

**Case Study**

|  |  |
| --- | --- |
| TITLE | Cyclistic Members vs. Casual Riders: A Data-Driven Analysis |
| AUTHORED BY | Filippo Pericoli |
| DATE | 07/15/2025 |

|  |  |
| --- | --- |
| BUSINESS TASK | I analyze the company's bike riders, with a particular focus on differences in bike usage between annual members and casual riders. Understanding the distinct characteristics and behaviors of these rider groups is intended to inform a marketing campaign aimed at converting casual riders into annual members, with the goal of increasing business profitability. |
| DATA SOURCES USED | I downloaded data from <https://divvy-tripdata.s3.amazonaws.com/index.html>, covering the four quarterly datasets from Q1 2019 to Q4 2019. The data is provided in compressed .zip files, which I extracted and stored on my local hard drive. Each decompressed file is in .csv format and corresponds to a specific month. Every .csv file contains several variables describing individual bike rides, including:   * Ride identifier * Start time of the ride * End time of the ride * Bike identifier * Duration of the trip (in seconds) * Start station ID * Start station name * End station ID * End station name * Membership type (casual or member) * Gender * Year of birth of the rider |
| DATA CLEANING AND MANIPULATION | To read, clean, and analyze the data, I used and adapted the template provided in the **Google Data Analytics** course on Coursera to fit my dataset.  The main steps included in the R Markdown script are as follows:   1. **Importing the Data:** I imported into R four .csv files, each corresponding to one quarter of the year 2019. 2. **Selecting Relevant Variables:** I selected only the variables relevant to our analysis. Specifically, we are interested in the number of rides and the ride duration, both broken down by membership type. Therefore, the analysis was limited to the following variables: ride ID, start time, end time, trip duration, and membership type. Since the four imported data frames used different variable names, I had to standardize both the variable names and types. Once harmonized, I combined the datasets vertically to create a single, clean data frame containing all the company’s bike rides from 2019. 3. **Data Cleaning and Feature Engineering:** Upon inspection, I noticed that the *membership type* variable was categorized into four different values, even though there were only two actual categories. I therefore standardized these strings. Additionally, using the *start time* variable, I extracted new features such as the month, day, year, and day of the week for each ride. These variables were used in subsequent visualizations. 4. **Descriptive Analysis:** I conducted a descriptive analysis by computing key statistics (mean, median, maximum, and minimum) for ride duration. These metrics were also calculated separately for each day of the week and for each user type (annual members and casual riders). Finally, I created two main visualizations to summarize and present these statistics clearly. |
| SUPPORTING VISUALIZATIONS AND KEY FINDINGS | The main visualizations are enclosed below:  **Visualization #1**    From the visualization, we observe that on weekdays, the number of rides by **annual members** is significantly higher than those by **casual riders**. However, this difference is much less pronounced during the weekends. In fact, on Saturdays and Sundays, the number of casual riders is only slightly lower than that of members.  This pattern suggests that annual members likely use the bikes for commuting or other routine weekday activities, while casual riders tend to use the service more for leisure, particularly on weekends.  **Visualization #2**    From the visualization, we observe a significant difference in ride duration between the two categories of riders. Specifically, **casual riders** consistently have ride durations of around one hour across all days of the week, while **annual members** have much shorter rides, averaging about 15 minutes.  This supports the earlier insight that **casual riders** likely use bikes for leisure activities such as exercise, park outings, or other relaxing purposes. In contrast, **annual members** appear to use bikes for more purposeful, routine trips—most likely related to commuting or other work-related activities. |
| TOP RECOMMENDATIONS | Based on the analysis of behavioral differences between casual riders and annual members, we conclude that a targeted marketing campaign could be effective in encouraging casual riders to convert to annual memberships, thereby increasing overall profitability.  Such a campaign should focus on non-working individuals—such as students, teenagers, or retirees—who are more likely to use bikes during leisure time. For example, promotional stands could be set up in public parks or near museums on weekends, offering special discounts or incentives for becoming annual members.  The campaign could emphasize the health benefits of regular cycling and promote features like the mobile app, which simplifies the process of locating and unlocking bikes at nearby stations. These practical and motivational arguments can help position the annual membership as both a convenient and beneficial choice.  Overall, the evidence suggests that there is a significant segment of casual riders who could be converted into annual members, offering a clear opportunity to boost revenue and user engagement. |