Recommending Neighborhoods in Boston, Massachusetts for Opening a Restaurant

Applied Data Science Capstone Project

Surbhi Bector

Introduction

Aspiring business owners hold a significant amount of risk when evaluating decisions to create the ideal model and strategies to kickstart a successful business. Restaurant owners specifically handle a lot of planning and research, due to the complex nature of the operation, to better their chances for success.

According to Entrepreneur, one of the most challenging parts of starting a restaurant is the pre-planning that goes along with it. Many newly opened restaurants do not profitize in the first year of opening due to failure in planning. One crucial factor in planning a restaurant is the location to open it. Owners have many factors to consider such as, how many restaurants are near the location, is the location popular, are there other attractions within this location. Business owners can take advantage of research and data available on any location to then analyze what area would be the best location for their restaurant. Ideally, owners will want a location that is popular, has complementary attractions, and less competition.

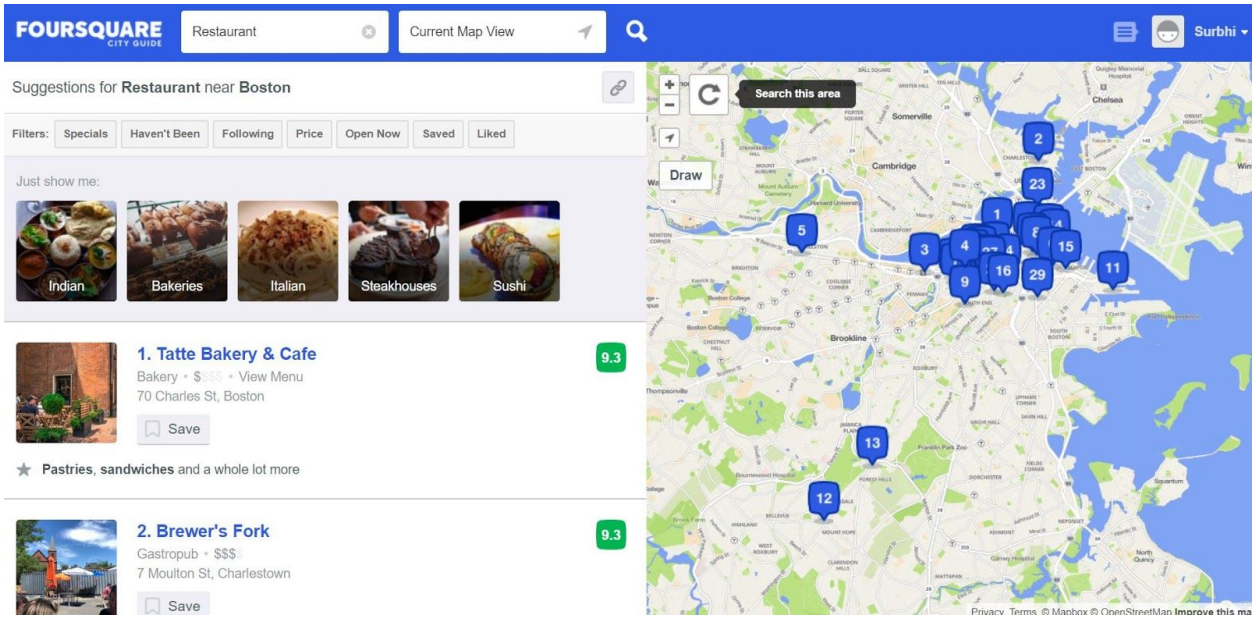
According to Fox Business, one of the top 10 locations to start a restaurant in the U.S. is Boston, Massachusetts, due to the high amount of annual restaurant sales and high median income level in the city.

The goal of this project is to recommend top neighborhoods in Boston for aspiring restaurant owners to open a restaurant based on existing data available for the area. This research will assist new restaurant owners and owners looking to expand franchises by providing advantageous information to plan their restaurant location.

Data

I will utilize Foursquare API to extract information on attractions in different neighborhoods in Boston, such as their location and ratings. The data leveraged by the API will include information on venues such as restaurants, theaters, cinemas,

malls, etc. This information will indicate the advantages and disadvantages of the neighborhoods. For example, a disadvantage would be many more restaurants and therefore competition in the area. An advantage would be complementary attractions to a restaurant, such as a movie theater or shopping mall.



Web scraping data on neighborhoods is being done through https://en.wikipedia.org/wiki/Neighborhoods_in_Boston. This page contains a list of official neighborhoods in the Boston area that can be used to find coordinates of each neighborhood.

The GeoPy library is used to obtain latitude and longitude for neighborhoods in Boston to visualize the maps of the locations using Folium. The K-means algorithm will be used to cluster neighborhoods. Data cleaning and wrangling will be used to examine clusters and their most common venues. Based on these defining categories, the clusters with the more advantageous location will be recommended to business owners.