Case study: How does a bike-share navigate speedy success?

Introduction

Cyclistic is a Chicago-based bike-share program with a fleet of over 5,800 bikes and 692 docking stations. Cyclistic offers flexible pricing options, including single-ride passes, day passes, and annual memberships. While casual riders account for a significant portion of trips, annual members provide greater long-term profitability. This case study explores how Cyclistic can better understand the differences between casual riders and members to convert casual riders into annual members.

Methodology

The study followed a structured data analysis process:

1. **Ask**: Identify differences in usage patterns between casual riders and annual members.
2. **Prepare**: Clean and preprocess Cyclistic trip data from 2019 and 2020, ensuring data consistency and usability.
3. **Process**: Calculate key metrics such as ride duration and analyze trends using tools like R and Excel.
4. **Analyze**: Explore behavioral differences through descriptive statistics and visualizations.
5. **Share**: Present findings and actionable recommendations.
6. **Act**: Provide strategic guidance for marketing campaigns.

Business Task

The primary task is to analyze how annual members and casual riders use Cyclistic bikes differently. The insights from this analysis will inform the design of targeted marketing strategies to convert casual riders into annual members, aligning with Cyclist's goal of maximizing profitability and fostering long-term customer loyalty.

Questions

- What patterns differentiate casual riders and annual members in terms of ride frequency, duration, and timing?
- What insights can help design targeted campaigns for casual riders?
- How can these findings support Cyclist's goal of boosting annual memberships?

Data Preparation Phase

Data cleaning and preparation steps included:

Data Overview

The dataset comprises 2019 and 2020 bike trip records provided by Cyclistic. Key fields include:

- **ride_id**: Unique identifier for each trip.
- **rideable_type**: Type of bike used (standard, electric, assistive).
- **member_casual**: User type (casual rider or annual member).
- **started_at** and **ended_at**: Trip start and end timestamps.
- **start_station_name** and **end_station_name**: Trip origin and destination.
- **start_lat** and **start_lng**: GPS coordinates for trip start.
- **end_lat** and **end_lng**: GPS coordinates for trip end.

Data Organization

The data is organized into CSV files, each representing a month of trip records. These files are standardized with consistent columns, enabling seamless integration for analysis.

Data Credibility

The dataset is:

- **Reliable**: Generated directly from Cyclistic's tracking system.
- **Original**: Not derived or modified from external sources.
- **Comprehensive**: Covers all trips over a full year.
- **Current**: Reflects recent trends.
- **Cited**: Provided under a license by Motivate International Inc.

Data Privacy and Security

To comply with privacy regulations:

- No personally identifiable information (PII) is included.
- Aggregated data is used for analysis, ensuring customer anonymity.

Preparation Steps

1. **Download and Store Data**
    - Organized the dataset into folders, separating raw and cleaned files.
2. **Initial Data Review**
    - Examined column consistency across files.
    - Verified data types and identified missing or invalid records.
3. **Data Cleaning**
    - Removed duplicate entries and outliers (e.g., rides with negative durations).

    - Standardized station names to address inconsistencies.
4. **Data Transformation**
    - Created new columns for **ride_length** (trip duration) and **day_of_week** (weekday of the trip).
    - Ensured proper formatting for date and time fields.

Data Challenges

- **Outliers**: Unusually long trip durations required careful filtering.
- **Missing Data**: A small percentage of records lacked station or GPS details and were excluded.
- **Station Name Variations**: Corrected inconsistencies in naming to enable reliable grouping.

The resulting dataset contained over 790,000 trip records, ready for aggregation and analysis.

Process Phase

1. Check Data for Errors

- **Downloaded Data**: 2019 and 2020 years trip data in CSV format.
- **Error Identification**:
    - Missing or null values in essential fields (e.g., `ride_id`, `started_at`).
    - Duplicate entries with identical `ride_id`.
    - Negative or zero ride durations (`ended_at < started_at`).

2. Tools Used

- **RStudio**: Selected for handling large datasets beyond the capacity of Excel/Google Sheets. Used for data transformation, cleaning, and visualization.
- **Microsoft Excel/Google Sheets**: Used for smaller-scale cleaning and preliminary checks.

3. Documentation of Data Cleaning Process

- **Step 1**: Loaded data into RStudio and examined structure with functions like `head()` and `str()`.
- **Step 2**: Removed rows with missing values using the `na.omit()` function.
- **Step 3**: Filtered out invalid records:
    - Rides with `ride_length` ≤ 0.
    - Duplicates using the `distinct()` function.
- **Step 4**: Standardized:
    - Station names and `rideable_type` using `mutate()` and `stringr::str_to_title()`.

    - Date formats with `lubridate::ymd_hms()`.
- **Step 5**: Verified data integrity:
    - Checked summary statistics (`summary()`).
    - Ensured no inconsistencies in categorical fields using `unique()`.

The cleaning and processing steps resulted in a structured and consistent dataset:

- **Ride Duration (`ride_length`)**: Validated and calculated for all trips.
- **Day of Week (`day_of_week`)**: Extracted for trend analysis.
- **Categorical Consistency**: Corrected station names and user type classifications.

Findings

1. Usage Patterns

- Casual riders predominantly use bikes on weekends, peaking on Sundays.
- Annual members maintain consistent usage across all days, reflecting regular commuting habits.

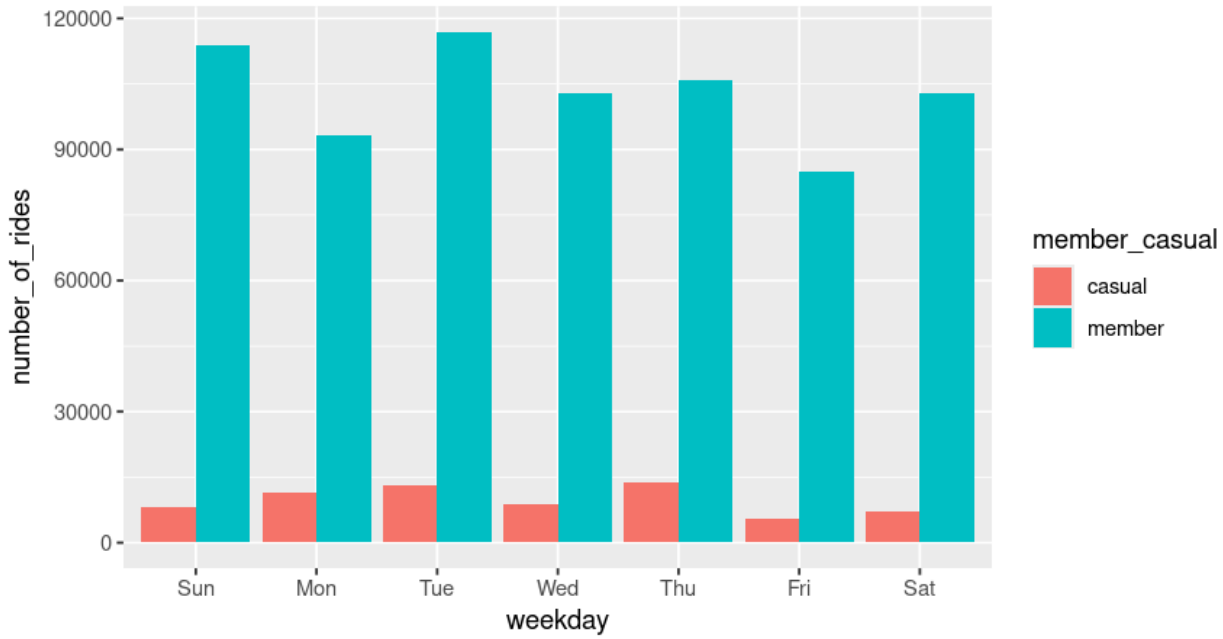
2. Ride Duration

- Casual riders' trips are significantly longer on average, particularly on weekends.
- Members' rides are shorter, indicative of focused utility-driven trips.

Visual Insights

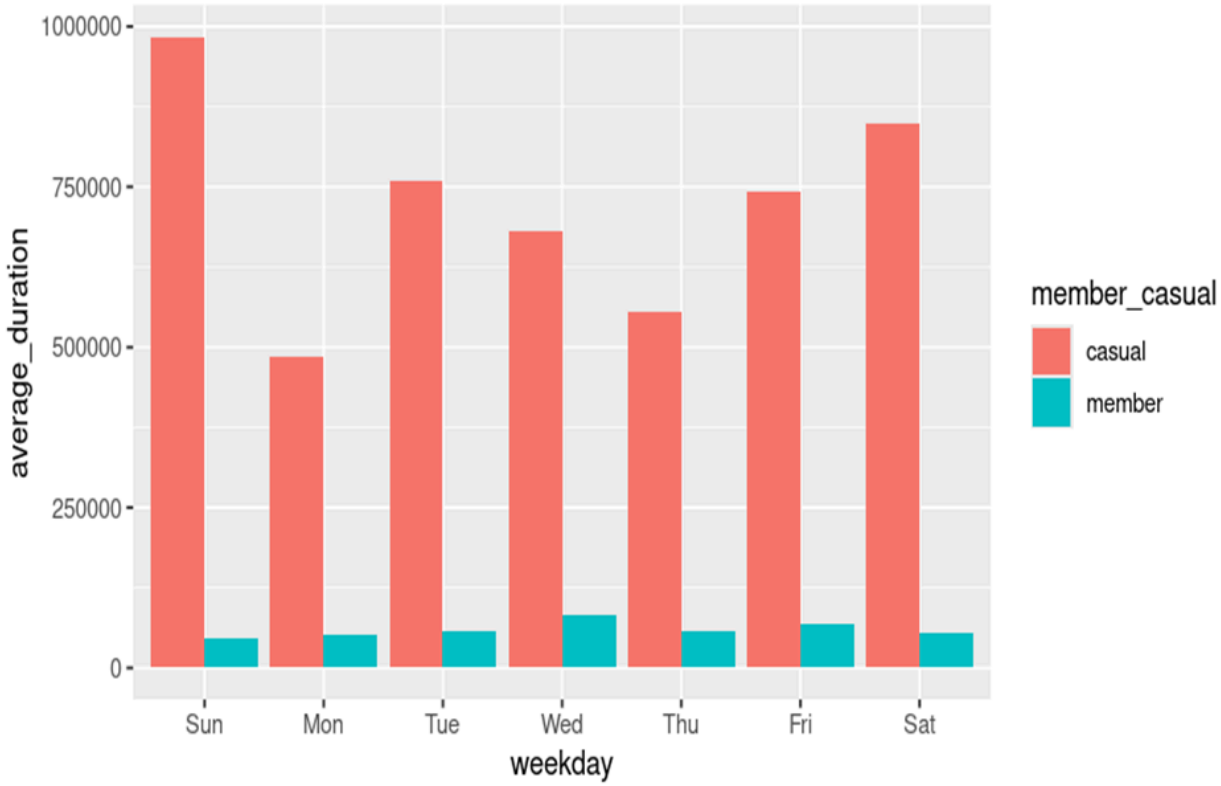
Number of Rides by Day of Week:

- The chart compares weekly ride trends for members and casual riders.
- **Members**: Consistent usage throughout the week, peaking on Tuesday.
- **Casual Riders**: Higher activity on weekends, indicating leisure-oriented usage.
- Members consistently outnumber casual riders across all days.
- Highlights distinct usage patterns between the two groups.



Average Ride Duration:

- Casual users have longer ride durations than members.
- Peak durations for casual users on weekends (Saturday, Sunday).
- Members show consistent, shorter durations across all days.
- Indicates casual users ride for leisure, members for routine trips.



Recommendations

To convert casual riders into annual members, Cyclistic should implement the following strategies:

1. **Weekend Promotions**: Offer exclusive weekend discounts on annual memberships to casual riders.
2. **Commute-Focused Marketing**: Highlight the cost-effectiveness and convenience of annual memberships for weekday commuters.
3. **Loyalty Programs**: Create rewards for frequent casual riders, including ride credits or trials of annual memberships.

Conclusion

By leveraging these insights, Cyclistic can design targeted marketing campaigns that address the specific behaviors and needs of casual riders. These strategies aim to enhance engagement, improve profitability, and foster long-term growth for the Cyclistic bike-share program.

Summary: Understanding rider behavior is key to increasing memberships.
Additional Insights:

- Annual members show patterns that align with daily commuting needs, emphasizing stability.
- Casual riders demonstrate a recreational approach, highlighting potential weekend-focused campaigns.

Next Steps: Implement targeted campaigns, refine data tracking for emerging trends, and monitor effectiveness of membership drives.

Future Opportunities

Enhanced Technology Integration: Leverage app-based notifications to promote membership benefits in real-time.
Sustainability Goals: Emphasize the environmental benefits of consistent bike-share usage.
Community Engagement: Partner with local businesses to offer discounts tied to memberships.