

Case Study

“How does a bike-share navigate speedy success?”



Scenario

You are a junior data analyst working on 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.

Phases

The phases or processes involved in this analysis are:

- Ask
- Prepare
- Process
- Analyze
- Share
- Act

Ask

Guiding questions

- **What is the problem you are trying to solve?**

“How do annual members and casual riders use Cyclistic bikes differently?”

- **How can your insights drive business decisions?**

Insights drawn by analyzing historical bike trip data can be used to explore trends and patterns in order to design marketing strategies aimed at converting casual bike riders to annual members.

Key Tasks

- **Identify the business task**

Convert casual riders into annual members since, annual members are much more profitable than casual riders.

- **Consider key stakeholders**

- **Lily Moreno:** The director of marketing and your manager.
- **Cyclistic marketing analytics team:** A team of data analysts who are responsible for collecting, analyzing, and reporting data that helps guide Cyclistic marketing strategy.
- **Cyclistic executive team:** The notoriously detail-oriented executive team will decide whether to approve the recommended marketing program.

Deliverable

- **A clear statement of the business task**

Explore 12 month bike trip data to analyze patterns which can contribute to determine and study factors affecting subscription purchasing behavior of riders which could be utilized to convert casual riders into members. This comes from the assumption of financial analysts that annual members are much more profitable than casual riders.

Prepare

Guiding questions

- **Where is data located?**

Data is located in a bucket of object storage service, S3 from Amazon.

- **How is the data organized?**

Data consists of .csv files for each month in each year from 2020 to 2024 (upto May), present in multiple zip folders. Each file consist of columns like ride_id, rideable_type, started_at, ended_at, start_station_name, start_station_id, end_station_name, end_station_id, start_lat, start_lng, end_lat, end_lng, member_casual.

- **Are there issues with bias or credibility in this data? Does this data ROCCC?**

ROCCC stands for Reliable, Original, Comprehensive, Cited, Current. By exploring the dataset columns, it can be stated that this dataset is reliable for seeking the differences in casual riders and members.

- **How are you addressing licensing, privacy, security, and accessibility?**

The data is sourced from an already existing trusted organization which provides licensed open source datasets for analysis work. For ensuring privacy, security and accessibility, its editing and viewing privileges are only shared among the data analyst team for referencing and feedback.

- **Are there any problems with the data?**

Yes, the data contains several missing values, which could be challenging to fill without more information about the data.

Deliverable

- **A description of all data sources used**

Only one data source is used, which is divytrips. They stored the data in s3 aws buckets, comprising multiple zip folders.

Process

Guiding questions

- **What tools are you choosing and why?**
 - **SQL:** Used for merging, cleaning, and transforming the data due to its powerful querying capabilities and ability to handle large datasets efficiently.
 - **Power BI:** Chosen for its robust data visualization capabilities, allowing me to create detailed and interactive dashboards that aid in deriving insights effectively.
 - **Excel:** Used pivot tables to effectively analyze data from separate months over a whole year.
- **Have you ensured your data's integrity?**
 - By merging monthly data tables into a single table, I ensured consistency and completeness of the data.
 - I backed up the original data before making any modifications, preserving the raw data for future reference or audits.
 - I handled null values and inconsistencies systematically to maintain the accuracy and reliability of the data.
- **What steps have you taken to ensure that your data is clean?**
 - Merged all monthly datasets into a single table using UNION ALL to facilitate comprehensive analysis.
 - Checked for and addressed null values in critical columns like start_station_name and end_station_name by updating them with dummy values to avoid data loss and bias.
 - Added and populated new columns (start_date, start_time, end_date, end_time) for ease of analysis.
 - Filled missing start_date values using the end_date where applicable, assuming most rides end on the same day.
 - Calculated average ride duration and used it to fill missing start_time values.

- **Have you documented your cleaning process so you can review and share those results?**
 - Detailed the SQL queries used for data cleaning and transformations.
 - Explained each step taken to clean and prepare the data, ensuring transparency and reproducibility.

Key Tasks

- **Check the data for errors.**
 - Identified and counted null values in key columns.
 - Reviewed the distribution of data across critical fields to spot any anomalies.
- **Choose your tools.**
 - Selected SQL for data merging, cleaning, and transformation due to its efficiency in handling large datasets.
 - Chose Power BI for its advanced visualization capabilities, enabling clear and interactive data presentation.
- **Transform the data so you can work with it effectively.**
 - Merged monthly datasets into a single table.
 - Addressed null values by assigning dummy values where necessary.
 - Added new columns for better data readability and future analysis.
 - Filled missing values using logical assumptions and calculations.
- **Document the cleaning process.**
 - Detailed the SQL queries and steps used in the data cleaning process.
 - Described the rationale behind each transformation and cleaning step.

Deliverable

- **Documentation of any cleaning or manipulation of data**
 - A copy of the cleaning log can be found [here](#).

Analyze

Guiding questions

- **How should you organize your data to perform analysis on it?**

Data should be organized by aggregating it into meaningful segments such as user type (casual riders vs. premium members), rideable type (classic bike, electric bike, docked bike), time of day, day of week, and month. This helps in comparing different segments and identifying trends.

- **Has your data been properly formatted?**

Yes, the data has been properly formatted. Missing values have been handled, new columns have been added for better readability, and additional columns have been derived using DAX in Power BI to facilitate detailed analysis.

- **What surprises did you discover in the data?**

- Casual riders have a higher average ride duration (10.64 minutes) compared to premium members (8.28 minutes).
- Both casual riders and premium members have their peak riding hour at 5 PM.
- Casual riders prefer electric bikes, while premium members prefer classic bikes.
- Docked bikes are rarely used and only by casual riders.

- **What trends or relationships did you find in the data?**

- Ride Duration: Casual riders tend to have longer ride durations than premium members.
- Peak Hours: Both user groups have peak riding activity at 5 PM, but premium members have the lowest activity at 3 PM.
- Day Preference: Both groups prefer riding on weekends (Saturday and Sunday).
- Bike Type Preference: Casual riders prefer electric bikes, while premium members prefer classic bikes.

- Seasonal Trends: July is the most preferred month for casual riders, while August is the most preferred month for premium members.
- Station Usage: Specific stations are highly preferred by each group, indicating potential hotspots for targeted promotions.

- **How will these insights help answer your business questions?**

The insights from the 12-month bike trip data reveal key factors that can inform strategies to convert casual riders into premium members. Casual riders have longer ride durations, prefer electric bikes, and ride mostly on weekends and during peak hours at 5 PM, suggesting recreational use. Targeted promotions during these times and at popular stations like Streeter Dr & Grand Ave can highlight the benefits of premium membership. Seasonal trends show casual riders peak in July, indicating summer promotions could be effective. Additionally, highlighting the cost savings and convenience of premium membership for frequent short trips, including the 4.67% of rides that start and end at the same station, can further entice casual riders to upgrade.

Key tasks

- **Aggregate your data so it's useful and accessible.**

Data has been aggregated by user type, bike type, time of day, day of the week, and month to facilitate meaningful comparisons and analysis.

- **Organize and format your data.**

Data has been organized into a single table with clearly defined columns for different attributes. New columns have been added for better readability and analysis.

- **Perform calculations.**

Key calculations include average ride durations, peak hours, popular days, and preferences for bike types and stations.

- **Identify trends and relationships.**

Identified trends in ride duration, peak hours, day preferences, bike type preferences, seasonal trends, and station usage.

Deliverable

- **A summary of your analysis**
 - **Total Rides:** *5.74 million* rides in the dataset, with *2.05 million* rides by casual riders and *3.69 million* rides by premium members.
 - **Average Ride Duration:** Casual riders have an average ride duration of *10.64 minutes*, while premium members have an average ride duration of *8.28 minutes*.
 - **Peak Riding Hours:** Both casual riders and premium members have their peak riding hour at *5 PM*. Premium members have the lowest number of rides at *3 PM*.
 - **Popular Days:** Both casual riders and premium members prefer riding on weekends (*Saturday and Sunday*).
 - **Top Stations:**
 - Casual Riders: Top starting and ending stations are *Streeter Dr & Grand Ave* and *Dusable Lake Shore Dr & Monroe St*.
 - Premium Members: Top starting and ending stations are *Clinton St & Washington Blvd*, *Kingsbury St & Kinzie St*, and *Clark St & Elm St*.
 - **Bike Type Preferences:** Casual riders prefer *electric bikes*, while premium members prefer *classic bikes*. On the other hand, *docked bikes* are used only by some casual riders.
 - **Seasonal Trends:** *July* is the most preferred month for casual riders, while *August* is the most preferred month for premium members. Both groups ride the least in *January*.
 - **Same Station Rides:** Approximately *4.67%* of total rides start and end at the same station.
 - **Frequent Ride Durations:** The most frequent ride duration for casual riders is *4-5 minutes*, while premium members' most frequent ride duration is *9 minutes*.

Share

Guiding questions

- **Were you able to answer the question of how annual members and casual riders use Cyclistic bikes differently?**

Yes, the data clearly indicates differences in ride duration, peak hours, day preferences, bike type preferences, and seasonal trends between annual members and casual riders.

- **What story does your data tell?**

The data reveals that casual riders typically use bikes for longer, more leisurely rides, often on weekends and prefer electric bikes. In contrast, premium members use bikes more consistently, with shorter ride durations and a preference for classic bikes, indicating usage for commuting and routine trips. Both groups have peak riding hours in the late afternoon, but premium members are less active in the early afternoon.

- **How do your findings relate to your original question?**

The findings directly address the original question by highlighting specific usage patterns and preferences that can be targeted to convert casual riders into premium members, thus increasing profitability.

- **Who is your audience? What is the best way to communicate with them?**

The audience includes Cyclistic's marketing team, financial analysts, and business strategists. The best way to communicate with them is through a combination of clear, concise reports and visually appealing data visualizations.

- **Can data visualization help you share your findings?**

Yes, data visualizations can effectively convey complex patterns and trends in an accessible and engaging way, making it easier for the audience to understand and act on the insights.

- **Is your presentation accessible to your audience?**

Ensuring the presentation uses clear language, intuitive visuals, and provides context for all data points will make it accessible to all members of the audience.

Key tasks

- **Determine the best way to share your findings.**

A comprehensive report supplemented by a presentation with interactive visualizations in Power BI can effectively communicate the insights.

- **Create effective data visualizations.**

Utilized various charts and graphs such as line charts, bar charts, and pie charts to present the key findings clearly. Ensured all visualizations are labeled correctly and include legends and annotations where necessary.

- **Present your findings.**

Structured the presentation to tell a coherent story: started with the business question, moved to the data preparation and analysis process, and concluded with key insights and recommendations.

- **Ensure your work is accessible.**

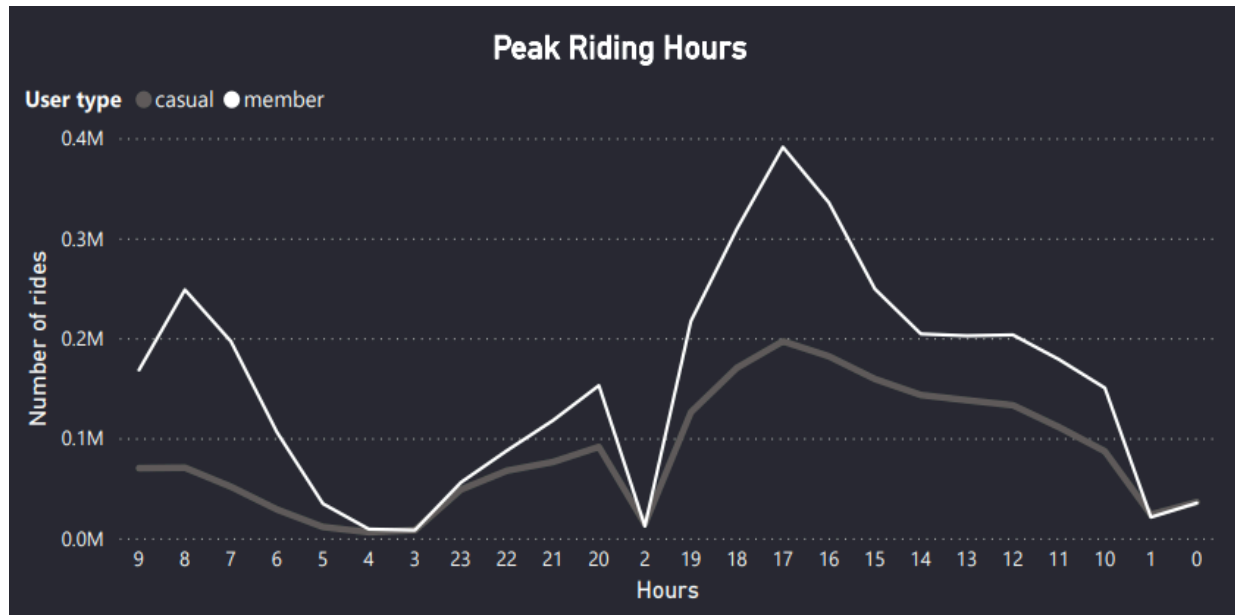
Provided a written report with the visualizations embedded, and ensured that the presentation is designed with accessibility in mind, using high-contrast colors and large fonts.

Deliverable

- **Supporting visualizations and key findings**

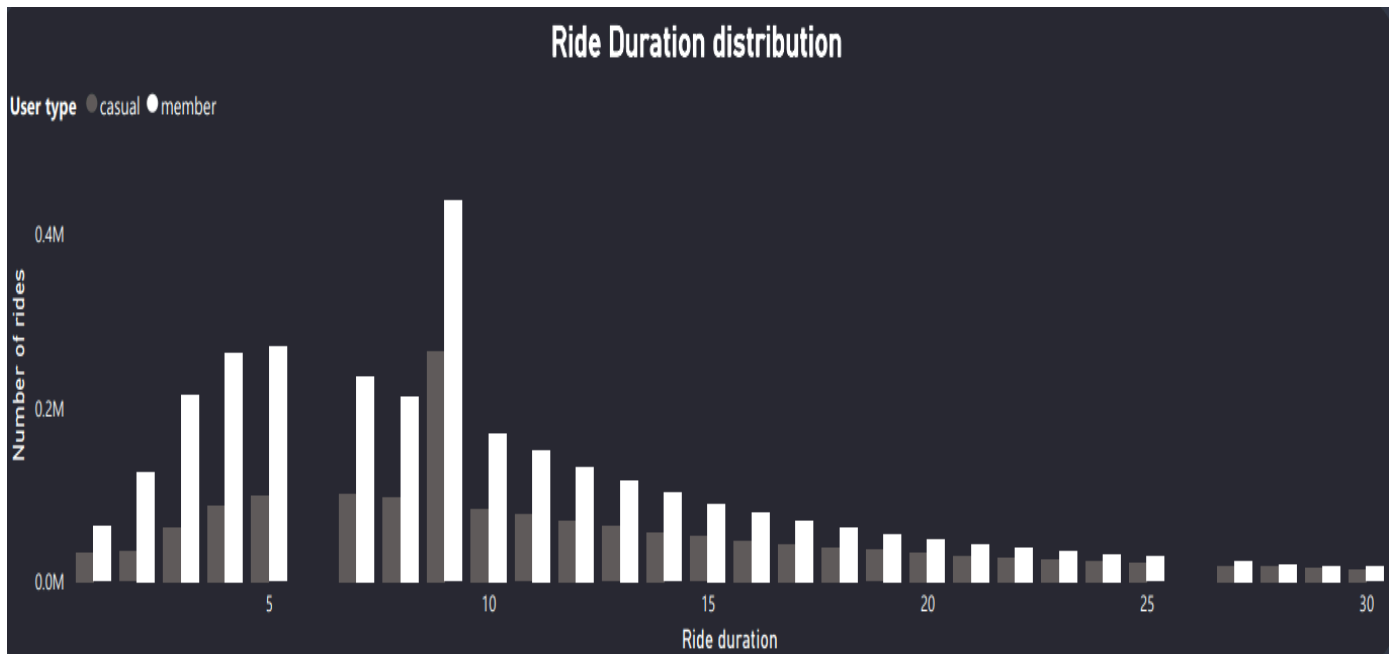
- **Peak Riding Hours:**

- Line chart showing the distribution of rides across different hours of the day for both user groups.

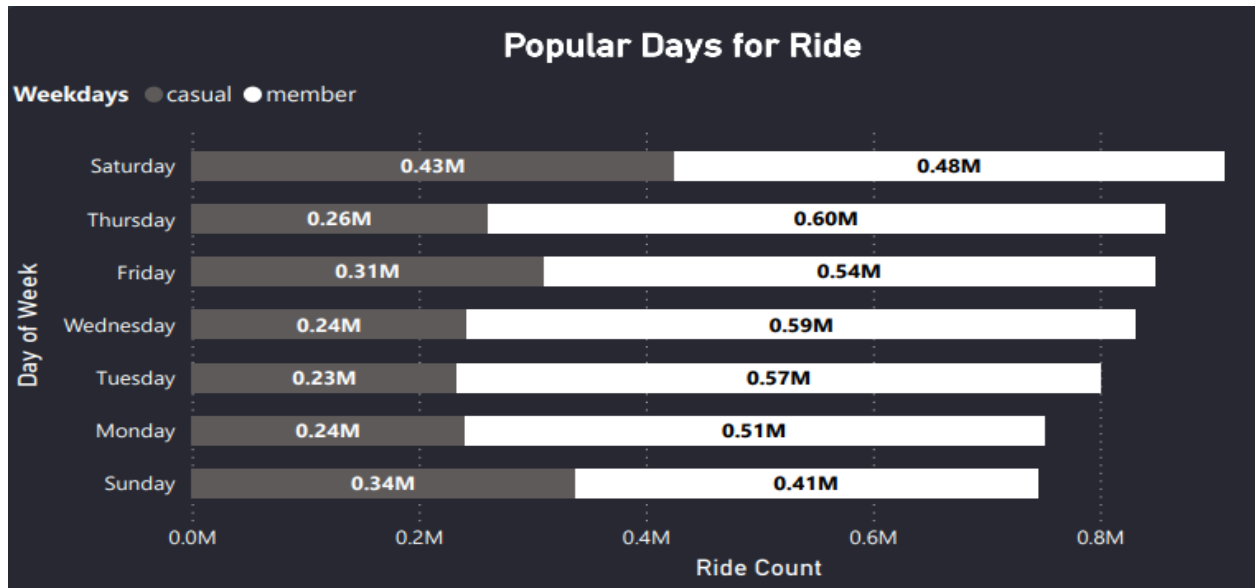


○ Ride Duration:

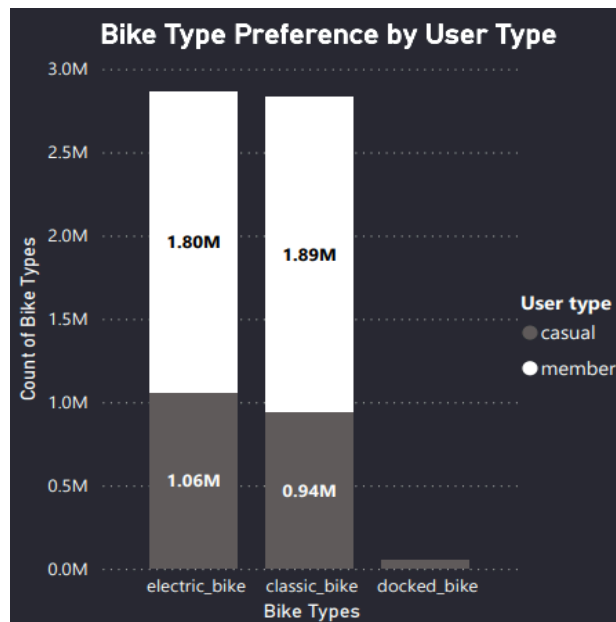
- Bar chart comparing ride duration distribution of casual riders and premium members.



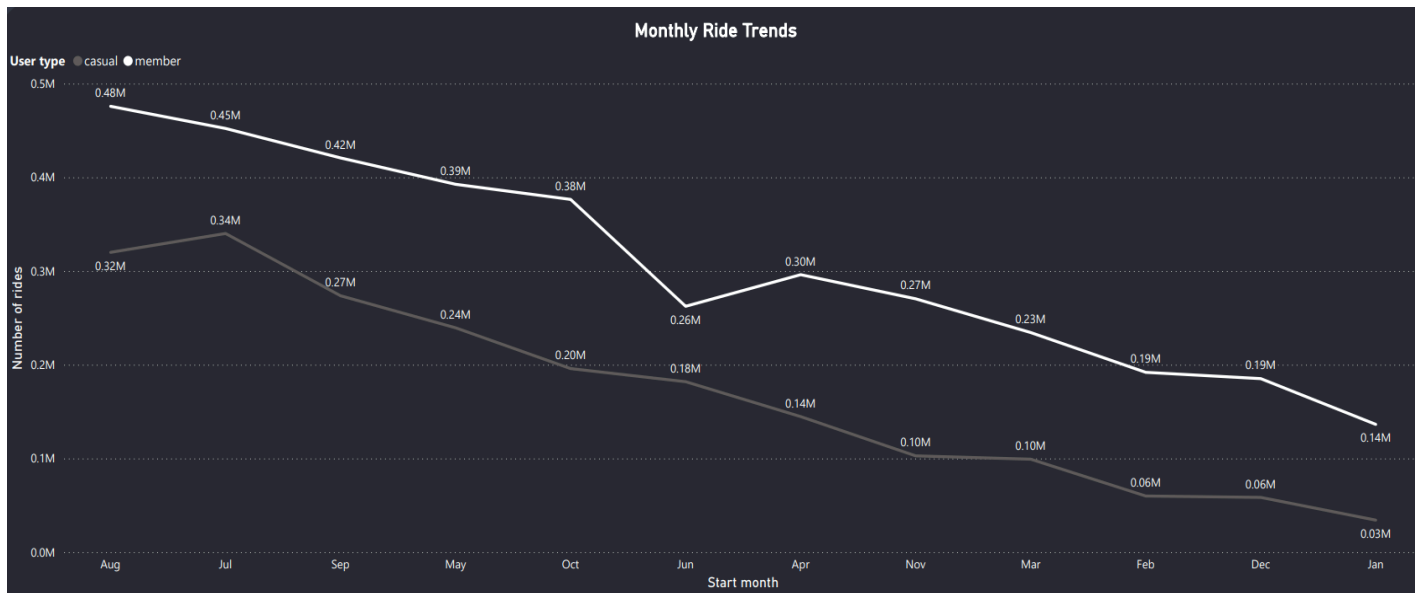
- **Day Preferences:**
 - Bar chart highlighting the most popular days for rides among casual riders and premium members.



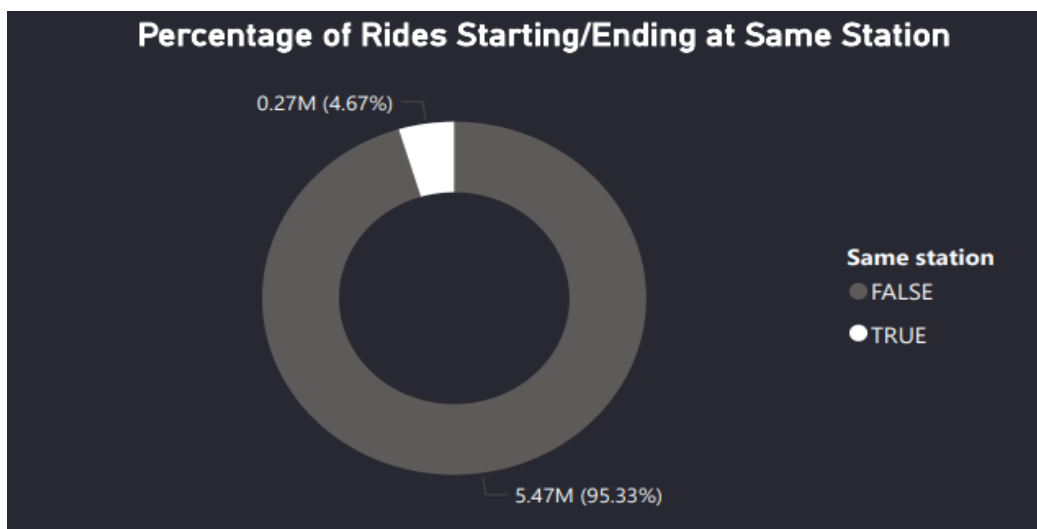
- **Bike Type Preferences:**
 - Stacked bar chart comparing the preferences for classic, electric, and docked bikes between casual riders and premium members.



- **Seasonal Trends:**
 - Line chart showing monthly ride counts for casual riders and premium members.



- **Same Station Rides:**
 - Pie chart showing the percentage of rides starting and ending at the same station for both user groups.



Act

Guiding questions

- **What is your final conclusion based on your analysis?**

Casual riders prefer longer, leisurely rides primarily on weekends and favor electric bikes, while premium members have shorter, more routine rides mostly during weekdays, preferring classic bikes. This indicates that casual riders use the service for recreation, whereas premium members use it for commuting and regular travel.

- **How could your team and business apply your insights?**

Cyclistic can tailor marketing campaigns, pricing strategies, and service offerings to target casual riders by emphasizing the benefits of premium membership for frequent and efficient travel. Additionally, optimizing bike availability and promotions during peak times and at popular stations can enhance user experience and encourage conversions.

- **What next steps would you or your stakeholders take based on your findings?**

- Develop targeted marketing campaigns for casual riders, particularly highlighting the convenience and cost savings of premium membership.
- Introduce promotional offers during peak usage times and popular days (weekends for casual riders).
- Enhance bike availability and service quality at top stations and during peak hours.
- Conduct user surveys to gather more detailed preferences and feedback to further refine marketing and service strategies.

- **Is there additional data you could use to expand on your findings?**

Additional data on user demographics, detailed trip purposes, and feedback from casual riders about their hesitations towards premium membership can provide deeper insights. Seasonal weather data and special event schedules could also help understand fluctuations in ride patterns.

Deliverable

- **Your top three recommendations based on your analysis?**
 - **Targeted Marketing Campaigns:**

Launch campaigns emphasizing the benefits of premium membership, such as cost savings for frequent riders, convenience, and exclusive perks. Focus on promoting these benefits during peak usage times and popular days for casual riders.
 - **Optimized Bike Availability and Promotions:**

Ensure high availability of electric bikes at top stations during weekends and peak hours to cater to casual riders' preferences. Offer limited-time promotions or discounts for upgrading to premium membership, particularly during the summer months when casual ridership peaks.
 - **Enhanced User Experience at Popular Stations:**

Improve amenities and services at the most frequented stations for both casual and premium members. Consider adding features like quick check-in/out kiosks, better signage, and promotional material highlighting the advantages of premium membership.