Finding Optimal Coordinate to Add New Hubway Stations

Riken Maharjan, Biken Maharjan

Introduction

Hubway is a bicycle sharing company where guest and subscribed users can lend a bike for certain amount of time. The purpose of this project is to utilize clustering algorithm like hierarchal and k-mean to find the optimal coordinates to place the new bike stations.

Datasets

1. We gathered Hubway station data, which showed the station present with longitude and latitude throughout Boston, and Cambridge.
2. We also collected the data, which showed total rides for three months.

Algorithm

Goal: Find coordinates where new bike station would be appropriate to install based on the popularity of older stations.
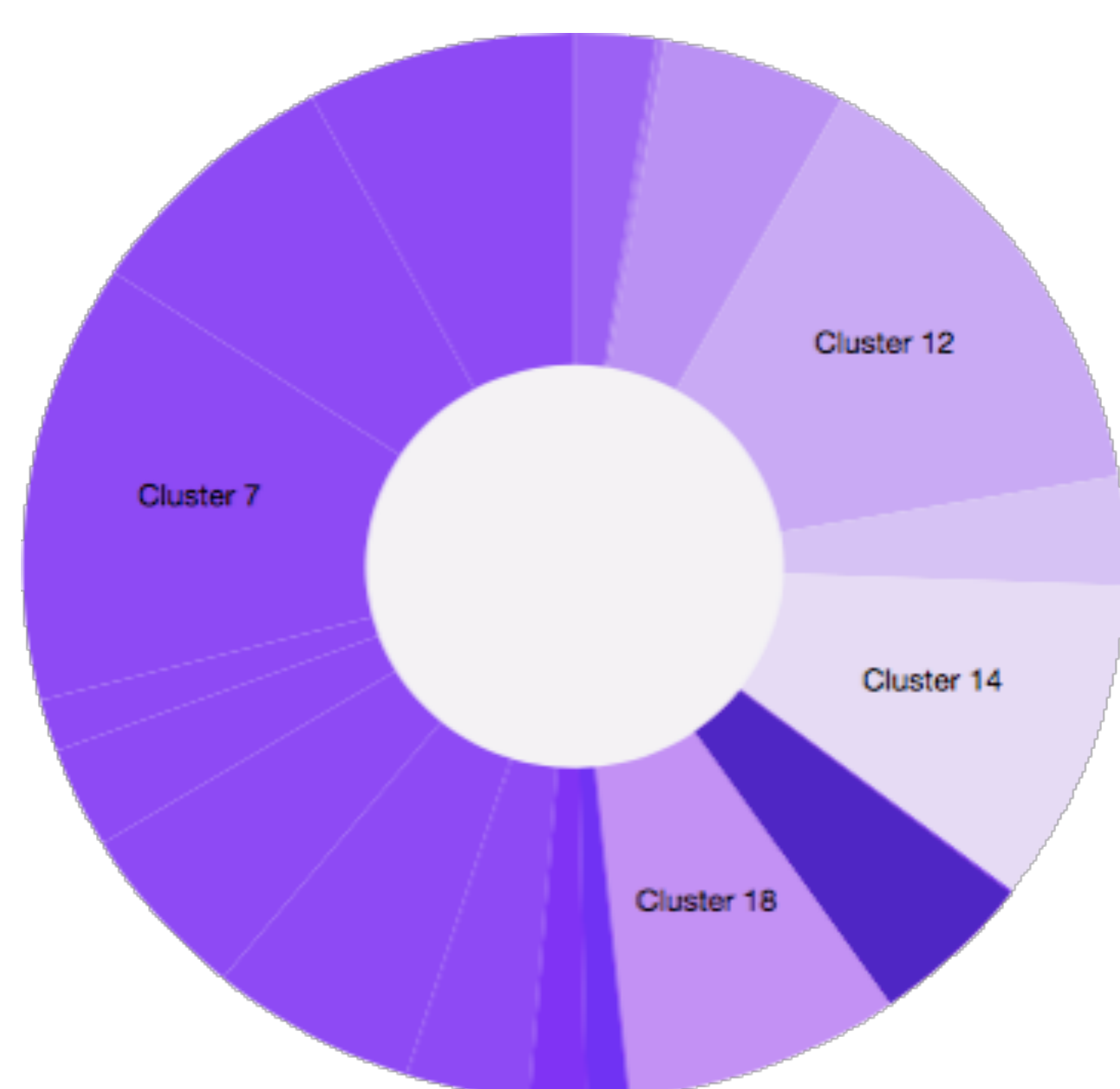

Fig 1.0 average popularity of all 20 clusters

- Used **Hierarchical clustering** to cluster data in into groups with respect to the distance.
- Calculated **average popularity** of the each cluster.
- Put the station at the center of the cluster with the lowest **average popularity**.
- Use **k-mean** in the cluster itself if more than one stop are to be installed in a particular cluster.
- Used **restful API to built web server** using python flask.
- Use frontend technologies **like D3 and javascript** to built an interactive website.
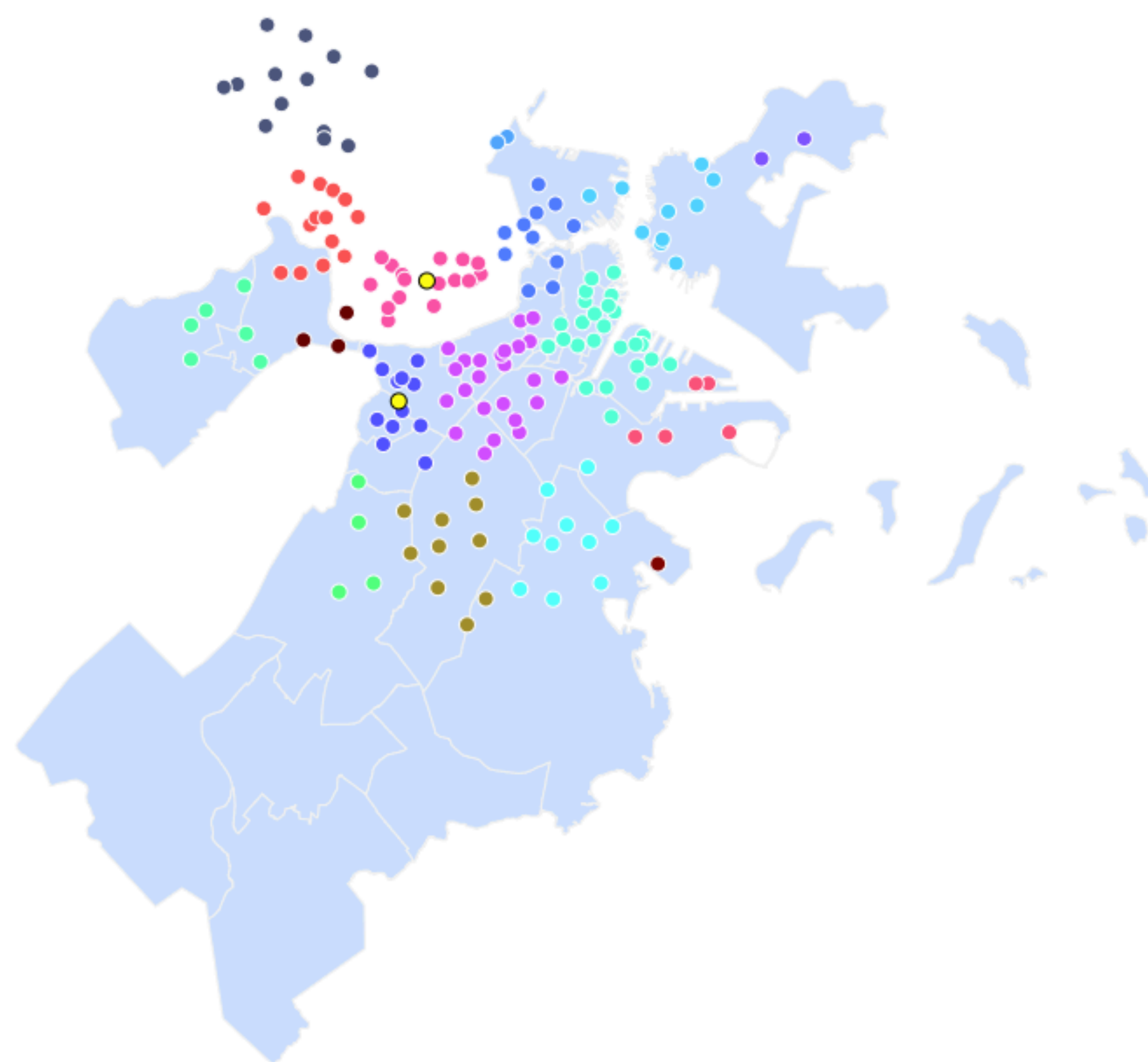
Results

Fig 2.0 : Cluster of Bike stations with added stations (Yellow points)

We found out that there are most likely 20 clusters present in the map using **Hierarchical clustering**. We used **Dendrogram** of this cluster to analyze there are 20 clusters most likely cluster present in the map. We did not use K mean here, as K needs to be decided even before clustering is done.
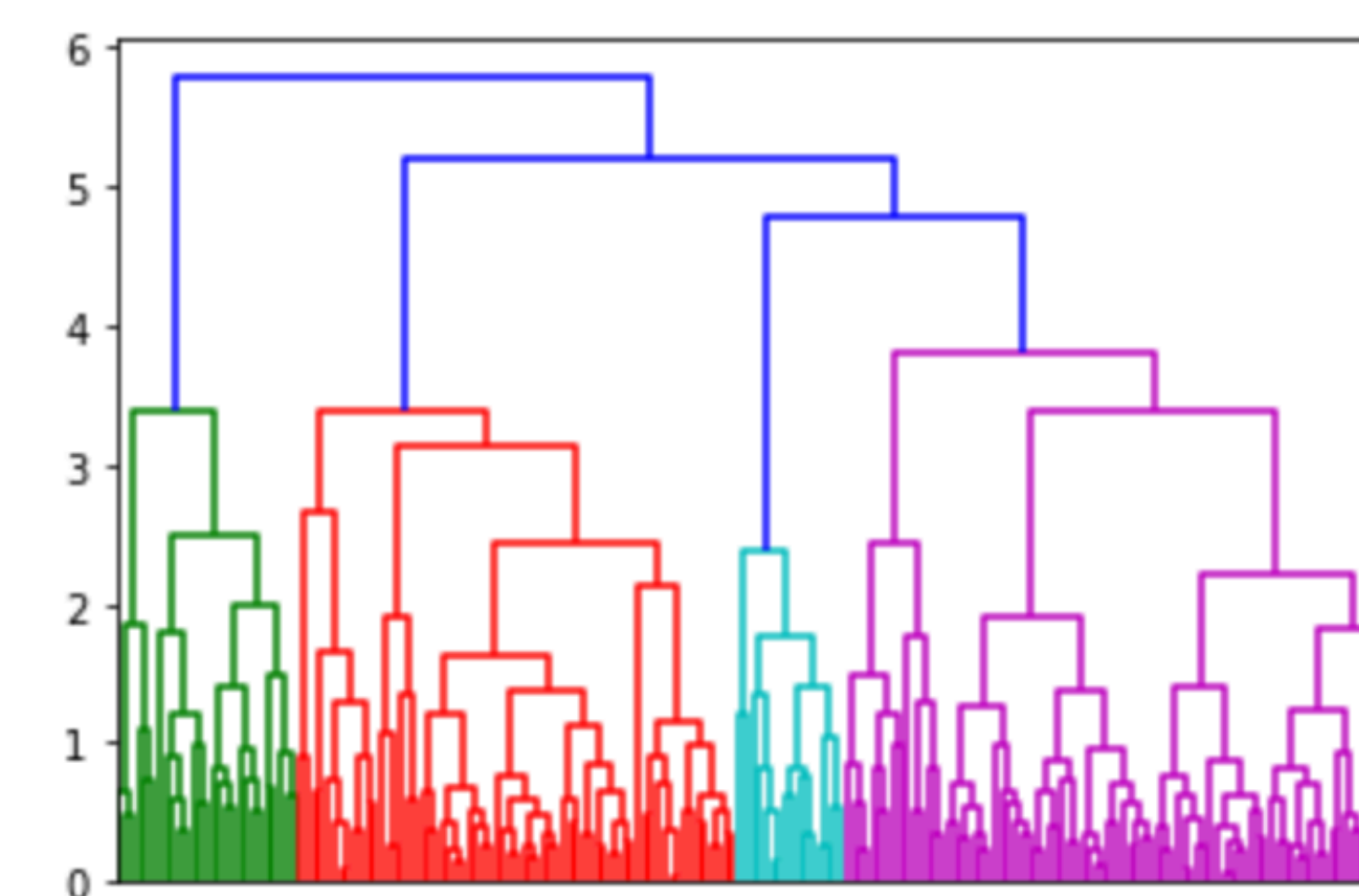

Fig 3.0: Dendrogram of our cluster

Conclusions

- Built a tool finding appropriate coordinates to put the new bike station in the city based on the popularity of the older stations using Hierarchical clustering and K-mean clustering
- Some of the cluster had very high popularity.

Future Work

Moving forward, there are many things that we can improve on, some of which are:
- Use algorithms like k-center to put stations in optimal places
- Expand datasets to include other factors like income, population instead of just popularity when clustering