Optimal Restaurant in Toronto

Data Acquisition

Toronto's census Data

- ***https://www.toronto.ca/city-government/data-research-maps/open-data/open-data-catalogue/#8c732154-5012-9afe-d0cd-ba3ffc813d5a***

- ***https://www.toronto.ca/city-government/data-research-maps/open-data/open-data-catalogue/#a45bd45a-ede8-730e-1abc-93105b2c439f***

Wikipedia

- https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

Methodology

- Toronto Census information combined with income, population and competitor
- Wikipedia for poster code
- K- cluster means

Analysis

- Data identification:

I using Wikipedia to search and identify the relevant data source about Toronto, Canada. I removed all the redundant data to make it clean and ready to use

- Longitude and Latitude

Leverage folium map to visualize and combine all dataset with Latitude and Longitude

- Explore the Toronto's neighborhoods

Using Latitude and Longitude data as well as using Foursquare API to see competitors present in Toronto. Filter the top venues

- Clustering

Use K-cluster algorithm to come up with 5 clusters

Results

After a careful investigation and cleaning up data, I had found the results as indicated below

```
print('We have {} boroughs and {} neighborhoods.'.format(
        len(Nebr_ungrp['Borough'].unique()),
        Nebr_ungrp.shape[0]
    )
)

Nebr_ungrp.dropna(inplace =True)
Nebr_ungrp.index = pd.RangeIndex(len(Nebr_ungrp.index))

address = 'Toronto, Ontario, Canada'

geolocator = Nominatim(user_agent="ES1234")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinate of Toronto are {}, {}.'.format(latitude, longitude))
```

```
We have 4 boroughs and 74 neighborhoods.
The geograpical coordinate of Toronto are 43.653963, -79.387207.
```

Conclusion

This will helpful for client who is planning to open a new restaurant in Toronto by comparing different neighbourhoods in Toronto. Foursquare data identified boroughs that justify further analysis. Clustering locations was performed to create major zones of interest. However, it may not cover all variables (attractiveness) so it can’t be used as a single decision-making tool.