```
In [18]: import pandas as pd
           import numpy as np
           import matplotlib.pyplot as plt
In [19]: | df = pd.read_csv('./Seattle_Police.csv')
           1 - CAD CDW ID
In [20]: m = df.loc[df['CAD CDW ID'].isna()]
           df['CAD CDW ID'][m.index]
Out[20]: 144 NaN
                 NaN
           145
           146 NaN
           147 NaN
           148 NaN
           149 NaN
           Name: CAD CDW ID, dtype: float64
           Lets look at the values above and below these NaN values
In [21]: df['CAD CDW ID'][142:152]
Out[21]: 142
                   15882.0
                   15883.0
           143
           144
                        NaN
           145
                        NaN
           146
                        NaN
           147
                        NaN
           148
                        NaN
           149
                        NaN
           150
                  15890.0
                   15891.0
           151
           Name: CAD CDW ID, dtype: float64
           These values are in a range with every value increasing by 1. Hence I will use the interpolate method to fill the values
In [22]: | df['CAD CDW ID'] = df['CAD CDW ID'].interpolate()
In [23]: df['CAD CDW ID'][142:152] # As you can see they are all filled now. Lets look for more erroneous data.
Out[23]: 142
                   15882.0
           143
                   15883.0
           144 15884.0
           145
                 15885.0
           146
                  15886.0
           147
                   15887.0
           148
                   15888.0
           149
                   15889.0
           150
                   15890.0
           151
                  15891.0
           Name: CAD CDW ID, dtype: float64
           Looking for values that are out of range or erroneous. e.g values with more than 7 digits
In [24]: | df['CAD CDW ID'] = df['CAD CDW ID'].astype(str)
In [25]: df['CAD CDW ID'].str.len().value_counts() # These values don't match the pattern(>7 digits) and are erroneous value
           S.
Out[25]: 7
                 9864
                 136
           Name: CAD CDW ID, dtype: int64
In [26]: #Lets set these values to NaN first
           a = df[df['CAD CDW ID'].str.len()>7]
           df['CAD CDW ID'].iloc[a.index] = np.nan
In [27]: # Replaced all with NaN
           df['CAD CDW ID'][a.index]
Out[27]: 48
                     NaN
                     NaN
           104
                     NaN
           189
                     NaN
           254
                    NaN
           9854
                    NaN
           9855
                    NaN
           9861
                    NaN
           9917
                    NaN
           9994
           Name: CAD CDW ID, Length: 136, dtype: object
In [28]: df['CAD CDW ID'] = df['CAD CDW ID'].astype(float)
In [29]: # Updating the NaN by adding 1 to their previous value
           for i in a.index:
                df['CAD CDW ID'].iloc[i] = (df['CAD CDW ID'].iloc[i-1])+1
In [30]: df['CAD CDW ID'][a.index]
Out[30]: 48
                    15786.0
           69
                    15807.0
           104
                    15842.0
           189
                    15930.0
                    15997.0
                      . . .
           9854
                    25780.0
           9855
                    25781.0
           9861
                    25787.0
           9917
                    25845.0
           9994
                    25923.0
           Name: CAD CDW ID, Length: 136, dtype: float64
In [31]: df['CAD CDW ID'].isna().sum() # RECHECK
Out[31]: 0
           2 CAD Event Number
           1 - Copied (bfill) the value from General Offense Number to Null values in CAD event Number.
           2 - Replacing "2016" with 10000 and "2010" with 10000 in the column to conform to the pattern. Using REGEX
In [32]: a2 = df[df['CAD Event Number'].isna()].head()
           a2
Out[32]:
                                                                                                                   Hundred
                           CAD
                                   General
                                                                                                         Event
                                               Event
                                                              Event
                                                                      Event Clearance
                                                                                      Event Clearance
                                                                                                      Clearance
                          Event
                                   Offense Clearance
                                                          Clearance
                                                                                                                     Block
                                                                                                                           District/Sector Zone/Beat
                CDW ID
                                                                           SubGroup
                                                                                               Group
                        Number
                                   Number
                                                Code
                                                         Description
                                                                                                          Date
                                                                                                                   Location
                                                                                                                 9XX BLOCK
                                                                                         SUSPICIOUS
                                                                                                        7/17/10
                                                        SUSPICIOUS
                                                                         SUSPICIOUS
            29 15766.0
                           NaN 2010246426
                                                280.0
                                                           PERSON CIRCUMSTANCES CIRCUMSTANCES
                                                                                                         20:46
                                                           PARKING
                                                          VIOLATION
                                                                            PARKING
                                                                                            TRAFFIC
                                                                                                        7/17/10
            30 15767.0
                           NaN 2010246213
                                                470.0
                                                           (EXCEPT
                                                                                                                OF TAYLOR
                                                                                                                                      D
                                                                         VIOLATIONS
                                                                                      RELATED CALLS
                                                                                                          20:48
                                                        ABANDONED
                                                         VEHICLES)
                                                                                                                 2XX BLOCK
                                                                                                        7/17/10
                                                      DISTURBANCE,
                                                                                                                       OF
            31 15768.0
                           NaN 2010246429
                                                                      DISTURBANCES DISTURBANCES
                                                                                                                  THOMAS
                                                            LIQUOR
                                                                                                                 1XX BLOCK
                                                                            LIQUOR
                                                                                             LIQUOR
                                                                                                        7/17/10
                                                        VIOLATION -
            32 15769.0
                           NaN 2010246463
                                                177.0
                                                       INTOXICATED
                                                                                          VIOLATIONS
                                                                                                                BROADWAY
                                                                         VIOLATIONS
                                                                                                          20:49
                                                           PERSON
                                                                                                                 3XX BLOCK
                                                                                                        7/17/10
                                                                                                                       OF
                           NaN 2010246305
                                                160.0
                                                          PROWLER
                                                                           PROWLER
                                                                                           PROWLER
            33 15770.0
                                                                                                                HARRISON
In [33]: cols = ['CAD Event Number', 'General Offense Number']
           df.loc[:,cols] = df.loc[:,cols].bfill(axis = 1)
In [34]: df[28:33] # As you can see the values have been copied. Now lets pluck out the starting digits.
Out[34]:
                                                                                                                      Hundred
                                       General
                                                   Event
                                                                 Event
                                                                                                             Event
                  CAD
                         CAD Event
                                                                         Event Clearance
                                                                                          Event Clearance
                                       Offense
                                                              Clearance
                                                                                                         Clearance
                                                                                                                        Block District/Sector Zone/Beat
                                               Clearance
                                                                               SubGroup
                CDW ID
                                                                                                   Group
                           Number
                                                             Description
                                                                                                              Date
                                                                                                                      Location
                                       Number
                                                   Code
                                                                                                                         21XX
                                                                                             NARCOTICS
                                                            NARCOTICS,
                                                                             NARCOTICS
                                                                                                            7/17/10
                                                                                                                    BLOCK OF
            28 15765.0 1.00002e+10 2010246224
                                                                            COMPLAINTS
                                                                                            COMPLAINTS
                                                                                                                     ALASKAN
                                                                 OTHER
                                                                                                             20:42
                                                                                                                    9XX BLOCK
                                                                            SUSPICIOUS
                                                                                                            7/17/10
                                                            SUSPICIOUS
                                                                                             SUSPICIOUS
            29 15766.0 2010246426 2010246426
                                                                                                                        OF NE
                                                               PERSON CIRCUMSTANCES CIRCUMSTANCES
                                                                                                                       50TH ST
                                                               PARKING
                                                             VIOLATION
                                                                                                                    3XX BLOCK
                                                                                                TRAFFIC
                                                                                                            7/17/10
                                                                               PARKING
            30 15767.0 2010246213 2010246213
                                                   470.0
                                                               (EXCEPT
                                                                                                                    OF TAYLOR
                                                                             VIOLATIONS RELATED CALLS
                                                           ABANDONED
                                                                                                                         AVE N
                                                             VEHICLES)
                                                                                                                    2XX BLOCK
                                                         DISTURBANCE,
                                                                                                            7/17/10
            31 15768.0 2010246429 2010246429
                                                                         DISTURBANCES DISTURBANCES
                                                                OTHER
                                                                                                                      THOMAS
                                                                                                              20:48
                                                                                                                           ST
                                                                LIQUOR
                                                                                                                    1XX BLOCK
                                                            VIOLATION -
                                                                                LIQUOR
                                                                                                 LIQUOR
                                                                                                            7/17/10
            32 15769.0 2010246463 2010246463
                                                           INTOXICATED
                                                                             VIOLATIONS
                                                                                             VIOLATIONS
                                                                                                                   BROADWAY
                                                                                                              20:49
                                                               PERSON
In [35]: |#Will replace only '2010' & 2016' from the entire digits and replace with 10000 and 160000 respectively
           df['CAD Event Number'] = df['CAD Event Number'].replace({'2010':'10000', '2016':'160000'}, regex=True)
In [36]: df[28:33] #check it out to confirm.
Out[36]:
                                                   Event
                                                                  Event
                                                                                                                       Hundred
                                        General
                                                                                                              Event
                  CAD
                         CAD Event
                                                                          Event Clearance
                                                                                          Event Clearance
                                                              Clearance
                                                                                                          Clearance
                                       Offense Clearance
                                                                                                                         Block District/Sector Zone/Beat
                                                                               SubGroup
                CDW ID
                                                                                                   Group
                            Number
                                        Number
                                                             Description
                                                                                                                       Location
                                                    Code
                                                                                                                         21XX
                                                                                              NARCOTICS
                                                                                                            7/17/10
                                                            NARCOTICS,
                                                                             NARCOTICS
                                                                                                                     BLOCK OF
            28 15765.0 1.00002e+10 2010246224
                                                                            COMPLAINTS
                                                                                             COMPLAINTS
                                                                                                              20:42
                                                                 OTHER
                                                                                                                      ALASKAN
                                                                                                                    9XX BLOCK
                                                                            SUSPICIOUS
                                                                                             SUSPICIOUS
                                                            SUSPICIOUS
                                                                                                            7/17/10
            29 15766.0 10000246426 2010246426
                                                                                                                        OF NE
                                                                PERSON CIRCUMSTANCES CIRCUMSTANCES
                                                                                                                       50TH ST
                                                               PARKING
                                                              VIOLATION
                                                                                                                    3XX BLOCK
                                                                                PARKING
                                                                                                 TRAFFIC
                                                                                                            7/17/10
            30 15767.0 10000246213 2010246213
                                                               (EXCEPT
                                                                                                                    OF TAYLOR
                                                                              VIOLATIONS RELATED CALLS
                                                            ABANDONED
                                                                                                                         AVE N
                                                             VEHICLES)
                                                                                                                    2XX BLOCK
                                                          DISTURBANCE,
                                                                                                             7/17/10
            31 15768.0 10000246429 2010246429
                                                                          DISTURBANCES DISTURBANCES
                                                                 OTHER
                                                                                                              20:48
                                                                                                                       THOMAS
                                                                LIQUOR
                                                                                                                    1XX BLOCK
                                                                                                 LIQUOR
                                                                                                            7/17/10
                                                             VIOLATION -
                                                                                 LIQUOR
                                                                                                                           OF
            32 15769.0 10000246463 2010246463
                                                                                                              20:49
                                                                                                                    BROADWAY
                                                           INTOXICATED
                                                                             VIOLATIONS
                                                                                              VIOLATIONS
                                                                PERSON
In [37]: df['CAD Event Number'].isna().sum()
Out[37]: 0
           3 Event Clearance Code
           MAP() to fill the NAs
           1- Creating a mapping of Event Clearance description and Event Clearance Code.
           2- Then, using this mapping to fill the NA values in column: Event Clearance description
In [38]: | a3 = df[df['Event Clearance Code'].isna()].head()
Out[38]:
                                       General
                                                   Event
                                                                                                            Event
                                                                                                                     Hundred
                  CAD
                         CAD Event
                                                           Event Clearance Event Clearance Event Clearance
                                                                                                         Clearance
                                                                                                                       Block District/Sector Zone/Beat
                                       Offense
                                               Clearance
                           Number
                                                               Description
                                                                               SubGroup
                                                                                                  Group
                                       Number
                                                   Code
                                                                                                              Date
                                                                                                                     Location
                                                                                                                        56XX
                                                                 TRAFFIC
                                                                                TRAFFIC
                                                                                                TRAFFIC
                                                                                                           7/17/10
                                                                                                                   BLOCK OF
             6 15743.0 1.00002e+10 2010246472
                                                                (MOVING)
                                                                                RELATED
                                                                                               RELATED
                                                    NaN
                                                                                                             21:05 BROOKLYN
                                                               VIOLATION
                                                                                  CALLS
                                                                                                 CALLS
                                                                                                                      AVE NE
                                                                 LIQUOR
                                                                                                                        15XX
                                                              VIOLATION -
                                                                                 LIQUOR
                                                                                                LIQUOR
                                                                                                           7/17/10
             7 15744.0 1.00002e+10 2010246493
                                                                                                                    BLOCK OF
                                                             INTOXICATED
                                                                              VIOLATIONS
                                                                                             VIOLATIONS
                                                                                                             21:09
                                                                                                                     4TH AVE
                                                                 PERSON
                                                                                                                       100XX
                                                                 MENTAL
                                                                                                MENTAL
                                                                                                           7/17/10
                                                                                                                    BLOCK OF
                                                                            MENTAL CALL
             8 15745.0 1.00002e+10 2010246478
                                                              COMPLAINT
                                                                                                HEALTH
                                                                                                            21:17
                                                                                                                    COLLEGE
                                                                                                                        46XX
                                                                                                 OTHER
                                                                                                           7/17/10
                                                                                                                   BLOCK OF
             9 15746.0 1.00002e+10 2010246397
                                                                                  THEFT
                                                          MISCELLANEOUS
                                                                                              PROPERTY
                                                                                                            21:18
                                                                                                                    S AUSTIN
                                                                                                                          ST
                                                                                                                         3XX
                                                           DISTURBANCE,
                                                                                                           7/17/10
                                                                                                                   BLOCK OF
                                                                          DISTURBANCES DISTURBANCES
            10 15747.0 1.00002e+10 2010246518
                                                                  OTHER
                                                                                                             21:28
                                                                                                                  HARRISON
In [39]: | desc_code = df.dropna(subset=['Event Clearance Code']).drop_duplicates('Event Clearance Description').set_index('Event Clearance Description').set_in
           nt Clearance Description')['Event Clearance Code']
           desc_code
Out[39]: Event Clearance Description
           THEFT - MISCELLANEOUS
                                                                                    65.0
           MISCHIEF, NUISANCE COMPLAINTS
                                                                                   250.0
           TRAFFIC (MOVING) VIOLATION
                                                                                   460.0
           SUSPICIOUS VEHICLE
                                                                                   281.0
           MENTAL COMPLAINT
                                                                                   220.0
                                                                                   . . .
           HARBOR - ASSIST BOATER (NON EMERGENCY)
                                                                                   343.0
           SEX OFFENDER - FAILURE TO REGISTER
                                                                                  143.0
           SOAP (STAY OUT OF AREA OF PROSTITUTION) ORDER VIOLATION
                                                                                  127.0
           HOMICIDE
                                                                                   10.0
           MARIJUANA PUBLIC USE (NOT DISPENSARY)
                                                                                  188.0
           Name: Event Clearance Code, Length: 95, dtype: float64
In [40]: df['Event Clearance Code'] = df['Event Clearance Code'].fillna(df['Event Clearance Description'].map(desc_code))
In [41]: df[5:11] # Confirm to see if the right codes have been filled
Out[41]:
                                       General
                                                   Event
                                                                                                                     Hundred
                  CAD
                         CAD Event
                                                          Event Clearance Event Clearance
                                                                                         Event Clearance
                                       Offense Clearance
                                                                                                         Clearance
                                                                                                                       Block District/Sector Zone/Beat
                                                                               SubGroup
                           Number
                                                               Description
                                                                                                  Group
                                       Number
                                                                                                              Date
                                                                                                                     Location
                                                                 MENTAL
                                                                                                MENTAL
                                                                                                           7/17/10
                                                                                                                   BLOCK OF
             5 15742.0 1.00002e+10 2010246436
                                                                            MENTAL CALL
                                                   220.0
                                                              COMPLAINT
                                                                                                HEALTH
                                                                                                             21:04
                                                                                                                    57TH AVE
                                                                                                                        56XX
                                                                 TRAFFIC
                                                                                TRAFFIC
                                                                                                TRAFFIC
                                                                                                           7/17/10
                                                                                                                   BLOCK OF
             6 15743.0 1.00002e+10 2010246472
                                                   460.0
                                                                (MOVING)
                                                                                RELATED
                                                                                               RELATED
                                                                                                                                        U
                                                                                                             21:05 BROOKLYN
                                                               VIOLATION
                                                                                  CALLS
                                                                                                 CALLS
                                                                 LIQUOR
                                                                                                                        15XX
                                                              VIOLATION -
                                                                                 LIQUOR
                                                                                                LIQUOR
                                                                                                           7/17/10
             7 15744.0 1.00002e+10 2010246493
                                                                                                                    BLOCK OF
                                                             INTOXICATED
                                                                              VIOLATIONS
                                                                                             VIOLATIONS
                                                                                                             21:09
                                                                                                                     4TH AVE
                                                                 PERSON
                                                                                                                       100XX
                                                                  MENTAL
                                                                                                MENTAL
                                                                                                           7/17/10
                                                                                                                    BLOCK OF
             8 15745.0 1.00002e+10 2010246478
                                                                            MENTAL CALL
                                                   220.0
                                                              COMPLAINT
                                                                                                HEALTH
                                                                                                            21:17
                                                                                                                    COLLEGE
                                                                                                                        46XX
                                                                                                           7/17/10
                                                                                                                   BLOCK OF
             9 15746.0 1.00002e+10 2010246397
                                                                                  THEFT
                                                         MISCELLANEOUS
                                                                                              PROPERTY
                                                                                                            21:18
                                                                                                                    S AUSTIN
                                                                                                                          ST
                                                                                                                         3XX
                                                                                                                   BLOCK OF
                                                           DISTURBANCE,
                                                                                                           7/17/10
                                                                          DISTURBANCES DISTURBANCES
            10 15747.0 1.00002e+10 2010246518
                                                                  OTHER
                                                                                                             21:28
                                                                                                                   HARRISON
In [42]: df['Event Clearance Code'].isna().sum()
Out[42]: 0
           4 Event Clearance Group
           MAP()
In [43]: | map1= df.dropna(subset=['Event Clearance Group']).drop_duplicates('Event Clearance SubGroup').set_index('Event Clear
           ance SubGroup')['Event Clearance Group']
In [44]: | df['Event Clearance Group'] = df['Event Clearance Group'].fillna(df['Event Clearance SubGroup'].map(map1))
In [45]: | df['Event Clearance Group'].isna().sum()
Out[45]: 0
           Event Clearance SubGroup
           MAP()
In [46]: subgroup_desc = df.dropna(subset=['Event Clearance SubGroup']).drop_duplicates('Event Clearance Description').set_in
           dex('Event Clearance Description')['Event Clearance SubGroup']
In [47]: | df['Event Clearance SubGroup'] = df['Event Clearance SubGroup'].fillna(df['Event Clearance Description'].map(subgroup')
           p_desc))
In [48]: | df['Event Clearance SubGroup'].isna().sum()
Out[48]: 0
           Event Clearance Description'
In [49]: c_desc = df.dropna(subset=['Event Clearance Description']).drop_duplicates(['Event Clearance Code']).set_index('Even
           t Clearance Code')['Event Clearance Description']
In [50]: df['Event Clearance Description'] = df['Event Clearance Description'].fillna(df['Event Clearance Code'].map(c_desc))
           Hundred Block Location
In [51]: print("Null Values before")
           df['Hundred Block Location'].isna().sum()
           Null Values before
Out[51]: 6
In [52]: | incid_loc_hblock = df.dropna(subset=['Hundred Block Location']).drop_duplicates(['Incident Location']).set_index('Incident Location']).
           cident Location')['Hundred Block Location']
In [53]: df['Hundred Block Location'] = df['Hundred Block Location'].fillna(df['Incident Location'].map(incid_loc_hblock))
In [54]: df['Hundred Block Location'].isna().sum() #Only 2 NAs removed. No duplicate values for other to create mappings. The
           se 4 values are unique.
Out[54]: 4
           Zone/Beat
In [55]: df['Zone/Beat'].isna().sum()
Out[55]: 26
In [56]: df['Zone/Beat'].unique()
Out[56]: array(['Q1', 'M2', 'D1', 'R3', 'U3', 'K1', 'N3', 'S2', 'K2', 'F3', 'S3',
                    'Q3', 'D2', 'J2', 'L3', 'J3', 'W1', 'U2', 'R2', 'M3', 'E1', 'L1',
                    'E2', nan, 'W3', 'N2', 'R1', 'L2', 'J1', 'G3', 'G2', 'C2', 'B1',
                    'S1', 'M1', 'F1', 'G1', 'B2', 'K3', 'B3', 'C1', 'O1', 'C3', 'U1',
                    'D3', 'Q2', 'N1', 'O2', 'W2', 'E3', 'F2', 'O3', '99'], dtype=object)
In [57]: | df['Zone/Beat'] = df['Zone/Beat'].replace('99', np.nan)
In [58]: Census_Zone = df.dropna(subset=['Zone/Beat']).drop_duplicates(['Census Tract']).set_index('Census Tract')['Zone/Bea
           df['Zone/Beat'] = df['Zone/Beat'].fillna(df['Census Tract'].map(Census_Zone))
In [59]: df['Zone/Beat'].isna().sum() # Some left since no matching values in mapping group for these values.
Out[59]: 6
           District/Sector
In [60]: #df['District/Sector'] = df['District/Sector'].astype(str)
           df['District/Sector'].isna().sum()
Out[60]: 14
In [61]: | df['District/Sector'].unique()
Out[61]: array(['Q', 'M', 'D', 'R', 'U', 'K', 'N', 'S', 'F', 'J', 'L', 'W', 'E',
                    'C', 'G', 'B', 'O', nan, '99'], dtype=object)
In [62]: df['District/Sector'] = df['District/Sector'].replace('99', np.nan)
           df['District/Sector'].isna().sum()
Out[62]: 31
In [63]: | Zone_District = df.dropna(subset=['District/Sector']).drop_duplicates('Zone/Beat').set_index(['Zone/Beat'])['Distric
           df['District/Sector'] = df['District/Sector'].fillna(df['Zone/Beat'].map(Zone_District))
In [64]: df['District/Sector'].isna().sum()
Out[64]: 0
           Incident Location
           **Since the Incident location is a combination of Latitude and Longitude, let's join them and put inplace of NAs in Incident Location Column
In [65]: df['Incident Location'].isna().sum()
Out[65]: 2469
In [66]: a = df[df['Incident Location'].isna()].index
In [67]: df.iloc[a].head()
Out[67]:
                                       General
                                                   Event
                                                                                                            Event
                                                                                                                     Hundred
                  CAD
                         CAD Event
                                                           Event Clearance Event Clearance Event Clearance
                                       Offense
                                               Clearance
                                                                                                         Clearance
                                                                                                                       Block District/Sector Zone/Beat
                CDW ID
                           Number
                                                               Description
                                                                               SubGroup
                                                                                                  Group
                                       Number
                                                   Code
                                                                                                              Date
                                                                                                                     Location
                                                                                                                        56XX
                                                                 TRAFFIC
                                                                                TRAFFIC
                                                                                                TRAFFIC
                                                                                                           7/17/10
                                                                                                                   BLOCK OF
             6 15743.0 1.00002e+10 2010246472
                                                   460.0
                                                                (MOVING)
                                                                                RELATED
                                                                                               RELATED
                                                                                                             21:05
                                                                                                                  BROOKLYN
                                                               VIOLATION
                                                                                  CALLS
                                                                                                 CALLS
                                                                                                                      AVE NE
                                                                 LIQUOR
                                                                                                                        15XX
                                                              VIOLATION -
                                                                                 LIQUOR
                                                                                                LIQUOR
                                                                                                           7/17/10
             7 15744.0 1.00002e+10 2010246493
                                                   177.0
                                                                                                                    BLOCK OF
                                                             INTOXICATED
                                                                              VIOLATIONS
                                                                                             VIOLATIONS
                                                                                                             21:09
                                                                                                                     4TH AVE
                                                                 PERSON
                                                                                                                       100XX
                                                                 MENTAL
                                                                                                MENTAL
                                                                                                           7/17/10
                                                                                                                   BLOCK OF
                                                                            MENTAL CALL
             8 15745.0 1.00002e+10 2010246478
                                                   220.0
                                                              COMPLAINT
                                                                                                HEALTH
                                                                                                            21:17
                                                                                                                    COLLEGE
                                                                                                                        46XX
                                                                                                 OTHER
                                                                                                           7/17/10
                                                                                                                   BLOCK OF
             9 15746.0 1.00002e+10 2010246397
                                                                                  THEFT
                                                         MISCELLANEOUS
                                                                                              PROPERTY
                                                                                                            21:18
                                                                                                                    S AUSTIN
                                                                                                                         3XX
                                                           DISTURBANCE,
                                                                                                           7/17/10
                                                                                                                   BLOCK OF
                                                                          DISTURBANCES DISTURBANCES
            10 15747.0 1.00002e+10 2010246518
                                                                  OTHER
                                                                                                             21:28 HARRISON
In [68]: | df['Incident Location'] = df['Incident Location'].fillna(( '(' +df['Latitude'].astype(str) + ',-' + df['Longitude'].
           astype(str)+ ')' ))
In [69]: df.iloc[a].head()
Out[69]:
                                       General
                                                   Event
                                                                                                            Event
                                                                                                                     Hundred
                  CAD CAD Event
                                                           Event Clearance Event Clearance Event Clearance
                                       Offense
                                                                                                         Clearance
                                                                                                                       Block District/Sector Zone/Beat
                                               Clearance
                CDW ID
                           Number
                                                                               SubGroup
                                                                                                  Group
                                                               Description
                                       Number
                                                   Code
                                                                                                              Date
                                                                                                                     Location
                                                                                                                        56XX
                                                                                TRAFFIC
                                                                                                TRAFFIC
                                                                 TRAFFIC
                                                                                                           7/17/10 BLOCK OF
             6 15743.0 1.00002e+10 2010246472
                                                   460.0
                                                                (MOVING)
                                                                                RELATED
                                                                                               RELATED
                                                                                                                                        U
                                                                                                             21:05 BROOKLYN
                                                               VIOLATION
                                                                                  CALLS
                                                                                                 CALLS
                                                                 LIQUOR
                                                              VIOLATION -
                                                                                 LIQUOR
                                                                                                LIQUOR
                                                                                                           7/17/10
             7 15744.0 1.00002e+10 2010246493
                                                                                                                    BLOCK OF
                                                                              VIOLATIONS
                                                                                            VIOLATIONS
                                                             INTOXICATED
                                                                                                            21:09
                                                                 PERSON
                                                                                                                       100XX
                                                                 MENTAL
                                                                                                MENTAL
                                                                                                           7/17/10 BLOCK OF
                                                                            MENTAL CALL
             8 15745.0 1.00002e+10 2010246478
                                                              COMPLAINT
                                                                                                HEALTH
                                                                                                            21:17
                                                                                                                    COLLEGE
                                                                                                                        46XX
                                                                                                 OTHER
                                                                                                           7/17/10 BLOCK OF
             9 15746.0 1.00002e+10 2010246397
                                                                                  THEFT
                                                                                                                                        S
                                                         MISCELLANEOUS
                                                                                              PROPERTY
                                                                                                             21:18
                                                                                                                    S AUSTIN
                                                                                                                         3XX
                                                           DISTURBANCE,
                                                                                                           7/17/10 BLOCK OF
            10 15747.0 1.00002e+10 2010246518
                                                                          DISTURBANCES DISTURBANCES
                                                                                                                                        D
                                                                  OTHER
                                                                                                            21:28 HARRISON
In [70]: | df['Incident Location'].isna().sum() # All values filled
Out[70]: 0
           General Offense Number
             • There are a few string values in this column, lets get rid of them

    I converted the column to numeric so that strings will be set as NaN

             • Then I filled the NaN with values from CAD Event Number
             • Finally changed the starting digits from 10000 to 2010....etc using REGEX
In [71]: df['General Offense Number'] = pd.to_numeric(df['General Offense Number'], errors='coerce')
In [72]: a5 = df[df['General Offense Number'].isna()]
           a5.head(3)
Out[72]:
                                    General
                                                Event
                                                                            Event
                                                                                          Event
                                                                                                    Event
                                                                                                              Hundred
                                                       Event Clearance
                  CAD CAD Event
                                    Offense Clearance
                                                                         Clearance
                                                                                      Clearance Clearance
                                                                                                                Block District/Sector Zone/Beat
                CDW ID
                           Number
                                                           Description
                                                                         SubGroup
                                                Code
                                                                                         Group
                                                                                                     Date
                                                                                                             Location
                                                                                                                 36XX
                                                                                                   7/17/30
                                                                                                           BLOCK OF
                                                               THEFT -
                                                                                        OTHER
                                                 65.0 MISCELLANEOUS
             0 15737.0 1.00002e+10
                                                                            THEFT
                                                                                                                                          Q1 5700.1012 -1
                                                                                     PROPERTY
                                                                                                    20:50 DISCOVERY
                                                                                                           PARK BLVD
                                                                                                   7/17/40 5XX BLOCK
            27 15764.0 1.00002e+10
                                                             SHOPLIFT
                                                                            THEFT SHOPLIFTING
                                                                                                                                           K1 8200.1004 -1
                                       NaN
                                                 64.0
                                                                                                          OF PINE ST
                                                                                                    20:39
                                                               LIQUOR
                                                                                                           8XX BLOCK
                                                                                        LIQUOR
                                                                                                   7/17/10
                                                           VIOLATION -
                                                                           LIQUOR
            94 15833.0 1.00002e+10
                                                177.0
                                                                                                            OF DENNY
                                                                                                                                          M2 7200.1039 -1
                                                          INTOXICATED VIOLATIONS
                                                                                    VIOLATIONS
                                                                                                    22:50
                                                              PERSON
In [73]: c = ['CAD Event Number', 'General Offense Number']
In [74]: df.loc[:,c] = df.loc[:,c].ffill(axis = 1)
In [75]: | df['General Offense Number'] = df['General Offense Number'].astype(str)
           df['General Offense Number'] = df['General Offense Number'].replace({'10000':'2010','160000':'2016'}, regex = True)
In [76]: | df['General Offense Number'].isna().sum()
Out[76]: 0
           Event Clearance Date
           Looked for dates in range and found invalid values which were not in range
In [77]: df['Event Clearance Date'] = pd.to_datetime(df['Event Clearance Date'], errors = 'coerce')
In [78]: | a = df['Event Clearance Date']
           Startdate = pd.to_datetime('1/1/2000')
           Enddate = pd.to_datetime('1/1/2018')
In [79]: error_dates1 = df.loc[df['Event Clearance Date'] < Startdate, : ]</pre>
           error_dates2 = df.loc[df['Event Clearance Date'] > Enddate, :]
```

Cens

U3 5301.20

D1 7200.10

D1 7100.20

E1 7400.30

D1 7100.20

M3 800

U3 530

D1 720

D1 710

E1 740

M3 80

U3 53

D1 72

D1 71

E1 74

U3 440

K1 810

N3 130

S2 1110

D1 710

R3 1020

U3 440

K1 810

N3 130

S2 1110

D1 710

U3 440

K1 810

N3 130

S2 1110

D1 710

U3 440

K1 810

N3 130

S2 1110

D1 710

Census

Tra

In [86]: |df['Event Clearance Date'].iloc[error_dates2.index] # As you can see, the dates are now in proper range Out[86]: 0 2010-07-17 20:50:00 27 2010-07-17 20:39:00 90 2010-07-18 00:12:00 2010-07-18 00:04:00 Name: Event Clearance Date, dtype: object

Hundred Block Location - Invalid(anomalies)

In [87]: | df['Hundred Block Location'].isna().sum()

1680 17462.0 1.00002e+10 2010249050.0

In [89]: df.loc[mask.index]['Hundred Block Location']

Canada

Alaska

Name: Hundred Block Location, dtype: object

In [92]: | df['Hundred Block Location'].iloc[mask.index] = np.nan

Pennsylvania Brooklyn

Washington DC

dred Block Location']

Manhattan

Looked for any string less than length of 10, which resulted in a list of country/city names

In [80]: | df['Event Clearance Date'] = df['Event Clearance Date'].astype(str)

_list()

Out[81]: ['2030-07-17 20:50:00',

), regex = **True**)

Out[87]: 4

Out[89]: 11

1680

5718

7352 7359

7826

In [97]: df.isna().sum()

Zone/Beat Census Tract

Longitude Latitude

their value.

dtype: int64

Incident Location

Event Clearance Group Event Clearance Date Hundred Block Location

District/Sector

Incident Location

import seaborn as sns

In [103]: sns.set(style="darkgrid")

In [102]: c = df['District/Sector'].value_counts()

dtype: int64

In [101]: import pandas as pd

Zone/Beat Census Tract Longitude Latitude

Out[97]: CAD CDW ID

In [81]: error # List of all Erroneous Dates

'2040-07-17 20:39:00', '2033-07-18 00:12:00', '2021-07-18 00:04:00' '1710-07-25 07:12:00']

error = df['Event Clearance Date'][error_dates2.index].to_list() + df['Event Clearance Date'][error_dates1.index].to

Event Hundred

Date Location

Canada

Alaska

Block District/Sector Zone/Beat

G

K2 9100.

G3 8800.

Clearance

2010-07-

21:28:00

23:18:00

17

19

Event Clearance

Group

TRESPASS

SUSPICIOUS

SUSPICIOUS

PERSON CIRCUMSTANCES CIRCUMSTANCES

In [84]: | df['Event Clearance Date'] = df['Event Clearance Date'].replace(dict.fromkeys(['2030', '2040', '2033', '2021'], '2010'

```
In [88]: | mask = (df[df['Hundred Block Location'].str.len() <14])</pre>
          mask.head(2)
Out[88]:
                                        General
                                                   Event
                                                               Event
                   CAD
                         CAD Event
                                                                       Event Clearance
                                        Offense Clearance
                                                            Clearance
                CDW ID
                           Number
                                                                            SubGroup
                                        Number
                                                           Description
             11 15748.0 1.00002e+10 2010246470.0
                                                    161.0
                                                           TRESPASS
                                                                            TRESPASS
```

		For the last 4 columns, there are so many NA values and they little in common with other columns to retrieve the lost data Hence will lose this data										
Out[95]:	: 4										
In [95]:	<pre>df['Hundred Block Location'].isna().sum() mappings group.</pre>	#	4 values	left,	as th	hey ar	e unique	and	don't	a match	in the
In [[94]: df['Hundred Block Location'] = df['Hundred Block Location'].fillna(df['Longitude'].map(Long_HBloc))											

In [96]: | df = df.drop(['Initial Type Description', 'Initial Type Subgroup', 'Initial Type Group', 'At Scene Time'], axis = 1)

In [93]: Long_HBloc = df.dropna(subset=['Hundred Block Location']).drop_duplicates(['Longitude']).set_index('Longitude')['Hun

SUSPICIOUS

280.0

```
CAD Event Number
General Offense Number
Event Clearance Code
Event Clearance Description
Event Clearance SubGroup
Event Clearance Group
Event Clearance Date
Hundred Block Location
District/Sector
```

18

0

"Melt" the dataset to "long-form" or "tidy" representation iris = pd.melt(df, "District/Sector", var_name="measurement")

In [98]: df = df.dropna() # Only 28 values lost out of 10,000

In [100]: | df.to_csv('/Users/qas/Desktop/cleanedSeattle.csv')

sns.set(style="whitegrid", palette="muted")

sns.barplot(c.index, c.values, alpha=0.9) plt.title('Frequency of Disctrict/Sector')

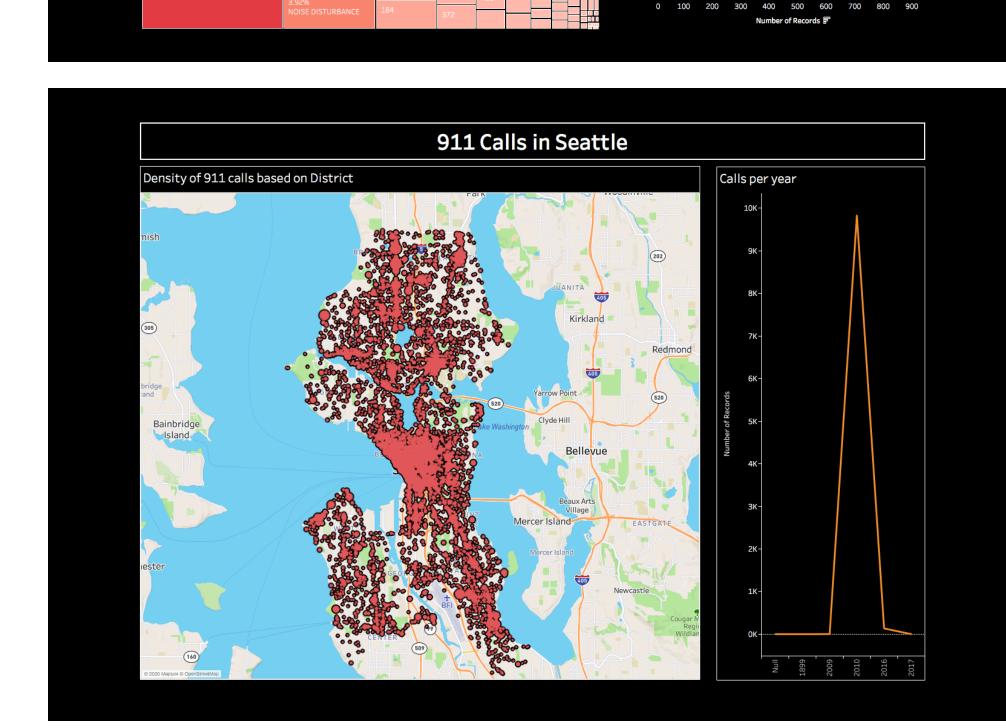
Since there are only a few null values, we will drop these values, given they are unique and have no match in any of the mappings-group to find

```
plt.ylabel('Number of Occurrences', fontsize=12)
plt.show()
                 Frequency of Disctrict/Sector
   600
Ź 200
Visualization in Tableau
                                               911 Calls in Seattle
```

911 Calls per District/Sector

911 as per Event Code

Percentage Frequency of Event Descriptions



In [99]: df.isna().sum() Out[99]: CAD CDW ID 0 0 CAD Event Number General Offense Number Event Clearance Code Event Clearance Description Event Clearance SubGroup