Opening a Shopping Mall in Rome

December 16, 2020

1. Introduction
1.1 Background

- Shopping malls are the central attraction of cities nowadays. Malls are meant for shopping various stuffs and spend some time and relax and get refreshed.
- Malls are very helpful because you can find almost everything in one building like clothing stores, electronic items, book stores, theatres etc. There is no need to go to various places and search for the things.
- The location of the mall is the prime factor for planning to build a mall.

1.2 Business Problem

- In this Capstone Project, the objective is to find which are the prime places or spots to build a new shopping mall in the city of Rome. We will take various factors into consideration.
- We will use the K-Means clustering algorithm to help and recommend the building developer that which place would be the best to open a shopping mall.

2. Data Requirements

There are 3 following types of data which will be required:

- List of neighbourhoods of Rome: This list can be found out from the link https://en.wikipedia.org/wiki/Category:Subdivisions_of_Rome
- Latitude and Longitude: Getting the latitude and longitude of Rome by using Geocoder python library package.
- Venues: Venue data for neighbourhood of Rome could be found out with the help of Foursquare API (www.foursquare.com)

3. Methodology

- Firstly, we will get the data of neighborhoods in Rome. The data is available in https://en.wikipedia.org/wiki/Category:Subdivisions_of_Rome.
- We will use the python library BeautifulSoup to get the data and turn that data into a dataframe using pandas library.
- Next, we will use geocoder which is another python library package, in order to get the latitude and longitude of Rome and its respective neighborhoods.
- We will use Folium library to visualize the city of Rome and its neighborhoods in a map.

- Next, we will use Foursquare API to get the venues. From those we will get top 100 venues that are within a radius of 2000.
- We will filter 'Shopping Mall' as venue category.
- We will perform clustering on the data by using a machine learning algorithm k-Means clustering.
- We will cluster the neighborhoods into 3 clusters based on their frequency of occurrence of Shopping Mall.
- The result will show us that which clusters have higher or lower concentration of Shopping Malls around it.
- Based on the outcome, it will help the stakeholder to decide which location will be the best to open a new Shopping Mall.

4. Results

The outcome of the k-Means clustering categorizes into following 3 clusters:
- Cluster 0- The red colored marker represents moderate number of shopping malls.
- Cluster 1- The purple/blue colored marker represents no shopping malls around that locality.
- Cluster 2- The mint-colored marker represents high concentration of shopping malls in the locality/neighborhood.
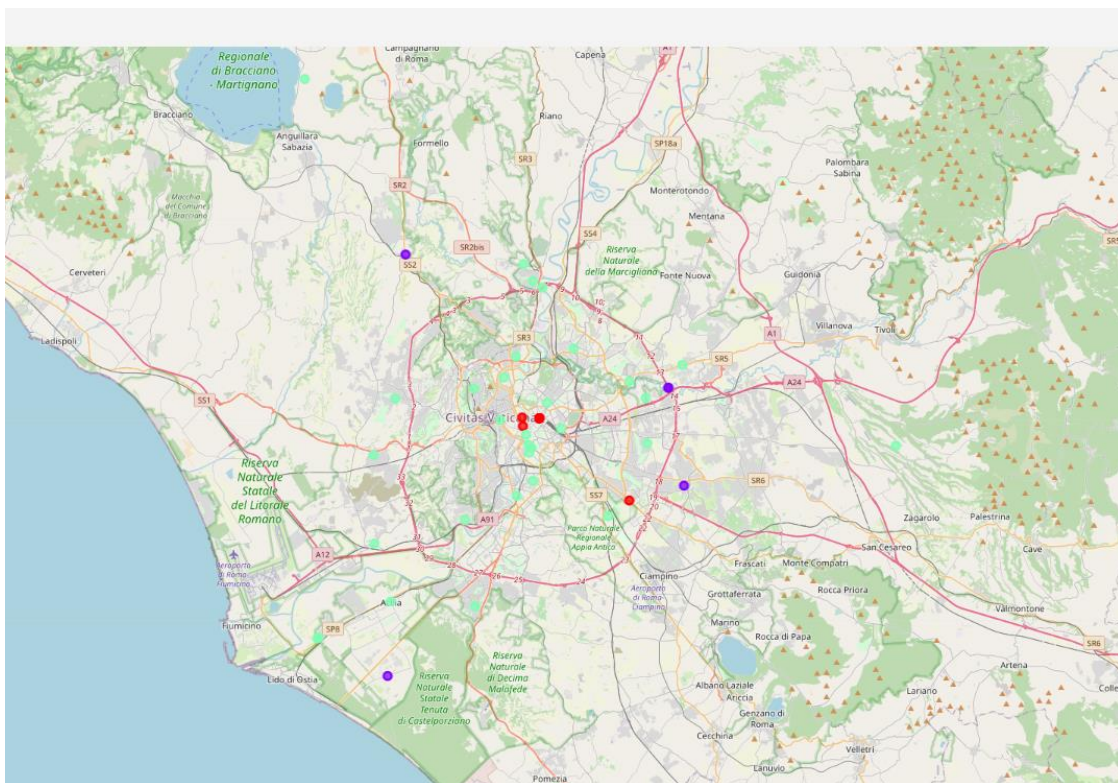
5. Discussion

- From the outcome, it can be observed that for the stakeholders the best place to open a new shopping mall will be Cluster 1 as there are no other Shopping malls nearby, which will ultimately turn into profit because there are no other competitors.
- Cluster 0 could also be a better option as there are less competitors.
- Cluster 2 would be the worst option as already there are large amount of shopping malls there and therefore face an intense competition.

6. Conclusion

In this project we have gone through fetching of data, analyzing it, cleaning it, visualizing it on a map and applying machine learning algorithm k-Means clustering in order to cluster the malls. We have divided the malls in 3 categories- no malls, moderate concentration of malls and high concentration of malls.

This project will provide recommendations to the stakeholders which would be the best location to open a new shopping mall that will make some profit. The locations in cluster 1 will be the best to open a new shopping mall.

7. References

- The list of neighborhoods in Rome:
  https://en.wikipedia.org/wiki/Category:Subdivisions_of_Rome