# Employability Outcomes of Engineering Graduates in India


AUGUST 22

**Coursera Applied Data Science Capstone Project**
**Authored by: Kandasamy Ramanujam**

# Problem Statement

According to All India Council for Technical Education, more than 1.8 Million students are enrolled in engineering courses across 10,000+ engineering institutions in India. Of these, only around 42% of them are placed in jobs at the time of completion of their engineering education. Additional information can be found at the AICTE website here.

The objective of this project is to understand the factors that influence the employability of engineers – specifically indicated by the initial salary offered to the engineering graduates. Some of the factors that influence this could be academic performance – both in engineering institution and prior to that in the schools, demographic factors such as gender, location of the college, proficiency in English, other factors such as aptitude for quantitative skills.

# Description of Data

Aspiring Minds Employability Outcomes 2015 (AMEO 2015) is a unique data set that contains engineering graduates' employability outcomes (salary, job title, city of employment) along with data on assessment scores and other demographic data. This includes the following:

- Scores from school final exams – $10^{th}$ and $12^{th}$ standard
- Scores from Engineering course
- Engineering branch
- Tier of college and the city in which the college is located
- Demographic data such as gender, state
- Scores on English, logical ability, quantitative aptitude, and Computer Programming from standardized assessment test conducted by Aspiring Minds

The training data set has 3998 entries with 39 columns. There is a test data set with 1500 entries. However, the salary information is not provided here. So, this cannot be effectively used to compare the predictions from the models built using training data. The test data set is not used in this project.

The data can be downloaded from the Aspiring Minds website. Location given below.
http://research.aspiringminds.com/resources/#ameo

# Methodology for Analysis

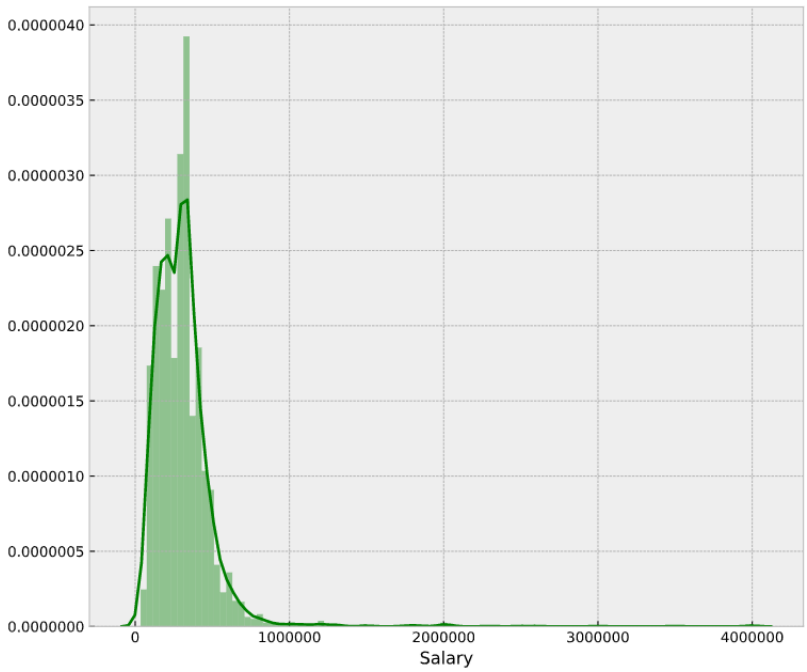
The proposed approach is as below.

- Exploratory Data Analysis to understand the data: Based on the initial analysis done, the data has non-null values on all rows and columns.
- Identify types of modeling to be used: As the outcome (salary) is a continuous variable, regression analysis is an obvious candidate. However, multi-class classification can be considered if the salary could be grouped into a small number of buckets and the problems is reframed as predicting the salary range.
- Build and test models: A subset of training data to be used for testing the models.

# Exploratory Data Analysis

Highlights of exploratory data analysis is given below:

- Shape of data: 3998 rows and 39 columns – all non-null values
- Salary ranges from 0 to 4 million (Indian Rupees). However, most values are less than 1 million.



- Though all the data is non-null, data distribution charts of numeric values indicates that several columns have zeros in many rows. Based on this only a subset of columns (10percentage, 12percentage, Domain, English, Logical, Quant) are chosen for regression analysis. Some of the other columns with complete data ( conscientiousness, extraversion, neuroticism, openness_to_experience) are not chosen because of lack of detailed description of the meaning of this data.