Three questions will guide the future marketing program:

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

The primary difference between annual members and causal riders is their usages. The annual members seem to be targeted towards individuals who use the Cylistic bikes for work purposes. They may be first time office goers such as university students or middle class lower level employees that may not be able to afford mediums such as cars, bikes or public transport.

The casual riders, on the other hand, seem to be a customer segment that consists of individuals who bike for leisure. This can include tourists or travelers who wish to travel Chicago through the roads, office goers who may have their vehicles in repair or might have to renew their train cards.

2. Why would casual riders buy Cyclistic annual memberships?

Casual riders who have used Cylistic products more than once in a short span of time may recognize the frequency of their cycle use. They may notice that instead of taking an expensive full-day or single-ride pass to work everyday, it is much more cost-efficient to consider Cylistic’s annual membership.

Here are some benefits that casual riders can get if they take the membership model:

* Cylistic would provide external benefits such as ride discounts, benefits such as personalized route navigation for members, preferential cycles with better quality materials like softer seaters, smoother gears for members, discounted on affiliated brands which may include clothing , cosmetics or entertainment businesses.

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

Cylistic can make use of social media advertising as a tool to influence riders to become members. They can post rider testimonials and show how their current members benefit from Cylistic’s annual membership plan. They can show statistics for their annual member customer group wherein they can showcase the finances annual members save, how they preserve the environment( to provide an environmental backing) and some membership specific perks.

**Main Question: How do annual members and casual riders use Cyclistic bikes differently?**

You will produce a report with the following deliverables:

1. A clear statement of the business task

2. A description of all data sources used

3. Documentation of any cleaning or manipulation of data

4. A summary of your analysis

5. Supporting visualizations and key findings

6. Your top three recommendations based on your analysis

**CASE STUDY ROADMAP:**

**ASK**

PRIMARY BUSINESS TASK:

Creating a marketing strategy that helps users to fulfill three goals. Firstly, Cylistic has to determine the cycle usage differences between annual members and casual riders in which they determine the type of user, how much they are willing to pay for the bikes and their frequency. Secondly, Moreno and the marketing need to identify incentives for Cylistic casual riders to consider annual membership. This includes possible monetary benefits, improved convenience and more specific service for the annual members. Finally, the team needs to find how they can use digital media outlets such as Instagram, Snapchat , Twitter and so on to entice casual riders to become annual members.

CONSIDER KEY STAKEHOLDERS:

The key stakeholders for this new marketing strategy are the customers, the marketing team at Cylistic and the director of the marketing team, Lily Moreno. These new marketing campaigns and slogans will be targeted towards pre-existing casual riders as well as prospective members looking for a means to travel to work in a cost-efficient and environmental manner. To develop this new marketing strategy, we need to research more about Cylistic’s target audience for this campaign which are the casual riders. Cylistic needs to conduct a comprehensive research about the current marketing strategy which is arguably

TYPE OF PROBLEM:

The type of data analysis problem we are facing in this case is called identifying themes. It is important to understand the categorization of the annual members and the casual riders. Cylistic could categorize both these types of customers even further by adding criteria such as additional revenue streams generated by each type of customer, frequency of travel by each type of customer and business costs as a result of annual members vs casual riders. Diving into analysis on these topics will help me, the data analyst, our data analyst team and Lily Moreno realize the potential structure and content of the new marketing campaign

**Guiding Questions:**

1. What is the problem you are trying to solve?

The main problem to solve is to identify whether Cylistic benefits more from the annual members or from casual riders. If they benefit from annual members, how can the casual riders be effectively converted to gain that extra revenue.

1. How can your insights drive business decisions?

The data analysis and subsequent insights will help Moreno and the marketing team make informed decisions for their upcoming marketing campaign where they can finally decide to shift their goal from general awareness to ads/posters targeted towards casual riders to convert to annual members.

**PREPARE**

User data from the past 12 months, April 2020 - December 2021 has been used for this analysis. Each data set is in csv format and details every ride logged by Cyclistic customers. This data has been made publicly available via license by Motivate International Inc. and the city of Chicago available [here](https://ride.divvybikes.com/data-license-agreement). All user’s personal data has been scrubbed for privacy.

Guiding Questions:

1. Where is your data located?

The data that I downloaded from the Prepare section is located on the Windows SSD(C:) Drive on my computer.

1. How is the data organized?

All the monthly bike data is stored in the “Capstone” folder within the main Google Capstone Project folder. The excel sheets are organized according to their respective dates starting with the data from April 2020 until December 2021.

1. Are there issues with bias or credibility in this data? Does your data ROCCC?

To check the ROCCC of the data:

Reliability: The data is reliable as the sample size is all 5824 bikes recorded and owned by Cylistic.

Originality : As Cylistic is a fictional company created for the use of the case study, the data is **not original** to Cylistic and has been originally created by Lyft Bikes and Scooters Inc running within the state of Chicago.

Comprehensiveness : The data is quite **comprehensive**. It contains 13 different columns consisting of data such as ride\_id to uniquely identify each of the 5824 across their start and end stations denoted through start\_station\_name and start\_station\_id and the end\_station\_name and end\_station\_id.Data entries go up to 34000 entries within one month.Because of privacy concerns, Lyft has not provided transaction data and hence financial transactions have not been recorded. This may reduce a little comprehensiveness.

Current: The data is recorded for the months between April 2020 - December 2021and hence this is fairly current data. This data is **current**.

Cited:This data is **cited** because it has been collected from Motivate International Inc and the city of Chicago and has a license agreement attached in the description above as well.

1. How are you addressing licensing, privacy, security, and accessibility?

The data will be saved on a local folder on my Computer and will also be saved in a file created on Google Drive. It will not be shared in its raw form and will only be displayed in the final report as code snippets, tibbles , visualizations and graphs.

1. How does it help you answer your question?

The data is the key to answering questions. Through analysis, we can find answers to many effective questions that will help the marketing team and Lily Moreno understand the usage differences between annual members of Cylistic and the casual riders.

1. Are there any problems with the data?

No, but there are problems such as duplicate entries, blank fields and a lack of data organization which will be tackled when we clean and prepare the data in R.Other than that, the data is credible because of the use of start and end coordinates. The main challenge is to find substitutes for missing data entries in the columns. The need to be estimated from whatever other information is available in that row.

Key Tasks:

1. Check the data for errors
2. Choose your tools
3. Transform the data so that you can work with it effectively
4. Document the cleaning process

**ANALYZE**

**Guiding Questions**

**● How should you organize your data to perform analysis on it?**

I have performed certain functions to ensure that the data is formatted and organized properly before analysis is performed. This includes creating a data frame which binds all the individual datasets.

● **Has your data been properly formatted?**

To ensure proper formatting, I have used the janitor and tidyr packages. This includes functions such as remove\_empty and remove\_missing. I have also used the distinct function to remove any duplicate values as well.

● **What surprises did you discover in the data?**

The most surprising element about the data is the sheer lack of formatting. I was surprised to see that I had to create additional columns, clean a great deal of rows and columns to finally get a proper clean dataset to analyze. I also saw surprises in data such as the formatting of the start\_at and end\_at columns.

● **What trends or relationships did you find in the data?**

I found a lot of important trends in the data. I first found that there was a trend in the total rides on the weekdays and weekends where weekends(mostly Saturday) sees more ride traffic. Member rider frequencies usually peek out at mid-week(Wednesday and Thursday) while Casual riders are seen usually at 460k to 530k during Saturday and Sunday. I also see more rides after May and peeks at June and July. Ride lengths in general and significantly greater(almost double) for member rides with average rides lasting about 30 minutes while casual riders usually cap out at 15-16 minutes. Furthermore, we see a trend of classic and electric bikes among the members.

**● How will these insights help answer your business questions?**

All these insights, relationships and trends are very useful for determining an effective marketing strategy for Cylistic. This will help Lily Moreno and the team to create an advertising campaign for the existing members as well as find effective marketing months, weekdays for casual riders to convert to Cylistic members.

Key tasks:

1. Determine the best way to share your findings.

2. Create effective data visualizations.

3. Present your findings.

4. Ensure your work is accessible.

**SHARE**

**Guiding Questions**

* **Were you able to answer the question of how annual members and casual riders use Cyclistic bikes differently?**

Yes, through performing the data analysis of the Cylistic Data Frame, I could get a more clearer idea into the primary guiding notion of this analysis: finding the discernible differences between Cylistics’s annual members and casual riders and finding out which customer group was more profitable and beneficial for Cylistic and finally devising an appropriate marketing strategy for those targeting that group in an effective manner.

* **What story does your data tell?**

The data shares a story of a bike-sharing company that is in a marketing dilemma. The data indicates that Cylistic

* **How do your findings relate to your original question?**
* **Who is your audience? What is the best way to communicate with them?**
* **Can data visualization help you share your findings?**
* **Is your presentation accessible to your audience?**

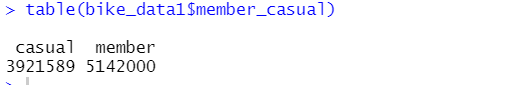
**KEY FINDINGS:**

**Laying the foundation:**

1. **Summary of the ride\_length field** 

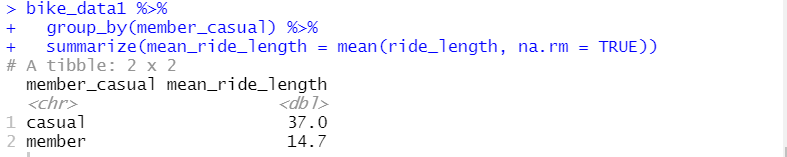
* This function provides the five-point summary including the mean ride length of 24.33 and the median of 12.92. This means that the data is a positively skewed graph with a longer tail on the right hand side. This generally means that ride length has a lot of outliers in the first quarters suggesting a general tendency of shorter ride times

1. **Table showing distribution of rider types between member and casual**



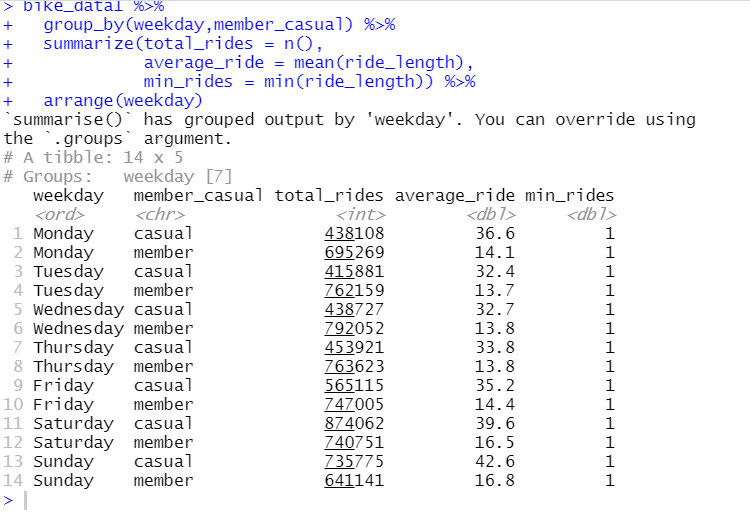
* This table function is integral to our findings as it suggests the distribution of data among the two main customer groups for Cylistic. It is clear that Cylistic has more annual members than casual riders. Further analysis looks at the benefits/costs of converting the casual riders into members.

1. **Mean ride lengths for members and casual riders**



* This is another important data category for the analysis as it looks at the mean ride lengths for casual riders and members. It is apparent that the casual riders have, on average, over 150% higher ride times than the annual members. This might suggest that annual riders use the transport for fixed routes such as going to the office or the grocery malls and casual riders have a more random riding pattern. Marketing can focus on providing distance based discounts to casual riders who opt for the annual membership and also give discounts to them for their first 5 rides.

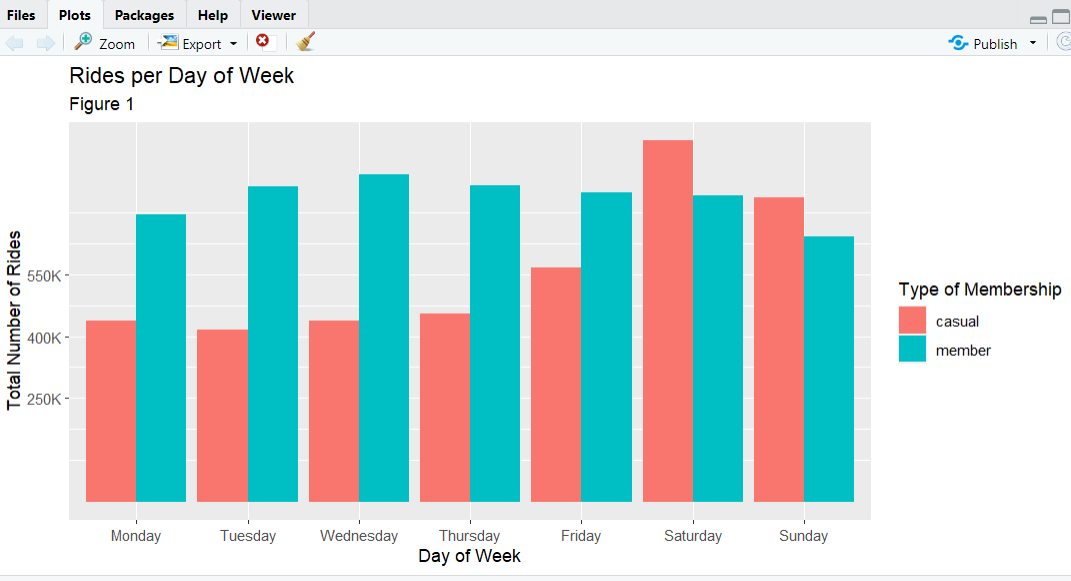
1. **Total Rides over the Weekdays**



* This data table showcases the total\_rides , average\_ride and min\_rides for all the 7 days of the week. The data is categorized by the classification between annual members and casual riders. As apparent from the mean ride lengths demographic above, annual members almost always have more total\_rides compared to the casual riders.

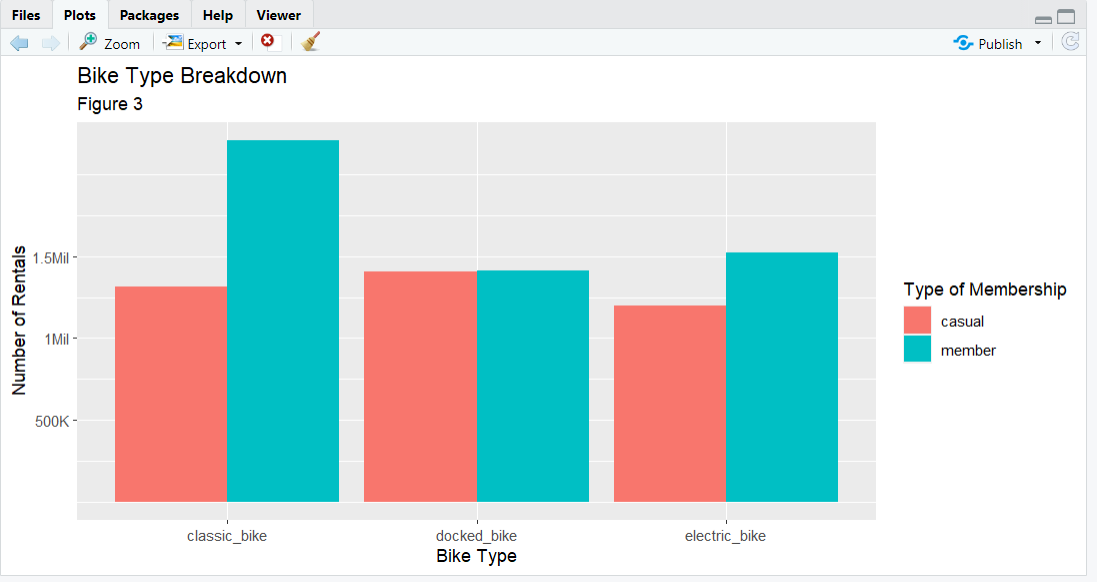
| Day | Higher total\_rides category | Percentage Difference |
| --- | --- | --- |
| Monday | Members | 58 |
| Tuesday | Members | 83 |
| Wednesday | Members | 80 |
| Thursday | Members | 68 |
| Friday | Members | 32 |
| Saturday | Casuals | 18 |
| Sunday | Casuals | 15 |

The table above summarizes the consolidated findings from the dplyr visualization snippet pasted above. The table shows the changes in total\_rides among annual members and casual riders as a pattern throughout the week. In the first 4 days until Thursday, we see significantly higher total\_rides among members which drops to 32% on Friday. In contrast, on the weekends we see higher total\_rides among casual riders. This is indicative of the weekly work traffic from Monday to Friday and a sudden increase in leisure traffic on the weekends.

****

**Fig 1: Data Viz of Rides per Day of Week**

The graph above shows the visualization of the total rides throughout the week. It also clearly shows higher total rides from Monday through Friday and higher total rides among casual riders on Saturday and Sunday



**Fig 2: Data Viz of Bike Type Breakdown**

* The visualization above shows the types of bikes normally preferred by either the casual riders or the annual members. In general, annual members prefer the classic bikes by more than 1 million in rentals, the members and casual riders are about equal for docked bikes at 1.8 million each and members prefer the electric bikes for rentals by 250k.

Marketing Recommendations:

* Cylistic can provide deals on single-passes and day-passes during the weekdays(Monday to Friday) to boost casual rider rides to match up with the annual member rides
* Cylistic can target casual riders on the end of the spectrum(ones with the higher ride times) with distance discounts for their first 10 rides if they join as a member
* They can include flat rates per distance in their current pricing model to entice casual riders who want to travel farther distances to actually join the annual membership plan