Report on Comprehensive Analysis of Candidate Attributes in the Electoral Dataset

Student Name: Pranay Reddy Bandharapu

Student number: 23028438

Git Hub: https://github.com/Pranayr5/pranay-23028438

Introduction

This report presents an in-depth analysis of the dataset candidates_with_phase.csv, which contains detailed information about candidates, including their phases, constituencies, ages, genders, parties, criminal cases, education levels, and assets. This dataset is essential for understanding various attributes and patterns among political candidates across different phases and constituencies.

2. Descriptive Statistics

The summary statistics for the numeric columns in the dataset are as follows:

Phase: Ranges from 1 to 8, with an average value of approximately 3.43.

Constituency No: Ranges from 1 to 72, with a mean of about 36.51.

Age: Candidates' ages range from 25 to 93, with an average age of around 52.73.

Criminal Cases: The number of criminal cases ranges from 0 to 22, with a mean of 0.83.

Assets: The declared assets range from 0 to 13.25 billion, with an average of approximately 116.4 million.

3. Visual Analysis: 3.1 Distribution of Phase

The bar chart below illustrates the distribution of candidates across different phases:
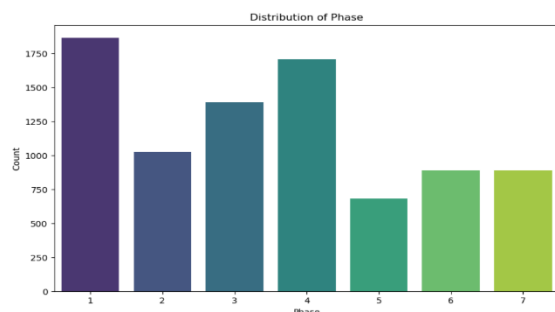
This visualization shows the count of candidates in each phase, highlighting which phases have more or fewer candidates. The distribution helps in understanding how candidates are spread across different phases. **3.2 Scatter Plot of 'Constituency No' vs 'Age' with 'Phase' as**
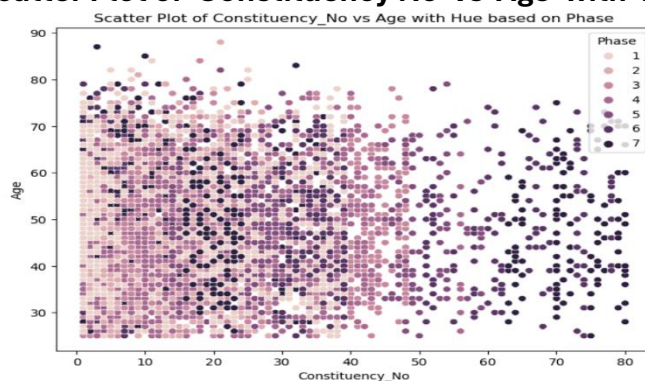

The scatter plot below depicts the relationship between constituency number and candidate age, coloured by phase:

This plot provides insights into the age distribution of candidates across various constituencies and phases. It helps identify any patterns or anomalies in candidate ages relative to their constituencies and phases.

3.3 Correlation Matrix Heatmap

The heatmap below visualizes the correlation matrix:
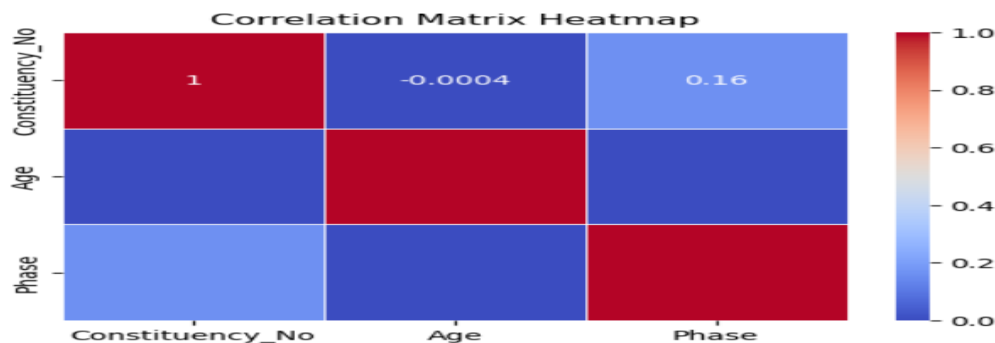
This heatmap displays the correlations between numeric attributes. It helps in identifying any strong or weak correlations visually, making it easier to interpret the relationships among different attributes.

Summary and Insights

Age Distribution: The average age of candidates is 52.73 years, with most candidates falling within the middle-aged category. This suggests that political participation is prevalent among experienced individuals.**Criminal Cases:** On average, candidates have less than one criminal case against them. However, some candidates have as many as 22 cases. This variance indicates a significant disparity in the legal backgrounds of candidates. **Correlations:** The weak correlations between numeric attributes suggest that there are no strong linear relationships among them. This indicates that attributes such as age, constituency number, and criminal cases are independent of each other in this dataset. **Gender and Education:** Additional analysis can be conducted to explore the distribution of gender and education levels among candidates. This can provide further insights into the demographic and educational backgrounds of political candidates. **Party Affiliation:** Examining the distribution of candidates across different political parties can reveal trends and patterns in party representation. This analysis can help understand which parties have more candidates with criminal cases or significant assets. **State-wise Analysis:** Analysing the data at the state level can uncover regional patterns and differences in candidate attributes. This can help identify states with higher or lower average ages, more criminal cases, or greater wealth disparities among candidates.

Conclusion

This report provides a comprehensive analysis of the candidates_with_phase.csv dataset. The dataset reveals significant insights into the demographics, legal backgrounds, and financial statuses of political candidates. Visualizations and correlation analyses help in understanding the relationships and distributions among various attributes. Further analysis can expand on these findings to uncover deeper patterns and trends in the data.