Ask

What is the problem you are trying to solve?

The main objective of this project is to construct profiles for the different users, annual and casual, and perform analysis on these profiles. This is to help build marketing strategies that are effective in turning the casual riders into annual members.

How can your insights drive business decisions?

The insights will help explain the reasons behind the casual ridership and give guidance to the marketing team for ways to increase annual members. Our project is focused on building profiles and analyze them, therefore, the insights can also help with decisions based on the different users.

Prepare

Where is your data located?

The data is located in an AWS S3 Bucket that was linked through the case study. The data is also available in a Kaggle dataset.

How is the data organized?

The directory for the data includes datasets for each month and includes all data of the year 2021. All are in a csv format.

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

There isn't a problem with bias because the users are the clients for this company and the clients represent the whole population for the dataset. Credibility is not an issue because of the same reason. With revision, it is also evident that the data is reliable, original, comprehensive, current, and cited.

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

The company owns the license for the datasets, and there isn't any personal information about the riders in the datasets; there aren't issues regarding privacy, security, and accessibility.

How did you verify the data's integrity?

All the datasets have consistent columns even the most updated ones. Also, each column has the correct data type.

How does it help you answer your question?

There may be some insights on the riders and their preferred method for renting.

Are there any problems with the data?

There aren't any key problems with the data in regards to our main objective. We have the key resources to produce an insightful project. However, climate data, station data could be useful for insights.

Process

What tools are you choosing and why?

Because of the large dataset, I will be using R for this project and also to gain experience.

Have you ensured your data's integrity?

Yes, data integrity is ensured with consistent columns and correct data types.

What steps have you taken to ensure that your data is clean?

Explored through the dataset and removed the duplicated values. Also, some of the columns were formatted to the correct format.

How can you verify that your data is clean and ready to analyze?

The notebook can verify that the data is clean and ready for analysis.

Have you documented your cleaning process so you can review and share those results?

All steps were documented in the R notebook and can be easily referenced back.

Analyze

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

In order to ensure consistency, we have organized the dataset into a single CSV file by concatenating all the datasets.

Has your data been properly formatted?

All columns are consistent with the correct data type.

What surprises did you discover in the data?

Some notable surprises were the differences between members and casual users, and how they differ when using our bikes. Other surprises were with ride length and day of week.

What trends or relationships did you find in the data?

There are more rides by members than casual users, however, there are more casual users during the weekend (Saturday and Sunday). There is a consistent number of users throughout the week, and it seems as if they use them on a schedule such as for transportation for work. Members have a shorter average ride length than the casual users. There are fewer rides during the colder months especially for casual riders; there are still more rides from members during colder months.

How will these insights help answer your business questions?

We have a better understanding of how members differ from casual users. We have also built profiles based on their preferences and their riding habits.

Share

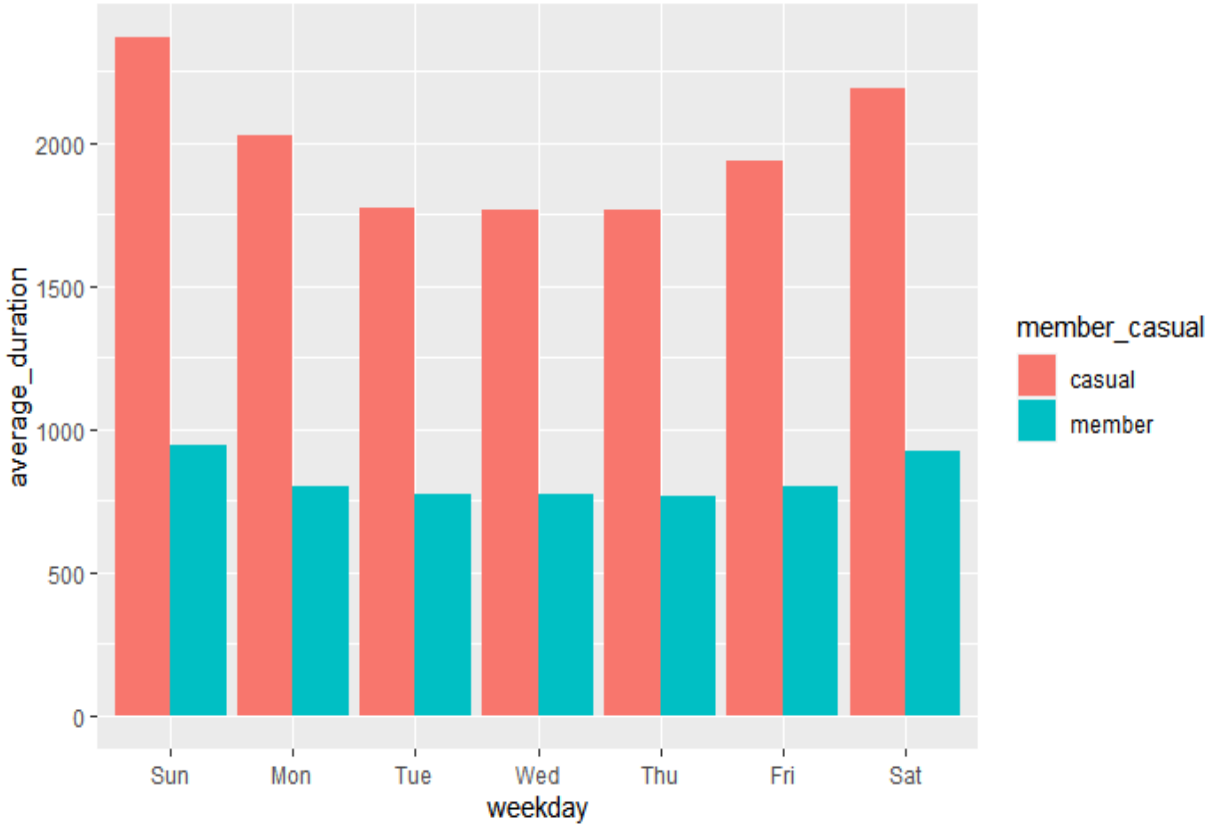
Were you able to answer the question of how annual members and casual riders use Cyclistic bikes differently?

Analysis from the data clearly shows the differences between the users and how they use our service.

What story does your data tell?

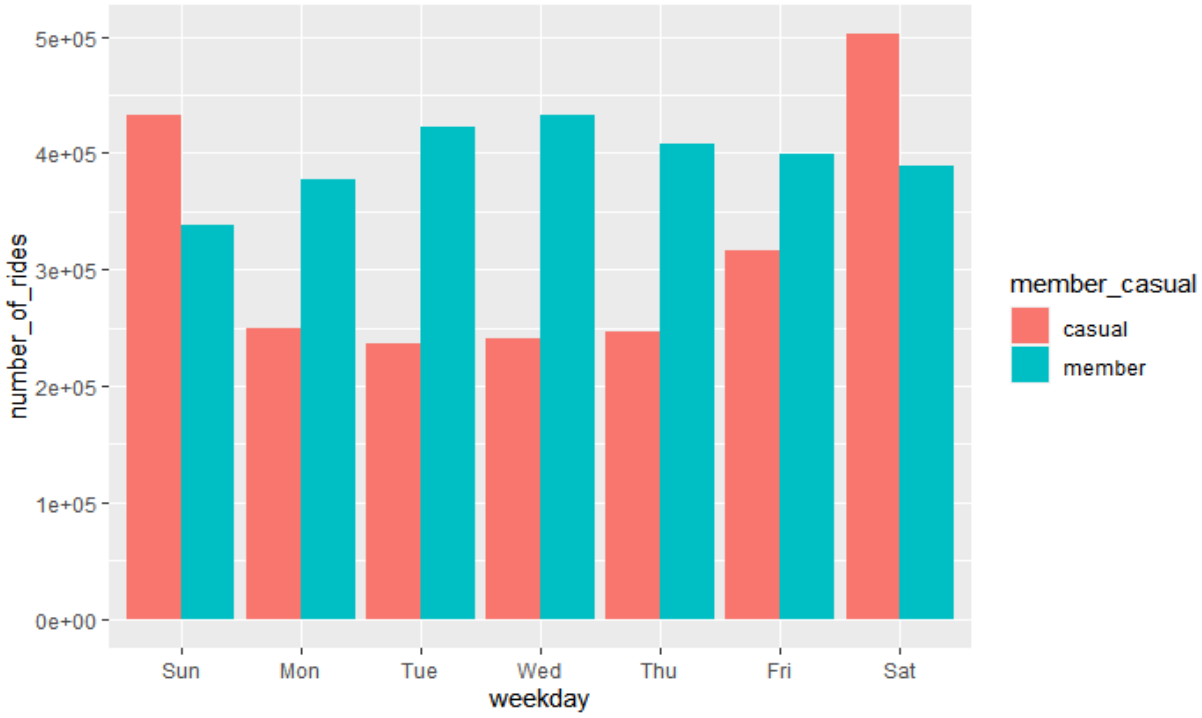
The graph below (Chart 1) explains how members have consistent ride durations and usually a lot lower than casual users. This explains that members are using our service as a means for transportation and are used for specific schedules for their needed route (like going to/from work).

Member/Casual – Day of Week and Average Duration (Chart 1)



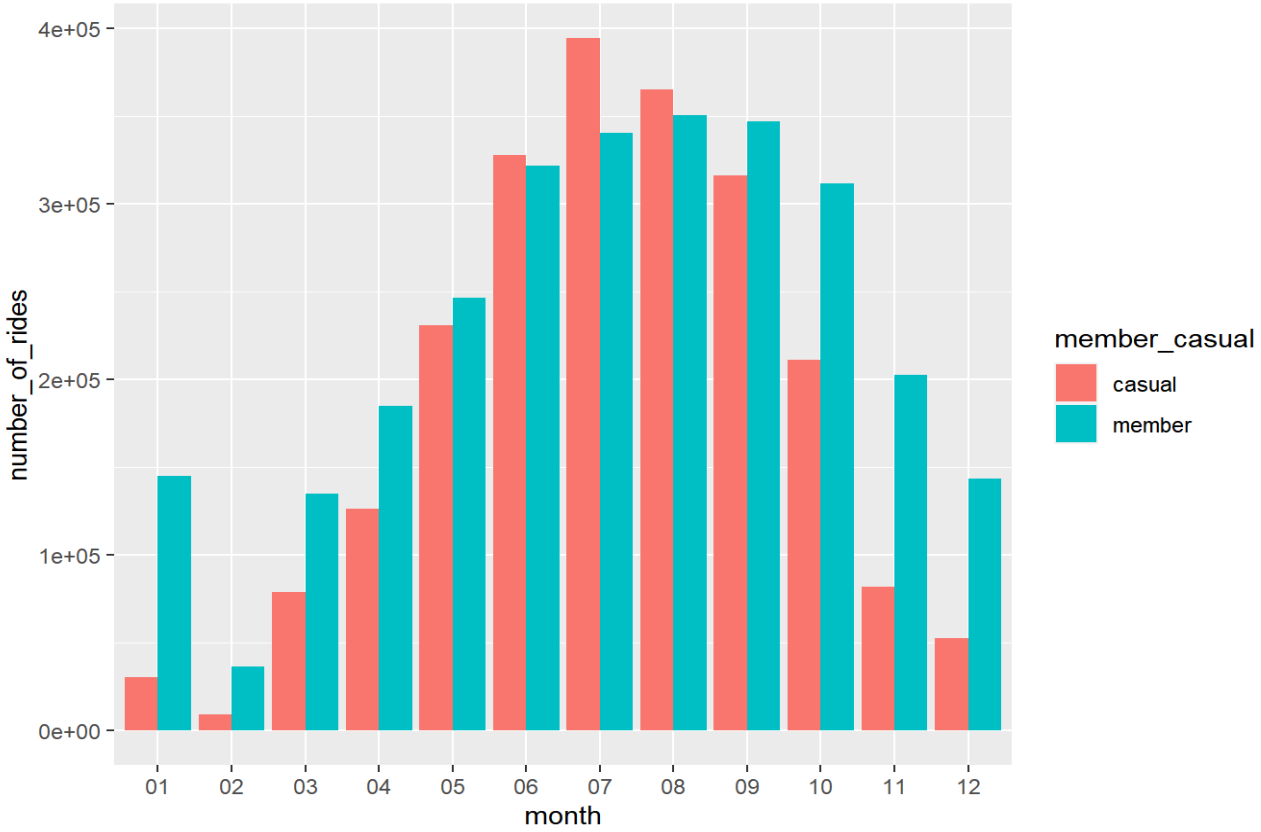
The graph below (Chart 2) shows that there are more casual users during the weekends (Saturday and Sunday) and a consistent number of member users throughout the week. This also explains how members are using our service for transportation to work/home. There aren't as many casual riders during the weekdays compared to the weekends, which shows that casual riders are usually using our bikes for fun, entertainment, sightseeing, and transportation rather than for their job or work.

Member/Casual – Day of Week and Number of Rides (Chart 2)



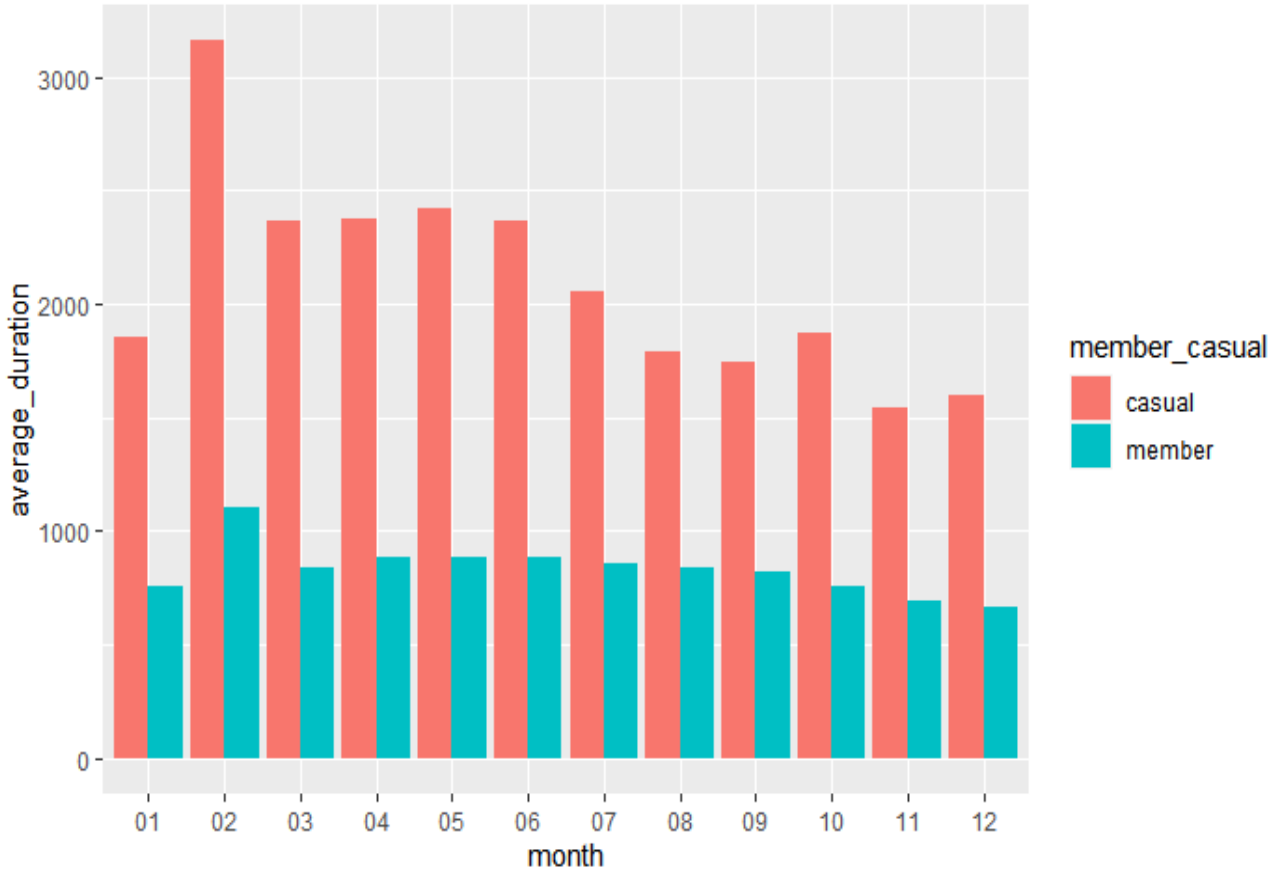
The graph (Chart 3) below shows the month and number of rides by the type of user. We can see here that member usage is usually consistent within each season; for example, the number of rides in June is similar throughout the entire summer. However, with casual riders, there is a large spike around May, June, and July (with June being the highest), but the other months have low usage.

Member/Casual - Month and Number of Rides (Chart 3)



The graph (Chart 4) below shows the average duration of casual and member rides by each month. This graph tells us that most rides by members have a consistent duration, and the season (weather) doesn't seem to affect the duration. This also justifies how members are utilizing our services as a means of transportation, while casuals seem to ride longer in general. The average duration in February for casuals seems a little over exaggerated, but we need to also remember that there weren't as many casual rides in February (with January and February having the fewest casual rides throughout 2021)

Member/Casual - Month and Average Duration (Chart 4)



How do your findings relate to your original question?

The findings help us identify the usage patterns and reasons for different users, which gives us profiles for them.

Who is your audience? What is the best way to communicate with them?

The main audience is the Cyclistic marketing analytics team, the marketing team because the profiles we have built will help them with their marketing strategies. The best way to communicate these findings and profiles is to construct a power point presentation.

Can data visualization help you share your findings?

Data visualization will help communicate these findings and also help the audience understand and grasp the differences and profiles of users.

Is your presentation accessible to your audience?

Yes, the necessary labels, contrasting colors will help identify and understand the visualizations.

Act

What is your final conclusion based on your analysis?

Members and casual riders are using our service for different purposes and reasons, and the analysis explains those reasons. Because of these different reasons, marketing strategies need to be refocused based on the users' different preferences.

How could your team and business apply your insights?

These insights can be taken into account when planning and preparing the next marketing campaign. There can be different promotions for the members as well as other services for weekend riders. Because of the profiles we have built, the team can focus on the different users and their needs.

What next steps would you or your stakeholders take based on your findings?

There could always be improvement through further analysis such as finding other meaningful, consistent data or applying different analysis strategies. The marketing team can use this information to help build a new marketing campaign as well.

Is there additional data you could use to expand on your findings?

Further analysis could help expand our findings. We could add more detailed climate data for usage patterns in different weather. More accurate mobility data, speed and location data, etc.

Top three recommendations based on analysis:

1. Construct a marketing campaign that highlights the benefits of using bikes as transportation for work by helping maintain the planet green and not having to deal with long traffic jams. Explain the benefits of riding by the amount of daily exercise we could practice. Promotions for members.
2. Build ads on how riding bikes can be great for recreation during the weekends and have a free trial for membership, so people can see how they can save more by becoming a member.
3. Give out coupons for the colder months and have a campaign, where members can earn more points during the winter, where they can use the points for a raffle, discount, etc.