INNOMATICS®
RESEARCH LABS

INNOVATION. AUTOMATION. ANALYTICS

PROJECT ON

Exploratory Data Analysis on AMEO Dataset

Deepraj Vadhwane

INNOMATICS
RESEARCH LABS

About me

After completing my BTech from MGM College of Engineering in Nanded, Maharashtra, I found my passion for Data Science amidst exploration in various technological streams. Following graduation, I opted to enhance my expertise by joining Innomatics Research Lab in Hyderabad. Here, I engaged in intensive learning, addressing assignments, quizzes, and projects spanning Python, SQL, Power BI, Data Analysis, and Machine Learning, including tasks like web scraping and data analysis.

My fascination with Data Science lies in its transformative impact on modern society. I am intrigued by how AI expedites processes once time-consuming, allowing for actionable insights from data and the creation of predictive models. Currently serving as a Data Scientist intern at Innomatics Research Lab, I continue to delve deeper into this field, tackling real-time case studies and exploring new concepts.

To me, D**ata is the fuel that propels businesses forward.**

https://www.linkedin.com/in/deeprajvadhwane/

https://github.com/DeeprajVadhwane

INNOMATICS RESEARCH LABS

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 Steps***
b. ***Data Manipulation Steps***
c. ***Univariate Analysis Steps***
d. ***Bivariate Analysis Steps***

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

INNOMATICS RESEARCH LABS

Business Problem Statement

The engineering industry is currently confronted with the challenge of optimizing the career trajectory of its graduates. Despite possessing diverse cognitive, technical, and personality skills, many engineering graduates struggle in a fiercely competitive job market.

In response, we are undertaking an extensive data exploration and analysis initiative. Through thorough examination of a comprehensive dataset comprising employment outcomes, standardized scores from cognitive, technical, and personality assessments, as well as demographic information, our goal is to unveil the keys to career advancement.

Our mission is unequivocal: to unearth actionable insights that will transform the employment landscape for engineering graduates, guiding them towards fulfilling and prosperous professional paths.

Objective of project

- Exploration of Relationships: Investigate the relationships between independent variables and the target and Dependent variable
- Pattern and Trend Identification: Uncover any discernible patterns or trends present within the data, which may offer valuable insights into the underlying dynamics of employment outcomes.
- Research 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.
- And Is there a relationship between gender and specialization? (i.e. Does the preference of Specialisation depend on the Gender?)

Workflow:-

INNOMATICS RESEARCH LABS

Univariate Categorical Analysis

Job City Insights:

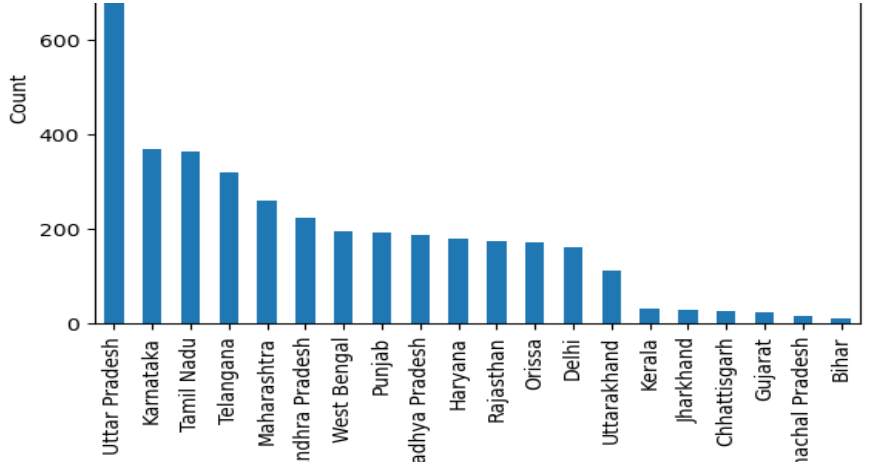
- The top job cities based on the dataset are Bangalore, Noida, Pune, Gurgaon, Mumbai, Lucknow, Mysore, Navi Mumbai, and Delhi.
- This suggests that these cities are major hubs for employment opportunities, with a concentration of job opportunities in various sectors.

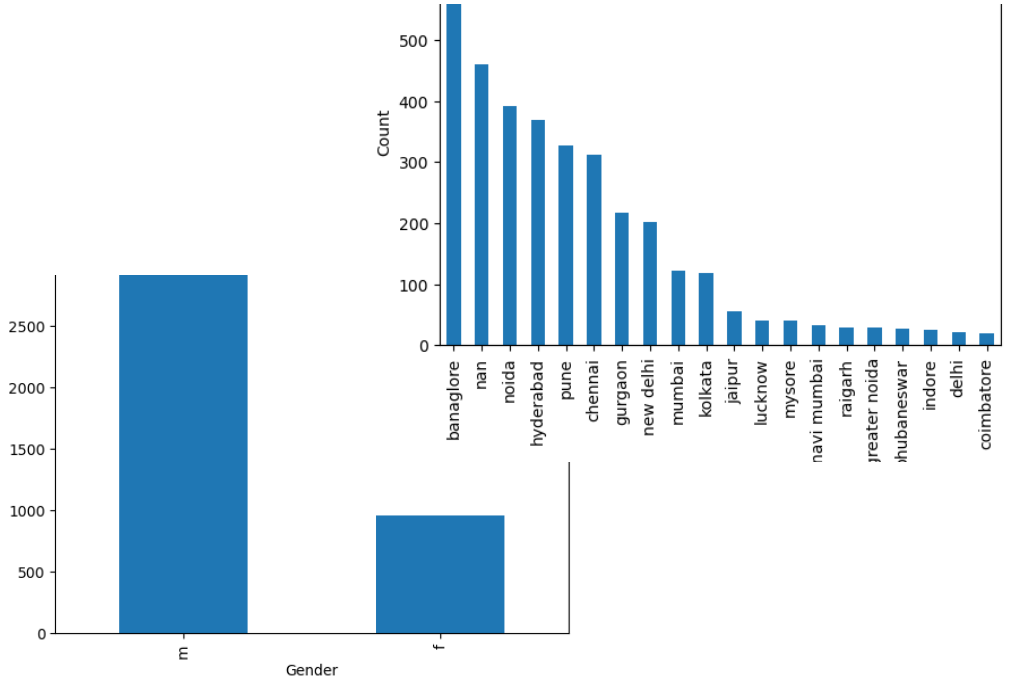
College State Insights:

- The top college states based on the dataset are Uttar Pradesh (UP), Karnataka, Tamil Nadu, Maharashtra, West Bengal, Punjab, Madhya Pradesh, Haryana, and Delhi.
- This indicates that these states have a significant presence of educational institutions, contributing to the pool of skilled graduates entering the workforce.

Gender Distribution Insights:

- The dataset shows a gender distribution where males constitute approximately 70% of the candidates, while females constitute about 30%.
- This suggests a gender imbalance in the dataset, with a higher representation of males compared to females.
- It's important to note the gender distribution for analyzing gender diversity and ensuring inclusivity in recruitment processes and workplace environments.





INNOMATICS RESEARCH LABS

(10th percentage, 12th percentage, and CGPA) vs Salary

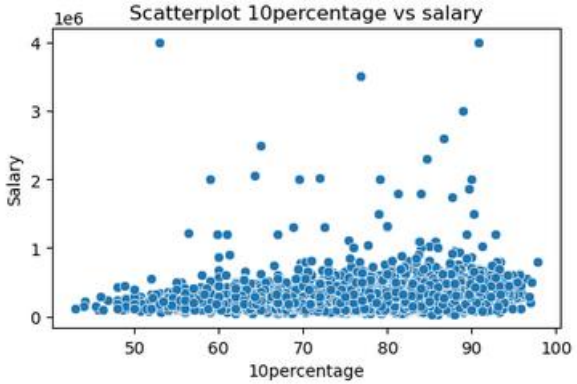
10th and 12th: There is a significant positive correlation between 10th and 12th percentages and Salary (correlation coefficients of 0.177 and 0.169 respectively, p-value=0.000). Higher academic performance, particularly in 10th and 12th grades, tends to correspond with higher Salary.

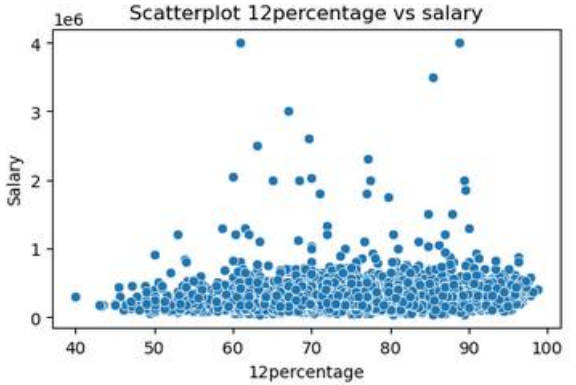
CGPA as a Determinant: The analysis also reveals a significant correlation between CGPA and Salary (correlation coefficient of 0.146, p-value=0.000). Candidates with a CGPA above 60% are more likely to secure jobs and negotiate higher Salaries compared to those with lower CGPA.

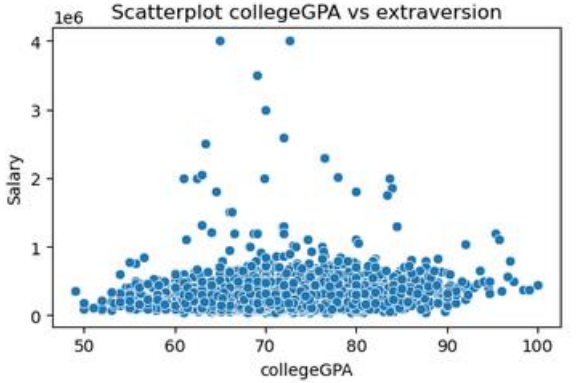
Threshold Effect of CGPA: There appears to be a threshold effect, suggesting that a minimum CGPA requirement (e.g., 60%) is necessary to enhance job prospects and Salary negotiation.

Consideration in Hiring Decisions: Employers may prioritize academic performance, particularly CGPA and 10th/12th percentages, when making hiring decisions. Candidates with stronger academic backgrounds may be perceived as having higher potential value to the organization.

Holistic Evaluation Needed: While academic performance plays a significant role, it is not the sole determinant of Salary. Employers should consider other factors such as work experience, skills, and cultural fit to make well-informed hiring decisions and ensure optimal utilization of human capital.







INNOMATICS RESEARCH LABS

(Conscientiousness, Neuroticism, Extraversion) vs Salary

Conscientiousness vs Salary:

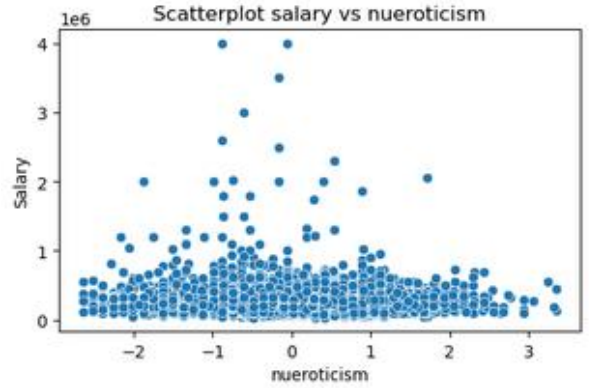
- Conscientiousness reflects organization, responsibility, and diligence.
- Weak negative correlation (coefficient: -0.063, p-value: 0.000) with Salary.
- Individuals with higher conscientiousness tend to have higher salaries.

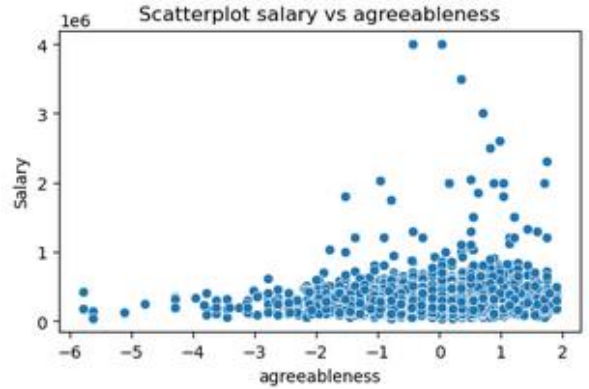
Neuroticism vs Salary:

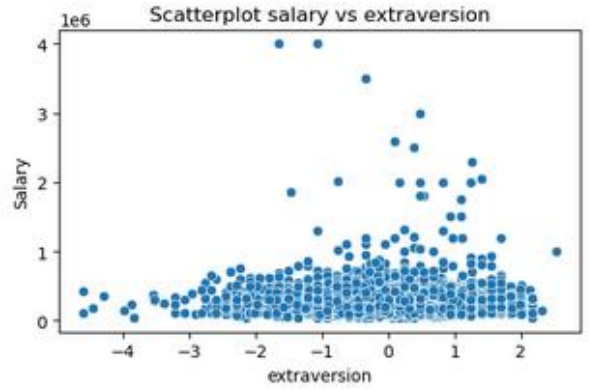
- Neuroticism entails negative emotions like anxiety and vulnerability.
- Significant salary difference based on neuroticism levels.
- Lower neuroticism linked to higher salaries, indicating an impact on earning potential.
- Neuroticism may affect job performance, leading to decreased productivity and career advancement challenges.

Extraversion vs Salary:

- Weak and insignificant correlation (coefficient: -0.010, p-value: 0.534) with Salary.
- Individuals with high extraversion may still have opportunities for higher salaries due to networking, leadership, and ambition.



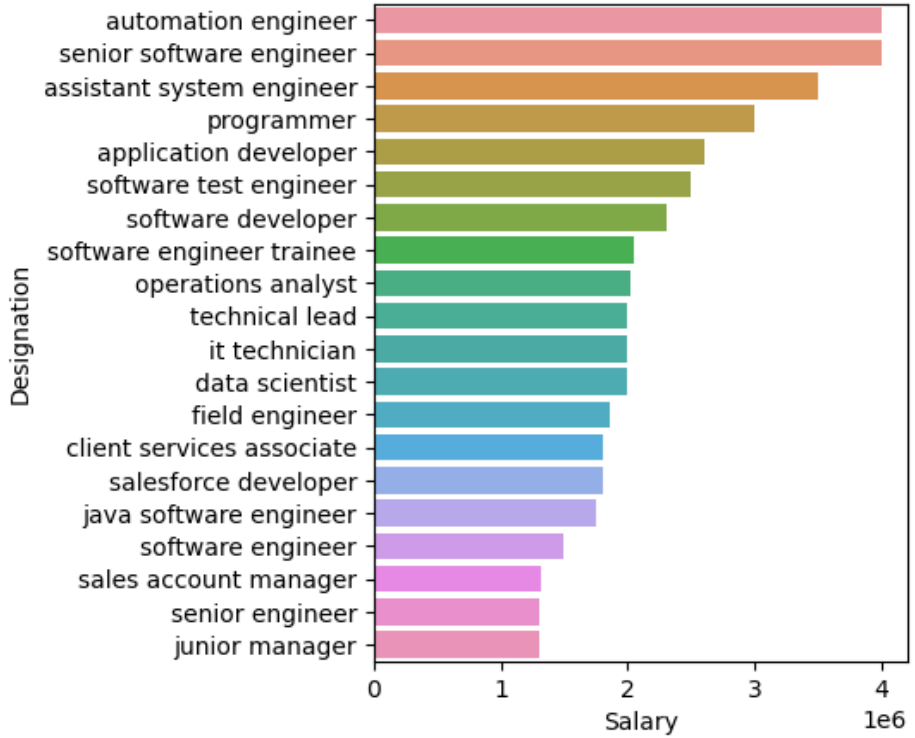




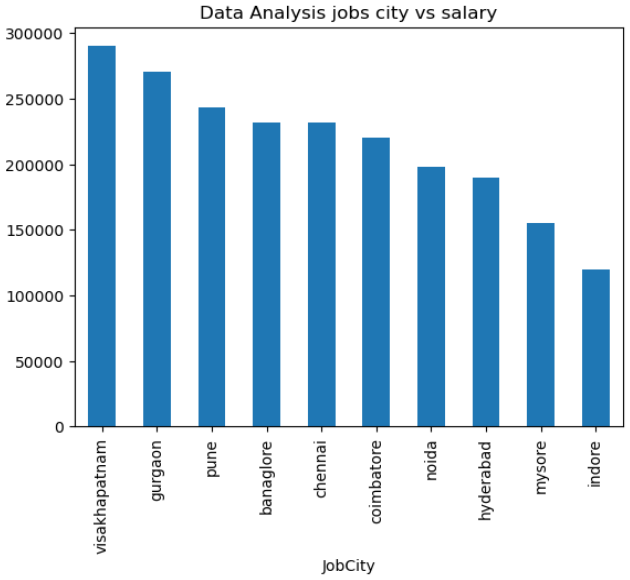
INNOMATICS RESEARCH LABS

Which top 20 jobs Designation has more salary in IT companies ?

- **High-Paying Roles:** Analysis reveals that job titles such as Data Scientist, Senior Developer, and Technology Lead are among the top 20 positions commanding higher salaries within
- **Leadership and Expertise:** Positions like Branch Manager, Research Scientist, and Sales Account Manager also feature prominently in the list, indicating that leadership roles and specialized expertise contribute significantly to salary levels within the IT industry.

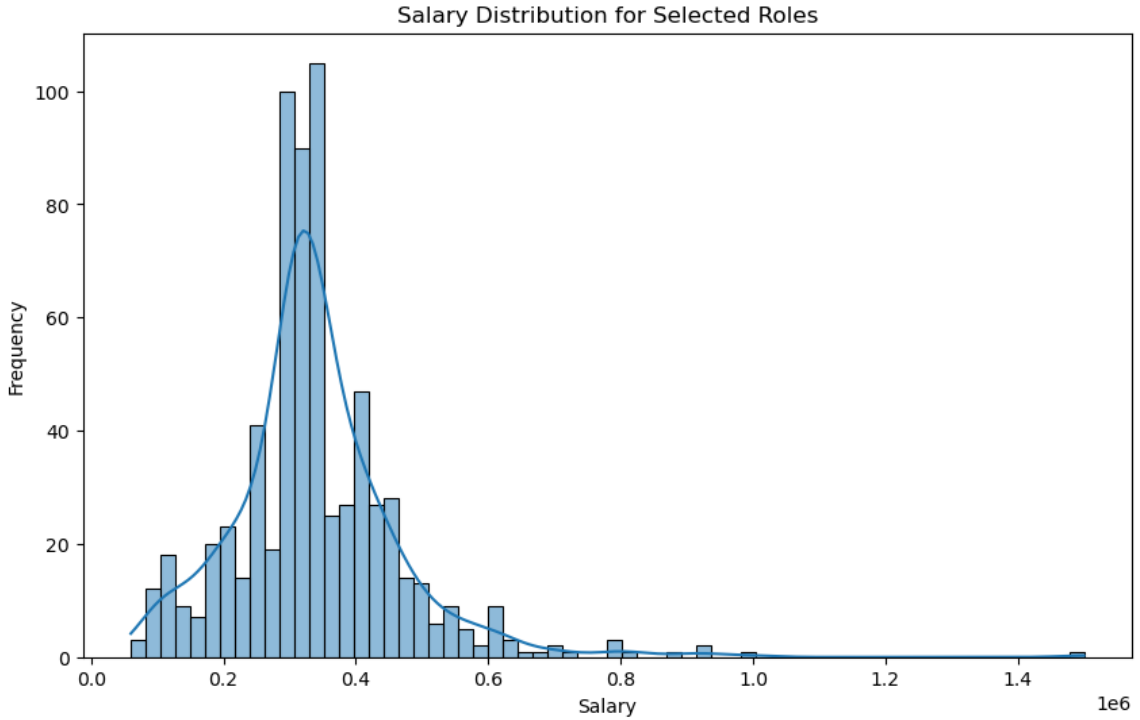


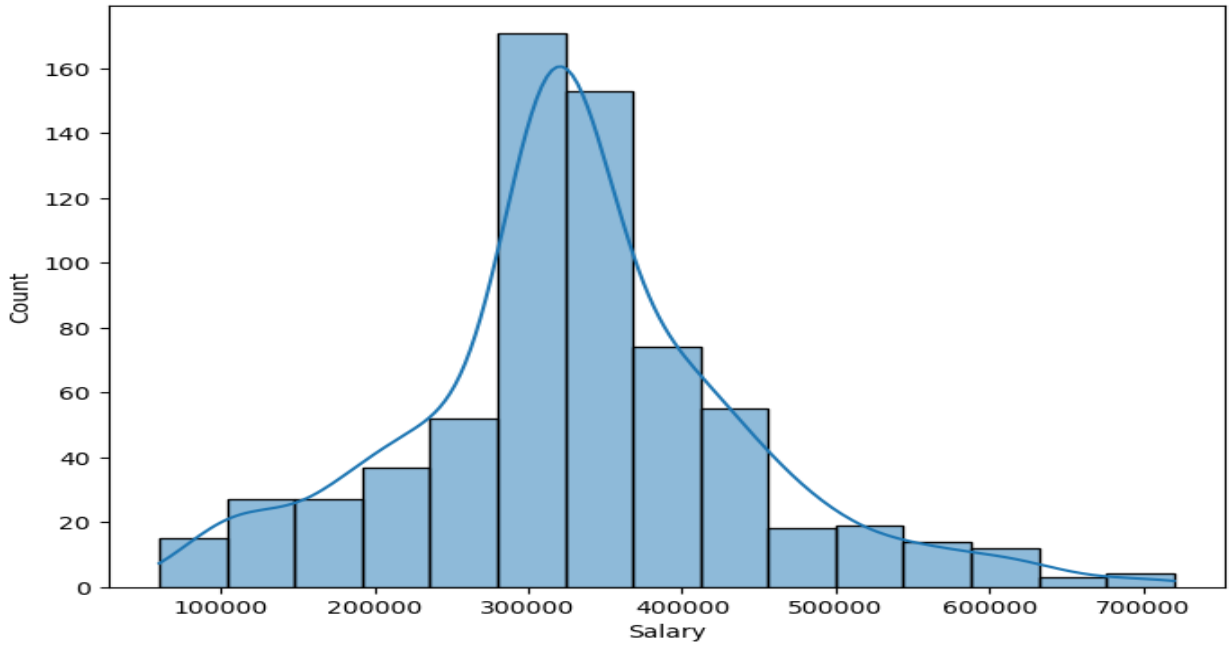
Which are top 20 job cities offer higher salaries for Data Analysis Designation ?



- **Key Employment Hubs:** Cities such as Gurgaon, Bangalore, Pune, Noida, and Hyderabad emerge as prominent destinations for Data Analysis roles with lucrative salary packages.
- **Growing Opportunities:** Emerging job markets like Visakhapatnam, Chennai, Coimbatore, Mysore, and Indore also feature in the list, indicating a growing demand for data analysis professionals in these regions.

INNOMATICS RESEARCH LABS





Software Engineer, Hardware Engineer and Associate Engineer you can earn up to 2.5-3 lakhs as a fresh graduate..5-3 lakhs as a fresh graduate. Test this claim with the data given to you.

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

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

```
:
1 from scipy.stats import chi2_contingency
2
3 gen_spec_df = df[df['Specialization'] != 'Other']
4 cont_tab = pd.crosstab(index=gen_spec_df['Gender'], columns=gen_spec_df['Specialization Category'])
5 chi2 = chi2_contingency(cont_tab, correction=False)
6 print("statistic:", chi2.statistic)
7 print("p-value:", chi2.pvalue)
```

```
statistic: 48.36336842977407
p-value: 1.782144335086325e-10
```

The p-value associated with the statistic (p = 1.782e-10) exceeds the chosen significance level, suggesting that the variables **gender** and **specialization** are independent of each other.

Therefore, there is no significant association between gender and specialization

Final Conclusion

Academic Performance Insights:

- Both 10th and 12th percentages exhibit a clustered distribution between 65% to 90%, with a significant portion of students scoring above 65%.
- College GPA distribution also shows a concentration between 65 to 85, with fewer students having exceptionally high GPAs above 95.

CGPA as a Determinant:

- CGPA above 60% is associated with enhanced job prospects and salary negotiation power.
- A threshold effect suggests a minimum CGPA requirement for favorable employment opportunities.

Gender Distribution Insights:

- Males constitute around 70% of the dataset, indicating a gender imbalance.
- Gender distribution is vital for ensuring inclusivity in hiring processes and workplace environments.

Correlation Analysis:

- Significant positive correlations exist between 10th/12th percentages, college GPA, and salary.
- Candidates with higher academic performance tend to negotiate higher salaries.

Final Conclusion

Job City and College State:

- Major job hubs include Bangalore, Noida, Pune, Gurgaon, and Mumbai, while top college states comprise Uttar Pradesh, Karnataka, Tamil Nadu, and Maharashtra.
- These locations indicate strong employment and educational infrastructure, respectively.

Employment :

- High-paying roles like Data Scientist and Senior Developer command lucrative salaries within IT firms.
- Leadership positions such as Branch Manager also correlate with higher compensation.

Key Employment:

- Cities like Gurgaon, Bangalore, and Pune are prominent destinations for high-paying data analysis roles.
- Emerging markets such as Visakhapatnam and Coimbatore offer growing opportunities in the field.

Conclusion on Salary Claim:

- The claim that fresh graduates can earn up to 2.5-3 lakhs is not supported by the data, based on statistical analysis of the variables.

Gender-Specialization Association:

- The p-value exceeds the significance level, indicating no significant association between gender and specialization.

THANK YOU

INNOMATICS RESEARCH LABS