# About this project

\*\*Cyclistic\*\* is a successful bike-share company. Until now, there marketing strategy focused on broad awareness. The company offers flexibility in its pricing. Customers who purchase one-day passes or full-day passes are considered casual riders. Customers who purchase annual memberships are considered members.

Management believes the future of the company will depend on maximizing annual memberships and have set a clear goal: Design marketing strategies aimed at converting casual riders into annual memberships.

To achieve this goal, I will analyze last month’s usage data to see how casual riders and members use our services differently. With this information, the marketing team will have a better understanding how our customers use our services and devise there marketing efforts accordingly.

# Data Cleaning Process

Before I began my analysis, I want to ensure the data I have is relevant for my tasks. As I am only looking at data for the month of April 2020, I confirm I have the correct datasets.

All usage data for the month has been collected and is confirmed. I have additional data for the latitude and longitude of the start and end stations. As riders do not have to ride directly to an end station, this data cannot be used to measure mileage of the rental sessions and thus, is irrelevant and won't be used in this analysis.

Next, I want to confirm the data's integrity by checking for duplicates, formatting, mistakes, any incorrect values, missing values, and any outliers that would skew the results of my analysis.

I make a copy of the original worksheet before modifying.

Each rental session has a unique ride id. I need to ensure there are no duplicates in this column. There were none.

Next, I want to ensure the formatting is consistent throughout. All data is formatted correctly, but I’ll check for extra spaces in the data with the trim function as well as check for proper casing

Next, I’ll check for incorrect or missing data.

There were 28 entries were rentals started in April but did not end until May. As casual users can only rent a full day pass or single ride pass, I believe these entries are made in error. If this were an actual project for a company, I would check with management about protocol for this situation. For the scope of this project, I will remove them from this analysis. I will show this information in a sheet color coded light blue

There are 125 records where the drop off station data was missing. As I still have the start and end times of the rental, I believe this data is still valid and useful for our goal. I will keep them for this analysis; however, I would check with analytics team as to why this data was not accurately recorded. This information will be displayed in my workbook color coded blue

Now, I want to ensure there are no outliers in the data that will skew my analysis.

Now I will aggregate the data to gather more insights. All my calculations in the worksheet and columns will be color coded green.

I determine the ride duration by finding the difference in end and start times.

Upon doing this, I have now discovered that there are 51 rental times that were not recorded correctly. Some have the exact same start and end time, but the start station and end station are different. This tells me the time of the rental was not recorded correctly. I will remove these from the analysis. I will let the analytics team know there may be an issue with our rental bikes or data collection efforts.

I create a listing matching the name of the day of the week (Monday-Sunday) to the number date. dd/mm/yyyy

I want to gather the max value of the duration of the rentals.

I now see we have 5 outliers that have kept the rentals for over 390 hours. Some of these entries are from casual riders who can only get a single ride pass or full day pass, so these records are made in error. There are outlier entries for members as well, but these seem to be a rare occurrence. These results would skew the results of the analysis. I will remove these from the analysis as well

After looking for the min of rental times, I see we have over 8000 records with rides less than 5 minutes. These will skew the results of the analysis as well. They will be removed. I recommend these records be analyzed separately.

Now that I have cleaned the data and confirmed its integrity, I can begin to analyze the data to see how casual riders and members use our services differently

# Analyze

After charting the total amount of rentals for the month, we are averaging 2,533 a day. I find that casual riders rented bikes 22,756 times in the month while members rented 53,237 times. Casual users make up 30% of rentals. Member’s rentals make up 70%. Members are using our services more frequently than casual users.

Next, I wanted to see how long our customers were using our rental bikes. Casual riders spent over 20,390 hours with our rentals. Members spent just over 18,960 hours with our rentals. **This tells us that while members rent more frequently their rental times are very low, Casual users initiate rentals less often, but are spending longer with the rentals**.

Next, I wanted to see if the day of the week made a difference in how our customers use our services. I can see that **casual users and members both use our services consistently throughout the week.**

**Saturday is the most popular day for both type of customers.**

I have also identified the top ten most used start stations for both casual riders and members. St. Clair St & Erie St is the most used for members while Clark St and Elm St are the most used for casual riders.

I have also identified the least used stations as well.

## Conclusion

Based on the available data of last month’s usage data I can conclude:

members are initiating rentals more often than casual users.

Casual riders are renting less often than members

members Ride duration for this month was on avg 0:35:51 per rental

casual riders avg 1:13:04 per rental

Casual riders are spending over twice as long riding when they do initiate a ride.

Both casual riders and members use our services consistently throughout the week.

Saturday is our busiest day for both members and casual riders

# Recommendation

I have identified how casual riders and members use your services differently. This information is vital for a marketing team to make effective advertising campaigns targeting casual riders for upgrades to annual memberships.

# Marketing team

While I would need to know the specific capabilities and culture of a particular company to understand the scope and limits of what our marketing team can do, I have compiled four strategies on how we can convert casual riders to members.

Most members are biking throughout the week for short distances, they prefer the flexibility of being able to get a rental any time and any number of times thought the week.

1. Because Casual riders purchase single ride or full day passes, they tend to keep the rentals longer. We should focus our marketing efforts on explaining the benefits of being able to get a rental at any time

2. Our busiest times of the month occurred on Saturdays. As this is the weekend, I propose another option, annual weekend passes. This will encourage casual riders to use our services more frequently on weekends.

3. Physical location of the customer may play a role as members and casual riders for the most part prefer different start stations. I have identified the top ten start stations for casual users. I believe we should focus our advertising at these locations.

4. We can use digital media to create ads that explain the benefits of using our services as a method of getting to and from work. These ads would explain the advantages members have of being able to rent bikes multiple times a day.

# Note on marketing Recommendations

With the marketing campaign recommendations, there are some limitations with the data. The dataset only contains usage data for one month. If more data is collected throughout the entire year, we would have a better understanding of customer preferences.

To more accurately conclude why members, ride for shorter times and why casual riders ride for much longer, it would be helpful to have more data. I suggest a survey to find out how our customers prefer to use our rentals. Transportation, Exercise, or leisure? The customer responses will increase understanding.

There are many factors that could affect how are bikes are rented. This may include traffic data for the month of April. This could also include weather data for the month of April. Access to those datasets would allow me to check for correlations and see if traffic or weather is playing a role in how customers use our products.

The dataset included a unique Ride id; however, there was no identifier for an individual user. The ride id tells when a bike is rented, but it doesn’t tell us if a specific customer is initiating the rental. With care for customer data and privacy, I suggest we collect encrypted user ids. This would allow us to see if the same members are renting more than once a day. As well as if casual users are renting multiple times a day further increasing our understanding of customer usage.

Thank you,

Jason