INNOMATICS®
RESEARCH LABS

INNOVATION. AUTOMATION. ANALYTICS

PROJECT ON

Exploratory Data Analysis of AMCAT Data

By
Divya Sri

INNOMATICS
RESEARCH LABS

About Me

I am Divya Sri who pursued B Tech in Electrical Engineering.

Data Science brings the opportunity to develop a broad and diverse skill set that can be applied in many different contexts. I enjoy problem solving and statistical analysis. I have a passion for using data to draw meaningful conclusions which led to career in Data Science.

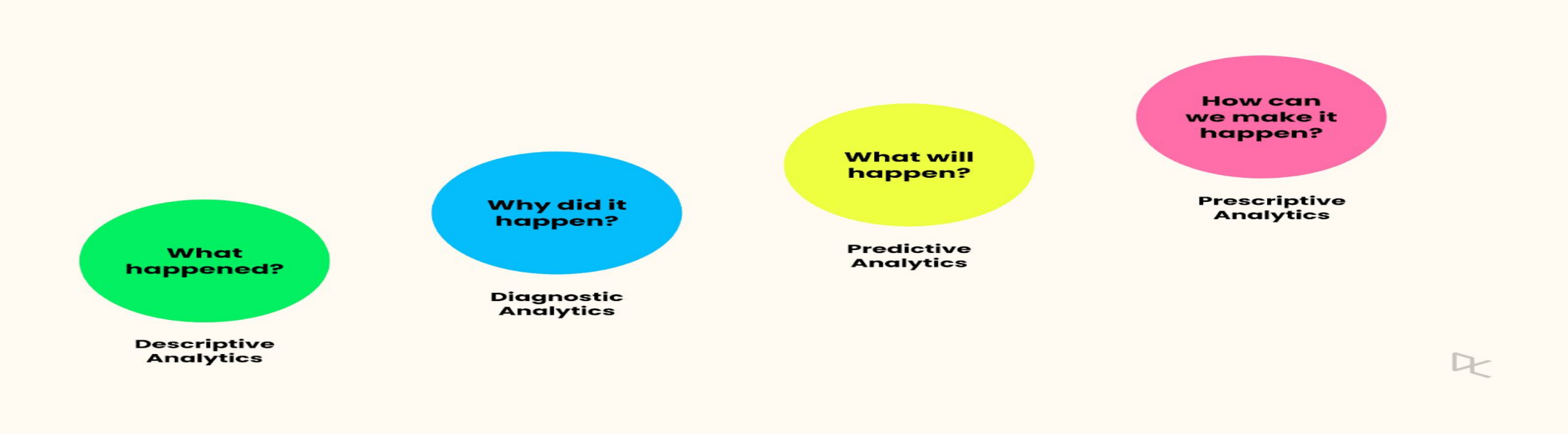
I have work experience for more than two years.

Data Analysis

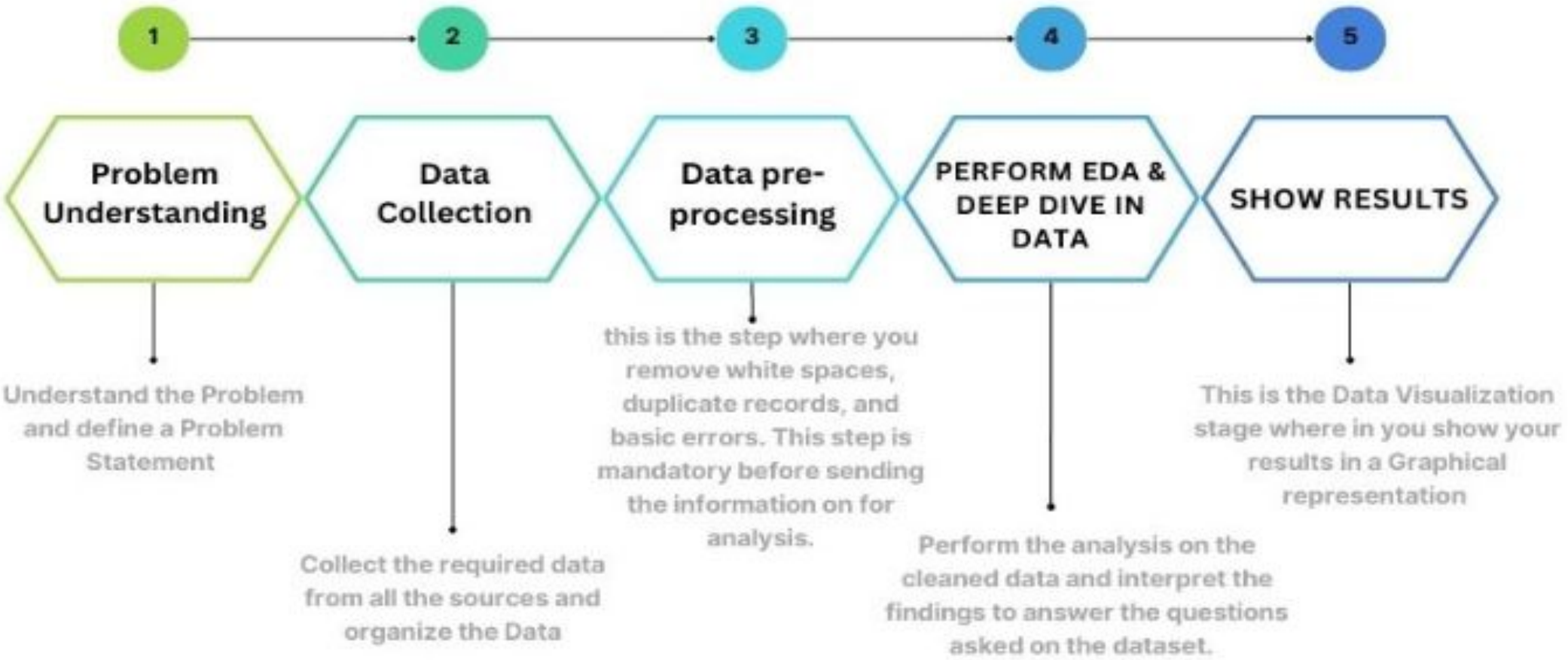
.It is the practice of working with data to glean useful information, which can then be used to make informed decisions.

Types of Data Analysis

Data analysis can be categorized into four main types, each serving a unique purpose and providing different insights. These are descriptive, diagnostic, predictive, and prescriptive analysis.



Data Analysis Steps



INNOMATICS RESEARCH LABS

AMCAT

AMCAT known as Aspiring Minds Computer Adaptive Test is an AI-based computer adaptive test which evaluates job applicants on critical areas like communication skills, logical reasoning, quantitative skills, and job-specific domain skills thereby helping recruiters identify the suitability of a candidate for different job roles.

Business Question

- Times of India article dated Jan 18, 2019 states that "*After doing your Computer Science Engineering if you take up jobs as a Programming Analyst, Software Engineer, Hardware Engineer and Associate Engineer you can earn up to 2.5-3 lakhs as a fresh graduate.*" Test this claim with the data given to you.

- Is there a relationship between gender and specialization? (i.e. Does the preference of Specialisation depend on the Gender?)

Objective of the Project

Univariate and Bivariate Analysis of Variables

Answers and Conclusions to relevant hypothesis questions.

Summary of the Data

```
# Displaying the top five rows of the datset
df.head()
```

	Unnamed: 0	ID	Salary	DOJ	DOL	Designation	JobCity	Gender	DOB	10percentage	...	ComputerScience	MechanicalEngg	E
0	train	203097	420000.0	6/1/12 0:00	present	senior quality engineer	Bangalore	f	2/19/90 0:00	84.3	...	-1	-1	
1	train	579905	500000.0	9/1/13 0:00	present	assistant manager	Indore	m	10/4/89 0:00	85.4	...	-1	-1	
2	train	810601	325000.0	6/1/14 0:00	present	systems engineer	Chennai	f	8/3/92 0:00	85.0	...	-1	-1	
3	train	267447	1100000.0	7/1/11 0:00	present	senior software engineer	Gurgaon	m	12/5/89 0:00	85.6	...	-1	-1	
4	train	343523	200000.0	3/1/14 0:00	3/1/15 0:00	get	Manesar	m	2/27/91 0:00	78.0	...	-1	-1	

5 rows × 39 columns

Data Transformation

- The DOJ and DOB columns needs to be converted from object to the date type.
- The DOL column though it contains date values would be left in the object type since it contains 'present' string values which indicates that the candidate still works at a company.
- The 'Unnamed: 0','CollegID','CollegeCityID' column appears to be irrelevant for this exploratory data analysis, and hence would need to be removed or dropped.
- There appear to be no null values (na) in any columns; however, some columns contain -1 and other negative values, which indicates that these values are not available and needs to be replaced with NaN instead.

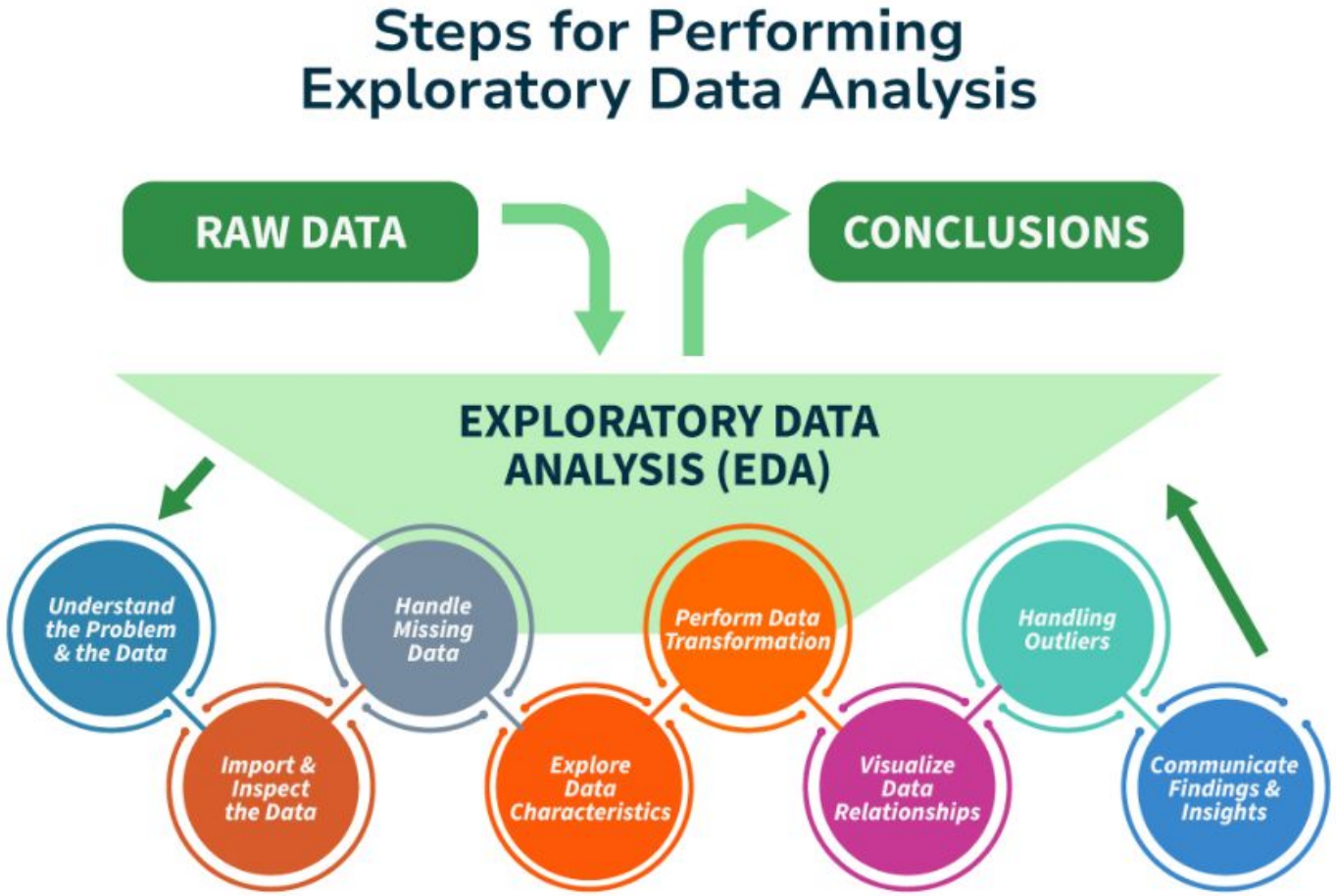
```
# Checking the columns characteristics
df.info()
```

```
<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   float64
 3    DOJ                       3998 non-null   object
 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   object
 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   Colle[No Title]           3998 non-null   int64
 20   Colle                     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   openess_to_experience     3998 non-null   float64
dtypes: float64(10), int64(17), object(12)
memory usage: 1.2+ MB
```

INNOMATICS RESEARCH LABS

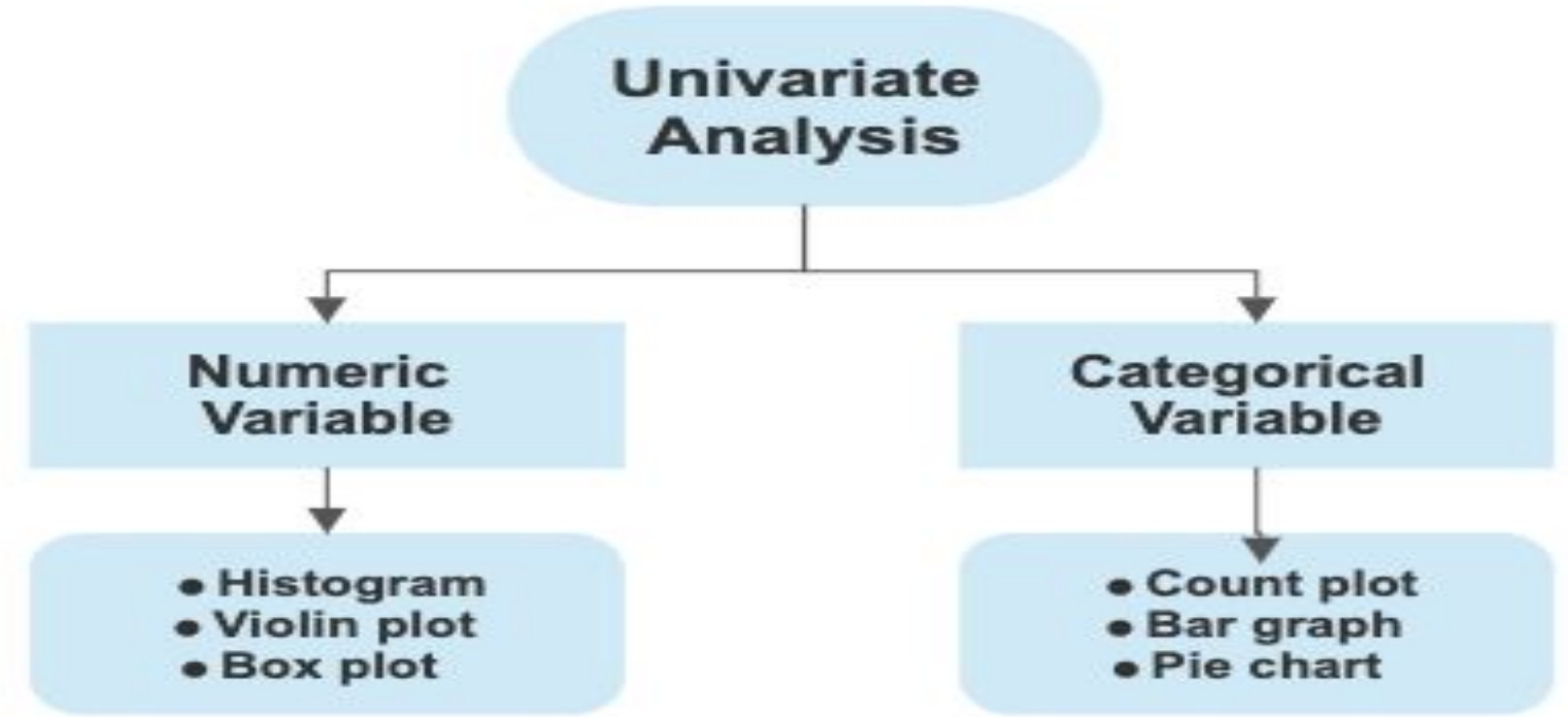
Exploratory Data Analysis

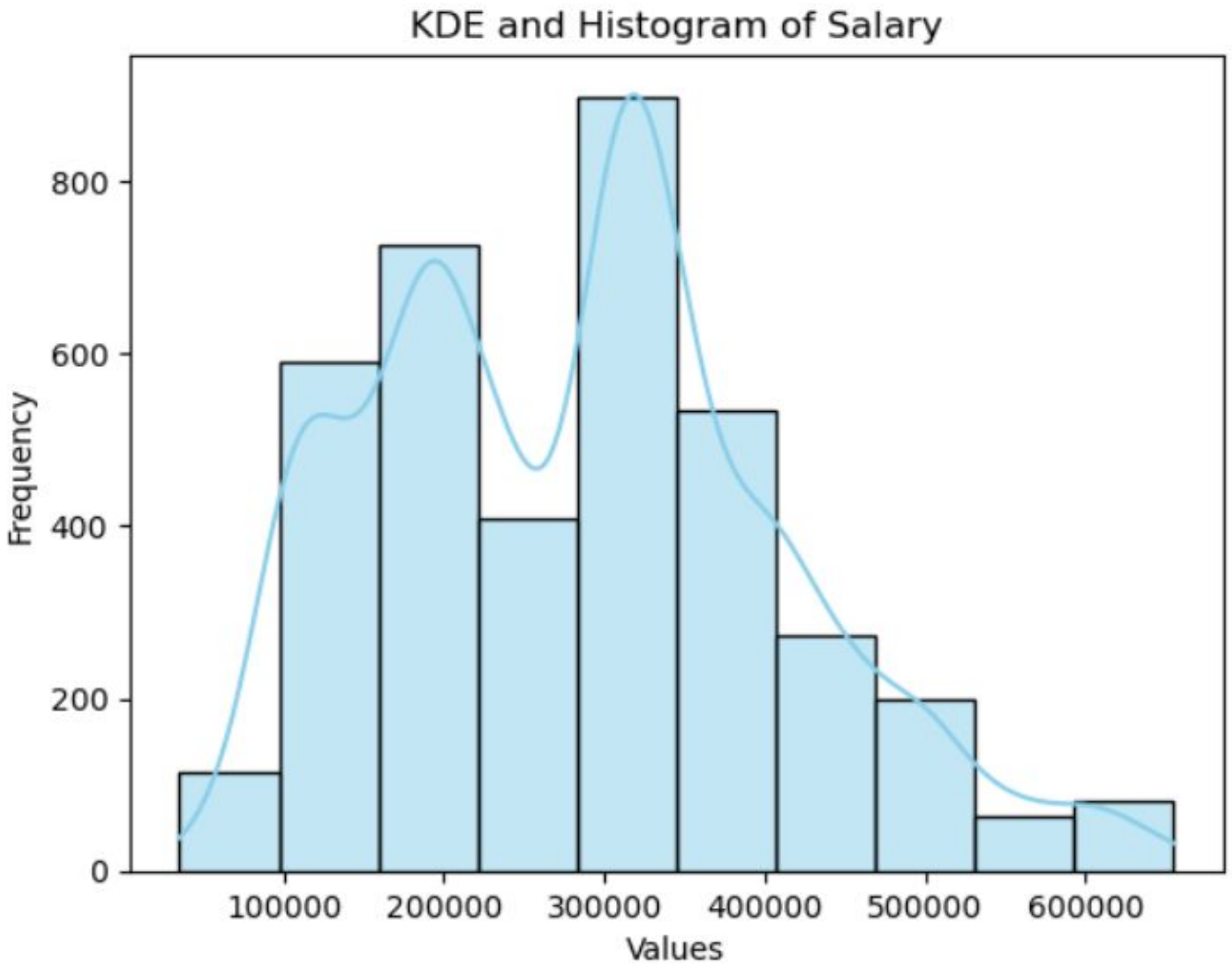
It is an approach for analyzing and summarizing data that allows analysts to identify patterns, trends, and relationships within the data.

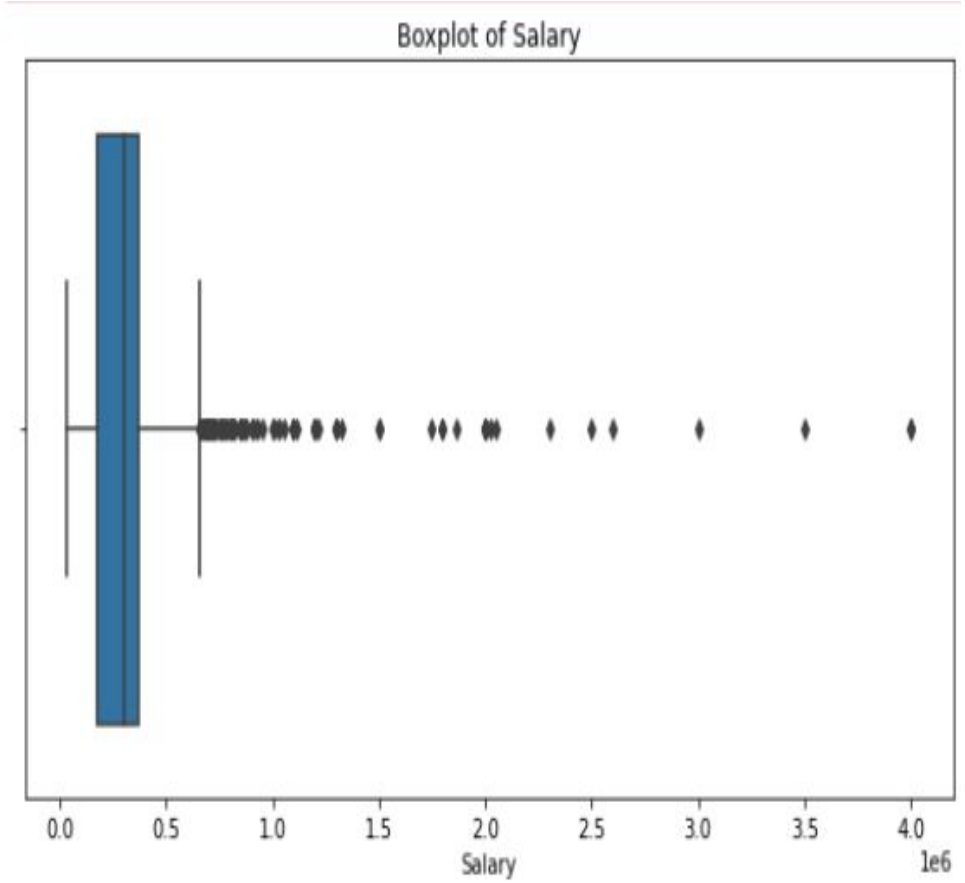


Univariate Analysis

It is a type of data visualization where we visualize only a single variable at a time.



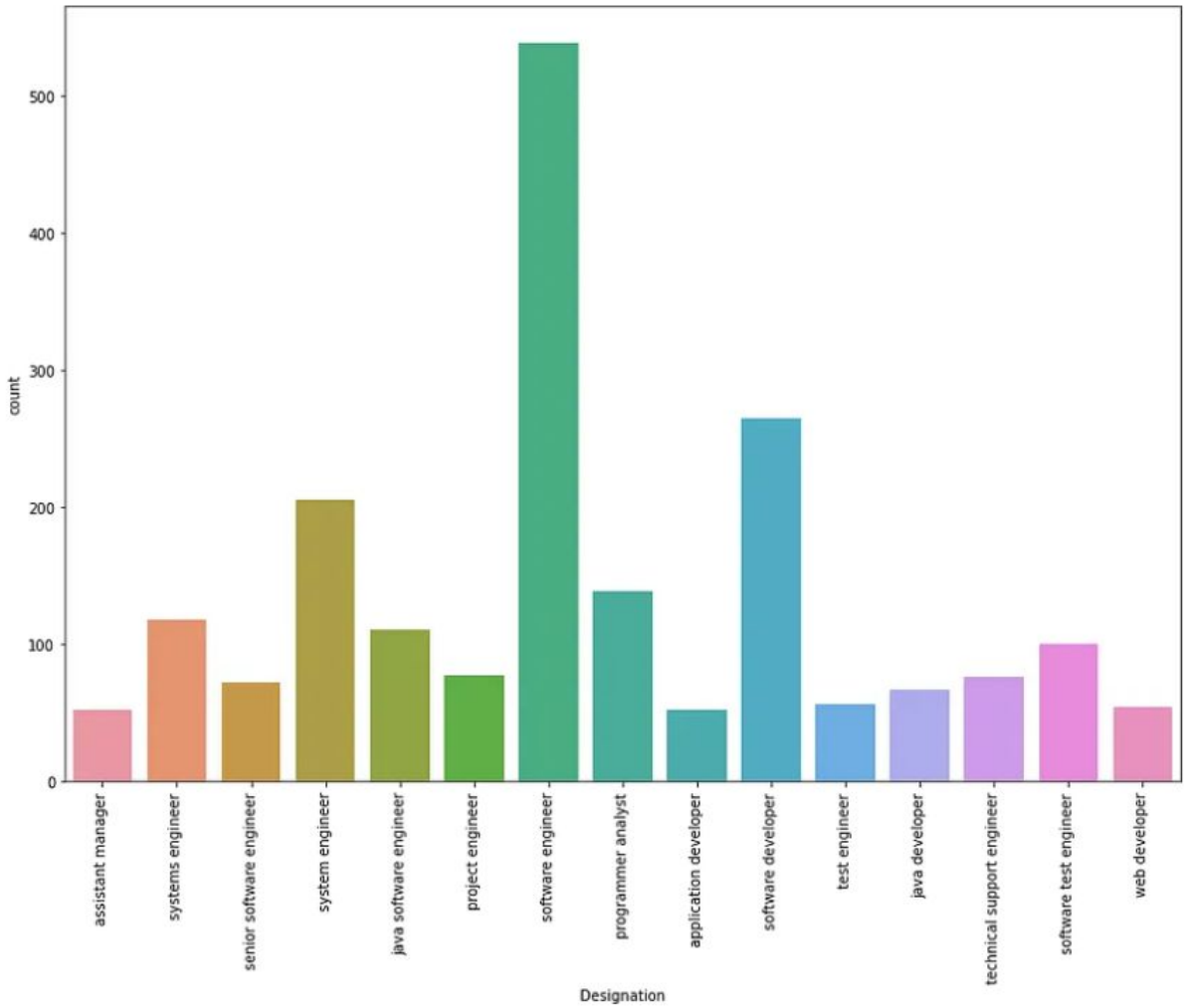


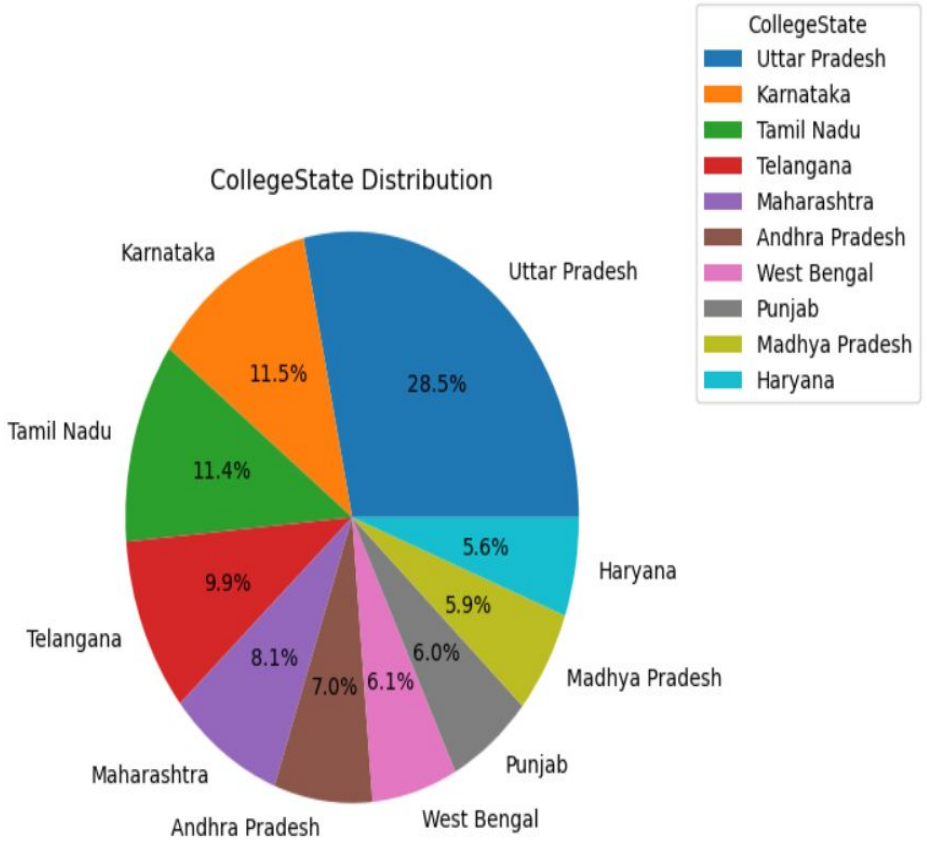


The distribution of Salary is right skewed as most of them earn low salary. Outliers suggest specific roles or location warrant further analysis.

Most of them salary ranges from 3,00,000 Lakhs to 3,50,000 lakhs

INNOMATICS RESEARCH LABS



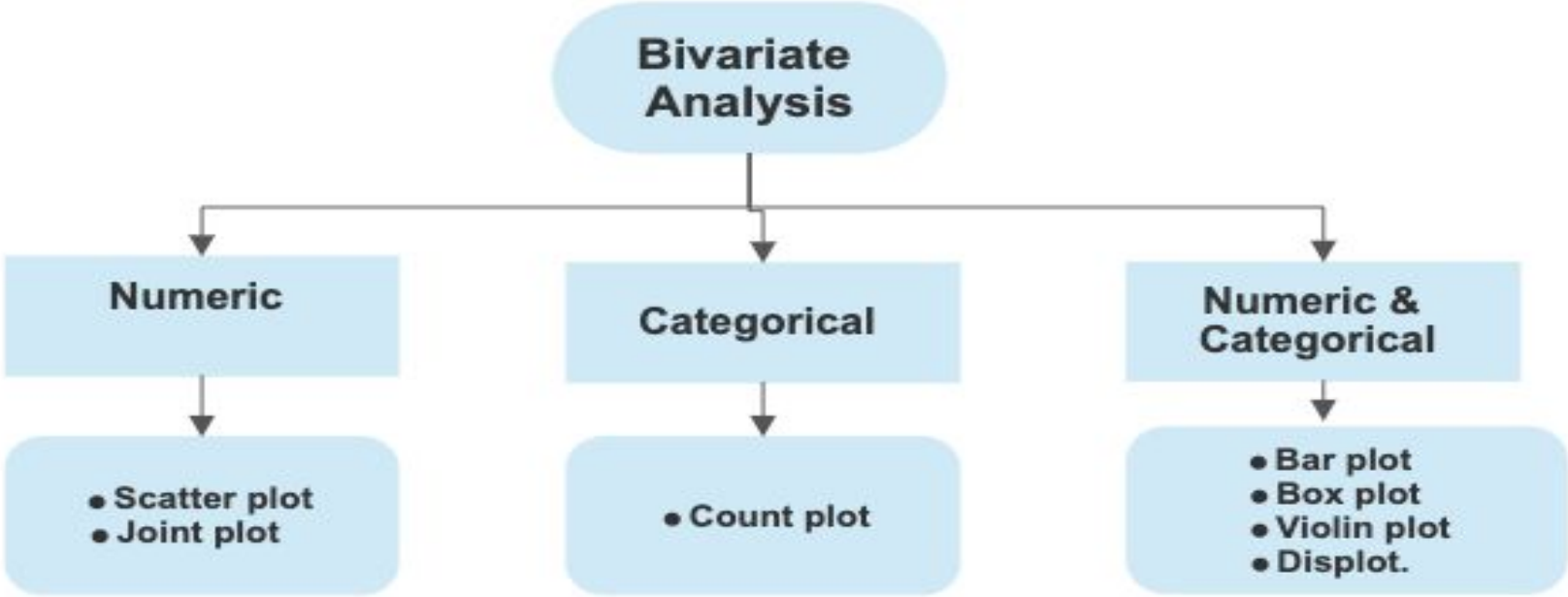


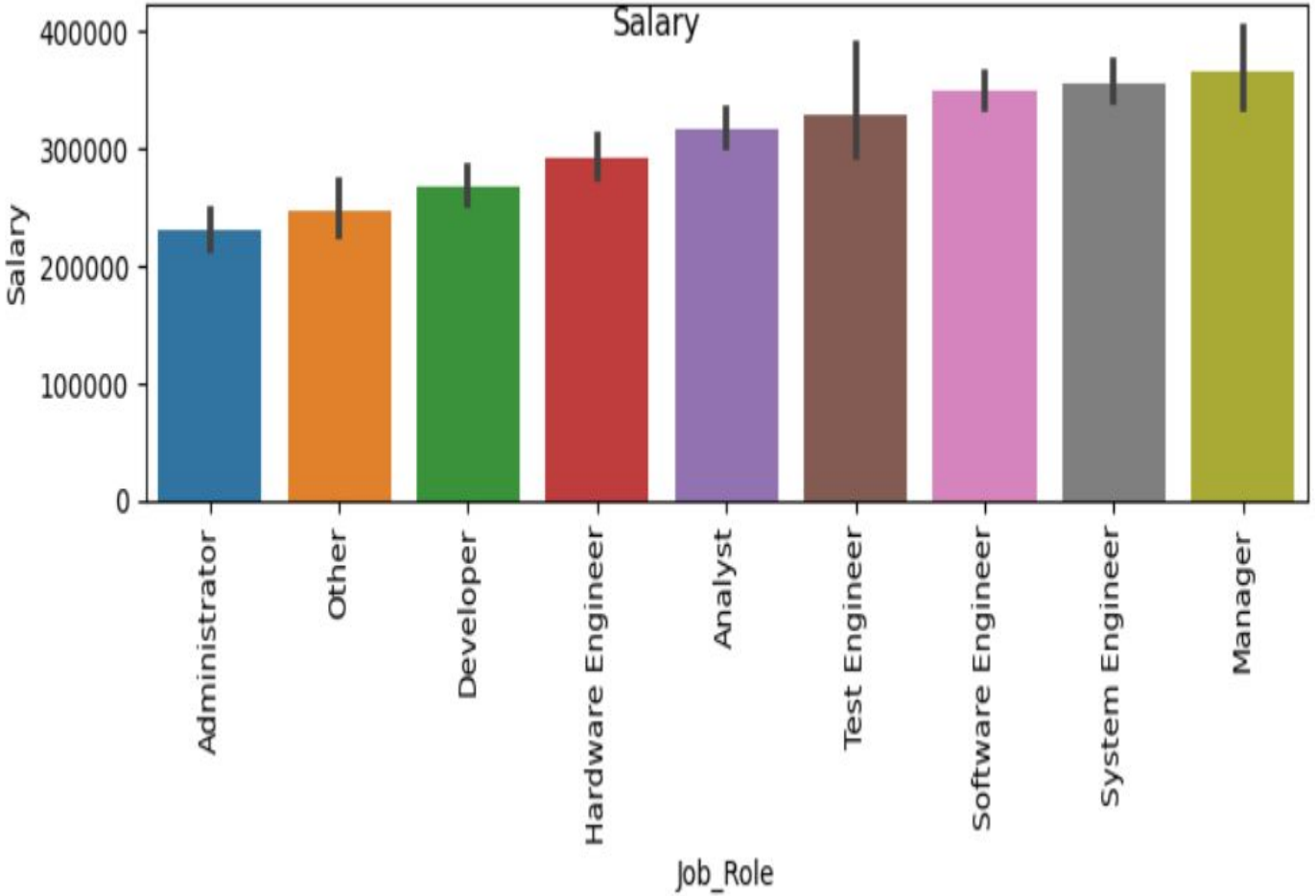
Most popular designation is Software Engineer.
Most students are from Uttar Pradesh followed by Karnataka and TamilNadu States.

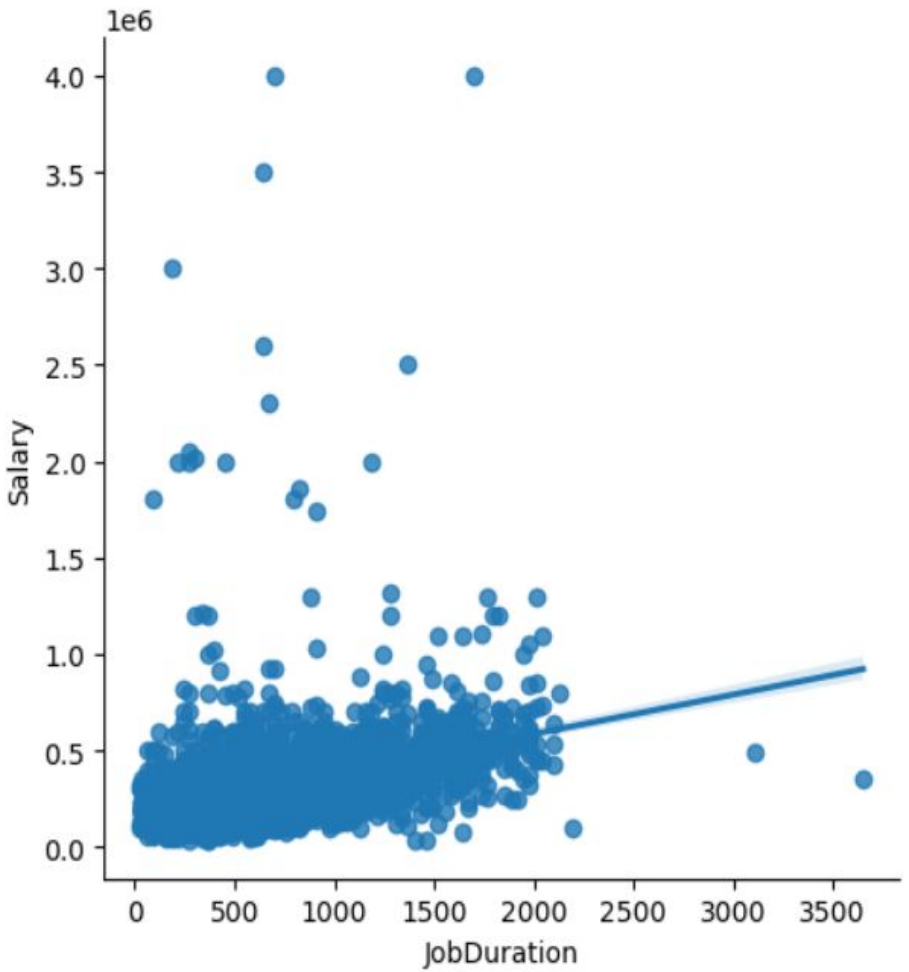
INNOMATICS RESEARCH LABS

Bivariate Analysis

It involves two different variables, and the analysis of this type of data focuses on understanding the relationship or association between these two variables.







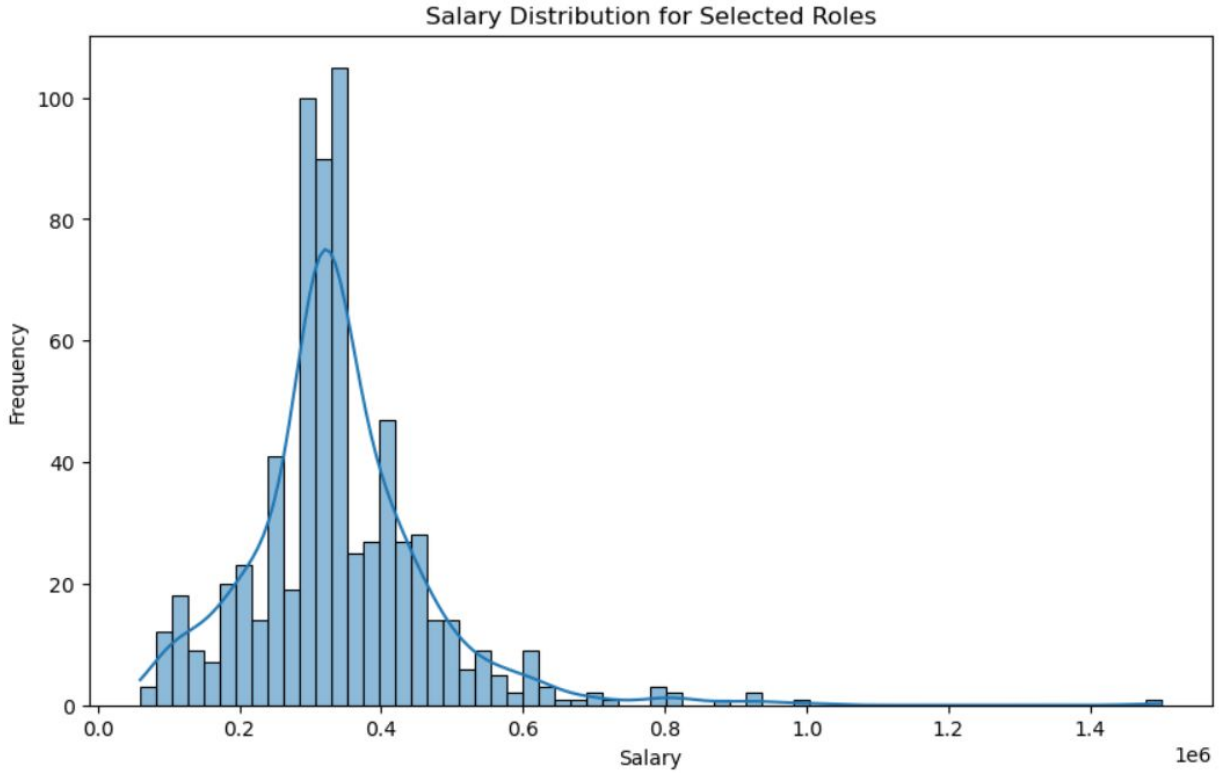
High paying job roles are Manager followed by System Engineer and Software Engineer.

Higher the experience results in higher salary.

INNOMATICS RESEARCH LABS

Business Questions

After doing your Computer Science Engineering if you take up jobs as a Programming Analyst, Software Engineer, Hardware Engineer and Associate Engineer you can earn up to 2.5-3 lakhs as a fresh graduate.



The claim that fresh graduates can earn up to 2.5-3 lakhs is not supported by the data.

INNOMATICS RESEARCH LABS

Is there a relationship between Gender and Specialization?

```
from scipy.stats import chi2_contingency

# Creating a contingency table
contingency_table = pd.crosstab(df['Speciality'], df['Gender'])

# Perform chi-square test for independence
chi2, p, dof, expected = chi2_contingency(contingency_table)
print("Chi-square statistic:", chi2)
print("p-value:", p)
```

```
Chi-square statistic: 20.899113570502816
p-value: 0.00011047893617613491
```

As the p-value is much smaller than the typical significance level of 0.05, I reject the null hypothesis - H0. Therefore, there is a relationship between Gender and Specialization

Conclusions / Findings

01 70% of them are Male candidates compared to female candidates in 2015 study

02 The majority of graduates find employment as software engineers and developers.

03 The strongest correlation with salary is the duration of employment at the company.

04 Graduates from Tier 1 colleges tend to earn higher salaries

INNOMATICS RESEARCH LABS

THANK YOU

INNOMATICS RESEARCH LABS