Assignment for Data QA &QC Internship

# Summary:

Steps followed during data cleaning process:

1. **Load the Data:-**

* Load the dataset using its path

2. **Inspect the Data:-**

* Examined the data to understand its structure and identify the types of errors and inconsistencies present.
* It consists of following columns

1.Name

2.Age

3.Email

4.Join Date

5.Salary

6.Department

3. **Handle Missing Values:-**

* Identifying and handling missing values in the dataset.
* 13% of the total data is missing
* Drop missing values and columns

4. **Remove Duplicates:-**

* Identifying and removing duplicate rows to ensure each record is unique.

5. **Correct Email Formats:-**

* Identifying and correcting invalid email formats.
* Defined a regex pattern for valid emails and found out valid emails.
* Filter out valid emails only

6. **Clean Name Fields:-**

* Remove any noise added to names
* Checked all names are in string datatype.
* Converted to title case.
* Removed special characters.

7. **Standardise Date Formats:-**

* Ensuring all dates in the 'Join Date' column follow a consistent format YYYY-MM-DD).

8. **Correct Department Names:-**

* Identify and correct typos in the 'Department' column.

9. **Handle Salary Noise**:-

* Identify and handle any noise in the 'Salary' column.
* Ensuring salary values are within a reasonable range

After Cleaning process saved the data to a csv file for further analysis.