**Spreadsheet Process - Dirty to Clean**

To work with the extensive dataset from "How does a bike-share navigate speedy success?" efficiently, I opted to focus on a subset of the data. I downloaded files spanning December 2023 to February 2024 from [Case Study 1\_How does a bike-shared navigate speedy success.](https://www.coursera.org/api/rest/v1/asset/download/pdf/1XKhm37HS9iPXHfAIEBaRQ?pageStart=&pageEnd=) And analyzes to derive insights specific to each month.

1. **Managing Data for Analysis: A Step-by-Step Guide**
2. **Download** [**case study**](https://www.coursera.org/api/rest/v1/asset/download/pdf/1XKhm37HS9iPXHfAIEBaRQ?pageStart=&pageEnd=)

Get the files from December 2023 to February 2024.

1. **Organize Files:**

Unzip them and put them in a folder named with "CSV" for clarity.

1. **Folder Location:**

Create a folder on your drive and name it appropriately.

Put the CSV folder inside it.

This setup helps manage the data for further analysis using SQL, R, or Tableau. We'll focus on analyzing the three-month data separately and comparing the results later

**2. Selecting and Cleaning data:***Checking if the sheet has any missing data by manually identifying and removing rows.*

*This process is per situation, and normally stakeholders are involved with the decision on what to do with empty cells.*

***1. Select All Fields***

1. Open Your Google Sheet:

* Navigate to the Google Sheet you are working on.

2. Select All Fields:

* Click the top-left corner of the sheet (the small square above row 1 and to the left of column A). This action will select all cells in the sheet.

***2. Identify Blank Cells***

1. Open the "Go To" Window:

* Press F5 or hold down CTRL+G to open the "Go To" window.

1. Select Blank Cells:

* In the "Go To" window, click on Special.
* Choose Blanks and then click OK.
* This may take a few minutes depending on the size of your sheet and the number of blank cells.

***3. Remove Rows with Blank Cells***

1. Identify Highlighted Cells:

* After selecting blanks, scroll through the sheet to find any highlighted cells indicating blank entries.

1. Delete Rows with Blank Cells:

* Right-click on any highlighted cell.
* Select Delete from the context menu.
* Choose Entire row to delete the entire row containing the blank cell.
* Confirm any warnings that appear.

1. Repeat if Necessary:

* If there are multiple sections with blank cells, repeat the process until all rows with blank cells are removed.
* It might take a few iterations, and the sheet might temporarily freeze during the process.

**3. Step-by-Step Guide to Add and Format "ride\_length" in Google Sheets**

***1. Add "ride\_length" Column***

1.Open Your Google Sheet:

* Navigate to the Google Sheet you are working on.

2. Insert the "ride\_length" Column:

* + Go to Column N.
  + Click on cell N1 and enter "ride\_length" to name the column.

***2. Calculate Ride Length***

1. Enter the Formula to Calculate Ride Length:

* Click on cell N2.
* Enter the formula =D2-C2 to calculate the ride length by subtracting the start time (Column C) from the end time (Column D).

***3. Format the Ride Length as HH:MM***

1. Select the Cell with the Formula:

* Click on cell N2.

1. Open Format Cells Window:

* Right-click on cell N2.
* Select Format cells.

1. Change to Time Format:

* In the "Format cells" window, go to the Number tab.
* Select Time from the "Category:" list.
* Choose the time format 37:30:55 from the "Type:" list.
* Click OK.

1. ***Autofill the Column with the Formula***

1. Copy the Formula Down the Column:

* Select cell N2.
* Press CTRL+C to copy the formula.
* Select cell N3, then hold CTRL+SHIFT+down-arrow key to select the entire column down to the last row with data.
* Press CTRL+V to paste the formula into all selected cells.

**4. Steps to Add a "day\_of\_week" Column in Google Sheets**

***1. Create the "day\_of\_week" Column***

1. Open Your Google Sheet:

* Navigate to the Google Sheet you are working on.

2. Create a New Column:

* Scroll to Column O (or the next available empty column).
* In cell O1, type day\_of\_week to name the new column.

***2. Enter the Formula***

1. Enter the Formula in O2:

* In cell O2, enter the formula =WEEKDAY(C2, 1).
* This formula will return a number representing the day of the week (1 = Sunday, 7 = Saturday).

**5. Creating Pivot table**

***1. Create a Pivot Table***

* Open Your Google Sheet:Insert a Pivot Table:
* Click anywhere on your data sheet.
* Go to the top menu and click on Data > Pivot table.

1. ***Set Up the Pivot Table***

1. Add Rows and Columns:

* In the Pivot table editor, under Rows, add ‘member\_casual’ and ‘day\_of\_week’
* In the Values area, add ‘ride\_length’.

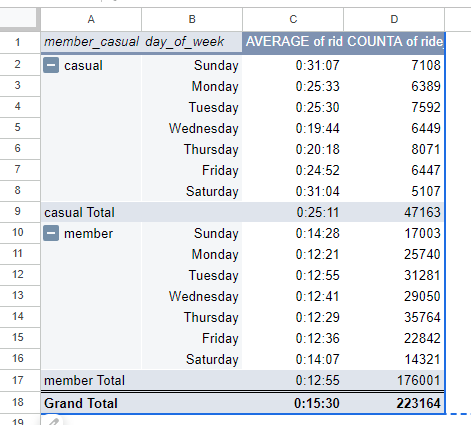
1. ***Add Values:***

* In the Values area, add ride\_length:Click on Summarize by and select AVERAGE to calculate the average ride length.
* Add ride\_id to the Values area:Click on Summarize by and select COUNTA to count the number of rides.

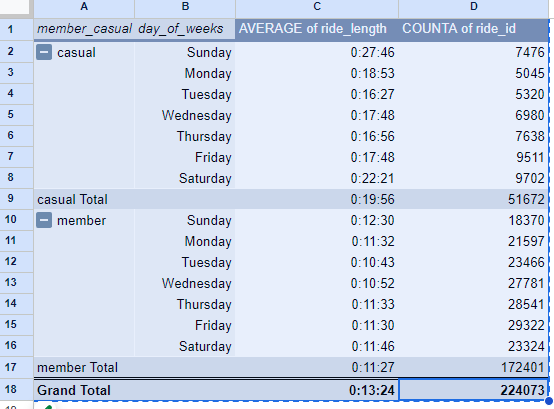
**6. Summary:**

Due to the amount of content, the summary is too large to post on GitHub; a .PNG file has to be done.

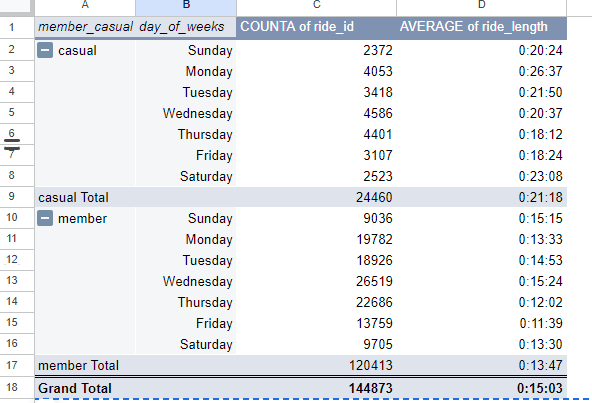
Bike share data pivot table for February 2024



Bike share data pivot table for December 2023



Bike share data pivot table for January 2024



**7. Fin:**

As per stakeholder request:

* Downloaded all CSV files from December 2023 to February 2024.
* Ensured all files were error-free and cleaned them up as needed.
* Added new columns to the data using formulas to combine relevant information.
* Created multiple pivot tables based on stakeholder requests to analyze different aspects of the data.
* Developed a summary report highlighting key findings and insights from the data analysis.
* Provided clean, organized, and easy-to-understand data for the stakeholder to make informed decisions.