Capstone Project - The Battle of Neighborhoods (Week 2)

REPORT

Health food related initiatives and health food retail growth plan in the city of Toronto

Introduction/Business Problem

The collaboration between a non-profit organisation that is working to improve the acesability to healthy food and a healthy food retailer provided an opportunity for the study of food access and food systems within Toronto. Specifically to identify food desert areas where a food desert is a residential neighborhood with little or no access to healthy food, to inventorie and characterize the network of engaged actors, and make recommendations for effective and comprehensive strategies in communication, coordination, and education for food related initiatives. Based on the given by the healthy food retailer, based on previous success in other markets, the objective is to locate, which neighborhood(s) of Toronto will be the best choice to support their growth plan. The information gained will assist in choosing the right location by providing data about the population of each neighborhood, in addition to other established venues present in these areas. Additionally, this information could be of interest to other stakeholders looking to identify food deserts in relationship to neighborhood demographics in Toronto.

Data

The necessary information needed for the study will come from the following sources: **City of Toronto Neighborhood Profiles** for providing an overview of the neighborhoods in Toronto (https://www.toronto.ca/city-government/data-research-maps/neighbourhoods-communities/neighbourhood-profiles/)

City of Toronto Open Data Catalogue : The Census of Population is held across Canada every five years (the last being in 2016), and collects data about age and sex, families and households, language, immigration and internal migration, ethnocultural diversity, Aboriginal peoples, housing, education, income, and labor. City of Toronto Neighborhood Profiles use this Census data to provide a portrait of the demographic, social and economic characteristics of the people and households in each City of Toronto neighborhood. The profiles present selected highlights from the data, but these accompanying data files provide the full data set assembled for each neighborhood.(https://www.toronto.ca/city-government/data-research-maps/open-data/open-data-catalogue/#8c732154-5012-9afe-d0cd-ba3ffc813d5a)

City of Toronto Neighborhood Shapes for mapping : GeoJSON File (https://open.toronto.ca/dataset/neighbourhoods/)

Wikipedia for Toronto Neighborhood Borough Designation : Each of the 140 social planning neighborhoods of Toronto reside within a defined borough. While the City of Toronto is a singular municipality, the 140 neighborhoods are still grouped into six distinct boroughs. (https://en.wikipedia.org/wiki/List_of_city-designated_neighbourhoods_in_Toronto)

Foursquare API to collect information on other venues/competitors in the neighborhoods of Toronto (https://developer.foursquare.com/)

Methodology

The Methodology section will describe the main components of our analysis and predication system. The Methodology section comprises four stages:

1. Collect Inspection Data
2. Explore and Understand Data
3. Data preparation and preprocessing
4. Modeling

Exploratory Data Analysis

Part 1: Identifying Neighborhoods
We will use Postal Codes of different regions to find the list of neighborhoods. We will essentially obtain our information
from https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M and then process the table inside this site. Images from dataframes and also from maps will be provided in the presentation. Here we only present our strategy and how we got the mission accomplished.

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

Part 2: Connecting to Foursquare and Retrieving Locational Data for Each Venue in Every Neighborhood
After finding the list of neighborhoods, we then connect to the Foursquare API to gather information about venues inside each and every neighborhood. For each neighborhood, we have chosen the radius to be 1000 meter. It means that we have asked Foursquare to find venues that are at most 1000 meter far from the center of the neighborhood. (I think distance is measured by latitude and longitude of venues and neighborhoods, and it is not the walking distance for venues.)

Part 3: Processing the Retrieved Data and Creating a DataFrome for All the Venues
When the data is completely gathered, we will perform processing on that raw data to find our desirable features for each venue. Our main feature is the category of that venue. After this stage, the column "Venue's Category" wil be One-hot encoded and different venues will have different feature-columns. After On-hot encoding we will integrate all restaurant columns to one column "Total Restaurants" and all food joint columns to "Total Joints" column. We assumed that different resaturants use the Same raw groceries. This assumption is made for simplicity and due to not having a very detailed dataset about different venues.

Now, the dataset is fully ready to be used for machine learning (and statistical analysis) purposes.

Applying one of Machine Learning Techniques (K-Means Clustering)

Here we cluster neighborhoods via K-means clustering method. We expect that 5 clusters is enough and can cover the complexity of our problem. After clustering we will update our dataset and create a column representing the group for each neighborhood.

Findings

We focus on the centers of clusters and compare them for their "Total Restaurants" and their "Total Joints". The group which its center has the highest "Total Sum" will be our best recommendation to the contractor. {Note: Total Sum = Total Restaurants + Total Joints + Other Venues.}

	PostalCode	Borough	Neighborhood	Latitude	Longitude	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Cluster Labels										
0.0	7	7	7	7	7	7	7	7	7	7
1.0	1	1	1	1	1	1	1	1	1	1
2.0	7	7	7	7	7	7	7	7	7	7
3.0	1	1	1	1	1	1	1	1	1	1
4.0	1	1	1	1	1	1	1	1	1	1

	Neighborhood	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
0	Berczy Park	Seafood Restaurant	Restaurant	French Restaurant	Eastern European Restaurant	Greek Restaurant	Thai Restaurant	Vegetarian / Vegan Restaurant	Japanese Restaurant
1	Central Bay Street	Italian Restaurant	Thai Restaurant	Japanese Restaurant	Middle Eastern Restaurant	Falafel Restaurant	Indian Restaurant	Vegetarian / Vegan Restaurant	Korean Restaurant
2	Christie	Italian Restaurant	Restaurant	Doner Restaurant	Gluten-free Restaurant	German Restaurant	French Restaurant	Filipino Restaurant	Fast Food Restaurant
3	Church and Wellesley	Sushi Restaurant	Japanese Restaurant	Restaurant	Mediterranean Restaurant	Italian Restaurant	Mexican Restaurant	American Restaurant	Caribbean Restaurant
4	Commerce Court, Victoria Hotel	Restaurant	American Restaurant	Italian Restaurant	Seafood Restaurant	Japanese Restaurant	Thai Restaurant	Vegetarian / Vegan Restaurant	French Restaurant
5	First Canadian Place, Underground	Restaurant	Japanese Restaurant	American Restaurant	Asian Restaurant	Seafood Restaurant	Thai Restaurant	Sushi Restaurant	Italian Restaurant

Conclusions

Based on this analysis, the best recommended neighborhood will be:
{'Neighborhood': 'Agincourt',
'Postal Code': 'M1S',

'Neighborhood Latitude': 43.7942003,

'Neighborhood Longitude': -79.26202940000002}