Bivariate Analysis

1. Demographic Features

- Gender ($Chi^2 = 0.48$, $p = 0.486$): No significant relationship with churn
- Senior Citizen ($Chi^2 = 159.43$, $p < 0.001$): Senior citizens show higher churn rates
- Partner ($Chi^2 = 158.73$, $p < 0.001$): Customers with partners show lower churn rates
- Dependents ($Chi^2 = 189.13$, $p < 0.001$): Customers with dependents have significantly lower churn rates

Strategic Insight: Family-oriented customers are more stable, suggesting family plans or bundled services could improve retention.

2.Account Info Features

- Contract Type ($Chi^2 = 1,184.60$, $p < 0.001$)
  Strongest predictor of churn:
  - Month-to-month: Very high churn rate (~43%)
  - One-year: Moderate churn rate (~11%)
  - Two-year: Lowest churn rate (~3%)
- Payment ($Chi^2 = 648.14$, $p < 0.001$)
  - Electronic check: Highest churn risk payment method
  - Automatic payment methods: Lower churn rates
- Paperless billing ($Chi^2 = 258.28$, $p < 0.001$):
  Customers with paperless billing show higher churn rates

Strategic Insight: Contract length is the most powerful predictor of churn. Incentivizing longer contracts could dramatically improve retention.

3.Service Features

- Internet Service ($Chi^2 = 732.31$, $p < 0.001$)
  - DSL customers: Lowest churn rate (~20%)
  - Fiber optic customers: Moderate churn rate (~42%)
  - No internet service: Extremely low churn rate (~7%)

Strategic Insight: Fiber optic customers represent the highest churn risk segment, requiring targeted retention strategies.

- Online Services Impact
    - OnlineSecurity (Chi^2 = 850.00, $p < 0.001$): Customers without online security show dramatically higher churn rates
    - Online Backup (Chi^2 = 601.81, $p < 0.001$): Similar pattern - lack of backup services correlates with higher churn
    - Tech Support (Chi^2 = 828.20, $p < 0.001$): Absence of tech support strongly predicts churn

Strategic Insight: Value-added services act as retention tools. Customers without these services are significantly more likely to churn.

- Entertainment Services
    - Streaming TV (Chi^2 = 374.20, $p < 0.001$): Moderate impact on churn
    - Streaming Movies (Chi^2 = 375.66, $p < 0.001$): Similar moderate impact

    Both services show higher retention for customers who subscribe

Correlation Matrix Analysis

- Tenure and Churn (-0.35) - longer-tenured customers are less likely to churn
- Total Charges and Churn (-0.20) - customers with higher total charges are less likely to churn

Insights: Customer retention improves with tenure, suggesting importance of early customer satisfaction. Monthly charges show weak correlation with churn, indicating price sensitivity may not be the primary driver

Feature Engineering Success

custom feature engineering revealed two powerful predictive indicators:

- Services Score (0-4 scale):Chi^2 = 338.97
    - Measures customer engagement with add-on services
    - Distribution:40% of customers have 0 services, only 7% have all 4 services
    - Business Impact: Higher service adoption correlates with lower churn risk

- Vulnerability Score (0-8 scale):Chi^2 = 1307.20,
  Composite risk metric based on customer demographics and contract type
  Risk Categories:
    - ❖ 0-1: Low vulnerability (married, long-term contracts)
    - ❖ 2-4: Moderate vulnerability
    - ❖ 5-8: High vulnerability (seniors, single, month-to-month contracts)

Model Performance & Reliability

Model Comparison Results:

- Random Forest: Robust performance with excellent interpretability

- XGBoost: Superior predictive accuracy with gradient boosting

- CatBoost:Strong performance on categorical features

Key Performance Metrics:

- F1score: Random forest(0.78), XGBoost(0.79), CatBoost(0.80)

- ROC-AUC: 0.85+ indicating strong predictive power

- Feature Importance: Consistent ranking across models validates findings

Most Important Predictive Features:

1. Contract type and tenure

2. Monthly charges and total charges

3. Services Score and Vulnerability Score

4. Internet service type

5. Payment method

Recommendations

- Contract migration campaigns for month-to-month customers
- Enhanced onboarding for new customers
- Service bundling optimization
- Senior customer support programs
- Real-time risk scoring implementation