**Data Cleaning with Pandas**

# **Dataset Overview**

The data was scrapped from glassdoor's website. There are two versions of the data one is uncleaned and another one is cleaned. Web scrapping, Data cleaning & EDA code are added in the code section.

In the cleaned version the columns explanation are as follows,

* **Job** **Title**: Title of the job posting
* **Salary** **Estimation**: Salary range for that particular job
* **Job** **Description**: This contains the full description of that job
* **Rating**: Rating of that post
* **Company**: Name of company
* **Location**: Location of the company
* **Headquarter**: Location of the headquater
* **Size**: Total employee in that company
* **Type of ownership**: Describes the company type i.e non-profit/public/private farm etc
* **Industry, Sector**: Field applicant will work in
* **Revenue**: Total revenue of the company
* **min\_salary,max\_salary,avg**\_**salary**: Refers to the minimum, maximum and average salary for that post
* **job\_state**: State where the applicant will work
* **same\_state**: Same state as headquarter or not(Boolean)
* **company\_age**: Age of company
* **python,excel,hadoop,spark,aws,tableau,big\_data:** Some most appeared skills in boolean columns form
* **job\_simp**: Job type
* **seniority**: if job type is senior or not (Boolean)

# **Dataset Cleaning Process**

Documented in the NOTEBook [here](Assignment.ipynb).