How Does a Bike-Share Navigate Speedy Success?

Cyclistic Case Study

Michael Karanikolas

2023-10-02

Introduction

Cyclistic is a leading bike-sharing company based in Chicago. The company has seen tremendous growth since its inception in 2016, however, the company's marketing director believes the future success lies in maximising the number of annual memberships. This study aims to provide insights for converting casual riders into annual members, a key factor in driving the company's profitability and sustainability.

The analysis that follows looks carefully at Cyclistic's trip data, highlighting usage differences between casual and annual members. The goal is to use this information to improve the company's marketing and increase the number of annual members. The project showcases how important data is in making good business decisions. It demonstrates the process of turning raw data into clear insights that can help a business grow.

This report will initially cover the data sources followed by the process of data cleaning and manipulation. This will be followed by an in-depth analysis of the data and the key findings from this analysis. Finally, actionable recommendations based on these findings will guide Cyclistic's future marketing strategies.

Data Process and Analysis

Data Sources

The case study focuses on a hypothetical company named 'Cyclistic,' a leading bike-sharing service based in Chicago. The data used in this study is a 12-month collection of trip records from Motivate International Inc.'s Divvy bike-sharing service in Chicago, providing a practical real-world reference. The data includes details such as start and end times, type of bike used, location coordinates, and type of membership. All data usage is in compliance with the terms of the Data License Agreement, with no attempt to correlate with personally identifiable information or utilise the data unlawfully. This analysis does not imply endorsement or affiliation with Lyft Bikes and Scooters, LLC, or the City of Chicago.

Data Preparation and Cleaning

Microsoft Excel

Figure 1 depicts a small part of the two extra columns that were created. A new column named '*ride_length*' was developed to calculate the length of each ride, determined by the difference between the '*started_at*' and '*ended_at*' timestamps. This new format, shown as HH:MM:SS, was key to studying how long the rides lasted.

In addition, a '*day_of_week*' column was introduced to indicate the day of the week each ride took place. This was derived using the 'WEEKDAY' function, providing a valuable perspective on patterns of ridership throughout the week.

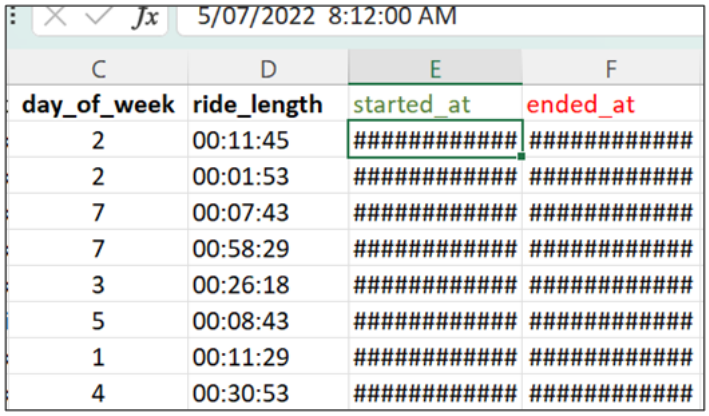

Figure 1: Screenshot of "*day_of_week*", "*ride_length*", "*started_at*" and "*ended_at*" columns from a sample dataset.

SQL

Since each file contained data for each separate month of the year, they were all combined into one table to make data cleaning and manipulation easier. The code is as follows:

```
CREATE TABLE `cyclistic-374321.cyclistic.combined_trips` AS
SELECT * FROM `cyclistic-374321.cyclistic.january_data` UNION ALL
SELECT * FROM `cyclistic-374321.cyclistic.february_data` UNION ALL
SELECT * FROM `cyclistic-374321.cyclistic.march_data` UNION ALL
SELECT * FROM `cyclistic-374321.cyclistic.april_data` UNION ALL
SELECT * FROM `cyclistic-374321.cyclistic.may_data` UNION ALL
SELECT * FROM `cyclistic-374321.cyclistic.june_data` UNION ALL
SELECT * FROM `cyclistic-374321.cyclistic.july_data` UNION ALL
SELECT * FROM `cyclistic-374321.cyclistic.august_data` UNION ALL
SELECT * FROM `cyclistic-374321.cyclistic.september_data` UNION ALL
SELECT * FROM `cyclistic-374321.cyclistic.october_data` UNION ALL
SELECT * FROM `cyclistic-374321.cyclistic.november_data` UNION ALL
SELECT * FROM `cyclistic-374321.cyclistic.december_data`;
```

A new column, named *ride_length_minutes*, was created in the *combined_trips* SQL table to store the ride length in minutes for easier data analysis and understanding.

```
ALTER TABLE `cyclistic-374321.cyclistic.combined_trips`
ADD COLUMN ride_length_minutes FLOAT64;
```

Converting the *ride_length* to minutes in the *combined_trips* table and updating the new *ride_length_minutes* column with these values.

```
UPDATE `cyclistic-374321.cyclistic.combined_trips`
SET ride_length_minutes = (EXTRACT(HOUR FROM ride_length) * 60)
+ EXTRACT(MINUTE FROM ride_length)
+ (EXTRACT(SECOND FROM ride_length) / 60)
WHERE 1=1;
```

Deleting rows from the *combined_trips* table where the trip duration is negative or zero, as these instances are not meaningful for the analysis and may distort the results.

```sql
DELETE FROM `cyclistic-374321.cyclistic.combined_trips`
WHERE TIMESTAMP_DIFF(ended_at, started_at, SECOND) <= 0;
```

Removing rows from the *combined_trips* table that contain blank or NULL values for the *ride_id* or *member_casual* columns. This is to ensure that each trip has a valid ID and that we know the membership status of each rider, which are crucial for the analysis.

```sql
DELETE
FROM `cyclistic-374321.cyclistic.combined_trips`
WHERE ride_id IS NULL OR ride_id = '' OR member_casual IS NULL OR member_casual = '';
```

Checking for duplicate rows in the *combined_trips* table using the *ride_id* and *ride_length* as identifiers, to ensure the data's uniqueness and integrity.

```sql
SELECT ride_id,  ride_length
FROM `cyclistic-374321.cyclistic.combined_trips`
GROUP BY ride_id, ride_length
HAVING COUNT(*) > 1;
```

Removing rides from the *combined_trips* table that have a *ride_length_minutes* that falls below the 1st percentile or above the 99th percentile. This is to remove outlier values, which are unlikely to represent typical rider behaviour and might skew the results of the analysis. Based on an initial examination of the data distribution, the 1st and 99th percentiles appeared to effectively isolate the most extreme values that seemed inconsistent with typical rider behaviour. Since the dataset is large, removing 2% still leaves a significant amount of data for analysis, ensuring robustness in the findings.

```sql
SELECT
ride_length_minutes,
  PERCENTILE_CONT(ride_length_minutes, 0.01) OVER () AS percentile_01,
  PERCENTILE_CONT(ride_length_minutes, 0.99) OVER () AS percentile_99
FROM `cyclistic-374321.cyclistic.combined_trips`
```

```sql
DELETE FROM `cyclistic-374321.cyclistic.combined_trips`
WHERE ride_length_minutes < '1st percentile'
 OR ride_length_minutes > '99th percentile'
;
```

Calculations and Analysis

All of the calculations below assisted in the visualisation part of the case study. That way useful graphs and charts could be created to give a grater insight on how riders use bikes. Calculating the average ride length per month for each member type (casual or member) in the *combined_trips* table. This provides insights into how ride durations differ between casual and member riders over time.

```sql
SELECT
   FORMAT_DATE('%m %Y', started_at) AS month_year,
  AVG(CASE WHEN member_casual = 'casual' THEN ride_length_minutes END) AS avg_ride_length_casual,
  AVG(CASE WHEN member_casual = 'member' THEN ride_length_minutes END) AS avg_ride_length_member
FROM `cyclistic-374321.cyclistic.combined_trips`
GROUP BY
  month_year
ORDER BY
  PARSE_DATE('%m %Y', month_year);
```

Calculating the total number of distinct casual and member rides in the *combined_trips* table. This gives a sense of the relative sizes of the two groups.

```
SELECT
  COUNT (DISTINCT CASE WHEN member_casual = 'casual' THEN ride_id END) AS casual_members,
  COUNT (DISTINCT CASE WHEN member_casual = 'member' THEN ride_id END) AS annual_members
FROM `cyclistic-374321.cyclistic.combined_trips`;
```

Calculating the monthly number of rides taken by casual and annual riders in the *combined_trips* table. This provides insights into how ride frequencies vary between the two groups over time.

```
SELECT
  FORMAT_DATE('%m %Y', started_at) AS month_year,
  SUM(CASE WHEN member_casual = 'casual' THEN 1 ELSE 0 END) AS num_riders_casual,
  SUM(CASE WHEN member_casual = 'member' THEN 1 ELSE 0 END) AS num_riders_member
FROM `cyclistic-374321.cyclistic.combined_trips`
GROUP BY
  month_year
ORDER BY
  PARSE_DATE('%m %Y', month_year);
```

Counting the number of distinct rides taken by casual and member riders on each day of the week in the *combined_trips* table. This helps in understanding how ride frequencies differ between the two groups by day of the week.

```
SELECT
  member_casual,
  weekday,
  COUNT(DISTINCT ride_id) as riders
FROM `cyclistic-374321.cyclistic.combined_trips`
GROUP BY member_casual, weekday
ORDER BY
  member_casual,
  CASE
    WHEN weekday = 'Monday' THEN 1
    WHEN weekday = 'Tuesday' THEN 2
    WHEN weekday = 'Wednesday' THEN 3
    WHEN weekday = 'Thursday' THEN 4
    WHEN weekday = 'Friday' THEN 5
    WHEN weekday = 'Saturday' THEN 6
    WHEN weekday = 'Sunday' THEN 7
  END;
```

Calculating the average number of rides per hour of the day for casual and member riders in the *combined_trips* table. This can help in identifying peak riding hours for each group.

```
SELECT
  EXTRACT(HOUR FROM started_at) AS hour_of_day,
  COUNT(CASE WHEN member_casual = 'casual' THEN 1 ELSE NULL END)/365 AS avg_rides_per_hour_casual,
  COUNT(CASE WHEN member_casual = 'member' THEN 1 ELSE NULL END)/365 AS avg_rides_per_hour_annual
FROM `cyclistic-374321.cyclistic.combined_trips`
GROUP BY hour_of_day
ORDER BY hour_of_day ASC;
```

Calculating the average ride duration for each day for casual and member riders in the *combined_trips* table. This can provide insights into how ride durations differ between the two groups on a daily basis.

```
SELECT
  EXTRACT(DATE FROM started_at) AS ride_date,
  AVG(CASE WHEN member_casual = 'casual' THEN ride_length_minutes END) AS avg_ride_duration_casual,
  AVG(CASE WHEN member_casual = 'member' THEN ride_length_minutes END) AS avg_ride_duration_annual
FROM `cyclistic-374321.cyclistic.combined_trips`
GROUP BY ride_date
ORDER BY ride_date;
```

Identifying the type of bike most frequently used by casual and member riders in the *combined_trips* table, excluding docked bikes. This gives insight into the bike preferences of the two groups.

```
SELECT
  member_casual,
  rideable_type,
  MAX(riders_count) AS riders
FROM (
  SELECT
    member_casual,
    rideable_type,
    COUNT(DISTINCT ride_id) AS riders_count
  FROM `cyclistic-374321.cyclistic.combined_trips`
  WHERE rideable_type != 'docked_bike'
  GROUP BY member_casual, rideable_type
)
GROUP BY 1, 2
ORDER BY 1, 2;
```

Extracting the coordinates and names of the start and end stations for rides taken by casual and annual members in the *combined_trips* table. This data can be useful for visualising and analysing the most popular routes or areas for casual riders.

```
SELECT
  start_lat,
  start_lng,
  start_station_name,
  end_lat,
  end_lng,
  end_station_name
FROM `cyclistic-374321.cyclistic.combined_trips`
WHERE member_casual = 'casual';
```

Exploratory Data Analysis

In the exploratory data analysis, several key categories of data were analysed:

1. Average ride length per day from beginning to end of data collected.
2. Average ride length per month of the year.
3. Number of riders per month of the year.
4. Number of annual and casual members.

5. Number of electric and classic bikes used by casual and annual members respectively.
6. Number of riders per day of the week.
7. Number of riders by hour.

Upon closer examination of each category, the following observations have been made. The analysis yielded several significant insights. Figure 2 shows that casual riders have a longer ride length than annual members. It seems that ride lengths fluctuate more for casual members depending on the day of the week and month of the year. On the other hand, the ride length of annual members does not change as greatly throughout the year.

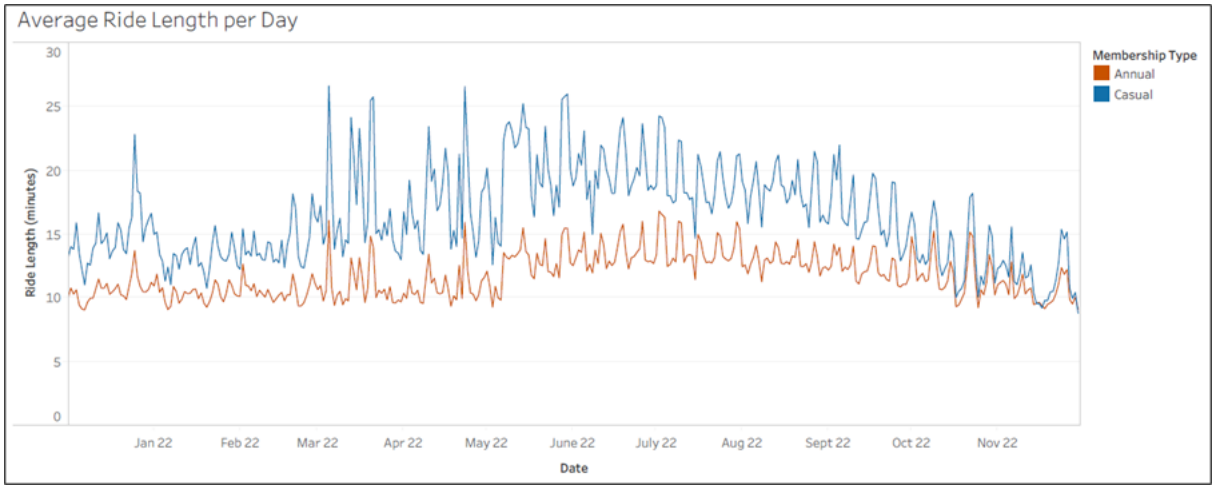

Figure 2: Average ride length per day over the data collection period.

There is a clearer image on how the average ride length varies across each month of the year for each type of member by looking at figure 3. Again, casual members have longer ride length than annual members throughout the year. The ride length for annual members peaks from May to August, while for casual members, it peaks from March to July. Therefore, both members' ride lengths start peaking during spring and drop during and after summer.

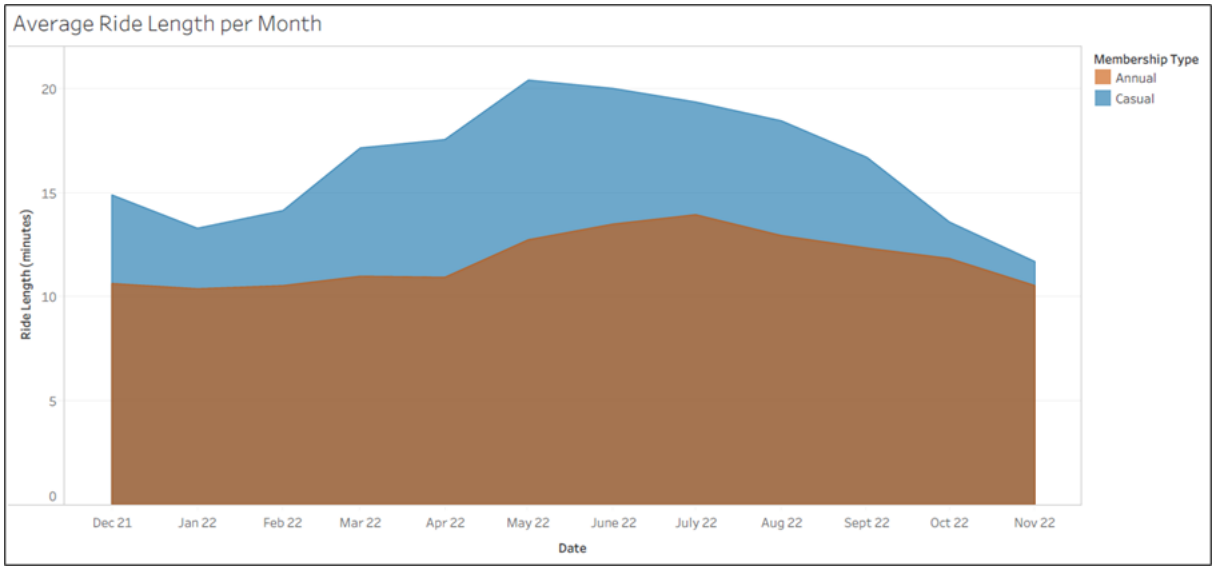

Figure 3: Average ride length for each month of the year.

In figure 4 the data reveals a higher overall number of annual riders compared to casual riders. The annual ridership displays a smoother curve from April to November, suggesting consistent use throughout these months. In contrast, casual ridership shows a less smooth trend, with a notable peak in July. This could indicate a seasonal surge of casual riders from April to June, followed by a significant drop thereafter.

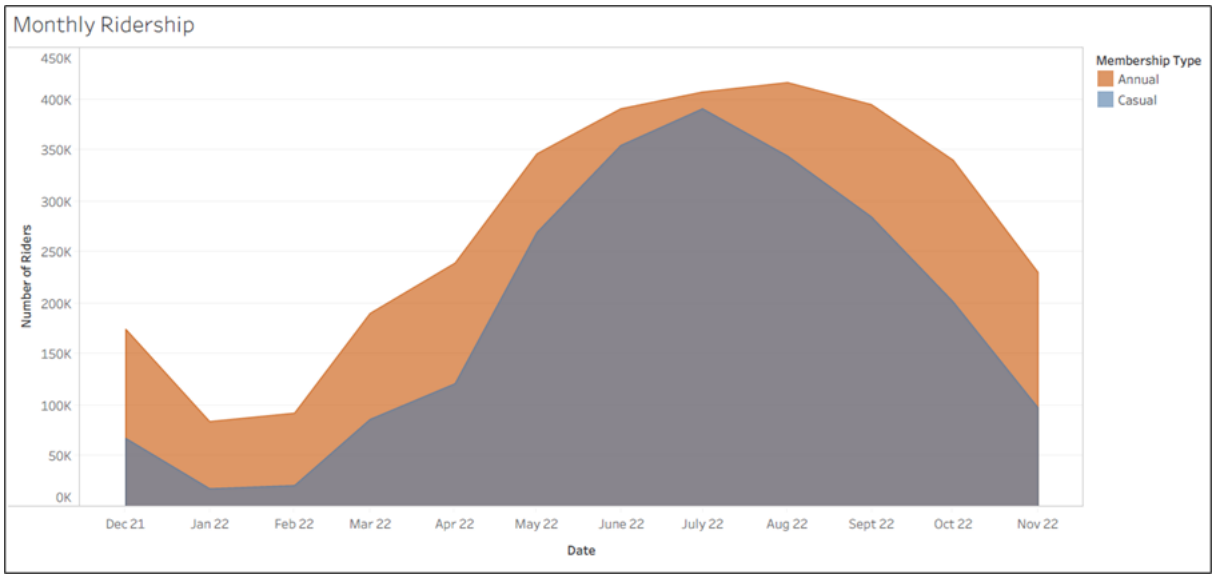

Figure 4: Number of riders for each month of the year.

Figure 5 clearly indicates a higher number of annual members compared to casual members.

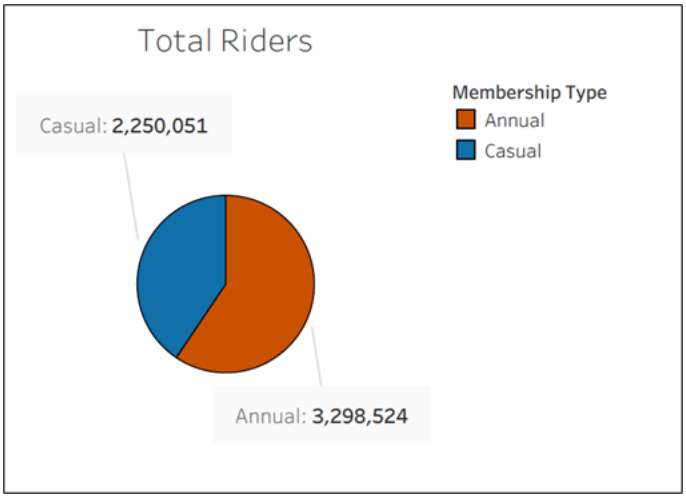

Figure 5: Number of annual and casual members.

Based on figure 6 there isn't a notable difference in the riders' preference between electric and classic bikes. However, classic bikes seem to be slightly more popular. To the contrary, electric bikes are more popular for casual members.

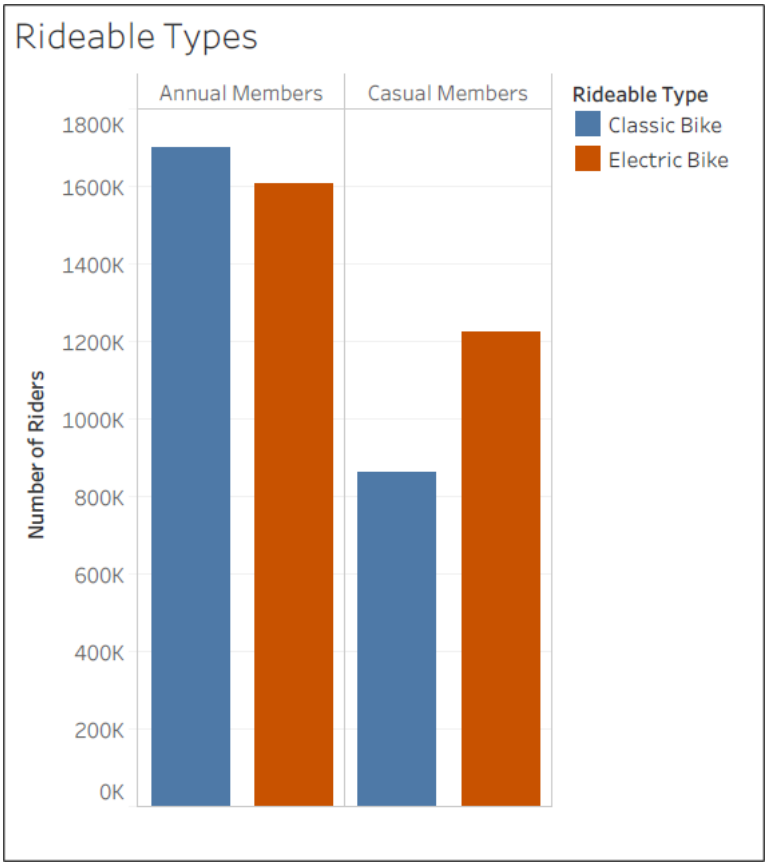

Figure 6: Comparison of electric and classic bikes usage by casual versus annual members.

From Monday through Friday, there are more annual riders than casual. Figure 7 illustrates that the number of casual riders slightly surpasses that of annual riders. On Sunday the amount of annual and casual riders is almost the same, however, there are slightly more annual riders than casual.

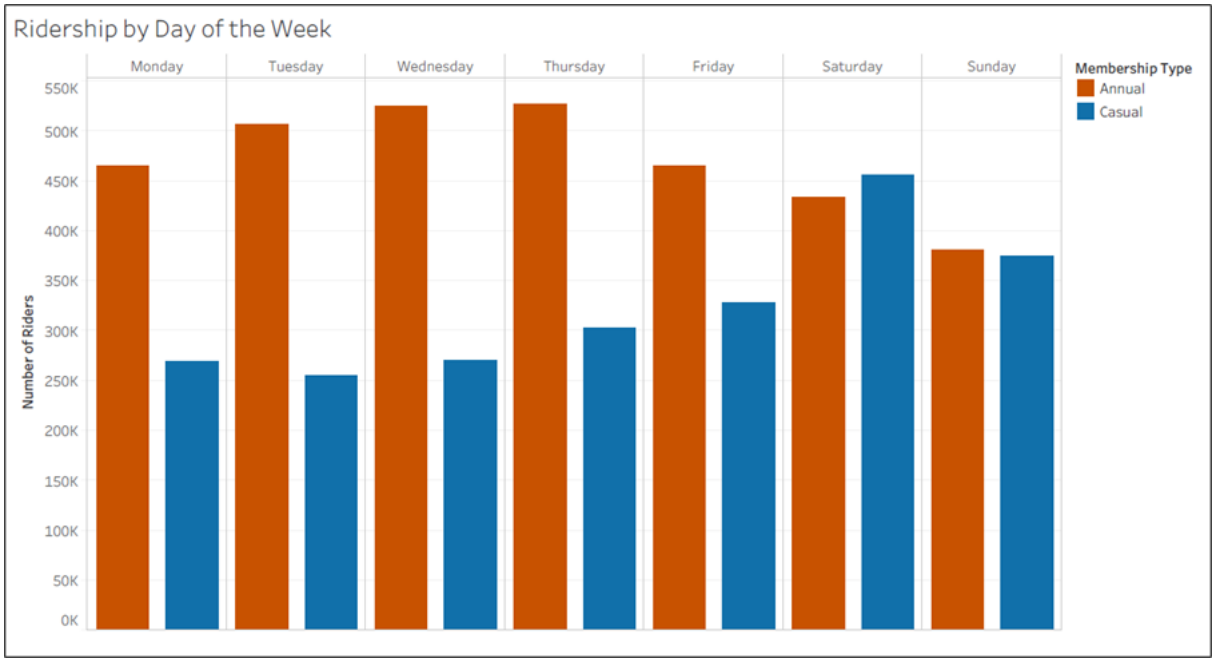

Figure 7: Number of riders for each day of the week.

Figure 8 demonstrates a higher number of annual riders compared to casual riders from 4 a.m. to 10 p.m., with peak ridership occurring at 8 a.m., 12 p.m., and 5 p.m. Conversely, casual riders experience their peak ridership only at 5 p.m.and there are slightly more riders during the times of 10 p.m. and 4 a.m.

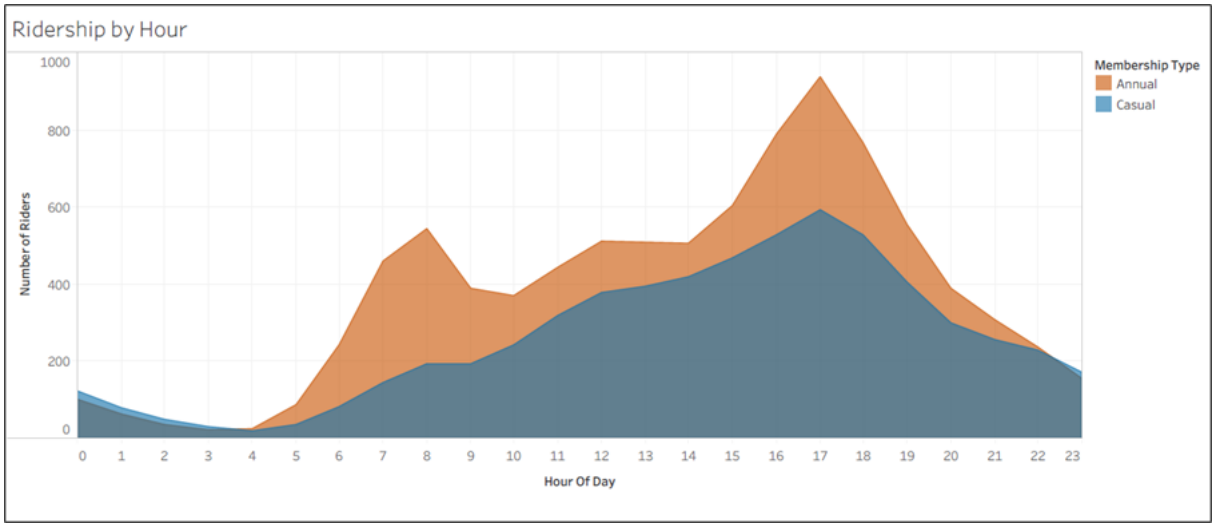

Figure 8: Number of riders by hour throughout the day.

Figures 9 and 10 depict the most popular start and end stations for casual cyclists. There seems to be a pattern where the top start stations are the same as the end stations. This indicates that casual riders return to their departure location at the end of their rides.

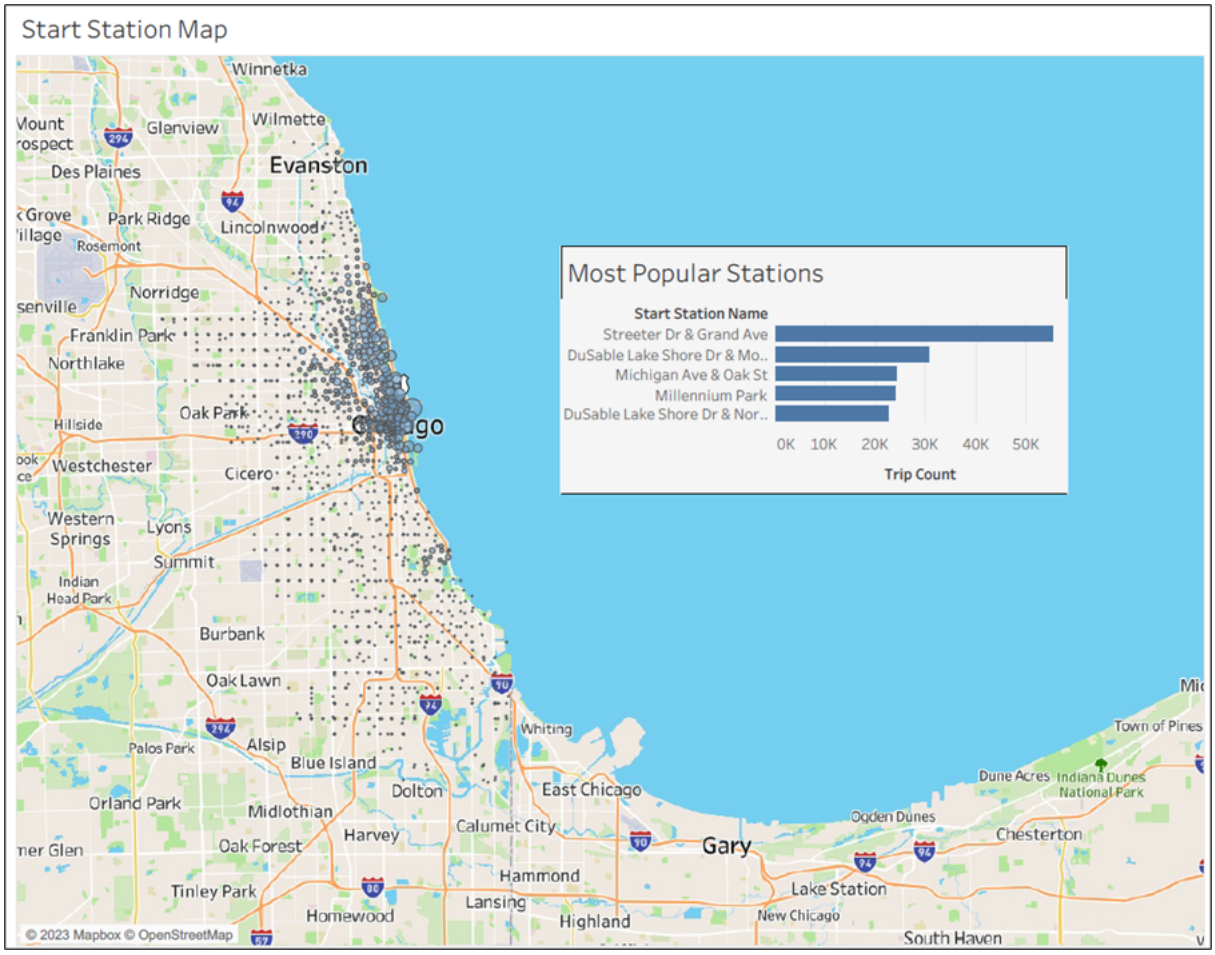

Figure 9: Most popular start stations for casual riders in Chicago.

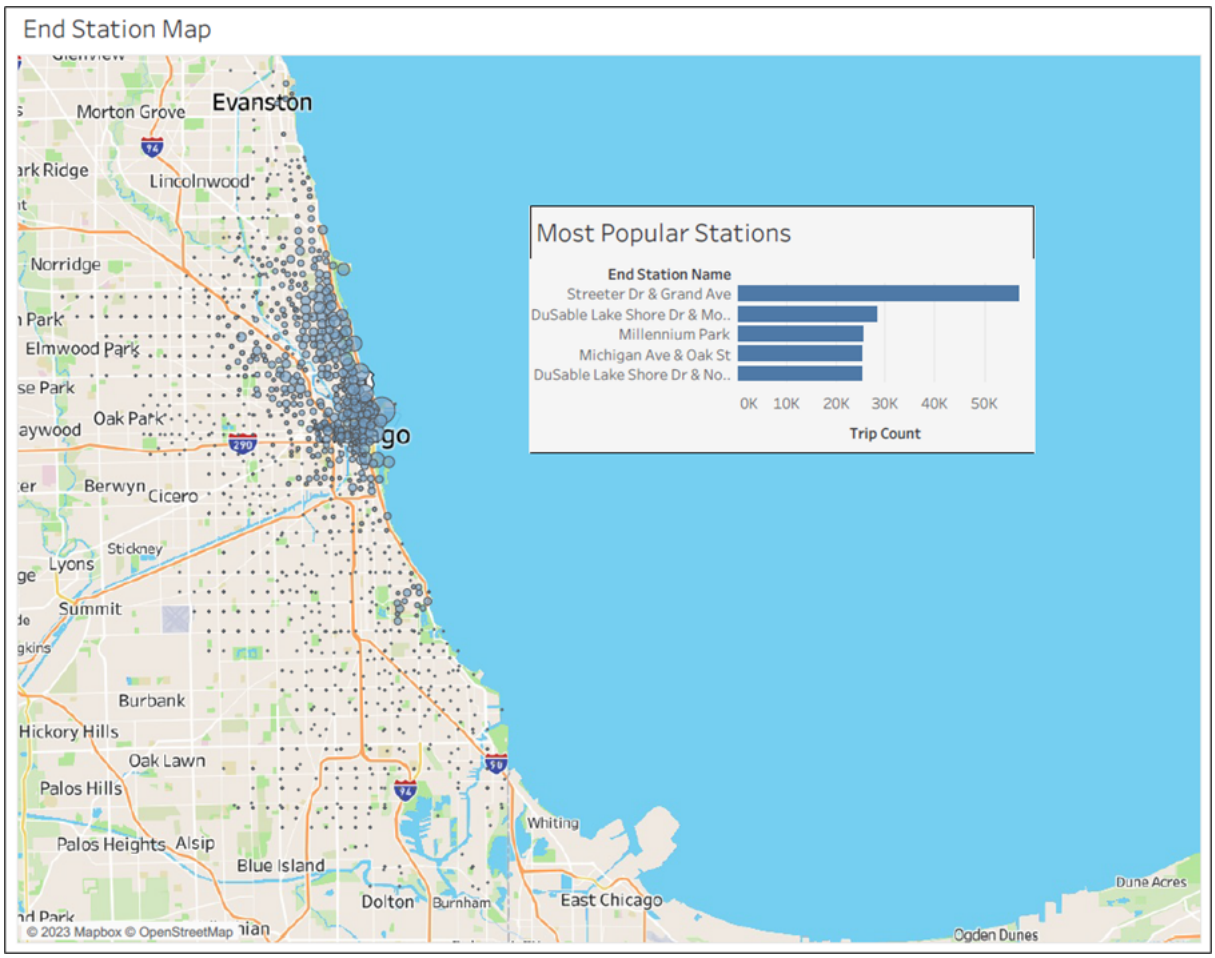

Figure 10: Most popular end stations for casual riders in Chicago.

Key Insights from Exploratory Data Analysis

Main Point	Description	Source
Casuals ride for longer	The ride length is longer for casual members.Bikes are ridden the longest on March and May-July.	Figure 2 Figure 3
Fewer casual riders	Casual rider numbers surge in spring/summer but are overall less than annual members.	Figure 4 Figure 5
Casual riders prefer electric bikes	Casual riders tend to use electric bikes more than classic bikes.	Figure 6
Casual peak during weekends	More casual riders on Saturdays than annuals, with over 300,000 riders from Friday to Sunday.	Figure 7
Casual's peak riding time	Their peak riding time is at 5 p.m. In contrast, annual members have multiple peak times.	Figure 8
Casuals return to start stations	Top start stations align with top end stations, indicating casuals often end where they start.	Figure 9Figure 10

Results

As indicated by the results of this study, the number of annual members is higher than that of casual yet casual riders prefer electric bikes and tend to ride for longer periods, especially from March and May through July. Additionally, there is a higher number of casual riders from Friday to Sunday centring around 5 pm. Such data can assist the marketing team in creating strategies for casual riders to switch to an annual membership. This conclusion takes into account the cleaning steps that were undertaken to complete the analysis.

Recommendations

The results uncovered that there are distinct riding patterns made by casual riders. The findings represent riders' data coming only for the Chicago-based bike-share company Cyclistic. Since the aim is to encourage Cyclistic's casual riders to switch to an annual membership, bike-share data from other companies were not analysed. The data indicate that casual members ride longer and more often during specific months of the year. That gives room for the marketing team to promote and advertise the annual membership during the off-peak period.

Some of the raw data had to be eliminated to make the results more accurate. For instance, changing the format of the time columns, deleting duplicate rows and eliminating outliers by using a combination of trimming and Winsorization. These steps might have affected the outcome of the results. Limiting the data to only that of Cyclistic may not provide a wider picture of distinct patterns among casual riders from other bike-share companies. Excluding other companies' data narrows the scope of the results and trends among casual riders.

As stated by the company, annual memberships are more profitable. Switching to an annual membership might be a more cost-effective option for casual members as they could ride more often and get better value. That way both Cyclistic and the annual members will be financially advantaged. Additionally, for future studies, it would be recommended to add data from other bike-share companies. This approach could provide more insight into casual rider habits across the industry and reveal potential influences that could motivate

them to switch to an annual membership. The analysis could be extended to different seasons or years to investigate whether the observed patterns hold true over longer periods.

A possible recommendation is to advertise Cyclistic's annual membership to potential casual members through social media platforms such as Instagram, Facebook, TikTok and Twitter. Information such as age range, location, gender (if available and appropriate for advertising), and interests (inferred from their activity on social media platforms) can be used to refine the target audience. Then the ads could be tailored based on other information gathered from casual riders. Here are some Ad recommendations:

1. An ad offering a discount for new riders or existing casual riders to switch to an annual membership. The ad could be released during the off-peak months of Winter and Autumn. That way the ad can motivate more people to ride during the off-peak months. It should be mentioned that the months of November through February are cold and possible snow may prevent riders to go for cycling so it's best to release the ad during Spring and right after Summer where the weather conditions are still desirable for a bike ride.
2. An ad promoting the unlimited use of electric bikes when someone becomes an annual member of Cyclistic. On that note, it is assumed that an annual membership will provide unlimited use of the bikes.
3. An ad highlighting the positive impacts using a bike has on the environment.
4. A referral promotion where both the referring member and the new member get a discount.
5. An ad that states the flexibility of use when becoming an annual member. For example, they can use the bike for leisure, work, or day-to-day activities if they become an annual member.
6. A round-trip promotion where annual riders can get a discount, rewards or points by returning to their departure station.

Any marketing strategies employed should comply with local regulations, and respect user privacy and data protection laws.

Ad Offering Discount

For the timing of release, it's recommended to air the ad primarily during transitional periods from off-peak to peak seasons. Specifically, introducing the discount ad at the end of Winter (late February) can encourage riders to re-establish their routines as Spring approaches. Similarly, launching promotions at the end of Summer (late August) might encourage ridership during the colder months. However, while retaining ridership in the colder seasons is important, it's crucial to understand that from November through February, there's typically a dip in riders due to colder conditions. As a result, even with end-of-Summer promotions, setting realistic expectations for uptake during these colder months is essential.

Ad Promoting Unlimited Use of Electric Bikes

The analysis shows that casual ridership peaks from Friday to Sunday around 5 pm. Therefore, offering special promotions during these times could encourage casual riders to consider annual memberships. For instance, offering a 'weekend warrior' discount could be an effective strategy to convert casual riders into annual members. This ad could also emphasise the unlimited use of electric bikes when someone becomes an annual member of Cyclistic.

Ad Highlighting Positive Environmental Impact

The analysis revealed a marked preference for electric bikes among casual riders. It is thus recommended that Cyclistic launches an advertising campaign offering unlimited access to electric bikes for users transitioning to annual memberships. This advertisement could prominently feature the environmental benefits of choosing electric bikes over traditional modes of transportation, such as cars, buses and motorcycles. By highlighting the large decrease in carbon emissions achieved by using electric bikes, the campaign can appeal to environmentally conscious riders and encourage them to upgrade to an annual membership.

Referral Promotion

A referral scheme where both the referring member and the new member get a discount could be beneficial. Referrals are a strong source of new customer acquisition in many businesses. It is recommended to design

an engaging referral scheme, potentially offering both parties a month of free riding or a significant discount on their next subscription.

Ad Promoting Flexibility of Use

Since casual riders seem to be riding for longer periods of time, they could be particularly valuable in spreading positive word-of-mouth about Cyclistic's service and attracting new members. To utilise this insight, the company should consider promotional efforts highlighting the flexibility and extended ride times. By doing so, casual riders can be motivated to share their positive experiences. Encouraging them with rewards for bringing in new members can further amplify this effect.

Round-Trip Promotion

Observations from the data indicate that casual riders frequently begin and conclude their journeys at the same station. Based on this behaviour, a recommendation would be to introduce a reward scheme for annual members who complete round trips. By accumulating points for each completed round trip, members could redeem them for discounts or perhaps even gift vouchers. This not only encourages repeated use but also potentially attracts new members interested in joining this unique rewards system.

Conclusion

The analysis has revealed key insights into Cyclistic's casual riders. Despite a higher number of annual members, casual riders exhibit longer ride durations, a preference for electric bikes and peak activity during weekends around 5 pm, especially from March to July. These findings are crucial for creating strategic marketing campaigns.

The main goal was to derive strategies that can motivate casual riders to switch to annual memberships, enhancing Cyclistic's profitability. Based on the data, campaigns highlighting discounts, unlimited electric bike usage, environmental benefits, convenience and a referral program could be particularly influential.

It's essential to note that some raw data was removed during cleaning and the analysis was restricted to Cyclistic's records, potentially excluding broader industry trends. Future research could incorporate data from other bike-share entities for a more holistic view.

In spite of these constraints, this study offers actionable recommendations for Cyclistic's marketing endeavours, aiming to transition casual riders to annual members. All marketing strategies should remain cognizant of regulations, user privacy, and data protection norms.