**Google Data Analytics Certification – Case Study 1**

**Description**

Cyclistic is a bike-share company based in Chicago that offers over 5,800 bicycles and 600 docking stations. The company differentiates itself by providing not only traditional bikes but also assistive options such as reclining bikes, hand tricycles, and cargo bikes, making it an inclusive service for people with disabilities. While a significant number of users ride for leisure, about 30% use the service for daily commuting. The company offers three pricing plans: single-ride passes, full-day passes, and annual memberships. Casual riders, who purchase single-ride or full-day passes, make up a considerable portion of users, but financial analysts have determined that annual members generate more revenue. To drive long-term profitability, Cyclistic's director of marketing, Lily Moreno, wants to shift the focus toward converting casual riders into annual members. To achieve this, the marketing team needs a data-driven strategy that highlights the key differences between casual riders and annual members.

The marketing analytics team is tasked with analyzing historical bike trip data to uncover trends and insights. By identifying patterns in rider behavior, the team aims to determine why casual riders might choose a membership and how digital marketing efforts could influence their decision. The analysis will follow a structured data process, including asking key business questions, preparing and processing the data, analyzing trends, and sharing insights through professional data visualizations. The findings will help design targeted marketing campaigns that encourage casual riders to become long-term members. Before implementation, the recommendations must be reviewed and approved by Cyclistic’s executive team, known for their detail-oriented approach. If successful, this strategy could enhance customer retention, improve revenue, and solidify Cyclistic’s market position.

**Approach**

To gain the necessary insights, I will go through the six phases of the data analysis process: Ask, Prepare, Process, Analyze, Share, and Act. Each phase will help me systematically explore the data, clean and process it, and uncover meaningful trends. By following this structured approach, I can develop data-driven recommendations for the marketing strategy. Finally, I will present my findings and suggest actionable steps to convert casual riders into annual members.

**1. Ask**

Business Task

The business task is to analyze the differences in usage patterns between annual members and casual riders to develop a data-driven marketing strategy that encourages casual riders to become annual members. This includes identifying key trends in trip duration, ride frequency, and peak usage times to understand the behaviors and preferences of both user groups. The ultimate goal is to provide actionable insights that help Cyclistic increase its number of annual memberships, which are more profitable than casual rides.

Primary Stakeholders

1. Cyclistic Executive Team – Evaluates the findings and decides whether to approve and implement the proposed marketing strategies.
2. Lily Moreno (Director of Marketing) – Oversees marketing initiatives and is responsible for developing campaigns that increase membership conversions.

Secondary Stakeholders

1. Cyclistic Marketing Analytics Team – Responsible for collecting, analyzing, and interpreting data to guide the marketing strategy.

**2. Prepare**

For this analysis, I will examine Cyclistic historical trip data to identify trends and differences in how casual riders and annual members use the bike-sharing service. To ensure a comprehensive study, I have downloaded the past 12 months of trip data from January 2024 to December 2024 using the provided public link. The dataset is made available by Motivate International Inc. under a specific license, ensuring transparency and accessibility. This dataset includes crucial details such as ride duration, trip start and end times, station locations, and user type classification (casual or annual member). The data will enable me to answer the key business questions outlined in this case study by uncovering patterns in user behavior.

The dataset for this analysis consists of 12 CSV files, each representing one month of Cyclistic's trip data. The dataset contains key details about bike trips, including ride ID, rideable type (classic or electric bike), trip start and end timestamps, and station information (name and ID for both start and end stations). Additionally, it includes latitude and longitude coordinates for trip origins and destinations, allowing for geospatial analysis. A crucial column in the dataset is member\_casual, which categorizes users as either "member" (annual subscriber) or "casual" (pay-per-ride user). This structured dataset enables us to analyze riding patterns, trip durations, preferred bike types, and station usage trends, which will help identify differences in behavior between casual riders and annual members.

To ensure the credibility and integrity of the dataset, I will apply the ROCCC framework, which evaluates data based on Reliability, Originality, Comprehensiveness, Currentness, and Citation. The dataset is reliable, as it comes from a well-documented source and includes a large sample size of trip records, making it suitable for analysis. It is also original, as it was collected directly by Cyclistic, ensuring authenticity. The dataset is comprehensive, containing crucial trip details such as ride duration, timestamps, station locations, and user type (member or casual). However, due to data privacy restrictions, personally identifiable information is not included, which prevents linking casual riders to specific purchases, potentially introducing sampling bias. The dataset is current, covering the past 12 months, making it relevant for analysis. Additionally, the data is properly cited, as it has been made publicly available by Motivate International Inc. under a recognized license. While the dataset provides valuable insights, potential biases must be considered, particularly regarding casual riders who may have taken multiple trips without being individually identifiable. Despite this limitation, the data remains a strong foundation for uncovering trends and guiding Cyclistic’s marketing strategies.

**3. Process**

In this crucial process phase, I have conducted data cleaning and transformation procedures to ensure the dataset's integrity and usability for analysis. All operations were executed within the R environment, leveraging its powerful capabilities for handling large datasets, which proved to be more efficient than SQL or spreadsheet applications for this particular task.

1. Data Merging and Initial Exploration: Combined 12 monthly datasets into a single file with 5,860,568 observations and 13 variables. The dataset includes ride details like IDs, bike types, start/end times and locations, and member status, spanning from January to December 2024.
2. Duplicate Removal: Identified and removed 422 duplicate ride entries based on unique ride\_id, ensuring data integrity. The cleaned dataset now contains 5,860,357 rows.
3. Missing Value Handling: Checked for missing values in the ride\_id column (none found) and removed any rows with missing ride\_id values to maintain data completeness.
4. Data Type Conversion: Converted 'rideable\_type' from character to factor for easier analysis. Additionally, consolidated 'electric\_scooter' category into 'electric\_bike' for simplification.
5. Ride Duration Calculation: Created a new 'ride\_duration' column, calculated as the difference between ended\_at and started\_at in minutes, to facilitate analysis of ride lengths.
6. Outlier Removal: Filtered out rides with durations less than 1 minute or greater than 180 minutes to remove potential data errors and extreme outliers.
7. Time-based Column Creation: Added 'ride\_month', 'ride\_weekday', and 'ride\_hour' columns derived from the 'started\_at' timestamp and converted these new columns to factors for easier analysis.
8. Station Data Conversion: Converted start and end station names and IDs to factors, enabling categorical analysis of station-related data.
9. Distance Calculation: Added 'distance\_between\_stations\_km' column using the distGeo function from the geosphere package, calculating the geographical distance between start and end coordinates of each ride.
10. Final Data Preparation: Converted the 'member\_casual' column to a factor and saved the cleaned dataset as a new CSV file for further analysis, ensuring all necessary transformations were applied for comprehensive bike-sharing system analysis.

All these steps have been cited in “ Data Cleaning.rmd” file. we can move on to the analysis phase, where we will explore and uncover insights related to the behavior of casual members compared to Cyclistic annual members. By examining the cleaned data, we aim to identify patterns, trends, and key differences that can provide valuable information for understanding how each group interacts with the service, ultimately guiding marketing and product strategies.\

**4. Analyse**

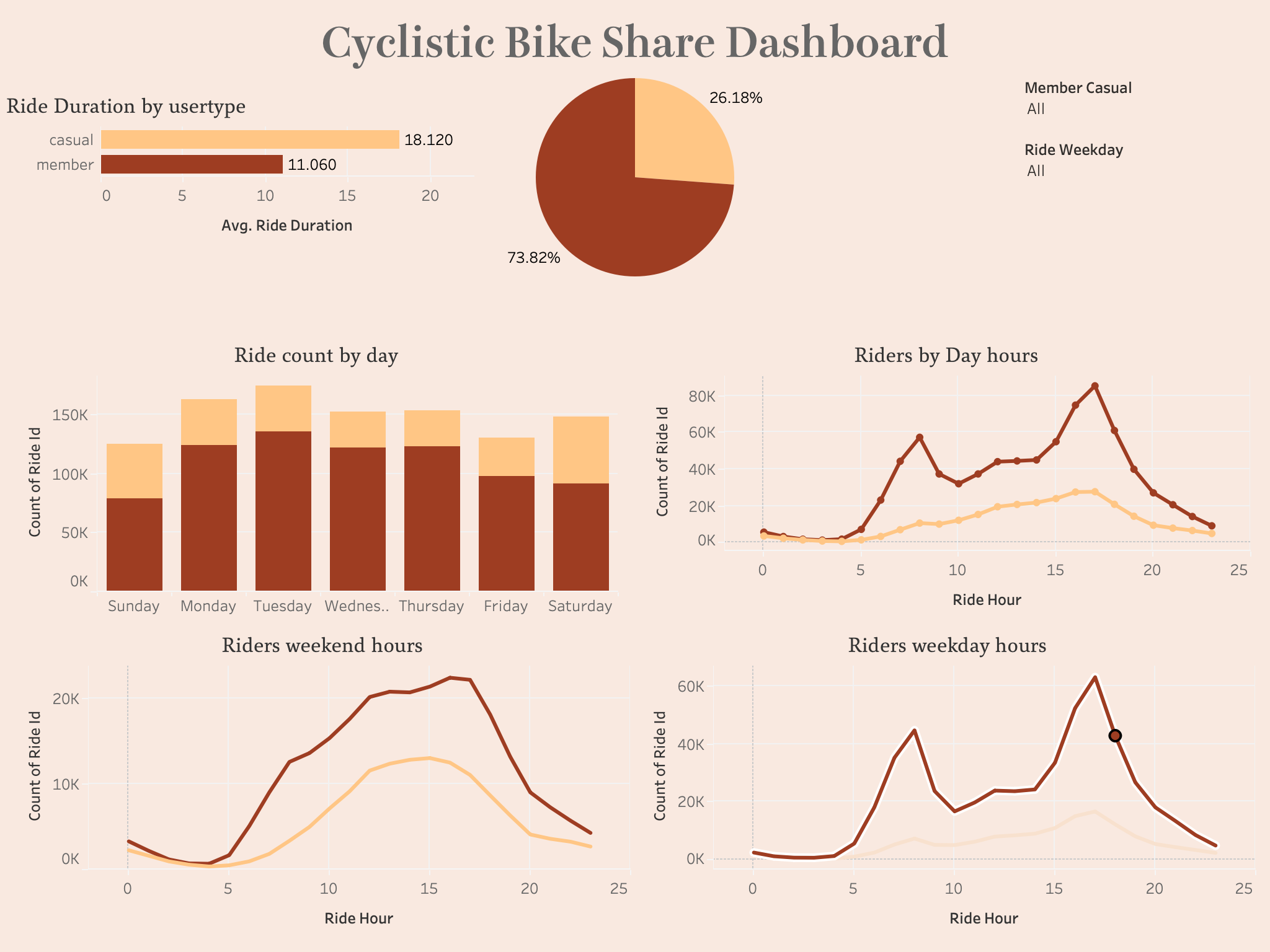
I conducted an in-depth data analysis on the cleaned dataset to gain valuable insights into the behavior of Cyclistic’s casual and annual members. Using R, I applied various analytical techniques to explore the data and identify significant patterns. Additionally, I created visualizations to effectively communicate the findings, helping to highlight key trends and differences between the two member groups. These visualizations provided a clear, visual representation of the insights, allowing for a deeper understanding of user behaviors and preferences.

1. Distribution Analysis: Annual members account for 63.7% of total rides, while casual riders make up 36.3%, indicating a higher usage by members.
2. Monthly Ride Patterns: Both member and casual rides peak during summer months and are lowest in winter, with members consistently taking more rides throughout the year.
3. Weekly Ride Patterns: Member ridership is highest on Wednesdays, while casual ridership peaks on weekends, suggesting different usage patterns between the two groups.
4. Daily Ride Patterns (Weekdays): Both groups show peak ridership in late afternoon/early evening, with members having an additional morning peak, indicating commute patterns.
5. Daily Ride Patterns (Weekends): Member ridership peaks significantly around 5-6 PM on weekends, while casual ridership shows a less pronounced afternoon peak.
6. Ride Duration Analysis: Casual riders consistently have longer average ride durations throughout the week, with both groups having longer durations on weekends.
7. Hourly Ride Duration for Casual Riders: Casual riders have the shortest rides in early morning hours and longest rides in late morning/early afternoon, suggesting leisure or tourism-related usage.
8. Popular Stations for Casual Riders: Analysis of top 10 stations most frequently used by casual riders provides insights into preferred starting locations for this user group.

I gained valuable insights and identified key trends from the data. All the analysis has been cited in the “Data analysis.rmd” file. These findings provided a deeper understanding of the behavior of casual and annual members. Now, the next step is to present these insights effectively through visualizations, which will highlight the trends and patterns in a clear and engaging way. The visualizations will help to convey the key takeaways from the data, making it easier to communicate the findings and inform decision-making.

**5. Share**

The visualizations of the insights gained from the analysis phase are presented to stakeholders through interactive dashboard in Tableau. This dashboard effectively showcases key trends and comparisons between casual and annual members, making the data more accessible and actionable. By leveraging Tableau’s dynamic features, stakeholders can explore the findings in a clear and engaging manner, aiding data-driven decision-making.



The dashboard reveals that casual riders have a significantly higher average ride duration (18.120) compared to members (11.060), suggesting different usage patterns. Members account for the majority (73.82%) of all rides, while casual riders represent 26.18% of the total ridership. Ride counts remain relatively consistent across weekdays, indicating stable demand throughout the week. Examining ridership by hour reveals peak activity in the late afternoon/early evening (around 5 PM), likely due to commuters and leisure riders. On weekends ridership peaks at 3 pm, On weekdays the highest ridership during mornings is at 8 AM and the evening it is 5 pm. These insights into ride duration, rider type distribution, and hourly ridership trends enable Cyclistic to tailor marketing strategies and optimize resource allocation to better serve both member and casual rider segments.

**6. Act**

Data-Driven Recommendations

Based on our analysis of Cyclistic’s casual and annual members, we propose the following recommendations to convert casual riders into annual members and enhance user engagement:

1. Target Commuters with Membership Benefits: Our analysis reveals that both casual and annual members experience peak ride activity during heavy workweek commute hours. Casual riders who frequently use Cyclistic bikes for commuting should be targeted with a campaign emphasizing the cost-effectiveness of a membership compared to single rides or day passes. Additionally, since casual riders prefer e-bikes during these times, marketing efforts should highlight the lower e-bike rates available to members. To maximize impact, ad placements should be strategically positioned at stations and locations with high commuter traffic.
2. Engage Leisure Riders with Seasonal Promotions: Casual riders often use Cyclistic bikes for leisure and exercise, primarily during weekends and favorable weather conditions. To capitalize on this trend, a targeted marketing campaign should be launched at the beginning of spring when riding conditions improve. This campaign should position Cyclistic as a key player in promoting an active and healthy lifestyle. To effectively reach this audience, a combination of on-the-ground marketing and social media advertisements should be used, focusing on Chicago residents who ride for recreation.
3. Promote Cost Savings for Longer Rides: Casual riders tend to take longer trips, making it essential to highlight the financial advantages of a membership. Specifically, the campaign should promote the lower per-minute rates for members and the 45 minutes of free ride time for classic bikes. These cost-saving benefits can serve as a strong incentive for casual riders to switch to an annual membership. To maximize engagement, this strategy should be promoted through advertisements at popular casual riding areas, such as downtown and along Lake Michigan, as well as through targeted social media campaigns.

By implementing these strategies, Cyclistic can effectively convert casual riders into loyal annual members while enhancing the overall user experience.