Exploring venues in Houston, TX, USA leveraging Foursquare API and Zomato API

by Deep Jayantibhai Patel

Finding the best Dining place is important when you are at a new place

- Houston is spread over 669 square miles. Being one of the largest tourist attractions due to Houston Museum of Natural Science, NASA's Space Centre and Horse Riding, visitors spend a lot of money in food too.

- **Each new visitor is naturally expected to have a confusion “What to eat?” and “Where to eat?” visiting a new place.**

- Thus, this project explores various venues with the attributes like location, rating and even price for two to have a meal in a particular restaurant.

Combining the location of the venues in the city with their price and rating information would surely help visitors in a city make better informed decisions about the places they should visit according to their cuisine preferences.

Data Preparation

Step 1: Using geolocator, coordinates for the location of interest is fetched in Latitude and longitude values.

Step2: Feeding the coordinates to the Foursquare's Explore API, details about the nearby recommended venues within a specified radius range i.e. 5 miles are fetched as shown below (Total 240 venues):

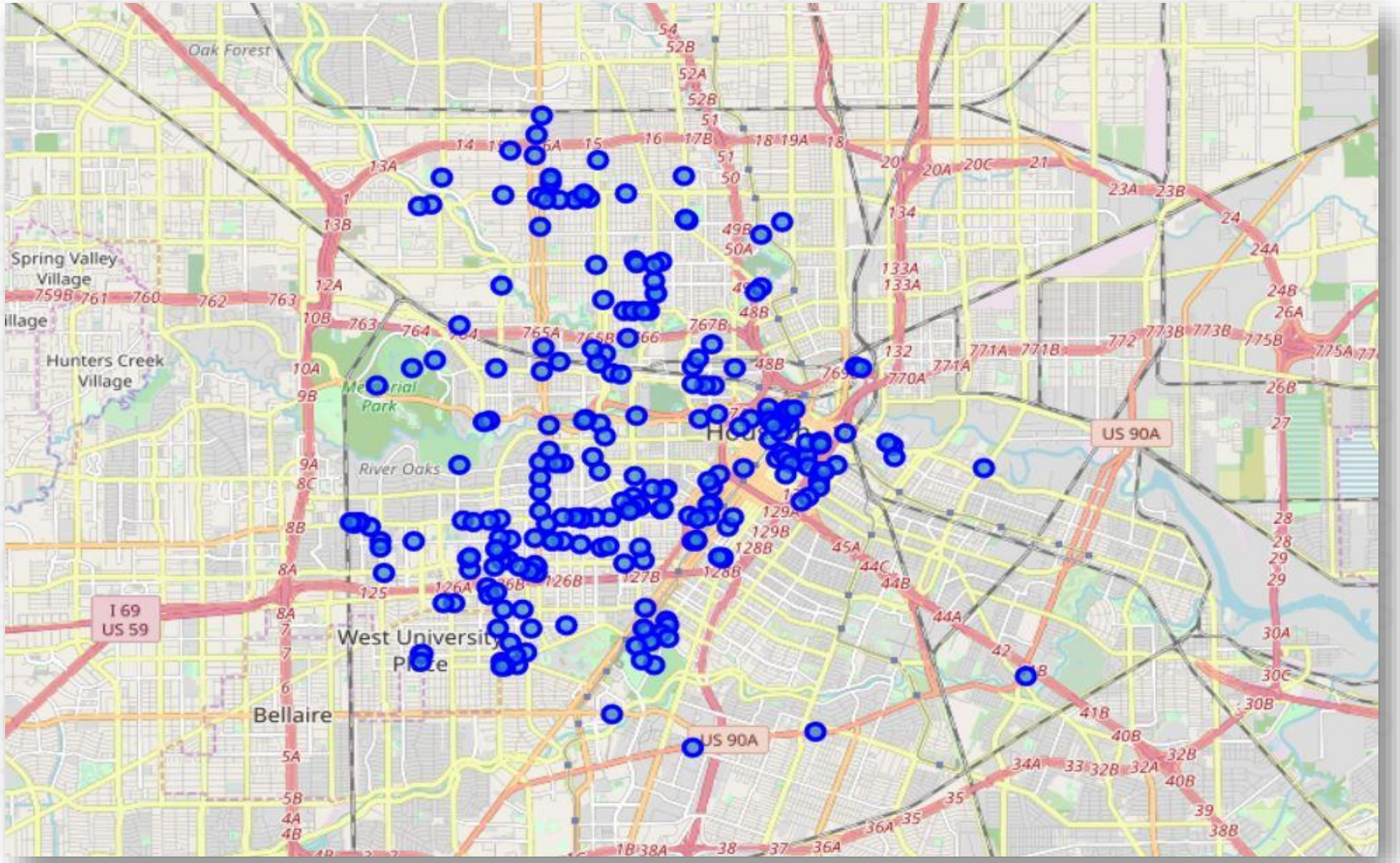
```
foursquare_venues
```

```
(240, 5)
```

Out[8]:

	name	categories	lat	lng	venue
0	Hobby Center for the Performing Arts	Performing Arts Venue	29.761526	-95.369376	Hobby Center for the Performing Arts
1	Alley Theatre	Theater	29.761671	-95.365313	Alley Theatre
2	Wortham Theater Center	Theater	29.763353	-95.365663	Wortham Theater Center
3	Jason's Deli	Deli / Bodega	29.757464	-95.365543	Jason's Deli
4	House of Blues	Music Venue	29.753836	-95.363932	House of Blues
5	Conservatory	Beer Garden	29.760427	-95.361570	Conservatory
6	Pappas Bros. Steakhouse	Steakhouse	29.755270	-95.363040	Pappas Bros. Steakhouse
7	Sam Houston Park	Park	29.759876	-95.371211	Sam Houston Park
8	Flying Saucer Draught Emporium	Beer Bar	29.759116	-95.363216	Flying Saucer Draught Emporium
9	Buffalo Bayou Walk	Trail	29.762177	-95.375844	Buffalo Bayou Walk

Step 3: Map of the venues from Foursquare API:



Step 4: For the fetched venues in Second step, details about the ratings and price range are collected for each venue leveraging the Zomato API (Total 240 venues)

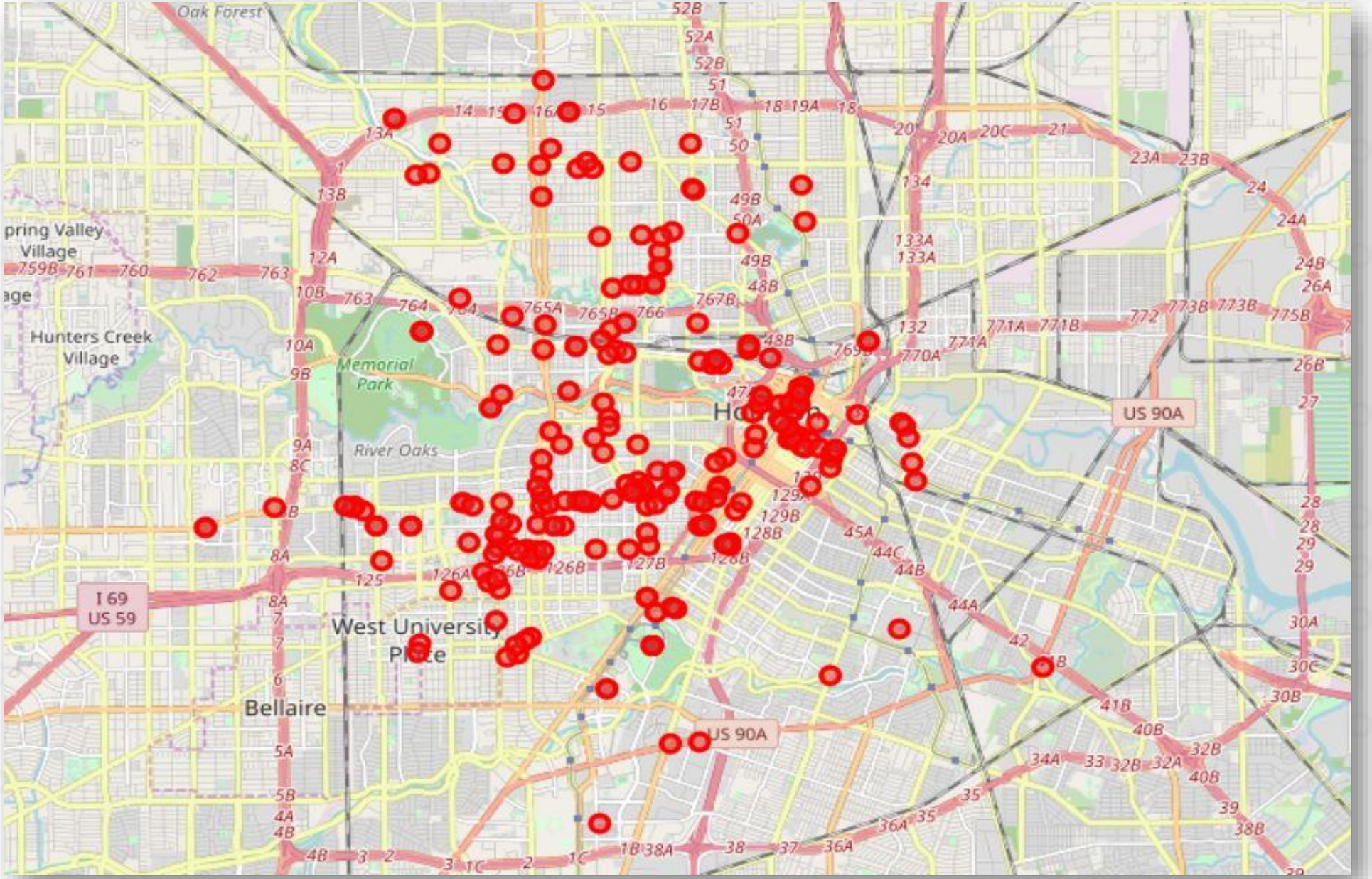
```
zomato_venues
```

```
(240, 7)
```

Out[12]:

	venue	latitude	longitude	price_for_two	price_range	rating	address
0	Artista	29.761770	-95.368790	90.0	4.0	3.8	800 Bagby Street, Suite 400, Hobby Center of t...
1	0	0.000000	0.000000	0.0	0.0	0	None
2	Artista	29.761770	-95.368790	90.0	4.0	3.8	800 Bagby Street, Suite 400, Hobby Center of t...
3	Jason's Deli	29.757541	-95.365495	30.0	3.0	0	901 McKinney Street, Houston 77002
4	House of Blues Restaurant & Bar	29.753910	-95.363580	70.0	4.0	2.8	1204 Caroline Street 77002
5	Conservatory Underground Beer Garden & Food Hall	29.760620	-95.361910	0.0	1.0	3.2	1010 Prairie Street 77002
6	Pappas Bros. Steakhouse	29.755270	-95.363040	0.0	1.0	3.3	1200 McKinney Street 77010
7	Skyline Houston Deli	29.758716	-95.370620	25.0	2.0	0	1111 Bagby St, Houston 77002
8	Flying Saucer Draught Emporium	29.758893	-95.363106	50.0	4.0	4.4	705 Main Street, Suite A, Houston 77002

Step 5: Map of the venues from Zomato API:



Data Cleaning

Step 1: Filter the venues to keep only restaurants and food places
In the above dataset fetched using Zomato API, for the venues in the FOURSQUARE API dataset which were NOT restaurants or food places, all the values are set to 0 or none. So, this step can be achieved by deleting the rows in Zomato Venues dataset having name = 0
(234 venues left now).

Step 2: Merge Foursquare venues and Zomato venues to create a final dataset
Both the datasets are merged as Inner Join having the same name of venues as the matching criteria (129 venues left now).

Step 3: Remove the fruitless venues i.e. Venues with no ratings, venues with no price details
(101 venues left now).

Step 4: Drop the columns that are common in both the API's result sets
Columns such as Latitude and Longitude were present in both the datasets. Thus, one of each is dropped (101 venues left now).

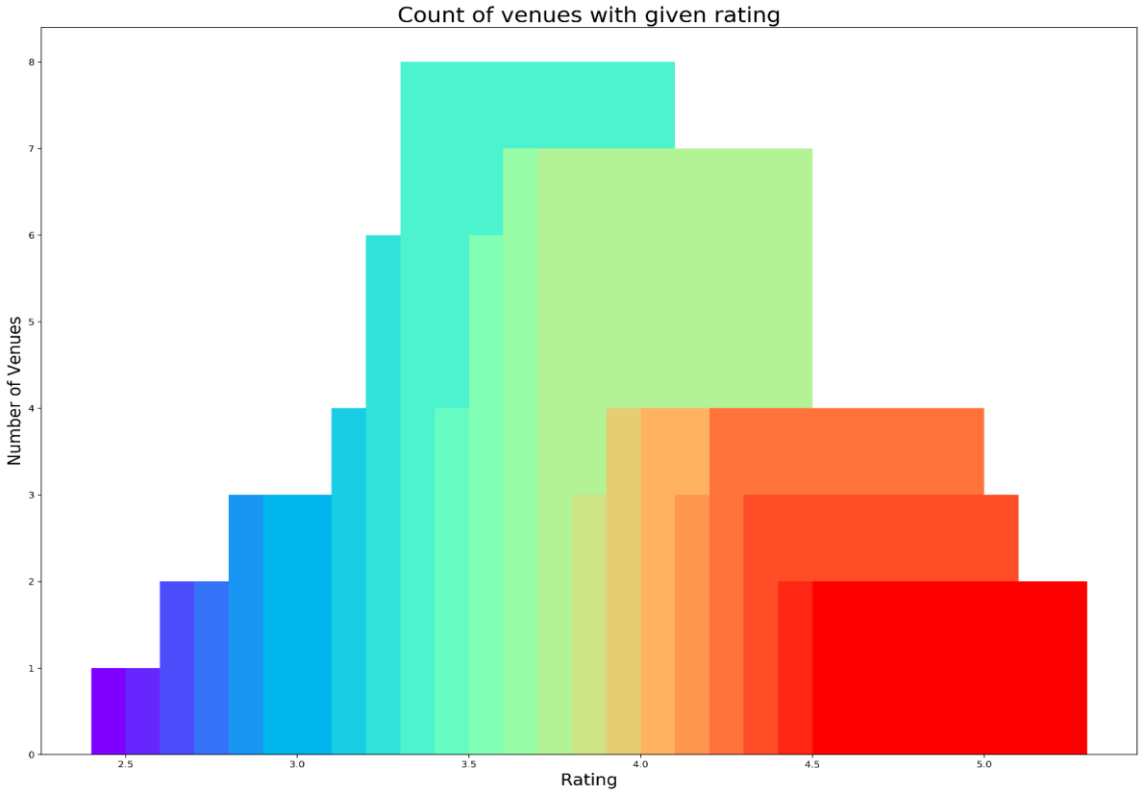
Step 5: Drop duplicates if any
(79 venues left now).

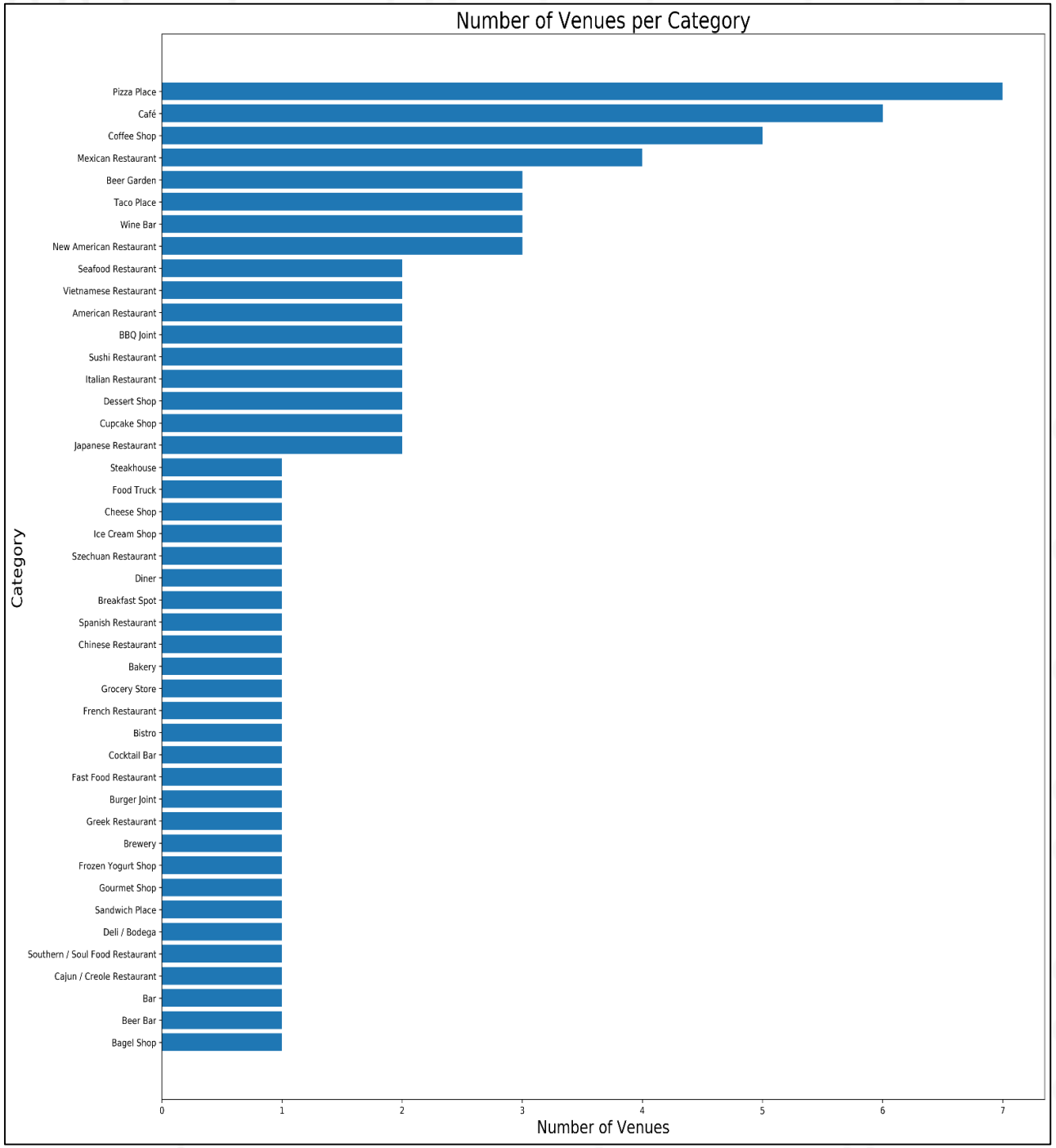
After performing all the above mentioned 5 data cleaning steps, final dataset was ready for further analysis having 79 venues with 8 features.

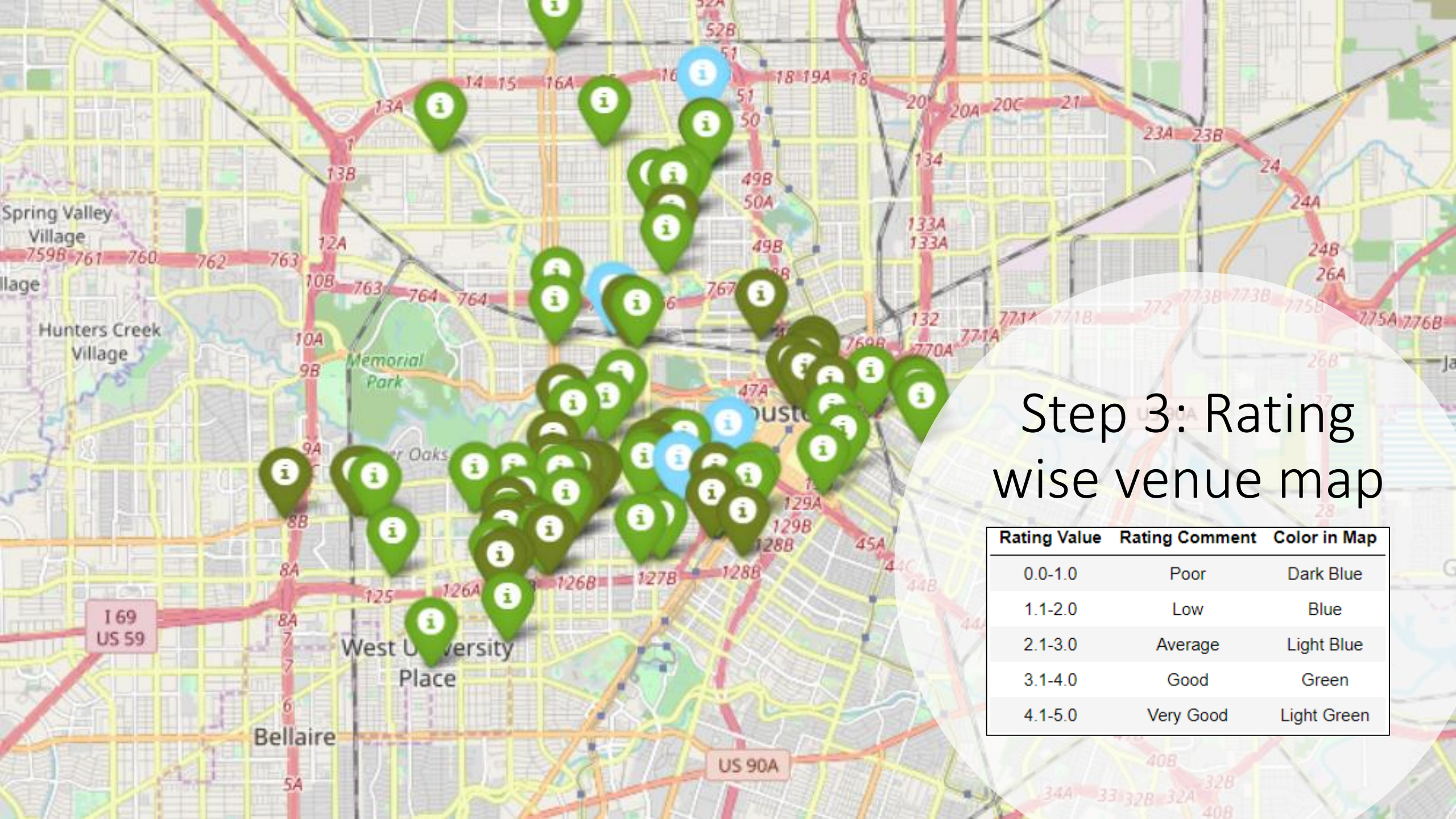
Data Analysis

Step 1: Venue's `Category` based analysis

Step 2: Venue's `Rating` based analysis



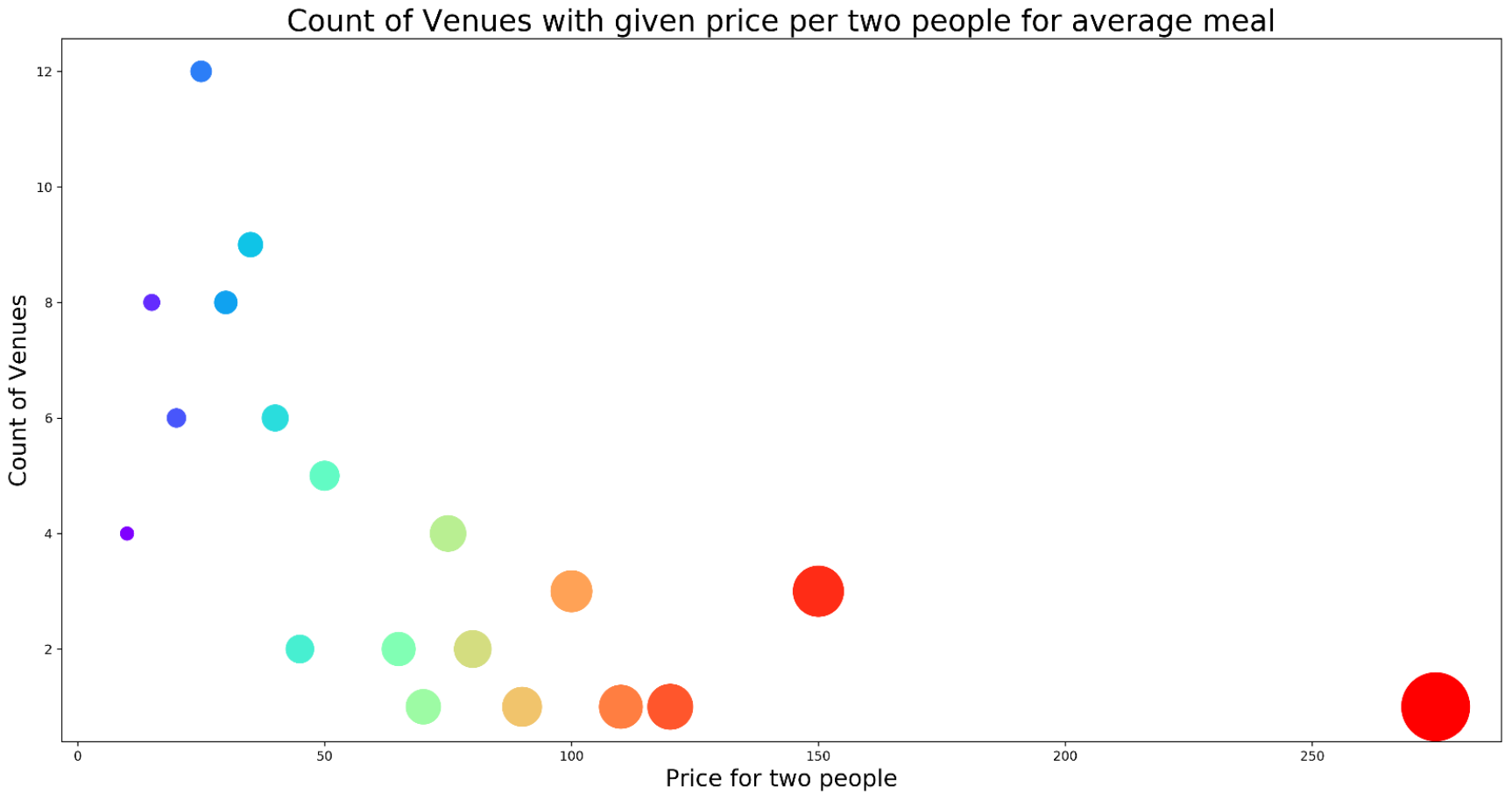


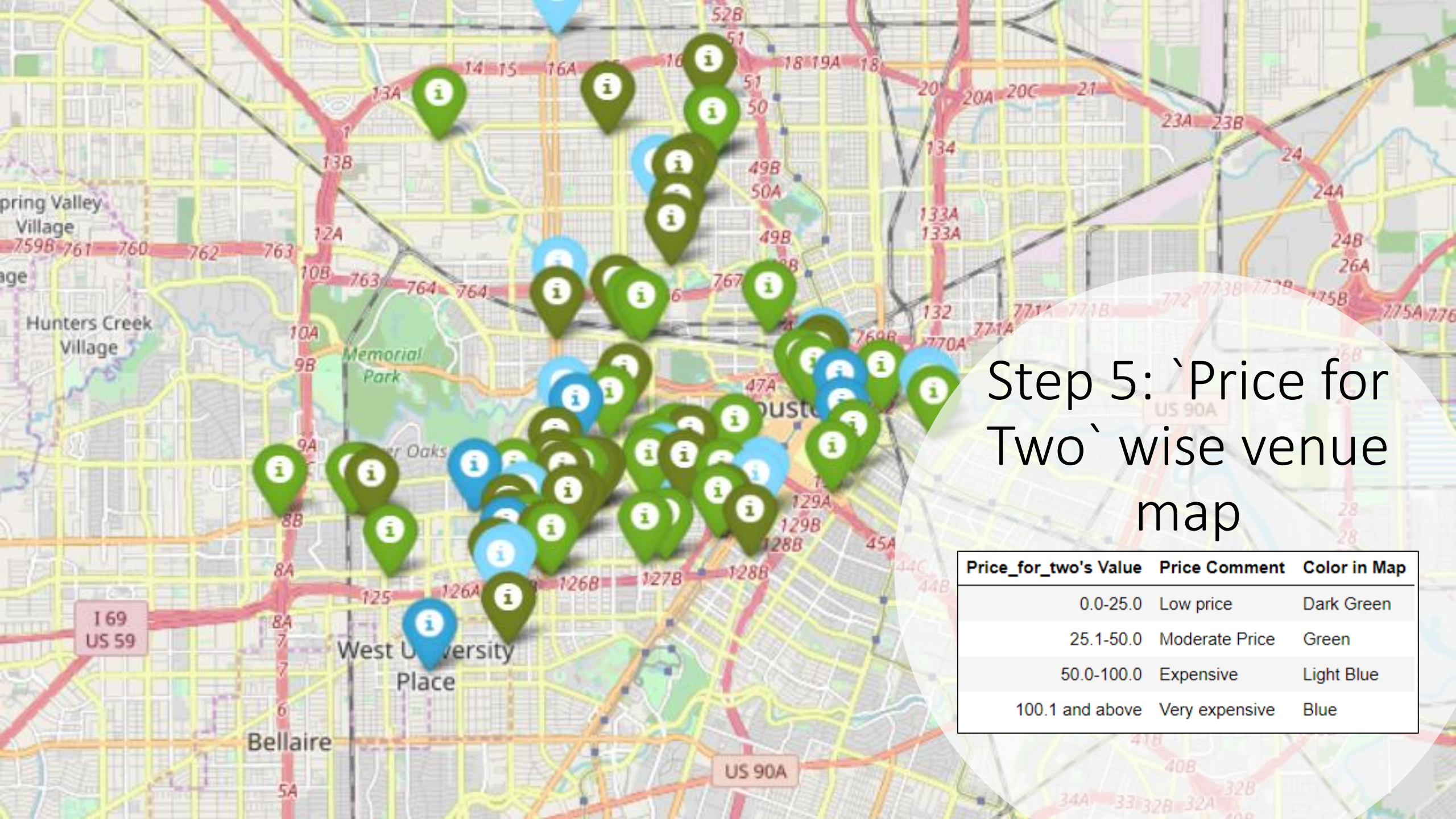


Step 3: Rating wise venue map

Rating Value	Rating Comment	Color in Map
0.0-1.0	Poor	Dark Blue
1.1-2.0	Low	Blue
2.1-3.0	Average	Light Blue
3.1-4.0	Good	Green
4.1-5.0	Very Good	Light Green

Step 4: Venue's `Price for Two` based analysis

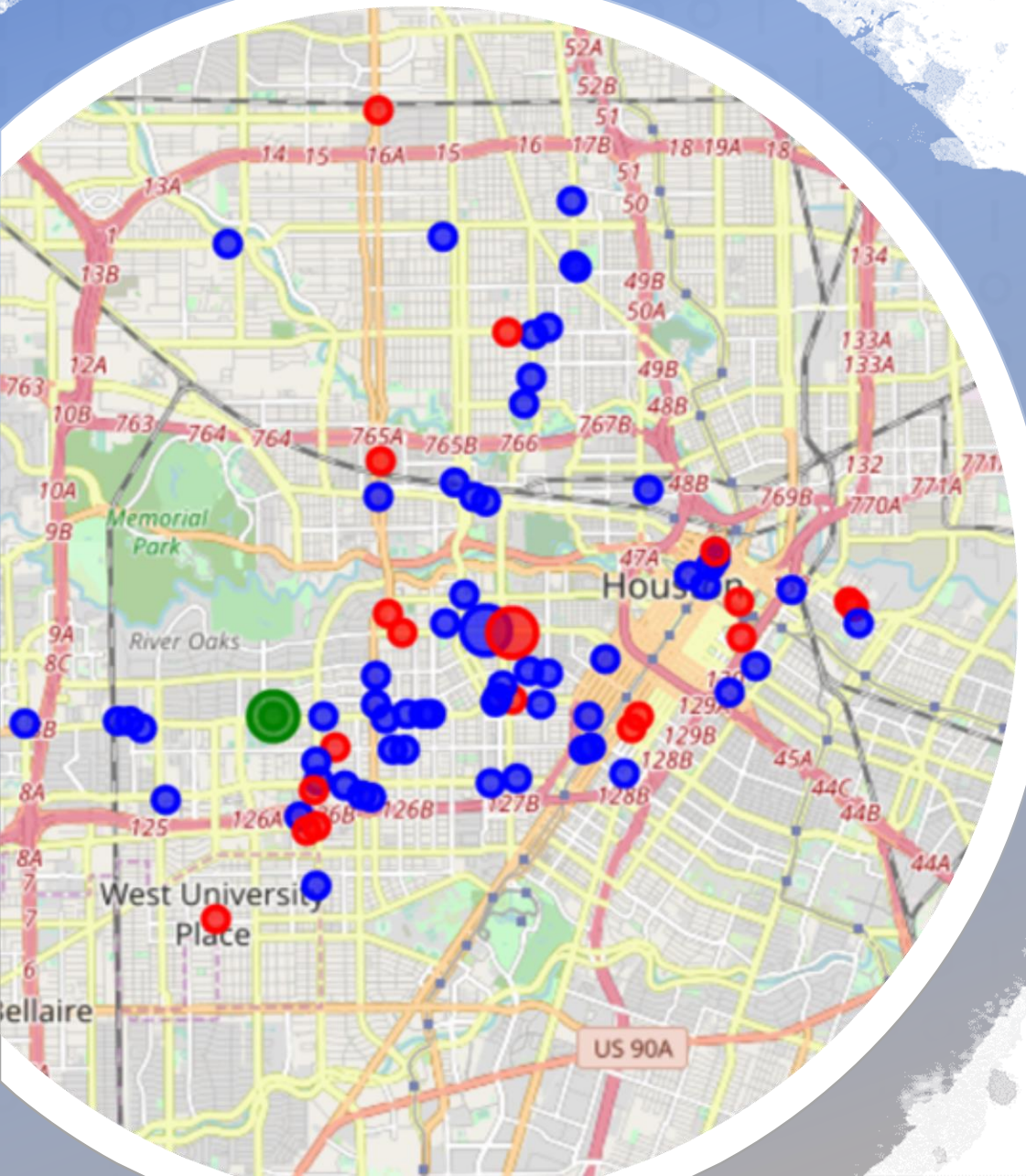




Step 5: `Price for Two` wise venue map

Price_for_two's Value	Price Comment	Color in Map
0.0-25.0	Low price	Dark Green
25.1-50.0	Moderate Price	Green
50.0-100.0	Expensive	Light Blue
100.1 and above	Very expensive	Blue

Step 6: K-means Clustering of the venues



- The venues present in CLUSTER 1 (BLUE) have average PRICE FOR TWO= 28.58 (min=10.00, max=50.00) with the average RATING= 3.91 (min=2.80, max=4.90).

- The venues present in CLUSTER 2 (RED) have average PRICE FOR TWO= 96.11 (min=65.00, max=150.00) with the average RATING= 4.02 (min=3.20, max=4.80).

- The venue in CLUSTER 3 (GREEN) is like an Outlier with the highest PRICE FOR TWO=275.0 and RATING=3.5.

Results

After collecting data from the Foursquare and Zomato APIs for Houston, Texas; we got a list of 240 different venues. However, not all venues from the two different APIs were identical. Hence, after inspecting the two datasets separately, two datasets were merged based on matching names of the venues. Latitude and Longitude can also be used as matching criteria, but in this case, they were not used because of the differences in their values. Completion of the Data cleaning process produced a FINAL dataset with 79 venues.

Following are the results of the analysis:

1. From the Category analysis, it is identified that majority of the venues were Pizza Places and Cafes followed by Beer Gardens and Mexican restaurants and Indian Restaurants.

2. Rating analysis clearly depicts that majority of the venues have good ratings i.e. 3.5 to 4.0 and at least 3rd quarter of the venues have their ratings above 3.5. Mapping the venues with the corresponding Rating Indicator, it is found that all the venues between Westheimer Road and Bissonnette Street have ‘Very Good’ and ‘Good’ rating values i.e. above 3.0 for sure.

3. From the Price for two analysis, it is deduced that majority of the venues can provide average meal for two people at the cost lower than 50 and very few venues can be considered expensive. Mapping the venues with the corresponding Price Indicator, one can say that venues between West Gary Street and Southwest Freeway are pocket friendly.

Discussion

Discussion on the clusters formed:

- Cluster 1: High Value Pocket Friendly food:
  60 of the venues falls under this cluster having very good rating averaged to 3.9 and that too at very pocket friendly price of 50 for two with average price of 28.58

- Cluster 2: Delicious food at Reasonable Price:
  18 out of the total venues can be the best choices for food lovers being averagely rated at 4.02 with the expected average cost of 96 for two people starting at as low as 65.

- Cluster 3: Expensive Mistake:
  Only 1 restaurant having price for two as high as 275 with the moderate rating of 3.5. Only few businessmen will try to opt for the venues under this cluster.

CONCLUSION

Purpose of this project was to leverage FOURSQUARE API to get the venues within 5 miles of the center of Houston, Texas. All these venues were fed to ZOMATO API to get the ratings and price details for the restaurants and food serving venues. These were later on analyzed and grouped into 3 clusters shown above. **These findings can help the visitor's find the best dining place for him/her as per their venue rating and price preferences.**

This project can even be extended as **a Restaurant Recommendation System** to recommend the restaurants on the basis of mentioned Rating's range and Price's Range by the user. All needed is to **feed the location i.e. "city name, state" in the highlighted location cell.**