

```
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
```

```
sns.set()
```

```
from google.colab import drive
drive.mount('/content/drive/')
```

Drive already mounted at /content/drive/; to attempt to forcibly remount, call drive.mount("/content/drive/", force_remount=True).

```
raw_data = pd.read_csv('/content/drive/MyDrive/Colab Notebooks/customer_churn_processed.csv')
data = raw_data.copy()
data
```

	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	OnlineBackup	DeviceProtection	Te
0	1	0	1	1	1	1	0	1	0	1	0	
1	0	0	0	1	34	0	1	1	1	0	1	
2	0	0	0	1	2	0	1	1	1	1	0	
3	0	0	0	1	45	1	0	1	1	0	1	
4	1	0	0	1	2	0	1	2	0	0	0	
...	
7005	0	0	1	0	24	0	2	1	1	0	1	
7006	1	0	1	0	72	0	2	2	0	1	1	
7007	1	0	1	0	11	1	0	1	1	0	0	
7008	0	1	1	1	4	0	2	2	0	0	0	
7009	0	0	0	1	66	0	1	2	1	0	1	

7010 rows × 20 columns

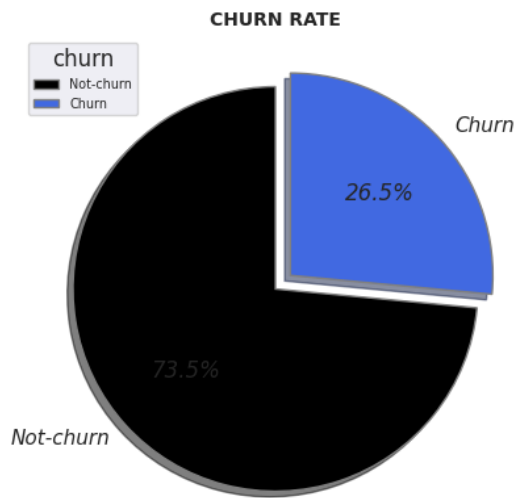
✦ Exploratory Data Analysis

```
colors = ['black', 'royalblue', 'brown', 'yellow', 'biege']
```

```
churn = data['Churn'].value_counts()
```

```
churn = data['Churn'].value_counts()
fig = plt.figure(figsize=(20,5))
churn_pie_label = ['Not-churn', 'Churn']
wp = {'linewidth':1, 'edgecolor':'grey', 'antialiased':True}
textprops = {'fontstyle':'italic'}
explode = [0,0.1]
plt.pie(churn, colors=colors, autopct="%0.1f%%", labels=churn_pie_label, startangle=90, explode=explode, shadow=True, wedgeprops=wp, textprops=textprops)
plt.legend(title='churn', loc='upper left',fontsize='xx-small')
plt.title('CHURN RATE', fontsize='small', fontweight='bold')
plt.show
```

```
<function matplotlib.pyplot.show(close=None, block=None)>
```



the chart above shows that the data is not balanced, it has a ratio of 1:3 which is not ideal for training the model

the data variable will be divided into three major parts depending on the value they provide

Customer personal information

Gender|Senior citizen|Partner|Dependents

```
gender_pie = data[data['Churn'] == 1]['gender'].value_counts()
senior_citizenship_pie = data[data['Churn'] == 1]['SeniorCitizen'].value_counts()
partner_pie = data[data['Churn'] == 1]['Partner'].value_counts()
dependents = data[data['Churn'] == 1]['Dependents'].value_counts()
```

```
fig, (ax1, ax2, ax3, ax4) = plt.subplots(nrows=1, ncols=4, figsize=(20,5))
```

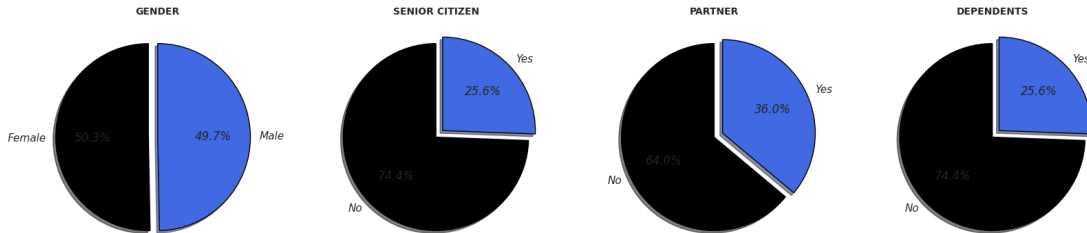
```
pie_label = ['Female', 'Male']
explode = [0.1,0]
wp = {'linewidth':1, 'edgecolor':'black', 'antialiased':True}
textprops = {'fontstyle':'italic'}
ax1.pie(gender_pie, colors=colors, autopct="%0.1f%%", labels=pie_label, startangle=90, explode=explode, shadow=True, wedgeprops=wp, textprops=textprops)
ax1.set_title('GENDER', fontsize='small', fontweight='bold')
```

```
pie_label = ['No', 'Yes']
wp = {'linewidth':1, 'edgecolor':'black', 'antialiased':True}
textprops = {'fontstyle':'italic'}
explode = [0,0.1]
ax2.pie(senior_citizenship_pie, colors=colors, autopct="%0.1f%%", labels=pie_label, startangle=90, explode=explode, shadow=True, wedgeprops=wp, textprops=textprops)
ax2.set_title('SENIOR CITIZEN', fontsize='small', fontweight='bold')
```

```
pie_label = ['No', 'Yes']
wp = {'linewidth':1, 'edgecolor':'black', 'antialiased':True}
textprops = {'fontstyle':'italic'}
explode = [0,0.1]
ax3.pie(partner_pie, colors=colors, autopct="%0.1f%%", labels=pie_label, startangle=90, explode=explode, shadow=True, wedgeprops=wp, textprops=textprops)
ax3.set_title('PARTNER', fontsize='small', fontweight='bold')
```

```
pie_label = ['No', 'Yes']
wp = {'linewidth':1, 'edgecolor':'black', 'antialiased':True}
textprops = {'fontstyle':'italic'}
explode = [0,0.1]
ax4.pie(senior_citizenship_pie, colors=colors, autopct="%0.1f%%", labels=pie_label, startangle=90, explode=explode, shadow=True, wedgeprops=wp, textprops=textprops)
ax4.set_title('DEPENDENTS', fontsize='small', fontweight='bold')
plt.show
```

```
<function matplotlib.pyplot.show(close=None, block=None)>
```



Gender

from the above data we can see that there is a nearly 50% to 50% distribution of male and female customers that churn, which tells us gender does not really play a role in customer churn.

Senior Citizen

as seen from the chart 74.4% of churn customers are not senior citizens this is an information that can be vital to the company

Partner & Dependents

from partner and dependent data, 69.2 of the churned were living by themselves

services subscribed to by Customer

Phone services|Multiple lines|Internet service|Streaming Tv|Streaming movies

```
phoneservice_pie = data[data['Churn'] == 1]['PhoneService'].value_counts()
multiplelines_pie = data[data['Churn'] == 1]['MultipleLines'].value_counts()
internetservice_pie = data[data['Churn'] == 1]['InternetService'].value_counts()
streamingtv_pie = data[data['Churn'] == 1]['StreamingTV'].value_counts()
streamingmovies_pie = data[data['Churn'] == 1]['StreamingMovies'].value_counts()

fig, (ax1, ax2, ax3, ax4, ax5) = plt.subplots(nrows=1, ncols=5, figsize=(20,4))

labels = [ 'Yes', 'No']
explode = (0,0.1)
textprops = {'fontstyle':'italic'}
wp={'linewidth':1, 'edgecolor':'black', 'antialiased':True}
ax1.set_title('PHONE SERVICE')
ax1.pie(phoneservice_pie, labels=labels, explode=explode, colors=colors, autopct='%0.1f%%', shadow=True, textprops=textprops, wedgeprops=wp, startangle=90)

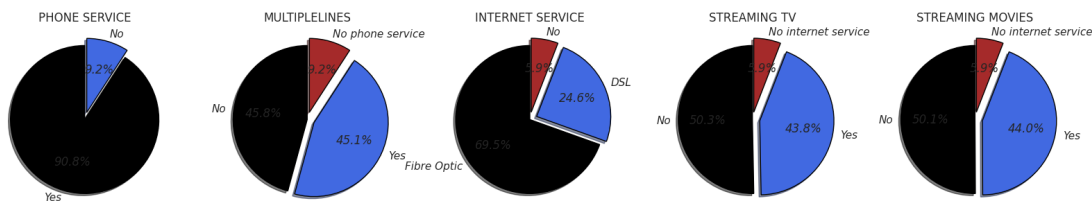
labels = ['No', 'Yes', 'No phone service']
explode = (0,0.1,0.1)
textprops = {'fontstyle':'italic'}
wp={'linewidth':1, 'edgecolor':'black', 'antialiased':True}
ax2.set_title('MULTIPLE LINES')
ax2.pie(multiplelines_pie, labels=labels, explode=explode, colors=colors, autopct='%0.1f%%', shadow=True, textprops=textprops, wedgeprops=wp, startangle=90)

labels = ['Fibre Optic', 'DSL', 'No']
explode = (0,0.1,0.1)
textprops = {'fontstyle':'italic'}
wp={'linewidth':1, 'edgecolor':'black', 'antialiased':True}
ax3.set_title('INTERNET SERVICE')
ax3.pie(internetservice_pie, labels=labels, explode=explode, colors=colors, autopct='%0.1f%%', shadow=True, textprops=textprops, wedgeprops=wp, startangle=90)

labels = ['No', 'Yes', 'No internet service']
explode = (0,0.1,0.1)
textprops = {'fontstyle':'italic'}
wp={'linewidth':1, 'edgecolor':'black', 'antialiased':True}
ax4.set_title('STREAMING TV')
ax4.pie(streamingtv_pie, labels=labels, explode=explode, colors=colors, autopct='%0.1f%%', shadow=True, textprops=textprops, wedgeprops=wp, startangle=90)

labels = ['No', 'Yes', 'No internet service']
explode = (0,0.1,0.1)
textprops = {'fontstyle':'italic'}
wp={'linewidth':1, 'edgecolor':'black', 'antialiased':True}
ax5.set_title('STREAMING MOVIES')
ax5.pie(streamingmovies_pie, labels=labels, explode=explode, colors=colors, autopct='%0.1f%%', shadow=True, textprops=textprops, wedgeprops=wp, startangle=90)

plt.show()
```



Phone Service

despite having phone service on 90.8% of customer churn which is quite high

Multiple Lines

availability of multiple lines did not matter as customers churn regardless of it

Internet Service

upto 69.5 customers churn opted out of the fibre optic cables, this shows low appreciation for the service

Streaming Tv & Streaming Movies

from the streaming tv and streaming movies data we can see that an average of 43.9 customers churn despite using this service.

services subscribed to by customer

Online security|Online backup|Device protection|Tech support

```

onlinesecurity_pie = data[data['Churn'] == 1]['OnlineSecurity'].value_counts()
onlinebackup_pie = data[data['Churn'] == 1]['OnlineBackup'].value_counts()
deviceprotection_pie = data[data['Churn'] == 1]['DeviceProtection'].value_counts()
techsupport_pie = data[data['Churn'] == 1]['TechSupport'].value_counts()

fig,(ax1,ax2,ax3,ax4) = plt.subplots(1,4,figsize=(20,4))

labels = ['No', 'Yes', 'No internet service']
explode = [0,0.1,0.1]
textprops = {'fontstyle':'italic'}
wp={'linewidth':1, 'edgecolor':'black', 'antialiased':True}
ax1.pie(onlinesecurity_pie, colors=colors, autopct='%0.1f%%', labels=labels, shadow=True, textprops=textprops, explode=explode, wedgeprops=wp, startangle=90)
ax1.set_title('ONLINE SECURITY')

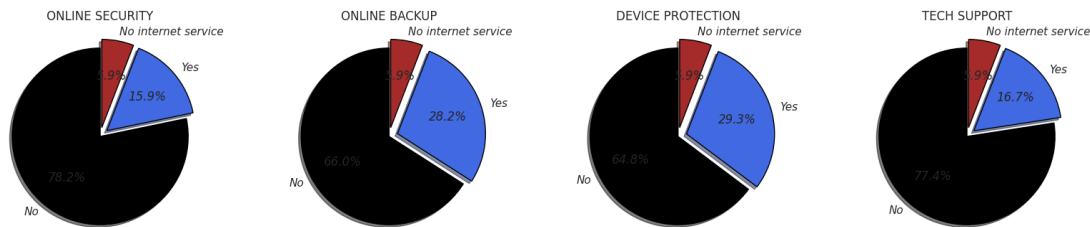
labels = ['No', 'Yes', 'No internet service']
explode = [0,0.1,0.1]
textprops = {'fontstyle':'italic'}
wp={'linewidth':1, 'edgecolor':'black', 'antialiased':True}
ax2.pie(onlinebackup_pie, colors=colors, autopct='%0.1f%%', labels=labels, shadow=True, textprops=textprops, explode=explode, wedgeprops=wp, startangle=90)
ax2.set_title('ONLINE BACKUP')

labels = ['No', 'Yes', 'No internet service']
explode = [0,0.1,0.1]
textprops = {'fontstyle':'italic'}
wp={'linewidth':1, 'edgecolor':'black', 'antialiased':True}
ax3.pie(deviceprotection_pie, colors=colors, autopct='%0.1f%%', labels=labels, shadow=True, textprops=textprops, explode=explode, wedgeprops=wp, startangle=90)
ax3.set_title('DEVICE PROTECTION')

labels = ['No', 'Yes', 'No internet service']
explode = [0,0.1,0.1]
textprops = {'fontstyle':'italic'}
wp={'linewidth':1, 'edgecolor':'black', 'antialiased':True}
ax4.pie(techsupport_pie, colors=colors, autopct='%0.1f%%', labels=labels, shadow=True, textprops=textprops, explode=explode, wedgeprops=wp, startangle=90)
ax4.set_title('TECH SUPPORT')

plt.show()

```



Online Security, Online Backup, Device Protection & Tech Support
from the chart 71.6% of customers churn as a lack of this services.

Customer contracts

Contract|Paperless billing|Payment method

```
contract_pie = data[data['Churn'] == 1]['Contract'].value_counts()
paperlessbilling_pie = data[data['Churn'] == 1]['PaperlessBilling'].value_counts()
paymentmethod_pie = data[data['Churn'] == 1]['PaymentMethod'].value_counts()
```

```
fig,(ax1,ax2,ax3) = plt.subplots(1, 3, figsize=(20,4))
```

```
labels = ['Month to Month', 'One year', 'Two year']
explode = [0,0.1,0.1]
textprops = {'fontstyle':'italic'}
wp={'linewidth':1, 'edgecolor':'black', 'antialiased':True}
ax1.pie(contract_pie, colors=colors, autopct='%0.1f%%', labels=labels, shadow=True, textprops=textprops, explode=explode, wedgeprops=wp, startangle=90)
ax1.set_title('CONTRACT')
```

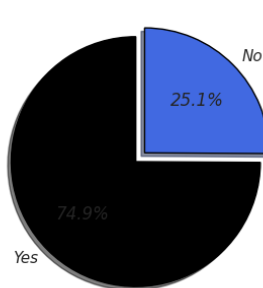
```
labels = ['Yes', 'No']
explode = [0,0.1]
textprops = {'fontstyle':'italic'}
wp={'linewidth':1, 'edgecolor':'black', 'antialiased':True}
ax2.pie(paperlessbilling_pie, colors=colors, autopct='%0.1f%%', labels=labels, shadow=True, textprops=textprops, explode=explode, wedgeprops=wp, startangle=90)
ax2.set_title('PAPERLESS BILLING')
```

```
labels = ['Electronic check', 'Mailed check', 'Bank tranfer (automatic)', 'Credit card (automatic)']
explode = [0,0.1,0.1,0.1]
textprops = {'fontstyle':'italic'}
wp={'linewidth':1, 'edgecolor':'black', 'antialiased':True}
ax3.pie(paymentmethod_pie, colors=colors, autopct='%0.1f%%', labels=labels, shadow=True, textprops=textprops, explode=explode, wedgeprops=wp, startangle=90)
ax3.set_title('PAYMENT METHOD')
```

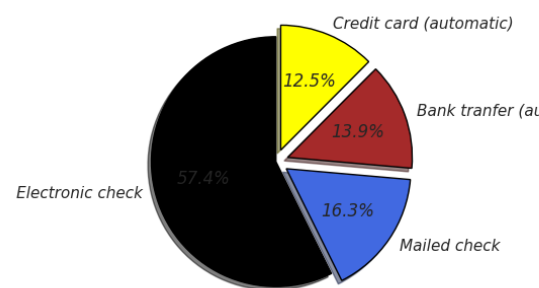
Text(0.5, 1.0, 'PAYMENT METHOD')



PAPERLESS BILLING



PAYMENT METHOD



Month to Month

when it comes to churning its clear that month to month has the highest share with 88.5%

Paperless Billing

74.9% have paper billing and it does matter

Electronic check

up to 57.4 of customers churn while subscribed to the electronic check, this needs to be sorted

Distribution of Numerical features

```
# Subsetting the data for churn customers
churn_data = data[data['Churn'] == 1]

# Defining the numeric columns
numeric_data = ['tenure', 'MonthlyCharges', 'TotalCharges']

# Creating subplots with 1 row and 3 columns
fig, ax = plt.subplots(nrows=1, ncols=3, figsize=(15, 5))

# Iterating over the numeric data columns
for i in range(len(numeric_data)):
    # Plotting a distribution plot (displot) for the current numeric column
    sns.histplot(data=churn_data[numeric_data[i]], kde=True, ax=ax[i])

    # Setting title for the current subplot
    ax[i].set_title('Distribution of ' + numeric_data[i] + ' for Churn Customers')

plt.tight layout()
```