

# Business venture to maximize revenue



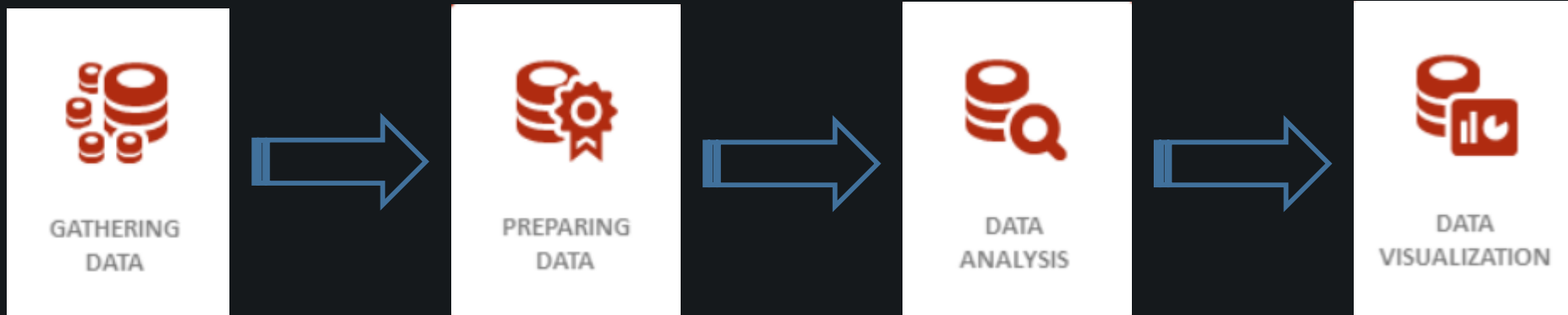
## Problem Statement

- In any over-crowded city, coming up with a food venture is like a double-edged sword.
- On one hand we see a high number of restaurants and eateries which would mean a high competition and on the other hand an increasing population would mean an opportunity to venture into a profitable business.
- How can data science help in understanding the market and give advice on the right venture to get into ?



# Approach

- Following block diagram displays the different sections in the data analysis.



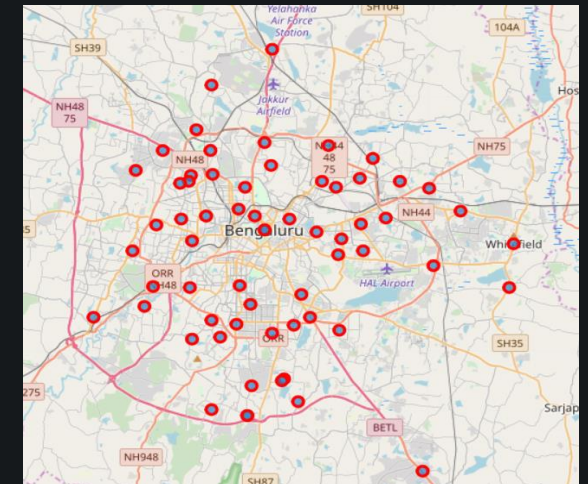


GATHERING  
DATA

- Scrape through publicly available Wiki page to obtain the neighborhood and region data.
- Use geopy to identify the geographical co-ordinates of each of the location.
- User FourSquare API to collect nearby venue details for each of the location.
- Below screenshots displays the details of the neighborhood and its corresponding plot on a city map.

Neighbourhood	
Region	
Central	11
Eastern	8
North-Eastern	6
Northern	8
South-Eastern	7
Southern	8
Southern suburbs	6
Western	9

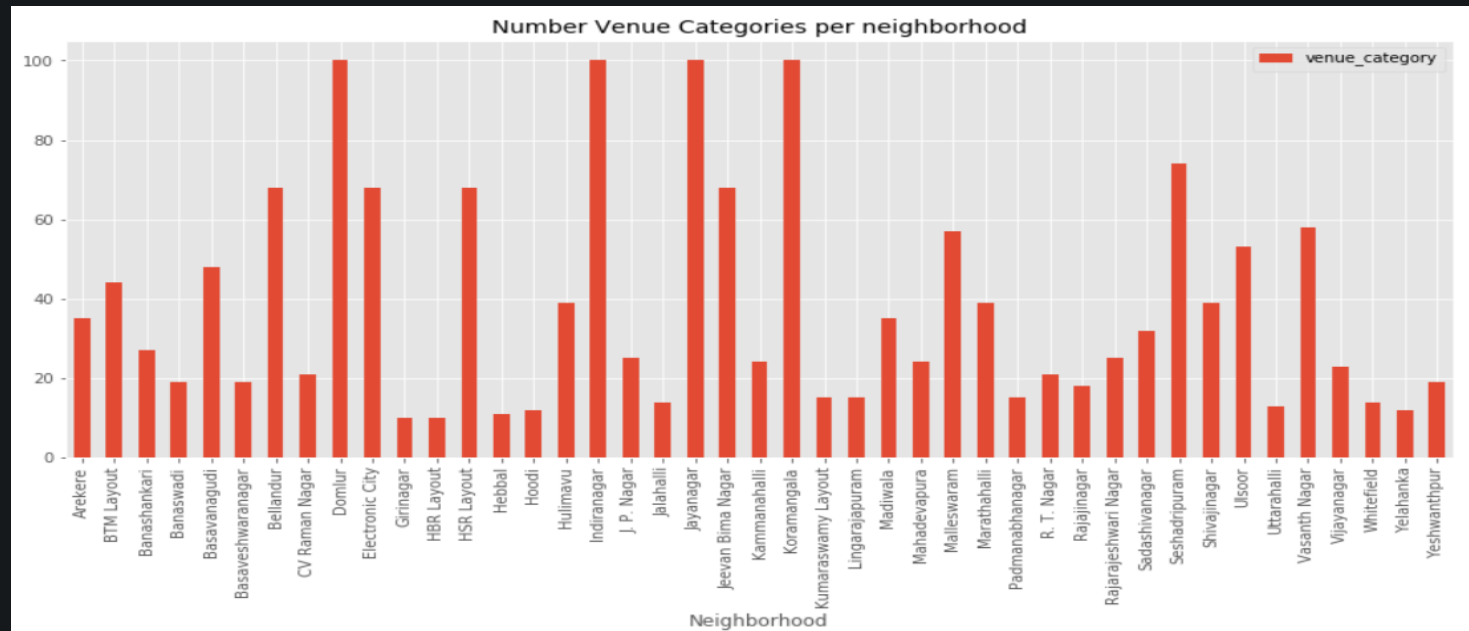
	Neighbourhood	Region	latitude	longitude
0	Cantonment area	Central	13.0196	77.5096
1	Domlur	Central	12.9625	77.6382
2	Indiranagar	Central	12.9733	77.6405
3	Malleswaram	Central	13.0163	77.5587
4	Rajajinagar	Central	12.9882	77.5549





PREPARING  
DATA

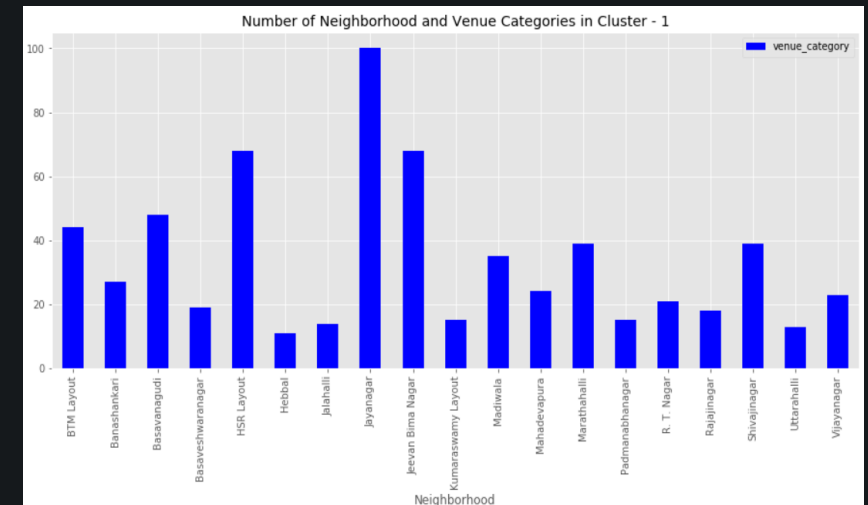
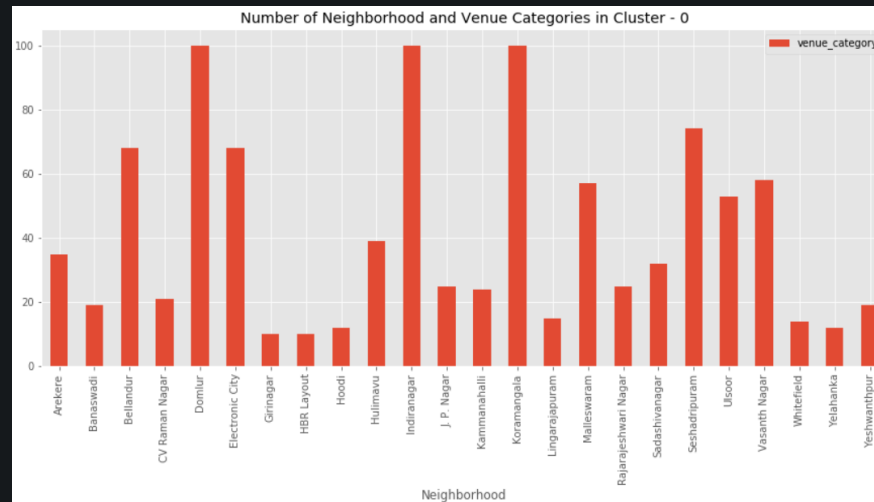
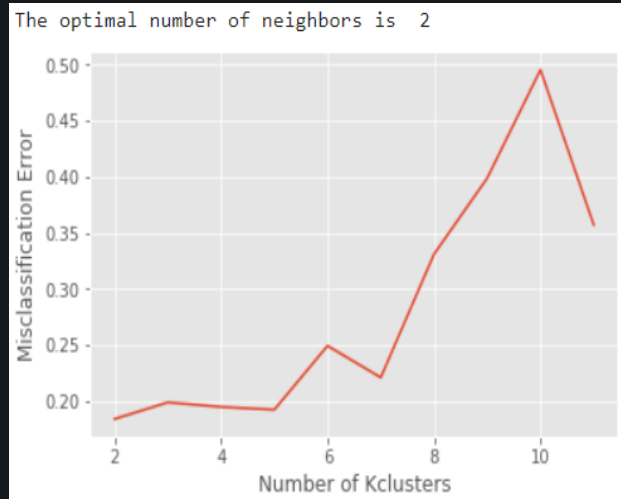
- Merge data collected from Wiki page and Four square API.
- Review and correct the co-ordinates for the location.
- Clean data with neighborhoods having less than 10 venues.
- Identify unique venue categories for the entire dataset
- Following screenshot displays the number of venue categories post data clean up.





DATA  
ANALYSIS

- Use knn algorithm to identify optimal cluster count with highest accuracy.
- KNeighborsClassifier and cross\_val\_score from sklearn.model\_selection was used to estimate the errors with each of the cluster values.
- Following screenshots displays the error rates per cluster value and neighborhood clustering post applying the kmeans algorithm on the dataset.







## Takeaways / Observations

- Neighborhoods in both the clusters have a high presence
  - **Indian Restaurants**
  - **Café**
- Restaurant catering to other cuisines are very less.
- Categories related to Ice-creams, Pizza are moderately available in each of the clusters.