Applied Data Science Capstone

by IBM (COURSERA)

Capstone Project - The Battle of Neighborhoods

SURVEY OF VENUES IN HONOLULU, HAWAII

Anubha Bhaik

24 April , 2020

1. INTRODUCTION:

For this capstone project, I have selected the capital of Hawaii that is Honolulu since it is home to a large number of tourists every year and I myself am very eager to visit these pretty islands of USA. Honolulu means "sheltered harbor" or "calm port" in Hawaiian It is popular for dining, nightlife and shopping, backed by palms and high – rise hotels with a beautiful skyline. It is the largest city in the middle of Pacific Ocean with a population of over one million people.

In 2017 alone, according to state government data, there were over 9.4 million visitors to the Hawaiian Islands with expenditures of over $16 billion. Tourism makes up 21% of the state's economy, with many of Hawaii's largest industries revolving around the constant flow of tourists. (From https://en.wikipedia.org/wiki/Tourism_in_Hawaii)

Since tourism is a major aspect of the city, many people visiting Honolulu want to know about the best places to dine in the city. People would want to know about the food and rating of restaurants around them so that they can easily decide where to eat. The FourSquare API along with the Zomato API will help people visiting Honolulu know about the places to eat.

1.1 BUSINESS PROBLEM

The objective of this Capstone Project is finding the most suitable place to eat when one visits Honolulu, USA. There are many venues (including restaurants, cafes) that can be explored. It helps us to find the best place to eat based on average price and user ratings. Using methods learnt over the course of this specialization such as clustering, this project aims to find the solution to the problem: Where to eat when I visit Honolulu.

1.2 TARGET AUDIENCE

This project is particularly useful for anyone who is visiting Honolulu for the first time. They can use the maps and plots illustrated clearly through the project to have a glance and decide quickly where to dine based on ratings and reviews from other users using Foursquare and Zomato. Also, other companies who want to provide services to these popular eating locations such as companies wanting to initiate fast delivery services can use this data to increase their service.

2. DATA

Data for this project is retrieved from two API's namely the Foursquare and Zomato. Venues located within a distance of a particular radius will be used.

Latitudes and longitudes of Honolulu can be taken directly from Wikipedia to plot the maps of the city with the suitable venues.

2.1 SOURCES OF DATA:

Wikipedia will be used to get the coordinates of Honolulu (such as the latitude and longitude). This can also be scrapped using the Geocoder package.

From Foursquare API (https://developers.zomato.com/api), I retrieved the following for each venue:

- **Name:** The name of the venue.
- **Category:** The category type as defined by the API.
- **Latitude:** The latitude value of the venue.
- **Longitude:** The longitude value of the venue.

From Zomato API(https://developers.zomato.com/api), I retrieved the following for each venue:

- **Name:** The name of the venue.
- **Address:** The complete address of the venue.
- **Rating:** The ratings as provided by many users.
- **Price range:** The price range the venue belongs to as defined by Zomato.
- **Price for two:** The average cost for two people dining at the place. I later convert the same to average price per person by dividing by 2.
- **Latitude:** The latitude value of the venue.
- **Longitude:** The longitude value of the venue.

```
Data for the following venue is being fetched: 1
Data for the following venue is being fetched: 2
Data for the following venue is being fetched: 3
Data for the following venue is being fetched: 4
Data for the following venue is being fetched: 5
Data for the following venue is being fetched: 6
Data for the following venue is being fetched: 7
Data for the following venue is being fetched: 8
Data for the following venue is being fetched: 9
Data for the following venue is being fetched: 10
Data for the following venue is being fetched: 11
Data for the following venue is being fetched: 12
Data for the following venue is being fetched: 13
Data for the following venue is being fetched: 14
Data for the following venue is being fetched: 15
```

Figure 1. Data being fetched from the FoursquareAPI

2.2 DATA CLEANING:

Out[12]:

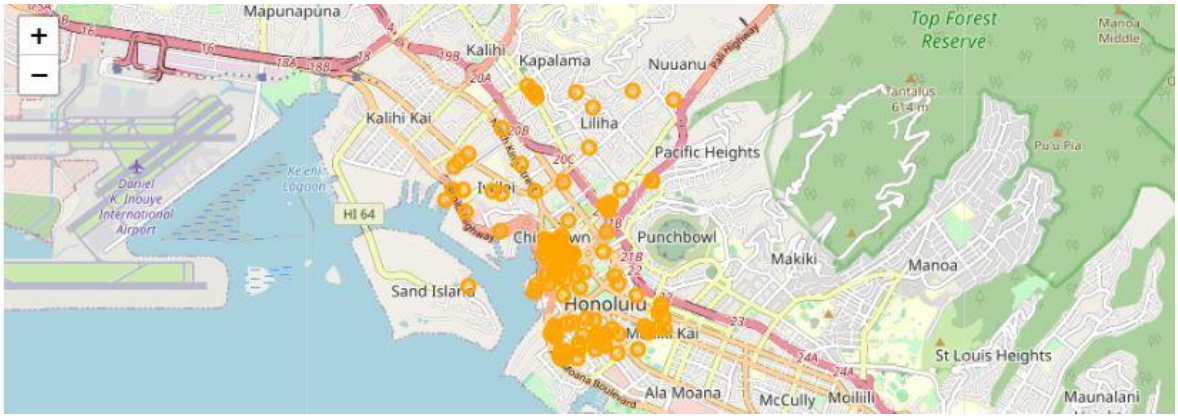

Figure 2 : Map of venues fetched by FourSquare API

Out[13]:

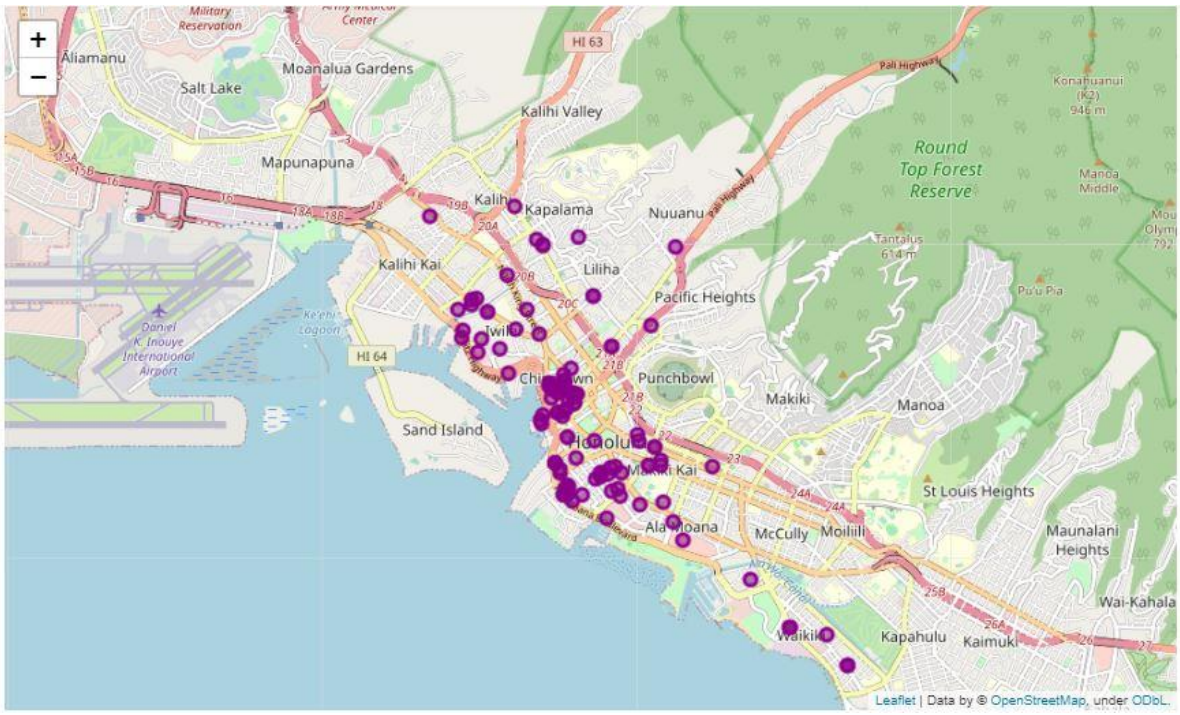

Figure 3: Map of venues fetched by Zomato API

From figure 2 and figure 3, we can clearly see that some venues from the two APIs do not align with each other. Thus they are combined using their latitude and longitude values.

After combining the data, the final dataset is obtained by combining the data obtained from both the APIs as shown in Figure 4

Out[42]:

	name	categories	lat	long	lng	venue	latitude	longitude	price_for_two	price_range	rating	address	lat_diff	lng_diff
0	Royal Kitchen	Chinese Restaurant	21.3149	NaN	-157.8606	Royal Kitchen	21.3142	-157.8609	10.0	1.0	3.6	100 N Beretania St, Honolulu 96817	-0.0007	-0.0003
1	JJ Dolan's	Pizza Place	21.3110	NaN	-157.8605	J.J. Dolan's	21.3110	-157.8605	25.0	2.0	4.1	1147 Bethel Street, Honolulu 96813	0.0000	0.0000
2	Char Hung Sut	Chinese Restaurant	21.3124	NaN	-157.8619	Char Hung Sut	21.3125	-157.8618	10.0	1.0	3.8	64 N Pauahi St, Honolulu 96817	0.0001	0.0001
3	Manifest	Lounge	21.3115	NaN	-157.8623	The Manifest	21.3116	-157.8621	10.0	1.0	3.1	32 North Hotel Street, Honolulu 96817	0.0001	0.0002
4	Livestock Tavern	American Restaurant	21.3117	NaN	-157.8626	Livestock Tavern	21.3115	-157.8626	0.0	1.0	0	49 N Hotel Street, Honolulu 96817	-0.0002	0.0000

Figure 4 : Final Dataset

3. METHODOLOGY:

The location data is extracted from the Foursquare API for all venues up to a distance of 2km. from the center of Honolulu. Using this, venue information including price and rating data from Zomato API is fetched.

The dataset from the two APIs will be combined based on the venue names, latitude, and longitude values. The final data will include the venue name, category, address, latitude, longitude, rating, price range, and average cost per person.

The venues on the map using the final dataset are now explored. This will allow us to better understand the location of various venues and the places where many venues co-exist. Venues are also explored based on the ratings and price range of various venues. The venues will be plot using proper color coding such that a simple glance at the map would reveal the location of the venues as well as give information about them. These venues will then be clustered to draw meaningful information out of what kind of venues exist in Honolulu.

As a final step, these plots will be analyzed and conclusions will be drawn on what places can be recommended to visitors.

The analysis is broken down into various factors such as categories of venues in Honolulu, rating and price.

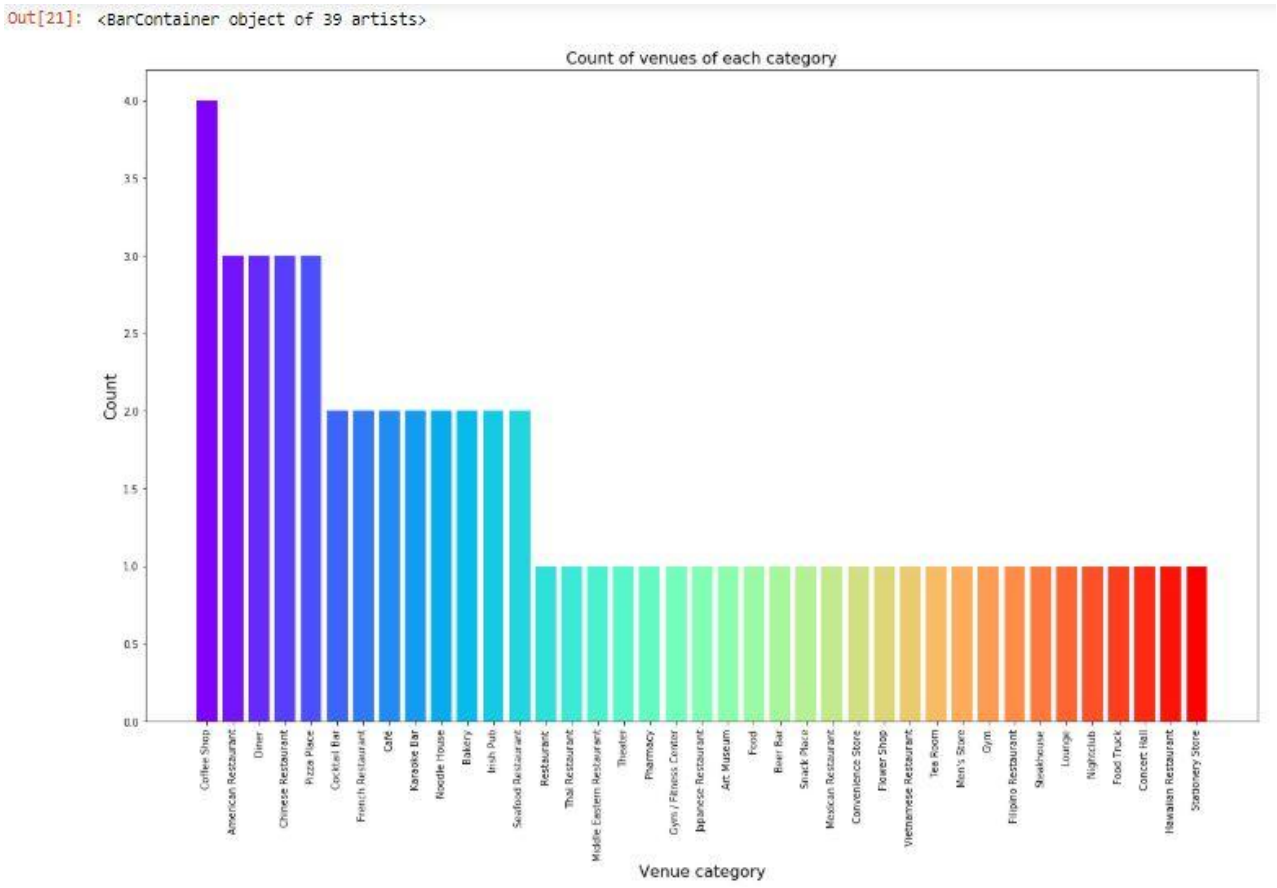

Figure 5: Count of various types of venues in Honolulu

From figure 4, we see that the majority venues are Coffee Shops. This is followed by American Restaurants, Diner, Chinese Restaurants and Pizza Places. Thus people who are fond of coffee and coffee shops will surely find a lot of places to go to.

The ratings of various venues in Honolulu are now explored. A bar chart with x-axis as the rating from 1 to 5 and the y-axis as the count of venues with that rating is now plotted to clearly visualize the results.

While the whole range of rating of venues might stretch from 0 to 5, the average rating is spread across 3 with maximum number of venues scoring a rating of around 3.

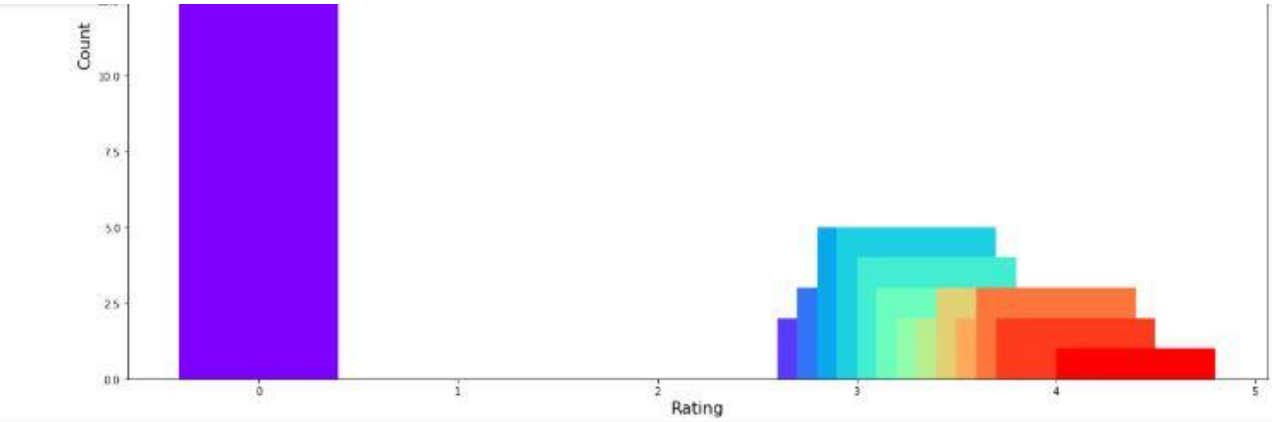

Figure 6: Rating and count of venues with that rating

Next, the average prices of all venues are explored for one person using a scatter plot along with the count of venues with that average price per person.

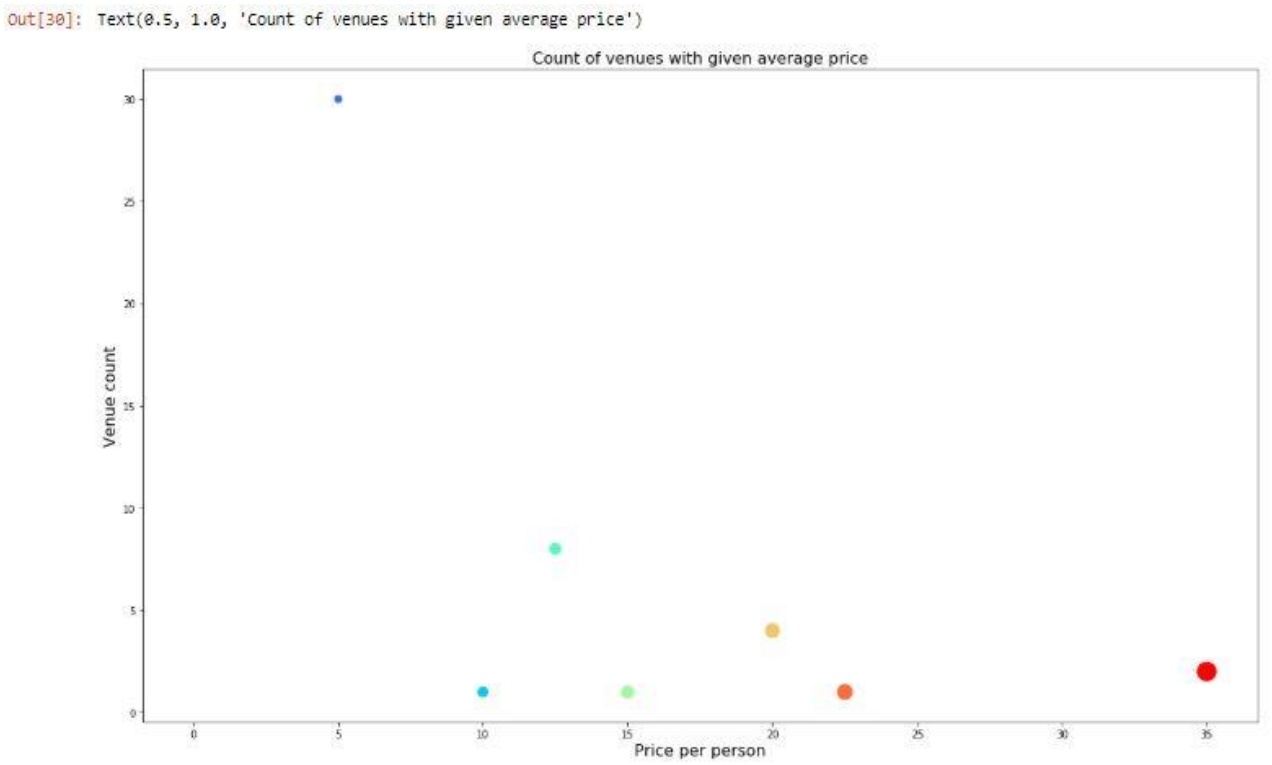

Figure 7: Price per person with count of venues with that price

Clustering:

Finally, all the venues are clustered based on their price range, location and more to identify similar venues and the relationship amongst them. I used KMeans clustering and decided to cluster the venues into two separate groups.

Out[33]:

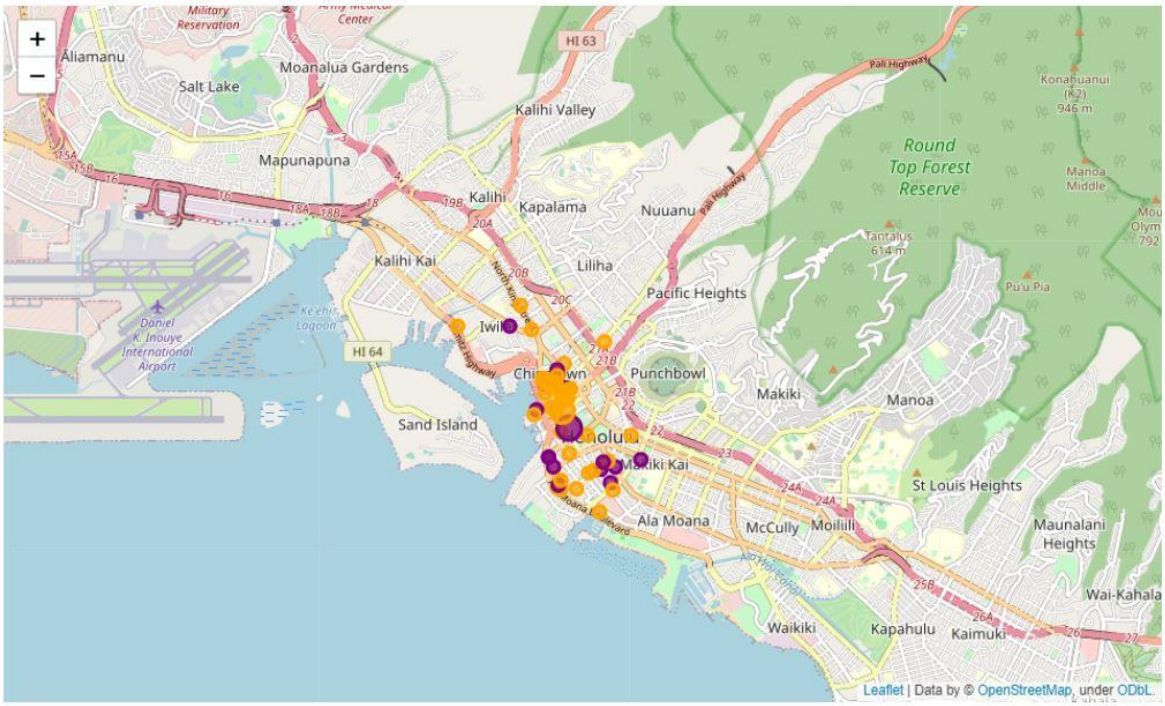

Figure 8: Clusters of venues

In figure 8, we see the two clusters:

1. The first cluster (orange) is spread across the whole city and includes the majority venues. These venues have mean price range of 2.62 and rating spread around 2.96.
2. The second cluster (dark magenta) is very sparsely spread and has very limited venues. These venues have mean price range of 1.02 and rating spread around 1.98.

4 . RESULTS and DISCUSSIONS:

After collecting data from the Foursquare and Zomato APIs, and combining the data, we could analyze it to find meaningful conclusions.

We identified that from the total set of venues, majority of them were Coffee shops. A visitor who loves Coffee Shops would surely find a number of places to visit in Honolulu.

Majority venues have ratings in the range from 3-4. This means that most restaurants provide good quality food which is liked by the people of the city, thus indicating the high rating. When we plot these venues on the map, we discover that there are clusters of venues around "The Arts District" , "Kaka'ako" and "Makiki Kai".

1. If you're looking for cheap places with relatively high rating, you should visit Chinese Restaurants like the 'Royal Kitchen'.
2. If you're looking for the best places, with the highest rating but might also carry a high price tag, you should visit Pizza Places like 'JJ Dolan's'.

5. CONCLUSION

The purpose of this project was to explore the places that a tourist visiting Honolulu could explore. The venues have been identified using Foursquare and Zomato API and have been plotted on the map. The map reveals that there are three major areas a person can visit: Sector **"The Arts District", "Kaka'ako"** and **"Makiki Kai"**. Based on the visitor's venue rating and price preferences, he/she can choose amongst the three places.