

Cluster Analysis of Railway Stations Along the Mumbai Suburban Rail Network

Introduction

The Mumbai suburban rail network, which is the commuter rail network of the city of Mumbai, India, is one of the busiest commuter rail networks in the world. It carries more than 7.5 million passengers in a normal day and it is spread over 340km of track length. It has about 150 railway stations covering the city and the suburban areas.

Most of the customers of this rail network are the middle class and working class population and there is an economy surrounding such commuters. Many small time vendors and shopkeepers depend on the traffic at each station and such businesses form one of the major part of the informal economy of India.

The railways stations are also a major crowd management hotspots. Thousands of people converge in a small area in a short time and they disperse to various places. Being the backbone of the entire transport network of the city, study of the traffic flow at these railway stations are important in both large scale and locality level city planning.

When the network and stations were originally planned, the city was different. Today the city has grown and many of the once remote stations have now become the centres of new developments. An analysis of how each railway station has developed will help in better urban planning.

In this regard, the Mumbai suburban railway stations are studied using clustering to see how similar or how different each station has grown to in terms of establishments that have come up around them. The popular venues are used to cluster the railway stations, and the popularity is dependent on two factors at the least: (1) establishments around the stations and (2) type of travelers who frequent one station. Since the types of establishments that have come up at a location will also be related the type of people who frequent the stations, the popular locations should be a good parameter to classify the stations with. It is hoped that this analysis will help in better city planning around each station.

Intended Target Audience:

1. City planners
2. Railway station management teams
3. Railways' business management team

Business Problem:

How different or how similar are the railway stations along the suburban lines of the city of Mumbai, in terms of the popular establishments around them.

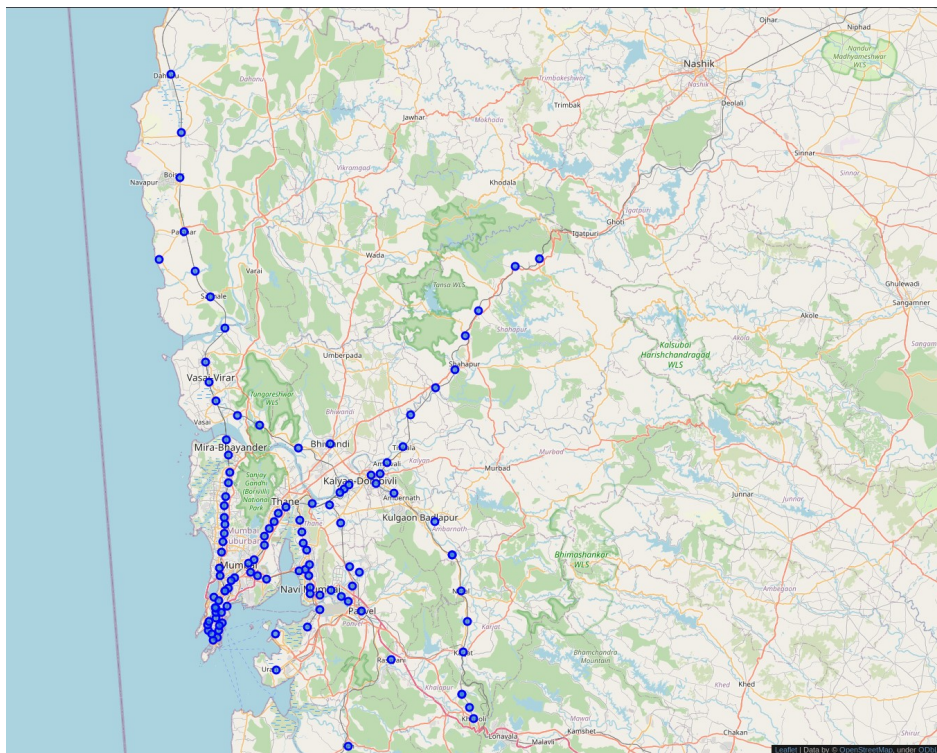
Data

For this project the following data are be used:

1. List of suburban railway stations: source: Wikipedia Link: https://en.wikipedia.org/wiki/List_of_Mumbai_Suburban_Railway_stations
2. Coordinates of each railway stations: from different sources using geopy library
3. Venues data: source: Foursquare

Data cleaning and preprocessing:

First a list of railway stations along Mumbai suburban rail network is downloaded from the Wikipedia and converted into a pandas data frame. Then for each station, the geographical



Map 1: Map showing the locations of railway stations in Mumbai
coordinates are retrieved using geopy library. Using the coordinates thus obtained, nearby venues for each station were obtained using Foursquare.

A list of 151 stations were obtained from the Wikipedia, for which the lat long values were obtained for 134 stations. The remaining stations were discarded from the analyses. This 135

stations are spread across the metropolitan, suburban and satellite town areas. List of venues were obtained for all these 134 stations.

After the list of venues were obtained, the histogram of the number of venues was plotted to get an idea about how many number of venues were available around stations. One can see that most of the stations more than 4 venues around them. Thus we proceed with the top 4 venues based on popularity for each station.

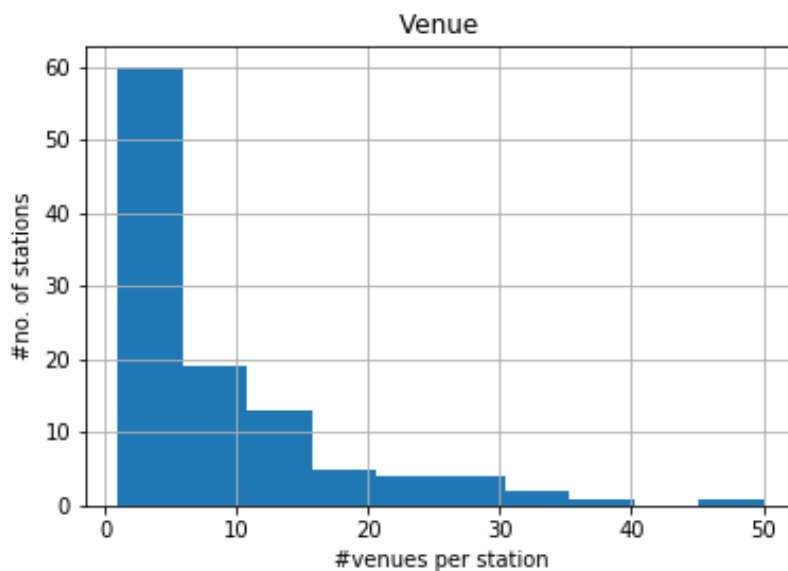


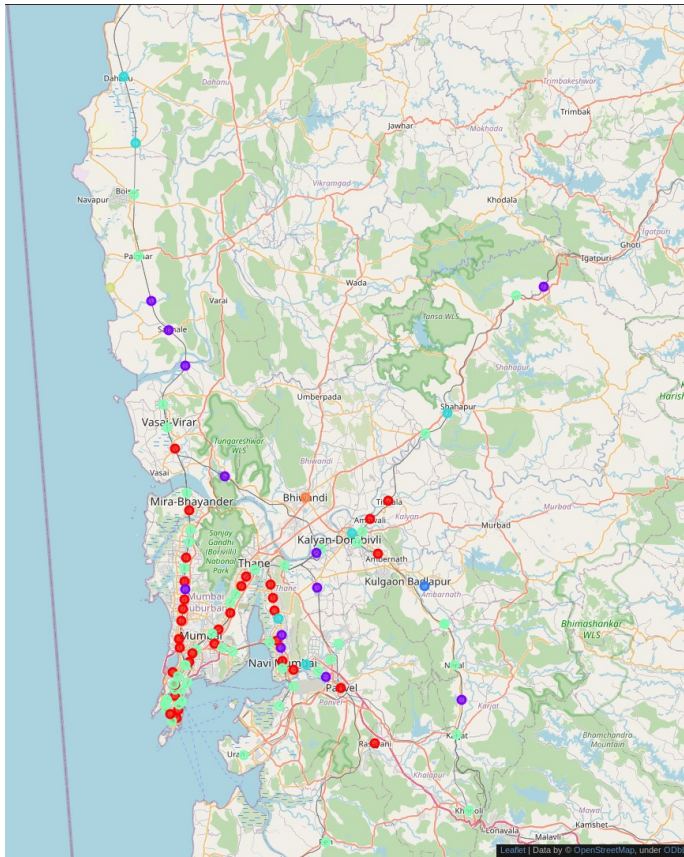
Illustration 1: Histogram showing number of stations with a particular number of popular venues around them

The venues are one-hot encoded first and then a list of 4 top venues are selected for each station. Further analyses were applied on this data set.

Methodology

K-means clustering is done on the data to see how the data is clustered. But for this one has to determine the number of clusters the data need to be divided into. The ideal number of clusters was found by trial and error and the clusters need to have some physical meaning which could be manually checked. The best number of clusters thus selected was 7.

Once the data is clustered, it is plotted on a map to get a better perspective.



Map 2: Map showing railway stations in Mumbai divided into 7 clusters and their location

Results

The number of stations that fall into the 7 category are given below in table (1). From inspecting the data in each category, the type of stations that characterize each category is also summarized in the table.

As can be seen the 0-th category stations are mostly known major stations or junctions. The most popular venue around these is Indian Restaurant. Most of these stations are located at commercial centers or are major transit points.

Category 1 stations are mainly seen at the city outskirts and characterized by the venue “Train Stations”. This is mainly because the businesses serving the passengers are mostly in the railway station itself. This include small tea shops and restaurants.

Category 2, 5 and 6 have only one station each and they are rather unique being a satellite town (2), a beach station(5) and a satellite industrial hub(6).

Category 3 stations are distributed mainly in the residential areas of the Navi(New) Mumbai area like Khargar and Kopar Khairane which are newly developed parts of the city. Some stations at the outskirts of Mumbai also fall into these category. These stations are characterized mainly by fast food restaurants.

Category 4 stations are distributed everywhere, but mostly clustered around the metro area of Mumbai and Navi Mumbai. These stations are crowded areas without major business venues near them like Category 0 stations.

Category No.	No. of stations	Characterized by venue	Location
0	33	Indian Restaurant	Known major stations in metro area and suburbs
1	13	Train station	Mainly at city outskirts
2	1	Grocery store	Satellite Town
3	6	Fast food restaurant	Residential suburbs of the Navi Mumbai(Newer City), Outskirts
4	54	Variable type	Metro area, suburbs, outskirts
5	1	Beach	Outer suburbs
6	1	Theater	Satellite industrial hub

Discussion

This study is based on the popularity of venues as reported by a portion of the people who goes to the venues around the stations. It does have all the problems associated with a self reported data set. Still this is one of the most extant data that one can avail concerning the popularity of the venues. As can be seen from the results, there are some patterns in this data. The stations can be divided mainly into 7 types as shown in table 1.

First thing to note is that the most popular venues in the city and suburban areas are the restaurants around the stations. But in remote areas, the most popular venue is mostly given as ‘Train Station’, which shows that most people spend time at establishments inside the railway station themselves. This has two implications:

(1) The crowds are mostly inside a remote station and are outside at city/suburban stations. This is a major point to note in designing the station or surroundings from a crowd management perspective.

(2) railways can get more revenue by leasing out more shops inside a remote station compared to leasing out a shop at city/suburban area where most of the passengers simply go out of the station to other venues.

Another observation is that in one station – Mahim is near a beach and most of the people near the station visit the beach. So there is a potential for tourism development at the station by providing more information and facilities to beach-goers at the station. This include sale of items like packed food, beach/swim wear etc. and more easy access to the beach from the station.

One station serving the satellite industrial hub of Bhivandi has a theater as a most popular venue nearby. There is scope for further business development at the station.

Conclusion

- The stations along the Mumbai Suburban Rail Network can be divided into 7 types based on the popular venues around them.
- Most popular venues in city area are restaurants and that in remote areas are mainly establishments serving inside the station themselves.
- More crowd outside the stations at category 0 and 3 stations. More crowd inside the stations at category 1 stations.
- Major flow between station and beach in category 5 station.
- Scope of development of establishments in category 6 station serving a satellite industrial hub