Crimes in Chicago Data Analysis

Group Members:

Apeksha Shetty

Shraddha Shahane

Bhavesh Yadav

Problem Statement: Analyzing the factors that affect the Crimes in Chicago based on five years data(2012-2017)

This dataset reflects reported incidents of crime (with the exception of murders where data exists for each victim) that occurred in the City of Chicago from 2012 to 2017. Data is extracted from the Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) system. In order to protect the privacy of crime victims, addresses are shown at the block level only and specific locations are not identified. This data includes unverified reports supplied to the Police Department.The dataset consists of 23 columns.

Parameters

ID - Unique identifier for the record.

Case Number - The Chicago Police Department RD Number (Records Division Number), which is unique to the incident.

Date - Date when the incident occurred. this is sometimes a best estimate.

Block - The partially redacted address where the incident occurred, placing it on the same block as the actual address.

Primary Type - The primary description of the IUCR code.

Description - The secondary description of the IUCR code, a subcategory of the primary description.

Location Description - Description of the location where the incident occurred.

Arrest - Indicates whether an arrest was made.

Domestic - Indicates whether the incident was domestic-related as defined by the Illinois Domestic Violence Act.

Updated On - Date and time the record was last updated.

Latitude - The latitude of the location where the incident occurred. This location is shifted from the actual location for partial redaction but falls on the same block.

Longitude - The longitude of the location where the incident occurred. This location is shifted from the actual location for partial redaction but falls on the same block.

Target Audience

We analyze this data from a perspective of a Chicago resident who wants to know more about their area so they can better navigate their way through the city and it can also help the Police Department to ensure the safety of the residents.

Questions:

(1)How has crime in Chicago changed across years?Is it Increasing or Decreasing?

(2)Which types of crimes are more likely to happen in specific locations or specific time of the day or specific day of the week than other types of crimes?

(3)Which crimes are most common among the top 20 most frequent crime types?

Importing Libraries

In [71]:

```
import numpy as np
from io import StringIO
import pandas as pd
import folium
import matplotlib.pyplot as plt
%matplotlib inline
plt.style.use('seaborn')
import seaborn as sns
```

Reading the dataset

In [79]:

```
crimes = pd.read_csv("C:\\Users\\APEKSHA\\Desktop\\Chicago_Crimes_2012_to_2017.csv")
crimes
```

Out[79]:

	Unnamed: 0	ID	Case Number	Date	Block	IUCR	Primary Type	Description	
0	3	10508693	HZ250496	05/03/2016 11:40:00 PM	013XX S SAWYER AVE	0486	BATTERY	DOMESTIC BATTERY SIMPLE	A
1	89	10508695	HZ250409	05/03/2016 09:40:00 PM	061XX S DREXEL AVE	0486	BATTERY	DOMESTIC BATTERY SIMPLE	R
2	197	10508697	HZ250503	05/03/2016 11:31:00 PM	053XX W CHICAGO AVE	0470	PUBLIC PEACE VIOLATION	RECKLESS CONDUCT	S
3	673	10508698	HZ250424	05/03/2016 10:10:00 PM	049XX W FULTON ST	0460	BATTERY	SIMPLE	S
4	911	10508699	HZ250455	05/03/2016 10:00:00 PM	003XX N LOTUS AVE	0820	THEFT	$500 AND UNDER	R
5	1108	10508702	HZ250447	05/03/2016 10:35:00 PM	082XX S MARYLAND AVE	041A	BATTERY	AGGRAVATED: HANDGUN	S
6	1130	10508703	HZ250489	05/03/2016 10:30:00 PM	027XX S STATE ST	0460	BATTERY	SIMPLE	C H
7	1801	10508704	HZ250514	05/03/2016 09:30:00 PM	002XX E 46TH ST	0460	BATTERY	SIMPLE	R
8	1868	10508709	HZ250523	05/03/2016 04:00:00 PM	014XX W DEVON AVE	0460	BATTERY	SIMPLE	S
9	1891	10508982	HZ250667	05/03/2016 10:30:00 PM	069XX S ASHLAND AVE	0486	BATTERY	DOMESTIC BATTERY SIMPLE	S
10	1935	10508710	HZ250469	05/03/2016 09:44:00 PM	074XX S SOUTH SHORE DR	143A	WEAPONS VIOLATION	UNLAWFUL POSS OF HANDGUN	V
				05/03/2016	006XX N				

	Unnamed: 0	ID	Case Number	Date	Block	IUCR	Primary Type	Description	
11	2150	10508715	HZ250541	11:11:00 PM	WABASH AVE	0486	BATTERY	DOMESTIC BATTERY SIMPLE	S
12	2193	10508717	HZ250415	05/03/2016 05:30:00 PM	011XX W JACKSON BLVD	0890	THEFT	FROM BUILDING	C
13	2279	10508724	HZ250513	05/03/2016 09:00:00 AM	028XX S DR MARTIN LUTHER KING JR DR	0820	THEFT	$500 AND UNDER	S
14	2477	10508728	HZ250505	05/03/2016 10:08:00 PM	016XX N CLAREMONT AVE	0810	THEFT	OVER $500	S
15	2847	10508732	HZ250535	05/03/2016 04:00:00 PM	072XX S RICHMOND ST	0486	BATTERY	DOMESTIC BATTERY SIMPLE	R
16	3023	10508738	HZ250440	05/03/2016 09:45:00 PM	020XX W LE MOYNE ST	0810	THEFT	OVER $500	S
17	3088	10508741	HZ250587	05/03/2016 10:00:00 PM	055XX S STATE ST	0313	ROBBERY	ARMED: OTHER DANGEROUS WEAPON	S
18	3242	10508747	HZ250577	05/03/2016 08:00:00 PM	100XX S SANGAMON ST	0910	MOTOR VEHICLE THEFT	AUTOMOBILE	S
19	3264	10508752	HZ250500	05/03/2016 11:00:00 PM	043XX S ELLIS AVE	0486	BATTERY	DOMESTIC BATTERY SIMPLE	A
20	3307	10508755	HZ250531	05/03/2016 11:37:00 PM	015XX N DAMEN AVE	0560	ASSAULT	SIMPLE	S
21	3416	10508757	HZ250601	05/03/2016 01:30:00 PM	057XX S COTTAGE GROVE AVE	0560	ASSAULT	SIMPLE	C
22	4333	10508987	HZ250698	05/03/2016 06:30:00 PM	025XX N GREENVIEW AVE	0810	THEFT	OVER $500	R
23	4377	10509011	HZ250748	05/03/2016 09:00:00 PM	068XX W HIGHLAND AVE	0820	THEFT	$500 AND UNDER	S
24	4571	10509016	HZ250610	05/03/2016 11:30:00 PM	003XX W HUBBARD ST	5002	OTHER OFFENSE	OTHER VEHICLE OFFENSE	S
25	4701	10509030	HZ250659	05/03/2016 07:00:00 AM	010XX W 95TH ST	0820	THEFT	$500 AND UNDER	V
26	5138	10509071	HZ250713	05/03/2016 08:30:00 PM	012XX W FULLERTON AVE	0820	THEFT	$500 AND UNDER	P L
27	5225	10509077	HZ250776	05/03/2016 12:01:00 AM	002XX W 33RD ST	1130	DECEPTIVE PRACTICE	FRAUD OR CONFIDENCE GAME	R
28	5423	10509078	HZ250611	05/03/2016 12:01:00 AM	043XX S SAWYER AVE	1153	DECEPTIVE PRACTICE	FINANCIAL IDENTITY THEFT OVER $ 300	R
29	5770	10509094	HZ250662	05/03/2016 11:00:00 PM	084XX S MORGAN ST	0910	MOTOR VEHICLE THEFT	AUTOMOBILE	S

...	Unnamed: 0	... ID	... Case Number	... Date	... Block	... IUCR	... Primary Type	... Description	..
1456684	6243883	10508639	HZ250458	05/03/2016 10:32:00 PM	037XX S HERMITAGE AVE	051A	ASSAULT	AGGRAVATED: HANDGUN	R
1456685	6245003	10508640	HZ250462	05/03/2016 10:07:00 PM	035XX W ROOSEVELT RD	1811	NARCOTICS	POSS: CANNABIS 30GMS OR LESS	S
1456686	6245288	10508641	HZ250498	05/03/2016 10:31:00 PM	066XX S WOLCOTT AVE	143A	WEAPONS VIOLATION	UNLAWFUL POSS OF HANDGUN	R
1456687	6245484	10508642	HZ250510	05/03/2016 10:45:00 PM	027XX W FLOURNOY ST	051A	ASSAULT	AGGRAVATED: HANDGUN	A
1456688	6245506	10508643	HZ250401	05/03/2016 09:00:00 PM	075XX S PARNELL AVE	0560	ASSAULT	SIMPLE	R
1456689	6245678	10508644	HZ250244	05/03/2016 07:13:00 PM	076XX S HALSTED ST	0460	BATTERY	SIMPLE	S
1456690	6245722	10508646	HZ250293	05/03/2016 07:45:00 PM	080XX S VERNON AVE	0460	BATTERY	SIMPLE	A
1456691	6246046	10508647	HZ250404	05/03/2016 08:56:00 PM	063XX S GREENWOOD AVE	0460	BATTERY	SIMPLE	S
1456692	6246219	10508648	HZ250445	05/03/2016 10:10:00 PM	073XX S MERRILL AVE	0520	ASSAULT	AGGRAVATED:KNIFE/CUTTING INSTR	A
1456693	6246370	10508649	HZ250442	05/03/2016 10:15:00 PM	041XX W CERMAK RD	4387	OTHER OFFENSE	VIOLATE ORDER OF PROTECTION	O
1456694	6246478	10508650	HZ250022	05/03/2016 05:00:00 PM	078XX S MARSHFIELD AVE	031A	ROBBERY	ARMED: HANDGUN	S
1456695	6246521	10508653	HZ250512	05/03/2016 11:58:00 PM	026XX W LE MOYNE ST	0520	ASSAULT	AGGRAVATED:KNIFE/CUTTING INSTR	R (F
1456696	6247563	10508656	HZ250476	05/03/2016 03:15:00 PM	014XX N OGDEN AVE	1720	OFFENSE INVOLVING CHILDREN	CONTRIBUTE DELINQUENCY OF A CHILD	S
1456697	6247798	10508658	HZ250506	05/03/2016 11:50:00 PM	018XX S KEDZIE AVE	4625	OTHER OFFENSE	PAROLE VIOLATION	S
1456698	6248016	10508659	HZ250499	05/03/2016 11:38:00 PM	038XX S PRINCETON AVE	0460	BATTERY	SIMPLE	C
1456699	6248192	10508661	HZ250344	05/03/2016 08:44:00 PM	070XX S WABASH AVE	041A	BATTERY	AGGRAVATED: HANDGUN	S
1456700	6248278	10508662	HZ250477	05/03/2016 08:00:00 AM	057XX S MICHIGAN AVE	5001	OTHER OFFENSE	OTHER CRIME INVOLVING PROPERTY	A
1456701	6248713	10508663	HZ250466	05/03/2016 10:10:00 PM	033XX W MARQUETTE RD	1563	SEX OFFENSE	CRIMINAL SEXUAL ABUSE	A
1456702	6248779	10508664	HZ250486	05/03/2016 11:35:00	083XX S CARPENTER	0560	ASSAULT	SIMPLE	R

	Unnamed: 0	ID	Case Number	Date	Block	IUCR	Primary Type	Description	
1456703	6248999	10508665	HZ250448	05/03/2016 10:15:00 PM	095XX S LOOMIS ST	1310	CRIMINAL DAMAGE	TO PROPERTY	R
1456704	6249417	10508666	HZ250497	05/03/2016 11:30:00 PM	053XX S PULASKI RD	0320	ROBBERY	STRONGARM - NO WEAPON	S
1456705	6249592	10508671	HZ250526	05/03/2016 11:50:00 PM	036XX E 106TH ST	502P	OTHER OFFENSE	FALSE/STOLEN/ALTERED TRP	A
1456706	6249615	10508672	HZ250441	05/03/2016 10:25:00 PM	071XX S MOZART ST	0460	BATTERY	SIMPLE	S
1456707	6249936	10508675	HZ250502	05/03/2016 11:00:00 PM	085XX S MAY ST	0320	ROBBERY	STRONGARM - NO WEAPON	S
1456708	6250154	10508678	HZ250481	05/03/2016 11:28:00 PM	088XX S LAFLIN ST	041A	BATTERY	AGGRAVATED: HANDGUN	S
1456709	6250330	10508679	HZ250507	05/03/2016 11:33:00 PM	026XX W 23RD PL	0486	BATTERY	DOMESTIC BATTERY SIMPLE	A
1456710	6251089	10508680	HZ250491	05/03/2016 11:30:00 PM	073XX S HARVARD AVE	1310	CRIMINAL DAMAGE	TO PROPERTY	A
1456711	6251349	10508681	HZ250479	05/03/2016 12:15:00 AM	024XX W 63RD ST	041A	BATTERY	AGGRAVATED: HANDGUN	S
1456712	6253257	10508690	HZ250370	05/03/2016 09:07:00 PM	082XX S EXCHANGE AVE	0486	BATTERY	DOMESTIC BATTERY SIMPLE	S
1456713	6253474	10508692	HZ250517	05/03/2016 11:38:00 PM	001XX E 75TH ST	5007	OTHER OFFENSE	OTHER WEAPONS VIOLATION	P L

1456714 rows × 23 columns

For simplicity, display subset of the datset using head() function

In [3]:

```
crimes.head()
```

Out[3]:

	Unnamed: 0	ID	Case Number	Date	Block	IUCR	Primary Type	Description	Location Description	Arrest	...	Ward	C
0	3	10508693	HZ250496	05/03/2016 11:40:00 PM	013XX S SAWYER AVE	0486	BATTERY	DOMESTIC BATTERY SIMPLE	APARTMENT	True	...	24.0	29
1	89	10508695	HZ250409	05/03/2016 09:40:00 PM	061XX S DREXEL AVE	0486	BATTERY	DOMESTIC BATTERY SIMPLE	RESIDENCE	False	...	20.0	42
2	197	10508697	HZ250503	05/03/2016 11:31:00 PM	053XX W CHICAGO AVE	0470	PUBLIC PEACE VIOLATION	RECKLESS CONDUCT	STREET	False	...	37.0	25
3	673	10508698	HZ250424	05/03/2016 10:10:00	049XX W FULTON	0460	BATTERY	SIMPLE	SIDEWALK	False	...	28.0	25

	Unnamed: 0	ID	Case Number	Date	Block	IUCR	Primary Type	Description	Location Description	Arrest	...	Ward	Co
3	673	10508698	HZ250424	10:10:00 PM	FULTON ST	0460	BATTERY	SIMPLE	SIDEWALK	False	...	28.0	25
4	911	10508699	HZ250455	05/03/2016 10:00:00 PM	003XX N LOTUS AVE	0820	THEFT	$500 AND UNDER	RESIDENCE	False	...	28.0	25

5 rows × 23 columns

To print the index of columns of the dataset

In [4]:

```
crimes.columns
```

Out[4]:

```
Index(['Unnamed: 0', 'ID', 'Case Number', 'Date', 'Block', 'IUCR',
       'Primary Type', 'Description', 'Location Description', 'Arrest',
       'Domestic', 'Beat', 'District', 'Ward', 'Community Area', 'FBI Code',
       'X Coordinate', 'Y Coordinate', 'Year', 'Updated On', 'Latitude',
       'Longitude', 'Location'],
      dtype='object')
```

Data Preprocessing and Cleaning

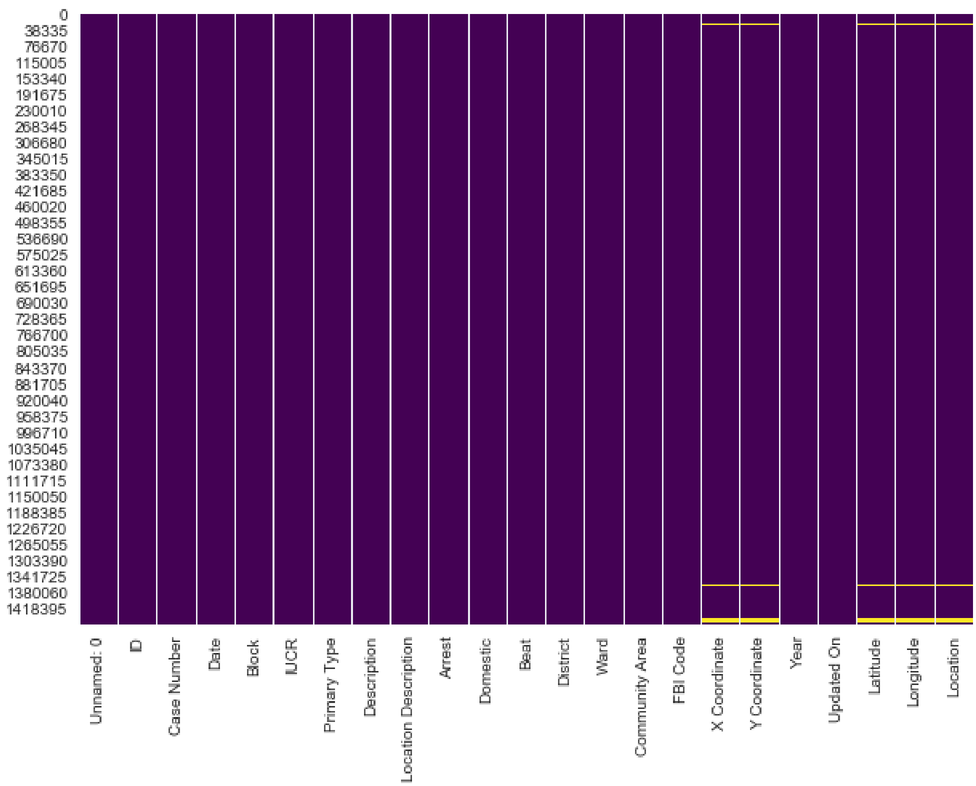
Determining null values in our dataset.

In [5]:

```
plt.figure(figsize=(10,7))
sns.heatmap(crimes.isnull(), cbar = False, cmap = 'viridis')
```

Out[5]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x2490dba2ef0>
```



We will not be using any of those columns in our analysis; so we filter them out.

In [6]:

```
crimes.drop(['Unnamed: 0', 'Case Number', 'IUCR', 'X Coordinate', 'Y Coordinate','Updated On','Year', 'FBI Code', 'Beat','Ward','Community Area', 'Location', 'District'], inplace=True, axis=1)
```

We will use the 'Date' column to explore temporal patterns, 'Primary Type' and 'Location Description' to investigate their relationship with time (month of the year, time of the day, hour of the day, .. etc).

We need to convert the 'Date' column into a date format that is understandable by Python (and pandas).

In [7]:

```
# convert dates to pandas datetime format
crimes.Date = pd.to_datetime(crimes.Date, format='%m/%d/%Y %I:%M:%S %p')
# setting the index to be the date will help us a lot later on
crimes.index = pd.DatetimeIndex(crimes.Date)
```

To display records and its corresponding features

In [8]:

```
crimes.shape
```

Out[8]:

```
(1456714, 10)
```

In [9]:

```
crimes.info()
```

```
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 1456714 entries, 2016-05-03 23:40:00 to 2016-05-03 23:38:00
Data columns (total 10 columns):
ID                      1456714 non-null int64
Date                    1456714 non-null datetime64[ns]
Block                   1456714 non-null object
Primary Type            1456714 non-null object
Description             1456714 non-null object
Location Description    1455056 non-null object
Arrest                  1456714 non-null bool
Domestic                1456714 non-null bool
Latitude                1419631 non-null float64
Longitude               1419631 non-null float64
dtypes: bool(2), datetime64[ns](1), float64(2), int64(1), object(4)
memory usage: 102.8+ MB
```

As 'Location Description', 'Description' and 'Primary Type' columns are actually categorical columns, we will only keep the most frequent categories and then cast them to a categorical type.

In [10]:

```
loc_to_change  = list(crimes['Location Description'].value_counts()[20:].index)
desc_to_change = list(crimes['Description'].value_counts()[20:].index)
#type_to_change = list(crimes['Primary Type'].value_counts()[20:].index)

crimes.loc[crimes['Location Description'].isin(loc_to_change) , crimes.columns=='Location Description'] = 'OTHER'
crimes.loc[crimes['Description'].isin(desc_to_change) , crimes.columns=='Description'] = 'OTHER'
#crimes.loc[crimes['Primary Type'].isin(type_to_change) , crimes.columns=='Primary Type'] = 'OTHER'
```

In [11]:

```
# we convert those 3 columns into 'Categorical' types -- works like 'factor' in R
crimes['Primary Type']          = pd.Categorical(crimes['Primary Type'])
crimes['Location Description'] = pd.Categorical(crimes['Location Description'])
crimes['Description']           = pd.Categorical(crimes['Description'])
```

```
crimes['Description']          = pd.Categorical(crimes['Description'])
```

Data Exploration and Visualization

In [12]:

```
pd.value_counts(crimes['Location Description'])[:10]
```

Out[12]:

```
STREET                            330471
RESIDENCE                         233530
OTHER                             202047
APARTMENT                         185023
SIDEWALK                          160891
PARKING LOT/GARAGE(NON.RESID.)     41768
ALLEY                              31771
RESIDENTIAL YARD (FRONT/BACK)      30645
SMALL RETAIL STORE                 28803
SCHOOL, PUBLIC, BUILDING           25959
Name: Location Description, dtype: int64
```

In [13]:

```
pd.value_counts(crimes['Primary Type'])[:10]
```

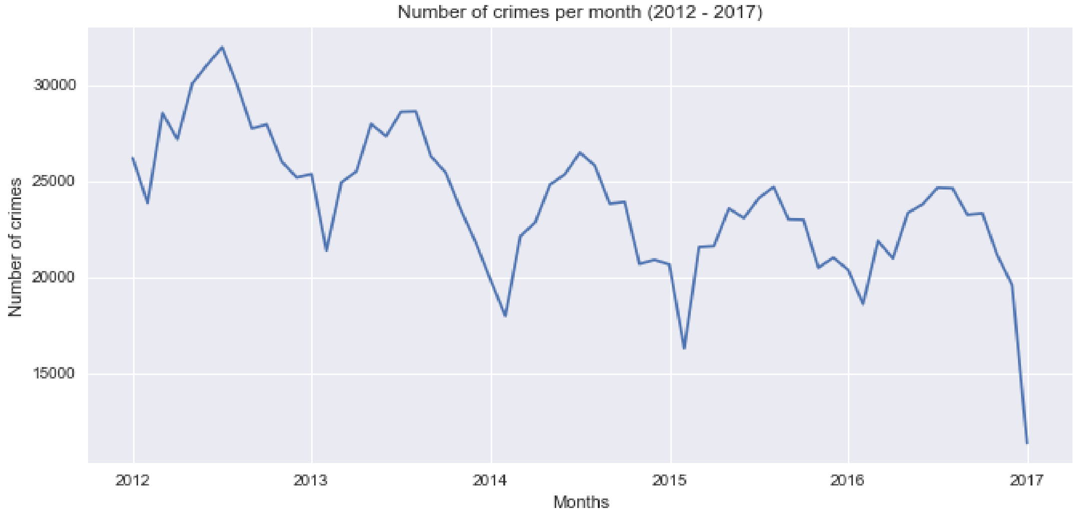
Out[13]:

```
THEFT                  329460
BATTERY                263700
CRIMINAL DAMAGE        155455
NARCOTICS              135240
ASSAULT                 91289
OTHER OFFENSE           87874
BURGLARY                83397
DECEPTIVE PRACTICE      75495
MOTOR VEHICLE THEFT     61138
ROBBERY                 57313
Name: Primary Type, dtype: int64
```

To plot the number of crimes per month

In [14]:

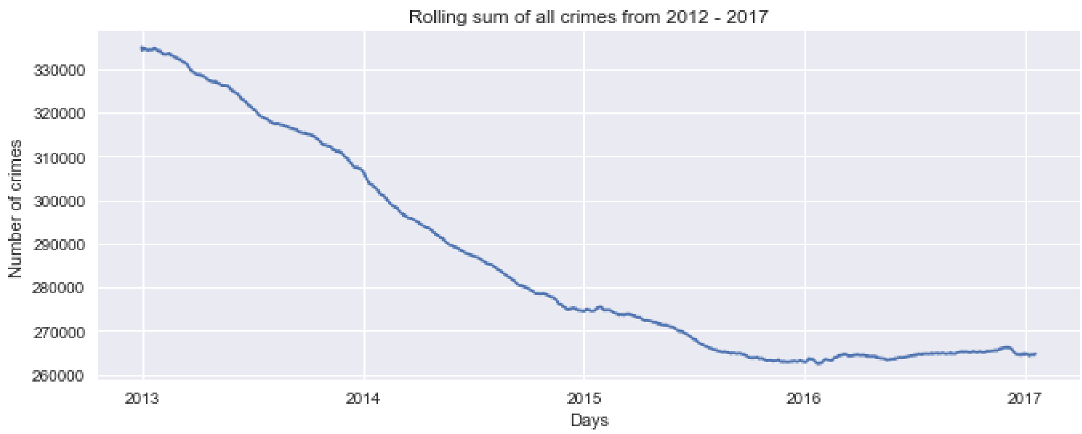
```
plt.figure(figsize=(11,5))
crimes.resample('M').size().plot(legend=False)
plt.title('Number of crimes per month (2012 - 2017)')
plt.xlabel('Months')
plt.ylabel('Number of crimes')
plt.show()
```



The previous graph shows monthly crime records.For more finer results,we take into consideration the rolling sum of crimes of the past year.The idea is, for each day, we calculate the sum of crimes of the past year. If this rolling sum is decreasing, then we know for sure that crime rates have been decreasing during that year. On the other hand, if the rolling sum stays the same during a given year, then we can conclude that crime rates stayed the same.

In [15]:

```
plt.figure(figsize=(11,4))
crimes.resample('D').size().rolling(365).sum().plot()
plt.title('Rolling sum of all crimes from 2012 - 2017')
plt.ylabel('Number of crimes')
plt.xlabel('Days')
plt.show()
```

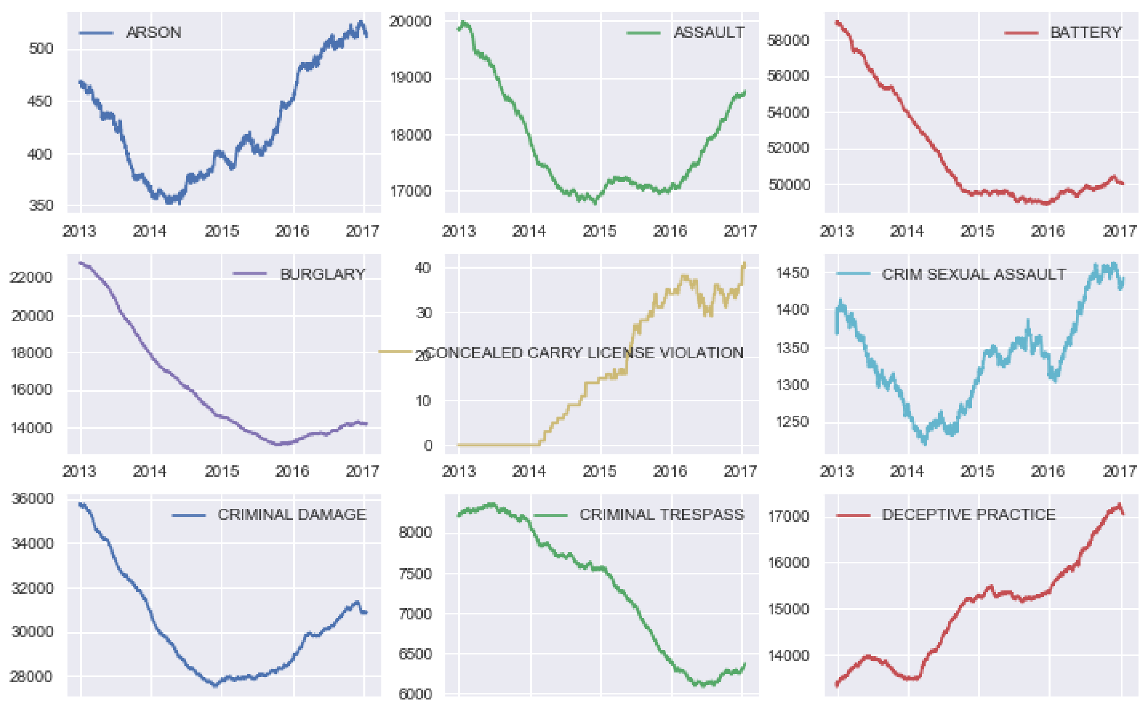


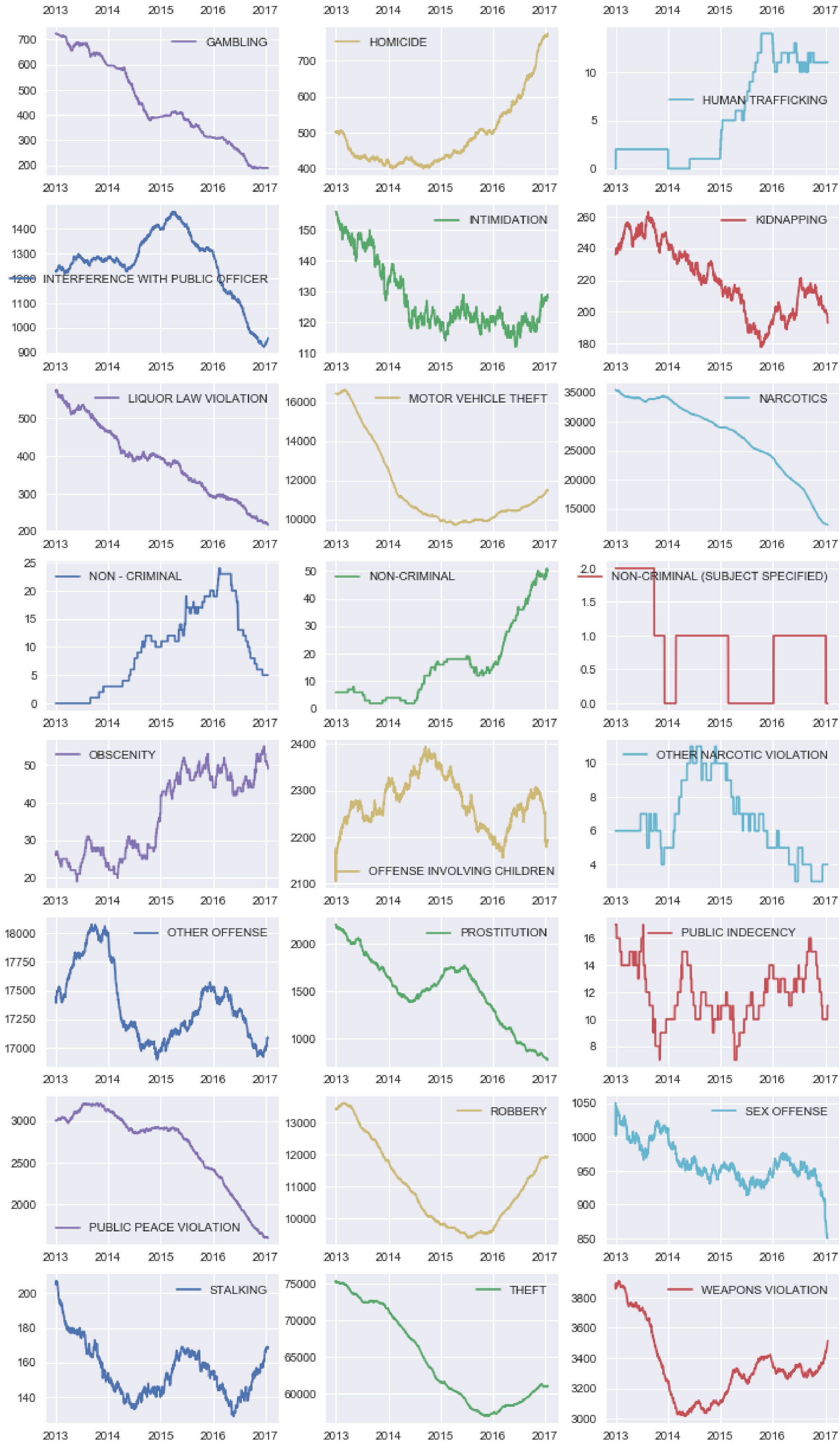
We see the line decreasing from 2013 up to some point around 2017 after which it stays around the same number of crimes. This all means that 2017 is really no better than 2016, but both years show a much better crime record (in total) than the previous years.

Separating crimes by type

In [16]:

```
crimes_count_date = crimes.pivot_table('ID', aggfunc=np.size, columns='Primary Type', index=crimes.
index.date, fill_value=0)
crimes_count_date.index = pd.DatetimeIndex(crimes_count_date.index)
plo = crimes_count_date.rolling(365).sum().plot(figsize=(12, 30), subplots=True, layout=(-1, 3), sh
arex=False, sharey=False)
```



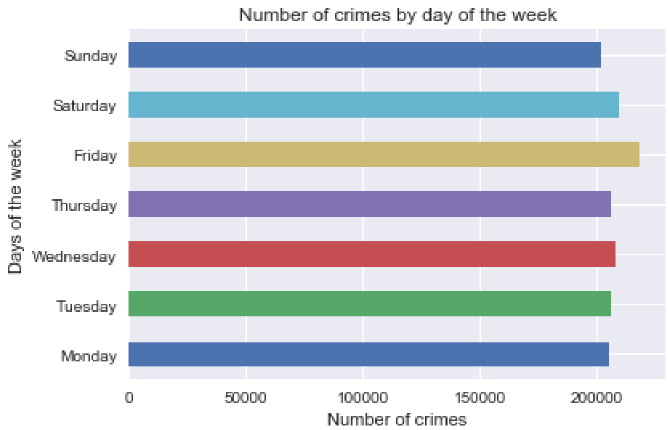


Differentiating crime records by time,type and location

Is there a difference in the number of crimes during specific days of the week.Are there more crimes during weekdays or weekend?

In [17]:

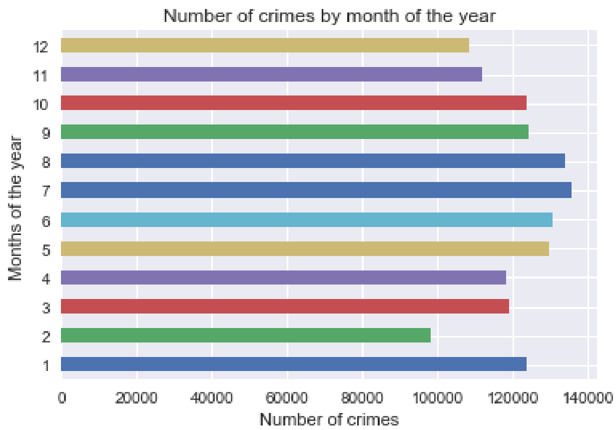
```
days = ['Monday','Tuesday','Wednesday',  'Thursday', 'Friday', 'Saturday', 'Sunday']
crimes.groupby([crimes.index.dayofweek]).size().plot(kind='barh')
plt.ylabel('Days of the week')
plt.yticks(np.arange(7), days)
plt.xlabel('Number of crimes')
plt.title('Number of crimes by day of the week')
plt.show()
```



Here we can see that the highest number of crimes occur on Friday.

In [18]:

```
crimes.groupby([crimes.index.month]).size().plot(kind='barh')
plt.ylabel('Months of the year')
plt.xlabel('Number of crimes')
plt.title('Number of crimes by month of the year')
plt.show()
```



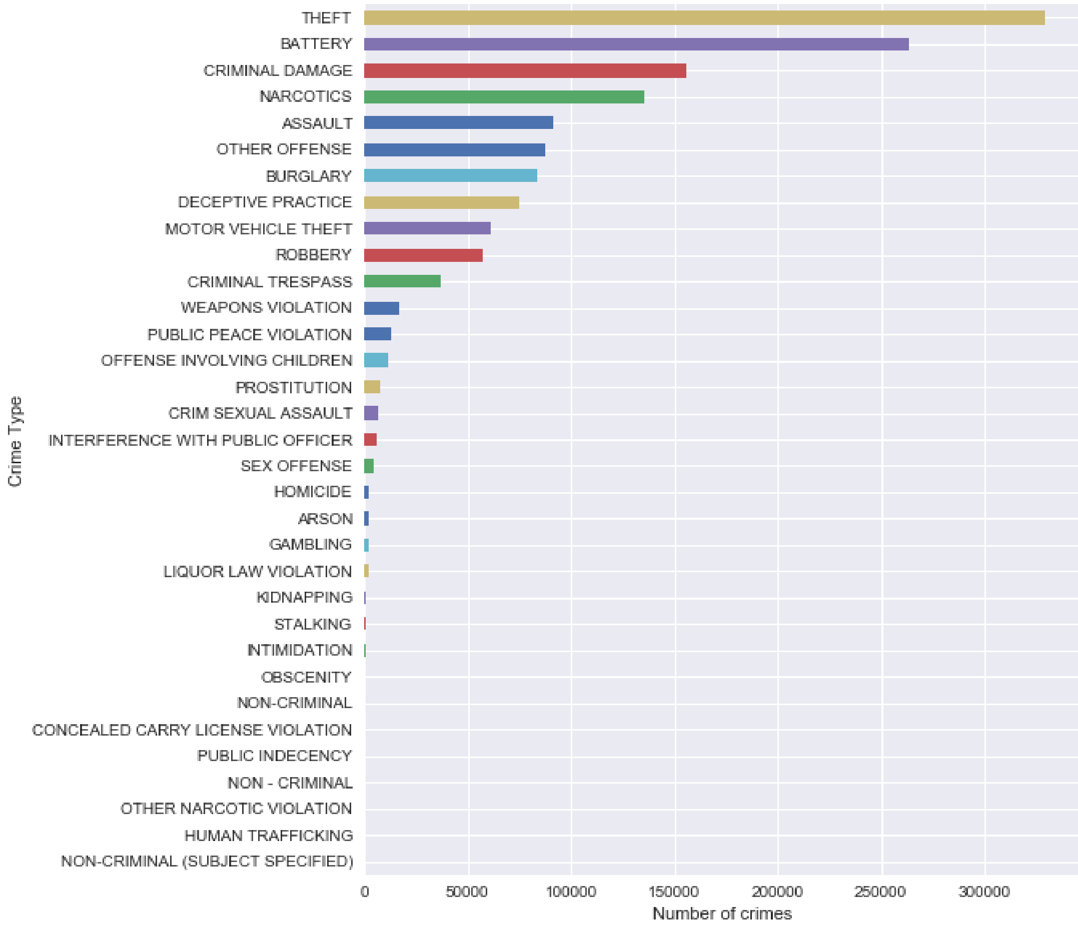
Here we can see that July has the highest no of crimes per month,

Which crimes are most common among the top 20 most frequent crime types?

In [19]:

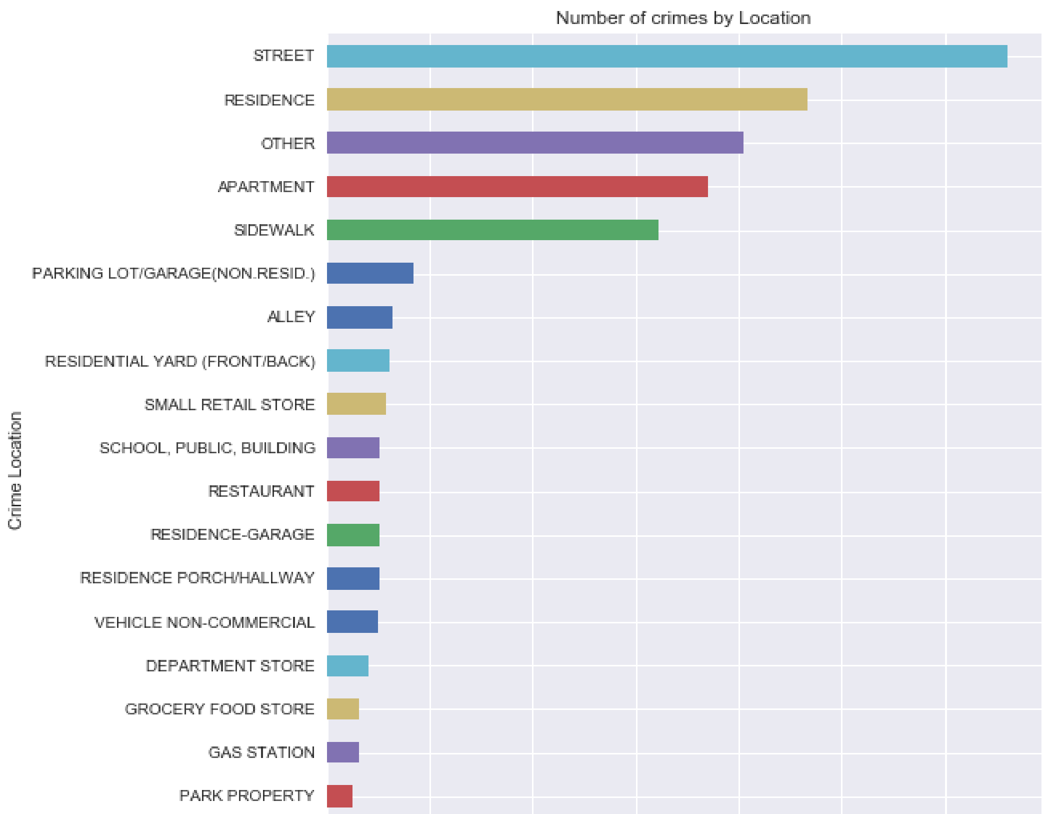
```
plt.figure(figsize=(8,10))
crimes.groupby([crimes['Primary Type']]).size().sort_values(ascending=True).plot(kind='barh')
plt.title('Number of crimes by type')
plt.ylabel('Crime Type')
plt.xlabel('Number of crimes')
plt.show()
```

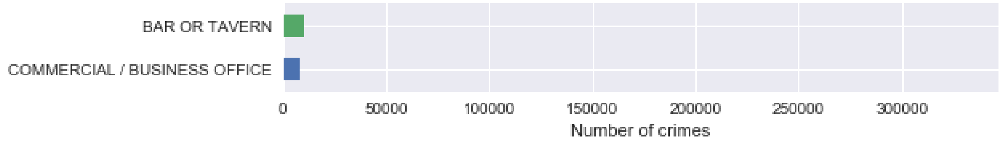
Number of crimes by type



In [20]:

```
plt.figure(figsize=(8,10))
crimes.groupby([crimes['Location Description']]).size().sort_values(ascending=True).plot(kind='bar
h')
plt.title('Number of crimes by Location')
plt.ylabel('Crime Location')
plt.xlabel('Number of crimes')
plt.show()
```





Creating Heatmaps for each crime to get an idea of where the crimes were generally commited.

In [86]:

```
x = crimes.sample(30000)
```

In [87]:

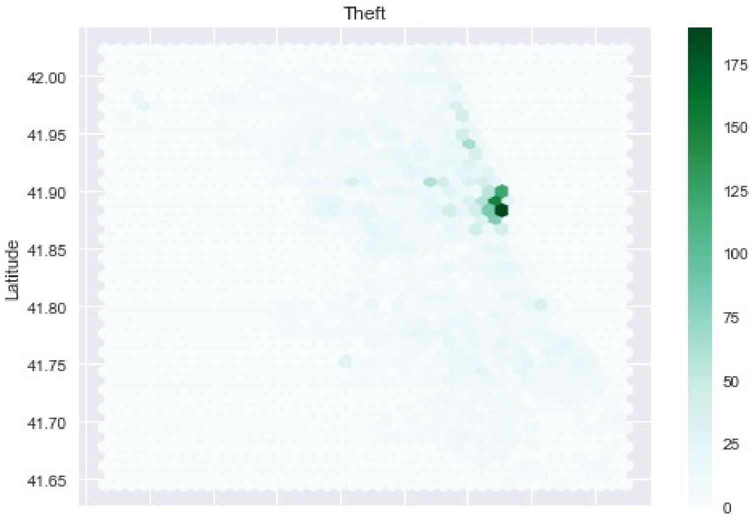
```
x_theft = x[x['Primary Type'] == "THEFT"]
x_battery = x[x['Primary Type'] == "BATTERY"]
x_cd = x[x['Primary Type'] == "CRIMINAL DAMAGE"]
x_narc = x[(x['Primary Type'] == "NARCOTICS")]
```

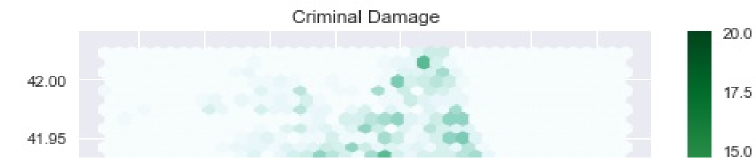
In [89]:

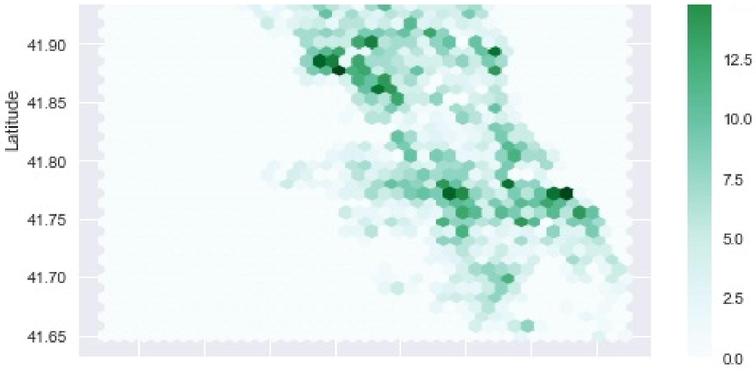
```
print("Heat map over cooridantes of crimes")
x.plot.hexbin(x='Longitude', y='Latitude', gridsize=40)
plt.title("Overall crimes")
plt.show()

x_theft.plot.hexbin(x='Longitude', y='Latitude', gridsize=40)
plt.title("Theft")

x_battery.plot.hexbin(x='Longitude', y='Latitude', gridsize=40)
plt.title("Battery")

x_cd.plot.hexbin(x='Longitude', y='Latitude', gridsize=40)
plt.title("Criminal Damage")
plt.show()

x_narc.plot.hexbin(x='Longitude', y='Latitude', gridsize=40)
plt.title("Narcotics")
plt.show()

x_narc.plot.hexbin(x='Longitude', y='Latitude', gridsize=40)
plt.title("Robbery")
plt.show()

x_narc.plot.hexbin(x='Longitude', y='Latitude', gridsize=40)
plt.title("Public Peace Violation")
plt.show()

x_narc.plot.hexbin(x='Longitude', y='Latitude', gridsize=40)
plt.title("Weapons Violation")
plt.show()

x_narc.plot.hexbin(x='Longitude', y='Latitude', gridsize=40)
plt.title("Motor Vehicle Theft")
plt.show()

x_narc.plot.hexbin(x='Longitude', y='Latitude', gridsize=40)
plt.title("Other Offense")
plt.show()

x_narc.plot.hexbin(x='Longitude', y='Latitude', gridsize=40)
plt.title("Deceptive Practice")
plt.show()

x_narc.plot.hexbin(x='Longitude', y='Latitude', gridsize=40)
plt.title("Offense Involving Children")
plt.show()

x_narc.plot.hexbin(x='Longitude', y='Latitude', gridsize=40)
plt.title("Sex Offense")
plt.show()
```

```python
x_narc.plot.hexbin(x='Longitude', y='Latitude', gridsize=40)
plt.title("Criminal Damage")
plt.show()
```

```
Heat map over cooridantes of crimes
```







Narcotics

Robbery

Public Peace Violation

Weapons Violation

Motor Vehicle Theft

Other Offense

Deceptive Practice

Analyzing thefts

In [22]:

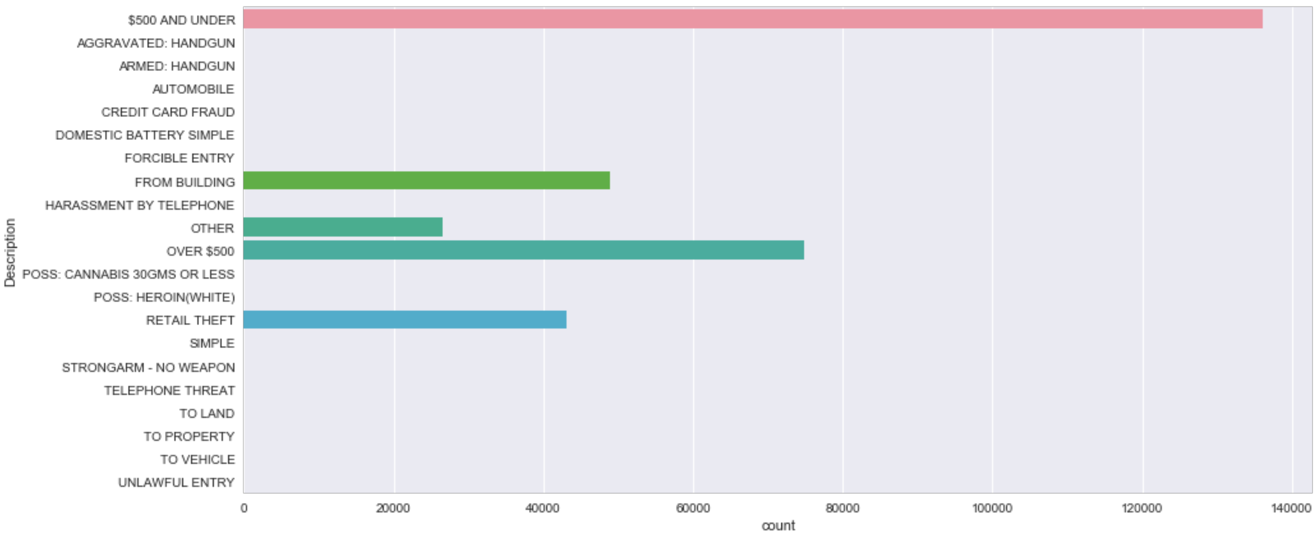
```
crimes_theft = crimes[crimes['Primary Type'] == 'THEFT']
```

In [23]:

```
plt.figure(figsize = (15, 7))
sns.countplot(y = crimes_theft['Description'])
```

Out[23]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x2490fb14320>
```



In [24]:

```
crimes_theft_data = pd.DataFrame({"Counts": crimes_theft['Description'].value_counts(),
"Description" : crimes_theft['Description'].value_counts().index})
```

In [25]:

```
crimes_theft_data.reset_index(inplace=True)
```

In [26]:

```
crimes_theft_data = crimes_theft_data.drop(columns=['index'], axis = 1)
crimes_theft_data.head()
```

Out[26]:

	Counts	Description
0	136036	$500 AND UNDER
1	74906	OVER $500
2	48835	FROM BUILDING
3	43109	RETAIL THEFT
4	26574	OTHER

In [27]:

```
%%time
crimes_theft['Date'] = pd.to_datetime(crimes_theft['Date'])
```

```
Wall time: 146 ms
```

```
C:\ProgramData\Anaconda3\lib\site-packages\ipykernel_launcher.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
```

```
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy
  """Entry point for launching an IPython kernel.
```

In [28]:

```
crimes_theft['Month'] = crimes_theft['Date'].apply(lambda x : x.month)
```

```
C:\ProgramData\Anaconda3\lib\site-packages\ipykernel_launcher.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html#indexing-view-versus-copy
  """Entry point for launching an IPython kernel.
```

In [29]:

```
theft_in_months = pd.DataFrame({"thefts" : crimes_theft['Month'].value_counts(), "month" : crimes_theft["Month"].value_counts().index}, index = range(12))
```

In [30]:

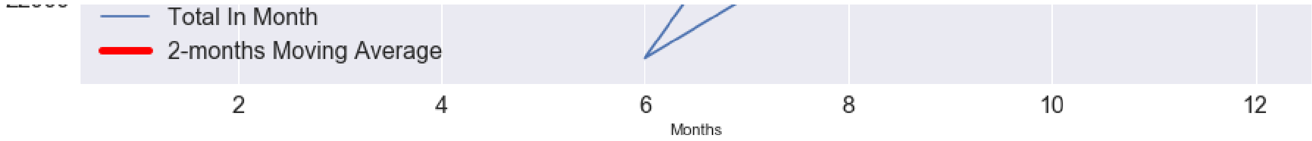
```
theft_in_months.head()
```

Out[30]:

	thefts	month
0	NaN	7
1	26982.0	8
2	20667.0	6
3	24702.0	9
4	25686.0	10

In [31]:

```
plt.figure(figsize = (15,7))
plt.plot(theft_in_months['month'],theft_in_months['thefts'], label = 'Total In Month')
plt.plot(theft_in_months['month'],theft_in_months['thefts'].rolling(window = 2).mean(),color='red', linewidth=5, label='2-months Moving Average' )

plt.title('Thefts per month', fontsize=16)
plt.xlabel('Months')
plt.legend(prop={'size':16})
plt.tick_params(labelsize=16);
```



In [32]:

```
print(max(crimes_theft['Date']))
print(min(crimes_theft['Date']))
```

```
2017-01-18 23:00:00
2012-01-01 00:00:00
```

In [33]:

```
crimes_theft['Date'].iloc[0].date()
```

Out[33]:

```
datetime.date(2016, 5, 3)
```

Conclusion

According to our analysis,

1.It can be concluded that the Crime has decreased over the years from 2012-2017.

2.Theft is the highest committed crime.

3.Weekdays encountered more number of crimes as compared to weekends.

4.It is observed that crimes committed were highest on street and lowest on Business/Commercial Office.