



INNOVATION. AUTOMATION. ANALYTICS

PROJECT ON

AMCAT – Exploratory Data Analysis

About me :

My self **Suryapakula Bhanuprakash** (IN9240486). I had a Masters Degree in **Statistics** from Pondicherry University ,passed out in the year May'2023.

As a statistics student, I want to learn data science because how and where the domain knowledge is utilized in large amounts of data and in turn it helps me to survive in the job market, opens up many exciting job opportunities, and allows me to contribute my findings in solving undefined challenges.

I had worked as a MIS-Executive in Kalamandir and now I'm looking for better opportunities. And finally I thank **Innomatics** for providing an opportunity to grow & upskill my knowledge and to assist my career path.

Contact Links.....

- LinkedIn : <https://www.linkedin.com/in/bhanux18/>
- GitHub : <https://github.com/Bhanux18>



Business Problem :

The main aim of this project is with respect to current job market over higher competition, how likely do freshers as well as experienced candidates can get the job in concerned domain of interest. In which how the salary will deviate with respect to the Job location, Specialization, Skills, Exam Scores, College Tiers ,Experience and more parameters are involved.

Objective of the Project :

The initial goal of an AMCAT (Aspiring Minds Computer Adaptive Test) project, especially when focusing on data analysis, is to conduct an Exploratory Data Analysis (EDA) on the dataset. This process aims to identify patterns and relationships within the data, offering valuable insights into the employment outcomes of engineering graduates.

Summary of the Data :

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.

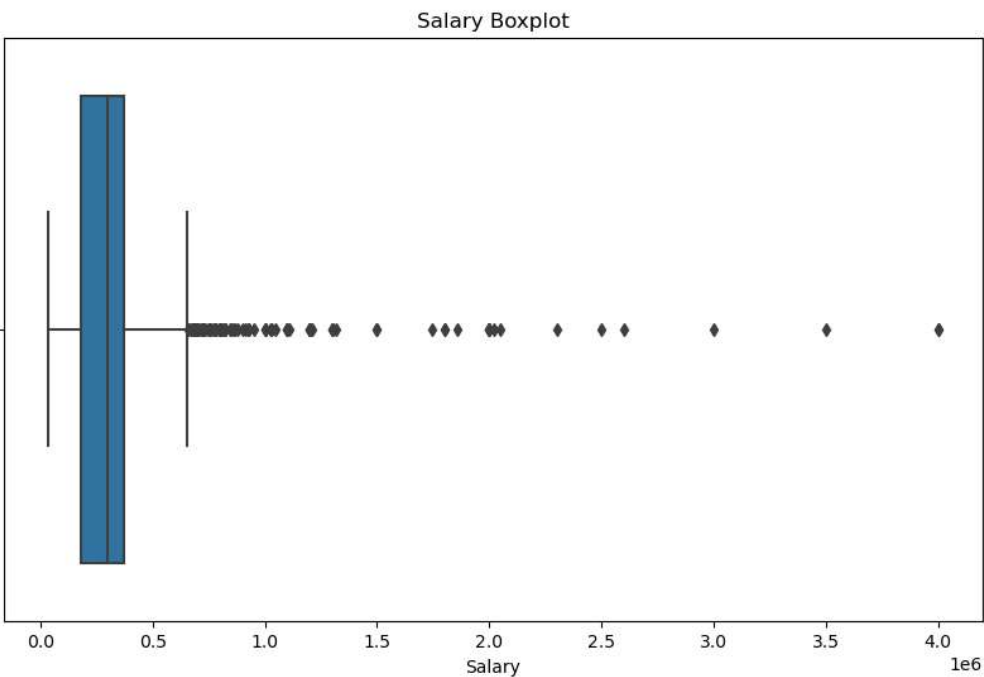
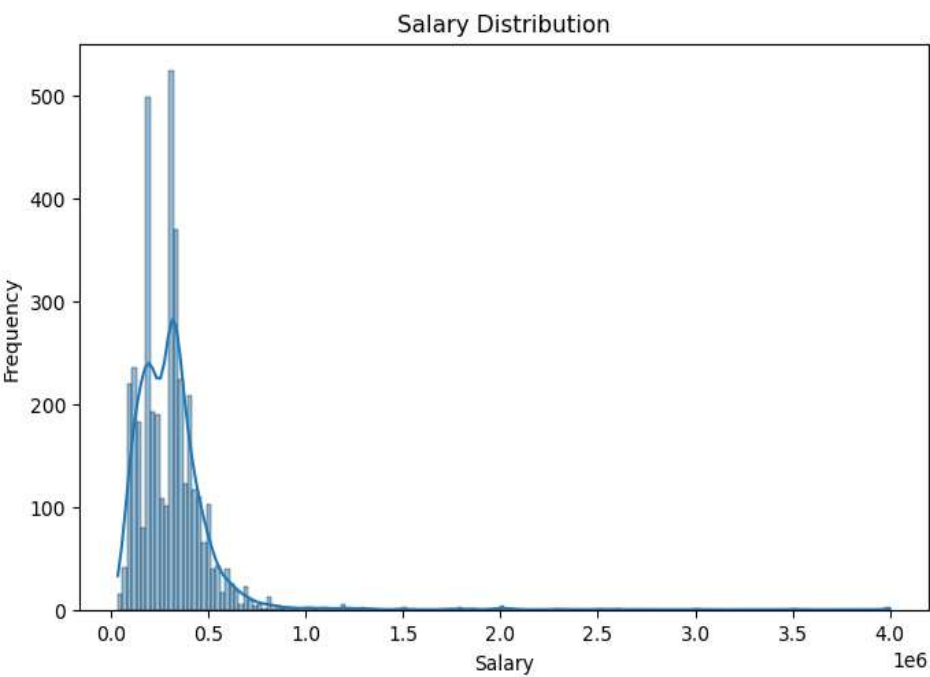
Exploratory Data Analysis:

This process involves concerned points in order extract insights through Visual and Non-Visual approaches.

- ✓ **Data Cleaning Steps**
 - Removing duplicates, handling missing values, correcting errors.
- ✓ **Data Manipulation Steps**
 - Sorting data, creating new variables, merging datasets.
- ✓ **Univariate Analysis Steps**
 - Non-Visual: Calculating mean, median, mode, and standard deviation.
 - Visual: Plotting histograms and box plots to visualize data distribution.
- ✓ **Bivariate Analysis Steps**
 - Non-Visual: Calculating correlation coefficients, performing regression analysis.
 - Visual: Creating scatter plots and heatmaps to explore relationships between variables.

Univariate Analysis :

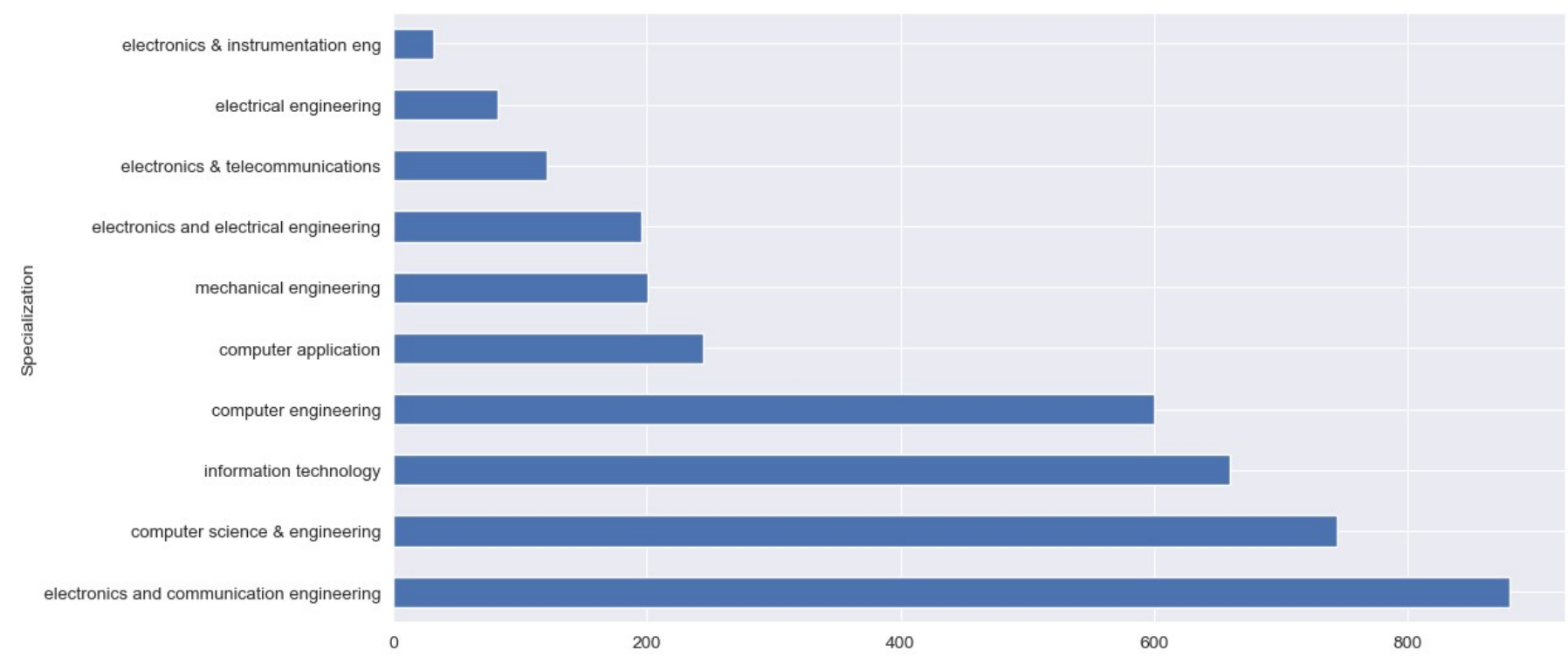
What is the distribution of a Target Column ?



The targeted salary attribute having outliers that are skewed right . It indicates that most of the candidates having higher salaries than the average.

Univariate Analysis :

What are the top 10 most common specializations among candidates?

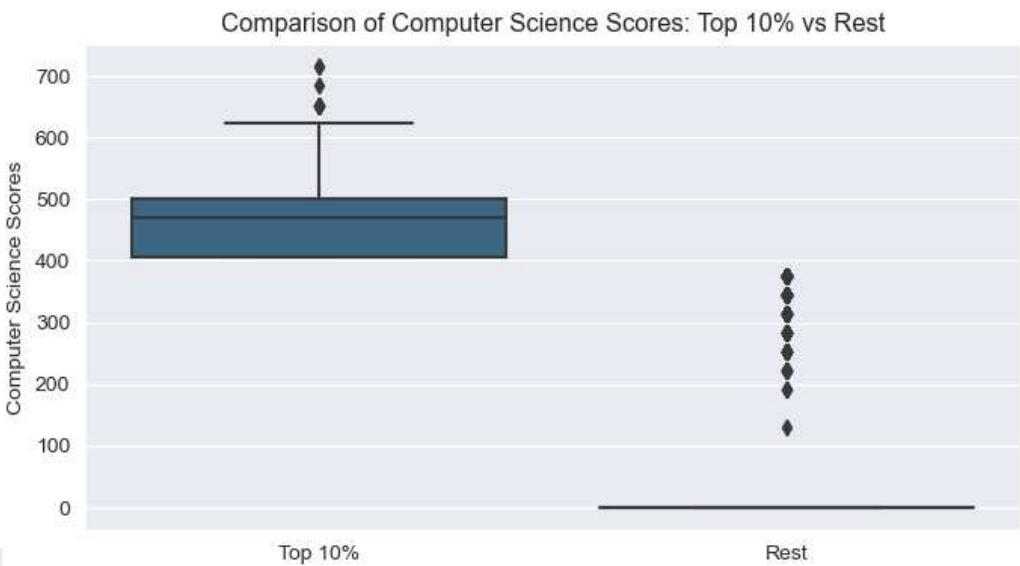
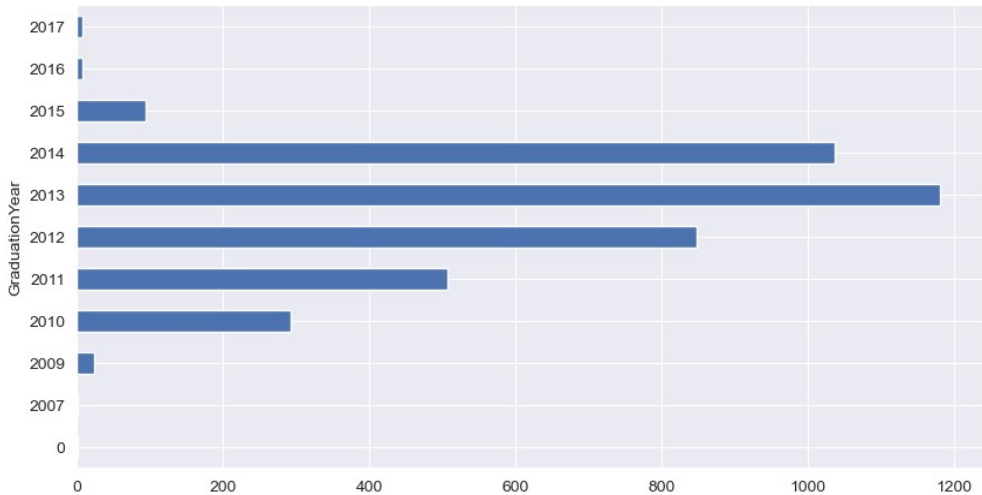


From the above plot it is observed that on top 10 Specializations, Electronics and Communication Engineering got topped in first with 800 candidates for this Specialization and the least admitted course was Electronics and Instrumental engineering holds 32 candidates.

Univariate Analysis :

How do the top 10% of candidates in Computer Science scores compare to the rest?

This comparison highlights that the top 10% of candidates not only score higher but also have more consistent performance in Computer Science, because of higher median scores with less variability. Unlike more variability and a lower median score of the rest scores.

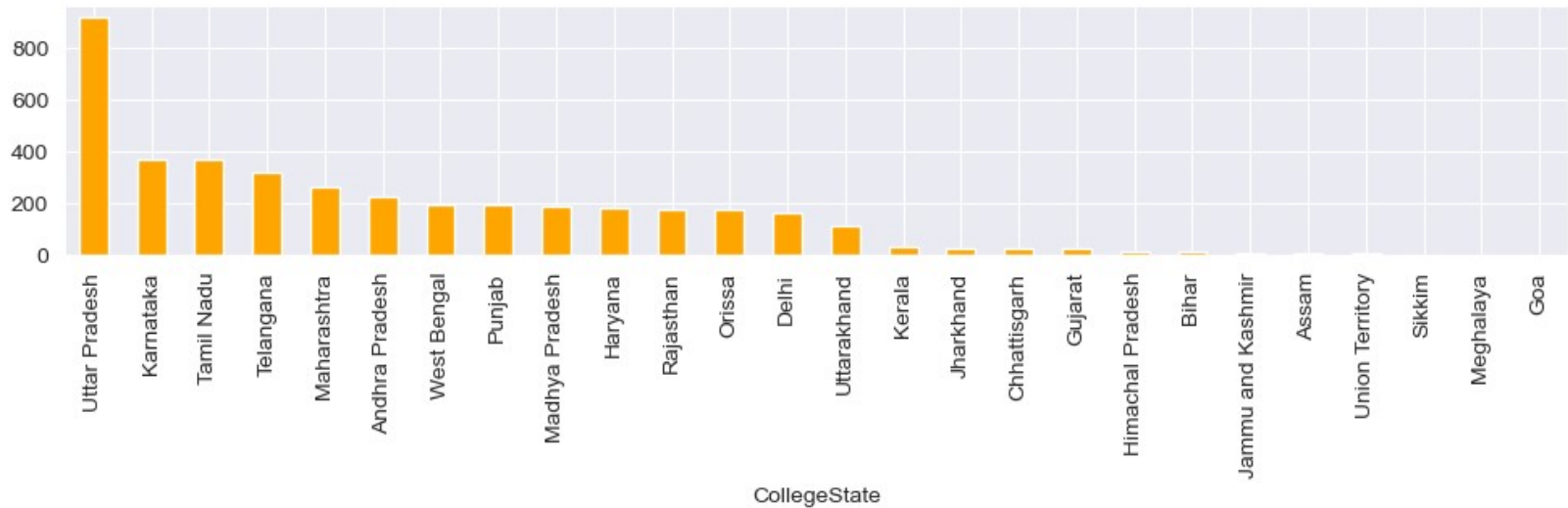


How many candidates graduated in each year?

From the top 10 records , this bar plot highlights that students from Uttar Pradesh about 915 are in higher than other states taken AMCAT Exam.

Univariate Analysis :

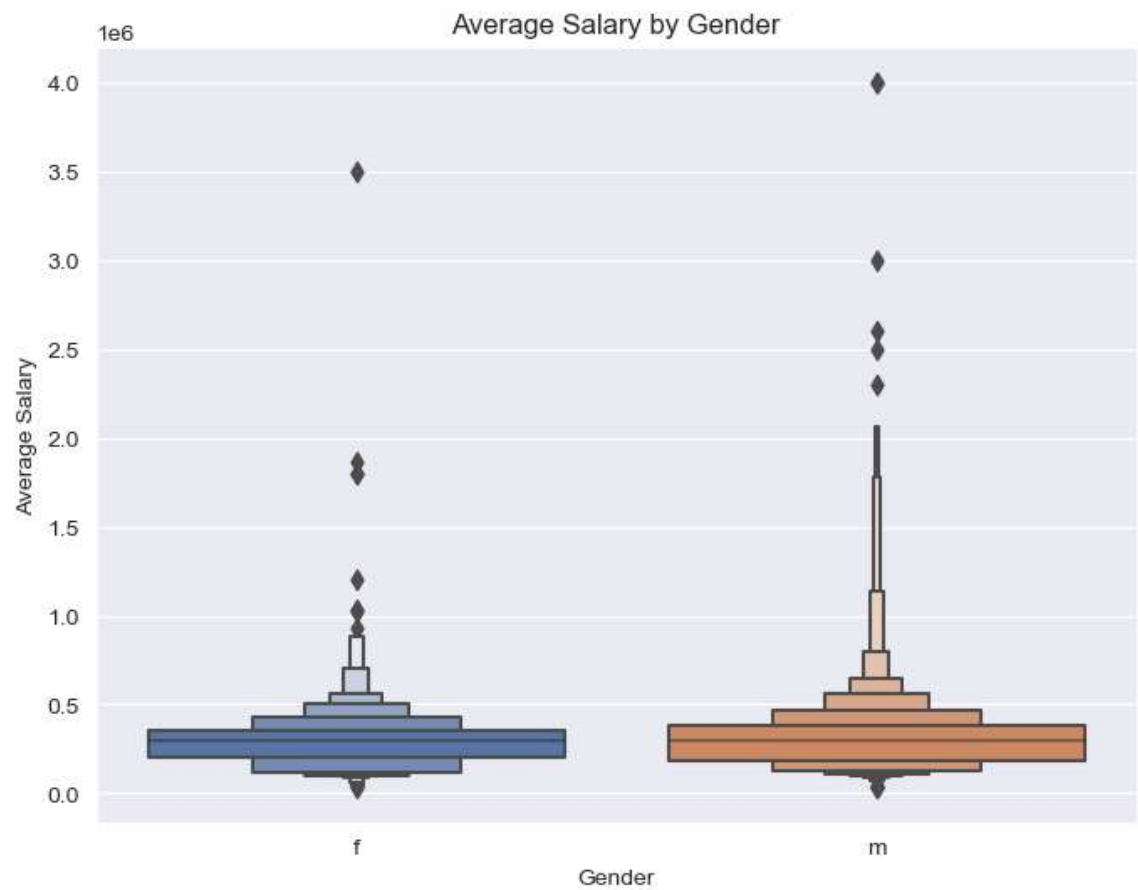
How many college students from each state ?



From the top 10 records, this bar plot highlights that students from Uttar Pradesh about 915 are in higher than other states taken AMCAT Exam

Bivariate Analysis :

Do male and female candidates have different average salaries?

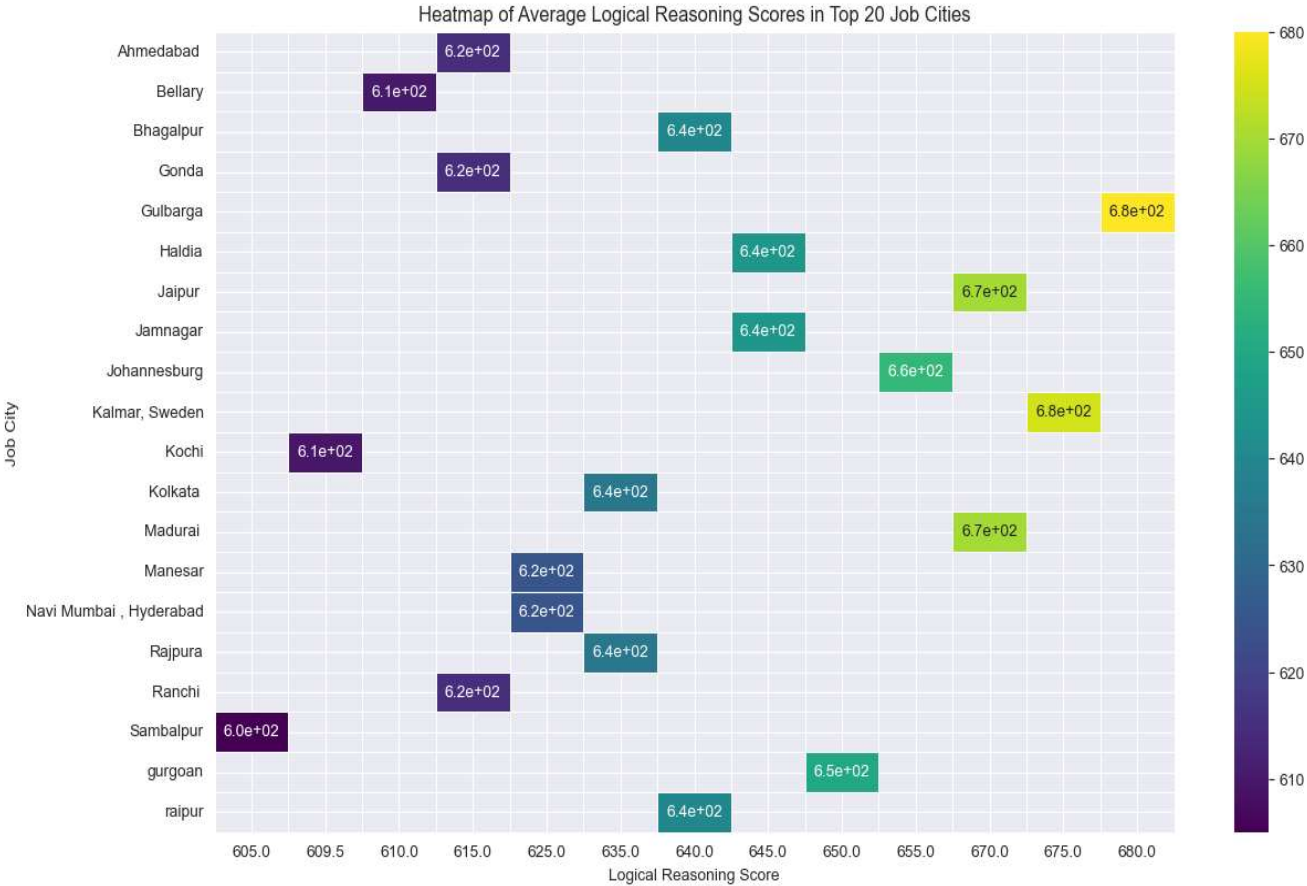


Yes, We can see that males are tend to earn a higher distributed salaries compared to the females though the average salary tend to approximately same.

Bivariate Analysis :

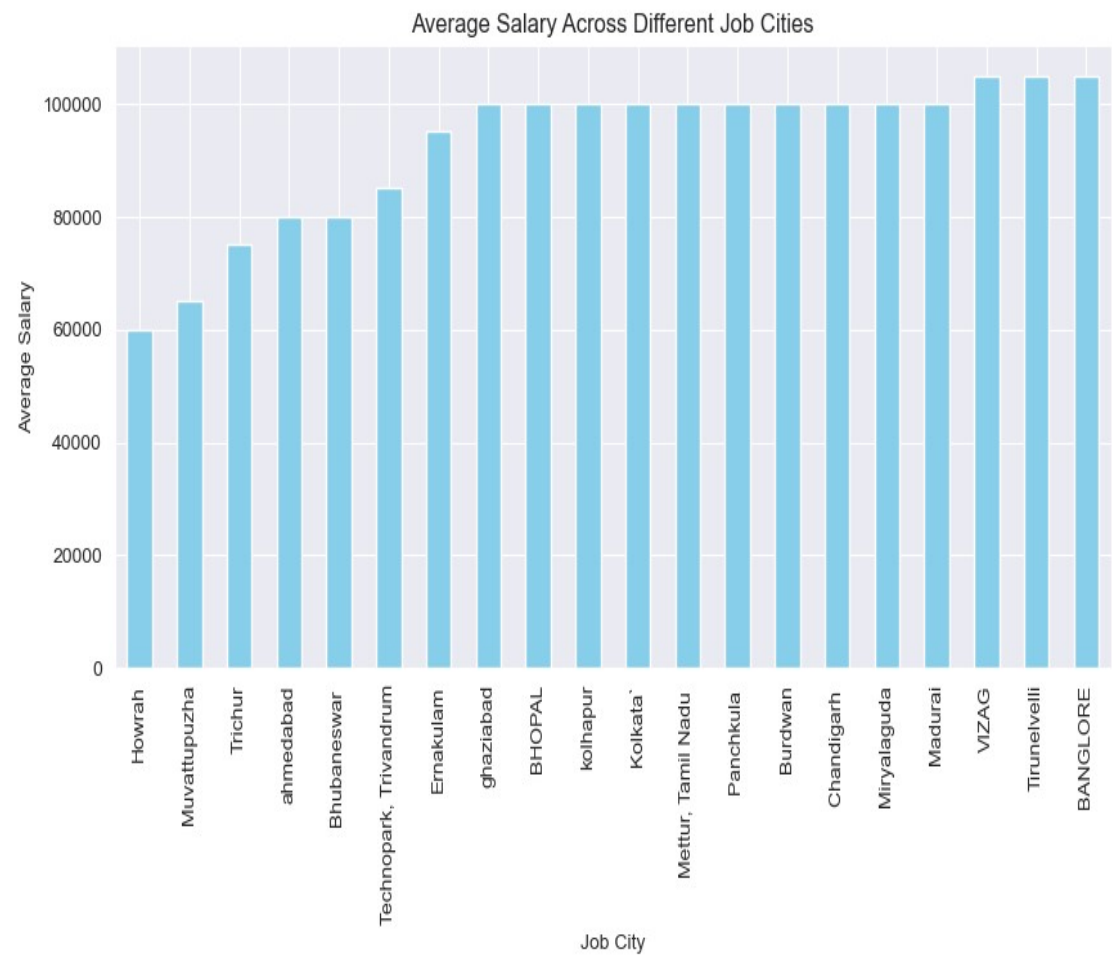
How do logical reasoning scores vary across different job cities ? Get top 20

For the top 20 job cities, Gulbarga city had the higher logical reasoning score of 680 and lower score was found in Sambalpur city respectively.



Bivariate Analysis :

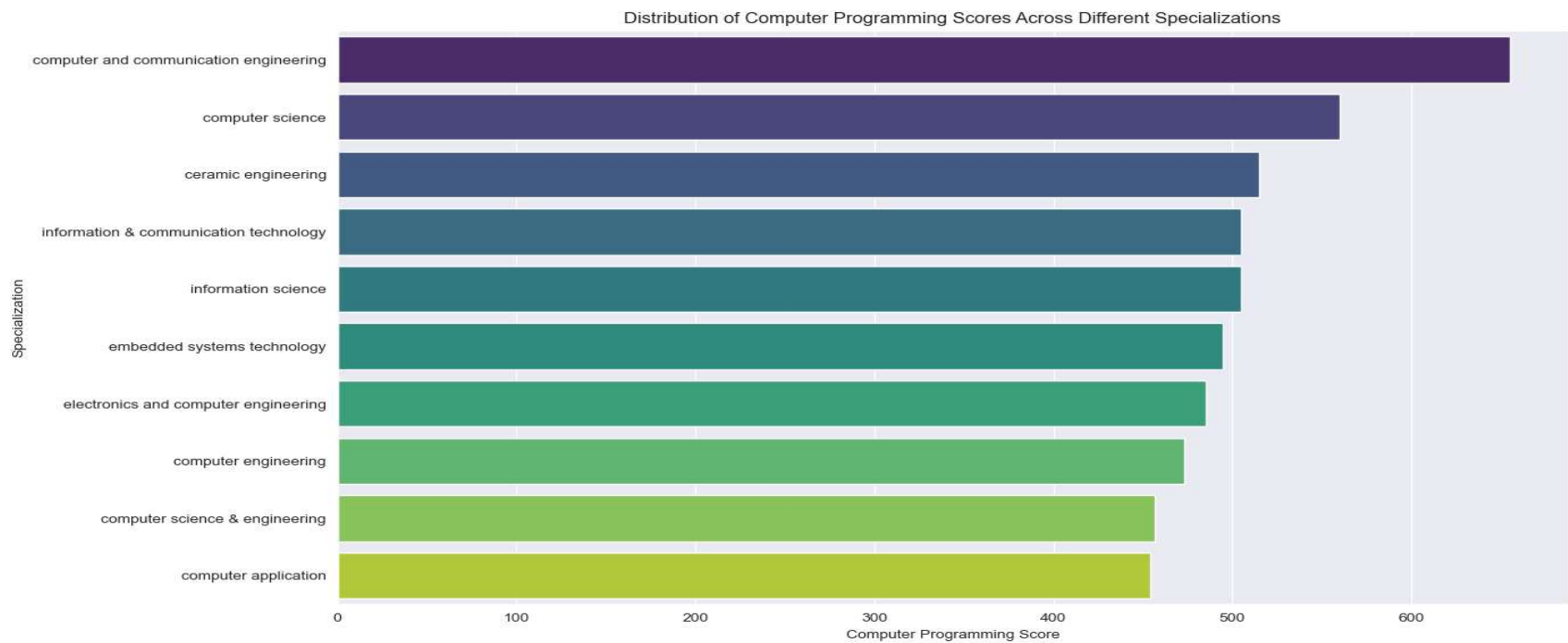
What is the distribution of salaries across different job cities?



Out of top 20 job cities, cities like VIZAG, Tirunelveli , BANGLORE holds the same average salary package of 1,05,000.In which Howrah is the one with lowest average salary package.

Bivariate Analysis :

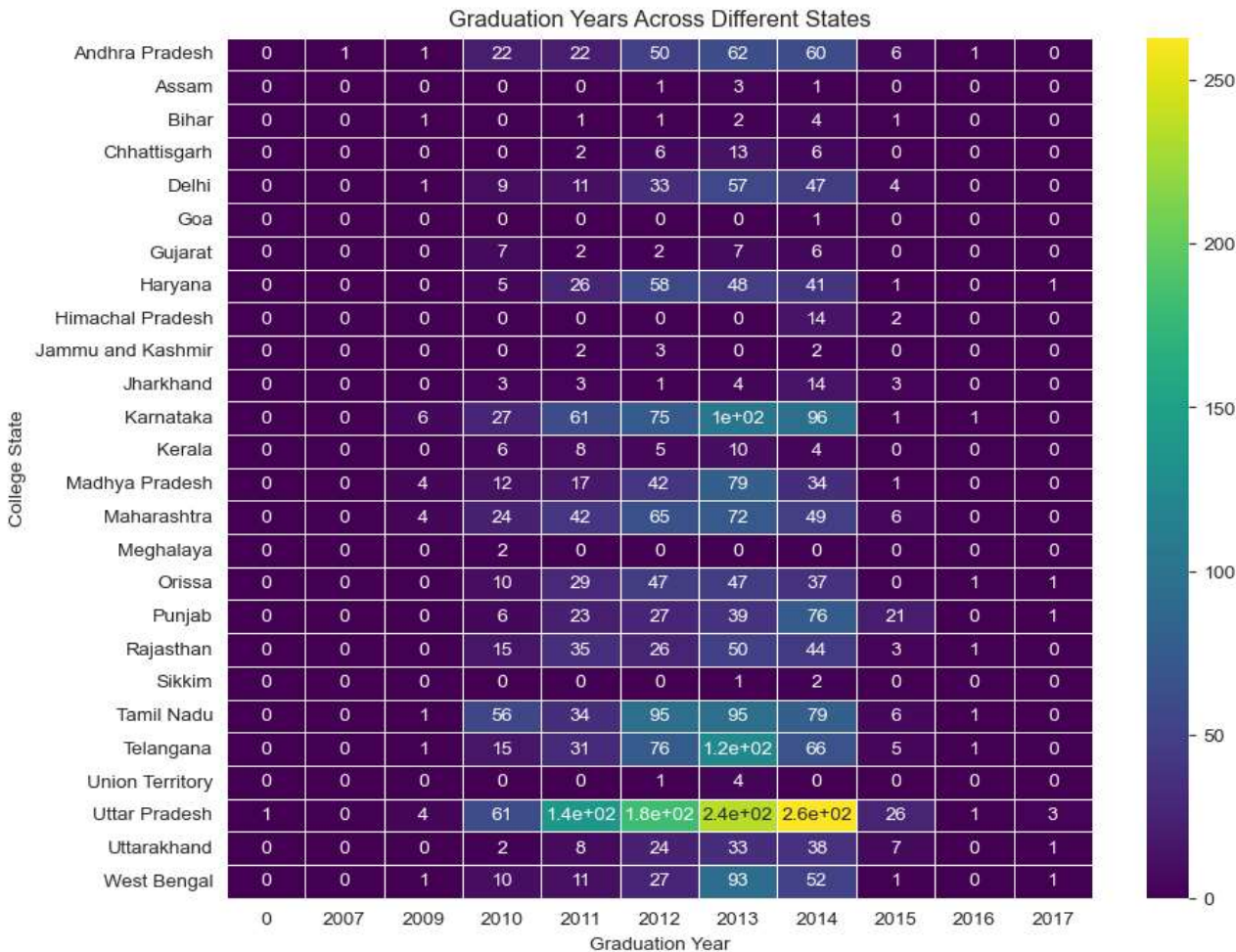
How do computer programming scores vary across different specializations?



This bar graph of top 10 records indicates that computer and communication engineering specialization topped among the all with a score of computer programming is 655, in which computer applications scores less around 470.

Bivariate Analysis :

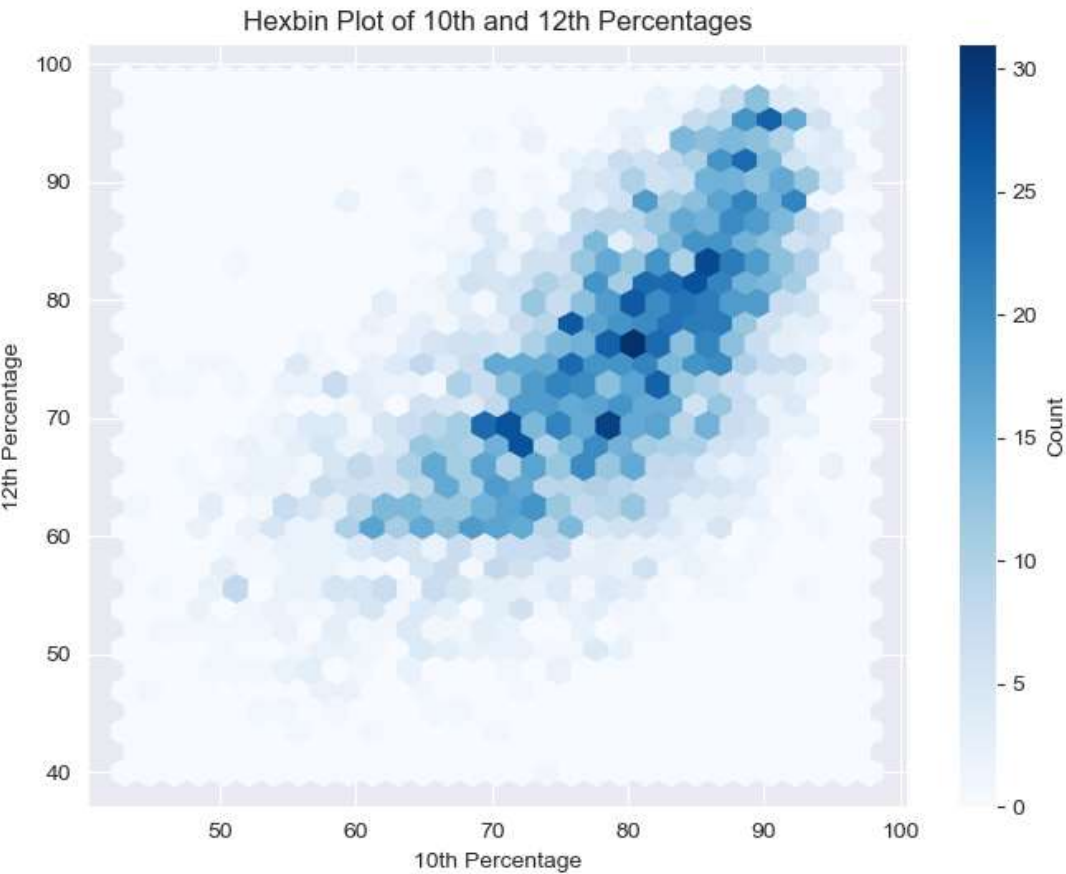
How does the distribution of graduation years vary across different states?



Concerned heatmap shows that 263 candidates form Uttar Pradesh got graduated in 2014 and 234 candidates in 2013, which correlates with the facilities available in the cities to topped among all.

Bivariate Analysis :

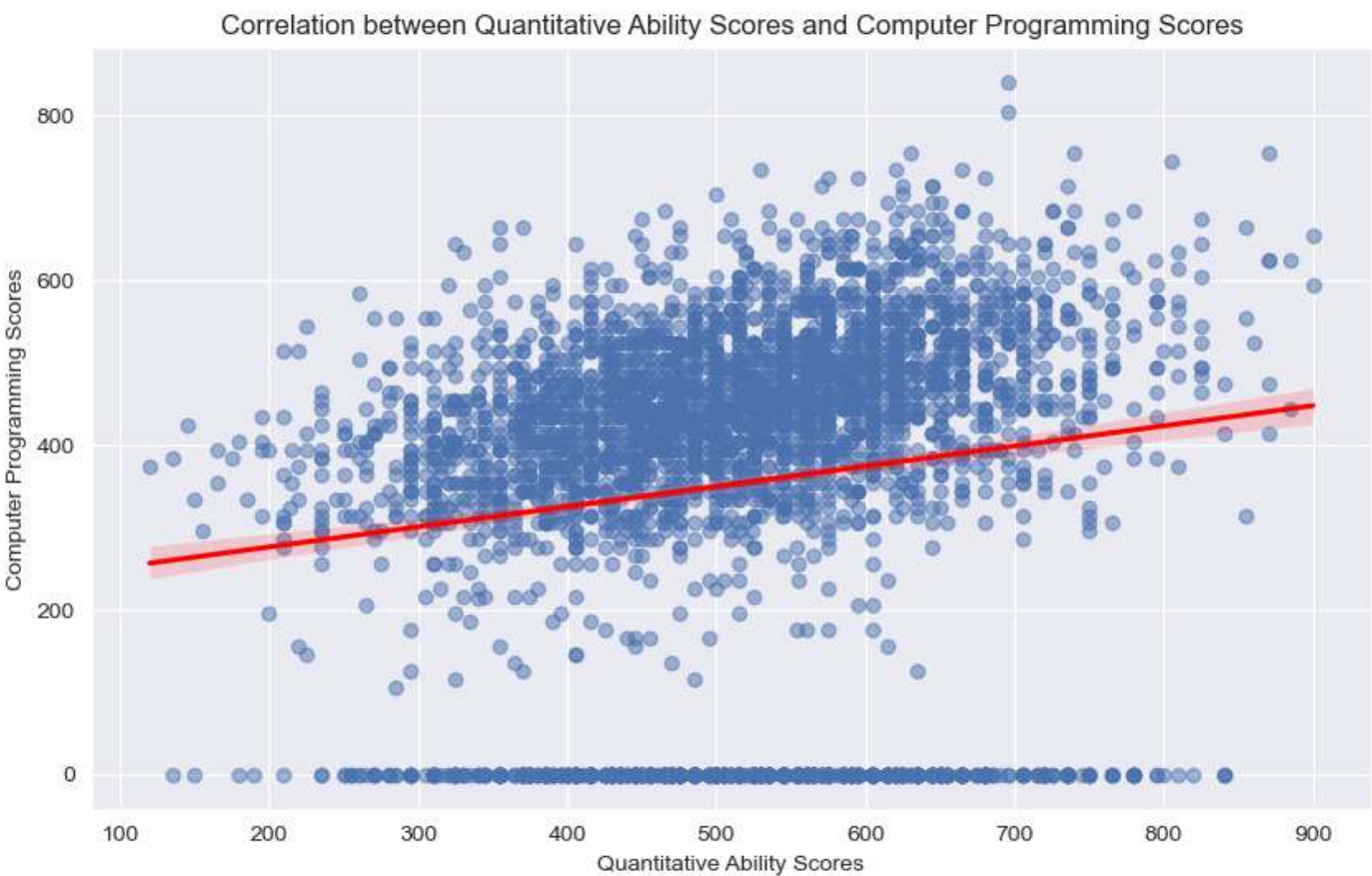
Is there a correlation between 10th and 12th percentages?



*We can see the strong positive correlation between the 10th and 12th percentages with Pearson value of correlation 0.64 .
If the student who scores better in 10th will definitely score in 12th grade too.*

Bivariate Analysis :

How do quantitative ability scores correlate with computer programming scores?



A strong positive correlation value of 0.64 can be seen that there is linear relationship between programming scores and quantitative scores . And this type of candidates have good effect on salary & skillset.

Conclusion :

Through this entire Exploratory Data Analysis, the drawn insights deals with the salary field which gives the hypothetical ideas about the how the particular parameters contributes the job opportunities. Few key points are...

Salary Distribution:

The salary attribute shows right-skewed outliers, indicating that a few candidates earn significantly more than the average salary.

Engineering Scores:

Fields like Computer Science, Mechanical Engineering, Civil Engineering, Electrical Engineering, and Telecom Engineering have right-skewed outliers, suggesting higher scores pulling the mean to the right.

Specializations:

Electronics and Communication Engineering has the highest number of candidates (800), while Electronics and Instrumentation Engineering has the least (32).

State-wise Participation:

Uttar Pradesh has the highest number of candidates (915) taking the AMCAT exam, compared to other states.

Hypothesis Analysis:

We Reject the null hypothesis as " There is a significant difference between 10th and 12th percentages. It means that there is a statistically significant difference between 10th and 12th percentages. Similarly , quantitative & computer programming have a strong significance of association between them, since the p-value is less than at 0.05 significance level.

THANK YOU

