Predictive Modelling for Customer Deposit: Identifying Churn for Better Promotion Allocation

Dipesh Yadav

Aim and Objective:

- To develop a predictive model that accurately identifies customer deposit behaviour within a 24-hour window.
- Improve prediction score to form relation between bonus allocation and customer deposit behaviour.
- Support business decisions for customer engagement, bonus allocation, and marketing strategy by identifying patterns in customer behaviour.

Data Classification and Sample Selection:

- The binary classification provided superior class separation compared to the multiclass model, improving overall prediction accuracy.
- Aligning the target variable ratio with the population dataset, rather than using a random sample, led to better model performance and more reliable outputs.
- $one\ to\ zero\ ratio = \dfrac{class\ one\ count}{class\ zero\ count}$

Model (Class)	Precision	Recall	F1-Score	KS	Gain	Lift
Random Forest (0)	0.91	0.82	0.86	52.35	79.92	1.60
Random Forest (1)	0.52	0.69	0.59			
Light GBM (0)	0.91	0.82	0.86	52.96	80.26	1.61
Light GBM (1)	0.52	0.70	0.60			
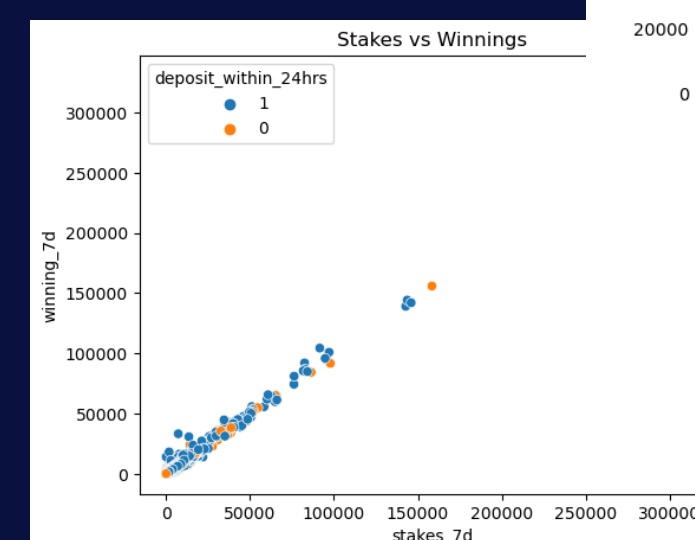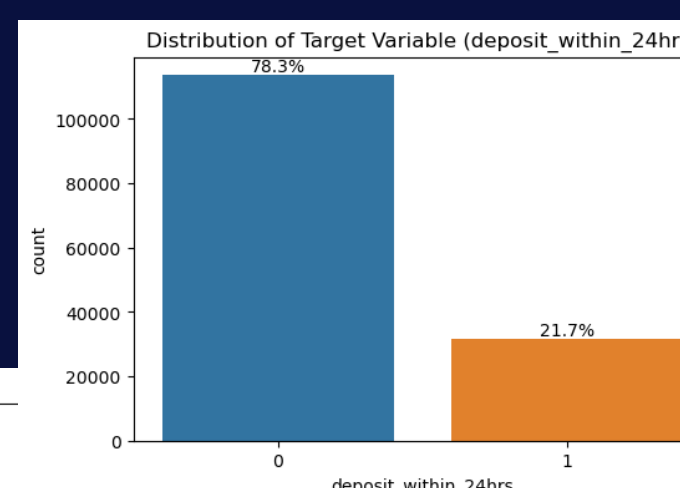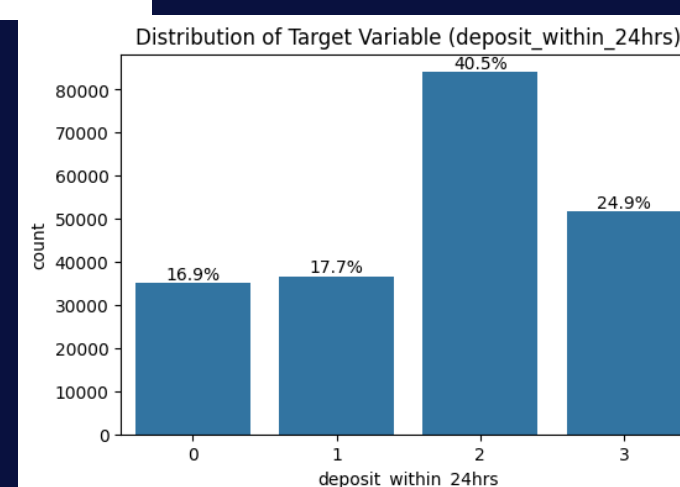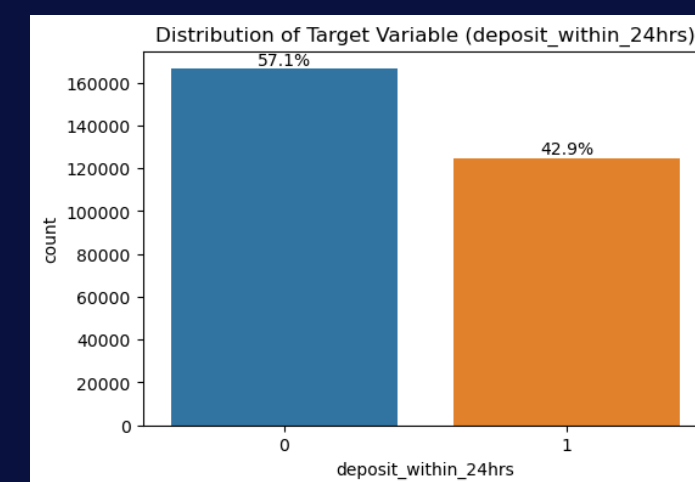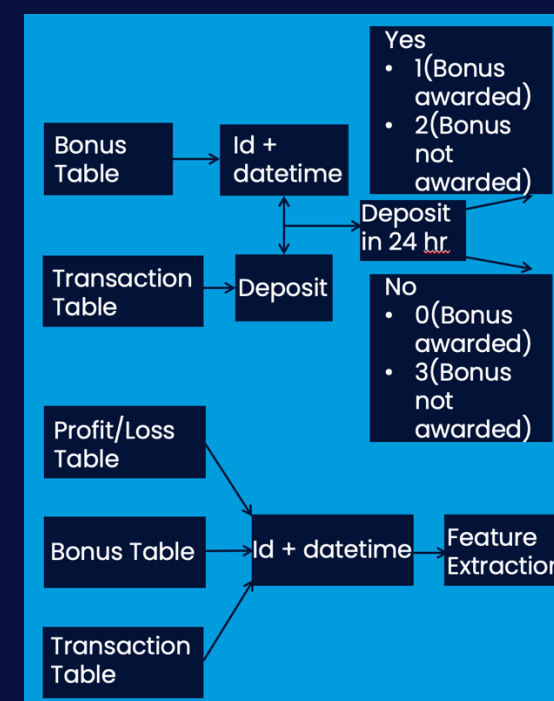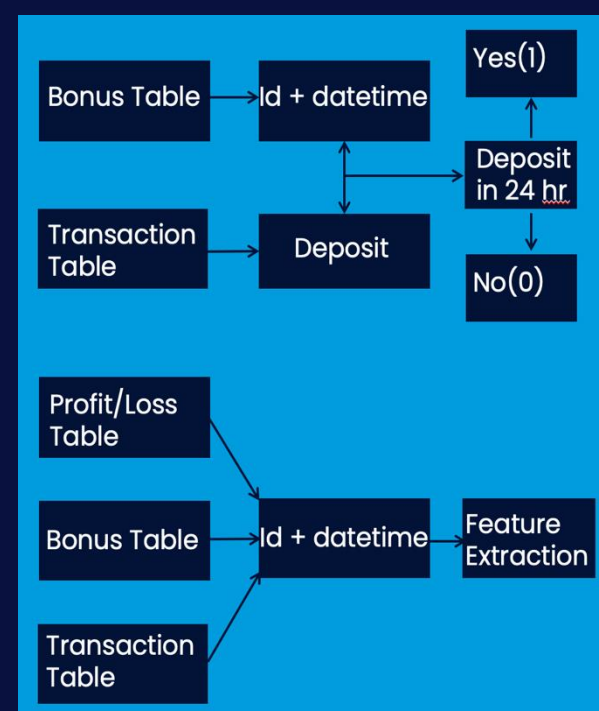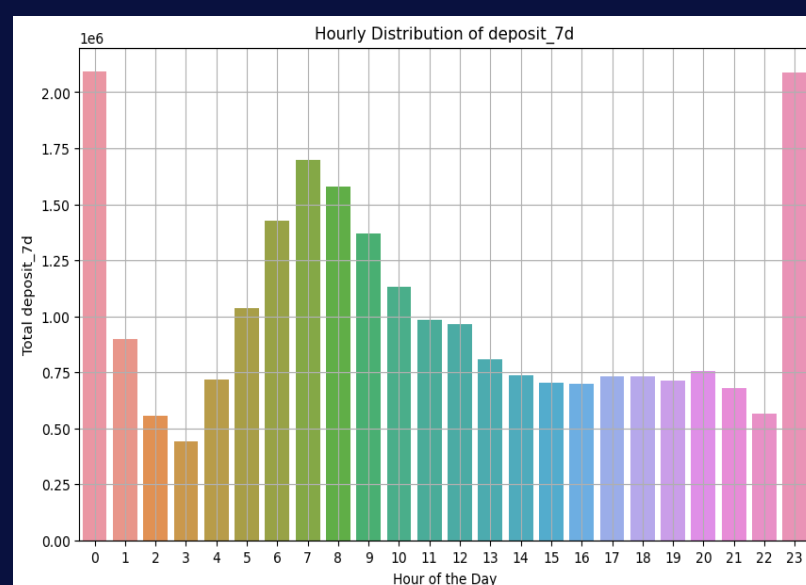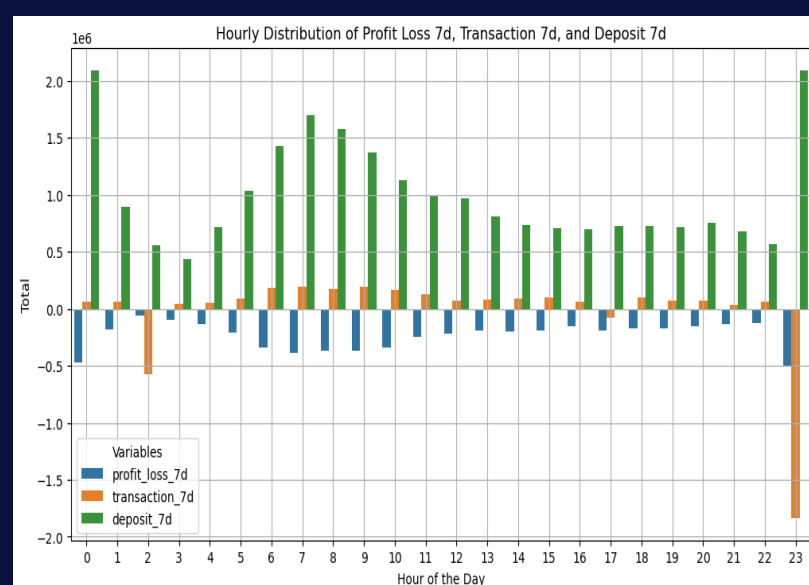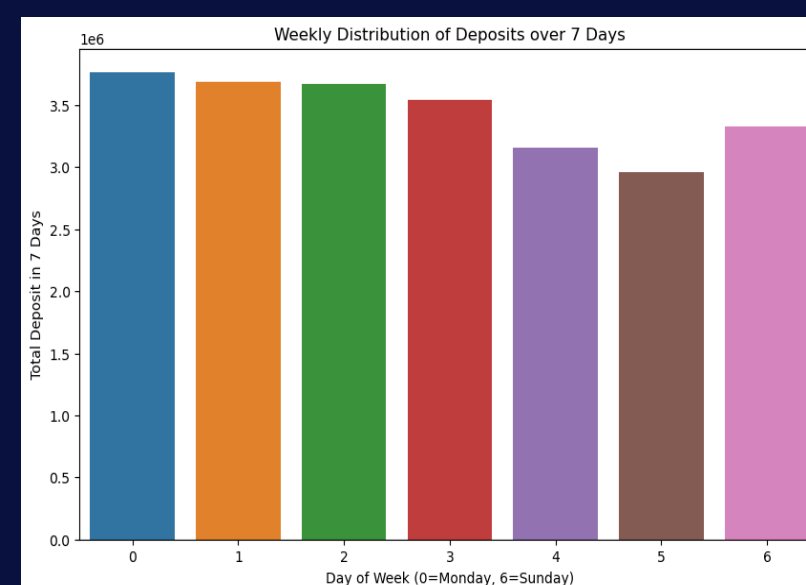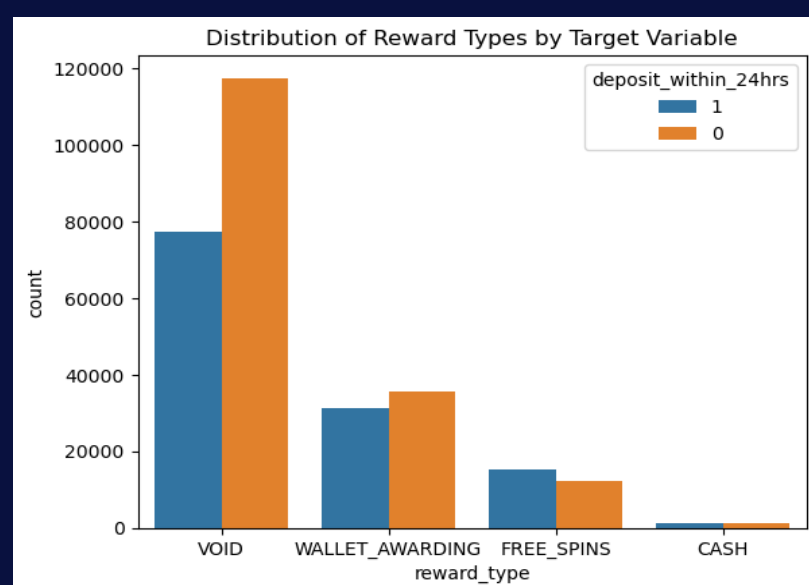
Model Selection And Performance:

- Ensemble systems gives better class separation for imbalanced target variable.
- The precision is initially 47 % but hyperparameter tuning using grid search increased the precision to 52%. Resulting overall ROC-AUC to 76%.
- The performance of bagging and boosting models showed similar outputs.
- KS, Gain and Lift determined a slight edge of boosting models over bagging for better class separation.
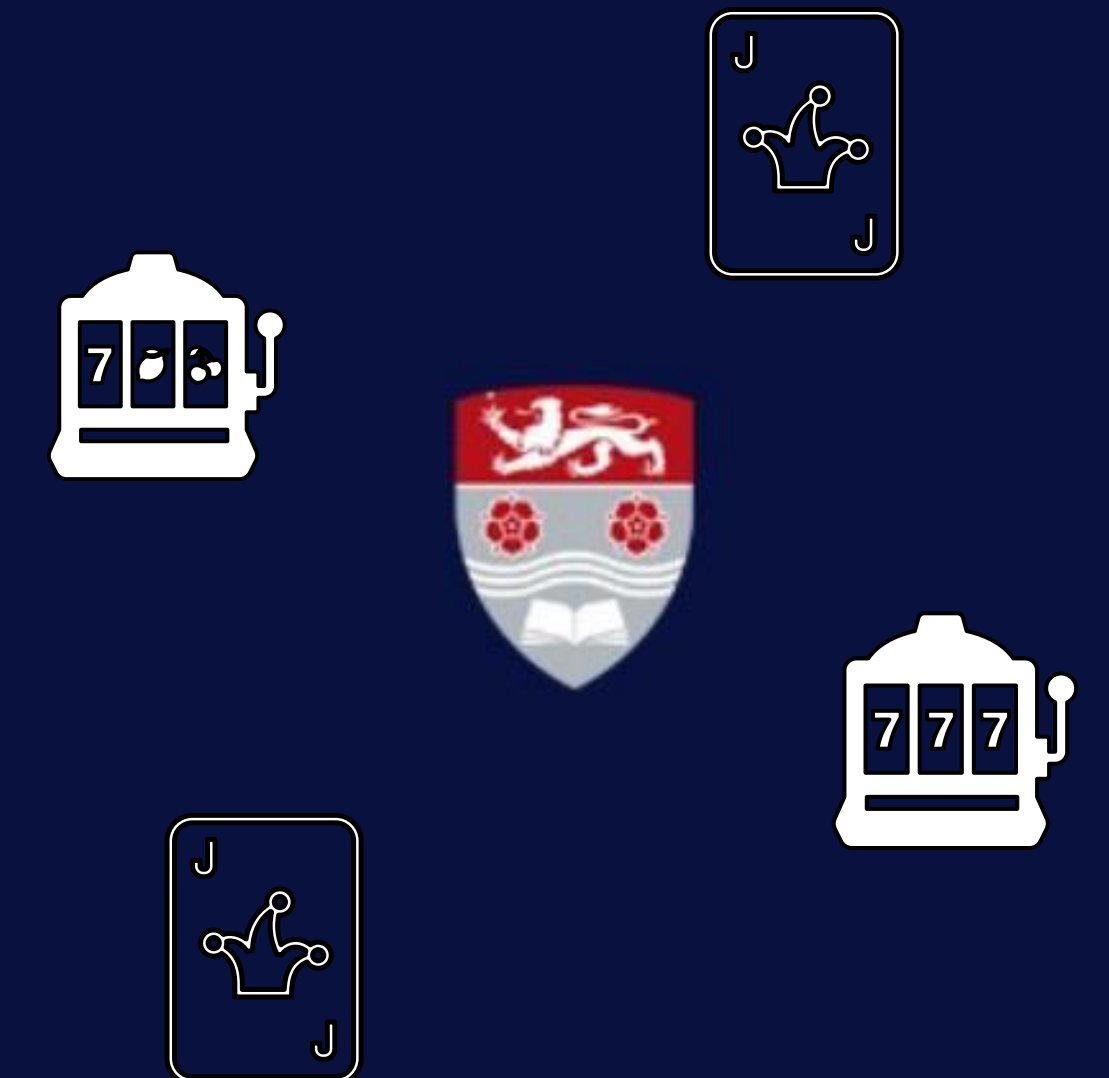
Data Analysis:

- **Target Variable**: Deposits within 24 hours analyzed as binary (deposit/no deposit), with bonus interactions also considered.
- **Time Patterns**: Significant hourly, weekly, and monthly variations in deposit behaviour. Business insights about customer higher interaction timelines.
- **Feature Correlations**: Strong correlation between stakes and winnings; deposit behaviour influenced by bonus reward type.
- **Class Distribution**: Imbalanced target with a higher non-deposit frequency, leading to careful model evaluation.

Conclusion and Result:

- The balancing target variable weak class with respect to population presents better result than random sampling.
- Statistical oversampling does not create bias for this dataset.
- KS, Gain and Lift provide better insight on model selection for similar results.
- Bonus, time of the day and profit/loss are the most influential for deposit behaviour of the customer.
- Free spins reward type presents high customer involvement in deposits after reward allocation.