Capstone Project - The Battle of Neighborhoods

Opening a new Artisanal Coffee Shop in Toronto

Data Description

Outline:

To achieve the objective set out by the client to identify an ideal neighborhood (that meet their criteria) for their coffee shop:

i) A reliable data source needs to be identified for Toronto neighborhoods
ii) Leverage Foursquare for businesses and competition data (coffee shops and cafes)
iii) Use techniques like k-means clustering to identify similar neighborhoods
iv) Use demographics data to make a recommendation on an ideal neighborhood from a candidate cluster

Data Sources:

The following data sources will be used to produce all the insights that will lead to a reliable recommendation.

1. Toronto Open Data Portal (https://open.toronto.ca/)
    a. Neighborhoods (https://open.toronto.ca/dataset/neighbourhoods/)
        i. Data extract: CSV directly into Python Notebook
        ii. Data:
            1. Neighborhood name
            2. Neighborhood latitude
            3. Neighborhood longitude

	Neighborhood	Longitude	Latitude
0	Wychwood	-79.425515	43.676919
1	Yonge-Eglinton	-79.403590	43.704689
2	Yonge-St.Clair	-79.397871	43.687859
3	York University Heights	-79.488883	43.765736
4	Yorkdale-Glen Park	-79.457108	43.714672

b. Neighborhood Profiles (https://open.toronto.ca/dataset/neighbourhood-profiles/)
   i. Data Extract: neighbourhood-profiles-2016-csv directly into Python Notebook
   ii. Data:
      1. Neighborhood name
      2. Average income of resident by neighborhood (census data)
      3. Total count of working age population by neighborhood (census data)

Neighborhood	Income	Work Age Pop
Agincourt North	30,414	11,305
Agincourt South-Malvern West	31,825	9,965
Alderwood	47,709	5,220
Annex	112,766	15,040
Banbury-Don Mills	67,757	10,810

c. Foursquare API
   i. Data: Coffee Shops, Cafes, Offices/Businesses and Colleges within each neighborhood in 1 km radius
      1. Neighborhood name
      2. Neighborhood latitude
      3. Neighborhood longitude
      4. Venue
      5. Venue latitude
      6. Venue longitude
      7. Venue category

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Wychwood	43.676919	-79.425515	CocoaLatte	43.681768	-79.425158	Café
1	Wychwood	43.676919	-79.425515	Baker and Scone	43.681614	-79.426075	Café
2	Wychwood	43.676919	-79.425515	Contra Cafe	43.669107	-79.426105	Café
3	Wychwood	43.676919	-79.425515	Krave Coffee	43.680740	-79.429417	Coffee Shop
4	Wychwood	43.676919	-79.425515	Starbucks	43.671530	-79.421400	Coffee Shop

Data utilization:

The neighborhood csv from Toronto OD Portal provides neighborhood names and coordinates, which will be used in Foursquare to identify relevant venues (demand drivers like offices/businesses and competition like other coffee shops) in each of the 140 neighborhoods in Toronto. This data will then be used to identify top venues in each neighborhood, cleaned, prepped (one hot encoding, means) and fed into a K-Means algorithm to identify neighborhood clusters. The neighborhood clusters on a map, overlaid with a coffee shop heatmap help identify the most appropriate cluster.

Once a cluster is identified, the most relevant neighborhood will be identified using a weighted value for average income (high) and working age population (high). The current density of coffee shops by population in the neighborhood is also a consideration.

Conclusion:

The data and the method listed above will help objectively assess neighborhoods to identify one that will be an ideal fit for the client's coffee shop meeting their criteria of customers (office workers), a robust population with decent income and reasonably low competition.