Reddit Reactions vs Reality: Online Perceptions of Neighborhood Safety in Washington, D.C.

Rebecca Wagner[1] and Elizabeth Bagley[1]

[1]**Georgetown University, McCourt School of Public Policy**

ABSTRACT

Cities all over the United States are battling the process of gentrification within their neighborhoods. This process causes both housing prices to increase, and crime rates to decrease. Additionally, public perceptions of crime have generally increased in many cities, despite crime rates falling nationwide. In this project, we attempt to better understand the relationship between perceptions of crime, true reported rates of crime, and housing markets, in an effort to untangle the many factors at play in gentrification. To do so, we focus specifically on the city of Washington, D.C, and utilize public datasets on crime rates and housing prices. We also collect our own dataset from social media platform Reddit through its API to create a quantification of public safety sentiment. We ultimately find that trends in perceived safety appear to lag behind changes in true crime rates, and that online safety sentiments and housing prices appear to be correlated, though more analysis is needed to parse the nature of this relationship.

Keywords: Data Science, Sentiment Analysis, Crime, Gentrification

INTRODUCTION

The issue of gentrification has become a common discussion in urban centers and regarding urban policies. Washington D.C. has not escaped the nation-wide gentrification crisis, causing the price of housing to skyrocket. An increasing number of studies argue that gentrification and crime are closely related, and in D.C. specifically increasing housing prices and falling poverty (as proxies for gentrification) appear to coincide with a drop in crime rates [Law13]. Of course, even as crime rates fall in many major cities—including the district—public sentiments may not follow suit. In fact, people often feel as though crime rates are higher than they are [Pog24].

This study aims to explore the relationship between these three factors—gentrification, crime, and public sentiment—across different neighborhoods in Washington, D.C. Does the housing market appear to respond to rates of crime or public perceptions of crime? Do perceptions align with actual crime rates, or is there a lag in changing perceptions as residents' neighborhoods become gentrified? We utilize data from Zillow on increasing housing prices as a proxy for gentrification, and data sourced from Open Data DC to measure true rates of crime in D.C. neighborhoods. To measure public perceptions of crime, we scrape data from D.C. centric communities on Reddit and perform sentiment analysis. By combining these three data sources, we are able to identify the ways in which online safety sentiments lag behind true crime rates, and how these indicators correlate with housing prices.

1 DATA

This paper utilizes data from three sources: Zillow, Open Data DC, and Reddit.

1.1 Acquisition

Zillow Zillow is a well-known leader in real estate. From buying and selling homes to rental properties, many people find a place to call home through Zillow. The research arm of Zillow provides unbiased and transparent data on trends in the housing market. We utilized two types of data from Zillow for this project, the first being pre existing data on median home prices and median rent prices, and the second being geographical boundary data.

The Zillow Home Value Index (ZHVI) represents the "typical" home value in a region. A typical value falls between the 35th and 65th percentile. We used the smoothed, seasonally-adjusted data, broken by neighborhood for all homes, to explore home value changes from 2019-2024.

The Zillow Observed Rent Index (ZORI) measures the "typical" market rate of rent in a region. According to Zillow, "The index is dollar-denominated by computing the mean of listed rents that fall into the 35th to 65th percentile range for all homes and apartments in a given region" [Res24]. We use this data, broken down by zip code, to help inform the change in rent price across D.C. neighborhoods from 2019-2024.

We also used a geodatabase provided by Zillow that contains neighborhood geography boundaries for over 650 cities in the U.S. This database contained shape boundaries for each of the most prominent neighborhoods in Washington, D.C, which we used to perform a spatial merging of crime data into neighborhoods.

Open Data DC Open Data D.C. is an official resource made public by the local government of Washington D.C. containing a wide range of data on the district. We made use of two types of data from this resource.

First, we accessed data regarding crime in D.C.. For each year of interest (2019-2024) Open Data DC makes available "locations and attributes of incidents reported in the ASAP (Analytical Services Application) crime report database by the District of Columbia Metropolitan Police Department (MPD)" [DC24] The crime incident data is reported by neighborhood cluster but also contains data that maps these incidents to XBlock and YBlock which contains X and Y coordinates for mapping data by block. This helps coordinate the data to more specific neighborhoods than the neighborhood clusters.

Second, we utilized shape files containing boundaries for both zip codes and neighborhood clusters to map these statistics on crime into geographical bounds. These files were last updated in 2022 and 2024 respectively. The Office of Planning defined these "neighborhood clusters" in the early 2000s, and does not provide distinct neighborhood boundaries as discussed under data limitations.

Reddit To measure public perceptions on neighborhood safety, we utilized the Reddit API and the Python Reddit API Wrapper (PRAW) to scrape data from three Washington, D.C. communities: r/washingtondc, r/washdc, and r/DCforRent. The largest of these subreddits is r/washingtondc with 329k members, then r/washdc with 49k, and finally r/DCforRent with 5.9k We chose to query multiple subreddits for simple breadth of data.

Reddit's API does not allow for querying of all comments across a subreddit, but rather of all threads or all comments within a single thread. For this reason, our data scraping occurred in two steps. First, we performed a basic query across all threads within a subreddit for any thread that mentioned at least one D.C. Neighborhood and one word indicating crime or safety such as "crime", "safe", or "safety". Second, we collected all the comments from each of those identified threads, as well as the author, date, id, and other comment identifiers. In all, we collected just over 32k comments across the three communities.

Utilizing these comments, we classified each as either relevant or not relevant to our interest in neighborhood safety, and then performed sentiment analysis in order to quantify the dataset. This process is described further in Methods.

1.2 Limitations

Notably, Washington D.C. has not defined official boundaries for its neighborhoods, as neighborhoods are fluid and susceptible to change over time. Instead, the district's Office of Planning defined a set of "neighborhood clusters" with distinct boundaries for data analysis. For this reason, we were unable to source neighborhood boundary data from Open Data DC.

However, we argue that residents of the District do not have a strong conception of these neighborhood clusters, and instead refer to neighborhoods individually in their online discussions, and subsequently in the data we collected from Reddit. For this reason, we have opted to perform our analysis on individual neighborhoods, and utilize the boundaries provided by Zillow in their neighborhood database. This database appears to have been last updated in 2017, meaning that the neighborhood boundaries may not be representative of how residents truly conceptualize them today. This would certainly account for a measure of error in our data.

2 METHODS

2.1 Classification

Due to the fact that we could not scrape comments from Reddit across threads, and instead had to search within them, the preliminary dataset on public sentiment contained many comments unrelated to neighborhood safety. To identify usable data points from this set, we made us of OpenAI's API and large language models. This process took three steps

Test Set creation We began by manually coding a small set of testing data as either relating to public safety (code 1) or not relating to public safety (code 0). This test set ultimately contained 130 datapoints, approxomately half of which were coded 1 and half coded 0. 52 of these points were from r/washingtondc, 50 from r/washdc, and 28 from r/DCforRent.

Prompt Engineering This same set of comments was then used to test a variation of prompts sent to OpenAI's GPT-4o mini model. GPT's response was then compared with our manual codes to determine the accuracy of GPT's classification. The final prompt was "Is this text about neighborhood safety in DC? Answer only with a number: 1 if about safety in DC, 0 if not. The test set was coded by OpenAI's model with 82.5 percent accuracy, which we deemed successful for a relatively small dataset.

Final Classification This final version of the prompt was eventually combined with an engineered sentiment analysis prompt to classify all data points in our original scraped dataset. Ultimately, 6445 data points were coded as safety related, and would form the beginning of our final dataset.

2.2 Sentiment Analysis

After each comment was coded as either 1 (related to crime in D.C.) or 0 (not related to crime in D.C.), we turned again to OpenAI's LLM, Chat GPT, to assist in analyzing the sentiment of each crime comment. As in the previous step, working with Chat GPT required us to first engineer a prompt. After prompt engineering, we were able to run our related-to-crime dataset through the Chat GPT API in batches of 10,000 comments.

Prompt Engineering Prior to running the classification and sentiment analysis on all 30,000+ Reddit comments through OpenAI's API, we worked with a set of 10 Reddit comments and OpenAI's GPT-4o mini model on a web browser. This allowed us to try a few different iterations of the classification and sentiment prompts to produce the most optimized results. The final engineered prompt asked the GPT model "Is this text about neighborhood safety in DC? Answer only with a number: 1 if about safety in DC, 0 if not. If the answer is 1, how safe do you think the author feels? Answer with only a number on a scale between 1 (very unsafe) and 10 (very safe) If you are answering both questions, please split the numbers with only a comma. Here is the text: ". As in the classification step, the text attached to the end of the prompt is the comment from Reddit.

Final Classification with Sentiment Analysis The final prompt listed above was applied to every comment we scraped from Reddit and returned a list of either a zero or a 1 and a number from 1 to 10 separated by a comma. We were able to merge this list into our dataset with the Reddit comments, remove all 0 values, separate the 1 and the sentiment, and add a column called sentiment score. This partially cleaned csv file provided the base for our analysis.

2.3 Geospacial Mapping

In order to aggregate crimes across D.C. neighborhoods, we leveraged the geopandas library in python, which allows for spacial joining. Each record in our crime data had an associated longitude and latitude indicating where exactly the crime took place. We then converted these longitudinal points into geographies with geopandas, and preformed a spacial join with the neighborhood boundaries distributed by Zillow. This allowed us to identify the neighborhood where each crime took place and then aggregate.

We additionally created a mapping of neighborhood cluster to zip code. Unfortunately, neighborhoods and neighborhood cluster boundaries do not align cleanly, resulting in neighborhood clusters that fall within more than one zip code. To map each cluster to a single zip code, we used OpenDataDC shape files to generate the intersection of each zip code and each neighborhood cluster using OpenDataDC shape files. From there, we identified for each neighborhood cluster the zip code with the largest overlap. This allowed us to map clusters to zip code. Even with this method, zip codes often contain several neighborhood clusters.

2.4 Data Cleaning and Merging

Reddit Neighborhood Identification Using our 6445 safety related data points scraped from Reddit, we then identified the neighborhood associated with each comment. This was done in two steps. First, we parsed the comment text, and searched for any neighborhood name or abbreviation from a a predefined list. If the comment text contained the name of a neighborhood, we assigned that comment to the mentioned neighborhood.

From there, we found that a subset of our comments contained no reference to any neighborhood. To identify a neighborhood for these comments, we matched the comment with its parent thread, and searched the title and thread text for neighborhoods in the same manor. In this way, comments and their sentiment scores were assigned to neighborhoods based on their parent thread. Ultimately, we were left with 4067 comments matched with neighborhoods for our final analysis.

This matching process certainly leaves room for error in neighborhood identification, particularly for comments that had to be matched based on their parent thread. The nature of Reddit is that comments may have been referring to an entirely different neighborhood than the original thread, based on some prior comment in the conversation. Reddit's API does not have built in functionality for identifying these chains of conversations. However, based on the distribution of sentiment scores across neighborhoods (shown in Figure 3) and our prior knowledge of neighborhood stereotypes, we are confident that sentiments are adequately matched to neighborhoods for analysis

Final Dataset After acquiring and performing spacial analysis on the three sources of data, we merged each of the datasets together, resulting in a final dataset with crime rates, home values, and safety sentiments across both neighborhoods and years. Below is an example of the final merged data:

Neighborhood	Year	Average Safety Sentiment	Crime Rate	Median Home Value
Shaw	2024	5.66	1394	786305.343135
Columbia Heights	2024	4.83	1545	638632.690910
Capital Hill	2024	5.83	1138	907238.986881
Adams Morgan	2024	487	686	581093.788565

We then used this dataset to make several comparisons, as described in our Analysis. To determine how true crime vs. perceived crime potentially impacted home values, we looked at average sentiment in a neighborhood from 2019 to 2024 and compared it with the ZHVI for those years. We also compared the trends of true crime with the trends in the average sentiment about crime in a particular neighborhood from 2019 to 2024. Finally, we ran a regression to determine whether sentiment about crime, true crime or both predicted changes in housing prices while controlling for neighborhood.

3 ANALYSIS

Our primary research questions seeks to uncover any relationship between perceived crime and the change in home value in D.C. neighborhoods. For the purposes of descriptive analysis, we will focus on two neighborhoods: Shaw and Columbia Heights. Through our analysis, we will show that there is a relationship between perceived safety and housing prices but the strength of that relationship still needs further exploration, which we discuss in the Limitations and Next Steps section.

3.1 Housing Trends by Neighborhood

To get a basic understanding of housing trends in neighborhoods across Washington D.C. we graphed the ZHVI (Median Home Value) for the top 8 neighborhoods mentioned in the Reddit analysis and Anacostia. Figure 1 shows general home value trends from January 2019 through November 2024.

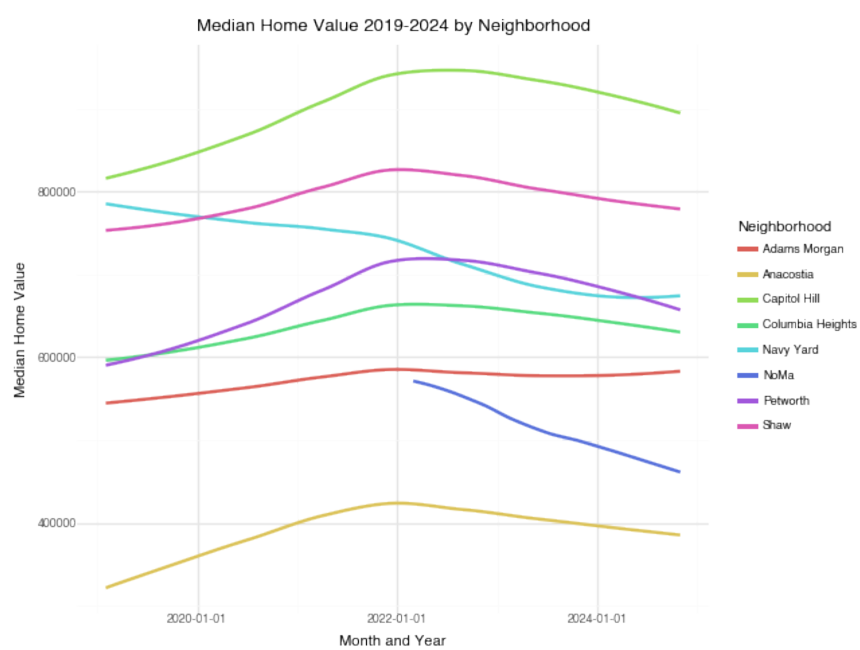

Figure 1. ZHVI by Month and Year for Different Neighborhoods

Unsurprisingly, the end of 2021 and early part of 2022 had the highest home values for most of the neighborhoods. This increase occurred nationally as home values soared in 2021 due to lower interest rates. Some houses sold in just hours. We realize that the housing market is not impacted by one factor alone and wanted to begin our analysis with this in mind.

Shaw (pink line) and Columbia Heights (dark green line) have nearly identically shaped trend lines over the five year period. Home values in Shaw are consistently approximately 200,000 dollars more than those in Columbia Heights. This sets the baseline for reviewing how perceptions on safety may impact home values.

3.2 Crime Trends by Neighborhood

There are many factors that impact the public opinion on crime in an area, including actual crimes committed. To get an idea of the actual incidents of crime in each of our neighborhoods of interest, we took the crime data from OpenDataDC and graphed them by offense for 2019 through 2024. Figure 2 shows these results.

According to the data, Shaw and Columbia Heights consistently have some of the highest total crime incidents across D.C. This should indicate that these neighborhoods have some of the lowest sentiments on safety. However, as we discuss, this is not the case. Further, if true crime predicted home values, these two neighborhoods should have similar ZHVIs. As we noted before, their trend lines take the same shape, but home prices in Shaw are higher overall.

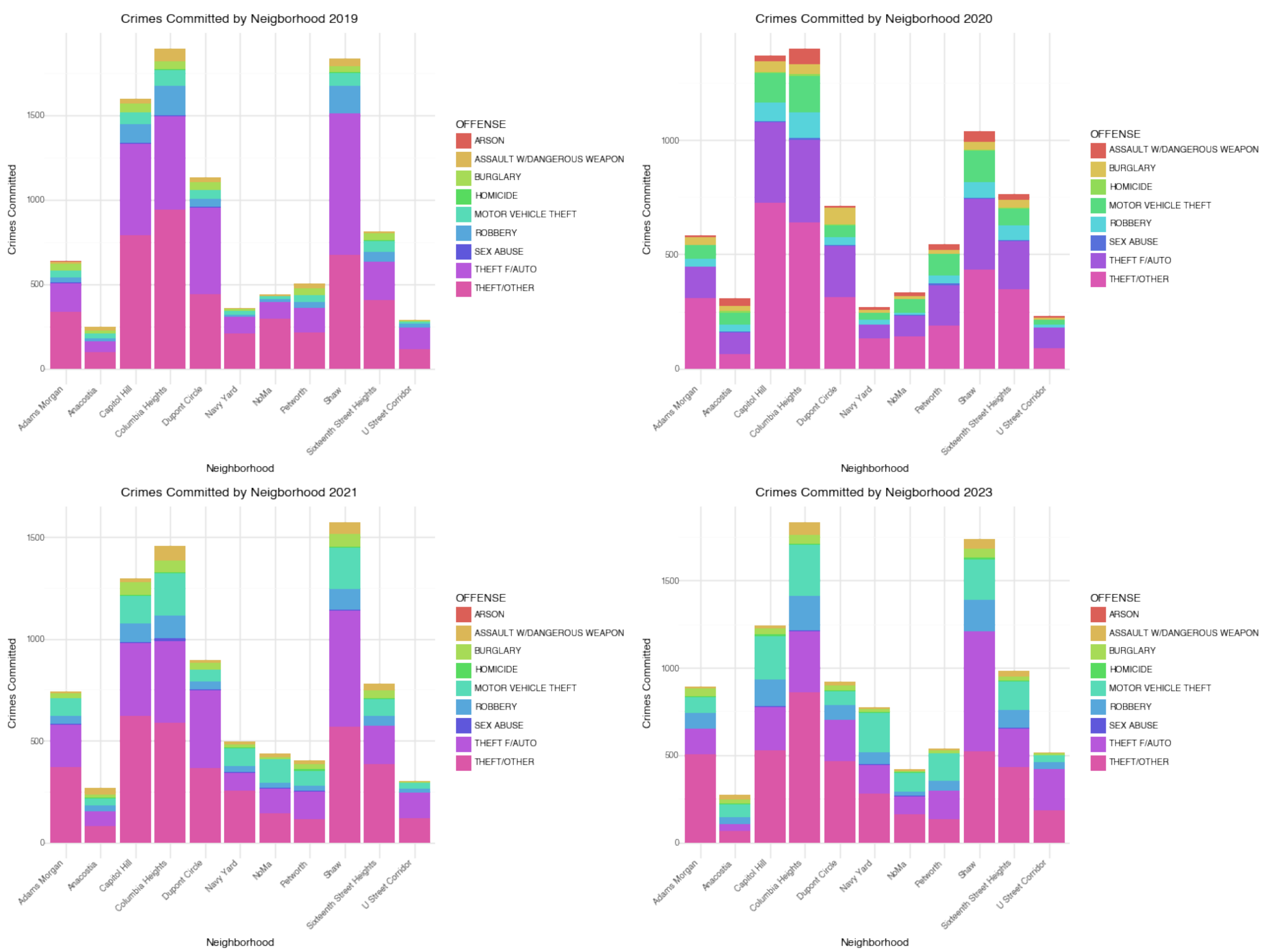

Figure 2. Crime Incidents by Year and Neighborhood (2019-2023)

3.3 Sentiment by Neighborhood

Figure 3 shows the distribution of safety sentiment scores generated from our Reddit data across the top 15 mentioned neighborhoods. Note that the number of data points varies between neighborhoods, and these averages span all years of data collected

Unsurprisingly, average sentiments cluster around a value of 5, the mid-point of our sentiment metric. However, we do see that neighborhoods such as Adams Morgan and Dupont have a range that reaches much higher in sentiment, where Anacostia's range reaches much lower. We posit that these ranges, more so than average scores, represent general online sentiments of safety, due to the tendency of online spaces to represent extreme opinions in both directions, resulting in a normalized mean.

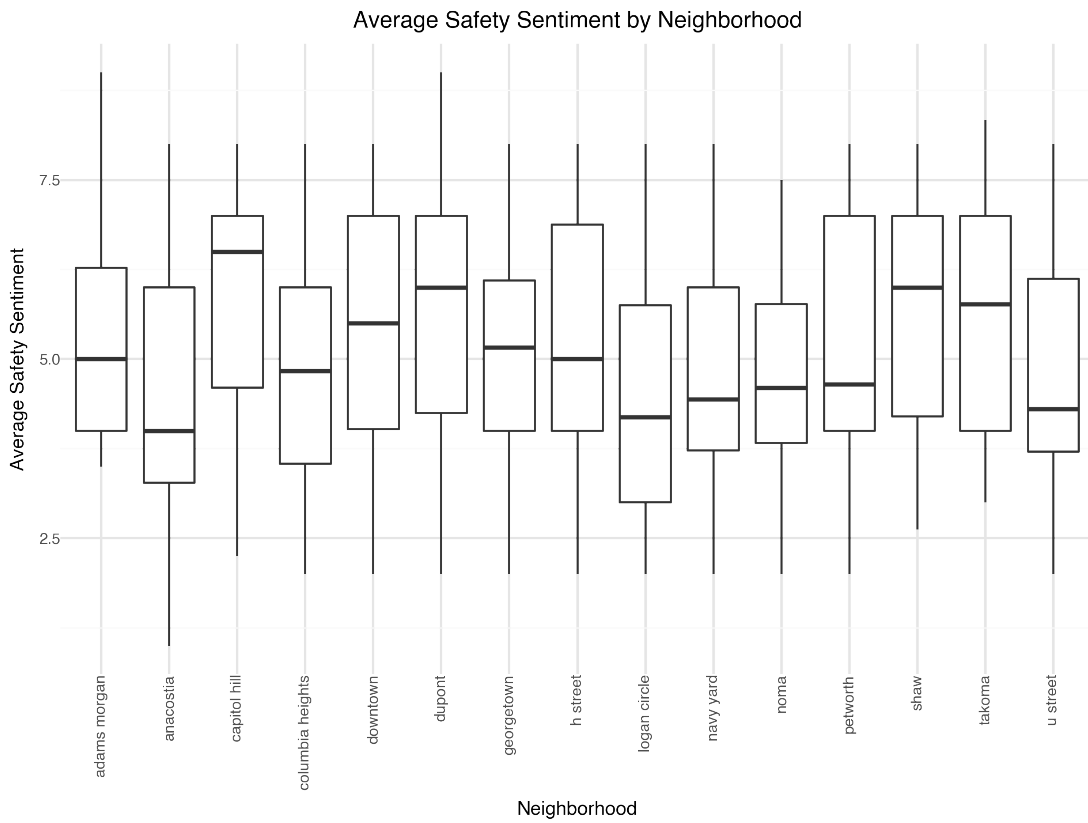

Figure 3. Reddit Safety Sentiments Across Top 15 Mentioned Neighborhoods

3.4 Perceived Crime Trends vs Actual Crime Trends

As previously mentioned, perceived crime and actual instances of crime in a neighborhood can often be out of sync. Below are graphs that show the trends in average sentiment on safety in a neighborhood on a scale of 1 to 10 (1 is very unsafe and 10 is very safe) and trends in true crime incidents in that neighborhood (scaled down by 1,000). We analyzed these discrepancies in Shaw and Columbia Heights.

The left-hand graph in Figure 4 shows the trends in perceived and actual crime in Shaw. The trends in average sentiment change quite a bit year over year but the overall crime incidents do not change much. The largest dip in crime for Shaw is in 2020 which could also be attributed to the national drop in crime and policing during the COVID-19 pandemic. Generally, crime perception is right around 6 which indicates that people feel neutral to slightly safe in Shaw. While the crime incidents stay around 1,800 for all years, the sentiment certainly fluctuates and there does seem to be a bit of a lag between true crime incidents and changes in perceived crime.

The right-hand graph in Figure 4 shows the trends in perceived and actual crime in Columbia Heights. The trends in average sentiment are less volatile than Shaw but are lingering right around 4. This indicates that people's opinion on safety in Columbia Heights is neutral to slightly unsafe. True crime, however, drops in 2020 as in Shaw but stays low for a few years until a spike in 2023. Sentiment on crime lags slightly behind this dip as sentiment begins rising in 2023 despite the increase in overall crime incidents that year.

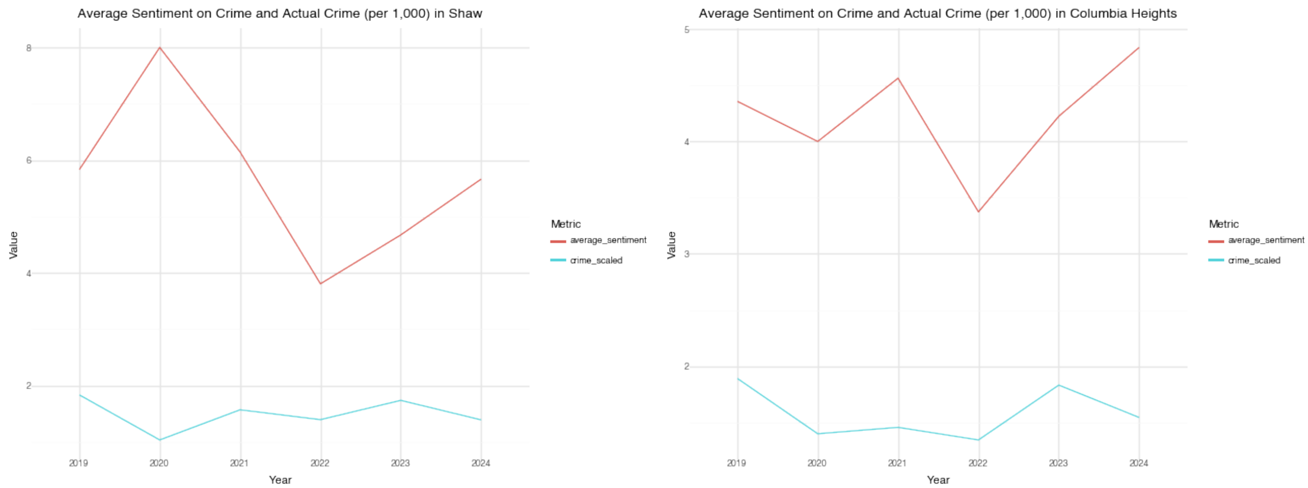

Figure 4. Perceived Crime Trends vs Actual Crime Trends (2019-2024)

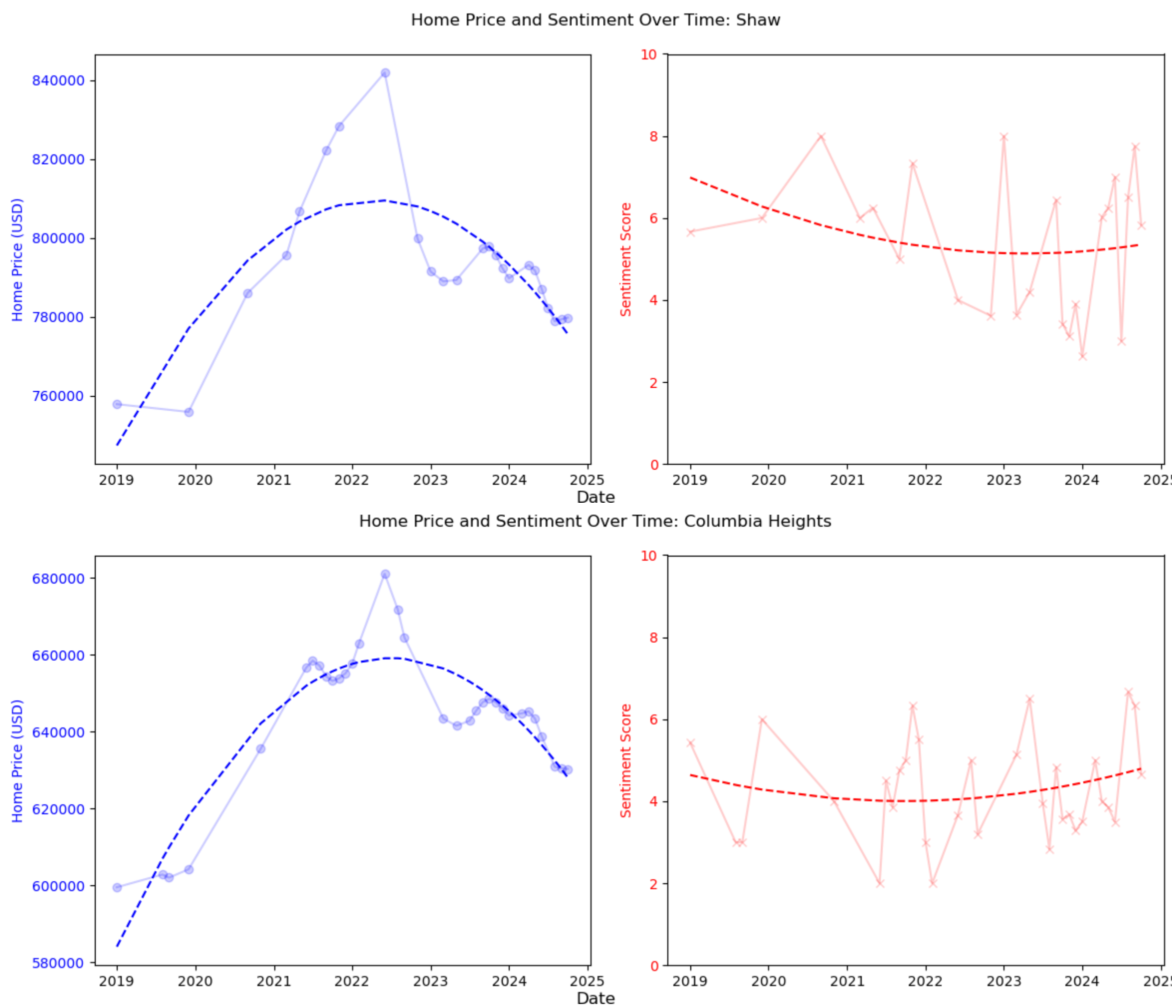

Figure 5. Housing Trends vs. Sentiment Trends (2019-2024)

3.5 Housing Trends vs. Sentiment

Figure 5 compares online safety sentiment and housing price trends over time for both Shaw and Columbia Heights. Trends lines are generated with a second degree polynomial in an effort to avoid over-fitting.

As previously noted, the two neighborhoods exhibit similar trends in price across the 5 years, peaking between 2022 and 2023, though Columbia Heights is on average, less expensive. We can see that the two neighborhoods also exhibit somewhat similar trends in sentiment, both decreasing slightly and then upturning slightly, though Shaw's decrease is more drastic.

We would expect that increasing housing prices are correlated with decreasing crime rates, which would then spur an increase in perceptions of safety. These figures provide some indication that perceptions of safety may lag behind an housing prices, with prices peaking in the early 2020s, and sentiments only just starting to increase for both Shaw and Columbia Heights. This makes conceptual sense, as people take time to adjust their perceptions as crime rates change. It is noteworthy, however, that the data underlying the sentiment trend lines is widely variable, and does not in itself fit the trends well, which speaks to the need for more data on public sentiments in general.

3.6 Putting it All Together

To go beyond descriptive analysis, we created a lag OLS regression model that took average ZHVI as a function of average sentiment on crime, actual crime, lagged average ZHVI (lagged by 1 year) and controlled for neighborhood. Below is the equation of the regression:

$$y = \beta_0 + \beta_1 AvgHomeValue(-1) + \beta_2 AverageSentiment + \beta_3 TotalNeighborhoodCrime + \beta_4 Neighborhood + \varepsilon$$

where neighborhood is a categorical variable that uses Adams Morgan as the reference category.

For ease of comparison, all variables are standardized. Therefore, a one standard deviation increase in average median home value is associated with a 0.1394 (β_2) standard deviation decrease in average sentiment holding everything else constant. This value is statistically significant with a t-statistic of -4.542 and a p-value of 0. While this model may have low power, we still argue that the statistical significance makes this finding important.

Dep. Variable:	MedianHomeValue_avg	**R-squared:**	0.986
Model:	OLS	**Adj. R-squared:**	0.980
Method:	Least Squares	**F-statistic:**	151.9
Date:	Tue, 17 Dec 2024	**Prob (F-statistic):**	2.89e-17
Time:	14:24:19	**Log-Likelihood:**	23.318
No. Observations:	32	**AIC:**	-24.64
Df Residuals:	21	**BIC:**	-8.512
Df Model:	10		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
const	8.327e-16	0.025	3.27e-14	1.000	-0.053	0.053
AvgHomeValue_lagged	0.9241	0.194	4.754	0.000	0.520	1.328
average_sentiment	-0.1394	0.031	-4.542	0.000	-0.203	-0.076
TotalNeighborhoodCrime	-0.1590	0.089	-1.793	0.087	-0.343	0.025
neighborhood_anacostia	-0.0567	0.079	-0.716	0.482	-0.221	0.108
neighborhood_capitol hill	0.1868	0.191	0.980	0.338	-0.210	0.583
neighborhood_columbia heights	0.0755	0.082	0.918	0.369	-0.096	0.247
neighborhood_navy yard	0.0937	0.062	1.506	0.147	-0.036	0.223
neighborhood_noma	0.0348	0.048	0.723	0.478	-0.065	0.135
neighborhood_petworth	-0.0004	0.058	-0.007	0.994	-0.120	0.120
neighborhood_shaw	0.1879	0.144	1.301	0.207	-0.112	0.488

Omnibus:	1.322	**Durbin-Watson:**	1.842
Prob(Omnibus):	0.516	**Jarque-Bera (JB):**	0.876
Skew:	-0.405	**Prob(JB):**	0.645
Kurtosis:	2.955	**Cond. No.**	21.6

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Figure 6. Regression Results on the Effect of Perceived Crime on Housing Price

The other independent variable of interest, total neighborhood crime, reveals that a one standard deviation increase in average median home value is associated with a 0.1590 (β_3) standard deviation decrease in crime incidents holding all else constant. This value is just barely not significant at an (α) of 0.05 with a t-statistic of -1.793 and a p-value of 0.087. However, due to the power of this model there is a chance of type II error (false negative) at this level of significance.

Overall, these results were a bit surprising. Our hypothesis stated that an increase in public sentiment on safety in a neighborhood would increase the prices of homes in that neighborhood, a positive correlation. Even accounting for a one year lag in home prices, we still observed a decrease in sentiment causing an increase in home value, a negative correlation. Further, our model suggests that actual crime values have no significant effect on housing prices in D.C. neighborhoods.

While the results were surprising, there are likely some factors that help explain this. First, the final sample size for the OLS model was 32, this is very small and can be an issue in determining true causal relationship. Second, we are not controlling for other factors in the housing market such as interest rates, the COVID-19 pandemic, and the number of houses on the market in that neighborhood. These are all likely sources of endogeneity we did not have the data to control for, resulting in a biased model. Finally, there may be heterogeneous treatment effects. Because no neighborhood is the same in D.C., our model makes it difficult to see if different neighborhoods are effected by sentiment differently. While the results of this model are preliminarily informative, there are serious upgrades that need to be made before making any final conclusions.

4 DISCUSSION

4.1 Conclusion

Our descriptive analysis helped us visualize the shifts in housing prices by neighborhood, the number of crime incidents in each neighborhood, and some of the ways we can compare them. By looking at true crime trends on the same graph as trends in average sentiment for a given neighborhood, we were able to note the likelihood of a lag between crimes happening and the public's feelings on crime in a neighborhood. A comparison of housing prices and sentiments over time supports this conclusion. We were also able to draw comparisons between two neighborhoods—Shaw and Columbia Heights—that were similar in housing trends and total crime incidents but different in the public's perception. While these results did not quantify a specific relationship between crime perceptions and housing prices, we were able to visualize any anomalies or similarities.

Our preliminary regression helped us attempt to quantify the potential relationship between perceived crime, true crime incidents, and housing prices controlling for D.C. neighborhoods. While our results were not what we expected, the standardized model with a lag on yearly average ZHVI revealed a statistically significant negative relationship between feelings on safety and average ZHVI. However, we believe the model to be underpowered due to a small sample size, and containing many sources of endogeneity, so these results should be considered with caution.

4.2 Limitations

One of the main limitations in our results were the number of observations used to build our OLS model. Due to inconsistencies in how often public opinions are shared on specific neighborhoods, we needed to work with average sentiment values across years. For consistency we then studied the average ZHVI for every neighborhood each year. This left us with 32 total observations, which calls into question the power of our model. Although sentiment is statistically significant in predicting future home value, we would like to get more data to support this.

As already mentioned, the process for collecting and processing Reddit data for sentiment analysis could also be improved. As opposed to searching threads and collecting all comments, we believe the ability to search all comments in a subreddit would improve the relevance of the original collected data. Additionally, we would like to improve the process for identifying neighborhoods associated with each comment, and mitigate the chance that comment sentiments are attributed to incorrect neighborhoods.

Finally, as the District does not define official neighborhood boundaries, spacial mapping between zip codes, neighborhood clusters, and neighborhood likely contains some error. We are also limited to analysis based on the units provided by Zillow for housing prices, such as average rent data, which is aggregated by zip code as opposed to neighborhood. This prevents us from achieving workable granularity in our data.

4.3 Next Steps

Other sources of public sentiment data that we would be interested in collecting are Google Reviews, news articles, and Yelp. Google and Yelp Reviews put a different lens on sentiments on neighborhood safety. A person may review a business or restaurant based on location just as much as food or service. In collecting some of this data, we are able to lean on a different subset on public sentiments on safety. News articles help us see how public sentiment might be shaped. If one neighborhood's crime gets reported on more often than another, despite similar crime rates, this could shape public opinion. By widening the scope of public sentiment data, can collect more robust information.

Another area we would be interested in exploring more deeply are the impacts of crime and sentiment on rental prices using ZORI. We began some analysis on this, but found some limitations in comparing zip codes to neighborhoods. Working to get this data cleaned and ready to analyze would be a good next step, especially since many people who actually live in D.C. are renters.

Thirdly, diving into data on AirBnB rentals in D.C. would be another interesting expansion. By exploring the number of rentals in a neighborhood and tracking the average price to rent an AirBnB in that neighborhood, we could possibly see how public sentiment changes where people choose to stay on vacations. There are many ways to track public sentiment on safety and its impact on housing and we think it would be valuable to explore more of them.

Finally, we would like to explore this relationship in other cities. D.C. is a bit of an anomaly in that it is not a state, but it is more than just a city. To ensure our results are reproducible and interesting, we would like to apply this model to other U.S. cities like Chicago, Boston, and Seattle.

REFERENCES

[Law13] Will Lawrence. "Does Gentrification Lower Crime: A Look at Granger Causality in Washington, D.C." In: *The Park Place Economist* 21 (2013).

[DC24] Open Data DC. *Open Data for Washington, D.C.* 2024. URL: https://opendata.dc.gov/.

[Pog24] David Pogue. "Reality Check: What We Get Wrong About Crime". In: *CBS News* (2024).

[Res24] Zillow Research. *Zillow Research Data*. 2024. URL: https://www.zillow.com/research/data/.