



INNOVATION. AUTOMATION. ANALYTICS

PROJECT ON

Exploratory Data Analysis of AMEO Dataset

Thirumalai G (IN9240382)

2ND OCTOBER 2024

About me

Hello everyone, I'm **Thirumalai G**, a pre-final year student pursuing a B.Tech in Artificial Intelligence & Data Science at **Panimalar Engineering College**. Throughout my academic journey, I've built a solid foundation in AI, data science, and software development, applying these skills to tackle real-world problems.

One of my key projects is a **DeepFake Detection System**, where I utilized video data from Meta's Deepfake Detection Challenge Dataset and trained a custom deep learning model using PyTorch. This project demonstrates my expertise in AI frameworks, model training, and data preprocessing, achieving impressive accuracy. I've also worked on an Object Detection System using YOLOv8, which efficiently detects and labels objects in real-time video streams.

Technically, I am proficient in **Python, C, C++, and Java**, with hands-on experience in using libraries like Pandas, NumPy, and Scikit-learn for data analysis, and frameworks like TensorFlow and PyTorch for deep learning. I have developed various AI-powered applications, including sales prediction systems and real-time video object detection tools.

I have secured the runner-up position in the SRMIST Ideathon 2024 for my contributions to AI projects and have served as a coordinator for paper conferences, where I managed communication between participants and the organizing committee. Additionally, I have participated in national-level hackathons and coding competitions, consistently ranking among the top teams.

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.

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.

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.
- ✓ **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.
- ✓ **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.

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.

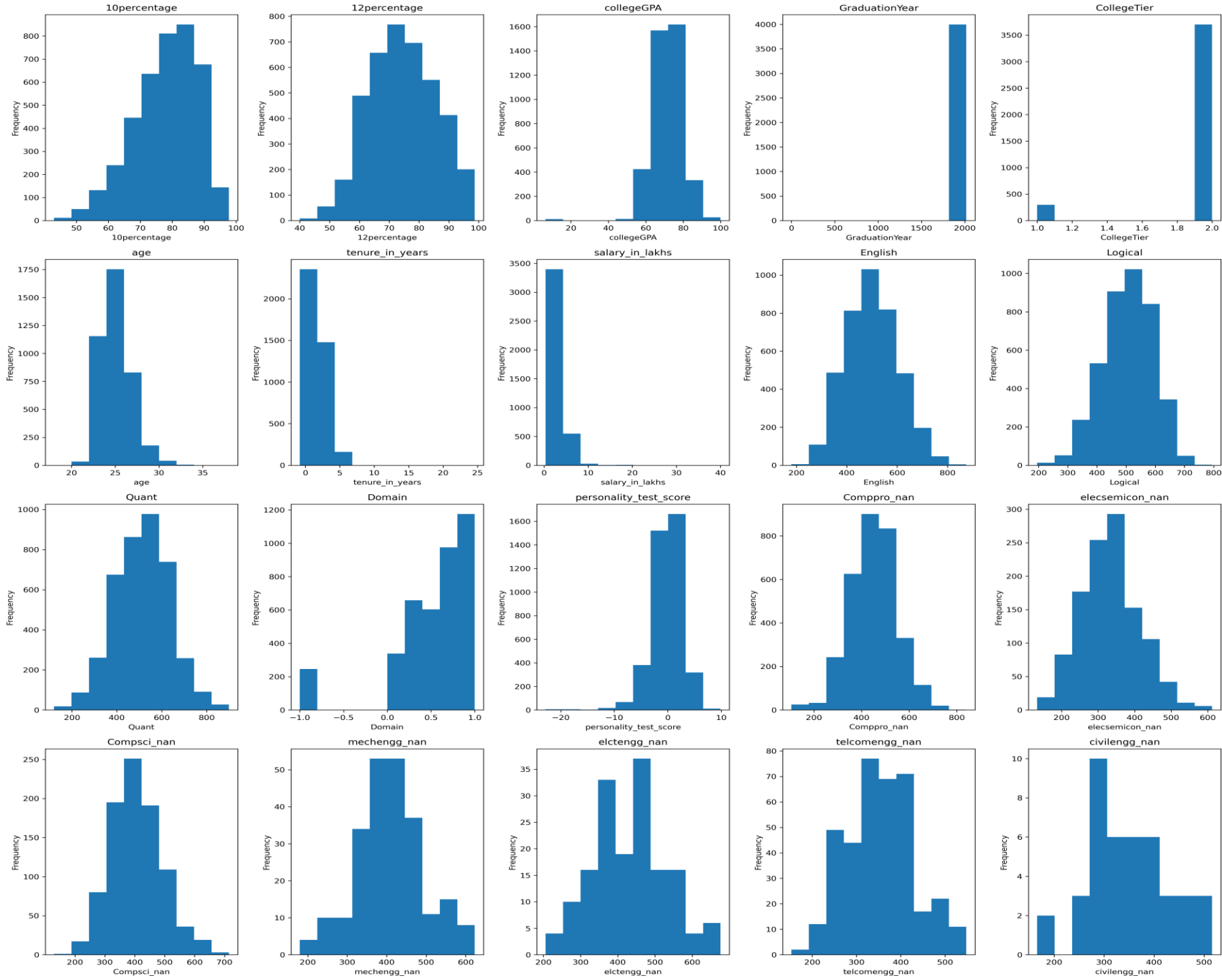
✓ Data Cleaning and Preprocessing:

Handling Missing Values: Utilized imputation techniques for missing data in certain columns.

Removing Duplicates: Ensured data integrity by removing duplicate entries.

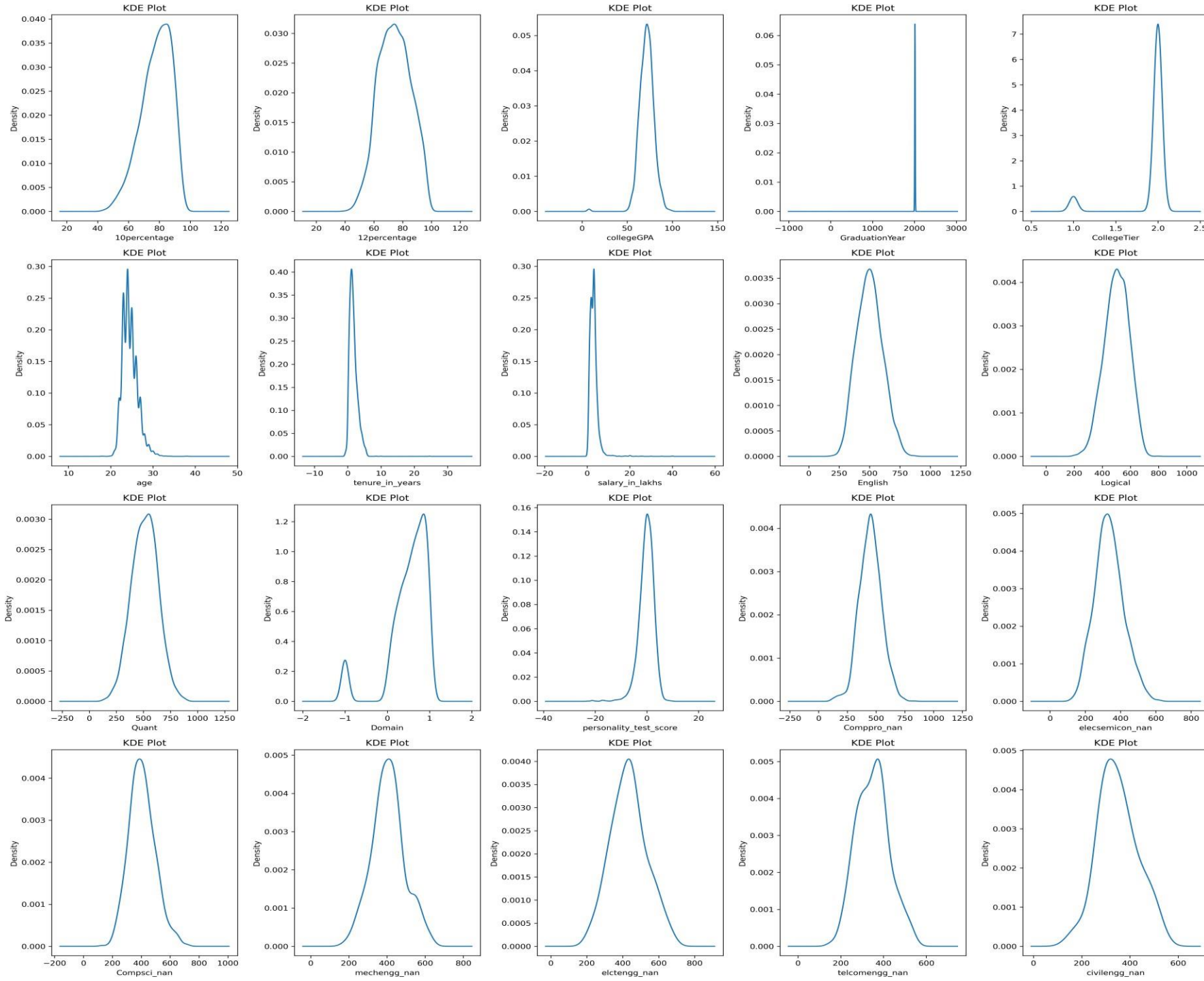
Standardization: Standardized data formats and encoded categorical variables for analysis.

Outlier Detection: Identified and addressed outliers in relevant columns for accurate analysis.



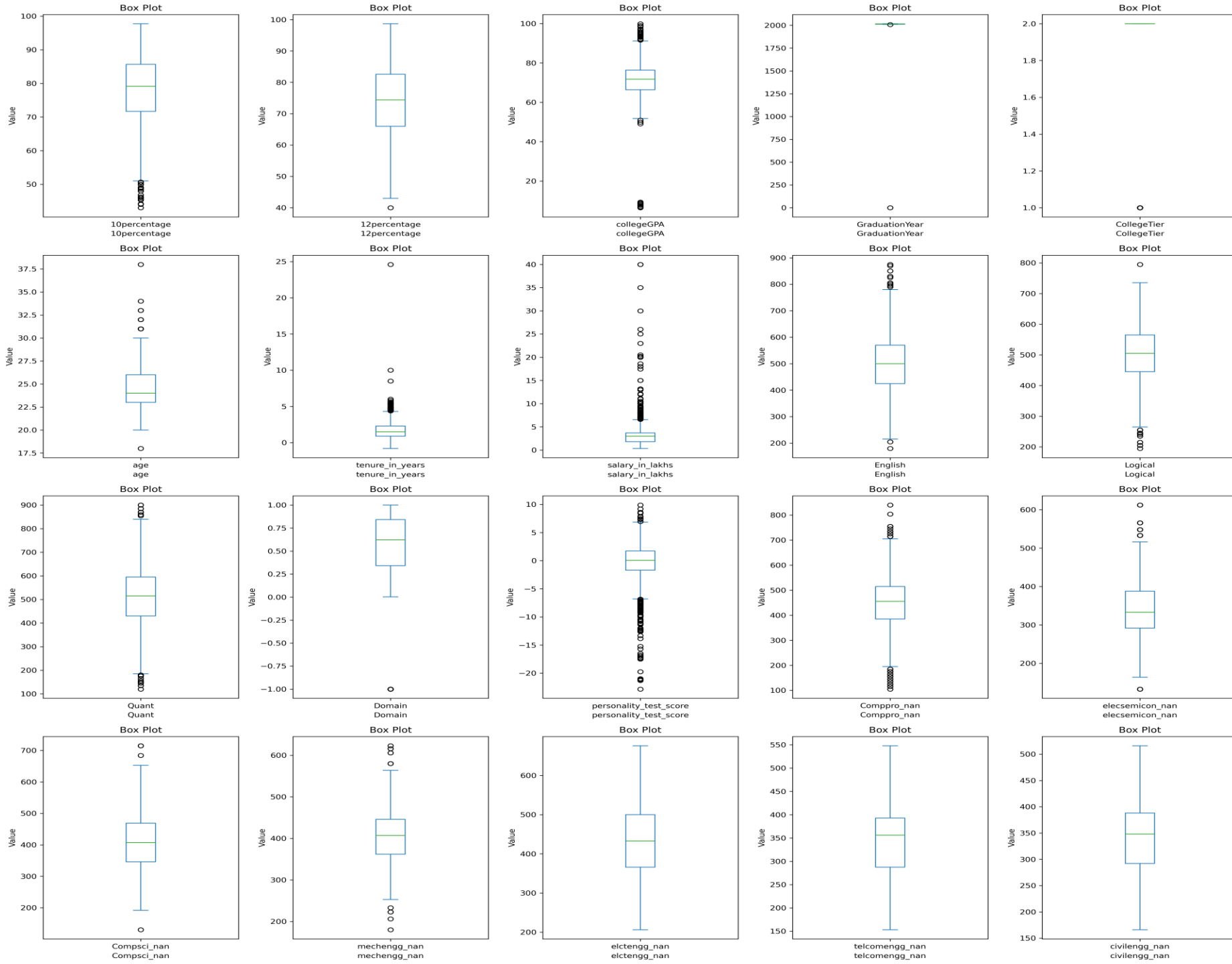
Univariate Numerical Variables Distributions

Histograms of various numerical columns present in the dataset



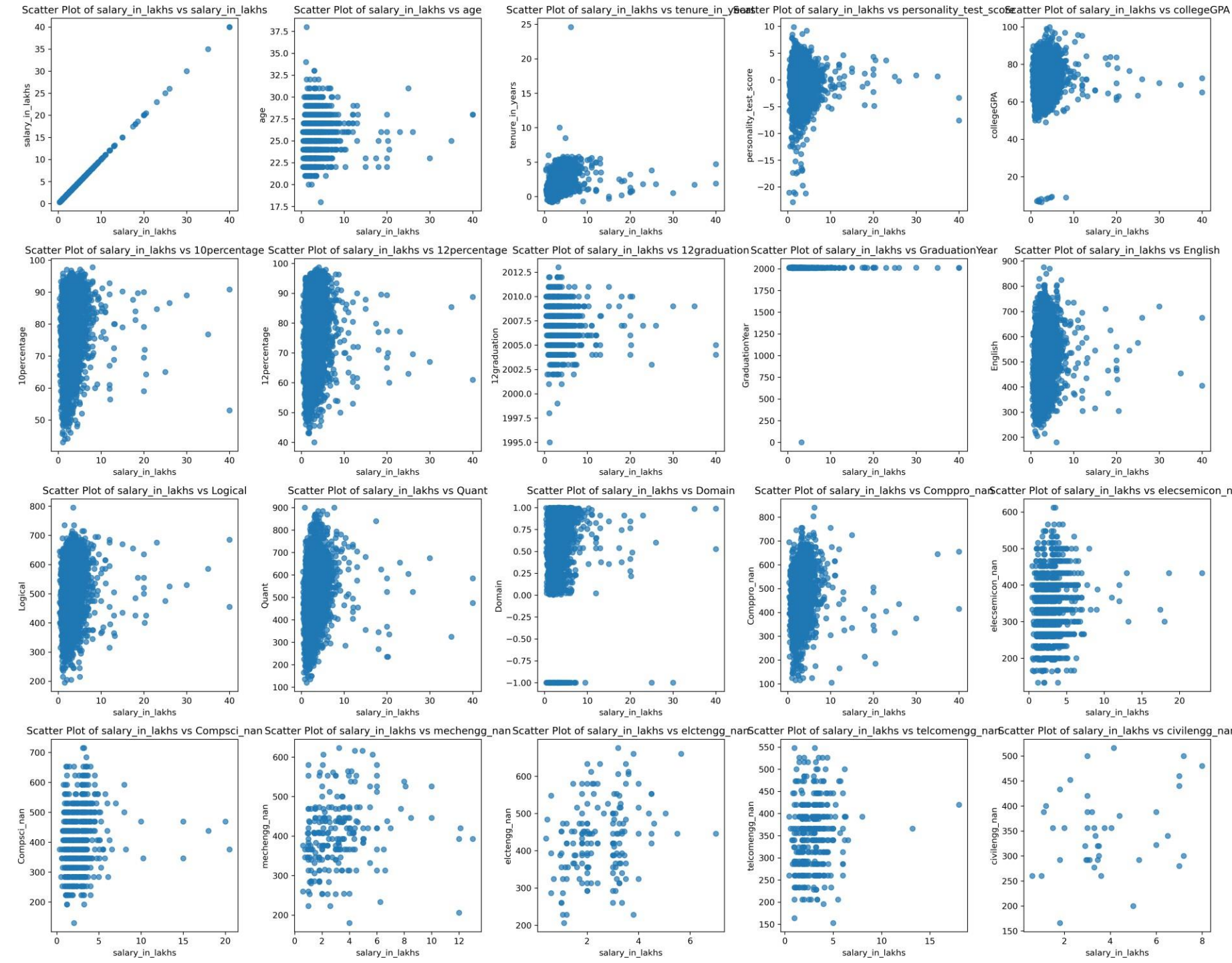
Univariate Numerical Variables Distributions

KDE of various numerical columns present in the dataset



Univariate Numerical Variables Distributions

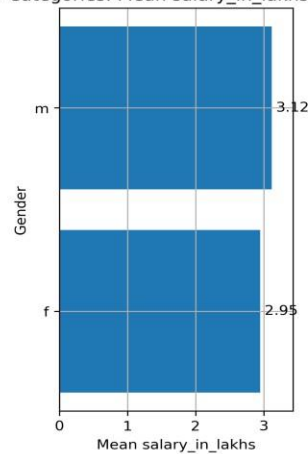
Box plots of various numerical columns present in the dataset



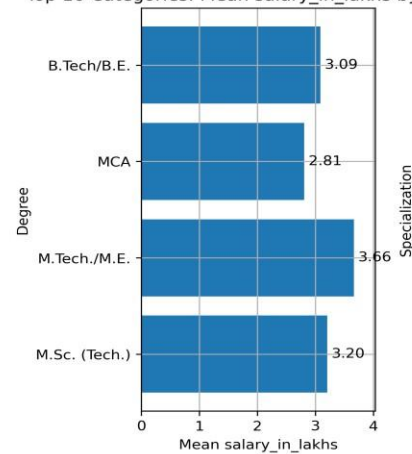
Various Features & Their Relationships With The Target Variable, Salary

Bivariate Analysis Of Salary With Other Numerical Variables

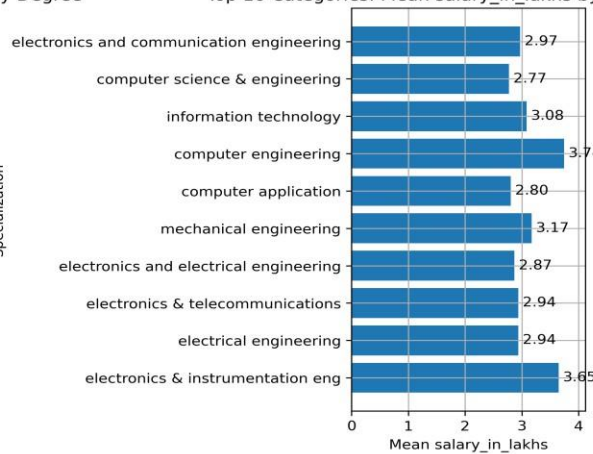
Top 10 Categories: Mean salary_in_lakhs by Gender



Top 10 Categories: Mean salary_in_lakhs by Degree



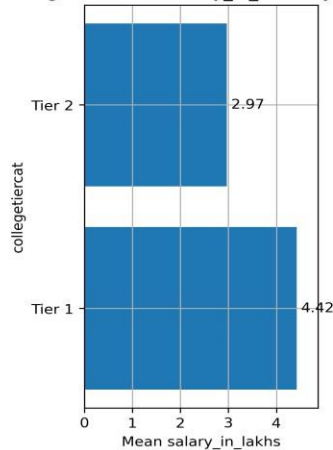
Top 10 Categories: Mean salary_in_lakhs by Specialization



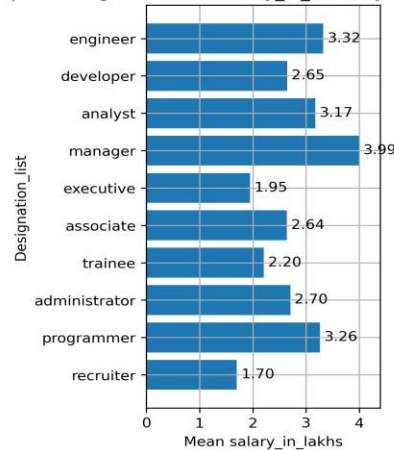
Various Features & Their Relationships With The Target Variable, Salary

Bivariate Analysis Of Salary With Other Categorical Variables

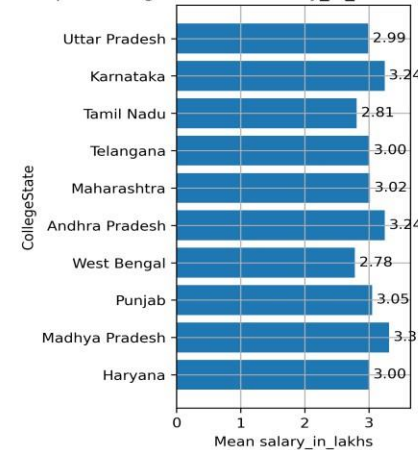
Top 10 Categories: Mean salary_in_lakhs by collegetiercat



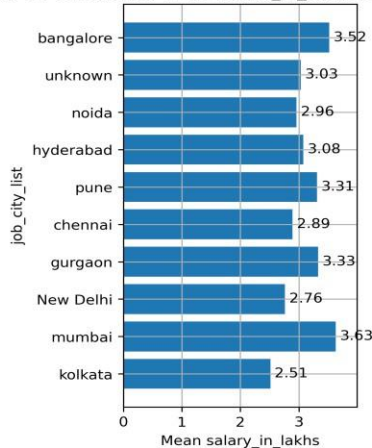
Top 10 Categories: Mean salary_in_lakhs by Designation_list



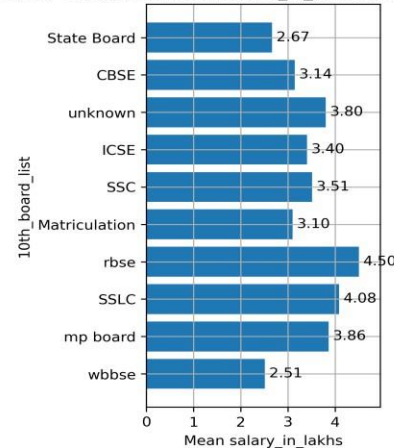
Top 10 Categories: Mean salary_in_lakhs by CollegeState



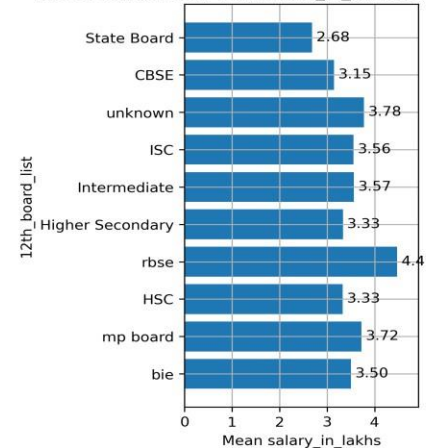
Top 10 Categories: Mean salary_in_lakhs by job_city_list

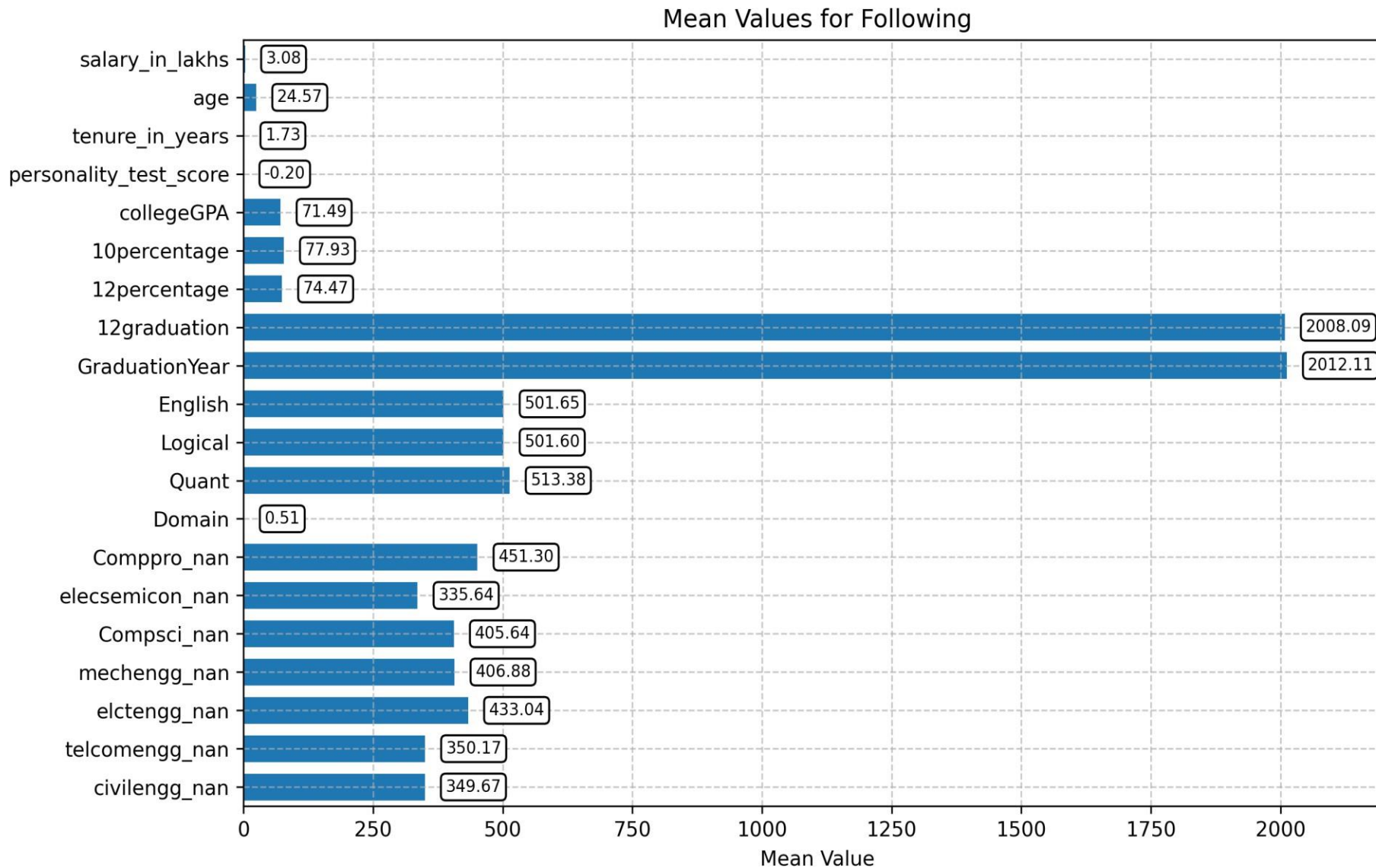


Top 10 Categories: Mean salary_in_lakhs by 10th_board_list



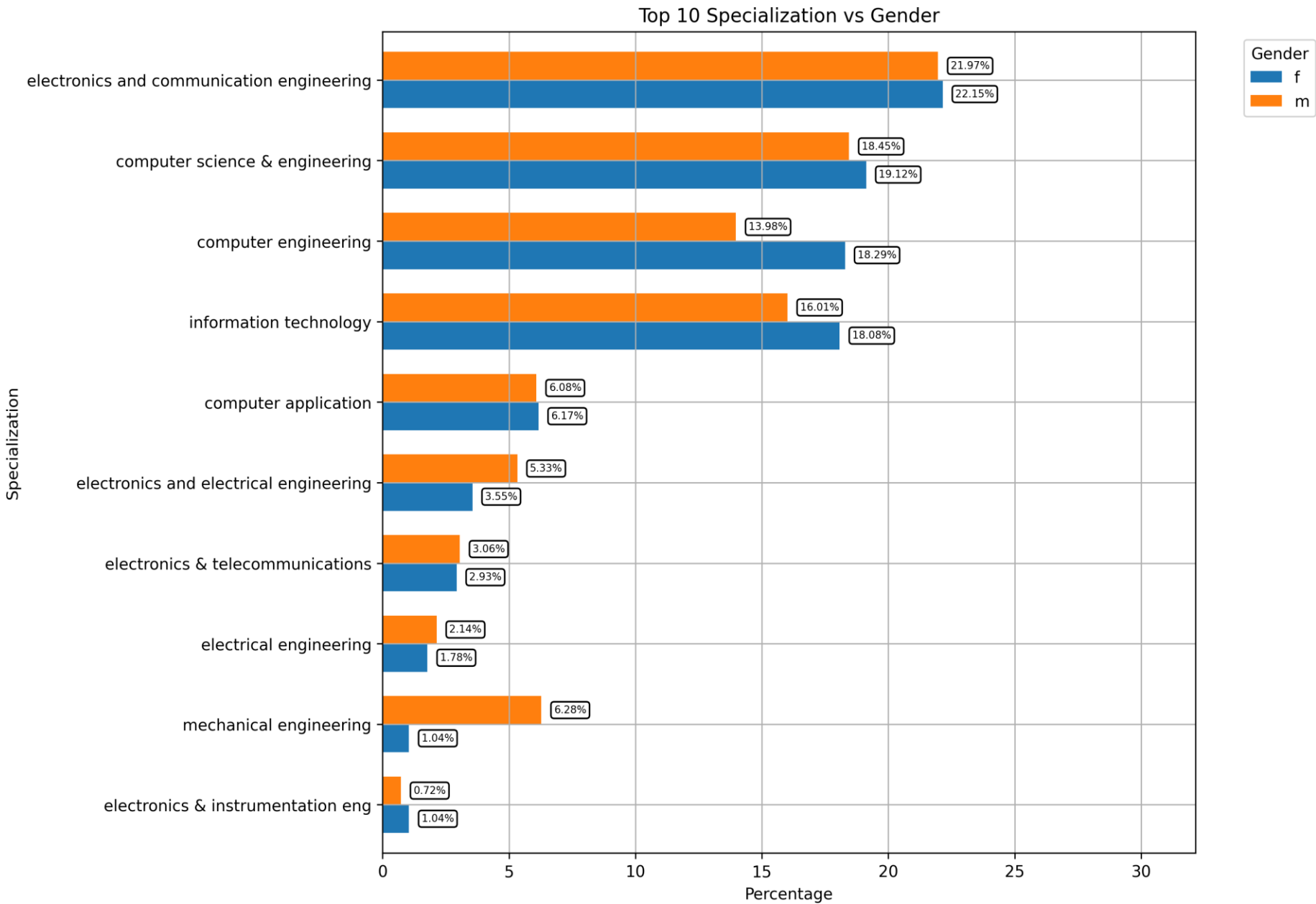
Top 10 Categories: Mean salary_in_lakhs by 12th_board_list





Various Numerical Features & Their Mean Values

Bivariate Analysis Of Salary
With Other Numerical Variables

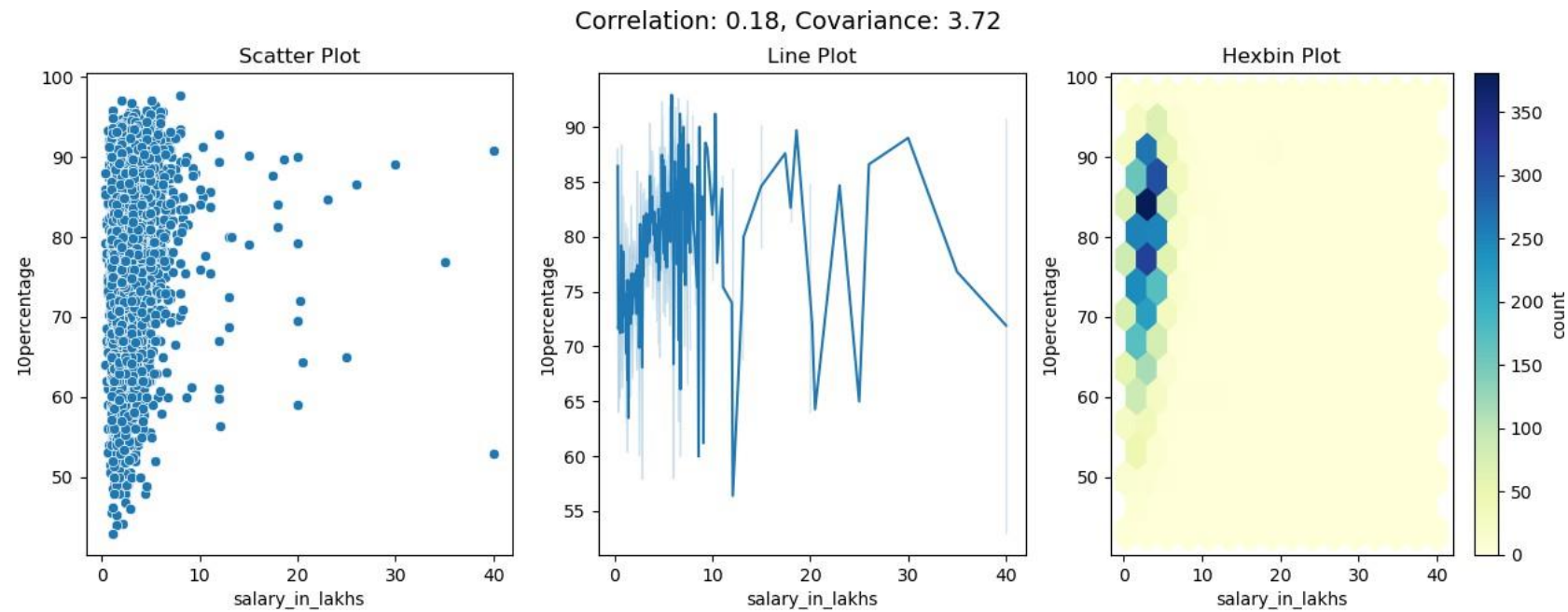


Question

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

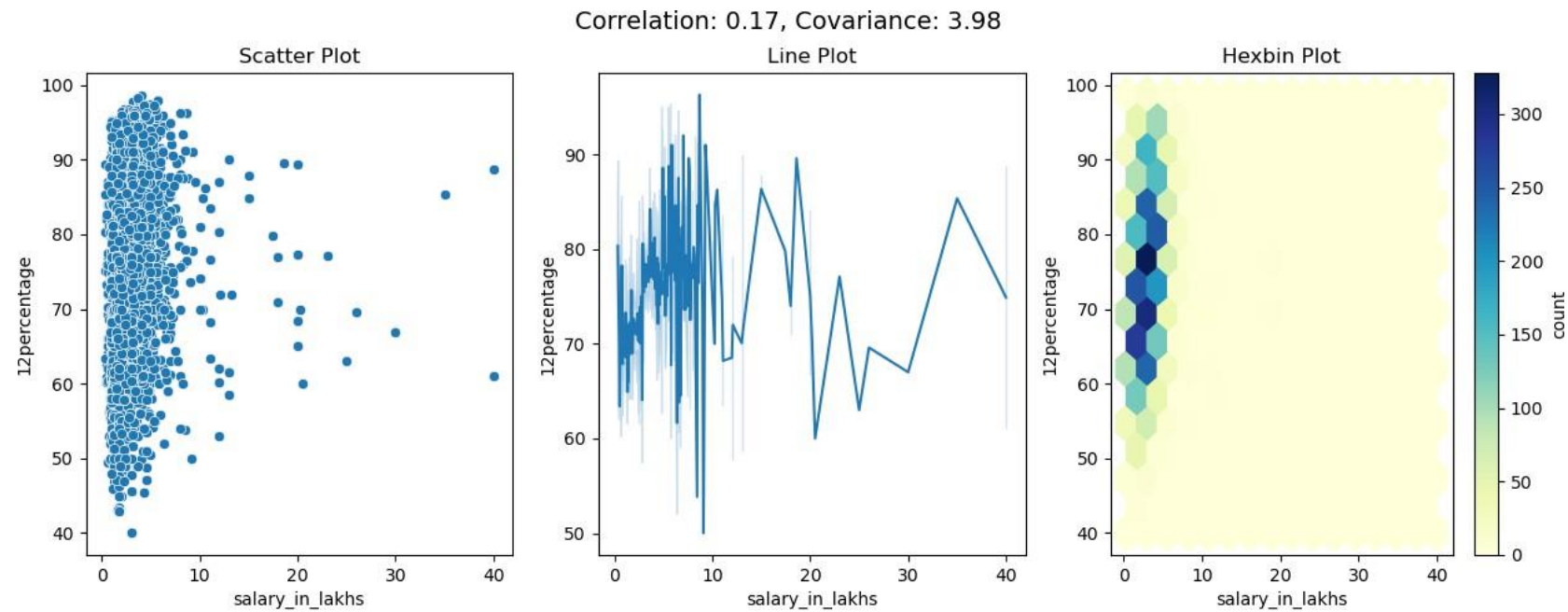
Questions

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



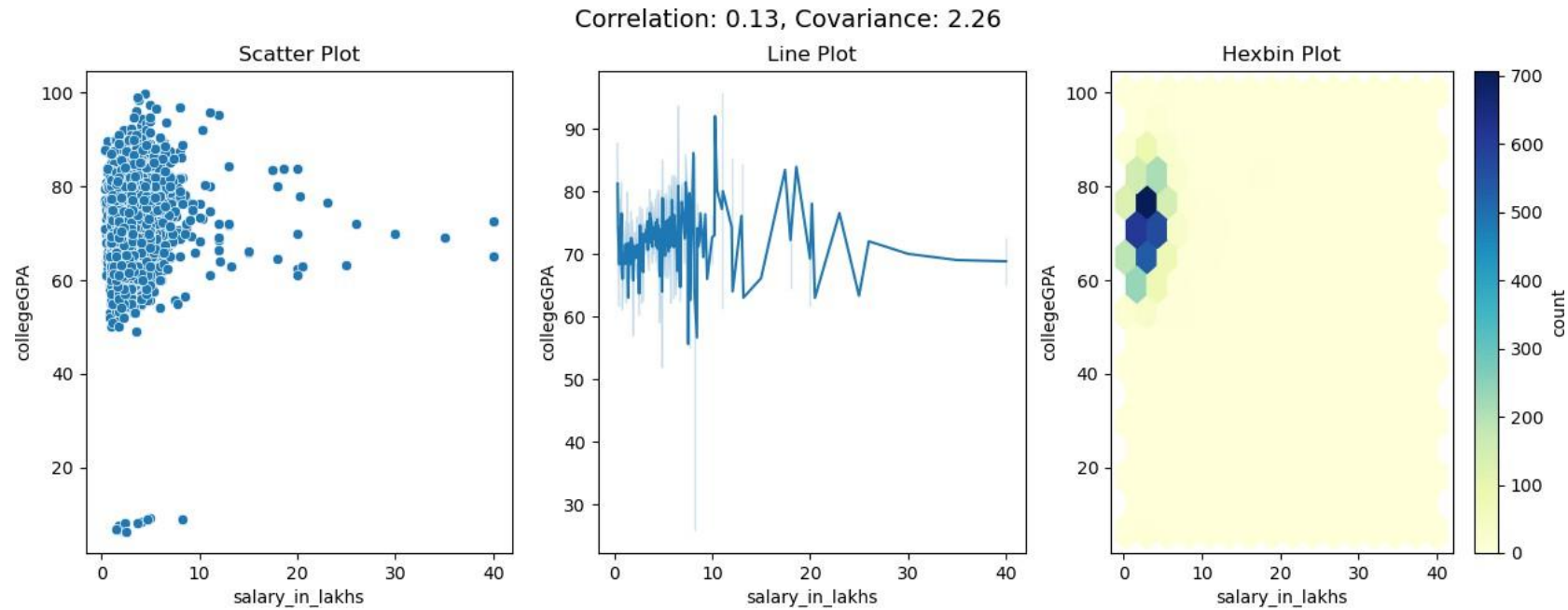
Questions

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

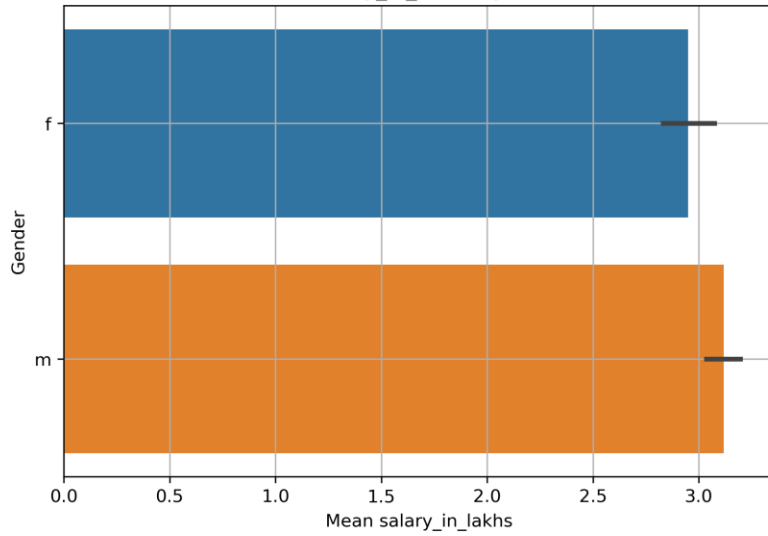


Questions

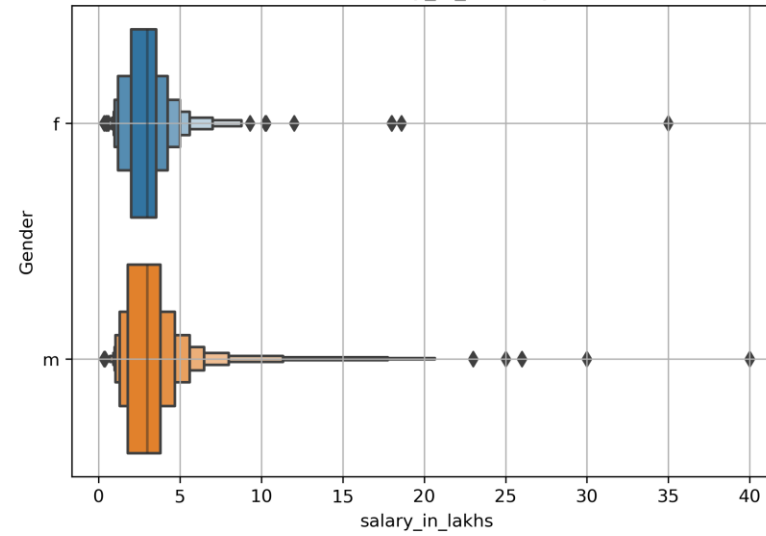
2. Relationship between education and earning potential (Salary Vs CollegeGPA)



Mean salary_in_lakhs by Gender



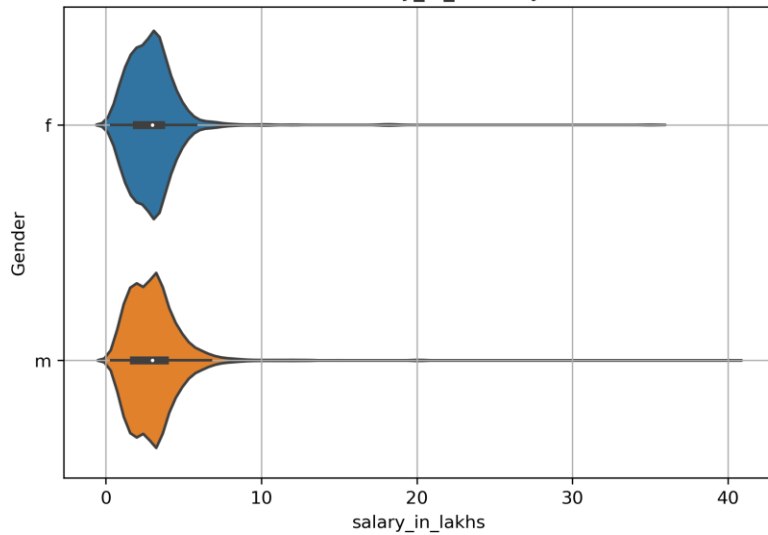
Boxen Plot of salary_in_lakhs by Gender



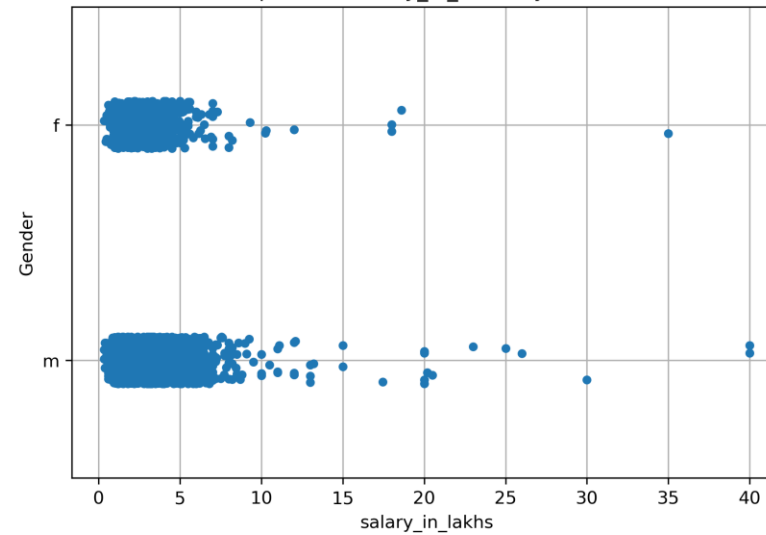
Question

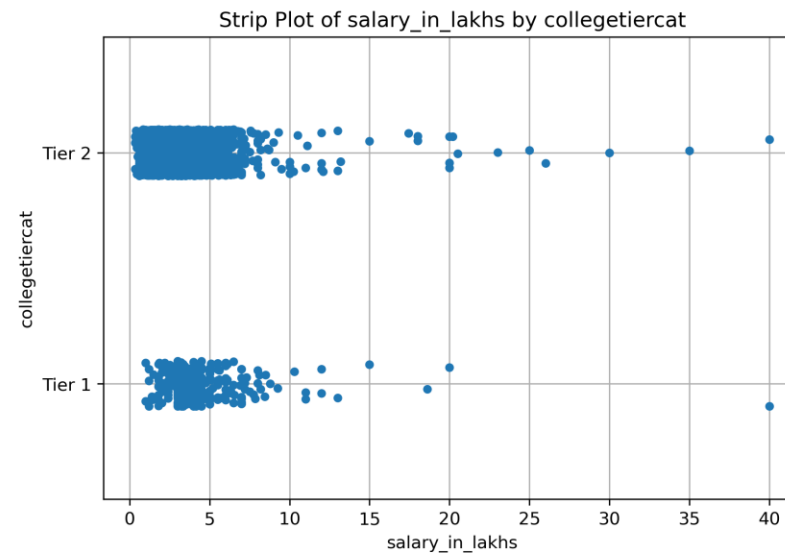
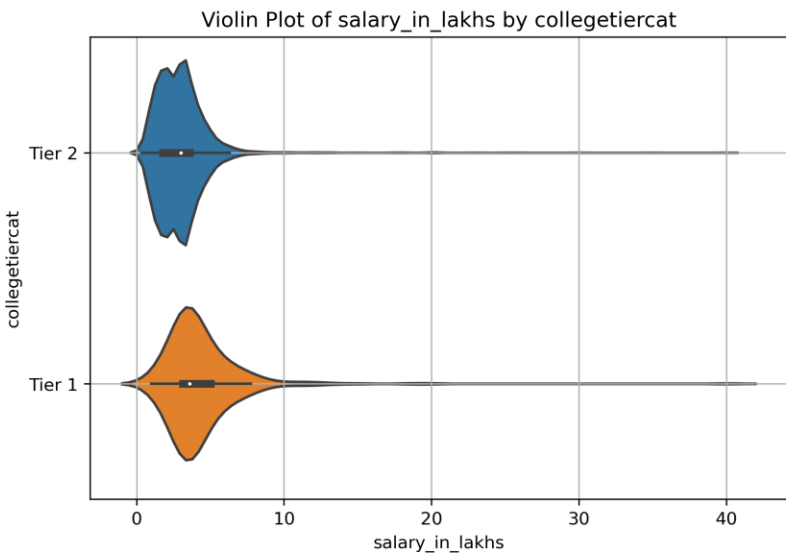
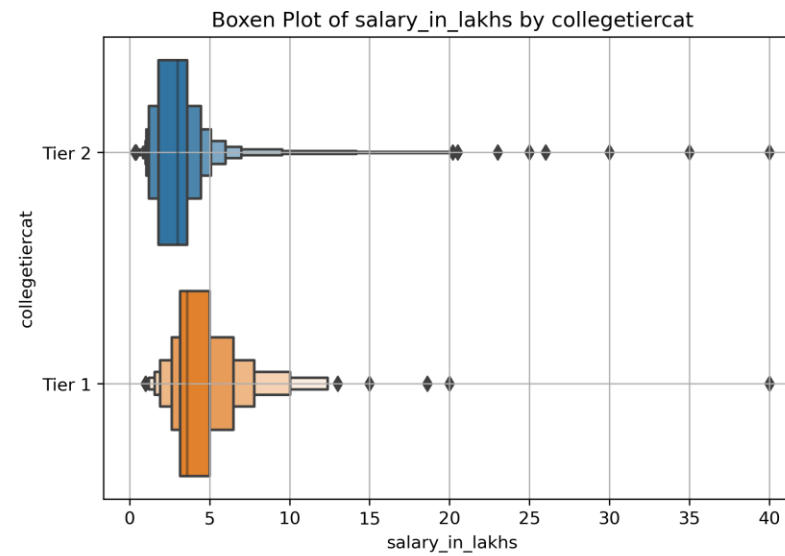
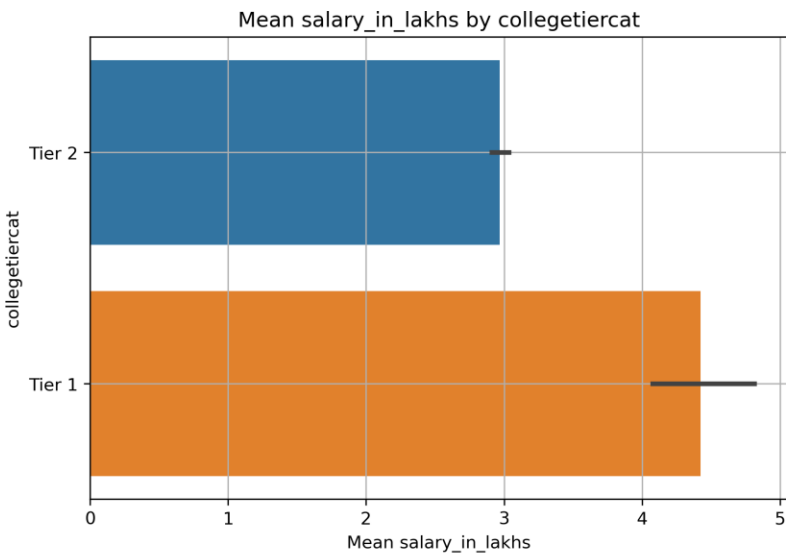
2. Explored the relationship between gender and salary

Violin Plot of salary_in_lakhs by Gender



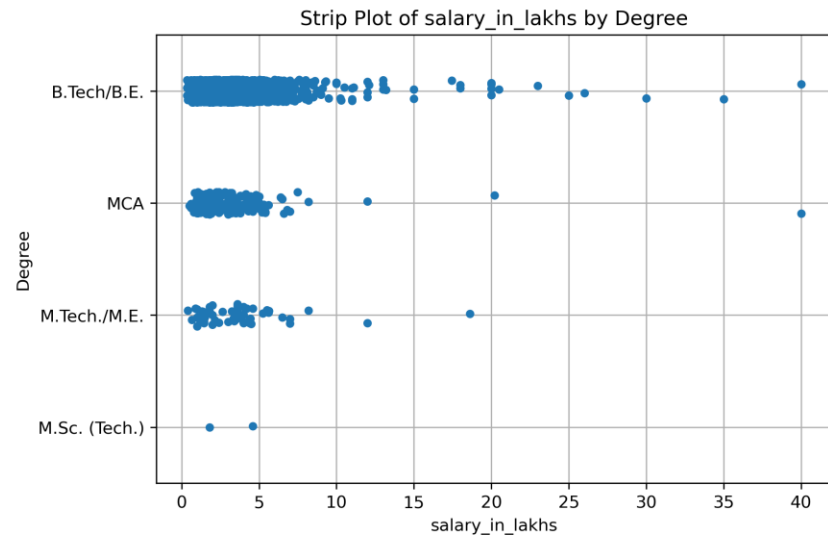
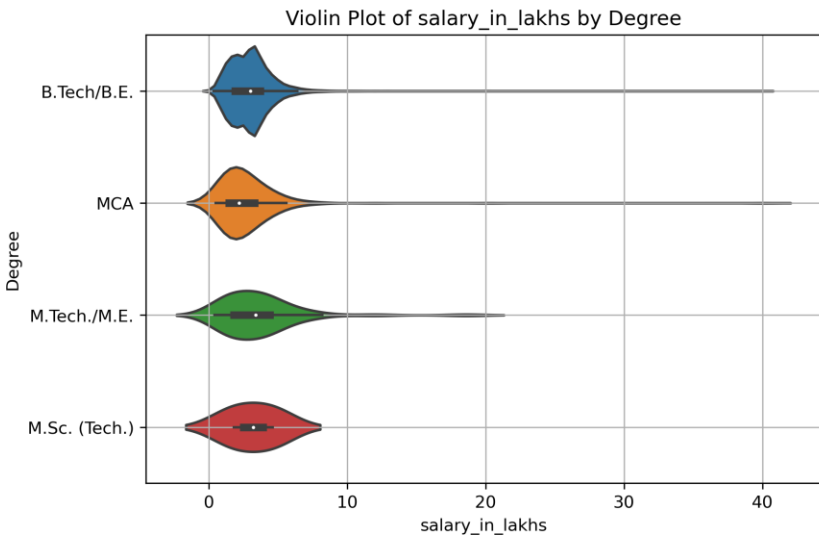
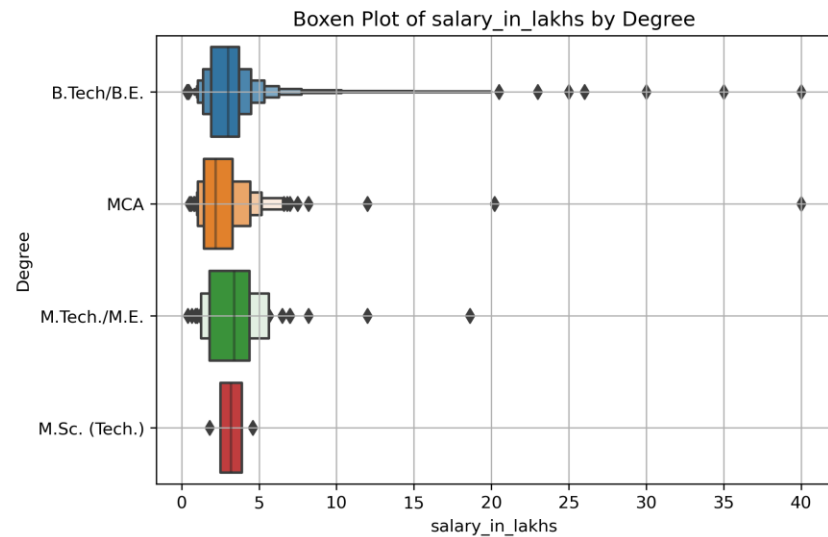
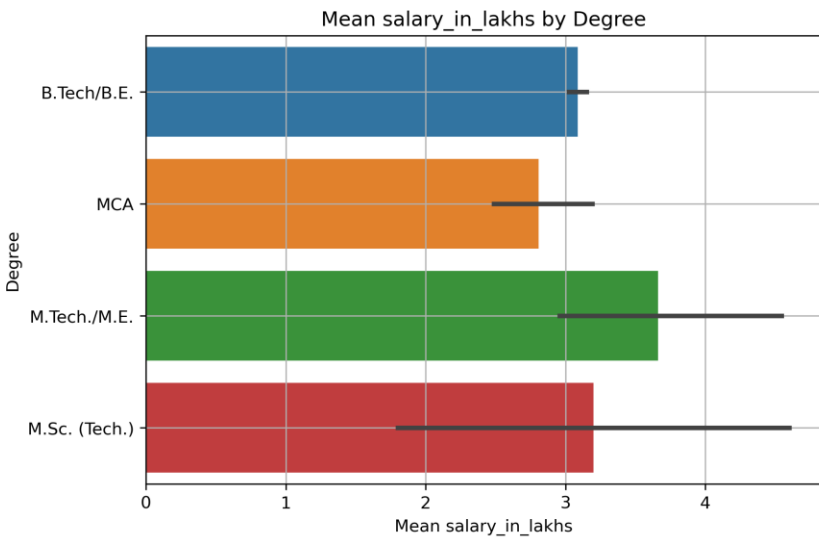
Strip Plot of salary_in_lakhs by Gender





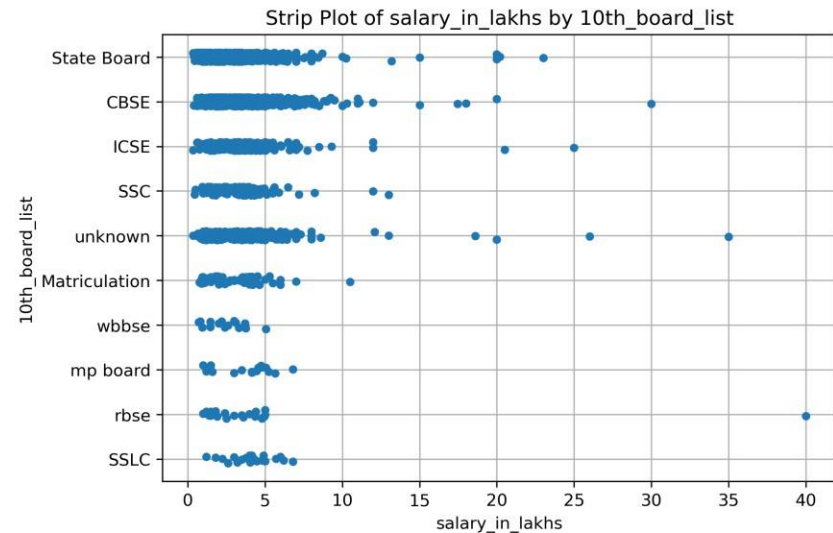
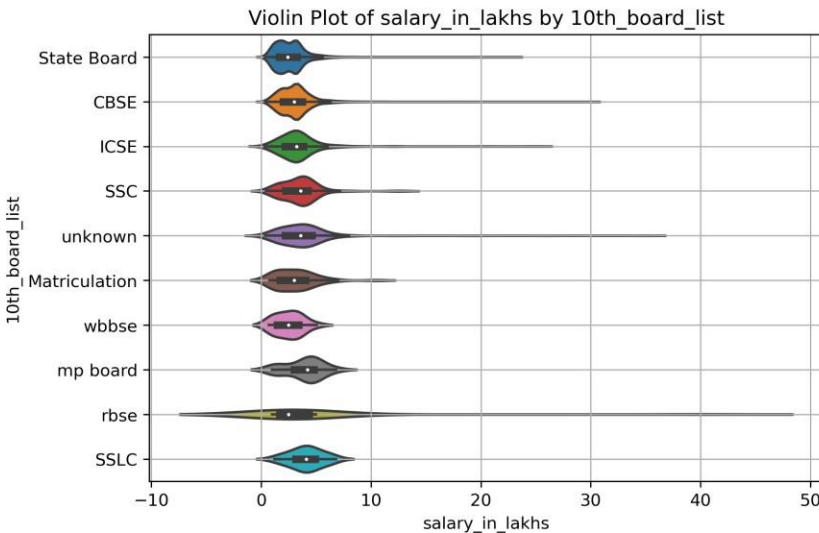
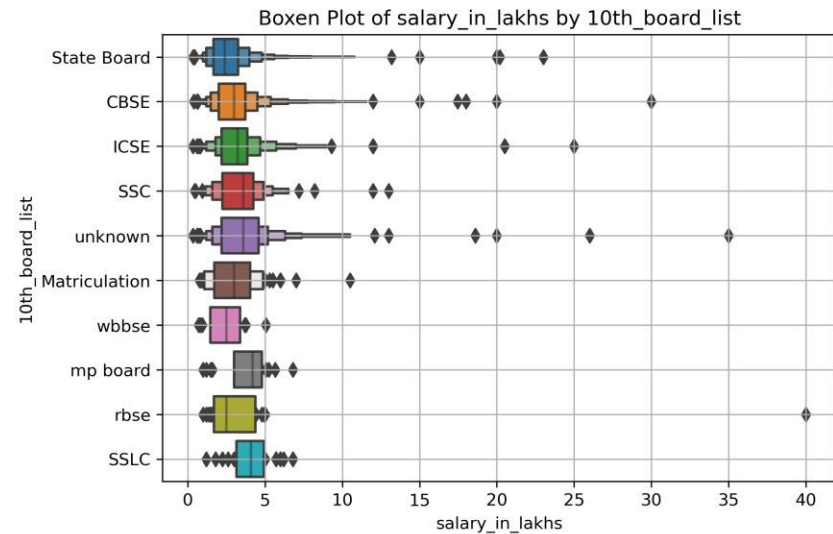
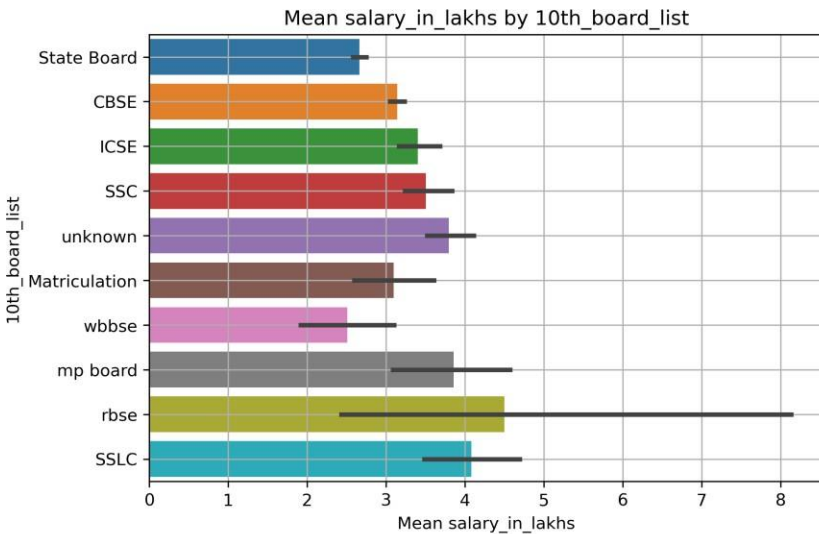
Question

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



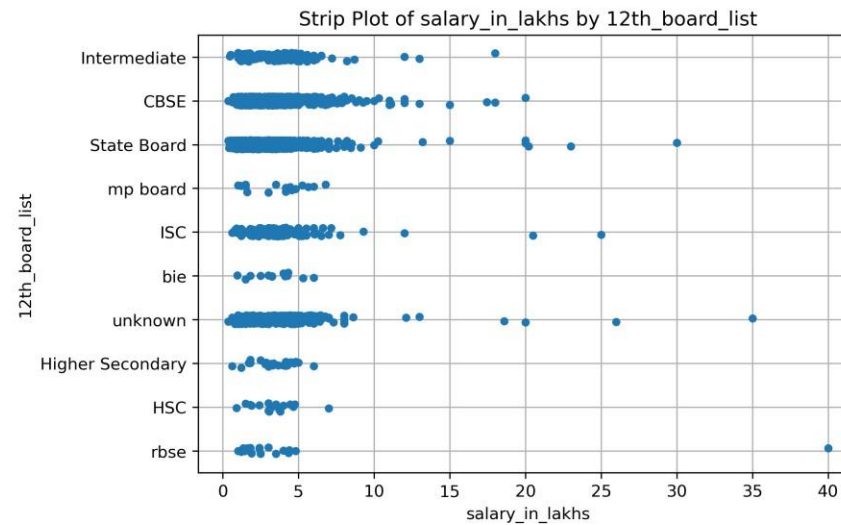
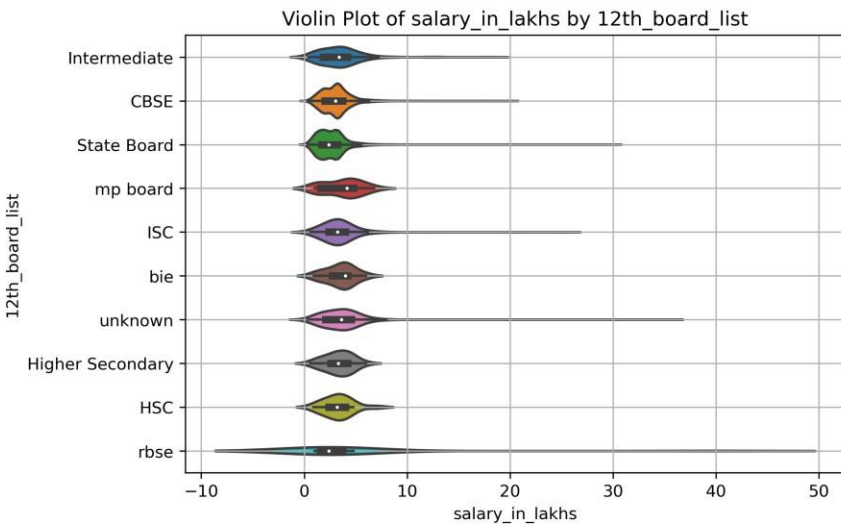
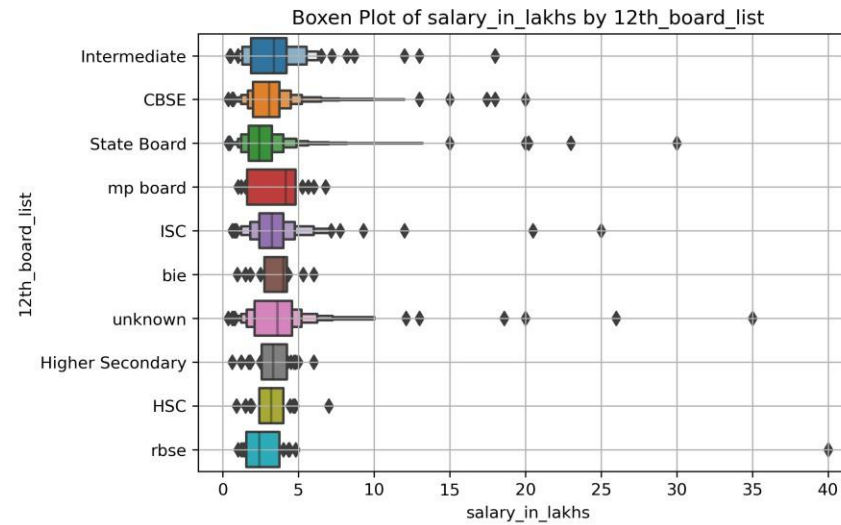
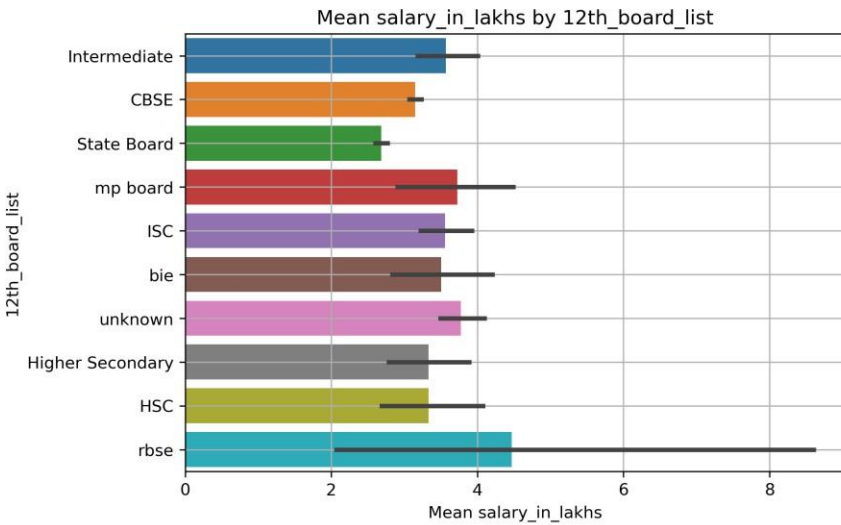
Question

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



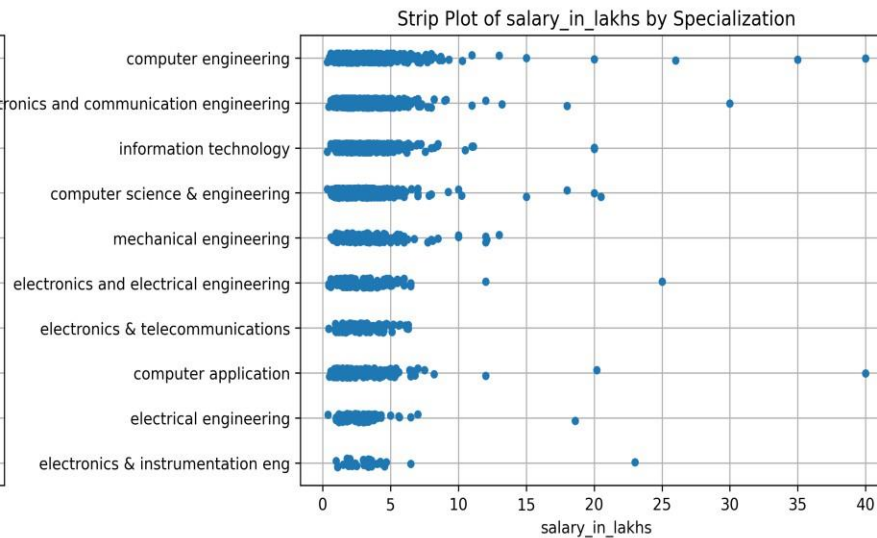
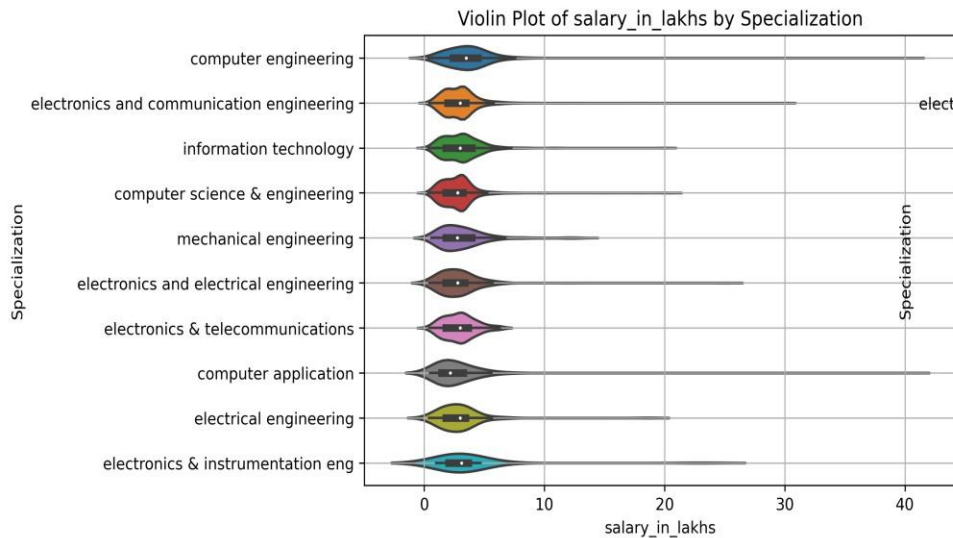
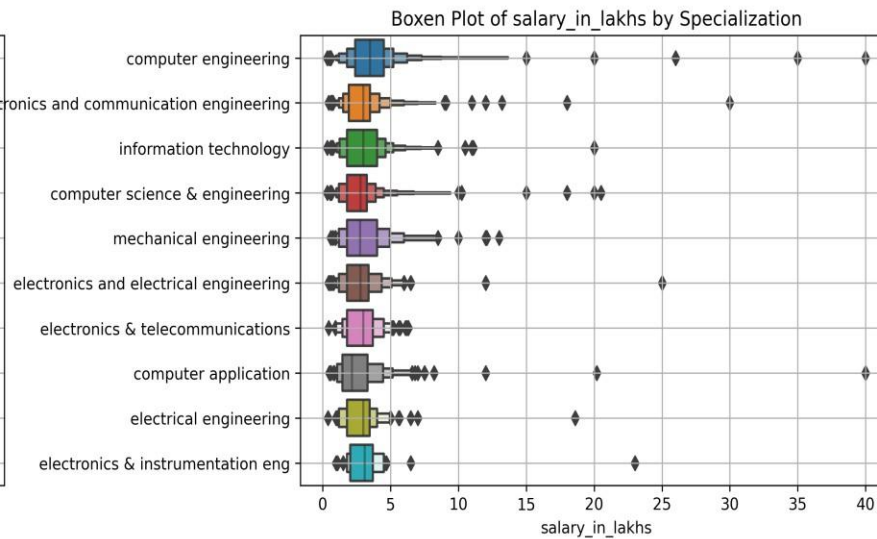
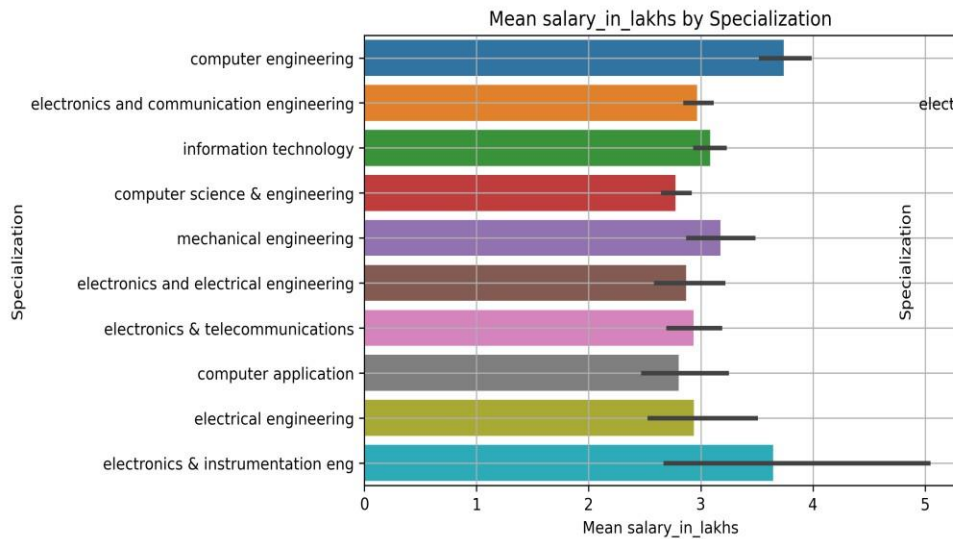
Question

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



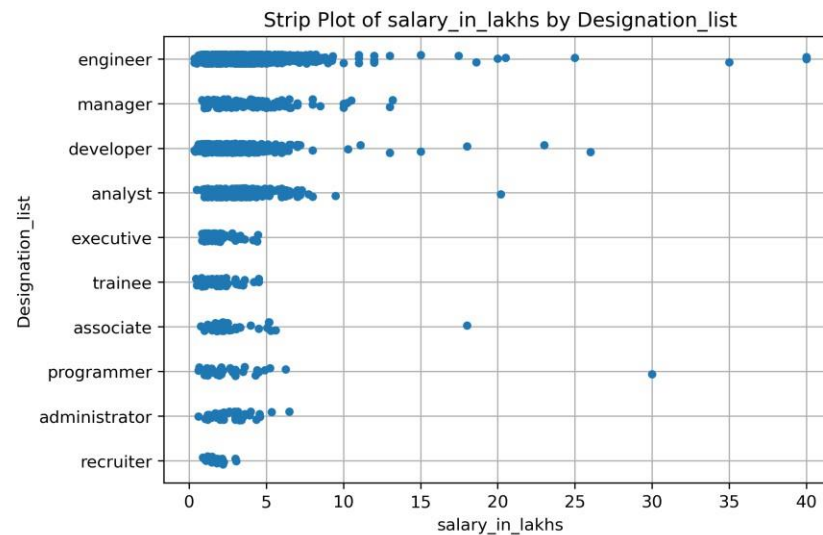
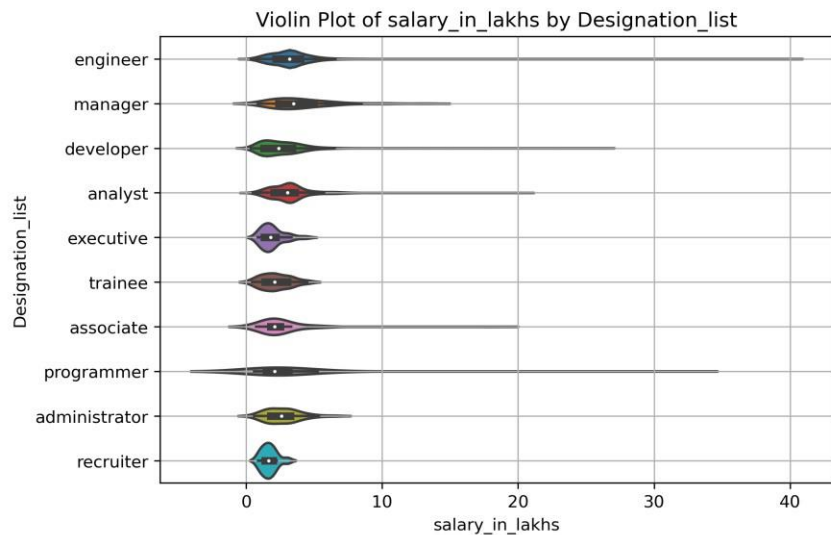
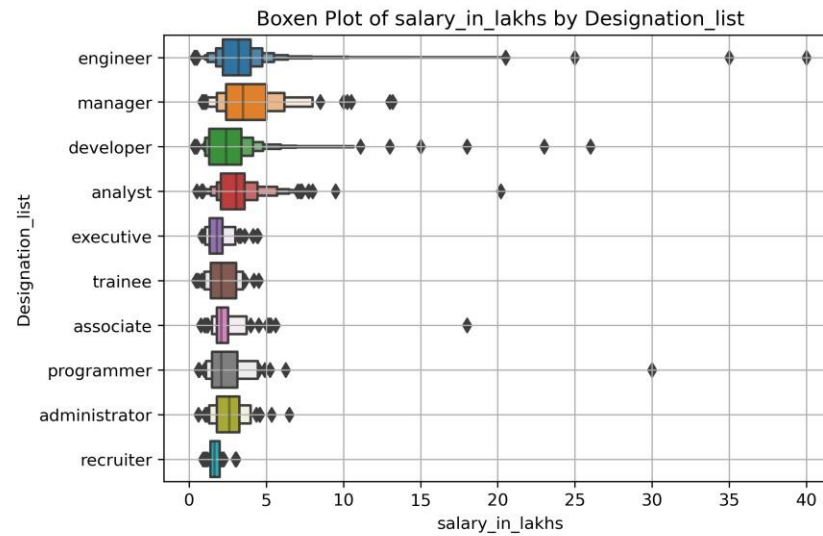
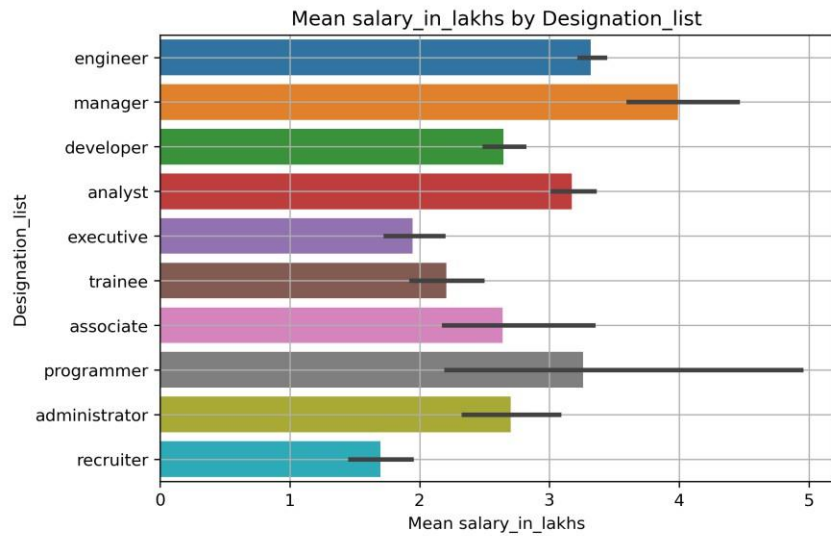
Question

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



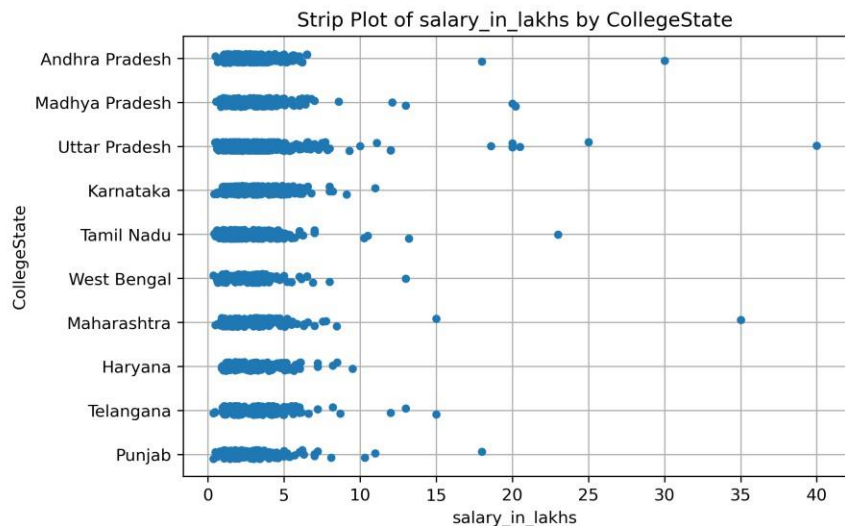
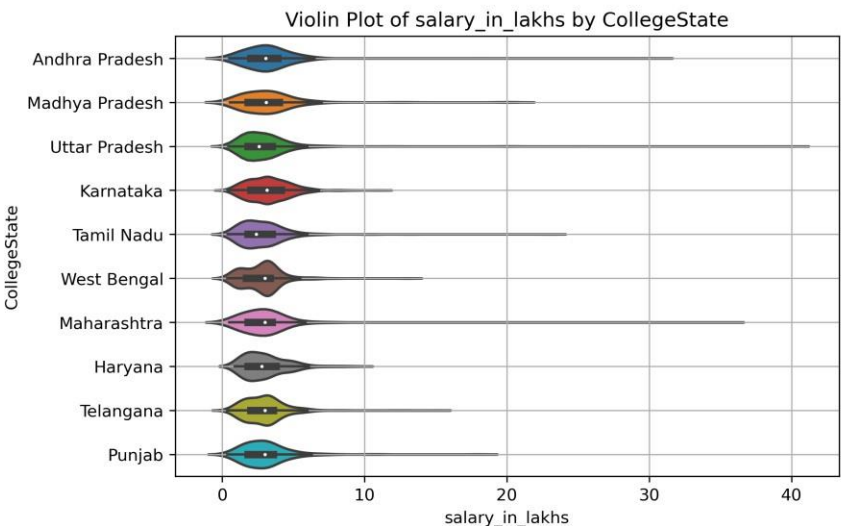
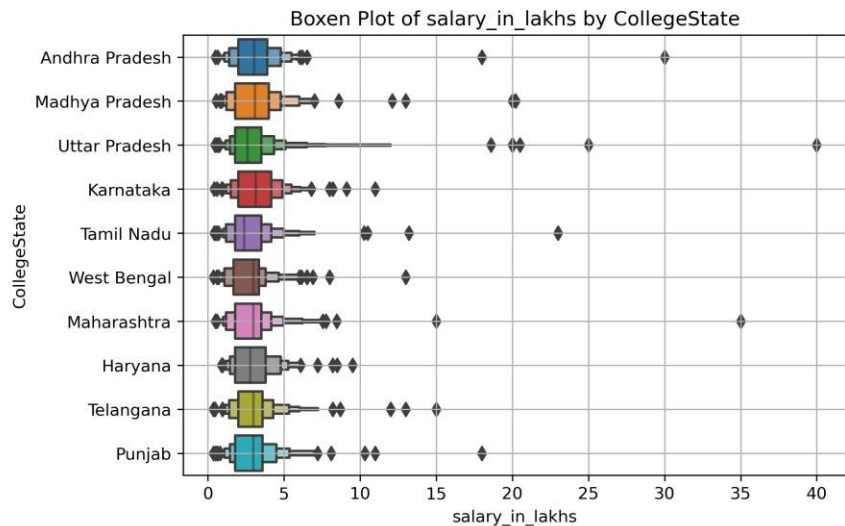
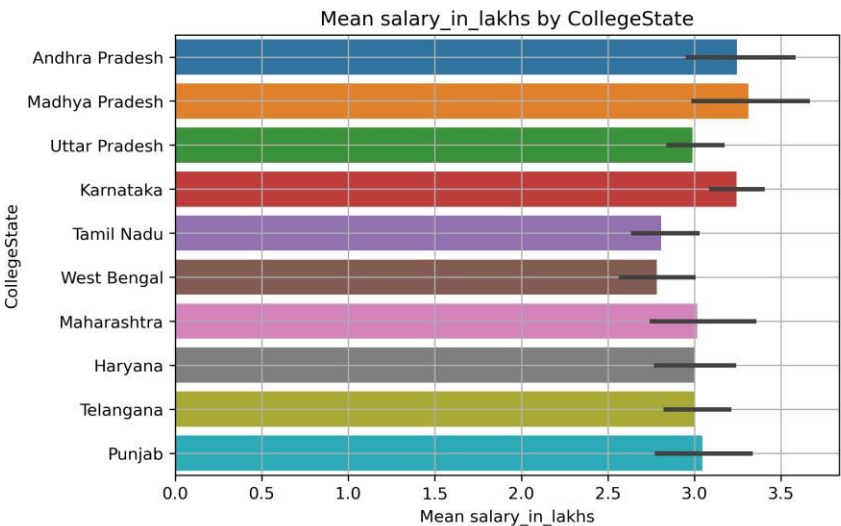
Salary Vs Specializations

Top 10



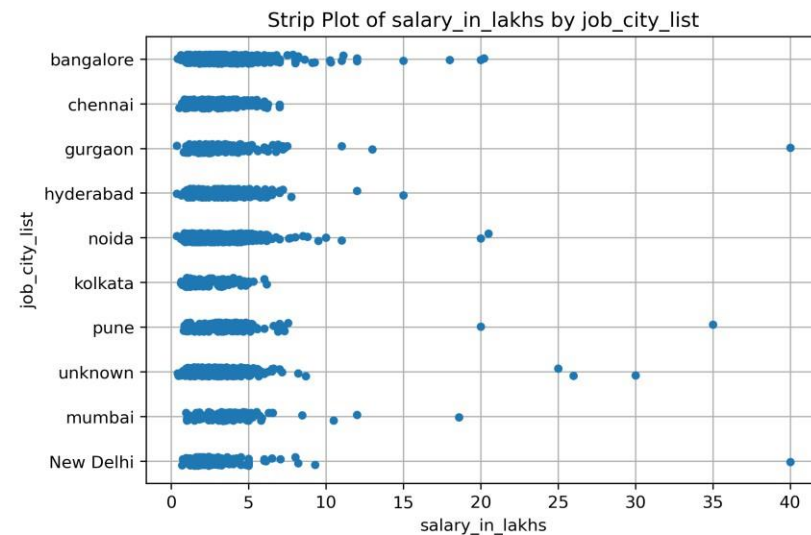
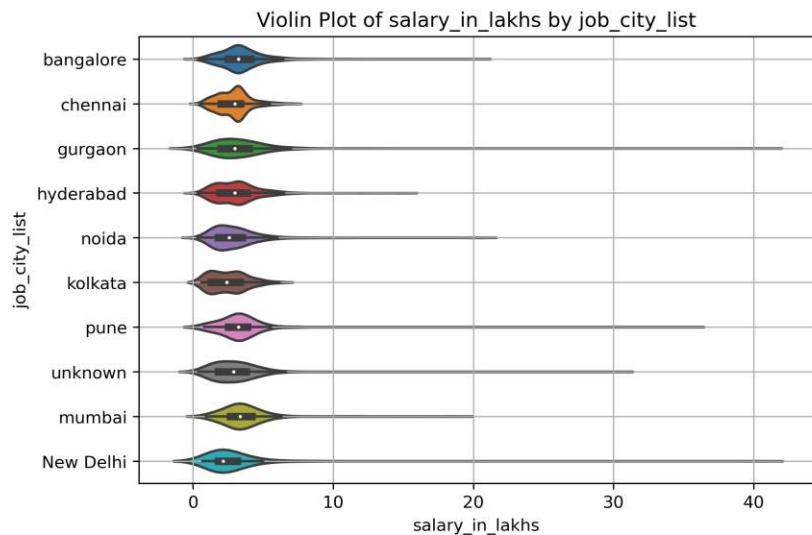
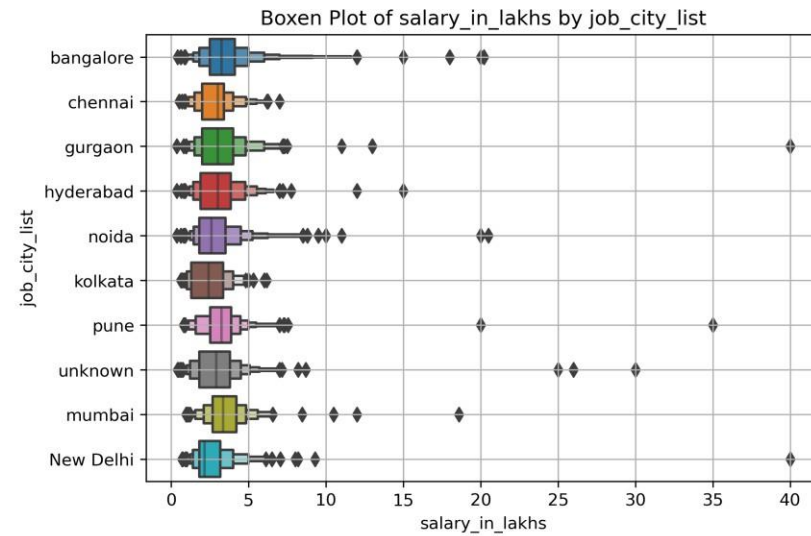
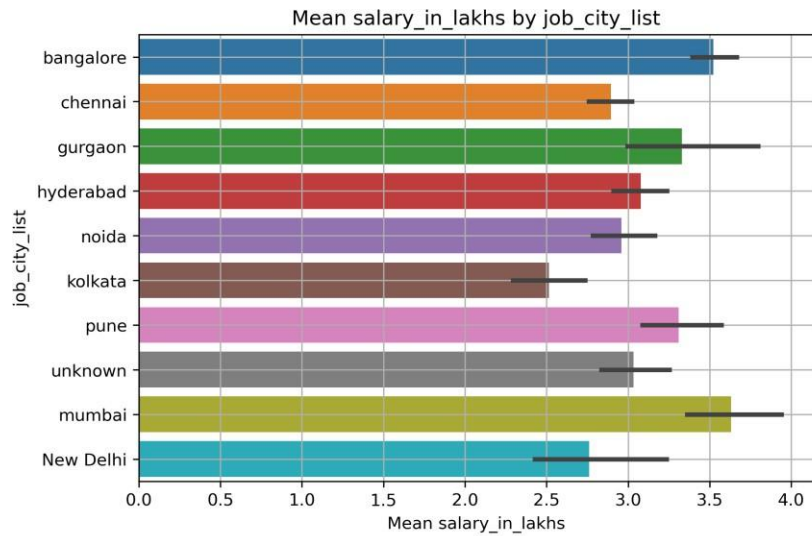
Salary Vs Designations

Top 10



Salary Vs CollegeStates

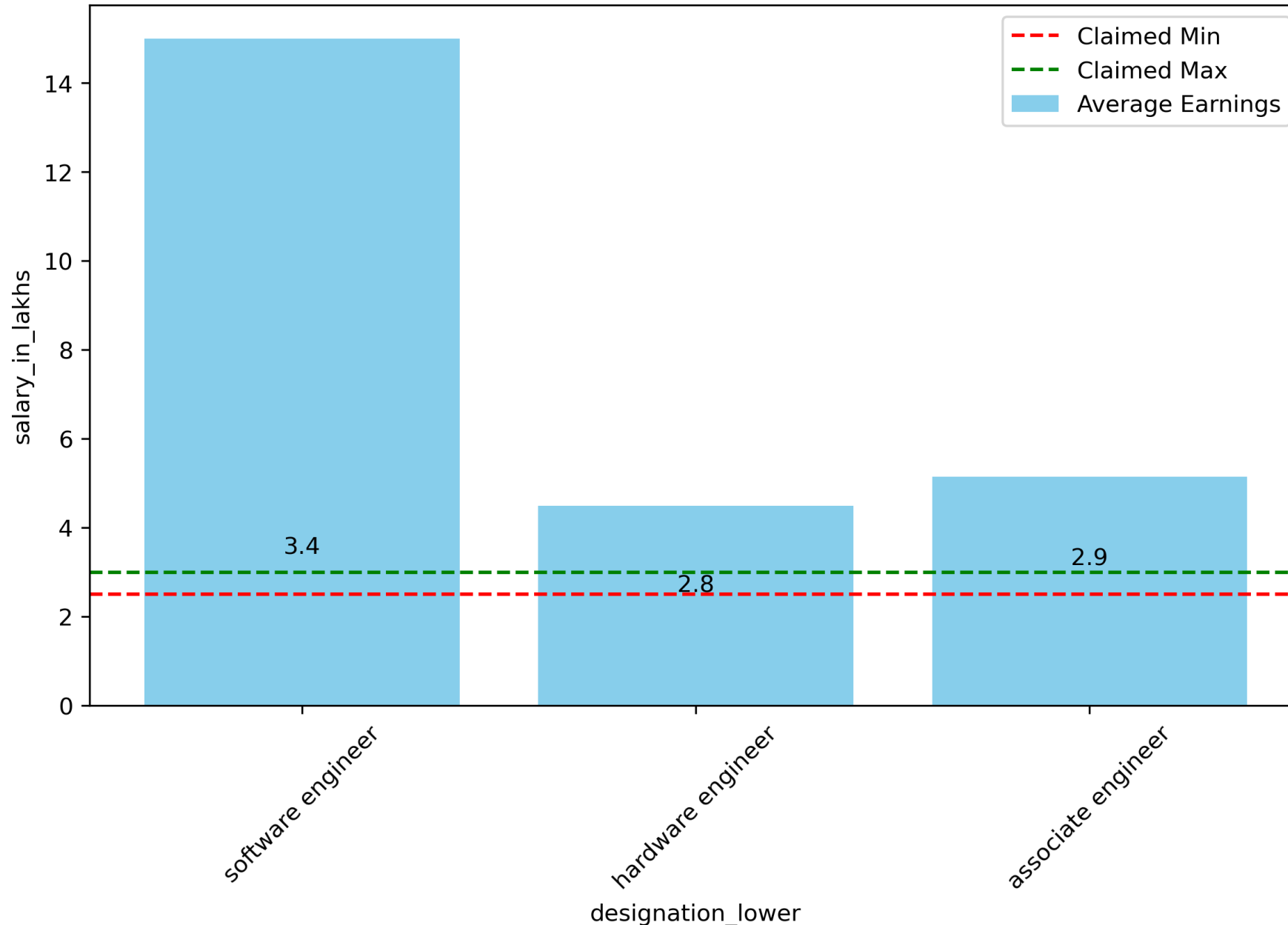
Top 10



Salary Vs JobCity

Top 10

Average Salary_in_lakhs for Designation_lower vs. Claimed Range



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

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.

THANK YOU