

# Data Cleaning (Pandas)

This dataset helps you to increase the data-cleaning process using the pure Python pandas library.

## Columns :

- Age
- Salary
- Rating
- Location
- Established
- Easy Apply

Here are some questions that you can follow to perform data cleaning and manipulation using the given dataset.

### 1. Missing Values:

Question: Are there any missing values in the dataset, and if so, how should they be handled for each indicator?

### 2. Data Types:

Question: What are the data types of each indicator, and do they align with their expected types (e.g., numerical, categorical)?

### 3. Outliers:

Question: Identify potential outliers in numerical indicators (e.g., Age, Salary, Rating). Should outliers be removed or adjusted?

### 4. Salary Formatting:

Question: Examine the format of the Salary column. Does it require any formatting or standardization for consistent analysis?

### 5. Location Standardization:

Question: Check the consistency of location entries. Do they need standardization, and how can this be achieved?

### 6. Established Column:

Question: Explore the Established column. Are there any inconsistencies or anomalies that need to be addressed?

**7. Easy Apply Indicator:**

Question: Analyze the Easy Apply column. Does it contain boolean values or need transformation for better analysis?

**8. Rating Range:**

Question: Investigate the range of values in the Rating column. Does it fall within expected rating scales, and how should outliers be treated?

**9. Age Distribution:**

Question: Check the distribution of values in the Age column. Are there any unusual entries, and how might they impact analysis?

**10. Handling Special Characters:**

Question: Examine all text-based columns (e.g., Location). Are there special characters or inconsistencies that need cleaning?

**11. Data Integrity:**

Question: Ensure data integrity by cross-referencing entries. For instance, does the Established column align with the Age column?

**12. Easy Apply Transformation:**

Question: If the Easy Apply column contains non-boolean values, how can it be transformed into a usable format?

**13. Location Accuracy:**

Question: Assess the accuracy of location entries. Are there misspelled or ambiguous locations that require correction?

**14. Handling Categorical Data:**

Question: For categorical indicators, consider encoding or transforming them into a format suitable for analysis.

**15. Consistent Rating Scale:**

**Question:** Ensure a consistent rating scale in the Rating column. Should it be normalized or adjusted for uniform analysis?

These questions cover various aspects of data cleaning and manipulation, guiding you through the process to enhance the quality and reliability of the dataset. Feel free to explore and find hidden patterns in the data.

