Analysis of AMCAT Data

(Understanding Employment Outcomes of Engineering Graduates)

Submitted by

Jai Chaudhrani

(IN1240779)



Date: 16-02-2024

**INTRODUCTION**

The dataset used for analysis is the Aspiring Minds Employment Outcome 2015 (AMEO) dataset obtained from Aspiring Minds. It consists of approximately 4000 rows and 40 columns, containing employment outcomes of engineering graduates. The dataset encompasses various dependent variables such as Designation, JobCity, 10percentage, Salary, etc., and comprises both numerical and categorical data.

**DATA MANIPULATION**

1. **Cleaning the Dataset:**
   * Removed duplicates:

Identified and removed duplicate rows from the dataset to ensure each observation is unique.

* + Removed unwanted rows and columns:

Eliminated irrelevant or redundant rows and columns that do not contribute to the analysis or contain redundant information.

1. **Handling Missing Values:**
   * Checked for NaN values:

Conducted a thorough examination of the dataset to identify any missing or NaN values in the data.

* + Imputed or removed NaN values:

Implemented appropriate strategies to handle missing values, such as imputation techniques or removal of rows/columns with excessive missing data.

**DATA VISUALIZATION**

**Univariate Analysis:**

* KDE plot of Educational Percentages:
  + 10th Percentage:

Left-skewed distribution with the majority of employees having around 80%.

* + 12th Percentage:

Normally distributed with the majority falling between 70-80%.

* + College GPA:

Normally distributed with the majority ranging from 65-75%.

* Histograms:
  + Distribution of pass-out years for 12th and Graduation, indicating most employees passed out in 2009 and 2013, respectively.
* Boxplots:
  + English, Logical, and Quantitative Ability columns show many outliers, both high and low extremes.
* Countplot:
  + JobCity: Indicates most employees work in Bangalore, with fewer in Bhopal.

**BIVARIATE ANALYSIS**

1. **Scatterplot on Specialization and Salary Comparing Gender:**
   * This scatter plot visually depicts the relationship between an employee's specialization and their salary, while also considering gender.
   * It provides insights into the distribution of salaries across different specializations and the gender breakdown within each specialization.
2. **Hexbinplot on Salary and College GPA:**
   * The hexbin plot represents the joint distribution of salary and college GPA.
   * It visualizes the density of data points, allowing for the identification of patterns or clusters in the relationship between salary and college GPA.
3. **Categorical and Numerical Columns:**
   * Swarmplot on Degree and College GPA:
     + This swarm plot illustrates the distribution of college GPA among different degree types.
     + It highlights the concentration of data points for each degree type and identifies any potential outliers.
   * Barplot on Salary and Degree:
     + This bar plot displays the average salary for each degree type.
     + It offers insights into the salary distribution across different educational backgrounds.
   * Boxplot on Degree and College GPA:
     + The box plot showcases the distribution of college GPA across different degree types.
     + It highlights any variations in GPA distribution among different degree categories and identifies potential outliers.

**CONCLUSION**

* **Data Understanding:**

The dataset provides insights into the employment outcomes of engineering graduates, including salary, job titles, and locations, along with standardized scores in cognitive, technical, and personality skills.

* **Data Manipulation:**

Initial observations revealed the need for data cleaning to remove duplicates and unwanted rows/columns. NaN values were checked, and the cleaned dataset was used for visualization.

* **Data Visualization:**

Utilized various plots for univariate analysis (PDF, histograms, boxplots, countplots) to understand probability and frequency distributions. Bivariate analysis (scatterplot) was conducted to explore relationships between variables.

* **Observations:**

Identified outliers using boxplots and analyzed the distribution of employees across different job cities.