Coursera Capstone

IBM Applied Data Science Capstone

Opening a New Restaurant in Portland, Oregon

By : **Tuan Nguyen**

July 2020

Introduction:

A great city always has many restaurants, from high-end and luxury, to close and friendly, anyone can pick up a restaurant and enjoy the dinner weekday or weekend. For shopkeepers, the central location and the large crowd near workplaces are great places to open their business. So you can easily find a restaurant in downtown Portland. Opening a restaurant can earn a lot of money, but also the rent is high. For that, the location of the restaurant is one of the most important decisions that will determine whether the mall will be success or failure.

Business Problem

The objective of this capstone is to analyse and select the best locations in Downtown Portland to open a new restaurant. Using data science methodology and machine learning techniques like clustering.This project aims to provide a solution the question of in Downtown Portland, where to open a new restaurant?.

Data:

- List of neighborhoods in Portland. Downtown Portland specializes in South West and North West Portland.

- 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 restaurants. We will use this data to perform clustering on the neighborhoods.

Source of data:

The Wikipedia page (https://en.wikipedia.org/wiki/Neighborhoods_of_Portland,_Oregon) contains a list of neighborhoods in Portland, with a total of 29 neighborhoods in Downtown. We will use web scraping techniques to extract the data from the Wikipedia page using Beautifulsoup packages. Then we will get the geographical coordinates of the neighborhoods using the 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. The API will provide many categories of the venue data, we are particularly interested in the Restaurant category in order to help us to solve the business problem.

Methodology

The list of neighborhoods in Portland is in Wikipedia page

(https://en.wikipedia.org/wiki/Neighborhoods_of_Portland,_Oregon) . We will do web scraping

with Python requests and BeautifulSoup packages to get the data. Then We need to use the

Foursquare API and geocoder to get all the information we need like latitude and longitude and

other places like restaurants, schools, workplace and so on.

We will look at top 100 venues that are within a radius 1000 meters. Using the API will return a

json file that contains the venue data, then venue name, category, latitude and longitude will be

extracted. Since we are analyzing "Restaurant" data, we will filter the "Restaurant" as a venue

category for the neighborhood. Because the API returns different types of restaurants: American

Restaurant, Chinese/Japanese/Mexican Restaurant, those categories will be grouped into 1

restaurant category.

Then we will perform clustering on the data by using k-means clustering

Result.

The results from the k-means clustering show that we can categorize the neighborhoods into 3

clusters based on the frequency of occurrence for "Restaurant":
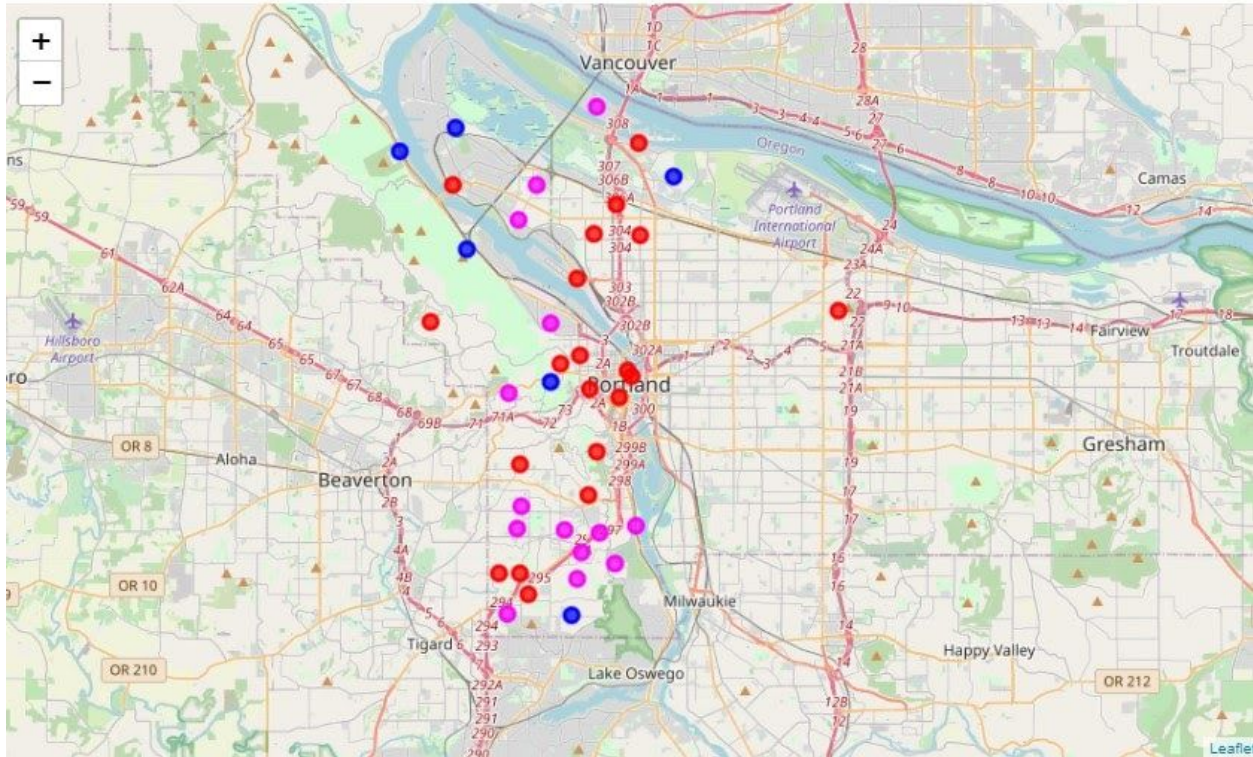
- Cluster 0(Pink): Neighborhoods with moderate number of restaurants

- Cluster 1(Red): Neighborhoods with high number of restaurants

- Cluster 2(Blue): Neighborhoods with low number of restaurants

Discussion

As observations noted from the note, most of the restaurants are concentrated in cluster 0 and

0. On the other hand cluster 0 has lower frequency of restaurants than cluster 1. While cluster 2

shows a very low number of restaurants. This also shows it is more competitive to open a

restaurant in cluster 1.Therefore this project recommends shopkeepers to capitalize on these

findings to open a new restaurant in neighborhoods in cluster 0, then 1, but not 2.

Limitations:

In this project, we only consider 1 factor: frequency of occurrence of restaurants, there are other factors such as residents, schools and bars. However, the data for that are not available to the neighborhood level for this project.

Conclusion:

In this project, we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing machine learning by clustering the data into 3 clusters based on their similarity, and providing recommendations to shopkeepers. To answer the business questions, the answer by this project is: The neighborhoods in cluster 0 is the most preferred location to open a new restaurant because of moderate frequency, unlikely cluster 1 is very competitive.

References:

Neighborhoods of Portland, Oregon. *Wikipedia*. Retrieved from:

https://en.wikipedia.org/wiki/Neighborhoods_of_Portland,_Oregon

Foursquare Developers Documentation. *Foursquare.* Retrieved from:

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

Appendix:

	Neighborhood
7	Collins View
10	Far Southwest
13	Hayden Island
14	Hayhurst (includes Vermont Hills)
20	Maplewood
21	Markham
22	Marshall Park
23	Multnomah (includes Multnomah Village)
26	Northwest Industrial
32	Portsmouth
33	South Burlingame
34	South Portland (includes Corbett, Fulton, Lair...
37	Sylvan-Highlands
38	University Park

Cluster 0

	Neighborhood
0	Arbor Lodge
3	Ashcreek
4	Bridgeton
5	Bridlemile (includes Glencullen)
6	Cathedral Park
8	Crestwood
12	Goose Hollow
15	Hillsdale
16	Hillside
17	Homestead
18	Kenton
24	Northwest District (includes Uptown, Nob Hill,...
25	Northwest Heights
27	Old Town Chinatown
28	Overlook
29	Pearl District
30	Piedmont
31	Portland Downtown
35	Southwest Hills, Portland, Oregon
39	West Portland Park (includes Capitol Hill)

Cluster 1

	Neighborhood
1	Arlington Heights
2	Arnold Creek
9	East Columbia
11	Forest Park
19	Linnton
36	St. Johns

Cluster 2