Summer Peaks and Winter Lows: An Analysis of Toronto Island Ferry Ticket Counts*

What Six Years of Data Reveal About Ferry Usage, Visitor Patterns, and Seasonal Consumer Behaviour

Krishna Kumar

September 25, 2024

This report looks at the ferry services to the Toronto Islands, a significant recreational destination accessible exclusively by water. We analyze a data set of Ferry Ticket Counts from May 2018 to September 2024 in order to find patterns and trends in ferry usage and the effects of external factors. Using time series analysis and data visualization, the data shows that there exist definitive seasonal trends, with sharp declines during the pandemic period. Our findings suggest that ferry usage is highly seasonal and very vulnerable to external shocks. The paper concludes by discussing broader implications for tourism, urban and contingency planning, and public transportation management.

1 Introduction

The Toronto Islands are a collection of small islands located off the southern shore of Toronto, Canada. Known for their scenic beauty and recreational activities, they serve as a popular vacation spot for tourists and locals alike. The islands feature beaches, parks, restaurants, and other attractions, making them a prime destination. Access to the Toronto islands is exclusively provided (except private water taxis) by the ferry services operated by the Toronto Parks, Forestry, and Recreational department. These ferries carry thousands of visitors daily, especially during peak season, and are essential in the functioning of the islands.

This paper analyzes the ferry ticket sales and redemption data from the Toronto Islands Ferry Service over the following time period: 1st of January, 2018 to the 20th of September, 2024. The Toronto Islands serve not only as a beloved spot for locals, but also as a popular

*Code and data are available at: https://github.com/krishnak30/Toronto-Island-Ferry-Ticket-Counts-Krishna-K.git.

tourist destination, attracting hundreds of thousands of visitors annually. Given the ferry's role as the primary means of transportation to the islands, understanding its usage patterns is critical for officials, policymakers, stakeholders, and business owners in order to improve service efficiency, prepare for fluctuating demand, accommodate future growth, and develop contingency plans. The significance of this research goes well beyond the ferry system, having impacts on issues of urban mobility, tourism management, and even economic recovery in a post-pandemic society.

The paper begins with a comprehensive overview of the chosen data set, detailing the data collection methods, critical data cleaning details, and statistical tools/resources employed for the analysis. Then, the paper presents a series of visualizations such as graphs and tables to illustrate the trend of tickets over time, including a detailed month-by-month analysis of ferry ticket counts. Our findings revealed significant seasonal trends, with notable peaks in ticket counts during the summer months (June to August), confirming our claim. Lastly, we discuss the implications of our findings and the broader context it serves, focusing on the importance of visitor patterns in improving service efficiency and and addressing potential future challenges.

The paper is structured as follows: Section 2 provides an in-depth discussion of the data, including its source, the critical aspects of the cleaning process, preparation for analysis, as well as discussion of the broader context. Section 3 presents a detailed analysis of ticket sales and redemption patterns, highlighting key trends, seasonal variations, and the impact of external factors such as the COVID-19 pandemic. Finally, section 4 discusses the implications of these findings, explores potential applications, and addresses the limitations of this study, offering suggestions for future research.

2 Data

2.1 Overview

The data used in this paper is derived from Open Data Toronto and is read into this paper using the library 'opendatatoronto' (Gelfand 2022). The data were cleaned and analyzed using the statistical programming language 'R' (R Core Team 2024). Simulating, cleaning, and testing the data was done with the help of the following packages: 'tidyverse' (Wickham et al. 2019), 'janitor' (Firke 2023), 'knitr' (Xie 2014), 'kableExtra' (Zhu 2024), 'dplyr' (Wickham et al. 2023), 'readr' (Wickham, Hester, and Bryan 2024), 'here' (Müller 2020), and 'lubridate' (Grolemund and Wickham 2011). The graphs and plots were made using 'ggplot2' (Wickham 2016).

The data is published by Parks, Forestry, and Recreation and is updated on an hourly basis, providing real-time data and updates. The data provides information regarding the ferries that takes visitors to the Toronto Island Park, which is considered a popular summer spot for tourists and locals alike. These ferries carry passengers to and from Centre Island, Hanlan's

Point, and Ward's Island. The raw data features the counts for the total number of tickets redeemed and tickets sold in particular time periods. The time stamps are shows in 15 minute intervals and contain the data for redemptions and sales in the past 15 minutes. The time period for the data set ranges from the 1st of January, 2018 to the 20th of September, 2024 and contains around 229,000 rows/data entries.

Hence, the chosen variables for the data were ID (unique identifier for each entry), data and time stamps, number of tickets redeemed for ferry access, and the number of tickets purchased. The time stamp column was broken down into two new 'Date of transaction' and 'Time of transaction' columns for better and deeper analysis.

The variable 'Number of Redemption Tickets' accounted for the number of people who redeemed their tickets to use the ferry to reach the Toronto Islands. The variable 'Number of Sales Tickets' accounted for the number of people who purchased tickets from in-person POS-Kiosks or online platforms, and encompass all product types.

A sample of the cleaned data set can be found in Table 1 showcasing the variables and the structure of the data.

Table 1: Sample of cleaned Ferry Ticket Counts Data

ID	Time	Tickets Redeemed	Tickets Sold	Date
317	23:45:00	0	13	2024-09-20
318	23:30:00	7	21	2024-09-20
319	23:15:00	1	5	2024-09-20
320	23:00:00	5	10	2024-09-20
321	22:45:00	1	12	2024-09-20
322	22:30:00	5	8	2024-09-20
323	22:15:00	2	26	2024-09-20
324	22:00:00	10	26	2024-09-20
325	21:45:00	13	23	2024-09-20
326	21:30:00	14	18	2024-09-20

2.2 Data Cleaning

The data cleaning process involved breaking down the 'timestamp' column into two 'Date' and 'Time' columns for a better and more clear analysis. Along with changing names of the columns, deleting duplicate columns, and removing entries from the data that served as outliers. The data was filtered to only analyse data from 2018 to 2024, removing years 2015 - 2018. This was because the data set was too large to analyse altogether efficiently and the new filtered data set allowed for a closer analysis of the impact of COVID-19.

2.3 Broader Context

This data was the only of its kind on the Open Data Toronto library, providing the numbers for the ticket counts of ferry use for the Toronto Islands. This data was chosen due to the broader context and the potential for discussion. This data set is essential in understanding the urban mobility and recreational trends in Toronto.

Firstly, as Toronto Islands Park is a major recreational area, it affects millions of visitors annually. The ferry contributes to the local economy through tourism-related spending. Understanding these patterns can help us gain insights into seasonal fluctuations in tourism and help local businesses (close to Toronto Islands) optimize their offerings. For example, increased ferry usage during the summer may correlate with higher foot traffic for restaurants and vendors on the islands. Secondly, while private water taxis are available to get to the islands, a large majority prefers to use the ferry due to the relatively lower price point and the fact that the ticket is free on the way back. By analyzing redemption and sales tickets, city planers and transportation officials can better understand the demand patterns, which can help them schedule service improvements. Lastly, analyzing ticket counts and patterns of increase usage can be important when designing public safety protocols. By understanding the usage patters, officials are able to have tailored contingency plans. For example, at the event of a fire at the islands and when evacuation is needed, less support might be required in the winter season compared to the summer.

2.4 Basic Summary Statistics

The summary statistics are provided in Table 2 and allow us to get an understanding of the data in a very comprehensible manner. The high volume of tickets sold and redeemed indicates a healthy and significant demand of the ferry and the closeness of the values conveys strong customer engagement and a reliable service. On average, 46.92 tickets were sold every 15 minutes during working hours, and 47.37 tickets were redeemed, indicating that most customers utilized their tickets. However, the small gap implies some no-shows or unused tickets. The maximum numbers of redemption and sales tickets (7,216 and 7,229) showcases the ferry's ability to handle high customer volumes and the closeness of these values tells us that, on busy days, the ferry usually operates at capacity.

Moreover, the relationships between the variables allow us to understand more about the patterns and usage. The high correlation between the number of tickets redeemed and sold ($\sim$ 98% of sold tickets were redeemed!) suggests that the customer satisfaction is high and that the ferry is able to meet demands. The averages and the maximum numbers show us that there is high correlation between the number of tickets sold and redeemed, suggesting good operational viability. Lastly, the minimum values of 0 for both types of tickets suggest that there are periods of little-to-no demand, which typically occur during off-peak hours and off-seasons (which we will analyse soon).

Table 2: Summary Statistics of Ticket Counts

Avg Redeemed	Avg Sold	Max Redeemed	Max Sold	Count
47.36944	46.91741	7216	7229	168691

2.5 Ferry Ticket Counts Over Time

After studying the summary statistics, we have an idea of the capacity and performance of the ferry, along with the relationships between the variables. Next, with the use ofFigure 1, we can study the patterns of ticket counts from 2018 to 2024. For figure 1, we can see that there is a definitive pattern in the usage numbers when looking at each year. We see that there are spikes during the mid-year months for each year, indicating higher popularity of the Toronto Islands during those times and suggesting that the islands serve as a significant recreational destination.

We can also see that for every year, there is a slow climb from the beginning of the year (which starts very low) indicating an off-season and the start of the tourist season. Conversely, there is a gradual decline during the end of the year (also indicating an off-season) which confirms a seasonal drop in demand as the cold weather deters visitors. The highs in each year are steady at 800 - 1000 for the number of tickets sold in one interval, however, the highs for the number of tickets redeemed are usually lower, but ever-fluctuating. This fluctuation and discrepancy could indicate that not all passengers redeem their tickets within the same time frame or that there are differences in purchasing behavior.

Notably, Figure 1 also shows an anomaly during the years of 2020 and 2021, conveyed by highly depressed sales and redemption ticket counts. This can be attributed to the COVID-19 pandemic, which lead to widespread public facility closures and reduced operational capacities. The peak ticket count in 2020, reaching only about 200, is in high contrast with every other year, highlighting the pandemic's massive impact on public transportation and tourism.

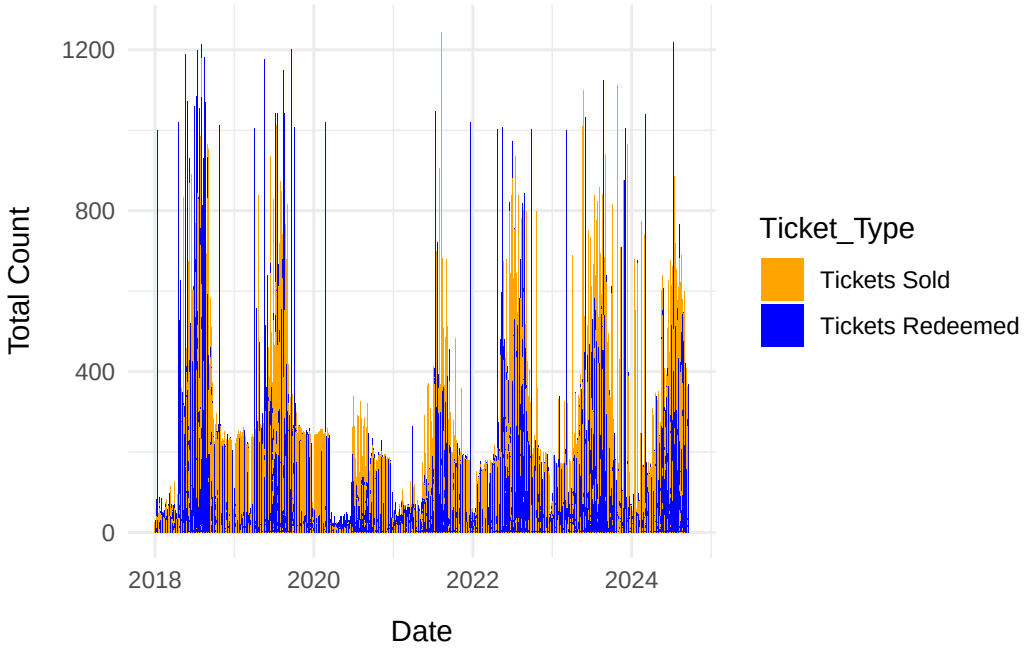

Figure 1: Tickets Sold and Redeemed per 15 min Interval

2.6 Month-by-Month Analysis

To build on our understanding of overall ticket trends observed in the previous section, it is essential to analyze the month-by-month usage to better understand usage patterns in each year. These patterns may not be visible in the previous section due to the large time range of the data.

Figure 2 shows a monthly analysis of redemption and sales ticket counts for each year. The data in this figure has been grouped together and summed. i.e. the ticket counts for each of the 12 months in every year from 2018 to 2024 has been summed and aggregated so that we can see the monthly trends over the past 6 years.

The graph shows distinct peaks in ticket sales and redemption during the summer months (June to August), aligning with the typical tourist season and confirming the ferry's role in providing transportation to the islands. In addition to the tourist season, individuals might prefer to go to the islands in these months due to the beach activities and the availability of restaurants and vendors. Conversely, the months of October through April show noticeably lower ticket counts, consistent with weather conditions in Toronto where temperatures often drop below zero degree Celsius. Hence, deterring visitors from the islands.

Interestingly, every month shows a close correlation between the sales and redemption tickets and what implications this has for the ferry service. The values for sales and redemption ticket

counts are extremely close together for all months as displayed in Figure 2.

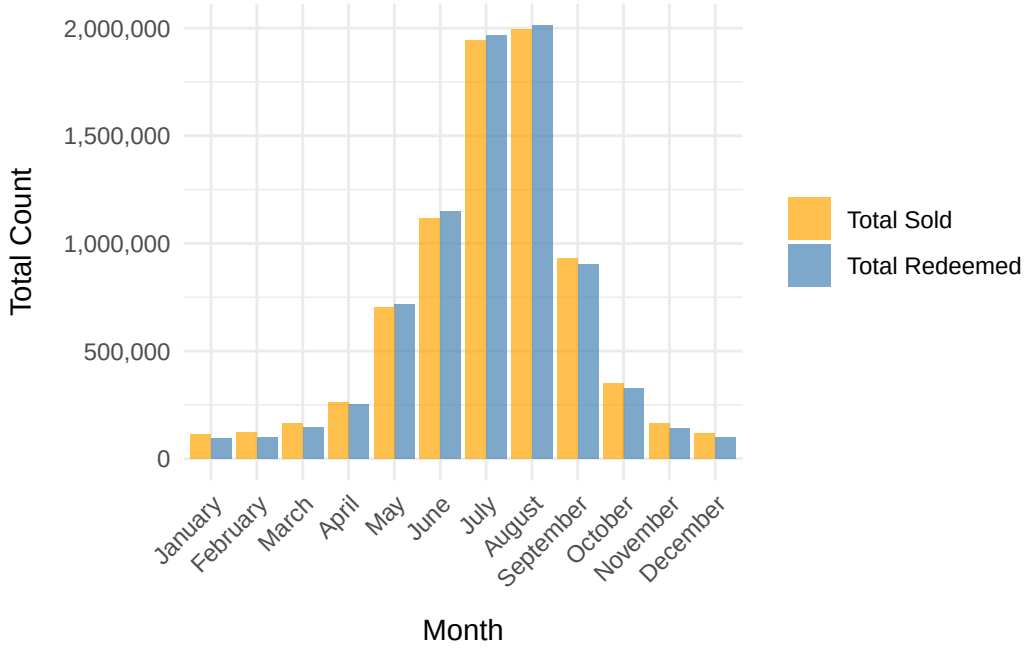

Figure 2: Aggregate Monthly Ticket Counts (Sold and Redeemed)

3 Discussion

This paper ferry ticket data from Open Data Toronto, focusing on the trends of ticket sales and redemptions from the 1st of January, 2018 to the 20th of September, 2024. Through summary statistics and visualizations, we identified distinct patters in ferry usage and how external factors have influenced these trends. The time series analysis illustrated a consistent peak in ferry activity during the summer months, while the month-by-month backed up this claim and showed how seasonal variations impacted ticket counts.

The analysis that ticket sales and redemptions are highly seasonal due to the nature of the Toronto Islands and the activities present. We saw peaks during the summer months, which align with the tourist season in Toronto, showing that the ferry plays an important role in facilitating access to the islands. We concluded that the ferry is not only a transportation service, but also contributes to the local economy by supporting tourism-related activities. Understanding these patterns allows stakeholders, business owners, and officials to plan for peal periods and allocate resources accordingly.

The ferry and the number of visits to the Toronto islands is highly affected by external factors, as we saw the notable decline in ticket counts during the COVID 19 pandemic years (2020

- 2021). This insight is essential for policymakers and service providers to consider as it emphasizes the need for contingency planning and adapting in unprecedented circumstances.

3.1 Limitations and Next Steps

While the analysis provides valuable insights into ferry ticket trends and Toronto Islands Popularity, it does not come without limitations. The data set is primarily focused on ticket sales and redemptions, which may not fully capture the complexity of factors influencing ferry usage and the number of visits to Toronto Islands. For example, weather conditions, public events, changes in tourism trends, or construction were not incorporated in this research. Furthermore, just because we saw an increase in ferry usage in the summer months doesn't necessarily mean that these are due to the tourists. Due to the cold weather, people (even locals) prefer to visit Toronto islands in the summer. Hence, it is difficult to determine whether the jump in ferry tickets were due to the increase in tourism or just because of the summer months.

Moreover, I believe that this research can greatly benefit by complementing with another data set such as the number of tourists visiting Toronto, or analyzing the activities at the Toronto Islands and determining which season they're better for. Incorporating other data sets would allow us to reach a more definitive conclusion about a claim.

Overall, this paper provides a great analysis of the ferry ticket counts for Toronto islands and serves as a solid foundation for future research.

References

Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://CRAN.R-project.org/package=janitor.

Gelfand, Sharla. 2022. *Opendatatoronto: Access the City of Toronto Open Data Portal*. https://CRAN.R-project.org/package=opendatatoronto.

Grolemund, Garrett, and Hadley Wickham. 2011. "Dates and Times Made Easy with lubridate." *Journal of Statistical Software* 40 (3): 1–25. https://www.jstatsoft.org/v40/i03/.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

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.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2024. *Readr: Read Rectangular Text Data*. https://CRAN.R-project.org/package=readr.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC.

Zhu, Hao. 2024. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.