

# Sales Data Cleaning and Preprocessing – Task Report

In this task, the main focus was to clean and organize a sales dataset so it's easier to work with for future analysis. I started by loading the data and taking a look at how it was structured. There was an extra column called "Unnamed: 0" that didn't serve any purpose, so I removed it. I also checked for any missing values and cleaned those up to avoid issues later on. One important column, 'Order Date', had to be converted into a proper date format, and I got rid of any rows that had invalid or empty dates. To make sure the data was accurate, I also removed any duplicate records that were unnecessarily repeated.

After that, I fixed the data types for a few key columns like 'Quantity Ordered', 'Price Each', 'Sales', 'Month', and 'Hour' so that they were in the right numerical format. This helps make calculations and analysis smoother. Then, I added some new columns using the order date—like the specific day, the day of the week, and the year—so it would be easier to study sales patterns over time.

Once the data was cleaned and organized, I did some basic analysis to find out which products were sold the most and which ones made the most money. These kinds of insights can be super helpful for understanding what's working well in sales. In the end, I saved the clean and updated dataset into a new file called "Cleaned\_Sales\_Data.csv". Overall, this task was all about transforming messy raw data into a neat and useful format that's ready for deeper analysis or reporting.