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.

1 Introduction

You can and should cross-reference sections and sub-sections. We use R Core Team (2023) and Wickham et al. (2019).

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

---

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

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

X_id	Timestamp	Redemption_Count	Sales_Count
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

Table 2: Sample of Modified Data with Total Sales and Time Features

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 Patterns and Seasonality Analysis

Analyzing daily ferry ticket sales provides insight into visitor flow over the course of a year. As shown in Figure 1, ticket sales for the year 2023 tend to peak during weekends, highlighting the demand for ferry services during leisure periods. Additionally, Figure 2 illustrates the monthly sales trend, where ferry ticket sales are highest during the summer months (July and August), which corresponds to peak tourist season. The seasonal variation in ferry usage is clearly visible in the monthly trends, with significant spikes in demand during the summer months. This seasonality reflects both local and tourist demand for access to the Toronto Islands during warmer weather. Additionally, holidays, including long weekends, show increased ferry traffic.

3.2 Passenger Flow: Peak Time Identification

The analysis of peak times reveals the busiest hours for ferry ticket sales, which are typically concentrated in the late morning to early afternoon Figure 3. These insights can assist ferry operators in planning for peak loads and adjusting schedules accordingly to reduce wait times and improve the overall passenger experience.
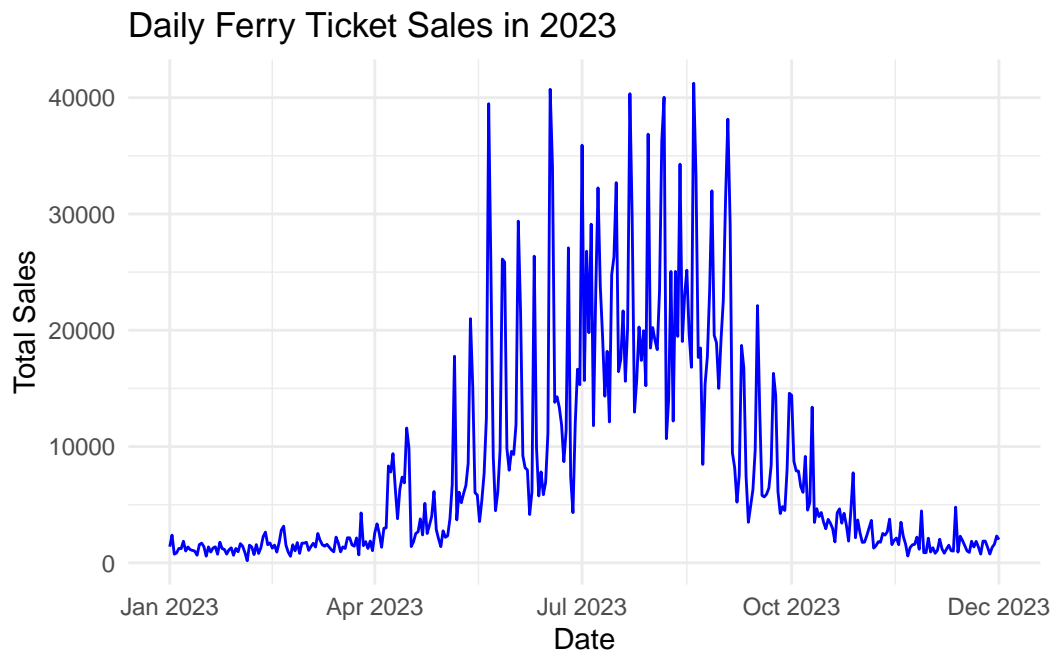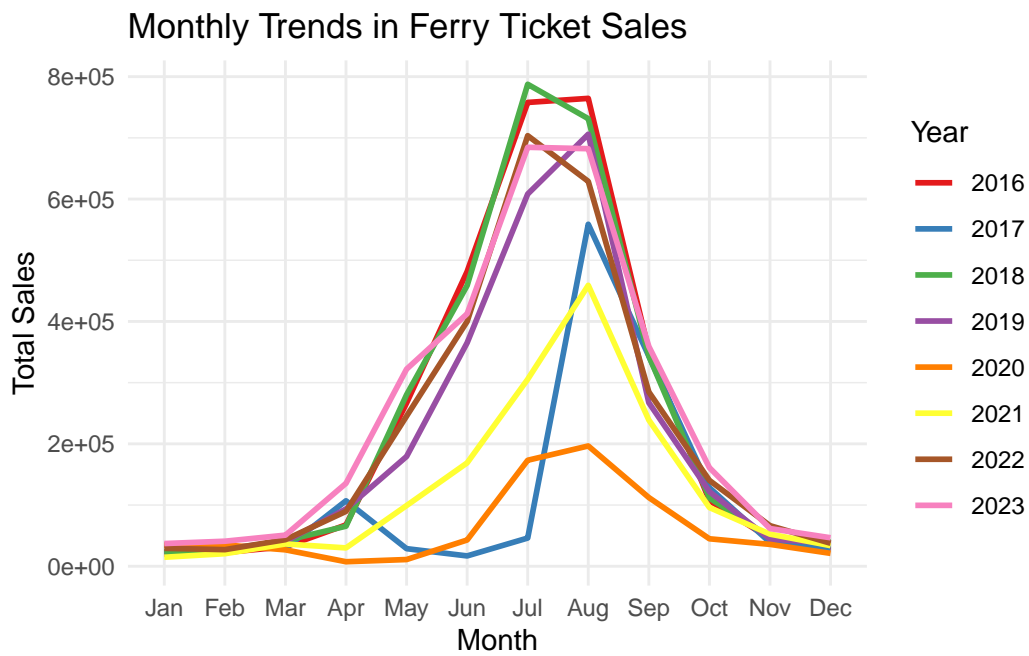
Figure 1: Daily Ferry Ticket Sales in 2023
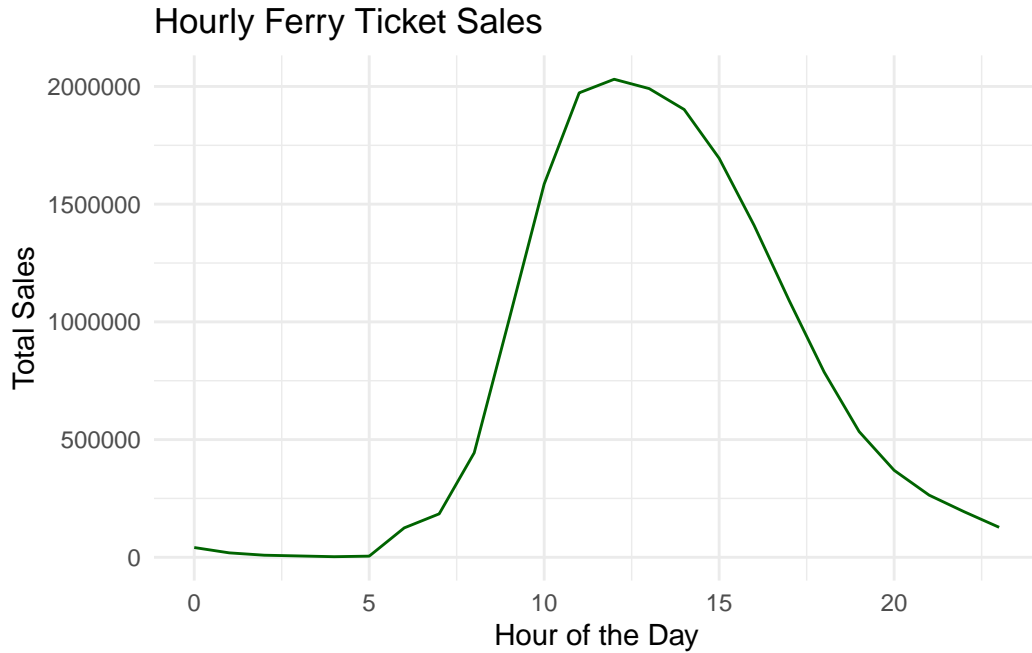

Figure 2: Monthly Trends in Ferry Ticket Sales

Figure 3: Hourly Ferry Ticket Sales

3.3 Day-of-Week Analysis

Figure 4 shows that weekends consistently see higher ferry ticket sales than weekdays, with peak demand on Saturdays and Sundays. This pattern is consistent across all years in the dataset, reinforcing the importance of scheduling and staffing adjustments on weekends to meet the heightened demand.

3.4 Peak Day Analysis

This analysis highlights the specific days with the highest ferry ticket sales for each year in the dataset. Table 3 presents the peak days for each year based on total sales. These peaks typically align with holidays, long weekends, or major summer events, further emphasizing the need for operational planning during these periods.

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

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

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 First discussion point

1. Seasonal Peaks: The analysis confirms that ferry ticket sales are highest during the summer months, as expected. This seasonal demand corresponds to increased tourism and recreational visits to the Toronto Islands.
2. Peak Times: The busiest hours of the day occur between late morning and early afternoon, with a secondary peak in the late afternoon. These insights can guide ferry scheduling to minimize congestion and improve service during high-demand hours.
3. Day-of-Week Trends: Weekends see the highest ferry traffic, which calls for operational adjustments during these periods to handle increased passenger loads.
4. Peak Day Analysis: The peak day analysis reveals that the busiest ferry days are typically aligned with holidays, long weekends, or special summer events. This provides clear guidance for ferry operations to plan for surge capacity during these specific times.

4.2 Weaknesses and next steps

Weaknesses and next steps should also be included. One limitation of the analysis is the absence of detailed weather data, which could potentially influence daily ferry usage. Future research could integrate weather patterns to further refine the analysis. Additionally, incorporating demographic data of ferry passengers could offer more insights into passenger behavior and preferences.

Another potential next step would be to explore predictive modeling to forecast future ferry demand based on historical data, combined with external factors such as event schedules, weather forecasts, and holidays. Such predictive models could allow the ferry operators to better anticipate high-traffic days and ensure adequate resources and staffing are available.
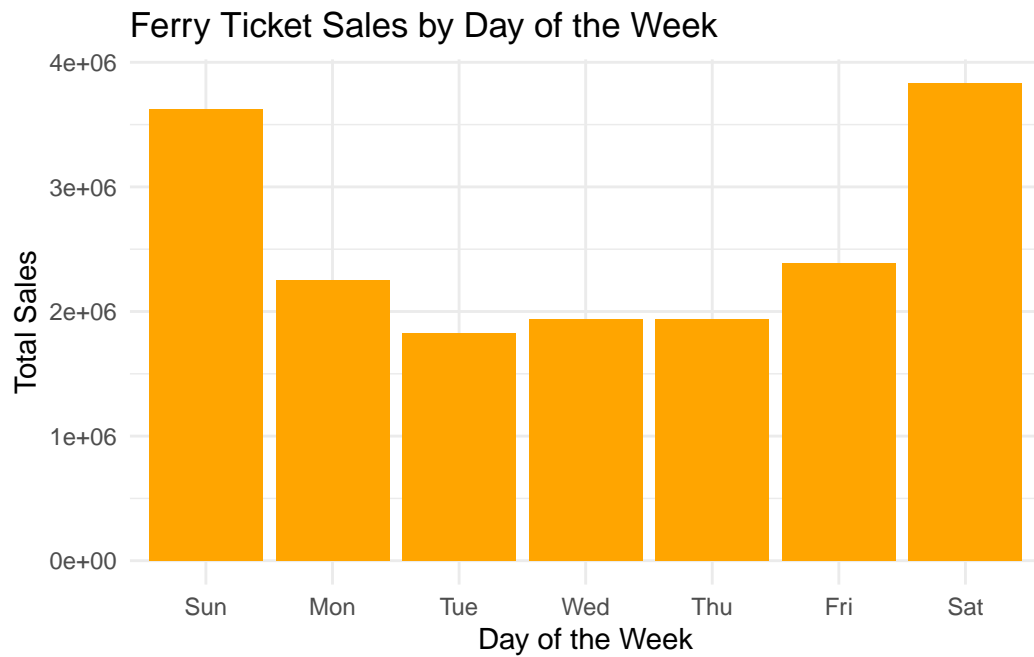
Figure 4: Ferry Ticket Sales by Day of the Week

Appendix

A Additional data details

B Model details

References

R Core Team. 2023. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.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.