

# Applied Data Science Capstone Project

---

**Location analysis to open a Bistro & Wine Bar in the City of Boston**



# OVERVIEW

---

1. Introduction
2. Business Problem
3. Data Description
4. Methodology
  - 4.1 Data acquisition
  - 4.2 Exploratory data analysis
    - 4.2.1 Search parameters
    - 4.2.2 Filtering data results
    - 4.2.3 Venue ratings
    - 4.2.4 Category analysis
5. Results
6. Discussion section
7. Conclusion

## 1. Introduction

---

- A group of investors is interested in opening a Bistro & Wine Bar in the City of Boston, and is needing a thorough location analysis before moving ahead with other important business considerations.
- The investors are primarily professionals from the banking industry and wine enthusiasts.
- As experienced business people they have the knowledge and the means to evaluate the opportunities and risks of starting a new business using a structured decision making process.
- It is important to categorize the bistro from restaurants, grills, sports bars and pubs.
- Although they serve food and drinks, a bistro is defined as a small, inexpensive restaurant which serves food at moderate prices in a casual or modest setting. The atmosphere inside the premises is relaxed and informal.
- The plan is to have a nice Mediterranean interior design, a reduced but top quality food menu and a great selection from major wineries from California, France, Italy and Spain.
- The goal is to make it a new landmark in the city!

## 2. Business Problem

---

- One key element of the entrepreneurial success for any brick & mortar business starts with choosing the right location for the venture. The old quote “location, location, location” is still the number one rule in real estate.
- The real estate location analysis considers the aspects of the attractiveness of the neighborhood, population density, demographics, average income, built-in traffic, tourism expenditure, economic trends, and public safety indicators among others.
- This preliminary study will identify the existence of direct or indirect competition in the same target area or neighborhood.
- The favorite areas under consideration by the investors are the Boston Downtown Business District, Back Bay, Beacon Hill and Waterfront, as shown in the city map below within the red squares.
- In summary, this analysis will address the following to the client investors:
  - Where are the best locations to establish the Bistro & Wine Bar?
  - Where and who are the closest competitors?

### 3. Data Description

---

- The location analytics will rely on data supplied by official sources, such as local government planning and development agencies, and data companies offering free access to data sets. All available datasets will be scrutinized for its reliable sources and current data, which includes:
  - Neighborhoods and postal codes
  - Geo coordinates
  - Venue data
- The City of Boston data hub ANALYZE BOSTON, offers 161 datasets for public access, which includes a variety of topics such as zoning districts, crime incident reports and economic indicators that will be used in this study.
- Reference places in the preferred investor's neighborhoods are Sheraton Boston, Hilton, at the Back Bay and the Marriott Long Wharf and Renaissance Hotel at the Waterfront.

- **City of Boston**

Population: 679,413  
Median income: \$65,883  
Median home: \$487,300  
Total crimes per 100k/people: 2,638  
Total area: 48 sq miles = 124km<sup>2</sup>

- **Back Bay**

Population: 17,557  
Median income: \$88,469

- **Beacon Hill**

Population: 9,305  
Median income: \$93,033

- **Downtown**

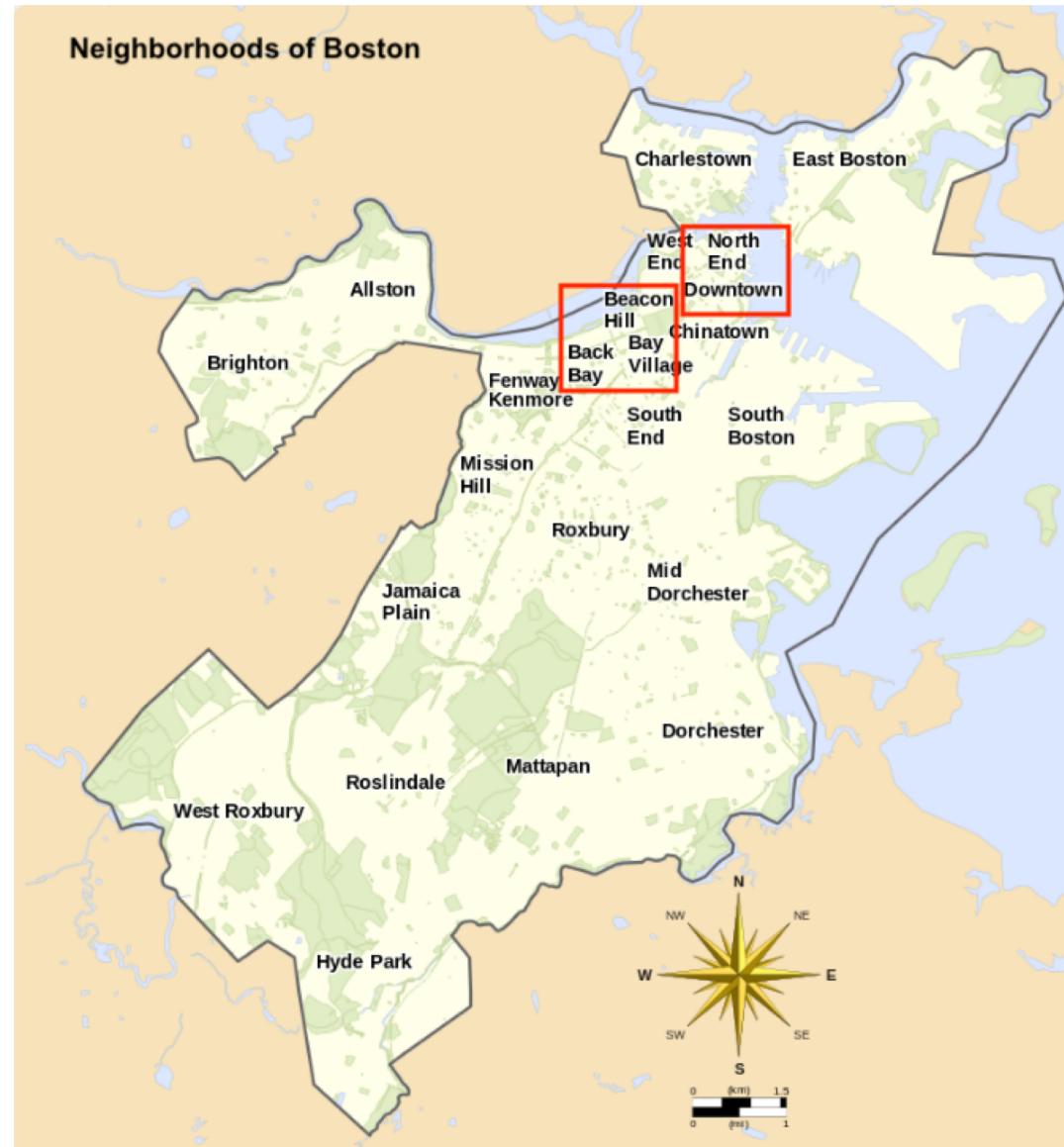
Population: 9,305  
Median income: \$93,033

- **North End**

Population: 9,107  
Median income: \$82,965

- **Waterfront**

Population: 9,107  
Median income: \$82,965



## Data sources and uses:

<https://data.boston.gov/>

The City of Boston data hub – ANALYZE BOSTON

<https://datausa.io/profile/geo/boston-ma/>

Created by MIT MediaLab and Deloitte for mining and visualizing US government open data

<https://www.topozone.com/massachusetts/suffolk-ma/>

## Boston Topographic maps

<https://www.findlatitudeandlongitude.com/>

## Maps and geo coordinates

<https://www.homesnacks.net/about-us/>

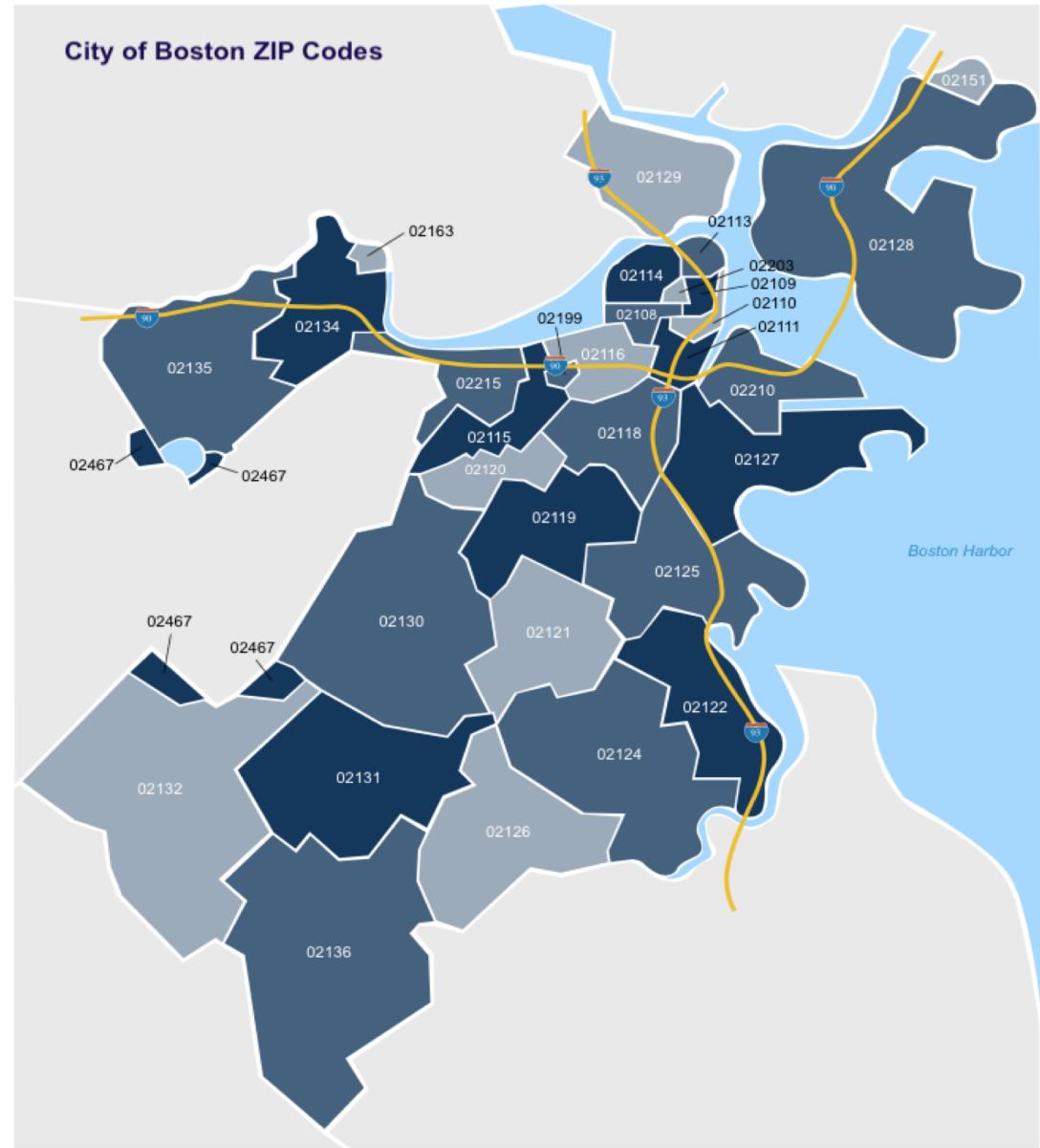
Combines recent data from the Census, FBI, OpenStreetMaps

<https://www.neighborhoodscout.com/ma/boston/crime>

## Public safety and crime incident

<https://public.opendatasoft.com/explore/dataset/us-zip-code-lat-and-long/>

## Postal codes and geo coordinates



## 4. Methodology

---

### 4.1 Data acquisition

- Initial search to select reliable and current datasets containing the City of Boston neighborhoods, postal codes and geo coordinates. This is the required data to be used to plot and identify locations in the maps and calculate distances between reference points.
- Acquire the datasets, through downloads or web scraping and combine them in useful format in pandas dataframes. The first built dataframe [neighborhood name, zip postal code, and geo coordinates] was the result of the concatenation of data coming from separate sources.
- Data cleaning step to verify missing, incomplete, duplicate or mismatched data. A reality check is necessary to ensure a reliable and accurate data table to be used in the subsequent steps.

## 4. Methodology

---

### 4.2 Exploratory data analysis

- The Foursquare API is the main tool used to locate and identify venues in the selected neighborhoods in the city. It is used to find venues, ratings and number of tips. Venue categories will be grouped and counted. A Developer's account was open in order to get access rights to the use of the data.

#### 4.2.1 Search parameters

- Interactive search with the keywords 'wine bar' and 'bistro' is used in the venue category. It is important to differentiate the nature of the business from regular restaurants, pubs, grills and sports bars. In case of any of these categories appear in the search result, additional exploratory analysis is made to determine whether it represents a direct or indirect competition.
- The other Foursquare parameter influencing the dataframe results is radius = 1000 meters from the reference points, after trying the value of 500 meters that did not bring enough representative venues. The parameter "Limit =" that limits the number of return results was set to 30.
- As the investors are mainly interested in upscale areas, the reference points selected were Sheraton Boston in the Back Bay and Marriott Long Wharf at the Waterfront.

## 4. Methodology

---

### 4.2 Exploratory data analysis

#### Filtering results

- The results table requires a sanity check and some cleaning. The words "wine" and "bar" in the search query retrieved venues with the "bar" in the name such as Barnes & Noble (bookstore) and barbershop. The Genius Bar, is the name used by the Apple for its tech support station located inside Apple's retail stores. Besides that, wine retail stores were also part of the returned list. A visual inspection determined the drop list of all these deviations to create the dataframe filter.
- The other Foursquare parameter influencing the dataframe results is radius = 1000 meters from the reference points, after trying the value of 500 meters that did not bring enough representative venues.  
As the investors are mainly interested in upscale areas, the reference points selected were Sheraton Boston in the Back Bay and Marriott Long Wharf at the Waterfront.

## **4. Methodology**

---

### **4.2 Exploratory data analysis**

#### **4.2.2 Mapping results**

- Folium is used to map the results loaded into the dataframe obtained from Foursquare, and superimposing the locations over the neighborhood map within a radius of 1000 meters from the reference points.

#### **4.2.3 Venue ratings**

- The venue ratings is a ten-point scale used to provide users with information about the best places around. It is a result of direct users' feedback and one indication how well the venue is performing. This will be used to get us a glimpse of the place in cases where it is believed to be a direct or close competitor.

#### **4.2.4 Category analysis**

- Group identical returned categories and plot the total number for each one.

## 5. Results

---

- The generated maps provide a good overview of the distribution of the filtered business establishments within a radius of 1000 meters of the reference points, which also covers neighbor areas such as Beacon Hill, Bay Village in the first case, and Downtown and North End in the second case.

Reference point 1:

Sheraton Boston Hotel at the Back Bay  
39 Dalton Street, Boston, MA 02199

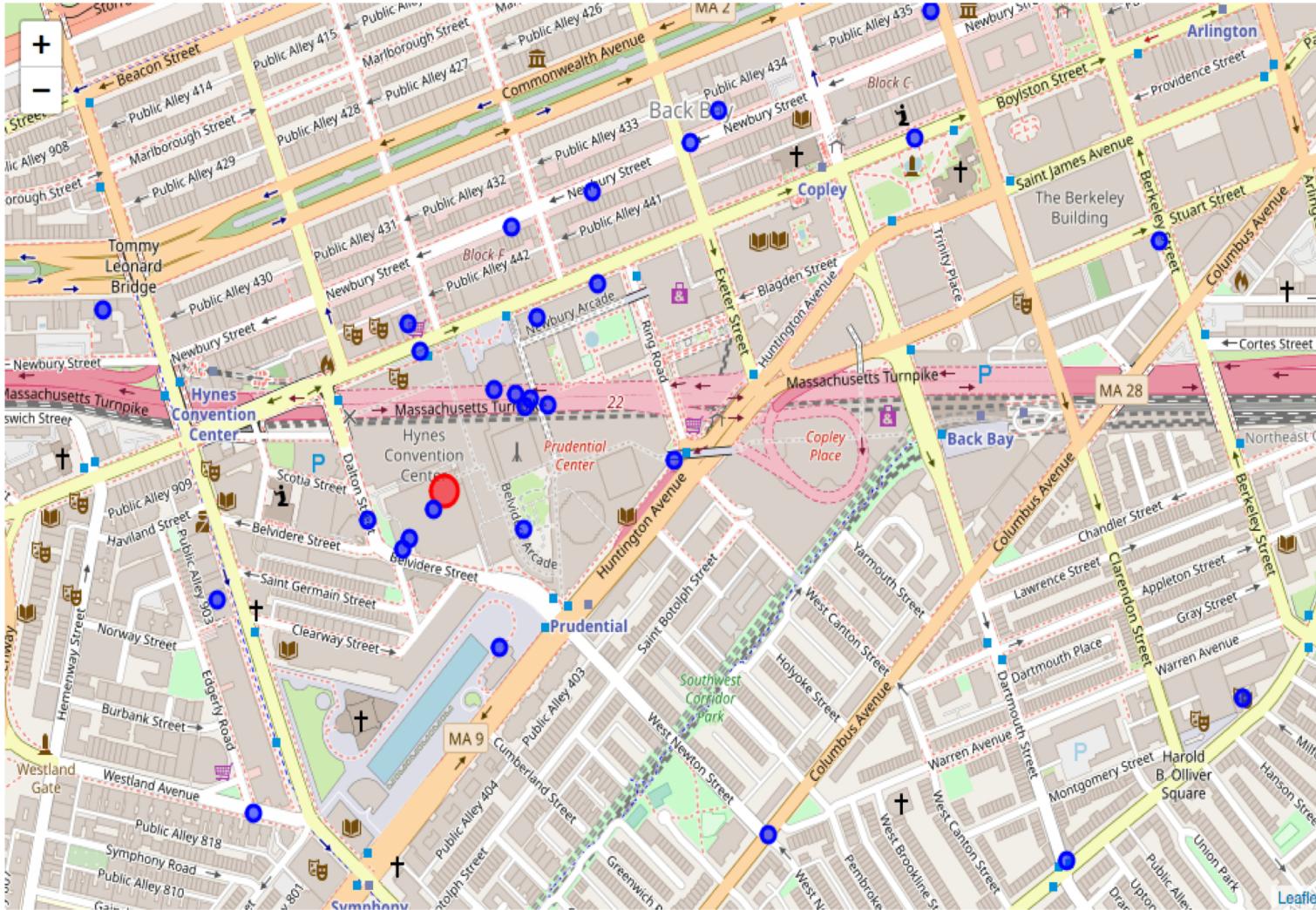
Reference point 2:

Boston Marriott Hotel at the Long Wharf  
296 State Street, Boston, MA 02109



## 5. Results

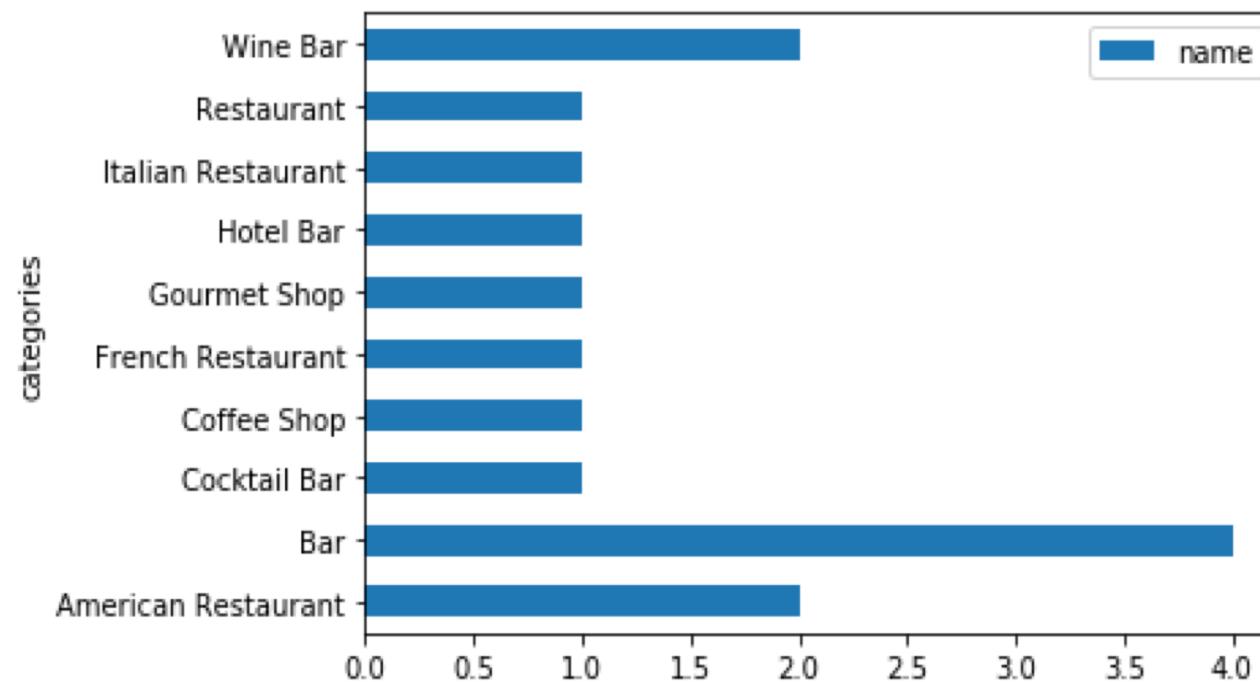
Reference point: Sheraton Boston at Back Bay – Venues map



## 5. Results

---

Reference point: Sheraton Boston at Back Bay – Venue categories



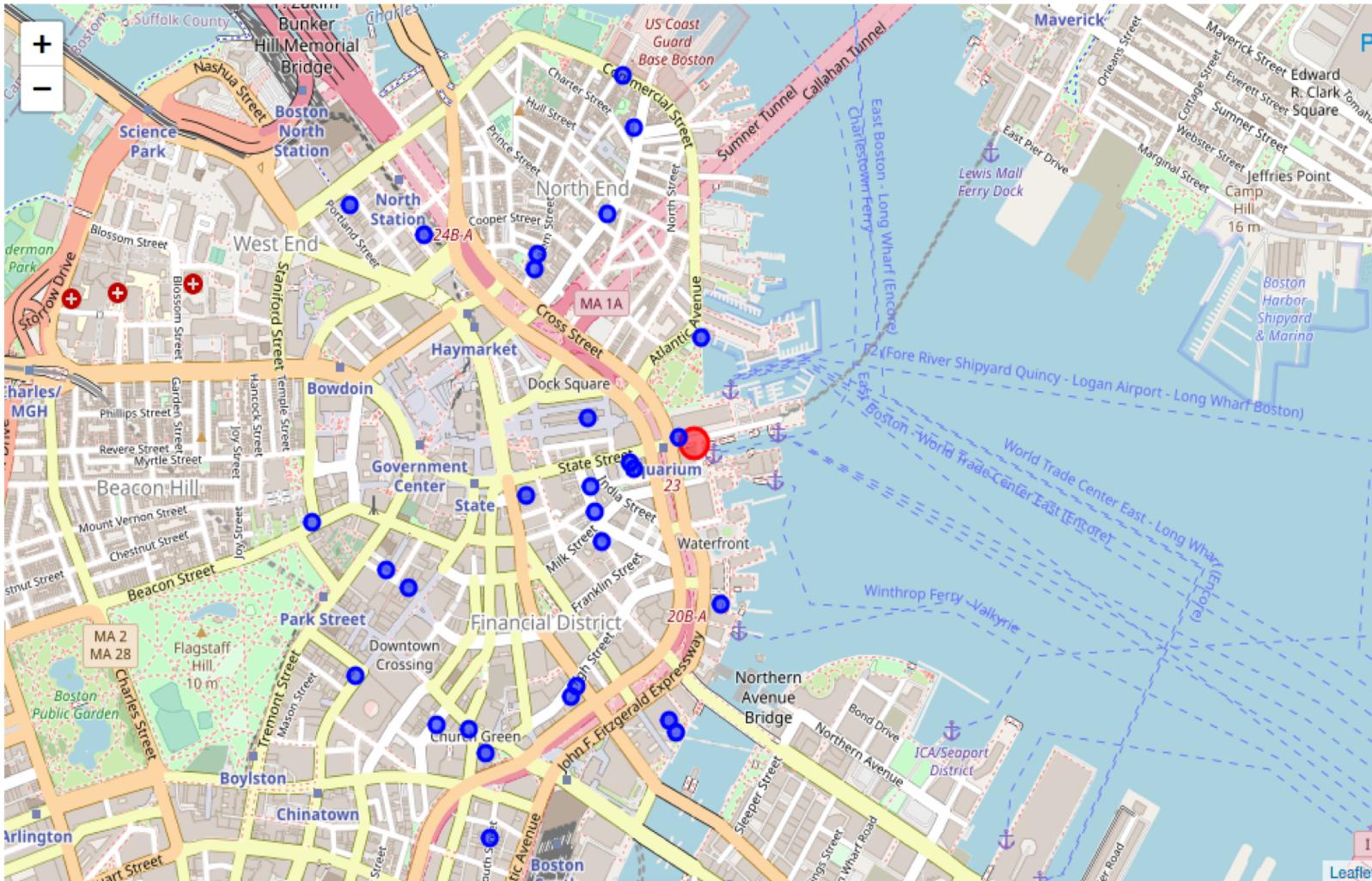
## 5. Results

Reference point: Sheraton Boston at Back Bay – List of filtered venues

	name	categories	address	distance	lat	lng
0	Barcelona Wine Bar	Wine Bar	525 Tremont St	1075	42.344622	-71.070919
1	Jaho Coffee Roaster & Wine Bar	Coffee Shop	116 Huntington Ave	303	42.347077	-71.079997
2	Piattini Wine Cafe	Italian Restaurant	226-228 Newbury St	394	42.349857	-71.081315
3	Side Bar Grille	Restaurant	39 Dalton St	26	42.346573	-71.083847
4	Forty Dalton Bar	Hotel Bar	40 Dalton St	106	42.346466	-71.084892
5	Pour House Bar & Grill	Bar	907 Boylston St	162	42.348197	-71.084051
6	SideBar & Grille	Bar	39 Dalton St	72	42.346269	-71.084227
7	Sheraton Shame Bar	Cocktail Bar		NaN	42.346166	-71.084326
8	Earls Kitchen + Bar	American Restaurant	800 Boylston St	167	42.347656	-71.082018
9	Eataly Boston	Gourmet Shop	800 Boylston St	153	42.347707	-71.082297
10	The Great Bitter Bar	Bar	800 Boylston St	143	42.347641	-71.082390
11	Bar Boulud	French Restaurant	776 Boylston St	310	42.348895	-71.081231
12	Cello Bar	Wine Bar		NaN	42.347759	-71.082525
13	McCarthy's Bar and Grill	Bar	903 Boylston St	198	42.348495	-71.084251
14	Precinct Kitchen & Bar	American Restaurant	154 Berkeley St	982	42.349350	-71.072251

## 5. Results

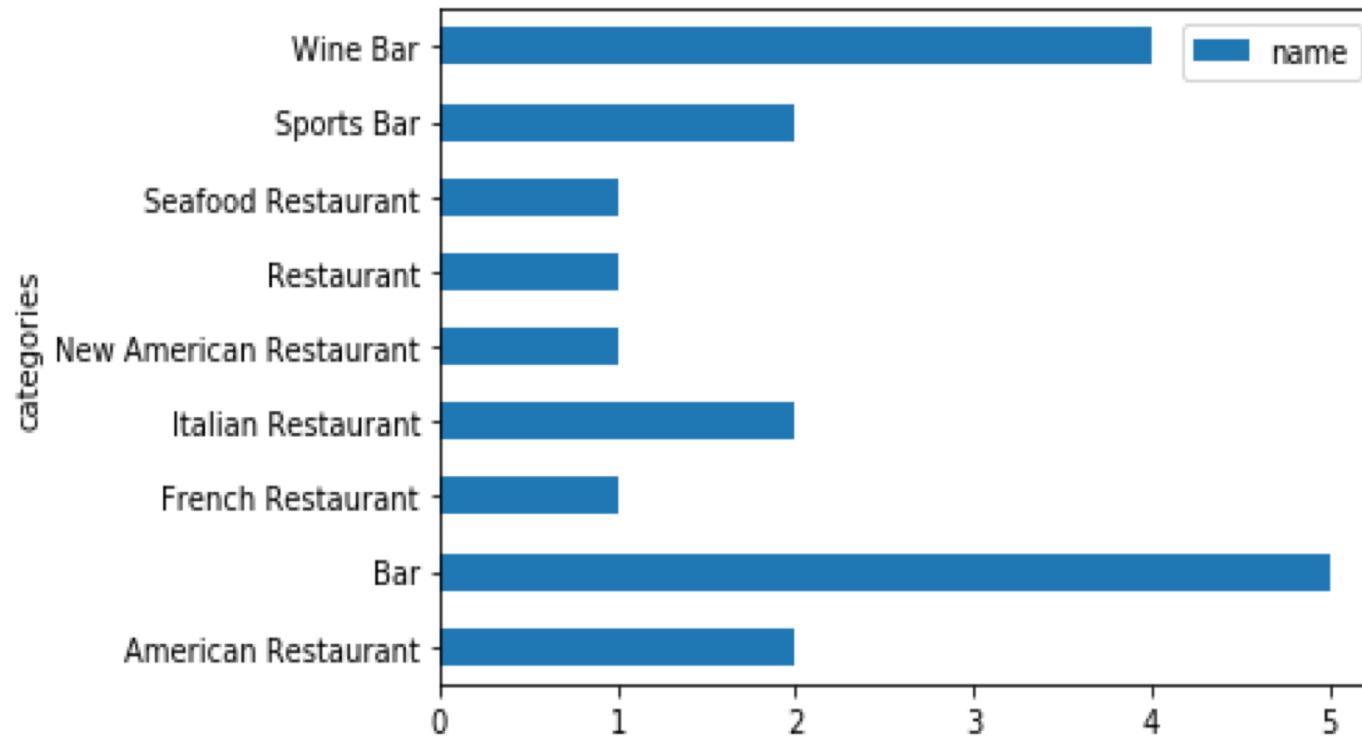
Reference point: Boston Marriott Long Wharf at the Waterfront – Venues map



## 5. Results

---

Reference point: Boston Marriott Long Wharf at the Waterfront – Venue categories



## 5. Results

Reference point: Boston Marriott Long Wharf at the Waterfront – List of filtered venues

	name	categories	address	distance	lat	lng	
0	Les Zygomatics Wine Bar Bistro	Wine Bar	129 South St	1067	42.351254	-71.057344	
1	Taste Wine Bar & Cafe	Wine Bar		Nan	882	42.353578	-71.057971
2	Sarah's Rooftop Wine Bar	Wine Bar	414 Hanover St	762	42.366367	-71.052933	
3	The Wine Bike Pop Up Wine Bar	Wine Bar	125 Summer St	904	42.353051	-71.057487	
4	Custom House Bar	Bar		Nan	40	42.359791	-71.051529
5	Joe's American Bar And Grill	American Restaurant	100 Atlantic Ave	252	42.361916	-71.050861	
6	Warehouse Bar & Grill	American Restaurant	40 Broad Street	297	42.358193	-71.054111	
7	Miel Brasserie Provençale	French Restaurant	510 Atlantic Ave	684	42.353517	-71.051640	
8	Kingston Grille & Bar (KGB Bar)	New American Restaurant	25 Kingston St	930	42.353659	-71.058965	
9	Wet Bar	Bar	183 State St	167	42.359257	-71.053040	
10	Rowes Wharf Bar	Bar	70 Rowes Wharf	387	42.356224	-71.050255	
11	Blake's Kitchen + Bar	Restaurant	276 Friend St	1034	42.364723	-71.061610	
12	Dockside Sports Bar	Sports Bar	183 State St	162	42.359113	-71.052911	
13	Rocco's Cucina & Bar	Sports Bar	450 Commercial St	889	42.367481	-71.053261	
14	Oyster Bar	Seafood Restaurant	1 Faneuil Hall Market Pl	275	42.360182	-71.054343	
15	Pulcinella Mozzarella Bar and Restaurant	Italian Restaurant	78 Salem St	598	42.363693	-71.055872	
16	Rabia's Seafood/Oyster Bar & Italian Restaurant	Italian Restaurant	73 Salem St	578	42.363377	-71.055977	
17	Side Bar Food & Spirits	Bar	14 Bromfield St	798	42.356574	-71.059848	
18	RumBa Rum Bar & Champagne Lounge	Bar	510 Atlantic Ave	658	42.353766	-71.051834	

## 5. Discussion Section

---

- The factors that will influence the decision based on the definition of the original problem are:
  - No significant number of direct or indirect competition, ideally no similar type of venue
  - Proximity to upscale city landmarks
- The resulting dataframes and the mapping of the locations in the preferred investor's neighborhoods, Back Bay and Waterfront, present no head-to-head competition for the kind of venue the investors have in mind. There are however other traditional business establishments like restaurants and bars in these areas that should be taken into account when selecting the actual street location for the new venue.
- One positive sign is that the chosen areas are not overcrowded with these existing businesses. It can be considered a low restaurant density area making it easier to find a spot at a reasonable distance from these potential competitors.

## 7. Conclusion

---

- The conclusion based on this preliminary analysis is that there is room for the Bistro & Wine Bar in the two selected areas, without colliding with any existing similar business. Each area has its own attractiveness with luxury shops and upscale entertainment and hospitality venues.
- The actual location short list will be decided based on other marketing considerations and the cost of the commercial real estate. These considerations are out of the scope of this present analysis.