The Battle of Neighborhoods

Which venues are coming to your area?

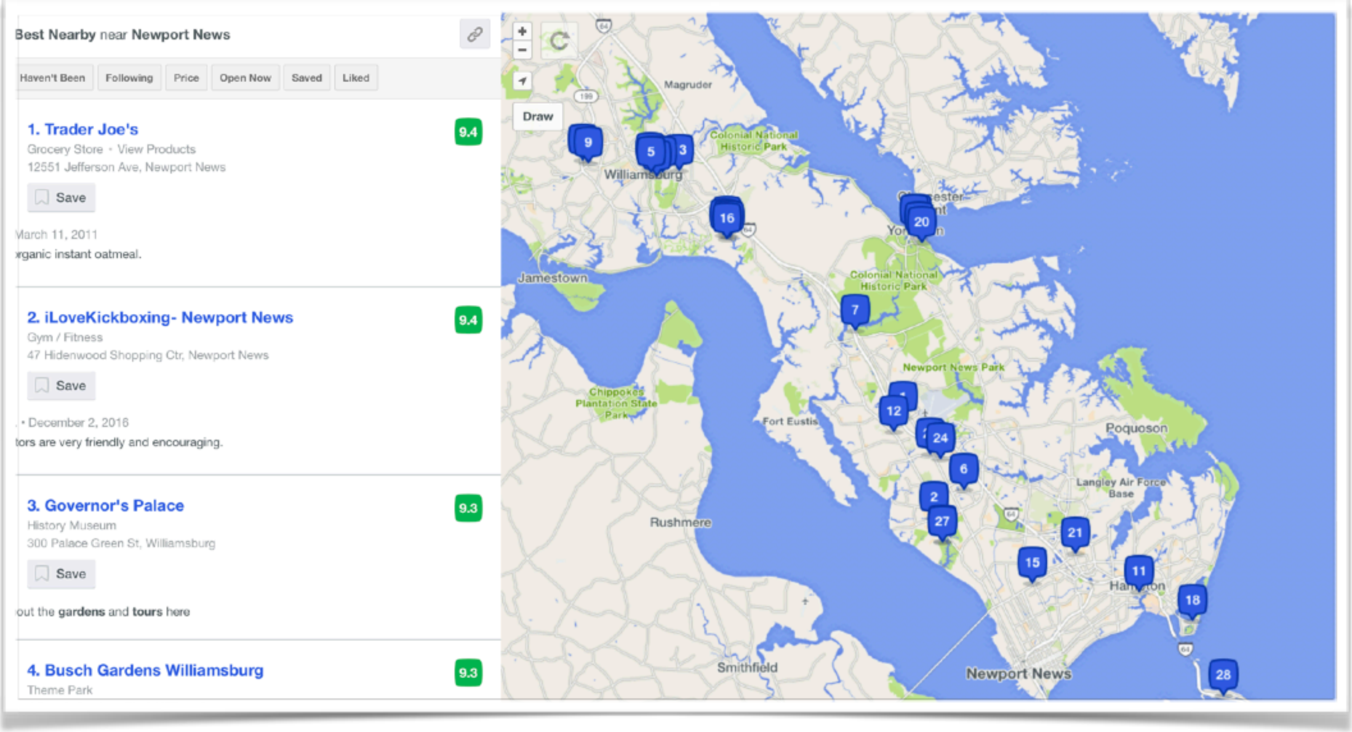
Jason Merten - February 10, 2020

Introduction

For the Coursera Capstone, I decided to do an analysis of what future restaurants and/or businesses would be recommended to open in a certain area based on popular restaurants and businesses in another area with a similar population. The idea is to use geographically dispersed cities and towns that are still relatively close together to reduce outliers (such as In-N-Out being recommended for the East Coast). Hopefully this project will provide some insight into the types of popular restaurants and attractions in both cities/regions and be of some use to sort-of predict what kind of restaurant/attraction will open next in an area.

Data

Data will be collected using Foursquare to find the top restaurants and businesses in the target city and use k-means clustering to create groups that we can view on a map and analyze. The same technique will be conducted for a second city of similar population in the same regional area (no more than 2 states away) to find any similarities. Based on the data collected, we can compare the clusters of each city to determine what restaurant and/or business would be recommended to open in the target city. The recommendation would be based on most popular restaurants and businesses from each city.



Best Nearby near Newport News

1. Trader Joe's 9.4
Grocery Store · View Products
12551 Jefferson Ave, Newport News

March 11, 2011
rganic instant oatmeal.

2. iLoveKickboxing- Newport News 9.4
Gym / Fitness
47 Hidenwood Shopping Ctr, Newport News

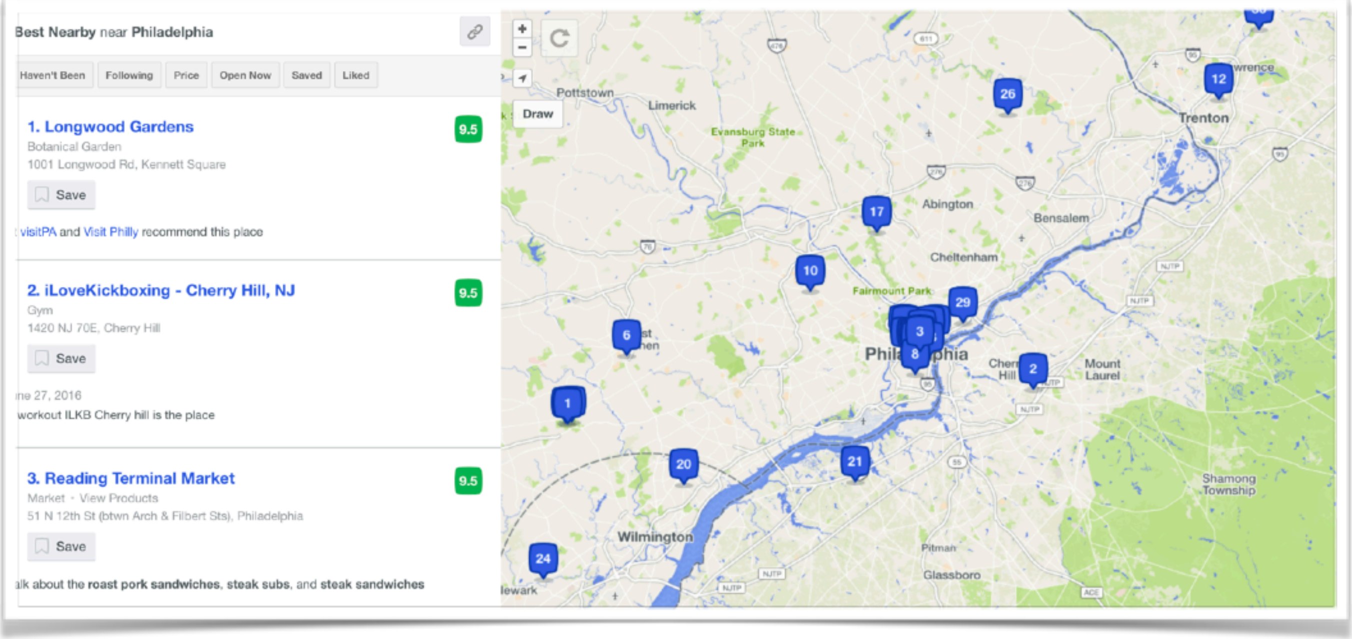
. • December 2, 2016
tors are very friendly and encouraging.

3. Governor's Palace 9.3
History Museum
300 Palace Green St, Williamsburg

out the **gardens** and **tours** here

4. Busch Gardens Williamsburg 9.3
Theme Park

Foursquare Results for Hampton Roads, VA



Best Nearby near Philadelphia

1. Longwood Gardens 9.5
Botanical Garden
1001 Longwood Rd, Kennett Square
Save
visitPA and Visit Philly recommend this place

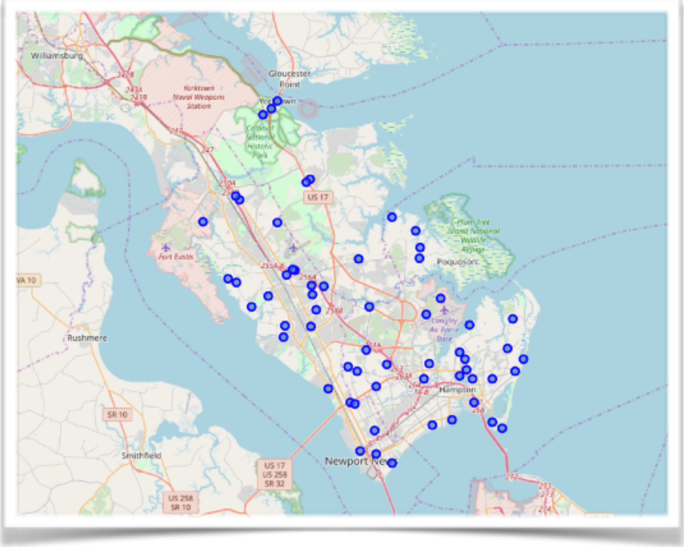
2. iLoveKickboxing - Cherry Hill, NJ 9.5
Gym
1420 NJ 70E, Cherry Hill
Save
ne 27, 2016
workout ILKB Cherry hill is the place

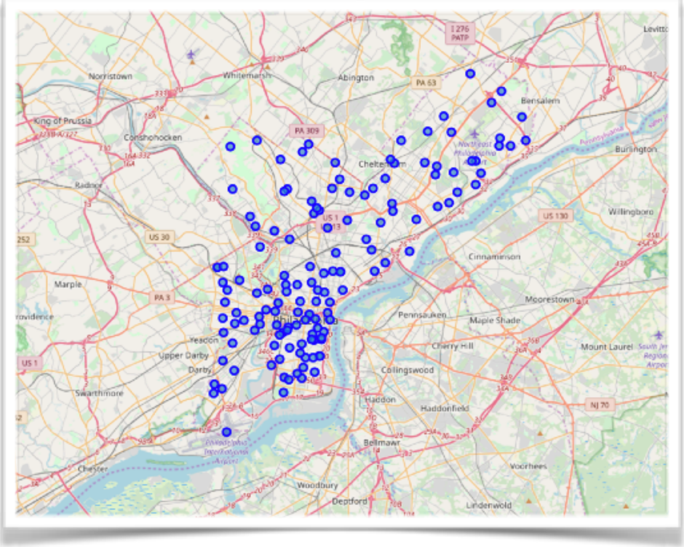
3. Reading Terminal Market 9.5
Market · View Products
51 N 12th St (btwn Arch & Filbert Sts), Philadelphia
Save
alk about the **roast pork sandwiches**, **steak subs**, and **steak sandwiches**

Foursquare Results for Philadelphia, PA

All of the neighborhood data was pulled by scraping public webpages and running the results through Nominatim to find the Latitude/Longitude of each neighborhood. I then used the Foursquare API to find the top 30 venues for each neighborhood and store them in a new DataFrame for analysis. Due to some issues with Nominatim, some of the initial neighborhood data had to be dropped but only accounted for 13% of the total number of neighborhoods and can be considered insignificant.

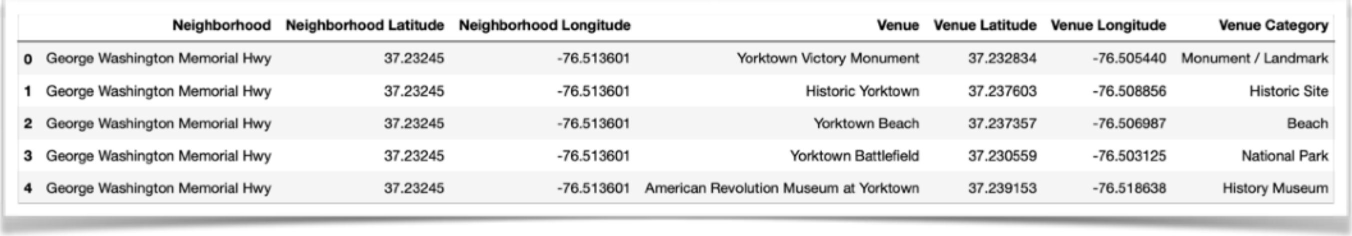
The final neighborhood data resulted in the following maps:





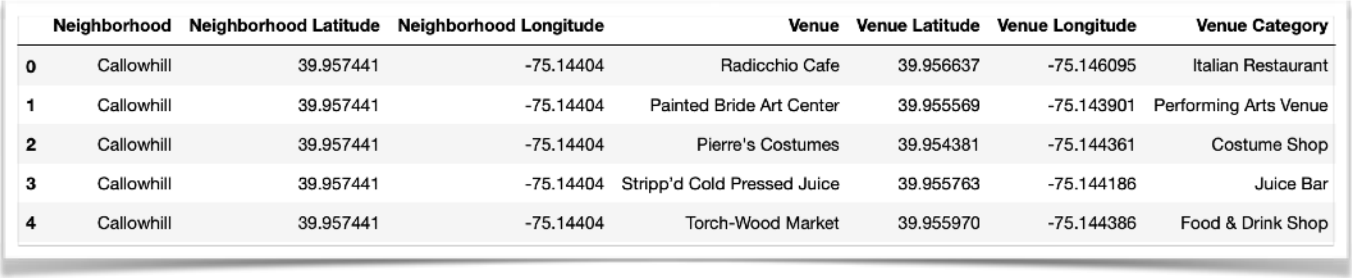
Pulling the venue data using the neighborhood data was pretty straightforward using Foursquare. In order to process the increased amount of Foursquare calls, I needed to upgrade my developer account to the Personal tier which allows for 99,500 normal calls per day. The expansion of each neighborhood DataFrame was roughly 16x the original DataFrame row size, so the Personal tier was justified.

The venue search resulted in the following DataFrames (first 5 rows shown):



	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	George Washington Memorial Hwy	37.23245	-76.513601	Yorktown Victory Monument	37.232834	-76.505440	Monument / Landmark
1	George Washington Memorial Hwy	37.23245	-76.513601	Historic Yorktown	37.237603	-76.508856	Historic Site
2	George Washington Memorial Hwy	37.23245	-76.513601	Yorktown Beach	37.237357	-76.506987	Beach
3	George Washington Memorial Hwy	37.23245	-76.513601	Yorktown Battlefield	37.230559	-76.503125	National Park
4	George Washington Memorial Hwy	37.23245	-76.513601	American Revolution Museum at Yorktown	37.239153	-76.518638	History Museum

Hampton Roads Venue DataFrame

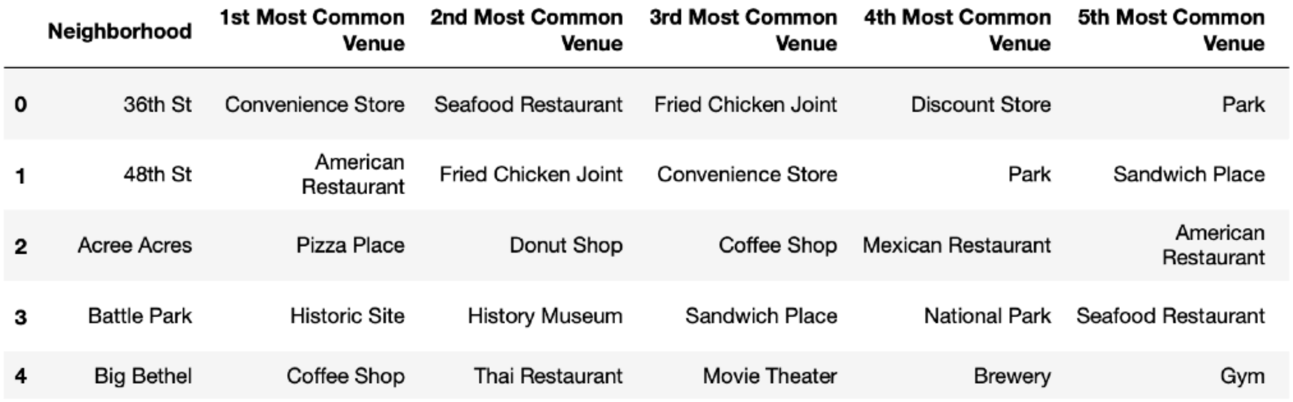


	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Callowhill	39.957441	-75.14404	Radicchio Cafe	39.956637	-75.146095	Italian Restaurant
1	Callowhill	39.957441	-75.14404	Painted Bride Art Center	39.955569	-75.143901	Performing Arts Venue
2	Callowhill	39.957441	-75.14404	Pierre's Costumes	39.954381	-75.144361	Costume Shop
3	Callowhill	39.957441	-75.14404	Stripp'd Cold Pressed Juice	39.955763	-75.144186	Juice Bar
4	Callowhill	39.957441	-75.14404	Torch-Wood Market	39.955970	-75.144386	Food & Drink Shop

Philadelphia Venue DataFrame

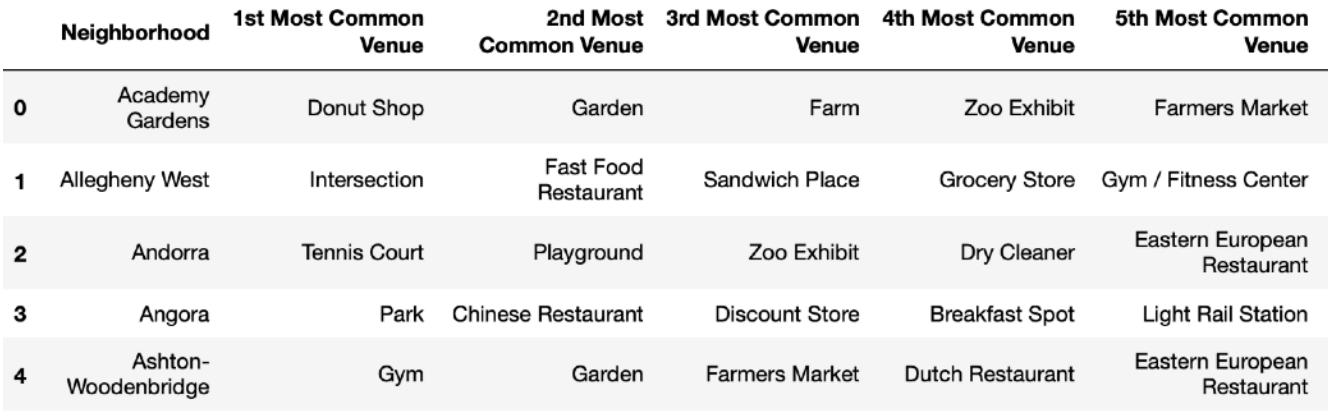
Methodology

To analyze the data that I collected, I grouped the venue DataFrames by venue category and did some one-hot encoding to determine the frequency of each category, by neighborhood, for both cities. The resulting DataFrames are below (limited to the first 5 rows):



	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	36th St	Convenience Store	Seafood Restaurant	Fried Chicken Joint	Discount Store	Park
1	48th St	American Restaurant	Fried Chicken Joint	Convenience Store	Park	Sandwich Place
2	Acree Acres	Pizza Place	Donut Shop	Coffee Shop	Mexican Restaurant	American Restaurant
3	Battle Park	Historic Site	History Museum	Sandwich Place	National Park	Seafood Restaurant
4	Big Bethel	Coffee Shop	Thai Restaurant	Movie Theater	Brewery	Gym

Hampton Roads Top Venue Categories



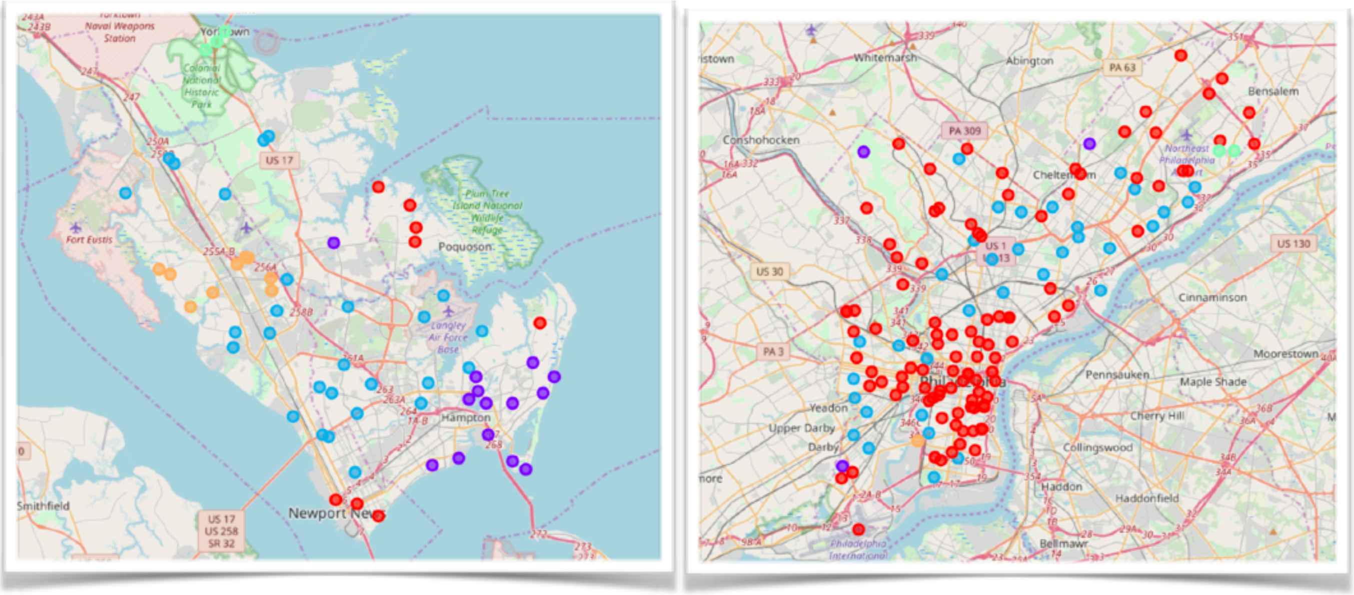
	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Academy Gardens	Donut Shop	Garden	Farm	Zoo Exhibit	Farmers Market
1	Allegheny West	Intersection	Fast Food Restaurant	Sandwich Place	Grocery Store	Gym / Fitness Center
2	Andorra	Tennis Court	Playground	Zoo Exhibit	Dry Cleaner	Eastern European Restaurant
3	Angora	Park	Chinese Restaurant	Discount Store	Breakfast Spot	Light Rail Station
4	Ashton-Woodenbridge	Gym	Garden	Farmers Market	Dutch Restaurant	Eastern European Restaurant

Philadelphia Top Venue Categories

After the top venue categories were identified, I decided the next step was to use k-means clustering to determine if there were any similar category groups/neighborhoods in each city that could help provide some insight into the problem.

Results

After the top venue category results were passed through the k-means clustering, the following city maps were developed to help provide a visual representation of the data:



Hampton Roads / Philadelphia Cluster Analysis Maps

It's fairly obvious that there are a lot more data points for Philadelphia compared to Hampton Roads. This is why I decided when I was pulling data from Foursquare to use a 5km radius for Hampton Roads vs a 500m radius for Philadelphia. I predicted that the much more urban Philadelphia would have a lot more popular restaurants within the search radius and wanted to compensate for that disparity.

Discussion

After all of the analysis was conducted, I'm not able to give specific venue brands/names that are most likely to open in the Hampton Roads area. I am, however, able to estimate based on cluster size what venue category has the potential to grow. Below is the cluster analysis for each city:

Hampton Roads	Cluster Name	Size
1	Convenience Store / Fast Food	Medium
2	Restaurant / Beach / Brewery	Medium
3	Miscellaneous	Large
4	Historic Site / Museum	Small
5	Pizza Restaurant	Medium

Philadelphia	Cluster Name	Size
1	Miscellaneous	Large
2	Playground	Small
3	Parks / Sports	Large
4	Cafe / Fast Food	Small
5	Art Gallery	Small

The venue category most likely to grow in the near future in the Hampton Roads is the Parks/Sports category. Most of the venues that were pulled from Foursquare for the Hampton Roads area, based on popularity, were restaurants, breweries, and historic sites. Also, considering that the Hampton Roads area has a lot of military bases of each service, parks and sports venues also seems to make sense.

However, having finished this project and experienced all of the roadblocks along the way, I'm convinced that there was something that I missed when planning this out that could've made the results more valuable. Perhaps an analysis of the brands would've been a better use of resources and time compared to just analyzing the venue categories of each city or the number of reviews written for each venue would weight the value of that particular venue on the recommendation. If I was given more time with this project, I would dive deeper into some of these ideas.

Conclusion

The Coursera Capstone project has been a very interesting project to work on and leverage everything that I've learned over the last few months. Hopefully all of the work that I put into this will have amounted to a relatively accurate prediction, but time will tell.