Analysis and Prediction of Shared Bicycle Usage at the University of Toronto St. George Campus*

Main Finding TBD

Haowei Fan

November 26, 2024

Students and staff at the University of Toronto's St. George campus often face difficulties finding a bike-sharing parking spot when arriving at campus, and finding a bike when leaving, which highlights the need for effective prediction of future bike-sharing demand to support campus commuting infrastructure. This study uses Bayesian Poisson regression to predict the usage of bike-sharing stations at 27 locations on campus for specific years, months, and four-hour intervals of the day. The results indicate that, on September 26, 2025, from 4:00 to 8:00 AM, usage at 8 stations will increase by 3, 14 stations by 2, 3 stations by 1, 1 station remains unchanged, and 1 station will decrease by 1 compared to the same period in 2024.

1 Introduction

Bike-sharing services have undoubtedly improved commuting efficiency, especially for short trips within a campus. For example, a study on bike-sharing in Xiamen, China, highlighted how data-driven analyses can inform policymakers about system performance, thereby guiding investments and policy decisions to enhance urban mobility[https://arxiv.org/abs/2401.03987?utm_source=chatgpt.com]. However, in the Toronto region, despite a study by the University of Toronto's School of Cities showing that ridership increased from about 665,000 trips in 2015 to over 4.5 million in 2022[https://schoolofcities.github.io/bike-share-toronto/growth], there has been little research focused on small-scale areas, such as universities, within Toronto. Furthermore, a study from Hangzhou, China, confirmed the significant differences in traveling characteristics between cities and university campuses[https://file.techscience.com/uploads/attached/file/20200916/20200916072445_75

*Code and data are available at: https://github.com/HaoweiFan0912/Bikeshare-Forecast.git

Therefore, this paper takes the University of Toronto as a case study to analyze bike-sharing usage within the campus.

The core of this study is to extract a sample of 27 bike-sharing stations within the University of Toronto campus from all bike-sharing usage data between January 1, 2017, and September 30, 2024. The total usage of bike-sharing at each station is calculated for each 4-hour interval. A Bayesian Poisson regression model is then employed to predict the usage based on the station, year, month, date, and the specific 4-hour interval of the day.

This study found that although the usage at all stations within the University of Toronto campus has increased rapidly each year, there are significant seasonal differences. During winter, usage at all stations, regardless of the year, remains close to zero. Additionally, from January to September each year, the usage shows an upward trend, which then gradually declines over the following months. There are also significant differences in usage across different times of the day. The period from 8 am to 8 pm accounts for an average of 79.8% of daily usage, while the period from midnight to 8 am accounts for only 6.8%. It is noteworthy that this project forecasts an average increase of 2 uses across 27 stations between 4:00 AM and 8:00 AM on September 26, 2025, compared to the same time period in 2024. Specifically, usage at 8 stations is expected to increase by 3, 14 stations by 2, and 3 stations by 1, while 1 station will remain unchanged and another will decrease by 1.

Compared to the general statistics on bike-sharing usage in the Greater Toronto Area, these findings offer more detailed recommendations specifically for transportation planning and policy adjustments within the University of Toronto campus. They also provide guidance on how to better allocate, deploy, or store shared bikes.

Structure: TBD

1.1 Estimand

This study aims to estimate the usage of a specific bike-sharing station within the University of Toronto campus during a particular 4-hour interval. By converting time into year, month, day, and the specific 4-hour interval of the day, it accounts for the overall trend, seasonal effects, and hourly variation in station usage. The core objective is to explore the temporal changes in the usage of 27 bike-sharing stations within the campus, thereby providing policymakers with recommendations for bike allocation to improve commuting efficiency on campus.

2 Data

2.1 Overview

The dataset used in this study comes from opendatatoronto [opendatatoronto], uploaded by Toronto Parking Authority and collected by Bike Share Toronto. It records every bike-sharing

usage in the Toronto area from 2015 to September 30, 2024, with a total of 28,017,329 records. The variables included in the data differ across years, but they all contain the following variables: Trip ID, Trip Duration, Trip Start Station ID, Trip Start Time, Trip Start Station Location, Trip End Station ID, Trip End Time, Trip End Station Location, Bike ID, and User Type.

This study follows the workflow of Telling Stories with Data [Telling Stories with Data], using its initial folder structure and part of its code. Data downloading, cleaning, modeling, and visualization were carried out using R [R]. The following R libraries were also used alongside R:

TBD

2.2 Measurement

To predict the relationship between bike-sharing usage and time within the University of Toronto's St. George campus, this study requires a time series to describe the changes in usage at different stations over time. Therefore, the dataset published by opendatatoronto [opendatatoronto], which records every instance of bike-sharing usage in the Toronto area since 2015, is ideal for this study. Additionally, this data is ideal because of its reliability—due to the commercial nature of bike-sharing, the specific time and location of each bike's use and return are accurately recorded. However, the raw data cannot be used directly in this project; sophisticated data cleaning is required, with specific steps and reasons as follows:

First, data from 2015-2016 was excluded because the collection and recording methods for those years differ significantly from later years and do not include the time and station variables required for this study. Next, data for stations located within the University of Toronto's St. George campus was extracted from all remaining samples, and the usage of each station was calculated for every four-hour interval. Finally, the data format was standardized for subsequent analysis. The cleaned data is in the format of Table 1.

Table 1: Samples of the dataset used for analysis

station_name	time	count
Willcocks St / St. George St	2024-09-29 12:00:00	1
Willcocks St / St. George St	2024-09-29 16:00:00	1
Willcocks St / St. George St	2024-09-30 04:00:00	1
Willcocks St / St. George St	2024-09-30 08:00:00	1
Willcocks St / St. George St	2024-09-30 12:00:00	7
Willcocks St / St. George St	2024-09-30 16:00:00	4

2.3 Variables

This study focuses on the following variables:

- `count`: The dependent variable of the study, a non-negative integer. It describes the total usage of bike-sharing at a particular station during a specific 4-hour interval.
- `time`: An independent variable representing the time interval. For example, "2024-09-29 12:00:00" represents the time interval from 12 pm to 4 pm on September 29, 2024, in a 24-hour format. The earliest `time` is "2017-01-01 00:00:00," and the latest is "2024-09-30 24:00:00."
- `station_name`: An independent variable representing the unique name of one of the 27 stations within the University of Toronto's St. George campus.

Figure 1 shows the daily usage totals of 27 shared bicycle stations from January 1, 2017, to September 30, 2024. It can be observed that the overall usage exhibits a significant upward trend, particularly after 2021, where usage fluctuations increased noticeably. In 2024, several peaks in daily usage reached historical highs, indicating a substantial growth in user demand. Additionally, the data displays some seasonal variations, with noticeable declines during the winter months and increases during spring and summer.

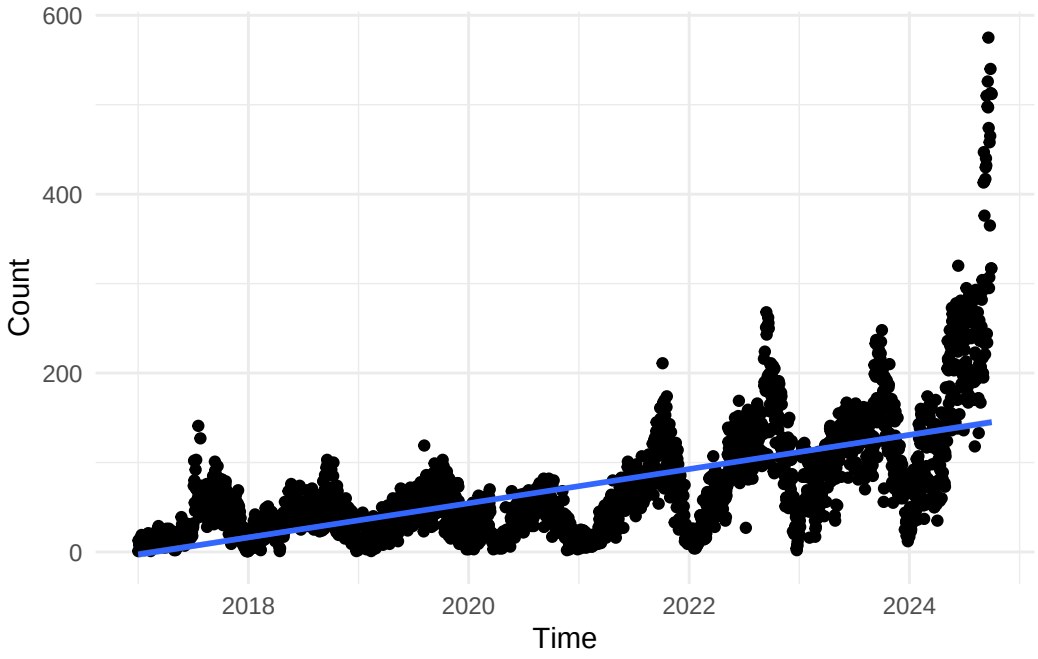

Figure 1: Daily usage from January 2017 to September 2024

hows the total usage and percentage of shared bicycles during different time intervals in a day from January 1, 2017, to September 30, 2024. It can be observed that the peak usage occurs

between 12:00 to 16:00 (28.5%) and 16:00 to 20:00 (31.4%). The usage between 08:00 to 12:00 accounts for 19.9%, while the usage from 20:00 to 00:00 accounts for 13.4%. The time intervals with the lowest usage are 00:00 to 04:00 (3.6%) and 04:00 to 08:00 (3.2%).

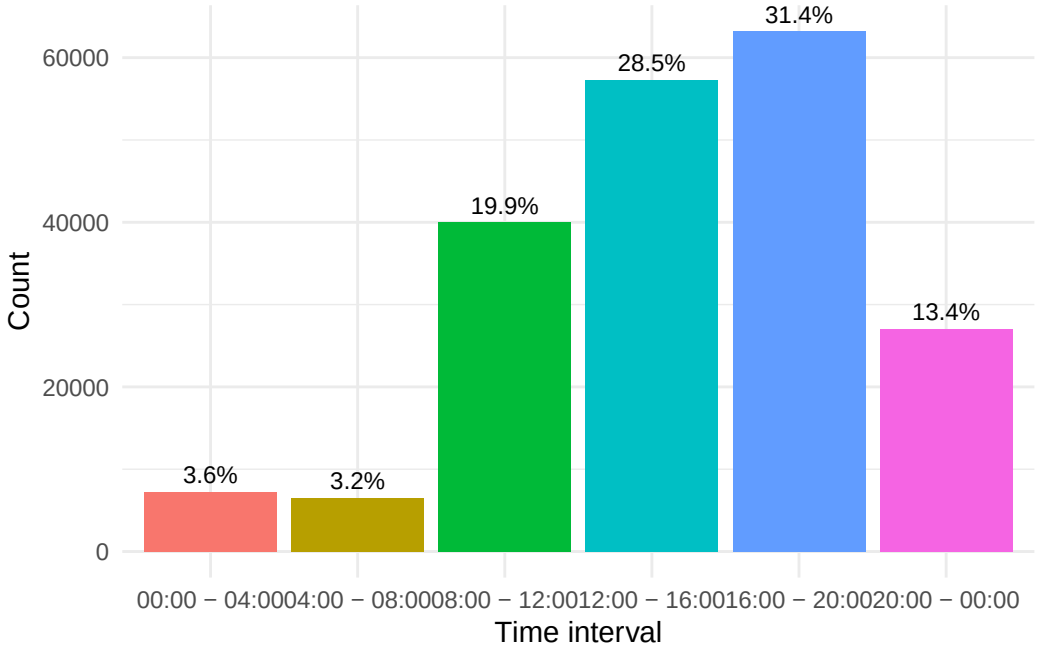

Figure 2: Total usage during different time periods from January 2017 to September 2024

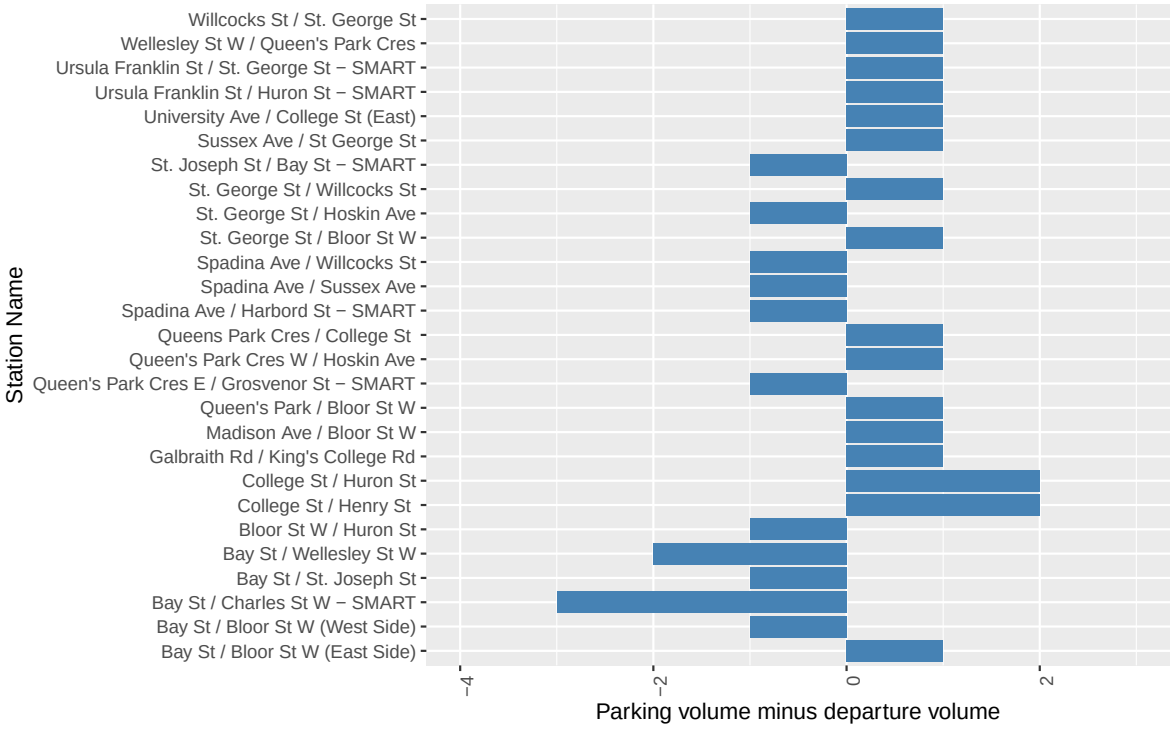

Figure 3: Difference of daily average of bicycle parking and departure volumes for each station