CUSTOMERS CHURN ANALYSIS WITH PYTHON

Uncovering Drivers of Customer Retention and Reducing Churn through Data Insights

OVERVIEW & OBJECTIVES

DATASET:
7,043 customer records from multiple sources (Customer, Internet, Churn data).

OBJECTIVE:
Understand churn patterns and identify factors driving customer retention or loss.

APPROACH:
Data Cleaning → Exploratory Data Analysis (EDA) → Insights → Recommendations.

PROJECT GOAL

MEASURE CUSTOMER CHURN RATE

Establish the overall percentage of customers leaving the service, providing a baseline metric to evaluate the business health.

SEGMENT AT-RISK CUSTOMERS

Pinpoint groups most vulnerable to churn, such as new customers, month-to-month subscribers, or high-bill customers.

IDENTIFY KEY DRIVERS OF CHURN

Examine demographic, contractual, and service-related variables to determine which factors most influence customer decisions.

SUPPORT STRATEGIC DECISION-MAKING

Translate insights into practical strategies for marketing, customer service, and retention programs.

DEVELOP ACTIONABLE RECOMMENDATIONS

Suggest targeted actions like loyalty programs, bundled services, and contract incentives to retain customers.

BUSINESS REQUIREMENTS

1. IDENTIFY AT-RISK CUSTOMERS

The business must be able to detect customers who are most likely to churn based on patterns in service usage, contract type, and payment methods. Early detection enables proactive interventions.

2. UNDERSTAND KEY INFLUENCING FACTORS

Management needs clear visibility into the major factors driving churn, such as high monthly charges, lack of value-added services (e.g., Tech Support), or short contract commitments. This understanding helps in targeting the right pain points.

3. ENABLE DATA-DRIVEN RETENTION STRATEGIES

The business requires insights that directly translate into practical strategies. These include offering better incentives for long-term contracts, promoting bundled services, and tailoring loyalty rewards to high-risk segments.

TOOLS & TECHNOLOGIES USED

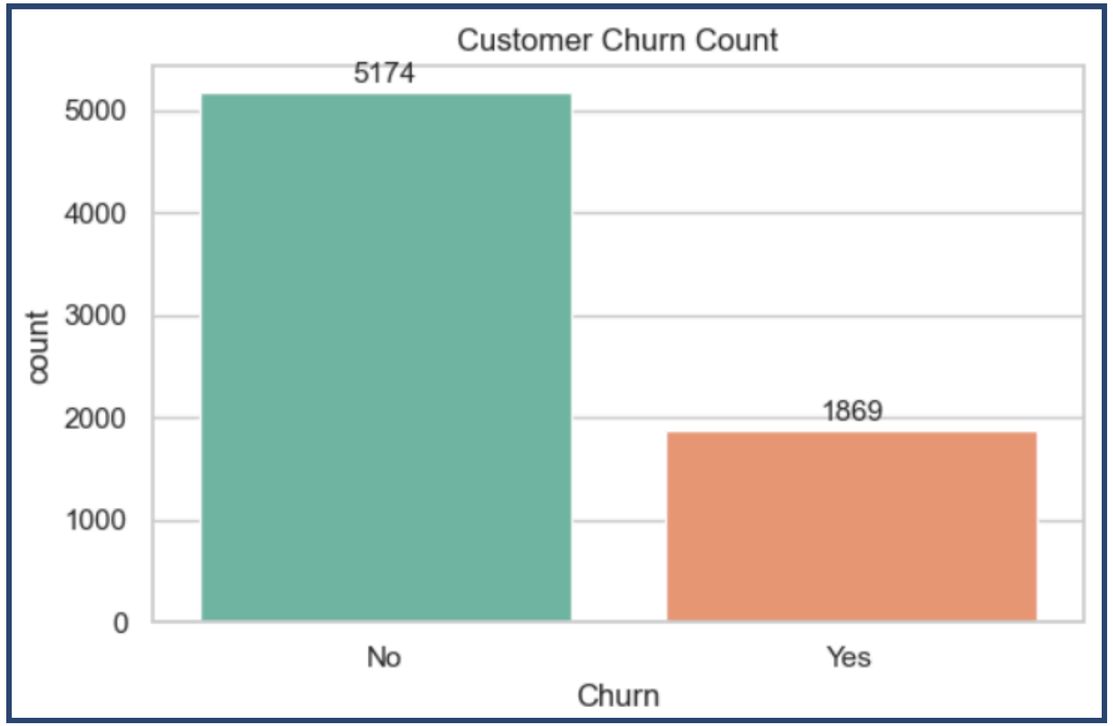
- **Programming Language: Python**
- **Environment: Jupyter Notebook**
- **Libraries:**
  - Pandas, NumPy (Data handling)
  - Matplotlib, Seaborn (Visualization)
  - Warnings filter for clean outputs

EDA - UNIVARIATE ANALYSIS

INSIGHTS:

Most customers did not churn; churners are a minority in the dataset.

- The bar for “No” is much taller, indicating far more retained customers than churned ones.
- Approximate counts show around 5.1k non-churn vs 1.9k churn, so retention is roughly 3 times higher than churn.
- Focus areas should prioritize understanding why the 1.9k left and replicating retention factors present in the larger non-churn group.

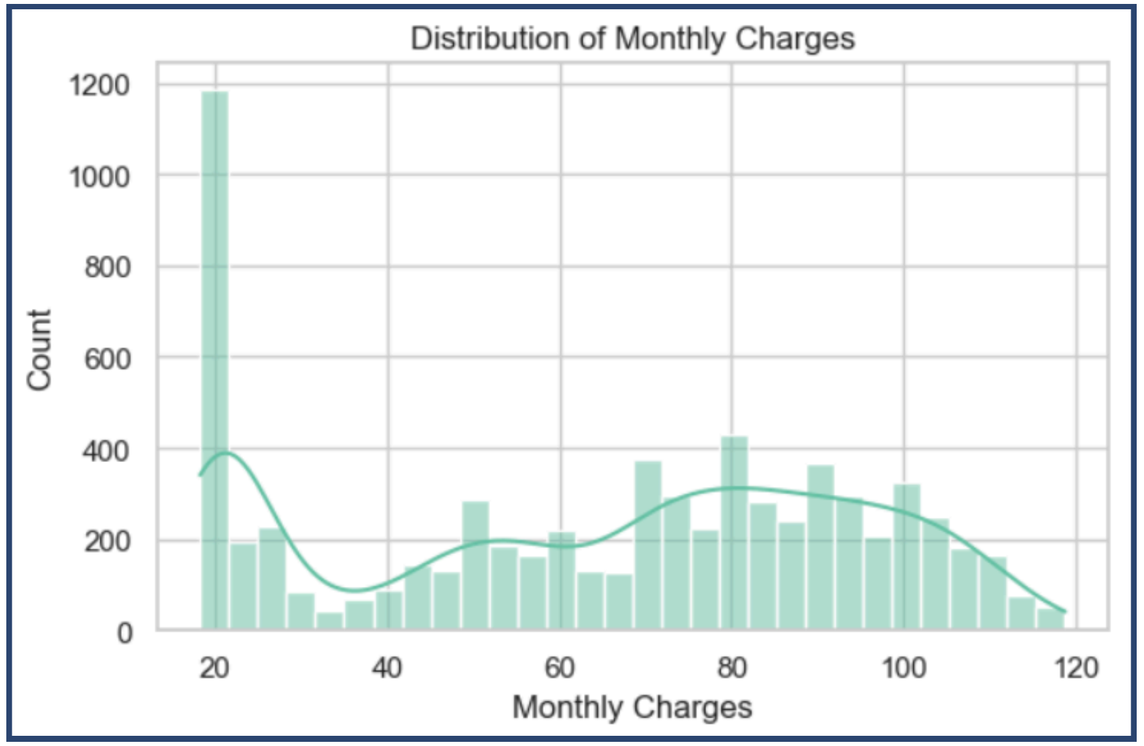


EDA - UNIVARIATE ANALYSIS

INSIGHTS:

Most customers are clustered around lower-to-mid monthly charges, with a sharp spike near the low end around ₹20–₹25, then a broad hump between roughly ₹60–₹100.

- The distribution is right-skewed: many low-charge customers, fewer very high-charge ones toward ₹110–₹120.
- A prominent peak at the low-charge bin suggests a basic plan is very popular.
- Widest density lies in mid-tier charges (~₹70–₹95), indicating substantial adoption of standard packages.
- High-charge customers exist but taper off, implying premium plans are less common.

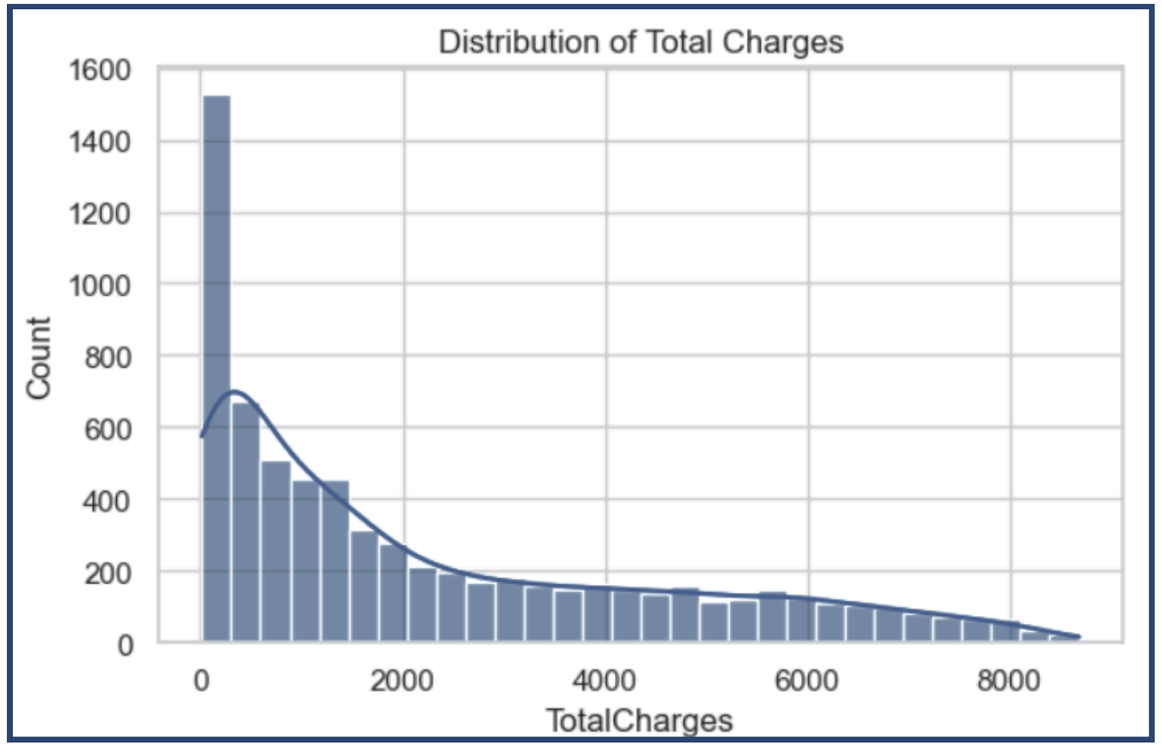


EDA - UNIVARIATE ANALYSIS

INSIGHTS:

Most customers have low total charges, with counts dropping steadily as totals increase.

- The distribution is strongly right-skewed: many customers near the lower totals, few with very high totals.
- A clear peak occurs at the smallest total charges, then a long tail extends beyond 8000, indicating a small segment of long-tenure or heavy-usage customers.
- This suggests customer base is dominated by newer or lighter users, while high-total customers are rare but potentially high value.

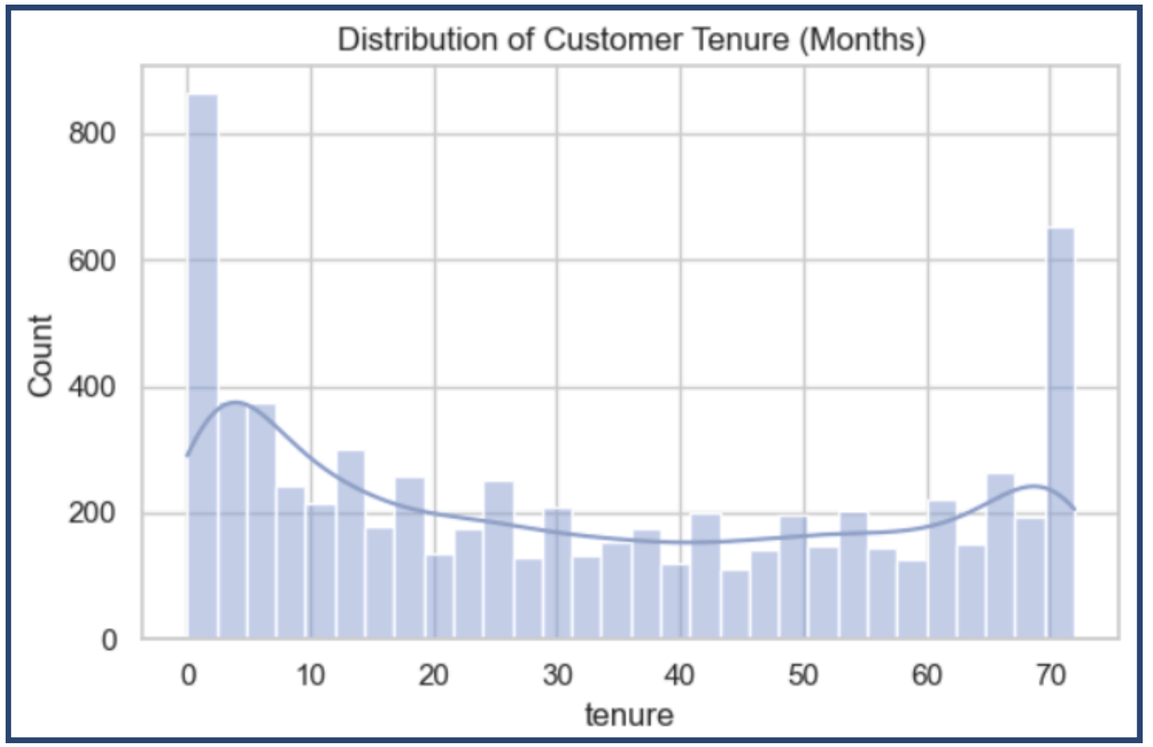


EDA - UNIVARIATE ANALYSIS

INSIGHTS:

Customer tenure is concentrated at the very beginning and end of the range, with many new customers and a noticeable spike near 70–72 months.

- The distribution is U-shaped: high counts at 0–2 months and again at long tenures, with fewer customers in the mid-tenure range.
- Early spike suggests frequent recent acquisitions, while the late spike indicates a loyal segment staying 6+ years.
- Mid-range dips imply higher churn risk or fewer conversions persisting through the middle months.

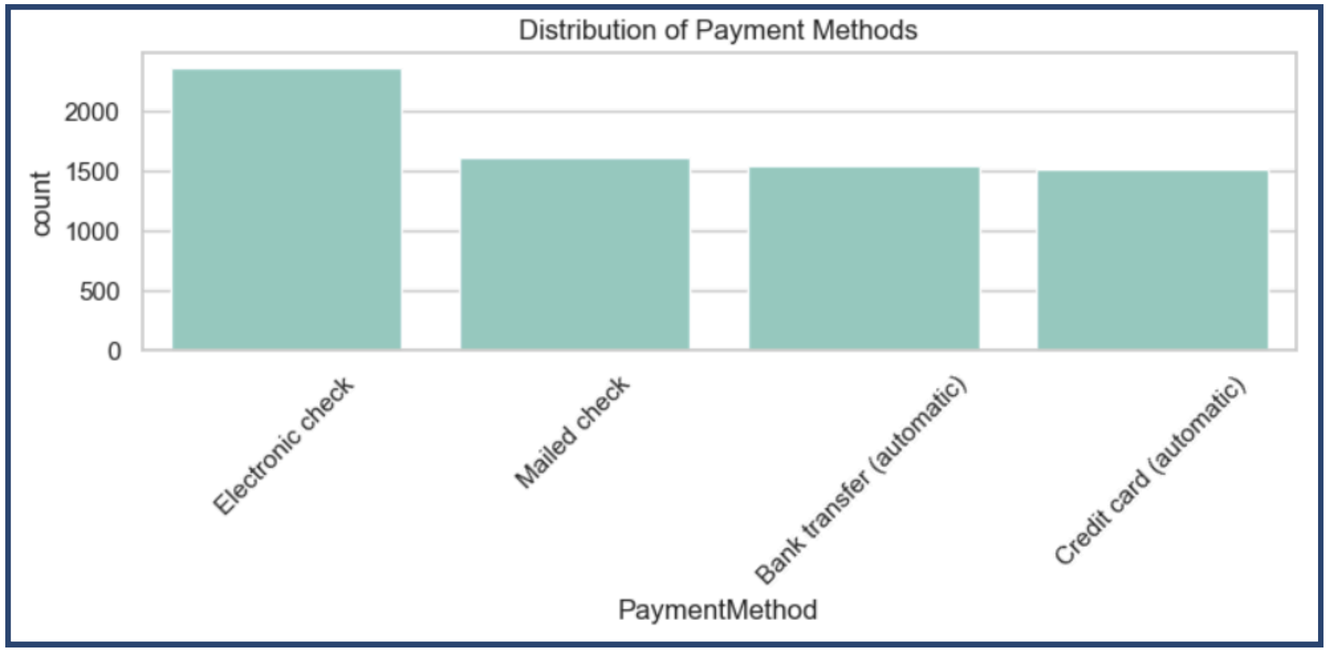


EDA - UNIVARIATE ANALYSIS

INSIGHTS:

Electronic check is the most used payment method, while the three others have similar and lower adoption.

- Electronic check leads by a clear margin, indicating strong preference for quick digital payments without cards.
- Mailed check, bank transfer (automatic), and credit card (automatic) are each moderately used and relatively close in count.
- Encouraging automatic methods could reduce late payments and improve retention versus manual checks.

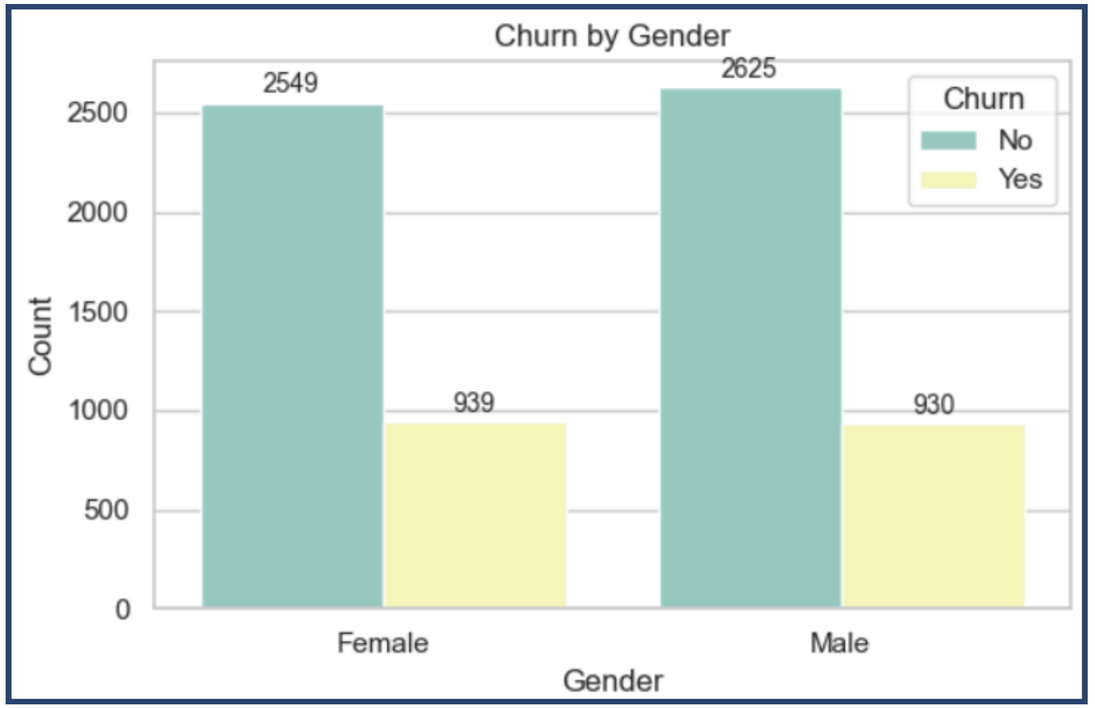


EDA - BIVARIATE ANALYSIS

INSIGHTS:

Churn rates look similar for males and females, with both groups showing many more retained customers than churned ones.

- Female: ~2549 no-churn vs ~939 churn, indicating higher retention than churn.
- Male: ~2625 no-churn vs ~930 churn, very close to females and also retention-dominant.
- Gender does not appear to be a strong driver of churn differences in this dataset.

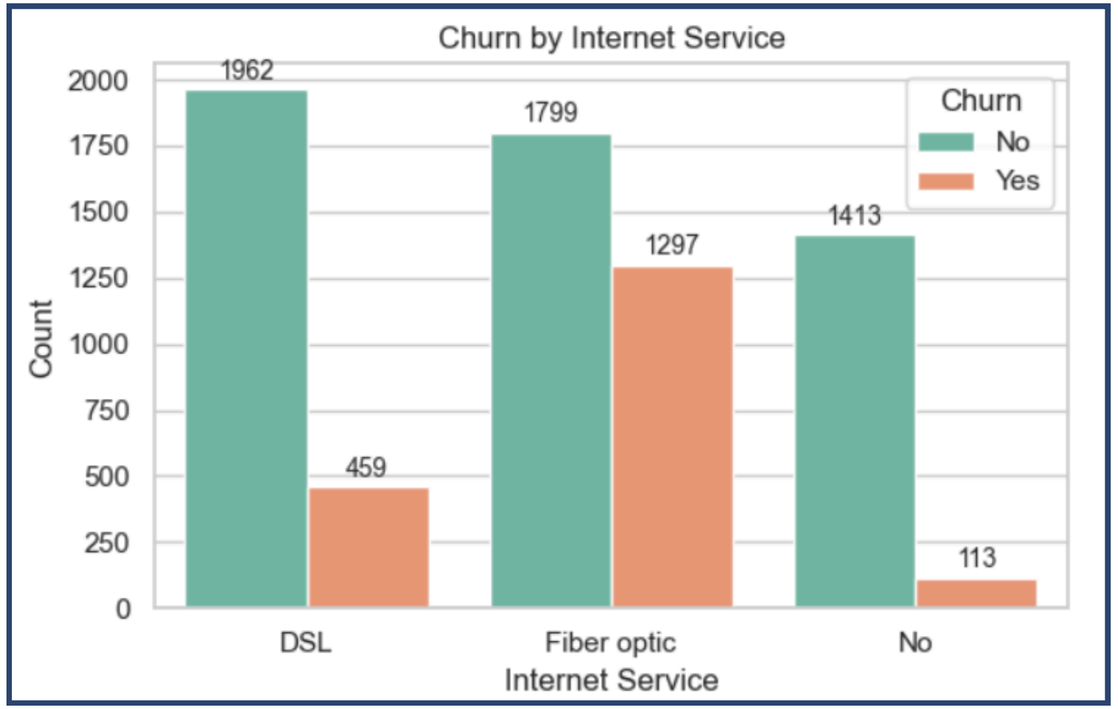


EDA - BIVARIATE ANALYSIS

INSIGHTS:

Fiber optic users show the highest churn counts, while DSL users churn less and non-internet users churn the least.

- DSL: Many more retained (1962) than churned (459), indicating relatively stable customers.
- Fiber optic: Retained (1799) and churned (1297) are closer, signaling higher churn risk in this segment.
- No internet: Very low churn (113) with strong retention (1413), likely due to fewer service touchpoints or different product mix.

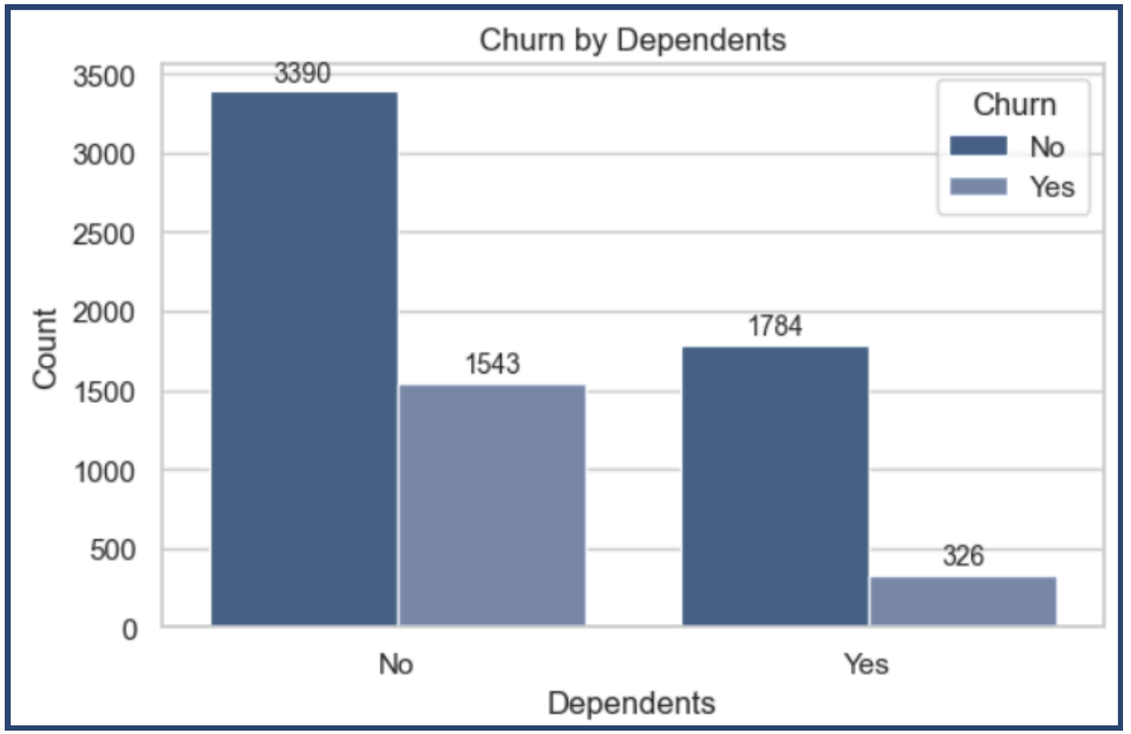


EDA - BIVARIATE ANALYSIS

INSIGHTS:

Customers without dependents churn much more than those with dependents, though retention exceeds churn in both groups.

- No dependents: ~3390 retained vs ~1543 churned, indicating higher churn pressure in this segment.
- With dependents: ~1784 retained vs ~326 churned, showing notably lower churn.
- Target churn reduction efforts on customers without dependents, as they are significantly more likely to leave.

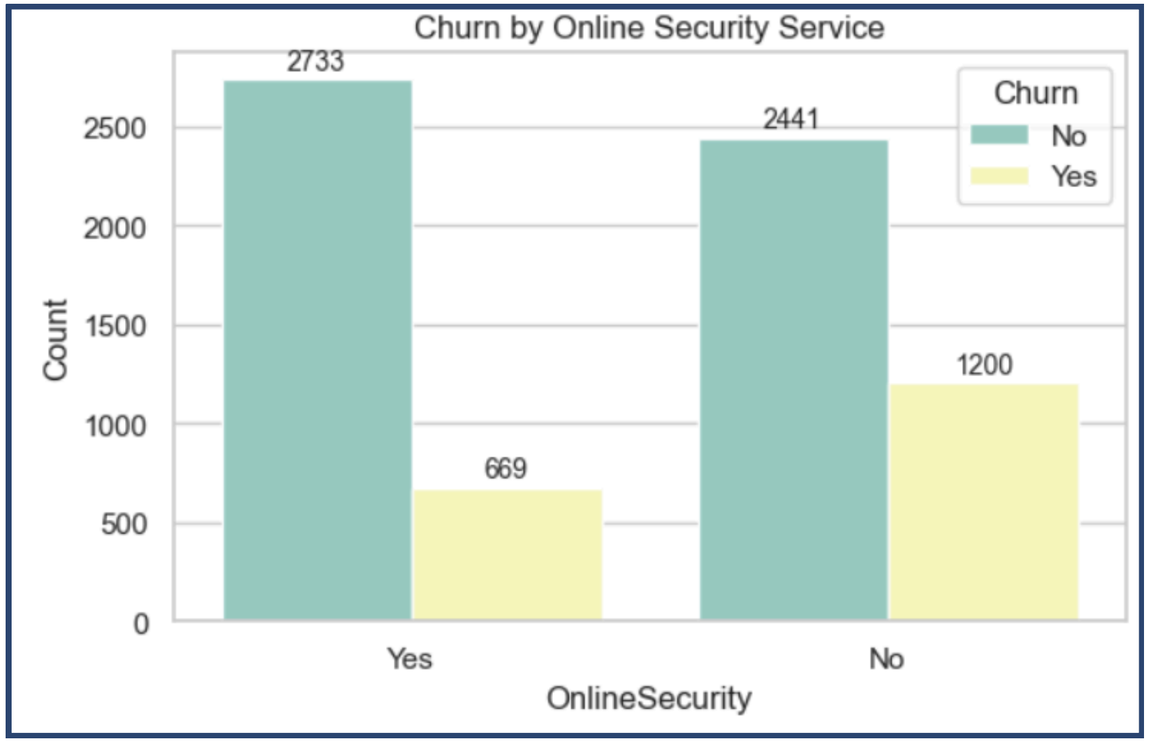


EDA - BIVARIATE ANALYSIS

INSIGHTS:

Customers with online security churn far less than those without it.

- With security: ~2733 retained vs ~669 churned, indicating strong retention.
- Without security: ~2441 retained vs ~1200 churned, showing markedly higher churn.
- Promoting online security add-ons may reduce churn by shifting customers into the lower-risk group.

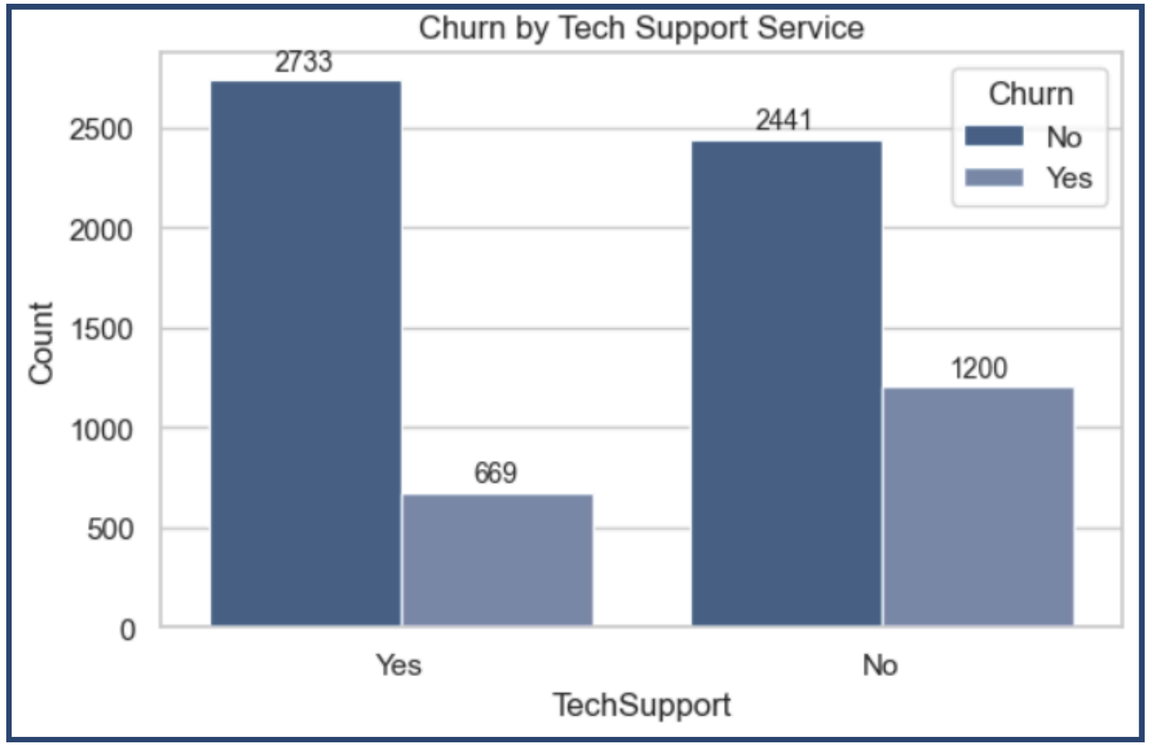


EDA - BIVARIATE ANALYSIS

INSIGHTS:

Customers with tech support churn far less than those without it.

- With tech support: ~2733 retained vs ~669 churned, indicating strong retention.
- Without tech support: ~2441 retained vs ~1200 churned, showing significantly higher churn.
- Offering or bundling tech support could be an effective lever to reduce churn.

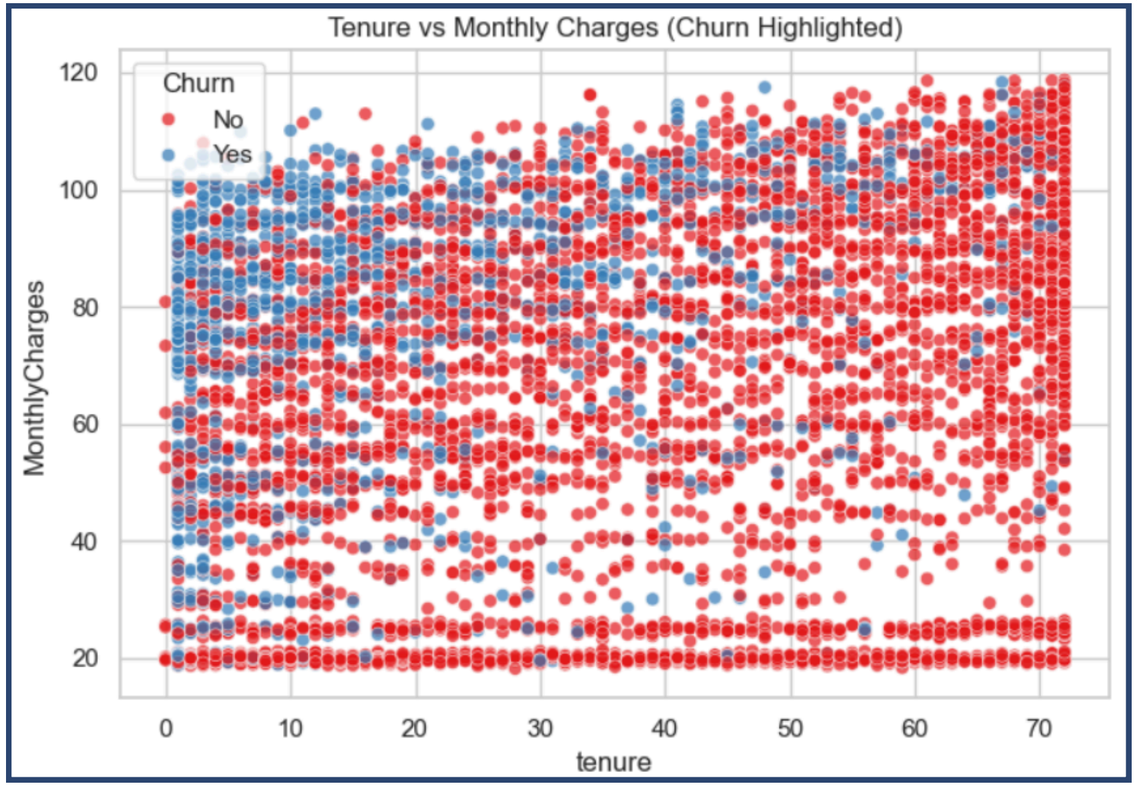


EDA - MULTIVARIATE ANALYSIS

INSIGHTS:

Churn is concentrated among newer customers with a range of monthly charges, while long-tenure customers show far fewer churn points across all charge levels.

- Early tenure (0–12 months) has many blue churn points across both low and high charges, indicating onboarding and early value perception are critical.
- As tenure increases, churn points thin out, suggesting retention improves substantially after the first year.
- High monthly charges alone do not guarantee churn; long-tenure high-payers mostly stay, pointing to service satisfaction or bundled value.

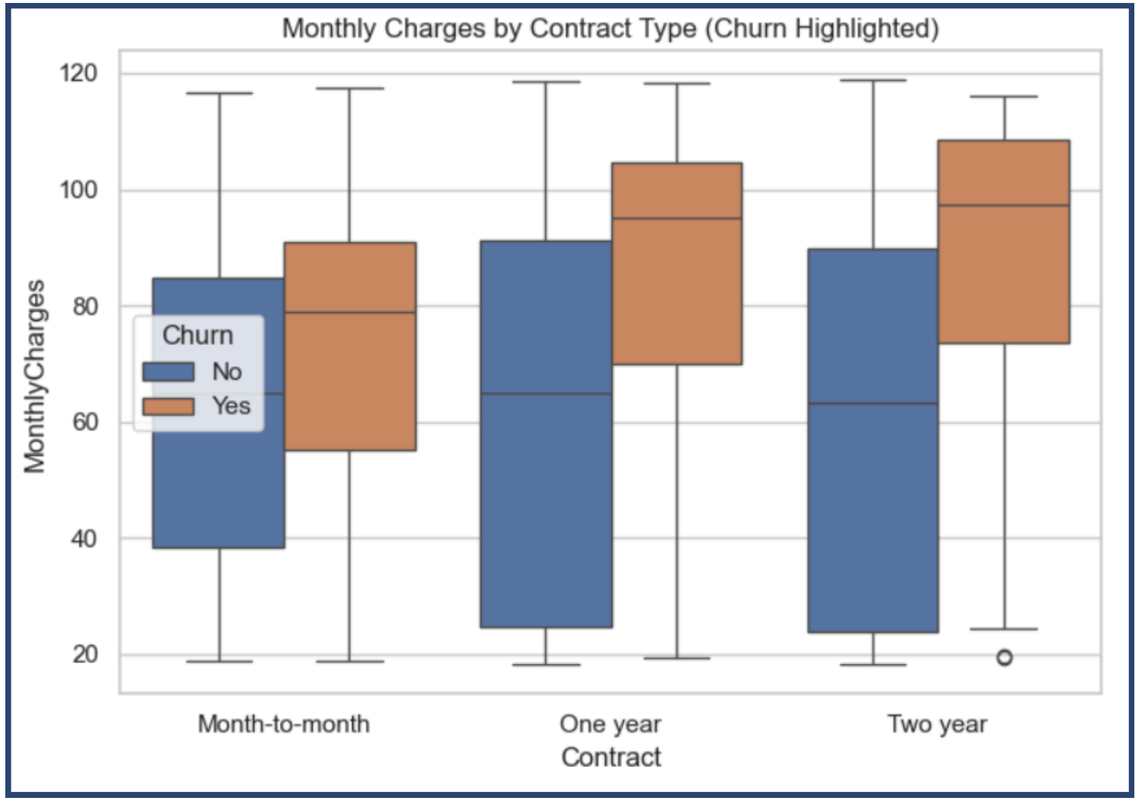


EDA - MULTIVARIATE ANALYSIS

INSIGHTS:

Churners pay higher monthly charges than non-churners across all contract types, and longer contracts tend to have higher charge levels overall.

- Early tenure (0–12 months) has many blue churn points across both low and high charges, indicating onboarding and early value perception are critical.
- One-year: Both groups shift higher than month-to-month, with churn still above non-churn, suggesting premium plan uptake among leavers.
- Two-year: Highest charge levels overall; churners remain priced above non-churners, though non-churn IQR is broad, implying mixed plan tiers.

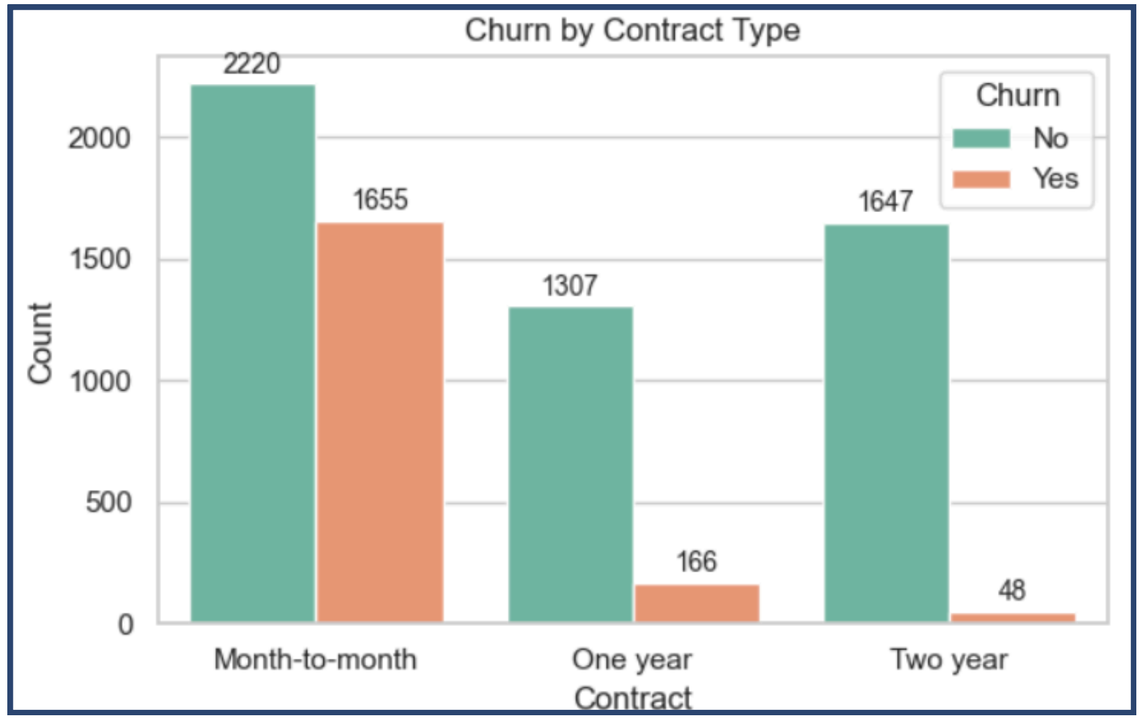


EDA - CUSTOMER CHURN ANALYSIS

INSIGHTS:

Month-to-month contracts have the highest churn, while one-year and two-year contracts exhibit very low churn.

- Month-to-month: ~2220 retained vs ~1655 churned, indicating elevated risk among flexible plans.
- One-year: ~1307 retained vs ~166 churned, showing strong retention with annual commitment.
- Two-year: Highest charge levels overall; churners remain priced above non-churners, though non-churn IQR is broad, implying mixed plan tiers.

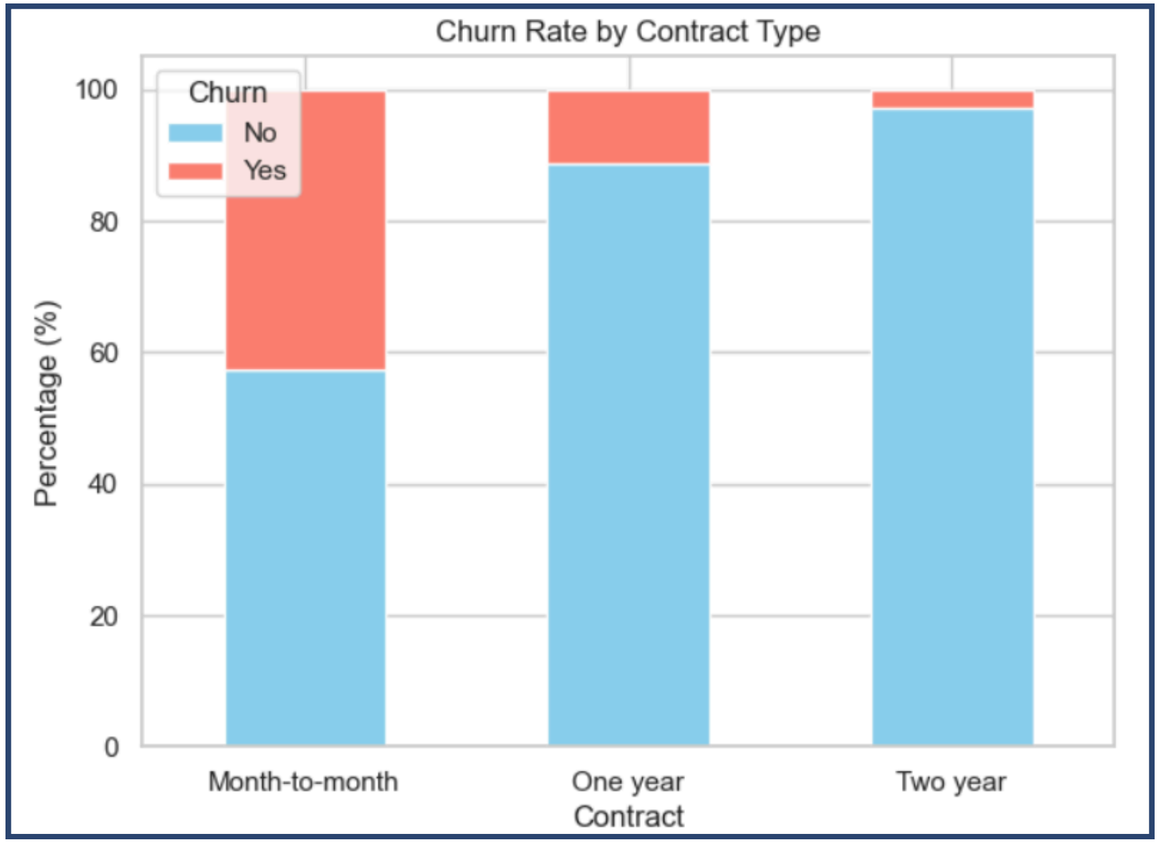


EDA - MULTIVARIATE ANALYSIS

INSIGHTS:

Longer contracts have much lower churn rates, with month-to-month showing the highest percentage of churners.

- Month-to-month: A large share of the bar is churn, indicating the riskiest contract type by percentage.
- One-year: Churn slice is small, reflecting strong retention relative to flexible plans.
- Two-year: Churn slice is minimal, the best retention rate among all contract types.

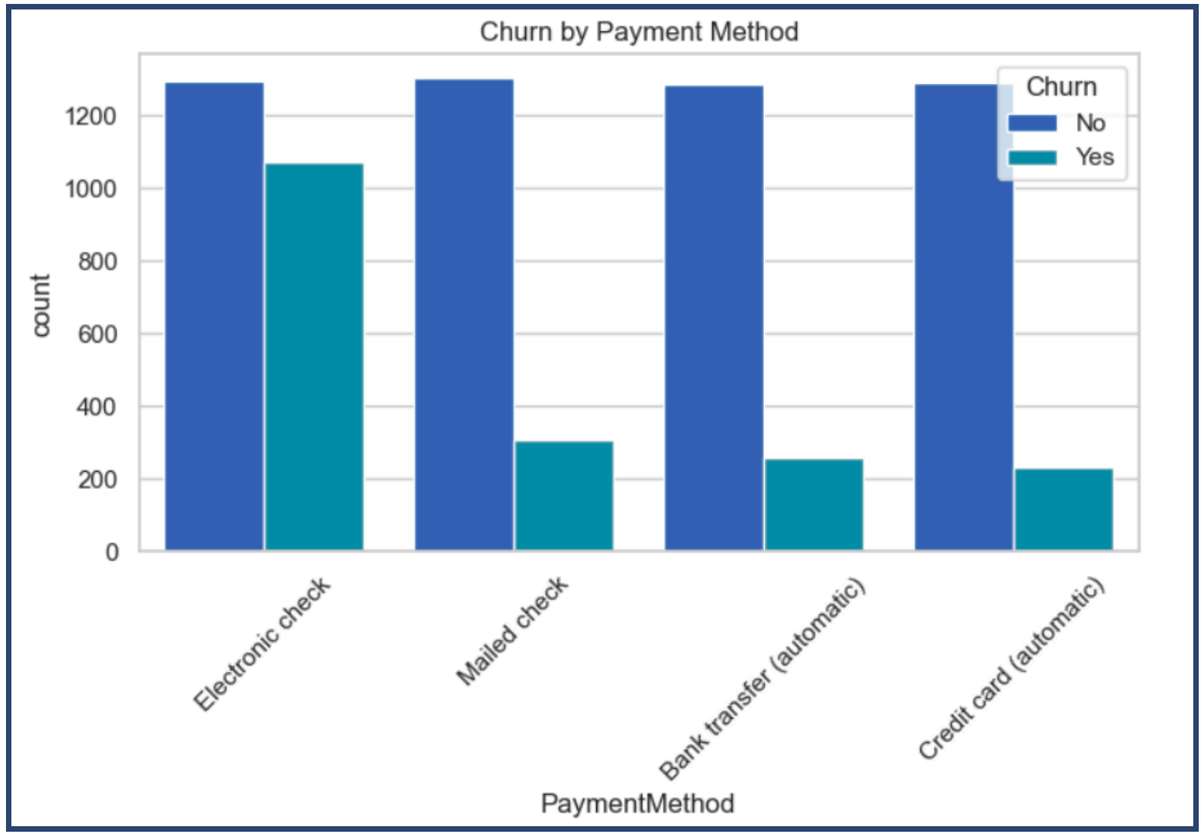


EDA - MULTIVARIATE ANALYSIS

INSIGHTS:

Electronic check is the most used payment method, while the three others have similar and lower adoption.

- Electronic check leads by a clear margin, indicating strong preference for quick digital payments without cards.
- Mailed check, bank transfer (automatic), and credit card (automatic) are each moderately used and relatively close in count.
- Encouraging automatic methods could reduce late payments and improve retention versus manual checks.

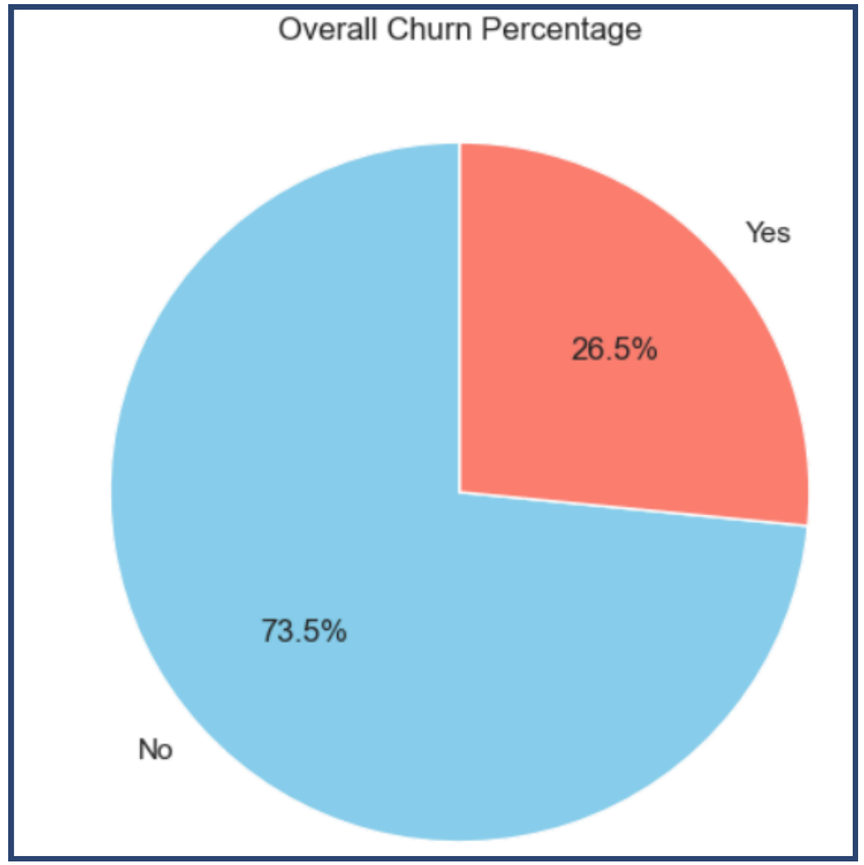


EDA - MULTIVARIATE ANALYSIS

INSIGHTS:

Overall churn is about 26.5%, meaning roughly three-quarters of customers are retained.

- Retained customers (No) make up around 73.5% of the base.
- Churned customers (Yes) account for about one-quarter, indicating meaningful but manageable attrition.
- Monitoring this rate over time and segmenting by contract, payment, and services can pinpoint improvement levers.

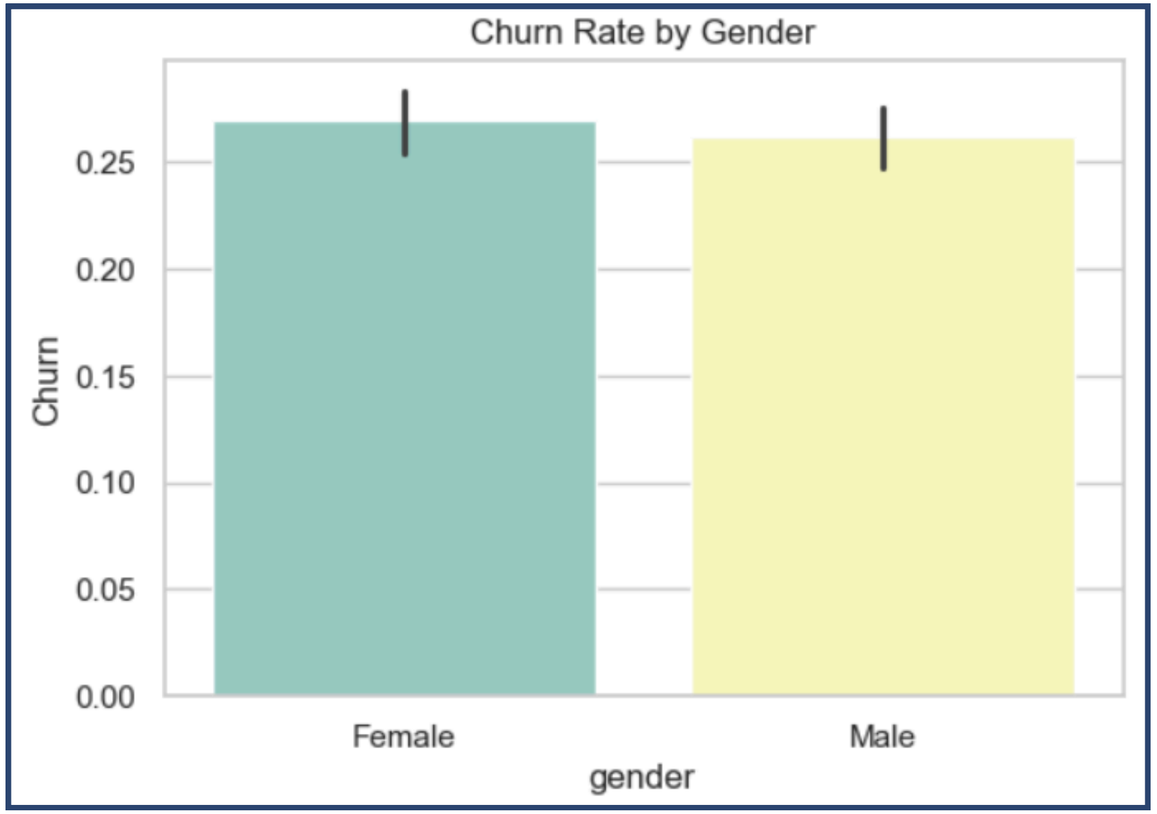


EDA - MULTIVARIATE ANALYSIS

INSIGHTS:

Churn rates by gender are nearly identical, with only a slight, likely insignificant difference.

- Female churn proportion is just above 0.26, very close to male.
- Male churn proportion is just below female's bar, and the error bars overlap, indicating no strong gap.
- Gender is not a meaningful predictor of churn in this dataset.

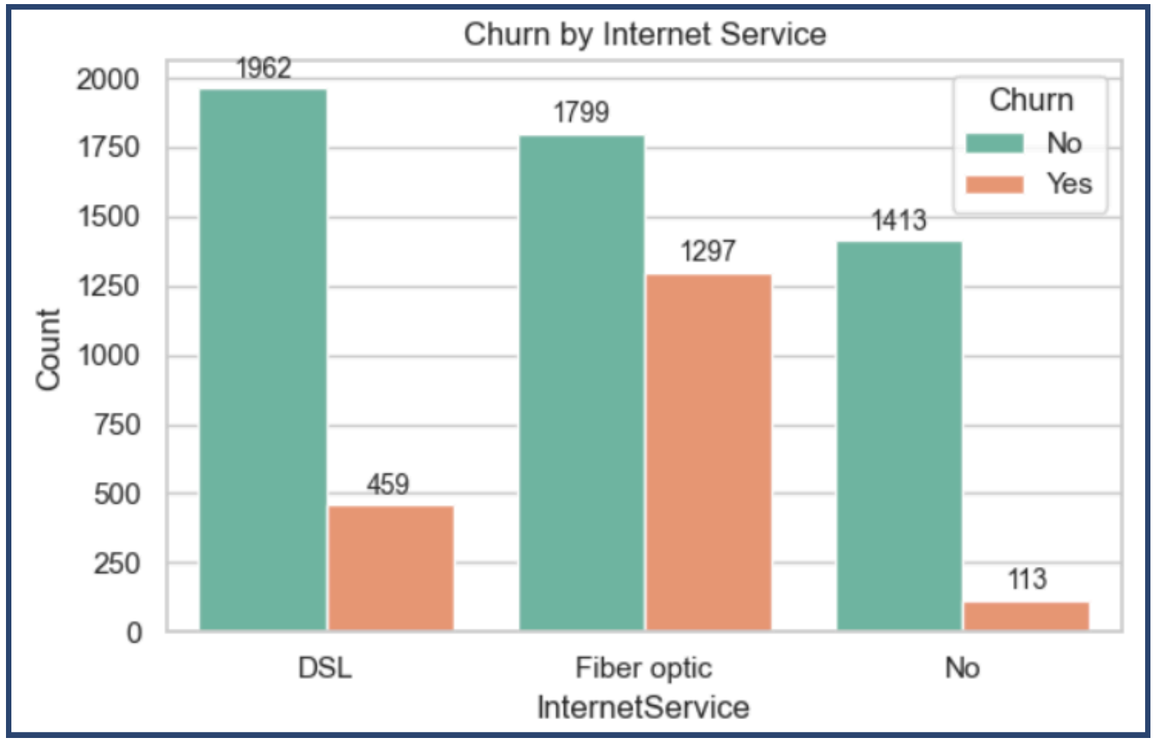


EDA - MULTIVARIATE ANALYSIS

INSIGHTS:

Fiber optic customers churn the most, DSL customers churn less, and customers without internet service churn the least.

- DSL: Strong retention (1962) with comparatively low churn (459).
- Fiber optic: Retention (1799) is close to churn (1297), indicating higher risk in this segment.
- No internet: Highest retention (1413) with minimal churn (113), likely due to fewer service issues or a different product mix.

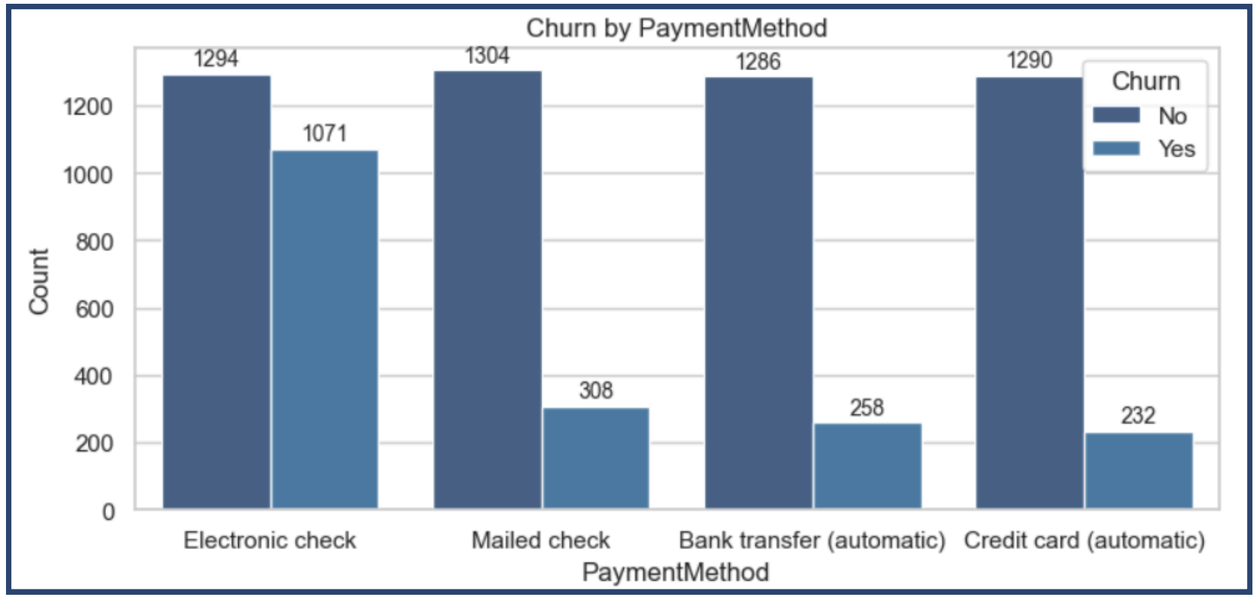


EDA - MULTIVARIATE ANALYSIS

INSIGHTS:

Electronic check users churn far more than users of other payment methods, while automatic bank/credit methods have the lowest churn.

- Electronic check: High churn count (1071) close to retention (1294), marking it the riskiest segment.
- Mailed check: Churn is much lower (308) versus retained (1304), indicating greater stability.
- Bank transfer (auto): Low churn (258) with strong retention (1286), suggesting auto-pay improves stickiness.
- Credit card (auto): Lowest churn (232) with high retention (1290), the safest payment group.

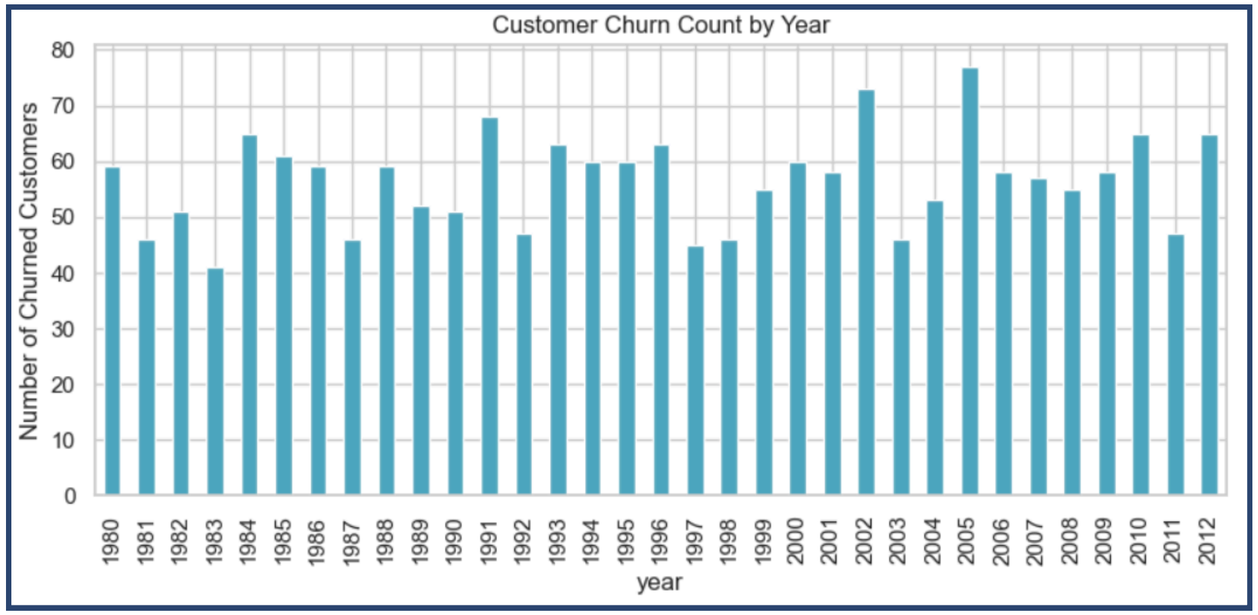


EDA - MULTIVARIATE ANALYSIS

INSIGHTS:

Annual churn counts fluctuate between roughly 40 and 75 per year, with notable peaks around 2002 and 2005.

- Early years (1980s–1990s) show moderate variation, mostly in the 45–65 range.
- Spikes occur in the early 2000s, especially 2002 and 2005, indicating periods of elevated attrition.
- Post-2005, churn remains elevated but variable, suggesting ongoing external or pricing/service factors influencing exits.

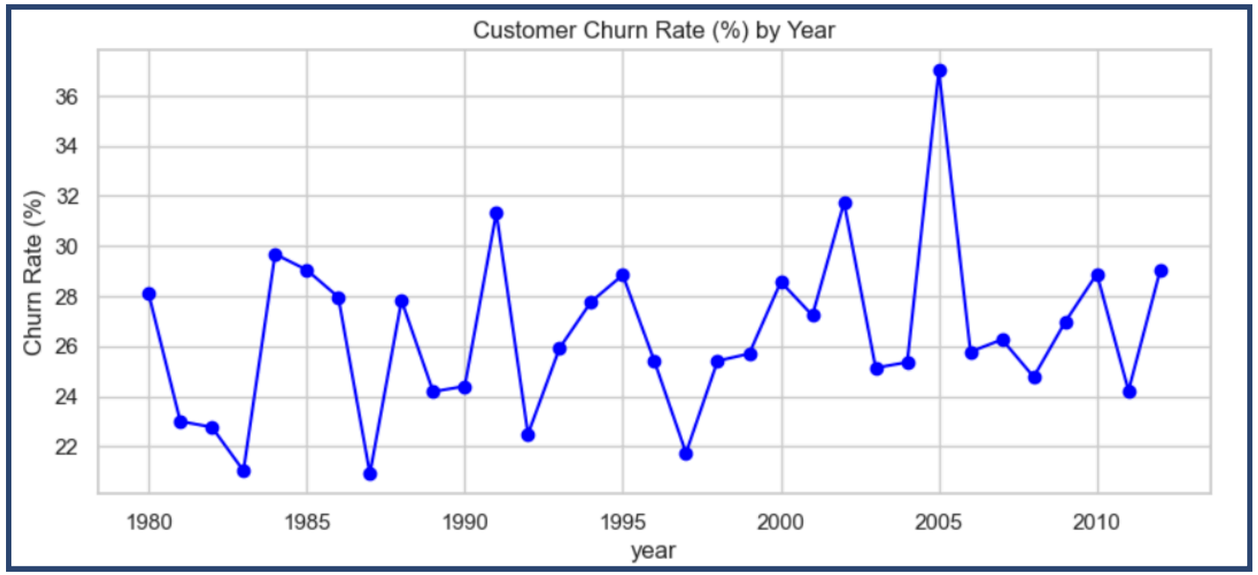


EDA - MULTIVARIATE ANALYSIS

INSIGHTS:

Churn rate fluctuates year to year, with a prominent spike around 2005 and smaller peaks near 1991, 1995, and 2002.

- Early 1980s saw declines from ~28% toward low 20s, followed by a jump near 1984–1985.
- The 1990s feature oscillations between ~21% and ~31%, with local highs around 1991 and 1995.
- Early 2000s rise to a high near 32% in 2002 and the series peaks around 37% in 2005 before easing back to mid–20s to high–20s later.



FINAL RECOMMENDATIONS

Final Here are concise, high-impact actions to reduce churn:

- ☑ Push annual and two-year contracts with limited-time discounts or loyalty perks; convert month-to-month users first.
- ☑ Target early-tenure customers (first 12–18 months) with onboarding, check-ins, and usage nudges; this is the highest-risk window.
- ☑ Offer value relief for high monthly charges: bundle security/tech support, tiered discounts, or loyalty credits for high-bill segments.
- ☑ Promote automatic payments (bank/credit) to lower churn risk versus electronic checks; run opt-in campaigns with incentives.
- ☑ Cross-sell Online Security and Tech Support; subscribers to these add-ons churn less.
- ☑ Prioritize fiber users for proactive care and save offers; churn is highest in this internet segment.
- ☑ Focus retention on customers without dependents/partners; tailor pricing and perks to this higher-risk group.
- ☑ Monitor churn spikes over time; set alerts around known peak periods and launch preventative offers in advance.
- ☑ Track and report churn rate alongside tenure, contract type, payment method, and add-on adoption to guide weekly actions.

CONCLUSION

- Customer churn is strongly tied to contract type, service options, and tenure.
- Month-to-month plans and high charges increase churn probability.
- Implementing retention-focused strategies can significantly reduce churn.
- The analysis provides a foundation for predictive modeling in the future.

THANK YOU

bhaskarpal1707.github.io/portfolio/

bhaskarpal.official@gmail.com