Customer Subscription-Based Digital Product: A Machine Learning to Predict Churn

BY TIARA FITRIYANI

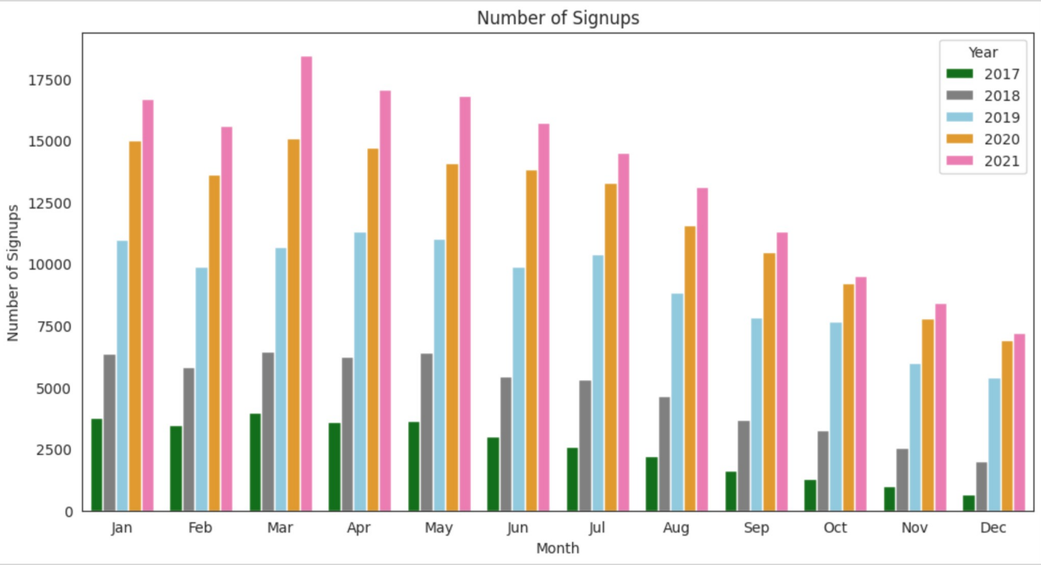
About Dataset

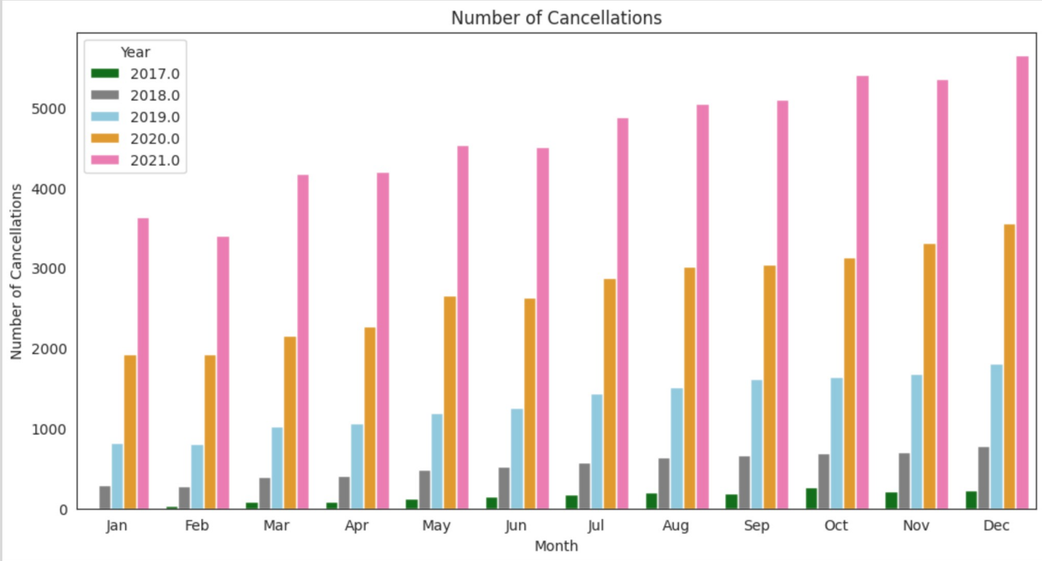
This dataset contains valuable information about a subscription-based digital product that offers financial advisory services, encompassing a wide range of features such as newsletters, webinars, and investment recommendations.

The product caters to various preferences, providing both annual and monthly subscription options. To enhance customer satisfaction, the platform offers daytime support, enabling customers to connect with a dedicated care team for product-related inquiries, as well as assistance with signup and cancellation processes.

Business Problem

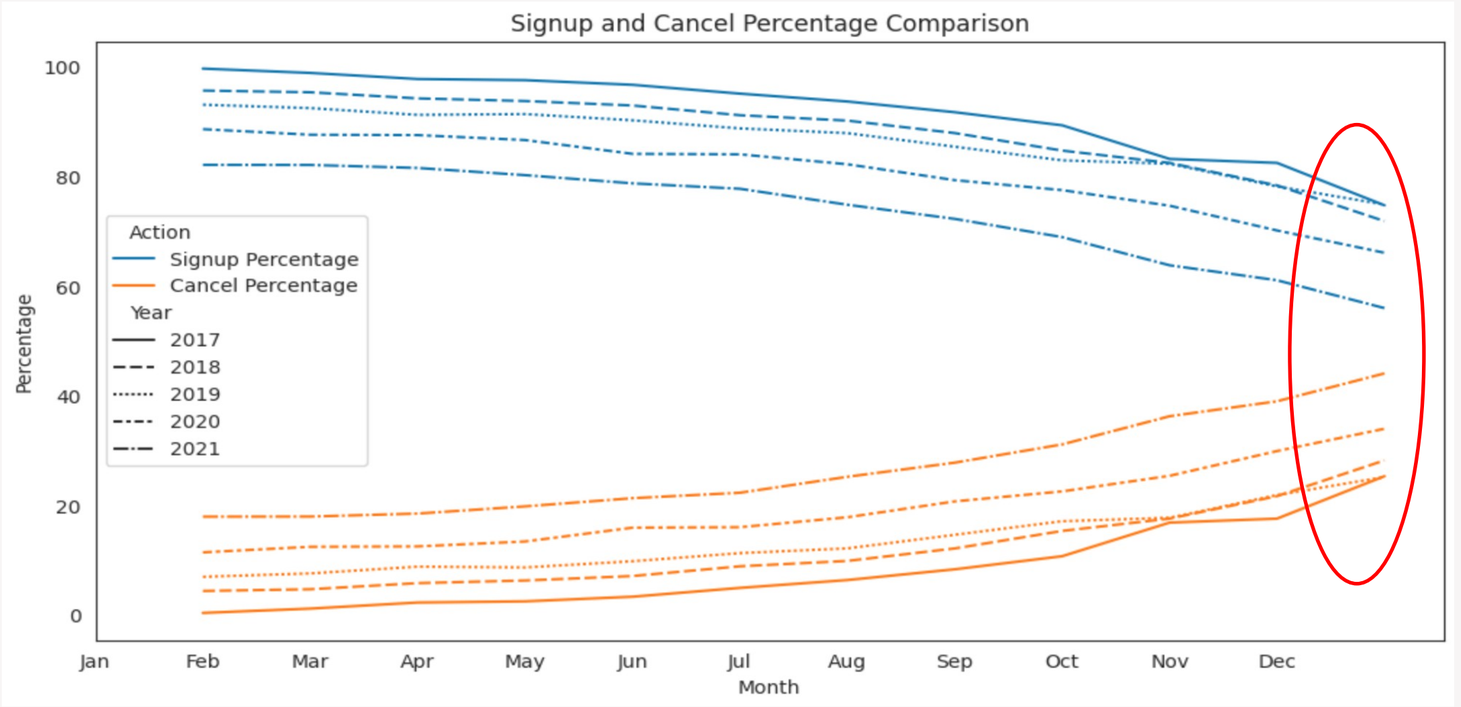
The number of signups appears to be most significant during the **first to third quarters in each year**, and despite a decrease in signups towards the end of the year, the overall trend shows a **consistent increase in the number of signups from year to year**. This is quite promising, isn't it?





The number of **cancellations** has **also increased**, and in fact, they consistently show an **upward trend every month and every year!**

Let's take a look at the **percentage comparison.**



As the years progress, **the percentage of signups continues to decline, while the percentage of churn keeps rising**. In fact, in the year 2021, both percentages are nearly equal! This is a significant concern. Hence, the company needs to predict customer churn behavior.

Objectives

Analysis

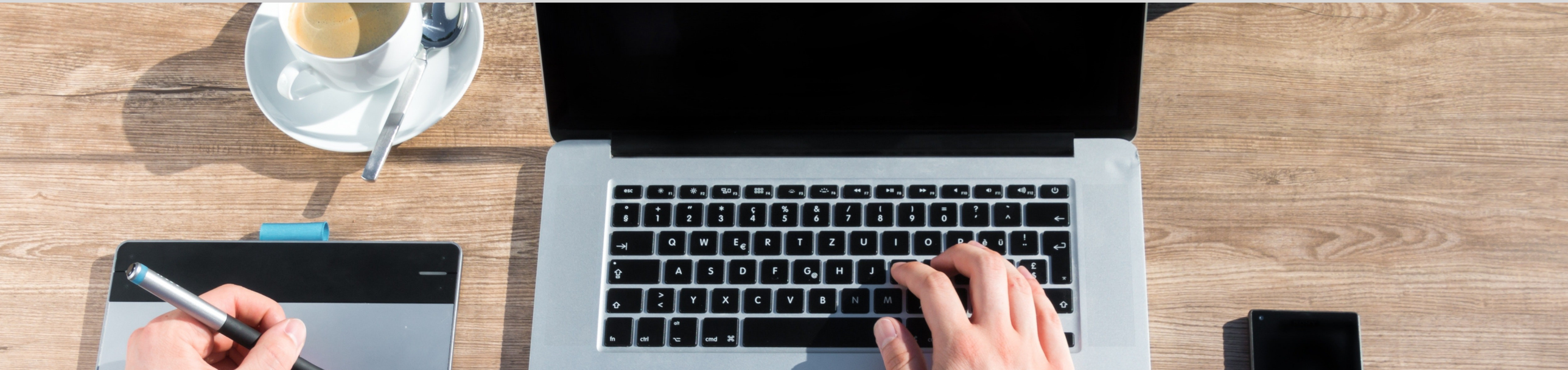
Understand more about customer behavior that occur on the platform.

Recommended Action

Product enhancement and maximizing case services based on exploratory data analysis results.

Prediction

To predict customer churn.



Data Understanding

This dataset contains 4 tables; Customer Case, Customer Info, Customer Product and Product Info.

Customer Case

Tabel customer case berisikan data customer dan case yang di alami. Terdapat 330.512 data dan 5 kolom (case_id, date_time, customer_id, channel, reason)

Customer Info

Tabel customer info berisikan data informasi customer. Terdapat 508.932 data dan 3 kolom (customer_id, age, gender).

Customer Product

Tabel customer product berisi data customer dan product yang digunakan. Terdapat 508.932 data dan 4 kolom (customer_id, product, signup_date_time, cancel_date_time).

Product Info

Tabel product info berisi data informasi mengenai product yang ditawarkan perusahaan. Terdapat 2 data dan 4 kolom (product_id, name, price, billing_cycle).

Data Preprocessing

Feature Integration by Merge Data

All tables are merged using the 'merge' method, resulting in a comprehensive single DataFrame with a total of 508,932 rows and 13 columns with no duplicated data.

Missing Value

After merged, all of feature has no missing value except '*cancel_date_time*' because not all of the customers canceled their subscription.

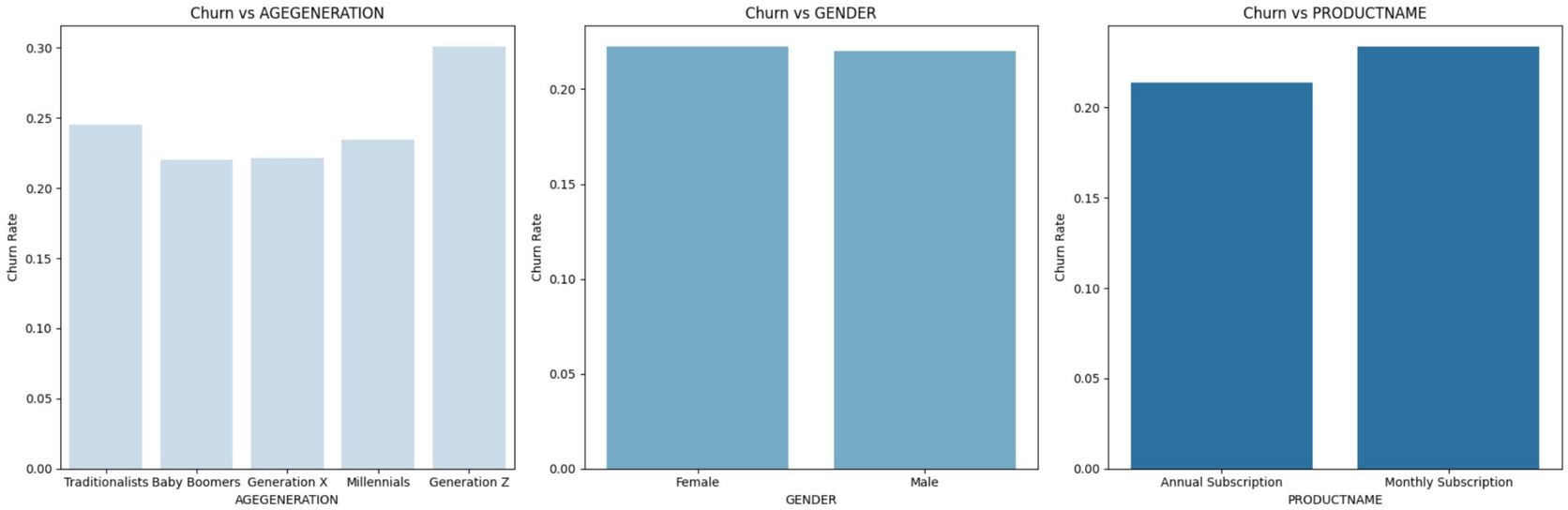
Exploratory Data Analysis & Visualization

Univariate, bivariate and multivariate analysis.

Feature Engineering

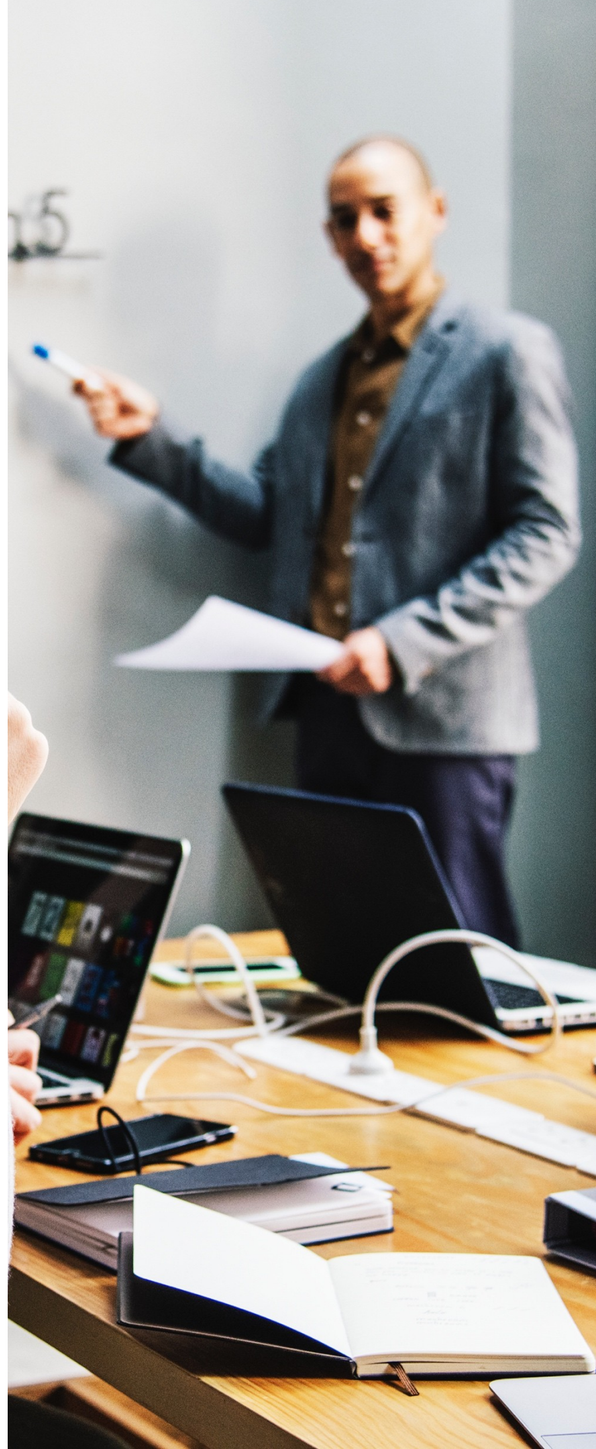
Converting categorical data into numeric using label encoding method.

Influence of Age Generation, Gender, and Products on Churn Behavior

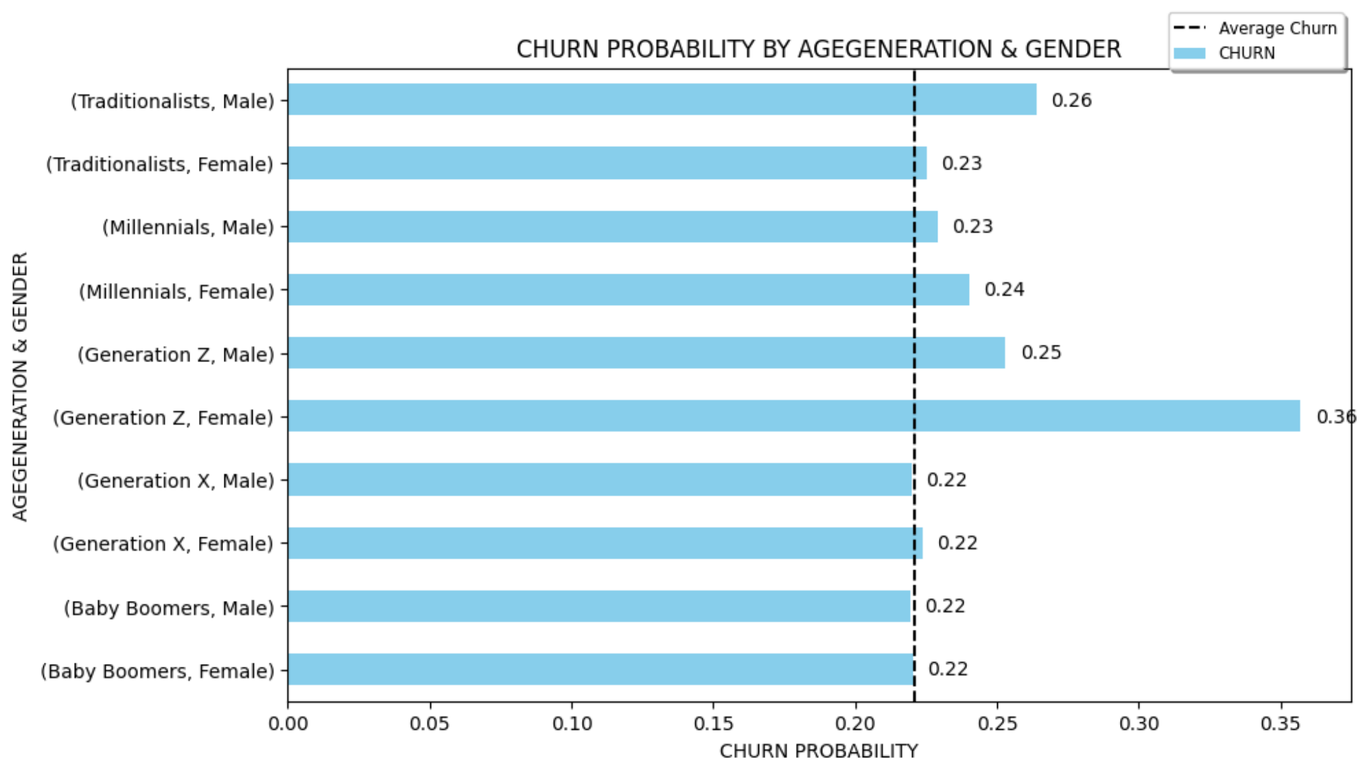


The churn rate appears to be **higher among customers belonging to Generation Z, female, and monthly subscription product.**

Let’s check them out!



Churn Probability by Age Generation and Gender



Fyi:
Baby boomers = 57 – 75 yo
Traditionalist = 41 – 56 yo
Millennials = 25 – 40 yo
Gen Z = 9 – 24 yo
Alpha = 10 - now

Action

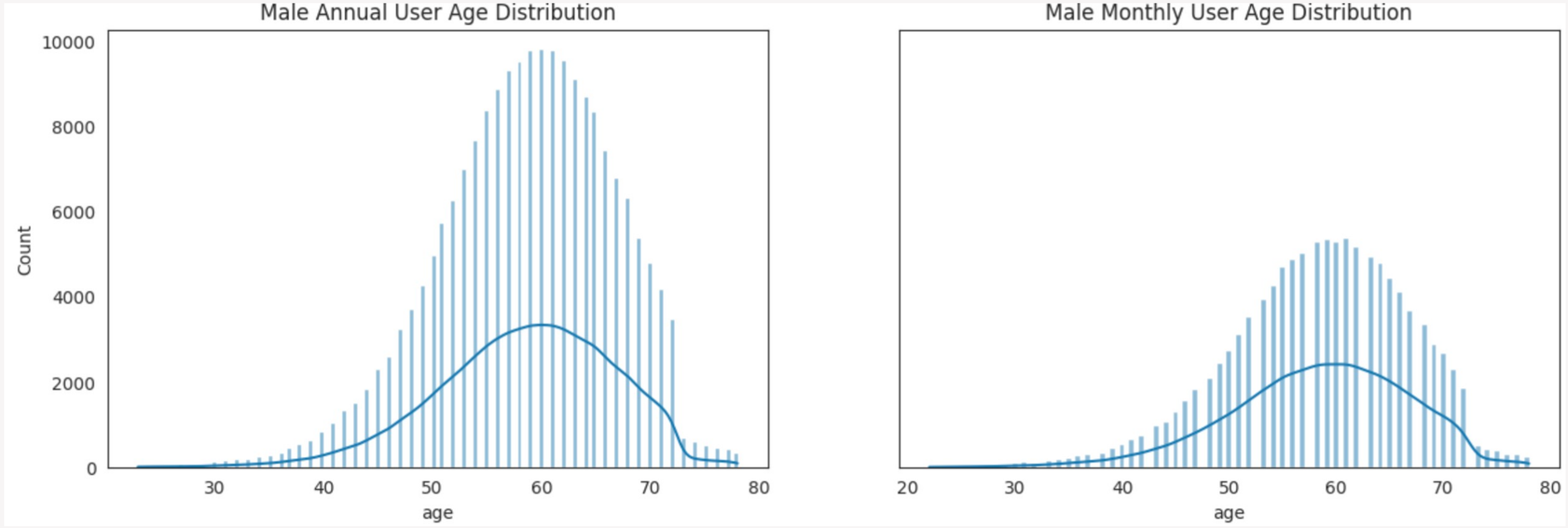
Develop targeted marketing campaigns and offers tailored to the preferences of these segments.

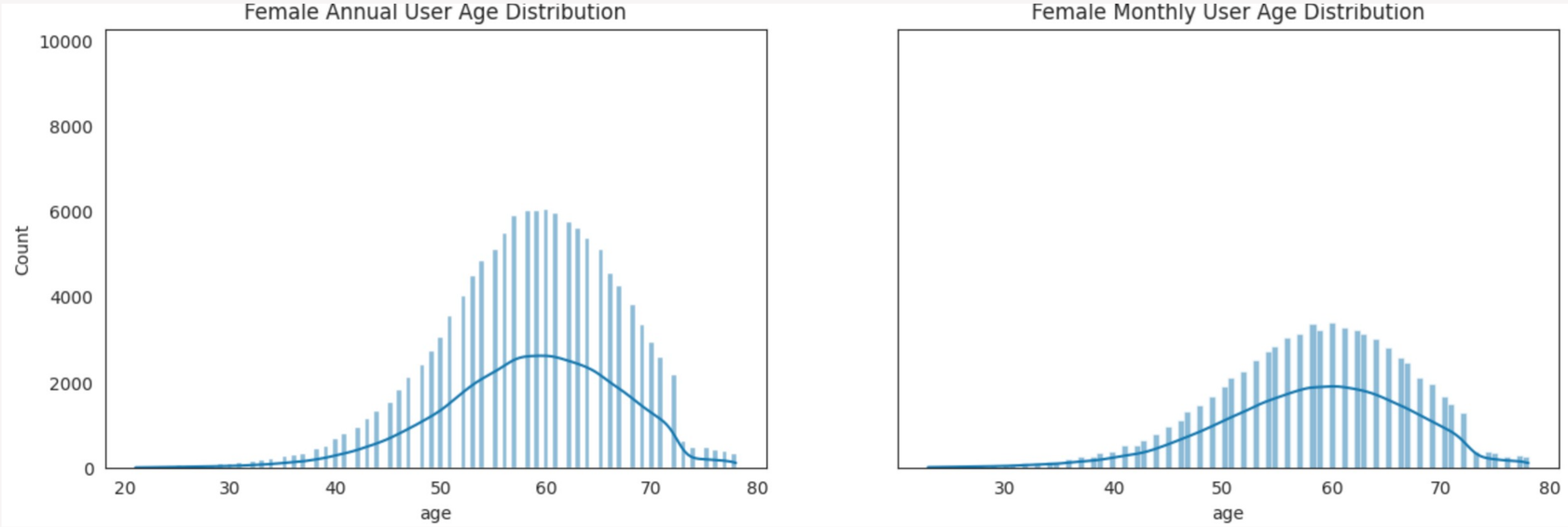
Both **male and female Gen Z customers exhibit a relatively high likelihood of churn**, particularly among females. Additionally, Traditionalist and Millennial generations also experience churn probabilities above the overall average.

Product Market Share

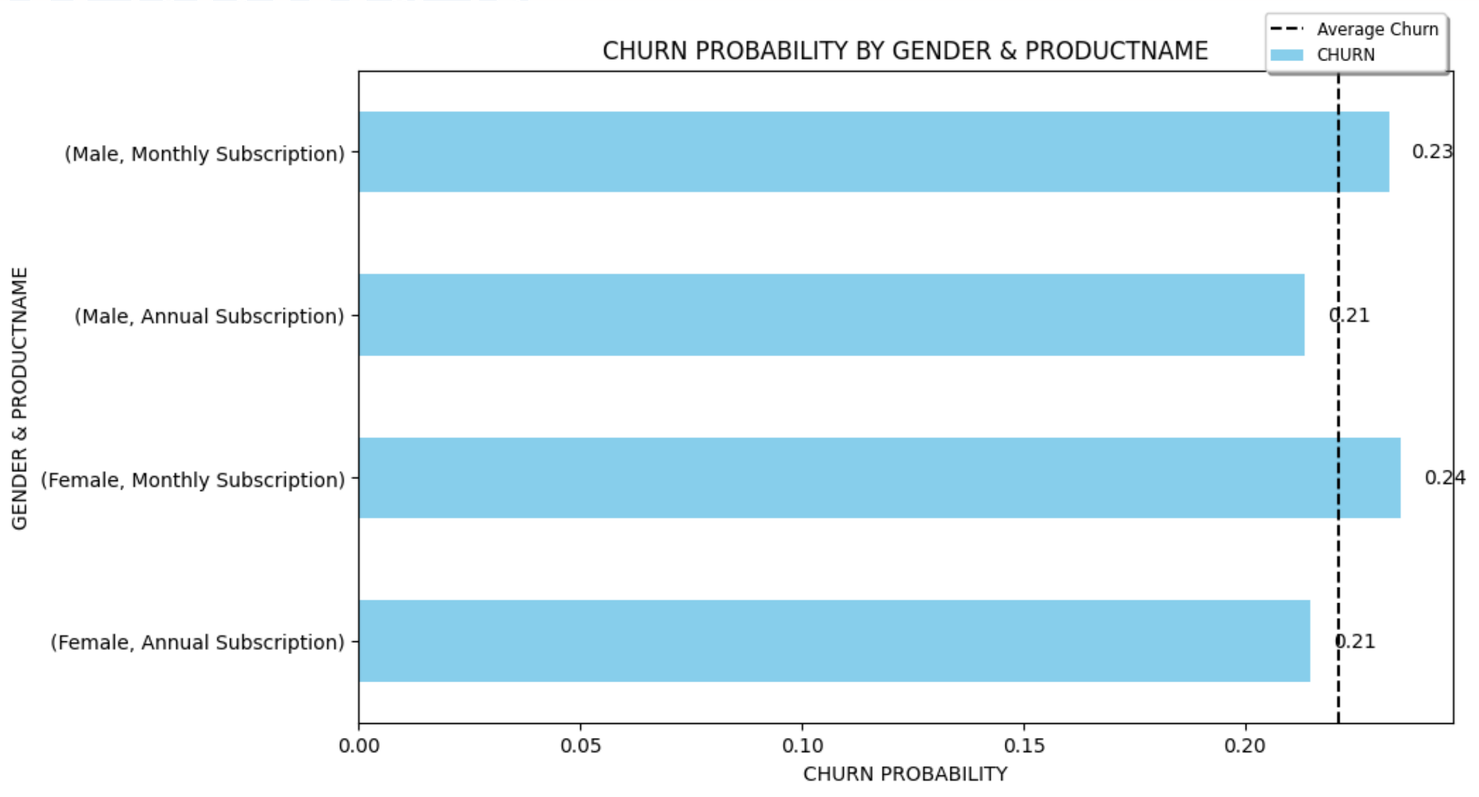
It appears that the company's offerings are more appealing **to customers aged 50-70.**

The majority of customers are **male** and the most frequently used product is **the annual subscription.**





Churn Probability by Gender and Products



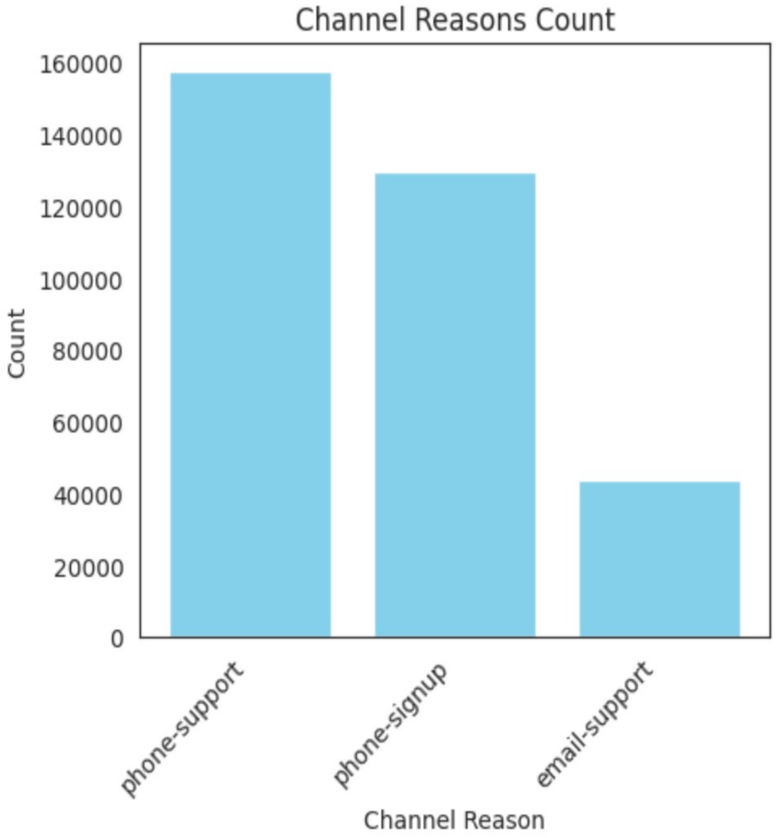
Both male and female customers with a **monthly subscription product** have a higher churn rates than the overall average churn percentage.

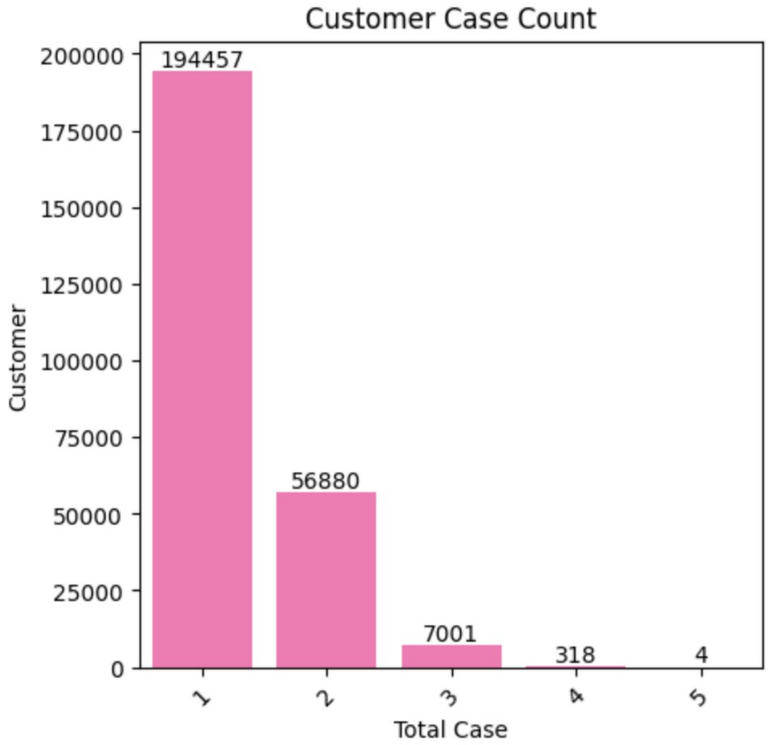
Action

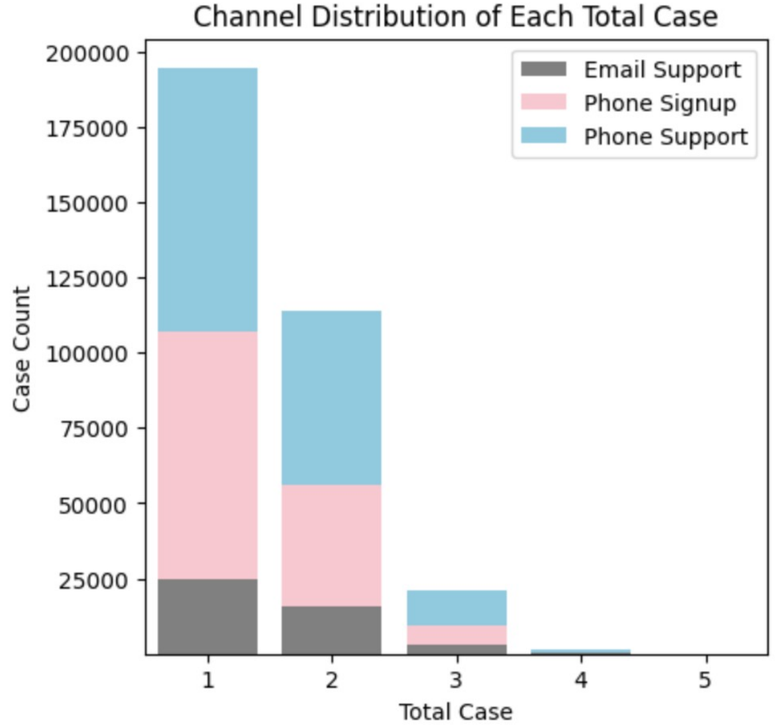
Intensify marketing efforts and tailor offers specifically for this age group. However, consider expanding to a broader market and appealing to other age groups.

Understanding the different types of Channel-Reason in Customer Cases

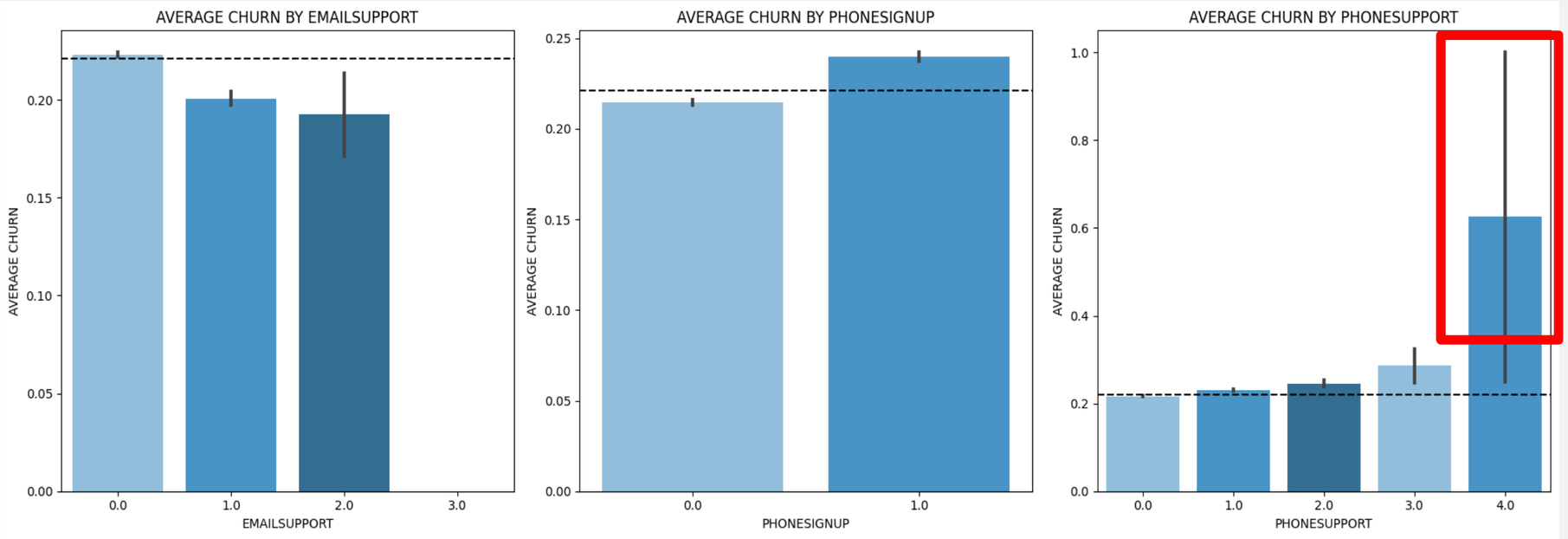
There are **three types of channel reasons** that customers use, with each customer having a **minimum of one and a maximum of five cases**, and the **most commonly used channel reason is phone support.**



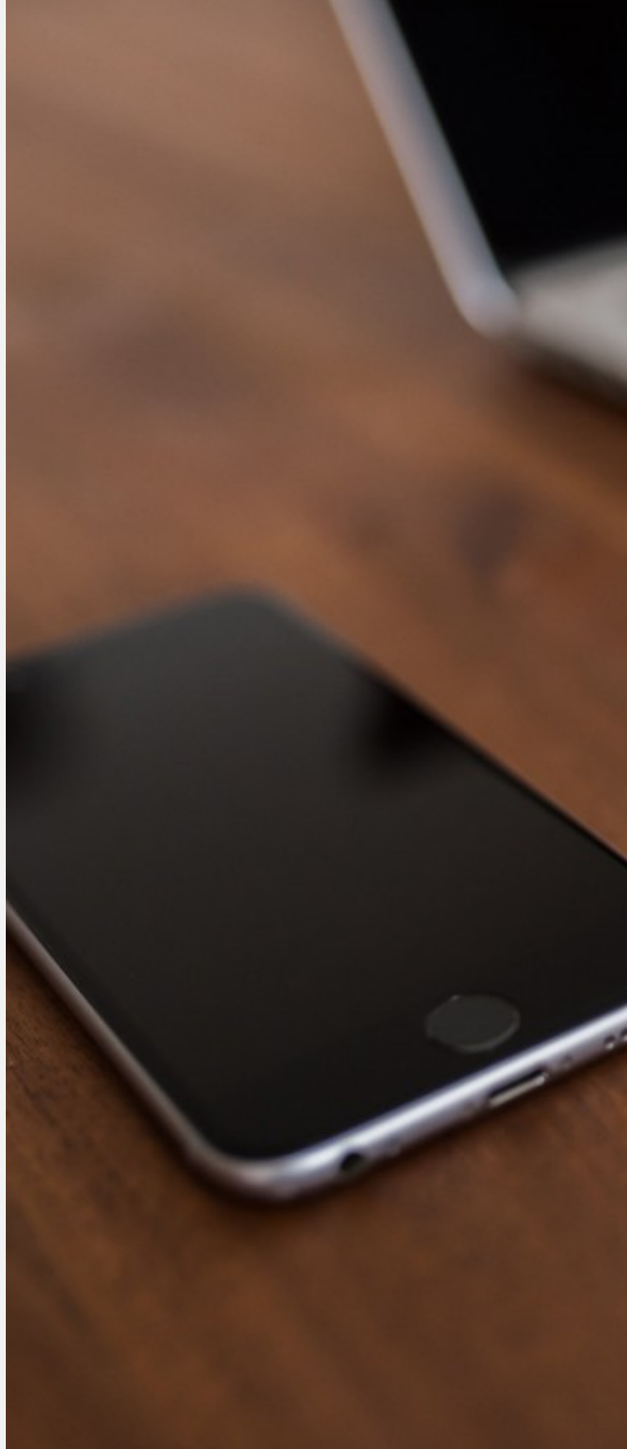


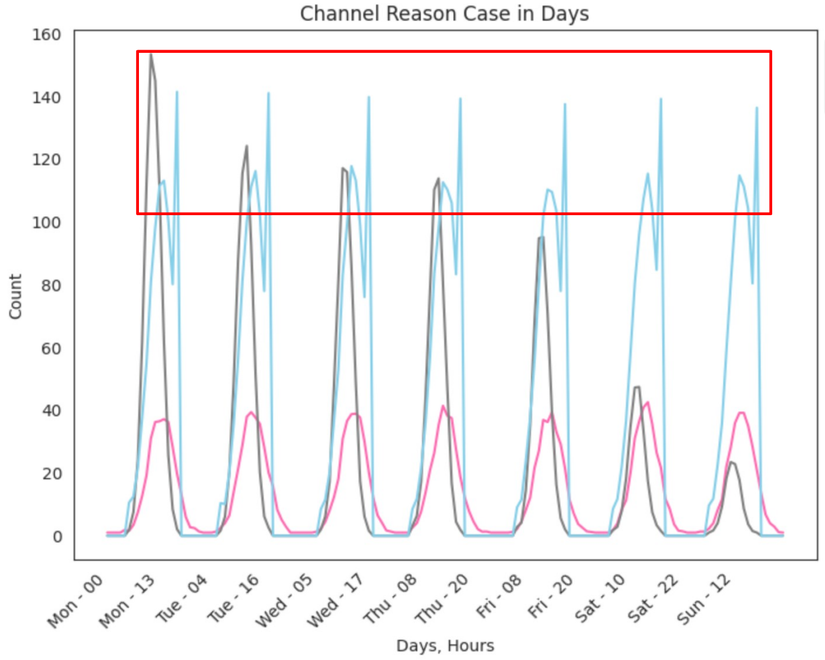


Unveiling Churn Insights via Customer Case Analysis



This clearly indicates that **phone support services require more attention**, **as the higher the number of cases, the higher the churn rate observed.**



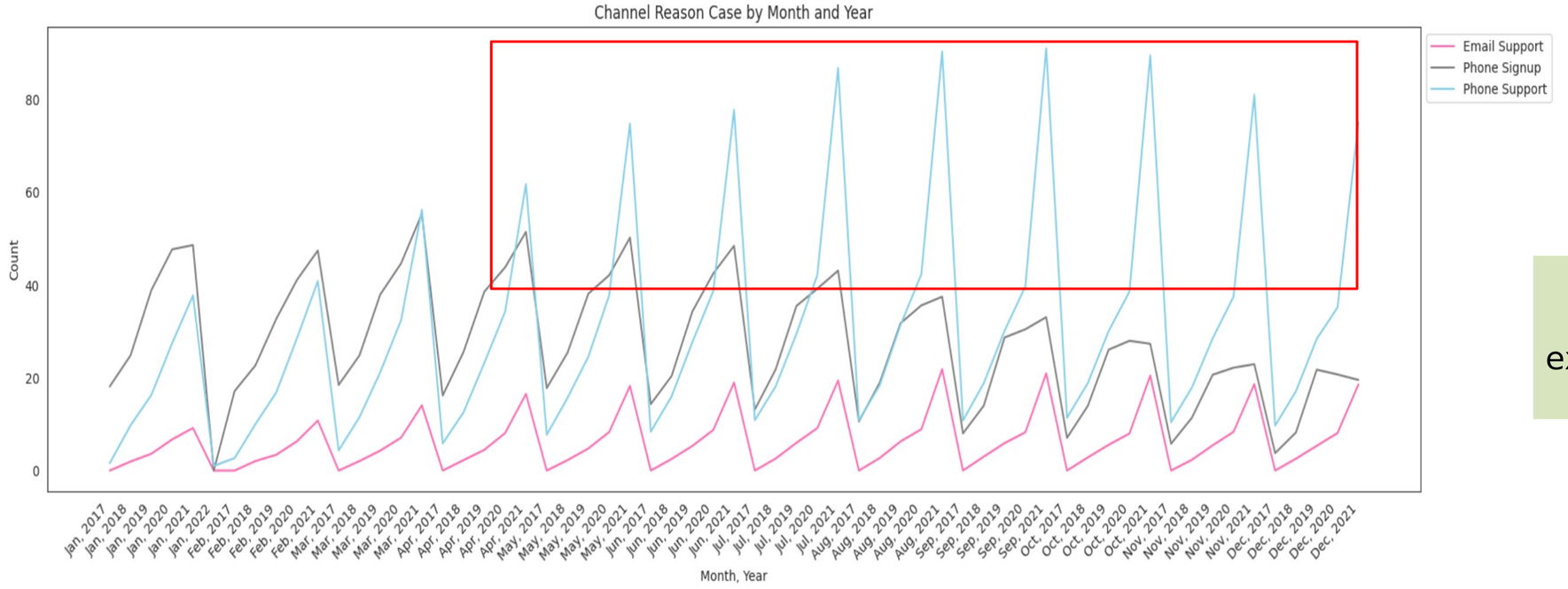


The timing pattern of Customer Cases

The number of cases consistently **peaks in the afternoon and evening hours**, specifically on the **phone support channel**.

Customers **tend to sign up on weekdays** and show an **increase in the first half of each year**.

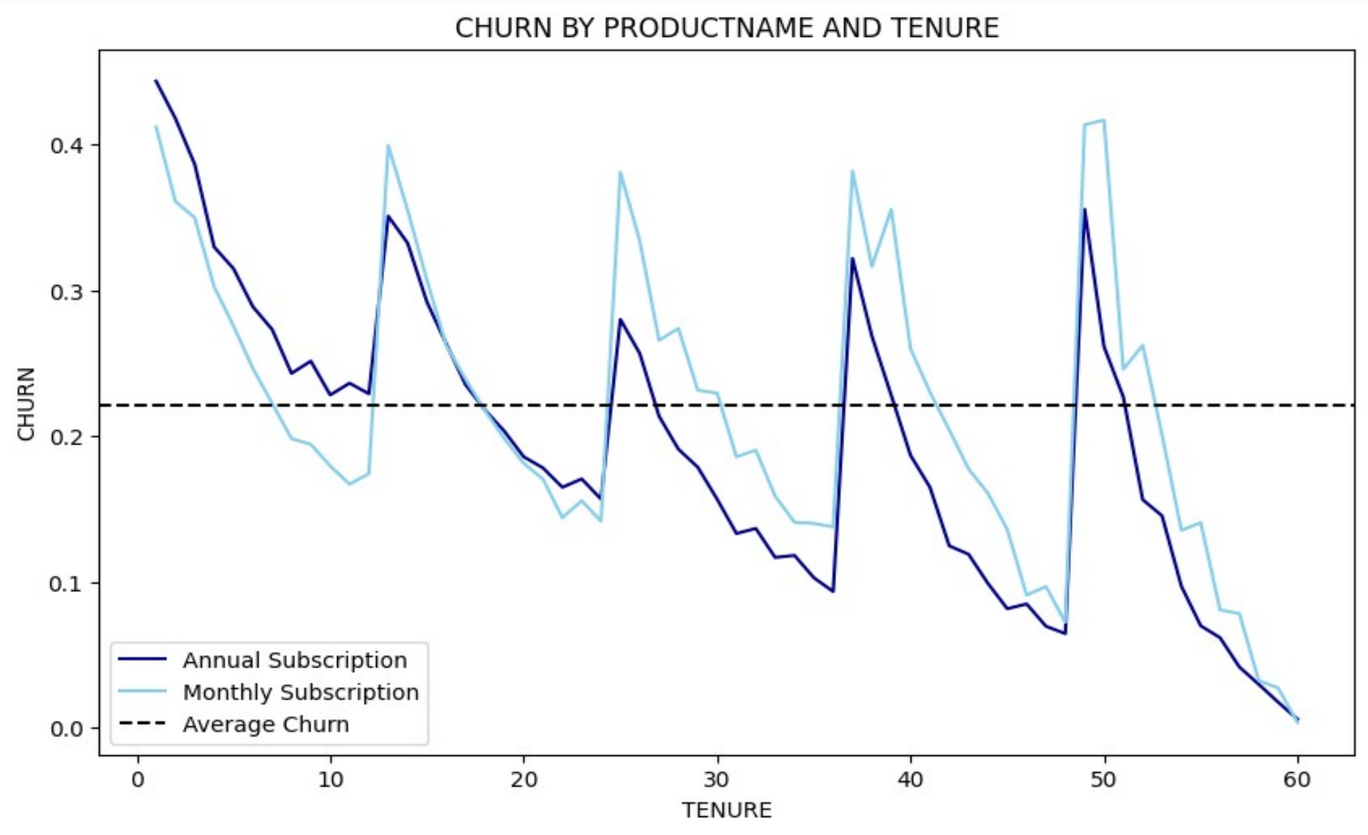
Every year, **the cases experience a rise during the mid-year period, extending towards the end of the year.**



Action

Strengthen the team and service to address these challenges. Ensure extended working hours and adequate training for customer service staff.

Customer Product and Tenure Role in Churn



A significant portion of annual subscription customers make only one purchase and subsequently discontinue their subscription in the subsequent year.

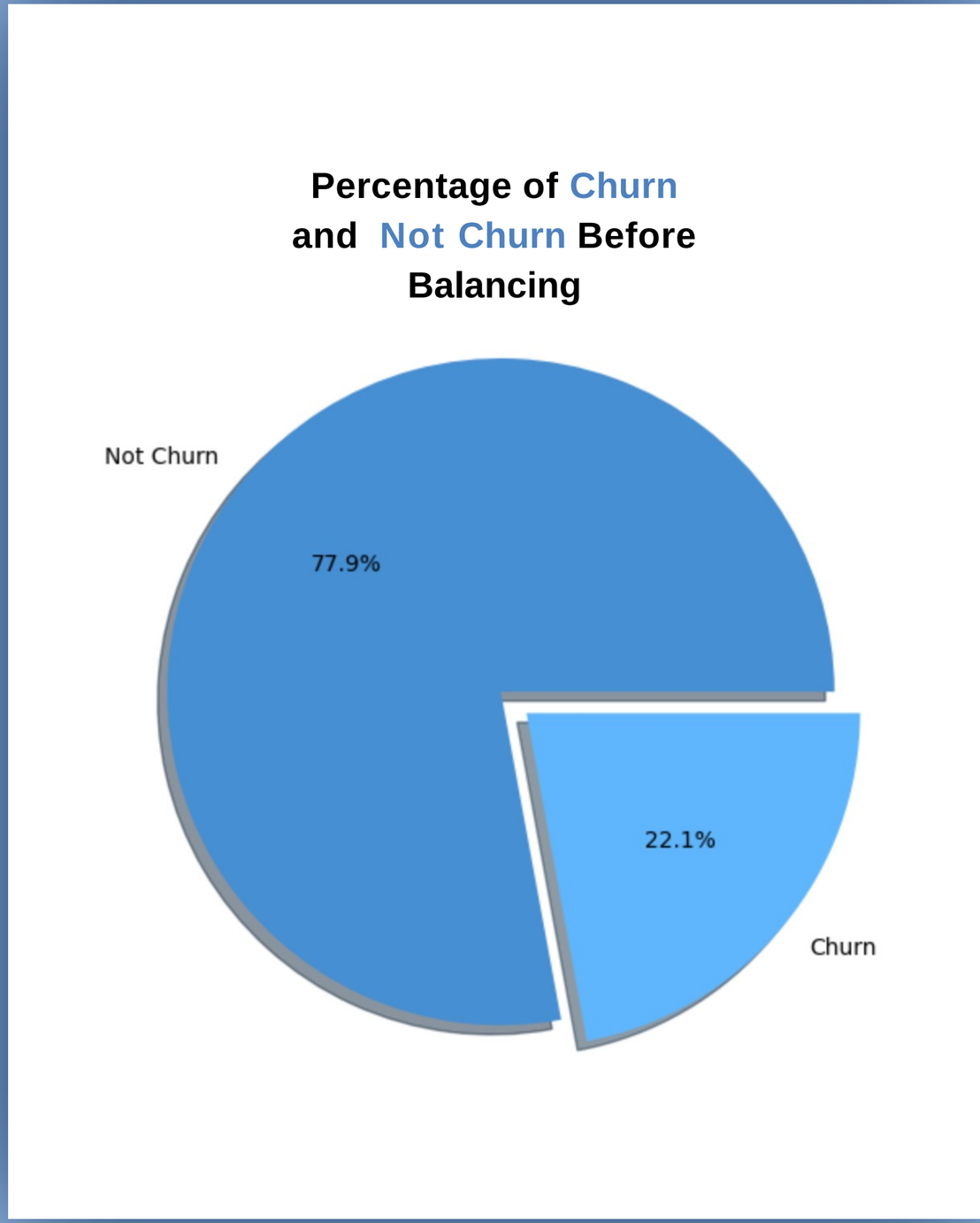
Action

Re-evaluate the annual and monthly subscription products. Consider flexible strategies that align more closely with customer preferences. To prevent customers from making single purchases and discontinuing their subscriptions (especially in annual subscription product), consider additional offers, discounts or exclusive benefits that encourage customers to renew their subscriptions.

Machine Learning Model

Training and Addressing Imbalanced Data

The training process involved utilizing 80% of the dataset, and to tackle the issue of imbalanced data, the SMOTE oversampling technique was implemented. This method is utilized to increase the minority class synthetically.



Model Evaluation and Selection: Prioritizing Recall for Effective Churn Prediction

Recall measures correctly identified churn cases out of the total actual churn cases. The goal is to capture as many churn instances as possible, even if non-churn cases are occasionally misclassified. I prioritize reducing missed churns over incorrect non-churn classifications. Simply put, it's better to mistakenly label a non-churning customer as churned than to miss an actual churning customer.

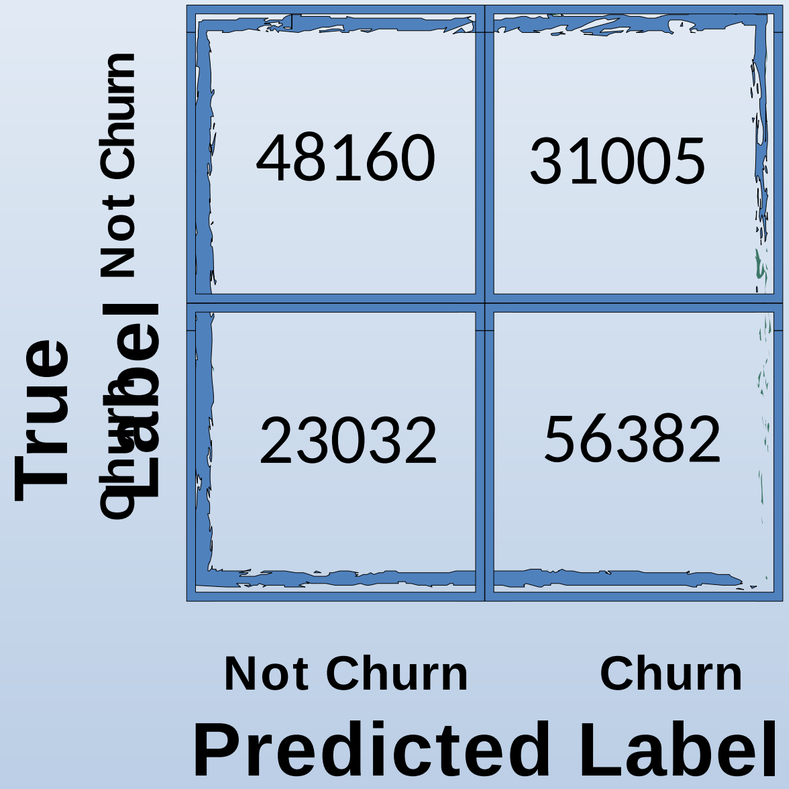
Model	Accuracy	Recall	AUC	Precision	F1 Score
dummy_classifier	0.499215	0.000000	0.500000	0.000000	0.000000
k_nearest_neighbors	0.626262	0.617788	0.626275	0.629189	0.623436
logistic_regression	0.585323	0.638766	0.585239	0.577763	0.606735
random_forest	0.684025	0.687952	0.684019	0.683263	0.685599
gradient_boosting	0.638597	0.689035	0.638517	0.626541	0.656304
xgboost	0.677807	0.677211	0.677808	0.678706	0.677958

Best Possible Gradient Boosting Classifier Model with Hyperparameter Tuning using GridSearchCV

	precision	recall	f1-score	support
0	0.68	0.61	0.64	79165
1	0.65	0.71	0.68	79414

The model effectively recognizes 71% of customers who genuinely churn, amounting to 56,382 customers. Additionally, it labels 61% of non-churning customers as churn, aiming to focus on retaining those who might genuinely churn.

Recall Score = 0.71



Modeling Conclusion

Through the application of SMOTE to our imbalanced dataset, we effectively tackled the problem of uneven labels, leading to an increased count of churn as opposed to non-churn.

The utilization of HYPERPARAMETER TUNING in the GRADIENT BOOSTING CLASSIFIER modeling has led to the model achieving favorable results, correctly predicting 61% of non-churn instances as churn, and accurately identifying 71% of churn cases.

Profound Impact

1. Strategic expansion into market segments encompassing diverse customer generations.
2. Enhancing customer experience through top-notch service to elevate customer satisfaction.
3. Optimizing product and service quality to meet customer expectations.
4. Prevent 61% customers to churn and ensure their loyalty.

The battle against churn necessitates steadfast dedication to harness the most recent methodologies and strategies to uphold the stability of customer

An ounce of prevention is worth a pound of cure.

– Benjamin Franklin