CYCLISTIC BIKE-SHARE CASE STUDY

CAPSTONE PROJECT

ASK PHASE

Business Task

Determine how annual members and casual riders use Cyclistic bikes differently so we can design digital marketing tactics to convert casual riders into yearly members.

Stakeholders

Lily Moreno, Director of Marketing at Cyclistic, and the executive team, who will use the insights to guide marketing strategies for converting casual riders into annual members.

PREPARE PHASE

Data Source

All trip data was downloaded from the official Divvy/Cyclistic data portal:
https://divvy-tripdata.s3.amazonaws.com/index.html

Date downloaded: October 29, 2025

Files Overview & Metadata

File Name	Size (approx.)	Rows
2024_11-divvy-tripdata.csv	250 MB	700,000
2024_12-divvy-tripdata.csv	245 MB	690,000
2025_01-divvy-tripdata.csv	210 MB	640,000
2025_02-divvy-tripdata.csv	230 MB	670,000
2025_03-divvy-tripdata.csv	265 MB	710,000
2025_04-divvy-tripdata.csv	280 MB	720,000
2025_05-divvy-tripdata.csv	310 MB	780,000
2025_06-divvy-tripdata.csv	320 MB	800,000
2025_07-divvy-tripdata.csv	330 MB	810,000
2025_08-divvy-tripdata.csv	340 MB	825,000
2025_09-divvy-tripdata.csv	350 MB	840,000
2025_10-divvy-tripdata.csv	355 MB	845,000
Total	**3.46 GB**	**≈9 million rows**

PROCESS PHASE

Data Consistency and Schema Variations

- Some earlier files used member_type instead of member_casual.
- Occasional missing values for start_station_name and end_station_name.
- A few records lacked end_lat and end_lng coordinates.
- No Personally Identifiable Information (PII) available.
- After standardization, all columns were aligned and consistent across files.

Data Cleaning & Preparation Summary

Objective

Prepare and standardize the last 12 months of Cyclistic trip data to ensure accurate, consistent, and analysis-ready datasets for comparing casual and member rider behaviors.

1. Data Import and Merging

- Imported monthly CSV files (2019 Q1, 2020 Q1) into R using the **read_csv()** function.

- Combined datasets into one consolidated dataframe using **bind_rows()** from the **dplyr** package.

- Standardized column names and saved the processed data in a separate folder simultaneously.
- Verified that all columns aligned correctly before merging.

2. Data Type Conversions

- Converted **start_time** and **end_time** columns to a standard **datetime** format.

- Changed **trip_id** to **character** type for analysis consistency.

3. New Variables Created

New Column	Description	Formula / Method
ride_length_min	Total duration of each ride in minutes	= (end_time(sec) - start_time(sec)) / 60
day_of_week	Day name of ride start	= weekdays(start_time, label=TRUE)
start_hour	Hour of ride start (0–23)	= format(start_time,"%H")
is_weekend	Indicates if ride occurred on Sat/Sun	Based on day_of_week
ride_season	Group rides by season (Winter, Spring, etc.)	Derived from the month field

4. Data Cleaning Steps

- Removed rows with missing **start_time, end_time**, or station name values and invalid rows having a negative duration where end_time < start_time.

- Filtered out rides with **negative or zero duration** (invalid records).

- Excluded outlier rides lasting over **24 hours**.

- Standardized **member_type** labels — unified all variations (**member, Subscriber, Customer**) into two categories: **member** and **casual**.

- Checked for and removed duplicate records based on **trip_id**.

- Verified data integrity after cleaning (row counts and column structure).

5. Data Validation

- Confirmed datetime consistency (**end_time** always occurs after **start_time**).

- Checked distributions of ride duration and ride start times for outliers.

- Cross-verified that total ride counts matched between the combined and original monthly datasets.

6. Tools Used

- **RStudio** with the **tidyverse** package for data cleaning and transformation.

- **Google Sheets** and **Excel** for spot checks, filters, and verification.

"Thorough data cleaning ensured high-quality inputs for analysis, allowing reliable comparisons between member and casual rider behavior."

ANALYZE PHASE

Data Analysis & Visualization Summary

Objective

Analyze cleaned trip data to uncover key differences between casual riders and annual members, focusing on usage patterns, ride duration, and time-of-day trends.

1. Descriptive Analysis

- Calculated **total rides**, **average ride length**, and **median ride length** for each rider type.

- Identified daily and hourly ride distribution patterns.

- Determined peak riding periods (weekday vs. weekend).

- Evaluated seasonality by grouping trips by month.

Example summary metrics:

Metric	Casual Riders	Members	Insight
Average Ride Length (min)	38.37	11.41	Casuals ride ~3× longer → leisure pattern
Median Ride Length (min)	23.13	8.47	Members have shorter, consistent rides
% Weekend Rides	47.3 %	16.60 %	Casuals prefer weekends
Peak Time of Day	1 – 4 PM	7 – 9 AM & 4 – 6 PM	Members commute; casuals leisure

2. Comparative Analysis

- Compared **ride-length distributions** using mean, median, and standard deviation.

- Created grouped summaries by `day_of_week` and `member_casual`.

- Observed higher variability in casual riders' durations.

- Conducted basic **t-tests** confirming that differences in mean ride lengths between rider types were statistically significant.

3. Visualization Overview

To communicate findings effectively, multiple charts were created in R (ggplot2) and Tableau:

Chart Type	Purpose / Insight
Bar Chart - Average Ride Length by Day of Week grouped by User Type	Compare ride duration between casual and member riders by day of the week.
Bar Chart - Rides by Day of Week grouped by User Type	Show daily trends by user type.
Bar Chart - Monthly Rides Analysis by Member Type	Shows monthly ride volumes between members and casual users.
Line Chart - Monthly Ride Volume	Identify seasonal usage patterns.
Line Chart - Hourly Ride Distribution by Member Type	Shows ride activity variations by hour for each member type.
Line Chart - Hourly Ride Distribution by DayType	Shows changes in ride activity by hour on weekdays vs. weekends.

Each visualization included clear labels, titles, and short takeaway descriptions in this report, such as:

"Casual riders take longer rides concentrated on weekends, while members show weekday commuting peaks."

4. Key Findings

- **Members** primarily ride on weekdays during commuting hours, with short, consistent trips.

- **Casuals ride approximately 3 times longer**, showing leisure use.

- **Weekend** use is much higher for **casuals** (~47% vs 17%).

- **Peak times differ**: casuals ride most in the afternoon, members during commute hours.

- Overall, **members = commuters**, **casuals = recreational users**.

SHARE PHASE

Share Results & Recommendations

Objective

Communicate the key findings clearly to Cyclistic's Director of Marketing and the executive team, and present recommendations based on the data.

1. Summary of Findings Shared

- **Casual riders** take longer trips and ride mostly on **weekends and afternoons**.

- **Members** ride mainly on **weekdays**, especially during **morning and evening commute hours**.

- Usage increases for both groups during the **summer months**.

2. Key Insights Communicated

The following insights were highlighted through charts, tables, and summaries:

- **Ride Duration Differences**
  Casual riders take roughly thrice as long per ride as members, suggesting recreational usage.

- **Time-of-Day Patterns**
  Members peak during commute hours (7–9 AM, 4–6 PM), whereas casual riders peak midday (1–4 PM).

- **Day-of-Week Patterns**
  Casual riders heavily favor weekends, whereas members ride mostly on weekdays.

- **Seasonal Behavior**
  Both groups exhibit higher usage during warmer months.

3. How Results Were Communicated

- A short slide presentation summarizing the business task, cleaned data, visualizations, and key insights.

- A written report with charts (ride duration, day-of-week patterns, hourly usage, station hotspots) and concise explanations.

Both formats focused on clear visuals and simple takeaways.

4. Recommendations

- **Weekend Promotions for Casual Riders**
  - Since casual riders are heavy weekend users, offer discounted trial memberships or weekend-focused ads.
- **Convert High-Frequency Casual Riders**
  - Identify casual riders with frequent usage and send personalized membership savings messages through email or app notifications.
- **Commuter Campaigns for Members**
  - Highlight the cost and time savings of an annual membership during commuter peak times.

5. Measurement Strategy

To track the success of marketing efforts, recommend monitoring:
- Monthly conversion rate (casual → annual member)

- Membership retention rates

- Cost-per-acquisition (CPA) for marketing campaigns

- Ride usage changes after promotional periods

6. Conclusion

The insights and recommendations provide a clear roadmap for increasing Cyclistic's annual membership base. By tailoring marketing strategies to the distinct behaviors of casual and member riders, the company can convert high-value users, improve customer retention, and enhance long-term revenue.

ACT PHASE

Final Recommendations & Next Steps

Objective

Turn insights into clear actions Cyclistic can take to increase annual memberships based on rider behavior patterns.

1. Final Recommendations

- Target Weekend Casual Riders
  - Casual riders show strong weekend and leisure usage. Launch weekend-focused membership promotions and location-based ads around parks and tourist areas.
- Convert Frequent Casual Users
  - Identify casual riders who ride multiple times per month. Send personalized messages highlighting membership savings and convenience.

- Strengthen Commuter Campaigns
  - Members ride during weekday commute hours. Promote annual memberships with messaging focused on cost savings, reliability, and time efficiency, especially near business districts.

2. Next Steps

- **Pilot Campaigns**: Run small A/B tests for each recommendation (weekend promo, commuter messaging, personalized offers).

- **Track KPIs:**

  - Casual → member conversion rate

  - Cost per acquisition

  - Membership retention

  - Ride volume changes after each campaign

- **Refine Strategy**: Adjust campaigns based on performance data from pilot tests.

- **Expand Insights:** Consider analyzing demographics, route popularity, and bike type preferences to further personalize marketing.

3. Conclusion

By tailoring marketing strategies to the distinct behaviors of casual and member riders, Cyclistic can effectively increase annual memberships, improve customer retention, and maximize long-term revenue.