통계모델링 프로젝트

Analysis of Bus Traffic in Gwangju Metropolitan City

2024년 11월 12일

청주대학교 데이터사이언스학과

2021012800 베네딕터스 에스라 헤르노오

Table of Contents

1. Introduction

1.1. Overview of the Project

The purpose of this project is to analyze bus traffic patterns in Gwangju using real-world data from Public Data provided by South Korea's government. By examining the flow of passengers throughout the city, we can better understand usage trends and peak traffic hours, providing information into how people move across different areas of Gwangju. This projects offers an insight of transportation needs, which can be valuable for urban planning, resource allocation, and enhancing the overall efficiency of the city's bus system

1.2. Importance of the Project

Public transportation is essential for many residents in Gwangju, providing a reliable and affordable way to navigate the city. Understanding how traffic varies by time, day and station can help the city to make decisions based on the data to improve the bus transportation system. For example, if we know the busiest hours for certain stations, we could optimize the bus schedule to reduce wait times and improve the service quality. In addition, this analysis can highlight trends in boarding and alighting patterns, which could be important in managing crowds during the peak hours.

1.3. Project Scope and Limitations

This project focuses on Gwangju's bus traffic data over a defined period of time, with attention to passenger counts at various time intervals and stations. The dataset includes data fields like date (날짜), station code (역번호), station name (역명), boarding or alighting status (구분), time period (시간대), and traffic count (traffic_count). However, there are some limitations. This analysis does not account for holidays, weather conditions, or other external factors that could impact bus usage. Furthermore, since the dataset is limited to specific stations and time periods, conclusions drawn here might not fully generalize to the entire bus network.

2. Background

2.1. Introduction to Gwangju Bus System

Public transportation services, particularly the bus system, are very important to the people of Gwangju as they provide easy movement about the city and to the neighboring regions. The bus system is under the local authority, and as such, it serves a great number of interrelated functions for residents and tourists alike by linking residential areas with workstation zones, educational and recreational facilities. In Gwangju, the buses run on a timetable that is properly followed, covering most regions with fairly high regular and frequent services.

The network is made up of several routes along which many people and their stations are served, operating at least thousands of people in a day. The Gwangju bus system has several issued bus lines – such as express, rapid and standard – each serving a given public’s need within the commuting system. Its purpose is to provide efficient service to all the system users no matter their geographical location even if it is reducing the distances for the urban centers and pushing out to the suburban ones.

In this context, we emphasize the study of specific parameters such as the number of passengers boarding (승차) and alighting (하차) the different stations. This fragment of the data provides information on the most loaded routes on which hours of operation, days, and season the majority of the passengers use them and provides an overall passenger profile. This enables us to model the functioning of the bus transport system of the city as a function of demand, the knowledge of which will be pivotal in revising the existing system and allocating new resources for the public transport system of Gwangju.

3. Data Collection

3.1. Sources of Data

The data for this analysis was obtained from South Korea's Public Data Portal (data.go.kr), a government platform that offers open access to a variety of public datasets. The data provided from the portal was covering the period from January 1, 2024, to October 31, 2024.

This data was collected for having the latest data input of passenger traffic across Gwangju's bus system, that can be used to examine trends over an extended period.

3.2. Data Overview

A data.frame: 6 × 28

	날짜	역번호	역명	구분	X03.04시	X04.05시	X05.06시	X06.07시	X07.08시	X08.09시	⋯	X17.18시	X18.19시	X19.20시	X20.21시	X21.22시	X22.23시	X23.00시	X00.01시	X01.02시	X02.03시
	<chr>	<int>	<chr>	<chr>	<int>	<int>	<int>	<int>	<int>	<int>	⋯	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>
1	2024-01-01	1101	판암	승차	0	66	61	55	50	87	⋯	120	129	54	47	33	13	13	2	0	0
2	2024-01-01	1101	판암	하차	0	53	50	74	48	51	⋯	145	130	111	85	73	94	62	17	0	0
3	2024-01-01	1102	신흥	승차	0	0	18	106	21	40	⋯	41	37	30	15	16	8	3	0	0	0
4	2024-01-01	1102	신흥	하차	0	0	9	104	17	15	⋯	45	69	39	37	27	39	20	6	0	0
5	2024-01-01	1103	대동	승차	0	27	94	31	51	77	⋯	116	84	102	52	49	37	28	0	0	0
6	2024-01-01	1103	대동	하차	0	25	85	58	70	32	⋯	103	112	85	98	85	106	99	14	0	0

The dataset incorporates multiple variables that elaborate on the daily traffic counts at station and temporal levels. Among important variables are the following:

- 날짜: The date recorded in the yyyy-mm-dd format to indicate the day of the data collection.
- 역번호: The bus station code that acts as an identification number for each bus station.
- 역명: The bus station title, which means the designated bus stop's name.
- 구분: This denotes the boarding (승차) or alighting (하차) related data, thus enabling both the in passenger traffic and out passenger traffic analysis.
- 03-04시 to 02-03시: Traffic counts by hourly basis for the entire day, that is, early in the morning to very late in the night showing how bus usage varies with different hours of the day.

In total, the data set contains 12,716 variables out of which each column represents an hour of the day or some other field regarding the number of passengers. This ensures an in depth traffic analysis both across time and space in turn establishing the basis for extensive research.

4. Data Preprocessing

4.1. Data Formatting

The first step, data required cleaning to make it more suitable for analysis. The columns consist of “시” suffix was remove from the time period to standardize the format. This will make the time data as numeric rather than string-based, making it easier to manipulate and analyze. After that, data consistency was ensured across column names and checked for any irregularities that could interfere with the analysis.

4.2. Data Reshape

After the data was cleaned, the data once again reprocessed by reshaping it into long format using the pivot_longer function in R. This transformation involved consolidating the hourly columns into two new columns, time_period and traffic_count. The time_period column now consists of each hour of the day variable, while traffic_count holds the associated passenger count. Reshaping the data in this way makes analyzing patterns better over time without the constraints of multiple hourly columns.

4.3. Creating New Dataset

날짜	역번호	역명	구분	time_period	traffic_count
<chr>	<int>	<chr>	<chr>	<chr>	<dbl>
2024-01-01	1101	판암	승차	03	0
2024-01-01	1101	판암	승차	04	66
2024-01-01	1101	판암	승차	05	61
2024-01-01	1101	판암	승차	06	55
2024-01-01	1101	판암	승차	07	50

After reshaping, the new dataset (referred to as *data_clean*) was structured with the following columns:

- 날짜: The date of the data record.
- 역번호: Unique station code.
- 역명: Station name.
- 구분: Boarding (승차) or alighting (하차) indicator.
- time_period: The time of day in hourly intervals.
- traffic_count: The count of passengers for that specific hour.

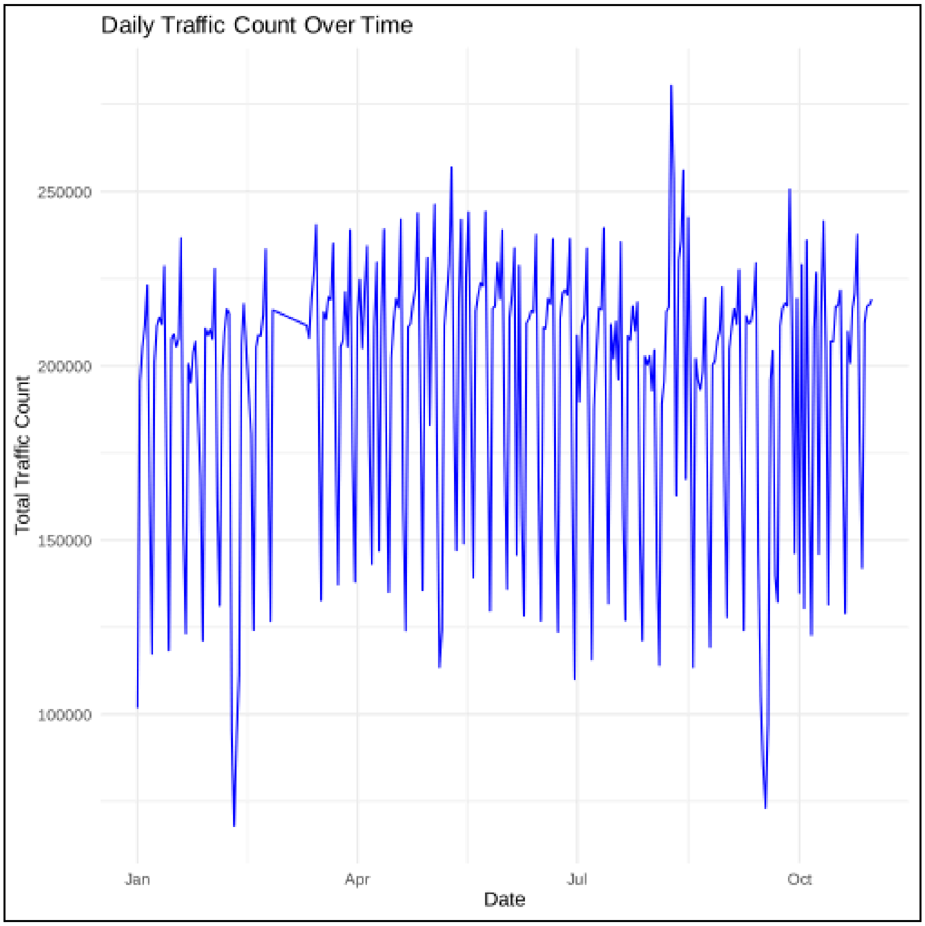
This structure makes the data more accessible for time-based and station-based analyses, setting the stage for focused exploration of boarding and alighting patterns across time and location.

4.4. Handling Missing Values

Upon inspection, there were no missing values (NA) in the traffic_count column, so no further action was required for handling missing data. With this inspection, the dataset was ready for analysis.

5. Exploratory Data Analysis

5.1. Daily Traffic Count over Time



To begin with the analysis, the total traffic count over the time period covered by the data period of time, January to October 2024, was examined. This helped to observe any fluctuations in trends in bus usage throughout the year.

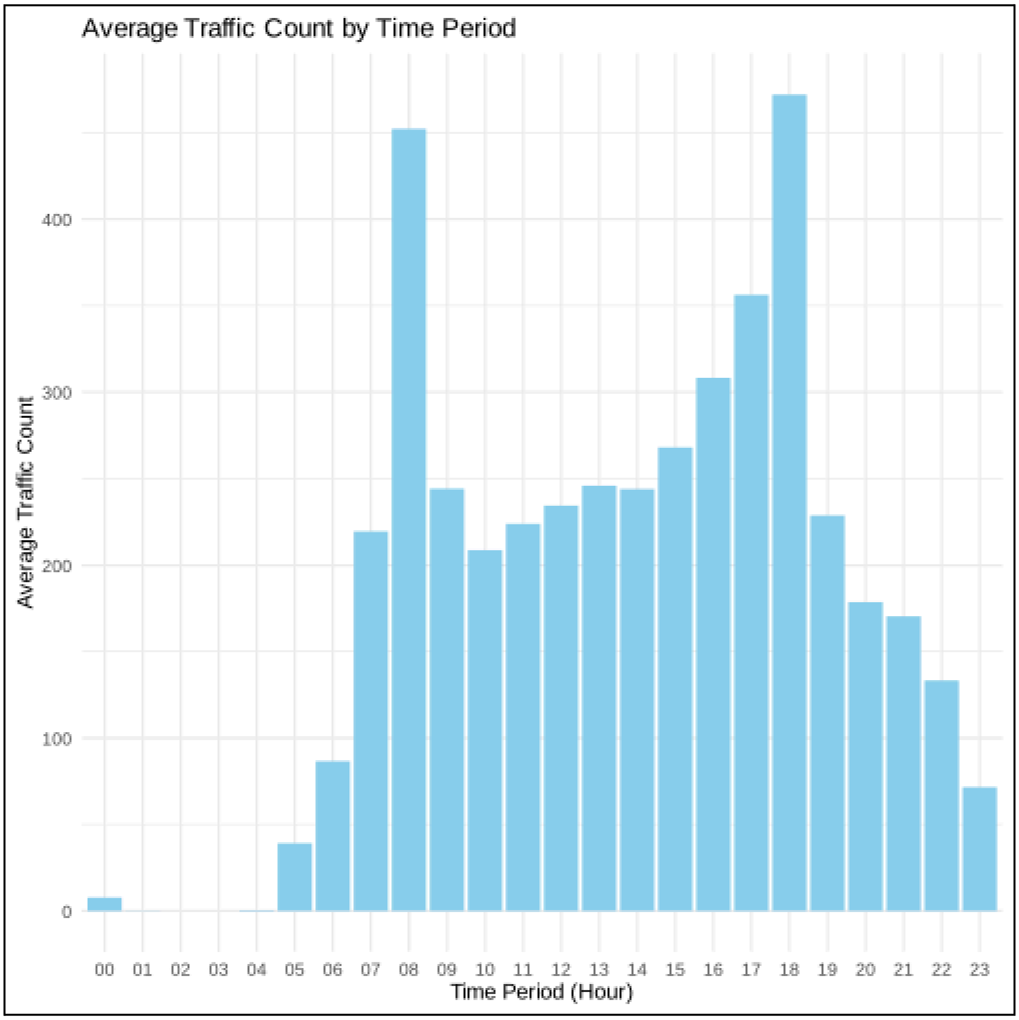
During this period, we can see that there are several significant spikes and drop:

- **February** and **September**: These months showed a significant drop in traffic count, deviating from the average usual monthly trend.
- **August:** On the other hand, this month showed a spike, with higher traffic counts compared to other months, passing the average trend line of monthly trend.

Other Months' traffic remained relatively stable across the year, generally fluctuating between 100,000 passengers to 250,0000 passengers monthly.

These extreme fluctuations are likely caused by several factors such as seasonal events, public holidays, or maybe changes in commuting patterns. The cause of spikes and drops needed further analysis to conclude the reason.

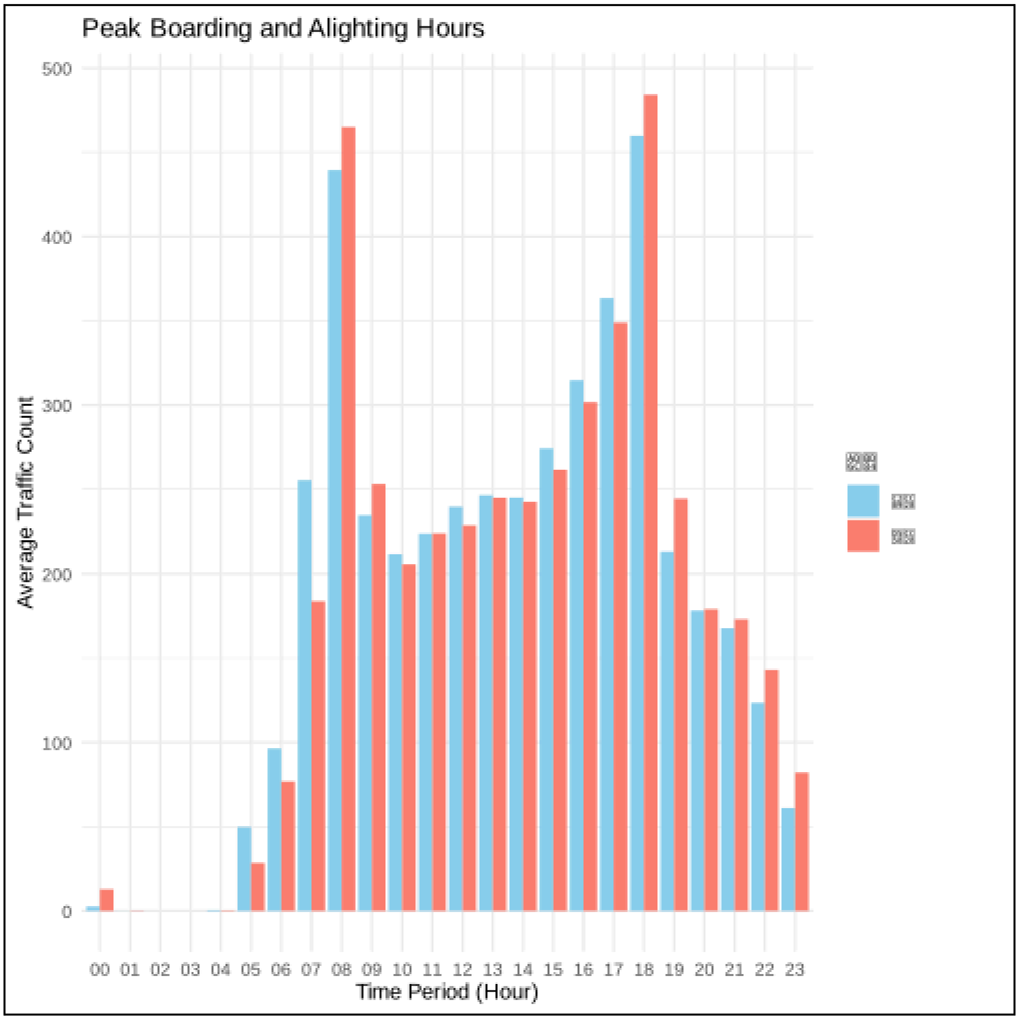
5.2. Traffic Distribution by Time Period



Traffic distribution analysis for the different hourly intervals is dissimilar, reflecting peak hours of commute and consistent usage during working hours. One of the easily recognizable peaks is at 8:00 AM, where the traffic count crosses 450, probably because of the morning commute as people go to work and school. It is around 6 PM when another prominent peak is attained in the evening, slightly over 460 are attained in this peak of the evening. This evening peak is contributed by passengers who pass in a gradual incline from about 3 PM to around 350 at 5 PM before giving the peak at 6 PM.

Over the whole day, from 9 AM to 5 PM, the volume of traffic does not go below 200, demonstrating a consistent movement of passengers aboard. On this route to and from work, it could also be shared by workers and students, among other people using public means of transport. These are hourly trends that may further provide information on the daily fluctuations of the demand for bus services.

5.3. Peak Boarding and Alighting Hours

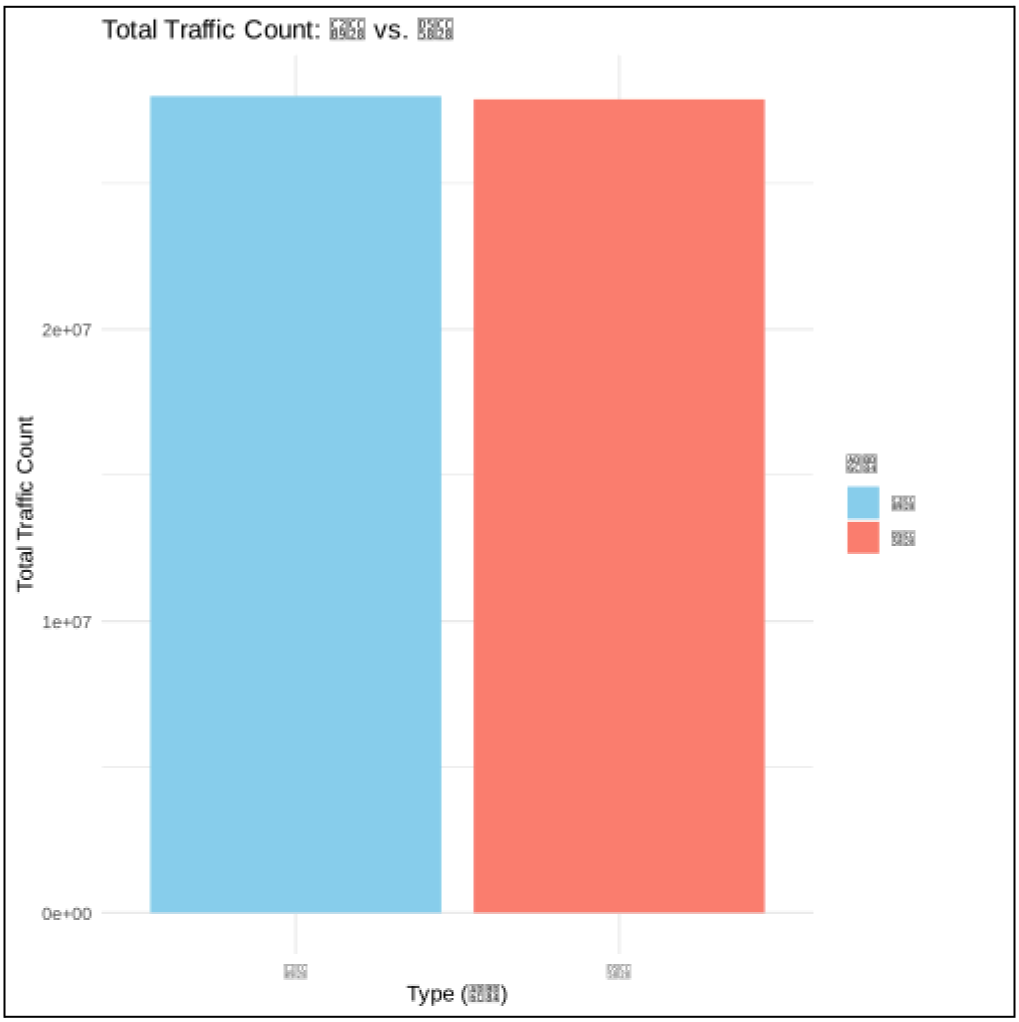


The results indicate that the most active periods are 8 AM and 6 PM, which coincides with the main working hours. The boarding volume at 8 am is about 439 persons while alighting is even higher by approximately 465 persons. This disparity is apparent as quite a number of people are expected to be getting off the buses at this time making their way to the offices, schools, or any other morning engagements.

The same case occurs in the evening at 6 pm with a high count of about 460 passengers boarding the bus while the count of those alighting is fairly high at around 484. This trend is characteristic of the evening peak period when there is a traffic buildup as passengers scramble to board the buses to go home or to other places after work.

These peaks for both boarding and alighting called for a lot of emphasis in travel demand management and capacity utilization on the morning and evening peak hours. Also, the small disparity between the figures for boarding and alighting of the passengers during these peak hours points out the normal flow trends, for instance, more passengers alighting in the morning and more boarding in evenings.

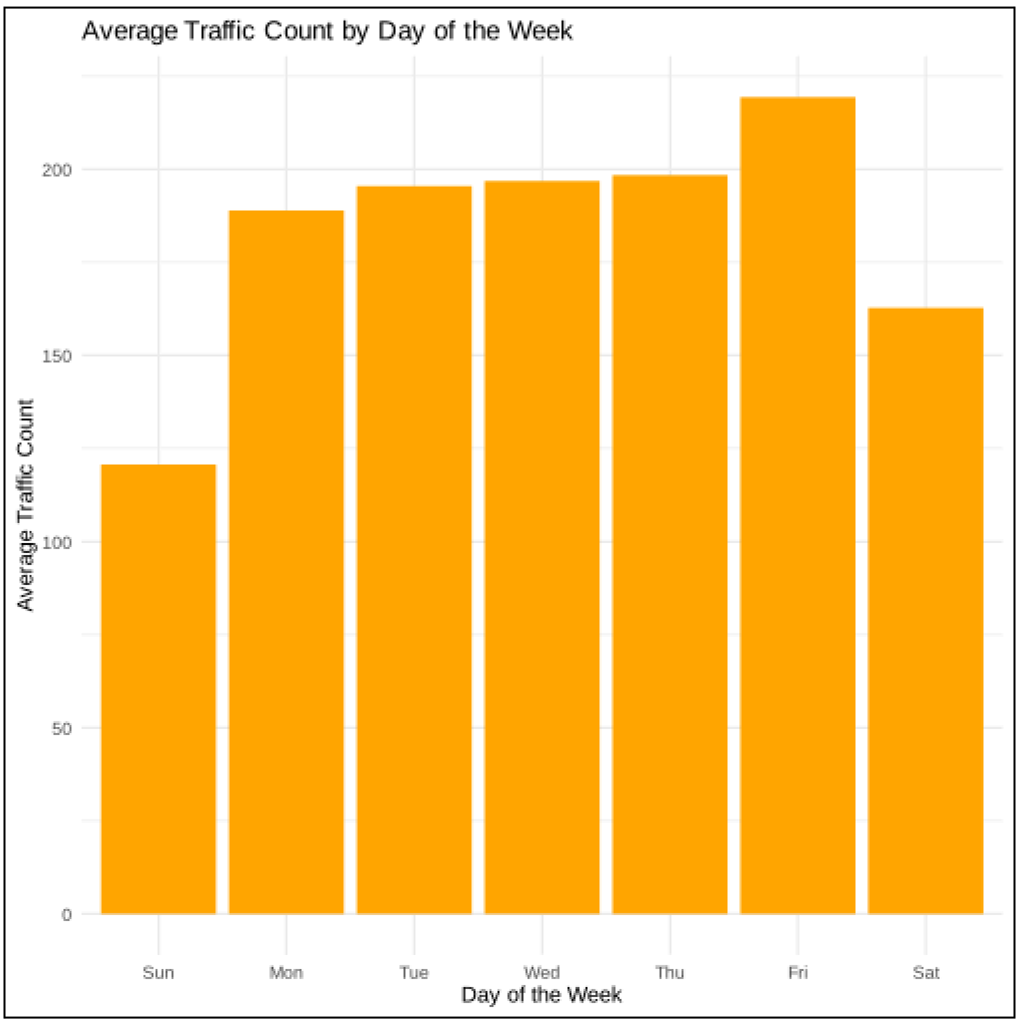
5.4. Total Traffic: 승차 vs 하차



The total count of boarding (승차) and alighting (하차) over the observed period within the Gwangju bus system. The total boarding count reached 27,985,335, while the alighting count closely followed at 27,878,543.

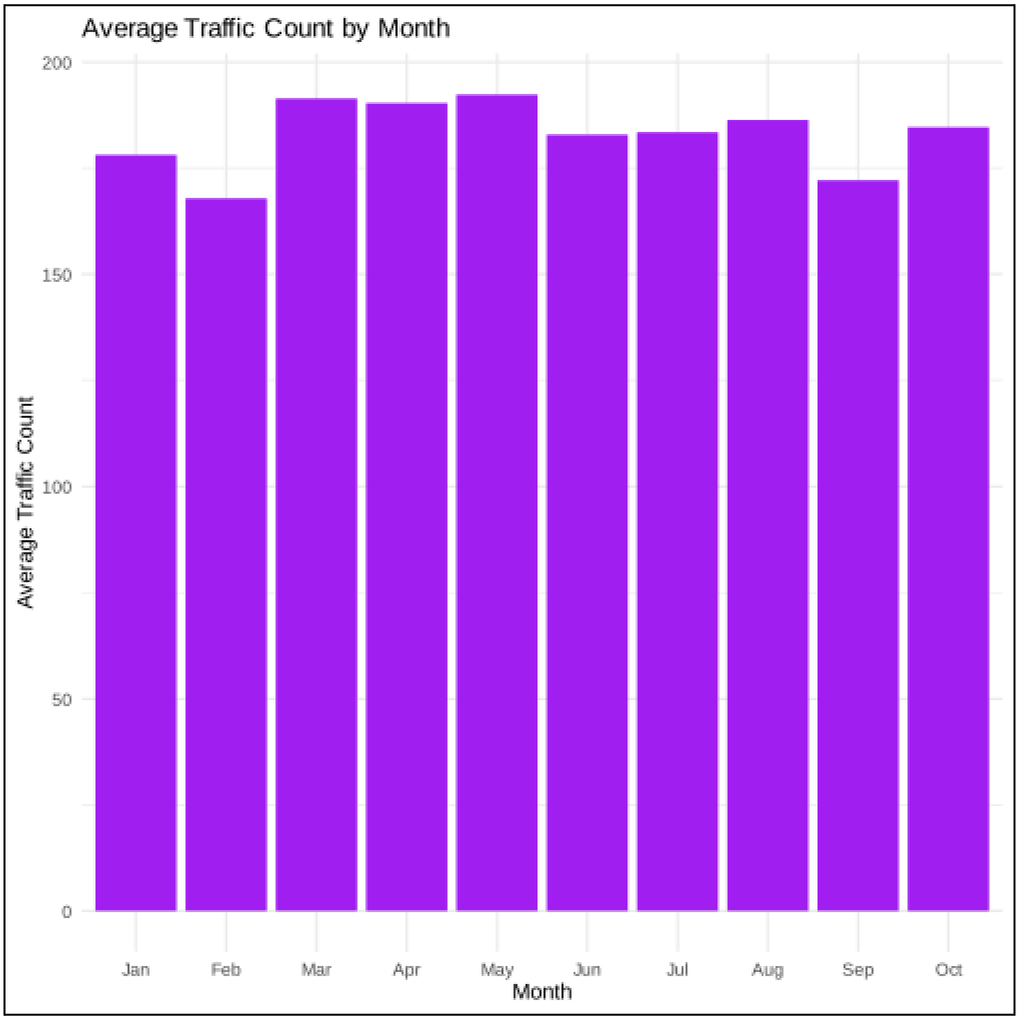
The near-equivalence between boarding and alighting counts suggests a well-balanced distribution of passengers entering and exiting the system. This balance is a positive indicator of effective route coverage and system utilization, as it implies that most passengers are completing their journeys within the network, likely without substantial overflow or underuse on particular routes.

5.5. Traffic Count by Day of the Week



From the data, we can see that the average passenger count from Monday to Thursday remains relatively stable, ranging from approximately 175 to 200 passengers. This suggests a consistent demand likely influenced by regular weekday activities such as commuting to work or school. Notably, Friday experiences a peak in passenger count, surpassing 213 passengers on average. This increase may be attributed to the additional activities that occur at the end of the workweek, possibly including social outings and extended operating hours for some services.

5.6. Average Traffic Count by Month

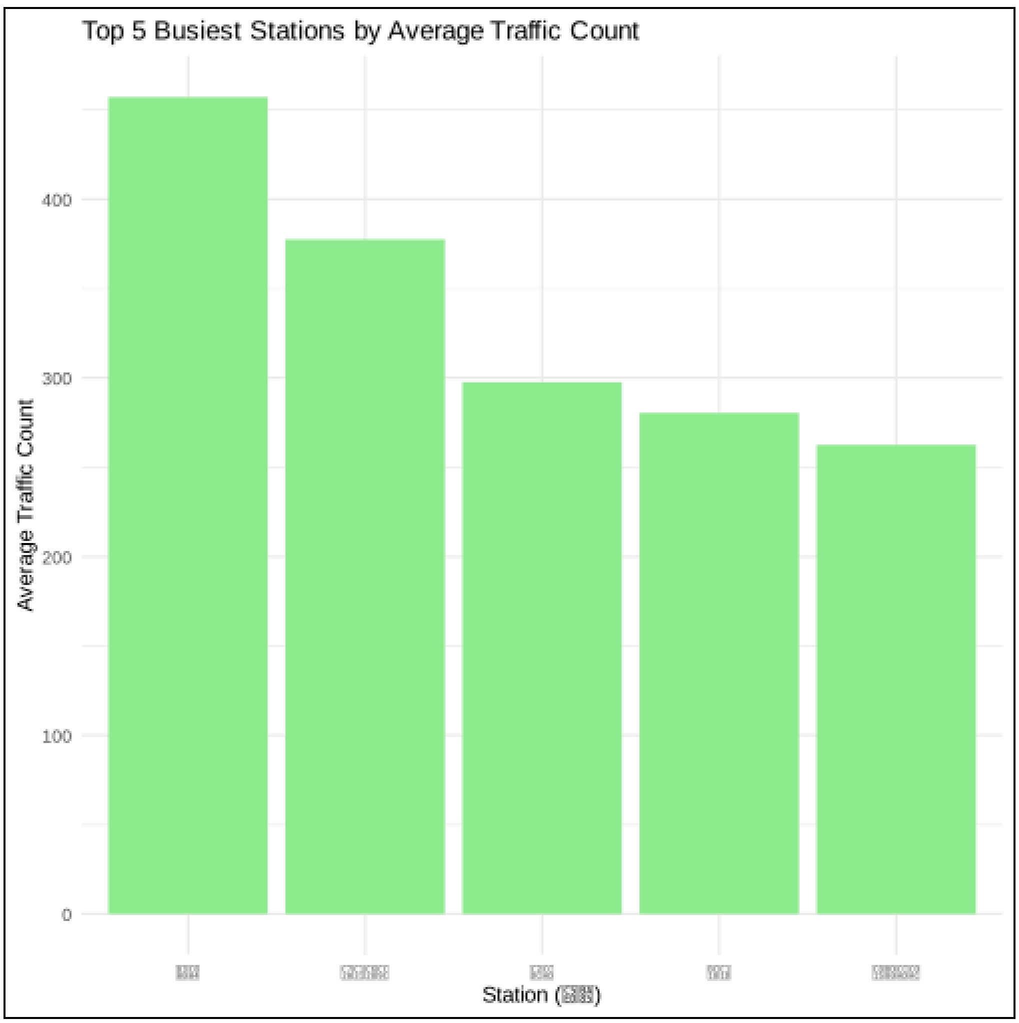


The data on the average monthly traffic shows that there exists a consistent year round demand for the Gwangju bus system with the passengers using the system standing at a figure ranging between 175 to 185 passengers per month. The stability of the bus system in usage implies that, as a whole, the bus system has a relatively the same level of usage throughout the year, regardless of any major seasonal changes.

Nonetheless, February and September seem to show a bit of traffic reduction . Monthly traffic for these months come at slightly lower than 175. This slump can be explained by certain characteristics of such months, for example February being a winter month probably not ideal for travel or September being a month of seasons for schools or holidays in the public sector.

As far as the general number of monthly traffic concerning passengers is concerned, the consistency and stability of the monthly traffic emails that planning and management of resources should follow a more or less the same pattern for the greater part of the year apart from the slight depression of the months in February and September which merits some attention mainly for the purpose of investigating and understanding the issues.

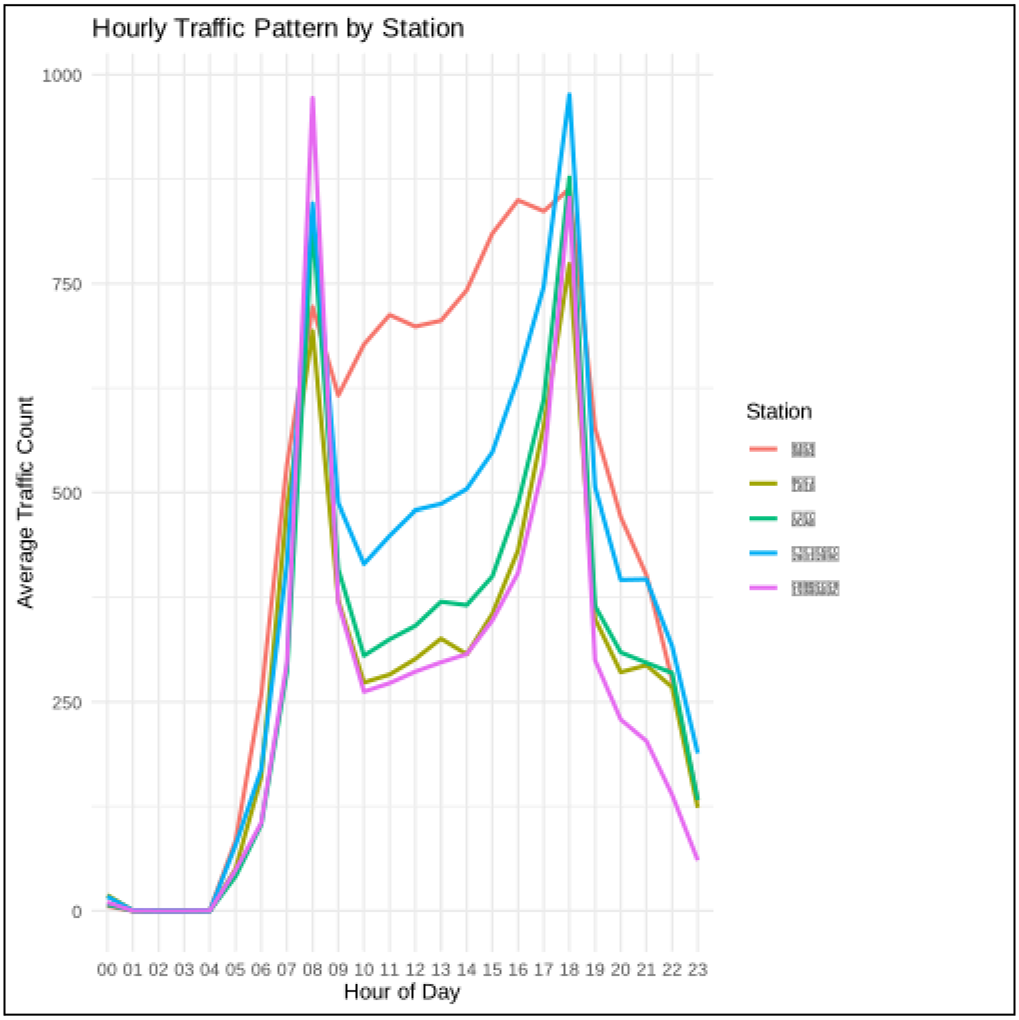
5.7. Top 5 Busiest Stations by Average Traffic Count



The analysis of average traffic count by station identifies the top 5 busiest locations, highlighting key areas with the highest passenger activity. These stations are crucial nodes within the transportation network and represent significant points of transit demand. The top 5 stations by average traffic count are as follows:

- 대전: This station holds the highest average traffic count, with an impressive 457.15 passengers. As a central hub, this station likely serves a high volume of daily commuters.
- 유성온천: With an average traffic count of 377.48, this station is the second busiest, suggesting its importance as a key transit location.
- 시청: The station at 시청, or City Hall, sees an average traffic count of 297.49, reflecting its role in connecting people to administrative and central business district areas.
- 반석: This station records an average of 280.30 passengers, positioning it as the fourth busiest station in the dataset.
- 정부청사: Finally, the Government Complex station averages 262.39 passengers, making it another high-traffic location likely frequented by both government employees and the public.

5.8. Hourly Traffic Pattern by Station



The hourly traffic patterns at the top 5 busiest stations—대전, 유성온천, 시청, 반석, and 정부청사—reveal distinct peaks and flows, particularly during commuting hours. 대전, the busiest station, maintains consistently high traffic throughout the day, with clear spikes around 7-8 AM and 6-7 PM, aligning with typical work commute times. 유성온천 follows a similar pattern, though with slightly lower peak counts, indicating its role as another key commuter hub. At 시청, traffic increases significantly in the morning and late afternoon, reflecting its position as a central location during business hours. 반석 sees a more steady flow, with less fluctuation throughout the day, while 정부청사 experiences regular peaks in the morning and late afternoon, consistent with standard office hours. These patterns highlight the importance of adjusting transportation resources at these critical times and locations to meet commuter demand effectively.

6. Results

6.1. Interpretation of Results and Insights Gained

Several insights were gained from the analysis regarding bus traffic patterns in the city of Gwangju. To begin with, a traffic count that spanned the months of January to October 2024 indicated that passenger volumes remained largely constant with sporadic exceptional increases. In the months of August, October and the months of February and September traffic volume decreased significantly owing to seasonal traveling behavior or particular activities. The time distribution of traffic also reinforced the high bus patronage during the peak times, which was evident with sharp upward trends around the hours of 7 am and 6 pm. These periods of peak bus utilization could be attributed to the cyclical nature of activities of most individuals who leave their homes for certain hours of the daytime primarily to go to work or schools and return to their houses at another set time. There is therefore a clear indication that the use of public transport systems during these times is very high.

Analyzing boarding (승차) versus alighting (하차) revealed that peak hours for both activities occur at 8 AM and 6 PM, with slightly higher alighting counts in the morning and boarding counts in the evening. This suggests that certain areas serve as arrival hubs in the morning and departure points in the evening. The top 5 busiest stations—대전, 유성온천, 시청, 반석, and 정부청사—further highlight important transit hubs with high passenger volumes. 대전, the station with the highest average traffic, experiences consistent usage throughout the day, especially during commuting hours, underscoring its role as a central hub. Hourly traffic patterns at these stations revealed that while some stations maintain steady usage, others fluctuate, aligning with peak commute hours and reflecting the stations' roles in Gwangju's transit network.

6.2. Limitations and Assumptions

Several limitations and assumptions were encountered during this analysis. First, the data only covers January to October 2024, so seasonal patterns or unusual events in the missing months might not be accounted for. Additionally, we assume that all missing values were valid zeros due to no actual null values present. Another limitation is that our analysis did not account for external factors such as weather, holidays, or special events, which could significantly impact traffic patterns. Lastly, the analysis assumes a direct correlation between traffic counts and general public demand, although other factors like route changes or schedule adjustments may also influence these numbers.

7. Conclusion

The study aimed to assess Gwangju bus traffic data obtained from January to October 2024 to identify trends in passenger demand across different times, locations, and modes. It was particularly evident that there were increased traffic volumes during the early morning and evening peak times, specifically at 7 a.m. and 6 p.m. Key transport hubs like 대전 and 유성온천 recorded deliberately increased volumes whenever their services were operational, illustrating the importance of these areas in Gwangju's public transport network. Furthermore, differences that were noticeable in the settlement boarding patterns and the settlement alighting patterns suggested the presence of waged migration, with certain stations being notably frequented in the morning for entry and in the evening for exit.

In summary, this study highlights the significant reliance of Gwangju's population on bus services for commuting and suggests potential adjustments to schedules during peak times for these services. While this study presents important findings, future research could concentrate on incorporating additional data sources such as weather information or event management systems to enhance the understanding of external factors affecting public transport usage.

References

- Public Data Portal (data.go.kr). "Traffic accident statistics by accident type." Retrieved from: https://data.go.kr
- Hadley Wickham, Garrett Grolemund. *R for Data Science*. O'Reilly Media, 2016. This book provides foundational concepts in data wrangling, data visualization, and exploratory data analysis using R.
- Hyun Kyung Kim, et al. (2018). "The Effects of Weather on Public Transit Ridership in Seoul, South Korea." *Journal of Transport Geography*, 69, pp. 95-108. This article discusses factors influencing public transit patterns and could be useful in contextualizing results.
- City of Gwangju Transportation Department. (2023). *Gwangju Public Transport Usage Report*. Retrieved from Gwangju.go.kr. This report provides detailed insights into transportation trends, relevant for discussing contextual background.
- James, G., et al. (2013). *An Introduction to Statistical Learning with Applications in R*. Springer Texts in Statistics. A comprehensive guide on statistical analysis and modeling, offering insights useful for interpreting traffic data.

Appendix

A. Data Source

A.1. Data

https://www.data.go.kr/en/data/15060591/fileData.do#/

A.2. Data table

A data.frame: 6 × 28

	날짜	역번호	역명	구분	X03.04시	X04.05시	X05.06시	X06.07시	X07.08시	X08.09시	⋯	X17.18시	X18.19시	X19.20시	X20.21시	X21.22시	X22.23시	X23.00시	X00.01시	X01.02시	X02.03시
	<chr>	<int>	<chr>	<chr>	<int>	<int>	<int>	<int>	<int>	<int>	⋯	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>	<int>
1	2024-01-01	1101	판암	승차	0	66	61	55	50	87	⋯	120	129	54	47	33	13	13	2	0	0
2	2024-01-01	1101	판암	하차	0	53	50	74	48	51	⋯	145	130	111	85	73	94	62	17	0	0
3	2024-01-01	1102	신흥	승차	0	0	18	106	21	40	⋯	41	37	30	15	16	8	3	0	0	0
4	2024-01-01	1102	신흥	하차	0	0	9	104	17	15	⋯	45	69	39	37	27	39	20	6	0	0
5	2024-01-01	1103	대동	승차	0	27	94	31	51	77	⋯	116	84	102	52	49	37	28	0	0	0
6	2024-01-01	1103	대동	하차	0	25	85	58	70	32	⋯	103	112	85	98	85	106	99	14	0	0

B. Source Code

```
# Data Preparation
## Packages load
# Load Packages
library(tidyverse)
library(ggplot2)
library(dplyr)
library(readr)
## Dataset Load
#Load Dataset
data <- read.csv("data.csv", fileEncoding= "EUC-KR")
head(data)
str(data)
colnames(data)
## Data Cleaning
# Standardize column names
colnames(data) <- gsub("시", "", colnames(data))
colnames(data) <- gsub("\\..*", "", colnames(data))
# Reshape data to long format
data_long <- data %>%
```

```
  pivot_longer(cols = starts_with("X"),
               names_to = "time_period",
               values_to = "traffic_count")
# Clean data suffix
data_long$time_period <- gsub("X", "", data_long$time_period)
data_long$time_period <- gsub("\\.", "-", data_long$time_period)
data_long$traffic_count <- as.numeric(data_long$traffic_count)
# Clean Data with NA Values
data_clean <- data_long %>%
  filter(!is.na(traffic_count))
colnames(data_clean)
str(data_clean)
## Handling Missing Values
# Confirm for NA Values
sum(is.na(data_clean))
summary(data_clean)
# Explarotary Data Analysis
## Traffic Count Over Time
data <- data_clean
# Making new group by feature
daily_traffic <- data %>%
  group_by(날짜) %>%
  summarise(total_traffic = sum(traffic_count, na.rm = TRUE))

# Convert 날짜 to Date format
daily_traffic$날짜 <- as.Date(daily_traffic$날짜)

# Data Visualization
ggplot(daily_traffic, aes(x = 날짜, y = total_traffic)) +
  geom_line(color = "blue") +
  labs(title = "Daily Traffic Count Over Time",
       x = "Date",
       y = "Total Traffic Count") +
  theme_minimal()
## Traffic Distribution by Time Period
```

```
# Group by time_period and average traffic count
time_period_traffic <- data %>%
  group_by(time_period) %>%
  summarise(avg_traffic = mean(traffic_count, na.rm = TRUE))

# Data Visualization
ggplot(time_period_traffic, aes(x = time_period, y = avg_traffic)) +
  geom_bar(stat = "identity", fill = "skyblue") +
  labs(title = "Average Traffic Count by Time Period",
       x = "Time Period (Hour)",
       y = "Average Traffic Count") +
  theme_minimal()
## 승차 vs. 하차 Comparasion
# Group by 구분 and sum the total traffic
boarding_alighting <- data %>%
  group_by(구분) %>%
  summarise(total_traffic = sum(traffic_count, na.rm = TRUE))
boarding_alighting

# Data Visualization
ggplot(boarding_alighting, aes(x = 구분, y = total_traffic, fill = 구분)) +
  geom_bar(stat = "identity") +
  labs(title = "Total Traffic Count: 승차 vs. 하차",
       x = "Type (구분)",
       y = "Total Traffic Count") +
  theme_minimal() +
  scale_fill_manual(values = c("skyblue", "salmon"))

## Top 5 Busiest Stations
# Group by 역명 and calculate the average traffic_count
busiest_stations <- data %>%
  group_by(역명) %>%
  summarise(avg_traffic = mean(traffic_count, na.rm = TRUE)) %>%
  arrange(desc(avg_traffic)) %>%
  slice(1:5)  # Select top 5
```

```
busiest_stations

# Plot the top 5 busiest stations
ggplot(busiest_stations, aes(x = reorder(역명, -avg_traffic), y =
avg_traffic)) +
  geom_bar(stat = "identity", fill = "lightgreen") +
  labs(title = "Top 5 Busiest Stations by Average Traffic Count",
       x = "Station (역명)",
       y = "Average Traffic Count") +
  theme_minimal()
## Average Traffic Count by Day of the Week
# Load necessary library for date manipulation
library(lubridate)

# Add a new column for the day of the week
data <- data %>%
  mutate(day_of_week = wday(날짜, label = TRUE, abbr = TRUE))

# Group by day_of_week and calculate the average traffic_count
weekly_traffic <- data %>%
  group_by(day_of_week) %>%
  summarise(avg_traffic = mean(traffic_count, na.rm = TRUE))

# Plot the average traffic by day of the week
ggplot(weekly_traffic, aes(x = day_of_week, y = avg_traffic)) +
  geom_bar(stat = "identity", fill = "orange") +
  labs(title = "Average Traffic Count by Day of the Week",
       x = "Day of the Week",
       y = "Average Traffic Count") +
  theme_minimal()
## Hourly Traffic Pattern by Station
selected_stations <- busiest_stations$역명[1:5]

# Filter data for selected stations and group by 역명 and time_period
hourly_station_traffic <- data %>%
```

```
  filter(역명 %in% selected_stations) %>%
  group_by(역명, time_period) %>%
  summarise(avg_traffic = mean(traffic_count, na.rm = TRUE))

# Plot the hourly traffic pattern for selected stations
ggplot(hourly_station_traffic, aes(x = time_period, y = avg_traffic,
color = 역명, group = 역명)) +
  geom_line(size = 1) +
  labs(title = "Hourly Traffic Pattern by Station",
       x = "Hour of Day",
       y = "Average Traffic Count",
       color = "Station") +
  theme_minimal()
## Monthly Traffic Trend
# Extract the month from 날짜 and add as a new column
data <- data %>%
  mutate(month = month(날짜, label = TRUE, abbr = TRUE))

# Group by month and calculate the average traffic count
monthly_traffic <- data %>%
  group_by(month) %>%
  summarise(avg_traffic = mean(traffic_count, na.rm = TRUE))

# Plot the average traffic count by month
ggplot(monthly_traffic, aes(x = month, y = avg_traffic)) +
  geom_bar(stat = "identity", fill = "purple") +
  labs(title = "Average Traffic Count by Month",
       x = "Month",
       y = "Average Traffic Count") +
  theme_minimal()
## Peak boarding vs Alighting
# Group by 구분 and time_period to calculate average traffic_count
peak_hours <- data %>%
  group_by(구분, time_period) %>%
  summarise(avg_traffic = mean(traffic_count, na.rm = TRUE))
```

```
peak_hours


# Plot boarding and alighting traffic count by time period
ggplot(peak_hours, aes(x = time_period, y = avg_traffic, fill = 구분)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Peak Boarding and Alighting Hours",
       x = "Time Period (Hour)",
       y = "Average Traffic Count") +
  theme_minimal() +
  scale_fill_manual(values = c("승차" = "skyblue", "하차" = "salmon"))
## Traffic Pattern on Weekdays vs Weekends
# Add a new column for weekday vs weekend
data <- data %>%
  mutate(day_type = ifelse(wday(날짜) %in% c(1, 7), "Weekend",
"Weekday"))

# Group by day_type and calculate the average traffic count
day_type_traffic <- data %>%
  group_by(day_type) %>%
  summarise(avg_traffic = mean(traffic_count, na.rm = TRUE))

# Plot the average traffic for weekdays vs weekends
ggplot(day_type_traffic, aes(x = day_type, y = avg_traffic, fill =
day_type)) +
  geom_bar(stat = "identity") +
  labs(title = "Average Traffic Count: Weekdays vs Weekends",
       x = "Day Type",
       y = "Average Traffic Count") +
  theme_minimal() +
  scale_fill_manual(values = c("Weekday" = "lightblue", "Weekend" =
"orange"))
```