**Project Report**

**Analysis of AMCAT Data**

**Abstract**

This project presents a thorough exploration of a dataset comprising information about individuals' educational background, career details, and salaries. The analysis is centered around two principal research questions: testing the claim regarding salaries for fresh graduates in specific roles and investigating the relationship between gender and specialization.

1. **Introduction**

In recent years, understanding the dynamics of salaries for fresh graduates in the field of Computer Science Engineering has become crucial. Additionally, exploring gender-based preferences in specialization can provide valuable insights into diversity and career choices.

**2. Dataset Overview**

2.1. Data Source

The dataset was obtained from [Specify the source], consisting of 3998 entries and 39 columns.

2.2. Data Description

The dataset includes various features such as educational qualifications, gender, salary information, and specialization details.

**3. Methodology**

3.1. Data Cleaning and Preprocessing

Checked for missing values and applied appropriate strategies for imputation or removal.

Ensured consistency in column names, addressing any discrepancies.

Verified the integrity of numerical data and handled outliers appropriately.

3.2. Univariate Analysis

3.2.1. Probability Density Functions (PDF) and Histograms

Conducted detailed analyses of the PDFs and histograms for key numerical features, providing insights into the distribution patterns.

3.2.2. Boxplots

Investigated the spread and presence of outliers in numerical columns through boxplots, enabling a robust understanding of the data's statistical characteristics.

3.3. Bivariate Analysis

3.3.1. Scatter Plots and Hexbin Plots

Explored relationships between numerical features, specifically investigating correlations between GPA, test scores, and salary.

3.3.2. Swarm Plots and Boxplots

Examined the interplay between gender and salary, degree, and specialization, shedding light on potential disparities and trends.

3.3.3. Stacked Bar Plots

Delved into the relationships between different categorical variables, such as gender and degree, offering valuable insights into the distribution patterns.

3.4. Multivariate Analysis

3.4.1. Heatmaps and Multivariate Statistics

Utilized heatmaps and multivariate statistical tests to uncover complex relationships between multiple variables simultaneously.

**4. Analysis Results**

4.1. Testing the Claim about Salaries

4.1.1. Challenges Encountered

The analysis faced challenges due to discrepancies in role names specified in the Times of India article and the actual dataset entries.

Subsequent investigation revealed a need for refinement in the filtering criteria.

4.1.2. Recommendations

Further refinement of data filtering criteria to align with the exact role names specified in the article.

Exploration of additional factors influencing salary, such as experience, location, and industry sector.

4.2. Relationship between Gender and Specialization

4.2.1. Inconclusive Findings

While initial analysis explored the distribution of specializations across genders, the statistical tests did not provide conclusive evidence of a significant relationship.

4.2.2. Future Directions

Further investigation into the nuances of gender preferences in specialization, considering additional demographic factors.

Collection of more granular data to enable a more in-depth analysis of gender-related trends.

4.3. Multidimensional Insights

4.3.1. Integrating Additional Dimensions

Introduced a multidimensional analysis approach to better understand interactions between multiple variables simultaneously.

4.3.2. Advanced Visualizations

Employed advanced visualization techniques, such as 3D plots or interactive visualizations, for a more nuanced representation of complex relationships.

**5. Conclusion**

This comprehensive analysis provides valuable insights into the dataset, uncovering patterns and challenges. The findings contribute to the broader understanding of factors influencing salaries and gender-specialization dynamics in the field of Computer Science Engineering.

**6. Recommendations and Future Work**

* Continued refinement of data filtering criteria and exploration of additional factors influencing salaries.
* In-depth exploration of gender-specialization dynamics with a focus on intersectionality.
* Consideration of external factors such as industry trends, economic conditions, and regional variations.

**submitted by :**

**Name:**  Amit Jadhav

**Intern ID:** IN1240852