ESSEX CRIME DATA ANALYSIS

In [1]:
```
!pip install jovian numpy pandas matplotlib seaborn plotly folium --quiet
```

In [2]:
```
import jovian
import pandas as pd
import numpy as np
import os
import folium
from folium import plugins
%matplotlib inline
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
from plotly.offline import init_notebook_mode
```

Reading the data set

In [3]:
```
filepath = (r'C:\Users\F.O.T\Desktop\python cww\\')
filelist = os.listdir(filepath)
filepath
```

Out[3]:
```
'C:\\Users\\F.O.T\\Desktop\\python cww\\\\'
```

The Python library being imported

In [4]:
```
essexdata1 = pd.DataFrame()
for file in filelist:
    cw_file = pd.read_csv(f'{filepath}{file}')
    essexdata1 = essexdata1.append(cw_file, ignore_index = True)
```

```
C:\Users\F.O.T\AppData\Local\Temp\ipykernel_10904\3738991387.py:4: FutureWarning:

The frame.append method is deprecated and will be removed from pandas in a future versio
n. Use pandas.concat instead.

C:\Users\F.O.T\AppData\Local\Temp\ipykernel_10904\3738991387.py:4: FutureWarning:

The frame.append method is deprecated and will be removed from pandas in a future versio
n. Use pandas.concat instead.

C:\Users\F.O.T\AppData\Local\Temp\ipykernel_10904\3738991387.py:4: FutureWarning:

The frame.append method is deprecated and will be removed from pandas in a future versio
n. Use pandas.concat instead.

C:\Users\F.O.T\AppData\Local\Temp\ipykernel_10904\3738991387.py:4: FutureWarning:

The frame.append method is deprecated and will be removed from pandas in a future versio
n. Use pandas.concat instead.

C:\Users\F.O.T\AppData\Local\Temp\ipykernel_10904\3738991387.py:4: FutureWarning:

The frame.append method is deprecated and will be removed from pandas in a future versio
n. Use pandas.concat instead.
```

```
C:\Users\F.O.T\AppData\Local\Temp\ipykernel_10904\3738991387.py:4: FutureWarning:

The frame.append method is deprecated and will be removed from pandas in a future versio
n. Use pandas.concat instead.

C:\Users\F.O.T\AppData\Local\Temp\ipykernel_10904\3738991387.py:4: FutureWarning:

The frame.append method is deprecated and will be removed from pandas in a future versio
n. Use pandas.concat instead.

C:\Users\F.O.T\AppData\Local\Temp\ipykernel_10904\3738991387.py:4: FutureWarning:

The frame.append method is deprecated and will be removed from pandas in a future versio
n. Use pandas.concat instead.

C:\Users\F.O.T\AppData\Local\Temp\ipykernel_10904\3738991387.py:4: FutureWarning:

The frame.append method is deprecated and will be removed from pandas in a future versio
n. Use pandas.concat instead.

C:\Users\F.O.T\AppData\Local\Temp\ipykernel_10904\3738991387.py:4: FutureWarning:

The frame.append method is deprecated and will be removed from pandas in a future versio
n. Use pandas.concat instead.

C:\Users\F.O.T\AppData\Local\Temp\ipykernel_10904\3738991387.py:4: FutureWarning:

The frame.append method is deprecated and will be removed from pandas in a future versio
n. Use pandas.concat instead.

C:\Users\F.O.T\AppData\Local\Temp\ipykernel_10904\3738991387.py:4: FutureWarning:

The frame.append method is deprecated and will be removed from pandas in a future versio
n. Use pandas.concat instead.

C:\Users\F.O.T\AppData\Local\Temp\ipykernel_10904\3738991387.py:4: FutureWarning:

The frame.append method is deprecated and will be removed from pandas in a future versio
n. Use pandas.concat instead.

C:\Users\F.O.T\AppData\Local\Temp\ipykernel_10904\3738991387.py:4: FutureWarning:

The frame.append method is deprecated and will be removed from pandas in a future versio
n. Use pandas.concat instead.

C:\Users\F.O.T\AppData\Local\Temp\ipykernel_10904\3738991387.py:4: FutureWarning:

The frame.append method is deprecated and will be removed from pandas in a future versio
n. Use pandas.concat instead.

C:\Users\F.O.T\AppData\Local\Temp\ipykernel_10904\3738991387.py:4: FutureWarning:

The frame.append method is deprecated and will be removed from pandas in a future versio
n. Use pandas.concat instead.

C:\Users\F.O.T\AppData\Local\Temp\ipykernel_10904\3738991387.py:4: FutureWarning:

The frame.append method is deprecated and will be removed from pandas in a future versio
n. Use pandas.concat instead.

C:\Users\F.O.T\AppData\Local\Temp\ipykernel_10904\3738991387.py:4: FutureWarning:

The frame.append method is deprecated and will be removed from pandas in a future versio
n. Use pandas.concat instead.

C:\Users\F.O.T\AppData\Local\Temp\ipykernel_10904\3738991387.py:4: FutureWarning:
```

```
The frame.append method is deprecated and will be removed from pandas in a future versio
n. Use pandas.concat instead.

C:\Users\F.O.T\AppData\Local\Temp\ipykernel_10904\3738991387.py:4: FutureWarning:

The frame.append method is deprecated and will be removed from pandas in a future versio
n. Use pandas.concat instead.

C:\Users\F.O.T\AppData\Local\Temp\ipykernel_10904\3738991387.py:4: FutureWarning:

The frame.append method is deprecated and will be removed from pandas in a future versio
n. Use pandas.concat instead.

C:\Users\F.O.T\AppData\Local\Temp\ipykernel_10904\3738991387.py:4: FutureWarning:

The frame.append method is deprecated and will be removed from pandas in a future versio
n. Use pandas.concat instead.
```

Data characterization

In [5]:
```
essexdata1
```

Out[5]:

	Crime ID	Month	Reported by	Falls within	Longitude	Latitude	
0	d50762ec94c695996eba977489d5e313926f9b15cb7c36...	2021-01	Essex Police	Essex Police	0.439541	51.639160	O
1	d535ca1a5da12cb0dfc82a45eb96cbdce41ee6e638b382...	2021-01	Essex Police	Essex Police	0.436841	51.638676	O
2	dcbab04b53a2bd4d3553b5bde3fd15dfb9fe20193dfe02...	2021-01	Essex Police	Essex Police	0.436841	51.638676	O
3	NaN	2021-01	Essex Police	Essex Police	0.439730	51.629479	O Bra
4	7df5b0b06b4b7b4db73a679ab3b15d42f0b6c529961217...	2021-01	Essex Police	Essex Police	0.437891	51.626712	O
...	...	...	...	...	...	...	
363432	879c4243fbf09b88b45f94122036123f007202f455ec11...	2022-10	Essex Police	Essex Police	NaN	NaN	No
363433	5d83265c55442b3a03def4ea36297361badcdbc279473d...	2022-10	Essex Police	Essex Police	NaN	NaN	No
363434	327cfdc300ef275be9cb22a19123b1e3b95e4d8eb8e2b0...	2022-10	Essex Police	Essex Police	NaN	NaN	No
363435	fc0f4b97857d228d38b4f6d967b7615ddc4e1e380223f3...	2022-	Essex	Essex	NaN	NaN	No

			10	Police	Police		
363436	bffb127b5e05bb526c52f73d9eaacdf1f00ce7ef90fcdc...	2022-10	Essex Police	Essex Police	NaN	NaN	No

363437 rows × 12 columns

checking for null values in the dataset

In [6]:
```
essexdata1.isnull().sum()
```

Out[6]:
```
Crime ID                  63652
Month                         0
Reported by                   0
Falls within                  0
Longitude                  2917
Latitude                   2917
Location                      0
LSOA code                  2917
LSOA name                  2917
Crime type                    0
Last outcome category     63652
Context                  363437
dtype: int64
```

removing null values in the dataset

In [7]:
```
essexdata=essexdata1.dropna(axis=1)
essexdata
```

Out[7]:

	Month	Reported by	Falls within	Location	Crime type
0	2021-01	Essex Police	Essex Police	On or near Smythe Road	Criminal damage and arson
1	2021-01	Essex Police	Essex Police	On or near Sadlers Close	Violence and sexual offences
2	2021-01	Essex Police	Essex Police	On or near Sadlers Close	Violence and sexual offences
3	2021-01	Essex Police	Essex Police	On or near Brackendale	Anti-social behaviour
4	2021-01	Essex Police	Essex Police	On or near Hillway	Drugs
...	...	...	...	...	...
363432	2022-10	Essex Police	Essex Police	No Location	Violence and sexual offences
363433	2022-10	Essex Police	Essex Police	No Location	Violence and sexual offences
363434	2022-10	Essex Police	Essex Police	No Location	Other crime
363435	2022-10	Essex Police	Essex Police	No Location	Other crime
363436	2022-10	Essex Police	Essex Police	No Location	Other crime

363437 rows × 5 columns

Describing the data's shape

In [8]:
```
essexdata.shape
```

```
(363437, 5)
```

Out[8]:

Viewing data from the head

In [9]:
```
essexdata.head(20)
```

Out[9]:

	Month	Reported by	Falls within	Location	Crime type
0	2021-01	Essex Police	Essex Police	On or near Smythe Road	Criminal damage and arson
1	2021-01	Essex Police	Essex Police	On or near Sadlers Close	Violence and sexual offences
2	2021-01	Essex Police	Essex Police	On or near Sadlers Close	Violence and sexual offences
3	2021-01	Essex Police	Essex Police	On or near Brackendale	Anti-social behaviour
4	2021-01	Essex Police	Essex Police	On or near Hillway	Drugs
5	2021-01	Essex Police	Essex Police	On or near Outwood Farm Close	Public order
6	2021-01	Essex Police	Essex Police	On or near Outwood Farm Close	Violence and sexual offences
7	2021-01	Essex Police	Essex Police	On or near Mount View	Violence and sexual offences
8	2021-01	Essex Police	Essex Police	On or near Sussex Way	Anti-social behaviour
9	2021-01	Essex Police	Essex Police	On or near Sussex Way	Public order
10	2021-01	Essex Police	Essex Police	On or near Springfield Road	Violence and sexual offences
11	2021-01	Essex Police	Essex Police	On or near Lampern Close	Anti-social behaviour
12	2021-01	Essex Police	Essex Police	On or near Graham Close	Burglary
13	2021-01	Essex Police	Essex Police	On or near Sports/Recreation Area	Anti-social behaviour
14	2021-01	Essex Police	Essex Police	On or near Mallow Gardens	Anti-social behaviour
15	2021-01	Essex Police	Essex Police	On or near Dolphin Gardens	Anti-social behaviour
16	2021-01	Essex Police	Essex Police	On or near Oakley Drive	Anti-social behaviour
17	2021-01	Essex Police	Essex Police	On or near Upland Drive	Anti-social behaviour
18	2021-01	Essex Police	Essex Police	On or near Lorrimore Close	Anti-social behaviour
19	2021-01	Essex Police	Essex Police	On or near Temple Close	Criminal damage and arson

In [11]:
```
essexdata.head()
```

Out[11]:

	Month	Reported by	Falls within	Location	Crime type
0	2021-01	Essex Police	Essex Police	On or near Smythe Road	Criminal damage and arson
1	2021-01	Essex Police	Essex Police	On or near Sadlers Close	Violence and sexual offences
2	2021-01	Essex Police	Essex Police	On or near Sadlers Close	Violence and sexual offences
3	2021-01	Essex Police	Essex Police	On or near Brackendale	Anti-social behaviour
4	2021-01	Essex Police	Essex Police	On or near Hillway	Drugs

Data Viewing from the Tail

In [12]:
```
essexdata.tail()
```

Out[12]:

	Month	Reported by	Falls within	Location	Crime type
363432	2022-10	Essex Police	Essex Police	No Location	Violence and sexual offences
363433	2022-10	Essex Police	Essex Police	No Location	Violence and sexual offences
363434	2022-10	Essex Police	Essex Police	No Location	Other crime
363435	2022-10	Essex Police	Essex Police	No Location	Other crime
363436	2022-10	Essex Police	Essex Police	No Location	Other crime

Data information

In [9]:
```
essexdata.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 363437 entries, 0 to 363436
Data columns (total 5 columns):
 #   Column         Non-Null Count   Dtype 
---  ------         --------------   ----- 
 0   Month          363437 non-null  object
 1   Reported by    363437 non-null  object
 2   Falls within   363437 non-null  object
 3   Location       363437 non-null  object
 4   Crime type     363437 non-null  object
dtypes: object(5)
memory usage: 13.9+ MB
```

In [14]:
```
essexdata.describe()
```

Out[14]:

	Month	Reported by	Falls within	Location	Crime type
count	363437	363437	363437	363437	363437
unique	22	1	1	14540	14
top	2021-07	Essex Police	Essex Police	On or near	Violence and sexual offences
freq	18387	363437	363437	12432	142681

In [15]:
```
essexdata.columns
```

Out[15]:
```
Index(['Month', 'Reported by', 'Falls within', 'Location', 'Crime type'], dtype='objec
t')
```

In [16]:
```
essexdata.nunique()
```

Out[16]:
```
Month              22
Reported by         1
Falls within        1
Location        14540
Crime type         14
dtype: int64
```

In [17]:
```
essexdata['Crime type'].unique()
```

Out[17]:
```
array(['Criminal damage and arson', 'Violence and sexual offences',
       'Anti-social behaviour', 'Drugs', 'Public order', 'Burglary',
       'Other theft', 'Other crime', 'Vehicle crime', 'Shoplifting',
       'Theft from the person', 'Possession of weapons', 'Robbery',
       'Bicycle theft'], dtype=object)
```

Determine whether the data collection contains any null values.

In [18]:
```
essexdata.isnull()
```

Out[18]:

	Month	Reported by	Falls within	Location	Crime type
0	False	False	False	False	False
1	False	False	False	False	False
2	False	False	False	False	False
3	False	False	False	False	False
4	False	False	False	False	False
...	...	...	...	...	...
363432	False	False	False	False	False
363433	False	False	False	False	False
363434	False	False	False	False	False
363435	False	False	False	False	False
363436	False	False	False	False	False

363437 rows × 5 columns

In [10]:
```
essexdata.isnull().sum()
```

Out[10]:
```
Month           0
Reported by     0
Falls within    0
Location        0
Crime type      0
dtype: int64
```

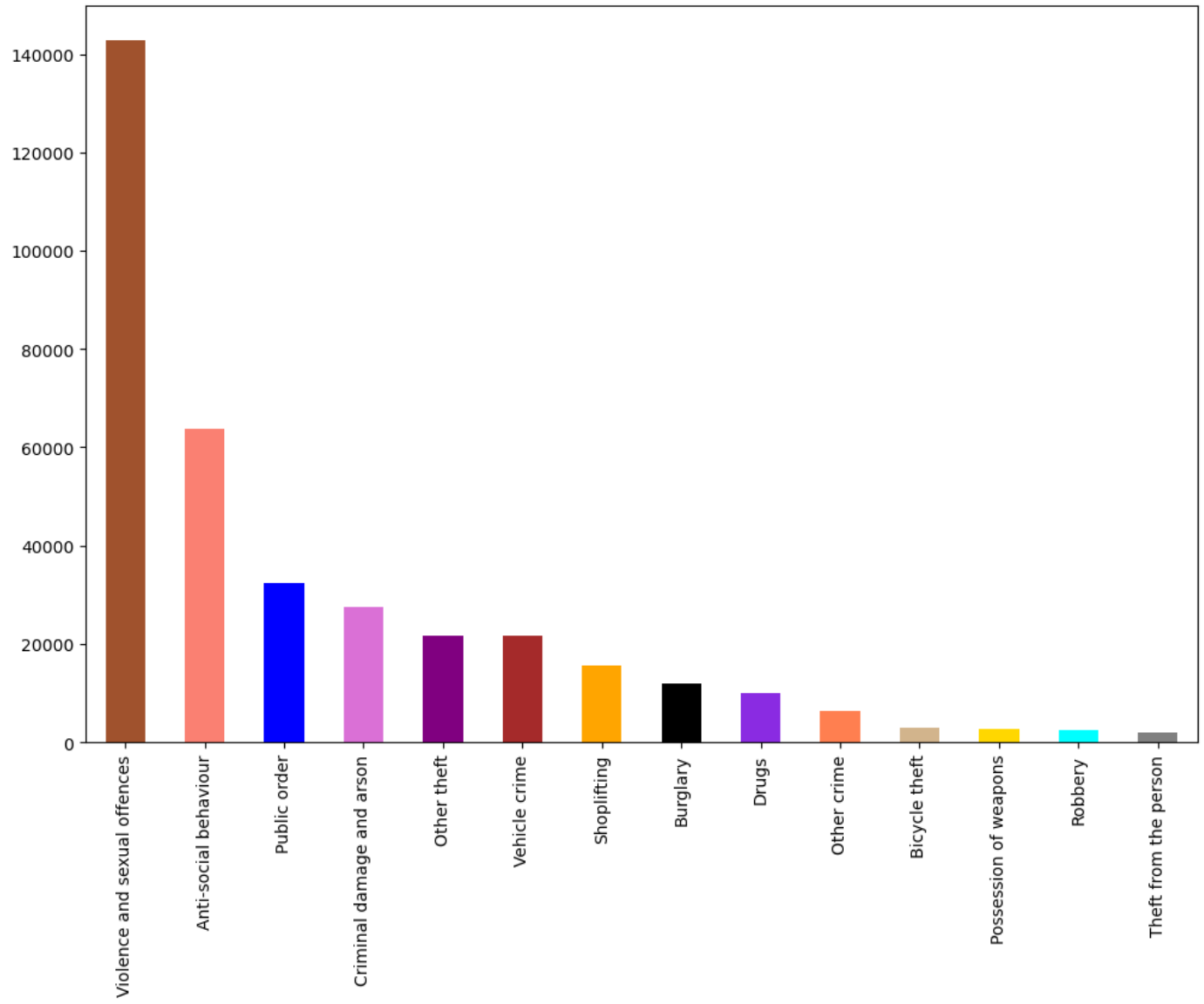
In [19]:
```
essexdata["Crime type"].value_counts()
```

Out[19]:
```
Violence and sexual offences    142681
Anti-social behaviour            63652
Public order                     32320
Criminal damage and arson        27572
Other theft                      21626
Vehicle crime                    21624
Shoplifting                      15576
Burglary                         11895
Drugs                             9987
Other crime                       6405
Bicycle theft                     2870
Possession of weapons             2596
Robbery                           2565
Theft from the person             2068
Name: Crime type, dtype: int64
```

Bar chat

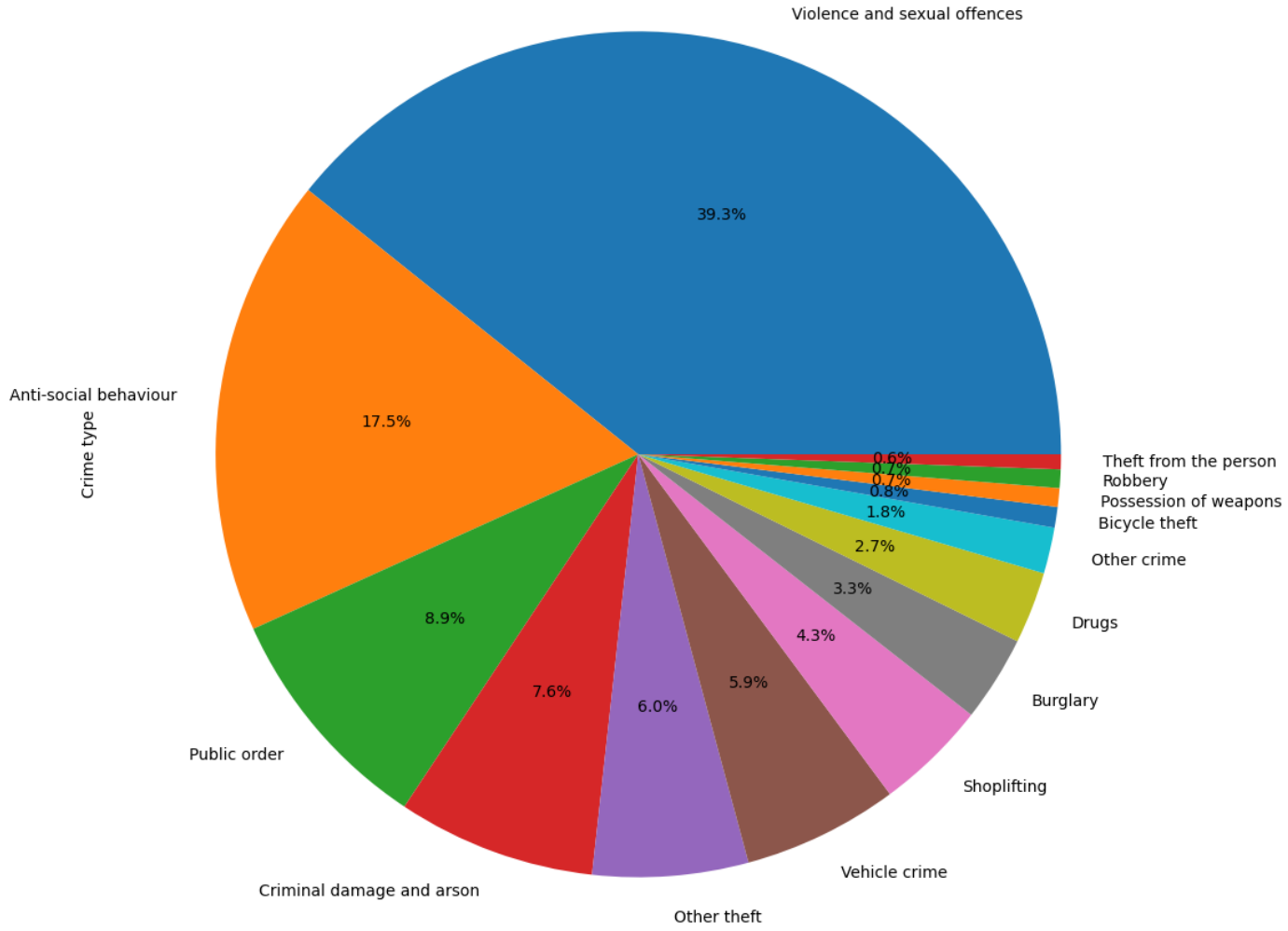
In [20]:
```
plt.figure(figsize=(12,8))
```

```
essexdata["Crime type"].value_counts().plot(kind="bar",color=('sienna', 'salmon', 'blue'
plt.show()
```



Pie chart

In [21]:
```
plt.figure(figsize=(12,12))
essexdata["Crime type"].value_counts().plot(kind="pie", autopct="%1.1f%%")
plt.show()
```

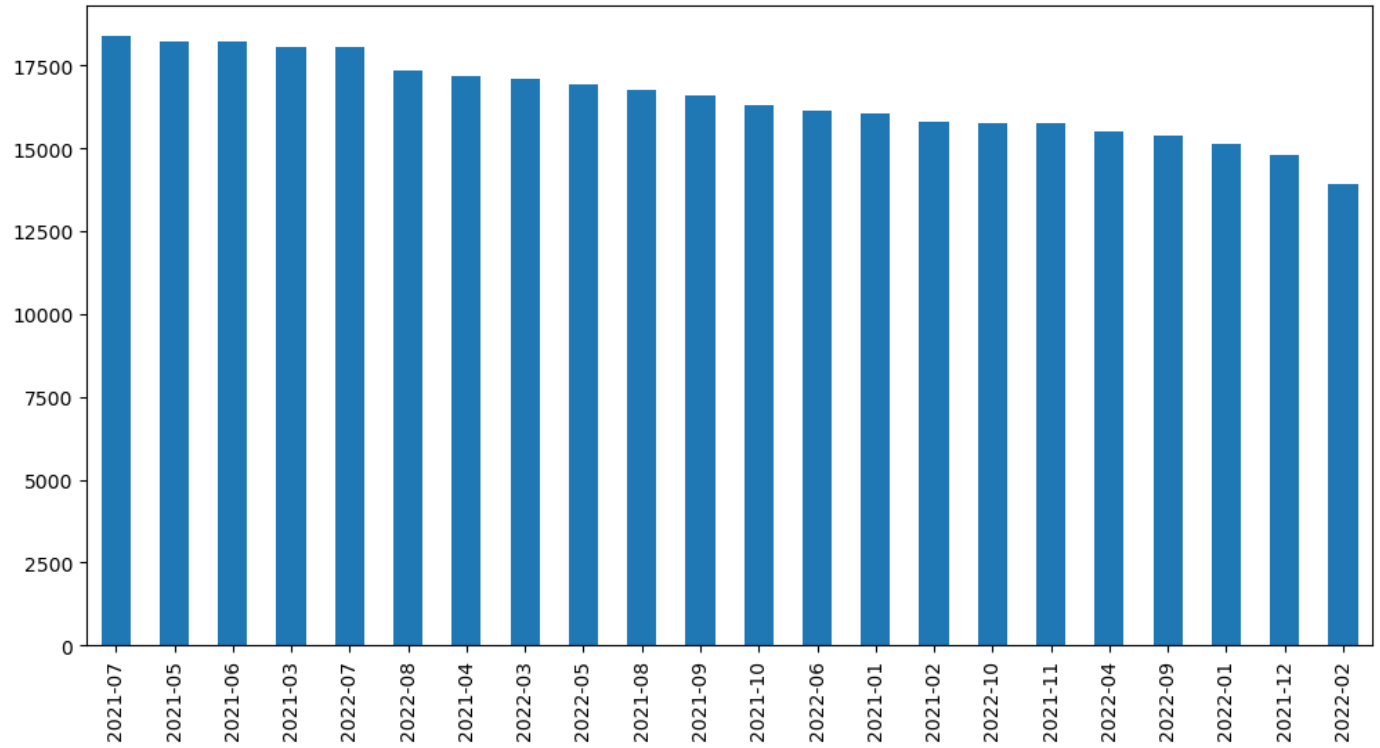


In [22]:
```
essexdata["Month"].value_counts()
```

Out[22]:
```
2021-07    18387
2021-05    18207
2021-06    18206
2021-03    18065
2022-07    18062
2022-08    17344
2021-04    17166
2022-03    17099
2022-05    16927
2021-08    16751
2021-09    16609
2021-10    16324
2022-06    16121
2021-01    16065
2021-02    15796
2022-10    15754
2021-11    15740
2022-04    15530
2022-09    15402
2022-01    15125
2021-12    14815
2022-02    13942
Name: Month, dtype: int64
```

In [23]:
```
plt.figure(figsize=(12,6))
essexdata["Month"].value_counts().plot(kind="bar")
```

```
plt.show()
```



In [24]:
```
# locations and total crimes occured on that location
pd.value_counts(essexdata['Location'])[:1000]
```

Out[24]:
```
On or near                    12432
On or near Parking Area       10465
On or near Supermarket         9994
On or near Petrol Station      6569
On or near Shopping Area       6397
                              ...
On or near Pyms Road             62
On or near Moore Avenue          62
On or near Atlas Road            62
On or near Ashdown Way           62
On or near Orsett End            62
Name: Location, Length: 1000, dtype: int64
```

In [25]:
```
# List of type of crime and there occurence
pd.value_counts(essexdata['Crime type'])[:15]
```

Out[25]:
```
Violence and sexual offences    142681
Anti-social behaviour            63652
Public order                     32320
Criminal damage and arson        27572
Other theft                      21626
Vehicle crime                    21624
Shoplifting                      15576
Burglary                         11895
Drugs                             9987
Other crime                       6405
Bicycle theft                     2870
Possession of weapons             2596
Robbery                           2565
Theft from the person             2068
Name: Crime type, dtype: int64
```

In [34]:
```
# count of offenses for everymonth
#each number corresponds to the respective month
```

```
crime_type_counts = essexdata.groupby(['Month', 'Crime type']).size().reset_index(name='
crime_type_counts
```

Out[34]:

	Month	Crime type	count
0	2021-01	Anti-social behaviour	5509
1	2021-01	Bicycle theft	68
2	2021-01	Burglary	498
3	2021-01	Criminal damage and arson	912
4	2021-01	Drugs	504
...	...	...	...
303	2022-10	Robbery	164
304	2022-10	Shoplifting	826
305	2022-10	Theft from the person	103
306	2022-10	Vehicle crime	1231
307	2022-10	Violence and sexual offences	6309

308 rows × 3 columns

In [16]:
```
essexcrime = essexdata1.iloc[:,[1,4,5,9]]
essexcrime.info()
essexcrime.head()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 363437 entries, 0 to 363436
Data columns (total 4 columns):
 #   Column      Non-Null Count   Dtype  
---  ------      --------------   -----  
 0   Month       363437 non-null  object 
 1   Longitude   360520 non-null  float64
 2   Latitude    360520 non-null  float64
 3   Crime type  363437 non-null  object 
dtypes: float64(2), object(2)
memory usage: 11.1+ MB
```

Out[16]:

	Month	Longitude	Latitude	Crime type
0	2021-01	0.439541	51.639160	Criminal damage and arson
1	2021-01	0.436841	51.638676	Violence and sexual offences
2	2021-01	0.436841	51.638676	Violence and sexual offences
3	2021-01	0.439730	51.629479	Anti-social behaviour
4	2021-01	0.437891	51.626712	Drugs

In [17]:
```
essexcrime2=essexcrime.dropna(axis=0)
essexcrime2
```

Out[17]:

	Month	Longitude	Latitude	Crime type
0	2021-01	0.439541	51.639160	Criminal damage and arson
1	2021-01	0.436841	51.638676	Violence and sexual offences
2	2021-01	0.436841	51.638676	Violence and sexual offences
3	2021-01	0.439730	51.629479	Anti-social behaviour

4	2021-01	0.437891	51.626712	Drugs
...	...	...	...	...
363279	2022-10	0.310917	51.795756	Violence and sexual offences
363280	2022-10	0.314202	51.795044	Violence and sexual offences
363281	2022-10	0.314202	51.795044	Violence and sexual offences
363282	2022-10	0.325673	51.831669	Violence and sexual offences
363283	2022-10	0.325673	51.831669	Violence and sexual offences

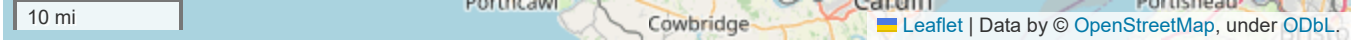
360520 rows × 4 columns

In [18]:
```
countcrime = dict(essexcrime["Crime type"].value_counts())
countcrime
```

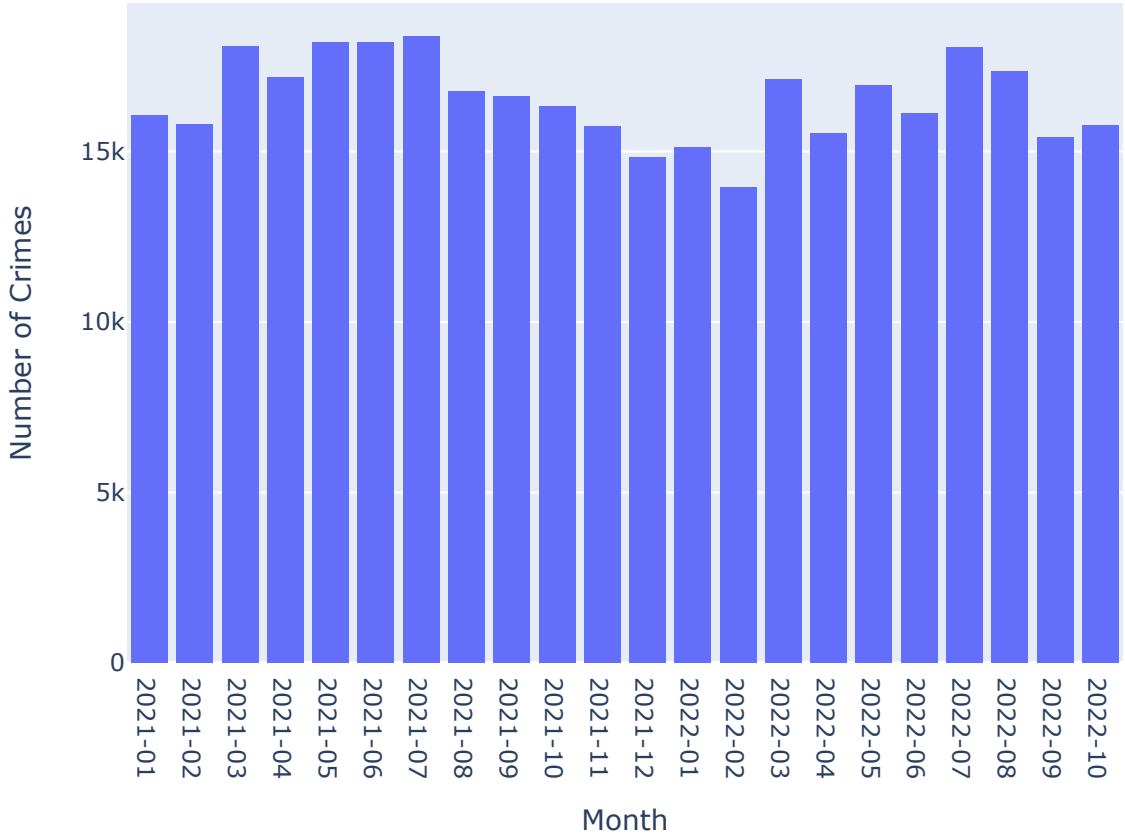
Out[18]:
```
{'Violence and sexual offences': 142681,
 'Anti-social behaviour': 63652,
 'Public order': 32320,
 'Criminal damage and arson': 27572,
 'Other theft': 21626,
 'Vehicle crime': 21624,
 'Shoplifting': 15576,
 'Burglary': 11895,
 'Drugs': 9987,
 'Other crime': 6405,
 'Bicycle theft': 2870,
 'Possession of weapons': 2596,
 'Robbery': 2565,
 'Theft from the person': 2068}
```

In [19]:
```
crime_area = list(zip(essexcrime2.Latitude, essexcrime2.Longitude))
def generateBaseMap(default_location=[51.8, -3.5], default_zoom_start=9):
    base_map = folium.Map(location=default_location, control_scale=True, zoom_start=defa
    heatmap = plugins.HeatMap(crime_area, radius=5, blur=2)
    base_map.add_child(heatmap)
    return base_map
generateBaseMap()
```

Out[19]:



In [20]:
```
fig = px.histogram(essexdata,
             x="Month",
             title="Levels of Crime in essex")
fig.update_layout(yaxis_title="Number of Crimes")
fig.update_xaxes(type='category')
fig.show()
```



In [23]:
```
crime_by_neighbourhoods_totals = essexdata1.groupby(['LSOA code','LSOA name']).size().re
crime_by_neighbourhoods_totals = crime_by_neighbourhoods_totals.sort_values(by='count',
crime_by_neighbourhoods_totals.head()
```

Out[23]:

	LSOA code	LSOA name	count
1077	E01023453	East Hertfordshire 003B	1
1092	E01030099	St Edmundsbury 010A	1
1091	E01029906	Babergh 009D	1
1090	E01029905	Babergh 010G	1
1086	E01028033	Bassetlaw 015C	1

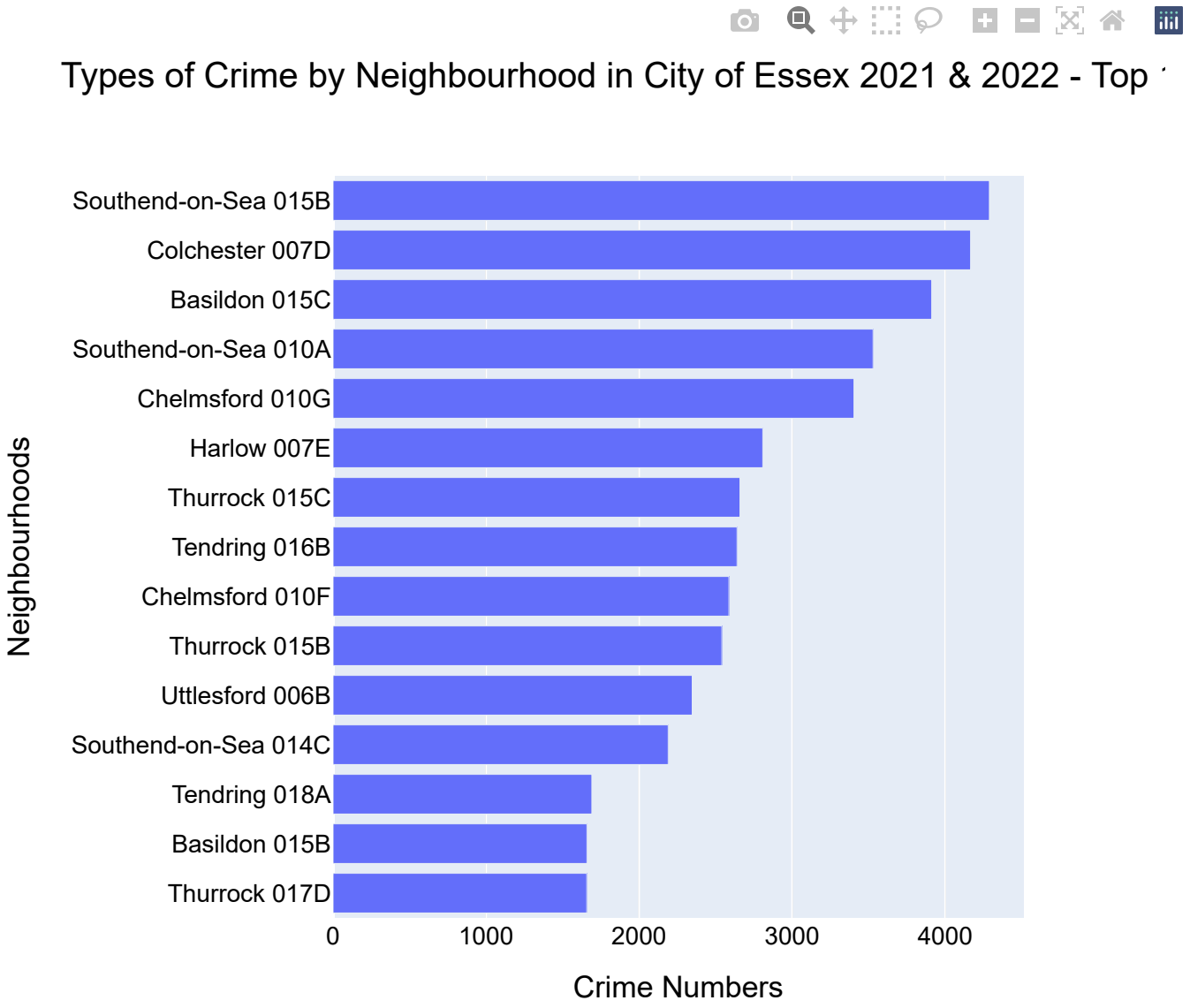
In [24]:
```
# Visualising crime by LSOA_code could use a horizontal bar chart
# The visualisation was limited to the top 10 neighbourhoods for clarity

fig = px.bar(crime_by_neighbourhoods_totals[-15:], x="count", y="LSOA name", orientation
             hover_data=["LSOA code", "LSOA name","count"],
```

```
            height=600,
            title='Types of Crime by Neighbourhood in City of Essex 2021 & 2022 - Top 1
fig.update_layout(

    xaxis_title="Crime Numbers",
    yaxis_title="Neighbourhoods",
    font=dict(
        family="Arial, monospace",
        size=14,
        color="Black"
    ))

fig.show()
```



In [56]:
```
# Another pandas dataframe is needed to visualise the neighbourhood crime by type

type_by_neighbourhoods_totals = essexdata1.groupby(['Crime type','LSOA code','LSOA name'
type_by_neighbourhoods_totals = type_by_neighbourhoods_totals.sort_values(by=['count','L
type_by_neighbourhoods_totals = type_by_neighbourhoods_totals.query
type_by_neighbourhoods_totals
```

Out[56]:
```
<bound method DataFrame.query of                                Crime type  LSOA code
    LSOA name  count
2752                          Burglary  E01029885          Babergh 002C      1
8845                     Public order  E01029880          Babergh 006A      1
12722  Violence and sexual offences  E01029880          Babergh 006A      1
1065           Anti-social behaviour  E01029906          Babergh 009D      1
```

```
12723  Violence and sexual offences  E01029905          Babergh 010G      1
...                             ...         ...               ...    ...
12737  Violence and sexual offences  E01033141       Chelmsford 010G   1273
11951  Violence and sexual offences  E01021318         Basildon 015C   1360
11751  Violence and sexual offences  E01015895  Southend-on-Sea 010A   1413
12273  Violence and sexual offences  E01021649       Colchester 007D   1473
11708  Violence and sexual offences  E01015852  Southend-on-Sea 015B   1584

[12757 rows x 4 columns]>
```

In [ ]: