



INNOMATICS[®]
RESEARCH LABS

INNOVATION. AUTOMATION. ANALYTICS

PROJECT ON
AMEO dataset EDA

About Me

K. Sunitha Devi

M.C.A

- **I have around one and half years of experience as a Test Engineer.**
- **I just want to update myself with the new technology, that is the reason I have chosen Data Science.**

Problem Statement

- AMEO dataset provides anonymized bio data information along with their respective skill scores and employment outcome information. As part of the challenge, the aim is to find the relationship between salary vs remaining features to know which are more affecting the salary column. Based on the analysis finding the insights.

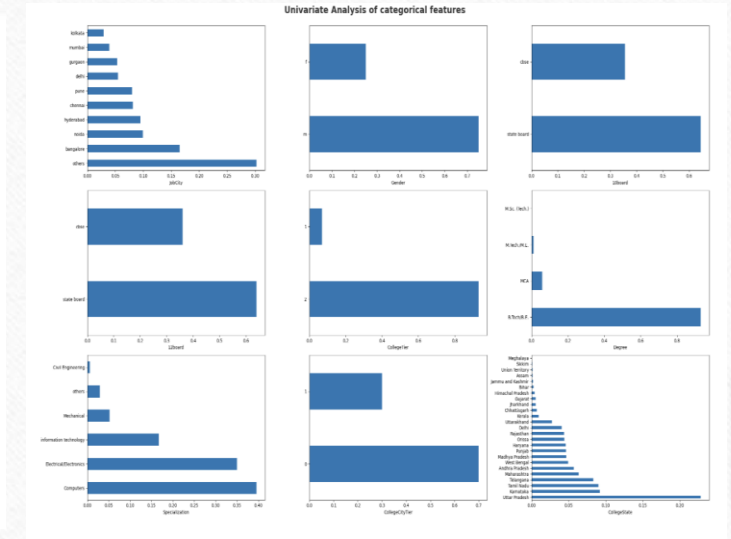
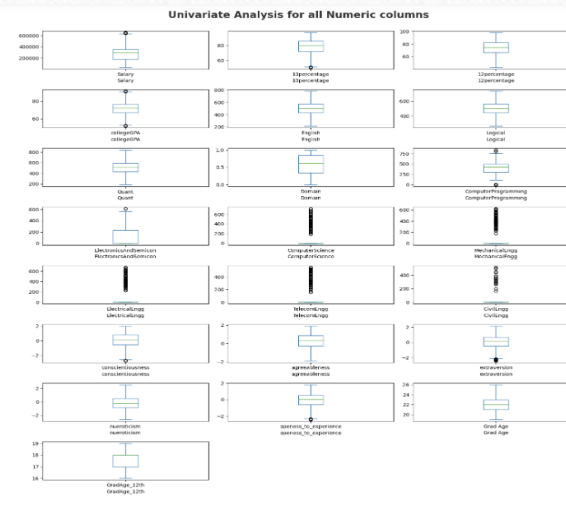
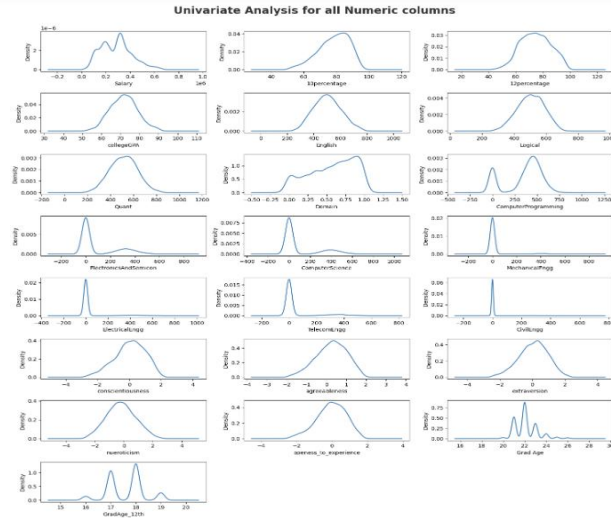
Brief About Data

- Most of the engineering students in the data are males.
- Majority of the students in 10th and 12th standard are from state board followed by CBSE board.
- Tier 2 colleges dominate the data just like they dominate in the real world meaning that only a few colleges are tier 1 colleges
- B Tech/ B.E clearly dominates other degrees in terms of frequency of occurrence
- The specialization has been manually categorized into 5 buckets with majority of students specializing in computer science, followed by Electronics/Electrical, IT, Mechanical and others.
- Majority of the data belongs to students who went to a college in Uttar Pradesh Data was checked for all the assumptions of linear regression and all of them were duly satisfied!

Data Cleaning and Manipulation

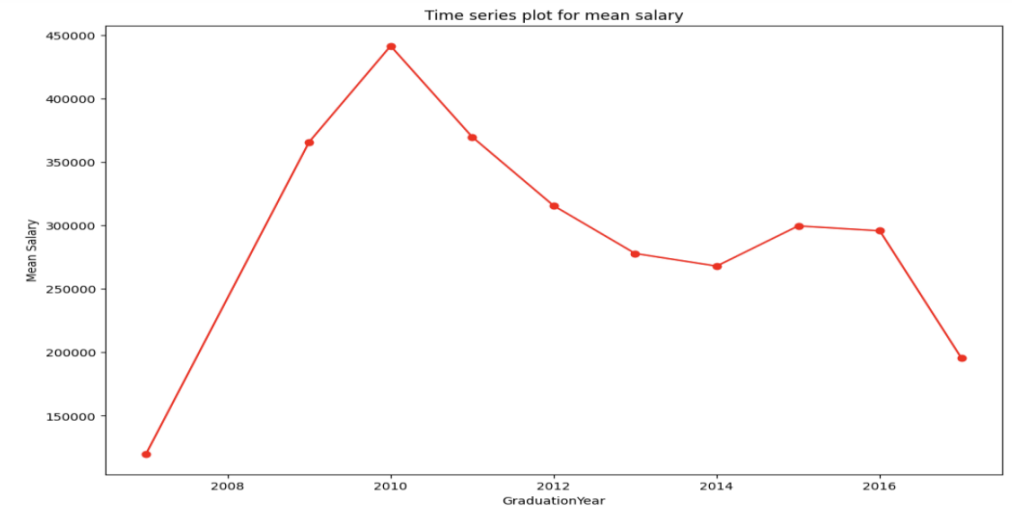
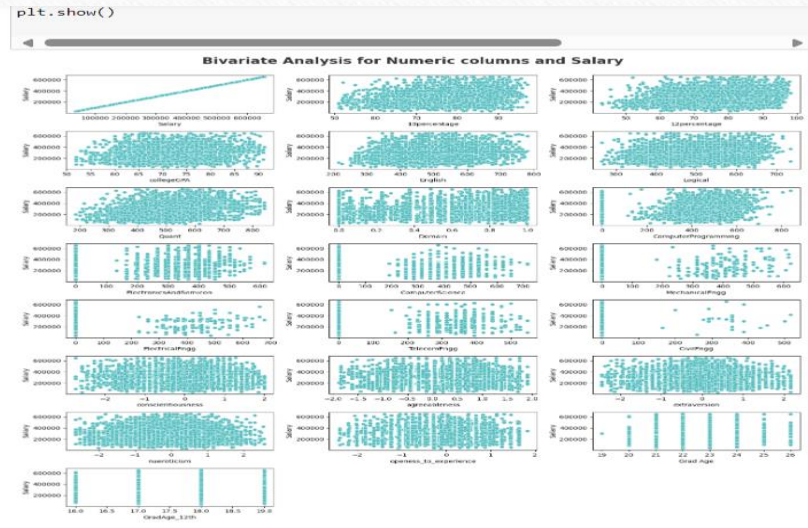
- Dropping College and college city ID' -- they are same (around 1350 unique id's)
- Dropping columns which are unnecessary or may not be known prior to receiving a job offer.
- Changing the data type as required.
- If a student hasn't opted/attempted an optional test, the same has been mentioned as -1. I will be replacing the same with 0.
- On observing the data, some graduation years were mentioned as 0, replacing graduation year 0 with the modal graduation year.
- we have many boards so i have converted these multi class to binary as cbse or state board.
- We have many specializations . I have to converted these many classes to 6 specializations.

Univariate Analysis



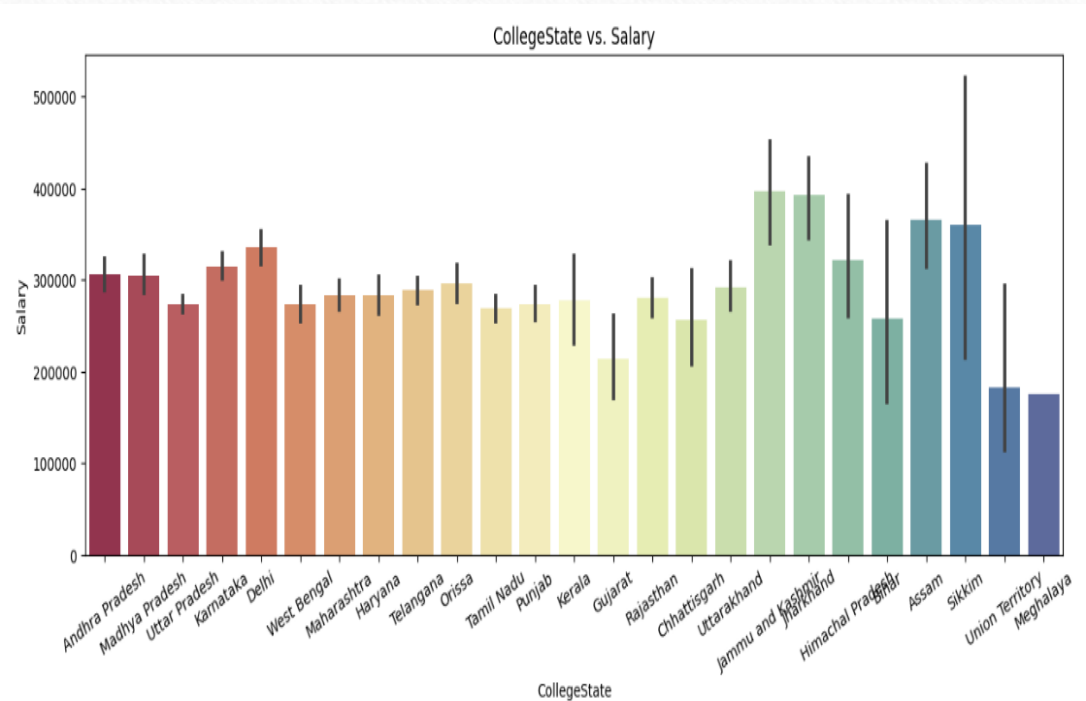
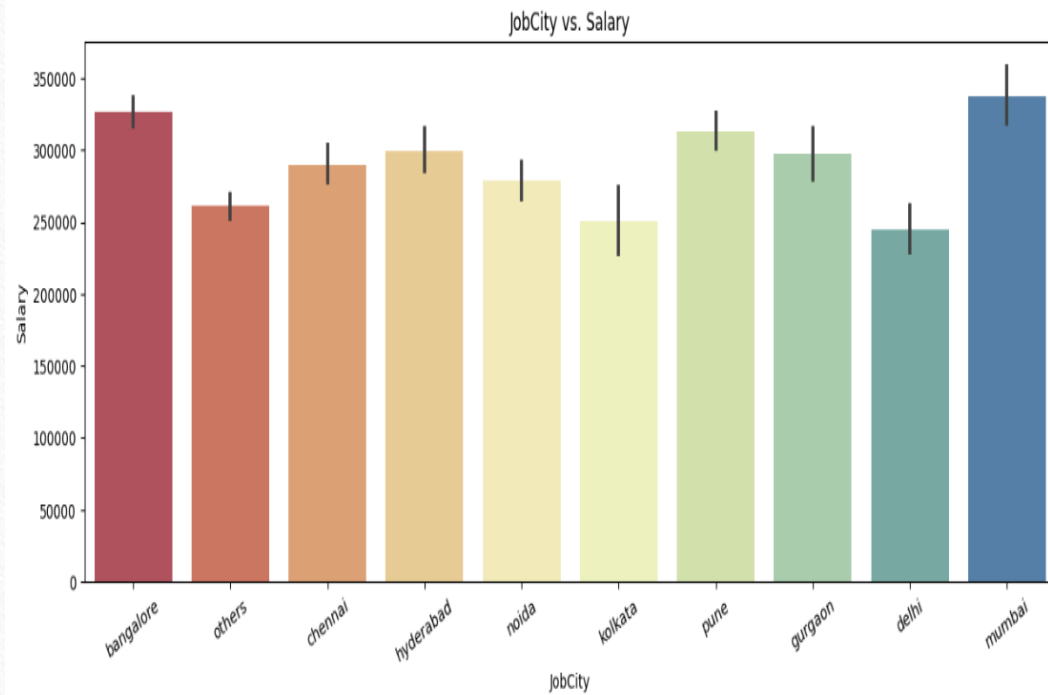
- Most of the engineering students in the data are males
- Majority of the students in 10th and 12th standard are from the state board followed central board.
- Tier 2 colleges dominate the data just like they dominate in the real world meaning that only a few colleges are tier 1
- B Tech/ B.E clearly dominates other degrees
- The specialization has been manually categorized into 6 with majority of students specializing in computer science, followed by Electronics/Electrical, IT, Mechanical and others
- Majority of the data belongs to students who went to a college in Uttar Pradesh

Bivariate Analysis



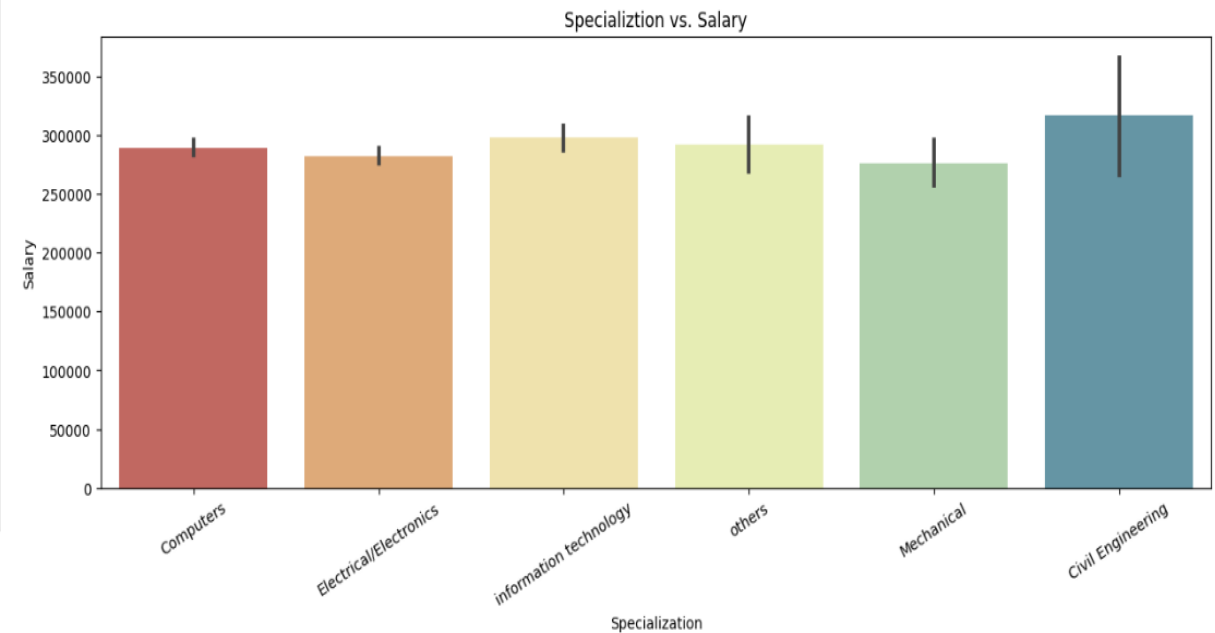
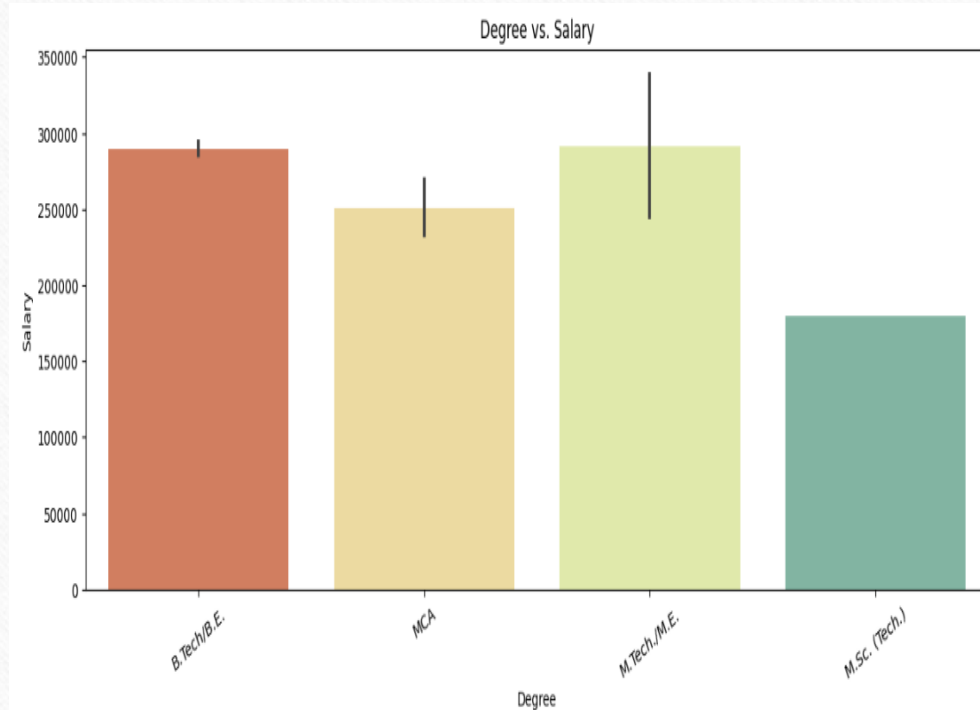
- A rise in entry level average salaries of students till 2010. since 2010 there has been a sharp decline in the average salary every year upto 2014. The salaries increase again from 2014 to 2015. A sharp decline again between 2016 and 2017.
- None of the features is showing a very strong linear relationship with the target variable salary.
- Out of all the given variables, it appears that quant skills are the most important.
- 10th percentile is positively correlated with 12th percentile.

Bivariate Analysis



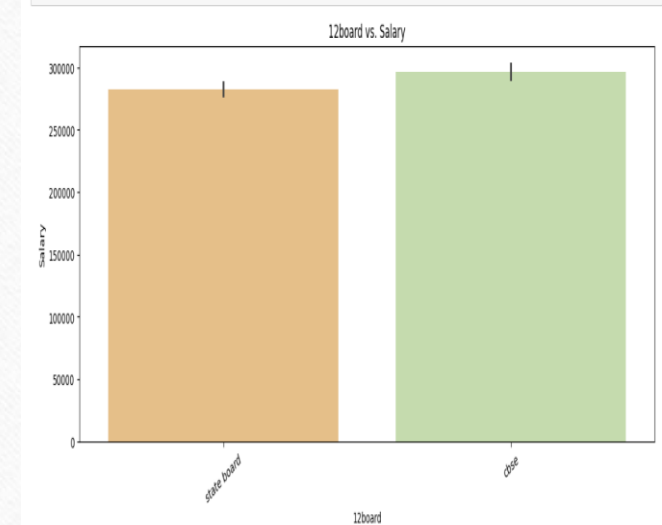
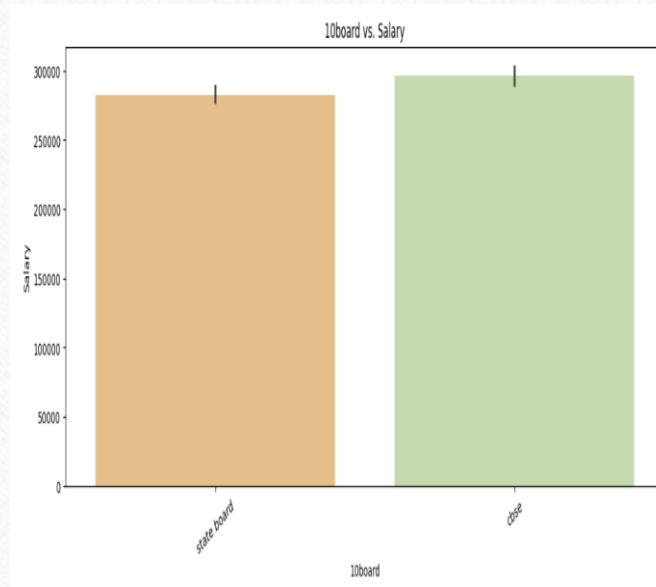
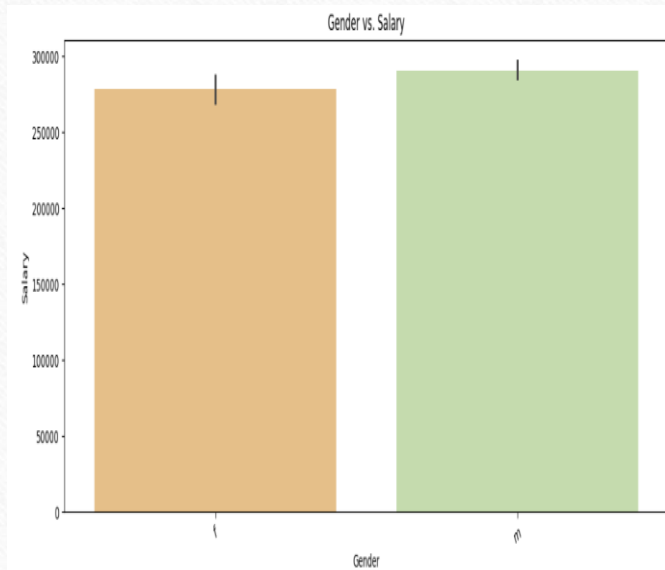
Mumbai, Bangalore, Pune and Hyderabad have the highest entry level mean salary while it is surprising to note that cities like Delhi and Kolkata have the lowest entry level salaries for engineering graduates.

Bivariate Analysis



- Mean salary of Mtech / M.E and Btech / B.E students appear to be slightly higher than other degrees.
- Average Starting salary of students having Specialization in Civil Engineering is higher relative to other fields.

Bivariate Analysis



- Mean salary of males appear slightly more than females.
- Board appears to be having significant difference in the salary of students with CBSE having higher mean salary.

Insights

- Out of all the given variables, it appears that quant skills are the most important in terms of Pearson's correlation coefficient. This means higher the score of candidate in quant, higher are the chances of having better salary prospects!
- 10th percentage is positively correlated with 12th percentage
- Scores in AMCAT's Logical aptitude test is having a positive correlation (moderate) with English and Quant tests.
- In the AMCAT's personality test, it appears that openness to experience is having a 44% positive correlation with agreeableness in the candidates.
- Mumbai, Bangalore, Pune and Hyderabad have the highest entry level salary while it is surprising to note that cities like Delhi and Kolkata have the lowest entry level salaries for engineering graduates. This was a shocker since these are the metropolitans!
- Mean salary of M.tech / M.E and B.tech / B.E students appear to be slightly higher than other degrees.
- Another interesting insight is that lower salary quartile (first quartile) of civil engineering students is higher than any other specialization. So do advice your juniors to choose civil engineering !

THANK
YOU

