Objective: Predicting Customer Churn

Business Objective: Minimize customer churn by proactively identifying high-risk customers and implementing targeted retention interventions.

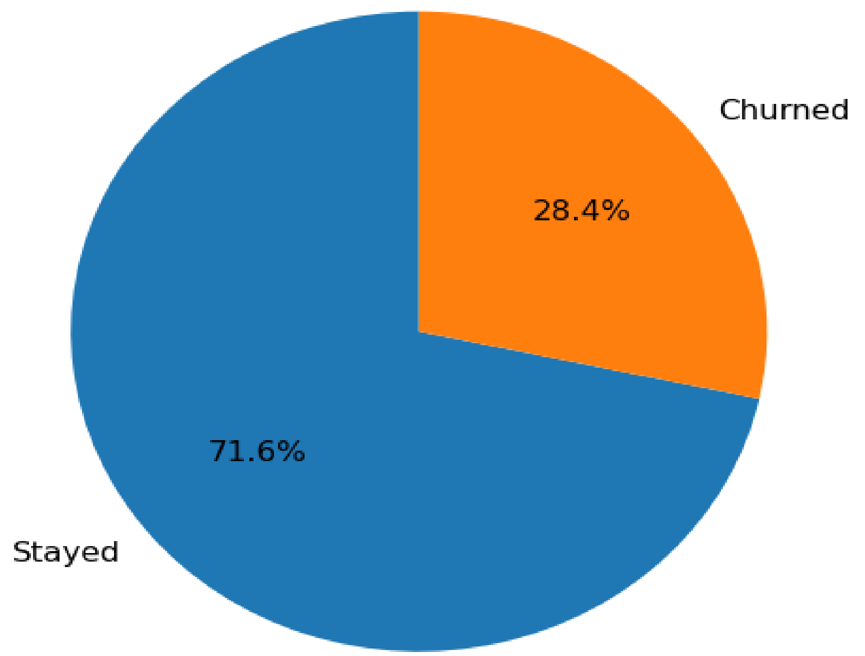
Analytical Objective: Develop a predictive model for churn (binary classification: Churned vs. Stayed) using customer behavioral and transactional data.

Strategic Value: Retaining existing customers is significantly more cost-efficient than acquiring new ones; leveraging predictive insights enables data-driven retention strategies, improves ROI, and strengthens customer lifetime value.

Data & Modeling Overview

- **Dataset:** Customer transactional and behavioral data (7,000+ customers).
- **Features Used:** Tenure in months, monthly spend, total charges, referrals, demographics, service types, payment methods, etc.
- **Preprocessing:**
  I. Handled missing values
  II. Encoded categorical variables
  III. Scaled/normalized numeric features where needed
- **Train/Test Split:** 70% training, 30% testing.
- **Class Imbalance Handling:** Weighted models to account for fewer churners.
- **Models Trained:** Logistic Regression, Random Forest, Gradient Boosting (XGBoost).

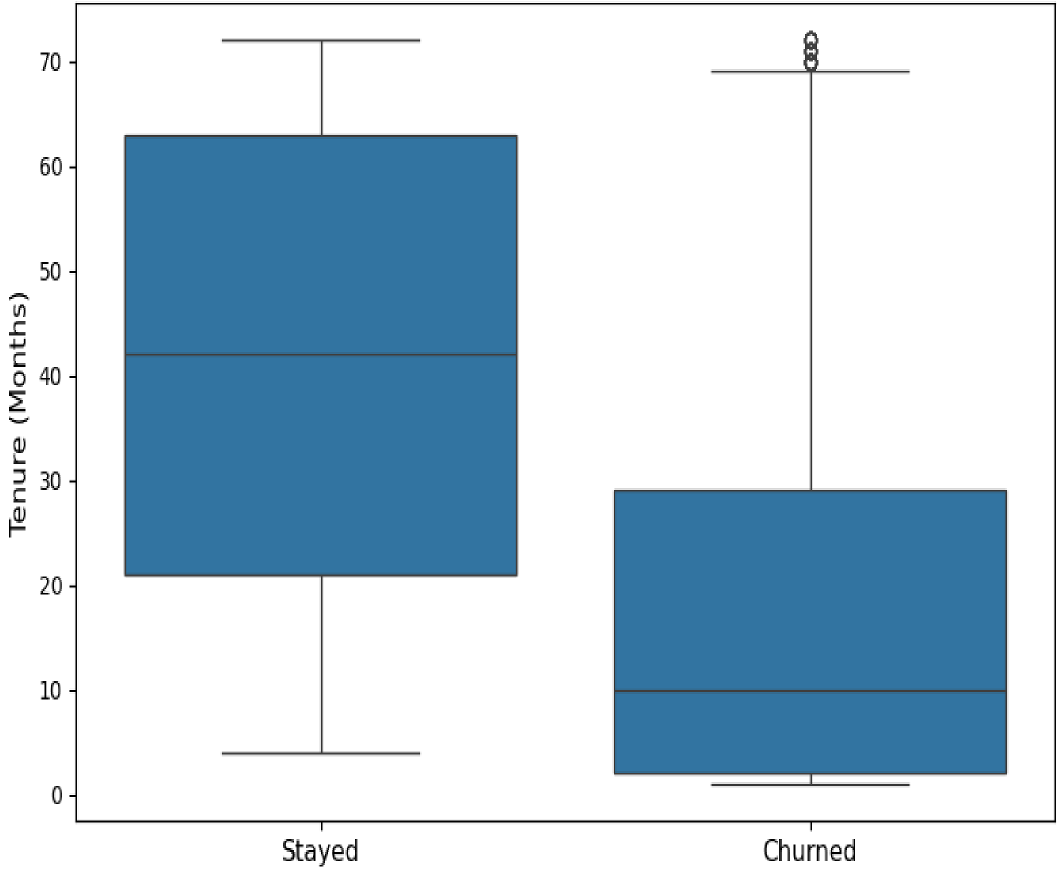
Customer Churn Distribution



- **Total Customers Analyzed**: **1,977** (Churners + Non-Churners).
- **Churn Rate**: 561 customers (**28%**) left AllLife.
- **Non-Churned Customers**: 1,416 customers (**72%**) remained.
- **Class Imbalance**: Majority are non-churners, fewer churners.
- **Implication**: Imbalance dataset ⟹ applied *scale_pos_weight* to improve model performance.

⟹ **Key Takeaway**: Roughly **3 out of 10** customers leave churn is a significant business risk that can not be ignored.

Impact of Tenure on Customer Churn



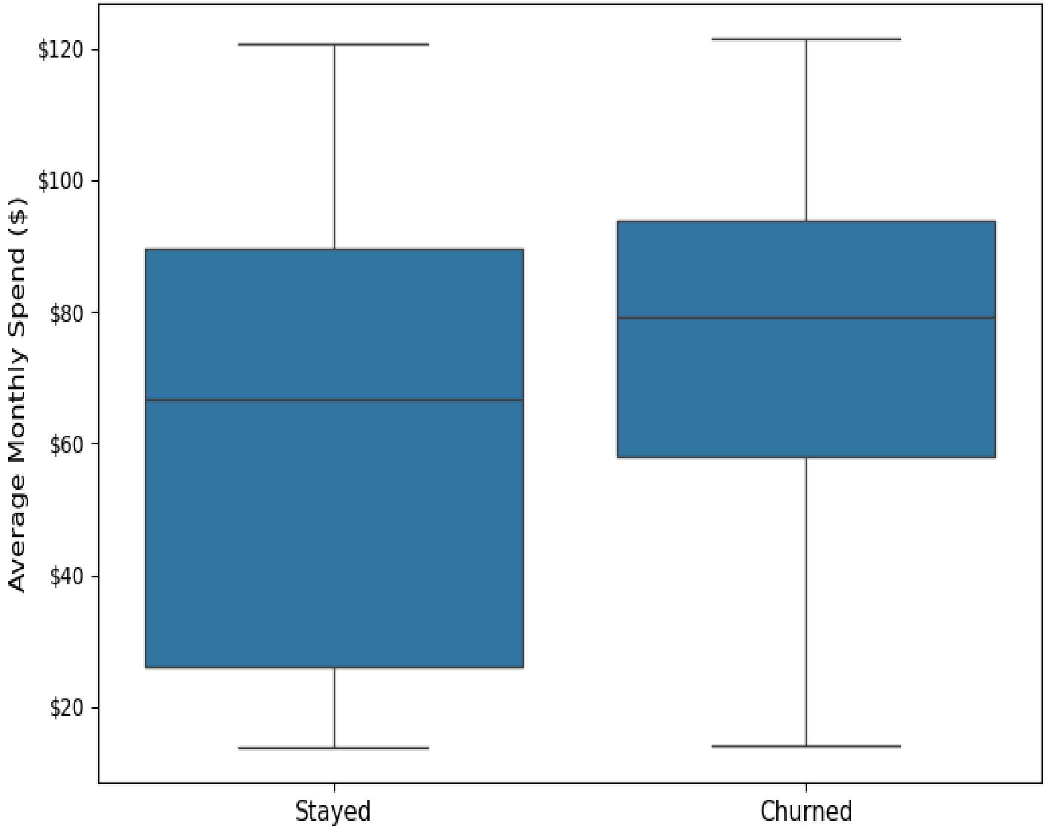
- ❑ **Stayed customers**: Median tenure **~42 months**; most remain **20-65** months.
- ❑ **Churned customers**: Median tenure **~10 months**; majority leave in the first **1-2** years.
- ❑ **Critical window**: First **12** months are the highest-risk period.

➡ **Key Takeaway**: Tenure is a strong predictor of churn.

➡ **Business action**: Focus retention campaigns on early-tenure customers (onboarding, discounts, offers, loyalty rewards).

AllLife

Impact of Monthly Spend on Customer Churn



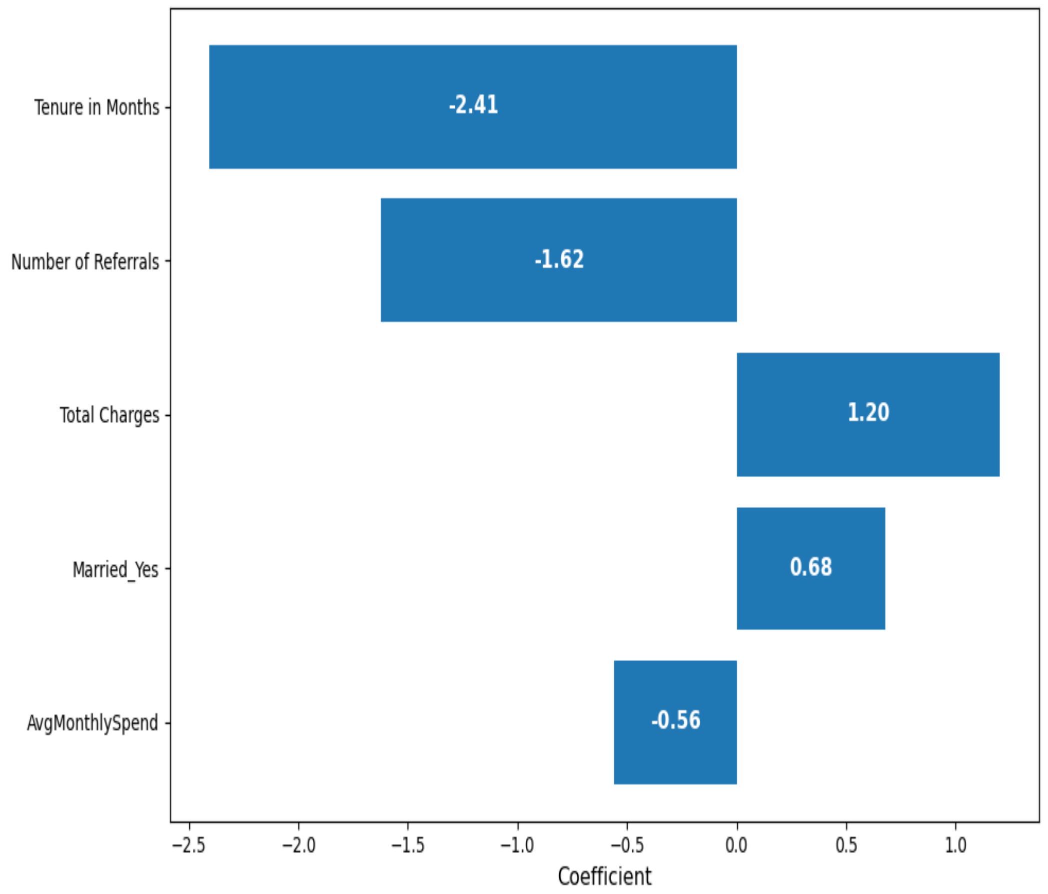
- **Stayed Customers**: Median monthly spend ~$65; most spend between **$25–$90**.
- **Churned Customers**: Median monthly spend ~$80; very high spenders **($60+)** are most likely to leave.
- **Critical Insights**: Higher monthly spend higher likelihood of churn.

➡ **Key Takeaway**: Monthly spend is a **strong** predictor for churn.

➡ **Business action**: Focus retention efforts on high-spending, early-tenure customers (personalized offers, loyalty rewards, VIP supports).

AllLife

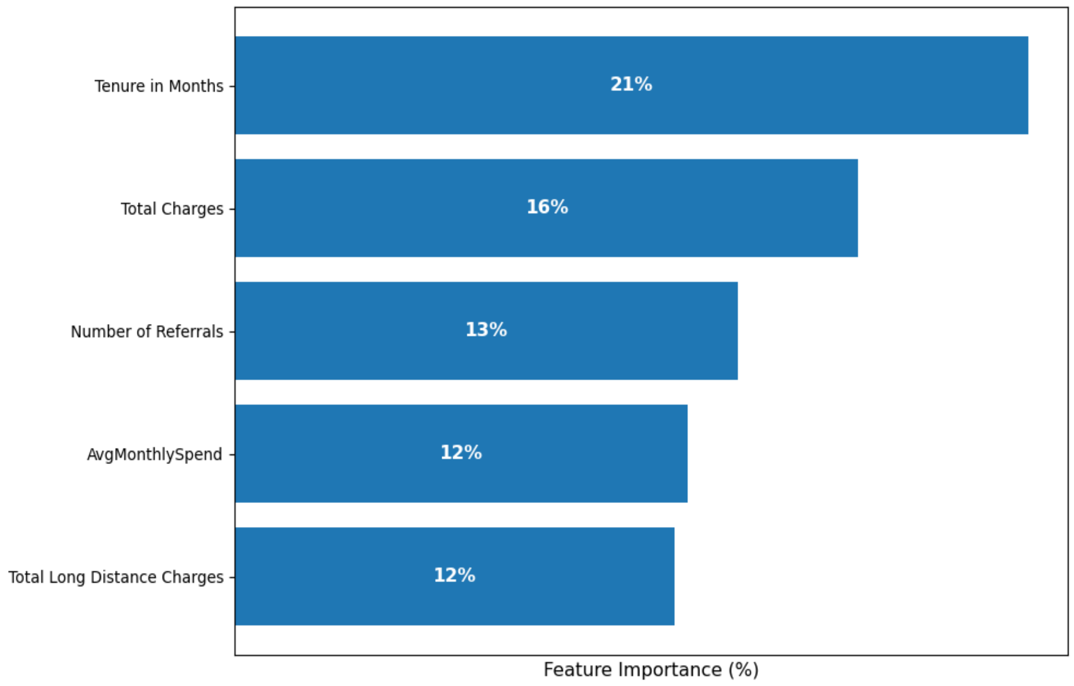
Top Predictors of Churn - Logistic Regression Model



- **Tenure**: Longer-tenure customers are much **less** likely to churn.
- **Number of Referrals**: Customers who provide more referrals have **lower** churn risk.
- **Total Charges**: **High** total charges strongly **increases** the likelihood of churn.
- **Other Key Drivers:** Features with **positive** coefficients indicate **higher** churn risk; **negative** coefficients indicate **retention** factors.
- **Critical Insights**: Both usage **patterns** and financial **metrics** significantly impact churn probability.

➡ **Key Takeaway**: Logistic regression quantifies which features most influence churn, helping target **high-risk** customers for retention offers.

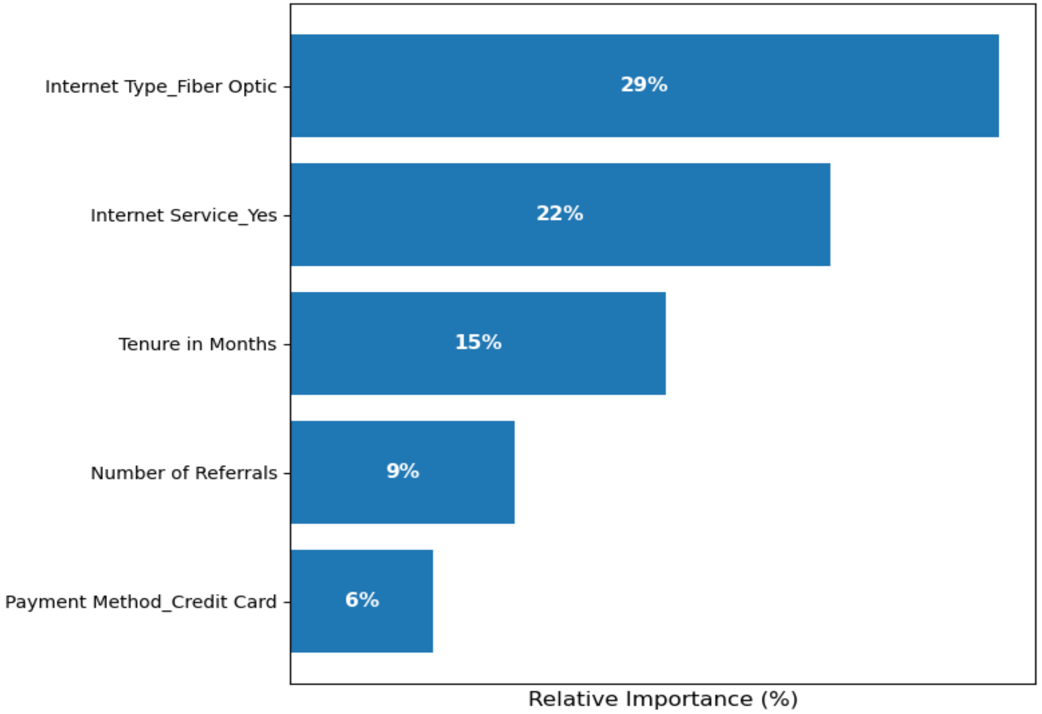
Key Drivers of Churn - Random Forest Model



- **Tenure in Months (21%)**: Shorter-tenure customers are much **more** likely to churn.
- **Total Charges (16%)**: Customer with **high** total spend show **higher** churn risk.
- **Number of Referrals (13%)**: Referrals strongly **reduce** churn risk.
- **Average Monthly Spend (12%): High** monthly bills **increase** churn likelihood.

→ **Key Takeaway**: Retention strategies should focus on **new, high-spending customers who are not yet engaged enough (e.g., no referrals).**

Top Predictors of Churn – Gradient Boosting

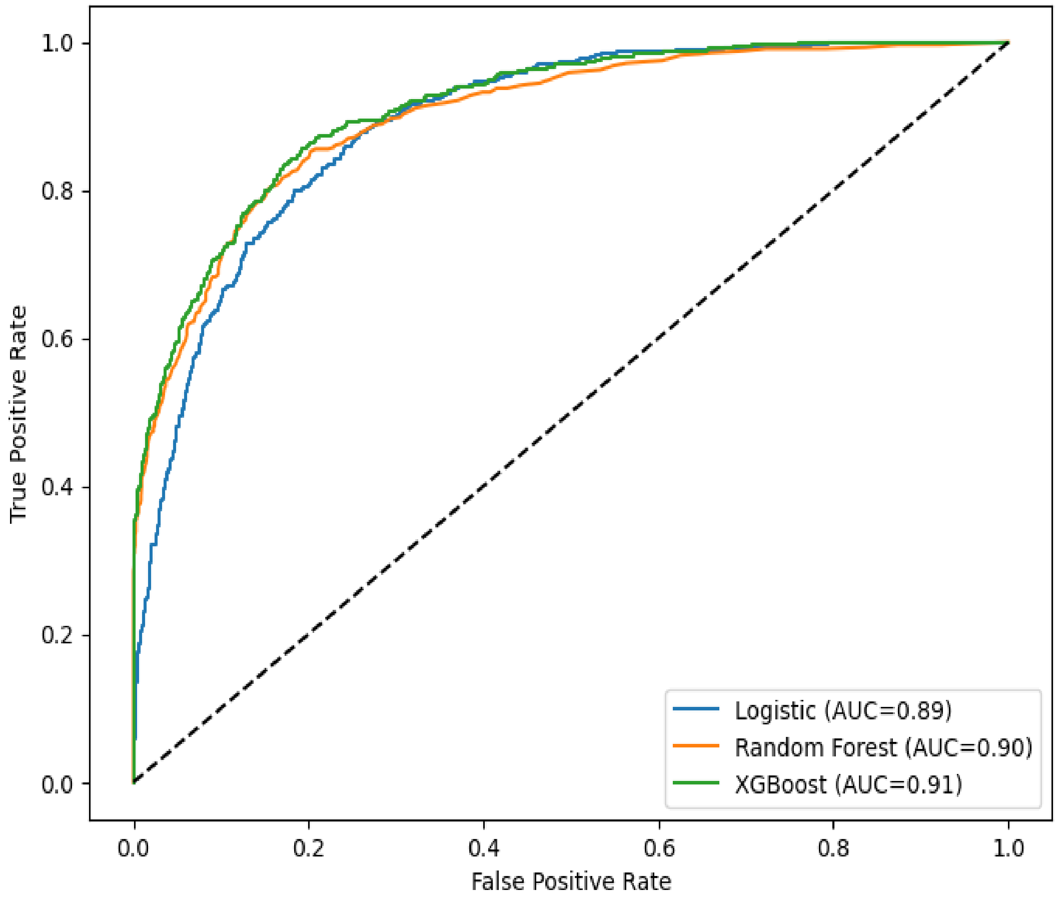


- **Top Drivers:** Fiber Optic internet (**29%**), Internet Service (**22%**), and Tenure (**15%**) are the strongest predictors of churn.
- **Tenure Effect:** Newer customers are more likely to churn.
- **Internet Plan Impact**: High-end internet plans (**Fiber Optic**) have higher churn probability.
- **Key Insight**: Both service type and customer behavior drive **churn** risk.

Key Takeaway: Retention efforts should target new, high-spending, or Fiber Optic customers.

AllLife

Model Performance Comparison (ROC Curves)

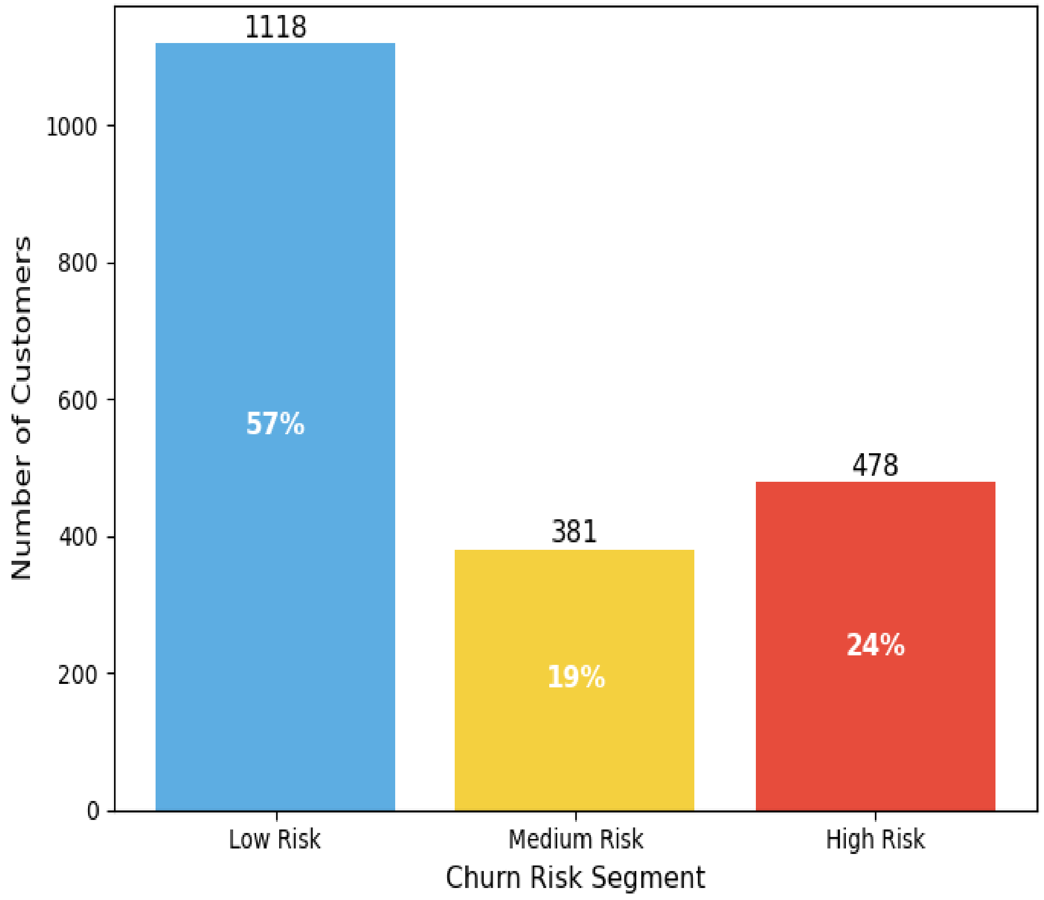


Model	Accuracy	Precision(Churn)	Recall(Churn)	F1-score(Churn)	ROC-AUC
Logistic Regression	0.83	0.72	0.65	0.68	0.890
Random Forest	0.84	0.81	0.58	0.68	0.901
XGBoost	0.83	0.67	0.81	0.73	0.912

- ❑ **Logistic Regression:** Interpretable, decent overall performance; balanced metrics.
- ❑ **Random Forest:** Higher precision (fewer false positives) but lower recall (misses some churners).
- ❑ **XGBoost:** Best recall (0.81) and highest ROC-AUC (0.912), **most effective** at identifying churners.
- ❑ **Business Implication:** XGBoost is recommended for retention campaigns targeting high-risk customers.
- ❑ **Key Insight:** Model choice depends on business goals here, catching as many churners as possible is most important.

AllLife

Customer Churn Risk Segmentation



- **Low Risk**: 57% of customers – unlikely to churn soon.
- **Medium Risk**: 19% of customers – moderate probability of churn.
- **High Risk:** 24% of customers – highest likelihood to churn; priority for retention campaigns.
- **Insight**: Churn probability is evenly distributed across segments, highlighting the need for targeted retention strategies.

Actionable Takeaway: Focus personalized offers, loyalty incentives, and proactive engagement on high-risk customers to reduce overall churn.

Model Choice: XGBoost provides reliable predicted probabilities for segmentation, enabling effective prioritization.

Improving Model Accuracy & Impact

- ❑ **Retraining:** update model periodically with new customer data to capture changing behaviors.
- ❑ **Feature engineering:** create derived features (e.g., usage trends, customer lifetime value, product interactions).
- ❑ **Hyperparameter tuning:** optimize XGBoost parameters for better recall and precision (like *max_depth*, *learning_rate, n_estimators, scale_pos_weight*).
- ❑ **Additional data sources:** incorporate external demographic, behavioral, or social data.
- ❑ **Performance monitoring:** track KPIs like churn rate reduction, retention ROI, and prediction accuracy over time.

Recommended Retention Strategy

- **Prioritize high-risk customers** with personalized retention campaigns/offers.
- **Engage medium-risk** customers to prevent escalation.
- **Maintain satisfaction for low-risk** customers to prevent future churn.
- **Leverage key drivers** (tenure in months, total charges, referrals, monthly spend) to design customized interventions.
- **Optimize marketing resources** by focusing efforts where churn likelihood is highest.

Key Takeaways & Next Steps

- ❑ **XGBoost** is the most effective model for identifying churners.
- ❑ **Top predictors**: Tenure in months, total charges, referrals, monthly spend.
- ❑ **Risk segmentation** allows prioritization of retention resources.
- ❑ **Continuous improvement**: Retrain model with new data and monitor performance.

➡ **Next Steps**: Integrate into marketing workflow, track KPIs, refine strategies based on results.

THANK YOU!