INNOMATICS®
RESEARCH LABS

INNOVATION. AUTOMATION. ANALYTICS

PROJECT ON

Exploratory Data Analysis on AMEO Dataset

Yuvasai Uppala

23rd February 2024

About me

I hold a Bachelor of Science (BSc) degree in Chemistry. Despite my non-technical background, my passion for understanding patterns and uncovering insights led me to explore data-driven fields. I embarked on a journey into the fascinating world of data analysis and data science and realized that it offer a unique blend of scientific inquiry and practical problem-solving.

I believe that data-driven decisions can transform businesses, research, and society. By mastering data analysis techniques, I aim to contribute to informed decision-making and drive positive change.

I have a hands on project experience in using MySQL, Excel in different projects across different domains like banking, social media, automobile, human resources and Entertainment industry.

LinkedIn profile : LinkedIn Profile
Link to project Repo: EDA on AMEO DATA

INNOMATICS RESEARCH LABS

1. About the Project

This project is about doing EDA on the given AMEO dataset, particularly identifying relationships between various variable and the target variable which is Salary.

2. Objective

- Explore the dataset and describe/understand the data.
- Find the relationships between the variables and target variable.
- Identify trends/patterns in the data.
- Identify any outliers from the dataset.

3. Summary of Data:

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.

4. Data Cleaning:

- Removed Unnecessary columns.
- Searched for nulls and duplicates.
- Corrected date columns datatypes.
- Checked whether DOL is earlier than DOJ and drops those rows if true.

5. Exploratory Data Analysis

5i) Univariate Analysis

a)Statistical Non visual analysis

The following are the observations from the analysis of **categorical variables:

- There are 417 different designations, among them Software Engineer is the most occupied designation/position.

- Out of 338 different job cities, Banglore provided the most number of jobs.

- Males occupied the jobs 3 times more compared to females.

- Most of the people had cbse as their 10 and 12 board.

- People from B.Tech/ B.E with specialization in electronics and communication engineering received more jobs.

- Uttar Pradesh college students grabs more job opportunities compared to students from other state colleges.

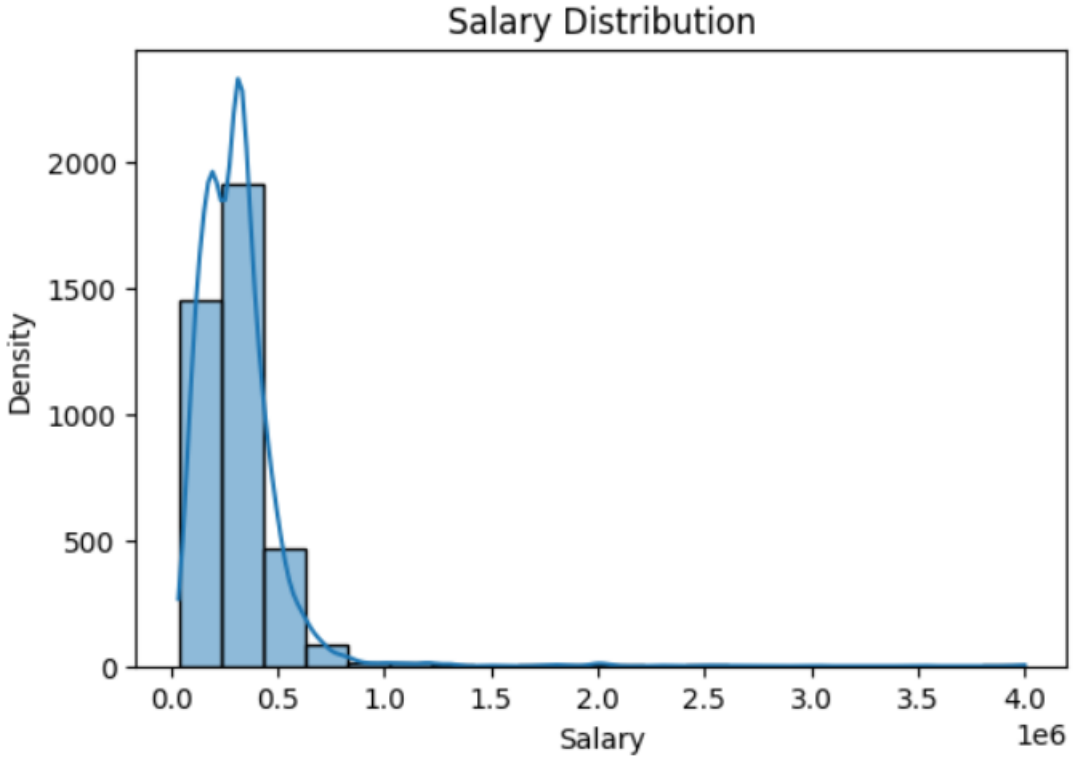
The following are the observations from the analysis of **discrete numerical variables:

- People graduated in 2013 grab more job opportunities.

- People from CollegeCityTier 0 and CollegeTier 2 received more jobs.

b) Visual Analysis

b1) Salary

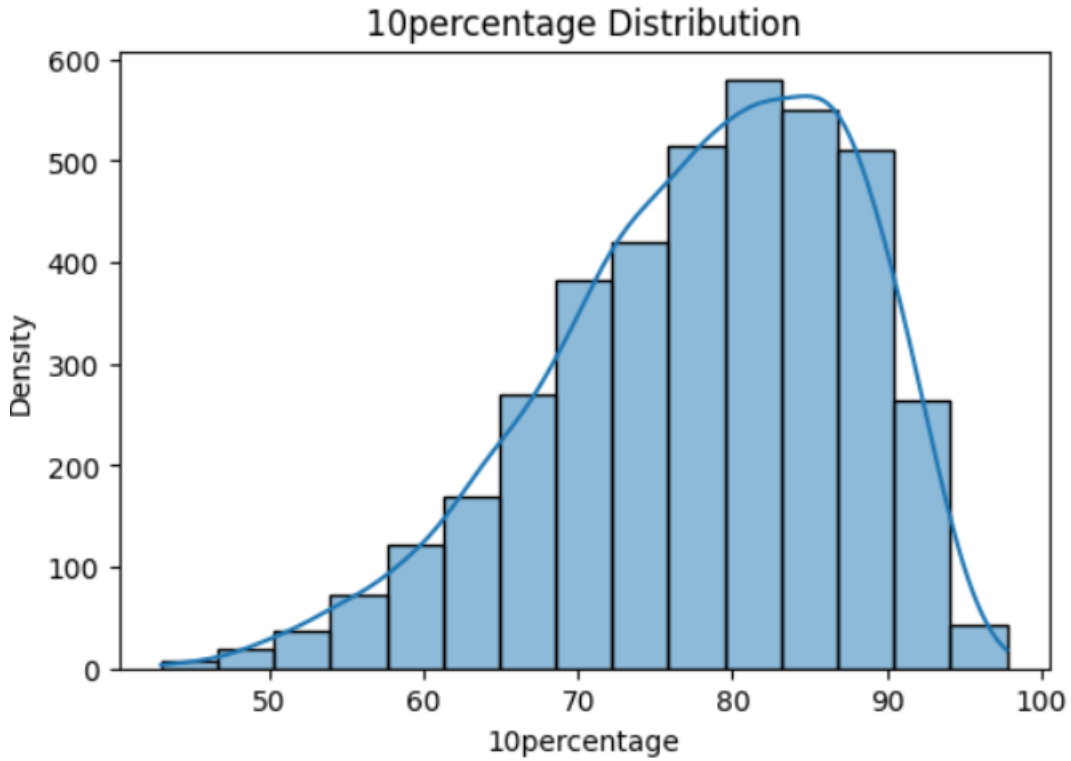


- Most of the peoples salaries falls under 0.5 million(5 lakhs).

- As salaries increase beyond this range, the frequency of individuals decreases significantly.

- The distribution is right-skewed, indicating that there are relatively fewer high earners compared to the middle-income group.

b2) 10percentage

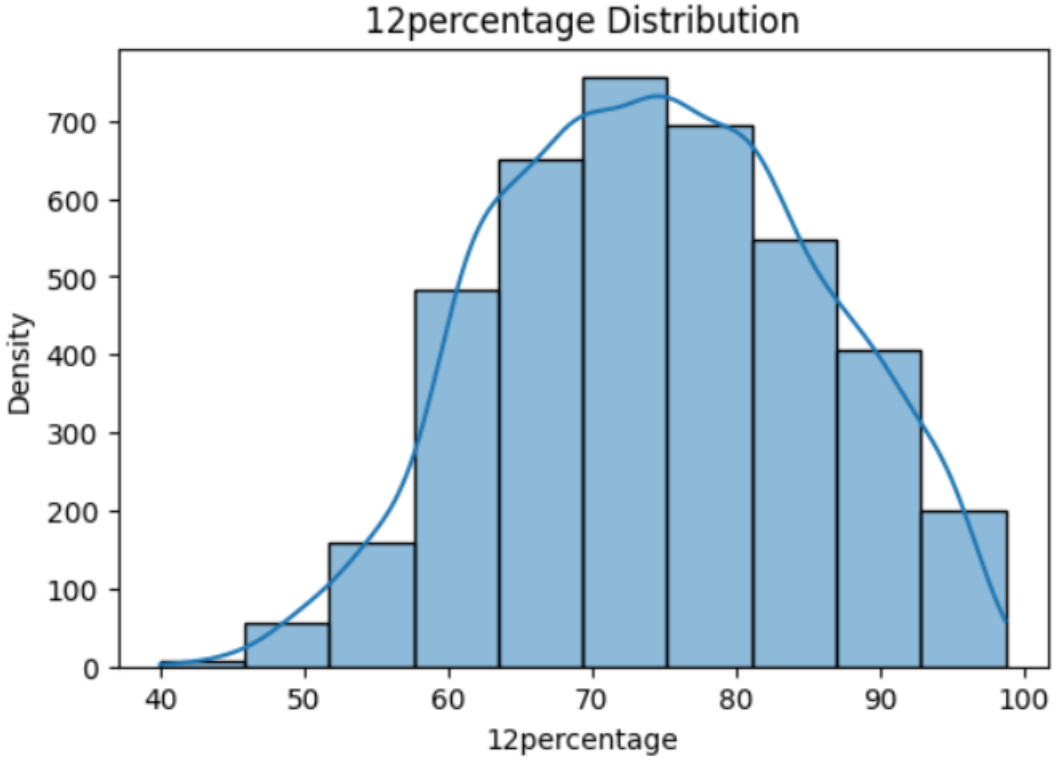


INNOMATICS RESEARCH LABS

- Very few people scored less than 50% during their 10th class and majority have scored in between 75%-90%.

- 80% is where the peak touched.

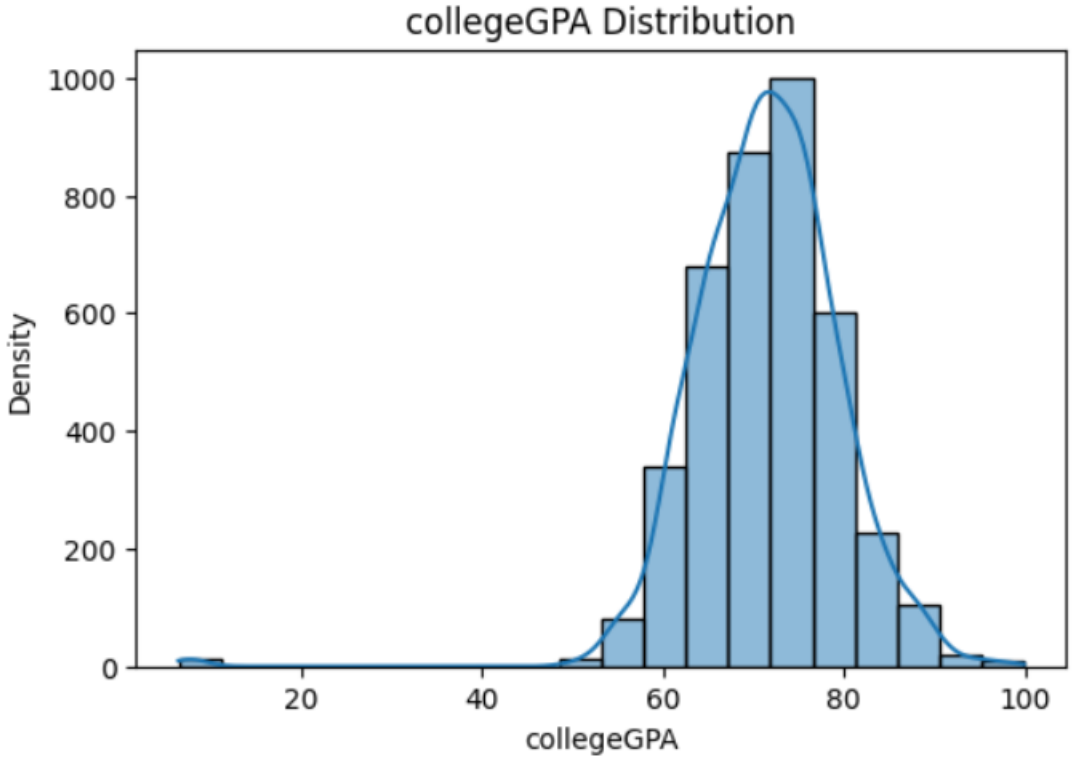
b3) 12percentage



- Very few people scored less than 50% during their 12th class and majority have scored in between 64%-81%.

- 75% is where the peak touched.

b4) collegeGPA



- Very few people scored less than 58% during their college and majority have scored in between 62%-81%.

- 72% is where the peak touched.

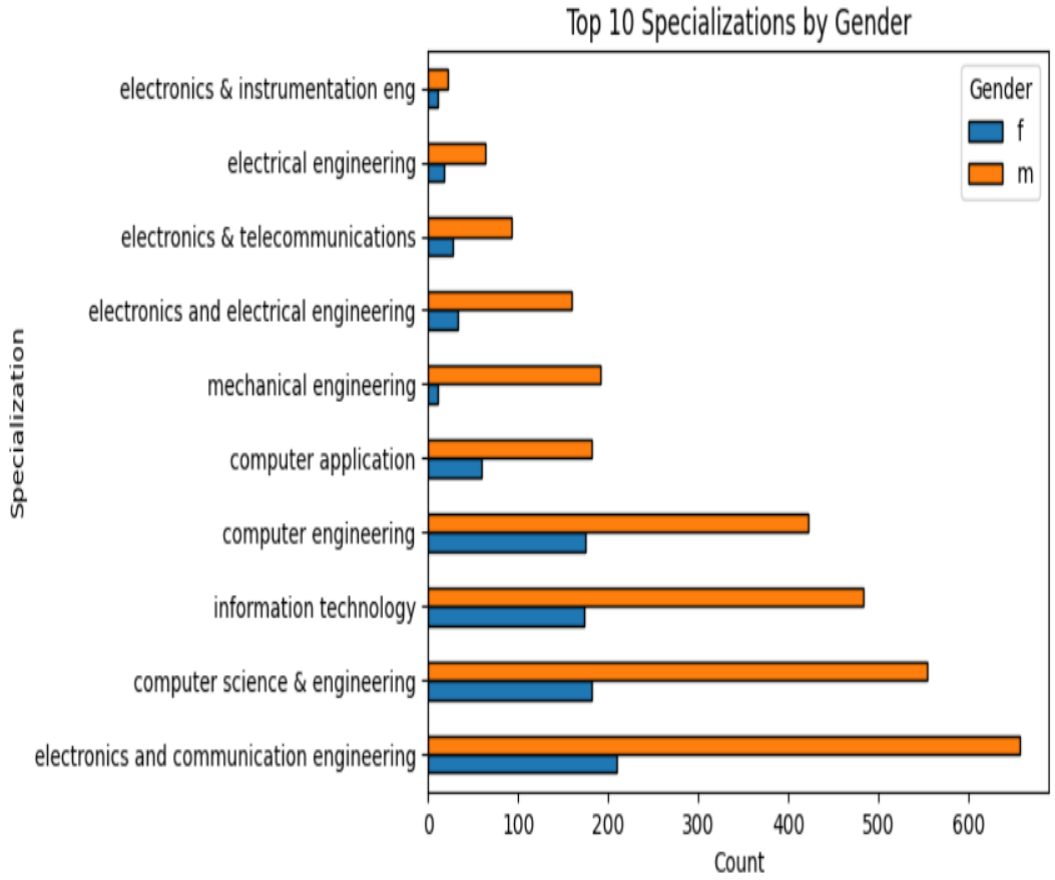
5ii) Bivariate Analysis

a) Numerical vs Numerical data

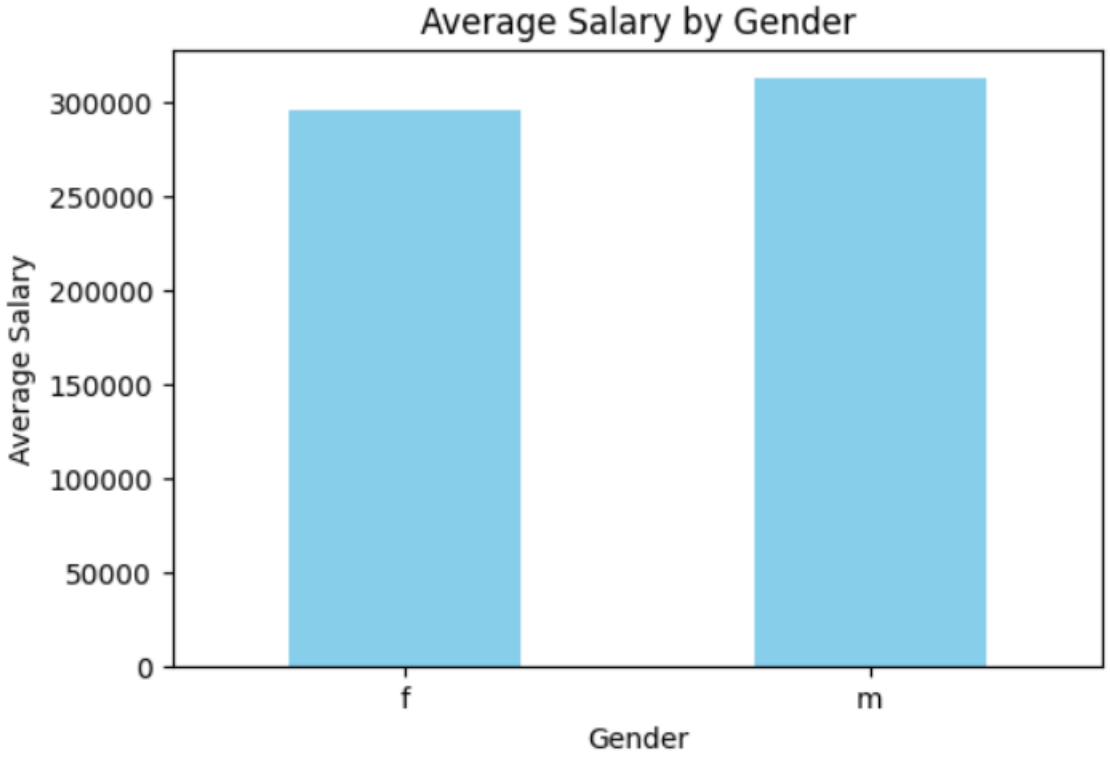
There exists **no correlation** between salary and other numerical columns like 10percentage, 12percentage, collegeGPA etc.,

b) Categorical vs categorical data

-Males occupied 3 times more than females in every specialization.



C) categorical vs Numerical data



-The avg salary for both male and female are approximately equal.

Conclusion:

Males occupied jobs more than females in every specialization. People graduated in 2013 grab more job opportunities and software engineer is the most occupied position. People from B.Tech/ B.E with specialization in electronics and communication engineering received more jobs.Uttar Pradesh college students grabs more job opportunities compared to students from other state colleges.

THANK YOU

INNOMATICS RESEARCH LABS