**Divvy\_Exercise\_Full\_Year\_Analysis:**

The purpose of this script is to consolidate downloaded Divvy data into a single dataframe and then conduct simple analysis to help answer the key question: “In what ways do members and casual riders use Divvy bikes differently?”

**Statement of the business task**

* What is the problem you are trying to solve?
* How do annual members and casual riders use Divvy Cyclistic bikes differently?
* Why would casual riders buy Divvy Cyclistic annual memberships?
* How can Divvy Cyclistic use digital media to influence casual riders to become members?
* How can your insights drive business decisions?

**Data Source Used:**

Downloaded the raw data for previous 12 months of cyclitic trip data from

[https://divvy-tripdata.s3.amazonaws.com/index.html](%20https:/divvy-tripdata.s3.amazonaws.com/index.html)

The files are zipped and the format is csv files. Downloaded the files and unzipped it to excel files with appropriate file-naming conventions. Organized as 12 months data files, the data needs to cleaned and filtered.

**Clean the data:**

* Checked for duplicate data and found no duplicates in the file 220201-divvy-tripdata
* Checked for blank data – there are few blanks’ data in station starting and ending point. But ignored it for now as we need only starting and ending date and time.
* Changed the started\_date and ended\_Date format as time – US – date and time format

**Process data in R**

* As the file size is large, decided to complete the project in R
* Read the files in R Studio and cleaned the extra columns and ensured column names and data types are consistent in each file.
* Removed columns that are not in use.
* Checked for column value consistency
* Created Date, year, month, day of the week column for calculations
* Created calculated field column for length of the ride and made it a numeric field for calcualtions and remove bad data having negative values in ride\_length column
* Created Version 2 for further analysis

**Descriptive Analysis**

* Conducted descriptive analysis on ride length, calculating mean, median, max and min values.
* Explored summary statistics
* Aggregated the mean, median and max values for member and casual members
* Calculated average ride time for each member vs casual riders also calculated average ride time as per the days of the week.
* Created visuals of the number of rides by rider type and for average duration per member type.

Summary:

As per the analysis the number of rides is more for members than casual riders especially on the middle of the week. Whereas the average duration of ride is more for casual riders than members. Thus the usage of casual and member riders are analyzed and visuals are created to support the analysis.