Exploring Toronto - Best Neighbourhoods for Best Chinese Food

Introduction
1. Description of the problem and background
2. Cleaning, Wrangling, Feature Selection, Feature Engineering
3. Exploratory Data Analysis
4. Predictive Modeling
5. Results and Discussion
6. Conclusion

```
# Import libraries
from PIL import Image
from os.path import join, isfile, dirname
from wordcloud import WordCloud, STOPWORDS
import numpy as np, pandas as pd, matplotlib as mpl, matplotlib.pyplot as plt, seaborn as sns, folium, sqlite3 # etc
```

Introduction: Exploring Toronto - Best Neighbourhoods for Best Chinese Food

In short, in this project we will go through step by step process to make a decision where is the for ease of access to chinese restaurants.

As a part of the first MS Capstone Project, we will work on the data science project about the objectives of the food establishments in a recent dataset of Toronto. The dataset is from the City of Toronto Open Data Portal. The dataset includes information about food establishments in Toronto and includes information about their location (lat/long), and information about their capacity. In an ideal world, we would have information on the demand for each food establishment, but we do not have that information. We will focus on the demand for Chinese food.

In this project, I will go through all the processes in a step by step manner from problem description, data preparation to free analysis. In the end, I will provide my conclusions and discuss the results. I will try to make it easier for the reader to follow the technical diagrams which are hard to optimize for experience.

1. Description of the problem and background

Concerning living in Toronto and eating out on chinese restaurants

Toronto is the capital city of the province of Ontario. It is the most populous city in Canada. Toronto is a multicultural city. Approximately 52% of the population in the Greater Toronto Area (GTA) are immigrants, including Canadian citizens. Approximately 1 million people in Toronto are Chinese, Korean, Vietnamese people (Wikipedia, Cumulative, accessed 2020). Chinese people are the second largest ethnic group in Toronto. Chinese people are the second largest ethnic group in Toronto. Chinese people are the second largest ethnic group in Toronto.

```
# Create map of Toronto using latitude and longitude.
map_toronto = folium.Map(location=[latitude, longitude])
# add markers to map
for lat, lng, neighborhood in zip(df_toronto['Latitude'], df_toronto['Longitude'], df_toronto['Neighborhood']):
    label = '{}'.format(neighborhood)
    folium.CircleMarker(
        [lat, lng],
        radius=5,
        fill=True,
        color='blue',
        fill_color='cyan',
        fill_opacity=0.7,
        parse_html=False).add_to(map_toronto)
map_toronto
```

2.2 Relationship between neighborhood and chinese restaurants

First we will extract the neighborhoods and chinese restaurant from the above toronto dataframe for further analysis

```
# Define Foursquare Credentials
CLIENT_ID = 'YOUR_FOURSQUARE_ID'
CLIENT_SECRET = 'YOUR_FOURSQUARE_SECRET'
VERSION = '20180605'
# Chinese Restaurants
print('Chinese Restaurants, Latitude, Longitude, Venue Category, Venue Name, Venue Address,Distance')
LIMIT = 100
radius = 500
for index, row in df_toronto.iterrows():
    try:
    # Foursquare API
    search_query = 'Chinese Restaurant'
    print('.',end='')
    url = 'https://api.foursquare.com/v2/venues/search?client_id={}&client_secret={}&ll={},{}&v={}&query={}&radius={}&limit={}'.format(CLIENT_ID, CLIENT_SECRET, row['Latitude'], row['Longitude'], VERSION, search_query, radius, LIMIT)
    results = requests.get(url).ison()['response']['venues']
    for venue in results:
        print(row['Latitude'],row['Longitude'],search_query,venue['name'],venue['location']['address'],venue['location']['distance'],sep=',')
    except:
        print('Error processing {}'.format(row['Neighborhood']))
```

```
# Analysis of North and South of Toronto Chinese Restaurants,of Latitude,Longitude,Venue Category,Venue Name,Venue Address,Distance
df_chinese = pd.read_csv('Chinese_Restaurant_Toronto.csv',header=None,sep=',')
df_chinese.columns=['Neighborhood',
'Latitude',
'Longitude',
'Venue Category',
'Venue Name',
'Venue Address',
'Distance']
df_chinese.shape
```

```
(23,6)
```

```
# Analysis of North and South of Toronto Chinese Restaurants,of Latitude,Longitude,Venue Category,Venue Name,Venue Address,Distance
df_chinese.drop_duplicates(subset=['Neighborhood', 'Venue Name'], inplace=True)
df_chinese.shape
```

```
(23,6)
```

	Neighborhood	Latitude	Longitude	Venue Category	Venue Name	Venue Address	Distance
0	Wexford / Maryvale	43.7542828	-79.295754	Chinese Restaurant	Bamboo		0.000000
1	Wexford / Maryvale	43.7542828	-79.295754	Chinese Restaurant	Golden Dragon		0.000000
2	Wexford / Maryvale	43.7542828	-79.295754	Chinese Restaurant	Chinese Restaurant		0.000000
3	Wexford / Maryvale	43.7542828	-79.295754	Chinese Restaurant	Chinese Restaurant		0.000000
4	Bathurst Manor, Wilson Heights, Downsview North	43.7543011	-79.442259	Chinese Restaurant	Little Shop		0.000000
5	Bathurst Manor, Wilson Heights, Downsview North	43.7543011	-79.442259	Chinese Restaurant	Chinese Restaurant		0.000000
6	Northwood Park, York University, Finch West	43.7537413	-79.506944	Chinese Restaurant	Chinese Restaurant		0.000000
7	Northwood Park, York University, Finch West	43.7537413	-79.506944	Chinese Restaurant	Chinese Restaurant		0.000000
8	Downsview	43.7370665	-79.465763	Chinese Restaurant	Chinese Restaurant		0.000000
9	Humber Bay	43.6382095	-79.498509	Chinese Restaurant	Chinese Restaurant		0.000000
10	Humber Bay	43.6382095	-79.498509	Chinese Restaurant	Chinese Restaurant		0.000000
11	Humber Bay	43.6382095	-79.498509	Chinese Restaurant	Chinese Restaurant		0.000000
12	Humber Bay	43.6382095	-79.498509	Chinese Restaurant	Chinese Restaurant		0.000000
13	Humber Bay	43.6382095	-79.498509	Chinese Restaurant	Chinese Restaurant		0.000000
14	Humber Bay	43.6382095	-79.498509	Chinese Restaurant	Chinese Restaurant		0.000000
15	Humber Bay	43.6382095	-79.498509	Chinese Restaurant	Chinese Restaurant		0.000000
16	Humber Bay	43.6382095	-79.498509	Chinese Restaurant	Chinese Restaurant		0.000000
17	Humber Bay	43.6382095	-79.498509	Chinese Restaurant	Chinese Restaurant		0.000000
18	Humber Bay	43.6382095	-79.498509	Chinese Restaurant	Chinese Restaurant		0.000000
19	Humber Bay	43.6382095	-79.498509	Chinese Restaurant	Chinese Restaurant		0.000000
20	Humber Bay	43.6382095	-79.498509	Chinese Restaurant	Chinese Restaurant		0.000000
21	Humber Bay	43.6382095	-79.498509	Chinese Restaurant	Chinese Restaurant		0.000000
22	Humber Bay	43.6382095	-79.498509	Chinese Restaurant	Chinese Restaurant		0.000000

23 rows × 6 columns

```
df_chinese.groupby('Neighborhood').count()
```

Neighborhood	Latitude	Longitude	Venue Category	Venue Name	Venue Address	Distance
Bathurst Manor, Wilson Heights, Downsview North	2	2	2	2	0	2
Downsview	1	1	1	1	0	1
Humber Bay	17	17	17	17	0	17
Northwood Park, York University, Finch West	2	2	2	2	0	2
Wexford / Maryvale	4	4	4	4	0	4

```
df_chinese.drop(['Venue Address'],axis=1,inplace=True)
df_chinese.shape
```

```
(23,5)
```

```
# Analysis of North and South of Toronto Chinese Restaurants,of Latitude,Longitude,Venue Category,Venue Name,Distance
df_chinese.rename(columns={'Venue Category':'Venue Type'},inplace=True)
df_chinese.groupby('Venue Type').count
```

```
<bmethod_DataFrame.groupby of                                Neighborhood  Latitude  Longitude  Venue Type  Venue Name  Distance
Venue Type
Chinese Restaurant                                 23        23         23          23          23        23>
```

```
# Analysis of North and South of Toronto Chinese Restaurants,of Latitude,Longitude,Venue Type,Venue Name,Distance
df_chinese.shape
```

```
(23,5)
```

```
# Analysis of North and South of Toronto Chinese Restaurants,of Latitude,Longitude,Venue Type,Venue Name,Distance
df_chinese.groupby('Neighborhood').count()
```

Neighborhood	Latitude	Longitude	Venue Type	Venue Name	Distance
Bathurst Manor, Wilson Heights, Downsview North	2	2	2	2	2
Downsview	1	1	1	1	1
Humber Bay	17	17	17	17	17
Northwood Park, York University, Finch West	2	2	2	2	2
Wexford / Maryvale	4	4	4	4	4

```
# Analysis of North and South of Toronto Chinese Restaurants,of Latitude,Longitude,Venue Type,Venue Name,Distance
df_chinese.groupby('Venue Type').count()
```

Venue Type	Neighborhood	Latitude	Longitude	Venue Name	Distance
Chinese Restaurant	23	23	23	23	23

```
# import folium library
import folium
# import MarkerCluster and MiniMap plugins
from folium.plugins import MarkerCluster, MiniMap

mapOfCities = folium.Map(location=[43.6532, -79.3832], zoom_start=10, width='100%', height='70%')

# Instead of the usual circle, create a custom icon using a piece of text instead.
# Here a dice icon with pips ranging from 1 to 6.
for lat, lon, pips in zip(df_clean['latitude'], df_clean['longitude'], df_clean['TOL_MAX']):
    folium.Marker([lat, lon], icon=folium.DivIcon(html=f'<div style="font-size: 24pt">{chr(9855+pips)}</div>')
                  popup=f'{pips} pips').add_to(mapOfCities)

# Add a MiniMap
MiniMap(width=150, height=130).add_to(mapOfCities)

# Show the map
mapOfCities
```

```
# import folium library
import folium
# import MarkerCluster and MiniMap plugins
from folium.plugins import MarkerCluster, MiniMap

mapOfCities = folium.Map(location=[43.6532, -79.3832], zoom_start=10, width='100%', height='70%')

# Instead of the usual circle, create a custom icon using a piece of text instead.
# Here a dice icon with pips ranging from 1 to 6.
for lat, lon, pips in zip(df_clean['latitude'], df_clean['longitude'], df_clean['TOL_MAX']):
    folium.Marker([lat, lon], icon=folium.DivIcon(html=f'<div style="font-size: 24pt">{chr(9855+pips)}</div>')
                  popup=f'{pips} pips').add_to(mapOfCities)

# Add a MiniMap
MiniMap(width=150, height=130).add_to(mapOfCities)

# Show the map
mapOfCities
```

Cluster 7
Scarborough

```
df_cluster7.get_item(0)
```

```
Neighborhood                 East Birchmount Park, Ionview, Kennedy Park
Borough                      Scarborough
Cluster Labels                           7
Name: 0, dtype: object
```

Scarborough is the city's largest, in high density of both chinese population and chinese restaurants.

```
# import folium library
import folium
# import MarkerCluster and MiniMap plugins
from folium.plugins import MarkerCluster, MiniMap

mapOfCities = folium.Map(location=[43.6532, -79.3832], zoom_start=10, width='100%', height='70%')

# Instead of the usual circle, create a custom icon using a piece of text instead.
# Here a dice icon with pips ranging from 1 to 6.
for lat, lon, pips in zip(df_clean['latitude'], df_clean['longitude'], df_clean['TOL_MAX']):
    folium.Marker([lat, lon], icon=folium.DivIcon(html=f'<div style="font-size: 24pt">{chr(9855+pips)}</div>')
                  popup=f'{pips} pips').add_to(mapOfCities)

# Add a MiniMap
MiniMap(width=150, height=130).add_to(mapOfCities)

# Show the map
mapOfCities
```

```
# import folium library
import folium
# import MarkerCluster and MiniMap plugins
from folium.plugins import MarkerCluster, MiniMap

mapOfCities = folium.Map(location=[43.6532, -79.3832], zoom_start=10, width='100%', height='70%')

# Instead of the usual circle, create a custom icon using a piece of text instead.
# Here a dice icon with pips ranging from 1 to 6.
for lat, lon, pips in zip(df_clean['latitude'], df_clean['longitude'], df_clean['TOL_MAX']):
    folium.Marker([lat, lon], icon=folium.DivIcon(html=f'<div style="font-size: 24pt">{chr(9855+pips)}</div>')
                  popup=f'{pips} pips').add_to(mapOfCities)

# Add a MiniMap
MiniMap(width=150, height=130).add_to(mapOfCities)

# Show the map
mapOfCities
```

4.3 Best spot and a not-so-good cluster intersection

We have found two areas where the ideal spot and the cluster of chinese cuisine seem to be optimal spots for maximum amount of chinese restaurants.

Out of these four cities, we have into interesting intersections between clusters, where the density of chinese restaurants is highest/lowest and psychological dimensions are maximized/minimized: North Toronto and Etobicoke Toronto cluster.

Intersecting circles

```
# import folium library
import folium
# import MarkerCluster and MiniMap plugins
from folium.plugins import MarkerCluster, MiniMap

mapOfCities = folium.Map(location=[43.6532, -79.3832], zoom_start=10, width='100%', height='70%')

# Instead of the usual circle, create a custom icon using a piece of text instead.
# Here a dice icon with pips ranging from 1 to 6.
for lat, lon, pips in zip(df_clean['latitude'], df_clean['longitude'], df_clean['TOL_MAX']):
    folium.Marker([lat, lon], icon=folium.DivIcon(html=f'<div style="font-size: 24pt">{chr(9855+pips)}</div>')
                  popup=f'{pips} pips').add_to(mapOfCities)

# Add a MiniMap
MiniMap(width=150, height=130).add_to(mapOfCities)

# Show the map
mapOfCities
```

```
# import folium library
import folium
# import MarkerCluster and MiniMap plugins
from folium.plugins import MarkerCluster, MiniMap

mapOfCities = folium.Map(location=[43.6532, -79.3832], zoom_start=10, width='100%', height='70%')

# Instead of the usual circle, create a custom icon using a piece of text instead.
# Here a dice icon with pips ranging from 1 to 6.
for lat, lon, pips in zip(df_clean['latitude'], df_clean['longitude'], df_clean['TOL_MAX']):
    folium.Marker([lat, lon], icon=folium.DivIcon(html=f'<div style="font-size: 24pt">{chr(9855+pips)}</div>')
                  popup=f'{pips} pips').add_to(mapOfCities)

# Add a MiniMap
MiniMap(width=150, height=130).add_to(mapOfCities)

# Show the map
mapOfCities
```


North Lake and scarborough intersection is the footpaths to A+. It is called quality of life.
Ethnic Expansion Toronto Small and Etobicoke Toronto Large have similar in the medium quality type.

5. Results and Discussion

We have reached the end of the analysis. In the result section we can see the footprints from above clustering and visualization of the data.

The focus of this project was to identify a good neighborhood where we can find a nice chinese restaurant.
We looked into all the neighborhoods in Toronto, analyzed the chinese population in each neighborhood and the spread/clustering of chinese restaurants in these neighborhoods, to come to a conclusion where which neighborhood would be a better spot for Chinese restaurant.
We then took past four-item recommender lists (lifegoals, geographic coordinates of Toronto neighborhoods, and Foursquare API, to set up a recommender system.

5.1 Results

In the boroughs of Toronto we identified, with the help of ratio between number of chinese restaurants in boroughs of Toronto, that only Scarborough, North York, Mississauga, and to a lesser extent Etobicoke Toronto, Etobicoke Heavy (high amount of chinese restaurants/high density).

While exploring clusters and ratio we found that Scarborough, North York, Mississauga are densely populated with chinese restaurants and chinese population. Etobicoke Toronto Small and Etobicoke Toronto Large are similar to other three cluster and to a lesser extent to a borough in the image where the restaurant location are spacious.

After careful consideration the two best places to find a good chinese restaurant which also is in the same neighborhood in the borough of Scarborough, North York, Mississauga, and to a lesser extent Etobicoke Toronto, Etobicoke Heavy (high amount of chinese restaurants/high density) is a higher population group among the chinese restaurants, and something to strive for as a customer.

According to this analysis, **Scarborough borough will be the ultimate best neighborhood to live in**. Also looking at the population distribution from the 2 or similar neighborhood with chinese people living in the city/borough by providing high confidence and possibility.

The reason could definitely be a perfect place for a quality chinese restaurant that money wise/cost is lower.

5.2 Discussion and drawbacks

This project only potentially depend on the type of clustering techniques that we use to examine the data.
Although we have used the most popular clustering technique, the k-means algorithm, it may not be the best or the best value. However, many these clusters should be too much. But we think this is a geographical clustering problem based on neighborhoods in a geographical area.

Another drawbacks of the analysis are to find the clustering is completely based on only data obtained from Foursquare API, thus the chinese population distribution in each neighborhood is also based on the 2010 census which is not up to date. Thus population distribution could have definitely changed by 2021.

Since population distribution of chinese cuisine in each neighborhood and number of chinese restaurants are the major features in the clustering algorithm, the results of the analysis could be biased. Other factors such as the food type, prices, and quality are also of the important. However, it certainly provides us with some good insights, preliminary information or possibilities and a head start into the problem of finding the best neighborhoods for us.

6. Conclusion

In this project, we have found out, if one likes chinese food and need to decide where to reside in Toronto, **Scarborough is the optimal place to live in**. We are ready to let the people know that, if they want to open up a small grocery or not the population density and purchasing data tells us.
We have found out a way to be another young man with little chinese cuisine and also the relatively. We have used python and various libraries to find the data, and manipulate the contents and to analyze and visualize those datasets.

We have made use of Foursquare API to explore the venues in neighborhoods of Toronto, then we just gained amount of data from Foursquare API, and we could explore the venues in each neighborhood. We have used the clustering technique, k-means algorithm, and folium libraries to visualize the data against machine learning technique to provide the output given the data we used found in a reasonable way.

Some of the drawbacks or areas of improvement in this project consist on that the analysis can further be improved with help of more-more data and different machine learning techniques. Similarly we can also explore the analysis after problem statement such as map and visualization of the data, and also the data itself, to find more insights and better results.

Above all, Toronto seems to be a nice place to live. It is an international centre of business, finance, arts, and culture, and is recognized as one of the more multicultural and cosmopolitan cities in the world.
Why this analysis shows is that the chances to find a nice chinese restaurant is very good in several boroughs of Toronto.

Zou Ka Lai Adrian as **Data SCIENTIST**. *In Fulltime/ Unemployed. Ready to start ASAP.*