DSCI 510: Principles of Programming for Data Science

Name: Yu-Chen Lu
USC ID: 3952-1526-52
GitHub: j12321

Final Project Report

1. Introduction

This project, **"Grocery Store Affluence Map: Correlating Store Locations with LA Property Prices"** examines whether the locations of certain grocery stores are linked to rental costs in Los Angeles. It focuses on five grocery store chains: Erewhon, Whole Foods, Trader Joe's, Ralphs, and Food 4 Less.

Grocery stores often target specific groups of people. For example, Erewhon is known for its upscale organic products, which might appeal to wealthier customers. The goal of this project is to see if the type of grocery store in an area is reflected in local rental prices. To analyze this, the project uses zip codes to define locations and focuses only on Los Angeles properties and stores.

2. Data Collection

Data was collected from the following resources:

I. Apartments.com (Web Scraping)
- Focused on the one-bedroom listings in Los Angeles.
- Extracted details such as property name, price range, zip code.

II. Grocery Store Websites (Web Scraping)
- Erewhon, Whole Foods, Trader Joe's, Ralphs
- Extracted store names and addresses.

III. LA County Zip Codes
- Download a GeoJSON file for LA zip code boundaries.

IV. US Zip Codes with Latitude and Longitude
- Download a CSV file with coordinates for zip codes.

The approach of data collecting:

I. Crawlbase CrawlingAPI
II. Beautiful Soup
III. Selenium

Some challenges during web scraping:

I. Erewhon, Whole Foods, and Ralphs websites load content using JavaScript, which cannot be captured with BeautifulSoup alone. To handle this, use Selenium to simulate webpage interactions.

II. Whole Foods' website does not update its URL when searching locations in LA, so interactions with the site were simulated to extract location data.

The dataset was collected from seven sources and includes 680 rental listings, 12 Erewhon stores, 7 Whole Foods stores, 11 Trader Joe's stores, and 17 Ralphs and Food 4 Less stores. Initially, the plan was to analyze all types of properties, but the focus was narrowed to one-bedroom listings from Apartments.com. This decision simplifies the analysis, ensures the data is more comparable, and makes the results clearer and easier to interpret.

3. Data Cleaning

Rental Data (Apartments.com):

- Extracted the minimum and maximum prices from each price range.
- Removed listings without a specific price (e.g., "Call for Rent") or a clear price range (single-price listings).
- Converted prices from string format (e.g., "$1,200") to integers.

Grocery Store Data:

- Extracted the zip code from each store's address (usually found at the end of the address).
- Ralphs' website includes Food 4 Less locations. Separated them based on the webpage structure (Food 4 Less has an extra link).

Zip Code Coordinates:

- Filtered out any zip codes that are not within Los Angeles.

4. Data Analysis

Methodology:

I. Used the Geopy package to calculate distances between zip codes.
II. For each grocery store location, identified apartment listings within 5 miles.
III. Calculated the average minimum and maximum rent prices of all listings within 5 miles.
IV. If no listings were found within the radius, the result was recorded as: "No listings around this location."

Output:

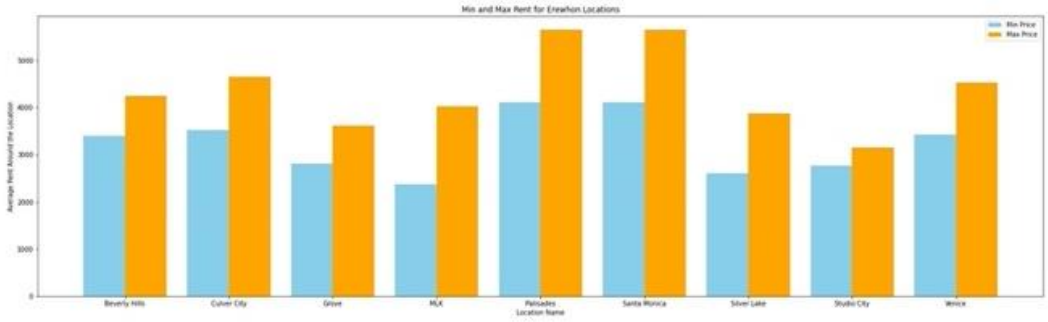
The analysis produced a list of average minimum and maximum rental prices for each grocery store location.

5. Visualizations

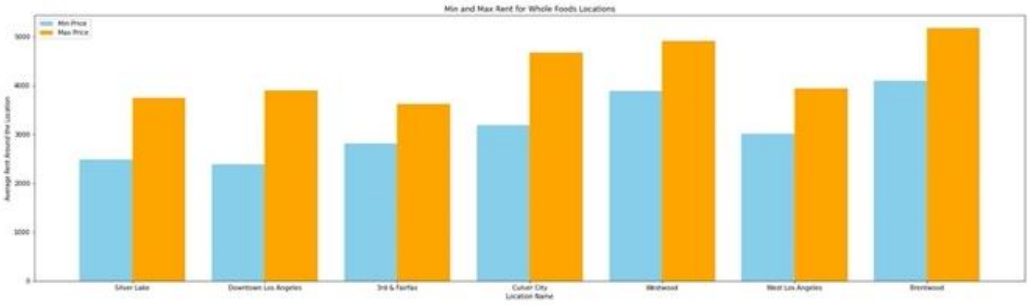
I. Bar Chart

Used bar chart to visualize the average minimum and maximum rent prices for each grocery store location. Separate charts were created for each chain:

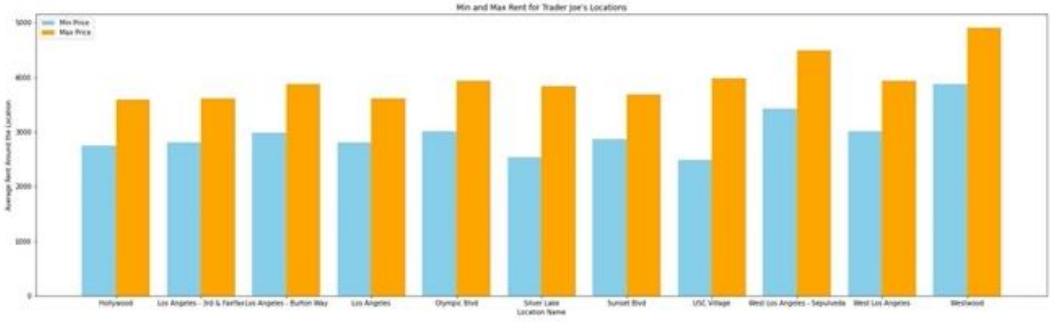
- Erewhon



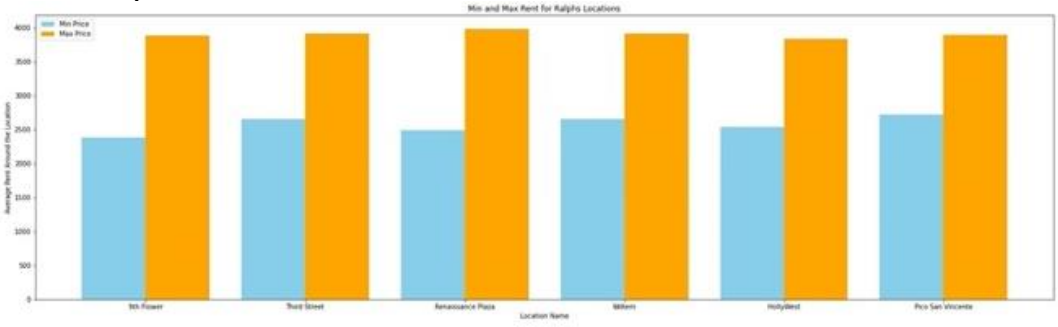
- Whole Foods:



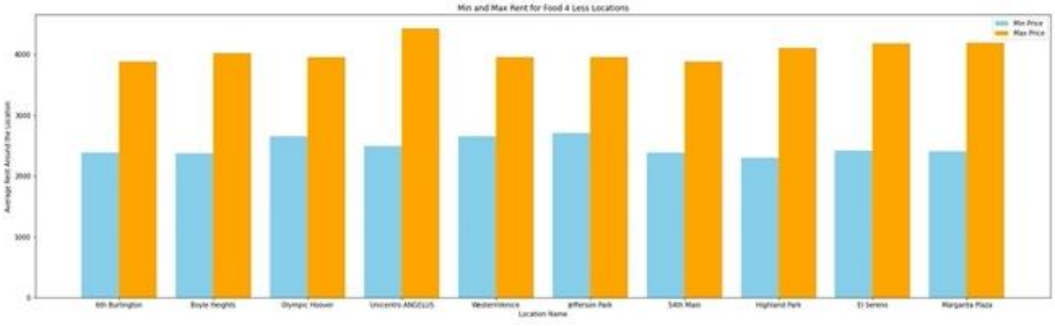
- Trader Joe's



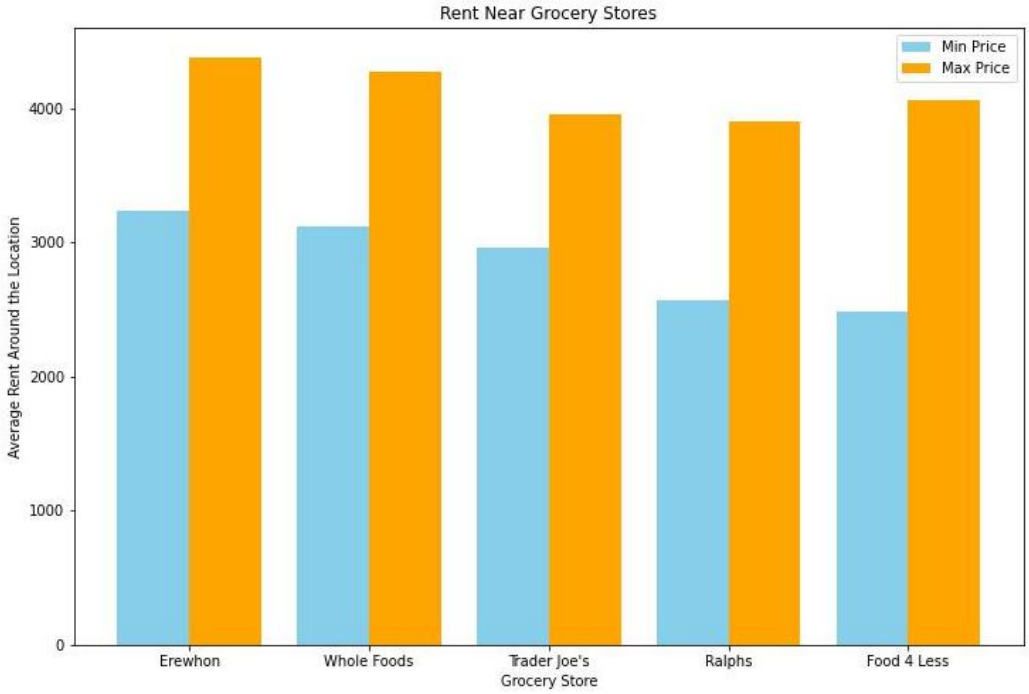
- Ralphs



- Food 4 Less



Additionally, a comparative bar chart shows the differences in rental prices across all grocery store chains.



II. Heat Map
    - Used the Plotly package to create heat maps showing the distribution of grocery stores and rental prices across Los Angeles.
    - Defined zip code boundaries using LA County Zip Codes GeoJSON file.
    - Grocery store locations were mapped based on zip code centroids using US Zip Codes CSV file. Note that the displayed locations are the centroid of zip code areas, not the exact store addresses.

6. Conclusion

The analysis reveals a clear pattern in rental prices based on proximity to different grocery store chains. For minimum rent prices, the ranking is: Erewhon > Whole Foods > Trader Joe's > Ralphs > Food 4 Less. For maximum rent prices, the ranking is: Erewhon > Whole Foods > Food 4 Less > Trader Joe's > Ralphs. These findings align with the hypothesis that grocery store locations correlate with rental costs. For instance, Erewhon stores are located in areas with the highest rental prices. Additionally, Whole Foods, known for its higher retail prices compared to Trader Joe's, Ralphs, and Food 4 Less, is associated with higher local rents.

7. Future Work

If given more time, I would expand the dataset to include more rental data. Currently, Apartments.com limits the results to 18 pages (around 700 listings), while there are over 11,000 rentals in Los Angeles. This limited dataset may not fully capture the city's rental landscape. The heat maps also show that the collected listings cover only a small portion of Los Angeles. I aim to gather rental information for all areas within the city. Additionally, I would use the exact addresses of properties and grocery stores to calculate distances and plot their locations on the map, instead of relying on zip code centroids. This would provide a more accurate representation of the spatial relationships between grocery stores and rental properties.