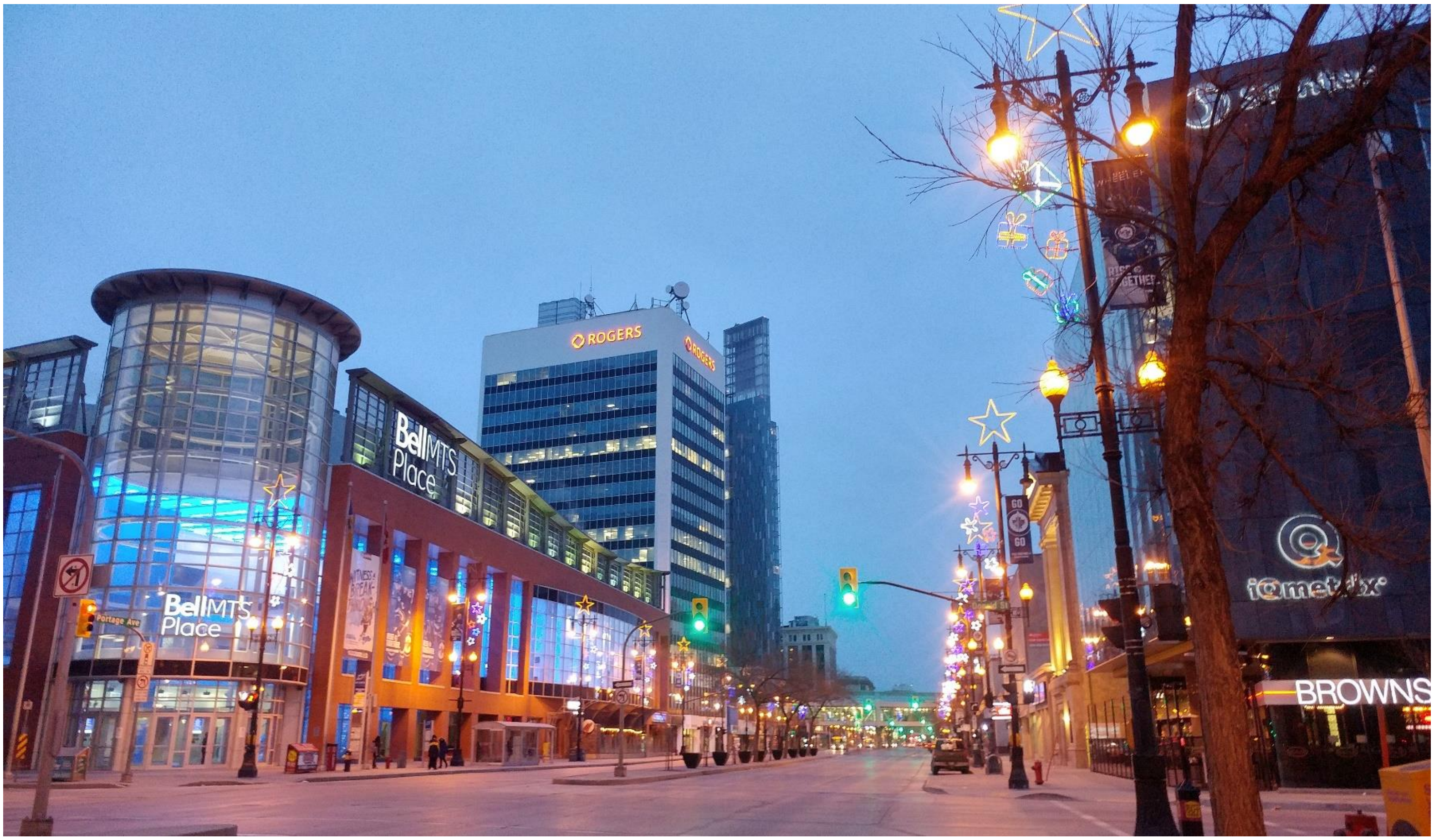
Battle of the Neighborhoods

IBM Applied Data Science Capstone



Date: December 4th, 2019

By Coursera student: Syed Faaris Razi

Table of Contents

INTRODUCTION – BUSINESS PROBLEM

This report presents findings of a Data Science project with the objectives of finding optimal neighborhood locations of launching a new Lebanese restaurant in. **Winnipeg**, a city located in the heart of Canada - Manitoba, will be the focus of this project as we explore through this city and apply various analysis to help in deciding on the right location of opening our restaurant. Stakeholders interested in launching a Lebanese/Middle Eastern type restaurant in the Peg would find this project compatible with their business decisions.

Before diving into the project, we had a few location factors to consider. We would first want to **setup near the city's hub** (Downtown Winnipeg). An important factor would be to look at **areas with good levels of personal safety** (low crime activity). Watching out for competition, we would be interested in places with **less restaurants and no Middle Eastern and Mediterranean restaurants** as these themes match a Lebanese theme. And also being **close to food markets** would make a timely regulation of fresh food items in restaurant service.

THE DATASET

The dataset for this project were acquired from open data sources on the web, you may find their links in the reference section. These datasets were preprocessed for analysis:

- Geo data of the 12 Communities in Winnipeg in Polygon-geometry (Head display - left),
- The data on annual crimes per Community (Head and Tail display - right):

	Communities	geometry
0	Assiniboine South	POLYGON ((-97.349126716065 49.80929140998, -97...
1	Downtown	POLYGON ((-97.206175382249 49.883646902492, -9...
2	Fort Garry	POLYGON ((-97.221079689052 49.814369167808, -9...
3	Inkster	POLYGON ((-97.22998265245 49.960776658412, -97...
4	Point Douglas	POLYGON ((-97.170754851168 49.932652218021, -9...
5	River East	POLYGON ((-97.126886803072 49.915029353449, -9...

	Communities	Total_Crimes
0	Assiniboine South	1339
1	Downtown	17223
2	Fort Garry	3688
3	Inkster	3005
4	Point Douglas	7656
5	River East	4847

	Communities	Total_Crimes
6	River Heights	5365
7	Seven Oaks	4040
8	St. Boniface	3639
9	St. James - Assiniboia	4129
10	St. Vital	2760
11	Transcona	2001

- Geo data of the 237 Neighborhoods in Winnipeg in Point-geometry, Latitude and Longitude values obtained with arcGIS geocoding and aligned with their Community areas (Head & Tail display):

	Neighborhoods	Community	Latitude	Longitude	geometry
0	River West Park	Assiniboine South	49.86853	-97.31940	POINT (-97.31940 49.86853)
1	Roblin Park	Assiniboine South	49.85159	-97.29022	POINT (-97.29022 49.85159)
2	Ridgewood South	Assiniboine South	49.84381	-97.29843	POINT (-97.29843 49.84381)
3	Edgeland	Assiniboine South	49.86811	-97.20859	POINT (-97.20859 49.86811)
4	Old Tuxedo	Assiniboine South	49.86997	-97.21478	POINT (-97.21478 49.86997)

	Neighborhoods	Community	Latitude	Longitude	geometry
232	Peguis	Transcona	49.90462	-97.04713	POINT (-97.04713 49.90462)
233	Griffin	Transcona	49.91322	-97.00243	POINT (-97.00243 49.91322)
234	Regent	Transcona	49.89791	-97.05693	POINT (-97.05693 49.89791)
235	Transcona Yards	Transcona	49.89122	-97.01723	POINT (-97.01723 49.89122)
236	Transcona North	Transcona	49.91284	-96.97961	POINT (-96.97961 49.91284)

- Venues surrounded around each neighborhood at 500 meter radius, that would be obtained by using Foursquare API after segmenting Communities and Neighborhoods to get our priority areas for analysis.

	Community	Neighborhoods	Latitude	Longitude	Venue	Venue_Lat	Venue_Long	Category
0	Assiniboine South	Edgeland	49.86811	-97.20859	Starbucks	49.867033	-97.211984	Coffee Shop
1	Assiniboine South	Edgeland	49.86811	-97.20859	Rady Jewish Community Centre	49.872103	-97.208858	Gym / Fitness Center
2	Assiniboine South	Edgeland	49.86811	-97.20859	Rumor's Restaurant & Comedy Club	49.867542	-97.213786	Comedy Club
3	Assiniboine South	Edgeland	49.86811	-97.20859	Safeway Tuxedo	49.866843	-97.212029	Grocery Store
4	Assiniboine South	Edgeland	49.86811	-97.20859	TD Canada Trust	49.866066	-97.212180	Bank

METHODOLOGY

With the datasets preprocessed and prepared for analysis, we will apply filtering procedures on our communities and neighborhoods to search for those that fit our business problem conditions.
We will first **limit our analysis to an area of 6km from Downtown Winnipeg**. The lat-long values of the Neighborhoods will be used in measuring their distance from Downtown with the Haversine formula, that finds the great-circle distance between two points on earth with lat-long values. This will help filter out Neighborhoods not within the radius surrounding Downtown.

The next step would be to ignore areas with high crime density. With the crime-levels per Community data we will use a **choropleth map to distinguish communities of varying crime severity levels**. Lat-Long values belonging in certain Communities will then be filtered out and selected accordingly.

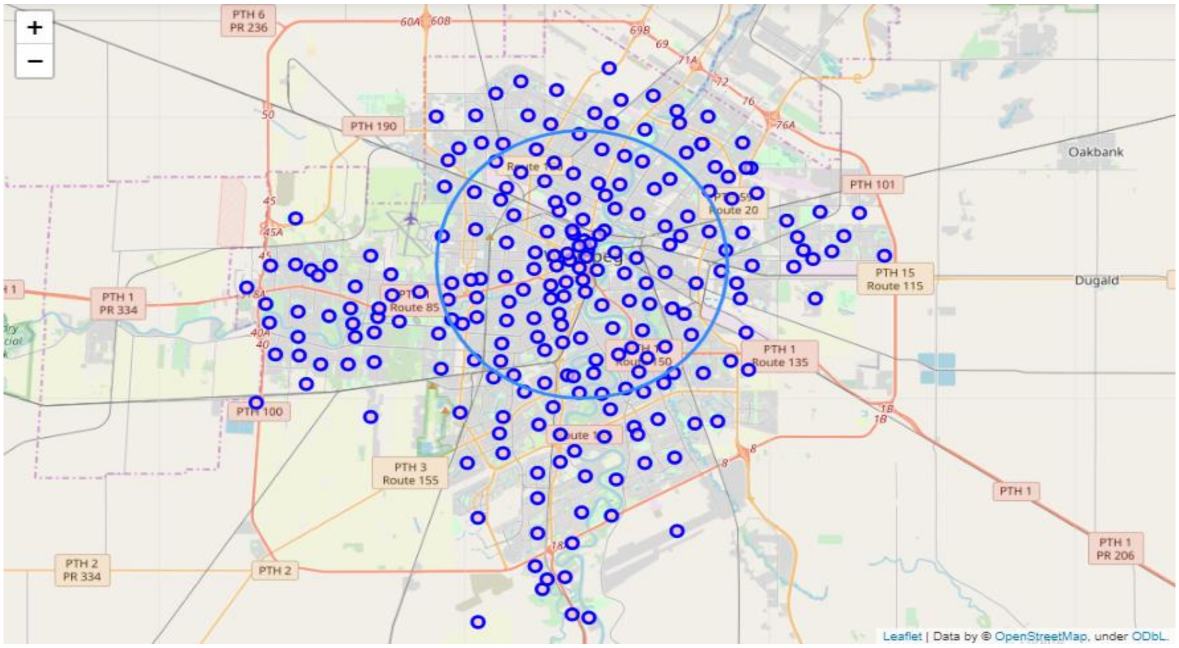
In the 3rd step we use **Foursquare API in identifying venues at a radius of 500m from each Neighborhood**. Here we will explore the most common venues within Neighborhoods, doing exploratory analysis by comparing Neighborhoods **with high or low density in restaurants and food markets**, with the condition of ***detecting areas with Lebanese, Middle Eastern or Mediterranian restaurants**.

In our final step, **K-means clustering** will be used in detecting similar trends between Neighborhoods based on their most common venues. Neighborhoods will be clustered into groups similar to each other, from which we will see the pros and cons of setting up in Neighborhoods that follow a certain trend of venues.

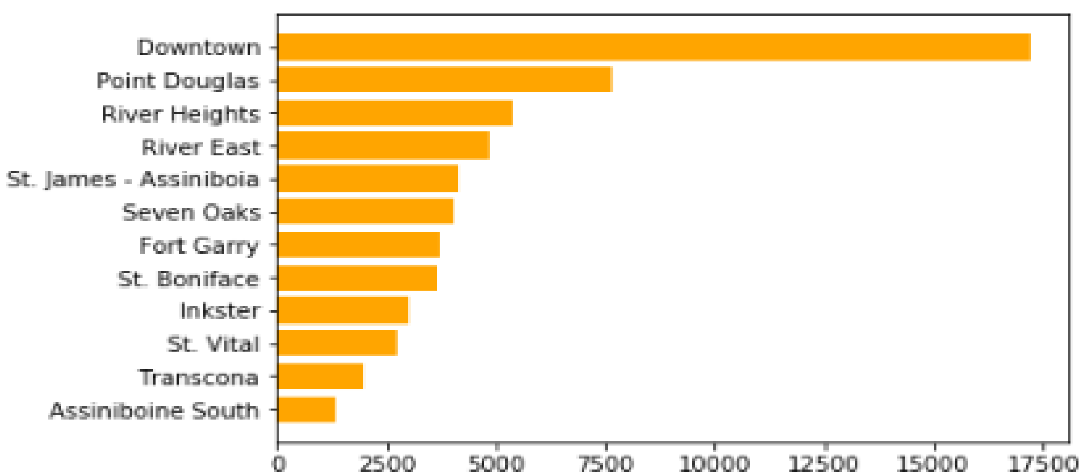
ANALYSIS

P.1) Segmentation of Neighborhoods

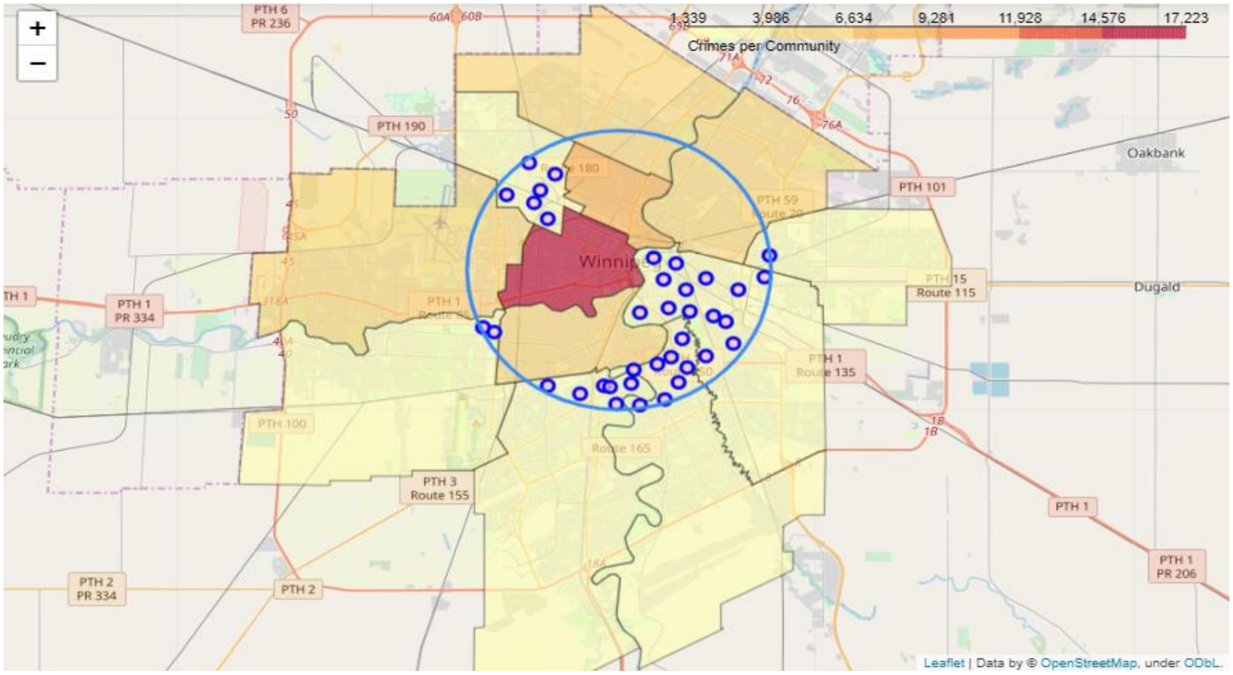
In this section we dive into segmenting our datasets with respect to the business problem requirements. We start by focusing on Neighborhoods within 6km radius of Downtown Winnipeg. Here’s a folium map of the neighborhoods in Winnipeg, with a circle boundary that describes the region to be focused on (centering Downtown):



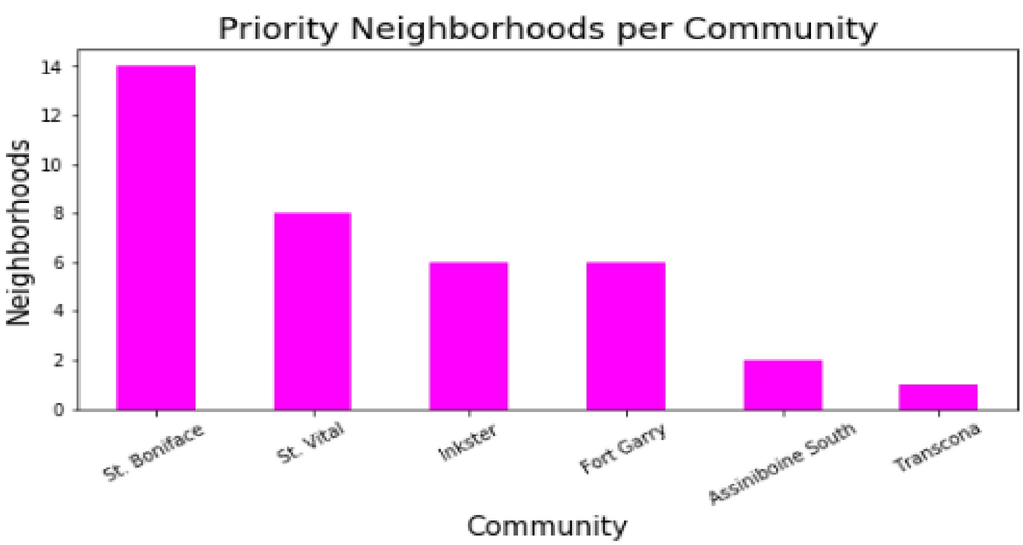
The Haversine formula is applied for filtering out Neighborhoods outside of the circular region. We then check which areas have significant crime-severity levels (horizontal barplot), the top 6 communities have significant crime-levels:



Downtown, **Point Douglas**, **River Heights**, **River East**, **St. James – Assiniboia** and **Seven Oaks** are filtered out. These communities are presented by the choropleth map below, where regions have more color and filtered Neighborhoods (empty regions). And as mentioned, Neighborhoods outside of the circular region were filtered out. The following Neighborhood points will be used for our analysis:

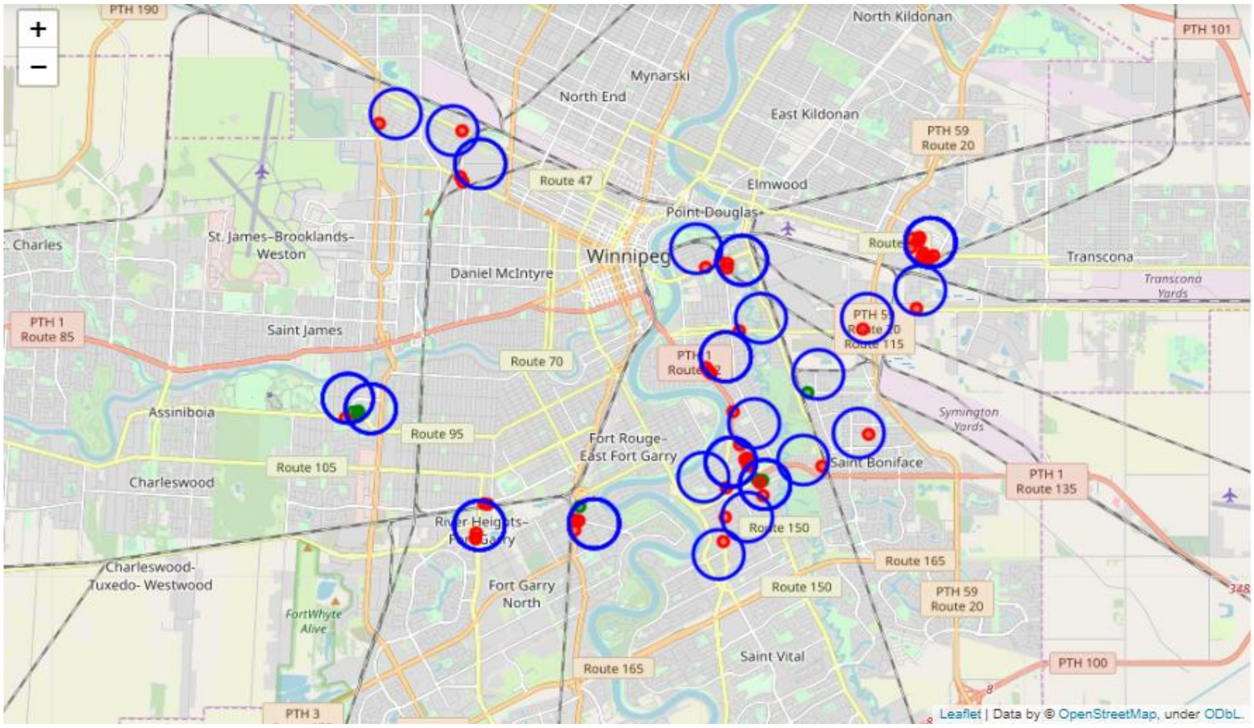


The counts of the priority Neighborhoods for analysis are summarized by Community in the bar plot below:



P.2) Density distribution of Restaurants and Markets

We now focus our attention on analyzing which areas have high or low restaurants and market densities. With the data of venues within (500m of) each Neighborhood collected from Foursquare API, we can see the densities within each Neighborhood with a folium map; red markers describe Restaurants, green for Suppliers (food markets) and the blue rings are centered on each Neighborhood at 500m radius:



Out of all the restaurants, we find that there lies one Mediterranean restaurant within one of our Neighborhoods:

```
Venues list identified by foursquare API have restaurants of type:
0 - Lebanese, 0 - Middle Eastern, and 1 - Mediterranean
```

The location of the Mediterranean restaurant found in our venues list is:

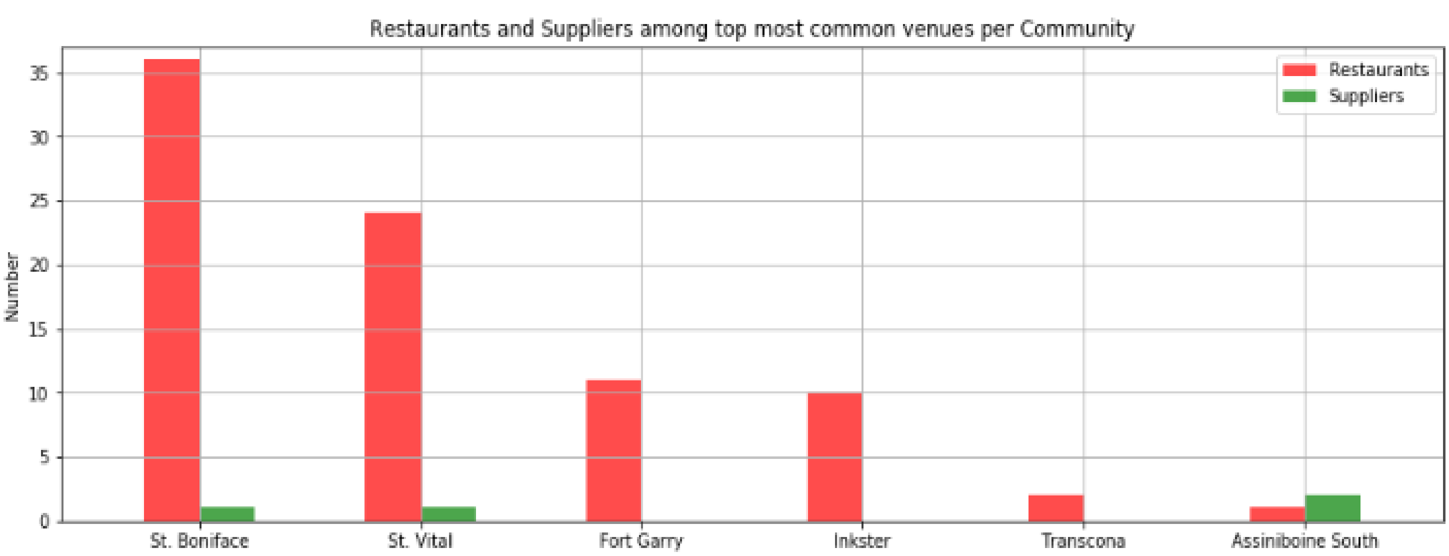
```
medit_res
```

	Community	Neighborhoods	Latitude	Longitude	Venue	Venue_Lat	Venue_Long	Category
51	Fort Garry	Parker	49.84718	-97.17913	Piazza Di Nardi	49.851063	-97.177872	Mediterranean Restaurant

Now we we develop a table listing the top 10 most common venues per Neighborhood. First we apply one-hot encoding that assigns dummy variable values on all venue categories, then we find the mean of all the encoded values. We then get some dataframe manipulation with pandas done, and here we have the desired table (showing the first 10 rows and 12/14 columns):

	Community	Neighborhoods	Latitude	Longitude	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 M Comm Ver
0	Assiniboine South	Edgeland	49.868110	-97.20859	Grocery Store	Café	Gym / Fitness Center	Bank	Pet Store	Pharmacy	Playground	Shipp St
1	Assiniboine South	Old Tuxedo	49.869970	-97.21478	Grocery Store	Gym / Fitness Center	Shipping Store	Greek Restaurant	Liquor Store	Comedy Club	Pet Store	Pharm
2	Fort Garry	Parker	49.847180	-97.17913	Gym / Fitness Center	Italian Restaurant	Coffee Shop	Gourmet Shop	Mediterranean Restaurant	Sandwich Place	Breakfast Spot	Boutic
3	Fort Garry	Point Road	49.847500	-97.14810	Coffee Shop	Pizza Place	Sporting Goods Shop	Breakfast Spot	Fast Food Restaurant	Spa	Ramen Restaurant	Convenier St
4	Fort Garry	South Pointe	49.846777	-97.14436	Athletics & Sports	Tennis Stadium	Wings Joint	Field	Convenience Store	Cosmetics Shop	Curling Ice	Departm St
5	Fort Garry	Wildwood	49.848230	-97.13313	Construction & Landscaping	Park	Fast Food Restaurant	Convenience Store	Cosmetics Shop	Curling Ice	Department Store	Di
6	Inkster	Brooklands	49.920970	-97.20179	Convenience Store	Recreation Center	Sandwich Place	Wings Joint	Coffee Shop	Construction & Landscaping	Cosmetics Shop	Curling
7	Inkster	Pacific Industrial	49.912010	-97.17893	Coffee Shop	Clothing Store	Office	Burger Joint	Filipino Restaurant	Wings Joint	Event Space	Curling
8	Inkster	Shaughnessy Park	49.929130	-97.17461	Gym / Fitness Center	Comedy Club	Convenience Store	Cosmetics Shop	Curling Ice	Department Store	Diner	Disco St
9	Inkster	Weston	49.917970	-97.18641	Butcher	Sports Bar	Pizza Place	Wings Joint	Event Space	Construction & Landscaping	Convenience Store	Cosme St

From the above table we can absorb a lot of information about venue trends of the Neighborhoods, neat. Then we go on to do some exploratory analysis, we wanted to look at Restaurants and Suppliers that are among the top most common venues in their respective Neighborhoods. We sum up all restaurant and food-market type categories within most common venues per Neighborhood, and then we get a double bar plot grouped by Communities:



P.3) K-Means clustering

As a final part of our analysis we work with K-means clustering, an unsupervised learning algorithm that helps connect datasets into cluster groups without the need of labelled or known data. So Neighborhoods will be clustered into groups based on their top most common venue trends, which would give us some insights on venue trend similarities between Neighborhoods. We specify the number of clusters K = 5, and using the sci-kit learn's K-Means clustering module we get the following cluster groups:

- **Cluster 0**: Most Neighborhoods from every Community (all from Assiniboine South and the one from Transcona) fall in this cluster. These Neighborhoods can be described as business areas, since common venues include Coffee shops, Banks, Grocery stores, Pharmacies, Fitness Centers, and a mix of other venues. (Showing only the first 9 out of 26 rows/Neighborhoods, columns 1-15)

```
common_clusters[common_clusters.Cluster_group == 0].iloc[:,1:15] #Cluster 0
```

	Community	Neighborhoods	Latitude	Longitude	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 M Comn Ve
0	Assiniboine South	Edgeland	49.868110	-97.20859	Grocery Store	Café	Gym / Fitness Center	Bank	Pet Store	Pharmacy	Playground	Shipp S
1	St. Vital	Varennes	49.858540	-97.11100	Convenience Store	Gas Station	Thai Restaurant	Sushi Restaurant	Burger Joint	Sandwich Place	Fast Food Restaurant	Wings J
2	St. Vital	St. George	49.848680	-97.10658	Coffee Shop	Board Shop	Pizza Place	Wings Joint	Fast Food Restaurant	Convenience Store	Cosmetics Shop	Curling
3	St. Vital	Norberry	49.841960	-97.11431	Gym / Fitness Center	Bakery	Breakfast Spot	Fast Food Restaurant	Field	Cosmetics Shop	Curling Ice	Departm S
4	St. Vital	Glenwood	49.865460	-97.10469	Argentinian Restaurant	Karaoke Bar	Sporting Goods Shop	Maternity Clinic	Wings Joint	Field	Cosmetics Shop	Curling
5	St. Vital	Elm Park	49.855830	-97.11844	Gym / Fitness Center	Soccer Field	Convenience Store	Bus Station	Mexican Restaurant	Fried Chicken Joint	Electronics Store	Golf Cou
6	St. Vital	Alpine Place	49.854600	-97.10174	Grocery Store	Curling Ice	Golf Course	Hotel	Coffee Shop	Gas Station	Fried Chicken Joint	Bak
7	St. Boniface	Windsor Park	49.863720	-97.07636	Convenience Store	Restaurant	Automotive Shop	Event Space	Wings Joint	Fast Food Restaurant	Cosmetics Shop	Curling
8	St. Boniface	Tissot	49.894730	-97.10799	Gym / Fitness Center	Spa	Bowling Alley	Breakfast Spot	Coffee Shop	French Restaurant	Ice Cream Shop	Sandw Pl

- **Cluster 1**: Describes a unique Neighborhood in St. Vital where the top most common venue would be Pet stores, followed by restaurants and several stores.

```
common_clusters[common_clusters.Cluster_group == 1].iloc[:,1:15] #Cluster 1
```

	Community	Neighborhoods	Latitude	Longitude	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	10tl Co
26	St. Vital	Victoria Crescent	49.83966	-97.12818	Pet Store	Wings Joint	Fast Food Restaurant	Convenience Store	Cosmetics Shop	Curling Ice	Department Store	Diner	Discount Store	Elec

- **Cluster 2**: This group has top common venues as Parks, Construction sites, followed by restaurants, Convenience stores and various other stores.

```
common_clusters[common_clusters.Cluster_group == 2].iloc[:,1:15] #Cluster 2
```

	Community	Neighborhoods	Latitude	Longitude	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 Mc Comm Ven
27	St. Boniface	Norwood West	49.87543	-97.12797	Park	Wings Joint	Fast Food Restaurant	Convenience Store	Cosmetics Shop	Curling Ice	Department Store	Diner	Discou Stc
28	Fort Garry	Wildwood	49.84823	-97.13313	Construction & Landscaping	Park	Fast Food Restaurant	Convenience Store	Cosmetics Shop	Curling Ice	Department Store	Diner	Discou Stc
29	St. Boniface	North St. Boniface	49.89881	-97.12040	Park	Coffee Shop	Italian Restaurant	Fast Food Restaurant	Convenience Store	Cosmetics Shop	Curling Ice	Department Store	Dir

- **Cluster 3**: Shaughnessy Park has a unique trend. We see that top venues revolve around Gym/Fitness Center, followed by Comedy club and stores.

```
common_clusters[common_clusters.Cluster_group == 3].iloc[:,1:15] #Cluster 3
```

	Community	Neighborhoods	Latitude	Longitude	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	Con V
30	Inkster	Shaughnessy Park	49.92913	-97.17461	Gym / Fitness Center	Comedy Club	Convenience Store	Cosmetics Shop	Curling Ice	Department Store	Diner	Discount Store	Electronics Store	I S

- **Cluster 4**: Another unique neighborhood trend in St.Vital, where Bus Stations are the 1st most common venue, followed by fields, restaurants and stores.

```
common_clusters[common_clusters.Cluster_group == 4].iloc[:,1:15] #Cluster 4
```

	Community	Neighborhoods	Latitude	Longitude	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	10tl Co
31	St. Vital	Kingston Crescent	49.85368	-97.13147	Bus Station	Field	Fast Food Restaurant	Convenience Store	Cosmetics Shop	Curling Ice	Department Store	Diner	Discount Store	Elec

RESULTS AND DISCUSSION

Now that we have reached the end of our analysis, we have gathered interesting results from our selected Communities and Neighborhoods. With the many valuable tools like the geocoder library, scikit learn's K-Means clustering, folium maps and foursquare API, we have exercised data exploration and analysis with our business problem conditions, and have found the following:

- One Mediterranian restaurant was found in Parker (Fort Garry), where as other neighborhoods don't have similar restaurant themes to Lebanese.

- The community St. Bonafice dominates in the density of restaurants among it's top most common venues, followed by St. Vital. Where as Assiniboine South ranks the lowest in this case.

- Assiniboine South has the highest density of food markets among it's top most common venues. St. Bonafice and St. Vital have food markets in very low density compared to restaurants among their most common venues, where as other Neighborhoods have none among their common venues.

- Some Neighborhoods belonging to all the Communities fall under cluster 0, where all of Assiniboine South's Neighboods are in.

From the above insights, we can see that the Assiniboine South community would have the least competition for the launch of a new restaurant, as it has Grocery stores among it's Neighborhood's top most common venues, with Greek Restaurant being the 4th most common venue in the neighborhood - Old Tuxedo. The top 3 competitive Neighborhoods would be of those in St. Bonafice, St.Vital and Fort Garry. Not only do most Neighborhoods have average restaurant density in St.Bonafice and St.Vital, they are also close to each other and accumulate to a high density level in restaurants with less supply assets (groceries/markets) among their most common venues. Looking at the Neighborhoods in Fort Garry, the Mediterranian restaurant would clash with our Lebanese theme, posing as competition. The Neighborhoods in Inkster and one in Transcona have an average density in restaurants as common venues, however there aren't any supply assets among their common venues. So these Communities would also pose as competition.

We can say that the Neighborhoods in Assiniboine South would seem to be the best candidates for opening a restaurant from our analysis. With K-means clustering we gethered information on similarities in the most common venues between Neighborhoods. An advantage to opening a restaurant in Assiniboine South is that it belongs to the cluster group that identifies as a business area, where you would be surrounded by Banks, Shopping Malls and Playgrounds. There would be a good frequency of potential customers around such places.

Some drawbacks of this analysis — Due to limited data, the focus of our analysis involved only factors such as personal safety per location (crime levels per community) and venues obtained from Foursquare data. We're nearing the end of 2019, the crimes data obtained from Winnipeg Police's annual report is based on 2018. Crime levels per community might have changed significantly this year but the annual report for 2019 or any up-to date Winnipeg crime report is yet to be presented. There are many other factors that can affect the success of businesses, such as land price, distance of venues from public transport stations, population of potential customers in a neighborhood/community etc.

The recommended Neighborhood/Community in this project should only be looked at as a starting point for more detailed analysis. Having all other relevant factors met, we would have more realistic and favorable insights that may fit all of our business problem conditions in opening a restaurant.

CONCLUSION

The purpose of this project was to identify Neighborhoods in Winnipeg near Downtown with good safety levels, low density of restaurants (with no similar themes as our restaurant) and near food markets that would help stakeholders in deciding on the ideal location for opening a new Lebanese restaurant. Calculating distances of Neighborhoods using the Haversine formula was a valuable tool in segmenting Neighborhoods. Data on crimes per community obtained from Winnipeg Police's Annual report gave us an idea of the levels of personal safety among communities. Venues data obtained from Foursquare API helped identify the most common venues near Neighborhoods, from which we brought exploratory analysis of Neighborhoods and Communities based on density distribution of restaurants and food markets. And finally applying K-means clustering for finding similar venue trends between Neighborhoods gave us an idea of what other benefits a candidate Neighborhood/Community would have for us in opening a restaurant based on it's venue trends.

This project gave us some important preliminary information on the possibilities of opening a new restaurant around a recommended area. There was a detailed discussion on the drawbacks of having limited data for analysis, where with more factors presented for our analysis would bring more realistic results and a clearer picture of the performance of a potential business in a chosen area.

This project was a real eye-opener for me, I feel rewarded with the efforts, time and money spent. I got a good exposure of how Data Science tools can be used in solving real life situations that would have an impact on business performance. I would highly recommend anyone to get the same Data Science experience in this program. A big thanks to the Coursera Team and fellow Students!

REFERENCES

1. Geospatial data for Winnipeg Communities: https://data.winnipeg.ca/City-Planning/Community-Characterization-Area/d9gq-xua6

2. Geospatial data for Winnipeg Neighborhoods: https://data.winnipeg.ca/City-Planning/Neighbourhood/fen6-iygi

3. Crimes per Community data table reference from Winnipeg Police's 2018 annual report (pg. 13/38): https://winnipeg.ca/police/AnnualReports/2018/2018_wps_annual_report_english.pdf

4. Haversine Formula definition: https://en.wikipedia.org/wiki/Haversine_formula

5. Foursquare API: https://developer.foursquare.com/

6. K-Means Clustering: https://en.wikipedia.org/wiki/K-means_clustering

7. Project codes – Github: https://github.com/SyedFaarisRazi/Coursera_Capstone/blob/master/Capstone%20-%20Battle%20of%20the%20Neighborhoods%20-%20Faaris.ipynb