Cyclistic Case Study: How Does a Bike-Share Navigate Speedy Success

Author: Pavan Chirumalla  
Tool Used: R Programming Language  
Project Type: Google Data Analytics Professional Certificate Capstone Case Study

Ask Phase: Business Task Statement

Business Task:  
Cyclistic, a leading bike-share company in Chicago, is seeking to boost the number of annual memberships, which are proven to be more profitable than casual rides. The marketing team, led by Lily Moreno, has identified an opportunity to convert existing casual riders into long-term users.

To support this strategy, the first critical step is to analyze how annual members, and casual riders use Cyclistic bikes differently. This analysis will uncover behavioral trends, usage patterns, and engagement differences that can inform targeted marketing efforts aimed at driving membership growth.

Key Stakeholders:

* Lily Moreno (Director of Marketing) – owns campaign strategy and implementation
* Cyclistic Marketing Analytics Team – provides insights for decision-making
* Cyclistic Executive Team – evaluates and approves proposed marketing strategies
* Casual Riders – the target audience for membership conversion

Problem to Solve:

Identify the key usage differences between casual riders and annual members to build a focused marketing strategy that encourages casual users to purchase an annual membership.

Goal:

Deliver actionable insights supported by data analysis and visualizations to help Cyclistic leadership make an informed decision about launching a marketing campaign for membership conversion.

Prepare Phase: Data Source:

Data was collected from the [Divvy Bike Share System](https://divvy-tripdata.s3.amazonaws.com/index.html) for the period April 2024 to March 2025, consisting of 12 individual CSV files. Each file contains trip-level details including:

* Trip time
* Bike types
* Rider type (member vs casual)
* Station address

The data was publicly available, free of personally identifiable information (PII), and verified for accuracy and completeness. Licensing was respected per the data provider’s policy.

Data was prepared/consolidated using R programing:

* Installed and used basic packages: tidyverse, lubridate, janitor, skimr, data.table, readr.
* Loaded the 12 months trip data using read.csv
* Combined all CSVs into one dataframe using bind\_rows()
* Made sure all the data is accurate

The code:

install.packages(c("tidyverse","lubridate","janitor","skimr","data.table","readr"))

library(tidyverse)

library(lubridate)

library(janitor)

library(skimr)

library(data.table)

library(readr)

# setting the directory

setwd("C:/Users/Zedi/Downloads/Cyclistic Project/Data")

#Load Data

data\_202402 <- read.csv("C:/Users/Zedi/Downloads/Cyclistic Project/Data/202402-divvy-tripdata.csv")

data\_202403 <- read.csv("C:/Users/Zedi/Downloads/Cyclistic Project/Data/202403-divvy-tripdata.csv")

data\_202404 <- read.csv("C:/Users/Zedi/Downloads/Cyclistic Project/Data/202404-divvy-tripdata.csv")

data\_202405 <- read.csv("C:/Users/Zedi/Downloads/Cyclistic Project/Data/202405-divvy-tripdata.csv")

data\_202406 <- read.csv("C:/Users/Zedi/Downloads/Cyclistic Project/Data/202406-divvy-tripdata.csv")

data\_202407 <- read.csv("C:/Users/Zedi/Downloads/Cyclistic Project/Data/202407-divvy-tripdata.csv")

data\_202408 <- read.csv("C:/Users/Zedi/Downloads/Cyclistic Project/Data/202408-divvy-tripdata.csv")

data\_202409 <- read.csv("C:/Users/Zedi/Downloads/Cyclistic Project/Data/202409-divvy-tripdata.csv")

data\_202410 <- read.csv("C:/Users/Zedi/Downloads/Cyclistic Project/Data/202410-divvy-tripdata.csv")

data\_202411 <- read.csv("C:/Users/Zedi/Downloads/Cyclistic Project/Data/202411-divvy-tripdata.csv")

data\_202412 <- read.csv("C:/Users/Zedi/Downloads/Cyclistic Project/Data/202412-divvy-tripdata.csv")

data\_202501 <- read.csv("C:/Users/Zedi/Downloads/Cyclistic Project/Data/202501-divvy-tripdata.csv")

data\_202502 <- read.csv("C:/Users/Zedi/Downloads/Cyclistic Project/Data/202502-divvy-tripdata.csv")

data\_202503 <- read.csv("C:/Users/Zedi/Downloads/Cyclistic Project/Data/202503-divvy-tripdata.csv")

#consolidating all the trips data for last 12 months

consolidate\_trips <- bind\_rows(data\_202404,data\_202405,data\_202406,data\_202407,data\_202408,data\_202409,data\_202410,data\_202411,data\_202412,data\_202501,data\_202502,data\_202503)

Process Phase: Data Cleaning & Transformation

All cleaning and transformation were performed using R with the following steps:

* Performed data transformations in R programming for descriptive analysis
* Formatted the start and end times of the rider to show hours, minutes and seconds, which will then be used to calculate the trip time of the rider
* Created custom fields into the consolidated data frame:
  + Ride length: difference between the start time and end time in minutes
  + Day\_of\_week: Finding the week of the trip day by a rider
* This process ensured the dataset was complete, reliable, and ready for analysis

The code:

#formatting time and creating custom columns

consolidate\_trips <- consolidate\_trips %>%

mutate(started\_at = ymd\_hms(started\_at),

ended\_at = ymd\_hms(ended\_at),

ride\_length = as.numeric(difftime(ended\_at,started\_at,units = "mins")),

day\_of\_week = wday(started\_at))

Analyze Phase: To provide insights to the stake holders for making business decisions

Summarized stats, grouped by the rider type – member/casual. This is done to analyze and answer the business question to drive memberships – to find differences between casual riders and annual members and to build a focused marketing strategy that encourages casual users to purchase an annual membership.

* Finding the average trip time, Maximum trip time, and total rides grouped by rider type – member/casual
* Analyzing the trend through the week – Day Usage. Grouped the data by rider type and weekday

The code:

#aggregating the data & Analyzing the data

summary\_stats <- consolidate\_trips %>%

group\_by(member\_casual) %>%

summarise(mean\_ride = mean(ride\_length),

median\_ride = median(ride\_length),

max\_ride = max(ride\_length),

mode\_ride = mode(ride\_length),

total\_rides = n()

)

day\_usage <- consolidate\_trips %>%

group\_by(member\_casual,day\_of\_week) %>%

summarise(avg\_duration = mean(ride\_length),

ride\_count = n())

Share Phase: Visualizations and Key Findings

Two visualizations were created using ggplot2 to highlight core differences:

1. Average Ride Duration by Day of Week – distinguishes which days casual vs member rides are longest
2. Total Ride Count by Day of Week – shows volume trends by rider type and weekday

The visuals use clear labels, contrasting colors, and minimal themes to present insights to executive stakeholders in an accessible and impactful format.

The code:

#Visualization

ggplot(day\_usage,aes(x = day\_of\_week, y=avg\_duration, fill = member\_casual)) +

geom\_col(position = "dodge") +

labs(title = "Average Ride Duration By Day",

x = "Week Day", y = "Duration (mins)",

fill = "User Type") +

theme\_minimal()

ggplot(day\_usage,aes(x = day\_of\_week, y=ride\_count, fill = member\_casual)) +

geom\_col(position = "dodge") +

labs(title = "Riders Per Day",

x = "Week Day", y = "Duration (mins)",

fill = "User Type") +

theme\_minimal()

Key Insights:

1. Ride Duration:
   * Casual riders average longer ride durations compared to members
   * Members tend to take shorter, more frequent rides
2. Day of Week Patterns:
   * Casual riders show peak usage during weekends, indicating leisure behavior
   * Members show consistent usage across weekdays, suggesting commuting patterns
3. Total Ride Count:
   * Members ride more overall but casual riders dominate during weekend

Act Phase: Strategic Recommendations

Based on the analysis, the following actions are recommended:

1. Weekend Incentives:
   * Offer discounts or bundle packages for weekend riders to encourage frequent use and membership upgrades
2. Commuter Campaigns:
   * Run digital ads during weekday morning and evening commute hours targeting casual riders with cost/time savings of memberships
3. Referral/Loyalty Program:
   * Reward casual riders for frequent use or referrals with membership discounts, encouraging them to convert to long-term users

These data-backed strategies provide clear next steps for Cyclistic’s marketing team to pursue its growth goals.

Outcome

This analysis reveals actionable behavioral trends and supports the development of an evidence-based marketing strategy. The findings have been formatted for easy integration into executive presentations and portfolio materials.

Case Study for Portfolio

This project demonstrates end-to-end proficiency in:

* Data collection, cleaning, and merging
* Exploratory data analysis
* Visualization and communication of findings
* Strategic business recommendation development

Prepared as part of the Google Data Analytics Professional Certificate Program.