Telco Customer Churn Analysis

Aziz Prabowo

About the project

Background

In the highly competitive telecommunications industry, retaining existing customers is as important as acquiring new ones. This year, the company aims to boost revenue by expanding its customer base and strengthening loyalty. However, customer churn presents a major challenge, directly impacting revenue growth. Understanding churn drivers is crucial for improving retention efforts and safeguarding both short-term revenue and long-term brand reputation.

Goal

The goal of this project is to analyze customer churn data to identify key patterns distinguishing churned customers from loyal ones. This will enable the company to implement targeted retention strategies, optimize services, and refine pricing models to reduce churn, enhance customer satisfaction, and drive long-term revenue growth.

Objective

Pinpoint the most significant factors contributing to customer churn.

Dataset: Telco Customer Churn
Notebook: Jupyter Notebook

About the data

There are 11 rows with missing values in the **TotalCharges** columns. These rows are removed.

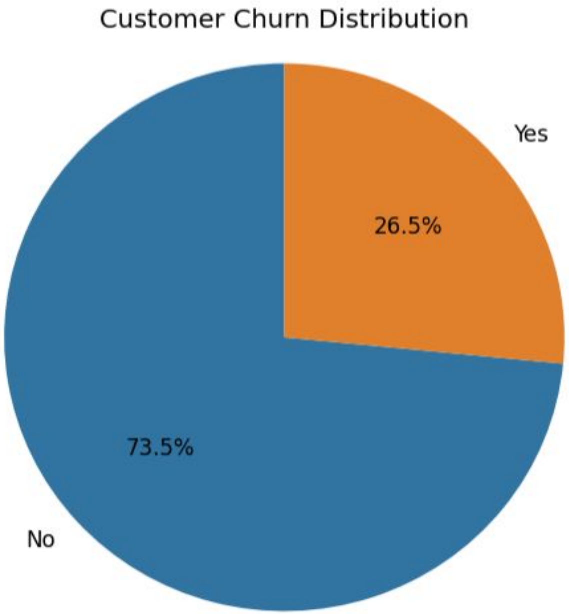
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7043 entries, 0 to 7042
Data columns (total 21 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   customerID        7043 non-null   object
 1   gender            7043 non-null   object
 2   SeniorCitizen     7043 non-null   int64
 3   Partner           7043 non-null   object
 4   Dependents        7043 non-null   object
 5   tenure            7043 non-null   int64
 6   PhoneService      7043 non-null   object
 7   MultipleLines     7043 non-null   object
 8   InternetService   7043 non-null   object
 9   OnlineSecurity    7043 non-null   object
 10  OnlineBackup      7043 non-null   object
 11  DeviceProtection  7043 non-null   object
 12  TechSupport       7043 non-null   object
 13  StreamingTV       7043 non-null   object
 14  StreamingMovies   7043 non-null   object
 15  Contract          7043 non-null   object
 16  PaperlessBilling  7043 non-null   object
 17  PaymentMethod     7043 non-null   object
 18  MonthlyCharges    7043 non-null   float64
 19  TotalCharges      7032 non-null   object
 20  Churn             7043 non-null   object
dtypes: float64(1), int64(2), object(18)
memory usage: 1.1+ MB
```

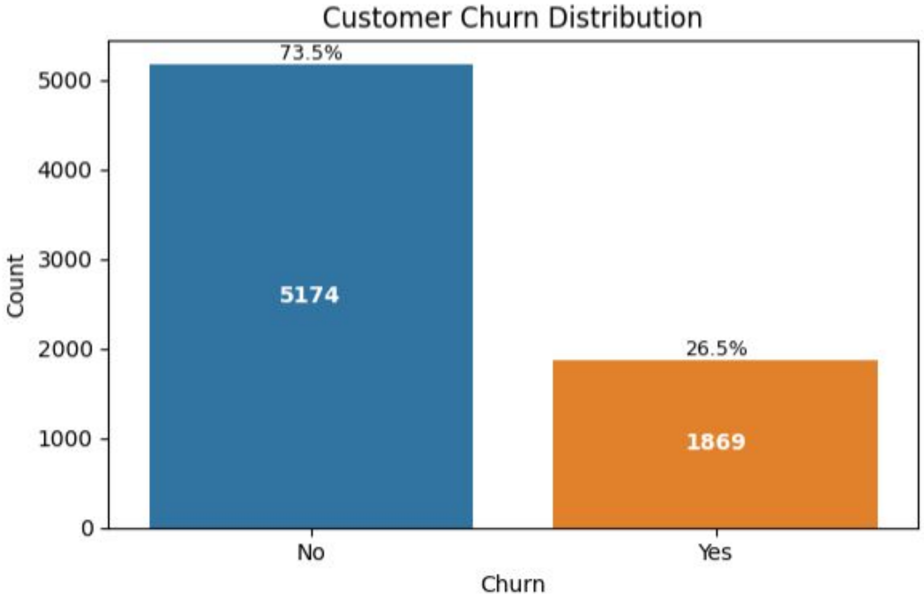
```
Null Values:

customerID          0
gender              0
SeniorCitizen       0
Partner             0
Dependents          0
tenure              0
PhoneService        0
MultipleLines       0
InternetService     0
OnlineSecurity      0
OnlineBackup        0
DeviceProtection    0
TechSupport         0
StreamingTV         0
StreamingMovies     0
Contract            0
PaperlessBilling    0
PaymentMethod       0
MonthlyCharges      0
TotalCharges       11
Churn               0
dtype: int64
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 7032 entries, 0 to 7042
Data columns (total 20 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   gender            7032 non-null   object
 1   SeniorCitizen     7032 non-null   int64
 2   Partner           7032 non-null   object
 3   Dependents        7032 non-null   object
 4   tenure            7032 non-null   int64
 5   PhoneService      7032 non-null   object
 6   MultipleLines     7032 non-null   object
 7   InternetService   7032 non-null   object
 8   OnlineSecurity    7032 non-null   object
 9   OnlineBackup      7032 non-null   object
 10  DeviceProtection  7032 non-null   object
 11  TechSupport       7032 non-null   object
 12  StreamingTV       7032 non-null   object
 13  StreamingMovies   7032 non-null   object
 14  Contract          7032 non-null   object
 15  PaperlessBilling  7032 non-null   object
 16  PaymentMethod     7032 non-null   object
 17  MonthlyCharges    7032 non-null   float64
 18  TotalCharges      7032 non-null   object
 19  Churn             7032 non-null   object
dtypes: float64(1), int64(2), object(17)
memory usage: 1.1+ MB
```

There are **26.5%** customer churned from the company's services.

1869 (26.5%) out of **7032** customers are churned. This number is actually quite small, but it is important to know the reasons for the churn.

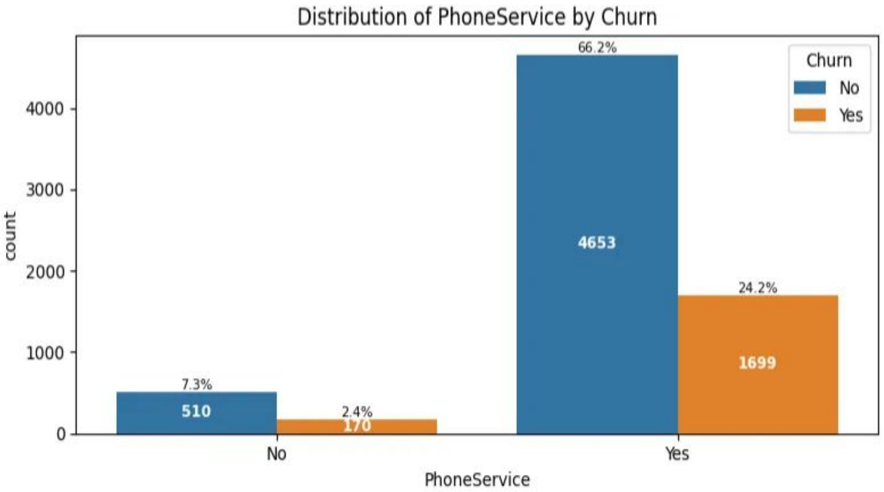


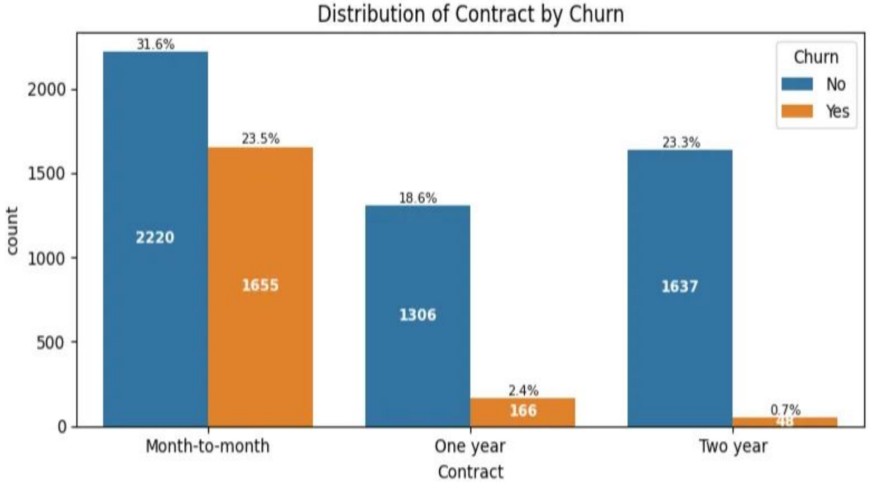


Follow-up question: *Which variable is associated with the highest number of churn?*

PhoneService has the highest churn rate, driven by user volume, while Month-to-month contract closely follow.

Among all customer segments, those with **PhoneService** show the highest churn rate at **24.2%**. This is particularly impactful because most customers are in this category, meaning this group contributes the largest share of total churn. Another major churn driver is **contract** type. Customers on **Month-to-month** plans have a churn rate of **23.5%**, nearly matching that of **PhoneService** users. Meanwhile, **One-year** and **Two-year** contracts show dramatically lower churn rates at **2.4%** and **0.7%**, respectively.

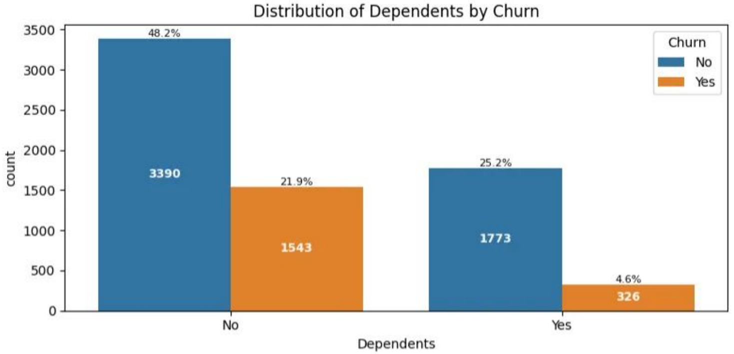


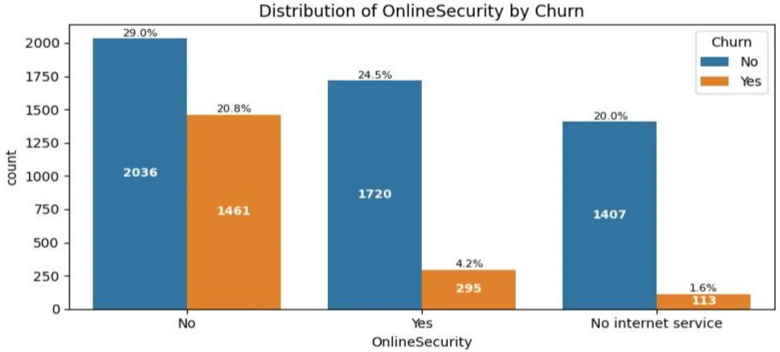


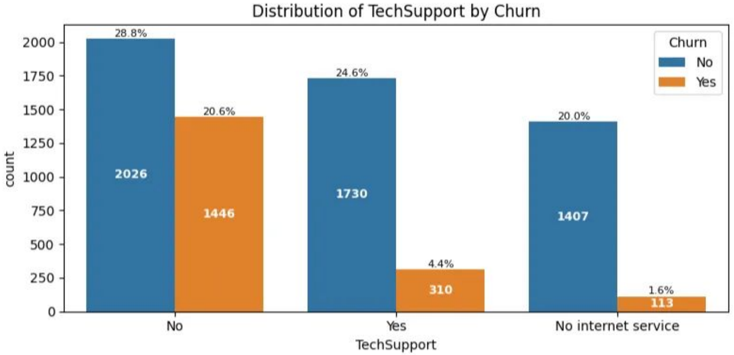
Follow-up question: *Which other variables are associated with the factor of churn?*

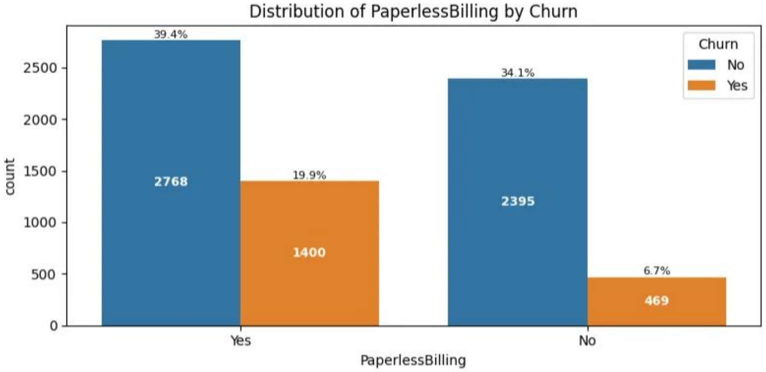
Dependents, OnlineSecurity, TechSupport, and PaperlessBilling were also identified as potential churn factors.

We found that variables such as **Dependents**, **OnlineSecurity**, **TechSupport**, and **PaperlessBilling** could potentially impact churn based on their population distribution. Churn risk appears to concentrate more in **larger, underserved segments** (e.g., no dependents, no support). This insight highlights an opportunity to reduce churn by expanding service features to these groups.



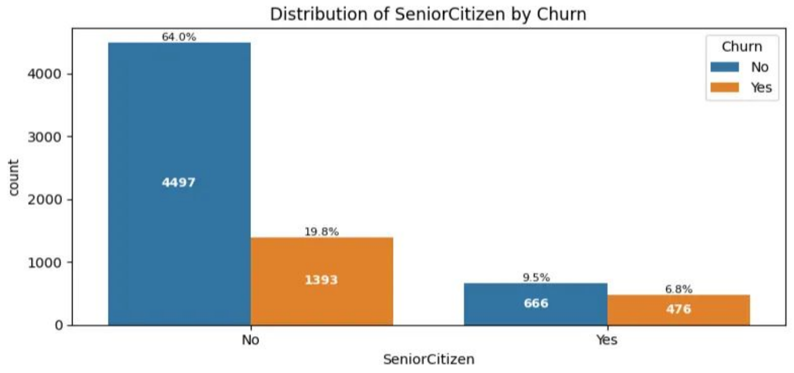


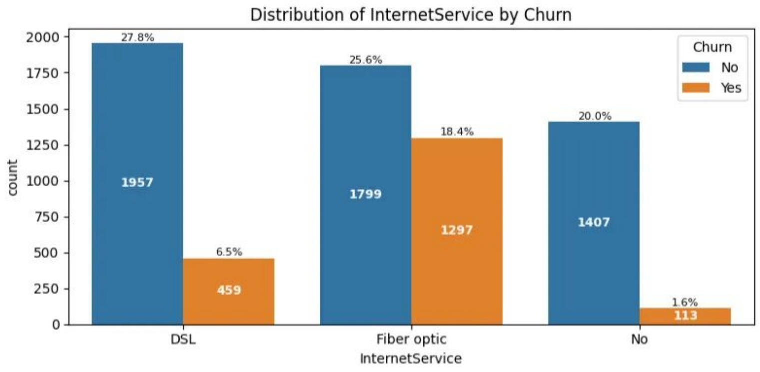


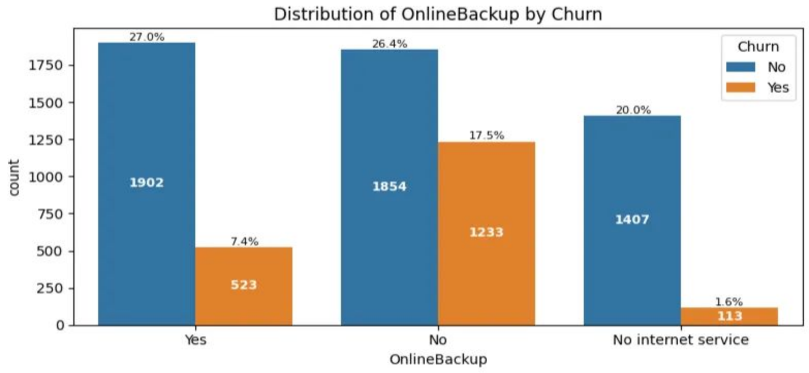


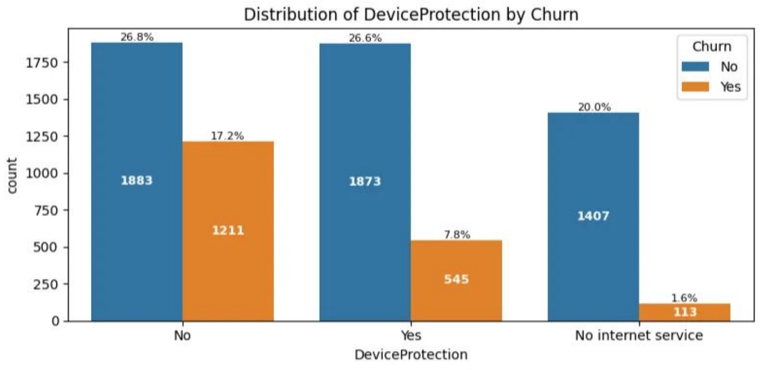
Additional churn risks found in SeniorCitizen, InternetService type, and lack of support features.

Senior Citizens, **InternetService**, **OnlineBackup,** and **DeviceProtection** were also identified as other factors contributing to churn. This suggests that the absence of added-value services and specific customer profiles also impact churn rates.



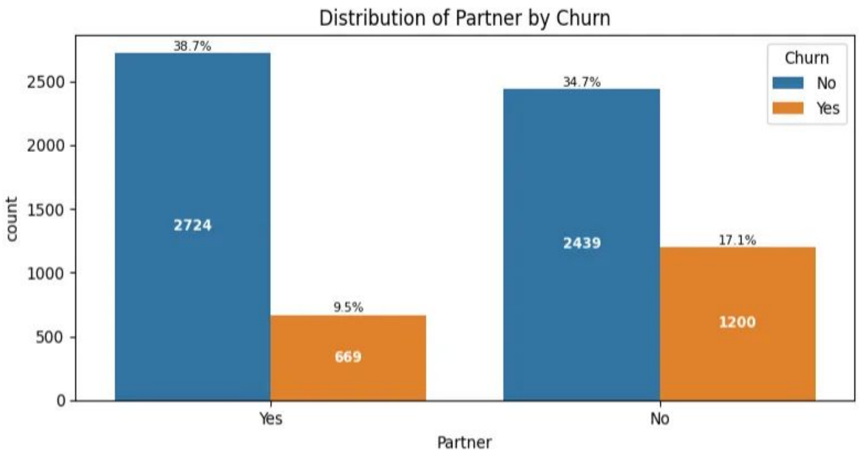


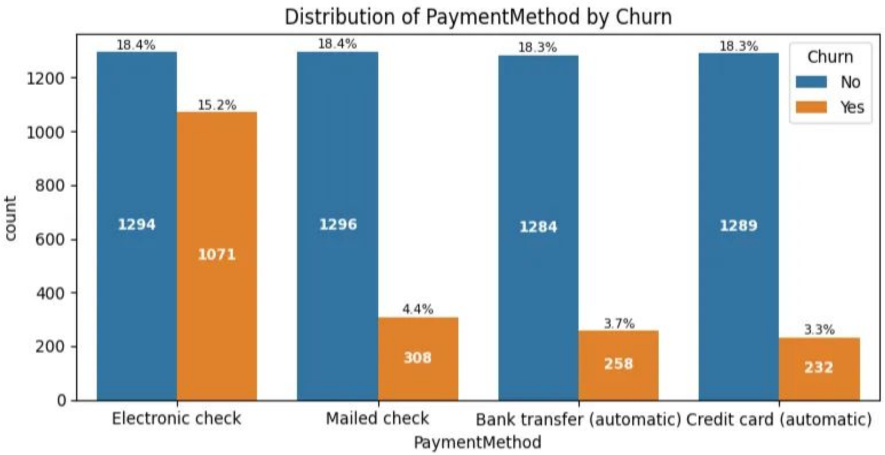




Partner and PaymentMethod were the last two categorical variables identified as potential churn factors.

Partner and **PaymentMethod** were two additional variables identified as impacting churn. This suggests that churn risk is influenced more by behavior and preferences than just group size, and larger segments can still exhibit higher churn depending on specific attributes.



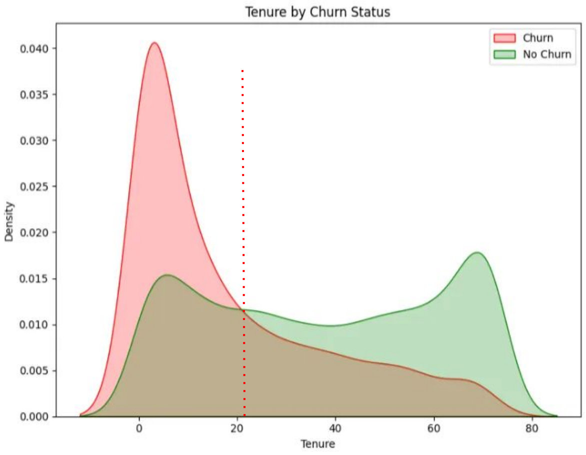


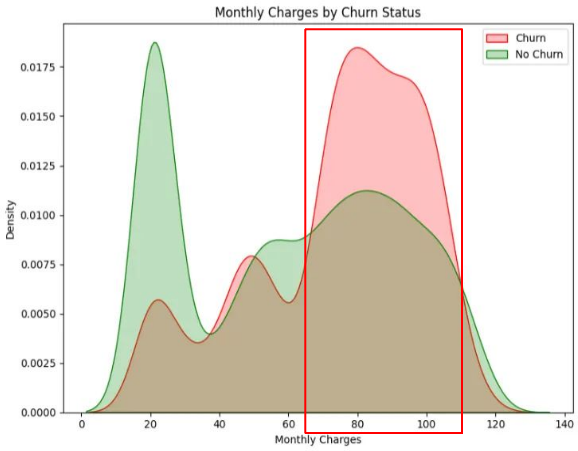
Follow-up question: *How about the numerical variables?*

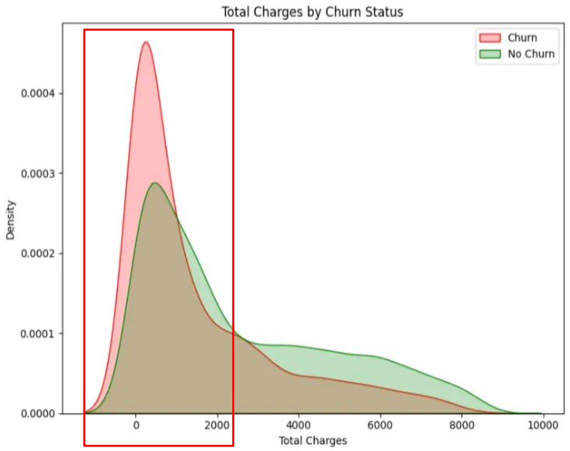
High monthly charges, low tenure, and low total charges are likely make customers churn.

Several key insights emerge from the chart:

1. **Customers with longer tenure (> 20 months)** are more likely to stay, suggesting that **early experiences** play a significant role in driving churn.
2. **Higher monthly charges** may be associated with a higher churn risk, as customers could be unwilling to pay more for their services.
3. **Churn is more common among customers with lower total charges**, reinforcing the idea that **newer or lower-value customers** are more likely to leave.







Follow-up question: *So, what's the most correlated variables with Churn in statistical way?*

Preparing for Correlation Analysis

Before diving into correlation analysis with the **Churn** variable, we've selected a focused set of features based on prior observations:

- **Categorical variables**: *Partner*, *InternetService*, *Contract*, and *OnlineBackup*.

  These were chosen because, based on their distributions, they showed **clear differences in churn rates** across their categories, indicating potential non-linear relationships worth exploring.

- **Numerical variables**: *MonthlyCharges*, *TotalCharges*, and *Tenure*.

  In particular, *MonthlyCharges* and *Tenure* show **varying churn behavior across different ranges**, not directly tied to population size. *TotalCharges* is included for completeness and to examine possible cumulative effects.

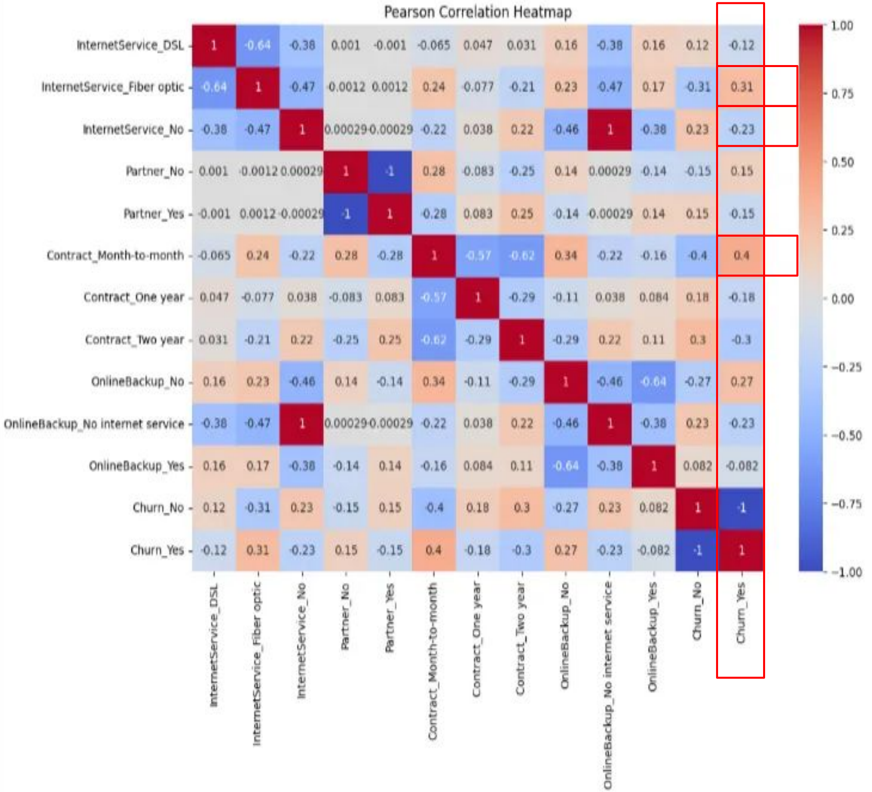
Target Variable: Churn

Selected Variables:
[Partner, InternetService, Contract, OnlineBackup, MonthlyCharges, TotalCharges, Tenure]

Contract Type, especially Month-to-Month and Internet Service Type Strongly Influence Churn Likelihood

Among categorical variables, two key drivers of churn stand out:

- **Customers on Month-to-Month contracts** show a much higher likelihood of churn (**correlation: +0.40**) compared to those on longer-term contracts, particularly **Two-Year contracts**, which are associated with reduced churn (**correlation: –0.30**).
- **Fiber Optic internet users** are also more likely to churn (**correlation: +0.31**), possibly due to higher costs or service expectations.
- In contrast, customers **without internet service** are less likely to churn (**correlation: –0.23**), which could reflect low engagement or inactive accounts that naturally have lower churn risk.

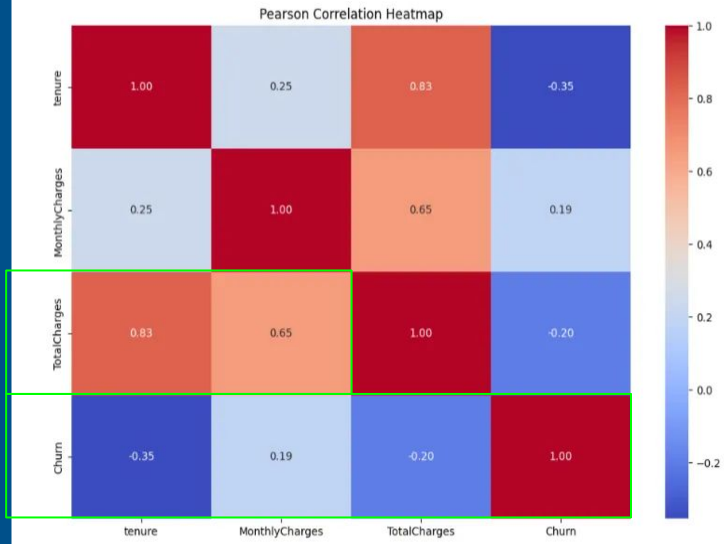


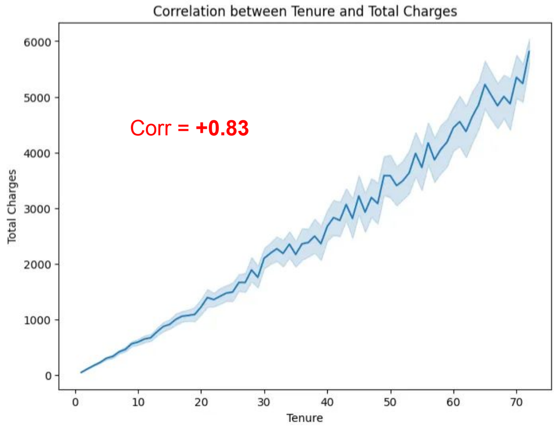
Follow-up question: *How about the numerical variables?*

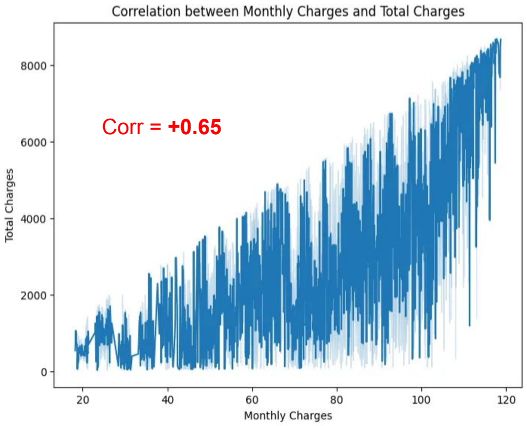
Customers with Higher Monthly Charges Are More Likely to Churn, While Longer Tenure Reduces Churn Risk

Among numerical variables, the key insights are:

- **Higher Monthly Charges** are associated with an increased likelihood of churn (**correlation: +0.19**), suggesting that cost sensitivity may be a factor in customer retention.
- **Tenure** shows a strong **negative correlation with churn (-0.35)**, indicating that long-term customers are significantly less likely to leave.
- As expected, **Tenure and Total Charges** are highly correlated (**+0.83**), since longer-tenured customers have paid more over time.
- **Monthly Charges** also show a positive relationship with **Total Charges** (**+0.65**), reflecting the cumulative nature of billing.







Summary

- **26.5% of customers churned** (1,869 out of 7,032).
- **Contract type**, especially **Month-to-Month**, significantly impacts churn. Customers on **longer contracts (One Year, Two Year)** are much less likely to churn.
- **Internet service type** also plays a role. Customers using **Fiber Optic** are more likely to churn than those using other services.
- Customers with **higher monthly charges** are more likely to churn, while those with **longer tenure** are less likely.
- While **Total Charges** doesn't directly influence churn, it has a strong correlation with both **Monthly Charges** and **Tenure**, and helps explain long-term value relationships.

Recommendations

1. **Promote longer-term contracts** (One Year, Two Year) to both new and existing customers to reduce churn.
2. **Target Month-to-Month customers** with incentives or exclusive offers to encourage switching to longer contracts.
3. **Introduce loyalty or rewards programs**, especially for long-tenured customers, to strengthen long-term engagement.
4. **Evaluate Fiber Optic service quality and customer satisfaction**, consider positioning improvements or offering alternatives with lower churn risk.
5. **Bundle value-added services** like **Tech Support, Online Security, and Device Protection** as part of retention packages.
6. **Review and optimize monthly pricing plans**, more flexible and affordable options may help retain price-sensitive customers.
7. **Revisit the Total Charges structure** to ensure pricing reflects customer-perceived value, particularly for high-value or long-term users.
8. **Investigate issues with Electronic Check payments** and consider promoting alternative payment methods with lower churn rates.

Thanks