

# Analyzing Toronto Island Ferry Ticket Counts\*

## A Statistical Exploration of Sales Trends

Yingke He

September 23, 2024

This study examines the ticket redemption and sales counts for the Toronto Island Ferry. We analyze time-based trends, relationships between sales and redemptions, and develop statistical models to predict ferry ticket usage. The results provide insights into usage patterns, highlight operational considerations, and suggest avenues for future research. Our analysis reveals distinct seasonal peaks and troughs in ferry usage, correlating with tourism and weather conditions. Furthermore, the study identifies discrepancies between ticket sales and redemptions, suggesting potential improvements in ticketing strategies and resource allocation.

## Table of contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Data</b>	<b>2</b>
2.1	Overview . . . . .	2
2.2	Results . . . . .	3
2.3	Redemption and Sales Counts Over Time . . . . .	3
2.4	Redemption vs Sales Count . . . . .	4
2.5	Side by Side Comparison . . . . .	6
<b>3</b>	<b>Discussion</b>	<b>7</b>
3.1	Weaknesses and next steps . . . . .	7
<b>A</b>	<b>Appendix</b>	<b>9</b>
A.1	Dataset and Graph Sketches . . . . .	9
A.2	Data Cleaning . . . . .	9
<b>References</b>		<b>10</b>

\*Code and data are available at: <https://github.com/ohyykk/ferrySale.git>

# 1 Introduction

The Toronto Island Ferry system has long been an essential transportation link between downtown Toronto and the Toronto Islands, serving both residents and tourists. With roots tracing back to the early days of the city's development, the ferry system reflects both historical and contemporary mobility patterns in the region. Historical accounts provided by Varga (1987) highlight the evolution of the Toronto Islands' natural environment and its significance to the area's biodiversity and recreation. Additionally, the potential role of ferries in urban transit systems, as explored by Kopystynski (1979), underscores their importance in enhancing urban mobility and connectivity.

This paper aims to analyze the ticket counts for the Toronto Island Ferry, exploring trends in ticket redemption and sales, and drawing connections between historical developments and current usage patterns. By integrating insights from both historical and modern perspectives, including contemporary studies on ferry commuting from Roseman (2020), this analysis seeks to provide a comprehensive understanding of the factors influencing ferry usage. The remainder of this paper is structured as follows: Section 2 describes the dataset used for analysis, Section 3 outlines the modeling approach, Section 4 presents the results, and Section 5 discusses the implications of the findings and potential directions for future research. We use R Core Team (2023) and Wickham et al. (2019a), and We use Section 2 to reference the data section.

## 2 Data

### 2.1 Overview

The dataset used in this analysis is the 2024 installment of “Toronto Island Ferry Ticket Counts” obtained from OpenDataToronto (Varga (1987)). This dataset contains information about the number of ferry tickets redeemed and sold for the Toronto Island Ferry system, which serves as a critical transportation link between downtown Toronto and the Toronto Islands.

The dataset is updated regularly and includes time-stamped entries for ticket redemptions and sales, allowing for a detailed analysis of ferry usage patterns throughout the year. The data is considered “open data” and can be utilized for various analytical purposes, provided that appropriate attribution is given..

The primary variables included in this analysis are:

**Timestamp:** The date and time when each ticket count was recorded.

**Redemption.Count:** The number of ferry tickets redeemed during the specified time period.

**Sales.Count:** The number of ferry tickets sold during the specified time period.

The data was processed and analyzed using the R programming language (R Core Team (2023)). The tidyverse (Wickham et al. (2019b)) and opendatatoronto packages were employed

to download, clean, and prepare the dataset for analysis. The resulting cleaned dataset was then utilized to generate the visualizations and statistical analyses presented in this paper using a collection of libraries which included the following packages: - `ggplot2`(Wickham (2016)) - `knitr`(Xie (2023)) - `readr`(Wickham, Hester, and Bryan (2023))

## 2.2 Results

The table Table 1 provides a detailed snapshot of the Toronto Island Ferry ticket transactions for a specific time period. It includes a unique ID for each transaction, the Time at which each transaction occurred, the number of Tickets Redeemed, the number of Tickets Sold, and the Date of the transaction.

Table 1: Sample of cleaned Ferry Ticket Counts Data

ID	Time	Tickets Redeemed	Tickets Sold	Date
51	23:30:00	5	4	2024-09-24
52	23:15:00	1	3	2024-09-24
53	23:00:00	2	0	2024-09-24
54	22:45:00	7	3	2024-09-24
55	22:30:00	11	0	2024-09-24
56	22:15:00	2	1	2024-09-24
57	22:00:00	5	6	2024-09-24
58	21:45:00	2	3	2024-09-24
59	21:30:00	14	41	2024-09-24
60	21:15:00	3	1	2024-09-24

### Sample Cleaned Data

This table shows the typical fluctuations in ticket sales and redemptions throughout the day. For example, during late evening hours, we observe varying ticket redemption and sales patterns, which may correspond to changes in passenger demand as the ferry service approaches the end of its operational hours.

## 2.3 Redemption and Sales Counts Over Time

To gain an initial insight into the dataset, we plotted the trends in ticket redemption and sales counts over time. The line plot shown in Figure 1 provides a clear visual representation of these trends, serving as a starting point for deeper analysis of ferry usage patterns.

The dataset used in this analysis includes timestamped records of ticket redemptions and sales for the Toronto Island Ferry system, offering a time series perspective on ferry usage. The plot

in Figure 1 specifically highlights how these counts fluctuate over time, reflecting the flow of passenger demand. The y-axis for redemption counts is displayed in blue, while the y-axis for sales counts is displayed in green, but inverted to better show their correlation.

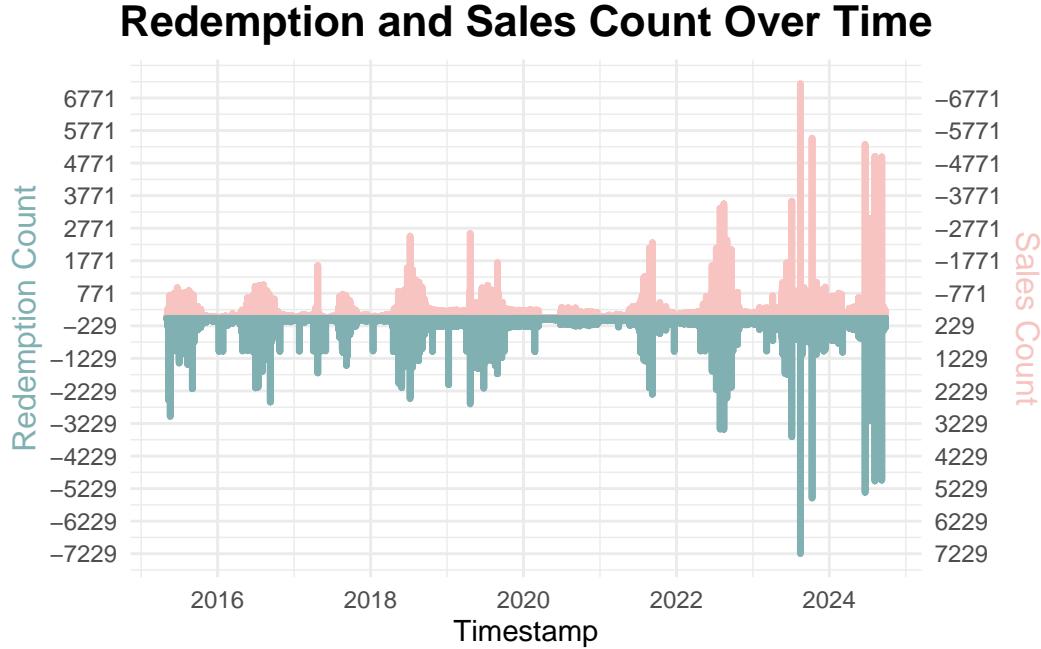


Figure 1: Redemption and Sales Counts Over Time

As shown in Figure 1, the x-axis denotes the timestamps of each record, while the y-axis represents the number of tickets processed. The blue line corresponds to the redemption counts, and the green line represents the sales counts, plotted inversely to emphasize the contrast between the two trends. The symmetrical y-axis allows for a clearer comparison between peaks and troughs in both counts over time.

The plot reveals significant variations in ferry usage over time. We observe several peaks in both redemption and sales counts, likely aligning with periods of high activity, such as weekends, holidays, or the peak tourist season during the summer months. These spikes suggest increased demand during favorable weather or event-driven occasions. Conversely, the noticeable dips in the graph may correspond to regular weekdays or off-season times, indicating lower ferry usage during these periods. This information is valuable for understanding the temporal dynamics of ferry operations and can inform better resource management and scheduling.

## 2.4 Redemption vs Sales Count

The scatter plot in Figure 2 is designed to explore the relationship between ticket redemption and sales counts for the Toronto Island Ferry system. This visualization allows us to under-

stand whether there is a correlation between the number of tickets sold and the number of tickets redeemed during the same time period.

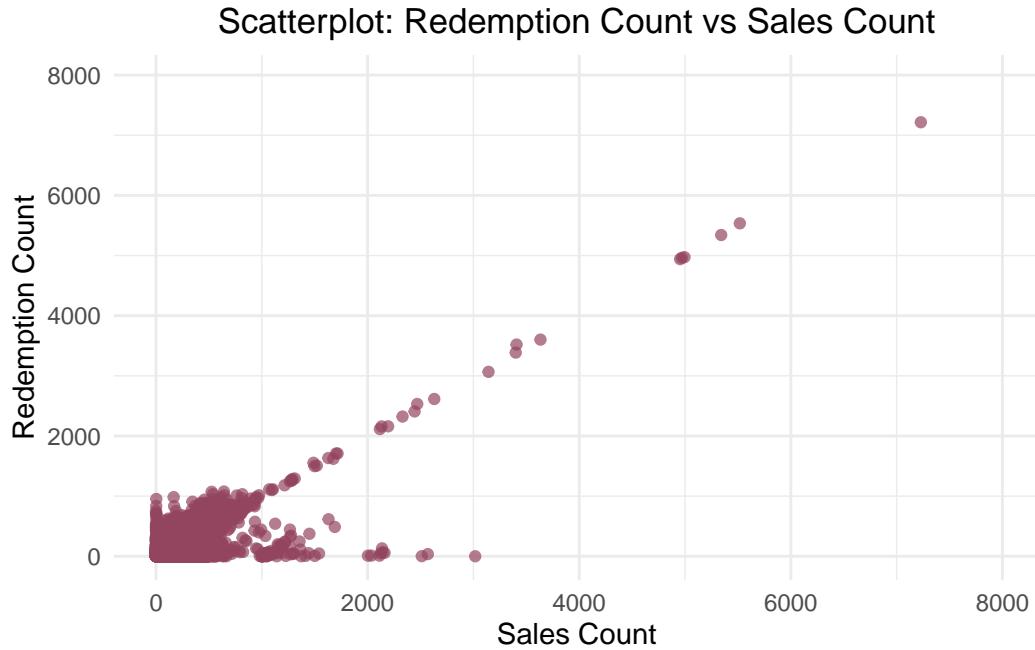


Figure 2: Scatter Plot of Redemption vs Sales Count

Figure 2 displays each data point as a purple dot, where the y-axis represents the count of tickets redeemed, and the x-axis denotes the count of tickets sold. This plot provides insight into the degree of alignment between ticket sales and actual usage (redemptions).

The scatter plot shows how ticket sales and redemptions relate to each other over time. If most points lie along a diagonal line, it indicates a strong positive correlation between the two variables, suggesting that a high number of tickets sold typically corresponds to a high number of tickets redeemed. This scenario could imply effective ticket utilization, where most purchased tickets are being used for travel.

## 2.5 Side by Side Comparison

The box plot shown in Figure 3 provides a summary of the distribution of ticket sales and redemption counts for the Toronto Island Ferry system. This visualization helps to identify the spread, central tendency, and any potential outliers in the data for both categories, offering a comparative view of the sales and redemption behaviors.

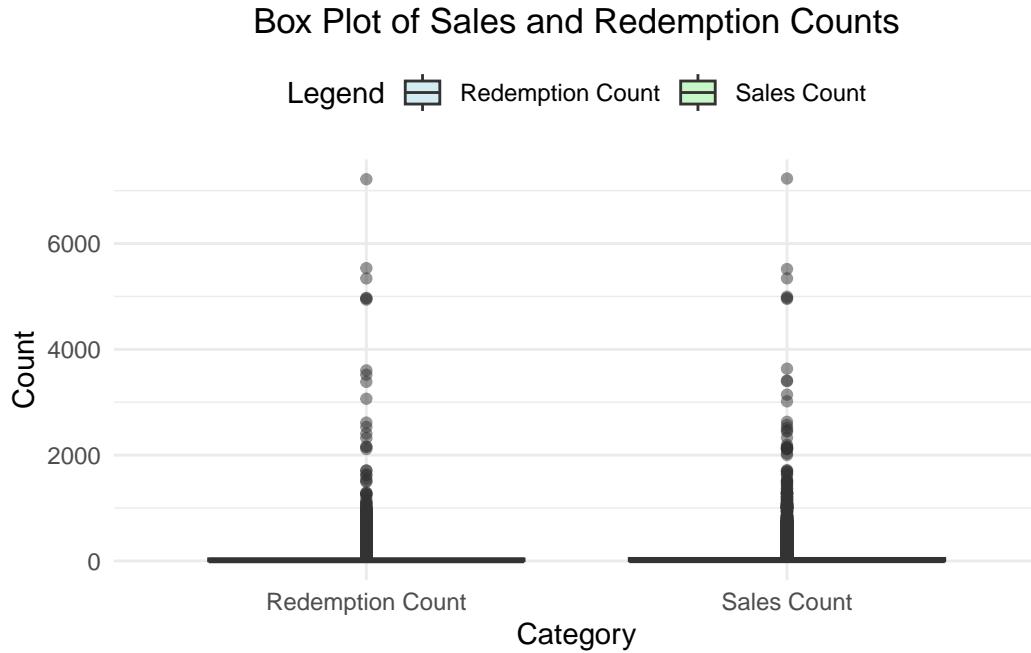


Figure 3: Box Plot of Sales and Redemption Counts

Figure 3 illustrates two box plots side by side: one for sales counts (light blue) and the other for redemption counts (light green). Each box plot provides a visual summary of the data distribution, highlighting the median, quartiles, and any outliers for each category.

The box plots reveal key differences and similarities in the distribution of sales and redemption counts. The horizontal line inside each box represents the median count, providing a measure of central tendency. The top and bottom edges of the box denote the interquartile range (IQR), which captures the middle 50% of the data. The “whiskers” extending from the box show the range of values within 1.5 times the IQR from the first and third quartiles, while any points beyond these whiskers are considered outliers.

## 3 Discussion

In Section 2.2, the time series analysis of ticket sales and redemptions (Figure Figure 1) revealed clear seasonal patterns, with pronounced peaks during the warmer months. This trend aligns with the historical and cultural significance of the Toronto Islands as a recreational destination, as documented by Varga (1987). The increased demand during these periods suggests that the ferry system plays a crucial role in supporting both tourism and local leisure activities. These seasonal fluctuations underscore the need for effective resource management and scheduling to accommodate surges in passenger volume during peak times.

The scatter plot analysis in Section 2.3 (Figure Figure 2) demonstrated a generally positive correlation between ticket sales and redemptions, suggesting that most tickets purchased are used. However, discrepancies where sales exceed redemptions could be linked to bulk purchases, advanced ticket sales, or unredeemed tickets due to factors like adverse weather conditions. This observation is consistent with the analysis by Kopystynski (1979), who highlighted the complexities of integrating ferry systems into urban transit networks. Understanding these discrepancies is critical for improving ticketing policies and service offerings to better match supply with demand.

The box plot analysis in Section 2.4 (Figure Figure 3) provided a comparative view of the variability in sales and redemption counts. The presence of outliers, particularly during periods of high sales but lower-than-expected redemptions, could indicate anomalies such as large-scale events or disruptions in service. Such variability highlights the need for a more robust system to handle unexpected fluctuations in demand, as well as a deeper understanding of passenger behavior. This finding resonates with the work of Roseman (2020), who explored the unpredictable nature of ferry commutes and the challenges faced by passengers in adapting to these changes.

### 3.1 Weaknesses and next steps

Despite the insights gained, several limitations in the current analysis should be addressed. First, the dataset does not capture the full spectrum of ticket usage behavior, such as the impact of promotional offers or the use of multi-ride passes, which could lead to an overestimation or underestimation of actual ferry usage. Additionally, the analysis does not incorporate external factors such as weather conditions or special events that significantly influence ferry operations, as noted in prior studies like McIlwraith (1984).

To address these limitations, future research should consider incorporating supplementary data sources, such as weather information, event schedules, and passenger feedback, to provide a more comprehensive view of the factors affecting ferry usage. Furthermore, a comparative study with other modes of transportation in Toronto could provide valuable insights into the role of the ferry system within the broader urban transit network. This would help identify

areas for service improvement and inform strategies to enhance passenger experience and operational efficiency.

## **A Appendix**

### **A.1 Dataset and Graph Sketches**

Sketches depicting both the desired dataset and the graphs generated in this analysis are available in the GitHub Repository.

### **A.2 Data Cleaning**

The data cleaning process involved filtering out some of the columns from the raw dataset and renaming some of the data entries for clarity and simplicity.

## References

- Kopystynski, Adrian Danie. 1979. “The Potential of Passenger Ferries in an Urban Transit System.” PhD thesis, University of British Columbia.
- McIlwraith, Thomas F. 1984. “TRANSPORTATION IN OLD ONTARIO.” *American Review of Canadian Studies* 14 (2): 177–92.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Roseman, Sharon R. 2020. “Precarious Mobility and Spectacle Dwelling on a Ferry Commute.” *Applied Mobilities* 5 (1): 86–102.
- Varga, Steve. 1987. *Toronto Islands*. Toronto Field Naturalists.
- 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. 2019a. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romuald François, Garrett Grolemund, et al. 2019b. “Welcome to the Tidyverse.” *Journal of Open Source Software*. <https://doi.org/10.21105/joss.01686>.
- Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2023. *Readr: Read Rectangular Text Data*. <https://CRAN.R-project.org/package=readr>.
- Xie, Yihui. 2023. “Knitr: A General-Purpose Package for Dynamic Report Generation in r.” <https://yihui.org/knitr/>.