In [1]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
```

In [2]:

```
crime_data = pd.read_csv('Crimes.csv')
crime_data.head(5)
```

Out[2]:

	ID	Case Number	Date	Block	IUCR	Primary Type	Description	Location Description
0	11798117	JC397040	8/17/19 23:59	058XX W EDDY ST	820	THEFT	$500 AND UNDER	STREET
1	11798223	JC397365	8/17/19 23:59	007XX E 87TH ST	1310	CRIMINAL DAMAGE	TO PROPERTY	TAVERN/LIQUOR STORE
2	11797883	JC396936	8/17/19 23:57	022XX W 50TH PL	143A	WEAPONS VIOLATION	UNLAWFUL POSS OF HANDGUN	SIDEWALK
3	11797909	JC397003	8/17/19 23:55	086XX S PHILLIPS AVE	143A	WEAPONS VIOLATION	UNLAWFUL POSS OF HANDGUN	ALLEY
4	11797926	JC396947	8/17/19 23:45	079XX S SOUTH CHICAGO AVE	460	BATTERY	SIMPLE	SIDEWALK

In [3]:

```
crime_data.shape
```

Out[3]:

```
(161942, 14)
```

In [4]:

```
1 crime_data.head(5)
```

Out[4]:

	ID	Case Number	Date	Block	IUCR	Primary Type	Description	Location Description
0	11798117	JC397040	8/17/19 23:59	058XX W EDDY ST	820	THEFT	$500 AND UNDER	STREET
1	11798223	JC397365	8/17/19 23:59	007XX E 87TH ST	1310	CRIMINAL DAMAGE	TO PROPERTY	TAVERN/LIQUOR STORE
2	11797883	JC396936	8/17/19 23:57	022XX W 50TH PL	143A	WEAPONS VIOLATION	UNLAWFUL POSS OF HANDGUN	SIDEWALK
3	11797909	JC397003	8/17/19 23:55	086XX S PHILLIPS AVE	143A	WEAPONS VIOLATION	UNLAWFUL POSS OF HANDGUN	ALLEY
4	11797926	JC396947	8/17/19 23:45	079XX S SOUTH CHICAGO AVE	460	BATTERY	SIMPLE	SIDEWALK

In [5]:

```
1 crime_data.dtypes
```

Out[5]:

```
ID                      int64
Case Number            object
Date                   object
Block                  object
IUCR                   object
Primary Type           object
Description            object
Location Description   object
Arrest                   bool
Domestic                 bool
Beat                    int64
District                int64
Ward                  float64
Community Area          int64
dtype: object
```

In [6]:

```
1 crime_data.isnull().sum()
```

Out[6]:

```
ID                        0
Case Number               0
Date                      0
Block                     0
IUCR                      0
Primary Type              0
Description               0
Location Description    569
Arrest                    0
Domestic                  0
Beat                      0
District                  0
Ward                      7
Community Area            0
dtype: int64
```

In [7]:

```
1 #drop the null valus
2 crime_data = crime_data.dropna()
```

In [8]:

```
1 crime_data.shape
```

Out[8]:

```
(161366, 14)
```

In [9]:

```
1 crime_data.isnull().sum()
```

Out[9]:

```
ID                      0
Case Number             0
Date                    0
Block                   0
IUCR                    0
Primary Type            0
Description             0
Location Description    0
Arrest                  0
Domestic                0
Beat                    0
District                0
Ward                    0
Community Area          0
dtype: int64
```

In [10]:

```
1 crime_data.columns
```

Out[10]:

```
Index(['ID', 'Case Number', 'Date', 'Block', 'IUCR', 'Primary Type',
       'Description', 'Location Description', 'Arrest', 'Domestic', 'Bea
t',
       'District', 'Ward', 'Community Area'],
      dtype='object')
```

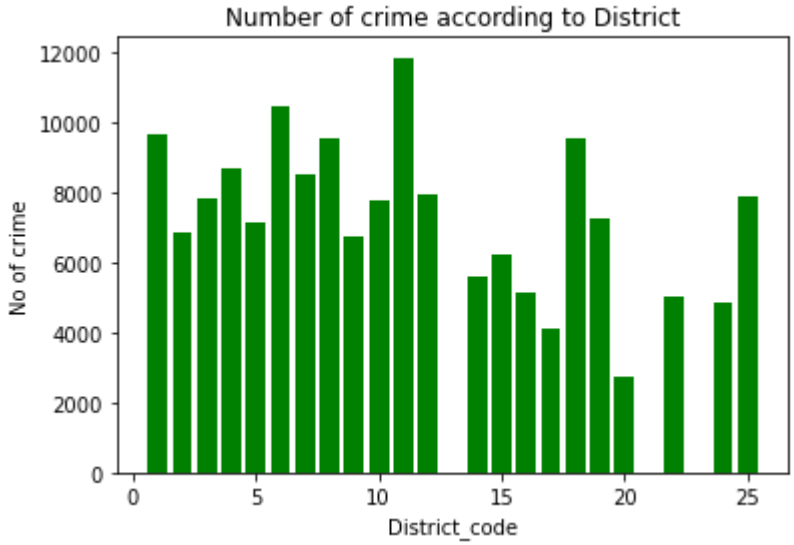
1.According to the district count the no of crime

In [11]:

```
count = crime_data['District'].value_counts()
print(count)
plt.bar(count.index,count.values,color = 'green')
plt.title('Number of crime according to District')
plt.xlabel('District_code')
plt.ylabel('No of crime')

plt.show()
```

```
11    11829
6     10451
1      9630
18     9558
8      9544
4      8703
7      8493
12     7957
25     7869
3      7844
10     7788
19     7246
5      7136
2      6847
9      6747
15     6242
14     5614
16     5119
22     5043
24     4869
17     4099
20     2738
Name: District, dtype: int64
```



2.month Vs Crime_count

In [12]:

```
crime_data['Date'] = pd.to_datetime(crime_data['Date'])
print(crime_data['Date'].dtype)
crime_data['month'] = crime_data['Date'].dt.month
month_counts = crime_data['month'].value_counts().sort_index()
print(month_counts)
plt.plot(month_counts.index,month_counts.values,color = 'red')
plt.title('month VS crime_count')
plt.xlabel('Month')
plt.ylabel('No of crime')

plt.show()
```

```
datetime64[ns]
1    19263
2    18047
3    20100
4    20604
5    23242
6    22996
7    24151
8    12963
Name: month, dtype: int64
```

3 According to Primary Type crime_count

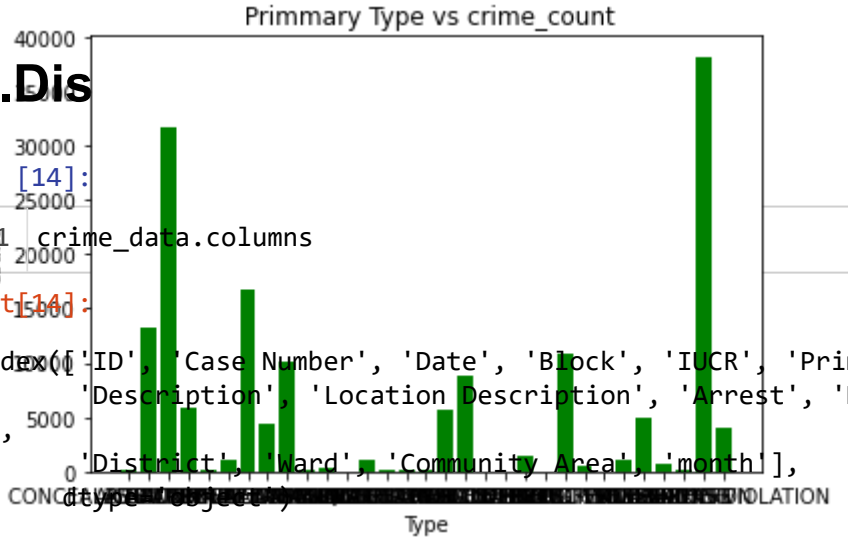
In [13]:

```
crime_count = crime_data['Primary Type'].value_counts().sort_index()
print(crime_count)
plt.bar(crime_count.index,crime_count.values,color ='green')

plt.xlabel("Type")

plt.ylabel("count")
plt.title(" Primmary Type vs crime_count")
plt.show()
```

```
ARSON                                 221
ASSAULT                             13218
BATTERY                             31727
BURGLARY                             5824
CONCEALED CARRY LICENSE VIOLATION     133
CRIM SEXUAL ASSAULT                   996
CRIMINAL DAMAGE                     16806
CRIMINAL TRESPASS                    4298
DECEPTIVE PRACTICE                  10063
GAMBLING                               92
HOMICIDE                              311
HUMAN TRAFFICKING                       8
INTERFERENCE WITH PUBLIC OFFICER     1016
INTIMIDATION                          103
KIDNAPPING                            112
LIQUOR LAW VIOLATION                  146
MOTOR VEHICLE THEFT                  5712
NARCOTICS                            8767
NON-CRIMINAL                            1
OBSCENITY                              31
OFFENSE INVOLVING CHILDREN           1478
OTHER NARCOTIC VIOLATION                4
OTHER OFFENSE                       10761
PROSTITUTION                          455
PUBLIC INDECENCY                        4
PUBLIC PEACE VIOLATION               1003
ROBBERY                              4903
SEX OFFENSE                           771
STALKING                              141
THEFT                               38199
WEAPONS VIOLATION                    4062
Name: Primary Type, dtype: int64
```



4.Dis

In [14]:

```
crime_data.columns
```

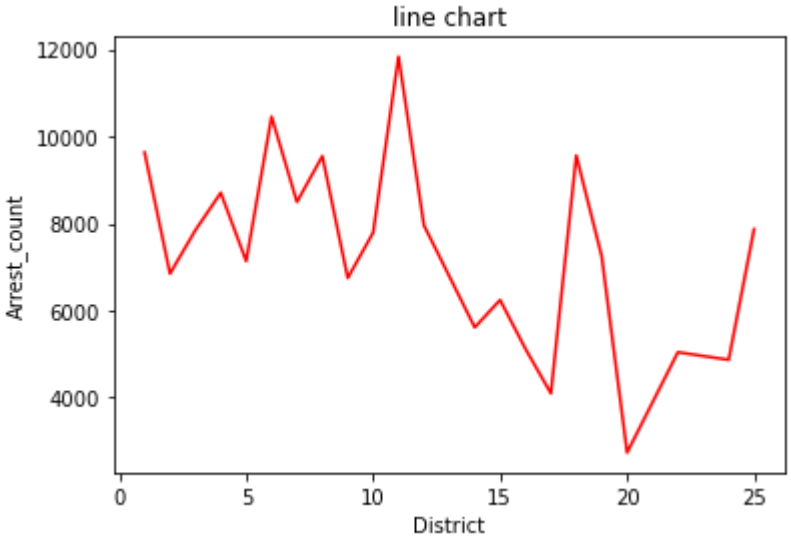
Out[14]:

```
Index(['ID', 'Case Number', 'Date', 'Block', 'IUCR', 'Primary Type',
       'Description', 'Location Description', 'Arrest', 'Domestic', 'Bea
t',
       'District', 'Ward', 'Community Area', 'month'],
      dtype='object')
```

In [43]:

```
Arrest_data = crime_data[crime_data['Arrest'] == True]
Arrest_count = crime_data.groupby(['District']).agg({'Arrest' : 'count'}).sort_index
print(Arrest_count)
plt.plot(Arrest_count.index, Arrest_count.values,color = 'red')
plt.xlabel('District')
plt.ylabel('Arrest_count')
plt.title('line chart')
plt.show()
```

```
          Arrest
District
1           9630
2           6847
3           7844
4           8703
5           7136
6          10451
7           8493
8           9544
9           6747
10          7788
11         11829
12          7957
14          5614
15          6242
16          5119
17          4099
18          9558
19          7246
20          2738
22          5043
24          4869
25          7869
```



In [ ]:

```
1
```

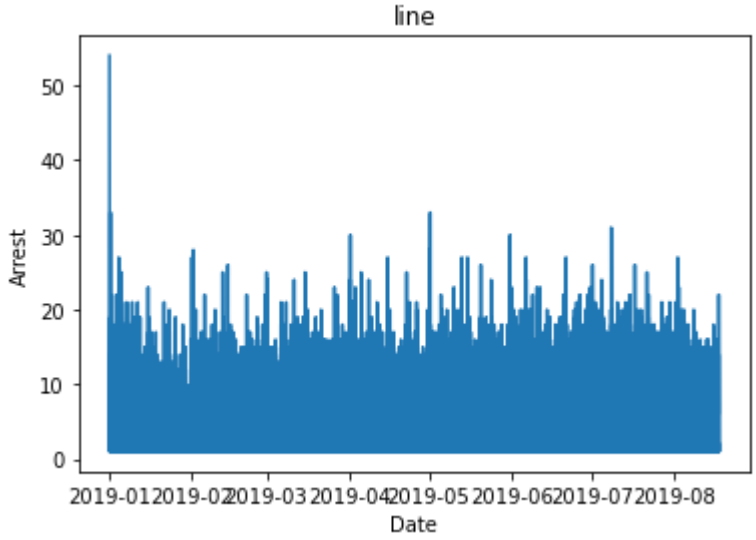
5.Date wise crime_count

In [16]:

```
Arrest_count_by_date = crime_data.groupby(['Date']).agg({'Arrest' : 'count'})
print(Arrest_count_by_date)
plt.plot(Arrest_count_by_date.index, Arrest_count_by_date.values)
plt.xlabel('Date')
plt.ylabel('Arrest')
plt.title('line')
plt.show()
```

```
                     Arrest
Date
2019-01-01 00:00:00      54
2019-01-01 00:01:00      41
2019-01-01 00:02:00       3
2019-01-01 00:03:00       2
2019-01-01 00:05:00       3
...                     ...
2019-08-17 23:42:00       2
2019-08-17 23:45:00       1
2019-08-17 23:55:00       1
2019-08-17 23:57:00       1
2019-08-17 23:59:00       2

[79284 rows x 1 columns]
```

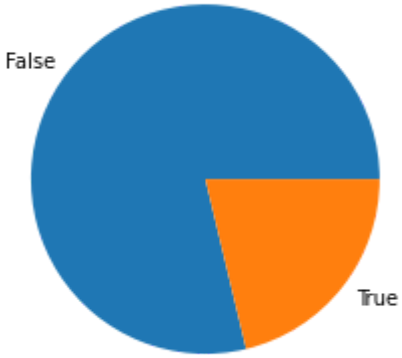


6. Arrest Vs Not_arrest

In [17]:

```
True_false_value_count = crime_data['Arrest'].value_counts().sort_index()
print(True_false_value_count)
d = dict(True_false_value_count)
plt.pie(d.values(), labels=d.keys())
plt.show()
```

```
False    127096
True      34270
Name: Arrest, dtype: int64
```

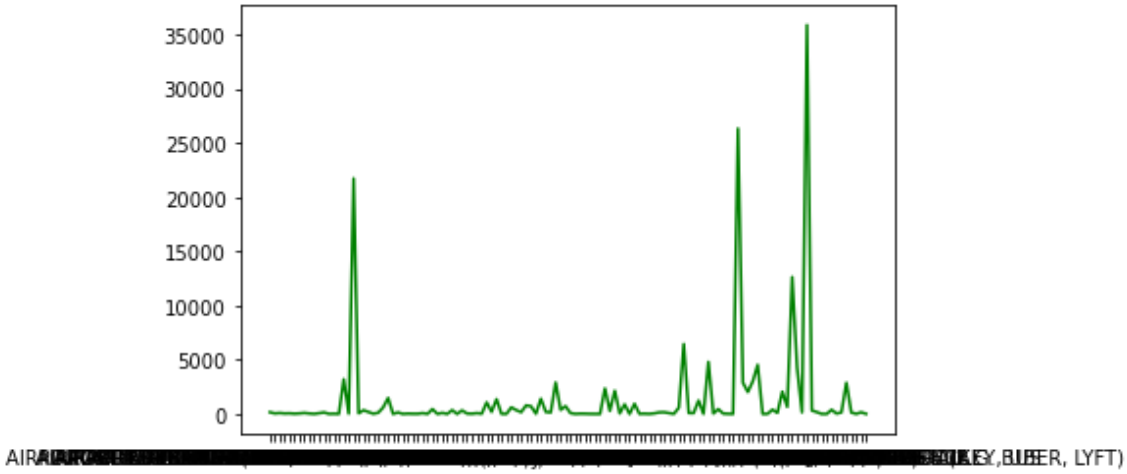


7.Location Description vs crime_count

In [18]:

```
location_wise = crime_data['Location Description'].value_counts().sort_index()
print(location_wise)
plt.plot(location_wise.index, location_wise.values,color = 'green')
plt.show()
```

```
ABANDONED BUILDING                                  161
AIRCRAFT                                             48
AIRPORT BUILDING NON-TERMINAL - NON-SECURE AREA      88
AIRPORT BUILDING NON-TERMINAL - SECURE AREA          44
AIRPORT EXTERIOR - NON-SECURE AREA                   58
                                                   ...
VEHICLE NON-COMMERCIAL                             2876
VEHICLE-COMMERCIAL                                  135
VEHICLE-COMMERCIAL - TROLLEY BUS                      3
WAREHOUSE                                           156
YARD                                                  6
Name: Location Description, Length: 122, dtype: int64
```

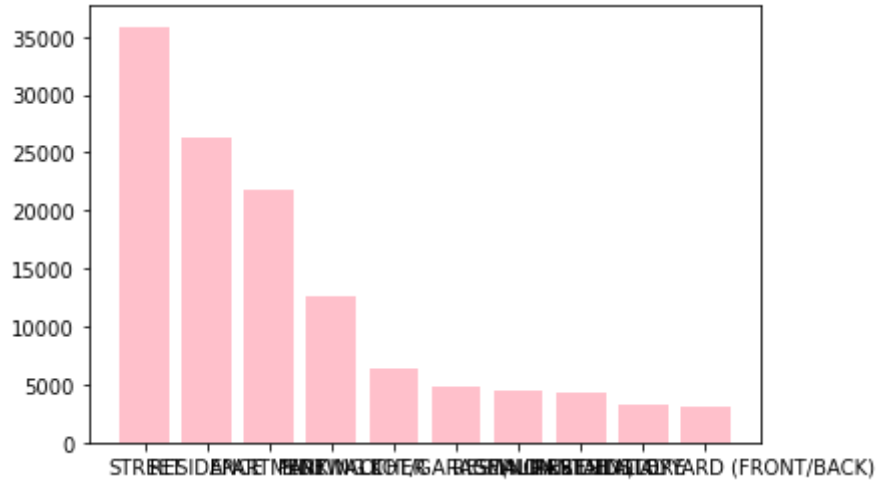


8.top 5 location wise crime

In [19]:

```
location_wise = crime_data['Location Description'].value_counts().sort_values(ascend
n_location_wise = location_wise.head(10)
print(n_location_wise)
plt.bar(n_location_wise.index,n_location_wise.values,color = 'pink')
plt.show()
```

```
STREET                            35875
RESIDENCE                         26326
APARTMENT                         21723
SIDEWALK                          12634
OTHER                              6451
PARKING LOT/GARAGE(NON.RESID.)     4800
RESTAURANT                         4529
SMALL RETAIL STORE                 4265
ALLEY                              3201
RESIDENTIAL YARD (FRONT/BACK)      2999
Name: Location Description, dtype: int64
```



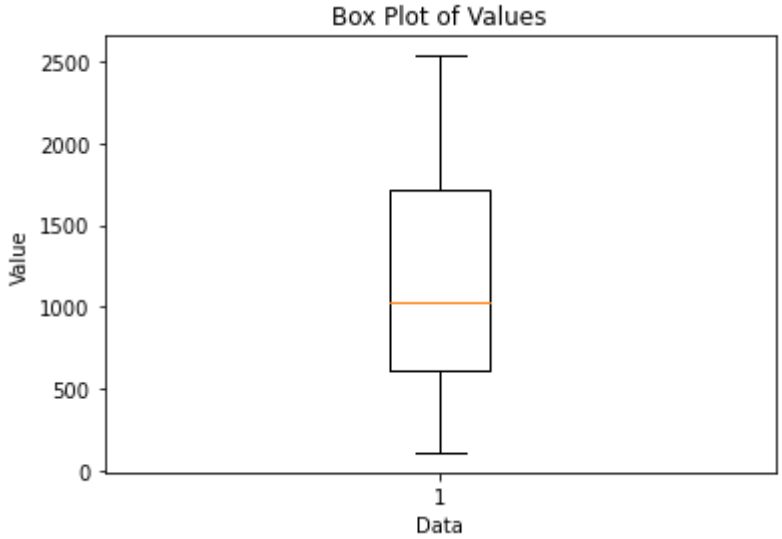
9.Create a box plot by using Beat

In [20]:

```
plt.boxplot(crime_data['Beat'])

plt.title('Box Plot of Values')
plt.xlabel('Data')
plt.ylabel('Value')

plt.show()
```



Type *Markdown* and LaTeX: α^2

In [21]:

```
crime_data['Date'] = pd.to_datetime(crime_data['Date'])
```

In [22]:

```
crime_data.dtypes
```

Out[22]:

```
ID                             int64
Case Number                   object
Date                  datetime64[ns]
Block                         object
IUCR                          object
Primary Type                  object
Description                   object
Location Description          object
Arrest                          bool
Domestic                        bool
Beat                           int64
District                       int64
Ward                         float64
Community Area                 int64
month                          int64
dtype: object
```

Extract only date

In [23]:

```
crime_data['Date'] = pd.to_datetime(crime_data['Date'], format='%Y-%m-%d %H:%M:%S')
crime_data['Date'] = crime_data['Date'].dt.date
print(crime_data.head())
```

```
        ID Case Number        Date                      Block  IUCR  \
0  11798117    JC397040  2019-08-17            058XX W EDDY ST   820
1  11798223    JC397365  2019-08-17            007XX E 87TH ST  1310
2  11797883    JC396936  2019-08-17            022XX W 50TH PL  143A
3  11797909    JC397003  2019-08-17       086XX S PHILLIPS AVE  143A
4  11797926    JC396947  2019-08-17  079XX S SOUTH CHICAGO AVE   460

        Primary Type               Description Location Description  Arres
t  \
0              THEFT            $500 AND UNDER               STREET    Tru
e
1    CRIMINAL DAMAGE               TO PROPERTY  TAVERN/LIQUOR STORE   Fals
e
2  WEAPONS VIOLATION  UNLAWFUL POSS OF HANDGUN             SIDEWALK   Fals
e
3  WEAPONS VIOLATION  UNLAWFUL POSS OF HANDGUN                ALLEY   Fals
e
4            BATTERY                    SIMPLE             SIDEWALK    Tru
e

   Domestic  Beat  District  Ward  Community Area  month
0     False  1633        16  36.0              15      8
1     False   632         6   6.0              44      8
2     False   931         9  15.0              63      8
3     False   423         4   7.0              46      8
4     False   411         4   8.0              46      8
```

In [24]:

```
1 crime_data
```

Out[24]:

	ID	Case Number	Date	Block	IUCR	Primary Type	Description	Des
0	11798117	JC397040	2019-08-17	058XX W EDDY ST	820	THEFT	$500 AND UNDER	
1	11798223	JC397365	2019-08-17	007XX E 87TH ST	1310	CRIMINAL DAMAGE	TO PROPERTY	TAVERN
2	11797883	JC396936	2019-08-17	022XX W 50TH PL	143A	WEAPONS VIOLATION	UNLAWFUL POSS OF HANDGUN	SI
3	11797909	JC397003	2019-08-17	086XX S PHILLIPS AVE	143A	WEAPONS VIOLATION	UNLAWFUL POSS OF HANDGUN	
4	11797926	JC396947	2019-08-17	079XX S SOUTH CHICAGO AVE	460	BATTERY	SIMPLE	SI
...	...	...	...	...	...	...	...	
161937	11649827	JC218222	2019-01-01	027XX S TRIPP AVE	1562	SEX OFFENSE	AGG CRIMINAL SEXUAL ABUSE	RES
161938	11673881	JC246843	2019-01-01	013XX W 72ND PL	1153	DECEPTIVE PRACTICE	FINANCIAL IDENTITY THEFT OVER $ 300	RES
161939	11643551	JC210380	2019-01-01	045XX N BEACON ST	1153	DECEPTIVE PRACTICE	FINANCIAL IDENTITY THEFT OVER $ 300	APA
161940	11583562	JC137815	2019-01-01	045XX N GREENVIEW AVE	1153	DECEPTIVE PRACTICE	FINANCIAL IDENTITY THEFT OVER $ 300	
161941	11552674	JC100085	2019-01-01	092XX S NORMAL AVE	910	MOTOR VEHICLE THEFT	AUTOMOBILE	

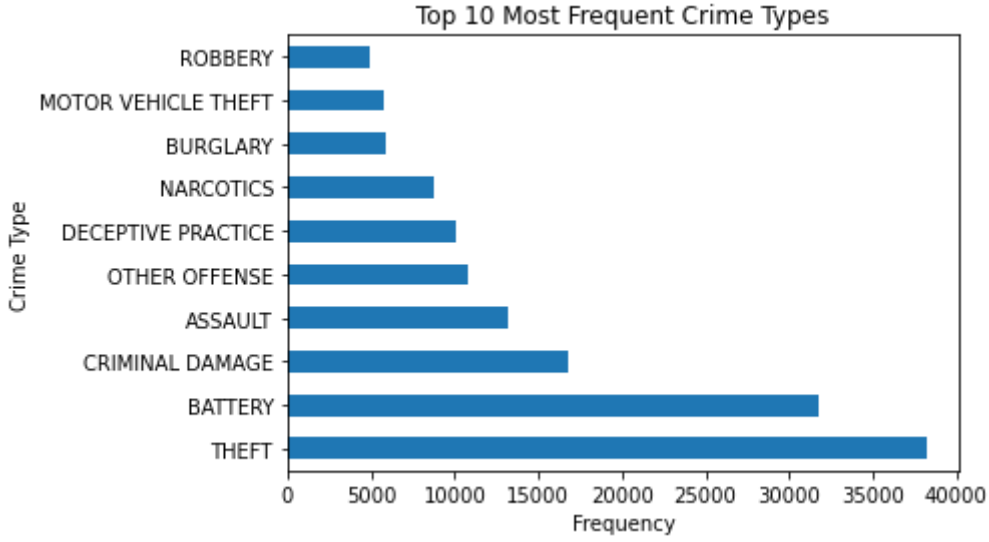
161366 rows × 15 columns

1.Crime frequency

In [25]:

```
crime_freq = crime_data['Primary Type'].value_counts()
print(crime_freq.head(10))
crime_freq.head(10).plot(kind='barh')
plt.title('Top 10 Most Frequent Crime Types')
plt.xlabel('Frequency')
plt.ylabel('Crime Type')
plt.show()
```

```
THEFT                38199
BATTERY              31727
CRIMINAL DAMAGE      16806
ASSAULT              13218
OTHER OFFENSE        10761
DECEPTIVE PRACTICE   10063
NARCOTICS             8767
BURGLARY              5824
MOTOR VEHICLE THEFT   5712
ROBBERY               4903
Name: Primary Type, dtype: int64
```

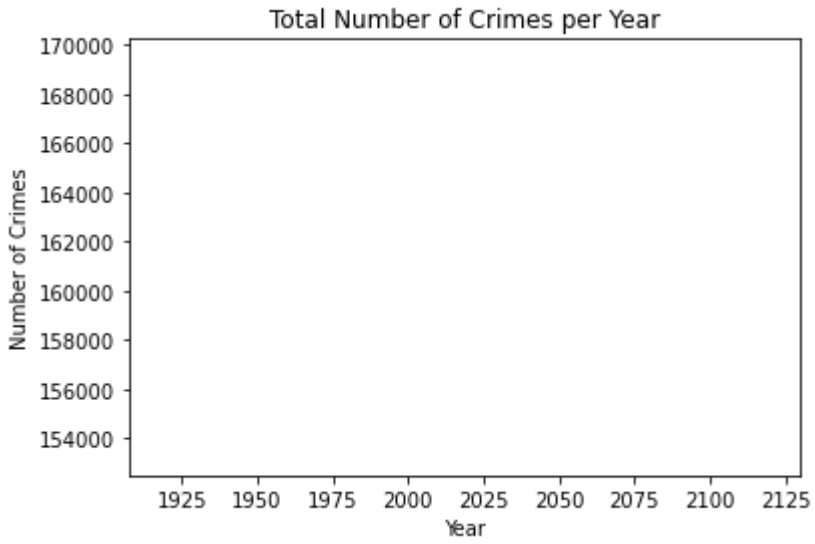


Crime trends over time

In [36]:

```
crime_data['Date'] = pd.to_datetime(crime_data['Date'], format='%Y-%m-%d %H:%M:%S')
crime_counts = crime_data.groupby(crime_data['Date'].dt.year).size()
print(crime_counts)
plt.plot(crime_counts.index, crime_counts.values)
plt.title('Total Number of Crimes per Year')
plt.xlabel('Year')
plt.ylabel('Number of Crimes')
plt.show()
```

```
Date
2019    161366
dtype: int64
```



In [49]:

```
from wordcloud import WordCloud

crime_descriptions = ' '.join(crime_data['Description'].dropna())

# create a word cloud of the most common crime descriptions
wordcloud = WordCloud(background_color='white', max_words=200).generate(crime_descri

# plot the word cloud
plt.figure(figsize=(10, 8))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')

# display the chart
plt.show()
```

```
---------------------------------------------------------------------------
-
ModuleNotFoundError                       Traceback (most recent call las
t)
~\AppData\Local\Temp/ipykernel_5460/1826285184.py in <module>
----> 1 from wordcloud import WordCloud
      2
      3 crime_descriptions = ' '.join(crime_data['Description'].dropna())
      4
      5 # create a word cloud of the most common crime descriptions

ModuleNotFoundError: No module named 'wordcloud'
```

In [ ]:

```

```

In [ ]:

```

```