INNOMATICS®
RESEARCH LABS

INNOVATION. AUTOMATION. ANALYTICS

PROJECT ON

Exploratory Data Analysis on AMEO Data

Done by – Gayathri Salapu

About Me

Hi there, My name is Gayathri Salapu currently enrolled as an Intern in Data Science course at 'INNOMATICS RESEARCH LABS'. I came from Electrical background wanted to explore Data science.

Iam actively seeking oppurtunities in data science and transform my career in to this field. This internship is a huge opportunity to achieve my dream in carrer transition. This provides me problem solving skills, hands on experience to apply my analytical skills and apply my engineering mind to this world.

My goal is not just to explore but to fully immerse myself in the challenges and innovations that data science presents. I am committed to acquiring new skills, embracing change, and carving a niche for myself in a completely different field. This journey signifies my dedication to growth, learning, and the pursuit of a career that aligns with my evolving interests and aspirations.

Project Objective :

This analysis seeks to delve into the provided dataset, with a focus on unraveling the nuanced relationships between different features and the designated target variable—Salary. The main goals include:

Thoroughly understanding the dataset and its individual features. Detecting inherent patterns or trends within the data.

Exploring the dynamic interactions between independent variables and the target variable (Salary). Spotting any outliers or anomalies that may be present in the dataset.

Data Description :

The Aspiring Mind Employment Outcome 2015 (AMEO) dataset, released by Aspiring Minds, focuses on employment outcomes of engineering graduates. Specifically tailored to students within engineering disciplines, the dataset encompasses information about salary, job titles, and job locations, serving as dependent variables. Additionally, it includes standardized scores in cognitive skills, technical skills, and personality skills.

Comprising approximately 40 independent variables and 4000 data points, these independent variables exhibit a mix of continuous and categorical characteristics. The dataset also incorporates demographic features, contributing to a comprehensive understanding of the factors influencing the employment outcomes of engineering graduates. Each candidate in the dataset is uniquely identified for reference and analysis purposes.

INNOMATICS RESEARCH LABS

Data Cleaning and Preprocessing

➢ Datatypes Conversion:

To ensure precision and uniformity in our analysis, a crucial step involved converting the data types of 'Date of Joining' (DOJ) and 'Date of Leaving' (DOL) fields from their original format to datetime objects.
Considering the survey's 2015 date, respondents indicating 'present' for DOL were assumed to have left the company by the latest survey date (2015-12-31).

➢ Handling Null Values:

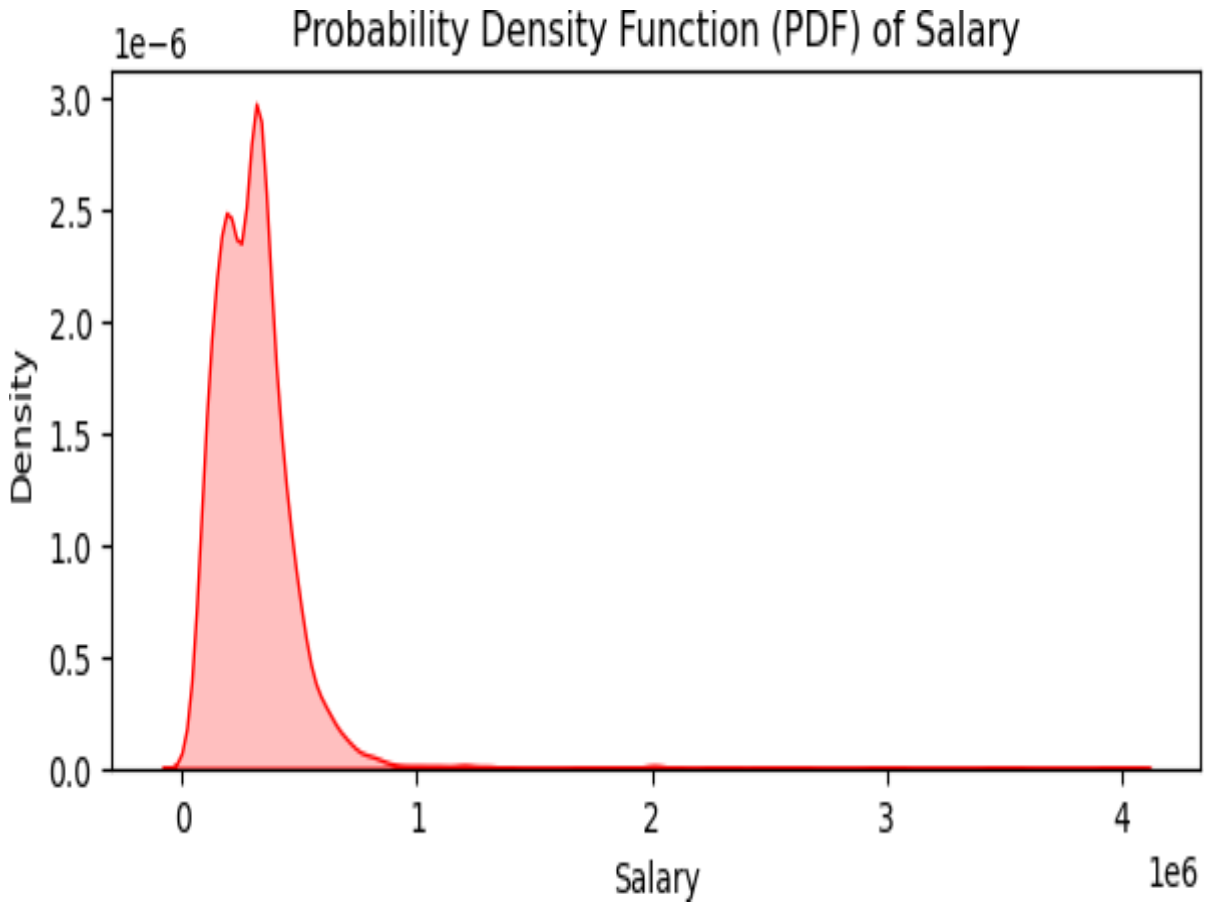
In the initial data cleaning phase, particular attention was given to managing null values represented by 0 or -1. Columns such as '10board,' '12board,' 'GraduationYear,' 'JobCity,' and 'Domain' underwent careful processing. Subsequently, columns with over 80% -1 values, such as 'MechanicalEngg,' 'ElectricalEngg,' 'TelecomEngg,' and 'CivilEngg,' were excluded from further analysis. Optional subject columns, namely 'ElectronicsAndSemicon' and 'ComputerScience,' had -1 values replaced with 0 to denote non-pursuit.

➢ Category Simplification:

As a concluding step, the dataset underwent refinement by retaining only the top 10 most frequently occurring categories within specific columns. Categories beyond this selection were amalgamated into an 'other' category, streamlining the dataset and facilitating a more focused and insightful analysis.

INNOMATICS RESEARCH LABS

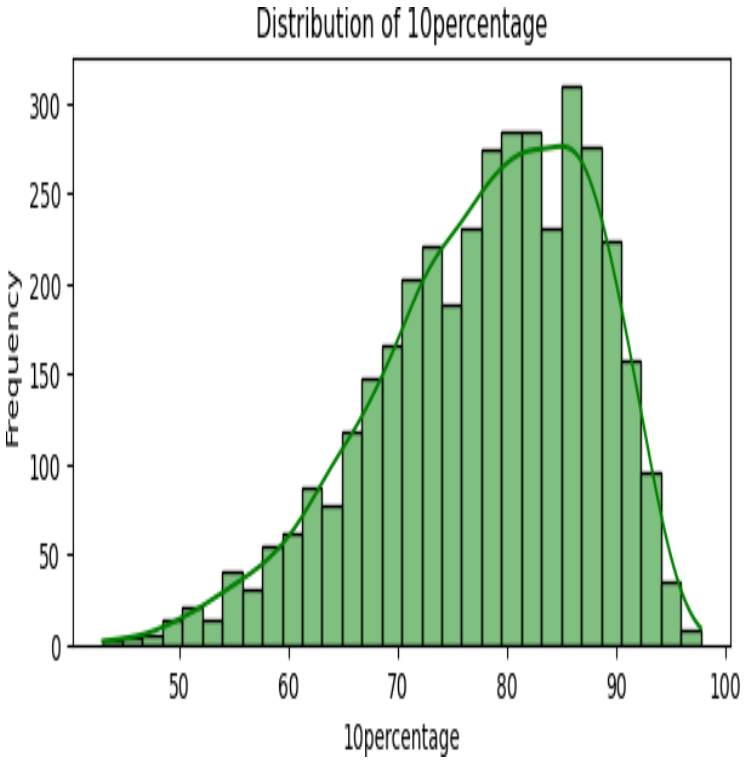
Univariate Analysis on 'Salary'

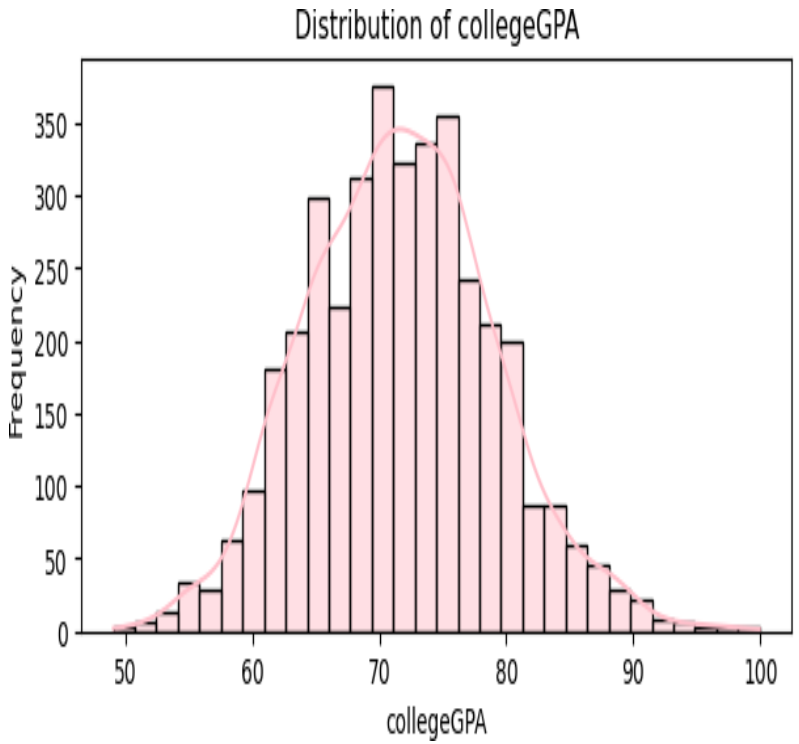


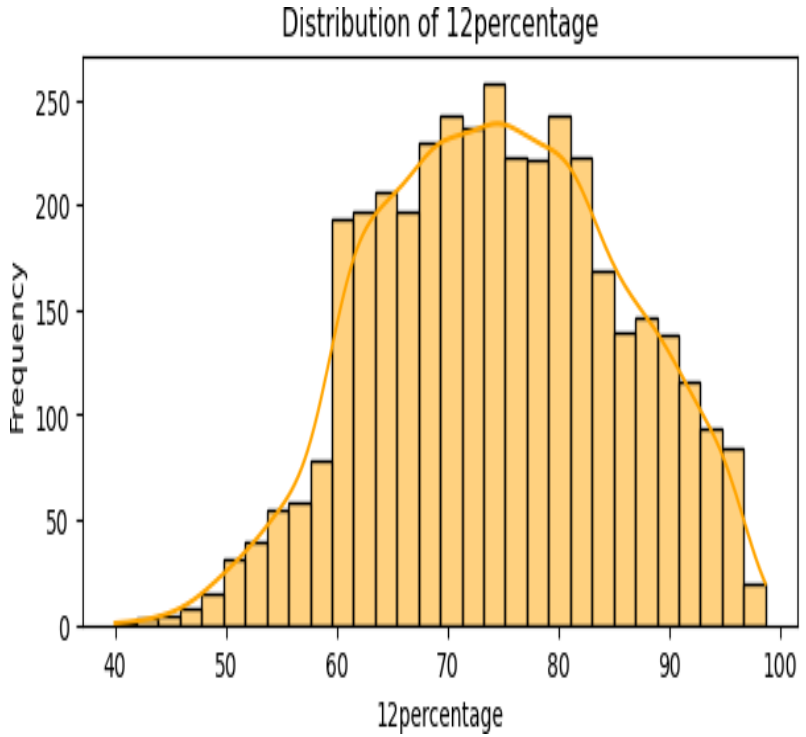
➢ **Conclusion:** *From the above graph, we observe the presence of an outlier where salaries exceeding 10,00,000 are rare, particularly in the first job.*

INNOMATICS RESEARCH LABS

Univariate Analysis on Numerical columns

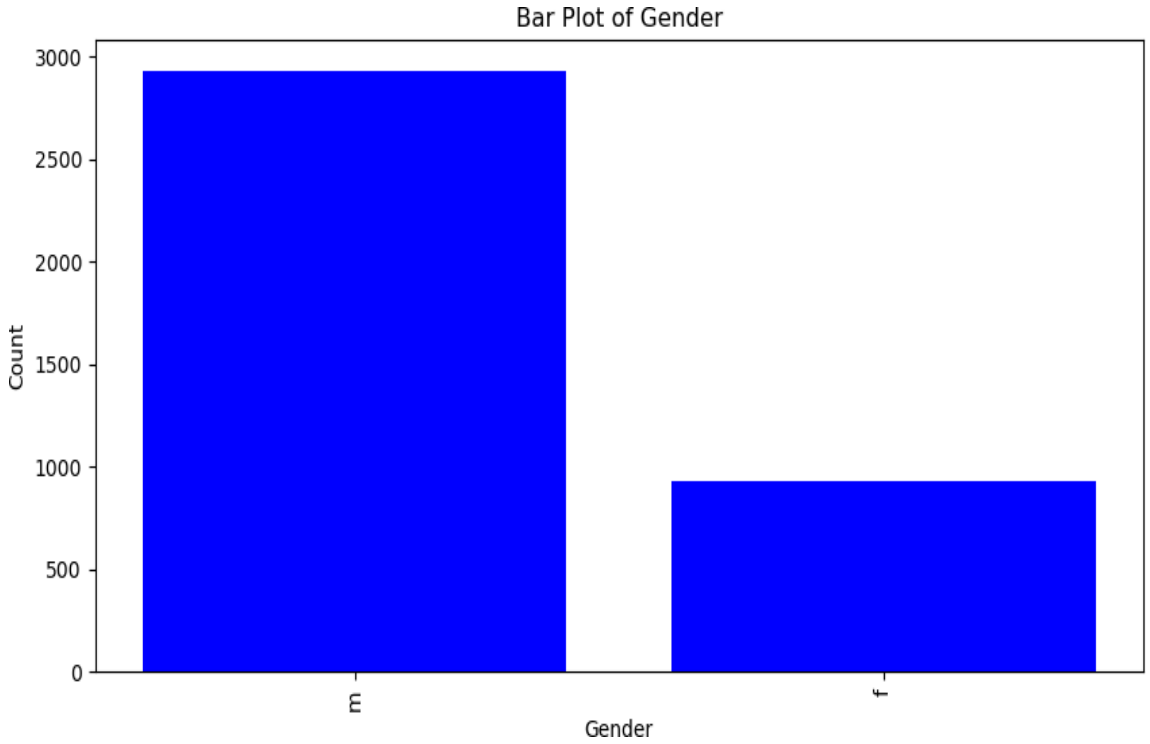


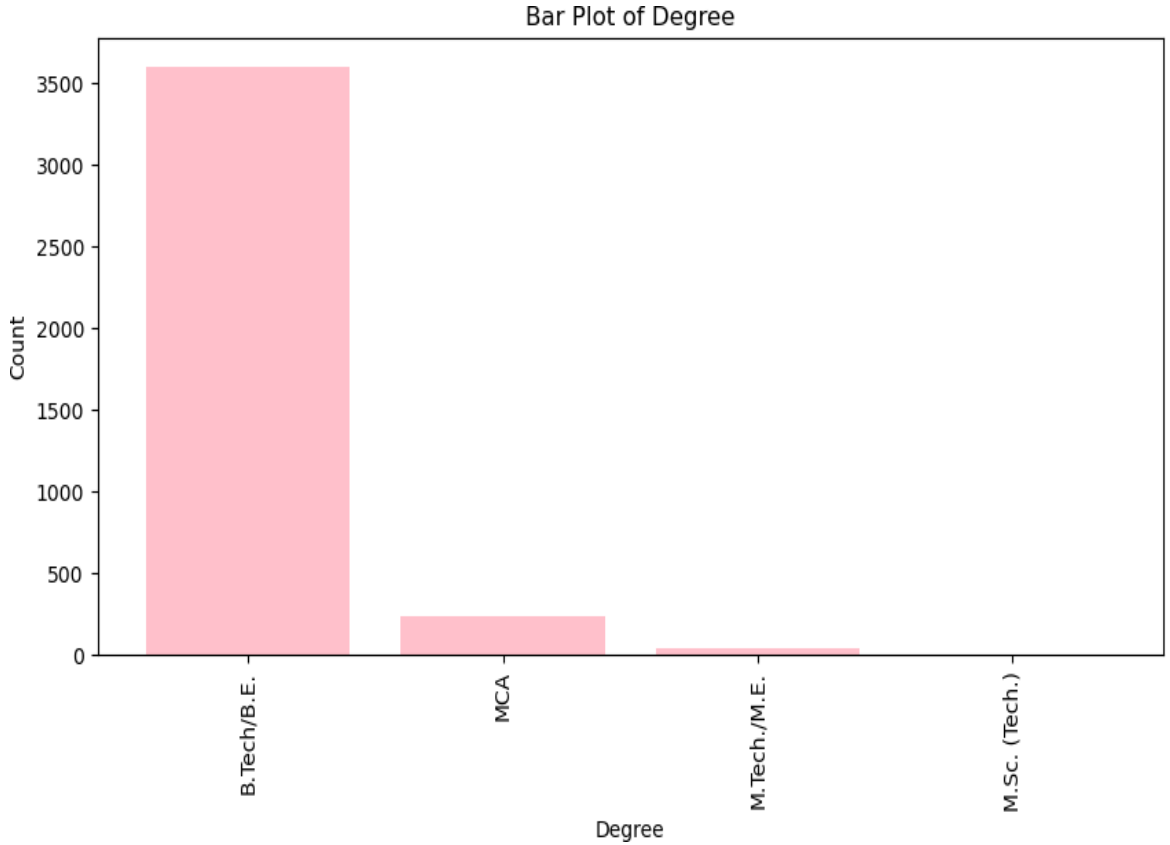




- **Conclusion:** *I conducted a detailed analysis on three academic metrics—10th percentage, 12th percentage, and college CGPA. Using graphs, I visually depicted their distribution patterns, revealing insights into students' academic achievements and potential employability. The identified trends are crucial indicators, enhancing our understanding of the relationship between academic performance and employability.*

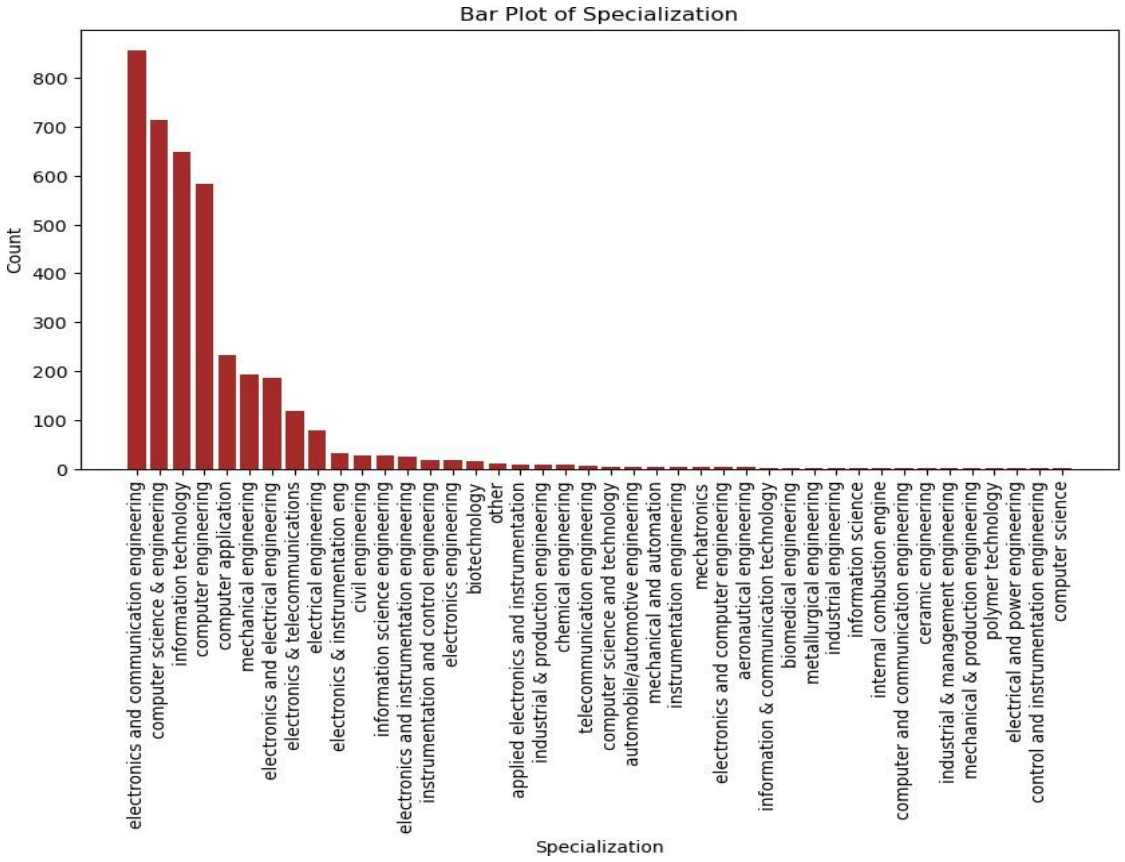
Univariate Visual Analysis on ‘Gender’, ‘Degree’

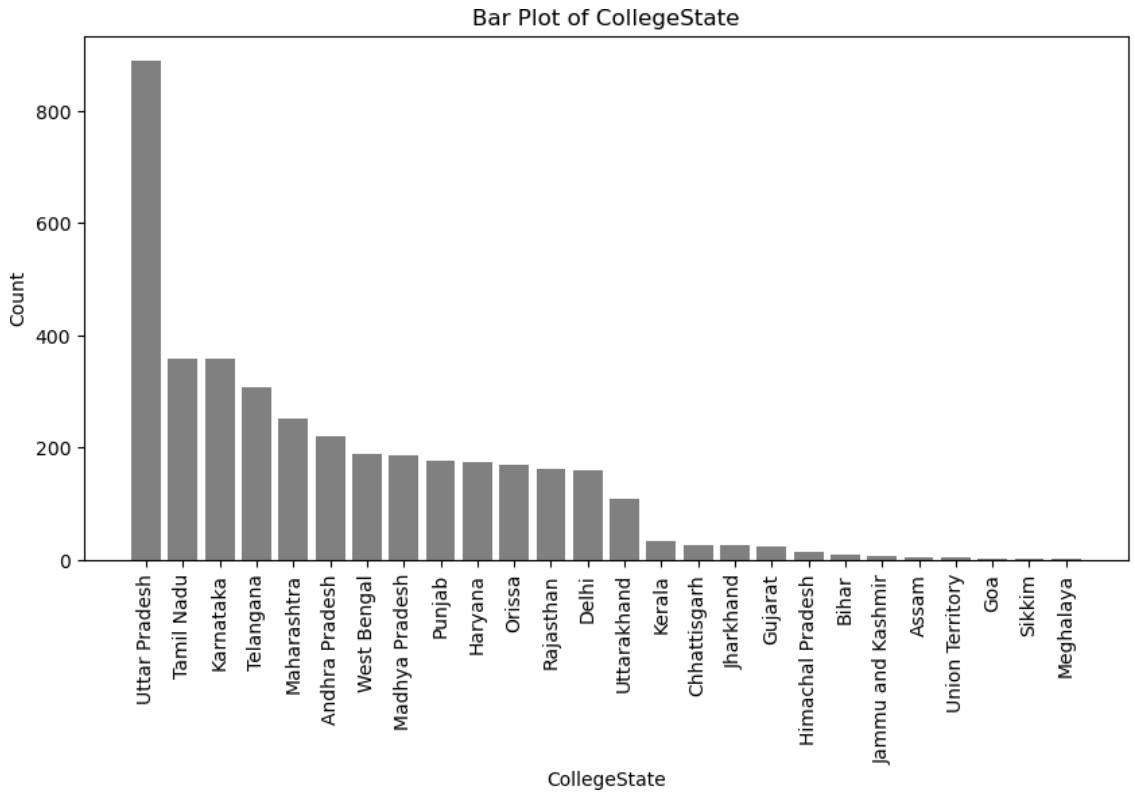




- **Conclusion:** *The data suggests a higher representation of males compared to females, and a prevalence of individuals with B.Tech/B.E degrees over other academic qualifications.*

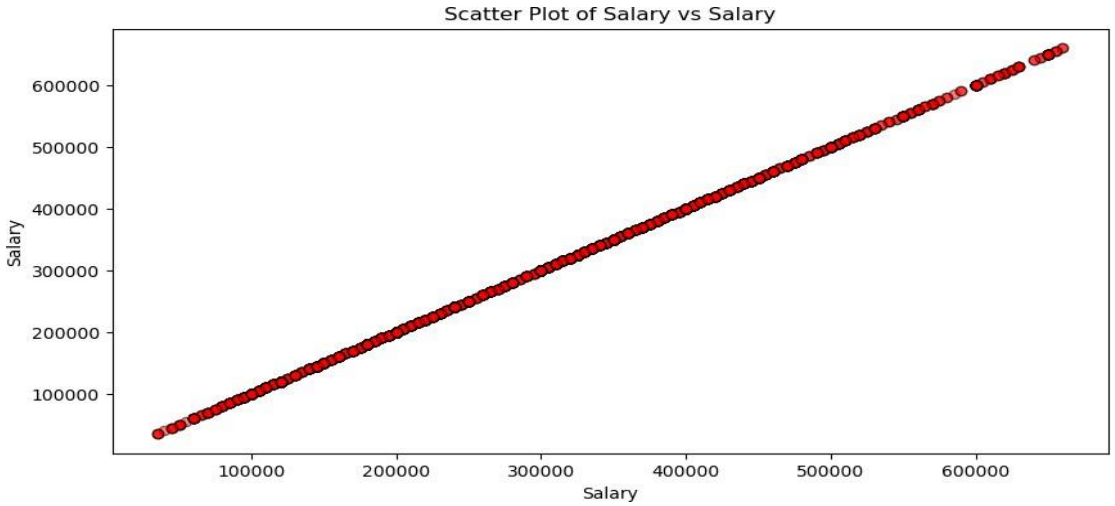
Univariate Analysis on ‘Specialization’ & ‘College State’

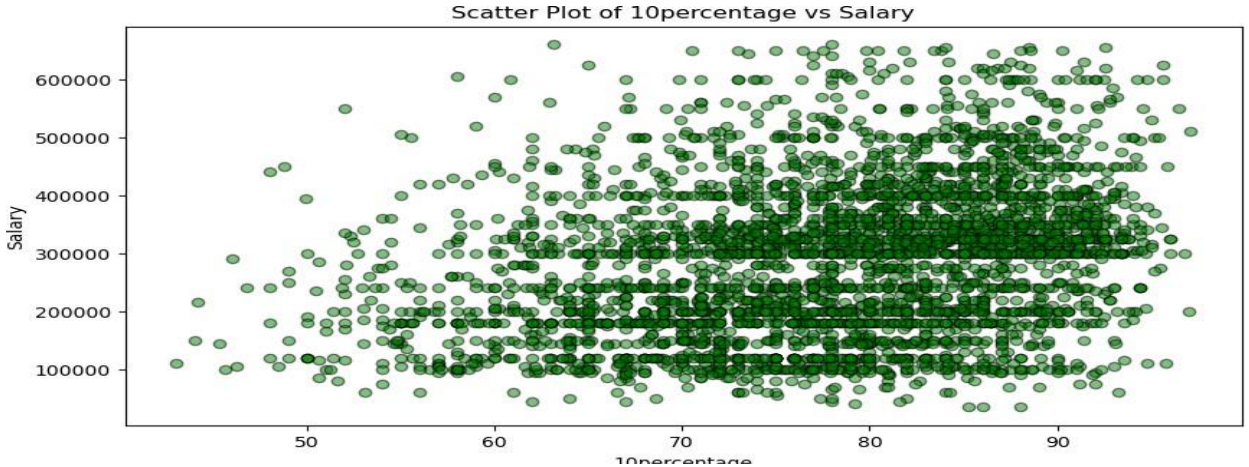


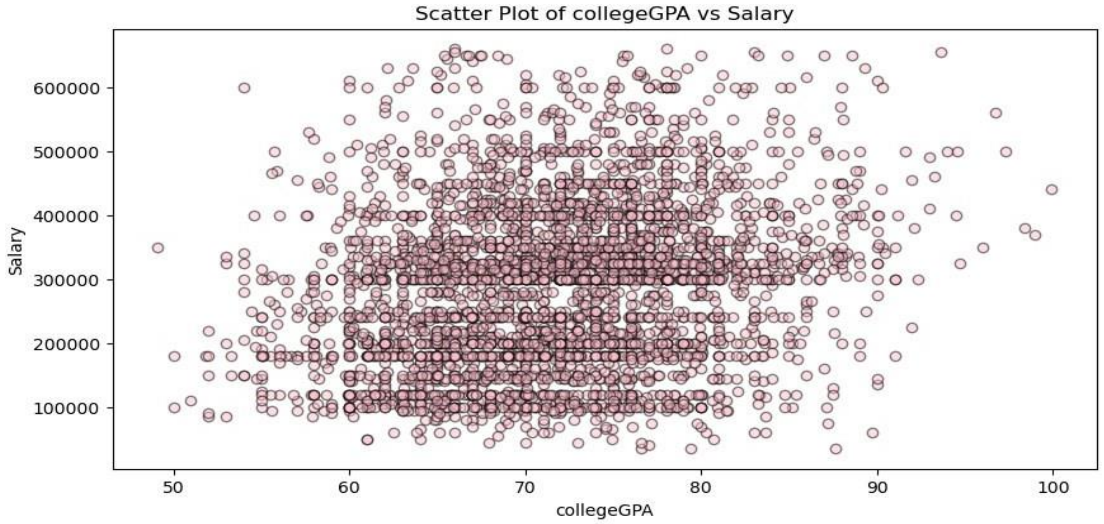


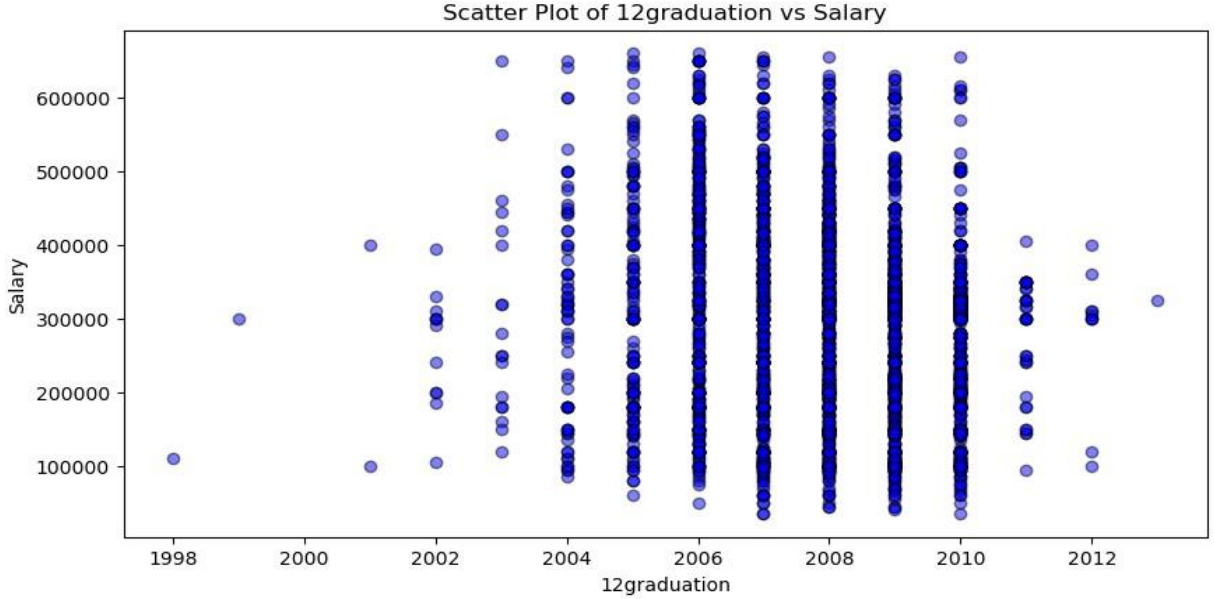
- **Conclusion:** *Here, it is evident that the most common Specialization is Electronics Communication engineering. And also clear that the highest College State count is Uttar Pradesh.*

Bivariate Analysis between 'Salary' & Other columns



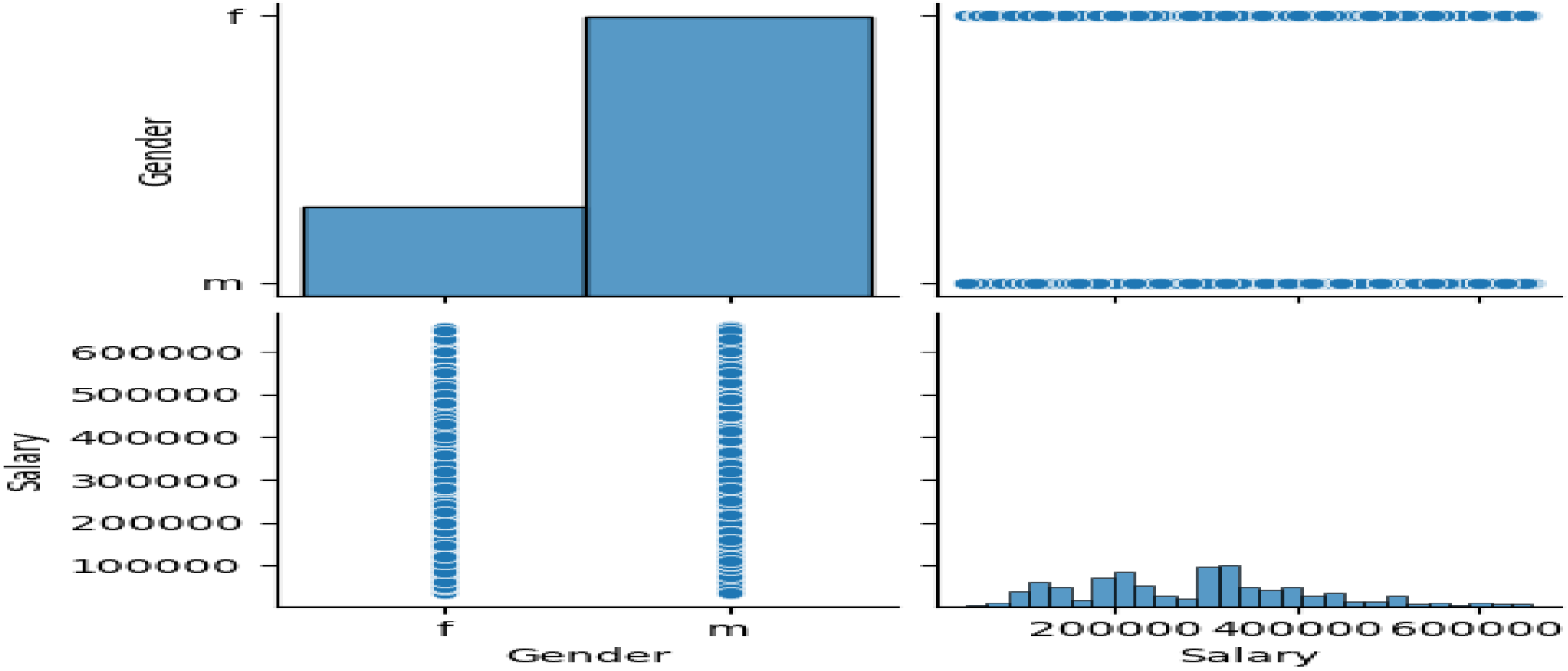






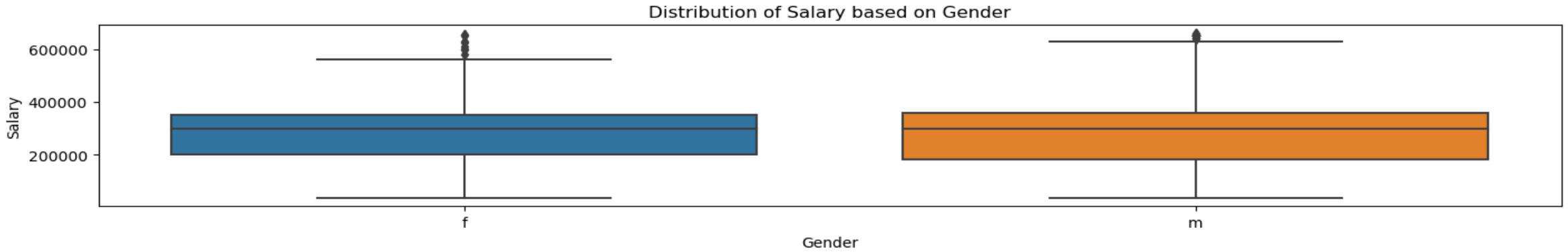
INNOMATICS RESEARCH LABS

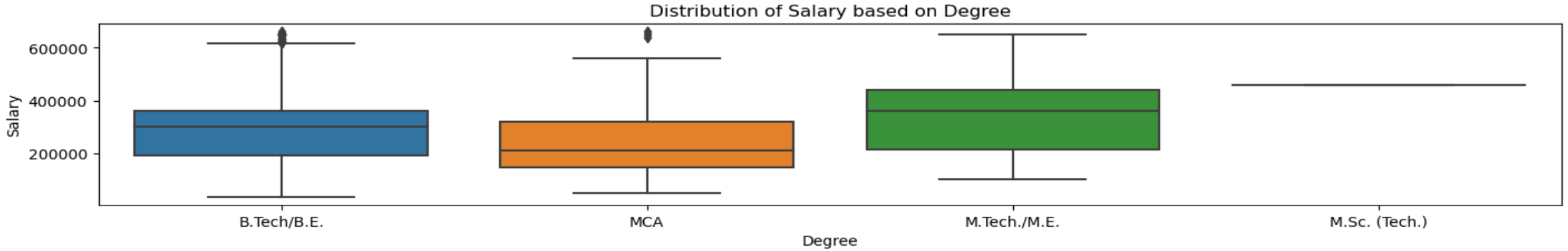
Bivariate Analysis between 'Salary' & 'Gender'

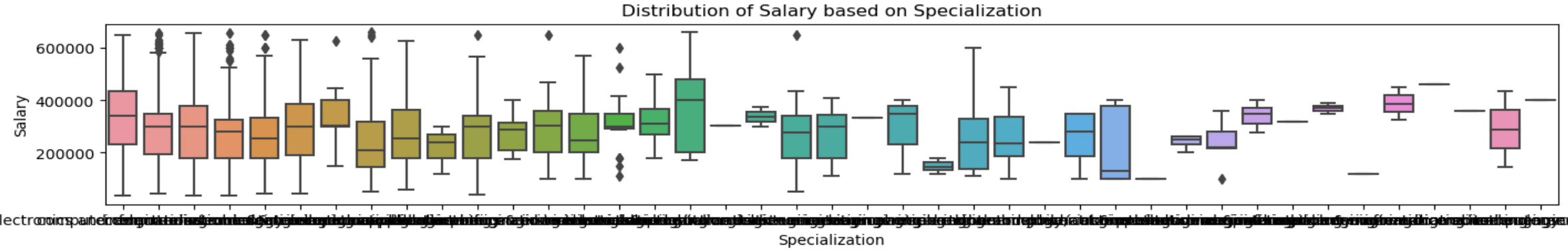


➢ **Conclusion:** *Here we can see Males and females take the same salary*.

Bivariate Analysis between 'Salary' with other columns

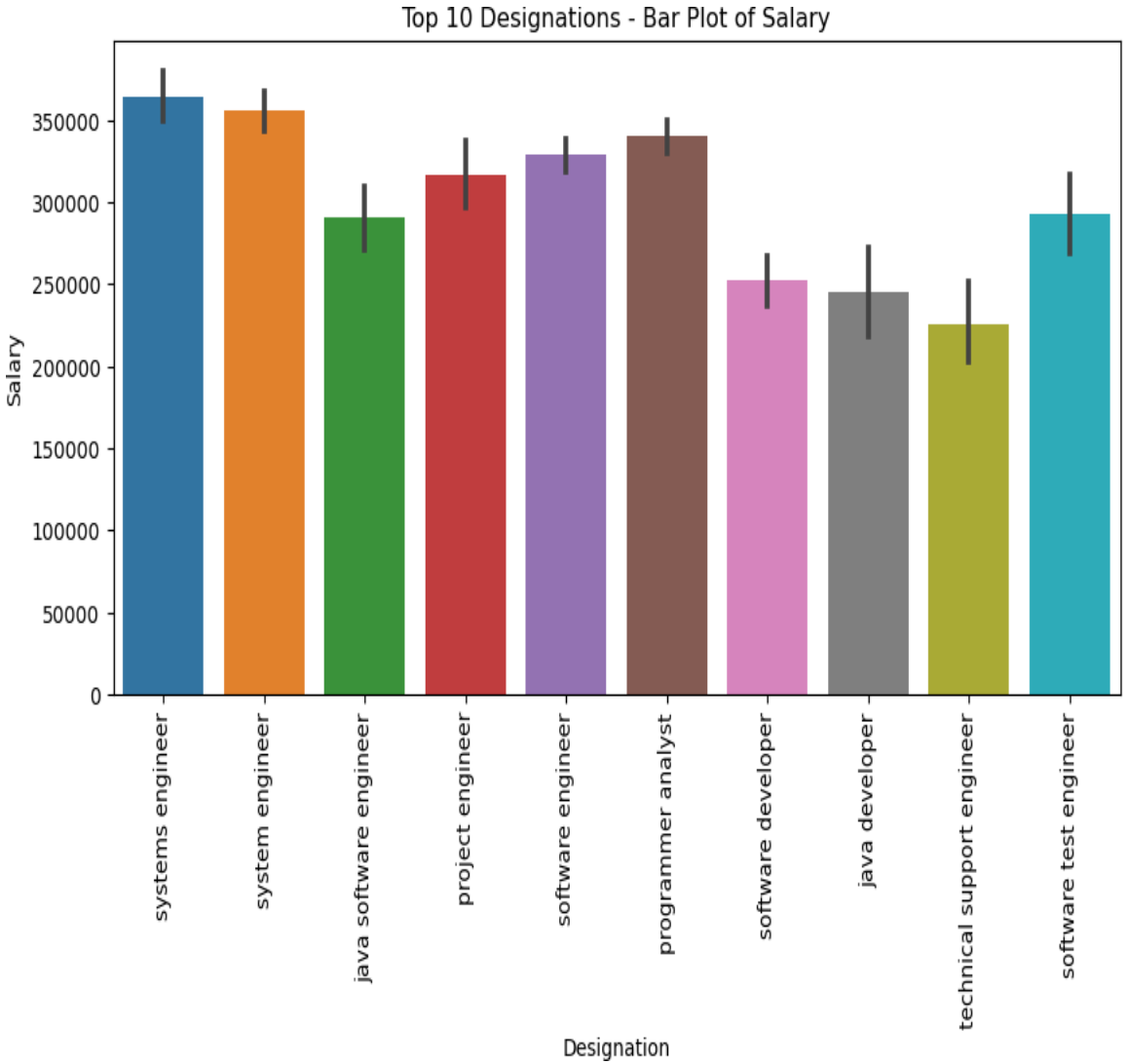


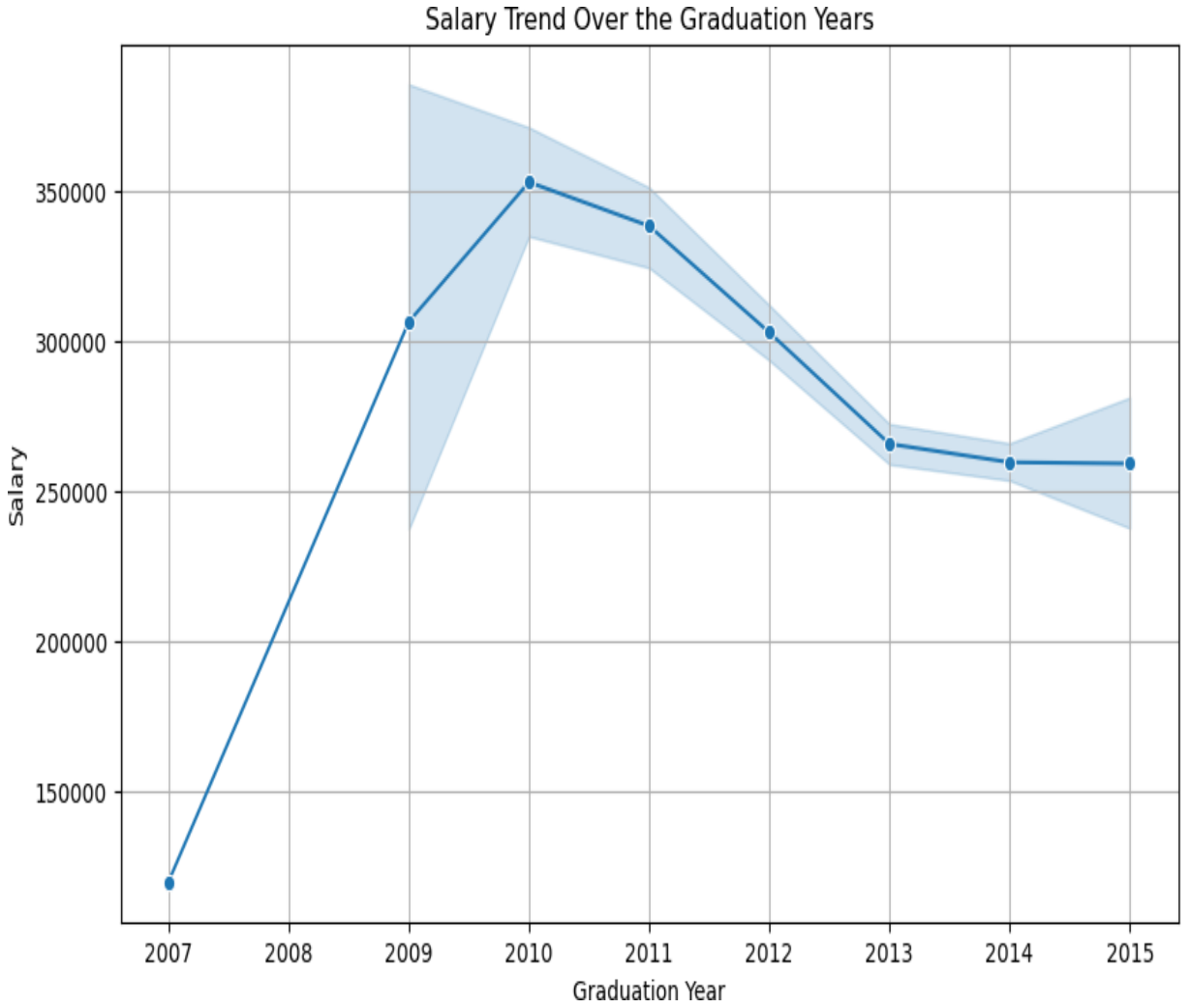




INNOMATICS RESEARCH LABS

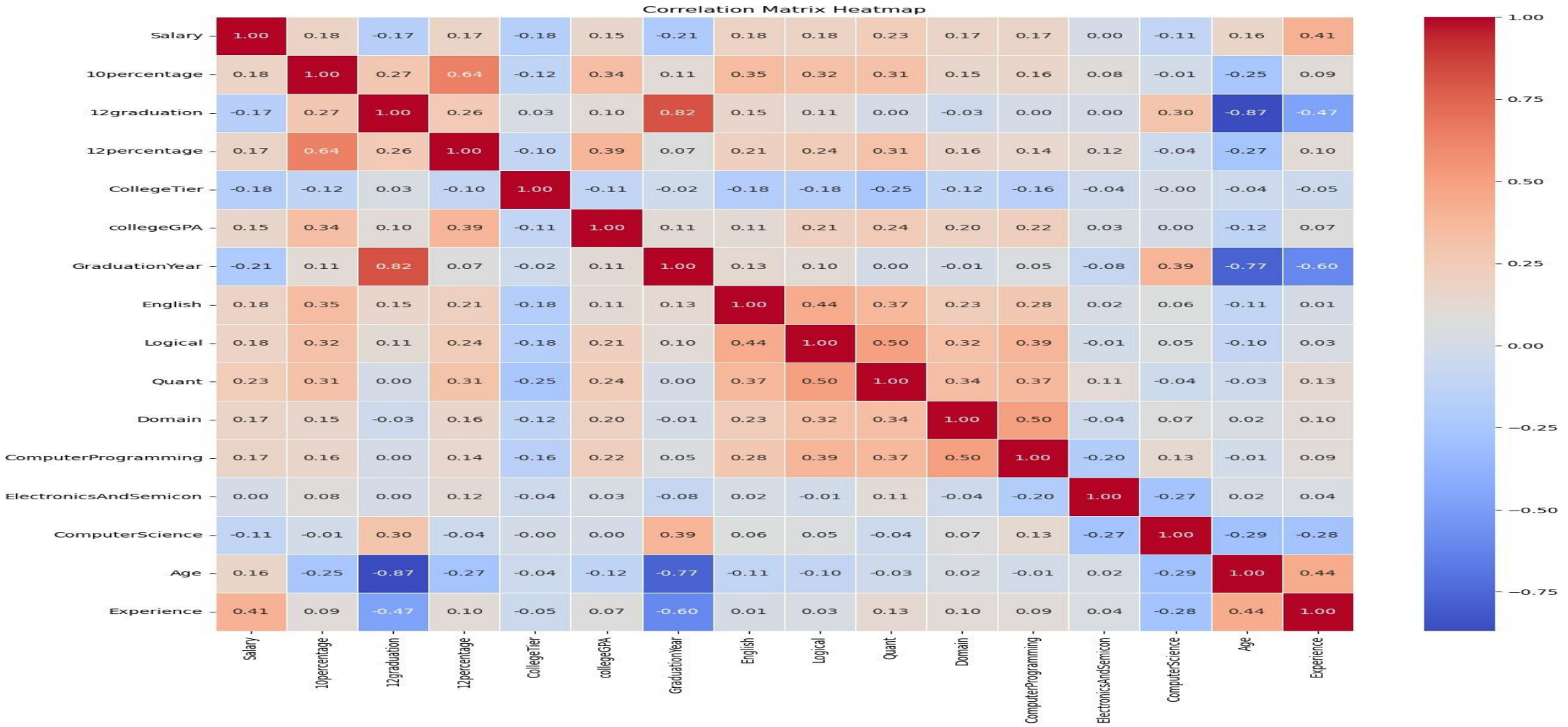
Bivariate Analysis between 'Salary' & other columns





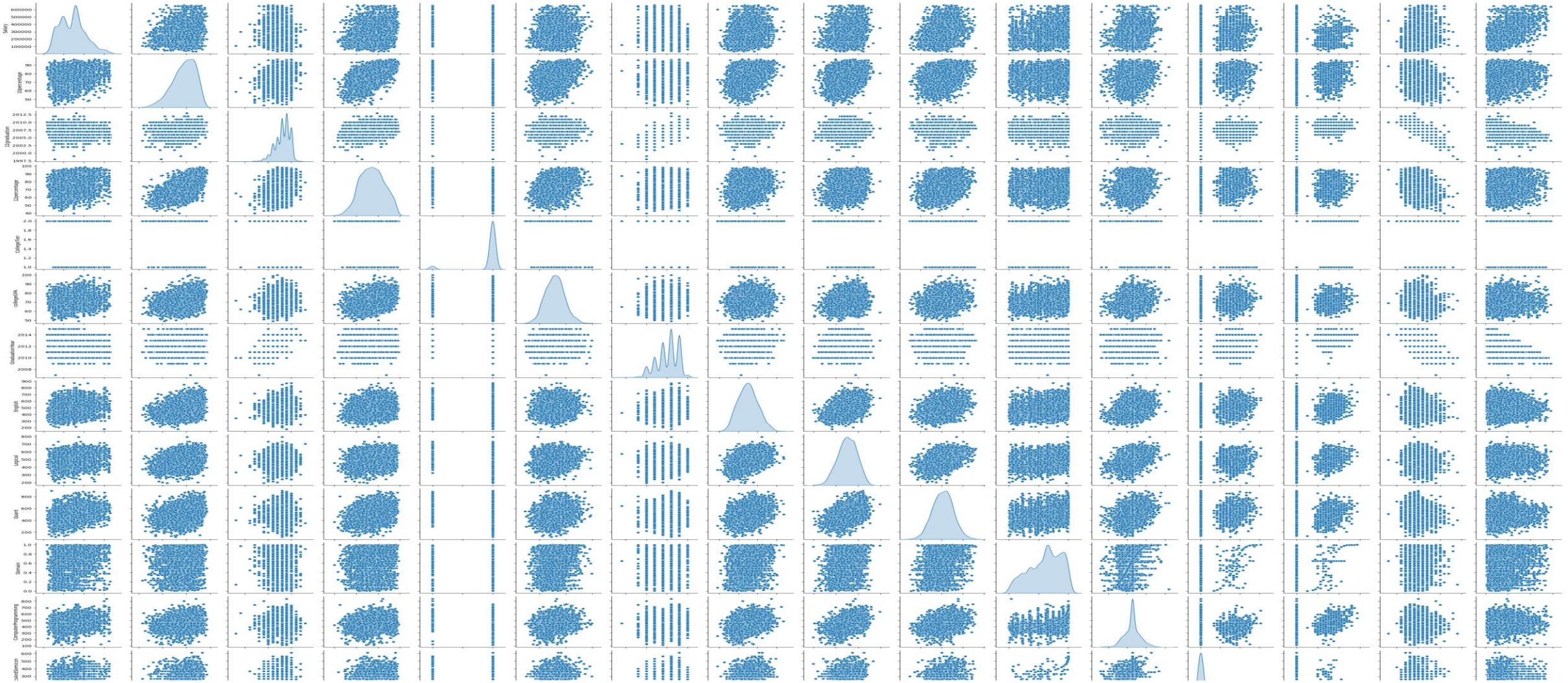
INNOMATICS RESEARCH LABS

Correlation Heatmap



➢ **Conclusion:** *Heat map show the correlation ship between all the numerical features*.

Pair Plot



- **Conclusion:** *Pair Plot show relation ship between all the features*.

INNOMATICS RESEARCH LABS

OBSERVATIONS

- *Graduates with a BTech/B.E or MCA degree experienced the highest salaries and job placements during the period between 2010 and 2014.*
- *Among branches, candidates from Computer Science (CS), Information Technology (IT), and Electronics and Telecommunication (ETC) received the highest salaries.*
- *Remarkably, even candidates with low to intermediate scores in the AMCAT exam secured lucrative salary packages.*
- *Occupations such as Software Engineer, Software Test Engineer, System Engineer, and Software Developer consistently commanded the highest salary packages.*
- *While Delhi and Pune offered the highest salaries, Bangalore, Noida, Hyderabad, and Pune provided a plethora of job opportunities with varying salary ranges.*
- *The year 2014 witnessed the highest number of job placements.*

INNOMATICS RESEARCH LABS

INNOMATICS
RESEARCH LABS