**Title: Credit Risk Modeling Analysis Report**

**Section 1: Overview** This report presents a comprehensive credit risk modeling workflow using a real-world dataset. The analysis includes feature distribution, information value (IV), model comparison, feature interpretability, and risk segmentation. Two models, Logistic Regression and XGBoost, are evaluated to predict loan default probabilities.

**Section 2: Categorical Feature Analysis**

*Employment Type Distribution:*

* Self-employed individuals are more represented than salaried ones.
* This suggests that income stability might vary, potentially impacting default probability.

*Information Value (IV) Summary:*

* perform\_cns.score.description has the highest predictive power (IV > 0.05), indicating strong ability to separate defaulters from non-defaulters.
* employment.type has moderate IV, useful in modeling.
* passport\_flag and pan\_flag are not informative (IV < 0.01).

**Section 3: Model Performance (ROC Curve & KS Statistic)**

* **XGBoost** outperforms **Logistic Regression** in distinguishing defaulters based on ROC curves.
* ROC Curve: XGBoost has a larger AUC (Area Under Curve), indicating better classification.
* KS Statistic confirms the higher separation power of XGBoost.

**Section 4: Feature Importance with SHAP Values**

SHAP summary plot reveals:

* ltv, perform\_cns.score, and disbursed\_amount are most influential features.
* High values of ltv and low credit scores increase default risk.
* Feature effects are visualized using color gradients (red = high, blue = low).

**Section 5: Partial Dependence Analysis**

*Key Insights:*

* **Disbursed Amount**: Higher loan amounts increase default probability.
* **Asset Cost**: Shows a non-linear relationship with risk.
* **UniqueID** should not be used in modeling—it is an identifier, not a predictive feature.

**Section 6: Risk Segmentation Strategy**

Borrowers are segmented based on scores:

* **High Risk** segment: ~26% default rate.
* **Medium Risk**: ~22% default rate.
* **Low Risk**: ~17% default rate.

This confirms effective segmentation for targeted decision-making (e.g., setting interest rates or offering credit limits).

**Conclusion**

The analysis confirms:

* XGBoost is superior for this task.
* Certain features like ltv, perform\_cns.score, and loan amount are highly predictive.
* Segmentation helps prioritize and tailor risk strategies.

This workflow can be used in credit underwriting systems to improve prediction accuracy and reduce loan defaults.