PROJECT ON

Analysis on AMCAT DATA

About me:

My name is **PAVANI GOGULA**. I currently pursuing BTech in Computer Science and Data Science Engineering from **Vignana Bharathi institute of technology**. I have a keen interest in **Machine Learning**, **NLP**, and **Generative AI**. This exciting journey of Data Science allows me to apply mathematics, logic, and computing in practical ways. My goal is to become a proficient Data Scientist, specializing in data and insights generation.

About the data:

The dataset was released by Aspiring Minds as part of their 2015 Employment Outcome Study (AMEO). It mainly focuses on engineering students. It includes information about the employment results of engineering graduates, such as salary, job titles, and job locations, which are the dependent variables. Additionally, it has standardized scores for cognitive, technical, and personality skills, as well as demographic details. The dataset contains about 40 independent variables and 4,000 data entries.

Objectives:

Aim & Objectives

Aim

To explore and analyze the **AMCAT participants' salary** data to identify key factors influencing their salaries and provide actionable insights.

Objectives

- Explore the data and **understand each feature** related to AMCAT participants.
- Analyze the **relationships** between features and their impact on salary.
- Identify the features that **significantly impact** participants' salaries.
- Extract **insightful observations** from the dataset for better understanding and future strategies.

Data Description:

The dataset has 4,000 entries and 40 columns. The features are a mix of categorical, numeric discrete, and numeric continuous types. Here's how the columns are categorized:

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

- **Numerical Continuous Features**: 'Salary', '10percentage', '12percentage', 'collegeGPA', 'Domain', 'conscientiousness', 'agreeableness', 'extraversion', 'neuroticism', 'openness_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 en gineering 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 lakhs.

- Many other students have completed their studies from Tier 2 colleges.
- Presence of outliers in almost all numerical continuous features.
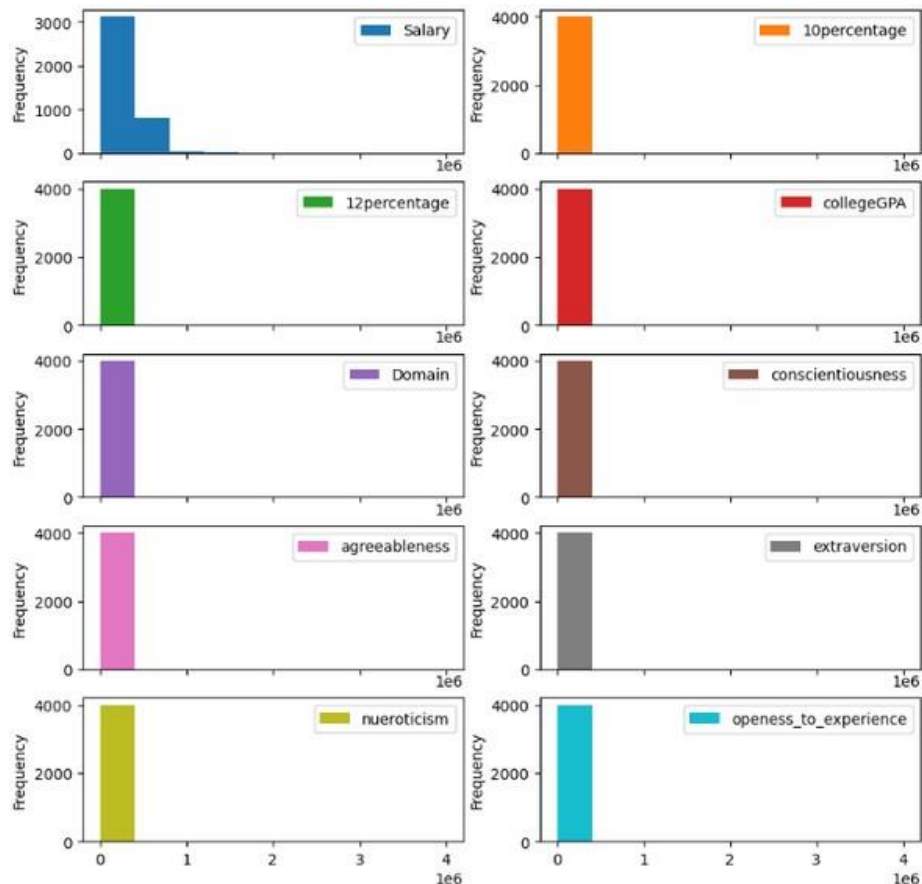
- **Salary Distribution:** The majority of salaries are concentrated at the lower end of the distribution, with very few participants earning higher salaries.
- **10th and 12th Percentage:** Both the 10th and 12th-grade percentages are consistently high across the dataset.
- **College GPA:** Similar 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 college GPA all shows positivecorrelation with the salary of the students. That indicates higher thepercentage 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.
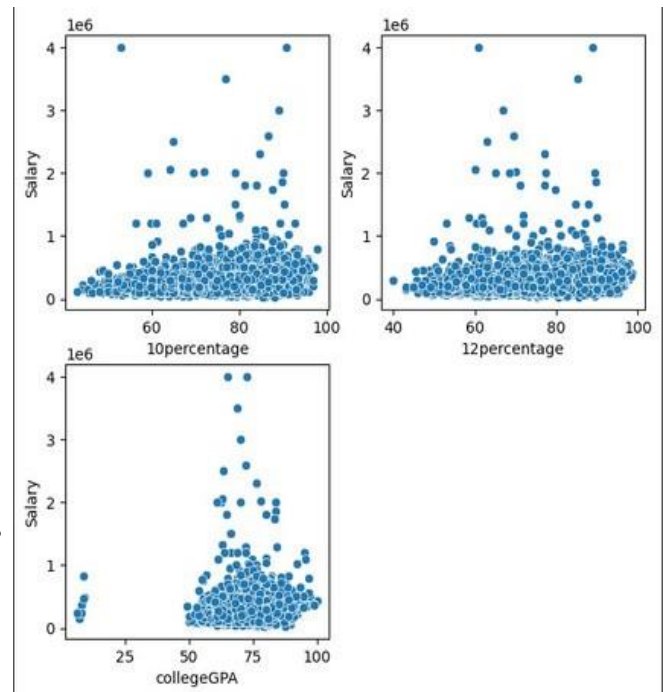
- Candidates with B.Tech degree are earning more compared to other degrees. So, there is impact of degree also on Salary.
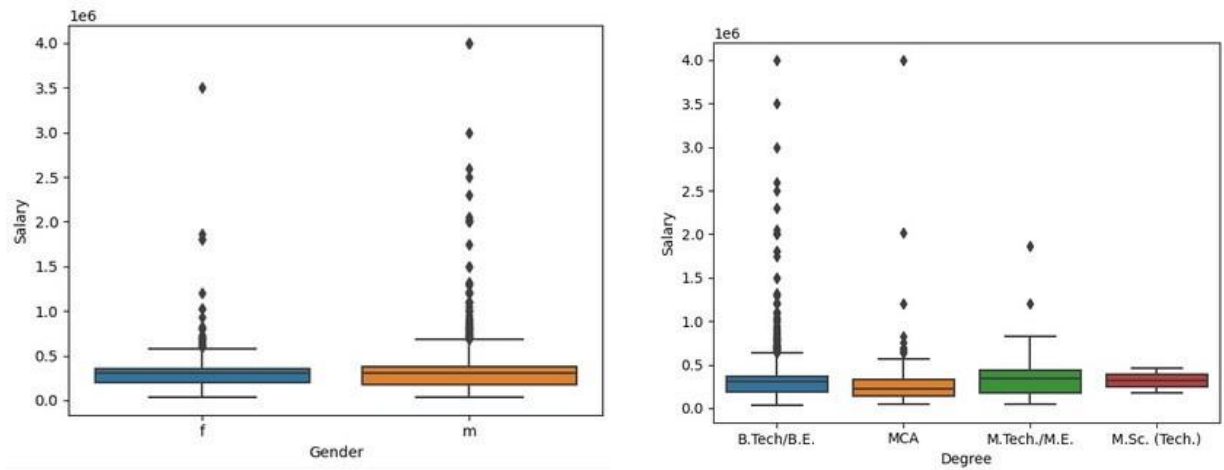
Research Question:

'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.