Finding a Suitable Restaurant Location in an Unfamiliar City

Coursera IBM Data Science Capstone Project : Final Presentation

by Ryan Herchig

May 30, 2020

Content

- Introduction & Background
- Preliminary Data Analysis
- Methodology
- Results
- Discussion
- Conclusion

Introduction & Background

Problem Statement

The business problem I have chosen to address is to determine suitable placement of a new restaurant within a city chosen by the user.

Target Users

- Business prospector looking to open a new restaurant in an unfamiliar city
- Person looking for area with large number of restaurants in a new city

Benefits to Users

- Could give advantage to restaurant (small business) owner to help them remain in business
- Could be useful for people who travel to new places and enjoy trying new cuisine

Introduction & Background

Use Case 1

A new restaurant owner who is looking to bring a new type of cuisine to a city might want to open their restaurant in an area with a high density of restaurants. This will bring people to the area, but the competition could be slightly reduced due to the difference in the style of food served.

Use Case 2

An entrepreneur looking to establish a new restaurant of the same or similar type already prevalent throughout the city might want to search for areas with low densities of restaurants to reduce competition.

Use Case 3

A person traveling to an unfamiliar city looking for a bite to eat would also want to find areas with a high density of restaurants.

Preliminary Data Analysis

- As a preliminary data analysis, 5 different cities of various sizes and populations were investigated and compared.
- The cities used were Los Angeles California, New York New York, Miami Florida, Houston Texas, and Atlanta Georgia.
- Information about the sizes of the cities in square kilometers as well as population sizes was retrieved using SPARQL queries.

Los Angeles was chosen as a case study due to the fact that it has a moderate size and number of inhabitants compared to the other cities which were investigated. Only physical zip codes which fall inside the city limits were considered.

Preliminary Data Analysis

Area & Population Statistics

A preliminary analysis of 5 large cities in the United States shows that the city with the largest area is Houston Texas while New York City has the most inhabitants. Miami and Atlanta are both relatively small cities compared to the others, both in terms of square kilometers and population.

Table: City Statistics

Address	Area (km^2)	Population (million residents)	Population Density (residents / km^2)	Number of Venues	Restaurant Density (restaurants / km^2)
Los Angeles, CA	1302.0	4.0	3095.93	701	0.538
New York, NY	1214.0	8.6	7040.56	58	0.048
Miami, FL	143.1	0.4	2790.50	668	4.666
Houston, TX	1625.0	2.1	1291.81	1694	1.042
Atlanta, GA	347.1	0.46	1336.44	625	1.801

More meaningful metrics for comparison are the population density and restaurant density (restaurants per square kilometer).

Preliminary Data Analysis

Number of Venues, Population & Restaurant Densities

- New York City has the largest population density, followed by Los Angeles California.
- The FourSquare API returned the largest number of restaurant venues for Houston Texas.
- Only 58 venues were returned for New York.
- New York has a very low restaurant density while Miami's is extremely high.

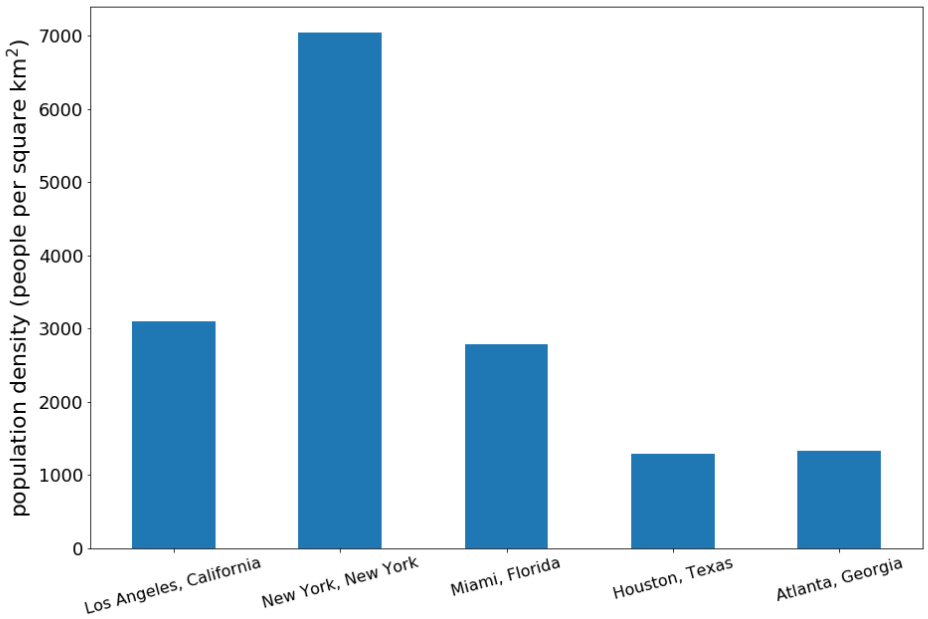
Table: City Statistics

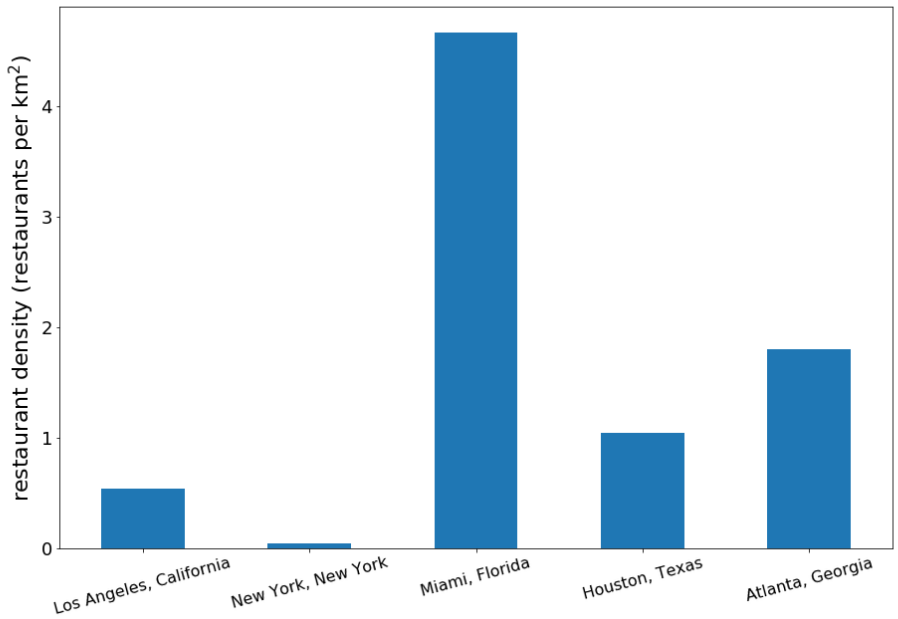
Address	Area (km^2)	Population (million residents)	Population Density (residents / km^2)	Number of Venues	Restaurant Density (restaurants / km^2)
Los Angeles, CA	1302.0	4.0	3095.93	701	0.538
New York, NY	1214.0	8.6	7040.56	58	0.048
Miami, FL	143.1	0.4	2790.50	668	4.666
Houston, TX	1625.0	2.1	1291.81	1694	1.042
Atlanta, GA	347.1	0.46	1336.44	625	1.801

Preliminary Data Analysis

City Statistics Bar Charts

The figures below show a graphical representation of the population and restaurant densities. It is clear from the first chart that New York City has the highest population density while Houston Texas with its extremely large area has the smallest population density. Miami Florida has by far the largest number of restaurants per square kilometer.





Methodology

Data Collection Procedure

- Queries to FourSquare API for each city using latitude and longitude of the physical zip code locations as the coordinates
- Search queries were run iteratively with the search radius parameter being varied for each iteration
- Search radii used are 5000, 8000, and 12000 meters
- If the search radius is made too small, venues will be missed in the intermediate space.
- To ensure all discoverable venues are found, the search radius is made excessively large
- After the query is ran for each zip code, the list is purged of duplicates
- This results in a unique list of restaurant venues for each city

Methodology

Zip Code Binning Procedure

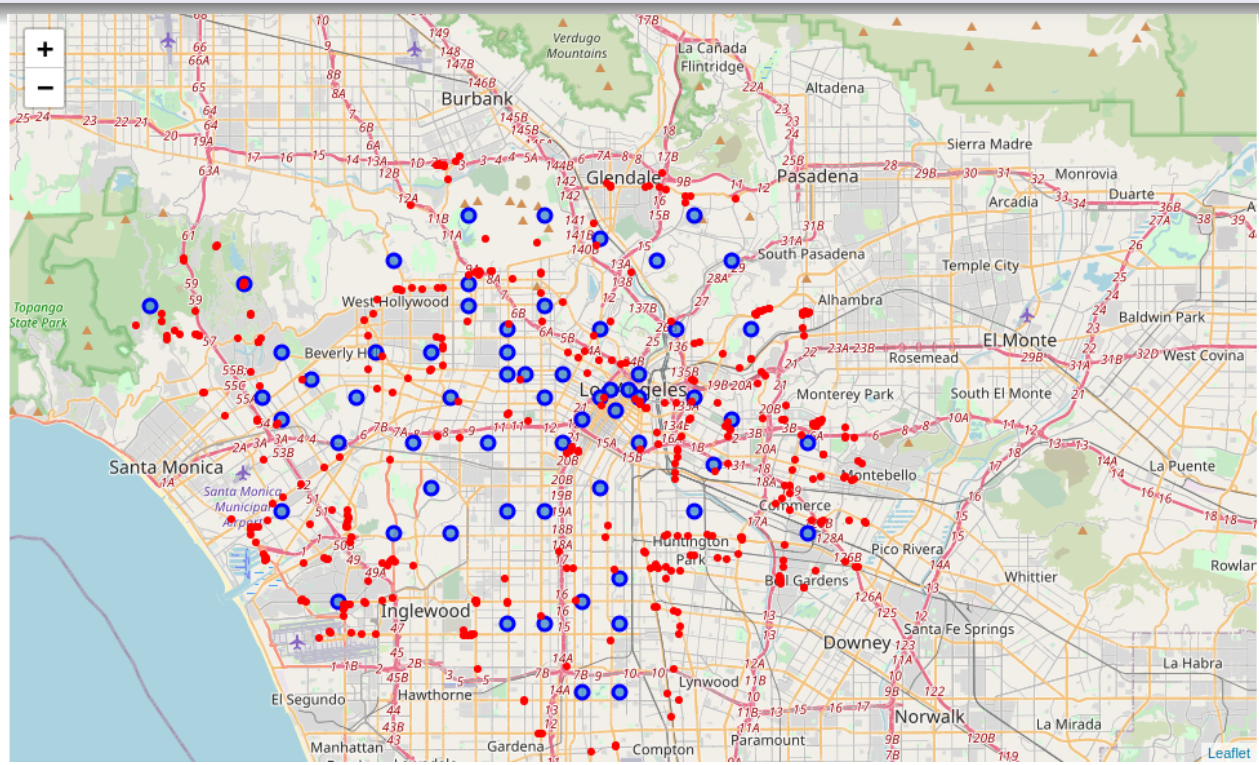
- Query restuarant locations & sort by zip code
- Zip codes were be placed into bins with the number of bins being $\approx 10\%$ of the number of zip codes
- For Los Angeles, bins were calculated using 0 as the absolute minimum and 30 as the absolute maximum
- Total number of zip codes was 52 with 10% of this being 5, rounded to the nearest integer
- Dividing values for the histogram bins were then calculated to be
- Cluster 1 includes all zip codes between 0.0 and 6.0 (rounded to the nearest integer), cluster 2 includes all zip codes greater than 6.0 and less than or equal to 12.0, etc.

Results

Venues & Zip Codes Map

Figure shows the latitude and longitude coordinates plotted on a Folium Leaflet map corresponding to the zip code locations (blue circles) and the restaurant venues (red circles). The figure shows streets with large concentrations of restaurants and other areas which are devoid of restaurants.

Areas with relatively low concentrations of restaurants could be residential areas in which businesses are not allowed to be established.



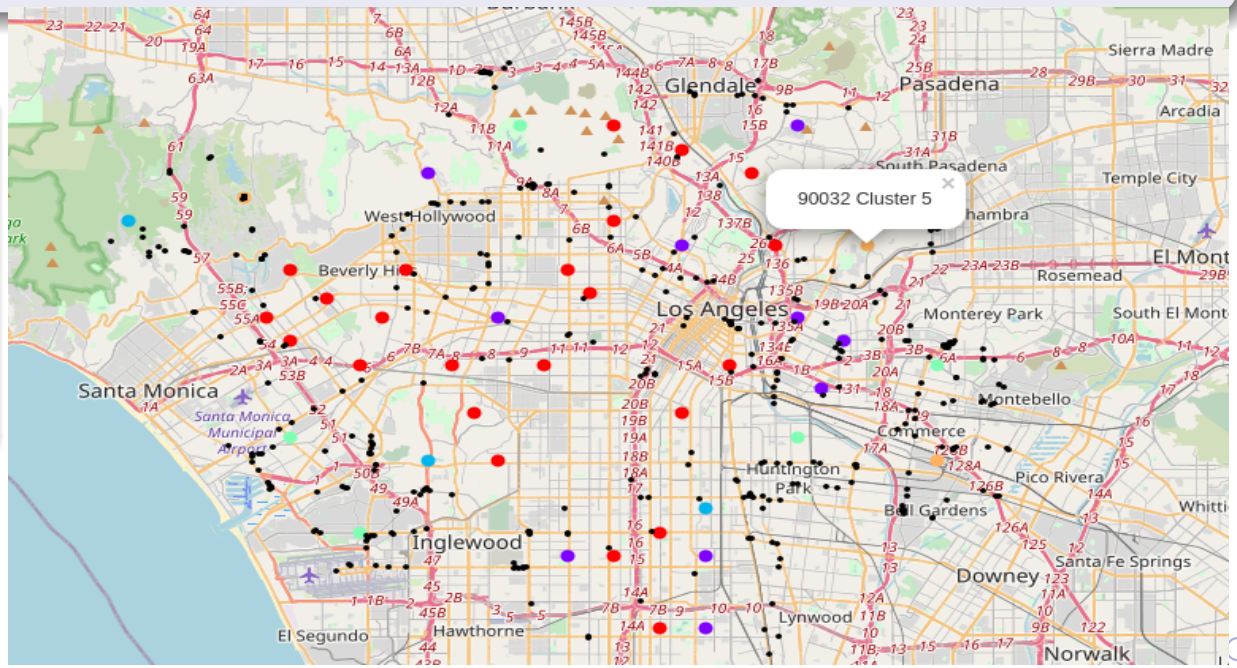
Results

Venues & Binned Areas

The Leaflet map shows the results of running the previously described binning function on the data returned from the FourSquare API for Los Angeles. The restaurant locations are represented by small black circles while the different zip code centers are depicted by the larger circles of various colors. Cluster 1 includes all zip codes where the number of restaurants falls between 0 and 6, cluster 2 includes all zip codes for which the number is greater than 6 and less than or equal to 12, etc.

Map Key

bin 1 = [0 : 6] red
bin 2 = (6 : 12] purple
bin 3 = (12 : 18] blue
bin 4 = (18 : 24] light green
bin 5 = (24 : 30] orange



Discussion

Use Case 1 - Restaurant Owner Opens in High Restaurant Density Region

- Area on the map in the vicinity of Commerce and Montebello would make excellent locations
- Orange marker (labeled by the pop-up which reads "90032 Cluster 5") exists in the northeastern area of the city
- Orange marker in the northwestern corner of the city which seems to be in an outdoor/forested area as compared to intercity Los Angeles

Use Case 2 - Restaurant Owner Opens in Low Restaurant Density Region

- The red markers denote areas with few restaurants in their immediate vicinity
- User would need to be conscious of the fact that some areas may be devoid of restaurants simply because the zoning restrictions do not allow businesses/commercial establishments

Discussion

Use Case 3 - Casual User Looking for Area with Many Restaurant Options

- Same areas mentioned for Use Case 1 would be of interest
- neighborhoods of Huntington Park, Montebello, and the area east of Topango State Park would all be good choices for someone looking for a variety of restaurant choices in a small area

General Observations

The areas which seem to have the highest concentration of restaurant venues are those near highways and main roads and which lie on the outskirts of town in the suburbs. The lowest concentrations seem to be found in what appear to be highly residential areas.

Conclusions

- Using Los Angeles as a specific case study, several areas were identified as promising due to either their high concentration of restaurant venues or low concentration.
- The areas with high concentrations of restaurants are also attractive areas for a person using the software to find areas in an unfamiliar city which have many restaurant choices in a small geographic region.
- Though Los Angeles was used as an example in this report, the application has been written in a general way such that similar results could be found for any city supported by the FourSquare database.

The End