Suggesting Stations for the Silver Line project, Kerala

Sreenath B S

May 24, 2020

1. Introduction

1.1. Background

Kerala is a small state with a total population of 3.45 crore people. The existing railway network and highways in the state are not amenable to faster travel. Average speed on the roads and trains is among the lowest of all regions in the country. Due to excessive number of level crossings and sharp curves, the fastest train takes about 12 hours for traversing a total length of 560kms with an average speed of 45km/hr.The train traffic in the existing double line between Thiruvananthapuram and Kasaragod has increased manifold and some of the sections have capacity utilization of more than 115%. With the above in mind, the Government has decided to build a Semi High-Speed Rail (Silver Line) corridor. It is very important for the project to limit the number of stations and connect popular sites around the state by proper distribution of stations.

1.2. Problem

This project aims to limit the number of stations without compromise on the connectivity across popular sites around the state.

2. Data acquisition and cleaning

2.1. Data sources

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)

2.2. Data Cleaning

The data collected from foursquare api contained some missing values, which are added manually. The api returned several duplicate copies of the same location with different names and category values, which were dropped off. Some locations turned out to be in the neighbouring locations of multiple districts, which are handled appropriately based on the distance from the district centre.

2.3. Feature Selection

Inorder to give appropriate weightage to different categories of locations, a weight formula was devised. It takes into account the predefined weights of the category, distance to the location and its longitude from the minimum longitude of the district.

*weight_of_location = weight_of_category * (100000.0 / distance) * (x + 0.01)*

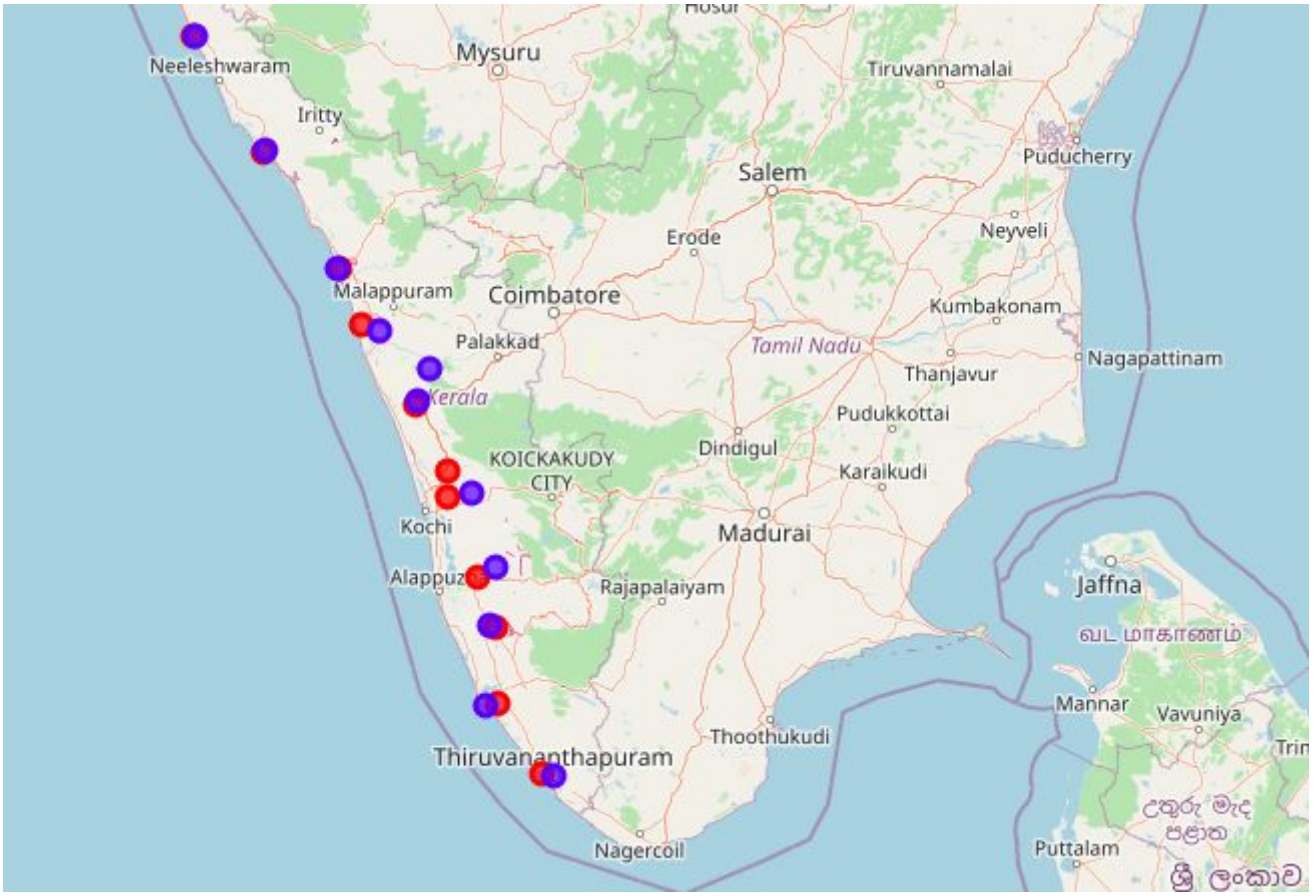
where:

weight_of_category = predefined weight of the category
distance = distance to the location from the district centre
x = (longitude < min_logitude + 1) ? 1 : 0

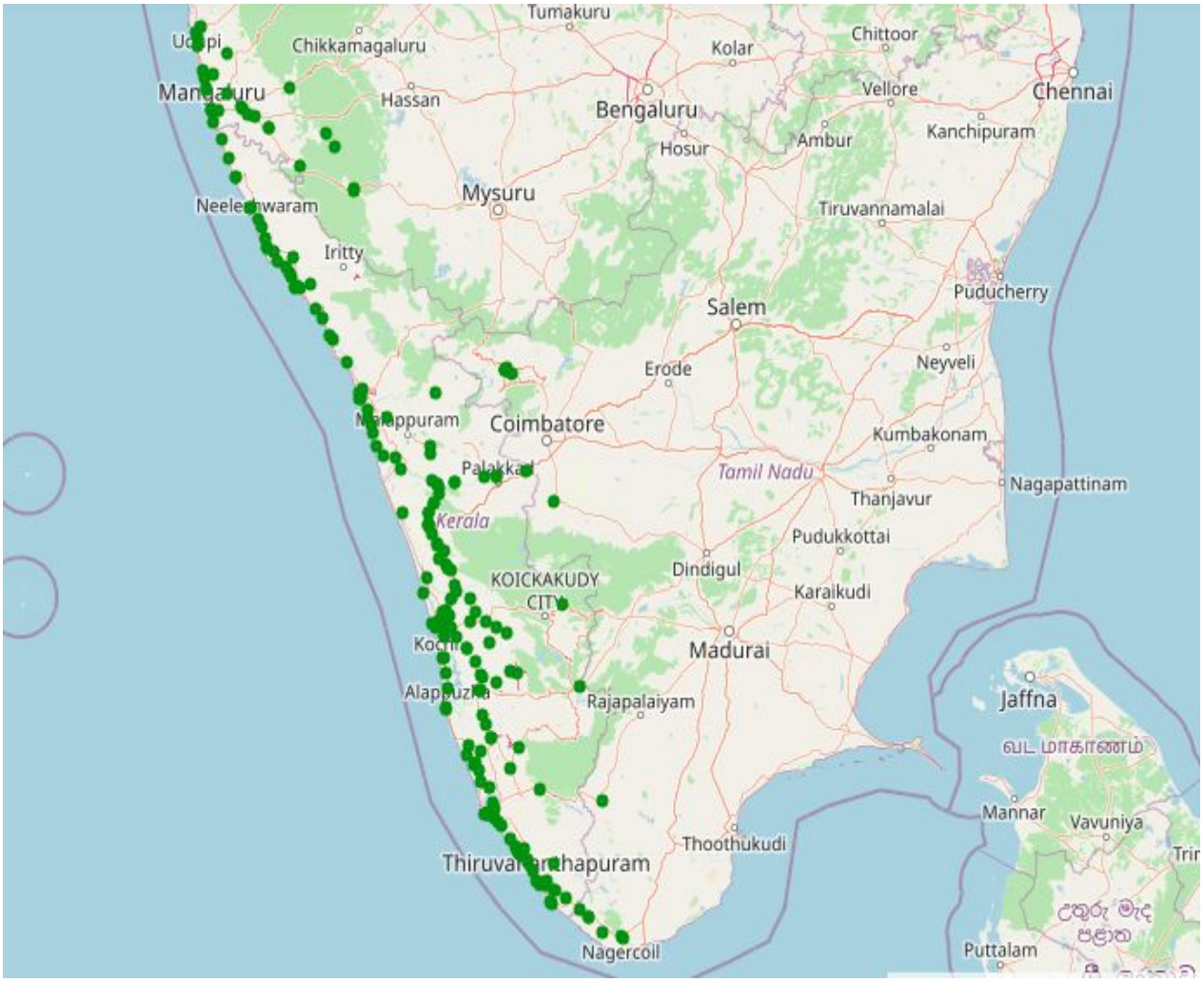
3. Exploratory Data Analysis

3.1. Plot of Current Stations vs District Centres

Red: Current Station
Blue: District Centre



3.2. Plot of popular sites



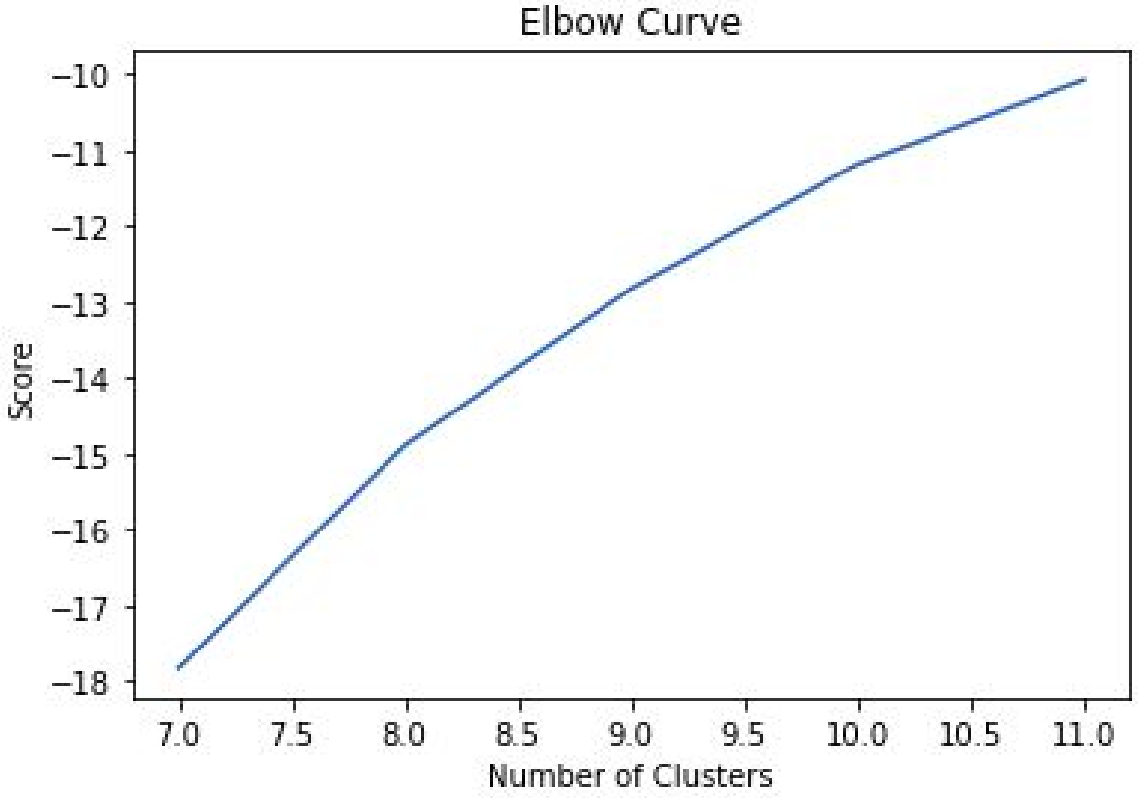
3.3. Explore Districts through which K-Rail runs

Rail_District	District_Latitude	District_Longitude	Venue	Venue_Latitude	Venue_Longitude	Distance	Venue_Category	weight
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

3.4. Determining the number of Stations

Elbow method can be used to determine the number of clusters in k-means clustering. For this project, minimising the number of clusters is one of the established aims. The plot below shows an elbow at around 10 in the x-axis. Therefore, the number of clusters is fixed as 10.

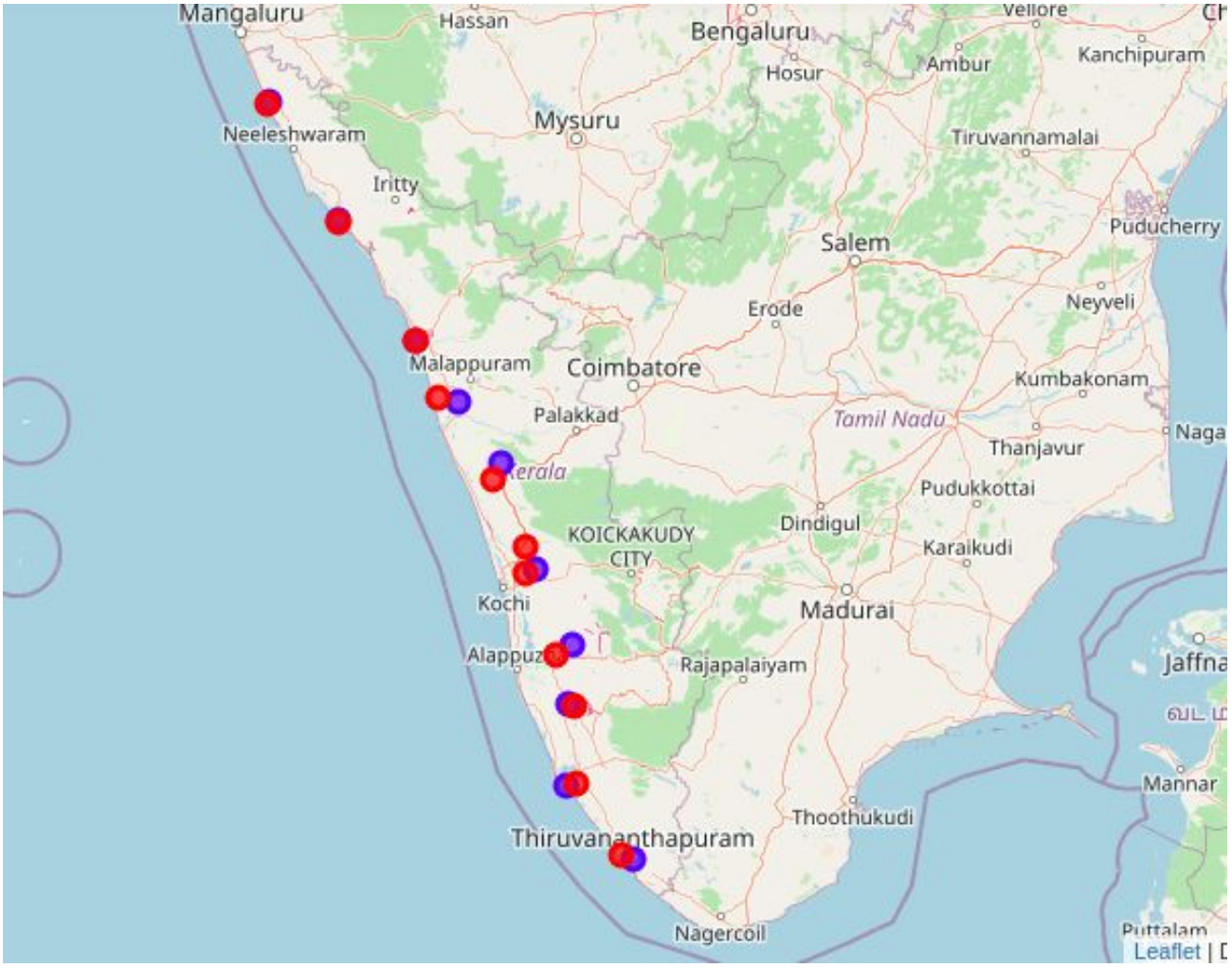
Since we have the number of clusters as 10, it is required to remove one district from the list. Upon comparing the locations with the alignment of the Silver Line project, Shornur is removed.



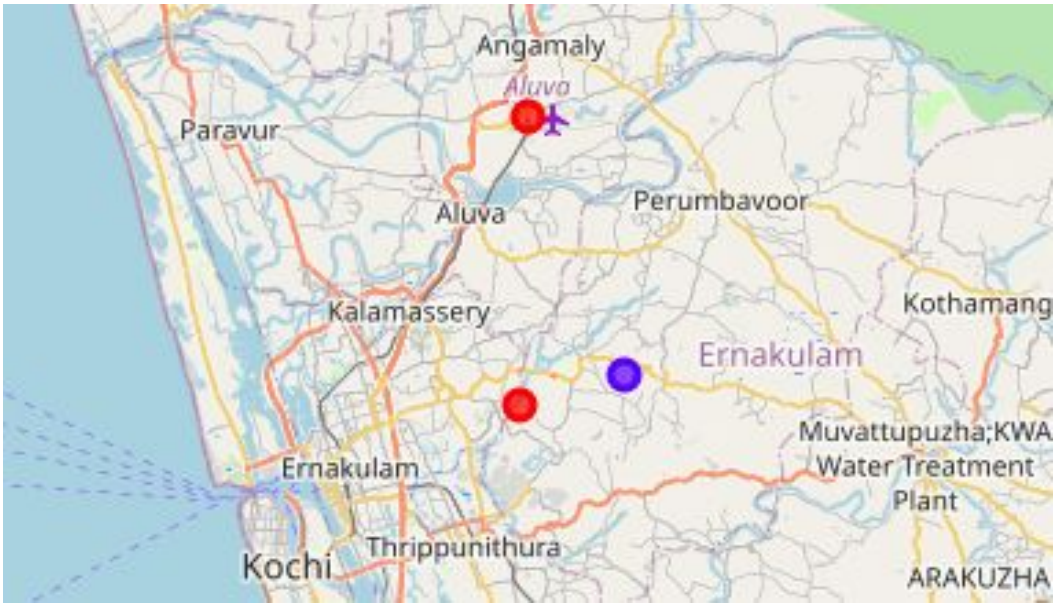
4. Results of K-means Clustering

Now that we have the number of clusters. It's now time to apply kmeans clustering to the data points in hand. Use the district centres as initial cluster centroids, and the calculated weights as the weightage of each record in clustering. Assign cluster names to each record and visualize the cluster centroids on the graph.

Red: Current Station
Blue: Suggested Station



5. Discussions



Stations proposed by the K-Rail team based on their aerial study has 2 nearby stations in the district Ernakulam (see the two red circles in Ernakulam district in the map above). This is an extra overhead to the system. The solution proposed here overcomes this issue by scientifically clustering the main locations throughout the state into 10 clusters. Also, the number of stations proposed by the team is one more than our suggestion, which is the one mentioned above. Following the clustering pattern in this project, this extra station could be removed, thus saving a lot of time and money.

6. Conclusions

In this study, I analyzed the popular locations in the state and clustered them into 10 groups. The centroid of each group is then identified as the location of the station. The stations are selected in such a way that connectivity to each cluster is maximised, giving proper preferences to airports, sea ports, railway stations and so on.