



Yujia Wang  
yujia8@bu.edu

Yinzhu Su  
yzsu@bu.edu

Sang Jae Park  
park4s@bu.edu

## 1 Data Collection

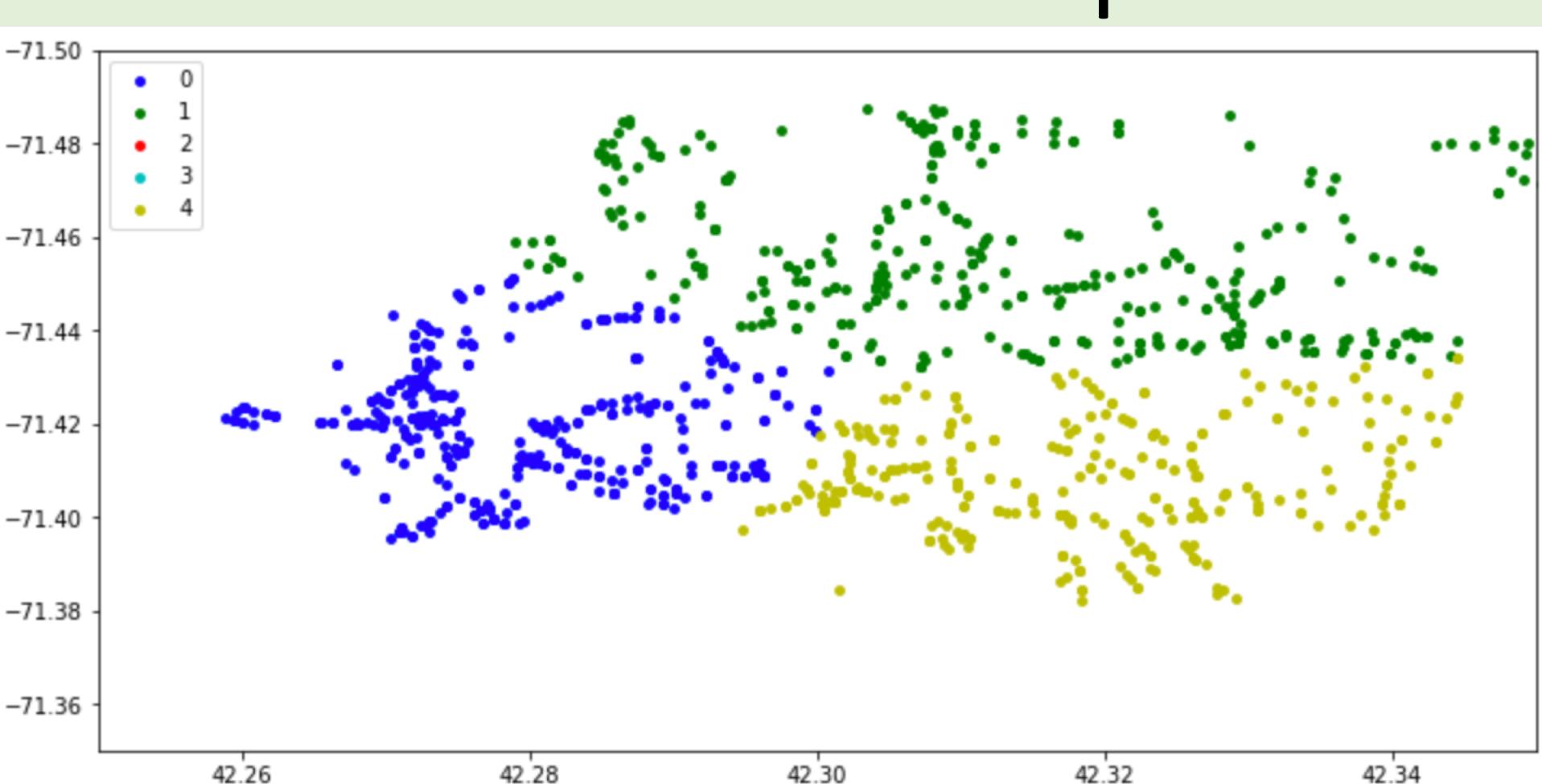
- 2017-2018 Framingham Bus Data
- X Public Schools in Framingham
- X Students in total
- Elementary, Middle, High

## 2 Data Cleaning

- Remove useless columns
- Filter rows whose data is not integrity
- Sort the data by school and time

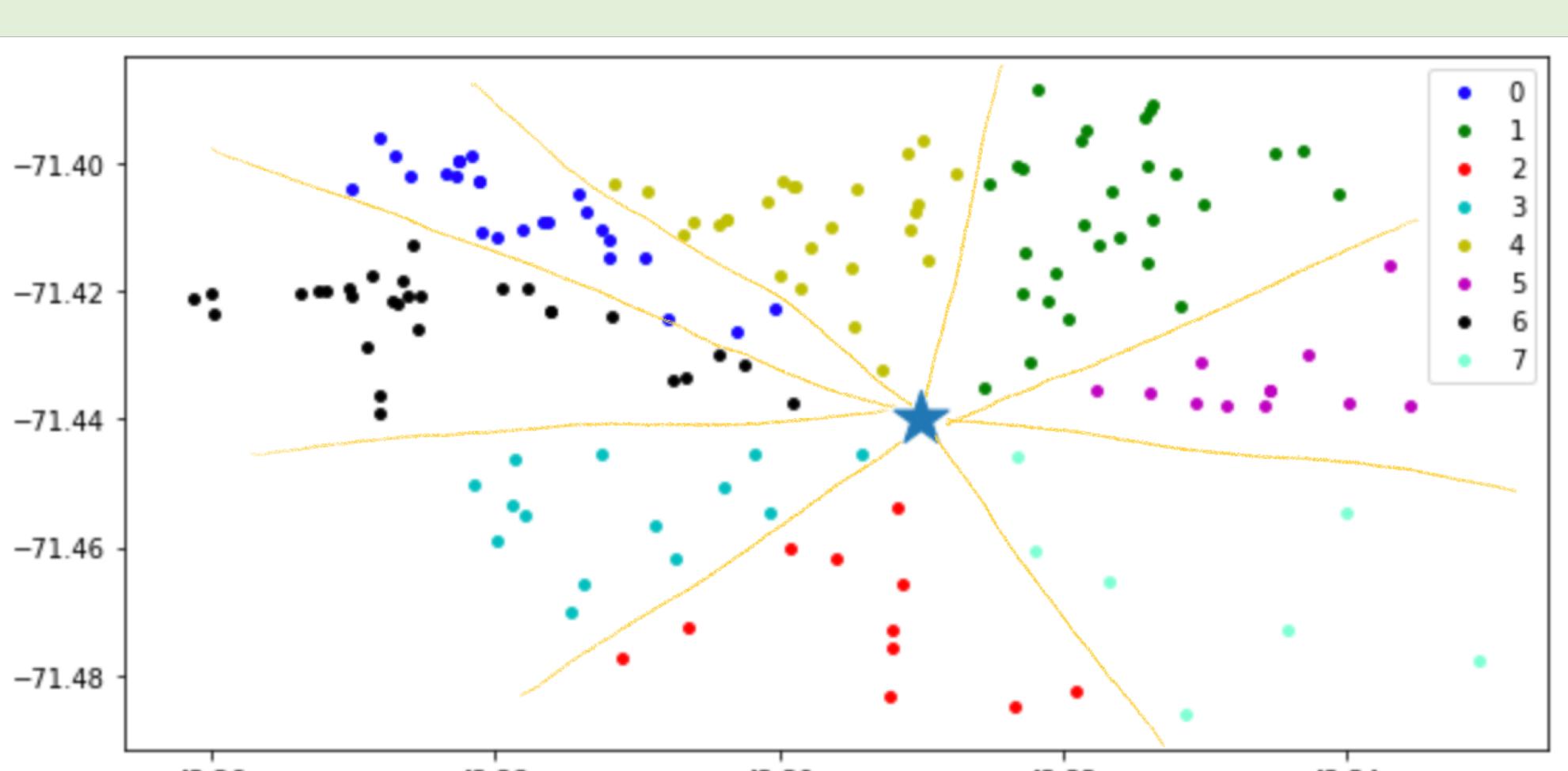
## 3 Stops Visualization

- Using Google Map API
- Get the latitude and longitude of all schools and stops
- Plot schools and stops



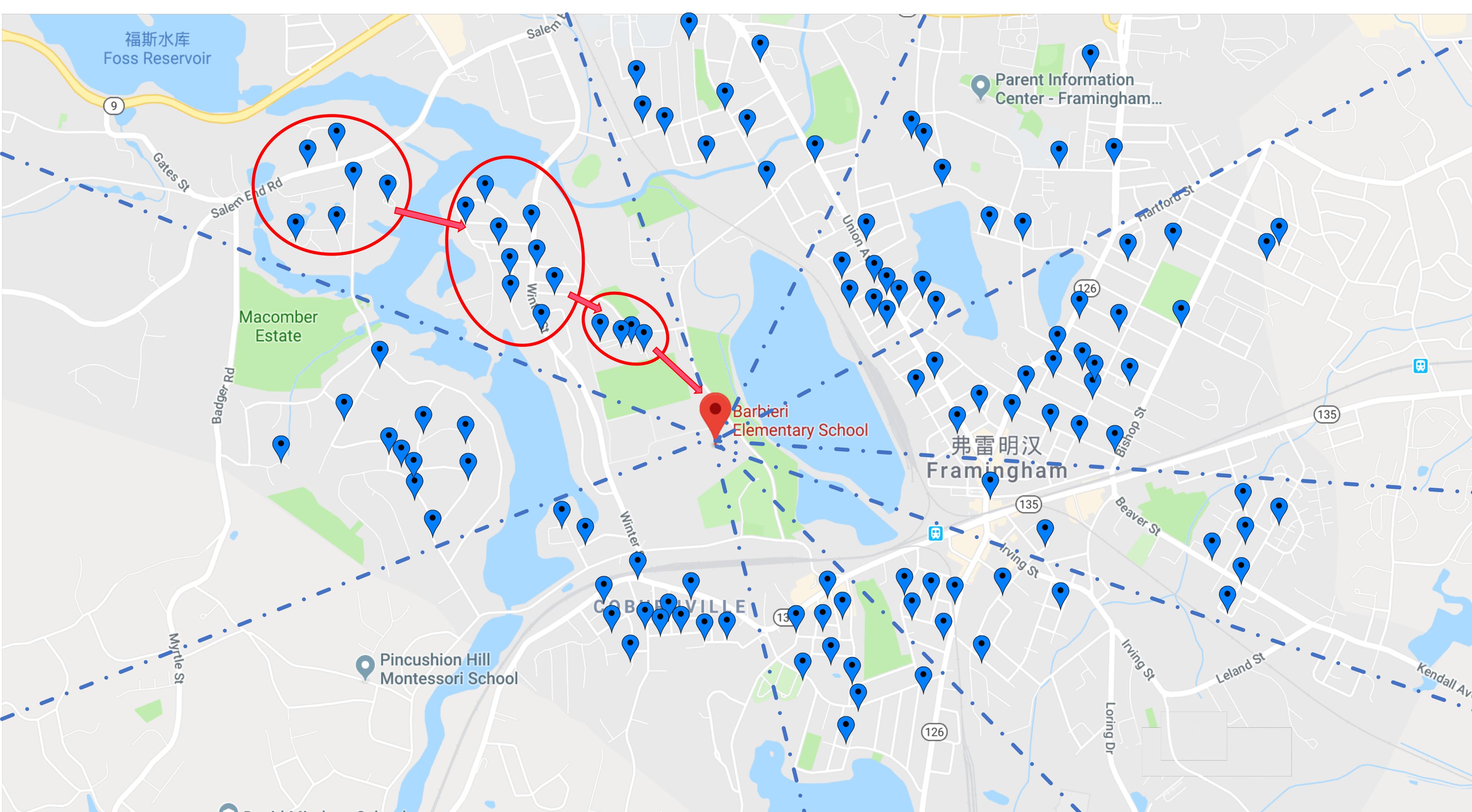
## 4 Make the Pizza!

- Using K-Means algorithm
- Cluster the Stops by angles of their Polar Coordinates
- School is in the central
- Just like a large piazza!



## MAIN GOALS

School buses are expensive to run, we need to find an efficient way to reform the bus routes for minimizing the quantity of buses needed in the overall system without sacrificing timely student arrivals to school and drop-off at home.



## MAIN TECHNIQUE

### K-Means Clustering

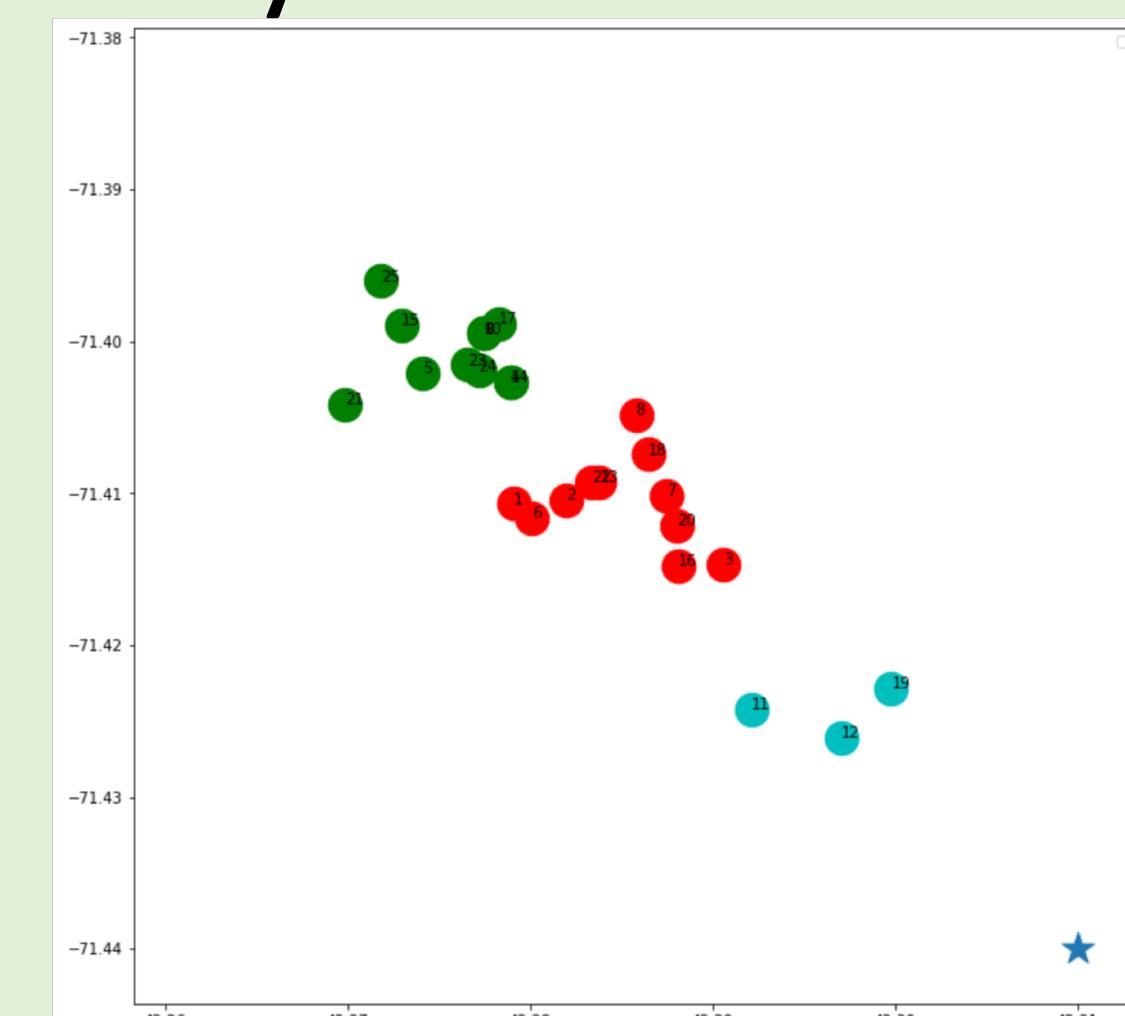
- K-means clustering is a type of unsupervised learning, which is used when you have unlabeled data (i.e., data without defined categories or groups). The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable  $K$ . The algorithm works iteratively to assign each data point to one of  $K$  groups based on the features that are provided.

### Google Maps API

- We mainly used two APIs in Google Maps APIs: Geocoding API and Distance Matrix API.
- The Geocoding is the process of converting addresses (like a street address) into geographic coordinates (like latitude and longitude), which you can use to place markers on a map, or position the map. Reverse geocoding is the process of converting geographic coordinates into a human-readable address.
- The Distance Matrix API is a service that provides travel distance and time for a matrix of origins and destinations. The API returns information based on the recommended route between start and end points, as calculated by the Google Maps API, and consists of rows containing duration and distance values for each pair.

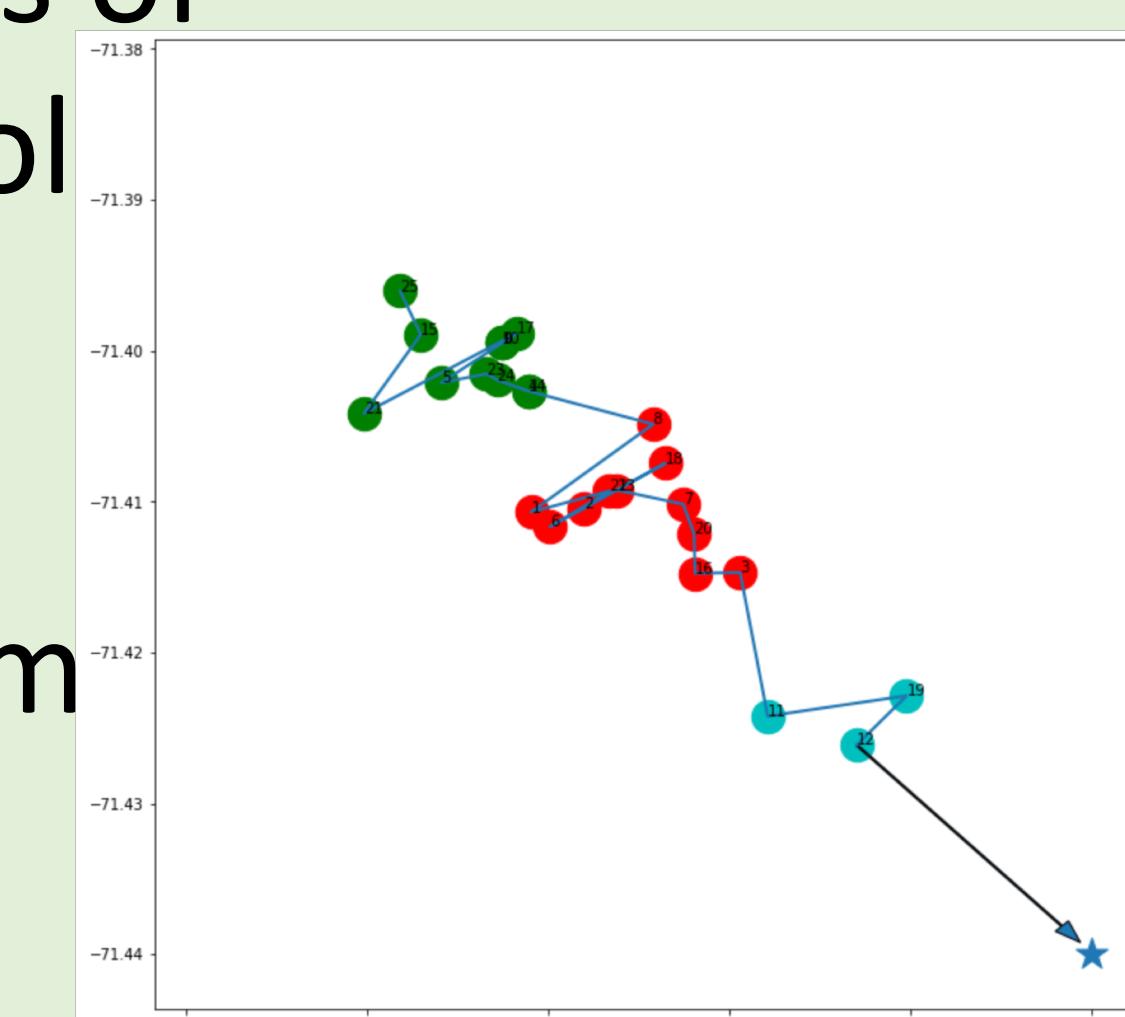
## 5 Cluster the Stops

- For one piece of pizza
- Cluster it by its Orthogonal Coordinates(3 clusters for each piece of pizza)
- Sort these three clusters by their central distance to school



## 6 Calculate the Route

- Calculate the distances of each stop to the school
- Sort the stops by the distances
- Connect the stops from the farthest to school



## 7 New Stops Schedule

- Using Google Map API to calculate the duration of driving between to stops
- Given the Drop Out time, reversely calculate each stops corresponding time

new_am_stptime
497
495
158
569
568
259
410
08:26:50
08:26:50
08:29:38
08:32:14
08:32:14
08:34:23
08:35:25

## 8 Rearrange the Buses

- Allocate the amount of buses for schools according to the amount of their students
- Arrange buses to each school according to their capacities