Nathan Alanmanou
DATA 11900

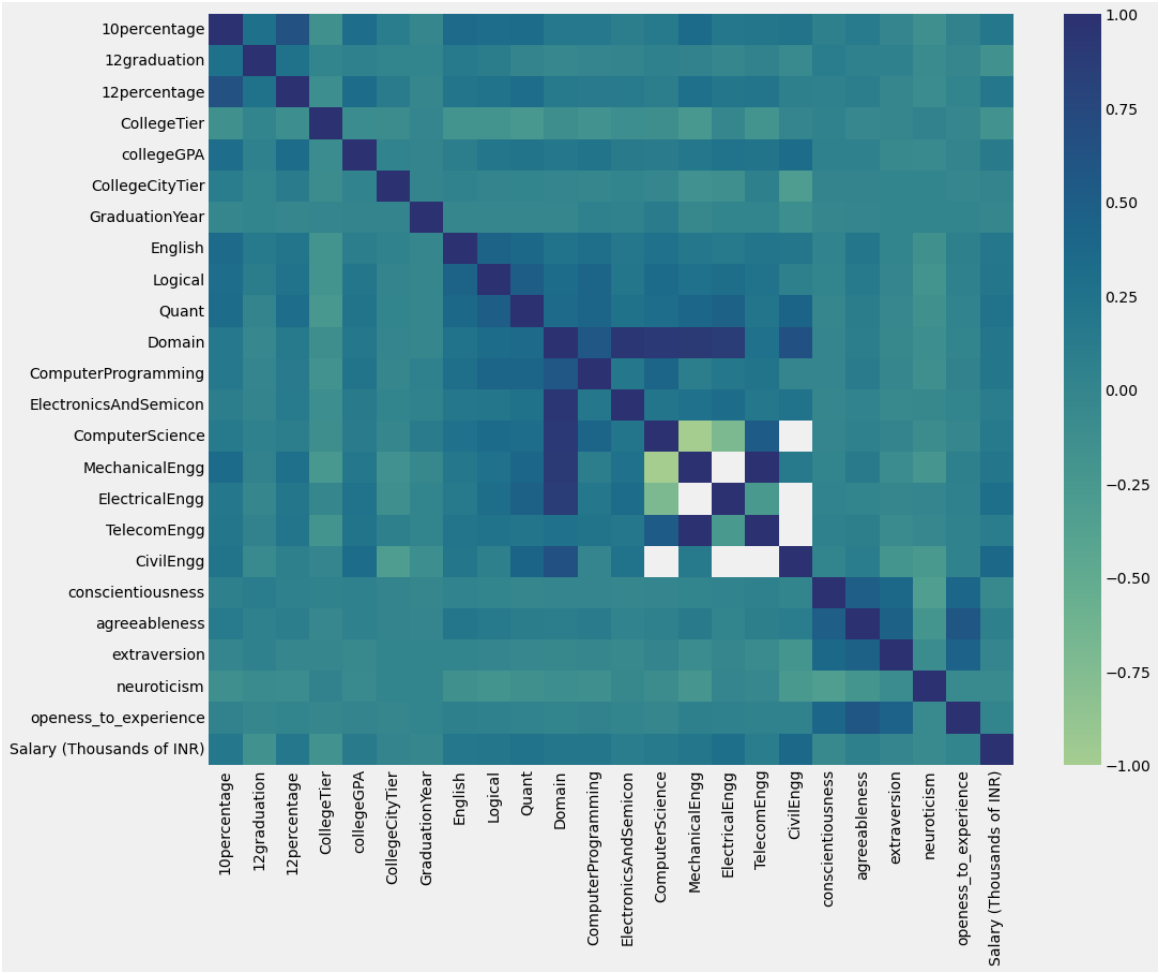
Analyzing the Demand of Various Metrics in the Engineering Job Market

This project will be an exploratory data analysis on the salaries that Indian engineering students are offered upon graduating. The full dataset consists of 2,998 students and 34 columns, including their specialization, college GPA, test scores, and starting salary, among others. The dataset used in this project was accessed through Kaggle.com on February 11, 2023. It can be accessed through the bibliography located after the conclusion. Analysis of the dataset will include organizing, visualizing, and applying data to report findings on the relationship between different variables and the graduates' salary.

An analysis of the information in this dataset is important in providing a quantitative insight on what characteristics are most important to engineering employers. The strength of the correlation between each characteristic in the dataset with the output, graduate engineering salary, will demonstrate to someone who is interested in the job market for engineers what traits make someone the most competitive in the industry.

The dataset used in this project was cleaned and modified to make it better suited for analysis. The three columns, 'ID', 'CollegeID', and 'CollegeCityID' were dropped from the dataset as they were not applicable to the purposes of this project. The value '-1' was used in the columns for test scores to represent students who did not take that test. The '-1' values in the dataset were converted to 'NaN' values to allow for comparison of the quantitative variables. The 'Salary' variable was renamed to 'Salary (Thousands of INR)', and the values in the column were divided by 1000 to reduce the number of zeroes and clarify that the currency is in Indian rupees. Finally, the variable 'nueroticism' was renamed to 'neuroticism' to fix a spelling error in the dataset. Each variable label will be described as they are mentioned, and a list of the descriptions of all of the variables are available in the link to the dataset in the bibliography.

It is important to find the correlation between the variables in the dataset to determine which characteristics are associated with others, and which have the strongest influence on engineer graduate salary. On the next page is a heatmap of the 24 quantitative variables in the dataset.

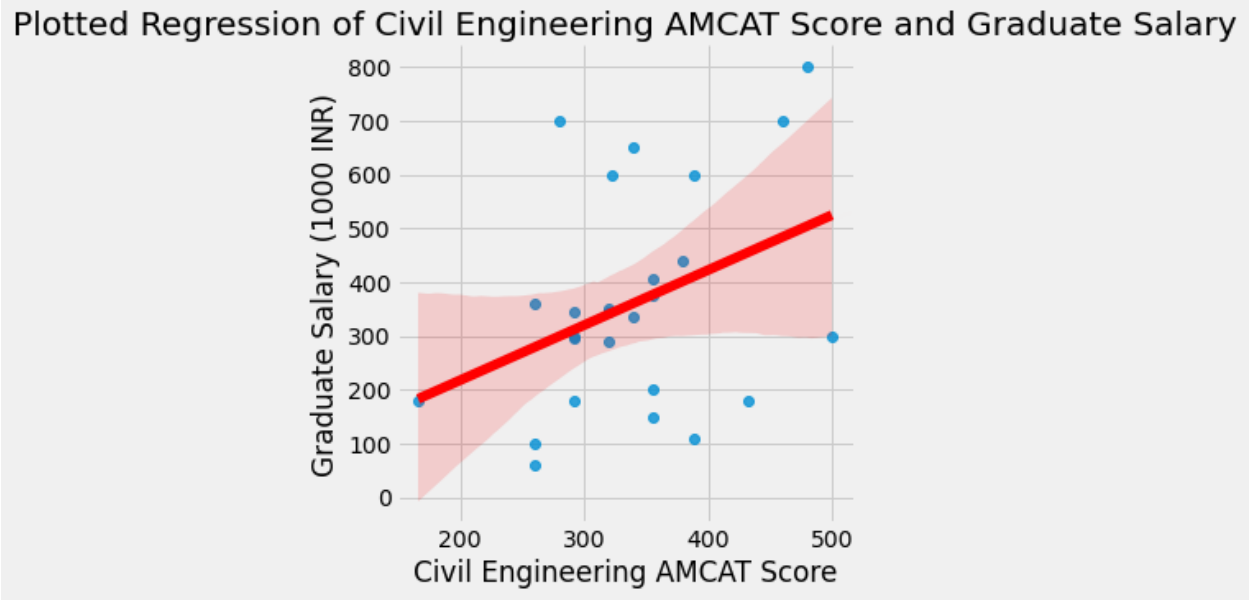


The cells in the heatmap represent the correlation coefficients of the associated row and column. The cell values range from -1.00, represented by a light green color and indicating a perfect negative linear correlation, to 1.00, represented by a dark blue color and indicating a perfect positive linear correlation. A cell value of 0 represents no linear correlation. The white squares represent non-applicable values. An example of this is that students who took the civil engineering exam (‘CivilEngg’) did not take the electrical engineering exam (‘ElectricalEngg’).

There are two significant takeaways from the heatmap. The dataset indicates that students’ performances in their exams, represented by the rows and columns from ‘English’ to ‘CivilEngg’, are positively correlated with each other. This means that students that performed well in one exam are more likely to have performed well in the other exams, and the students who performed poorly in one exam are more likely to have performed poorly in the others. The correlation coefficient values of the cells from the ‘English’ to ‘CivilEngg’ rows and columns, excluding the ‘1’ and ‘NaN’ values, have a median value of 0.273, representing a light to moderate positive correlation.

The second significant takeaway is that the metrics used in the dataset share a broad influence on the graduate salary, meaning that no one variable has an outstanding effect on an engineering graduate's salary. The median correlation coefficient of the 23 non-salary quantitative metrics with the graduate salary is 0.132, representing a weak positive correlation. The variable with the strongest correlation with graduate salary is performance on the civil engineering exam ('CivilEngg'), with a correlation coefficient of 0.37.

With performance on the civil engineering exam having the largest influence on the output of interest, graduate salary for engineers, it is appropriate to further explore the relationship between the two variables. Below is a visualization of the regression line between students' scores on the Civil Engineering AMCAT, a standardized exam, and their respective graduate salary in thousands of Indian rupees.



The regression line shows a notable positive relationship between performance on the Civil Engineering AMCAT and graduate salary. This finding is significant in illustrating the quantitative variable with the largest influence on graduate salary for engineers.

To evaluate the performance of the regression model, cross-validation is used, dividing the dataset into multiple folds, and using each fold as a validation set while the rest of the data is used for training. The regression model is trained on different random splits of the training and test sets, and the test set error of these iterations are then averaged. The cross-validated test set error of the regression line between the Civil Engineering AMCAT scores and their respective graduate salaries (in thousands of Rupees) when using 24 folds is 3.90*10^4. The test set error indicates that there is some margin of error on the strength of the positive correlation between the AMCAT scores and graduate salaries.

Qualitative variables such as gender may also influence the graduate salary of engineering students. Below is a boxplot of the salary distributions, differentiated by gender.

The box plots visualize the distributions of the graduate salaries. Half of the male graduate salaries lie between the range of 180K INR and 380K INR, with a median of 300K INR. Half of the female graduate salaries lie between the range of 190K INR and 350K INR, with a median of 300K INR. The distributions of the salaries for male and female engineering graduates are similar, although male engineering graduates do have a higher average salary, with an average male graduate salary of 309.9K INR compared to the average female graduate salary of 290.4K INR.

The exploratory data analysis in this project on the relationship between various quantitative and qualitative variables to salary for engineering graduates is limited to this dataset, but there are additional analyses from other sources that come to interesting conclusions. In an article from Caleb M. Bowyer on the factors that affect the salary of engineers in the United States based on an analysis of salary data provided by the US Bureau of Labor Statistics, he found that two significant variables that influence engineer salaries are their specialization and degree of education. Other findings in the analysis concluded that geographic location and the size of the company will also influence an engineer's salary.

This project was an analysis on how various characteristics influence the salary for Indian engineering graduates. The analysis found that the variables in the dataset share a broad influence over graduate salary, and that no one variable has an outstanding influence that dwarfed the importance of the other variables. It was found that male engineering graduates have a slightly higher average salary than female engineering graduates. The most significant takeaway from the analysis of the dataset is that the market value of an engineering graduate is

based on a multidisciplinary evaluation, and that the ideal engineering candidate will show proficiency in multiple areas of consideration, as opposed to one significant characteristic.

Bibliography:

Engineering Graduate Salary Prediction, version 1 of 1 – accessed on February 11, 2023
Author: Manish KC
https://www.kaggle.com/datasets/manishkc06/engineering-graduate-salary-prediction

Caleb M. Bowyer, Ph.D. Candidate. “Factors Affecting Salaries of Engineers in America and How to Earn More.” *Medium*, Level Up Coding, 15 Sept. 2022, https://levelup.gitconnected.com/factors-affecting-salaries-of-engineers-in-america-and-how-to-earn-more-52e47b0dff33.