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

Then get the locations using OpenCage API - https://opencagedata.com/api

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

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

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