1 Introduction

Throughout the year, ferries depart from Jack Layton Ferry Terminal to Toronto Island Park. In order to better understand and grasp the number and status of Toronto Island Ferry tickets, I decided to count the relevant ticket information. Before actually downloading the data, I first simulated the possible forms of the data and their distribution through a deep understanding of the subject. This pre-processing process helped me form a preliminary concept of the overall structure of the data and laid the foundation for subsequent data analysis. Then, I found relevant data about Toronto Island Ferry tickets by visiting the Open Data Toronto platform. This platform provides near-real-time ferry ticket sales and ticket redemption information, and the data is updated every hour at 15-minute intervals. This real-time data provides a valuable basis for me to analyze the sales trend and redemption of ferry tickets. After successfully downloading the data, I entered the stage of data cleaning. Data cleaning is an indispensable step in the data analysis process. In this step, I eliminated irrelevant or duplicate information and organized the data to make it more organized and readable. This process not only simplified the data structure, but also ensured the accuracy of my subsequent analysis. After careful cleaning and organization, the data became more concise and clear, and more suitable for further processing and analysis. To ensure the reliability and accuracy of the data, I conducted comprehensive tests on the cleaned data. Through testing, I confirmed that the integrity and consistency of the data met the expected standards, which provided a good foundation for my next step of data visualization and analysis. Finally, I used the code written in R language to generate a series of charts based on this data. These charts intuitively show the number and status trends of Toronto Island Ferry tickets. Through in-depth analysis of these images, I obtained detailed conclusions about ticket sales and redemption, providing valuable insights for understanding the operation of Toronto Island Ferry.

2 Data

All the data comes from from Gelfand (2022).

```
Rows: 32000 Columns: 3
-- Column specification --------------------------------------------------------
Delimiter: ","
dbl  (2): Redemption Count, Sales Count
dttm (1): Timestamp

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.

Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
i Please use `linewidth` instead.
```
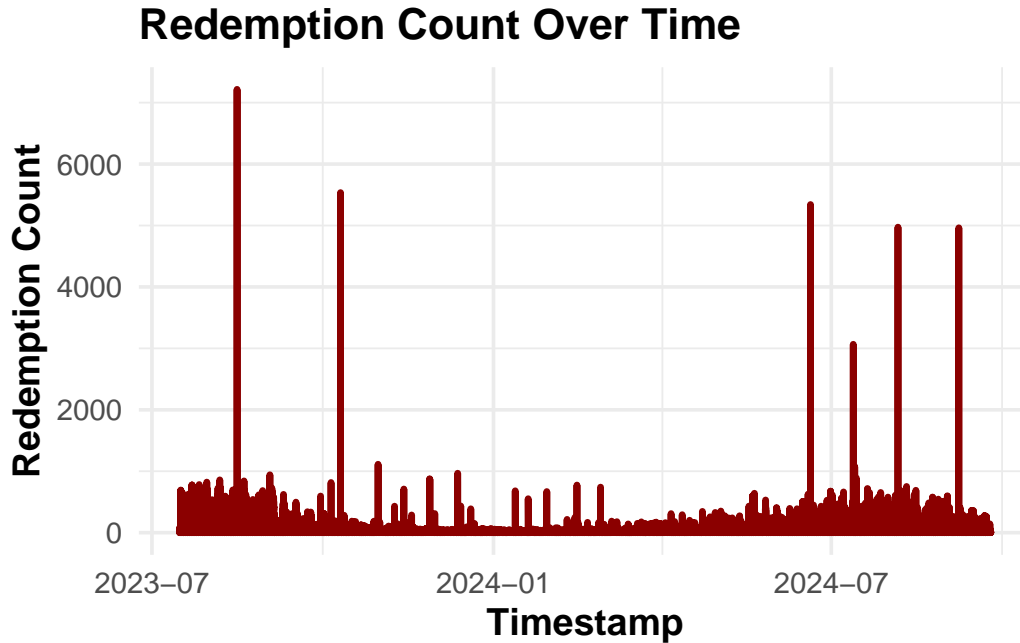
Redemption Count Over Time

Figure 1: Redemption Count Over Time

This graph shows the change in the number of redemptions over time@fig-redemption. Redemption behavior is relatively stable throughout the time period, but there are some significant peaks, especially on certain days, when the number of redemptions rises sharply. This may reflect promotions associated with the sale or certain redemption events

This is the distribution of the number of redemptions@fig-distribution1. The vast majority of redemption activities are concentrated in the low redemption frequency area, and a few events with a large number of redemptions pull up the tail of the distribution. The distribution of the number of redemptions also shows a skewed distribution, similar to the distribution of the number of sales.

This line chart shows how the number of sales changes over time@fig-sales. It can be seen that the number of sales is relatively stable throughout the period, with occasional sales peaks, especially at a few specific time points. This may reflect a short-term sales surge caused by some promotional activities or special events.

This is a bar chart showing the distribution of sales@fig-distribution2. Most sales are concentrated in a small range, and the frequency drops rapidly as the number of sales increases. The vast majority of sales are less than 1,000, and sales over 2,000 are almost non-existent. This shows that the distribution of sales is highly skewed, with a small number of sales accounting for the majority of sales.
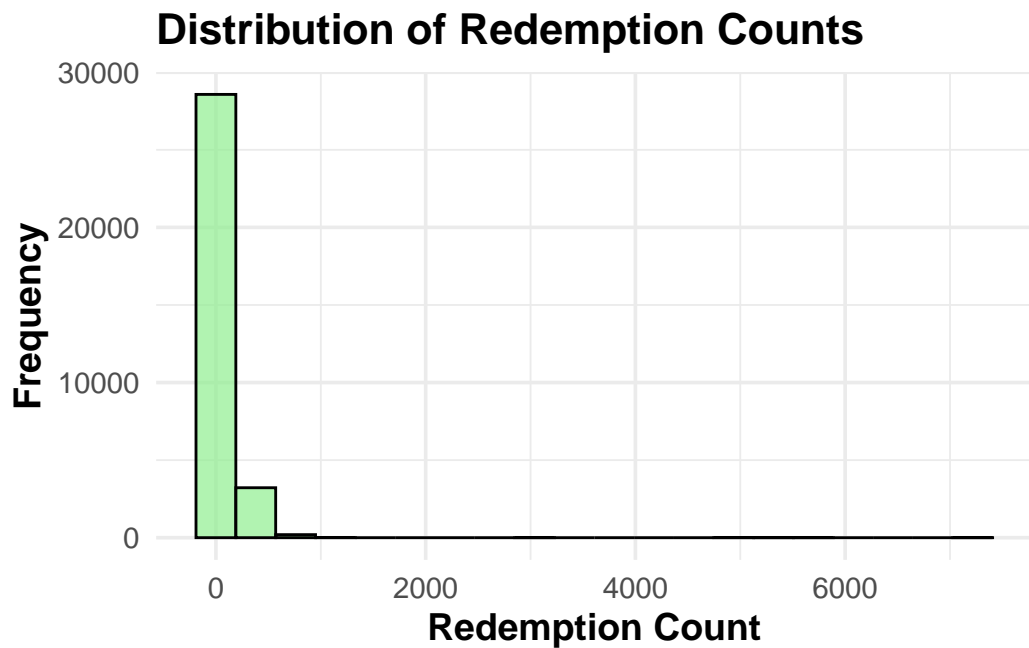
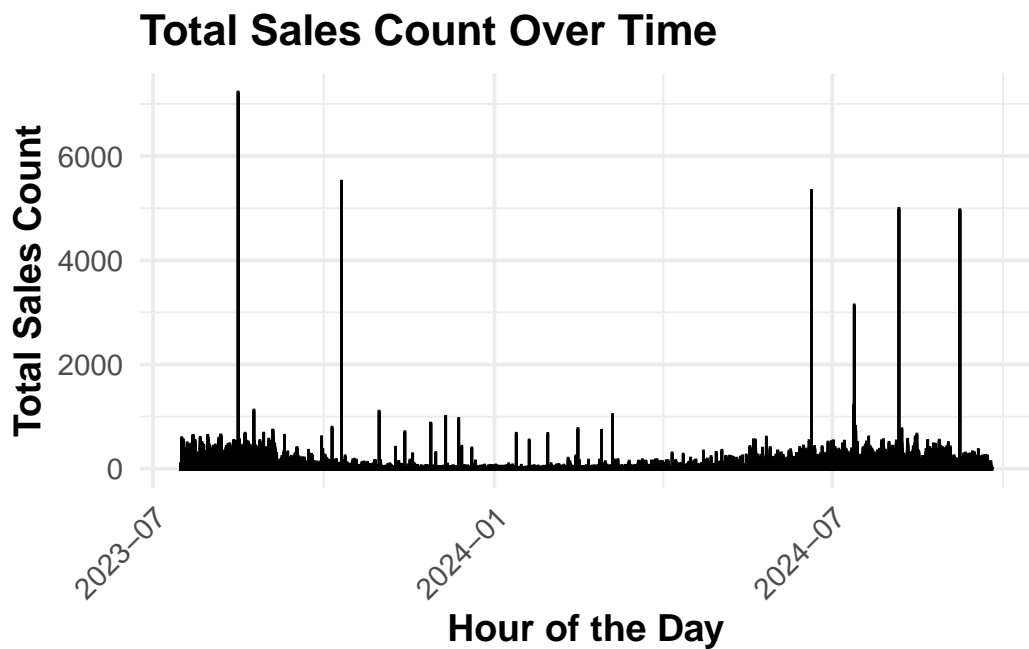Figure 2: Distribution of Redemption Counts

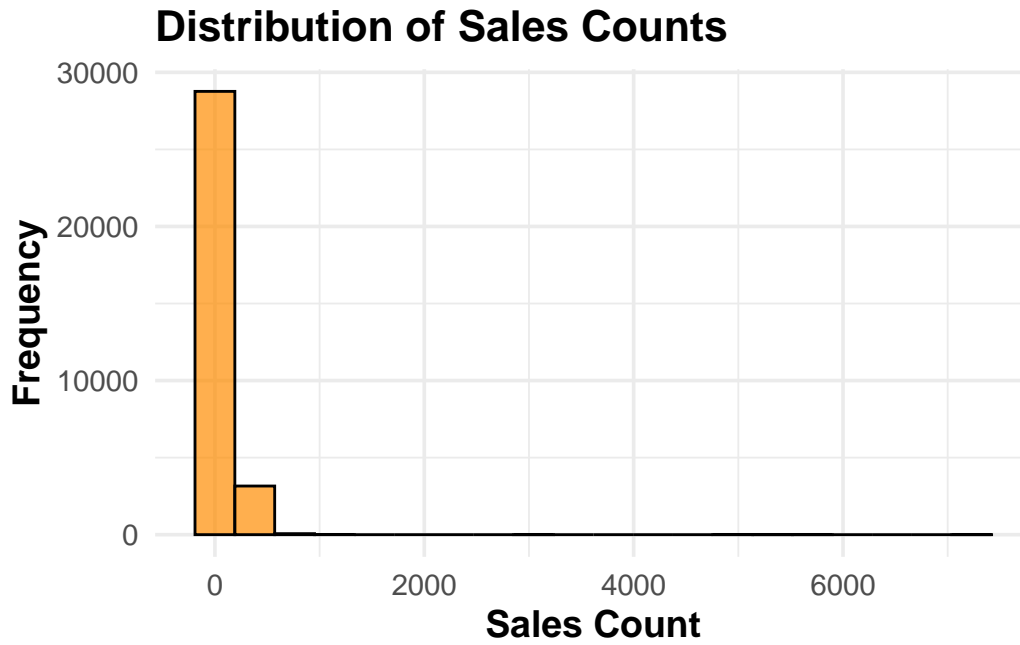Figure 3: Total Sales Count Over Time

Distribution of Sales Counts

Figure 4: Distribution of Sales Counts

3 Conclusion

Overall, these four graphs show the distribution and temporal trends of the number of sales and redemptions. Both sales and redemptions exhibit highly skewed distributions. In addition, sales and redemption behavior can fluctuate dramatically at specific points in time, which may be related to special activities or events.

References

Gelfand, Sharla. 2022. *Opendatatoronto: Access the City of Toronto Open Data Portal.* https://CRAN.R-project.org/package=opendatatoronto.