

IBM Data Science Professional Certificate

Coursera Capstone – IBM Applied Data Science Capstone

Best Suburb to Open a Cafeteria in Melbourne, Australia

Author: Piyal Banik
Date: 8th February, 2021



Introduction

In a busy city like Melbourne, cafeterias are becoming more and more popular due to its concept where the customer choose a food from a counter, pays for it and take it either to a table or take it away with them. Many business owners have already taken advantage of its popularity and have set up their cafeterias in a location where the demand is greater. But for people who are interested in opening up a new Cafe are highly uncertain about their various issues that they might face. One of them being choice of location. i.e. whether the location is good enough for running the business longer. They want to choose a location where there will be less competition i.e. have fewer Cafeterias present in the suburb, in order to minimise the risk of facing a loss after opening the Cafeteria.

Business Problem

In this project we aim to solve the above mentioned problem where a person willing to invest in a new cafe faces. The main goal of this capstone project is to collect and analyse data in order to select a location in Melbourne, Australia to open a Cafeteria. We want to help a business owner planning to open up a Cafeteria in a location by exploring better facilities around the Suburb in Melbourne.

Target Audience

This project is aimed at helping people willing to open up a Cafe in the city of

Melbourne, Australia. Since, with each day passing by there are greater chances of new cafeterias getting introduced in the city, the research performed is time constrained i.e. this research will be useful for readers in the year 2021-2022 only.

Data

We would need few different types of data in order to solve the problem in hand:

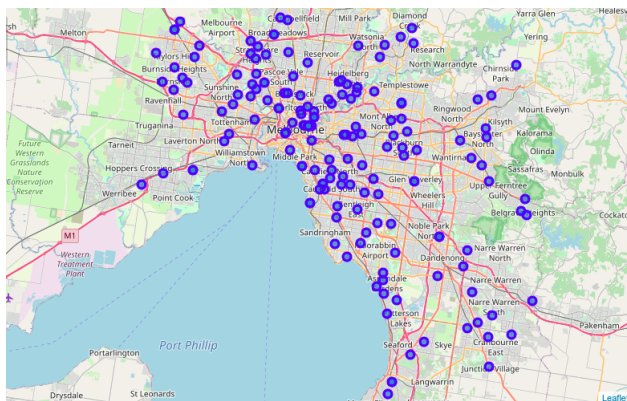
- List of Suburbs in Melbourne, Australia which we will extract from:
https://en.wikipedia.org/wiki/Category:Suburbs_of_Melbourne using web scrapping methods i.e use BeautifulSoup package and Request of python
- We will use Foursquare and Geocoder package of python to extract data about various venues present in different Suburbs of Melbourne along with Latitude, Longitude Data of those suburbs and venues

Methodology

The Wikipedia page contains list of suburbs in Melbourne. There are 212 suburbs in Melbourne which we extracted using web scraping technique with the help of Python BeautifulSoup and Request packages. After that, the geographical coordinates such as latitude and longitude of each suburb using Python's Geocoder package. Then, Foursquare API was used to extract details about the various venues present

in each suburb. But we are only interested in Cafe category which was obtained appropriately.

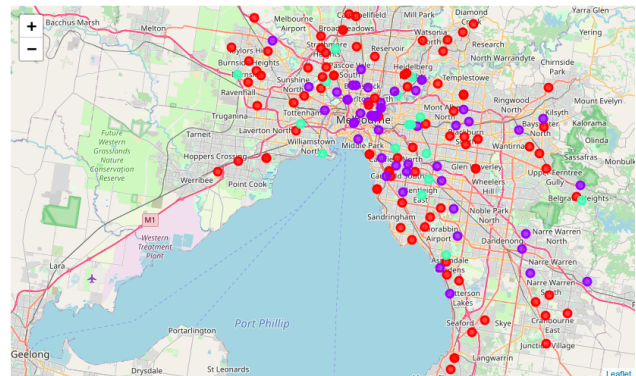
After the data has been extracted from the Wikipedia page, it was saved in a CSV file and stored in a pandas dataframe for further analysis. Once, the location data was extracted by using Geocoder, we used Folium package to visualise the data in a map. This also ensured us that the data we retrieved was correct. Foursquare API was used to obtain top 100 venues within a radius of 2000 meters.



We then converted the data into dummy variables which will be essential for performing clustering algorithm using `get_dummies` method of Pandas package. We then grouped the data as per Suburb and by also taking the mean of the frequency of occurrence of each category. After that we extracted the data of Cafeteria only.

Lastly, we performed clustering on the data using K-means clustering. We found out 3 clusters from the data based on the frequency of occurrence of Cafes in each suburb. This allowed us to find out the suburb which had the highest concentration of Cafes and also the

lowest concentration in various suburbs which we visualised in a map again.



Results

We categorised the data into 3 categories using K-means clustering based on frequency of occurrence for 'Cafe'.

- Cluster 0: Suburbs with low number of Cafeterias.
- Cluster 1: Suburbs with moderate number of Cafeterias.
- Cluster 2: Suburbs with high concentration of Cafeteria.

Discussion

As we can observe from the map of Melbourne after performing the clustering algorithm that most of the Cafeterias are concentrated in or around the Melbourne CBD which is displayed in Cyan color and is the cluster 2. Cluster 0 is displayed as red color and have the lowest number of cafes in those suburbs. Although this represents a greater opportunity and high potential but also suffer from the risk of having fewer customers as those areas are not as busy as the ones in Cluster 2 and 1. As a new business owner it wouldn't be wise

enough to choose cluster 2 . Therefore, we would recommend that cluster 1 should be chosen where there are medium competition but greater opportunity.

Conclusion

This project included quite a lot of steps which we needed to perform in order to achieve the result we intended at the start of the project. We first decided on the problem we will try to solve, then collected data from various sources, cleaned the data, visualised the data and then performed machine learning algorithm to come up with the solution that were looking for. This findings will help anyone willing to open up a new Cafeteria in Melbourne and can use the insights to choose a location accordingly. As per our suggestion new business owner should choose Cluster 1 as the perfect one for their dream Cafeteria.

References

https://en.wikipedia.org/wiki/Category:Suburbs_of_Melbourne

<https://developer.foursquare.com/docs/>