PROJECT - *Optimizing bus routes helps funnel cost savings back to schools*

INTRODUCTION

Public schools in the USA are often underfunded, and any money that can be redirected into core educational efforts is a boon for schools and students. Due to these pressures, Boston Public Schools (BPS) began to look for ways to reduce costs while improving educational outcomes. As part of this effort, the transportation department turned to its busing system to examine how it could make changes that could benefit the classroom, traffic and the budget. With an annual transportation budget of $120 million, compromising nearly 10% of the district's overall appropriation, any savings could have a significant effect [1]. BPS used Analytics to optimize its bus stops and bus routes, which resulted in huge savings. This report outlines the problem statement of BPS and a possible approach for an analytical solution that could solve the problem.

PROBLEM STATEMENT [1]

The oldest public school system in America, BPS operates 125 schools serving 57,000 students from pre-K through 12th grade. In 2016, the district provided transportation for 25,000 students via 650 buses across 45,000 miles. This adds up to 20,200 unique stops at nearly 5,000 locations each day. Students are transported in the morning and afternoon for close to an hour each day due to special classes, which was time consuming and added to the school's expense which was hard to maintain. The district used the same legacy bus stops year after year. The logistics behind these stops largely weren't adjusted over time, which was driving up costs. The district wanted to consolidate and streamline routes.

There are blanket rules when assigning bus stops laid by the district. For example, every student should walk less than half a mile to his or her bus stop. Student's age or neighborhood safety should also be considered while assigning them to stops. BPS required a solution that would automatically account for all of those factors without dramatically increasing costs.

ANALYTICAL SOLUTION OVERVIEW

In this section, I am giving an overview of a possible analytical solution to tackle the problem. The solution is discussed in detail in the upcoming sections of this report.

In order to make the transportation system for BPS more efficient, the first step would be to find an optimal way to place bus stops and assignment of students to it. Ideally, having minimum stops, yet covering all the students would greatly help in avoiding multiple stops, and reducing travel time. Secondly, if we have the optimal locations for bus stops, identifying the shortest route along the bus stops to the school would help in faster transportation and reducing fuel costs. In both cases, the safety of the children, and district associated blanket rules have to be accounted for.

MODEL 1: CLASSIFICATION OF SAFE NEIGHBORHOOD

For finding optimal locations for bus stops, we have to make sure that a specific neighborhood is safe. We could build a classification model that classifies if a new neighborhood is safe or not. Important data points to build this model would include below:

Data Collection:

- District crime statistics in a specific neighborhood that comes under BPS's radius including count of crimes, incident time, location, etc. Data can be gathered from reports that are accessible and allowed to be downloaded. [2].
- Average age of people living in a neighborhood. Census information would be a good start to gather this data [3].
- Number of people per household in the area.[4]
- Number of businesses in the area.[4]
- Zip codes of the zones in the neighborhood.
- Label indicating collected neighborhood safety

References indicate websites where data can be collected from. For the label data points, we could generate labels based on a threshold of crime rates in the regions and use these labels for training the model.

Given above data points, we can **use** a K-nearest neighbor classification (KNN) model **to** classify if a specific zone (identified by a postal code) is safe or not. As we know, KNN is a supervised machine learning algorithm, that identifies the k nearest neighbors that are similar to classify a label. In our scenario, we should start with a minimum of 5 nearest locations to consider the safety of a neighborhood. Based on the data provided above, the model will be trained to classify if a location is safe or not. This model can be used on all other neighborhood's that will have schools operated by BPS. This information would be used further in the other models.

MODEL 2: CLUSTERING OF STUDENTS

In order to identify optimal locations for bus stops, we should first group the students that are closely located to one another and go to same schools. This could be visualized as community of interconnected population. For this, we should create a network combining the zonal data of all students to form a community of regions around BPS. This can be done by using shortest distance path algorithms like Dijkstra's algorithm, which will get the different zonal codes and create paths connecting them with the shortest distance.

Data Collection: (Can be obtained from BPS school databases)

- Student Address (Zip codes/ Lat/Long)
- School Location (Zip codes/Lat/Long)
- School that student attends
- Any special courses attended at different schools
- Student Age
- Network of connected zones generated using distance algorithm mentioned above.

Given above data points, a Louvain Algorithm can be **used to cluster** and build a network of highly interconnected students and schools. Louvain Algorithm is a hierarchical clustering algorithm to detect communities in large networks. It tries to maximize the modularity, which quantifies the amount of similarity or connection between the nodes. In our case, with the different student data points above, Louvain algorithm will be used to find small interconnected student clusters within nearby schools that can be further used in the optimization model.

MODEL 3: OPTIMIZATION MODEL FOR LOCATING BUS STOPS

a) For the student communities identified, it will be ideal to place one bus stop at each interconnected community and assign students in that community to that specific bus stop. However, we also have to consider constraints like the maximum distance a student can walk to the stop, safety etc. This can be achieved by building an optimization model with the data available to minimize the placement of bus stops without affecting the transportation of students.

Data Collection:

- Network of interconnected students with respective location data (from previous step)
- Zip codes that are classified as safe (from the first classification model)
- Maximum distance allowed for students to walk (data from BPS according to district rules)
- Distance between a node(student) to a bus-stop location (a node within the community) (the same distance algorithm employed in Model 2 can be used to calculate distances).

Given the above data elements, we can **build** an optimization model **to** minimize the number of bus stops allocated while assigning every student to a nearest bus stop. The model will take below inputs:

Variables:

- A binary variable that can be used to choose a location(node) as bus stop or not.
- A binary variable to indicate if a student is assigned to bus stop.

Constraints

- Distance from student to bus stop should be lesser than the maximum distance allowed from student's location to stop
- Location of the stop should be classified as a safe location
- Every student should be assigned to one stop
- Every bus stop should have at least one student

Objective function

To minimize the total number of bus stops assigned in the community.
This optimization model will identify the list of optimal bus stop placement locations ensuring that all students are covered. The output of this model will be using in the next model.

MODEL 4: OPTIMIZATION MODEL TO IDENTIFY BEST ROUTE TO SCHOOL

After identifying optimal locations for bus stops, and assigning students to stops, we should build a model to find the best routes to reach the school by reducing travel time. BPS has mentioned that buses run in the morning and afternoon, so our model should consider both times of the day.

a) The bus stop locations, and schools can now be visualized as an interconnected community that could be clustered within a larger community.

Data Collection:

- Bus stop location data (from previous model)
- School locations (from BPS database)
- Distance between nodes(bus-stops/school) (can be calculated using distance algorithms).
- Time of travel (can be approximated from distance and average speed of allowed in school community).
- Available Number of buses for mornings/afternoon (data from BPS).

Given above data points, we can **use** Louvain Algorithm **to cluster** the bus stops and schools in the form of an interconnected community.

b) With the above network data, we can build an optimization model that would minimize the distance of each bus route.

Variables:

- A binary variable to choose a bus-stop(node) as a starting point
- A variable to represent the number of bus stops added to the route
- Distance from starting point to school
- Time of day (Morning/Afternoon)

Constraints:

- Bus stop should be in the route only once
- Route should not exceed maximum travel time (defined by BPS).
- Number of routes should not exceed maximum available buses for that time of the day.

Objective function:

Minimize the distance of route from starting point to school.

This objective function will identify the optimal routes connecting the best stops and ensuring that students are transported to their respective schools.

DATA REFRESH/ SIMULATION TO TEST THE MODELS

Since the proposed solution would create a huge impact on the daily bus schedule and transportation of students, the models have to be tested multiple times to measure possible delays in the route etc. A simulation model has to be built which would get the following inputs and access the behavior of the proposed system.

- Different number of bus stops that can be allocated
- Modifying constraints like maximum distance that a student can walk
- Changing maximum time allowed in travel.
- Changing availability of buses

The simulation model has to be run multiple times to identify delays or issues in the system that has to be corrected.

Also, data from BPS has to be collected yearly to account for new students/ students leaving schools every year and simulations have to be run to ensure that the simulations generate the expected outcome.

Using an analytical approach to optimizing the bus-stops and routes, BPS would be able to reduce their transportation expenses (reduction in number of buses running, fuel costs etc.), while safely transporting students to classes and their homes on time.

REFERENCES

[1] https://www.sas.com/en_us/customers/boston-public-schools.html

[2] https://data.boston.gov/dataset/crime-incident-reports-august-2015-to-date-source-new-system

[3] https://datausa.io/profile/geo/boston-ma/

[4] https://data.boston.gov/dataset/certified-business-directory/resource/3fc08ca2-9baf-4d77-b03a-aaed1cc936ed