

The Authentic Soul of New York

AMY RUTH'S

~ EST. 1998 ~

Home Style Southern Cuisine

COMFORT • CUISINE • COURTESY • CULTURE

Amy Ruth's Location Scouting

Introduction



“Amy Ruth's” restaurant in New York City wants to expand to a new location. It can either be in New York itself or any of the major US cities (1M or more inhabitants).

The goal of the project is to identify another neighborhood in an US city, which is similar to the one where “Amy Ruth’s” currently resides (Zip Code 10026).

Amy Ruth's Location Scouting

Data and Processing



The project uses three data sources:

- List of US cities (*Wikipedia*)
- Neighborhood definition: postal zip codes from “*uszipcode*” python package
- The data characterizing neighborhood (*Foursquare* venue data)

All zip codes are retrieved for the most populous cities in the US (≥ 1 M inhabitants) extracted from Wikipedia. Then venue data from Foursquare is added to that table as one row per zip code. A clustering will be performed and similar zip code areas as the original location are extracted and further processed.

Amy Ruth's Location Scouting

Data and Processing

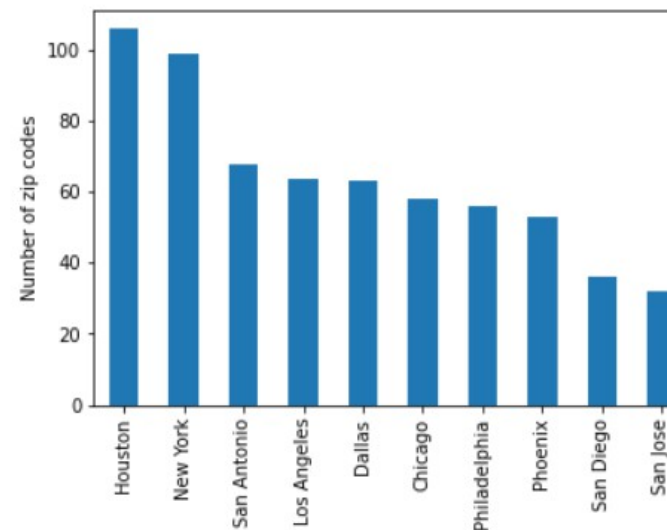


First glance at data:

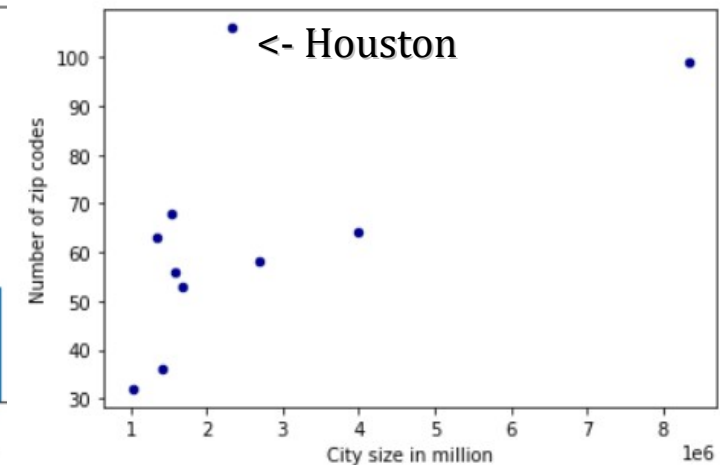
The most populated cities in the US according to the Wikipedia table are:

City	State	Size Estimate
New York	New York	8336817
Los Angeles	California	3979576
Chicago	Illinois	2693976
Houston	Texas	2320268
Phoenix	Arizona	1680992
Philadelphia	Pennsylvania	1584064
San Antonio	Texas	1547253
San Diego	California	1423851
Dallas	Texas	1343573
San Jose	California	1021795

Number of zip codes per city:



City size vs. Number of zip codes



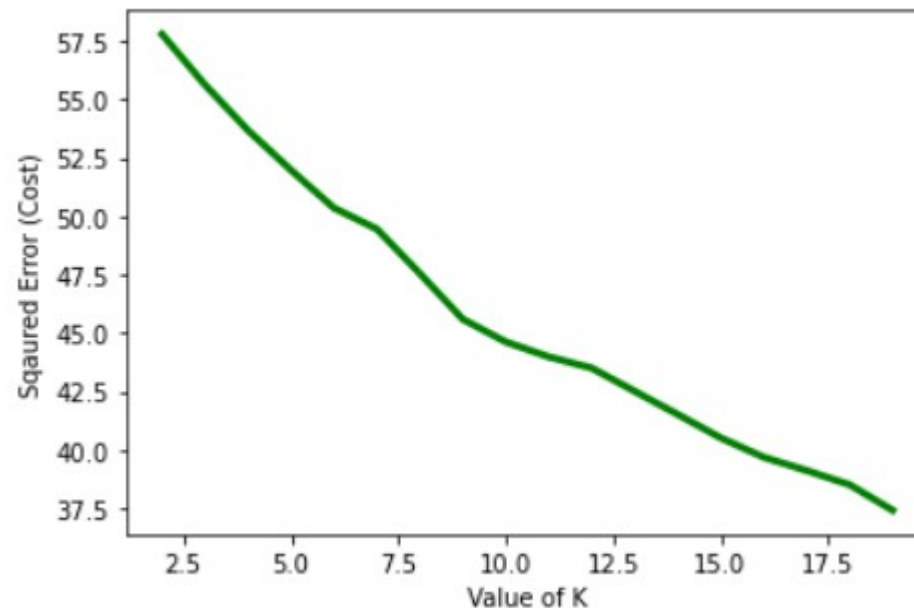
Houston is a bit of an outlier when comparing city size vs. number of zip codes

Amy Ruth's Location Scouting

Methodology - Clustering

Determining optimal number of clusters "k" value for clustering for one hot encoded venue data per zip code

Number of clusters "k" vs squared error (cost)



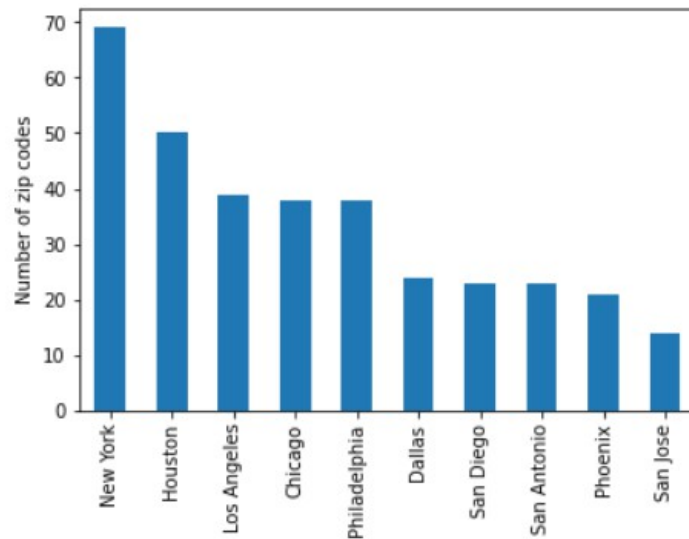
-> "weak" elbow for k=9

Amy Ruth's Location Scouting

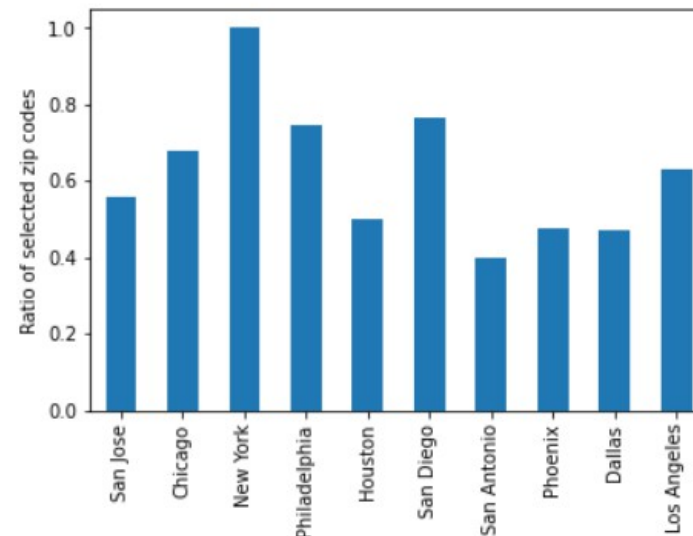
Methodology - Clustering

Neighborhood of interest is part of cluster "7" (contains zip code 10026 where Amy Ruth's is located), which contains a lot of zip codes (339):

Number of zip codes part of cluster "7"



Fraction of zip codes of individual cities part of cluster "7"



-> interestingly New York neighborhoods are almost all included in cluster "7", i.e. are more consistently similar compared to other cities

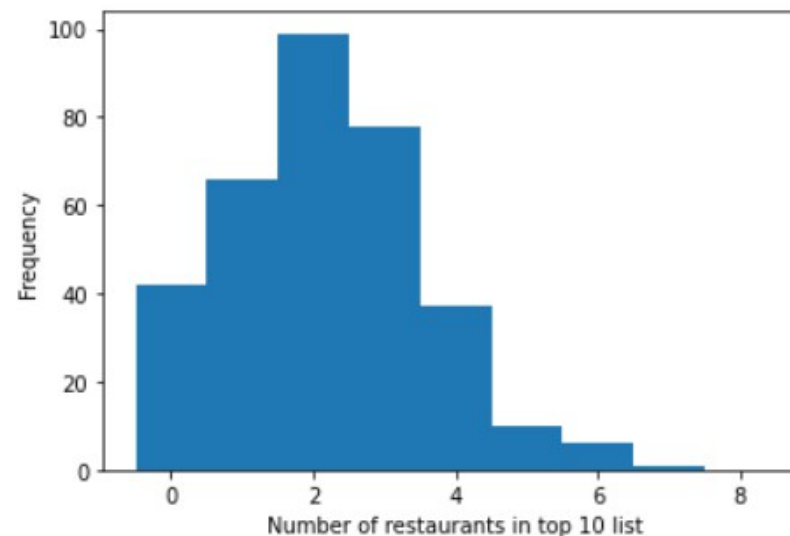
Amy Ruth's Location Scouting

Methodology - Postprocessing

Initial number of potential restaurant locations is large: remove zip codes if:

- "Southern / Soul Food Restaurant" in the top 10 list (avoid direct competition)
- No restaurants in the top 10 list (zip code might not be suitable, e.g. Industrial area)
- Contains more than the average number of restaurants (2.2) in the top 10 list

Histogram of number of restaurants in top 10 list for zip codes in cluster 7



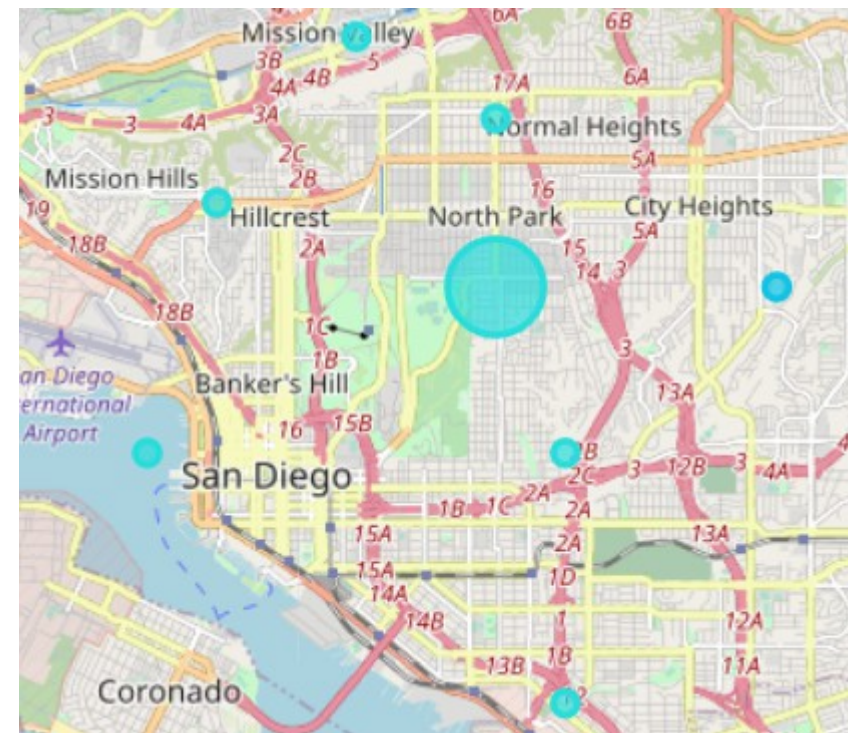
Amy Ruth's Location Scouting Results

For remaining zip codes identify the most similar zip code per city by using Euclidean distance metric on one hot encoded table when compared to 10026 zip code (Amy Ruth's location)

Zip code short list for potential restaurant areas:

City	State	Zipcode	Latitude	Longitude
New York	New York	10023	40.7800	-73.9800
Los Angeles	California	90042	34.1100	-118.1900
Chicago	Illinois	60647	41.9200	-87.7000
Houston	Texas	77201	29.7655	-95.3658
Phoenix	Arizona	85004	33.4500	-112.0700
Philadelphia	Pennsylvania	19107	39.9500	-75.1600
San Antonio	Texas	78229	29.5100	-98.5700
San Diego	California	92104	32.7400	-117.1300
Dallas	Texas	75226	32.7800	-96.7800
San Jose	California	95110	37.3400	-121.9100

Shortlisted areas highlighted on Folium map for selected cities (large circle):



Amy Ruth's Location Scouting

Discussion



- Clustering was a helpful methodology to group the 572 zip codes and come up with an initial set of interesting areas
- However, clustering alone provided too many potential locations (339)
- Post-processing of the initial selection obtained from clustering was possible using the Foursquare data to focus on areas where restaurants are underrepresented to avoid too stiff competition