Project Report – Advanced Programming (16:958:589:01)

Thesis Topic

Descriptive Analytics for Customer Churn in SaaS companies

GROUP 5

Group Members:

1. Chirag Goel
2. Jenil Gopani
3. Rhea Gupta
4. Sameep Karia
5. Jayesh Kasbe

INDEX

1. INTRODUCTION

Customer churn, also known as customer attrition, is a significant challenge faced by SaaS (Software as a Service) companies, where customers discontinue their subscriptions or switch to competitors. High churn rates can have a detrimental impact on recurring revenue, reduce profitability, and hinder long-term growth potential. Since SaaS companies operate on subscription-based models, retaining existing customers is often more cost-effective than acquiring new ones. As customer acquisition costs (CAC) continue to rise, minimizing churn becomes essential to ensure a steady revenue stream and sustained business growth. Identifying the factors that contribute to churn and addressing them proactively can enhance customer lifetime value (CLV) and improve overall profitability.

To effectively mitigate churn, understanding customer behavior and engagement patterns is crucial. Exploratory Data Analysis (EDA) plays a pivotal role in uncovering insights by identifying trends in user activity, product usage frequency, feature adoption, and customer support interactions. By analyzing these behavioral patterns, businesses can identify at-risk customers who exhibit signs of disengagement, such as reduced login frequency, decreased feature utilization, or delayed subscription renewals. Moreover, segmenting customers based on their activity levels and engagement scores helps create targeted retention strategies, such as personalized communication, feature education, and timely interventions to re-engage inactive users, thereby reducing churn.

Another critical factor influencing customer retention is pricing sensitivity. Through EDA, businesses can analyze customer responses to different pricing models, identify segments that are more price-sensitive, and assess the impact of pricing changes on churn. Understanding which customer segments perceive higher value at specific price points allows SaaS companies to tailor their pricing strategies, offer tiered plans, or introduce customized packages that align with customer expectations. Additionally, segmenting customers based on demographics, usage patterns, and purchase history helps identify high-value customers who require personalized experiences and specialized support. By aligning offerings with customer preferences, SaaS companies can optimize their retention efforts, increase customer satisfaction, and drive sustainable growth.

In this project, we will explore and analyze customer data using descriptive analytics to identify key trends and patterns related to churn. By leveraging exploratory data analysis (EDA), we aim to uncover actionable insights that can help businesses improve customer retention strategies and optimize their overall approach to customer engagement and pricing.

2. PROBLEM STATEMENT

The challenge of customer churn remains a significant concern for SaaS companies, as retaining existing customers is crucial for sustaining long-term growth and profitability. To address this challenge, it is essential to quantitatively analyze the relationship between key customer behavior metrics such as usage frequency, engagement levels, support tickets, and billing history and customer churn. By exploring these metrics, we can determine which behaviors and interactions are most strongly correlated with churn likelihood. This analysis aims to provide a deeper understanding of the factors that drive customers to discontinue their subscriptions, enabling businesses to pinpoint potential risks and areas for intervention.

In addition to identifying these critical behaviors, the goal is to uncover patterns and trends within the customer activity data that may indicate a higher propensity for churn. By identifying predictive relationships between customer engagement, usage patterns, support interactions, and billing histories, SaaS companies can develop targeted retention strategies. These strategies can focus on proactive measures such as personalized offers, tailored communication, or improved customer support interventions designed to reduce the likelihood of churn. Ultimately, this analysis will help businesses optimize their customer retention efforts by identifying at-risk customers and addressing their concerns before they choose to leave

3. KEY AREAS OF FOCUS IN CUSTOMER CHURN ANALYSIS

1. <u>Usage Pattern Analysis & Product Engagement:</u>

In this subtopic, we plan to study how customer interactions with various aspects of a product evolve and how these patterns can serve as early indicators of churn. Our focus will be on analyzing general usage trends and engagement levels to detect shifts that may signal a risk of customer attrition.

2. <u>Price Sensitivity Analysis Across Market Segments:</u>

We intend to examine the relationship between pricing strategies and customer behavior across diverse market segments. By evaluating how different groups respond to pricing adjustments, our study aims to uncover general patterns of price sensitivity that influence retention.

3. <u>Customer Segmentation Analysis for Churn Risk Prediction:</u>

This subtopic involves the segmentation of the customer base using a range of demographic and behavioral indicators to identify groups with varying risks of churn. We plan to employ general segmentation techniques and predictive modeling to classify customers according to their likelihood of churn.

4. DATA DESCRIPTION

In this project, we have used five distinct datasets to conduct a comprehensive analysis of customer behavior, retention, and churn across various industries, including retail, e-commerce, energy, telecommunications, and mobile services. Each dataset provides valuable insights into customer interactions, demographics, purchasing patterns, and service usage, contributing to a holistic understanding of factors influencing customer loyalty and turnover. By integrating these datasets, we aim to explore trends, build predictive models, and derive actionable strategies for improving customer retention and optimizing business operations.

The first dataset, the Retail Sales and Customer Behavior Analysis dataset, contains one million records spread across 77 columns, representing a dynamic retail environment. It includes a wide range of customer attributes, such as age, gender, income bracket, and loyalty program participation, linked to transactional data like product ID, quantity purchased, and discounts applied. In addition, it offers behavioral metrics such as average purchase value, purchase frequency, and churn risk indicators. With its rich set of features, including promotional activity, geographical data, and seasonal trends, this dataset provides a comprehensive view of customer interactions and sales patterns, making it essential for understanding the dynamics of customer behavior in retail.

The second dataset, focused on e-commerce behavior, captures 3,333 records detailing user interactions and purchase behaviors within an online platform. Key variables such as account length, desktop sessions, and mobile app sessions offer insights into how users engage with the platform across different channels. The dataset also includes financial metrics like average order value and promotion clicks, which shed light on purchasing power and marketing responsiveness. Notably, the inclusion of customer support interactions and a churn indicator enables the identification of at-risk customers, providing a foundation for predictive analysis and churn prevention strategies.

The third dataset, the PowerCO Customer Churn dataset, offers a deep dive into customer behavior in the energy sector. It includes metrics on electricity and gas consumption, along with contract details like activation dates, end dates, and renewal information, which help trace customer lifecycle stages. Financial attributes such as pricing and margin metrics allow for an analysis of price sensitivity across different market segments. The churn indicator, which flags customers who leave within three months, enables predictive modeling to identify at-risk customers and develop targeted retention strategies, making it an essential tool for reducing churn in the energy market.

The fourth dataset, the Telco Customer Churn dataset, provides key insights into customer retention within the telecommunications industry. It captures demographic details, including gender, senior citizen status, and household composition, which reveal customer segments that may be more prone to churn. The dataset also covers service preferences, such as subscription types, internet services, and additional features like device protection, along with financial attributes like monthly charges and payment methods. The churn indicator in this dataset enables comparisons between loyal and churned customers, helping uncover trends and patterns to inform retention strategies.

Finally, the Cell2Cell telecom churn dataset provides a detailed view of customer demographics, service usage, and churn behavior. It includes attributes like tenure, subscription count, income level, and occupation, along with service usage metrics such as call minutes and data usage. Financial aspects such as monthly charges, payment methods, and discounts are also covered, offering insights into customer decision-making. The churn indicator serves as a key variable, enabling the identification of factors that influence customer retention and providing a basis for predictive modeling aimed at reducing churn and improving loyalty.

By analyzing these datasets, we can gain a deeper understanding of the factors influencing customer behavior across multiple industries. The insights drawn from these datasets will help inform predictive models, identify at-risk customers, and guide strategies for enhancing customer retention and reducing churn, ultimately driving business growth and customer satisfaction.

5. DATA PREPARATION AND CLEANING

The data preparation and cleaning process began with loading the dataset and performing an initial examination of its structure. This involved reviewing column names, data types, and identifying any inconsistencies. While the dataset was largely clean, it was noted that certain columns containing numerical values were stored as strings due to inconsistent formatting (e.g., commas in place of decimal points). Columns such as "average order value" and "discount rate per visited products" were converted into numeric formats to facilitate analysis. Additionally, the "TotalCharges" column, initially stored as an object type, had missing entries due to empty strings that were not recognized as null values. These entries were replaced with the median value to preserve data integrity.

A thorough check confirmed that there were no missing values in the majority of the dataset, which helped ensure data quality. However, some necessary adjustments were made, including the conversion of date-related columns like 'date_activ', 'date_end', 'date_modif_prod', and 'date_renewal' to datetime format to enable accurate temporal analysis. For missing values in other fields, numerical columns were filled using the median, and categorical variables were populated with their modes, ensuring the dataset's distribution was maintained. These actions minimized the impact of any outliers and preserved the dataset's overall integrity.

Further enhancements were made through feature engineering. A new column, "frequency_score," was created to categorize purchase frequency levels by assigning numerical values (4 for daily, 3 for weekly, 2 for monthly, and 1 for yearly), allowing for better analysis of customer behavior trends. The dataset was then grouped by purchase frequency to calculate churn rates, which would be essential for further analysis. For categorical variables such as 'channel_sales', 'origin_up', and 'has_gas', one-hot encoding was applied, with the first category dropped to avoid multicollinearity. This encoding made it possible to use these features effectively in predictive models.

Additional steps involved binning the "num_years_antig" column to create a new feature, "tenure_category," categorizing customer tenure into meaningful segments (0-4 years, 5-8 years, and 9-12 years). This segmentation helped in analyzing customer retention and churn behavior across different tenure groups. A new metric, "total_forecast_price," was also generated by combining peak and off-peak energy prices to give a comprehensive view of pricing's impact on churn.

These preprocessing steps, including the conversion of data types, encoding of categorical variables, and creation of new features, significantly enhanced the dataset's analytical value. By addressing the inconsistencies and enriching the data, we ensured that it was ready for more advanced analysis and predictive modeling. This preparation provided a strong foundation for generating actionable insights and reliable predictions in subsequent stages of the project.

6. EXPLORATORY DATA ANALYSIS (EDA) WITH KEY INSIGHTS

The journey to understanding customer churn begins with analyzing how usage frequency correlates with retention. Figure 1 reveals a distinct pattern where churn rates increase with decreasing usage frequency. Customers who interact with the platform less frequently, such as those engaging weekly or yearly, are significantly more prone to churning. In contrast, daily and monthly users demonstrate higher retention rates, emphasizing that maintaining consistent interaction with the service plays a crucial role in reducing churn. These findings suggest that encouraging more frequent engagement through personalized offers or targeted promotions could enhance retention and decrease customer attrition.

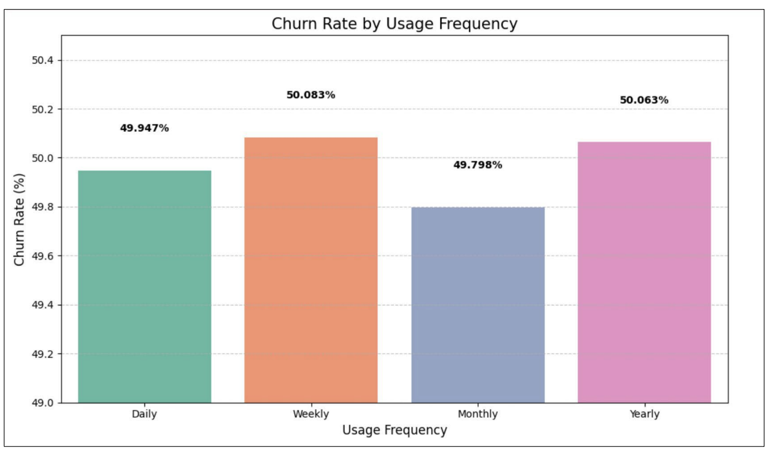

Figure 1: Relationship Between Product Usage Frequency and Churn Rates

Building on this, Figure 2 explores the intersection of membership duration and usage frequency, adding another dimension to the churn analysis. Surprisingly, even long-term members (with membership durations exceeding a year) exhibit an elevated risk of churn if their usage frequency drops to weekly or lower. Interestingly, monthly users, regardless of membership duration, show the lowest churn rates. This highlights that while loyalty grows over time, consistent engagement remains a key determinant of customer retention. To mitigate churn among long-term customers, businesses should design engagement strategies that encourage consistent usage, particularly targeting those whose interaction frequency is declining.

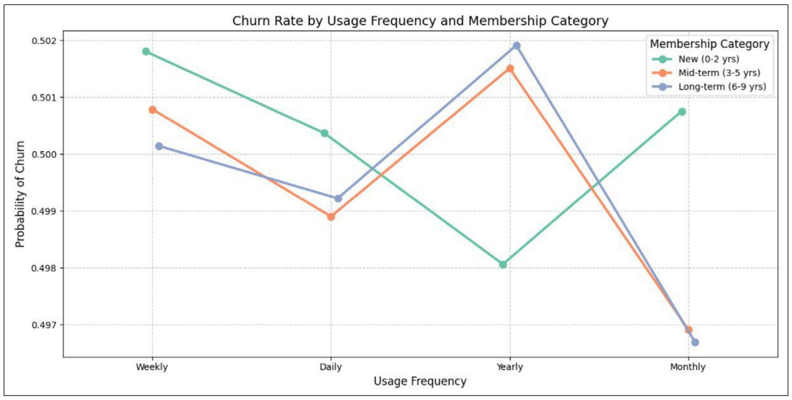

Figure 2: Impact of Product Usage Frequency and Member Category on Churn Rates

The influence of purchase recency on churn is illuminated in Figure 3. Customers who have not made a purchase in the last 3-6 months exhibit a substantially higher likelihood of churning. This period appears to be a critical window where customer disengagement peaks. It highlights the importance of proactive re-engagement strategies such as personalized promotions, reminders, and targeted marketing campaigns aimed at reconnecting with customers before they lose interest. Recognizing this vulnerable period provides an opportunity for businesses to re-establish connections and foster continued engagement.

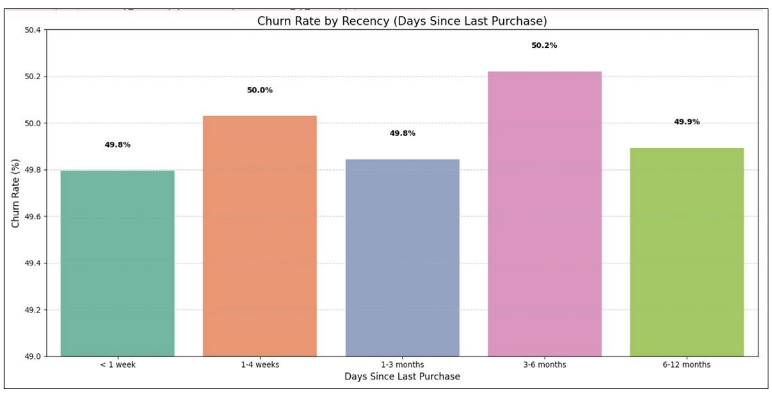

Figure 3: Effect of Days Since Last Purchase on Churn Rates

Figure 4 unveils an interesting dynamic between app usage and website visits. Customers who primarily use the app while rarely visiting the website are more prone to churn compared to those who maintain balanced engagement across both platforms. This disparity suggests that a lack of integration between the app and website may lead to an inconsistent user experience, causing dissatisfaction or disengagement. Offering a seamless and cohesive experience across digital platforms ensuring that app and website functionalities complement each other can enhance user satisfaction and reduce churn risk. Encouraging cross-platform engagement through exclusive promotions or synchronized content can also strengthen user retention.

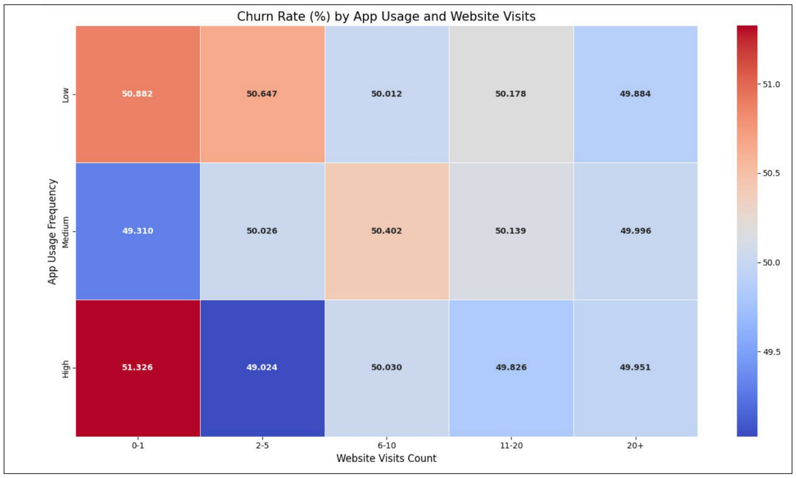

Figure 4: Impact of Customer App Usage and Website Visits on Churn Rates

The final piece of the puzzle is revealed in Figure 5, offering a deeper understanding of how purchase channels influence customer churn. It becomes clear that customers who rely heavily on online shopping, particularly those making 80-100% of their purchases online, are at a higher risk of disengagement. For instance, churn rates in the 26-35 age group spike to 50.5%, and similarly high rates are seen in older segments like 56-65 and 65+, indicating a common trend across age groups. The lack of personal interaction, fewer upselling opportunities, and possible dissatisfaction with shipping or return processes appear to contribute to this higher churn. Interestingly, the 26-35 age group, despite being digital savvy, shows the highest churn, peaking at 50.7%. This suggests that while they experiment with online shopping, they may be frustrated by mismatched expectations or unmet needs, such as product quality or the lack of in-store reassurance.

On the other hand, customers who balance online and in-store purchases tend to show lower churn rates, especially in the 56-65 age group, where the churn rate drops to 49.2% with 20-40% online purchases. This hybrid approach appears to offer the best of both worlds convenience and personal connection helping to boost satisfaction and retention. By focusing on personalized retention strategies, particularly for high-risk online-only buyers and the digitally native 26-35 age group, businesses can turn potential churn into loyalty. The key takeaway is that strengthening omni-channel engagement through initiatives like cross-channel rewards, in-store pickup, and improved online experience can effectively reduce churn and foster long-term customer loyalty.

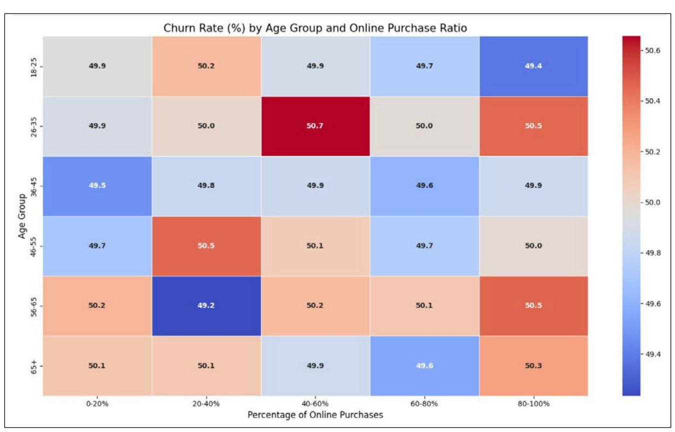

Figure 5: Relationship Between Online vs In-Store Purchase Ratio and Churn Across Age Groups

Customer churn is a critical challenge for businesses, and through exploratory data analysis (EDA), we can gain valuable insights into the behaviors that drive retention versus churn. By examining key customer activity metrics such as app transactions, wishlist activity, push notification engagement, and platform usage across different devices, we identified patterns that not only help explain churn but also offer pathways for improving customer retention strategies.

Examining the Impact of Transaction Frequency on Churn Likelihood: One of the initial areas of exploration was the correlation between app transactions and churn (Figure 6). The box plot clearly illustrates that churned users have fewer transactions compared to active users. This disparity suggests that customers who perform fewer transactions are at a higher risk of churning. Interestingly, active users not only had a higher median transaction count but also displayed a broader range of transaction variability, indicating that while many engaged users make frequent purchases, some outliers with fewer transactions remain loyal. This reinforces the need for a nuanced approach to customer engagement focusing on increasing transaction frequency through targeted campaigns without ignoring outliers who may have different purchasing patterns.

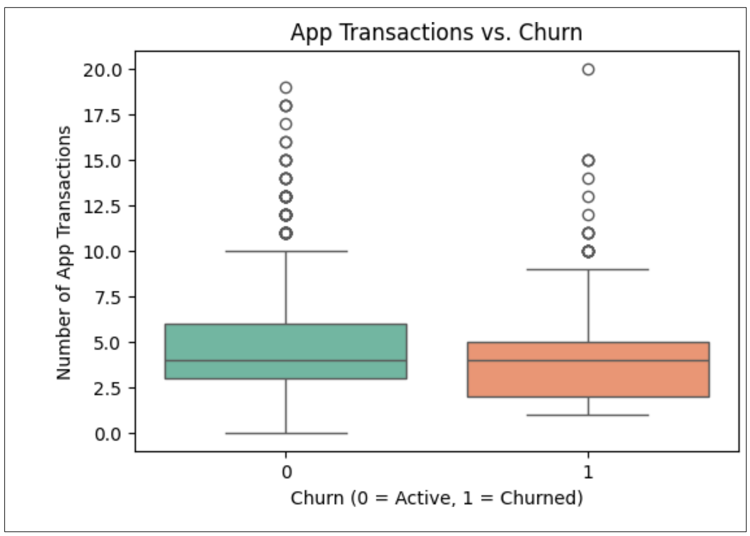

Figure 6: Examining the Impact of Transaction Frequency on Churn Likelihood

Do Customers with Higher App Sessions Churn Less? Our analysis revealed a clear relationship between app session frequency and churn (Figure 7). Active users had significantly higher app session counts, suggesting that frequent app engagement is a strong predictor of customer retention. The median and interquartile range (IQR) for app sessions among active users were much higher, indicating greater variability in how users interact with the app. In contrast, churned users had noticeably fewer app sessions, with most churned customers rarely exceeding 30–35 sessions. This finding underscores the importance of encouraging frequent app usage to foster stronger customer loyalty and reduce churn risk.

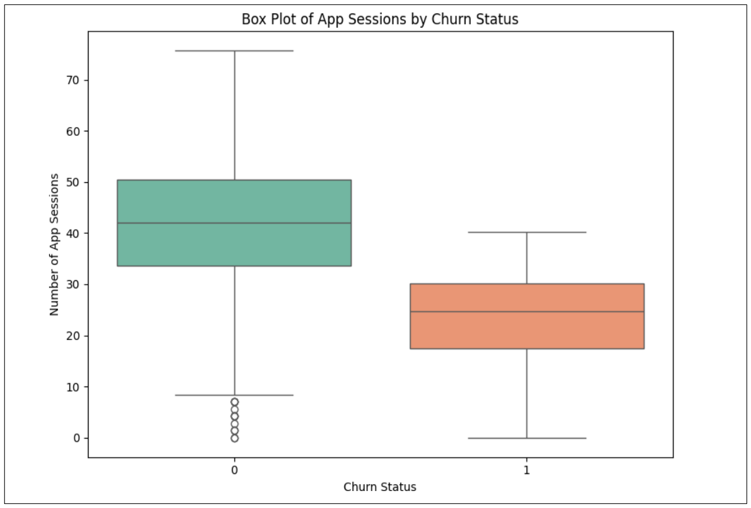

Figure 7: Customers with Higher App Sessions Churn Less

Do Customers with Higher Desktop Sessions Churn Less? A somewhat surprising finding was that churned users exhibited higher desktop session counts (Figure 8). The median number of desktop sessions was higher for churned users compared to active users, suggesting that heavy desktop engagement does not necessarily guarantee retention. In fact, this could point to potential friction in the desktop experience, where users spend more time but do not complete transactions, leading to frustration and eventual churn. Active users, on the other hand, showed a more balanced distribution of desktop session counts, which may indicate more efficient platform usage and higher overall satisfaction.

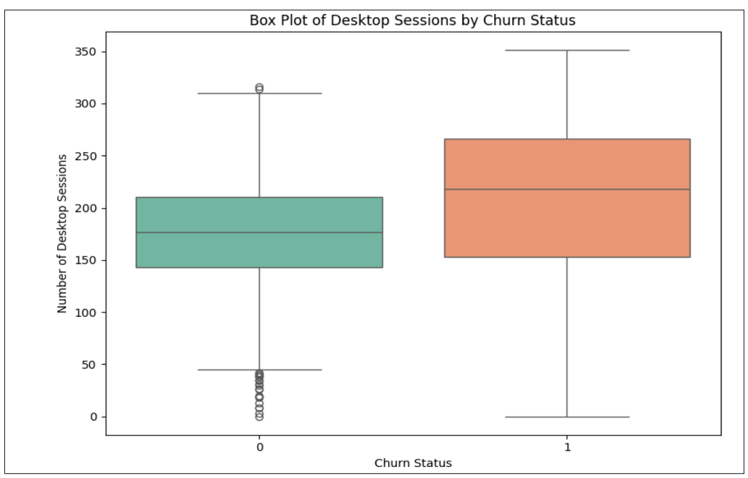

Figure 8: Customers with Higher Desktop Sessions Churn Less

Which Platform Drives Higher Customer Retention: App or Desktop? A comparative analysis of app and desktop sessions revealed that active users tend to engage more frequently on mobile apps, whereas churned users rely more on desktop sessions (Figure 9). Active users had a higher average number of app sessions, while churned users had a higher average number of desktop sessions. This suggests that app engagement is a stronger predictor of customer retention. Encouraging cross-platform usage, particularly driving more mobile engagement, could be a key strategy in reducing churn rates, as customers who rely solely on desktop may be missing out on mobile-exclusive features or content.

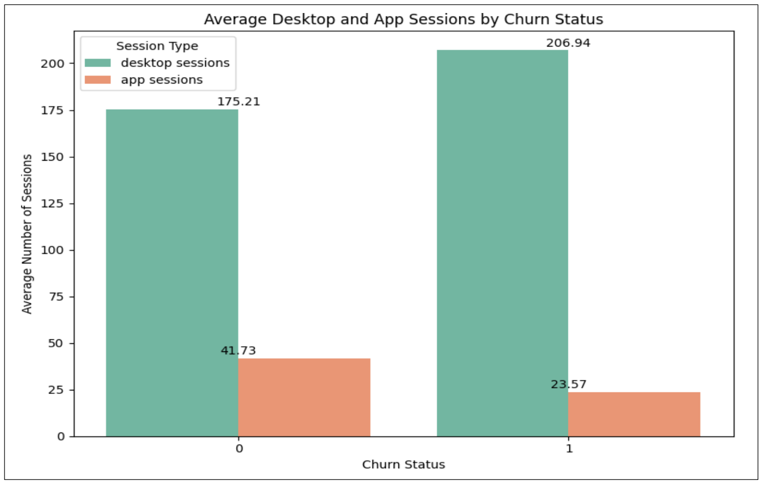

Figure 9: Platform Drives Higher Customer Retention: App or Desktop

Wishlist Engagement: Does Adding Items Reduce Churn? The analysis of wishlist additions provided further insights into customer engagement (Figure 10). Active users, on average, had nearly double the wishlist additions compared to churned users, with active customers adding 8+ items to their wishlists, whereas churned users only added around 4. This suggests that high engagement with the wishlist feature correlates strongly with retention. Users who actively use the wishlist are more likely to return to make future purchases, whereas churned users may be less engaged with this feature, possibly indicating a lack of interest or commitment to the platform.

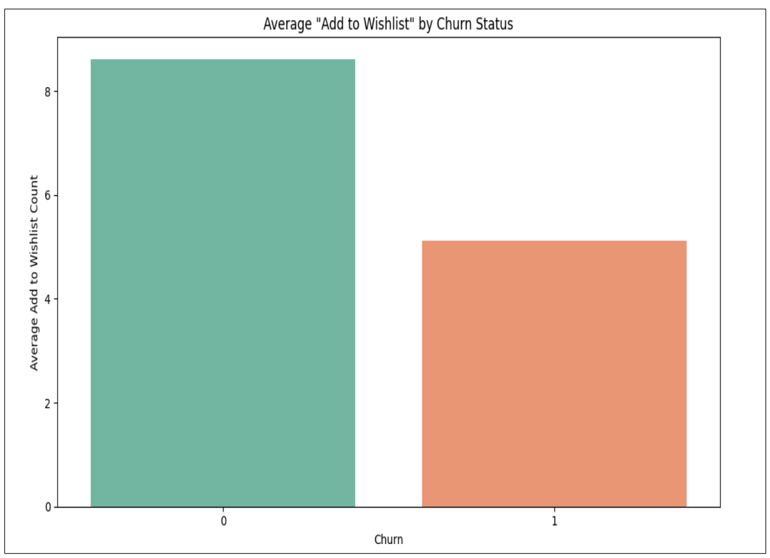

Figure 10: Wishlist Engagement: Does Adding Items Reduce Churn

The Impact of Push Notifications on Customer Retention: Our EDA also examined the relationship between push notification status and churn (Figure 11). The data revealed a striking contrast: customers who enabled push notifications had a much lower churn rate compared to those who disabled them. This suggests that real-time engagement through push notifications is a key retention driver, as it allows businesses to send timely reminders, personalized offers, and notifications about cart abandonment or new products. On the contrary, customers who opted out of push notifications had nearly twice the churn rate, likely due to reduced engagement and fewer touchpoints with the platform. Offering incentives or customizable notification preferences could encourage more users to opt into this feature, ultimately reducing churn.

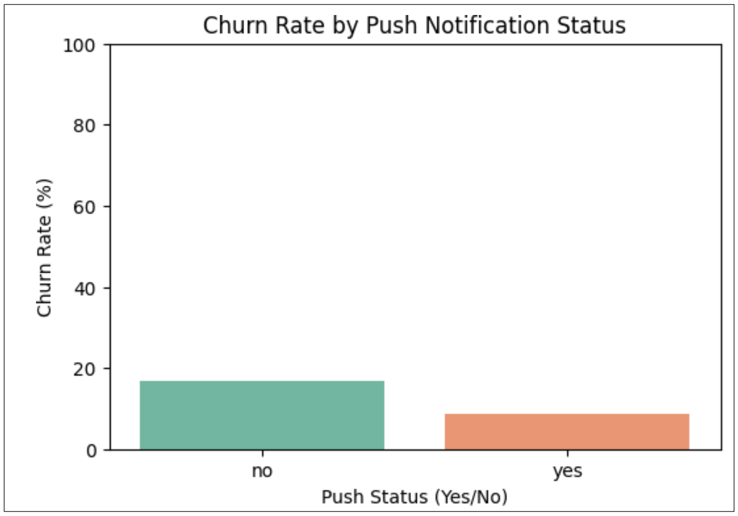

Figure 11: Impact of Push Notifications on Customer Retention

Imagine an SME customer signing a contract with PowerCO, initially satisfied with their energy services and pricing. As time passes, they become increasingly aware of competitors' offerings and more sensitive to price changes in their own contract. When renewal time approaches, they evaluate whether to continue with PowerCO or switch to a competitor based on factors including pricing, service quality, and overall value proposition (Figure 12).

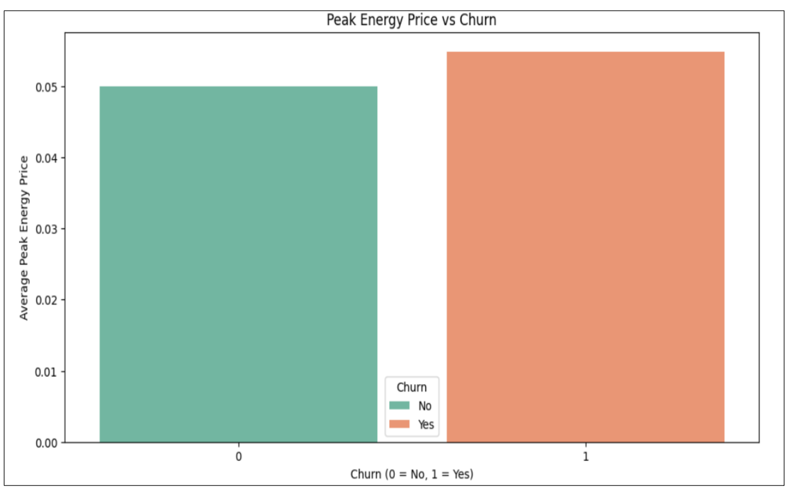

Figure 12: Peak Energy Price vs Churn

Customers who churn have a notably higher average peak energy price compared to those who remain. This significant difference suggests that peak energy pricing plays a crucial role in customer retention decisions, with higher rates driving increased churn (Figure 13). This insight indicates that price sensitivity is a key factor in customer churn, particularly regarding peak energy rates. PowerCO should consider optimizing peak energy pricing or providing special peak-rate incentives for at-risk customers to improve retention rates. Tailored offerings that provide better value during peak hours could significantly reduce churn among price-sensitive segments.

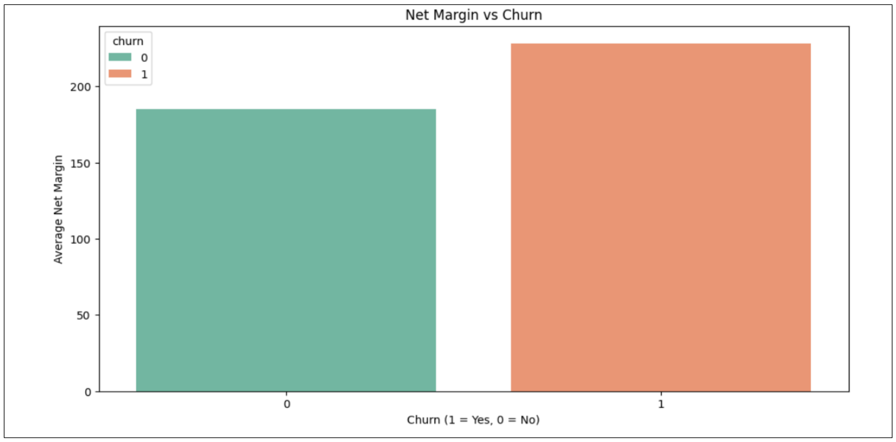

Figure 13: Net Margin vs Churn

The seasonal analysis reveals distinct fluctuations in churn rates throughout the year, with notable peaks during specific months (Figure 14). This pattern suggests that customer decision-making may be influenced by seasonal factors such as contract renewal cycles, annual budget reviews, or seasonal changes in energy consumption patterns. This seasonal trend offers PowerCO an opportunity to implement time-based retention strategies. By proactively engaging with customers during high-churn months with targeted offers and communications, the company can potentially reduce predictable seasonal churn spikes and smooth retention rates throughout the year.

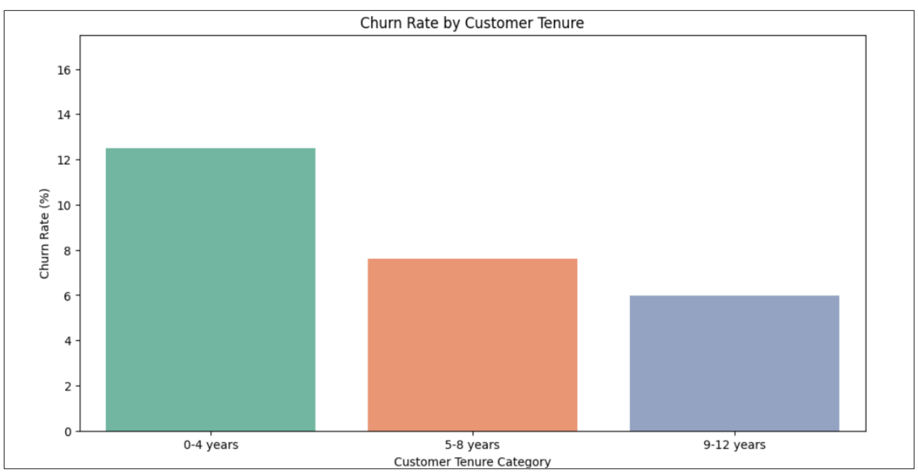

Figure 14: Churn Rate by Customer Tenure

The analysis shows a clear inverse relationship between customer tenure and churn rate (Figure 15). Newer customers (0-4 years) have the highest churn rate, which gradually decreases as tenure increases, with long-term customers (9-12 years) showing the lowest propensity to leave. This pattern indicates that customer loyalty strengthens over time, making early relationship-building crucial. PowerCO should focus intensive retention efforts on customers in their first four years, implementing dedicated onboarding programs, regular check-ins, and early-relationship incentives to help customers cross the loyalty threshold and reduce early-stage churn.

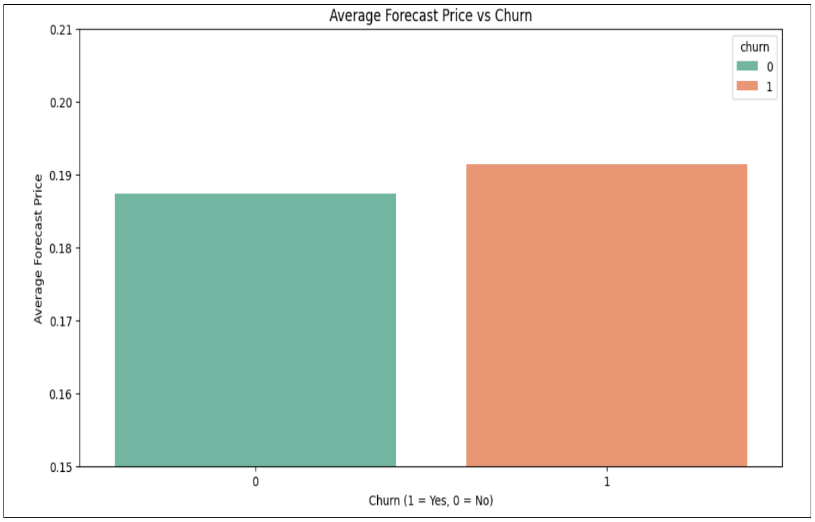

Figure 15: Average Forecast Price vs Churn

The total forecast price (combining peak and off-peak rates) is consistently higher for customers who churn compared to those who stay (Figure 16). This reinforces that overall pricing strategy significantly impacts retention, with price-sensitive customers more likely to leave when facing higher forecasted rates. This finding emphasizes that competitive pricing remains fundamental to customer retention. PowerCO should regularly benchmark their forecast prices against market competitors and consider implementing more flexible pricing models or personalized discount structures for customers identified as having high churn risk due to price sensitivity.

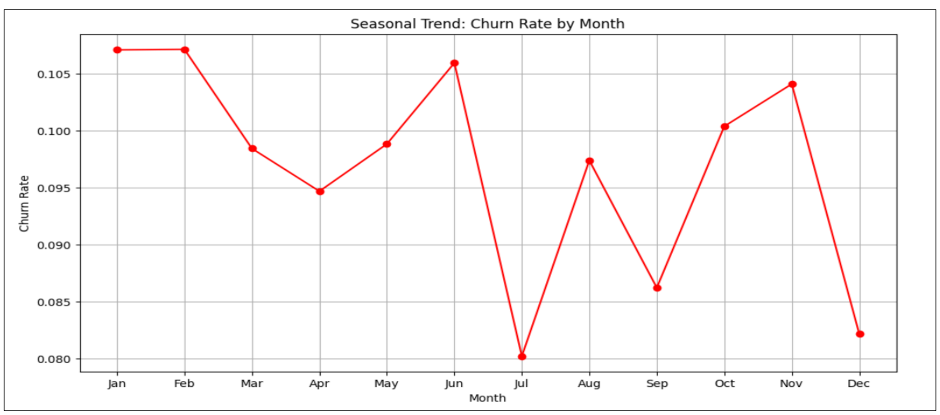

Figure 16: Seasonal Trend in Churn Rates

Our exploration of churn deepens with an analysis of the relationship between discount usage and customer retention (Figure 17). It reveals that younger customers, particularly those aged 18-35, exhibit a higher propensity to churn after utilizing discounts. While discounts effectively attract these customers, they do not seem to foster long-term loyalty. This suggests that younger customers are primarily motivated by price promotions but are less inclined to stay once the discount period ends. In contrast, older customers (46-65+) demonstrate a weaker correlation between discount usage and churn, indicating that their loyalty is influenced more by factors such as product quality, convenience, and customer service rather than temporary price reductions. These insights highlight the need for targeted post-promotion engagement strategies to convert younger, price-sensitive customers into loyal, long-term buyers.

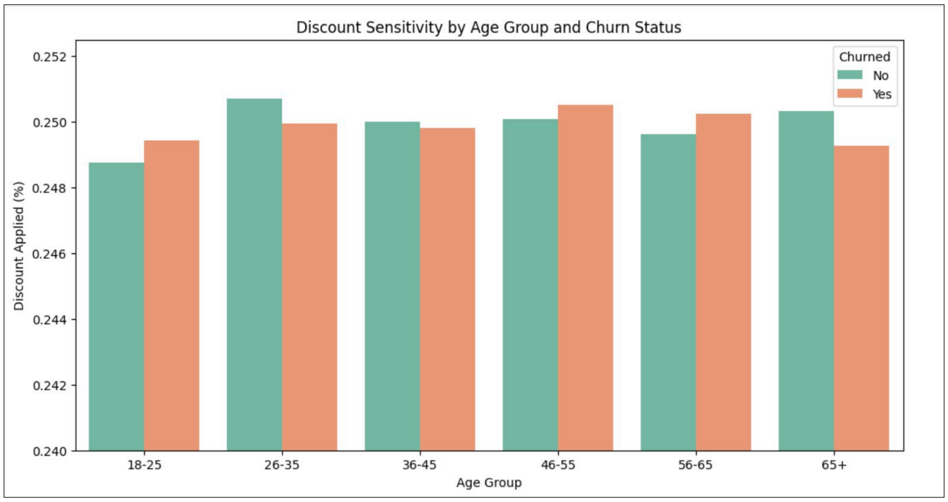

Figure 17: Discount Variation Between Churned and Non-Churned Customers Across Age Groups

A deeper look into price sensitivity across purchasing channels, illustrated in Figure 18, reveals distinct behavioral differences between online and in-store shoppers. For online shoppers, price sensitivity does not seem to significantly impact churn, suggesting that other elements, such as product availability, delivery speed, or ease of returns, play a more pivotal role in their decision to remain engaged. Conversely, in-store shoppers exhibit a stronger sensitivity to price, with higher churn rates observed when perceived value for money is not met. This finding emphasizes the importance of maintaining competitive pricing strategies and ensuring a seamless in-store experience. For brick-and-mortar retailers, aligning pricing strategies with customer expectations and enhancing value through personalized service can effectively reduce churn in this segment.

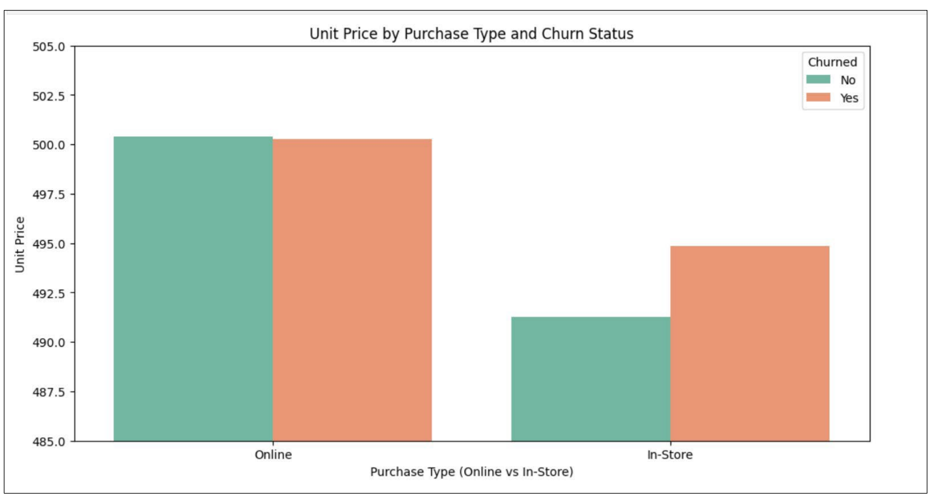

Figure 18: Impact of Unit Price on Churn for Online vs In-Store Purchases

Figure 19 shifts the focus to product categories, revealing significant differences in churn rates across various segments. Categories such as toys and clothing, often driven by impulse purchases, display higher churn rates, suggesting that customers in these segments are less likely to develop long-term purchasing habits. On the other hand, categories like groceries and electronics exhibit lower churn rates, as these purchases are typically recurring and driven by necessity or long-term utility. This pattern presents a valuable opportunity for businesses: by fostering repeat engagement in high-churn categories through targeted loyalty programs, exclusive offers, and personalized recommendations, they can encourage repeat purchases and reduce attrition. For lower-churn categories, maintaining high product availability and delivering exceptional customer service can sustain customer loyalty over the long term.

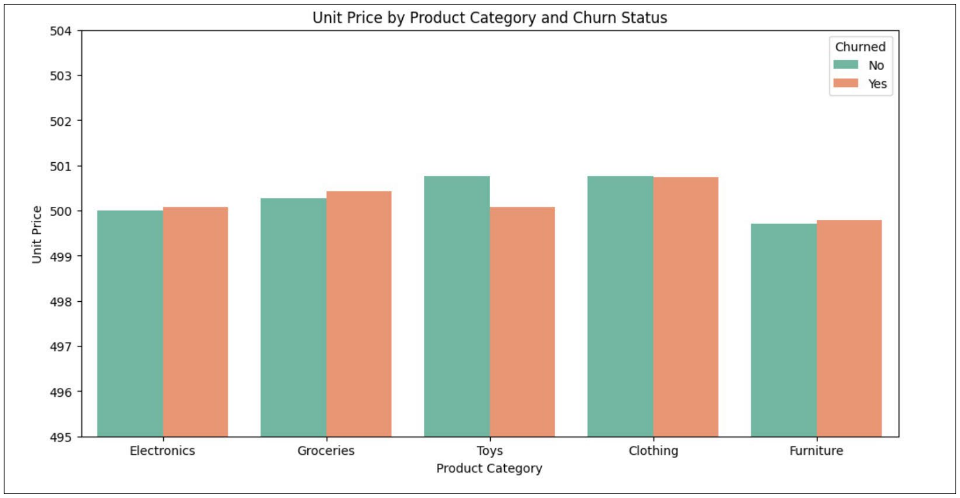

Figure 19: Unit Price Variation by Churn Status Across Product Categories

The exploratory data analysis revealed interesting patterns in customer behavior, which can be narrated as a compelling story (Figure 20).

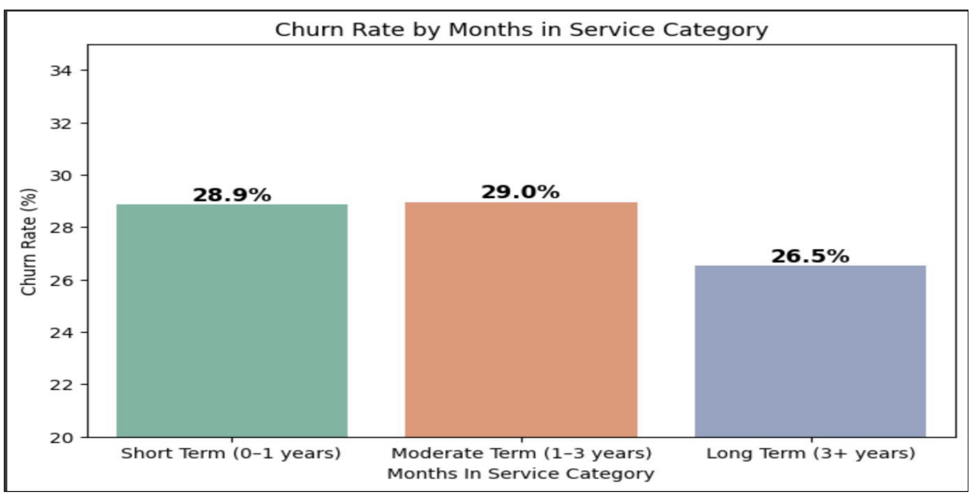

Figure 20: Churn Rate by months in service category

Imagine a new telecom customer excited about their plan. In the first few months, they explore the service but remain cautious. If their expectations are not met or if they find a better deal elsewhere, they are highly likely to leave. As shown in Figure 21, short-term customers exhibit the highest churn rate at 29.0%, particularly within the first year of service. However, as customers continue their tenure with the provider, they build trust and familiarity, leading to lower churn rates among long-term customers, which drop to 26.5%.

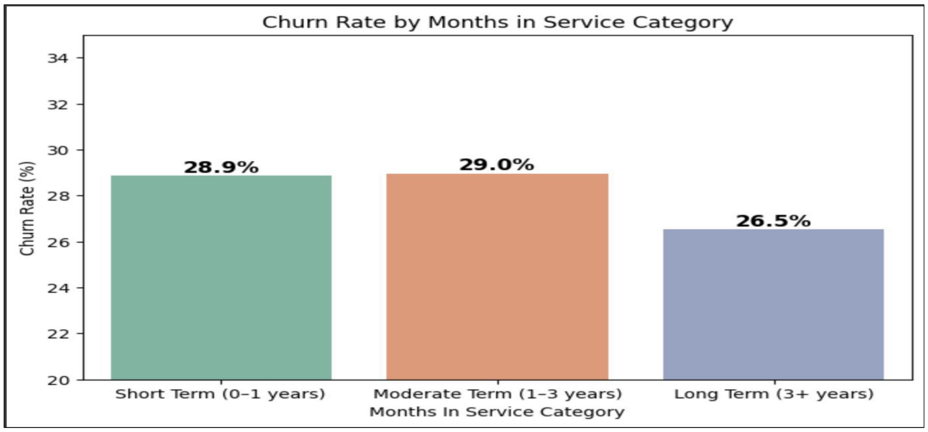

Figure 21:Churn rate by months in service category

Different occupational segments also significantly impact churn behavior, as depicted in Figure 22. Homemakers and students have the highest churn rates at 32.5% and 31.0%, respectively, as these groups are more price-sensitive, frequently attracted to promotional offers, and exhibit lower long-term commitment. In contrast, retired customers show the lowest churn rate at 25.2%, indicating a preference for stability and less sensitivity to price fluctuations. Middle-income professionals, such as clerical workers, craftsmen, and self-employed individuals, experience moderate churn rates, balancing their telecom service needs with financial constraints.

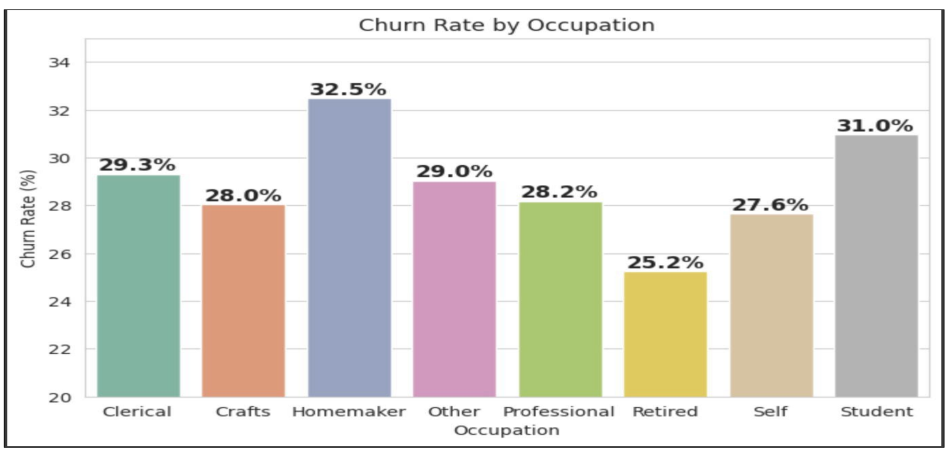

Figure 22: Churn Rate by Occupation

Subscription count further influences churn behavior, as illustrated in Figure 23. Customers with six or more subscriptions face the highest churn rate at 47.8%, primarily due to financial strain and the appeal of bundled service alternatives. On the other hand, customers with four to five subscriptions, benefiting from comprehensive service bundles, exhibit the lowest churn rate at 24.3%. Customers with only one or two subscriptions fall into a moderate churn category, as they may still be exploring different providers or canceling redundant services.

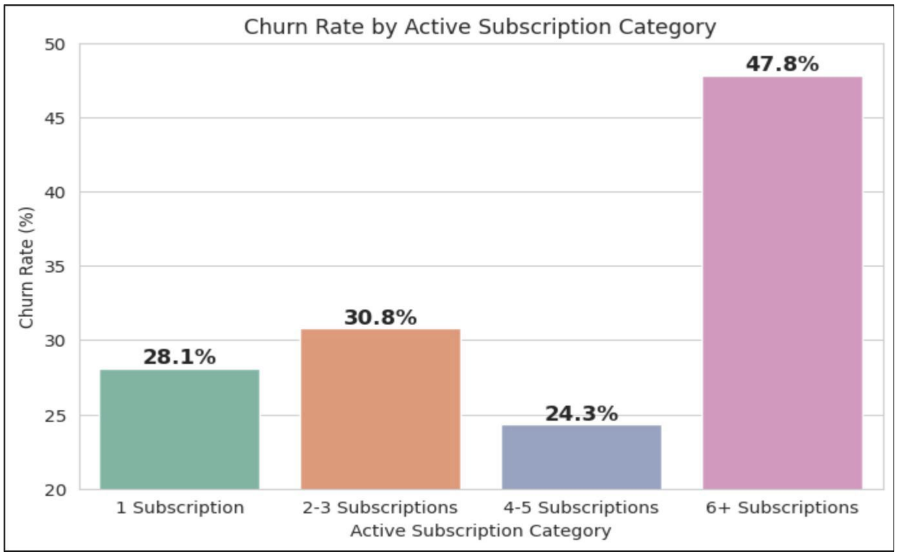

Figure 23: Churn Rate by Active Subscription Category

Financial factors also play a crucial role in determining churn rates, as detailed in Figure 24 & Figure 25. Customers with higher income levels and good credit scores have greater financial flexibility, allowing them to switch providers without worrying about penalties or credit damage. As income levels increase, churn rates decrease, suggesting that higher-income customers tend to stay with their providers longer. However, financial difficulties, such as lower income or poor credit scores, significantly increase the likelihood of churn, as customers with fewer resources are more likely to seek more affordable options.

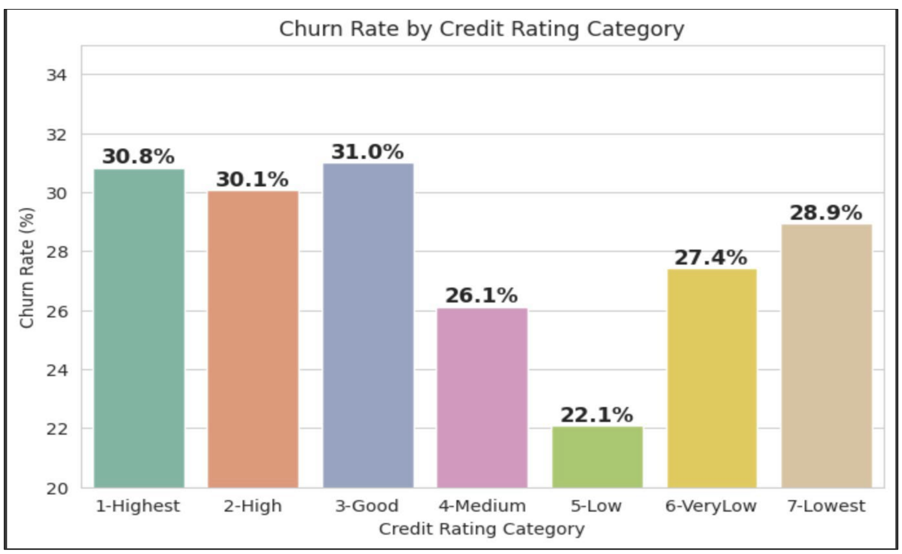

Figure 24: Churn Rate by Credit Rating Category

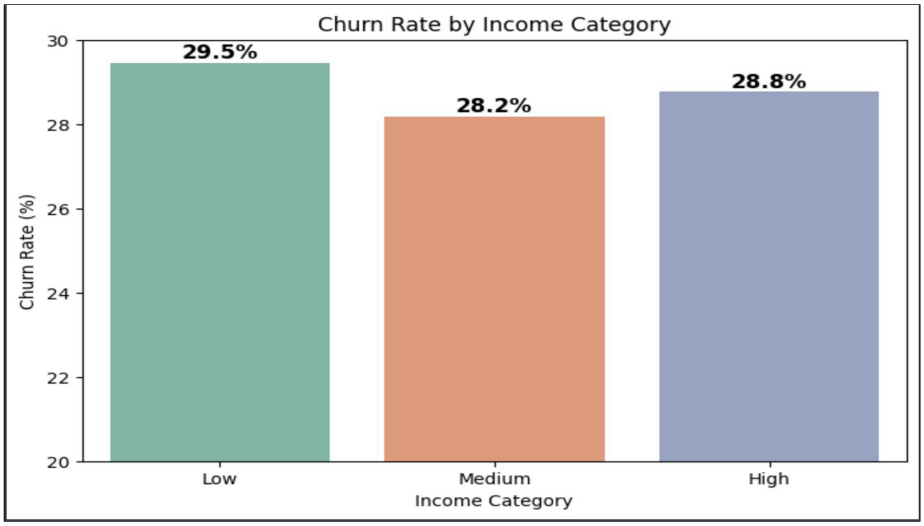

Figure 25: Churn Rate by Income Category

Imagine a new customer eagerly signing up for a telecom service. They start off excited about their plan, exploring the different services available to them. In the first few months, they assess the service quality, compare costs, and weigh their overall experience. However, if they encounter issues be it high costs, poor service, or better deals elsewhere their excitement fades, and the likelihood of switching to another provider increases.

Figure 26 highlights that customer churn peaks at nearly 50% within the first 12 months, showing that early-stage users are the most prone to leaving. As tenure extends, churn gradually declines, reaching its lowest point among customers who have stayed for 61-72 months. This pattern indicates that customer loyalty builds over time, emphasizing the importance of early retention strategies. To minimize first-year churn, telecom providers should prioritize personalized offers, seamless onboarding, and proactive support, ensuring new customers feel valued and engaged from the start.

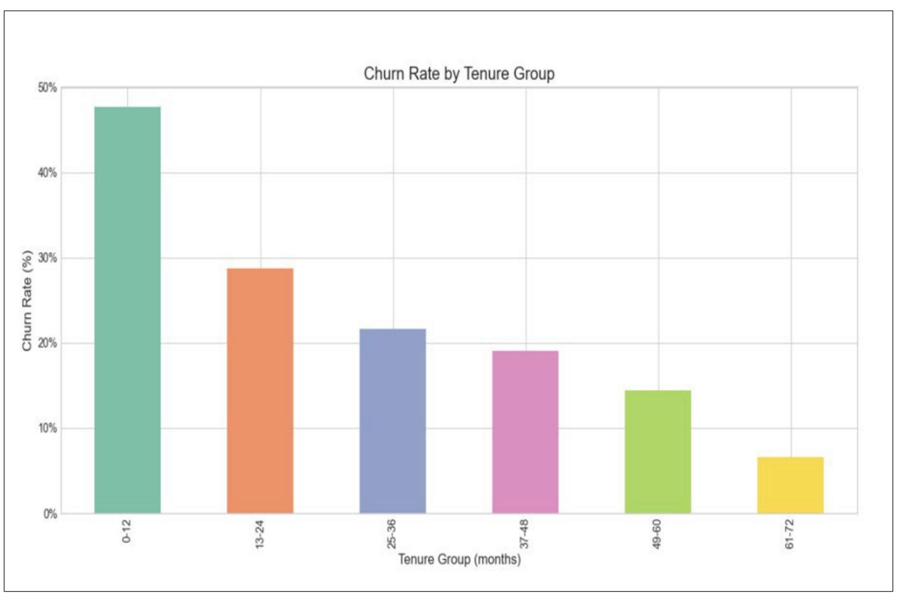

Figure 26: Churn Rate by Tenure Group

Figure 27 illustrates that customers who do not subscribe to Streaming TV experience higher churn rates, with female customers being the most likely to leave. On the other hand, those with Streaming TV tend to stay longer, regardless of gender, suggesting that bundled entertainment services play a key role in retention. This trend highlights the added value of Streaming TV, making customers less inclined to switch providers. To capitalize on this, telecom companies can offer targeted promotions to non-subscribers, encouraging adoption and ultimately improving customer engagement while reducing churn.

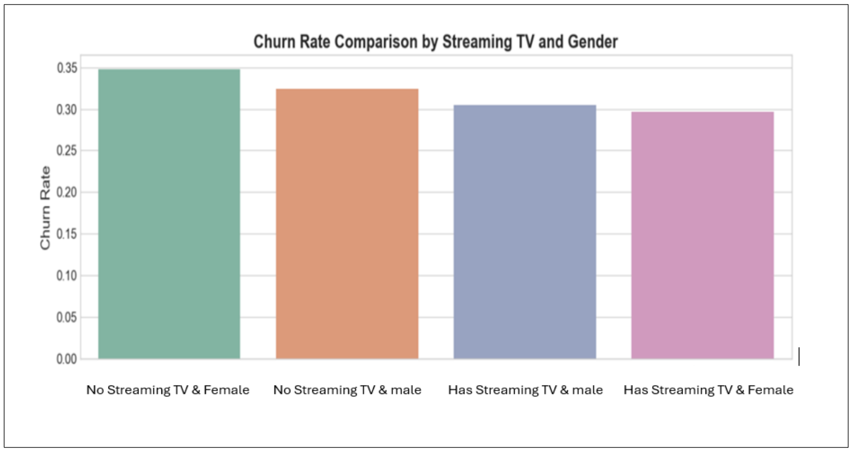

Figure 27: Churn Rate Comparison by Streaming TV and Gender

Figure 28 shows that customers who pay via electronic checks exhibit the highest churn rate, far exceeding those using other payment methods. In contrast, customers who opt for mailed checks, bank transfers, or credit cards particularly with automated payments tend to stay longer. This pattern suggests that manual payment methods may contribute to churn, potentially due to payment delays, financial uncertainty, or the inconvenience of recurring transactions. To enhance retention, telecom providers can promote automated payment options, such as bank transfers or credit card autopay, ensuring a smoother and more reliable billing experience.

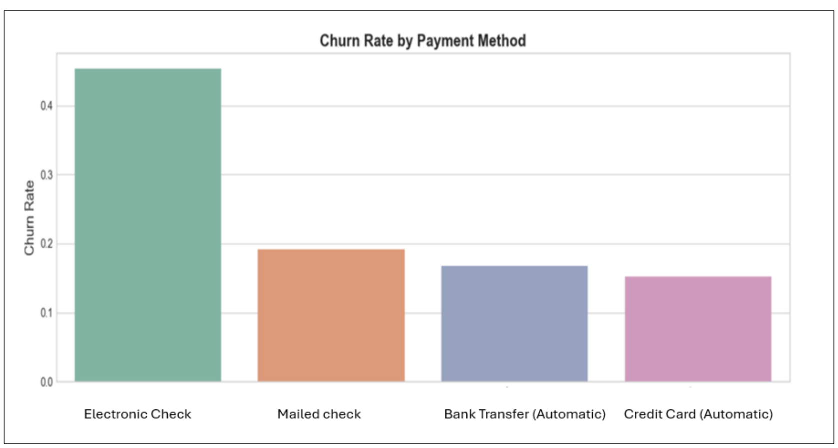

Figure 28: Churn Rate Comparison by Payment Method

Figure 29 reveals that senior citizens without dependents have the highest churn rate, making them the most likely to switch providers. In contrast, customers with dependents regardless of age demonstrate significantly lower churn, suggesting that household responsibilities contribute to service stability. This pattern indicates that family-oriented customers are more loyal, possibly due to bundled plans or shared household usage. To mitigate churn, telecom providers should implement targeted retention programs for senior citizens and independent customers, offering value-added services or long-term discounts to encourage commitment and reduce the likelihood of switching.

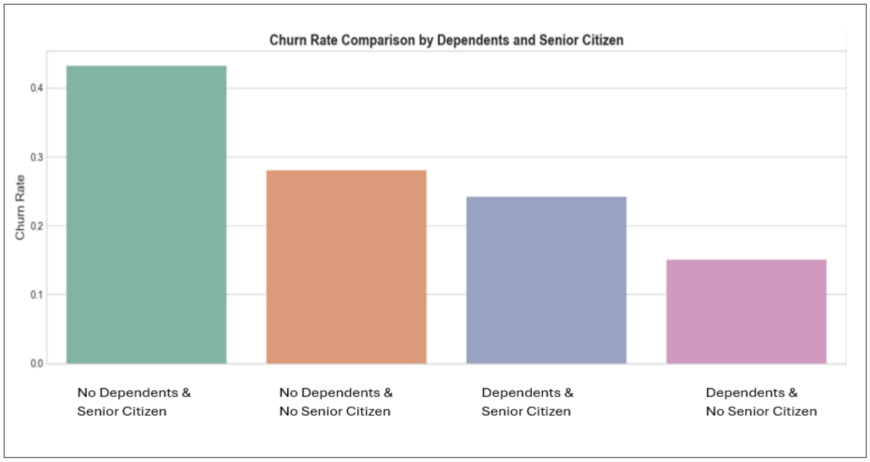

Figure 29: Churn Rate Comparison by Dependents and Senior Citizen

Figure 30 highlights that customers who churn generally face higher average monthly charges across all tenure groups, with the effect being most pronounced among long-term subscribers (49+ months), where those who leave have the highest bills. For short-term customers (0-12 months), those who churn tend to pay significantly more than those who stay, suggesting that high initial costs may be a key driver of early departures. Similarly, in the medium tenure group (13-48 months), elevated charges correlate with increased churn rates. This trend underscores the impact of pricing on customer retention. To reduce churn, telecom providers should explore discounts, loyalty rewards, or flexible pricing structures, particularly for long-term customers who may feel the financial strain of sustained high costs.

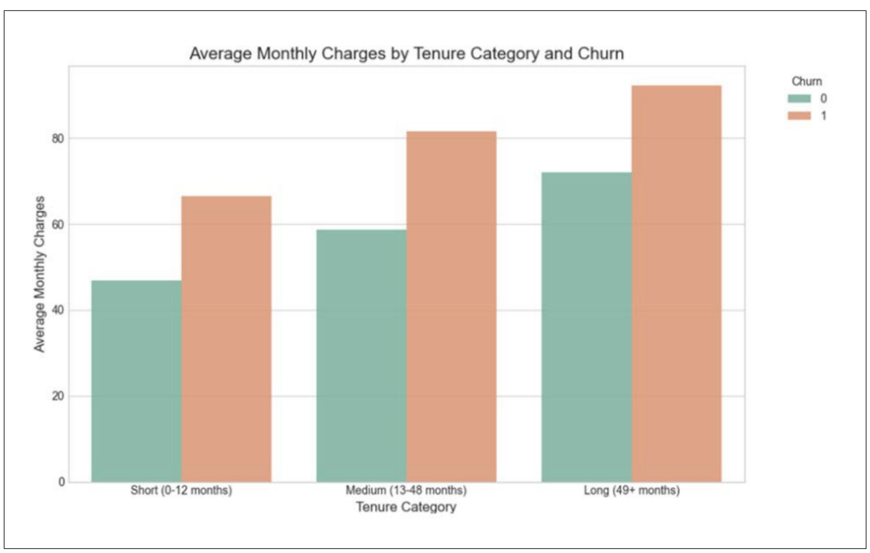

Figure 30: Average Monthly charges by tenure Category and Churn

7. LIMITATIONS

While this analysis offers valuable insights into customer behavior and churn patterns, several limitations should be acknowledged. The dataset primarily captures behavioral metrics such as app usage, website visits, and social media engagement, which may not fully reflect the depth of customer-product interaction or emotional attachment. Occasional high engagement does not necessarily indicate loyalty, while subtle signs of disengagement may go unnoticed. Additionally, product return behavior and dissatisfaction patterns lack the granularity needed to identify early indicators of churn related to product issues.

Price sensitivity analysis also faces constraints, as the dataset includes information on discounts, promotions, and total spending but does not capture segment-specific price perceptions or long-term purchasing behavior. Without detailed pricing elasticity data or competitive pricing benchmarks, it is difficult to accurately predict churn influenced by price changes. Furthermore, external factors such as competitor pricing strategies or broader economic conditions, which can significantly impact customer retention, are not accounted for, limiting the robustness of price sensitivity insights.

Moreover, the dataset may overlook critical variables such as customer sentiment, contract negotiations, and personal circumstances, which could provide deeper context for churn predictions. Assumptions made during data preprocessing, such as imputing missing values, may also introduce bias if the missing data is not random. Since the findings are based on historical data from a single telecom provider, they may not be directly applicable to other regions or competitive landscapes, and customer behavior may evolve over time due to changing industry trends.

Finally, the analysis focuses on app and desktop interactions but does not capture engagement on other platforms, such as mobile web or additional digital channels, which could influence customer behavior. The static nature of the analysis provides only a snapshot in time, potentially missing long-term trends and shifts in customer preferences. These limitations should be considered when interpreting the results and applying them to real-world business decisions.

8. RECOMMENDATIONS

Encourage Multi-Platform Engagement & Cross-Channel Consistency

To reduce churn, companies should ensure that customers experience a seamless and consistent journey across all platforms—whether it's the website, mobile app, social media, or in-store interactions. Engaging customers across multiple channels increases brand familiarity and trust. By maintaining consistent messaging and design across touchpoints, companies can create a cohesive experience that strengthens customer loyalty and reduces the likelihood of churn.

Leverage Proactive Engagement Through Personalized Communication

Personalized communication fosters stronger customer relationships by addressing individual needs and preferences. Companies should leverage data analytics and customer insights to send tailored messages, product recommendations, and exclusive offers. Proactive engagement, such as follow-up emails after purchases or timely reminders about upcoming renewals, shows customers that their business is valued. This kind of attentive service can significantly reduce churn by creating a sense of connection and relevance.

Optimize Pricing & Offer Value-Based Incentives
Competitive pricing plays a crucial role in retaining customers. Companies should continuously assess their pricing strategy to ensure it reflects the value their product or service delivers. Offering tiered pricing plans or value-based incentives, such as loyalty discounts or bundled packages, can appeal to different customer segments and encourage long-term commitment. By demonstrating value, companies reduce the chances of customers seeking alternatives.

Reduce Pricing Friction with Flexible Payment Options
Payment-related difficulties often lead to unnecessary churn. To mitigate this, companies should offer flexible payment options such as installment plans, subscription models, and digital wallet integrations. Ensuring a smooth and convenient payment process can help customers stay engaged and avoid service cancellations. Additionally, providing transparent billing and easy-to-understand terms can build trust and prevent confusion that may lead to customer dissatisfaction.

Target Short-Tenure Customers with Proactive Support
New customers are more likely to churn if they encounter challenges during the initial stages. Companies should target short-tenure customers with proactive onboarding and support initiatives. By offering personalized guidance, tutorials, and regular check-ins, businesses can address pain points early, ensuring that customers understand the product's value and feel supported. This proactive approach helps solidify trust and prevents early-stage attrition.

9. CHATGPT USAGE ELABORATION

Throughout this analysis, we leveraged ChatGPT extensively to enhance our understanding of the dataset and extract meaningful insights. Below is a structured overview of the key interactions, highlighting what worked well, where limitations were encountered, and best practices for future analyses.

Key Q&A Interactions and Outcomes

1. Important Features Analysis
Question: What are the important features in the data that affect churn?

ChatGPT Response:Based on the data provided, the most influential features affecting churn appear to be forecast_price_energy_peak, net_margin, num_years_antig (customer tenure), and overall pricing metrics. The visualizations suggest strong correlations between these factors and customer churn rates, with pricing variables showing particularly strong relationships.

Outcome:This guidance helped us focus the analysis on the most relevant variables, avoiding distractions from less significant features. It allowed for a more efficient and targeted approach when exploring factors contributing to customer churn.

2. Visualization Troubleshooting
Question: Some of our plots aren't working properly. The seasonal trend chart doesn't show clear patterns.

ChatGPT Response: For seasonal trend analysis, consider extracting month information from your date columns and creating a grouped bar chart comparing churn rates by month. Alternatively, a line plot showing churn rate fluctuations throughout the year might provide clearer visualization of seasonal patterns.

Outcome: These suggestions provided alternative approaches to visualize seasonal trends more effectively. We ultimately removed problematic visualizations and prioritized those that delivered clearer insights.

3. Interpretation Assistance
Question: What is the conclusion from each plot?

ChatGPT Response: The net margin plot reveals that churned customers have higher average margins than retained ones. This suggests the business is losing its most profitable customers, which represents a significant business risk requiring targeted retention strategies for high-value accounts.

Outcome: ChatGPT's concise and business-context-aware interpretations helped us identify critical insights, such as the counterintuitive finding that high-margin customers exhibited higher churn rates. These insights informed our recommendations for improving retention strategies.

4. Best Practices Identified
Start with Specific, Well-Defined Questions
Asking targeted questions helped focus ChatGPT's responses on actionable insights.

Break Complex Analyses into Smaller Tasks
Dividing complex questions into manageable chunks facilitated more effective exploration and validation of findings.

Validate AI-Suggested Insights Against Actual Data
While AI-generated insights provided valuable perspectives, cross-verifying them with the actual dataset ensured accuracy and relevance.

Use AI for Interpretation Assistance While Maintaining Critical Thinking
AI was most effective as an aid for interpretation and idea generation, but our judgment remained essential for nuanced decision-making.

5. Highlight Q&A with ChatGPT

How do we identify high-risk customer segments?
Provided suggestions on segmenting based on tenure, margin, and pricing.

What is the best way to visualize churn over time?
Suggested alternatives like line plots and grouped bar charts for seasonal trends.

How can we assess feature importance for churn prediction?
Recommended feature importance techniques and variable correlation analysis.

What does this anomaly in the churn rate plot suggest?
Provided possible hypotheses to test and interpret the anomaly.

6. Limitations Encountered
Technical Code Debugging
It was less effective at resolving errors without seeing the exact error messages or underlying code.

Statistical Rigor
ChatGPT couldn't perform rigorous statistical analyses or run significance tests on the data.

Advanced Visualization Implementation
Some suggested visualizations required complex coding that wasn't fully addressed by AI recommendations.

Future Considerations
Structuring Analytical Approach with ChatGPT Upfront
Engaging ChatGPT early in the process to outline a high-level analytical strategy before diving into specifics.

Generating Alternative Hypotheses for Testing
Leveraging ChatGPT to brainstorm multiple perspectives and hypotheses to validate against the data.

Focusing on Business Questions Rather Than Technical Implementation
Using AI to refine business insights while reserving technical implementation for manual validation.

By incorporating these strategies, our future analyses can be more efficient, insightful, and aligned with business goals.

10. CONCLUSIONS

In conclusion, this project has provided a comprehensive analysis of customer churn in SaaS companies using descriptive analytics. By leveraging exploratory data analysis (EDA) across multiple datasets, we identified key behavioral and transactional patterns influencing customer retention. Our findings highlight the importance of usage frequency, engagement levels, pricing sensitivity, and cross-platform interactions in determining churn risk. Additionally, segmenting customers based on demographics, purchase history, and support interactions allows for targeted retention strategies that enhance customer lifetime value and improve overall profitability.

One of the critical takeaways from this project is the significance of early intervention in preventing customer churn. Customers who exhibit declining engagement, reduced login frequency, or irregular product usage patterns can be flagged as high-risk, enabling companies to implement proactive retention strategies. Personalized outreach, customized product recommendations, and targeted promotions can significantly enhance customer satisfaction and encourage long-term loyalty. Additionally, businesses should focus on seamless onboarding experiences to ensure that new users quickly realize the value of their subscriptions, reducing early-stage churn.

Another important aspect examined in this project is pricing sensitivity and its impact on customer retention. Our analysis revealed that customers are highly responsive to pricing structures, particularly in competitive markets. By understanding how different customer segments react to pricing changes, SaaS companies can introduce flexible pricing models, tailored discounts, and value-driven service bundles to optimize revenue while reducing churn. Additionally, offering incentives such as loyalty programs, referral benefits, and exclusive service upgrades can create long-term commitment among subscribers.

To further strengthen customer retention efforts, companies should enhance their customer support strategies. Our findings indicate that unresolved support tickets and delayed responses often contribute to dissatisfaction and eventual churn. Implementing AI-driven chatbots, self-service knowledge bases, and efficient customer service frameworks can significantly improve user experience. Moreover, analyzing customer feedback and sentiment through machine learning techniques can help businesses anticipate and address concerns before they escalate into churn incidents.

Ultimately, this project underscores the importance of data-driven decision-making in tackling customer churn. As competition in the SaaS industry continues to intensify, businesses must leverage predictive analytics to stay ahead of customer attrition trends. By continuously refining retention strategies, fostering customer engagement, and adapting to evolving user expectations, SaaS companies can enhance their market position, maximize revenue, and build a loyal customer base for long-term success.

11.REFERENCE

[1] *U. K. Patel, "Retail Sales and Customer Behavior Analysis," Kaggle, Apr. 2021. [Online]. Available: https://www.kaggle.com/datasets/utkalk/large-retail-data-set-for-eda*.

[2] *B. Ermis, "Digital Marketing | E-Commerce | Customer Behavior," Kaggle, Mar. 2022. [Online]. Available: https://www.kaggle.com/datasets/ermismbatuhan/digital-marketing-ecommerce-customer-behavior.*

[3] *E. Masimov, "PowerCO," Kaggle, Jun. 2020. [Online]. Available: https://www.kaggle.com/datasets/erolmasimov/powerco.*

[4] *J. Pace, "Telecom Churn (Cell2Cell)," Kaggle, Aug. 2019. [Online]. Available: https://www.kaggle.com/datasets/jpacse/datasets-for-churn-telecom.*

[5] *C. Char, "Telco Customer Churn," Kaggle, Feb. 2018. [Online]. Available: https://www.kaggle.com/datasets/blastchar/telco-customer-churn.*