

# **IBM Data Science Professional Certificate - Capstone Project**

**Perth, Western Australia, where to open a new Bar?  
Battle of Neighborhoods**

July 2020

## Table of Content

Introduction	3
Business Problem Definition	3
Target Audience	4
Methodology	4
Data Extraction	5
Conclusion	6

## Introduction

Bars are known as centers of social interactions. Individuals or small group of people stop by Bars to use their open and comfy atmosphere to talk, congregate, watch sports, celebrating special occasions or spending time. Since a decade ago, Bars provides internet services to their customers that make their business more crowded. Some bars have become as a kind of club for their members where local musical bands perform and entertain visitors on specific occasions. Beside serving different type of alcohol, bars have become like as diner that serve food, sandwiches, soups, salad, snacks. Other businesses can take benefits of having bars nearby as they usually operate as a place for gathering and others to interact. Hence, the location of Bar is important for both the owner and other retailers.

## Business Problem Definition

The objective of this project is to help decision makers to find the potential place to open a new Bar in Perth, Australia by using data science methodology and machine learning techniques such as clustering. In this project our goal is to answer this question as: if a decision maker has a plan to open a new bar, where would be the potential place.

## Target Audience

This project is useful for both investors and individuals who are interested to open a bar in Perth, Australia. Moreover, this project is a comprehensive package for those who are interested of knowing data science and machine learning applications in our daily life. This project is a good example for students to deep dive into Python libraries, functions and scripts.

## Methodology

The following steps are taken to find out the best location to construct a new Bar in Perth, Australia:

- 1) All neighborhoods in Perth, Australia suburbs are listed.
- 2) Latitude and longitude (Geographical coordinates) of all neighborhoods obtained at first step are extracted. This information is used to plot map and get the venues.
- 3) Finally, venue data is grouped by neighborhoods and related venues to Bar are selected.

These data is used to cluster neighborhoods and find the best location for the new Bar.

## Data Extraction

- 1) The raw data is extracted from the following Wikipedia page that includes the list of Perth, West Australia neighborhood of total 72.
- 2) [https://en.wikipedia.org/wiki/Category:Suburbs\\_of\\_Perth,\\_Western\\_Australia](https://en.wikipedia.org/wiki/Category:Suburbs_of_Perth,_Western_Australia)
- 3) Then web scraping techniques are used to extract data from Wikipedia page using Python request library and beautiful soup package.
- 4) In next step, Python geocoder package is used to get latitude and longitude of all 72 neighborhoods.
- 5) 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 Bars, data cleaning and data wrangling are used to take the relevant venues and grouped the neighborhood.
- 6) Machine learning (K-mean) clustering and map visualization (Folium) are used to find the potential location to open Café bar.
- 7) The other details are explained in the project.

## Conclusion

Based on the summary page above, most of the Bars are in Clusters one and two. The frequency of Bar locations are more in cluster two but cluster 1 has higher number particularly in East Perth and West Perth with above 0.5.

Compare to Cluster two with moderate number, cluster 0 have very low number to no Bars like as Bayswater, Balcatta and so on.

This means there are high competition in East Perth and West Perth (Cluster 1). Neighborhood in cluster 0 are starving of having Bars. Also, the data illustrates that most of bars are concentrated in east and west 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 bars in neighborhood in cluster 2.