Bias Detection Report

Bias Detection Report: Agency Distribution in COMPAS Dataset

This report analyzes the distribution of assessments across agencies in the COMPAS dataset to determine if there is a dominant agency or if the distribution is fairly even.

Dataset Overview

The analysis was conducted on the COMPAS dataset, focusing on the 'Agency_Text' column, which represents the agencies performing assessments.

Data Cleaning

The 'Agency_Text' column was cleaned to remove any missing or invalid values, resulting in a cleaned dataset used for further analysis.
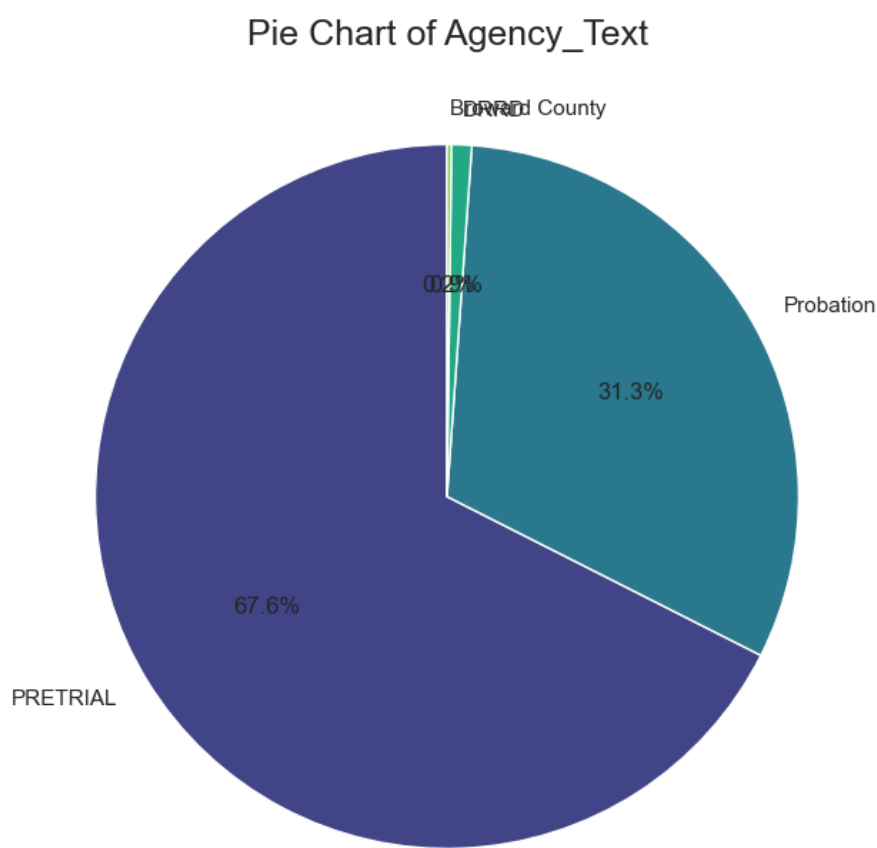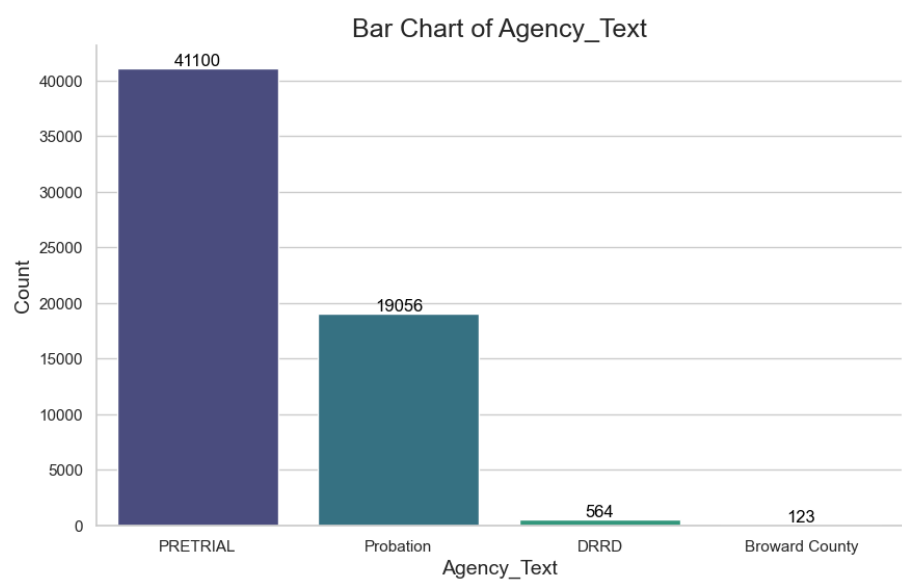
Bias Detection Methods

The following methods were used to assess the distribution bias:

- **Shannon Entropy and Balance Metric:** Balance = 0.4938 (Level 5: Extreme Bias)

- **Gini Index:** Adjusted Gini Index = 0.5940 (Level 4: Significant Bias)

- **Max/Min Ratio:** Ratio = 334.1463 (Level 5: Extreme Bias)

These metrics indicate a significant to extreme bias, suggesting a few agencies dominate the

assessment process.

Bar Chart of Agency_Text

Pie Chart of Agency_Text

Visual Analysis

The bar chart and pie chart illustrate the uneven distribution of assessments across agencies, visually supporting the quantitative findings of bias.

Conclusion

The analysis reveals a significant bias in the distribution of assessments, with certain agencies disproportionately conducting assessments. These findings indicate a need for further investigation or corrective measures to ensure a more balanced distribution.

Bias Type: Distribution

Relevant Feature: Agency_Text

Bias Levels:

- Shannon Entropy and Balance Metric: Level 5

- Gini Index: Level 4

- Max/Min Ratio: Level 5

Overall, the bias is categorized as extreme due to the dominance of certain agencies in assessment distribution.