EVALUATION OF THE BEST NEIGHBORHOODS TO OPEN A NEW HEALTHY FOOD STORE IN THE CITY OF TORONTO, CANADA

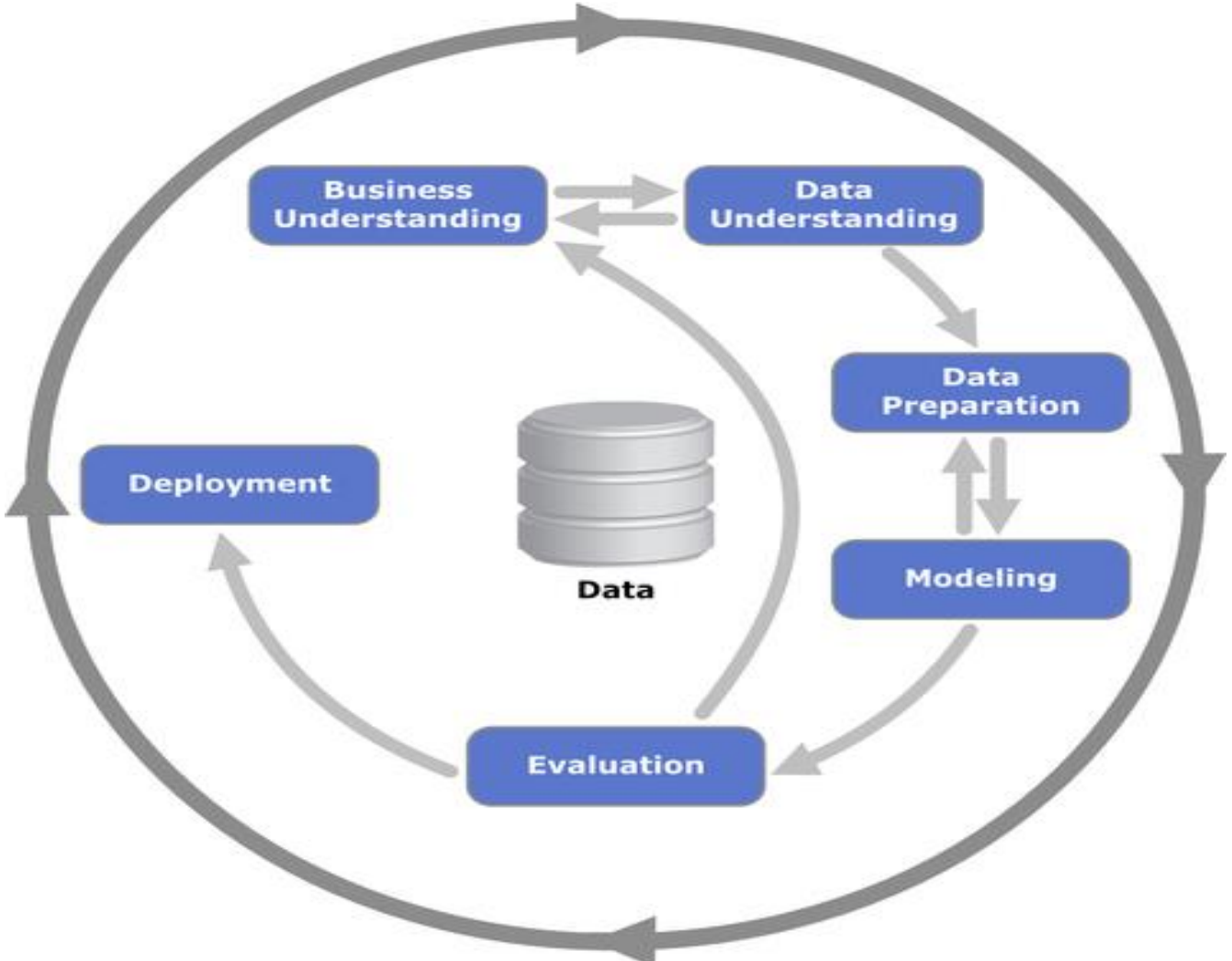
Health and Wellness food Market projections

Objectives

- Evaluating the neighborhood venues demands Exploring and classifying the venues of a neighborhood through geospatial data is an efficient way to evaluate all the services that such community has available.

- Additional considerations to open a new Healthy Food Store The study will also include the number of healthy food-related restaurants /venues in order to get insights about the lifestyle of the neighbors.

Approach and Methodology



CRISP-DM Methodology

1. **Business understanding.**
2. **Data Understanding**
3. **Data Preparation**
4. **Modeling**
5. **Deployment**

CRISP-DM Methodology

- **Business understanding:** Analysis of the problem to be solved. In our case, to get information about the best places to open new Healthy Food and wellness restaurants in Toronto.
-
- **Data Understanding:** The data needed for performing the study. In our case the data is the demographic and geographic data of the boroughs and neighborhoods in Toronto, the postal code and the respective geospatial coordinates for each neighborhood and the list of Venues for each location.
-
- **Data Preparation:** Prepare pandas data frame with the neighborhoods, boroughs and postal code (pandas data frame with the geospatial coordinates for each neighborhood, and the list of venues using the FoursquareAPI)
-
- **Modeling:** Create a map of Toronto's neighborhoods, a machine learning algorithm based on a Kmean approach to segment and cluster the list of venues in each neighborhood. Also we need to evaluate the model creating a map of the clusters and evaluate the accuracy of the kmean algorithm during the clustering process.
-
- **Deployment:** According to the frequency of the Healthy Food Stores and other venues like gyms or parks in each neighborhood, make a conclusion about which one would be the best to open a new store

DATA

- Geographic data about Toronto’s neighborhoods: https://en.wikipedia.org/wiki/List_of_postal_codes_of _Canada:_M
- Geospatial Coordinates information: Geocoder platform https://geocoder.readthedocs.io/index.html
- List of Venues available at the FoursquareAPI: (https://developer.foursquare.com/docs/api/venues/e xplore)
- Other tools and library used: Pandas library, Numpy library, Foursquare API, Folium library, Geopy library

Results

- Data frame generated with Panda

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	M5E	Downtown Toronto	Berczy Park	43.64531	-79.37368
1	M6K	West Toronto	Brockton, Parkdale Village, Exhibition Place	43.63941	-79.42676
2	M7Y	East Toronto	Business reply mail Processing Centre, South C...	43.64869	-79.38544
3	M5G	Downtown Toronto	Central Bay Street	43.65609	-79.38493
4	M6G	Downtown Toronto	Christie	43.66878	-79.42071
5	M4Y	Downtown Toronto	Church and Wellesley	43.66659	-79.38130
6	M5V	Downtown Toronto	CN Tower, King and Spadina, Railway Lands, Har...	43.64082	-79.39956
7	M5L	Downtown Toronto	Commerce Court, Victoria Hotel	43.64823	-79.37890
8	M4S	Central Toronto	Davisville	43.70340	-79.38596
9	M4P	Central Toronto	Davisville North	43.71276	-79.38851
10	M6H	West Toronto	Dufferin, Dovercourt Village	43.66509	-79.43871
11	M5X	Downtown Toronto	First Canadian Place, Underground city	43.64828	-79.38146
12	M5P	Central Toronto	Forest Hill North & West, Forest Hill Road Park	43.69479	-79.41440
13	M5B	Downtown Toronto	Garden District, Ryerson	43.65736	-79.37818

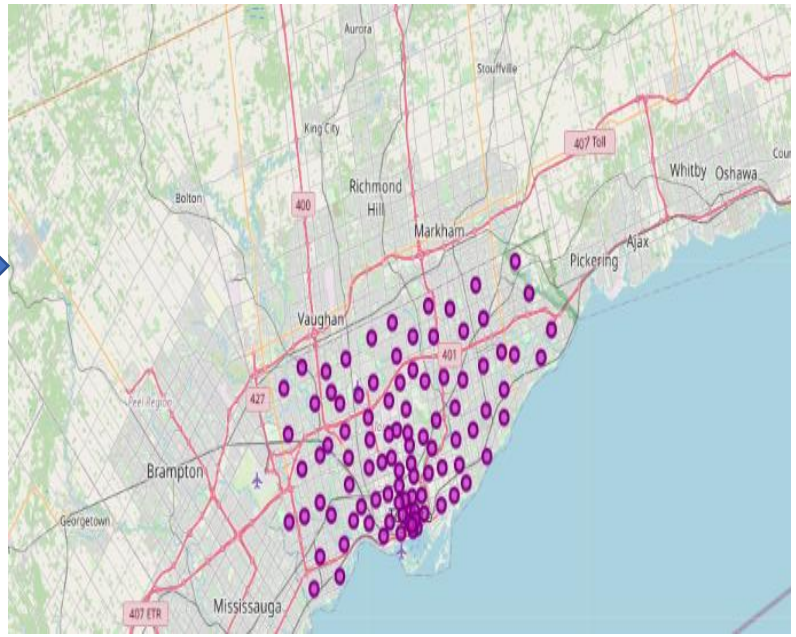
Results

- Code to generate a map of Toronto

```
address = 'Toronto, ON' #Use geopy library to get the latitude and longitude values of Toronto

geolocator = Nominatim(user_agent="toronto_explorer")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinate of Toronto are {}, {}.'.format(latitude, longitude))
```

```
The geograpical coordinate of Toronto are 43.6534817, -79.3839347.
```

```
map_toronto = folium.Map(location=[latitude, longitude], zoom_start=8)

# add markers to map
for lat, lng, borough, neighborhood in zip(df_toronto_coor['Latitude'], df_toronto_coor['Longitude'], df_toronto_coor['Borough'], df_toronto_coor['Neighborhood']):
    label = '{}, {}'.format(neighborhood, borough)
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=5,
        popup=label,
        color='purple',
        fill=True,
        fill_color='#c931cc',
        fill_opacity=0.7,
        parse_html=False).add_to(map_toronto)

map_toronto
```



Results

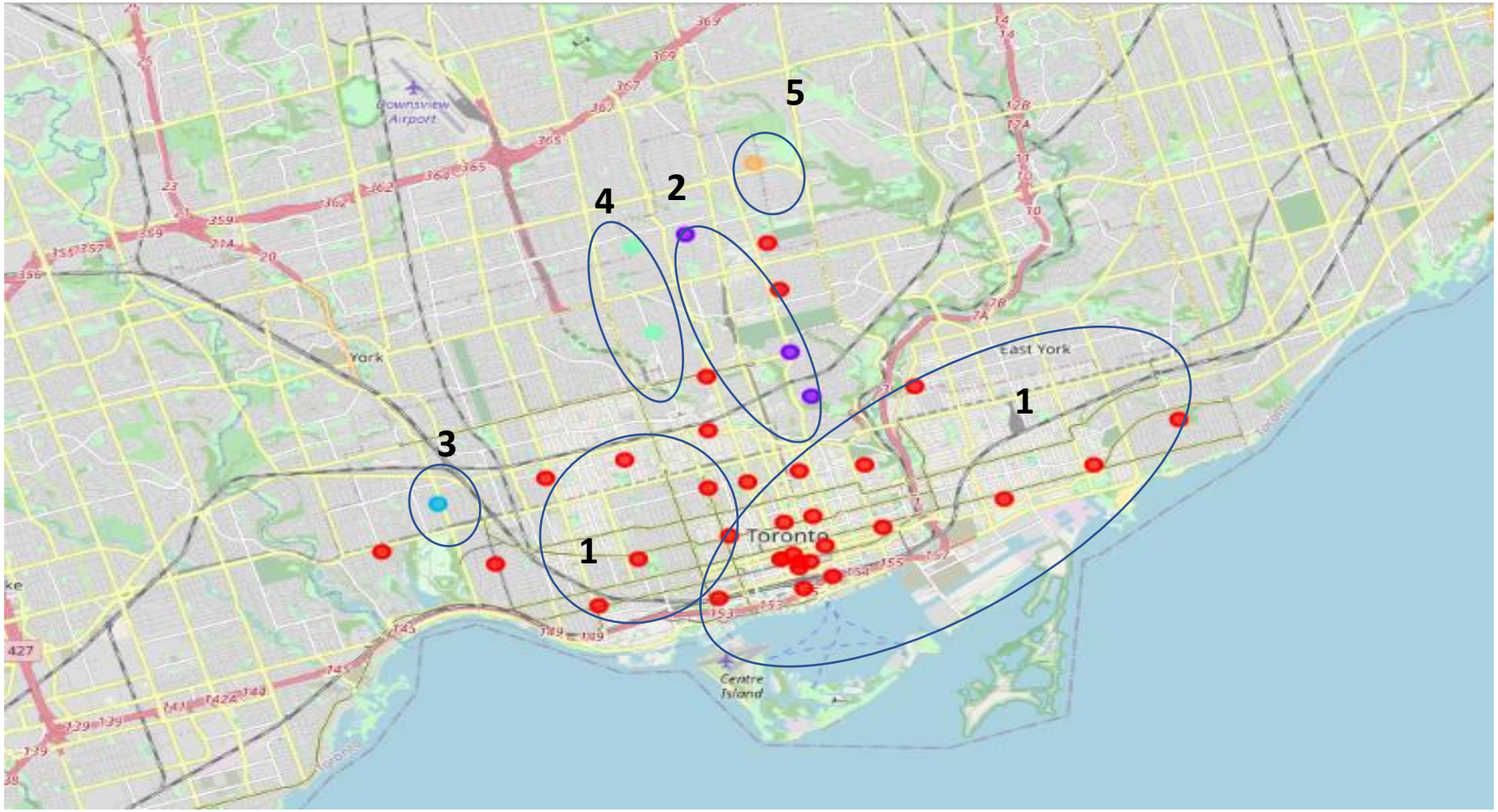
- Using the Explore of Foursquare query to create a list of venues

```
# Get the venues for each neighborhood
toronto_venues = getNearbyVenues(
                                   postalCodes=df_toronto_ngh['Postal Code'],
                                   boroughs=df_toronto_ngh['Borough'],
                                   neighbourhoods=df_toronto_ngh['Neighborhood'],
                                   latitudes=df_toronto_ngh['Latitude'],
                                   longitudes=df_toronto_ngh['Longitude']
                                  )

toronto_venues.head()
```

	Postal Code	Borough	Neighborhood	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	M5E	Downtown Toronto	Berczy Park	43.64531	-79.37368	The Keg Steakhouse + Bar - Esplanade	43.646712	-79.374768	Restaurant
1	M5E	Downtown Toronto	Berczy Park	43.64531	-79.37368	Fresh On Front	43.647815	-79.374453	Vegetarian / Vegan Restaurant
2	M5E	Downtown Toronto	Berczy Park	43.64531	-79.37368	Meridian Hall	43.646292	-79.376022	Concert Hall
3	M5E	Downtown Toronto	Berczy Park	43.64531	-79.37368	Goose Island Brewhouse	43.647329	-79.373541	Beer Bar
4	M5E	Downtown Toronto	Berczy Park	43.64531	-79.37368	LCBO	43.642944	-79.372440	Liquor Store

Results

- Machine learning (K-means) approach to cluster the neighborhoods



**All red points are cluster 1

Discusions

- **Cluster 1 (cluster label = 0) (Red color):** The Cluster 1 is formed by neighborhood with a high density of restaurants of a diverse range of options. It includes food from different places of the world, being the Asian food restaurants extremely popular. This cluster is probably has a lot of competitors but also a big population density

```
#examining the cluster 1
toronto_merged.loc[toronto_merged['Cluster Labels'] == 0]
```

	Postal Code	Borough	Neighborhood	Latitude	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
0	M5E	Downtown Toronto	Berczy Park	43.64531	-79.37368	0	Coffee Shop	Restaurant	Hotel	Breakfast Spot	Beer Bar	Italian Restaurant	Seafood Restaurant
1	M6K	West Toronto	Brockton, Parkdale Village, Exhibition Place	43.63941	-79.42676	0	Café	Coffee Shop	Bar	Restaurant	Nightclub	Bakery	Sandwich Place
2	M7Y	East Toronto	Business reply mail Processing Centre, South C...	43.64869	-79.38544	0	Coffee Shop	Hotel	Asian Restaurant	Restaurant	Italian Restaurant	Café	American Restaurant
3	M5G	Downtown Toronto	Central Bay Street	43.65609	-79.38493	0	Coffee Shop	Clothing Store	Plaza	Hotel	Middle Eastern Restaurant	Electronics Store	Italian Restaurant
4	M6G	Downtown Toronto	Christie	43.66878	-79.42071	0	Café	Grocery Store	Italian Restaurant	Baby Store	Coffee Shop	Athletics & Sports	Candy Store
5	M4Y	Downtown Toronto	Church and Wellesley	43.66659	-79.38130	0	Coffee Shop	Japanese Restaurant	Restaurant	Gay Bar	Sushi Restaurant	Men's Store	Bubble Tea Shop
6	M5V	Downtown Toronto	CN Tower, King and Spadina, Railway Lands, Har...	43.64082	-79.39956	0	Coffee Shop	Italian Restaurant	Café	Bar	Park	Gym / Fitness Center	Speakeasy

Discusions

- **Cluster 2 (cluster label = 1) (Purple color):** In this cluster we can see the venues of three neighborhoods where there are not so much restaurants but there are a lot of parks, playgrounds, gyms and sport venues that can be a good sign to invest in the healthy food industry.

```
#cluster 2
toronto_merged.loc[toronto_merged['Cluster Labels'] == 1]
```

	Postal Code	Borough	Neighborhood	Latitude	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
20	M4T	Central Toronto	Moore Park, Summerhill East	43.69066	-79.38356	1	Playground	Gym	Park	Tennis Court	Women's Store	Distribution Center	Farm	Falafel Restaurant
21	M4R	Central Toronto	North Toronto West, Lawrence Park	43.71452	-79.40696	1	Gym Pool	Playground	Park	Garden	Women's Store	Distribution Center	Farm	Falafel Restaurant
26	M4W	Downtown Toronto	Rosedale	43.68190	-79.37850	1	Japanese Restaurant	Playground	Park	Bike Trail	Dog Run	Farmers Market	Farm	Falafel Restaurant

Discusions

- **Cluster 3 (cluster label =2) (Blue color):** The Cluster 3 is formed by one neighborhood in which there are some restaurants and some options of leisure. Maybe is not best place to invest because low density population and a lot of competitors

```
#cluster 3
toronto_merged.loc[toronto_merged['Cluster Labels'] == 2]
```

	Postal Code	Borough	Neighborhood	Latitude	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
15	M6P	West Toronto	High Park, The Junction South	43.65994	-79.46302	2	Sandwich Place	Park	Residential Building (Apartment / Condo)	Women's Store	Farm	Falafel Restaurant	Event Space	Ethiopian Restaurant

Discusions

- **Cluster 4 (cluster label = 3) (Green color):** The Cluster 4 is formed by 2 neighborhood and it is characterized by having some places to eat and stores at the same proportion. There are not so much food options so this neighborhood could be a great investment opportunity

```
#cluster 4
toronto_merged.loc[toronto_merged['Cluster Labels'] == 3]
```

	Postal Code	Borough	Neighborhood	Latitude	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
12	M5P	Central Toronto	Forest Hill North & West, Forest Hill Road Park	43.69479	-79.41440	3	Home Service	Women's Store	Dog Run	Fast Food Restaurant	Farmers Market	Farm	Falafel Restaurant	Event Space
27	M5N	Central Toronto	Roselawn	43.71194	-79.41912	3	Home Service	Clothing Store	Women's Store	Dog Run	Fast Food Restaurant	Farmers Market	Farm	Falafel Restaurant

Discusions

- **Cluster 5 (cluster label = 4) (Orange color):** The Cluster 5 is also formed by one neighborhood in which there are few restaurants and few options of leisure. Probably there is a good place to open a healthy food restaurant but near to the park or dog run.

```
#cluster 5
toronto_merged.loc[toronto_merged['Cluster Labels'] == 4]
```

	Postal Code	Borough	Neighborhood	Latitude	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
18	M4N	Central Toronto	Lawrence Park	43.72898	-79.39173	4	Park	Electronics Store	Women's Store	Dog Run	Fast Food Restaurant	Farmers Market	Farm	Falafel Restaurant

Conclusions

- After preparing, modelling and analyzing the data, it can be concluded that the k-mean algorithm of the machine learning approach is very useful to find segmentation in real life application that involves studying the market, competitors, target, segments, etc.
- We found there are 2 neighborhoods(cluster 4 and 2) with great potential to open a healthy food and wellness restaurant or facilities and two more neighborhoods in which, even with a lot of competitors, the target customers can be found because of the great population density (cluster 1 and 5).