

Customer Success Score

Data source: Telco-Customer-Churn

Yao Xu

Data Information

- Customer demographics (gender, age range, partner/dependent status)
- Account information (tenure, contract type, payment method)
- Services each customer has signed up for (phone, internet, security, etc.)
- Churn status (whether the customer left within the last month)

Implement Steps

Data Preparation

- Clean data types
- Missing values handling

Data Visualization

- Correlation
- Statistical Test

Feature Engineering

- Data Preprocessing
- Feature Selection

Modeling

- Data imbalance Handling
- Normalization
- Model Comparison
 - Random Forest, XGBoost, CatBoost, Logistic Regression
- Performance Improvement

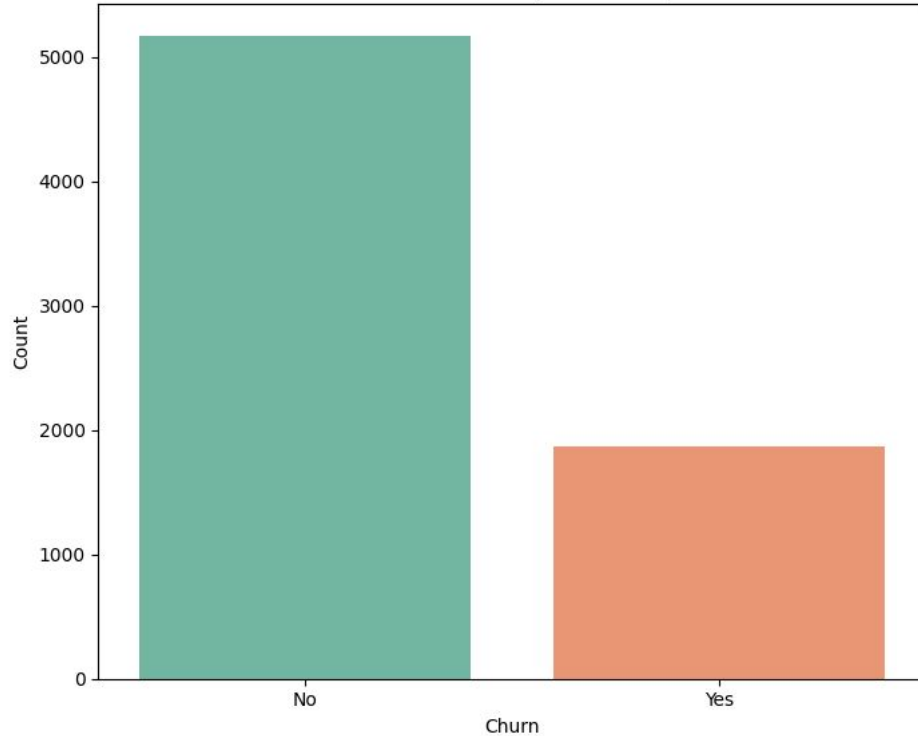
Customer Success Score

- Feature Coefficient
- Customer Success Score Normalization

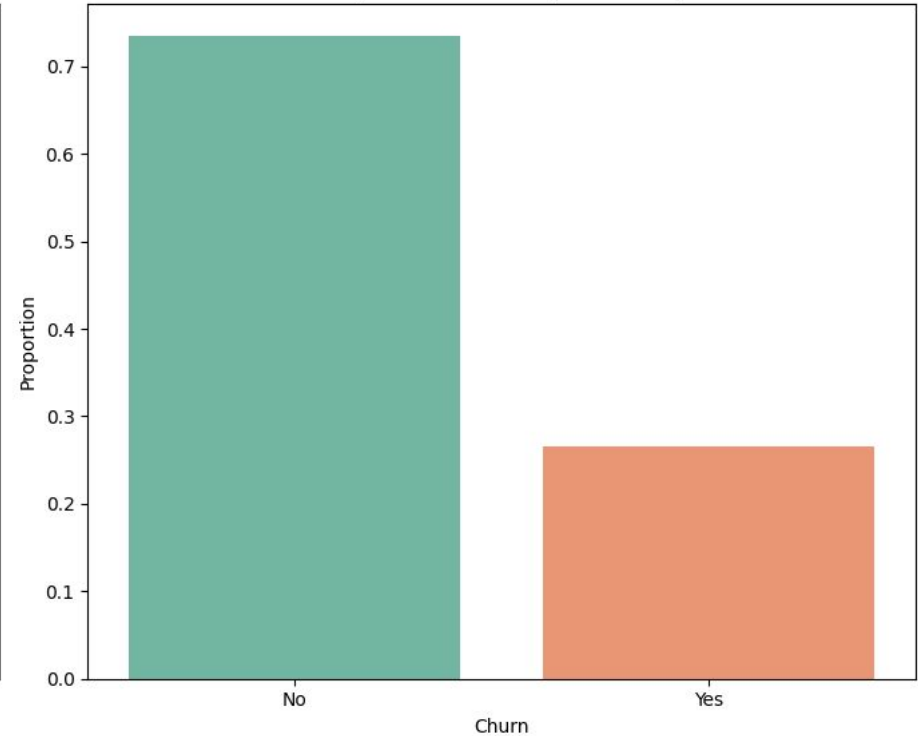
Data Visualization

- Churn Rate Over 70%

Counts of Churn (Yes and No)

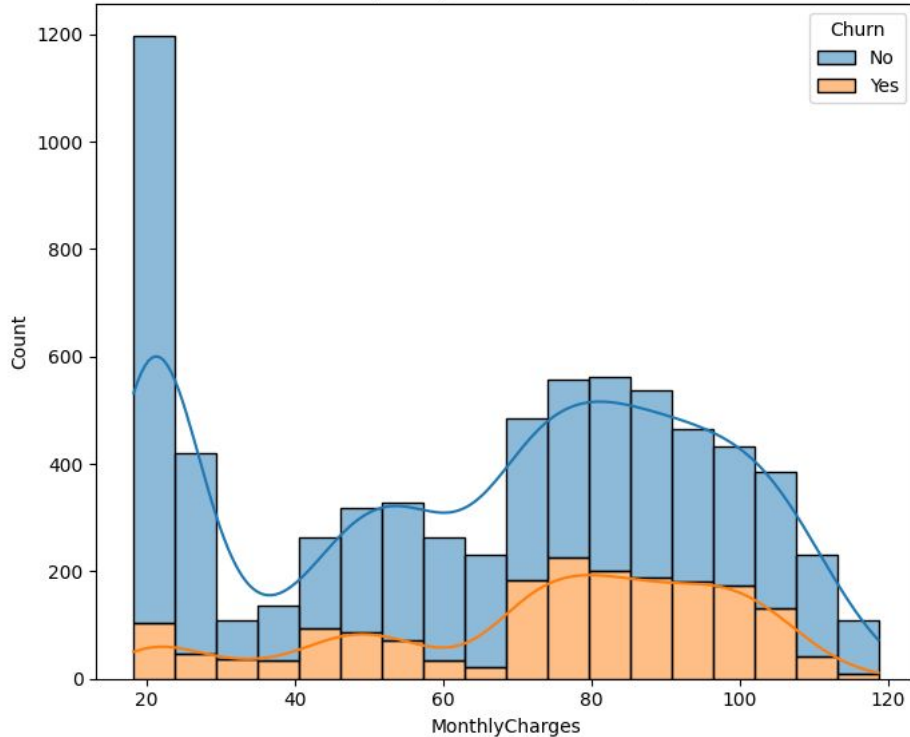


Proportion of Churn (Yes and No)

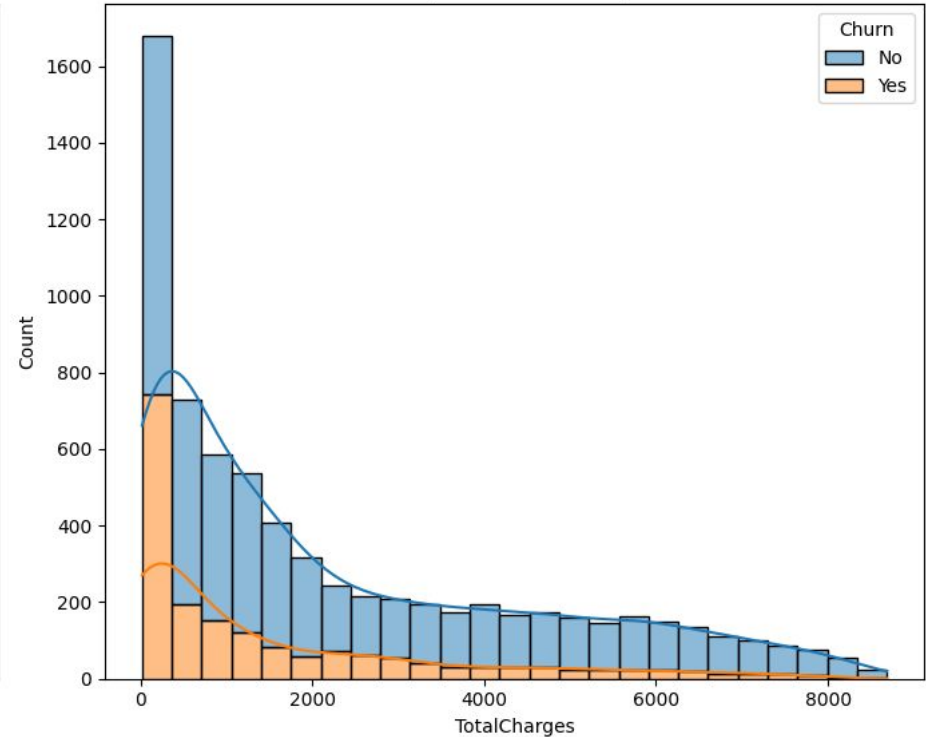


- Higher monthly charges (\$70-100) appear associated with increased churn risk.
- Low monthly charges (~\$20) show strong retention.

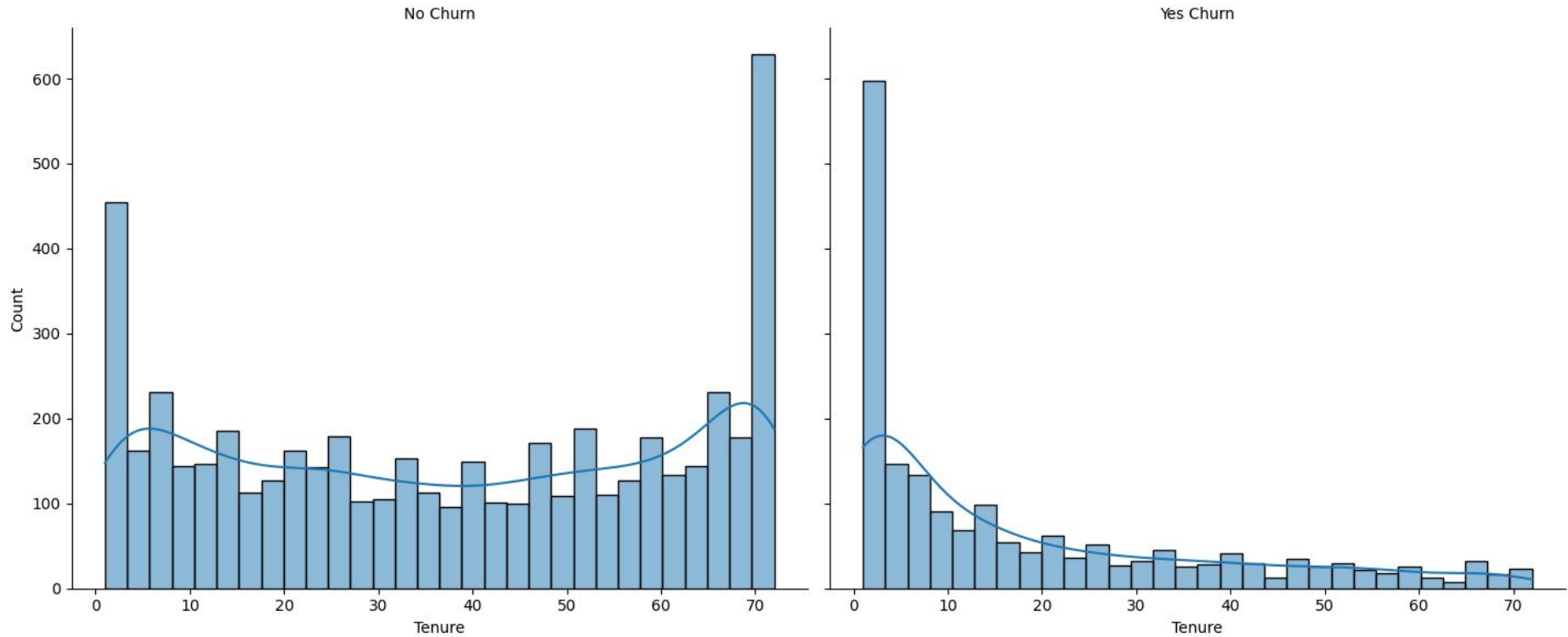
MonthlyCharges Distribution by Churn



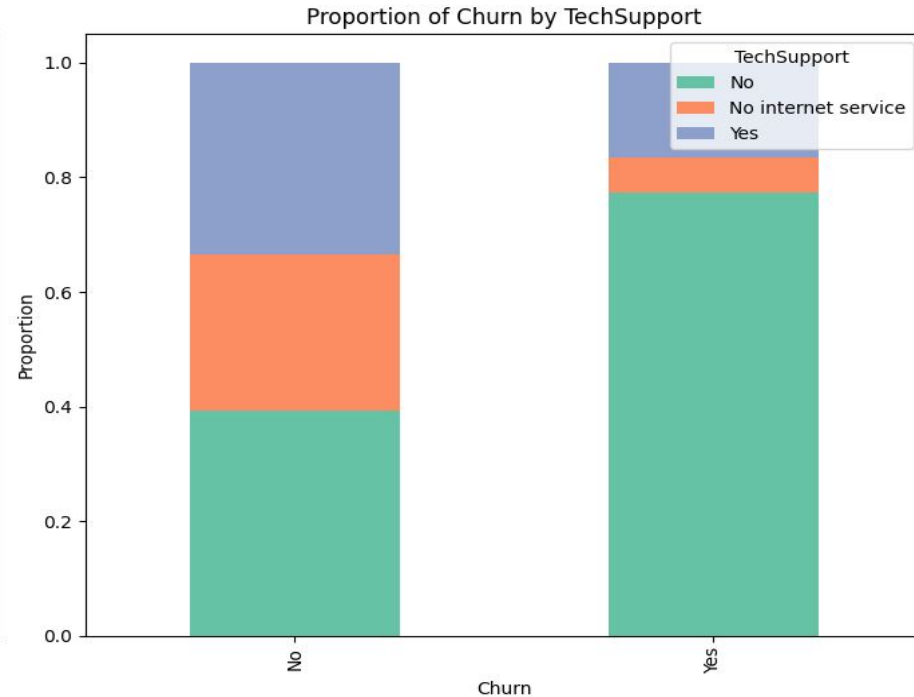
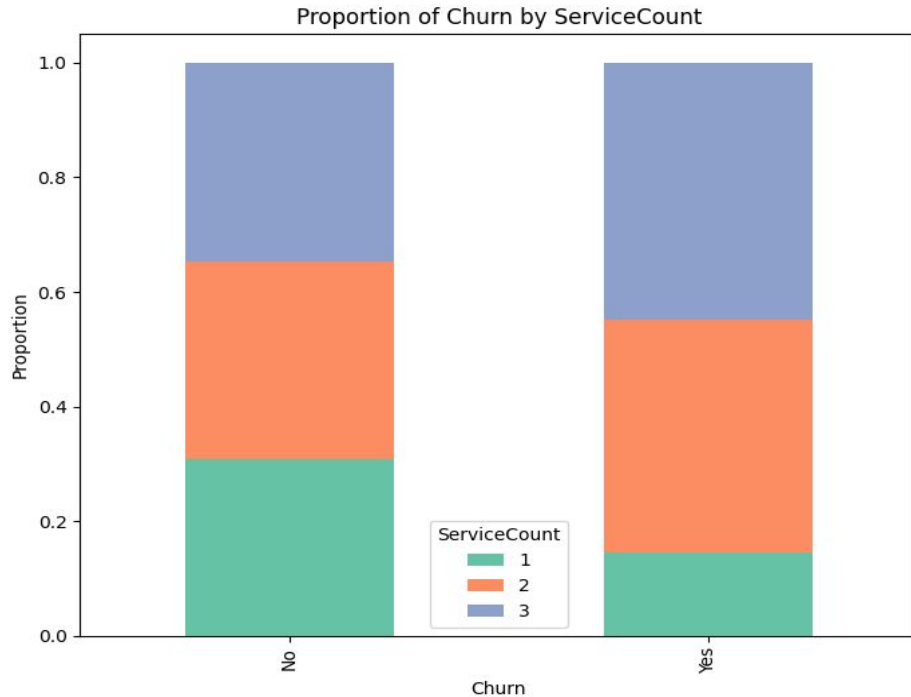
TotalCharges Distribution by Churn



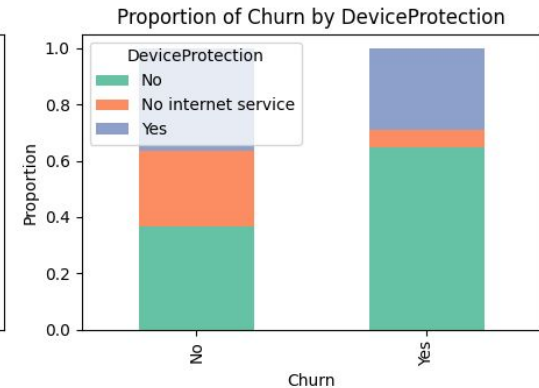
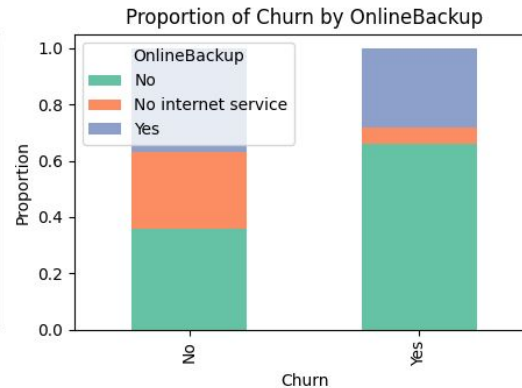
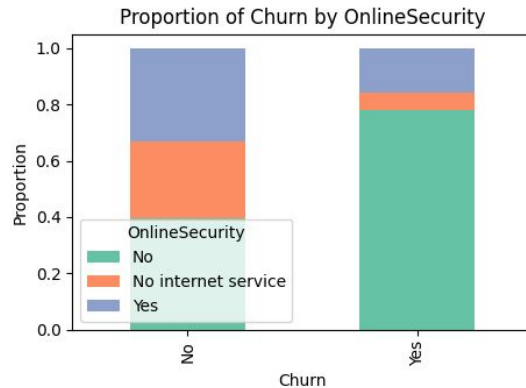
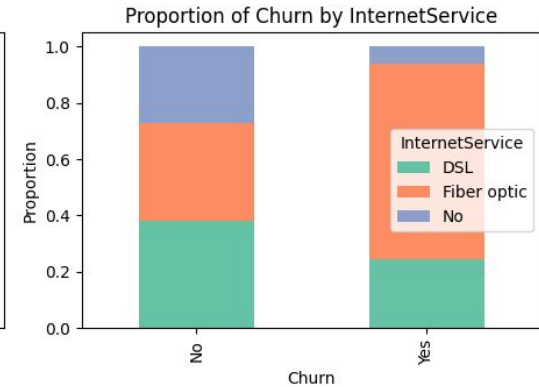
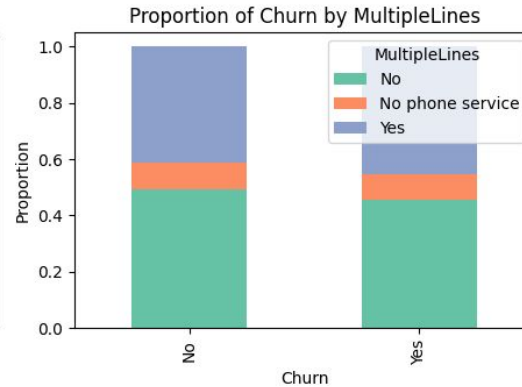
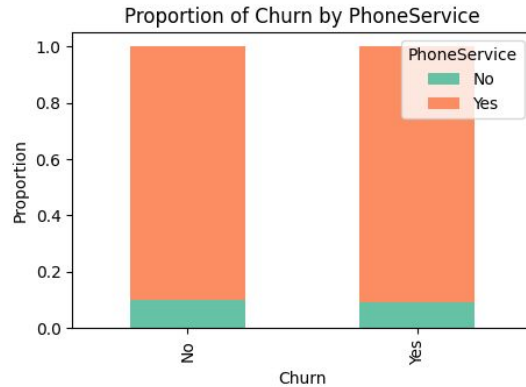
- Customers with longer tenures are less likely to churn



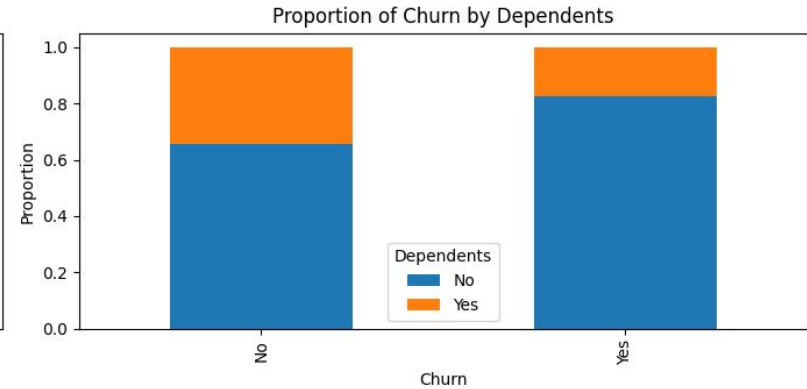
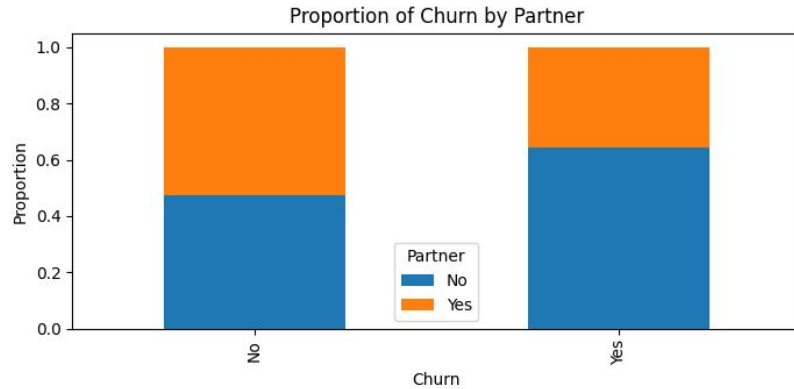
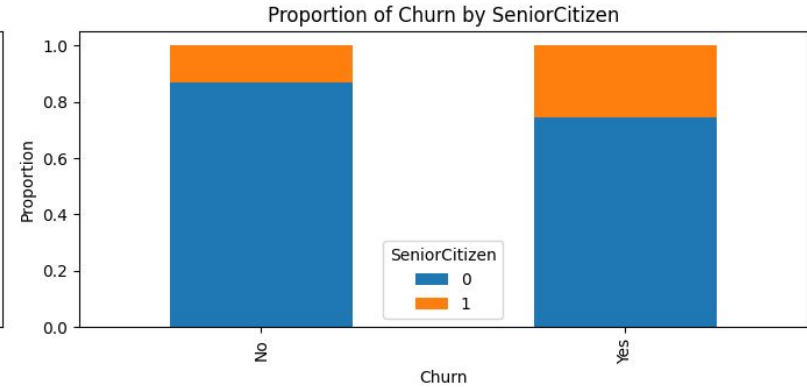
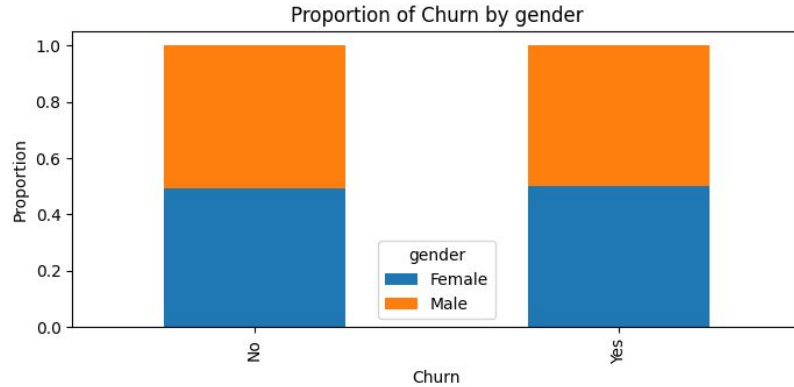
- Lack of tech support is strongly associated with churn.
- Customers with more services appear more likely to churn, contrary to what might be expected.



- Fiber optic customers are churning more - possibly due to price sensitivity or service issues;
- Supplementary services like online security, online backup, device protection, appear more important for retention



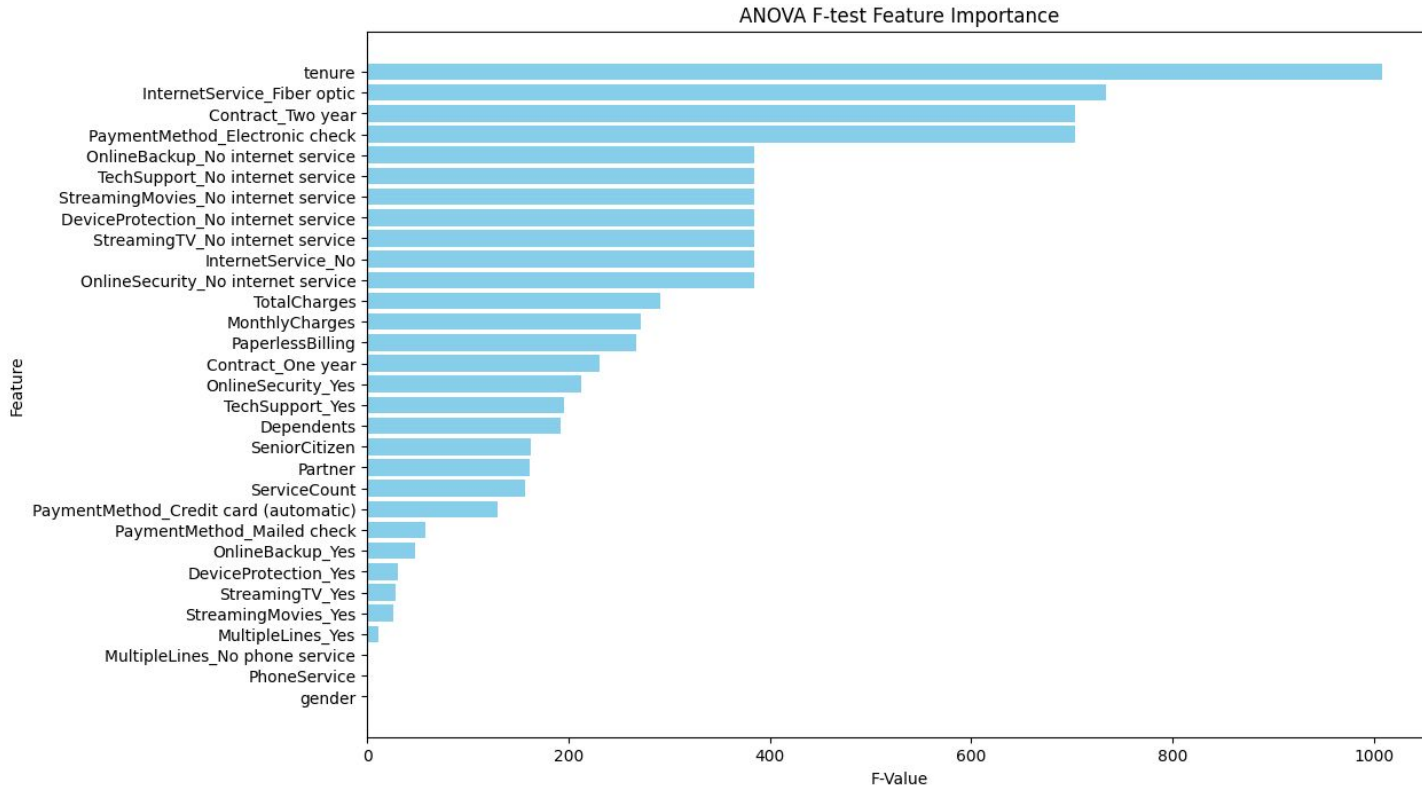
- Family-oriented customers (with partners/dependents) show more loyalty.



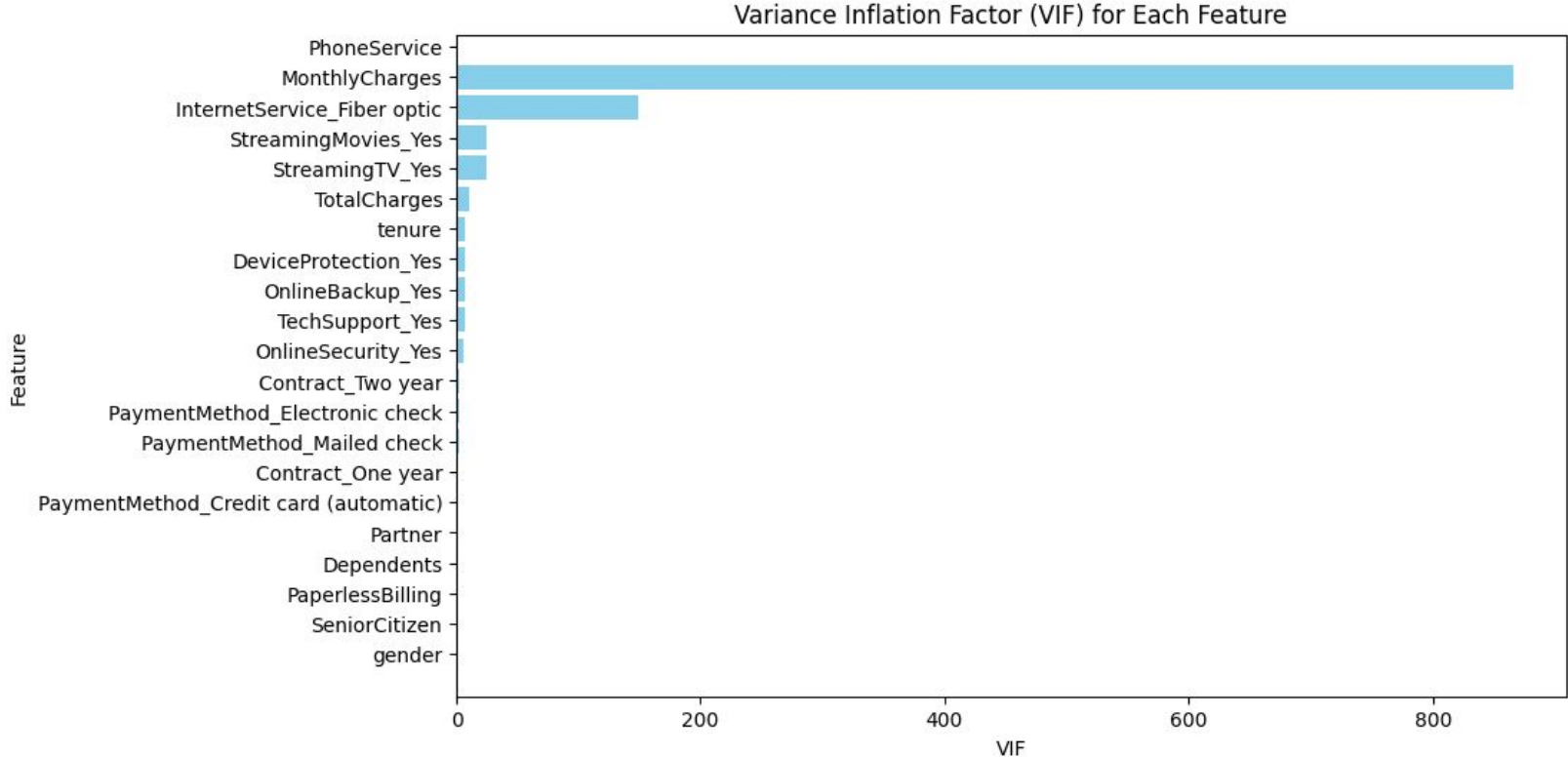
Feature Selection

Top Important Features:

Tenure, Contract_Two year, PaymentMethod_Electronic check



Consider Removing High Collinearity Variables: Phone Service, Monthly Charges



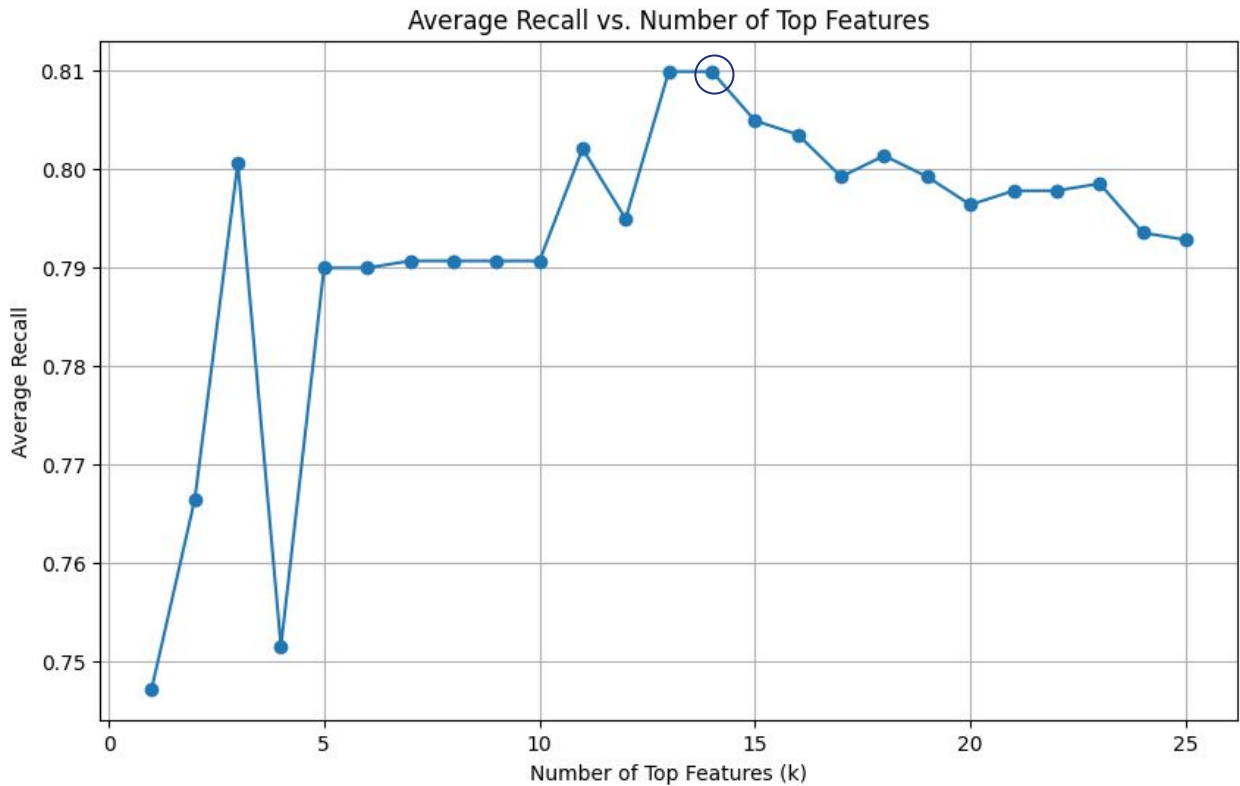
Model Comparison

Logistic Regression Shows Better Performance

	Random Forest	XGBoost	CatBoost	Logistic Regression
Recall	0.70	0.70	0.72	0.77
Accuracy	0.74	0.74	0.74	0.75

Recall is a metric that measures the model's ability to correctly identify all positive instances.

Model Enhance: K Important Features with Cross Validation

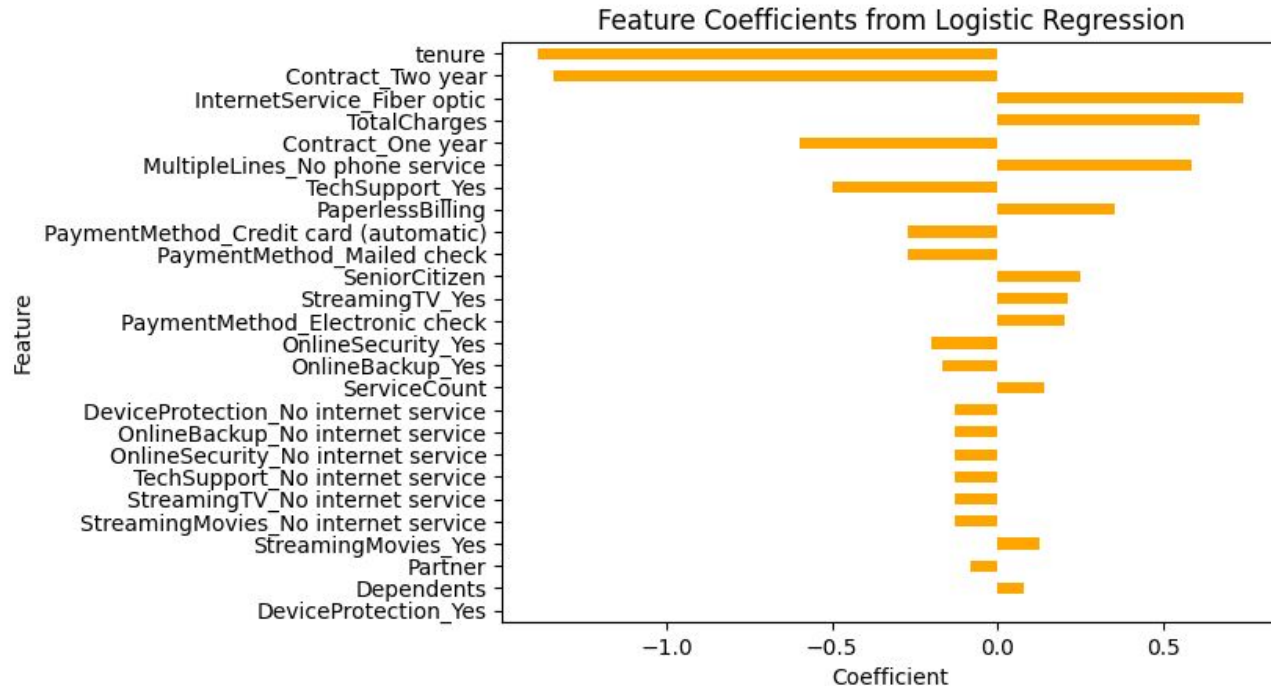


K=14	Random Forest
Recall	0.78
Accuracy	0.75

Given the negligible improvement in model performance when using only the top 14 features, we will proceed with our original model, which utilizes a more comprehensive feature set.

Customer Success Score

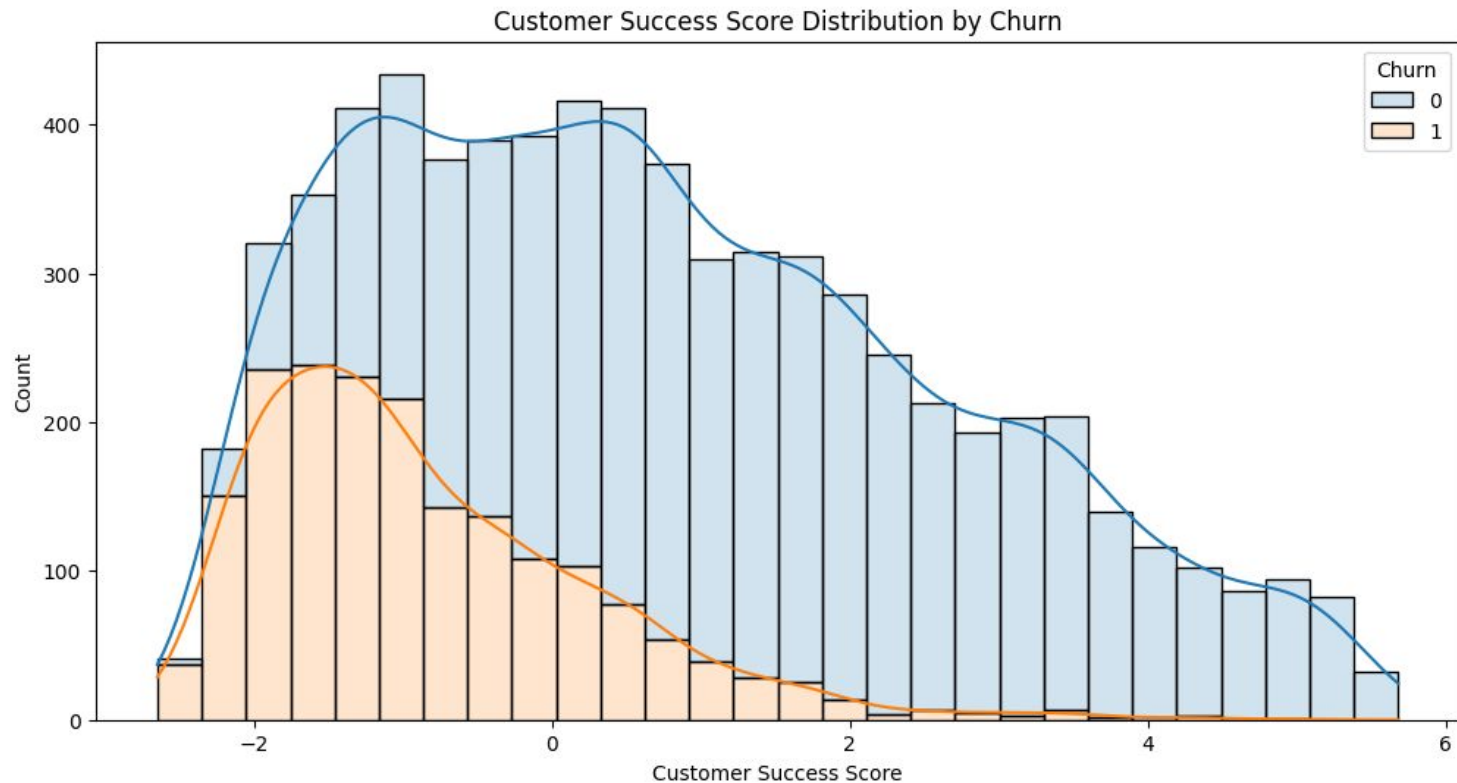
- Customers with long tenure, two-year contracts, and additional security/support services are less likely to churn.
- Customers with fiber optic internet, no phone service, and month-to-month contracts are more likely to churn; Higher Total Charges increases churn slightly.
- Payment method also plays a role, with electronic checks associated with higher churn.



Build Customer Success Score - Linear Regression

$$\text{Customer Success Score} = \text{intercept} + \sum(\text{customer features} \times \text{coefficients})$$

- Higher Customer Success Scores: 2~6
- Lower Customer Success Scores: -3~0
- Middle Region: 0-2



- High-Risk Group has more than half customers churned
- Other two groups have more moderate churn proportion
- The customer success score appears reasonable

Customer Segment	Score Group	Customer Count	Churn Proportion
High-Risk	$(-3, 0)$	2861	51.9%
Medium-Risk	$(0, 2)$	2357	14.8%
Low-Risk	$(2, 6)$	1814	1.9%

Thank You

