Steps were taken per each .csv file before combining into an .xlsx file

1. Created a table of the data.
2. Change the time data to match the formatting in R Programming.
   1. R programming on Kaggle could not handle the volume of data and incorrectly associated time data structures. We’ll be using Tableau instead of R programming.
3. Checked the ride\_id columns for duplicate entries as they are unique to each individual ride; each row should be unique.
   1. No duplicates returned.
4. Explored station names in Pivot Table to make sure the IDs were, in fact, unique to the station names.

I ran a spell check on the start\_station\_name and end\_station\_name columns. Spell-checked all station names flagged against streets in Chicago using Google Maps and Divvy map online. *Corrects for Grammar/Spelling were run last once the spreadsheets were in one notebook.*

<https://data.cityofchicago.org/Transportation/Divvy-Bicycle-Stations-Map/bk89-9dk7>

[www.googlemaps.com](http://www.googlemaps.com)

|  |  |  |
| --- | --- | --- |
| The Error | The Correct | Note |
| Mies van der Rohe Way & Chestnut St | Mies Van der Rohe Way & Chestnut St |  |
| Dusable Harbor | DuSable Harbor |  |
| DuSable Lake Shore Dr & Monroe St | DuSable Lake Shore Dr & Monroe St | Extra spacing error |
|  |  |  |
| Hastings LWS | Hastings Ln, Dr., St. | Unclear – not adjusted appeared on Chicago.gov |

Checked the member\_casual column to make sure it only produced member or casual.

1. The start\_lag and end\_lag column had an unsurmountable number of errors on individual cells that would not convert to numbers unless done by cell.
   1. Therefore, I copied the columns as start\_lag\_raw\_data and end\_lag\_raw\_data and created new start\_lag and end\_lag column using the =VALUE() formula.
   2. Later, I copied and pasted the columns as values only and removed start\_lag\_raw\_data and end\_lag\_raw\_data columns.
2. Time columns and number columns were formatted accordingly.

NOTE: Some started\_at times are later than ended\_at times; this is an error but occurs too frequently to just adjust without more information.

1. Each table was formatted uniformly and converted to range.
2. Sheets all combined into .xlsx workbook. Cyclistic\_range
3. Created two new columns: ride\_length =(ended\_at -started\_at) and weekday =WEEKDAY(C2,1)
4. ride\_length was not producing usable accurate numbers for negative time values. The formula changed to produce a blank for negative time values.
   1. ERROR: ################
   2. =IF(ended\_at<started\_at,””, ended\_at - started\_at)

NOTE: Previously produced the incorrect MAX, MIN, AVG of ride\_length

1. Formatting corrected.
2. Saved xlsx Workbook - Cyclistic\_range\_static\_with\_new\_columns