

Comparing Park Locations Across Cities

Mark Rossitto

December 11th, 2020

1.1 Introduction/Problem

After the initial outbreak of COVID-19, Many people have become housebound to a new extent. For many, just getting outside to on a daily basis is a challenge, causing adverse affects on the health of those affected. Part of this challenge in itself is just remaining safe while getting outdoors and risking possible exposure which complicated the usual means of exercise through the gym due to likely coming into close proximity being indoors. This makes parks a viable and safe alternative to those who would frequent the gym during this time.

1.2 Interest

For my Capstone project, I decided to look further into the the two cities of San Fran and New York in order to get a better idea on the parks available in each city. Many people are, when looking to move into a city from a small town or what have you, like to look at the different possible environments they could visit if choosing to move to this new city. In comparing these two cities, my project would give people a better idea as to which city offers the most for people who love the outdoors while still needing to call a city a home. In offering this, it may possibly help people who were previously on the fence about the decision who have a strong fixation on outdoor places they could visit close to home.

2.1 Data

As for the data used in my project, I relied heavily on leveraging the data returned by the Foursquare API for locational data. In using this, I was able to designate center points, venue/neighborhood locations, as well as set usable and realistic range/radius

to keep such places within a controlled and comparable environment. Additionally, I used several other libraries such as Folium for the mapping section of my work and a combination of others for the data manipulation, cleaning and formatting.

3.1 Methodology

In describing the methodology of my work, I began by importing all the necessary Python components to ensure the proper usage of my notebook. Following that, I implemented my foursquare client data to allow accurate collection of location and venue data through the Foursquare API. This allowed me to both set the radius I wanted for my work but also allowed me to specify the type of venue I was in search of. For this instance, I performed a search based on the Foursquare listed category ID for locations categorized as a 'Park'. After gathering my results from the API, I was able to cleanly table the parks listed I received back and then analyze the accuracy. While not entirely accurate, the results sufficed with 29/32 of the tabled locations correctly being labeled as a park for the New York section of my project (37/40 for the project total). I then used folium to map all the park locations with the set 750 meters from the geolocator generated center-point. I followed the completing of my New York exploration by reporting these same steps for San Francisco.

After making my finds between the two cities, I further explored New York as a whole and gather venue and location data of all of the neighborhoods within the city. After mapping the locations of all the neighborhoods, I further broke down the neighborhoods and the venues they consisted of. I arranged the data to display the neighborhoods top most frequent venues of which many carried the park category as a top five most frequent venue.

Lastly, to see which neighborhoods were considered to be similar to one another, I ran a kmeans clustering function on top of the neighborhood and found that most of the neighborhoods fell into the 0 and 3 cluster.

4.1 Results

Following the completion of exploring the park locations of the two cities, I found the results to show me that New York had more park locations listed with the set radius than that of San Francisco. Although the overall accuracy of the retuned park categorized locations was perfect, it remained true that the New York area had over 20 more park locations within 750 meters of the provided center point.

The results of further insight into New York City revealed that a park was one of the most frequent or most common venues for the many neighborhoods of the city. By reviewing the means data results, it was shown that much of these neighborhoods were very similar as well, with a few outliers here and far between.

5.1 Conclusions

I conclude my project after receiving the shocking results of New York providing a greater selection of park locations within my set parameters. Prior to doing this project, I would not have hypothesized this to be the case never the less, it seems as though New York is clear choice while only taking into account the tested parameters. It's not a given however that the results would hold if comparing the entirety of the city or by looking at the locations on a per square milage basis. However, it's not entirely realistic that the common resident would make use of every listed park location within their city, and that is why the set parameters were used. I would expect this data methodology to be similarly useful even when using different cities or center-points to determine the best cities offering the most park/outdoor locations.

In those choosing New York, the New York environment holds numerous neighborhoods that are not only similar in make but have many neighborhoods were parks comprise a majority of the venues within them. If park availability is of concern, it

will be more than easy enough to find neighborhood with both similar make-ups and numerous parks to be visited.