# Christian Campbell - Milestone 5

In [47]:
```
import sqlite3 as sql
import numpy as np
import requests
import re
import ssl
from bs4 import BeautifulSoup
import urllib.request, urllib.parse
from urllib.error import HTTPError, URLError
import socket
import json
import matplotlib.pyplot as plt
import ast
import seaborn as sns
import plotly.express as px
import pandas as pd
import plotly.graph_objects as go
import plotly.subplots as sp
```

# Importing website data

In [2]:
```
database = "website_crime_data"
connection = sql.connect(database)
```

In [3]:
```
# The code below checks if the connection open.
def is_opened(connection):
    try:
        connection.execute("SELECT * FROM website_crime_data LIMIT 1")
        return True
    except sql.ProgrammingError as e:
        print("Connection closed {}".format(e))
        return False
```

In [4]:
```
print(is_opened(connection))
```

```
True
```

In [7]:

```
#This code brings up the table
website_data="website_crime_data"
query = f"SELECT * FROM {website_data}"
try:
    web_data = pd.read_sql(query, connection)
    print(web_data)
except sql.Error as e:
    print(f"Error reading data from the table: {e}")
```

| | Location | Violent crime rate | Homicide rate | Rape rate \ |
|---|---|---|---|---|
| 0 | Alabama | 409.1 | 10.9 | 29.6 |
| 1 | Alaska | 758.9 | 9.5 | 134.0 |
| 2 | Arizona | 431.5 | 6.8 | 44.1 |
| 3 | Arkansas | 645.3 | 10.2 | 76.0 |
| 4 | California | 499.5 | 5.7 | 37.4 |
| 5 | Colorado | 492.5 | 6.4 | 63.4 |
| 6 | Connecticut | 150.0 | 3.8 | 18.1 |
| 7 | Delaware | 383.5 | 4.8 | 22.0 |
| 8 | District of Columbia | 812.3 | 29.3 | 41.5 |
| 9 | Florida | 258.9 | 5.0 | 30.2 |
| 10 | Georgia | 367.0 | 8.2 | 36.4 |
| 11 | Hawaii | 259.6 | 2.1 | 37.9 |
| 12 | Idaho | 241.4 | 2.7 | 48.7 |
| 13 | Illinois | 287.3 | 7.8 | 48.1 |
| 14 | Indiana | 306.2 | 6.2 | 32.8 |
| 15 | Iowa | 286.5 | 1.7 | 42.5 |
| 16 | Kansas | 414.6 | 4.6 | 45.5 |
| 17 | Kentucky | 214.1 | 6.8 | 33.8 |
| 18 | Louisiana | 628.6 | 16.1 | 43.0 |
| 19 | Maine | 103.3 | 2.2 | 32.0 |
| 20 | Maryland | 398.5 | 8.5 | 30.6 |
| 21 | Massachusetts | 322.0 | 2.1 | 29.1 |
| 22 | Michigan | 461.0 | 6.9 | 64.8 |
| 23 | Minnesota | 280.6 | 3.2 | 40.7 |
| 24 | Mississippi | 245.0 | 7.8 | 33.7 |
| 25 | Missouri | 488.0 | 10.1 | 48.9 |
| 26 | Montana | 417.9 | 4.5 | 54.4 |
| 27 | Nebraska | 282.8 | 3.2 | 55.3 |
| 28 | Nevada | 454.0 | 6.8 | 58.9 |
| 29 | New Hampshire | 125.6 | 1.8 | 39.6 |
| 30 | New Jersey | 202.9 | 3.1 | 16.8 |
| 31 | New Mexico | 780.5 | 12.0 | 54.6 |
| 32 | New York | 429.3 | 4.0 | 29.5 |
| 33 | North Carolina | 405.1 | 8.1 | 30.5 |
| 34 | North Dakota | 279.6 | 3.5 | 56.7 |
| 35 | Ohio | 293.6 | 6.1 | 48.4 |
| 36 | Oklahoma | 419.7 | 6.7 | 57.5 |
| 37 | Oregon | 342.4 | 4.5 | 40.6 |
| 38 | Pennsylvania | 279.9 | 7.9 | 29.5 |
| 39 | Rhode Island | 172.3 | 1.5 | 38.0 |
| 40 | South Carolina | 491.3 | 11.2 | 38.2 |
| 41 | South Dakota | 377.4 | 4.3 | 55.8 |

| | | | | |
|---|---|---|---|---|
| 42 | Tennessee | 621.6 | 8.6 | 38.2 |
| 43 | Texas | 431.9 | 6.7 | 50.0 |
| 44 | Utah | 241.8 | 2.0 | 59.5 |
| 45 | Vermont | 221.9 | 3.4 | 36.8 |
| 46 | Virginia | 234.0 | 7.3 | 30.2 |
| 47 | Washington | 375.6 | 5.0 | 39.2 |
| 48 | West Virginia | 277.9 | 4.6 | 44.4 |
| 49 | Wisconsin | 297.0 | 5.3 | 38.6 |
| 50 | Wyoming | 201.9 | 2.6 | 62.8 |

| | Robbery rate | Aggrevated Assault rate |
|---|---|---|
| 0 | 34.5 | 334.1 |
| 1 | 75.1 | 540.2 |
| 2 | 70.1 | 310.5 |
| 3 | 39.7 | 519.4 |
| 4 | 123.5 | 332.8 |
| 5 | 72.6 | 350.1 |
| 6 | 44.9 | 83.3 |
| 7 | 57.0 | 299.8 |
| 8 | 357.5 | 383.9 |
| 9 | 33.6 | 190.1 |
| 10 | 43.6 | 278.8 |
| 11 | 66.1 | 153.5 |
| 12 | 8.2 | 181.7 |
| 13 | 84.7 | 146.7 |
| 14 | 43.0 | 224.2 |
| 15 | 21.6 | 220.7 |
| 16 | 29.2 | 335.4 |
| 17 | 38.1 | 135.4 |
| 18 | 67.3 | 502.1 |
| 19 | 10.0 | 59.0 |
| 20 | 114.2 | 245.2 |
| 21 | 37.7 | 253.1 |
| 22 | 36.6 | 352.7 |
| 23 | 57.0 | 179.7 |
| 24 | 25.6 | 178.0 |
| 25 | 54.8 | 374.2 |
| 26 | 23.3 | 335.7 |
| 27 | 29.1 | 195.2 |
| 28 | 86.1 | 302.3 |
| 29 | 16.1 | 68.1 |
| 30 | 47.6 | 135.4 |
| 31 | 110.6 | 603.3 |

| | | |
|---|---|---|
| 32 | 112.0 | 283.8 |
| 33 | 54.9 | 311.6 |
| 34 | 27.6 | 191.8 |
| 35 | 53.1 | 185.9 |
| 36 | 40.6 | 314.8 |
| 37 | 68.6 | 228.7 |
| 38 | 68.1 | 174.5 |
| 39 | 24.6 | 108.3 |
| 40 | 40.6 | 401.3 |
| 41 | 25.3 | 292.0 |
| 42 | 67.1 | 507.6 |
| 43 | 70.5 | 304.7 |
| 44 | 29.6 | 150.7 |
| 45 | 13.3 | 168.5 |
| 46 | 38.4 | 158.1 |
| 47 | 86.8 | 244.7 |
| 48 | 10.0 | 218.9 |
| 49 | 39.4 | 213.7 |
| 50 | 7.9 | 128.7 |

In [43]:
```
# This code plots a pie chart for Violent crime rate in the US States.
# The chart is interactive.
fig = px.pie(web_data, names='Location', values='Violent crime rate', title='Violent Crime Rate (2022)')
fig.show()
```

## Importing flatfile data

In [9]:
```
# This code is to establish a connection to the flatfile database
database2 = "flatfile_crime_data"
conn = sql.connect(database2)
```

In [12]:
```
# The code below checks if the connection open.
def is_opened(conn):
    try:
        conn.execute("SELECT * FROM flatfile_crime_data LIMIT 1")
        return True
    except sql.ProgrammingError as e:
        print("Connection closed {}".format(e))
        return False
```

In [13]:
```
print(is_opened(conn))
```

```
True
```

In [50]:

```
#This code brings up the flatfile table
flat_data="flatfile_crime_data"
query = f"SELECT * FROM {flat_data}"
try:
    flat_data = pd.read_sql(query, conn)
    print(flat_data)
except sql.Error as e:
    print(f"Error reading data from the table: {e}")
```

```
      jurisdiction  includes_jails  year  prisoner_count  \
0          ALABAMA               0  2001           24741
1           ALASKA               1  2001            4570
2          ARIZONA               0  2001           27710
3         ARKANSAS               0  2001           11489
4       CALIFORNIA               0  2001          157142
..             ...             ...   ...             ...
795       VIRGINIA               0  2016           29882
796     WASHINGTON               0  2016           17228
797  WEST VIRGINIA               0  2016            5899
798      WISCONSIN               0  2016           23163
799        WYOMING               0  2016            2352

     crime_reporting_change  crimes_estimated  state_population  \
0                       0.0               0.0         4468912.0
1                       0.0               0.0          633630.0
2                       0.0               0.0         5306966.0
3                       0.0               0.0         2694698.0
4                       0.0               0.0        34600463.0
..                      ...               ...               ...
795                     0.0               0.0         8414380.0
796                     0.0               0.0         7280934.0
797                     0.0               0.0         1828637.0
798                     0.0               0.0         5772917.0
799                     0.0               0.0          584910.0

     violent_crime_total  murder_manslaughter  rape_legacy  robbery  \
0                19582.0                379.0       1369.0   5584.0
1                 3735.0                 39.0        501.0    514.0
2                28675.0                400.0       1518.0   8868.0
3                12190.0                148.0        892.0   2181.0
4               212867.0               2206.0       9960.0  64614.0
..                   ...                  ...          ...      ...
795              18495.0                482.0          NaN   4826.0
796              22101.0                195.0          NaN   5649.0
797               6633.0                 85.0          NaN    720.0
798              17716.0                232.0          NaN   4707.0
799               1431.0                 20.0          NaN     59.0

     agg_assault  property_crime_total  burglary   larceny  vehicle_theft
0        12250.0              173253.0   40642.0  119992.0        12619.0
1         2681.0               23160.0    3847.0   16695.0         2618.0
2        17889.0              293874.0   54821.0  186850.0        52203.0
```

```
3         8969.0              99106.0   22196.0   69590.0         7320.0
4       136087.0            1134189.0  232273.0  697739.0       204177.0
..           ...                  ...       ...       ...            ...
795      10357.0             157292.0   20159.0  127285.0         9848.0
796      13124.0             254994.0   49249.0  173423.0        32322.0
797       5144.0              37282.0    9127.0   25657.0         2498.0
798      10772.0             111911.0   19498.0   82455.0         9958.0
799       1146.0              11460.0    1771.0    8889.0          800.0

[800 rows x 16 columns]
```

In [16]:

```
#This code plots a bar graph for the US prisoner population for 2016
filtered_data = flat_data[flat_data['year'] == 2016]
fig = px.bar(filtered_data, x='jurisdiction', y='prisoner_count', title='Prisoner Count by Jurisdiction i
fig.show()
```

# Importing api data

In [32]:
```
# This code is to establish a connection to the api database
db3 = "api_crime_data"
conn2 = sql.connect(db3)
```

In [33]:
```
# The code below checks if the connection open.
def is_opened(conn2):
    try:
        conn2.execute("SELECT * FROM api_crime_data LIMIT 1")
        return True
    except sql.ProgrammingError as e:
        print("Connection closed {}".format(e))
        return False
```

In [34]:
```
print(is_opened(conn2))
```

```
True
```

In [45]:

```
#This code brings up the api table
api_data="api_crime_data"
query = f"SELECT * FROM {api_data}"
try:
    api_data = pd.read_sql(query, conn2)
    print(api_data)
except sql.Error as e:
    print(f"Error reading data from the table: {e}")
```

```
            State  Year  Aggravated Assault  Manslaughter by Negligence  \
0         Alabama  2016              4887.0                        13.0
1          Alaska  2016              1845.0                         5.0
2         Arizona  2016              9402.0                        42.0
3        Arkansas  2016              3929.0                        10.0
4      California  2016             87210.0                       224.0
..            ...   ...                ...                         ...
294      Virginia  2021              4957.0                        55.0
295    Washington  2021              6058.0                        16.0
296  West Virginia  2021              1246.0                         4.0
297     Wisconsin  2021              5394.0                        53.0
298       Wyoming  2021               279.0                         2.0

     Murder and Nonnegligent Manslaughter    Rape  Robbery  Simple Assault  \
0                                   358.0   407.0   1347.0         19419.0
1                                    41.0   130.0    282.0          4300.0
2                                   253.0   361.0   1893.0         27540.0
3                                   142.0   298.0    573.0         12248.0
4                                  1441.0  2568.0  15895.0         81092.0
..                                    ...     ...      ...             ...
294                                 387.0   503.0    977.0         27754.0
295                                 178.0   518.0   1489.0         23044.0
296                                  43.0    99.0     46.0          3849.0
297                                 194.0   911.0    756.0         15223.0
298                                  13.0    38.0     12.0          1494.0

     Human Trafficking - Commercial Sex Acts  \
0                                        0.0
1                                        0.0
2                                        0.0
3                                        0.0
4                                        0.0
..                                       ...
294                                     11.0
295                                     22.0
296                                      9.0
297                                     15.0
298                                      0.0

     Human Trafficking - Involuntary Servitude  \
0                                          0.0
1                                          0.0
2                                          0.0
```

```
3                                                 0.0
4                                                 0.0
..                                                ...
294                                               0.0
295                                               0.0
296                                               0.0
297                                               4.0
298                                               1.0

     Sex Offenses (Except Rape, and Prostitution and Commercialized Vice)
0                                                741.0
1                                                281.0
2                                               1472.0
3                                                124.0
4                                               9247.0
..                                                 ...
294                                              575.0
295                                              433.0
296                                               61.0
297                                              971.0
298                                               28.0

[299 rows x 11 columns]
```

In [46]:

```
# This code chunk plots the line chart for the following five states.
#The chart is interactive.
selected_states = ['New York', 'California', 'Texas', 'Georgia', 'Florida']
filtered_data = api_data[api_data['State'].isin(selected_states)]
fig = px.line(filtered_data, x='Year', y='Rape', color='State', markers=True,
              title='Rape Between 2016 and 2021')
fig.show()
```

# First Visualization across two sources

In [55]:

```
# In order to create this visual, I took values from the flat_data table and the api_data table
# Filter flat_data for CALIFORNIA and 2016
filtered_flat_data = flat_data[(flat_data['jurisdiction'] == 'CALIFORNIA') & (flat_data['year'] == 2016)]

# Filter api_data for California and 2016
filtered_api_data = api_data[(api_data['State'] == 'California') & (api_data['Year'] == 2016)]

# Create subplots with shared y-axis
fig = sp.make_subplots(rows=1, cols=2, subplot_titles=['Aggravated Assault', 'Simple Assault'],
                       shared_yaxes=True)

# Add bar chart for flat_data
fig.add_trace(
    go.Bar(x=filtered_flat_data['jurisdiction'], y=filtered_flat_data['agg_assault'], name='Agg Assault'),
    row=1, col=1
)

# Add bar chart for api_data
fig.add_trace(
    go.Bar(x=filtered_api_data['State'], y=filtered_api_data['Simple Assault'], name='Simple Assault'),
    row=1, col=2
)

# Update layout
fig.update_layout(title_text='Combined Bar Charts for Aggravated Assault and Simple Assault in California',
                  showlegend=False)

# Show the plot
fig.show()
```

# Second Visualization across two sources

In [62]:

```
# In order to create this visual, I took values from the flat_data table and api_data table.
# The graph is interactive.
filtered_flat = flat_data[(flat_data['jurisdiction'].isin(['CALIFORNIA', 'TEXAS', 'GEORGIA'])) & (flat_da

# Filter api_data for California, Texas, Florida
filtered_api = api_data[(api_data['State'].isin(['California', 'Texas', 'Georgia']))]

# Create a line graph
fig = go.Figure()

# Add line for flat_data
for jurisdiction in ['CALIFORNIA', 'TEXAS', 'GEORGIA']:
    fig.add_trace(
        go.Scatter(x=filtered_flat[filtered_flat['jurisdiction'] == jurisdiction]['year'],
                   y=filtered_flat[filtered_flat['jurisdiction'] == jurisdiction]['robbery'],
                   mode='lines+markers',
                   name=f'{jurisdiction} - flat_data')
    )

# Add line for api_data
for state in ['California', 'Texas', 'Georgia']:
    fig.add_trace(
        go.Scatter(x=filtered_api[filtered_api['State'] == state]['Year'],
                   y=filtered_api[filtered_api['State'] == state]['Robbery'],
                   mode='lines+markers',
                   name=f'{state} - api_data')
    )

# Update layout
fig.update_layout(title_text='Robbery from 2001 to 2021',
                  xaxis_title='Year',
                  yaxis_title='Robbery Count',
                  legend=dict(x=0, y=1, traceorder='normal'))

# Show the plot
fig.show()
```

# Merging all three tables

In [68]:

```
# I couldn't get this to work.
# Merge flat_data and api_data on 'jurisdiction' and 'State'
merged_data1 = pd.merge(flat_data, api_data, left_on='jurisdiction', right_on='State', how='outer')

# Merge the result with website_data on 'Location'
final_merged_data = pd.merge(merged_data1, website_data, left_on='jurisdiction', right_on='Location', how

# Drop redundant columns if needed
final_merged_data = final_merged_data.drop(['State'], axis=1)

# Display the final merged data
print(final_merged_data)
```

```
---------------------------------------------------------------------------
TypeError                                 Traceback (most recent call last)
Cell In[68], line 5
      2 merged_data1 = pd.merge(flat_data, api_data, left_on='jurisdiction', right_on='State', how='out
er')
      4 # Merge the result with website_data on 'Location'
----> 5 final_merged_data = pd.merge(merged_data1, website_data, left_on='jurisdiction', right_on='Loca
tion', how='outer')
      7 # Drop redundant columns if needed
      8 final_merged_data = final_merged_data.drop(['State'], axis=1)

File ~\anaconda3\lib\site-packages\pandas\core\reshape\merge.py:110, in merge(left, right, how, on, lef
t_on, right_on, left_index, right_index, sort, suffixes, copy, indicator, validate)
     93 @Substitution("\nleft : DataFrame or named Series")
     94 @Appender(_merge_doc, indents=0)
     95 def merge(
   (...)
    108     validate: str | None = None,
    109 ) -> DataFrame:
--> 110     op = _MergeOperation(
    111         left,
    112         right,
    113         how=how,
    114         on=on,
    115         left_on=left_on,
    116         right_on=right_on,
    117         left_index=left_index,
    118         right_index=right_index,
    119         sort=sort,
    120         suffixes=suffixes,
    121         indicator=indicator,
    122         validate=validate,
    123     )
    124     return op.get_result(copy=copy)

File ~\anaconda3\lib\site-packages\pandas\core\reshape\merge.py:645, in _MergeOperation.__init__(self,
left, right, how, on, left_on, right_on, axis, left_index, right_index, sort, suffixes, indicator, vali
date)
    628 def __init__(
    629     self,
    630     left: DataFrame | Series,
   (...)
    642     validate: str | None = None,
```

```
    643 ) -> None:
    644     _left = _validate_operand(left)
--> 645     _right = _validate_operand(right)
    646     self.left = self.orig_left = _left
    647     self.right = self.orig_right = _right

File ~\anaconda3\lib\site-packages\pandas\core\reshape\merge.py:2426, in _validate_operand(obj)
   2424         return obj.to_frame()
   2425 else:
-> 2426     raise TypeError(
   2427         f"Can only merge Series or DataFrame objects, a {type(obj)} was passed"
   2428     )

TypeError: Can only merge Series or DataFrame objects, a <class 'str'> was passed
```

In [ ]:

In [ ]: