

# 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

2)

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        | 2.583405  |
| min        | 3.000000  |
| 25%        | 4.500000  |
| 50%        | 6.500000  |
| 75%        | 7.000000  |
| max        | 20.000000 |

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

5)

| Designation                                   |          |
|---|----------|
| AI ML Engineer                                | 6.800000 |
| Analyst                                       | 6.250000 |
| Analyst Data Science                          | 4.800000 |
| Associate                                     | 5.800000 |
| Associate Consultant                          | 6.100000 |
| Business Development Engineer                 | 6.500000 |
| BI Analyst                                    | 6.800000 |
| Business Analyst                              | 7.300000 |
| Consultant                                    | 6.810000 |
| Consultant in Business Consulting Risk        | 6.200000 |
| Corporate Account Manager                     | 7.800000 |
| Data Analyst                                  | 6.800000 |
| Data Engineer                                 | 8.800000 |
| Data Scientist                                | 6.150000 |
| Deputy Unit Manager-Data Control Data Science | 6.800000 |
| Executive, Data Scientist                     | 6.500000 |
| Intern  | 7.800000 |
| Jr. Data Analyst                              | 8.800000 |
| Jr. Technical Consultant                      | 6.800000 |
| Junior Data Analyst                           | 5.800000 |

## 6) Companies with most number of students

|                                |    |
|--------------------------------|----|
| Uniconics                      | 11 |
| SE                             | 9  |
| SASTAR                         | 4  |
| Qualifox                       | 4  |
| Fun Math Company               | 3  |
| Facilities                     | 3  |
| Patorusoo                      | 3  |
| Indian Software                | 3  |
| Inference Labs Pvt. Ltd        | 3  |
| Bridge Health                  | 3  |
| TVS Credit                     | 2  |
| Skiloon                        | 2  |
| Valigon                        | 1  |
| Quasr                          | 1  |
| Landing Tree                   | 1  |
| Deloitte                       | 1  |
| Litmus World                   | 1  |
| Cognitia Analytics             | 1  |
| WSP                            | 1  |
| Uris Health                    | 1  |
| Lyra Information               | 1  |
| Finserve India   Finserve Bank | 1  |
| Capco Technologies             | 1  |
| ANIMATIONS PVT LTD INDIA       | 1  |
| Q&A Industries                 | 1  |
| TRIGATE SOLUTIONS              | 1  |
| THE SMART CUBE                 | 1  |
| IMMERSA                        | 1  |
| ALICATE                        | 1  |

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

| CTC in LPA |          |
|------------|----------|
| Location   |          |
| Bangalore  | 6.384628 |
| Chennai    | 6.250000 |
| Delhi      | 7.000000 |
| Gurgaon    | 7.000000 |
| Hyderabad  | 7.000000 |

Batch Data set :

-Batches with mean CTC and no of placements

| CTC in LPA Company |          |    |
|--------------------|----------|----|
| Batch              |          |    |
| Batch 10           | 6.405275 | 29 |
| Batch 2            | 6.374981 | 29 |
| Batch 3            | 6.815639 | 31 |
| Batch 4            | 5.340984 | 61 |
| Batch 5            | 4.812500 | 8  |
| Batch 6            | 6.493108 | 19 |
| Batch 7            | 6.740686 | 34 |
| Batch 8            | 5.908592 | 37 |
| Batch 9            | 8.826937 | 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             |           |
|---------------------|-----------|
| legendssoftwares    | 28.000000 |
| Amazon              | 12.000000 |
| Amnionlabs          | 15.000000 |
| Dailyhunt           | 11.000000 |
| Futurense           | 16.333333 |
| Climate Corporation | 16.170000 |
| Adobe               | 9.000000  |
| Litmusworld         | 1.250000  |
| E Cube Analytics    | 1.133333  |
| LAIT Introspect     | 1.000000  |

  

```

32]: d2['Company'].value_counts().head(10)
32]: Equifax      17
      EY           10
      Genpact      10
      Dharmics      9
      Nielsen      9
      Traneer       7
      EY India      7
      PCS           6
      Inference Labs 6
      Easy Rewards  6
      Name: Company, dtype: int64

```

-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  | 138.500000 | 6.284329   |
| std   | 76.931788  | 1.037233   |
| min   | 1.000000   | 3.160000   |
| 25%   | 67.250000  | 5.000000   |
| 50%   | 138.500000 | 6.284329   |
| 75%   | 199.750000 | 7.000000   |
| max   | 288.000000 | 29.000000  |

View Code for different visualisation cleaning and analysis

