PROJECT-02 REPORT

Professor Samantha Luckhardt

DATA 601 Introduction to Data Science

20th December,2023

Team members:

Hemanth Gorapalli

Sai Chaitanya Varma Indukuri

Revanth Kumar Chekuri

Shirisha Biyyala

Ashwini Koppula

Veerendra Narapareddy

Shanmukha Surya Sriteja Doddipatla

Premanth Alahari

Aim:

In the subsequent phases, the project ventures into the intricate interplay between neighborhood dynamics and crime rates in Baltimore. The overarching goal is to hypothesize connections among factors such as vacant building notices, rehabilitation initiatives, and gun offender registries. The aim is twofold: to discern whether these elements exert positive or negative influences on crime rates and to unravel unexpected insights that could reshape our comprehension of urban dynamics and their profound impact on community safety.

Tools Required

Online

- Google Collab

Offline

- Python Software
- Anaconda Navigator
  : Jupyter Notebook

Procedure

Step1: First load the dataset by using python libraries like python numpy, pandas, matplotlib and seaborn and load the data set by downloading your data and collab file in same location and mount it and run the code as shown below it will load the data.

```
Code:

import pandas as pd

import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# Load the dataset
file_path = "/content/drive/MyDrive/Colab Notebooks/Part_1_Crime_Data.csv"
```

```
df = pd.read_csv(file_path, parse_dates=['CrimeDateTime'])
```

Step2:

Filter the relevant which require to calculate the neighborhood crime rates and filter the data from 2017-2022 and extract the year from DateCrimeTime Column and proceed with the remaining steps

```
# Filter relevant columns
columns_of_interest = ['RowID', 'CrimeDateTime', 'CrimeCode', 'Description', 'Neighborhood']
df = df[columns_of_interest]

# Filter data for the years 2017 to 2022
df = df[(df['CrimeDateTime'].dt.year >= 2017) & (df['CrimeDateTime'].dt.year <= 2022)]

# Extract year from CrimeDateTime
df['Year'] = df['CrimeDateTime'].dt.year
```

```
# Group data by year and neighborhood, calculate counts
crime_counts = df.groupby(['Year', 'Neighborhood']).size().reset_index(name='Total_Incidents')

# Pivot the table to have years as columns
crime_counts_pivot = crime_counts.pivot(index='Neighborhood', columns='Year', values='Total_Incidents').fillna(0)
```

```
Step3:
Calculate the difference in counts from 2017 to 2022 and calculate the percentage change as shown below:
# Calculate the difference in counts from 2017 to 2022

print("Table showing the difference of counts from 2017 to 2022:")
crime_counts_pivot['Difference'] = crime_counts_pivot[2022] - crime_counts_pivot[2017]
print(crime_counts_pivot['Difference'])
```

```
# Calculate the percentage change
print("\nTable showing the percentage change from 2017 to 2022:")
crime_counts_pivot['PercentageChange'] = (crime_counts_pivot['Difference'] / crime_counts_pivot[2017]) * 100
```

```
print(crime_counts_pivot['PercentageChange'])
```

Step3:

```
Sort the neighbourhoods based on difference
# Sort neighborhoods based on the difference

sorted_neighborhoods = crime_counts_pivot.sort_values(by='Difference', ascending=False)
print(sorted_neighborhoods)
```

```
Select the top5 and bottom 5 neighbourhoods with highest increase and decrease

# Select top 5 and bottom 5 neighborhoods

top_5_neighborhoods = sorted_neighborhoods.head(5)
bottom_5_neighborhoods = sorted_neighborhoods.tail(5)

# Display summary tables
print("Top 5 Neighborhoods with Highest Increase:")
print(top_5_neighborhoods[['Difference', 'PercentageChange']])

print("\nTop 5 Neighborhoods with Highest Decrease:")
print(bottom_5_neighborhoods[['Difference', 'PercentageChange']])
```

Output:

```
Top 5 Neighborhoods with Highest Increase:
Year                       Difference  PercentageChange
Neighborhood
FRANKFORD                        97.0         80.165289
DOWNTOWN                         64.0         40.506329
UPTON                            50.0         40.322581
EAST BALTIMORE MIDWAY            34.0         53.968254
ELLWOOD PARK/MONUMENT            31.0         53.448276

Top 5 Neighborhoods with Highest Decrease:
Year                     Difference  PercentageChange
Neighborhood
PERKINS HOMES                 -27.0        -93.103448
BROADWAY EAST                 -28.0        -24.347826
SANDTOWN-WINCHESTER           -39.0        -18.840580
BALTIMORE HIGHLANDS           -41.0        -50.617284
BROOKLYN                      -49.0        -21.681416
```

Step4:

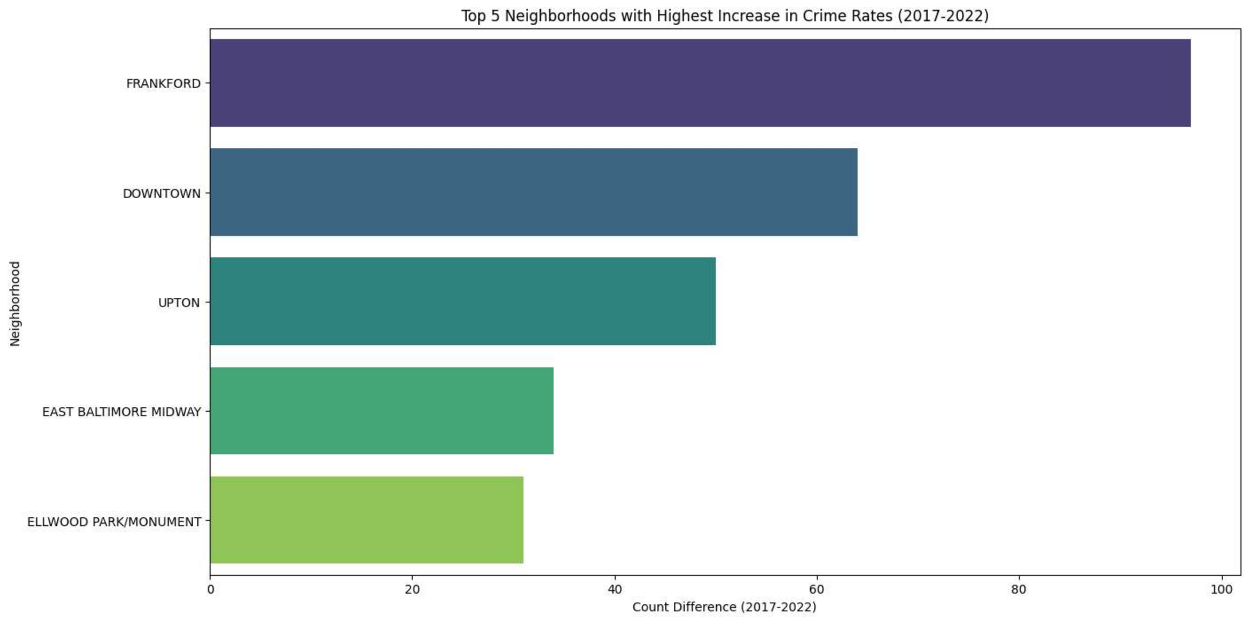
Creating and representing the visualizations which is important in the code I have implemented other visualizations too please check it.

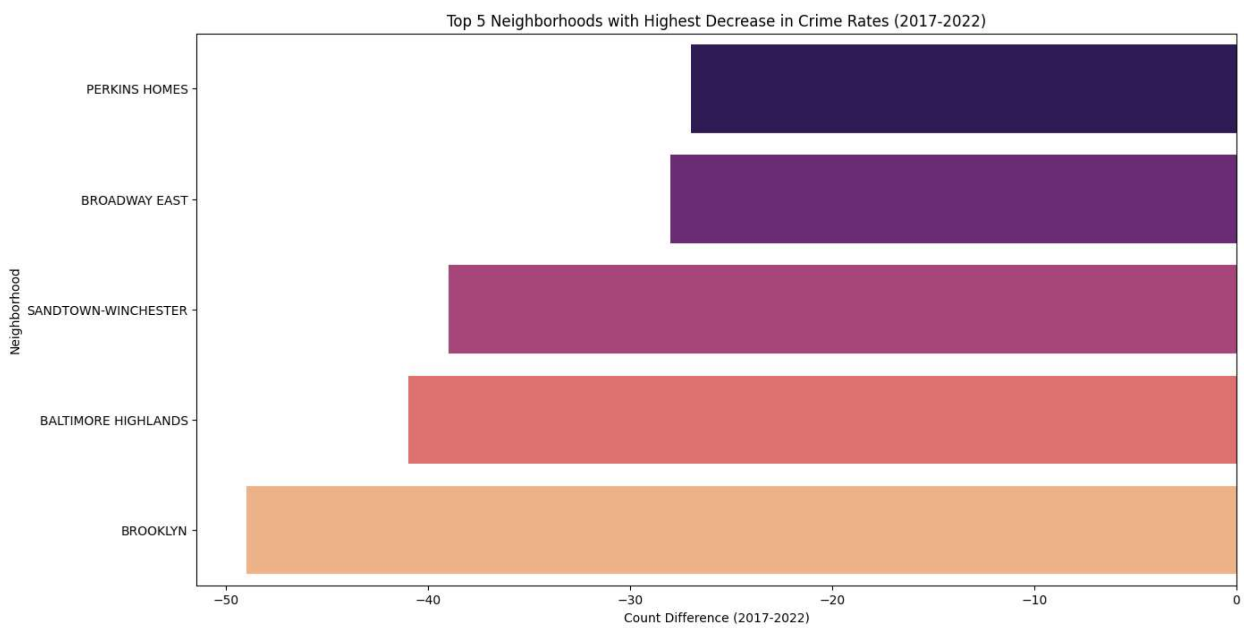
Visualization:

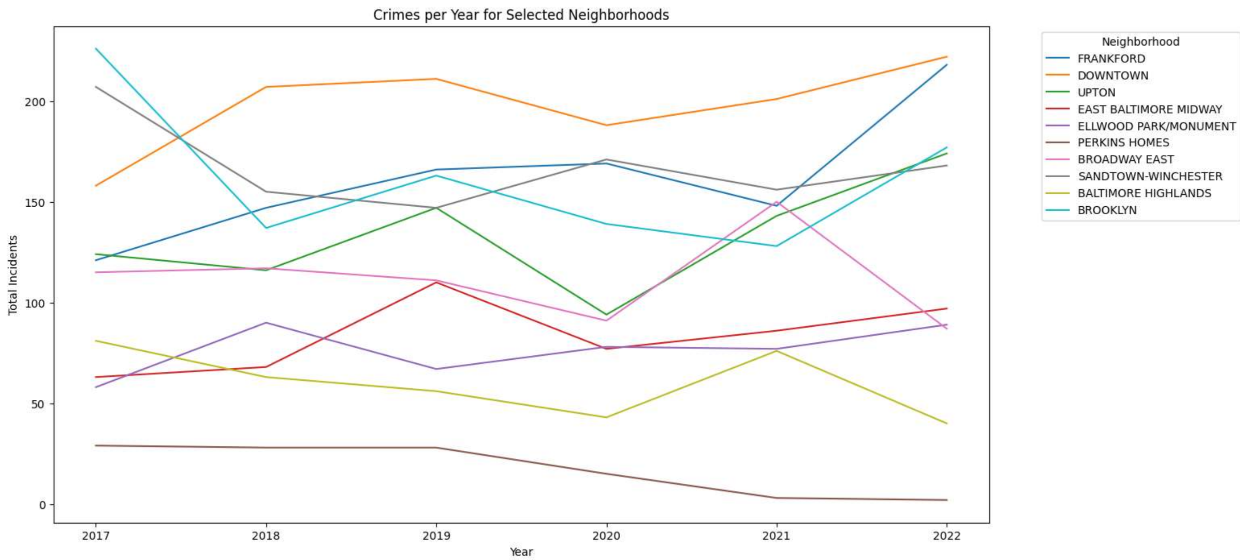
Employed the Matplotlib library to create visualizations for a better understanding of crime trends.
Utilized bar charts to illustrate the crime counts for specific neighborhoods and line graphs to depict changes over the years.

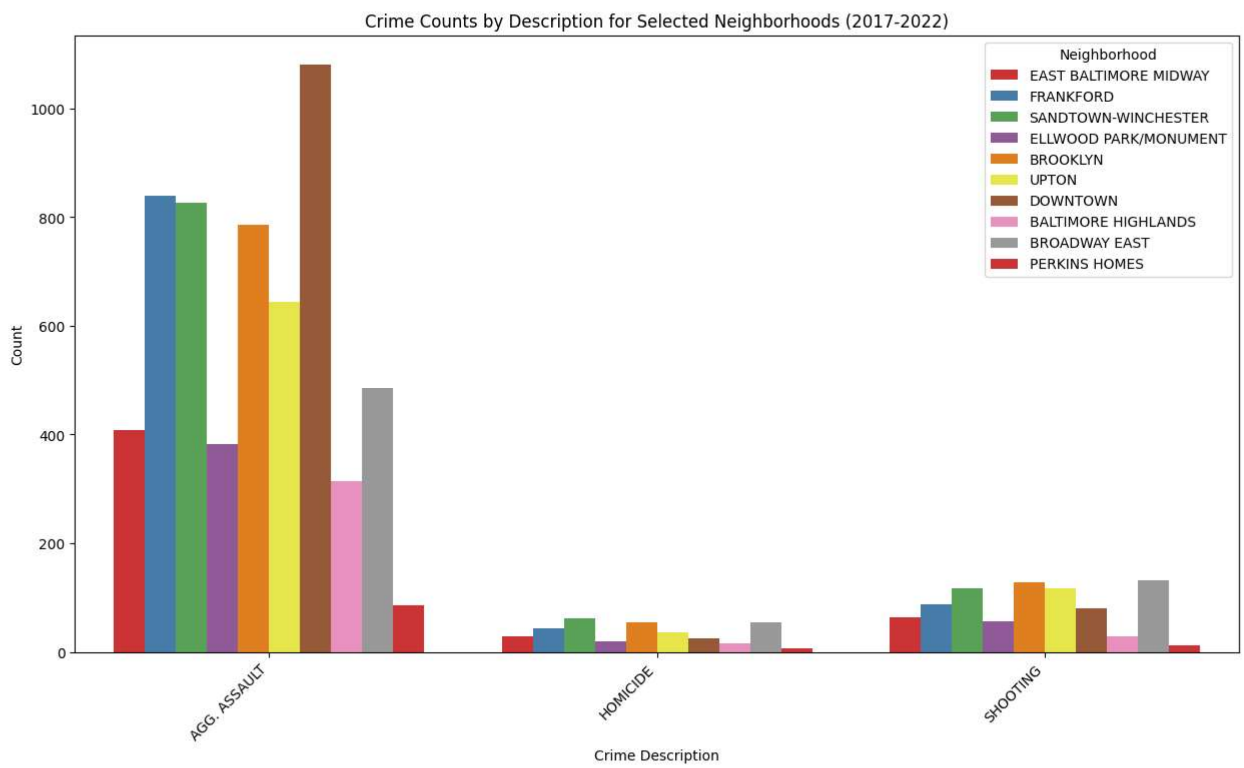
Statistical Analysis:

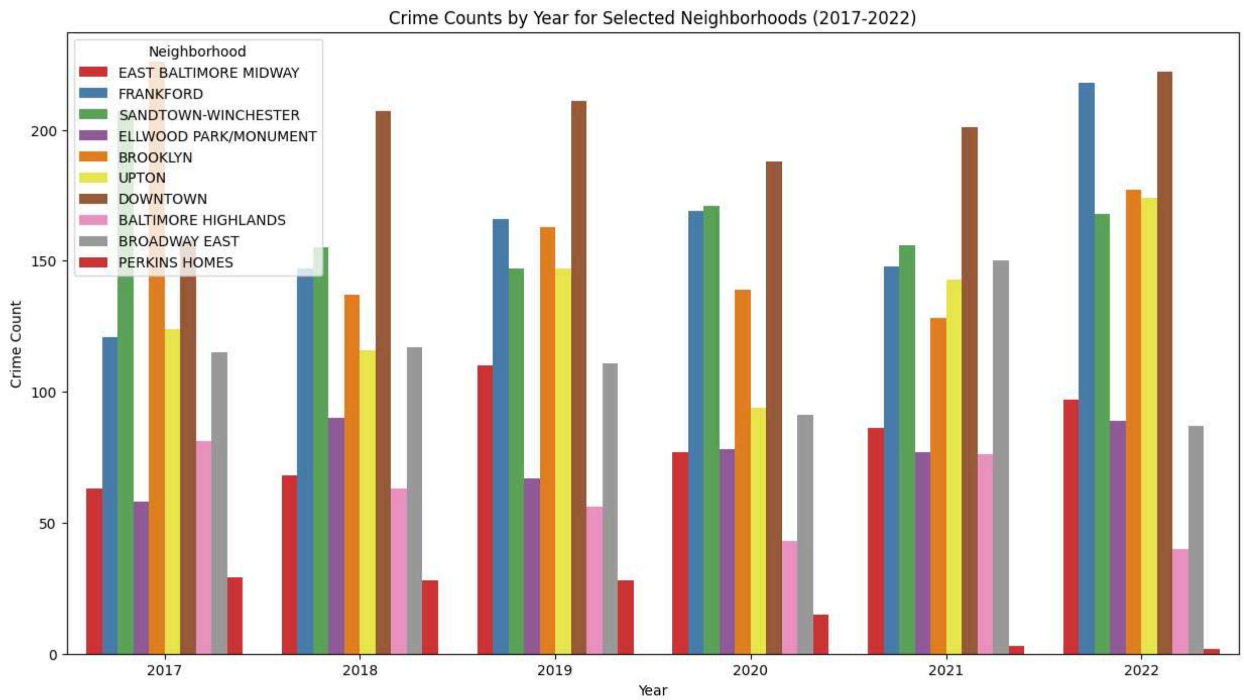
Calculated crime rate changes between consecutive years and percentage changes for selected neighborhoods.
Evaluated the overall change in crime rates across all neighborhoods.

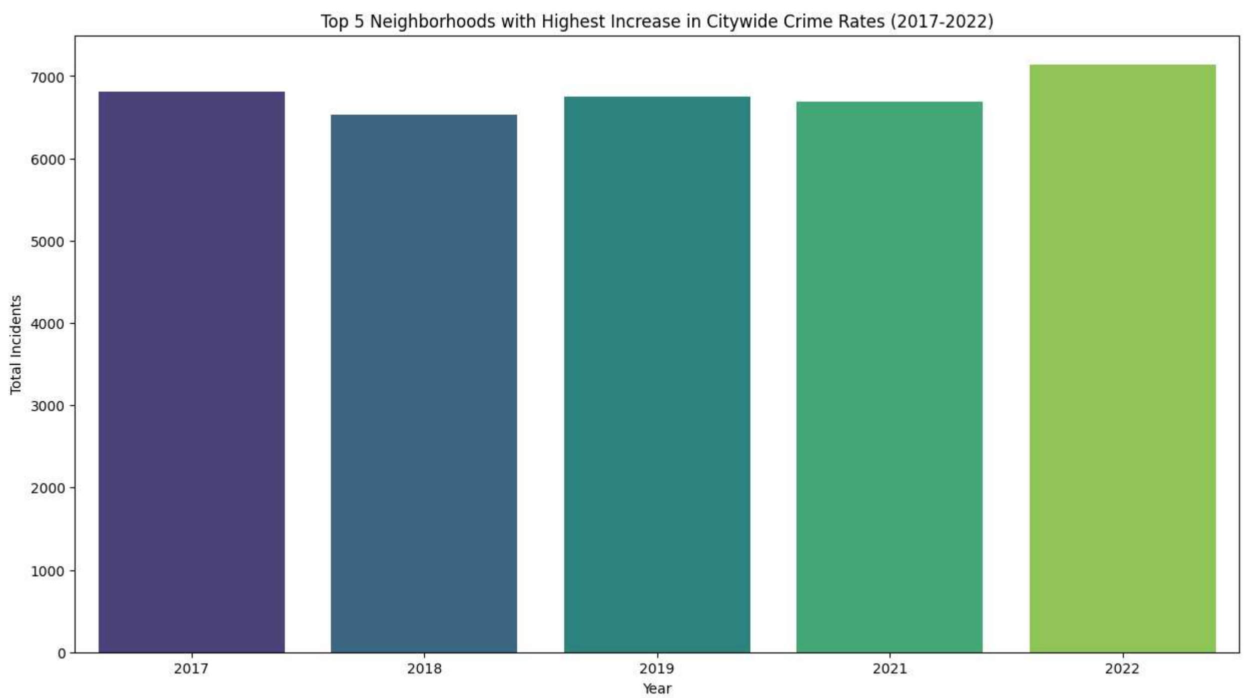

Top 5 Neighborhoods with Highest Increase in Crime Rates (2017-2022)
FRANKFORD
DOWNTOWN
UPTON
EAST BALTIMORE MIDWAY
ELLWOOD PARK/MONUMENT
Neighborhood
0
20
40
60
80
100
Count Difference (2017-2022)

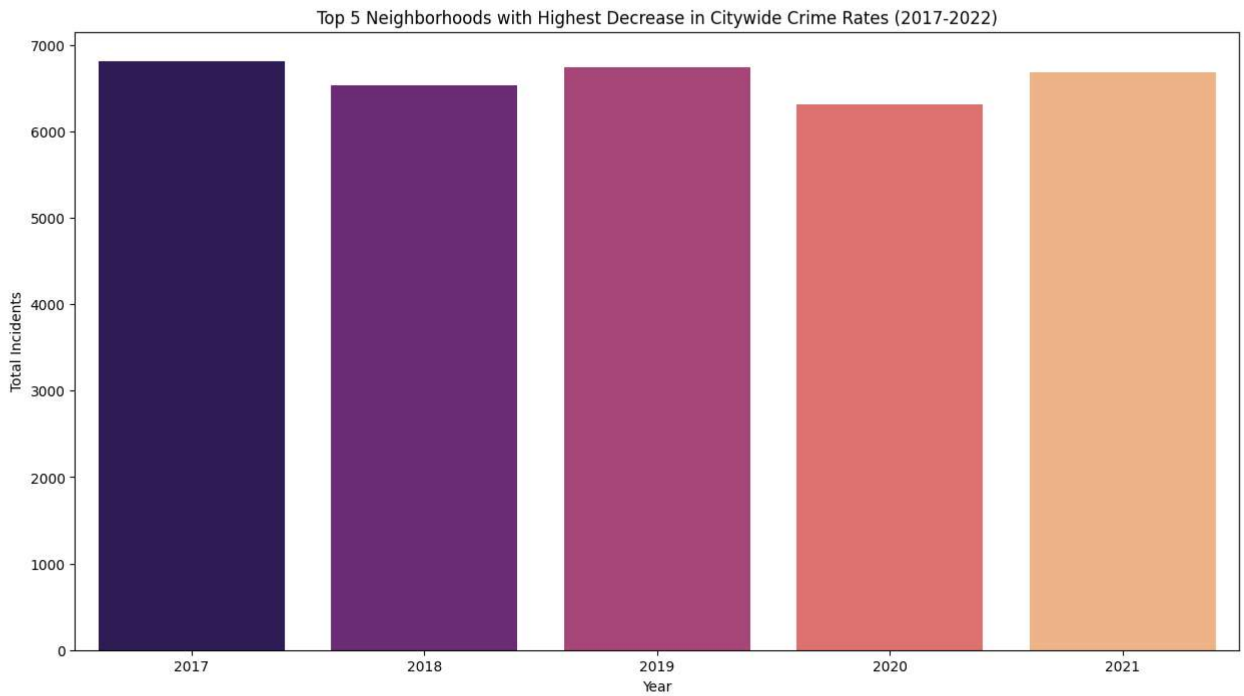

Top 5 Neighborhoods with Highest Decrease in Crime Rates (2017-2022)
PERKINS HOMES
BROADWAY EAST
SANDTOWN-WINCHESTER
BALTIMORE HIGHLANDS
BROOKLYN
Neighborhood
−50
−40
−30
−20
−10
0
Count Difference (2017-2022)

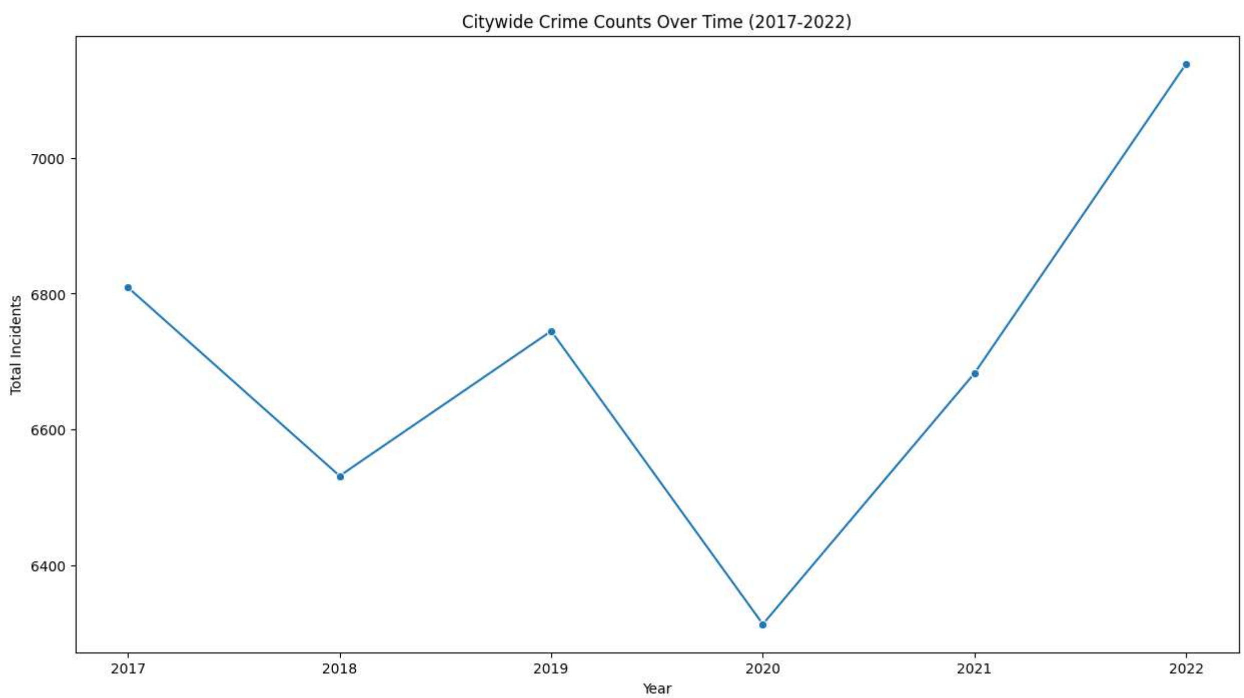

Crimes per Year for Selected Neighborhoods
Total Incidents
Year
0
50
100
150
200
2017
2018
2019
2020
2021
2022
Neighborhood
FRANKFORD
DOWNTOWN
UPTON
EAST BALTIMORE MIDWAY
ELLWOOD PARK/MONUMENT
PERKINS HOMES
BROADWAY EAST
SANDTOWN-WINCHESTER
BALTIMORE HIGHLANDS
BROOKLYN


Crime Counts by Description for Selected Neighborhoods (2017-2022)
Count
0
200
400
600
800
1000
AGG. ASSAULT
HOMICIDE
SHOOTING
Crime Description
Neighborhood
EAST BALTIMORE MIDWAY
FRANKFORD
SANDTOWN-WINCHESTER
ELLWOOD PARK/MONUMENT
BROOKLYN
UPTON
DOWNTOWN
BALTIMORE HIGHLANDS
BROADWAY EAST
PERKINS HOMES


Crime Counts by Year for Selected Neighborhoods (2017-2022)
Neighborhood
EAST BALTIMORE MIDWAY
FRANKFORD
SANDTOWN-WINCHESTER
ELLWOOD PARK/MONUMENT
BROOKLYN
UPTON
DOWNTOWN
BALTIMORE HIGHLANDS
BROADWAY EAST
PERKINS HOMES
Crime Count
0
50
100
150
200
2017
2018
2019
2020
2021
2022
Year


Top 5 Neighborhoods with Highest Increase in Citywide Crime Rates (2017-2022)
Total Incidents
0
1000
2000
3000
4000
5000
6000
7000
2017
2018
2019
2021
2022
Year


Top 5 Neighborhoods with Highest Decrease in Citywide Crime Rates (2017-2022)
Total Incidents
7000
6000
5000
4000
3000
2000
1000
0
2017
2018
2019
2020
2021
Year


Citywide Crime Counts Over Time (2017-2022)
Total Incidents
7000
6800
6600
6400
2017
2018
2019
2020
2021
2022
Year

Hypothesis Testing:
A statistical technique called hypothesis testing is essential for deriving conclusions about population parameters from the examination of sample data. This procedure involves creating hypotheses, gathering and analyzing data, and drawing inferences about the population from sample results that have been seen. The scientific method relies heavily on hypothesis testing, which is also essential to empirical research in a variety of fields.

The Most Important Parts of a Hypothesis Test:

1.Null Hypothesis (H0): The null hypothesis states that there is no influence, difference, or modification to the population parameter. It explains the situation as it stands and implies that sample fluctuations or random variation are the cause of observed differences. It usually states something equivalent, such population mean equality.

2.The alternative hypothesis (H1 or Ha) posits the existence of an effect, distinction, or alteration in opposition to the null hypothesis. It stands for the hypothesis that researchers seek to substantiate with data and is represented by the symbols H1 or Ha. Comparisons can be made bilaterally (not equal to) or unilaterally (more or less than).

3.The statistical term "alpha" (α) indicates the significance level. It shows the likelihood of mistakenly rejecting the Null Hypothesis when it is true. When choosing an alpha to determine the threshold for statistical significance, common values are 0.05, 0.01, or 0.10.

4. The type of data being analyzed and the hypothesis under investigation influence the test statistic selection. F-tests, chi-square tests, z-tests, and t-tests are examples of common test statistics. A particular probability distribution, such as the t-distribution or normal distribution, is followed by the chosen test statistic.

5. P-value: If the null hypothesis is true, the p-value represents the probability of receiving results that are as extreme as the observed outcomes. When determining whether to reject or keep the null hypothesis, a low p-value indicates strong evidence against it.

In the next steps, it delves into unraveling the intricate dance between neighborhood dynamics and crime rates in Baltimore. It hypothesizes connections between factors like vacant building notices, rehabilitation efforts, and gun offender registries, seeking to discern whether these elements sway crime rates positively or negatively. As it embarks on this investigative journey, the aim is to not only decipher correlations but also uncover unexpected insights that might reshape our understanding of urban dynamics and their impact on community safety.

Step 5:

In formulating the hypothesis, considering variables such as vacant building notices, vacant building rehabs, and the Gun Offender Registry. The objective is to investigate whether these factors exert any influence on the fluctuations in crime rates within the identified neighborhoods and citywide. This comprehensive analysis aims to discern patterns and correlations, shedding light on the potential impact of these specific elements on the dynamic landscape of crime in Baltimore.

This analytical process entails a systematic approach, involving the meticulous filtering of relevant columns, grouping the data by year and neighborhood, and subsequently calculating the differences in counts from 2017 to 2022. Moreover, computing the percentage change for each identified factor—vacant building notices, vacant building rehabs, and the Gun Offender Registry. Through these rigorous steps, the aim is to unravel meaningful insights into the potential correlation between these factors and the changes in crime rates over the specified timeframe.

Step 6:

The subsequent phase of this investigative journey involves the systematic sorting of neighborhoods based on the calculated differences in counts. Following this sorting process, the focus will shift to the meticulous selection of the top five and bottom five neighborhoods. This strategic approach aims to spotlight specific geographic areas that exhibit notable variations in factors such as vacant building notices, rehabs, and gun offender data. By honing in on these extremes, the group seeks to uncover nuanced patterns and potential correlations within Baltimore's urban landscape.

Step 7:

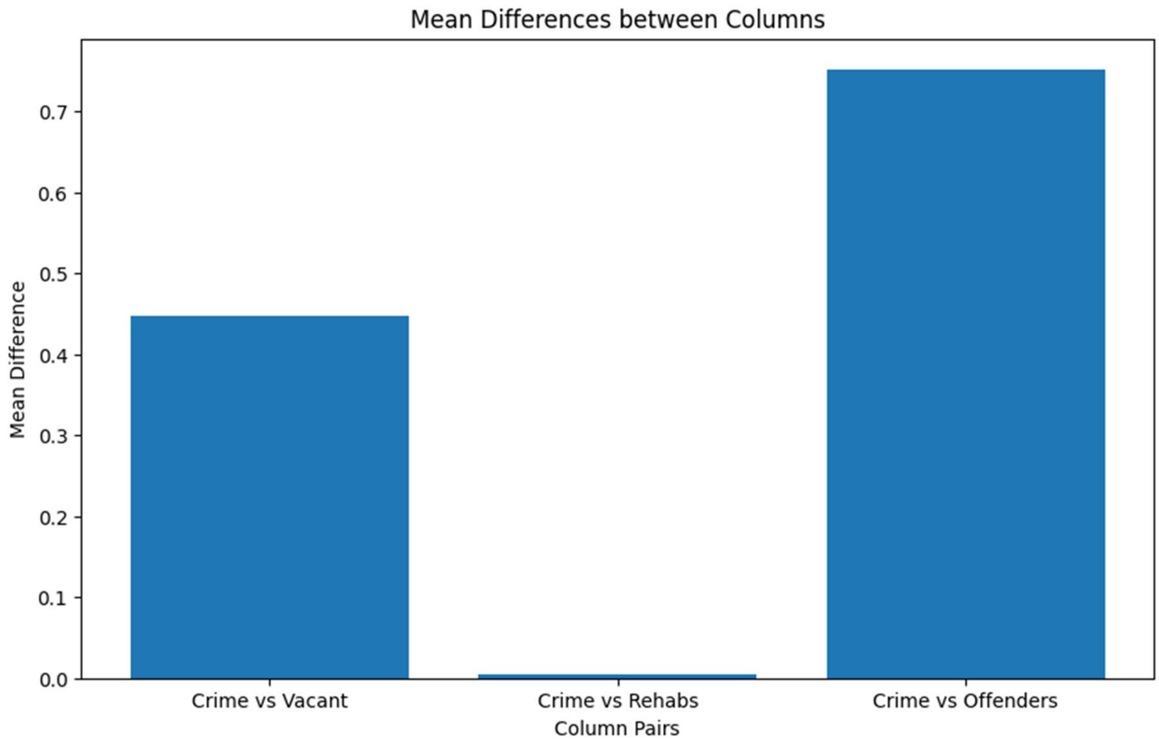
In the subsequent phase of our exploration, the focus shifts to visualization, starting with the presentation of data related to vacant building notices. This visualization endeavor aims to provide a clear and insightful depiction of how vacant building notices vary across neighborhoods and over the specified timeframe. By employing visual elements, the group intends to convey nuanced patterns, trends, and potential correlations within Baltimore's urban fabric. This visual representation serves as a crucial tool in fostering a deeper understanding of the dynamics between vacant building notices and changes in crime rates.

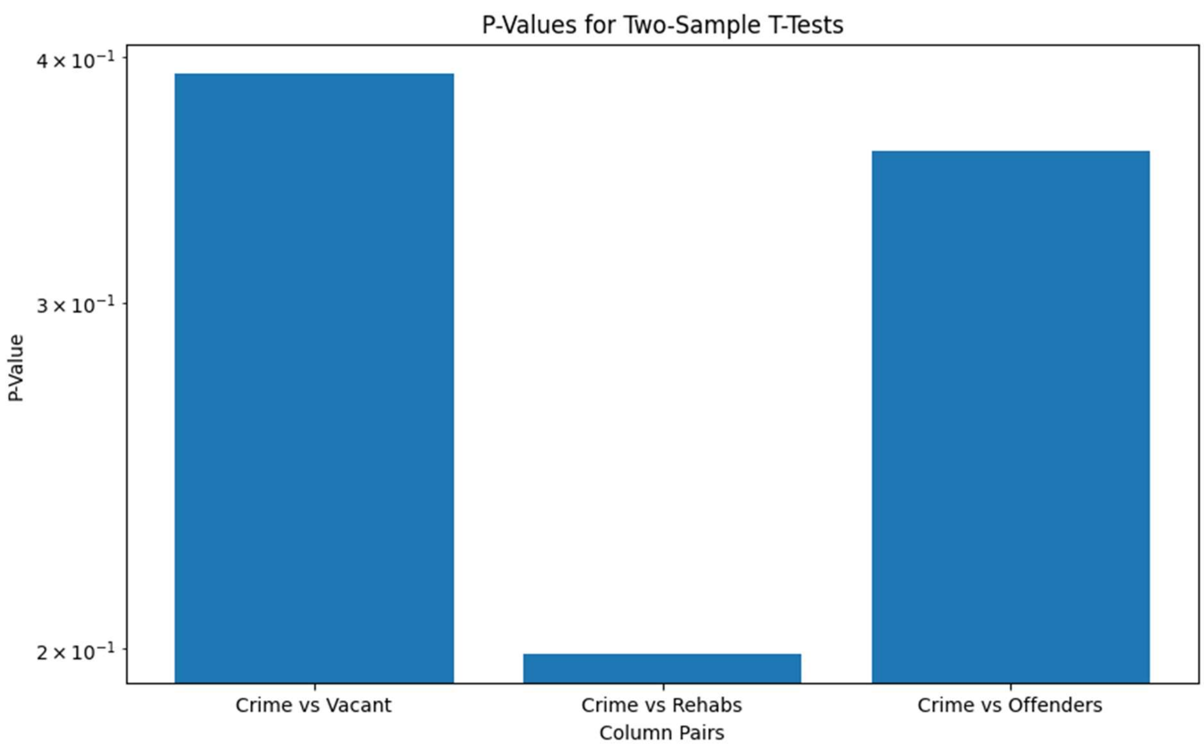
NoticeYear Neighborhood	2017	2018	2019	2020	2021	2022	difference_of_vancant
EAST BALTIMORE MIDWAY	16.0	12.0	16.0	24.0	22.0	55.0	39.0
CARROLLTON RIDGE	36.0	32.0	43.0	50.0	74.0	74.0	38.0
BROADWAY EAST	29.0	30.0	48.0	37.0	61.0	64.0	35.0
OLIVER	22.0	16.0	24.0	15.0	24.0	46.0	24.0
BROOKLYN	6.0	13.0	15.0	18.0	23.0	29.0	23.0
...	...	...	...	...	...	...	...
FRANKLINTOWN ROAD	10.0	10.0	9.0	7.0	9.0	8.0	-2.0
WINCHESTER	7.0	7.0	5.0	4.0	4.0	4.0	-3.0
BIDDLE STREET	12.0	8.0	8.0	4.0	6.0	8.0	-4.0
SHIPLEY HILL	20.0	4.0	25.0	16.0	17.0	12.0	-8.0
SANDTOWN-WINCHESTER	48.0	42.0	48.0	28.0	53.0	18.0	-30.0

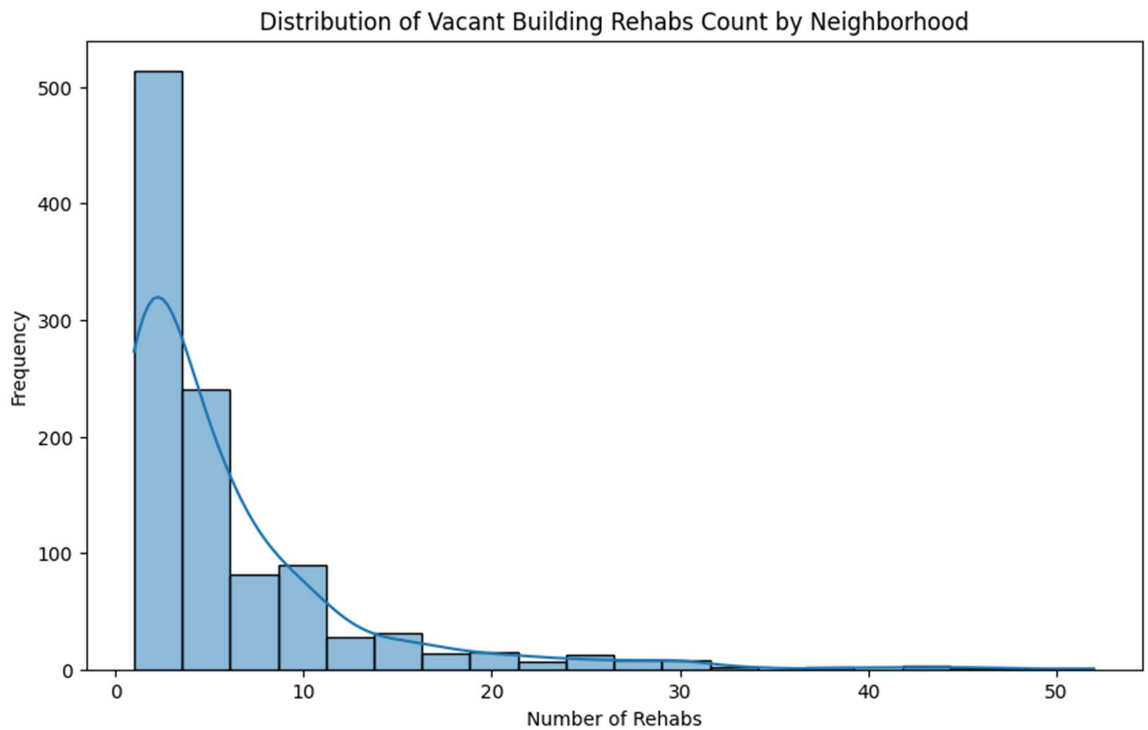
The coming visualizations show various representations of the hypothesis made.

Year Neighborhood	2017	2018	2019	2020	2021	2022	Difference	PercentageChange
FRANKFORD	121.0	147.0	166.0	169.0	148.0	218.0	97.0	80.165289
DOWNTOWN	158.0	207.0	211.0	188.0	201.0	222.0	64.0	40.506329
UPTON	124.0	116.0	147.0	94.0	143.0	174.0	50.0	40.322581
EAST BALTIMORE MIDWAY	63.0	68.0	110.0	77.0	86.0	97.0	34.0	53.968254
ELLWOOD PARK/MONUMENT	58.0	90.0	67.0	78.0	77.0	89.0	31.0	53.448276
...	...	...	...	...	...	...	...	...
PERKINS HOMES	29.0	28.0	28.0	15.0	3.0	2.0	-27.0	-93.103448
BROADWAY EAST	115.0	117.0	111.0	91.0	150.0	87.0	-28.0	-24.347826
SANDTOWN-WINCHESTER	207.0	155.0	147.0	171.0	156.0	168.0	-39.0	-18.840580
BALTIMORE HIGHLANDS	81.0	63.0	56.0	43.0	76.0	40.0	-41.0	-50.617284
BROOKLYN	226.0	137.0	163.0	139.0	128.0	177.0	-49.0	-21.681416

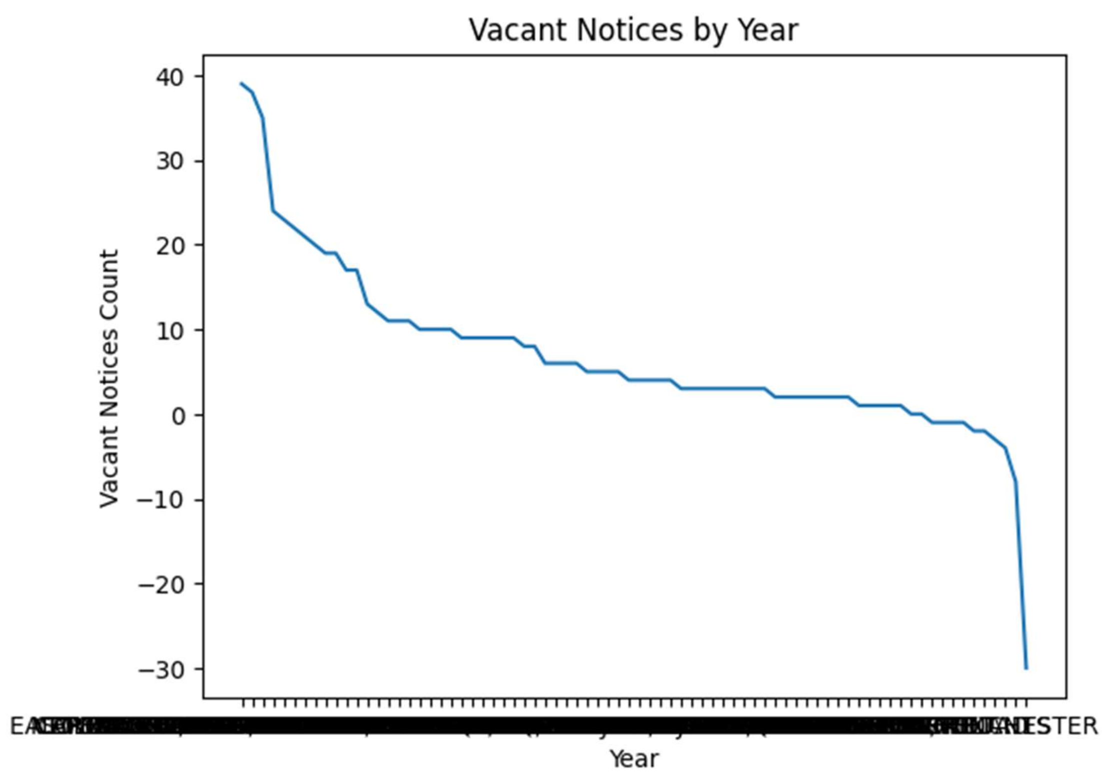
The visualizations which were made was taken based on the datasets used and various testings performed.


Mean Differences between Columns
Mean Difference
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
Crime vs Vacant
Crime vs Rehabs
Crime vs Offenders
Column Pairs

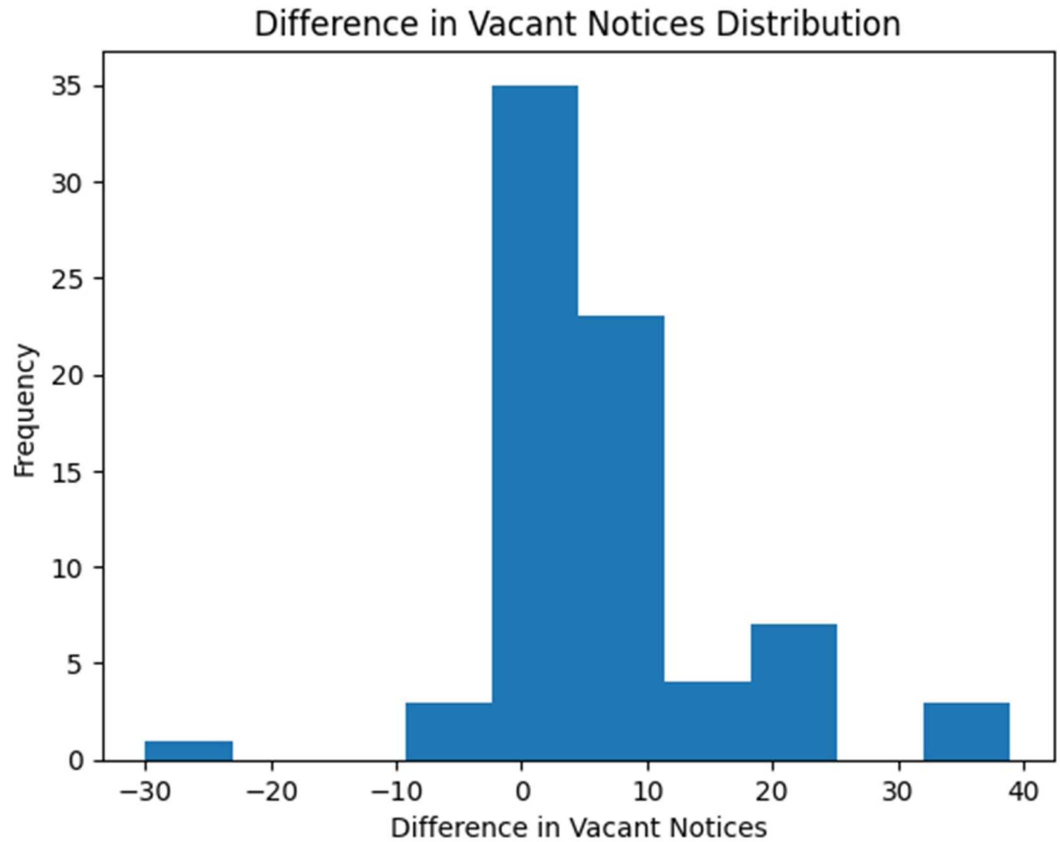

P-Values for Two-Sample T-Tests
P-Value
4×10^{-1}
3×10^{-1}
2×10^{-1}
Crime vs Vacant
Crime vs Rehabs
Crime vs Offenders
Column Pairs



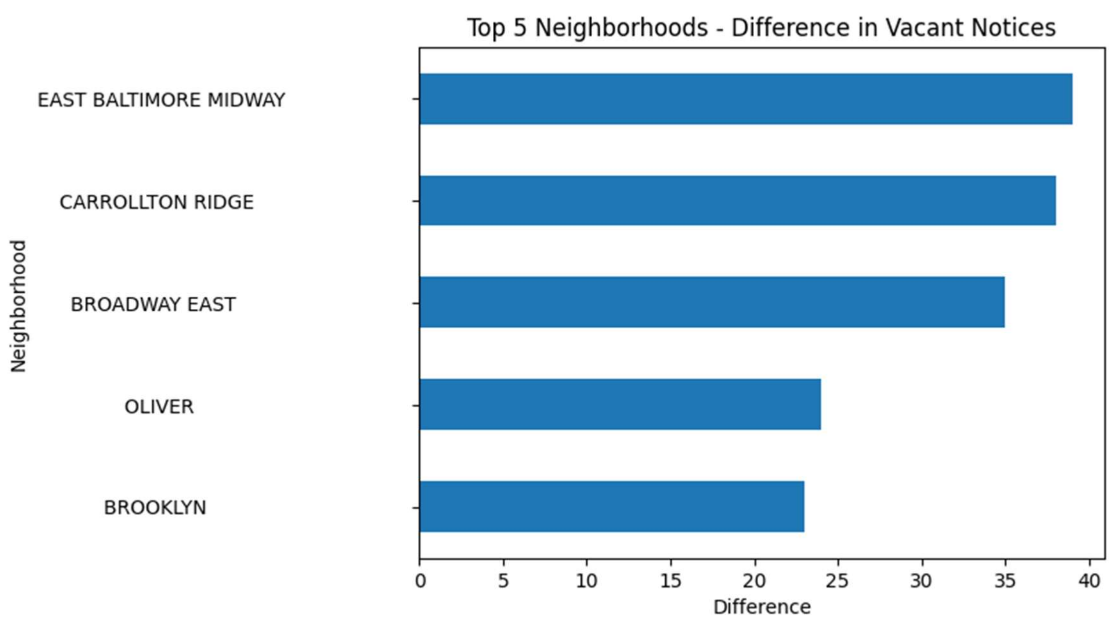
Distribution of Vacant Building Rehabs Count by Neighborhood
Frequency
Number of Rehabs



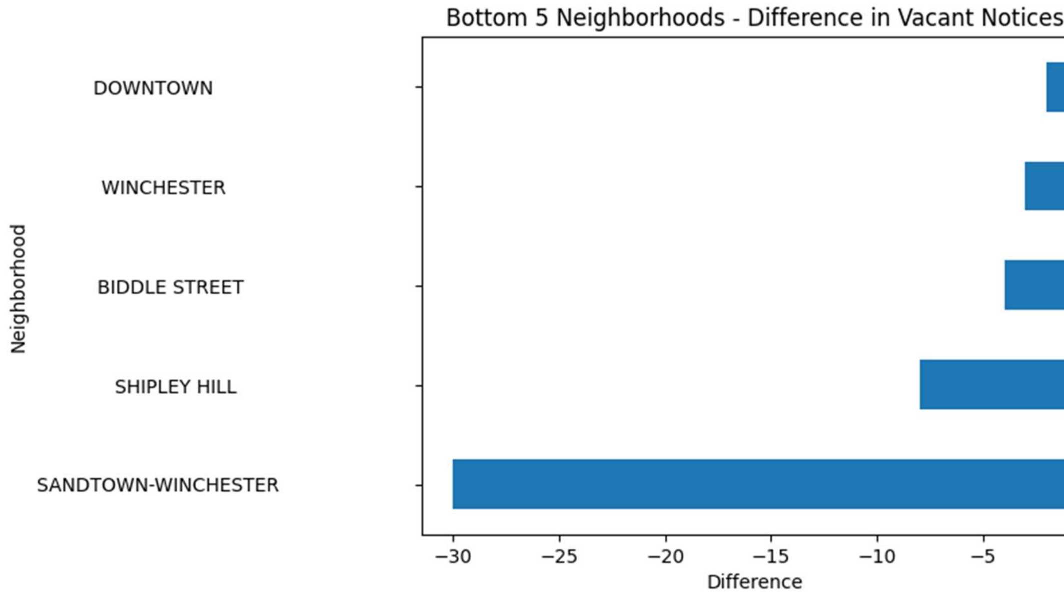
Vacant Notices by Year
Vacant Notices Count
Year

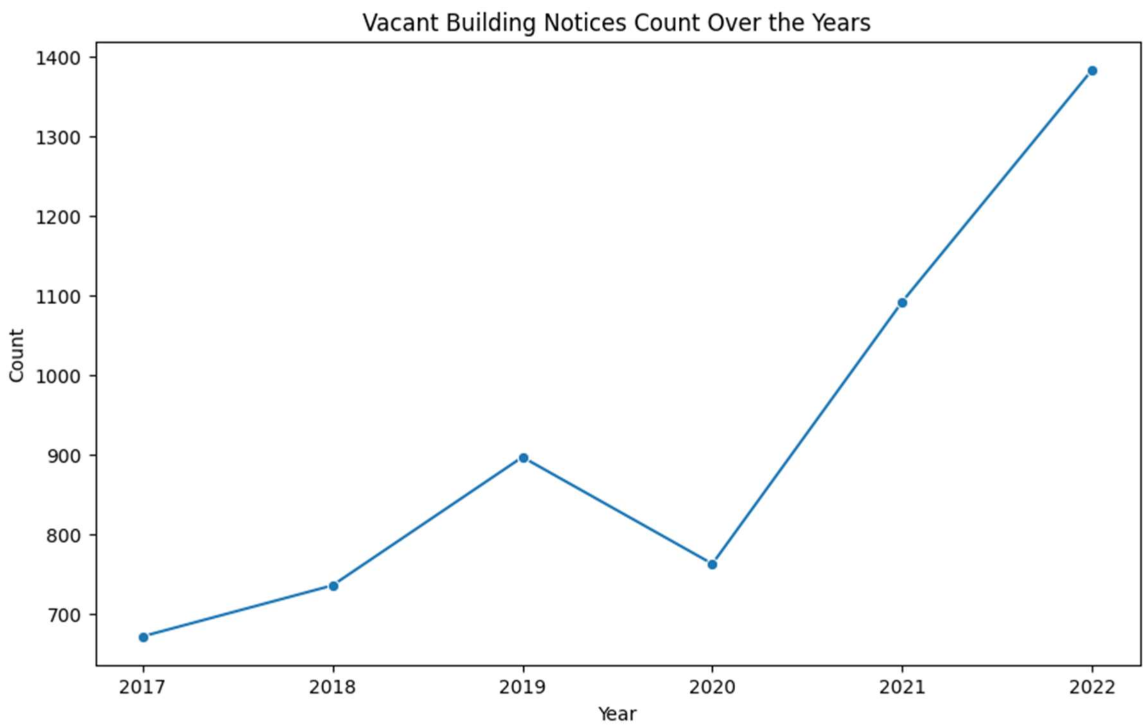


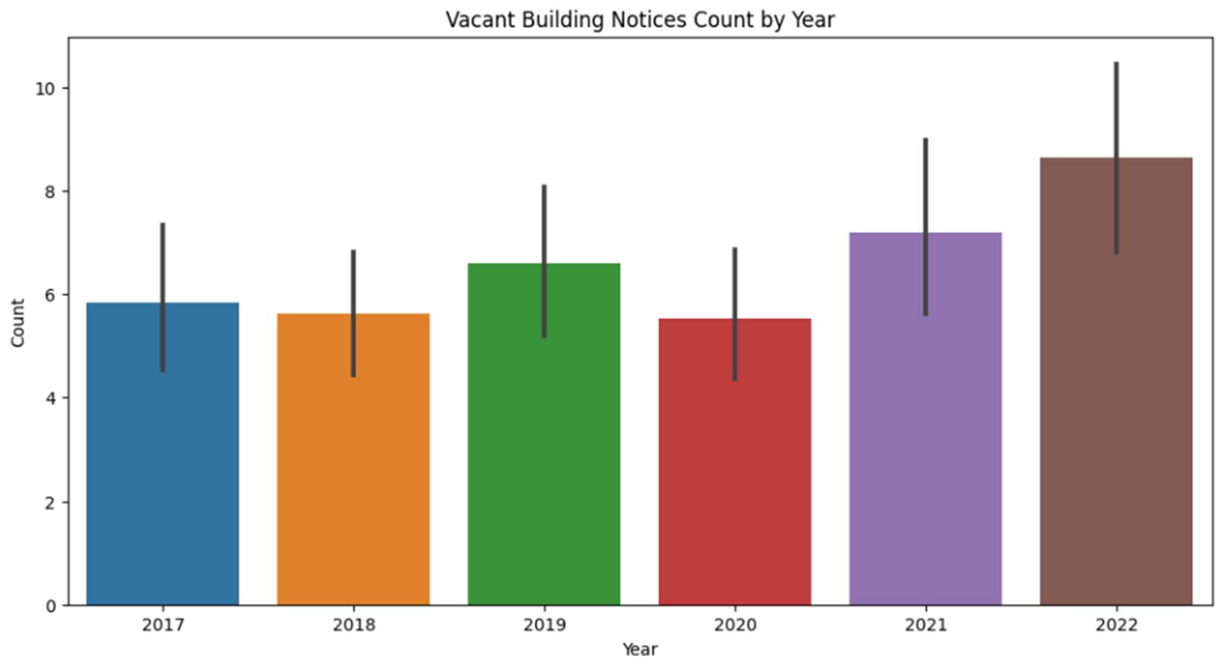
Difference in Vacant Notices Distribution
Frequency
35
30
25
20
15
10
5
0
−30
−20
−10
0
10
20
30
40
Difference in Vacant Notices

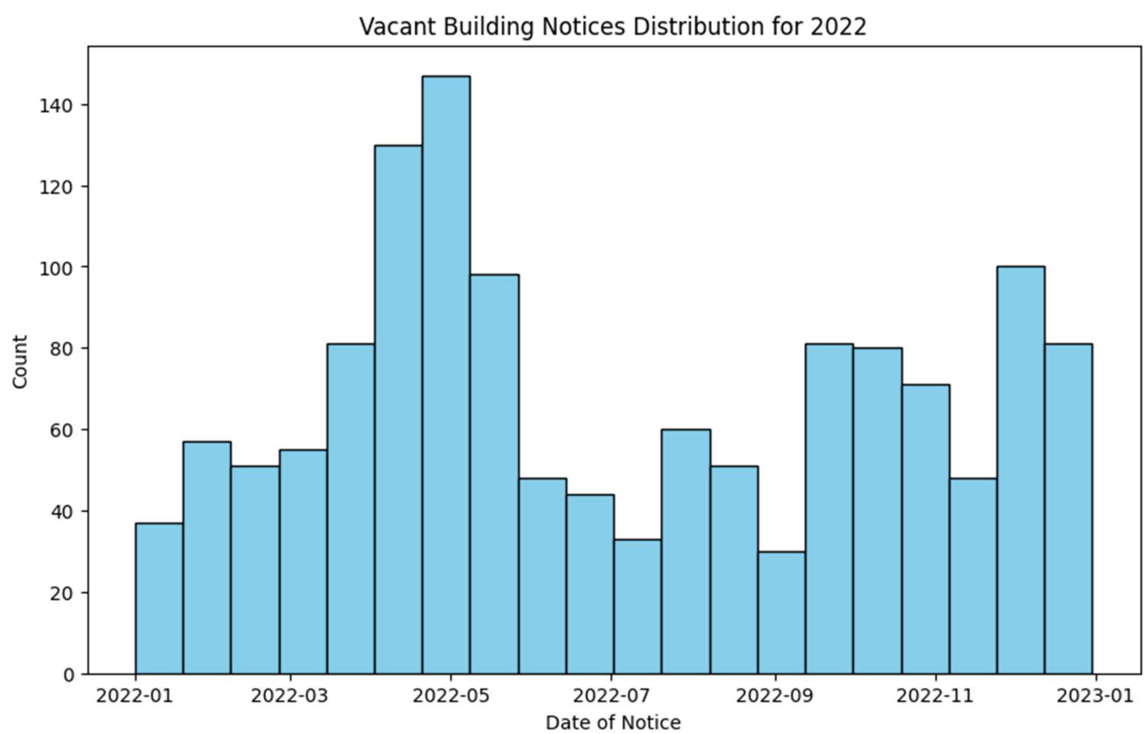


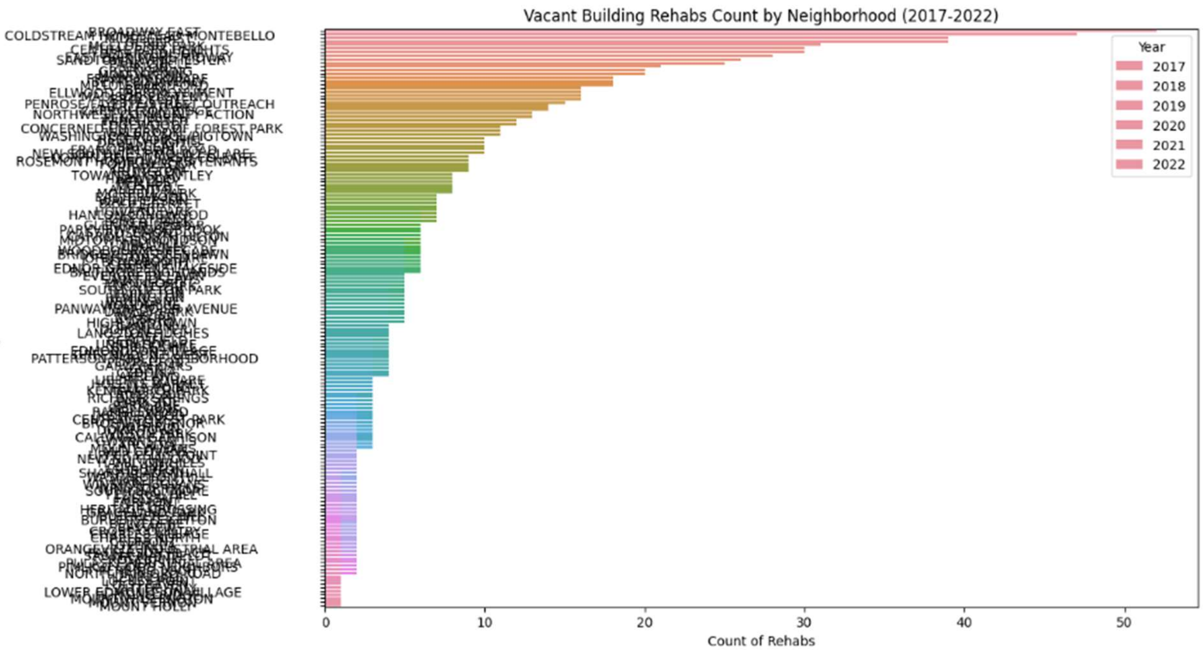
Top 5 Neighborhoods - Difference in Vacant Notices
Neighborhood
EAST BALTIMORE MIDWAY
CARROLLTON RIDGE
BROADWAY EAST
OLIVER
BROOKLYN
0
5
10
15
20
25
30
35
40
Difference


Bottom 5 Neighborhoods - Difference in Vacant Notices
DOWNTOWN
WINCHESTER
BIDDLE STREET
SHIPLEY HILL
SANDTOWN-WINCHESTER
Neighborhood
−30
−25
−20
−15
−10
−5
0
Difference

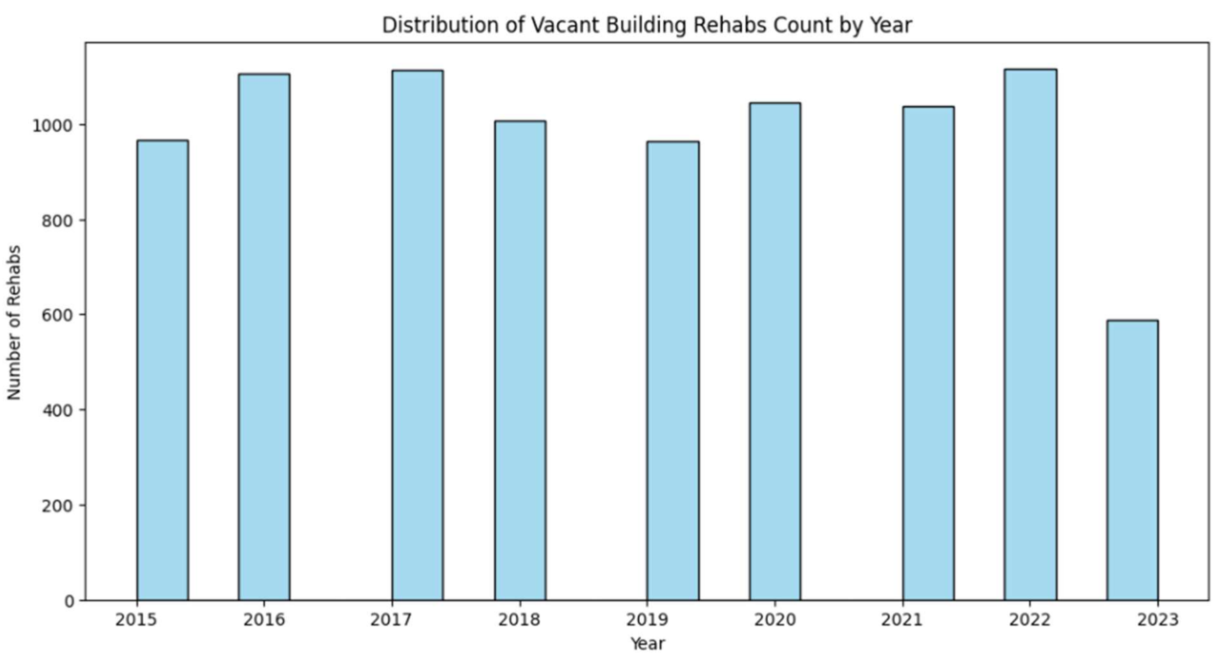

Vacant Building Notices Count Over the Years
1400
1300
1200
1100
1000
900
800
700
Count
2017
2018
2019
2020
2021
2022
Year


Vacant Building Notices Count by Year
Count
0
2
4
6
8
10
2017
2018
2019
2020
2021
2022
Year

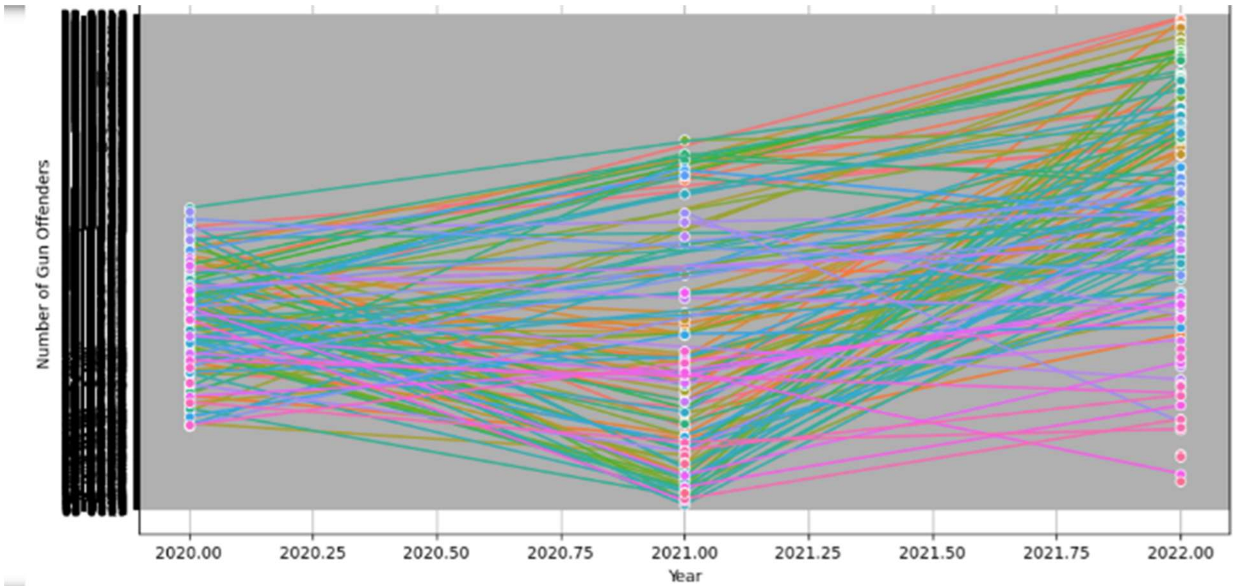

Vacant Building Notices Distribution for 2022
Count
0
20
40
60
80
100
120
140
2022-01
2022-03
2022-05
2022-07
2022-09
2022-11
2023-01
Date of Notice

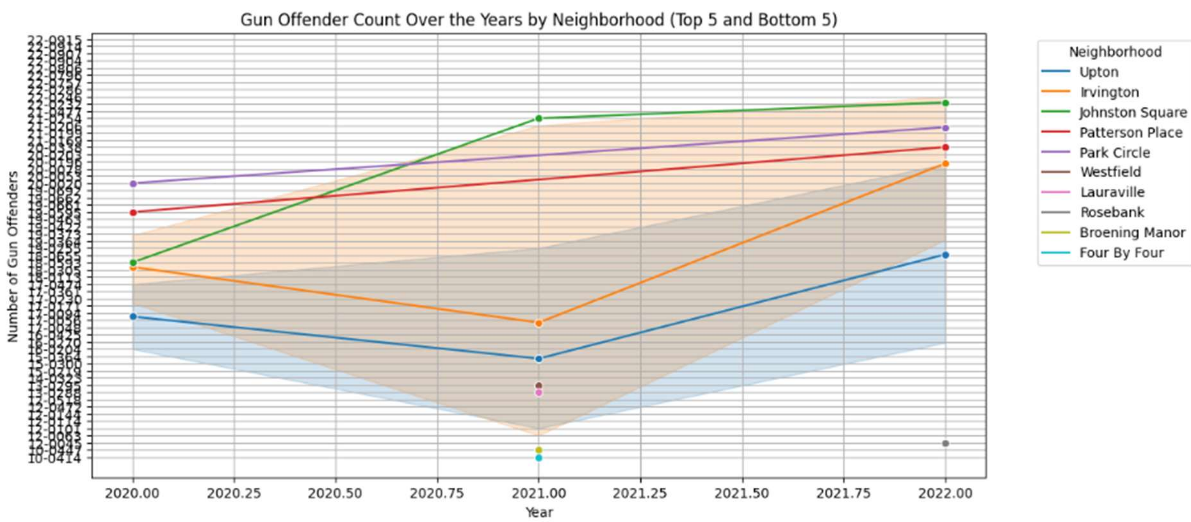


Vacant Building Rehabs Count by Neighborhood (2017-2022)
Neighborhood
Count of Rehabs
0
10
20
30
40
50
Year
2017
2018
2019
2020
2021
2022



Distribution of Vacant Building Rehabs Count by Year
Number of Rehabs
0
200
400
600
800
1000
Year
2015
2016
2017
2018
2019
2020
2021
2022
2023





Conclusion:

In unraveling Baltimore's crime tapestry, our analysis of diverse datasets spotlighted temporal crime shifts, scrutinized vacant building dynamics, and delved into the Gun Offender Registry's impact. Visual storytelling painted nuanced crime portraits, spotlighting fluctuating trends and high/low-performance neighborhoods. Statistical tests added precision, demystifying correlations. Despite limitations, this analysis ignites actionable insights, fueling targeted interventions. Following the analysis of Baltimore's crime landscape, our findings present an intricate narrative. The interplay between vacant building notices, rehabilitations, and crime rates remains elusive, lacking statistically significant correlations. However, a noteworthy revelation emerges concerning the Gun Offender Registry, indicating a substantial correlation with shifting crime dynamics. The observed positive link suggests that heightened gun offender counts may contribute to rising neighborhood crime. Policymakers and law enforcement should heed this insight, focusing on targeted initiatives for supervision and violence prevention within this demographic. While further research, incorporating socioeconomic nuances, promises deeper clarity, the current evidence underscores the pivotal role of gun offender registries in shaping Baltimore's recent crime trends.