Name: Shruti Rao
Email: raoshrut@usc.edu
USC ID: 6339285973
Course: DSCI 510 - Principles of Programming for Data Science

Analysis of Safe Housing For USC Students

Project Description

As a USC student, I know how difficult it can be to find safe and affordable housing near campus. When I started looking for off campus apartments, I realized there wasn't an easy way to compare neighborhoods based on both rent prices and safety. Some areas seemed affordable but had high crime rates, while others were safer but way too expensive. This project aims to help USC students make better housing decisions by analyzing the relationship between rental prices and crime rates in neighborhoods within a 3 mile radius of campus.My main goal was to answer the question: Are safer neighborhoods more expensive, and which areas offer the best value for USC students?

Data

Source 1: Craigslist apartment listings
Method: Web scraping using Python (requests and BeautifulSoup libraries)
Data Collected: 1,276 apartment listings from craigslist within a 3 mile radius of USC
Info extracted from craigslist: property name, rent price, number of bedrooms, location/zipcode, and listing URL
dataset: After cleaning and filtering to USC area zip codes, I had 532 usable listings across 9 zip codes

Source 2: LA city crime data
Method: Downloaded csvfile from LA Open Data portal (https://catalog.data.gov/dataset/crime-data-from-2020-to-present)
Data Collected: 816,028 crime records from 2020 to present
time period used: Filtered to crimes from 2022-2024 (last 3 years)
geographic filter: Focused on four police areas near USC (Central, Southwest, Olympic, Rampart)
Final dataset: 95,440 crime incidents, expanded to 425,466 records after mapping police areas to specific zip codes

I chose these data sources because I wanted my analysis to reflect the actual housing decisions students have to make in real life. Using Craigslist apartment listings allowed me to work with current rental prices that students actually see when searching for off campus housing, rather than relying on broad estimates or averages. I used the Los Angeles crime dataset to get a

measure of neighborhood safety instead of depending on personal perceptions, which is why I felt that some other sources I had considered such as Reddit would not be a reliable source.. Combining these two datasets made it possible for me to directly analyze the relationship between affordability and safety.

Data Cleaning, Analysis & Visualization

Data Cleaning

I began by cleaning the apartment and crime datasets so they could be accurately compared. The craigslist apartment data required significant cleaning because listing formats were inconsistent, so I standardized rent prices, extracted bedroom counts from listing titles, calculated rent per bedroom, and filtered listings to only zip codes within a 3-mile radius of USC. For the crime data, I filtered records to the most recent three years and limited the dataset to police areas surrounding USC. Since the crime data was reported by police areas rather than zip codes, I created a mapping between police areas and usc area zip codes so both datasets could be analyzed at the same geographic level.

Analysis

I sorted apartment listings by zip code to calculate average rent and average rent per bedroom for each neighborhood. I also used crime incidents by zip code to measure neighborhood safety over the same time period. To make comparisons fair across zip codes with different numbers of listings, I created a normalized crime metric by calculating crimes per listing and then analyzed the relationship between rent and crime by computing the correlation coefficient and by creating a combined best value score that balanced affordability and safety.

Visualization

I chose to use visualizations that highlighted key patterns in the data. A scatter plot of average rent versus crime rate helped visualize the overall relationship between affordability and safety across neighborhoods. Bar charts were used to compare crime rates and average rents by zip code, making it easy to identify the safest and most affordable areas. I used a chart highlighting the top five best neighborhoods, which summarized the results of the analysis in a way that is easy for students to interpret and apply to their own housing decisions.

Hypothesis and Conclusion

At the start of this project, my hypothesis was that neighborhoods closer to USC with lower crime rates would be significantly more expensive, making it difficult for students to find housing that is both affordable and safe. Based on personal experience and common perceptions

among students, I expected there to be a strong trade off between rent prices and neighborhood safety.

After analyzing the data, I found that while there is a relationship between rent and crime, it is not as extreme as I initially assumed. The correlation analysis showed a moderate negative correlation between average rent and crime rate, indicating that higher rent areas tend to have lower crime, but not always. The scatter plot visualization reinforced this finding by showing several neighborhoods that fall into a middle range, offering reasonable rent prices without extremely high crime rates.

Combining the rent and crime data showed that some neighborhoods provide a better balance of affordability and safety than others. Westlake (90057) and Koreatown (90006) stood out in the visualizations as offering strong overall value, while some higher priced neighborhoods were not necessarily the safest. Overall, the analysis and visualizations show that although safer neighborhoods often cost more, USC students do not have to choose strictly between affordability and safety, and data can be used to identify neighborhoods that offer both.

Changes to original Proposal:

I initially planned to use the US Census data for housing information, but I switched to scraping Craigslist because it provided real rental prices that students actually see when searching for housing. Another challenge was that crime data was reported by police areas while apartment data used zip codes, so I created a mapping between the two to make the datasets comparable. I also narrowed the project scope by focusing on overall crime rates instead of crime types or when those crimes took place during the day. Other than my sources, the goal of this project remains the same.

Mention of Future Work

If I had more time and resources, I would expand this project by incorporating additional factors such as proximity to campus, access to public transportation, and nearby amenities to create a more comprehensive housing score as these are all factors that students take into consideration when choosing between housing options.