DATA ANALYSIS USING PYTHON LIBRARIES ON JUPYTER NOTEBOOK

Points from the batch data set:

- 1). ON an average 33.33 people were placed in each batch
- 2) The Placement ratio on an average is 0.87
- 3) 79.6 percent of the total eligible candidates are placed
- 4) 20.3 percent are not placed
- 5) batches 8 and 4 have highest number of non eligible candidates /not opted for placement candidates
 - 6) 7 percent of the candidates enrolled have been opted out of placements
- 7) The avg CTC is 6.1, batch 9 has the highest avg ctc and a 100 percent placement ratio
- 8) A linear increasing trend between the no of students enrolled/eligible and the no of students placed
 - 9) decreasing trend on the number of people placed and avg CTC
- 10) batches 8 and 9 have the highest placement ratio above in 9 which means that 5 students not eligible for placements got placed here

Points from the other two data sets (PDGS and Batch)

1). A lot of missing data ,inconvenient data and null values were present in the data set which were cleaned and replaced appropriately

Batch 4 Batch 2 and Batch 8 have the highest number of people placed in companies

3). Descriptive stats of this is given below

	CTC in LPA
count	64.000000
mean	6.430781
std	2.593405
min	3.000000
25%	4.500000
50%	6.500000
76%	7.000000
max	20.000000

4). Intern, Consultant and Machine Learning Scientists are the highest designated jobs

5)

	Designation
6.860600	albà is escassi ancest clastes albă a bis ostină ALM, COMMING Uniferen
6.125800	Analyst
4.500000	Analysi Ceta Svience
6.800000	Assertate
6.790000	Associate Consultant
SIMBERRA	Material Scribson Ingland
5000000	N designal
7. 90000	Susiness Analysi
0.818:02	Consultant
6290900	Communities in Business Communiting Nick
7.860600	Evroporato Accessed Waterpoor
6.500000	Data Analysi
8.860800	Data Engineer
6.135800	Data Scientist
5.000000	Deputy Lint Manager- Man Control Cota Science
6.500000	Executive, Date Scientist
7 800000	Indeen
th 2000000	à. Bata Analysi
61960600	Jr. Technical Computant
5,860600	Junior Bata Anshut

6) Companies with most number of students

Onlowies 11
SC 9
SASTAR 4
Desiles 6
The Noth Company 5
Position 5
Position 5
Pateronee 5
Indian defense 5
Indian defense 5
Indian defense 6
Indian defense 6
Indian defense 7
Indian defense 7
Indian defense 8
Indian 6
Indian 7
Indian 6
Indian 1
In

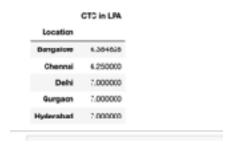
7). Senior Data Scientist, Senior data engineer, Software Engineer -Data Scientist receive the highest paid salary

8). Course data

PGDDS 10 28 PGDDS 8 27 PGDDS 9 20 PGDDS 7 2

Showing number of students per course

- 9). Most of the students placed are in Bangalore 92.2 percent
- 10)Mean CTC of Each city



Batch Data set:

-Batches with mean CTC and no of placements

	GTG in LPA	Gemoeny
Batch		
Batch 10	6.401275	29
Baton z	6.574589	39
Batch 2	6.013639	21
Batch 4	5.940984	61
Batch 5	4.812500	
Batch 3	6.460108	18
Batch 7	6.140695	34
Batch 3	5.908592	17
Batch 7	8.029007	18

-Male to female ratio almost 3:1 and the avg CTC for female being 0.15 times better

-Company info with CTC and most recurred company

	Company	
	Tilgent Softwares	21.000100
	Amazon	12.000000
	Annionlaha	12.000000
	Dailyhunt	11.000000
	Futurense	16.333333
	Climate Corporation	16,170600
	Adobe	9.000000
	Litmusworld	1.250100
	E Cube Analytics	8.133833
	LAT Infotech	4.000000
32]:	d2['Company'].	alue_coun
32]:	Equifax	17
	EY	10
	Senract	10
	Dhiomics	9
	Rielsen Gramener	9 7
	EY India	7
	TCS	6
	Inference Labs	6
	Easy Revardz	6

-In Batch 10 male female ratio is almost the same

[Batch 9 ,2,6 have highest CTC mean whereas 2,7,8 have highest number of students

-Batch 3 has the lowest male to female ratio inferred from the scatter plot $\,$

	Serial #	CTC in LPA
count	288,000000	288.000000
mean	133.500000	8.284329
std	76.931788	1.637233
min	1.000000	3.160000
25%	67.250000	5.000000
50%	133.500000	6.284329
75%	199.750000	7.000000
max	268,000000	20.000000

View Code for different visualisation cleaning and analysis