Unique values in Dataset

In [1]:
```
# Imports
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd

# reading and printing all the columns
df = pd.read_csv('F_PoliceIncident.csv')
pd.set_option('display.max_columns', None)
#pd.set_option('display.max_rows', None)
df.head()
```

```
C:\Users\jules\anaconda3\lib\site-packages\IPython\core\interactiveshell.py:344
4: DtypeWarning: Columns (1) have mixed types.Specify dtype option on import or
set low_memory=False.
  exec(code_obj, self.user_global_ns, self.user_ns)
```

Out[1]:

	Year of Incident	Watch	Call (911) Problem	Type of Incident	Type of Location	Reporting Area	Beat	Division
0	2021	3	15 - ASSIST OFFICER	ASSAULT -PUB SERV (PEACE OFFICER/JUDGE)	Commercial Property Occupied/Vacant	4381.0	745.0	SOUTH CENTRAL
1	2021	3	15 - ASSIST OFFICER	THEFT OF PROP <$2,500 2+PREV CONV (SHOPLFT-NOT...	Commercial Property Occupied/Vacant	4381.0	745.0	SOUTH CENTRAL
2	2019	1	DASV-DIST ACTIVE SHOOTER VEH	PUBLIC INTOXICATION	Highway, Street, Alley ETC	4042.0	424.0	SOUTHWEST
3	2020	3	6X - MAJOR DIST (VIOLENCE)	ASSAULT - BODILY INJURY ONLY	Gas or Service Station	4167.0	446.0	SOUTHWEST
4	2016	2	20 - ROBBERY	ROBBERY OF INDIVIDUAL (AGG)	Single Family Residence - Occupied	4308.0	733.0	SOUTH CENTRAL

In [2]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 848467 entries, 0 to 848466
Data columns (total 20 columns):
 #   Column                   Non-Null Count   Dtype  
---  ------                   --------------   -----  
 0   Year of Incident         848467 non-null  int64  
 1   Watch                    848467 non-null  object 
 2   Call (911) Problem       809734 non-null  object 
 3   Type of Incident         848467 non-null  object 
 4   Type of Location         847324 non-null  object 
 5   Reporting Area           847682 non-null  float64
 6   Beat                     848333 non-null  float64
 7   Division                 848333 non-null  object 
 8   Sector                   848282 non-null  float64
 9   Council District         846785 non-null  object 
 10  Year of Occurrence       848467 non-null  int64  
 11  Month of Occurence       848467 non-null  object 
 12  Day of the Week          848467 non-null  object 
 13  Time of Occurrence       848467 non-null  object 
 14  Person Involvement Type  817550 non-null  object 
 15  Victim Type              809575 non-null  object 
 16  Hate Crime Description   847270 non-null  object 
 17  Drug Related Incident    810528 non-null  object 
 18  Penal Code               848467 non-null  object 
 19  Zip Code                 845036 non-null  float64
dtypes: float64(4), int64(2), object(14)
memory usage: 129.5+ MB
```

Call (911) Problem

Type of Incident

Type Location

Council

In [3]:
```
df[df['Council District'] == 1]
```

Out[3]:

Year of Incident	Watch	Call (911) Problem	Type of Incident	Type of Location	Reporting Area	Beat	Division	Sector	Council District	Ye Occurr

In [4]:
```
group = df.groupby('Council District')
```

Renaming 1,8,9,10,11 to D1,D8,D9,D10,D11

In [5]:
```
councilDistrict = group.apply(lambda x: x['Sector'].unique())
```

In [6]:
```
councilDistrict
```

Out[6]:
```
Council District
1                                          [410.0, 440.0]
10                                  [230.0, 250.0, 240.0]
11            [620.0, 630.0, 250.0, 640.0, 650.0, 110.0]
8                    [350.0, 450.0, 750.0, 740.0, 730.0]
9                                   [220.0, 230.0, 110.0]
D1                          [440.0, 410.0, 430.0, 130.0]
D10                         [250.0, 240.0, 230.0, 210.0]
D11                  [630.0, 650.0, 620.0, 640.0, 250.0]
D12                                        [610.0, 620.0]
D13    [210.0, 650.0, 550.0, 640.0, 520.0, 530.0, 240.0]
D14    [150.0, 210.0, 130.0, 120.0, 540.0, 140.0, 110.0]
D2     [130.0, 520.0, 540.0, 510.0, 110.0, 150.0, 120...
D3     [430.0, 440.0, 450.0, 740.0, 410.0, 750.0, 730...
D4     [730.0, 450.0, 710.0, 720.0, 440.0, 740.0, 750...
D5                          [330.0, 320.0, 350.0, 310.0]
D6     [420.0, 520.0, 510.0, 550.0, 410.0, 530.0, 430...
D7     [340.0, 220.0, 350.0, 310.0, 320.0, 110.0, 230...
D8                   [740.0, 730.0, 750.0, 450.0, 350.0]
D9                          [230.0, 220.0, 240.0, 210.0]
dtype: object
```

In [7]:
```
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)
```

In [8]:
```
councilDistrict['D2']
```

Out[8]:
```
array([130., 520., 540., 510., 110., 150., 120., 140., 530., 310.])
```

In [9]:
```
councilDistrict['D3']
```

Out[9]:
```
array([430., 440., 450., 740., 410., 750., 730., 720., 420.])
```

In [10]:
```
councilDistrict['D4']
```

Out[10]:
```
array([730., 450., 710., 720., 440., 740., 750., 150.])
```

In [11]:
```
councilDistrict['D6']
```

Out[11]:
```
array([420., 520., 510., 550., 410., 530., 430., 130.])
```

In [12]:
```
councilDistrict['D7']
```

Out[12]:
```
array([340., 220., 350., 310., 320., 110., 230., 150., 730., 710., 330.])
```

In [ ]:

In [13]:
```
df2 = df.groupby('Council District').apply(lambda x: x['Division'].unique())
```

In [14]:
```
df2
```

```
1                                          [SouthWest]
10                                          [NorthEast]
11                [North Central, NorthEast, Central]
8               [SouthEast, SouthWest, South Central]
9                                 [NorthEast, Central]
D1                                [SOUTHWEST, CENTRAL]
D10                                         [NORTHEAST]
D11                         [NORTH CENTRAL, NORTHEAST]
D12                                   [NORTH CENTRAL]
D13             [NORTHEAST, NORTH CENTRAL, NORTHWEST]
D14                   [CENTRAL, NORTHEAST, NORTHWEST]
D2                    [CENTRAL, NORTHWEST, SOUTHEAST]
D3                         [SOUTHWEST, SOUTH CENTRAL]
D4                [SOUTH CENTRAL, SOUTHWEST, CENTRAL]
D5                                          [SOUTHEAST]
D6                    [SOUTHWEST, NORTHWEST, CENTRAL]
D7     [SOUTHEAST, NORTHEAST, CENTRAL, SOUTH CENTRAL]
D8              [SOUTH CENTRAL, SOUTHWEST, SOUTHEAST]
D9                                          [NORTHEAST]
dtype: object
```

Beat and council district

In [17]:
```
df2 = df.groupby('Council District').apply(lambda x: x['Beat'].unique())
```

In [70]:
```
df2
```

Out[70]:
```
Council District
1              [411.0, 413.0, 414.0, 443.0, 444.0, 417.0]
10     [231.0, 251.0, 252.0, 253.0, 248.0, 247.0, 243...
11     [625.0, 633.0, 251.0, 644.0, 641.0, 635.0, 653...
8      [356.0, 451.0, 756.0, 745.0, 736.0, 355.0, 456...
9             [229.0, 235.0, 233.0, 227.0, 111.0, 232.0]
D1     [446.0, 411.0, 417.0, 432.0, 414.0, 413.0, 441...
D10    [255.0, 252.0, 245.0, 257.0, 246.0, 253.0, 244...
D11    [631.0, 651.0, 632.0, 624.0, 642.0, 625.0, 634...
D12    [614.0, 623.0, 613.0, 611.0, 622.0, 612.0, 621...
D13    [211.0, 216.0, 214.0, 652.0, 552.0, 641.0, 213...
D14    [155.0, 218.0, 131.0, 122.0, 546.0, 154.0, 121...
D2     [134.0, 523.0, 542.0, 514.0, 116.0, 151.0, 541...
D3     [436.0, 445.0, 433.0, 456.0, 742.0, 743.0, 435...
D4     [733.0, 453.0, 711.0, 734.0, 725.0, 717.0, 726...
D5     [331.0, 327.0, 336.0, 324.0, 328.0, 332.0, 334...
D6     [424.0, 521.0, 517.0, 516.0, 512.0, 551.0, 422...
D7     [344.0, 221.0, 346.0, 351.0, 341.0, 318.0, 322...
D8     [745.0, 744.0, 736.0, 752.0, 452.0, 757.0, 748...
D9     [234.0, 233.0, 227.0, 225.0, 236.0, 231.0, 228...
dtype: object
```

In [69]:
```
df2.to_csv('COuncilandBeat.csv', encoding='utf-8',sep=',', index=False)
```

In [73]:
```
df2 = df.groupby('Beat').apply(lambda x: x['Council District'].unique())
```

In [74]:
```
df2
```

Out[74]:
```
Beat
111.0           [D2, nan, 9, 11]
112.0             [D14, D2, nan]
113.0                  [D2, nan]
114.0             [D2, D14, nan]
115.0              [D7, D2, nan]
116.0                  [D2, nan]
121.0                 [D14, nan]
122.0             [D14, D2, nan]
123.0             [D2, D14, nan]
124.0                 [D14, nan]
125.0                 [D14, nan]
131.0             [D14, D2, nan]
132.0                 [D14, nan]
133.0             [D14, D2, nan]
134.0             [D2, D14, nan]
135.0                  [D2, nan]
136.0          [D2, D1, D6, nan]
141.0                 [D14, nan]
```

In [ ]:

In [ ]:

In [ ]:

Beat and Division

In [71]:
```
df2 = df.groupby('Beat').apply(lambda x: x['Division'].unique())
```

In [72]:
```
df2
```

```
714.0    [SOUTH CENTRAL, South Central]
715.0    [SOUTH CENTRAL, South Central]
716.0    [SOUTH CENTRAL, South Central]
717.0    [SOUTH CENTRAL, South Central]
721.0    [SOUTH CENTRAL, South Central]
722.0    [SOUTH CENTRAL, South Central]
723.0    [SOUTH CENTRAL, South Central]
724.0    [SOUTH CENTRAL, South Central]
725.0    [SOUTH CENTRAL, South Central]
726.0    [SOUTH CENTRAL, South Central]
727.0    [SOUTH CENTRAL, South Central]
728.0    [SOUTH CENTRAL, South Central]
731.0    [SOUTH CENTRAL, South Central]
732.0    [SOUTH CENTRAL, South Central]
733.0    [SOUTH CENTRAL, South Central]
734.0    [SOUTH CENTRAL, South Central]
735.0    [SOUTH CENTRAL, South Central]
736.0    [SOUTH CENTRAL, South Central]
737.0    [SOUTH CENTRAL, South Central]
741.0    [SOUTH CENTRAL, South Central]
```

Zip and Beat

In [77]:
```
df2 = df.groupby('Zip Code').apply(lambda x: x['Beat'].unique())
```

In [78]:
```
df2
```

Out[78]:
```
Zip Code
48232.0                                                [741.0]
57238.0                                                [215.0]
63145.0                                                [122.0]
72220.0                                                [521.0]
72231.0                                                [244.0]
73237.0                                                [745.0]
74204.0                                                [122.0]
74243.0                                                [243.0]
75001.0                     [151.0, nan, 625.0, 622.0, 623.0]
75006.0    [134.0, 315.0, 133.0, 237.0, 539.0, 613.0, nan...
75007.0                          [611.0, 612.0, 613.0, 553.0]
75010.0                                                 [nan]
75014.0                                                [125.0]
75016.0                                                [219.0]
75019.0              [539.0, 357.0, 356.0, 151.0, 233.0, nan]
75023.0                   [454.0, 432.0, 212.0, 455.0, 447.0]
75024.0                                        [524.0, nan]
75025.0                                                [151.0]
75026.0                                        [543.0, 218.0]
```

Beat and reporting area

In [83]:
```
df2 = df.groupby('Reporting Area').apply(lambda x: x['Beat'].unique())
```

In [84]:
```
df2
```

Out[84]:
```
Reporting Area
1001.0                                          [622.0]
1002.0                                          [624.0]
1003.0                                          [624.0]
1004.0                                          [622.0]
1005.0                                          [624.0]
1006.0                                          [625.0]
1007.0                                          [624.0]
1008.0                                [631.0, 724.0]
1009.0                                [632.0, 132.0]
1010.0                                       [632.0]
1011.0                                       [632.0]
1012.0                                       [632.0]
1013.0                                [635.0, 121.0]
1014.0                                       [635.0]
1015.0                                       [635.0]
1016.0                                       [635.0]
1017.0                                       [633.0]
1018.0                                       [635.0]
```

Beat and Sector

In [85]:
```
df2 = df.groupby('Sector').apply(lambda x: x['Beat'].unique())
```

In [86]:
```
df2
```

Out[86]:
```
Sector
110.0           [116.0, 114.0, 113.0, 115.0, 112.0, 111.0]
120.0                  [122.0, 121.0, 124.0, 123.0, 125.0]
130.0           [134.0, 131.0, 132.0, 136.0, 135.0, 133.0]
140.0           [143.0, 141.0, 142.0, 145.0, 146.0, 144.0]
150.0           [155.0, 151.0, 153.0, 154.0, 156.0, 152.0]
210.0    [211.0, 218.0, 216.0, 214.0, 219.0, 213.0, 212...
220.0    [221.0, 227.0, 229.0, 225.0, 223.0, 228.0, 226...
230.0    [234.0, 233.0, 231.0, 236.0, 232.0, 235.0, 237...
240.0    [245.0, 246.0, 244.0, 242.0, 248.0, 247.0, 243...
250.0    [255.0, 252.0, 257.0, 253.0, 251.0, 254.0, 256...
310.0    [318.0, 313.0, 311.0, 312.0, 317.0, 315.0, 314...
320.0    [327.0, 324.0, 328.0, 322.0, 326.0, 323.0, 325...
330.0    [331.0, 336.0, 332.0, 334.0, 335.0, 338.0, 337...
340.0    [344.0, 346.0, 341.0, 342.0, 347.0, 345.0, 343...
350.0    [351.0, 352.0, 357.0, 355.0, 354.0, 356.0, 353.0]
410.0    [411.0, 417.0, 414.0, 413.0, 416.0, 412.0, 415.0]
420.0           [424.0, 422.0, 423.0, 426.0, 421.0, 425.0]
430.0    [436.0, 433.0, 432.0, 435.0, 437.0, 431.0, 434.0]
440.0    [446.0, 445.0, 441.0, 442.0, 443.0, 444.0, 447.0]
450.0           [453.0, 452.0, 456.0, 451.0, 454.0, 455.0]
510.0           [517.0, 514.0, 516.0, 513.0, 512.0, 515.0]
520.0           [521.0, 523.0, 522.0, 524.0, 526.0, 525.0]
530.0    [538.0, 532.0, 534.0, 535.0, 536.0, 533.0, 537...
540.0           [542.0, 541.0, 546.0, 544.0, 543.0, 545.0]
550.0           [551.0, 552.0, 553.0, 554.0, 556.0, 555.0]
610.0                         [614.0, 613.0, 611.0, 612.0]
620.0                 [624.0, 623.0, 625.0, 622.0, 621.0]
630.0                 [631.0, 632.0, 634.0, 635.0, 633.0]
640.0                        [642.0, 644.0, 643.0, 641.0]
650.0                        [651.0, 652.0, 654.0, 653.0]
710.0    [711.0, 717.0, 716.0, 714.0, 713.0, 712.0, 715.0]
720.0    [725.0, 726.0, 724.0, 722.0, 723.0, 727.0, 728...
730.0    [733.0, 736.0, 734.0, 735.0, 737.0, 732.0, 731.0]
740.0    [745.0, 744.0, 742.0, 748.0, 746.0, 743.0, 741...
750.0    [752.0, 757.0, 754.0, 751.0, 755.0, 753.0, 756.0]
dtype: object
```

sECTOR AND coUNCIL

In [87]:
```
df2 = df.groupby('Sector').apply(lambda x: x['Council District'].unique())
```

In [88]:
```
df2
```

Out[88]:
```
Sector
110.0    [D2, D7, D14, nan, 9, 11]
120.0               [D14, D2, nan]
130.0       [D2, D14, D1, nan, D6]
140.0               [D14, D2, nan]
150.0       [D14, D2, D7, nan, D4]
210.0     [D13, D14, D9, D10, nan]
220.0              [D7, D9, 9, nan]
230.0    [D9, 10, D10, D7, nan, 9]
240.0      [D10, D9, D13, nan, 10]
250.0      [D10, D11, 10, nan, 11]
310.0            [D7, D5, D2, nan]
320.0                [D5, D7, nan]
330.0                [D5, D7, nan]
340.0                    [D7, nan]
350.0         [D7, D8, D5, 8, nan]
410.0         [D1, D6, D3, nan, 1]
420.0                [D6, D3, nan]
430.0           [D3, D1, nan, D6]
440.0         [D1, D3, D4, nan, 1]
450.0         [D4, D8, D3, 8, nan]
510.0                [D6, D2, nan]
520.0           [D6, D2, D13, nan]
530.0           [D2, D6, D13, nan]
540.0               [D2, D14, nan]
550.0               [D6, D13, nan]
610.0                   [D12, nan]
620.0          [D11, D12, 11, nan]
630.0               [D11, 11, nan]
640.0          [D11, D13, nan, 11]
650.0          [D11, D13, nan, 11]
710.0                [D4, D7, nan]
720.0                [D4, D3, nan]
730.0     [D4, D8, D3, D7, nan, 8]
740.0         [D8, D3, D4, nan, 8]
750.0         [D8, D4, D3, 8, nan]
dtype: object
```

In [ ]:
```

```

In [91]:
```
df2 = df.groupby('Zip Code').apply(lambda x: x['Council District'].unique())
```

In [92]:
```
df2
```

```
75203.0                    [D4, D1, nan, D2, D7, D6, D8, 1]
75204.0                            [D14, D2, nan, D12, D10]
75205.0                             [D14, D13, nan, D2, D5]
75206.0                [D14, D2, D9, nan, D13, D7, D10, D11]
75207.0                [D2, D6, D1, D14, nan, D5, 1, D3, D4]
75208.0           [D6, D1, D4, D7, D3, nan, 1, D5, D12, D11]
75209.0                             [D13, D2, nan, D7, D10]
75210.0                  [D7, nan, D2, D5, D14, D8, D6, D13]
75211.0           [D3, D1, D6, D2, nan, D7, D4, D14, D8, D5]
75212.0                   [D6, D3, D4, nan, D7, D14, D1, D2]
75213.0                                        [D2, D4, D10]
75214.0     [D2, D14, D9, nan, D6, D7, D10, D8, D5, D13, D4]
75215.0    [D7, D2, D5, D8, D3, nan, D4, D14, D10, D1, D1...
75216.0    [D4, D2, D3, D8, D7, D11, nan, D14, D1, 8, D10...
75217.0       [D5, D8, D7, D14, D2, nan, 8, D10, D4, D3, D1]
75218.0          [D9, D2, D6, D14, nan, 9, D10, D13, D5, D7]
75219.0                        [D14, D2, nan, D6, D13, D12]
75220.0                  [D6, nan, D2, D13, D8, D5, D9, D11]
75221.0                                                [D2]
75223.0              [D2, D7, D14, nan, D1, D5, D8, D13, D4]
```