INNOMATICS®
RESEARCH LABS

INNOVATION. AUTOMATION. ANALYTICS

PROJECT ON

AMCAT Data Analysis

Manoj Kumar Sahoo

About me

1. Introduction:

- Name: Manoj Kumar Sahoo
- Academic Status: MCA
- College:Trident Academy Of Technology,BBSR

2. Why Data Science:

- Passionate about leveraging technology to solve real-world problems
- Keen interest in exploring the vast field of Data Science
- Excited about the potential of data-driven decision-making

3. Work Experience:

- No formal work experience yet, currently focused on academic pursuits

4. LinkedIn
5. GitHub
6. Conclusion:

- Aspiring to build a career in Data Science
- Committed to continuous learning and growth

INNOMATICS RESEARCH LABS

Exploratory Data Analysis

a. Data Cleaning Steps:

- Identify columns with missing values.
- Decide on the appropriate method to handle missing values (imputation, removal, or other techniques).
- Check for and remove any duplicate rows in the dataset.
- Identify outliers in numerical columns.
- Decide on the approach to handle outliers (removal, transformation, or other techniques).
- Assess whether numerical columns need standardization or normalization.
- Ensure that the data types of columns are appropriate.
- Convert data types if needed (e.g., converting string dates to datetime objects).
- Check for any inconsistencies in categorical data (e.g., different spellings of the same category).
- Standardize categories if necessary.

Exploratory Data Analysis (Cont.)

b. Data Manipulation Steps:

- Create newfeatures if needed based on existing columns.
- Convert categorical variables into numerical format using techniques like one-hot encoding or label
- encoding.
- Calculate or derive new variables if they add value to the analysis.
- Aggregate data if needed for a higher-level analysis.
- If building predictive models, split the dataset into training and testing sets.
- Perform a final check to ensure the dataset is in the desired format for analysis or modeling. Validate that all steps have been appropriately applied.

Exploratory Data Analysis (Cont.)

c. Univariate Analysis Steps:

- For numerical columns: Generate PDFs (Probability Density Functions), histograms, boxplots to find outliers, and understand the probability and frequency distribution.
- For categorical columns: Create countplotsto understand the frequency distribution.

d. Bivariate Analysis Steps:

- For relationships between numerical columns, we'll use scatter plots, hexbinplots, and pair plots to discover patterns.
- To identify patterns between categorical and numerical columns, we'll use swarm plots, box plots, and bar plots.
- For relationships between categorical columns, we'll use stacked bar plots.

Key Business Questions

1. Test the claim that after doing Computer Science Engineering, taking up jobs as a Programming Analyst, Software Engineer, Hardware Engineer, and Associate Engineer can earn up to 2.5-3 lakhs as a fresh graduate.

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

3. Does graduating from a Tier 1 college significantly affect the starting salary compared to graduating from a Tier 2 college? This analysis could highlight the importance of college prestige on career outcomes.

4. Is there a correlation between academic performance (measured by GPA and percentages in 10th and 12th grades) and professional success (measured by salary)? This could challenge or confirm the notion that higher academic achievements directly translate to better career opportunities.

5. Are certain job roles within the technology and engineering sectors dominated by a particular gender? This question aims to identify potential gender biases or preferences in specific technology and engineering job roles.

6. Does the job location (JobCity) influence the salary and job roles offered to graduates? This analysis could uncover regional disparities in career opportunities and salaries within the tech and engineering sectors.

Conclusions/Key Findings

1. The analysis of numerical columns revealed the presence of outliers in salary and academic scores, as well as variations in GPA. The categorical analysis highlighted the predominance of male, engineering graduates from Tier 2 colleges. These insights can be useful for further analysis, including predictive modelingor understanding the factors influencing job outcomes and salaries.

2.

- Numerical-Numerical Relationships: We observed that academic performance in the 10th and 12th grades tends to be positively correlated, as is the correlation between 12th-grade performance and college GPA. However, higher academic scores do not necessarily correlate with higher salaries.
- Categorical-Numerical Relationships: The analysis highlighted a gender disparity in salary distribution, with males generally having a broader salary range and more representation in higher-paying jobs.
- Categorical-Categorical Relationships: The gender distribution across college tiers further emphasized the gender imbalance within the dataset, particularly in Tier 2 colleges.

Conclusions/Key Findings (Cont.)

3.

- For Computer Science Engineering graduates taking up jobs as Programming Analyst, Software Engineer, Hardware Engineer, and Associate Engineer, the salary analysis reveals:
- The mean salary is approximately ₹332,711, which is above the 2.5-3 lakhs range mentioned in the claim.
- The median salary (50% percentile) is ₹315,000, also slightly above the upper limit of the claim.
- Minimum and maximum salaries range from ₹85,000 to ₹1,000,000, indicating a wide variance in compensation.
- Out of the filtered group, 25 individuals (approximately 17.6%) fall within the 2.5-3 lakhs salary range.
-

4.

- High Participation in Certain Fields: Fields like 'computer science & engineering', 'electronics and communication engineering', and 'information technology' have high participation from both genders, with males generally outnumbering females.
- Gender-Specific Concentrations: Certain specializations such as 'biotechnology' and 'electronics & telecommunications' have relatively higher female participation compared to other engineering specializations.
- Fields with Low Female Representation: Specializations like 'mechanical engineering' and 'civil engineering' have a significantly higher number of males, with females being less represented.

Conclusions/Key Findings (Cont.)

5.

- Prestige vs. Performance: College tier may have a significant impact on initial career outcomes, but individual performance and skills (e.g., programming or analytical abilities) could play a more critical role in long-term career success.
- Diverse Career Paths: Graduates from technology and engineering fields have a wide range of career paths available, not limited to traditional roles but expanding into interdisciplinary fields that combine technical expertise with business, design,or other areas.
- Gender Disparities: While gender disparities in certain specializations and job roles are evident, there's a growing trend of gender diversification in traditionally male-dominated fields, suggesting ongoing changes in societal and professional landscapes.
- Regional Opportunities: Job location can significantly affect salary and career growth opportunities, highlighting the importance of geographical mobility and flexibility for career advancement in the tech and engineering sectors.
- Evolving Industry Needs: The demand for certain specializations changes over time, reflecting industry trends, technological advancements, and global economic factors. Continuous learning and adaptation are key to maintaining relevance and success in the fast-paced tech and engineering industries.

THANK YOU

INNOMATICS RESEARCH LABS