INNOMATICS®
RESEARCH LABS

INNOVATION. AUTOMATION. ANALYTICS

PROJECT ON

EDA - AMCAT Data Analysis Project.

- Presented by: Shivam Dubey

About me

- **Background ? (B-tech)**

- **I am a recent graduate with a B.Tech in Computer Science and Engineering. My passion for technology and data has driven me to specialize in Artificial Intelligence and Machine Learning. I possess strong proficiency in programming languages and frameworks such as Python, TensorFlow, and scikit-learn, which I have applied to develop and deploy AI-driven applications.**

- **Why you want to learn Data Science**

I am passionate about extracting insights from data and making data-driven decisions. I believe that data science combines my love for mathematics and programming, allowing me to contribute to solving real-world problems and enhancing business operations.

- **Share your linkedin profile urls**

About me

- Previous work experience

- **AI/ML Intern**

- Antihak.AI (Feb 2024 – May 2024)

- Developed "SafeTransact," a secure transaction management system to combat financial fraud using Python and machine learning.

- **Data Science Intern**
- Suvidha Foundation NGO (Dec 2023 – Feb 2024)

- Contributed to the Digital Summarizer Project, focusing on text summarization and Seq2seq model training in NLP and Deep Learning.

- Share github profile urls

INNOMATICS RESEARCH LABS

Agenda (This should be the PPT flow)

- **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***

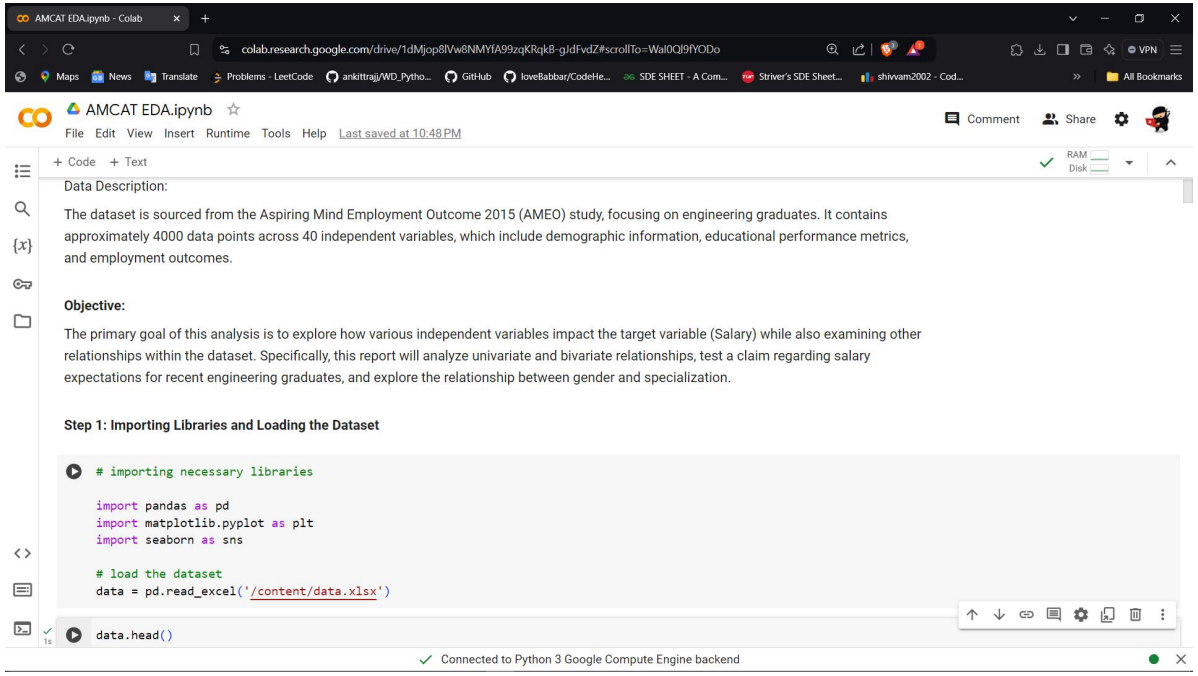
- **Key Business Question**
- **Conclusion (Key finding overall)**
- **Q&A Slide**
- **Your Experience/Challenges working on Web Scraping – Data Analysis Project.**

INNOMATICS RESEARCH LABS

Business Problem and Use Case Understanding:

The project aims to perform Exploratory Data Analysis (EDA) on the AMCAT dataset to uncover insights regarding employment outcomes for engineering graduates. This will help in understanding factors affecting salaries and job placements.

Use Case: Analyzing AMCAT data to provide insights for students and educational institutions.



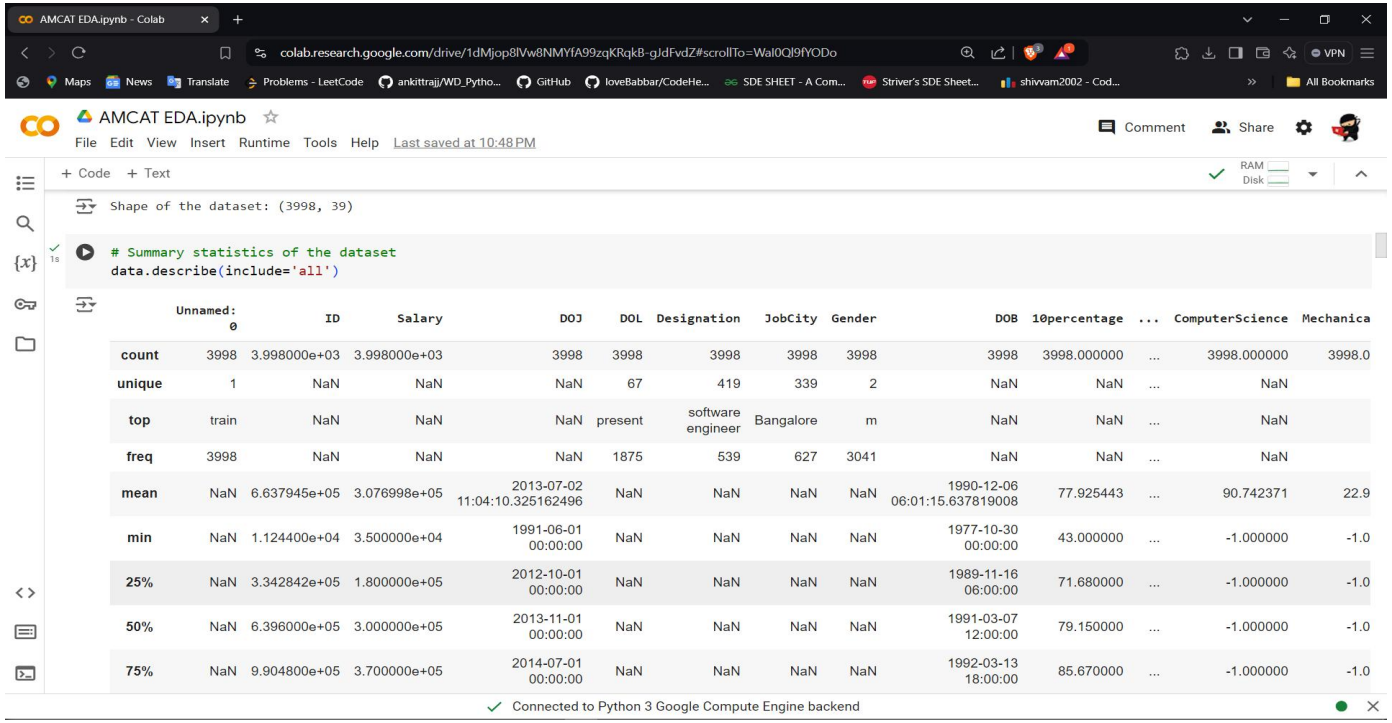
INNOMATICS RESEARCH LABS

Objective of the Project

- To perform Exploratory Data Analysis (EDA) on the AMCAT dataset, focusing on the target variable 'Salary' and identifying key trends and insights.

Summary of the Data

- The AMCAT dataset contains approximately 4,000 candidates with around 40 variables, including demographic details, academic performance, AMCAT test scores, and job outcomes.



	Unnamed: 0	ID	Salary	DOJ	DOL	Designation	JobCity	Gender	DOB	10percentage	...	ComputerScience	Mechanica
count	3998	3.998000e+03	3.998000e+03	3998	3998	3998	3998	3998	3998	3998.000000	...	3998.000000	3998.0
unique	1	NaN	NaN	NaN	67	419	339	2	NaN	NaN	...	NaN	
top	train	NaN	NaN	NaN	present	software engineer	Bangalore	m	NaN	NaN	...	NaN	
freq	3998	NaN	NaN	NaN	1875	539	627	3041	NaN	NaN	...	NaN	
mean	NaN	6.637945e+05	3.076998e+05	2013-07-02 11:04:10.325162496	NaN	NaN	NaN	NaN	1990-12-06 06:01:15.637819008	77.925443	...	90.742371	22.9
min	NaN	1.124400e+04	3.500000e+04	1991-06-01 00:00:00	NaN	NaN	NaN	NaN	1977-10-30 00:00:00	43.000000	...	-1.000000	-1.0
25%	NaN	3.342842e+05	1.800000e+05	2012-10-01 00:00:00	NaN	NaN	NaN	NaN	1989-11-16 06:00:00	71.680000	...	-1.000000	-1.0
50%	NaN	6.396000e+05	3.000000e+05	2013-11-01 00:00:00	NaN	NaN	NaN	NaN	1991-03-07 12:00:00	79.150000	...	-1.000000	-1.0
75%	NaN	9.904800e+05	3.700000e+05	2014-07-01 00:00:00	NaN	NaN	NaN	NaN	1992-03-13 18:00:00	85.670000	...	-1.000000	-1.0

INNOMATICS RESEARCH LABS

Exploratory Data Analysis

- Data Cleaning Steps: [Briefly outline your data cleaning steps]
- Data Manipulation Steps: [Outline how you manipulated the data for analysis]

- **Exploratory Data Analysis:**

a. Data Cleaning Steps

AMCAT EDA.ipynb - Colab

colab.research.google.com/drive/1dMjop8lVw8NMYfA99zqKRqkB-gJdFvdZ#scrollTo=VX2JCzfZZhXU

AMCAT EDA.ipynb

File Edit View Insert Runtime Tools Help Last saved at October 4

```
# Summary statistics of the dataset
data.describe(include='all')
```

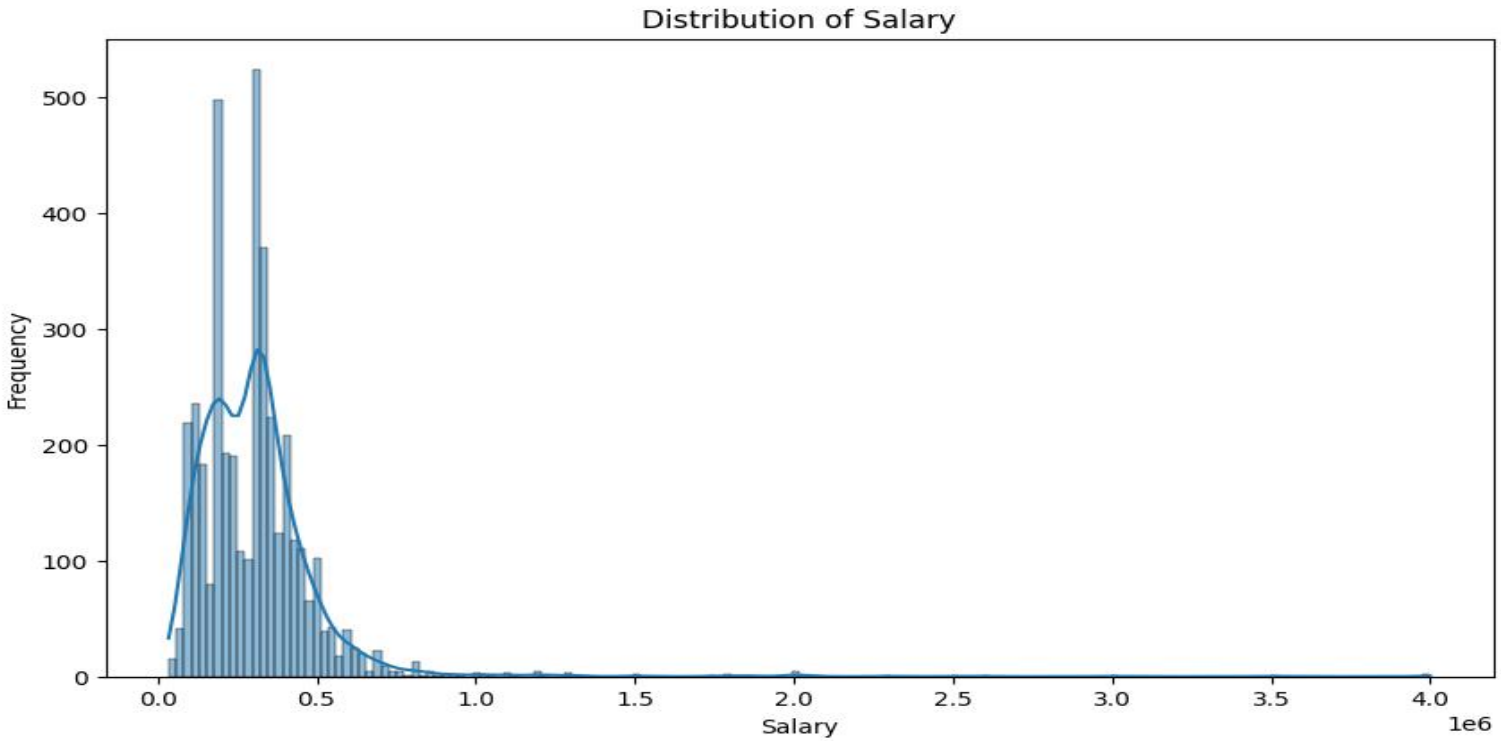
	Unnamed: 0	ID	Salary	DOJ	DOL	Designation	JobCity	Gender	DOB	10percentage	...	ComputerScience	Mechanica
count	3998	3.998000e+03	3.998000e+03	3998	3998	3998	3998	3998	3998	3998.000000	...	3998.000000	3998.0
unique	1	NaN	NaN	NaN	67	419	339	2	NaN	NaN	...	NaN	
top	train	NaN	NaN	NaN	present	software engineer	Bangalore	m	NaN	NaN	...	NaN	
freq	3998	NaN	NaN	NaN	1875	539	627	3041	NaN	NaN	...	NaN	
mean	NaN	6.637945e+05	3.076998e+05	2013-07-02 11:04:10.325162496	NaN	NaN	NaN	NaN	1990-12-06 06:01:15.637819008	77.925443	...	90.742371	22.9
min	NaN	1.124400e+04	3.500000e+04	1991-06-01 00:00:00	NaN	NaN	NaN	NaN	1977-10-30 00:00:00	43.000000	...	-1.000000	-1.0
25%	NaN	3.342842e+05	1.800000e+05	2012-10-01 00:00:00	NaN	NaN	NaN	NaN	1989-11-16 06:00:00	71.680000	...	-1.000000	-1.0
50%	NaN	6.396000e+05	3.000000e+05	2013-11-01 00:00:00	NaN	NaN	NaN	NaN	1991-03-07 12:00:00	79.150000	...	-1.000000	-1.0
75%	NaN	9.904800e+05	3.700000e+05	2014-07-01 00:00:00	NaN	NaN	NaN	NaN	1992-03-13 18:00:00	85.670000	...	-1.000000	-1.0
max	NaN	1.298275e+06	4.000000e+06	2015-12-01 00:00:00	NaN	NaN	NaN	NaN	1997-05-27 00:00:00	97.760000	...	715.000000	623.0

Connected to Python 3 Google Compute Engine backend

INNOMATICS RESEARCH LABS

- **Exploratory Data Analysis:**

B) Data Manipulation Steps

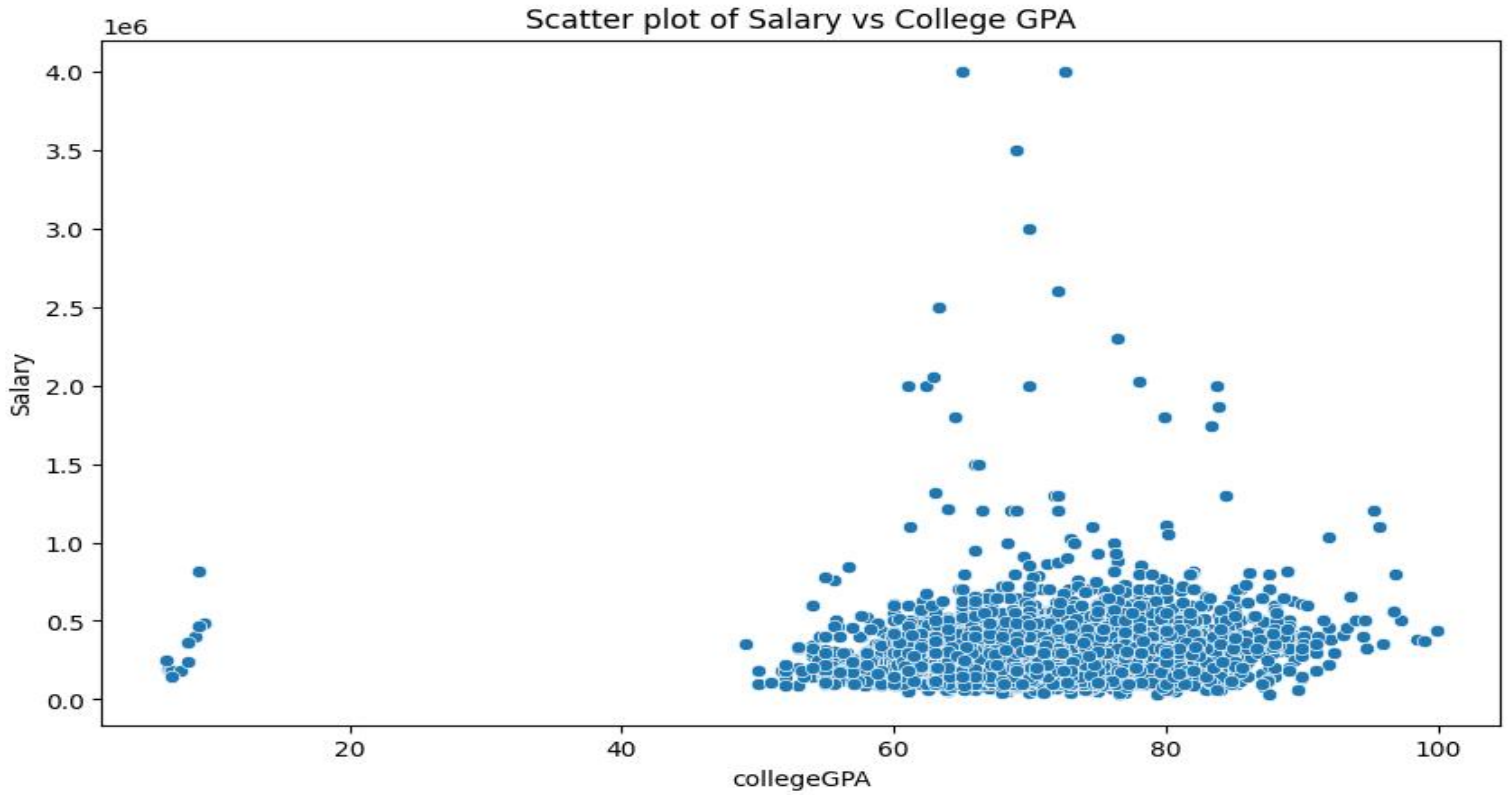


INNOMATICS RESEARCH LABS

- **Exploratory Data Analysis:**

C. Univariate Analysis Steps

Conducted analysis using histograms, boxplots, and PDFs to understand the distribution of each feature.



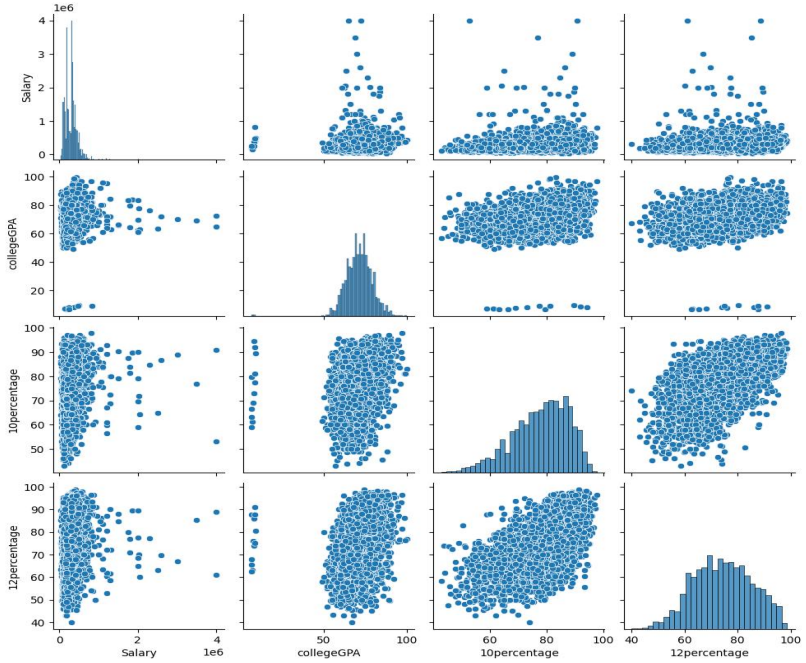
INNOMATICS RESEARCH LABS

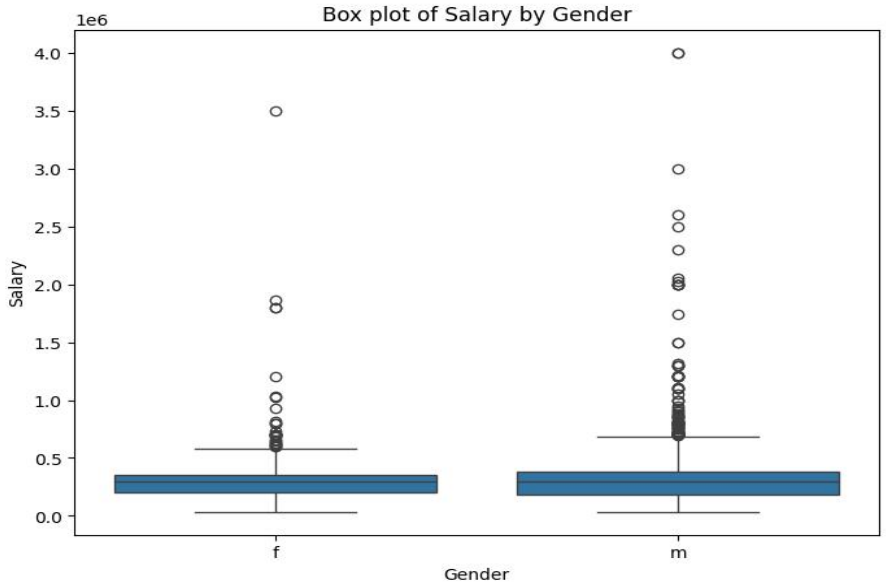
- **Exploratory Data Analysis:**

D) . Bivariate Analysis Steps

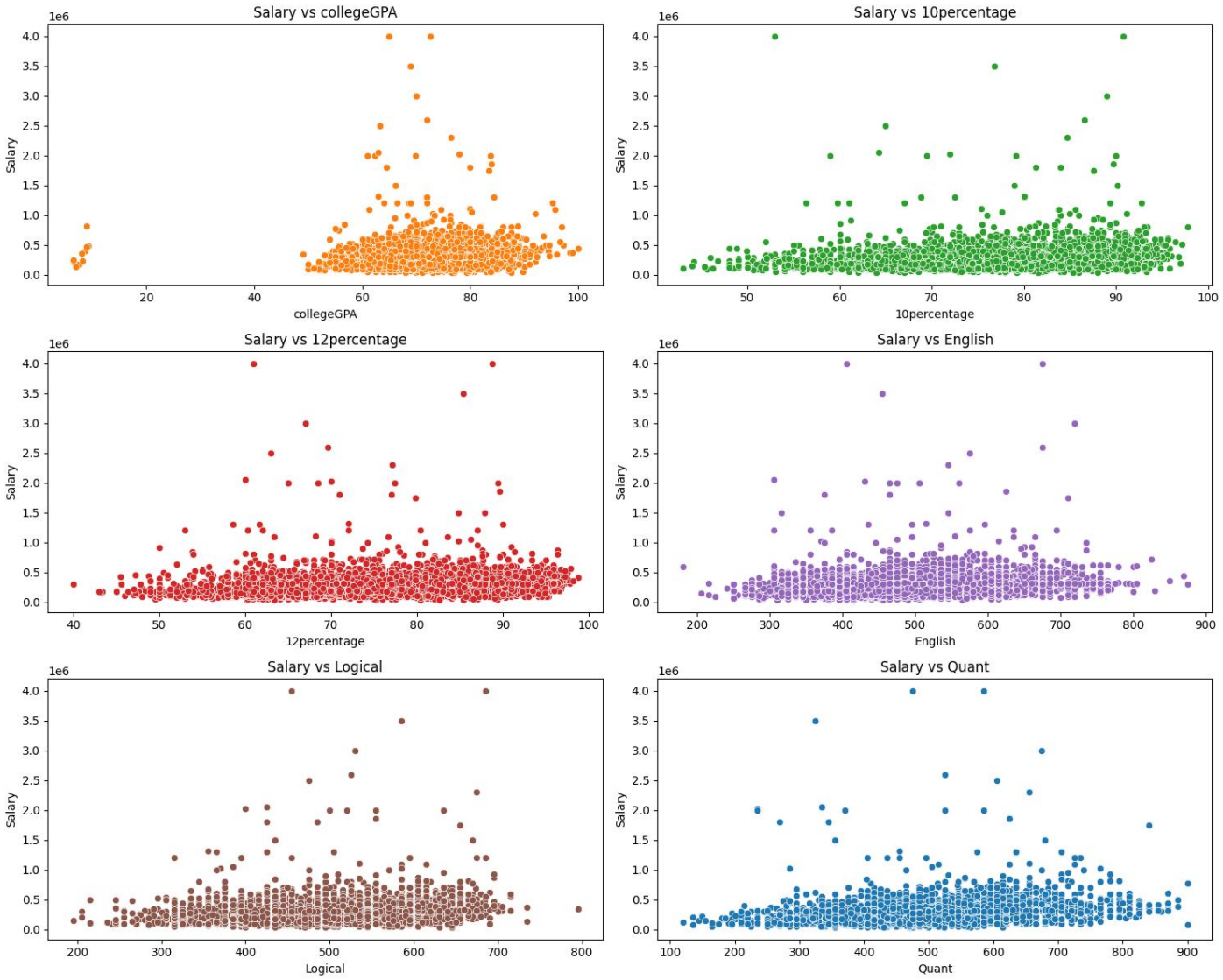
Used scatter plots, correlation matrices, and boxplots to explore relationships between the target variable and other features.

Analyzed the impact of categorical variables on salary.

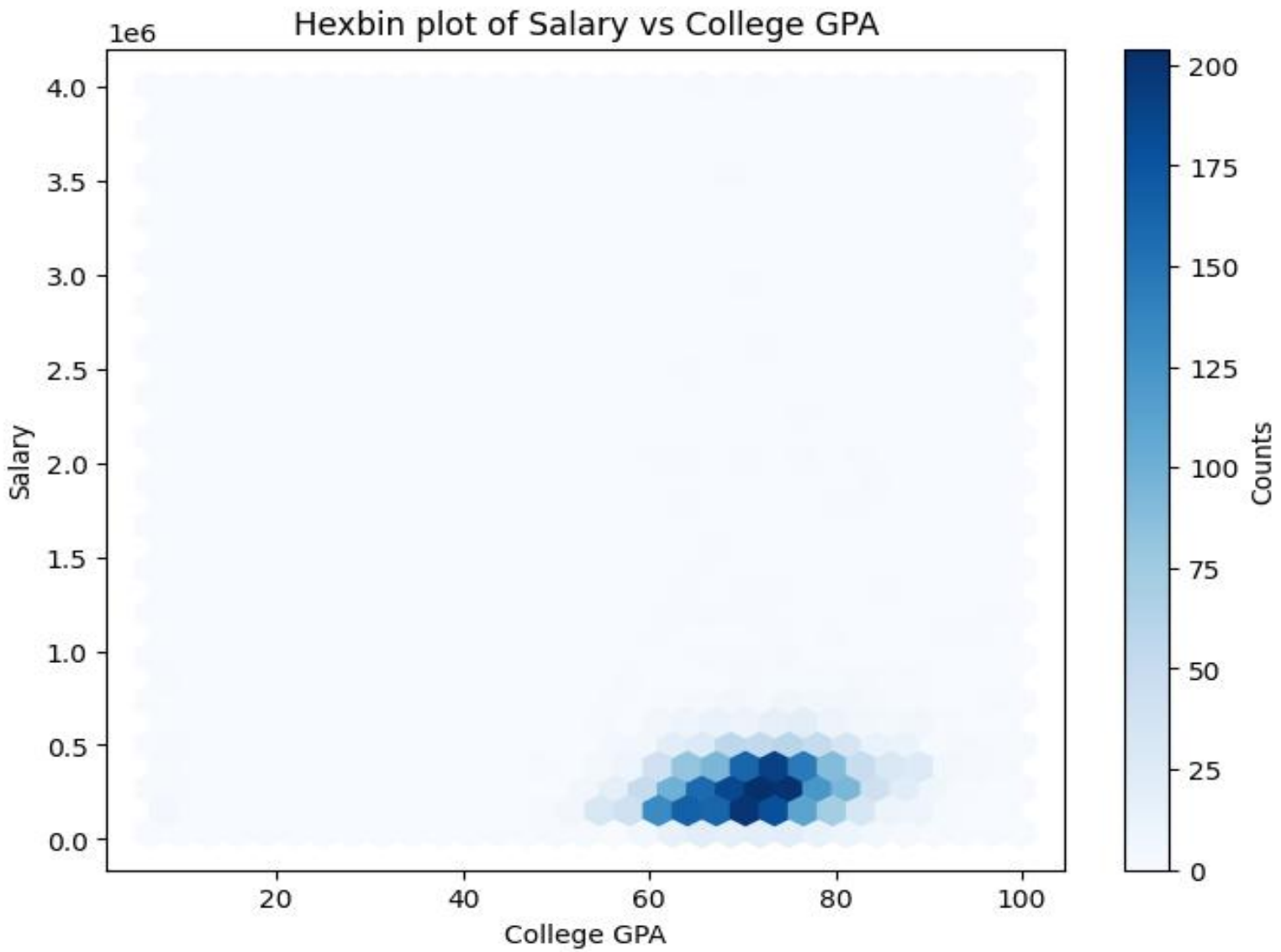




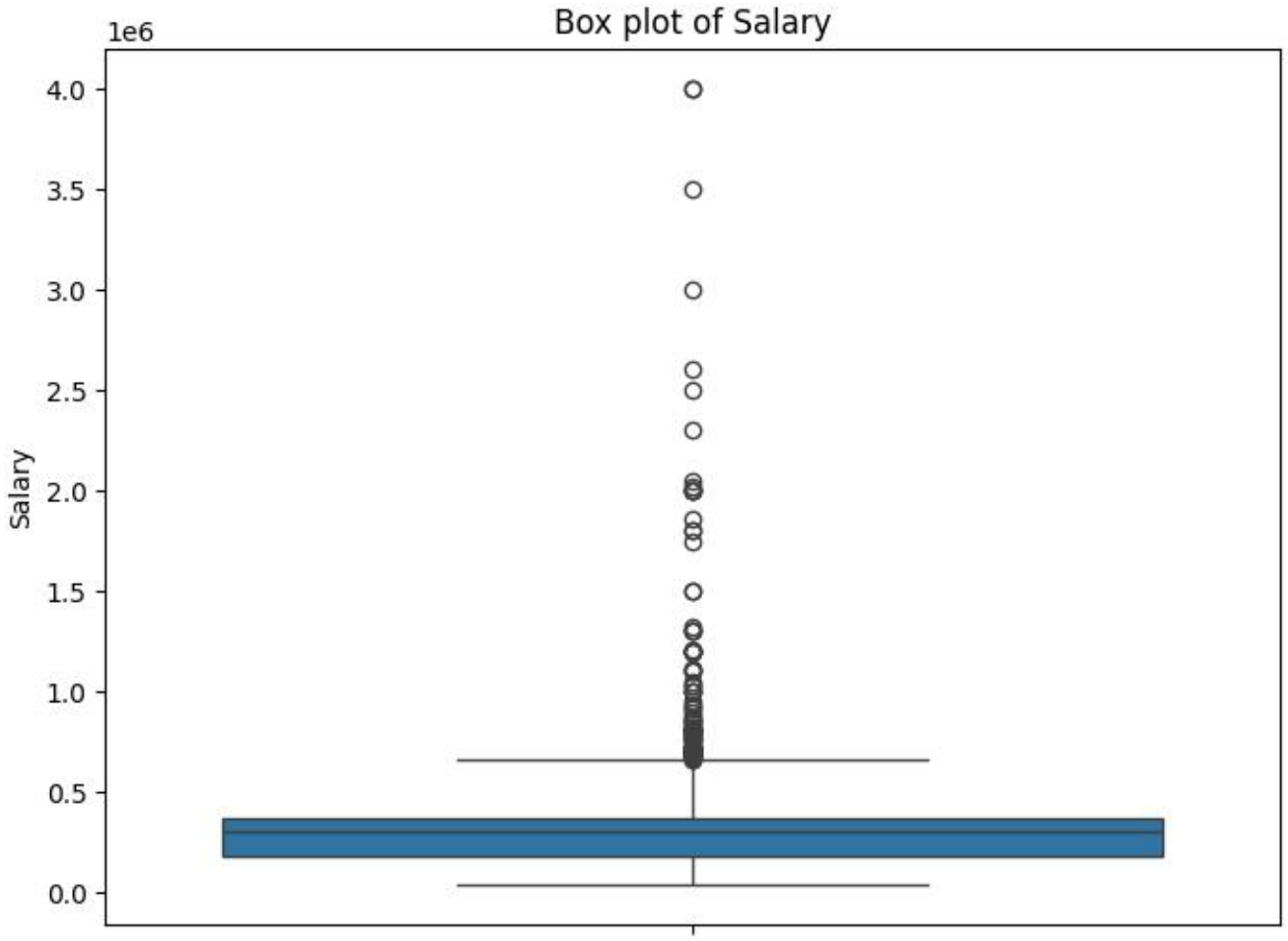
INNOMATICS RESEARCH LABS



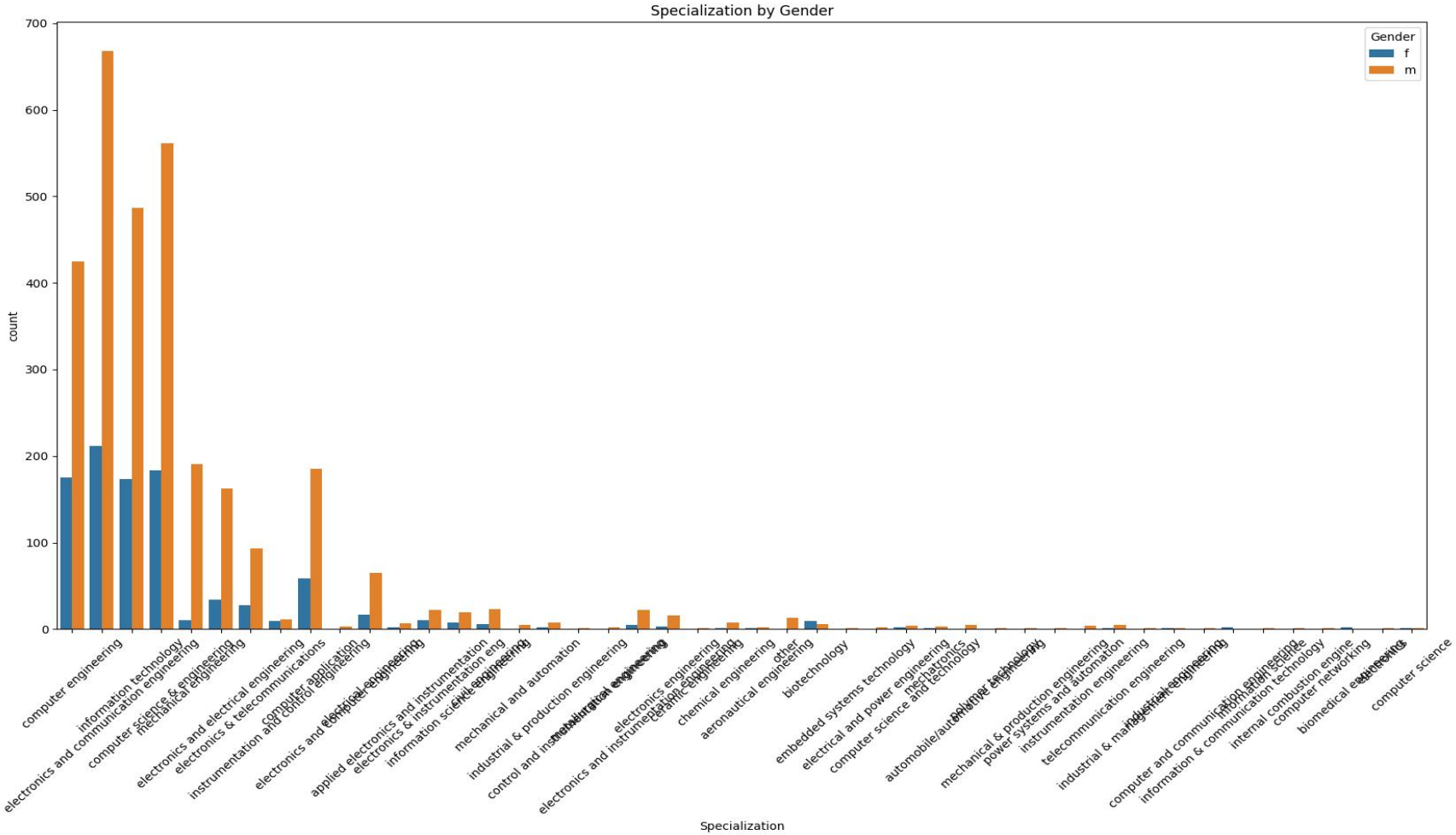
INNOMATICS RESEARCH LABS



INNOMATICS
RESEARCH LABS



INNOMATICS
RESEARCH LABS



INNOMATICS RESEARCH LABS

Conclusion (Key finding overall)

Open the floor for questions and discussions regarding the project.

THANK YOU

INNOMATICS
RESEARCH LABS