Coursera Capstone Project

IBM Data Science Professional Certificate - Capstone Project

Clustering Neighborhoods of Hamilton New Zealand to find the potential location to open Café bar

July 2020

Table of Content

Introduction

Caffeehouses (Café) are known as centers of social interactions. Café bars provide an open atmosphere for individuals or small group to talk, congregate, read, write, entertain others or spending time. Since Wi-Fi is invented, Café bars have been transformed to cozy places for individuals to access to the internet using their personal laptops or smartphones. Some Café bars acts as a kind of club for their members where local musical bands perform and entertain visitors on specific occasions. Beside serving different type of coffees, Café bars have become like as diner that serve food, sandwiches, soups, salad, snacks and pastries. Some Café bars operates as personal business and some others are like as franchise business model. Other businesses can take benefits of having café bars nearby as they usually operate as a place for passengers and visitors to interact and get togethers. To add to all above, Café bar whereabout is an important factor that needs to be considered.

Business Problem Definition

The objective of this project is to help project managers to find out the best locations in Hamilton New Zealand suburbs to open a new Café bars by using data science methodology and machine learning techniques such as clustering. In this project tries to answer this question as: if project manager wishes to open a new café bar, where would be the most appropriate location.

Target Audience

This project is useful for project management and investors in construction sections who are looking for potential place to open a café bar in Hamilton New Zealand. Also this project is a real case study for individuals who are interested of knowing data science and machine leaning applications in our daily life. This project is a good example for students to deep dive into Python libraries, functions and scripts.

Methodology

To find out the potential location to build a new Café bar in Hamilton, New Zealand, the following data are required:

1) List of all neighborhoods in Hamilton, New Zealand suburbs.

2) Geographical coordinates (Latitude and longitude) of neighborhood obtained at first step. This information is used to plot map and get the venues.

3) Finally, venue data is grouped by neighborhoods and related venues to Café bar are selected. These data is used to cluster neighborhoods and find the best location for the new café bar.

Data Extraction

1) The raw data is extracted from the following Wikipedia page that includes the list of Hamilton, New Zealand neighborhood of total 50.

https://en.wikipedia.org/wiki/Category:Suburbs_of_Hamilton,_New_Zealand

2) Then web scraping techniques are used to extract data from Wikipedia page using Python request library and beautiful soup package.

3) In next step, Python geocoder package is used to get latitude and longitude of all 49 neighborhoods.

4) To get venues data for all those neighborhood Foursquare API is used. Foursquare API provide many venue categories. As in this project our focus is on Café bars, data cleaning and data wrangling are used to take the relevant venues and grouped the neighborhood.

5) Machine leaning (K-mean) clustering and map visualization (Folium) are used to find the potential location to open Café bar.

6) The other details are explained in the project.

Conclusion

Based on the summery page above, most of the Cafe bars are in Cluster in Melville and Thornton areas with highest number.

Compare to Custer 1 with moderate number, cluster 0 have very low number to no Cafe bar like as Temple View, Rotokauri and so on.

This means there are high competition in Melville and Thornton (Cluster 2) are another neighborhood in cluster 0 are starving of having Cafe bar. Also, the data illustrates that most of cafe bars are concentrated in central areas and not suburb area.

Therefore, the project recommends construction developers rely on these findings and invest of suburb areas more. Although those investors who are looking into more competition can also open cafe bars in neighborhood in cluster 1.