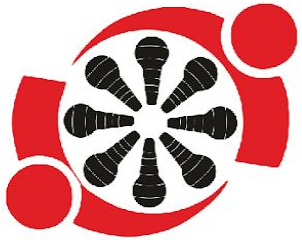


INNOMATİCS®
RESEARCH LABS

INNOVATION. AUTOMATION. ANALYTICS

PROJECT ON

AMCAT Exploratory Data Analysis(EDA)

About me

Parmod(IN9240744)

Background:

- Recently, I'm done M.Sc in Computer Science from MDU Rohtak.
- Fundamental understanding of computer science principles.
- Passionate for fusion of that knowledge with AI.

Motivation for Data Science:

- What is truly amazing is how such complex datasets are transformed into something actionable through data science.
- Made to use data-driven decisioning for the creation of innovative solutions in engineering and finance.
- He is fascinated by the dynamic nature of Data Science and continuous learning, new tools, and methodologies adaptation.

INNOMATICS RESEARCH LABS

Work Experience

Intern, Data Science, Acmegrade

- It constructed a movie recommendation system.
- Forecasted future sales.
- I learned about machine learning and time series analysis.
- I learned how data facilitates businesses.

Presently, Data Science Intern at the Innmatics Research Labs.

- To learn to use data in solving problems.
- It merges computer skills with data skills.
- Work begins on new and challenging projects.

GitHub : Parmod2310
LinkedIn : Parmod2310

INNOMATICS RESEARCH LABS

Objective:

Understand Dataset Structure:

- Explore the overall structure of the AMCAT dataset, including the types and distribution of variables related to candidates' profiles.
- Gender and Specialization Analysis: Analyze the relationship between gender and specialization to uncover any trends or preferences in candidate specialization choices.
- Salary as Target Variable: Investigate the factors influencing salary, focusing on identifying trends, correlations, and patterns between independent variables (such as specialization, gender, etc.) and the target variable (Salary).
- Feature Distribution and Outliers: Visualize and interpret the distribution of various features (e.g., experience, skills, etc.), identifying outliers or unusual patterns in the dataset.
- Insights and Correlations: Summarize key insights and relationships, particularly looking for dependencies between variables like gender, specialization, and their impact on salary.

Description:

- The dataset contains information on 3,998 individuals, spanning across 39 columns.
- Each row represents a unique individual, while each column provides specific details about their employment and educational background.
- The dataset offers a rich source of information for exploring employment outcomes, educational backgrounds, and personality traits among engineering graduates.
- With a diverse range of variables, it provides ample opportunities for in-depth analysis and insights into factors influencing career trajectories.

Process of the Project

1. Importing the libraries ,data and finding the shape and info of the data

2. Univariate Analysis

3. Bivariate Analysis

4. Research Questions

5. Research Questions

Importing the libraries ,data and finding the shape and info of the data

In first step we have imported the libraries required for the analysis and then imported the dataset.
The shape and info() of the dataset for further analysis.

```
# Loading Data
df=pd.read_csv("data.csv")
```

```
display (df.shape)  # Number of records- Shape
```

```
(3998, 39)
```

```
display(df.head())
```

	Unnamed: 0	ID	Salary	DOJ	DOL	Designation
0	train	203097	420000.0	6/1/12 0:00	present	senior quality engineer
1	train	579905	500000.0	9/1/13 0:00	present	assistant manager
2	train	810601	325000.0	6/1/14 0:00	present	systems engineer
3	train	267447	1100000.0	7/1/11 0:00	present	senior software engineer

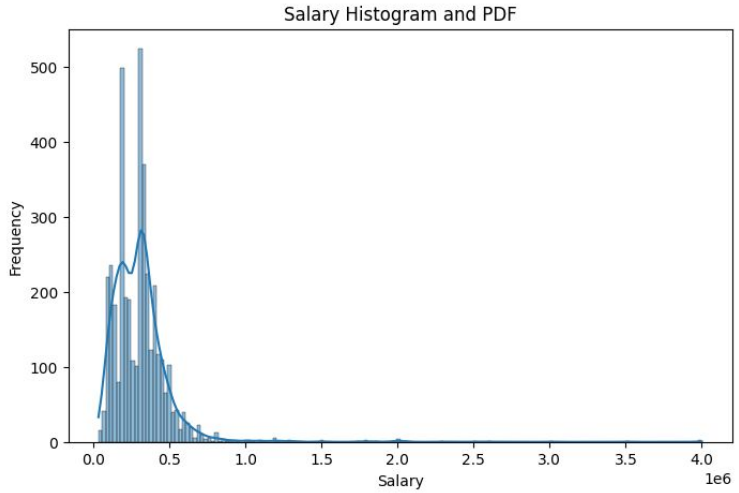
```
print (df.info()) # Data set details - Info
```

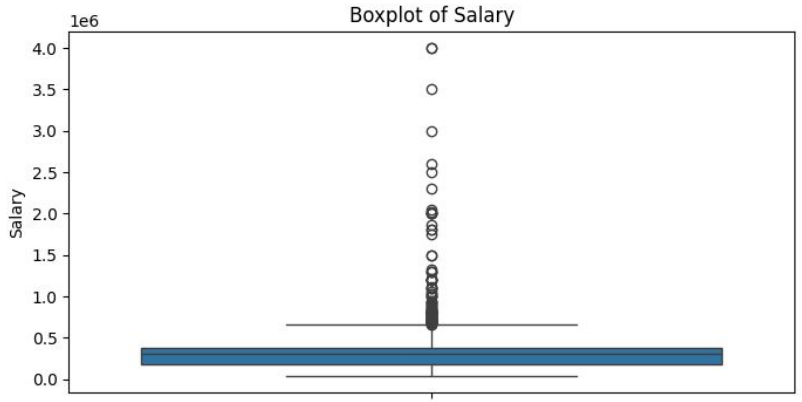
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3998 entries, 0 to 3997
Data columns (total 39 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   Unnamed: 0             3998 non-null   object
 1   ID                     3998 non-null   int64
 2   Salary                 3998 non-null   float64
 3   DOJ                    3998 non-null   object
 4   DOL                    3998 non-null   object
 5   Designation            3998 non-null   object
 6   JobCity                3998 non-null   object
```

INNOMATICS RESEARCH LABS

Univariate Analysis:

Boxplot : In the salary column we can see that the large number of outliers are present, to make the data more clear we can remove the data that is > 2.0
Histplot : The data is right skewed



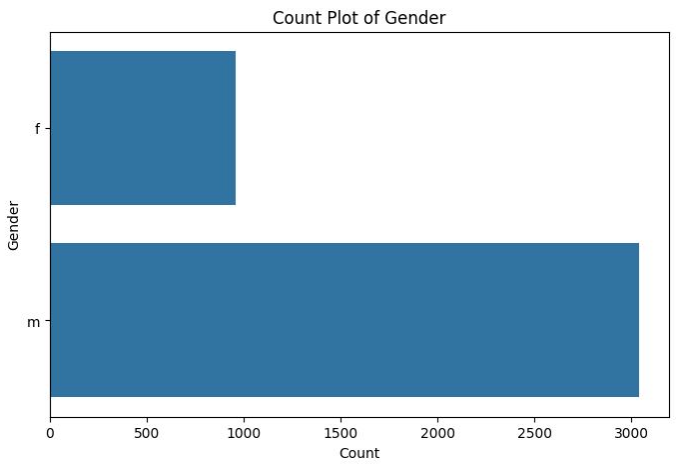


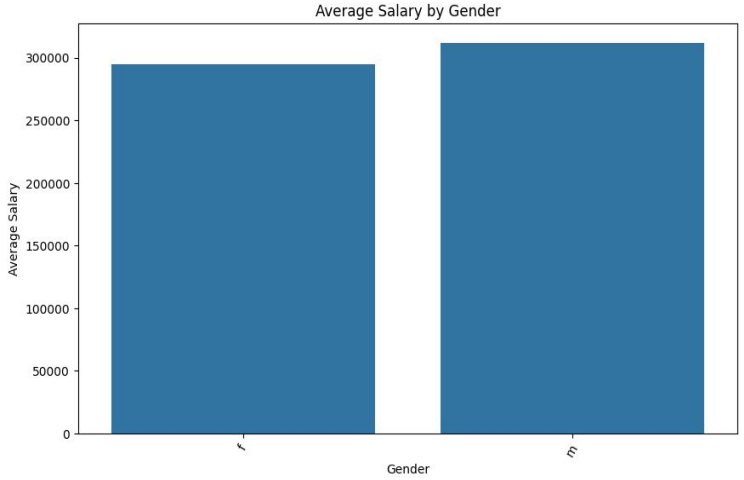
Bar plot

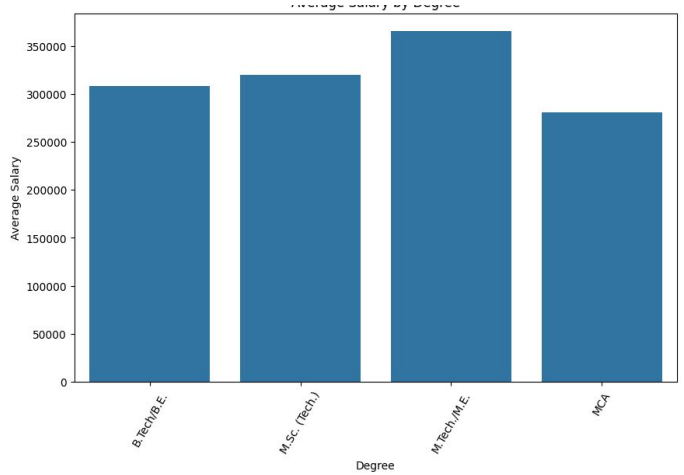
Analysis of the Gender column reveals a substantial gender gap, with a proportion of approximately 1 female for every 3 males.

We can observe that compare to Female Male has the higher salary.

Senior software engineer has the highest salary followed by M.Tech/M.E.

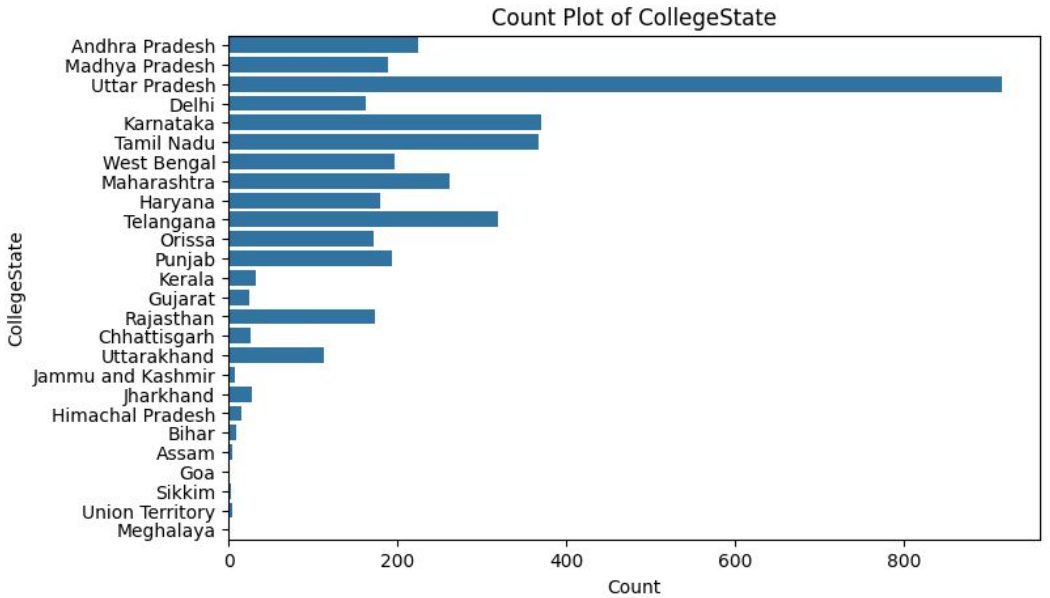






INNOMATICS RESEARCH LABS

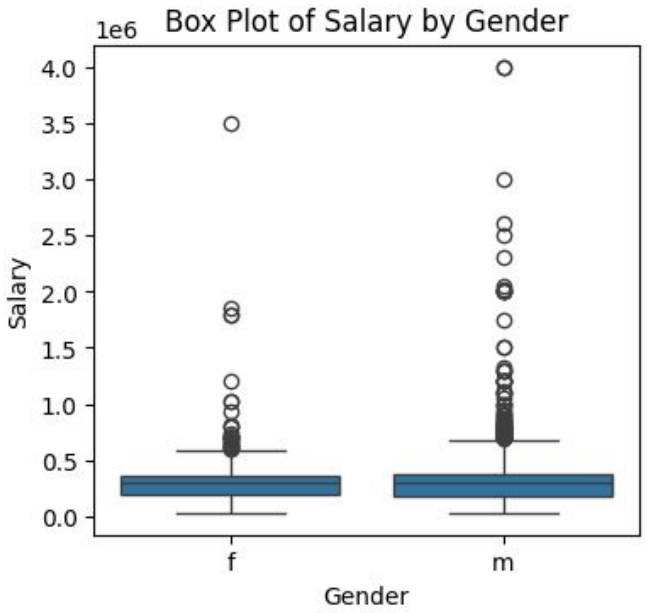
- A significant majority,Collegestate from Uttar Pradesh dominate, indicating a strong presence of graduates from this region.
- • Following Uttar Pradesh, Karnataka and Tamil Nadu are notable for their college representation.
- • Conversely, Meghalaya and Goa have fewer graduates, suggesting lower college participation rates.

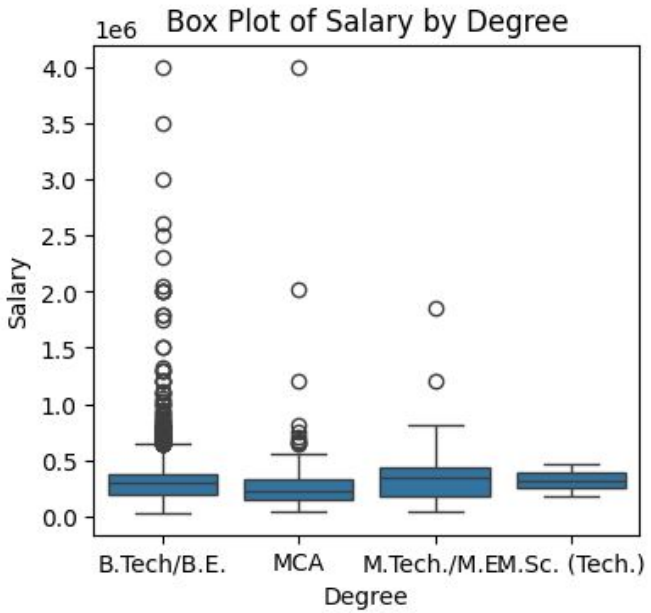


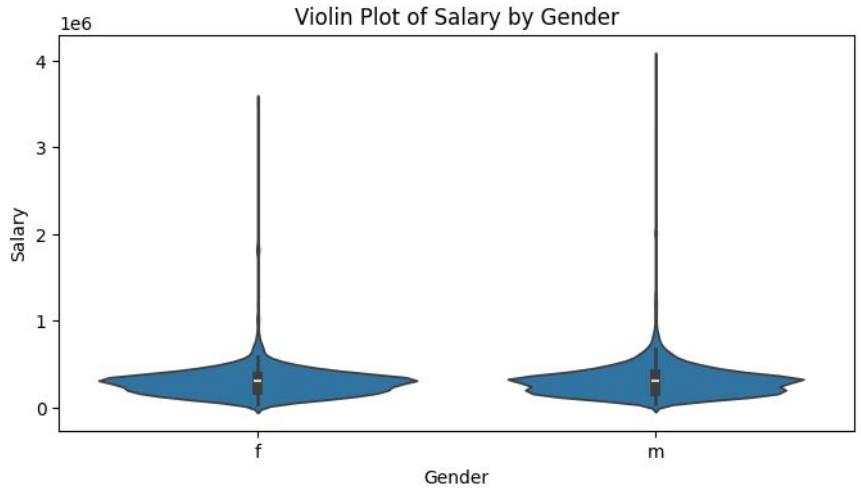
INNOMATICS RESEARCH LABS

Box Plot & Violin Plot

Men are earning more than women.
M.Tech/M.E graduates generally earn higher average salaries.







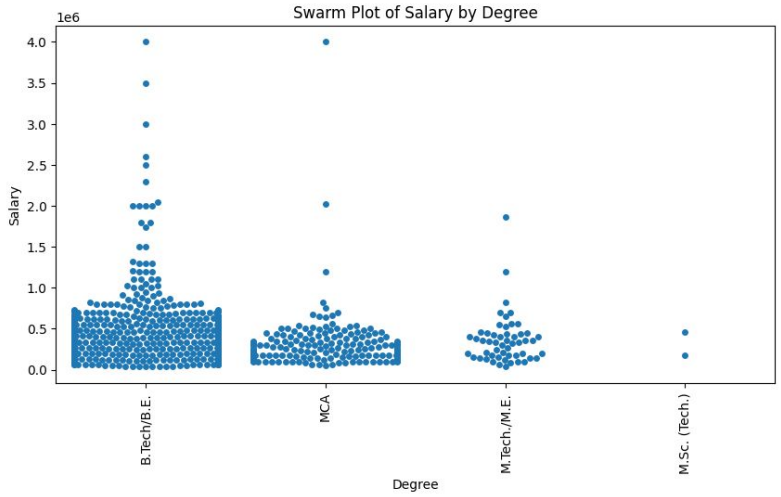
INNOMATICS RESEARCH LABS

Bivariate Analysis

Swarm Plot

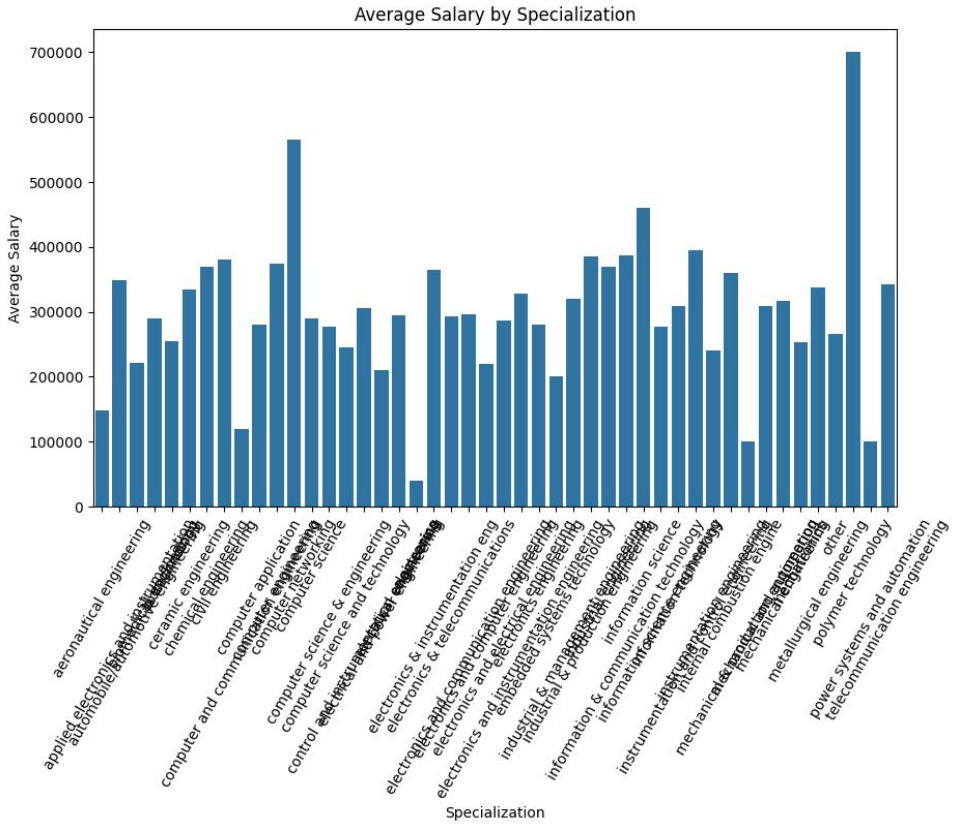
Salary Vs Degree & Salary Vs Experience

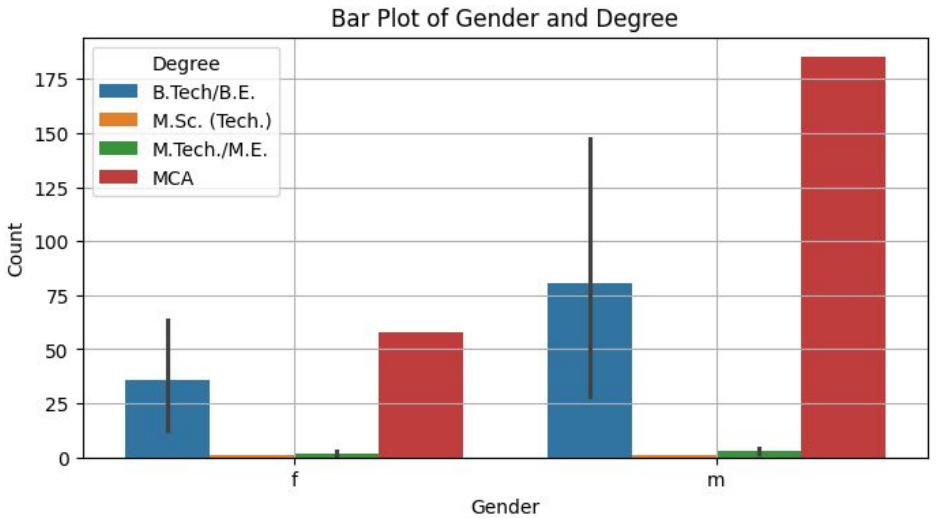
- M.Tech/M.E graduates generally earn higher average salaries.
- Salaries tend to increase with more years of experience



INNOMATICS RESEARCH LABS

- Computer Application represent the second highest earned underscoring the significant earning potential associated with technical roles.
- Higher enrollment of females in the MCA program compared to their male counterparts, reflecting a growing trend of female participation in technology and computer applications

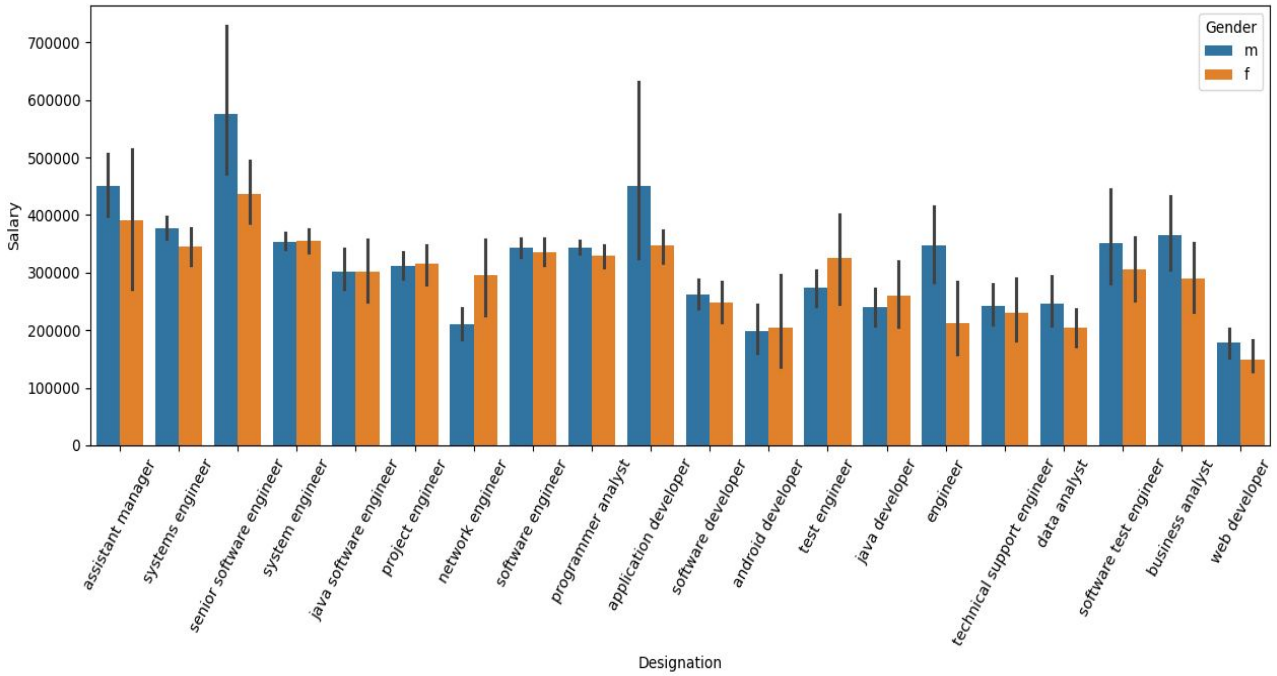


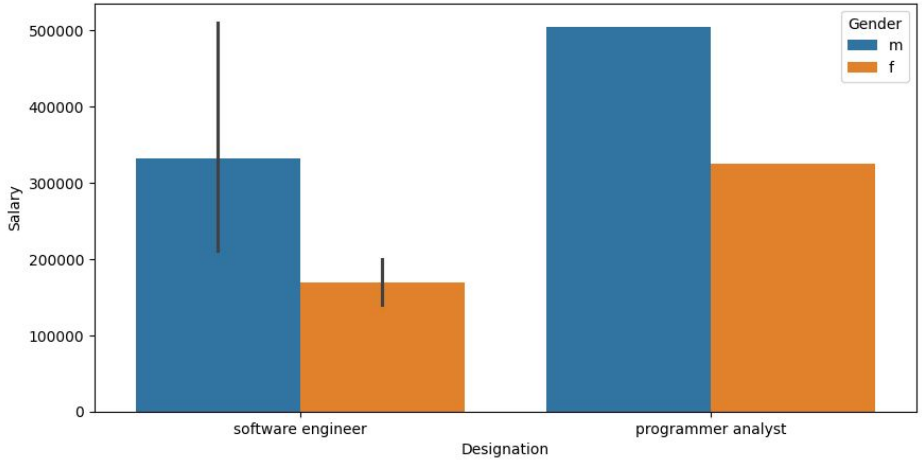


INNOMATICS RESEARCH LABS

Designation Vs Salary

Senior software engineer has the highest salary followed by system engineer.





INNOMATICS RESEARCH LABS

Conclusions:

- Based on the Analysis made and we have found the relationship between the target variable Salary and other variables. Here are some insights
- Male has the higher salary compared to female - Senior software Engineer has the highest salary compared to other Designations - Most of the software engineers are from Specilization Computer science , Electronics , Information Technology.

THANK YOU

INNOMATICS RESEARCH LABS