INNOMATICS®
RESEARCH LABS

INNOVATION. AUTOMATION. ANALYTICS

# PROJECT ON

## Analysis on AMCAT DATA

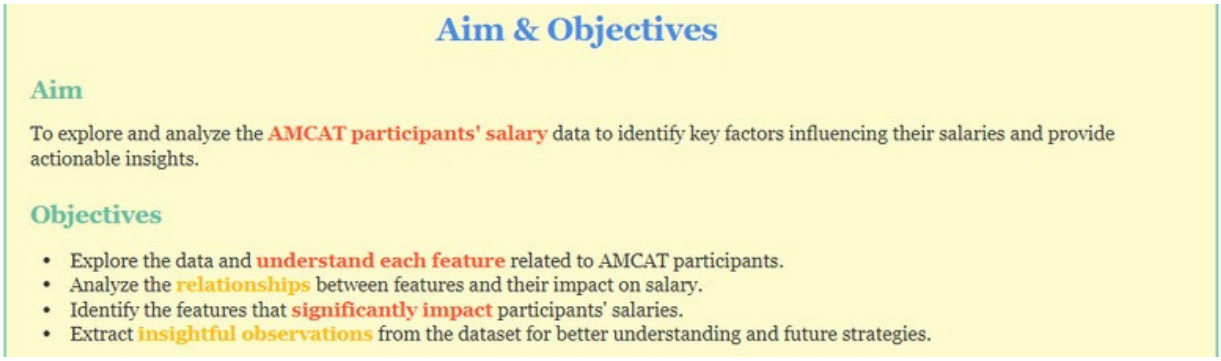
### About me:

My name is **MD QAMAR**. I completed my **Master's in Computer Application** from **National Institute of Technology**. I'm particularly fascinated by the fields of Machine Learning, Natural Language Processing (NLP), and Generative AI. I love exploring how data can be transformed into actionable insights, using my skills in mathematics and computing. My aspiration is to become an expert Data Scientist, dedicated to uncovering meaningful patterns in data to drive innovation and decision-making.

### About the data:

The dataset was released by Aspiring Minds from the Aspiring Mind Employment Outcome 2015 (AMEO). The study is primarily limited only to students with engineering disciplines. The dataset contains the employment outcomes of engineering graduates as dependent variables (Salary, Job Titles, and Job Locations) along with the standardized scores from three different areas – cognitive skills, technical skills and personality skills. The dataset also contains demographic features. The dataset contains around 40 independent variables and 4000 data points

### Objectives:



## Data Description:

The dataset consists of 4000 data points with 40 columns. The features are of type categorical, numeric discrete and numeric continuous. Following are the categories in which the different columns come.

Numerical Discrete Features : 'English', 'Logical', 'Quant', 'ComputerProgramming', 'ElectronicsAndSemicon', 'ComputerScience', 'MechanicalEngg', 'ElectricalEngg', 'TelecomEngg', 'CivilEngg'.

Numerical Continuous Features: 'Salary', '10percentage', '12percentage', 'collegeGPA', 'Domain', 'conscientiousness', 'agreeableness', 'extraversion', 'nueroticism', 'openess_to_experience'.

Categorical Features: 'Designation', 'JobCity', 'Gender', '10board', '12board', 'Degree', 'Specialization', 'CollegeState', 'CollegeTier','CollegeCityTier'.

## Data Cleaning:

The datatypes are fixed based on the features identity. There are no missing values in the data provided.

## Analysis:

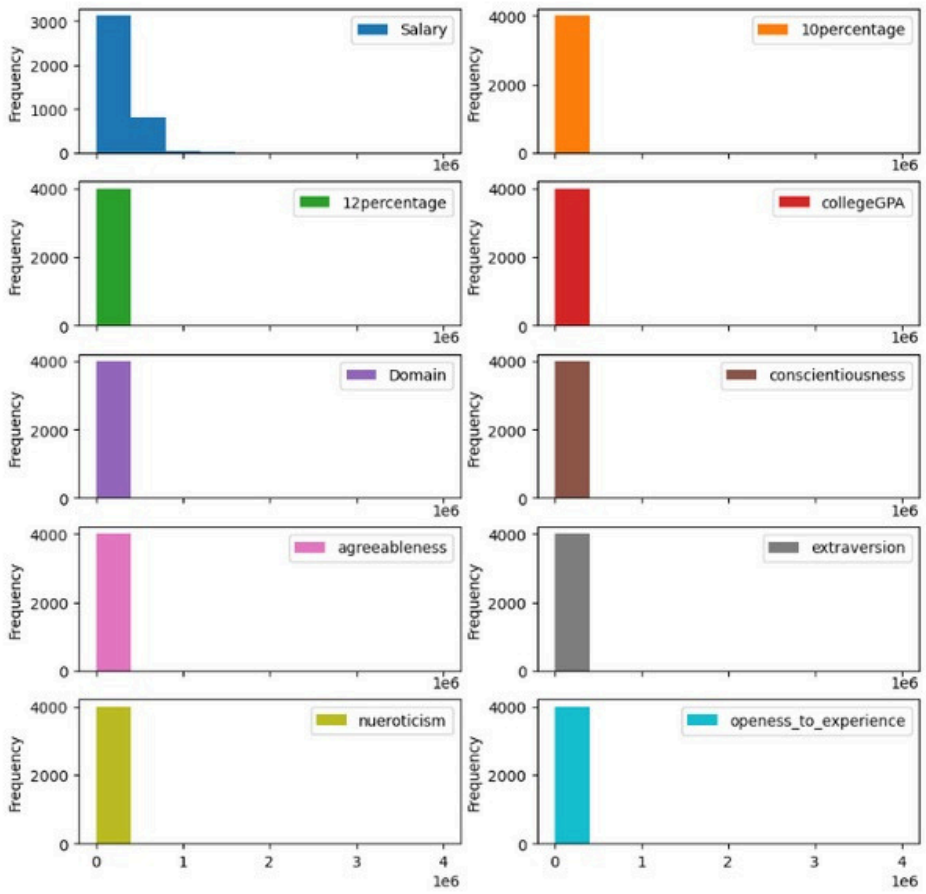
- To summarize and describe the each features univariate analysis needs to be performed.
- To find the relationship between each features bivariate analysis needs to be performed.

## Outcome of Univariate Analysis

- Most of the students are from CBSE board both for 10th and 12th.
- After the secondary education maximum students completed their degree in B.Tech /B.E.
- Maximum number of students has taken electronics and communication engineering as their specialization.
- Many of the students have completed their college in the UP state.

- Maximum salary earned by a student as per data is 4 crore and the mean salary is 3.0 laks

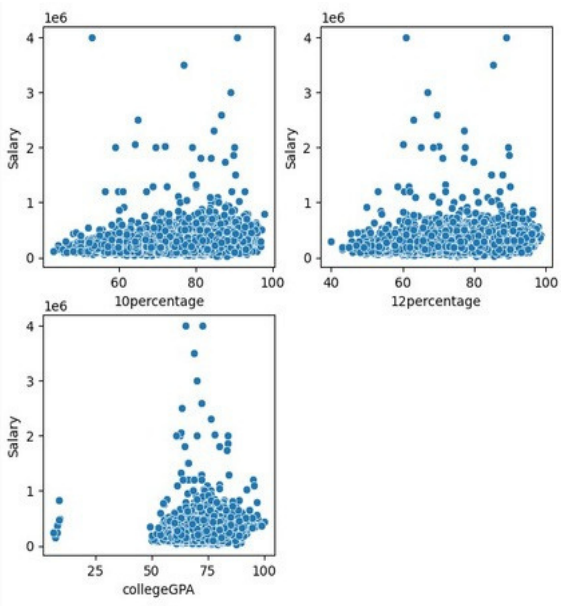
- Many other students have completed their studies from Tier 2
- colleges. Presence of outliers in almost all numerical continuous features.




- **Salary Distribution:** The majority of salaries are concentrated at the lower end of the distribution, with very few participants earning higher salaries.
- **10th and 12th Percentage:** Both the 10th and 12th-grade percentages are consistently high across the dataset.
- **College GPA:** Similar to the 10th and 12th percentages, college GPAs are concentrated at high values.

- **Personality Traits:** Traits like conscientiousness, agreeableness, extraversion, neuroticism, and openness to experience seem to show little variation, indicating that most participants have similar levels for these personality features.
  **Domain:** The field of study or domain also shows little variation, suggesting that most participants are from a similar background.

## Outcome of Bivariate Analysis:




- Based on the correlation analysis

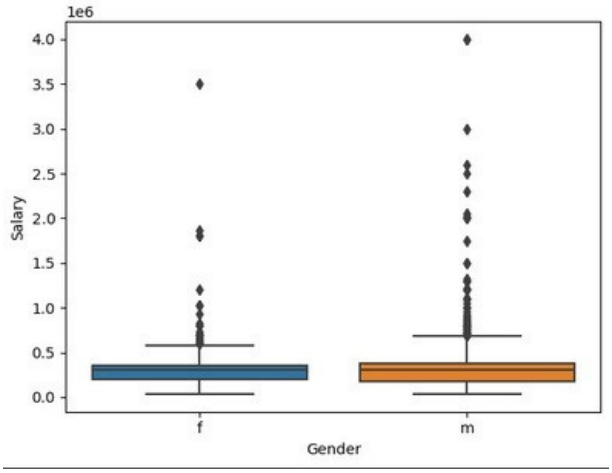
  10th percentage, 12th percentage and collegeGPA all shows positive correlation with the salary of the students. That indicates higher the percentage earned by the student more the salary is.

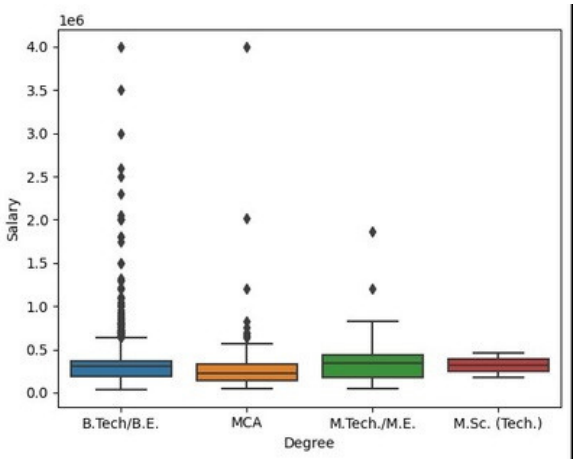
- After performing the cross tabulation, It is observed that maximum of the females have taken biomedical engineering, Information and communication technology. It is also observed that there are Specializations which are only taken by males Such as industrial and management Engineering, instrumentation engineering etc.

- Mean salary earned by female and male

  Candidates are almost same but the maximum is quite high for male students.

- Candidates with B.Tech degree are earning more compared to other degrees. So there is impact of degree also on Salary.




## Research Question:

'Times of India article dated Jan 18, 2019 states that "After doing your Computer Science Engineering if you take up jobs as a Programming Analyst, Software Engineer, Hardwar Engineer and Associate Engineer you can earn up to 2.5-3 lakhs as a fresh graduate."Test this claim with the data given to you.'

To test this claim, let's use the method of Hypothesis testing with given mean and standard deviation unknown given by the following formula,

$$t = \frac{\bar{x} - \mu_0}{s \,/\, \sqrt{n}}$$

Null hypothesis here is the claim, where mean = 2.7 lakhs Alternative hypothesis here is to reject the claim. By applying the t-stats testing on the sample made according to the claim from
the overall sample it is found that the claim is true.

```
Significance level: 0.05
Degrees of freedom: 73
Test statistic: -0.2272
Critical value: -1.993
Reject H0: False

False
```

This claims that a fresher with Computer Science engineering can choose his career as Programmer analyst, software or hardware engineer and associate engineer and can earn up to 2.5-3.0 lakhs.

## Conclusions:

- Specializations are gender specific. The grades obtained by students have
- impact on their Salary. Salary also depends on the degrees obtained by the
- students. In few cases males are earning higher than the females. Based on
- the research question a student with computer science engineer can earn up
- to 2.5 to 3.0 lakhs as a fresher.