

Capstone Project Final Report

Introduction

In this Project, I will try to find best location for a Thai Restaurant in Toronto.

Business Understanding

Problem: I want to open a new Thai Restaurant in Toronto, Canada.

Question: If Toronto needs a Thai Restaurant, which area is the most needy?

Method: Business Acumen and Data Science Methodology

Algorithm: K-means clustering

Data

- Neighbourhoods and their geographical data (Latitude and Longitude) in Toronto
- Venue data of Thai Restaurants

Source

- Neighbourhoods on Wikipedia and Geocodes (Geocoder Package)
- Foursquare API

Methodology

- 1. Get the list of neighbourhoods from Wikipedia.
- 2. Scraping the data with Pandas HTML.
- 3. Match the coordinates of Toronto Neighbourhoods with Geocoder Package.
- 4. Visualize the map of Toronto using Folium Package to verify correct coordinates.
- 5. Create Foursquare developer account.
- 6. Use Foursquare API to pull the list of top 100 venues within 500 meters Radius.
- 7. Analyze each neighborhood by grouping the rows by neighborhood and take the mean on the frequency of occurrence of each venue category.
- 8. Look for Thai Restaurant.
- 9. Perform the clustering method by using KMeans.

- 10. Cluster the neighborhoods in Toronto into three clusters based on their frequency of occurrence for Italian Food .
 - 11. Recommend the ideal location to open the restaurant.
-
- Cluster 0 : Neighborhoods with no Thai restaurants.
 - Cluster 1 : Neighborhoods with the more number of Thai restaurants.
 - Cluster 2 : Neighborhoods with the less number of Thai restaurants.

Recommendations

- Most of the Indian restaurants are in cluster 1.
- Lowest in Cluster 0 which are in North Toronto. Also, there are good opportunities to open.
- Looking at nearby venues it seems cluster 2 might be a good location as there are not a lot of Thai restaurants in these areas.
- Therefore, this project recommends the entrepreneur to open an Thai restaurant in these locations.