

# DATA HUT QA ASSIGNMENT

## STEPS PERFORMED FOR CLEANING THE DATA:

### 1. LOAD THE DATA

The dataset is downloaded as csv format and it is read using `read_csv()` as a dataframe.

### 2. INSPECT THE DATA

- a. Checking for null values
- b. Checking the duplicate values
- c. Checking the shape of the dataset
- d. Checking the columns

### 3. DATA PREPROCESSING

- a. Dropped a column "Unnamed" which was irrelevant
- b. Dropped the duplicate rows
- c. Dropped rows with more null values in "Name" ,"Join date
- d. Cleaned "Department Field Errors"
- e. Replaced null values in Age and Salary Column using Mean
- f. Replaced null values in Department column using Mode
- g. Split the name field into First name, Middle Name and Last Name for easy inspection
- h. Dropped the rows with middle Name more than a specific length which indicated presence of unwanted characters
- i. Dropped First name Middle name and Last Name columns
- j. Corrected Errors in Email and Date Format

k. Handle Noise in “Salary

#### **4. DOWNLOAD THE DATA**