

Final Assignment - Data Sciences Capstone

Project - The Battle of the Neighborhoods



Applied Data Science Capstone by IBM/Coursera

Table of Contents

1. Introduction: Business Problem
2. Data
3. Methodology
4. Analysis
5. Results and Discussion
6. Conclusion



Introduction: Business Problem

Minneapolis–Saint Paul is a major metropolitan area built around the Mississippi, Minnesota and St. Croix rivers in east central Minnesota with a population of about 4 million currently. The area is commonly known as the Twin Cities after its two largest cities, Minneapolis, the most populous city in the state, and Saint Paul, the state capital. Minnesotans living outside of Minneapolis and Saint Paul often refer to the two together as "The Cities".

In 2017, Minneapolis-St. Paul-Bloomington, MN-WI had a population of 3.6M people with a median age of 37 and a median household income of \$76,856. The student population is about 270,000.

The population in the Twin /cities is incredibly diverse and with a high focus on healthy lifestyle. The numbers and diversity of the restaurants are notable but I'd like to check if it would be possible to open a French café or French take-out Food / market could have a place in what is currently being offered. What neighborhood could best accommodate this business? I would like to see if there's more populated neighborhoods that would be more receptive to the addition of such a food place and where the population dynamics would benefit of a French café (simple food but healthy food with French recipes).

I need also to consider the income of the population where the business will be established, and it needs to have an easy access and parking space for quick pickup. I'm looking for a sponsor to launch this kind of enterprise and need to recommend a location (neighborhood) for this business.



Description of the data and how it will be used to solve the problem

Based on definition of our problem, factors that will influence our decision are:

- Number of existing restaurants in the neighborhoods (any kind of restaurant / café)
- Number of and distance to French restaurants/Business in the neighborhoods.

I decided to use the identification of Neighborhood across Minneapolis and St Paul (not the counties). Following data sources will be needed to extract/generate the required information:

- Capture the neighborhoods from this source:
<https://www.mncompass.org/profiles/neighborhoods/minneapolis-saint-paul#!community-areas> and using this file MSP Neighborhoods_2013-2017.csv (location of neighborhood will be determined by coordinates added using geocoder)
- Number of restaurants and type as well as location in every neighborhood will be obtained using Foursquare API.

Methodology

To accomplish analysis , extract and prepare the data, the main following packages were used:

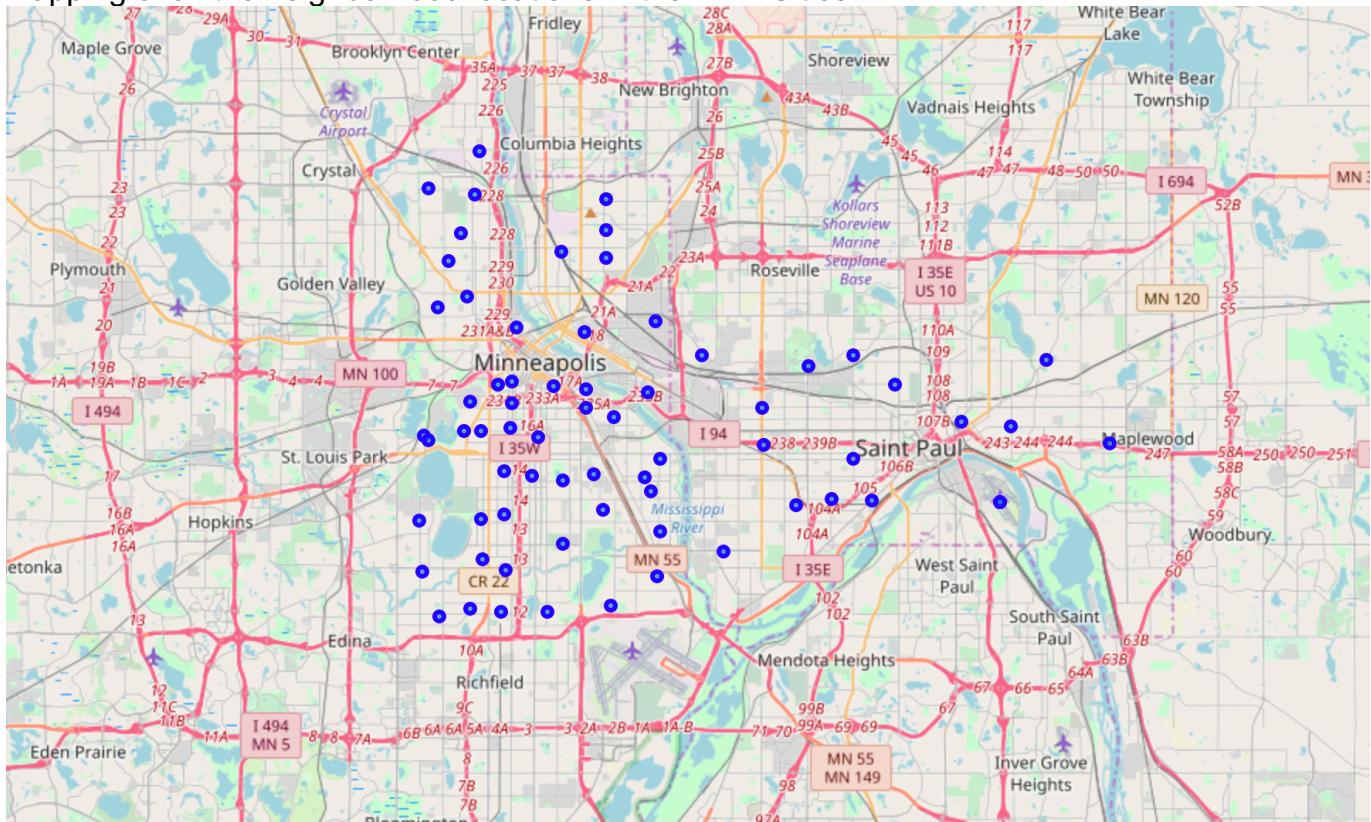
- Pandas – Library for Data Analysis
- numpy - Library to handle data in a vectorized manner
- geopy.geocoders (with Nominatim) - convert an address into latitude and longitude values
- Matplotlib – Library for Matplotlib and associated plotting modules
- Sklearn – Python Machine Learning Library
- Folium – Map rendering Library

Work Flow:

- Obtained first the file with all the neighborhoods name from a link on internet.
- Based on this file and the names of the neighborhood I extracted the coordinates (latitude and longitudes) using geocoder and create a map with folium of all the neighborhood locations.
- Once I gatherer this information in a data frame I ran Foursquare API to gather information on the various venues. I ran it first against any kind of venues and re-ran the model with only venues related to food (but I could not select only food related but it was overall build using key words such as Shop, market etc.)
- I used the Elbow method to identify the best K means value on both set s of data as mentioned above.
- Then I used Folium to create the maps to visualize the neighborhoods clusters using K=3 for both sets of data.
- Unsupervised machine learning algorithm K-mean clustering has been applied to form clusters of different categories of places located in and around the neighborhood areas. These clusters for the second scenario (only food related venue) have been reviewed and analyzed individually and compared to obtain the conclusions.

Analysis

Mapping of all the neighborhood locations in the Twin Cities



Identify the best K using the Elbow method

(1) based on all venue categories



(2) based on only food venue categories

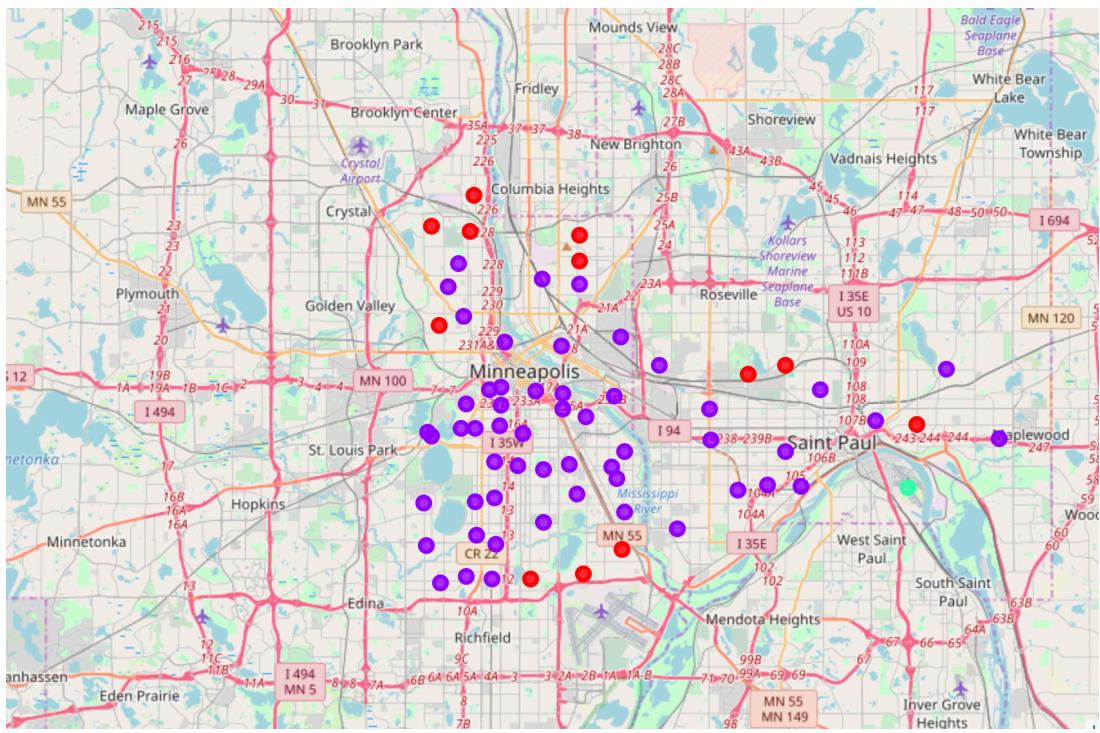


Based on the 2 results, I chose to use K=3 for both situations.

Results and Discussion

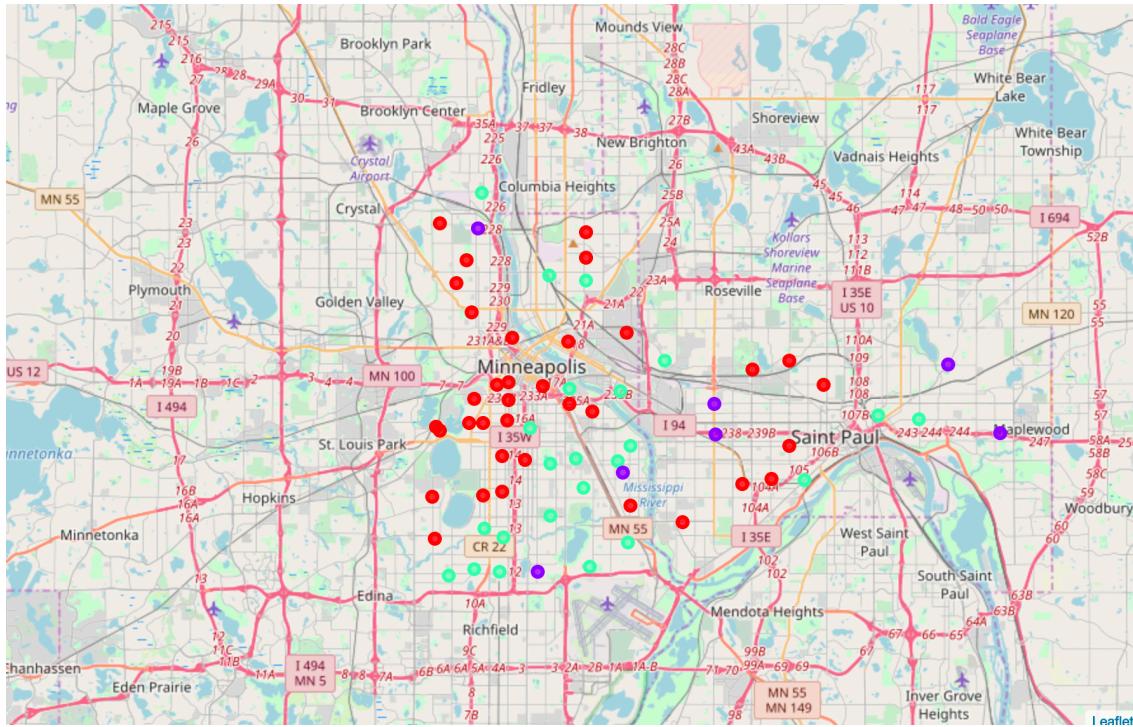
(1) based on all venue categories

We see one of the cluster with only one value and the other one with a large number. I did not select this case. Too much disparity!



(2) based on only food venue categories

This is the case I selected to obtain details on the 3 clusters.



Neighborhoods with most French restaurants:

	Neighborhood	French Restaurant
0	Armatage	0.040000
13	Downtown West	0.010000
15	East Isles	0.010000
19	Fulton	0.030303
36	Lynnurst	0.037037
53	Summit Hill	0.013333
54	Summit-University	0.025000
55	Tangletown	0.015385

Based on the 3 clusters we can note the following:

- **Cluster 0 (35 instances)** : dominance of American restaurants (although 2 of the neighborhoods with highest number of French restaurants, Fulton and Summit University) – extract below

geography	longitude	Cluster Labels	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
East Harriet	-93.291487	0	Performing Arts Venue	Spanish Restaurant	Playground	Coffee Shop	Carpet Store	Cuban Restaurant	New American Restaurant	Arts & Crafts Store	Garden
Cedar-Isles-Dean	-93.316987	0	Park	New American Restaurant	Trail	American Restaurant	Cosmetics Shop	Sushi Restaurant	Coffee Shop	Beach	Chinese Restaurant
Folwell	-93.300384	0	American Restaurant	Intersection	Asian Restaurant	Restaurant	Vietnamese Restaurant	Fish & Chips Shop	Empanada Restaurant	English Restaurant	Ethiopian Restaurant
East Isles	-93.299393	0	American Restaurant	Italian Restaurant	Asian Restaurant	Indian Restaurant	Mexican Restaurant	Thai Restaurant	French Restaurant	Eastern European Restaurant	Latin American Restaurant
Victory	-93.315190	0	American Restaurant	Ramen Restaurant	Vietnamese Restaurant	Fish & Chips Shop	Eastern European Restaurant	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Falafel Restaurant
Near North	-93.298131	0	Fast Food Restaurant	American Restaurant	Chinese Restaurant	Restaurant	Southern / Soul Food Restaurant	Vietnamese Restaurant	Eastern European Restaurant	Empanada Restaurant	English Restaurant
Jordan	-93.305963	0	American Restaurant	Asian Restaurant	Seafood Restaurant	Restaurant	Vietnamese Restaurant	Eastern European Restaurant	Empanada Restaurant	English Restaurant	Ethiopian Restaurant

- **Cluster 1 (7 instances)**: the more diversified ethnically (beside fast food, you can find many Asian restaurant - Chinese, Vietnamese, Asian, Korean, Thai - but also Turkish, Russian, Greek)

geography	longitude	Cluster Labels	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
Webber-Camden	-93.294827	1	Fast Food Restaurant	Vietnamese Restaurant	Cuban Restaurant	Eastern European Restaurant	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Falafel Restaurant	Fish & Chips Shop	Fondue Restaurant
Diamond Lake	-93.262468	1	New American Restaurant	Italian Restaurant	Asian Restaurant	Fast Food Restaurant	Vietnamese Restaurant	Fish & Chips Shop	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Falafel Restaurant
Howe	-93.215948	1	American Restaurant	Asian Restaurant	Fast Food Restaurant	Vietnamese Restaurant	Fondue Restaurant	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Falafel Restaurant	Fish & Chips Shop
Hamline-Midway	-93.166927	1	Fast Food Restaurant	Asian Restaurant	Korean Restaurant	Turkish Restaurant	Thai Restaurant	American Restaurant	Ethiopian Restaurant	New American Restaurant	Vietnamese Restaurant	Eastern European Restaurant
Union Park	-93.166244	1	Fast Food Restaurant	Chinese Restaurant	Thai Restaurant	Greek Restaurant	Asian Restaurant	Ethiopian Restaurant	Russian Restaurant	Middle Eastern Restaurant	American Restaurant	Furniture Home Store
Battle Creek-Highwood	-93.012256	1	Fast Food Restaurant	Chinese Restaurant	Korean Restaurant	Vietnamese Restaurant	Fish & Chips Shop	Eastern European Restaurant	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Falafel Restaurant
Greater East Side	-93.040384	1	Asian Restaurant	Fast Food Restaurant	Vietnamese Restaurant	Fondue Restaurant	Eastern European Restaurant	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Falafel Restaurant	Fish & Chips Shop

- Cluster 2 (23 instances):** Mexican restaurants are predominant in this cluster (but again 2 neighborhoods with highest # of French food are reported here Armatage and Lynnhurst)

geography	longitude	Cluster Labels	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 Co
Corcoran	-93.241356	2	Mexican Restaurant	Fast Food Restaurant	Japanese Restaurant	Liquor Store	Coffee Shop	Rental Car Location	Chinese Restaurant	Caribbean Restaurant	Café Rest	
Kenny	-93.296505	2	Gym / Fitness Center	Italian Restaurant	Mexican Restaurant	Fast Food Restaurant	Brewery	Garden	Furniture / Home Store	French Restaurant	Food	Di
Cooper	-93.212392	2	Mexican Restaurant	Burger Joint	Latin American Restaurant	Furniture / Home Store	Italian Restaurant	Asian Restaurant	Beach	Coffee Shop	Fast Food Restaurant	E Rest
Lind-Bohanon	-93.292166	2	Fast Food Restaurant	Food	Mexican Restaurant	Chinese Restaurant	Restaurant	Park	American Restaurant	French Restaurant	Furniture / Home Store	F Rest
Holland	-93.255927	2	Mexican Restaurant	American Restaurant	Thai Restaurant	Vietnamese Restaurant	South American Restaurant	Dive Bar	Fast Food Restaurant	Garden Center	German Restaurant	I E Rest
Wenonah	-93.234312	2	Mexican Restaurant	Fast Food Restaurant	Vietnamese Restaurant	Cuban Restaurant	Eastern European Restaurant	Empanada Restaurant	English Restaurant	Ethiopian Restaurant	Falafel Restaurant	
Tangletown	-93.280911	2	Mexican Restaurant	Chinese Restaurant	Soccer Stadium	Italian Restaurant	Other Great Outdoors	English Restaurant	Latin American Restaurant	Fast Food Restaurant	Fish & Chips Shop	Am Rest
Phillips West	-93.266298	2	Mexican Restaurant	Vietnamese Restaurant	American Restaurant	African Restaurant	Thai Restaurant	Middle Eastern Restaurant	Asian Restaurant	Caribbean Restaurant	Restaurant	Am Rest
Northrop	-93.255178	2	Chinese Restaurant	Comfort Food Restaurant	Gluten-free Restaurant	Mexican Restaurant	Italian Restaurant	Sushi Restaurant	Restaurant	Fast Food Restaurant	Falafel Restaurant	Emp Rest
Minnehaha	-93.213695	2	Seafood Restaurant	Gluten-free Restaurant	American Restaurant	Mexican Restaurant	Italian Restaurant	Chinese Restaurant	Fast Food Restaurant	Vietnamese Restaurant	Empanada Restaurant	E Rest

Conclusion

This is not conclusive at this point, but I would not recommend opening a French restaurant in the neighborhoods from cluster 1 (too diverse ethnically) but it makes more sense to me in neighborhoods from cluster 0 or cluster 2.

More research needs to be done regarding other criteria: accessibility (public transportation, parking), distance to Students locations, Revenue by household for example.

Finally, it could also be more like a market offering various food as a take out. the other option would be to look at combining French and Gluten Free options to attract different populations.