INNOMATİCS®
RESEARCH LABS

INNOVATION. AUTOMATION. ANALYTICS

PROJECT ON

Exploratory Data Analysis on AMEO Dataset

Ramesh
M.Sc. Applied Statistics
2017

INNOMATICS
RESEARCH LABS

OBJECTIVE OF THE PROJECT :

This analysis aims to gain insights and understanding from the provided dataset, particularly focusing on the relationship between various features and the target variable, which is Salary.
Specifically, the goals of this analysis include:

- Describing the dataset and its features comprehensively.
- Identifying any patterns or trends present in the data.
- Exploring the relationships between independent and Salary.
- Identifying any outliers or anomalies in the data.

SUMMARY OF DATA :

The Aspiring Mind Employment Outcome 2015 (AMEO) dataset, released by Aspiring Minds, focuses on employment outcomes for engineering graduates. It includes dependent variables such as Salary, Job Titles, and Job Locations, along with standardized scores in cognitive skills, technical skills, and personality skills. With around 40 independent variables and 4000 data points, these variables encompass both continuous and categorical data. The dataset also includes demographic features and unique identifiers for each candidate.

Data Cleaning :

Assuming 'data' is your Data Frame and 'columns to check' are the columns with -1 values

Computer Programming 21.710855
ElectronicsAndSemicon 71.385693
Computer Science 77.438719
Mechanical Engineering 94.122061
Electrical Engineering 95.972986
Telecom Engineering 90.645323
Civil Engineering 98.949475

In the given columns, some columns have null values exceeding 70%, while the remaining columns have zero null values.

Replace -1 with Nan for numerical & objective columns

Datatype Conversion

```
df1['DOB'] =
pd.to_datetime(df1['DOB'])
df1['DOL'] =
pd.to_datetime(df1['DOL'])
df1['DOJ'] =
pd.to_datetime(df1['DOJ'])
```

Outlier:

count 3.998000e+03
mean 3.076998e+05
std 2.127375e+05
min 3.500000e+04
25% 1.800000e+05
50% 3.000000e+05
75% 3.700000e+05
max 4.000000e+06
Name: Salary,
dtype: float64

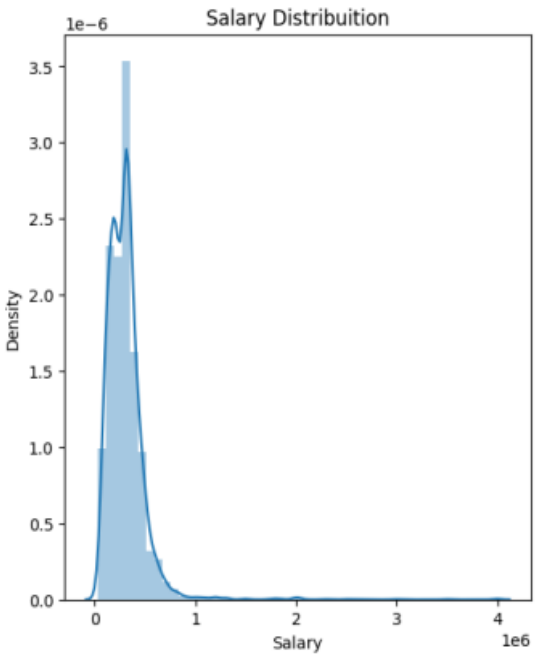
Outlier Remove:

```
count 3889.000000
mean 285447.158653
std 126514.542129
min 35000.000000
25% 180000.000000
50% 300000.000000
75% 360000.000000
max 655000.000000
Name: Salary,
dtype: float64
```

UNI variate :

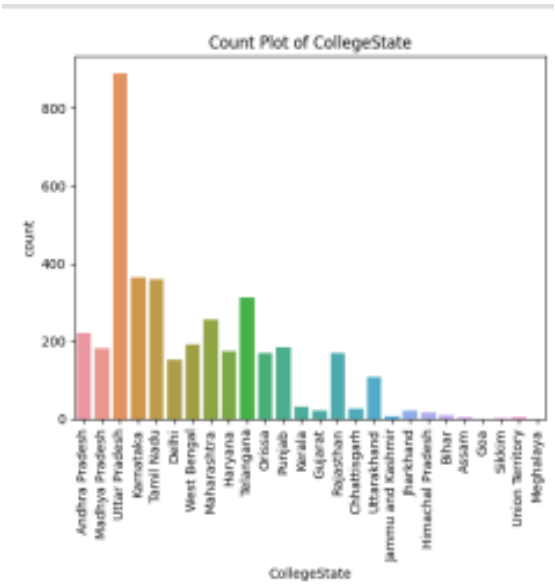
Continuous Features:

Salary :
The summary plot indicates considerable variation, while the histogram reveals significant positive skewness



Categorical Features:

College State:
Most of students studying in Uttar Pradesh

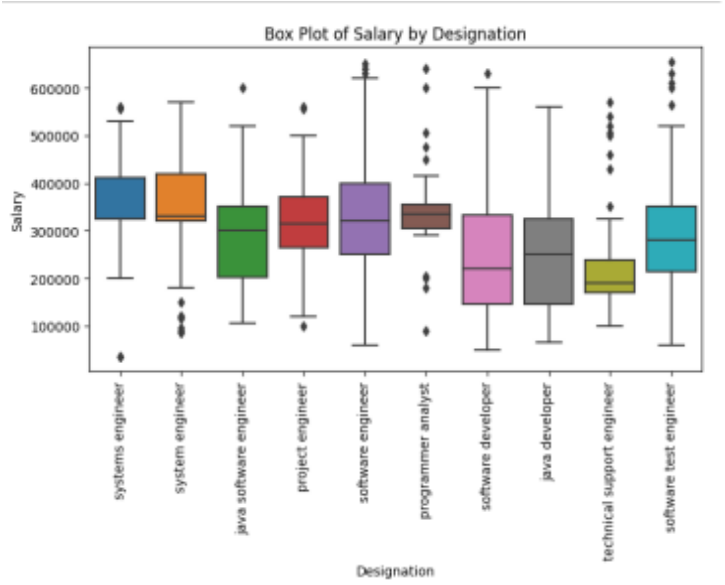


INNOMATICS RESEARCH LABS

Bivariate analysis:

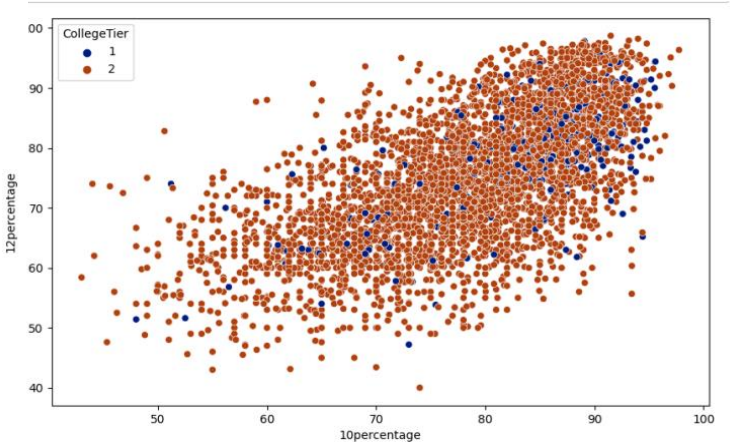
NUM vs Num:

10percentage vs 12percentage: correlation plot and scatterplot, it is evident that the 10th and 12th grades are positively correlated



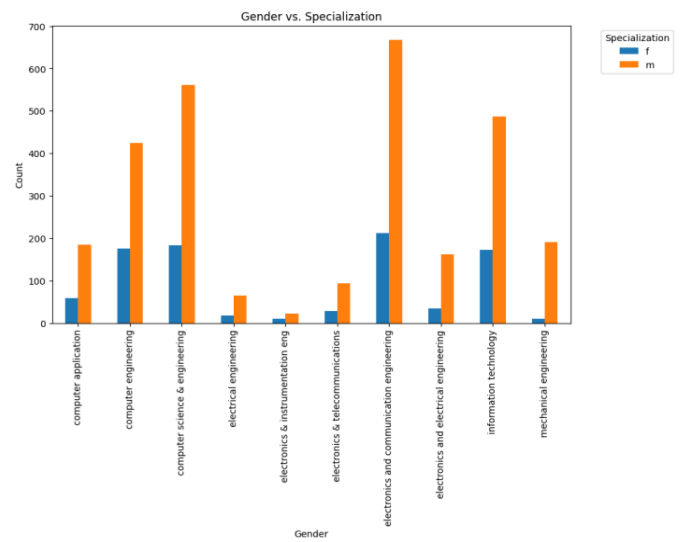
NUM vs CAT:

Designation vs Salary: In the plot, senior software engineers earn more than other software employees.



CAT vs CAT

Specialization vs Gender: females in every specialization. Also, there are very less number of females who opted for mechanical



RESEARCH OUTCOMES

"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."

With an average salary of approximately 332,943.26 INR, the data suggests that fresh graduates in Computer Science Engineering working in the specified job roles (Programming Analyst, Software Engineer, Hardware Engineer and Associate Engineer

Is there a relationship between gender and specialization? (i.e. Does the preference of Specialization depend on the Gender?)

Chi-squared statistic: 104.46891913608455 P-value: 1.2453868176976918e-06
The small p-value suggests that the observed distribution of specializations is unlikely to have occurred by chance if there were no actual relationship between gender and specialization.

CONCLUSION:

Senior Software Engineers earn the highest salaries but with more variability, while Software Developers and Technical Support Engineers earn below the average. Gender has a minor impact on average salary, with females earning less. Academic performance, based on 10th, 12th, and college GPA scores, does not show a clear correlation with pay.