Analyzing Toronto Islands Ferry Tickets Sales and Redemption Count Data

Yun Chu

September 23, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

Table of contents

1 Introduction

Ferries to Toronto Island Park is a service that operates all year round. Ferries carry passenger from Jack Layton Ferry Terminal to Center Island, Hanlan's Point and Ward's Island or from the three islands to the terminal. They are popular places to visit for tourists and people who live in Toronto. Most people take the ferries to and from Toronto Islands while others take water taxis. By analyzing the sales and redemption count data of the ferry to Toronto Islands, we can better understand visitor trends to help with maximizing resource allocation and analyze revenue data.

Through identifying peak seasons for ferry usage, it would help understand at what times are the islands most popular. This would inform details about resource allocation and operational planning information for governments. For example, the information could be used to design the departure and returning schedule of the ferry and the number of staff on duty in each season. By examining sales data over time, revenue trends can be identified and thus help assess the financial health of the ferry service and guide pricing strategies.

This paper reveals that the sales and redemption data are highly correlated and is highly likely to have a linear relationship. It also finds that in summer months, the number of redemption is much higher compared with winter months. Moreover, the 95% of redemption is below 16,076 per day.

The remainder of this paper is structured as follows. Section 2 discusses about the features of raw data and the cleaned data, as well as some summary of the cleaned data. Section 3 discusses about the details of investigating the relationship between Sales and Redemption Data through graphs, linear regression and correlation. Besides, the section also presents and analyses the distribution of the redemption count data based on frequency and month. Section 4 elaborates further on the implication and application of the analysis in Section 3. It also discusses about the weakness of the analysis and the ways to improve.

2 Data

2.1 Raw Data

The data used in this paper is taken from Toronto Open Data and is read into the paper using the opendatatoronto library (Gelfand 2022). The raw data contains the timestamp data per 15 minutes and the sales and redemption count data for each 15-minute time (Toronto 2024).The it contains data from July 11st, 2023 to September 20, 2024. The data is updated hourly.

The dataset is downloaded, cleaned and analysed using R (R Core Team 2023). The following packages are also utilized:

- tidyverse (Wickham 2021)

- lubridate (Grolemund & Wickham, 2011)
- dplyr (Wickham et al. 2023)
- ggplot2 (Wickham 2016)
- knitr (Xie, 2015)
- scales (Wickham, 2018)

2.2 Cleaned Data

A sample of the cleaned data (Table 1) and a line plot (Figure 1) showcasing every sales and redemption count data by date 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

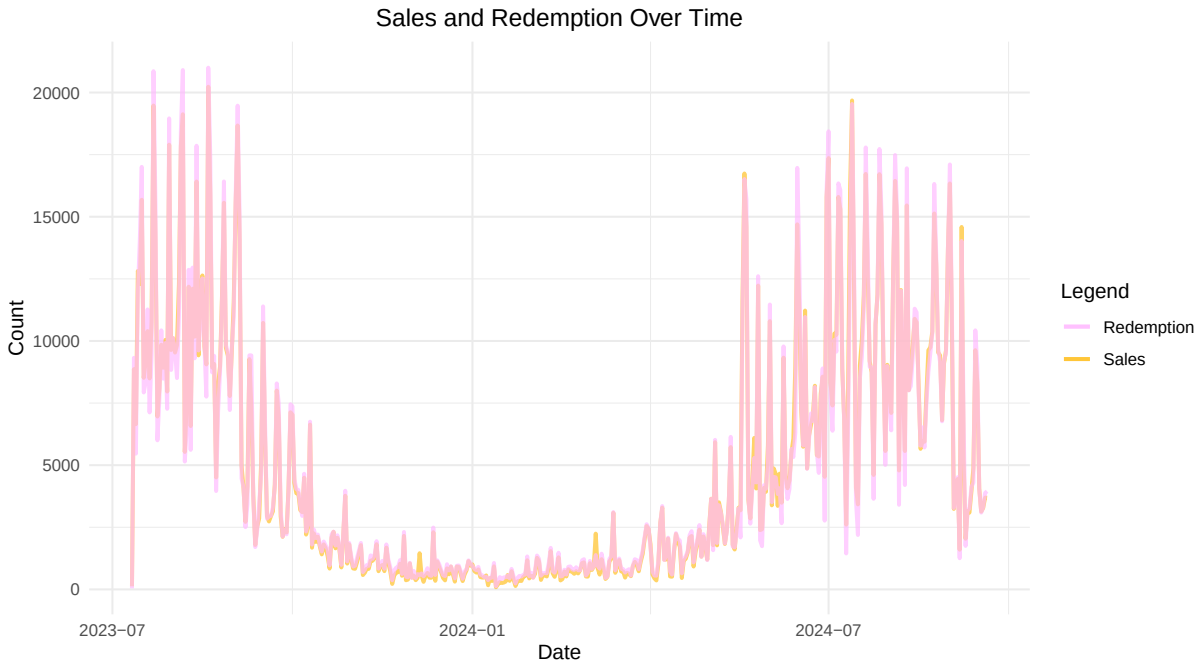

Figure 1: line plot of sales and redemption data with respect to time

From the line graph, we can see that the sales and redemption data is highly overlapped and that the values are much bigger in summer compared with winter.

2.3 Summary of Data

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

Table 2: Mean Sales and Redemption Count

Mean Sales	Mean Redemption
4746.505	4785.342

The mean sales is very close to mean redemption. From both Figure 1 and Table 2, the sales and redemption data look very alike. The analysis of the relationship between sales and redemption is done in the next section.

3 Result

3.1 Investigating the Relationship between the Sales and Redemption Data

To investigate the relationship between sales and redemption data directly, we first plot Figure 2 where it describe the relationship between the variables visually.

It is obvious to tell that they have a linear relationship. To further investigate, we do a linear regression with redemption count as the response variable and sales count as the predictor variable. As Table 3 shows, the R-squared statistic is 0.9882 and the correlation between redemption count and sales count is 0.994. These values indicates that the redemption count and sales count are indeed highly correlated.

In this paper we'll be mainly focusing on analyzing the redemption count data since it is the direct reflection of the visitor data and using it to analyze revenue is accurate enough since the two variables are higly overlapped and correlated.

Table 3: R-squared and Correlation Statistic

Statistics	Value
R-Squared	0.9882393
Correlation	0.9941023

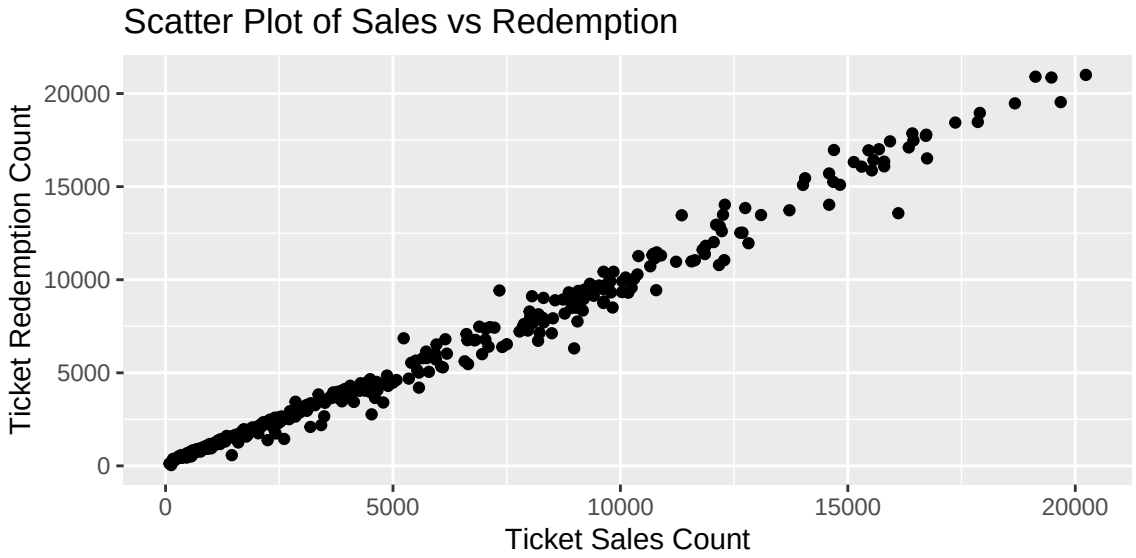

Figure 2: Relationship betwwen Sales and Redemption

3.2 Investigating the Frequency of Redemption Count Data

Figure 3 shows that for most of the days, the redemption count is below 10,000 times, while the rest of days has redemption count up to 20234 times in one day. For our purpose of using the information to help with resource allocation, monthly data would be sufficient enough so that we can create monthly schedule for staff. Creating daily schedules would be not as feasible and efficient as monthly schedules.

Figure 4 showcases the redemption count over each month. The dividing line in each bar is to indicate each day's redemption count and show how they add up together. It is evident that in Winter months, the redemption counts are very small, while in summer months, the redemption counts are very high. In August 2023, the redemption count is almost 350,000.

3.3 Investigating the Monthly Redemption Data

talk more.

4 Discussion

4.1 First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

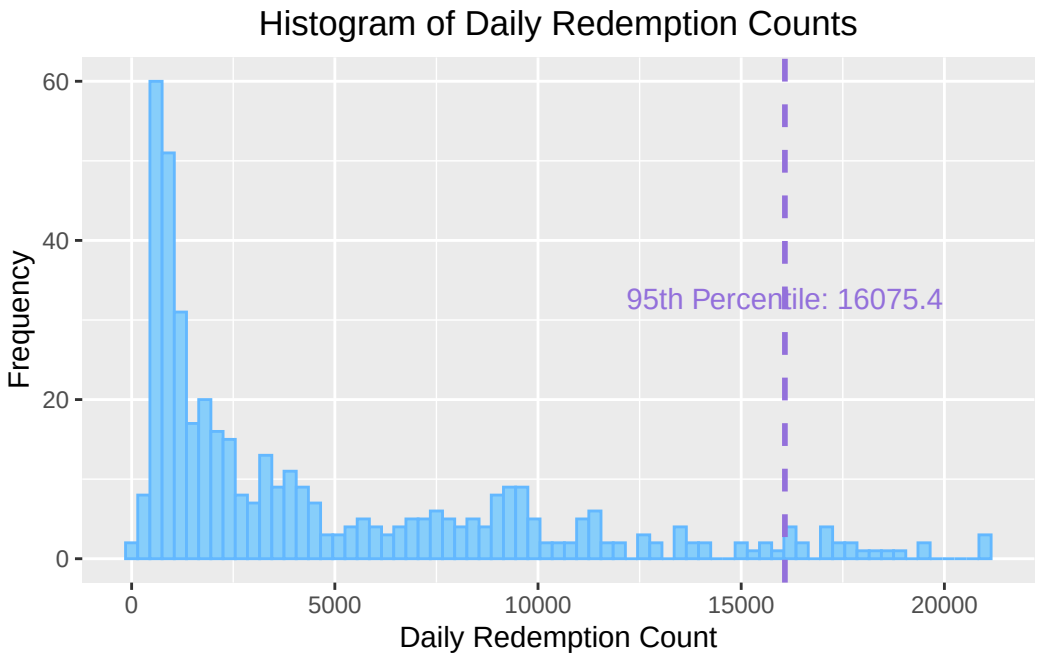

Figure 3: Histogram of Daily Redemption count data

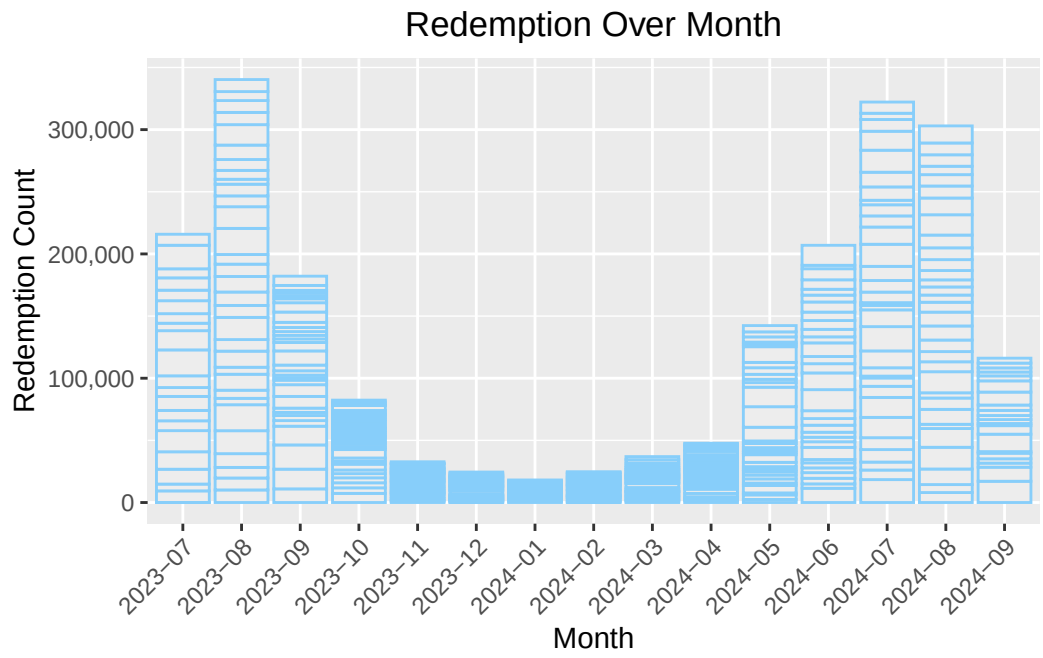

Figure 4: Redemption Count of Each Month

4.2 Second discussion point

4.3 Third discussion point

4.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

A Additional data details

B References