Data analysis of venues in Bangalore & Hyderabad, India

Ramachandra N
July 30, 2020

A. Introduction

A.1. Description & Discussion of the Background

In Asia, growth is increasingly being driven by successful cities or regions within countries. Two of Asia's three fastest-growing cities are **Bangalore** and **Hyderabad** in southern India. According to Oxford Economics, Bangalore should achieve average annual real GDP growth of **9.9%** over 2020-2024 – far above aggregate growth for India of 6.8% and average growth for Asian cities of 3.9% over the same period. The third fastest-growing Asian city is Hyderabad with average GDP growth of **7.8%**.

This dynamic nature of cities attracts the visitors from all across the world. Due to its dynamic changes, the venues in the cities also change dynamically. This makes it difficult to obtain information about venues in cities for visitors based on their interests.

Hence, to give a clear picture about the cities for visitors, we created maps & applied data analytics on cities data. This makes it easy for visitors to choose the best city for their exploration based on their interests. We are considering only two cities (Bangalore & Hyderabad) for this study.

A.2. Data description

To solve this problem, we extract data from,

1. Google maps: Used to extract the geographical coordinates (latitude & longitude) of cities Bangalore & Hyderabad.
2. Foursquare API: Used to extract the venue data centered around the cities Bangalore & Hyderabad.

The Venue data includes, venue name, id, categories, location coordinates, rating, address, Like count, ratingSignals, tip count. Exploratory data analysis would be performed on Venue data to find out how venue categories such as Restaurants, bars,

pubs, shopping malls, hotels, historical places etc are spread across the city within specified radius. From this data best venue categories would be identified & clustered using K-means clustering algorithm. Some visitors are Foodie, so we recommend a best city based on the number of Restaurants, their rating, user likes etc. Some visitors may like shopping , so we recommend a city which is best for shopping based on the number of shopping malls & their rating. Some visitors may like to visit historical places like monuments, rivers, museums, so we recommend a city based on the number of historical places available & their rating. Like the best city can be recommended based on visitors interests. This analyzed data can be used to build recommended systems for chosen cities.

B. Methodology

The goal of the project is to **recommend a best city to visitors based on their interests using City venue data**. Considering only two cities **Bangalore** & **Hyderabad** for this study.

As a first step, **Foursquare API** is used to extract the venue data for the cities Bangalore & Hyderabad within the radius of **5** km. The fetched venue data contains 203 venues & 6 attributes for Bangalore & 93 venues & 6 attributes for Hyderabad.

The venue data that has main attributes Venue ID, Venue name, Venue categories, address, latitude, & longitude for each city.

For Bangalore,

	id	name	categories	lat	lng	address
0	4bcd805cfb84c9b61512223e	UB City	Shopping Mall	12.971709	77.595905	at Vittal Mallya Rd
1	51d1245e498ef93fd0e713bb	JW Marriott Hotel Bengaluru	Hotel	12.972362	77.595051	24/1 Vittal Mallya Road
2	520390f3498ebe0f1dd63e9e	Smoke House Deli	Deli / Bodega	12.971656	77.598254	52,53
3	4bc1cd90b492d13a4e74a660	Toscano	Italian Restaurant	12.971980	77.596066	UB City Level 2 Concorde Block
4	4bbc4513e45295217db855a4	Cubbon Park	Park	12.977042	77.595277	M G Road

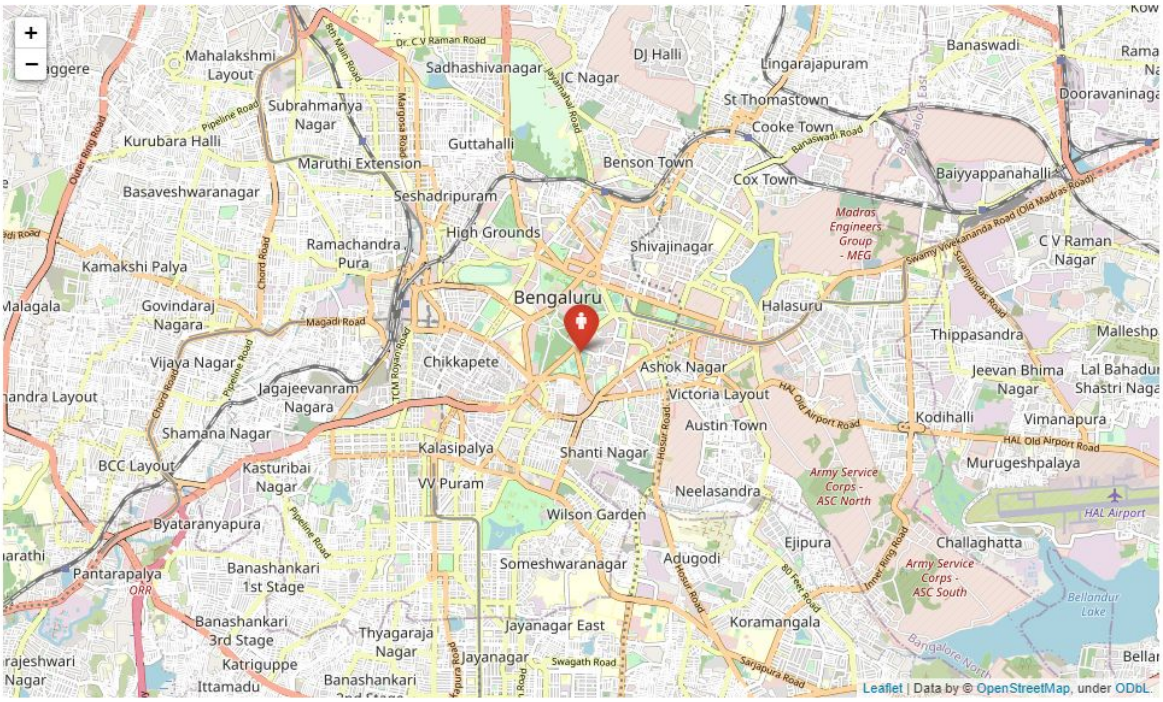
For Hyderabad,

	id	name	categories	lat	lng	address
0	4dc4a65e18506de4adc5d5e5	Pragati	South Indian Restaurant	17.388088	78.481134	Kandaswamy Lane, Sultan Bazaar, Koti Andhra Pr...
1	4d3d4eca14aa8cfaa6d6b15e	Santosh Dhaba	Indian Restaurant	17.388485	78.479509	NaN
2	4df9c65c62e1e9a24367f9e5	King & Cardinal	Burger Joint	17.400678	78.488575	Himayatnagar
3	4cdd08d4fc973704fe47d905	Mayur Pan Shop	Juice Bar	17.388894	78.480578	Hanuman Tekdi Rd.
4	4e009da21838cb6a1acb225e	Minerva Coffee Shop	Indian Restaurant	17.401410	78.485672	Himyathnagar

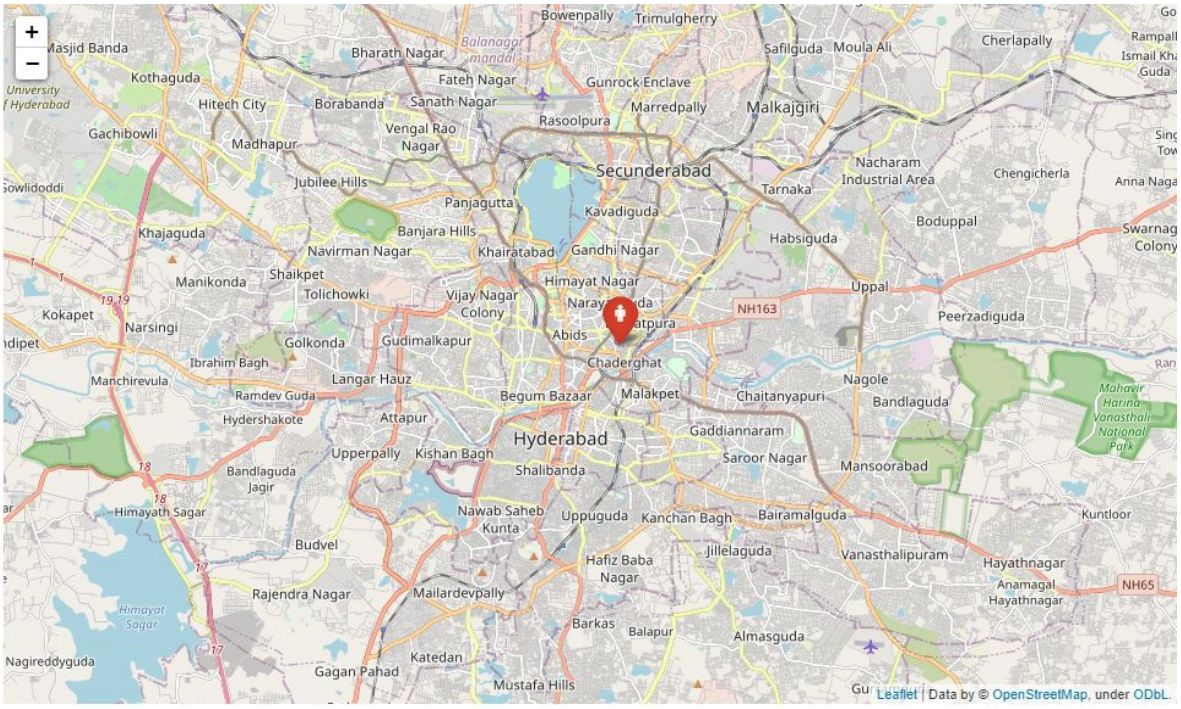
I used a python **folium** library to visualize the geographical details of cities. I used Google maps to extract the geographical details of the cities. I created a folium map centered around the cities using its latitude & longitude values.

Here are the maps,

For Bangalore,

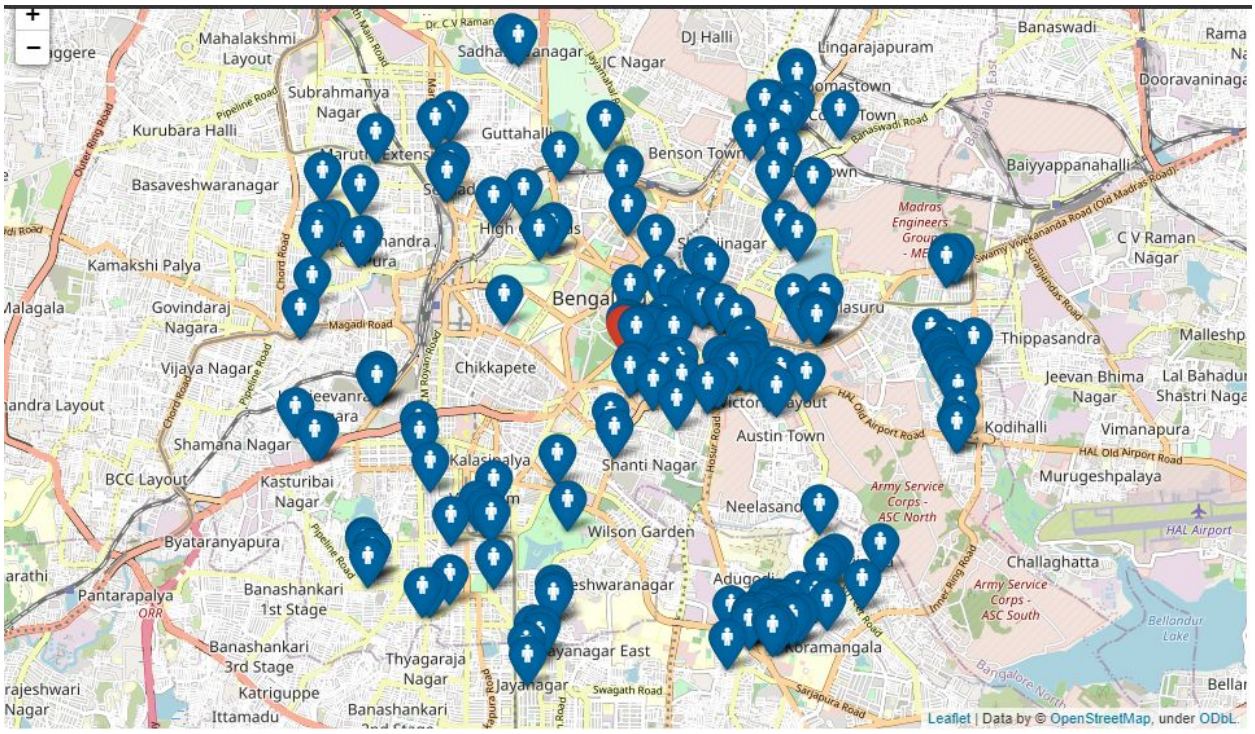


For Hyderabad,

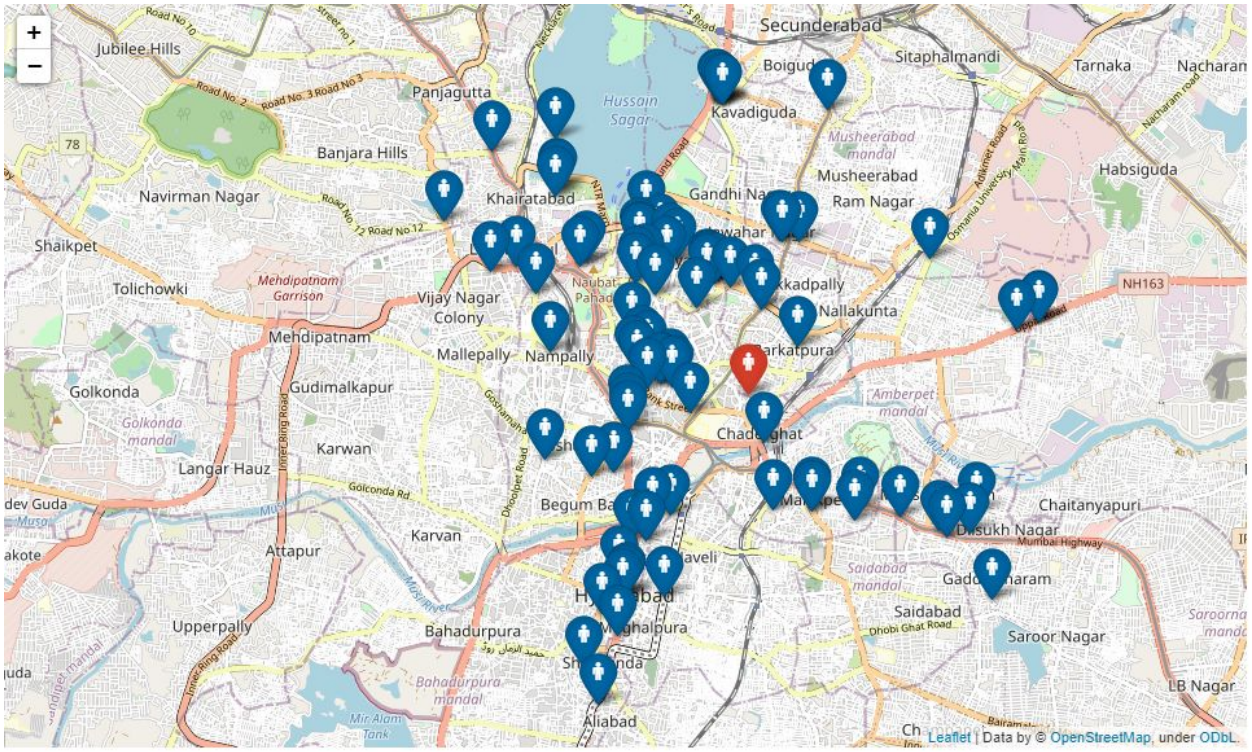


I created a folium map of Bangalore & Hyderabad with venue categories superimposed on top. Here are visuals created using geo coordinates,

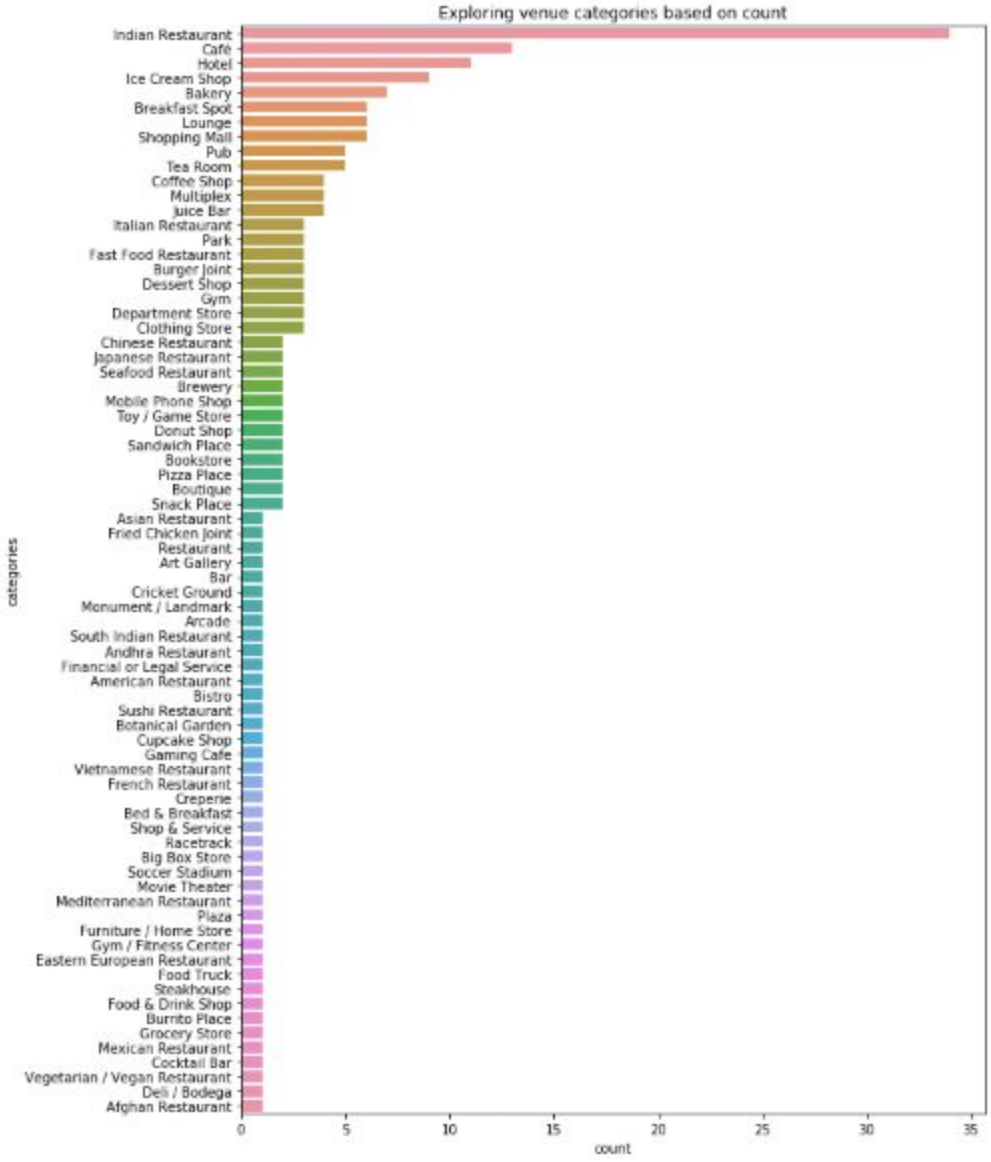
For Bangalore,



For Hyderabad,

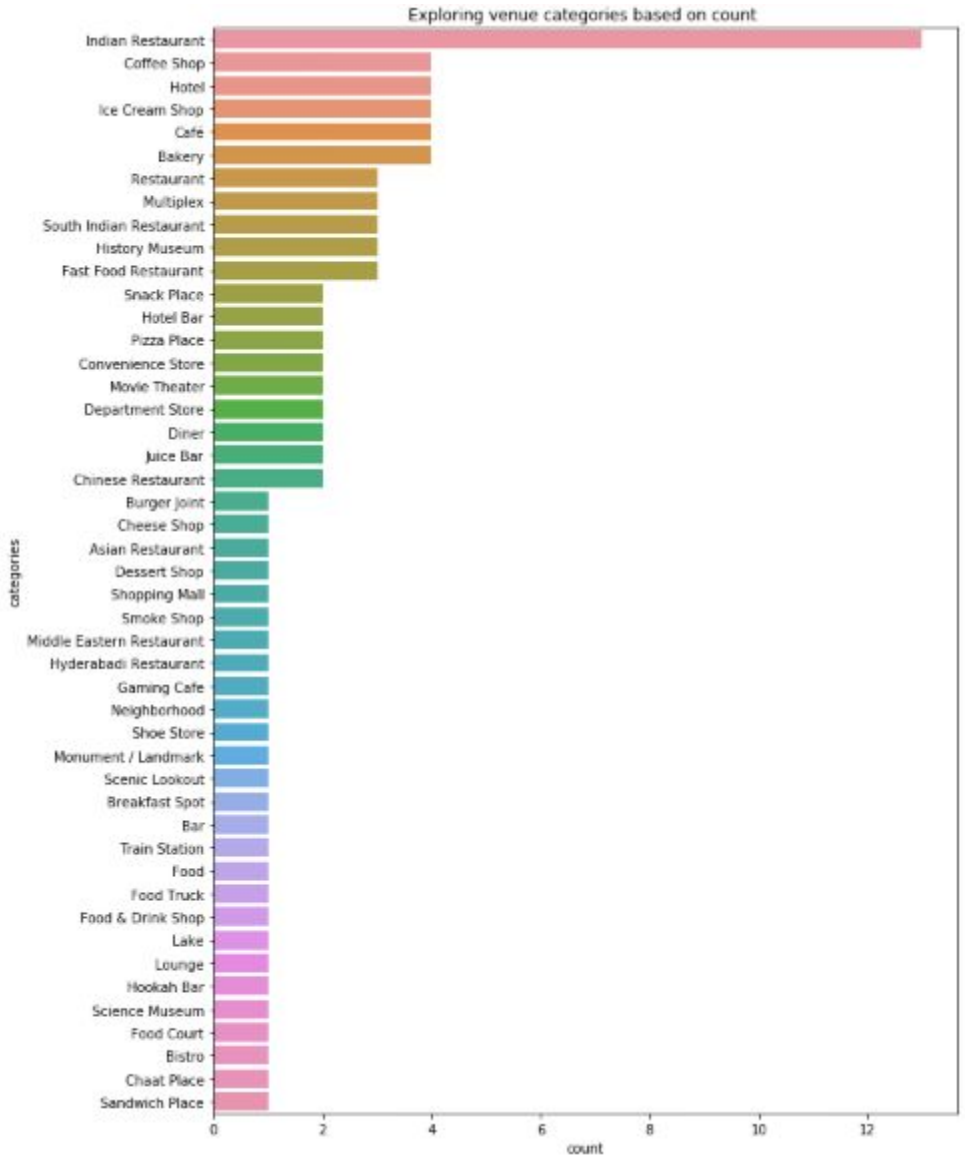


I used **seaborn** to create a count plot of **Venue categories** for both cities.
Here are the plots,
For Bangalore,



A wide variety of cuisines are available such as Indian, Japanese, Mexico, French, American, Burger joint, Seafood & a lot more. Bangalore has the highest number of Indian restaurants than any other restaurants. If you are a foodie, Bangalore is best for you.

For Hyderabad,



We could see that the number of restaurants are less around Hyderabad compared to Bangalore, including South Indian, Hyderabadi, Indian, Asian, Chinese, Fast Food & Middle Eastern. The majority of the venues are Indian Restaurants. If you love Indian recipes, especially Hyderabadi Biryani, Hyderabad is best for you.

I have also extracted the data that contains **tipcount**, **like count**, **rating** & **ratingSignals** for each venue category using Foursquare API.
Here,
tipcount - Number of tips given by user
like count - Number of users who liked that venue

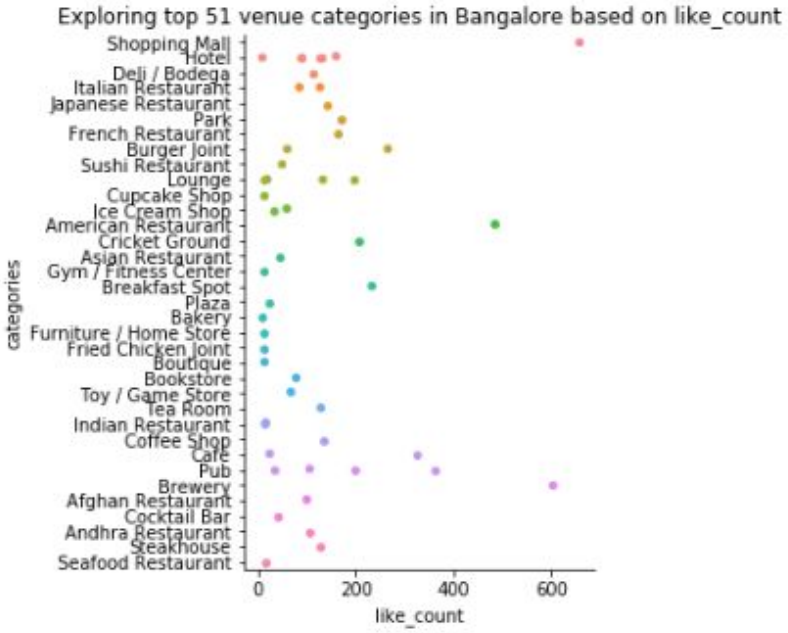
rating - Rating of Venue category given by user
ratingSignals - Number of ratings received from the user
Due to limited API calls, I have extracted this data only for **top 51 venue categories** & merged with venue data. Here are the final data in pandas dataframe,
For Bangalore,

	id	name	categories	lat	lng	address	tipcount	like_count	rating	ratingSignals
0	4bcd805cfb84c9b61512223e	UB City	Shopping Mall	12.971709	77.595905	at Vittal Mallya Rd	104	657	9.0	802
1	51d1245e498ef93fd0e713bb	JW Marriott Hotel Bengaluru	Hotel	12.972362	77.595051	24/1 Vittal Mallya Road	36	128	8.7	166
2	520390f3498ebe0f1dd63e9e	Smoke House Deli	Deli / Bodega	12.971656	77.598254	52,53	37	111	8.9	147
3	4bc1cd90b492d13a4e74a660	Toscano	Italian Restaurant	12.971980	77.596066	UB City Level 2 Concorde Block	65	127	8.6	188
4	4b895510f964a520442c32e3	Shiro	Japanese Restaurant	12.971900	77.596236	3rd Flr., UB City, Vittal Mallya Rd.	70	140	8.4	204

For Hyderabad,

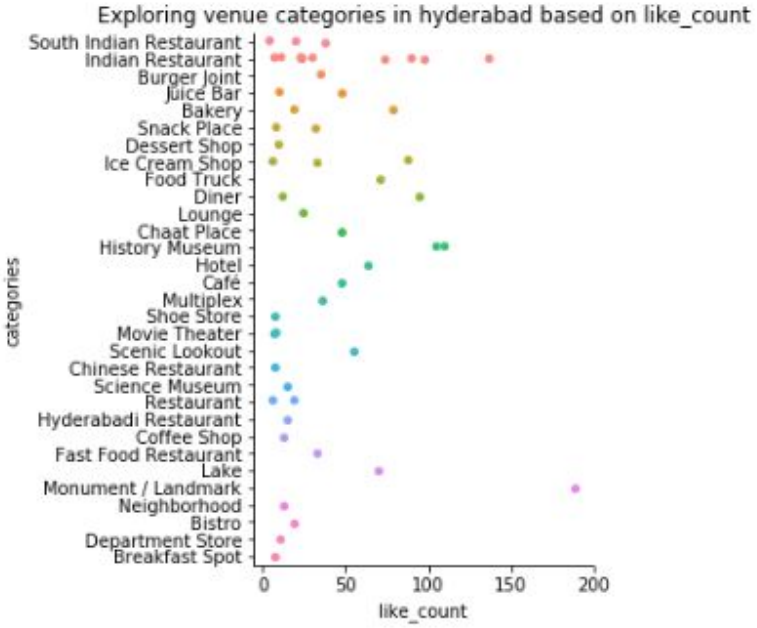
	id	name	categories	lat	lng	address	tipcount	like_count	rating	ratingSignals
0	4dc4a65e18506de4adc5d5e5	Pragati	South Indian Restaurant	17.388088	78.481134	Kandaswamy Lane, Sultan Bazaar, Koti Andhra Pr...	15	38	8.3	55
1	4d3d4eca14aa8cfaa6d6b15e	Santosh Dhaba	Indian Restaurant	17.388485	78.479509	NaN	15	23	8.2	40
2	4df9c65c62e1e9a24367f9e5	King & Cardinal	Burger Joint	17.400678	78.488575	Himayatnagar	20	35	8.4	57
3	4cdd08d4fc973704fe47d905	Mayur Pan Shop	Juice Bar	17.388894	78.480578	Hanuman Tekdi Rd.	21	48	7.9	71
4	4e009da21838cb6a1acb225e	Minerva Coffee Shop	Indian Restaurant	17.401410	78.485672	Himyathnagar	31	74	8.3	108

I created a categorical plot between **venue categories** & **like count** to check how like count is spreaded across all categories.
He are the visuals,
For Bangalore,



From the above plot , we could say that the most liked venue categories are Shopping Mall & Brewery and the least liked categories are Seafood restaurant, Gym/Fitness center, Bakery, & a Hotel. All other restaurants except American got average number of likes.

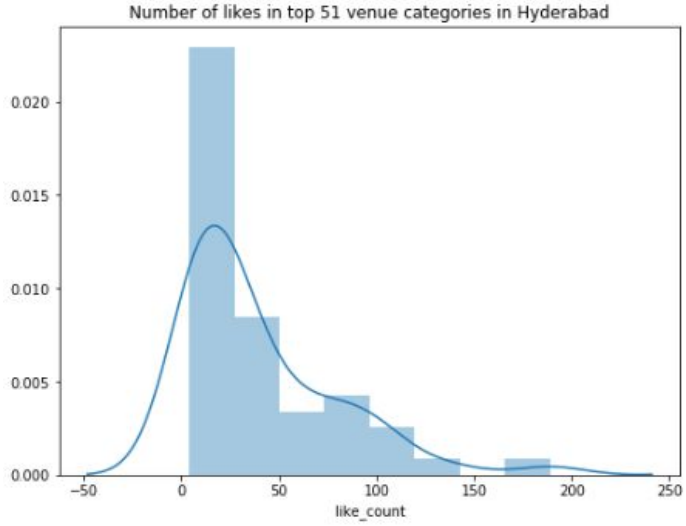
For Hyderabad,



In Hyderabad, the most liked categories are Monument/Landmark & Indian restaurant and least liked category is Ice Cream Shop.

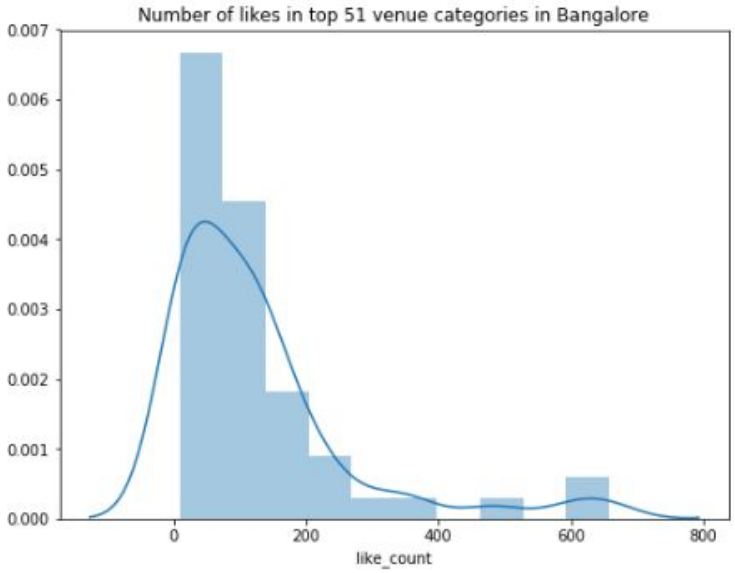
Also created a distribution plot for **like count** to check the average number of users that have liked that venue. Here are the plots,

For Hyderabad,



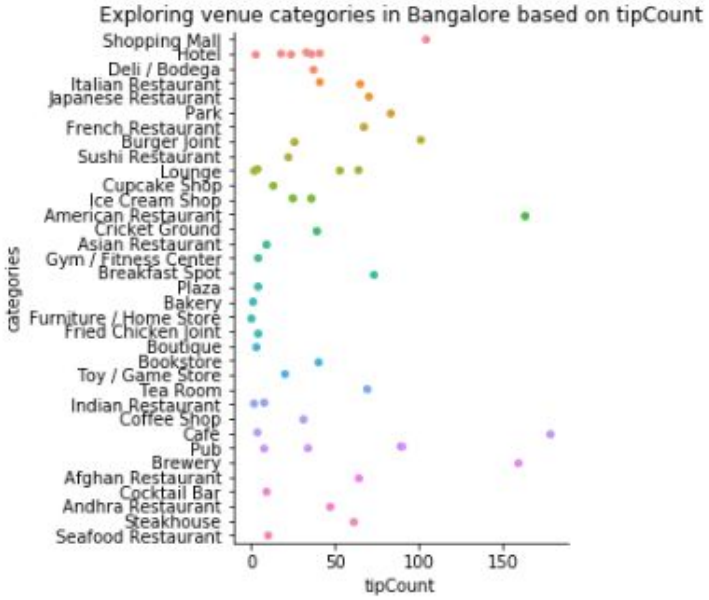
It appears that the average number of user likes that have been reported in top 51 venue categories in Hyderabad is 40.

For Bangalore,



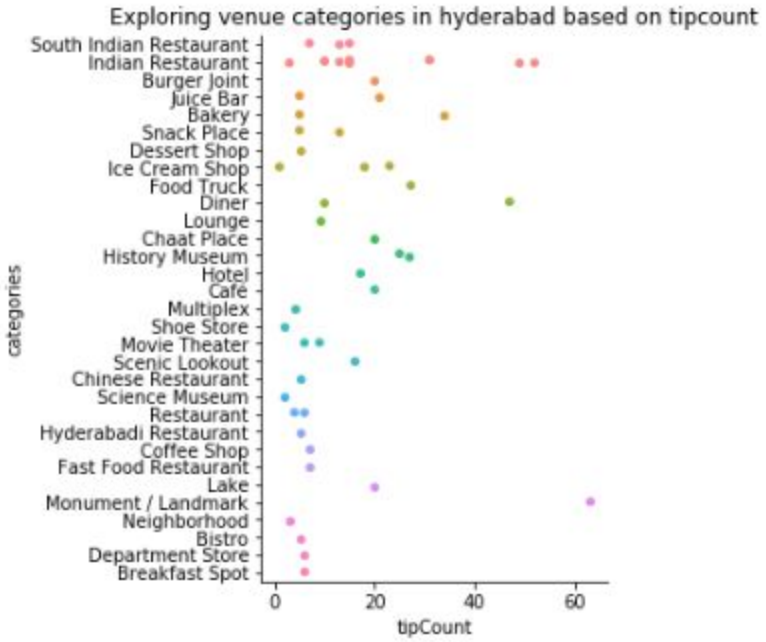
It seems that the average number of user likes that have been reported in top 51 venue categories in Bangalore is 125.

I created a categorical plot between **venue categories** & **tipcount** to check how tip count is distributed across all categories.
For Bangalore,

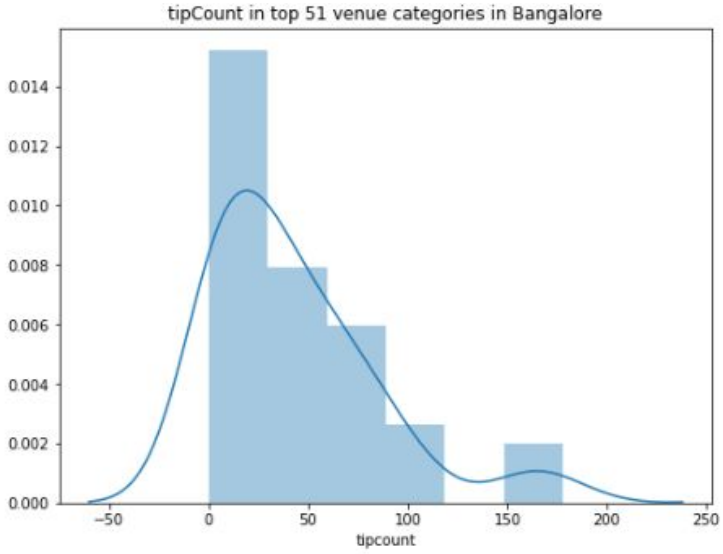


It is evident that categories Cafe, Brewery, & American restaurant have received highest tips & Lounge, Furniture/Home Store, & a Hotel have received fewer tips.

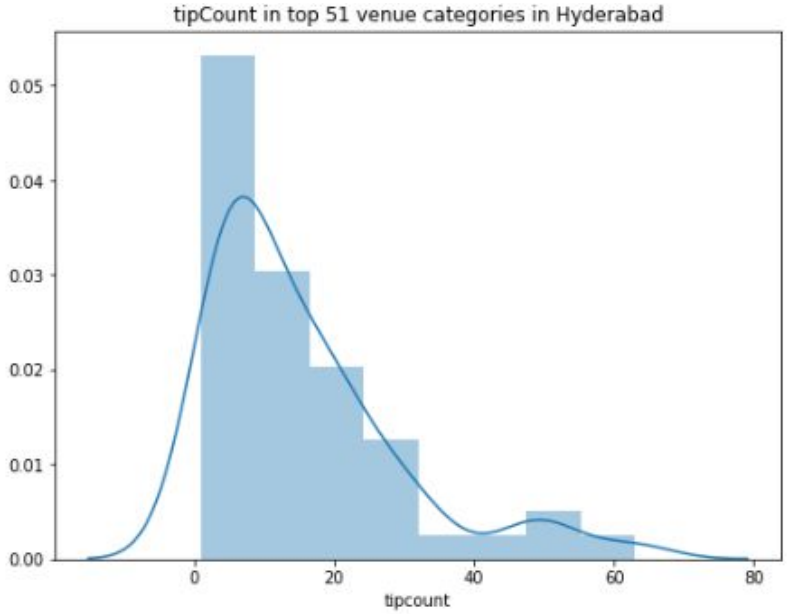
For Hyderabad,



It shows that, Monument/Landmark category received the highest tips & the Ice Cream Shop category received the lowest tips for Hyderabad city.
Also created a distribution plot for **tipcount** to check the average number of tips given by users across all categories.
Here are dist plots,
For Bangalore,

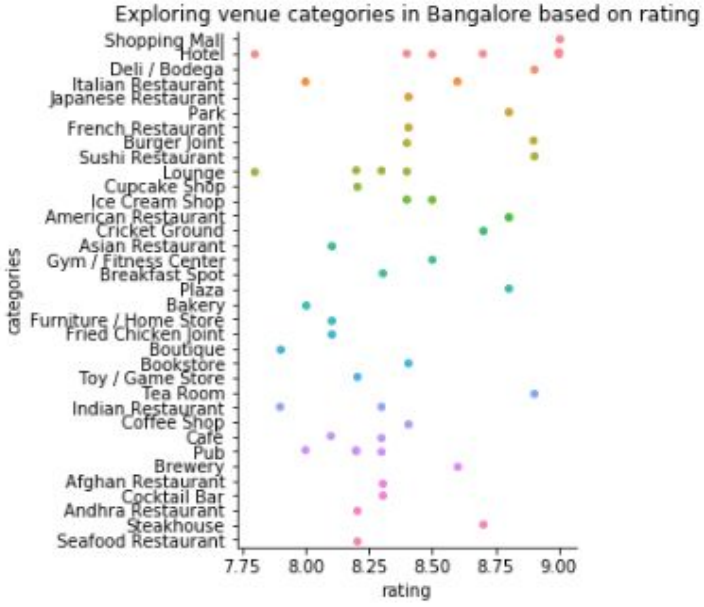


It seems that the average number of user tips that have been reported in top 51 venue categories in Bangalore is 42.
For Hyderabad,



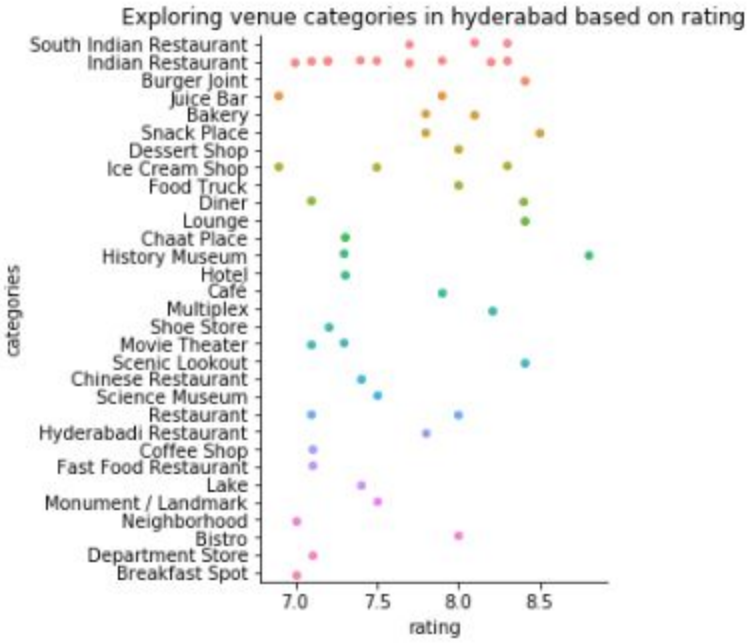
It appears from the plot that the average number of user tips that have been reported in top 51 venue categories in Hyderabad is 15.

I created a categorical plot between **venue categories** & **rating** to check how rating is distributed across all categories. Here are the plots,
For Bangalore,



From the plot, it appears that users have given the highest rating to Shopping Mall and given least rating to Lounge. All restaurants got above average ratings.

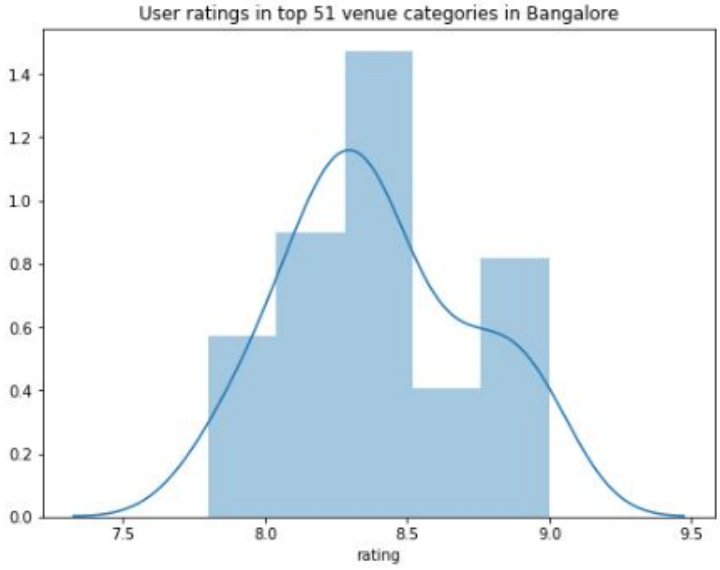
For Hyderabad,



It seems from the above plot, History Museum got the highest rating & Ice Cream got the lowest rating. ALL restaurants got above average ratings.

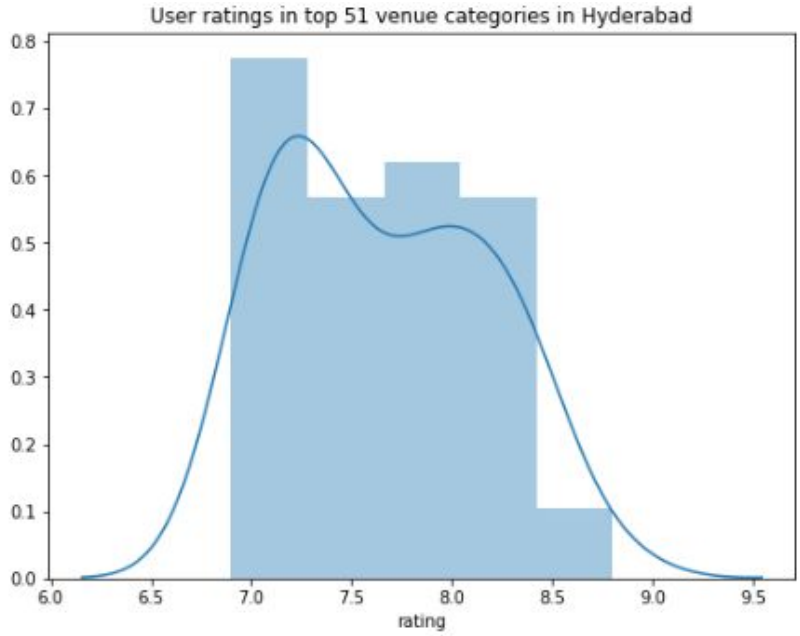
Also created a distribution plot for **rating** to check the average number of ratings given by users across all categories. Here the dist plots,

For Bangalore,



The above plot shows that, the average number of user ratings that have been reported in top 51 venue categories in Bangalore is 8.0.

For Hyderabad,



It is evident from above the plot that the Average number of user ratings that have been reported in top 51 venue categories in Hyderabad is 7.

Now let's create bins for different ratings & plot them in different colors on map. Divide the rating into 3 bins. Here are the bins,

Bin range	Category
6.0-7.0	Okay
7.0-8.0	Good
8.0-9.0	Very Good

The Resulting bins category is merged into the original dataframe showing Okay, Good, Very Good in rating_bins column for each category based on rating.

Rating_bins are mapped by colors for folium map,

Okay - 'dark blue'

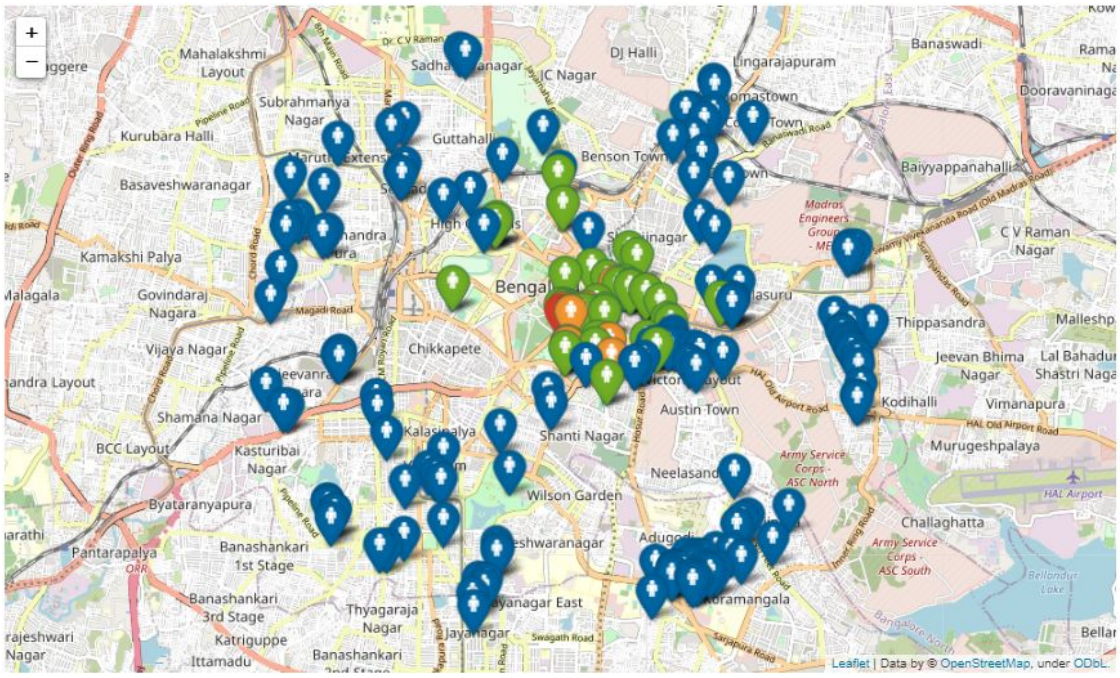
Good - 'orange'

Very Good - 'green'

For Bangalore,

	id	name	categories	lat	lng	address	tipcount	like_count	rating	ratingSignals	rating_bins
0	4bcd805cfb84c9b61512223e	UB City	Shopping Mall	12.971709	77.595905	at Vittal Mallya Rd	104	657	9.0	802	Very Good
1	51d1245e498ef93fd0e713bb	JW Marriott Hotel Bengaluru	Hotel	12.972362	77.595051	24/1 Vittal Mallya Road	36	128	8.7	166	Very Good
2	520390f3498ebe0f1dd63e9e	Smoke House Deli	Deli / Bodega	12.971656	77.598254	52,53	37	111	8.9	147	Very Good
3	4bc1cd90b492d13a4e74a660	Toscano	Italian Restaurant	12.971980	77.596066	UB City Level 2 Concorde Block	65	127	8.6	188	Very Good
4	4b895510f964a520442c32e3	Shiro	Japanese Restaurant	12.971900	77.596236	3rd Flr., UB City, Vittal Mallya Rd.	70	140	8.4	204	Very Good

The venue categories are plotted on map colored by rating_bins as shown in below,

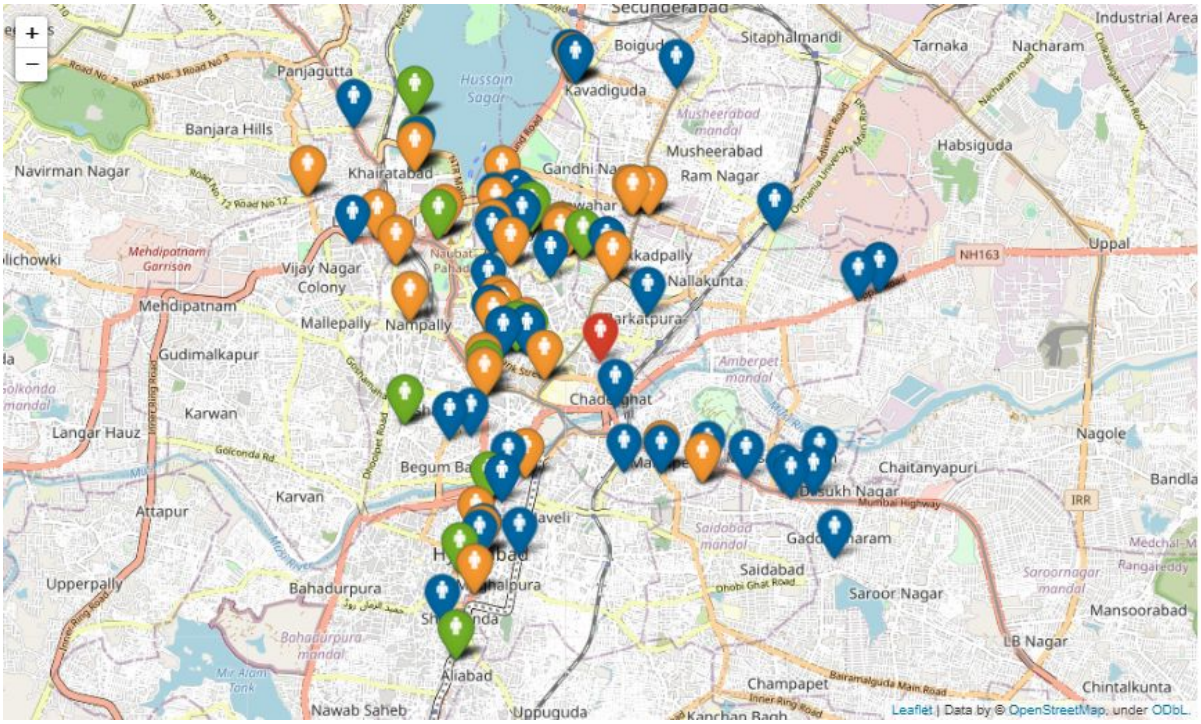


As you can see that, high rated venue categories (between 8 & 9) are clustered around Bangalore club, Vittal Mallya road & Church street road.

For Hyderabad,

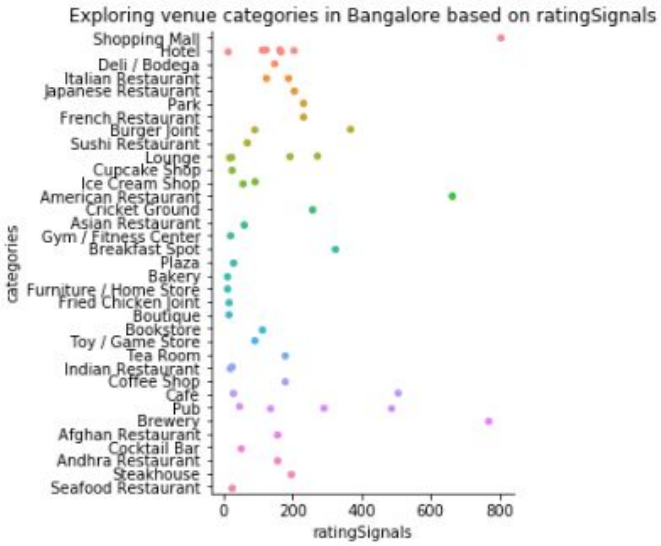
	id	name	categories	lat	lng	address	tipcount	like_count	rating	ratingSignals	rating_bins
0	4dc4a65e18506de4adc5d5e5	Pragati	South Indian Restaurant	17.388088	78.481134	Kandaswamy Lane, Sultan Bazaar, Koti Andhra Pr...	15	38	8.3	55	Very Good
1	4d3d4eca14aa8cfaa6d6b15e	Santosh Dhaba	Indian Restaurant	17.388485	78.479509	NaN	15	23	8.2	40	Very Good
2	4df9c65c62e1e9a24367f9e5	King & Cardinal	Burger Joint	17.400678	78.488575	Himayatnagar	20	35	8.4	57	Very Good
3	4cdd08d4fc973704fe47d905	Mayur Pan Shop	Juice Bar	17.388894	78.480578	Hanuman Tekdi Rd.	21	48	7.9	71	Good
4	4e009da21838cb6a1acb225e	Minerva Coffee Shop	Indian Restaurant	17.401410	78.485672	Himyathnagar	31	74	8.3	108	Very Good

The venue categories are plotted on map colored by rating_bins for Hyderabad as shown in below,

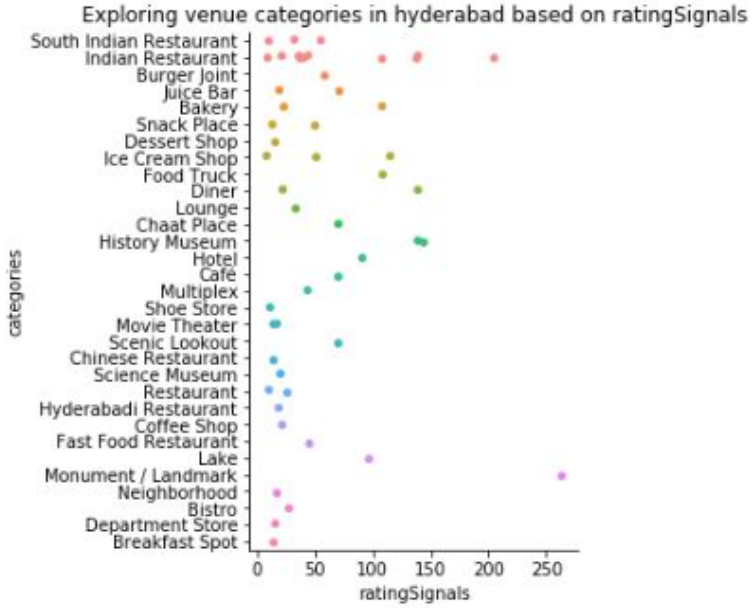


As you can see that the high rated venue categories (between 7 & 8 and 8 & 9) are situated in Himayat Nagar, Abids, Basheer Bagh flyover road & Charminar road.

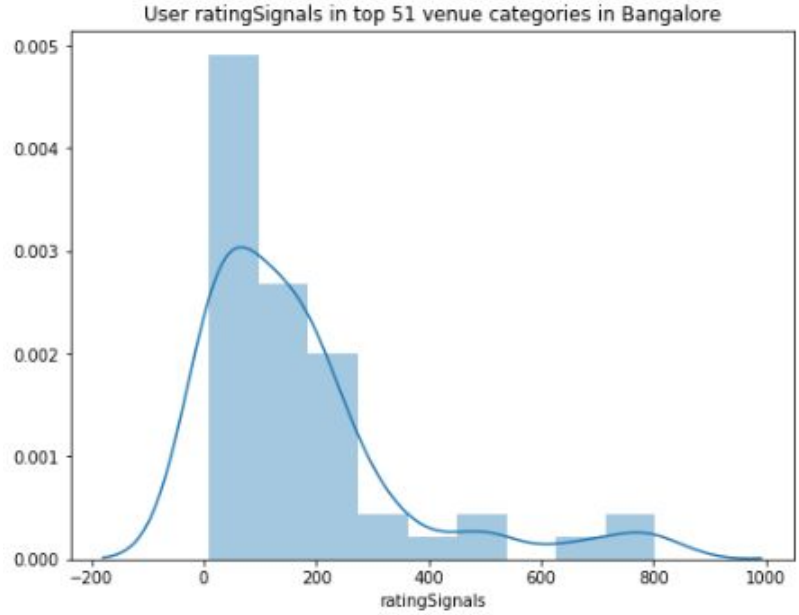
I created a categorical plot between venue **categories** & **ratingSignals** to check how ratingSignals are spread across all categories.
For Bangalore,



It appears from the plot that, Brewery & Shopping mall category received the highest number of ratingSignals & least number of ratingSignals received by a Hotel.
For Hyderabad,



Monument/Landmark & Ice cream shop categories have received the highest & lowest number of ratingSignals respectively in Hyderabad.
Also created a distribution plot for **ratingSignals** to check average of rating Signals across all categories.
For Bangalore,



Average number of user ratingSignals that have been reported in top 51 venue categories in Bangalore is 170 & for Hyderabad is 57.

Clustering of Venue Categories & Key takeaways

Now we will cluster venue categories based on geo coordinates, like count, tipcount, rating & ratingSignals using **K-means clustering**. Number of clusters considered are 3 for clustering.

The resulting clustered labels are inserted into the dataframe to create a folium map centered around cluster centers for each city.

The resulted dataframe is,

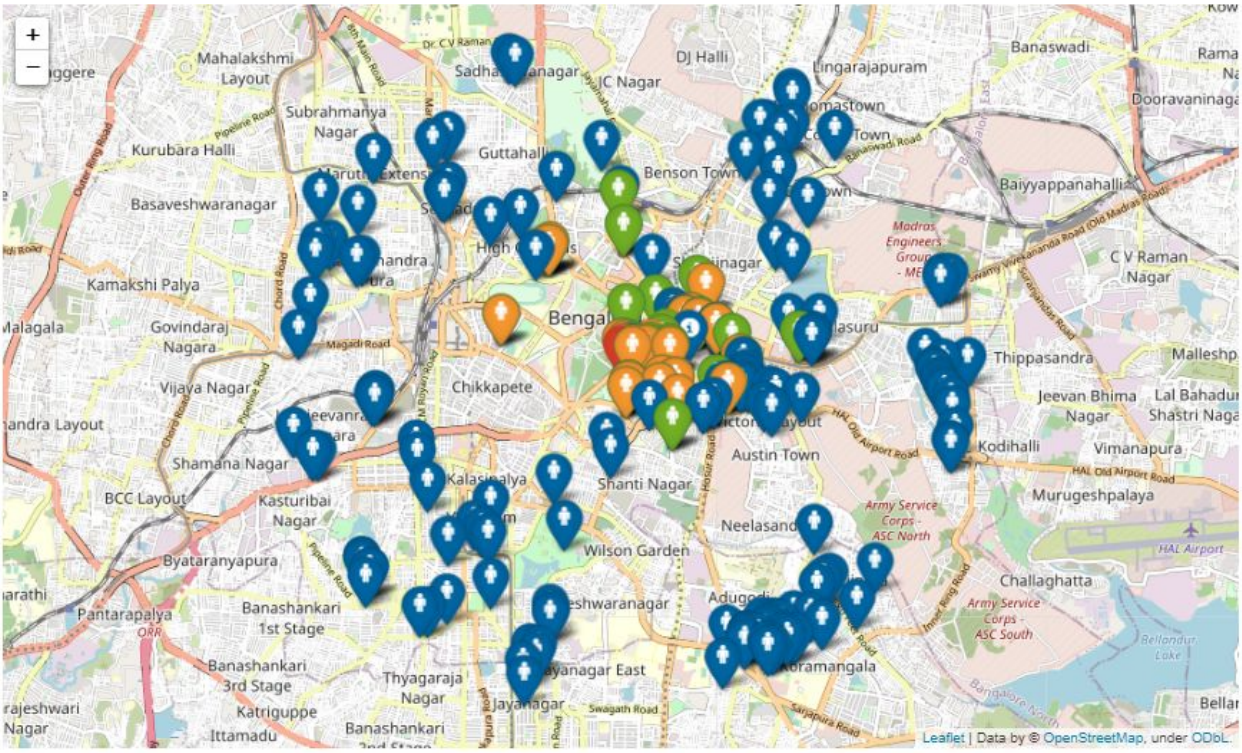
For Bangalore,

	Cluster_labels	id	name	categories	lat	lng	address	tipcount	like_count	rating	ratingSignals	rating_bins
0	0	4bcd805cfb84c9b61512223e	UB City	Shopping Mall	12.971709	77.595905	at Vittal Mallya Rd	104	657	9.0	802	Very Good
1	2	51d1245e498ef93fd0e713bb	JW Marriott Hotel Bengaluru	Hotel	12.972362	77.595051	24/1 Vittal Mallya Road	36	128	8.7	166	Very Good
2	2	520390f3498ebe0f1dd63e9e	Smoke House Deli	Deli / Bodega	12.971656	77.598254	52,53	37	111	8.9	147	Very Good
3	2	4bc1cd90b492d13a4e74a660	Toscano	Italian Restaurant	12.971980	77.596066	UB City Level 2 Concorde Block	65	127	8.6	188	Very Good
4	2	4b895510f964a520442c32e3	Shiro	Japanese Restaurant	12.971900	77.596236	3rd Flr., UB City, Vittal Mallya Rd.	70	140	8.4	204	Very Good

For Hyderabad,

	Cluster_labels	id	name	categories	lat	lng	address	tipcount	like_count	rating	ratingSignals	rating_bi
0	0	4dc4a65e18506de4adc5d5e5	Pragati	South Indian Restaurant	17.388088	78.481134	Kandaswamy Lane, Sultan Bazaar, Koti Andhra Pr...	15	38	8.3	55	Very Go
1	0	4d3d4eca14aa8cfaa6d6b15e	Santosh Dhaba	Indian Restaurant	17.388485	78.479509	NaN	15	23	8.2	40	Very Go
2	0	4df9c65c62e1e9a24367f9e5	King & Cardinal	Burger Joint	17.400678	78.488575	Himayatnagar	20	35	8.4	57	Very Go
3	0	4cdd08d4fc973704fe47d905	Mayur Pan Shop	Juice Bar	17.388894	78.480578	Hanuman Tekdi Rd.	21	48	7.9	71	Go
4	1	4e009da21838cb6a1acb225e	Minerva Coffee Shop	Indian Restaurant	17.401410	78.485672	Himyathnagar	31	74	8.3	108	Very Go

Now plot the cluster labels on map,
In the map cluster **0** , **1** , & **2** are represented by **dark blue**, **orange**, & **green** color respectively.
For Bangalore,



Key takeaways

From the map, we could say that,

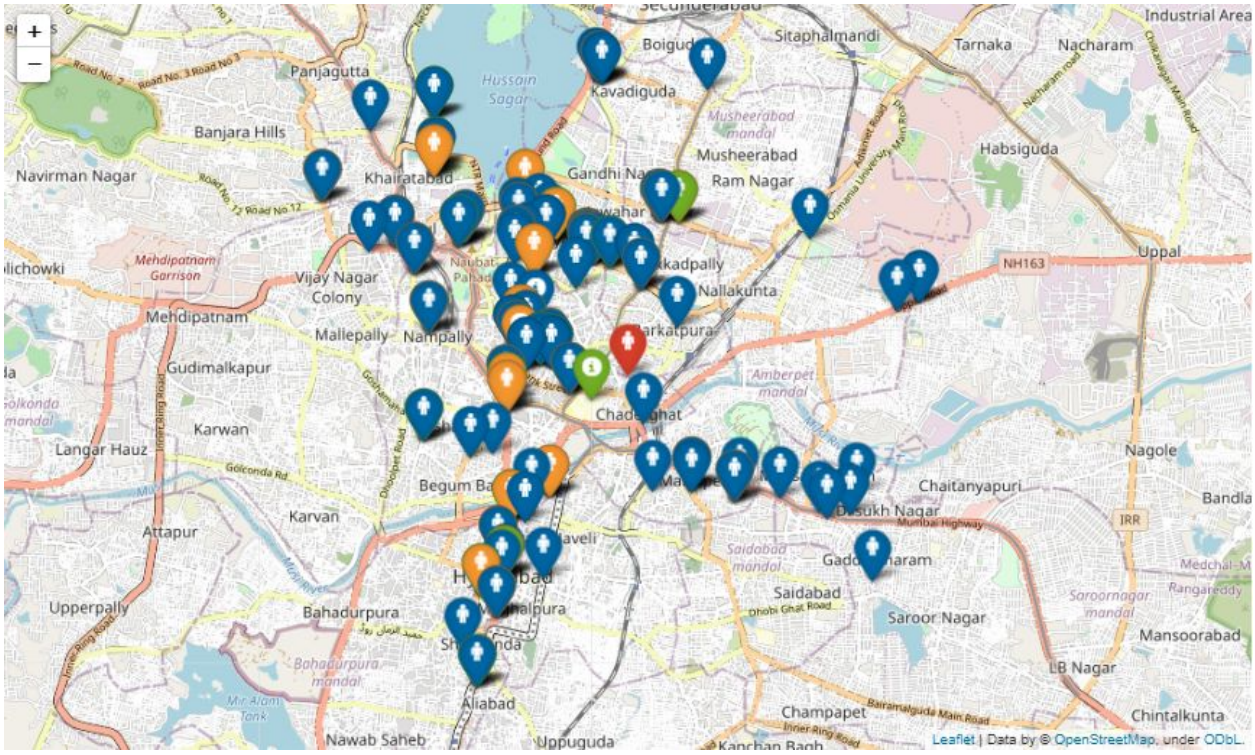
1. Cluster first is spreaded across the whole city & has majority venues.
2. Cluster second & three are spreaded at the city center & have limited venues

For Hyderabad,

Key takeaways

From below plot, we could say that,

1. Cluster first is spreaded across the whole city & has majority venues.
2. Cluster second & third are less sparsely spreaded & have limited venues



C. Results

In this section I would like to describe the results obtained from the above analysis.
From above analysis, key results are,

- Number of venue categories in Bangalore more than Hyderabad within a radius of 5 km.
- A wide variety of cuisines are available such as Indian, Japanese, Mexico, French, American, Burger joint, Seafood in Bangalore than Hyderabad. This makes Bangalore unique in the restaurant category.
- Most liked venue categories in Bangalore are Shopping Mall & Brewery & in Hyderabad are Monument/Landmark & Indian restaurant.
- Most rated venue category in Bangalore is the Shopping Mall & in Hyderabad is the History Museum.
- Number of highest rated venue categories are more in Bangalore than Hyderabad.
- Highest rated venue categories are situated at the city center in Bangalore, but in Hyderabad spreaded across the city.

D. Discussion

As I mentioned earlier, this study was carried out by considering only top 51 venue categories due to limited API calls. So results may vary if we consider all venue categories within a radius of 5km.

Also, I considered geo coordinates at the center of cities, results may vary if we consider geo coordinates all across the city. I used K-means clustering for clustering. Other clustering methods can be used for better results.

Other attributes like price, distance, user details etc can be added to the above data to get more meaningful insights.

I ended this project by visualizing data & clustering categories on maps centered around cities. This work can be carried out to build a recommender system for visitors.

E. Conclusions

From above insights, I would recommend to explore,

1. Bangalore, if you are a foodie, want to try & explore a wide variety of cuisines such as Japanese, French, American, Indian & a lot more. Also if you love shopping, Bangalore has the highest rated shopping malls.

2. Hyderabad, if you want to try Indian cuisines, specially Hyderabadi Biryani & would love to visit historical places such as Museums, Monument (Charminar), river (Hussain sagar), which have rated highest & got maximum likes & tips.

3. Bangalore, if you want high rated & most liked venue categories spread across the city.

G. References

1. .https://foursquare.com/
2. .https://www.google.com/maps
3. https://www.financialexpress.com/money/bangalore-hyderabad-to-be-two-fastest-growing-asian-cities-over-2020-24/1822445/
4. https://www.linkedin.com/pulse/housing-sales-prices-venues-data-analysis-ofistanbul-sercan-y%C4%B1ld%C4%B1z/
5. https://en.wikipedia.org/wiki/Bangalore
6. https://en.wikipedia.org/wiki/Hyderabad