Identifying Strategies to help Prevent Crime in the US

February 3, 2023

1 Identifying Strategies to help Prevent Crime in the US

[1]:
```
#Importing Libraries & Setting Parameters
import pandas as pd
import plotly.express as px
import seaborn as sns
import plotly.graph_objects as go
from plotly.subplots import make_subplots
from scipy.interpolate import interp1d
import matplotlib.pyplot as plt

pd.set_option('display.max_rows', None)
pd.set_option('display.max_columns', None)
```

1.1 Setting the Context

Preventing crime is just as important as detecting crime once it has occurred. **A crime that doesn't occur is a victim that hasn't been created**. By having a sustained focus on prevention, we can reduce the damage caused by crime, both in economic terms and to the lives it ruins.

USA has started witnessing an increase in volume of crimes after a long period of steady decline. Now is the time to devise policies and take action to prevent crimes and curb this uptick. USA has had a violent history in crime, and the purpose of this report is to identify strategies that have worked in controlling crime in the history to advise current decision makers.

The report begins by organising & checking the data in the **'Data Cleaning'** section. This is followed by initial exploration of crime trends in the US History in the **'Initial Exploration'** section. Once we identify the pattern of crime, reasons driving this pattern are explored in the **'Investigating Socio-Economic Factors'** section.

[2]:
```
#Plotting Context Graph
introduction = pd.read_csv('./us_new_data.txt', header=0,sep='\t',encoding =␣
↪"ISO-8859-1")
fig = px.line(introduction[introduction['year']>=2000], x="year", y=["crimes"])
fig.update_layout(title_text="# of Crimes increasing in the US in recent years")
fig.update_yaxes(title_text="# of Crimes")
fig.update_yaxes(title_text="Year")
```

```
fig.show()
```

1.2 Data Cleaning

[3]:
```
#Data Source - https://www.kaggle.com/datasets/marshallproject/crime-rates
#Importing Data for Analysis
crimes_raw = pd.read_csv('./crime_report_us.txt', header=0,sep='\t',encoding =␣
 ↪"ISO-8859-1")
region_data = pd.read_csv('./us_region_map.txt', header=0,sep='\t',encoding =␣
 ↪"ISO-8859-1")

print('Historical Crime Data Imported')
```

```
Historical Crime Data Imported
```

[4]:
```
#Checking Data Schema
crimes_raw.head()
```

[4]:

	report_year	agency_jurisdiction	population	violent_crimes	homicides	rapes	assaults	robberies
0	1975	Albuquerque, NM	286238	2383	30	181	1353	819
1	1976	Albuquerque, NM	292265	2420	28	186	1335	871
2	1977	Albuquerque, NM	292341	2390	31	207	1398	754
3	1978	Albuquerque, NM	291834	2434	37	187	1516	694
4	1979	Albuquerque, NM	302120	2679	47	215	1602	815

[5]:
```
crimes_raw['state'] = crimes_raw['agency_jurisdiction'].apply(lambda s: s.
 ↪split(', ')[1])
crimes_raw['city'] = crimes_raw['agency_jurisdiction'].apply(lambda s: s.
 ↪split(', ')[0])
print('Segregated States and Cities')
```

```
Segregated States and Cities
```

[6]:
```
list_of_columns = list(crimes_raw.columns)
list_of_columns.extend(['region','state_name'])
crimes_data = crimes_raw.merge(region_data,␣
 ↪how='left',left_on='state',right_on='state_code')[list_of_columns]
print('Added regions and state names to original data set')
```

```
Added regions and state names to original data set
```

[7]:
```
print('The dataset has '+str(crimes_data['report_year'].count())+' rows.')
```

```
The dataset has 2753 rows.
```

[8]:
```
#Checking for missing values
for column_name in crimes_data.columns:
    if crimes_data[column_name].count()==crimes_data.shape[0]:
        print('No missing values detected in '+column_name)
    else:
        print('Missing values detected in '+column_name)
```

```
No missing values detected in report_year
No missing values detected in agency_jurisdiction
No missing values detected in population
No missing values detected in violent_crimes
No missing values detected in homicides
No missing values detected in rapes
No missing values detected in assaults
No missing values detected in robberies
No missing values detected in state
No missing values detected in city
No missing values detected in region
No missing values detected in state_name
```

1.3 Initial Exploration - Checking for Trends in USA's History

[9]:
```
#Checking if crime was consistent across states
fig = make_subplots(rows=1, cols=2)
fig.add_trace(go.Histogram(y=crimes_data['state_name'],␣
 ↪x=crimes_data['population'], name="Population",
                           orientation='h', histfunc='sum'),row=1, col=1)
fig.add_trace(go.Histogram(y=crimes_data['state_name'],␣
 ↪x=crimes_data['violent_crimes'],name="Crimes",
                           orientation='h', histfunc='sum'),row=1, col=2)
fig.update_layout(height=800, width=950, title_text="Volume of Crimes across␣
 ↪States: 1975-2015")
fig.show()
print('''New York, California and Texas have the highest volume of crimes.
However, these are also among the most populous states.''')
```

```
New York, California and Texas have the highest volume of crimes.
However, these are also among the most populous states.
```

[10]:
```
crimes_data_state_year=crimes_data.groupby(['state_name','report_year']).
 ↪agg('sum').reset_index()
print('Aggregated Data to State + Year Level')
```

```
Aggregated Data to State + Year Level
```

[11]:
```
fig = px.line(crimes_data_state_year, x="report_year",␣
 ↪y=["violent_crimes"],color = "state_name")
fig.update_layout(title_text="Total Crimes over the Years across States")
fig.update_yaxes(title_text="# of Crimes")
fig.update_yaxes(title_text="Year")
fig.show()
print('''While some states have experienced lesser variation over time compared␣
 ↪to others, it is clear that
most states have experienced a peak between 1990 and 1995 and a steep decline␣
 ↪since then.''')
```

```
While some states have experienced lesser variation over time compared to
others, it is clear that
most states have experienced a peak between 1990 and 1995 and a steep decline
since then.
```

[12]:
```
crimes_data_state=crimes_data.groupby(['state_name']).agg('sum').reset_index()
crimes_data_state["violent_crimes_percapita"]=crimes_data_state["violent_crimes"]/
 ↪crimes_data_state["population"]
print('Aggregated Data to State Level')
```

```
Aggregated Data to State Level
```

[13]:
```
fig = make_subplots(rows=1, cols=1)
fig.add_trace(go.Histogram(y=crimes_data_state['state_name'],␣
 ↪x=crimes_data_state['violent_crimes_percapita'],
                           name="violent_crimes_percapita",orientation='h',␣
 ↪histfunc='avg'),row=1, col=1)
fig.update_layout(height=1200, width=950, title_text="Per Capita Crime across␣
 ↪States: 1975-2015")
fig.update_xaxes(title_text="Year")
fig.show()

print('''Looking at Per Capita Crime changes the picture, as we take the␣
 ↪population out of the picture,
we see that Georgia, New Jersey & Miami have the per capita highest crime rate.
 ↪''')
```

```
Looking at Per Capita Crime changes the picture, as we take the population out
of the picture,
we see that Georgia, New Jersey & Miami have the per capita highest crime rate.
```

[14]:
```
crimes_data_state_year["violent_crimes_percapita"]=crimes_data_state_year["violent_crimes"]/
 ↪crimes_data_state_year["population"]
fig = px.line(crimes_data_state_year, x="report_year",␣
 ↪y=["violent_crimes_percapita"],color = "state_name")
fig.update_layout(title_text="Per Capita Crimes over the Years across States")
```

```
fig.update_yaxes(title_text="Per Capita Crimes")
fig.update_yaxes(title_text="Year")
graph1 = fig
fig.show()
print('''The decline in the 90s is clear even when we look at per capita crime␣
 ↪across states ''')
```

```
The decline in the 90s is clear even when we look at per capita crime across
states
```

[15]:
```
#Checking how National Crimes changes over time
crimes_data_nation=crimes_data.groupby(['report_year']).agg('sum').reset_index()
fig = px.line(crimes_data_nation, x="report_year",␣
 ↪y=["violent_crimes","homicides","rapes","assaults",
                                                    "robberies"])
fig.update_layout(title_text="National Crimes over the Years")
fig.update_yaxes(title_text="Volume of Violent Crimes")
fig.show()
print('''The same trend is noted for the US as a whole. This seems to be␣
 ↪majorly driven by assaults and robberies.''')
```

```
The same trend is noted for the US as a whole. This seems to be majorly driven
by assaults and robberies.
```

1.4 Investigating Socio-Economic Factors (Focus - Overall US)

Data Sources - Bureau of Labour Statistics, Bureau of Justice Statistics, World Prison Brief, US Census

[16]:
```
#Importing Socio-Economic Data
soc_econ_data = pd.read_csv('./us_soc_eco_data.txt', header=0,sep='\t',encoding␣
 ↪= "ISO-8859-1")
crimes_data_nation_merged = crimes_data_nation.merge(soc_econ_data,␣
 ↪how='inner',left_on='report_year',right_on='year')
crimes_data_nation_merged['police_force_size_percapita'] =␣
 ↪crimes_data_nation_merged['police_force_size']/
 ↪crimes_data_nation_merged['population']
print('National Socio-Economic Data imported and merged')
```

```
National Socio-Economic Data imported and merged
```

[17]:
```
#Defining list of factors to be analysed
dict_of_factors = {'unemp_rate' : 'Unemployment Rate',
                   'cons_price_index' : 'Consumer Price Index',
                   'police_force_size_percapita' : 'Size of Police Force per␣
 ↪Capita',
                   'prison_population' : 'Prison Population',
                   'inflation' : 'Inflation',
```

```
                    'poverty_population' : 'Population under Poverty'}
dict_of_rows = {'unemp_rate' : 1,
                    'cons_price_index' : 1,
                    'police_force_size_percapita' : 2,
                    'prison_population' : 2,
                    'inflation' : 3,
                    'poverty_population' : 3}
dict_of_cols = {'unemp_rate' : 1,
                    'cons_price_index' : 2,
                    'police_force_size_percapita' : 1,
                    'prison_population' : 2,
                    'inflation' : 1,
                    'poverty_population' : 2}
```

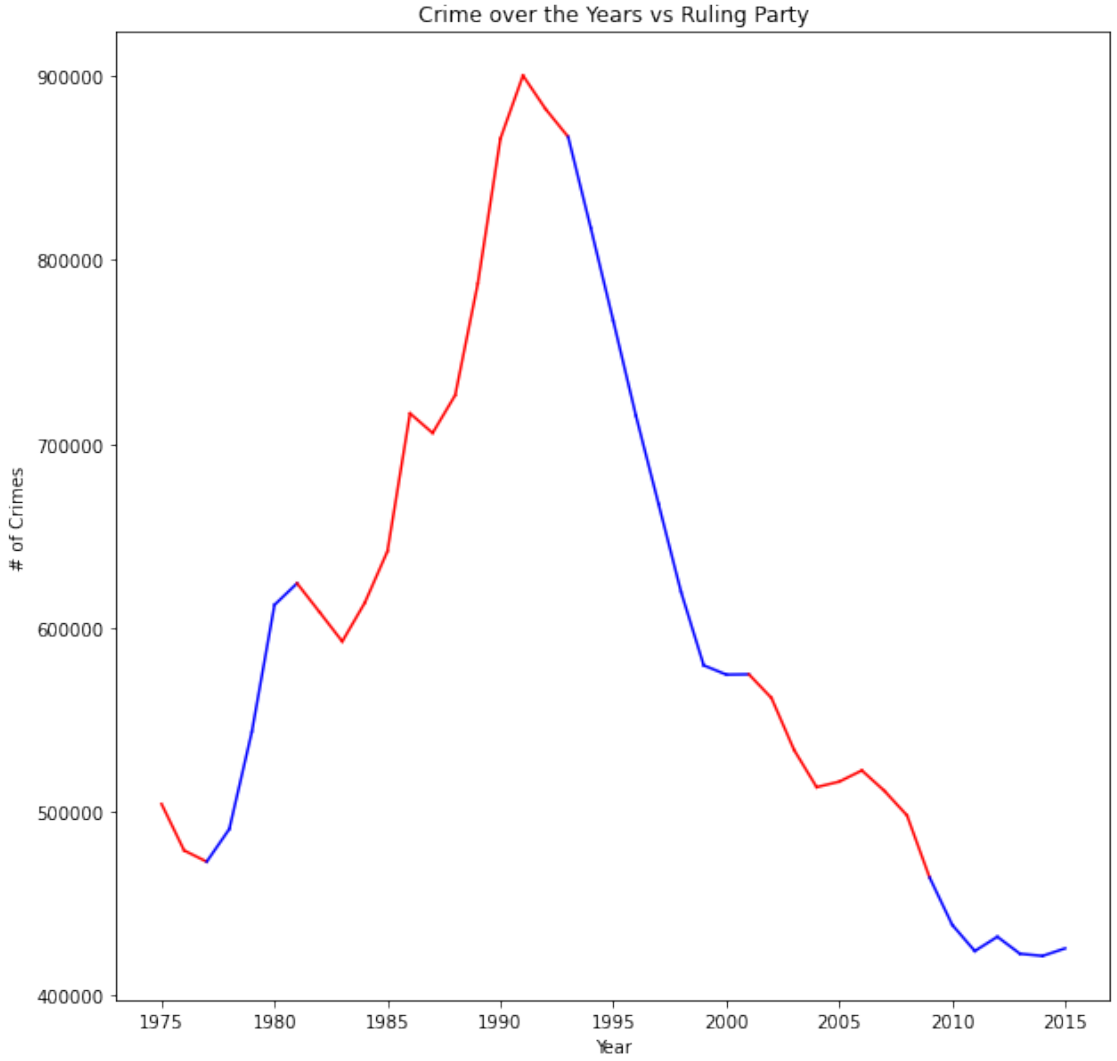
[18]:
```
def plotting_factors(dataframe,crime,factor_dict,rows,cols):
    fig = make_subplots(specs=[[{"secondary_y": True},{"secondary_y": True}],
                                       [{"secondary_y": True},{"secondary_y":␣
 ↪True}],
                                       [{"secondary_y": True},{"secondary_y":␣
 ↪True}]],
                            rows = rows, cols = cols,
                        subplot_titles=list(factor_dict.values()))
    fig.update_layout(title_text="National Crimes & Socio-Economic Factors over␣
 ↪the Years",height=600, width=1000)
    fig.
 ↪update_layout(legend=dict(orientation="h",itemwidth=70,yanchor="bottom",y=-0.
 ↪2,xanchor="right",x=0.9))
    for factor in list(dict_of_factors.keys()):
        fig.add_trace(go.Scatter(x=dataframe['report_year'], y=dataframe[crime],
                         name='# of Crimes',showlegend=False),␣
 ↪secondary_y=False,row=dict_of_rows[factor],
                      col=dict_of_cols[factor])
        fig.add_trace(go.Scatter(x=dataframe['report_year'],␣
 ↪y=dataframe[factor],
                                 ␣
 ↪name=factor_dict[factor]),secondary_y=True,row=dict_of_rows[factor],
                      col=dict_of_cols[factor])
        fig.update_traces(connectgaps=True)

    return fig
```

[19]:
```
#Plotting Socio-Economic Factors
graph2 =␣
 ↪plotting_factors(crimes_data_nation_merged,'violent_crimes',dict_of_factors,3,2)
graph2
```

[20]:
```
#Plotting Ruling Parties:

colours = {'Republican': 'red', 'Democrat':'blue'}
plt.figure(figsize=(10, 10))
for data in range(crimes_data_nation_merged.shape[0]-1):
    plt.plot([crimes_data_nation_merged['report_year'][data],
                                crimes_data_nation_merged['report_year'][data+1]],
             [crimes_data_nation_merged['violent_crimes'][data],
                                ␣
 ↪crimes_data_nation_merged['violent_crimes'][data+1]],
             color=colours[crimes_data_nation_merged['party'][data]])
plt.xlabel("Year")
plt.ylabel("# of Crimes")
plt.title('Crime over the Years vs Ruling Party')
plt.show()
```



2 Mini Report (246 Words)

2.0.1 Problem Statement:

Followed by a consistent decline, USA has started witnessing an increase in volume of crimes. Now is the time to devise policies and take action to curb this uptick. The purpose of this report is to understand USA's history of battling crime & identifying strategies that have succeeded to advise current decision makers.

2.0.2 Insights Drawn:

Break it down by region While the trend may be similar, some regions might have a higher contribution to crime compared to others. A drilldown by region can help in creating targeted strategies and larger impact by prioritising the high contribution states. (Graph 1)

Factor Police Force per Capita While increasing the size of the police force is not the most economical solution, a temporary focus on staffing might prove to be helpful. This was seen in the 1990s (Graph 2). Rise in the per capita police coverage accompanied a dip in volume of crimes and a steeper increase in the prison population. Once the crime rate normalized, the per capita coverage was slowly brought down, the same strategy may be applied in the present

Identify the type of crime and tackle the associated economic factors As inflation & unemployment was controlled, a dip was noted in volume of crimes. Robberies played a big role in the increase in crime. Inflation & unemployment act as a source of motivation for robberies. Naturally, identifying the type of crime and tackling related economic parameters, might be effective. (Graph 2)

[21]:
```
graph1
```

[22]:
```
graph2
```