Battle of Neighborhoods : Mumbai

IBM APPLIED DATA SCIENCE CAPSTONE PROJECT

BY SHIVANI MAHADDALKAR

AUGUST, 2020

Overview

- Introduction
- Business Problem
- Data
- Methodology
- Results
- Discussion
- Conclusion

Introduction

- Mumbai is India's financial capital and seventh most populous city in the world.
- Restaurants are a growing business with Indian and Chinese cuisines being the most popular amongst the people of Mumbai.
- This project aims to narrow down the pool of potential neighborhoods in which a restaurant with a fusion of both Indian and Chinese can become a profitable venture.
- The project uses an unsupervised machine learning algorithm to identify such neighborhoods.

Business Problem

- For an investor, looking to capitalize on the growing reliance on take-aways and fine dining experiences, and an affinity to both Indian and Chinese cuisines, an Indo-Chinese restaurant could be an interesting venture.
- With a numerous places with Indian and Chinese cuisines, but limited options for both cuisines in one place in one meal, can prove to give this idea a fairly unique selling point
- Such a business plan requires extensive planning and decisions between varied choices.
- In such a huge city, narrowing down the choices for the location, can be a huge competitive advantage.

Data

The data required for the analysis is obtained in three steps.

1. List of neighborhoods in Mumbai. They are categorized by their pin codes in a csv file obtained from a Mumbai Tour Guide page.
2. The geographical coordinates of the neighborhoods using the Python Geocoder package.
3. Venues present in these neighborhoods within 500 m radius of their geographical coordinates using Foursquare API.

Methodology

Data Extraction

- The data is in a csv file obtained from Mumbai Guide website(https://mumbai7.com/postal-codes-in-mumbai/), which contains the neighborhood's name and the pin code of the neighborhood.
- The Geocoder package is used to obtain the latitudinal and longitudinal coordinates from the pin codes of the neighborhoods in the data-frame.
- The latitudinal and longitudinal coordinates are used to extract information about venues within 500 m of the neighborhood using the Foursquare API.

Methodology

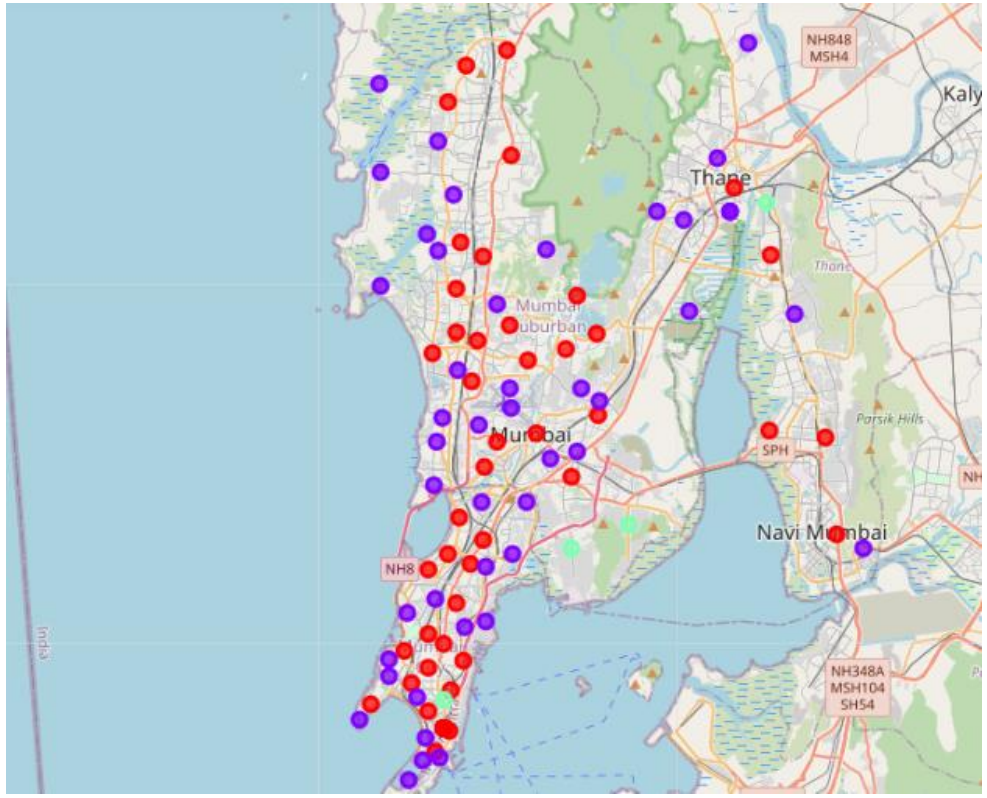
Data Analysis

- The venues within 500 m of the coordinates for every neighborhood are obtained.
- The frequency of each venue category in every neighborhood is calculated.
- The data is filtered by venue categories Indian Restaurant and Chinese Restaurant.
- The frequencies are summed and populated in a data-frame with the neighborhood names.
- The neighborhoods are divided into 3 clusters, where they are classified into clusters with centroids closest to the data points.
- The three clusters are high, moderate and low frequencies of Indian and Chinese restaurants in the neighborhood.

Methodology

Data Visualization

- The three clusters are visualized on the map of Mumbai city using the Folium package.
- Red, purple, and mint green rings represent the neighborhoods present in cluster 0, 1 and 2 respectively.



Results

The clusters are classified into high, medium and low concentration of Indian and Chinese restaurants. The clusters are as follows:

- Cluster 0: Medium frequency of Indian and Chinese restaurants
- Cluster 1: Low frequency of Indian and Chinese restaurants
- Cluster 2: High frequency of Indian and Chinese restaurants

Discussion

- In the cluster 0, moderate number of Indian and Chinese restaurants are present in these neighborhoods. The competition in these areas is decent and therefore it is prudent to avoid these neighborhoods.
- In cluster 1, low number of Indian and Chinese restaurants are present. The competition in these neighborhoods is very low, and picking one of the neighborhoods in this cluster will be a good decision.
- In cluster 2, high number of Indian and Chinese restaurants are present. The number of neighborhoods in this cluster are less and the high frequency of existing restaurants is a disadvantage to future investors looking to open restaurants in these neighborhoods.

Conclusion

- With a venture like this, with a lot of decisions to be made in starting such a business, a well-researched location for a restaurant can be a good starting point in planning to invest in an Indo-Chinese restaurant.
- The investors could add in more constraints to narrow down the pool further from the present one present in cluster 1.

Thank You!