Ask:

1. The problem I am trying to solve is: how can the company use digital marketing to get customers to convert from casual riders to annual membership holders.
2. My insights can drive business decisions because the reasons customers pick to be casual riders as opposed to annual is in the data, I just need to put it together.

Prepare:

1. The data is located in the Cyclistics database and is organized uniformly by rows and columns.
2. The data is ROCCC ready because it comes from the company itself and there is enough of a sample size to make sure there isn't sample bias. It takes into account both types of riders.
3. The data comes from a secure location, and I have been given permission to use it. I linked the license in the presentation to make this clear.
4. There don't seem to be any problems with the data right now, I won't know until I start filtering and sorting to make sure it's clean.

Process:

1. The tools I will be using to manipulate my data will be google sheets at first, to get a view for the data I am working with, then R if the dataset is too large and requires a more widescope approach.
2. I am going to use sheets at first to fill in any missing columns of data I think should be included, in this case I will add a ride_length column to calculate in HH:MM:SS how long each ride was (stared_at (-) ended_at). Then I will add a day_of_the_week column to tabulate on what day each ride occurred using the weekday() function. 1 represents Sunday and 7 represents Saturday.
3. I then make sure to remove any duplicates and trim white space. It is important to double check the data I am removing or trimming to make sure that it does not impact my sample size, and also if it was crucial data to the analysis I would find a way to retrieve it. There were some 0 and negative values in my ride length column, so I made sure to remove those rows because they would make no sense and cause an error in the analysis.
4. The packages I am going to be using to clean and organize my data in R are
   install.packages("ggplot2")
   install.packages("tidyverse")
   install.packages("readr")
   install.packages("rlang")
   library(ggplot2)

```
library(tidyverse)
library(readr)
library("rlang")
install.packages("here")
install.packages("skimr")
install.packages("janitor")
install.packages("dplyr")
library(here)
library(skimr)
library(janitor)
library(dplyr)
```

5. - head(), colnames(), glimpse(),clean_names,rename()
6. I have used the above functions and have sifted through the data to make sure it is clean enough so that when I go to arrange and filter it I can do so with ease.
7. The documentation of any manipulation of data will be in the code (R markdown document) and explained in the presentation.
8. Now that I have used spreadsheets to manipulate my data the way I need, and cleaned it in both spreadsheets and double checked it in R, it's time to analyze.

Analyze:

1. Since I have already cleaned and formatted my data, I am going to start analyzing it by deriving some values I think would be useful to have when evaluating the differences between casual riders and annual members.
2. I create a pivot table using spreadsheets because it is one of the best tools for quick analysis on a large dataset, and I use it, and some other formulas to find key values.
3. I calculated the:
   - **mean ride length** for all the rides to find out on average how long members and casual riders take on their trips.
   - **max ride length**, to see the longest trip that was taken.
   - **mode of the day of the week**, to see which day was the most popular for people to take rides on.
   - **mean ride length for casual riders and annual members separately**, to see the differences in ride length between casual riders and annual members
   - **average ride length for users by the day of the week** to see on which days casual riders and annual members spend the longest
   - **number of riders per user by day of the week**, to figure out the days in which casual riders and annual members are the most active.

4. After getting a birds eye view for the data, I am going to move into R because I must analyze all 4 quarters of bike trip data and it would be much more time efficient to use R for that magnitude of analysis.
5. I will include an R markdown file of my analysis.

Share:

1. I was able to answer the question regarding differences between casual riders and annual members.
2. Trip durations for casual riders are almost double that of annual members on a day to day basis.
3. Casual riders do not go on rides as frequently, so they go on longer trips to make it worthwhile
4. Annual members go on a more consistent basis and frequently go on bike trips, and have shorter ride times.
5. Casual Riders' most active days are Sat and Sun
6. Casual Riders' least active days are the weekdays Mon-Fri
7. Annual Members' most active days are the weekdays Mon-Fri
8. Annual Members' least active days are Sat and Sun

Act:

1. The final conclusion based on my analysis is that yes there are some key differences between the 2 groups that can be used to impact their likelihood of converting from casual riders to annual members.
2. If we were to run a marketing campaign it would need to emphasize benefits for coming during the week, as that is when casual members are choosing to come the most.
3. Since casual members tend to enjoy longer trips, maybe adding incentives to the annual memberships based on trip length might be helpful.
4. Creating a survey and sending it out to casual riders might also give us more insight as to why they choose to not go for the annual membership and what would be some lucrative that could entice them.