Patterns of Toronto Islands Ferries Tickets Sales and Redemption and potential Solutions*

Yun Chu

September 26, 2024

The paper firstly discussed the background of the dataset, then presented and analyzed the cleaned dataset. In the result section, the relationship between the sales and redemption data, the distribution of redemption count, and the trend of redemption count with respect to time is analyzed. The discussion section comes up with some suggestions to Toronto Islands ferries based on the findings and discuss the limilation of the analysis.

Table of contents

*code and data supporting this analaysis can be found here: https://github.com/chuyun2024/Analyzing-Toronto-Islands-Ferry-Tickets-Sales-and-Redemption-Count-Data

1 Introduction

Ferries to Toronto Island Park operate year-round, transporting passengers between the Jack Layton Ferry Terminal and Center Island, Hanlan's Point, and Ward's Island, or in the reverse direction from the islands to the terminal. Currently, the ferries operate on four different schedules, each corresponding to one of the four seasons of the year (City of Toronto, n.d.).

These islands are popular destinations for both tourists and residents of Toronto. By analyzing the sales and redemption count data of the ferry service, we can gain insights into visitor trends, helping to optimize resource allocation and revenue management.Identifying peak ferry usage seasons enables a better understanding of when the islands are most popular. This information can be instrumental for governments in resource allocation and operational planning, such as adjusting ferry schedules and staffing levels throughout the year. Additionally, analyzing sales data over time helps uncover revenue trends, providing valuable insights into the financial health of the ferry service and informing pricing strategies.

This paper reveals a strong correlation between sales and redemption data, suggesting a likely linear relationship. It also finds that during the summer months, redemptions are significantly higher compared to winter months. Moreover, 95% of daily redemptions falls below 16,076 Figure 3.

The remainder of this paper is structured as follows. Section 2 discusses the features of the raw and cleaned data, along with a summary of the cleaned data. Section 3 investigates the relationship between sales and redemption data using graphs, linear regression, and correlation analysis. It also explores the distribution of redemption counts by frequency and counts. Section 4 further elaborates on the implications and applications of the findings from Section 3, addresses the limitations of the analysis, and suggests potential improvements.

2 Data

2.1 Raw Data & Measurement

The data used in this paper is sourced from the Toronto Open Data portal and was imported using the opendatatoronto library (Gelfand, 2022). The raw dataset contains sales and redemption count data, recorded at 15-minute intervals, collected from both online and POS-kiosk sales. The dataset contains four variables: id, timestamp, sales count and redemption count. The id variable is a sequential number that records the order of the data entries. The dataset

spans from July 11, 2023, to September 20, 2024, at the time of download (City of Toronto 2024b). Similar datasets were not found online.

The data was downloaded, cleaned, and analyzed using R (R Core Team 2023). Several R packages were utilized throughout the analysis:

- tidyverse (Wickham 2021)
- lubridate (Grolemund and Wickham 2011)
- dplyr (Wickham et al. 2023)
- knitr (Xie 2015)
- scales (Wickham 2018)

2.2 Cleaned Data

A sample of the cleaned data (Table 1) and a line plot (Figure 1) illustrating the sales and redemption counts over time are shown below.

Table 1: Sample of Cleaned Sales and Redemption Data

Date	Sales	Redemption
2023-07-11	124	44
2023-07-12	8866	9322
2023-07-13	6651	5465
2023-07-14	12817	11960
2023-07-15	12296	14028
2023-07-16	15684	17008

The sales and redemption data largely overlap, with significantly higher values in the summer compared to the winter months (Figure 1).

2.3 Summary Statistic

To have a better understanding of the data, the minimum, maximum, mean and standard deviation of sales and redemption counts are calculated in Table 2.

The summary statistics for sales and redemption counts, presented in Table Table 2, reveal notable similarities across all metrics.

Given these similarities, it is worthwhile to investigate the relationship between sales and redemption further. This analysis will be explored in Section 3.1.

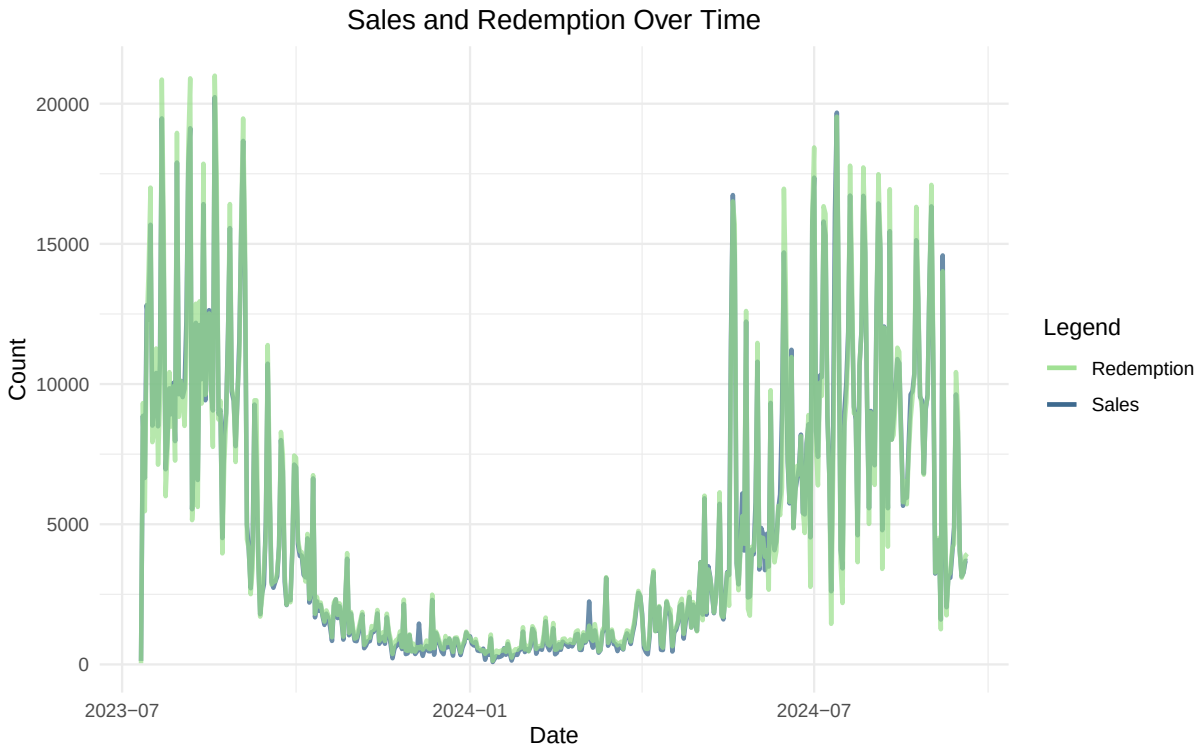

Figure 1: line plot of sales and redemption data with respect to time

Table 2

Table 3: Summary Statistics for Sales and Redemption Count

	Min	Max	Mean	Standard Deviation
Sales	86	20234	4746.505	4815.811
Redemption	44	21001	4785.342	4928.666

3 Result

3.1 Investigating the Relationship between the Sales and Redemption Data

To investigate the relationship between sales and redemption data, we first create a plot (Figure 2) that visually describes the relationship between these variables.

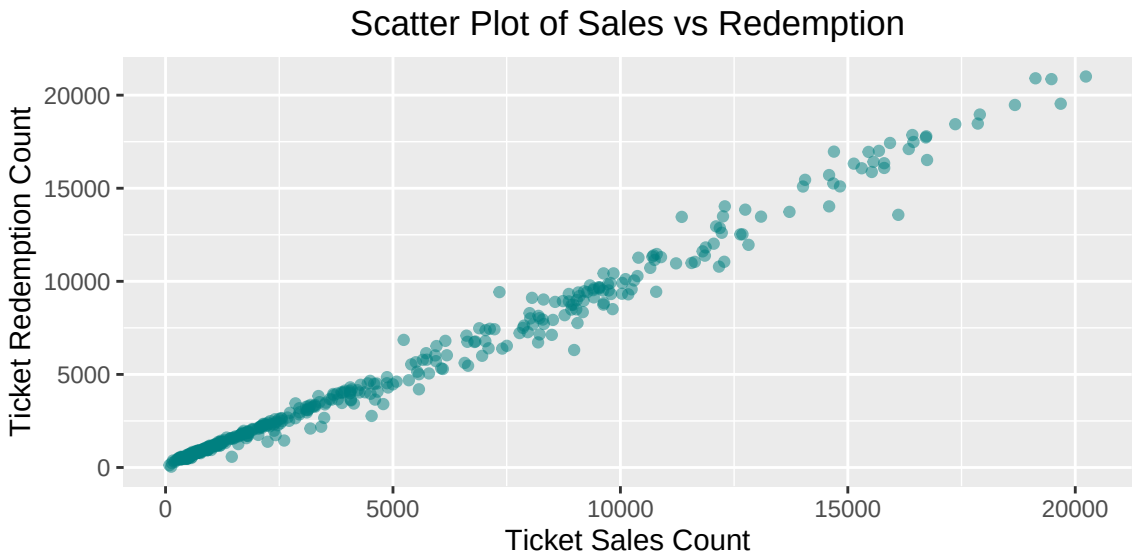

Figure 2: Relationship betwwen Sales and Redemption

A linear relationship can be seen from the scatter plot. To further investigate this relationship, we performed a linear regression with redemption count as the response variable and sales count as the predictor variable. As shown in Table Table 4, the R-squared statistic is 0.9882, and the correlation coefficient between redemption count and sales count is 0.994. These values indicate a strong linear correlation between the two variables.

In this paper, we will primarily focus on analyzing the redemption count data, as it directly reflects visitor data. Utilizing redemption counts to analyze revenue is sufficiently accurate due to the high overlap and correlation between the two variables.

Table 4: R-squared and Correlation Statistic

Statistics	Value
R-Squared	0.9882393
Correlation	0.9941023

3.2 Investigating the Frequency of Redemption Count Data

Figure 3 illustrates that for most days, the redemption count remains below 10,000. However, there are some days where the redemption count peaks at 20,234.

For our resource allocation needs, utilizing monthly data would be sufficient to create an effective staffing schedule. Daily scheduling may not be as feasible or efficient, making a monthly approach more practical for our purposes. The monthly redemption distribution is analysed **?@sec-monred**.

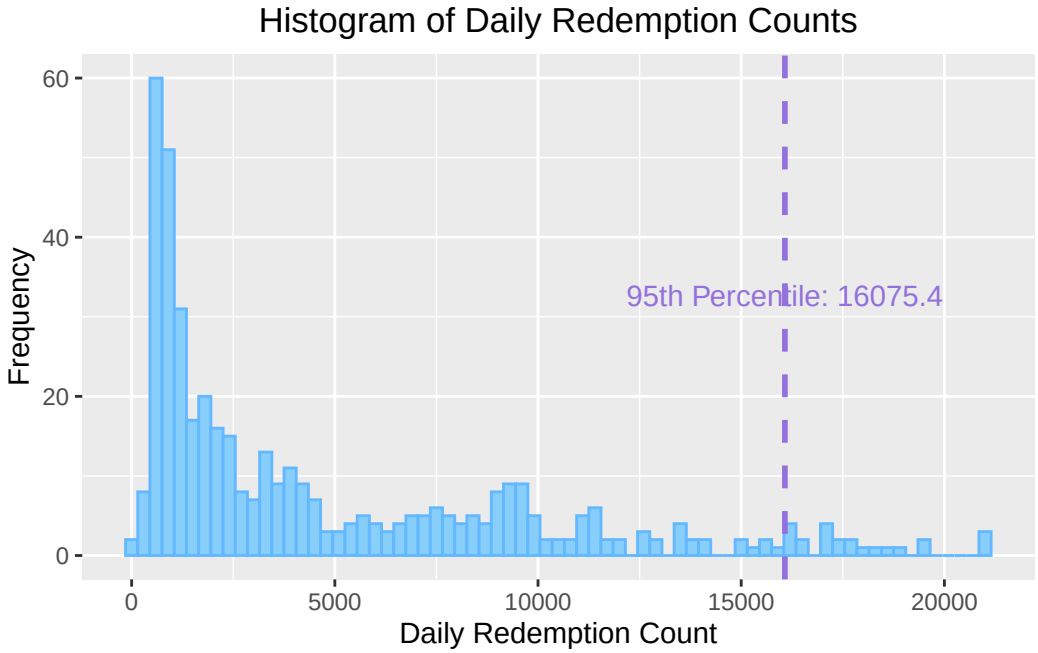

Figure 3: Histogram of Daily Redemption count data

Figure 4 illustrates the redemption counts for each month, with dividing lines in each bar indicating daily redemption counts and their cumulative total. The redemption counts are notably low during the winter months, while summer months exhibit significantly higher counts. Notably, in August 2023, the redemption count approached 350,000.

From July to September 2023 and from May to September 2024, the monthly redemption counts exceed 100,000. In contrast, from October 2023 to April 2024, the redemption counts fall below this threshold each month.

Given these observations, it would be prudent to design two distinct staffing schedules for the ferries: one for peak periods when redemption counts are high and another for off-peak periods when counts are lower. This approach enables more efficient staff allocation based on expected demand.

3.3 Investigating the Monthly Redemption Data{sec-monred}

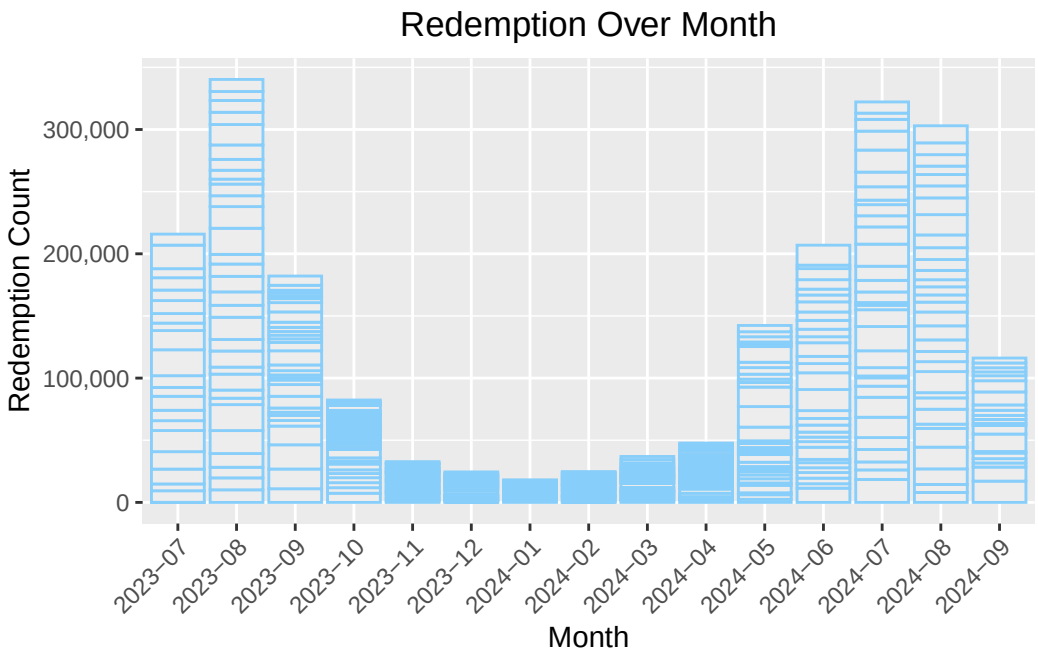

Figure 4: Redemption Count of Each Month

4 Discussion

4.1 Impact of Pricing Strategy on Revenue Optimization

The high correlation and linear relationship between sales and redemption data in Section 3.1 suggest that the flat pricing model for ferry tickets has a significant impact. If the fare were adjusted to be lower during the off-peak season, it could potentially enhance revenue during the winter months. While this may result in some revenue loss during the summer, the overall impact might not be as detrimental due to the typically higher volume of sales during that period.

4.2 Efficient Staffing Solutions for Peak and Off-Peak Seasons

Based on Figure 3, one measure that Toronto Islands ferries could implement is to integrate staff schedules during peak seasons to prepare for large visitor volumes. For instance, having janitors on-call during busy periods would ensure that adequate cleaning services are available when the number of visitors increases significantly.

Additionally, it is essential to develop an on-call schedule that aligns with peak and off-peak seasons. For example, in the winter months, preparing a full staff contingent to manage more than 20,000 visitors per day would be an inefficient use of resources. Instead, a more tailored approach would allow for better resource allocation based on anticipated demand.

4.3 Evaluation of Staffing Schedules in Relation to Visitor Demand

From Figure 4, a noticeable difference in the height of the bars can be observed between May and September compared to the rest of the year. Upon closer examination of the graph, July and August exhibit significantly higher redemption counts compared to other summer months.

Currently, the Toronto Island ferries have adapted four different schedules: Summer (May 16th - September 15th), Fall (September 16th - October 14th), Winter (October 15th - an unspecified date in 2025), and Spring (an unspecified date - an unspecified date in 2025) (City of Toronto 2024a). However, without an analysis of the specific details of these schedules, it remains unclear whether the current allocation of resources is optimal based on the redemption data.

Given the extremely high volume of visitors in July and August, adjusting the staffing schedules during these peak months may be more beneficial. This approach would align more closely with visitor demand and help avoid resource wastage during periods of lower activity.

4.4 Weaknesses and Next Steps

There are a few limitations to address in this analysis.

Firstly, while the examination of the linear relationship between redemption and sales data in Section 3.1 focuses on redemption data for simplicity and as a direct representation of actual visitor volume on the ferries, it is important to note that incorporating sales data into the analysis would yield a more accurate understanding of revenue-related factors. This additional data would provide a clearer and more comprehensive perspective.

Moreover, without a detailed analysis of the current ferry schedules, it is challenging to offer specific recommendations for adjusting these schedules to optimize resource allocation effectively.

References

City of Toronto. 2024a. "Toronto Island Ferry Schedules." https://www.toronto.ca/explore-enjoy/parks-recreation/places-spaces/beaches-gardens-attractions/toronto-island-park/all-ferry-schedules/.

———. 2024b. "Toronto Island Ferry Ticket Counts." https://open.toronto.ca/dataset/toronto-island-ferry-ticket-counts/.

———. n.d. *Ferries to Toronto Island Park*. https://www.toronto.ca/explore-enjoy/parks-recreation/places-spaces/beaches-gardens-attractions/toronto-island-park/ferries-to-toronto-island-park/.

Grolemund, Garrett, and Hadley Wickham. 2011. *Lubridate: Make Dealing with Dates a Little Easier*. https://CRAN.R-project.org/package=lubridate.

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. 2018. *Scales: Scale Functions for Visualization*. https://CRAN.R-project.org/package=scales.

———. 2021. *Tidyverse: Easily Install and Load the 'Tidyverse'*. https://CRAN.R-project.org/package=tidyverse.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Xie, Yihui. 2015. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://CRAN.R-project.org/package=knitr.