Boston Public School Transportation Challenge

School Bus Routing System Optimizations

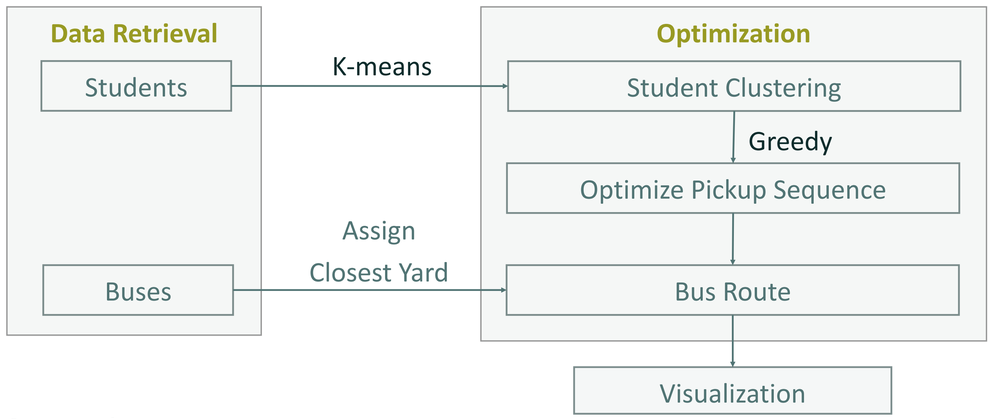
Lingyi Gu
Wei Wei
Haoyu(Jerry) Wu

OBJECTIVE

In 2016, transportation costs accounted for for **$110 million**. Among them, Boston Public School's transportation cost is the **#2** highest. Our team utilizes the **data mechanics** techniques to solve this challenge. Our main objectives include:

- **Address the inefficiencies and rising costs** in current transportation system
- **Improve the school bus routing efficiency** to decrease students riding time
- **Reduce operational cost** by systemically allocating resources

ALGORITHM OVERVIEW



Assumption:

- Students are picked up from their home location
- Each bus has unlimited capacity
- Each yard has unlimited bus
- The distance between two points is Euclidean distance

GROUPING STUDENTS

In order to make sure each bus has shortest route, it should pick up students that live near to each other.

- Group students by school using **map reduce paradigm**
- Implement the **k-mean clustering** algorithm to group the students based on their location

REFERENCE

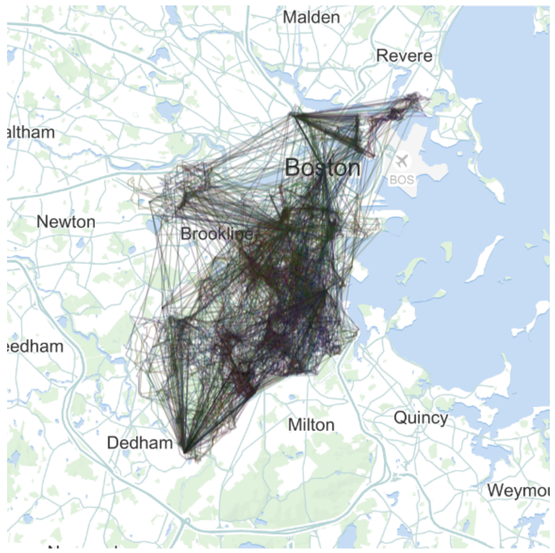
[1] K-means Clustering Algorithm http://cs-people.bu.edu/lapets/591/s.php#2.2
[2] Travelling Salesman Problem Solver https://github.com/dmishin/tsp-solver

OPTIMIZE BUS ROUTES

We propose three solutions to optimize the bus routes:

- **Prim's algorithm**: Find the Minimum Spanning Tree for each bus to pick all the students in that cluster. However, this algorithm may lead to buses go back and forth for every subtree and cause redundant traveling route
- **Dynamic Programming (DP)**: Find the shortest route among every permutation of all the possible routes. Although DP yields the optimal result, the running time is too high and grows exponentially when bus capacity increases. Since we assume that bus capacity is generally larger than 50, we don't choose the DP solution in the end
- **Greedy Algorithm (Nearest Neighbor)**: The algorithm randomly selects the starting point and consistently chooses nearest students as the next stop. This algorithm is the optimal solution for our project because it yields shorter routes than Prim's algorithm and a lower running time than DP

VISUALIZATION



LIMITATION & FUTURE WORK

- K-means algorithm **can't guarantee** that the number of students assigned to **each cluster is equal**
- Incorporate the **real road information** and expend the current solution by using **Dijkstra algorithm**
- Creating **gathering points** for nearby students to furtherly reduce bus stops and the bus waiting time