# 1. Duplicate Records

* **Duplicates in Employee\_ID:**
  + **Output:** 0 duplicate entries.
  + **Interpretation:** There are no duplicate employee records, ensuring that each employee has unique data in the dataset. No additional cleaning is required for Employee\_ID.

### **2. Categorical Variables**

* **Employment\_Type:**
  + **Distinct Types:** 2 categories: In-Office and Remote.
  + **Occurrences:**
    - **In-Office:** 510 employees.
    - **Remote:** 490 employees.
  + **Interpretation:** The dataset has an almost even distribution of employees working in-office and remotely. This balance could allow for meaningful comparisons between employment types when analysing their impact on productivity and well-being.

### **3. Missing Data**

* **Check for Missing Values:**
  + **Output:**
    - **Employee\_ID: 0 missing values.**
    - **Employment\_Type: 0 missing values.**
    - **Hours\_Worked\_Per\_Week: 0 missing values.**
    - **Interpretation: There are no missing values in the dataset, which indicates data completeness and reliability. No further imputation or cleaning is necessary.**

### **4. Data Types**

* **Data Types in Each Column:**
  + **Employee\_ID: int64 (Integer)**
  + **Employment\_Type: object (Categorical)**
  + **Hours\_Worked\_Per\_Week: int64 (Integer)**
  + **Productivity\_Score: int64 (Integer)**
  + **Well\_Being\_Score: int64 (Integer)**
  + **Interpretation:**
    - **The data types are generally appropriate.**
    - **Employment\_Type is stored as an object, which should ideally be converted to a category type, though it will not significantly impact the analysis.**
    - **The other numerical variables (Hours\_Worked\_Per\_Week, Productivity\_Score, Well\_Being\_Score) are correctly stored as integers.**

### **5. Statistical Summary of Numerical Variables**

**The summary statistics provide key insights into the central tendency, spread, and range of the data.**

* **Hours\_Worked\_Per\_Week:**
  + **Mean: 39.72 hours.**
  + **Std. Deviation: 8.04.**
  + **Min - Max: 16 - 64 hours.**
  + **Interquartile Range (IQR): 34.75 (25th percentile) to 45.00 (75th percentile).**
  + **Interpretation: The average employee works approximately 40 hours per week, with a minimum of 16 and a maximum of 64 hours. The standard deviation of 8.04 hours suggests moderate variability in working hours.**
* **Productivity\_Score:**
  + **Mean: 68.60.**
  + **Std. Deviation: 12.24.**
  + **Min - Max: 33 - 112.**
  + **Interpretation: The average productivity score is around 68.6, with scores ranging from 33 to 112. A relatively high standard deviation of 12.24 indicates some variability in productivity among employees.**
* **Well\_Being\_Score:**
  + **Mean: 63.98.**
  + **Std. Deviation: 13.87.**
  + **Min - Max: 14 - 104.**
  + **Interpretation: The average well-being score is approximately 64, with a range between 14 and 104. The well-being scores have a higher standard deviation, indicating greater variability in employee well-being.**

### **6. Correlation Analysis**

**The correlation matrix measures the strength and direction of the relationships between numerical variables.**

* **Correlation Matrix:**
  + **Hours\_Worked\_Per\_Week vs. Productivity\_Score: -0.25 (negative correlation).**
  + **Hours\_Worked\_Per\_Week vs. Well\_Being\_Score: -0.25 (negative correlation).**
  + **Productivity\_Score vs. Well\_Being\_Score: 0.14 (weak positive correlation).**
* **Interpretation:**
  + **Negative Correlation (Hours Worked vs. Well-Being): Employees who work more hours per week tend to have lower well-being scores. The correlation of -0.25 suggests that longer working hours may be associated with reduced well-being, though the relationship is moderate.**
  + **Negative Correlation (Hours Worked vs. Productivity): Employees working longer hours tend to have slightly lower productivity, with a correlation of -0.25. This may indicate diminishing returns on productivity with increased working hours.**
  + **Weak Positive Correlation (Productivity vs. Well-Being): There is a weak positive correlation (0.14) between productivity and well-being, meaning that employees with higher well-being scores tend to have slightly higher productivity, though the relationship is not strong.**

### **7. Key Insights**

* **Balance of Employment Types: With a near-even split between in-office and remote employees, comparisons between the two groups are statistically viable.**
* **Impact of Hours Worked on Well-Being: There is a moderate negative correlation between hours worked and well-being, suggesting that working longer hours may negatively impact employee well-being.**
* **Moderate Negative Impact on Productivity: Similarly, longer working hours are associated with a slight decline in productivity, indicating that employees might not be more productive by simply working more hours.**
* **Productivity and Well-Being: While the correlation is weak, higher well-being tends to be associated with higher productivity. Enhancing employee well-being could potentially have a positive impact on productivity.**

### **8. Conclusion**

**This analysis reveals important insights into the relationship between working hours, productivity, and well-being. While increased working hours have a moderate negative impact on both well-being and productivity, enhancing well-being might lead to slightly better productivity outcomes. Further analysis could involve grouping data by employment type (In-Office vs. Remote) to explore whether the impact of hours worked differs across these groups.**