Data Science Capstone Project – Where to Open a New Coffee Shop in Toronto?

Introduction

An International coffee chain is looking to expand to Canada and wants to open its first location in Toronto. It wants to ensure that its first location is a success, and is looking for an area that fits the following criteria:

- **Large customer base:** Densely populated area with many residents, offices, and schools
- **Underdeveloped coffee scene:** Not many existing coffee shops in the area

Data

This analysis was performed by integrating the following data sets:

- Neighbourhood information for the City of Toronto that was obtained from Wikipedia
- Venue related information for the City of Toronto that will be obtained from the Foursquare API

This data was converted into a final pandas dataframe for analysis, and Folium was used to visualize the results. Finally, k-means clustering was used to cluster similar neighbourhoods together, and uncover the areas that are best suited to opening a new coffee shop

Methodology

After using the Beautiful Soup library to extract the neighbourhood information data from Wikipedia, that data was converted into a pandas data frame.

Afterwards, data from Foursquare related to the different venues in the city of Toronto was downloaded and merged into the same data frame as the neighbourhood information. This created one dataset which showed the different venues (e.g. offices, parks, cafes) found in each Toronto neighbourhood (represented by postal code).

A machine learning model (k-means) was then used to group all Toronto neighbourhoods into similar clusters. The areas deemed most attractive for the coffee chain were those who did not have café's or coffee shops within a 1-kilometer radius.

Results were then plotted on a map to visually represent where these neighbourhoods were located.

Results

Most of the areas best suited to opening a coffee shop were in city's west and east areas.

Discussion

It makes sense that underserved coffee areas are located on the city's borders vs. in the downtown core, where there are already a large number of coffee chains.

After more closely examining the results it appears there are more underserved coffee areas in the west vs. the east end, with several underserved areas in the city's northwest area. As a result, I would recommend that the client start conducting more thorough analysis in the city's northwest end to nail down the perfect launch location. The lack of options in that area gives the client a unique opportunity to quickly establish several locations and become the area's best known coffee brand

Conclusion

This analysis identified a geographic area (Toronto's northwest) that is currently underserved by other coffee chains. The company should conduct further analysis to determine the best location for its first coffee shop in this large area by examining data such as:

- Household income
- Foot traffic
- Food/drink preferences via a survey of area residents