

Analyzing Temporal Patterns and Passenger Flow in Toronto Island Ferry Ticket Sales*

A Time Series Approach

John Zhang

September 24, 2024

This paper analyzes Toronto Island ferry ticket sales over a specific period, focusing on time series patterns and passenger flow trends. By examining both sales and redemptions data at 15-minute intervals, we uncover daily and seasonal trends in ferry usage. These insights can help optimize ferry operations and enhance visitor experience. The analysis highlights peak times of demand and usage variations over different seasons.

Table of contents

1	Introduction	1
2	Data	2
2.1	Raw Data	2
2.2	Modifications and Reasoning	3
2.2.1	Filtering for Full-Year Data (2016-2023)	3
2.2.2	Adding a “Total Sales” Column	3
2.2.3	Extracting Date Components from Timestamps	3
3	Result	4
3.1	Trend Detection: Daily, Monthly, and Seasonal Patterns	4
3.2	Passenger Flow: Peak Time Identification	6
3.3	Day-of-Week Analysis	6
3.4	Peak Day Analysis	7

*Code and data are available at: https://github.com/Clearsky21z/Toronto_Island_Ferry_Ticket_Sale_Analysis.

4	Discussion	8
4.1	Key Insights and Implications	8
4.2	Weaknesses and Future Directions	9
A	Appendix	10
A.1	Dataset and Graph Sketches	10
A.2	Data Cleaning	10
A.3	Data Testing	10
A.4	Attribution Statement	11
	References	12

1 Introduction

Toronto Island Park is a popular recreational destination for residents and tourists alike, situated in Lake Ontario just off the coast of downtown Toronto. Accessible primarily via ferries, the park includes several islands such as Centre Island, Hanlan’s Point, and Ward’s Island. These islands offer a variety of activities, including beaches, nature reserves, trails, and historical landmarks like the Gibraltar Point Lighthouse. The park’s popularity peaks during the summer months, when visitors flock to its beaches and recreational facilities.

Ferry transportation to and from the islands is operated year-round by the Jack Layton Ferry Terminal. Passengers can purchase ferry tickets either online or at point-of-sale (POS) kiosks at the terminal, with the tickets covering both the outbound and return journeys. Online ticket holders benefit from separate boarding lines during peak times, which helps ease congestion. Although ferries run throughout the year, demand is highest on weekends, holidays, and during the summer months, with the busiest travel periods typically between 10 a.m. and 2 p.m. for trips to the islands and from 5:30 p.m. to 9 p.m. for return trips to the city. The ferry route from Centre Island sees the most traffic, and wait times for the return trip can exceed an hour during peak periods.

Given the significant variation in ferry demand throughout the year and at different times of the day, analyzing ferry ticket sales data provides valuable insights into temporal patterns of ferry usage. Understanding these patterns can help optimize ferry schedules, manage peak loads, and improve the overall passenger experience. This paper uses time series analysis to investigate ferry ticket sales and redemptions, exploring daily, weekly, and monthly trends, seasonal patterns, peak usage times, and the busiest days of the year.

The remainder of the paper is structured as follows. Section 2 introduces the dataset and its source, as well as key variables used in the analysis. Section 3 presents the findings from the time series analysis, including trend detection, seasonality, and peak times of usage. Section 4 provides a discussion of these results, highlighting key insights and offering suggestions for future work.

2 Data

The dataset for this analysis was obtained from Open Data Toronto, which provides real-time ferry ticket sales and redemption counts for the Toronto Island Ferry system. The data covers ticket sales and redemptions at 15-minute intervals, including both online and point-of-sale (POS) kiosk sales. It contains information on ferry activity to and from Centre Island, Hanlan’s Point, and Ward’s Island. This dataset is particularly valuable for understanding ferry demand patterns and optimizing operations to meet passenger flow during peak times.

2.1 Raw Data

The raw dataset starts from July 2015 and continues to the most recent available date. For the purpose of this analysis, I focused on complete years of data between 2016 and 2023, to avoid incomplete yearly trends. The original raw dataset consists of four columns:

1. `X_id`: An identifier for each record.
2. `Timestamp`: The date and time when the ferry transaction was recorded, stored at 15-minute intervals.
3. `Redemption_Count`: The number of ferry tickets redeemed in each 15-minute interval.
4. `Sales_Count`: The number of ferry tickets sold in each 15-minute interval.

Table 1 is a sample of the raw data using `head()` to show the structure and contents of the dataset. As seen, each row represents the ferry ticket activity for a specific 15-minute interval, showing the timestamp, number of redemptions, and the number of sales.

Table 1: Sample of Raw Data

<code>X_id</code>	<code>Timestamp</code>	<code>Redemption_Count</code>	<code>Sales_Count</code>
1	2024-09-23 16:45:00	0	38
2	2024-09-23 16:30:00	60	70
3	2024-09-23 16:15:00	48	38
4	2024-09-23 16:00:00	51	36
5	2024-09-23 15:45:00	52	49
6	2024-09-23 15:30:00	36	33

2.2 Modifications and Reasoning

2.2.1 Filtering for Full-Year Data (2016-2023)

To provide a more comprehensive analysis and to avoid the issues that arise from incomplete data, such as partially recorded years, I filtered the dataset to include only the full-year data

from 2016 to 2023. This allows for more consistent year-over-year comparisons and avoids the distortion of trends that could result from partial data (e.g., incomplete months in 2015 or the current year).

2.2.2 Adding a “Total Sales” Column

In the original dataset, sales and redemption data were recorded in separate columns, which made it harder to quickly assess the total ferry ticket activity. To streamline the analysis, I created an additional column called `Total_Sales`, which is simply the sum of `Redemption_Count` and `Sales_Count`. This new column allows for faster aggregate comparisons and simplifies the time series analysis.

2.2.3 Extracting Date Components from Timestamps

To facilitate time series analysis, I extracted several useful features from the `Timestamp` column, such as year, month, week, day of the week, and hour. These features allow for granular analysis of patterns over time, including daily, weekly, and seasonal trends. This extraction makes it easier to analyze peak times and demand variations by breaking down the timestamp into relevant components.

Table 2 shows a sample of modified data.

Table 2: Sample of Modified Data

ID	Timestamp	Redemption Count	Sales Count	Total Sales	Date	Year	Month	Week	Day of Week	Hour
19562	2023-12-31 23:30:00	2	1	3	2023- 12-31	2023	Dec	53	Sun	23
19563	2023-12-31 23:15:00	6	3	9	2023- 12-31	2023	Dec	53	Sun	23
19564	2023-12-31 23:00:00	26	10	36	2023- 12-31	2023	Dec	53	Sun	23
19565	2023-12-31 22:45:00	15	8	23	2023- 12-31	2023	Dec	53	Sun	22
19566	2023-12-31 22:30:00	3	0	3	2023- 12-31	2023	Dec	53	Sun	22
19567	2023-12-31 22:15:00	14	2	16	2023- 12-31	2023	Dec	53	Sun	22

3 Result

3.1 Trend Detection: Daily, Monthly, and Seasonal Patterns

The analysis of daily ferry ticket sales reveals clear patterns of visitor flow throughout the year. As illustrated in Figure 1, ticket sales for 2023 peak prominently during the summer months of June to August, with noticeable spikes during weekends. This trend emphasizes the importance of the Toronto Islands as a leisure destination, particularly during the warmer months when outdoor activities are in high demand.

In Figure 2, the monthly trends further confirm the significant increase in ferry usage during the summer months. The highest peaks are observed in July and August across multiple years, reinforcing the idea that ferry demand is closely linked to the tourism and recreational seasons. This pattern is consistent year-over-year, with slight variations depending on external factors such as public events, holidays, and the influence of the COVID-19 pandemic in certain years.

The seasonality of ferry ticket sales is a vital insight for operational planning, as it highlights the months where additional resources and staffing are required to meet the increased demand. Conversely, the sharp decline in sales during the winter months suggests that resources can be scaled back during the off-season to optimize operational efficiency.

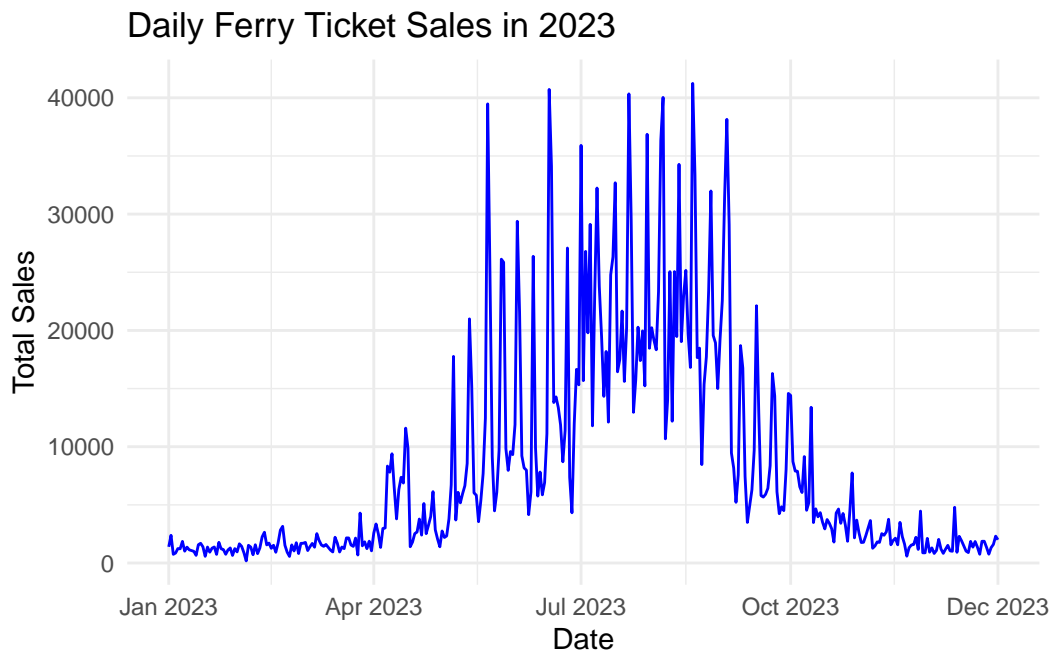


Figure 1: Daily Ferry Ticket Sales in 2023

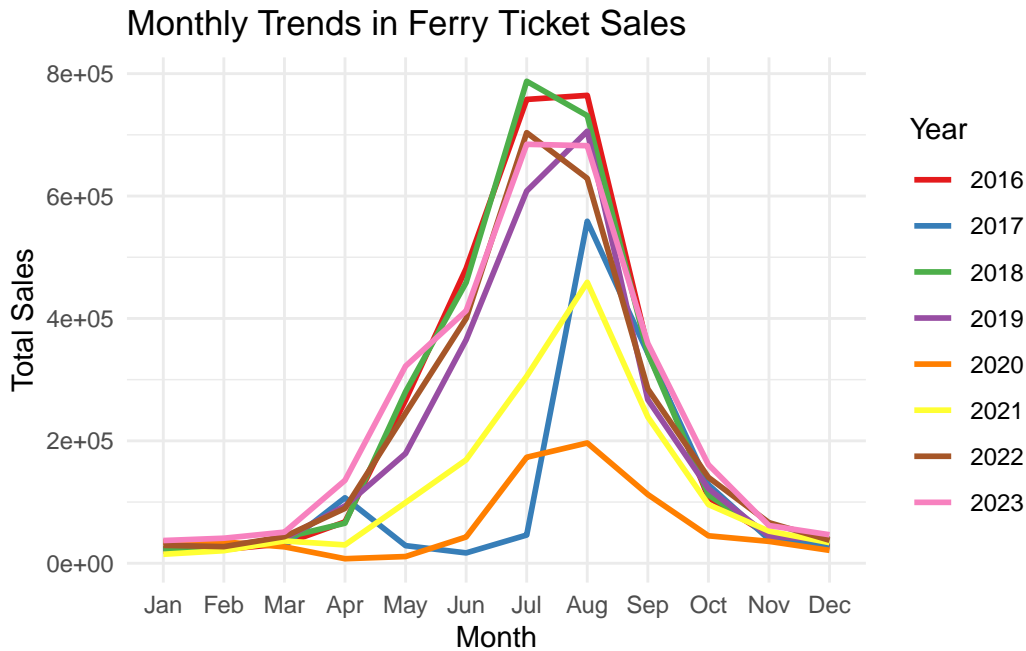


Figure 2: Monthly Trends in Ferry Ticket Sales

3.2 Passenger Flow: Peak Time Identification

The analysis of hourly ferry ticket sales, as shown in Figure 3, provides a detailed view of passenger flow throughout the day. The peak time for ferry ticket sales occurs between 10 a.m. and 3 p.m., reflecting the tendency of visitors to travel to the islands during late mornings and early afternoons. A secondary, smaller peak occurs in the late afternoon, as visitors return to the city.

This information is critical for ferry operators to adjust their schedules and staffing levels to accommodate peak passenger loads, especially during summer weekends. By anticipating these high-traffic periods, ferry operations can be more responsive to demand, reducing wait times and improving the overall passenger experience.

3.3 Day-of-Week Analysis

As shown in Figure 4, ferry usage is consistently higher on weekends, with Saturdays and Sundays being the busiest days across all years. This pattern is driven by recreational visitors, families, and tourists who tend to visit the Toronto Islands during their leisure time.

This day-of-week analysis emphasizes the need for strategic scheduling and additional ferry services during weekends to avoid congestion and manage the large influx of passengers. Moreover,

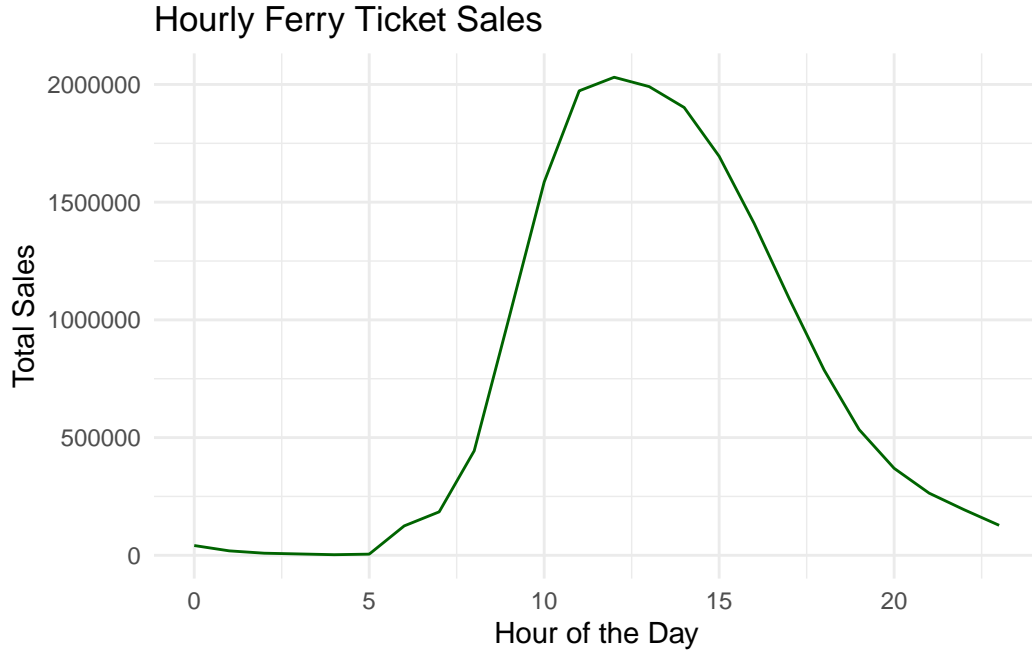


Figure 3: Hourly Ferry Ticket Sales

understanding the day-of-week patterns can help operators allocate resources more effectively and improve customer service during peak times.

3.4 Peak Day Analysis

The peak day analysis, presented in Table 3, highlights the specific days of the year with the highest ferry ticket sales. These peak days often coincide with public holidays, long weekends, or major summer events. For instance, August 1, 2016, and August 20, 2017, mark notable peaks in ferry ticket sales, likely driven by holiday-related traffic.

Understanding these peak days is crucial for operational planning. During these high-demand days, ferry operators can anticipate higher passenger loads and prepare accordingly by increasing the number of ferry trips, deploying additional staff, and improving crowd management strategies. By proactively addressing these surges in demand, ferry operations can ensure a smoother and more enjoyable experience for passengers.

Table 3: Peak Day for Each Year Based on Total Sales

Year	Peak Day	Total Sales
2016	2016-08-01	49513
2017	2017-08-20	36225

Table 3: Peak Day for Each Year Based on Total Sales

Year	Peak Day	Total Sales
2018	2018-07-01	49477
2019	2019-08-04	42431
2020	2020-07-25	10844
2021	2021-09-04	26860
2022	2022-07-16	43296
2023	2023-08-19	41235

4 Discussion

4.1 Key Insights and Implications

1. Seasonal and Monthly Trends: The analysis confirms the strong seasonality in ferry ticket sales, with the highest demand occurring in the summer months of July and August. This seasonal spike is closely tied to tourism and recreational activities on the Toronto Islands, particularly during holidays and weekends. The sharp contrast in sales between the summer and winter months highlights the importance of efficient resource allocation, ensuring that operations can scale up during peak seasons and down during the off-season.
2. Daily and Hourly Passenger Flow: Peak ferry traffic is concentrated between late morning and early afternoon, with a significant surge in sales between 10 a.m. and 3 p.m. on weekends. These insights provide actionable recommendations for ferry operators to adjust their schedules to minimize wait times during these high-demand hours. By aligning ferry trips with peak demand, operators can ensure that passengers experience shorter lines and faster service.
3. Day-of-Week Trends: The consistent pattern of increased ferry sales on Saturdays and Sundays suggests that weekends remain the primary travel time for both local visitors and tourists. This insight underscores the need for operational adjustments on weekends, including more frequent ferry trips, extended service hours, and enhanced crowd management.
4. Peak Day Analysis: The identification of peak days, typically aligned with holidays or major summer events, provides valuable information for managing extraordinary surges in passenger demand. By knowing in advance which days are likely to experience the highest traffic, ferry operators can implement targeted strategies to handle the increased load, such as adding extra ferry trips or implementing temporary crowd control measures.

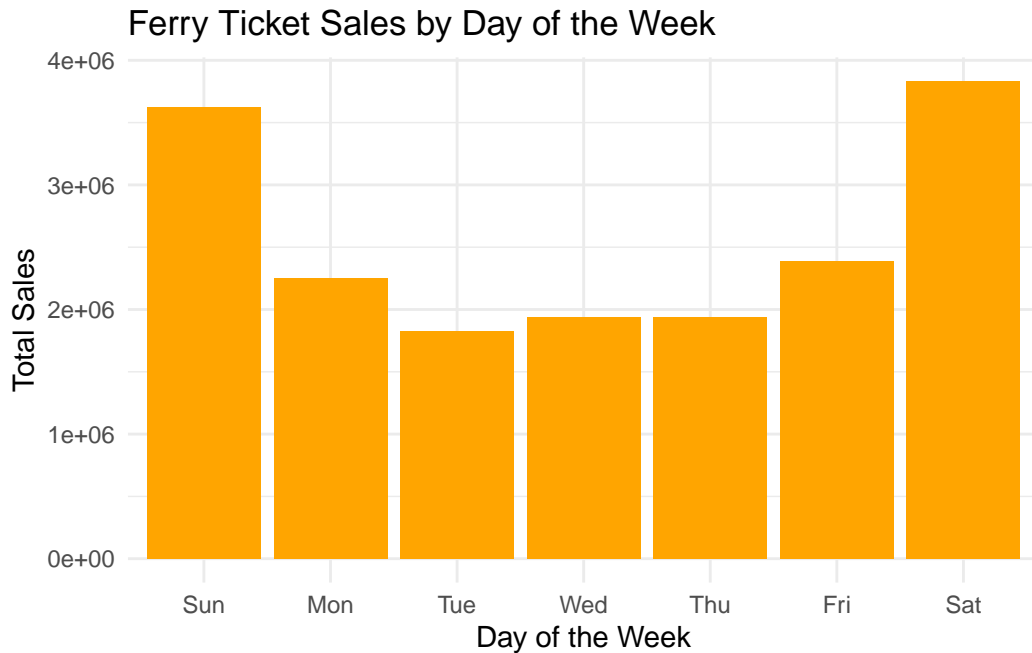


Figure 4: Ferry Ticket Sales by Day of the Week

4.2 Weaknesses and Future Directions

One notable limitation of this analysis is the absence of external factors such as weather conditions, which can significantly influence daily ferry usage. For example, rainy days or extreme heat may deter visitors, while sunny weekends could lead to unexpectedly high demand. Incorporating weather data into future analyses could provide more accurate predictions of ferry traffic patterns.

Additionally, the lack of demographic data prevents a deeper understanding of the types of passengers using the ferry service. By integrating demographic insights, such as age groups, tourist vs. local usage, or family vs. individual travelers, ferry operators could tailor their services to meet the specific needs of different customer segments.

Future research should also explore predictive modeling, using historical data to forecast future ferry demand based on a combination of seasonal patterns, event schedules, and weather forecasts. Such models could help operators anticipate peak days and hours more accurately, allowing for better preparedness and service delivery.

By addressing these gaps and building on the existing analysis, ferry operators can continue to enhance their services, improving the overall experience for visitors to the Toronto Islands.

A Appendix

A.1 Dataset and Graph Sketches

Sketches of the desired dataset structure and the visualizations created for analysis are available in the GitHub repository. These sketches illustrate the different time intervals and patterns analyzed in the report.

A.2 Data Cleaning

The data cleaning process involved the following steps to ensure the accuracy and integrity of the dataset:

Duplicate Removal: Any duplicate entries, particularly those with identical timestamps, were removed to ensure data accuracy. **Missing Values:** Rows with missing values in key fields (timestamp, redemption count, and sales count) were eliminated to maintain the completeness of the dataset. **Validation of Positive Values:** Ensured that both `redemption_count` and `sales_count` contained only non-negative integers. No modifications to the actual data content were made beyond these cleaning steps.

A.3 Data Testing

The following tests were applied to the cleaned dataset to ensure its quality and consistency:

- **Redemption and sales counts:** All `redemption_count` and `sales_count` values were checked to ensure they are non-negative.
- **Total sales validation:** A derived column, `total_sales`, was checked to ensure that all values are non-negative, as it represents the sum of `redemption_count` and `sales_count`.
- **Timestamp validation:** Ensured that all timestamps fall within the expected range (from 2016 to 2023).
- **Duplicate records check:** Verified that there are no duplicate entries for the same timestamp.
- **Missing values check:** Ensured that there are no missing values in the key columns: `timestamp`, `redemption_count`, and `sales_count`. Results from the tests are shown in Table 4 below:

Table 4: Data Testing Results

Test Description	Result
All redemption counts non-negative	Passed
All sales counts non-negative	Passed

Table 4: Data Testing Results

Test Description	Result
All total sales non-negative	Passed
All timestamps within valid range	Passed
No duplicate timestamps	Passed
No missing values in key columns	Passed

A.4 Attribution Statement

“Contains information licensed under the Open Government Licence – Toronto” (City of Toronto, n.d.b).

This appendix clearly outlines the data cleaning and testing process, ensuring the dataset’s reliability for subsequent analysis without mentioning any data modification.

References

- 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/>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Toronto, City of. 2024. *Toronto Island Park*. City of Toronto. <https://www.toronto.ca/explore-enjoy/parks-recreation/places-spaces/beaches-gardens-attractions/toronto-island-park/>.
- 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/>.