catboost) (3.1.2)
Requirement already satisfied: tenacity>=6.2.0 in
/usr/local/lib/python3.10/dist-packages (from plotly->catboost) (8.3.0)
Installing collected packages: catboost
Successfully installed catboost-1.2.5
```

[183]:

```python
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.neighbors import KNeighborsClassifier
from sklearn.naive_bayes import BernoulliNB
from xgboost import XGBClassifier
import catboost
import lightgbm as lgb
from sklearn.metrics import classification_report,accuracy_score
from sklearn.metrics import confusion_matrix,ConfusionMatrixDisplay
```

[184]:

```python
dec = DecisionTreeClassifier(criterion="entropy",random_state=42)
rfc = RandomForestClassifier(n_estimators= 30,random_state=42)
knn=KNeighborsClassifier(n_neighbors=7)
naive=BernoulliNB()
xgb = XGBClassifier()
clf = catboost.CatBoostClassifier()
lgb_c = lgb.LGBMClassifier(force_row_wise=True)
lst_model = [dec,rfc,xgb,clf,lgb_c,knn,naive]
```

[185]:

```python
for i in lst_model:
    print(i)
    print("-"*75)
    i.fit(x_train,y_train)
    y_pred=i.predict(x_test)
    print(classification_report(y_test,y_pred))
    print("Accuracy score of",i," ",accuracy_score(y_test,y_pred))
    labels=[0,1,2]
    result=confusion_matrix(y_test,y_pred)
    cmd=ConfusionMatrixDisplay(result,display_labels=labels)
    cmd.plot()
```

```
DecisionTreeClassifier(criterion='entropy', random_state=42)
---------------------------------------------------------------------------
              precision    recall  f1-score   support

           0       0.10      0.12      0.11      2145
           1       0.21      0.23      0.22     19089
           2       0.89      0.87      0.88    137267

    accuracy                           0.78    158501
   macro avg       0.40      0.41      0.40    158501
weighted avg       0.79      0.78      0.79    158501

Accuracy score of DecisionTreeClassifier(criterion='entropy', random_state=42)
0.7836669800190535
RandomForestClassifier(n_estimators=30, random_state=42)
---------------------------------------------------------------------------
              precision    recall  f1-score   support

           0       0.80      0.02      0.03      2145
           1       0.47      0.05      0.09     19089
           2       0.87      0.99      0.93    137267

    accuracy                           0.87    158501
   macro avg       0.71      0.35      0.35    158501
weighted avg       0.82      0.87      0.82    158501

Accuracy score of RandomForestClassifier(n_estimators=30, random_state=42)
0.8665118832057842
XGBClassifier(base_score=None, booster=None, callbacks=None,
              colsample_bylevel=None, colsample_bynode=None,
              colsample_bytree=None, device=None, early_stopping_rounds=None,
              enable_categorical=False, eval_metric=None, feature_types=None,
              gamma=None, grow_policy=None, importance_type=None,
              interaction_constraints=None, learning_rate=None, max_bin=None,
              max_cat_threshold=None, max_cat_to_onehot=None,
              max_delta_step=None, max_depth=None, max_leaves=None,
              min_child_weight=None, missing=nan, monotone_constraints=None,
              multi_strategy=None, n_estimators=None, n_jobs=None,
              num_parallel_tree=None, random_state=None, …)
---------------------------------------------------------------------------
              precision    recall  f1-score   support

           0       0.48      0.03      0.05      2145
           1       0.50      0.05      0.10     19089
           2       0.87      0.99      0.93    137267

    accuracy                           0.87    158501
   macro avg       0.62      0.36      0.36    158501
weighted avg       0.82      0.87      0.82    158501

Accuracy score of XGBClassifier(base_score=None, booster=None, callbacks=None,
              colsample_bylevel=None, colsample_bynode=None,
              colsample_bytree=None, device=None, early_stopping_rounds=None,
              enable_categorical=False, eval_metric=None, feature_types=None,
              gamma=None, grow_policy=None, importance_type=None,
              interaction_constraints=None, learning_rate=None, max_bin=None,
              max_cat_threshold=None, max_cat_to_onehot=None,
              max_delta_step=None, max_depth=None, max_leaves=None,
              min_child_weight=None, missing=nan, monotone_constraints=None,
              multi_strategy=None, n_estimators=None, n_jobs=None,
              num_parallel_tree=None, objective='multi:softprob', …)
0.867931432609258
<catboost.core.CatBoostClassifier object at 0x7901055cf880>
---------------------------------------------------------------------------
Learning rate set to 0.106977
0:      learn: 0.9660651        total: 369ms    remaining: 6m 8s
1:      learn: 0.8680253        total: 632ms    remaining: 5m 15s
2:      learn: 0.7916029        total: 941ms    remaining: 5m 12s
3:      learn: 0.7302687        total: 1.2s     remaining: 4m 58s
4:      learn: 0.6814309        total: 1.49s    remaining: 4m 57s
5:      learn: 0.6406897        total: 1.78s    remaining: 4m 54s
6:      learn: 0.6066317        total: 2.07s    remaining: 4m 54s
7:      learn: 0.5781040        total: 2.31s    remaining: 4m 46s
8:      learn: 0.5538412        total: 2.56s    remaining: 4m 42s
9:      learn: 0.5331905        total: 2.84s    remaining: 4m 41s
10:     learn: 0.5159336        total: 3.13s    remaining: 4m 41s
11:     learn: 0.5006074        total: 3.42s    remaining: 4m 41s
12:     learn: 0.4875244        total: 3.71s    remaining: 4m 41s
13:     learn: 0.4765281        total: 4s       remaining: 4m 42s
14:     learn: 0.4665462        total: 4.24s    remaining: 4m 38s
15:     learn: 0.4579670        total: 4.53s    remaining: 4m 38s
16:     learn: 0.4503100        total: 4.83s    remaining: 4m 39s
17:     learn: 0.4440554        total: 5.11s    remaining: 4m 38s
18:     learn: 0.4385606        total: 5.4s     remaining: 4m 38s
19:     learn: 0.4334943        total: 5.68s    remaining: 4m 38s
20:     learn: 0.4294343        total: 5.94s    remaining: 4m 36s
21:     learn: 0.4255730        total: 6.24s    remaining: 4m 37s
22:     learn: 0.4222851        total: 6.52s    remaining: 4m 36s
23:     learn: 0.4194118        total: 6.79s    remaining: 4m 36s
24:     learn: 0.4168109        total: 7.12s    remaining: 4m 37s
25:     learn: 0.4144481        total: 7.39s    remaining: 4m 36s
```

```
26:     learn: 0.4126343        total: 7.67s    remaining: 4m 36s
27:     learn: 0.4107996        total: 7.94s    remaining: 4m 35s
28:     learn: 0.4092858        total: 8.25s    remaining: 4m 36s
29:     learn: 0.4079142        total: 8.52s    remaining: 4m 35s
30:     learn: 0.4065167        total: 8.77s    remaining: 4m 34s
31:     learn: 0.4052202        total: 9.18s    remaining: 4m 37s
32:     learn: 0.4041555        total: 9.64s    remaining: 4m 42s
33:     learn: 0.4031572        total: 10.1s    remaining: 4m 47s
34:     learn: 0.4024563        total: 10.6s    remaining: 4m 52s
35:     learn: 0.4017424        total: 11.1s    remaining: 4m 56s
36:     learn: 0.4009958        total: 11.6s    remaining: 5m 1s
37:     learn: 0.4002289        total: 12.2s    remaining: 5m 8s
38:     learn: 0.3997829        total: 12.7s    remaining: 5m 11s
39:     learn: 0.3992514        total: 13.1s    remaining: 5m 13s
40:     learn: 0.3986324        total: 13.6s    remaining: 5m 16s
41:     learn: 0.3982106        total: 14.1s    remaining: 5m 21s
42:     learn: 0.3977777        total: 14.4s    remaining: 5m 21s
43:     learn: 0.3974646        total: 14.7s    remaining: 5m 20s
44:     learn: 0.3969598        total: 15s      remaining: 5m 18s
45:     learn: 0.3964612        total: 15.3s    remaining: 5m 18s
46:     learn: 0.3961988        total: 15.6s    remaining: 5m 16s
47:     learn: 0.3959293        total: 15.9s    remaining: 5m 15s
48:     learn: 0.3955824        total: 16.3s    remaining: 5m 15s
49:     learn: 0.3952919        total: 16.6s    remaining: 5m 14s
50:     learn: 0.3950838        total: 16.8s    remaining: 5m 13s
51:     learn: 0.3948204        total: 17.1s    remaining: 5m 11s
52:     learn: 0.3946407        total: 17.4s    remaining: 5m 10s
53:     learn: 0.3943589        total: 17.7s    remaining: 5m 9s
54:     learn: 0.3939681        total: 18s      remaining: 5m 8s
55:     learn: 0.3937421        total: 18.3s    remaining: 5m 7s
56:     learn: 0.3936220        total: 18.5s    remaining: 5m 6s
57:     learn: 0.3934308        total: 18.8s    remaining: 5m 5s
58:     learn: 0.3932521        total: 19.1s    remaining: 5m 4s
59:     learn: 0.3930936        total: 19.4s    remaining: 5m 3s
60:     learn: 0.3928499        total: 19.7s    remaining: 5m 2s
61:     learn: 0.3926331        total: 20s      remaining: 5m 1s
62:     learn: 0.3925108        total: 20.2s    remaining: 5m
63:     learn: 0.3922801        total: 20.5s    remaining: 4m 59s
64:     learn: 0.3919571        total: 20.8s    remaining: 4m 59s
65:     learn: 0.3918183        total: 21s      remaining: 4m 57s
66:     learn: 0.3916291        total: 21.3s    remaining: 4m 56s
67:     learn: 0.3914737        total: 21.6s    remaining: 4m 56s
68:     learn: 0.3913905        total: 21.9s    remaining: 4m 54s
69:     learn: 0.3912189        total: 22.1s    remaining: 4m 54s
70:     learn: 0.3910059        total: 22.4s    remaining: 4m 53s
71:     learn: 0.3908623        total: 22.7s    remaining: 4m 52s
72:     learn: 0.3907115        total: 23s      remaining: 4m 51s
73:     learn: 0.3905916        total: 23.3s    remaining: 4m 50s
```

```
74:     learn: 0.3904796        total: 23.5s    remaining: 4m 49s
75:     learn: 0.3902515        total: 23.7s    remaining: 4m 48s
76:     learn: 0.3901395        total: 24s      remaining: 4m 48s
77:     learn: 0.3899334        total: 24.3s    remaining: 4m 47s
78:     learn: 0.3897814        total: 24.9s    remaining: 4m 50s
79:     learn: 0.3896259        total: 25.5s    remaining: 4m 52s
80:     learn: 0.3894705        total: 26.1s    remaining: 4m 55s
81:     learn: 0.3893844        total: 26.5s    remaining: 4m 56s
82:     learn: 0.3893090        total: 26.9s    remaining: 4m 57s
83:     learn: 0.3892331        total: 27.4s    remaining: 4m 58s
84:     learn: 0.3891710        total: 27.9s    remaining: 4m 59s
85:     learn: 0.3890583        total: 28.3s    remaining: 5m 1s
86:     learn: 0.3889349        total: 28.9s    remaining: 5m 2s
87:     learn: 0.3888496        total: 29.4s    remaining: 5m 4s
88:     learn: 0.3888096        total: 29.6s    remaining: 5m 3s
89:     learn: 0.3887481        total: 29.9s    remaining: 5m 2s
90:     learn: 0.3886412        total: 30.1s    remaining: 5m
91:     learn: 0.3885424        total: 30.4s    remaining: 5m
92:     learn: 0.3884903        total: 30.7s    remaining: 4m 59s
93:     learn: 0.3884067        total: 30.9s    remaining: 4m 58s
94:     learn: 0.3883520        total: 31.2s    remaining: 4m 57s
95:     learn: 0.3882747        total: 31.4s    remaining: 4m 55s
96:     learn: 0.3881826        total: 31.7s    remaining: 4m 55s
97:     learn: 0.3880769        total: 32s      remaining: 4m 54s
98:     learn: 0.3880006        total: 32.3s    remaining: 4m 53s
99:     learn: 0.3879597        total: 32.5s    remaining: 4m 52s
100:    learn: 0.3877504        total: 32.8s    remaining: 4m 52s
101:    learn: 0.3876649        total: 33.1s    remaining: 4m 51s
102:    learn: 0.3875939        total: 33.4s    remaining: 4m 50s
103:    learn: 0.3875234        total: 33.7s    remaining: 4m 50s
104:    learn: 0.3873597        total: 34s      remaining: 4m 49s
105:    learn: 0.3872344        total: 34.3s    remaining: 4m 49s
106:    learn: 0.3871443        total: 34.6s    remaining: 4m 48s
107:    learn: 0.3870823        total: 34.9s    remaining: 4m 47s
108:    learn: 0.3869871        total: 35.1s    remaining: 4m 47s
109:    learn: 0.3869114        total: 35.4s    remaining: 4m 46s
110:    learn: 0.3867936        total: 35.7s    remaining: 4m 46s
111:    learn: 0.3866476        total: 36s      remaining: 4m 45s
112:    learn: 0.3865486        total: 36.3s    remaining: 4m 45s
113:    learn: 0.3864333        total: 36.6s    remaining: 4m 44s
114:    learn: 0.3863745        total: 36.8s    remaining: 4m 43s
115:    learn: 0.3862755        total: 37.1s    remaining: 4m 42s
116:    learn: 0.3860232        total: 37.4s    remaining: 4m 42s
117:    learn: 0.3859099        total: 37.7s    remaining: 4m 41s
118:    learn: 0.3858460        total: 37.9s    remaining: 4m 40s
119:    learn: 0.3857258        total: 38.2s    remaining: 4m 40s
120:    learn: 0.3856384        total: 38.4s    remaining: 4m 39s
121:    learn: 0.3855997        total: 38.7s    remaining: 4m 38s
```

```
122:    learn: 0.3854696        total: 39s      remaining: 4m 37s
123:    learn: 0.3854102        total: 39.3s    remaining: 4m 37s
124:    learn: 0.3853559        total: 39.5s    remaining: 4m 36s
125:    learn: 0.3852833        total: 40s      remaining: 4m 37s
126:    learn: 0.3852237        total: 40.4s    remaining: 4m 38s
127:    learn: 0.3851597        total: 41s      remaining: 4m 39s
128:    learn: 0.3851305        total: 41.3s    remaining: 4m 39s
129:    learn: 0.3850466        total: 41.7s    remaining: 4m 39s
130:    learn: 0.3849522        total: 42.3s    remaining: 4m 40s
131:    learn: 0.3848980        total: 42.7s    remaining: 4m 40s
132:    learn: 0.3848183        total: 43.2s    remaining: 4m 41s
133:    learn: 0.3847586        total: 43.6s    remaining: 4m 41s
134:    learn: 0.3846682        total: 44.1s    remaining: 4m 42s
135:    learn: 0.3846129        total: 44.6s    remaining: 4m 43s
136:    learn: 0.3845633        total: 45s      remaining: 4m 43s
137:    learn: 0.3844767        total: 45.3s    remaining: 4m 43s
138:    learn: 0.3843300        total: 45.6s    remaining: 4m 42s
139:    learn: 0.3842258        total: 45.9s    remaining: 4m 42s
140:    learn: 0.3841579        total: 46.2s    remaining: 4m 41s
141:    learn: 0.3841104        total: 46.4s    remaining: 4m 40s
142:    learn: 0.3840403        total: 46.7s    remaining: 4m 40s
143:    learn: 0.3839699        total: 47.1s    remaining: 4m 39s
144:    learn: 0.3839322        total: 47.3s    remaining: 4m 38s
145:    learn: 0.3838102        total: 47.6s    remaining: 4m 38s
146:    learn: 0.3837281        total: 48s      remaining: 4m 38s
147:    learn: 0.3836730        total: 48.3s    remaining: 4m 37s
148:    learn: 0.3836003        total: 48.5s    remaining: 4m 37s
149:    learn: 0.3834095        total: 48.8s    remaining: 4m 36s
150:    learn: 0.3833192        total: 49.1s    remaining: 4m 36s
151:    learn: 0.3832226        total: 49.4s    remaining: 4m 35s
152:    learn: 0.3831846        total: 49.6s    remaining: 4m 34s
153:    learn: 0.3831491        total: 49.9s    remaining: 4m 34s
154:    learn: 0.3830624        total: 50.2s    remaining: 4m 33s
155:    learn: 0.3830245        total: 50.4s    remaining: 4m 32s
156:    learn: 0.3829828        total: 50.7s    remaining: 4m 32s
157:    learn: 0.3829038        total: 51s      remaining: 4m 31s
158:    learn: 0.3828575        total: 51.2s    remaining: 4m 31s
159:    learn: 0.3827783        total: 51.5s    remaining: 4m 30s
160:    learn: 0.3826883        total: 51.8s    remaining: 4m 29s
161:    learn: 0.3826267        total: 52.1s    remaining: 4m 29s
162:    learn: 0.3825865        total: 52.4s    remaining: 4m 28s
163:    learn: 0.3825269        total: 52.6s    remaining: 4m 28s
164:    learn: 0.3824725        total: 52.9s    remaining: 4m 27s
165:    learn: 0.3824390        total: 53.2s    remaining: 4m 27s
166:    learn: 0.3823902        total: 53.5s    remaining: 4m 26s
167:    learn: 0.3823468        total: 53.7s    remaining: 4m 26s
168:    learn: 0.3822993        total: 54s      remaining: 4m 25s
169:    learn: 0.3822667        total: 54.3s    remaining: 4m 25s
```

```
170:    learn: 0.3821730        total: 54.6s    remaining: 4m 24s
171:    learn: 0.3821056        total: 54.9s    remaining: 4m 24s
172:    learn: 0.3820428        total: 55.3s    remaining: 4m 24s
173:    learn: 0.3820004        total: 55.8s    remaining: 4m 24s
174:    learn: 0.3819687        total: 56.2s    remaining: 4m 25s
175:    learn: 0.3819178        total: 56.8s    remaining: 4m 25s
176:    learn: 0.3818683        total: 57.3s    remaining: 4m 26s
177:    learn: 0.3818044        total: 57.8s    remaining: 4m 26s
178:    learn: 0.3817264        total: 58.3s    remaining: 4m 27s
179:    learn: 0.3816971        total: 58.7s    remaining: 4m 27s
180:    learn: 0.3816650        total: 59.1s    remaining: 4m 27s
181:    learn: 0.3816210        total: 59.5s    remaining: 4m 27s
182:    learn: 0.3815802        total: 60s      remaining: 4m 27s
183:    learn: 0.3815367        total: 1m       remaining: 4m 27s
184:    learn: 0.3814932        total: 1m       remaining: 4m 27s
185:    learn: 0.3814586        total: 1m       remaining: 4m 26s
186:    learn: 0.3813726        total: 1m 1s    remaining: 4m 26s
187:    learn: 0.3813381        total: 1m 1s    remaining: 4m 25s
188:    learn: 0.3812829        total: 1m 1s    remaining: 4m 25s
189:    learn: 0.3812547        total: 1m 2s    remaining: 4m 24s
190:    learn: 0.3811912        total: 1m 2s    remaining: 4m 23s
191:    learn: 0.3811610        total: 1m 2s    remaining: 4m 23s
192:    learn: 0.3811386        total: 1m 2s    remaining: 4m 22s
193:    learn: 0.3810980        total: 1m 3s    remaining: 4m 22s
194:    learn: 0.3810648        total: 1m 3s    remaining: 4m 21s
195:    learn: 0.3810201        total: 1m 3s    remaining: 4m 21s
196:    learn: 0.3809814        total: 1m 3s    remaining: 4m 20s
197:    learn: 0.3809142        total: 1m 4s    remaining: 4m 20s
198:    learn: 0.3808760        total: 1m 4s    remaining: 4m 19s
199:    learn: 0.3808003        total: 1m 4s    remaining: 4m 18s
200:    learn: 0.3807546        total: 1m 5s    remaining: 4m 18s
201:    learn: 0.3806967        total: 1m 5s    remaining: 4m 17s
202:    learn: 0.3806349        total: 1m 5s    remaining: 4m 17s
203:    learn: 0.3805930        total: 1m 5s    remaining: 4m 17s
204:    learn: 0.3805639        total: 1m 6s    remaining: 4m 16s
205:    learn: 0.3805371        total: 1m 6s    remaining: 4m 15s
206:    learn: 0.3804478        total: 1m 6s    remaining: 4m 15s
207:    learn: 0.3804168        total: 1m 6s    remaining: 4m 14s
208:    learn: 0.3803774        total: 1m 7s    remaining: 4m 14s
209:    learn: 0.3803266        total: 1m 7s    remaining: 4m 13s
210:    learn: 0.3802653        total: 1m 7s    remaining: 4m 13s
211:    learn: 0.3802103        total: 1m 8s    remaining: 4m 12s
212:    learn: 0.3801456        total: 1m 8s    remaining: 4m 12s
213:    learn: 0.3800914        total: 1m 8s    remaining: 4m 12s
214:    learn: 0.3800571        total: 1m 8s    remaining: 4m 11s
215:    learn: 0.3800145        total: 1m 9s    remaining: 4m 11s
216:    learn: 0.3799553        total: 1m 9s    remaining: 4m 10s
217:    learn: 0.3799203        total: 1m 9s    remaining: 4m 10s
```

```
218:    learn: 0.3798694        total: 1m 10s   remaining: 4m 9s
219:    learn: 0.3798495        total: 1m 10s   remaining: 4m 9s
220:    learn: 0.3797936        total: 1m 10s   remaining: 4m 9s
221:    learn: 0.3797680        total: 1m 11s   remaining: 4m 9s
222:    learn: 0.3797380        total: 1m 11s   remaining: 4m 9s
223:    learn: 0.3797184        total: 1m 12s   remaining: 4m 9s
224:    learn: 0.3796995        total: 1m 12s   remaining: 4m 9s
225:    learn: 0.3796548        total: 1m 12s   remaining: 4m 10s
226:    learn: 0.3796273        total: 1m 13s   remaining: 4m 10s
227:    learn: 0.3795970        total: 1m 13s   remaining: 4m 10s
228:    learn: 0.3795632        total: 1m 14s   remaining: 4m 10s
229:    learn: 0.3795302        total: 1m 14s   remaining: 4m 11s
230:    learn: 0.3795077        total: 1m 15s   remaining: 4m 10s
231:    learn: 0.3794822        total: 1m 15s   remaining: 4m 10s
232:    learn: 0.3794632        total: 1m 15s   remaining: 4m 9s
233:    learn: 0.3794329        total: 1m 16s   remaining: 4m 9s
234:    learn: 0.3794019        total: 1m 16s   remaining: 4m 8s
235:    learn: 0.3793624        total: 1m 16s   remaining: 4m 8s
236:    learn: 0.3793101        total: 1m 16s   remaining: 4m 7s
237:    learn: 0.3792837        total: 1m 17s   remaining: 4m 7s
238:    learn: 0.3792623        total: 1m 17s   remaining: 4m 6s
239:    learn: 0.3792383        total: 1m 17s   remaining: 4m 6s
240:    learn: 0.3792163        total: 1m 17s   remaining: 4m 5s
241:    learn: 0.3791887        total: 1m 18s   remaining: 4m 4s
242:    learn: 0.3791662        total: 1m 18s   remaining: 4m 4s
243:    learn: 0.3791223        total: 1m 18s   remaining: 4m 3s
244:    learn: 0.3790872        total: 1m 19s   remaining: 4m 3s
245:    learn: 0.3790512        total: 1m 19s   remaining: 4m 3s
246:    learn: 0.3790204        total: 1m 19s   remaining: 4m 2s
247:    learn: 0.3789922        total: 1m 19s   remaining: 4m 2s
248:    learn: 0.3789553        total: 1m 20s   remaining: 4m 1s
249:    learn: 0.3789291        total: 1m 20s   remaining: 4m
250:    learn: 0.3788954        total: 1m 20s   remaining: 4m
251:    learn: 0.3788593        total: 1m 20s   remaining: 3m 59s
252:    learn: 0.3788314        total: 1m 21s   remaining: 3m 59s
253:    learn: 0.3787983        total: 1m 21s   remaining: 3m 59s
254:    learn: 0.3787561        total: 1m 21s   remaining: 3m 58s
255:    learn: 0.3787210        total: 1m 21s   remaining: 3m 58s
256:    learn: 0.3786928        total: 1m 22s   remaining: 3m 57s
257:    learn: 0.3786451        total: 1m 22s   remaining: 3m 57s
258:    learn: 0.3786177        total: 1m 22s   remaining: 3m 56s
259:    learn: 0.3785791        total: 1m 23s   remaining: 3m 56s
260:    learn: 0.3784174        total: 1m 23s   remaining: 3m 55s
261:    learn: 0.3783879        total: 1m 23s   remaining: 3m 55s
262:    learn: 0.3783508        total: 1m 23s   remaining: 3m 54s
263:    learn: 0.3782915        total: 1m 24s   remaining: 3m 54s
264:    learn: 0.3782689        total: 1m 24s   remaining: 3m 53s
265:    learn: 0.3782423        total: 1m 24s   remaining: 3m 53s
```

```
266:    learn: 0.3782136        total: 1m 24s   remaining: 3m 52s
267:    learn: 0.3781943        total: 1m 25s   remaining: 3m 52s
268:    learn: 0.3781515        total: 1m 25s   remaining: 3m 52s
269:    learn: 0.3781211        total: 1m 25s   remaining: 3m 52s
270:    learn: 0.3781039        total: 1m 26s   remaining: 3m 52s
271:    learn: 0.3780821        total: 1m 26s   remaining: 3m 52s
272:    learn: 0.3780520        total: 1m 27s   remaining: 3m 52s
273:    learn: 0.3780190        total: 1m 27s   remaining: 3m 52s
274:    learn: 0.3779919        total: 1m 28s   remaining: 3m 52s
275:    learn: 0.3779571        total: 1m 28s   remaining: 3m 52s
276:    learn: 0.3779305        total: 1m 29s   remaining: 3m 52s
277:    learn: 0.3778817        total: 1m 29s   remaining: 3m 52s
278:    learn: 0.3778504        total: 1m 30s   remaining: 3m 53s
279:    learn: 0.3778119        total: 1m 30s   remaining: 3m 53s
280:    learn: 0.3777800        total: 1m 30s   remaining: 3m 52s
281:    learn: 0.3777551        total: 1m 31s   remaining: 3m 52s
282:    learn: 0.3777316        total: 1m 31s   remaining: 3m 51s
283:    learn: 0.3777039        total: 1m 31s   remaining: 3m 51s
284:    learn: 0.3776786        total: 1m 32s   remaining: 3m 50s
285:    learn: 0.3776585        total: 1m 32s   remaining: 3m 50s
286:    learn: 0.3776134        total: 1m 32s   remaining: 3m 49s
287:    learn: 0.3775880        total: 1m 32s   remaining: 3m 49s
288:    learn: 0.3775603        total: 1m 33s   remaining: 3m 48s
289:    learn: 0.3775408        total: 1m 33s   remaining: 3m 48s
290:    learn: 0.3775110        total: 1m 33s   remaining: 3m 47s
291:    learn: 0.3774845        total: 1m 33s   remaining: 3m 47s
292:    learn: 0.3774369        total: 1m 34s   remaining: 3m 47s
293:    learn: 0.3773832        total: 1m 34s   remaining: 3m 46s
294:    learn: 0.3773600        total: 1m 35s   remaining: 3m 47s
295:    learn: 0.3773333        total: 1m 35s   remaining: 3m 47s
296:    learn: 0.3773033        total: 1m 35s   remaining: 3m 47s
297:    learn: 0.3772800        total: 1m 36s   remaining: 3m 46s
298:    learn: 0.3772385        total: 1m 36s   remaining: 3m 46s
299:    learn: 0.3772058        total: 1m 36s   remaining: 3m 45s
300:    learn: 0.3771833        total: 1m 36s   remaining: 3m 45s
301:    learn: 0.3771539        total: 1m 37s   remaining: 3m 44s
302:    learn: 0.3771210        total: 1m 37s   remaining: 3m 44s
303:    learn: 0.3770895        total: 1m 37s   remaining: 3m 43s
304:    learn: 0.3770476        total: 1m 38s   remaining: 3m 43s
305:    learn: 0.3770140        total: 1m 38s   remaining: 3m 42s
306:    learn: 0.3769927        total: 1m 38s   remaining: 3m 42s
307:    learn: 0.3769675        total: 1m 38s   remaining: 3m 41s
308:    learn: 0.3769468        total: 1m 39s   remaining: 3m 41s
309:    learn: 0.3769136        total: 1m 39s   remaining: 3m 41s
310:    learn: 0.3768642        total: 1m 39s   remaining: 3m 40s
311:    learn: 0.3768342        total: 1m 39s   remaining: 3m 40s
312:    learn: 0.3768142        total: 1m 40s   remaining: 3m 39s
313:    learn: 0.3767820        total: 1m 40s   remaining: 3m 39s
```

```
314:    learn: 0.3767649        total: 1m 40s   remaining: 3m 38s
315:    learn: 0.3767339        total: 1m 41s   remaining: 3m 38s
316:    learn: 0.3766995        total: 1m 41s   remaining: 3m 38s
317:    learn: 0.3766796        total: 1m 42s   remaining: 3m 38s
318:    learn: 0.3766290        total: 1m 42s   remaining: 3m 38s
319:    learn: 0.3766029        total: 1m 43s   remaining: 3m 38s
320:    learn: 0.3765804        total: 1m 43s   remaining: 3m 38s
321:    learn: 0.3765543        total: 1m 43s   remaining: 3m 38s
322:    learn: 0.3765177        total: 1m 44s   remaining: 3m 38s
323:    learn: 0.3764785        total: 1m 44s   remaining: 3m 38s
324:    learn: 0.3764446        total: 1m 45s   remaining: 3m 38s
325:    learn: 0.3763998        total: 1m 45s   remaining: 3m 39s
326:    learn: 0.3763731        total: 1m 46s   remaining: 3m 38s
327:    learn: 0.3763620        total: 1m 46s   remaining: 3m 38s
328:    learn: 0.3763213        total: 1m 46s   remaining: 3m 37s
329:    learn: 0.3762984        total: 1m 47s   remaining: 3m 37s
330:    learn: 0.3762496        total: 1m 47s   remaining: 3m 36s
331:    learn: 0.3762162        total: 1m 47s   remaining: 3m 36s
332:    learn: 0.3761926        total: 1m 47s   remaining: 3m 36s
333:    learn: 0.3761508        total: 1m 48s   remaining: 3m 35s
334:    learn: 0.3761184        total: 1m 48s   remaining: 3m 35s
335:    learn: 0.3760950        total: 1m 48s   remaining: 3m 34s
336:    learn: 0.3760526        total: 1m 48s   remaining: 3m 34s
337:    learn: 0.3760208        total: 1m 49s   remaining: 3m 33s
338:    learn: 0.3759945        total: 1m 49s   remaining: 3m 33s
339:    learn: 0.3759543        total: 1m 49s   remaining: 3m 32s
340:    learn: 0.3759131        total: 1m 49s   remaining: 3m 32s
341:    learn: 0.3758688        total: 1m 50s   remaining: 3m 31s
342:    learn: 0.3758401        total: 1m 50s   remaining: 3m 31s
343:    learn: 0.3758108        total: 1m 50s   remaining: 3m 31s
344:    learn: 0.3757924        total: 1m 50s   remaining: 3m 30s
345:    learn: 0.3757659        total: 1m 51s   remaining: 3m 30s
346:    learn: 0.3757240        total: 1m 51s   remaining: 3m 29s
347:    learn: 0.3756657        total: 1m 51s   remaining: 3m 29s
348:    learn: 0.3756454        total: 1m 52s   remaining: 3m 29s
349:    learn: 0.3756156        total: 1m 52s   remaining: 3m 28s
350:    learn: 0.3755931        total: 1m 52s   remaining: 3m 28s
351:    learn: 0.3755696        total: 1m 52s   remaining: 3m 27s
352:    learn: 0.3755395        total: 1m 53s   remaining: 3m 27s
353:    learn: 0.3755117        total: 1m 53s   remaining: 3m 26s
354:    learn: 0.3754785        total: 1m 53s   remaining: 3m 26s
355:    learn: 0.3754555        total: 1m 53s   remaining: 3m 25s
356:    learn: 0.3754146        total: 1m 54s   remaining: 3m 25s
357:    learn: 0.3753839        total: 1m 54s   remaining: 3m 25s
358:    learn: 0.3753476        total: 1m 54s   remaining: 3m 24s
359:    learn: 0.3753225        total: 1m 54s   remaining: 3m 24s
360:    learn: 0.3752931        total: 1m 55s   remaining: 3m 23s
361:    learn: 0.3752614        total: 1m 55s   remaining: 3m 23s
```

```
362:    learn: 0.3752351        total: 1m 55s   remaining: 3m 23s
363:    learn: 0.3752188        total: 1m 55s   remaining: 3m 22s
364:    learn: 0.3751956        total: 1m 56s   remaining: 3m 22s
365:    learn: 0.3751842        total: 1m 56s   remaining: 3m 22s
366:    learn: 0.3751603        total: 1m 57s   remaining: 3m 21s
367:    learn: 0.3751254        total: 1m 57s   remaining: 3m 21s
368:    learn: 0.3750927        total: 1m 58s   remaining: 3m 21s
369:    learn: 0.3750441        total: 1m 58s   remaining: 3m 21s
370:    learn: 0.3750099        total: 1m 59s   remaining: 3m 21s
371:    learn: 0.3749878        total: 1m 59s   remaining: 3m 21s
372:    learn: 0.3749639        total: 2m       remaining: 3m 21s
373:    learn: 0.3749134        total: 2m       remaining: 3m 21s
374:    learn: 0.3748857        total: 2m       remaining: 3m 21s
375:    learn: 0.3748349        total: 2m 1s    remaining: 3m 21s
376:    learn: 0.3747960        total: 2m 1s    remaining: 3m 21s
377:    learn: 0.3747777        total: 2m 2s    remaining: 3m 20s
378:    learn: 0.3747444        total: 2m 2s    remaining: 3m 20s
379:    learn: 0.3747078        total: 2m 2s    remaining: 3m 19s
380:    learn: 0.3746805        total: 2m 2s    remaining: 3m 19s
381:    learn: 0.3746548        total: 2m 3s    remaining: 3m 19s
382:    learn: 0.3746332        total: 2m 3s    remaining: 3m 18s
383:    learn: 0.3746069        total: 2m 3s    remaining: 3m 18s
384:    learn: 0.3745730        total: 2m 3s    remaining: 3m 18s
385:    learn: 0.3745414        total: 2m 4s    remaining: 3m 17s
386:    learn: 0.3744903        total: 2m 4s    remaining: 3m 17s
387:    learn: 0.3744701        total: 2m 4s    remaining: 3m 16s
388:    learn: 0.3744525        total: 2m 5s    remaining: 3m 16s
389:    learn: 0.3744279        total: 2m 5s    remaining: 3m 16s
390:    learn: 0.3744010        total: 2m 5s    remaining: 3m 15s
391:    learn: 0.3743710        total: 2m 5s    remaining: 3m 15s
392:    learn: 0.3743409        total: 2m 6s    remaining: 3m 14s
393:    learn: 0.3743131        total: 2m 6s    remaining: 3m 14s
394:    learn: 0.3742960        total: 2m 6s    remaining: 3m 14s
395:    learn: 0.3742737        total: 2m 6s    remaining: 3m 13s
396:    learn: 0.3742119        total: 2m 7s    remaining: 3m 13s
397:    learn: 0.3741720        total: 2m 7s    remaining: 3m 12s
398:    learn: 0.3741476        total: 2m 7s    remaining: 3m 12s
399:    learn: 0.3741004        total: 2m 8s    remaining: 3m 12s
400:    learn: 0.3740762        total: 2m 8s    remaining: 3m 11s
401:    learn: 0.3740545        total: 2m 8s    remaining: 3m 11s
402:    learn: 0.3740164        total: 2m 8s    remaining: 3m 10s
403:    learn: 0.3739887        total: 2m 9s    remaining: 3m 10s
404:    learn: 0.3739705        total: 2m 9s    remaining: 3m 10s
405:    learn: 0.3739480        total: 2m 9s    remaining: 3m 9s
406:    learn: 0.3739232        total: 2m 9s    remaining: 3m 9s
407:    learn: 0.3738805        total: 2m 10s   remaining: 3m 8s
408:    learn: 0.3738545        total: 2m 10s   remaining: 3m 8s
409:    learn: 0.3738196        total: 2m 10s   remaining: 3m 8s
```

```
410:    learn: 0.3738013        total: 2m 10s   remaining: 3m 7s
411:    learn: 0.3737535        total: 2m 11s   remaining: 3m 7s
412:    learn: 0.3737253        total: 2m 11s   remaining: 3m 7s
413:    learn: 0.3736998        total: 2m 12s   remaining: 3m 7s
414:    learn: 0.3736511        total: 2m 12s   remaining: 3m 6s
415:    learn: 0.3736378        total: 2m 13s   remaining: 3m 6s
416:    learn: 0.3736195        total: 2m 13s   remaining: 3m 6s
417:    learn: 0.3735919        total: 2m 13s   remaining: 3m 6s
418:    learn: 0.3735458        total: 2m 14s   remaining: 3m 6s
419:    learn: 0.3735178        total: 2m 14s   remaining: 3m 6s
420:    learn: 0.3734875        total: 2m 15s   remaining: 3m 6s
421:    learn: 0.3734635        total: 2m 15s   remaining: 3m 6s
422:    learn: 0.3734407        total: 2m 16s   remaining: 3m 6s
423:    learn: 0.3734074        total: 2m 16s   remaining: 3m 6s
424:    learn: 0.3733686        total: 2m 17s   remaining: 3m 5s
425:    learn: 0.3733456        total: 2m 17s   remaining: 3m 5s
426:    learn: 0.3733062        total: 2m 17s   remaining: 3m 4s
427:    learn: 0.3732888        total: 2m 18s   remaining: 3m 4s
428:    learn: 0.3732634        total: 2m 18s   remaining: 3m 4s
429:    learn: 0.3732239        total: 2m 18s   remaining: 3m 3s
430:    learn: 0.3731933        total: 2m 18s   remaining: 3m 3s
431:    learn: 0.3731557        total: 2m 19s   remaining: 3m 2s
432:    learn: 0.3731254        total: 2m 19s   remaining: 3m 2s
433:    learn: 0.3730995        total: 2m 19s   remaining: 3m 2s
434:    learn: 0.3730761        total: 2m 19s   remaining: 3m 1s
435:    learn: 0.3730457        total: 2m 20s   remaining: 3m 1s
436:    learn: 0.3730167        total: 2m 20s   remaining: 3m
437:    learn: 0.3730048        total: 2m 20s   remaining: 3m
438:    learn: 0.3729721        total: 2m 20s   remaining: 3m
439:    learn: 0.3729381        total: 2m 21s   remaining: 2m 59s
440:    learn: 0.3729228        total: 2m 21s   remaining: 2m 59s
441:    learn: 0.3729039        total: 2m 21s   remaining: 2m 58s
442:    learn: 0.3728733        total: 2m 22s   remaining: 2m 58s
443:    learn: 0.3728420        total: 2m 22s   remaining: 2m 58s
444:    learn: 0.3728182        total: 2m 22s   remaining: 2m 57s
445:    learn: 0.3727970        total: 2m 22s   remaining: 2m 57s
446:    learn: 0.3727724        total: 2m 23s   remaining: 2m 56s
447:    learn: 0.3727582        total: 2m 23s   remaining: 2m 56s
448:    learn: 0.3727345        total: 2m 23s   remaining: 2m 56s
449:    learn: 0.3727188        total: 2m 23s   remaining: 2m 55s
450:    learn: 0.3726839        total: 2m 24s   remaining: 2m 55s
451:    learn: 0.3726609        total: 2m 24s   remaining: 2m 55s
452:    learn: 0.3726434        total: 2m 24s   remaining: 2m 54s
453:    learn: 0.3726206        total: 2m 24s   remaining: 2m 54s
454:    learn: 0.3725966        total: 2m 25s   remaining: 2m 53s
455:    learn: 0.3725723        total: 2m 25s   remaining: 2m 53s
456:    learn: 0.3725494        total: 2m 25s   remaining: 2m 53s
457:    learn: 0.3725085        total: 2m 25s   remaining: 2m 52s
```

```
458:    learn: 0.3724762        total: 2m 26s   remaining: 2m 52s
459:    learn: 0.3724489        total: 2m 26s   remaining: 2m 51s
460:    learn: 0.3724299        total: 2m 26s   remaining: 2m 51s
461:    learn: 0.3724014        total: 2m 26s   remaining: 2m 51s
462:    learn: 0.3723660        total: 2m 27s   remaining: 2m 50s
463:    learn: 0.3723436        total: 2m 27s   remaining: 2m 50s
464:    learn: 0.3723161        total: 2m 28s   remaining: 2m 50s
465:    learn: 0.3722889        total: 2m 28s   remaining: 2m 50s
466:    learn: 0.3722283        total: 2m 29s   remaining: 2m 50s
467:    learn: 0.3722052        total: 2m 29s   remaining: 2m 50s
468:    learn: 0.3721895        total: 2m 30s   remaining: 2m 50s
469:    learn: 0.3721753        total: 2m 30s   remaining: 2m 49s
470:    learn: 0.3721513        total: 2m 31s   remaining: 2m 49s
471:    learn: 0.3721309        total: 2m 31s   remaining: 2m 49s
472:    learn: 0.3721060        total: 2m 32s   remaining: 2m 49s
473:    learn: 0.3720840        total: 2m 32s   remaining: 2m 49s
474:    learn: 0.3720654        total: 2m 32s   remaining: 2m 49s
475:    learn: 0.3720368        total: 2m 33s   remaining: 2m 48s
476:    learn: 0.3720116        total: 2m 33s   remaining: 2m 48s
477:    learn: 0.3719872        total: 2m 33s   remaining: 2m 47s
478:    learn: 0.3719625        total: 2m 34s   remaining: 2m 47s
479:    learn: 0.3719414        total: 2m 34s   remaining: 2m 47s
480:    learn: 0.3719211        total: 2m 34s   remaining: 2m 46s
481:    learn: 0.3719004        total: 2m 34s   remaining: 2m 46s
482:    learn: 0.3718717        total: 2m 35s   remaining: 2m 46s
483:    learn: 0.3718534        total: 2m 35s   remaining: 2m 45s
484:    learn: 0.3718331        total: 2m 35s   remaining: 2m 45s
485:    learn: 0.3718041        total: 2m 35s   remaining: 2m 44s
486:    learn: 0.3717768        total: 2m 36s   remaining: 2m 44s
487:    learn: 0.3717557        total: 2m 36s   remaining: 2m 44s
488:    learn: 0.3717259        total: 2m 36s   remaining: 2m 43s
489:    learn: 0.3717023        total: 2m 36s   remaining: 2m 43s
490:    learn: 0.3716429        total: 2m 37s   remaining: 2m 43s
491:    learn: 0.3716177        total: 2m 37s   remaining: 2m 42s
492:    learn: 0.3715973        total: 2m 37s   remaining: 2m 42s
493:    learn: 0.3715743        total: 2m 38s   remaining: 2m 41s
494:    learn: 0.3715542        total: 2m 38s   remaining: 2m 41s
495:    learn: 0.3715359        total: 2m 38s   remaining: 2m 41s
496:    learn: 0.3714944        total: 2m 38s   remaining: 2m 40s
497:    learn: 0.3714449        total: 2m 39s   remaining: 2m 40s
498:    learn: 0.3714238        total: 2m 39s   remaining: 2m 40s
499:    learn: 0.3714062        total: 2m 39s   remaining: 2m 39s
500:    learn: 0.3713824        total: 2m 39s   remaining: 2m 39s
501:    learn: 0.3713504        total: 2m 40s   remaining: 2m 38s
502:    learn: 0.3713321        total: 2m 40s   remaining: 2m 38s
503:    learn: 0.3713202        total: 2m 40s   remaining: 2m 38s
504:    learn: 0.3713087        total: 2m 40s   remaining: 2m 37s
505:    learn: 0.3712864        total: 2m 41s   remaining: 2m 37s
```

```
506:    learn: 0.3712583        total: 2m 41s   remaining: 2m 36s
507:    learn: 0.3712309        total: 2m 41s   remaining: 2m 36s
508:    learn: 0.3712006        total: 2m 41s   remaining: 2m 36s
509:    learn: 0.3711882        total: 2m 42s   remaining: 2m 35s
510:    learn: 0.3711613        total: 2m 42s   remaining: 2m 35s
511:    learn: 0.3711376        total: 2m 42s   remaining: 2m 35s
512:    learn: 0.3711156        total: 2m 43s   remaining: 2m 35s
513:    learn: 0.3710773        total: 2m 43s   remaining: 2m 34s
514:    learn: 0.3710358        total: 2m 44s   remaining: 2m 34s
515:    learn: 0.3709984        total: 2m 44s   remaining: 2m 34s
516:    learn: 0.3709781        total: 2m 45s   remaining: 2m 34s
517:    learn: 0.3709630        total: 2m 45s   remaining: 2m 34s
518:    learn: 0.3709510        total: 2m 46s   remaining: 2m 34s
519:    learn: 0.3709354        total: 2m 46s   remaining: 2m 33s
520:    learn: 0.3709077        total: 2m 47s   remaining: 2m 33s
521:    learn: 0.3708722        total: 2m 47s   remaining: 2m 33s
522:    learn: 0.3708521        total: 2m 48s   remaining: 2m 33s
523:    learn: 0.3708262        total: 2m 48s   remaining: 2m 32s
524:    learn: 0.3708042        total: 2m 48s   remaining: 2m 32s
525:    learn: 0.3707918        total: 2m 48s   remaining: 2m 32s
526:    learn: 0.3707693        total: 2m 49s   remaining: 2m 31s
527:    learn: 0.3707371        total: 2m 49s   remaining: 2m 31s
528:    learn: 0.3707119        total: 2m 49s   remaining: 2m 31s
529:    learn: 0.3707008        total: 2m 49s   remaining: 2m 30s
530:    learn: 0.3706866        total: 2m 50s   remaining: 2m 30s
531:    learn: 0.3706594        total: 2m 50s   remaining: 2m 29s
532:    learn: 0.3706254        total: 2m 50s   remaining: 2m 29s
533:    learn: 0.3705987        total: 2m 50s   remaining: 2m 29s
534:    learn: 0.3705797        total: 2m 51s   remaining: 2m 28s
535:    learn: 0.3705447        total: 2m 51s   remaining: 2m 28s
536:    learn: 0.3705258        total: 2m 51s   remaining: 2m 28s
537:    learn: 0.3705055        total: 2m 52s   remaining: 2m 27s
538:    learn: 0.3704809        total: 2m 52s   remaining: 2m 27s
539:    learn: 0.3704517        total: 2m 52s   remaining: 2m 27s
540:    learn: 0.3704290        total: 2m 52s   remaining: 2m 26s
541:    learn: 0.3704088        total: 2m 53s   remaining: 2m 26s
542:    learn: 0.3703947        total: 2m 53s   remaining: 2m 25s
543:    learn: 0.3703741        total: 2m 53s   remaining: 2m 25s
544:    learn: 0.3703556        total: 2m 53s   remaining: 2m 25s
545:    learn: 0.3703322        total: 2m 54s   remaining: 2m 24s
546:    learn: 0.3703164        total: 2m 54s   remaining: 2m 24s
547:    learn: 0.3702989        total: 2m 54s   remaining: 2m 24s
548:    learn: 0.3702824        total: 2m 54s   remaining: 2m 23s
549:    learn: 0.3702369        total: 2m 55s   remaining: 2m 23s
550:    learn: 0.3702181        total: 2m 55s   remaining: 2m 23s
551:    learn: 0.3702056        total: 2m 55s   remaining: 2m 22s
552:    learn: 0.3701586        total: 2m 56s   remaining: 2m 22s
553:    learn: 0.3701337        total: 2m 56s   remaining: 2m 21s
```

```
554:    learn: 0.3701103        total: 2m 56s   remaining: 2m 21s
555:    learn: 0.3700826        total: 2m 56s   remaining: 2m 21s
556:    learn: 0.3700657        total: 2m 57s   remaining: 2m 20s
557:    learn: 0.3700271        total: 2m 57s   remaining: 2m 20s
558:    learn: 0.3700142        total: 2m 57s   remaining: 2m 20s
559:    learn: 0.3700014        total: 2m 57s   remaining: 2m 19s
560:    learn: 0.3699883        total: 2m 58s   remaining: 2m 19s
561:    learn: 0.3699727        total: 2m 58s   remaining: 2m 19s
562:    learn: 0.3699516        total: 2m 59s   remaining: 2m 19s
563:    learn: 0.3699274        total: 2m 59s   remaining: 2m 18s
564:    learn: 0.3698995        total: 3m       remaining: 2m 18s
565:    learn: 0.3698804        total: 3m       remaining: 2m 18s
566:    learn: 0.3698545        total: 3m 1s    remaining: 2m 18s
567:    learn: 0.3698276        total: 3m 1s    remaining: 2m 18s
568:    learn: 0.3698121        total: 3m 2s    remaining: 2m 17s
569:    learn: 0.3697864        total: 3m 2s    remaining: 2m 17s
570:    learn: 0.3697498        total: 3m 3s    remaining: 2m 17s
571:    learn: 0.3697226        total: 3m 3s    remaining: 2m 17s
572:    learn: 0.3697036        total: 3m 3s    remaining: 2m 16s
573:    learn: 0.3696788        total: 3m 4s    remaining: 2m 16s
574:    learn: 0.3696549        total: 3m 4s    remaining: 2m 16s
575:    learn: 0.3696267        total: 3m 4s    remaining: 2m 15s
576:    learn: 0.3695996        total: 3m 4s    remaining: 2m 15s
577:    learn: 0.3695768        total: 3m 5s    remaining: 2m 15s
578:    learn: 0.3695669        total: 3m 5s    remaining: 2m 14s
579:    learn: 0.3695460        total: 3m 5s    remaining: 2m 14s
580:    learn: 0.3695068        total: 3m 5s    remaining: 2m 14s
581:    learn: 0.3694806        total: 3m 6s    remaining: 2m 13s
582:    learn: 0.3694609        total: 3m 6s    remaining: 2m 13s
583:    learn: 0.3694334        total: 3m 6s    remaining: 2m 13s
584:    learn: 0.3694254        total: 3m 7s    remaining: 2m 12s
585:    learn: 0.3694050        total: 3m 7s    remaining: 2m 12s
586:    learn: 0.3693924        total: 3m 7s    remaining: 2m 12s
587:    learn: 0.3693724        total: 3m 7s    remaining: 2m 11s
588:    learn: 0.3693422        total: 3m 8s    remaining: 2m 11s
589:    learn: 0.3693215        total: 3m 8s    remaining: 2m 11s
590:    learn: 0.3693035        total: 3m 8s    remaining: 2m 10s
591:    learn: 0.3692934        total: 3m 9s    remaining: 2m 10s
592:    learn: 0.3692686        total: 3m 9s    remaining: 2m 9s
593:    learn: 0.3692412        total: 3m 9s    remaining: 2m 9s
594:    learn: 0.3692128        total: 3m 9s    remaining: 2m 9s
595:    learn: 0.3692042        total: 3m 10s   remaining: 2m 8s
596:    learn: 0.3691775        total: 3m 10s   remaining: 2m 8s
597:    learn: 0.3691551        total: 3m 10s   remaining: 2m 8s
598:    learn: 0.3691180        total: 3m 10s   remaining: 2m 7s
599:    learn: 0.3691052        total: 3m 11s   remaining: 2m 7s
600:    learn: 0.3690753        total: 3m 11s   remaining: 2m 7s
601:    learn: 0.3690485        total: 3m 11s   remaining: 2m 6s
```

```
602:    learn: 0.3690223        total: 3m 11s   remaining: 2m 6s
603:    learn: 0.3690054        total: 3m 12s   remaining: 2m 6s
604:    learn: 0.3689783        total: 3m 12s   remaining: 2m 5s
605:    learn: 0.3689635        total: 3m 12s   remaining: 2m 5s
606:    learn: 0.3689386        total: 3m 13s   remaining: 2m 4s
607:    learn: 0.3689110        total: 3m 13s   remaining: 2m 4s
608:    learn: 0.3688757        total: 3m 13s   remaining: 2m 4s
609:    learn: 0.3688585        total: 3m 14s   remaining: 2m 4s
610:    learn: 0.3688296        total: 3m 14s   remaining: 2m 3s
611:    learn: 0.3688190        total: 3m 15s   remaining: 2m 3s
612:    learn: 0.3687859        total: 3m 15s   remaining: 2m 3s
613:    learn: 0.3687702        total: 3m 16s   remaining: 2m 3s
614:    learn: 0.3687331        total: 3m 16s   remaining: 2m 3s
615:    learn: 0.3687187        total: 3m 17s   remaining: 2m 2s
616:    learn: 0.3686944        total: 3m 17s   remaining: 2m 2s
617:    learn: 0.3686648        total: 3m 18s   remaining: 2m 2s
618:    learn: 0.3686432        total: 3m 18s   remaining: 2m 2s
619:    learn: 0.3686207        total: 3m 18s   remaining: 2m 1s
620:    learn: 0.3685896        total: 3m 19s   remaining: 2m 1s
621:    learn: 0.3685547        total: 3m 19s   remaining: 2m 1s
622:    learn: 0.3685321        total: 3m 19s   remaining: 2m
623:    learn: 0.3685152        total: 3m 20s   remaining: 2m
624:    learn: 0.3684916        total: 3m 20s   remaining: 2m
625:    learn: 0.3684654        total: 3m 20s   remaining: 1m 59s
626:    learn: 0.3684342        total: 3m 20s   remaining: 1m 59s
627:    learn: 0.3684132        total: 3m 21s   remaining: 1m 59s
628:    learn: 0.3683957        total: 3m 21s   remaining: 1m 58s
629:    learn: 0.3683756        total: 3m 22s   remaining: 1m 58s
630:    learn: 0.3683504        total: 3m 22s   remaining: 1m 58s
631:    learn: 0.3683306        total: 3m 22s   remaining: 1m 58s
632:    learn: 0.3683051        total: 3m 23s   remaining: 1m 57s
633:    learn: 0.3682925        total: 3m 23s   remaining: 1m 57s
634:    learn: 0.3682739        total: 3m 23s   remaining: 1m 57s
635:    learn: 0.3682520        total: 3m 23s   remaining: 1m 56s
636:    learn: 0.3682341        total: 3m 24s   remaining: 1m 56s
637:    learn: 0.3681962        total: 3m 24s   remaining: 1m 56s
638:    learn: 0.3681693        total: 3m 24s   remaining: 1m 55s
639:    learn: 0.3681543        total: 3m 25s   remaining: 1m 55s
640:    learn: 0.3681331        total: 3m 25s   remaining: 1m 54s
641:    learn: 0.3681160        total: 3m 25s   remaining: 1m 54s
642:    learn: 0.3680866        total: 3m 25s   remaining: 1m 54s
643:    learn: 0.3680679        total: 3m 26s   remaining: 1m 53s
644:    learn: 0.3680347        total: 3m 26s   remaining: 1m 53s
645:    learn: 0.3680153        total: 3m 26s   remaining: 1m 53s
646:    learn: 0.3679954        total: 3m 27s   remaining: 1m 52s
647:    learn: 0.3679830        total: 3m 27s   remaining: 1m 52s
648:    learn: 0.3679516        total: 3m 27s   remaining: 1m 52s
649:    learn: 0.3679430        total: 3m 27s   remaining: 1m 51s
```

```
650:    learn: 0.3679099        total: 3m 28s   remaining: 1m 51s
651:    learn: 0.3678820        total: 3m 28s   remaining: 1m 51s
652:    learn: 0.3678550        total: 3m 28s   remaining: 1m 50s
653:    learn: 0.3678278        total: 3m 28s   remaining: 1m 50s
654:    learn: 0.3678056        total: 3m 29s   remaining: 1m 50s
655:    learn: 0.3677827        total: 3m 29s   remaining: 1m 50s
656:    learn: 0.3677661        total: 3m 30s   remaining: 1m 49s
657:    learn: 0.3677490        total: 3m 30s   remaining: 1m 49s
658:    learn: 0.3677250        total: 3m 31s   remaining: 1m 49s
659:    learn: 0.3676638        total: 3m 31s   remaining: 1m 49s
660:    learn: 0.3676439        total: 3m 32s   remaining: 1m 48s
661:    learn: 0.3676386        total: 3m 32s   remaining: 1m 48s
662:    learn: 0.3676169        total: 3m 33s   remaining: 1m 48s
663:    learn: 0.3675858        total: 3m 33s   remaining: 1m 48s
664:    learn: 0.3675714        total: 3m 34s   remaining: 1m 47s
665:    learn: 0.3675546        total: 3m 34s   remaining: 1m 47s
666:    learn: 0.3675417        total: 3m 34s   remaining: 1m 47s
667:    learn: 0.3675243        total: 3m 35s   remaining: 1m 46s
668:    learn: 0.3674963        total: 3m 35s   remaining: 1m 46s
669:    learn: 0.3674749        total: 3m 35s   remaining: 1m 46s
670:    learn: 0.3674524        total: 3m 35s   remaining: 1m 45s
671:    learn: 0.3674372        total: 3m 36s   remaining: 1m 45s
672:    learn: 0.3674130        total: 3m 36s   remaining: 1m 45s
673:    learn: 0.3674040        total: 3m 36s   remaining: 1m 44s
674:    learn: 0.3673898        total: 3m 36s   remaining: 1m 44s
675:    learn: 0.3673671        total: 3m 37s   remaining: 1m 44s
676:    learn: 0.3673507        total: 3m 37s   remaining: 1m 43s
677:    learn: 0.3673324        total: 3m 37s   remaining: 1m 43s
678:    learn: 0.3673084        total: 3m 37s   remaining: 1m 43s
679:    learn: 0.3672814        total: 3m 38s   remaining: 1m 42s
680:    learn: 0.3672475        total: 3m 38s   remaining: 1m 42s
681:    learn: 0.3672192        total: 3m 38s   remaining: 1m 41s
682:    learn: 0.3671798        total: 3m 39s   remaining: 1m 41s
683:    learn: 0.3671337        total: 3m 39s   remaining: 1m 41s
684:    learn: 0.3671116        total: 3m 39s   remaining: 1m 40s
685:    learn: 0.3670974        total: 3m 39s   remaining: 1m 40s
686:    learn: 0.3670737        total: 3m 40s   remaining: 1m 40s
687:    learn: 0.3670650        total: 3m 40s   remaining: 1m 39s
688:    learn: 0.3670477        total: 3m 40s   remaining: 1m 39s
689:    learn: 0.3670247        total: 3m 40s   remaining: 1m 39s
690:    learn: 0.3670025        total: 3m 41s   remaining: 1m 38s
691:    learn: 0.3669675        total: 3m 41s   remaining: 1m 38s
692:    learn: 0.3669444        total: 3m 41s   remaining: 1m 38s
693:    learn: 0.3669262        total: 3m 41s   remaining: 1m 37s
694:    learn: 0.3668976        total: 3m 42s   remaining: 1m 37s
695:    learn: 0.3668768        total: 3m 42s   remaining: 1m 37s
696:    learn: 0.3668513        total: 3m 42s   remaining: 1m 36s
697:    learn: 0.3668289        total: 3m 42s   remaining: 1m 36s
```

```
698:    learn: 0.3668055        total: 3m 43s   remaining: 1m 36s
699:    learn: 0.3667876        total: 3m 43s   remaining: 1m 35s
700:    learn: 0.3667544        total: 3m 43s   remaining: 1m 35s
701:    learn: 0.3667358        total: 3m 44s   remaining: 1m 35s
702:    learn: 0.3667195        total: 3m 44s   remaining: 1m 34s
703:    learn: 0.3667009        total: 3m 44s   remaining: 1m 34s
704:    learn: 0.3666822        total: 3m 45s   remaining: 1m 34s
705:    learn: 0.3666398        total: 3m 45s   remaining: 1m 34s
706:    learn: 0.3666205        total: 3m 46s   remaining: 1m 33s
707:    learn: 0.3665989        total: 3m 46s   remaining: 1m 33s
708:    learn: 0.3665700        total: 3m 47s   remaining: 1m 33s
709:    learn: 0.3665398        total: 3m 47s   remaining: 1m 33s
710:    learn: 0.3665023        total: 3m 48s   remaining: 1m 32s
711:    learn: 0.3664655        total: 3m 48s   remaining: 1m 32s
712:    learn: 0.3664422        total: 3m 49s   remaining: 1m 32s
713:    learn: 0.3664088        total: 3m 50s   remaining: 1m 32s
714:    learn: 0.3663849        total: 3m 50s   remaining: 1m 31s
715:    learn: 0.3663573        total: 3m 51s   remaining: 1m 31s
716:    learn: 0.3663389        total: 3m 51s   remaining: 1m 31s
717:    learn: 0.3663302        total: 3m 52s   remaining: 1m 31s
718:    learn: 0.3663223        total: 3m 52s   remaining: 1m 30s
719:    learn: 0.3663080        total: 3m 53s   remaining: 1m 30s
720:    learn: 0.3662845        total: 3m 53s   remaining: 1m 30s
721:    learn: 0.3662636        total: 3m 53s   remaining: 1m 30s
722:    learn: 0.3662503        total: 3m 54s   remaining: 1m 29s
723:    learn: 0.3662291        total: 3m 54s   remaining: 1m 29s
724:    learn: 0.3662080        total: 3m 54s   remaining: 1m 29s
725:    learn: 0.3661905        total: 3m 54s   remaining: 1m 28s
726:    learn: 0.3661607        total: 3m 55s   remaining: 1m 28s
727:    learn: 0.3661266        total: 3m 55s   remaining: 1m 27s
728:    learn: 0.3660943        total: 3m 55s   remaining: 1m 27s
729:    learn: 0.3660744        total: 3m 56s   remaining: 1m 27s
730:    learn: 0.3660549        total: 3m 56s   remaining: 1m 26s
731:    learn: 0.3659854        total: 3m 56s   remaining: 1m 26s
732:    learn: 0.3659655        total: 3m 56s   remaining: 1m 26s
733:    learn: 0.3659408        total: 3m 57s   remaining: 1m 25s
734:    learn: 0.3659200        total: 3m 57s   remaining: 1m 25s
735:    learn: 0.3658998        total: 3m 57s   remaining: 1m 25s
736:    learn: 0.3658865        total: 3m 57s   remaining: 1m 24s
737:    learn: 0.3658581        total: 3m 58s   remaining: 1m 24s
738:    learn: 0.3658387        total: 3m 58s   remaining: 1m 24s
739:    learn: 0.3658171        total: 3m 58s   remaining: 1m 23s
740:    learn: 0.3658020        total: 3m 58s   remaining: 1m 23s
741:    learn: 0.3657779        total: 3m 59s   remaining: 1m 23s
742:    learn: 0.3657469        total: 3m 59s   remaining: 1m 22s
743:    learn: 0.3657335        total: 3m 59s   remaining: 1m 22s
744:    learn: 0.3657079        total: 4m       remaining: 1m 22s
745:    learn: 0.3656819        total: 4m       remaining: 1m 21s
```

```
746:    learn: 0.3656552        total: 4m       remaining: 1m 21s
747:    learn: 0.3656309        total: 4m       remaining: 1m 21s
748:    learn: 0.3656220        total: 4m 1s    remaining: 1m 20s
749:    learn: 0.3656002        total: 4m 1s    remaining: 1m 20s
750:    learn: 0.3655830        total: 4m 1s    remaining: 1m 20s
751:    learn: 0.3655608        total: 4m 1s    remaining: 1m 19s
752:    learn: 0.3655443        total: 4m 2s    remaining: 1m 19s
753:    learn: 0.3655263        total: 4m 2s    remaining: 1m 19s
754:    learn: 0.3654999        total: 4m 2s    remaining: 1m 18s
755:    learn: 0.3654831        total: 4m 2s    remaining: 1m 18s
756:    learn: 0.3654572        total: 4m 3s    remaining: 1m 18s
757:    learn: 0.3654316        total: 4m 3s    remaining: 1m 17s
758:    learn: 0.3654115        total: 4m 4s    remaining: 1m 17s
759:    learn: 0.3653881        total: 4m 4s    remaining: 1m 17s
760:    learn: 0.3653770        total: 4m 5s    remaining: 1m 17s
761:    learn: 0.3653615        total: 4m 5s    remaining: 1m 16s
762:    learn: 0.3653405        total: 4m 6s    remaining: 1m 16s
763:    learn: 0.3653297        total: 4m 6s    remaining: 1m 16s
764:    learn: 0.3652831        total: 4m 7s    remaining: 1m 15s
765:    learn: 0.3652680        total: 4m 7s    remaining: 1m 15s
766:    learn: 0.3652324        total: 4m 8s    remaining: 1m 15s
767:    learn: 0.3652144        total: 4m 8s    remaining: 1m 15s
768:    learn: 0.3651980        total: 4m 8s    remaining: 1m 14s
769:    learn: 0.3651689        total: 4m 9s    remaining: 1m 14s
770:    learn: 0.3651511        total: 4m 9s    remaining: 1m 14s
771:    learn: 0.3651353        total: 4m 9s    remaining: 1m 13s
772:    learn: 0.3651122        total: 4m 9s    remaining: 1m 13s
773:    learn: 0.3650937        total: 4m 10s   remaining: 1m 13s
774:    learn: 0.3650827        total: 4m 10s   remaining: 1m 12s
775:    learn: 0.3650704        total: 4m 10s   remaining: 1m 12s
776:    learn: 0.3650525        total: 4m 10s   remaining: 1m 11s
777:    learn: 0.3650267        total: 4m 11s   remaining: 1m 11s
778:    learn: 0.3650026        total: 4m 11s   remaining: 1m 11s
779:    learn: 0.3649824        total: 4m 11s   remaining: 1m 10s
780:    learn: 0.3649553        total: 4m 11s   remaining: 1m 10s
781:    learn: 0.3649337        total: 4m 12s   remaining: 1m 10s
782:    learn: 0.3649135        total: 4m 12s   remaining: 1m 9s
783:    learn: 0.3648805        total: 4m 12s   remaining: 1m 9s
784:    learn: 0.3648551        total: 4m 13s   remaining: 1m 9s
785:    learn: 0.3648419        total: 4m 13s   remaining: 1m 8s
786:    learn: 0.3648123        total: 4m 13s   remaining: 1m 8s
787:    learn: 0.3647930        total: 4m 13s   remaining: 1m 8s
788:    learn: 0.3647727        total: 4m 14s   remaining: 1m 7s
789:    learn: 0.3647475        total: 4m 14s   remaining: 1m 7s
790:    learn: 0.3647300        total: 4m 14s   remaining: 1m 7s
791:    learn: 0.3647123        total: 4m 14s   remaining: 1m 6s
792:    learn: 0.3646830        total: 4m 15s   remaining: 1m 6s
793:    learn: 0.3646576        total: 4m 15s   remaining: 1m 6s
```

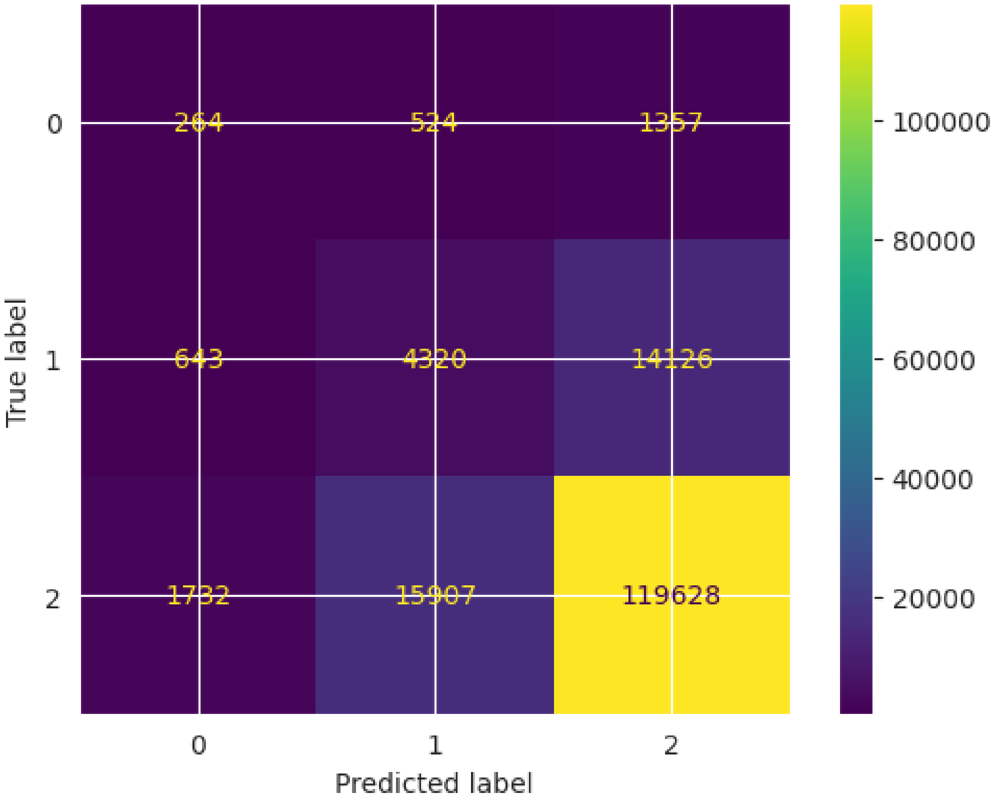
```
794:    learn: 0.3646449        total: 4m 15s   remaining: 1m 5s
795:    learn: 0.3646215        total: 4m 15s   remaining: 1m 5s
796:    learn: 0.3645920        total: 4m 16s   remaining: 1m 5s
797:    learn: 0.3645648        total: 4m 16s   remaining: 1m 4s
798:    learn: 0.3645470        total: 4m 16s   remaining: 1m 4s
799:    learn: 0.3645376        total: 4m 17s   remaining: 1m 4s
800:    learn: 0.3645143        total: 4m 17s   remaining: 1m 3s
801:    learn: 0.3644979        total: 4m 17s   remaining: 1m 3s
802:    learn: 0.3644651        total: 4m 17s   remaining: 1m 3s
803:    learn: 0.3644389        total: 4m 18s   remaining: 1m 2s
804:    learn: 0.3644244        total: 4m 18s   remaining: 1m 2s
805:    learn: 0.3643988        total: 4m 19s   remaining: 1m 2s
806:    learn: 0.3643840        total: 4m 19s   remaining: 1m 2s
807:    learn: 0.3643691        total: 4m 19s   remaining: 1m 1s
808:    learn: 0.3643480        total: 4m 20s   remaining: 1m 1s
809:    learn: 0.3643233        total: 4m 20s   remaining: 1m 1s
810:    learn: 0.3643089        total: 4m 21s   remaining: 1m
811:    learn: 0.3642943        total: 4m 22s   remaining: 1m
812:    learn: 0.3642599        total: 4m 22s   remaining: 1m
813:    learn: 0.3642320        total: 4m 22s   remaining: 1m
814:    learn: 0.3642110        total: 4m 23s   remaining: 59.8s
815:    learn: 0.3641940        total: 4m 23s   remaining: 59.4s
816:    learn: 0.3641736        total: 4m 23s   remaining: 59.1s
817:    learn: 0.3641558        total: 4m 24s   remaining: 58.7s
818:    learn: 0.3641368        total: 4m 24s   remaining: 58.4s
819:    learn: 0.3641308        total: 4m 24s   remaining: 58.1s
820:    learn: 0.3641105        total: 4m 24s   remaining: 57.7s
821:    learn: 0.3640858        total: 4m 25s   remaining: 57.4s
822:    learn: 0.3640677        total: 4m 25s   remaining: 57.1s
823:    learn: 0.3640520        total: 4m 25s   remaining: 56.7s
824:    learn: 0.3640370        total: 4m 25s   remaining: 56.4s
825:    learn: 0.3640191        total: 4m 26s   remaining: 56.1s
826:    learn: 0.3639988        total: 4m 26s   remaining: 55.7s
827:    learn: 0.3639888        total: 4m 26s   remaining: 55.4s
828:    learn: 0.3639679        total: 4m 26s   remaining: 55.1s
829:    learn: 0.3639531        total: 4m 27s   remaining: 54.7s
830:    learn: 0.3639375        total: 4m 27s   remaining: 54.4s
831:    learn: 0.3639261        total: 4m 27s   remaining: 54.1s
832:    learn: 0.3639023        total: 4m 27s   remaining: 53.7s
833:    learn: 0.3638822        total: 4m 28s   remaining: 53.4s
834:    learn: 0.3638608        total: 4m 28s   remaining: 53.1s
835:    learn: 0.3638437        total: 4m 28s   remaining: 52.7s
836:    learn: 0.3638202        total: 4m 29s   remaining: 52.4s
837:    learn: 0.3638017        total: 4m 29s   remaining: 52.1s
838:    learn: 0.3637862        total: 4m 29s   remaining: 51.7s
839:    learn: 0.3637640        total: 4m 29s   remaining: 51.4s
840:    learn: 0.3637360        total: 4m 30s   remaining: 51.1s
841:    learn: 0.3637091        total: 4m 30s   remaining: 50.7s
```

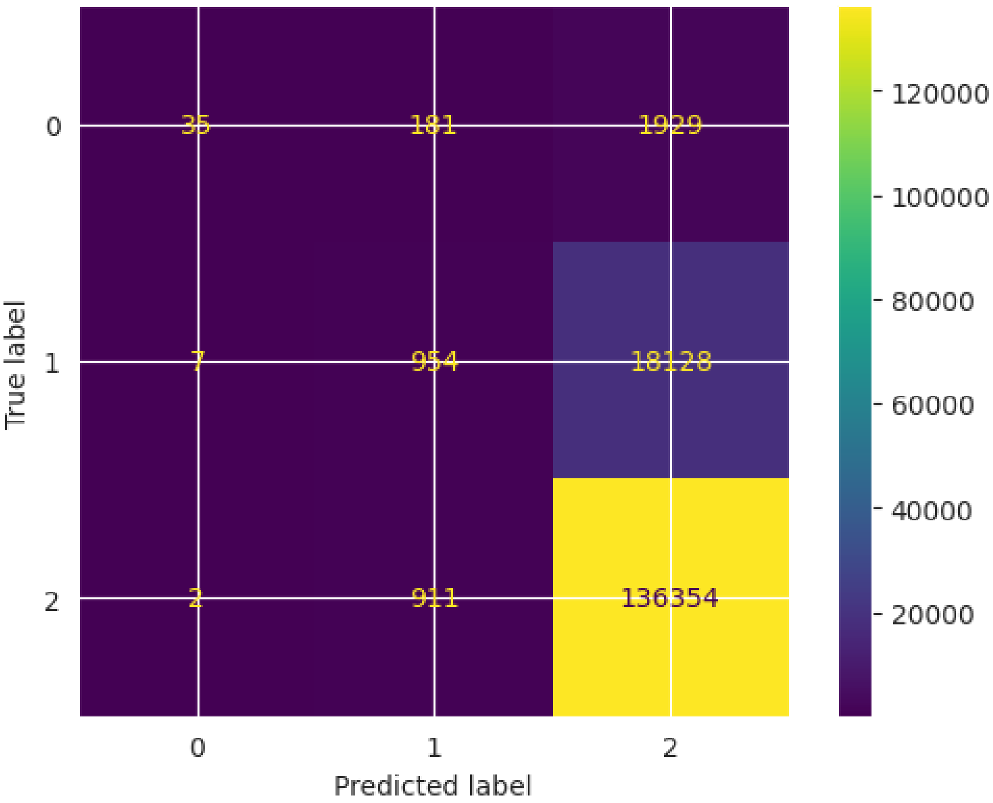
```
842:    learn: 0.3636845        total: 4m 30s   remaining: 50.4s
843:    learn: 0.3636724        total: 4m 30s   remaining: 50.1s
844:    learn: 0.3636388        total: 4m 31s   remaining: 49.8s
845:    learn: 0.3636127        total: 4m 31s   remaining: 49.4s
846:    learn: 0.3635764        total: 4m 31s   remaining: 49.1s
847:    learn: 0.3635607        total: 4m 32s   remaining: 48.8s
848:    learn: 0.3635447        total: 4m 32s   remaining: 48.4s
849:    learn: 0.3635189        total: 4m 32s   remaining: 48.1s
850:    learn: 0.3635044        total: 4m 32s   remaining: 47.8s
851:    learn: 0.3634900        total: 4m 33s   remaining: 47.4s
852:    learn: 0.3634660        total: 4m 33s   remaining: 47.1s
853:    learn: 0.3634374        total: 4m 33s   remaining: 46.8s
854:    learn: 0.3633958        total: 4m 34s   remaining: 46.6s
855:    learn: 0.3633686        total: 4m 35s   remaining: 46.3s
856:    learn: 0.3633413        total: 4m 35s   remaining: 46s
857:    learn: 0.3633304        total: 4m 36s   remaining: 45.7s
858:    learn: 0.3633252        total: 4m 36s   remaining: 45.4s
859:    learn: 0.3632992        total: 4m 36s   remaining: 45.1s
860:    learn: 0.3632725        total: 4m 37s   remaining: 44.8s
861:    learn: 0.3632549        total: 4m 37s   remaining: 44.5s
862:    learn: 0.3632205        total: 4m 38s   remaining: 44.2s
863:    learn: 0.3632013        total: 4m 38s   remaining: 43.9s
864:    learn: 0.3631851        total: 4m 38s   remaining: 43.5s
865:    learn: 0.3631514        total: 4m 39s   remaining: 43.2s
866:    learn: 0.3631371        total: 4m 39s   remaining: 42.9s
867:    learn: 0.3631176        total: 4m 39s   remaining: 42.5s
868:    learn: 0.3630920        total: 4m 39s   remaining: 42.2s
869:    learn: 0.3630812        total: 4m 40s   remaining: 41.9s
870:    learn: 0.3630408        total: 4m 40s   remaining: 41.5s
871:    learn: 0.3630094        total: 4m 40s   remaining: 41.2s
872:    learn: 0.3629873        total: 4m 41s   remaining: 40.9s
873:    learn: 0.3629727        total: 4m 41s   remaining: 40.6s
874:    learn: 0.3629518        total: 4m 41s   remaining: 40.2s
875:    learn: 0.3629264        total: 4m 41s   remaining: 39.9s
876:    learn: 0.3629112        total: 4m 42s   remaining: 39.6s
877:    learn: 0.3628994        total: 4m 42s   remaining: 39.2s
878:    learn: 0.3628809        total: 4m 42s   remaining: 38.9s
879:    learn: 0.3628621        total: 4m 43s   remaining: 38.6s
880:    learn: 0.3628505        total: 4m 43s   remaining: 38.3s
881:    learn: 0.3628373        total: 4m 43s   remaining: 37.9s
882:    learn: 0.3628248        total: 4m 43s   remaining: 37.6s
883:    learn: 0.3628093        total: 4m 44s   remaining: 37.3s
884:    learn: 0.3627882        total: 4m 44s   remaining: 36.9s
885:    learn: 0.3627729        total: 4m 44s   remaining: 36.6s
886:    learn: 0.3627518        total: 4m 44s   remaining: 36.3s
887:    learn: 0.3627226        total: 4m 45s   remaining: 36s
888:    learn: 0.3627118        total: 4m 45s   remaining: 35.6s
889:    learn: 0.3626941        total: 4m 45s   remaining: 35.3s
```

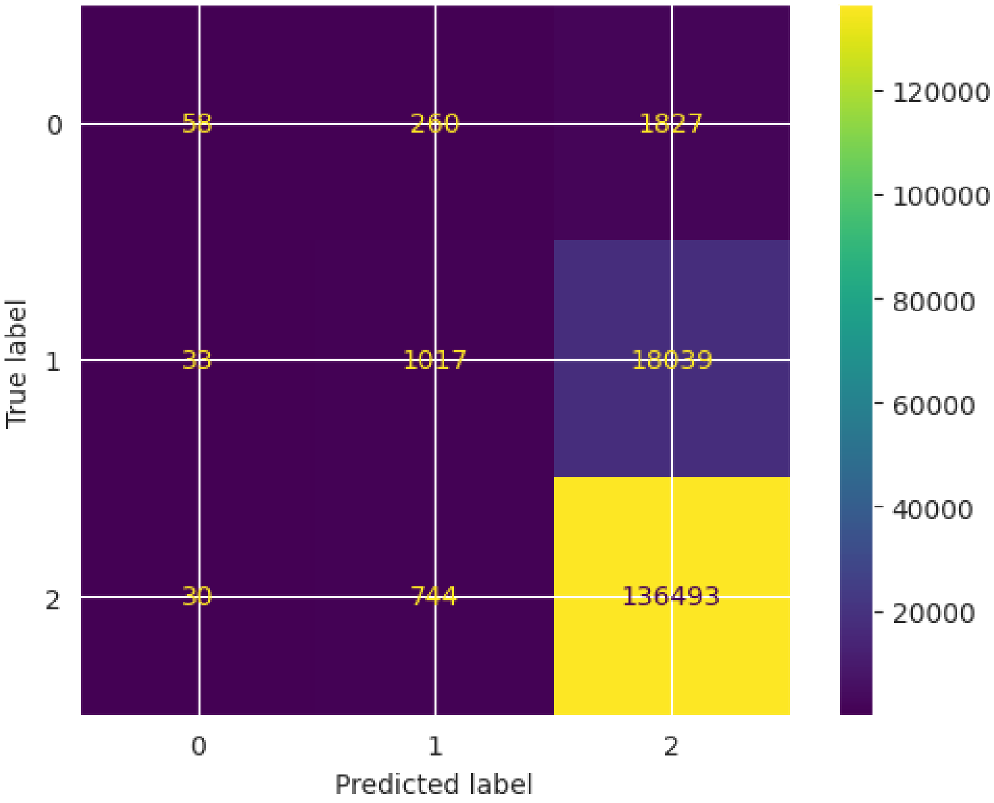
```
890:    learn: 0.3626773        total: 4m 45s   remaining: 35s
891:    learn: 0.3626591        total: 4m 46s   remaining: 34.7s
892:    learn: 0.3626364        total: 4m 46s   remaining: 34.3s
893:    learn: 0.3626306        total: 4m 46s   remaining: 34s
894:    learn: 0.3626194        total: 4m 46s   remaining: 33.7s
895:    learn: 0.3625964        total: 4m 47s   remaining: 33.3s
896:    learn: 0.3625678        total: 4m 47s   remaining: 33s
897:    learn: 0.3625604        total: 4m 47s   remaining: 32.7s
898:    learn: 0.3625438        total: 4m 48s   remaining: 32.4s
899:    learn: 0.3625305        total: 4m 48s   remaining: 32s
900:    learn: 0.3625096        total: 4m 48s   remaining: 31.7s
901:    learn: 0.3624825        total: 4m 48s   remaining: 31.4s
902:    learn: 0.3624577        total: 4m 49s   remaining: 31.1s
903:    learn: 0.3624393        total: 4m 49s   remaining: 30.8s
904:    learn: 0.3624158        total: 4m 50s   remaining: 30.5s
905:    learn: 0.3623841        total: 4m 50s   remaining: 30.2s
906:    learn: 0.3623639        total: 4m 51s   remaining: 29.9s
907:    learn: 0.3623524        total: 4m 51s   remaining: 29.6s
908:    learn: 0.3623277        total: 4m 52s   remaining: 29.3s
909:    learn: 0.3623109        total: 4m 52s   remaining: 29s
910:    learn: 0.3622833        total: 4m 53s   remaining: 28.6s
911:    learn: 0.3622578        total: 4m 53s   remaining: 28.3s
912:    learn: 0.3622273        total: 4m 54s   remaining: 28s
913:    learn: 0.3622007        total: 4m 54s   remaining: 27.7s
914:    learn: 0.3621852        total: 4m 54s   remaining: 27.4s
915:    learn: 0.3621634        total: 4m 54s   remaining: 27.1s
916:    learn: 0.3621223        total: 4m 55s   remaining: 26.7s
917:    learn: 0.3620982        total: 4m 55s   remaining: 26.4s
918:    learn: 0.3620851        total: 4m 55s   remaining: 26.1s
919:    learn: 0.3620636        total: 4m 56s   remaining: 25.7s
920:    learn: 0.3620154        total: 4m 56s   remaining: 25.4s
921:    learn: 0.3619902        total: 4m 56s   remaining: 25.1s
922:    learn: 0.3619686        total: 4m 56s   remaining: 24.8s
923:    learn: 0.3619316        total: 4m 57s   remaining: 24.4s
924:    learn: 0.3619085        total: 4m 57s   remaining: 24.1s
925:    learn: 0.3618938        total: 4m 57s   remaining: 23.8s
926:    learn: 0.3618673        total: 4m 58s   remaining: 23.5s
927:    learn: 0.3618341        total: 4m 58s   remaining: 23.1s
928:    learn: 0.3618081        total: 4m 58s   remaining: 22.8s
929:    learn: 0.3617840        total: 4m 58s   remaining: 22.5s
930:    learn: 0.3617659        total: 4m 59s   remaining: 22.2s
931:    learn: 0.3617492        total: 4m 59s   remaining: 21.8s
932:    learn: 0.3617243        total: 4m 59s   remaining: 21.5s
933:    learn: 0.3617008        total: 4m 59s   remaining: 21.2s
934:    learn: 0.3616820        total: 5m       remaining: 20.9s
935:    learn: 0.3616614        total: 5m       remaining: 20.5s
936:    learn: 0.3616508        total: 5m       remaining: 20.2s
937:    learn: 0.3616264        total: 5m 1s    remaining: 19.9s
```

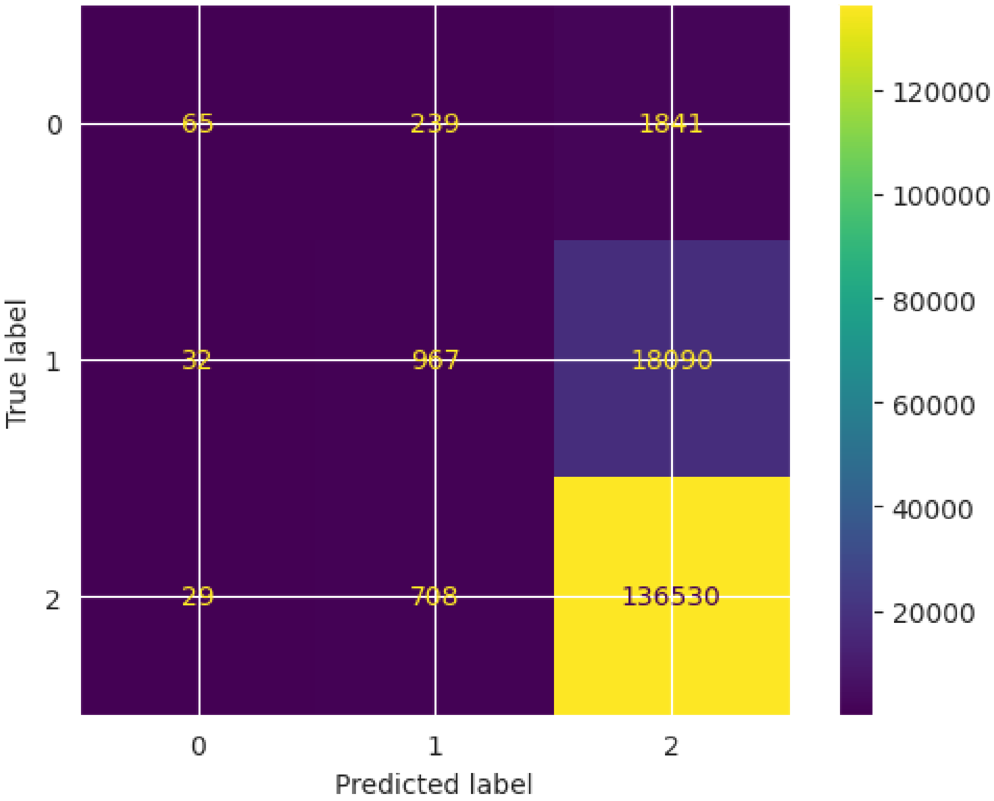
```
938:    learn: 0.3616118        total: 5m 1s    remaining: 19.6s
939:    learn: 0.3615991        total: 5m 1s    remaining: 19.2s
940:    learn: 0.3615857        total: 5m 1s    remaining: 18.9s
941:    learn: 0.3615686        total: 5m 2s    remaining: 18.6s
942:    learn: 0.3615503        total: 5m 2s    remaining: 18.3s
943:    learn: 0.3615368        total: 5m 2s    remaining: 18s
944:    learn: 0.3615243        total: 5m 2s    remaining: 17.6s
945:    learn: 0.3615118        total: 5m 3s    remaining: 17.3s
946:    learn: 0.3614972        total: 5m 3s    remaining: 17s
947:    learn: 0.3614820        total: 5m 3s    remaining: 16.7s
948:    learn: 0.3614625        total: 5m 3s    remaining: 16.3s
949:    learn: 0.3614413        total: 5m 4s    remaining: 16s
950:    learn: 0.3614203        total: 5m 4s    remaining: 15.7s
951:    learn: 0.3613974        total: 5m 5s    remaining: 15.4s
952:    learn: 0.3613868        total: 5m 5s    remaining: 15.1s
953:    learn: 0.3613656        total: 5m 6s    remaining: 14.8s
954:    learn: 0.3613543        total: 5m 6s    remaining: 14.4s
955:    learn: 0.3613363        total: 5m 6s    remaining: 14.1s
956:    learn: 0.3613117        total: 5m 7s    remaining: 13.8s
957:    learn: 0.3613007        total: 5m 7s    remaining: 13.5s
958:    learn: 0.3612863        total: 5m 8s    remaining: 13.2s
959:    learn: 0.3612569        total: 5m 8s    remaining: 12.9s
960:    learn: 0.3612430        total: 5m 9s    remaining: 12.6s
961:    learn: 0.3612247        total: 5m 9s    remaining: 12.2s
962:    learn: 0.3612021        total: 5m 9s    remaining: 11.9s
963:    learn: 0.3611688        total: 5m 10s   remaining: 11.6s
964:    learn: 0.3611448        total: 5m 10s   remaining: 11.3s
965:    learn: 0.3611262        total: 5m 10s   remaining: 10.9s
966:    learn: 0.3611213        total: 5m 10s   remaining: 10.6s
967:    learn: 0.3610963        total: 5m 11s   remaining: 10.3s
968:    learn: 0.3610740        total: 5m 11s   remaining: 9.96s
969:    learn: 0.3610433        total: 5m 11s   remaining: 9.64s
970:    learn: 0.3610299        total: 5m 12s   remaining: 9.32s
971:    learn: 0.3610176        total: 5m 12s   remaining: 9s
972:    learn: 0.3609870        total: 5m 12s   remaining: 8.68s
973:    learn: 0.3609685        total: 5m 12s   remaining: 8.35s
974:    learn: 0.3609519        total: 5m 13s   remaining: 8.03s
975:    learn: 0.3609279        total: 5m 13s   remaining: 7.71s
976:    learn: 0.3609122        total: 5m 13s   remaining: 7.39s
977:    learn: 0.3608819        total: 5m 14s   remaining: 7.06s
978:    learn: 0.3608734        total: 5m 14s   remaining: 6.74s
979:    learn: 0.3608587        total: 5m 14s   remaining: 6.42s
980:    learn: 0.3608390        total: 5m 14s   remaining: 6.1s
981:    learn: 0.3608125        total: 5m 15s   remaining: 5.78s
982:    learn: 0.3607881        total: 5m 15s   remaining: 5.45s
983:    learn: 0.3607708        total: 5m 15s   remaining: 5.13s
984:    learn: 0.3607366        total: 5m 15s   remaining: 4.81s
985:    learn: 0.3607136        total: 5m 16s   remaining: 4.49s
```

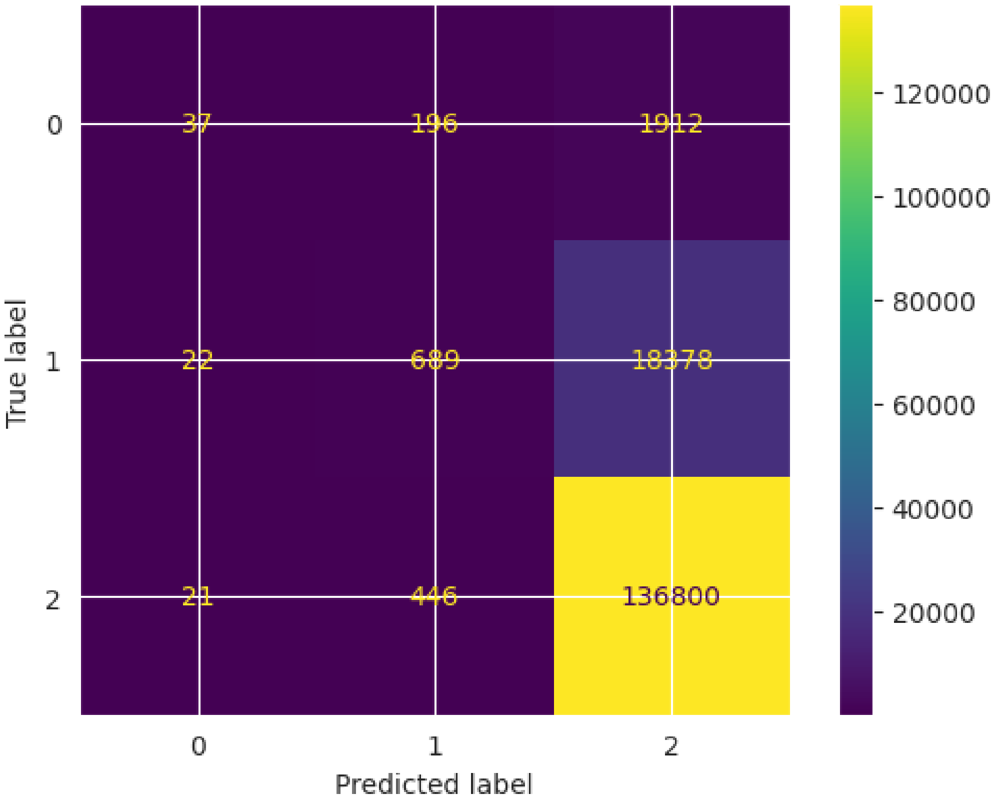
```
986:    learn: 0.3607006        total: 5m 16s   remaining: 4.17s
987:    learn: 0.3606760        total: 5m 16s   remaining: 3.85s
988:    learn: 0.3606591        total: 5m 17s   remaining: 3.53s
989:    learn: 0.3606451        total: 5m 17s   remaining: 3.2s
990:    learn: 0.3606278        total: 5m 17s   remaining: 2.88s
991:    learn: 0.3606097        total: 5m 17s   remaining: 2.56s
992:    learn: 0.3605937        total: 5m 18s   remaining: 2.24s
993:    learn: 0.3605682        total: 5m 18s   remaining: 1.92s
994:    learn: 0.3605511        total: 5m 18s   remaining: 1.6s
995:    learn: 0.3605338        total: 5m 18s   remaining: 1.28s
996:    learn: 0.3605216        total: 5m 19s   remaining: 960ms
997:    learn: 0.3604977        total: 5m 19s   remaining: 640ms
998:    learn: 0.3604823        total: 5m 19s   remaining: 320ms
999:    learn: 0.3604559        total: 5m 20s   remaining: 0us
              precision    recall  f1-score   support

           0       0.52      0.03      0.06      2145
           1       0.51      0.05      0.09     19089
           2       0.87      0.99      0.93    137267

    accuracy                           0.87    158501
   macro avg       0.63      0.36      0.36    158501
weighted avg       0.82      0.87      0.82    158501

Accuracy score of <catboost.core.CatBoostClassifier object at 0x7901055cf880>
0.8678935779584986
LGBMClassifier(force_row_wise=True)
---------------------------------------------------------------------------
[LightGBM] [Info] Total Bins 2377
[LightGBM] [Info] Number of data points in the train set: 369834, number of used
features: 43
[LightGBM] [Info] Start training from score -4.282638
[LightGBM] [Info] Start training from score -2.110311
[LightGBM] [Info] Start training from score -0.145033
              precision    recall  f1-score   support

           0       0.46      0.02      0.03      2145
           1       0.52      0.04      0.07     19089
           2       0.87      1.00      0.93    137267

    accuracy                           0.87    158501
   macro avg       0.62      0.35      0.34    158501
weighted avg       0.82      0.87      0.81    158501

Accuracy score of LGBMClassifier(force_row_wise=True)   0.8676664500539428
KNeighborsClassifier(n_neighbors=7)
---------------------------------------------------------------------------
              precision    recall  f1-score   support
           0       0.27      0.02      0.03      2145
           1       0.32      0.06      0.11     19089
           2       0.87      0.98      0.92    137267

    accuracy                           0.86    158501
   macro avg       0.49      0.35      0.35    158501
weighted avg       0.80      0.86      0.81    158501

Accuracy score of KNeighborsClassifier(n_neighbors=7)   0.8588589346439455
BernoulliNB()
---------------------------------------------------------------------------
              precision    recall  f1-score   support

           0       0.08      0.16      0.11      2145
           1       0.34      0.03      0.06     19089
           2       0.87      0.97      0.92    137267

    accuracy                           0.85    158501
   macro avg       0.43      0.39      0.36    158501
weighted avg       0.80      0.85      0.81    158501

Accuracy score of BernoulliNB()   0.8462280995072586
```

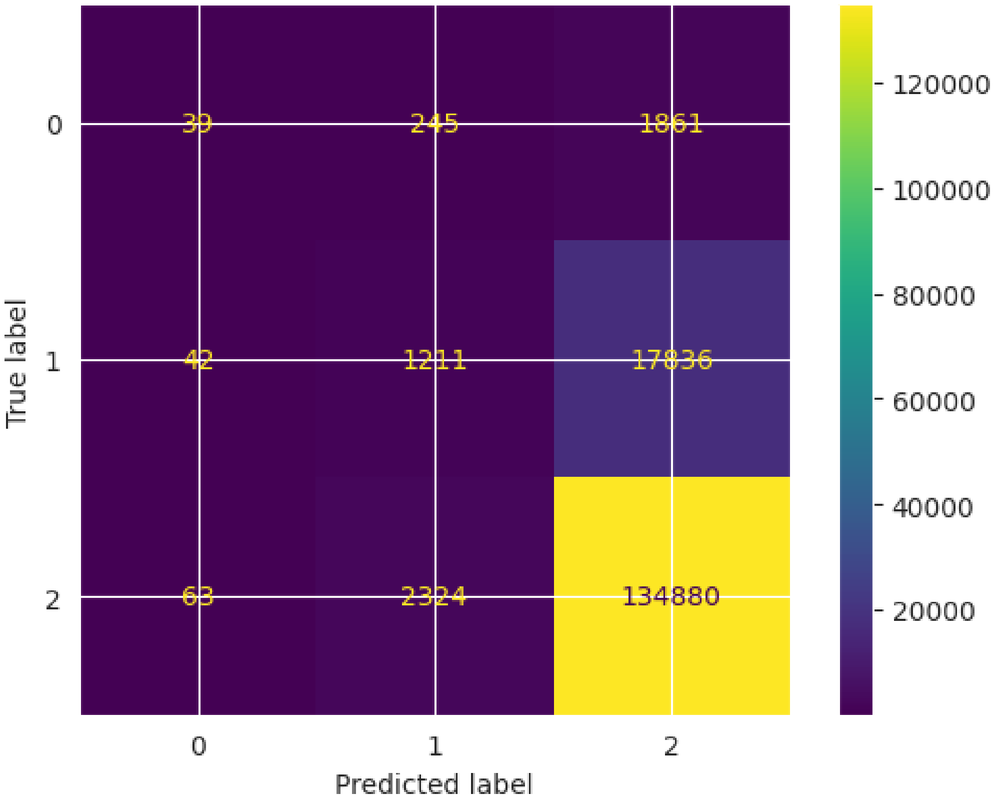


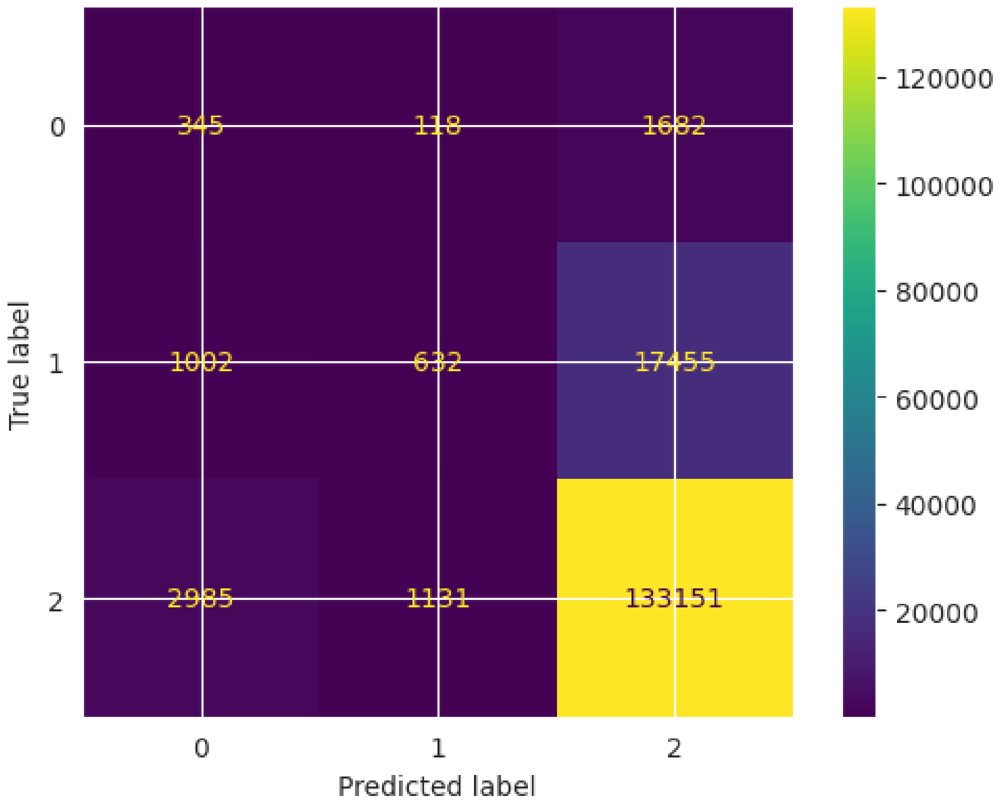

0
1
2
True label
35
181
1929
7
954
18128
2
911
136354
0
1
2
Predicted label
120000
100000
80000
60000
40000
20000


True label
0
1
2
58
260
1827
33
1017
18039
30
744
136493
0
1
2
Predicted label
120000
100000
80000
60000
40000
20000


0
1
2
True label
65
239
1841
32
967
18090
29
708
136530
0
1
2
Predicted label
120000
100000
80000
60000
40000
20000


0
1
2
True label
37
196
1912
22
689
18378
21
446
136800
0
1
2
Predicted label
120000
100000
80000
60000
40000
20000


0
1
2
True label
39
245
1861
42
1211
17836
63
2324
134880
0
1
2
Predicted label
120000
100000
80000
60000
40000
20000

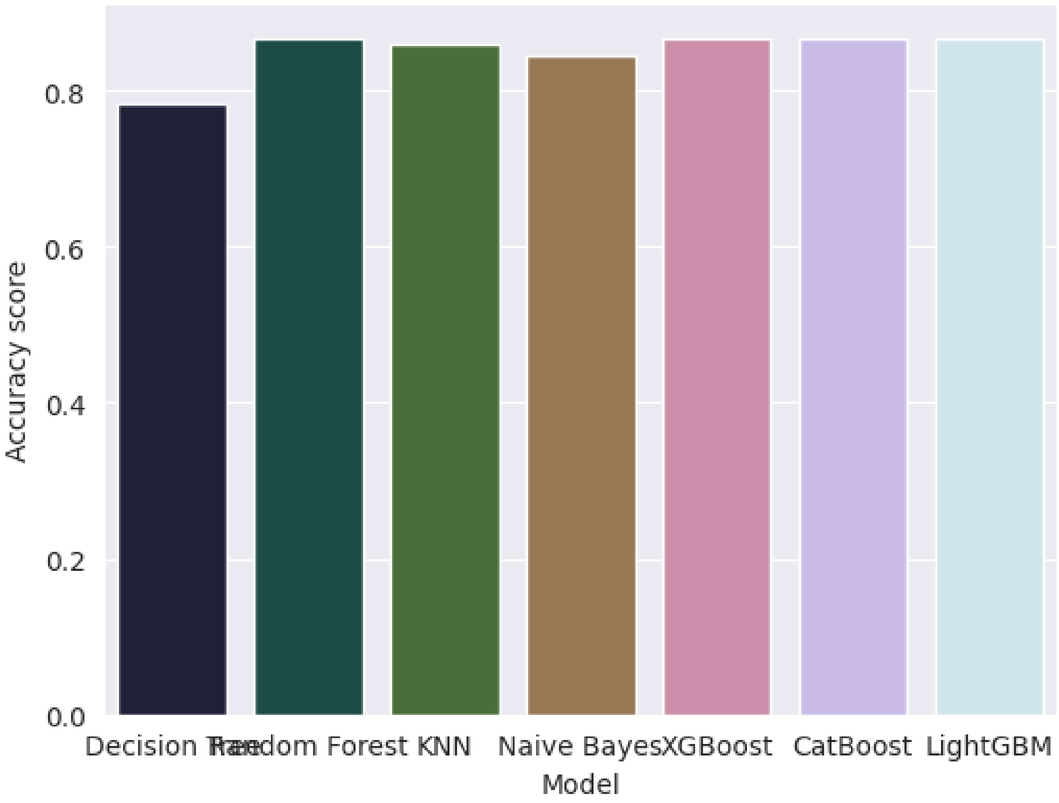


```
[186]: dic={'Model':["Decision Tree","Random Forest","KNN","Naive␣
       ↪Bayes","XGBoost","CatBoost","LightGBM"],'Accuracy score':[0.7836,0.8665,0.
       ↪8588,0.8462,0.8679,0.8678,0.8676]}
       result=pd.DataFrame(dic)
       result
```

```
[186]:            Model  Accuracy score
       0  Decision Tree          0.7836
       1  Random Forest          0.8665
       2            KNN          0.8588
       3    Naive Bayes          0.8462
       4        XGBoost          0.8679
       5       CatBoost          0.8678
       6       LightGBM          0.8676
```

```
[187]: sns.barplot(x='Model',y='Accuracy score',data=dic,ci=None,palette='cubehelix')
```

```
[187]: <Axes: xlabel='Model', ylabel='Accuracy score'>
```


0.0
0.2
0.4
0.6
0.8
Accuracy score
Decision Tree
Random Forest
KNN
Naive Bayes
XGBoost
CatBoost
LightGBM
Model