

# **Cluster Analysis of Railway Stations Along the Mumbai Suburban Rail Network**

# Introduction

- Mumbai suburban rail network
  - ~7.5 Million passengers per day
  - ~150 railway stations
  - ~340 km. of track length
- Backbone of Mumbai's transport network
- Centers of many small businesses, informal economy
- A nightmare to any crowd management team
- How similar/different are the stations? Are the establishments around them are similar? Travelers are similar?

# Business Problem

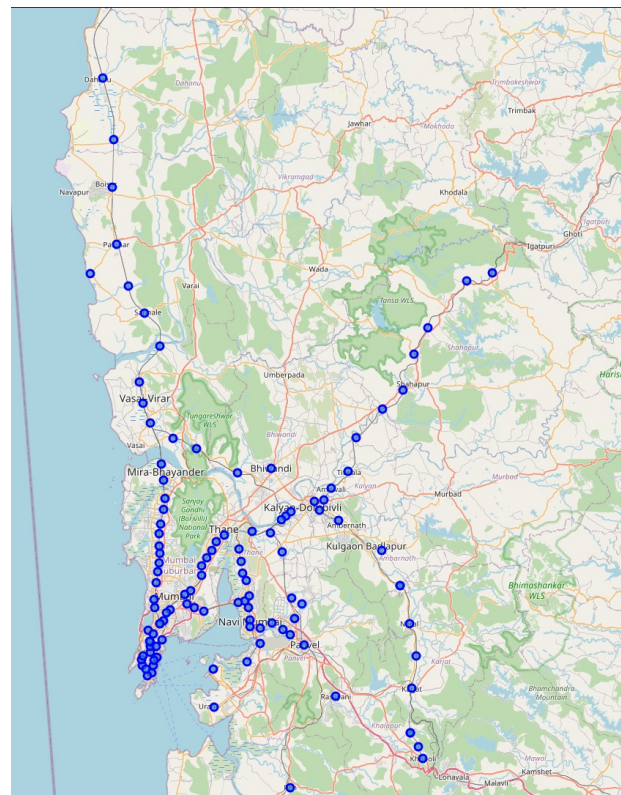
- How many different types of stations are there based on the type of popular establishments around them?
- What are the takeaway points from such a classification, for city development?

## **Target Audience:**

- City planners
- Railway business development teams
- Railway station managers

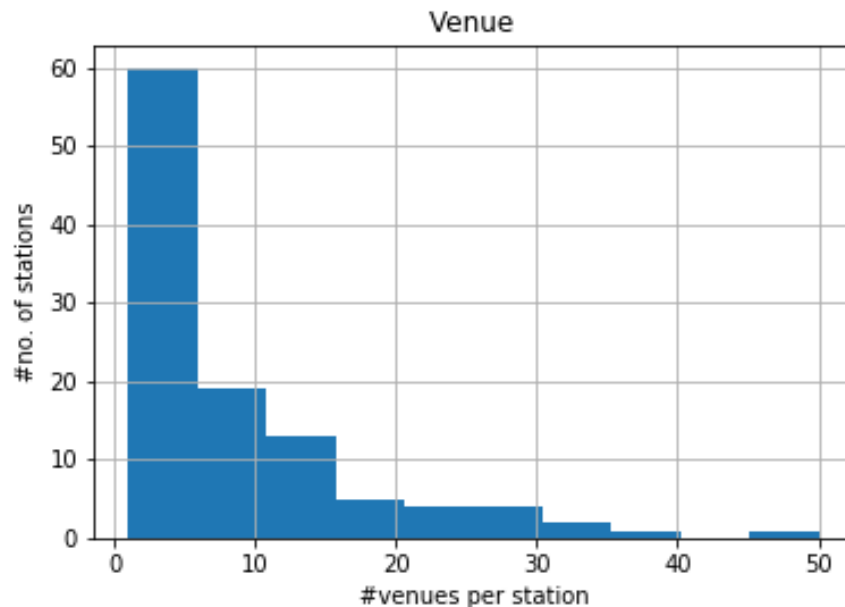
# Data

- Sources
  - Wikipedia: list of stations
  - Python's geopy library: coordinates of stations
  - Foursquare: list of popular venues around stations



# Data

- ~ 150 stations from Wikipedia
- ~135 stations with lat-long values
- 109 stations with at least 4 venues around them
  - So use the top 4 venues for classification
- Finally: 109 stations distributed over city/metro areas, suburban areas, remote or satellite towns and city outskirts, having at least 4 popular venues around them.

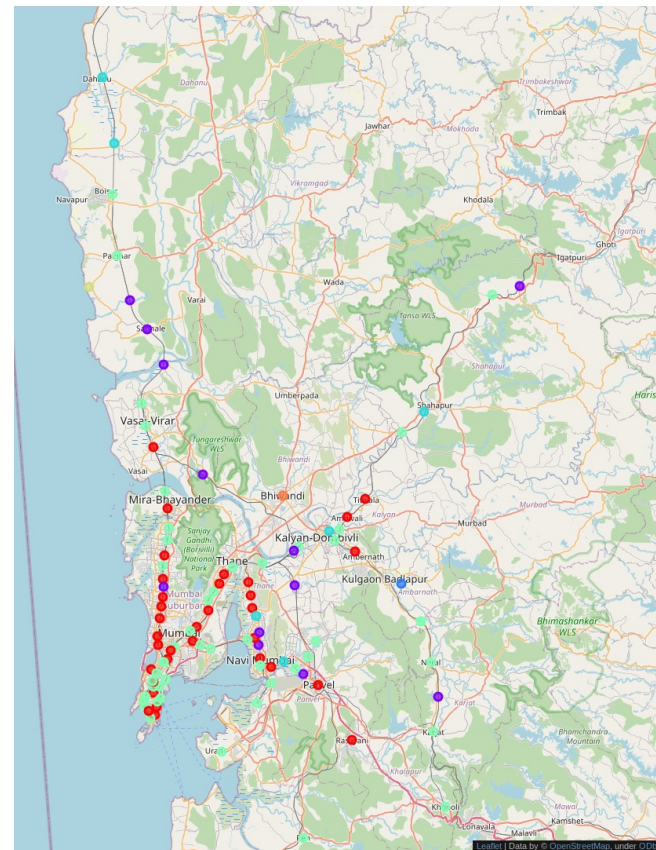


# Methodology

- Get the list of stations from Wikipedia
- Get the latitude and longitude for each station using geopy
- Get the list of popular venues from Foursquare
- One-hot encode the venues
- Get the top 10 venues for each station
- Cluster using k-means clustering algorithm
  - Determine number of clusters – here it was by trial and error
  - Classify using the algorithm
- Analyze the result – the first 2/3 venues might have got more weights in clustering.

# Results

- 7 clusters
- Category 0 to category 6 stations
- 2 categories have only one station each



# Results:

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 Mumbaiv(Newer City), Outskirts
4	54	Variable type	Metro area, suburbs, outskirts
5	1	Beach	Outer suburbs
6	1	Theater	Satellite industrial hub



# Discussion

- Most popular locations are restaurants around the station at city centers and shops inside the railway stations at remote locations
  - Crowd+business is more outside city/suburban restaurants and more inside stations at remote locations
  - implications in crowd management and revenue development

# Discussion

- Scope for business development at
  - One beach station(Mahim): beach related business
  - One satellite manufacturing hub(Bhivandi): restaurants
- Better to lease more floor area to shops at remote locations
  - Most people spend time inside the railway stations at remote locations
  - At city centers and suburban areas, people go out to other businesses

# Conclusion

- Location and venue based analyses are helpful in railway station development
- For the Mumbai suburban network, the stations could be classified into 7 types based on popular venues around them
- This helps in understanding the crowd patterns and venue preference of the crowd
- Helpful to identify lack of certain businesses at certain stations => scope for further business development
- This analysis will be more helpful if the data on the number of passengers are also available