Bhaskar Pal

CUSTOMER CHURN ANALYSIS IN THE TELECOM INDUSTRY

Identifying Key Factors Behind Churn and Retention Strategies

Bhaskar Pal

PROJECT OVERVIEW

Objective:

- **Analyze customer churn patterns in the telecom industry.**
- **Identify key factors influencing churn rates.**
- **Provide data-driven recommendations for customer retention.**

Tech Stack:

-- Language: Python

-- Libraries: Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn

-- IDE: Jupyter Notebook

Bhaskar Pal

DATASET OVERVIEW

Dataset Description: Contains customer data with multiple attributes.

Key Features Analyzed:

- **Contract Types: Month-to-month, one-year, two-year**
- **Payment Methods: Electronic checks, credit cards, bank transfers**
- **Tenure: Customer relationship duration**

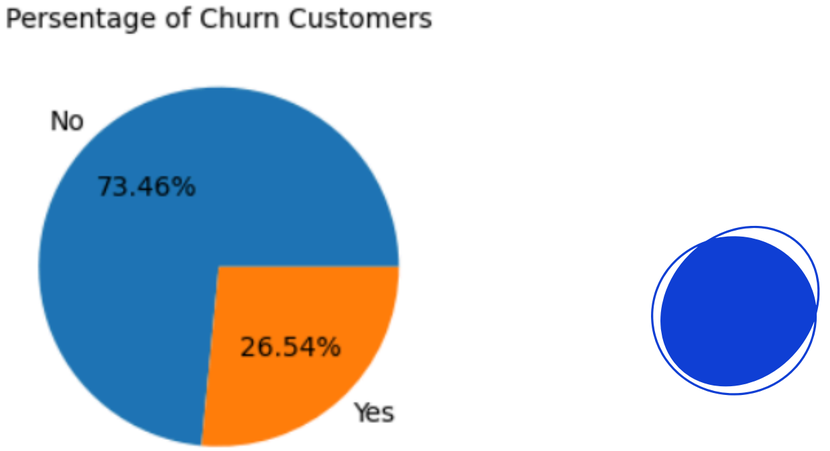
Demographics: Age group, senior citizens

Bhaskar Pal

DATA PREPROCESSING

- **Data Cleaning: Handled missing values and corrected data inconsistencies.**
- **Feature Engineering: Converted categorical variables to numerical using encoding techniques.**
- **Data Normalization: Scaled data for consistent model performance.**

```
plt.figure(figsize = (3,4))
gb = df.groupby("Churn").agg({'Churn':"count"})
plt.pie(gb['Churn'], labels = gb.index, autopct = "%1.2f%%")
plt.title("Persentage of Churn Customers", fontsize = 10)
plt.show()
```



Converted 0 and 1 values of SeniorCitizen to yes/no to make it easier to understand

```
# Change SeniorCitizen value 0,1 to yes,no:
def conv(value):
    if value == 1:
        return "yes"
    else:
        return "no"

df['SeniorCitizen'] = df["SeniorCitizen"].apply(conv)
```

Bhaskar Pal

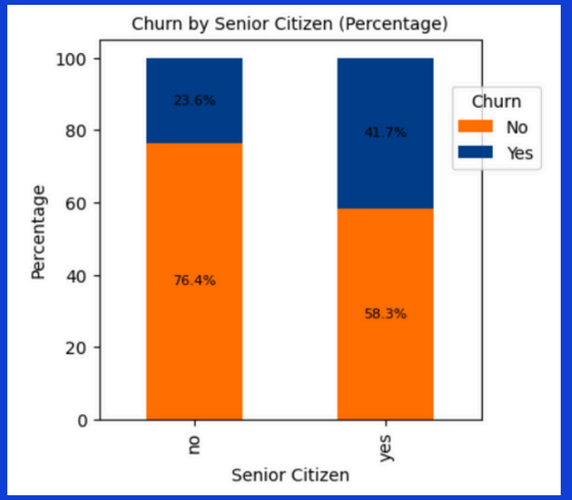
EXPLORATORY DATA ANALYSIS (EDA)

Contract Type vs Churn:

- **42% churn rate for month-to-month contracts.**
- **11% churn for one-year contracts.**
- **3% churn for two-year contracts.**

Implication: Longer contracts reduce churn risk.

```python
# Calculate percentage
df_grouped = df.groupby(["SeniorCitizen", "Churn"]).size().unstack()
df_percent = df_grouped.div(df_grouped.sum(axis=1), axis=0) * 100

# Define custom colors
colors = ["#FF6F00", "#003F88"]  # Deep orange and deep blue

# Plot stacked bar chart
ax = df_percent.plot(kind="bar", stacked=True, figsize=(4, 4), color=colors)

# Add labels
for bars in ax.containers:
    ax.bar_label(bars, fmt="%.1f%%", label_type="center", fontsize=8, color="black")

plt.title("Churn by Senior Citizen (Percentage)", fontsize=10)
plt.xlabel("Senior Citizen")
plt.ylabel("Percentage")
plt.legend(title="Churn", bbox_to_anchor =(0.9,0.9))
plt.show()
```



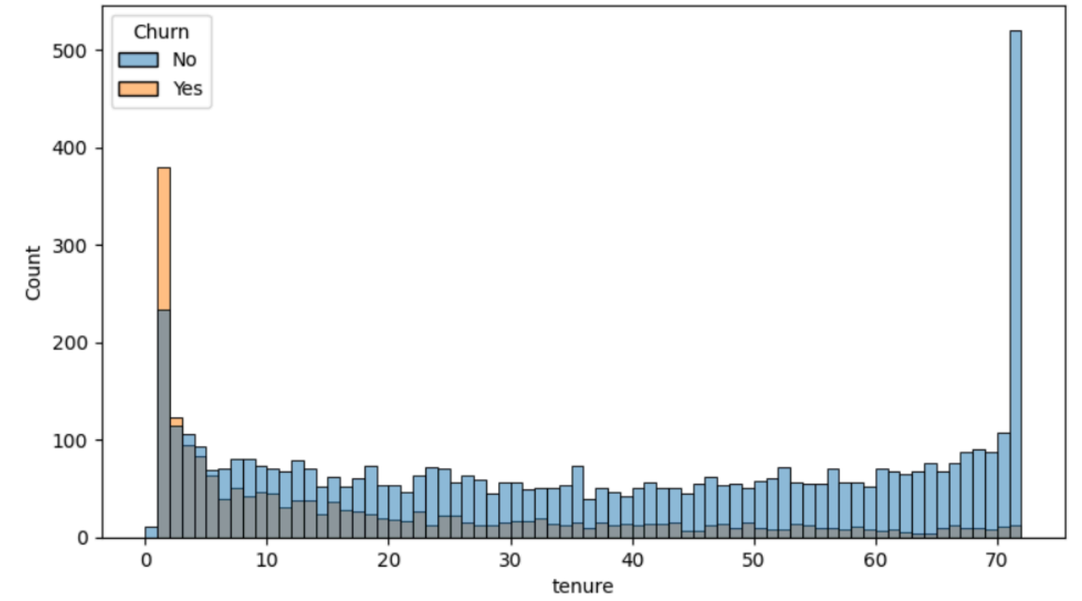
Bhaskar Pal

PAYMENT METHODS & CHURN

- **Churn Rates by Payment Methods:45% churn for electronic checks.**
- **15-18% churn for credit cards, bank transfers, and mailed checks.**
- **Implication:Encouraging customers to switch to stable payment methods can reduce churn.**

Comparatively a greated persentage of people in senior citi churned.

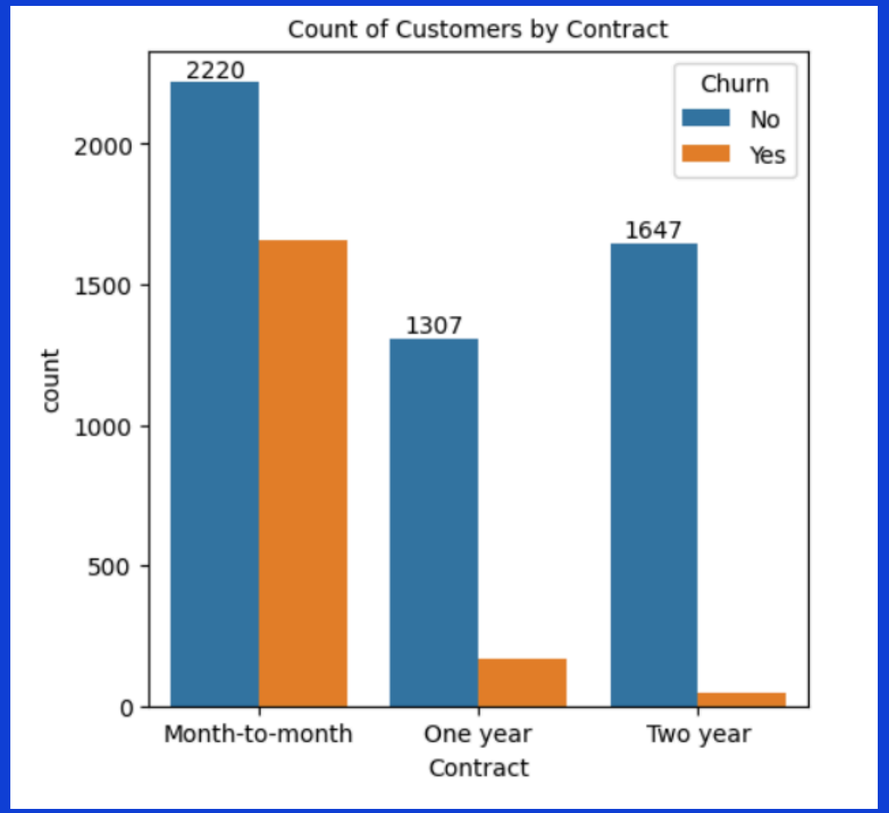
```
plt.figure(figsize = (9,5))
sns.histplot(x = "tenure", data = df, bins = 72, hue = "Churn")
plt.show()
```



Bhaskar Pal

TENURE & CHURN

- **Churn by Tenure: 50% churn for customers with <1 year tenure.**
- **35% churn for 1-3 years.**
- **15% churn for >3 years.**
- **Implication: Early engagement is essential for customer retention.**



People who have used our services for a long time have stayed and people who have used oue services 1 or 2 months have churned.

```
plt.figure(figsize = (5,5))
ax = sns.countplot(x = "Contract", data = df, hue = "Churn")
ax.bar_label(ax.containers[0])
plt.title("Count of Customers by Contract", fontsize = 10)
plt.show()
```

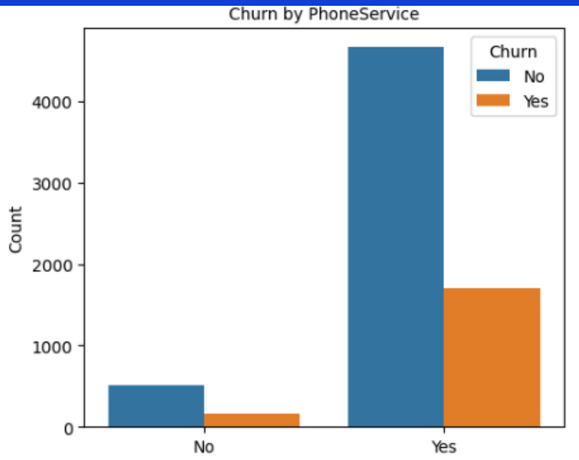
People who have month to month contract are likely to churn then from those who have 1 or 2 years of contract.

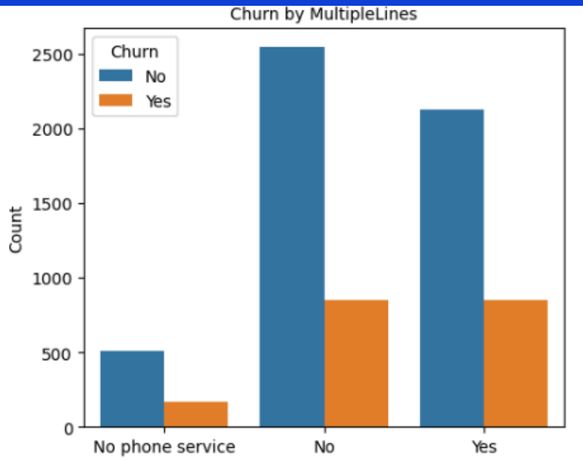
```
df.columns.values
```

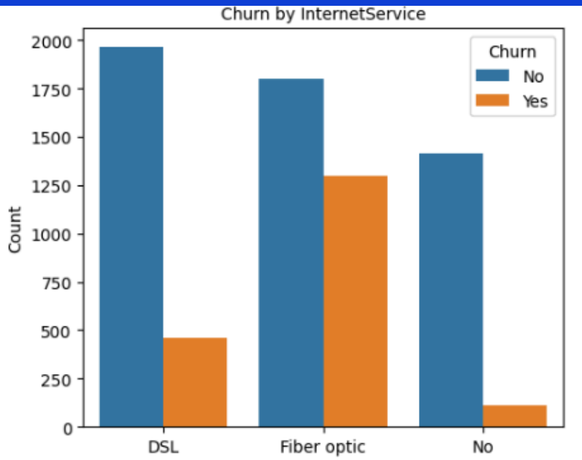
```
array(['customerID', 'gender', 'SeniorCitizen', 'Partner', 'Dependents',
       'tenure', 'PhoneService', 'MultipleLines', 'InternetService',
       'OnlineSecurity', 'OnlineBackup', 'DeviceProtection',
       'TechSupport', 'StreamingTV', 'StreamingMovies', 'Contract',
       'PaperlessBilling', 'PaymentMethod', 'MonthlyCharges',
       'TotalCharges', 'Churn'], dtype=object)
```

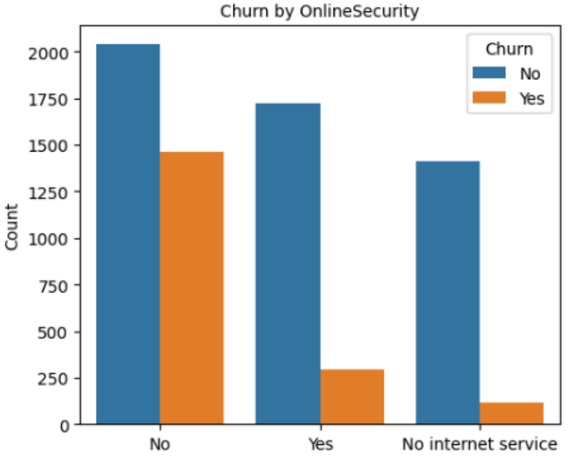
```
# Define the columns to plot
columns = ['PhoneService', 'MultipleLines', 'InternetService',
           'OnlineSecurity', 'OnlineBackup', 'DeviceProtection',
           'TechSupport', 'StreamingTV', 'StreamingMovies']

# Create subplots
fig, axes = plt.subplots(3, 3, figsize=(15, 12))  # Adjust grid size as needed
axes = axes.flatten()  # Flatten to easily iterate

# Plot countplots
for i, col in enumerate(columns):
    sns.countplot(x=col, data=df, hue="Churn", ax=axes[i])
    axes[i].set_title(f"Churn by {col}", fontsize=10)
    axes[i].set_xlabel("")
    axes[i].set_ylabel("Count")

# Adjust layout
plt.tight_layout()
plt.show()
```

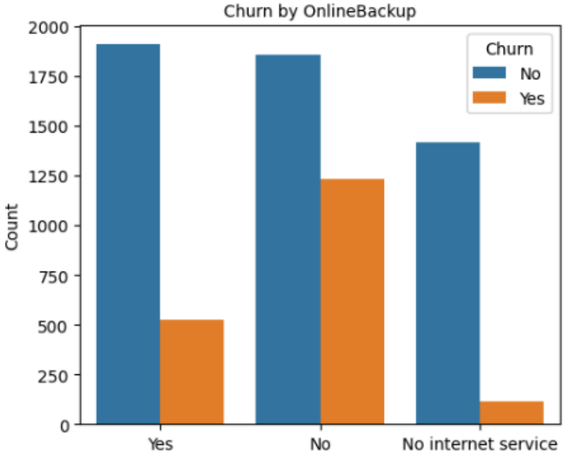
Bhaskar Pal

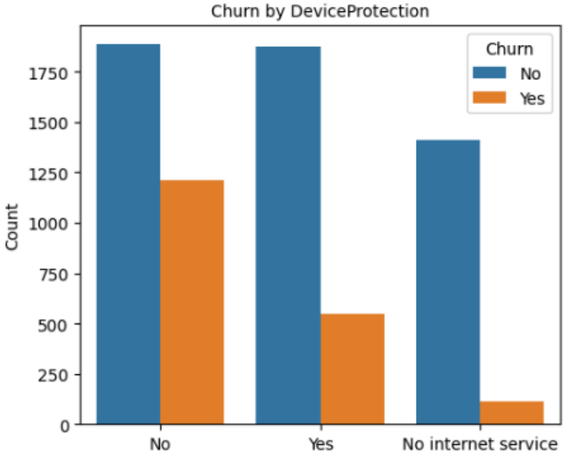

Churn by PhoneService
Churn
No
Yes
Count
0
1000
2000
3000
4000
No
Yes

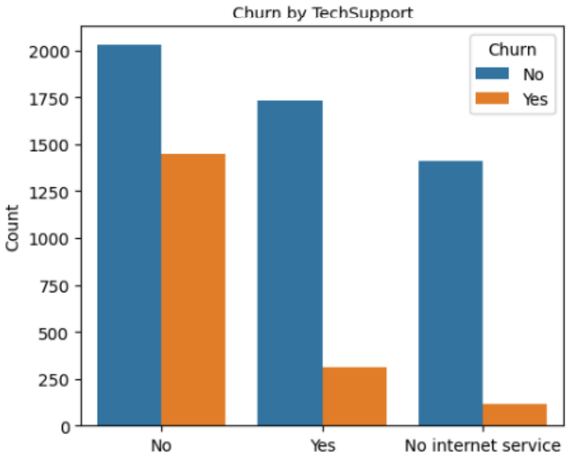

Churn by MultipleLines
Churn
No
Yes
Count
0
500
1000
1500
2000
2500
No phone service
No
Yes

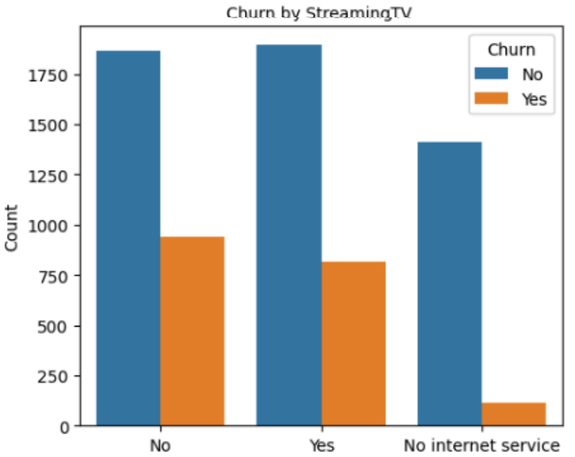

Churn by InternetService
Churn
No
Yes
Count
0
250
500
750
1000
1250
1500
1750
2000
DSL
Fiber optic
No

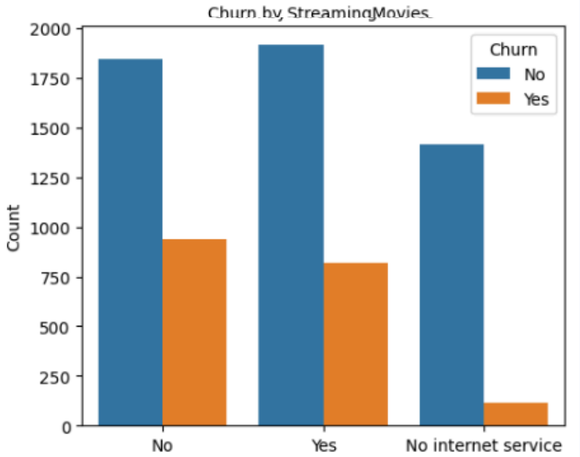

Churn by OnlineSecurity
Churn
No
Yes
Count
0
250
500
750
1000
1250
1500
1750
2000
No
Yes
No internet service


Churn by OnlineBackup
Churn
No
Yes
Count
0
250
500
750
1000
1250
1500
1750
2000
Yes
No
No internet service


Churn by DeviceProtection
Churn
No
Yes
Count
0
250
500
750
1000
1250
1500
1750
No
Yes
No internet service


Churn by TechSupport
Churn
No
Yes
Count
0
250
500
750
1000
1250
1500
1750
2000
No
Yes
No internet service


Churn by StreamingTV
Churn
No
Yes
Count
0
250
500
750
1000
1250
1500
1750
No
Yes
No internet service


Churn by StreamingMovies
Churn
No
Yes
Count
0
250
500
750
1000
1250
1500
1750
2000
No
Yes
No internet service

Bhaskar Pal

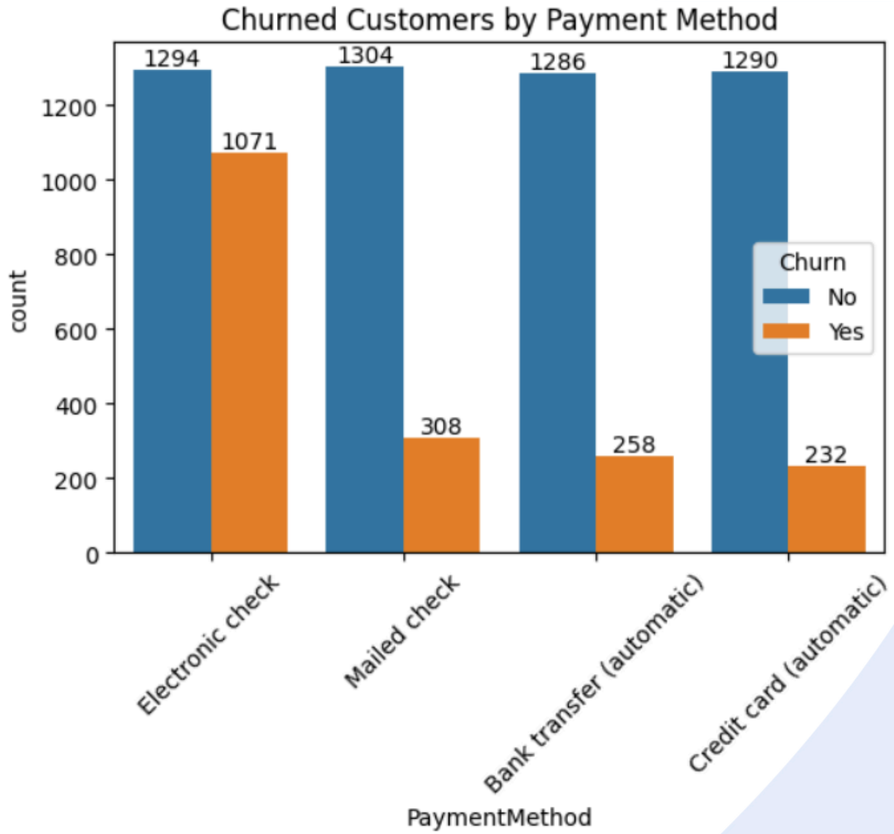
MODEL & PREDICTIONS

-- Models Used:

- Logistic Regression
- Decision Tree

-- Results:

- Achieved high accuracy in predicting churn likelihood.
- Identified key factors influencing customer attrition.



The majority of customers who do not churn tend to have services like PhoneService, InternetService (particularly DSL), and OnlineSecurity enabled. For services like OnlineBackup, TechSupport, and StreamingTV, churn rates are noticeably higher when these services are not used or are unavailable.

```
plt.figure(figsize = (6,4))
ax = sns.countplot(x = "PaymentMethod", data = df, hue = "Churn")
ax.bar_label(ax.containers[0])
ax.bar_label(ax.containers[1])
plt.title("Churned Customers by Payment Method")
plt.xticks(rotation = 45)
plt.show()
```

Bhaskar Pal

CONCLUSION & RECOMMENDATIONS

-- Key Takeaways:

- **Longer contracts reduce churn.**
- **Promoting stable payment methods improves retention.**
- **Early engagement with new customers is crucial.**
- **Service quality improvements retain Fiber Optic users.**

-- Future Scope:

- **Apply more advanced models (e.g., Random Forest, XGBoost).**
- **Use customer feedback data for sentiment analysis.**

Bhaskar Pal

THANK YOU