Coursera Course: IBM Data Science Capstone Project Recommendation System for Restaurants in Hyderabad under GHMC.
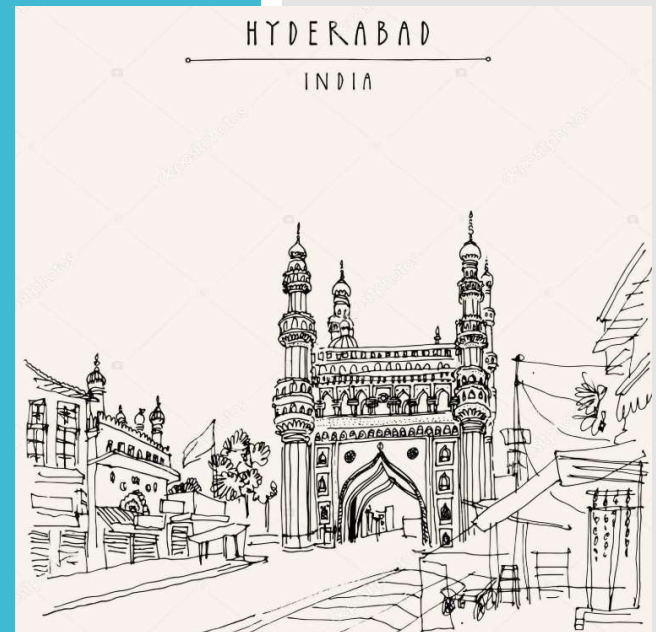
By Devender Madura

HYDERABAD
INDIA

Introduction
## Business Problem

- Hyderabad?
- Opening a Restaurant?

- The Greater Hyderabad Municipal Corporation (GHMC) is the civic body that oversees Hyderabad, the capital and largest city of the Indian state of Telangana.

- It's evident that the city of Hyderabad has a great affinity with varieties of food and is a great city and presents lot of opportunities for opening restaurants & eateries.

- While there are many factors to consider while opening a restaurant like concept, budget, cost, time, staff, vendors, licenses etc. We shall be considering 2 major factors Location and Cuisine that need to be thought through in detail to be able to open a successful restaurant. We shall be trying to use some Data Science methodologies to try and get some key findings that can help make this decision.

Description of Data & Sources

- **Hyderabad Locations:** We can identify and pull all the neighborhoods from Wikipedia link below, where they are referred to as Wards using a simple gethtml function
https://en.wikipedia.org/wiki/List_of_Hyderabad_Corporation_wards

- **Hyderabad Restaurants:** For the venues we will be pulling the data from FourSquare API and Zomato API.
  - From FourSquare API we are able to source only 200 restaurants
  - From Zomato API we are able to source and identify 1050 restaurants along with rating, average cost and voters information.

Methodology
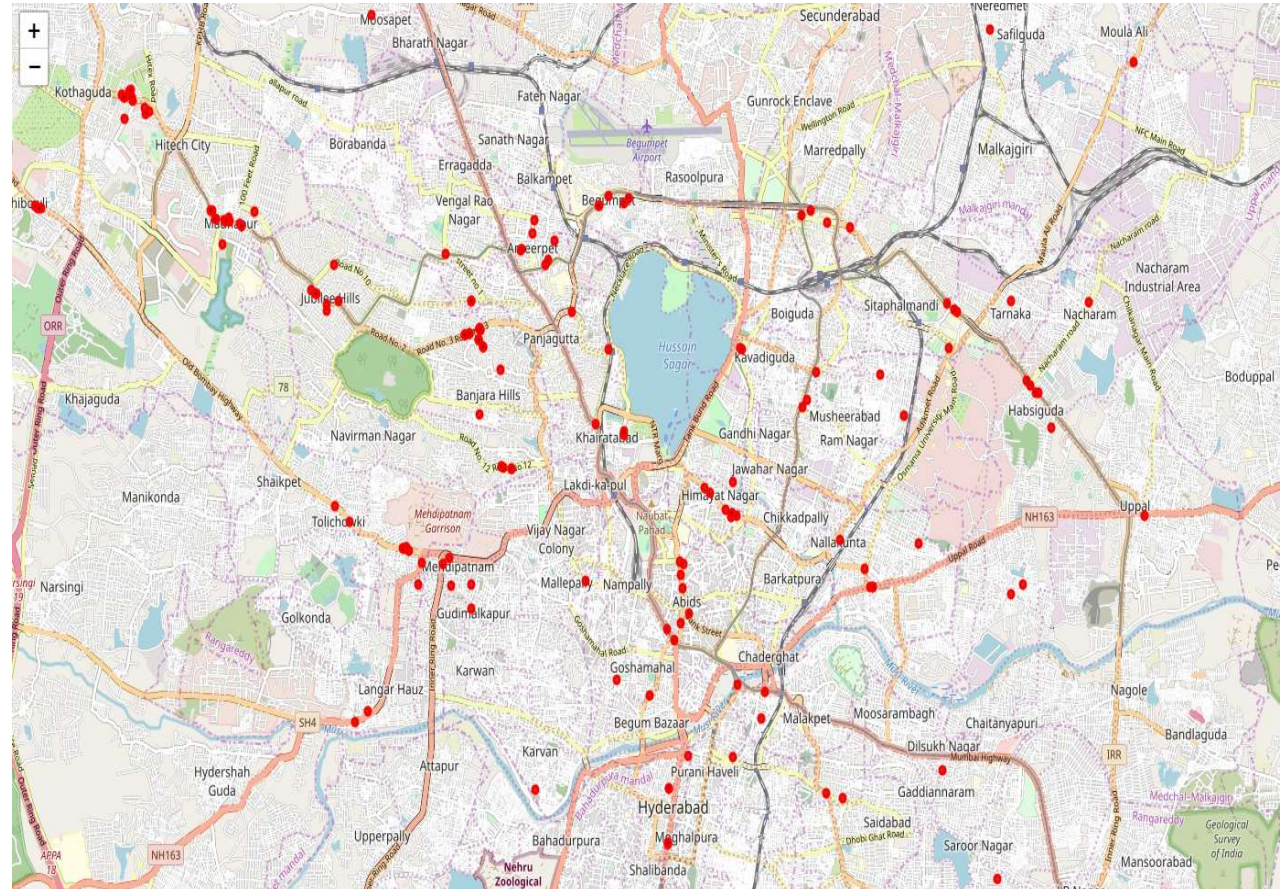**Exploratory Data Analysis**

All Hyderabad Ward Locations
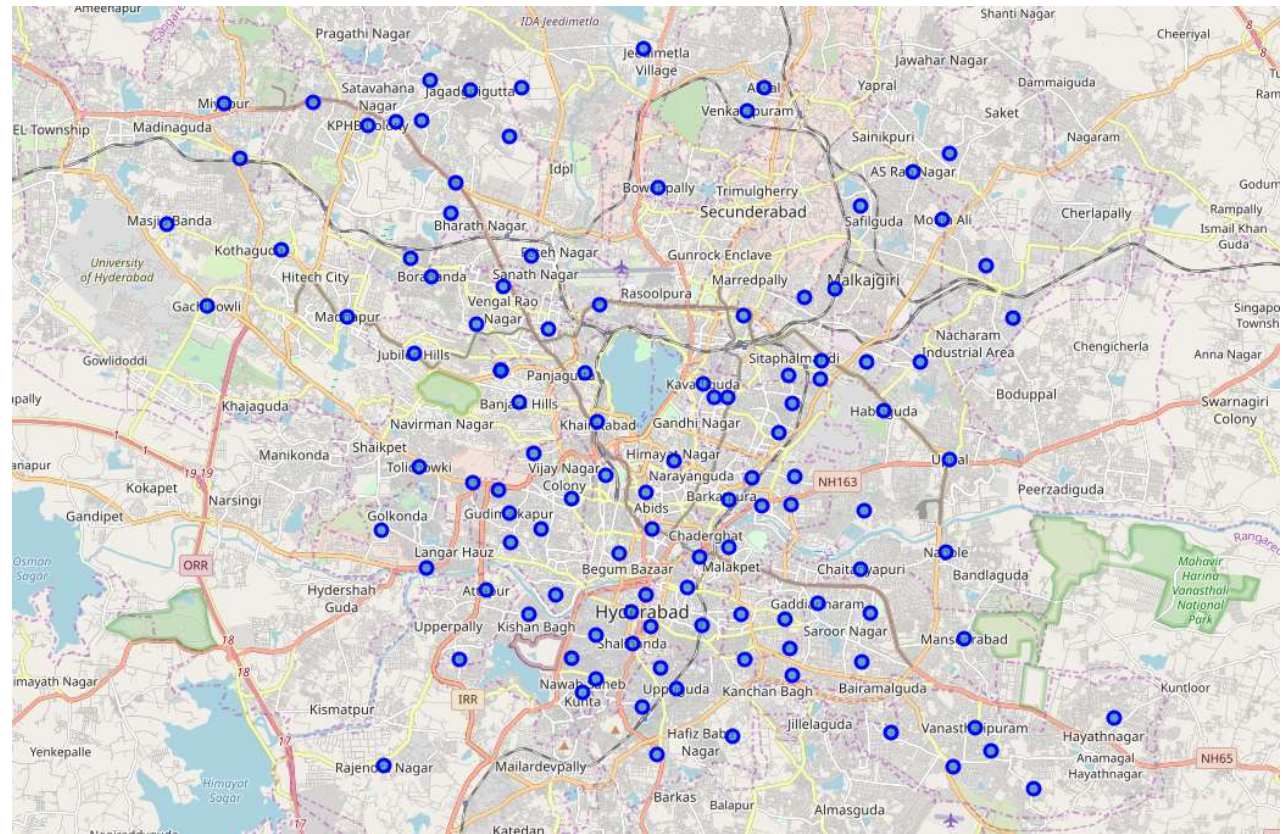
Methodology
Exploratory Data Analysis

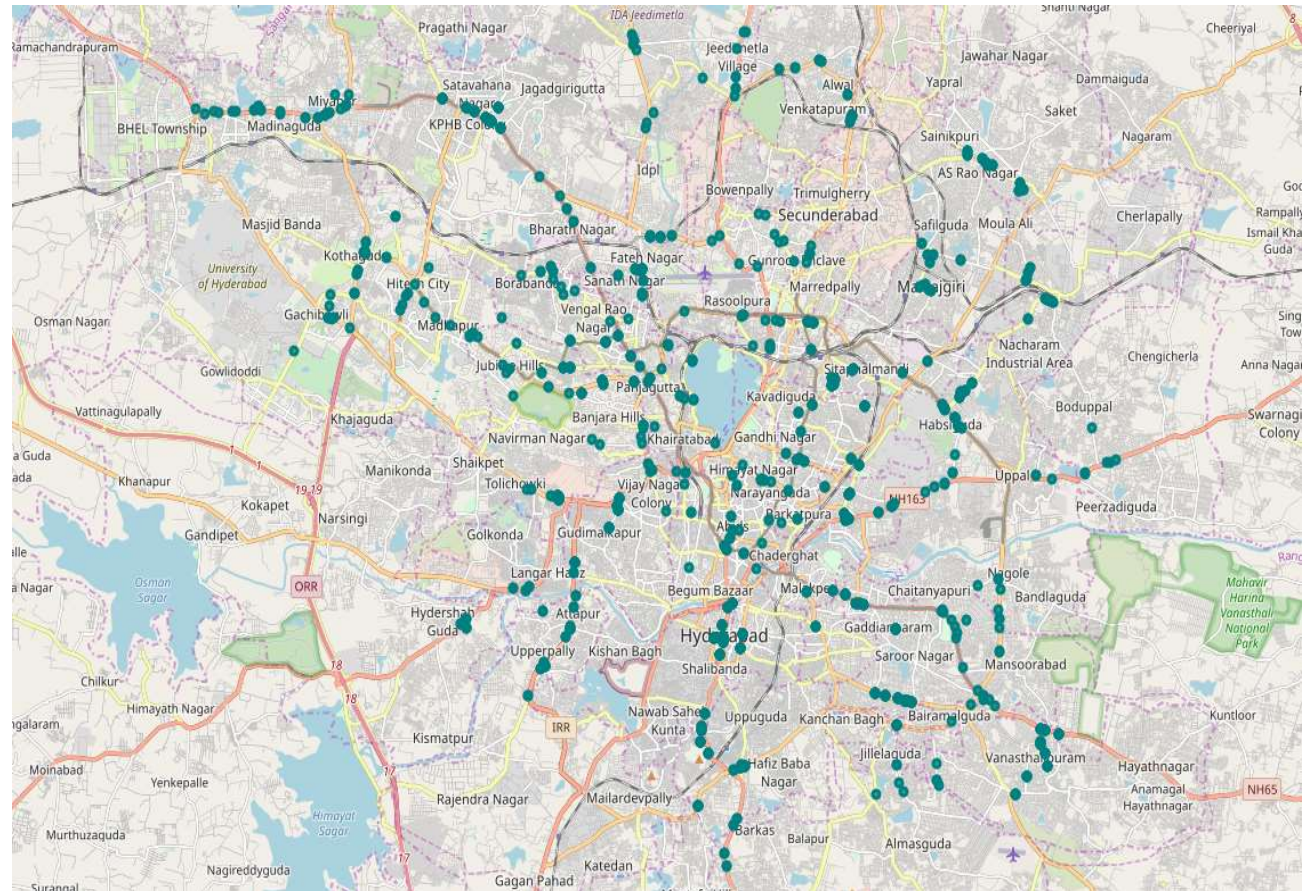**Hyderabad Restaurants from Foursquare API**

Methodology
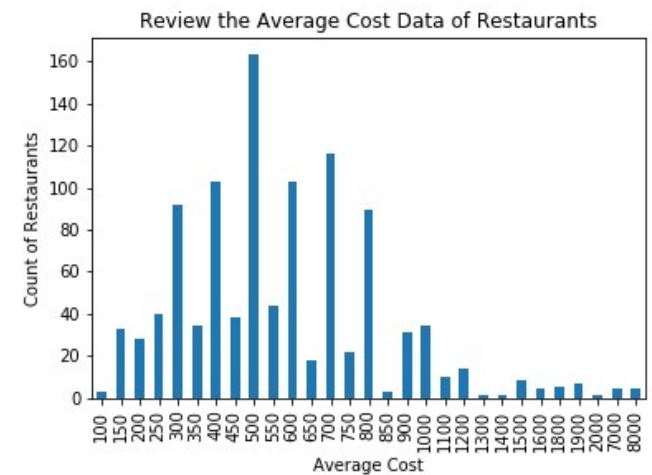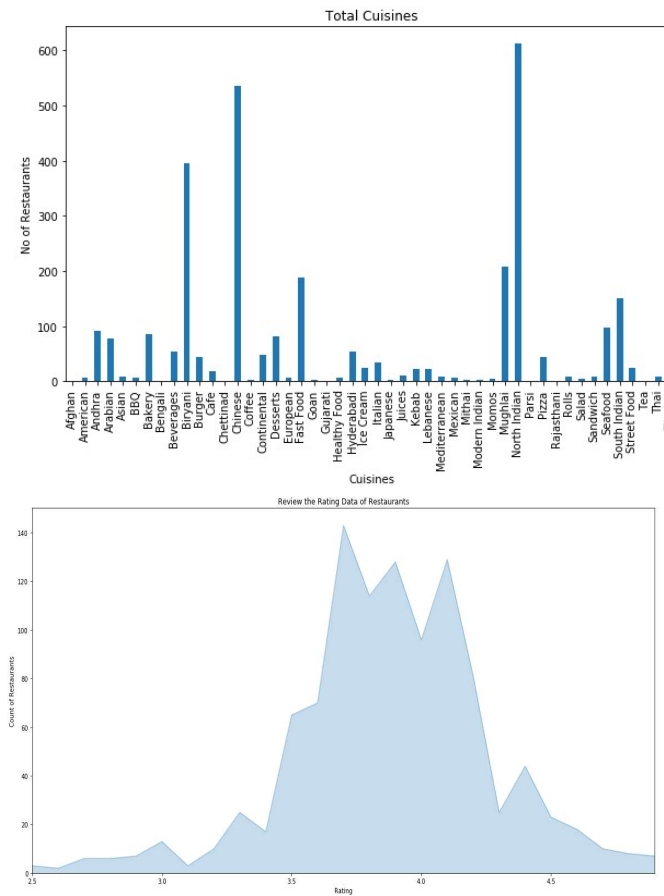## Exploratory Data Analysis

Hyderabad Restaurants from Zomato API

- Using Zomato API sourced data for further Analysis.

Methodology
Exploratory
Data Analysis

Additional Insights
from Zomato API

Total Cuisines

Review the Average Cost Data of Restaurants

Review the Rating Data of Restaurants

Methodology K-Means Clustering

- First One Hot Encoding is used to convert all the cuisines into columns which are the top 10 most common cuisines

- KMeans is fitted iteratively and using silhouette score the best K is identified.

- The Number of clusters is set to 5 and data frame is merged with original data to identify a cluster to each restaurant.

- All the clusters are reviewed and analyzed to identify insights.

- The Clusters are plotted to a color coded markers on a Hyderabad Map

Methodology
K-Means
Clustering
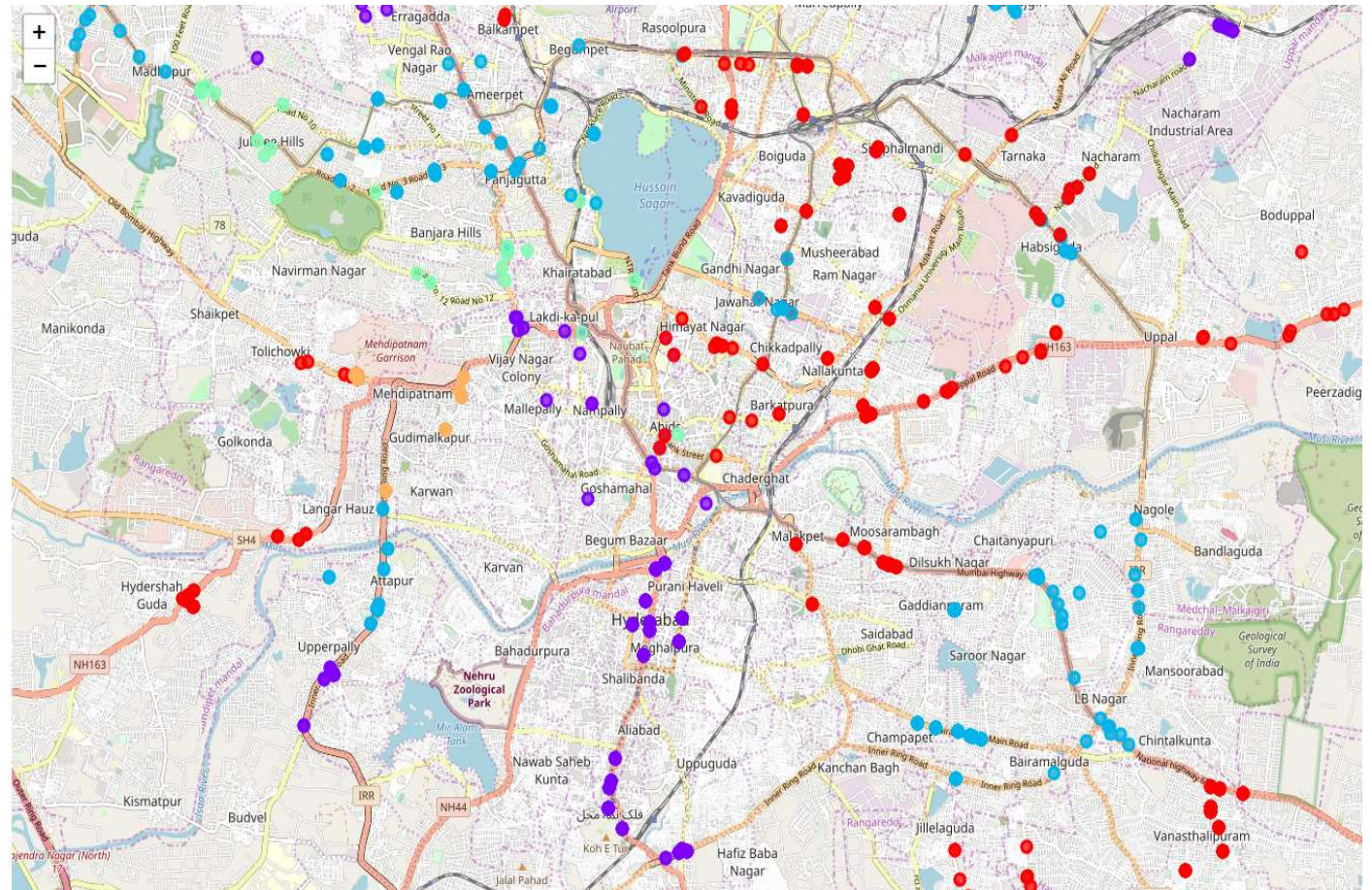
Results

- Analyze the 5 clusters to identify the Most Common and Least Common Cuisines that can used to help recommend a suitable location and cuisine for a new restaurant.

	Most Common Cuisines	Least Common Cuisines
Cluster1	North Indian, Chinese	Beverages, Burger
Cluster2	Biryani, Bakery, FastFood	Burger, Chettinad
Cluster3	North Indian, Chinese	Bengali, Continental
Cluster4	North Indian	Bengali, Biryani
Cluster5	North Indian	Healthy Food

Discussion

- The above identified cuisines and the location clusters based on the wards and neighborhoods can be used to identify a suitable combination of the cuisine and location for a restaurant recommendation

- If there is a preferred location identified the corresponding cluster of that location can be reviewed and either the least common cuisine can be opened there to improve probability of success and profits

- If there is a preferred cuisine identified all the clusters can be reviewed to identify the clusters with least number of restaurants of this cuisine and choose neighborhood\ward that can improve probability of success and profits.

Conclusion

- The Conclusion is that we are able to build a recommendation system for restaurants based on cuisines and location

- But with limited data and a simplistic model. There is a lot of scope of improvement and improvisation on this model.

- We have only sourced top 10 restaurants from Zomato; and there are opportunities of getting all of the location which can go to 100+ per neighborhood.

- Additionally Zomato API has availability of rating, cost, voters, and details of restaurants like delivery, dine in offers etc., which all can help provide additional details for more complex data model and more comprehensive recommendation model.

- Plan to continue this project to try and elaborate this into a more complete restaurant recommendation system, maybe even with some fixed parameters to improve accuracy and maybe also bring in regression and prediction.