### **Introduction-**

### **Description & Discussion of the Background**

Houston is the fourth most populous city in the United States with a median age of 31. The demographic offers an exciting opportunity for entrepreneurs to open a Bar in the city. However, owning and operating a bar can be a daunting task. Several factors must be considered, such as foot traffic, property prices, etc.

The most important factor in opening a bar is the location of the bar. A residential area with many schools, for instance, would not be a suitable location and might lead to problems such as noise complaints and issues obtaining a liquor license. It is important to pick a location where other bars are operating successfully.

Another important decision is the type of bar. There are several categories of bars such as breweries, pubs, etc. and each kind of bar has unique requirements. For instance, a Sports Bar would require additional investment in quality screens and audio systems for patrons to watch events.

In this project, we will try to find an ideal location for a Bar. We will also provide recommendations for a category of the bar that would be most profitable in the area. The target audience is an investor looking to open a bar in Houston.

### **Data and Solving problem:**

The Data that will be used to solve this problem is outlined below:

* I obtained a list of zip codes in Houston scrapped from the website – <https://www.zip-codes.com/city/tx-houston.asp> for analysis.
* I used FourSquare API and provided each zip code as input to fetch all the different bars within a 5000 meters radius.
* Using the category Id sent as part of the FourSquare response, I provide a breakdown of the kinds of bars in the area.

### **Methodology**

I used the GitHub repository for this project. The primary data frame has information such as zip code, county, population, etc. for which I have used – <https://www.zip-codes.com/city/tx-houston.asp> to pull the basic zip code information.

-- Screen of zip code data frame

A screenshot of a cell phone

Description automatically generated

This is used as input to FourSquare API to find a list of all Bars in each zip code within a 5000-meter radius. The request is made to fetch venues of category ‘Bar’, which includes a set of subcategories such as Sports Bar, Lounge, etc.

### **Sample FourSquare data looks like:**

-- FourSquare Raw Data Below in DataFrame

A screenshot of a cell phone

Description automatically generated

-- FourSquare data below after cleaning up removing “NaN” from DataFrame

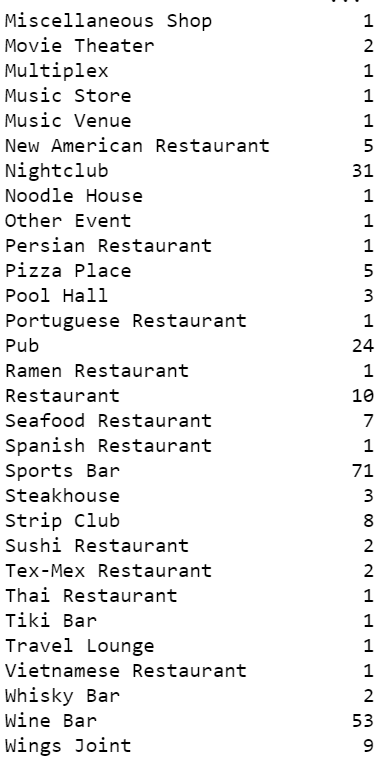
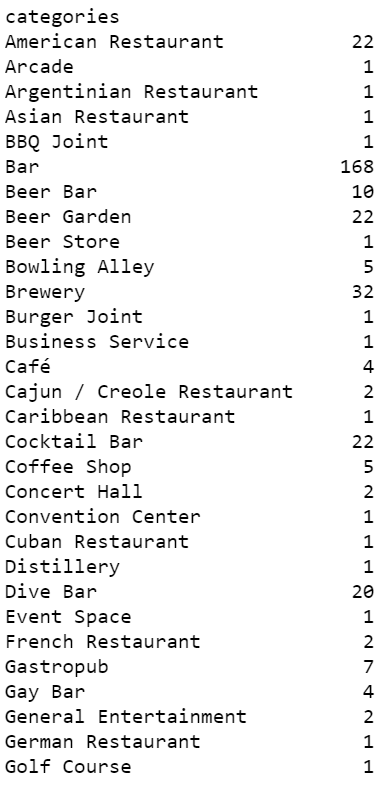
I have cleaned the data by removing the “NaN” where the address and zip codes are not pulled from FourSquare API.

A screenshot of a cell phone

Description automatically generated

Furthermore, I have created a DataFrame to categorized the type of bars and count for the same to get an overview; there are 73 different types/categories of bars in the Dataframe:

--Category List and Total Number of Categories



I used Python Folium library to visualize geographic details of the different bars and I generated a map of Houston with the bar information superimposed on its top. The latitude, longitude information is used for visualization as below:

-- Map Screenshot

A picture containing text, map

Description automatically generated

The FourSquare API that I used to pull the on the longitude and latitude is using **5000 meters** radius to pull bar data that is in Houston limit or is under Harris county. \*Data marked on the above map is clean with no duplication.

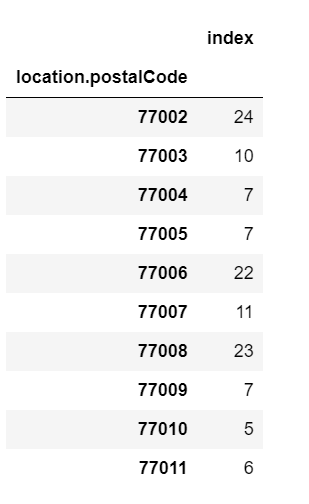
Below is the sample size from the population DataFrame and Bar Chart that will illustrate the total number of bars in each Zip code:

--Total Bars in Zip code DataFrame

A picture containing object

Description automatically generated

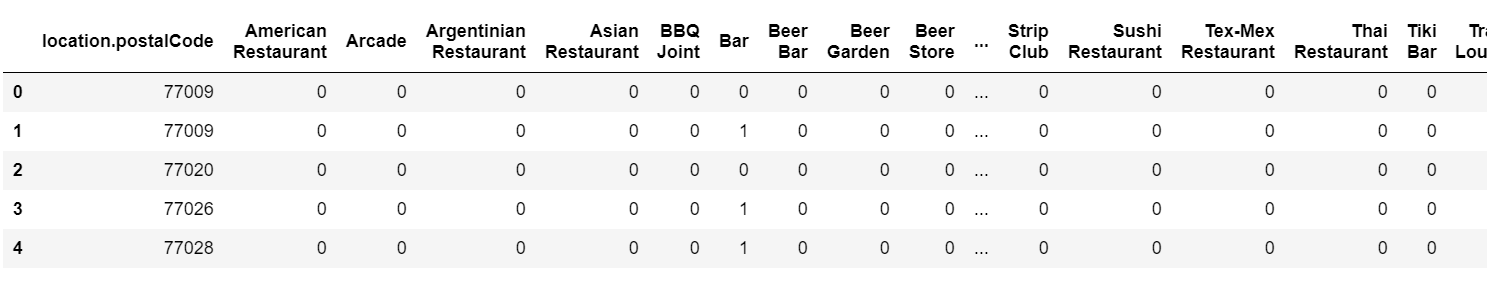
--Sample DataFrame



For further analysis, I will analyze the categories of bars in these locations. I have u-sed two methods to find that. First is the sum total and the other is a simple statistical method **Mean** which will give us the number of categories and categories by zip code.

--Breakdown of Bar Types

#1 Total Number by Category:



#2 Mean by Category and Groupby Zip code:

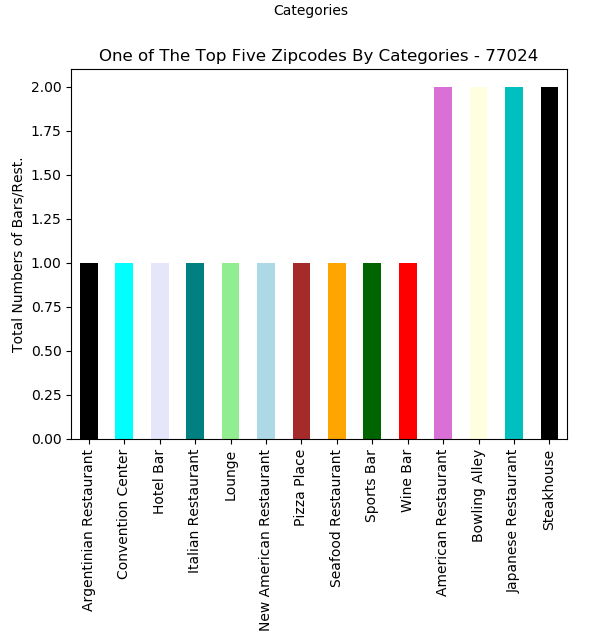
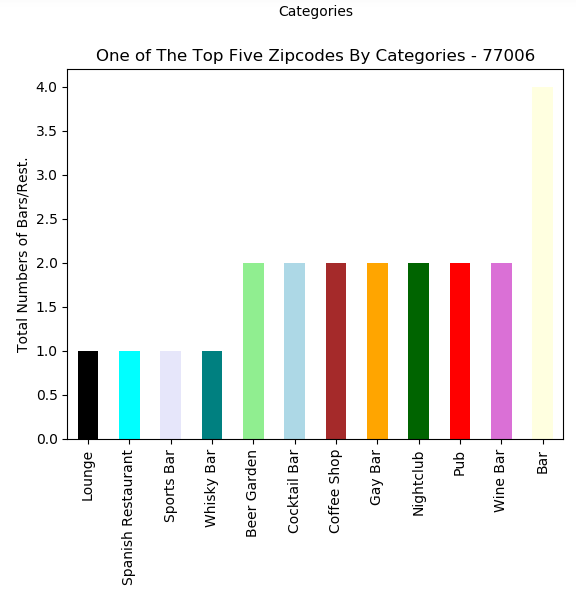
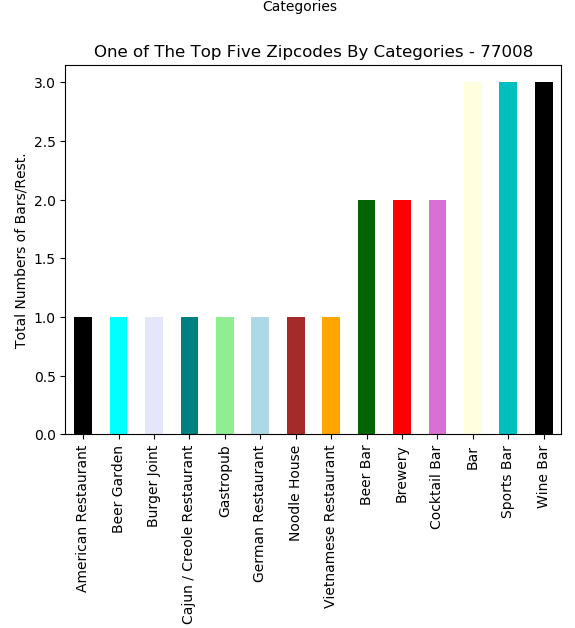
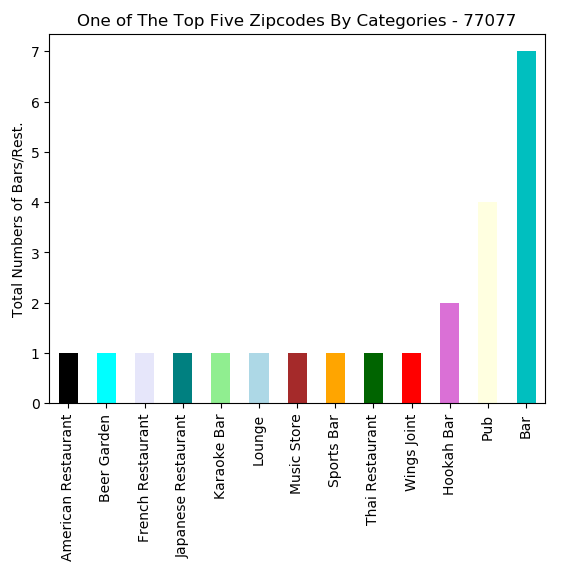
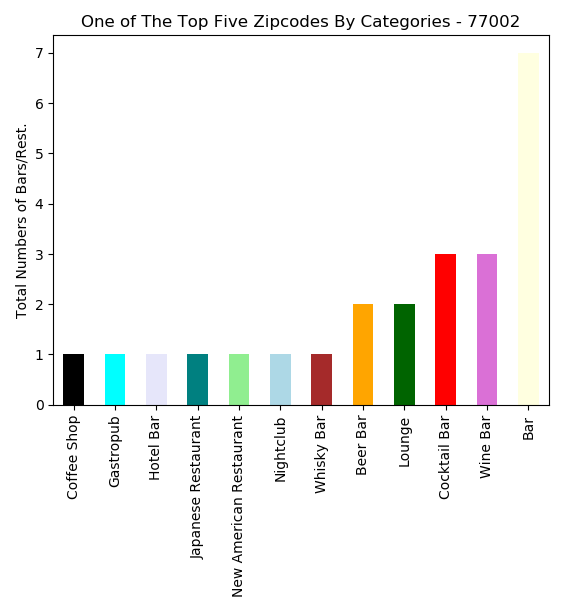
A close up of a piece of paper

Description automatically generated

For better analysis and presentation, I have converted the above DataFrame to show top 5 categories by Zip Code(\*\*Sample Size DataFrame).



After analyzing the data I have drilled down to the Top 5 zip codes and their categories on Bar graph:



### **Results:**

Using the Marker Cluster and the Folium method in maps I have representation of the bars in the topmost 5 zip codes is as follows:

A close up of a map

Description automatically generated

On top of it, I have also provided an option to drill down the DataFrame by user input which will give the Analyst/Developer/Scientist to provide the information on his/her hands just by entering the Zip code.

A screenshot of a cell phone

Description automatically generated

### **Discussion:**

From our analysis, we observe that the top 5 zip codes in Houston with the most number of bars are as follows: 77002', '77077', '77008', '77006', '77024'

Since these locations have a proven history of bars being operated, they provide the least amount of risk for the new investor.

Also, we can recommend that after a zip code location is selected, the type of bar to be opened can be selected from either the 3rd, 4th or 5th most common venue as shown in the graph below. For instance, in zip code 77007 there are fewer Lounge and Gastropub and that can be a niche market that can be targeted:

A screenshot of a cell phone

Description automatically generated

### **Conclusion:**

With a large amount of data at their disposal, entrepreneurs can make better-informed decisions. From our analysis we can see that Houston offers great opportunities to potential investors. It can also be explored further with consideration of population in Houston by zip codes/neighborhood.