INNOMATİCS®
RESEARCH LABS

INNOVATION. AUTOMATION. ANALYTICS

PROJECT ON

Exploratory Data Analysis of AMEO Dataset

KADIR A (IN9240213)

30TH SEPTEMBER 2024

About me

Hello everyone, I'm A. Kadir, a Pre-final-year student pursuing a B.Tech in Artificial Intelligence & Data Science at Panimalar engineering college. Throughout my academic journey, I have built a strong foundation in data science, machine learning, and software development, applying these skills to solve real-world challenges.

One of my key projects is a **Real-Time Speech Emotion Recognition System**, where I integrated a CNN model with Flask, achieving 82% accuracy. This project highlights my ability to work with advanced AI frameworks, including model training, data preprocessing, and real-time deployment. Additionally, I've developed an **Object Detection System** using the YOLO architecture, which accurately identifies objects in video streams.

Technically, I'm proficient in Python, SQL, and C++, with hands-on experience using libraries such as Pandas, NumPy, Matplotlib, and Scikit-learn for data analysis. I've also worked extensively with deep learning frameworks like TensorFlow, Keras, and PyTorch for building predictive models and deploying them in real-time applications.

I have participated in national hackathons, securing recognition for my contributions to **AI and machine learning** projects. My innovative work includes projects such as **NLP-based Voice Assistants, IPL Match Prediction Systems, and Automated Document Verification Systems**. Additionally, I've worked on various real-time web applications, showcasing my skills in full-stack development using Flask and JavaScript.

As I look forward to advancing my career, I am excited to apply my expertise in AI, machine learning, and software engineering to innovative projects, while continuing to learn and grow within the industry.

INNOMATICS RESEARCH LABS

Business Problem

Overview

The job market for engineering graduates is complex, with salary prospects shaped by several factors like skill sets, job roles, geographic locations, and demographic attributes. Data from Aspiring Minds' AMEO offers key insights into these aspects, helping to analyze and improve salary expectations.

Specific Challenges

- **Complexity of Factors: The interaction between cognitive abilities, technical skills, personality characteristics, job roles, and geographic locations makes it difficult to determine salary outcomes.**
- **Data Variability: The presence of various continuous and categorical variables requires in-depth analysis to identify significant trends.**
- **Predictive Modeling: Estimating salaries accurately while accounting for fluctuating market conditions poses significant challenges..**

Solution

Perform comprehensive exploratory data analysis (EDA) and feature engineering, utilize advanced machine learning algorithms for predictive modeling, and convert findings into practical recommendations to enhance employability and support salary negotiations.

INNOMATICS RESEARCH LABS

Use Case Domain Understanding

Overview

The use case domain revolves around understanding employment outcomes and salary determinants specifically tailored for engineering graduates. It encompasses a range of factors that influence salary outcomes, including skills, job titles, locations, demographics, and industry trends.

Importance

- ✓ **Strategic Decision-Making:** Informed decisions in recruitment, talent management, and career development.
- ✓ **Competitive Advantage:** Enhances talent acquisition and retention strategies.
- ✓ **Career Development:** Guides graduates in negotiating better salaries and enhancing employability.

Key Components

- ✓ **Salary Determinants:** Factors impacting salaries like skills, job roles, industries, and locations.
- ✓ **Employment Trends:** Analysing industry-specific trends and emerging roles.
- ✓ **Skill Enhancement:** Recommending skill development and career advancement opportunities.
- ✓ **Salary Negotiation:** Providing tactics for effective salary negotiation.

INNOMATICS RESEARCH LABS

Objective of the Project

The project aims to delve into the factors influencing salary outcomes for engineering graduates using the Aspiring Mind Employment Outcome 2015 (AMEO) dataset. The primary objectives are to identify key determinants of salary, develop predictive models for salary estimation, and provide actionable recommendations for skill enhancement and career advancement.

- ✓ **Identify Key Salary Determinants:** Explore cognitive skills, technical expertise, personality traits, job titles, locations, and demographics influencing salary outcomes. Uncover correlations and patterns within the dataset to understand the impact of these factors.

- ✓ **Provide Actionable Recommendations:** Translate analysis findings into actionable insights and recommendations. Offer strategies for skill enhancement, training programs, certifications, and career growth tailored to engineering graduates.

- ✓ **Enhance Employability and Salary Negotiation:** Empower engineering graduates with the knowledge and tools to negotiate competitive salary packages. Facilitate career development and improve employability in the dynamic job market.

Project Overview

Overview

The project aims to analyse the factors influencing salary outcomes for engineering graduates using the Aspiring Mind Employment Outcome 2015 (AMEO) dataset. Key objectives include identifying salary determinants, developing predictive models, and providing actionable recommendations for career enhancement.

Key Components

Data Collection: Obtain and preprocess the AMEO dataset containing information on cognitive skills, technical expertise, personality traits, job titles, locations, demographics, and salary outcomes.

Exploratory Data Analysis (EDA): Conduct thorough EDA to uncover correlations, trends, and patterns related to salary outcomes and key influencing factors.

Actionable Insights: Translate analysis findings into actionable recommendations for skill enhancement, career development, and salary negotiation strategies tailored to engineering graduates.

Summary of the Data

✓ **Dataset Overview:**

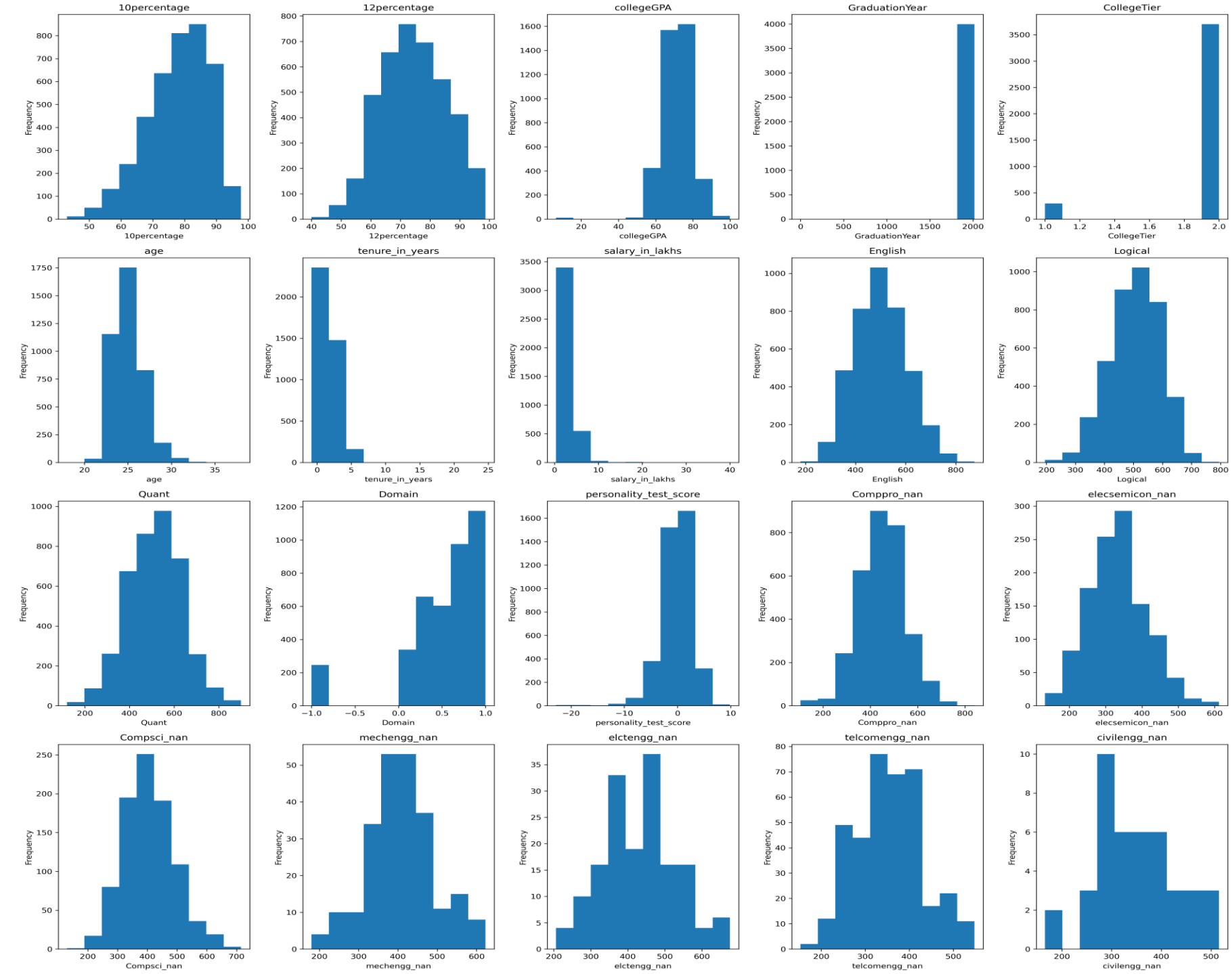
Total data points: 3,998
Variables: 39 columns, including ID, Salary, Date of Joining (DOJ), Date of Leaving (DOL), Designation, Job City, Gender, Date of Birth (DOB), academic performance (10th and 12th percentages), college details (college ID, tier, state), degree information, specialization, GPA, graduation year, and various skill scores (English, Logical, Quantitative, Domain, Computer Programming, etc.).

✓ **Key Data Insights:**

Salary Distribution: Variability in salaries based on designations, job cities, and specialization areas.
Demographic Impact: Gender differences, academic background influence on salary prospects.
Skill Scores Analysis: Correlation between skill scores (English, Logical, Quantitative, etc.) and salary outcomes.
Personality Traits: Influence of personality traits (conscientiousness, agreeableness, etc.) on job roles and salary levels.

✓ **Data Cleaning and Preprocessing:**

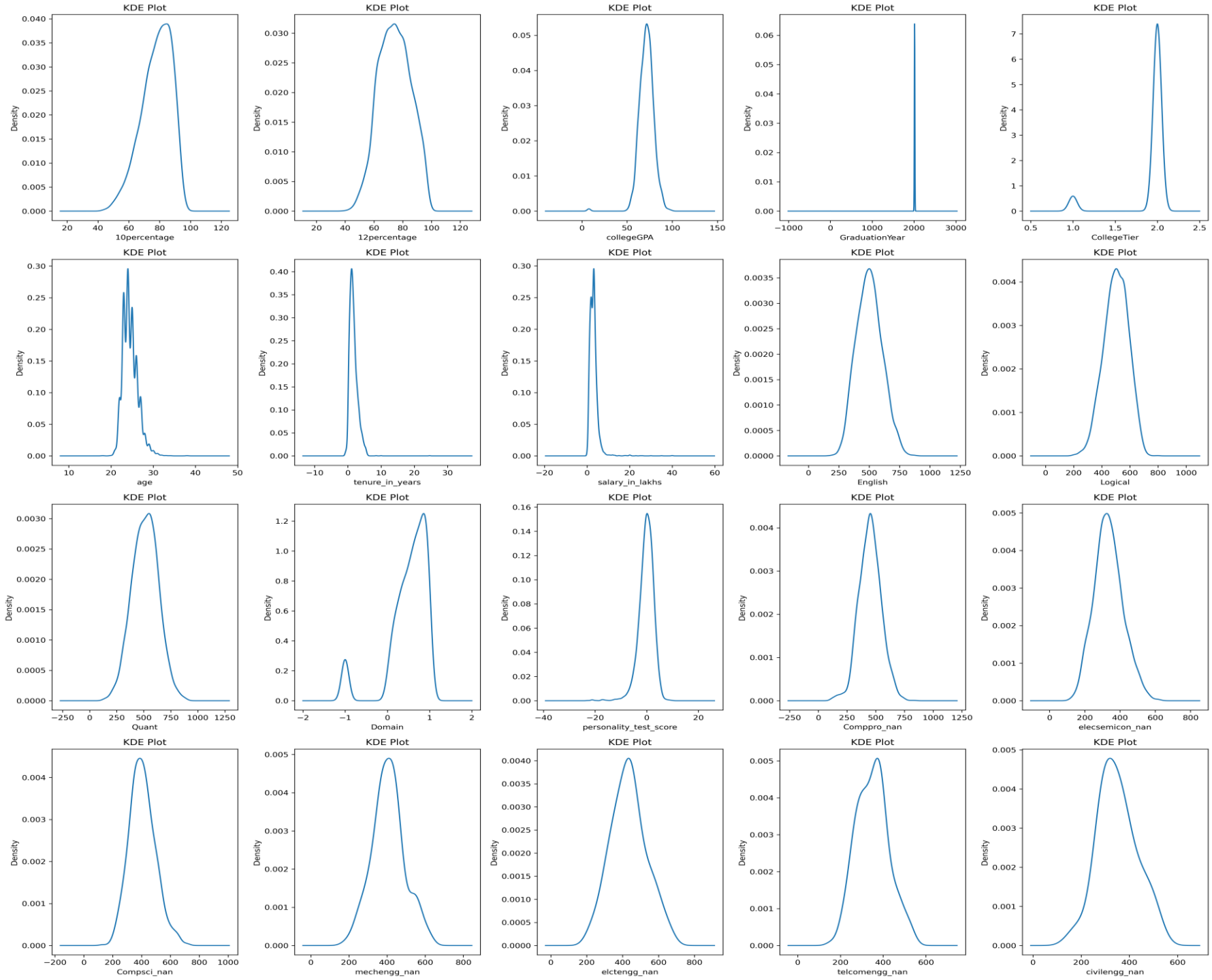
Handling Missing Values: Utilized imputation techniques for missing data in certain columns.
Removing Duplicates: Ensured data integrity by removing duplicate entries.
Standardization: Standardized data formats and encoded categorical variables for analysis.
Outlier Detection: Identified and addressed outliers in relevant columns for accurate analysis.



Univariate Numerical Variables Distributions

Histograms of various numerical columns present in the dataset

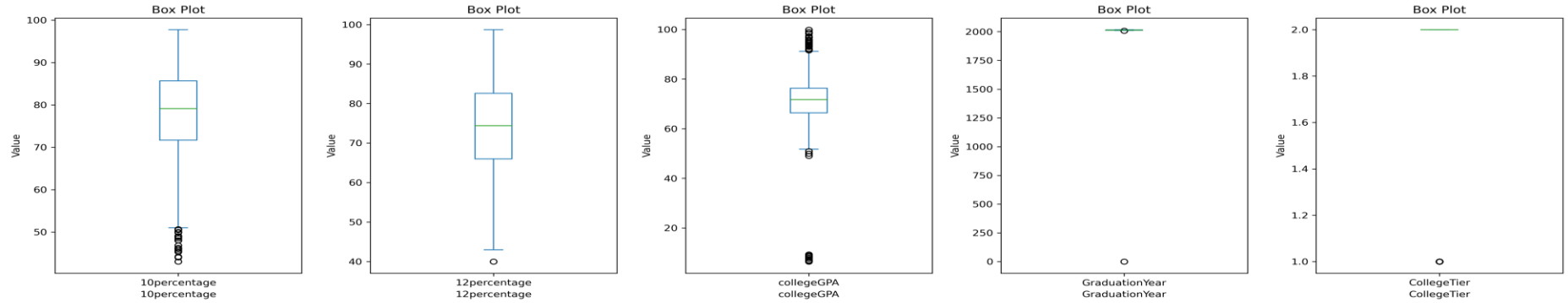
INNOMATICS RESEARCH LABS

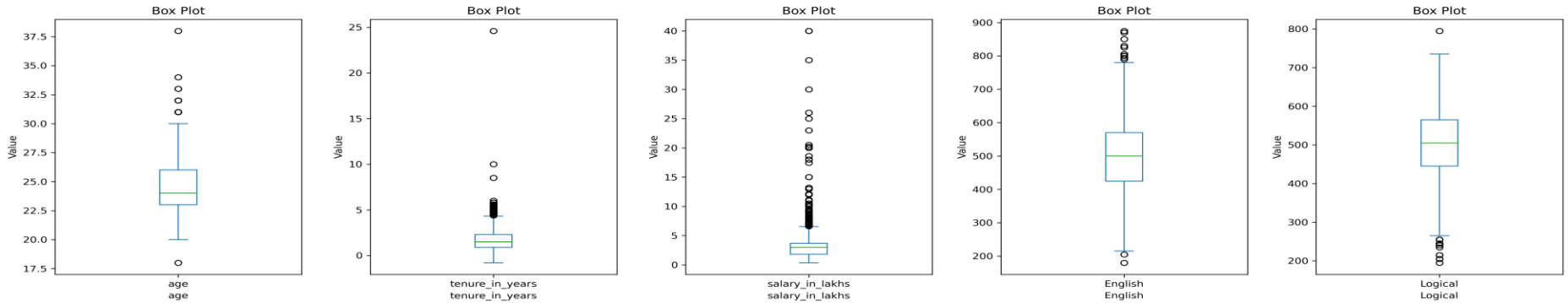


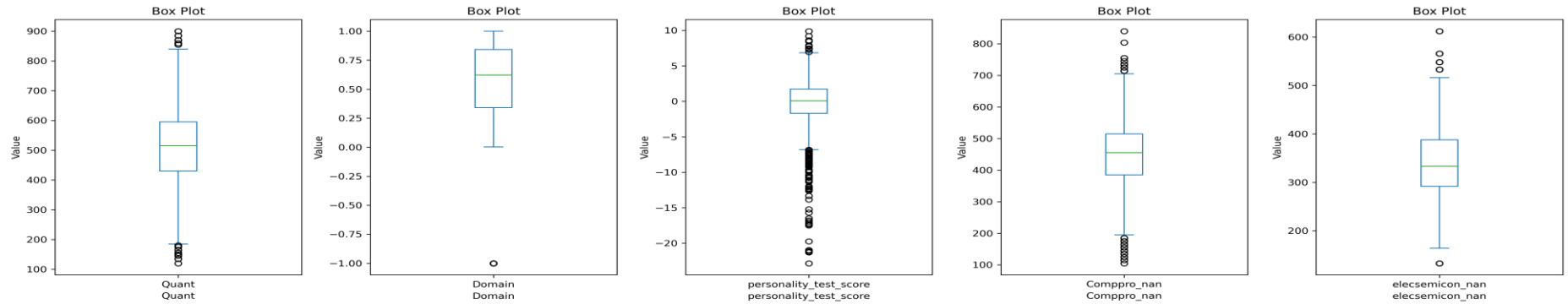
Univariate Numerical Variables Distributions

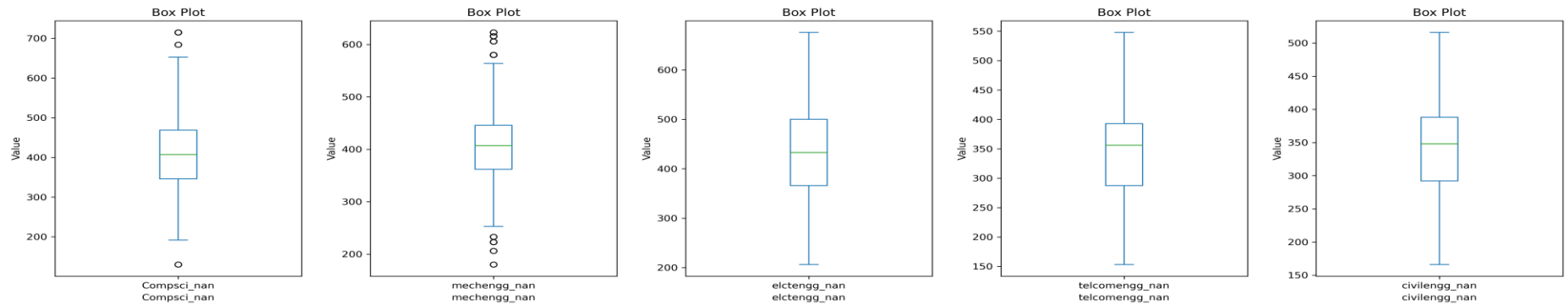
KDE of various numerical columns present in the dataset

INNOMATICS RESEARCH LABS





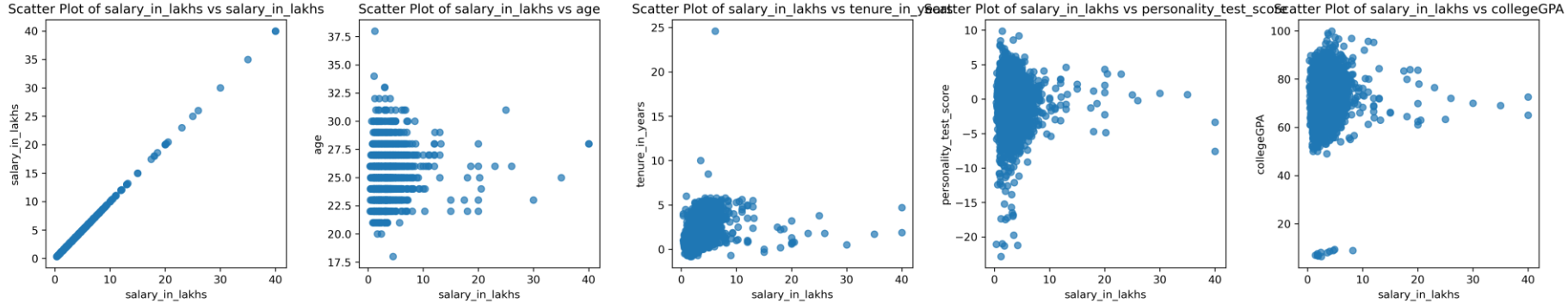


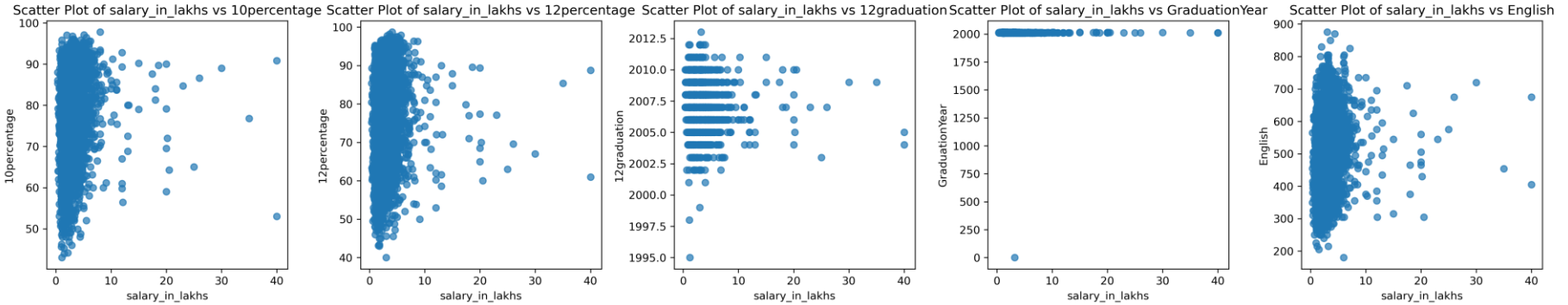


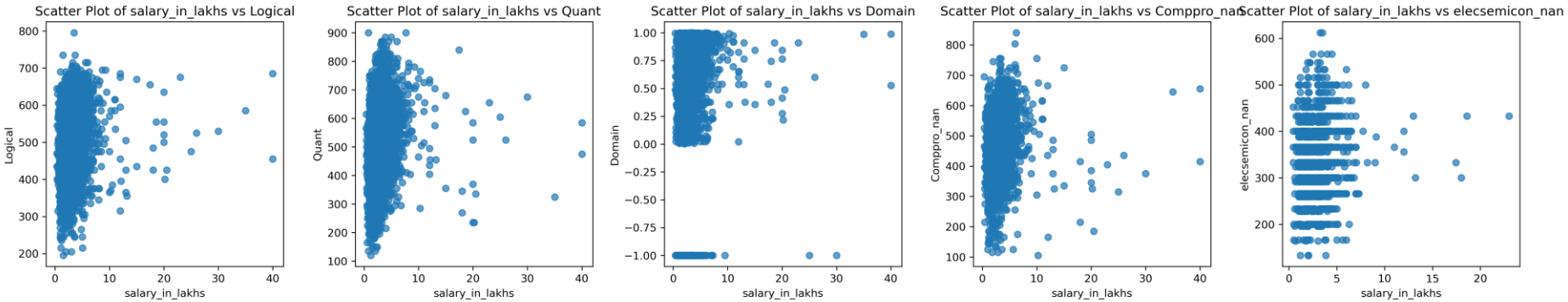
Univariate Numerical Variables Distributions

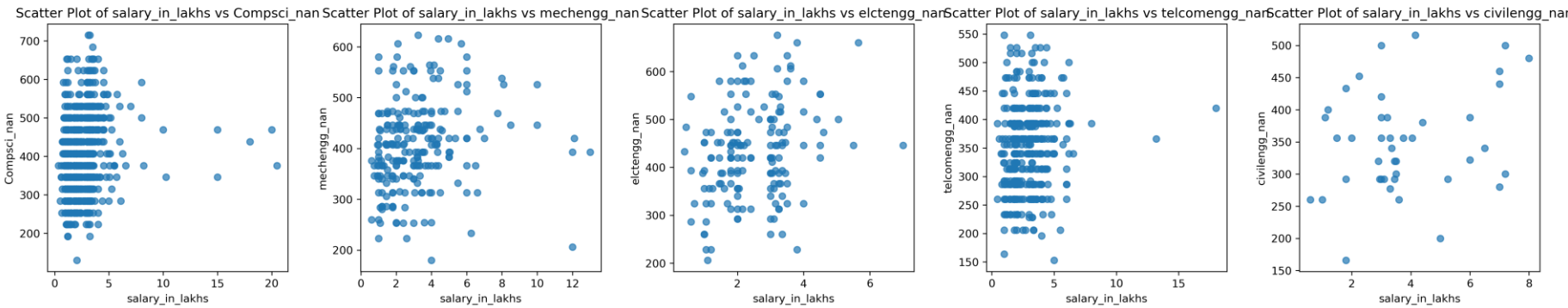
Box plots of various numerical columns present in the dataset

INNOMATICS RESEARCH LABS





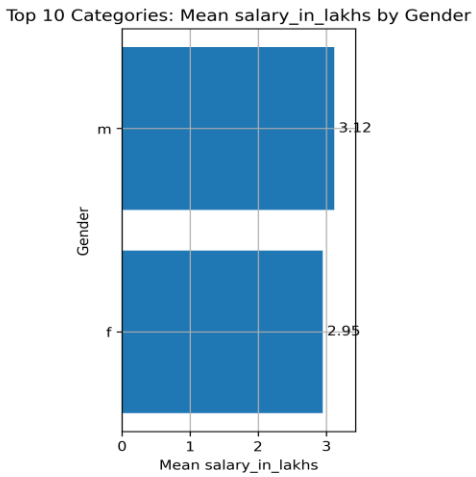


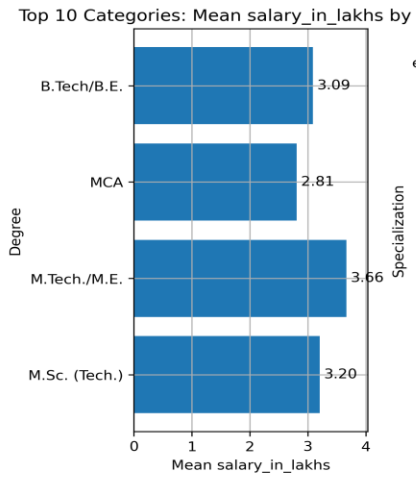


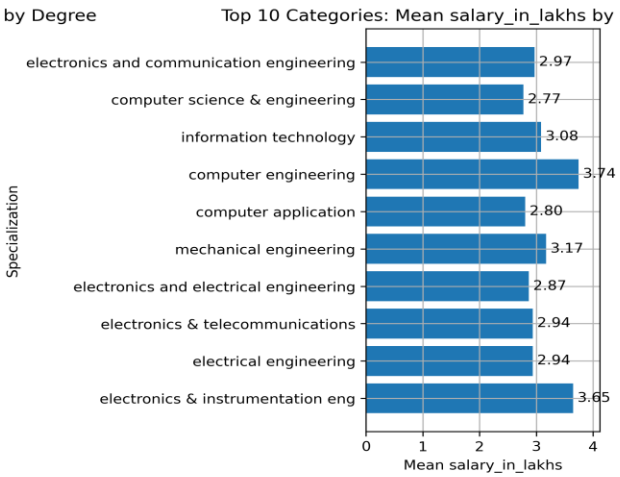
Various Features & Their Relationships With The Target Variable, Salary

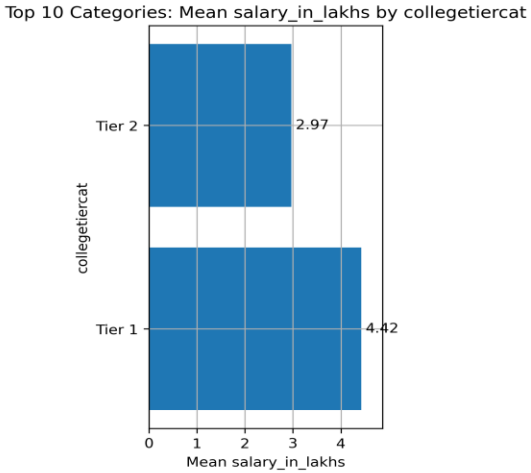
Bivariate Analysis Of Salary With Other Numerical Variables

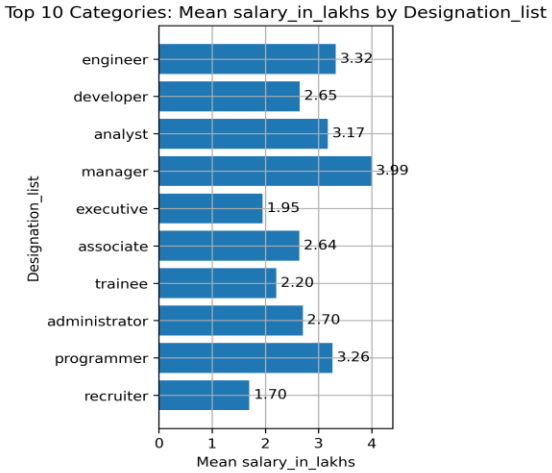
INNOMATICS RESEARCH LABS

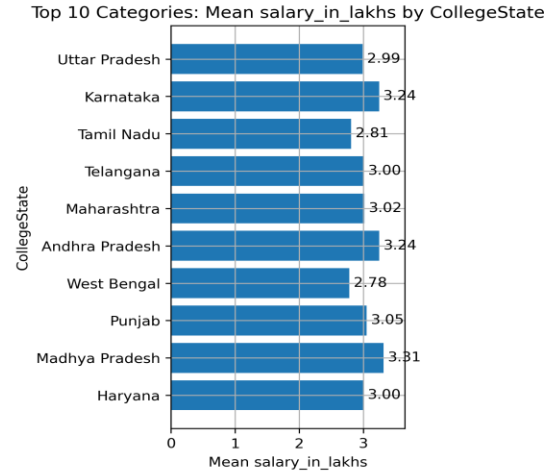


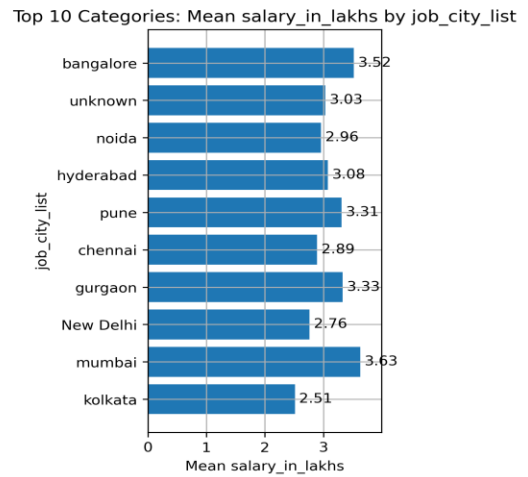


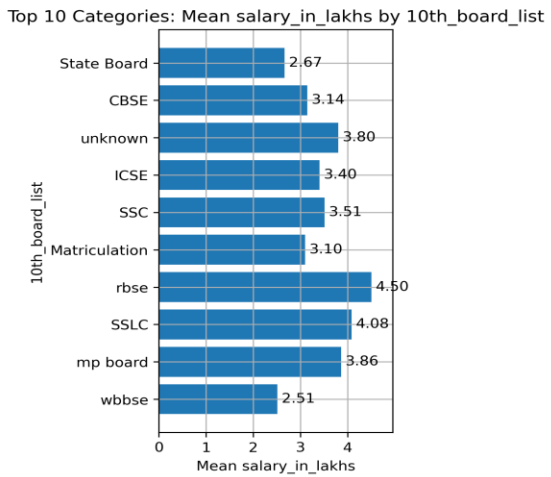


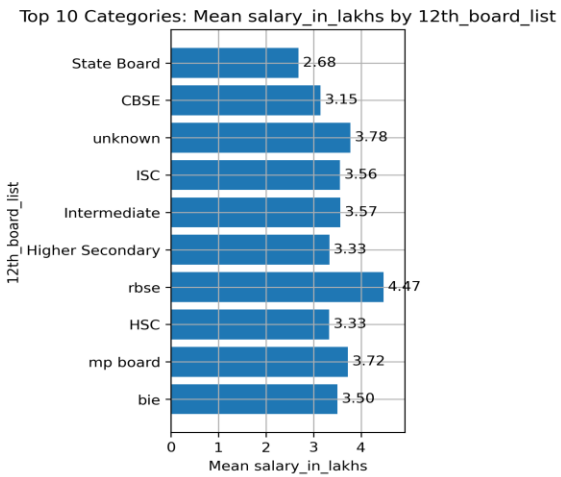








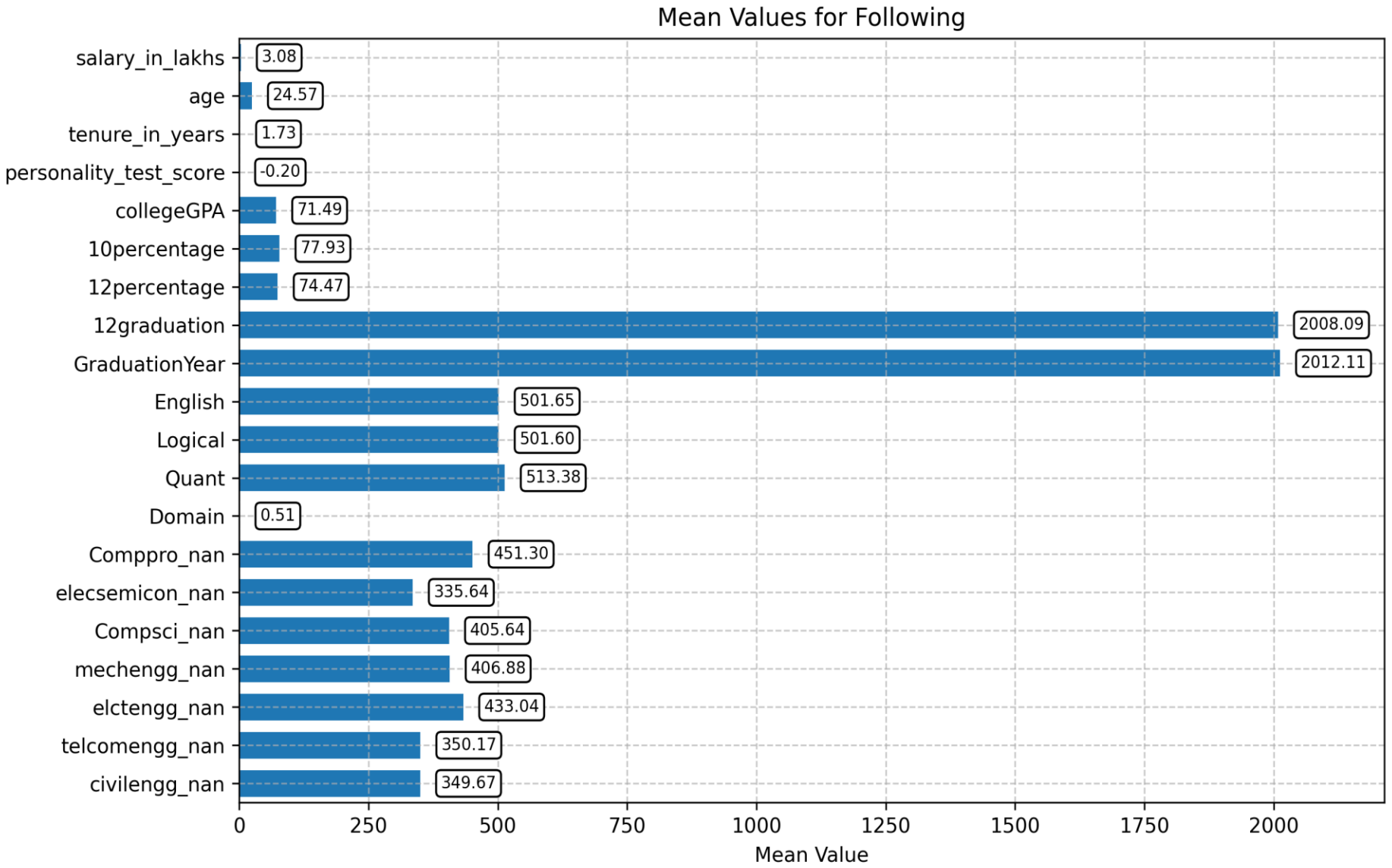




Various Features & Their Relationships With The Target Variable, Salary

Bivariate Analysis Of Salary With Other Categorical Variables

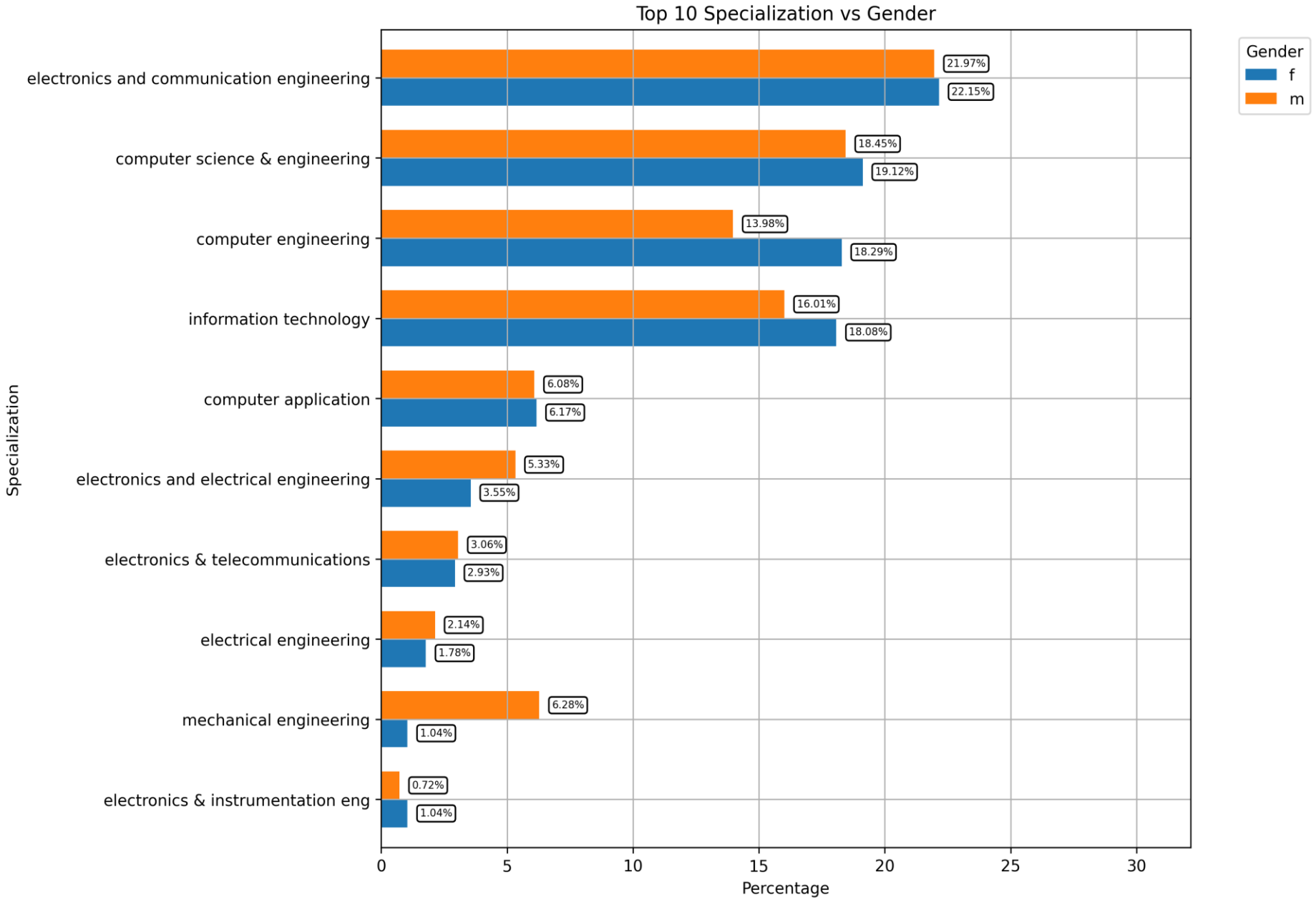
INNOMATICS RESEARCH LABS



Various Numerical Features & Their Mean Values

Bivariate Analysis Of Salary With Other Numerical Variables

INNOMATICS RESEARCH LABS



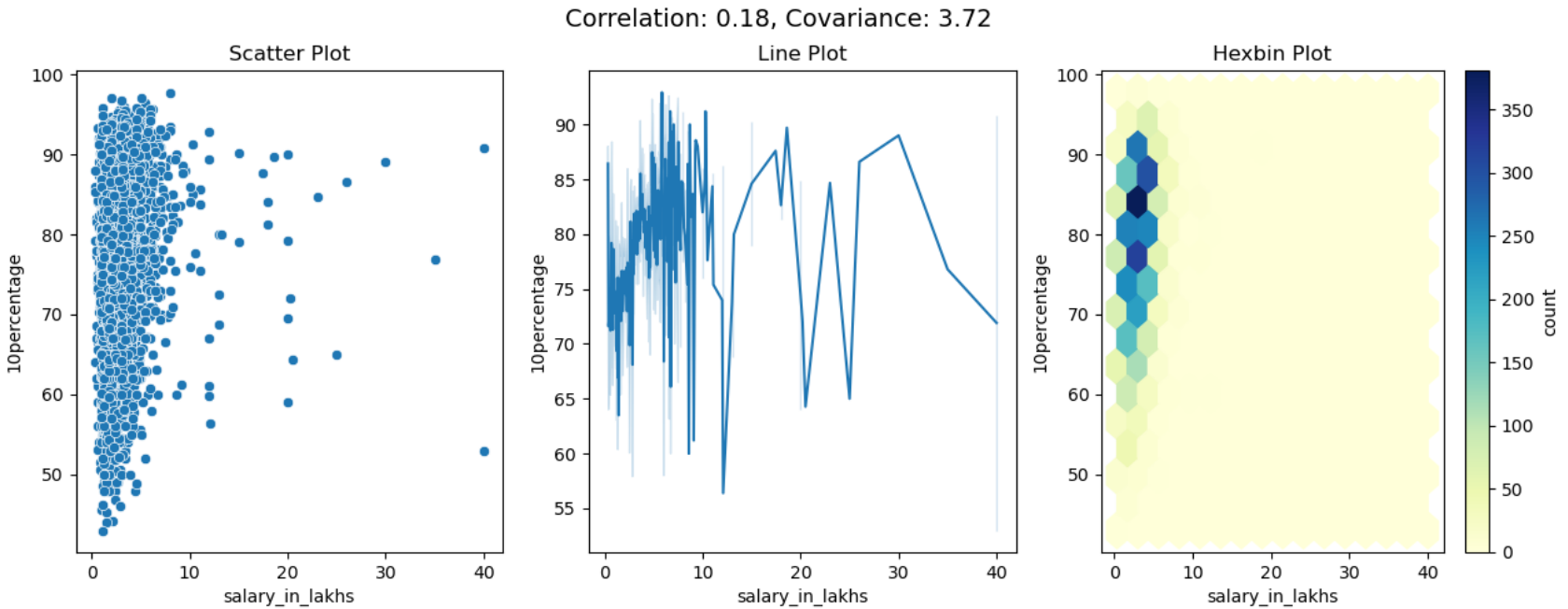
Question

1. Explored the relationship between gender and specialization (Top 10 Specializations)

INNOMATICS RESEARCH LABS

Questions

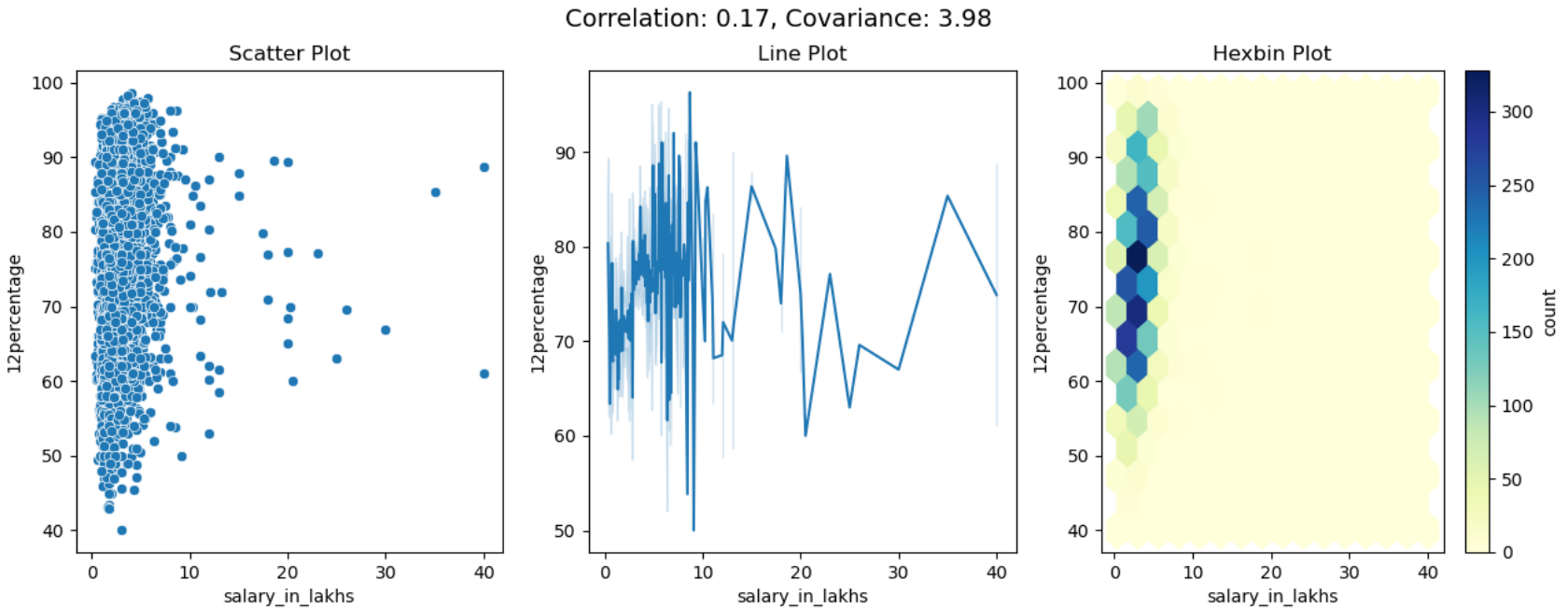
2. Relationship between education and earning potential (Salary Vs 10th Percentages)



INNOMATICS RESEARCH LABS

Questions

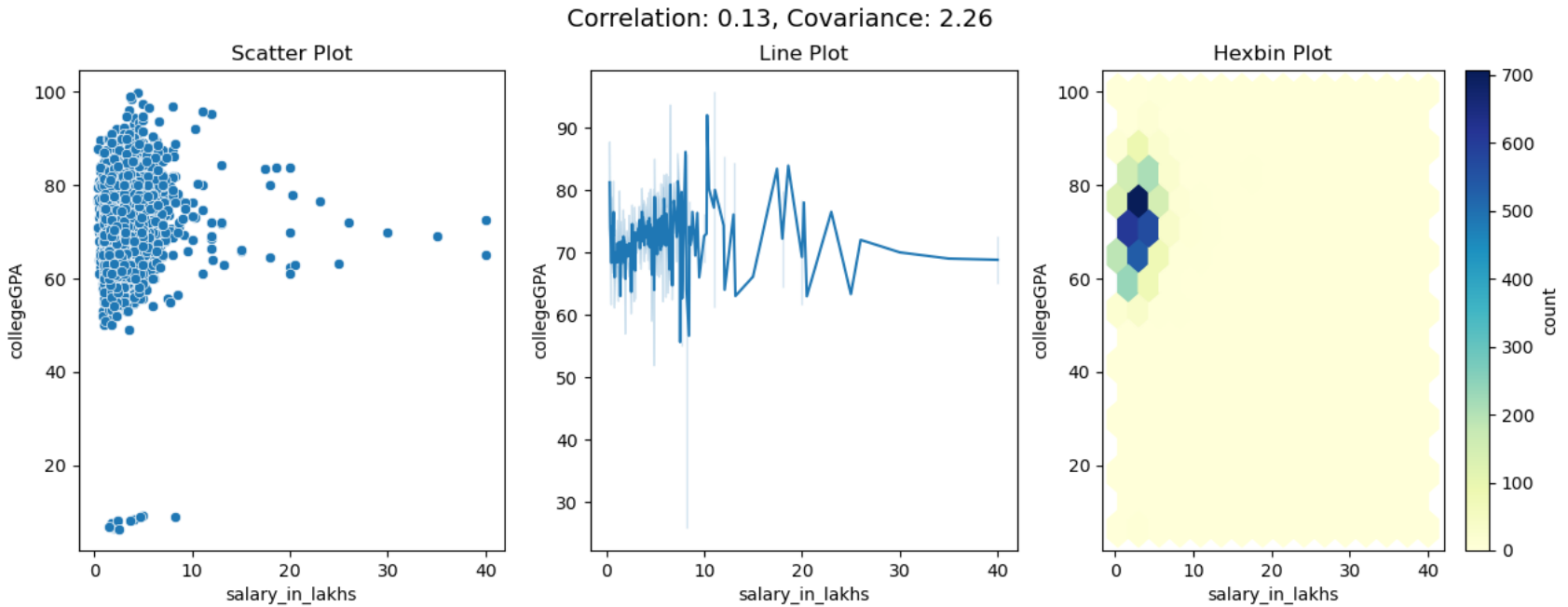
2. Relationship between education and earning potential (Salary Vs 12th Percentages)



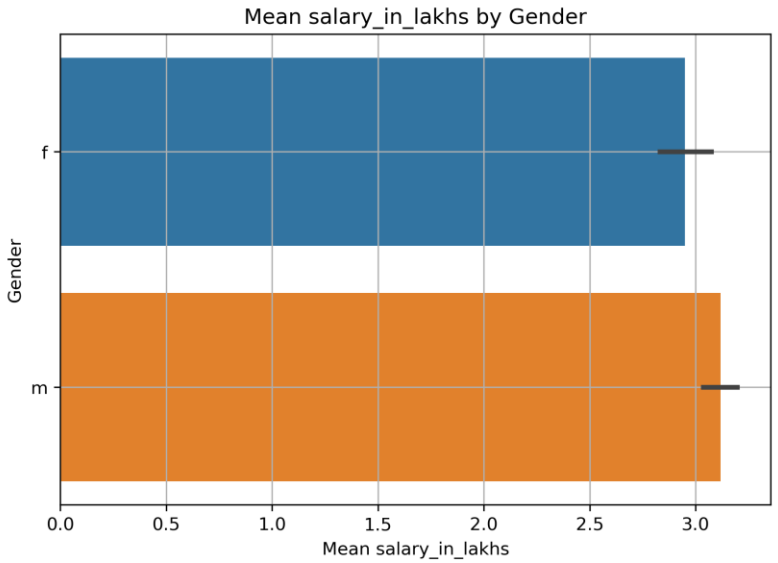
INNOMATICS RESEARCH LABS

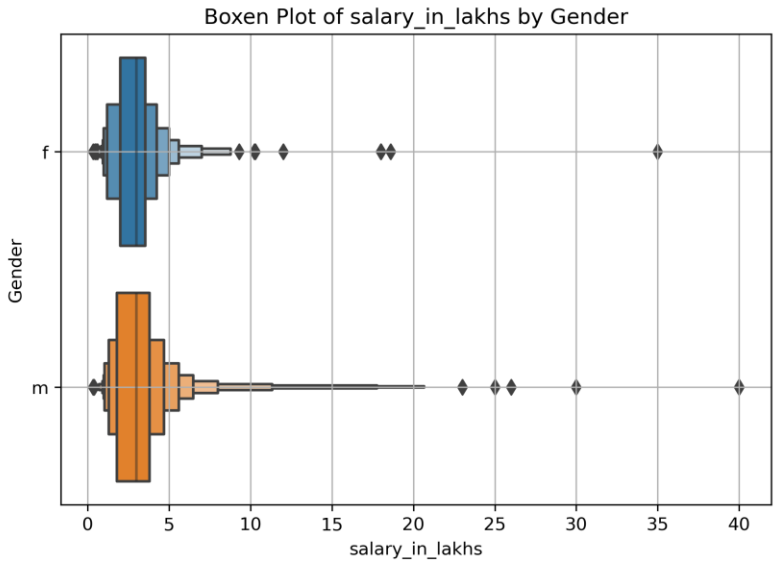
Questions

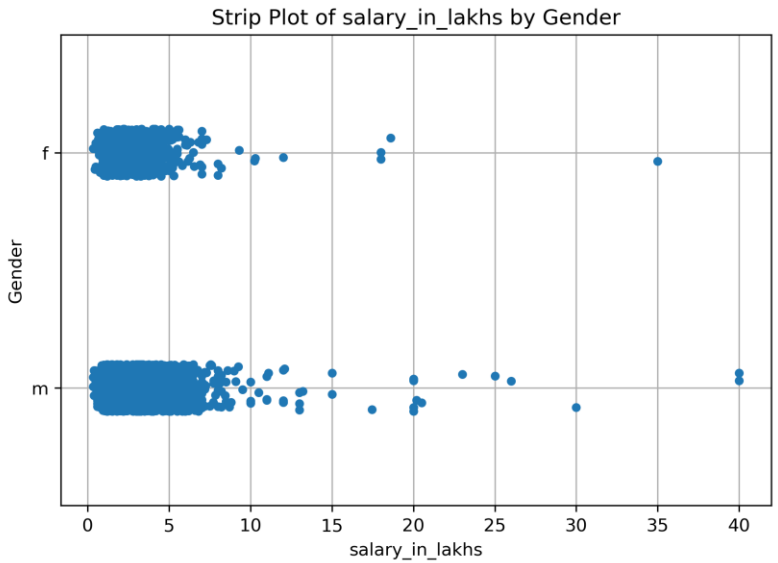
2. Relationship between education and earning potential (Salary Vs CollegeGPA)



INNOMATICS RESEARCH LABS



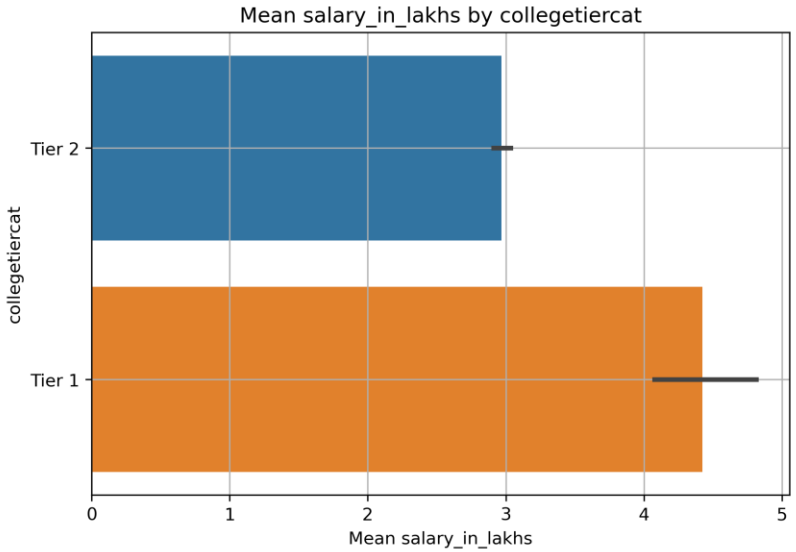


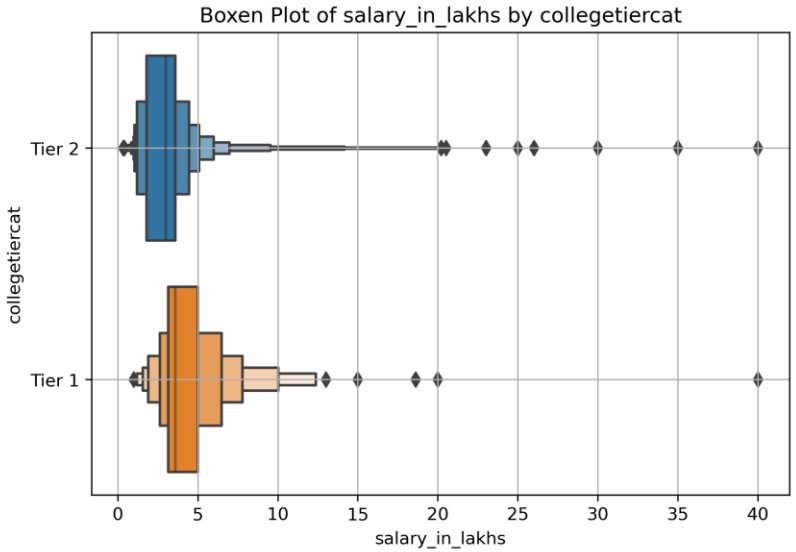


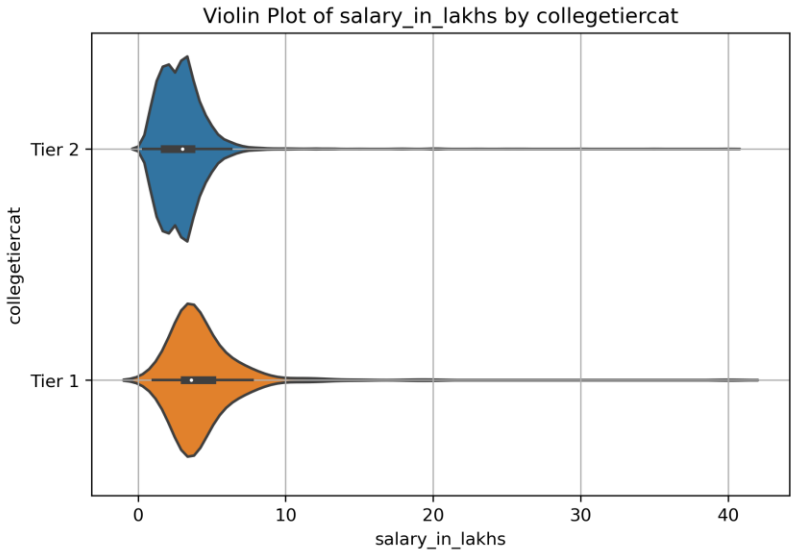
Question

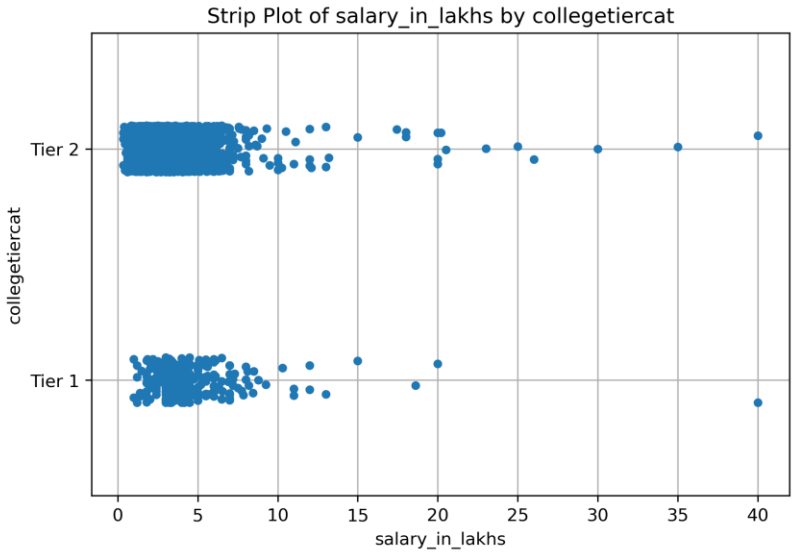
2. Explored the relationship between gender and salary

INNOMATICS RESEARCH LABS





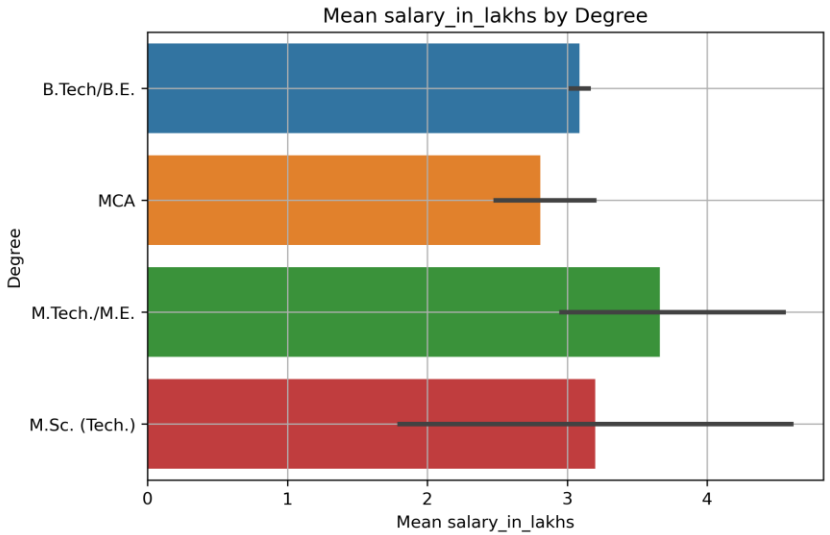


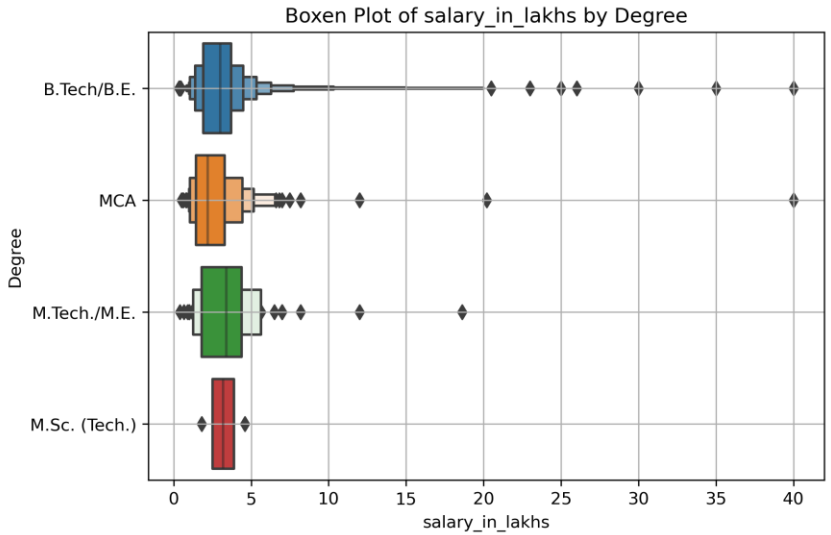


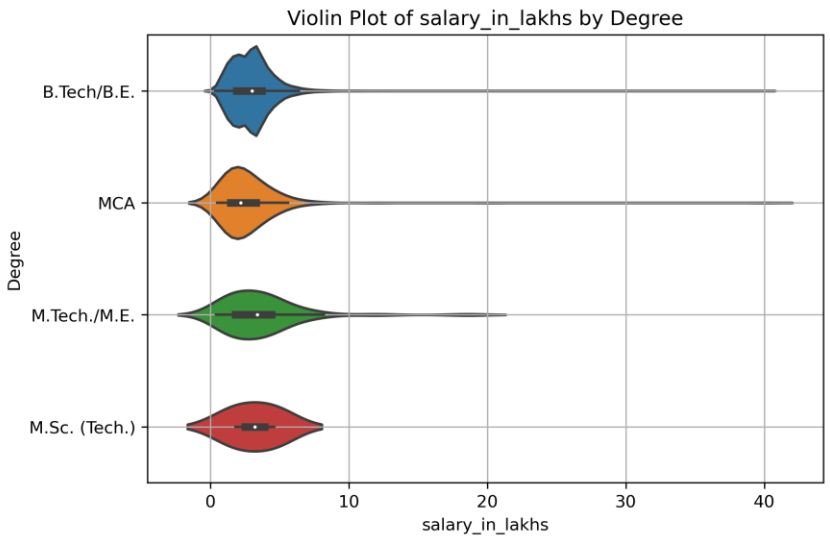
Question

2. Relationship between education and earning potential (Salary Vs CollegeTier)

INNOMATICS RESEARCH LABS



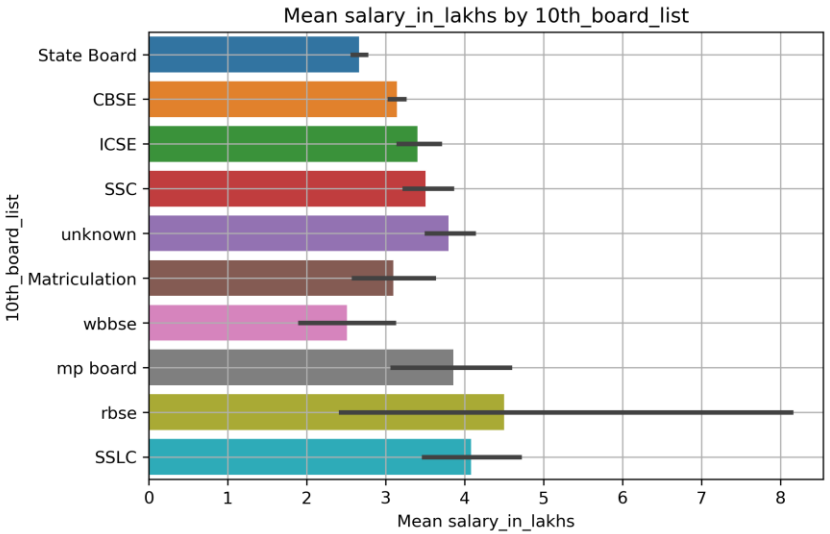


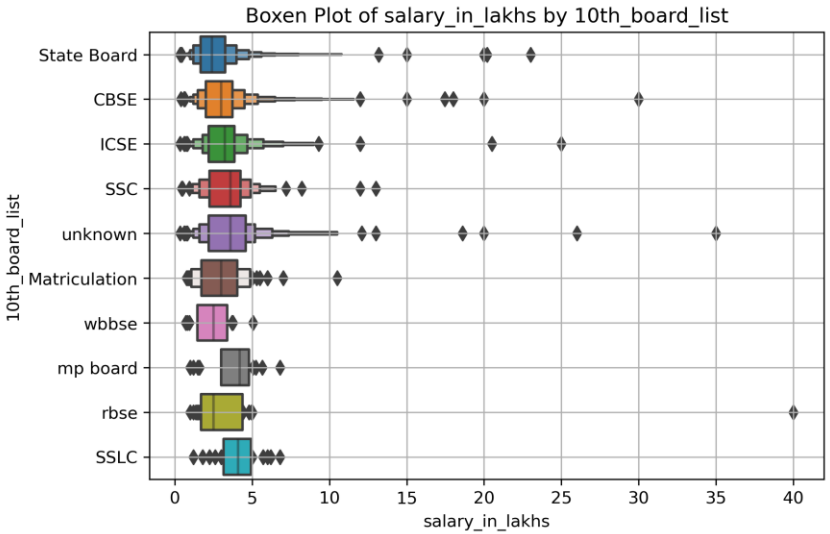


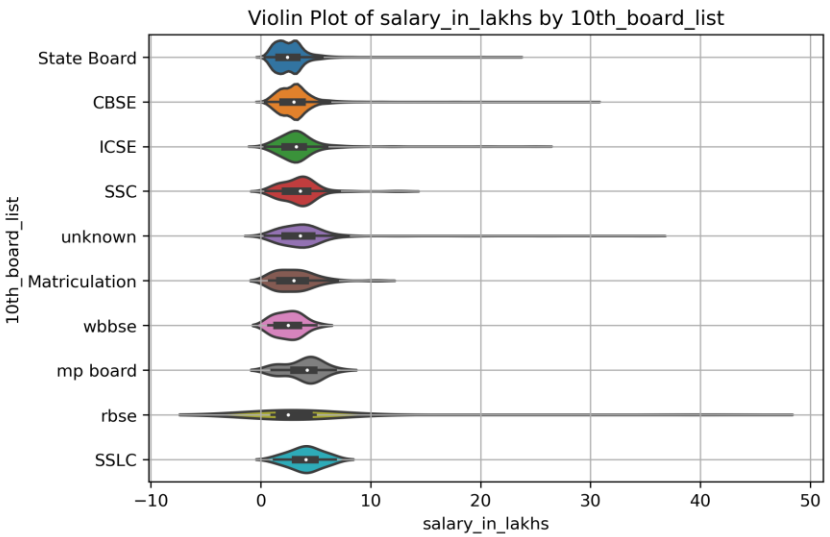
Question

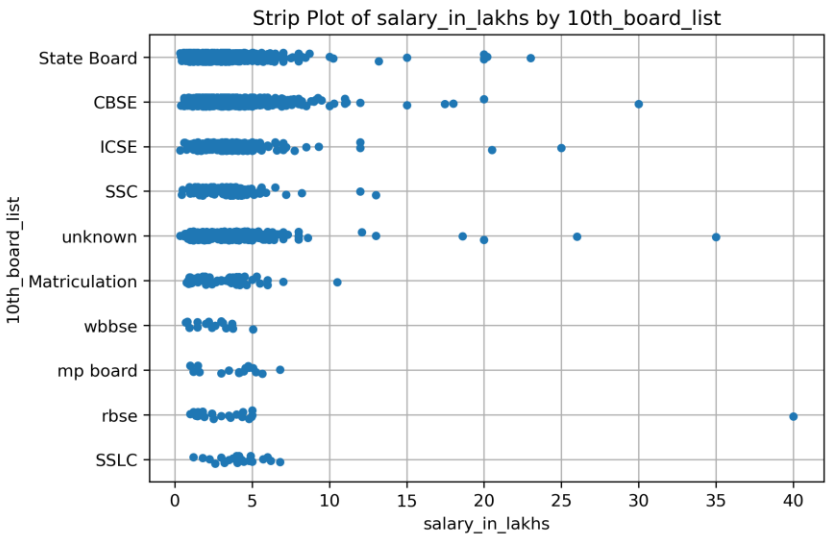
2. Relationship between education and earning potential (Salary Vs Degree)

INNOMATICS RESEARCH LABS





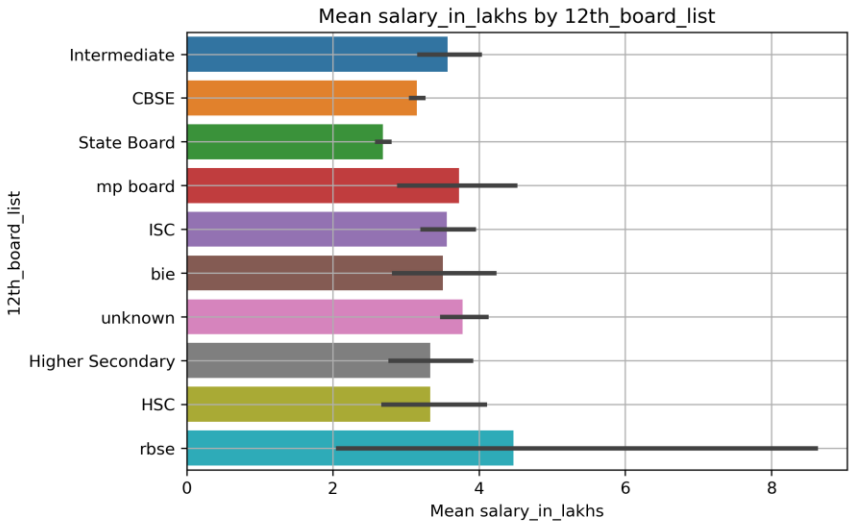


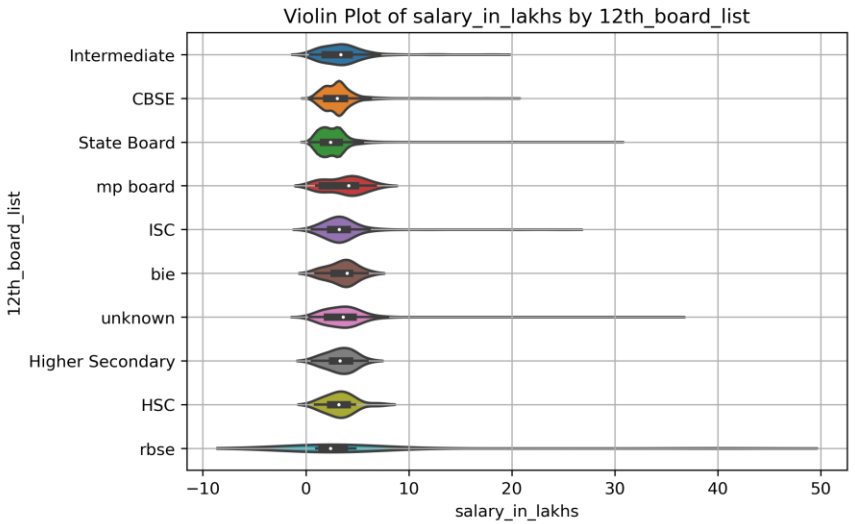


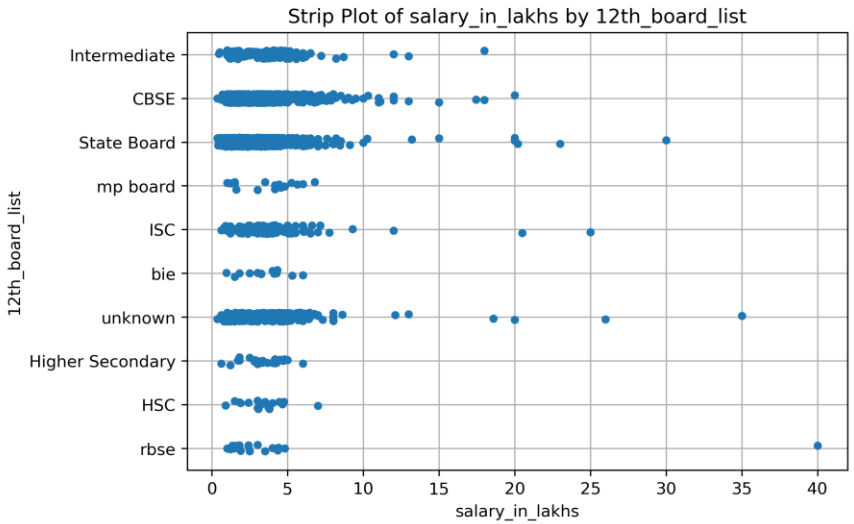
Question

2. Relationship between education and earning potential (Salary Vs 10th Boards) Top10

INNOMATICS RESEARCH LABS







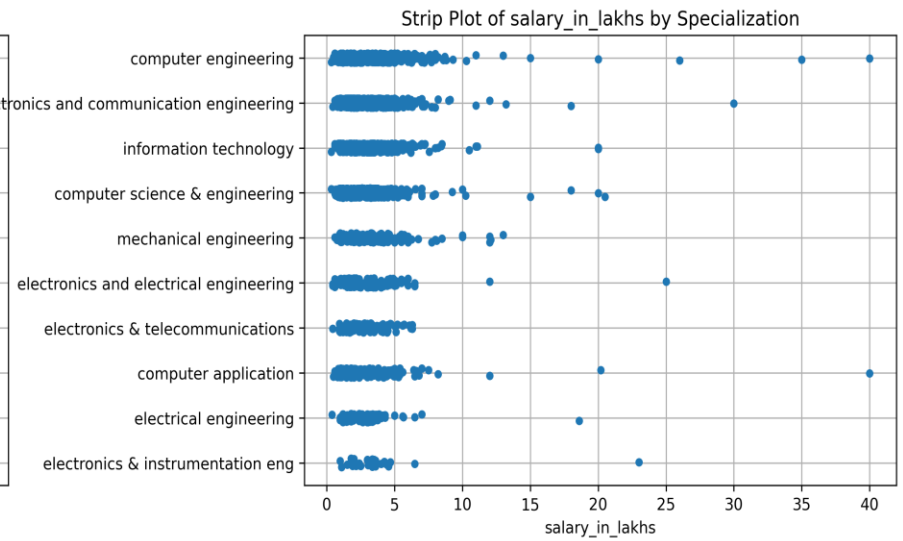
Question

2. Relationship between education and earning potential (Salary Vs 12th Boards) Top 10

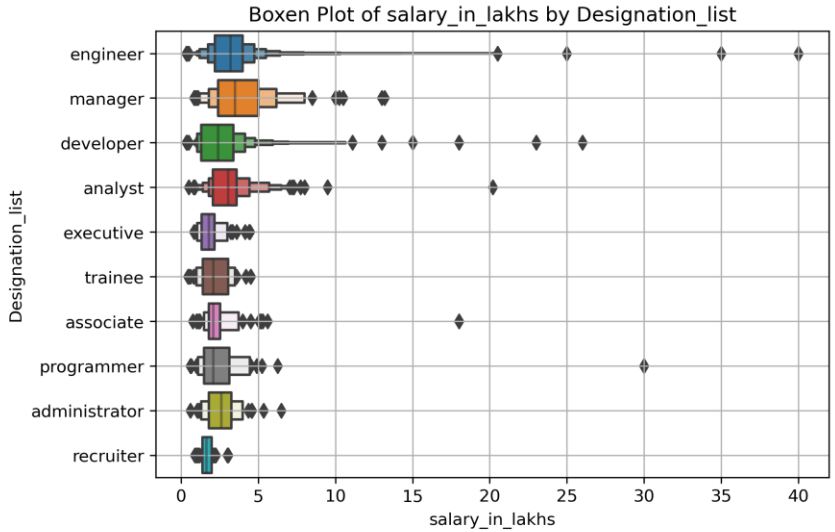
INNOMATICS RESEARCH LABS

Salary Vs Specializations

Top 10

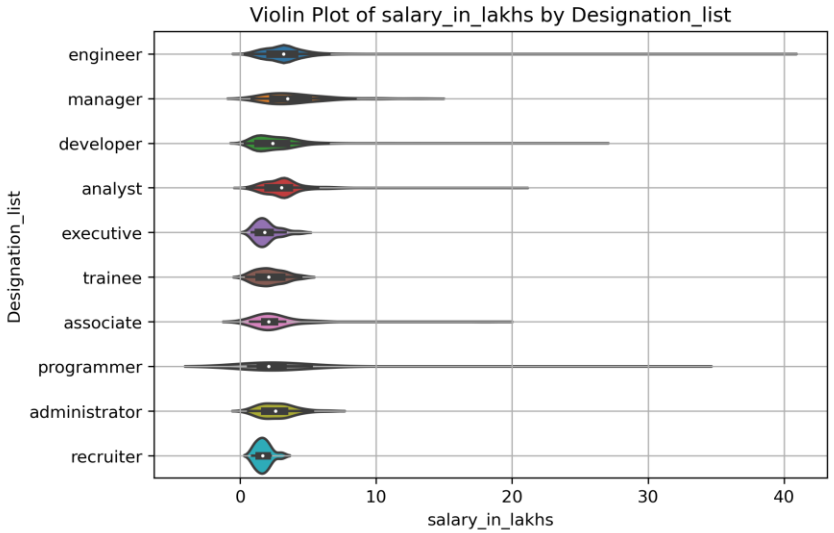


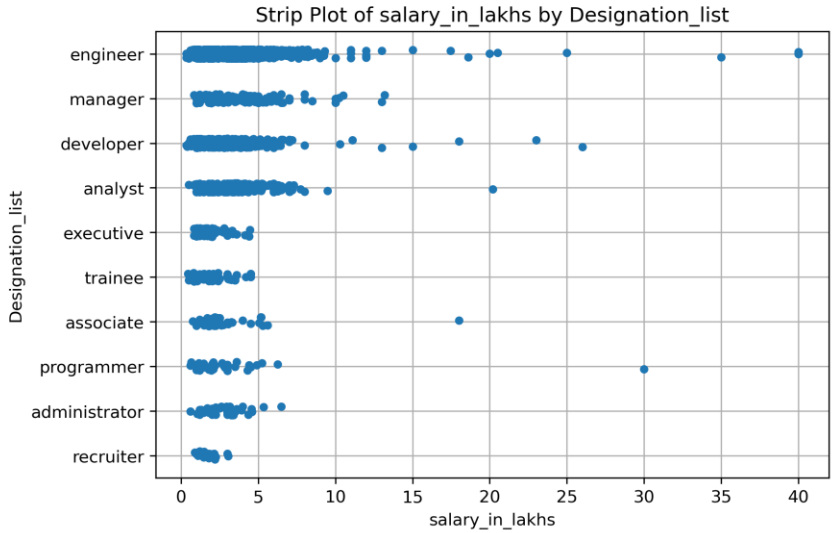
INNOMATICS RESEARCH LABS



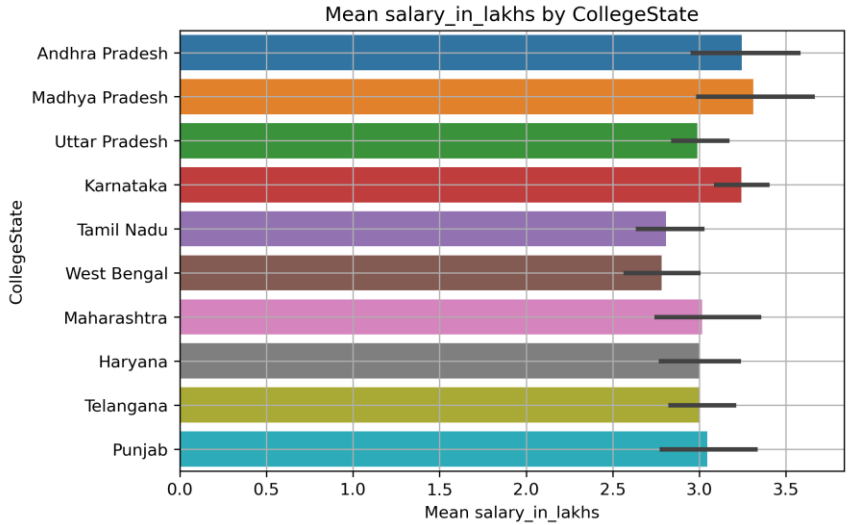
Salary Vs Designations

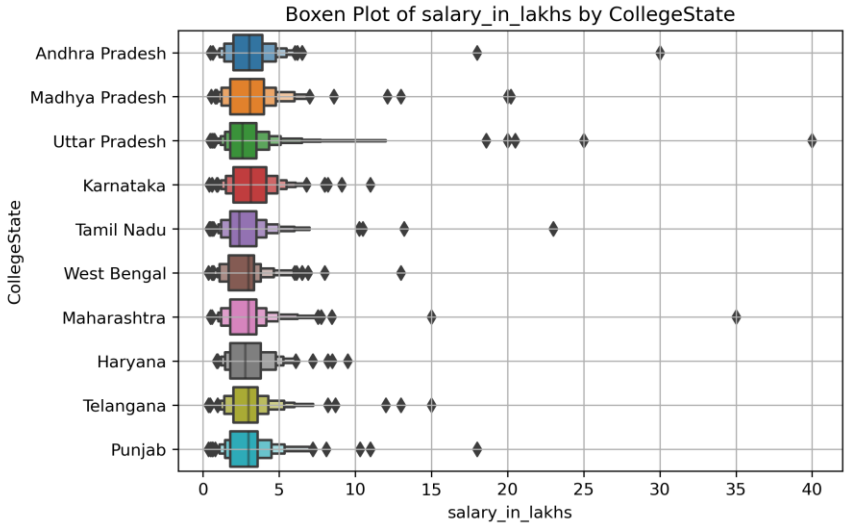
Top 10





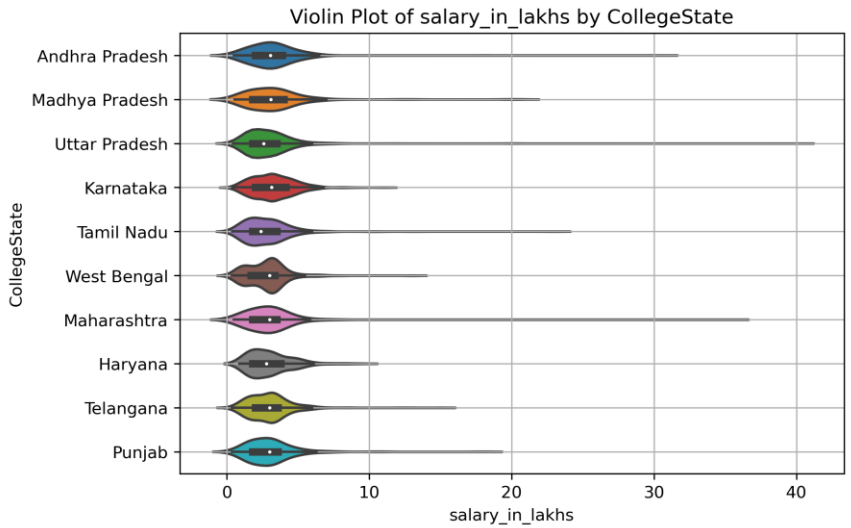
INNOMATICS RESEARCH LABS

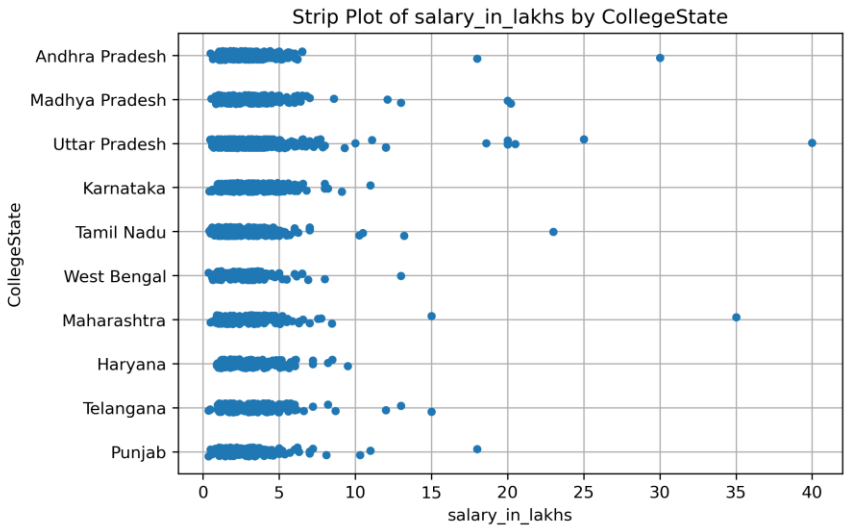




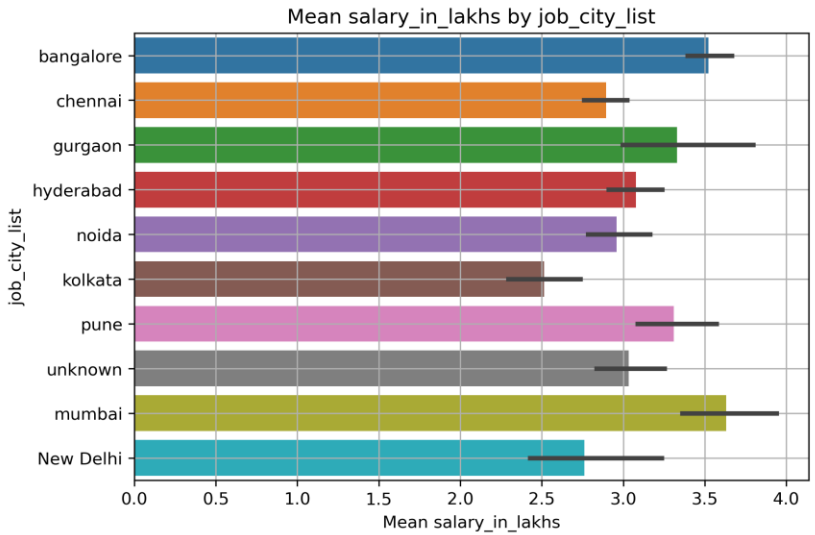
Salary Vs CollegeStates

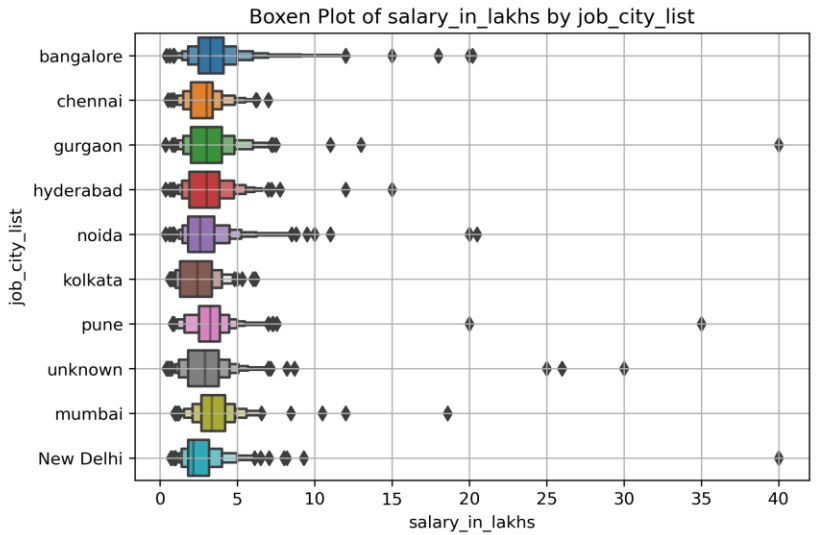
Top 10





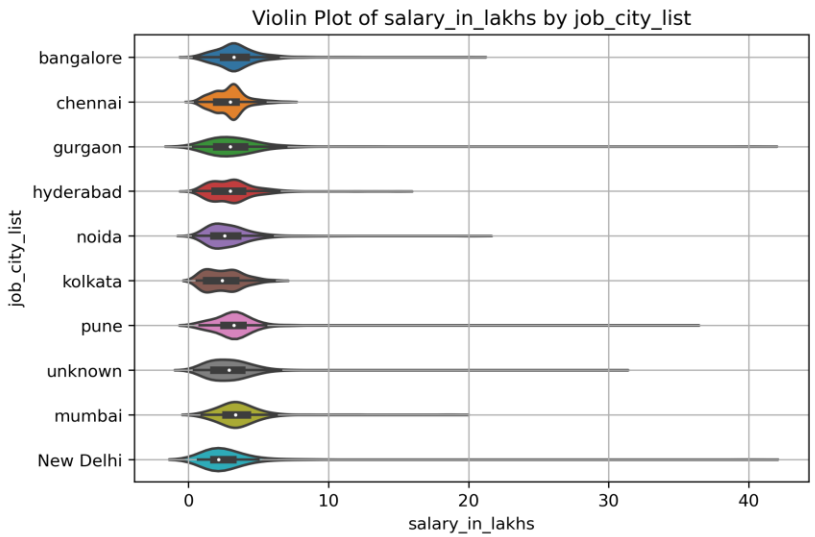
INNOMATICS RESEARCH LABS



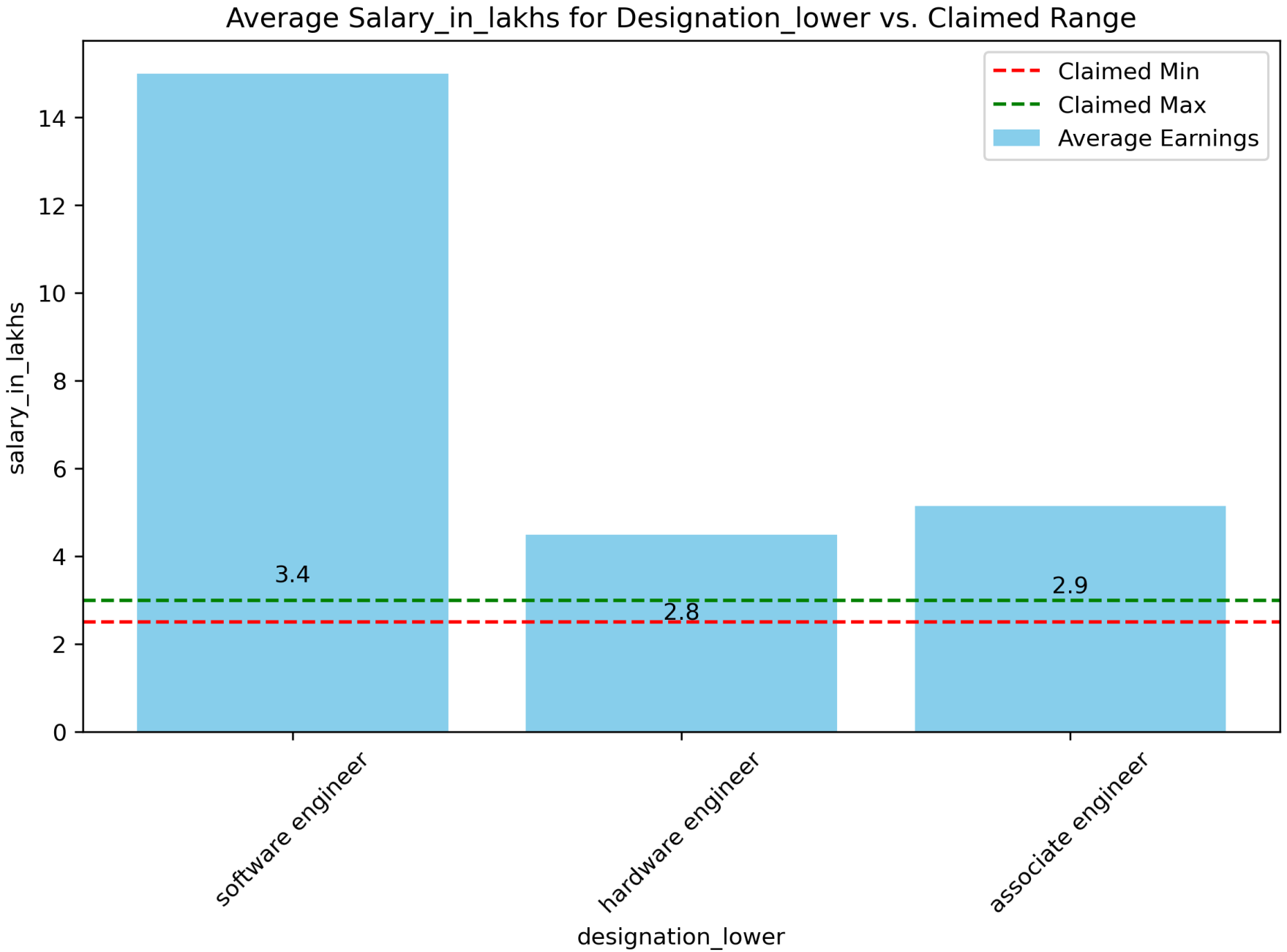


Salary Vs JobCity

Top 10



INNOMATICS RESEARCH LABS



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.

Claim Is True, They Earn More

INNOMATICS RESEARCH LABS

Conclusion

Successfully conducted data analysis on the AMEO dataset to understand salary determinants for engineering graduates. Key findings and actionable recommendations were derived, empowering graduates with insights for career growth and salary negotiation. Future steps involve refining models and sharing insights for broader impact.

- ✓ **Cognitive Skills and Technical Expertise:** Both cognitive skills and technical expertise have a significant impact on salary outcomes for engineering graduates. Continuous skill development in these areas is crucial for enhancing employability and salary prospects.

- ✓ **Personality Traits:** Certain personality traits also contribute to salary variations. Building strong communication, leadership, and problem-solving skills can positively influence salary negotiations.

- ✓ **Demographic Factors:** Job titles, locations, and industry sectors play a crucial role in determining salary prospects. It's important for graduates to consider these factors when planning their careers.

- ✓ **Predictive Modelling:** The developed predictive model accurately estimates salaries based on candidate attributes, providing actionable insights for salary negotiation and career planning.

Impact and Future Directions:

- ✓ The project's insights can empower engineering graduates to make informed career decisions, negotiate better salary packages, and navigate the competitive job market.

- ✓ Future directions may include expanding the analysis to include additional variables or exploring industry-specific salary trends for more targeted recommendations.

INNOMATICS RESEARCH LABS

THANK YOU

INNOMATICS
RESEARCH LABS