Business tasks:

1. Identify trends in usage minutes/duration for member and casual riders, to help convert casual riders to annual members.

Data sources used:

* The data source is publicly available.
* The data source is from year 2018 which divided by quarter

Cleaning log & manipulation:

*Note: below are some bulk cleaning and modification description. But, personally I’d very like to log each bit of changes in data into a matrix template which will enable me to apply or reverse those changes at some point where needed.*

* Below columns/variables added:

1. Quarter; to give us the ability to see the data by quarter
2. Ride\_length; is the duration of ride between two stations
3. Day\_of\_week; is the week day for each ride, this will be used for analysis

* In addition, I’ve tweaked the variable names just to be aligned with each other, and this will help me a lot when binding/merging all quarters together. And I did the alignment when reading the datasets.
* Also, there is a comma in ride duration column, so I removed that and changed the variable type from character to number, just to make the data analysis easier and accurate.
* Outliers removed: there was quite some outliers in all quarters, I’m removing them to avoid any further inconsistency in the calculation.

Exploratory Data Analysis (EDA):

* Just to learn a bit more about data, I’m testing my data for NAs if any, so I should be aware of it.
* Taking the descriptive analysis, which helps to know more about duration of rides
* Eventually; tidying data for running the final analysis and visualization
* Finally; doing some data visualization using ggplot, just to see if there is any trend to help us to answer the main question.

Summary of analysis:

*Based on the analysis result, it looks there is no relationship between customers and subscribers in general. But, overall subscribers mostly take the rides for shorter distances but many rides. Despite that, customers take the rides for longer distances but few rides. This is consistent for overall analysis and also disaggregated by quarter.*

Conclusion: Based on the analysis result, customers are taking long rides but fewer times, it will be effective if we offer a membership focusing on distance or length of ride rather than number of rides per day/week or even year. Also, it seems already subscribers like to ride shorter distance but more times, so this means, we should keep that membership as it is, but create another type of membership to target customers.