Final Project Report

Date: December 20,2023

Subject: DATA 601

Introduction to Data Science

Under the guidance of -

Prof. Samantha Luckhardt

Team members-

Name	Campus ID
Ayushi Bhujade	ZG28331
Charan Sai Gudivada	OA48212
Sree Sai PreethamKadiyam	FG32258
Rami Reddy Kancharla	OC12613
Simhagiri Patnaik Sekharamahanti	MB37094
Namruth Goud Thimmapuram	ES58614
Chandrikarani LNU	TT75490

TASK 1: *Produce an analysis on the crime data of Baltimore over the years 2017-2022 for the violent crimes- murders, shootings and aggravated assaults. Identify the top 5 neighborhoods both with the highest increase and greatest reduction of violent crimes.*

TASK 2: *Use hypothesis testing method to find the correlation among the crime rate change and the outside factors given in the following list datasets to explore whether or not these factors have any relationship to the increase and decreasing crime in the neighborhoods identified in Task 1*

1. *DATA 1: Vacant Building Rehabs*
2. *DATA 2: BPD (Baltimore Police Department) Arrests*
3. *DATA 3: Gun Offender Registry*

Analysis 1) **The overall change in the crime rate for the city of Baltimore**

The below table summarizes the overall change in crime rates for the entire city over the years 2017-2022 for the violent crimes- murders, shootings and aggravated assaults.

	Year	Total_Incidents	Percentage_Change
0	2017	3824	0.000000
1	2018	4016	5.020921
2	2019	4347	8.242032
3	2020	3898	-10.328963
4	2021	4880	25.192406
5	2022	5633	15.430328

Fig (i) Summary table

This can be visually represented in a combination chart of bar graph and line graph for better comparison , where the bars represent the count of the crime for the corresponding year while the line represent the percentage change of the crime over the years
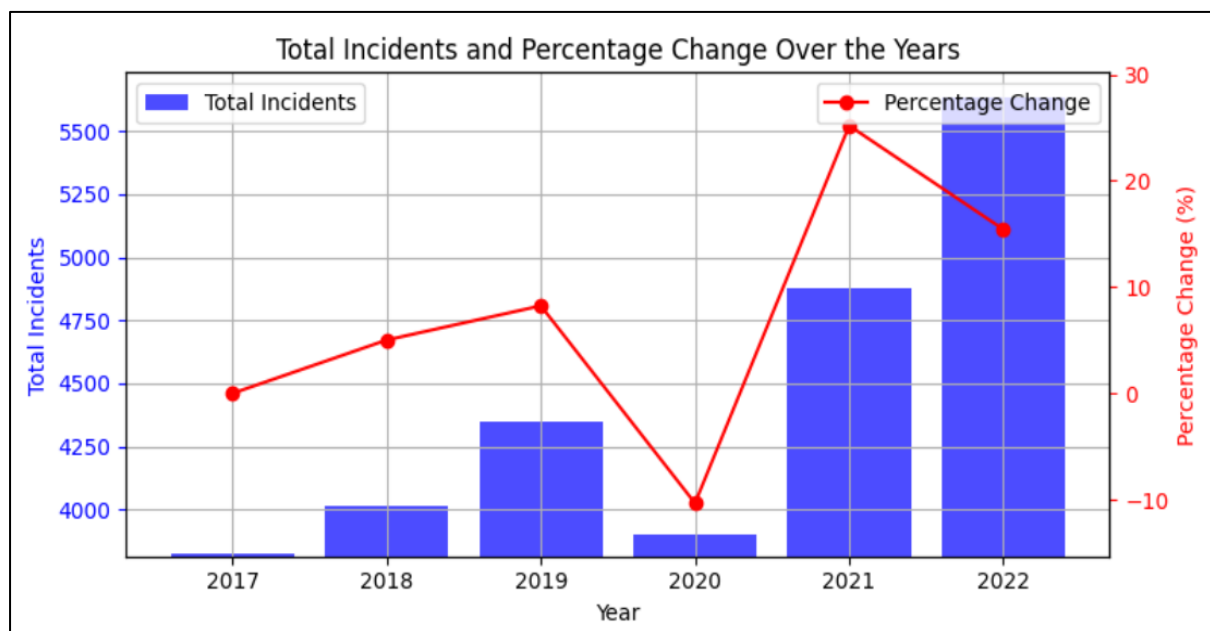

Fig (ii) Comparison chart

Analysis 2) **Top 5 neighborhoods with INCREASE in crime**

The top 5 neighborhoods that faced the increase in crime since 2017 are – *Frankford, Downtown, Upton, Morrell Park, Belair-Edison.* The corresponding crime rate change is between the years 2017 and 2022.

The below table gives a summary of the count over the period of five years for the top 5 neighborhoods with the INCREASE in crime.

Year Neighborhood	2017	2018	2019	2020	2021	2022	Crime_change	Crime_rate_change
FRANKFORD	57.0	88.0	108.0	110.0	117.0	174.0	117.0	205.263158
DOWNTOWN	91.0	136.0	118.0	103.0	147.0	169.0	78.0	85.714286
UPTON	76.0	76.0	98.0	74.0	109.0	154.0	78.0	102.631579
MORRELL PARK	32.0	25.0	36.0	35.0	46.0	85.0	53.0	165.625000
BELAIR-EDISON	102.0	86.0	105.0	95.0	108.0	145.0	43.0	42.156863

Fig(iii) Summary Table

This can be visually represented using a bar chart, giving the comparison of the crime statistics over the five years for the top 5 neighborhoods with the INCREASE in crime rates
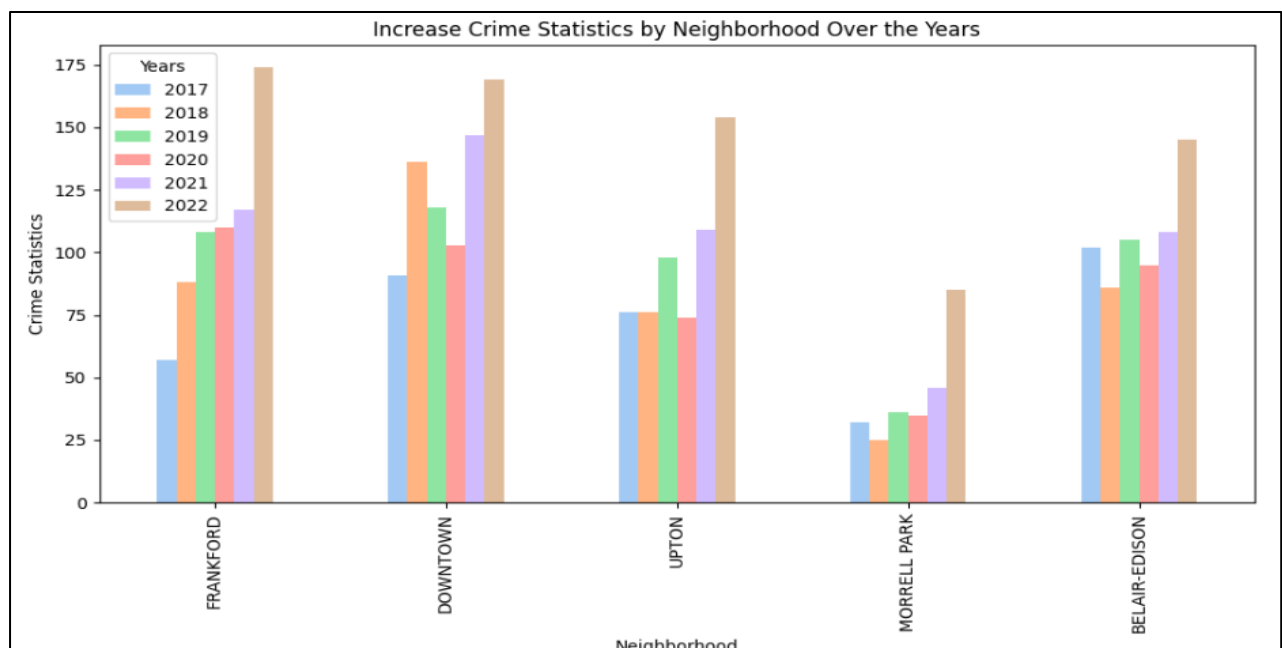

Fig (iv) Comparison chart

Analysis 3) **Top 5 neighborhoods with DECREASE in crime**

The top 5 neighborhoods that faced the decrease in crime since 2017 are – *Perkin Homes, Boyd-Booth, Johnston Square, Homeland, Greenmount West.* The corresponding crime rate change is between the years 2017 and 2022.

The below table gives a summary of the count over the period of five years for the top 5 neighborhoods with the DECREASE in crime.

Year	2017	2018	2019	2020	2021	2022	Crime_change	Crime_rate_change
Neighborhood								
PERKINS HOMES	16.0	13.0	15.0	13.0	3.0	1.0	-15.0	-93.750000
BOYD-BOOTH	24.0	6.0	15.0	14.0	12.0	13.0	-11.0	-45.833333
JOHNSTON SQUARE	39.0	28.0	30.0	19.0	23.0	29.0	-10.0	-25.641026
HOMELAND	13.0	10.0	9.0	0.0	2.0	4.0	-9.0	-69.230769
GREENMOUNT WEST	14.0	12.0	12.0	10.0	16.0	5.0	-9.0	-64.285714

Fig(v) Summary Table

This can be visually represented using a bar chart, giving the comparison of the crime statistics over the five years for the top 5 neighborhoods with the DECREASE in crime rates
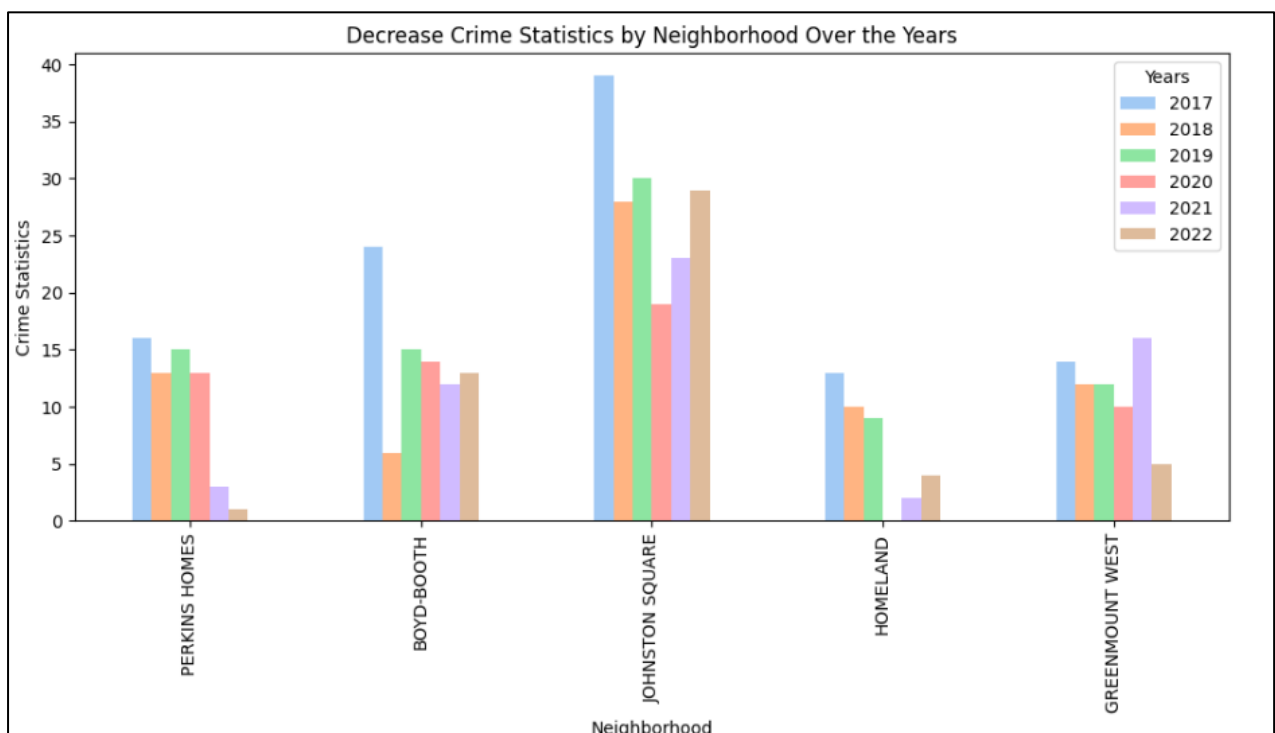

Fig (vi) Comparison chart

DATA 1: VACANT BUILDING REHABS

This dataset represents the location of vacant buildings rehabs located throughout the City of Baltimore

Analysis for the neighborhoods with INCREASE in crime

Analysis 1) A summary table

The below table summarizes the crime count and the rehab count for the neighborhoods with INCREASE in crime over the years 2017-2022

Neighborhood_x	Year_x	crime_count	rehab_count
BELAIR-EDISON	2017	102	43.0
BELAIR-EDISON	2018	86	44.0
BELAIR-EDISON	2019	105	46.0
BELAIR-EDISON	2020	95	26.0
BELAIR-EDISON	2021	108	34.0
BELAIR-EDISON	2022	145	30.0
DOWNTOWN	2017	91	3.0
DOWNTOWN	2018	136	1.0
DOWNTOWN	2019	118	6.0
DOWNTOWN	2020	103	3.0
DOWNTOWN	2022	169	2.0
FRANKFORD	2017	57	21.0
FRANKFORD	2018	88	22.0
FRANKFORD	2019	108	12.0
FRANKFORD	2020	110	8.0
FRANKFORD	2021	117	9.0
FRANKFORD	2022	174	5.0
MORRELL PARK	2017	32	15.0
MORRELL PARK	2018	25	7.0
MORRELL PARK	2019	36	7.0
MORRELL PARK	2020	35	7.0
MORRELL PARK	2021	46	4.0
MORRELL PARK	2022	85	8.0

Fig(vii) Summary Table

Analysis 2) **Visualize results**

The below graphs are for the 5 different neighborhoods with INCREASE in crime. Every graph is a line chart comparison between the crime count and the rehab count in those neighborhoods over the years 2017-2022.

(The missing years in the graph represents the missing data for that year)
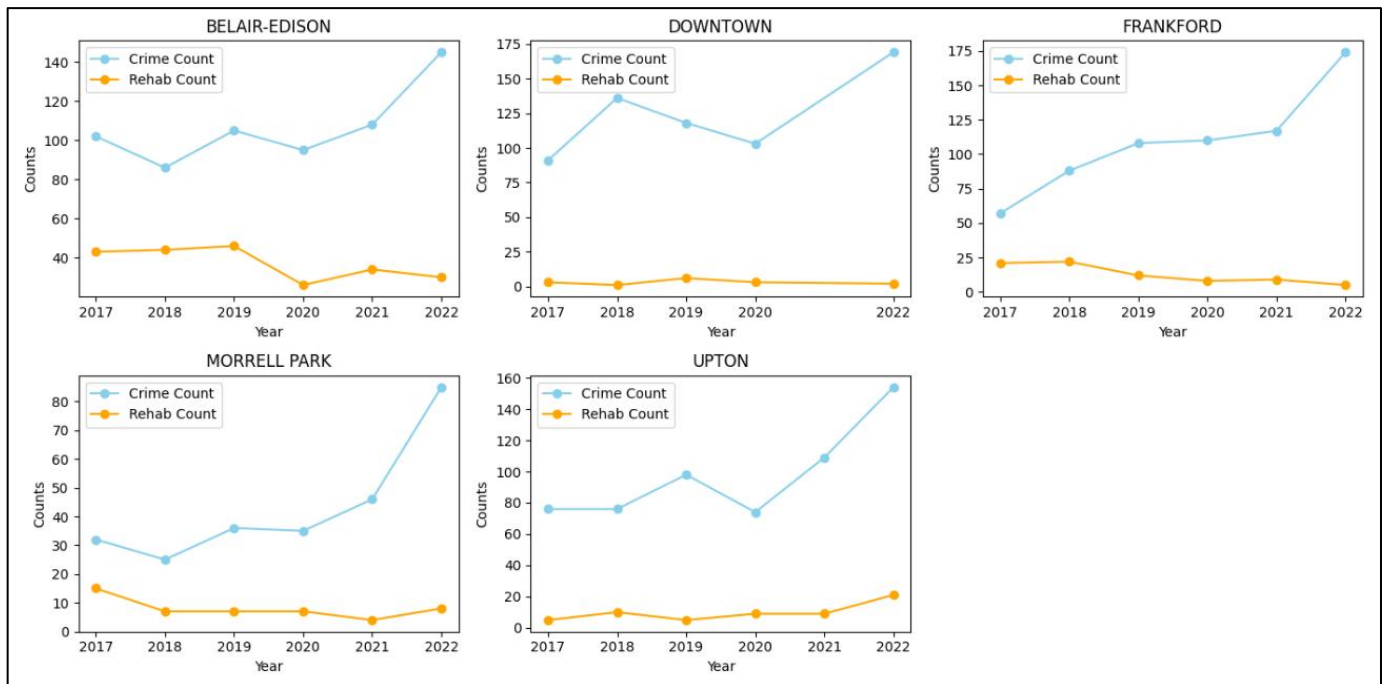

Fig (viii) Comparison chart

Analysis 3) **Hypothesis testing**

Since we need to compare the numerical data like 'crime_count' and 'rehab_count' across different neighborhoods and years, a paired T-test would be more suitable for comparing means or differences between these variables within each neighborhood

Hypothesis test: **T-test**

T-statistic: `10.72448959525234`

P-value: `2.0101177091110666e-11`

Result: **Reject Null Hypothesis.** There is a significant difference between crime_count and rehab_count

Pearson Correlation coefficient: `0.06043446371717581`

Interpretation: There is weak or no linear relationship between the crime count and the rehab count in the neighborhoods with INCREASE in crime

Analysis 1) **A summary table**

The below table summarizes the crime count and the rehab count for the neighborhoods with DECREASE in crime over the years 2017-2022

Neighborhood_x	Year_x	crime_count	rehab_count
BOYD-BOOTH	2017	24	2.0
BOYD-BOOTH	2018	6	3.0
BOYD-BOOTH	2019	15	1.0
BOYD-BOOTH	2021	12	4.0
BOYD-BOOTH	2022	13	5.0
GREENMOUNT WEST	2017	14	13.0
GREENMOUNT WEST	2018	12	13.0
GREENMOUNT WEST	2019	12	10.0
GREENMOUNT WEST	2020	10	6.0
GREENMOUNT WEST	2021	16	3.0
GREENMOUNT WEST	2022	5	3.0
HOMELAND	2018	10	1.0
HOMELAND	2019	9	1.0
HOMELAND	2021	2	1.0
JOHNSTON SQUARE	2017	39	5.0
JOHNSTON SQUARE	2018	28	11.0
JOHNSTON SQUARE	2019	30	10.0
JOHNSTON SQUARE	2020	19	4.0
JOHNSTON SQUARE	2021	23	12.0
JOHNSTON SQUARE	2022	29	5.0

Fig(ix) Summary Table

Analysis 2) **Visualize results**

The below graphs are for the 5 different neighborhoods with DECREASE in crime. Every graph is a line chart comparison between the crime count and the rehab count in those neighborhoods over the years 2017-2022.

(The missing years in the graph represents the missing data for that year)
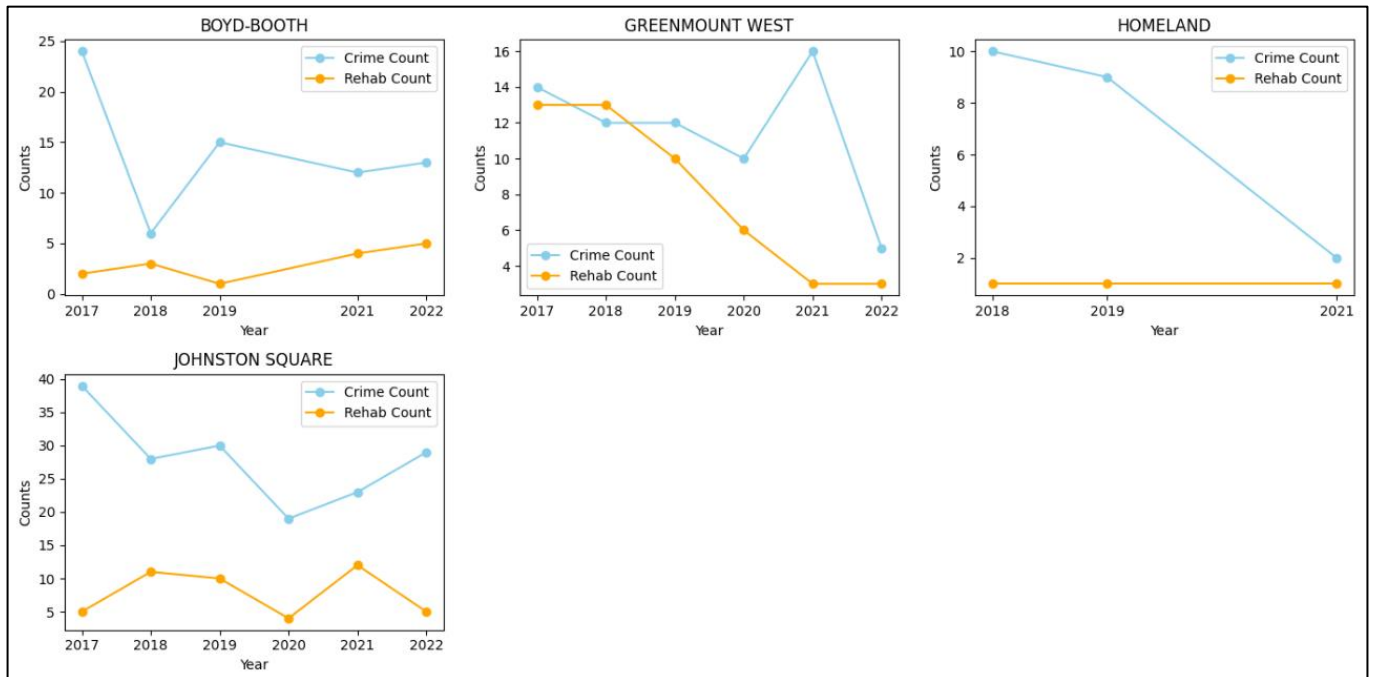

Fig (x) Comparison chart

Analysis 3) **Hypothesis testing**

Since we need to compare the numerical data like 'crime_count' and 'rehab_count' across different neighborhoods and years, a paired T-test would be more suitable for comparing means or differences between these variables within each neighborhood

Hypothesis test: **T-test**

T-statistic: `5.21653511200186`

P-value: `4.912781772276238e-05`

Result: **Reject Null Hypothesis.** There is a significant difference between crime_count and rehab_count

Pearson Correlation coefficient: `0.31298644919160684`

Interpretation: There is positive but weak linear relationship between the crime count and the rehab count in the neighborhoods with DECREASE in crime.

DATA 2: BPD (Baltimore Police Department) ARRESTS

This dataset represents the arrest made by the Baltimore Police Department.

Analysis for the neighborhoods with INCREASE in crime

Analysis 1) A summary table

The below table summarizes the crime count and the arrest count for the neighborhoods with INCREASE in crime over the years 2017-2022

Neighborhood_x	Year_x	crime_count	arrest_count
BOYD-BOOTH	2017.0	24.0	122
BOYD-BOOTH	2018.0	6.0	53
BOYD-BOOTH	2019.0	15.0	46
BOYD-BOOTH	2020.0	14.0	18
BOYD-BOOTH	2021.0	12.0	9
BOYD-BOOTH	2022.0	13.0	20
GREENMOUNT WEST	2017.0	14.0	30
GREENMOUNT WEST	2018.0	12.0	13
GREENMOUNT WEST	2019.0	12.0	10
GREENMOUNT WEST	2020.0	10.0	9
GREENMOUNT WEST	2021.0	16.0	3
GREENMOUNT WEST	2022.0	5.0	6
HOMELAND	2017.0	13.0	52
HOMELAND	2018.0	10.0	36
HOMELAND	2019.0	9.0	23
HOMELAND	2021.0	2.0	17
HOMELAND	2022.0	4.0	4
JOHNSTON SQUARE	2017.0	39.0	52
JOHNSTON SQUARE	2018.0	28.0	31
JOHNSTON SQUARE	2019.0	30.0	37
JOHNSTON SQUARE	2020.0	19.0	19
JOHNSTON SQUARE	2021.0	23.0	14
JOHNSTON SQUARE	2022.0	29.0	23
PERKINS HOMES	2017.0	16.0	20

Fig(xi) Summary Table

Analysis 2) **Visualize results**

The below graphs are for the 5 different neighborhoods with INCREASE in crime. Every graph is a line chart comparison between the crime count and the arrest count in those neighborhoods over the years 2017-2022.

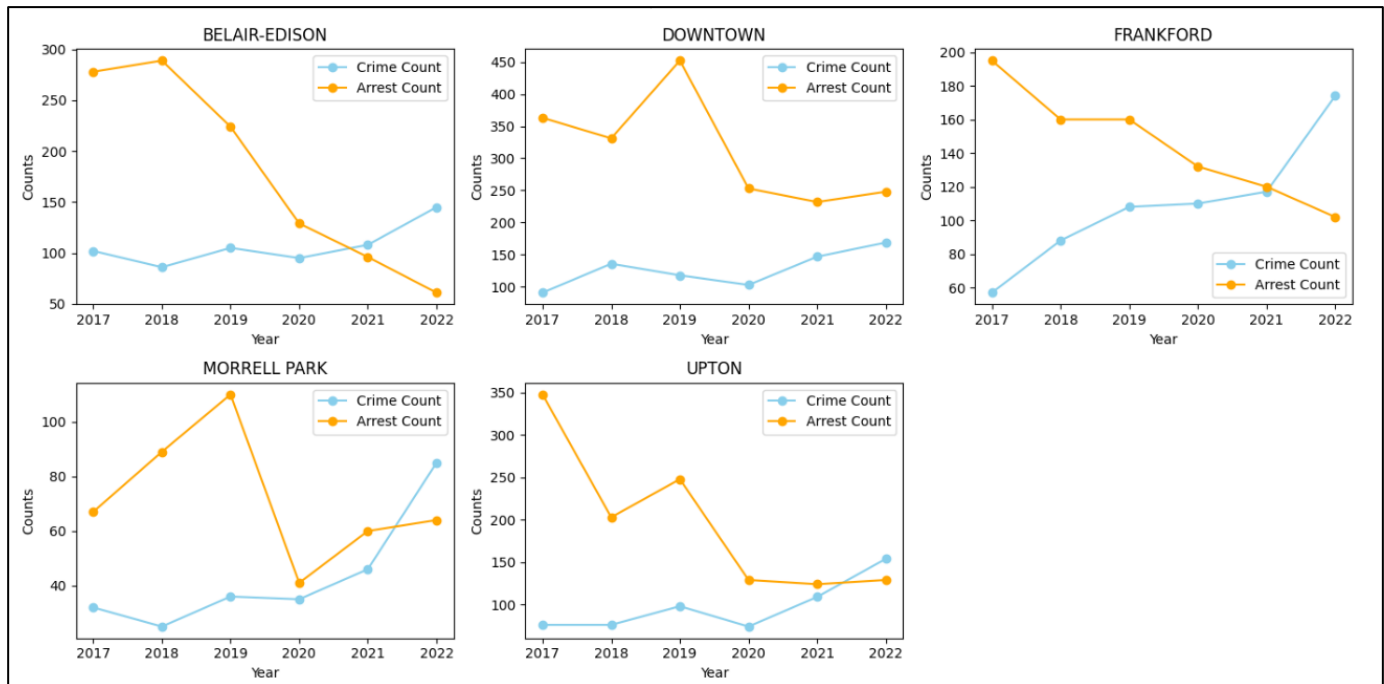*(The missing years in the graph represents the missing data for that year)*

Fig (xii) Comparison chart

Analysis 3) **Hypothesis testing**

Since we need to compare the numerical data like 'crime_count' and 'arrest_count' across different neighborhoods and years, a paired T-test would be more suitable for comparing means or differences between these variables within each neighborhood

Hypothesis test: **T-test**

T-statistic: `-4.536775324129274`

P-value: `9.17193551592652e-05`

Result: **Reject Null Hypothesis.** There is a significant difference between crime_count and arrest_count

Pearson Correlation coefficient: `0.25888728672647443`

Interpretation: There is positive but weak linear relationship between the crime count and the arrest count in the neighborhoods with INCREASE in crime.

Analysis 1) **A summary table**

The below table summarizes the crime count and the arrest count for the neighborhoods with DECREASE in crime over the years 2017-2022

Neighborhood_x	Year_x	crime_count	arrest_count
BOYD-BOOTH	2017.0	24.0	122
BOYD-BOOTH	2018.0	6.0	53
BOYD-BOOTH	2019.0	15.0	46
BOYD-BOOTH	2020.0	14.0	18
BOYD-BOOTH	2021.0	12.0	9
BOYD-BOOTH	2022.0	13.0	20
GREENMOUNT WEST	2017.0	14.0	30
GREENMOUNT WEST	2018.0	12.0	13
GREENMOUNT WEST	2019.0	12.0	10
GREENMOUNT WEST	2020.0	10.0	9
GREENMOUNT WEST	2021.0	16.0	3
GREENMOUNT WEST	2022.0	5.0	6
HOMELAND	2017.0	13.0	52
HOMELAND	2018.0	10.0	36
HOMELAND	2019.0	9.0	23
HOMELAND	2021.0	2.0	17
HOMELAND	2022.0	4.0	4
JOHNSTON SQUARE	2017.0	39.0	52
JOHNSTON SQUARE	2018.0	28.0	31
JOHNSTON SQUARE	2019.0	30.0	37
JOHNSTON SQUARE	2020.0	19.0	19
JOHNSTON SQUARE	2021.0	23.0	14
JOHNSTON SQUARE	2022.0	29.0	23

Fig(xiii) Summary Table

Analysis 2) **Visualize results**

The below graphs are for the 5 different neighborhoods with DECREASE in crime. Every graph is a line chart comparison between the crime count and the arrest count in those neighborhoods over the years 2017-2022.

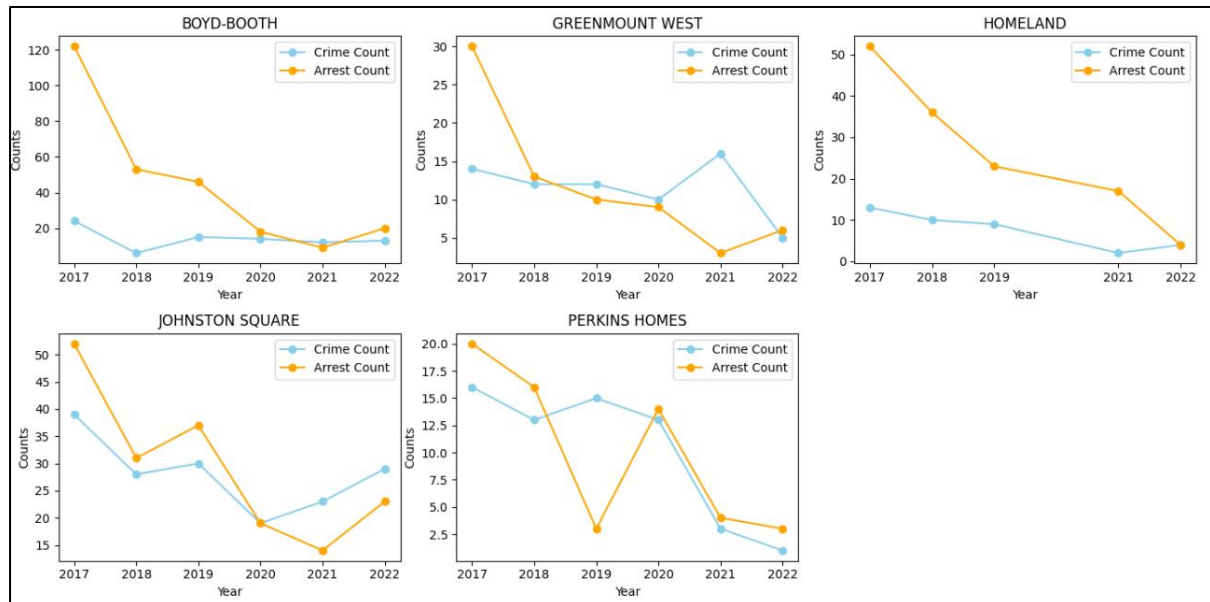*(The missing years in the graph represents the missing data for that year)*

Fig (xiv) Comparison chart

Analysis 3) **Hypothesis testing**

Since we need to compare the numerical data like 'crime_count' and 'arrest_count' across different neighborhoods and years, a paired T-test would be more suitable for comparing means or differences between these variables within each neighborhood

Hypothesis test: **T-test**

T-statistic: `-2.420276411440015`

P-value: `0.022247189301684227`

Result: **Reject Null Hypothesis.** There is a significant difference between crime_count and arrest_count

Pearson Correlation coefficient `0.42525100182998576`

Interpretation: There is positive but low linear relationship between the crime count and the arrest count in the neighborhoods with DECREASE in crime.

DATA 3: GUN OFFENDER REGISTRY

This dataset represents the Gun Offender Registry for persons convicted of at least one gun-related offense and are required to register their name and address with police

Analysis for the neighborhoods with INCREASE in crime

Analysis 1) A summary table

The below table summarizes the crime count and the gun count for the neighborhoods with INCREASE in crime over the years 2017-2022

Neighborhood_x	Year_x	crime_count	gun_count
BELAIR-EDISON	2017	102	6.0
BELAIR-EDISON	2018	86	8.0
BELAIR-EDISON	2019	105	13.0
BELAIR-EDISON	2020	95	15.0
BELAIR-EDISON	2021	108	30.0
BELAIR-EDISON	2022	145	44.0
DOWNTOWN	2020	103	1.0
DOWNTOWN	2021	147	1.0
FRANKFORD	2017	57	4.0
FRANKFORD	2018	88	6.0
FRANKFORD	2019	108	6.0
FRANKFORD	2020	110	8.0
FRANKFORD	2021	117	23.0
FRANKFORD	2022	174	30.0
MORRELL PARK	2017	32	1.0
MORRELL PARK	2019	36	2.0
MORRELL PARK	2020	35	2.0
MORRELL PARK	2021	46	4.0
MORRELL PARK	2022	85	6.0
UPTON	2017	76	2.0
UPTON	2018	76	3.0
UPTON	2019	98	1.0
UPTON	2020	74	2.0
UPTON	2021	109	15.0

Fig(xv) Summary Table

Analysis 2) **Visualize results**

The below graphs are for the 5 different neighborhoods with INCREASE in crime. Every graph is a line chart comparison between the crime count and the gun count in those neighborhoods over the years 2017-2022.

(The missing years in the graph represents the missing data for that year)
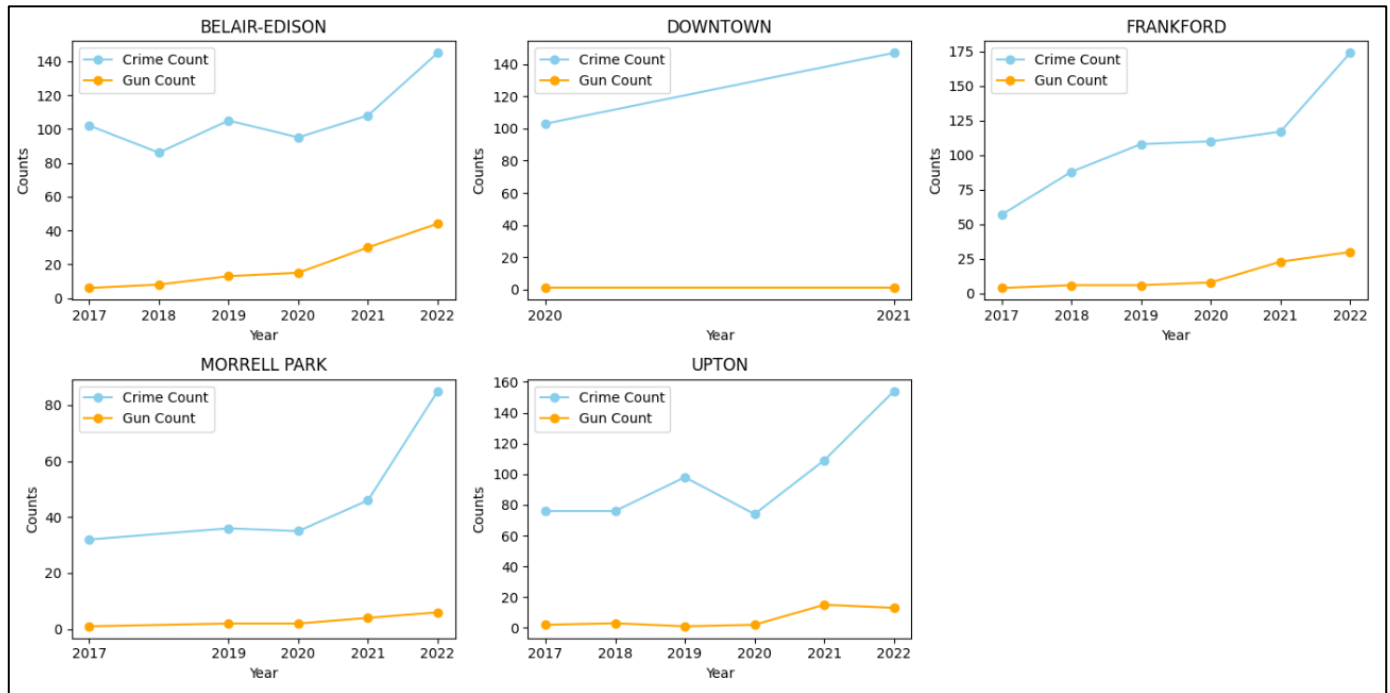

Fig (xvi) Comparison chart

Analysis 3) **Hypothesis testing**

Since we need to compare the numerical data like 'crime_count' and 'gun_count' across different neighborhoods and years, a paired T-test would be more suitable for comparing means or differences between these variables within each neighborhood

Hypothesis test: **T-test**

T-statistic: 13.512049839304584

P-value: 1.031688286888082e-12

Result: **Reject Null Hypothesis.** There is a significant difference between crime_count and gun_count

Pearson Correlation coefficient: 0.6099997085909663

Interpretation: There is positive but moderate linear relationship between the crime count and the gun count in the neighborhoods with INCREASE in crime.

Analysis for the neighborhoods with DECREASE in crime

Analysis 1) A summary table

The below table summarizes the crime count and the gun count for the neighborhoods with DECREASE in crime over the years 2017-2022

Neighborhood_x	Year_x	crime_count	gun_count
BOYD-BOOTH	2020	14	1.0
BOYD-BOOTH	2021	12	2.0
BOYD-BOOTH	2022	13	5.0
GREENMOUNT WEST	2017	14	2.0
GREENMOUNT WEST	2019	12	1.0
GREENMOUNT WEST	2021	16	1.0
GREENMOUNT WEST	2022	5	1.0
HOMELAND	2019	9	1.0
HOMELAND	2021	2	1.0
HOMELAND	2022	4	1.0
JOHNSTON SQUARE	2017	39	2.0
JOHNSTON SQUARE	2018	28	1.0
JOHNSTON SQUARE	2020	19	2.0
JOHNSTON SQUARE	2021	23	6.0
JOHNSTON SQUARE	2022	29	8.0
PERKINS HOMES	2018	13	2.0
PERKINS HOMES	2019	15	2.0
PERKINS HOMES	2020	13	1.0
PERKINS HOMES	2021	3	1.0

Fig(xvii) Summary Table

Analysis 2) **Visualize results**

The below graphs are for the 5 different neighborhoods with DECREASE in crime. Every graph is a line chart comparison between the crime count and the gun count in those neighborhoods over the years 2017-2022.

(The missing years in the graph represents the missing data for that year)

Fig (xviii) Comparison chart

Analysis 3) **Hypothesis testing**

Since we need to compare the numerical data like 'crime_count' and 'gun_count' across different neighborhoods and years, a paired T-test would be more suitable for comparing means or differences between these variables within each neighborhood

Hypothesis test: **T-test**

T-statistic: `6.327653420677774`

P-value: `5.801497690843811e-06`

Result: **Reject Null Hypothesis.** There is a significant difference between crime_count and gun_count

Pearson Correlation coefficient: `0.45600531583952514`

Interpretation: There is positive but moderate linear relationship between the crime count and the gun count in the neighborhoods with INCREASE in crime.