Bias Detection Report

Bias Detection Report: Native Country Distribution in the Adult Dataset

Introduction: This report analyzes the distribution bias in the 'native-country' feature of the Adult dataset to determine if there are any notable patterns or imbalances in geographic representation.

Methodology: Four different methods were used to assess distribution bias: Max/Min Ratio, Shannon Entropy & Balance, Gini Coefficient, and Normalized Entropy. These methods provide a comprehensive view of the distribution's diversity and inequality.

Results:

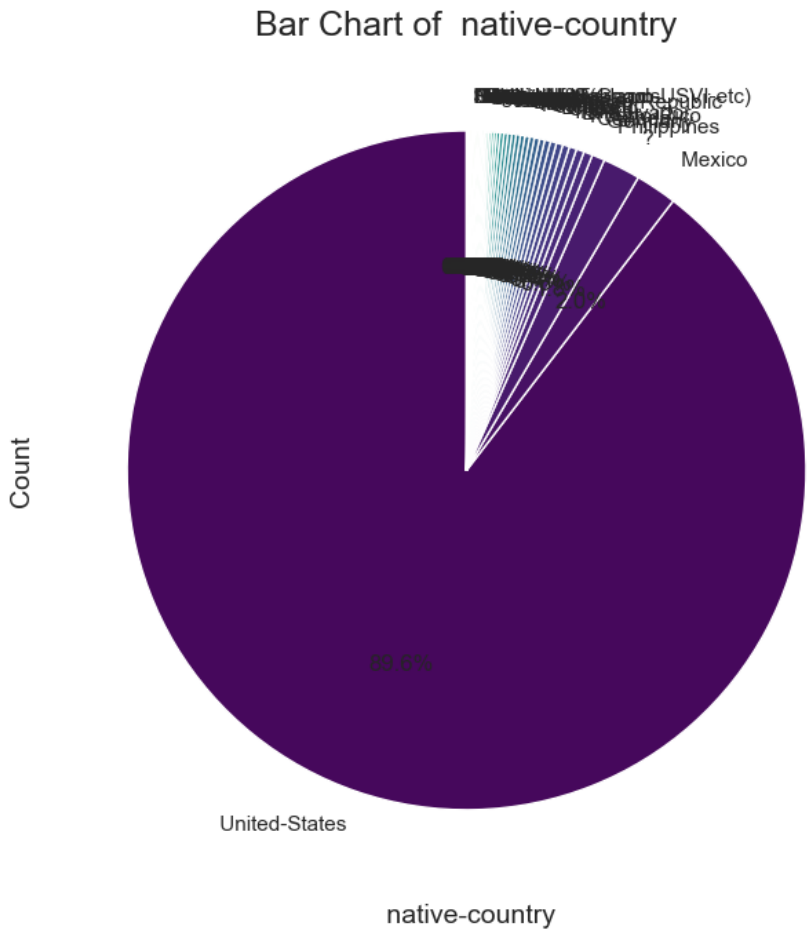
1. Max/Min Ratio: The extreme bias level indicates a highly uneven distribution of the 'native-country' attribute, with a ratio of 29170.0.

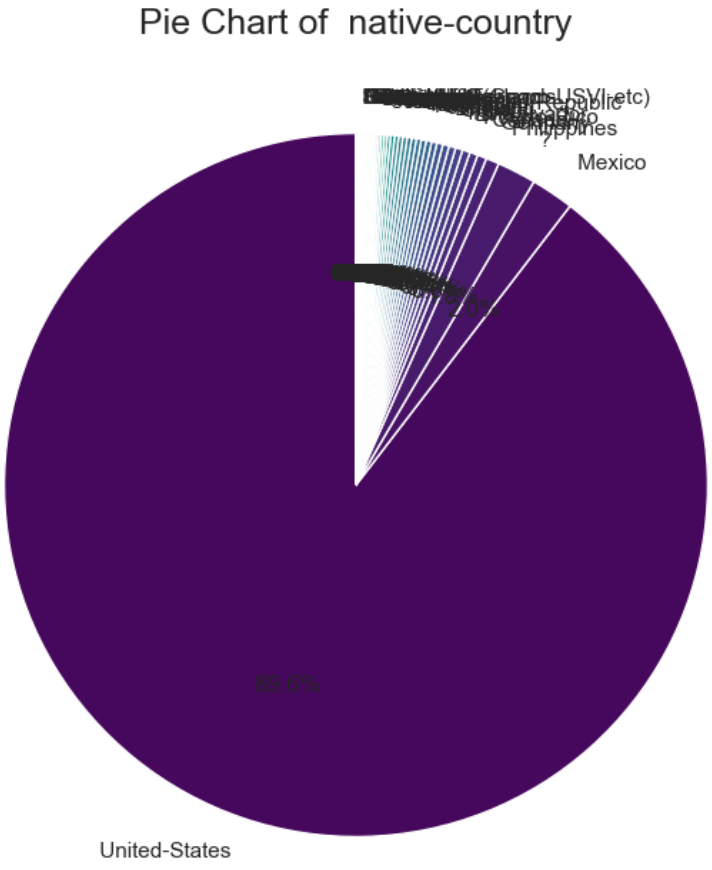
2. Shannon Entropy & Balance: Both metrics suggest an extreme bias level, with a Shannon Entropy of 0.9438 and a Balance metric of 0.1750.

3. Gini Coefficient: The significant bias level is represented by a Corrected Gini Index of 0.1986 and an Adjusted Gini Index of 0.2034.

4. Normalized Entropy: Consistent with other measures, it indicates an extreme bias level with a Normalized Entropy of 0.1750.

Conclusion: The analysis reveals a strong skew in the 'native-country' distribution, highlighting potential biases in geographic representation. The dataset shows extreme bias levels in most metrics, suggesting a need for careful consideration in applications relying on geographic diversity.


Bar Chart of native-country
Count
Mexico
89.6%
United-States
native-country


Pie Chart of native-country
Mexico
89.6%
United-States