Customer Churn Prediction

BA305 Team 5

Joshua Nahm, Jiwon Choi, Jaejoong Kim, Shinyeong Park, Seokhoon Shin

Agenda

01 Intro to Our Dataset

02 EDA

03 Model Evaluation

04 Findings Implications

05 Conclusion

The significant impact of reducing churn motivated us to create a predictive model

for the Telecom Industry

Potential Dataset

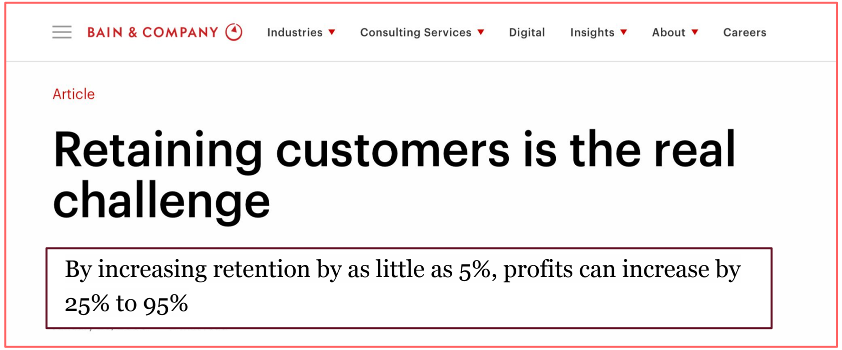
Retail Data

E-commerce Data

Craft beer Data

Historical sales and

Active Inventory Data

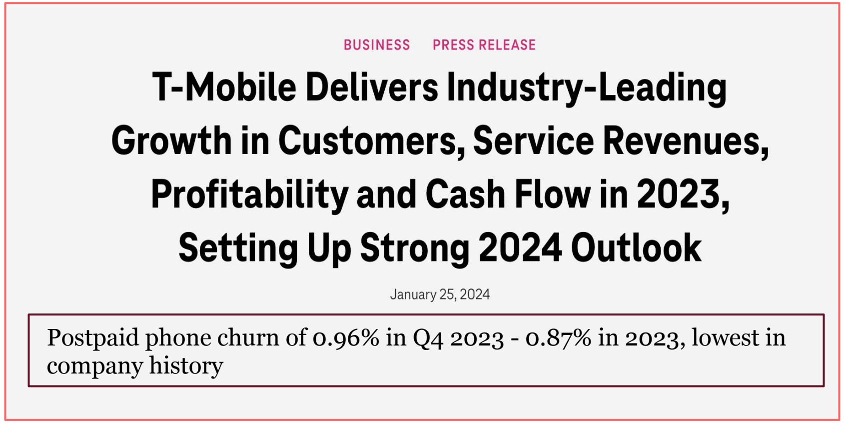


BAIN & COMPANY Industries Consulting Services Digital Insights About Careers

Article

Retaining customers is the real challenge

By increasing retention by as little as 5%, profits can increase by 25% to 95%



BUSINESS PRESS RELEASE

T-Mobile Delivers Industry-Leading Growth in Customers, Service Revenues, Profitability and Cash Flow in 2023, Setting Up Strong 2024 Outlook

January 25, 2024

Postpaid phone churn of 0.96% in Q4 2023 - 0.87% in 2023, lowest in company history

Our main goal is to develop a churn prediction model

that ultimately leads the business to long-term profitability and consistent growth

1. **Analyze the behavior of the customers to timely and accurately predict the possible churners**
2. **Understand the key factors driving their likelihood to churn and improve the overall customer experience**
3. **Effectively optimize resource allocation and can design retention policies, develop efficient customer management strategies**

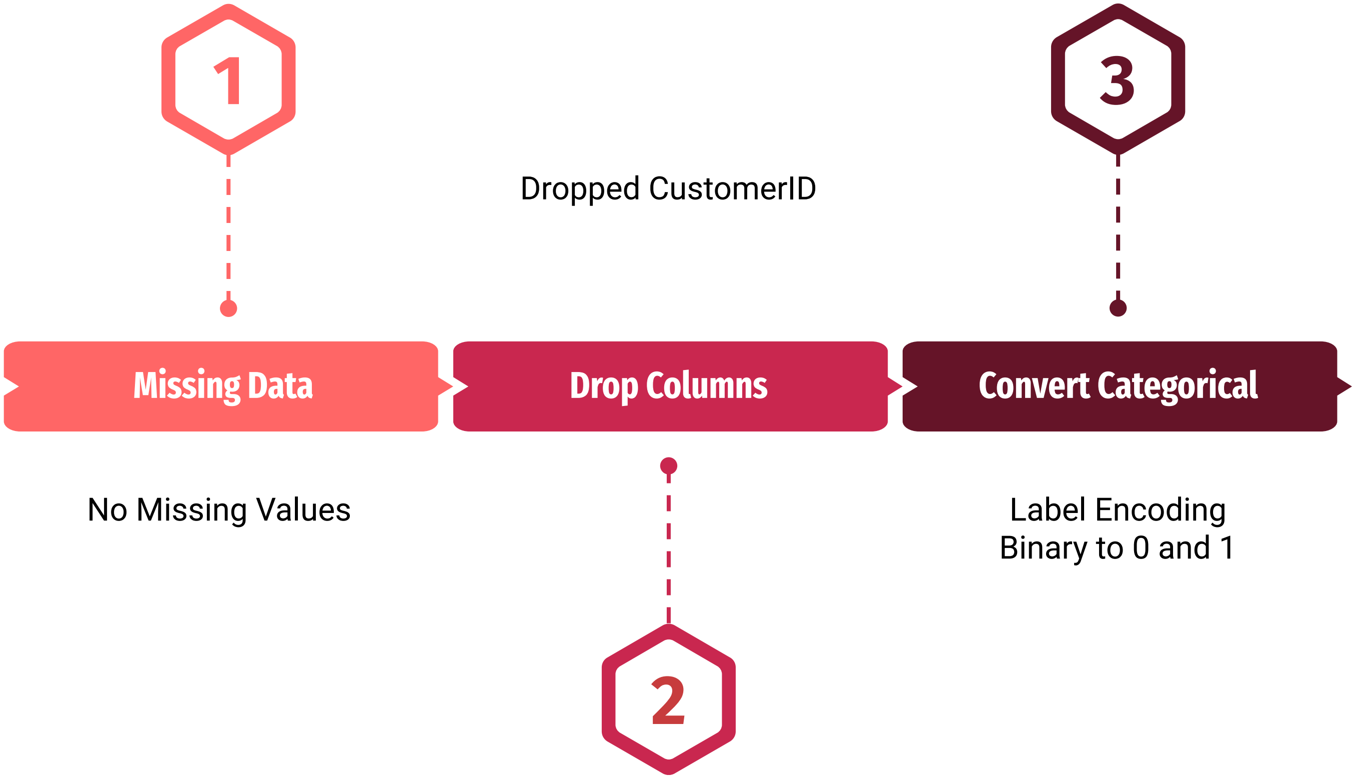
- Long-term Profitability
- Consistent Growth

We used the Telco Customer Churn dataset

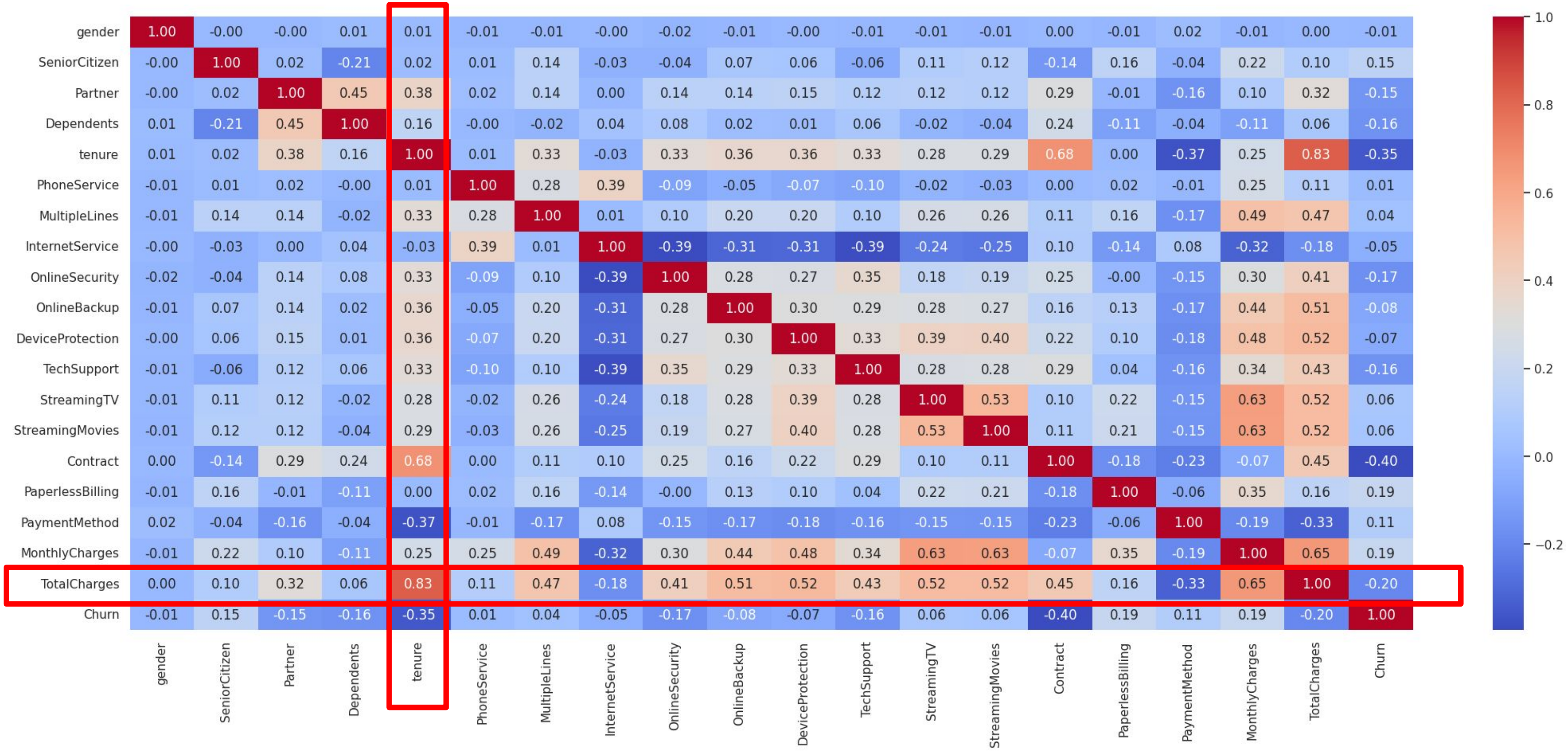
21 Features **7043 Observations**

CustomerID	Gender	Senior Citizen	Partner	Dependents	tenure	Phone Service
Multiple Lines	Internet Service	Online Security	Online Backup	Device Protection	Tech Support	Streaming TV
Streaming Movies	Contract	Paperless Billing	Payment Method	Monthly Charges	Total Charges	Churn

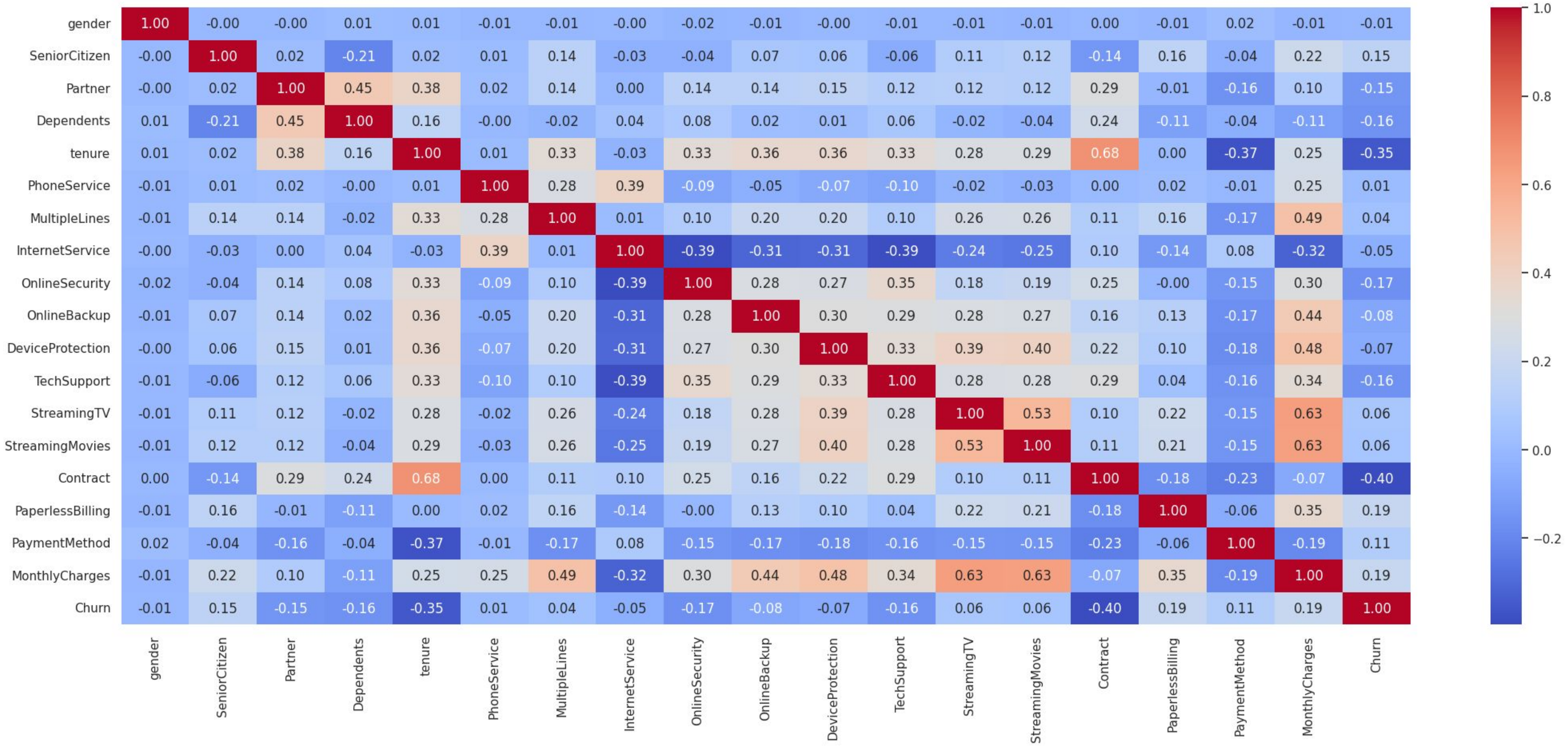
This was our data preprocessing step



There is a high correlation between TotalCharges and Tenure

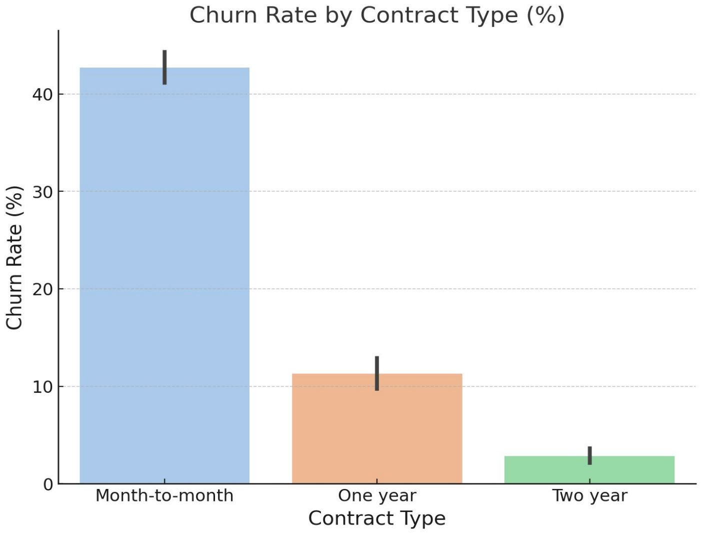


There is no strong correlation after dropping TotalCharges

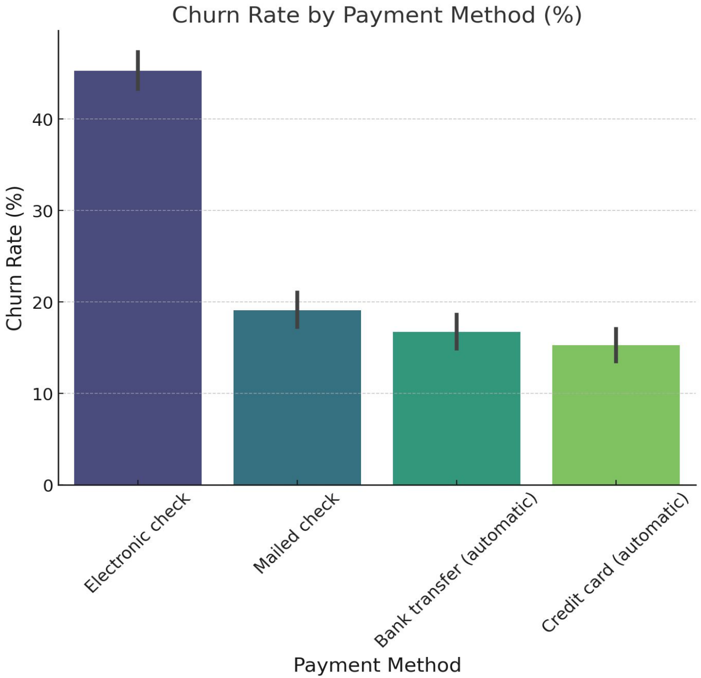


Customers on month-to-month contracts who pay with electronic checks

are more likely to churn

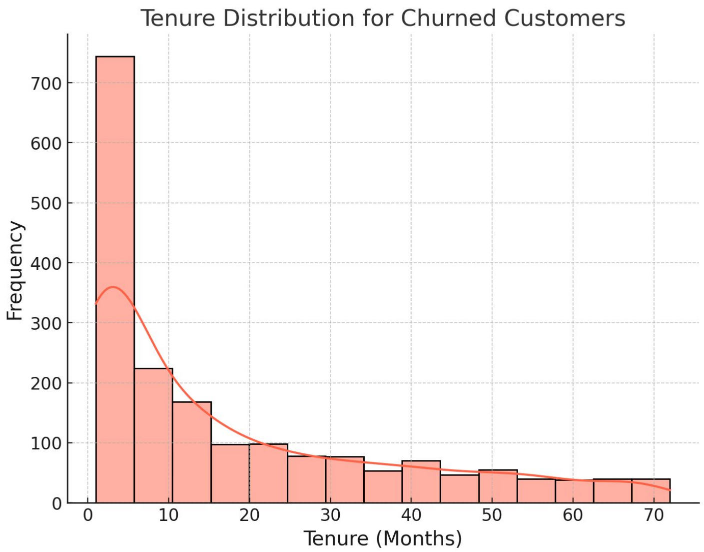


Customers on **month-to-month** contracts account for **~89%** of churned customers

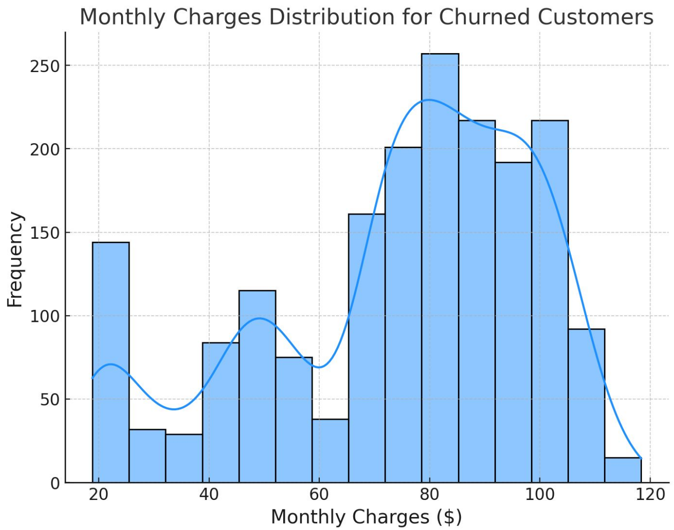


Customers who pay with **electronic checks** account for **~57%** of churned customers

Customers with short tenures and higher monthly charges
are at a greater risk



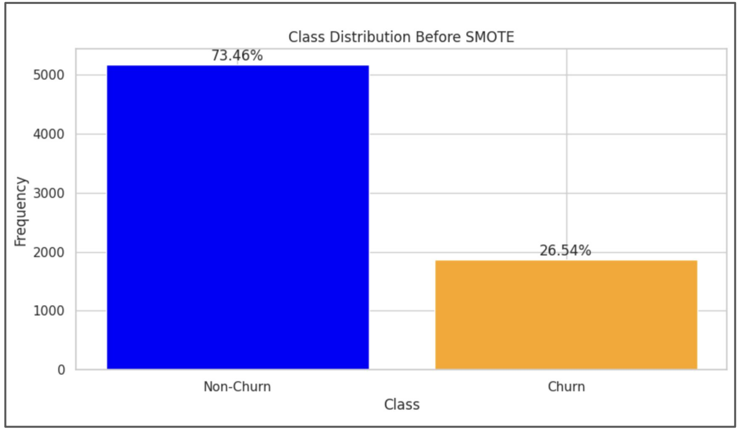
Customers with **short tenures (0-6 months)** are at higher risk **(~43%)**



Higher monthly charges correlate with **increased** churn likelihood
Customers with **lower charges (<$40)** show **lower** churn rates

We used SMOTE to address class imbalance

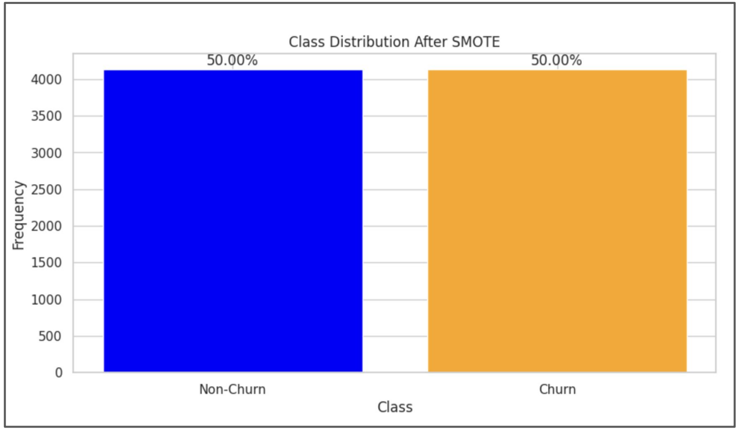
Before



Non-Churn : 73.46 %

Churn : 26.54 %

After



Non-Churn : 50 %

Churn : 50 %

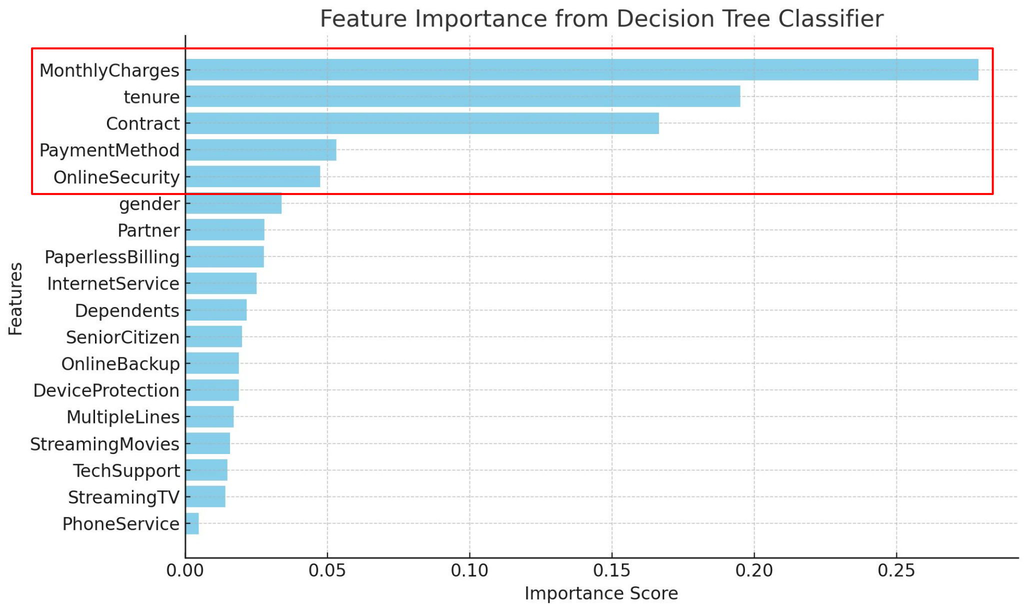
★ So What is SMOTE?

➔ A technique to generate synthetic examples for the **minority class (churned customers)**, helping models **overcome bias** and better predict the minority class.

Determine the Most Important Features

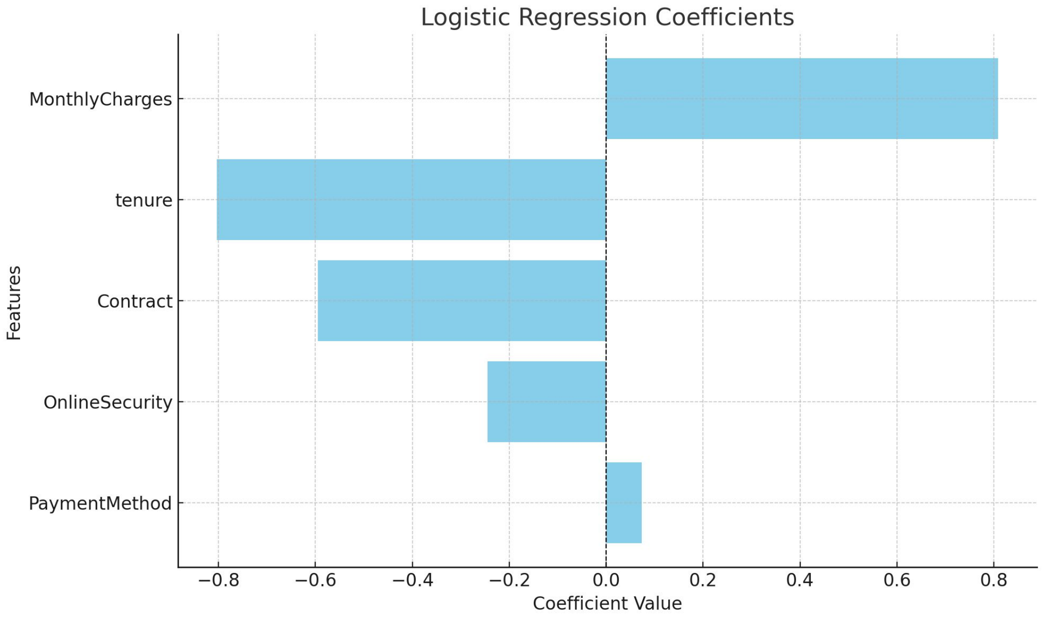
Decision Tree Classification to Assess Important Features

- **MonthlyCharges**: Most important feature, with high influence on churn. Higher charges significantly increase churn likelihood.
- **Tenure**: Second most important, with shorter tenures strongly associated with churn.
- **Contract**: Longer-term contracts reduce churn risk, making this a crucial retention strategy.
- **PaymentMethod**: Certain methods like electronic checks are associated with higher churn.
- **OnlineSecurity**: Customers subscribing to security add-ons are less likely to churn.



Logistic Regression to Highlight Direction and Magnitude

- **MonthlyCharges**: Strong positive relationship with churn.
  - Higher charges significantly increase churn likelihood.
- **Tenure**: Negative impact on churn.
  - Longer-tenured customers are less likely to churn, but early-tenure customers are at higher risk.
- **Contract**: Negative impact on churn.
  - Customers with long-term contracts are far less likely to churn compared to those on month-to-month plans.
- **OnlineSecurity**: Negative relationship with churn.
  - Customers who subscribe to security services are less likely to leave.
- **PaymentMethod**: Slight positive effect.
  - Customers using electronic checks show a marginally higher likelihood of churning.



"

We Built Models to Identify Churn Customers

in Advance

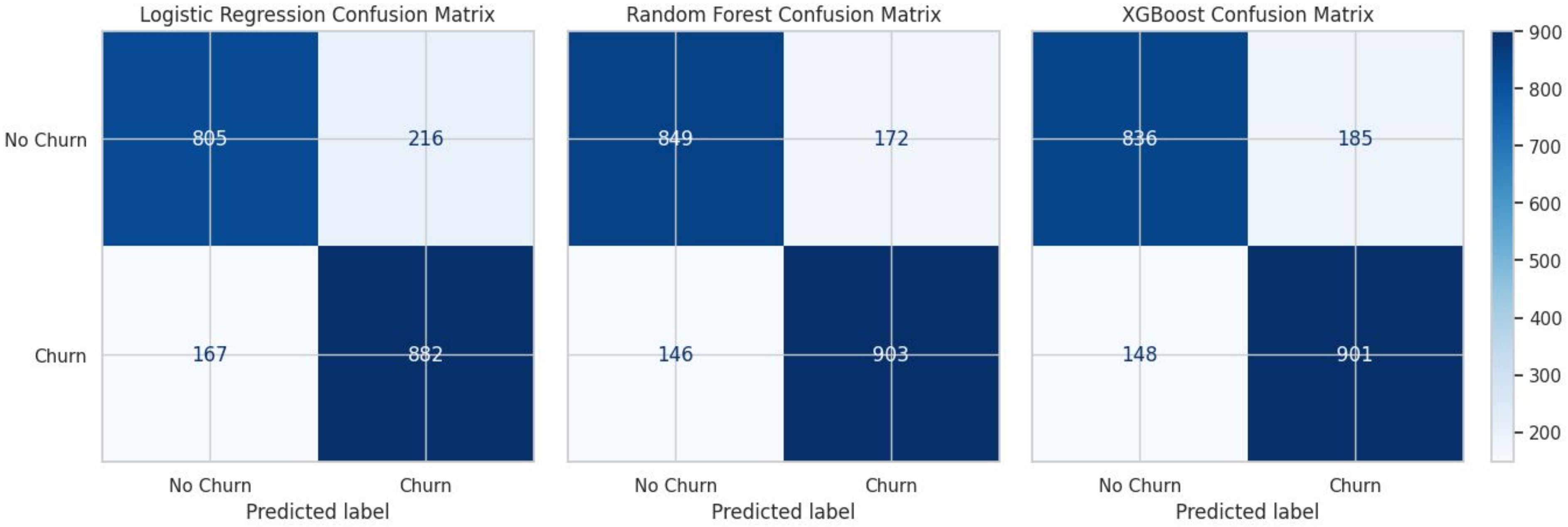
We selected five different models and tested their accuracy

all higher than 73.46% (Naive Rule)

	Logistic Regression	Random Forest	XGBoost	CatBoost	LightGBM
PROS	Simple, Interpretable	Robust, Interpretable, Resistant to overfitting	Captures complex relationships	Excellent handling categorical variables, less preprocessing	Fast training, highly scalable, efficient for larger datasets
CONS	Limited ability to model complex patterns	Computationally intensive with larger datasets	Computationally heavy	Slower training compared to LightGBM	Less interpretable compared to Random Forest
RESULTS	Accuracy = 0.81 F1-Score = 0.81	Accuracy = 0.85 F1-Score = 0.85	Accuracy = 0.84 F1-Score = 0.84	Accuracy = 0.84 F1-Score = 0.85	Accuracy = 0.84 F1-Score = 0.84

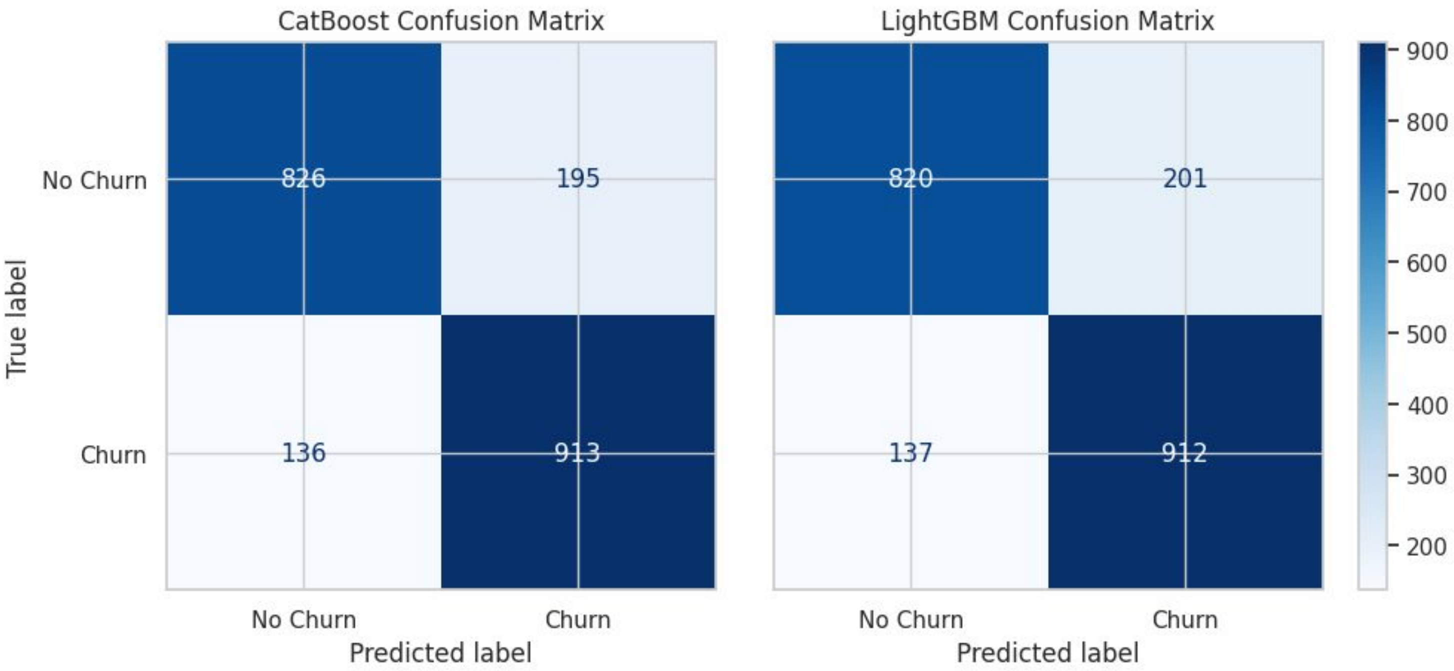
Let’s deeply dive into each model via Confusion Matrix

Logistic Regression | **Random Forest** | **XGBoost**



CatBoost

LightGBM



Random Forest & LightGBM turns out to be the best models

here are five reasons why...

01

Robustness & Interpretability

Random Forest minimizes overfitting, enhances interpretability

02

Efficiency

LightGBM offers high computational speed and low memory

03

Balanced Error Management

Optimizes false positives and negatives handling

04

Scalability

LightGBM handles large datasets effectively

05

Accuracy

Consistent, reliable performance on new data

Our training and test set accuracies confirm that

our models do not overfit

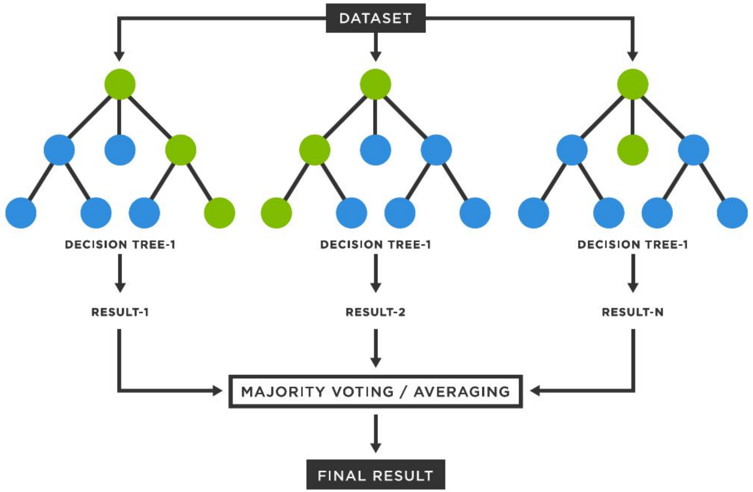
Random Forest	LightGBM
Training: 86%	Training: 85%
Test: 85%	Test: 84%

Specific interpretations of Random Forest

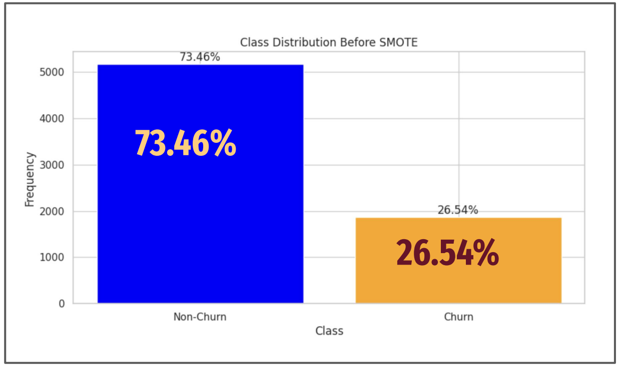
which shows why it is the best model...

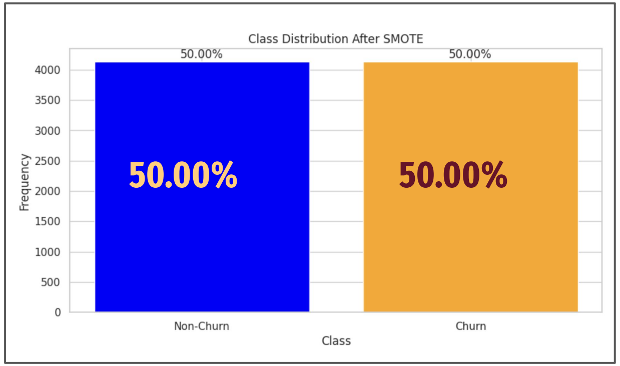
	Random Forest Classifiers with a random state of 42
PHASE 1	Trained on the standardized and scaled training data through Standard Scaler
PHASE 2	Demonstrated a high level of accuracy at approximately 85% on the test set
PHASE 3	Effectively balances FP & FN minimizing false positives and detecting churners. (172 False Positives, 146 False Negatives)
PHASE 4	Continue to fine tune the Random Forest model for enhanced performance

Random Forest



Addressing Class Imbalance: The Role of SMOTE





What is the impact?

Before :

01 Models favored non-churn customers due to class imbalance.

02 Low recall for churned customers resulted in missed retention opportunities.

After :

03 Improved **recall** for churned customers and overall performance metrics.

04 Example Improvements : **Random Forest** -> 85 % accuracy with balanced F1-scores.

CatBoost & XGBoost -> Better identification of Churned Customers.

Example Cases

Company A

Customer 1 - Joshua

Monthly Charges :	Tenure :	Contract Type :	Payment Method :
$55	**3 Months**	**Month-to-Month**	**Electronic Check**

Customer 2 - Jason

Monthly Charges :	Tenure :	Contract Type :	Payment Method :
$35	**8 Months**	**Month-to-Month**	**Credit Card**

Customer 3 - Jay

Monthly Charges :	Tenure :	Contract Type :	Payment Method :
$40	**18 Months**	**2 year contract**	**Bank Transfer**

Example Cases

Company A

Customer 1 - Joshua

75 % to CHURN

Monthly Charges : **$55**

Tenure : **3 Months**

Contract Type : **Month-to-Month**

Payment Method : **Electronic Check**

Customer 2 - Jason

58 % to CHURN

Monthly Charges : **$35**

Tenure : **8 Months**

Contract Type : **Month-to-Month**

Payment Method : **Credit Card**

Customer 3 - Jay

15 % to CHURN

Monthly Charges : **$40**

Tenure : **18 Months**

Contract Type : **2 year contract**

Payment Method : **Bank Transfer**

What we can do for Future Studies...

01

INCREASE focus on Customer Engagement for Month-to-Month Contracts

02

INCREASE focus on Monthly Charges and Tenure

03

Up-to-DATE Teleco's Dataset
* Changes in how customers use services

04

Integrating External Datasets

What we can do for Future Studies...

02

04

★ Apply Promotions to the Customers
- One Month FREE
- Provide Discount COUPON

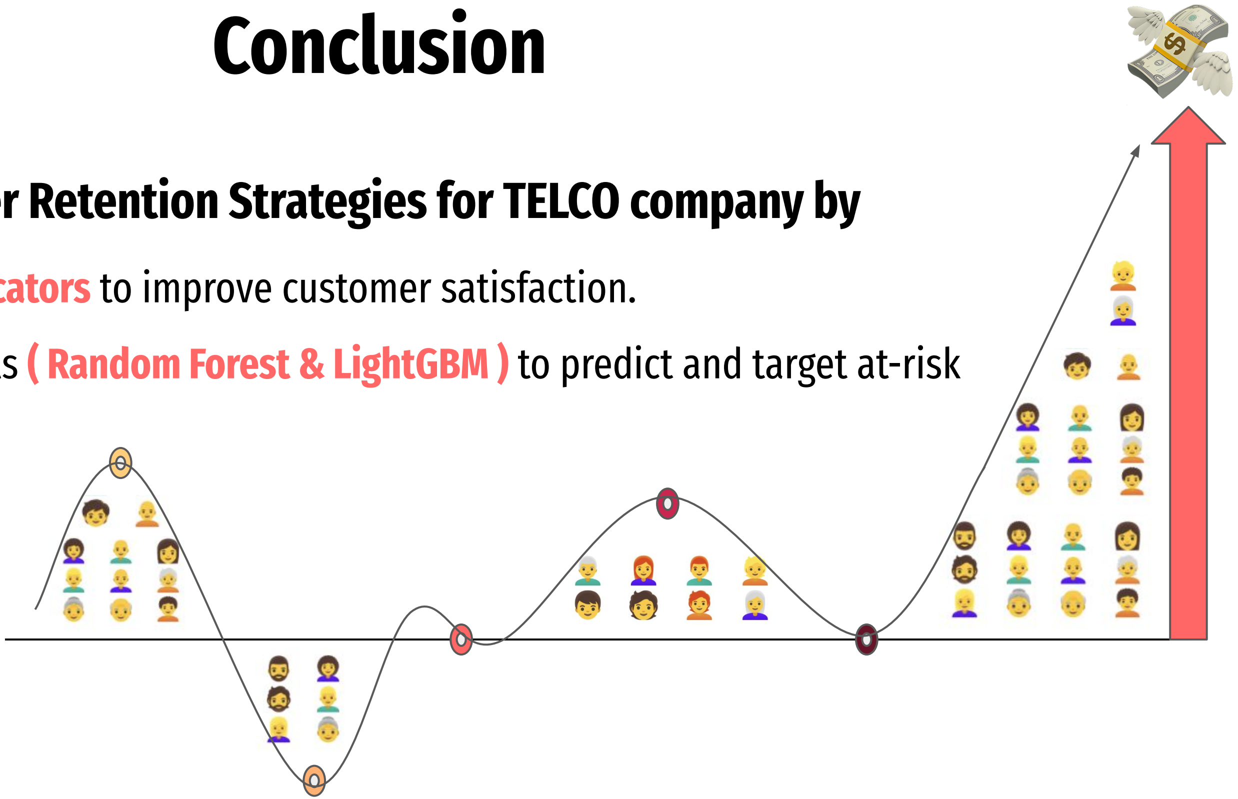
IN

Contracts

Conclusion

We have designed Customer Retention Strategies for TELCO company by

1. Identifying **Key Churn indicators** to improve customer satisfaction.
2. Using high-accuracy models **(Random Forest & LightGBM)** to predict and target at-risk customers.



★ **This project allowed us to apply what we have learned in class to a real-world situation, teaching us how to analyze customer behavior and communicate insights to drive business strategies.**

Thank You

Any Questions?