

Google Data Analytics Capstone Cyclistic Bike Data Report

Tools Used:

SQL Server, Python, Docker, Git, Excel

Link to Project Files:

([Github](#))

Business Task

The primary business task in this case study is to understand and articulate the differences in usage patterns between annual members and casual riders of Cyclistic, a Chicago-based bike-share company. This analysis is crucial for developing targeted marketing strategies aimed at converting casual riders into annual members, a key objective for the company's future growth and profitability. By examining various aspects of bike usage, such as frequency, duration, routes, time of day, and types of bikes used, the goal is to identify distinct behaviors and preferences between these two groups of riders. This understanding will inform the creation of tailored marketing initiatives that address the specific needs and motivations of casual riders, encouraging them to become annual members.

Data Sources/ Tools Used

In this analysis, we utilized user data from Cyclistic, sourced from the publicly available Divvy Trip dataset for the period of December 2022 to December 2023. This dataset, accessible at [[Divvy Trip Data](#)], provides a comprehensive view of bike-share usage patterns. It includes several key variables: 'Ride_id', a unique identifier for each trip; 'Rideable_type', specifying the type of bike used; 'started_at' and 'ended_at', indicating the start and end times of each ride; 'start_station_name' and 'end_station_name', along with their respective station IDs ('start_station_id', 'end_station_id'), detailing the locations for the beginning and end of each trip. Additionally, the dataset encompasses geographical coordinates ('start_lat', 'start_lng', 'end_lat', 'end_lng') for each station, offering insights into geographical patterns of bike usage. Crucially, it differentiates between 'member' and 'casual' riders under the 'member_casual' field, allowing for a focused analysis of the differing usage patterns between these two categories of Cyclistic's users.

Cleaning/Manipulation of Data

In the process of preparing the dataset for analysis, a crucial step involved data cleaning and manipulation to ensure accuracy and reliability of the results. After aggregating the

collected data from December 2022 to December 2023 into a singular, large CSV file, a thorough examination was conducted to identify any inconsistencies or missing values. It was observed that certain rows contained null or corrupted data points. Given the structured nature of the dataset, where each row represented a complete record of an individual trip, the integrity of the entire row was compromised if even one value was missing or incorrect. To maintain the quality and consistency of the dataset, rows with such flawed data were entirely removed from the analysis. This step was essential to prevent any skewness or inaccuracies in subsequent analyses, ensuring that the insights derived from the dataset are based on complete and reliable data. This approach aligns with standard data cleaning practices, prioritizing the accuracy and reliability of the dataset for meaningful analysis.

Analysis

In this analysis of Cyclistic's bike-share data, we uncovered several key insights about the behavior of both members and casual riders. Our analysis was focused on different aspects, including bike type preferences, usage patterns, and ride characteristics.

The analysis of the bike-share data reveals intriguing insights into the ride lengths of both member and casual riders. Notably, the distribution of ride lengths for both categories displays a significant concentration around the 2000-2500 meter mark. This similarity in ride distances suggests that, regardless of user type, there is a common range within which most of the bike journeys fall.

The data revealed that members and casual riders exhibit distinct preferences for the type of bike they use. Members tend to prefer classic bikes slightly more (54.86%) than electric bikes (45.14%). In contrast, casual riders showed a slight inclination towards electric bikes (51.69%) over classic bikes (48.31%). These percentages indicate a relatively balanced distribution of preferences for both groups, albeit with a slight leaning towards specific types of bikes depending on the user category.

Regarding spatial usage patterns, we observed that the heatmaps for start locations of both member and casual users were remarkably similar. This suggests that both groups of riders tend to start their journeys from similar areas, indicating shared popular locations or perhaps similar needs in terms of starting points.

In terms of temporal usage patterns, we noted distinct differences in the highest use times for members and casual riders. Members showed the highest usage during weekdays, specifically on Fridays, Thursdays, and Wednesdays, and during the winter months of December, January, February, and March. On the other hand, casual riders were more active during weekends, with the highest usage on Saturdays, Sundays, and Mondays, and in the warmer months of July and April. This contrast in usage patterns may reflect different motivations and needs between the two groups, such as members using bikes for regular

commutes and casual riders using them more for leisure activities during weekends and warmer weather.

Additionally, our analysis of ride length revealed significant differences between the two user groups. The average ride length for members was approximately 11.89 minutes, with a median of 8 minutes and a mode of 5 minutes. In contrast, casual riders had an average ride length of 27.47 minutes, a median of 11 minutes, and a mode of 6 minutes. These figures highlight that casual riders tend to have longer ride durations compared to members, which could be indicative of different usage behaviors like exploration or leisurely rides, as opposed to the more routine or utilitarian use by members.

Overall, these insights provide a comprehensive understanding of Cyclistic's user behaviors and preferences, which can be instrumental in shaping targeted marketing strategies and operational decisions to cater to the distinct needs of each user group.

Recommendations Based on Analysis

Based on the comprehensive analysis of Cyclistic's bike-share data, we have derived several actionable recommendations that could significantly enhance the company's marketing strategy and user engagement.

Firstly, our analysis underscores the importance of focused physical advertising in areas most frequented by Cyclistic users. The heatmaps have highlighted the Chicago downtown area, stretching from Evanston in the north to Blue Island in the west and Bridgeview in the south, as key hotspots for both member and casual riders. Capitalizing on these areas with targeted advertising campaigns can greatly increase visibility and potentially drive up membership sign-ups. Offering special discounts and promotions for new members in these high-traffic areas can incentivize trial and adoption of the service.

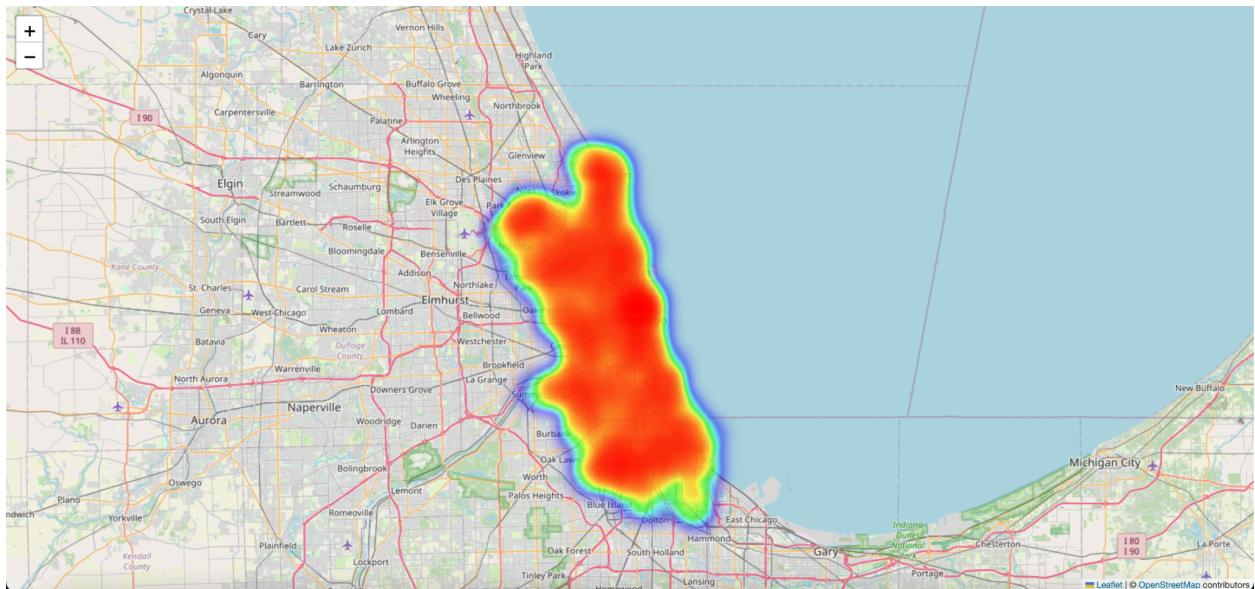


Figure 1, Casual Start Location Heatmap ([linked here](#))

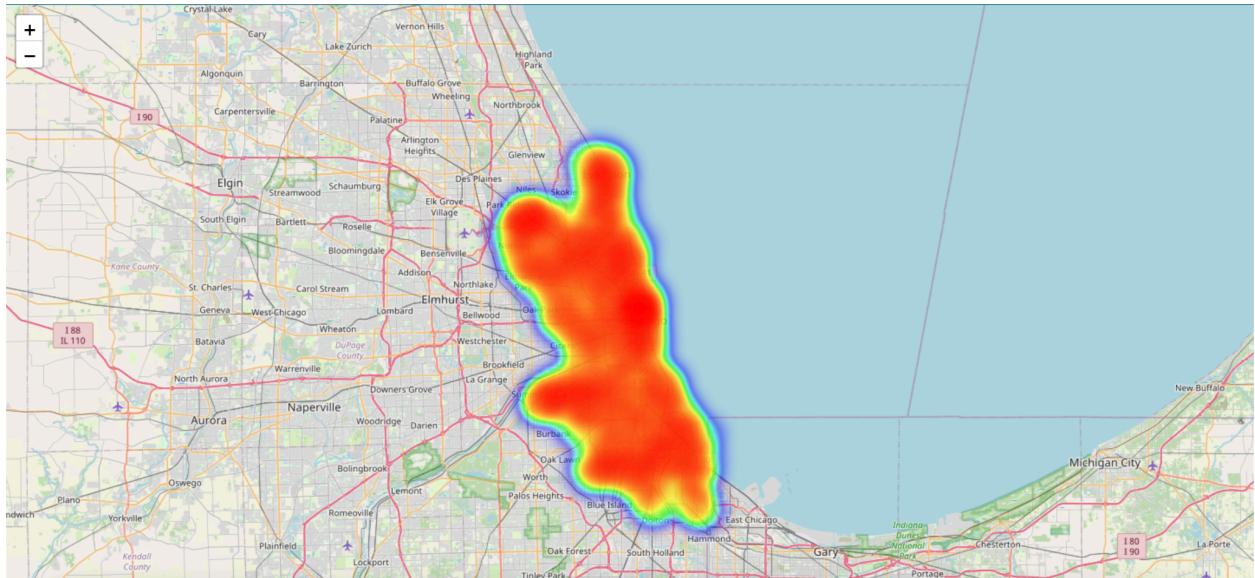


Figure 2, Member Start Location Heatmap ([linked here](#))

Secondly, we recommend deploying targeted physical advertising efforts towards casual riders during the months they are most active. The data shows that casual ridership peaks during the weekends and warmer months, particularly in July and April. Tailoring marketing campaigns and special offers during these months can attract more casual riders. This approach can include outdoor ads in popular recreational spots or events that align with the casual riders' increased activity during these warmer months. Additionally, promotions like discounted day passes or special weekend rates could be effective in converting casual riders into regular users or members.

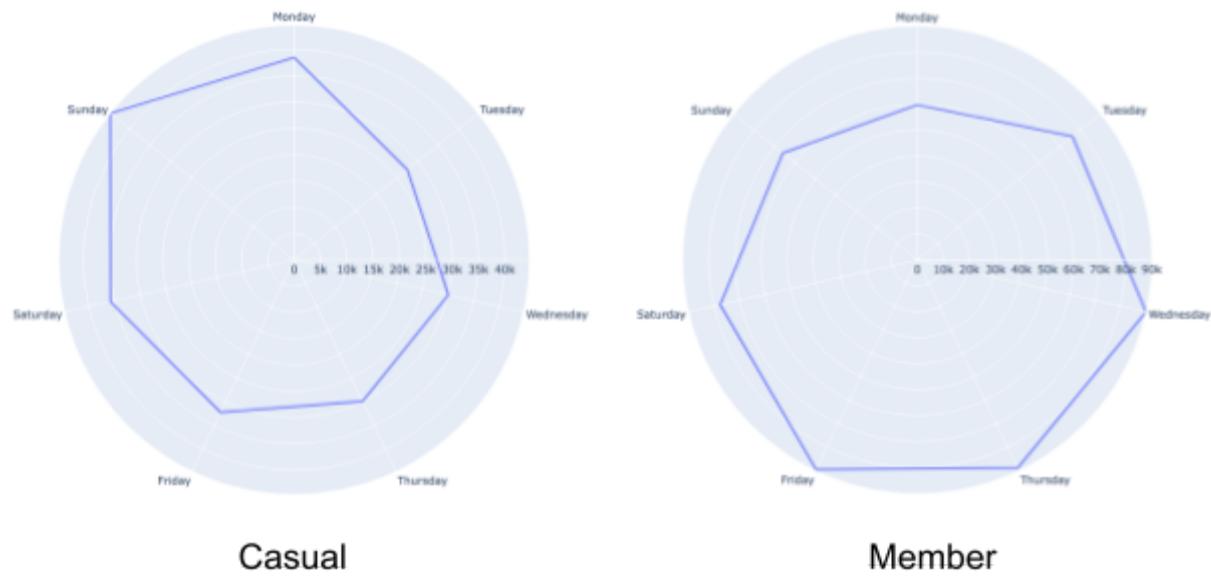


Figure 3: Member vs Casual Days of Week Counts Spider Plots

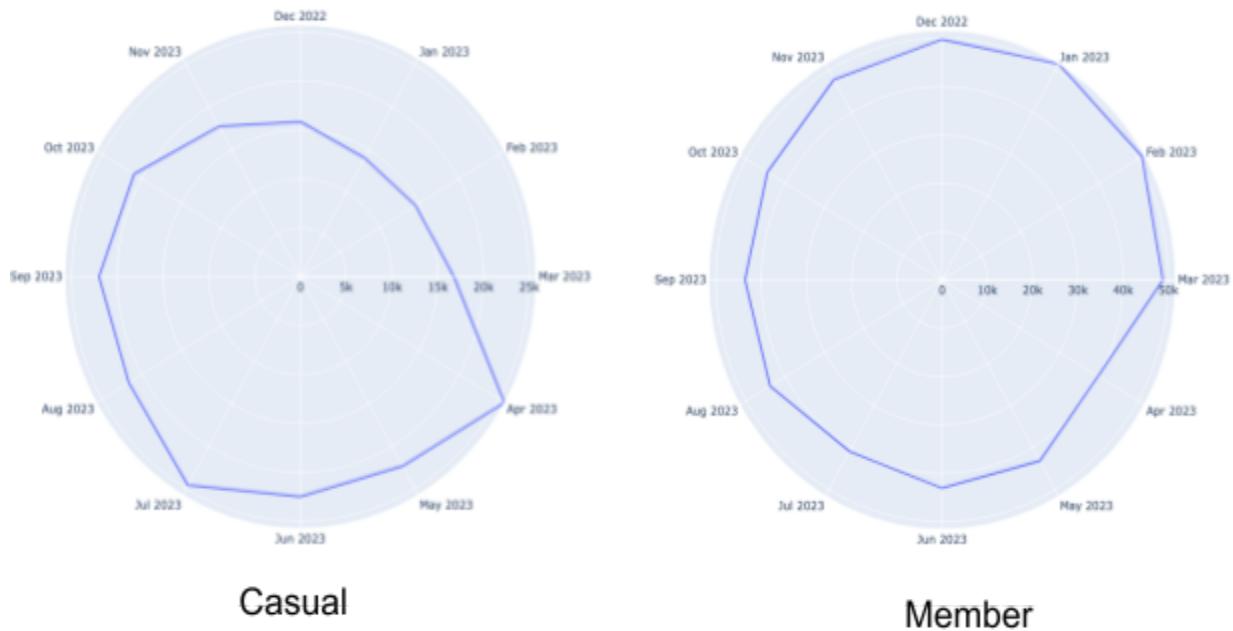


Figure 4: Member vs Casual Days of Month Counts Spider Plots

Lastly, considering the differences in ride lengths and preferences between member and casual riders, there is an opportunity to customize the riding experience to better cater to each group's needs. Members, who predominantly use the service for shorter, more routine

commutes, might benefit from enhanced membership plans with perks like guaranteed bike availability during rush hours or reward points for frequent usage. Conversely, for casual riders who tend to prefer longer, leisurely rides, introducing scenic route recommendations, partnered discounts at local attractions, or special leisure-oriented packages could enhance their experience. This targeted approach, respecting each user group's distinct behavior, can not only improve user satisfaction but also encourage casual riders to consider regular memberships.

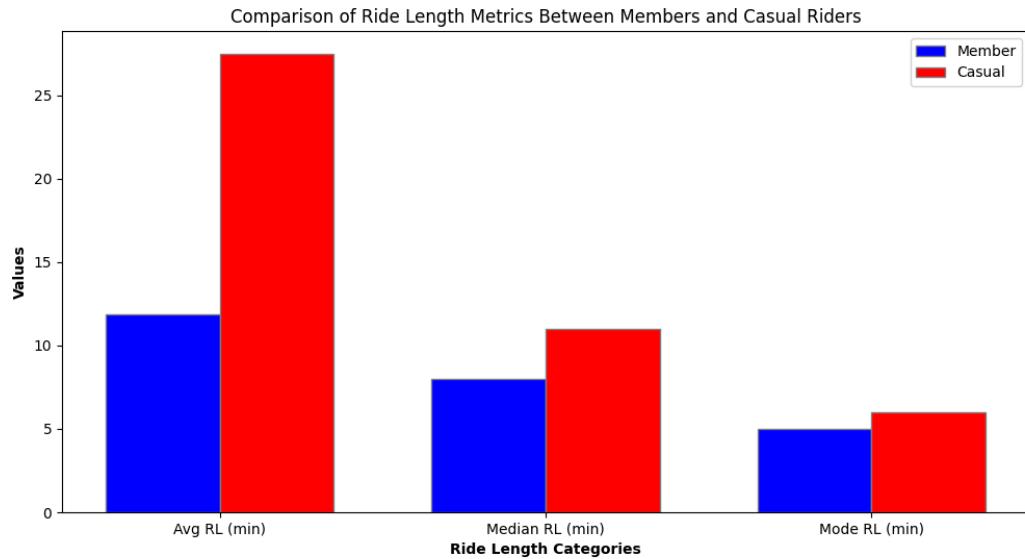


Figure 6: Comparison of Ride Length Metrics Between Members and Casual Riders

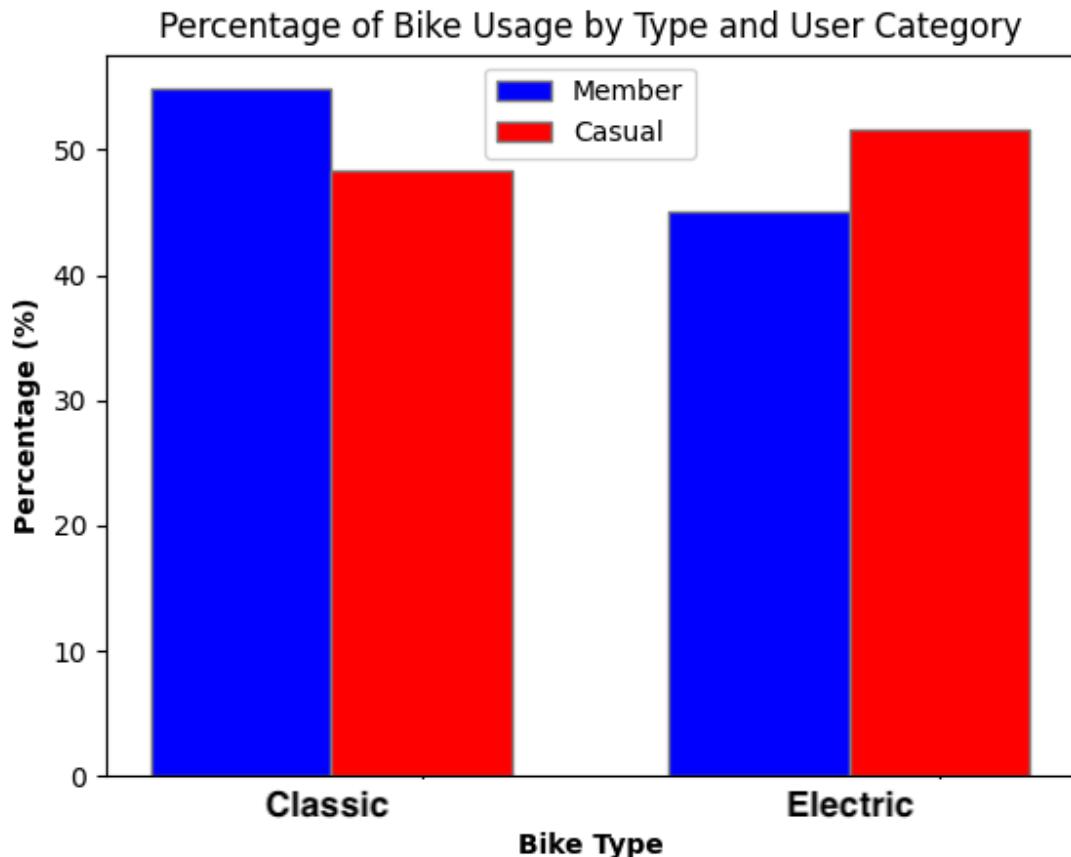


Figure 5: Percentage Of Bike Usage by Type and User Category Graph

Implementing these recommendations requires a strategic approach, ensuring that each initiative aligns with Cyclistic's overall business goals and enhances the user experience for both members and casual riders. Through focused marketing and tailored services, Cyclistic can leverage its insights to strengthen its market position and foster growth.

Conclusions and Further Thoughts

The analysis of Cyclistic's bike-share data offers a multifaceted view of user behavior and preferences, highlighting significant opportunities for targeted marketing and service enhancement. The convergence in ride lengths among members and casual riders, coupled with distinct geographic and temporal usage patterns, provides a strategic roadmap for the business. The insights gained can guide the team in refining the marketing approach, optimizing service offerings, and enhancing overall user experience.

Applying these insights, the team should focus on geo-targeted marketing strategies in high-density areas like the Chicago downtown region, utilizing the identified hotspots to maximize membership conversion. The alignment of marketing campaigns with peak usage times—weekdays for members and weekends for casual users—can further enhance the effectiveness of promotional efforts. Additionally, considering the uniformity in ride distances, service offerings can be tailored to support these common journey lengths, such as strategically placing bike stations and ensuring bike availability. The next steps would involve implementing these strategies, closely monitoring their impact, and iteratively refining the approach based on user feedback and data analysis.

To expand on these findings, additional data could be immensely valuable. For instance, surveying users to understand their motivations, satisfaction levels, and barriers to membership conversion could offer deeper insights into user needs and preferences. Data on weather patterns could also be analyzed to understand its impact on bike usage. Moreover, integrating city infrastructure and public transportation data could aid in optimizing station locations and enhancing service accessibility. Continuously augmenting the data analysis with qualitative insights and external datasets will enable Cyclistic to stay attuned to its users' evolving needs and maintain a competitive edge in the bike-share market.