Income Prediction Analysis Report

# 1. Introduction

This report encompasses an analysis performed on the 'adult' dataset, aiming to explore the data and develop a predictive model determining whether an individual earns above 50K a year. The dataset contains demographic and employment-related variables for several individuals.

# 2. Data Exploration

The initial exploration involved understanding the dataset's structure, evaluating basic statistics, and identifying potential missing values. Advanced exploration provided insights into variable distributions and their relationships with income levels.

# 3. Modeling

A logistic regression model was constructed to predict the binary outcome of whether an individual earns above 50K annually. The model was subjected to evaluation metrics to ensure robustness and accuracy.

## 3.1 Model Metrics

Accuracy: 85.69%  
Gini Coefficient: 82.18%  
Somers' D: 41.09%  
AUC: 91.09%

# 4. Conclusion

The model exhibited satisfactory performance, meeting the predefined criteria and providing valuable insights into the factors influencing income levels. Further analysis and model tuning can be explored to enhance predictive capabilities.

# 5. Appendix

The Python code utilized for the analysis is provided separately, adhering to functional programming principles and PEP-8 standards.