IBM DATA SCIENCE CAPSTONE PROJECT

OPENING NEW AUTOMOBILE WORKSHOP IN ATLANTA, GA, USA

By:

BALASUBRAMANIAN S

Introduction:

For many vehicle owners, servicing the vehicle periodically and keeping it in good condition is a headache. Since the showroom services are costlier, people are in search of local mechanic, who will do the work in much cheaper way. So, if you are a good mechanic who got lot of experience in that field and can do in an affordable way, are much needed in the society. Establishing the workshop is not an easy go job, you have to do some exploration work and analysis. The location where the workshop located plays a major and the location partially defines the success of the workshop.

Business Problem:

The main objective of the project is to analyse the problem and select the best location for the Automobile workshop in Atlanta. Using Data Science methodology and machine learning techniques like 'Clustering', the project aims to solve the business problem which is, '*If someone is a Mechanic who wants to establish his own workshop, where would you recommend him to open'*

Business or stake holders who are targeted by this project:

This project approach was not only applicable for someone who wants to open new workshop, it can be used by anyone who wants to open their shop like restaurants, service centres, …

Data:

Types of Data needed to solve this problem:

- The most important data are the location data of every workshop all over the city like latitude and longitude data
- Another data which helps us to find the better location is the customer rating of those shops, which helps to find out the good and bad service.

Sources of data:

Foursquare: The Foursquare Places API provides location-based experiences with diverse information about venues, users, photos, and check-ins. The API supports real time access to places, Snap-to-Place that assigns users to specific locations, and Geo-tag. Additionally, Foursquare allows developers to build audience segments for analysis and measurement

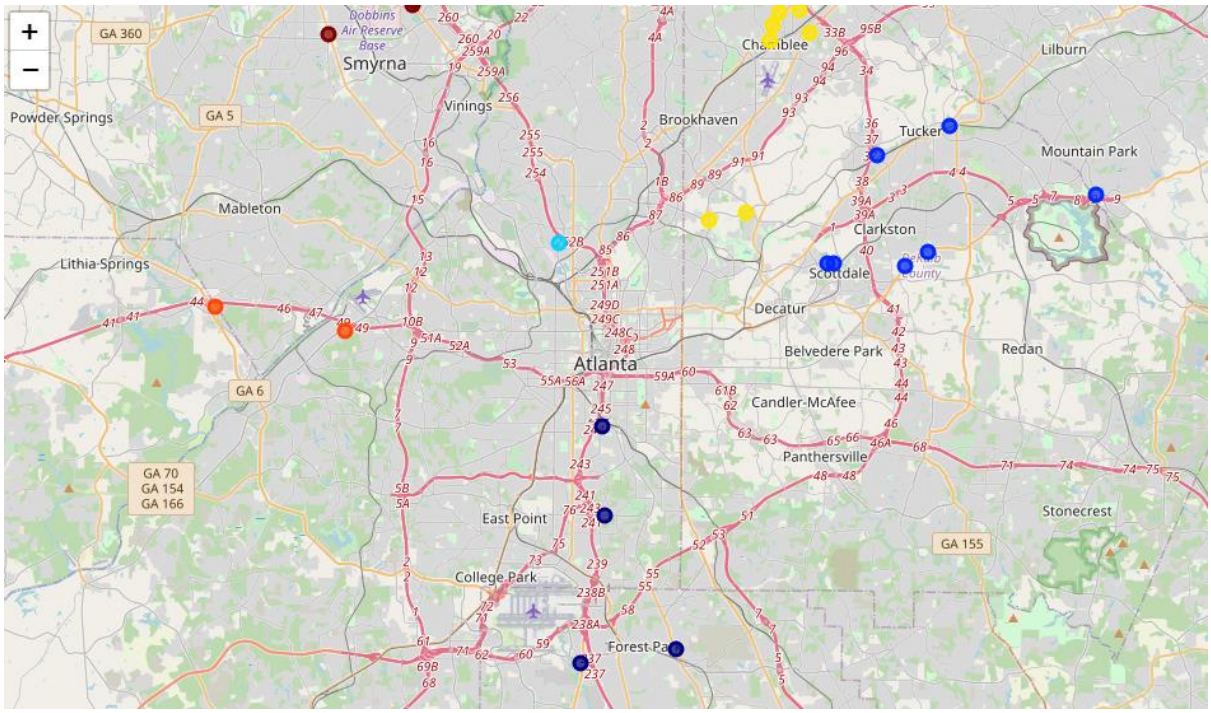
'geopy.geocoders' python library: This library provides the latitude and longitude data, by just giving it the name of location.

Methodology:

The Machine learning algorithm we used to solve in this project is ‘K-Means Clustering’, which is used to cluster the workshop over the region of area. This algorithm makes the data into clusters based on the location where the data points are available, it uses Euclidian distance formula to calculate the distance between the data points.

After grouping the workshop into clusters, we used mean formula to find the average customer ratings of the shop for each cluster.

Clustered Map:



Result:

After clustering the workshop into 7 clusters, it comes down that cluster 4 and 7 as the least. Among that while checking for customer reviews, it comes down that cluster 4 as the least average customer rating, which is 6.0.

Discussion Section:

From the above result we can see that, the group 4 and 7 has the least shop, and from that the cluster 4 has the least rating, so I would recommend the Business owner to open the shop in that region, which will be that first step for his success.

Conclusion:

The aim of this project is to find the best place to open the workshop in Atlanta. With the help of some Data scrapping, cleaning and analysing, and with the use of Machine learning technique we are able to find the best place to open the shop which result in the