Data Cleaning of Customer Personality Analysis Dataset

For this task, I worked with the **Customer Personality Analysis** dataset, which I downloaded from Kaggle. The dataset had **over 2,000 rows and 29 columns**, containing details about customers such as their age, income, spending habits, and buying behavior. My goal was to clean this raw dataset using Excel so that it becomes accurate, organized, and ready for analysis.

I started by fixing the column names. Many of them were long or inconsistent, so I changed them to simple lowercase names that were easier to work with. After organizing the headers, I checked how many missing values were present in each column by using the COUNTBLANK function. This helped me understand which columns needed the most attention.

The Income column had several missing values, so instead of leaving them empty, I filled them using the **median income** of each marital status group. This made the data more realistic because different marital groups can have different income levels. Next, I cleaned the text columns by removing extra spaces and correcting formatting using the TRIM and PROPER functions. I also fixed spelling errors in the education column. In the marital_status column, values like "yolo," "absurd," and "alone" did not make sense as separate categories, so I changed them all to "single" to keep things consistent.

I also looked for outliers in the Income column by sorting the values from highest to lowest. Some income values were extremely high and unrealistic, so I removed them to avoid incorrect results in future analysis.

After completing all these steps, the dataset became much cleaner and easier to understand. The final cleaned version is more consistent, has fewer errors, and is now ready for further analysis or visualization. This task helped me practice real data-cleaning techniques and understand how important it is to prepare data properly before doing any kind of analysis.