Data Cleaning Process

On the first column EmpID, there are duplicate entries with some data appearing twice.

A screenshot of a computer

AI-generated content may be incorrect.I have used conditional formatting to identify any duplicates by highlighting them as red.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.Removed duplicates to ensure each row only has unique data.

There are missing values in the Gender column which I have identified using the filter.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.These blank rows are removed from Gender Column as Gender can be an important feature for data analysis.

A screenshot of a computer

AI-generated content may be incorrect.The Department Column has inconsistent naming of departments. For example, Research & Development and R&D, HR and Human Resources.

A screenshot of a computer

AI-generated content may be incorrect.A screenshot of a computer

AI-generated content may be incorrect.A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

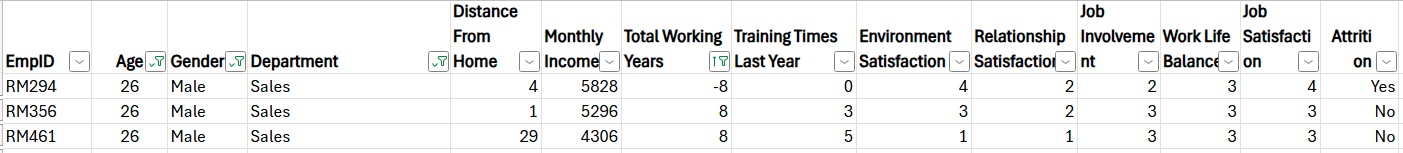
AI-generated content may be incorrect.

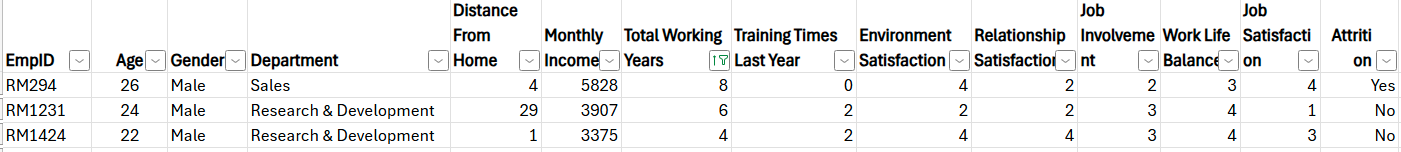
A screenshot of a computer

AI-generated content may be incorrect.Standardised the name of Research & Development and Human Resources for consistency.

A screenshot of a computer

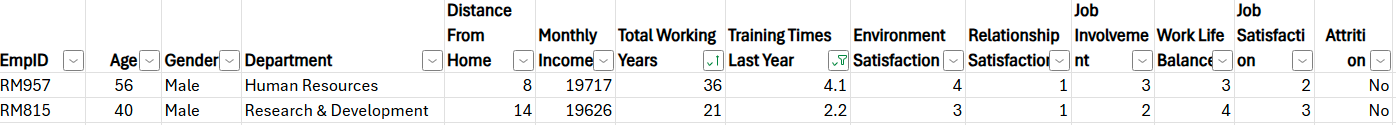
AI-generated content may be incorrect.A screenshot of a computer

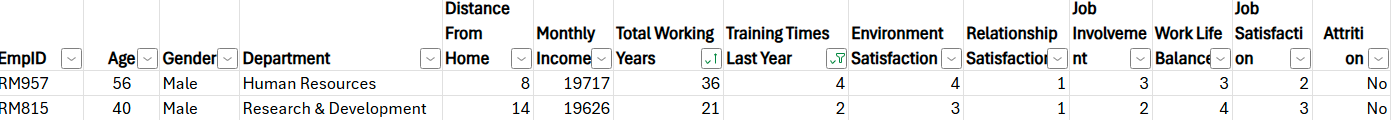
AI-generated content may be incorrect.There are negative values in total working years where the total working years should always be a positive value. I have made the value of Age, Gender, Department and Total Working Years to be the same as the negative value. Since the salary of the negative values is within the salary range of the positive values, I can deduce logically that the negative sign might have been a typo.

I have changed the negative values -8, -6, -4 to 8, 6, 4 respectively.

A screenshot of a computer

AI-generated content may be incorrect.Employees with 0 work years are preserved, as this likely reflects new employees.

Training Time Last Year has float data types where it should have been an integer.

I rounded off the float values to the nearest integer.