

Toronto Island Traffic Analysis*

-Bo Tang

September 27, 2024

This post analyzes data for Toronto Island Ferry Ticket Counts from July 2023 to September 2024 provided by open data Toronto. The counts were found to have a very distinct time factor distribution. Suggests that ferry schedules should be adjusted relative to the time factor.

Table of contents

1	Introduction	1
2	Data	2
2.1	Raw Data	2
2.2	basic data information	2
2.3	Monthly data distribution	3
2.4	weekly data distribution	4
2.5	hourly data distribution	4
2.6	Results	5
3	Discussion	5
3.1	Key Findings	5
3.2	Limitations and Future Work	6
	References	7

1 Introduction

The Toronto Island, Often referred to as Toronto Island, the Toronto Islands are a picturesque archipelago near Toronto’s downtown waterfront. You can enjoy the beautiful surroundings

*Code and data are available at: [Toronto Island Traffic Analysis](#)

and get away from the noise of the city. It is a place that can help cleanse the soul, and for this reason a large number of tourists as well as locals come to this place every year. There are three main ways to get to Toronto Islands, you can take a private water taxi, a tender and the main and most popular way to get to Toronto Islands is by ferry. This paper focuses on ticket sales and redemption data after 2023. due to how much the Toronto Islands are loved. This has led to a particularly high volume of traffic, and since ferries are the main way to get to the Toronto Islands, this has led to a relatively large amount of revenue for the ferries. However, this has also led to long waits for the ferry and traffic jams around the ferry. The article wants to analyze the data from September 2023 to September 2024 and the timing of the ferry. This will help people to have a more enjoyable trip to Toronto Island, as well as alleviate traffic congestion.

2 Data

2.1 Raw Data

The data for this analysis comes from Toronto's Open Data portal (Gelfand (2022)). The article utilizes rcode and the following major packages under processing data, R code (R Core Team (2023)), Tidyverse (Wickham et al. (2019)), Dplyr (Wickham et al. (2023)), Lubridate (Grolemund and Wickham (2011)), and Knitr (Xie (2024)) and ggplot2 (Wickham (2016)).

Toronto Island Ferry Ticket Counts: The data we are allowed to download, this packet records every 15 minutes ticket counts from July 2023 to September 23, 2024, which contains two kinds of tickets namely Redemption and Sales. A total of 32,000 pieces of data are included.

2.2 basic data information

A bit of basic information about the year for both tickets, including mean, median, variance and standard deviation. The first thing we can tell is that REDEMPTION and SALES COUNTS are similar, not much different, reaching about 65 votes per 15 minutes. It can also be noticed that the data is spread out a lot if we just look at the dimension of one year

Table 1: Basic Statistics of Redemption and Sales Counts

Redemption Mean	Redemption Median	Redemption Variance	Redemption SD	Sales Mean	Sales Median	Sales Variance	Sales SD
64.18319	17	17616.75	132.7281	63.65659	22	15234.06	123.4264

Total data analysis

2.3 Monthly data distribution

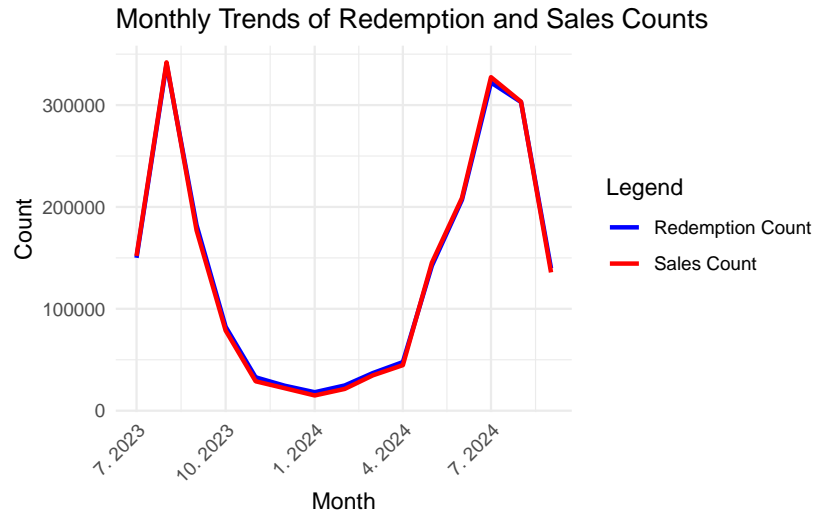


Figure 1: monthly data analysis

The data is then categorized by month and the Redemption and Sales counts are added up and it is found that in the months of July to September 2023 and May to September 2024 their counts are more than 100,000 whereas in the other months they are not more than 83,000. This is one of the reasons for the large standard deviation of the data. Then we compare the data for each month and make a line graph, we can observe the image and conclude that the amount of redemption and sales is about the same and the same as the conclusion drawn by the basic data. In this graph we can also see that Redemption and Sales counts are concentrated in the months of May to September in a year, with the highest counts in July and August.

2.4 weekly data distribution

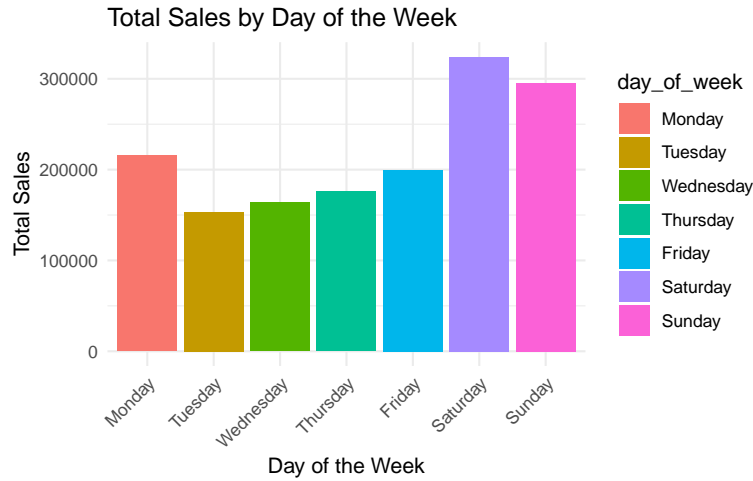


Figure 2: weekly data analysis

The data was next broken down even more, integrating the data for each of the Monday through weekday days of the year. We find that Saturdays and Sundays are very high, which makes more sense because people don't have a lot of time to hang out on weekdays. Meanwhile, we can see from Figure 2 that the sales count on Monday is more than the other four weekdays

2.5 hourly data distribution

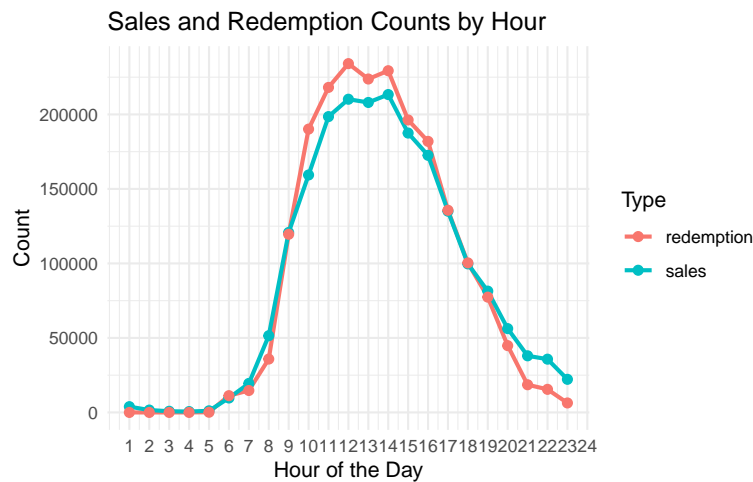


Figure 3: hour data analysis

Breaking the data down to each hour and making a trend graph reveals that the times people do the ferry are clustered between 9am and 18pm

2.6 Results

After understanding this year's data, the first thing that can be clearly found is that Redemption and Sales Counts are concentrated in the months of May to September, and the temperature in Toronto from May to September is above 10 degrees Celsius, which means that Redemption and Sales Counts are mainly concentrated in the summer months, and the summer vacation cycle around the world is basically the same. This also allows more people to travel than would normally be the case. I think these two factors are the main reasons that make counts concentrated in these months. Again, based on WEEKLY DATA it can be found that many more people ride the ferry on weekends than on weekdays, and on weekdays there are many more people riding on Mondays compared to other weekdays. A comparison of Redemption and Sales Counts between 24 hours shows that foot traffic is highest between 9am and 6pm. Now we know a few time factors for high traffic.

3 Discussion

3.1 Key Findings

After summarizing the data and reviewing the ferry frequency across different seasons, an area for improvement becomes clear—the summer, when there is a significantly higher number of ferry riders. Despite this increased demand, there are only 22 ferry trips per day, which is nearly the same as the 21 trips provided in the fall and winter (City of Toronto (2024)). Additionally, the frequency of ferry trips remains fixed throughout the week, regardless of demand.

To improve efficiency, the frequency of ferry trips during the summer should be increased. Specifically, there should be an increase in the number of summer trips, while the number of trips during the fall and winter could be reduced. Increasing the ferry frequency during the summer, especially on weekends and during peak hours from 9 AM to 6 PM, would reduce passenger wait times and alleviate the stress caused by heavy foot traffic. This would, in turn, help ease congestion.

On the other hand, during off-peak hours and seasons with lower demand, the ferry frequency can be reduced. While this might cause some inconvenience for passengers who travel during these less busy periods, the overall system would become more efficient by reallocating resources when they are most needed.

3.2 Limitations and Future Work

Since the only data available for use in this article is for the interval July 2023 to September 23, 2024, there is no way to fully guarantee the generalizability of this data, whether it would be equally applicable in other years, and some of the data is not very accurate, as you can see in my hourly data that there is a COUNT before 6:00 a.m. but the earliest time that the ferries have departed all year long is 6:30 a.m. It could be an error caused by entering the data.

References

- City of Toronto. 2024. “All Ferry Schedules - Toronto Island Park.” <https://www.toronto.ca/explore-enjoy/parks-recreation/places-spaces/beaches-gardens-attractions/toronto-island-park/all-ferry-schedules/>.
- 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/>.
- 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>.
- Xie, Yihui. 2024. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. <https://yihui.org/knitr/>.