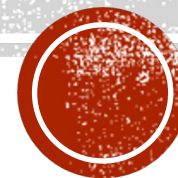








# CUSTOMER CHURN ANALYSIS



# OUTLINE

- 
- 
- 
- 

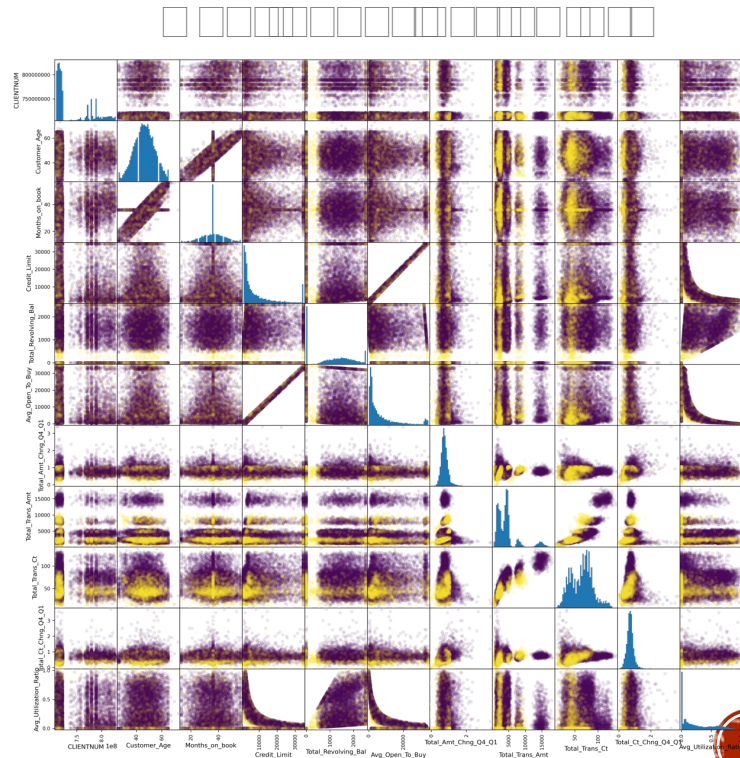
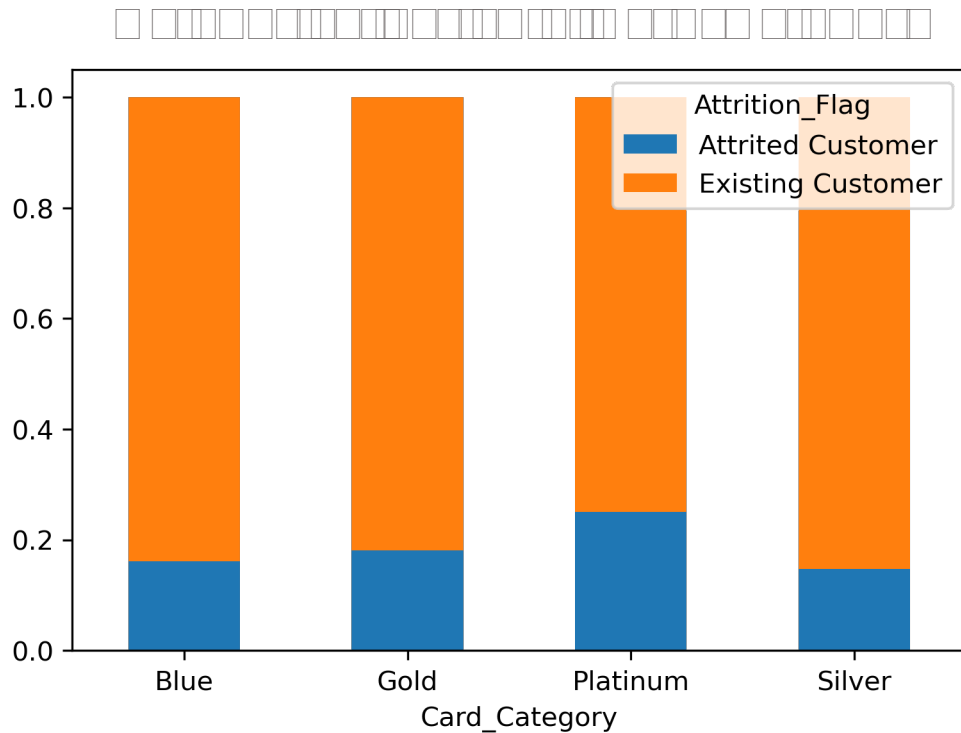


# BACKGROUND

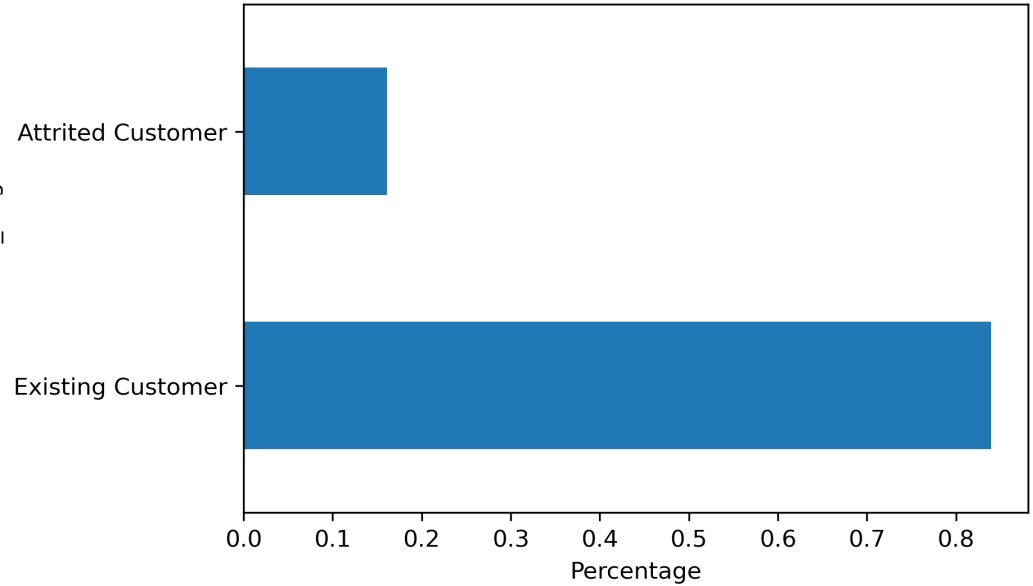
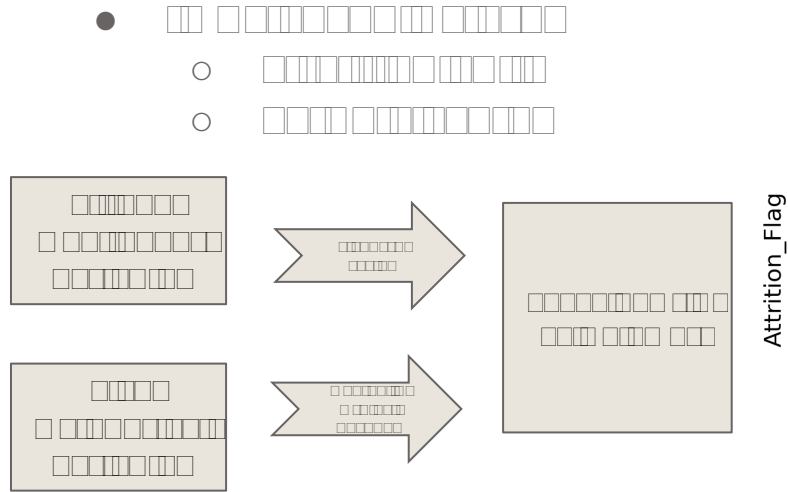
- **Customer Churn Prediction Analysis:** This analysis aims to identify factors that lead to customer churn, such as low engagement, poor customer service, or lack of product usage. By understanding these factors, businesses can implement targeted interventions to reduce churn and improve customer retention.
- **Importance and Data Overview:** Customer churn is a critical metric for businesses, as it directly impacts revenue and growth. This section provides an overview of the data used in the analysis, including customer demographics, usage patterns, and service history.



# EDA



# PREPROCESSING



# CROSS VALIDATION PIPELINE

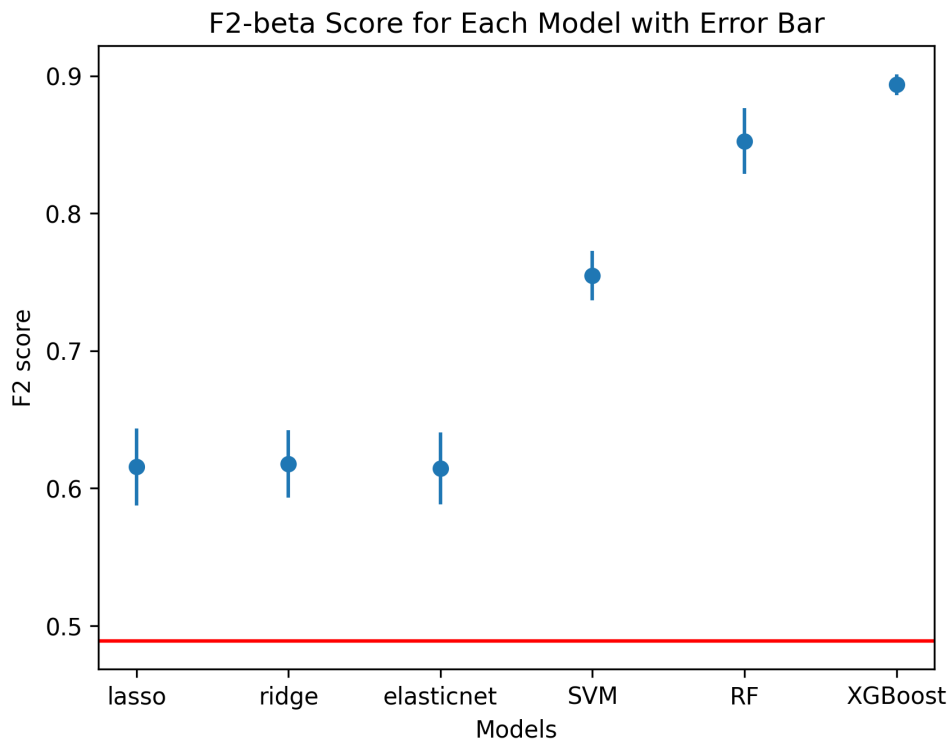
- **Random State Selection** 1. 1000개의 랜덤 상태를 생성한다. 2. 각 랜덤 상태에 대해 데이터를 랜덤으로 분할한다. 3. 분할된 데이터를 사용하여 모델을 학습하고 평가한다.
- **Data Splitting** 1. 데이터를 랜덤으로 분할한다. 2. 분할된 데이터를 사용하여 모델을 학습하고 평가한다. 3. 분할된 데이터를 사용하여 모델을 학습하고 평가한다.
- **Preprocessing** 1. 데이터를 전처리한다. 2. 전처리된 데이터를 사용하여 모델을 학습하고 평가한다. 3. 전처리된 데이터를 사용하여 모델을 학습하고 평가한다.
- **Hyperparameter Tuning** 1. 하이퍼파라미터를 조정한다. 2. 조정된 하이퍼파라미터를 사용하여 모델을 학습하고 평가한다. 3. 조정된 하이퍼파라미터를 사용하여 모델을 학습하고 평가한다.
- **Model Evaluation** 1. 모델을 평가한다. 2. 평가된 모델을 사용하여 데이터를 예측한다. 3. 예측된 데이터를 사용하여 모델을 학습하고 평가한다.



[illegible]

# RESULTS

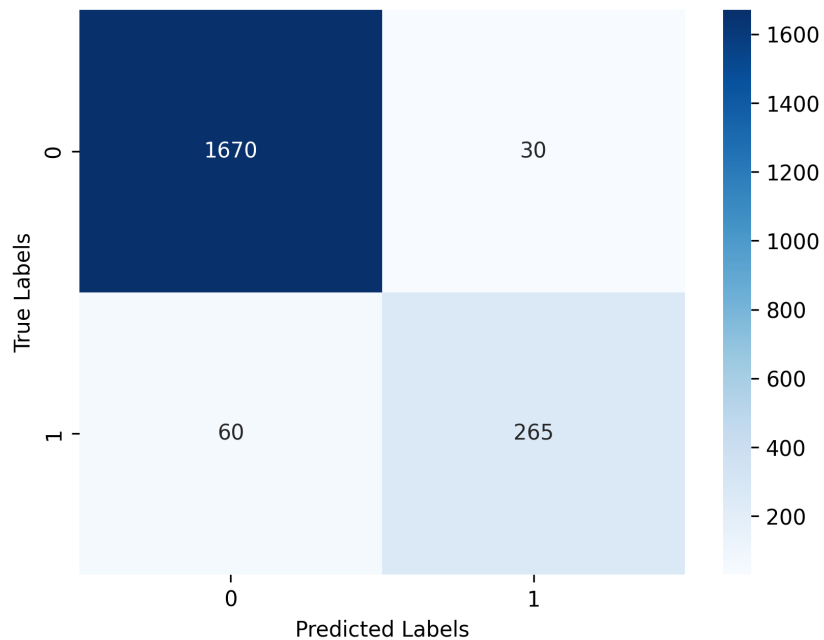
	Model	F2 Beta Score (Mean $\pm$ Std)
0	Lasso Logistic Regression	0.616 $\pm$ 0.031
1	Ridge Logistic Regression	0.618 $\pm$ 0.027
2	SVM	0.755 $\pm$ 0.02
3	Random Forest	0.853 $\pm$ 0.027
4	XGBoost	0.894 $\pm$ 0.008
5	Baseline	0.489



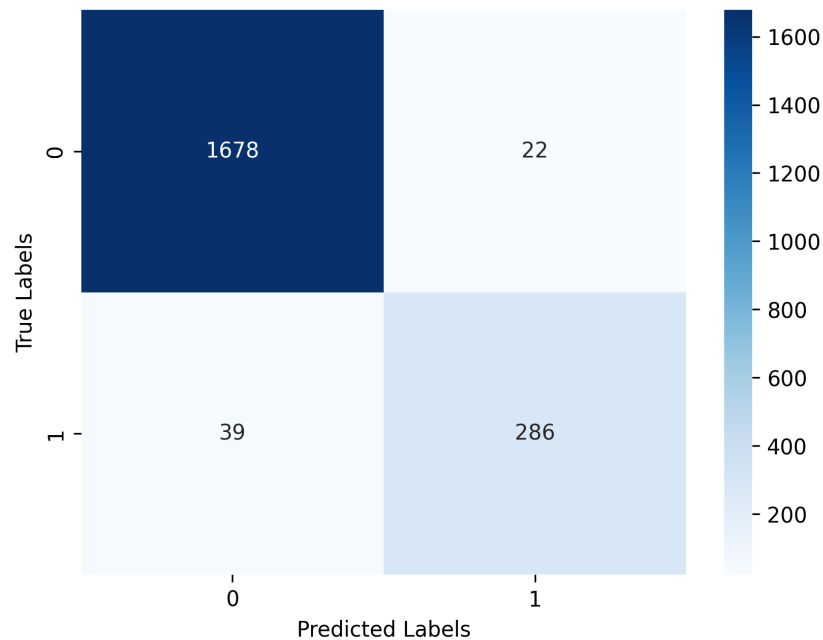


# RESULTS CONTINUE

Confusion Matrix for Random Forest

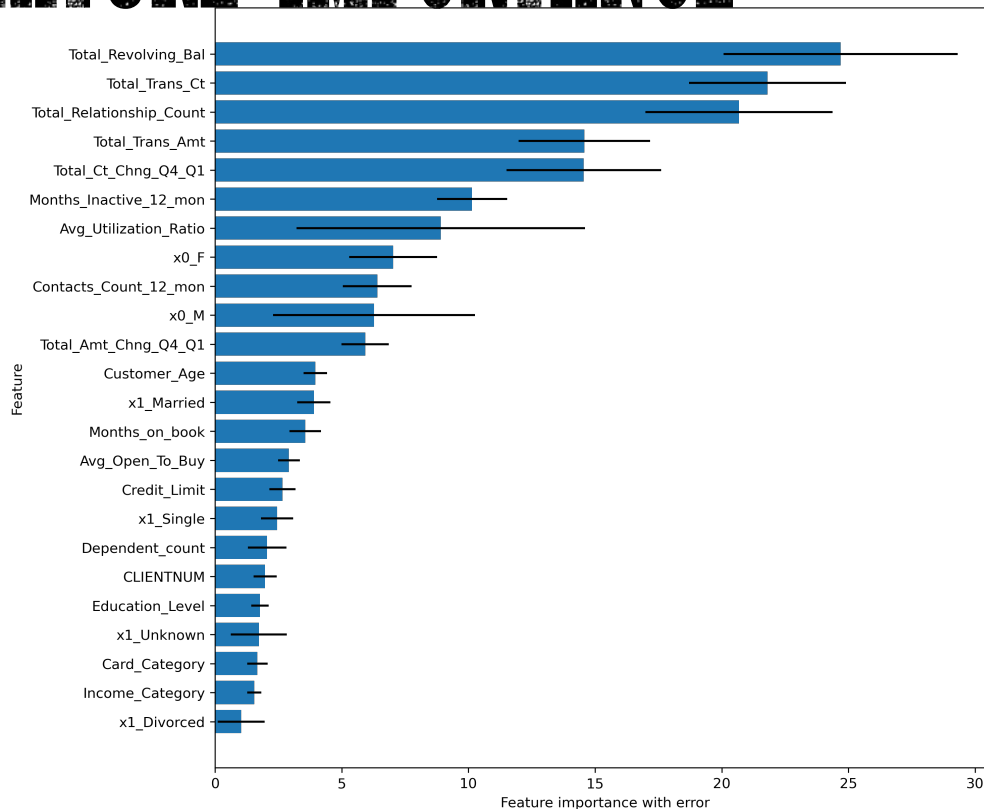


Confusion Matrix for XGBoost

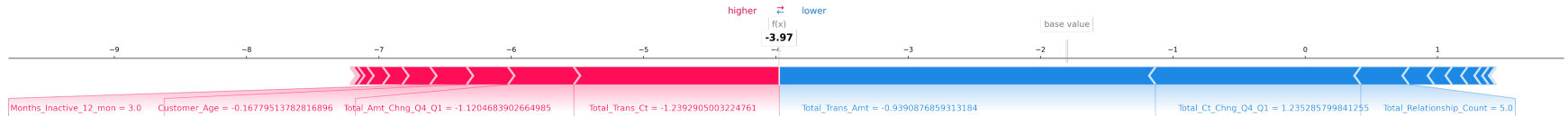
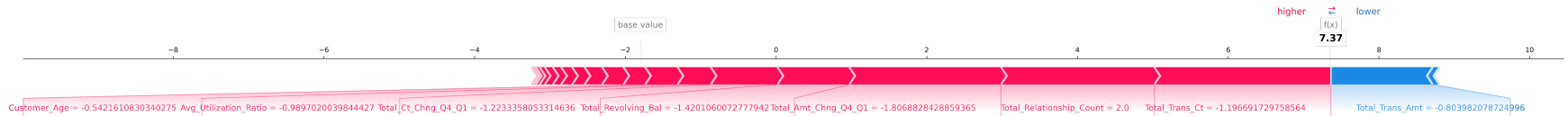


# GLOBAL FEATURE IMPORTANCE

- Total\_Revolving\_Bal is the most influential feature when predicting whether a customer will churn using XGBoost and 'Gain' as the importance metric.



# LOCAL IMPORTANCE



# OUTLOOK

- **Expand Dataset:** Collect more data from various sources to improve model performance and generalization.
- **Enhanced Hyperparameter Tuning:** Implement advanced optimization techniques to find the best model configuration.
- **Advanced Models:** Explore state-of-the-art models and architectures to achieve better results.
- **Interpretability:** Develop methods to understand and explain the model's predictions and decisions.



# Q & A

