Introduction of the problem and a discussion of the background.

Problem: Chicago Police Department facing funding cuts

Chicago is one of the largest cities of Illinois, and is widely believed to be one of the most dangerous cities in the US. Although Chicago accounts for only 0.8% if the US population (roughly 2.7 million), Chicago accounts for nearly half of the increase in 2016's homicides (Sanburn, 2016). Especially in the category of violent crimes, it leads the national ranking with numbers well above the national average (City-Data.com, 2018).

Target Audience: Chicago Police Department and policy makers

The study can help the CPD and the policy maker of the city better understand the problems the city's police officers deal with, and better allocate the resources in the future. It can also help the public better rely on the resources that the police department is spending.

Who and why should they care?

Although measure have been taken to curtail this rampant increase, a large portion of the efforts have been towards spending more on the police budget. The most recent numbers suggest that Chicago spends nearly 1.6 billion dollars on just over 13,000 police officers which work out to be about 123,000 dollars per personnel. This comes under the 2.7 million dollars that the city spends on public safety. And yet the number have not been promising over the years.

But the solution of more and specialized police has not been able to tackle the rise in crime over the last decade. Mayor Rahm Emanuel disbanded the Chicago Police Department's anti-gang unit in 2012 in order to focus on beat patrols, which he said would have a more long-term solution to violence than anti-gang units (Wbez.org, 2012). This has been largely criticized.

Owing to the recent support for redistributing police funds to more appropriate segments of public safety, the Chicago Police department would be facing budget cuts in the future, which the department has resisted in the past. With a corporate fund of only 4.5 billion dollars (2020), the move however seems inevitable (chicago.gov, 2020).

This however does not guarantee a safer neighborhood or a reduction in the crime rates. So, moving forward, the police department will have to come up with more prudent ways to spend their fund.

Discussion: Better distribution of police funds to tackle crime

The objective is to study the kind of crimes occurring in the city, their location and see if the police can be better prepared to tackle the rise in crime. The Study undertaken should also recommend the kind of policing required and the timings that they would be most in demand. This can enable the department to be better prepared in terms of the kind of personnel to train and the facilitate better scheduling for the available ones.

This study can prove valuable not just to the city of Chicago, but also form as a better means of reallocating the funds across the state or even on a federal level to better plan and prepare the officers of the law. The study could focus also on the easy interpretation of its findings so as to convince the boards and the government to enable easy and fast approvals for the decisions made in the future.

On terms of simple changes being suggested, the study could focus on the location data to let officers know about the places that need more resources. The output of the study could act as recommender system for the resources at the department's disposal.

In [ ]:

Description of the data and how it will be used to solve the problem.

Data: https://data.cityofchicago.org/Public-Safety/Crimes-2019/w98m-zvie

The data used in the study has been gathers by 'data.cityofchicago.org'. This data will be leverage to better understand the location of the crime using the foursquare API. This will be crucial in classifying the location based on the type of crimes. **Eg: do more fights break-out near nightclubs? Are there more DUIs near a bar?**

Examples of the data:

- ID: 12109210
- Case number: JD301317
- Date: 04/30/2019 02:53:00 AM
- Block: 064XX S BISHOP ST
- ICUR: 1752
- Primary type: OFFENSE INVOLVING CHILDREN
- Description: AGGRAVATED CRIMINAL SEXUAL ABUSE BY FAMILY MEMBER
- Location: APARTMENT
- Arrest: false
- Domestic: true

- Beat: 0725
- District: 007
- Ward: 16
- Community Area: 67
- AFBI Code: 17
- X-Coordinates: 1167773
- Y-Coordinates: 1862010
- Year: 2019
- Updated-on: 10/20/2020 03:56:59 PM
- Latitude: 41.776897256
- Longitude: -87.660498495
- Location: "(41.776897256, -87.660498495)"

It has been recently updated on the 24th of October with nearly 260,000 records. The website that hosts the data is the government and the source for the data has been cited to be the Chicago Police department with the following disclaimer: These crimes may be based upon preliminary information supplied to the Police Department by the reporting parties that have not been verified. The website also provides information about research help through RandD@chicagopolice.org. The data is updated from Tuesday to Sunday and has had nearly 3400 downloads.

The city of Chicago has been able to maintain a constant record of crime statistics over the past decade, however the changing nature of the crime space, has limited the study to be effectively conducted only with the data from the last year which is the most recent information about the crime statistics. This has been proven useful to researches who study crimes in these cities to better equip the police and law makers of the kind of policies that are to be made.

A brief introduction into the data:

1: **ID** Unique identifier for the record (Number)

2: **Case Number** The Chicago Police Department RD Number (Records Division Number), which is unique to the incident (Plain Text)

3: **Date** Date when the incident occurred. this is sometimes a best estimate (Date & Time)

4: **Block** The partially redacted address where the incident occurred, placing it on the same block as the actual address (Plain Text)

5: **IUCR** The Illinois Unifrom Crime Reporting code. This is directly linked to the Primary Type and Description. See the list of IUCR codes at https://data.cityofchicago.org/d/c7ck-438e (Plain Text)

6: **Primary Type** The primary description of the IUCR code (Plain Text)

7: **Description** The secondary description of the IUCR code, a subcategory of the primary description (Plain Text)

8: **Location Description** Description of the location where the incident occurred (Plain Text)

9: **Arrest** Indicates whether an arrest was made (Checkbox)

10: **Domestic** Indicates whether the incident was domestic-related as defined by the Illinois Domestic Violence Act (Checkbox)

11: **Beat** Indicates the beat where the incident occurred. A beat is the smallest police geographic area – each beat has a dedicated police beat car. Three to five beats make up a police sector, and three sectors make up a police district. The Chicago Police Department has 22 police districts. See the beats at https://data.cityofchicago.org/d/aerh-rz74 (Plain Text)

12: **District** Indicates the police district where the incident occurred. See the districts at https://data.cityofchicago.org/d/fthy-xz3r (Plain Text)

13: **Ward** The ward (City Council district) where the incident occurred. See the wards at https://data.cityofchicago.org/d/sp34-6z76 (Number)

14: **Community Area** Indicates the community area where the incident occurred. Chicago has 77 community areas. See the community areas at https://data.cityofchicago.org/d/cauq-8yn6 (Plain Text)

15: **FBI Code** Indicates the crime classification as outlined in the FBI's National Incident-Based Reporting System (NIBRS). See the Chicago Police Department listing of these classifications at http://gis.chicagopolice.org/clearmap_crime_sums/crime_types.html (Plain Text)

16: **X Coordinate** The x coordinate of the location where the incident occurred in State Plane Illinois East NAD 1983 projection. This location is shifted from the actual location for partial redaction but falls on the same block (Number)

17: **Y Coordinate** The y coordinate of the location where the incident occurred in State Plane Illinois East NAD 1983 projection. This location is shifted from the actual location for partial redaction but falls on the same block (Number)

18: **Year** Year the incident occurred (Year)

19: **Updated On** Date and time the record was last updated (Date & Time)

20: **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 (Number)

21: **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 (Number)

22: **Location** The location where the incident occurred in a format that allows for creation of maps and other geographic operations on this data portal. This location is shifted from the actual location for partial redaction but falls on the same block (Location)

Features Extractable / Solution capability:

- The data consists of the time, location and the type of crime One of the key areas of exploration will be the location. The partial location from the block and the ward can be easily visualized to better understand the problematic areas of the city.
- Time data can explore into better scheduling for the city. This will help better plan the city's response with adequate number of officers.
- These two data can then be seen together to see if there is any co-relation between the time of day/month and the location in which these crimes occur. This will equip the police department to better allocate the resources at hand efficiently.
- The description of the crime is another area of exploration. Something simple as a word cloud could provide inside into the type of crime beyond just the title, and also describe the frequency with which these crimes occur. The type of personnel to deal with these crimes can then be trained in the future.
- Find other simple understand insights.

Methodology

The goal is to see the co-relation bewteen **places** the crime occurs, the **time** when they occur and the **similarity of the crimes** themselves.

This way, the department can start to spend less on the similar resources and start to increase spending on personnel/resoruces that are specialized to the tasks that can handle more problems in the city. An overview of where the crime is concentrated and during which time is expected. This can lead to better planning of the avaliable resources by the department.

Steps to achieve the goals

The dataframe is assumed to cleaned before the start of the tasks.

- ### Find co-relation between location and crime

The location data (both latitude and longitude) and the crime type (called as the primary type) will have to be co-related. this can be done by using the map data. The Idea is to use the data available tovisualize the crime on the maps themselves so as to start performing more complicated tasks on them.

- ### Find co-relation between time of day/month and crime

The dataframe is assumed to have easily accessible data about the time/day/month of the recorded crime and the primary type of the crime. This can be tabulated to see if there is any co-relation between the data. One important point of investigation is to check if the time of the month (beginning or end) has more crime comparted to teh other.

- ### Find the co-relation between the type of crime and the locations in which these crimes occur

The resources allocated will have to be managed according to the needs of the city. This can be improved if the data can show some co-relation between the location of the crime and crimes themselves. Simple query like: Is th eplace next to teh bar more or less likely to have DUIs. This will also have to be visualized.

- ### Find similarities between the crime themselves

The crimes themselves could have a pattern. To to chech this the crimes will have to be clustered into different groups. To perfrom this K-Means method will have to be leveraged. The number of clusters will have to be iteratively found.

- ### Any additional insight into the crime

Any suggestions for teh improved of better resource handling is also explored.

Importing the required packages in Python

In [1]:

```
import pandas as pd
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)
from pandasgui import show

import numpy as np
import requests
from bs4 import BeautifulSoup
from IPython.display import Image
```

```
import json
from wordcloud import WordCloud, STOPWORDS

from geopy.geocoders import Nominatim
from pandas.io.json import json_normalize
import matplotlib.cm as cm
import matplotlib.colors as colors
import matplotlib.pyplot as plt

from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler

import folium

print('Importing completed!')
```

```
Importing completed!
```

In [2]:

```
df = pd.read_csv('Crime.csv')
#df = pd.read_csv('https://ibm.box.com/shared/static/svflyugsr9zbqy5bmowgswqemfpm1x7f.csv')

#Dropping unusable data points and resetting the index
df.dropna(inplace=True)
df.reset_index(inplace=True)
df.head(3)
```

Out[2]:

	index	ID	CASE_NUMBER	DATE	BLOCK	IUCR	PRIMARY_TYPE	DESCRIPTION	LOCATION_DESCRIPTION	ARREST	DOMESTIC	B
0	0	3512276	HK587712	08/28/2004 05:50:56 PM	047XX S KEDZIE AVE	890	THEFT	FROM BUILDING	SMALL RETAIL STORE	False	False	
1	1	3406613	HK456306	06/26/2004 12:40:00 PM	009XX N CENTRAL PARK AVE	820	THEFT	$500 AND UNDER	OTHER	False	False	1
2	2	8002131	HT233595	04/04/2011 05:45:00 AM	043XX S WABASH AVE	820	THEFT	$500 AND UNDER	NURSING HOME/RETIREMENT HOME	False	False	

Splitting the data and time from one column intot individual columns.

This was proved to be helpful during the visualization of different co-relations

In [3]:

```
df['DATES'] = pd.to_datetime(df['DATE']).dt.date
df['WEEKDAY'] = pd.to_datetime(df['DATE']).dt.day_name()
df['HOUR'] = pd.to_datetime(df['DATE']).dt.hour
df['YEAR'] = pd.DatetimeIndex(df['DATES']).year
df['MONTH'] = pd.DatetimeIndex(df['DATES']).month
df['DAY'] = pd.DatetimeIndex(df['DATES']).day
df['TIME'] = pd.to_datetime(df['DATE']).dt.time
df.drop('DATE', axis=1, inplace=True)
df.head(3)
```

Out[3]:

	index	ID	CASE_NUMBER	BLOCK	IUCR	PRIMARY_TYPE	DESCRIPTION	LOCATION_DESCRIPTION	ARREST	DOMESTIC	BEAT	DISTRIC
0	0	3512276	HK587712	047XX S KEDZIE AVE	890	THEFT	FROM BUILDING	SMALL RETAIL STORE	False	False	911	
1	1	3406613	HK456306	009XX N CENTRAL PARK AVE	820	THEFT	$500 AND UNDER	OTHER	False	False	1112	1
2	2	8002131	HT233595	043XX S WABASH AVE	820	THEFT	$500 AND UNDER	NURSING HOME/RETIREMENT HOME	False	False	221	

In [4]:

```
df.shape # Overall shape of the dataframe
```

Out[4]:

```
(487, 28)
```

Map of Chicago with all crime superimposed

The latitude and longitude of Chicago was obtained using the geocode package

In [5]:

```
address = 'Chicago, IL'

geolocator = Nominatim(user_agent="to_explorer")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
latitude, longitude
```

Out[5]:

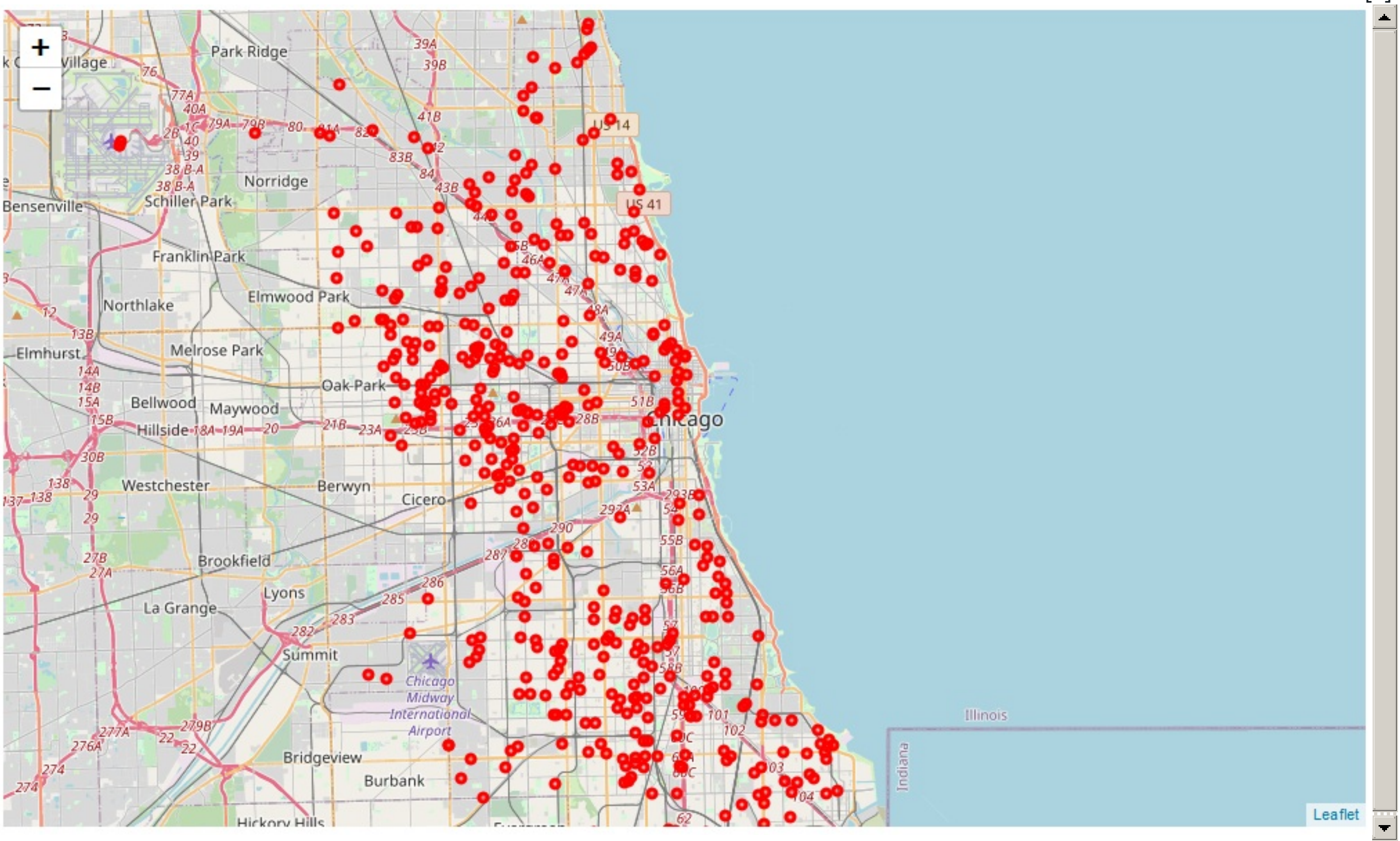
```
(41.8755616, -87.6244212)
```

The map was made using Folium

In [6]:

```
map_chicago = folium.Map(location=[latitude, longitude], zoom_start=10.5)

for lat, lng, label in zip(df['LATITUDE'], df['LONGITUDE'], df['IUCR']):
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=3,
        popup=label,
        color='red',
        fill=True,
        fill_color='red',
        fill_opacity=0.3,
        parse_html=False).add_to(map_chicago)

map_chicago # Can be sepearetely printed, but for easy of readability in GitHub this methid was chosen
Image(filename = "map.png", width = 1000, height = 600)
```

Out[6]:



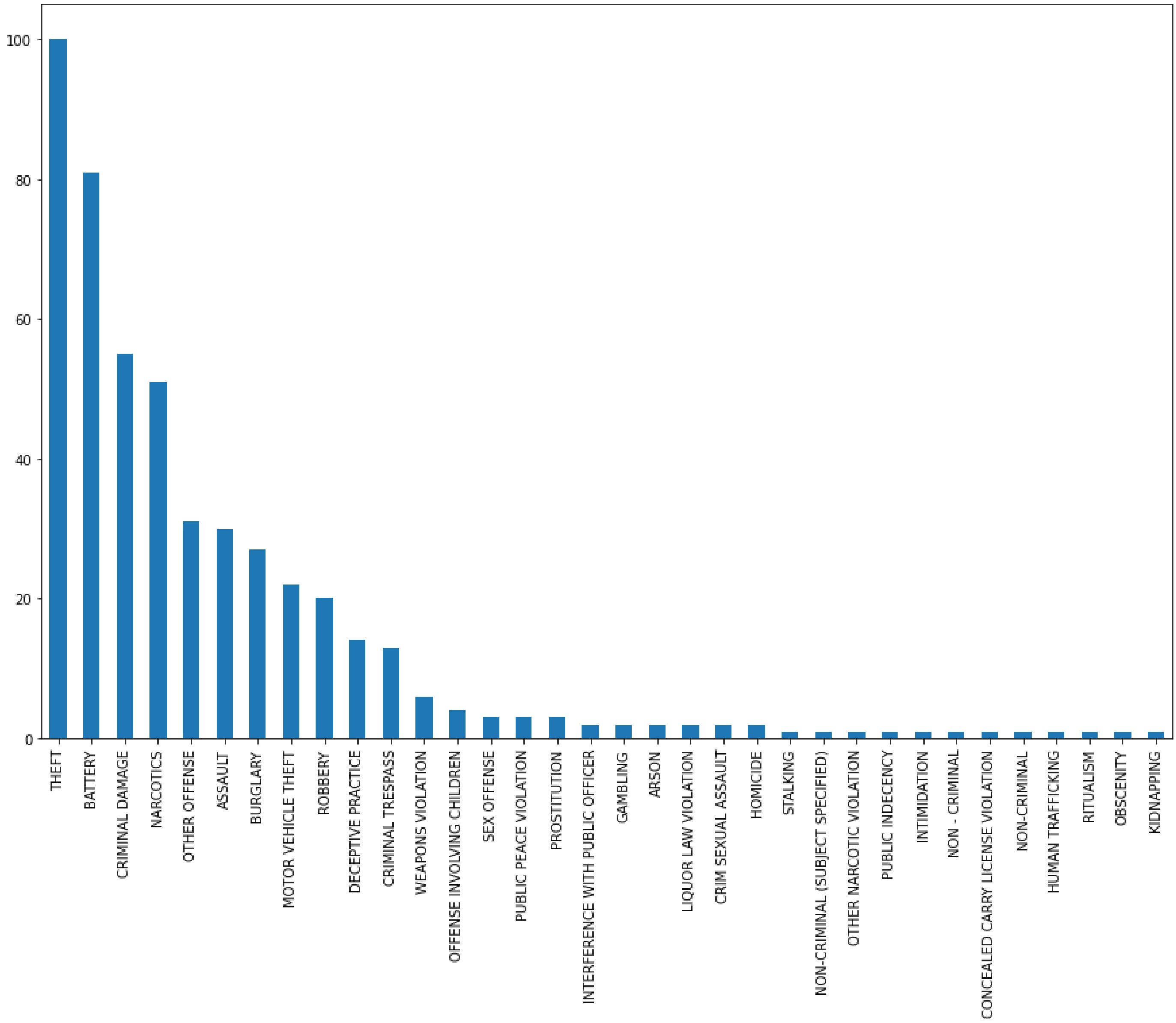
Finding out about the count of individual crimes commited (bar graph)

In [7]:

```
df['PRIMARY_TYPE'].value_counts().plot(kind='bar', figsize=(15,10))
```

Out[7]:

```
<AxesSubplot:>
```



Very apparent that the top four crimes account for the majority of the crimes commited, with theft leading the chart by a large margin. The resources available there for can be pivoted towards preventing these crimes through regular patrols. It is also to note that three of the top four are essentially victimless although danmage to public property is incurred.

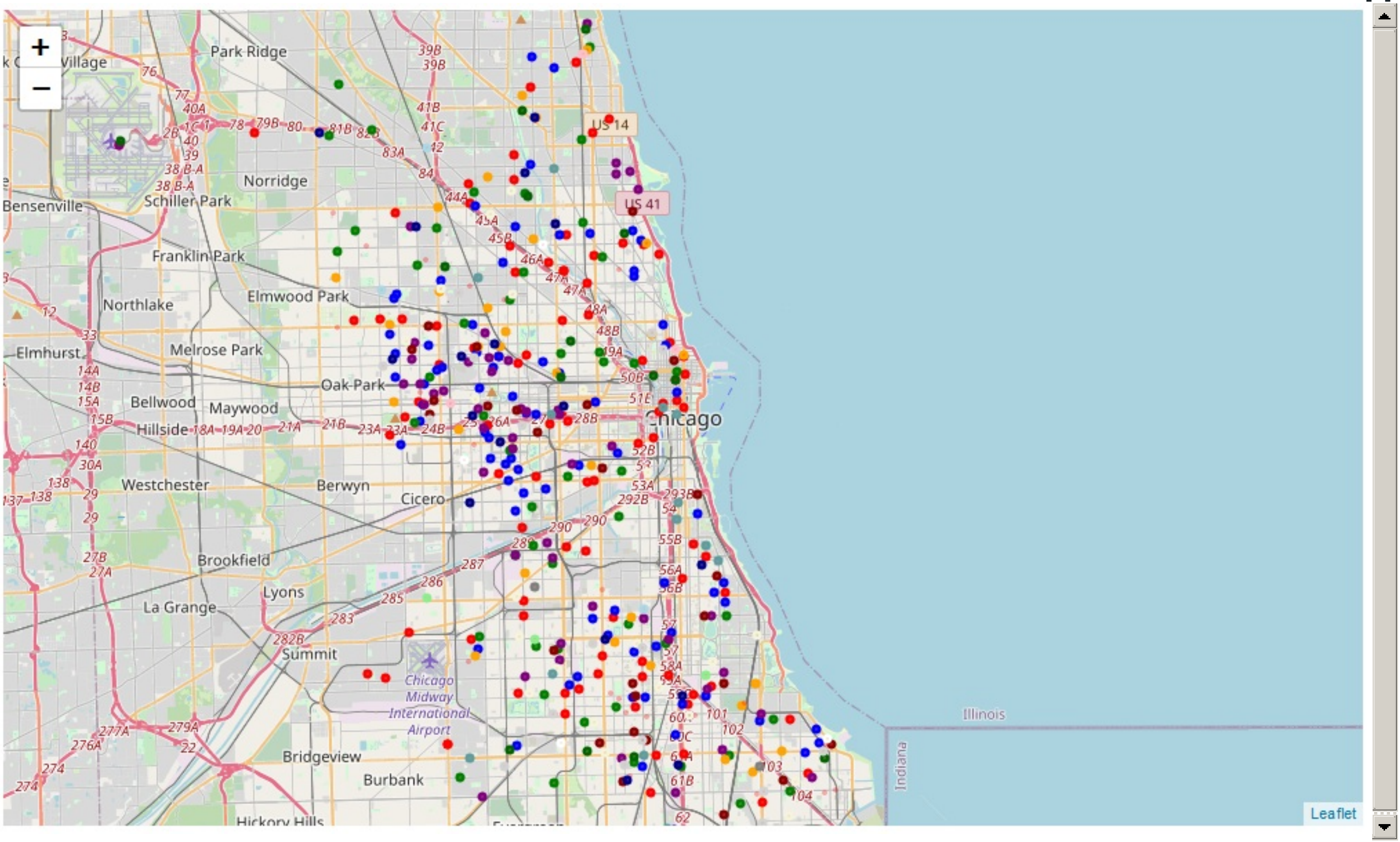
Classifying crime on the map to privde insight into the relatioship between the type of crime and the location

In [8]:

```
map_chicago = folium.Map(location=[latitude, longitude], zoom_start=10.5)

p_type = list(df[['PRIMARY_TYPE']].groupby('PRIMARY_TYPE').count().index)
colors_array = cm.rainbow(np.linspace(0, 1, len(p_type)))
rainbow = [colors.rgb2hex(i) for i in colors_array]

for lat, lng,PT in zip(df['LATITUDE'], df['LONGITUDE'], df['PRIMARY_TYPE']):
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=2,
        popup=PT,
```

```
        color=rainbow[p_type.index(PT)],
        fill=True,
        fill_color='red',
        fill_opacity=0.3,
        parse_html=False).add_to(map_chicago)

map_chicago
Image(filename = "crimemap.png", width = 1000, height = 600)
```

Out[8]:




```
It is found that the secluded areas are more prone to theft. Of note are:
    * The airport surroundings
    * The North-east outer side ofthe city and
    * The US 42 HW along th waterfront

These areas are found to be more prone to theft. As theft is one of the highest number of
recorded crimes and there is a visible cluster, the resources available ought to be spend on
securing these areas. Regular surveillance and patrol are some of the suggestions to reduce the
occurances of theft in these areas.
```

A better view of the crimes relative to the locational data is given below

In [9]:

```
Image(filename = "crime_map.png", width = 1000, height = 600)
```

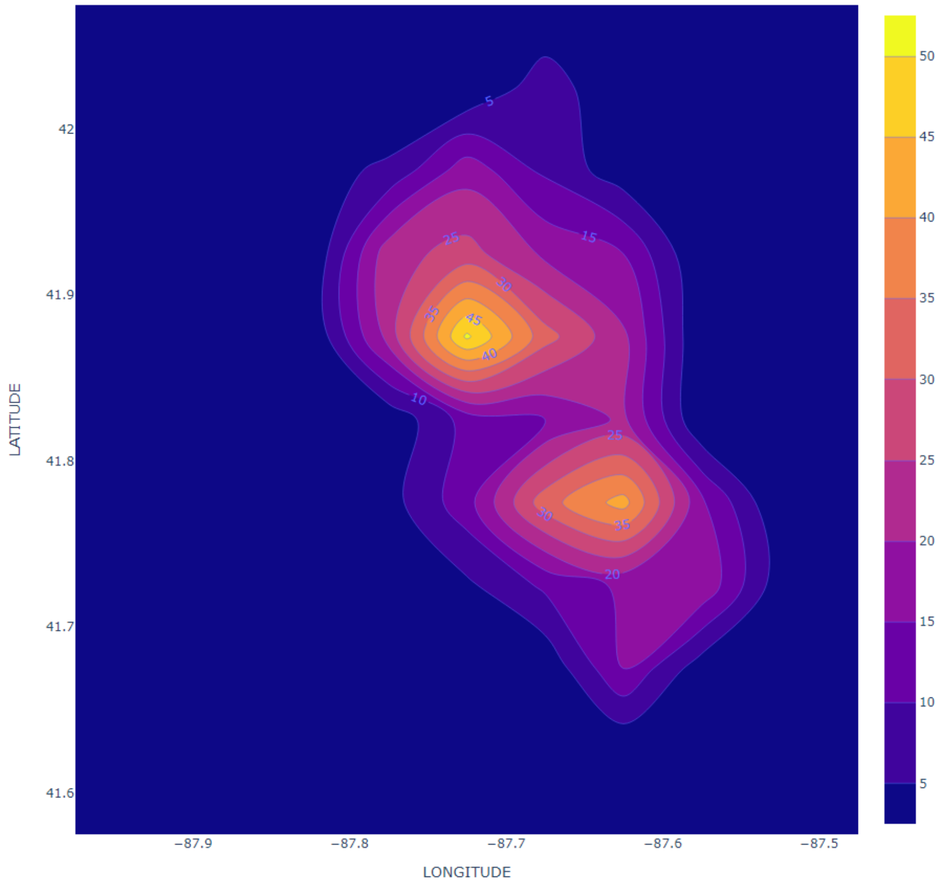
Out[9]:

It is found that there are two clusters city where there are crime is highly concentrated.

In [10]:

```
Image(filename = "crime_numbers.png", width = 600, height = 600)
```

Out[10]:



Oak park is however the most crime prone area in the city. With an average of 35% of the crimes occuring there. A simple suggestion would be to increase the outreach programs in this neighborhood. Further investivation leads to believe these are more gang related activities. This is a perfect area for incorporating crime/gangs task force.

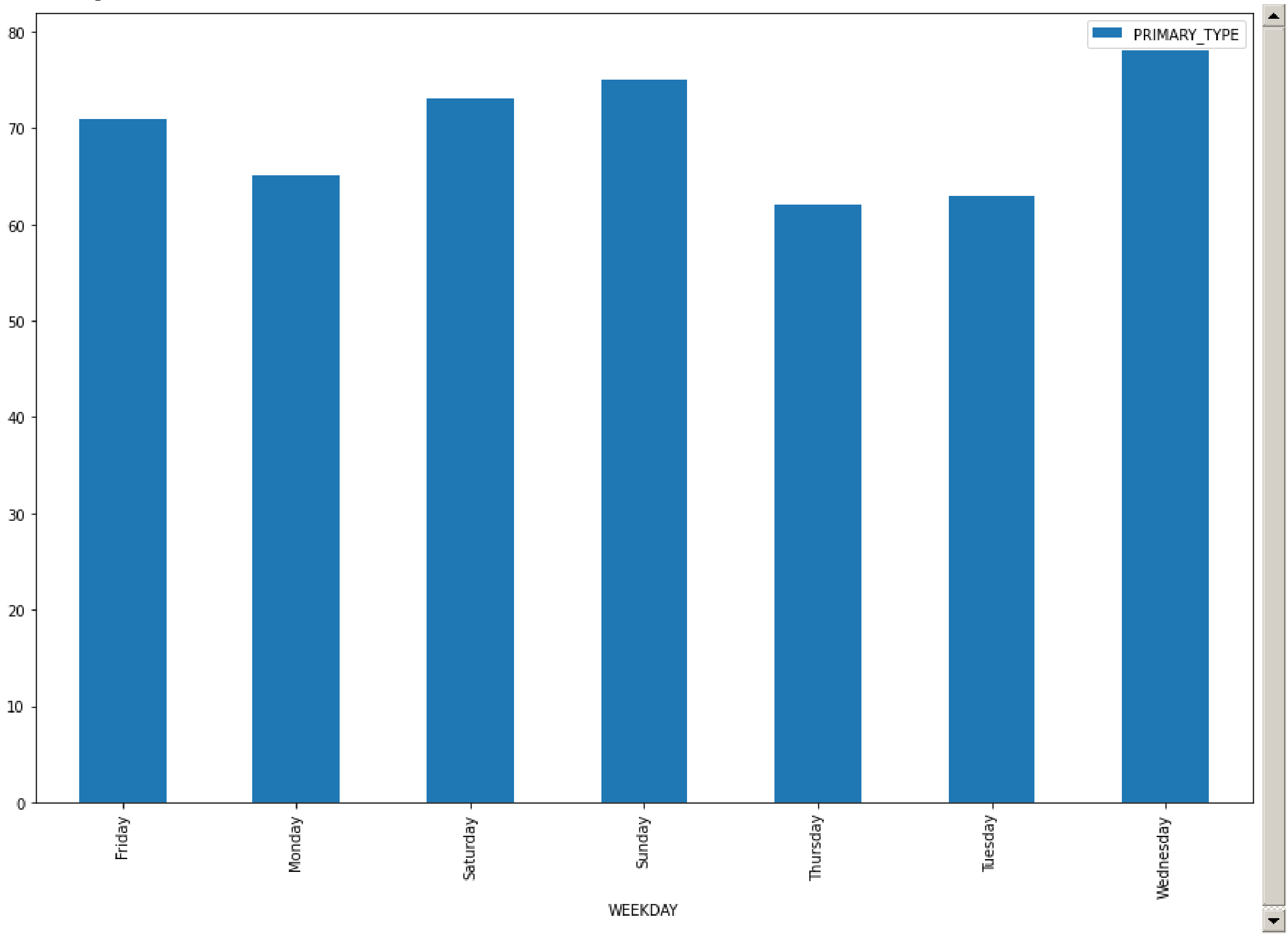
Co-relation between time of day/month/week and the crimes commited

In [11]:

```
df[['WEEKDAY','PRIMARY_TYPE']].groupby('WEEKDAY').count().plot(kind='bar', figsize=(15,10))
```

Out[11]:

```
<AxesSubplot:xlabel='WEEKDAY'>
```



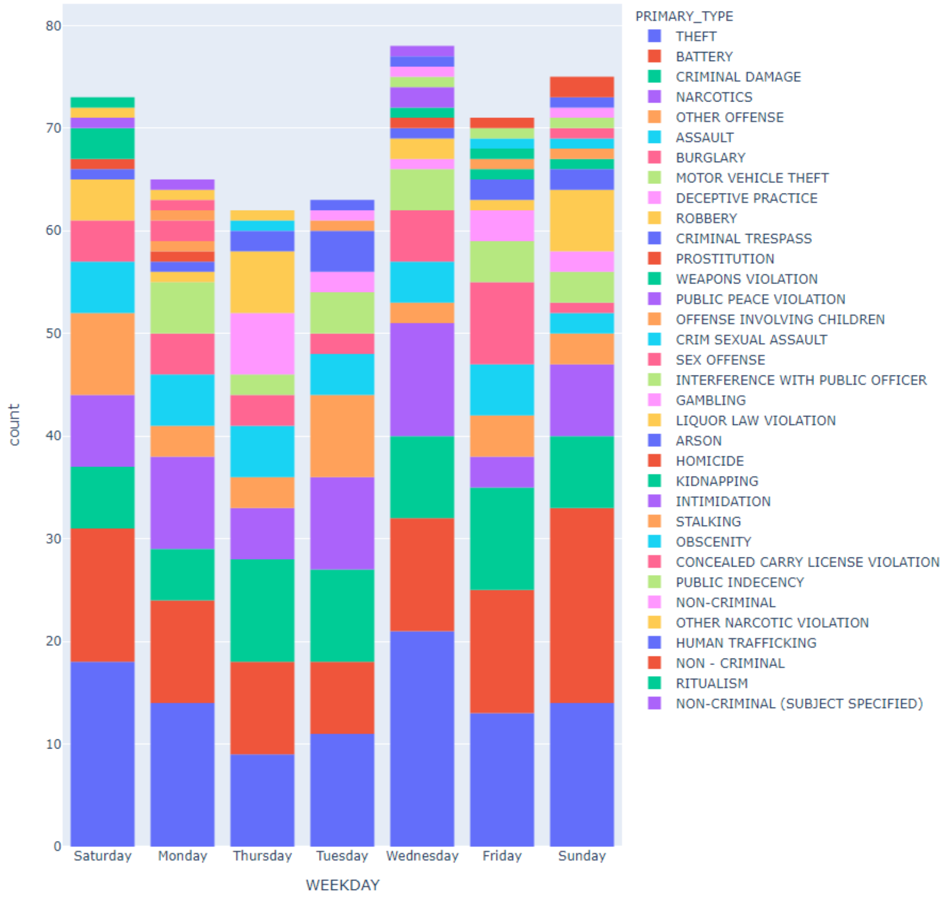

```
There appeares to be two spikes in the week with the first one in the weekend and the next
(much higher) one in the middle of the week. Thursday is one of the least crime reported times
during the week. A simple suggestion would be to co-ordinate resources so as to have more
experienced officiers working the high spike times
```

Co-relation been crime and day of week

In [12]:

```
#show(df[['WEEKDAY','PRIMARY_TYPE']])
Image(filename = "weekdays.png", width = 600, height = 600)
```

Out[12]:




```
A interesting insight in the crime data: There appears to be a co-relation been the theft
activity and narcotics related events. Another insight: While theft is the most reported crim
on any day of the week, during saturdays there is a sharp spike in the barrerty charges. This
is highly indicative of rowdy behaviour over the weekend. The police should prepare for battery
and other violent charges during the weekend.
```

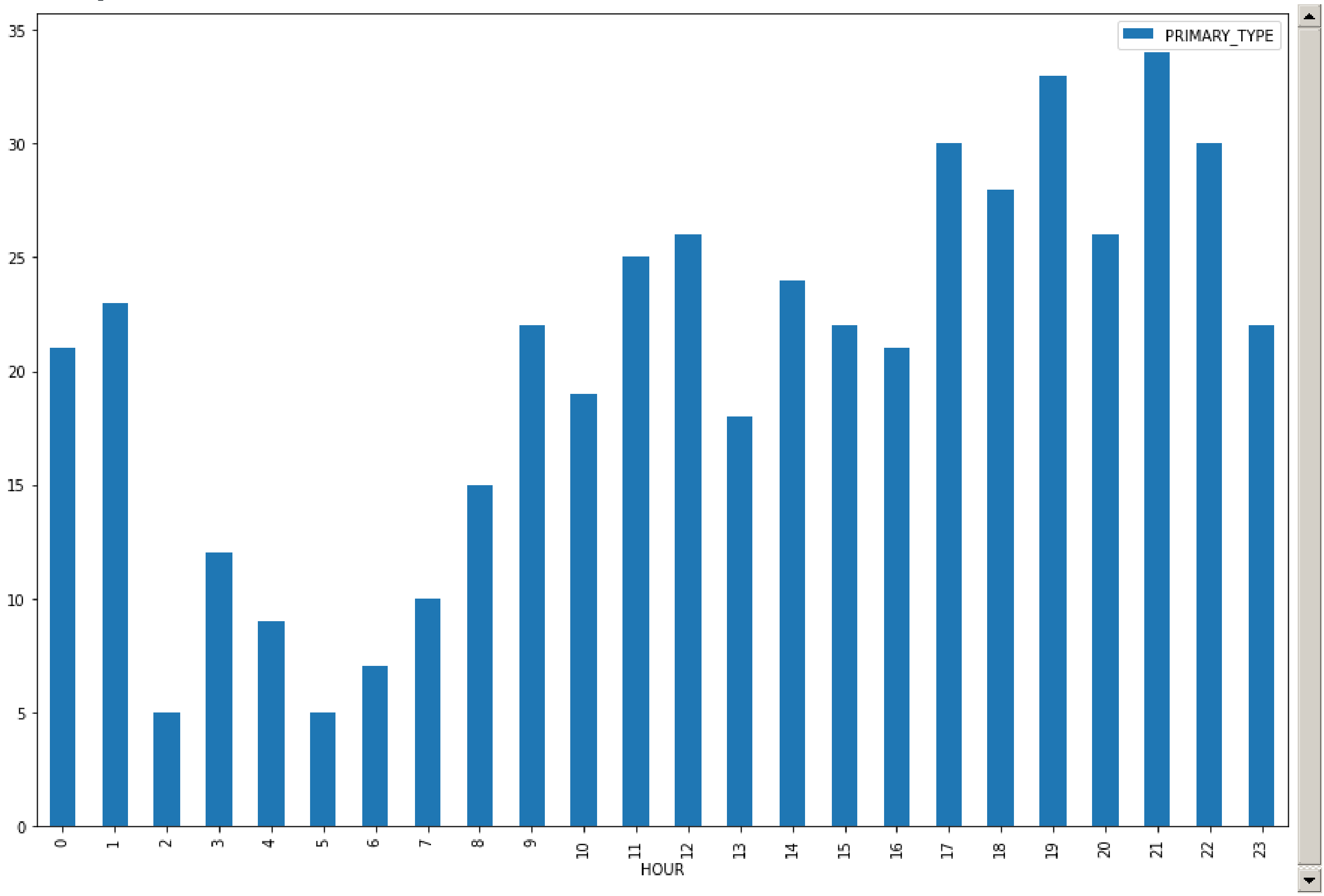
Co-relation been crime and time of day - I

In [13]:

```
df[['HOUR','PRIMARY_TYPE']].groupby('HOUR').count().plot(kind='bar', figsize=(15,10))
```

Out[13]:

```
<AxesSubplot:xlabel='HOUR'>
```



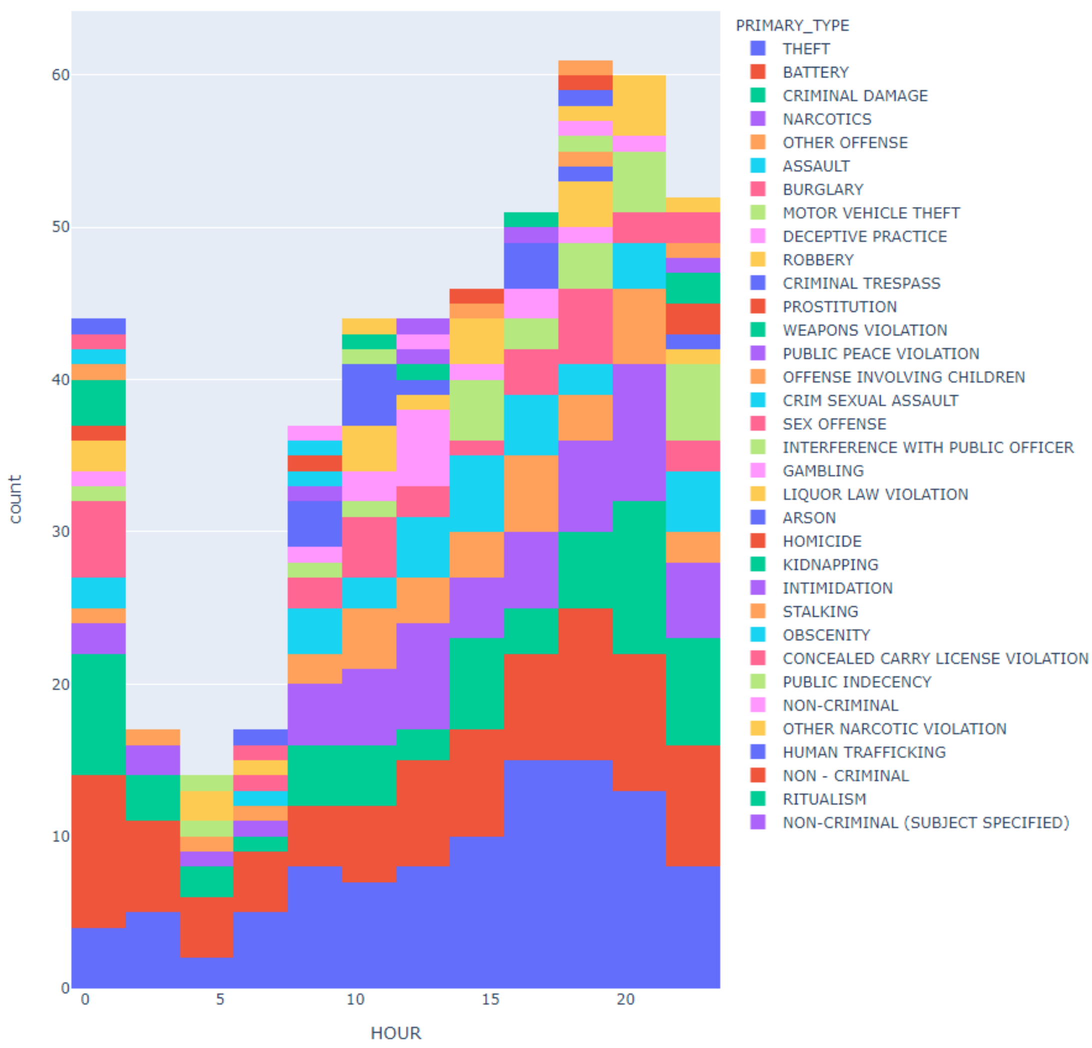

```
As expected the crime numbers are higher during the evenings and nights, and fall sharply
during the late night 'dog-watch' shifts. More insight into whether these are actual lulls of
crime or whether there is a fall in the reporting of these crimes in needed.
```

Co-relation been crime and time of day - II

In [14]:

```
#show(df[['HOUR','PRIMARY_TYPE']])
Image(filename = "hour.png", width = 700, height = 600)
```

Out[14]:



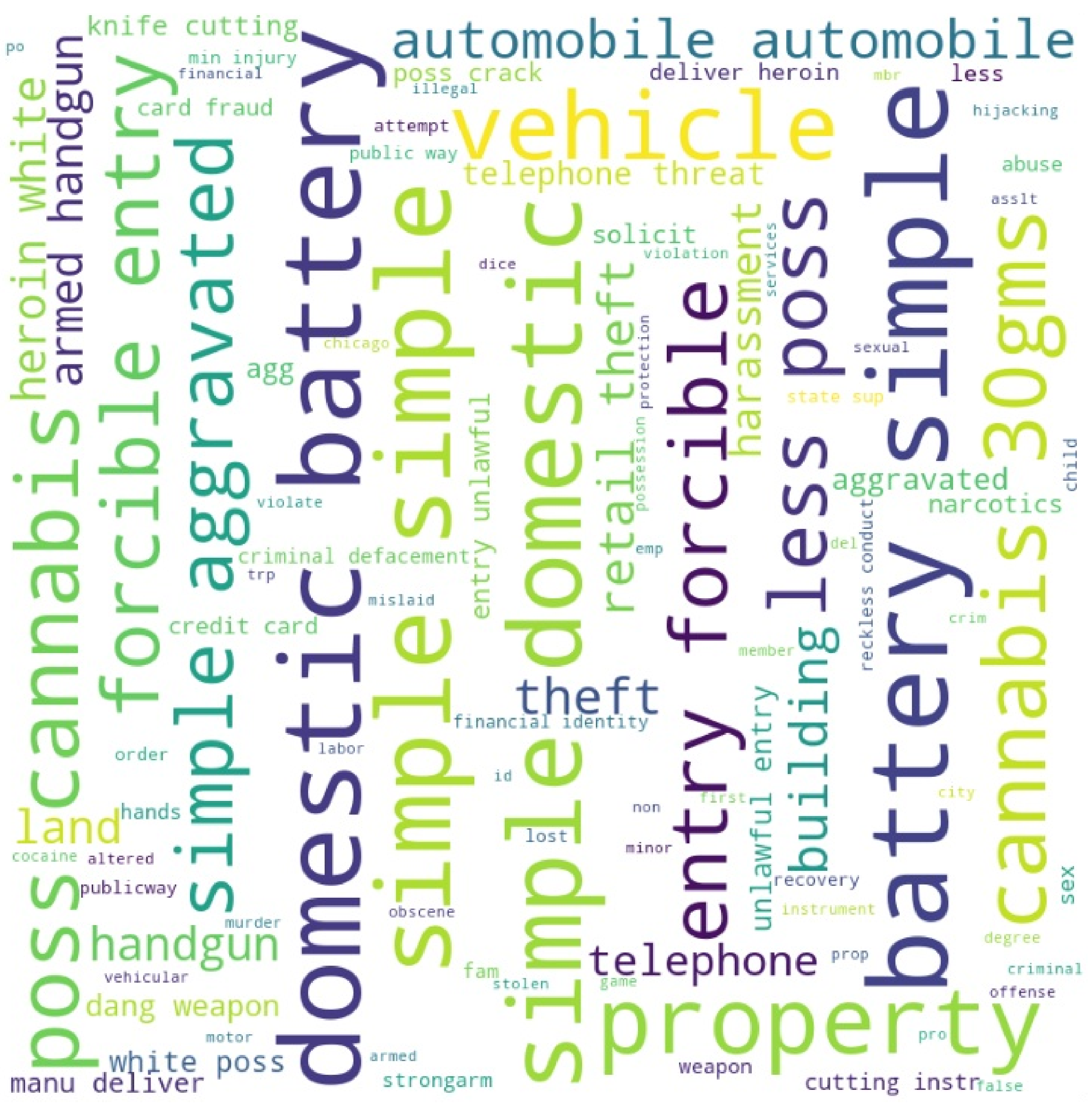

```
During the fall in crime at the 'dog-shift', there appears to be more battery charges than
thefts. As mentioned above Check whether this is due to under-reporting or actual fall in
crime. It is of note that crimes like narcotics and motor vehicular theft are also more
reported during the day than during night time. So, any resources that monitor these crimes
could be better allocated during the day time, rather than at night.
```

Insight into the types of crimes' descriptions

A word cloud of the crime description was created to check if there are any similarities among the crimes reported and recorded. It is found to be less than useful however, it is possible to see a pattern that better sutis the kind of crime data that has been studied.

In [15]:

```
comment_words = ''
stopwords = set(STOPWORDS)

for val in df['DESCRIPTION']:
    val = str(val.lower())
    tokens = val.split()
    for i in range(len(tokens)):
        tokens[i] = tokens[i].lower()
    comment_words += " ".join(tokens)+" "

wordcloud = WordCloud(width = 800, height = 800, background_color ='white', stopwords = stopwords, min_fo

plt.figure(figsize = (15, 10), facecolor = None)
plt.imshow(wordcloud)
plt.axis("off")
plt.tight_layout(pad = 0)

plt.show()
```



In [ ]:

Model building

The data is now going ot be manipulated into only have scalar values. It will than be normalized accoring to the preprocessing needed. The data will first be transfered into a dummy dataframe before these actions are performed. The data will then be used to build a K-Means model that finds similarites among the crimes. This is done to bsee if there are any better co-relations among the crimes than just the primary type of the crime.

The number of K-Means custers was iteratively performed to find the best solution that classified the crimes into appopriate clusters.

In [16]:

```
df_dummy = df.copy()
df_dummy.head(3)
```

Out[16]:

	index	ID	CASE_NUMBER	BLOCK	IUCR	PRIMARY_TYPE	DESCRIPTION	LOCATION_DESCRIPTION	ARREST	DOMESTIC	BEAT	DISTRIC
0	0	3512276	HK587712	047XX S KEDZIE AVE	890	THEFT	FROM BUILDING	SMALL RETAIL STORE	False	False	911	
1	1	3406613	HK456306	009XX N CENTRAL PARK AVE	820	THEFT	$500 AND UNDER	OTHER	False	False	1112	1
2	2	8002131	HT233595	043XX S WABASH AVE	820	THEFT	$500 AND UNDER	NURSING HOME/RETIREMENT HOME	False	False	221	

Preprocessing the dummy dataset:

- Dropping unnecessary columns
- Replacing the True and False values with 1 and 0
- Scaling the overall data using the StandardScalar() fucntion

In [17]:

```
df_dummy.drop(['index','ID','CASE_NUMBER','BLOCK','IUCR','DESCRIPTION','LOCATION_DESCRIPTION','BEAT','DIS

df_dummy.replace(to_replace=False,value=int(0),inplace=True)
df_dummy.replace(to_replace=True,value=int(1),inplace=True)
df_dummy.replace(to_replace='Sunday',value=int(0),inplace=True)
df_dummy.replace(to_replace='Monday',value=int(1),inplace=True)
df_dummy.replace(to_replace='Tuesday',value=int(2),inplace=True)
df_dummy.replace(to_replace='Wednesday',value=int(3),inplace=True)
df_dummy.replace(to_replace='Thursday',value=int(4),inplace=True)
df_dummy.replace(to_replace='Friday',value=int(5),inplace=True)
df_dummy.replace(to_replace='Saturday',value=int(6),inplace=True)

scaler = StandardScaler()
df_dummy.iloc[:,4:] = scaler.fit_transform(df_dummy.iloc[:,4:])
```

Creating One-Hot-Encoding for the types of crimes using get_dummies(). This will then be concatenated to the original dataframe

In [18]:

```
x = pd.get_dummies(df_dummy['PRIMARY_TYPE'])
x.head(3)
```

Out[18]:

	ARSON	ASSAULT	BATTERY	BURGLARY	CONCEALED CARRY LICENSE VIOLATION	CRIM SEXUAL ASSAULT	CRIMINAL DAMAGE	CRIMINAL TRESPASS	DECEPTIVE PRACTICE	GAMBLING	HOMICIDE	HUMAN TRAFFICKING	IN
0	0	0	0	0	0	0	0	0	0	0	0	0	
1	0	0	0	0	0	0	0	0	0	0	0	0	
2	0	0	0	0	0	0	0	0	0	0	0	0	

In [19]:

```
df_dummy = pd.concat([df_dummy,x],axis=1)
df_dummy.drop('PRIMARY_TYPE',axis=1, inplace=True)
df_dummy.head(3)
```

Out[19]:

	ARREST	DOMESTIC	YEAR	LATITUDE	LONGITUDE	WEEKDAY	HOUR	MONTH	DAY	ARSON	ASSAULT	BATTERY	BURGLARY	CON I VIC
0	0.0	0.0	2004	-0.435590	-0.428563	1.474956	0.508366	0.325812	1.435728	0	0	0	0	
1	0.0	0.0	2004	0.653060	-0.637615	1.474956	-0.235100	-0.290384	1.210071	0	0	0	0	
2	0.0	0.0	2011	-0.333811	0.903251	-0.991745	-1.275953	-0.906581	-1.272161	0	0	0	0	

In [20]:

```
## K-Means
```

In [24]:

```
kclusters = 12
kmeans = KMeans(init = "random", n_clusters = kclusters, n_init=10, max_iter=1000)
kmeans.fit(df_dummy)
kmeans.labels_[:5]
```

Out[24]:

```
array([ 1,  9,  7,  7, 11])
```

In [22]:

```
df_dummy['CLUSTER'] = kmeans.labels_
df_dummy.head(3)
```

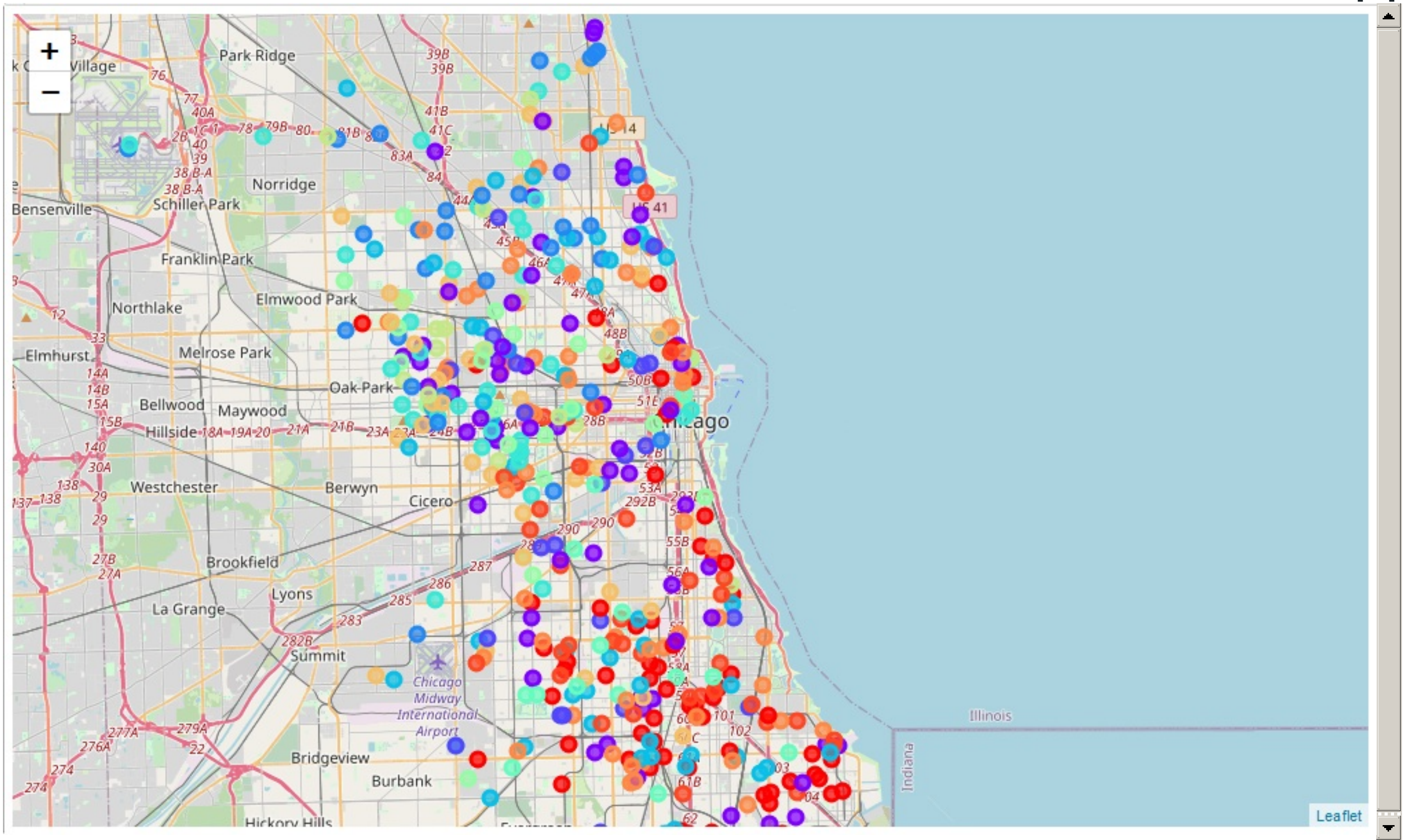
Out[22]:

	ARREST	DOMESTIC	YEAR	LATITUDE	LONGITUDE	WEEKDAY	HOUR	MONTH	DAY	ARSON	ASSAULT	BATTERY	BURGLARY	CON I VIC
0	0.0	0.0	2004	-0.435590	-0.428563	1.474956	0.508366	0.325812	1.435728	0	0	0	0	
1	0.0	0.0	2004	0.653060	-0.637615	1.474956	-0.235100	-0.290384	1.210071	0	0	0	0	
2	0.0	0.0	2011	-0.333811	0.903251	-0.991745	-1.275953	-0.906581	-1.272161	0	0	0	0	

In [23]:

```
map_clusters = folium.Map(location=[latitude, longitude], zoom_start=10.5)

x = np.arange(kclusters)
ys = [i + x + (i*x)**2 for i in range(kclusters)]
colors_array = cm.rainbow(np.linspace(0, 1, len(ys)))
rainbow = [colors.rgb2hex(i) for i in colors_array]

# add markers to the map
markers_colors = []
for lat, lon, cluster in zip(df['LATITUDE'], df['LONGITUDE'], df_dummy['CLUSTER']):
    label = folium.Popup('Cluster ' + str(cluster), parse_html=True)
    cluster = int(cluster)
    folium.CircleMarker(
        [lat, lon],
        radius=5,
        popup=label,
        color=rainbow[int(cluster)-1],
        fill=True,
        fill_color=rainbow[int(cluster)-1],
        fill_opacity=0.7).add_to(map_clusters)

map_clusters
Image(filename = "clustering.png", width = 1000, height = 600)
```

Out[23]:



It seems that, there is a particulat kind of crime (denoted in red) that is more prevalent in the lower part of the ity while the upper part of the city is filled with the teal and turquiose clusters.

In [ ]:

Results

The following are the results of the study:

- While the department spends a lot on gang tasks forces, military surplus gear and tactical training for officers, the majority of the crimes reported in Chicago is theft. especially in sparsely populated area of the airport and waterfront highway. Better surveillance, patrol and funds for the CI program could be a bteer place to spend the resources
- Oak park is one of the most important places in the city that needs attention. A reported 35% of the crimes occur here and better tackling of there crimes could reduce the overall city crime-rate
- Weekends and Wednesdays, during the third shift is the most important time for the officers to be vigilant. The steady rise in crime over the month could be crimes of need rather than crimes of passion.
- A suspiciously lower among of data is found for the crime reported during the late-night shift
- There appears to be clear distinction between the type of crimes that occur in the lower part of the city to the upper part of the city.

In [ ]:

Discussion and suggestions

The top four crimes account for the majority of the crimes commited, with theft leading the chart by a large margin. **The resources available there for can be pivoted towards preventing these crimes through regular patrols.** It is also to note that three of the top four are essentially victimless although danmage to public property is incurred.

Secluded areas such as the airport surroundings, North-east side of the city and US 42 HW along the waterfront are more prone to theft. As theft is one of the highest number of recorded crimes and there is a visible cluster, the resources available ought to be spend on securing these areas. **Regular surveillance and patrol are some of the suggestions to reduce the occurances of theft in these special interest areas.**

Oak park is the most crime prone area in the city qith nearly 35% of the crimes occuring there. **A simple suggestion would be to increase the outreach programs in this neighborhood. Further investivation leads to believe these are more gang related activities owing to the nature of the crimes commited there.** This is a perfect area for incorporating crime/gangs task force.

The two spikes in the week with the first one in the weekend and the next one in the middle of the week are to be carefully observed. Thursday is one of the least crime reported times during the week. **A suggestion would be to co-ordinate resources so as to have more experienced officers working the high spike times.**

The crimes of needs (theft commited over the later part of the month) oculd be tackled with better outreach programs that help the community

The co-relation been the theft activity and narcotics related events are to be further studied. And the fall in the number of cases of theft and the rise in the battery charges arew also to be watched. This is highly indicative of rowdy behaviour over the weekend. **The police should prepare for battery and other violent charges during the weekend.**

The police officers working th eareas on the lower half of the city need much different set of skills tha the once working onthe upper part of the city. This is clear from the clustering of the crimes on different part ofthe city, rather than by different primary crime types. It seems that, there is a particulat kind of crime (denoted in red) that is more prevalent in the lower part of the city while the upper part of the city is filled with the teal and turquiose clusters.

In [ ]: