IBM Capstone Project: exploring restaurants in an industrial district

By Nikita

Contents

Introduction

Etobicoke is an industrial district in Canada. Due to the fact that it is industrialised, the population density is lower than the surrounding districts. However, restaurants and shopping malls are located in this district.

Business problem

Given that Etobicoke is such an industrialised area, it is likely that the district is not saturated with restaurants. It is of interest to visualise whether or not the restaurants cluster together in such an industrialised district, as it has stretches of expansive industrial land. If restaurants are clustered, then as more restaurants open, the market share of the existing restaurants in the surrounding area decreases, which has the potential to impact profitability of these existing venues. If the restaurants are not clustered, there is less risk of losing market share.

Thus, the target audience of this study is existing restaurant owners looking to protect their existing business or to further expand their franchise.

Data

1. Neighbourhoods: these will be obtained from Wikipedia using webscraping tools in Python.
2. Geocoding: the geographic coordinates of each neighbourhood will be obtained using Google Geocoding API
3. Restaurants: the geographic coordinates of each restaurant will be obtained using Google Geocoding API and Foursquare API

Once the data is obtained, it will need to be cleaned. Missing values will result in an entry being removed. Furthermore, entries that have repeating information will be merged together.

After the data is cleaned, it will be presented in a dataframe for easy viewing and further manipulation.

Methodology

Etobicoke is located in Toronto, Canada. Thus, Etobicoke was extracted from a list of districts in the city of Toronto. This list can be found on Wikipedia.

To extract the list of neighbourhoods of Etobicoke, web scraping was used. This was done through Python. The data was cleaned and formatted as a dataframe.

The geographical coordinates of each neighbourhood was determined using Foursquare API and the Geocoder package. The data was formatted as a dataframe and visualised.

The nearest Etobicoke venues, in a radius of 500 metres, were extracted using the Foursquare API. API calls were made to Foursquare using a loop, and the venue data was returned in JSON format. Thus, the venue name, category and coordinates were obtained.

Furthermore, clustering was performed to analyse the presence of venues in a particular neighbourhood. The method used was k-means. K-means clustering identifies k number of centroids, and then allocates every data point to the nearest cluster.

The neighborhoods were clustered into 5 clusters. The results would indicate which neighborhoods have had a concentration of certain types of restaurants.

The results would help existing owners to better understand whether they should protect of expand their business.

Results
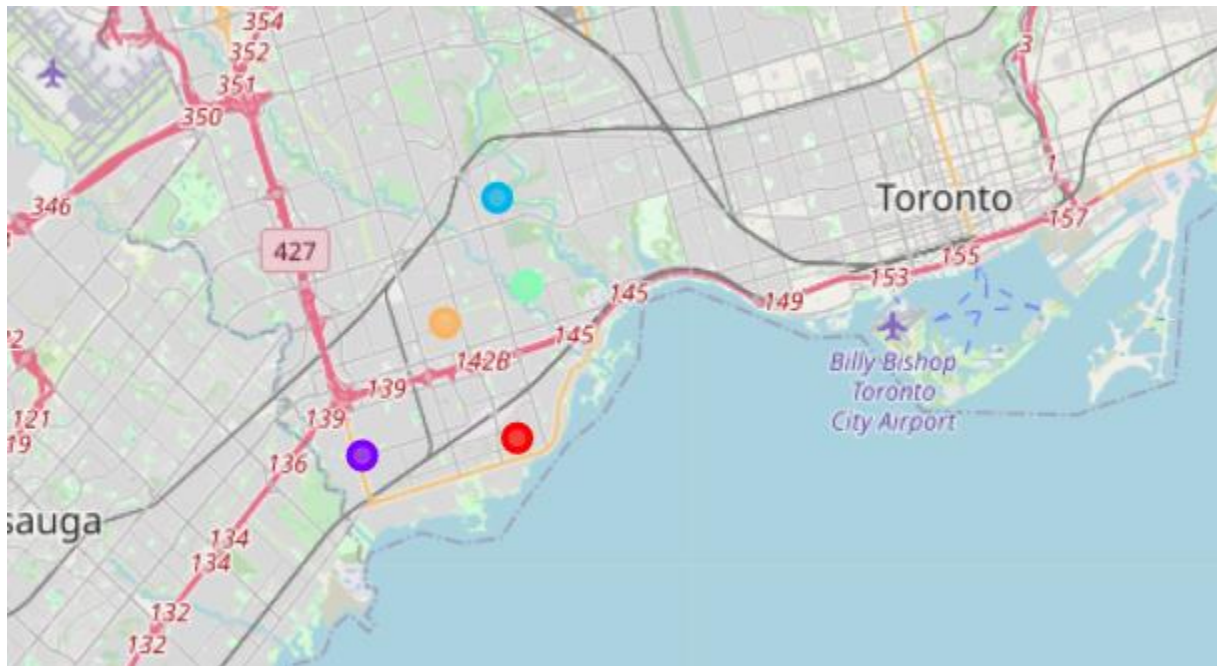
The results indicate that there are 5 clusters in Etobicoke where restaurants appear most frequently.

Based on the results, the most commonly occurring types of restaurants in these clusters include:

- Bakeries
- Wing joints
- Sandwich shops
- American restaurants

However, one cluster (green) was more populated by shopping venues rather than restaurants.

A conclusion based on these results is discussed in the next section.

Conclusion

The results of the study can help existing restaurant owners to make a more informed decision regarding whether to protect of expand their existing restaurant business, based on saturation of similar businesses in their current neighbourhood. A high saturation of a particular restaurant type implies that there it may be more difficult to gain market share, thus business owners should protect their established market share. Should a particular type of restaurant appear less frequently in an area, there is potential to expand into that area as it may be easier to gain market share.

Thus, based on the results observed, bakeries, wing joints, sandwich shops and American restaurants should aim to protect their existing market share within their neighbourhoods. A potential area for expansion, where almost no other restaurants exist, is in the area defined as cluster 3, and the most market share can be gained here if expanding or starting a new restaurant.