

Coursera Capstone

IBM Applied Data Science Capstone

Opening a new coffee shop in New York

-Shivang Agrawal

Introduction

For many people, coffee shops are places of recreation, relaxation, and act as places where they get an environment to work, write and many more things. To own a coffee shop in New York can be very profitable, but it matters where it is opened.

Business Problem

The objective of this project is to find the best neighbourhood to open a coffee shop in New York. Using data science and machine learning techniques like clustering, this project aims to fulfill the objective.

Target Audience

The primary audience is the one which is planning on opening a new coffee shop in New York, one who has to consider where to open it to gain maximum profits.

Data

Neighbourhoods; latitudes and longitudes of them; venue data of coffee shops.

Methodology

1. Build a dataframe of neighborhoods in New York
2. Get the geographical coordinates of the neighborhoods
3. Obtain the venue data for the neighborhoods from Foursquare API
4. Explore and cluster the neighborhoods
5. Select the best cluster to open a new coffee shop

Result and Discussion

Most of the coffee shops are concentrated in cluster 2 and moderate number in cluster 5. On the other hand, cluster 4 has a very low number to totally no coffee shop in the neighborhoods. This represents a great opportunity and high potential areas to open new coffee shop as there is very little to no competition from existing shops. Meanwhile, coffee shops in cluster 2 are likely suffering from intense competition due to oversupply and high concentration. Therefore, this project recommends coffee shop owners to capitalize on these findings to open new coffee shops in neighborhoods in cluster 4 with little to no competition. Coffee shop owners with unique selling propositions to stand out from the competition can also open new shopping malls in neighborhoods in cluster 1 or 3 with moderate competition. Lastly, coffee shop owners are advised to avoid neighborhoods in cluster 2 which already have high concentration of shopping malls and suffering from intense competition. Result

Conclusion

We identified the business problem, specified the data and then extracted it. Then we performed machine learning technique clustering, and then provided recommendations to coffee shop owners who are the stakeholders and the target audience. Finally, the results and discussions are given