

DSCI 510, Spring 2022
Homework 5- Final Project
Chia-Hsuan Lee

The Relationship Between Housing prices and Park Space

Project Description

This project seeks to find a relationship between housing prices and park space. What inspired me to start with this idea is a work I saw in an exhibition, *Becoming Los Angeles*, in the Natural History Museum. The work was the aerial photography of Los Angeles, capturing a disproportionate abundance of green and blue color in wealthier neighborhoods in the county. Therefore, I wonder if it is also true that the residents in wealthier neighborhoods have more park space to access.

Data Sources

1. Recreation and Parks Information

<https://data.lacity.org/Parks-Recreation/Recreation-and-Parks-Information/rwq7-yhp5>

Link to fetch json: <https://data.lacity.org/resource/rwq7-yhp5.json>

Using the “requests” library, I have accessed the json from the URL and created **park.csv** file. This dataset consists of 561 rows and 14 columns. From those columns, I have used the column “address” to extract the zip code from each row by using the “re” library, and the column “area_acres” to calculate the area for each zip code. Then, I created a dictionary with the zip codes for the keys and the sum of the area for the values.

2. California Zip Codes by Population

https://www.california-demographics.com/zip_codes_by_population

Using the “requests” and the “BeautifulSoup” libraries, I have scraped the web pages and created the second dataset called **pop.csv**. This dataset consists of 1650 rows and 3 columns, which are “Rank”, “Zipcode”, and “Population”. To obtain the value in regard to average park space per resident, I divided the sum of the area from the

dictionary (turned the unit from acre to square feet) which was created by the first dataset by the population for each zip code from the column “Population” of this dataset, (sum of the area/population). Then, I created a dictionary with the zip codes for the keys and the average park space per resident for the values.

3. Redfin House Price

<https://www.redfin.com/>

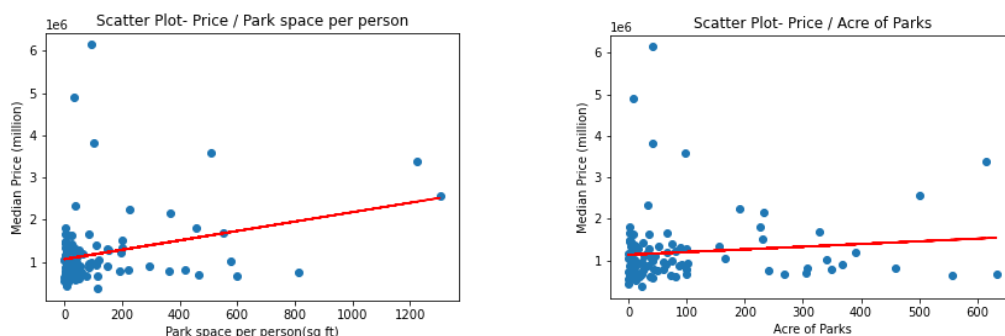
Using the “requests”, “BeautifulSoup”, and “re” libraries, I scraped the median prices of all house types for each zip code listed in the above second dictionary and created the third dataset called “price.csv”.

Using the datasets above, I created the final dataset called “final_dataset.csv”, including 4 columns, “zip code”, “median price”, “park space per person” and “Acre of parks” for further analysis.

Analysis

1. Scatter Plots

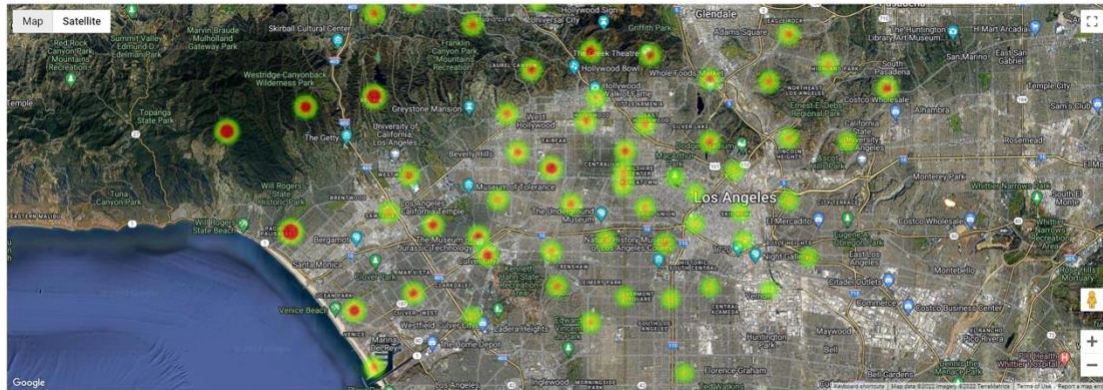
Since there are some outliers, I removed some data points that were more than 750 acres from the final dataset. To see if there is any correlation between house prices and park space, I made two scatter plots (plot1.png and plot2.png), showing the relationships between house prices and park space per capita and house prices and total park space, respectively. The red lines are the predictions for the price by using Linear Regression model.



Both plots above showed that there is no pattern to the datapoints, indicating that there is no connection between the housing prices and park space/ park space per capita.

2. Google Maps API

To examine if the greener the space, the higher the house price, I would like to visualize the relationship between the price and space. I found coordinates for each zip code of the final dataset and made a formula for weights= $\sqrt{\text{median price}/200}$. By using gmaps API, I made a heat map (export.html) where warmer colors indicated higher prices.



From the heat map above, we could see that the warmer markers show up in the greener spaces, such as northern mountains and areas to the west of the city. Therefore, the map roughly illustrates the possible relationship between house prices and green space.

Maintainability/Extensibility

The code creates the dataset of park space and house prices for each zip code in LA city. As a student from the MPPDS program, I am interested in the intersection of Public Policy and Data Science. With this dataset, one can extend the analysis to a variety of issue such as chronic disease, mental health, and the homeless. For example, if you want to know if park space can improve chronic disease or mental health, you can use this dataset and conduct an analysis with data provided by CDC. You can also use this dataset to see if the homeless distribution is related to park space.

However, there are some issue about maintainability with the code needed to be noticed. I found a few zip codes on Redfin lack for the median price, and I skipped those zip codes which may lead to losing important information sometimes. In addition, the HTML of Redfin may change. I revised the code of scraping from the website once since the HTML changed, since the previous code could not find the

median price from the website successfully, which failed to generate the final dataset. Moreover, some of the populations are calculated by combining several zip codes in the data source and I used just one of them to represent the neighborhood, which may lead to losing important information as well.

Conclusion

In this project, the scatter plots in the analysis shows no correlation between park space and the house prices. The dataset I collected covers only parks owned by the Los Angeles City Recreation and Parks Department, which limits the scope of the analysis. On the other hand, the heatmap visualizes the relationship between home prices and green space by showing home prices for each zip code on a map using color markers. We can still find evidence that there may be a relationship between house prices and green space. I suggest that datasets about parks or green space can be collected from more sources for a more complete and accurate analysis.

As the environment changes, residents pay more attention to air quality, and park space becomes a vital indicator of life. I used the final dataset to build a user interface. Users can enter any zip code in LA city, and they can get information about the size of the parks and park space per capita based on the entered zip code.