INNOMATICS®
RESEARCH LABS

INNOVATION. AUTOMATION. ANALYTICS

PROJECT ON

AMCAT EDA Data Analysis

Narayana Pagadala

OBJECTIVES OF THE PROJECT

The AMCAT data analysis focus on understanding various factors that target variable as salary in that dataset.

The analysis to gain insights and understanding from the provided dataset. Investigate on the relationship between the independent variables and dependent variable and salary column. Finally comes up with conclusion of insight.

Research Questions:

•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?)

INNOMATICS RESEARCH LABS

SUMMARY OF THE DATA

- **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 38 independent columns (29 numerical columns and 9 categorical columns) to find the individual impacts on salary and 3998 data points.**
- **The independent variables are both continuous and categorical in nature.**

INNOMATICS RESEARCH LABS

DATA PREPARATION AND CLEANING:

- Find out number of rows and columns
- checking duplicated values
- Statistical Measurements of each column
- Find out datatype of each column
- Unique values of each column
- Removing unnecessary columns
- Finding the numerical and categorical columns

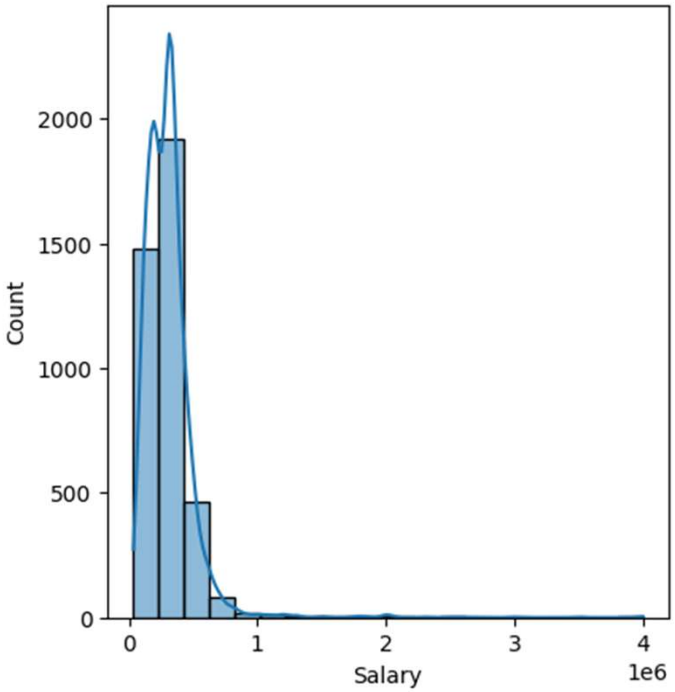
INNOMATICS
RESEARCH LABS

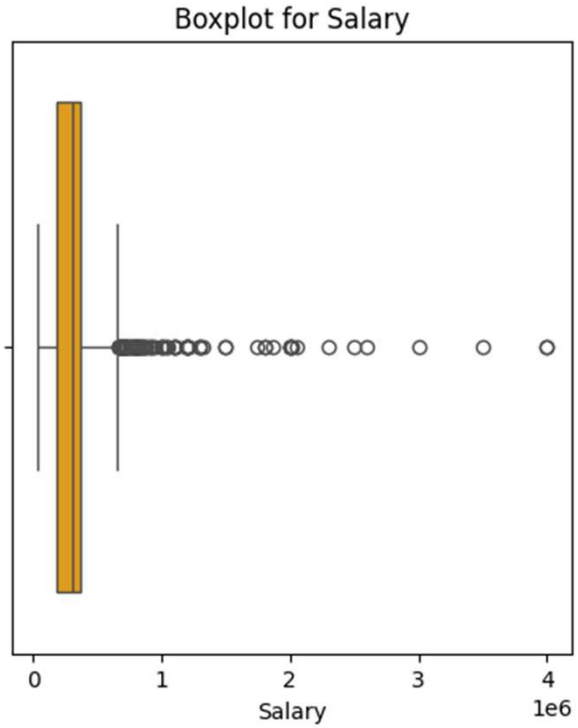
DATA VISUALIZATION:

- **Univariate Analysis**
- **Bivariate Analysis**
- **Multi Variate Analysis**

INNOMATICS
RESEARCH LABS

Univariate:

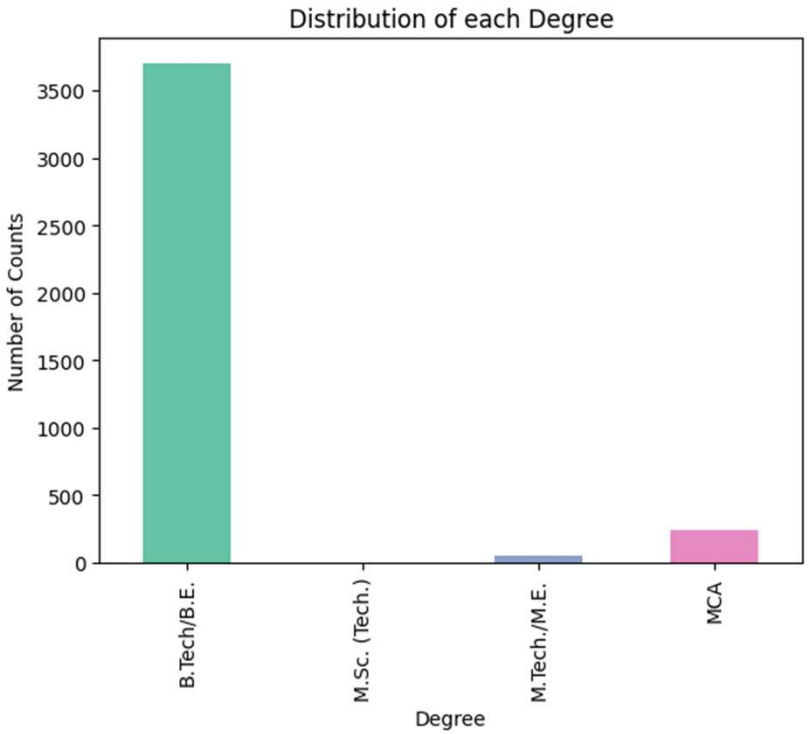


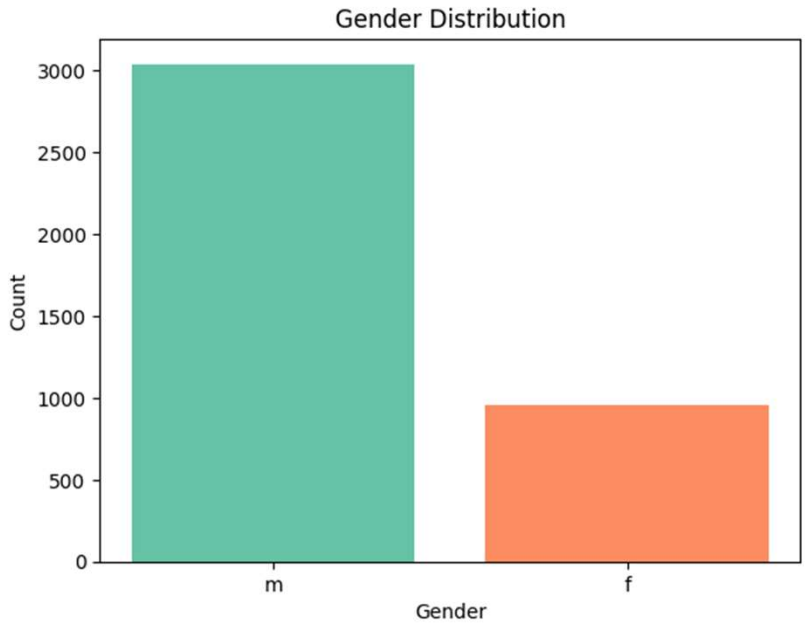


- **The salary distribution is highly right-skewed, with the majority of salaries concentrated below 20,00,000.**
- **A significantly peak is observed around 400,000,indicating a large number of employees earning within this image.**
- **The boxplot visualizes the distribution of salary values, indicating a wide range of salaries with several outliers towards the higher end**

INNOMATICS RESEARCH LABS

Univariate:



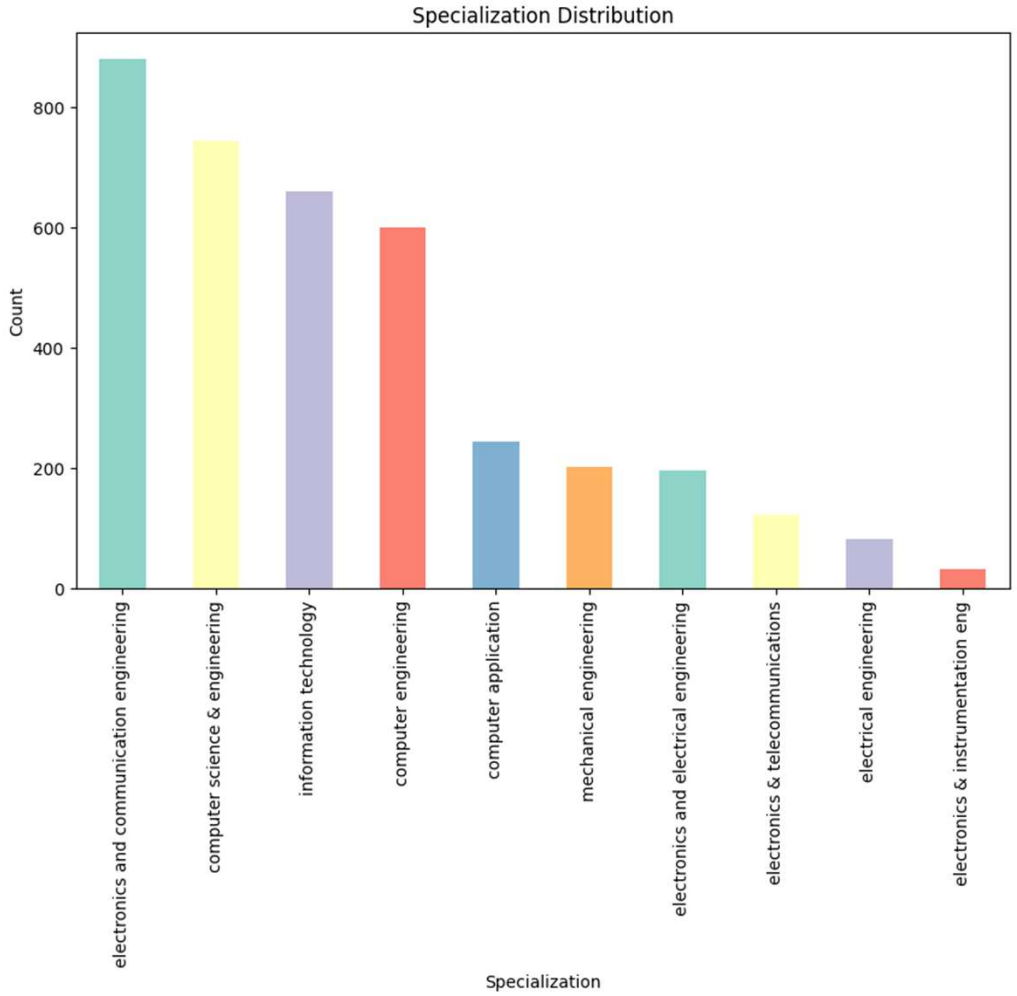


B.Tech / BE students are higher compare to the other Degrees.

Males are 3 times higher than to the females.

INNOMATICS RESEARCH LABS

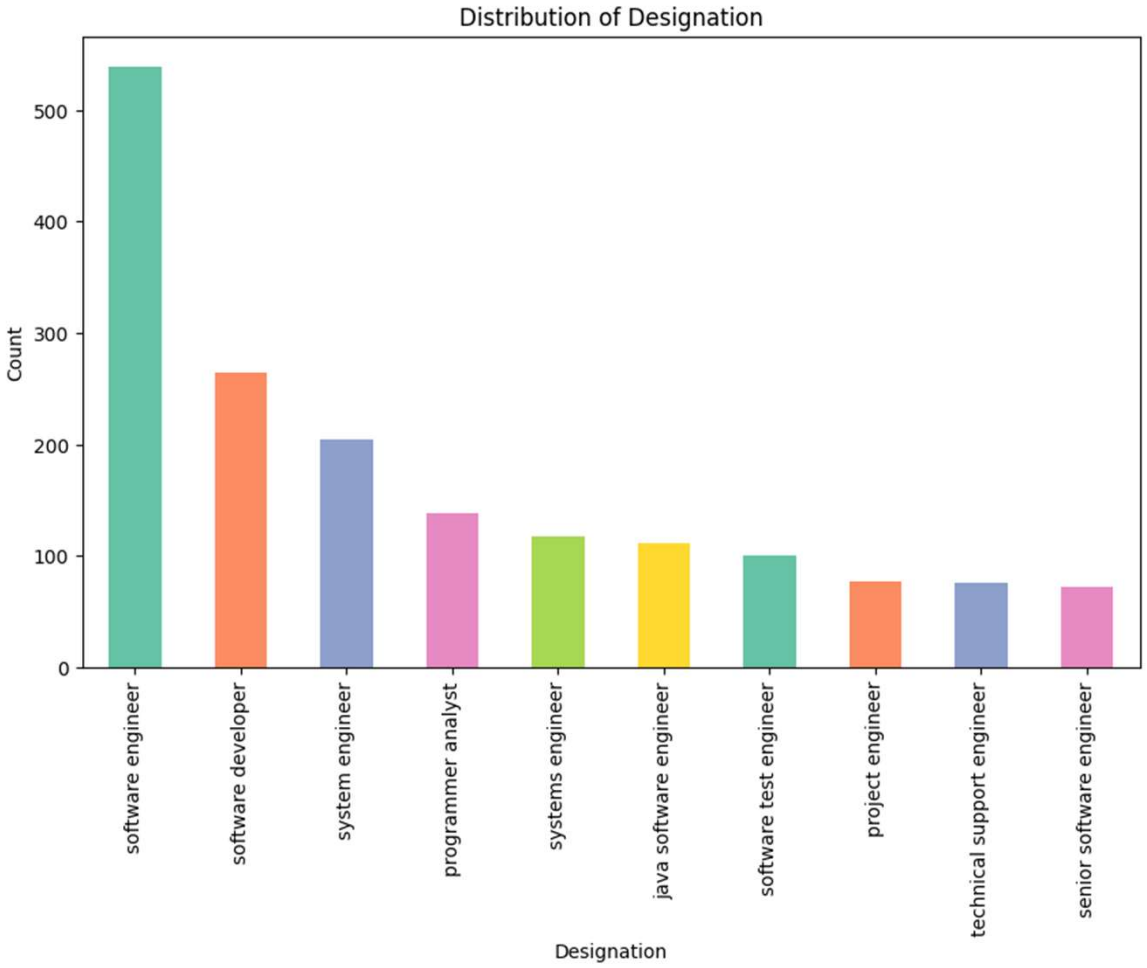
Univariate:



- **The electronic and communication engineering graduates are higher compare to the other specializations.**
- **The highest number of graduates is in computer science and engineering followed by Electronics and communication engineering.**

INNOMATICS RESEARCH LABS

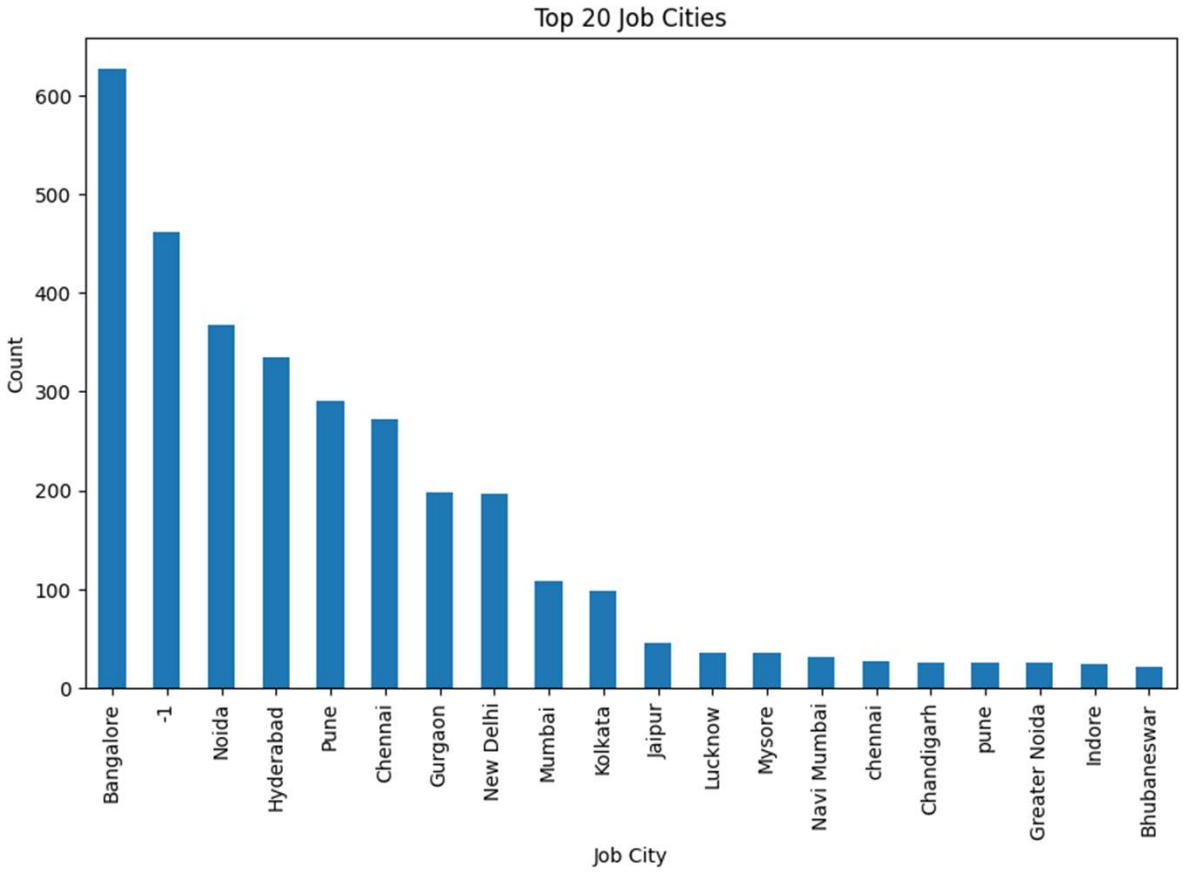
Univariate:



- **This bar plot the frequency distribution of the top 10 designation recorded in the dataset.**
- **The highest number designation is in software engineers compare to the other designations.**

INNOMATICS RESEARCH LABS

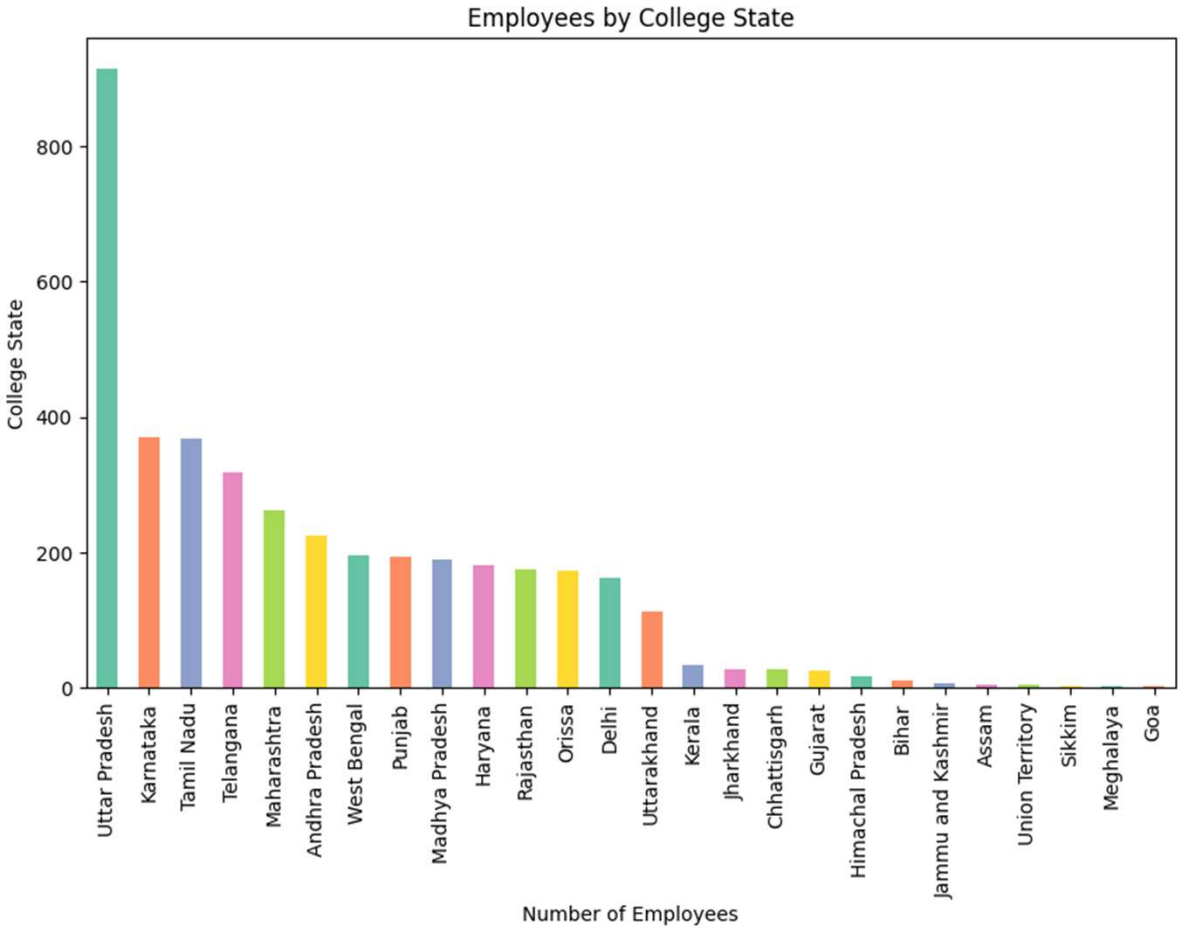
Univariate:



- **This bar plot the frequency distribution of the top 20 job cities recorded in the dataset.**
- **Top 5 cities are Bangalore , Noida, Hyderabad, Pune, and Chennai.**
- **“Bangalore” have the most number of employees.**
- **“-1” shows that the there exist some null values that needs to be fined and cleaned for further processing.**

INNOMATICS RESEARCH LABS

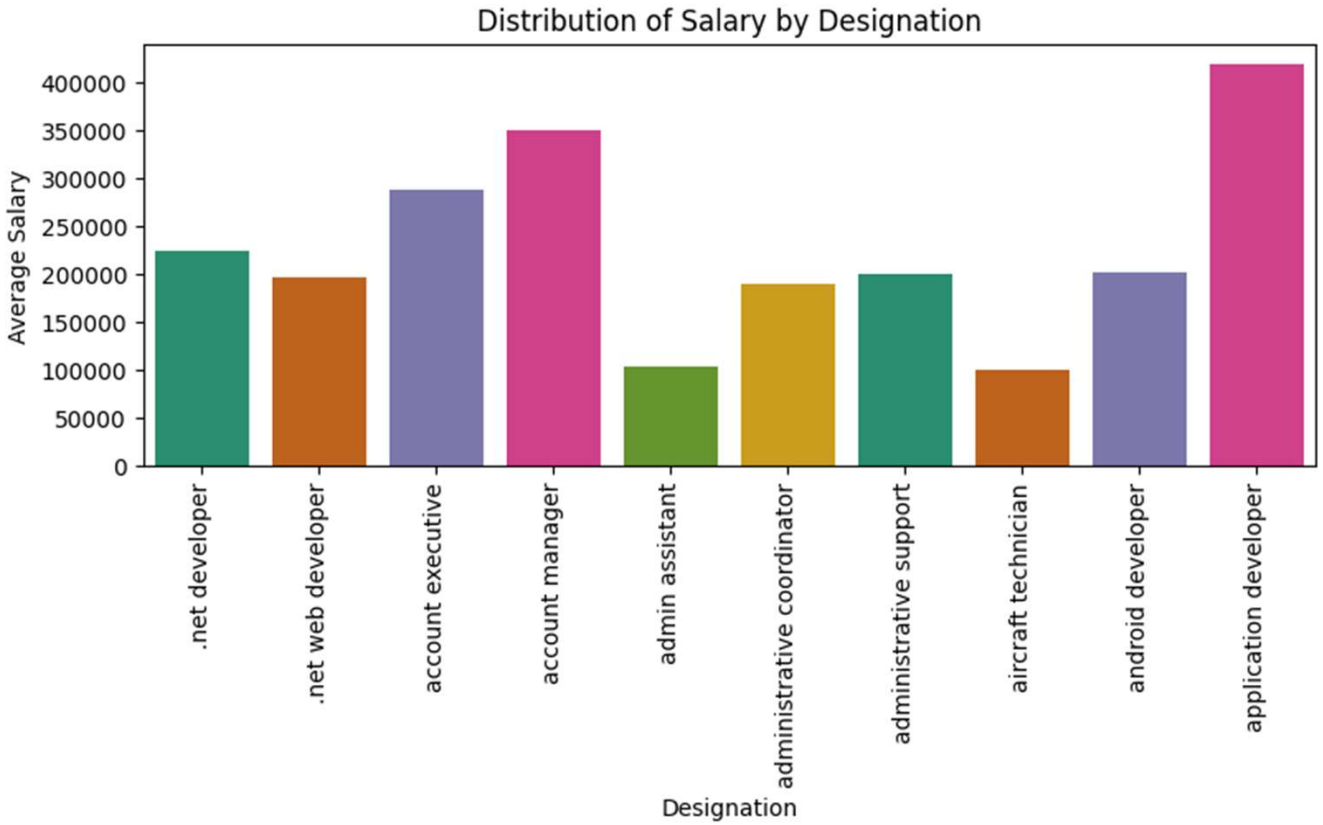
Univariate:



- **Top 5 College states are Uttar Pradesh, Karnataka, Tamil Nadu, Telangana and Maharashtra.**
- **“Uttar Pradesh” have the most number of the employees.**
- **And Uttar Pradesh have 2 times more higher employees compare to other college states.**

INNOMATICS RESEARCH LABS

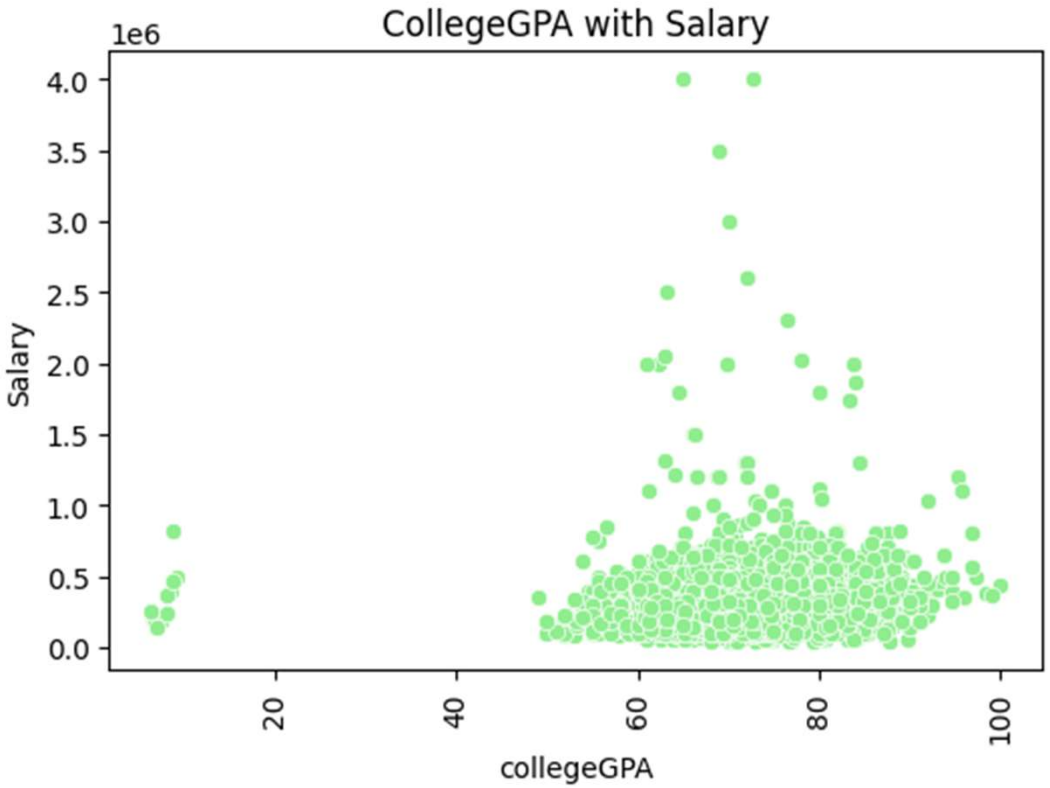
Bivariate:



- **The average salary of Application developer is more compared to other designations.**
- **The less salaries are admin assistant and aircraft technician.**

INNOMATICS RESEARCH LABS

Bivariate:



- **The More Salaries get the College GPA range between 45 to 99.**
- **The highest salary get the college GPA range between 60 to 80.**

INNOMATICS RESEARCH LABS

Determine whether fresh graduates earn 2.5-3 lakhs annually as stated in the article.

Maximum salary for fresh graduates in relevant designations: nan

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

INNOMATICS RESEARCH LABS

Does the preference of Specialisation depend on the Gender?

```
import pandas as pd
from scipy.stats import chi2_contingency

# Assuming 'df' with 'Gender' and 'Specialization'
columns
# Create a contingency table
cont_tab = pd.crosstab(df['Gender'],
df['Specialization'])

# Perform the Chi-squared test
chi2, p, do f, expected = chi2_contingency(cont_tab)

# Print the results
print(f" Chi-square statistic: {chi2}")
print(f" P-value: {p}")

# Interpret the results
alpha = 0.05  # Significance level
if p < alpha:
    print("There is a significant relationship between
gender and specialization.")
else:
    print("There is no significant relationship between
gender and specialization.")
```

- **There is a significant relationship between gender and specialization.**

```
Chi-square statistic: 104.46891913608455 P-value: 1.2453868176976918e-06 There is a significant
relationship between gender and specialization.
```

INNOMATICS RESEARCH LABS

Conclusion

- **Salary Distribution: The salary distribution shows a right skew , with outliers at the higher end, indicating a few individuals earn significantly more than others.**
- **Gender Disparity: Males tend to occupy more positions across cities, and there might be a salary gap favouring males.**
- **Salary claim validation: Based on the dataset, the average salary for computer science engineers does align with or slightly exceed the claim made in the time of India article.**
- **Job City Insights: certain cities, like Bangalore , have a higher concentration of employees, potentially offering better salary prospects.**

INNOMATICS RESEARCH LABS

THANK YOU

INNOMATICS
RESEARCH LABS