



INNOVATION. AUTOMATION. ANALYTICS

PROJECT ON

Exploratory Data Analysis on
AMCAT Data

By,
Sahithi Thakur
IN9240798

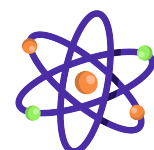
About:

My name is Sahithi Thakur, and I hold a B.Sc. in Computer Science from St. Francis College for Women. During my studies, I developed a deep passion for data science, which inspired me to explore and learn more about the field. Driven by this interest, I enrolled in a data science course at Innomatics Research Labs, where I worked on several hands-on projects that enhanced my practical skills. After completing the course, I joined a data science internship at Innomatics to further upskill myself and continue my journey in this exciting field.

As they say, 'The best way to predict the future is to create it.' This quote motivates me to constantly learn and adapt, as I aspire to create meaningful impact through data-driven insights.



<https://www.linkedin.com/in/sahithi-thakur-193ab6258/>



https://github.com/SahithiThakur/Data_Science-Internship

Problem statement:

Analysis AMCAT test data point to identify certificate performance trends ,strength and weakness job selection processes often lead to mismatches between candidates and positions, resulting in decreased productivity and increased turnover. This project aims to streamline the job selection process by identifying key criteria for successful optimize the use of AMCAT scores in the hiring candidate evaluation and implementing strategies to improve the efficiency and effectiveness of the selection process, ultimately enhancing organizational performance and employee satisfaction.

Objective of the Project:

To develop and implement a comprehensive job selection framework that enhances the efficiency, fairness, and effectiveness of the recruitment process. This framework should include standardized criteria for evaluating candidates, streamlined procedures for application and assessment, and tools for ensuring alignment between candidates' skills and organizational needs. The ultimate goal is to optimize hiring outcomes by selecting candidates who are best suited for the roles and contribute positively to the organization's success."

AGENDA

- Business Problem and Use case domain understanding(If Required)
- Objective of the Project
- Web Scraping –Details (Websites, Processor you followed)
- Summary of the Data

Exploratory Data Analysis:

- a. Data Cleaning
- b. Data Manipulation Steps
- c. Univariate Analysis Steps
- d. Bivariate Analysis Steps

a

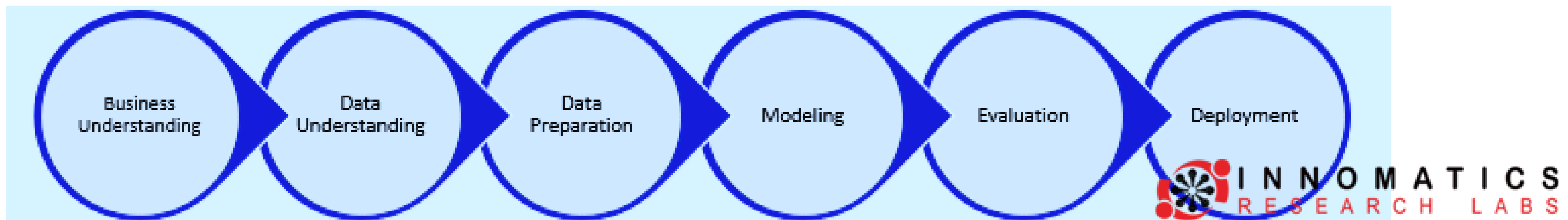
- Key Business Question
- Conclusion (Key finding overall)
- Q&A Slide
- Your Experience/Challenges working on Web Scraping –Data Analysis Project.



OBJECTIVE OF PROJECT

- Exploration of Relationships:
Investigate the relationships between independent variables and the target (dependent) variable, aiming to understand how various factors influence employment outcomes.
- Pattern and Trend Identification:
Identify any discernible patterns or trends in the dataset that could offer valuable insights into the dynamics of employment outcomes, particularly how skills, demographics, and education affect job placements and salaries.
- Research Question:
A Times of India article from January 18, 2019, claims that after completing Computer Science Engineering, fresh graduates in roles such as Programming Analyst, Software Engineer, Hardware Engineer, and Associate Engineer can earn between 2.5-3 lakhs per year. The task is to test this claim using the provided dataset.
Additionally, investigate whether there is a relationship between gender and specialization. Specifically, does a candidate's gender influence their preference for a particular specialization?

Workflow:-



DROPED UNWANTED COLUMNS

Removed 10 Un-important column

```
df.info()
```

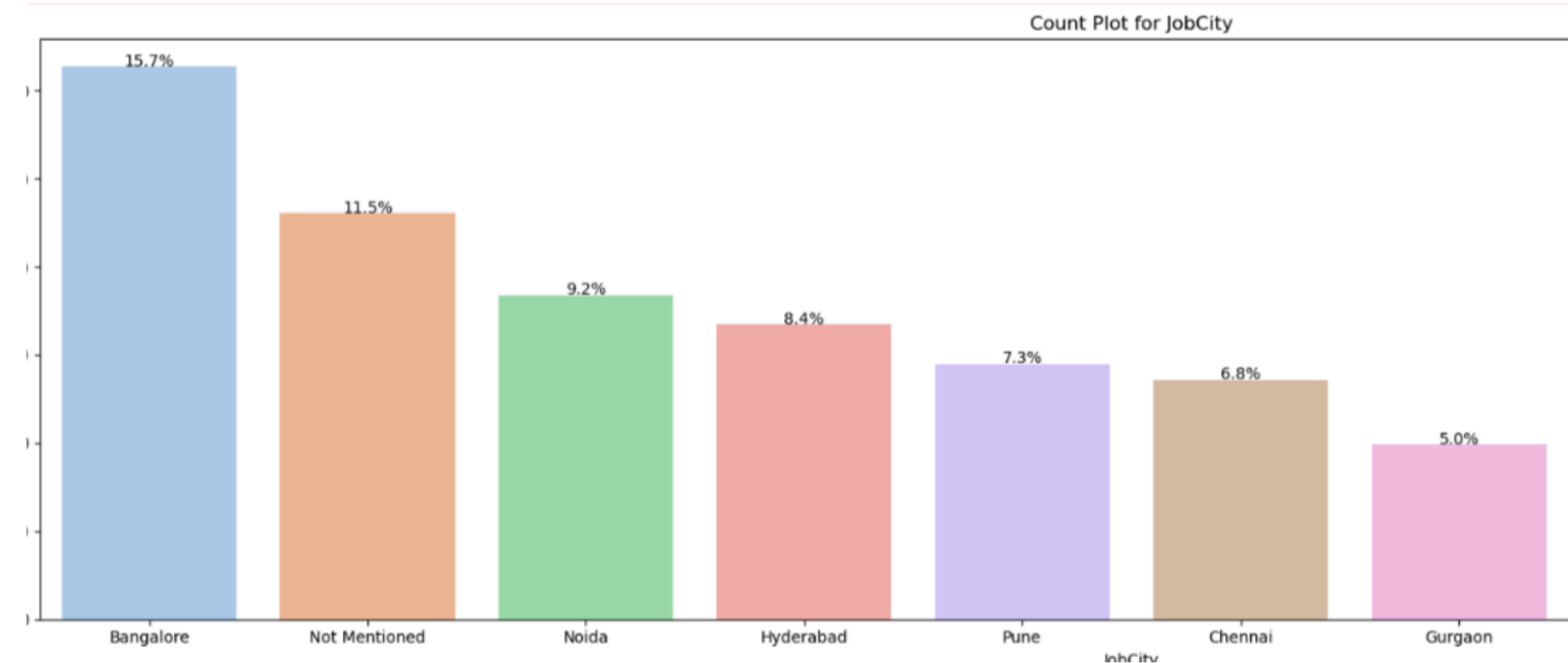
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3998 entries, 0 to 3997
Data columns (total 39 columns):
#   Column                                Non-Null Count  Dtype
---  ---                                -
0   Unnamed: 0                            3998 non-null   object
1   ID                                    3998 non-null   int64
2   Salary                              3998 non-null   int64
3   DOJ                                 3998 non-null   datetime64[ns]
4   DOL                                 3998 non-null   object
5   Designation                         3998 non-null   object
6   JobCity                            3998 non-null   object
7   Gender                             3998 non-null   object
8   DOB                                 3998 non-null   datetime64[ns]
9   10percentage                        3998 non-null   float64
10  10board                             3998 non-null   object
11  12graduation                        3998 non-null   int64
12  12percentage                        3998 non-null   float64
13  12board                             3998 non-null   object
14  CollegeID                           3998 non-null   int64
15  CollegeTier                         3998 non-null   int64
16  Degree                             3998 non-null   object
17  Specialization                     3998 non-null   object
18  collegeGPA                         3998 non-null   float64
19  CollegeCityID                      3998 non-null   int64
20  CollegeCityTier                    3998 non-null   int64
21  CollegeState                       3998 non-null   object
22  GraduationYear                     3998 non-null   int64
23  English                            3998 non-null   int64
24  Logical                            3998 non-null   int64
25  Quant                              3998 non-null   int64
26  Domain                             3998 non-null   float64
27  ComputerProgramming                3998 non-null   int64
28  ElectronicsAndSemicon              3998 non-null   int64
29  ComputerScience                    3998 non-null   int64
30  MechanicalEngg                     3998 non-null   int64
31  ElectricalEngg                     3998 non-null   int64
32  TelecomEngg                        3998 non-null   int64
33  CivilEngg                          3998 non-null   int64
34  conscientiousness                  3998 non-null   float64
35  agreeableness                      3998 non-null   float64
36  extraversion                       3998 non-null   float64
37  nueroticism                        3998 non-null   float64
38  openness_to_experience              3998 non-null   float64
dtypes: datetime64[ns](2), float64(9), int64(18), object(10)
memory usage: 1.2+ MB
```

```
df.info()
```

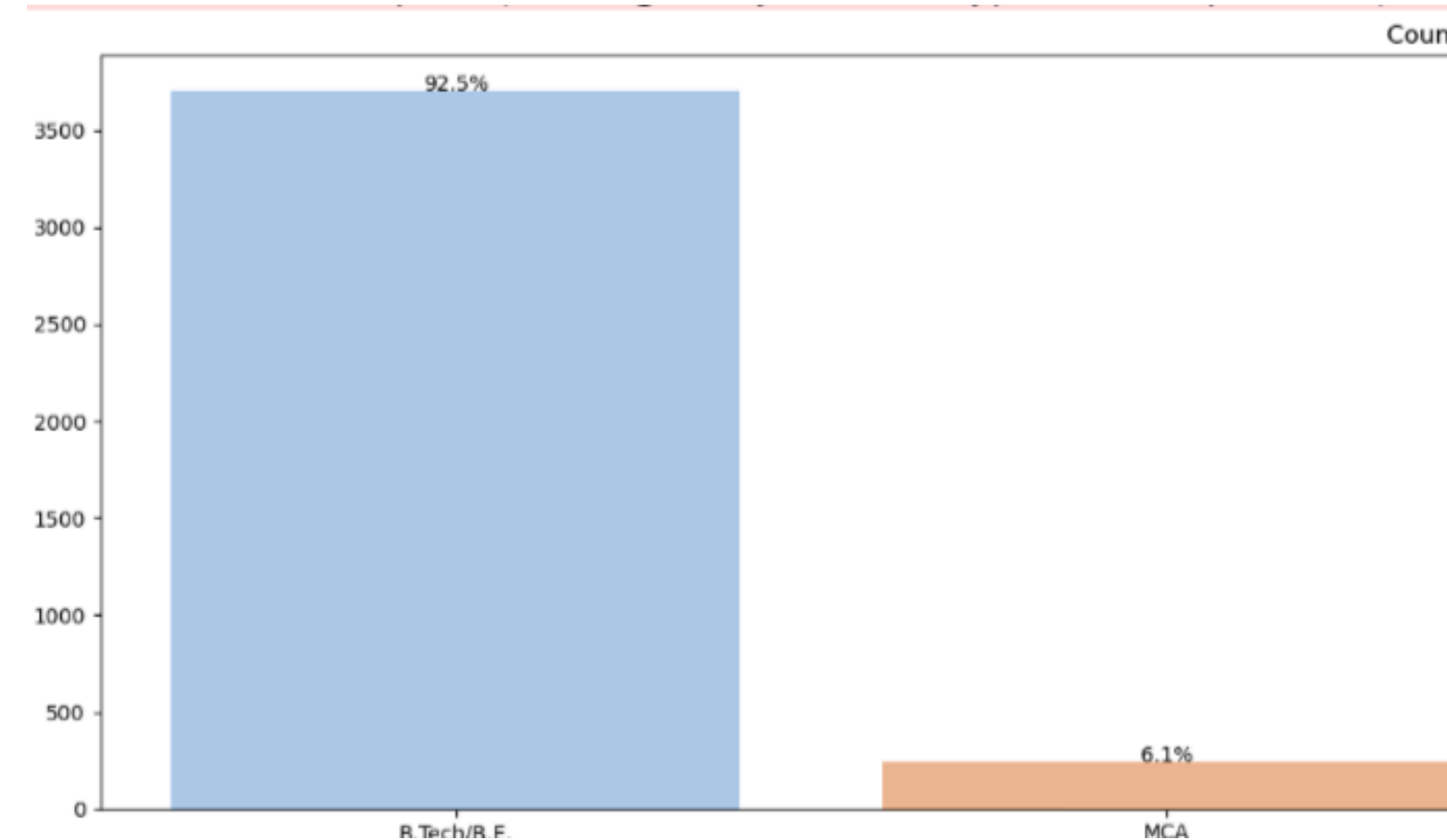
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3998 entries, 0 to 3997
Data columns (total 29 columns):
#   Column                                Non-Null Count  Dtype
---  ---                                -
0   Unnamed: 0                            3998 non-null   object
1   Salary                              3998 non-null   int64
2   DOJ                                 3998 non-null   datetime64[ns]
3   DOL                                 3998 non-null   object
4   Designation                         3998 non-null   object
5   JobCity                            3998 non-null   object
6   Gender                             3998 non-null   object
7   10percentage                        3998 non-null   float64
8   12percentage                        3998 non-null   float64
9   CollegeTier                         3998 non-null   int64
10  Degree                             3998 non-null   object
11  Specialization                     3998 non-null   object
12  collegeGPA                         3998 non-null   float64
13  English                            3998 non-null   int64
14  Logical                            3998 non-null   int64
15  Quant                              3998 non-null   int64
16  Domain                             3998 non-null   float64
17  ComputerProgramming                3998 non-null   int64
18  ElectronicsAndSemicon              3998 non-null   int64
19  ComputerScience                    3998 non-null   int64
20  MechanicalEngg                     3998 non-null   int64
21  ElectricalEngg                     3998 non-null   int64
22  TelecomEngg                        3998 non-null   int64
23  CivilEngg                          3998 non-null   int64
24  conscientiousness                  3998 non-null   float64
25  agreeableness                      3998 non-null   float64
26  extraversion                       3998 non-null   float64
27  nueroticism                        3998 non-null   float64
28  openness_to_experience              3998 non-null   float64
dtypes: datetime64[ns](1), float64(9), int64(12), object(7)
memory usage: 905.9+ KB
```



Univariate - Visual Analysis (Categorical Column)

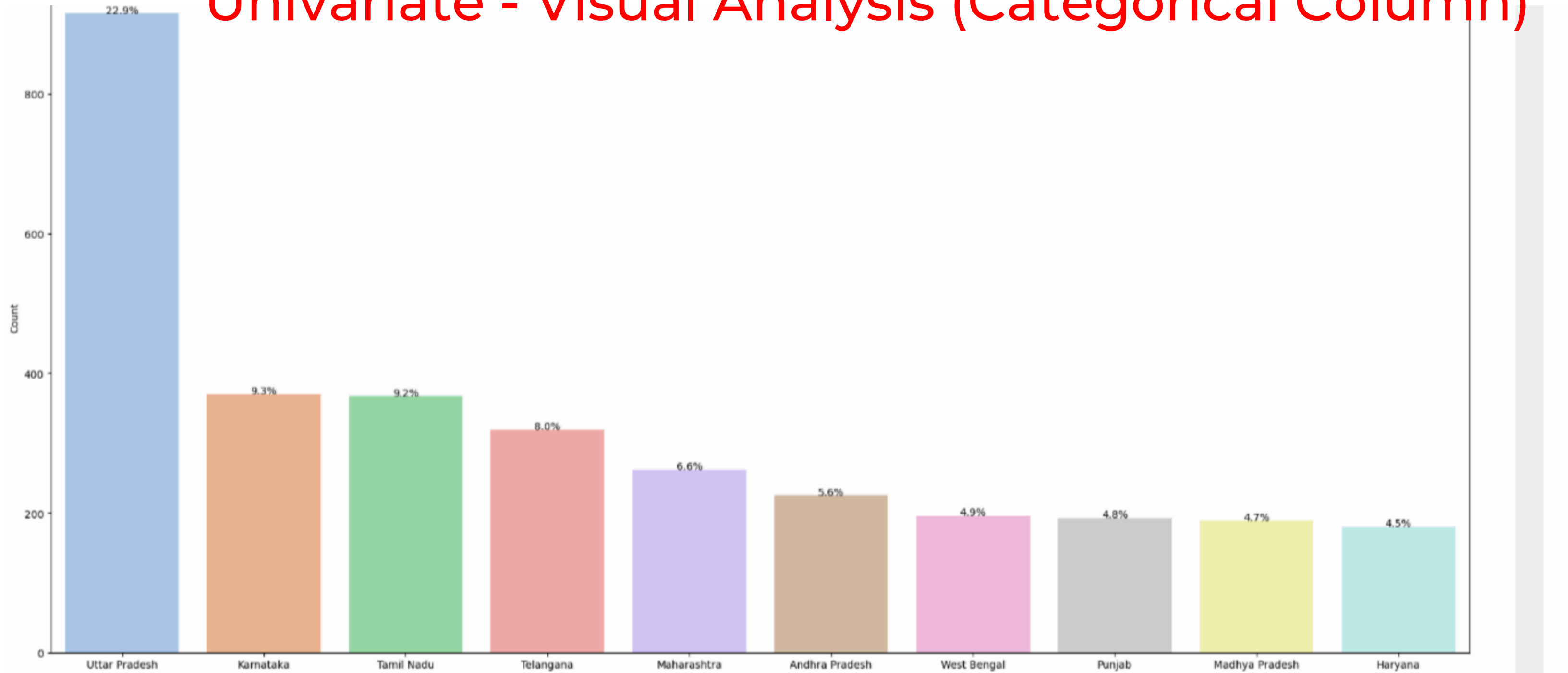


- This plot only shows seven ten Job Cities.
- Most of the students got job in Bangalore followed by Noida and Hyderabad.
- Almost 11.5% students who are employed are Not Mentioned their Job City



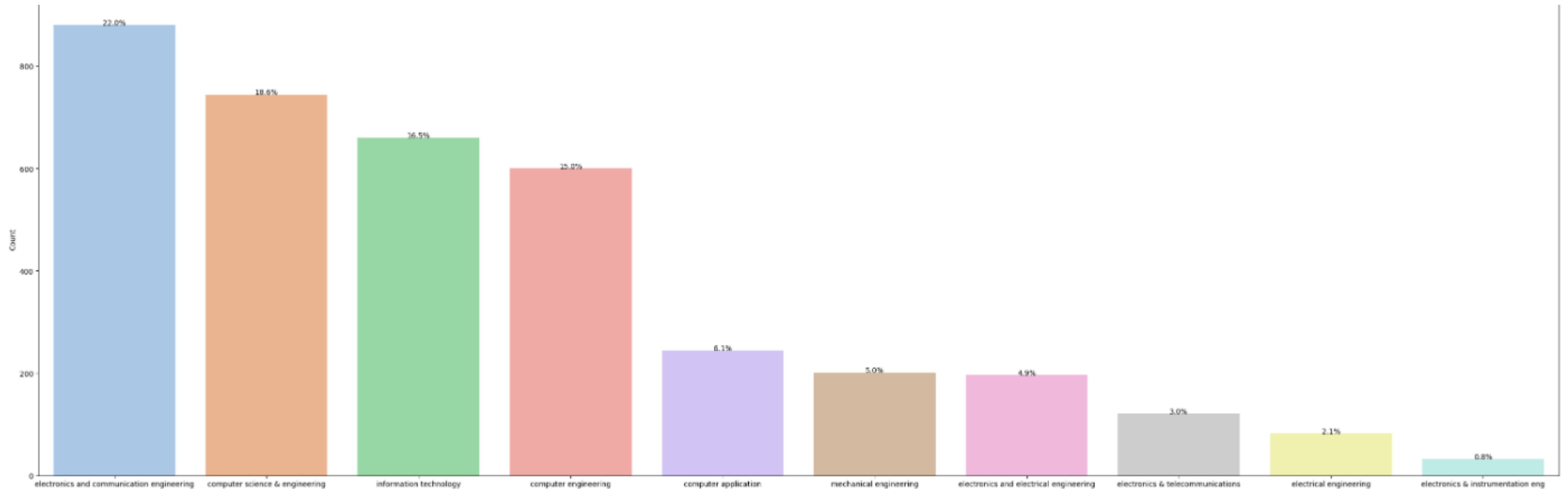
- Most of the students have their Degree is Btech with almost of 92.5% followed by MCA, M.Tech/M.E and M.Sc

Univariate - Visual Analysis (Categorical Column)



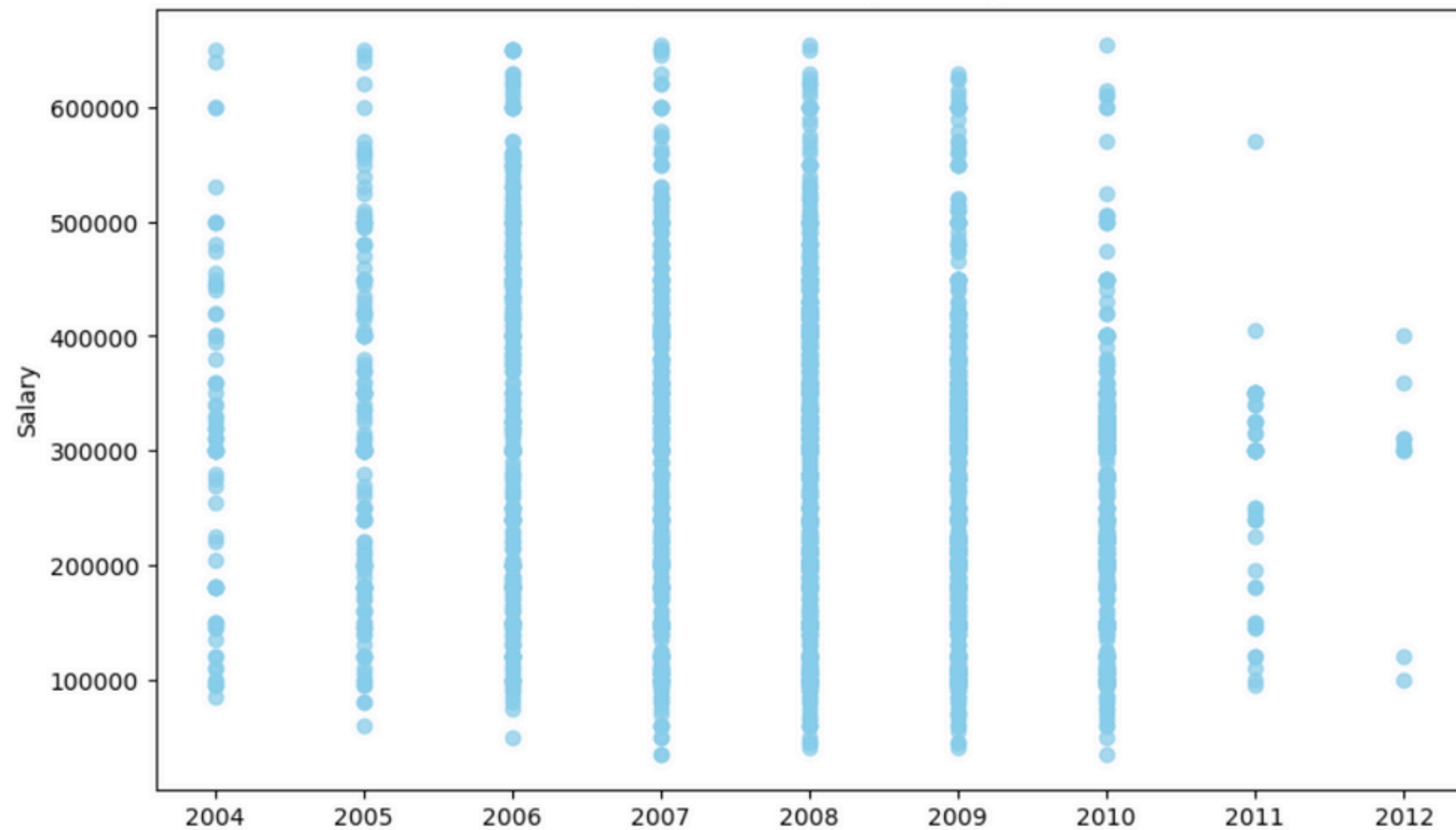
- This plot contains only top 10 CollegeState among many of them.
- Most of the students are from Uttar Pradesh with 22.9% followed by karnataka with 9.3% followed by Tamilnadu 9.2% so on

Univariate - Visual Analysis (Categorical Column)

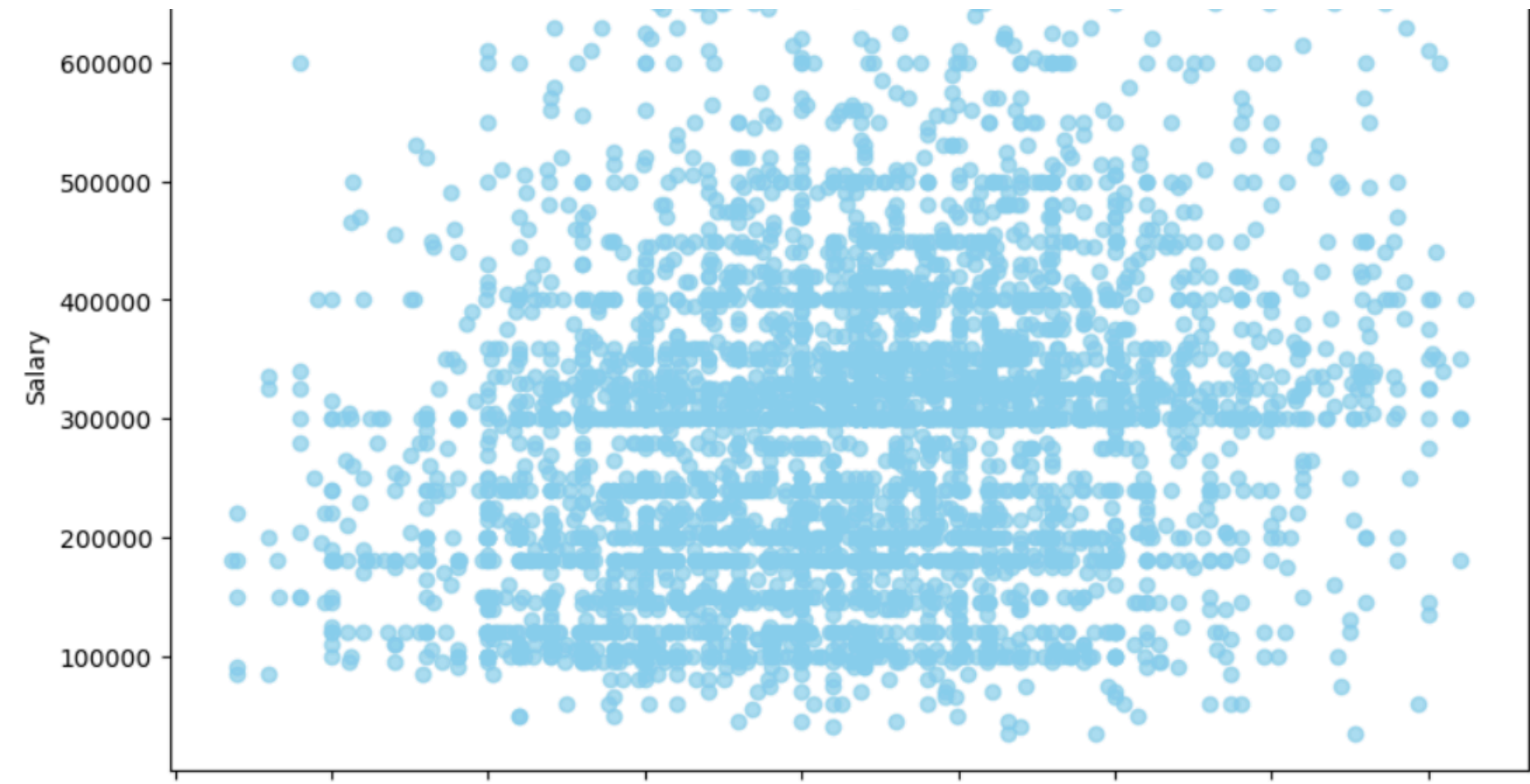


- This plot contains only top 10 Specializations among many of them.
- Most of the students with Degree Specialization is electronics and communication engineering with 22% followed by computer science and engineering with 18.6%, Information Technology with 16.5%

Bivariate analysis (numerical vs numerical)



- 2004 to 2010 12th graduation students have packages from 1lac to 6lac.
- 2012 graduation students are rarely got job that means data may not be there or who are 2012 12th passedout students are not employed in software industry

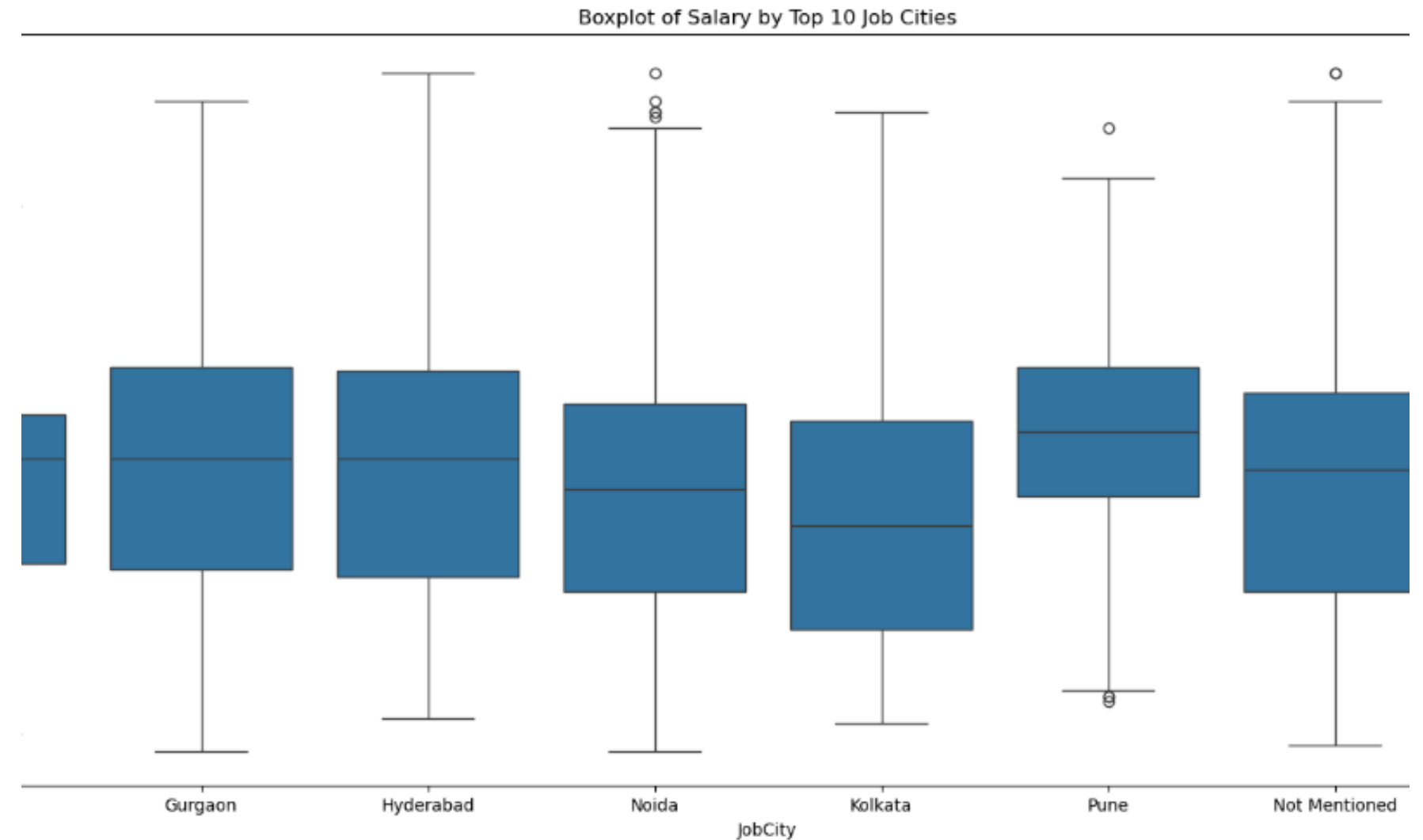


- who have 50 to 60 Degree Percentage have got rarely above 4lac package.
- who have 60 to 85 Degree percentage have got package from 1lac to 4lac and rarely 5 to 6lac package

Bivariate analysis (numerical vs numerical)

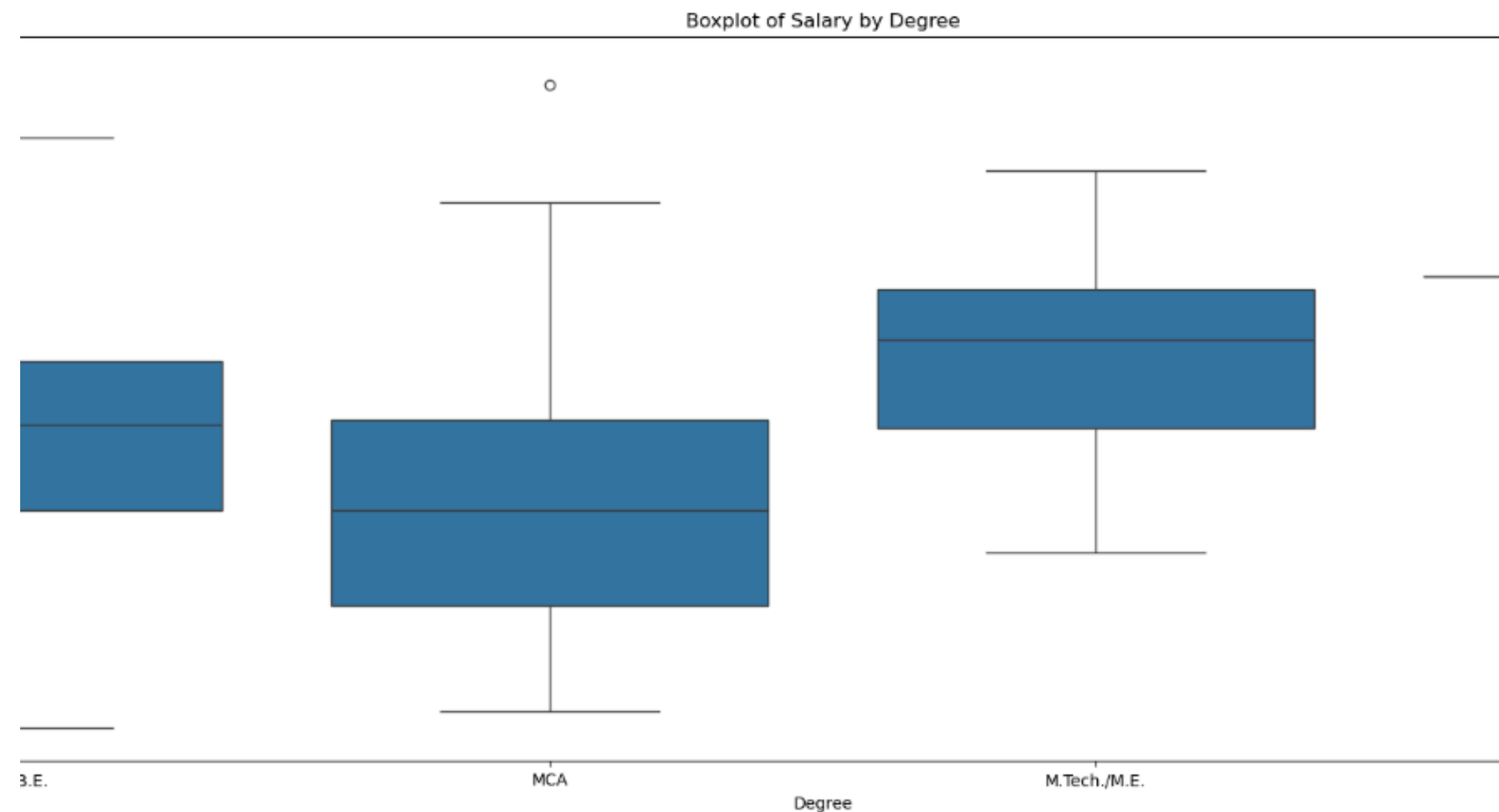


- Employees with Designation of senior software engineer have more salaries followed by system engineer.
- Employees with Designation of technical support engineer have less salaries.

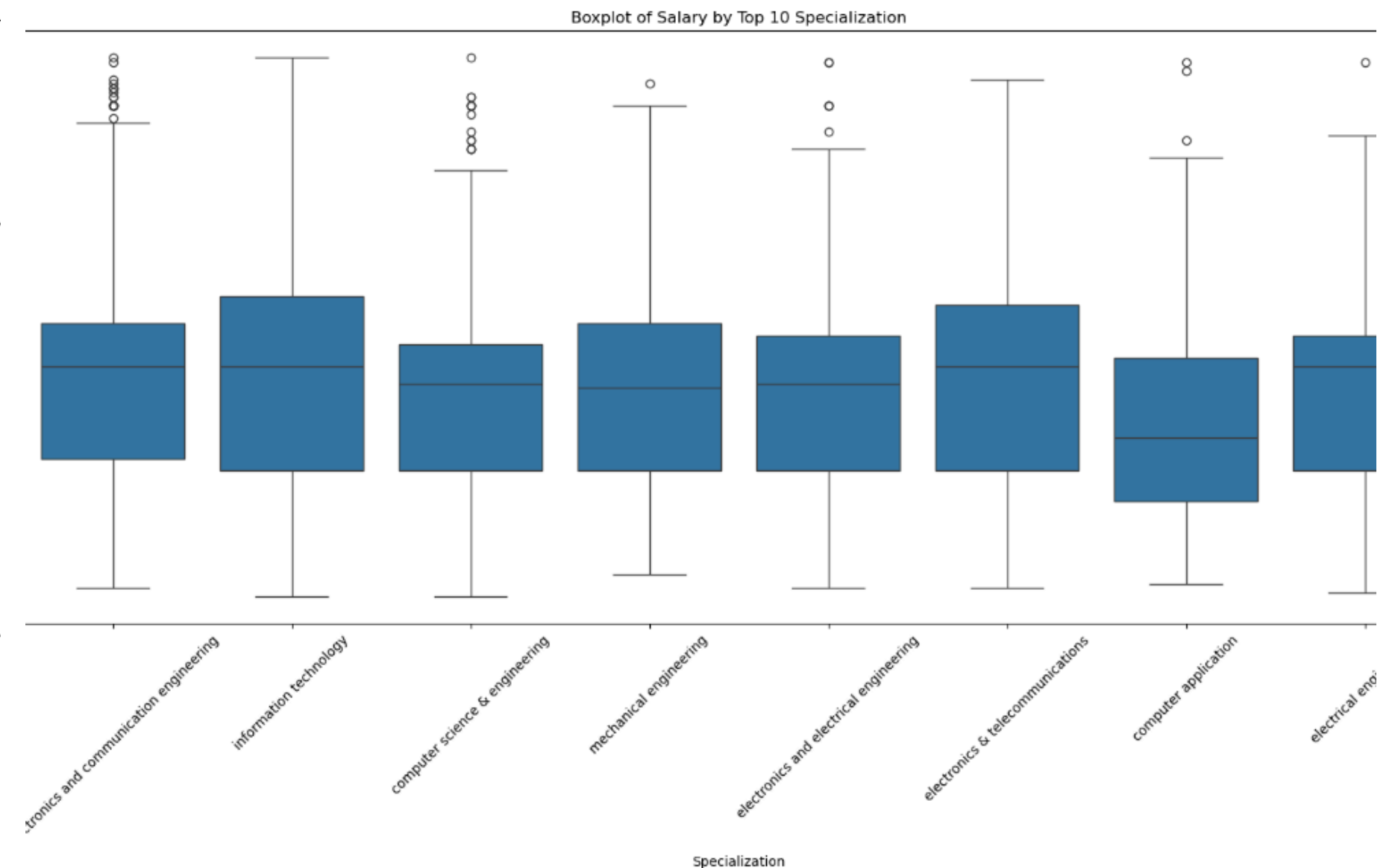


- Employees who are in Mumbai and Bangalore having highest salaries.
- Employees who are in New Delhi and kolkata have less salaries

Bivariate analysis (numerical vs numerical)



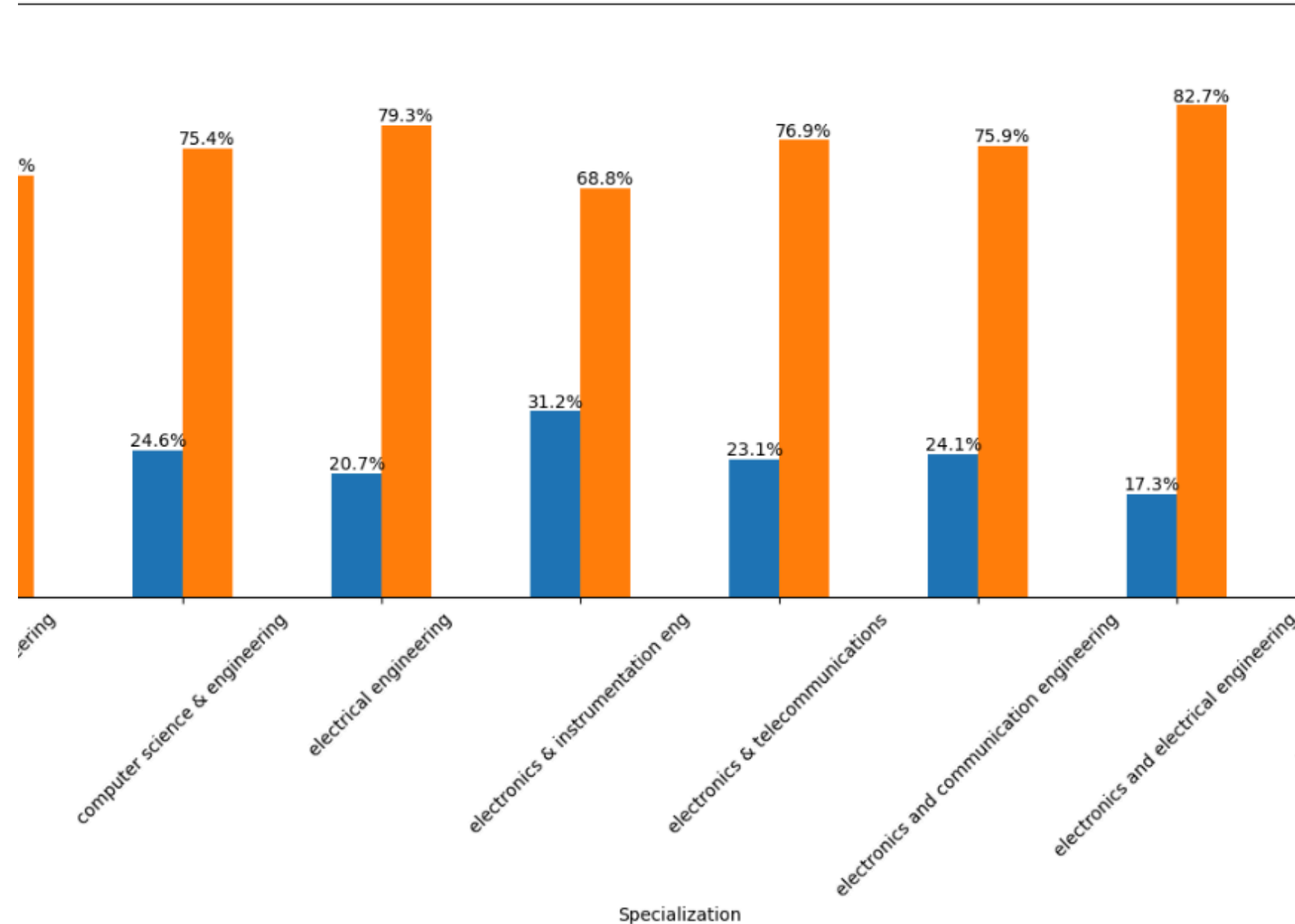
- Mtech/ME people having more salaries followed by Btech and MCA and MSc.
- MCA people have less salaries comparing to others.



- Employees with computer engineering specialization have more salaries.
- Employees with computer application specialization have less salaries

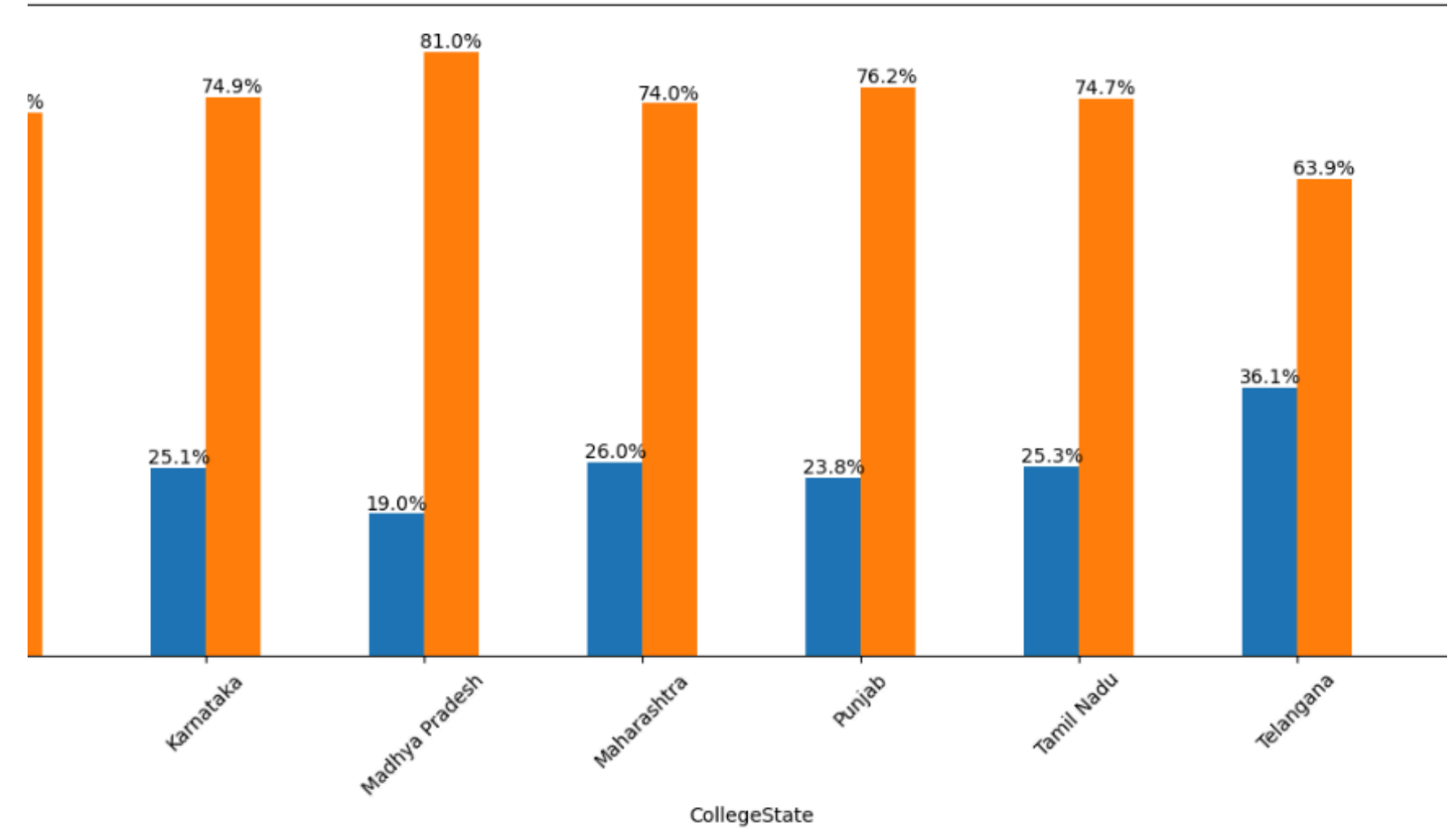
Bivariate analysis (categorical vs categorical)

Percentage of Persons by Specialization and Gender



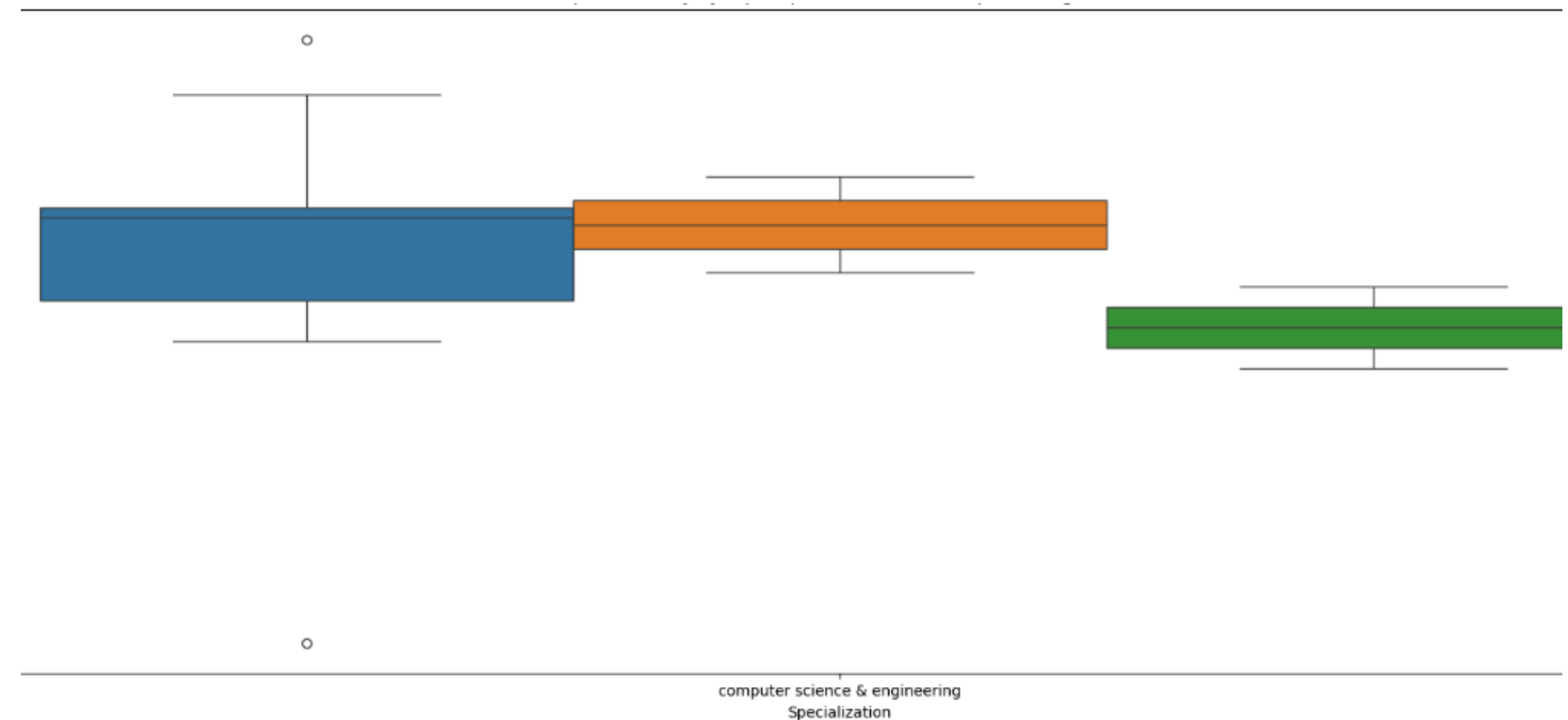
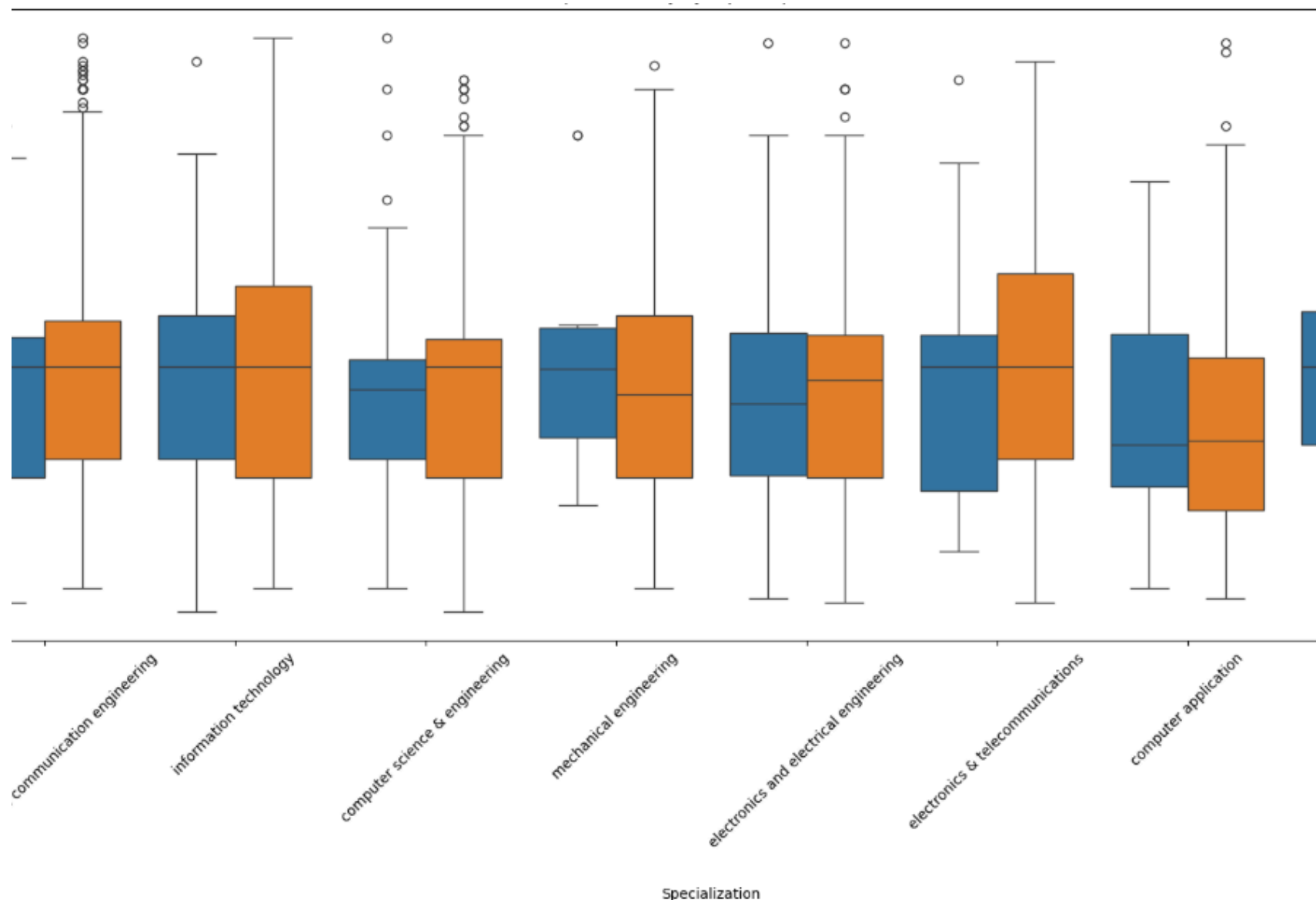
- There is a relationship between gender and specialization. Irrespective of specialization male candidates are more than female candidates.

Percentage of Persons by CollegeState and Gender



- There is a relationship between gender and CollegeState. Irrespective of specialization male candidates are more than female candidates,

Multivariate analysis



- There is a relationship between gender, Specialization and salary. Some specialization candidates belongs to male have more salary and some female candidates have more salary

- we have data from 2007 to 2014,so for specialization computer science and engineering, salary is between 2.8 to 3.5 lack for Programmer analyst, associate engineer and junior software engineer

CONCLUSION-1

Claim Testing: Can fresh graduates earn 2.5-3 lakhs in roles like Software Engineer, Hardware Engineer, or Associate Engineer?

- Conclusion: The data does not fully support the claim. Some roles like Software Engineer show slightly higher salary ranges than expected.



Gender and Specialization:

- Statistical analysis shows no significant relationship between gender and specialization preference ($p = 1.782e-10$).
- Specialization is independent of gender.

Degree and Salary Trends:

- Candidates with MTech/ME degrees tend to have the highest salaries, followed by BTech, MCA, and MSc graduates.

Gender and Salary:

- There is no significant difference in salary ranges between male and female candidates.

CONCLUSION-2

Specialization and Salary for Computer Science Graduates:

- For Computer Science and Engineering graduates from 2007 to 2014, the average salary for roles such as Programmer Analyst, Associate Engineer, and Junior Software Engineer ranges between 2.8 to 3.5 lakhs.

Job Roles:

- The majority of candidates (13.5%) are employed as Software Engineers, followed by System Engineers and Software Developers.

Job Locations:

- Bangalore is the top location for job placements, followed by Noida and Hyderabad.

Gender Employment Distribution:

- Male candidates dominate the workforce, making up 76.1% of employed candidates, while female candidates account for 23.9%.

Specialization Trends:

- Most students have a degree in Electronics and Communication Engineering (22%) followed by Computer Science and Engineering (18.6%), and Information Technology (16.5%).

THANK YOU

