

Research on Toronto Island Ferry Ticket Counts*

Ticketing data in relation to daily time periods

Mingjin Zhan

September 24, 2024

This paper aims to study the correlation between Toronto Island Ferry Ticket Counts and time, and analyzes the difference in sales volume in different time periods, holidays or working days from sale counts and redemption count. We found that the number of people visiting the island in summer and autumn was much higher than in winter and spring, and weekends and holidays were also higher than weekdays. Finally, we can understand the island time habits of people living in Toronto or visiting Toronto through this study that is People prefer to go to the island at noon during the warmer season. At the same time, suggestions can be made for optimizing ferry schedules

Table of contents

| | | |
|----------|--|----------|
| 1 | Introduction | 2 |
| 2 | Data | 2 |
| 2.1 | Overview | 3 |
| 2.2 | count on holidays and weekends compared with count on weekdays | 4 |
| 2.3 | Every hour of the day | 6 |
| 3 | Discussion | 7 |
| 4 | Conclusion | 8 |
| | References | 9 |

*Code and data are available at: <https://github.com/Re9shepp/Research-on-Toronto-Island-Ferry-Ticket-Counts.git>

1 Introduction

Toronto Island has always been a popular destination for tourists and local residents, offering a wealth of entertainment and leisure options. The main way to get to the island is by ferry. Although many people use ferry services every day, little research has been done on the relationship between ferry ticket demand and different times of the year and different points of the day. Gaining insight into these trends will not only help optimize ferry schedules, but also enhance the overall travel experience for visitors.

This study focuses on the analysis of ticket sales and redemption for the Toronto Island Ferry, with the aim of revealing how these data fluctuate across seasons, time periods, and between weekdays and weekends/holidays, and drawing conclusions from them to provide actionable optimization recommendations for ferry operations.

The study analysed ferry ticket sales and redemption data between August 2023 and August 2024, looking at how the data changed with the season, weekdays versus weekends, and different times of the day. The results show that ferry demand is highest during the warmer months of summer and autumn, especially on weekends and holidays, with the highest concentration of traffic during the midday hours. In contrast, demand decreased significantly in winter and spring, and there was little difference between weekdays and weekends. Based on these observations, this study further explores people’s travel preferences and makes recommendations for optimizing ferry frequencies.

The structure of this paper is as follows: Section 2 describes the data and methods used for the analysis, Section 3 discusses the results and their implications, and finally Section 4 gives a brief description of the meaning of the article.

2 Data

This article is based on OpenDataToronto(Gelfand 2022) -Toronto Island Ferry Ticket Counts. The original data has four columns of data, `_id`, `Timestamp`, `Redemption Count` and `Sales count`. `Id` represent row identifiers, which are not used in the research of this paper. `Timestamp` indicates the time distribution of the original data, with a line of 15 minutes. `Redemption Count` Indicates the number of redemptions every 15 minutes. The `Sales Count` represents the number of sales per 15 minutes. The study will focus on data from August 1, 2023 to August 1, 2024, and will combine the original 15-minute row of data into one hour for the study. Unless otherwise specified below, the unit of time is hour, and the unit of count is 1. The data was cleaned and analyzed using the R programming language (R Core Team 2024). Use the “tidyverse” package (Wickham et al. 2019) for cleaning. Subsequently, the software package “dplyr” (Wickham et al. 2022) is used for analysis, the software package “ggplot2” (Wickham 2016) is used for visualization, the software package “patchwork” (Pedersen 2020) used for

integrating the figure and the software package “lubridata” (Grolemund and Wickham 2023) is used for conversion to time format.

2.1 Overview

In the overview, the data used in this article are all time-frame data only, consistent with the study focus from August 1, 2023 to August 1, 2024. The relationship between redemption count and sales count and time during this period will be analyzed, and the general impression of redemption count and data will be obtained for subsequent analysis.

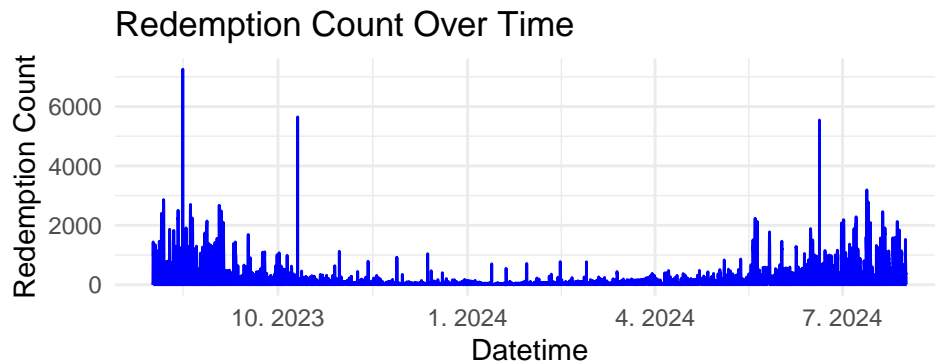


Figure 1: redemption counts over time, from August 2023 to August 2024

This Figure 1 describes the relationship between Redemption counts and time. We can find that this data is stable and high from August to October, 2016, and enters a low period after November. After May of the following year (2024), this data will start to recover to a relatively stable and high value.

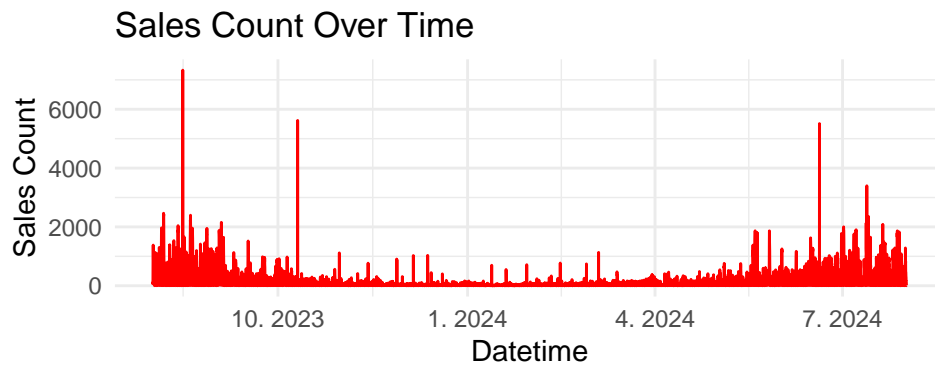


Figure 2: sales counts over time, from August 2023 to August 2024

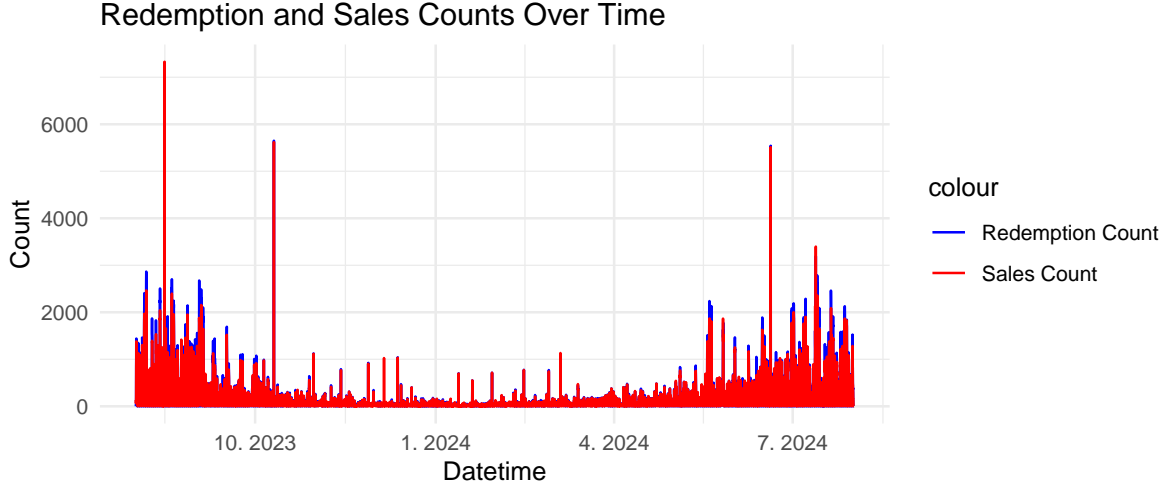


Figure 3: A comparison of Redemption Counts and Sales Counts against time, from August 2023 to August 2024

Figure 2 describes the relationship between sales counts and time. If we compare sales counts(blue) and redemption counts(red), as Figure 3, we can find a similar conclusion, that is, sales are higher in summer and autumn, but lower in winter and spring. Given the large time span and low precision of the data studied in these three figures, only one conjecture can be obtained: many tourists visit the island in summer and autumn, while most necessary people, such as work, visit the island in winter and spring.

Next, we will further refine the data and first verify our conclusion that there are more redemption counts and sales counts in summer and autumn, while fewer in winter and spring. Further analysis of the number of weekdays and holidays on the island to verify the conjecture

2.2 count on holidays and weekends compared with count on weekdays

In the following, the data will be divided into two groups, one group includes holidays and weekends in Toronto, and the other group includes working days. By comparing the differences between the two groups of data, we will verify our conjecture that most people who visit the island in winter will have something to do with their work.

In Figure 4, blue represents the relationship between redemption counts and time on working days, and red represents the relationship between redemption counts and time on weekends and holidays. According to previous findings, redemption counts and sales counts have similar distribution with time, so only redemption counts will be studied. As can be seen from the figure, excluding outliers, most of the redemption counts on weekends and holidays are higher than usual. Since the data is too rough, the data will be refined for further analysis.

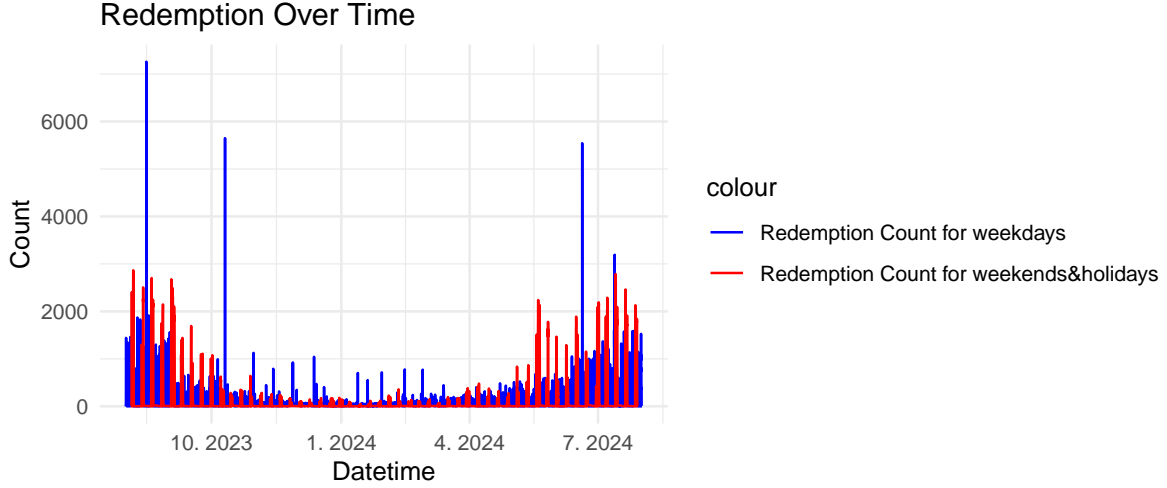


Figure 4: Compare the difference in redemption counts between weekdays and holidays, from August 2023 to August 2024

Table 1: The average redemption counts and sales counts for weekdays, from August 2023 to August 2024, divided into summer autumn and winter spring two groups

| season | average redemption | |
|--|--------------------|----------------------|
| | counts | average sales counts |
| Weekdays in Summer and Autumn | 237.73338 | 248.48688 |
| Weekdays in Winter and Spring | 50.61447 | 44.80999 |
| Weekends and Holidays in Summer and Autumn | 469.47743 | 447.98416 |
| Weekends and Holidays in Winter and Spring | 58.94128 | 54.96658 |

The above data, Table 1, is the data obtained after the data is divided into working days and rest days and then refined again. The data is divided into winter and spring and summer and fall, and this is because we want to test our hypothesis and remove the outliers, those with a count greater than 3000 per hour. The data in Table1 respectively represent the count of average working hours in summer and autumn, the count of average working hours in winter and spring, the count of average working hours in summer and autumn holidays and weekends, and the count of average working hours in winter and spring holidays and weekends. By comparing these four data, we find that there is a big difference between working days and non-working days in summer and autumn, namely, redemption count is about 238 on working days and 470 on non-working days, and sales count is about 249 on working days and 448 on non-working days. In winter and spring, there is little difference between working days

and non-working days, namely, redemption count is about 50 on non-working days and 59 on non-working days, and sales count is about 45 on working days and 55 on non-working days, respectively. To support our suspicion, most of the people who visit the island in winter and spring are people with essential business.

2.3 Every hour of the day

From the above, we can see that people go to the island differently in summer, autumn and winter and spring. The following will continue to discuss this data, and focus on different times of the day to explore people's living habits. In the following, the average redemption count and sales count of each time period throughout the year will be analyzed first, and then the two sets of data in summer, autumn and winter and spring will be compared and analyzed.

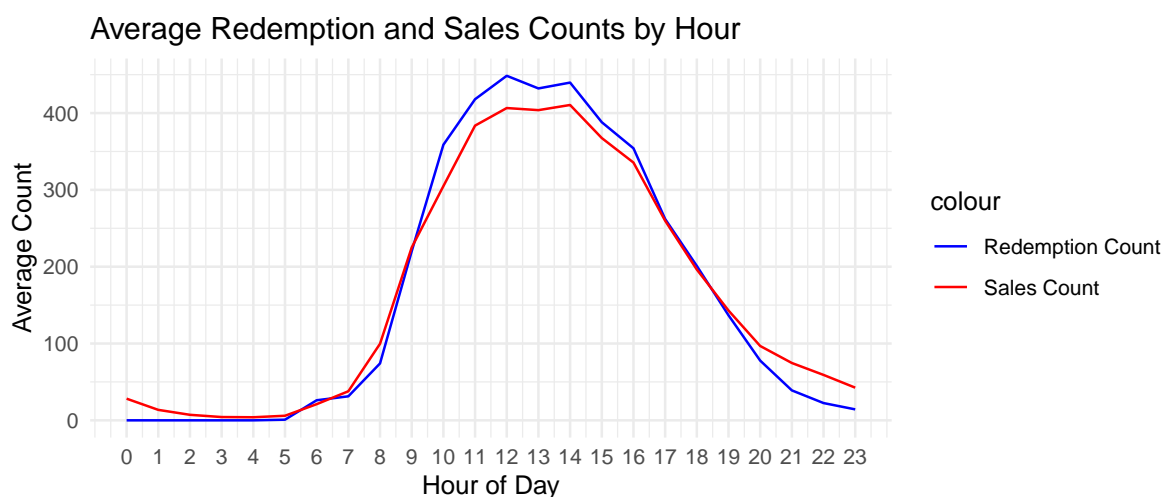


Figure 5: Annual hourly average of redemption count and sales count

From Figure 5, we can see that there is almost no count data after 8 am and 8 PM, and it peaks between 12 noon and 15 PM. A general conclusion can be drawn that most people will choose to go to the island between noon and afternoon, rather than morning or evening. Next, the data will be divided into winter&spring and summer&autumn with weekdays or weekends, and analyzed again.

Analysis Figure 6, for summer and autumn working days, the peak is reached at 10-12 o'clock, and the downturn continues from evening to evening. For weekends and holidays in the summer and fall, the peak lasts from 10 to 15, and the peak is even higher. For winter and spring working days, a lower peak is reached at 16pm, and the range of growth and decline is larger. For weekends and holidays in winter and spring, the peak value is also reached at 16:00, which is higher than that of working days in winter and spring, and the change range is equally drastic. Looking at the annual hourly averages from these four sets of data, only

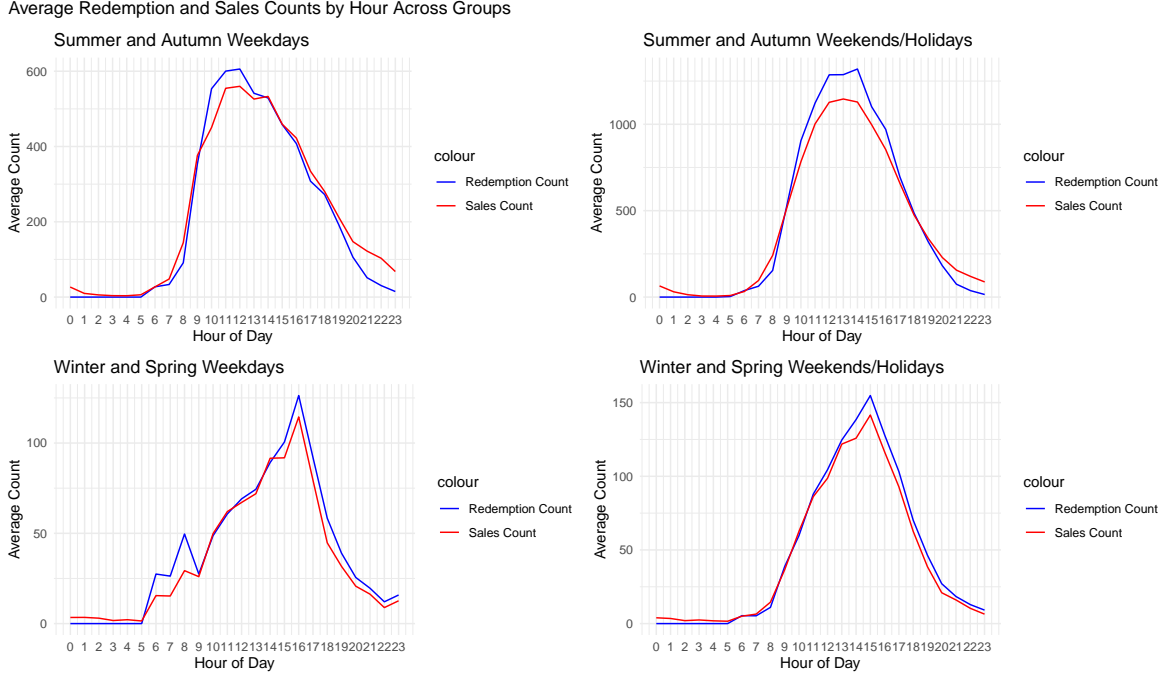


Figure 6: Average Redemption and Sales Counts by Hour Across Groups

one similar conclusion can be drawn: summer and fall are busier than winter and spring, and weekends and holidays are more in demand than weekdays.

3 Discussion

In Section 2, we first observed the distribution of data in general, and found that there were more data in summer and autumn than in winter and spring. Then we divided the data into working days and weekend holidays for comparison, and found that there were more weekend holidays than working days. Then we further refined the data and divided the data into four groups, namely summer and autumn working days, summer and autumn weekend holidays, winter and spring working days and winter and spring weekend holidays. Through comparison and analysis of the four groups of data, it is assumed that the proportion of tourists in summer and autumn is relatively high, and most of the people who visit the island in winter and spring have necessary reasons. Then, looking at the data from a different Angle and analyzing the annual average of each hour of the data, we found that the peaks were concentrated in the late morning and early afternoon. Finally, the average hourly data was analyzed again using the four previously sorted data sets. A conclusion similar to Section 2.2 is obtained. Through the analysis, we can conclude that most of the tourists who choose to visit the island will choose to visit the island when the weather is warmer in summer or autumn, and will choose to go

to the island near the end of the morning. In winter and spring, because the climate is colder, the time of people's activities on the island is shifted from noon to afternoon, which can also be shown that the change of season will affect people's activity. For Toronto Island crews, more ferries can be scheduled in the summer and fall, and a modest reduction in the winter and spring to balance the budget and traffic pressure. In the specific situation, more shifts can be arranged in the morning in summer and autumn to reduce the queuing pressure of the flow of people, and the shifts can be slowly reduced after the afternoon to balance the budget. In winter and spring, the frequency can be maintained at a relatively low level, and the schedule can be released in advance to avoid too long waiting for people on the island. No matter which season, the ferry frequency can be appropriately increased on better weather days to avoid unnecessary waiting for people on the island.

4 Conclusion

Through the research on redemption count and sales count of Toronto Island ferry in this paper, relevant personnel can understand tourists' travel preferences and help better arrange ferry schedules. For example, tourists like to go to the island in warm times, so increase the ferry frequency in better weather in advance. So that people can avoid unnecessary waiting, better enjoy the time on the island.

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. 2023. *Lubridate: Make Dealing with Dates a Little Easier*. <https://CRAN.R-project.org/package=lubridate>.
- Pedersen, Thomas Lin. 2020. *Patchwork: The Composer of Plots*. <https://CRAN.R-project.org/package=patchwork>.
- 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, and Kirill Müller. 2022. *Dplyr: A Grammar of Data Manipulation*. <https://CRAN.R-project.org/package=dplyr>.