Google Data Analytics Capstone Project: Cyclistic Bikeshare Case Study

BY: FAZLIE LATIB

INTRODUCTION

The capstone project is the final course in the Google Data Analytics Professional Certificate where all the knowledge from 7 main courses can be applied to understand and solve real-world problems through data analysis. To complete this project, I followed the steps of the data analysis process learned previously: ask, prepare, process, analyze, share, and act. I chose R to facilitate my data analysis processes.

CASE STUDY

"You are a **junior data analyst** working in the marketing analyst team at Cyclistic, a bike-share company in Chicago. The director of marketing believes the company's future success depends on maximizing the number of annual memberships. Therefore, your team wants to understand how casual riders and annual members use Cyclistic bikes differently. From these insights, your team will design a new marketing strategy to convert casual riders into annual members. But first, Cyclistic executives must approve your recommendations, so they must be backed up with compelling data insights and professional data visualizations."

COMPANY BACKGROUND

Since its inception in 2016, Cyclistic's popularity and usefulness amongst users has led to incredulous growth. Cyclistic has grown to a fleet of 5,824 bicycles that are geotracked and locked into a network of 692 stations across Chicago. The bikes can be unlocked from one station and returned to any other station in the system anytime.

Given this business model Cyclistic's marketing strategy relies on building general awareness and appealing to broad consumer segments. One approach that helped make these things possible was the flexibility of its pricing plans: single-ride passes, full-day passes, and annual memberships. Customers who purchase single-ride or full-day passes are referred to as **casual riders**. Customers who purchase annual memberships are **Cyclistic members**. In terms of financial viability, the finance analysts at Cyclistic have concluded that **memberships are more profitable than casual riders** due to the nature of their uses.

STEP 1: ASK - define business task, identify stakeholders

The task assigned to the junior analyst is to figure out **how annual members and casual riders use Cyclistic bikes differently**.

The stakeholders:

1. **Lily Moreno**: The director of marketing. Responsible for the development of campaigns and initiatives to promote the bike-share program. These may include email, social media, and other channels.

2. **Cyclistic marketing analytics team**: A team of data analysts who are responsible for collecting, analyzing, and reporting data that helps guide Cyclistic marketing strategy.
3. **Cyclistic executive team**: Responsible for undertaking executive decisions.

STEP 2: PREPARE - collect data, identify source, credibility and limitations

For this analysis, I have chosen the data from January to December 2022. Each month data was stored in one file of CSV format which contains hundreds of thousands of rows and 13 columns. Riders' personally identifiable information has been excluded from the data for privacy purposes.

The data quality is then checked using **ROCC** criteria (**R**eliable, **O**riginal, **C**urrent, **C**omprehensive, **C**ited).
Reliable - The bike trips data has a very large sample (more than 4 millions ride data). Therefore, it is reliable.
Original - The data is collected based on riders' usage of Cyclistic bikes. There is no third party involved in collecting the data, hence, we can consider this data as original.
Current - The data is not current as it is May 2023 now. It is not up to date, however, the data is not very old.
Comprehensive - The data contains a lot of information for each ride, which is great for analysis.
Cited - The data has been made available by Motivate International Inc. at this link under this license.

STEP 3: PROCESS - cleaning data

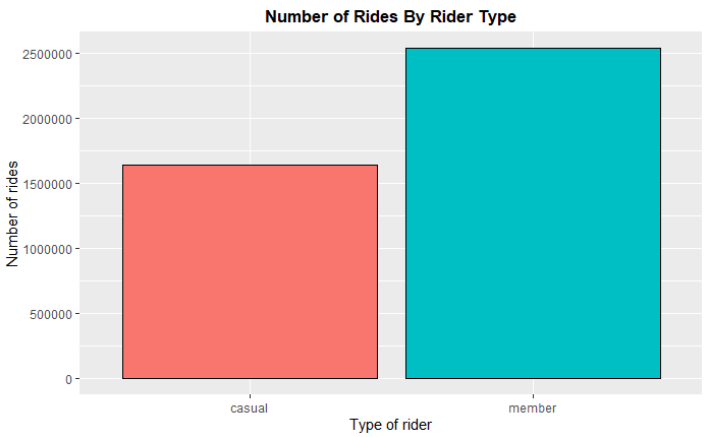
The processing of the data was done using R. The full R markdown file can be viewed here. In a real life situation, one should ask his/her supervisor or one of the primary shareholders of the project on what specific data cleaning procedures that are allowed. To keep the original "combined_trips" data, the cleaned data is inserted into a new dataframe "combined_trips_cleaned".

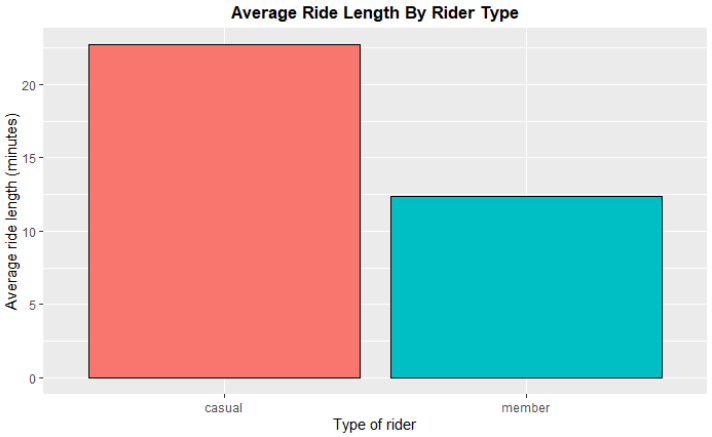
A summary of the steps involved are as follows:

1. Combined all twelve individual months into a single data frame called "combined_trips".
2. Created three new columns ("day","month" and "day_of_week) into "combined_trips" to extract the day, month, and day of the week of each ride.
3. Created a new column ("ride_length") where the length or time taken for each ride is recorded by calculating the difference between the recorded "end_time" and "start_time" of each ride.
4. Changed the "ride_length" to numeric datatype.
5. Deleted all the rows in the data frame with a null entry..
6. Deleted entries with ride length less than zero.
7. Created a new column where "started_at" and "ended_at" locations were combined to get a route (from and to) of each ride.

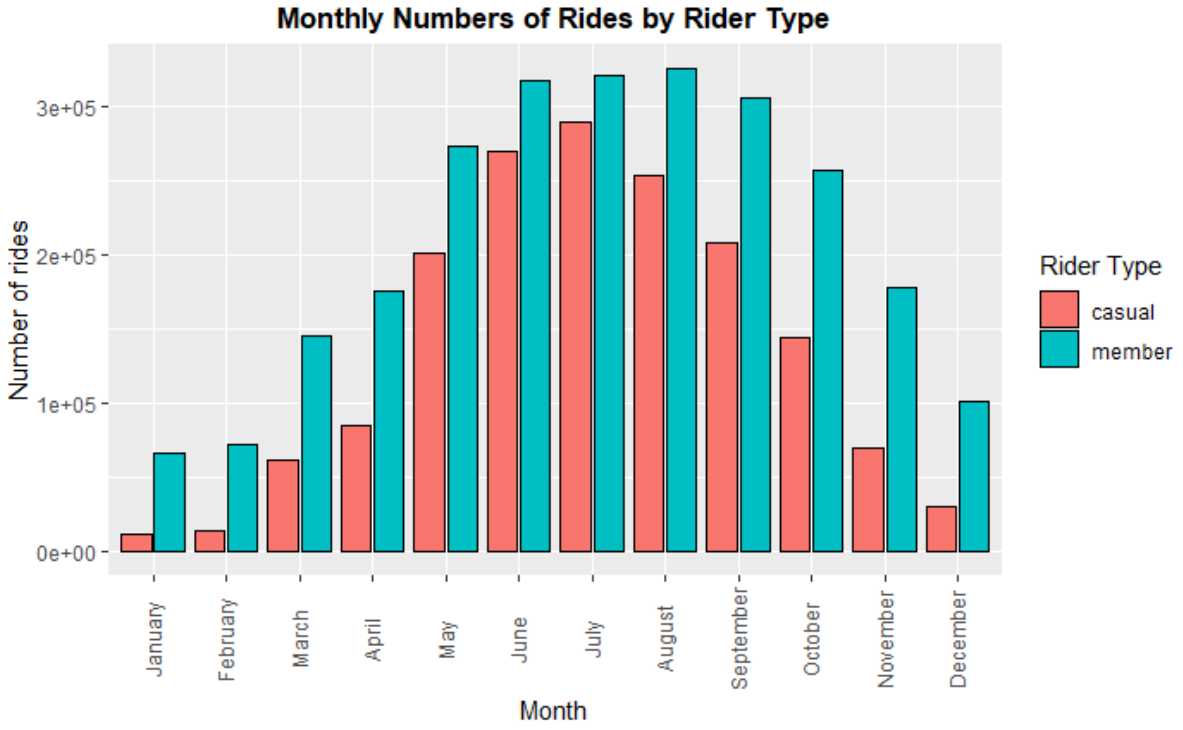
STEP 4 & 5: ANALYZE - find patterns, relationships and trends & SHARE - create visualizations

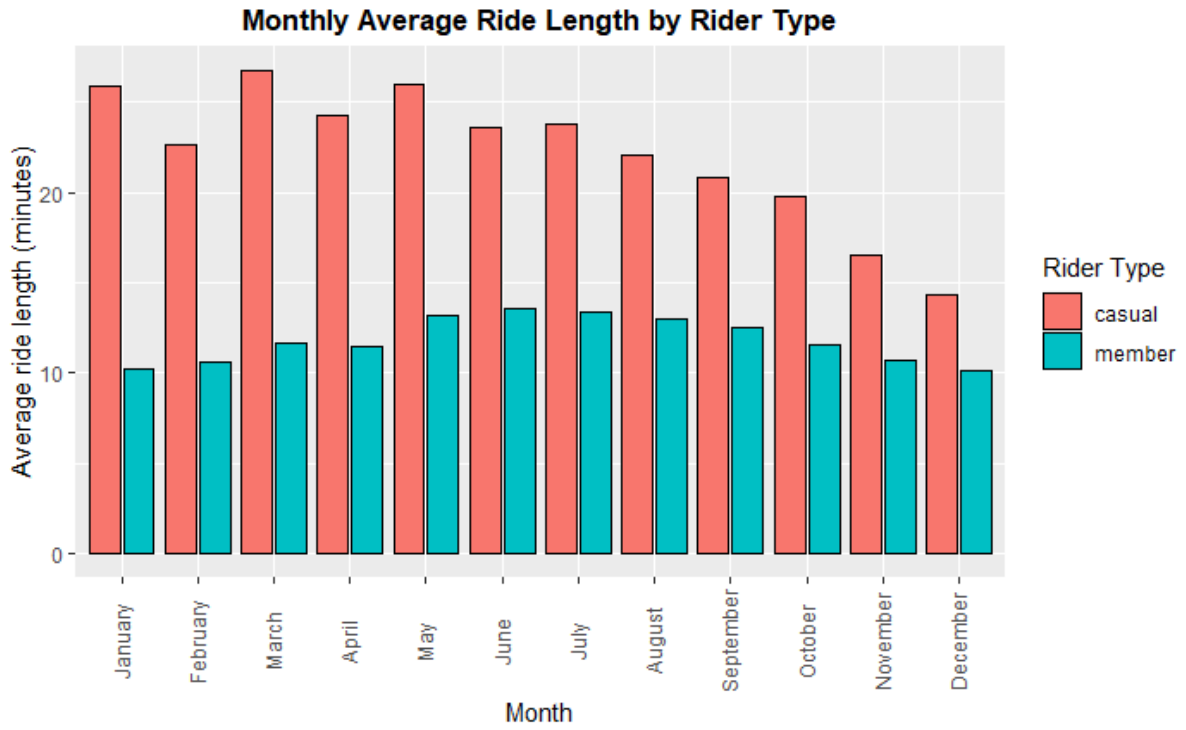
1. In 2022, there were 4168269 or 4.17 millions of rides recorded. Annual members accounted for 60.79% of those rides, however, casual riders took a longer time on each ride which averaged at 22.71 minutes which is about 9.6 minutes longer than average ride length for members.



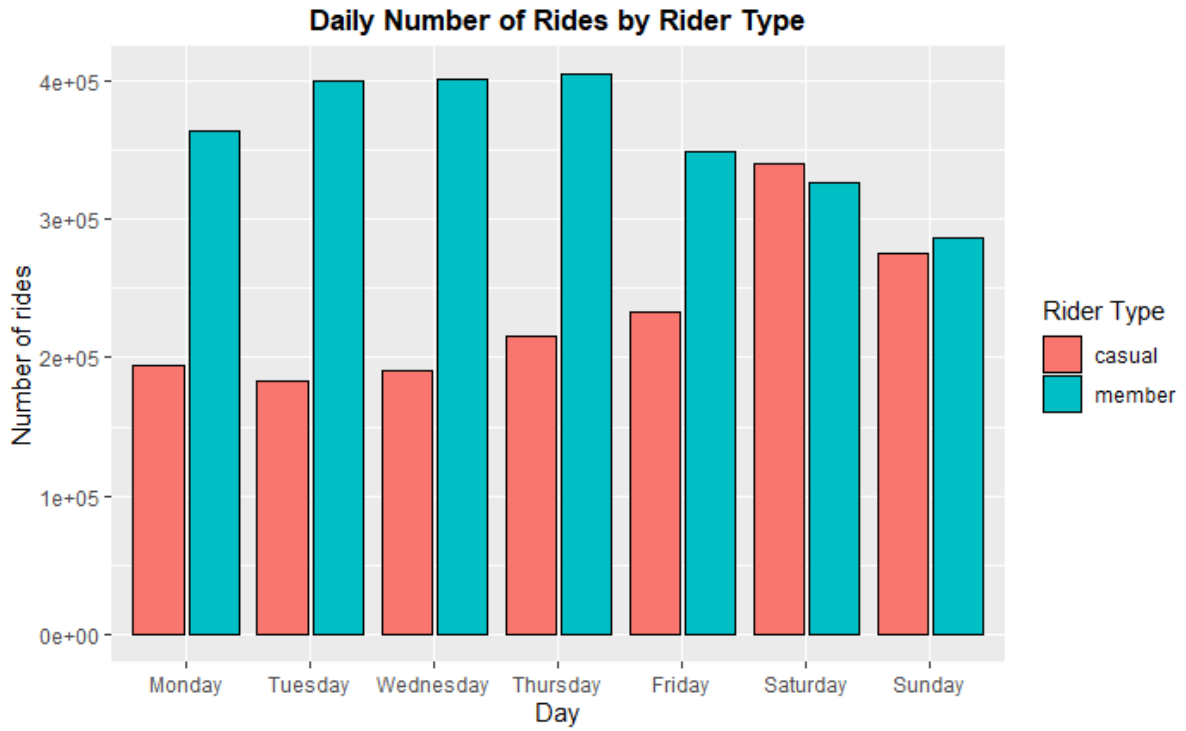


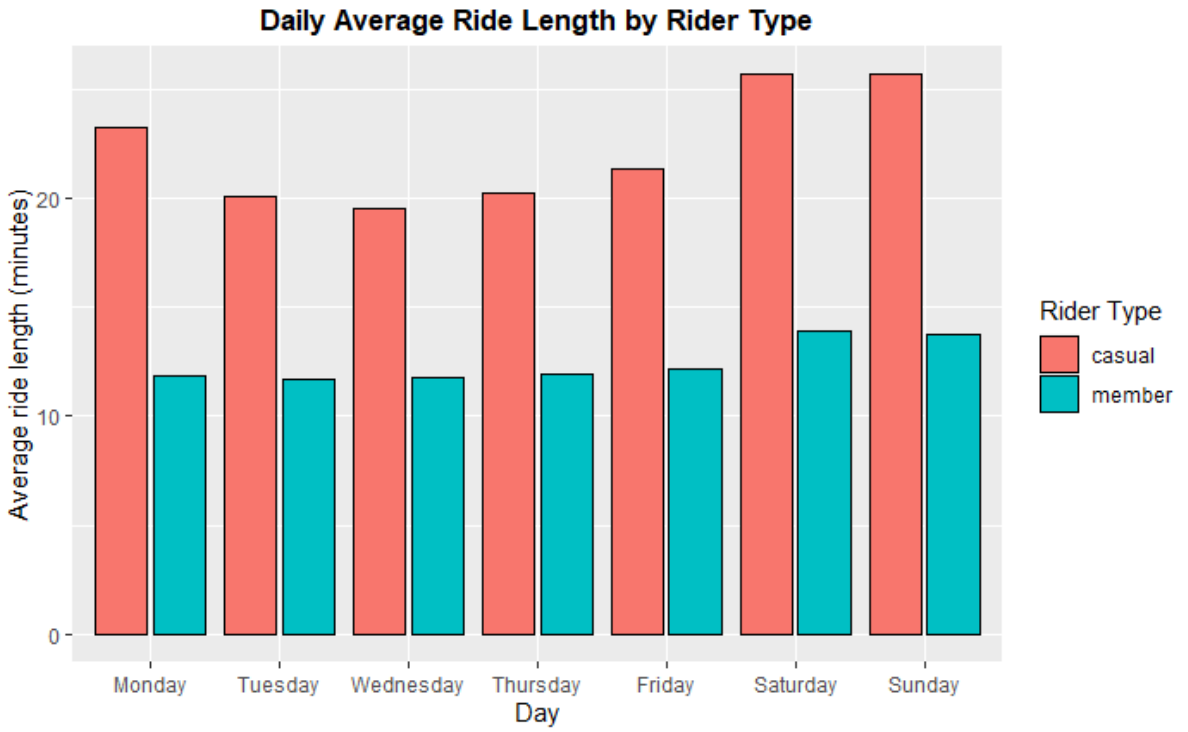
2. The number of bike rides varies throughout the year, with the highest usage during the summer months (June, July, August) and the lowest usage during the winter months (December, January, February). This pattern is likely due to the weather conditions and the seasonal availability of bikes. Annual members tend to use the bike share system more frequently than casual riders. In terms of ride length, casual riders spent more time on the bike on average in each month, where the average ride time is twice or more than members from January to April.



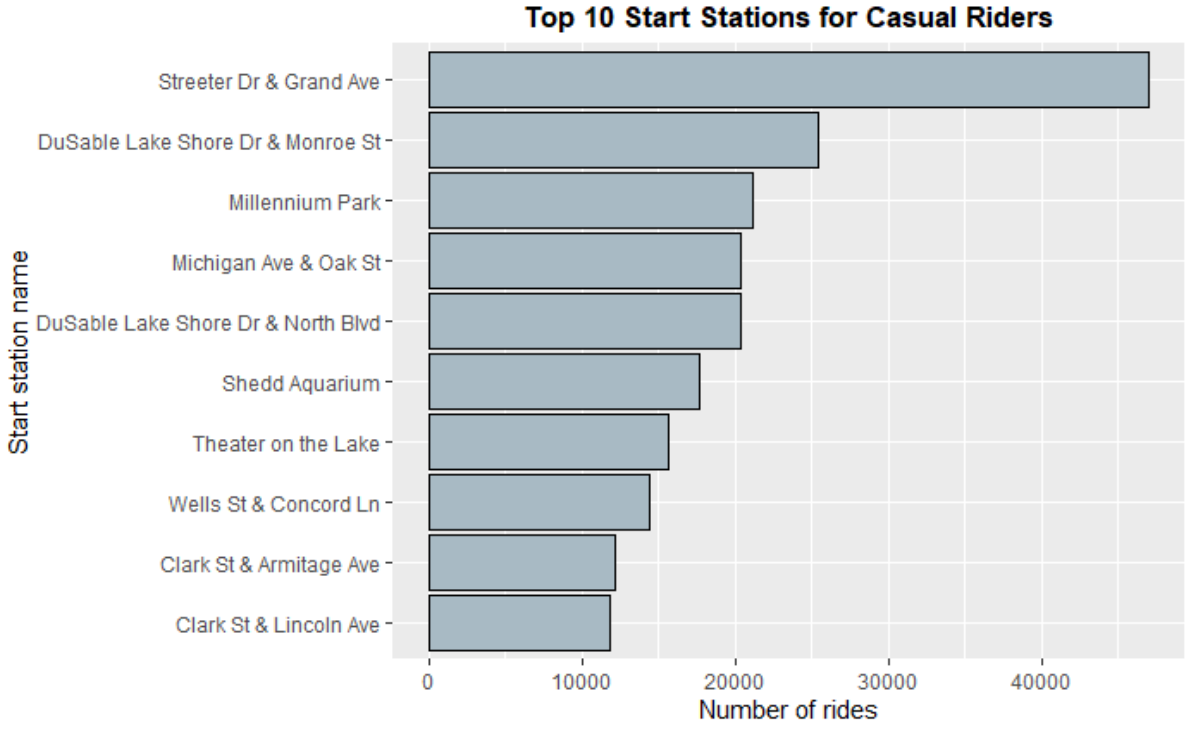


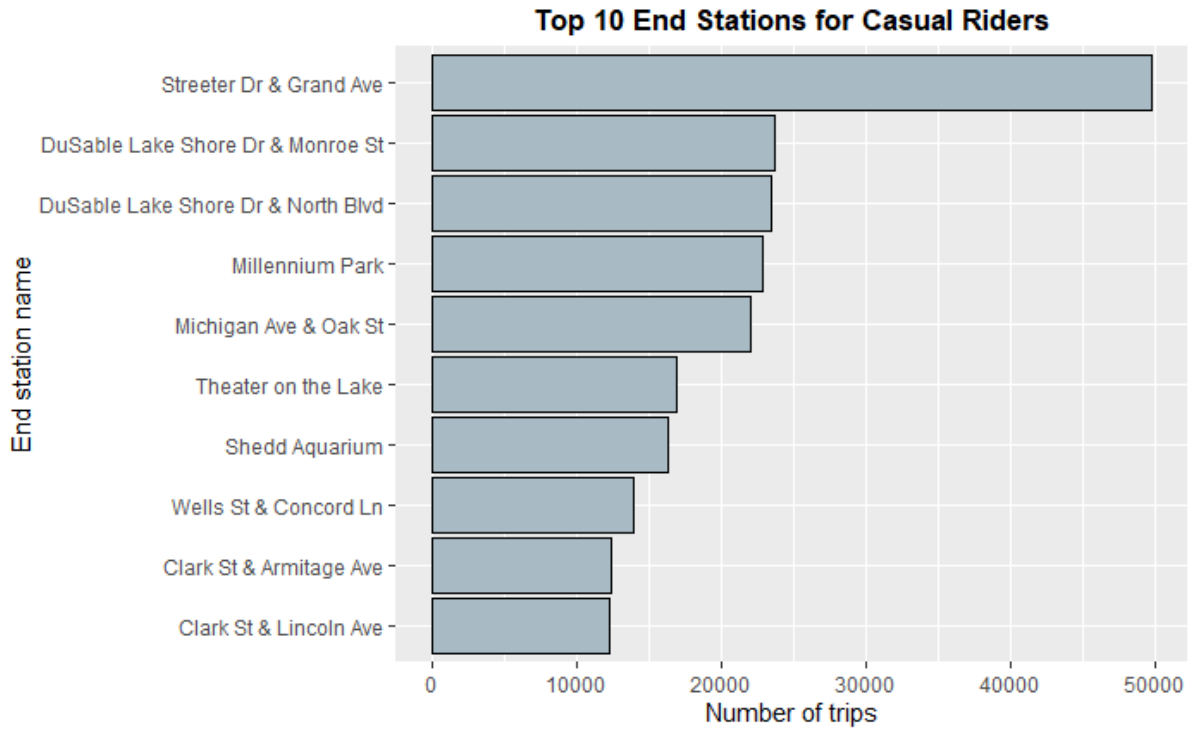
3. For annual members, the number of bike rides is consistently higher on weekdays than on weekends. This suggests that bike share systems may be more commonly used for commuting to and from work or other weekday activities by annual members. On the other hand, casual members tend to take more rides on weekends rather than weekdays. This shows that more casual members might just use the bike as a leisure activity on the weekends. Overall for each day, casual members would spend more time on the bikes than members.





4. The top 10 start and end stations for casual riders were identified.





STEP 6: ACT - recommend plans of actions

Key findings:

1. Bike share ridership in Chicago follows seasonal patterns, with peak usage in the summer months and lower usage in the winter.

2. Weekday usage is higher than weekend usage, indicating that the service is primarily used for commuting and errands.

3. Casual riders and annual members use the bikes differently, with annual members using the service more frequently while casual riders take longer rides than members.

4. The top stations are located in areas with high demand and should be prioritized for bike availability and maintenance.

Recommendation:

1. Offer promotions and incentives for annual membership sign-ups during the winter months or all year to increase usage and revenue.

2. Increase bike availability and maintenance at the top stations to improve user experience and satisfaction.

3. Develop targeted marketing campaigns to encourage casual riders to become annual members based on the benefits of longer rides and more frequent usage.

4. Monitor weather forecasts and adjust bike availability and maintenance schedules accordingly to minimize disruptions and optimize service delivery.

CONCLUSION

The bike share data analysis reveals important insights into the usage patterns and trends of bike share ridership in Chicago. By leveraging these insights, bike share providers can improve the user experience, increase annual memberships, and optimize service delivery to meet the needs of riders.