**Google Data Analytics Capstone: Cyclistic Case Study**

A blue circle with a person riding a bike

Description automatically generated

**Business Overview & Task Statement**

Cyclistic is a notable bike-sharing service with a network of 5,824 geo-tracked bikes located at 692 stations throughout Chicago. These bikes provide the convenience of being picked up from one station and returned to another. In its early marketing days, Cyclistic aimed to appeal to a broad spectrum of consumers, supported by its diverse pricing models: single-ride passes, daily passes, and annual memberships. Those who opt for single-ride or daily passes are categorized as casual riders, while the annual subscribers are termed Cyclistic members.

A recent assessment by the marketing and finance teams emphasized that the annual members play a vital role in boosting the company's profitability. As such, there is a growing consensus that future growth largely depends on augmenting the number of annual membership.

This report examines the usage trends from August 2022 to July 2023 through the process: ASK – PREPARE – PROCESS – ANALYZE – SHARE - ACT, and proposes marketing strategies to convert casual riders into annual members.

**ASK**

1. How do annual members and casual riders use Cyclistic bikes differently?

2. Why would casual riders buy Cyclistic annual memberships?

3. How can Cyclistic use digital media to influence casual riders to become members?

**Prepare(Excel, BigQuery)**

The dataset available in the "[Index of bucket "divvy-tripdata"](https://divvy-tripdata.s3.amazonaws.com/index.html) is a public offering generously shared by Motivate International Inc. under a specified license. Ensuring user privacy, the dataset excludes personal details, making it impossible to identify individual riders' usage patterns.

Columns in the dataset capture information regarding the start and end stations, the start and end times of the rides, and the categorization of riders as either members or casual users. Columns indicating the latitude and longitude of the stations also prove useful for visual data representation.

1. A screenshot of a computer

   Description automatically generatedTwelve distinct Excel files, containing data spanning from August 2022 to July 2023, were consolidated using the following procedure:
2. Data exploration

* The total number of rows was verified to be 5,555,694

A black arrow pointing to a black arrow

Description automatically generated

* The table provided below outlines all the columns along with their respective data types, and the primary key is ‘ride\_id’.

A screenshot of a computer

Description automatically generated

**Process(BigQuery)**

1. Rows containing any NULL values were removed, as executed with the following code:

A screen shot of a computer

Description automatically generated

1. All duplicate entries under the 'ride\_id' column were eliminated using the subsequent procedure:

A screen shot of a computer code

Description automatically generated

1. Verification was conducted to ensure that no duplicate 'ride\_id' exists within the 'CombinedData' table.



A screenshot of a computer

Description automatically generated

1. Columns for 'ride length for duration of the trip,' 'day of week,' and 'month' were added. Additionally, trips with durations shorter than a minute or longer than a day were filtered out to enhance the analysis.

A screenshot of a computer program

Description automatically generated

1. Data after cleaning

* The data now comprises 4,016,127 rows, indicating that 1,539,567 rows were deleted during the data cleaning process.

A black arrow pointing to a white box

Description automatically generated

* After the data cleaning process, we have 2,459,221 members and 1,471,203 casual riders.

A screenshot of a computer

Description automatically generated

* The following table presents the total number of trips for each bike type during the specified period.

A screenshot of a computer

Description automatically generated

**Analyze and Share(Tableau)**

A black and white text

Description automatically generated

A group of blue orange and black squares

Description automatically generatedA screenshot of a graph

Description automatically generatedA blue circle with white text

Description automatically generatedA pie chart with numbers and a number of percentages

Description automatically generated

Members comprise 63% of the total, with casual riders making up the remaining 37%. In the breakdown by bike type, each chart displays the percentage out of the overall total. The classic bike emerges as the most popular choice, succeeded by electric bikes. Docked bikes, on the other hand, are least preferred and are exclusively chosen by casual riders.

In the following sections of this report, a detailed examination will be undertaken on the distribution of trips, segmented by months, days of the week, and individual hours of the day.

A black and blue text

Description automatically generated

A graph showing a line of blue and orange lines

Description automatically generated with medium confidence

A graph with blue and orange lines

Description automatically generated

Months: Regarding trips taken on a monthly basis, both casual users and members show similar patterns, with an increased number of trips during the spring and summer months and a decline in the winter. The difference between casual and member trips is most narrow in July, a summer month. It is evident that, in terms of total trips per month, the number of total trips by members surpasses that of casual riders. However, based solely on this data, distinguishing the usage pattern between the two is challenging.

Days of Week: When breaking down the weekly travel trends, it is clear that members ride more on weekdays, while casual riders tend to prefer the weekends.

Hours of the Day: Members typically have two primary travel spikes: an early morning surge from 6 am to 8 am and an evening burst between 4 pm and 6 pm. In contrast, casual riders' trips gradually climb as the day progresses, peaking in the evening and then diminishing thereafter.

A screenshot of a phone

Description automatically generated

When evaluating the total ride duration (in minutes), casual users exhibit a ride length that is approximately 10% longer than members. Furthermore, the average ride duration for casual riders exceeds that of members by 11 minutes. Additionally, as illustrated in the preceding charts, the average ride duration for members remains relatively stable throughout the period. In contrast, casual riders exhibit more variation in their ride durations, with noticeable increases in the spring and summer seasons and declines during the fall and winter.

A graph showing the growth of the week

Description automatically generated with medium confidenceA graph of a line

Description automatically generated with medium confidence

Based on the earlier observations, we can surmise that members predominantly utilize bikes for commuting to and from work during weekdays, with distinct peak times in the morning and evening. On the other hand, casual riders appear to ride bikes at a steadier pace throughout the day, especially on weekends, likely for recreational activities. Furthermore, it can be deduced that while casual riders take trips less often, they tend to have longer ride durations when they do. Notably, both categories are most active during the summer and spring seasons.



A screenshot of a map

Description automatically generated

A group of blue orange and black squares

Description automatically generatedA map of a city with orange dots

Description automatically generated

To better comprehend the distinctions between casual and member riders in Chicago, the starting and ending station locations on the city's map should be analyzed. By applying filters, only stations with over 10,000 trips are considered to derive and observe clear distinction in usage pattern between two user groups.

On the map illustrating both starting and ending stations, it is observable that member usage is evenly spread across the city, especially around commercial areas, residential areas, and universities. On the other hand, casual riders primarily frequent stations close to the lakeside, such as those near museums, parks, and eateries.

**Key Takeaways**

Here is a recap of the response to the question: How do annual members and casual riders use Cyclistic bikes differently?

|  |  |  |
| --- | --- | --- |
|  | **Members** | **Casual Riders** |
| Bike Type | Traditional bikes represent 63%, while e-bikes make up the remaining 37%. | 50% are traditional bikes, 42% are e-bikes, and docked-bikes make up 8% |
| Total Trip per Month | The count of both user types rises in spring and summer but decreases in winter. | |
| Total Trip per Day, Hour | Utilize the service for commuting to work or school, indicating peak usage on weekdays, particularly between 6-8AM and 4-6PM. | Use the service on weekends, with a fairly consistent distribution throughout the afternoon hours |
| Ride Length(mins) | AVG ride length =12  Total ride length = 33.4 M | AVG ride length = 23  Total ride length = 30.5 M |
| Number of Uses per Location | Use the service all over the city, mainly near offices, residential areas, and schools. | Use the service near lakeside spots like museums, parks, and restaurants for weekend leisure. |

**Strategy Recommendations**

1. To examine the current pricing and develop a new pricing strategy, in order to encourage casual riders to join membership, ensuring the annual membership appears more cost-effective than single-ride and day passes.
2. To add membership benefits: advance reservation during peak time or season, late charge waiver, etc.
3. To consider another annual membership product development reflecting the needs of current casual riders, since, in terms of ride length, the total ride length of casual riders surpasses that of the current members.
4. Launch marketing campaigns or events at places where casual riders frequent to boost engagement. Eg) Streeter Dr & Grand Ave