Plagiarism Scan Report

11% Plagiarism	
0% Exact Match	
11% Partial Match	
89% Unique	
Words	920
Characters	5964
Sentences	80
Paragraphs	99
Read Time	5 minute(s)
Speak Time	7 minute(s)

Content Checked For Plagiarism

4. Churn Based on the Tenure

This code creates a boxplot to visualize the distribution of Tenure (how long a customer has been with the company) across two groups:

- Churn = 0 → Customers who stayed
- Churn = 1 → Customers who left
- Purpose
- plt.figure(figsize=(6,4)) :Sets the figure size for the plot
- sns.boxplot() :Creates a boxplot comparing tenure across churn labels
- x='Churn' :Groups by churn (0 = stayed, 1 = left)
- y='Tenure' :Plots the tenure duration
- palette='coolwarm' :Sets color scheme: blue for non-churned, red for churned
- plt.title() :Adds title
- plt.show() :Displays the plot

Churn Interpretation
0 (Non-Churned Customers) Tend to have higher and more spread-out tenures. The median is noticeably higher.
1 (Churned Customers) Have much lower tenure, concentrated around smaller values. Many are recent users.
Outliers: You can observe extreme values in both categories, especially a few long-term users in the churn group.
Insight: Customers with shorter tenure are more likely to churn.
5. Churn by Payment Method:

This code is using Seaborn and Matplotlib to plot a count plot showing the relationship between customer churn status and their payment method.

- plt.figure(figsize=(8,4)) Sets the size of the plot to 8 inches wide and 4 inches tall.
- sns.countplot(data=new_df, x='Payment', hue=new_df['Churn'].astype(str), palette='Set2')

Creates a count plot with:

- x='Payment': Payment method categories on the x-axis.
- hue=new_df['Churn'].astype(str): Colors split by Churn status (0 = Not Churned, 1 = Churned).
- palette='Set2': Uses Seaborn's 'Set2' color palette for better visual distinction.
- plt.title("Churn by Payment Method") Sets the title of the plot.
- plt.show() Displays the plot.

Output Summary

- A grouped bar chart is shown with payment methods on the x-axis and customer count on the y-axis.
- Each payment method has two bars:
  - o Orange = Not Churned (0)
  - o Green = Churned (1)
- Debit Card is the most common payment method and also has the highest churn count.
- Other methods like Credit Card Payment and UPI show lower overall usage and churn.
- Cash on Delivery and E-wallet have relatively balanced usage but with higher non-churn counts.

6. Service Score Distribution for Churned vs. Non-Churned Customers

sns.boxplot(): This creates a boxplot using Seaborn to visualize the distribution of Service_Score across churned (Churn = 1) and non-churned (Churn = 0) customers.

- x='Churn': The category on the x-axis (0 = not churned, 1 = churned).
- y='Service_Score': The numerical variable on the y-axis (values from 0 to 5).
- palette='coolwarm': Assigns colors to the boxes based on the "coolwarm" palette.
- plt.title(): Adds a descriptive title to the chart.
- plt.show(): Displays the plot.

A boxplot shows:

- Box: The interquartile range (IQR: 25th to 75th percentile).
- Line inside the box: Median value.
- Whiskers: Minimum and maximum values (excluding outliers).
- Dots: Outliers.

7. Revenue Growth vs. Churn

This code generates a boxplot using Seaborn to compare revenue growth (YoY) between:

- Churned customers (Churn = 1)
- Non-churned customers (Churn = 0)
- Boxplot Elements:
  - o Boxes show interquartile range (IQR)
  - o Middle line is the median
  - o Whiskers extend to show the full data range (except outliers)
  - o Dots beyond whiskers = outliers

Churn Status	Observation
Churn = 0 (Blue)	Slightly higher median and spread in revenue growth
Churn = 1 (Red)	Slightly lower median; more tightly distributed revenue growth

The correlation:

- Customers with higher revenue growth are less likely to churn. But the overlap between the groups means revenue growth alone isn't a strong predictor.

8. Feature Importance (XGBoost)

The goal is to:

1. Train an XGBoost classifier on training data.
2. Extract and display the importance of each feature in predicting customer churn.

XGBoost Classifier

- Brings in the XGBoost classifier algorithm for use.

Initializes an XGBoost classifier with:

- n_estimators=100: number of decision trees
- learning_rate=0.1: shrinkage rate
- random_state=42: ensures reproducibility

Trains the model using X_train and y_train.

- feature_importances_ provides a score for each feature based on how useful it was in splitting data.
- The Data Frame is sorted to show most important features at the top.
- A horizontal bar chart is plotted using Seaborn to visualize which features contributed most to the model.

Top Contributors:

- Tenure: most influential in predicting churn.
- Complain_ly: whether the customer complained last year.
- Marital_Status: possibly influencing service expectations or loyalty.

Least Important Features:

- Service_Score: surprisingly low, suggesting it might not help much in predicting churn (despite previous plots).
- AccountID: expected to be irrelevant as it's just an identifier.

9. Feature Importance (Random Forest)

The Goal is to:

1. Trains a Random Forest classifier.
2. Extracts and visualizes feature importance, showing which features contribute most to predicting customer churn.

Random Forest Classifier

Brings in the Random Forest model from scikit-learn's ensemble methods. Initializes a random forest with:

n_estimators=100: the forest will have 100 trees.

random_state=42: ensures reproducibility.

Fits the model to training data.

Retrieves feature importances from the trained model.

Sorts them from most to least important.

Uses a horizontal bar plot to display the importance of each feature using the magma color palette.

Rank	Feature	Interpretation
1	Tenure	Most important: Longer customer tenure correlates with churn likelihood.
2	cashback	Cashback benefits influence customer retention.
3	Day_Since_CC_connect	Recent interactions with customer care seem to matter.

This visualization helps in understanding which features are most influential in the Random Forest model and can be valuable for feature selection, model interpretation, and gaining insights into the underlying relationships in the data.

Matched Source

Similarity 3%

Title:Box Plot | Introduction to Statistics - JMP

The bottom and top of the box show the 25th and 75th quantiles, or percentiles. These two quantiles are also called quartiles because each cuts off a quarter (25%) of the data. The length of the box is the difference between these two percentiles and is called the interquartile range (IQR).

https://www.jmp.com/en/statistics-knowledge-portal/exploratory-data-analysis/box-plot

Similarity 3%

Title:How to Predict & Prevent Churn: 7 Factors & Indicators (2025)

Mar 30, 2023 � Customers with higher CLV can be less likely to churn because they have more investments in your solution or service. You can also prioritize�...

https://whatfix.com/blog/predicting-churn

Similarity 3%

Title:Sohaila Elbhnsawy's Post - Customer Churn Analysis - LinkedIn

Dec 21, 2024 � 5⃣Tenure Analysis ⏳: Insight: Customers with shorter tenure are more likely to churn, emphasizing the importance of early engagement strategies.

https://www.linkedin.com/posts/sohaila-elbhnsawy_customer-churn-analysis-powerbi-project-activity-7276310056738975744-Nbp7

Similarity 3%

Title:Debit Cards Remain Most Popular Payment Method

May 28, 2024 � Debit cards remain the most popular payment method for American consumers due to overall utilization and high customer satisfaction, according�...Missing: churn count.

https://www.bankingexchange.com/news-feed/item/9996-debit-cards-remain-most-popular-payment-method%3FItemid%3D777
