Pittsburgh Regional Transit Bus Scheduling

Annie Xie[1]
fumanx@andrew.cmu.edu

Marisa Yang[2]
yujuy@andrew.cmu.edu

Yuxuan Liu[3]
yuxuanl4@andrew.cmu.edu

December 8, 2023

[1]Conception, Data Cleaning, Optimization
[2]Conception, Data Cleaning, Optimization
[3]Optiguide, Optimization

Executive Summary

This report presents an approach to optimizing Pittsburgh Regional Transit's (PRT) bus scheduling system. The primary objective is to maximize profit and ridership numbers without compromising cost efficiency, an essential factor considering the significant government funding involved in public transportation. Utilizing datasets from the Western Pennsylvania Regional Data Center and the American Community Survey, we focus on evaluating bus stop usage, scheduled trip counts, and demographic data such as poverty and minority rates within Allegheny County at a census-tract level.

The project includes analyzing 95 routes and 3839 inbound stops to optimize bus schedules. Our model introduces decision variables X_{i,j_i,t_i} to determine whether each route stops at specific stops during its trip rides. The approach integrates constraints to 1) ensure that every stop is served at least once and 2) place additional emphasis on serving stops in areas with higher poverty and minority rates.

In summary, despite some limitations, our findings provide a strategic foundation for iteratively adjusting PRT's bus schedules on a seasonal basis, taking into account ridership in the previous season while focusing on efficiency, profitability, and equitable access, particularly for disadvantaged communities. We hope our work, incorporated with Optiguide, provides insightful policy implications, contributing to future public transit planning.

1 Problem Definition

Public transportation is an essential part of public infrastructure. With data provided by Pittsburgh Regional Transit (PRT), we want to optimize the current bus schedule. To be more specific, given the number of passengers at each stop on each route, we aim to determine whether the bus stop should stop at each stop throughout the day on a quarterly basis. We have also noticed that the bus routes have been modified due to the city's new construction, as well as service hours have been reduced, and ridership has dropped due to the Pandemic. All these factors contribute to the complexity of bus scheduling.

Furthermore, the Port Authority of Allegheny County emphasizes the importance of enhancing mobility for disadvantaged communities. This is also included in our project as we aim to extend beyond mere efficiency and cost reduction and emphasize the necessity to create bus schedules that are inclusive to all community members, particularly the minority and low-income groups who are highly dependent on public transportation.

The results of our project can serve as the foundation for iterative updates: merging less frequently used stops, merging shared visited stops used between routes and rerouting according to efficiency, and deleting redundant stops or routes.

2 Datasets

We will be using data from Western Pennsylvania Regional Data Center and American Community Survey(ACS). The dataset we are using includes all bus stops, active bus routes in PRT. We used data from the most recent quarter. For ACS, we used the API provided and accessed the census-tract-level demographics in the Allegheny County.

- Pittsburgh Regional Transit Bus Stop Usage [1] : $R_{i,j}$, J_i
- Pittsburgh Regional Transit Scheduled Trip Counts [2] :T_i, $Time_i$, I
- American Community Survey(ACS)[3]: Pov_j, Min_j

Below is a list of parameters that we will collect or estimate from our datasets and will be used in our model.

Parameter Definitions

C The operating costs per minute for rides

I The total number of routes

J_i The total number of stops for route i

Min_j The minority rate of the census tract where stop j is located, calculated by non-white population / total population in the census tract

Pov_j The poverty rate of the census tract where stop j is located, calculated by low income population / total population in the census tract

R_{i,j_i,t_i} The average number of riders on route i at bus stop j_i at the t_i-th ride

T_i The total number of rides each day for route i

$Time_i$ The total operation time for route i

Our decision variables are X_{i,j_i,t_i}: a list of binary variables that indicate whether route i should stop at bus stop j_i at the t_i-th ride.

In total, we have 95 routes and 3839 inbound stops as shown in Figure 1 with poverty rate and minority rate in the base map. We have 226783 decision variables.

[1] https://data.wprdc.org/dataset/prt-transit-stop-usage
[2] https://data.wprdc.org/dataset/prt-scheduled-trip-counts
[3] https://www.census.gov/data/developers/data-sets.html

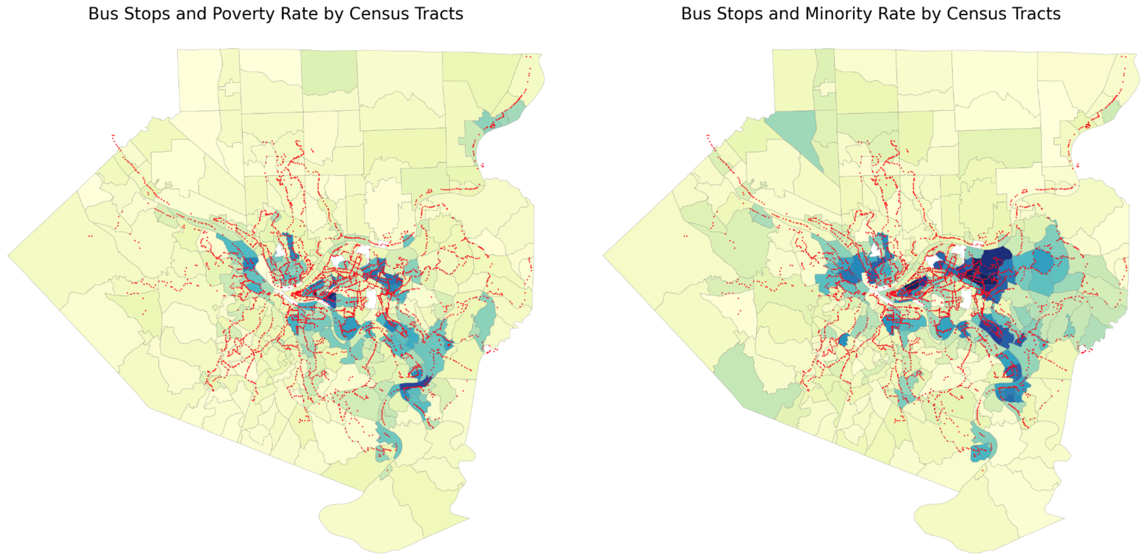

Figure 1: Bus Stop Distribution

3 Optimization Formulation

Objective function: To maximize profit or the number of riders without sacrificing cost efficiency, considering that public transportation is heavily funded by the government.

$$\text{Maximize} \sum_{i=1}^{I}\sum_{j_i=1}^{J_i}\sum_{t_i=1}^{T_i} R_{i,j_i,t_i}X_{i,j_i,t_i} * price - \sum_{i=1}^{I}\sum_{t_i=1}^{T_i}(\frac{\sum_{j_i=1}^{J_i} X_{i,j_i,t_i}}{J_i}) * C * Time_i$$

Decision variables: X_{i,j_i,t_i} for each route i at each stop j_i and each trip t_i.
Model specifications: An essential component of our model is to estimate the number of riders that go on route i at stop j_i on the t_i-th trip based on available data.

$$R_{i,j_i,t_i} = \frac{R_{i,j_i} * p}{T_i/5}\text{,where}$$

$$\begin{aligned} p &= 1/10, \text{ if } \frac{t_i}{T_i} < 0.2 \text{ or } > 0.8 \\ &= 2/10, \text{ if } 0.4 < \frac{t_i}{T_i} < 0.6 \\ &= 3/10, \text{ for other } \frac{t_i}{T_i} \end{aligned}$$

We assume that buses are dispatched at regular intervals, divided into five daily periods. The number of riders in these five periods (#1, #2, #3, #4, #5) are proportioned as 1, 3, 2, 3, and 1, respectively, to reflect morning and evening rush hours. Say that throughout the day there are 20 inbound trips for 61C. Then in each period, there will be 20 * 1/5 = 4 rides. And we have estimated there will be 40 riders that go on 61C at stop j. Then in period #1, there will be 40 * 1 / (1+3+2+3+1) = 4 riders, which means about 1 rider per ride. In period #2, there will be 40 * 3/10 = 12 riders, which means about 3 riders per ride.
Constraints:

1. Each stop must be passed by for at least once, i.e., for each stop j, $\sum_i \sum_{t_i=1}^{T_i} X_{i,j,t_i} > 0$

2. If a stop is located in a census tract with a poverty rate higher than 10%, it must be passed by for at least 10 times, i.e.,

$$\text{If } Pov_j > 0.1, \sum_i \sum_{t_i=1}^{T_i} X_{i,j,t_i} >= 10$$

3. If a stop is located in a census tract with a minority rate higher than 20%, it must be passed by for at least 10 times, i.e.,

$$\text{If } Min_j > 0.2, \sum_{i} \sum_{t_i=1}^{T_i} X_{i,j,t_i} >= 10$$

Assumptions:

1. We assume buses are dispatched at regular, fixed intervals.
2. We assume that the total operating hours of the bus are directly proportional to the number of stops it makes.
3. We assume that for stops that are visited in both inbound and outbound trips, half of the people are going inbound and the other half are going outbound.
4. We assume that the trip counts recorded in the dataset are the sum of inbound and outbound trips and should be evenly distributed between two types.

Results and Policy Implications:

For detailed results, please see our jupyter notebooks due to the page limits. Based on our optimization, we calculated the skip-rates for each route and the number of passes for each bus stop. We also provide a visualization for stop probability for a selected route as shown below. PRT can then quickly identify bus routes/stops that may need to be adjusted and make adjustments accordingly. With the help of Optiguide and an iterative approach, PRT can adjust the model with few efforts, which makes sure that they can provide the most needed services in a timely manner even with changing objectives and constraints.

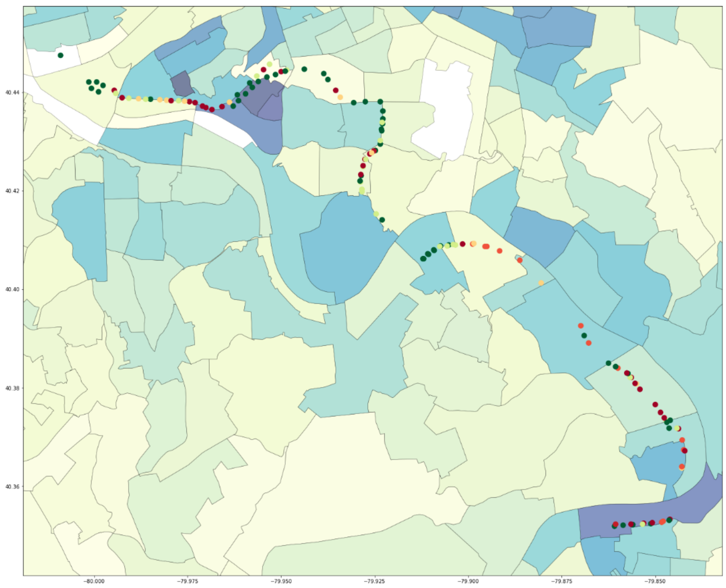

Figure 2: Stop Probability for 61C

Limitations:

1. Although PRT collects data regularly, the latest data we have access to were collected in April, 2021. So our results cannot reflect the current situation.
2. We only considered inbound trips in our analysis. We can do the same thing for outbound trips.
3. We currently lack a defined approach for determining our C. Given additional time, we could analyze the financial statements to develop a more accurate estimation.

4 OptiGuide

4.1 OptiGuide Q&A

Here are a few examples of answers given by LLMs after the Pittsburgh Regional Transit (PRT) model was optimized through OptiGuide. Except for some discrepancies between the answers provided by OptiGuide and the expected results on individual questions, the overall results were in line with the original experimental expectations.

Question1: If every stop has to go through at least two routes?

Answer1: The new constraint, which requires each stop to be passed by at least two routes, causes the model to be infeasible. In other words, it's not possible to satisfy all of the constraints at the same time. Looking at the conflicting constraints, one of them is 'minimum_route_constraint_E02430'. This indicates that it's not possible for stop 'E02430' to be passed by at least two routes. This issue could be due to a number of reasons, such as the stop being located in a route that isn't covered by at least two different transport routes. It's important to note that the new results differ from the original ones where the model was feasible with an optimal solution. The addition of the new constraint changed the problem from a feasible to an infeasible one.

Question2: What is the potential impact of a 29% increase in the proportion of poor locations?

Answer2: The new constraint, which requires each stop to be passed by at least two routes, causes the model to be infeasible. In other words, it's not possible to satisfy all of the constraints at the same time.Looking at the conflicting constraints, one of them is 'minimum_route_constraint_E02430'. This indicates that it's not possible for stop 'E02430' to be passed by at least two routes. This issue could be due to a number of reasons, such as the stop being located in a route that isn't covered by at least two different transport routes. It's important to note that the new results differ from the original ones where the model was feasible with an optimal solution. The addition of the new constraint changed the problem from a feasible to an infeasible one.

Question3: What impacts would occur if stops with minority rates higher than 20% and local poverty rates higher than 9% must be passed at least once?

Answer3: Based on the retrieved results, the optimal number of parking stops skipped with a threshold of local poverty rate at 5% was 43.38%. This is actually the same as the percentage when the poverty rate was set at 10%.It is evident the number of stops that are passed only once or less than three times is the same as the previous result, which was 27 and 103 respectively. This suggests that adjusting the local poverty rate from 10% to 5% didn't change the outcomes of the model optimization significantly. To put in other words, the number of bus stops that were skipped and the frequency of bus stops being passed by remained the same.Such result could be interpreted as that the initial model was already accounting well for areas with higher poverty rates, and the change in poverty threshold did not introduce substantial adjustments to the bus routing schedule.

Question4: If the local poverty rate is higher than 5% and each stop must pass at least 10 times, what are the potential impacts?

Answer4: If the local poverty rate is increased to 5% and each stop must pass at least 10 times, the model will add more constraints to it. This could potential leads to fewer stops being skipped and affect the maximum revenue of the bus scheduling. From the execution result, we can see that the optimal number of parking stops skipped is 43.38%, which is unchanged compared to the previous model. This means that, even by considering the higher poverty rate and adjusting each stop to pass at least 10 times, the optimized model could still maintain the same number of stopped being skipped which maintain the bus efficiency.For the percentage of stops being skipped on each route, we can also see the similar pattern, where each route still got almost the same percentage of stops being skipped. For the stops that only passed once or less than three times, the numbers are 27 and 103 respectively, which are unchanged from the original model. So, based on the result, the change of local poverty rate does not have significant impact on the bus scheduling optimization. This could be due to the flexible bus scheduling model and the data distribution that can adapt to these changes. However, by increasing the local poverty rate, we might potential provide better service to the poor area without decreasing the efficiency or the revenue of the bus scheduling.

Question5: What impacts would occur if stops with minority rates higher than 20% and local poverty rates higher than 9% must be passed at least once?

Answer5: The optimization result after changing the constraints shows that the optimal number of stops to be skipped decreases slightly from 43.38% to 43.26%. This means that with the new constraints, buses would skip slightly fewer stops, indicating that more stops are being serviced. It should be noted that all stops are passed at least once now, compared to the original 27 stops that were only passed once. Also, the number of stops that are passed less than three times has decreased from 103 to 75.Hence, modifying the constraints to ensure that all stops are passed at least once has improved bus service accessibility across all routes, particularly for stops in high minority and poverty rate areas. However, the extent of this impact varies across different routes.For instance, for routes like O5, 16, and 19L, the skip percentage remains unaffected. However, for some routes like 7, 18, 36, there is an increase in the service level as reflected in the decrease in skip percentages. Conversely, on routes like P69, 2, and 58, the skip percentage increased, suggesting that more stops are being skipped on these routes than before. This might be due to the redistribution of bus stop services to cover more stops as required by the new constraints. Overall, ensuring that all stops are passed at least once regardless of local poverty or minority rates results in a more evenly distributed bus service at the minor expense of optimality, contributing to improved social equity.

Question6: What would be the impact of removing stop constraints where the local minority rate is above 20% and the local poverty rate is above 9%?

Answer6: After lifting the constraints on stops in minority areas and poverty areas, the optimal percentage of parking stops skipped remained the same at 43.38%. The percentage of stops that can be skipped on each route also remained unchanged.In terms of the number of stops passed only once or less than three times, the results remain the same: 27 stops were only passed once, and 103 stops were passed less than three times. In summary, lifting these constraints didn't impact the model's solution. This indicates that, in the optimal solution, the stops in minority areas and poverty areas were already being served at least as much as required by the constraints.

Question7: How to ensure the fairness of each route through each stop?

Answer7: The addition of fairness constraints to ensure approximately equal visits by different bus routes to each stop has led to an infeasibility in the optimization problem. The conflicting constraints identified are 'pov_constraint_N28290' and 'fairness_N28290_O5_versus_4'. The 'pov_constraint_N28290' constraint is a pre-existing constraint that ensures a stop is passed at least 10 times if the local poverty rate is higher than 9%. While the 'fairness_N28290_O5_versus_4' constraint is a newly added one which tries to endorse fairness such that bus routes O5 and 4 pass the stop N28290 an approximately even number of times.Infeasibility conveys that there are contradictory requirements within the constraints that make it impossible to find a solution that satisfies all of them. In this particular case, it indicates that it's not possible for the bus routes O5 and 4 to visit the stop N28290 approximately an equal number of times (with an allowed variation of up to 50% between them), while also ensuring stop N28290 is visited at least 10 times, given its local poverty rate criterion. This poses a need to re-evaluate the constraints of the problem or the parameters therein, to reconcile the conflict and make the problem feasible again. As per the original optimization result, without the fairness constraints, the solution was feasible and the model skipped around 43.38% of stops to maximize revenue while satisfying the conditions. However, with the addition of new fairness constraints, there's yet no feasible solution that could lead to a comparison of the corresponding outcomes.

Question8: Why not just use one route through all the stops?

Answer8: Based on the optimized scheduler, approximately 43.38% of stops can be skipped which is potentially significant for improving the efficiency of bus operations. This means nearly half of the stops were determined as non-critical for servicing, which could reduce the bus dwelling time and overall travel time for the passengers. A summary of the percentage of stops to remain for each route is also generated. For example, the route 'O5' has the highest percentage of stops that can be skipped, which is around 88.12%, indicating high efficiency in this route design. In contrast, the route 'P13' only has around 18.57% stops which could be skipped, implying this route might have a high demand of stops and all the stops are relatively important. Moreover, there are also some stops that are passed only once (27 stops) or less than three times (103 stops) in the optimized solution. These stops might locate in remote areas or sparse population density, indicating a low demand for bus services. The detailed comparison of results for each route can provide insightful implications for bus operation management, such as route planning, scheduling, and frequency setting. Thus, it demonstrates that the Gurobi solver effectively helps to optimize the problem under the specific constraints and can make bus operation more efficient.

Question9: Q9:Can you make some suggestions for the current routes optimization results?

Answer9: Sorry. I cannot answer your question.

Analyse: OptiGuide tried to explain this sentence in code, causing OpenAI to generate the wrong code to insert into the original file, thus resulting in unanswerable analysis results.

Question10: What would be the impact if the ticket price were reduced to half of the

original price and the average cost dropped to 0.1?

Answer10: After reducing the ticket price to half of the original value and lowering the average cost to 0.1, the optimal solution suggests that 43.38% of stops will still be skipped, which is exactly the same as the original case. The distribution of percentage of skippable stops for each route also remains unchanged. Consequently, the changes in ticket price and average cost didn't affect the optimization result in terms of skipped stops and their distribution across routes. However, this doesn't necessarily mean that these changes have no impact at all. Despite the unchanged operations (i.e., which stops to skip and not), the reduced ticket price will inevitably lead to a decrease in the total revenue, while the lower average cost might help save some expenditure. Therefore, the overall profitability should be verified separately to fully understand the impact of these changes. Moreover, the results show that there are still 27 stops passed only once and 103 stops passed less than three times. This part of the result is also identical to the original situation. This continues to show the balance between efficiency (i.e., skipping as many stops as possible) and service quality (i.e., not skipping stops with particularly high demand or social requirements).

Analyse: The conclusion does not accord with the expectation. The reason may be that the code generated by OpenAI is different from the expected code.

4.2 OptiGuide Visual Optimization

By using the Streamlit component, we built an interactive visualization window that can display the response results of LLMs in stages. This method not only improves the efficiency and readability of OptiGuide's response display, but also enhances the user interaction experience, making it more intuitive and convenient to obtain information from complex language models. However, due to module optimization issues, when multiple questions are asked, the initial question results are overwritten. Therefore, we will further optimize the visualization window in the future to provide effective data saving and reading.

5 Appendix