CASE STUDY

HOW DOES A BIKE-SHARE NAVIGATE SPEEDY SUCCESS?

Diyashree Devadiga

Table of contents:

Statement of business task:

The task is to analyse and compare the usage patterns of casual riders and annual members to provide data-driven insights and recommendations for converting casual riders into annual members through a targeted marketing strategy.

Description of Data:

The dataset contains trip details for Cyclistic bike users over a one-year period. It includes:

- **User Type:** Categorized as either "Member" (annual subscribers) or "Casual" (pay-per-ride users).

- **Ride Length:** Duration of each bike trip in time format. This is a calculated column.

- **Start date and End date:** The month and year of each ride. This column is not present in all excel files.

- **Other Metadata:** Additional details like ride start time, end time, and ride id, start station, end station, start longitude, end longitude, start latitude, end latitude, rideable type.

The analysis focuses on understanding the behavioural differences between members and casual riders by examining ride duration, user trends over time, and variations in usage patterns.

Issues of Data and Data Transformation:

Data Issues:

All the columns were in general format. Columns had to be converted according to their correct data types. There were many missing values especially in start station and end station columns. Other issues were:

Extreme Ride Durations: Some rides lasted for **over 1400 minutes**, indicating possible paused rides rather than continuous trips.

Outliers: The presence of **extremely long rides** skewed the data distribution.

Data Transformation:

Excel transformation:

Calculated and converted ride length from general to time format.

Separated started_at column into 2 columns: started_date and started_time and same for ended_at for all excel files.

Calculated and created day_of_week column.

Pandas transformation:

- **Converted ride length from seconds to minutes** for better readability in visualizations.
- **Filtered out extreme outliers** (e.g., rides exceeding a reasonable duration threshold).
- **Grouped data by month** to observe seasonal trends.
- **Separated data by user type** to compare behavioral differences between members and casual riders.

Approach:

To analyze the differences in usage patterns between members and casual riders, the following steps were taken:

Excel:

Calculated ride length and day of week to get mean ride lengths and mode of day(maximum cycles used in a particular day).

Created pivot table and added summary of data for every excel file.

Pandas:

1. **Exploratory Data Analysis (EDA):** Examined ride duration, user distribution, and monthly trends.

2. **Visualizations:** Created bar charts, line charts, box plots, pie charts, and KDE plots.

3. **Key Metrics Comparison:** Compared average ride duration, total rides per month, and seasonal trends.

The goal was to identify key insights that can help Cyclistic optimize its service and improve user engagement.

Results:

Visualizations and Insights

1. Number of Members vs. Casual Riders Per Month (Bar Chart)

- **Members outnumbered casual riders in all months.**

- **Both groups showed an increase in ridership from January to September, then declined.**

- **Seasonality effect:** Higher usage in warmer months, decline in winter.

2. Ride Duration Trends (Line Chart)

- **Casual riders have longer average ride durations than members.**

- The **mean ride length of casual riders increased from January to May**, then suddenly dropped, but remained higher than members' average ride time.

- This suggests that **members use bikes for shorter, more consistent trips**, likely for commuting.

3. Ride Duration Variations (Box Plot)

- **Many outliers indicate unusually long rides**, likely due to paused rides.

- The first few months had a **high concentration of long ride times (>1400 minutes).**

- **Members have more consistent ride durations**, while casual riders exhibit more variation.

4. Distribution of Members vs. Casual Riders (Pie Chart)

- **62.8% of users are members**, while **37.2% are casual riders**.

- This reinforces that **Cyclistic primarily serves regular commuters.**

5. Ride Length Distribution (KDE Plot)

- **Casual riders tend to have longer ride durations compared to members.**

- **Members exhibit a sharper peak at lower ride durations**, indicating more consistent short trips.

Key Findings:

- Members ride more frequently and have shorter, consistent trips, suggesting daily commute usage.
- Casual riders take longer rides, possibly for leisure or tourism, and their numbers fluctuate seasonally.
- Seasonal trends show higher ridership in warmer months (peak in September), declining afterward.
- Outliers indicate some rides lasted excessively long due to paused rides or improper ride endings.
- Members form the majority (62.8%), reinforcing that Cyclistic is primarily used as a commuting service.

Conclusion:

- The data **confirms clear differences in how members and casual riders use Cyclistic bikes.**
- **Members use bikes for daily commutes with shorter and consistent ride durations.**
- **Casual riders take longer trips, possibly for recreational purposes.**
- The **seasonal variation** in ridership suggests an opportunity to attract more casual riders in peak months.
- **Ride duration outliers suggest some casual riders may be pausing rides rather than ending them.**

Top 3 Recommendations based on Analysis:

- **Encourage casual riders to become members**
    - Offer targeted promotions during peak months (e.g., **discounted annual memberships in summer**).
    - Provide incentives for first-time casual users to subscribe.
- **Optimize ride duration policies to prevent excessive long rides**
    - **Introduce ride-ending reminders** to reduce long-duration outliers.
    - Adjust pricing to **discourage paused rides** and encourage proper ride termination.
- **Leverage seasonality to boost engagement**

- Launch **seasonal campaigns in summer** when casual ridership is high.
- Consider **expanding bike availability in peak months** to accommodate demand.