INNOMATICS®
RESEARCH LABS

INNOVATION. AUTO**MAT**ION. ANALY**TICS**

PROJECT ON

EDA Project - AMCAT Data Analysis

About me

- **B.E. in Computer Engineering**
- **I wanna learn Data Science to equips the skills to make informed decisions, tackle complex problems, and stay competitive in a data-driven world.**
- **Github: https://github.com/Azay961**
- **LinkedIn:https://www.linkedin.com/in/chaudharyajay/**

Objective Of The Project

- Analyze the employment outcomes of engineering graduates based on Salary, Job Titles and Job Locations along with the standardized scores from three different areas – cognitive skills, technical skills and personality skills.

Summary Of The 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.

A. Data Cleaning Steps

- df.describe(): for describing the data(finding mean, std, quartiles, min, max, count)
- df.info(): for viewing total null values in each columns and data types of column values
- df.head(): Viewing top 5 data points for ensuring the data types of columns
- df.drop(): For dropping unwanted columns
- Fixing the columns name and typecasting it to appropriate data types

INNOMATICS
RESEARCH LABS

B. Data manipulation steps

- Identifying Null values in dataset and removing them
- Fuzzywuzzy: for correcting spellings in categorical columns

C. Univariate Analysis

- For numerical columns mean, median, std, min, max, skew, kurt were found using aggregate function for non visual analysis. And for visual analysis plots like HIstogram, KDE, Box, QQ are plotted. For statistical test Shapiro wilk and Kolmogorov-Smirnov test was performed.
- For categorical columns count, nunique, unique, value_counts were used for non visual analysis and for visual analysis count plot was used

D. Bivariate Analysis

- Numerical vs Numerical: Correlation, heatmap
- Numerical vs Categorical : Groupby, boxplot
- Categorical vs categorical : crosstab, stacked barplot

Key Business Question

1. Times of India article dated Jan 18, 2019 states that "After doing your Computer Science Engineering if you take up jobs as a Programming Analyst, Software Engineer, Hardware Engineer and Associate Engineer you can earn up to 2.5-3 lakhs as a fresh graduate."

Conclusion:

The boxplot aside proves the statement given by Times of India that fresh graduate can lakhs if you take up jobs as a Programming Analyst, Software Engineer, Hardware Engineer and Associate Engineer after doing your Computer Science Engineering.
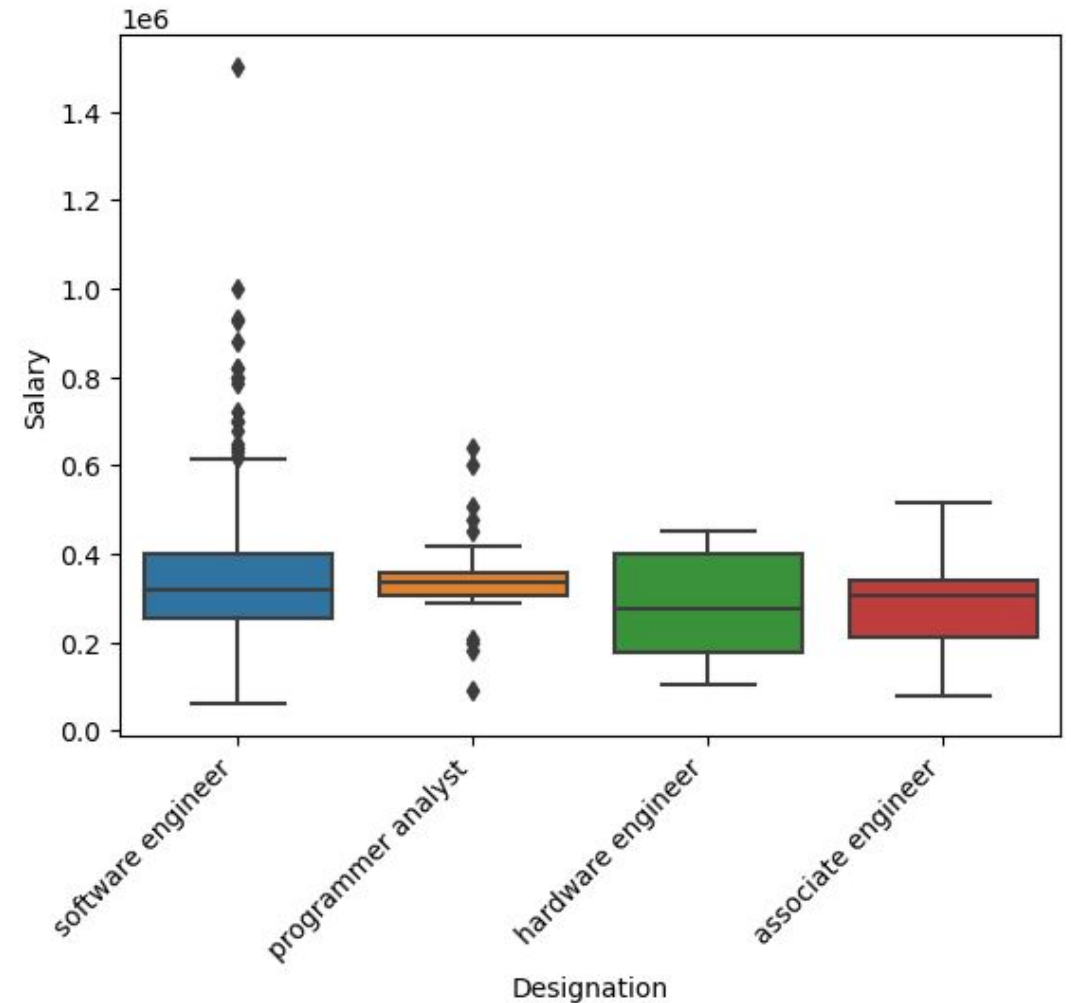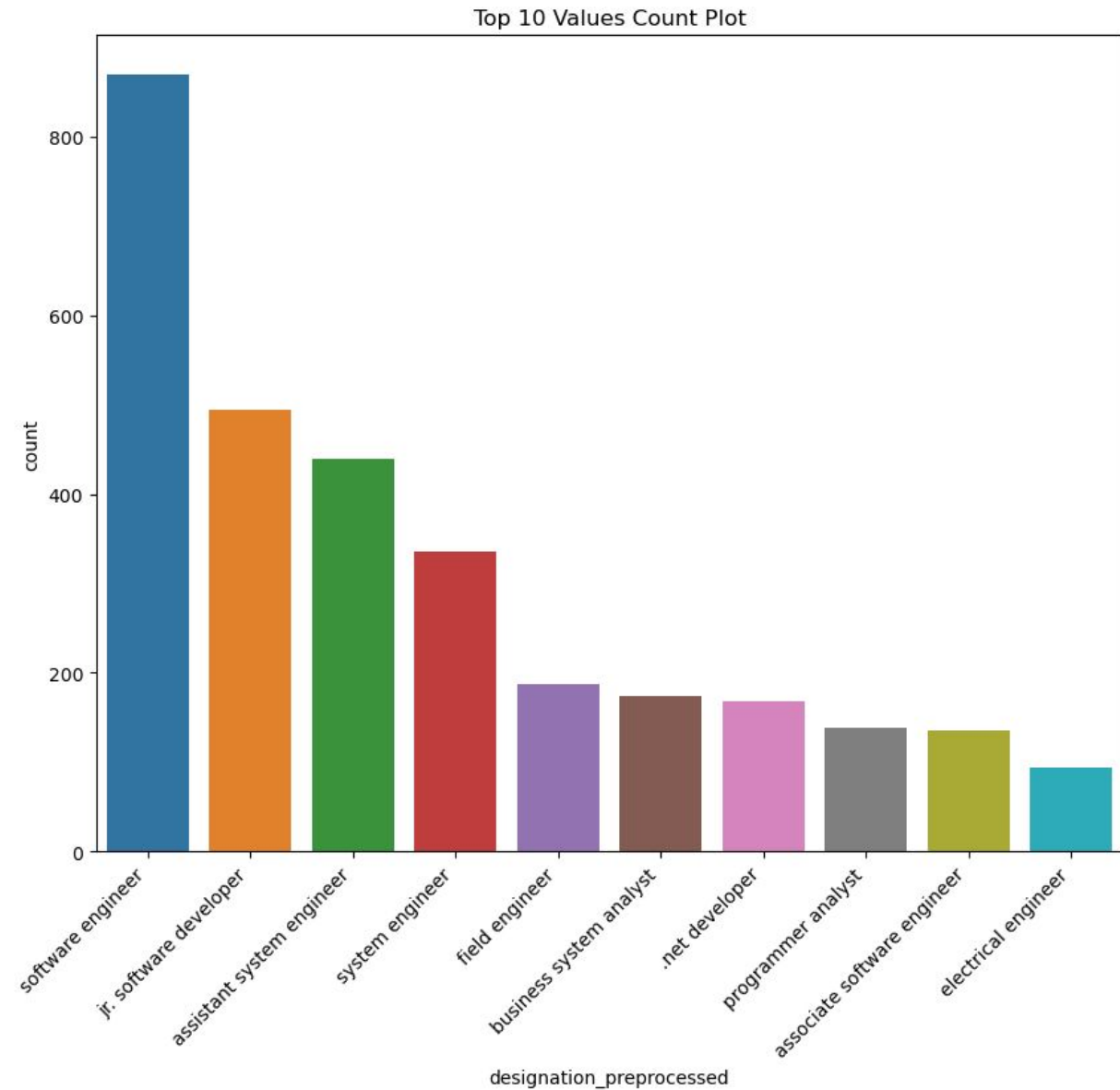
Analysis and Findings

1. Preferences of Gender on Specialization

This stacked bar plot shows that both gender prefers electronics and communication engineering as a specialization.
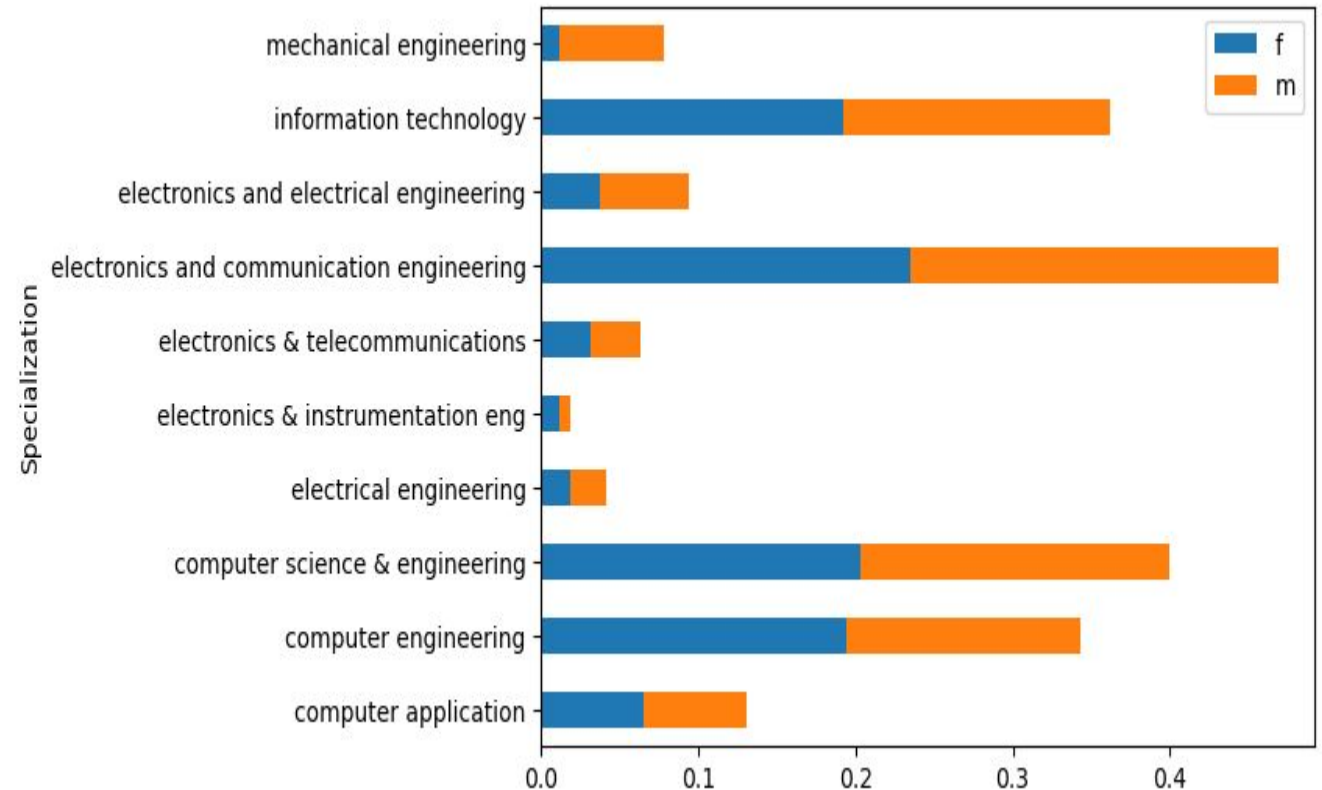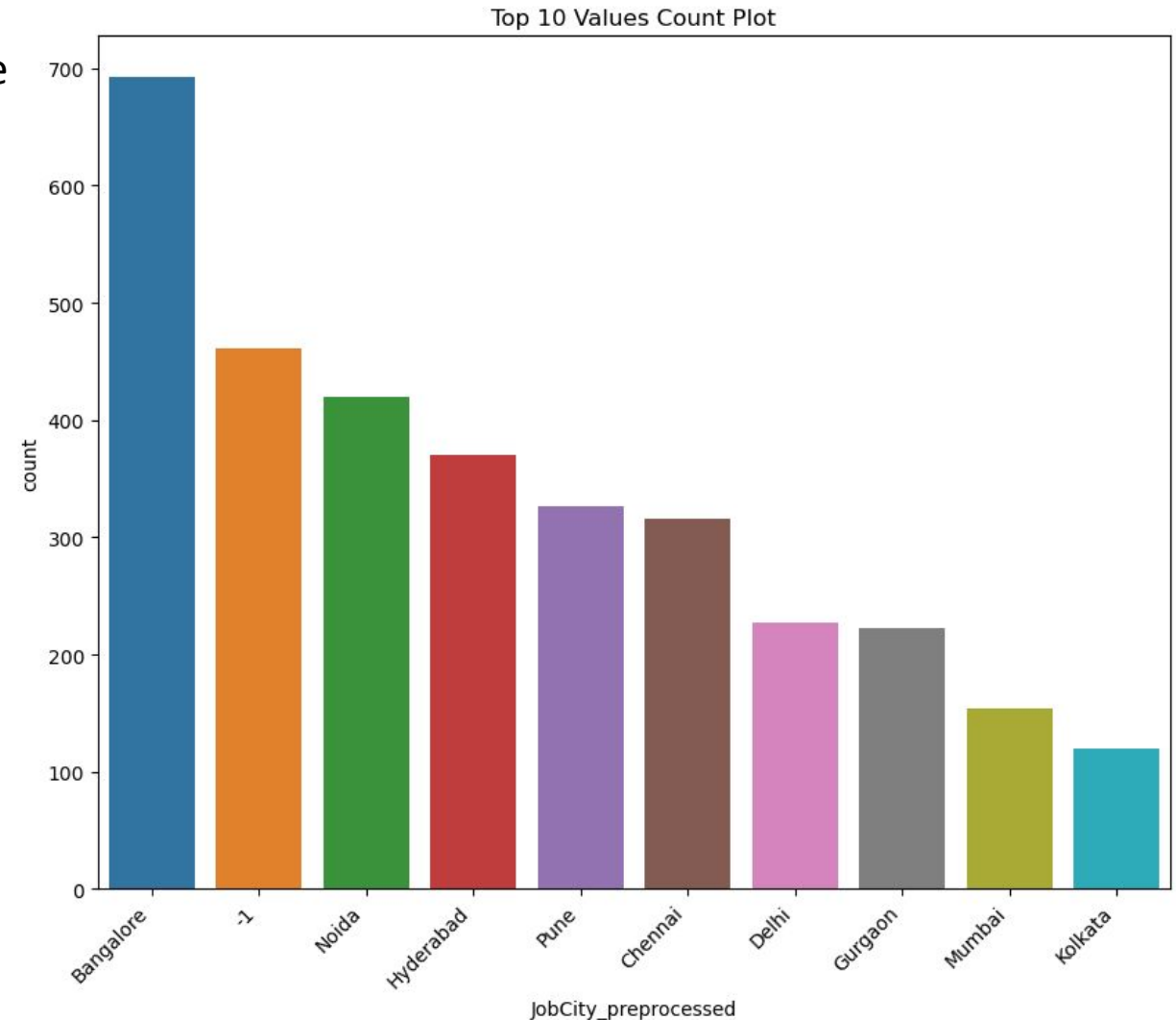
2. Top 10 job roles:

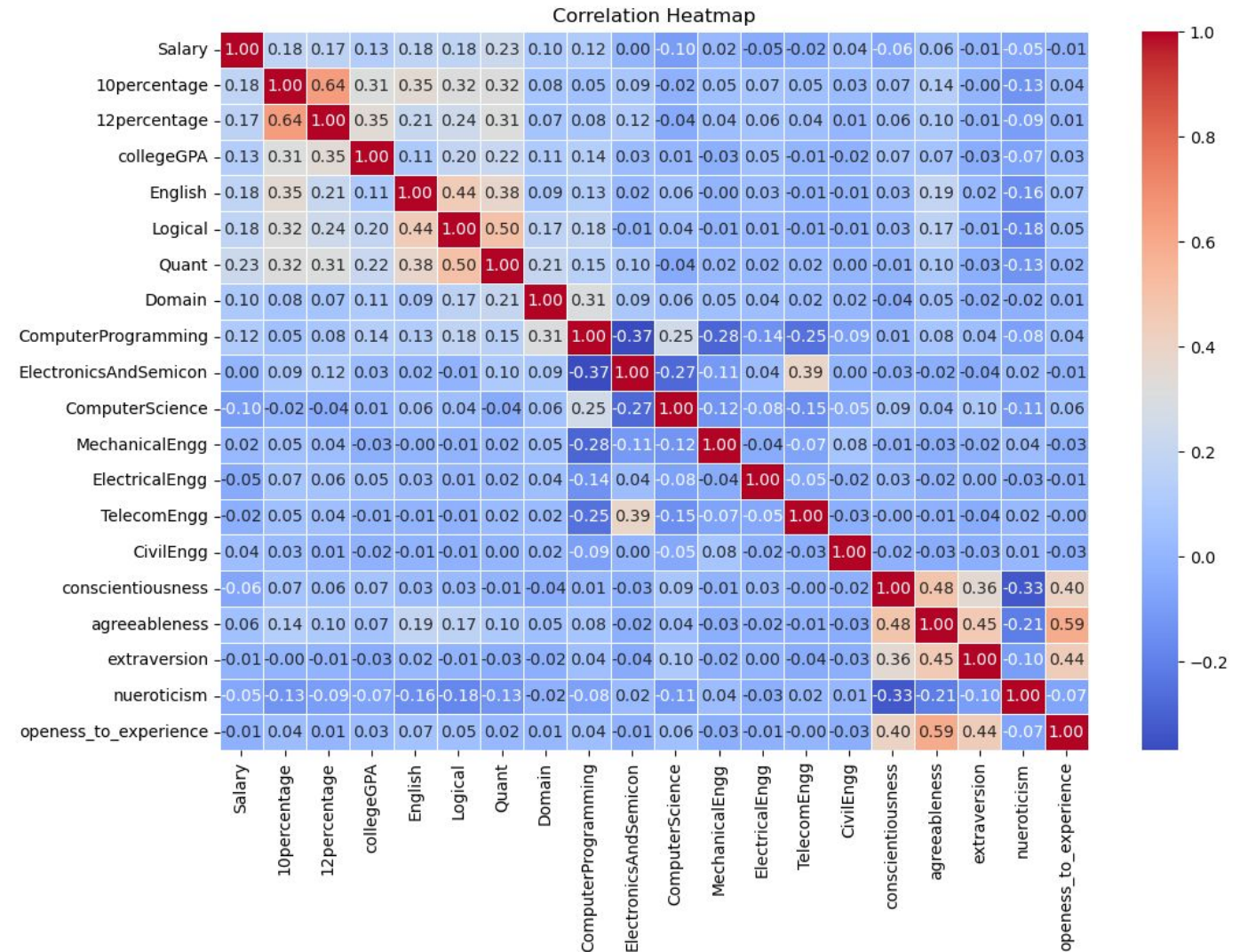The count plot shows that software engineer was the job role most of the candidates were doing.

3. Top 10 job city

According to count plot most of the candidate were doing job at Banglore. Here -1 represents the candidates who has not mentioned their job city which is second highest.

Top 10 Values Count Plot

INNOMATICS
RESEARCH LABS

4. Salary package with grades scored

According to heat map there is no strong relationship of salary with marks scored at different level.

Correlation Heatmap