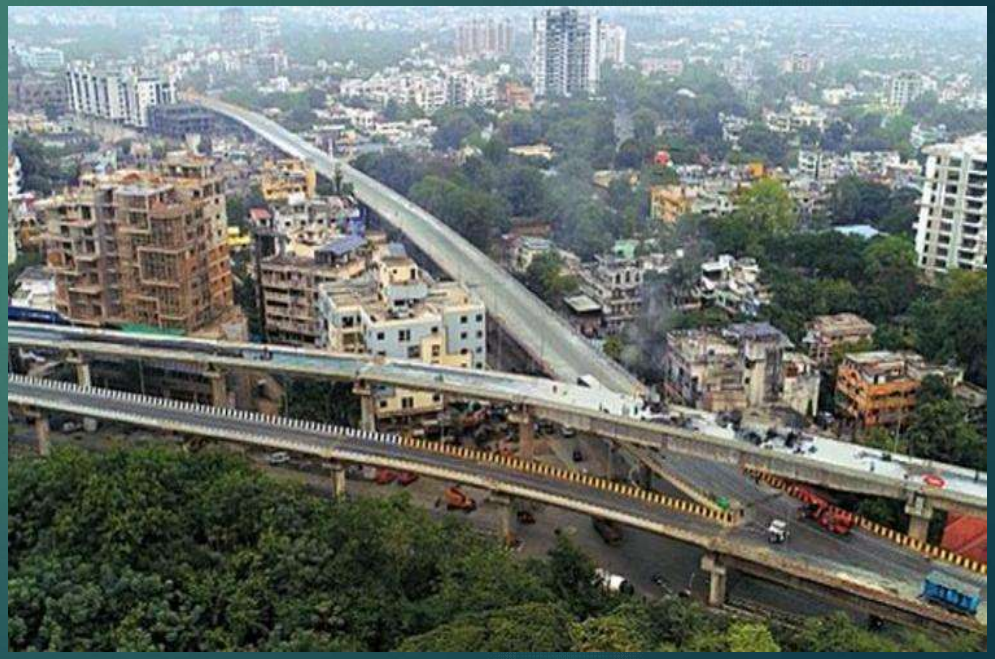
Clustering Localities of Nagpur



Introduction

Nagpur is the third largest city and the winter capital of the Indian state of Maharashtra and it is popularly known as the **Orange City** of India. It is the 13th largest city in India by population. There are around 42 localities in Nagpur. Each localities have their own popular venues and thus if we find similar localities based on their popular venues, it will help people who want to transfer to different localities in order to being close to their work offices.

Problem Statement

The main goal of this project is to cluster the localities of Nagpur based on their similarities. So if a person who want to shift to new locality as he/she wants to being as close as possible to their work offices but they also wants to shift in a locality where there are similar popular venues like cafes, restaurants etc. as their old localities.

Target Audience

So the target audience are the peoples who want to transfer to another localities in order to being as close as possible to their work offices with the similar popular venues like cafes, restaurants etc. as their old ones.

Outcome

The outcome of this project would be a good recommendation of a locality where a person can shift in order to get as close as possible to work office with the similar locality in terms of popular venues.

Data

Firstly we will fetch the localities of Nagpur. We can get this localities using the Wikipedia with the help of Web Scraping. We will scrape the webpage using the BeautifulSoup library of Python. Next we will need the coordinates of each localities that we can get using Geopy library of Python. Then at last we will pass this coordinates into the FourSquare API to fetch the nearby venues of each localities.

Required Libraries

- Pandas (For data analysis & data wrangling)
- Numpy (For handling data in a vectorized manner)
- JSON (For handling JSON files)
- Geopy (For converting an address into latitude and longitude values)
- Plotly (Data Visualization)
- Folium (Map Rendering)
- Sklearn (For using Unsupervised Clustering method namely K-Means)
- BeautifulSoup (For Web Scraping data from Wikipedia)

Methodology

- The main aim of this project is to find the new locality for the peoples who want to shift to new locality in order to get close to their work office and they also want a similar surroundings in terms of popular venues as their old locality
- Now in this project we are analyzing the localities of Nagpur city of Maharashtra and based on their similarities in terms of popular venues we are grouping them into clusters so the person who want to shift his/her residence to new locality can select the localities that are in the same cluster according to their convenience in terms of which is near to their work offices.

Exploratory Data Analysis - I

- We have the data of localities in Nagpur in a Wikipedia web page. So at first, we are going to perform web scraping of Wikipedia page in order to get the localities of Nagpur. We can do the web scraping using the BeautifulSoup library of Python.
- We are going to store the localities of Nagpur in Pandas Dataframe. During this process of storing in it we are also going to perform data cleaning where we are going to remove unwanted citations and information about the localities.
- After that, we are going to fetch the Geolocation of the localities in terms of Latitude and Longitudes which we will be used later to fetch nearby popular venues.
- We are going to perform this extraction of Geolocation using the Geopy Library of Python.

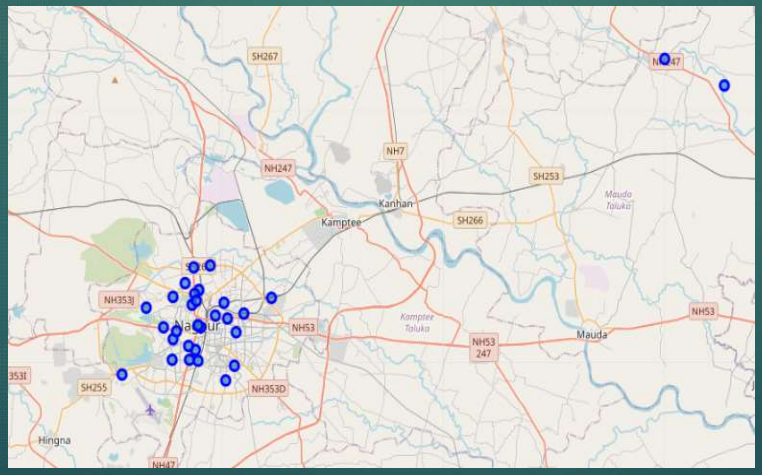
Exploratory Data Analysis - II

- Then we are going to show the map with the indication of different localities of Nagpur which we can perform using the Folium Library of Python.
- We are going to drop that localities which are far away from the Nagpur as people won't shift there.
- Then we are going to use the FourSquare API which will give the data of nearby venues in JSON format.
- Finally we are going to use the K-Means Clustering algorithm to cluster the localities of Nagpur based on their similarities.

- The Data frame with the localities as well as their Latitudes and Longitudes are as follows:

	Locality	Latitude	Longitude
0	Mahal	21.1452	79.1124
1	Sitabuldi	21.1483	79.0843
2	Dhantoli	21.1339	79.0806
3	Itwari	21.1575	79.1188
4	Mominpura	21.1561	79.0963

- The map with the localities of Nagpur as labels is as shown below:



- The data frame with the localities along with their venues and their respective categories is as follows:

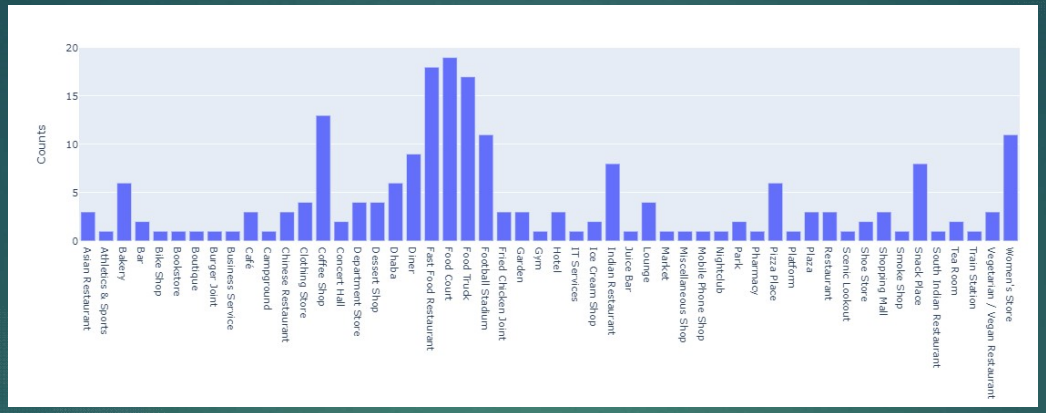
	Locality	Locality Latitude	Locality Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Sitabuldi	21.148329	79.084326	Eternity Mall	21.144854	79.082416	Shopping Mall
1	Sitabuldi	21.148329	79.084326	Naivedhyam Veg Restaurant	21.144874	79.082260	Vegetarian / Vegan Restaurant
2	Sitabuldi	21.148329	79.084326	Sitabuldi Market	21.144859	79.082403	Miscellaneous Shop
3	Sitabuldi	21.148329	79.084326	Charu Boutique	21.144923	79.082426	Women's Store
4	Sitabuldi	21.148329	79.084326	Apna Bazaar	21.144605	79.082371	Clothing Store
5	Sitabuldi	21.148329	79.084326	Ratan Plaza	21.150570	79.088102	Plaza
6	Sitabuldi	21.148329	79.084326	Ghugre Snacks Point	21.144600	79.081780	Snack Place
7	Sitabuldi	21.148329	79.084326	Big Bazaar	21.150600	79.088438	Department Store
8	Dhantoli	21.133944	79.080566	Hotel Centre Point	21.134719	79.076593	Hotel
9	Dhantoli	21.133944	79.080566	TDS	21.136808	79.078449	Restaurant
10	Dhantoli	21.133944	79.080566	Moksha	21.137272	79.078857	Lounge
11	Dhantoli	21.133944	79.080566	Big Bazaar	21.137971	79.079028	Shopping Mall

Results

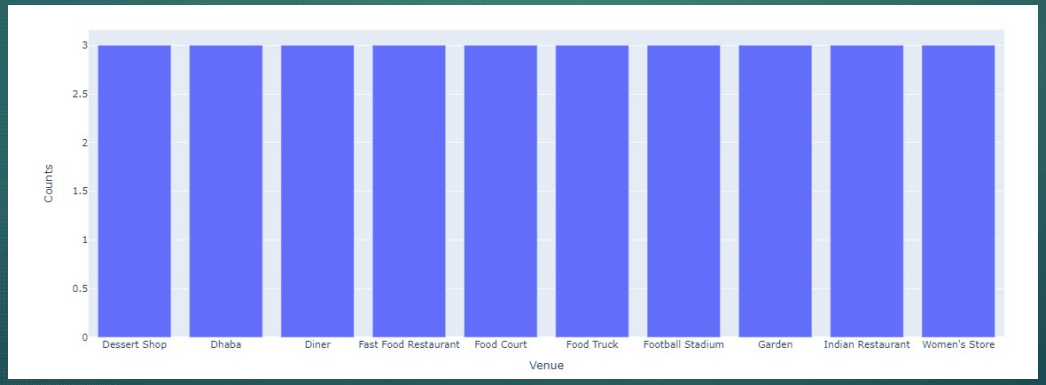
- There are 26 localities and 52 categories of nearby venues. As we can see that most of the localities falls into the Cluster 1 as they have similar category of popular venues across it. Also we have seen that other clusters i.e. Cluster 3 and Cluster 4 have only one locality in it and it's obvious because they are locality situated near hills.
- The Data frame with the cluster labels and their most common nearby venues is as follows:

	Locality	Latitude	Longitude	Cluster Labels	1st Most Common Venue	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	Sitabuldi	21.1483	79.0843	0	Women's Store	Clothing Store	Vegetarian / Vegan Restaurant	Snack Place	Shopping Mall	Department Store	Plaza	Miscellaneous Shop	Food Court	Fast Food Restaurant
2	Dhantoli	21.1339	79.0806	0	Indian Restaurant	Hotel	Shopping Mall	Juice Bar	Restaurant	Lounge	Women's Store	Coffee Shop	Food Court	Fast Food Restaurant
3	Itwari	21.1575	79.1188	0	Football Stadium	Bakery	Indian Restaurant	Fast Food Restaurant	Women's Store	Coffee Shop	Food Truck	Food Court	Diner	Dhaba
4	Mominpura	21.1561	79.0963	0	Hotel	Indian Restaurant	Shoe Store	Fried Chicken Joint	Women's Store	Coffee Shop	Food Truck	Food Court	Fast Food Restaurant	Diner
5	Dharampeth	21.141	79.0624	0	Indian Restaurant	Ice Cream Shop	Park	Shoe Store	Café	Restaurant	Market	Bookstore	Vegetarian / Vegan Restaurant	Snack Place

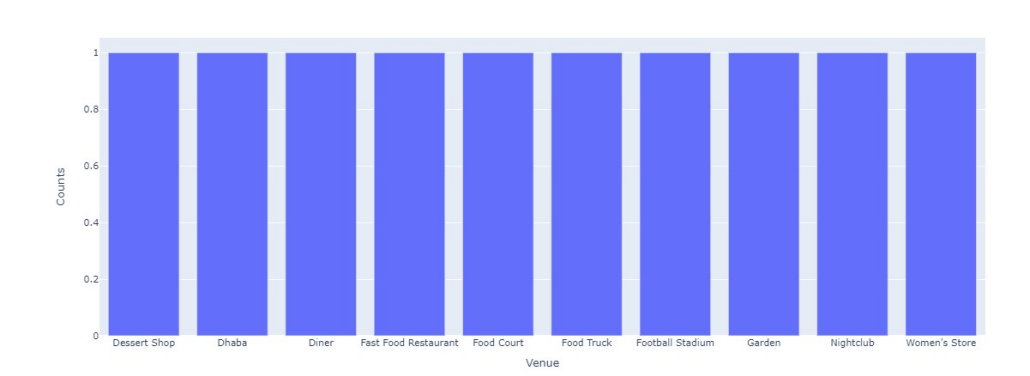
- **Cluster 1 Venues v/s Count**



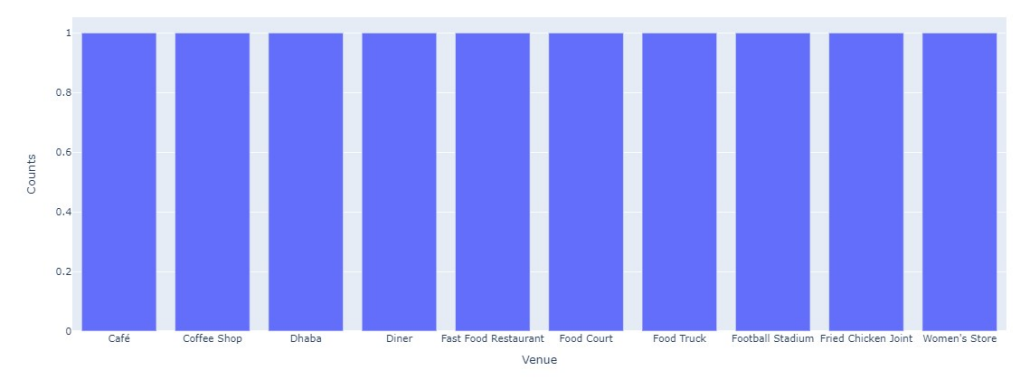
- **Cluster 2 Venues v/s Count**



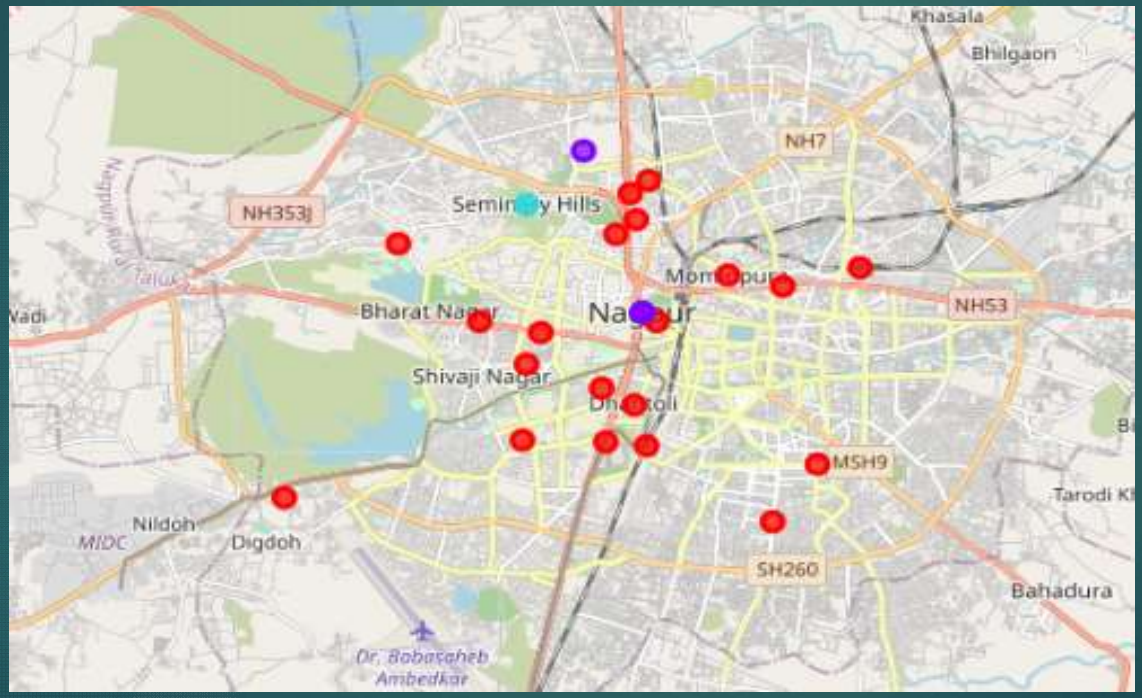
- **Cluster 3 Venues v/s Count**



- **Cluster 4 Venues v/s Count**



- The map with the different clusters of the localities of Nagpur is as follows:



Discussion

- As we can see most of the localities falls into the one cluster because most of them have common popular nearby venues.
- We can notice that some cluster have only one locality because they are located on the outer portion of the city and nearby hilly areas.
- Also we can see the clusters with the most localities in it almost contains all the categories because they are located in the inner part of the city where it is heavily populated compare to the outer regions.

Conclusion

This project is making clusters based on the limited data. So the accuracy of it can be increased by increasing the data. But by looking at the clusters, we can clearly see that the localities which are located in the inner part of Nagpur have been grouped into one cluster because they have most common nearby popular venues as mostly the inner part are densely populated compare to the outer regions. So to summarize it overly it is providing a good recommendation to a person in shifting to new locality which is similar to their old one.

Thank You!!!