Scaling offline business or where to place a new shop

IBM Applied Data Science Capstone Project report

by: Ivan Ilukhin

2020

Introduction

Many businesses very often face the next problem: when it begins to grow up, owners start to think about: where to open new outlets or found another place where you satisfy your customers' needs. This issue usually relates to offline companies like barbershops, veterinary clinics, retail stores, etc. There is a lot of parameters that impact on the decision: monthly rent, competitors, crime rates, a wealth of the people around and many others. This little research is focusing on the similarity of the boroughs or **How to split all districts into groups.**

The idea is based on the underlying assumption that we should found spots in similar places.

Business Problem

The objective of this research is to find similar districts and extract attributes that distinguish them. I selected the Greater Sydney Area for the region of study. It's one of the biggest agglomerations in the world, consists of almost about 650 suburbs that sprawl about 70 km to the west, 40 km to the north, and 60 to the south and generates approximately 24.1% of Australia's GDP.

Target audience of this project

The result of this project will be interesting, primarily, for the business owners and people who choose the location for expanding. It may also be of interest to investors, especially when you are presenting your startup or when you ask for money to grow your business.

Data

Feature selection

For classifying suburbs, we should get features for each district. In this research, I will use the next:

- latitude and longitude to show them on a map;

- venue data, amount of each type of venues;

- crime data, about of crimes grouped by type(murder, robbery, carjacking;

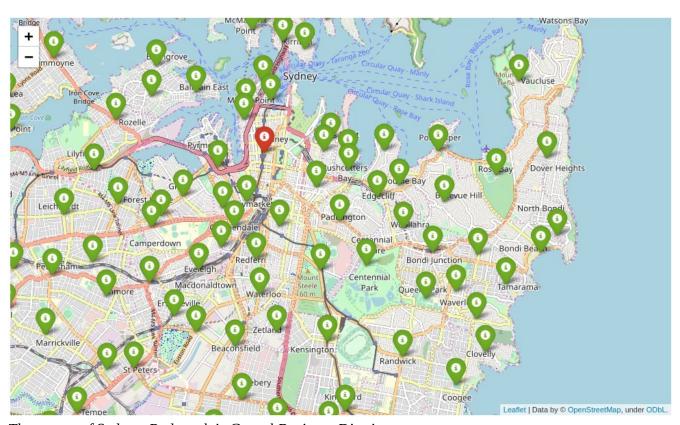- distance from the Sydney Central Business District (CBD).

Data sources

First we should get the list of suburbs by scrapping the wiki page - https://en.wikipedia.org/wiki/List_of_Sydney_suburbs

	suburb
0	Abbotsbury
1	Abbotsford
2	Acacia Gardens
3	Agnes Banks
4	Airds

```
suburbs_df.shape
```

```
(689, 1)
```

Then get the locations using the Google Maps Geocoder API

suburb	lat	lng
Abbotsbury	-33.87500	150.86200
Abbotsford	-33.85215	151.12726
Acacia Gardens	-33.73220	150.91700
Agnes Banks	-33.61529	150.71616
Airds	-34.09160	150.82490

Suburbs with latitudes and longitudes

The center of Sydney. Red mark is Cenral Business District.

Crude distance from the central district will be calculated based on the centres of suburbs, got on the second step.

suburb	lat	lng	dist_from_cbd
East Kurrajong	-33.515278	150.767778	56.711739
Yarramundi	-33.622200	150.670900	56.843291
Grose Wold	-33.598611	150.684722	57.072646
Tennyson	-33.536944	150.737500	57.172222
Wisemans Ferry	-33.381944	150.985000	58.018323
Grose Vale	-33.584000	150.674000	58.782859
The Slopes	-33.532801	150.706912	59.649490
Kurmond	-33.551000	150.690000	59.650771
Kurrajong	-33.550000	150.666700	61.464918
Kurrajong Hills	-33.533333	150.650000	63.804356

Farest from the CBD districts after removing outliners

Monthly crime reports from the Bureau of Crime Statistics and Research. There are reports for each suburb in New South Wales, so we should select only ones that we scrapped on the first step - https://www.bocsar.nsw.gov.au/Pages/bocsar_datasets/Datasets-.aspx

suburb	sum_crimes	lat	lng	dist_from_cbd
Abbotsbury	3715	-33.87500	150.86200	31.953973
Abbotsford	6009	-33.85215	151.12726	7.693985
Acacia Gardens	1716	-33.73220	150.91700	30.971957
Agnes Banks	947	-33.61529	150.71616	53.617942
Airds	22071	-34.09160	150.82490	42.992898

Suburbs with coodinates, distanse and summary amount of crimes

And the list of venues using Foursquare API - https://foursquare.com/

suburb	ATM	Accessories Store	Advertising Agency	Afghan Restaurant	Airport	Airport Terminal	American Restaurant	Aquarium	Arcade	Arepa Restaurant	...	Wine Bar	Wine Shop	V
Abbotsbury	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	
Abbotsford	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	1.0	
Acacia Gardens	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	
Airds	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	
Alexandria	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
Yagoona	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	
Yarrawarrah	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	
Yennora	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	
Yowie Bay	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	
Zetland	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	