COURSERA CAPSTONE PROJECT

Finding the best neighborhoods to visit in Boston, MA.

Introduction

In this project, I'll try to check if its possible to cluster the neighborhoods in Boston, MA by taking into account the quality and cost of the venues located per neighborhood.

MOTIVATION

Travelers nowadays are constantly trying to find destinations that offer high quality food and drink for a fair price. Businesses also have to keep track of their neighborhood in order to understand their competitors and make strategic location decisions.

OBJECTIVE

Determine the areas in the selected destination that would be a good option for visitors looking for cheap, high quality entertainment.

Data sets

I'll be using two datasets, the Foursquare API and a US zip codes dataset downloaded from the simplemaps website.

Foursquare API

General description:
Foursquare is a location service that allows users to explore the venues around them. Foursquare API allows developers to access that information.

Information used:

- Venues (name and location) per zip code in Boston.
- Venue details (price tier and rating)

US zip codes dataset

General description:
This dataset contains the information on all the zip codes in the US along with the city name, states, latitude and longitude.

Information used:

- Zip codes in the selected city.

Methodology

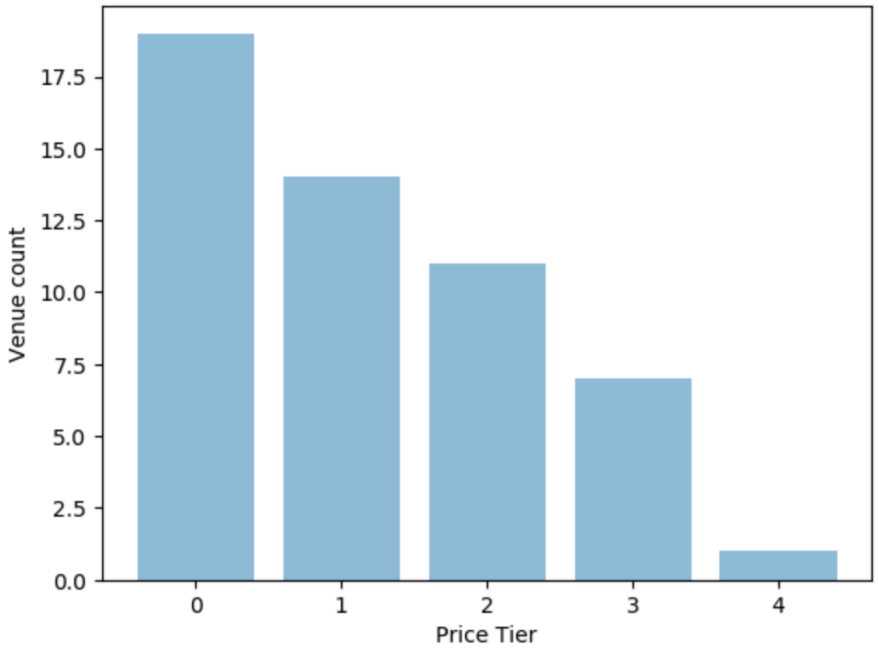
These are the general steps required to complete the project:

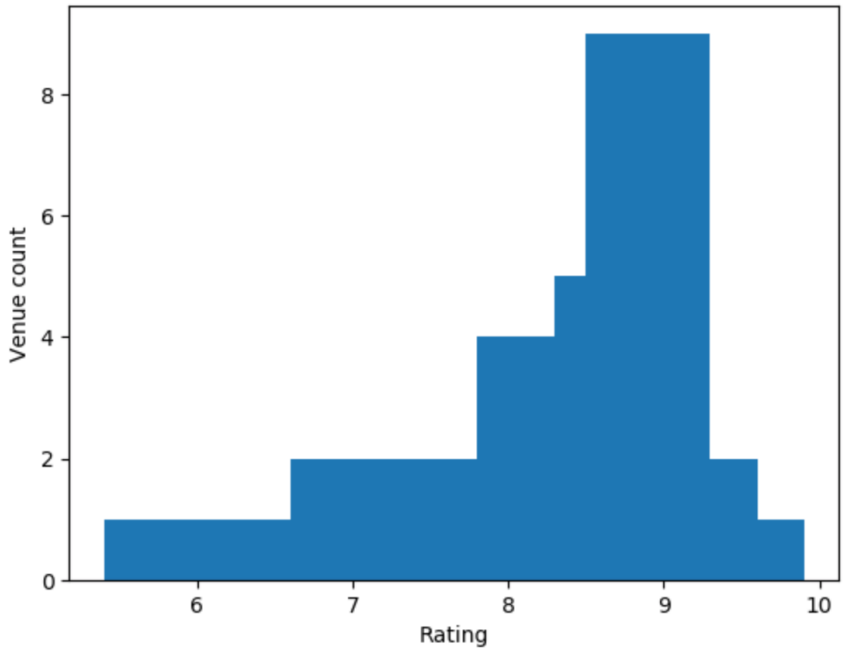
1. Obtain the data from the US zipcodes dataset.
2. Get the venues' names for the selected zip codes from the Foursquare API.
3. Arrange the data to fit it into the kmeans model.
4. Use kmeans clustering to cluster the neighborhoods based on the rating and prices of the venues.
5. Analyze the clustering results.

Results

I decided to run the kmeans clustering algorithm for three clusters. These are the resulting clusters' spread regarding the price tier and the rating.

CLUSTER 0





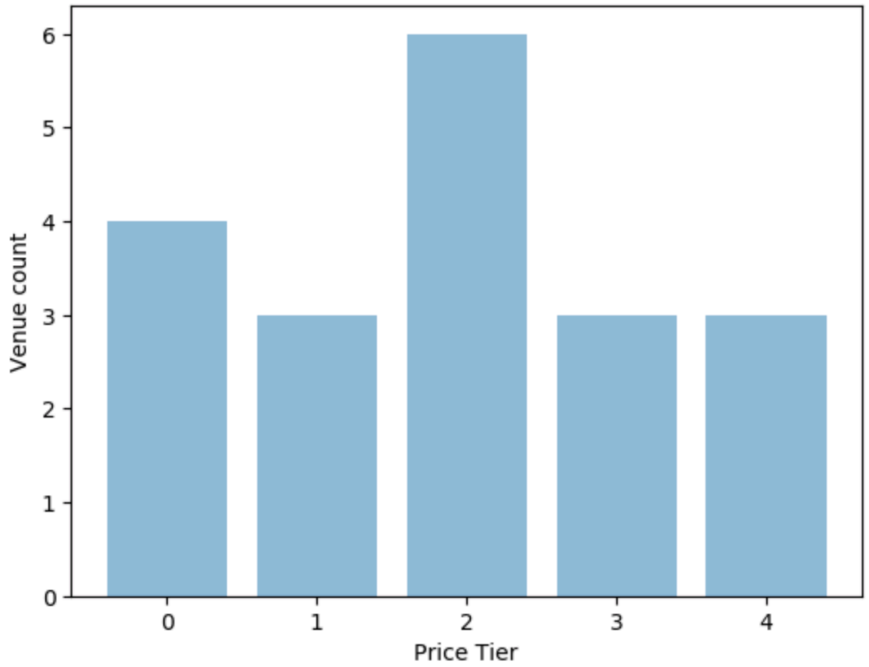
Notes:

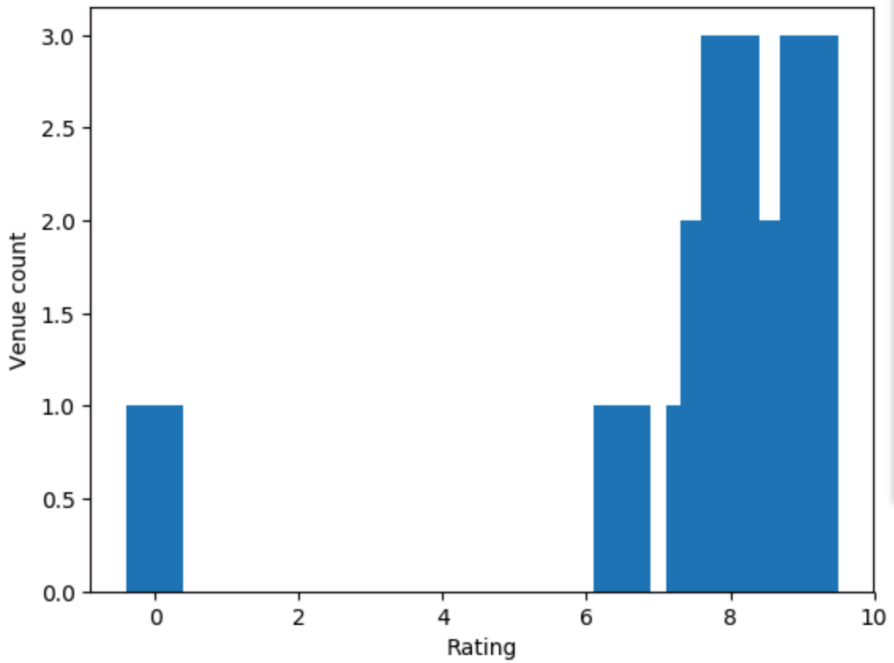
- In cluster 0 most venues are "cheap".
- Most venues in this cluster are scored between 8.5 and 9.5.

Results

I decided to run the kmeans clustering algorithm for three clusters. These are the resulting clusters' spread regarding the price tier and the rating.

CLUSTER 1





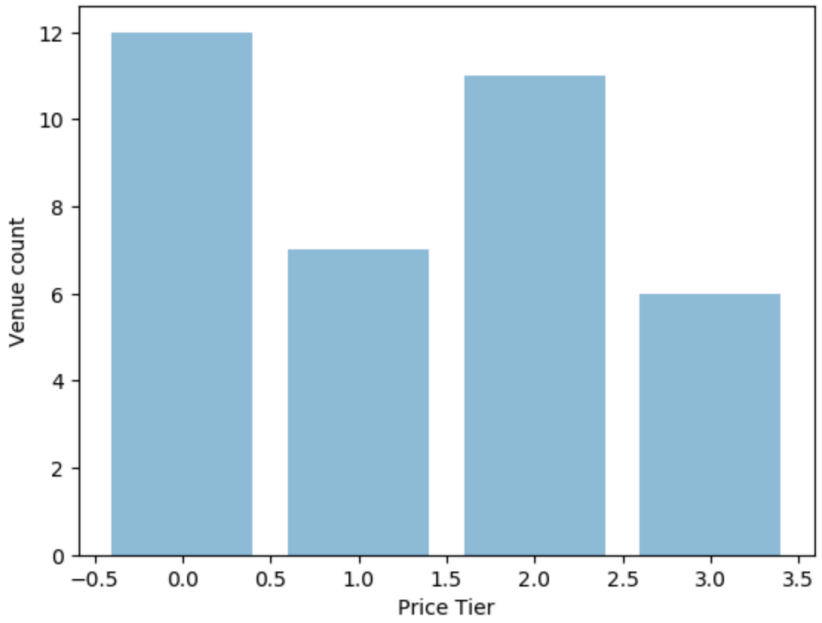
Notes:

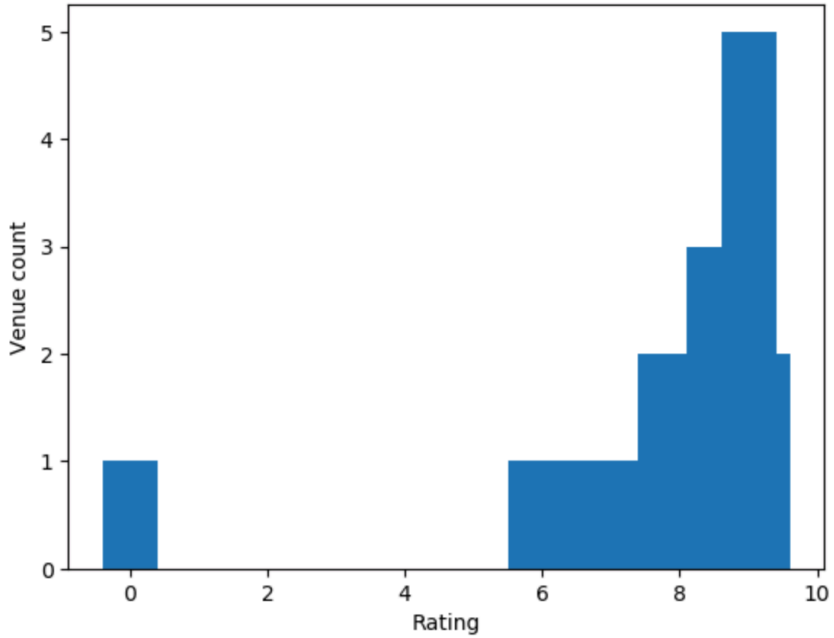
- Most venues in this cluster have an "average price".
- Most of these venues score around 8 and 9.5

Results

I decided to run the kmeans clustering algorithm for three clusters. These are the resulting clusters' spread regarding the price tier and the rating.

CLUSTER 2





Notes:

- in this cluster range from very cheap to average pricing.
- Most venues score at least 9 in this cluster.

Conclusion

With this project, I used Foursquare API data and US zipcodes dataset to cluster neighborhoods in Boston, MA based on the price and ratings of the venues in the neighborhoods.

After getting the data and preparing it, I split the data into three clusters. Based on this analysis, the neighborhoods in the third cluster (zip codes 2108, 2114, 2113 and 2109) seem to be the best option for visitors. It also seems to have major number of venues.

Future work could zoom into these neighborhoods to analyze the categories of the venues in these areas.