

The Battle of the Neighborhoods – Report

1. Introduction & Business Problem

Problem Background:

Home Away From Home[1]

When there is a need, whether from job, career or needs for long vacation break, you might need to travel and live in a different city for either a short or longer period of time. If you are the type of person that enjoy staying at similar neighborhood as the one you are currently living in to make yourself comfortable - a feeling of stay at **home away from home**. You will want to find a living place in the visiting city that has as similar neighborhood as your hometown.

[1]: Theme title of Lesson 1 in book "Best Lessons of a Chess Coach" by Sunil Weeramantry and Ed Ebusebi, 1993

Problem Description:

The problem is that prior to visiting the city, how do you find out which neighborhood of the visiting city to stay that has similar neighborhood as your hometown?

As an example, if you are living in Cupertino, California and plan to stay at New York city for one year. Which neighborhood in New York City should you choose to stay to experience similar neighborhood as Cupertino?

For this project, we will limit our scope of visiting city to New York city and your home city can be any city in the US. The tool developed could be further expanded to other visiting city and any home city in the world with the methodology adopted in this project.

Target Audience

Those travelers who want to feel "Home away from Home" during visit to other cities.

Success Criteria

The success criteria of the project will be a good recommendation of borough/Neighborhood choice to traveler based on neighborhood similarity between your home city and New York city.

2. Data

Home cities to be analyzed in this project:

Cupertino, CA.

Mountain View, CA.

Menlo Park, CA.

Tempe, AZ.

Dallas, TX

But you can input any US city name you want.

Visiting city to be analyzed in this project:

The neighborhoods will be limited to New York city only.

Data 1:

"uscitiesv1.4.csv" downloaded from <https://simplemaps.com/data/us-cities> , which contains geographical coordinates data of US cities. Home city name will be used to look up its corresponding latitude and longitude coordinates, which in turn will be used as input for the Foursquare API, that will be leveraged to provision venues information for your home city. We will use the Foursquare API to explore venues of your home city.

Below is the Pandas data frame with geographical coordinates data of US cities.

	city	city_ascii	state_id	state_name	county_fips	county_name	lat	lng
0	Prairie Ridge	Prairie Ridge	WA	Washington	53053	Pierce	47.1443	-122.1408
1	Edison	Edison	WA	Washington	53057	Skagit	48.5602	-122.4311
2	Packwood	Packwood	WA	Washington	53041	Lewis	46.6085	-121.6702
3	Wautauga Beach	Wautauga Beach	WA	Washington	53035	Kitsap	47.5862	-122.5482
4	Harper	Harper	WA	Washington	53035	Kitsap	47.5207	-122.5196

Data 2:

Visiting city - New York city has a total of 306 neighborhoods. In order to identify neighborhoods and explore them, we will need a dataset that contains all the neighborhoods as well as the latitude and longitude coordinates of each neighborhood.

This dataset exists for free on the web. Link to the dataset is:

https://geo.nyu.edu/catalog/nyu_2451_34572

Data 3:

New York city geographical coordinates data will be used as input for the Foursquare API, that will be leveraged to provision venues information for each neighborhood. We will use the Foursquare API to explore neighborhoods in New York City. The below is image of the Foursquare API data.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Wakefield	40.894705	-73.847201	Lollipops Gelato	40.894123	-73.845892	Dessert Shop
1	Wakefield	40.894705	-73.847201	Rite Aid	40.896649	-73.844846	Pharmacy
2	Wakefield	40.894705	-73.847201	Carvel Ice Cream	40.890487	-73.848568	Ice Cream Shop
3	Wakefield	40.894705	-73.847201	Dunkin Donuts	40.890631	-73.849027	Donut Shop
4	Wakefield	40.894705	-73.847201	Cooler Runnings Jamaican Restaurant Inc	40.898276	-73.850381	Caribbean Restaurant

Data 4:

Hometown city geographical coordinates data will be used as input for the Foursquare API, that will be leveraged to provision venues information as a matching base for visiting city neighborhoods. The below is image of the Foursquare API data of hometown city – Cupertino.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Cupertino	37.3167	-122.0465	Apple Fitness Center Results Way	37.317490	-122.051129	Gym
1	Cupertino	37.3167	-122.0465	Fujitsu Planetarium De Anza College	37.318837	-122.046217	Planetarium
2	Cupertino	37.3167	-122.0465	Ike's Place	37.322801	-122.041220	Sandwich Place
3	Cupertino	37.3167	-122.0465	Shane Co.	37.323249	-122.046537	Jewelry Store
4	Cupertino	37.3167	-122.0465	Whole Foods Market	37.323478	-122.039759	Grocery Store

3. Methodology

What Machine Learnings should we use?

When it comes to what machine learning algorithm should be used for the target problem, it really depends on what kind of problem you are solving and what goal you want to achieve. Our goal here is to recommend a neighborhood of New York city that has close similarity to our hometown. There is no golden labeled data here that can be used to train for similarity, so supervised learning methods like classification and regression will not be suitable to solve the problem. Un-supervised learning seems a good candidate, such as k-mean or DBSCAN clustering. But here K will be difficult to select because we have 306 neighborhoods and we only want to pick, say top 5.

Since the goal is to recommend something that has close similarity to something else, recommendation system fit naturally to solve this problem.

Here we use default method used in content-based recommendation system.

The mapping of how our data to content-based recommendation is as following.

- Hometown venues data is used as user profile.
- Venues data of 306 New York city neighborhoods are used as items matrix.
- The recommended list will be based on sorted results of items matrix multiply by user profile.

A simplified method is used in this project and the results look satisfactory. The simplified method only use TF instead of TF-IDF and use pairwise cosine similarity in matrix multiplication.

Exploratory Data Analysis:

TBD

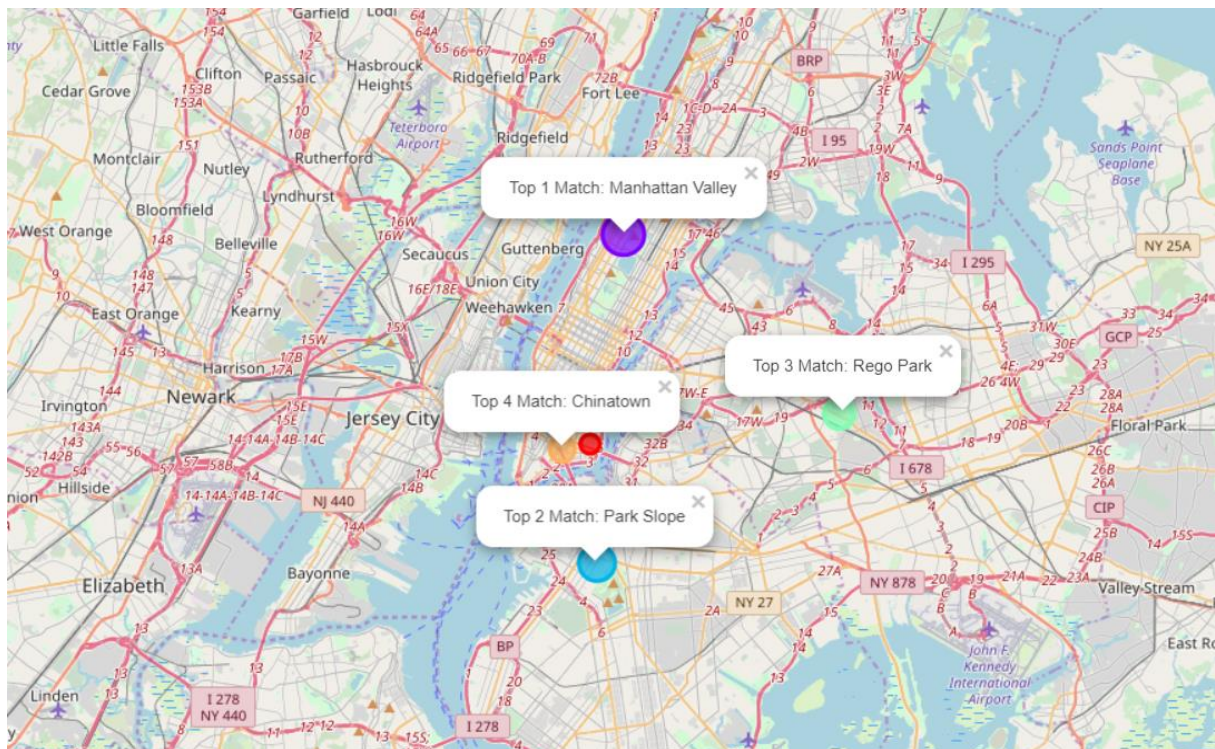
4. RESULTS

This section presents the top 5 recommendation neighborhoods for 5 different hometowns. The top 10 common venues of hometown and recommended neighborhoods are listed in table form followed by their locations on New York City map.

By examining the top 10 Common Venues between hometown and recommended neighborhoods, there a good amount of overlap/matching. The higher the rank, the more the venue overlaps between the two. And the results are valid for all 5 selected hometowns. Therefore, the results look reasonably good.

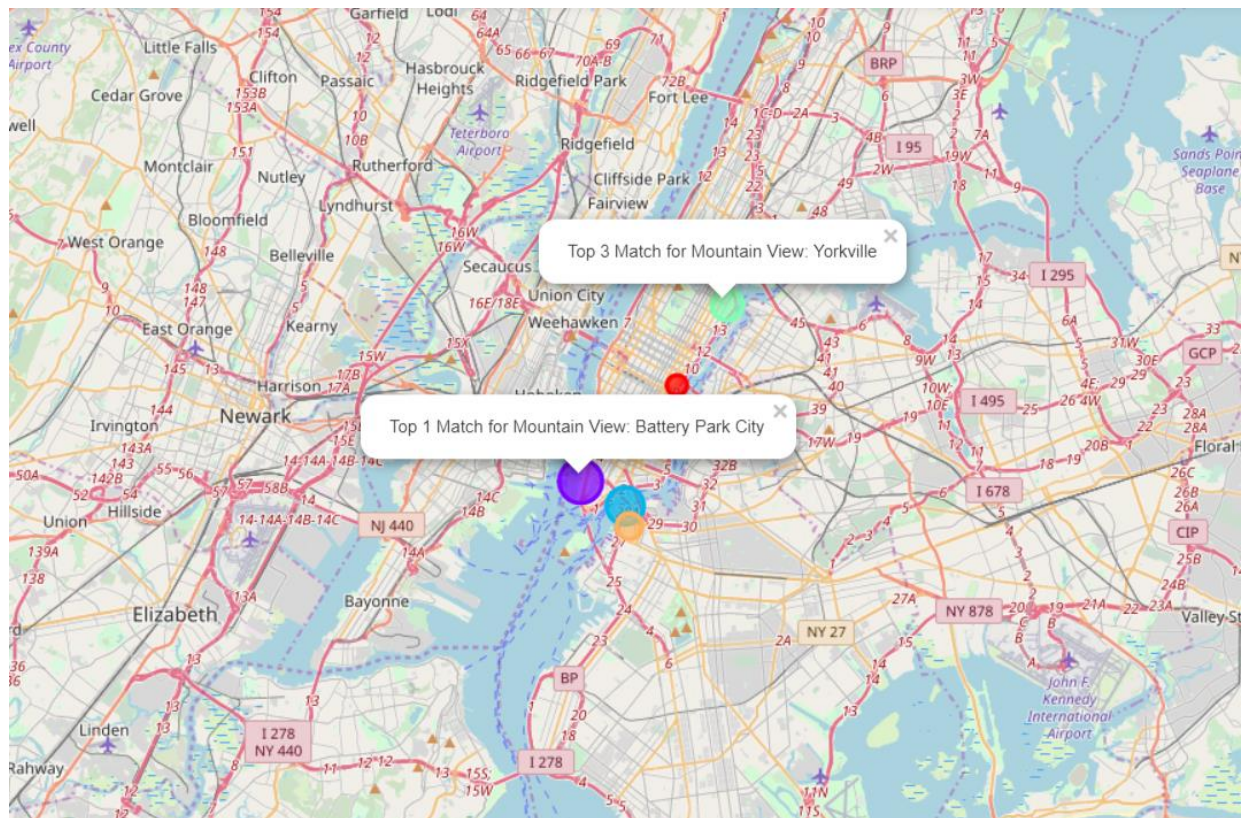
1. Hometown: Cupertino, CA

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Cupertino	Coffee Shop	Chinese Restaurant	Bakery	Sandwich Place	Park	Vietnamese Restaurant	Grocery Store	American Restaurant	Bubble Tea Shop	Mexican Restaurant
1	Manhattan Valley	Pizza Place	Coffee Shop	Yoga Studio	Café	French Restaurant	Mexican Restaurant	Bar	Chinese Restaurant	Thai Restaurant	Burger Joint
2	Park Slope	Coffee Shop	Pizza Place	American Restaurant	Italian Restaurant	Burger Joint	Bookstore	Japanese Restaurant	Pub	Spa	Bagel Shop
3	Rego Park	Bakery	Grocery Store	Pharmacy	Pizza Place	Sushi Restaurant	Chinese Restaurant	Sandwich Place	Restaurant	Donut Shop	Cosmetics Shop
4	Chinatown	Chinese Restaurant	American Restaurant	Cocktail Bar	Dim Sum Restaurant	Vietnamese Restaurant	Bubble Tea Shop	Noodle House	Bakery	Salon / Barbershop	Ice Cream Shop
5	Lower East Side	Coffee Shop	Art Gallery	Chinese Restaurant	Café	Pizza Place	Ramen Restaurant	Cocktail Bar	Japanese Restaurant	Sandwich Place	Shoe Store



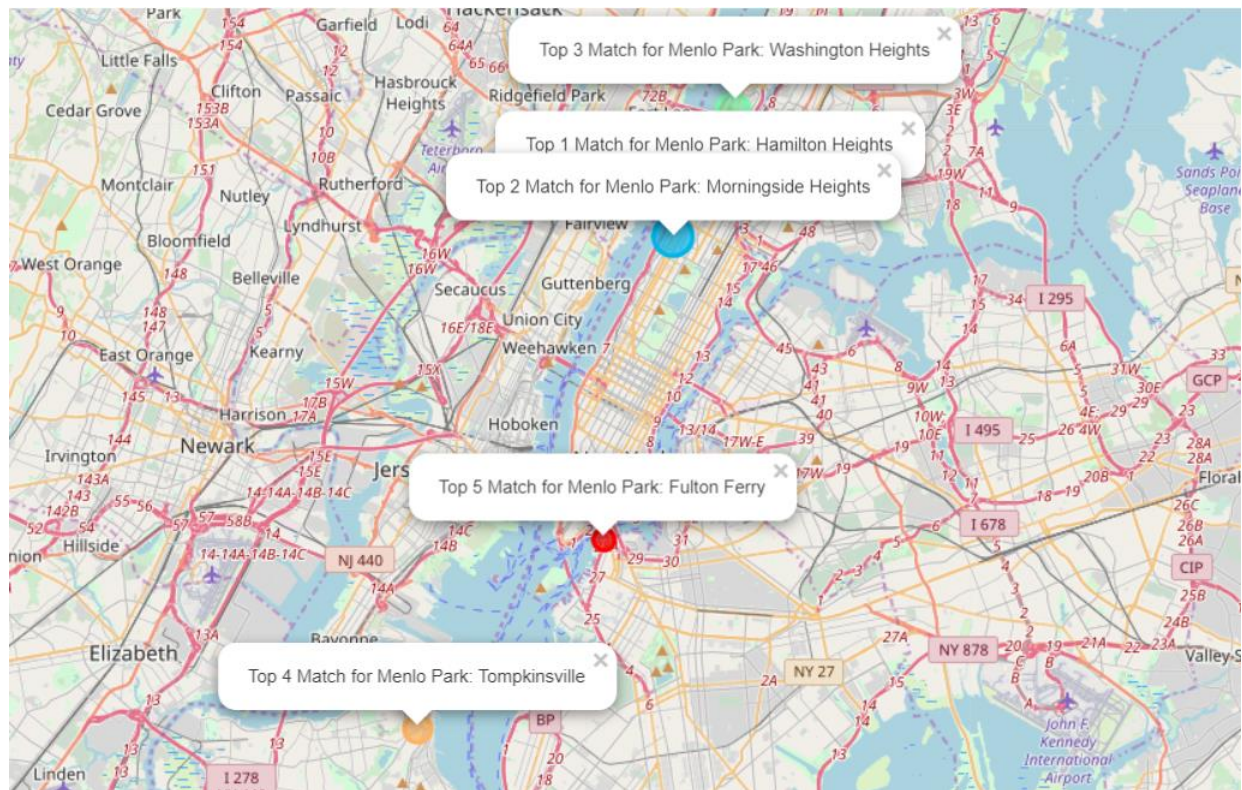
2. Howtown: Mountain View, CA

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Mountain View	Park	Coffee Shop	Mexican Restaurant	Japanese Restaurant	Ramen Restaurant	Pizza Place	Grocery Store	Sushi Restaurant	American Restaurant	Optical Shop
1	Battery Park City	Coffee Shop	Park	Hotel	Wine Shop	Italian Restaurant	Gym	Ice Cream Shop	Fountain	Plaza	BBQ Joint
2	Fulton Ferry	Park	Ice Cream Shop	American Restaurant	Scenic Lookout	Hotel Bar	Café	Bridge	Bakery	Playground	Coffee Shop
3	Yorkville	Italian Restaurant	Bar	Coffee Shop	Gym	Pizza Place	Sushi Restaurant	Deli / Bodega	Mexican Restaurant	Thai Restaurant	Diner
4	Brooklyn Heights	Yoga Studio	Park	Italian Restaurant	Deli / Bodega	Cosmetics Shop	Coffee Shop	Pizza Place	Gym	Plaza	Pet Store
5	Tudor City	Park	Mexican Restaurant	Café	Sushi Restaurant	Diner	Pizza Place	Hotel	Deli / Bodega	Greek Restaurant	Dog Run



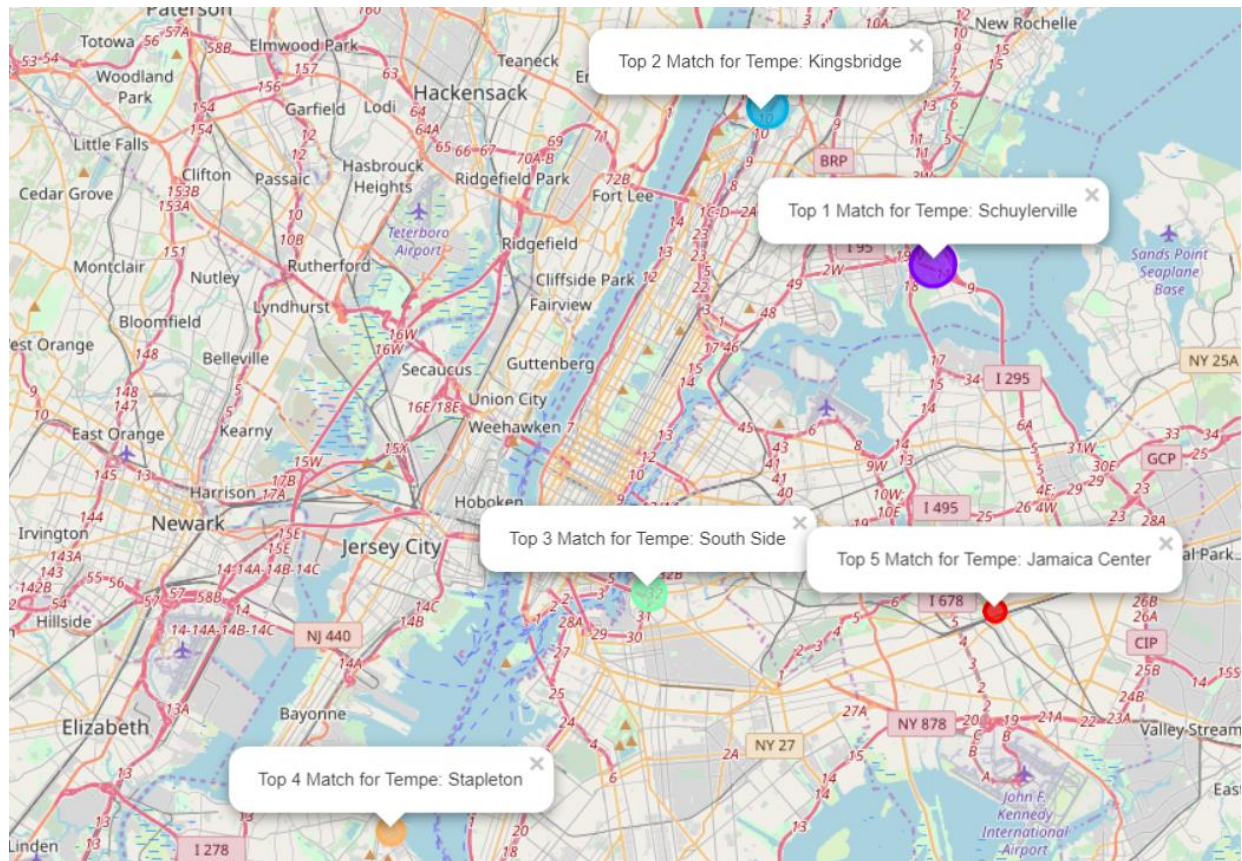
3. Hometown: Menlo Park, CA

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Menlo Park	Park	Café	Food Truck	Fish & Chips Shop	Event Space	Gym / Fitness Center	Grocery Store	Caribbean Restaurant	Coffee Shop	Performing Arts Venue
1	Hamilton Heights	Mexican Restaurant	Deli / Bodega	Café	Coffee Shop	Pizza Place	Bakery	Caribbean Restaurant	Chinese Restaurant	School	Sandwich Place
2	Morningside Heights	Coffee Shop	American Restaurant	Bookstore	Park	Tennis Court	Sandwich Place	Food Truck	Burger Joint	Deli / Bodega	Pub
3	Washington Heights	Café	Bakery	Sandwich Place	Mobile Phone Shop	Spanish Restaurant	Mexican Restaurant	Shoe Store	Grocery Store	New American Restaurant	Chinese Restaurant
4	Tompkinsville	Deli / Bodega	Pizza Place	Park	Rental Car Location	Spanish Restaurant	Thrift / Vintage Store	Grocery Store	Mexican Restaurant	Gastropub	Supermarket
5	Fulton Ferry	Park	American Restaurant	Ice Cream Shop	Scenic Lookout	Bridge	Coffee Shop	Bakery	Hotel Bar	Café	Playground



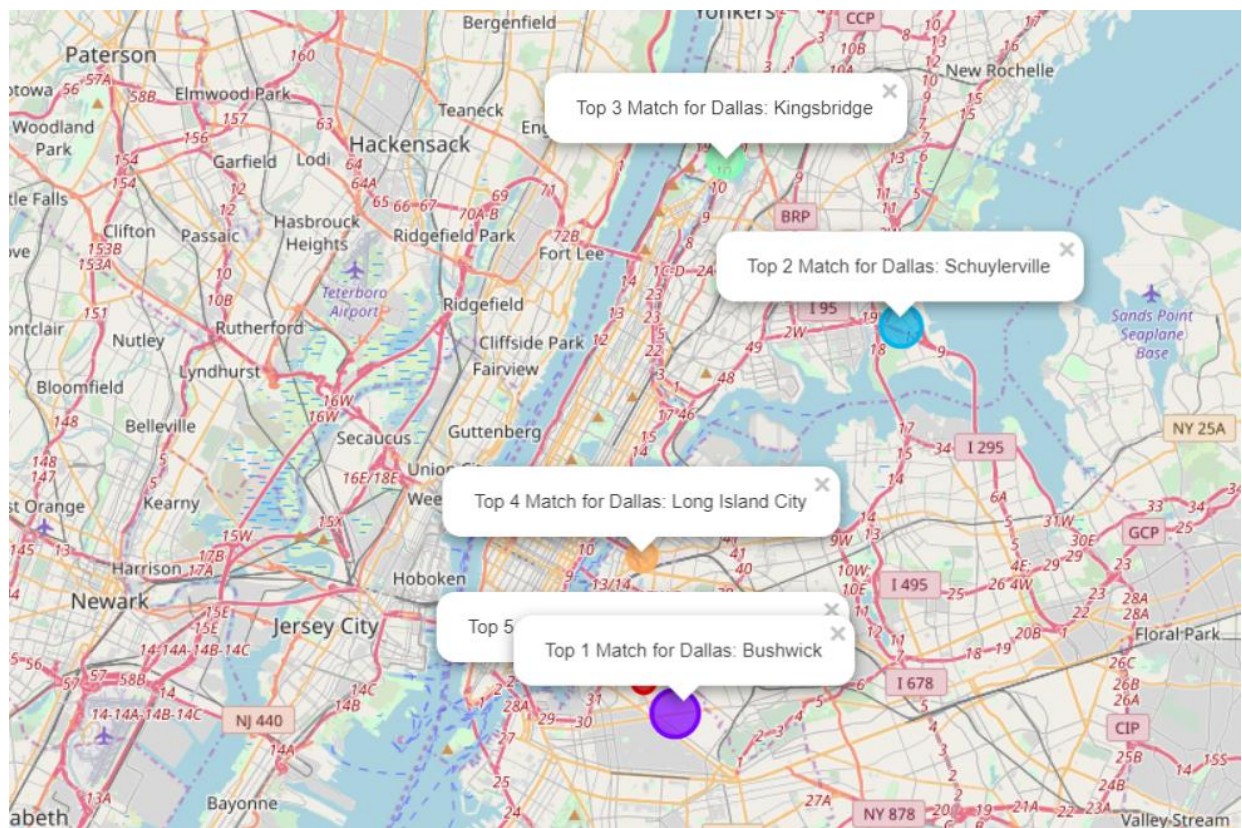
4. Howtown: Tempe, AZ

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Tempe	Sandwich Place	Coffee Shop	Pizza Place	Mexican Restaurant	Fast Food Restaurant	Bar	Grocery Store	Shipping Store	Chinese Restaurant	Asian Restaurant
1	Schuylerville	Pizza Place	Bar	Bank	Pharmacy	Mexican Restaurant	Fast Food Restaurant	Sandwich Place	Hookah Bar	Donut Shop	Latin American Restaurant
2	Kingsbridge	Pizza Place	Sandwich Place	Bar	Supermarket	Latin American Restaurant	Mexican Restaurant	Discount Store	Liquor Store	Spanish Restaurant	Bakery
3	South Side	Bar	Pizza Place	Coffee Shop	American Restaurant	Yoga Studio	Chinese Restaurant	Burger Joint	Breakfast Spot	Asian Restaurant	Dive Bar
4	Stapleton	Discount Store	Mexican Restaurant	Bank	Pizza Place	Sandwich Place	Café	Motorcycle Shop	Optical Shop	Skate Park	Fast Food Restaurant
5	Jamaica Center	Mobile Phone Shop	Mexican Restaurant	Sandwich Place	Fried Chicken Joint	Coffee Shop	Clothing Store	Caribbean Restaurant	American Restaurant	Donut Shop	Performing Arts Venue



5. Howtown: Dallas, TX

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Dallas	Mexican Restaurant	Bar	Fast Food Restaurant	Pizza Place	Taco Place	Convenience Store	Dive Bar	Sandwich Place	Coffee Shop	Smoke Shop
1	Bushwick	Bar	Mexican Restaurant	Pizza Place	Coffee Shop	Deli / Bodega	Bakery	Chinese Restaurant	Thrift / Vintage Store	Discount Store	Shoe Store
2	Schuylerville	Pizza Place	Bar	Bank	Pharmacy	Mexican Restaurant	Fast Food Restaurant	Sandwich Place	Hookah Bar	Donut Shop	Latin American Restaurant
3	Kingsbridge	Pizza Place	Sandwich Place	Bar	Supermarket	Latin American Restaurant	Mexican Restaurant	Discount Store	Liquor Store	Spanish Restaurant	Bakery
4	Long Island City	Hotel	Coffee Shop	Mexican Restaurant	Café	Pizza Place	Bar	Bus Station	Donut Shop	General Entertainment	Burger Joint
5	East Williamsburg	Bar	Cocktail Bar	Coffee Shop	Deli / Bodega	Mexican Restaurant	Bakery	Music Venue	Concert Hall	Gym / Fitness Center	Bagel Shop



5.DISCUSSION

Assumptions

Although the results look reasonably good in terms of good matching on the top common venues between hometown city and recommended neighborhoods, there are some assumptions have been made to get this result and there are a lot of room for improvements.

The main assumption has been made is that only neighborhood venue datum is critical and therefore is used to get the result. But there are different set of data that might also be as important. These data include but not limited to neighborhood safety (police crime record), housing/rental price, school district, or even personal data such as race, age, food preference.

Suggestion for further Improvement

The followings are suggestion for further improvement of model and results if time is available.

- Add more neighborhood data – such as police crime record, housing/rental price, traffic, weather, distance to work, demographic info, etc.
- Add more personal data – race/age/sex, favorite food, etc.
- Improve algorithm – use TF-IDF instead of TF only, advanced to use collaborate filtering if more data is available and when it's appropriate.
- Extend the coverage to more visiting cities – collect more neighborhood geographic coordinates of different visiting city for analysis.
- Extend the coverage of hometown cities – extend to whole world instead of based in US only.

6.CONCLUSION

With content-based filtering with pairwise cosine similarity, we are able to provide good neighborhood recommendation of your visiting city (New York) that make you feel comfortable as you are staying at your hometown. The assumption to get the result and suggestions for further improvement are discussed. I believe that with enough resource and time, this project could be developed into a very successful tool that will benefit a lot of people who want to feel “Home away from Home” during stay at other cities.

