Coursera Capstone

IBM Applied Data Science Capstone

Opening a New Coffee Shop in Cincinnati

By: Vineeth Kumar Kondamadugu

Introduction

once heard someone say *"A bad day with coffee is better than a good day without it".* For many People, visiting Coffee Shop is a great way to relax and enjoy themselves during weekends and holidays. They can make it as their hobby, dine at, spend some time out with friends, find peace and even work at coffee shops. Coffee shops are like a one-stop destination for all coffee lovers. Property developers are also taking advantage of this trend to build more coffee shops to cater to the demand. As a result, there are many best coffee shops in the city of Cincinnati and many more are being built. Opening coffee shops allows lease owners as well as owners to earn consistent rental income. Of course, as with any business decision, opening a new coffee shop requires serious consideration and is a lot more complicated than it seems. Particularly, the location of the coffee shop is one of the most important decisions that will determine whether the shop will be a success or a failure.

Business Problem

The objective of this capstone project is to analyze and select the best locations in the city of Cincinnati, Ohio to open a new Coffee Shop. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: In the city of Cincinnati, Ohio, if a property developer or retailer or owner himself is looking to open a new coffee shop, where would you recommend that they open it?

Target Audience of this project

This project is particularly useful to property developers, investors, retailers and individuals looking to open or invest in new shopping malls in the city of Cincinnati, Ohio, USA. Current trend says that coffee shops are more successful with the growing busy workstyle of people around globe not only in Cincinnati. Simply put, Cincinnati has an impressive variety of coffee shops. Whether you're interested in finding a place to get some work done, looking for a place to meet a friend, seeking out a place to enjoy a caffeinated beverage, or searching for a spot to host a meetup, Cincinnati has coffee shops sprinkled throughout the city where you can do all of these things.

Data

To solve the problem, we will need the following data:

- List of neighborhoods in Cincinnati. This defines the scope of this project which is confined to the city of Cincinnati, Ohio, United States of America.
- Latitude and longitude coordinates of those neighborhoods. This is required in order to plot the map and also to get the venue data.
- Venue data, particularly data related to Coffee Shops. We will use this data to perform clustering on the neighborhoods.

Sources of data and methods to extract them

This Wikipedia page (https://en.wikipedia.org/wiki/Category:Neighborhoods_in_Cincinnati) contains a list of neighborhoods in Cincinnati, Ohio, with a total of 52 neighborhoods. We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and beautifulsoup packages. Then we will get the geographical coordinates of the neighborhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the neighborhoods.

After that, we will use Foursquare API to get the venue data for those neighborhoods. Foursquare has one of the largest database of 105+ million places and is used by over 125,000 developers. Foursquare API will provide many categories of the venue data, we are particularly interested in the "Coffee Shop" category in order to help us to solve the business problem put forward. This is a project that will make use of many data science skills, from web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium). In the next section, we will present the Methodology section where we will discuss the steps taken in this project, the data analysis that we did and the machine learning technique that was used.