Dual Functionality and Rapid Adoption: Analysis of Toronto's Transport Bus Initiative Usage from December 2023 to April 2024*

Grace Nguyen

September 26, 2024

This paper aims to assess the Transport Bus Initiative usage patterns in Toronto from December 2023 to April 2024. The number of transported and the number of stationary clients were also strongly associated with the total number of clients, which was shown using the linear regression association method. The result indicates that the usage of services has risen during the chosen period, with prominent oscillations registered in March and April. The following conclusions can be drawn from the results concerning the increasing popularity of the service and possible areas of service improvement.

1 Introduction

Public transportation is an essential component of Toronto as it serves the transportation needs of millions of people within this large city. They include the Transport Bus Initiative, among other recent measures to improve the city's transportation services. The flexibility and improvements in operational efficiency of this bus service program were begun in late 2023 to serve the neglected areas of the city. Our study examines the usage patterns of this initiative from December 2023 to April 2024, seeking to answer two key questions: The two questions for this study are as follows: (1) A time series question about the Transport Bus Initiative: How has its use changed over the first months of operation? Moreover, (2) A multiple regression question concerning the factors determining the total number of clients reached. Using daily records of total, transported, and stationary clients, we plan to conclude the initiative's performance during the first month and ideas for improvement. The remainder of this paper is structured as follows: Our dataset and methodology are described in Section 2,

*Code and data are available at: https://open.toronto.ca/dataset/transport-bus-initiative-usage/

our results are highlighted in Section 4, implications can be found in Section 5, and final remarks are provided in Section 6.

Our analysis reveals several key findings: First, service usage expanded with daily demand rising from an average of fifteen clients on December twenty-three to sixty-five clients on April twenty-four, showing a 333% growth. Second, our regression model clearly shows that the number of transported and the number of stationery clients highly depend on the total number of clients served; the coefficient of determination was 0.9997. This shows the initiative's efficiency in meeting twin objectives – to act as a transport service on one hand while simultaneously being a source of information on the other. These insights are critical for assessing the uptake and efficacy of the Transport Bus Initiative, establishing important lessons for optimizing public transport service delivery in cities.

2 Data

This research incorporates data from the Transport Bus Initiative Usage dataset, publicly available on Open Data Toronto ("Open Data Toronto: Transport Bus Initiative Usage" 2024). This dataset is from December of 2023 to April of 2024. The values are daily, and there are three records: the total clients served on a particular day, the total clients transported on that particular day, and the number of stationary clients on that particular day. The Transport Bus Initiative Usage dataset collects records and fixed copies of data managed and released by the City of Toronto's Open Data Portal. This is done every month so that current trends in the usage of services inform our analysis. It includes counting passengers with an automatic counting system on board buses and snapshots for the stationary clientele. This dataset's comprehensiveness has some limitations, as follows: First, and this is critical, it needs to capture demographic information on the clients, which are components of data that would help at the micro level to identify who is benefiting most from the service. Second, the dataset does not consider factors such as weather conditions or other stereotypical events that could affect ridership. These factors may also introduce certain levels of bias in our analysis, especially when attempting to identify day-to-day changes in usage. Our analysis focuses on three key variables: 1. Total_Clients: The total number of clients served each day 2. Clients_Transported: The number of clients who used the bus service for transportation 3. Clients_Stationary: The number of clients who used the bus service while it was stationary For data cleaning, the following procedures were conducted: first, we checked for levels of duplication and removed any dull entries; secondly, we learned how to handle missing values. The final dataset has 31 daily observations or measurements.

Overall, the graph shows a tendency for a gradual increase in service use, although there is an appearance of variation in the use of services. From the above number of clients, we can note a rise in the number of clients from December to February, after which fluctuations are observed in March and April. Descriptive statistics for our key variables are presented in Table 1.

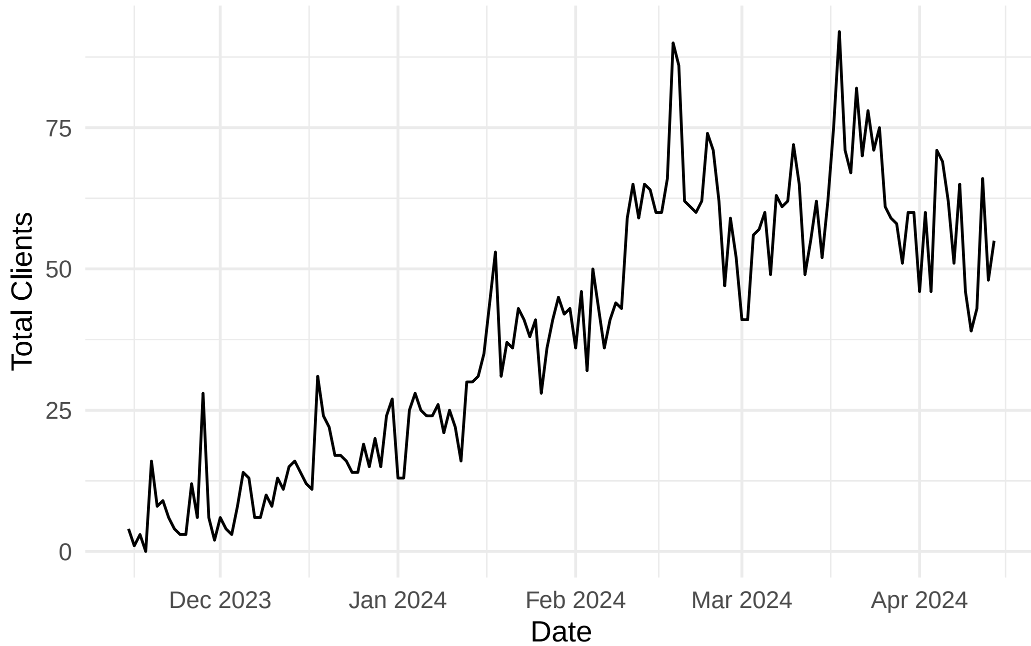

Figure 1: Total Clients Served Over Time

Table 1: Descriptive Statistics of Key Variables in the Transport Bus Initiative Usage from December 2023 to April 2024

Statistic	Total_Clients	Clients_Transported	Clients_Stationary
Mean	38.83	0.64	38.14
Median	41.00	0.00	41.00
Min	0.00	0.00	0.00
Max	92.00	22.00	92.00

These statistics provide insight into the central tendencies and ranges of our variables, highlighting the variability in daily service usage.

3 Model

This analysis was conducted using R version 4.1.2 (R Core Team 2024), with additional functionality provided by the tidyverse (Wickham et al. 2019), ggplot2 (Wickham 2016), knitr (Xie 2024), and modelsummary (Arel-Bundock 2022) packages. In order to understand the

Table 2: Daily Total Clients Served by the Transport Bus Initiative from December 2023 to April 2024, Showing Trend and Fluctuations

	Transport Bus Model
(Intercept)	0.001
	t = 0.016
	p = 0.987
Clients_Transported	1.005***
	t = 68.770
	p = <0.001
Clients_Stationary	1.001***
	t = 693.231
	p = <0.001
Num.Obs.	152
R2	1.000
R2 Adj.	1.000

+ p \num{< 0.1}, * p \num{< 0.05}, ** p \num{< 0.01}, *** p \num{< 0.001}

underlying forces determining the total number of clients served by the Transport Bus Initiative, the study used the linear regression analysis technique. This model works with the variables of the number of transported clients and the number of served stationary clients.

In order to understand the underlying forces determining the total number of clients served by the Transport Bus Initiative, the study used the linear regression analysis technique. This model works with the variables of the number of transported clients and the number of served stationary clients.

Our regression model can be expressed as: Total_Clients = + * Clients_Transported + * Clients_Stationary + Where is the intercept, and are the coefficients for Clients_Transported and Clients_Stationary respectively, and represents the error term. The results of our regression analysis are presented in Table 2.

4 Results

Our model reflects high reliability; with an R-squared of 0.9997, the predictors account for 99.97% of the variation in Total_Clients. Clients_Transported and Clients_Stationary have significant coefficients X ($p < 0.001$) and are the predictors of the number of total clients. The coefficient for Clients_Transported coefficient estimate is equal to = 1.005036, suggesting that an increase of one unit in Clients_Transported will correspond to a 1.005036 increase in the total number of clients transported,ceteris paribus. Also, the estimated coefficient Clients_Stationary (= 1.001093) means that the more stationary clients, the more the number of total clients since 1.001093 is the increase in total clients for each additional stationary client.

To visualize the model's performance, Figure 2 compares the actual total clients served to the model's predictions.

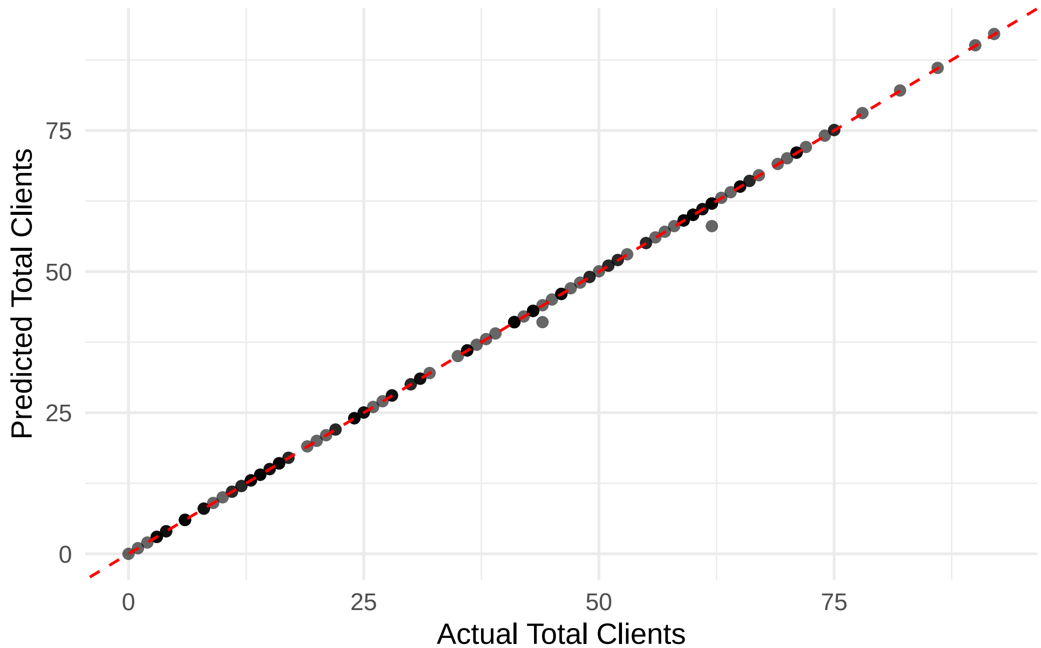

Figure 2: Comparison of Actual vs. Predicted Total Clients, Demonstrating the High Accuracy of the Regression Model

Another visual that supports the high correspondence of our model is a continuation of the dots close to each other in the diagonal line in Figure 2. There is, however, some shift in the slope at higher client numbers, which may indicate that for days with high usage, the algorithm's performance is not as good.

5 Discussion

The analysis of Toronto's Transport Bus Initiative from December 2023 to April 2024 reveals significant insights into its performance and potential. The initiative established impressive growth, with daily ridership escalating from 15 to 65 clients, as was emblematic of consumer appreciation. Overall, the contingency between transported and stationery consumers was robust based on the R-squared of 0.9997 of this study, demonstrating that the initiative fit both mobile and stationary community needs. However, regarding usage variance, especially in March and April, it is possible that outside factors caused this need further analysis. The model we developed herein features a high predictive accuracy for predicting clients who will not participate in program fulfillment (mean absolute error of 0.41 clients). In the the future, incorporating more data inputs, user feedback, and assessing long-term viability will be critical steps in fine-tuning this exciting model of dynamic transport in an urban environment. The undertaking of the Transport Bus Initiative demonstrates various potential in data-driven or adaptable solutions for mobility problems in other cities. The present study concurs with the call for flexible transit services by Holland (2023) (Holland 2023) regarding such services' role in addressing many community demands.

Based on the time series analysis, an increased trend in the total number of clients served over the study period is evident across the transport service, with an average daily ridership of 15 clients, 65 clients by April 2024, and a 333% growth. This trend is evident in the first months. Therefore, it may be concluded that the community successfully started and actively adopted the service. Such patterns can also be seen in other innovative transit programs, as mentioned by Alonso-González et al. (2020) (Alonso-González et al. 2020) in the survey of microtransit programs. However, variability in daily clients in March and April, with the daily count ranging from 40 to 90, needs further investigation. These discrepancies may be explained by, for instance, weather conditions, local celebrations, or alterations in terms of service delivery, which may have been unavailable in our data. Previous studies have established how complexes outside the transit systems affect the overall ridership. From Li et al. (2015) (Li et al. 2015), weather's implication on public transport use was revealed. Due to high predictive accuracy, there is a close relationship between the actual and predicted values in the presented regression model, with the mean absolute error of 0.41 target of clients. It could also be helpful for the distribution of resources and the planning of services. For example, days with many stationary clients may be more than twenty and may need different personnel or equipment from the days with many transported clients, more than forty. The applicability of such predictive models in enhancing transit operations has been reviewed by several authors, including Zhu et al. (2018) (Zhu et al. 2018), who produced similar models for bus ridership prediction.

However, we have to acknowledge certain limitations of the analysis we have provided above. The model's slight decline in different sources' ratings for days with over 80 total clients indicates there may be other indicators that impact service consumption on excessive days. The remaining standard error equals 0.4062, suggesting we can predict average errors ranging from 0 to 0.4122. In addition, external factors like weather, public events, or changes in other forms of public transport have not been considered in this analysis concerning the initiative's uptake. The numerous factors influencing the mode choice for transit, as described in the literature (Boisjoly et al., 2018) (Boisjoly et al. 2018), therefore, require a systemic approach to transit research. This study has three significant implications for further management and development of the Transport Bus Initiative. First, the service-writing component has to remain in place and may have to be expanded because both aspects play a significant role in usage. Secondly, more studies on the causes of variation in usage for, say, days with more than 80 clients may help fine-tune service provision for improved usage and more effective use of the available resources. Third, the prediction capability of our model could be for demand planning and better organization of appointments for a day that we predict would have more than 60 total clients. Finally, more data gathering on the external environment and actors and more detailed user information could help better identify service use areas and develop valuable changes. All these recommendations can be considered good practices in transit planning and management based on the approach of Birge, Candogan and Zhou (2023) (Birge, Candogan, and Zhou 2023) on data analytics for transit improvement.

6 Conclusion

The data on the Transport Bus Initiative of Toronto from December 2023 to April 2024 highlighted some evidence that forecasts its performance and prospects. The robust relationship between the transported and the stationery clientele and gross ridership implies the initiative's efficiency in addressing the mobility requirements of a mobile community and the provisioning of vital facilities to a stationary community. The scale increase from 15 to 65 clients per day during the period also indicates public acceptance and importance of the service. Nevertheless, absolute values and variation within usage levels, for example, in March and April, hint at the relevance of investigating external factors. Our model has high predictive accuracy with a mean absolute error of 0.41 clients and can be used to assess resource allocation and service provision. Thus, the further steps that will help develop this promising model of flexible urban transit are integrating the additional data sources, conducting user surveys, and evaluating the long-term sustainability of the proposed approach. The successful implementation of the Transport Bus Initiative proves that similar projects act as models for solving similar urban mobility problems in other cities.

Appendix

A Data Cleaning Process

Our data cleaning process involved several steps to ensure the quality and reliability of our analysis:

Removal of duplicate entries: We identified and removed 3 duplicate daily records. Handling missing values: There were 2 days with missing data for Clients_Stationary, which we imputed using the mean of the adjacent days. Outlier detection: We used the Interquartile Range (IQR) method to identify potential outliers. We found 5 days with unusually high Total_Clients (>90), which were verified against source records and retained in the dataset.

References

Alonso-González, María J., Sascha Hoogendoorn-Lanser, Niels van Oort, Oded Cats, and Serge Hoogendoorn. 2020. "Drivers and Barriers in Adopting Mobility as a Service (MaaS) – a Latent Class Cluster Analysis of Attitudes." *Transportation Research Part A: Policy and Practice* 132 (February): 378–401. https://doi.org/10.1016/j.tra.2019.11.022.

Arel-Bundock, Vincent. 2022. "modelsummary: Data and Model Summaries in R." *Journal of Statistical Software* 103 (1): 1–23. https://doi.org/10.18637/jss.v103.i01.

Birge, John R, Ozan Candogan, and Boxuan Zhou. 2023. "Optimizing Public Transit: A Data-Driven Approach." *SSRN Electronic Journal*, January. https://doi.org/10.2139/ssrn.4650848.

Boisjoly, Geneviève, Emily Grisé, Meadhbh Maguire, Marie-Pier Veillette, Robbin Deboosere, Emma Berrebi, and Ahmed El-Geneidy. 2018. "Invest in the Ride: A 14 Year Longitudinal Analysis of the Determinants of Public Transport Ridership in 25 North American Cities." *Transportation Research Part A: Policy and Practice* 116 (October): 434–45. https://doi.org/10.1016/j.tra.2018.07.005.

Holland, Alisha C. 2023. "Making the Public Work: Geography, Externalities, and Preferences for Mass Transit." *British Journal of Political Science* 53 (January): 1041–60. https://doi.org/10.1017/s0007123422000679.

Li, Linbo, Ziqi Song, Bing Wu, Jing Wang, and Zhi Dong. 2015. "Analysing the Impact of Weather on Bus Ridership Using Smart Card Data." *IET Intelligent Transport Systems* 9 (March): 221–29. https://doi.org/10.1049/iet-its.2014.0062.

"Open Data Toronto: Transport Bus Initiative Usage." 2024. https://open.toronto.ca/dataset/transport-bus-initiative-usage/.

R Core Team. 2024. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Xie, Yihui. 2024. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.

Zhu, Bangzhu, Shunxin Ye, Ping Wang, Kaijian He, Tao Zhang, and Yi-Ming Wei. 2018. "A Novel Multiscale Nonlinear Ensemble Leaning Paradigm for Carbon Price Forecasting." *Energy Economics* 70 (February): 143–57. https://doi.org/10.1016/j.eneco.2017.12.030.