IST-652 Scripting for Data Analytics

FINAL PROJECT

Analysis of crime data in Syracuse, from the year 2019 – 2022

Our chosen topic for review involves the examination of crime data available on Open Data Syracuse, a publicly accessible datasets collected by the City of Syracuse, which can be found at

Crime Data 2022 Part 1 Offenses
https://data.syr.gov/datasets/d3c98278e2864a2bbcd00e6e30358856_0/about

Crime Data 2021 Part 1 Offenses:
https://data.syr.gov/datasets/72933036db8a4341870278ae4298e381_0/about

Crime Data 2020 Part 1 Offenses:
https://data.syr.gov/datasets/41cb753474f241e4bbc8977a711fa108_0/about

Crime Data 2019 Part 1 Offenses:
https://data.syr.gov/datasets/4e0ea21c67ff43bdbb38ffecbfba8175_3/about

Our project's objectives include the analysis of various types of crimes occurring within Syracuse. We aim to determine which crimes have seen an increase or decrease in the past years and Which types of crimes are most frequently happening and which are least frequent.

A description of your data exploration

1. Our Syracuse Crime dataset will include all the below from the year 2019 – 2022

2. Date End - Date that the crime was reported. It could have happened earlier. This is in the format of DD-MON-YY (Ex. 01-Jan-22).

3. Time starts and time end - Listed in military time (2400) - Burglaries and larcenies are often a time frame.

4. Address - Where the crime occurred. All addresses are in the 100's because the Syracuse Police Department allows privacy for residents and only lists the block number.

5. Code Defined - Offense names are listed as crime categories group for ease of understanding. There may have been other offenses also, but the one displayed is the highest Unified Crime Reporting (UCR) category.

6. Arrest - Means that there was an arrest, but not necessarily for that crime.

7. Larceny Code - Indicates the type of larceny (Example: From Building or From Motor Vehicle) . It will be crucial in analyzing the crime data.

```
# Import the pandas library for data manipulation

import pandas as pd

# Loading the datasets into individual dataframes from CSV files
crime_data_2019 = pd.read_csv('Crime_Data_2019_(Part_1_Offenses).csv')
crime_data_2020 = pd.read_csv('Crime_Data_2020_(Part_1_Offenses).csv')
crime_data_2021 = pd.read_csv('Crime_Data_2021_(Part_1_Offenses).csv')
crime_data_2022 = pd.read_csv('Crime_Data_2022_(Part_1_Offenses).csv')

# Remove the 'Attempt' column from the 2020 crime dataset
crime_data_2020 = crime_data_2020.drop(columns=['Attempt'])

# Define standard column names for uniformity across datasets
standard_columns = ['CRIME DATE', 'TIME START', 'TIME END', 'ADDRESS', 'CRIME TYPE', 'ARREST MADE', 'LARCENY TYPE', 'ID']

# Rename columns in each dataset to standardize them
crime_data_2019.columns = standard_columns
crime_data_2020.columns = standard_columns
crime_data_2021.columns = standard_columns
crime_data_2022.columns = standard_columns

# Merge all dataframes into one combined DataFrame
all_crime_data = pd.concat([crime_data_2019, crime_data_2020, crime_data_2021, crime_data_2022], axis=0, ignore_index=True)

# Replace NaN values in 'ARREST MADE' column with 'No'
all_crime_data['ARREST MADE'] = all_crime_data['ARREST MADE'].fillna('No')

# Convert 'CRIME DATE' to datetime format and then extract just the date part
all_crime_data['CRIME DATE'] = pd.to_datetime(all_crime_data['CRIME DATE']).dt.date
```

```
# Inspecting the first few rows of the dataframe
print(all_crime_data.head())
```

```
   CRIME DATE TIME START  TIME END                ADDRESS  \
0  2019-01-01   01:20:00  01:24:00  100 BALLANTYNE RD
1  2019-01-01   18:19:00  07:00:00       100 DAWES AV
2  2019-01-01   01:15:00  02:45:00      100 FENTON ST
3  2019-01-01   19:00:00  19:00:00     100 MILDRED AV
4  2019-01-01   07:25:00  07:25:00    1400 MIDLAND AV

           CRIME TYPE ARREST MADE   LARCENY TYPE  ID
0  AGGRAVATED ASSAULT         Yes            NaN   1
1             LARCENY          No      All Other   2
2           MV THEFT          No  Motor Vehicle   3
3  AGGRAVATED ASSAULT         Yes            NaN   4
4  AGGRAVATED ASSAULT          No            NaN   5
```

```
# Check data types validations of all columns
print(all_crime_data.dtypes)

# Convert data types if necessary
# For example, if 'CRIME DATE' is not in datetime format, convert it:
all_crime_data['CRIME DATE'] = pd.to_datetime(all_crime_data['CRIME DATE'], errors='coerce')
```

```
CRIME DATE      object
TIME START      object
TIME END        object
ADDRESS         object
CRIME TYPE      object
ARREST MADE     object
LARCENY TYPE    object
ID               int64
dtype: object
```

Data acquisition and analysis -

Our project's core goals involve analyzing a range of crimes taking place in Syracuse. We aim to identify the most frequently occurring types of crimes, as well as those that are less commonly reported. Additionally, we will assess which categories of crime have experienced an increase or decrease over the years.

For actual data processing, we'll be simply downloading the data CSV from the source website.

Then, once read into Jupyter, we anticipate using the following packages:

- Pandas - which allows us to analyze data and make conclusions based on statistical theories. Pandas can clean messy data sets and make them readable and relevant.
- We will require CSV format for the data since it is expected to be available for download in that specific file format. This is essential for importing the data into our Jupiter Notebook for analysis, and subsequently, we will use this format to export the data into new CSV files that can be easily shared.

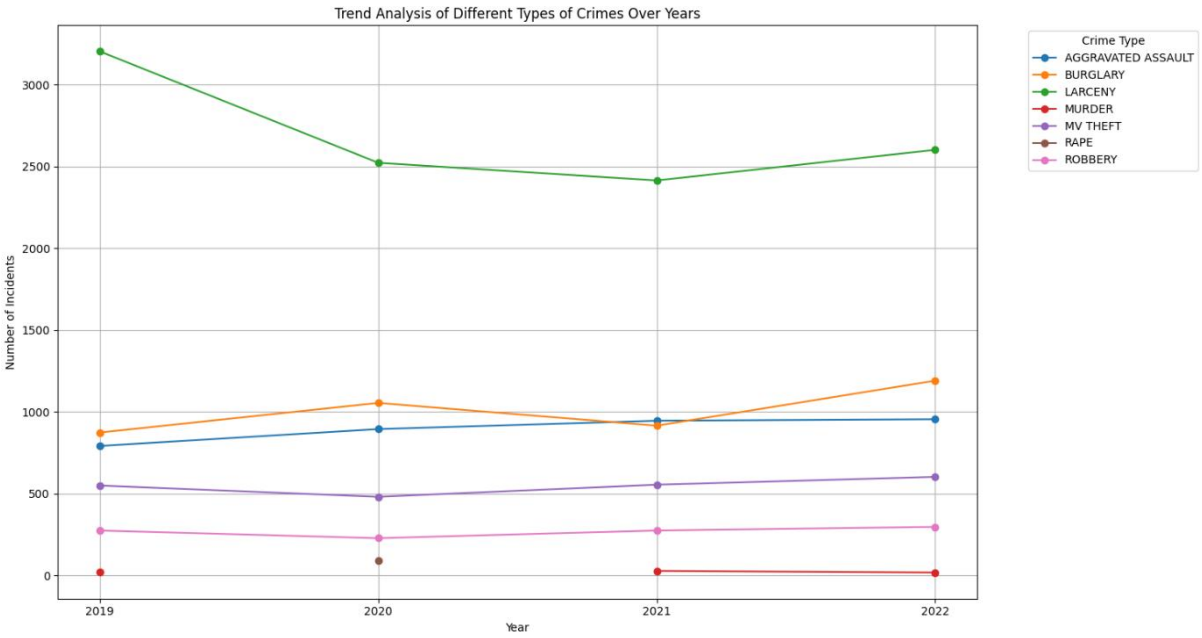
- Matplotlib - creating static, animated, and interactive visualizations and plots. It provides a wide range of functions and tools for generating various types of graphs, charts, and figures.

- NumPy - used to generate arrays from numerical data, which can subsequently be utilized to perform statistical analyses on the data, enabling the extraction of valuable insights and information.

- Folium – This library makes it easy to visualize data that's been manipulated in Python on an interactive leaflet map. It enables both the binding of data to a map for choropleth visualizations as well as passing rich vector/raster/HTML visualizations as markers on the map.

Insights And Analysis -

- **Trend Analysis of Crime Over Years at Syracuse from 2019 to 2022 –**



CRIME TYPE	AGGRAVATED ASSAULT	BURGLARY	LARCENY	MURDER	MV THEFT	RAPE	\
YEAR							
2019	791	873	3203	19	549	0	
2020	894	1054	2522	0	480	89	
2021	945	915	2414	27	554	0	
2022	954	1190	2601	17	602	0	

CRIME TYPE	ROBBERY
YEAR	
2019	274
2020	227
2021	274
2022	296

The Syracuse crime analysis from 2019 to 2022 reveals rising aggravated assault and motor vehicle theft, with burglary also up since 2020. Larceny generally decreased, despite a high in 2019. Murders peaked in 2021, while inconsistent rape data calls for cautious interpretation.

Robberies dipped in 2020 but rebounded, possibly linked to pandemic-related social changes. These trends, though insightful, are influenced by various factors and reporting practices, and thus should guide but not solely dictate policy and public safety strategies.

- **Method**: Group the data by 'CRIME TYPE' and year, then count the number of incidents for each type per year. Use line graphs or bar charts to visualize trends over time.
- **Question**: How has the frequency of different types of crimes (like theft, burglary, assault) changed from 2019 to 2022 in Syracuse?
- **Insight**: This analysis can highlight which crimes are becoming common, indicating shifts in the crime landscape of Syracuse.

- **Time Analysis of Crime Incidents –**

Crime Counts by Hour of the Day:

	Hour	Count
0	0	1250
1	1	739
2	2	724
3	3	626
4	4	530
5	5	407
6	6	448
16	16	1114
17	17	1178
18	18	1201
19	19	1030
20	20	992
21	21	1029
22	22	1149
23	23	1016

	Day of the Week	Count
0	Tuesday	3212
1	Monday	3210
2	Saturday	3109
3	Sunday	3091
4	Thursday	3072
5	Friday	3064
6	Wednesday	3006

Crime Counts by Month of the Year:

	Month	Count
0	August	2159
1	July	2082
2	June	2016
3	September	2012
4	October	1941
5	May	1867
6	November	1731

The crime data for Syracuse shows midnight crime spikes, tapering off towards 5 AM, then climbing again, with a noon uptick. Crimes are even across the week, with a small rise on Mondays and Tuesdays. Seasonally, August sees the most crime, with a drop in colder months,

bottoming out in February. These insights suggest night and summer require more police focus, with consistent weeklong patrols despite minor early week increases.

- **Method**: Extract the hour from 'TIME START' and 'TIME END' and day of the week from 'CRIME DATE'. Analyze the frequency of crimes by time and day.
- **Question**: What times of day and days of the week do most crimes occur in Syracuse? Are there patterns in timing for different types of crimes?

- **Insight**: Identifying peak crime times can aid in resource allocation for law enforcement and community awareness programs.

- **Correlation between Arrests and Types of Crimes –**

```
Correlation between Arrests and Types of Crimes:
            CRIME TYPE  TOTAL CRIMES  ARRESTS MADE  ARREST RATE
3               MURDER            63            44     0.698413
0  AGGRAVATED ASSAULT          3584          1987     0.554408
6              ROBBERY          1071           356     0.332400
5                 RAPE            89            17     0.191011
1             BURGLARY          4032           764     0.189484
4             MV THEFT          2185           389     0.178032
2              LARCENY         10740          1804     0.167970

Correlation coefficient between Total Crimes and Arrests Made:
-0.1407309127275592
```

The crime and arrest data correlation shows murders have the highest arrest rate at nearly 70%, while aggravated assault follows with over 55%. Lesser crimes like larceny, burglary, and vehicle theft have lower arrest rates, with larceny at nearly 17%. There's a slight negative correlation of -0.1407 between crime rates and arrests, hinting that higher crime rates may lead to a marginally lower arrest ratio, possibly due to limited resources or the complexities of the crimes.

- **Method**: Analyze the 'ARREST MADE' data in relation to 'CRIME TYPE'. Calculate the arrest rate for each type of crime.

- **Question**: Is there a correlation between types of crimes committed and the likelihood of an arrest being made?

- **Insight**: Understanding which crimes more often result in arrests can give insights into the effectiveness of law enforcement strategies for different crime types.

CONCLUSION –

- The crime report for Syracuse between 2019 and 2022 shows that some crimes went up and others went down. More assaults and car thefts happened, but fewer burglaries occurred after 2020. There were more murders in 2021, but the number of rapes has changed a lot and is hard to understand.

- Robberies went down in 2020 but then increased again, possibly because of the pandemic.

- Most crimes happen around midnight and drop off after 5 AM, but start to increase again in the daytime, especially at noon. Mondays and Tuesdays have slightly more crimes than other days. More crimes happen in August, and the number goes down in the colder months, with the fewest in February. This means police might need to work harder at night and in the summer, but they should still be around all week.

- When looking at arrests, murders have the most arrests, and assaults are next. But for theft, burglary, and car theft, not as many people get arrested, especially for theft.

- In short, Syracuse's crime numbers show patterns that can help make better safety plans. More police might be needed at night and in the summer, and they should look at why some crimes have lower arrest rates.

Below are the steps we perform to complete the project, as each step required exploring, discussing and generating results, we have mentioned names of the team members involved in each of the steps

1. **Setting up Insite -**

Defining research objective & questions

- Mahima Jayanth
- Siddarth Asati
- Kenyang Lual Mirag

2. Data cleaning & understanding

- Mahima Jayanth
- Siddarth Asati
- Kenyang Lual Mirag

3 Getting the Insights , Visualization , Coding –

- Siddarth Asati
- Kenyang Lual Mirag

4 Project Report and PPT –

- Mahima Jayanth