Customer Churn Analysis – EDA Project Documentation

1.Object of the Project

The objective of this project is to develop an end-to-end machine learning solution to predict customer churn in a telecommunications company using historical customer data. Customer churn refers to the situation where customers discontinue their services, which directly impacts the company's revenue and growth. By analysing customer demographics, service subscriptions, contract details, and billing information, this project aims to identify patterns and factors that influence a customer's decision to leave the service.

2. Tools and Technologies used

- **Programming Language:** Python
- **Libraries:**
  - Pandas-data manipulation and analysis
  - NumPy-numerical operations
  - Matplotlib-data visualization
- **Environment:** Google Collab / Jupiter Notebook
- **Dataset Format:** CSV file (churn_updated.csv)

3. Dataset Description

The Telco Customer Churn dataset contains information about a fictional telecommunications company that provided home phone and Internet services to 7,043 customers in California.

3.1. Customer Demographics

- **Customer ID**: Unique identifier for each customer.
- **gender**: Whether the customer is male or female (nearly 50/50 split).
- **Senior Citizen**: Whether the customer is a senior citizen (1) or not (0). Approximately 16% are seniors.
- **Partner**: Whether the customer has a partner (Yes/No).
- **Dependents**: Whether the customer has dependents (Yes/No)

3.2. Customer Account Information

- **Tenure**: Number of months the customer has stayed with the company (Average: ~32 months; Max: 72 months).
- **Contract**: The contract term (Month-to-month, One year, Two year). Over 50% are on month-to-month contracts.
- **Paperless Billing**: Whether the customer has paperless billing (Yes/No).
- **Payment Method**: Electronic check, mailed check, Bank transfer (automatic), Credit card (automatic).

- **Monthly Charges**: The amount charged to the customer monthly (Average: $64.76).
- **Total Charges**: The total amount charged to the customer (Average: $2,283.30). *Note:* Contains 11 missing values for new customers with 0 tenure.

3.3. Services Signed Up For

- **Phone Service**: Whether the customer has a phone service (Yes/No).
- **Multiple Lines**: Whether the customer has multiple lines (Yes, No, No phone service).
- **Internet Service**: Customer's internet service provider (DSL, Fiber optic, No).
- **Online Security, Online Backup, Device Protection, Tech Support, Streaming TV, Streaming Movies**: Additional services (Yes, No, No internet service).

3.4. Target Variable

- **Churn**: Whether the customer churned (Yes/No).
    - **No:** 5,174 (73.5%)
    - **Yes:** 1,869 (26.5%)

Statistical Summary of Numeric Features

Feature	Mean	Std Dev	Min	50% (Median)	Max
tenure	32.37	24.56	0	29	72
Monthly Charges	$64.76	$30.09	$18.25	$70.35	$118.75
Total Charges	$2283.30	$2266.77	$18.80	$1397.48	$8684.80

4. Visualizations, Graph Type & Understanding from Charts

4.1 Variables:

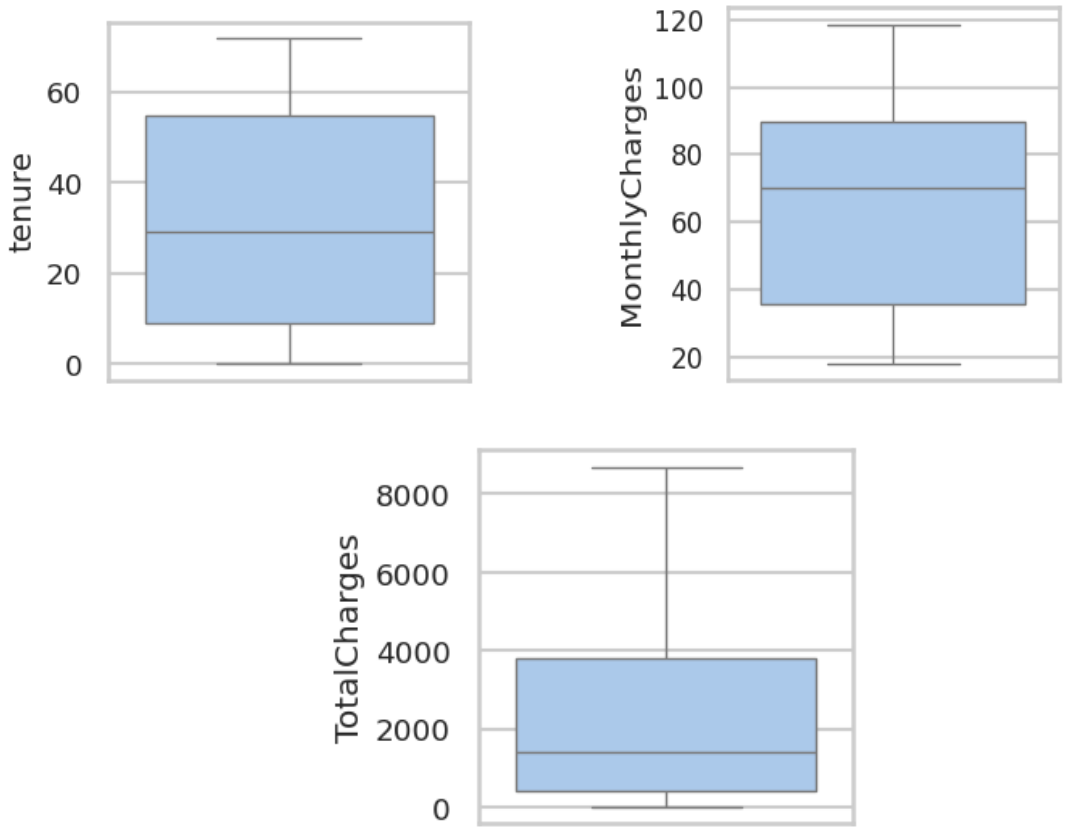
 Median (The horizontal line inside the box): Approximately **70**. This is the middle value of the dataset; 50% of the customers pay more than this, and 50% pay less.

 Interquartile Range (IQR) (The blue box): This represents the middle 50% of the data.

- **Third Quartile (Q_3 - Top of the box):** Around **90**. 75% of the data falls below this value.
- **First Quartile (Q_1 - Bottom of the box):** Around **35**. 25% of the data falls below this value.

 Whiskers (The vertical lines): These extend to the minimum and maximum values (excluding outliers).

- **Maximum:** Approximately **118-120**.
- **Minimum:** Approximately **18-20**.

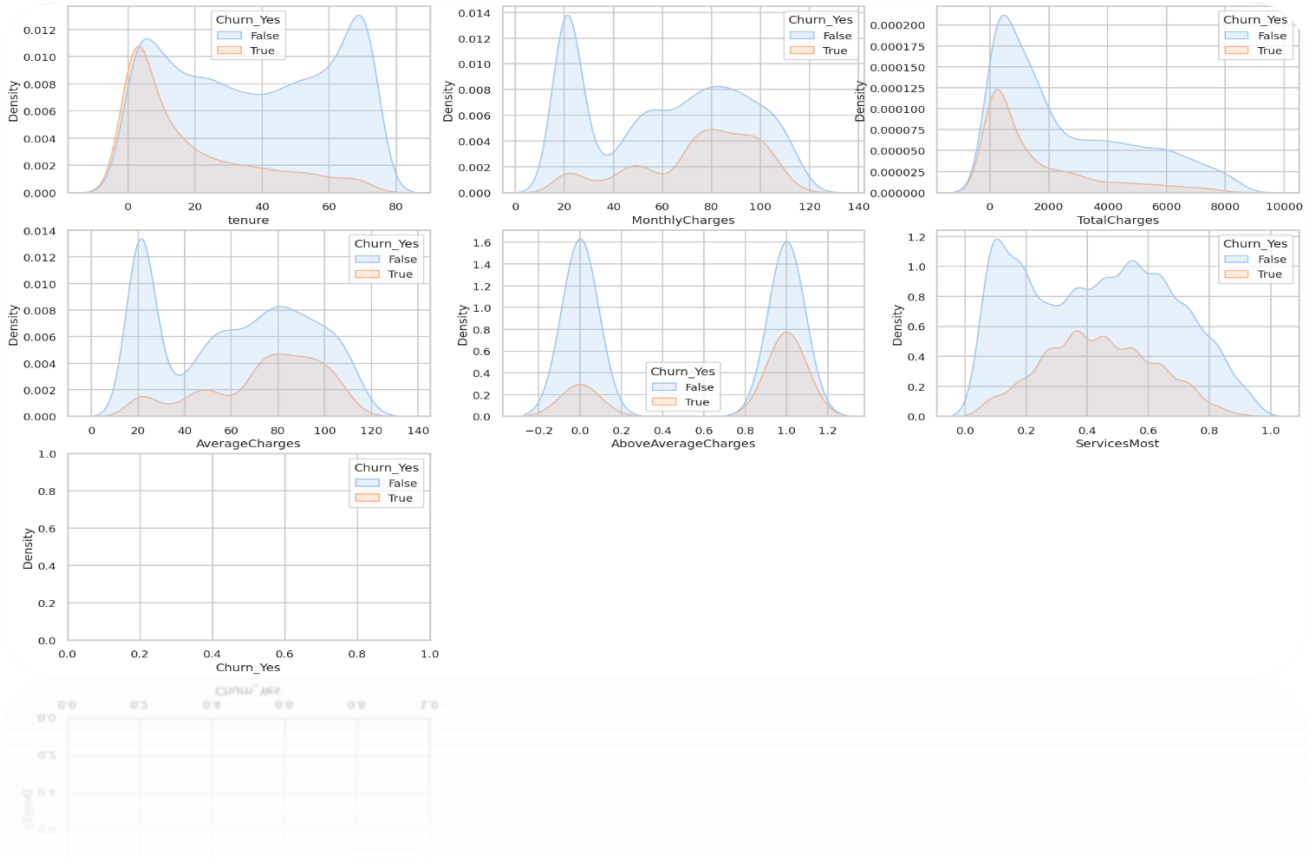


4.2 Density Distribution of Customer Metrics by Churn Status:

These are Kernel Density Estimate (KDE) plots. They compare the "shape" of the data for customers who stayed (False, blue) versus those who left (True, orange).

- **Tenure:** Customers who churn (orange) are heavily concentrated at the very beginning of their "tenure" (0–10 months). Long-term customers (60+ months) are much less likely to leave.
- **Monthly & Average Charges:** There is a significant orange peak around the 80–100 range. This suggests that customers with higher monthly bills are more likely to churn than those with low bills (peak at 20).

- **Total Charges:** The density for churned customers is highest at low total charges, likely because they leave before they can accumulate a high lifetime spend.
- **Above Average Charges:** This is a categorical comparison. You can see a higher density of "Churn = True" (orange) when the value is 1.0, confirming that being above the average bill is a risk factor for leaving.



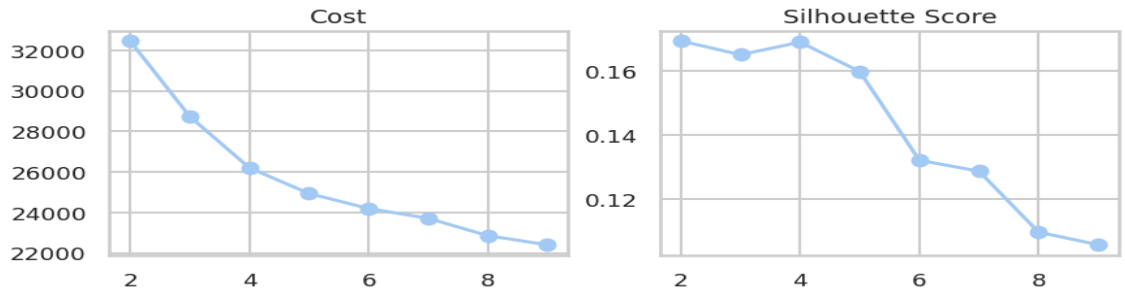
4.3 Exploratory Data Analysis and Cluster Optimization for Customer Retention:

 Central Tendency: The horizontal line in the middle of the box represents the **median** monthly charge, which is approximately **70**.

 Spread (IQR): The blue box represents the middle 50% of the data (the Interquartile Range). It shows that most customers are charged between approximately **35** (Q_1) and **90** (Q_3).

 Range: The vertical "whiskers" extend to the minimum (approx. **18**) and maximum (approx. **118**) values.

 Conclusion: The data has a wide spread, indicating a diverse range of service plans, with no visible outliers in this specific view.

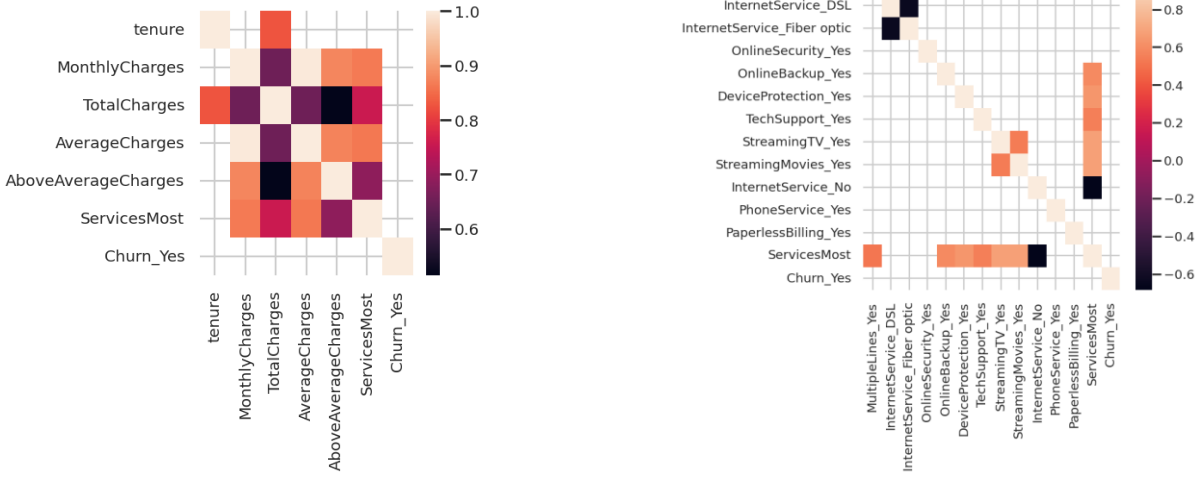


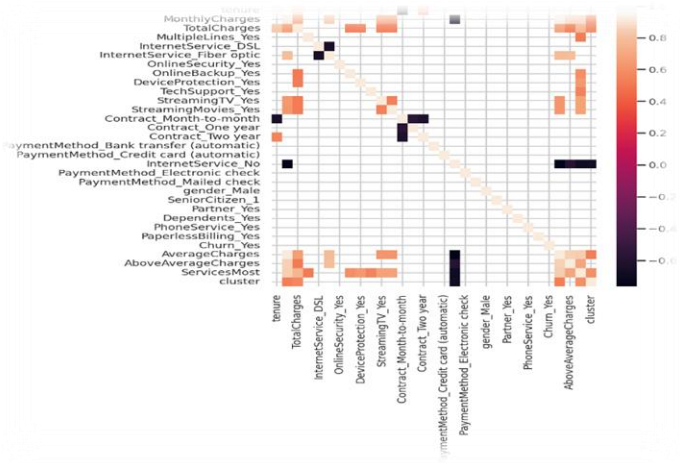
4.4 Customer Churn Analysis:

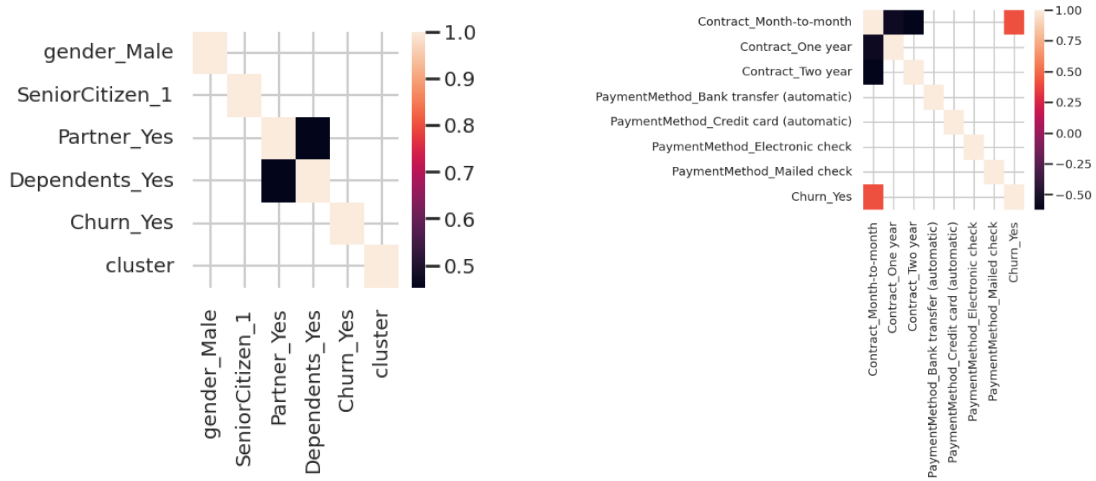
 Total Charges is strongly correlated with **tenure**, as long-term customers naturally accumulate higher total costs.

 Monthly Charges shows a strong relationship with **Streaming TV** and **Streaming Movies**, indicating that these add-on services drive up monthly bills.

 Average Charges and **Above Average Charges** are highly correlated with the **cluster** assignment, suggesting that billing amounts were a primary factor in grouping these customers.







4.5 Comprehensive Analysis of Customer Churn Drivers and Segmentation

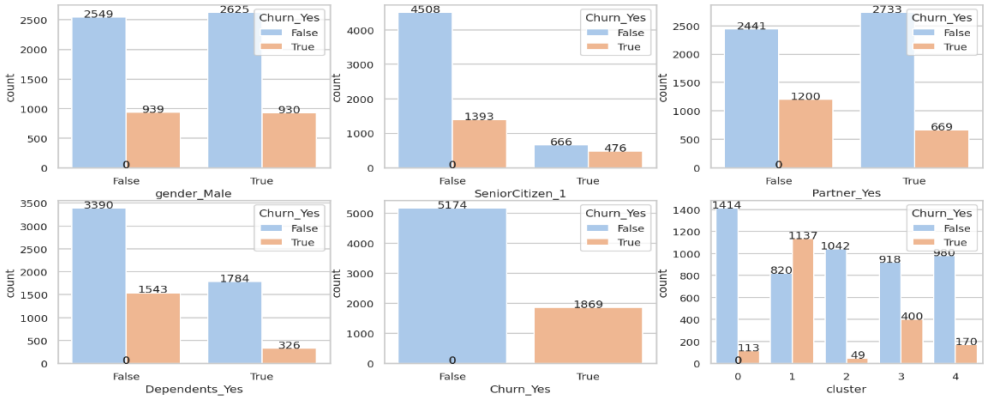
 Demographics (Gender, Senior Citizen, Partner, Dependents): * Gender: Churn rates are nearly identical between males and females, suggesting gender is not a primary driver of churn.

- **Senior Citizens:** While there are fewer senior citizens overall (SeniorCitizen_1 = True), a much higher proportion of them churn compared to non-seniors.
- **Dependents/Partners:** Customers without dependents or partners show a higher volume and proportion of churn compared to those with domestic ties.

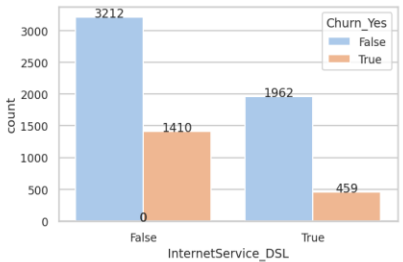
 Churn Volume: The central plot shows the total baseline: **5,174** customers stayed, while **1,869** churned.

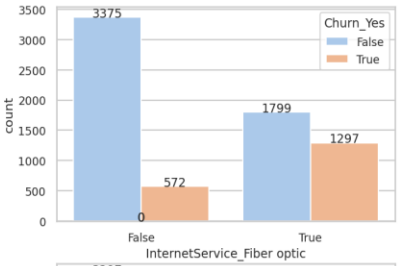
 Cluster Segmentation: This is the most actionable plot. It shows how the 4 clusters (plus a baseline group) behave differently:

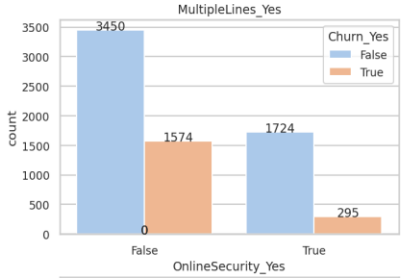
- **Cluster 1:** This group has the highest churn rate, with **1,137** customers leaving and only **820** staying.
- **Cluster 0 & 2:** These groups are highly loyal, with very few churned customers relative to the total.

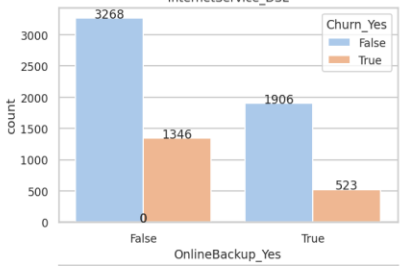


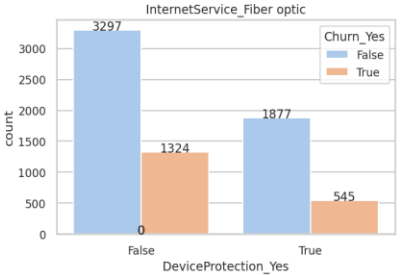

Churn_Yes
False
True
count
3053
1019
2121
850
0
False
True
MultipleLines_Yes

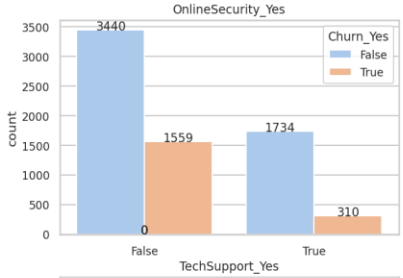

Churn_Yes
False
True
count
3212
1410
1962
459
0
False
True
InternetService_DSL

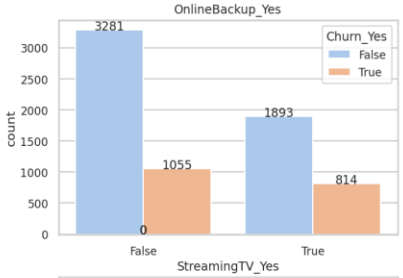

Churn_Yes
False
True
count
3375
572
1799
1297
0
False
True
InternetService_Fiber optic

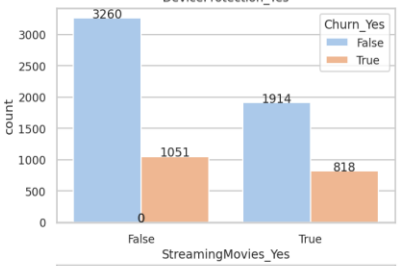

Churn_Yes
False
True
count
3450
1574
1724
295
0
False
True
OnlineSecurity_Yes

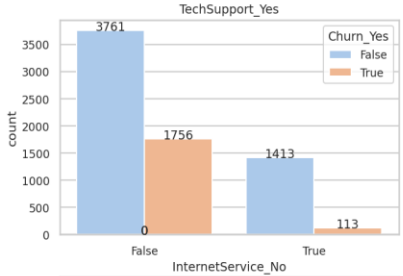

Churn_Yes
False
True
count
3268
1346
1906
523
0
False
True
OnlineBackup_Yes

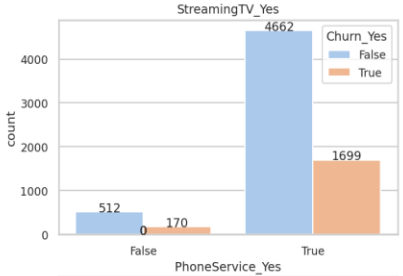

Churn_Yes
False
True
count
3297
1324
1877
545
0
False
True
DeviceProtection_Yes

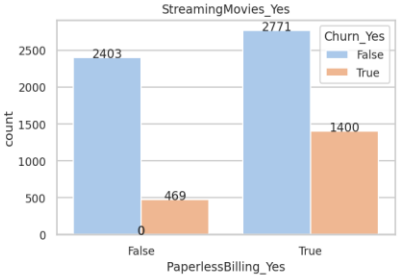

Churn_Yes
False
True
count
3440
1559
1734
310
0
False
True
TechSupport_Yes

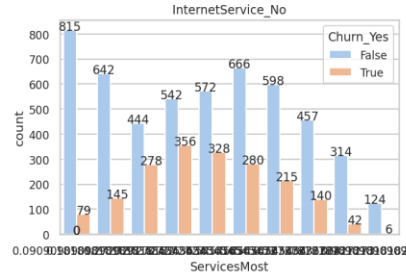

Churn_Yes
False
True
count
3281
1055
1893
814
0
False
True
StreamingTV_Yes

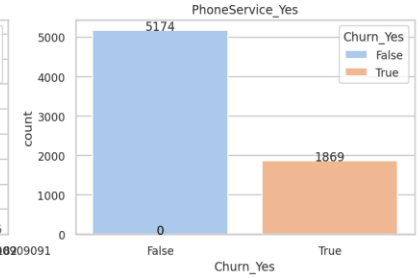

Churn_Yes
False
True
count
3260
1051
1914
818
0
False
True
StreamingMovies_Yes

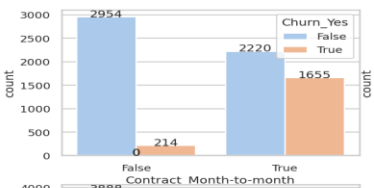

Churn_Yes
False
True
count
3761
1756
1413
113
0
False
True
InternetService_No

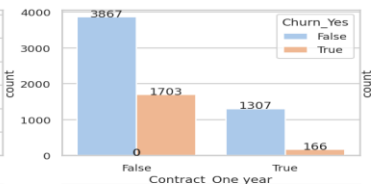

Churn_Yes
False
True
count
4662
512
170
1699
0
False
True
PhoneService_Yes

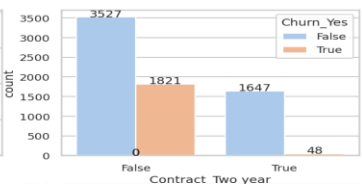

Churn_Yes
False
True
count
2403
2771
469
1400
0
False
True
PaperlessBilling_Yes

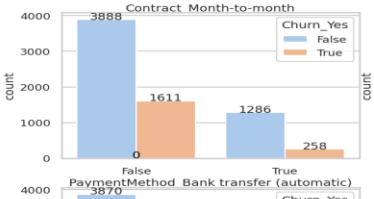

Churn_Yes
False
True
count
815
642
444
542
572
666
598
457
314
79
145
278
356
328
280
215
140
42
124
6
0
ServicesMost

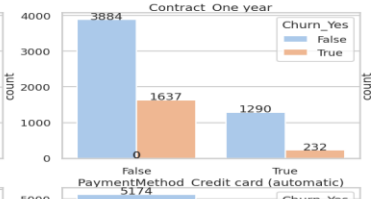

Churn_Yes
False
True
count
5174
1869
0
False
True
Churn_Yes

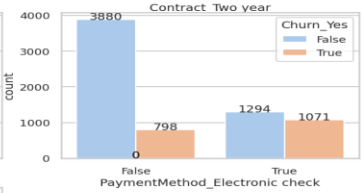

Churn_Yes
False
True
count
2954
2220
1655
214
0
False
True
Contract_Month-to-month

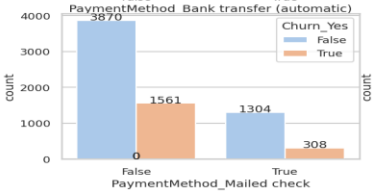

Churn_Yes
False
True
count
3867
1703
1307
166
0
False
True
Contract_One year

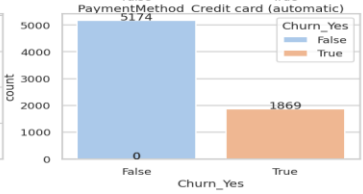

Churn_Yes
False
True
count
3527
1821
1647
48
0
False
True
Contract_Two year


Churn_Yes
False
True
count
3888
1611
1286
258
0
False
True
PaymentMethod_Bank transfer (automatic)

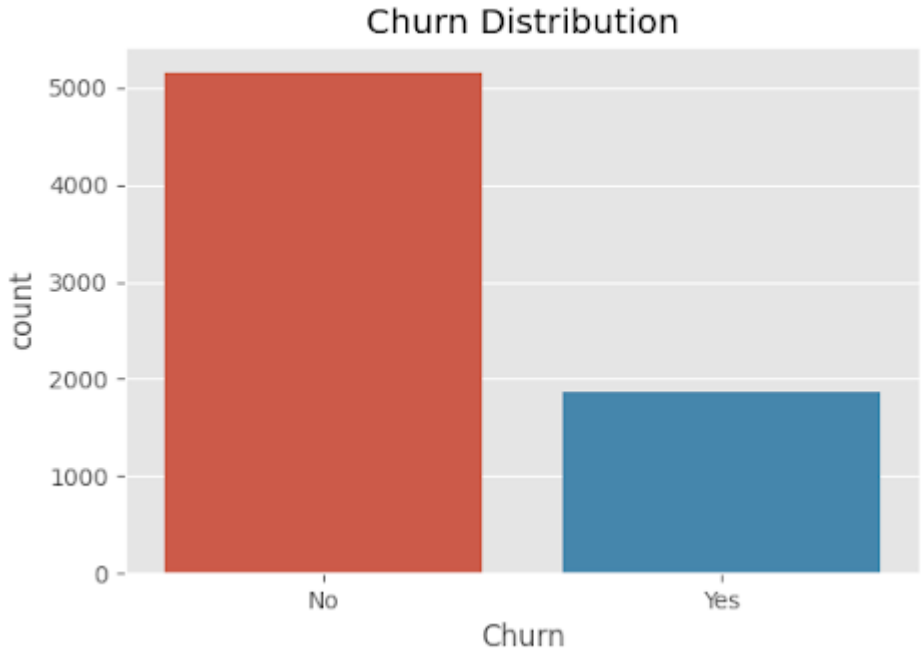

Churn_Yes
False
True
count
3884
1637
1290
232
0
False
True
PaymentMethod_Credit card (automatic)

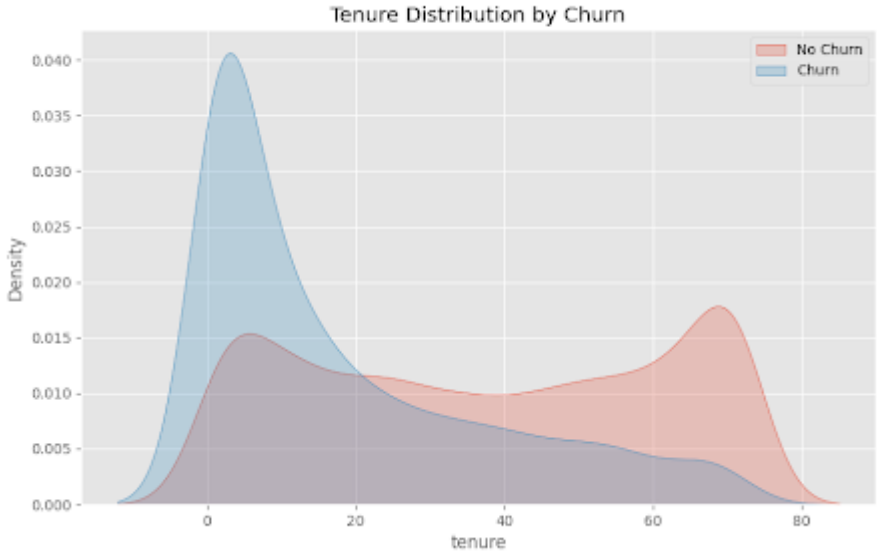

Churn_Yes
False
True
count
3880
798
1294
1071
0
False
True
PaymentMethod_Electronic check

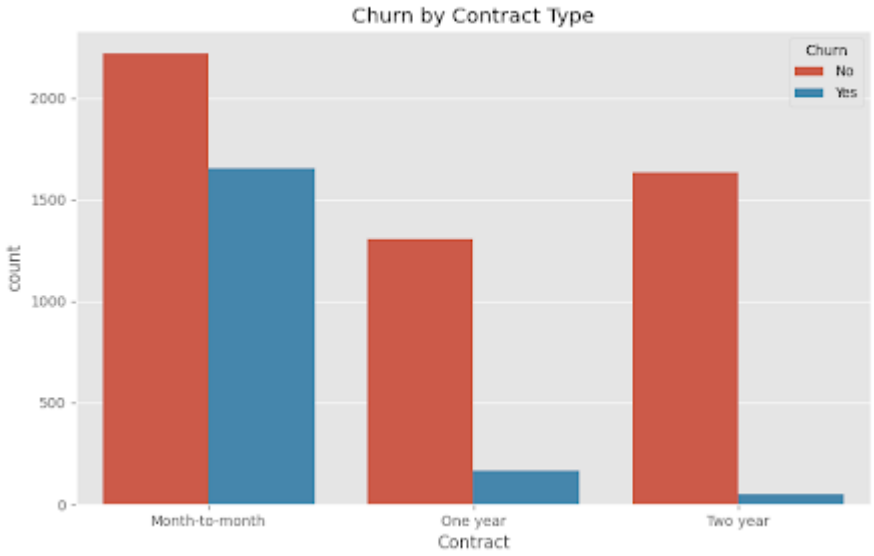

Churn_Yes
False
True
count
3870
1561
1304
308
0
False
True
PaymentMethod_Mailed check

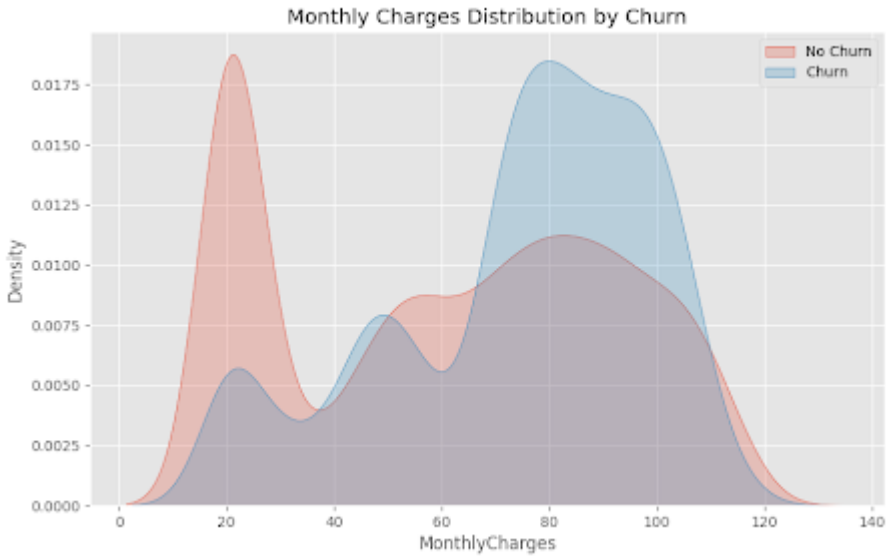

Churn_Yes
False
True
count
5174
1869
0
False
True
Churn_Yes

5. Overall Insights









Based on the analysis of the Telco Customer Churn dataset, here are the overall insights and key findings:

Executive Summary

- Overall Churn Rate: Approximately 26.6% of the customers in the dataset have churned.
- Total Customers: 7,043 (with 11 missing values for Total Charges, which were cleaned).

Key Drivers of Churn

- Contract Type: This is the strongest predictor. Customers on Month-to-month contracts are much more likely to churn (42.7%) compared to those on One-year (11.3%) or Two-year (2.8%) contracts.
- Internet Service: Customers using Fiber optic service show a significantly higher churn rate (41.9%) compared to DSL (19%) or no internet service.
- Tenure: New customers are at the highest risk. Churn is heavily concentrated in the first few months of service and decreases as the customer's tenure increases.
- Financials: Customers with higher monthly charges tend to churn more frequently. The density plot shows a peak in churn for customers paying between $\$70$ and $\$100$ per month.

Demographic Insights

- Senior Citizens: Senior citizens churn at a rate of 41.7%, which is nearly double the rate of non-senior citizens (23.7%).
- Dependents & Partners: Customers who are single (no partner) or have no dependents are more likely to churn.
- Gender: Interestingly, gender has almost no impact on churn, with both males and females churning at nearly identical rates (~26%).

Service-Related Findings

- Support Services: Customers who do not have Tech Support or Online Security are significantly more likely to churn (over 41%). Providing these services appears to be a strong retention factor.
- Billing: Paperless billing and certain payment methods (like Electronic Checks) are associated with higher churn rates.

Visualizations Overview

- Churn Distribution: Shows the imbalance between churned and retained customers.
- Tenure vs Churn: Highlights that churn occurs early in the customer lifecycle.
- Contract Type vs Churn: Visualizes how long-term contracts effectively lock in customers.
- Monthly Charges vs Churn: Illustrates the price sensitivity of customers.

Recommendations

1. Promote Long-term Contracts: Incentivize Month-to-month customers to move to 1 or 2-year plans.
2. Focus on Fiber Optic Customers: Investigate why fibre optic customers are leaving (e.g., service quality or pricing) and provide targeted support.
3. Enhance Support Services: Bundle Tech Support and Online Security as they are strongly linked to lower churn.
4. Senior Citizen Engagement: Develop specific retention programs or simplified plans for senior citizens.