1. A python notebook “preprocess.ipynb” is created to handle some errors in the data set.
2. An irrelevant column named “Unnamed: 4” was dropped
3. No duplicate entries found
4. There are 9 NaN in column "Mar Cap - Crore" and 123 NaN in "Sales Qtr - Crore", which are 1.8% and 25% of the total data volume. Since do additional data found, these NaN rows where market capital was missing were only removed.
5. Values in the dataset were in Cr. They have been transformed by multiplying 107 and let Tableau Public handle the exponent better.
6. The new data is saved as “data\_v2.csv”
7. In Tableau Public, “Name” is renamed to “Company”, “Mar Cap - Crore” is renamed to “Market Capital” and “Sales Qtr - Crore” is renamed to “Quarterly Capital”
8. S. No column was hidden which is not required