

Data Science Assignment

1. Read the data from “assign_chubb_engineering .json” and convert it into a table with columns:

Company uuid	Is score calculated
Company Name	List of companies they hired from
Employees Count	Employee count they hired
Fresher Count	List of companies their ex-employees work
IIT Count	Employee count their ex-employees work
IIT Fresher Count	Tier-1 College Density Score
NIT Count	Tier1 College Freshers Density Score
NIT Fresher Count	Tier-1 College Freshers Strength
IIT Count	Tier-1 College Strength
IIT Fresher Count	

2. Find number of Null values in each column
3. Add a column ‘List of companies where hiring companies hire from’. It contains companies where the hiring companies of current company hires its employees. For example, company A hire from companies (B, C, D) then ‘List of companies where hiring companies hire from’ for company A has all the companies where B or C or D companies hire
4. Based on Number of employees classify them into 9 categories and add a column with category number

Number of employees: Category Number

1-10 :1 ; 11-50 : 2 ; 51-100 :3 ; 101-250 :4 ; 251-500 :5 ; 510-1000: 6 ; 1001-5000: 7; 5001-10000:8 ;
10000+: 9

5. Find number of companies for each category
6. Find median and mean of tier-1 strength for each category
7. Display top 5 and bottom 5 companies based on Tier-1 College Freshers Strength for each category
8. Save the data category wise into separate excel sheet and name it as category_k.xlsx where k is the category number