Customer Churn Prediction

🚀 Key Takeaways:

- 🚨 **Senior citizens, customers without partners or dependents, and those using electronic checks are more likely to churn**. Consider targeted retention strategies for these groups.
- 🔍 **DSL users and customers with additional services like online security and tech support are less likely to churn**. Encourage more customers to adopt these services.

Analysis

Data Overview 📂 & Quality Checks 🔍

The dataset provides a comprehensive look at customer behaviors and demographics, including:

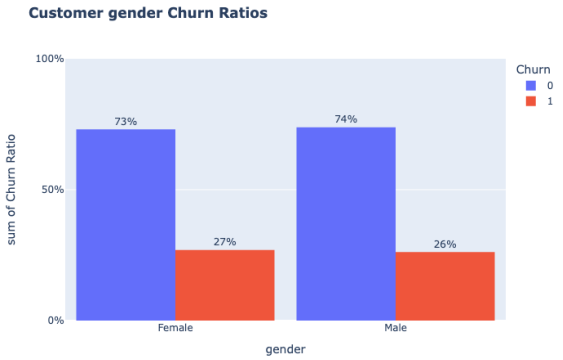
- 🏃 **Churn Information**: Identifies customers who left within the last month.
- 📋 **Service Subscriptions**: Details on phone, internet, and additional services.
- 💼 **Account Info**: Contract type, payment method, and billing details.
- 👨‍👧 **Demographics**: Gender, age, and family status.
- **Total Charges**: Null values were found where tenure is zero. This makes sense, as new customers haven't been billed yet.

Exploratory Data Analysis & Insights 📊

- **Churn Rate**: 26.6% of customers have churned.
- **Gender Balance**: The customer base is evenly split between females (49.5%) and males (50.5%).

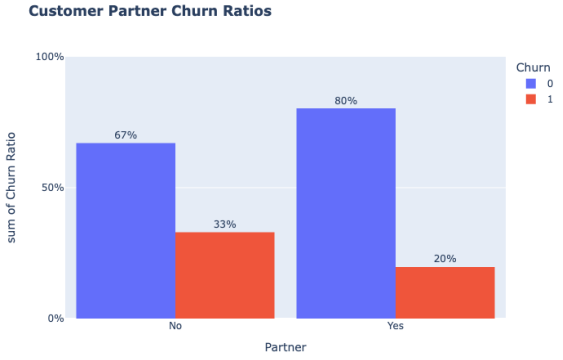
Churn Ratios by Categories

- **Gender**: Churn rates are similar for females (27%) and males (26%).

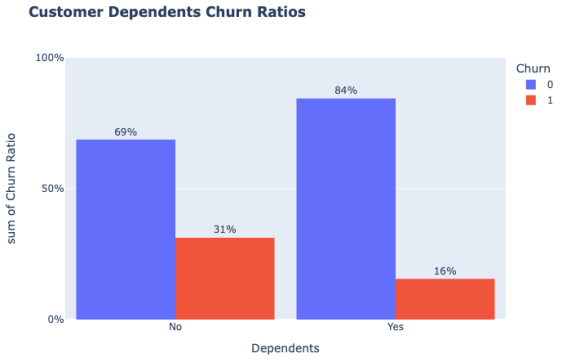


- **Senior Citizens**: **A significant 42% churn rate** compared to **24% for non-seniors**.
  - 💡 Implement targeted retention programs for senior citizens.

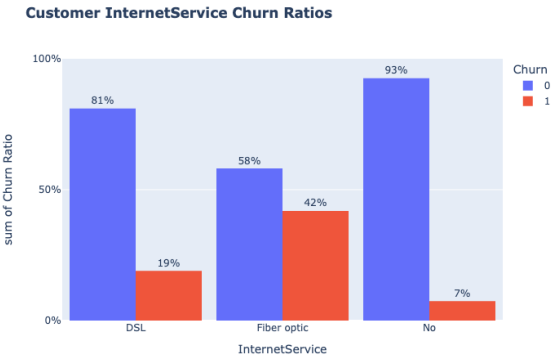
- **Partners**: **20% churn for those with partners** vs. **33% for those without.**
  - 💡 Encourage partner or family plans to reduce churn.



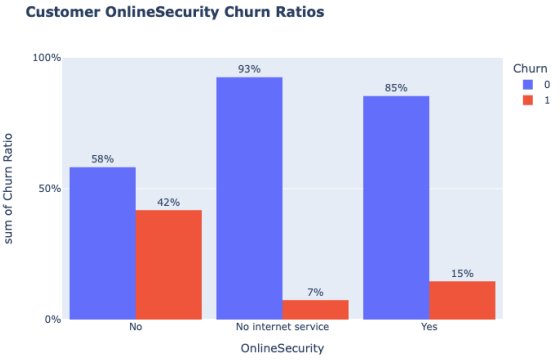
- **Dependents**: **Customers with dependents churn at 16%**, compared to **31% without**.
  - 💡 Promote family-friendly services to retain customers without dependents.



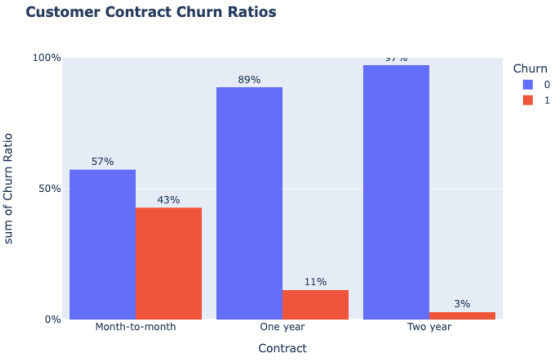
- **Internet Service**: **DSL users churn at 19%**, while **fiberoptic users churn at 42%.** 📶
  - 💡 Encourage customers to switch to DSL or enhance fiberoptic service satisfaction.



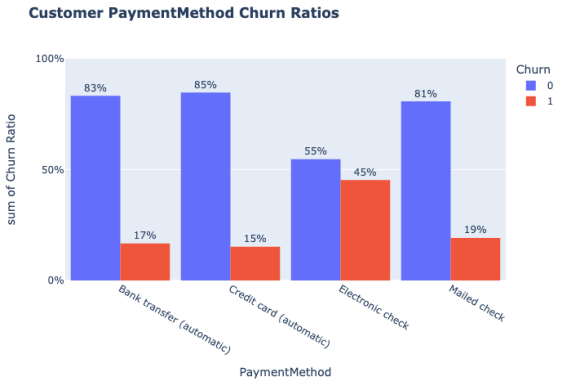
- **Online Security**: **Only 15% churn for users with online security,** but **42% for non-users**. 🔒
  - 💡 Try getting more customers to adopt online security.



- **Contracts**: **Month-to-month contracts see a 43% churn rate**, while **one-year and two-year contracts have much lower rates (11% and 3%, respectively).** 📆
  - 💡 Promote longer-term contracts with incentives.



- **Payment Method**: **Electronic check users have a 45% churn rate**, much higher than other methods.
    - 💡 Encourage switching to more stable payment methods like credit cards or bank transfers.



Machine Learning (ML) 🤖

Machine learning (ML) helps us predict customer churn by analysing patterns and trends in data. It's a **powerful tool for identifying at-risk customers and devising targeted retention strategies.**

- **Model Training**: The dataset was split into train, test, and holdout sets. Categorical columns were encoded, and numerical columns were standardised.

ML Performance

	ROC AUC	Accuracy	Precision	Recall
Gradient Boosting	0.856264	0.798720	0.651163	0.524064
Logistic Regression	0.853962	0.805832	0.662379	0.550802
Random Forest	0.832398	0.788051	0.627517	0.500000
XGBoost	0.824656	0.774538	0.587692	0.510695
Support Vector Machine	0.802051	0.793030	0.650909	0.478610
K-Nearest Neighbors	0.768020	0.751067	0.531746	0.537433

Top Models

Using ROC AUC score to decide on top performing model. This is a way to measure how good models are at distinguishing between customers who churn and those who don't. A score closer to 1 means the model is very good at

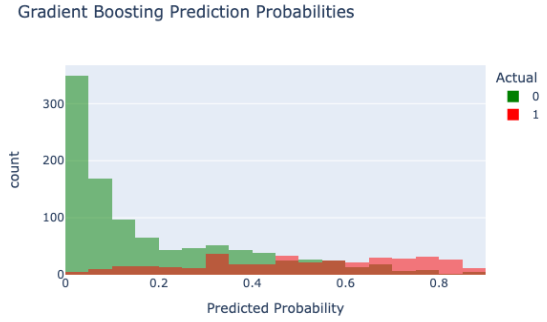
making this distinction. So, scores of 0.85 and 0.83 indicate strong performance.

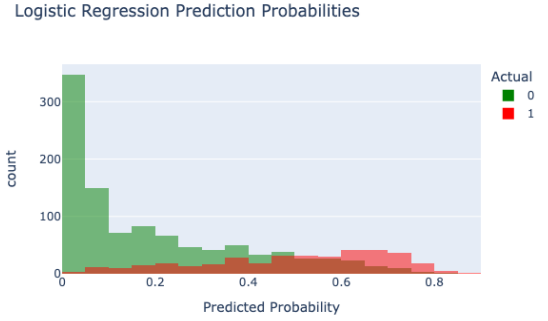
- **Gradient Boosting**: Think of this as a smart team of decision-makers. Each member learns from the mistakes of the previous one, gradually improving the team's decisions. It's like a relay race where each runner learns from the last to run faster and smoother. This method is great for handling complex patterns in data, and it achieved a score of 0.85 in our analysis.
- **Logistic Regression**: Imagine trying to draw a line that best separates two groups of dots on a graph. Logistic Regression does just that, using a mathematical approach to predict whether a customer will churn or not. It's straightforward and effective, scoring 0.83.

Visualising Predictions 📈

Below are visualizations that show how confident our models are in their predictions for each customer:

- **Prediction Probability Bars**: These charts display bars representing the likelihood (or probability) that each customer will churn (1) or not churn (0). The bars are colored based on the actual outcome—red for churned customers and green for non-churned ones.
- **Understanding Confusion**: By comparing the predicted probabilities to the actual outcomes, we can see where the model gets it right and where it might be confused. For example, if a red bar (churned) is low, but the actual outcome is churn, the model underestimated the risk. These insights help us refine the models to make more accurate predictions.





Recommendations and Next Steps 🎯

- **Targeted Retention**: Focus on high-risk groups identified in the analysis, such as senior citizens and electronic check users. Develop personalised offers or loyalty programs to retain these customers. Use the predictive model to proactively target and engage these at-risk customers.
- **Service Enhancements**: Encourage customers to switch to DSL and adopt additional services like online security and tech support, which have been shown to reduce churn rates significantly. Tailor marketing campaigns to highlight the benefits of these services.
- **Contract Strategies**: Promote longer-term contracts to stabilise the customer base and reduce churn. Highlight the benefits and savings of committing to longer terms. Consider offering incentives for customers who switch from month-to-month to annual contracts.
- **Payment Method Optimisation**: Encourage customers to switch from electronic checks to more stable payment methods like credit cards or bank transfers, which are associated with lower churn rates. Provide easy transition options and benefits for switching.
- **Further Analysis**: Continue refining models and explore additional features or data sources to improve predictions. Consider analyzing customer feedback or behavioural data for deeper insights. Regularly update the model to adapt to changing customer behaviours.
- **Model Deployment and Monitoring**: Deploy the predictive model into production to continuously monitor customer behaviour and identify those at risk of churning. Implement a feedback loop to capture model performance and make iterative improvements. This ongoing process ensures the model remains effective and aligned with business goals.