

**PRESENTED BY :**  
**HARRIET NGOMO**

# **TAXPAYER RISK CLASSIFICATION**

**FAIRTAX ANALYTICS  
COMPANY**

## OVERVIEW



### **Project aim:**

Classify taxpayers as high or low risk for compliance

### **Goal:**

Help tax authorities prioritize audits, improve revenue recovery and reduce costs



# BUSINESS & STAKEHOLDERS

Stakeholders :

- Tax Authority
- Audit Teams
- Policy Makers
- Taxpayers
- DataScientists

**We evaluated how the model  
benefits each group  
(focused audits,fair  
compliance)**



# DATASET

## Description of the dataset:

1000 records , 13 columns

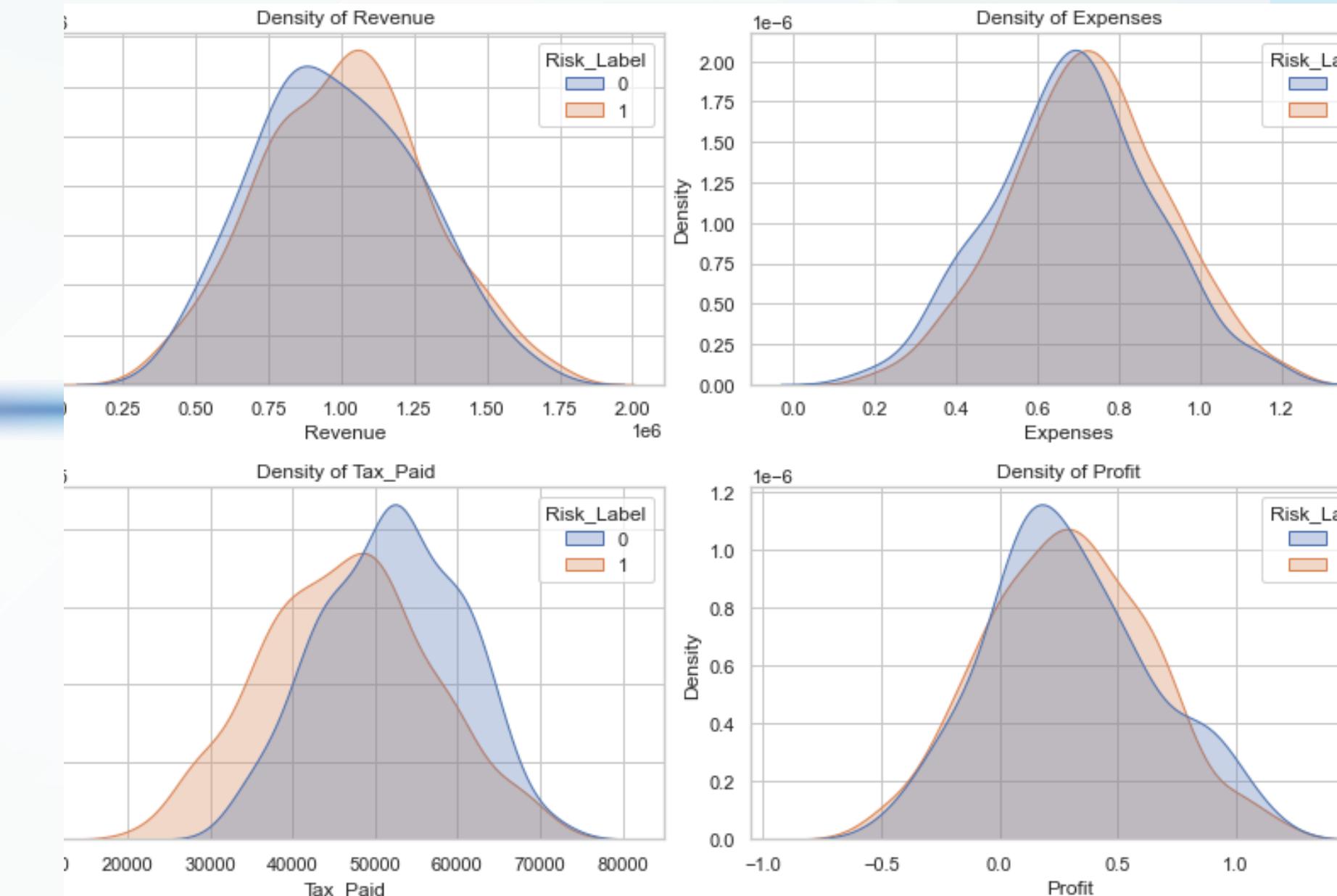
## Key features:

Financial data : revenue , tax paid

Behavioral data : Late Fillings

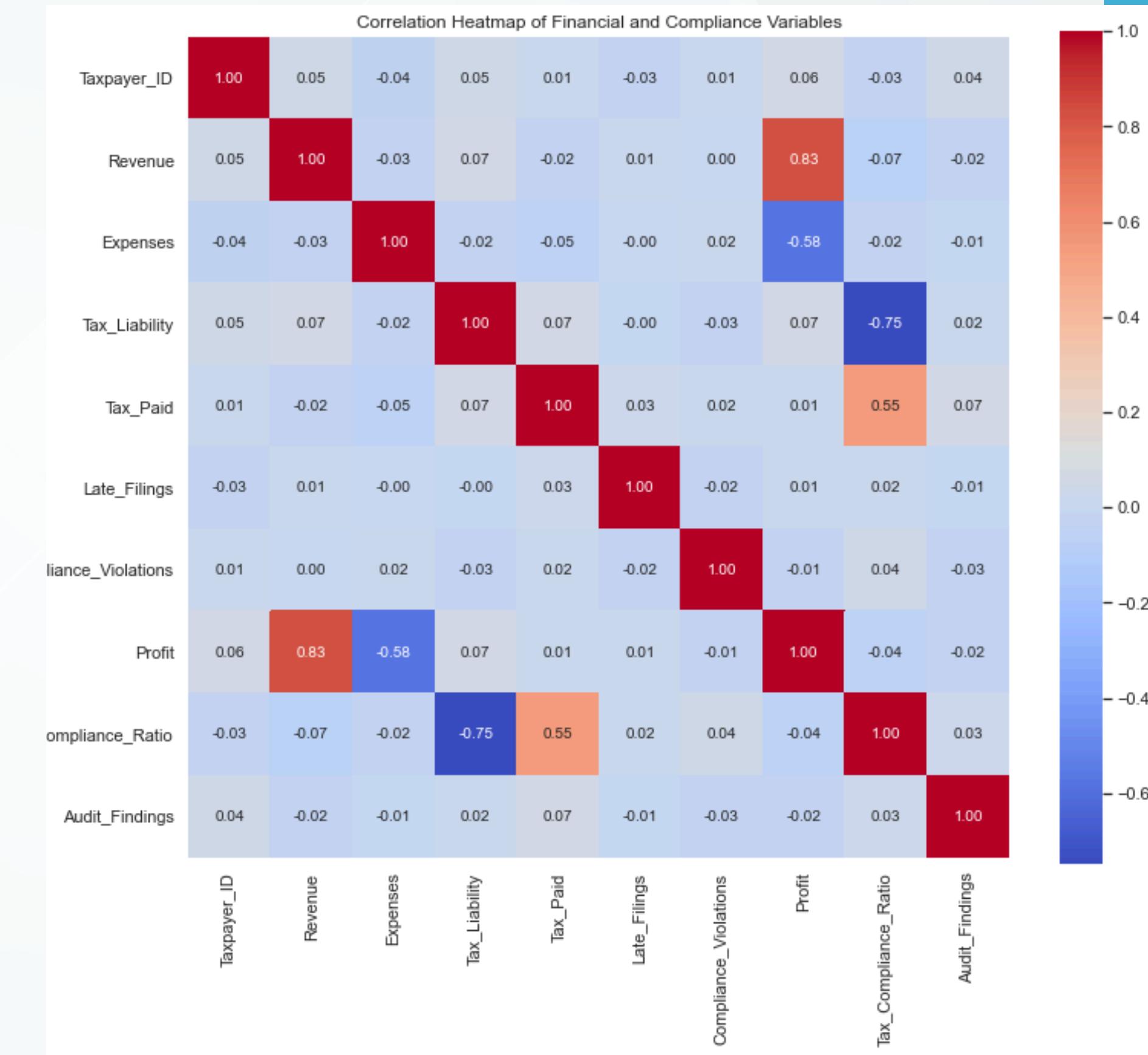
**Target Variable:** Risk Label -High/Low

Data Quality checks : No missing or  
duplicate values



# EXPLORATORY DATA ANALYSIS

**Visualizations** of key features  
& their relation to risk  
**Insights** : Revenue & Profit  
correlate strongly with high  
risk taxpayers



## MODELLING APPROACH

### Models

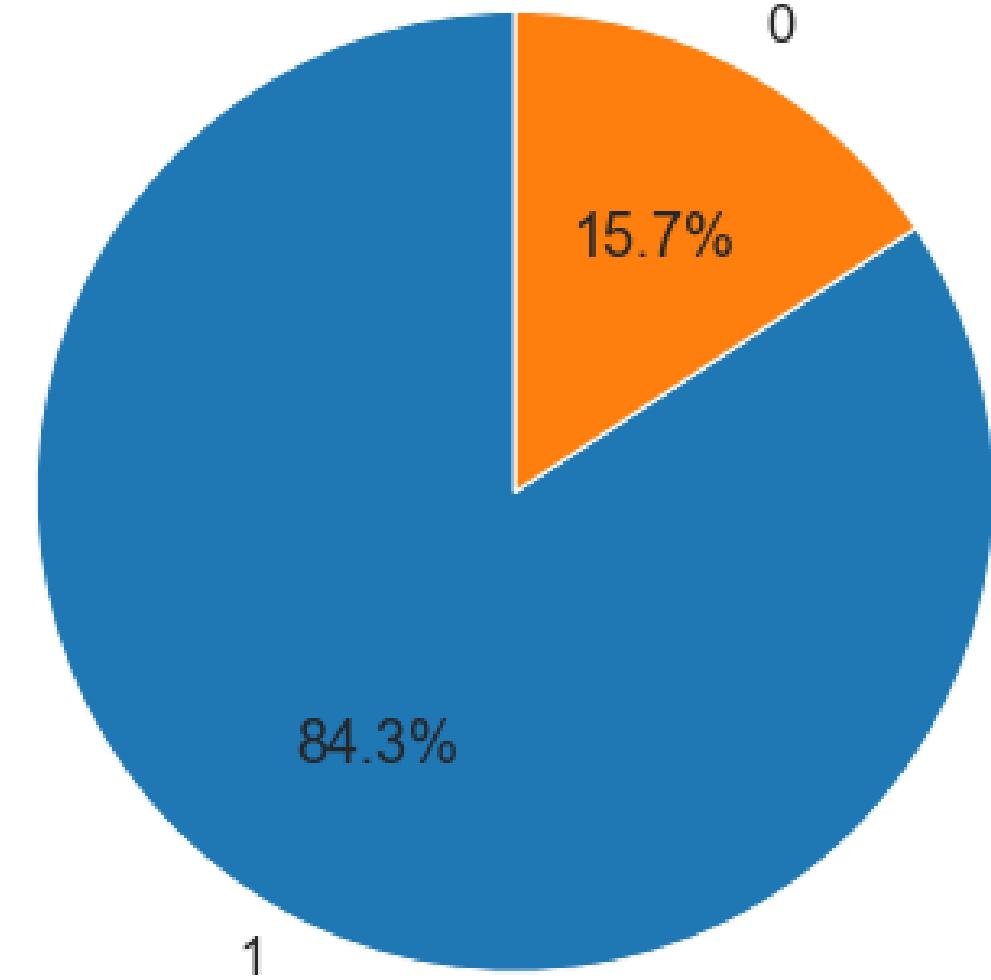
Logistic Regression

Decision Trees(untuned&tuned)

Tuned Random Forest

Handled Class Imbalance with SMOTE

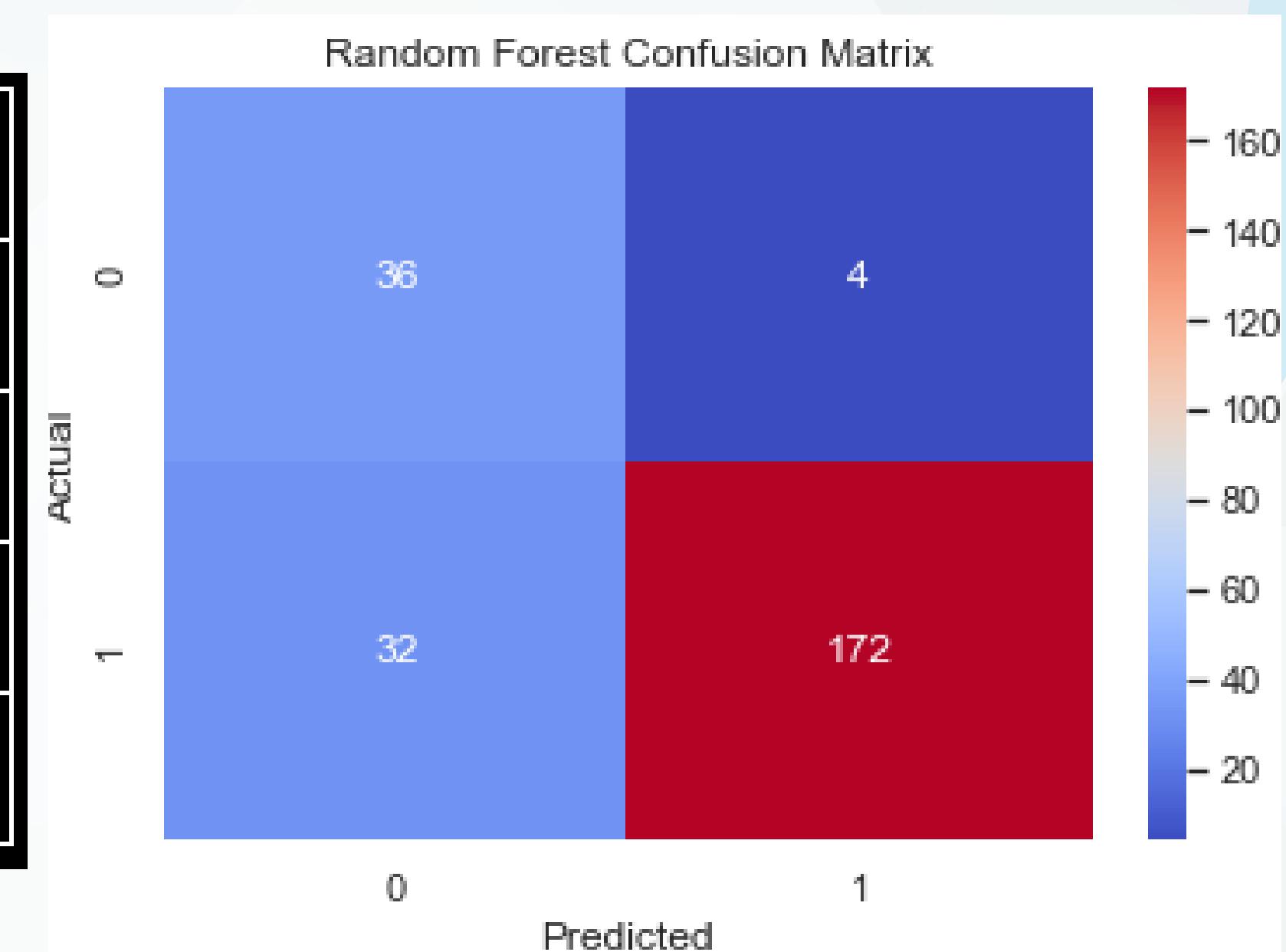
Distribution of Taxpayer Risk Levels



# MODEL PERFORMANCE COMPARISON

Model	ROC AUC
Logistic Regression	0.87
Untuned Decision Tree	0.82
GridCV-Tuned Decision Tree	0.90
Tuned Random Forest	0.94

Tuned Random Forest achieved perfect ROC AUC of 0.94



# EVALUATION METRICS & BUSINESS IMPACT

- For the taxpayer risk classification task, we chose **ROC AUC** as the primary evaluation metric because:
- It measures the model's ability to **distinguish** between **high risk and low risk taxpayers** across all classification thresholds
- **Tuned Random Forest** achieved a perfect score of 0.94 indicating flawless ability to distinguish high risk from low risk taxpayers .This supports **audit prioritization & Fair audits**



# CONCLUSIONS & RECOMMENDATIONS

## Recommendations:

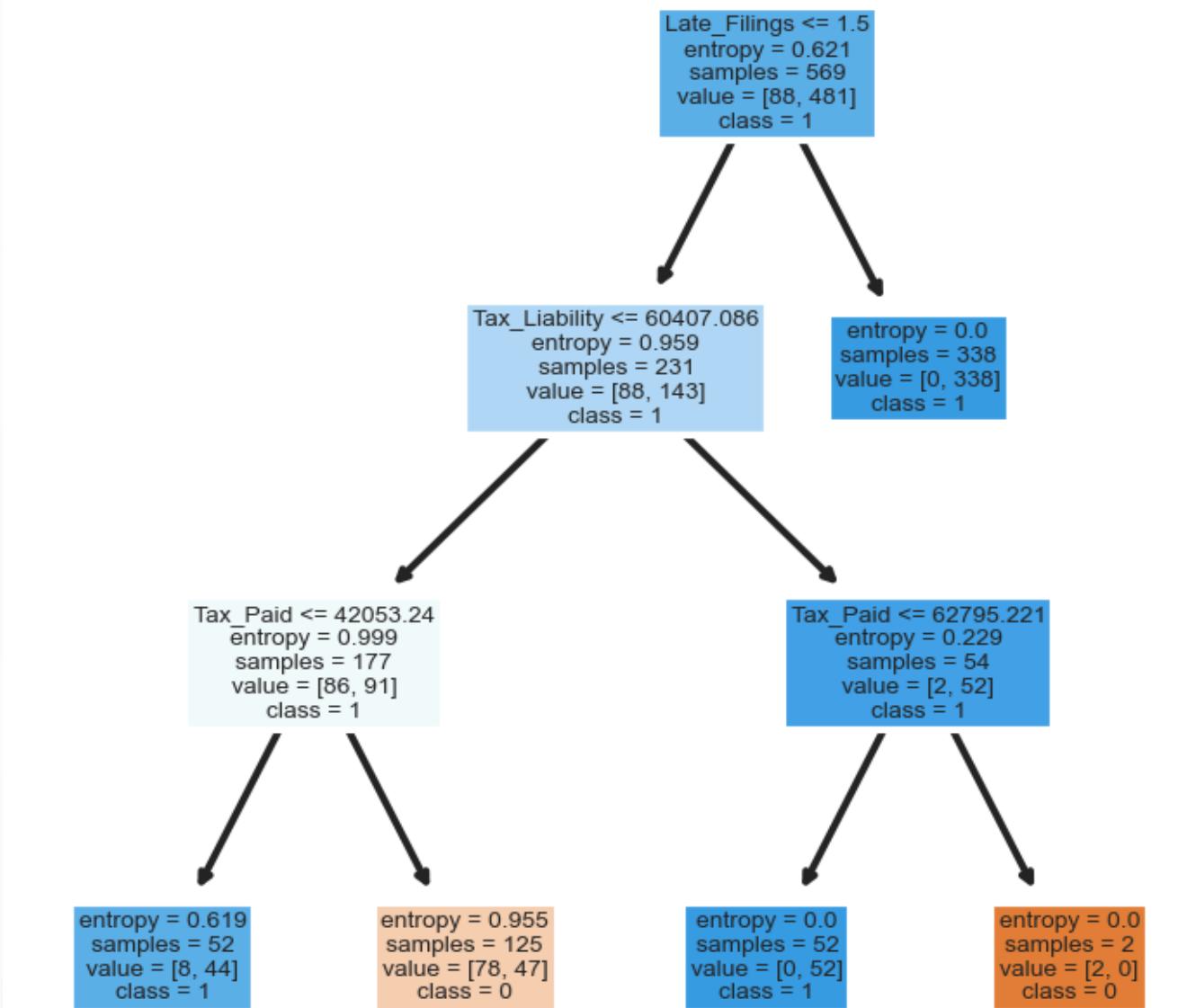
Tuned Random Forest is the best Model for deployment

## Next Steps:

Deploy model, build decision support dashboards , update data regularly , monitor for bias

## Future Exploration:

Ensemble methods like Gradient Boosting



PRESENTED BY :

HARRIET NGOMO

**THANK  
YOU**

**Q AND A SESSION**

**FAIRTAX ANALYTICS  
COMPANY**