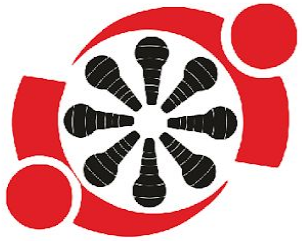


INNOMATICS®
RESEARCH LABS

INNOVATION. AUTOMATION. ANALYTICS

PROJECT ON

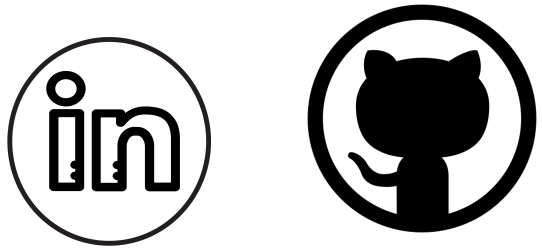
EDA Project - Analysis of AMCAT Data

By

Utsav Chaturvedi - IN9240497

About me

- I completed my bachelor ’s degree in Computer Science and Engineering from SRM University . I'm captivated by how data science can uncover meaningful patterns and insights from what might initially seem like chaotic data. It feels like having a unique ability to unveil the hidden structure of the world around us. I appreciate how data science empowers both individuals and organizations to make informed, data-driven decisions, leading to better outcomes across various domains. The constantly evolving field, with its continuous advancements and discoveries , keeps me engaged and fuels my desire for lifelong learning.

- Connect with ME 

INNOMATICS RESEARCH LABS

Agenda

1. Business Problem
2. Objective of the project
3. Summary of Data
4. Data Cleaning
5. Univariate Analysis
6. Bivariate Analysis
7. Key Business Question

Business Problem :

India produces 1.5 million engineers every year. A relevant question is what determines the salary and the jobs these engineers are offered right after graduation. Analyse the Aspiring Mind Employment Outcome 2015 Data to generate some insights.

Objective of this project :

The goal of this project is to conduct exploratory data analysis on AMEO data and verify the authenticity of The Times of India's report, as well as to answer other research issues.

Summary of Data:

The dataset was released by Aspiring Minds from the Aspiring Mind Employment Outcome 2015 (AMEO).The study is mostly limited to students studying engineering subjects. The dataset includes engineering graduates' career outcomes as dependent variables (salary, job titles, and job locations), as well as standardized scores in three areas: cognitive skills, technical skills, and personality skills. The dataset also includes demographic information. The dataset includes approximately 40 independent variables and 4000 data points.

Data Cleaning:

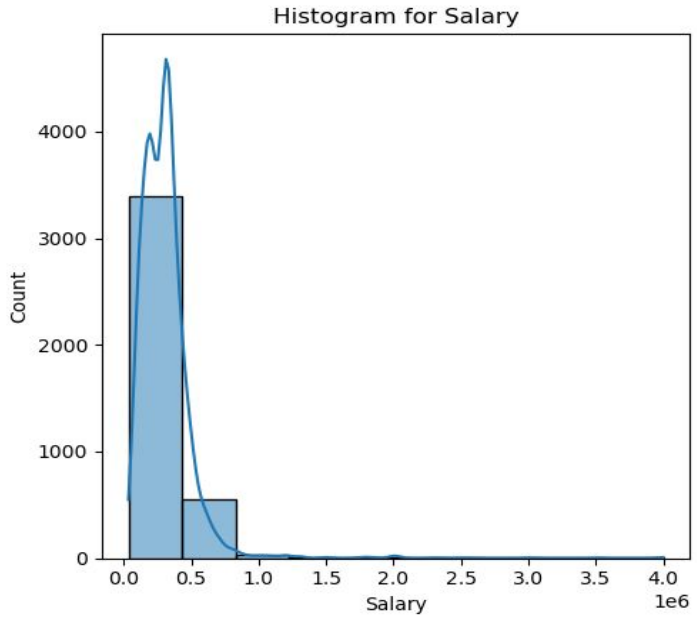
1. Dropped the unwanted columns.
2. Renamed the attribute names.
3. Changed the data type of some attributes to appropriate format.
4. Corrected the spelling errors present in the Job_city attribute.
5. Replaced the redundant values of the 10th Board and 12th Board column with 'State Board','CBSE','ICSE' .
6. Replaced each specialization with appropriate short form

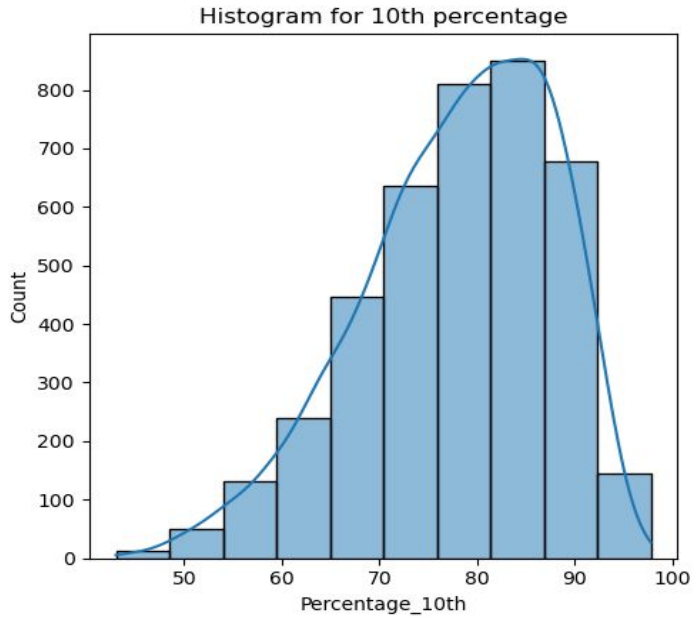
Univariate Analysis:

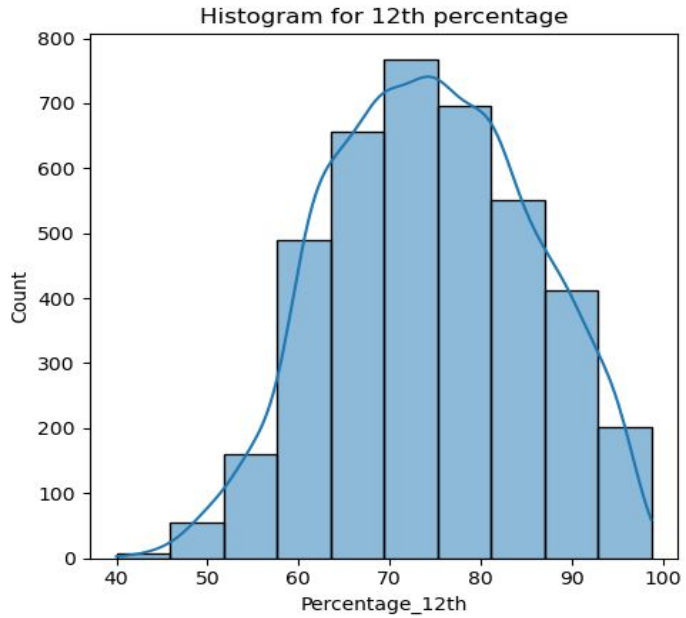
1. Univariate non-visual analysis was performed on the numerical attributes, displaying the minimum, maximum, mean, median, and standard deviation for each attribute.
2. Univariate visual analysis was applied to numerical attributes using visualizations such as histograms and kde plots. The insights obtained from each visual plot are described below.

Insights Generated from Univariate Analysis on numerical attributes :

1. Most of the people are earning below 1000000.
2. Most of the people scored in the range of 75 to 85 percent in their 10th Standard.
3. Most of the people scored in the range of 70 to 75 percent in their 12th Standard.
4. In AMCAT English, Logical, Quant Sections most people scored between 500 to 600 marks.
5. Most of the people scored less than 100 in Computer Science, Mechanical Engg,Electrical Engg Sections.
6. In the Personality Sections most of the people got scores in the range between -1 and 1.

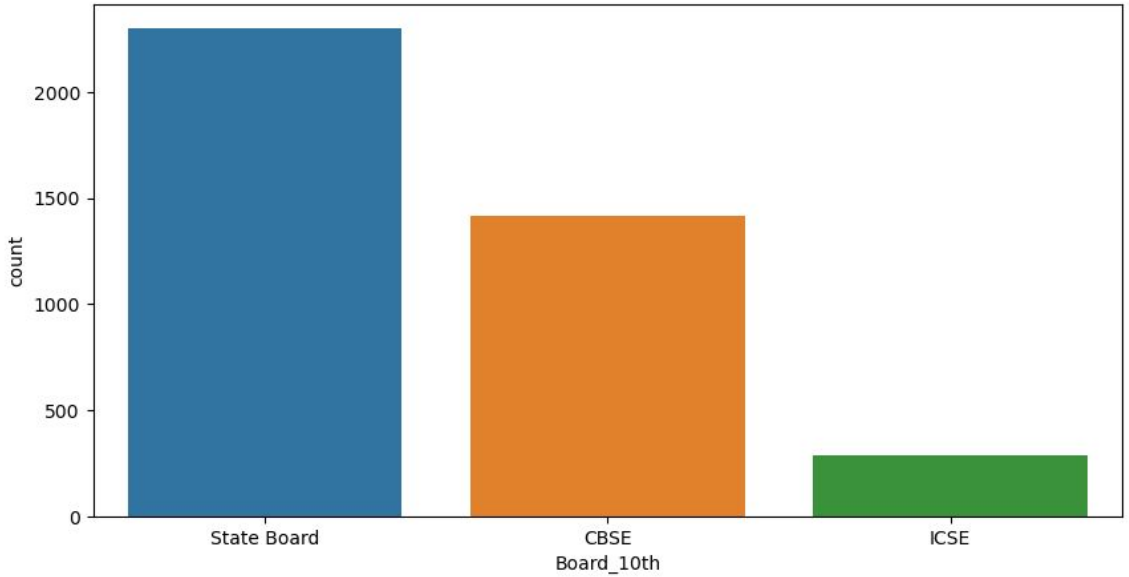


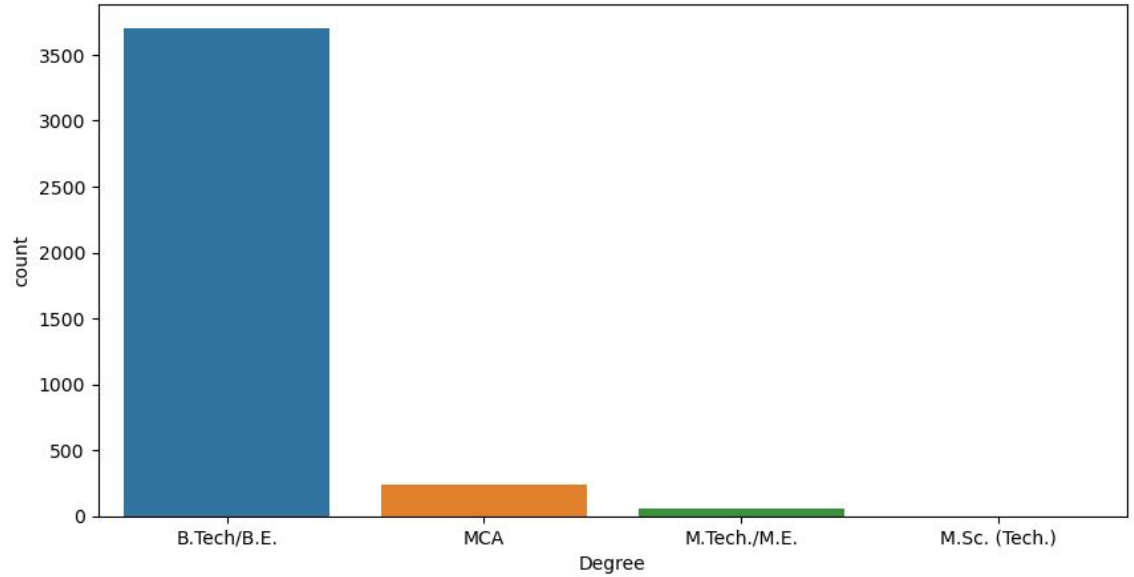


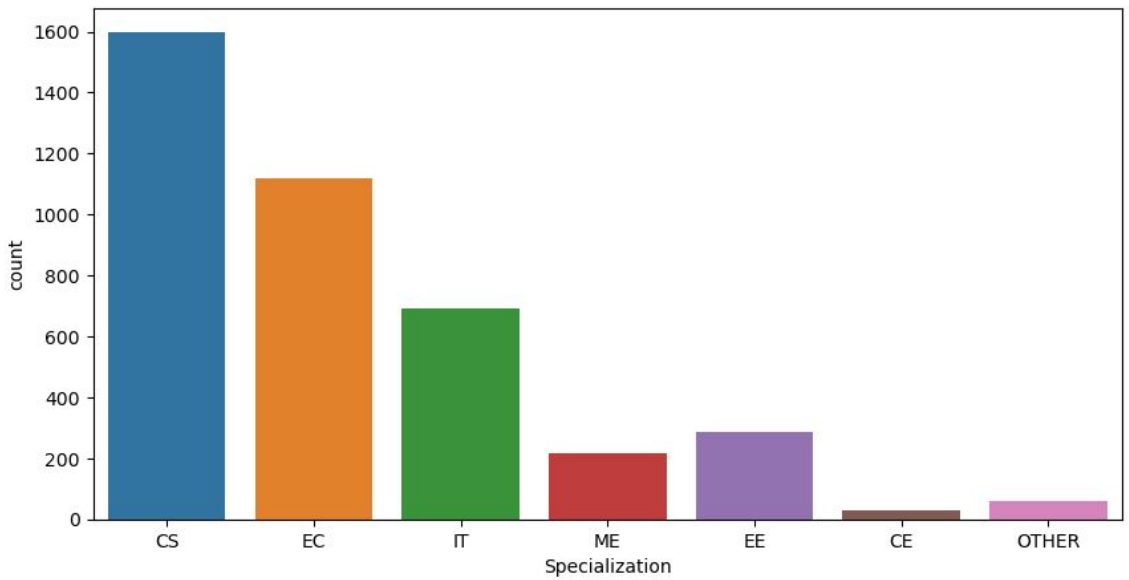


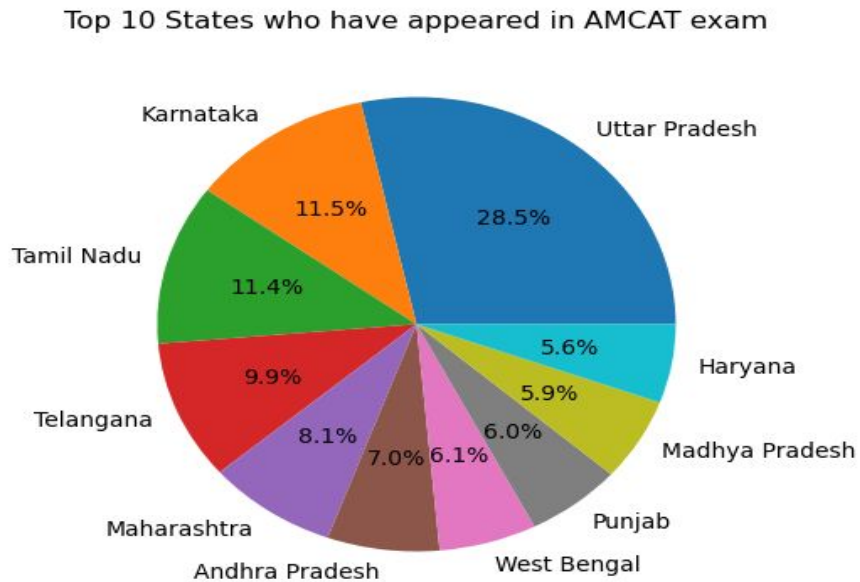
Insights generated from univariate analysis on categorical attributes

1. Male professionals are in large number when compared to that of female professionals.
2. Most of people completed their 10th standard from State Board or CBSE.
3. Most of people completed their 12th standard from State Board or CBSE.
4. Majority of the people enrolled in B.Tech/B.E Degree.
5. Most people are from Computer Science, Information Technology and Electronics Communication background.
6. Most of the people appeared for AMCAT exam are from Uttar Pradesh.









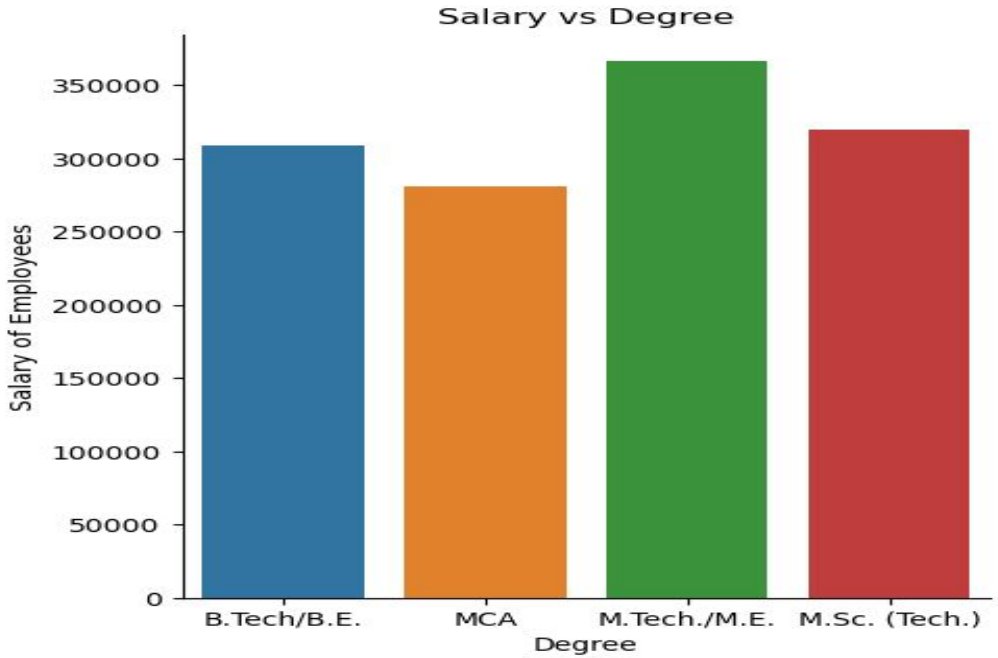
INNOMATICS RESEARCH LABS

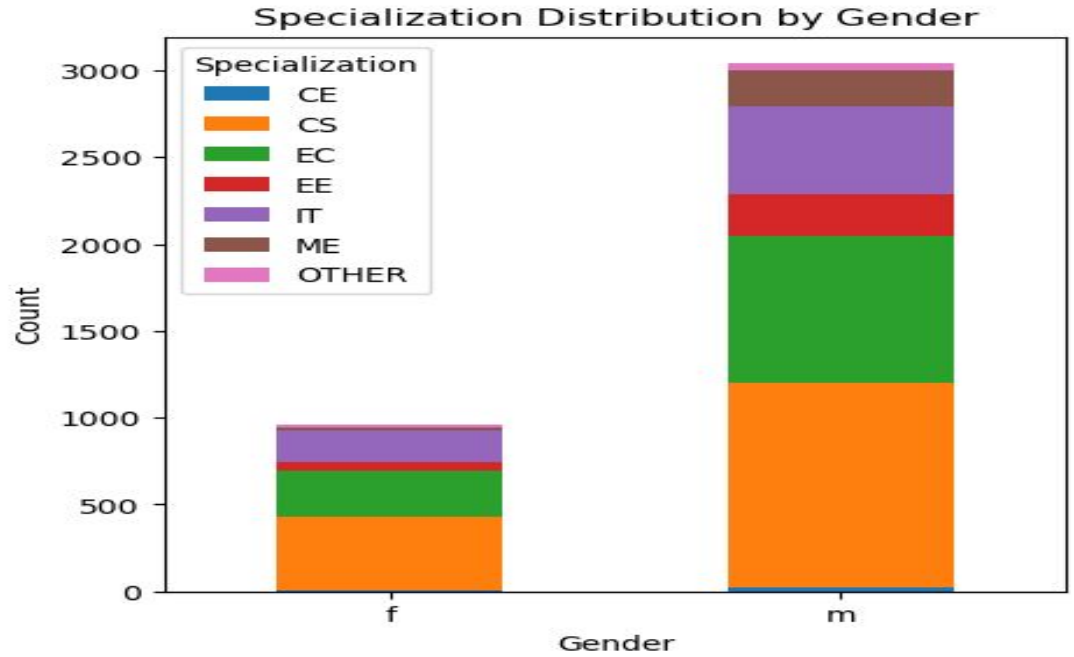
Bivariate Analysis:

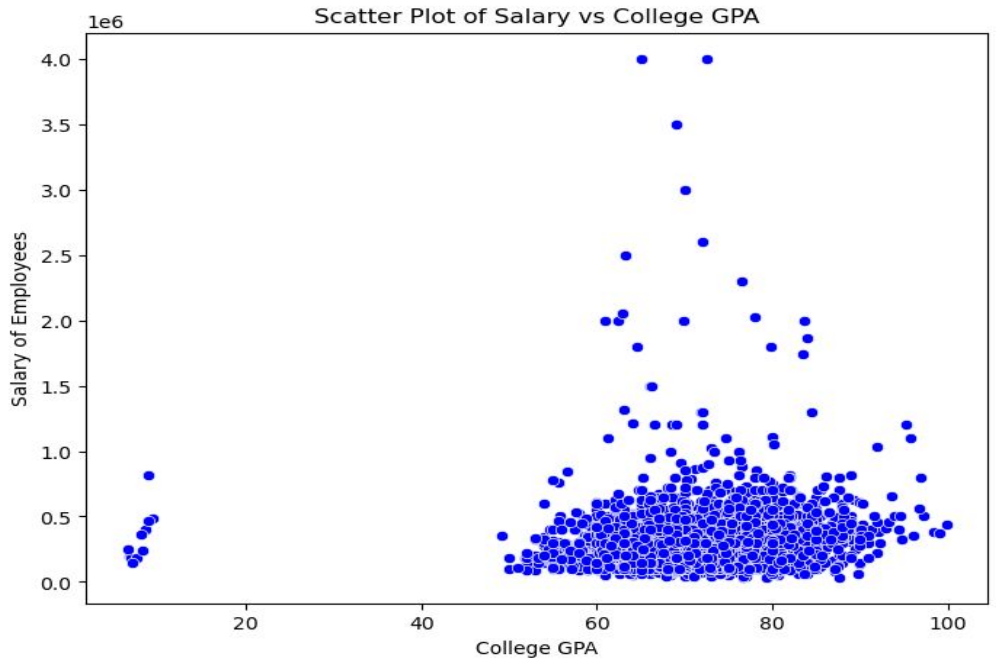
Bivariate analysis is performed on both numerical and categorical attributes using various visualizations, and the resulting insights are discussed below the visuals.

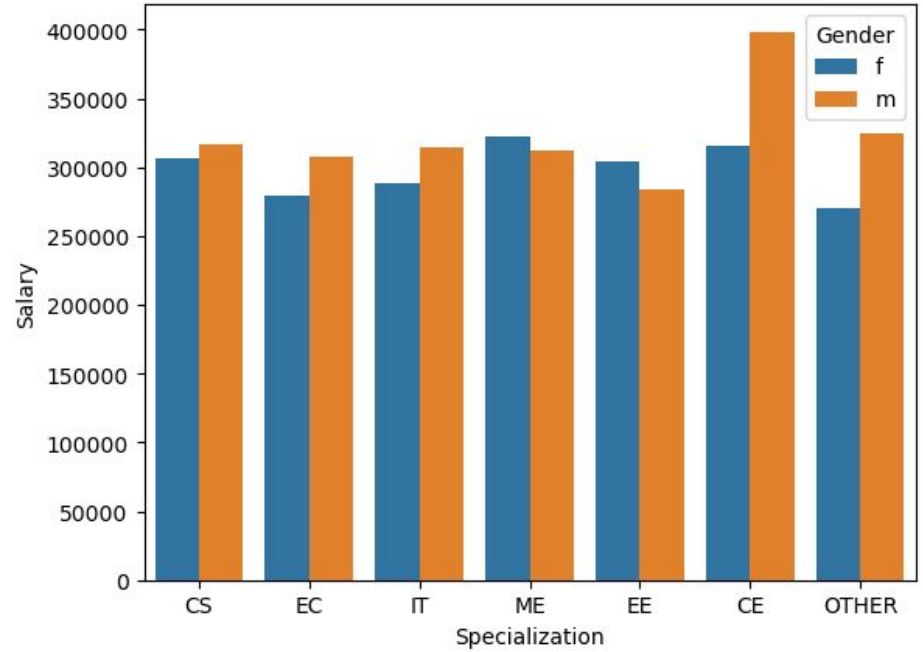
Generated insights are :

1. Majority of the people have their college GPA in between 60 to 90 and earn in the range of 2 Lakhs to 10 Lakhs per year.
2. People who Completed M.Tech./M.E. Degree Earn more.
3. More number of Male and Female candidates have studies in Computer Science followed by Electronics Communication.
4. Males have slightly better pay than Females.
5. People who did not provide their 12th Board and people from ICSE Board are earning more salary.
6. Males from Civil Engineering background are earning more.









Key Business Question:

1. **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, Hardware 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.**

   The data is subjected to a one-sample t-test to determine the article's validity.Based on the findings of the statistical test, we can conclude that the article is invalid.

2. **Is there a relationship between gender and specialization? (i.e. Does the preference of Specialisation depend on the Gender?)**

   The data is subjected to Chi square test to find if there is any relation between gender and specialization. Based on the findings, we can conclude that there is no relation between gender and specialization.

Creating a Research Question:

Is there any Relation between the College GPA and Salary.

Based on the results of the pearson's test, we can say that salary earned doesn't depend on the College GPA

THANK YOU

INNOMATICS RESEARCH LABS