**Observations**

**Record count before dropping duplicates:** 46,648,470  
**Record count after dropping duplicates:** 46,648,470 **Record count after dropping NA values:** 5,502,503

From the above, we can conclude that there are no duplicate records in the data. However, there is a significant number of NULL values in the dataset that need to be addressed.

**Questions:**

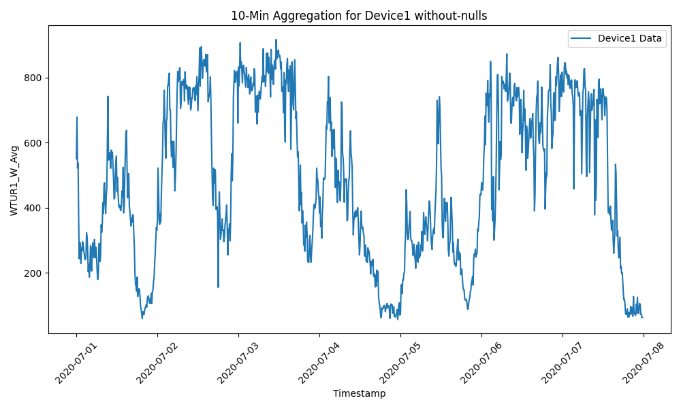
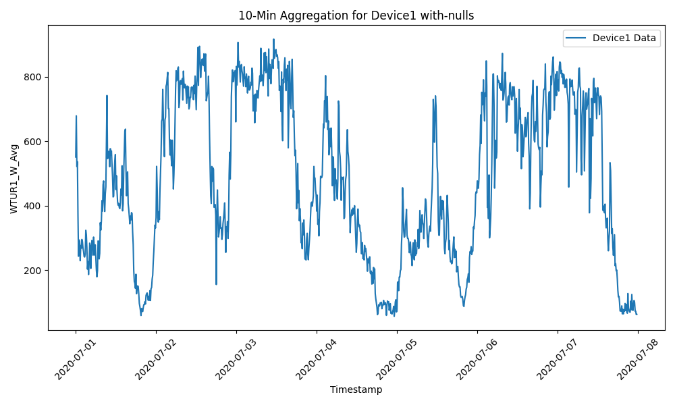
**What are some issues to be mindful of, and how do you measure data quality?**  
Maintaining data quality requires careful handling of NULL values, as they can lead to misleading results. In this dataset, since the data comes from multiple devices, it is better to remove NULL values. Assigning arbitrary values to NULLs may distort the results and lead to inaccuracies.

**Are there signals for which aggregation is not meaningful? Can you identify them in this dataset?**  
Aggregation is not meaningful for the “variable” column, as it doesn't represent numeric or aggregable data. Aggregation is relevant only for the “value” column, where different statistical operations can be applied to derive insights.

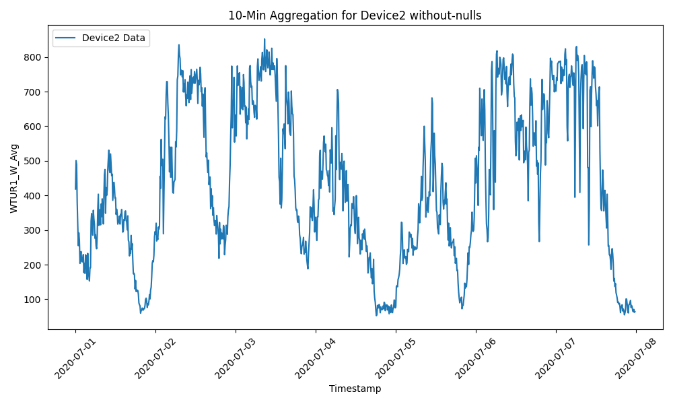
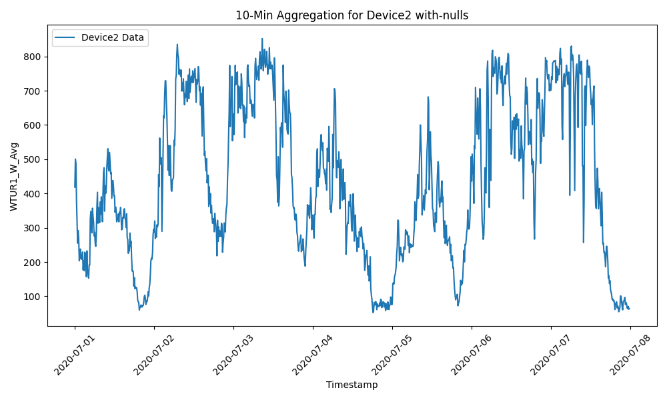
**How would you handle NaN values in the dataset?**  
NaN values can be removed to avoid inaccuracies, as imputing them with arbitrary values (such as averages or standard deviations) might misrepresent the actual data. While methods like replacing NaNs with averages or other metrics can be used, as a Data Engineer without a specialization in Data Analysis or Visualization, I would recommend leaving such decisions to domain experts or analysts.

**Visualisation Plots: (With and Without NULLs) – Zoom to View Clearly**

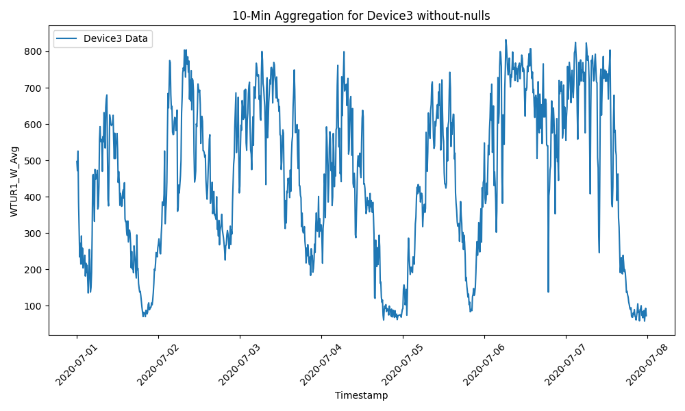
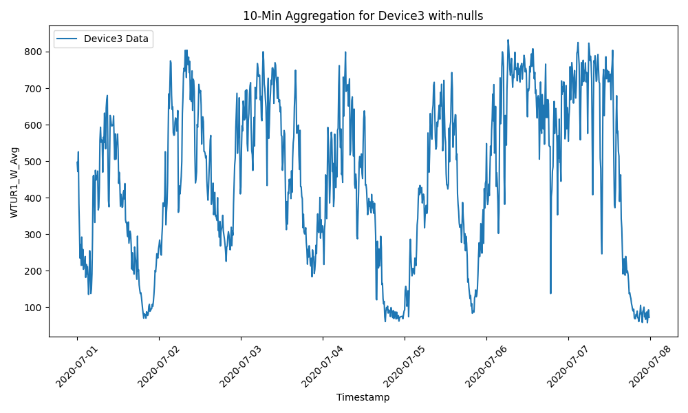
**Device1:**



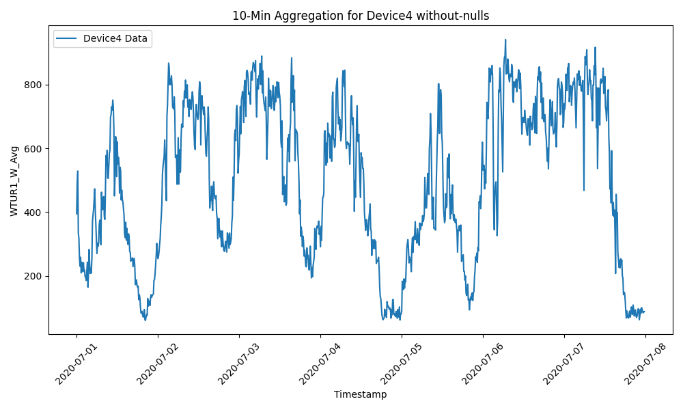
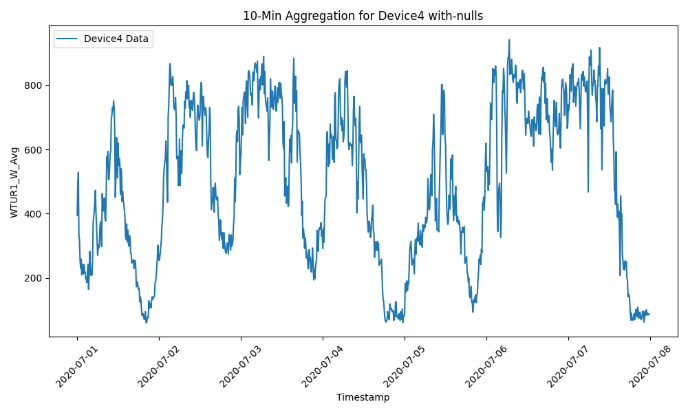
**Device2:**



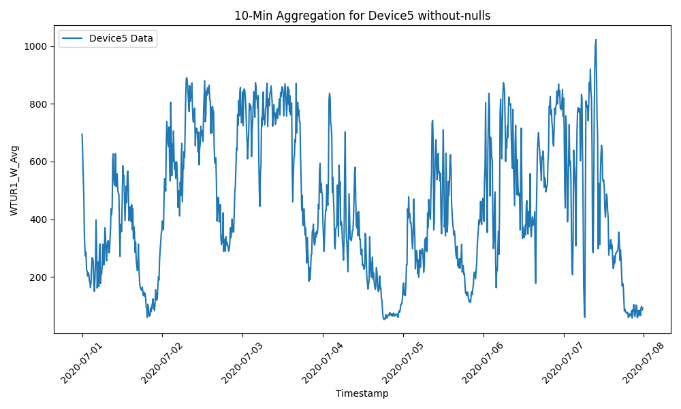
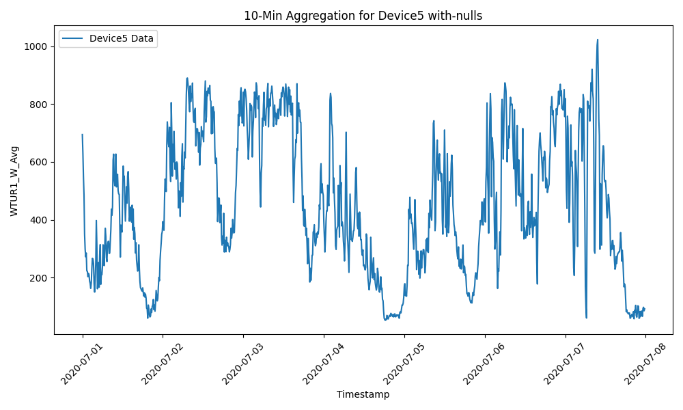
**Device3:**



**Device4:**



**Device5:**



**Device6:**

