IBM Applied Data Science Capstone Project – Battle of the Neighbourhoods Report

1. Introduction/Business Problem

Kochi, once in the past referred to in English as Cochin, and Arabic as Kashi, is a significant port city on the Malabar Coast of India lining the Laccadive Sea, which is a part of the Arabian Sea. It is an important part of the district of Ernakulam and is regularly alluded to as Ernakulam. Kochi is the most densely populated city in Kerala. It has a population of more than 2.1 million inside a space of 440 km², making it the biggest and the most crowded metropolitan region in Kerala. Kochi city is classified as a Tier-II city by the Government of India.

The plus points of Kochi include modern housing facilities with reliable supply to water and electricity, access to ample business opportunities, good educational facilities and transportation (including air, rail and sea). Kochi also has the largest shopping mall in Kerala, superb restaurants and theatres. But the downside of this city is its sewage facilities. During the rainy season, Kochi is submerged in water with drains overflowing and mosquitos happily breeding. Kochi, being the industrial hub of Kerala also has lower quality air compared to rest of the state. The traffic in Kochi is probably the worst thing about the city.

Kochi is a city where more young people from different parts of Kerala and even India are coming to find better opportunities and lead a better life.

In this project we will leverage infrastructure data of Kochi to find answers to the following questions:

1. Which are the locations in Kochi with the best infrastructure facilities and thus probably the best locations for people to settle down?

2. Which are the locations in Kochi with a bad infrastructure and thus the best options for investors looking for new opportunities?

We are going to use data of Kochi neighbourhood including pin codes, neighbourhood name etc obtained from the internet through web scraping for this project.

2. Data Description

We need the following datasets to solve our problem:

1. Pincode (zipcode) of the neighbourhoods of Kochi. This can be obtained from http://pincode.city/india/kerala/ernakulam/kochi.

2. Latitudes and longitudes of various neighbourhoods in Kochi. By using GeoPy https://github.com/geopy/geopy, we can find out the required coordinates.

3. Infrastructure details of neighbourhoods in Kochi. We can use Foursquare API for this purpose.

3. Methodology

3.1 Data Acquisition

We were able to acquire our required data from the sources mentioned in the Data Description section. This data was inserted into a dataframe. This data was cleaned and unnecessary components were removed.

3.2 Data Geocoding

In order to use the Foursquare API we need the latitudes and longitudes. This was obtained using GeoPy which was then populated into our dataframe.

3.3 Data Visualization

Using the Folium package, we were able to see all the neighbourhoods in the Kochi region.

3.4 Exploring the Neighbourhood

We take first neighbourhood & get the top 100 venues that are within a radius of 1km of the first neighbourhood. Then we repeat this process for all the neighbourhoods. By using one hot encoding, we find out the most frequented venues in each neighbourhood. We also find out the total infrastructure in all the neighbourhoods. The availability of various facilities like market, restaurants etc is a huge factor when it comes to settling in a location.

3.5 Clustering

The final part of this project is clustering the neighbourhoods into six clusters based on their infrastructural facilities. This will help us find out which regions are the most and the least suitable for living when it comes to infrastructure conditions. We will also use the Folium package to visualize the clusters.

4. Results

The following are the results obtained from this project:

1. Mattancherry Jetty is the best place to settle, when you look at the infrastructure facilities.
2. The worst places to settle down are Malipuram and Nayarambalam, when you look at infrastructure facilities.

In [49]:
```
quality_kochi_grouped[quality_kochi_grouped['Total infrastructure'] == quality_kochi_grouped['Total infrastructure'].min()].trans
```

Out[49]:

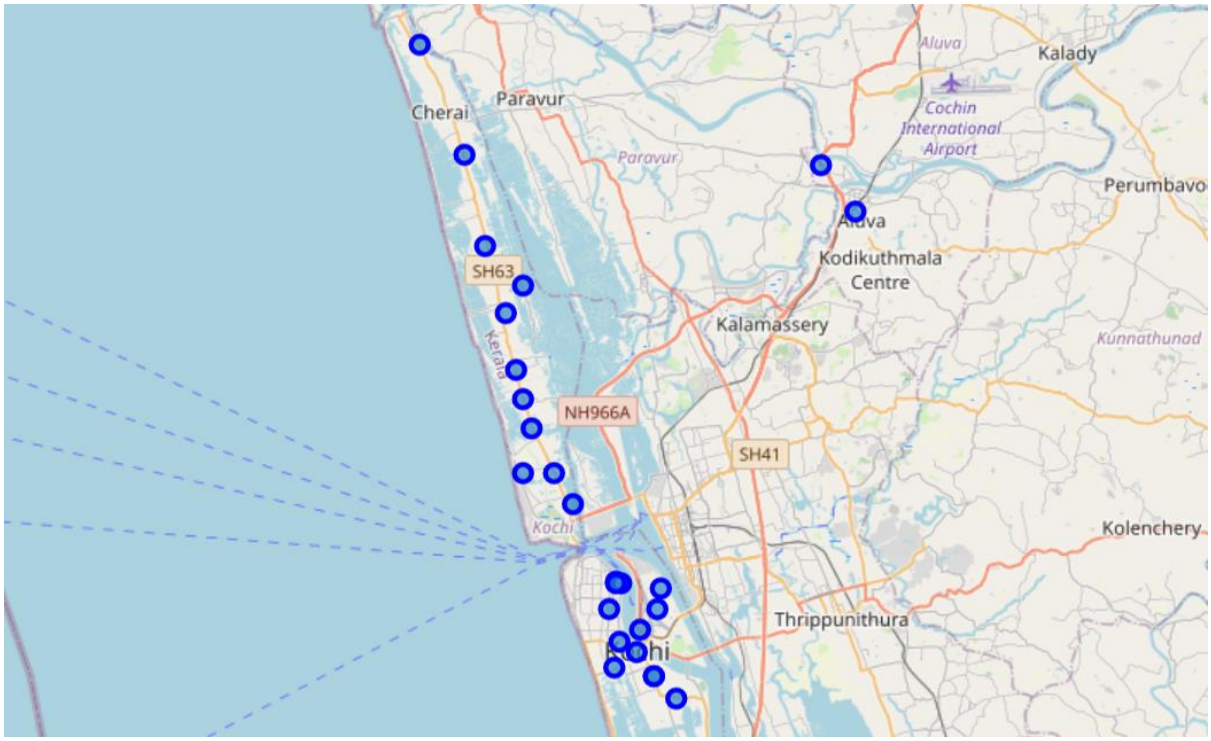
	9	16
Neighborhood	Malipuram	Nayarambalam
ATM	0	0
Airport	0	0
Antique Shop	0	0
Asian Restaurant	0	0
Auditorium	0	0
Auto Garage	0	0
Bakery	0	0
Bar	0	0
Beach	0	0
Bed & Breakfast	0	0
Boat or Ferry	0	0
Bus Line	0	0
Bus Station	0	0
Bus Stop	0	0
Café	0	0
Clothing Store	0	0
Coffee Shop	0	0
Convenience Store	0	0
Department Store	0	0
Fast Food Restaurant	0	0
Fish Market	0	0
Fishing Store	0	0
Flea Market	0	0
Food	0	0
Food & Drink Shop	0	0
Food Court	0	0
Fried Chicken Joint	0	0
Gift Shop	0	0
Gym	0	0
Halal Restaurant	0	0
Harbor / Marina	0	0
Historic Site	0	0
Hotel	0	0
Hotel Bar	0	0
Indian Restaurant	1	0
Indie Movie Theater	0	0
Jewelry Store	0	0
Kerala Restaurant	0	0
Lake	0	0
Light Rail Station	0	0
Lighthouse	0	0
Market	0	1
Mobile Phone Shop	0	0
Multiplex	0	0
Performing Arts Venue	0	0
Platform	0	0
Playground	0	0
Resort	0	0
Scenic Lookout	0	0
Shopping Mall	0	0
Smoke Shop	0	0
Spa	0	0
Vegetarian / Vegan Restaurant	0	0
Total infrastructure	1	1

In [48]:
```
quality_kochi_grouped[quality_kochi_grouped['Total infrastructure'] == quality_kochi_grouped['Total infrastructure'].max()].trans
```

Out[48]:

	12
Neighborhood	Mattancherry Jetty
ATM	0
Airport	0
Antique Shop	0
Asian Restaurant	1
Auditorium	1
Auto Garage	0
Bakery	0
Bar	0
Beach	0
Bed & Breakfast	0
Boat or Ferry	2
Bus Line	0
Bus Station	0
Bus Stop	0
Café	3
Clothing Store	0
Coffee Shop	0
Convenience Store	1
Department Store	0
Fast Food Restaurant	0
Fish Market	0
Fishing Store	0
Flea Market	0
Food	0
Food & Drink Shop	0
Food Court	0
Fried Chicken Joint	0
Gift Shop	0
Gym	0
Halal Restaurant	0
Harbor / Marina	0
Historic Site	1
Hotel	2
Hotel Bar	1
Indian Restaurant	3
Indie Movie Theater	0
Jewelry Store	0
Kerala Restaurant	0
Lake	0
Light Rail Station	0
Lighthouse	0
Market	1
Mobile Phone Shop	0
Multiplex	0
Performing Arts Venue	1
Platform	0
Playground	0
Resort	0
Scenic Lookout	0
Shopping Mall	0
Smoke Shop	0
Spa	0
Vegetarian / Vegan Restaurant	0
Total infrastructure	17

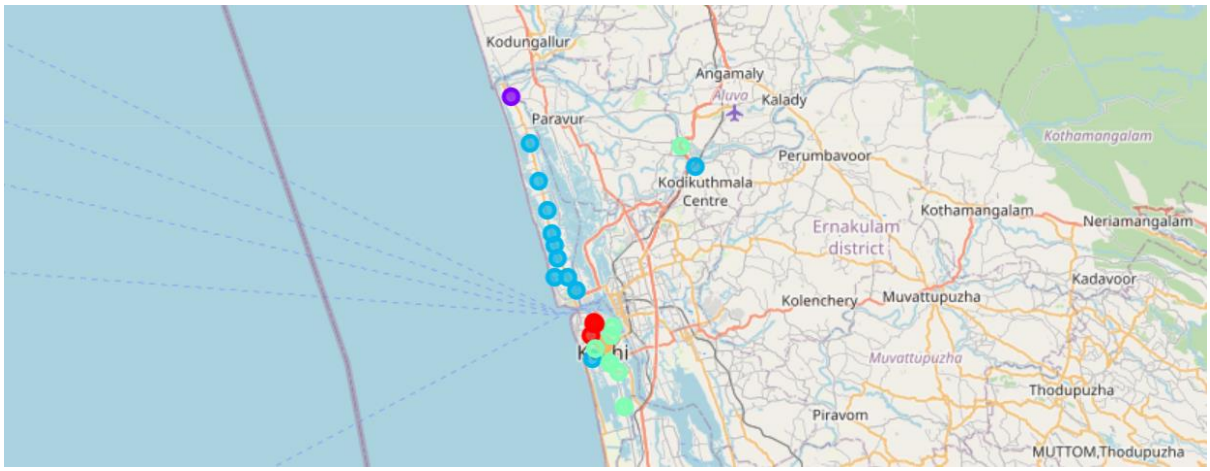
5. Discussion



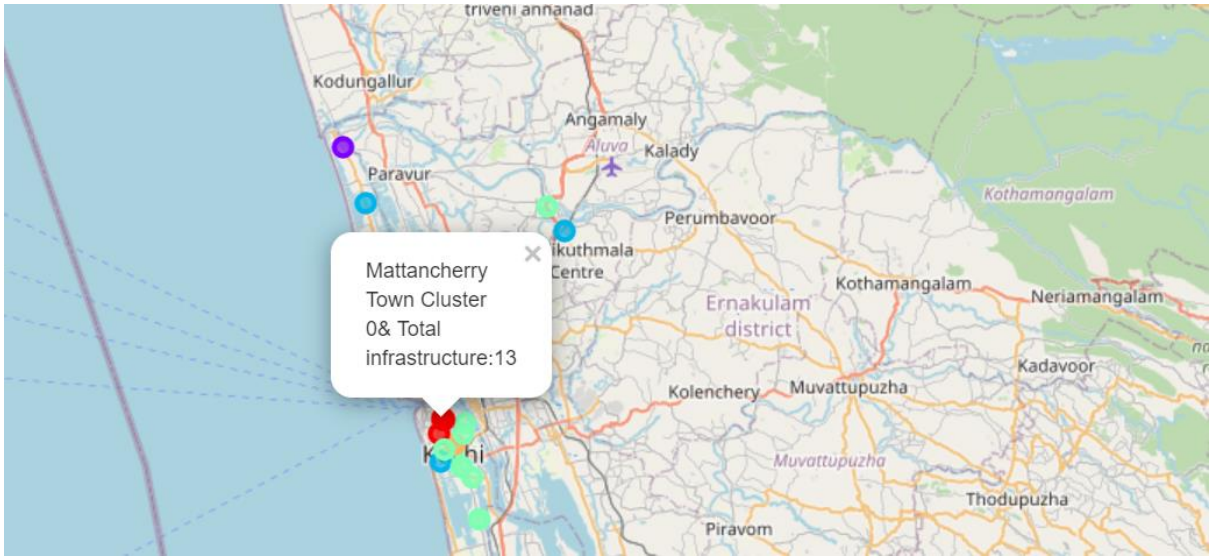
Before Clustering 1

The following observations can be made from this analysis:

1. The middle part of Kochi has the most infrastructure development and is the best place to live.
2. The southern part is the second-best choice to live.
3. The northern part is the worst choice to live.



After Clustering 1



After Clustering 2

6. Conclusion

We have utilized the methods of data scraping, data cleaning and kmeans clustering along with utilities like Foursquare API and GeoPy to find out which are the best places to live in the city of Kochi when it comes to infrastructure facilities.

There might be some anomalies in the overall analysis as Foursquare data for Kochi region is inadequate. So there is scope for further improvement.