IBM Applied Data Science Capstone

Chennai Metro Stations Clustering

By : Sanjay Kannu Periyasami

November 2020

1

Business Problem

- Due to high volume of traffic on road and very densely populated area, people often spend lot of time in traffic and avoiding traffic is main agenda for people.

- Staying near Chennai Metro is most convenient, especially in early morning while going to Office and School and other times when traffic on road is high.

- A residential location near Chennai Metro is most wanted.

- We must find out which metro station is best suited for residential purpose.

Data

- Data required
  - List of Metro Stations in Chennai
  - Latitude and longitude coordinates of Metro Stations in Chennai
  - Venue data of Metro Stations in Chennai

- Sources of data
  - Chennai Metro Rail Limited https://chennaimetrorail.org
  - Geocoder package for latitude and longitude coordinates
  - Foursquare API for venue data

Methodology

- Get latitude and longitude coordinates using Geocoder

- Use Foursquare API to get venue data

- Group data venue details of each metro station for all categories of our interest namely Food & Drink Shops, Schools, Medical Center, Fitness Center, Restaurants and Shopping Mall

- List out top 5 metro station for each venue category with number of venues

- Categorize all metro station in 3 categories (Red, Blue and Green). To categorize all metro station, we will use K-means algorithm

- Visualize the clusters in a map using Folium

Plot metro station on Chennai Map

- Using Foursquare API, plot metro station on Dubai Map
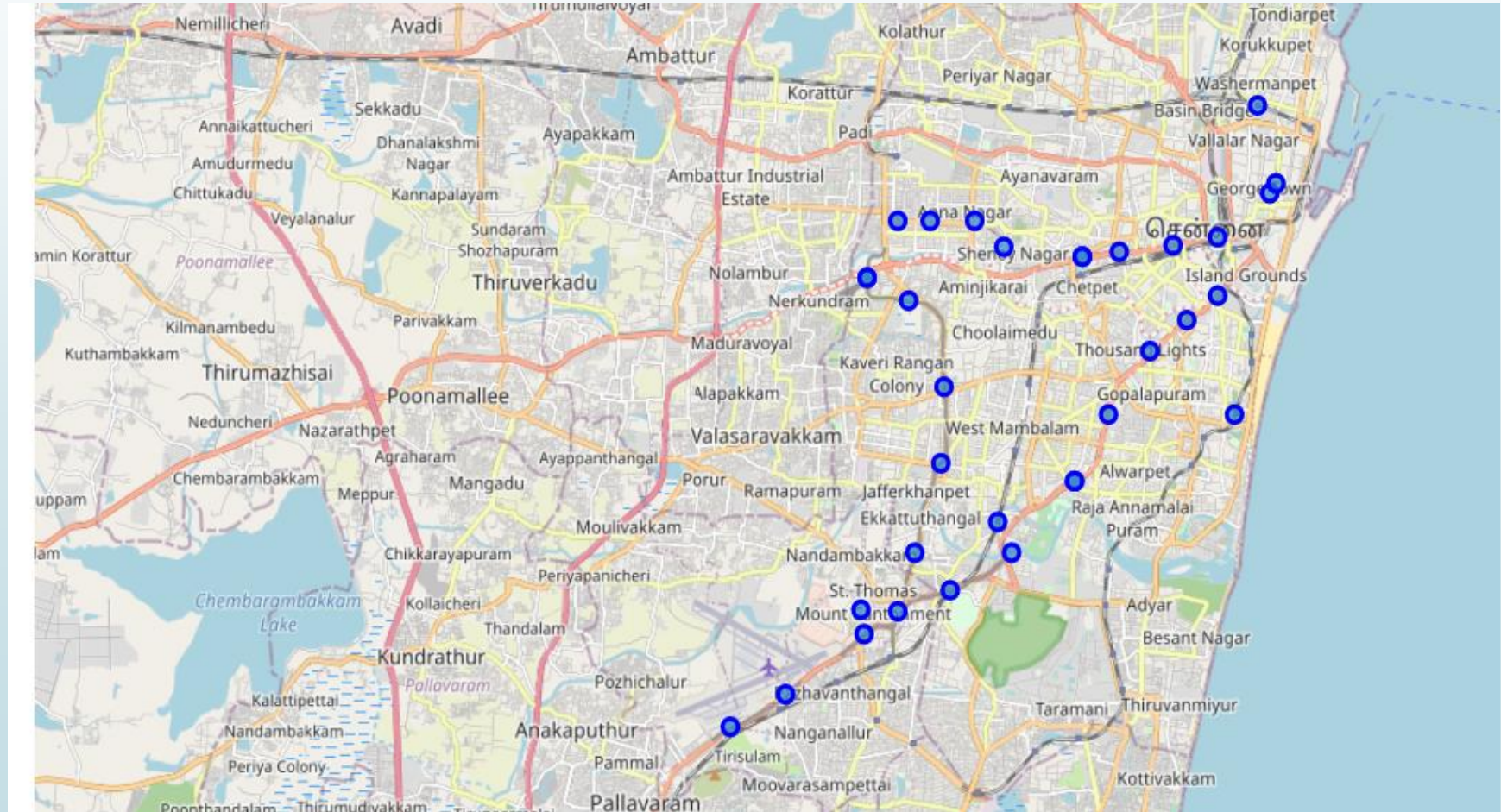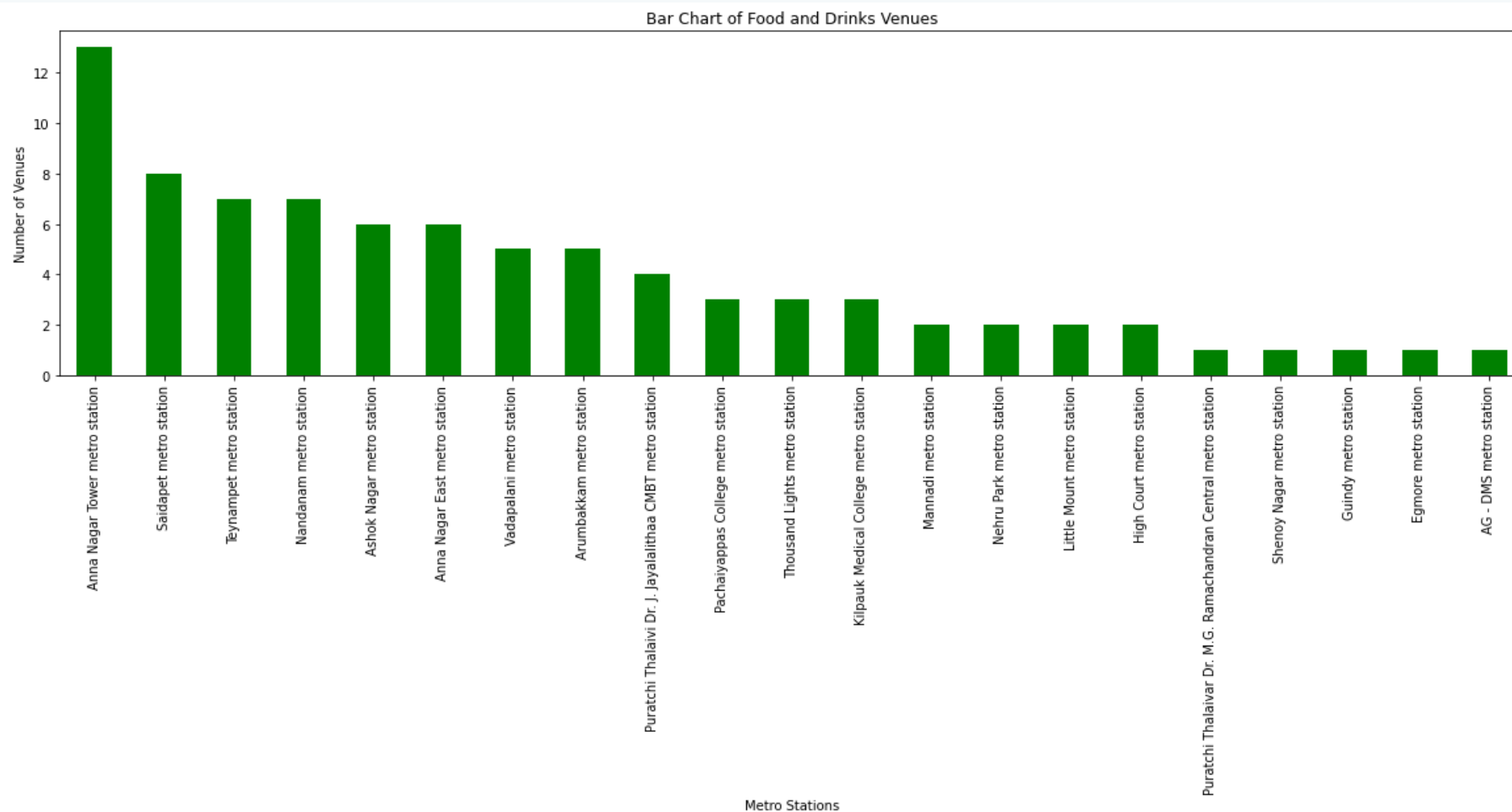
Cluster Chennai Metro Stations

➥ Cluster Chennai Metro stations based on 6 categories

- Food and Drinks

- Schools

- Medical Center
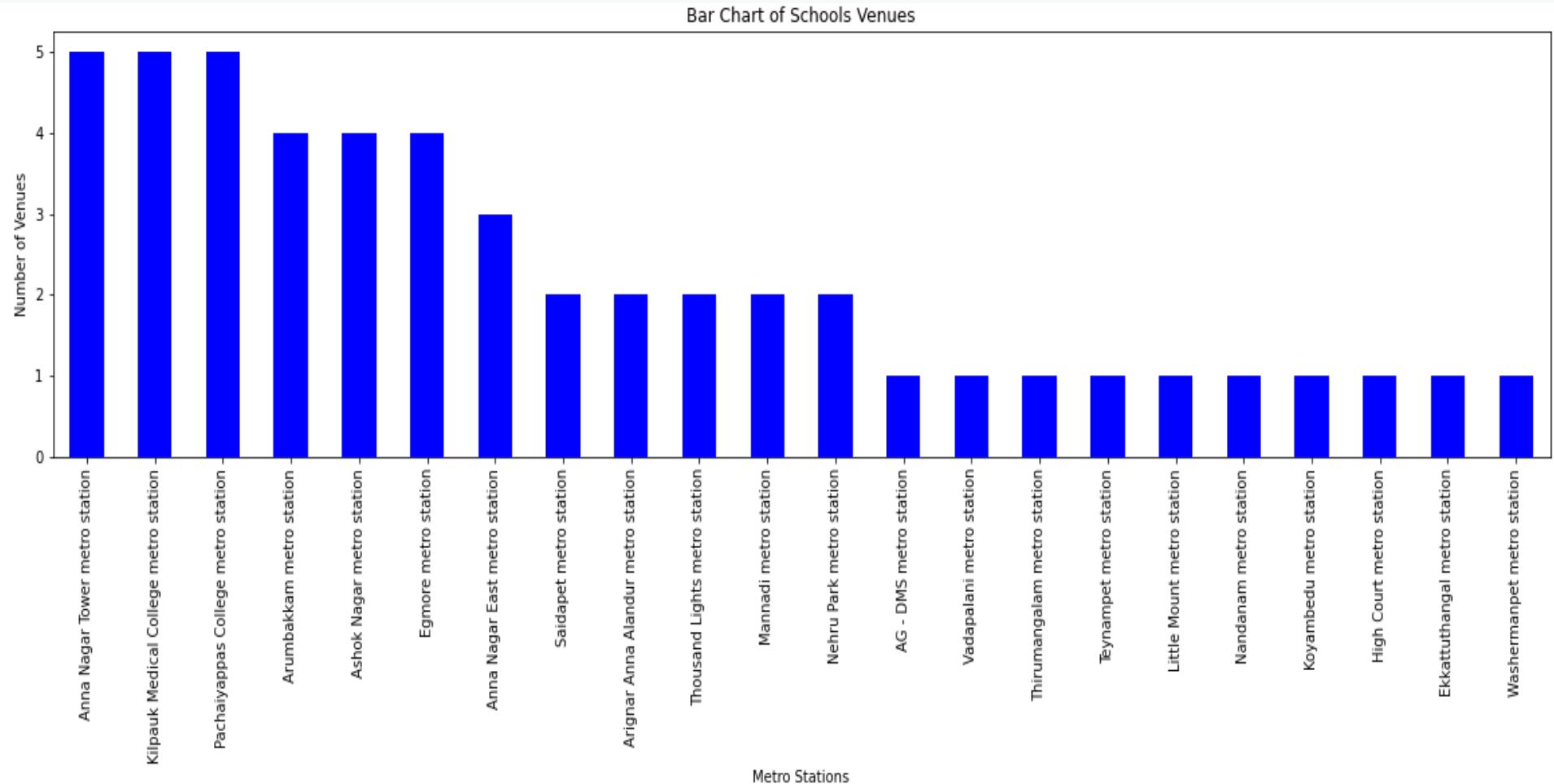
- Fitness Center

- Restaurant

- Shopping Mall

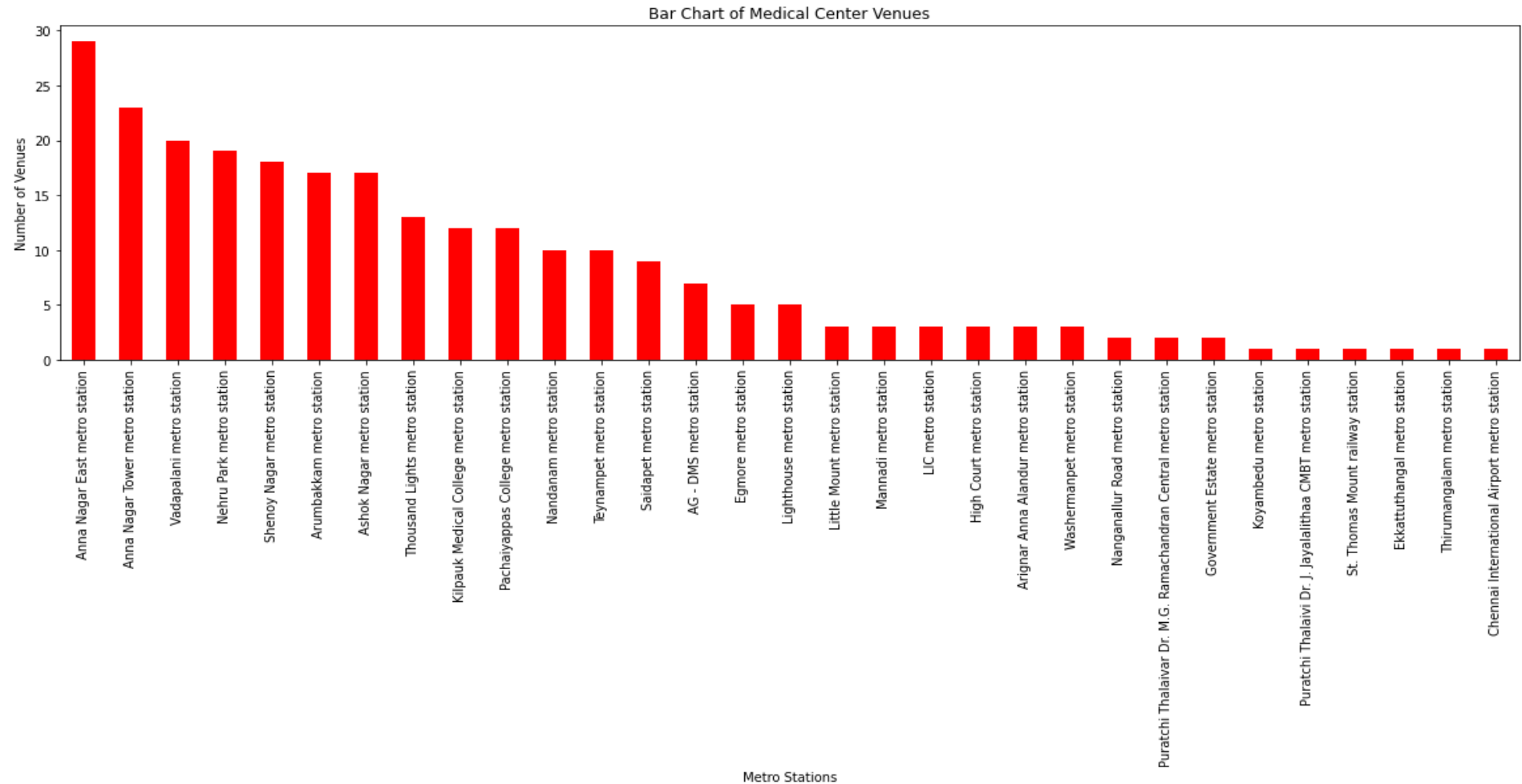# Prepare Bar Chart for each category

➤ Food and Drinks

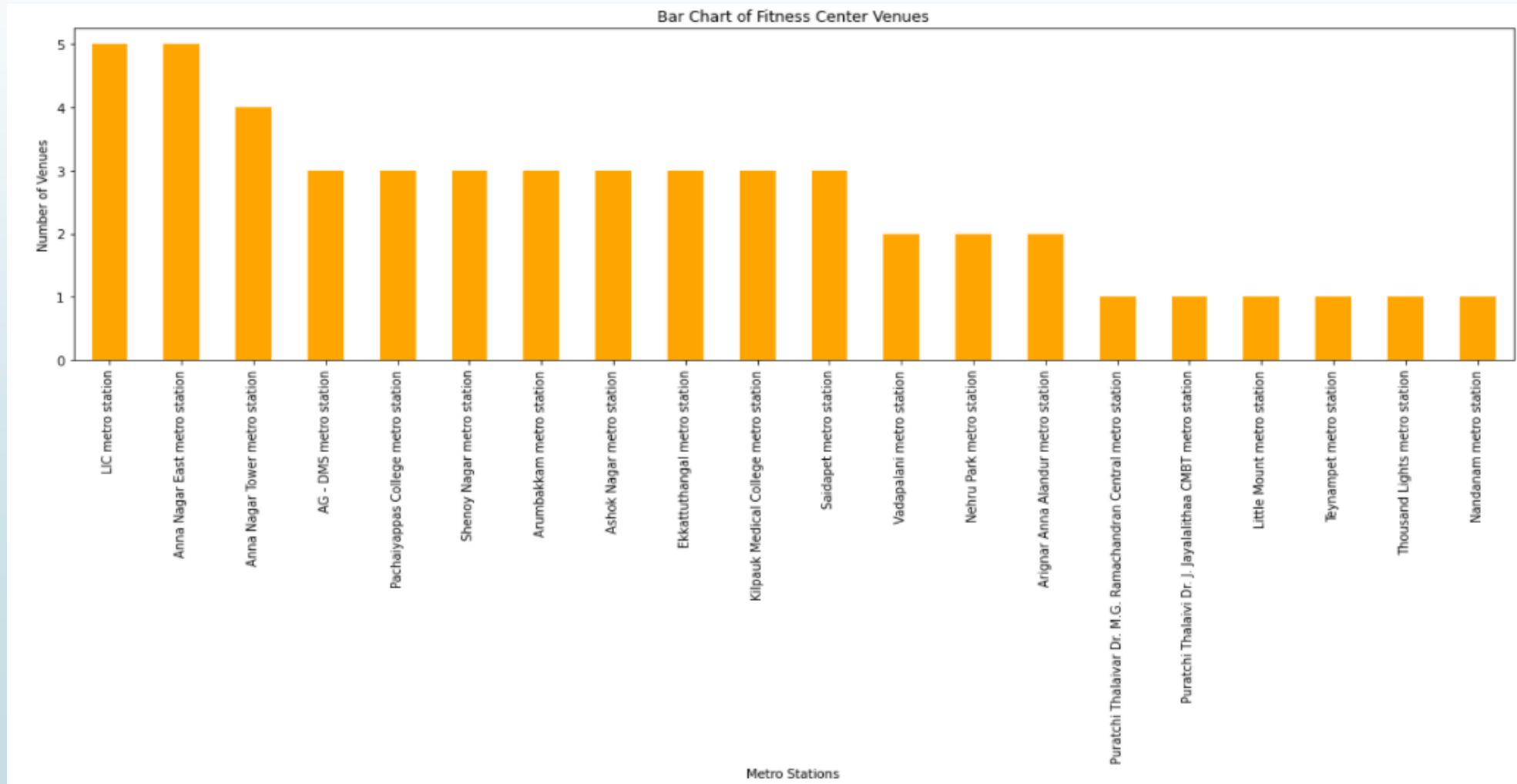# Prepare Bar Chart for each category

▪ Schools

Prepare Bar Chart for each category

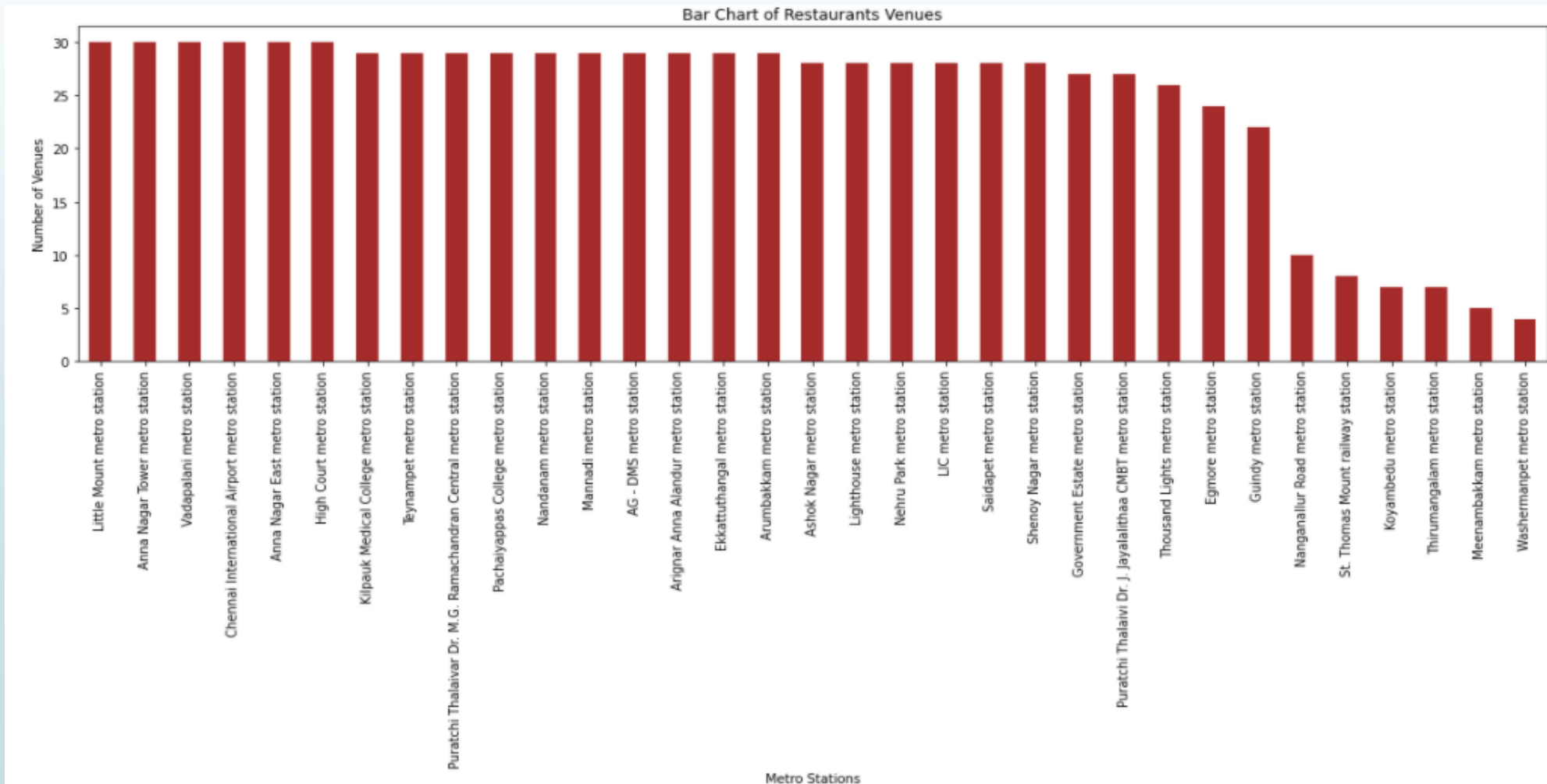➡ Medical Center

Bar Chart of Medical Center Venues

Prepare Bar Chart for each category

➤ Fitness Center

Prepare Bar Chart for each category

▶ Restaurants

Prepare Bar Chart for each category

➡ Shopping Malls

Bar Chart of Shopping Mall Venues
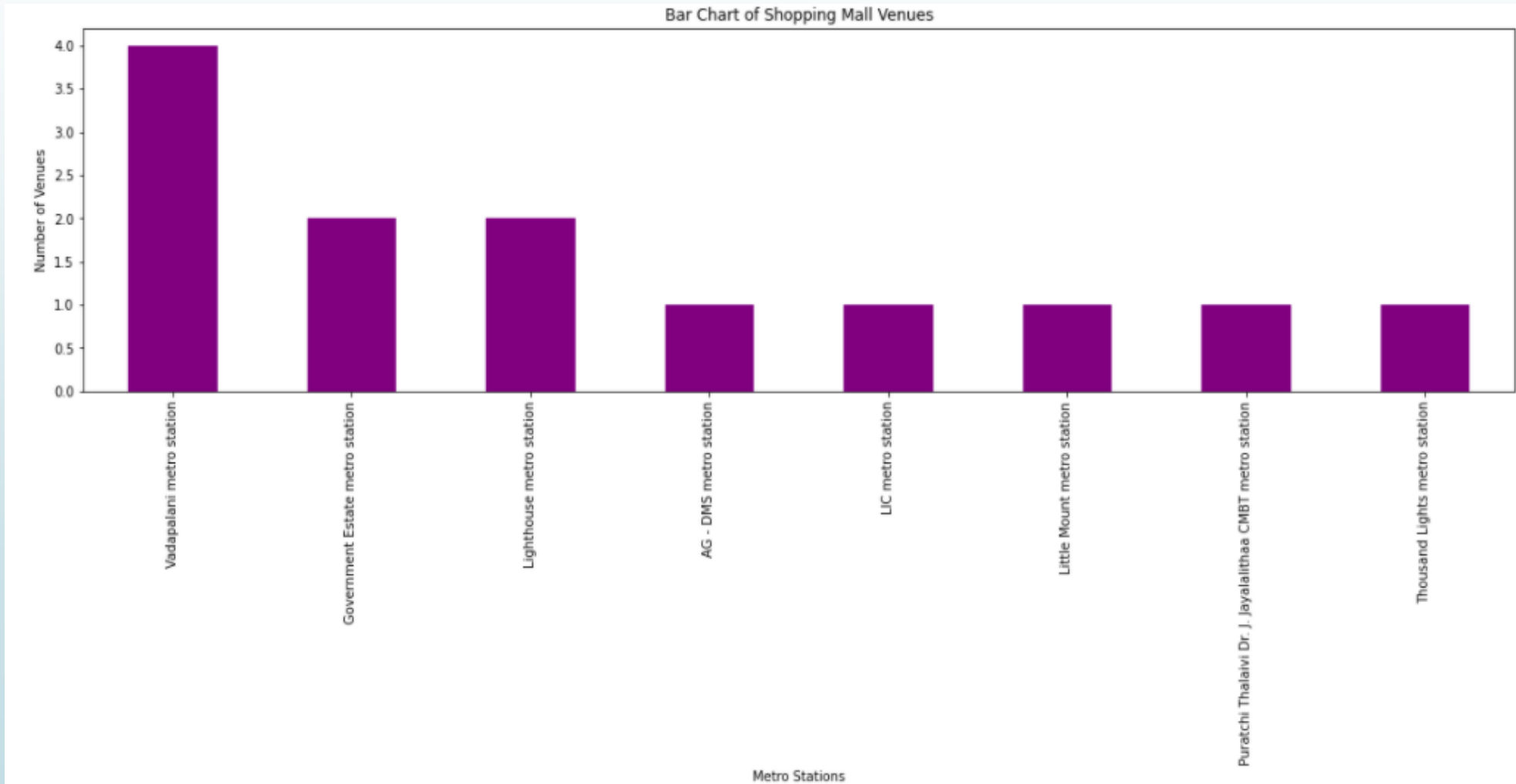
Execute K-means algorithm

- Categorize metro station using K-means algorithm with 3 cluster, Red/Blue/Green

Cluster Labels	Latitude	Longitude	Food and Drinks	Schools	Medical Center	Fitness Center	Restaurants	Shopping Mall
blue	13.061636	80.229068	5.307692	2.692308	16.076923	2.615385	28.692308	0.384615
green	13.041475	80.205945	0.000000	0.500000	1.333333	0.000000	6.833333	0.000000
red	13.047800	80.241820	1.000000	0.857143	2.785714	1.142857	27.928571	0.571429

Plot categorized metro station on Dubai map
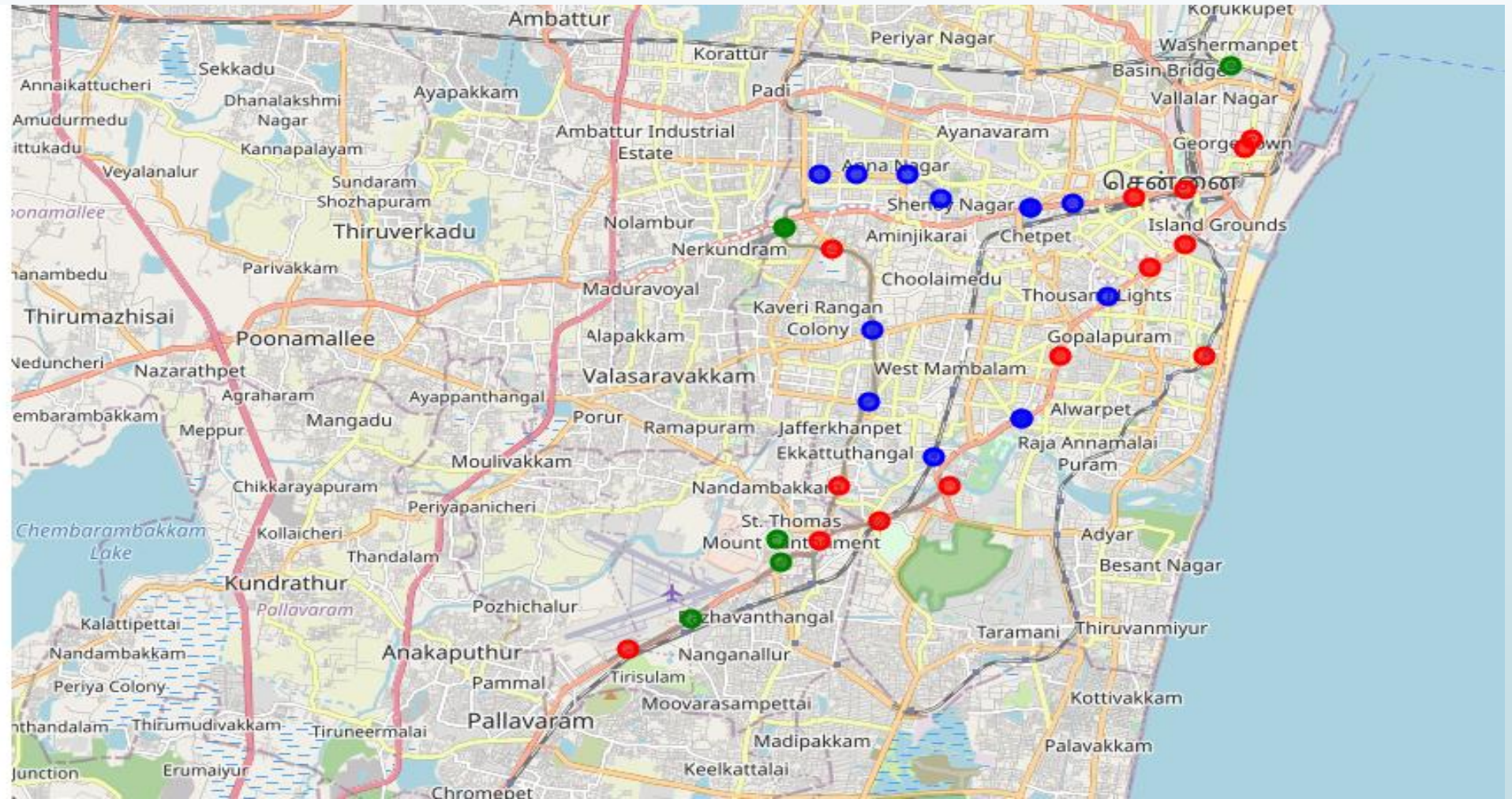
Discussion

- With above results we can identify that metro station with label Blue are most suitable for residential purpose as we have lot of options for each type of venues like Food and Drinks, Restaurant, Medical Center, Schools etc.

- And we also identified that metro station with label green has least number of venues across same category and may not be that desirable for residential purpose.

- But this report does not consider the availability of residential apartments and rent of residential apartment in these areas. These factors may affect the actual desirability of people to rent the apartment in these areas.

Conclusion

- We conclude with this report that people may consider options available of different venues near metro station while deciding to rent apartment.

- We also conclude that some metro station has higher number of options around them compare to other metro stations.

Thank You