



INNOVATION. AUTOMATION. ANALYTICS

PROJECT ON
Exploratory Data Analysis Project
on
AMCAT AMEO DATA

About me

I have completed my Bachelors of Technology (B-Tech) in Electrical and Electronics Engineer in 2011.

Over the span of eight years in the staffing industry, I honed my skills as a dedicated sourcer, specializing in identifying and attracting top talent for various positions. My journey in the staffing realm equipped me with a profound understanding of the intricacies involved in talent acquisition, allowing me to thrive in a dynamic and people-centric environment.

As my career evolved, I recognized the transformative potential of data in enhancing decision-making processes. Motivated by a desire to delve deeper into the analytical aspects, I embarked on a career transition into data science. Embracing data science has empowered me to harness the power of data for strategic decision-making. I have acquired proficiency in data analysis, machine learning, and statistical modeling, enabling me to extract meaningful insights from complex datasets.

My career transition reflects my commitment to staying at the forefront of industry trends and my dedication to utilizing technology and data to drive positive outcomes. I am excited about to enter into the data science world.

LinkedIn: <https://www.linkedin.com/in/sunkranthi90/>

Github: <https://github.com/sunkarikranthi>



Introduction:

The dataset was released by Aspiring Minds from the Aspiring Mind Employment Outcome 2015 (AMEO). The study is primarily limited only to students with engineering disciplines. The dataset contains the employment outcomes of engineering graduates as dependent variables (Salary, Job Titles, and Job Locations) along with the standardized scores from three different areas – cognitive skills, technical skills and personality skills. The dataset also contains demographic features. The dataset contains around 40 independent variables and 4000 data points. The independent variables are both continuous and categorical in nature. The dataset contains a unique identifier for each candidate. Below mentioned table contains the details for the original dataset.

Objective:

- ❖ The main goal is to perform EDA on the dataset and find the important relations between the Variables and Visualize the given data.
- ❖ Target Variable is Salary.

Exploratory Data Analysis

Univariate Analysis

Numerical Variables – Page 5-9

Categorical Data – Page 10-16

Bivariate Analysis

Numerical vs Numerical Variables – Page 17-19

Categorical vs Numerical – Page 20-21

Categorical vs Categorical Variables – Page 22-23

Research Question

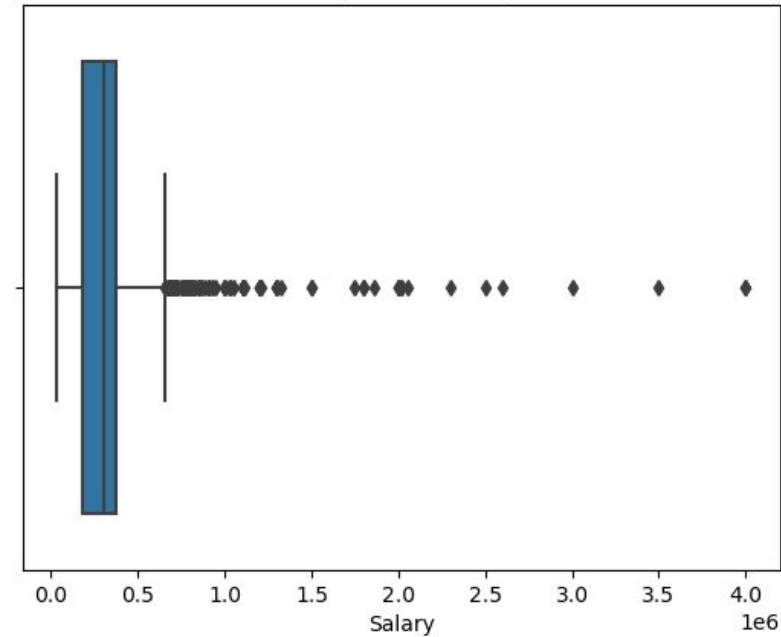
Page 24

Conclusion

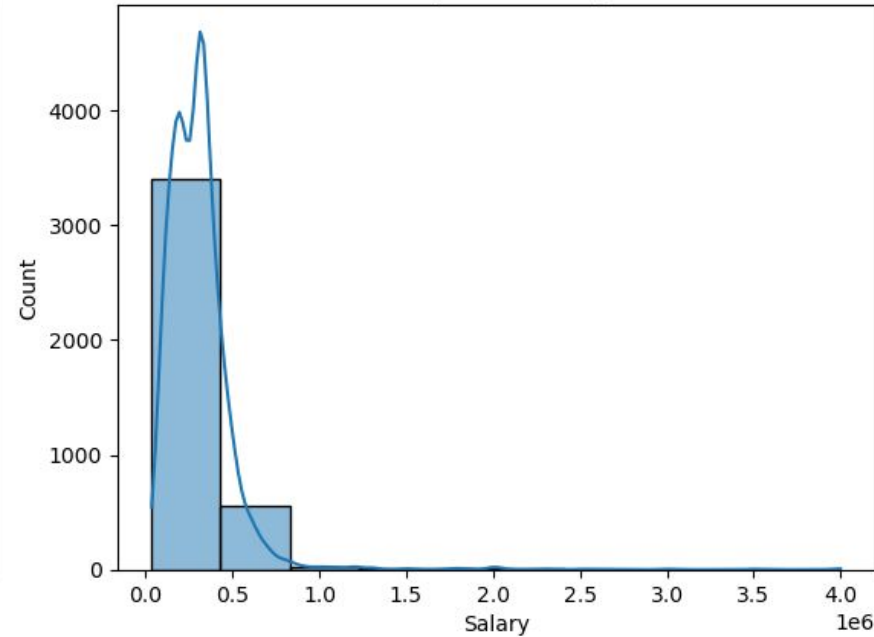
Page 25

Numerical Output Variable

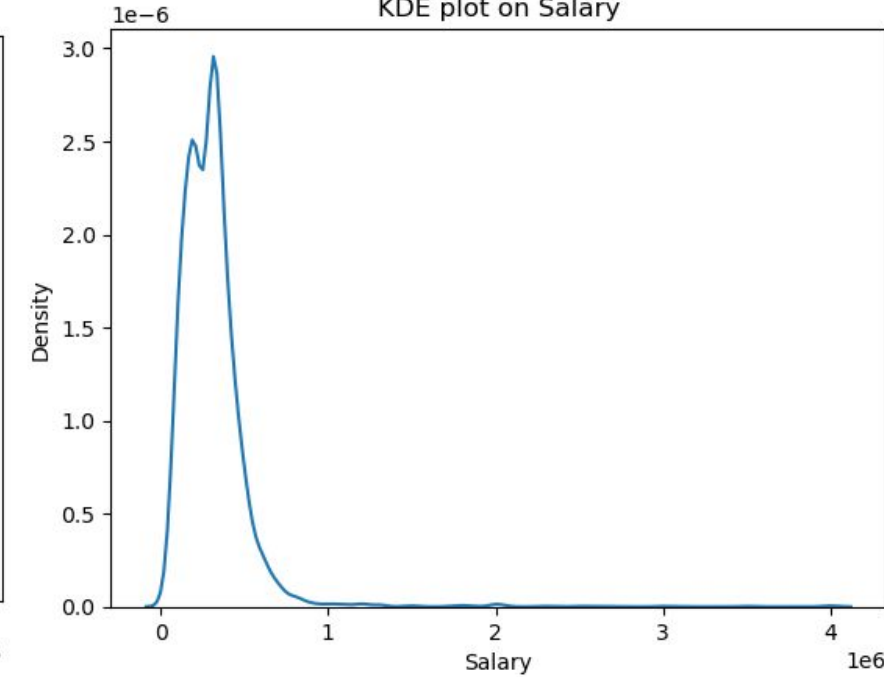
Box plot on Salary



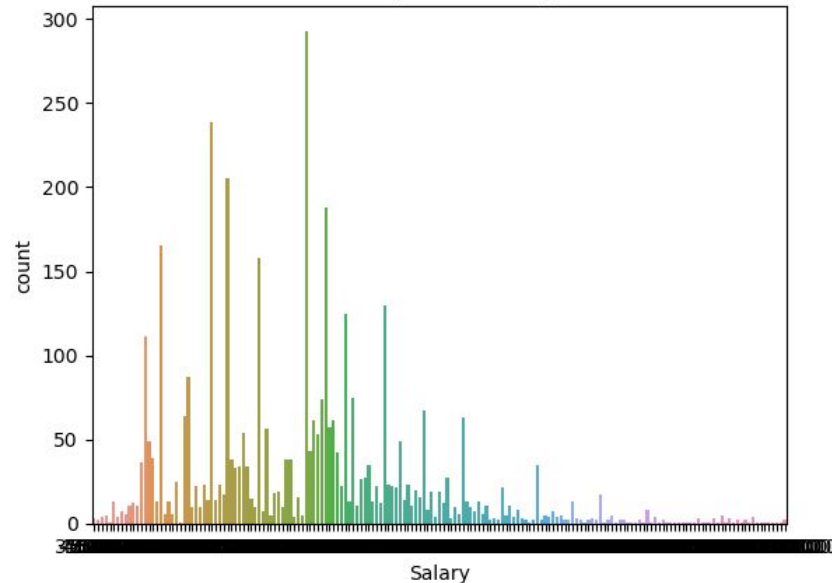
Hist plot on Salary



KDE plot on Salary



Count plot on Salary

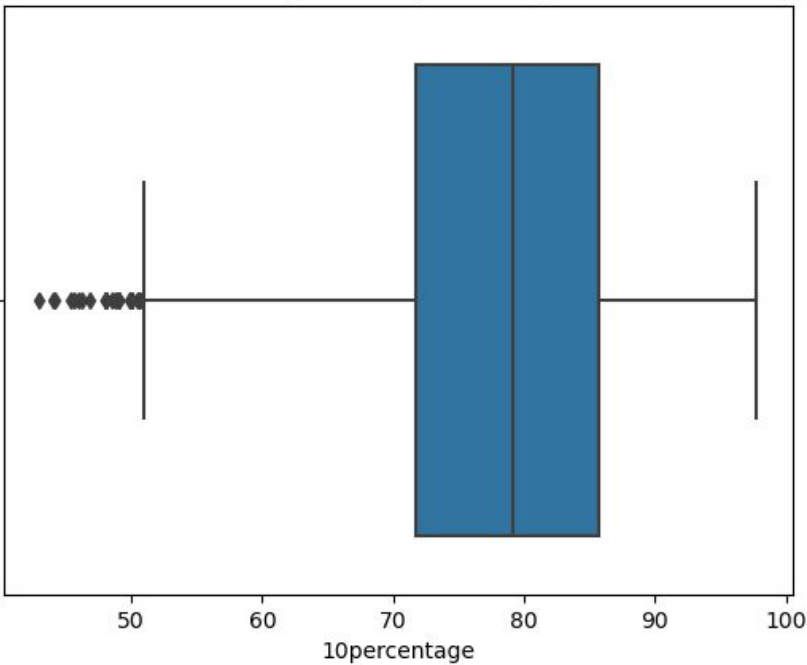


Observation

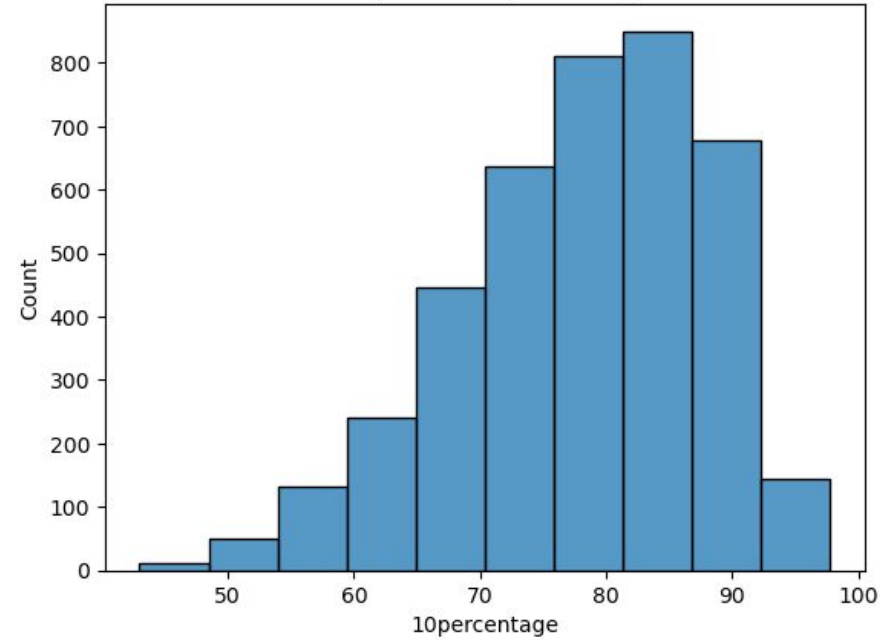
- **Boxplot** : We can clearly see that there are more number of outliers in the Salary Column, to make the data more clear we can remove the data that is > 2.0
- **Histplot** : The data is right skewed
- **countplot** : Nearly 300 is the highest value in the Salary

10 percentage

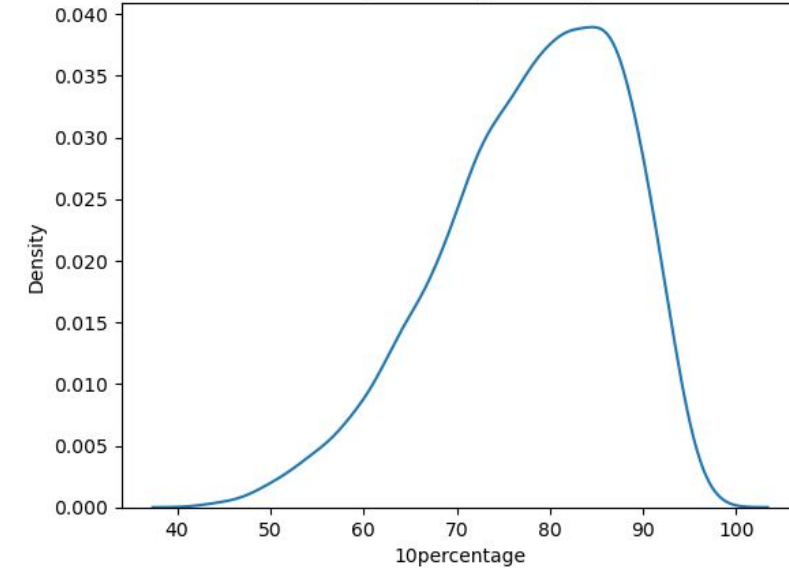
Box plot on 10percentage



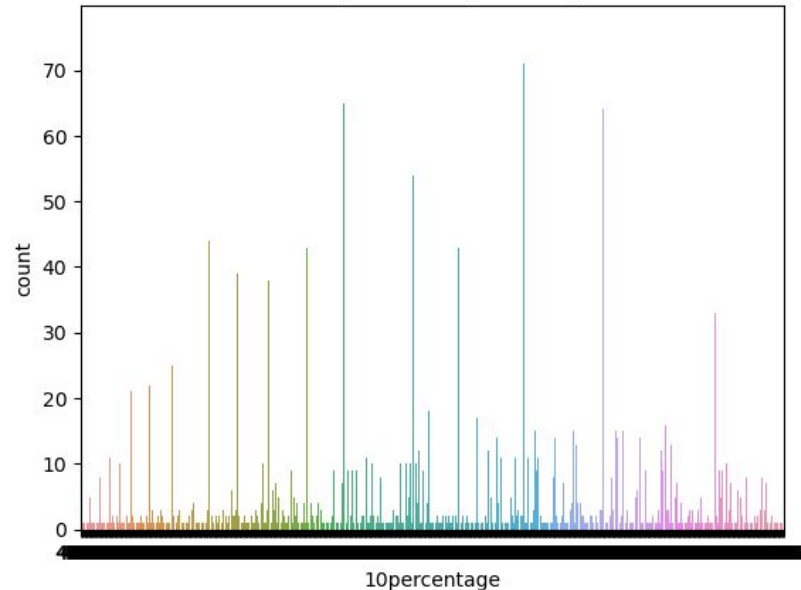
Hist plot on 10percentage



KDE plot on 10percentage



Count plot on 10percentage

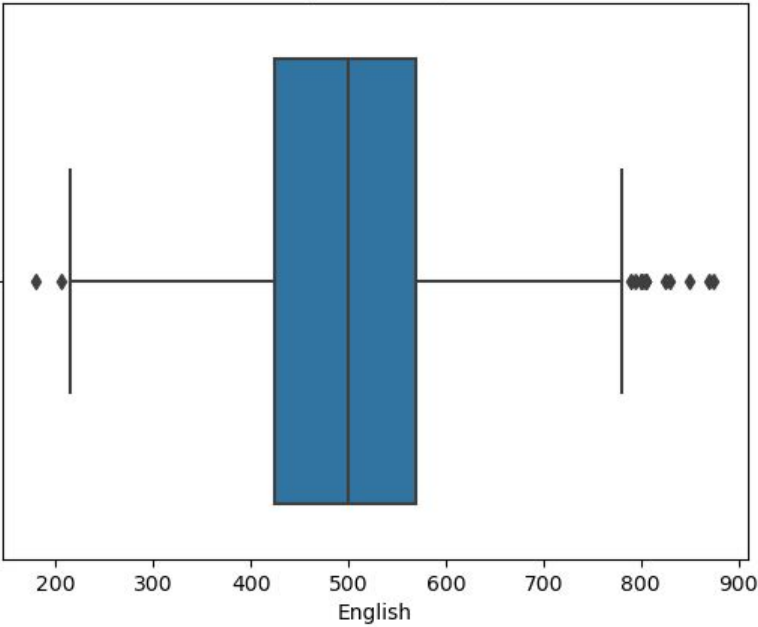


Observation

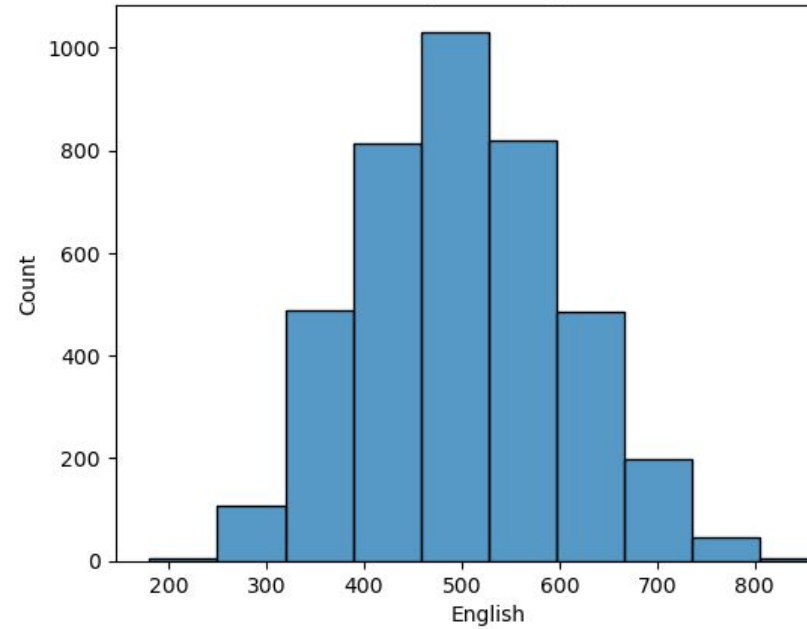
- **Boxplot** : We can say that average 10percentage is 80 and there are some outliers which are less than 50
- **Histplot** : The data is left skewed and from plot we can tell that avg lies between 80-90

English

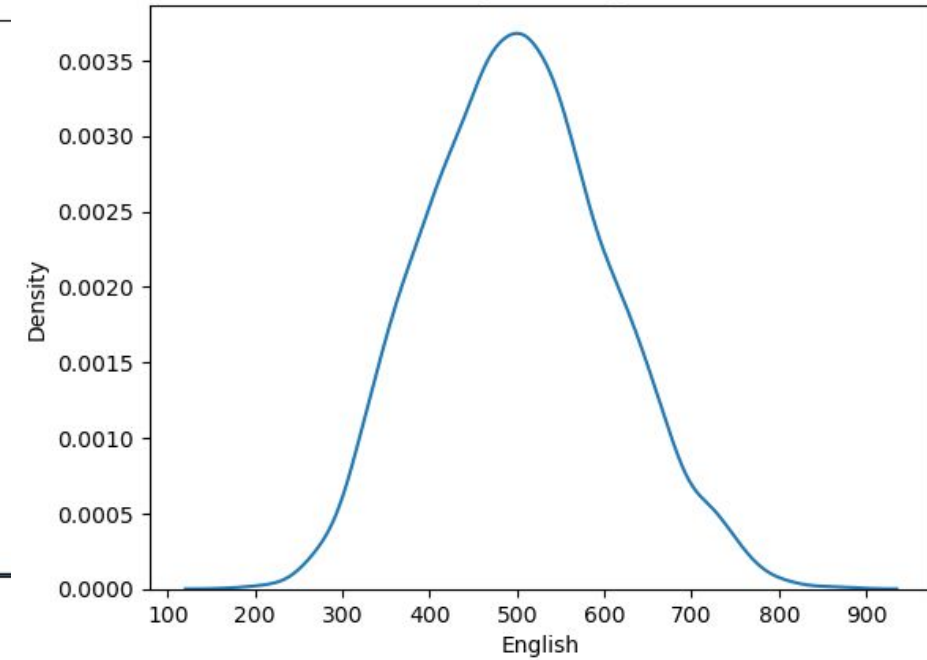
Box plot on English



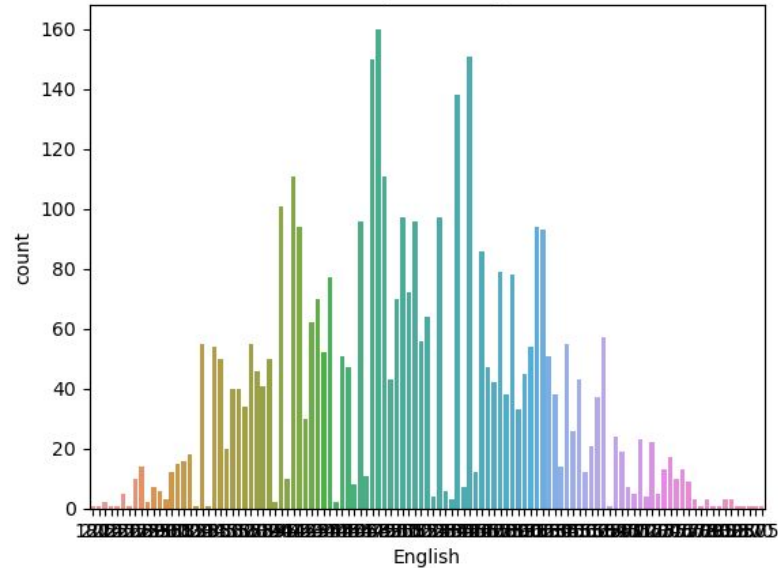
Hist plot on English



KDE plot on English



Count plot on English

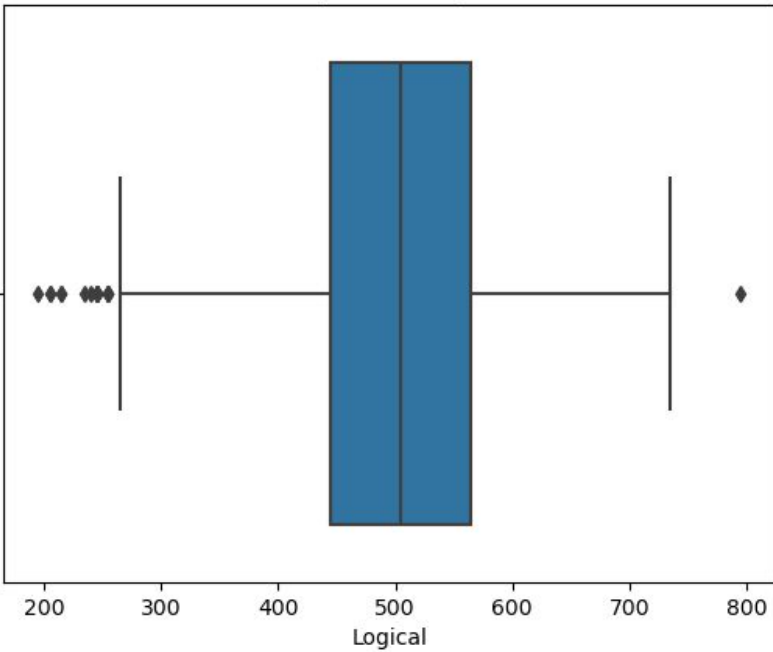


Observation

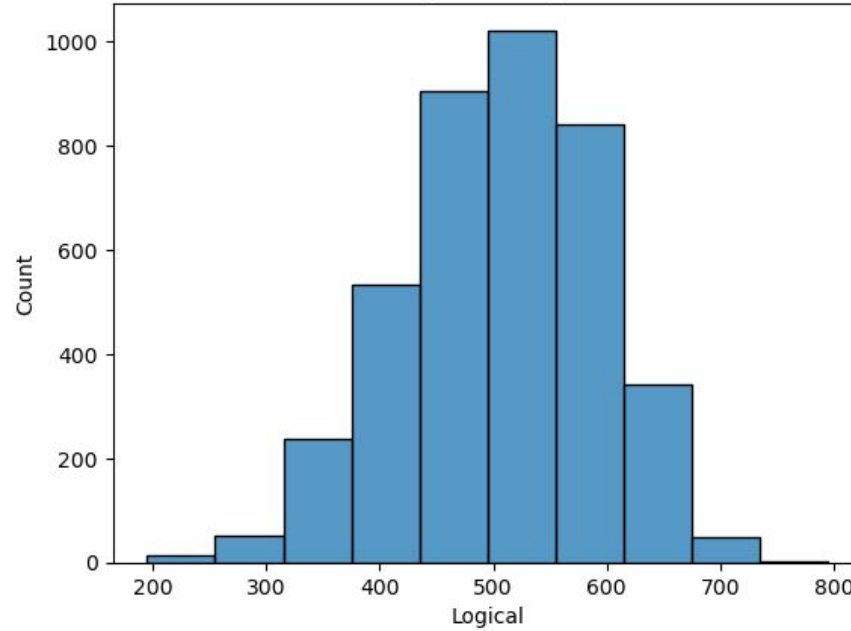
- **Boxplot** : The average value is at 500 and having more outliers which are greater than 800
- **Histplot** : This looks like a Normal Distribution

Logical

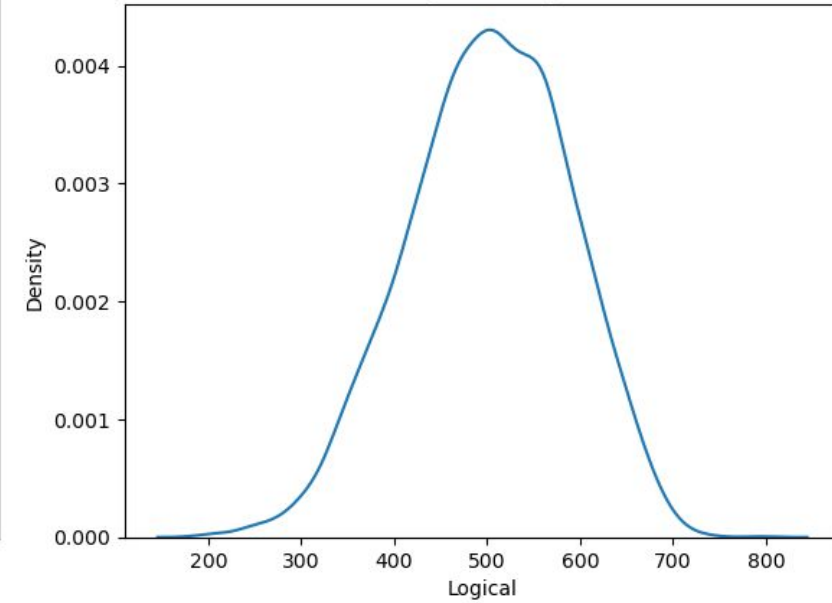
Box plot on Logical



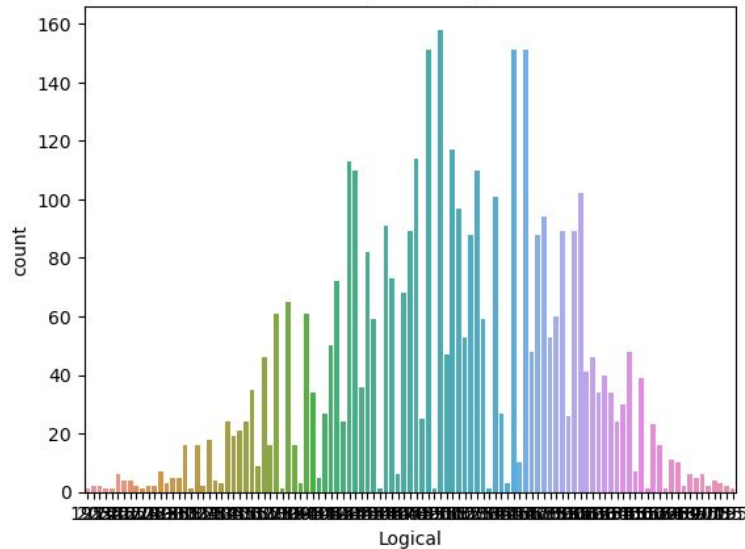
Hist plot on Logical



KDE plot on Logical



Count plot on Logical

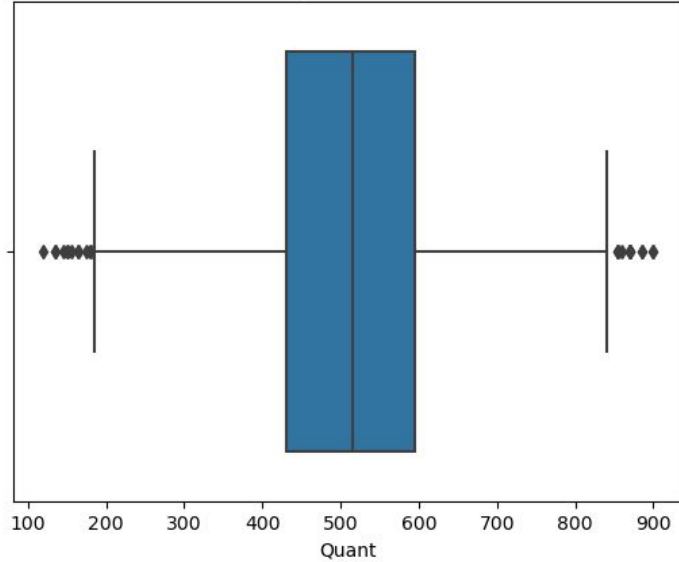


Observation

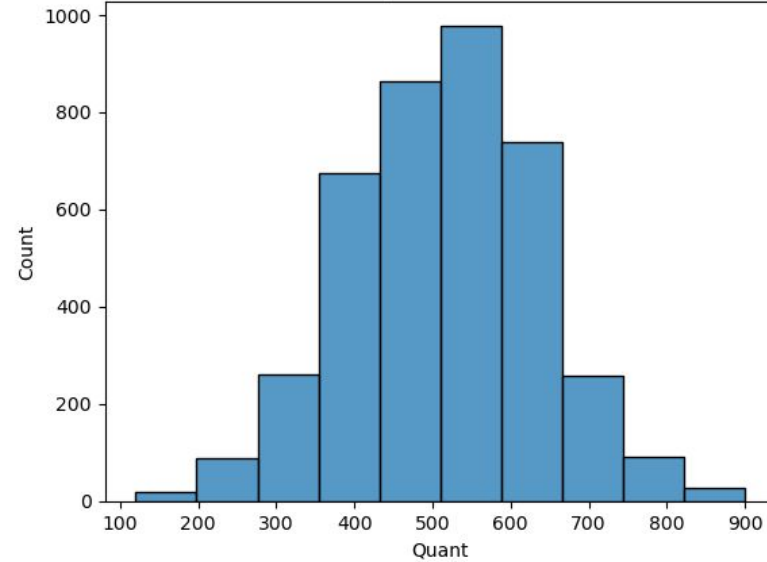
- **Boxplot** : The avg value is at 500 and there are more outliers which are greater than 800
- **Histplot** : This looks like a Normal Distribution

Quant

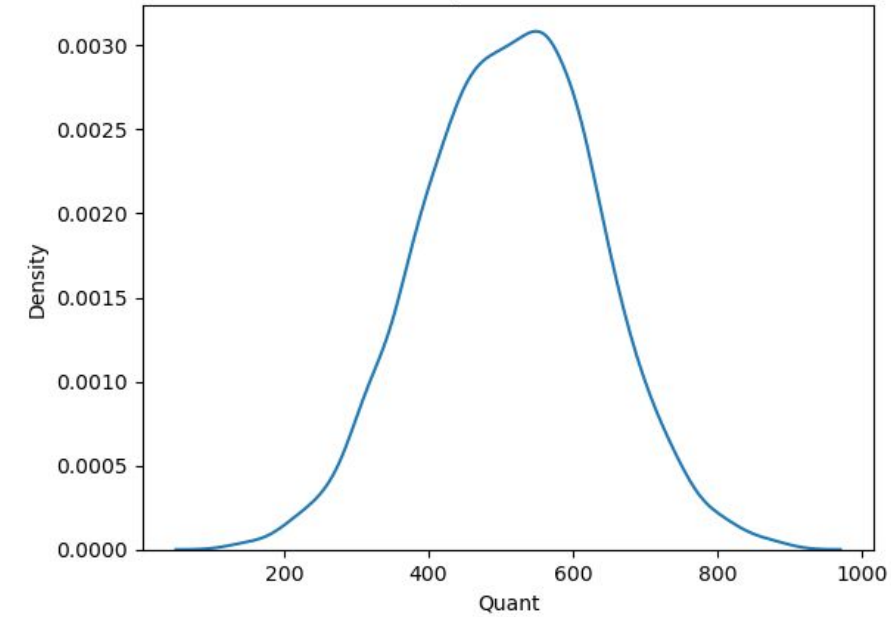
Box plot on Quant



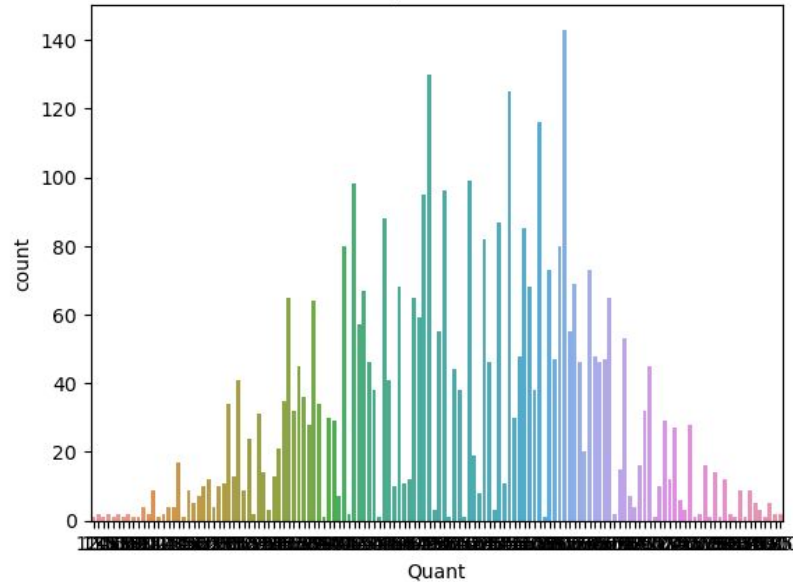
Hist plot on Quant



KDE plot on Quant



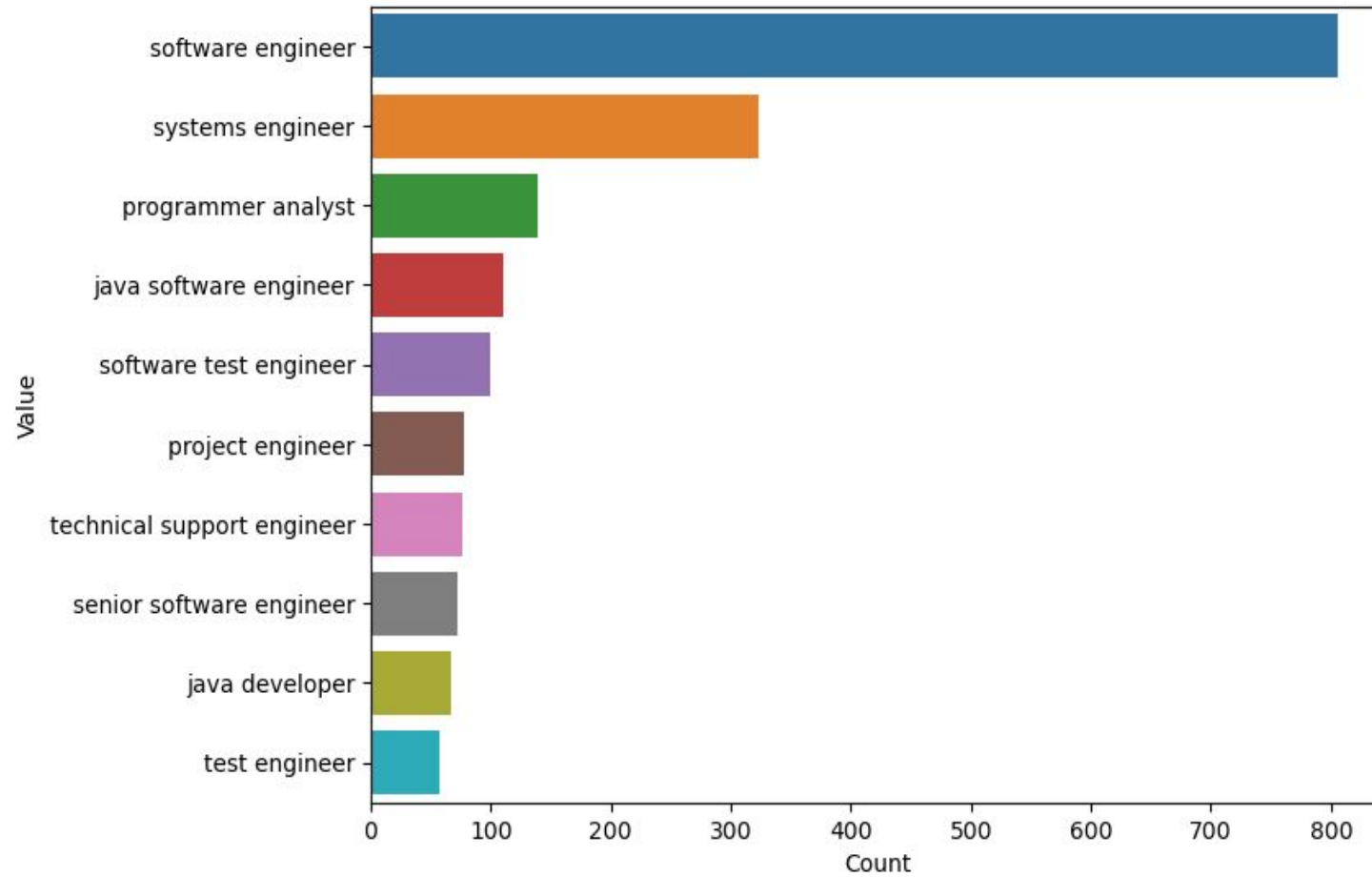
Count plot on Quant



Observation

- **Boxplot** : The avg value is at 500
- **Histplot** : This looks like a Normal Distribution

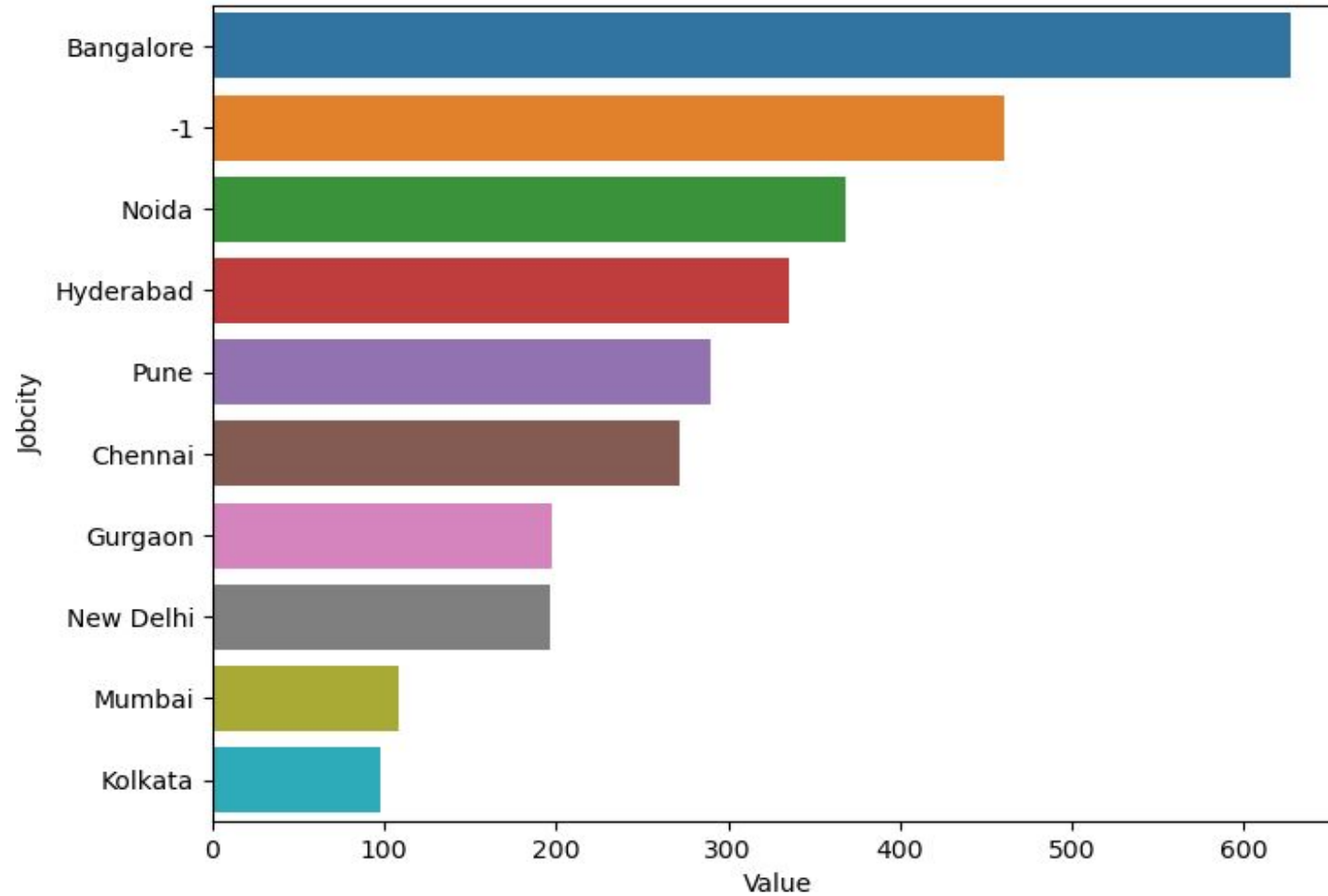
Designation



Observation

---- From the above bar plot we can observe that most of the members are Software Engineers

JobCity

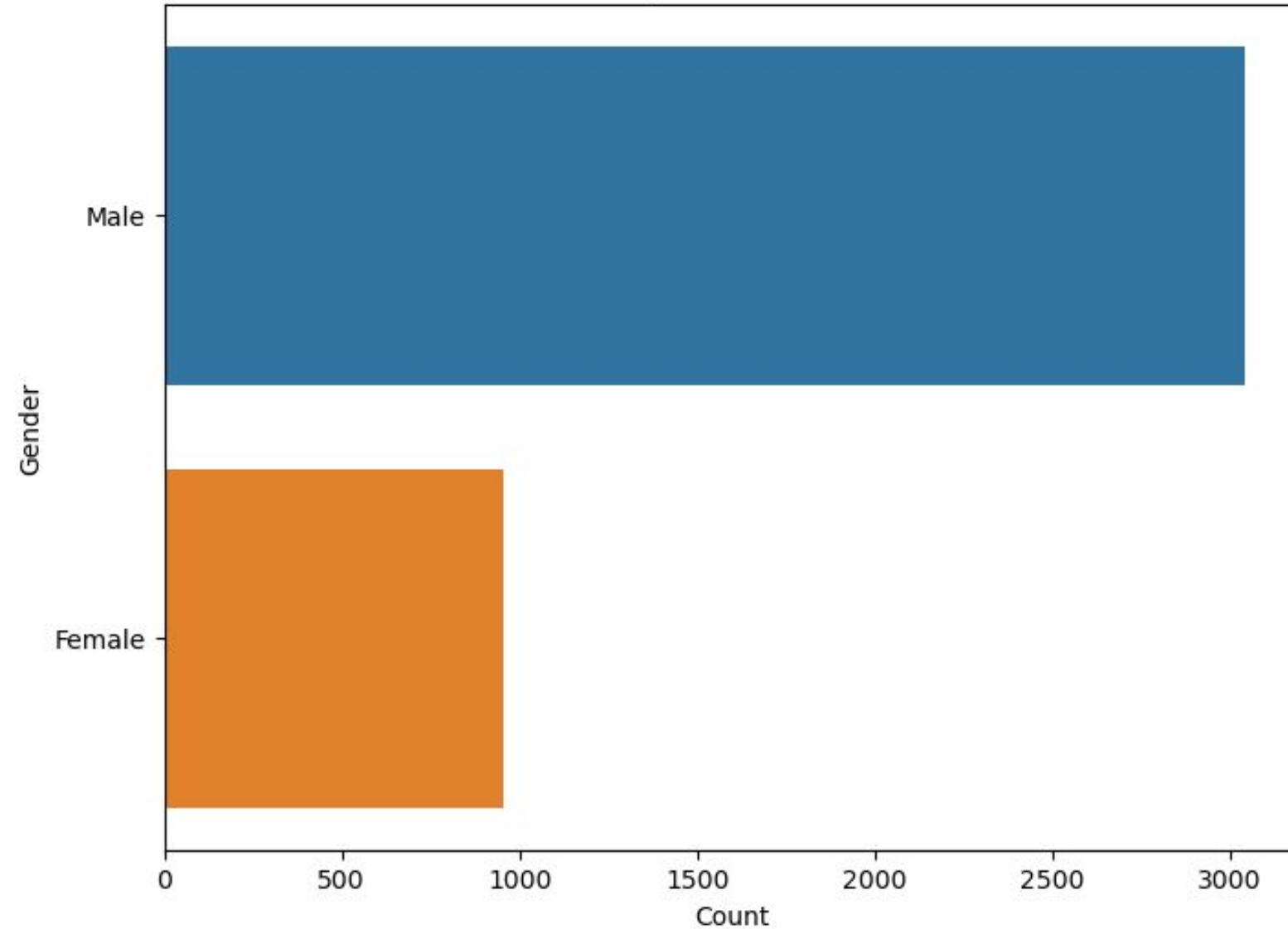


Observation

---- From the above bar plot we can tell that most of the members are working in Bangalore , Noida ,Hyderabad.

Gender

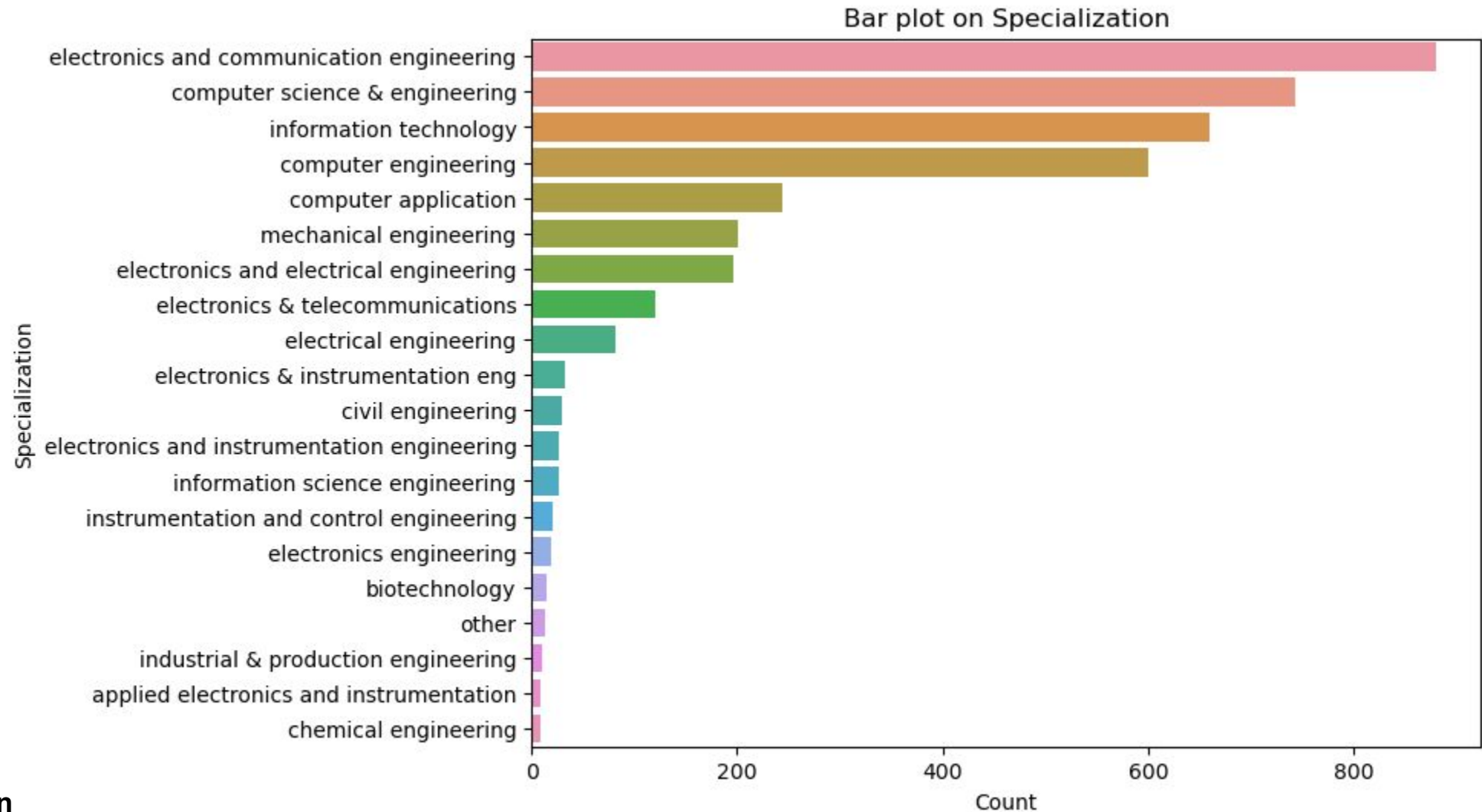
Bar plot on Gender



Observation

---- Male genders are more in number when compared to Female

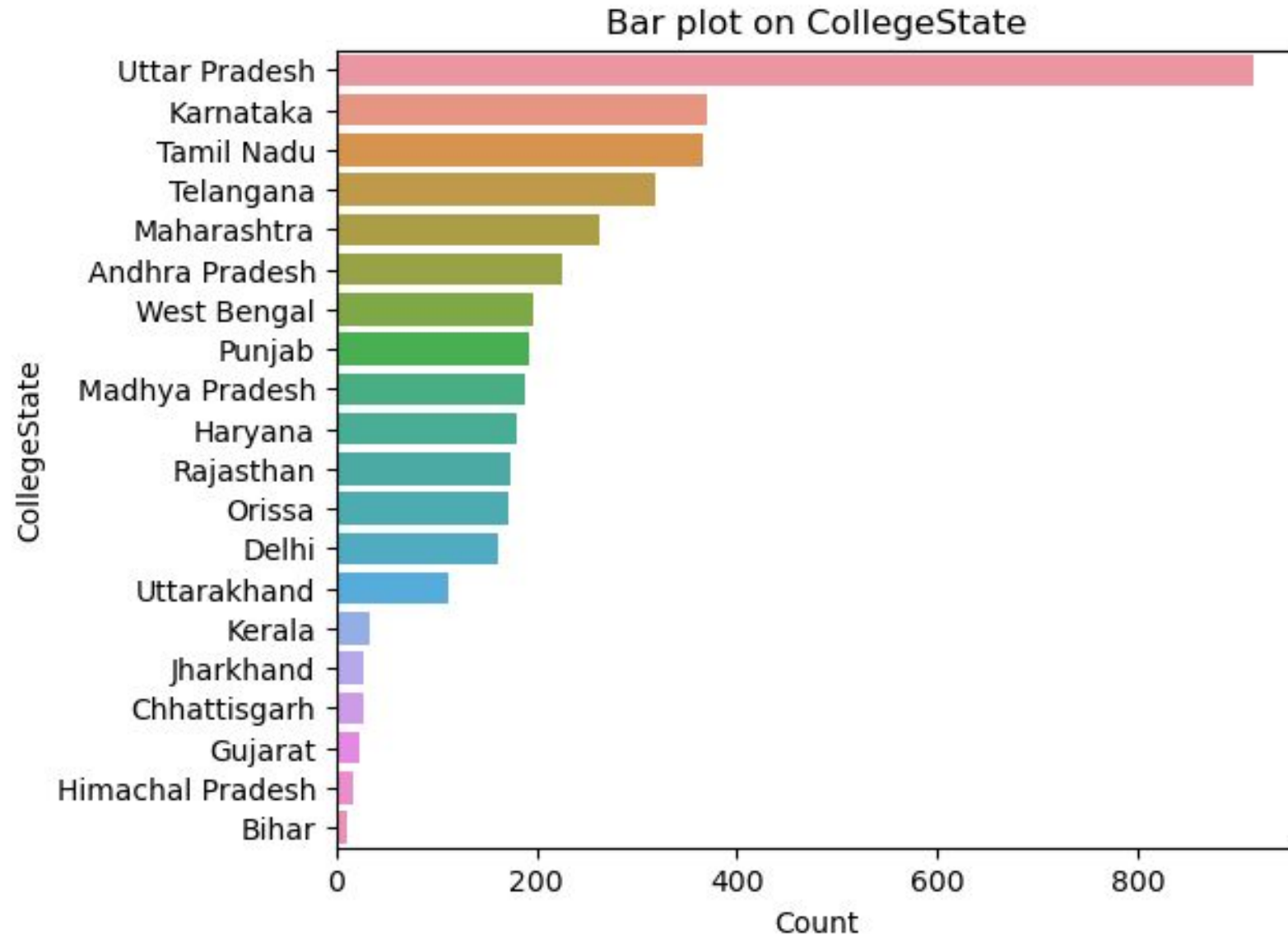
Specialization



Observation

---- From the above plot we can tell that most of them are from electronics and communication engineering (ece) , computer sciences

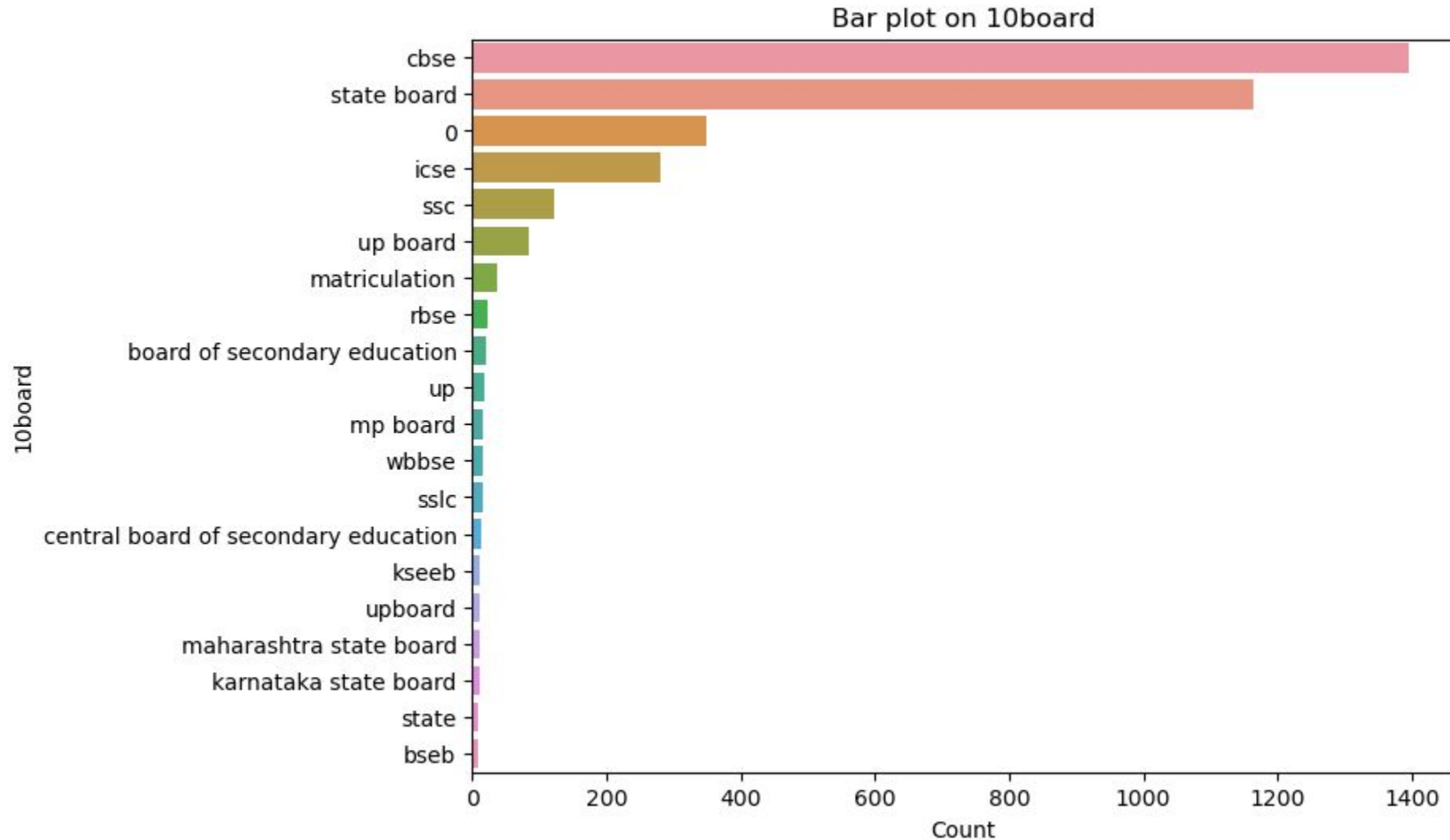
CollegeState



Observation

---- From the above plot we can say that there are most graduates from Uttar Pradesh (>800)

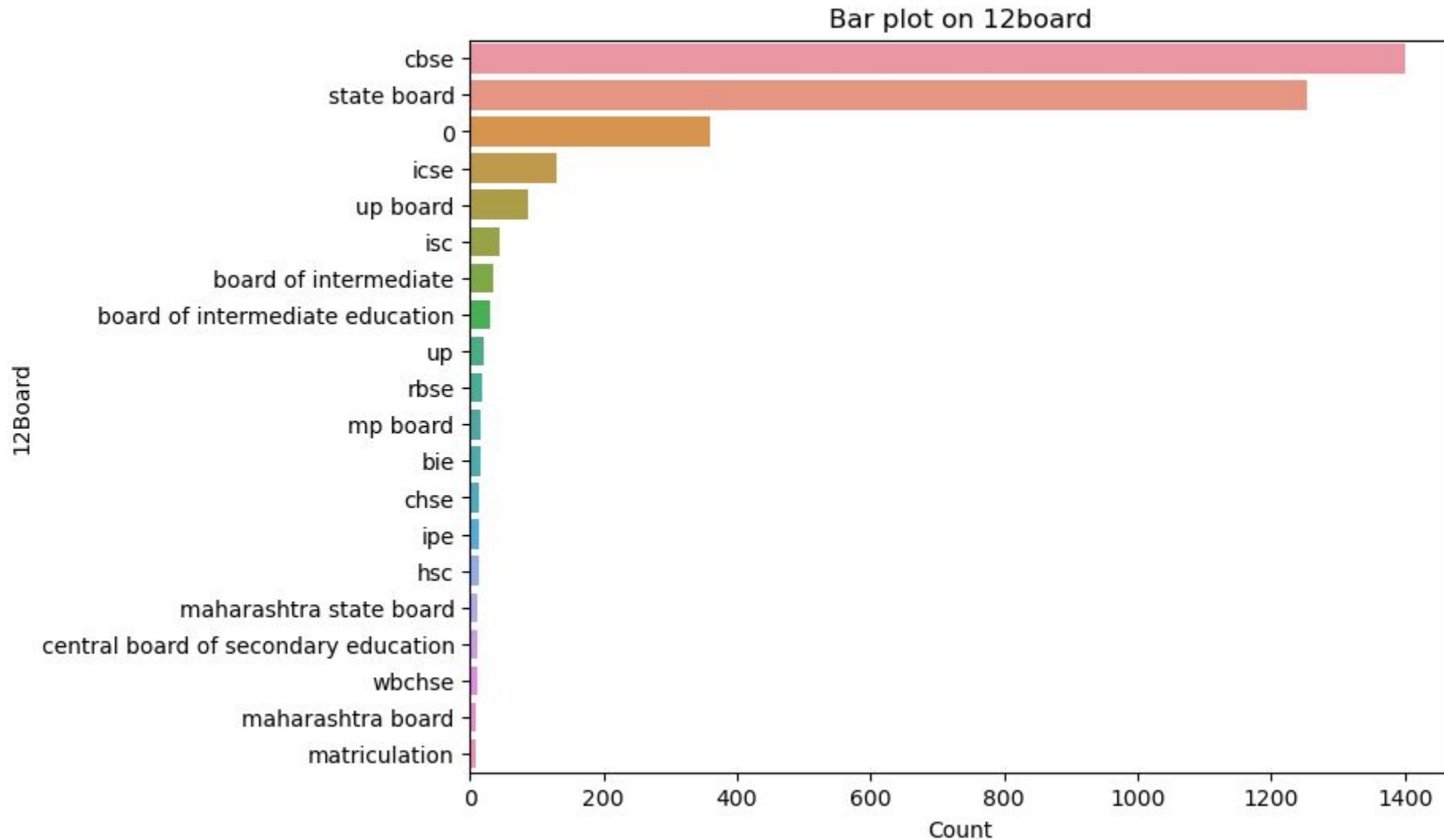
10board



Observation

---- From the above plot we can say that most of them have completed there 10th board under cbse and state board.

12board

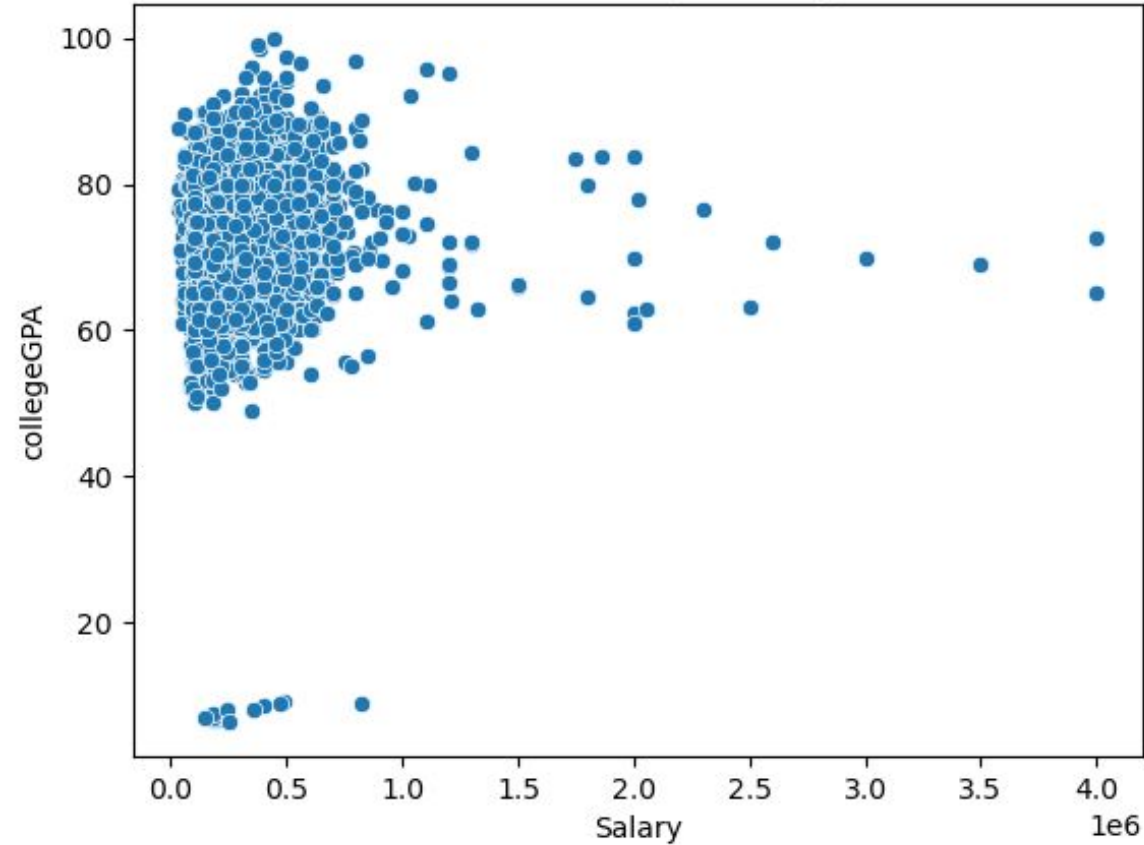


Observation

---- From the above plot we can say that most of them have completed there 10th board under cbse and state board

Salary vs collegeGPA

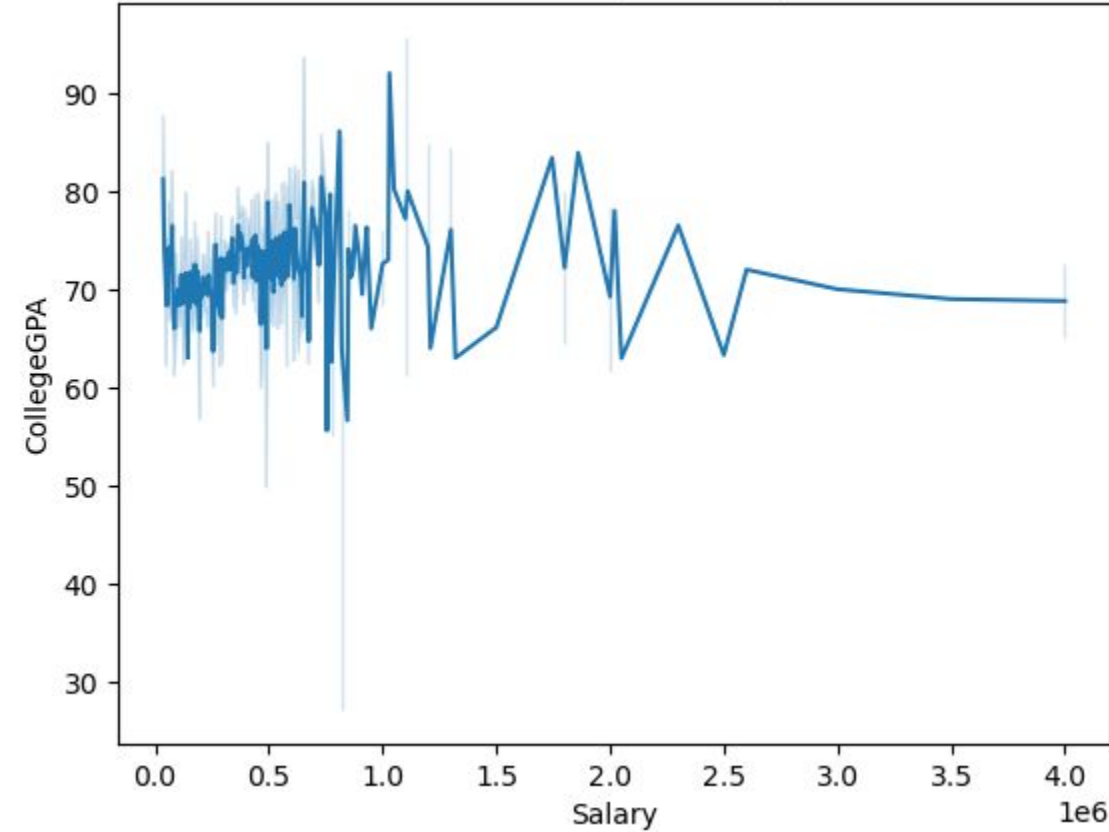
Scatter Plot - Salary vs collegeGPA



Observation

---- From the above scatter plot we can observe that college gpa of large number lies between 60 and 80 and most of the salary lies between 0 and 1

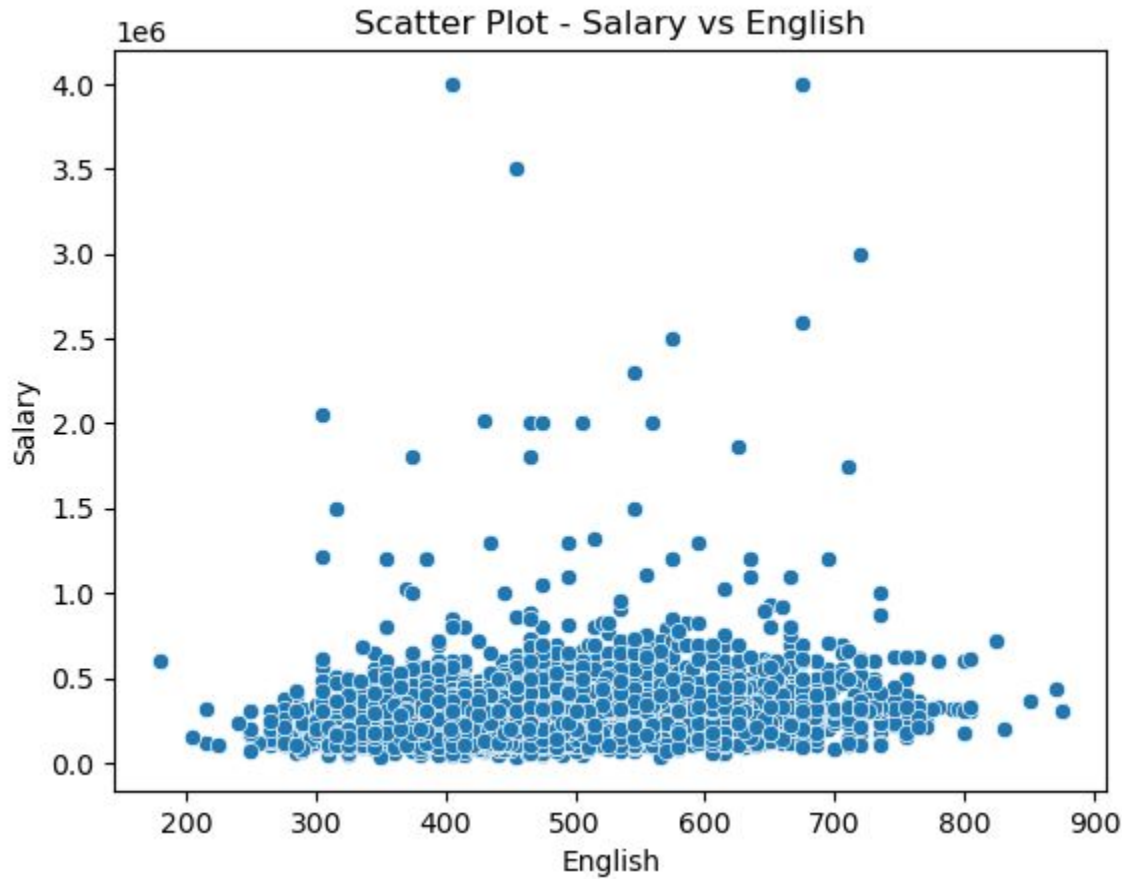
Line Plot - Salary vs collegeGPA



Observation

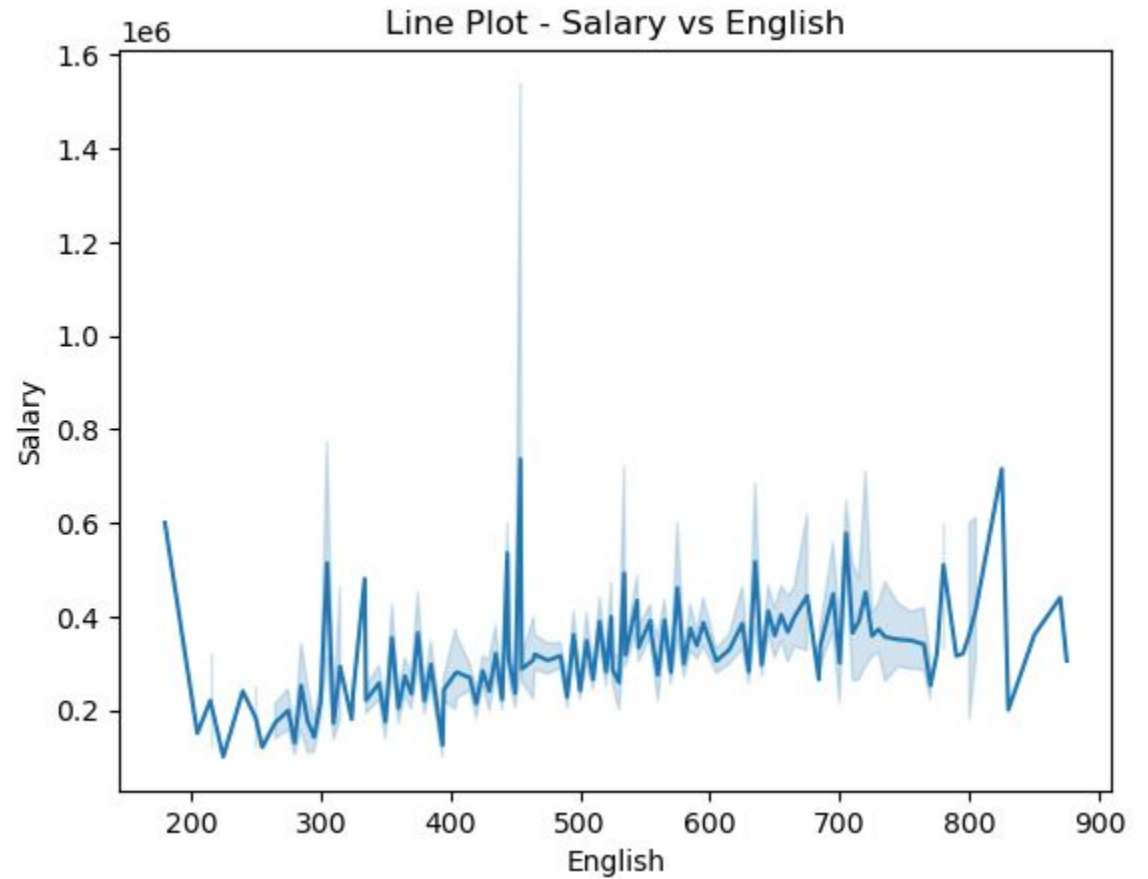
---- From the line plot we can observe that max gpa is 70 and highest salary is at 4.0

Salary vs English



Observation

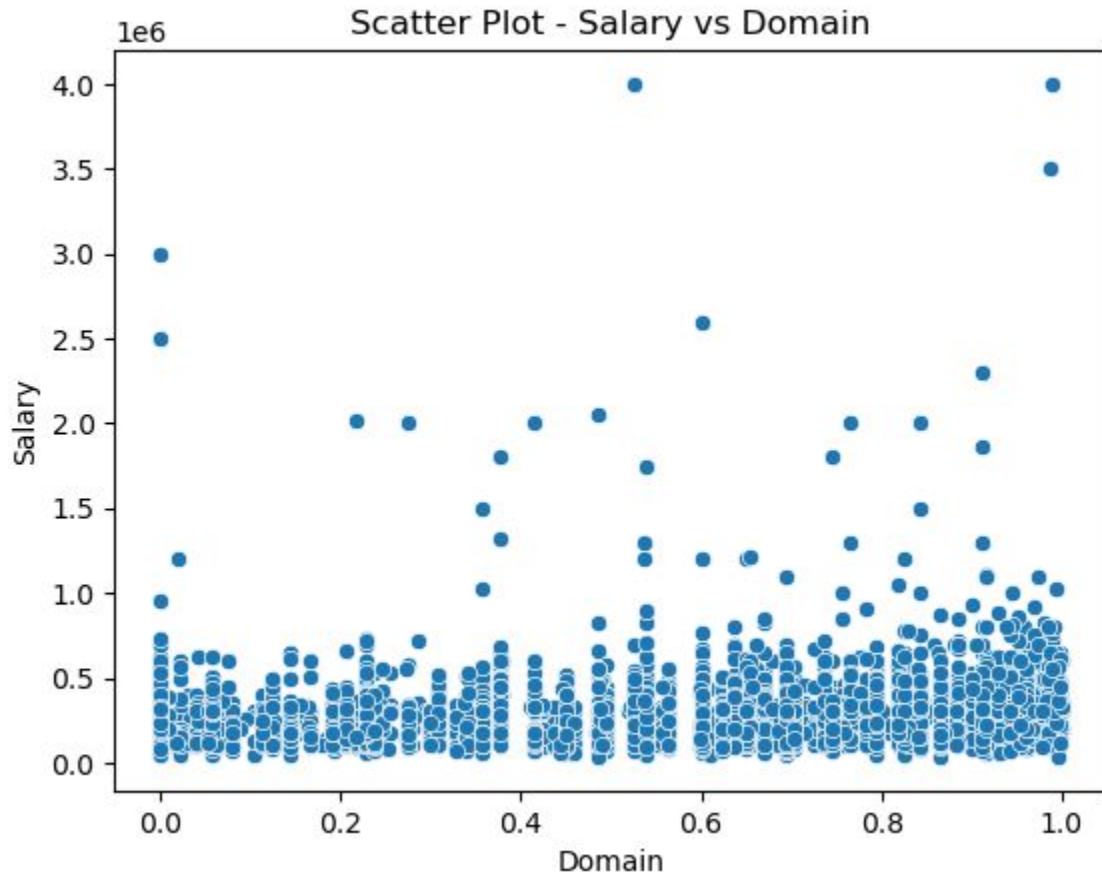
---- From the above plot we can observe that most salary range is in between 0 and 1 range



Observation

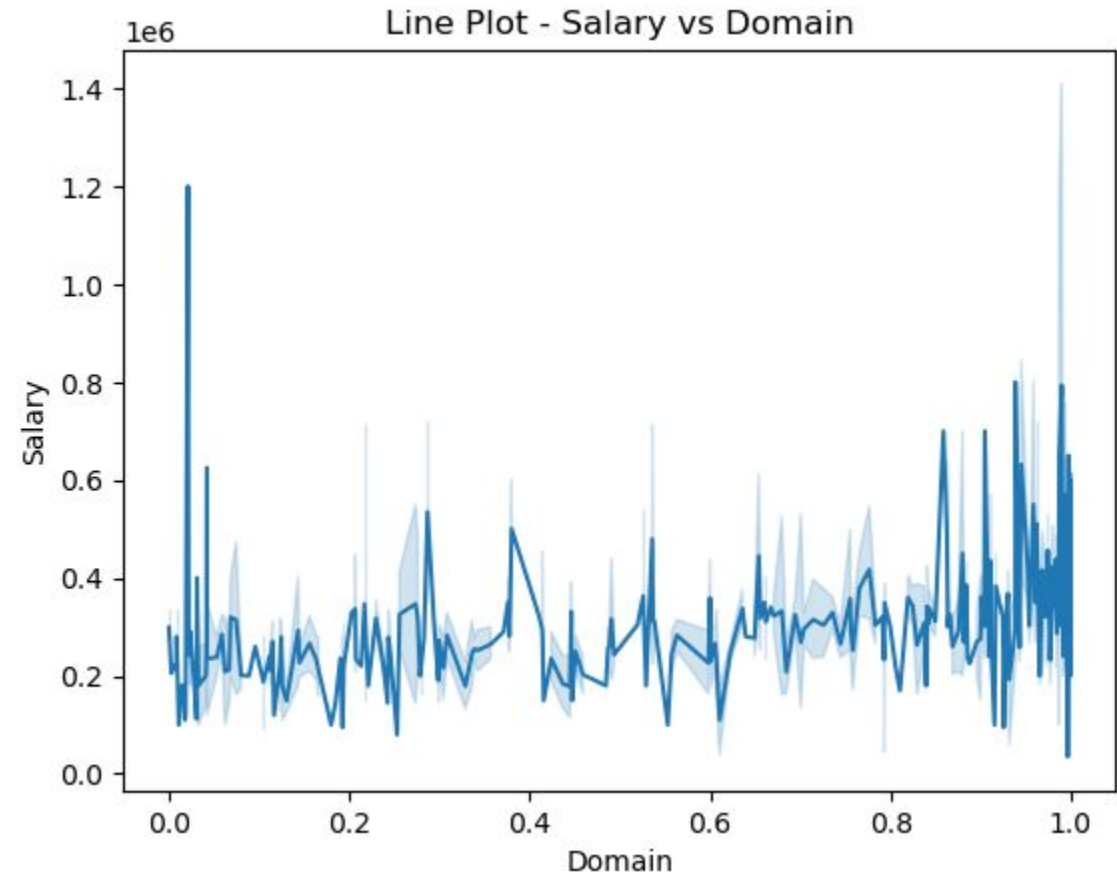
---- From the above Line Plot we can say that the people with least knowledge in english has salary in range of 0.2 to 0.6

Salary vs Domain



Observation

---- From the above scatter plot, we can say that there is no proper relation between Salary and Domain

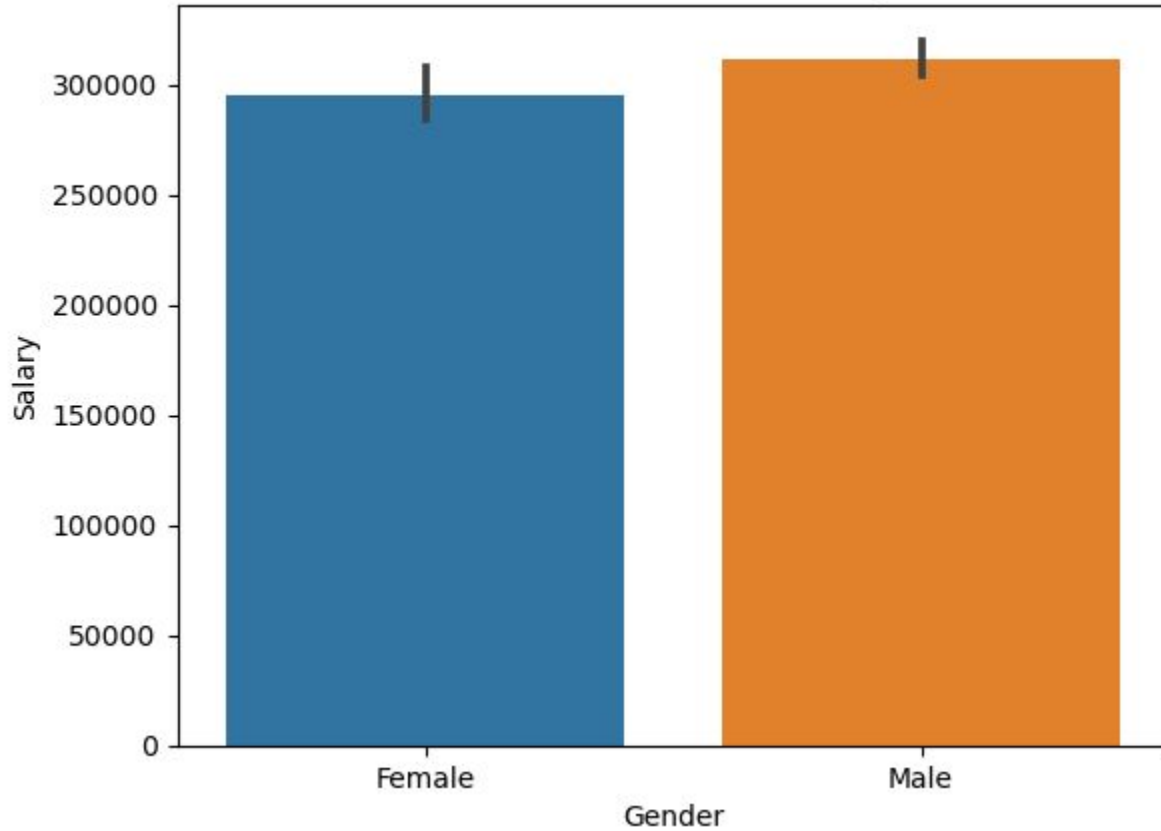


Observation

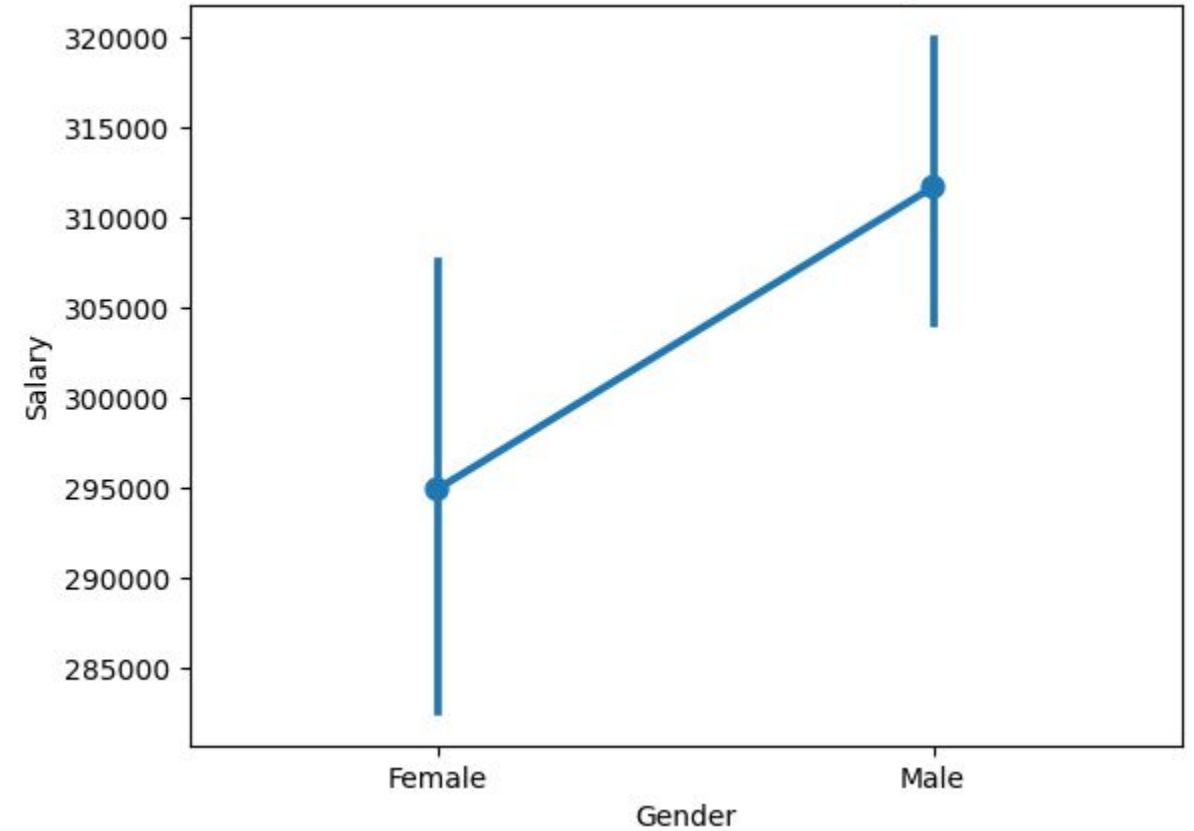
---- In domain we can observe more values in negative.
According to AMCAT -1 represents the absent status.

Gender vs Salary

Bar Plot - Gender vs Salary



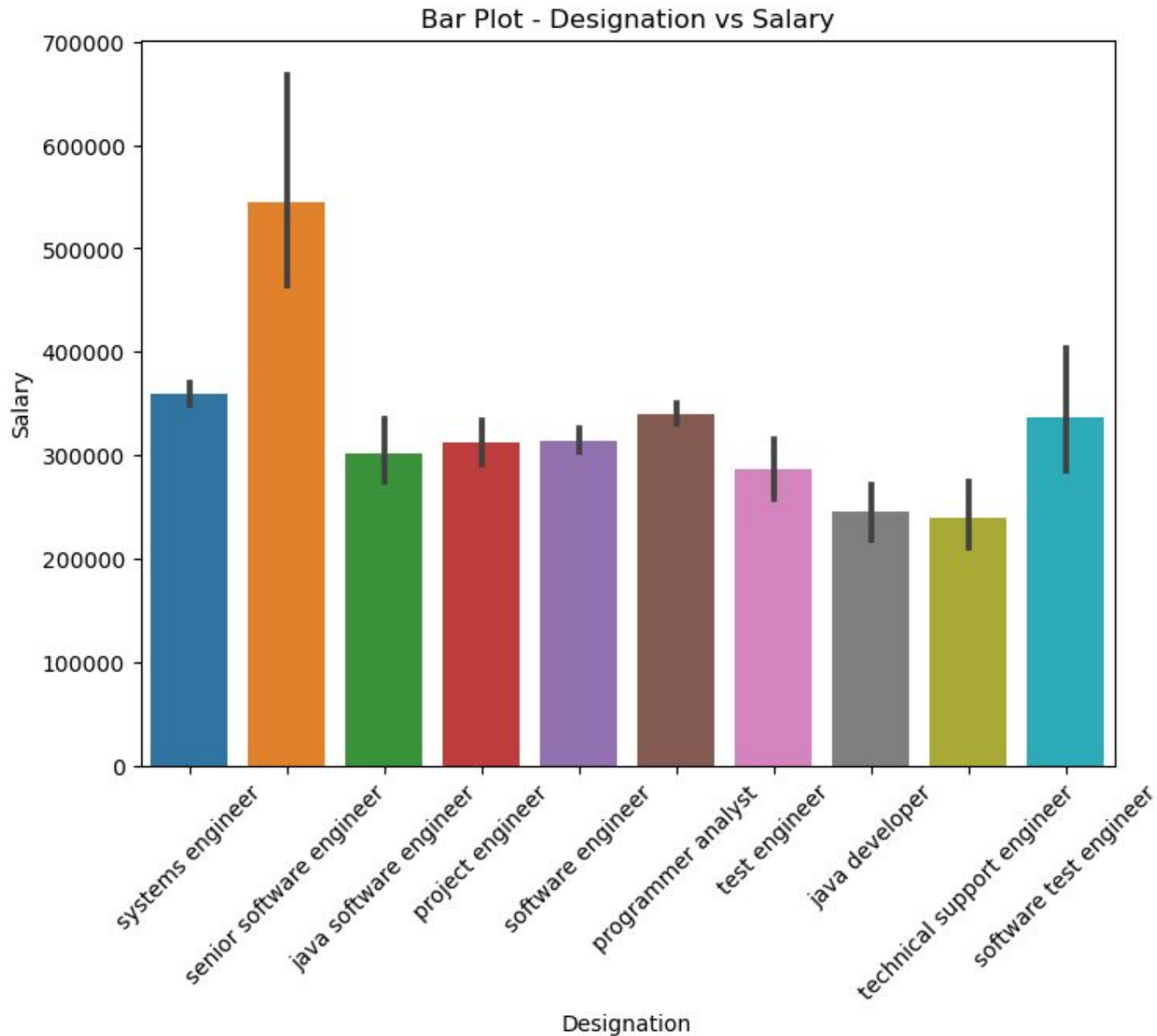
Point Plot - Gender vs Salary



Observation

- **Barplot** : We can observe that compare to Female Male has the higher salary
- **Pointplot** : We can observe that the max salary of Female is around 295000 and Male is 310000000

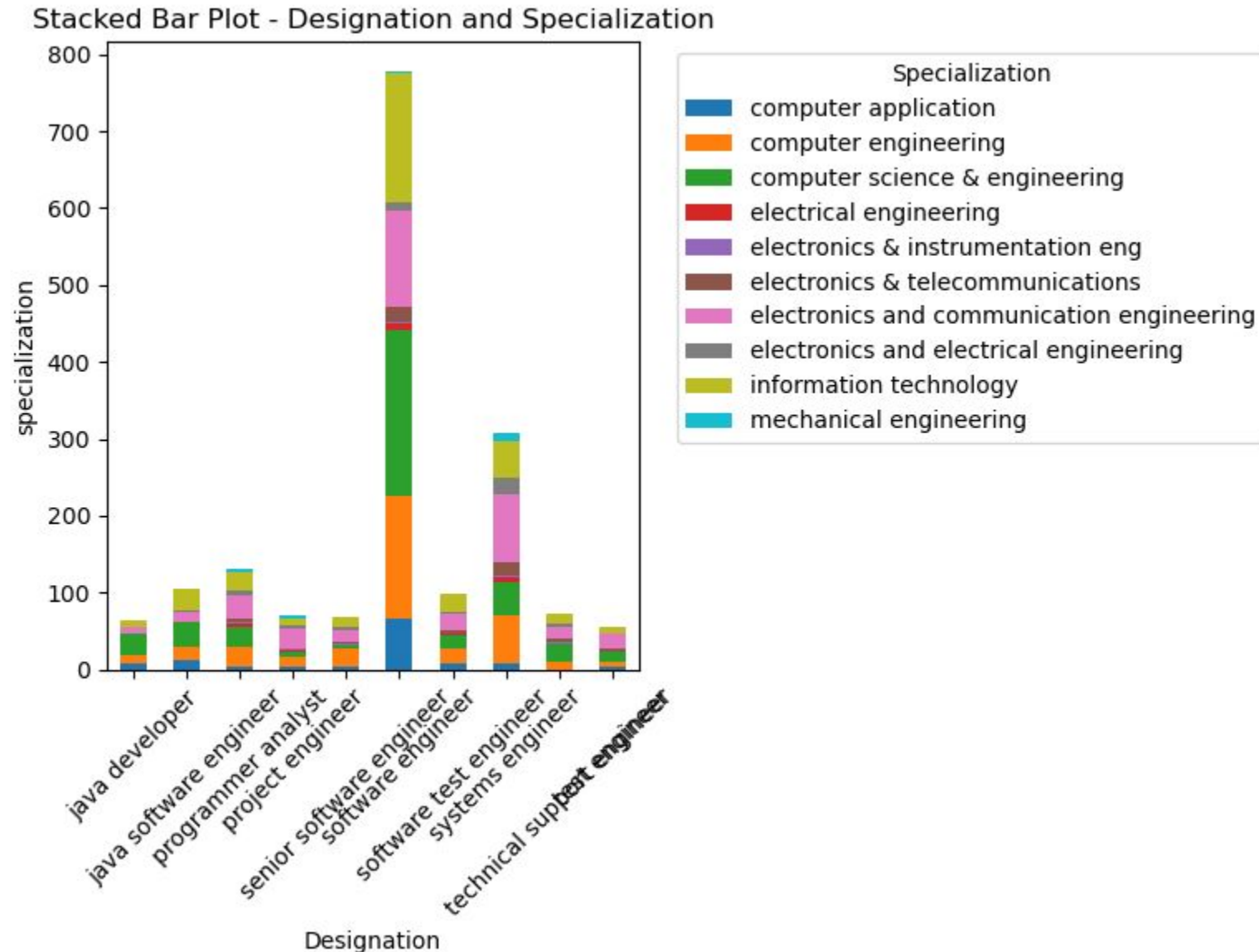
Designation vs Salary



Observation

---- Senior software engineer has the highest salary followed by system engineer

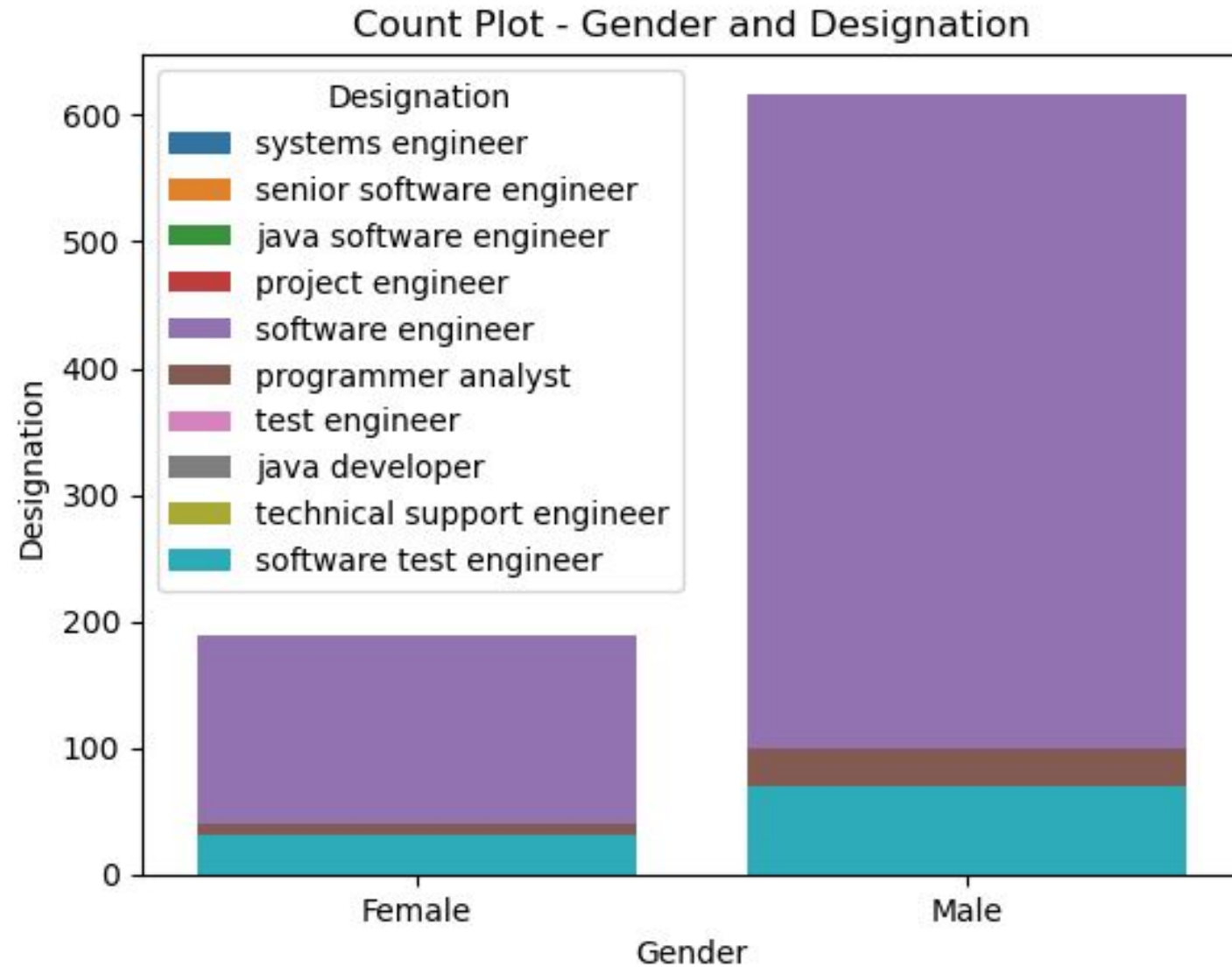
Designation and Specialization



Observation

---- From the above plot we can understand that most software developers are from computer science , electronics and information technology

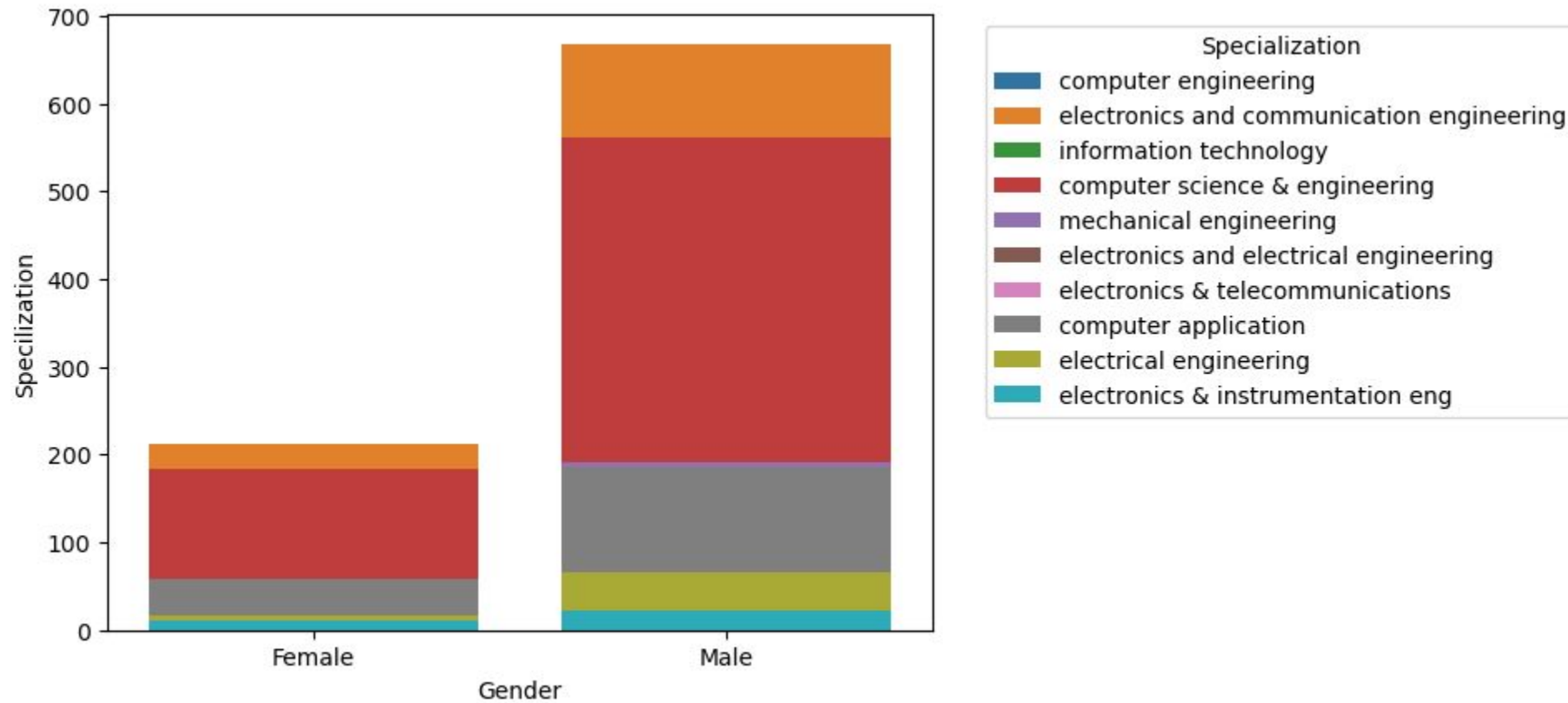
Gender and Designation



Observation

---- From the above plot we can tell that compared to Female large number of Male have the software engineer designation.

Research Question



Observation

**** We can clearly see that male has chosen computer science when compared to female

Conclusion

Based on the Analysis made on the AMEO data set and found the relationship between the Salary and other variables. Here are some insights

1. Computer Science specialization has the highest median salary.
2. Software Engineer are the most aimed profession.
3. Males have highest paid when compared to female.
4. Software Engineer Domain has the largest number of employees.
5. Graduates in Computer Science & Engineering show a strong alignment between their degree and specialization choices.

THANK
YOU

