
Coursera Capstone

IBM Certificate Data Science Course

Opening an Italian Restaurant in Mumbai, India

By Debjyoti Bardhan
June 2020



Introduction

Italian cuisine is prized and enjoyed the world over. It is the perfect food for a nice evening out, with a romantic partner, or with family or with a group of friends. A nice pizza or spaghetti with authentic cheese used in its making can be a part of a grand evening of dining, or a casual hangout session among friends. However, despite the proliferation of eateries - specifically pizzeria - in the city of Mumbai, there is a dearth of restaurants catering to genuine Italian cuisine. Since even eateries serving non-authentic Italian food are garnering good business, it can be surmised that there is genuine demand for such things in the market. The goal of this project is to determine the optimal locations for an Italian restaurant, whether it can be expanded into a chain and whether the possible business fate of the venture can be ascertained.

Business Problem

The primary question is this: In the Indian city of Mumbai, what are the optimal locations for a restaurants dedicated exclusively to Italian cuisine? The secondary question is to gauge whether the outlet will be a success or a failure, and the likelihood of expanding it as a restaurant chain in the city. In order to answer this, we shall use data for existing Italian restaurants and use various data science techniques such as k-means clustering to determine optimal locations.

Data

Data needed

The following data is needed:

1. The different neighbourhoods in Mumbai, along with other details, like population, and their geospatial data.
2. Locations of existing Italian restaurants

Obtaining the Data

Following the numbering used in the previous list:

1. The list of neighbourhoods and the sub-areas in each neighborhood can be obtained from Wikipedia. We can use web scraping to obtain the data from the relevant page, using packages such as BeautifulSoup, in a Python notebook. Wikipedia lists the

latitudes and longitudes of the neighbourhoods and these can be obtained similarly. This will utilise the datascience skills learnt in the course such as web-scraping.

2. One can use data from Zomato or the FourSquare API to explore the different Italian restaurants in each neighbourhood. Clustering of this data can be done using k-means clustering and visualised on a map using Folium.