

Suggesting Stations for the Silver Line project, Kerala

May, 2020



Why minimise the number of Stations?

- Low commute time
- Low implementation cost
- Low operational cost

How to select Stations?

- Optimal connectivity across the state
- Feasibility of construction
- Optimal separation between stations

Data acquisition

- Districts through which the railway corridor runs - (collected from <https://keralarail.com>)
- Main venues around the district which should have good connectivity to the Silver Line project - (collected from foursquare places api)

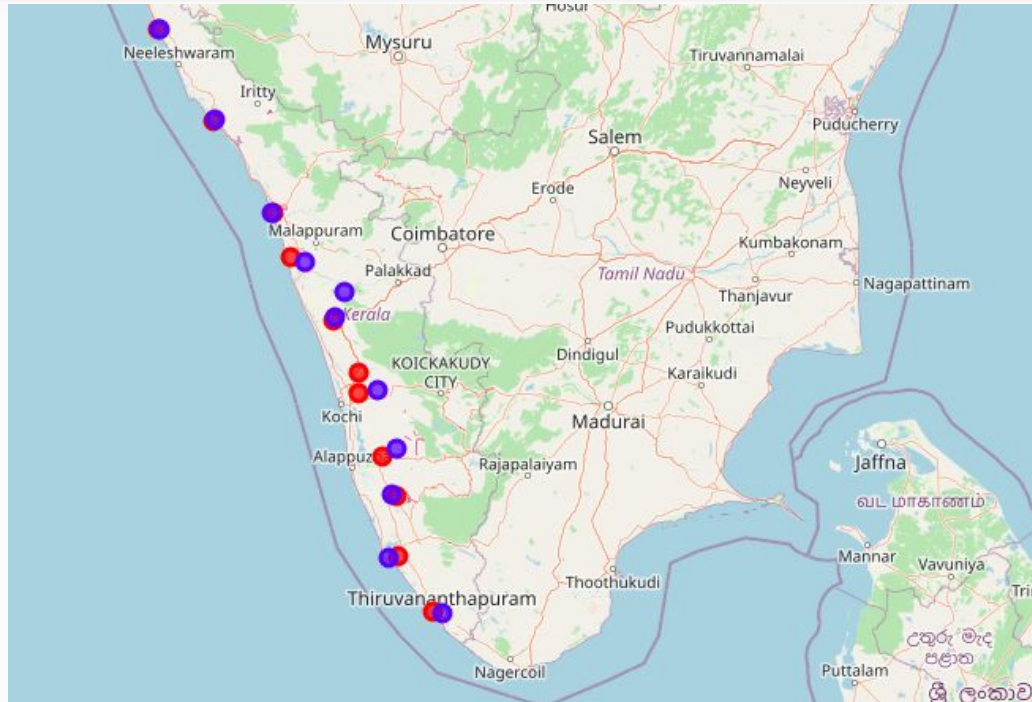
Data Cleaning

- Missing values manually added
- Duplicate entries dropped off
- Highly correlated data - treated based distance

Data Weights

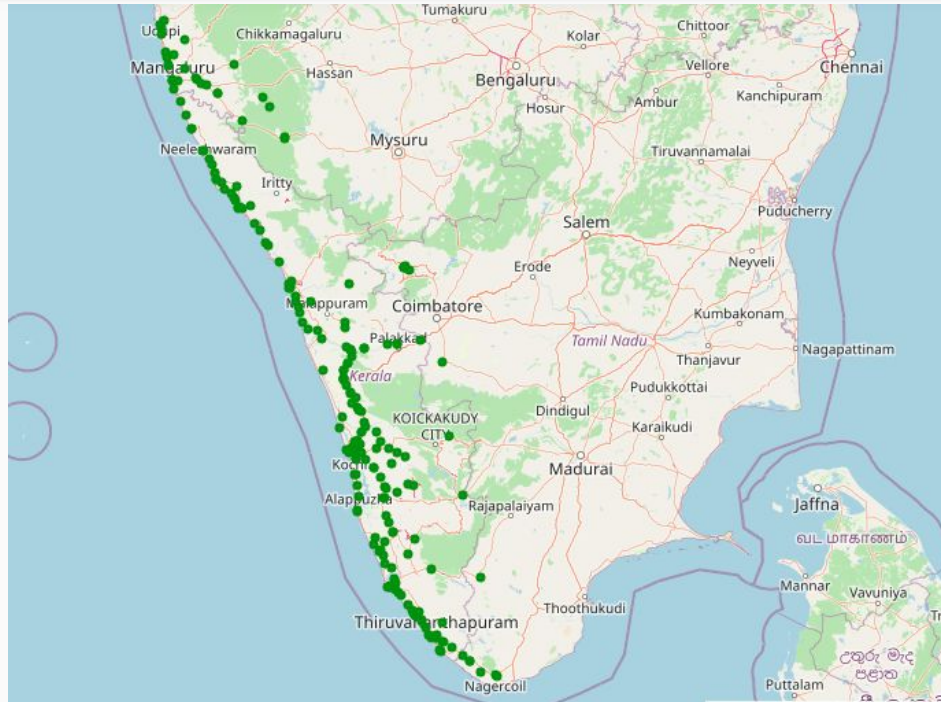
- Airport, Seaport, Railway Stations.etc are given appropriate weightage
- Weight of each record calculated using formula
 - ◆ $\text{weight_of_location} = \text{weight_of_category} * (100000.0 / \text{distance}) * (x + 0.01)$
- $x = (\text{longitude} < \text{min_logitude} + 1) ? 1 : 0$

Plot of Current Stations vs District Centres



- **Red: Current Station**
- **Blue: District Centre**

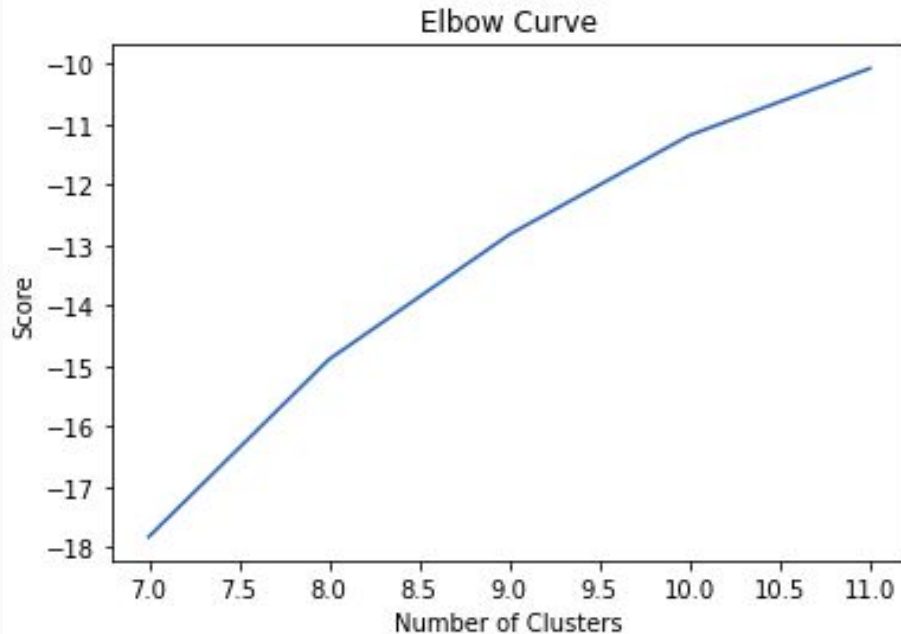
Plot of popular sites



Explore Districts through which K-Rail runs

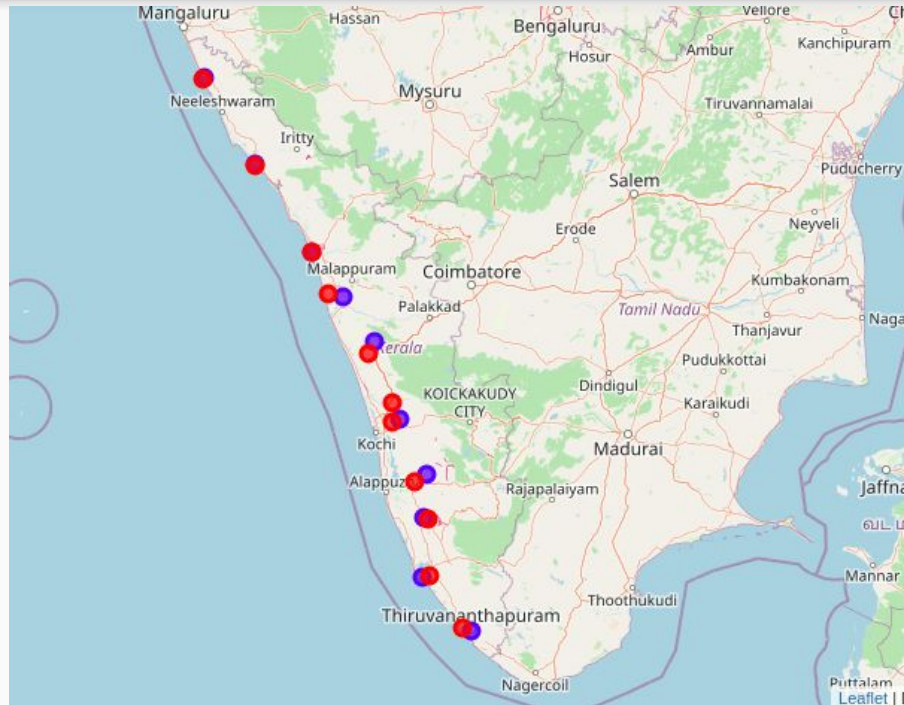
	District_Latitude	District_Longitude	Venue	Venue_Latitude	Venue_Longitude	Distance	Venue_Category	weight
Rail_District								
Chengannur	14	14	14	14	14	14	14	14
Ernakulam	39	39	39	39	39	39	39	39
Kannur	19	19	19	19	19	19	19	19
Kasargod	34	34	34	34	34	34	34	34
Kollam	15	15	15	15	15	15	15	15
Kottayam	16	16	16	16	16	16	16	16
Kozhikode	11	11	11	11	11	11	11	11
Shornur	17	17	17	17	17	17	17	17
Thiruvananthapuram	26	26	26	26	26	26	26	26
Thrissur	14	14	14	14	14	14	14	14
Tirur	9	9	9	9	9	9	9	9

Determining the number of Stations

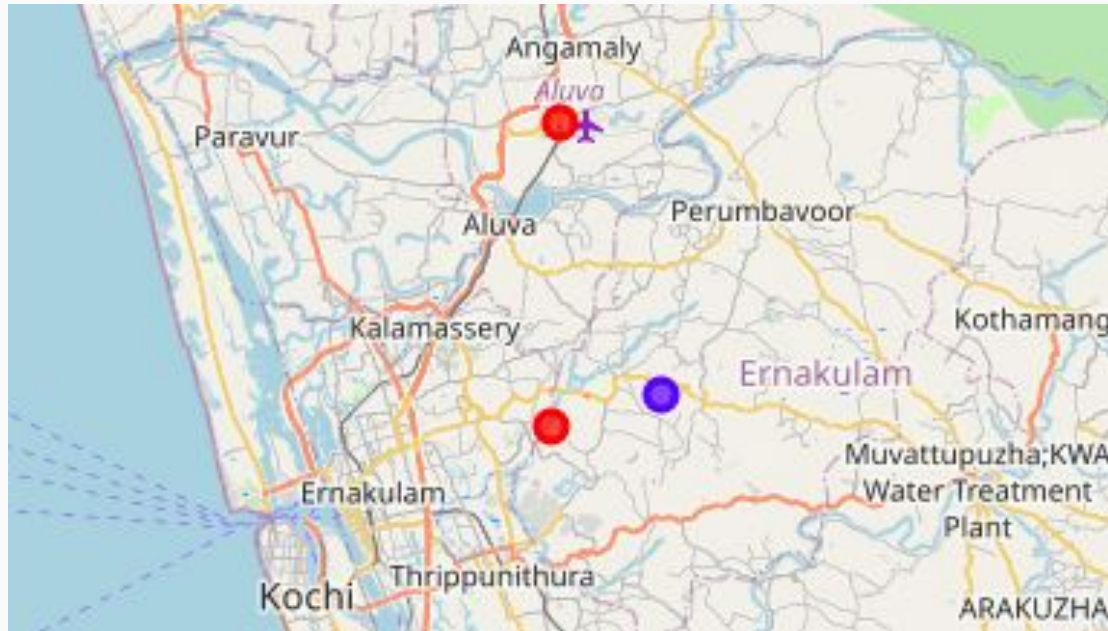


- Number of clusters, $k = 10$

Results of K-means Clustering



Discussions



- **Red: Current Station**
- **Blue: Suggested Station**

2 stations at very close distance in the district Ernakulam (stations by K-Rail team)

This project corrects it by a single station (blue circle)

Conclusions

- Popular locations across the state grouped into 10 clusters
- Each cluster with optimal connectivity
- No two stations in the same district