Coursera Capstone Project

SUCHETA MARATHE

3RD NOV 2019

Project – Predicting a location to open a next restaurant outlet

Aim of this project is to find out the next probable location to open new ‘Chipotle Mexican Grill’ outlet in USA

Background

‘Chipotle Mexican Grill’ is a casual restaurant chain in USA.

Currently it operates less than 2,500 restaurants, most of which are in the U.S. This number is very small as compared to the35000+ McDonald’s restaurants in 119 countries or 42000+ Subways across many countries, which are similar in type of ‘fast casual restaurant’ chains.

There are states in USA with very less or without Chipotle outlet!!

Although Chipotle is closing some of its outlets, it is expanding business in some areas as well. There is a tremendous growth opportunity for Chipotle to expand both domestically and internationally.

Problem

Aim of this project is to find out the next probable location where there's a potential to open new 'Chipotle Mexican Grill' outlet.

For which real data on every Chipotle is explored to identify their presence in the states of USA.

Then to find out the next potential 'Chipotle' locations, based on several important factors, such as states with no or very less Chipotle outlets, state population, proximity to shopping centers, proximity to universities and the distance from tourist attractions.

This study is presently based on limited factors but can certainly be expanded to many other factors such as travelers who prefer Mexican food over other fast foods, areas where customers opt for vegetarian options (which is offered by Chipotle) adds value, user suggested locations for new stores (Chipotle website collects this information), cost effectiveness compared to other casual restaurant chains etc.

Data Aquistion & Cleaning

Mainly two data sources are used in this analysis.

Dataset giving details of US states is downloaded. This data has US states, cities with GPS coordinates (latitude and longitude) and city population.

Foursquare location data is used to find the locations of ‘Chipotle Mexican Grill’ outlets in different cities.

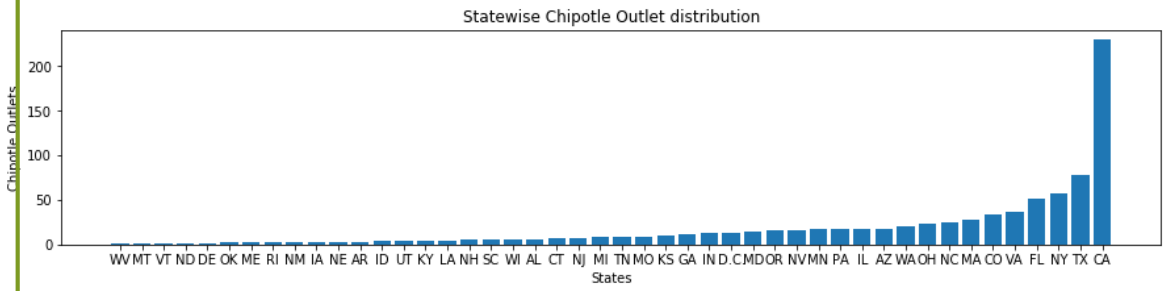
Locations data is further used to find Universities/Educational institutes and tourist attractions in those cities.

Population of cities is used further for analysis

Exploratory Data Analysis

Data is read from the available USA data source and then converted into pandas Dataframe. Data is filtered on the States with population >100,000

Using Foursquare API, chipotle outlets in different states of USA are fetched where population is >100,000 and store in pandas data frame. This data is filtered on states where there were no outlets or less than 3 outlets per state. Aim of this exercise is to find a city where Chipotle outlet can be set up.(Dataframe - df_r)



Missing values are treated as follows:

State id is replaced with state id of the city

Population is replaced with average population of the state

Predictive Modeling

Hierarchical clustering is used to find the prospective location for the new /next outlet. I have used the 'AgglomerativeClustering' function from scikit-learn library to cluster the dataset. The AgglomerativeClustering performs a hierarchical clustering using a bottom up approach. The linkage criteria determines the metric used for the merge strategy.

Dataframe after clustering

	location.state	location.city	No_of_Outlets	No_of_Universities	No_of_attractions	population	cluster_
0	DE	Dover	1.0	9.0	0.0	115352.0	0
1	ME	Portland	1.0	10.0	0.0	206023.0	0
2	ME	Westbrook	1.0	0.0	0.0	206023.0	1
3	MT	Billings	1.0	9.0	4.0	120800.0	0
4	ND	Fargo	1.0	9.0	0.0	204820.0	0
5	OK	Tulsa	2.0	10.0	1.0	672054.0	4
6	RI	Johnston	1.0	0.0	0.0	1206642.0	3
7	RI	Providence	1.0	9.0	2.0	1206642.0	2
8	VT	Burlington	1.0	8.0	0.0	108277.0	0
9	WV	Huntington	1.0	9.0	4.0	194155.0	0
10	MN	Moorhead	0.0	1.0	1.0	0.0	1
11	OK	Broken Arrow	0.0	5.0	1.0	108303.0	0
12	OK	Oklahoma City	0.0	9.0	1.0	955998.0	2
13	VT	South Burlington	0.0	2.0	0.0	108277.0	1
14	WV	Charleston	0.0	9.0	1.0	142858.0	0

Data cleaning

Population is replaced with average population of the state

Data after cleaning (sorted on States):

	location.state	location.city	No_of_Outlets	No_of_Universities	No_of_attractions	population
0	DE	Dover	1.0	9.0	0.0	115352.0
1	ME	Portland	1.0	10.0	0.0	206023.0
2	ME	Westbrook	1.0	0.0	0.0	206023.0
3	MT	Billings	1.0	9.0	4.0	120800.0
4	ND	Fargo	1.0	9.0	0.0	204820.0
5	OK	Tulsa	2.0	10.0	1.0	672054.0
6	RI	Johnston	1.0	0.0	0.0	1206642.0
7	RI	Providence	1.0	9.0	2.0	1206642.0
8	VT	Burlington	1.0	8.0	0.0	108277.0
9	WV	Huntington	1.0	9.0	4.0	194155.0
10	MN	Moorhead	0.0	1.0	1.0	0.0
11	OK	Broken Arrow	0.0	5.0	1.0	108303.0
12	OK	Oklahoma City	0.0	9.0	1.0	955998.0
13	VT	South Burlington	0.0	2.0	0.0	108277.0
14	WV	Charleston	0.0	9.0	1.0	142858.0

Discussion

The results are analyzed based on the current number of outlets, no of educational institutes (as students are good clients for fast casual food joints) and tourist attractions (Mexican food is a preferred choice for vegetarian tourists and also is a good alternative to the casual food served by restaurants like McDonalds and Subways). The additional influencing feature is the population of the city.

Cluster 0 (e.g. Portland and Charleston) which presently doesn't have Chipotle outlets and has 9+ educational institutes. However, there are hardly any tourist attractions and their respective city population is less than 0.2M

Cluster 1 and 3 have 1 or no outlets but very less educational institutes, tourist attractions and also less population.

Cluster 4 - Tulsa City in the state of OK which has large number of educational institutes, has a tourist attraction and a substantial population of .6M. However, it currently has 2 outlets.

Cluster 2 - looking at the cluster distribution, there are two cities (Oklahoma City, OK and Providence, RI) in Cluster 2 which have a potential to open outlets. This is because Oklahoma City in the state of OK, presently doesn't have any outlet. However, it has 9 universities with 1 tourist attraction and has a population of 955,000+. Providence in the state of RI has presently 1 outlet with 9 universities and 2 tourist attractions. Its population is more than 1.2 M.

Conclusion

Looking at the graph and the result of predictive model, **Oklahoma City** is the most probable city to open a next Chipotle Outlets. The other next probable location could be **Providece, RI**.