

Abdi-278 /
Phase-3-project

Code

Issues

Pull requests

Actions

Projects

Wiki

Security

[Phase-3-project](#) / [Desktop](#) / [bigfolder](#) / [phase 3 project](#) / [notebook PDF.pdf](#) 

Abdi-278 first commit

2ff318c · 13 minutes ago



5.94 MB



INTRODUCTION

As a telecommunications company, we are facing challenges in retaining customers. Customer churn, defined as the rate at which customers stop doing business with a company, is a significant issue impacting revenue and profitability. To address this, we aim to build a customer churn prediction model that will help identify customers likely to leave and enable us to take proactive retention measures.

The analysis uses historical customer data from SyriaTel to uncover patterns and insights that drive churn behavior and provide actionable strategies to improve customer retention.

BACKGROUND

According to industry research, customer churn is one of the largest challenges for telecom companies, resulting in billions of dollars in lost revenue annually. High competition, price wars, and poor customer experience often lead to customer dissatisfaction.

Telecom companies typically analyze customer demographics, service usage, and interactions with customer support to understand churn patterns. By identifying at-risk customers, businesses can implement strategies like loyalty programs, improved services, and better customer engagement.

This project leverages historical customer behavior data to predict churn, reduce revenue loss, and ensure long-term business growth.

BUSINESS PROBLEM

Customer churn results in significant revenue loss for telecommunications companies. SyriaTel currently faces increasing churn rates, which threaten its market position and financial stability.

The primary challenge is to identify customers at risk of churning and determine the underlying factors contributing to churn. SyriaTel lacks a systematic approach to predict churn and retain customers, making it essential to leverage data analytics to build a churn prediction model.

By addressing this problem, the company can reduce customer attrition, improve customer satisfaction, and increase profitability.

OBJECTIVES

1.Explore the key factors that influence customer churn. 2.Build a machine learning classifier to predict customer churn. 3.Identify actionable insights to reduce churn and

Data Preparation

```
In [1]: #Import the necessary libraries
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np
```

```
In [2]: #Loading the data
df=pd.read_csv('churn.csv')
df.head()
```

Out[2]:

	state	account length	area code	phone number	international plan	voice mail plan	number vmail messages	total day minutes	total day calls	total day charge
0	KS	128	415	382-4657	no	yes	25	265.1	110	45.0
1	OH	107	415	371-7191	no	yes	26	161.6	123	27.4
2	NJ	137	415	358-1921	no	no	0	243.4	114	41.3
3	OH	84	408	375-9999	yes	no	0	299.4	71	50.9
4	OK	75	415	330-6626	yes	no	0	166.7	113	28.3

5 rows × 21 columns



```
In [3]: #checking for how many columns
df.columns
```

```
Out[3]: Index(['state', 'account length', 'area code', 'phone number',
       'international plan', 'voice mail plan', 'number vmail messages',
       'total day minutes', 'total day calls', 'total day charge',
       'total eve minutes', 'total eve calls', 'total eve charge',
       'total night minutes', 'total night calls', 'total night charge',
       'total intl minutes', 'total intl calls', 'total intl charge',
       'customer service calls', 'churn'],
      dtype='object')
```

```
In [4]: #checking how many rows and columns
df.shape
```

Out[4]: (3333, 21)

```
In [5]: #checking for null values
df.isna().sum()
```

```
Out[5]: state          0
account length    0
area code         0
phone number      0
international plan 0
voice mail plan   0
```

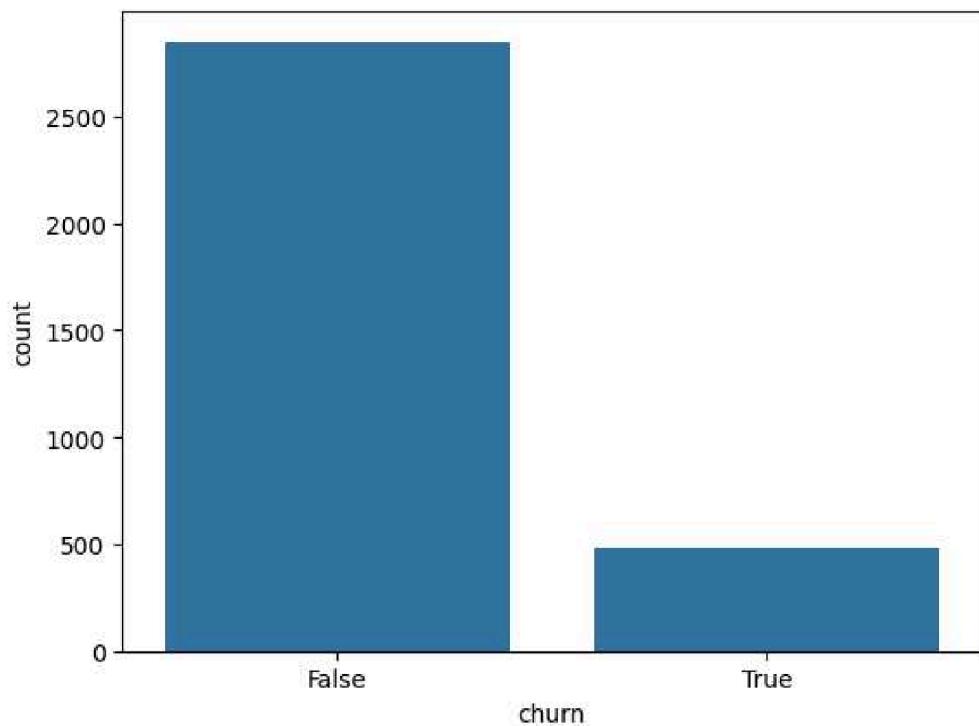
```
number vmail messages      0
total day minutes         0
total day calls           0
total day charge          0
total eve minutes         0
total eve calls           0
total eve charge          0
total night minutes       0
total night calls         0
total night charge        0
total intl minutes        0
total intl calls          0
total intl charge         0
customer service calls   0
churn                      0
dtype: int64
```

In [6]: # checking for column information
df.info()

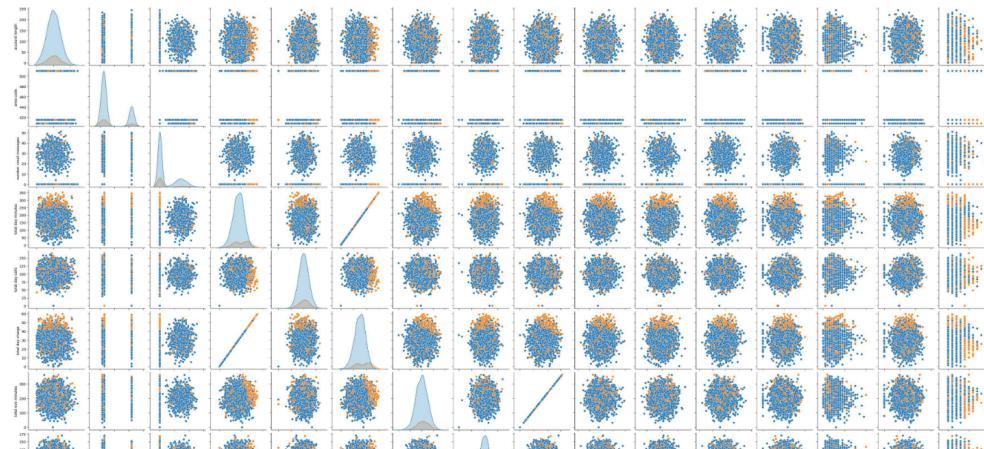
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3333 entries, 0 to 3332
Data columns (total 21 columns):
 #   Column            Non-Null Count  Dtype  
 ---  -- 
 0   state              3333 non-null   object 
 1   account length     3333 non-null   int64  
 2   area code          3333 non-null   int64  
 3   phone number       3333 non-null   object 
 4   international plan 3333 non-null   object 
 5   voice mail plan    3333 non-null   object 
 6   number vmail messages 3333 non-null   int64  
 7   total day minutes  3333 non-null   float64
 8   total day calls    3333 non-null   int64  
 9   total day charge   3333 non-null   float64
 10  total eve minutes  3333 non-null   float64
 11  total eve calls   3333 non-null   int64  
 12  total eve charge   3333 non-null   float64
 13  total night minutes 3333 non-null   float64
 14  total night calls  3333 non-null   int64  
 15  total night charge 3333 non-null   float64
 16  total intl minutes 3333 non-null   float64
 17  total intl calls   3333 non-null   int64  
 18  total intl charge   3333 non-null   float64
 19  customer service calls 3333 non-null   int64  
 20  churn               3333 non-null   bool  
dtypes: bool(1), float64(8), int64(8), object(4)
memory usage: 524.2+ KB
```

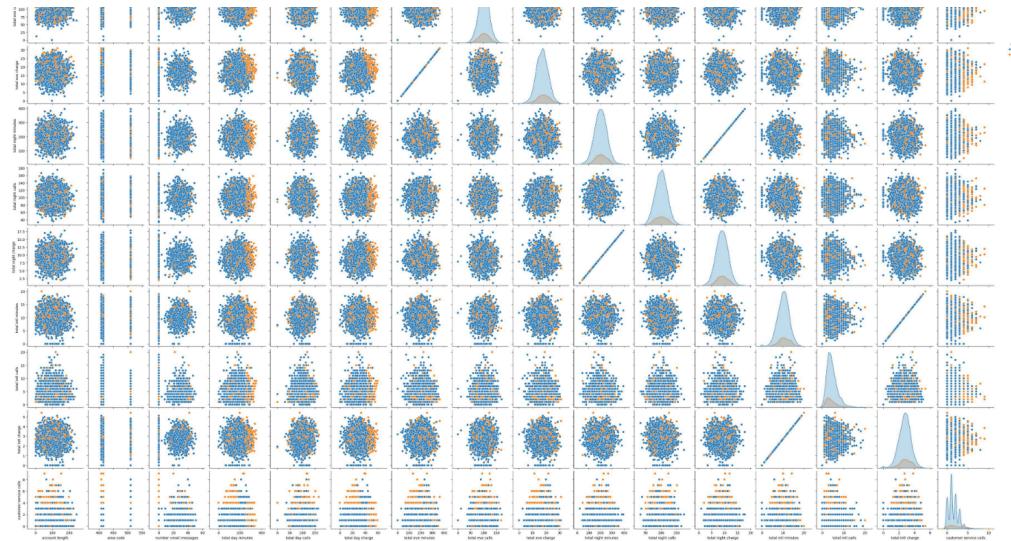
In [7]: # Create numeric & categorical lists
numeric_cols = ['account length','number vmail messages','total day minutes','total eve minutes','total eve calls','total eve charge','total night minutes','total night calls','total intl minutes','total intl charge']
categoric_cols = ['state','area code','international plan','voice mail plan']

In [8]: # Countplot of churn feature
sns.countplot(data=df, x='churn')
plt.show()



```
In [9]: sns.pairplot(df,hue="churn")
plt.show()
```





```
In [10]: sns.displot(df,x="area code",hue="churn",kde=True)  
plt.show()
```

More Pages

