The Battle of Neighborhoods – Vizag

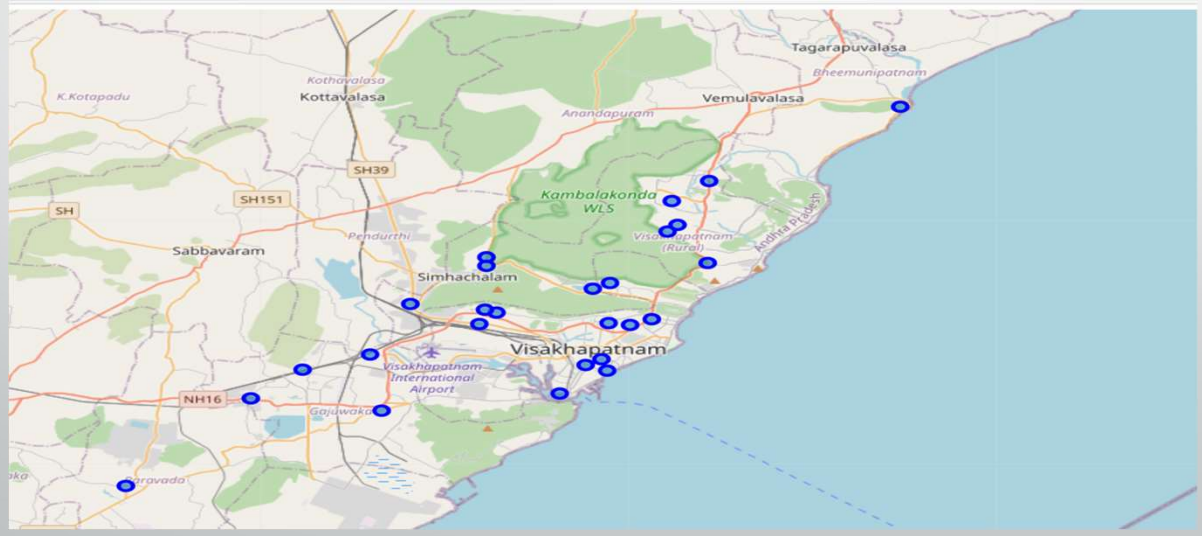
Applied Data Science by IBM on Coursera

-Suresh Malla

Introduction: Business Problem

- This project deals with major venue categories in the neighborhoods of Vizag, the proposed executive capital of the Indian state of Andhra Pradesh. It is most populated and one of the largest cities in state of Andhra Pradesh. As the city drawing an attention in recent times, this Project would specifically help investors to choose right business to start.

- Vizag



Data Requirement

- Vizag has multiple localities to explore. Extracted all the localities by web scrapping Wikipedia page which has list of localities in Vizag. Used geocoder to get coordinates of longitude and latitude for each locality.
- In order to obtain the venue details in each neighborhood Foursquare API is used.
- **Wiki page:**

  https://commons.wikimedia.org/wiki/Category:Suburbs_of_Visakhapatnam"
- **Foursquare API:**

  https://api.foursquare.com/v2/venues/explore?

Data points fetched from wiki and Foursquare API.

```
neig_df['Locality']
```

```
0                          Adavivaram
1                          Aganampudi
2                              Arilova
3                       Bakkannapalem
4                     Bheemunipatnam
5                           Boyapalem
6                              Duvvada
7                            Gajuwaka
8                  Jagadamba Centre
9                              Kommadi
10                       Maddilapalem
11                       Madhavadhara
12                        Madhurawada
13                       Maharanipeta
14                       Marikavalasa
15                         Marripalem
16         Midhilapuri VUDA Colony
17                       Mudasarlova
18                         MVP Colony
19                         Naiduthota
20                           Parawada
21        Port Area, Visakhapatnam
22                      Seethammadhara
23     Sheelanagar, Visakhapatnam
24                        Simhachalam
25       Suryabagh, Visakhapatnam
26                           Yendada
Name: Locality, dtype: object
```

```
print(Visakhapatnam_venues.shape)
Visakhapatnam_venues.head()
```

(62, 7)

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Aganampudi	17.68904	83.13988	Domino's Pizza	17.686542	83.139156	Pizza Place
1	Arilova	17.76843	83.31107	GSN PET CLINIC	17.765692	83.312883	Pet Service
2	Bakkannapalem	17.80368	83.33904	Ravi's	17.804363	83.342161	Smoke Shop
3	Bakkannapalem	17.80368	83.33904	ravi ka thota	17.807009	83.339605	Campground
4	Bakkannapalem	17.80368	83.33904	Apollo Pharmacy	17.804428	83.342990	Pharmacy

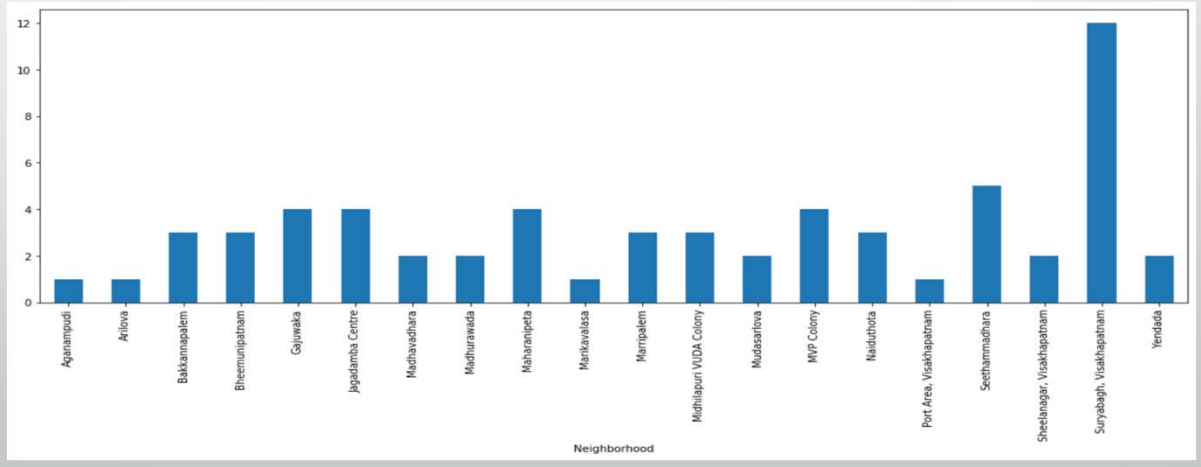
Methodology

- Now we have the neighborhoods data of Vizag (26 neighborhoods). We also have the most popular venues in each neighborhood obtained by using Foursquare API. A total of 62 venues have been obtained in whole city and 37 unique categories. Observation here is data with venues for each neighborhood is very less which varies from 1-12 venues.
- Perform one-hot encoding on the dataset and use it finding the 10 most common venue category in each neighborhood.
- K- Nearest Neighbor clustering technique have been used to find optimum number of clusters.

- Each cluster is analyzed to find major type of venue categories in each cluster.
- This data can be used to suggest investors to start suitable business and location based on the category.

ANALYSIS

- Looking into the dataset, venues returned are varying from 1- 12 range. We can see there is not enough data that foursquare API return which could lead to in accurate results. But for this capstone, I am proceeding with the same data.



Exploring the Neighborhood in all the localities with groupby

```
vskp_grouped = vskp_onehot.groupby('Neighborhood').mean().reset_index()
vskp_grouped
```

	Neighborhood	ATM	Bakery	Bookstore	Bus Station	Bus Stop	Campground	Chinese Restaurant	Clothing Store	Coffee Shop	...	Park	Pet Service	Pet Store	Pharmacy	Pizza Place	Rest
0	Aganampudi	0.00	0.00	0.0	0.00	0.0	0.000000	0.00	0.000000	0.00	...	0.0	0.0	0.0	0.000000	1.0	
1	Arilova	0.00	0.00	0.0	0.00	0.0	0.000000	0.00	0.000000	0.00	...	0.0	1.0	0.0	0.000000	0.0	
2	Bakkannapalem	0.00	0.00	0.0	0.00	0.0	0.333333	0.00	0.000000	0.00	...	0.0	0.0	0.0	0.333333	0.0	
3	Bheemunipatnam	0.00	0.00	0.0	0.00	0.0	0.000000	0.00	0.000000	0.00	...	0.0	0.0	0.0	0.000000	0.0	
4	Gajuwaka	0.00	0.00	0.0	0.00	0.2	0.000000	0.00	0.000000	0.00	...	0.0	0.0	0.0	0.200000	0.2	
5	Jagadamba Centre	0.00	0.00	0.2	0.00	0.0	0.000000	0.00	0.000000	0.00	...	0.0	0.0	0.2	0.000000	0.0	
6	MVP Colony	0.00	0.00	0.0	0.25	0.0	0.000000	0.00	0.000000	0.00	...	0.0	0.0	0.0	0.000000	0.0	
7	Madhavadhara	0.25	0.25	0.0	0.00	0.0	0.000000	0.00	0.000000	0.00	...	0.0	0.0	0.0	0.000000	0.0	
8	Madhurawada	0.00	0.00	0.0	0.00	0.0	0.000000	0.00	0.000000	0.00	...	0.0	0.0	0.0	0.500000	0.0	
9	Maharanipeta	0.00	0.00	0.0	0.00	0.0	0.000000	0.00	0.000000	0.00	...	0.0	0.0	0.0	0.000000	0.0	
10	Marikavalasa	0.00	0.00	0.0	0.00	0.0	0.000000	0.00	0.000000	0.00	...	0.0	0.0	0.0	0.000000	0.0	
11	Marripalem	0.50	0.00	0.0	0.00	0.0	0.000000	0.00	0.000000	0.00	...	0.0	0.0	0.0	0.500000	0.0	
12	Midhilapuri VUDA Colony	0.25	0.25	0.0	0.00	0.0	0.000000	0.00	0.000000	0.00	...	0.0	0.0	0.0	0.250000	0.0	
13	Mudasarlova	0.00	0.00	0.0	0.00	0.0	0.000000	0.00	0.000000	0.00	...	0.5	0.0	0.0	0.000000	0.0	
14	Naiduthota	0.00	0.00	0.0	0.00	0.0	0.000000	0.00	0.000000	0.00	...	0.0	0.0	0.0	0.000000	0.0	
15	Port Area, Visakhapatnam	0.00	0.00	0.0	0.00	0.0	0.000000	0.00	0.000000	0.00	...	0.0	0.0	0.0	0.000000	0.0	
16	Seethammadhara	0.00	0.00	0.0	0.00	0.0	0.000000	0.25	0.000000	0.25	...	0.0	0.0	0.0	0.000000	0.0	
17	Sheelanagar, Visakhapatnam	1.00	0.00	0.0	0.00	0.0	0.000000	0.00	0.000000	0.00	...	0.0	0.0	0.0	0.000000	0.0	
18	Suryabagh, Visakhapatnam	0.00	0.00	0.0	0.00	0.0	0.000000	0.00	0.090909	0.00	...	0.0	0.0	0.0	0.000000	0.0	
19	Yendada	0.00	0.00	0.0	0.00	0.0	0.000000	0.00	0.000000	0.00	...	0.0	0.0	0.0	0.000000	0.0	

20 rows × 39 columns

Apply K-Means to find the clusters

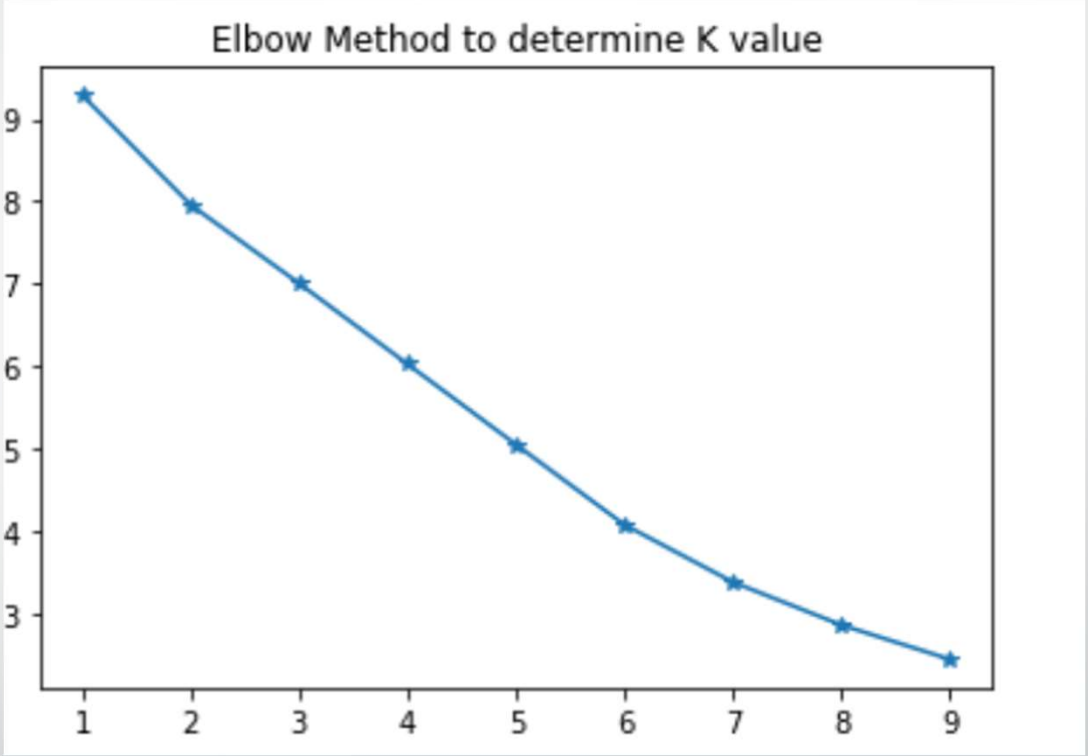
Elbow Method to find optimum number of K value

K- Nearest Neighbor clustering technique have been used to find optimum number of clusters.

Each cluster is analyzed to find major type of venue categories in each cluster.

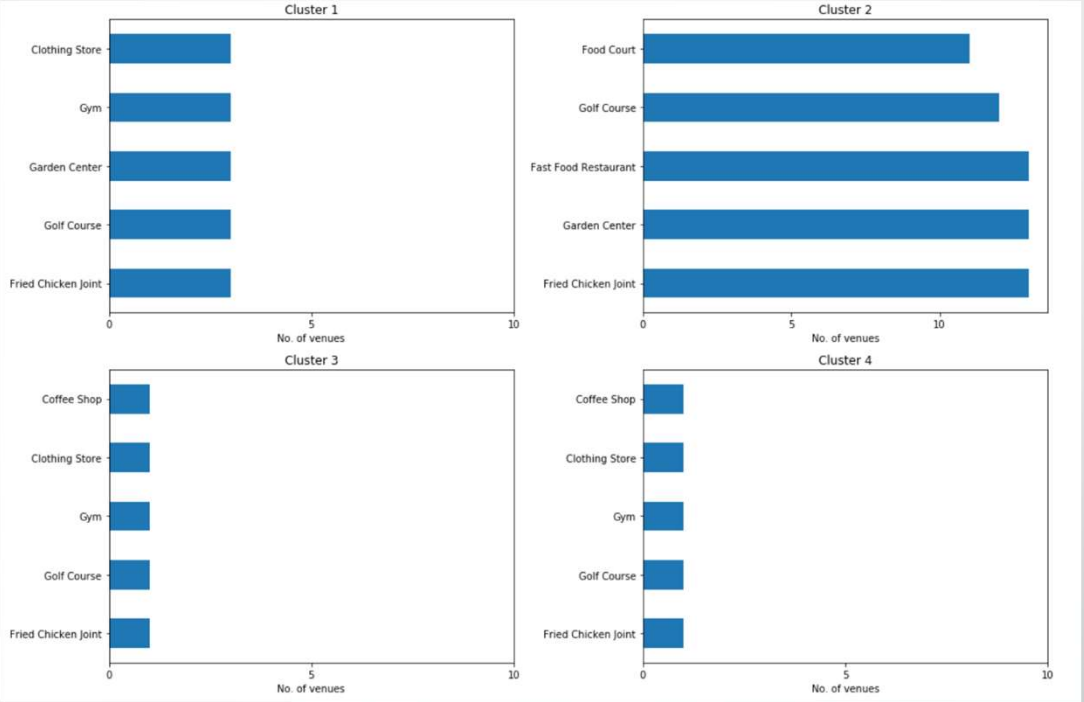
This data can be used to suggest investors to start suitable business and location based on the category.

A range of values from 1 to 10 was considered, KNN clustering was performed on the dataset and plotted a elbow plot. From the elbow plot we can see that a value of k-value of 6 provides the best score. This k-value is used for K-means clustering technique-means labels obtained were included in the top neighborhoods dataset for examining the characteristics of each cluster.



Results and Discussion

Using the clusters and top venue categories lets visualize the top 5 venue category in each cluster.



Cluster formed in City of Vizag
Top-3 Clusters

	Latitudes	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
3	17.80368	Pharmacy	Campground	Trail	Diner	Golf Course	Garden Center	Fried Chicken Joint	Food Court	Fast Food Restaurant
11	17.74773	Gym	Bakery	ATM	Restaurant	Coffee Shop	Garden Center	Fried Chicken Joint	Food Court	Fast Food Restaurant
12	17.80817	Pharmacy	Trail	Coffee Shop	Golf Course	Garden Center	Fried Chicken Joint	Food Court	Fast Food Restaurant	Electronics Store
15	17.74051	Pharmacy	Coffee Shop	Golf Course	Garden Center	Fried Chicken Joint	Food Court	Fast Food Restaurant	Electronics Store	Diner
16	17.75005	Bakery	Garden Center	Pharmacy	Diner	Golf Course	Fried Chicken Joint	Food Court	Fast Food Restaurant	Electronics Store
23	17.71927	Diner	Grocery Store	Golf Course	Garden Center	Fried Chicken Joint	Food Court	Fast Food Restaurant	Electronics Store	Coffee Shop

	Latitudes	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
4	17.889380	Trail	Coffee Shop	Golf Course	Garden Center	Fried Chicken Joint	Food Court	Fast Food Restaurant	Electronics Store	Diner
7	17.680830	IT Services	Pizza Place	Pharmacy	Bus Stop	Trail	Diner	Garden Center	Food Court	Fast Food Restaurant
8	17.715800	Bookstore	Hotel	Pet Store	Electronics Store	Trail	Diner	Garden Center	Fried Chicken Joint	Food Court
13	17.708290	Hotel	Indian Restaurant	Food Court	Grocery Store	Golf Course	Garden Center	Fried Chicken Joint	Fast Food Restaurant	Electronics Store
17	17.764530	Golf Course	Coffee Shop	Garden Center	Fried Chicken Joint	Food Court	Fast Food Restaurant	Electronics Store	Diner	Trail
18	17.743690	Ice Cream Shop	Indian Restaurant	Food Court	Trail	Electronics Store	Golf Course	Garden Center	Fried Chicken Joint	Fast Food Restaurant
19	17.754101	Indian Restaurant	Trail	Diner	Golf Course	Garden Center	Fried Chicken Joint	Food Court	Fast Food Restaurant	Electronics Store
22	17.740670	Indian Restaurant	Chinese Restaurant	Fast Food Restaurant	Diner	Golf Course	Garden Center	Fried Chicken Joint	Food Court	Electronics Store
25	17.711910	Hotel	Indian Restaurant	Indie Movie Theater	Movie Theater	Fast Food Restaurant	Clothing Store	Fried Chicken Joint	Food Court	Electronics Store
26	17.782230	Diner	Indian Restaurant	Grocery Store	Bus Station	Bus Stop	Campground	Chinese Restaurant	Clothing Store	Gym

	Latitudes	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	17.68934	Trail	Coffee Shop	Golf Course	Garden Center	Fried Chicken Joint	Food Court	Fast Food Restaurant	Electronics Store	Diner

Conclusion:

After going through the neighborhoods of Visakhapatnam, India and looking the cluster information, cluster1 is already occupied with good number of restaurants, grocery, food and women store but see an opportunity for someone to start off with a gym which is missing in the cluster. Whereas starting with restaurant, clothing business as an opportunity to start in cluser2,3,4 considering the less competition and uniqueness of the business.

The main challenge here with the analysis is Foursquare API has returned very less data points and this can be improved by trying with other API's which gives better data for our analysis to recommend business ideas in Vizag location.