The test project should provide tools and insights to assist the decision making for the following

Business Problem:

Company considering opening several restaurants in Toronto. There are many factors affecting the business decision and calculations. Amongst them

- Location of restaurants (Neighborhoods and Clusters of Neighborhoods)
- What type of cuisine (perhaps Italian?) would be most suitable
- Various levels of competition, i.e. existing vendors of the same type (i.e. Italian) vs. general competition from all restaurants etc.

Data:

1. The project will need a list of Toronto Neighborhoods (zip) which will be scraped from https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

	Postcode	Borough	Neighborhood
0	M1B	Scarborough	Malvern, Rouge
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union
2	M1E	Scarborough	Morningside, Guildwood, West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae

2. The project will need a geo data (using http://cocl.us/Geospatial_data)

	Postal Code	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476

3. The project will need a list of Foursquare labels (venue categories for and all subcategories for restaurant venues (Food) which will be obtained programmatically via Foursquare API https://developer.foursquare.com/docs/resources/categories as json


```
{
  "meta": {
    "code": 200,
    "requestId": "5854a3664434b94a0959cfb7"
  },
  "notifications": [
    {
      "item": {
        "unreadCount": 13
      },
      "type": "notificationTray"
    }
  ],
  "response": {
    "categories": [
      {
        "categories": [
          {
            "categories": [],
            "icon": {
              "prefix": "https://ss3.4sqi.net/img/categories_v2/arts_entertainment/default_",
              "suffix": ".png"
            },
            "id": "56aa371be4b08b9a8d5734db",
            "name": "Amphitheater",
            "pluralName": "Amphitheaters",
            "shortName": "Amphitheater"
          },
          {
            "categories": [],
            "icon": {
              "prefix": "https://ss3.4sqi.net/img/categories_v2/arts_entertainment/aquarium_",
              "suffix": ".png"
            },
            "id": "4fceea171983d5d06c3e9823",
            "name": "Aquarium",
            "pluralName": "Aquariums",
            "shortName": "Aquarium"
          },
          {
            "categories": [],
            "icon": {
              "prefix": "https://ss3.4sqi.net/img/categories_v2/arts_entertainment/arcade_",
              "suffix": ".png"
            },
            "id": "4bf58dd8d48988d1e1931735",
```

4. Finally, the project will rely on data queried from Foursquare API for the venues in all Food categories returned in json format like this


```
{
  "meta": {
    "code": 200,
    "requestId": "5ac51d7e6a607143d811cecb"
  },
  "response": {
    "venues": [
      {
        "id": "5642aef9498e51025cf4a7a5",
        "name": "Mr. Purple",
        "location": {
          "address": "180 Orchard St",
          "crossStreet": "btwn Houston & Stanton St",
          "lat": 40.72173744277209,
          "lng": -73.98800687282996,
          "labeledLatLngs": [
            {
              "label": "display",
              "lat": 40.72173744277209,
              "lng": -73.98800687282996
            }
          ],
          "distance": 8,
          "postalCode": "10002",
          "cc": "US",
          "city": "New York",
          "state": "NY",
          "country": "United States",
          "formattedAddress": [
            "180 Orchard St (btwn Houston & Stanton St)",
            "New York, NY 10002",
            "United States"
          ]
        },
        "categories": [
          {
            "id": "4bf58dd8d48988d1d5941735",
            "name": "Hotel Bar",
            "pluralName": "Hotel Bars",
            "shortName": "Hotel Bar",
            "icon": {
              "prefix": "https://ss3.4sqi.net/img/categories_v2/travel/hotel_bar_",
              "suffix": ".png"
            },
            "primary": true
          }
        ],
        "venuePage": {
          "id": "150747252"
        }
      }
```

Methodology

All, the coordinate points obtained in the data frame and visualized with folium map plotter with neighborhoods superimposed on top of map of greater Toronto, ON, Canada.

After, plotting the data is filtered out the coordinates which are not inside the main city of Toronto

Use the respective latitudes and longitudes to get the corresponding nearby venues for all the neighborhoods

Exploratory data analysis includes checking size of the resulting dataframe, number of venues were returned for each neighborhood, grouping rows by neighborhood and by taking the mean of the frequency of occurrence of each food (restaurant) category and resulting dataframe and display the top 10 venues for each neighborhood.

Generate one-hot vectors for the venue types in all the neighborhoods and generate most common venues.

Use the list of most common venues (frequency) to get the top10 venues

Then, K-means clustering is used to cluster the neighborhoods based on the similarity of the venue types.

Create new dataframe that includes the cluster as well as the top 10 venues for each neighborhood.

Explore and visualize clusters.

Data is further augmented with isolating the frequencies for a specific (desired) venue categories and presenting them (join) to the neighborhood clusters

Results

The resulting datasets indicate several neighborhoods with lower concentration of specific restaurants making them prospective candidate for potential new location for new restaurants. Additionally, tools are provided to easily examine various types of possible cuisine and how it would stack up against the current competition

Discussion

In general, the research shows that Toronto as a major world capital is very diverse and has a wide variety of Food venues. Further research is likely needed to estimate various factors like commercial real estate leasing data, neighborhood density, median incomes, trends, parking and transportation issues etc. But the provided research is a good first step for concentrating the search more specifically.

Conclusion

Hopefully this was indicative of my efforts and interest in acquiring further proficiency in Data Science.