

UNDERSTANDING CUSTOMER CHURN



Company Customer Churn Analysis

T by Tejeswar Reddy Nalijeni

Problem Description:

The data is related to a company where they have various subscription plans based on tenure and provide support calls to customers if they have any issues or if there is any churn. Our goal is to predict the reasons for customer churn (variable y) and analyze the data to increase customer subscriptions.

Business Understanding:

The main goal of this project is to predict the reasons behind customer churn in subscriptions using recorded data. This involves optimizing marketing efforts, improving customer engagement strategies, and ultimately boosting subscription numbers. By utilizing historical data for binary classification, the project seeks to accurately identify potential subscribers, understand the causes of churn, and increase customer retention through targeted and informed promotions.

Data Information:



Data Extraction and Analysis

In this dataset, we will use the test data to evaluate several machine learning models and identify the top-performing ones.



Model Evaluation and Selection

From these, we will select the best two or three models to train on the full dataset.



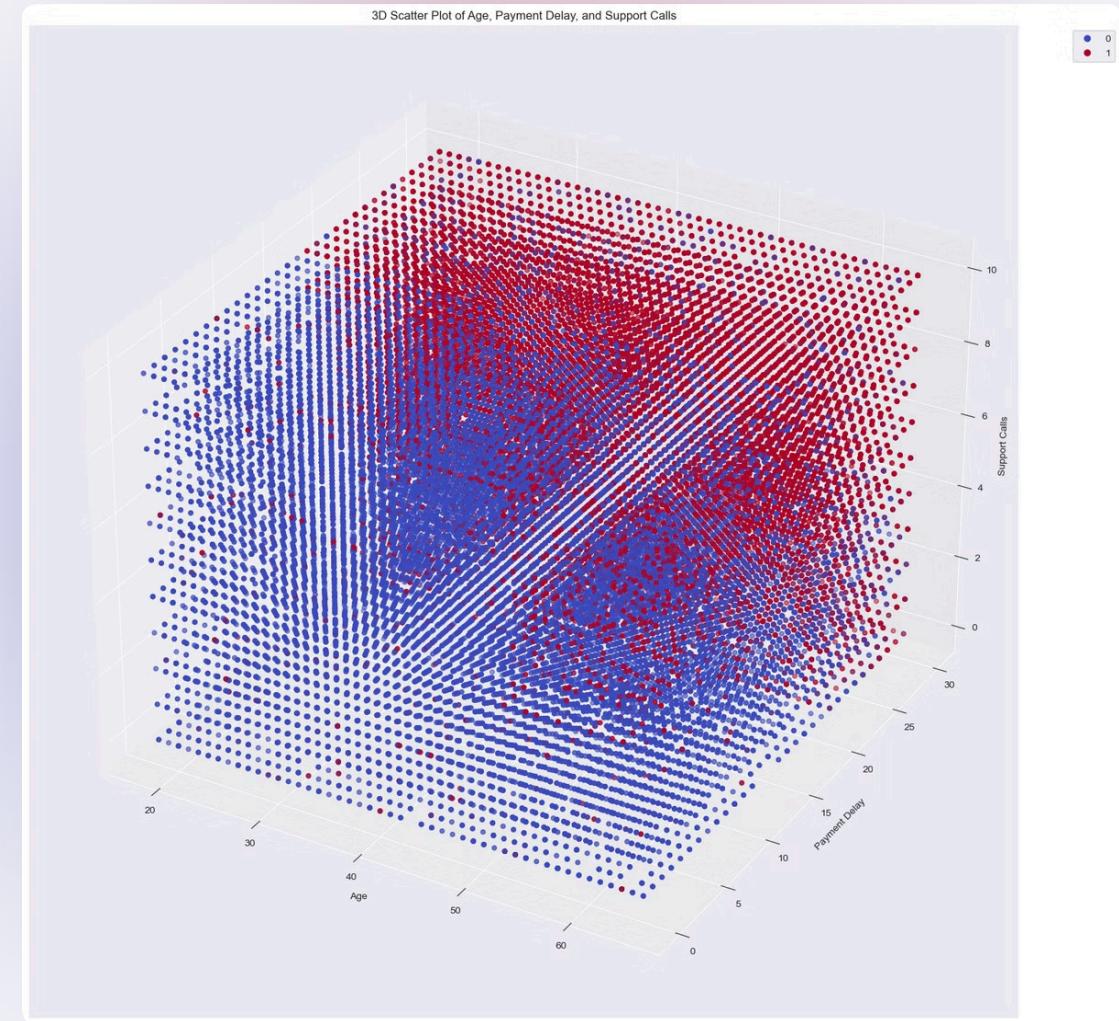
Accuracy Assessment and Model Selection

We will also use the test data to assess accuracy and determine the most effective model based on its performance score.

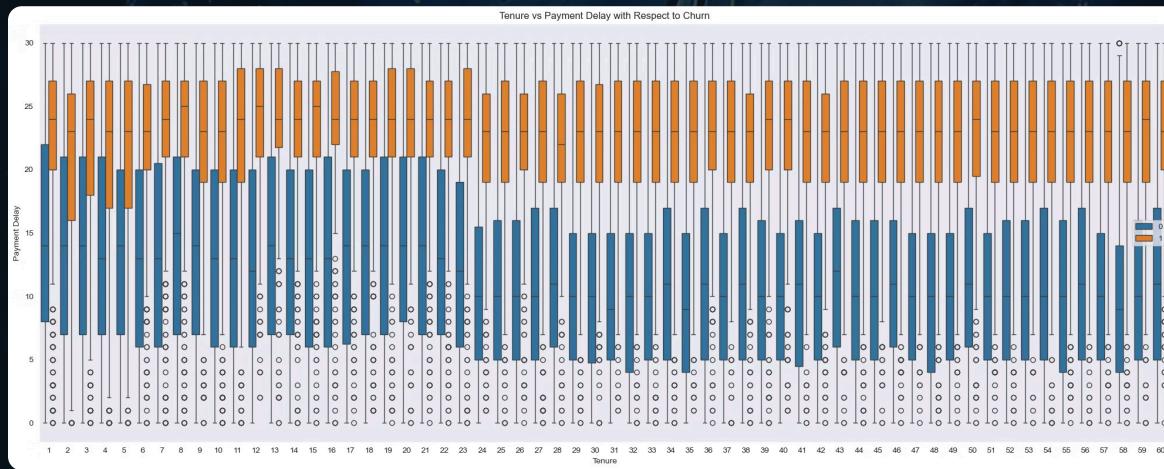
Here is a breakdown of the data columns:

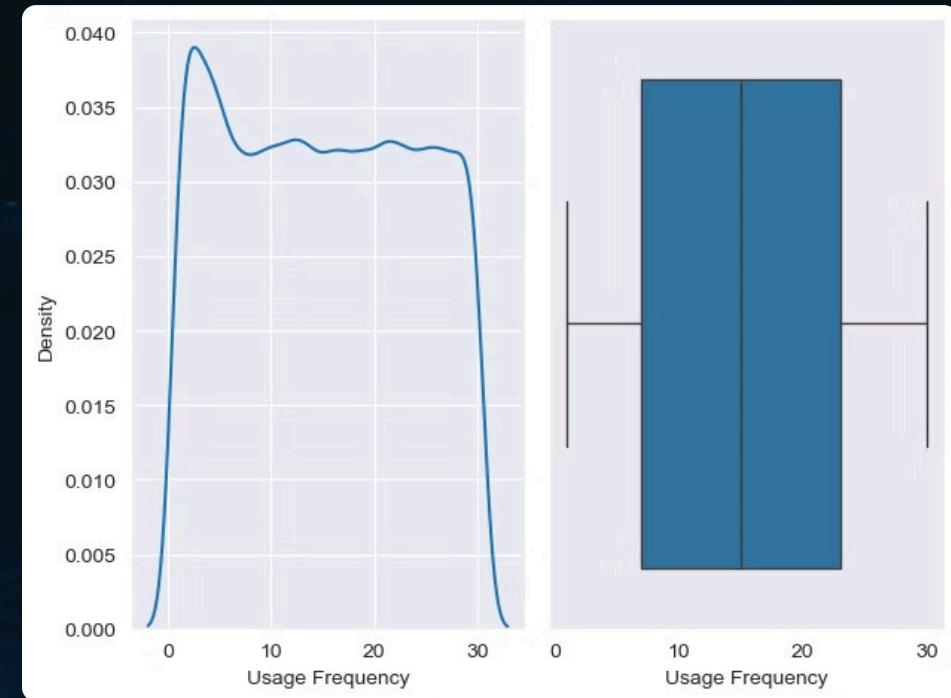
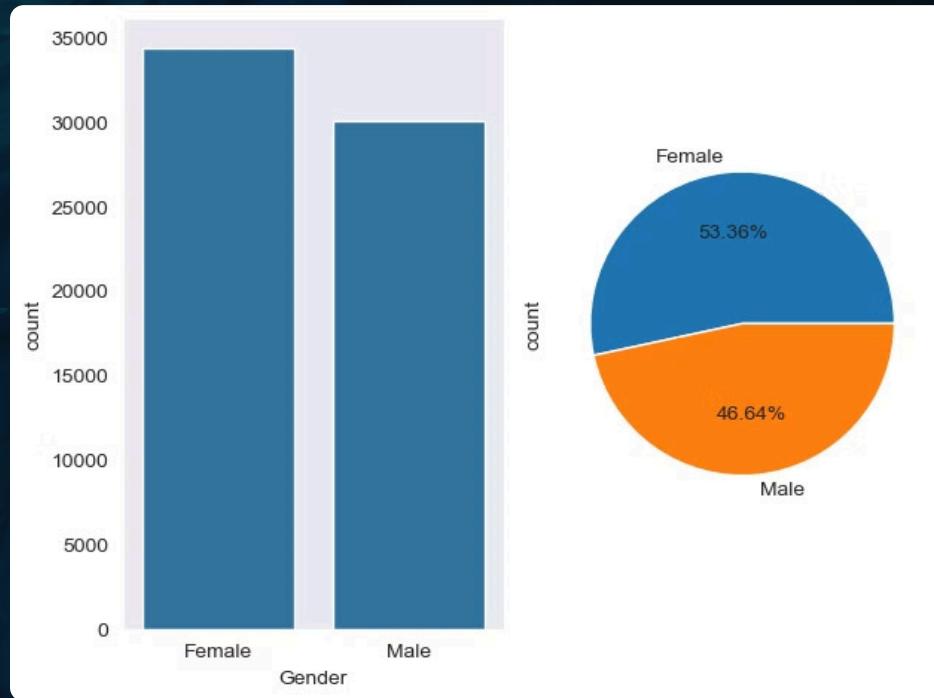
1. Customer ID: A unique identifier for each customer.
2. Age: The age of the customer.
3. Gender: The gender of the customer.
4. Tenure: The length of time the customer has been with the service.
5. Usage Frequency: How often the customer uses the service.
6. Support Calls: The number of support calls made by the customer.
7. Payment Delay: The number of times the customer has delayed payment.
8. Subscription Type: The type of subscription the customer has.
9. Contract Length: The duration of the customer's contract.
10. Total Spend: The total amount of money the customer has spent.
11. Last Interaction: The time since the customer's last interaction with the service.
12. Churn: Indicates whether the customer has churned (likely a binary column with values like 0 for 'No' and 1 for 'Yes').

3D scatter plot representation of
Age vs Payment Delay vs Support
Cells



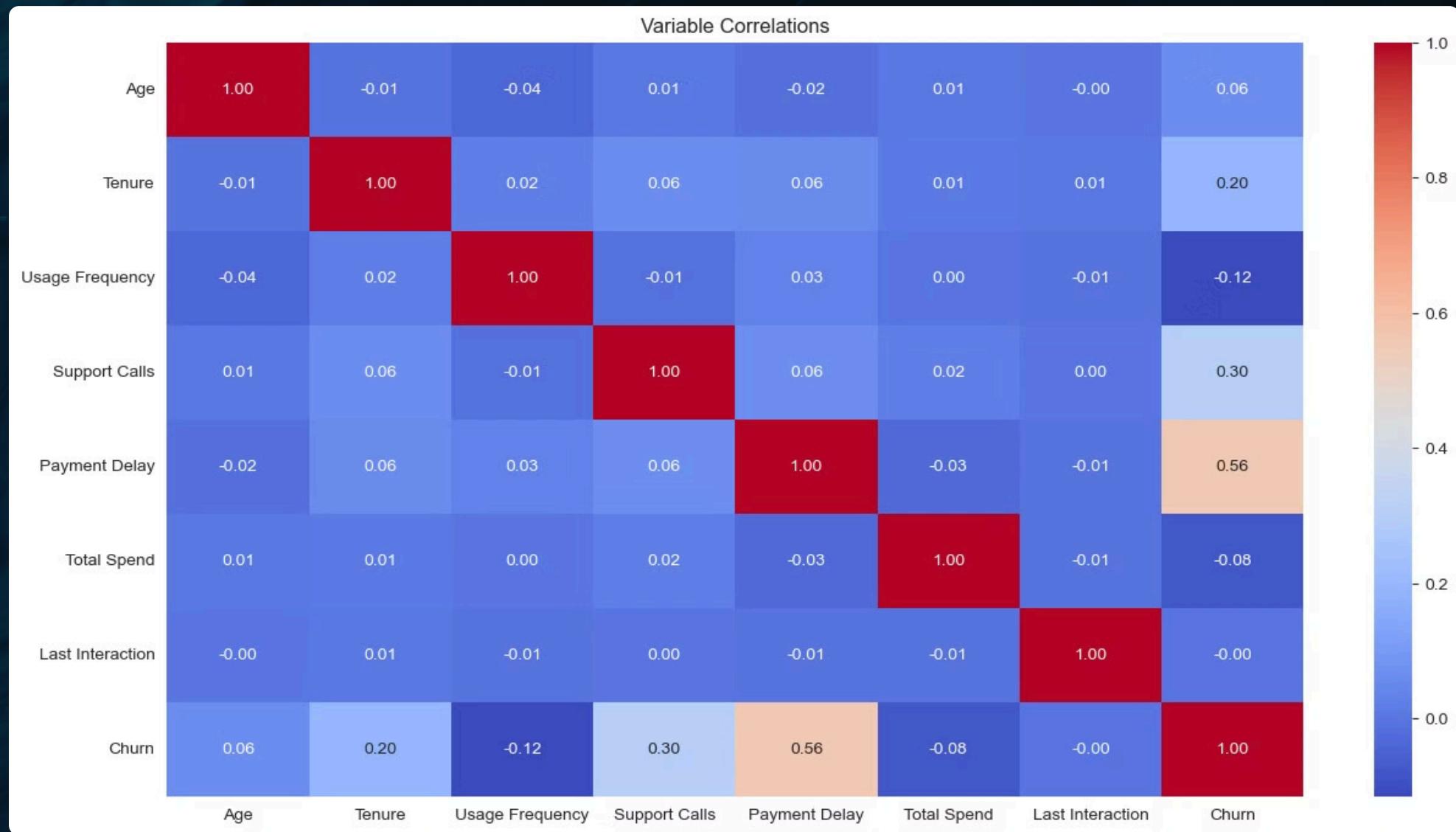
Tenure vs Payment Delay with Respect to Churn





- From the above graphs, there is 53.36 % of Female and 46.64% of Male based on Gender column data.

Overall Correlation





Variable Correlations

Variable	Correlation
Churn & Payment Delay	0.557
Churn & Support Calls	0.304
Churn & Tenure	0.195

Payment Delay (0.557) and **Support Calls** (0.304) have a significant positive impact on **Churn**, making them key factors to monitor in customer retention.

Model Interpretation

Models	Accuracy	Precision	Recall	F1 score
Logistic Regression	0.8279	0.8152	0.8250	0.8201
Gradient Boost	0.9969	0.9977	0.9958	0.9967
Random Forest	0.9992	0.9995	0.9989	0.9992
Decision Tree	0.9994	0.9993	0.9993	0.9993
SVM	0.9330	0.9143	0.9479	0.9308
KNN	0.9086	0.8774	0.9389	0.9071
XGB	0.9999	1.0000	0.9998	0.9999
LGBM	0.9998	0.9998	0.9998	0.9998

Top Features Affecting Churn

1

Payment Delay

Longer payment delays are strongly associated with higher churn rates, indicating that customers who delay payments are more likely to leave.

2

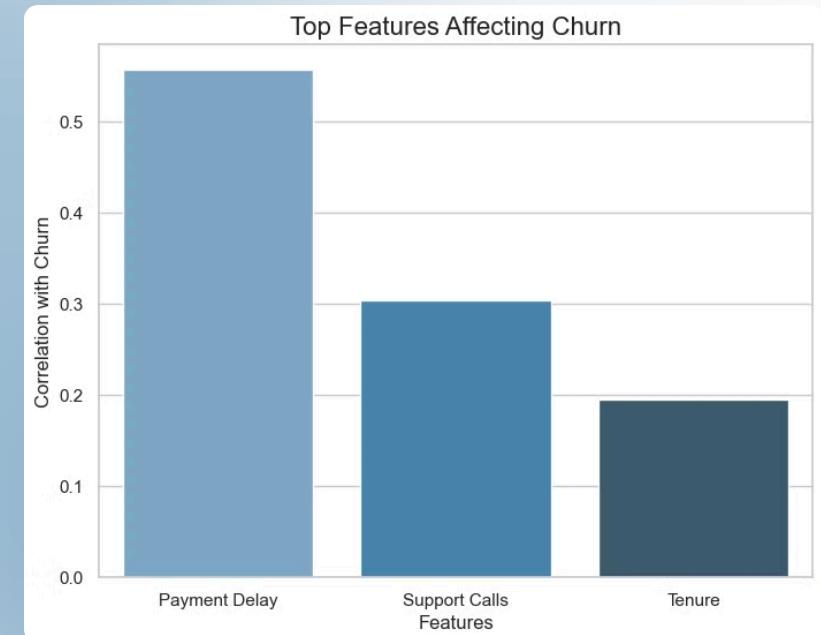
Support Calls

A higher number of support calls is weakly associated with churn, suggesting that customers needing frequent assistance may be more prone to churn.

3

Tenure

There is a weak positive correlation, indicating that customers with longer tenures are slightly more likely to churn, though the relationship is not very strong.





Conclusion and Recommendations



Focus on Payment Delay

Allocate more resources to resolving **payment delays**, as models like **XGB Classifier**, **Random Forest**, and **LGBM** consistently show that this is the most influential factor for predicting churn. Use predictive insights to proactively identify customers at risk of delayed payments and implement personalized reminders or flexible payment options to reduce churn.



Address Support Issues

Models such as **Gradient Boosting** and **SVM** highlight the correlation between **support calls** and churn. High support call volumes often indicate customer dissatisfaction. Enhance support services and provide faster issue resolution to minimize churn among customers needing frequent assistance.



Refine Approach Using Model Insights

Leverage the high accuracy and interpretability of models like **Random Forest** and **Decision Tree** to refine your churn prevention strategies. Use feature importance analysis to identify high-risk customers early and develop targeted retention campaigns. These models can help focus marketing efforts on the most critical factors, such as payment behaviors and support experiences, ensuring better resource allocation and improving overall retention.