In [2]:

```
import numpy as np
import pandas as pd

# for visualizations
import matplotlib.pyplot as plt
import seaborn as sns
import folium
import squarify
```

In [2]:

```
data = pd.read_csv('train.csv')

# check the shape of the data
data.shape
```

Out[2]:

```
(150500, 13)
```

In [3]:

```
data.head()
```

Out[3]:

n	Category	Descript	DayOfWeek	Date	Time	PdDistrict	Resolution	Addres
2	WEAPON LAWS	POSS OF PROHIBITED WEAPON	Friday	01/29/2016 12:00:00 AM	11:00	SOUTHERN	ARREST, BOOKED	800 Blo of BRYAN S
2	WEAPON LAWS	FIREARM, LOADED, IN VEHICLE, POSSESSION OR USE	Friday	01/29/2016 12:00:00 AM	11:00	SOUTHERN	ARREST, BOOKED	800 Blo of BRYAN S
3	WARRANTS	WARRANT ARREST	Monday	04/25/2016 12:00:00 AM	14:59	BAYVIEW	ARREST, BOOKED	KEITH ST SHAFTE A
2	NON-CRIMINAL	LOST PROPERTY	Tuesday	01/05/2016 12:00:00 AM	23:50	TENDERLOIN	NONE	JONES S OFARREL S
0	NON-CRIMINAL	LOST PROPERTY	Friday	01/01/2016 12:00:00 AM	00:30	MISSION	NONE	16TH ST MISSIO S

In [4]:

```
data.describe()
```

Out[4]:

	IncidntNum	X	Y	PdId
count	1.505000e+05	150500.000000	150500.000000	1.505000e+05
mean	1.616440e+08	-122.423599	37.768921	1.616440e+13
std	5.535976e+06	0.026210	0.023637	5.535976e+11
min	1.135121e+07	-122.513642	37.707922	1.135121e+12
25%	1.603283e+08	-122.434036	37.756486	1.603283e+13
50%	1.606541e+08	-122.416903	37.775421	1.606541e+13
75%	1.609764e+08	-122.406605	37.785063	1.609764e+13
max	9.910090e+08	-122.365565	37.819975	9.910090e+13

In [5]:

```
data.isnull().sum()
```

Out[5]:

```
IncidntNum    0
Category      0
Descript      0
DayOfWeek     0
Date          0
Time          0
PdDistrict    1
Resolution    0
Address       0
X             0
Y             0
Location      0
PdId          0
dtype: int64
```

In [6]:

```
data['PdDistrict'].fillna(data['PdDistrict'].mode()[0], inplace = True)

data.isnull().any().any()
```

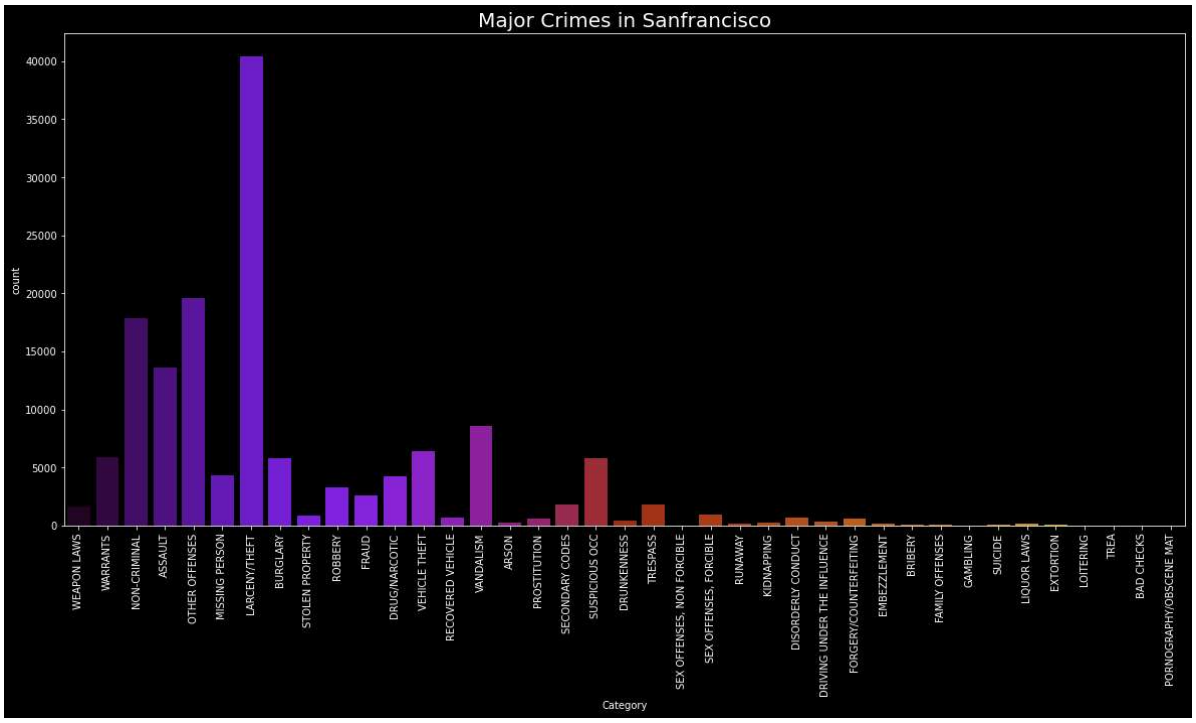
Out[6]:

```
False
```

In [17]:

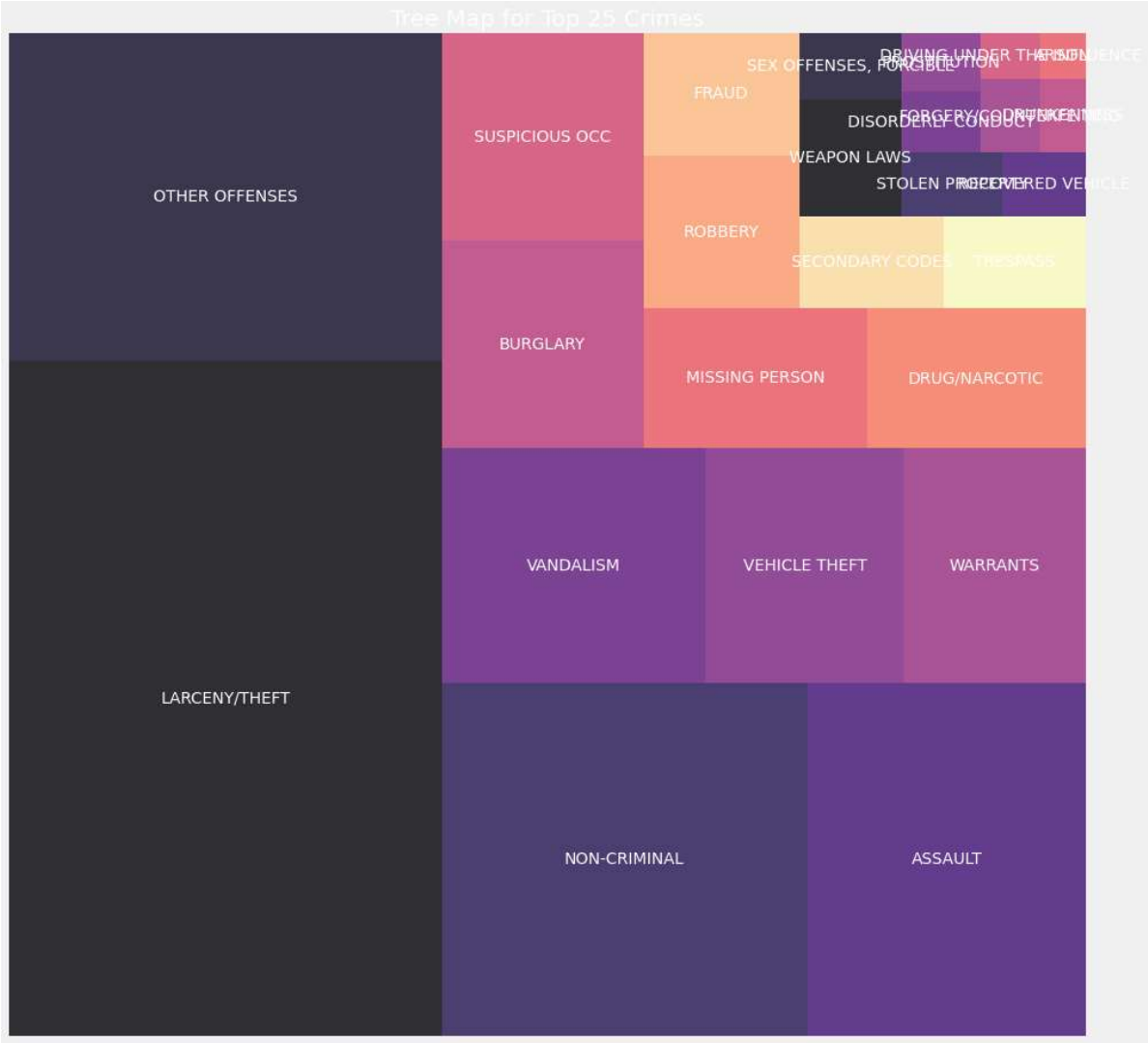
```
plt.rcParams['figure.figsize'] = (20, 9)
plt.style.use('dark_background')

sns.countplot(data['Category'], palette = 'gnuplot')

plt.title('Major Crimes in Sanfrancisco', fontweight = 30, fontsize = 20)
plt.xticks(rotation = 90)
plt.show()
```

```
C:\Users\hp\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWar
ning: Pass the following variable as a keyword arg: x. From version 0.12, th
e only valid positional argument will be `data`, and passing other arguments
without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
```



In [19]:

```
y = data['Category'].value_counts().head(25)

plt.rcParams['figure.figsize'] = (15, 15)
plt.style.use('fivethirtyeight')

color = plt.cm.magma(np.linspace(0, 1, 15))
squarify.plot(sizes = y.values, label = y.index, alpha=.8, color = color)
plt.title('Tree Map for Top 25 Crimes', fontsize = 20)

plt.axis('off')
plt.show()
```

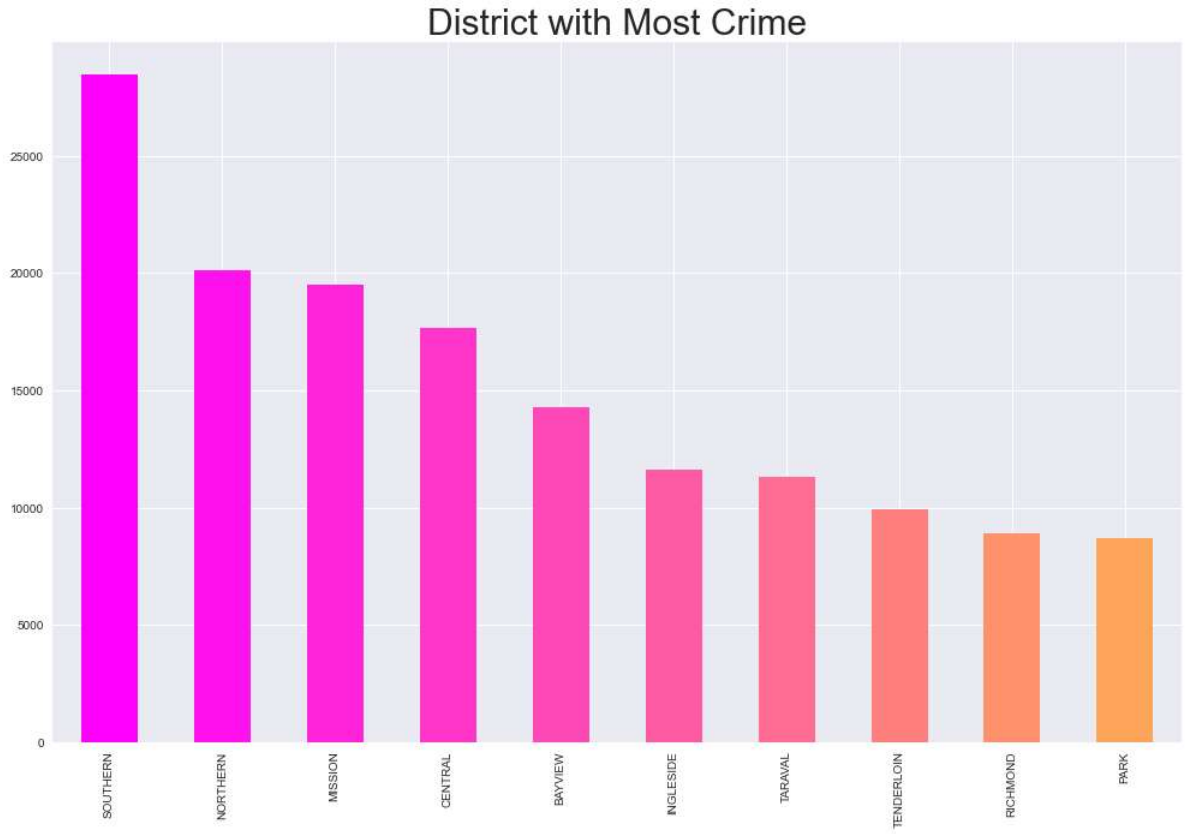


In [21]:

```
from wordcloud import WordCloud

plt.rcParams['figure.figsize'] = (15, 15)
plt.style.use('fast')

wc = WordCloud(background_color = 'orange', width = 1500, height = 1500).generate(str(data[
plt.title('Description of the Crime', fontsize = 20)

plt.imshow(wc)
plt.axis('off')
plt.show()
```

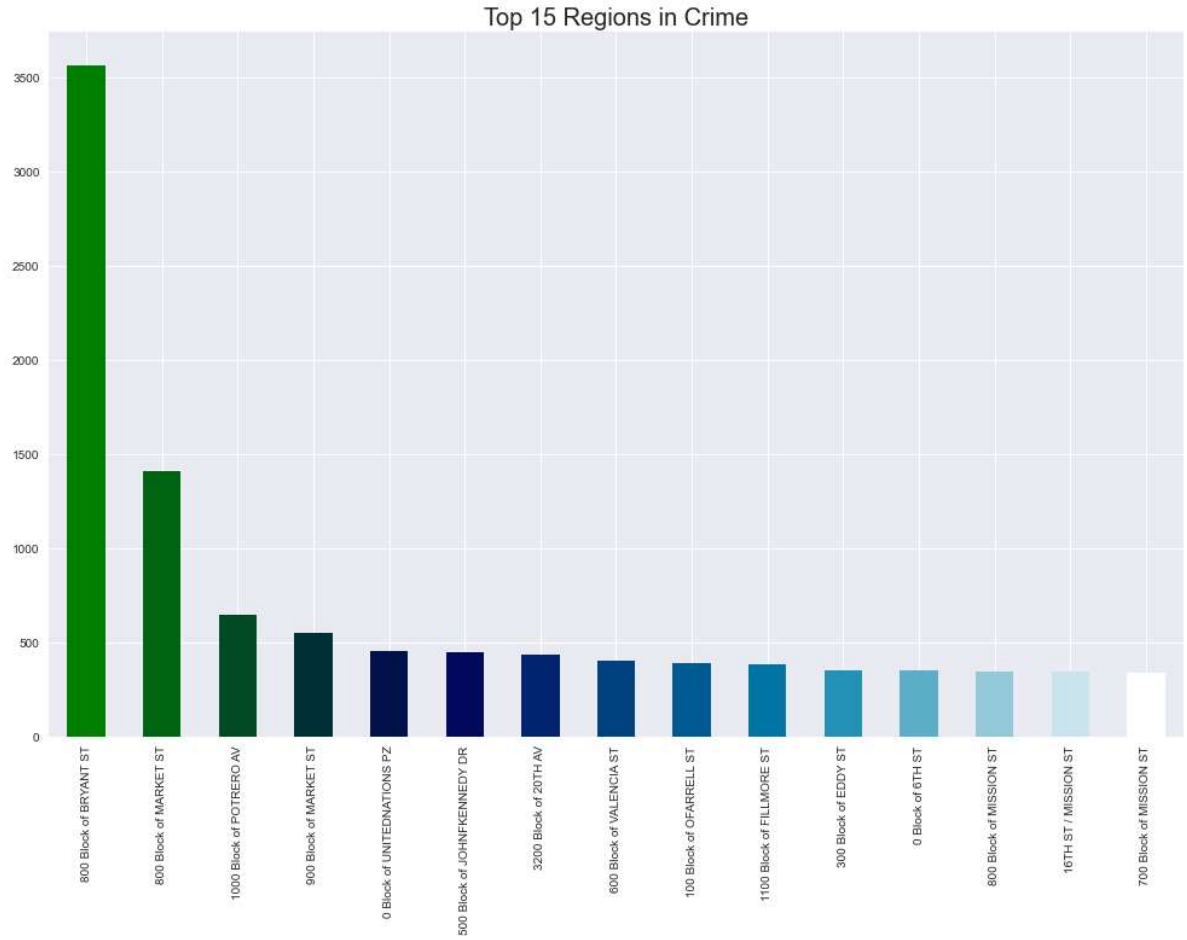
In [22]:

```
plt.rcParams['figure.figsize'] = (20, 9)
plt.style.use('seaborn')

color = plt.cm.spring(np.linspace(0, 1, 15))
data['PdDistrict'].value_counts().plot.bar(color = color, figsize = (15, 10))

plt.title('District with Most Crime',fontsize = 30)

plt.xticks(rotation = 90)
plt.show()
```



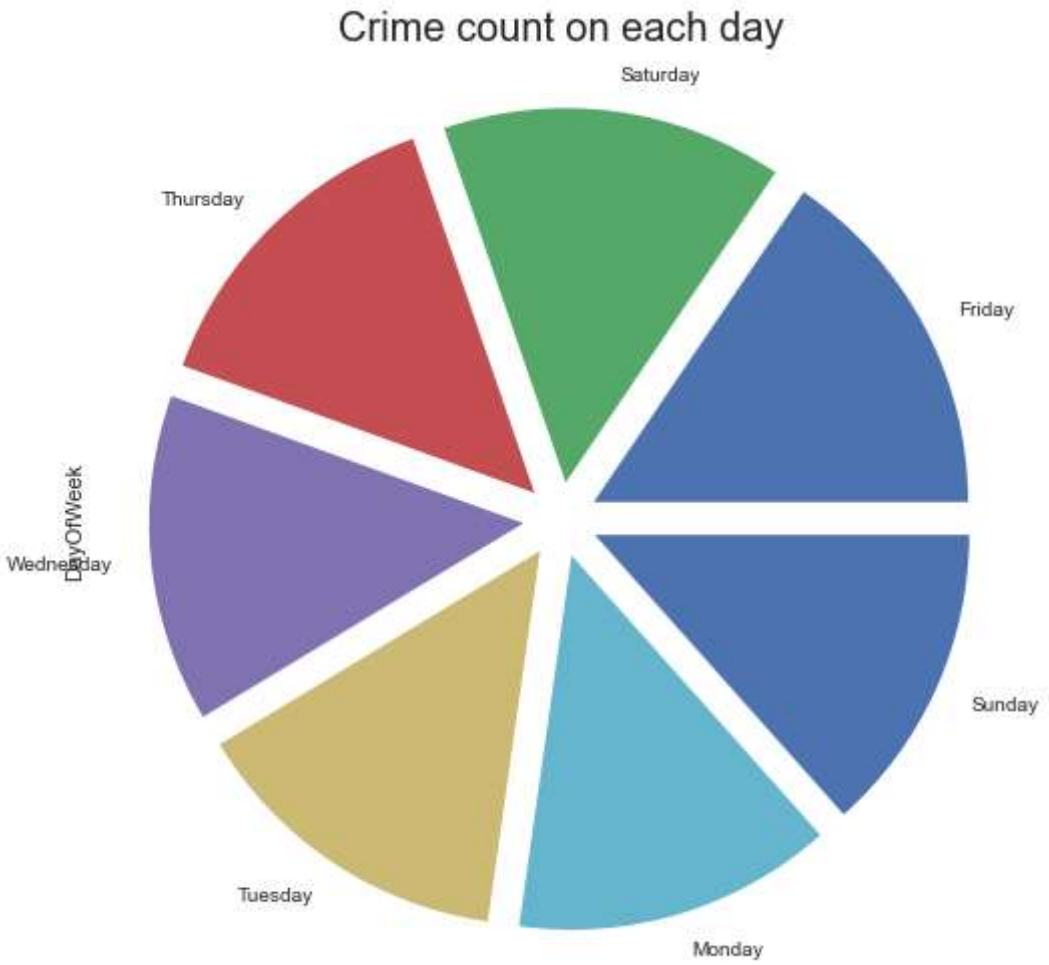
In [23]:

```
plt.rcParams['figure.figsize'] = (20, 9)
plt.style.use('seaborn')

color = plt.cm.ocean(np.linspace(0, 1, 15))
data['Address'].value_counts().head(15).plot.bar(color = color, figsize = (15, 10))

plt.title('Top 15 Regions in Crime',fontsize = 20)

plt.xticks(rotation = 90)
plt.show()
```



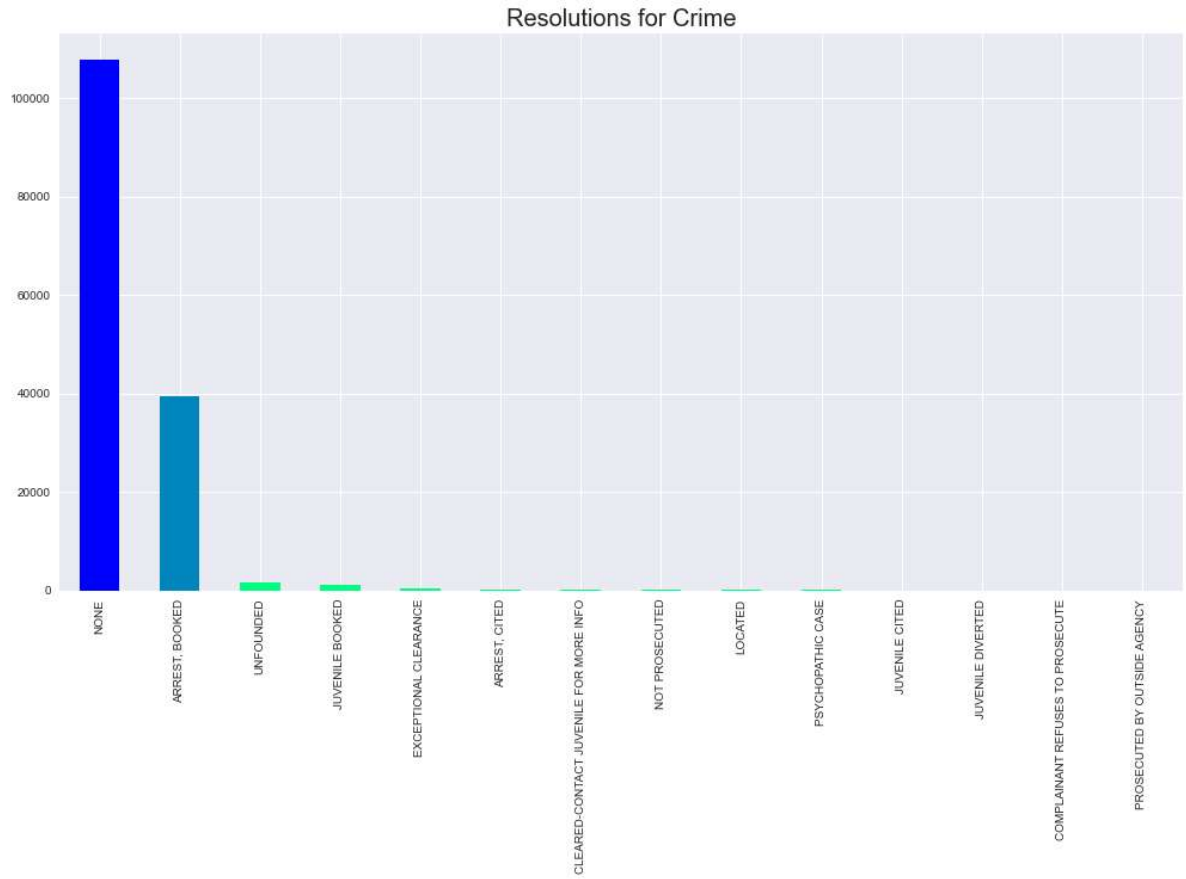
In [21]:

```
plt.style.use('seaborn')


data['DayOfWeek'].value_counts().head(15).plot.pie(figsize = (15, 8), explode = (0.1, 0.1,

plt.title('Crime count on each day',fontsize = 20)

plt.xticks(rotation = 90)
plt.show()
```



In [24]:

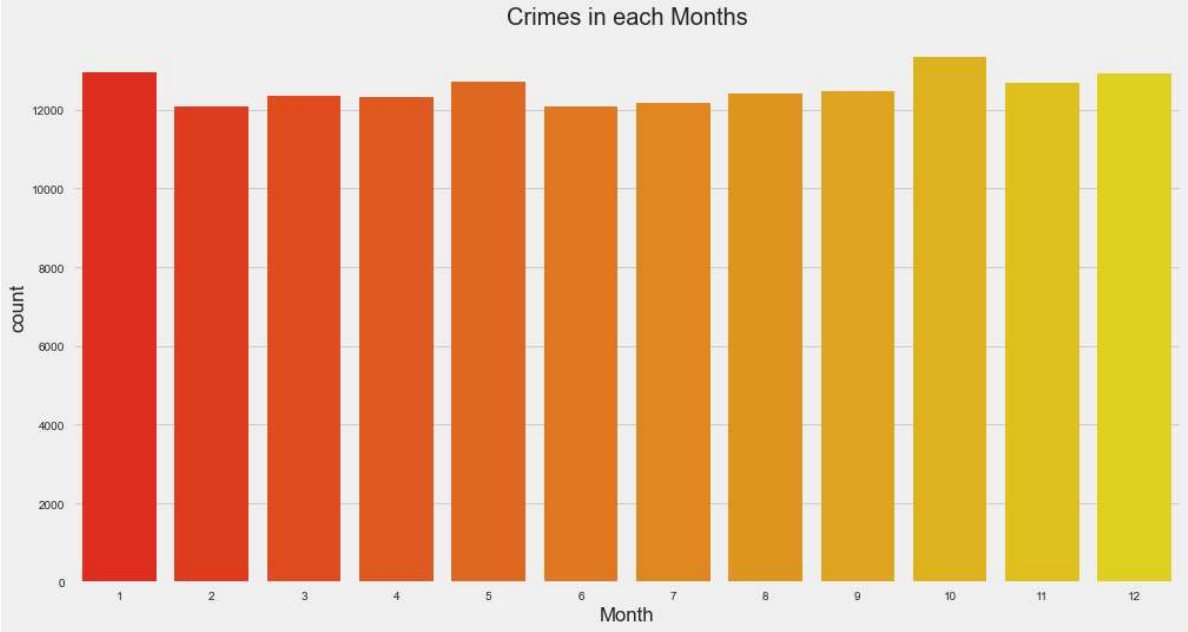
```
plt.style.use('seaborn')

color = plt.cm.winter(np.linspace(0, 10, 20))
data['Resolution'].value_counts().plot.bar(color = color, figsize = (15, 8))

plt.title('Resolutions for Crime',fontsize = 20)
plt.xticks(rotation = 90)
plt.show()
```



In [23]:

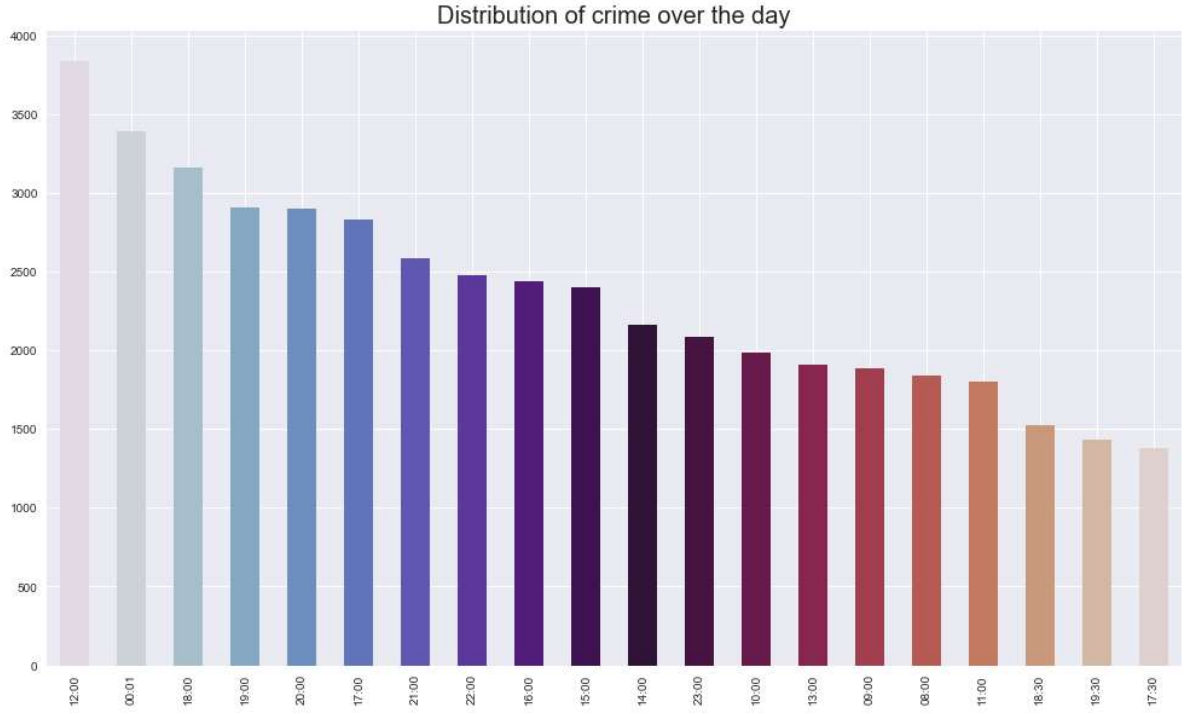
```
data['Date'] = pd.to_datetime(data['Date'])

data['Month'] = data['Date'].dt.month

plt.style.use('fivethirtyeight')
plt.rcParams['figure.figsize'] = (15, 8)

sns.countplot(data['Month'], palette = 'autumn',)
plt.title('Crimes in each Months', fontsize = 20)

plt.show()
```

```
C:\anaconda\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass
the following variable as a keyword arg: x. From version 0.12, the only vali
d positional argument will be `data`, and passing other arguments without an
explicit keyword will result in an error or misinterpretation.
  warnings.warn(
```



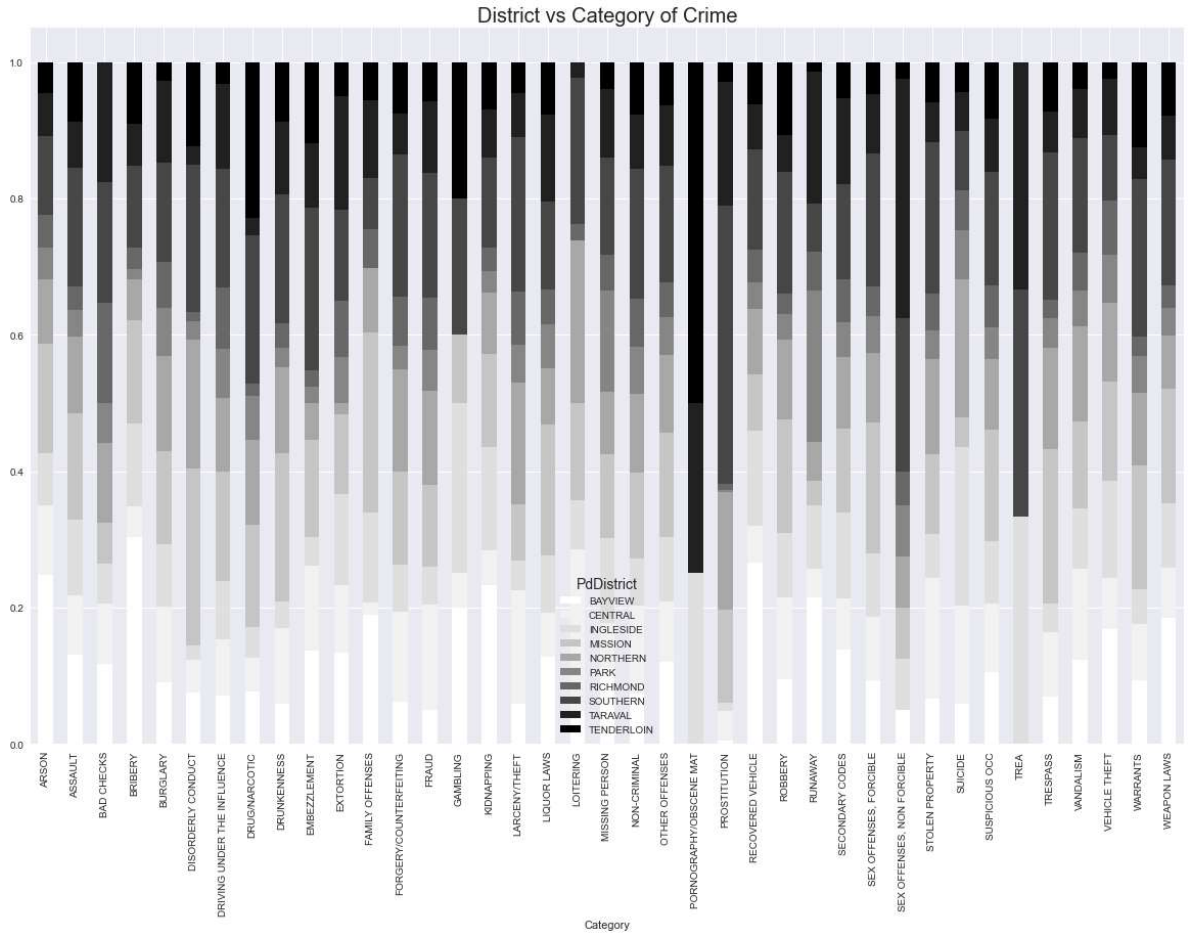
In [25]:

```
import warnings
warnings.filterwarnings('ignore')

color = plt.cm.twilight(np.linspace(0, 5, 100))
data['Time'].value_counts().head(20).plot.bar(color = color, figsize = (15, 9))

plt.title('Distribution of crime over the day', fontsize = 20)
plt.show()
```



In [26]:

```
df = pd.crosstab(data['Category'], data['PdDistrict'])
color = plt.cm.Greys(np.linspace(0, 1, 10))

df.div(df.sum(1).astype(float), axis = 0).plot.bar(stacked = True, color = color, figsize =
plt.title('District vs Category of Crime', fontweight = 30, fontsize = 20)

plt.xticks(rotation = 90)
plt.show()
```



In [27]:

```
t = data.PdDistrict.value_counts()
table = pd.DataFrame(data=t.values, index=t.index, columns=['Count'])
table = table.reindex(["CENTRAL", "NORTHERN", "PARK", "SOUTHERN", "MISSION", "TENDERLOIN",

table = table.reset_index()
table.rename({'index': 'Neighborhood'}, axis='columns', inplace=True)

table
```

Out[27]:

	Neighborhood	Count
0	CENTRAL	17666
1	NORTHERN	20100
2	PARK	8699
3	SOUTHERN	28446
4	MISSION	19503
5	TENDERLOIN	9942
6	RICHMOND	8922
7	TARAVAL	11325
8	INGLESIDE	11594
9	BAYVIEW	14303

In [ ]:

In [ ]: