INNOMATICS®
RESEARCH LABS

INNOVATION. AUTOMATION. ANALYTICS

PROJECT ON

Exploratory Data Analysis on AMEO Dataset

About me

I am a student who is currently pursuing BTech Computer Science.I am passionate about data analysis and its applications in various domains.I have always been fascinated by the power of data and how it can be used to solve real-world problems. I am interested in data analysis because I enjoy finding patterns, insights, and trends from large and complex datasets.

Connect with Me on Linkedin

INNOMATICS RESEARCH LABS

Objective

- **To explore the employment outcomes and skill assessments of engineering graduates in India, based on a dataset from Aspiring Minds.**
- **To identify the factors that influence the salary, job title, and job location of engineering graduates, such as college tier, degree, specialization, domain, and personality traits.**

About the data set

The dataset was released by Aspiring Minds from the Aspiring Mind Employment Outcome 2015 (AMEO). The study is primarily limited only to students with engineering disciplines. The dataset contains the employment outcomes of engineering graduates as dependent variables (Salary, Job Titles, and Job Locations) along with the standardized scores from three different areas – cognitive skills, technical skills and personality skills. The dataset also contains demographic features. The dataset contains around 40 independent variables and 4000 data points. The independent variables are both continuous and categorical in nature. The dataset contains a unique identifier for each candidate. Below mentioned table contains the details for the original dataset.

INNOMATICS RESEARCH LABS

Data cleaning and preprocessing

1.Used df.info(), df.shape, df.head(), df.describe(), and df.columns to get some basic information about the dataset, such as the number of rows, columns, data types, missing values, summary statistics, and column names.

2. Used pd.to_datetime() to convert the columns ‘DOJ’, ‘DOL’, ‘DOB’, and ‘12graduation’ to datetime format, which makes it easier to manipulate and analyze the data values.

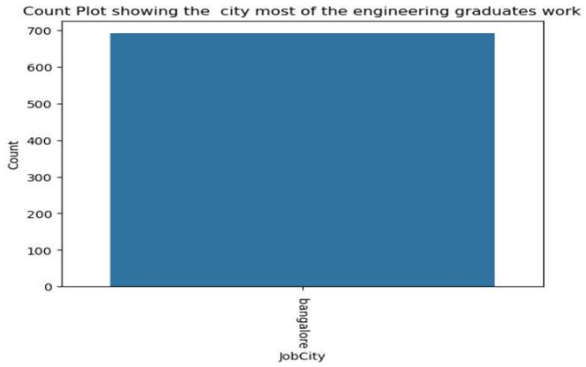
3. Used df.replace() and df.median() to replace the missing values (represented by -1) in the columns ‘ComputerProgramming’, ‘MechanicalEngg’, ‘ElectricalEngg’, ‘TelecomEngg’, and ‘CivilEngg’ with the median values of each column.

4. Used df.drop() to drop the columns ‘Unnamed: 0’, ‘ID’, ‘CollegeID’, which are either redundant or irrelevant for the analysis.

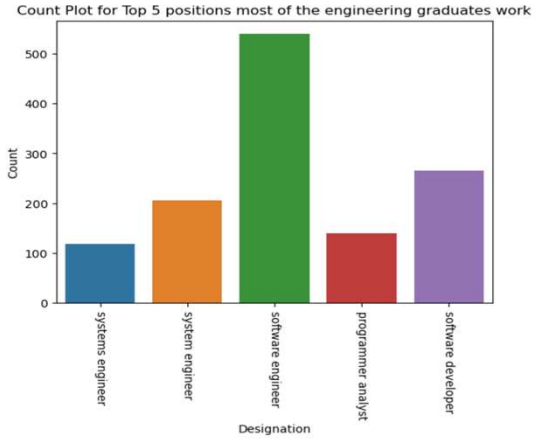
5.Used astype() to convert the columns ‘CollegeTier’ and ‘CollegeCityTier’ to string type, since they are categorical variables and not numerical.

Key Findings

- The average salary for engineering graduates is around 3 lakhs
- Most engineering graduates work in Bangalore

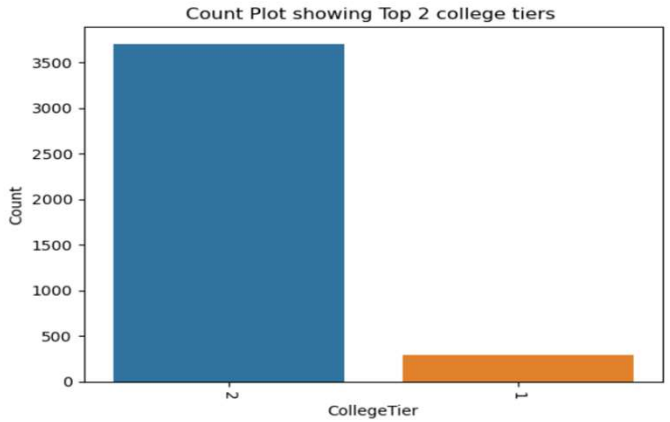


- Most of the engineering graduates work as Software Engineers



INNOMATICS RESEARCH LABS

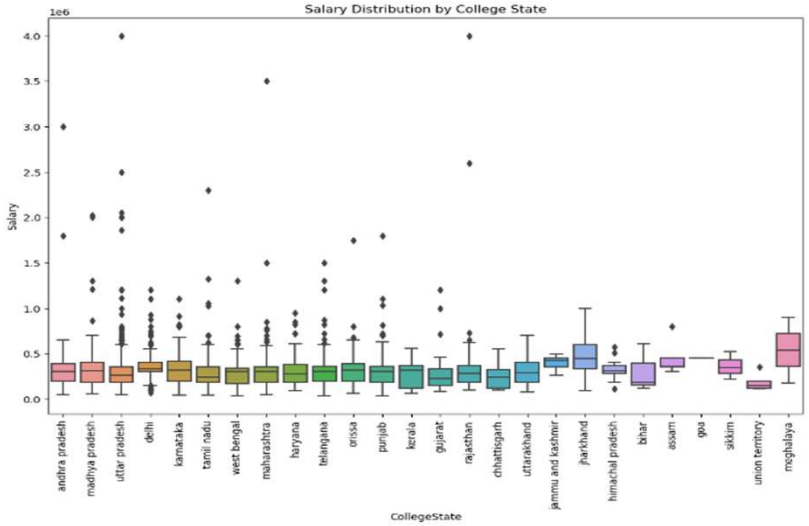
- Graduates from tier one college are tend to get paid more
- Graduates who got a job are mostly from tier 2 colleges



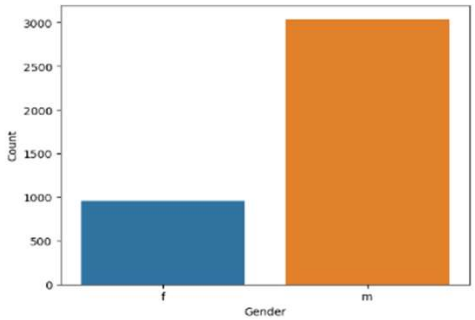
- The most common college city tier among the graduates is Tier 0
- The graduates from Tier 1 cities have a slightly higher average salary than those from Tier 0 cities.

INNOMATICS RESEARCH LABS

- The median salary is highest for Delhi, followed by Goa and Himachal Pradesh. The lowest median salary is for Jharkhand, followed by Assam and Bihar.



- Female graduates are less in number when compared to male graduates



INNOMATICS RESEARCH LABS

"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."

```
df = df[(df['Degree'] == 'B.Tech/B.E.') & (df['Specialization'].str.contains('computer')) & (df['Designation'].isin(['programming

# Calculate the mean salary for the filtered data
mean_salary = df['Salary'].mean()

# Print the mean salary and compare it with the hypothesis
print(f'The mean salary for computer engineering graduates with the given designations is {mean_salary:.2f}')
if mean_salary >= 250000 and mean_salary <= 300000:
    print('The hypothesis is supported by the data.')
else:
    print('The hypothesis is not supported by the data.')
```

```
The mean salary for computer engineering graduates with the given designations is 353534.14
The hypothesis is not supported by the data.
```

- The mean salary for computer engineering graduates with the given designations is 353534.14 The hypothesis is not supported by the data.
-

INNOMATICS RESEARCH LABS

Is there a relationship between gender and specialization?

```
df = df[(df['Degree'] == 'B.Tech/B.E.') & (df['Specialization'].str.contains('computer')) & (df['Designation'].isin(['programming

# Calculate the mean salary for the filtered data
mean_salary = df['Salary'].mean()

# Print the mean salary and compare it with the hypothesis
print(f'The mean salary for computer engineering graduates with the given designations is {mean_salary:.2f}')
if mean_salary >= 250000 and mean_salary <= 300000:
    print('The hypothesis is supported by the data.')
else:
    print('The hypothesis is not supported by the data.')
```

```
The mean salary for computer engineering graduates with the given designations is 353534.14
The hypothesis is not supported by the data.
```

- the p-value (.072) of the test is larger than the significance level (0.05) Hence we can say that gender and specialization are related

INNOMATICS RESEARCH LABS

THANK YOU

INNOMATICS
RESEARCH LABS