INNOMATICS®
RESEARCH LABS

INNOVATION. AUTOMATION. ANALYTICS

PROJECT ON

ANALYSIS OF AMCAT DATA

About me

- MSc
- Because I want to get more knowledge about data science. It will definitely help in future growth.
- I am a student.so no work experience on this field.
- LinkedIn Url: https://www.linkedin.com/in/shana-kk-055662251/
- Github Url : github.com/shana123kk

INNOMATICS RESEARCH LABS

Agenda

- **Business Problem and Use case domain understanding(If Required)**
- **Objective of the Project**
- **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

Domain Understanding

Understanding the employment outcomes of engineering graduates is crucial for various stakeholders, including educational institutions, employers, policymakers, and students themselves. By analyzing factors influencing employment outcomes, such as salary, we can gain valuable insights into the effectiveness of educational programs, the demand for specific skills in the job market, and potential areas for improvement in career development strategies.

Objective of the Project

The primary objective of this project is to perform Exploratory Data Analysis (EDA) on the provided dataset from the Aspiring Minds Employment Outcome 2015 (AMEO) study. Specifically, we aim to gain insights into factors influencing salary levels and employment patterns among engineering graduates. By conducting EDA, we can identify key variables that impact salary outcomes and understand the relationships between different factors, such as academic performance, specialization, college tier, and demographic variables.

Summary of the Data

The AMEO dataset contains comprehensive information about the employment outcomes of engineering graduates, with a focus on salary as the target variable. It includes data on approximately 40 independent variables and 4000 data points, covering both continuous and categorical variables. The dataset provides insights into various aspects of graduates' academic background, employment history, skill assessments, and personal attributes. A detailed description of the variables included in the dataset will be provided to facilitate the analysis and interpretation of the data.

Data Cleaning Steps

1. Handling Missing Values: Identify and handle missing values in the dataset, either by imputation or removal.
2. Outlier Detection and Treatment: Identify outliers in numerical variables and decide whether to remove them or treat them.
3. Data Type Conversion: Convert data types of variables if necessary (e.g., converting date columns to datetime format).
4. Handling Duplicates: Check for and handle any duplicate rows in the dataset.
5. Standardization or Normalization: Standardize or normalize numerical variables if required for analysis.

Data Manipulation Steps

1. Feature Engineering: Create new features from existing ones if needed to extract additional insights from the data.
2. Encoding Categorical Variables: Encode categorical variables into numerical format using techniques like one-hot encoding or label encoding.
3. Data Aggregation: Aggregate data if necessary to analyze at a higher level of granularity.
4. Data Transformation: Transform data using mathematical operations (e.g., logarithmic transformation) to make it more suitable for analysis.

Univariate Analysis Steps

1. Probability Density Functions (PDFs): Plot PDFs for numerical variables to visualize their distributions.
2. Histograms: Plot histograms for numerical variables to understand their frequency distributions.
3. Boxplots: Visualize the spread of numerical variables and identify outliers using boxplots.
4. Countplots: Plot countplots for categorical variables to visualize their frequency distributions.
5. Descriptive Statistics: Compute summary statistics (mean, median, mode, standard deviation, etc.) for numerical variables.

Bivariate Analysis Steps

1. Scatter Plots: Plot scatter plots to visualize relationships between two numerical variables.
2. Heatmaps: Create heatmaps to visualize the correlation matrix between numerical variables.
3. Pair Plots: Plot pair plots to visualize relationships between multiple numerical variables.
4. Bar Plots: Plot bar plots to compare the distribution of numerical variables across different categories of categorical variables.
5. Box Plots: Visualize the distribution of numerical variables across different categories of categorical variables using box plots.

Key Business Questions

- Testing specific claims about salary expectations for fresh computer science engineering graduates:
  Are fresh computer science engineering graduates able to earn salaries within the range of 2.5-3 lakhs, as claimed by the Times of India article dated Jan 18, 2019?
  How do the actual salary distributions for fresh computer science engineering graduates compare to the claimed salary range?
- Investigating the relationship between gender and specialization:
  Is there a significant difference in the choice of specialization between male and female engineering graduates?
  Are certain specializations more popular among one gender compared to the other?
  Does gender influence the distribution of salaries across different specializations?

Conclusion

In conclusion, the exploratory data analysis (EDA) of the Aspiring Minds Employment Outcome 2015 (AMEO) dataset provided valuable insights into the employment outcomes of engineering graduates. Here are the key findings:

Salary Distribution: The distribution of salaries among engineering graduates varied widely, with some graduates earning significantly higher salaries than others. Further analysis revealed factors such as specialization, college tier, and academic performance may influence salary levels.

Gender Disparit: There appeared to be a gender disparity in employment outcomes, with males dominating certain specializations and job roles. Understanding and addressing this gender gap is essential for promoting diversity and inclusivity in the engineering workforce.

Skill Scores: The dataset included standardized scores from various skill areas such as cognitive skills, technical skills, and personality traits. These scores could be further analyzed to understand their impact on employment outcomes and career trajectories.

Relationships Between Variables: Bivariate analysis revealed interesting relationships between different variables. For example, there seemed to be correlations between academic performance (e.g., GPA, percentage scores) and job roles or salary levels.

THANK YOU

INNOMATICS RESEARCH LABS