

# MARKET ANALYSIS FOR ITALIAN RESTAURANTS IN BOSTON



CAPSTONE PROJECT - HANNI ALSUFIANI

# 1- INTRODUCTION

- Boston is a highly dynamic city with a competitive market
- Opening an Italian restaurant needs a study of the market
- The aim is to understand the market and cluster similar neighborhoods to help knowing the market and the business opportunity



## 2- THE DATA

- In order to conduct the study we'll be needing the following data:
  1. Zip codes with their location coordinates
  2. Population in each zip code area
  3. Average household income in each zip code area
  4. number of Italian restaurants in each Zip code area
  5. GeoJSON data representing the geographical boundaries of each zip code area

## 2- THE DATA

- The data sources will be the following:
  1. For Zip codes, coordinates, population, and income data from the below link:  
<http://zipatlas.com/us/ma/boston/zip-code-comparison/median-household-income.htm>
  2. For Venues (Italian restaurants) in each location: using Foursquare API, this will help analyze the existing Italian restaurants.
  3. The GeoJSON data will be downloaded from  
[https://opendata.arcgis.com/datasets/53ea466a189b4f43b3dfb7b38fa7f3b6\\_1.geojson](https://opendata.arcgis.com/datasets/53ea466a189b4f43b3dfb7b38fa7f3b6_1.geojson)

## 3- METHODOLOGY

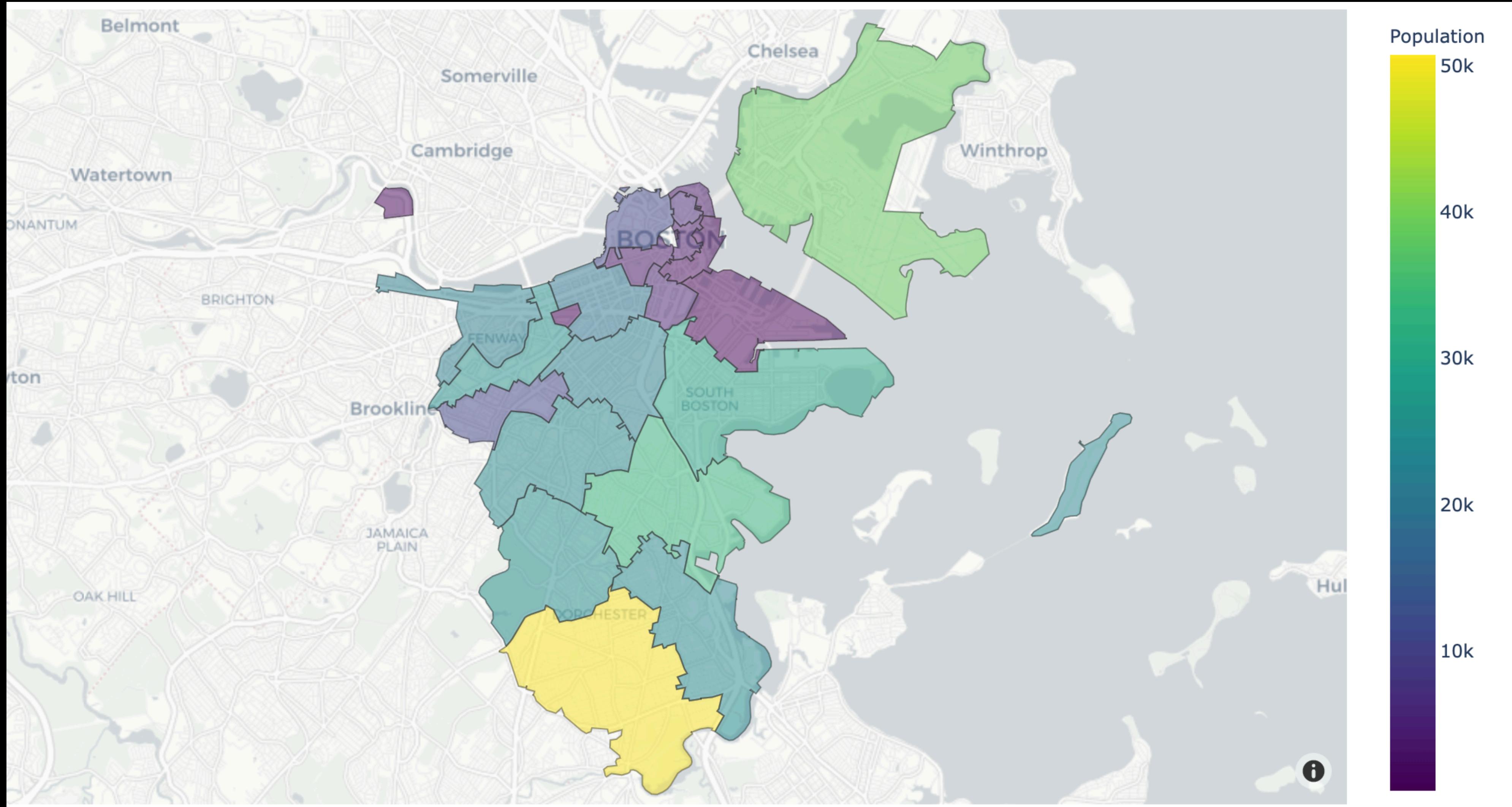
- Target Analysis

1.The target analysis is to cluster the zip code areas into similar groups and understand how are they similar

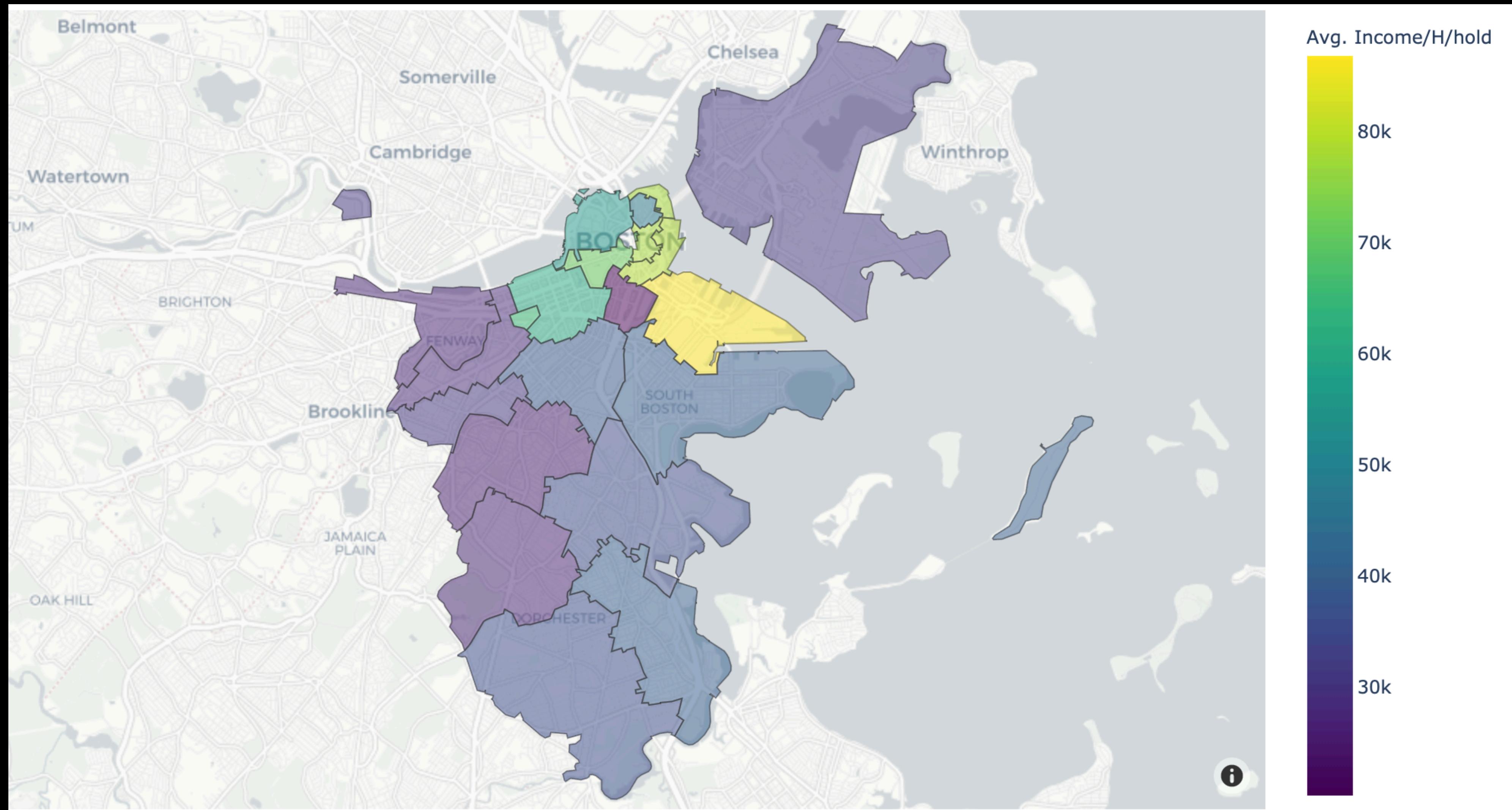
2.In the process to achieve the target analysis, we need to look over the following to understand the market:

- Population Distribution over the areas
- Average income per household for each area
- No. of Italian Restaurants in each area (the competition)

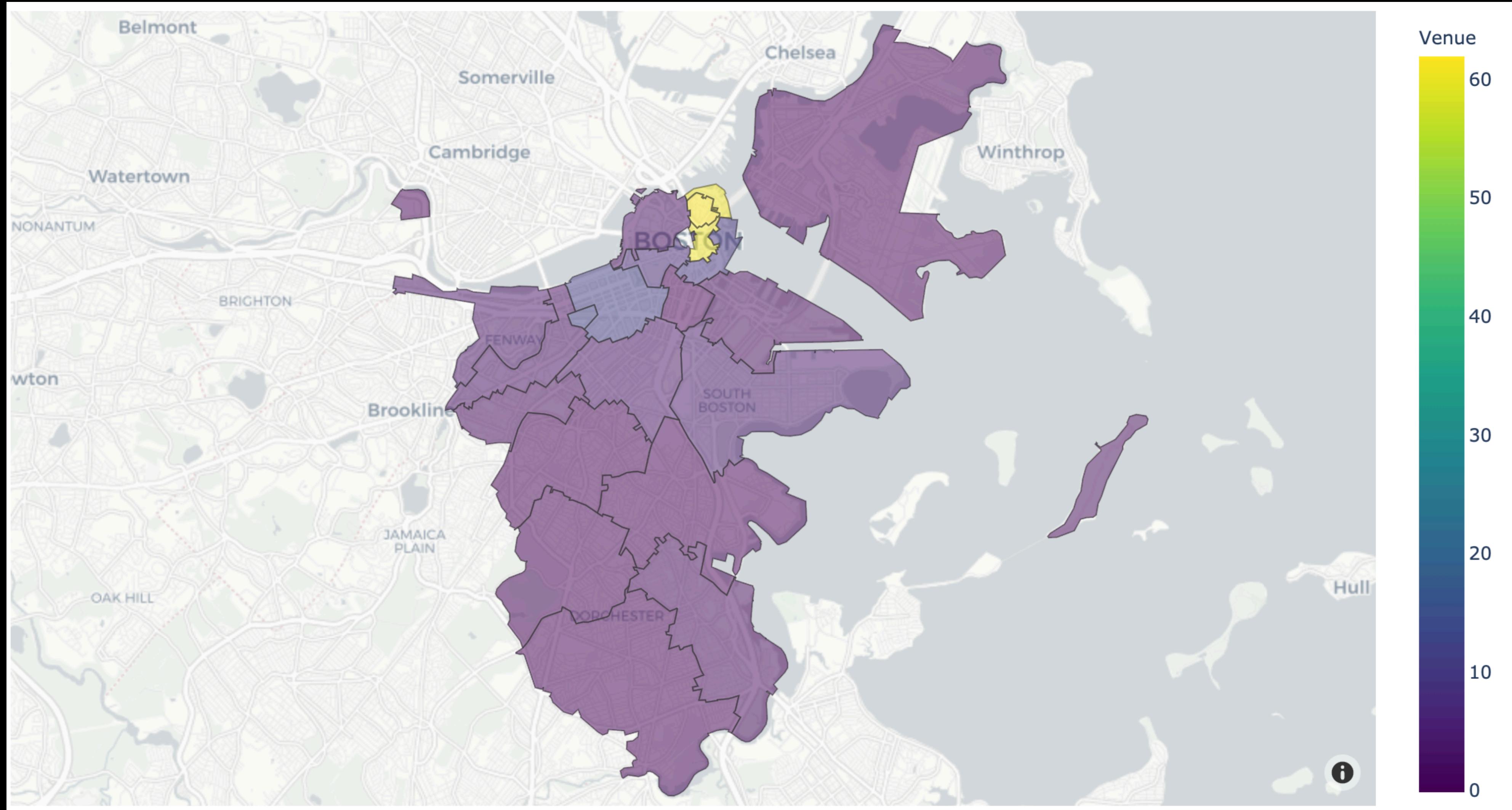
### 3.1- POPULATION BY ZIP CODE AREA



## 3.2- AVG. INCOME BY ZIP CODE AREA

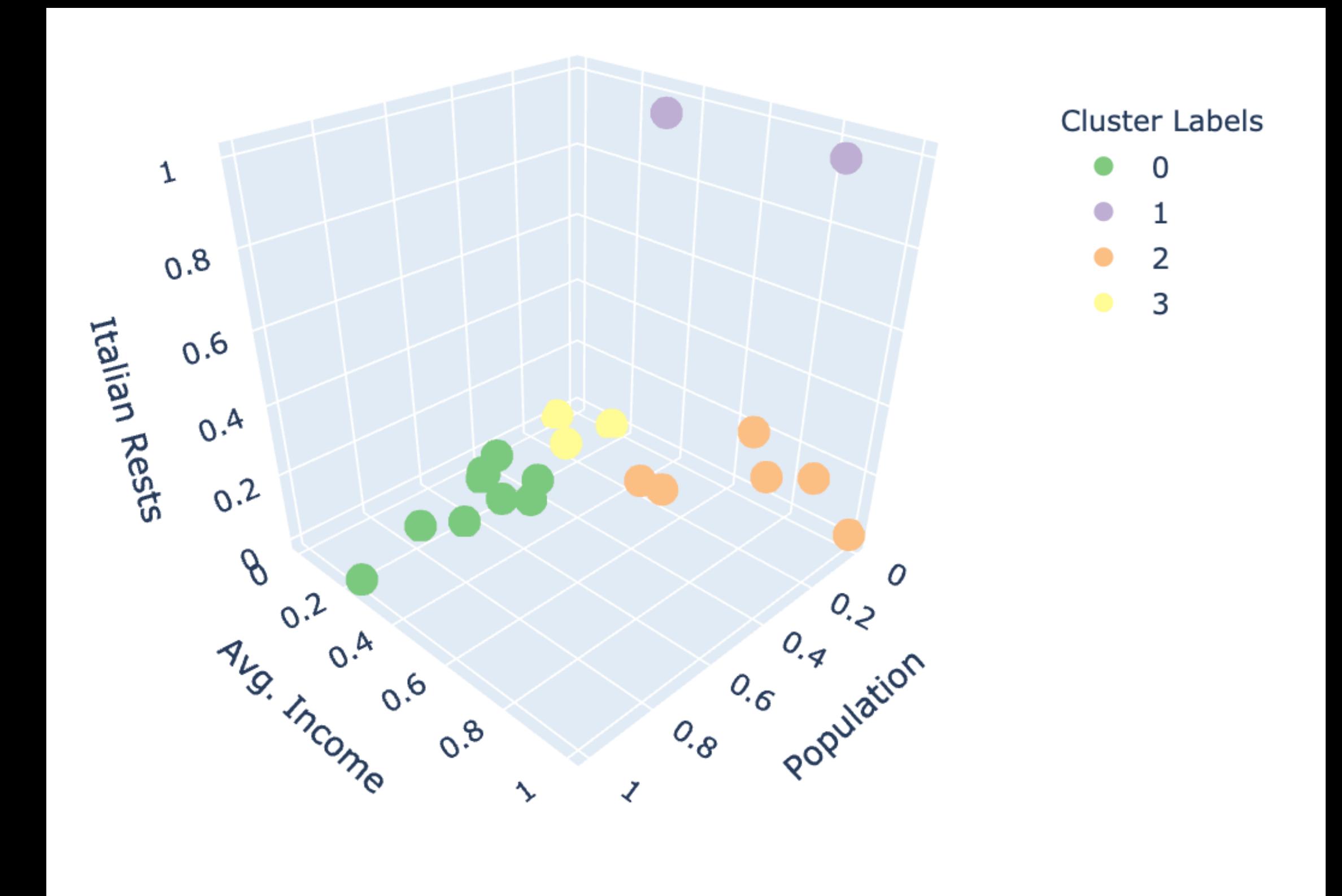


### 3.3- ITALIAN RESTAURANTS BY ZIP CODE AREA

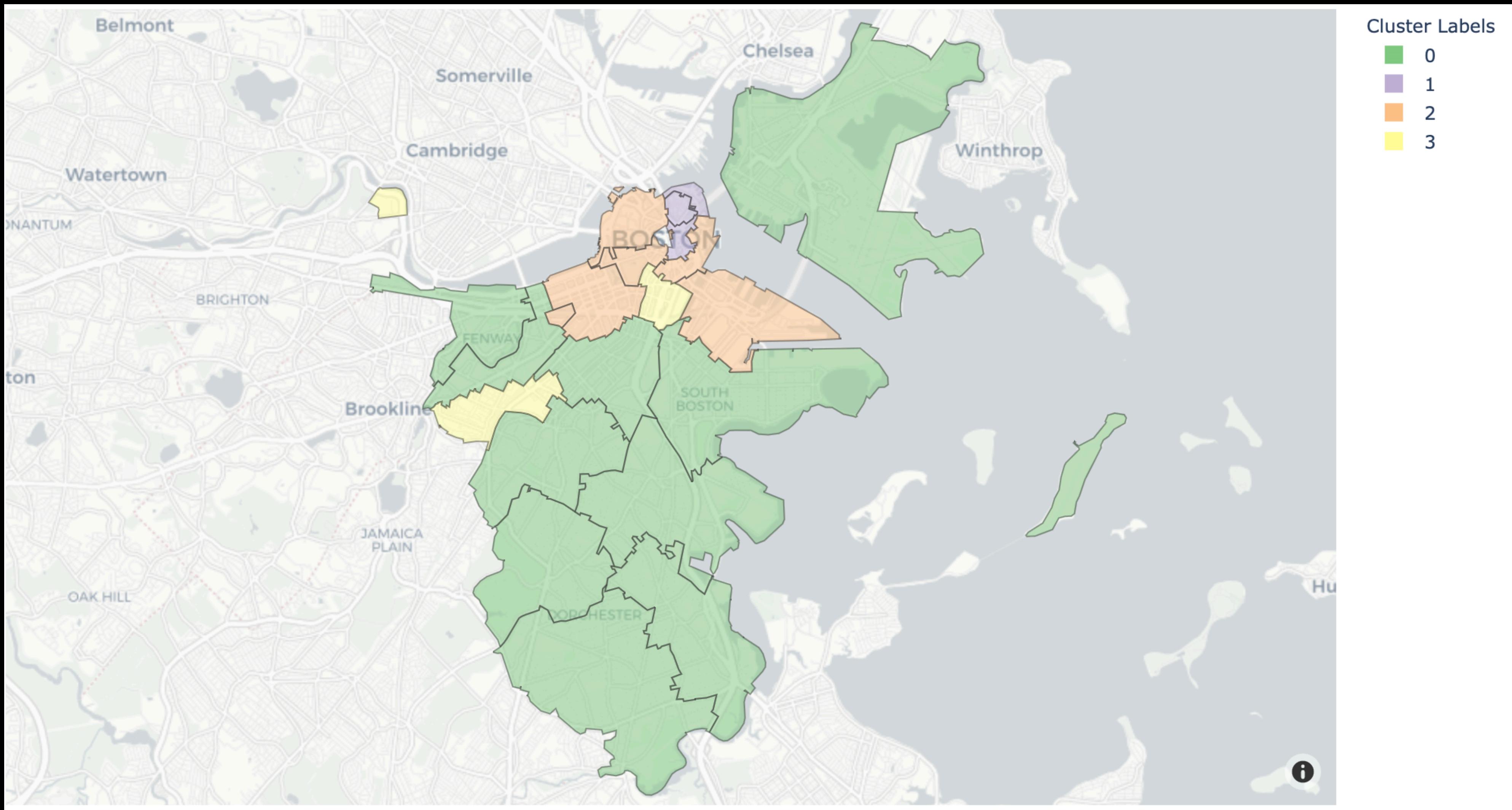


## 3.4 CLUSTERING THE ZIP CODES

- Since we're clustering using 3 features, we can plot them in a 3D Scatter plot
- This plot helps us understand what each cluster represents



## 3.4- CLUSTERING THE ZIP CODES - MAP



# 4- RESULTS & OBSERVATIONS

1. The cluster (1) with the purple dots:

1.1 Many Italian Restaurants with relatively low population

1.2 Areas difficult to penetrate due to the relatively high competition

2. The cluster (2) with the orange dots:

2.1 Similar profile of low population/high income areas but lower competition

2.2 Better business opportunity as it has lower competition

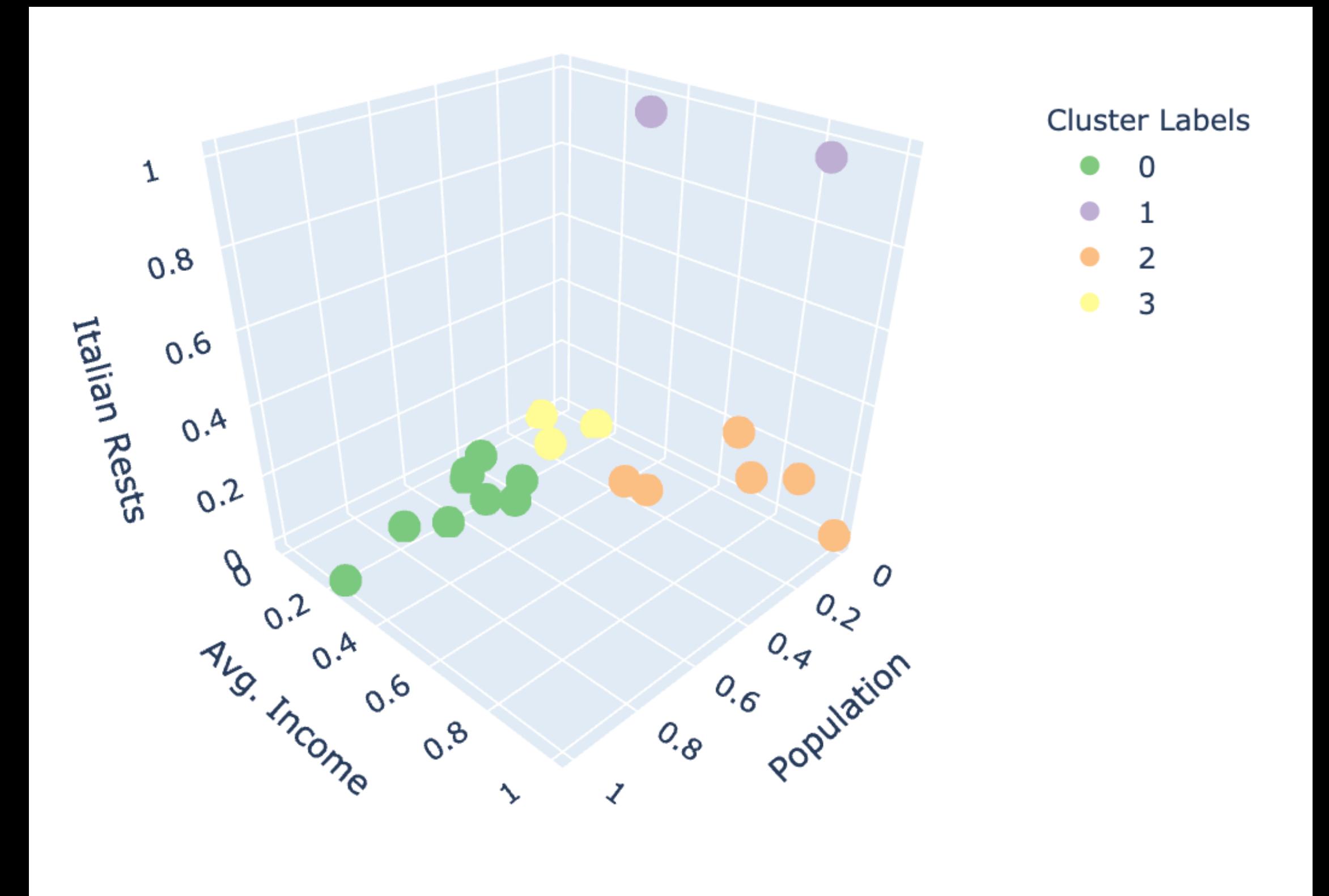
3. The cluster (3) with yellow dots:

3.1 Low population and Low income areas, less attractive as a location

4. The cluster (0) with green dots:

4.1 shows better locations than the yellow cluster since it's more populous.

5. When the green cluster (0) is compared to the orange cluster (2) we can see that the green cluster has relatively lower competition and higher population, however the avg. income is considerably lower



## 4.1 - RECOMMENDATIONS

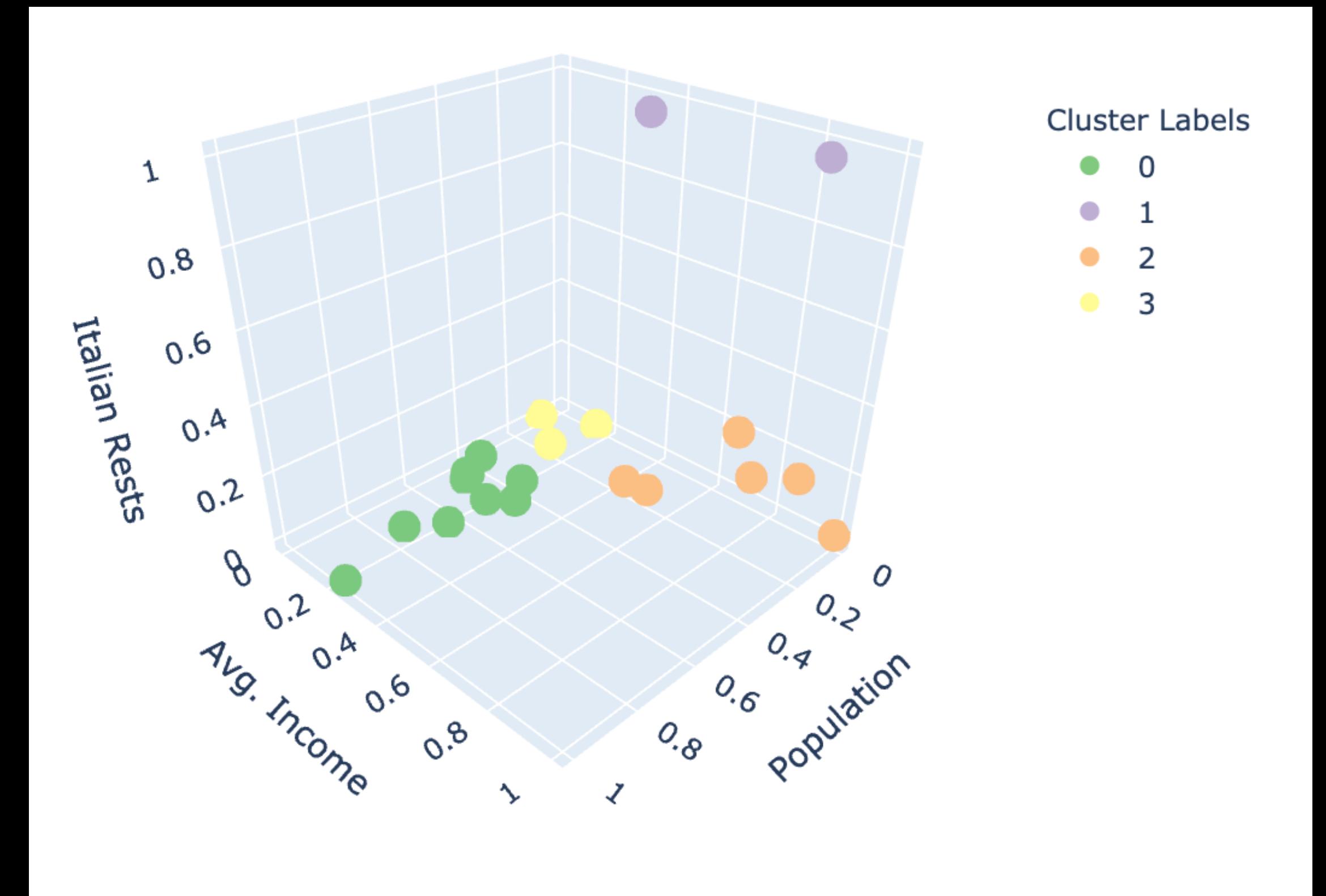
1. Two most appealing clusters are 0 & 2 (Green & Orange) and here are the recommendations for them:

2. Cluster 0 (Green):

1. More residential (relatively higher populations)
2. Relatively lower population
3. A bit less competition
4. **Suitable for Casual Dining and Fast Food Italian restaurants**

3. Cluster 2 (Orange):

1. Similar to Cluster 1 in terms of Population and Avg. Income, but less competition
2. Higher Avg. Income than Cluster 0 (Green)
3. **Fits the profile of a high-end Italian Restaurant**



## 5- CONCLUSION

- Analysis for the market status
- Looked at heat maps of Boston showing:
  - Population by area
  - Average income per household by area
  - No. of Italian Restaurants (competition) by area
- Clustered the areas into similar groups