Australasian Transport Research Forum 2023 Proceedings
29 November – 1 December, Perth, Australia
Publication website: http://www.atrf.info

Impact of off-peak fares on train patronage: a case study from DoT Victoria, pre and post COVID-19

Yohanes Puraditya [1], Adriana-Simona Mihaita[2], Yuming Ou[2], Tuo Mao[3]
[1] Under-graduate student, University of Technology, Sydney
[2,3] Senior Lecturer, University of Technology, Sydney
[3] Research Fellow, University of Technology, Sydney

Email for correspondence (presenting author): adriana-simona.mihaita@uts.edu.au

Abstract

The immense growth of public transportation users (patronage or ridership) makes public transportation experience excessive demands, especially during peak hours. All governments worldwide experience extreme demand problems in their public transport systems. Therefore, policies on managing demands on peak hours are made, such as pricing strategy (on/off-peak fares). With a pricing strategy, public transportation could utilize its capacity in a balanced manner. Since Australia's population is growing to 27.55 million in 2027, transportation users are also increasing with the growing population. This paper presents a detailed analysis of the impact of fare or policy pre/during the Covid-19 pandemic as a case study from DoT Victoria, Australia. In the first part of the paper, we start by analyzing the train network, choosing the most popular train line networks, and undergoing an analysis of the patronage impact of the off-peak fare during regular traffic periods throughout the COVID-19 pandemic. In the second part of the paper, we deploy a machine learning approach for predicting patronage numbers with and without major COVID-disruption and discuss the findings and future recommendations.

1. Introduction and related works

1.1 Large travel disruptions impact

Nowadays, people can commute between locations using a variety of modes of transportation, including cars, motorcycles, and public transportation. Most of the world's transportation systems are experiencing similar issues as time goes on. These include excessive demands (commuter volumes) on public transportation and traffic jams during peak highway hours (Lovrić et al. 2016). Many governments worldwide, particularly in Australia, have sought a solution to these issues. Most governments globally select on/off-peak policy or price tactics since they are more effective because investing in infrastructure will require substantial expenditures (Yen et al. 2015).

On-peak and off-peak measures have been used by numerous nations throughout the world, including Singapore, Indonesia, and Australia, to address these issues. To improve the efficiency of the public transportation system, on- and off-peak policies are used. It is envisaged that the on-peak and off-peak pricing techniques will redistribute excessive passenger loads in public transportation and ease congestion during on-peak times. As a result, public transportation could utilize its capacity in a balanced manner (Lovrić et al. 2016).

Many researchers worldwide have researched the impact of pricing strategies on various cities, countries, and transportation modes. Some of them use a fare-free program at certain times or periods (Štraub, 2020; Bull et al., 2021), discounted prices during particular hours or off-peak hours (Adnan et al., 2020; Huan et al., 2021), and free travel rewards for off-peak or certain hours after several trips (Yang & Tang, 2018; Yen et al., 2015). Ge et al. (2015) and Lin et al. (2019) wrote that managing travel behaviour from the passengers' side, including applying pricing strategies, is more promising for solving the excessive on-peak passengers rather than building more infrastructure and capacity. Huan et al. (2021) stated that time-dependent pricing policies could reduce approximately 13.97% of on-peak passengers in the Beijing Metro. Like Huan et al. (2021), Bull et al. (2021) also explained that a fare-free strategy in Santiago could reduce the total amount of trips by around 23%. In the Czech Republic, the fare-free public transport strategy could increase public transportation usage (Štraub, 2020). Halvorsen et al. (2016) conducted their research in Hong Kong and found that the off-peak price policy could decrease around 3% of peak hour trips. Moreover, they wrote that price change could impact ridership or patronage during on and off-peak hours by calculating the price elasticity.

While off-peak versus on-peak travel measures have been efficient in a pre-COVID-19 period, many countries around the world have reduced their public transport services after March 2020, especially during peak hours, due to a lockdown policy by their government to prevent the spreading virus (Beck et al., 2021). Similarly, Australia imposed several restrictions, and Victoria was one of the states with the most extended lockdown periods that have generated several movement disruptions across the state. Overall, studies report an up to 80% decline in public transport patronage compared to the pre-pandemic era (see Munawar et al. 2021 and Ou et al. 2021).

Therefore, in the first instance of our work, we aim to analyze the impact of the lockdown on off-peak travel behavior across the train network in Victoria and draw insights into the factors that could have led to a significant travel behavioral change. We make the observation that by large disruption we mainly refer to the “COVID-19 pandemic” and not other technical disruptions that might have occurred.

1.2 Patronage prediction under large disruptions

Secondly, we aim to study the potential of training several machine learning models on the available datasets regarding disrupted travel behavior and learn whether such models could cope with large-scale events in the future.

Kusonkhum et al. (2022) researched predicting the Thailand Underground Train's passengers with a machine-learning approach. Thailand Underground Train experienced a crowded ridership, creating a problem and reducing customer satisfaction. Therefore, the study aimed to develop a machine-learning model for predicting passenger demand over time. Standard data collection tools were also utilized to gather information from the Purple Line of the Metropolitan Rapid Transit (MRT). There were 16 stations along this line. The nine considered variables are the station's name, day, month, period, number of commuters, holidays, weekends, and weather. The analysis phase, classification, and regression algorithm were the analysis methods. However, the regression approach could be applied due to its poor accuracy. This investigation also used three categorization techniques: decision tree, random forest, and artificial neural networks. The results also presented that the artificial neural network has high prediction accuracy. The accuracy value is indicated to be greater than 0.85.

Park et al. (2022) studied and predicted subway passenger flow with machine learning. They used the daily traffic at each Seoul Metro station using a sizable dataset of smart card transactions. According to their passenger transportation patterns, they first grouped the stations into six categories. After that, they subsequently predicted the daily passenger volume for each cluster. By comparing their predicted results with the actual or absolute number of passengers, they demonstrated that the expected number of passengers based on the clustering results was more accurate than the result without considering the regional features. As a result of their data-driven strategy, which can alleviate congestion by adjusting train intervals based on passenger flow, the subway service plan could be improved. The predicted outcome could also be used to design a "smart city," which aims to achieve reduced travel times, comfortable riding, and environmental sustainability.

Zou et al. (2022) provided a method for deriving the station's passenger flow from various routes and used the XGBoost model to determine the contributions of multiple variables to the forecast of the station's passenger flow. They obtained the data from a connected bus system using smart card information. Because the competition and complementarity of other routes and buses on the same road can significantly impact the passenger flow of a station, they added the number of routes buses during the anticipated interval into the model to boost accuracy. In addition, XGBoost could achieve higher accuracy using fewer computational resources than the LSTM deep learning model. As can be observed, the most crucial factor in predicting passenger flow is time. Hence bus scheduling during peak hours needs to be improved dramatically, with or without significant disruptions affecting it.

2. Research objectives

As described in the previous section, the main research objectives of this work are to:

a) Analyze the impact of off-peak fares on people's movement inside the Victorian train network before and after the COVID-19 pandemic.
b) Understand what factors influence people's choices and behavior and whether the off-peak fares have helped shape the traffic demand across the network and influence people's behavior.
c) Build a machine learning framework for training patronage prediction under large disruptions and evaluate its efficiency.

3. Case study

3.1 Location

The location of our study is the V/line train in Victoria, as depicted in Fig. 1, which has 276 stations, 17 train lines, spreading across 79 LGAs (Local Governmental Areas), with a length of 1.712km and servicing a population of 6,68 Million people. The peak hours in Victoria are considered between 7–9 am for the morning AM peak and 4–6 pm for the afternoon peak hours. The off-peak discount fare currently in place holds under several conditions: a) leave or arrive in Melbourne between 9 am and 4 pm on weekdays, b) leave or arrive in Melbourne after 6 pm on weekdays (three zones or more), c) weekends and public holidays.

3.2 Data availability

The data is obtained from the Department of Transport (DOT) Victoria and covers the following periods:

Before the Covid-19 Pandemic (2019):

- 4 – 10 February (First week of February 2019)
- 25 – 31 March (Last week of March 2019)
- 19 – 25 August (Third week of August 2019)
- 18 – 24 November (Third week of November 2019)

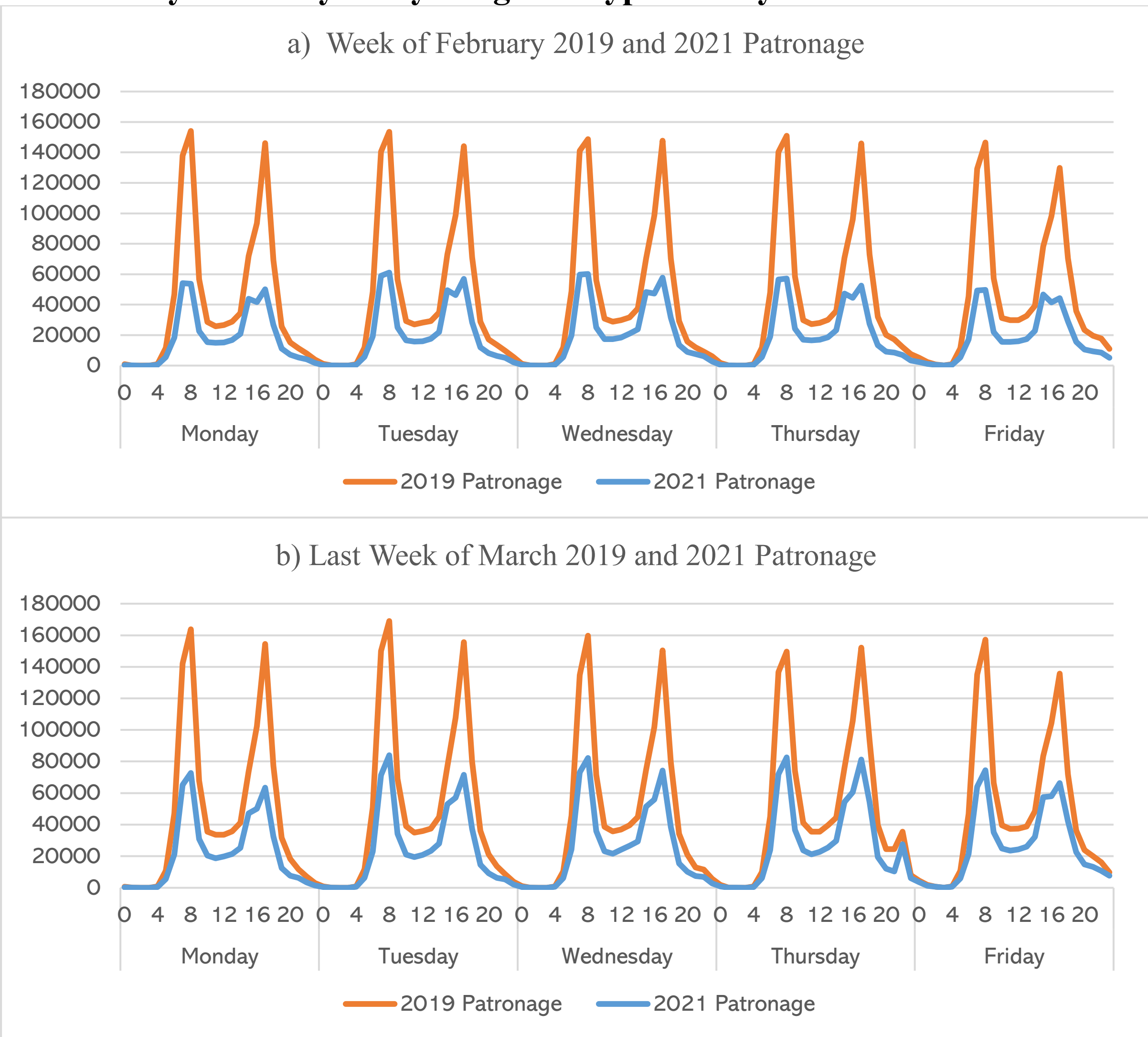
During Covid-19 Pandemic (2021):

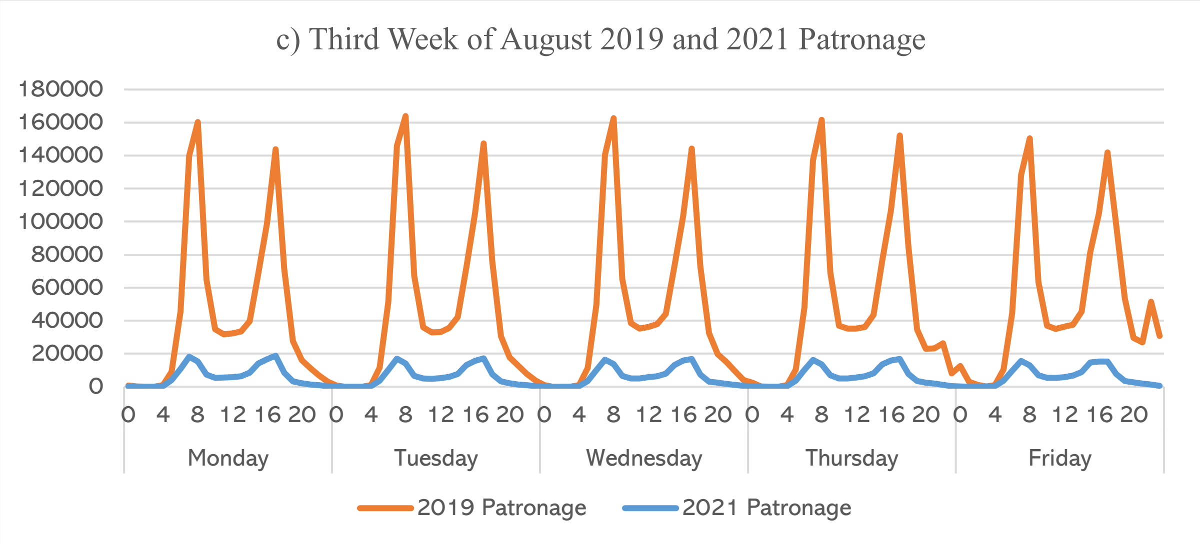
- 1 – 7 February (First week of February 2021)
- 22- 28 March (Last week of March 2021)
- 16 – 22 August (Third week of August 2021 – Lockdown)
- 15 – 21 November (Third week of November 2021 – ease of lockdown)

Figure 1: Victorian Train Network Map; Source: (Maps - Public Transport Victoria, n.d.).

4. Data mining on patronage impact – off-peak versus on-peak

4.1 Monthly and daily analysis against types of days





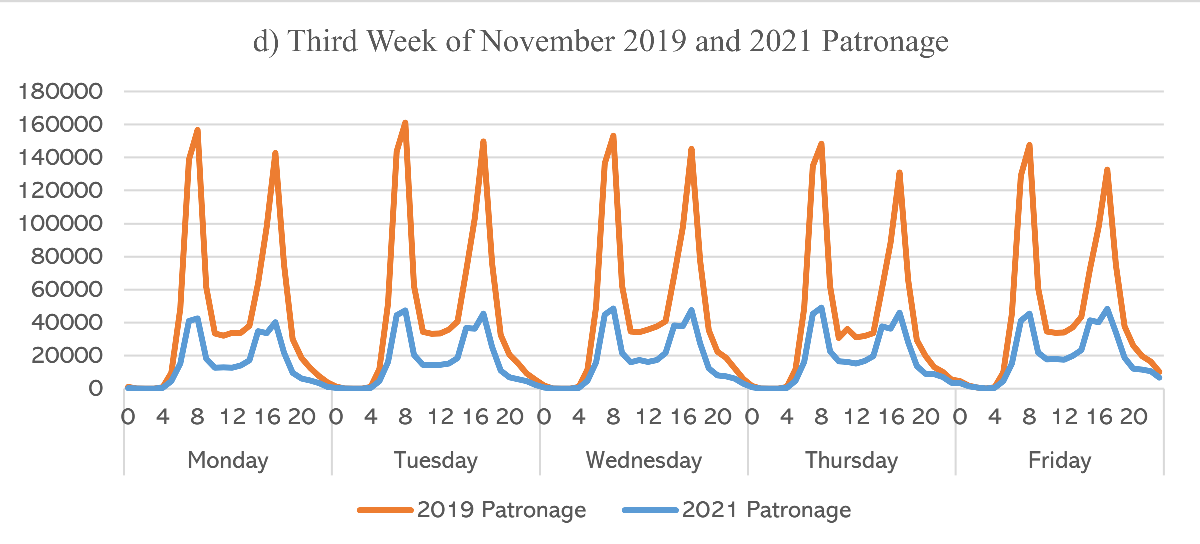
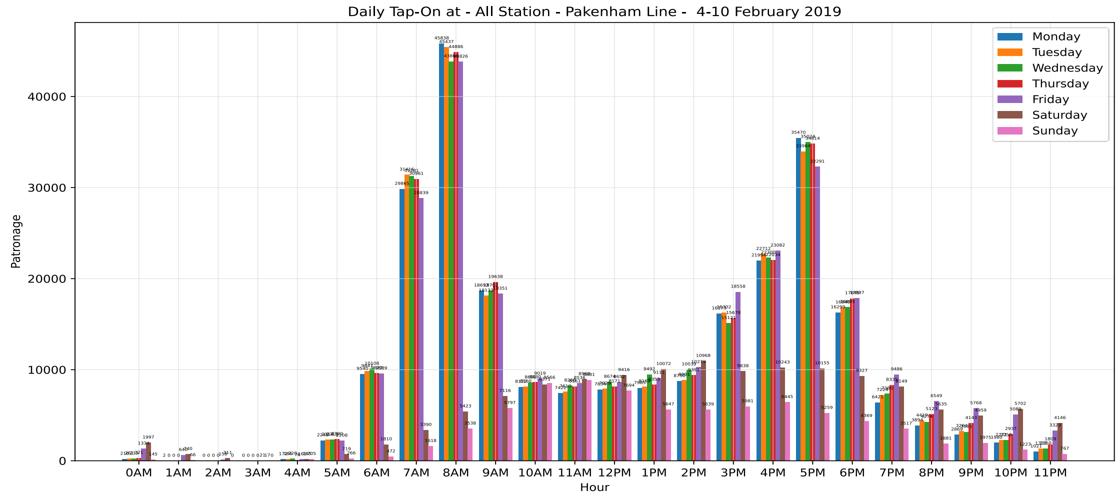


Figure 2: Monthly analysis over weekly days – 2019 versus 2021 patronage impact analysis.

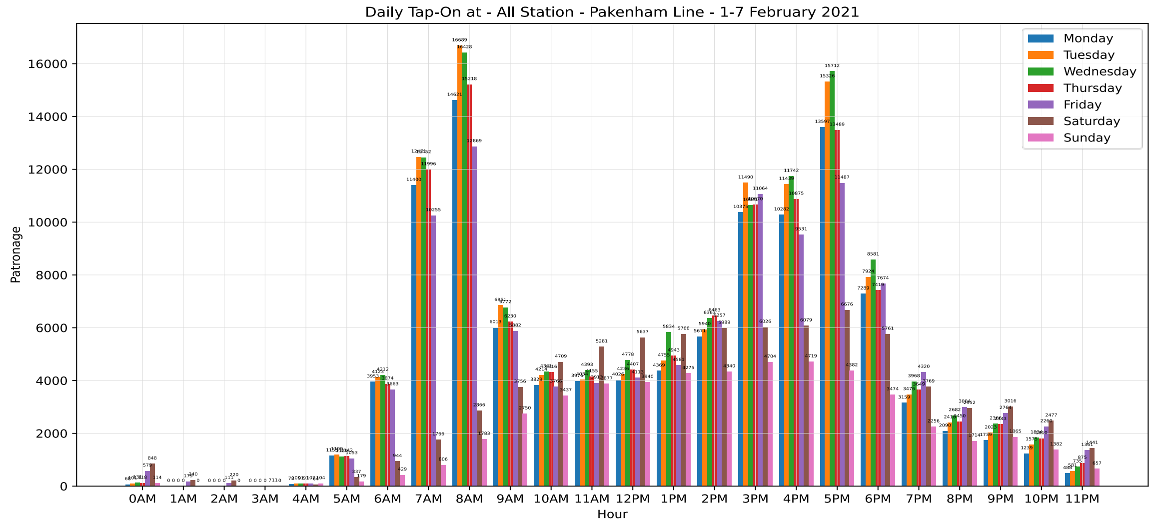
One of the first findings is related to the monthly comparison against the type of day of the week as seen in Fig 4a) – February 2019 versus February 2021, Fig. 4b) March 2019 versus March 2021, Fig. 4c) August 2019 versus August 2021 and Fig. 4d) November 2019 versus November 2021. While a significant patronage reduction reached almost 57%-60% in February and March 2021 against 2019, the most critical patronage drop was observed in August 2021 versus August 2019, when an 87.5% drop was recorded. This explanation is mainly related to the strict government lockdown in Victoria in August 2021, which led to a significant slowdown in people's movement across the public transport network.

An exciting travel pattern has been observed regarding off-peak versus on-peak travel trips. Despite being offered the possibility to travel during off-peak hours to maintain physical distancing and receive a 30% discount on fare, people still travelled more during the 7–9 am and 4-6 pm peak hours, even during the lockdown periods. This indicates that even such incentives can hardly change travel behaviour across the train network.

4.2 Daily peak versus off-peak patronage impact



a)



b)

Figure 3: Pakenham Line Daily Tap-On counts from Monday – Sunday, compared between February 2019 (a) and February 2021 (b).

In order to evaluate the off-peak versus on-peak travel demand, we have conducted an in-depth analysis of all train lines and selected as an example for discussion one of the most famous lines, which is the Pakenham Line (see Figure 3). When comparing the 24h patronage profiles across the Pakenham line between 2019 and 2021, we have observed that the travel patterns are still maintained throughout the day, but on a lower scale (45,838 passengers travelling at 8 am during a pre-pandemic regular Monday as compared to 16,689 trips during the COVID -19 pandemic, translating in a reduction of 63.5% decrease in travel demand during the most popular morning rush hour). The off-peak travel demand at 6 am in 2021 seems to represent 75% of the total on-peak demand from 8:00 am, whereas, in 2019, it was estimated at 77%. So, this is the opposite of the expected increased travel demand during the pandemic and the incentives provided and shows that the measures taken were not enough to convince travellers to an early start of their day, despite being during the lockdown period. The only period when the off-peak fares were successful was between 10 am and 1 pm (see Figure 3b) on Saturdays, but given the weekend travel behaviour, this does not pose a problem to the current train network system during the weekend but rather during the weekdays. The most affected stations by the patronage drop were Flagstaff, Westfall, Parliament, Richmond, and Malvern, most of which are central stations in the CBD. Similar trends have been recorded against all lines during the 24h time frame.

4.3 Yearly analysis of patronage decrease

While the off-peak fare has proved inefficient in decreasing patronage during the morning or afternoon rush hours, the overall travel restrictions have led to a decrease among all train lines, all months, and all types of peak hours (AM or PM). Figure 4 provides one overall visualisation of the decline recorded across all lines across several months from 2019 and 2021. The most significant decrease has been recorded for August 2021, where patronage suffered an 87.42% reduction due to severe lockdowns in Victoria. The rest of the months have seen a patronage reduction between 43% and 58%, making the service operating at half of its regular capacity to comply with travel distance between passengers and travel distance restrictions to local areas.

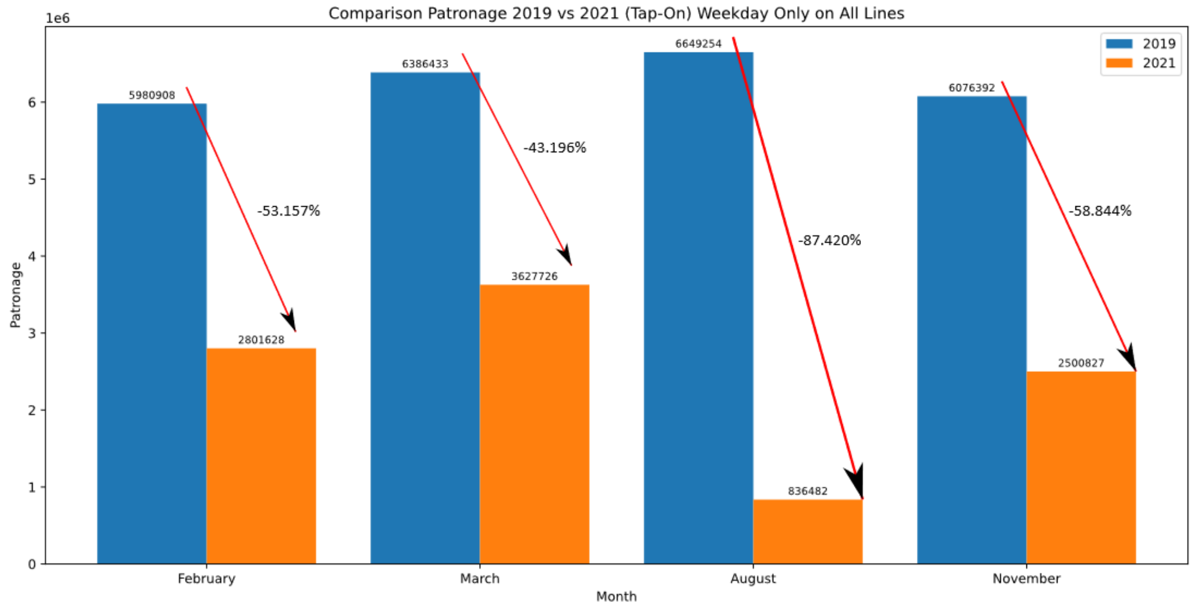


Figure 4: Patronage comparison on all lines pre and post-COVID.

5. Patronage prediction modelling

5.1 Modelling description

Another important research question for this project was to investigate the feasibility of training machine learning models in order to predict train patronage under such large-scale disruptions. For this work, we have trained several ML models (such as Linear Regression Models, Support Vector Machines, SVM Kernels, Gradient Boosted Decision Trees, Neural Networks, and Regression Trees) and evaluated them against several performance metrics, such as Root Mean Squared Error (RMSE) (Botchkarev 2019) and the Coefficient of Determination (R^2) (Plevris et al. 2022).

The workflow that we have followed is presented in Figure 5 and showcases the process from the patronage data filtering, building the matrix of features (see Table 1), data splitting for training, validation and testing, as well as the several ML models put to the test for this prediction modelling.

Features/ Explanation	Value dataset
Day of week	{1, 2, 3, 4, 5, 6, 7}
Hour of day	{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23}
Tap-off	Ranging from {0 – 169003}

Table 1: Feature matrix example.

We have validated the models via three different techniques such as a) the holdout validation (the composition of data separation is 60% training – 20% validation – 20% test), b) the 5-fold cross-validation (the percentage of data aside for testing is 20%. The rest of the data is used for training and cross-validation) and c) the 10=fold cross – validation (10% of the data set is kept for testing and the rest for training and validation purposes).

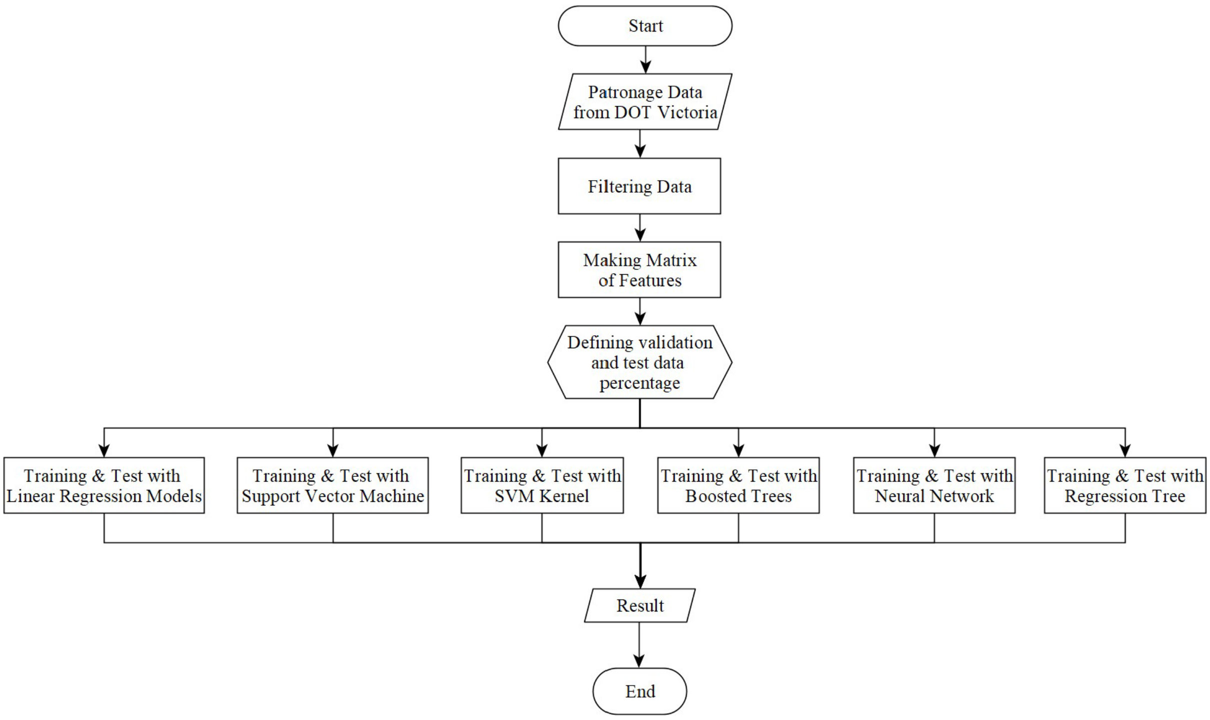


Figure 5: Flowchart of prediction modelling against several ML models.

Several model hyperparametric settings have been adopted for each model in order to obtain the best performance metrics evaluation. We make the observation that this research only predicts using the day, hour, and tap-off data. This analysis does not consider other factors, such as weather, temperature, or event that may force people to take a train. This represents a limitation and an opportunity for future work around this topic.

5.2 Comparison by Validation Method

Based on several comparisons and validation methods, the best-performing models were boosted trees, medium neural networks, and wide neural networks. In order to keep the results concise, we further focus the analysis on these models against all three validation methods (see Figure 6 and Figure 7).

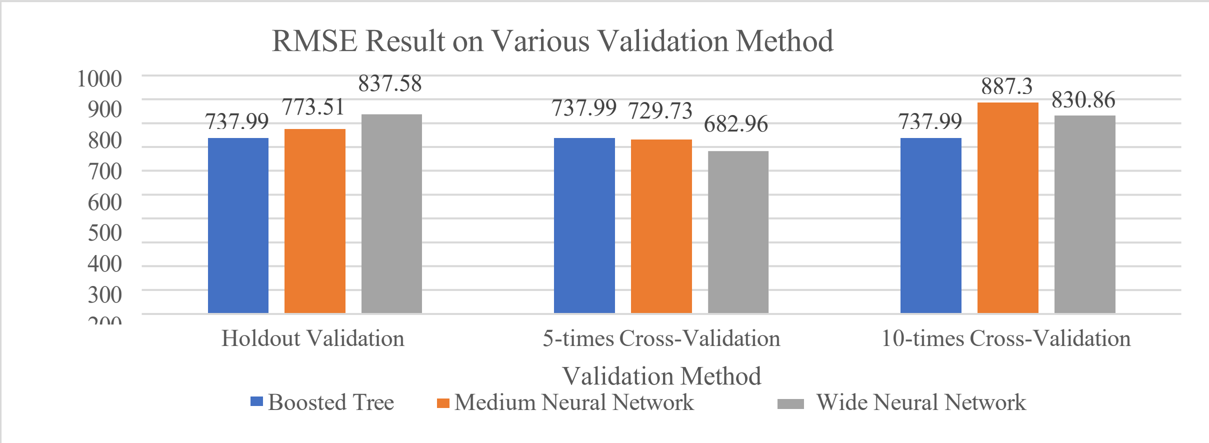


Figure 6: RMSE Result on Various Validation Methods.

Based on the figures above, the lowest RMSE results are obtained from the 5-times cross-validation method. Moreover, the highest R^2 values from the three figures above are obtained from the holdout validation and 5-times Cross-Validation. Therefore, the validation method that can get the best result is the 5-times Cross-Validation.

A similar condition was also experienced by Eertink et al. (2022), which stated that small datasets could suffer from significant uncertainty. Therefore, Cross-Validation is preferred for small datasets. Since this train patronage data is relatively small, Cross-Validation is more suitable than the holdout validation for prediction.

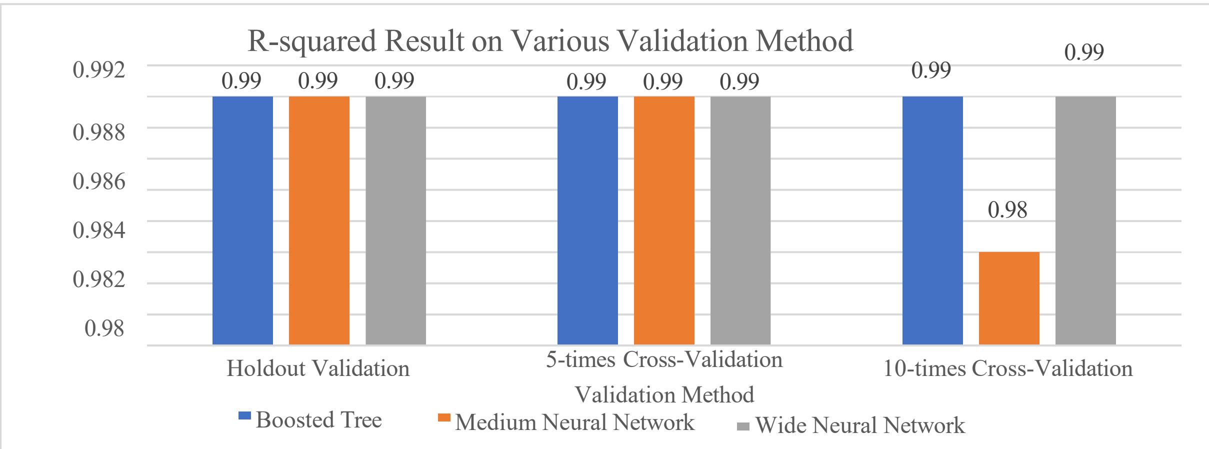


Figure 7: R-squared Result on Various Validation Methods

The accuracy of machine learning algorithms is not directly correlated with the value of *k* in the *k*-fold Cross-Validation. As a result, when determining the value of *k*, one must be careful

because a lower k value means less computational complexity and less variance but more bias. On the other hand, a higher value of *k* needs a higher computational complexity but has more variance and lower bias. Therefore, the choice of k must let the size of each validation set be adequate to offer a realistic approximation of the model's performance (Nti, Nyarko-Boateng & Aning 2021). The main role of Figure 7 was to prove which validation method is the best to apply and the very close results indicate that any of the three validation methods is working well with the data set.
In some other data sets, we need to test very carefully which validation method is more suitable and this is the main reason for us conducting this sensitivity analysis to have reassurance.
Therefore, both the 5-times Cross-Validation and 10-times Cross-Validation act similarly on the model's performance. However, this research with the received data set is more suitable for using a 5-times Cross-Validation rather than a 10-times Cross-Validation.

5.3 Comparison across all models

We have therefore applied a 5-fold cross validation and compared the overall performance of all models tested, as represented in Figure 8 and Figure 9 below. According to both the RMSE and the R^2 values, the best performing model was the Boosted Trees which seem to outperform the other models (most likely due to their handling of categorical features), while the worst performing model was the SVM kernel model.

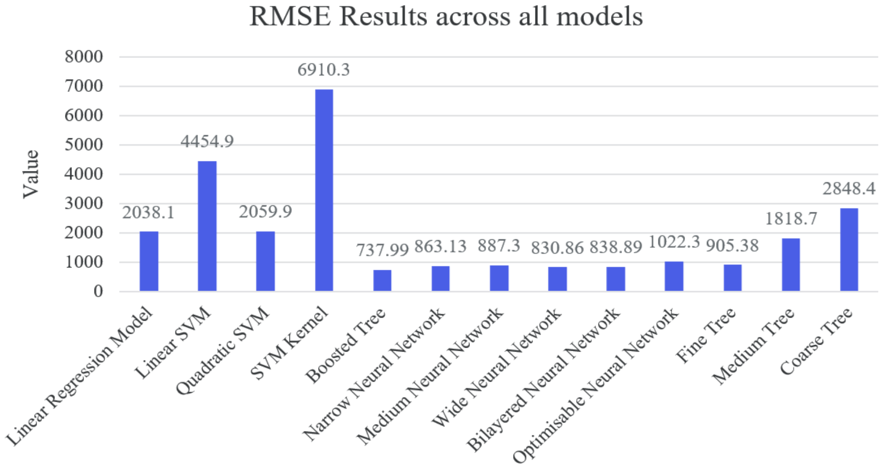


Figure 8: RMSE results across all ML models.

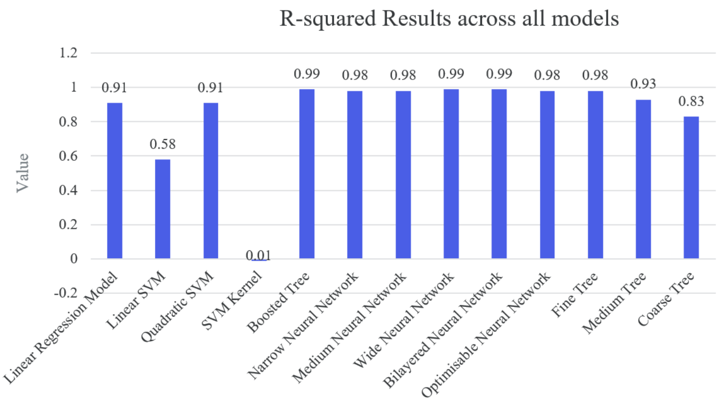


Figure 9: R-squared results across all ML models.

Observations: due to the small data set, the model's performance was not competitive in terms of other predicting tasks that one might envisage for this type of problem. This was mainly due to data restrictions from DOT Victoria's IT department. Ideally, we wanted to have access to at least 6 continuous months prior and 6 months during the pandemic. In such type of analysis however multiple years are needed – 2 years would have been ideal (2019 and 2018 for the travel patterns pre-pandemic and 2020-2021 for the post-pandemic recovery period). Also, given the large discrepancies between the trends from 2019 and 2021, we believe that the ML models have a hard time learning and respecting the patterns when it comes to predicting large-scale disruption situations such as pandemics. More data would need to be collected over several years and under different travelling circumstances to improve such prediction's accuracy under extreme conditions. This represents a future objective and a challenge related to large outlier prediction modelling.

6. Conclusions and findings

This work represents an initial data exploration undertaken as part of a student internship for DoT Victoria under iMOVE CRC funding. The project revealed the inefficiency of the off-peak travel incentives provided during the pandemic situation in 2021 and the lessons that could be learned for future network optimisation approaches to shift patronage. We make the observation that several factors can significantly affect the patronage demand, and some of these have been documented in several studies such as:

- Hygiene & crowd concern (Beck et al., 2021) – this is a study undertaken in 2021 by researchers from ITLS Australia surveying people from all around Australia, which showcased a strong correlation between hygiene concerns and patronage – due to the concrete facts around hygiene and the lack of time availability, we have not replicated the study for the DoT Victoria.
- High-risk areas for Covid-19 exposure (Munawar et al., 2021)
- Strict travelling restrictions by the government (Munawar et al., 2021),
- People worked or studied at home, where possible (Beck et al., 2021),
- Safe distancing restrictions (Beck et al., 2021),
- Change in days worked from home (Munawar et al., 2021).

Moreover, Yen et al. (2015) state that an ineffective off-peak policy indicates that the discount is not strong enough or the passengers do not have the flexibility to change their travel time. Lovrić et al. (2016), Adnan et al. (2020), and Halvorsen et al. (2016) stated previously that off-peak discount strategies do not become adequate for students, workers, and other daily and local commuters who have tight working schedules.

Significant findings regarding the impact of off-peak fares on patronage under large disruptions:

- Before and after the pandemic, people still travelled at peak hours, probably due to having a tight schedule.
- 30% off-peak fares are ineffective in flattening the patronage but could reduce the peak patronage – this is an important aspect that could have been further tested and explored by the transport representatives during the pandemic.
- The decrease in patronage in 2021 was caused by health concerns (crowd & hygiene) and external factors (capacity limitation, work-from-home policy)

- Since future patronage will increase, the government can apply a combination of on and off-peak pricing strategies.
- Flattening the curve can be difficult since it involves passengers who have a tight schedule.
- However, reducing patronage in peak hours to achieve optimum comfortability and utilization could be achieved by combining the on & off-peak pricing strategies.

Based on the data set received and the results obtained, we believe that the main factors that influence public transport patronage were:

- Health concerns – see survey results from (Beck et al., 2021), which highly emphasised the worry that travellers have concerning public transport and work hygiene.
- The nature of work and the primary commuting mode to get to work – where respondents rely only on public transport to get to work, then their travel behaviour is less likely to change as compared to travellers with more flexible travel patterns or with all facilities to work from home.
- The departure time is an essential feature that influences people's departing and arrival times to their workplaces during the day.
- The price reduction – should be more efficient if more incentives are provided, as a 30% price reduction is not significant enough to change entire travel patterns. However, more price policies could be tested (like 50% reductions outside peak hours) or even free travel in the first off-peak hours. We make the observation that the price reduction has not been considered so far in the previous surveying works and that it should be further integrated into future off-peak travel incentive policies that the local authorities should explore further.

We also believe that more factors can be explored in the future for testing the efficiency of such peak versus on-peak demand reduction strategies, such as price reduction incentives via multiple transportation modes, alternative transportation modes at more attractive pricing schemes (e-bikes, e-scooters, on-demand shared buses), free ride days during the week, and so on. These strategies must be tested on a real-life public transport system under several scenarios, with a supporting before and after data analysis to test their efficiency.

Major findings regarding the prediction modelling:

- The prediction modelling is challenging under small data sets and large-scale disruptions without precedent.
- More data granularity is recommended in order to train accurate models for future large-scale disruptions. Several ML models can be employed and fine-tuned based on the data availability.
- Several other external factors need to be included for prediction improvement, such as weather, population, events, etc.
- Regarding the cut-off values for the RMSE errors, we make the observation that this is highly dependent on the data sets that are being used to make the prediction. For specific prediction problems with large and highly reliable data sets the RMSE values can be below 100, while for other data sets which are more sparse, not continuous, and also not covering a large period of time, the RMSE results will be higher (in the hundreds). This is the main reason why the data set is usually critical in establishing very good prediction performance metrics. In our case, due to the limited data set, we expect the RMSE to be in the hundreds while the R2 value to be very close to 1, giving us reassurance that the models and indeed learning even from this limited data set.

Experiments have been made using a combination of MATLAB and Python. Should larger data sets be made available, we expect the RMSE errors to reduce further.

References

Adnan, M., Nahmias Biran, B., Baburajan, V., Basak, K., & Ben-Akiva, M. (2020). Examining impacts of time-based pricing strategies in public transportation: A study of Singapore. *Transportation Research Part A: Policy and Practice*, *140*(December 2019), 127–141. https://doi.org/10.1016/j.tra.2020.08.010

Beck, M. J., Hensher, D. A., & Nelson, J. D. (2021). Public transport trends in Australia during the COVID-19 pandemic: An investigation of the influence of bio-security concerns on trip behaviour. *Journal of Transport Geography*, *96*(March), 103167. https://doi.org/10.1016/j.jtrangeo.2021.103167

Boellaard, R. 2022, 'External validation: a simulation study to compare cross-validation versus holdout or external testing to assess the performance of clinical prediction models using PET data from DLBCL patients', *EJNMMI Research*, vol. 12, no. 1, pp. 4–11.

Botchkarev, A. 2019, 'A New Typology Design of Performance Metrics to Measure Errors in Machine Learning Regression Algorithms', *Interdisciplinary Journal of Information, Knowledge, and Management*, vol. 14, no. 9, pp. 045–76.

Bull, O., Muñoz, J. C., & Silva, H. E. (2021). The impact of fare-free public transport on travel behaviour: Evidence from a randomised controlled trial. *Regional Science and Urban Economics*, *86*(May 2020), 103616. https://doi.org/10.1016/j.regsciurbeco.2020.103616

Eertink, J.J., Heymans, M.W., Zwezerijnen, G.J.C., Zijlstra, J.M., de Vet, H.C.W. &

Ge, Y.-E., Prentkovskis, O., Tang, C., Saleh, W., Bell, M. G. H., & Junevičius, R. (2015). SOLVING TRAFFIC CONGESTION FROM THE DEMAND SIDE. *PROMET - Traffic&Transportation*, *27*(6), 529–538. https://doi.org/10.7307/ptt.v27i6.1734

Halvorsen, A., Koutsopoulos, H. N., Lau, S., Au, T., & Zhao, J. (2016). Reducing subway crowding: Analysis of an off-peak discount experiment in Hong Kong. *Transportation Research Record*, *2544*, 38–46. https://doi.org/10.3141/2544-05

Huan, N., Hess, S., Yao, E., & Liu, W. (2021). Time-dependent pricing strategies for metro lines considering peak avoidance behaviour of commuters. *Transportmetrica A: Transport Science*. https://doi.org/10.1080/23249935.2021.1946203

Kusonkhum, W., Srinavin, K., Leungbootnak, N. & Chaitongrat, T. 2022, 'Using a Machine Learning Approach to Predict the Thailand Underground Train's Passenger', *Journal of Advanced Transportation*, vol. 2022.

Lin, D. Y., Fang, J. H., & Huang, K. L. (2019). Passenger assignment and pricing strategy for a passenger railway transportation system. *Transportation Letters*, *11*(6), 320–331. https://doi.org/10.1080/19427867.2017.1343764

Lovrić, M., Raveau, S., Adnan, M., Pereira, F. C., Basak, K., Loganathan, H., & Ben-Akiva, M. (2016). Evaluating Off-Peak Pricing Strategies in Public Transportation with an Activity-Based Approach. *Transportation Research Record: Journal of the Transportation Research Board*, *2544*(1), 10–19. https://doi.org/10.3141/2544-02

Munawar, H. S., Khan, S. I., Qadir, Z., Kouzani, A. Z., & Mahmud, M. A. P. (2021). Insight into the impact of COVID-19 on Australian transportation sector: An economic and community-based perspective. *Sustainability (Switzerland)*, *13*(3), 1–24. https://doi.org/10.3390/su13031276

Nti, I.K., Nyarko-Boateng, O. & Aning, J. 2021, 'Performance of Machine Learning Algorithms with Different K Values in K-fold Cross Validation', *International Journal of Information Technology and Computer Science*, vol. 13, no. 6, pp. 61–71.

Ou, Yuming; Mihaita A. S.; Chen, Fang; "Big Data Processing and Analysis on the Impact of COVID-19 on Public Transport Delay", in Data Science for COVID-19, (Edited by Utku Kose, Deepak Gupta, Victor Hugo C. de Albuquerque, Ashish Khanna), Oct 2021.

Park, Y., Choi, Y., Kim, K. & Yoo, J.K. 2022, 'Machine learning approach for study on subway passenger flow', *Scientific Reports*, vol. 12, no. 1, pp. 1–20.

Plevris, V., Solorzano, G., Bakas, N. & Seghie, M.E.A. Ben 2022, 'Investigation of performance metrics in regression analysis and machine learning-based prediction models', *The 8th European Congress on Computational Methods in Applied Sciences and Engineering, 5–9 June 2022, Oslo, Norway*, no. June.

Štraub, D. (2020). The effects of fare-free public transport: A lesson from frýdek-místek (Czechia). *Sustainability (Switzerland)*, *12*(21), 1–15. https://doi.org/10.3390/su12219111

Štraub, D. (2020). The effects of fare-free public transport: A lesson from frýdek-místek (Czechia). *Sustainability (Switzerland)*, *12*(21), 1–15. https://doi.org/10.3390/su12219111

Yang, H., & Tang, Y. (2018). Managing rail transit peak-hour congestion with a fare-reward scheme. *Transportation Research Part B: Methodological*, *110*, 122–136. https://doi.org/10.1016/j.trb.2018.02.005

Yen, B. T. H., Tseng, W.-C., Chiou, Y.-C., Lan, L. W., Mulley, C., & Burke, M. (2015). Effects of Two Fare Policies on Public Transport Travel Behaviour: Evidence from South East Queensland, Australia. *Journal of the Eastern Asia Society for Transportation Studies*, *11*, 425–443. https://doi.org/10.11175/easts.11.425

Yen, B. T. H., Tseng, W.-C., Chiou, Y.-C., Lan, L. W., Mulley, C., & Burke, M. (2015). Effects of Two Fare Policies on Public Transport Travel Behaviour: Evidence from South East Queensland, Australia. *Journal of the Eastern Asia Society for Transportation Studies*, *11*, 425–443. https://doi.org/10.11175/easts.11.425

Zou, L., Shu, S., Lin, X., Lin, K., Zhu, J. & Li, L. 2022, 'Passenger Flow Prediction Using Smart Card Data from Connected Bus System Based on Interpretable XGBoost', *Wireless Communications and Mobile Computing*, vol. 2022.