Business venture to maximize revenue

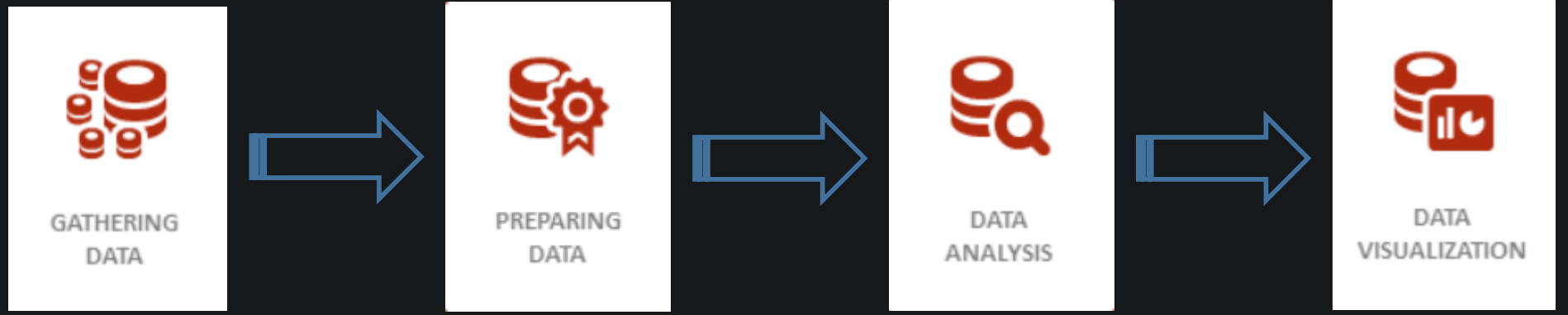
Problem Statement

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

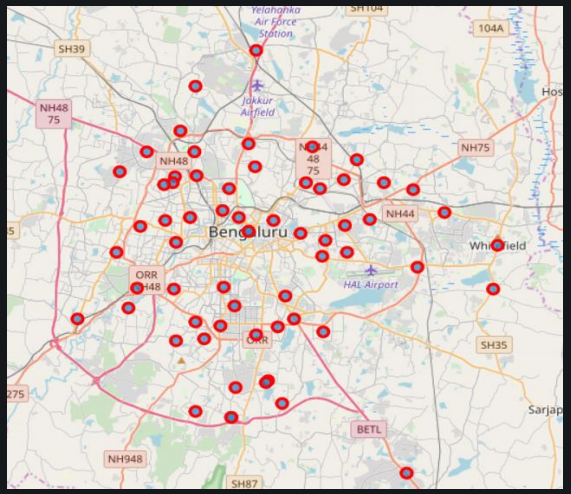

GATHERING DATA
PREPARING DATA
DATA ANALYSIS
DATA VISUALIZATION

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.

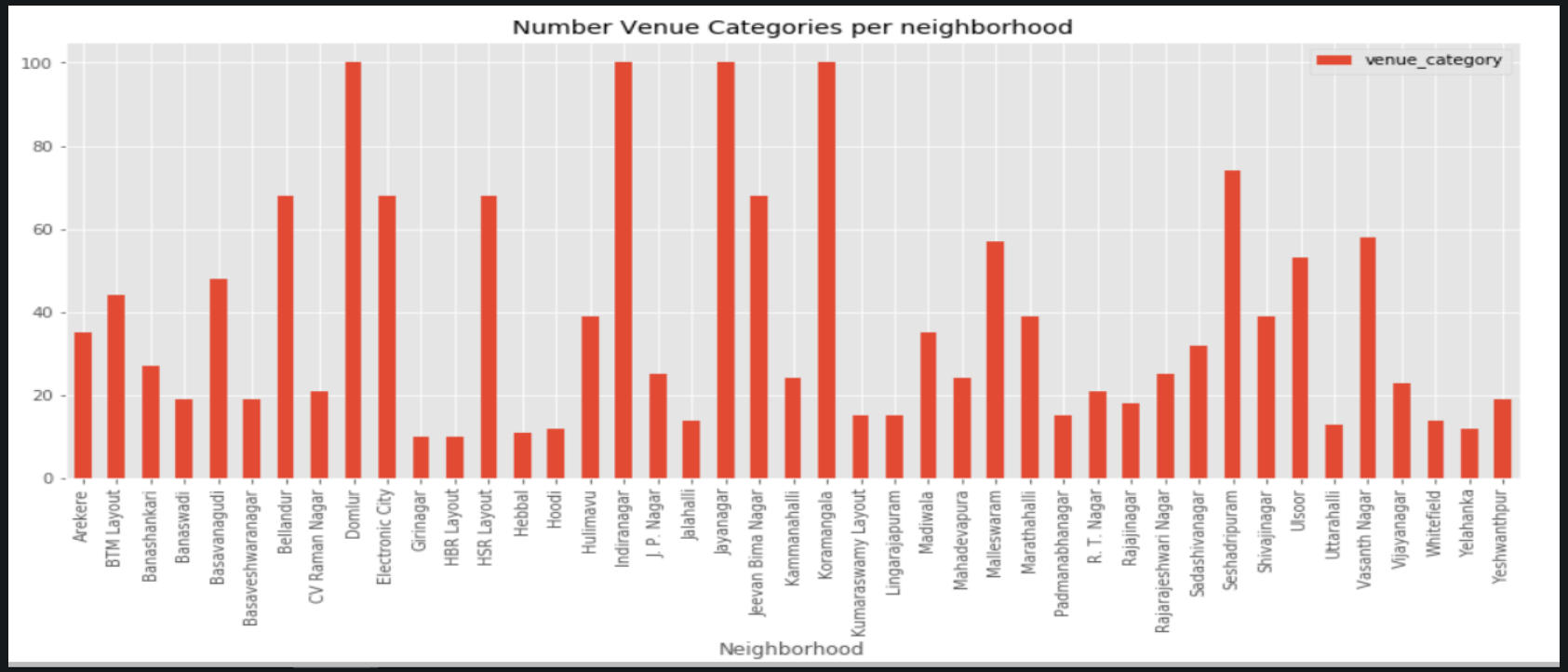
Region	Neighbourhood
Central	11
Eastern	8
North-Eastern	6
Northern	8
South-Eastern	7
Southern	8
Southern suburbs	6
Western	9

	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



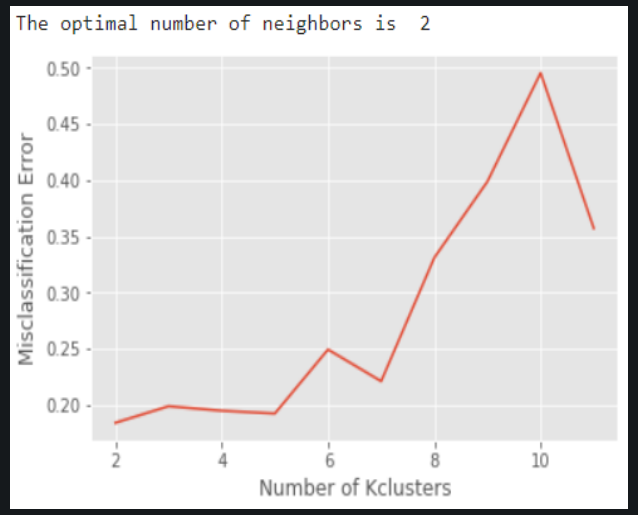
PREPARING DATA

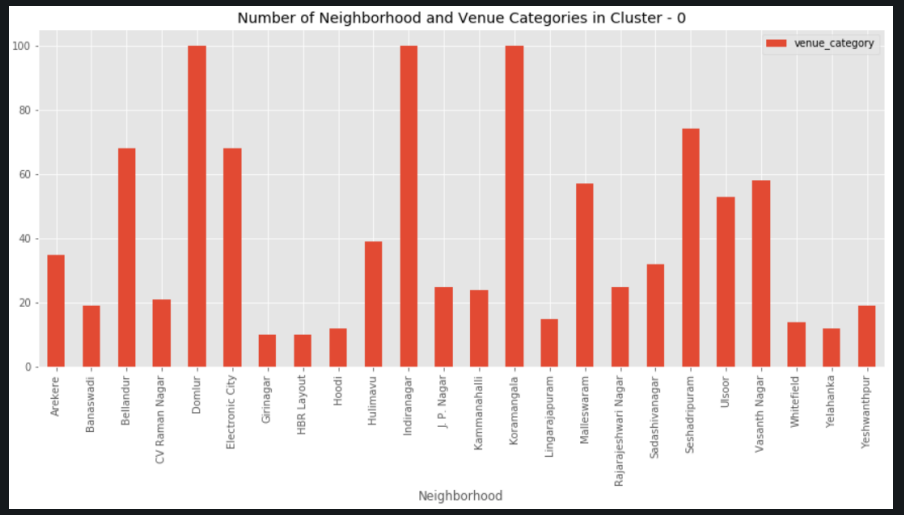
- Merge data collected from Wiki page and Four square API.
- Review and correct the co-ordinates for the location.
- Clean data with neighborhoods having less than 10 venues.
- Identify unique venue categories for the entire dataset

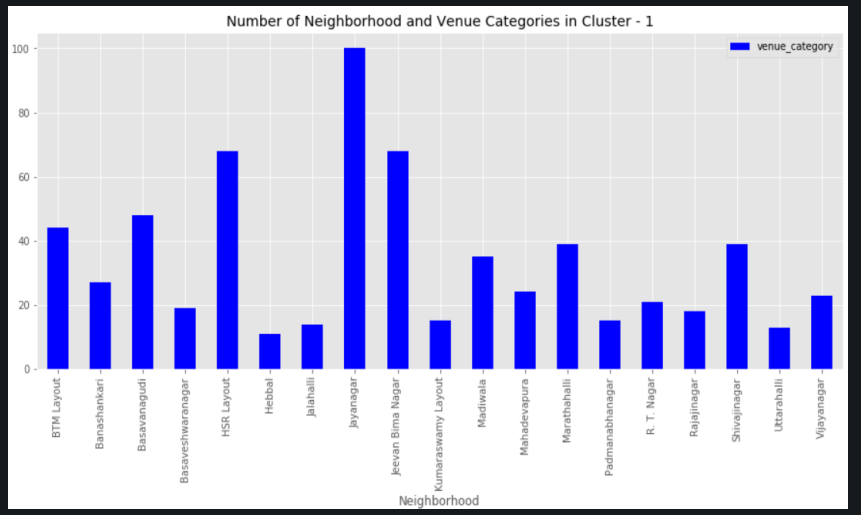


DATA ANALYSIS

- 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.







DATA VISUALIZATION

- Visualize data via maps, bar charts and word clouds to identify venue categories popular in a particular cluster