

라이브러리 로드

In [5]:

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

데이터 로드

In [6]:

```
df = pd.read_csv('Crime_Data_from_2010_to_Present.csv')
```

컬럼명 확인

In [3]:

```
df.columns
```

Out[3]:

```
Index(['DR_NO', 'Date Rptd', 'DATE OCC', 'TIME OCC', 'AREA ', 'AREA NAME',
      'Rpt Dist No', 'Part 1-2', 'Crm Cd', 'Crm Cd Desc', 'Mocodes',
      'Vict Age', 'Vict Sex', 'Vict Descent', 'Premis Cd', 'Premis Desc',
      'Weapon Used Cd', 'Weapon Desc', 'Status', 'Status Desc', 'Crm Cd 1',
      'Crm Cd 2', 'Crm Cd 3', 'Crm Cd 4', 'LOCATION', 'Cross Street', 'LAT',
      'LON'],
      dtype='object')
```

필요 없는 컬럼 제거

In [4]:

```
select_col = ['DATE OCC', 'TIME OCC', 'AREA NAME', 'Crm Cd', 'Crm Cd Desc', 'Vict Age', 'Vict Sex',
              'Vict Descent', 'Premis Desc',
              'LOCATION', 'LAT', 'LON']
```

In [5]:

```
data = df.loc[:, select_col]
```

2015년 부터의 데이터만 추출

In [6]:

```
data = data.iloc[997702:]
```

M,F,X 성별만 추출

In [7]:

```
data = data[data['Vict Sex'].isin(['M', 'F', 'X'])]
```

In [8]:

```
data = data.reset_index()
```

TIME OCC 컬럼에서 시간 앞에 0으로 채우기

In [9]:

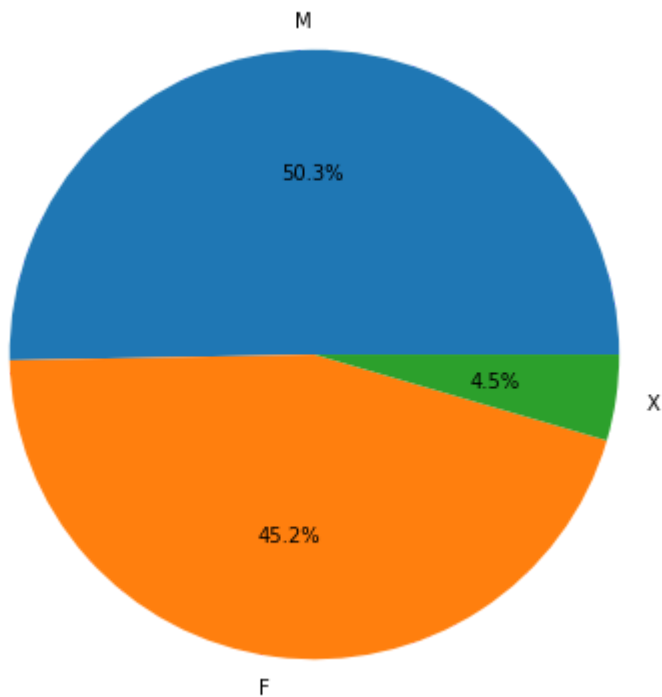
```
data['TIME OCC'] = [str(i).zfill(4) for i in data['TIME OCC']]
```

In [10]:

```
data.to_csv('ing_3.csv')
```

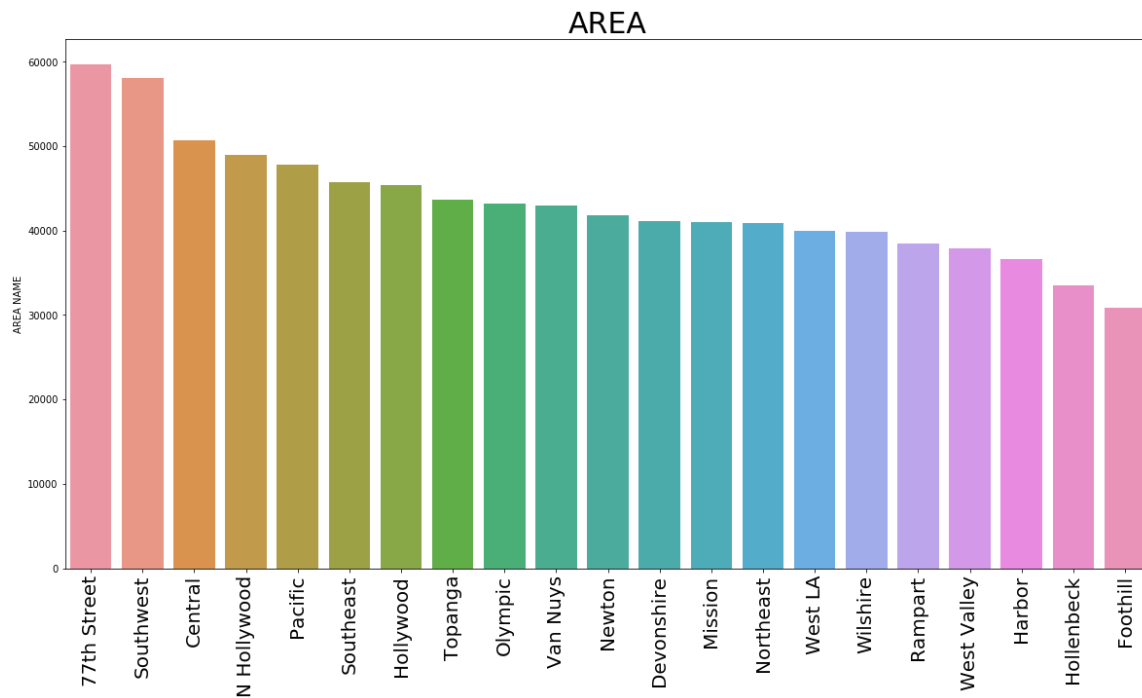
In [11]:

```
plt.figure(figsize=(7,7))  
plt.pie(data['Vict Sex'].value_counts(), labels = data['Vict Sex'].value_counts().index, autopct='%1.1f%%')  
plt.show()
```



In [12]:

```
plt.figure(figsize=(20,10))
sns.barplot(x=data['AREA NAME'].value_counts().index,y=data['AREA NAME'].value_counts())
plt.xticks(rotation=90,size=20)
plt.title("AREA",size=30)
plt.show()
```



In [20]:

```
lat_lon = ["[{}], {}".format(data['LON'][i], data['LAT'][i]) for i in range(len(data))]
```

In [21]:

```
with open('lon_lat.txt','w') as f:
    f.write('\n'.join(lat_lon))
```