**Cleaning phase**

* I added two columns in each file of 12 files in the dataset by excel , the first is ride\_length which is the duration of the ride which on form hh:mm:ss and calculated by subtraction of (end at) column from (start at) column , and the other column is called dayweek which represents the day of week has the ride happened in on form 1 for Sunday , 7 for Saturday and so on.
* Handling the format of (start at) column and (end at) column of files start from month 6 (May) to the end of files (December) in excel .
* Confirmed that all the dates in (end at) column are all after the dates in (start at) column and corrected the cells who are not satisfied with this in excel.
* Checked for duplicates in excel.
* Used filters in excel on column member\_casual to confirm that the column just has two types of users (member, casual)
* The ride length which exceeds 24 hours and the user type is casual I have capped the length to just 24 hours due to the biggest ticket is for just one day.
* Deleted the records have the ride length equal to zero and at the same time whether the start station or end station are null.
* deleted the records have the start station and end station are null and the coordinates are missing.
* deleted the records have the start station or end station are null and the coordinates at the start and the end are equal to each other.