Assignment: Project 2

Class: DATA 601

Date: 12/20/2023

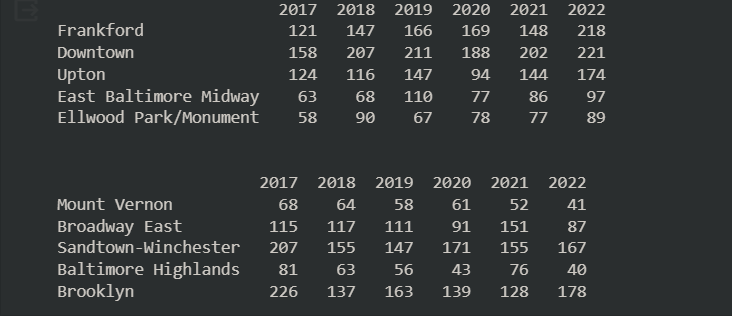
Group 2: Luis Vargas Ramirez, Nasim Aalemi Hari Pavan Bhuti, James Gilmore Jr, Soumitra Abhijit Joshi, Kushal Pasumarty, Aditya Phadtare, Joseph Schoenbaum

Final Project Report

Top Neighborhoods: FRANKFORD, DOWNTOWN, UPTON, EAST BALTIMORE MIDWAY, ELLWOOD PARK/MONUMENT

Bottom Neighborhoods: MOUNT VERNON, BROADWAY EAST, SANDTOWN-WINCHESTER, BALTIMORE HIGHLANDS, BROOKLYN

Crime Rates we used from project 1:



Joseph James Gun Offender Registry: <https://arcg.is/1K9fDH>

Luis Nasim Hari Vacant Building Notices: <https://arcg.is/>

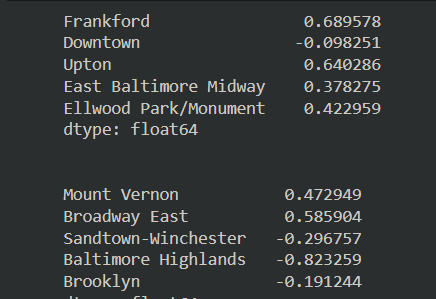
Aditya Kushal Soumitra BPD Arrests: <https://arcg.is/0SLeza0>

**Gun Offender Registry Factor Report**:

(James Gilmore Jr and Joseph Schoenbaum)

NULL HYPOTHESIS: Gun offender registry has no significant effect on crime rate within the top and bottom crime rate neighborhoods from 2017-2022.

The first thing we did was get the gun offender registry counts from 2017-2022 for the neighborhoods specified from project 1 (stated above). After that we checked correlation results to see if there was high correlation between gun offender registry and crime rate for our results.\



Above are the correlation matrices for crime rate and gun offender registry within the top and bottom crime rate neighborhoods in Baltimore from 2017-2022.

As shown in the correlation matrix, there is strong positive correlation between crime rate and gun offender registry for Frankford and Upton. A moderate positive correlation for East Baltimore Midway and Ellwood Park/Monument. And a weak negative correlation for Downtown. Overall I'd say there is a noticeable correlation between high crime neighborhoods and gun offender registry. This is an interesting piece of data that would be worth looking further into.

For the bottom 5 crime rate neighborhoods Baltimore Highlands has a strong negative correlation. Mount Vernon and Broadway East have a moderate positive correlation. Lastly, Brooklyn and Sandtown-Winchester have weak negative correlation. This is curious because while gun offender count went up in many of these neighborhoods, crime rate went down.

There could be many factors at play. For one, there is not a lot of gun offender registry data for these neighborhoods. In addition, most of the data tends to be recent. One idea could be there has been an emphasis by police to focus on gun offenses in recent years. This would increase gun offender registry overall.

Our objective is to find if gun offender registry has a linear relationship on the top and bottom crime count neighborhoods in Baltimore from 2017-2022. A linear regression model is a good way to test the relationship between two quantitative variables such as crime count and gun offender count. Through the summary we are also able to see the p-value for both the top and bottom neighborhoods.

For the top neighborhoods the p-value is .323 which is greater than .05. This means there was no statistically significant effect of gun offender registry on crime rate within these neighborhoods from 2017-2022.

For the bottom neighborhoods the p-value is .001 which is less than .05. This means there was a statistically significant effect of gun offender registry on crime rate within these neighborhoods from 2017-2022.

When we combine the data we get a p-value of .005 which is less than .05. This means there was a statistically significant effect of gun offender registry on crime rate within these neighborhoods from 2017-2022.

For all three results the R-squared value was relatively low. This means that the variability in crime counts cannot be explained by our linear relationship founded above.

Based on these findings we reject our null hypothesis that Gun offender registry has no significant effect on crime rate within the (combined) top and bottom crime rate neighborhoods from 2017-2022.

After analyzing our data, we could see that as the years progressed, the number of Gun Offender Registries steadily increased after every year. We also noticed that there was a significant increase in Registries from the year 2020 to 2021, with the number remaining relatively high in 2022. The steady increase in Gun Offender Registries is similar to the overall crime increase from the years 2020 to 2022. We could infer that outside factors such as the economic landscape, political officers, and a global pandemic play critical roles in the overall variance.

**BPD Arrest Report**

(Soumitra Abhijit Joshi, Kushal Pasumarty, Aditya Phadtare)

Null Hypothesis: BPD arrest has no significant effect on crime rate within the top and bottom crime rate neighborhoods from 2017-2022.

To begin with, we used BPD arrest data from 2017-2022 for the neighborhoods specified from neighborhoods specified from Project 1. We then checked correlation results to understand whether the BPD arrest factor affects the crime rate for our results.

Figure 1

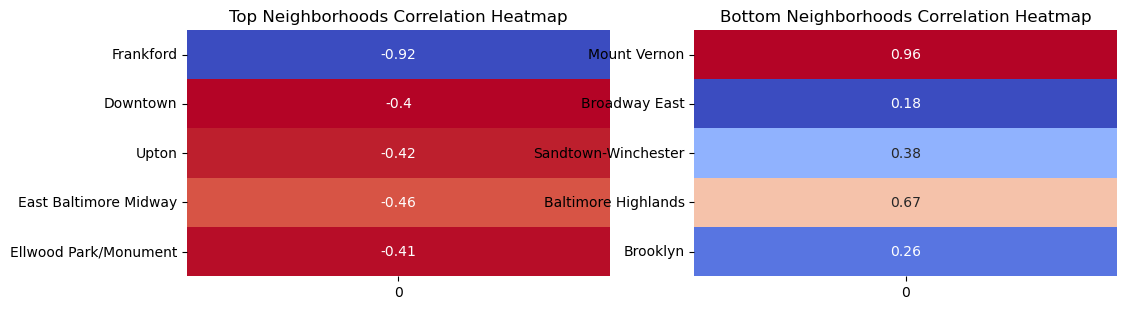
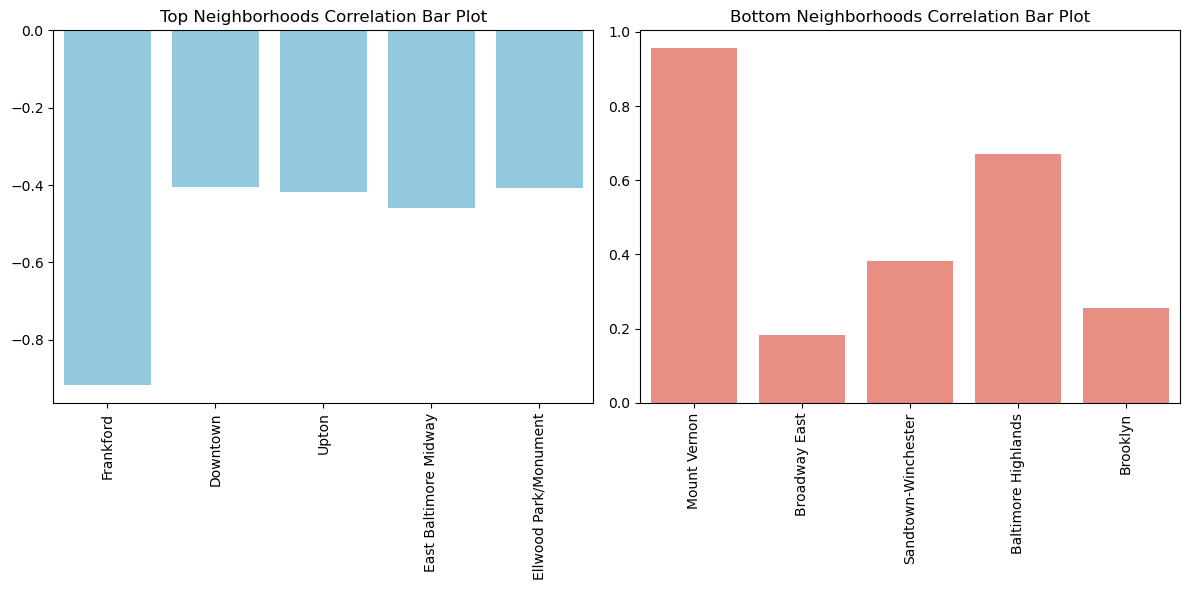


Figure 2



The above plots, heatmap (figure 1), and Bar graphs (figure 2), we can say:

**Highest Neighborhoods Correlation:**

Frankford -0.917130 -> Strong Negative Correlation

Downtown -0.404104 -> Medium Negative Correlation

Upton -0.417335 -> Medium Negative Correlation

East Baltimore Midway -0.458197 -> Medium Negative Correlation

Ellwood Park/Monument -0.408772 -> Medium Negative Correlation

**Lowest Neighborhoods Correlation:**

Mount Vernon 0.956233 -> Strong Positive Correlation

Baltimore Highlands 0.669747 -> Strong Positive Correlation

Sandtown-Winchester 0.382933 -> Medium Positive Correlation

Broadway East 0.181967 -> Weak Positive Correlation

Brooklyn 0.255906 -> Weak Positive Correlation

**P-value Analysis for Arrests and Neighborhoods with High Crime:**

P-value for Predictor Variable (x1): 0.079 (greater than 0.05)

Conclusion: There is no statistically significant effect of the predictor variable (x1) on the dependent variable (y) at the 0.05 significance level.

**P-value Analysis for Arrests and Neighborhoods with Low Crime:**

P-value for Predictor Variable (x1): 0.011 (less than 0.05)

Conclusion: There is a statistically significant effect of the predictor variable (x1) on the dependent variable (y) at the 0.05 significance level.

**P-value Analysis for Arrests and Neighborhoods with Both Low/High Crime:**

P-value for Predictor Variable (x1): 0.000 (less than 0.05)

Conclusion: There is a statistically significant effect of the predictor variable (x1) on the dependent variable (y) at the 0.05 significance level

For neighborhoods with high crime, there is no significant effect. For neighborhoods with low crime, there is a significant effect. For neighborhoods with both low and high crime, there is a significant effect. These results suggest that the impact of the predictor variable varies across different neighborhood crime levels, being significant in low and mixed-crime neighborhoods but not in high-crime neighborhoods.

**R-squared value analysis:**

Result 1: 0.106

Result 2: 0.914

Result 3: 0.840

Result 1 has a relatively low R-squared, suggesting limited explanatory power. Results 2 and 3 have high R-squared values, indicating a strong linear relationship. These findings suggest that the model performs well in explaining the variability in the dependent variable for Results 2 and 3, while Result 1 shows a weaker performance. It's essential to consider the R-squared values in conjunction with the p-values for a comprehensive understanding of the model's effectiveness.

**Decision on Null Hypothesis:**

Considering both analyses, it appears that there is a significant impact of the predictor variable on crime rates in neighborhoods with low and mixed crime. However, in neighborhoods with high crime, the impact is not statistically significant. Therefore, we may reject the null hypothesis for low and mixed-crime neighborhoods but fail to reject it for high-crime neighborhoods. The strength of the relationships is also evident in the high R-squared values for Results 2 and 3.

# **Factor: Vacant Building Notices**

(Luis Vargas Ramirez, Nasim Aalemi, Hari Pavan Bhuti)

# Vacant Building Notices (VBN) are a specific category given to buildings that are unoccupied inside Baltimore city due to many different factors. The purpose of analyzing this specific factor is to see if there is any relationship between VBN and the crime rate.

Null Hypothesis (H0): There is no significant difference in crime rates across different categories of vacant building notices for the neighborhoods.

Alternative Hypothesis (H1): There is a significant difference in crime rates across different categories of vacant building notices for the neighborhoods.

To investigate the relationship of crime rates when coupled with vacant building notices, one of the statistical approaches we can undertake is to study the correlation of VBN and the crime rates. Following the project 2 guidelines, we are also tasked to determine the correlation coefficients for this relationship but only for specific neighborhoods which were determined in project 1 and for the duration of 2017 to 2022 years. We have carried out the correlation analysis of vacant building notices vs. crime rates using python and have discussed the results in detail further down in this report.

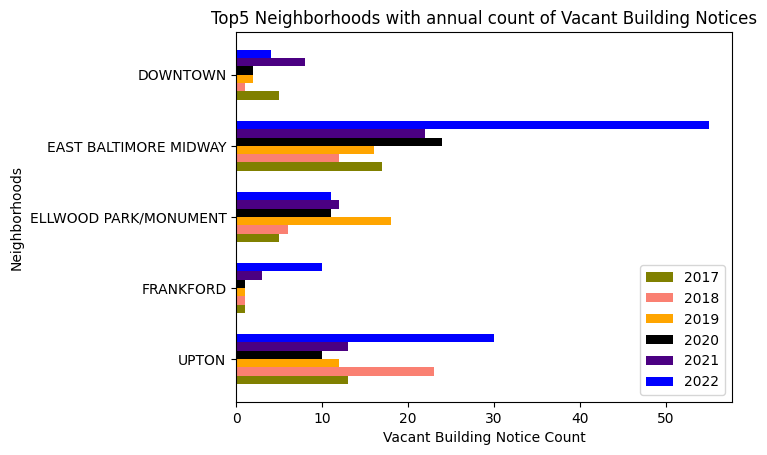
##### **Top 5 crime neighborhoods:**

Per project 2 guidelines, we have extracted counts of incidents over the years for specified neighborhoods, in this case Top5 neighborhoods, depicted below in a bar graph for 2017 to 2022 years.

The specified Top5 neighborhoods which were determined during project 1 are:

**'FRANKFORD', 'DOWNTOWN', 'UPTON', 'EAST BALTIMORE MIDWAY', 'ELLWOOD PARK/MONUMENT'.**

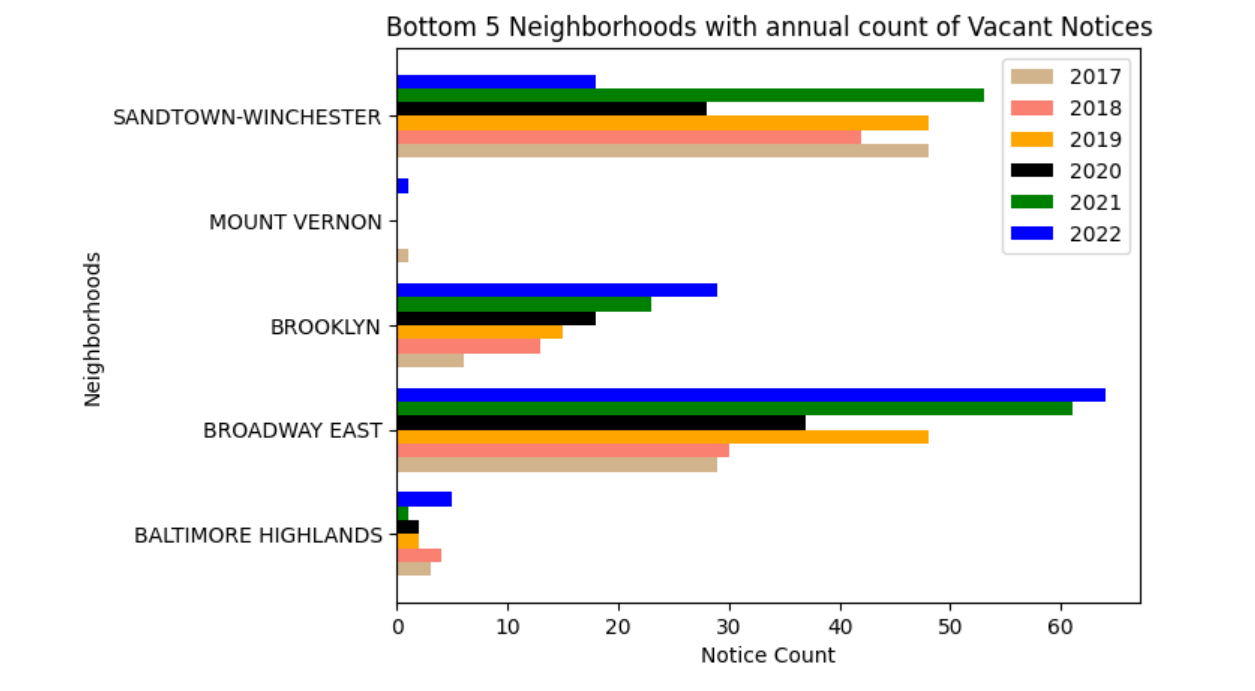
The graph visualization displays counts of the Vacant Building Notices (VBN) when combined with the Top5 neighborhoods that we determined in project1. From the analysis and visualizations, we can deduce that '**EAST BALTIMORE MIDWAY**' and '**UPTON**' had the highest count of VBN when combined with Top 5 neighborhoods from project 1 while '**FRANKFORD**' and '**DOWNTOWN**' had the lowest.



##### **Least 5 crime neighborhoods:**

A similar approach has been applied with the Bottom5 neighborhoods coupled with Vacant Building Notices. The specified Bottom5 neighborhoods are:

**‘MOUNT VERNON', 'BROADWAY EAST', 'SANDTOWN-WINCHESTER', 'BALTIMORE HIGHLANDS', 'BROOKLYN'**



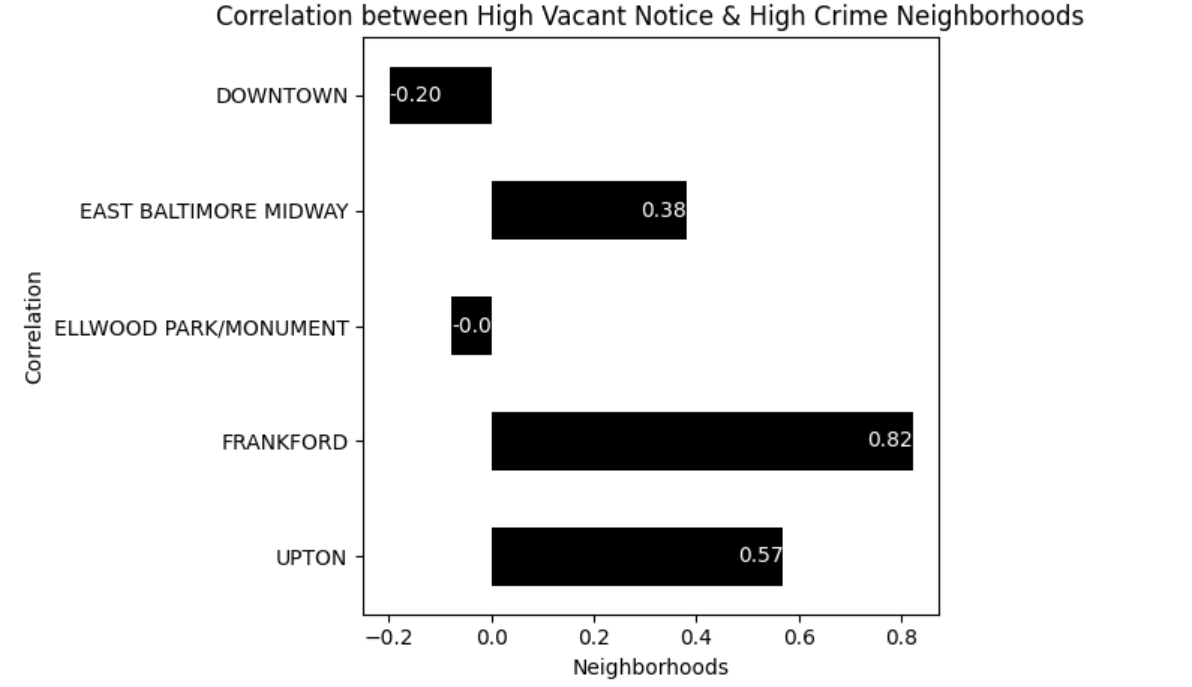
From the visualizations we can see that **‘SANDTOWN-WINCHESTER’** and **‘BROADWAY EAST’** have the highest VBN counts while **‘MOUNT VERNON',** and **'BALTIMORE HIGHLANDS'** have the lowest.

## **Correlation with Crime Rate and VBN:**

Correlation is a statistical measure that shows the relationship between two variables. In other words, it measures the extent to which a shift in one variable results in a change in another. When there is a positive correlation, it means that when one variable rises, the other also tends to rise, and when there is a negative correlation, it means that when one variable rises, the other tends to fall.

As per project-2 guidelines, before diving into the hypothesis testing it is a good method to do correlation and get a glimpse of what we can expect in this factor. However, it is best not to arrive at a possible conclusion and to determine the analysis solely based on the correlation. We applied correlation measures to determine the correlation between the variables Top5 high crime neighborhoods and vacant building notices.

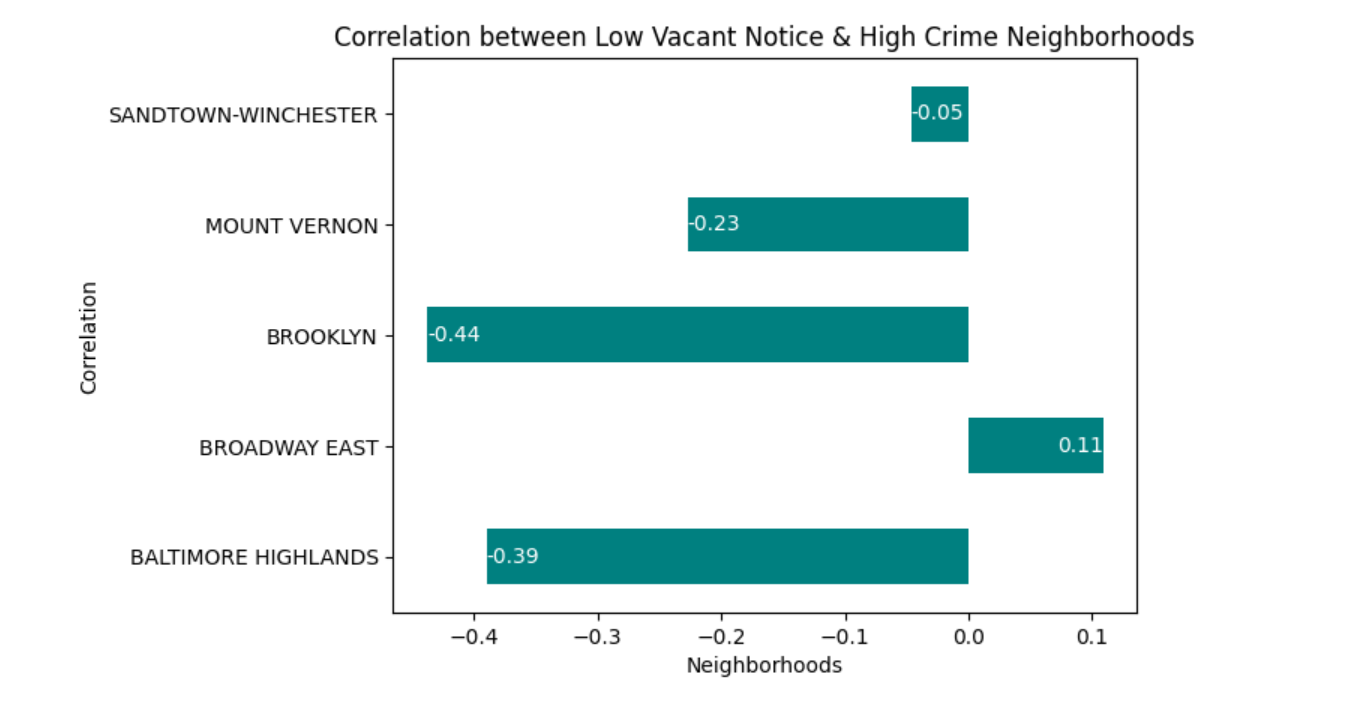
##### **Top5 High Crime Neighborhoods:**



From the correlation we can observe that **‘FRANKFORD’** and **‘UPTON’** have a strong positive correlation with the vacant notices whereas, ‘**EAST BALTIMORE MIDWAY’** has a good positive correlation.

Out of the 5 neighborhoods **‘DOWNTOWN’** has the most negative correlation while **‘ELLWOOD PARK / MONUMENT’** has a -0.0 correlation. Although **‘DOWNTOWN’** might be justified by the fact that there is a lot of commerce and tourism going on there, **ELLWOOD PARK / MONUMENT’** does not seem to have this kind of characteristics. However, it is important to keep in mind that this correlation alone does not conclusively establish a direct relationship between crime and this specific factor.

##### **Low crime neighborhoods:**



In this correlation graph, **‘BROOKLYN’** and **‘BALTIMORE HIGHLANDS’** have a strong negative correlation whereas, ‘**MOUNT VERNON’** has a good negative correlation and **‘SANDTOWN-WINCHESTER’** has a weak but still, negative correlation. From this it is evident that these 4 neighborhoods showed less crime as there are less vacant buildings.

Although ‘**BROADWAY EAST’** has a positive correlation, it is still too early to arrive at a conclusion as there are many other factors that may come into play and there is much more analysis that should be done.

## **Hypothesis Testing**

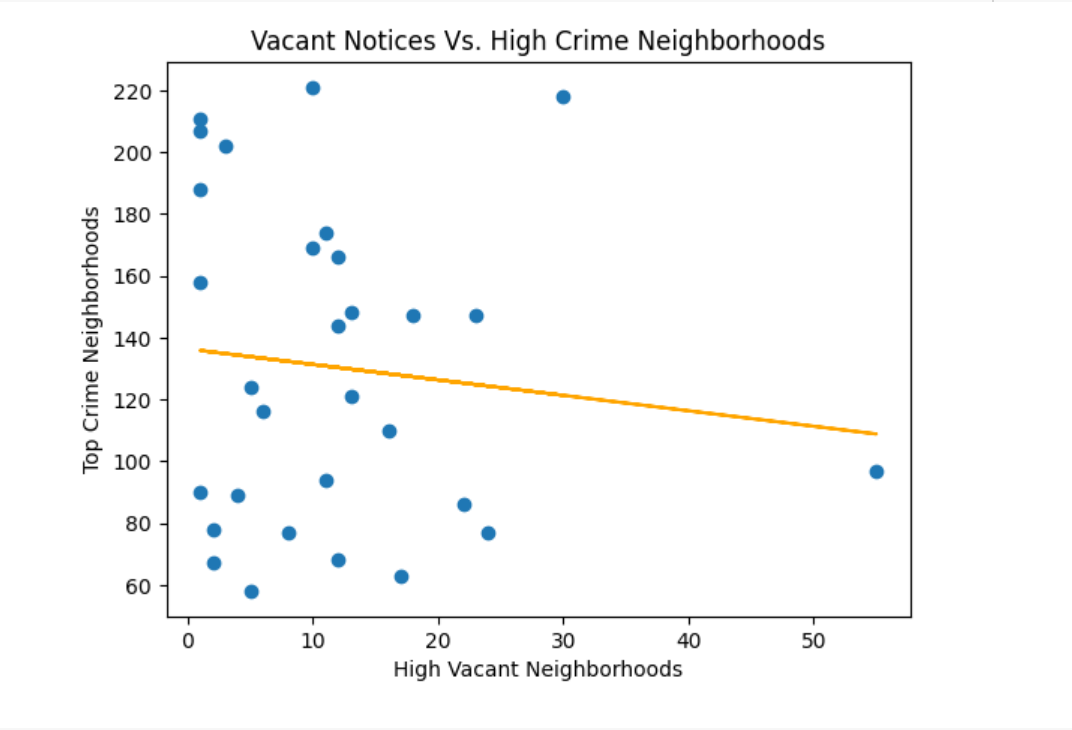
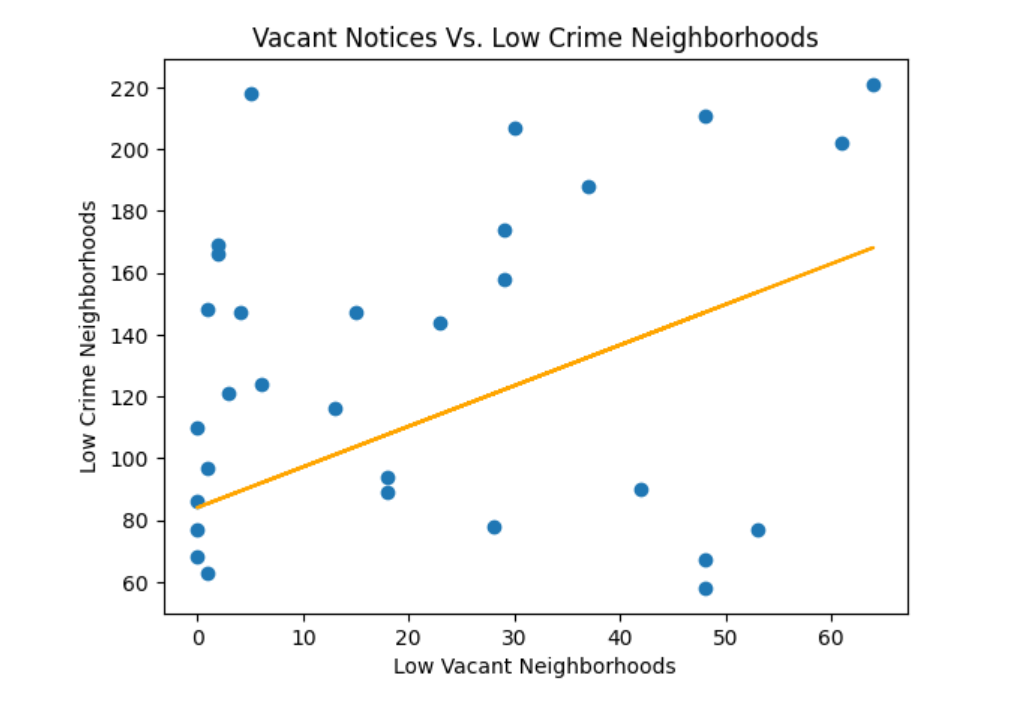
Here, our goal is to check if there is any linear relationship between vacant buildings and crime rate. For that, Linear regression analysis is a good approach to find a relationship between the two variables vacant building notices count and the crime rate count from the years 2017-2022.

We applied linear regression analysis for both top 5 crime rate neighborhoods and least 5 crime rated neighborhoods individually to study the results and to confirm if the linear regression analysis is the best approach for this specific factor.

### **Linear Regression Analysis:**

Linear regression analysis is a statistical tool for analyzing the connection between one or more independent variables and a dependent variable. Linear regression is used to determine whether there is a significant linear relationship between variables and to create predictions based on this relationship.

##### **High and Low crime Neighborhoods:**



After analyzing the graph the first thing we thought is that ‘it doesn't make any sense’.

For high crime neighborhoods:

P-value is 0.565 which is greater than 0.05, this means that there is no significant difference in crime rates across different categories of vacant notices for top5 neighborhoods from 2017 to 2022.

For low crime neighborhoods:

Conversely, the P-value is 0.03, which is less than 0.05, this indicates that there is a significant difference in the crime rates with respect to vacant buildings from the year 2017 to 2022.

Now, there can be other variables which may have made an influence on the results, one of them being the issue with multi-collinearity. This signifies that there are two or more highly correlated independent variables, making it difficult to separate the individual effect of each variable on the dependent variable.

We arrived at a conclusion that linear regression analysis is not the correct approach to follow for this specific factor and shifted to T-test analysis.

### **T - test:**

A t-test is a statistical test that compares the means of two groups to see if they differ substantially. The t-test is based on the t-distribution, which is comparable to the normal distribution but takes into account the additional uncertainty provided by working with small sample sizes. The t-distribution resembles the normal distribution as sample sizes increase.

After determining that linear regression analysis is not suitable for our variables, we started to apply the T-test analysis for high crime and low crime variables with respect to vacant buildings count.

##### **High and Low Crime Neighborhoods:**

From the data we collected, it's evident that there's a structured analysis contrasting high crime neighborhoods with vacant neighborhoods, both in the top and bottom tiers of crime rates.

The statistical measures, particularly the T-statistic and P-value, offer valuable insights into the significance of these differences. T-test analysis is what is better suited for our data and the result we got from it made more sense to the team.

For the high crime neighborhoods this is our results  
  
T-Statistic: -12.27845623588145   
P-Value: 9.03446944049104e-18

The T-statistic of -12.278 and an extremely low P-value of 9.034×10-18 unequivocally reject the null hypothesis. This robust rejection indicates a significant effect that the vacant building notices have on the crime rates when combined with the Top 5 high crime neighborhoods. Such a disparity underscores distinct characteristics or conditions between these neighborhoods, warranting further investigation.

For the low crime neighborhoods this is our results

T-Statistic: -8.75254394691726  
P-Value: 3.4031304198486973e-12

Similarly, with a T-statistic of -8.753 and a P-value of 3.403×10−12, the null hypothesis is again rejected. This reaffirms that there exists a highly significant effect on the Bottom 5 high crime neighborhoods when combined with Bottom 5 vacant neighborhoods. The statistical evidence here suggests that even within neighborhoods with lower crime rates, vacant areas have distinct features or challenges compared to their counterparts.

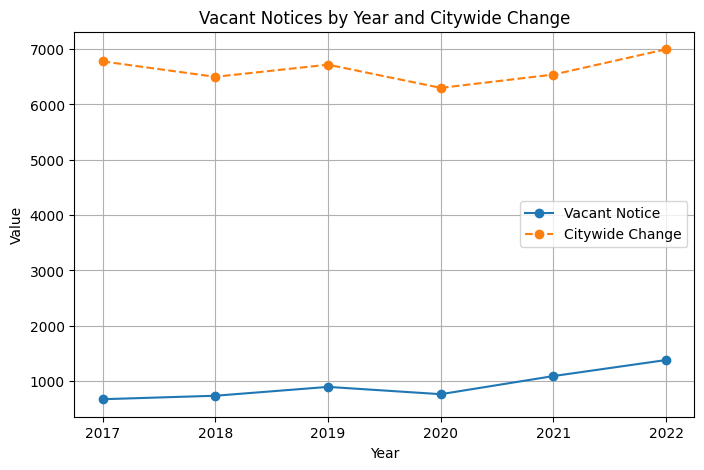
In essence, both sets of results indicate that neighborhoods characterized by high crime rates and vacant properties have discernible differences when compared to their counterparts. These findings hold significant implications for policymakers, urban planners, and community stakeholders, emphasizing the need for targeted interventions and strategies tailored to address the unique challenges and characteristics of these neighborhoods.

## **Citywide Analysis:**

To fully understand the effect of vacant building notices on crime rates throughout all neighborhoods—not simply the top and bottom 5—citywide analysis is crucial. As locations with abandoned buildings may be more prone to criminal activity, this aids in further evaluating if vacant building notices serve as a potential indicator in all of the neighborhoods across Baltimore city.

A citywide analysis of vacant building notices provides an understanding on their potential relationship with overall crime rates. By visualizing the changes over time, the city wide analysis allows us to identify potential trends and patterns within the data. Understanding the relationship between vacant building notifications and crime rates can help shape urban planning and development initiatives, assisting local governments in addressing issues associated with abandoned structures and, as a result, contributing to crime prevention efforts.

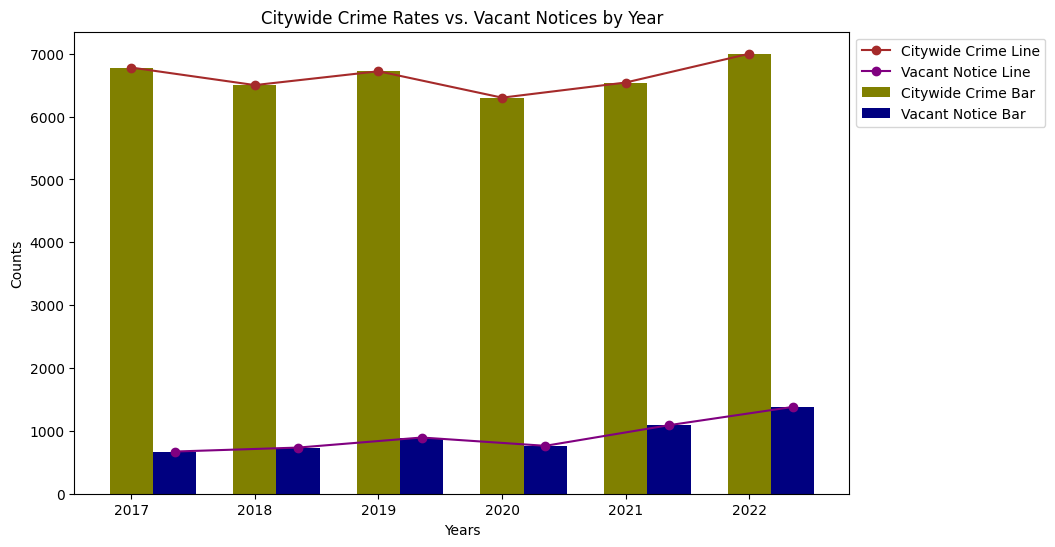
##### **Vacancy Count VS Citywide Crime Change:**



This graph shows an overview of the relationship between vacancy notices count compared to the crime counts in all of the neighborhoods with the years 2017 to 2022.

On observing the graph we can see a trend that if the number of vacancy notices increase or decrease each year there is also an increase in crime rate count most of the time (except in the year 2018). From this graph we can say that vacant buildings are directly proportional to the increase in crime count in the city of Baltimore.

##### **Citywide Crime Count VS Vacancy Notices Count (per year):**



This graph provides a detailed review of comparison between vacancy notices to the crime rate within the years 2017 to 2022 in Baltimore city.

Although, the vacancy notice count is far less compared to the crime count its impact is significant. This graph further solidifies our hypothesis that there is a significant influence by vacancy notices on the citywide crime rate over the years.

To conclude, our analysis provides valuable insights on the complex relationship between vacant buildings and crime rates in Baltimore. Adopting proper statistical tools, such as T-test analysis, revealed relevant patterns that can be used to drive governmental decisions and community development activities. However, given the diverse nature of urban dynamics, the necessity for thorough observation is needed for further understanding and addressing the crime-related challenges.