

IBM Applied Data Science Capstone

# Chennai Metro Stations Clustering

*Chennai, Tamil Nadu, India*

By: Sanjay Kannu Periyasami

November 2020



**CHENNAI METRO RAIL LIMITED**



## Table of Contents

Introduction .....	3
Business Problem: .....	3
Target Audience of this project .....	3
Data Section .....	3
Methodology.....	4
Results Section .....	5
Food and Drinks.....	5
Schools.....	6
Medical Center .....	7
Fitness Center .....	8
Restaurants.....	9
Shopping Mall.....	10
Metro Station Categorization .....	11
Discussion Section.....	13
Conclusion Section .....	14

## Introduction

**Chennai** is the capital of the Indian state of **Tamil Nadu**. The Quality of Living Survey rated **Chennai** as the safest **city** in India. As per the estimates of UN World Population Prospects, the Chennai Population in 2020 is **10.9 Million** (1.09 Crore). Chennai is fifth largest urban agglomeration city in India and 30th largest city in the World

The Chennai Metro Rail Limited (CMRL) started Chennai Metro in 2015 to handle its traffic problem. Chennai Metro is now lifeline of Chennai and its working population is depending on Chennai Metro.

Foursquare.com has very rich database of venues around the world it also provides tips and photos in detail. Foursquare provide APIs to developers to explore these data. We will use Foursquare APIs in this work to get Chennai Metro information like name/latitude/longitude and different venue details around each Chennai Metro stations.

## 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.

To decide best residential area, we have considered five categories listed below.

1. Food & Drink Shops
2. School
3. Medical Centre
4. Gym
5. Restaurants
6. Shopping Mall

In this project we must identify which metro station is best considering above 6 conditions and categorize them and plot them on map to visual identification.

## Target Audience of this project

This project is particularly useful for expatriates from other places of India looking for the residence in the neighborhoods of Chennai.

## Data Section

To solve this problem, we will use Foursquare APIs and collect data of Chennai Metro Stations.

Find out Chennai Metro Station Details including Name, Latitude, and Longitude. Using search endpoint with category id of “Metro Station” we will get all metro station details. We will convert JSON into dataframe. Once dataframe is received we will remove unwanted columns except Name, Latitude and Longitude.

Gathering all the Metro Station information, we have total 34 rows.

For each metro station we will find venue details of each category listed below with search endpoint using category id from Foursquare.

S. No.	Category	Category Id
1	Food & Drink Shops	4bf58dd8d48988d1f9941735
2	Schools	4bf58dd8d48988d13b941735
3	Medical Center	4bf58dd8d48988d104941735
4	Fitness Center	4bf58dd8d48988d175941735
5	Restaurants	4d4b7105d754a06374d81259
6	Shopping Mall	4bf58dd8d48988d1fd941735

Now we have metro station details and venue for each metro station in radius of 500 meters. We will group venue detail by Name of Metro station and Venue Category to get number of venues for each category at each metro station.

## Methodology

First, we will fetch all metro station using Foursquare API. It will provide Name, Latitude, and Longitude of all metro stations.

Once we have metro station details, we will fetch 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 using Foursquare API.

To check number of venues at each metro station we will plot Bar chart for each category. It will show us how many venues are there around each metro station for every category.

We will list out top 5 metro station for each venue category with number of venues.

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

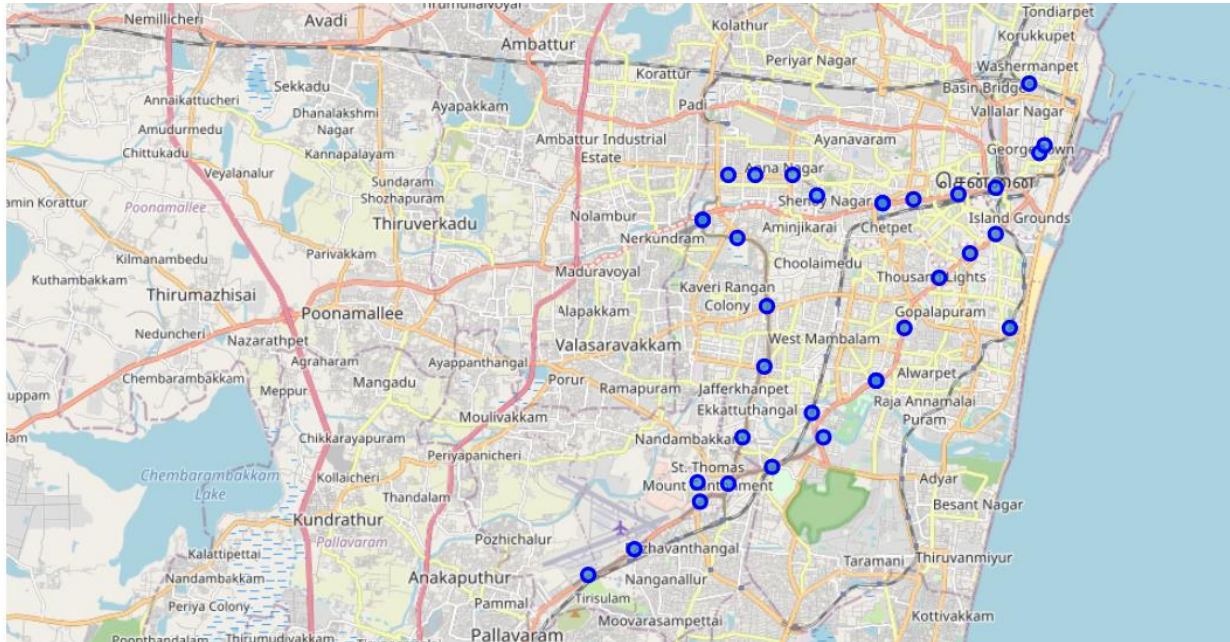
Using K-means algorithm we will label each metro station with its respective category and plot these categories on Chennai map using folium library.

Each category will be displayed with separate colour for better visualization.

## Results Section

When we fetched Chennai metro stations, we got 34 metro stations.

We also displayed all metro stations on Chennai map using folium library and shown in Figure 1



**Figure 1: Metro Stations on Chennai Map**

Once we fetched venue details for each category, we got below mentioned details.

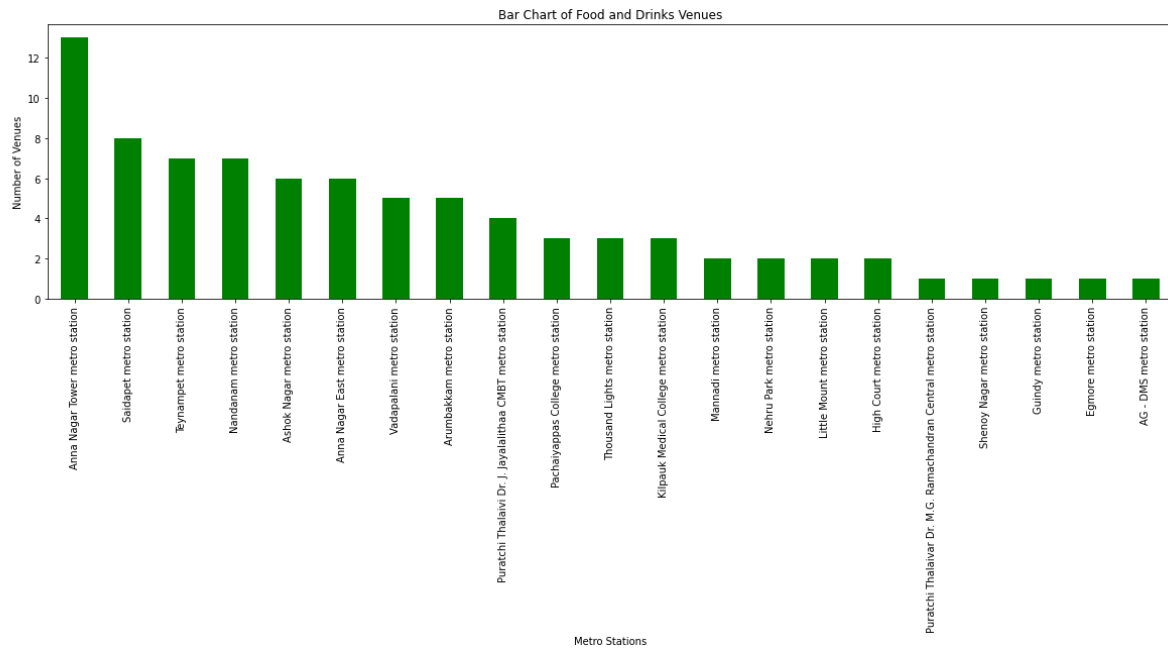
## Food and Drinks

Top 5 metro station in term of no of venues in **Food and Drinks** category are mentioned below.

Sl. No	Metro Station	Number of Venues
1	Anna Nagar Tower metro station	13
2	Saidapet metro station	8
3	Teynampet metro station	7
4	Nandanam metro station	7
5	Ashok Nagar metro station	6

We can see that Anna Nagar Tower metro station has maximum Food and Drinks venues in 500-meter radius.

Bar Plot of Food and Drinks venues against metro station is shown in Figure 2



**Figure: 2**

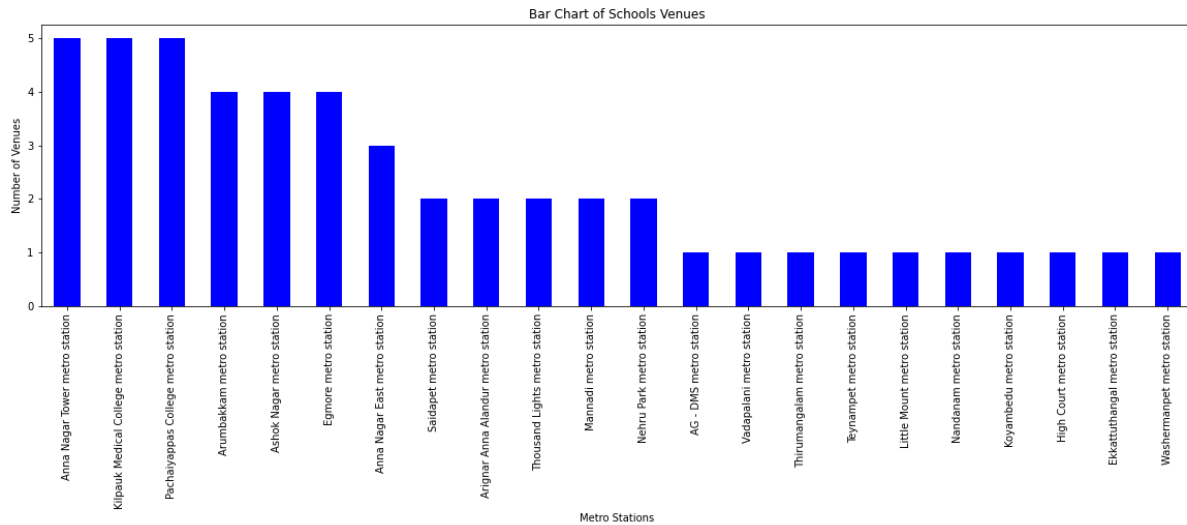
## Schools

Top 5 metro station in term of no of venues in **Schools** category are mentioned below.

Sl. No	Metro Station	Number of Venues
1	Anna Nagar Tower metro station	5
2	Kilpauk Medical College metro station	5
3	Pachaiyappas College metro station	5
4	Arumbakkam metro station	4
5	Ashok Nagar metro station	4

We can see that Anna Nagar Tower metro station has maximum Schools venues in 500-meter radius.

Bar Plot of Schools venues against metro station is shown in Figure 3



**Figure 3**

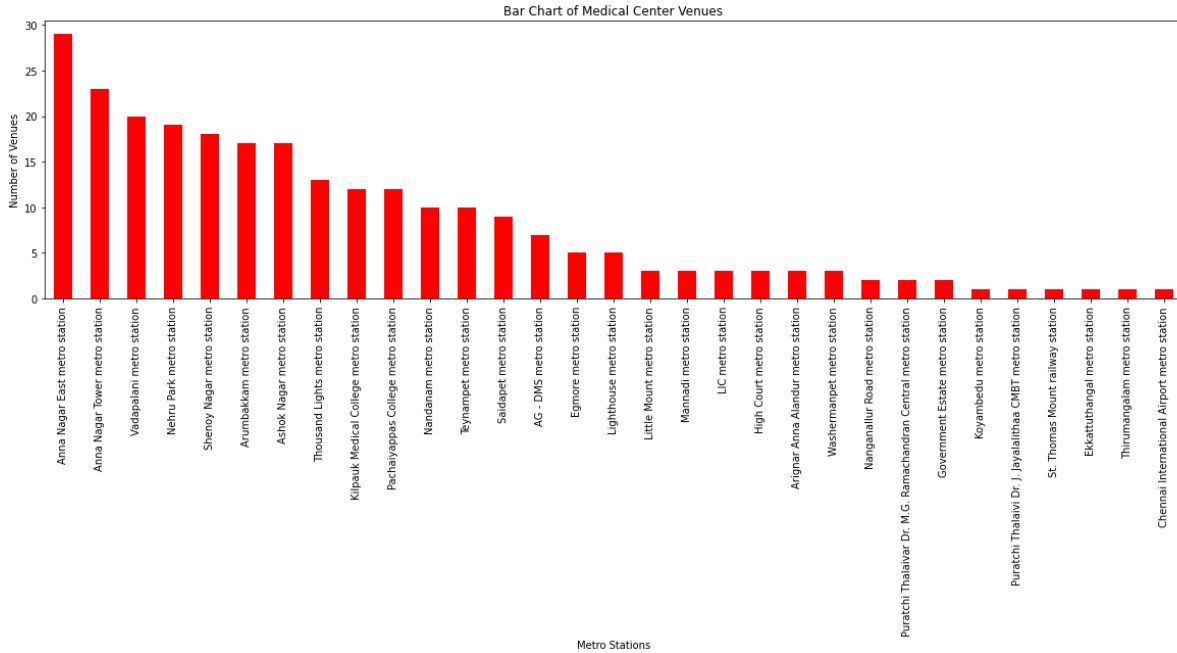
## Medical Center

Top 5 metro station in term of no of venues in **Medical Center** category are mentioned below.

Sl. No	Metro Station	Number of Venues
1	Anna Nagar East metro station	29
2	Anna Nagar Tower metro station	23
3	Vadapalani metro station	20
4	Nehru Park metro station	19
5	Shenoy Nagar metro station	18

We can see that Anna Nagar East metro station has maximum Medical Center venues in 500-meter radius.

Bar Plot of Medical Center venues against metro station is shown in Figure 4



**Figure 4**

## Fitness Center

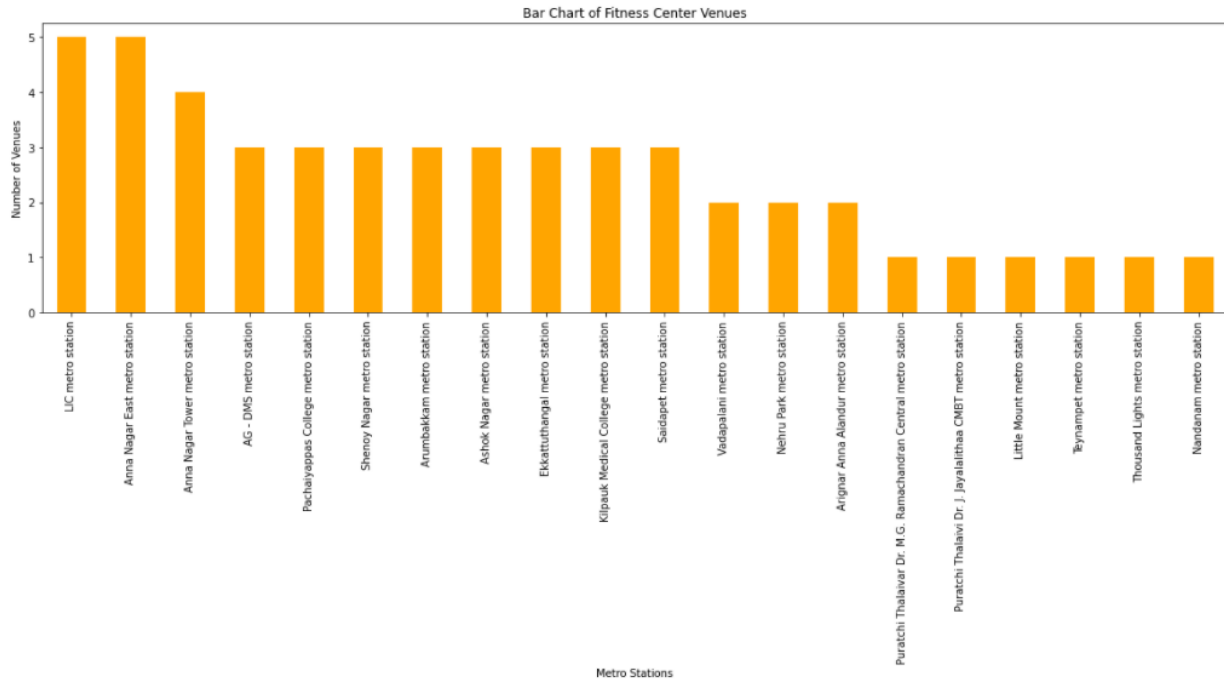
Top 5 metro station in term of no of venues in **Fitness Center** category are mentioned below.

Sl. No	Metro Station	Number of Venues
1	LIC metro station	5
2	Anna Nagar East metro station	5
3	Anna Nagar Tower metro station	4
4	AG - DMS metro station	3
5	Pachaiyappa's College metro station	3

We can see that LIC metro station and Anna Nagar East metro station has maximum Fitness Center venues in 500-meter radius.

Bar Plot of Fitness Center venues against metro station is shown in Figure 5





**Figure 5**

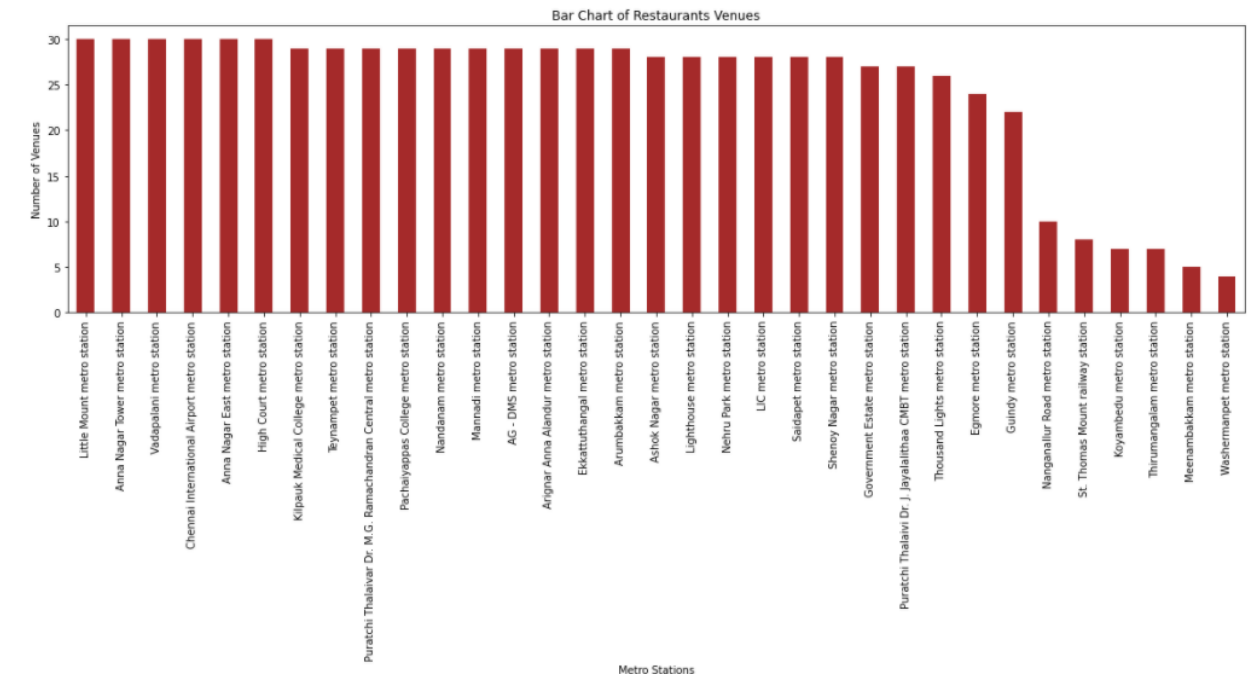
## Restaurants

Top 5 metro station in term of no of venues in **Restaurants** category are mentioned below.

Sl. No	Metro Station	Number of Venues
1	Little Mount metro station	30
2	Anna Nagar Tower metro station	30
3	Vadapalani metro station	30
4	Chennai International Airport metro station	30
5	Anna Nagar East metro station	30

We can see that Little Mount metro station, Anna Nagar Tower metro station, Vadapalani metro station, Chennai International Airport metro station, Anna Nagar East metro station have maximum Restaurants venues in 500-meter radius.

Bar Plot of Restaurants venues against metro station is shown in Figure 6.



**Figure 6**

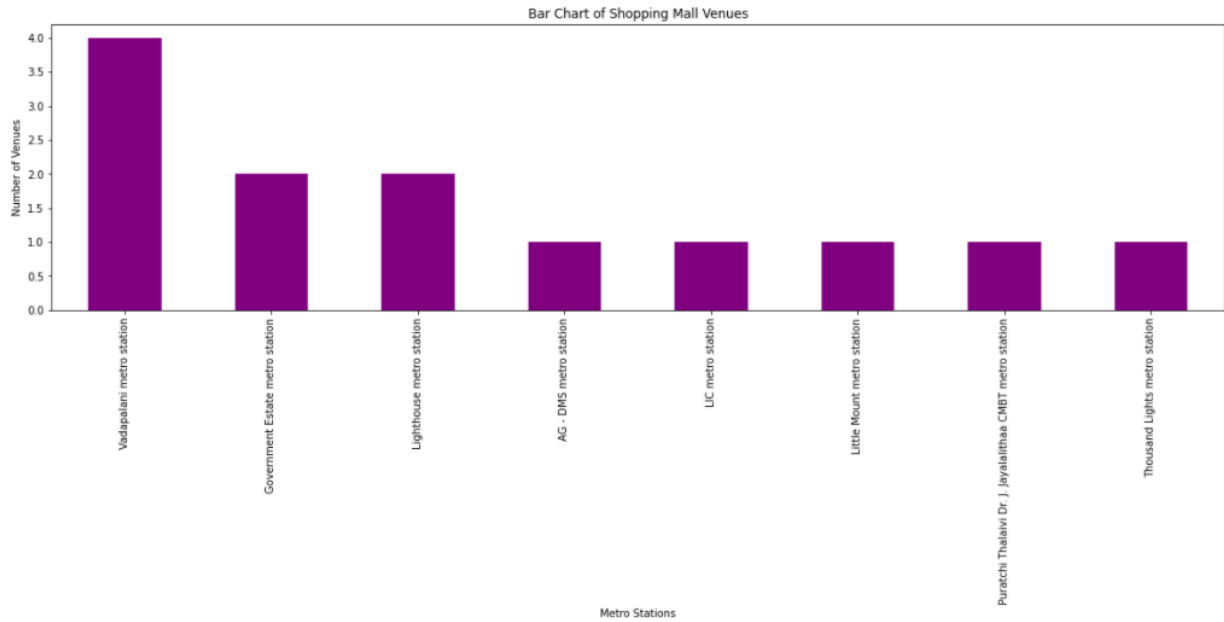
## Shopping Mall

Top 5 metro station in term of no of venues in **Shopping Mall** category are mentioned below.

Sl. No	Metro Station	Number of Venues
1	Vadapalani metro station	4
2	Government Estate metro station	2
3	Lighthouse metro station	2
4	AG - DMS metro station	1
5	LIC metro station	1

We can see that Vadapalani metro station has maximum **Shopping Mall** venues in 500-meter radius.

Bar Plot of Shopping Mall venues against metro station is shown in Figure 7.



**Figure 7**

## Metro Station Categorization

After fetching all data we created dataframe with features (categories) as column and ran K-means algorithm with k-cluster value as 3.

After categorization we labelled all metro station and changed the label names to **Red, Blue and Green**.

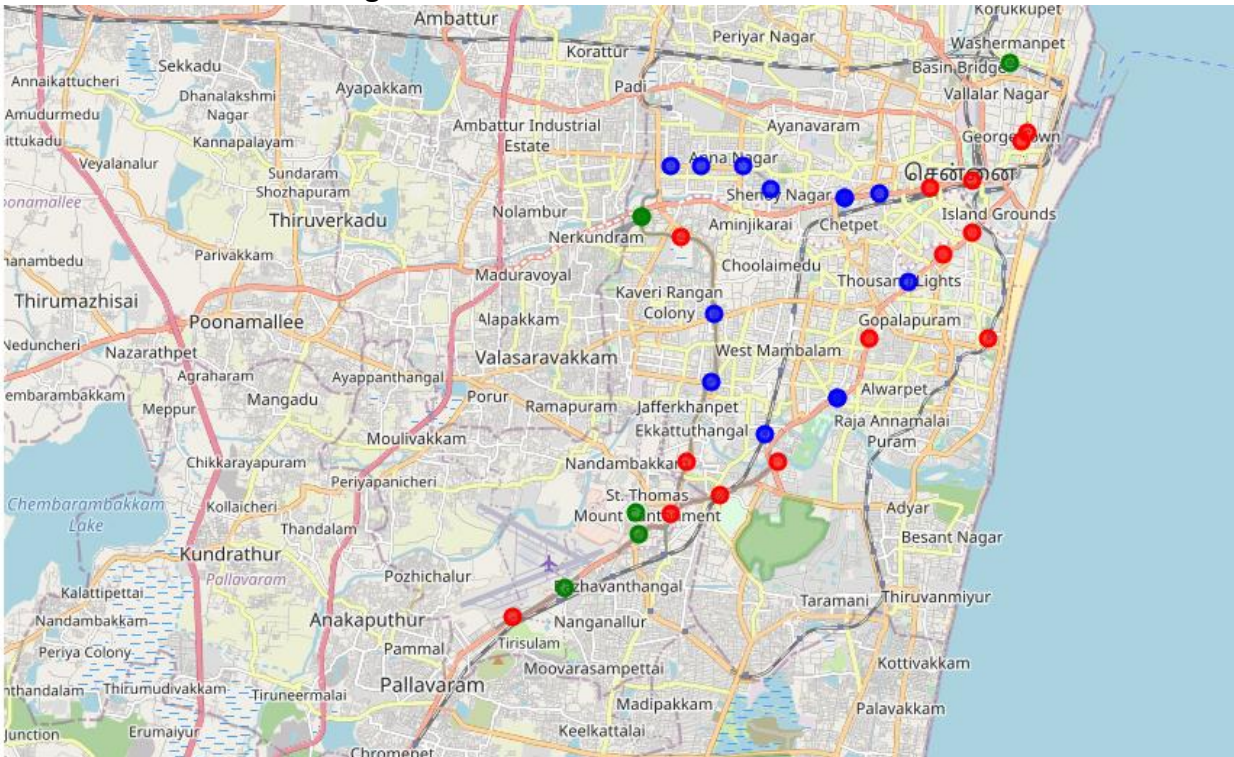
When we checked the mean values of all venues for each category we found values mentioned in **Figure 8**.

We can see that label **Blue** has maximum no of venues across different features followed by label **Red** and then label **Green**

	Latitude	Longitude	Food and Drinks	Schools	Medical Center	Fitness Center	Restaurants	Shopping Mall
Cluster Labels								
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

**Figure 8 : Mean values of all feature for each Label**

We also plot all metro stations on Chennai Map with its respective category colour to better visualization. It is shown in **Figure 9** below



**Figure 9: Metro station with label color on Chennai Map**

Label of all Metro Station and no of venues against each feature is shown in Figure 10.

Metro Station	Cluster Labels	Latitude	Longitude	Food and Drinks	Schools	Medical Center	Fitness Center	Restaurants	Shopping Mall
AG - DMS metro station	red	13.045515	80.248436	1	1	7	3	29	1
Anna Nagar East metro station	blue	13.085080	80.218646	6	3	29	5	30	0
Anna Nagar Tower metro station	blue	13.085226	80.208746	13	5	23	4	30	0
Arignar Anna Alandur metro station	red	13.004859	80.201483	0	2	3	2	29	0
Arumbakkam metro station	blue	13.085245	80.201590	5	4	17	3	29	0
Ashok Nagar metro station	blue	13.035452	80.211194	6	4	17	3	28	0
Chennai International Airport metro station	red	12.981076	80.164156	0	0	1	0	30	0
Egmore metro station	red	13.080281	80.263055	1	4	5	0	24	0
Ekkattuthangal metro station	red	13.016922	80.205424	0	1	1	3	29	0
Government Estate metro station	red	13.069770	80.272882	0	0	2	0	27	2
Guindy metro station	red	13.009275	80.213085	1	0	0	0	22	0
High Court metro station	red	13.090991	80.284594	2	1	3	0	30	0
Kilpauk Medical College metro station	blue	13.077782	80.242902	3	5	12	3	29	0
Koyambedu metro station	green	13.073450	80.194836	0	1	1	0	7	0
LIC metro station	red	13.064722	80.266092	0	0	3	5	28	1
Lighthouse metro station	red	13.045362	80.276728	0	0	5	0	28	2
Little Mount metro station	red	13.017068	80.226882	2	1	3	1	30	1
Mannadi metro station	red	13.092904	80.285907	2	2	3	0	29	0
Meenambakkam metro station	green	12.987673	80.176479	0	0	0	0	5	0
Nandanam metro station	blue	13.031709	80.241196	7	1	10	1	29	0
Nanganallur Road metro station	green	13.000195	80.194016	0	0	2	0	10	0
Nehru Park metro station	blue	13.078868	80.250907	2	2	19	2	28	0
Pachaiyappas College metro station	blue	13.077761	80.242841	3	5	12	3	29	0
Puratchi Thalaivar Dr. M.G. Ramachandran Central metro station	red	13.081689	80.272897	1	0	2	1	29	0
Puratchi Thalaivi Dr. J. Jayalalithaa CMBT metro station	red	13.068769	80.203860	4	0	1	1	27	1
Saidapet metro station	blue	13.023310	80.223719	8	2	9	3	28	0
Shenoy Nagar metro station	blue	13.079793	80.225245	1	0	18	3	28	0
St. Thomas Mount railway station	green	13.005115	80.193495	0	0	1	0	8	0
Teynampet metro station	blue	13.031709	80.241196	7	1	10	1	29	0
Thirumangalam metro station	green	13.073450	80.194836	0	1	1	0	7	0
Thousand Lights metro station	blue	13.058370	80.258000	3	2	13	1	26	1
Vadapalani metro station	blue	13.050962	80.211705	5	1	20	2	30	4
Washermanpet metro station	green	13.108965	80.282010	0	1	3	0	4	0

**Figure 10**

## Discussion Section

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 Section

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.