

Data exploration

```
In []: from google.colab import drive
      drive.mount('/content/drive')

Mounted at /content/drive

In []: #optional to adjust coding space
      import pandas as pd
      pd.set_option('display.width', 180)

In []: #optional to adjust frame height
      from google.colab import output
      from IPython.display import Javascript

      def resize_colab_cell():
          display(Javascript('google.colab.output.setIframeHeight(0, true, {maxHeight:
          5000})))

      get_ipython().events.register('pre_run_cell', resize_colab_cell)

In []: import pandas as pd

      df = pd.read_excel('/content/drive/MyDrive/Task14/Data/telecom_users.xlsx')
      # First 5 rows of the DataFrame
      print(df.head())

      customerID  gender  SeniorCitizen  Partner  Dependents  Tenure  PhoneService  MultipleLin
es InternetService OnlineSecurity  ... StreamingTV StreamingMovies  Contract \
0 7590-VHVEG Female 0 Yes No 1 No No phone servic
e DSL No ... No No Month-to-month
1 5575-GNVDE Male 0 No No 34 Yes No
DSL Yes ... No No One year
2 3668-QPYBK Male 0 No No 2 Yes No
DSL Yes ... No No Month-to-month
3 7795-CFOCW Male 0 No No 45 No No phone service
DSL Yes ... No No One year
4 9237-HQITU Female 0 No No 2 Yes No
Fiber optic No ... No No Month-to-month

      PaperlessBilling  PaymentMethod  MonthlyCharges  TotalCharges  Churn  Provider Ag
e
0 Yes Electronic check 29.85 29.85 No NaN NaN
1 No Mailed check 56.95 1889.5 No NaN NaN
2 Yes Mailed check 53.85 108.15 Yes NaN NaN
3 No Bank transfer (automatic) 42.30 1840.75 No NaN NaN
4 Yes Electronic check 70.70 151.65 Yes NaN NaN

[5 rows x 23 columns]

In []: import pandas as pd

      df = pd.read_excel('/content/drive/MyDrive/Task14/Data/telecom_users.xlsx')
```

```
# First 10 rows of the DataFrame
print(df.head(10))

customerID  gender  SeniorCitizen  Partner  Dependents  Tenure  PhoneService  MultipleLin
es InternetService OnlineSecurity  ... StreamingTV StreamingMovies  Contract \
0 7590-VHVEG  Female                0      Yes                No          1          No  No phone servic
e                DSL                No  ...                No          No  Month-to-month
1 5575-GNVDE  Male                0      No                No          34          Yes                No
DSL                Yes  ...                No          One year
2 3668-QPYBK  Male                0      No                No          2          Yes                No
DSL                Yes  ...                No  Month-to-month
3 7795-CFOCW  Male                0      No                No          45          No  No phone service
DSL                Yes  ...                No          One year
4 9237-HQITU  Female                0      No                No          2          Yes                No
Fiber optic                No  ...                No  Month-to-month
5 9305-CDSKC  Female                0      No                No          8          Yes                Yes
Fiber optic                No  ...                Yes  Yes  Month-to-month
6 1452-KIOVK  Male                0      No                Yes         22          Yes                Yes
Fiber optic                No  ...                Yes  No  Month-to-month
7 6713-OKOMC  Female                0      No                No          10          No  No phone servic
e                DSL                Yes  ...                No  Month-to-month
8 6388-TABGU  Male                0      No                Yes         62          Yes                No
DSL                Yes  ...                No          One year
9 9763-GRSKD  Male                0      Yes                Yes         13          Yes                No
DSL                Yes  ...                No  Month-to-month

PaperlessBilling  PaymentMethod  MonthlyCharges  TotalCharges  Churn  Provider  Ag
e
0                Yes          Electronic check          29.85          29.85  No      NaN  NaN
1                No          Mailed check          56.95          1889.5  No      NaN  NaN
2                Yes          Mailed check          53.85          108.15  Yes     NaN  NaN
3                No  Bank transfer (automatic)          42.30          1840.75  No      NaN  NaN
4                Yes          Electronic check          70.70          151.65  Yes     NaN  NaN
5                Yes          Electronic check          99.65           820.5  Yes     NaN  NaN
6                Yes  Credit card (automatic)          89.10          1949.4  No      NaN  NaN
7                No          Mailed check          29.75          301.9  No      NaN  NaN
8                No  Bank transfer (automatic)          56.15          3487.95  No      NaN  NaN
9                Yes          Mailed check          49.95          587.45  No      NaN  NaN

[10 rows x 23 columns]
In[]: import pandas as pd

df = pd.read_excel('/content/drive/MyDrive/Task14/Data/telecom_users.xlsx')

# Print the first 5 rows of the DataFrame
print(df.head())
```

	customerID	gender	SeniorCitizen	Partner	Dependents	Tenure	PhoneService	MultipleLin
es	InternetService		OnlineSecurity	...	StreamingTV	StreamingMovies	Contract \	
0	7590-VHVEG	Female	0	Yes	No	1	No	No phone servic
e	DSL		No	...	No	No	Month-to-month	
1	5575-GNVDE	Male	0	No	No	34	Yes	No
DSL	Yes	...	No		No	One year		
2	3668-QPYBK	Male	0	No	No	2	Yes	No
DSL	Yes	...	No		No	Month-to-month		
3	7795-CFOCW	Male	0	No	No	45	No	No phone service
DSL	Yes	...	No		No	One year		
4	9237-HQITU	Female	0	No	No	2	Yes	No
Fiber optic			No	...	No	No	Month-to-month	

	PaperlessBilling		PaymentMethod	MonthlyCharges	TotalCharges	Churn	Provider	Ag
0	Yes		Electronic check	29.85	29.85	No	NaN	NaN
1	No		Mailed check	56.95	1889.5	No	NaN	NaN
2	Yes		Mailed check	53.85	108.15	Yes	NaN	NaN
3	No	Bank transfer (automatic)		42.30	1840.75	No	NaN	NaN
4	Yes		Electronic check	70.70	151.65	Yes	NaN	NaN

```
[5 rows x 23 columns]
In[: print(df.tail(10))
```

	customerID	gender	SeniorCitizen	Partner	Dependents	Tenure	PhoneService	Multiple
Lines	InternetService		OnlineSecurity	...	StreamingTV	\		
6040	0871-OPBXW	Female	0	No	No	2	Yes	
No	No	No internet service	...	No internet service				
6041	3605-JISKB	Male	1	Yes	No	55	Yes	
Yes	DSL		Yes	...	No			
6042	6894-LFHLY	Male	1	No	No	1	Yes	
Yes	Fiber optic		No	...	No			
6043	9767-FFLEM	Male	0	No	No	38	Yes	
No	Fiber optic		No	...	No			
6044	8456-QDAVC	Male	0	No	No	19	Yes	
No	Fiber optic		No	...	Yes			
6045	6840-RESVB	Male	0	Yes	Yes	24	Yes	
Yes	DSL		Yes	...	Yes			
6046	2234-XADUH	Female	0	Yes	Yes	72	Yes	
Yes	Fiber optic		No	...	Yes			
6047	4801-JZAZL	Female	0	Yes	Yes	11	No	No phone ser
vice	DSL		Yes	...	No			
6048	8361-LTMKD	Male	1	Yes	No	4	Yes	
Yes	Fiber optic		No	...	No			
6049	3186-AJIEK	Male	0	No	No	66	Yes	
No	Fiber optic		Yes	...	Yes			

	StreamingMovies		Contract	PaperlessBilling		PaymentMethod	Monthly
Charges	TotalCharges	Churn	Provider	Age			
6040	No internet service	Month-to-month		Yes		Mailed check	
20.05	39.25	No	NaN	NaN			
6041		No	One year	No	Credit card (automatic)		

60.00	3316.1	No	NaN NaN		
6042		No	Month-to-month	Yes	Electronic check
75.75	75.75	Yes	NaN NaN		
6043		No	Month-to-month	Yes	Credit card (automatic)
69.50	2625.25	No	NaN NaN		
6044		No	Month-to-month	Yes	Bank transfer (automatic)
78.70	1495.1	No	NaN NaN		
6045		Yes	One year	Yes	Mailed check
84.80	1990.5	No	NaN NaN		
6046		Yes	One year	Yes	Credit card (automatic)
103.20	7362.9	No	NaN NaN		
6047		No	Month-to-month	Yes	Electronic check
29.60	346.45	No	NaN NaN		
6048		No	Month-to-month	Yes	Mailed check
74.40	306.6	Yes	NaN NaN		
6049		Yes	Two year	Yes	Bank transfer (automatic)
105.65	6844.5	No	NaN NaN		

[10 rows x 23 columns]

In[:]: # Information about the data:

print(df.info())

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6050 entries, 0 to 6049
Data columns (total 23 columns):
#   Column                Non-Null Count  Dtype
---  -
0   customerID            6050 non-null   object
1   gender                 6050 non-null   object
2   SeniorCitizen         6050 non-null   int64
3   Partner                6050 non-null   object
4   Dependents            6050 non-null   object
5   Tenure                 6050 non-null   int64
6   PhoneService          6050 non-null   object
7   MultipleLines         6050 non-null   object
8   InternetService       6050 non-null   object
9   OnlineSecurity        6050 non-null   object
10  OnlineBackup          6050 non-null   object
11  DeviceProtection      6050 non-null   object
12  TechSupport           6050 non-null   object
13  StreamingTV           6050 non-null   object
14  StreamingMovies       6050 non-null   object
15  Contract              6050 non-null   object
16  PaperlessBilling      6050 non-null   object
17  PaymentMethod         6050 non-null   object
18  MonthlyCharges        6050 non-null   float64
19  TotalCharges          6050 non-null   object
20  Churn                 6050 non-null   object
21  Provider              0 non-null      float64
22  Age                   0 non-null      float64
dtypes: float64(3), int64(2), object(18)
memory usage: 1.1+ MB
None
In[:]: # Descriptive statistics
df.describe()
```

Out[]:

	SeniorCitizen	Tenure	MonthlyCharges	Provider	Age
count	6050.000000	6050.000000	6050.000000	0.0	0.0
mean	0.162149	32.479174	64.836719	NaN	NaN
std	0.368618	24.517192	30.099479	NaN	NaN
min	0.000000	0.000000	18.250000	NaN	NaN
25%	0.000000	9.000000	35.750000	NaN	NaN
50%	0.000000	29.000000	70.400000	NaN	NaN
75%	0.000000	56.000000	89.900000	NaN	NaN
max	1.000000	72.000000	118.750000	NaN	NaN

In []: df.shape # Number of rows and columns

Out[]:(6050, 23)

In []: # Names of columns
df.columns

Out[]:Index(['customerID', 'gender', 'SeniorCitizen', 'Partner', 'Dependents', 'Tenure', 'PhoneService', 'MultipleLines', 'InternetService', 'OnlineSecurity', 'OnlineBackup', 'DeviceProtection', 'TechSupport', 'StreamingTV', 'StreamingMovies', 'Contract', 'PaperlessBilling', 'PaymentMethod', 'MonthlyCharges', 'TotalCharges', 'Churn', 'Provider', 'Age'], dtype='object')

In []: # Data types
df.dtypes

Out[]:customerID object
gender object
SeniorCitizen int64
Partner object
Dependents object
Tenure int64
PhoneService object
MultipleLines object
InternetService object
OnlineSecurity object
OnlineBackup object
DeviceProtection object
TechSupport object
StreamingTV object
StreamingMovies object
Contract object
PaperlessBilling object
PaymentMethod object
MonthlyCharges float64
TotalCharges object
Churn object
Provider float64
Age float64
dtype: object

In []: # Check a column

```
df['Churn'].dtype
Out[:dtype('O')
In[: df.describe()
```

Out[:	SeniorCitizen	Tenure	MonthlyCharges	Provider	Age
count	6050.000000	6050.000000	6050.000000	0.0	0.0
mean	0.162149	32.479174	64.836719	NaN	NaN
std	0.368618	24.517192	30.099479	NaN	NaN
min	0.000000	0.000000	18.250000	NaN	NaN
25%	0.000000	9.000000	35.750000	NaN	NaN
50%	0.000000	29.000000	70.400000	NaN	NaN
75%	0.000000	56.000000	89.900000	NaN	NaN
max	1.000000	72.000000	118.750000	NaN	NaN

```
In[: # Random sample of 10 rows.
df.sample(10)
```

Out[:	customerID	gender	SeniorCitizen	Partner	Dependents	Tenure	PhoneService	MultipleLines	InternetService
3333	8261-GWDBQ	Female	1	Yes	No	60	Yes	Yes	Fiber optic
414	6579-JPICP	Male	0	No	No	1	Yes	No	No
2532	5590-BYNII	Male	0	No	No	22	Yes	Yes	Fiber optic
828	9489-UTFKA	Male	1	Yes	No	8	Yes	No	DSL
4637	6005-OBZPH	Female	1	No	No	26	Yes	Yes	Fiber optic
2951	8566-YPRGL	Female	0	No	No	1	Yes	No	No
1314	2886-KEFUM	Female	0	Yes	No	63	Yes	Yes	Fiber optic
5041	6131-JLWZM	Female	0	No	No	13	Yes	No	No
5208	9526-BIHHD	Male	0	No	No	1	No	No phone service	DSL
4608	9637-EIHEQ	Female	0	No	No	1	Yes	No	DSL

10 rows × 23 columns

```
In [ ]: #Checking of unique values
df.nunique()
```

Out[]:	customerID	5986
	gender	2
	SeniorCitizen	2
	Partner	2
	Dependents	2
	Tenure	73
	PhoneService	2
	MultipleLines	3
	InternetService	3
	OnlineSecurity	3
	OnlineBackup	3
	DeviceProtection	3
	TechSupport	3
	StreamingTV	3
	StreamingMovies	3
	Contract	3
	PaperlessBilling	2
	PaymentMethod	4
	MonthlyCharges	1526
	TotalCharges	5611
	Churn	2
	Provider	0
	Age	0
	dtype:	int64

```
In [ ]: # Missing values
df.isna().sum()
```

Out[]:	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	0
	Churn	0
	Provider	6050
	Age	6050
	dtype:	int64

```
In []: #create a subset of not Senior Citizens
df_notsenior = df[(df["SeniorCitizen"] == 0)]
print(df_notsenior)
```

customerID	gender	SeniorCitizen	Partner	Dependents	Tenure	PhoneService	Multiple
Lines	InternetService	OnlineSecurity	... StreamingTV	StreamingMovies	Contract	\	
0	7590-VHVEG	Female	0	Yes	No	1	No No phone ser
vice	DSL	No	...	No	No	Month-to-month	
1	5575-GNVDE	Male	0	No	No	34	Yes
No	DSL	Yes	...	No	No	One year	
2	3668-QPYBK	Male	0	No	No	2	Yes
No	DSL	Yes	...	No	No	Month-to-month	
3	7795-CFOCW	Male	0	No	No	45	No No phone serv
ice	DSL	Yes	...	No	No	One year	
4	9237-HQITU	Female	0	No	No	2	Yes
No	Fiber optic	No	...	No	No	Month-to-month	
...
...
6044	8456-QDAVC	Male	0	No	No	19	Yes
No	Fiber optic	No	...	Yes	No	Month-to-month	
6045	6840-RESVB	Male	0	Yes	Yes	24	Yes
Yes	DSL	Yes	...	Yes	Yes	One year	
6046	2234-XADUH	Female	0	Yes	Yes	72	Yes
Yes	Fiber optic	No	...	Yes	Yes	One year	
6047	4801-JZAZL	Female	0	Yes	Yes	11	No No phone ser
vice	DSL	Yes	...	No	No	Month-to-month	
6049	3186-AJIEK	Male	0	No	No	66	Yes
No	Fiber optic	Yes	...	Yes	Yes	Two year	

	PaperlessBilling	PaymentMethod	MonthlyCharges	TotalCharges	Churn	Provider
Age						
0	Yes	Electronic check	29.85	29.85	No	NaN NaN
1	No	Mailed check	56.95	1889.5	No	NaN NaN
2	Yes	Mailed check	53.85	108.15	Yes	NaN NaN
3	No	Bank transfer (automatic)	42.30	1840.75	No	NaN
4	Yes	Electronic check	70.70	151.65	Yes	NaN
...
6044	Yes	Bank transfer (automatic)	78.70	1495.1	No	NaN
6045	Yes	Mailed check	84.80	1990.5	No	NaN NaN
6046	Yes	Credit card (automatic)	103.20	7362.9	No	NaN
6047	Yes	Electronic check	29.60	346.45	No	NaN
6049	Yes	Bank transfer (automatic)	105.65	6844.5	No	NaN

[5069 rows x 23 columns]

```
In []: # Males and Females (Not Senior Citizens)
df_notsenior_m = df_notsenior[(df_notsenior["gender"] == "Male")]
```



```
print(df_notsenior_m)

df_notsenior_f = df_notsenior[(df_notsenior["gender"] == "Female")]
print(df_notsenior_f)

customerID gender SeniorCitizen Partner Dependents Tenure PhoneService MultipleL
ines InternetService OnlineSecurity ... StreamingTV StreamingMovies Contract \
1 5575-GNVDE Male 0 No No 34 Yes
No DSL Yes ... No No One year
2 3668-QPYBK Male 0 No No 2 Yes
No DSL Yes ... No Month-to-month
3 7795-CFOCW Male 0 No No 45 No No phone servi
ce DSL Yes ... No One year
6 1452-KIOVK Male 0 No Yes 22 Yes Y
es Fiber optic No ... Yes Month-to-month
8 6388-TABGU Male 0 No Yes 62 Yes
No DSL Yes ... No One year
... ... ... ... ... ...
.. ... ... ...
6037 0550-DCXLH Male 0 No No 13 Yes
No DSL No ... Yes Month-to-month
6043 9767-FFLEM Male 0 No No 38 Yes
No Fiber optic No ... No Month-to-month
6044 8456-QDAVC Male 0 No No 19 Yes
No Fiber optic No ... Yes Month-to-month
6045 6840-RESVB Male 0 Yes Yes 24 Yes Y
es DSL Yes ... Yes One year
6049 3186-AJIEK Male 0 No No 66 Yes
No Fiber optic Yes ... Yes Two year

PaperlessBilling PaymentMethod MonthlyCharges TotalCharges Churn Provider
Age
1 No Mailed check 56.95 1889.5 No NaN N
aNaN
2 Yes Mailed check 53.85 108.15 Yes NaN N
aNaN
3 No Bank transfer (automatic) 42.30 1840.75 No NaN
NaN
6 Yes Credit card (automatic) 89.10 1949.4 No NaN
NaN
8 No Bank transfer (automatic) 56.15 3487.95 No NaN
NaN
... ... ... ... ...
..
6037 No Mailed check 73.35 931.55 No NaN N
aNaN
6043 Yes Credit card (automatic) 69.50 2625.25 No NaN
NaN
6044 Yes Bank transfer (automatic) 78.70 1495.1 No NaN
NaN
6045 Yes Mailed check 84.80 1990.5 No NaN N
aNaN
6049 Yes Bank transfer (automatic) 105.65 6844.5 No NaN
NaN

[2591 rows x 23 columns]

customerID gender SeniorCitizen Partner Dependents Tenure PhoneService Multiple
```

Lines	InternetService	OnlineSecurity	...	StreamingTV	\		
0	7590-VHVEG Female	0	Yes	No	1	No	No phone ser
vice	DSL	No	...	No	No		
4	9237-HQITU Female	0	No	No	2	Yes	
No	Fiber optic	No	...	No	No		
5	9305-CDSKC Female	0	No	No	8	Yes	
Yes	Fiber optic	No	...	Yes	Yes		
7	6713-OKOMC Female	0	No	No	10	No	No phone serv
ice	DSL	Yes	...	No	No		
13	3655-SNQYZ Female	0	Yes	Yes	69	Yes	
Yes	Fiber optic	Yes	...	Yes	Yes		
...	
...		
6036	8775-CEBBJ Female	0	No	No	9	Yes	
No	DSL	No	...	No	No		
6038	9281-CEDRU Female	0	Yes	No	68	Yes	
No	DSL	No	...	Yes	Yes		
6040	0871-OPBXW Female	0	No	No	2	Yes	
No	No No internet service	...	No internet service				
6046	2234-XADUH Female	0	Yes	Yes	72	Yes	
Yes	Fiber optic	No	...	Yes	Yes		
6047	4801-JZAZL Female	0	Yes	Yes	11	No	No phone ser
vice	DSL	Yes	...	No	No		

	StreamingMovies		Contract		PaperlessBilling		PaymentMethod		Monthly
Charges	TotalCharges	Churn	Provider	Age					
0		No	Month-to-month		Yes		Electronic	check	
29.85	29.85	No	NaN	NaN					
4		No	Month-to-month		Yes		Electronic	check	
70.70	151.65	Yes	NaN	NaN					
5		Yes	Month-to-month		Yes		Electronic	check	
99.65	820.5	Yes	NaN	NaN					
7		No	Month-to-month		No		Mailed	check	
29.75	301.9	No	NaN	NaN					
13		Yes	Two	year	No		Credit	card (automatic)	
113.25	7895.15	No	NaN	NaN					
...		
...					
6036		No	Month-to-month		Yes		Bank	transfer (automatic)	
44.20	403.35	Yes	NaN	NaN					
6038		No	Two	year	No		Bank	transfer (automatic)	
64.10	4326.25	No	NaN	NaN					
6040	No internet service		Month-to-month		Yes		Mailed	check	
20.05	39.25	No	NaN	NaN					
6046		Yes	One	year	Yes		Credit	card (automatic)	
103.20	7362.9	No	NaN	NaN					
6047		No	Month-to-month		Yes		Electronic	check	
29.60	346.45	No	NaN	NaN					

[2478 rows x 23 columns]

```
In []: #Males_Not_Senior_Phone
df_notsenior_m.groupby('PhoneService').count()
df_males_notsenior_phone = df_notsenior_m.groupby('PhoneService').count()
print(df_males_notsenior_phone)

customerID gender SeniorCitizen Partner Dependents Tenure MultipleLines
InternetService OnlineSecurity OnlineBackup ... StreamingTV StreamingMovies \
```

PhoneService

...

No	255	255	255	255	255	255	255
255	255	255	...	255	255	255	
Yes	2336	2336	2336	2336	2336	2336	2336
2336	2336	2336	...	2336	2336	2336	

Contract PaperlessBilling PaymentMethod MonthlyCharges TotalCharges Churn
Provider Age
PhoneService

No	255	255	255	255	255	255
0 0						
Yes	2336	2336	2336	2336	2336	2336
0 0						

[2 rows x 22 columns]

```
In[: df_males_notsenior_phone.to_excel('males_not_senior_phones.xlsx', index=False, sheet_name='Sheet1')
```

```
In[: #Males_Not_Senior_Internet
df_notsenior_m.groupby('InternetService').count()
df_males_notsenior_internet = df_notsenior_m.groupby('InternetService').count()
print(df_males_notsenior_internet)
```

customerID gender SeniorCitizen Partner Dependents Tenure PhoneService
MultipleLines OnlineSecurity OnlineBackup ... StreamingTV StreamingMovies \
InternetService

DSL	959	959	959	959	959	959	959
959	959	959	...	959	959	959	
Fiber optic	988	988	988	988	988	988	988
988	988	988	...	988	988	988	
No	644	644	644	644	644	644	644
644	644	644	...	644	644	644	

Contract PaperlessBilling PaymentMethod MonthlyCharges TotalCharges Churn
Provider Age
InternetService

DSL	959	959	959	959	959	95
9 0 0						
Fiber optic	988	988	988	988	988	98
8 0 0						
No	644	644	644	644	644	644
0 0						

[3 rows x 22 columns]

```
In[: df_males_notsenior_internet.to_excel('males_not_senior_internet.xlsx', index=False, sheet_name='Sheet1')
```

```
In[: #Males_Not_Senior_DeviceProtection
df_notsenior_m.groupby('DeviceProtection').count()
df_males_notsenior_device = df_notsenior_m.groupby('DeviceProtection').count()
print(df_males_notsenior_device)
```

customerID gender SeniorCitizen Partner Dependents Tenure PhoneSer

DeviceProtection		MultipleLines		InternetService		OnlineSecurity		...		StreamingTV		\	
....													
No				1089		1089			1089		1089		1089
1089		1089				1089		1089	...		1089		
No internet service				644		644			644			644	644
644		644				644		644	...		644		
Yes				858		858			858		858		858
858		858				858		858	...		858		
				StreamingMovies		Contract		PaperlessBilling		PaymentMethod		MonthlyCharge	
es	TotalCharges	Churn		Provider		Age							
DeviceProtection													
No						1089		1089			1089		1089
1089	1089		0	0									
No internet service						644		644			644		64
4		644	644			0	0						
Yes						858		858			858		858
858	858		0	0									

[3 rows x 22 columns]

```
In [ ]: df_males_notsenior_device.to_excel('males_not_senior_device.xlsx', index=False,
      sheet_name='Sheet1')

In [ ]: #Males_Not_Senior_StreamingTV
      df_notsenior_m.groupby('StreamingTV').count()
      df_males_notsenior_streaming = df_notsenior_m.groupby('StreamingTV').count()
      print(df_males_notsenior_streaming)
```

		customerID	gender	SeniorCitizen	Partner	Dependents	Tenure	PhoneSer
vice	MultipleLines	InternetService	OnlineSecurity	...	TechSupport	\		
StreamingTV								
...								
No		1014	1014	1014	1014	1014	1014	1
014	1014		1014	1014	...	1014		
No internet service		644	644	644	644	644	644	
644	644		644	644	...	644		
Yes		933	933	933	933	933	933	
933	933		933	933	...	933		
		StreamingMovies	Contract	PaperlessBilling	PaymentMethod	MonthlyCharg		
es	TotalCharges	Churn	Provider	Age				
StreamingTV								
No			1014	1014	1014	1014		1014
1014	1014	0	0					
No internet service			644	644	644	644		64
4	644	644	0	0				
Yes			933	933	933	933		933
933	933	0	0					

[3 rows x 22 columns]

```
In [ ]: df_males_notsenior_streaming.to_excel('males_not_senior_streaming.xlsx', index=False,
      sheet_name='Sheet1')

In [ ]: #Males_Not_Senior_PaperlessBilling
```

```
df_notsenior_m.groupby('PaperlessBilling').count()
df_males_notsenior_paperless = df_notsenior_m.groupby('PaperlessBilling').count()
print(df_males_notsenior_paperless)

customerID  gender  SeniorCitizen  Partner  Dependents  Tenure  PhoneService
MultipleLines  InternetService  OnlineSecurity  ...  TechSupport  StreamingTV  \
PaperlessBilling
...
No          1167    1167          1167    1167          1167    1167          1167
1167        1167          1167  ...          1167          1167
Yes          1424    1424          1424    1424          1424    1424          1424
1424        1424          1424  ...          1424          1424

StreamingMovies  Contract  PaymentMethod  MonthlyCharges  TotalCharges  Churn
Provider  Age
PaperlessBilling

No          1167    1167          1167          1167          1167    1167
0          0
Yes          1424    1424          1424          1424          1424    142
4          0      0

[2 rows x 22 columns]
In[: df_males_notsenior_paperless.to_excel('males_not_senior_paperless.xlsx', index=False,
sheet_name='Sheet1')

In[: #Females Not Senior Phone
df_notsenior_f.groupby('PhoneService').count()
df_females_notsenior_phone = df_notsenior_f.groupby('PhoneService').count()
print(df_females_notsenior_phone)

customerID  gender  SeniorCitizen  Partner  Dependents  Tenure  MultipleLines
InternetService  OnlineSecurity  OnlineBackup  ...  StreamingTV  StreamingMovies  \
PhoneService
...
No          254    254          254    254          254    254          254
254        254          254  ...          254          254
Yes          2224    2224          2224    2224          2224    2224          2224
2224        2224          2224  ...          2224          2224

Contract  PaperlessBilling  PaymentMethod  MonthlyCharges  TotalCharges  Churn
Provider  Age
PhoneService

No          254          254          254          254          254    254
0          0
Yes          2224          2224          2224          2224          2224    2224
0          0

[2 rows x 22 columns]
In[: df_females_notsenior_phone.to_excel('females_not_senior_phone.xlsx', index=False,
sheet_name='Sheet1')

In[: #Females Not Senior Internet
df_notsenior_f.groupby('InternetService').count()
df_females_notsenior_internet = df_notsenior_f.groupby('InternetService').count()
print(df_females_notsenior_internet)
```

		customerID	gender	SeniorCitizen	Partner	Dependents	Tenure	PhoneService
e	MultipleLines	OnlineSecurity	OnlineBackup	...	StreamingTV	StreamingMovies	\	InternetService
...								
DSL		909	909	909	909	909	909	909
909	909		909	...	909	909		
Fiber optic		954	954	954	954	954	954	954
954	954		954	...	954	954		
No		615	615	615	615	615	615	615
615	615		615	...	615	615		

		Contract	PaperlessBilling	PaymentMethod	MonthlyCharges	TotalCharges	Ch
urn	Provider	Age					
InternetService							
DSL		909	909	909	909	909	90
9	0	0					
Fiber optic		954	954	954	954	954	95
4	0	0					
No		615	615	615	615	615	615
0	0						

```
[3 rows x 22 columns]
In[: df_females_notsenior_internet.to_excel('females_not_senior_internet.xlsx',
      index=False, sheet_name='Sheet1')

In[: #Females Not Senior Device Protection
      df_notsenior_f.groupby('DeviceProtection').count()
      df_females_notsenior_device = df_notsenior_f.groupby('DeviceProtection').count()
      print(df_females_notsenior_device)
```

		customerID	gender	SeniorCitizen	Partner	Dependents	Tenure	PhoneSer
vice	MultipleLines	InternetService	OnlineSecurity	...	StreamingTV	\	DeviceProtection	
...								
No		1046	1046	1046	1046	1046	1046	1
046	1046		1046	1046	...	1046		
No internet service		615	615	615	615	615	615	
615	615		615	615	...	615		
Yes		817	817	817	817	817	817	
817	817		817	817	...	817		

		StreamingMovies		Contract	PaperlessBilling	PaymentMethod	MonthlyCharg
es	TotalCharges	Churn	Provider	Age			
DeviceProtection							
No			1046	1046	1046	1046	1046
1046	1046	0	0				
No internet service			615	615	615	615	61
5	615	615	0	0			
Yes			817	817	817	817	817
817	817	0	0				

```
[3 rows x 22 columns]
In[: df_females_notsenior_device.to_excel('females_not_senior_device.xlsx', index=False,
      sheet_name='Sheet1')
```

```
In[]: # Females Not Senior Streaming TV
df_notsenior_f.groupby('StreamingTV').count()
df_females_notsenior_streaming = df_notsenior_f.groupby('StreamingTV').count()
print(df_females_notsenior_streaming)
```

	customerID	gender	SeniorCitizen	Partner	Dependents	Tenure	PhoneSer
vice	MultipleLines	InternetService	OnlineSecurity	...	TechSupport	\	
StreamingTV							
...							
No		956	956		956	956	956
956	956		956	956	...	956	
No internet service		615	615		615	615	615
615	615		615	615	...	615	
Yes		907	907		907	907	907
907	907		907	907	...	907	

```
StreamingMovies Contract PaperlessBilling PaymentMethod MonthlyCharg
es TotalCharges Churn Provider Age
StreamingTV
```

No				956	956		956		956
956	956	0	0						
No internet service				615	615		615		61
5	615	615	0	0					
Yes				907	907		907		907
907	907	0	0						

[3 rows x 22 columns]

```
In[]: df_females_notsenior_streaming.to_excel('females_not_senior_streaming.xlsx',
index=False, sheet_name='Sheet1')
```

```
In[]: #Females Not Senior Paperless
df_notsenior_f.groupby('PaperlessBilling').count()
df_females_notsenior_paperless = df_notsenior_f.groupby('PaperlessBilling').count()
print(df_females_notsenior_paperless)
```

	customerID	gender	SeniorCitizen	Partner	Dependents	Tenure	PhoneServi
ce	MultipleLines	InternetService	OnlineSecurity	...	TechSupport	StreamingTV	\
PaperlessBilling							
...							
No		1077	1077		1077	1077	1077
1077	1077		1077	...	1077	1077	
Yes		1401	1401		1401	1401	1401
1401	1401		1401	...	1401	1401	

```
StreamingMovies Contract PaymentMethod MonthlyCharges TotalCharges Ch
urn Provider Age
PaperlessBilling
```

No				1077	1077		1077		1077
0	0								
Yes				1401	1401		1401		140
1	0	0							

[2 rows x 22 columns]

```
In[]: df_females_notsenior_paperless.to_excel('females_not_senior_paperless.xlsx',
index=False, sheet_name='Sheet1')
```

```
In []: #my own analysis
      # average monthly charges
      #create a subset of not Senior Citizens
      df_senior = df[(df["SeniorCitizen"] == 1)]
      print(df_senior)
```

	customerID	gender	SeniorCitizen	Partner	Dependents	Tenure	PhoneService	Multiple
Lines	InternetService	OnlineSecurity	...	StreamingTV	StreamingMovies	Contract	\	
16	8779-QRDMV	Male	1	No	No	1	No	No phone serv
ice	DSL	No	...	No	Yes	Month-to-month		
25	3841-NFECX	Female	1	Yes	No	71	Yes	
Yes	Fiber optic	Yes	...	No	No	Two year		
28	3413-BMNZE	Male	1	No	No	1	Yes	
No	DSL	No	...	No	No	Month-to-month		
43	6575-SUVOI	Female	1	Yes	No	25	Yes	
Yes	DSL	Yes	...	Yes	No	Month-to-month		
44	7495-OKKFY	Female	1	Yes	No	8	Yes	
Yes	Fiber optic	No	...	No	No	Month-to-month		
...	
...	
6027	0666-UXTJO	Male	1	Yes	No	34	Yes	
No	Fiber optic	No	...	Yes	No	Month-to-month		
6039	2235-DWLJU	Female	1	No	No	6	No	No phone ser
vice	DSL	No	...	Yes	Yes	Month-to-month		
6041	3605-JISKB	Male	1	Yes	No	55	Yes	
Yes	DSL	Yes	...	No	No	One year		
6042	6894-LFHLY	Male	1	No	No	1	Yes	
Yes	Fiber optic	No	...	No	No	Month-to-month		
6048	8361-LTMKD	Male	1	Yes	No	4	Yes	
Yes	Fiber optic	No	...	No	No	Month-to-month		

	PaperlessBilling	PaymentMethod	MonthlyCharges	TotalCharges	Churn	Provider
Age						
16	Yes	Electronic check	39.65	39.65	Yes	NaN
NaN						
25	Yes	Credit card (automatic)	96.35	6766.95	No	NaN
NaN						
28	No	Bank transfer (automatic)	45.25	45.25	No	NaN
NaN						
43	Yes	Credit card (automatic)	69.50	1752.65	No	NaN
NaN						
44	Yes	Credit card (automatic)	80.65	633.3	Yes	NaN
NaN						
...
..						
6027	Yes	Credit card (automatic)	85.20	2874.45	No	NaN
NaN						
6039	Yes	Electronic check	44.40	263.05	No	NaN
NaN						
6041	No	Credit card (automatic)	60.00	3316.1	No	NaN
NaN						
6042	Yes	Electronic check	75.75	75.75	Yes	NaN
NaN						
6048	Yes	Mailed check	74.40	306.6	Yes	NaN
NaN						

[981 rows x 23 columns]


```
In []: # Males and Females (Senior Citizens)
df_senior_m = df_senior[(df_senior["gender"] == "Male")]
print(df_senior_m)

df_senior_f = df_senior[(df_senior["gender"] == "Female")]
print(df_senior_f)
```

	customerID	gender	SeniorCitizen	Partner	Dependents	Tenure	PhoneService	MultipleLines
16	8779-QRDMV	Male	1	No	No	1	No	No phone service
28	3413-BMNZE	Male	1	No	No	1	Yes	DSL
46	1658-BYGOY	Male	1	No	No	18	Yes	DSL
48	5067-XJQFU	Male	1	Yes	Yes	66	Yes	Fiber optic
61	1891-QRQSA	Male	1	Yes	Yes	64	Yes	Fiber optic
...
6019	2274-XUATA	Male	1	Yes	No	72	No	DSL
6027	0666-UXTJO	Male	1	Yes	No	34	Yes	DSL
6041	3605-JISKB	Male	1	Yes	No	55	Yes	Fiber optic
6042	6894-LFHLY	Male	1	No	No	1	Yes	DSL
6048	8361-LTMKD	Male	1	Yes	No	4	Yes	Fiber optic

	PaperlessBilling	PaymentMethod	MonthlyCharges	TotalCharges	Churn	Provider
16	Yes	Electronic check	39.65	39.65	Yes	NaN
28	No	Bank transfer (automatic)	45.25	45.25	No	NaN
46	Yes	Electronic check	95.45	1752.55	Yes	NaN
48	Yes	Electronic check	108.45	7076.35	No	NaN
61	Yes	Bank transfer (automatic)	111.60	7099	No	NaN
...
6019	Yes	Bank transfer (automatic)	63.10	4685.55	No	NaN
6027	Yes	Credit card (automatic)	85.20	2874.45	No	NaN
6041	No	Credit card (automatic)	60.00	3316.1	No	NaN
6042	Yes	Electronic check	75.75	75.75	Yes	NaN
6048	Yes	Mailed check	74.40	306.6	Yes	NaN

[490 rows x 23 columns]

	customerID	gender	SeniorCitizen	Partner	Dependents	Tenure	PhoneService	Multiple
Lines	InternetService	OnlineSecurity	...	StreamingTV	StreamingMovies	Contract	\	
25	3841-NFECX	Female	1	Yes	No	71	Yes	
Yes	Fiber optic	Yes	...	No	No	Two year		
43	6575-SUVOI	Female	1	Yes	No	25	Yes	
Yes	DSL	Yes	...	Yes	No	Month-to-month		
44	7495-OKFY	Female	1	Yes	No	8	Yes	
Yes	Fiber optic	No	...	No	No	Month-to-month		
45	4667-QONEA	Female	1	Yes	Yes	60	Yes	
No	DSL	Yes	...	No	Yes	One year		
63	2673-CXQEU	Female	1	No	No	56	Yes	
Yes	Fiber optic	Yes	...	Yes	Yes	One year		
...	
...	
5987	9972-VAFJJ	Female	1	Yes	No	53	Yes	
No	Fiber optic	No	...	Yes	No	One year		
6001	5981-ZVXOT	Female	1	No	No	35	Yes	
Yes	Fiber optic	Yes	...	Yes	Yes	One year		
6020	1980-KXVPM	Female	1	No	No	3	Yes	
Yes	Fiber optic	No	...	No	No	Month-to-month		
6022	0723-DRCLG	Female	1	Yes	No	1	Yes	
Yes	Fiber optic	No	...	No	No	Month-to-month		
6039	2235-DWLJU	Female	1	No	No	6	No	No phone ser
vice	DSL	No	...	Yes	Yes	Month-to-month		

	PaperlessBilling	PaymentMethod	MonthlyCharges	TotalCharges	Churn	Provider	A
ge							
25	Yes	Credit card (automatic)	96.35	6766.95	No	NaN	Na
N							
43	Yes	Credit card (automatic)	69.50	1752.65	No	NaN	Na
N							
44	Yes	Credit card (automatic)	80.65	633.3	Yes	NaN	Na
N							
45	Yes	Credit card (automatic)	74.85	4456.35	No	NaN	Na
N							
63	No	Electronic check	110.50	6139.5	No	NaN	NaN
...
5987	Yes	Electronic check	94.00	4871.45	No	NaN	Na
N							
6001	Yes	Electronic check	108.35	3726.15	No	NaN	Na
N							
6020	Yes	Credit card (automatic)	75.05	256.25	Yes	NaN	Na
N							
6022	Yes	Electronic check	74.45	74.45	Yes	NaN	Na
N							
6039	Yes	Electronic check	44.40	263.05	No	NaN	Na
N							

[491 rows x 23 columns]

```
In [ ]: #Males_Senior_Phone
df_senior_m.groupby('PhoneService').count()
df_males_senior_phone = df_senior_m.groupby('PhoneService').count()
print(df_males_senior_phone)
```

	customerID	gender	SeniorCitizen	Partner	Dependents	Tenure	MultipleLines
InternetService	OnlineSecurity	OnlineBackup	...	StreamingTV	StreamingMovies	\	
PhoneService							
...							
No	52	52	52	52	52	52	52
52	52	52	...	52	52		
Yes	438	438	438	438	438	438	438
438	438	438	...	438	438		

	Contract	PaperlessBilling	PaymentMethod	MonthlyCharges	TotalCharges	Churn
Provider	Age					
PhoneService						
No	52	52	52	52	52	52
0	0					
Yes	438	438	438	438	438	438
0	0					

[2 rows x 22 columns]

```
In [ ]: df_males_senior_phone.to_excel('males_senior_phone.xlsx', index=False,
    sheet_name='Sheet1')

In [ ]: #Males_Senior_Internet
df_senior_m.groupby('InternetService').count()
df_males_senior_internet = df_senior_m.groupby('InternetService').count()
print(df_males_senior_internet)
```

	customerID	gender	SeniorCitizen	Partner	Dependents	Tenure	PhoneService
e	MultipleLines	OnlineSecurity	OnlineBackup	...	StreamingTV	StreamingMovies	\
InternetService							
...							
DSL	116	116	116	116	116	116	116
116	116	116	...	116	116		
Fiber optic	348	348	348	348	348	348	348
348	348	348	...	348	348		
No	26	26	26	26	26	26	26
26	26	26	...	26	26		

	Contract	PaperlessBilling	PaymentMethod	MonthlyCharges	TotalCharges	Churn
urn	Provider	Age				
InternetService						
DSL	116	116	116	116	116	11
6	0	0				
Fiber optic	348	348	348	348	348	34
8	0	0				
No	26	26	26	26	26	26
0	0					

[3 rows x 22 columns]

```
In [ ]: df_males_senior_internet.to_excel('males_senior_internet.xlsx', index=False,
    sheet_name='Sheet1')

In [ ]: #Males_Device_Protection_Phone
df_senior_m.groupby('DeviceProtection').count()
df_males_senior_device = df_senior_m.groupby('DeviceProtection').count()
print(df_males_senior_device)
```

	customerID	gender	SeniorCitizen	Partner	Dependents	Tenure	PhoneSer
vice MultipleLines	InternetService	OnlineSecurity	...	StreamingTV	\		
DeviceProtection							
...							
No	270	270	270	270	270	270	270
270	270	270	270	...	270		
No internet service	26	26	26	26	26	26	26
26	26	26	26	...	26		
Yes	194	194	194	194	194	194	194
194	194	194	194	...	194		

	StreamingMovies	Contract	PaperlessBilling	PaymentMethod	MonthlyCharg
es TotalCharges	Churn	Provider	Age		
DeviceProtection					
No	270	270	270	270	270
270	270	0	0		
No internet service	26	26	26	26	26
26	26	0	0		
Yes	194	194	194	194	194
194	194	0	0		

[3 rows x 22 columns]

```
In[: df_males_senior_device.to_excel('males_senior_device.xlsx', index=False, sheet_name='Sheet1')
```

```
In[: #Males_Senior_Streaming TV
df_senior_m.groupby('StreamingTV').count()
df_males_senior_streaming = df_senior_m.groupby('StreamingTV').count()
print(df_males_senior_streaming)
```

	customerID	gender	SeniorCitizen	Partner	Dependents	Tenure	PhoneSer
vice MultipleLines	InternetService	OnlineSecurity	...	TechSupport	\		
StreamingTV							
...							
No	225	225	225	225	225	225	225
225	225	225	225	...	225		
No internet service	26	26	26	26	26	26	26
26	26	26	26	...	26		
Yes	239	239	239	239	239	239	239
239	239	239	239	...	239		

	StreamingMovies	Contract	PaperlessBilling	PaymentMethod	MonthlyCharg
es TotalCharges	Churn	Provider	Age		
StreamingTV					
No	225	225	225	225	225
225	225	0	0		
No internet service	26	26	26	26	26
26	26	0	0		
Yes	239	239	239	239	239
239	239	0	0		

[3 rows x 22 columns]

```
In[: df_males_senior_streaming.to_excel('males_senior_streaming.xlsx', index=False, sheet_name='Sheet1')
```

```
In []: #Males_Senior_Paperless
df_senior_m.groupby('PaperlessBilling').count()
df_males_senior_paperless = df_senior_m.groupby('PaperlessBilling').count()
print(df_males_senior_paperless)
```

	customerID	gender	SeniorCitizen	Partner	Dependents	Tenure	PhoneService
PaperlessBilling	MultipleLines	InternetService	OnlineSecurity	...	TechSupport	StreamingTV	\
...							
No	117	117	117	117	117	117	117
117	117	117	...	117	117		
Yes	373	373	373	373	373	373	373
373	373	373	...	373	373		

	StreamingMovies	Contract	PaymentMethod	MonthlyCharges	TotalCharges	Churn
Provider	Age					
PaperlessBilling						
No	117	117	117	117	117	117
0	0					
Yes	373	373	373	373	373	37
3	0	0				

[2 rows x 22 columns]

```
In []: df_males_senior_paperless.to_excel('males_senior_paperless.xlsx', index=False,
sheet_name='Sheet1')
```

```
In []: #Females_Senior_Phone
df_senior_f.groupby('PhoneService').count()
df_females_senior_phone = df_senior_f.groupby('PhoneService').count()
print(df_females_senior_phone)
```

	customerID	gender	SeniorCitizen	Partner	Dependents	Tenure	MultipleLines
InternetService	OnlineSecurity	OnlineBackup	...	StreamingTV	StreamingMovies	\	
PhoneService							
...							
No	37	37	37	37	37	37	37
37	37	37	...	37	37		
Yes	454	454	454	454	454	454	454
454	454	454	...	454	454		

	Contract	PaperlessBilling	PaymentMethod	MonthlyCharges	TotalCharges	Churn
Provider	Age					
PhoneService						
No	37	37	37	37	37	37
0	0					
Yes	454	454	454	454	454	454
0	0					

[2 rows x 22 columns]

```
In []: df_females_senior_phone.to_excel('females_senior_phone.xlsx', index=False,
sheet_name='Sheet1')
```

```
In []: #Females_Senior_Internet
df_senior_f.groupby('InternetService').count()
df_females_senior_internet = df_senior_f.groupby('InternetService').count()
```

```
print(df_females_senior_internet)

customerID gender SeniorCitizen Partner Dependents Tenure PhoneService
MultipleLines OnlineSecurity OnlineBackup ... StreamingTV StreamingMovies \
InternetService
...
DSL 108 108 108 108 108 108 108
108 108 108 ... 108 108
Fiber optic 368 368 368 368 368 368 368
368 368 368 ... 368 368
No 15 15 15 15 15 15 15
15 15 15 ... 15 15

Contract PaperlessBilling PaymentMethod MonthlyCharges TotalCharges Churn
Provider Age
InternetService

DSL 108 108 108 108 108 10
8 0 0
Fiber optic 368 368 368 368 368 36
8 0 0
No 15 15 15 15 15 15
0 0

[3 rows x 22 columns]
In[: df_females_senior_internet.to_excel('females_senior_internet.xlsx', index=False,
sheet_name='Sheet1')

In[: #Females_Senior_Device_Protection
df_senior_f.groupby('DeviceProtection').count()
df_females_senior_device = df_senior_f.groupby('DeviceProtection').count()
print(df_females_senior_device)

customerID gender SeniorCitizen Partner Dependents Tenure PhoneSer
vice MultipleLines InternetService OnlineSecurity ... StreamingTV \
DeviceProtection
...
No 265 265 265 265 265 265
265 265 265 ... 265
No internet service 15 15 15 15 15 15
15 15 15 ... 15
Yes 211 211 211 211 211 211
211 211 211 ... 211

StreamingMovies Contract PaperlessBilling PaymentMethod MonthlyCharg
es TotalCharges Churn Provider Age
DeviceProtection

No 265 265 265 265 265
265 265 0 0
No internet service 15 15 15 15 15
15 15 0 0
Yes 211 211 211 211 211
211 211 0 0

[3 rows x 22 columns]
In[: df_females_senior_device.to_excel('females_senior_device.xlsx', index=False,
```

```
sheet_name='Sheet1')

In [ ]: #Females_Senior_Streaming_TV
df_senior_f.groupby('StreamingTV').count()
df_females_senior_streaming = df_senior_f.groupby('StreamingTV').count()
print(df_females_senior_streaming)

customerID gender SeniorCitizen Partner Dependents Tenure PhoneSer
vice MultipleLines InternetService OnlineSecurity ... TechSupport \
StreamingTV
...
No 223 223 223 223 223 223
223 223 223 223 ... 223
No internet service 15 15 15 15 15
15 15 15 15 ... 15
Yes 253 253 253 253 253 253
253 253 253 253 ... 253

StreamingMovies Contract PaperlessBilling PaymentMethod MonthlyCharg
es TotalCharges Churn Provider Age
StreamingTV

No 223 223 223 223 223
223 223 0 0 15 15 15 15
No internet service 15 15 15 15 15
15 15 0 0 253 253 253 253
Yes 253 253 253 253 253
253 253 0 0

[3 rows x 22 columns]

In [ ]: df_females_senior_streaming.to_excel('females_senior_streaming.xlsx', index=False,
sheet_name='Sheet1')

In [ ]: #Females_Senior_Paperless_Billing
df_senior_f.groupby('PaperlessBilling').count()
df_females_senior_paperless = df_senior_f.groupby('PaperlessBilling').count()
print(df_females_senior_paperless)

customerID gender SeniorCitizen Partner Dependents Tenure PhoneServi
ce MultipleLines InternetService OnlineSecurity ... TechSupport StreamingTV \
PaperlessBilling
...
No 116 116 116 116 116 116
116 116 116 ... 116 116
Yes 375 375 375 375 375 375
375 375 375 ... 375

StreamingMovies Contract PaymentMethod MonthlyCharges TotalCharges Ch
urn Provider Age
PaperlessBilling

No 116 116 116 116 116
0 0
Yes 375 375 375 375 375
5 0 0

[2 rows x 22 columns]

In [ ]:
```

```
df_females_senior_paperless.to_excel('females_senior_paperless.xlsx', index=False,
sheet_name='Sheet1')

In [ ]: #Payment method
df.groupby('PaymentMethod').count()

Out[ ]:
```

	customerID	gender	SeniorCitizen	Partner	Dependents	Tenure	PhoneService	MultipleLines	In
PaymentMethod									
Bank transfer (automatic)	1321	1321	1321	1321	1321	1321	1321	1321	13
Credit card (automatic)	1315	1315	1315	1315	1315	1315	1315	1315	13
Electronic check	2033	2033	2033	2033	2033	2033	2033	2033	20
Mailed check	1381	1381	1381	1381	1381	1381	1381	1381	13

4 rows × 22 columns

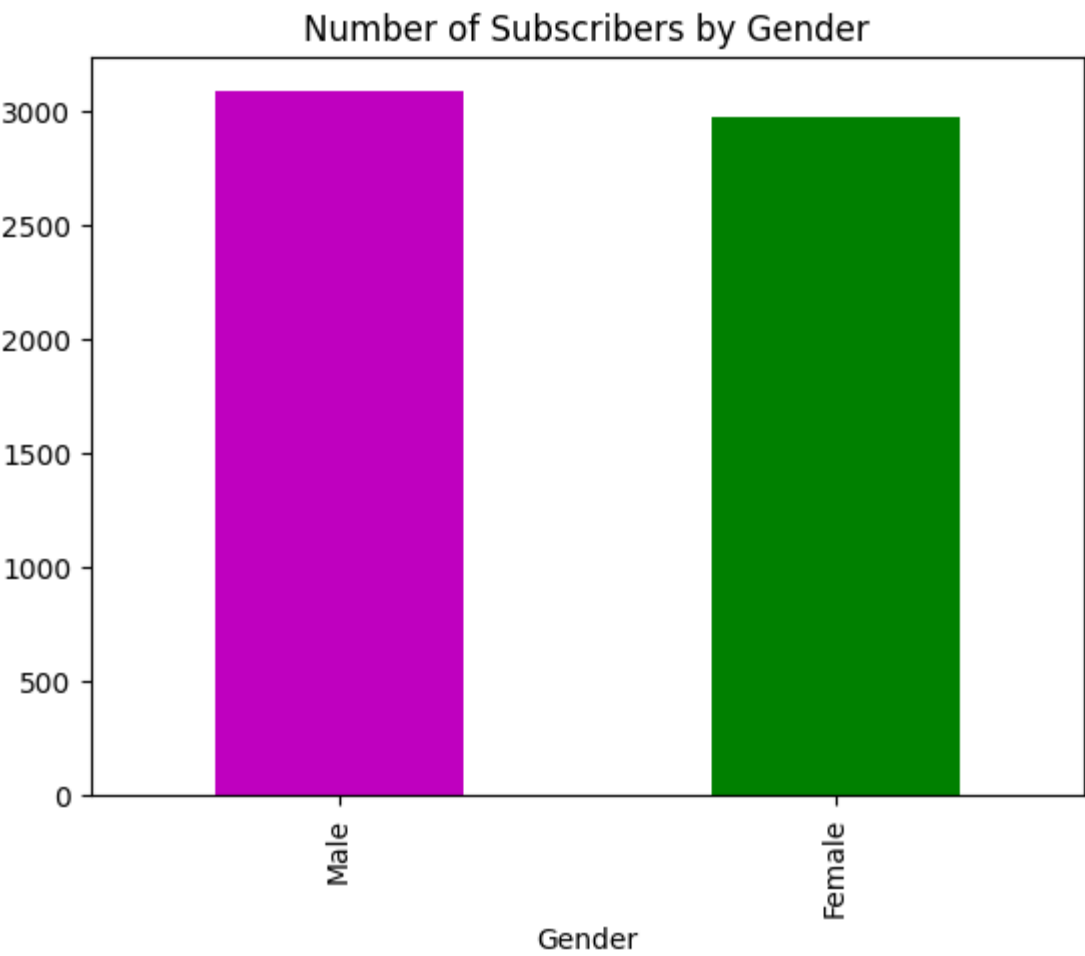
```
In [ ]: #Visualisation
import matplotlib.pyplot as plt

In [ ]: gender = df["gender"].value_counts()

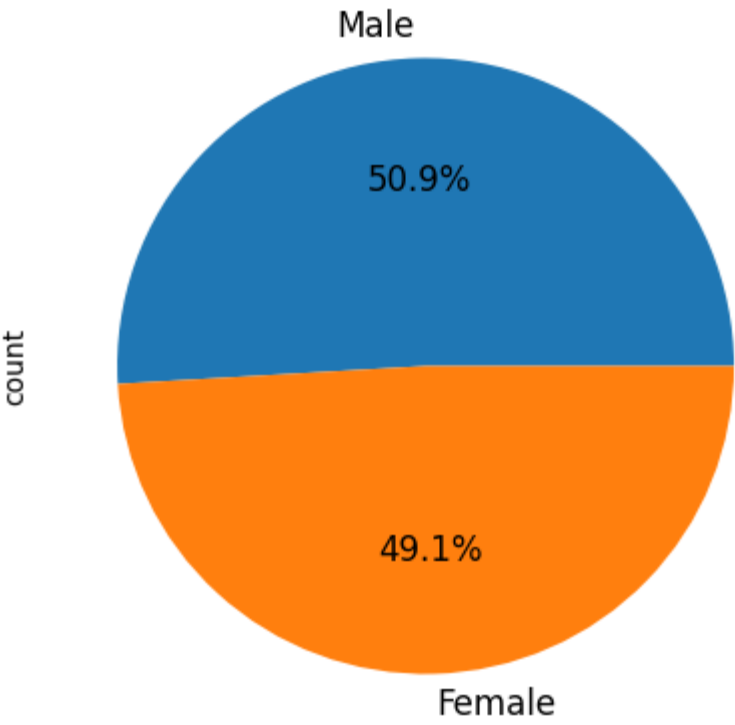
print(gender)

gender
Male      3081
Female    2969
Name: count, dtype: int64

In [ ]: #visualisation - gender
import matplotlib.pyplot as plt
gender.plot(kind = 'bar', color=['m', 'g'])
plt.title('Number of Subscribers by Gender')
plt.xlabel('Gender')
plt.show()
```

```
In[]: # Pie chart for gender distribution
axis = (df['gender'].value_counts() * 100.0 / len(df)).plot.pie(autopct='%.1f%%',
labels = ['Male', 'Female'] ,figsize =(5,5), fontsize = 12)
```



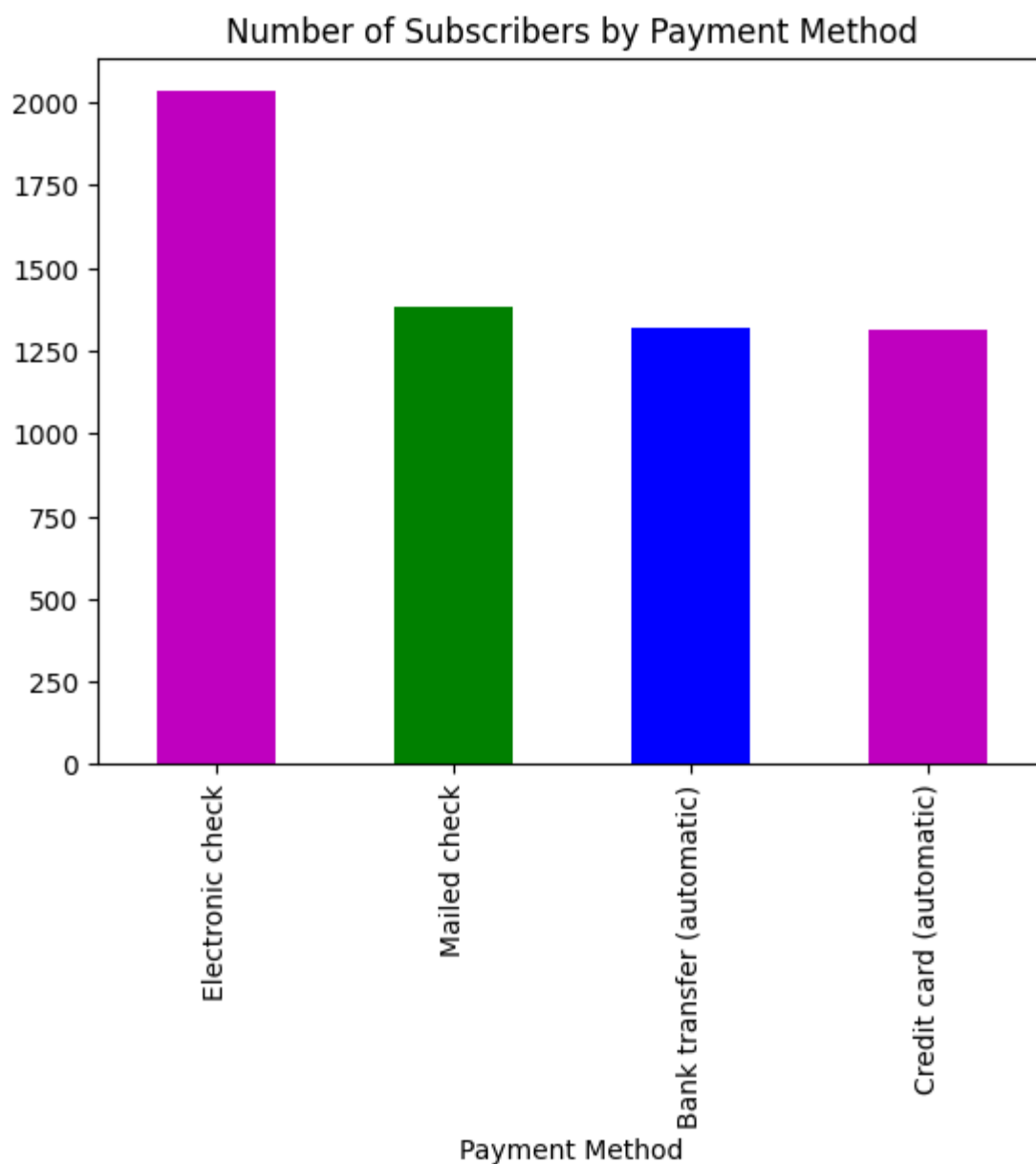
```
In [ ]: paymethod = df["PaymentMethod"].value_counts()
```

```
print(paymethod)
```

```
PaymentMethod
Electronic check      2033
Mailed check          1381
Bank transfer (automatic) 1321
Credit card (automatic) 1315
Name: count, dtype: int64
```

```
In [ ]: # visualisation payment method
```

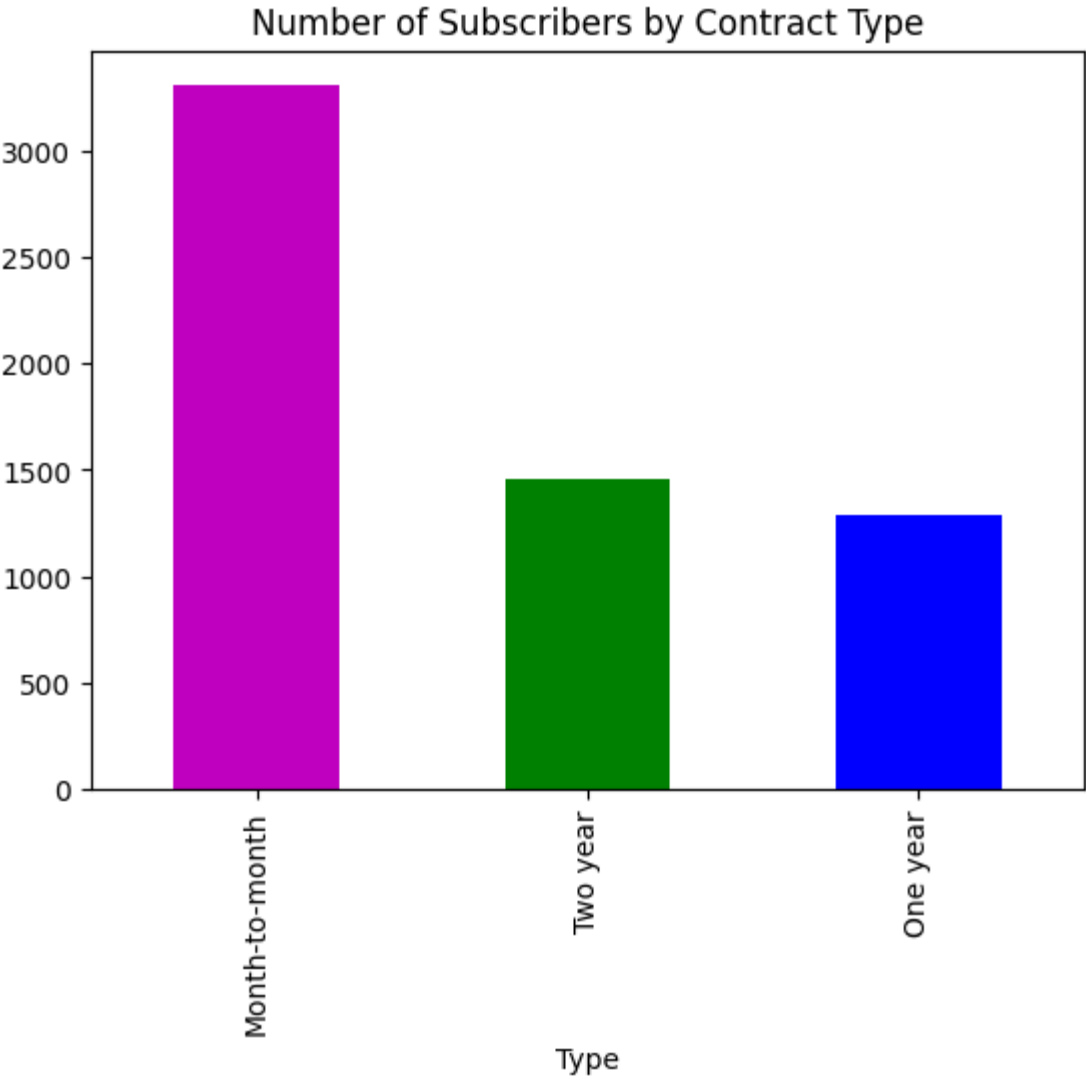
```
import matplotlib.pyplot as plt
paymethod.plot(kind = 'bar', color=['m', 'g', 'b'])
plt.title('Number of Subscribers by Payment Method')
plt.xlabel('Payment Method')
plt.show()
```



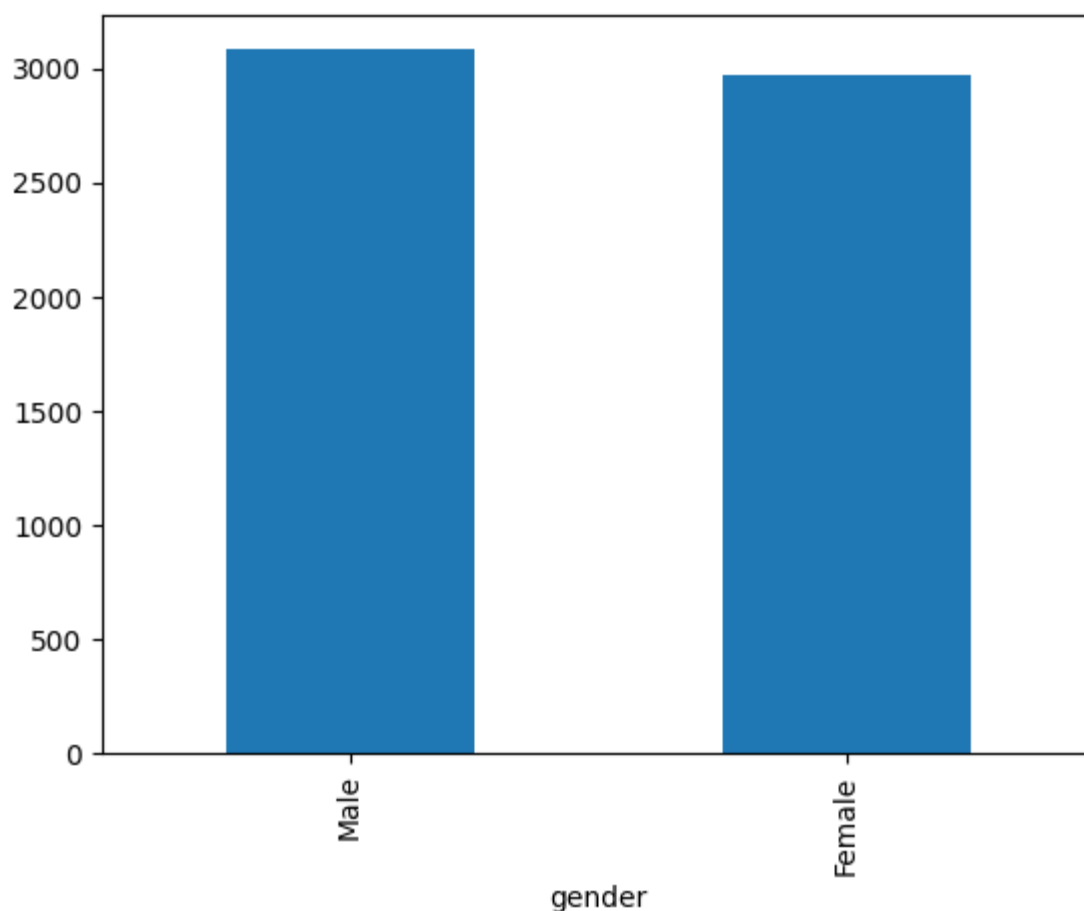
```
In [ ]: #contract
contract = df["Contract"].value_counts()
print(contract)
```

```
Contract
Month-to-month      3305
Two year            1457
One year            1288
Name: count, dtype: int64
```

```
In[:]: # visualisation payment method
import matplotlib.pyplot as plt
contract.plot(kind = 'bar', color=['m', 'g', 'b'])
plt.title('Number of Subscribers by Contract Type')
plt.xlabel('Type')
plt.show()
```



```
In[:]: #Visualisation
import matplotlib.pyplot as plt; ax = plt.subplots(1,1)
import matplotlib.pyplot as plt
gender.plot(kind = 'bar')
plt.show()
title = 'Gender Distribution'
```



```
In[: #customer type
      customer = df["SeniorCitizen"].value_counts()
      print(customer)
```

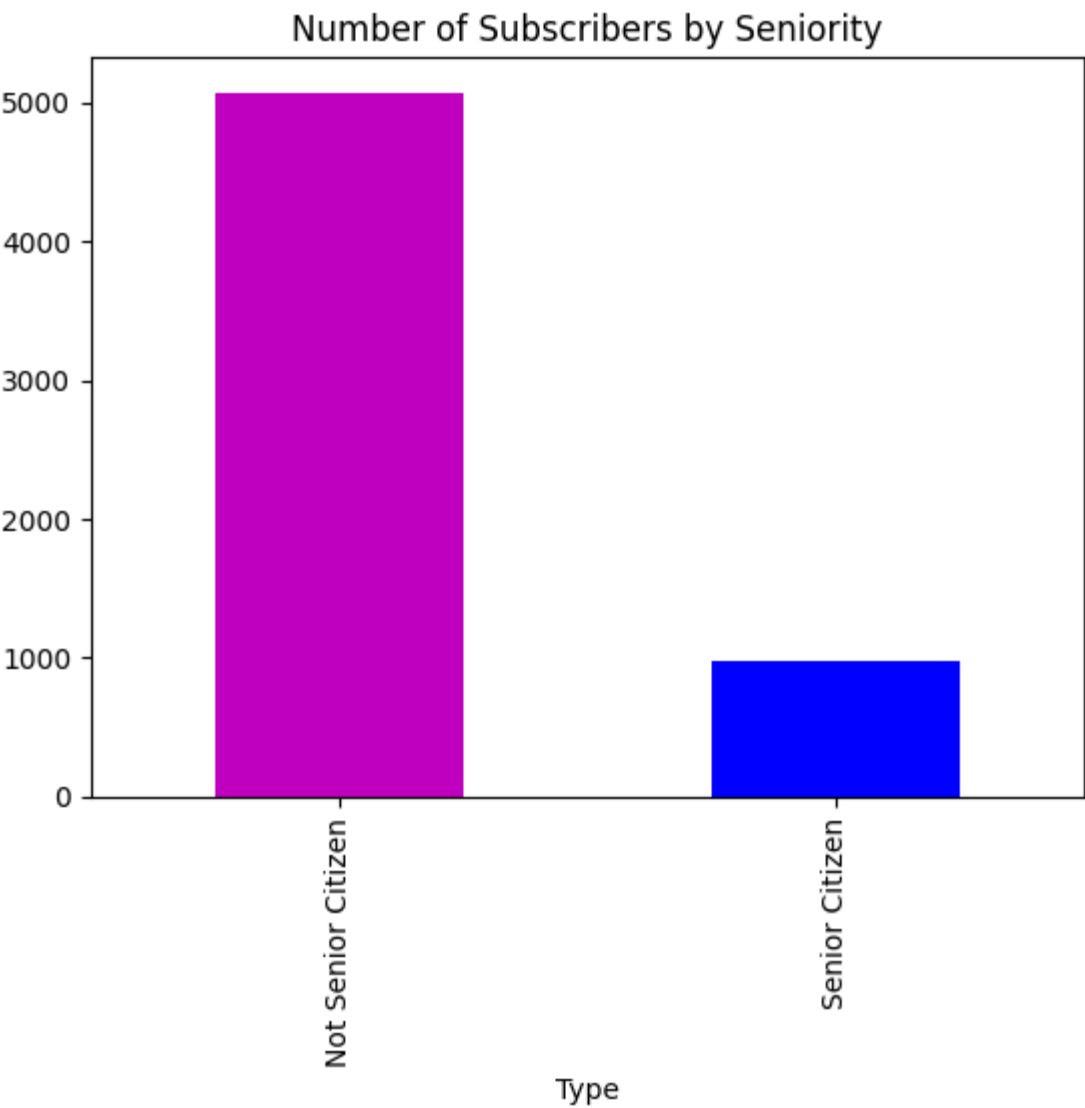
```
SeniorCitizen
0      5069
1       981
Name: count, dtype: int64
```

```
In[: # replace 1 and 0 with senior citizen and not senior citizen
      df['SeniorCitizen'].replace([0,1],['Not Senior Citizen','Senior
      Citizen'],inplace=True)
```

```
In[: #customer type
      customer = df["SeniorCitizen"].value_counts()
      print(customer)
```

```
SeniorCitizen
Not Senior Citizen    5069
Senior Citizen        981
Name: count, dtype: int64
```

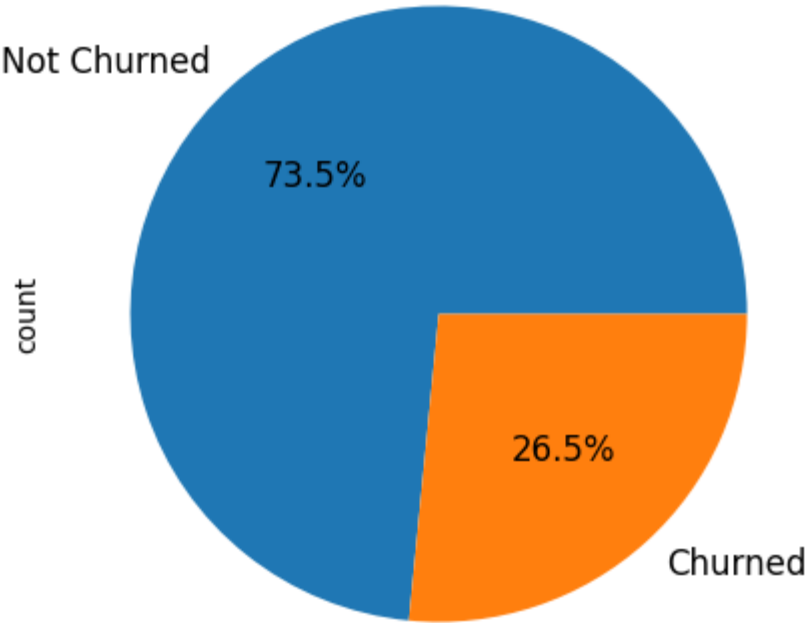
```
In[: # visualisation senior citizen
      import matplotlib.pyplot as plt
      customer.plot(kind = 'bar', color=['m', 'b'])
      plt.title('Number of Subscribers by Seniority')
      plt.xlabel('Type')
      plt.show()
```



```
In [ ]: #churn
        churn = df["Churn"].value_counts()
        print(churn)

Churn
No      4444
Yes     1606
Name: count, dtype: int64

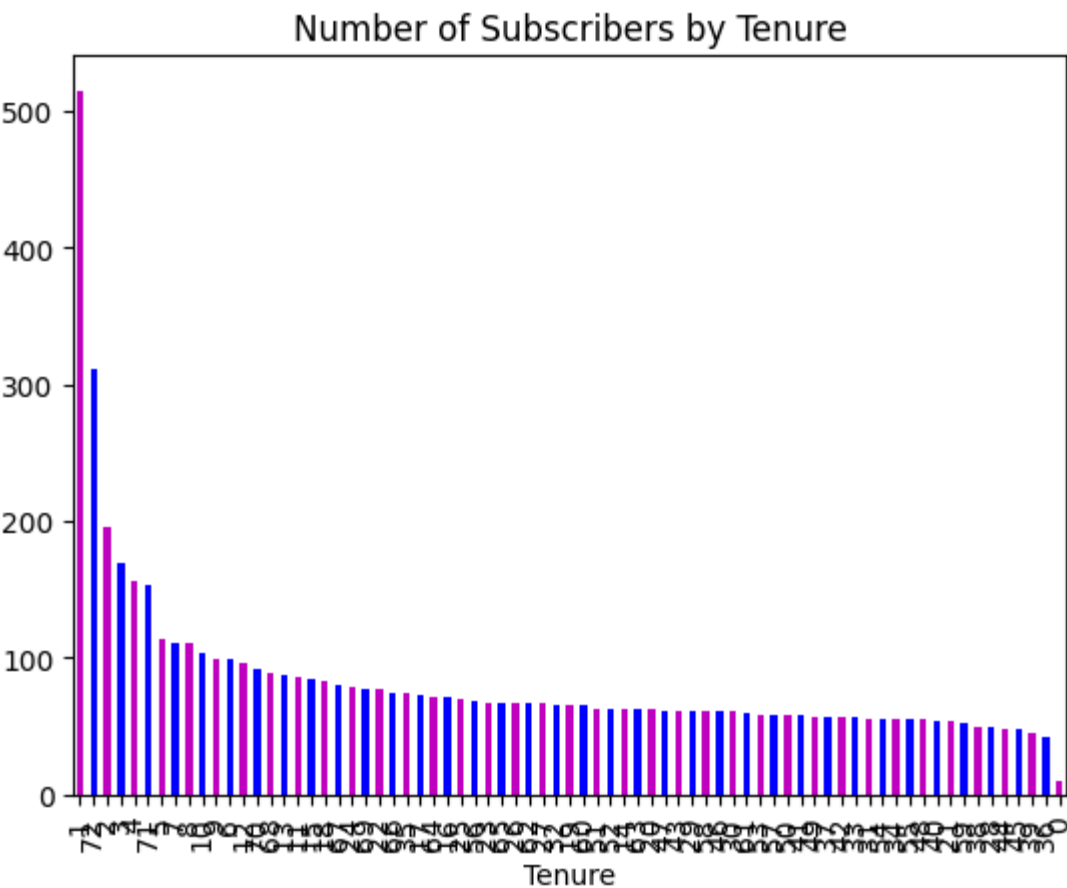
In [ ]: # Pie chart for Churn
        xis = (df['Churn'].value_counts() * 100.0 / len(df)).plot.pie(autopct='%.1f%%',
        labels = ['Not Churned', 'Churned'] ,figsize =(5,5), fontsize = 12)
```



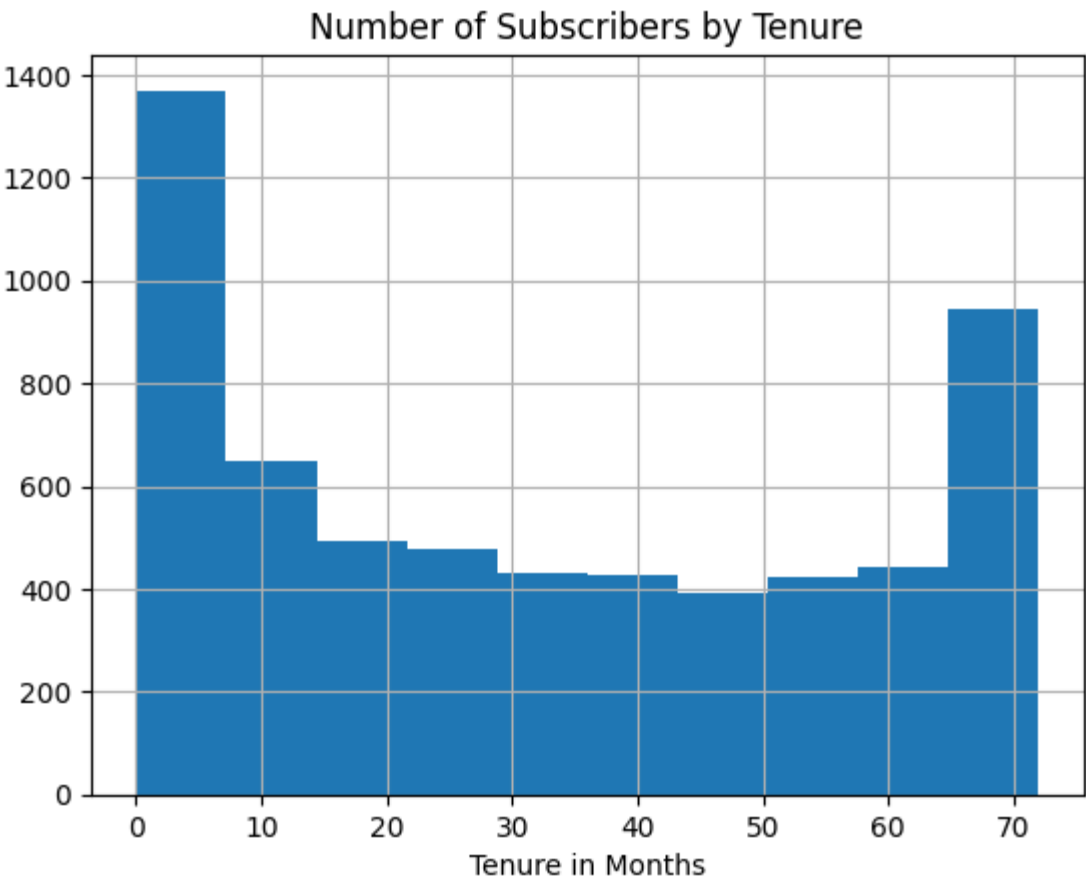
```
In [ ]: #tenure
        tenure = df["Tenure"].value_counts()
        print(tenure)

Tenure
1      514
72     311
2      196
3      170
4      156
...
44      48
45      48
39      45
36      43
0       10
Name: count, Length: 73, dtype: int64

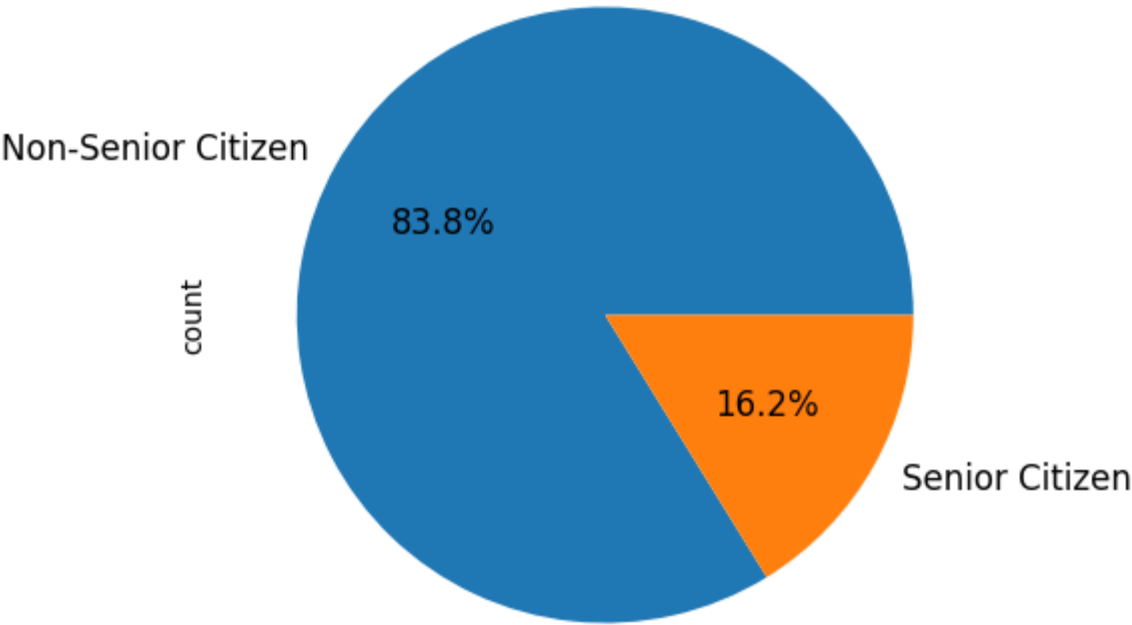
In [ ]: # visualisation tenure
        import matplotlib.pyplot as plt
        tenure.plot(kind = 'bar', color=['m', 'b'])
        plt.title('Number of Subscribers by Tenure')
        plt.xlabel('Tenure')
        plt.show()
```



```
In [ ]: # Histogram for tenure
df.hist(column='Tenure')
plt.title('Number of Subscribers by Tenure')
plt.xlabel('Tenure in Months')
plt.show()
```



```
In [ ]: # Pie chart for seniority
axis = (df['SeniorCitizen'].value_counts() * 100.0 /
len(df)).plot.pie(autopct='%.1f%%', labels = ['Non-Senior Citizen', 'Senior Citizen']
,figsize =(5,5), fontsize = 12)
```



```
In [ ]:
```



```
# Try group-by if it works - will be more efficient
# check columns
df.columns

Out[:]:Index(['customerID', 'gender', 'SeniorCitizen', 'Partner', 'Dependents', 'Tenure', 'PhoneService', 'MultipleLines', 'InternetService', 'OnlineSecurity', 'OnlineBackup',
             'DeviceProtection', 'TechSupport', 'StreamingTV', 'StreamingMovies', 'Contract', 'PaperlessBilling', 'PaymentMethod', 'MonthlyCharges', 'TotalCharges', 'Churn', 'Provider',
             'Age'],
            dtype='object')

In[:]: # grouping by select variables - trial
df.groupby('PhoneService').count()
```

Out[:]:

	customerID	gender	SeniorCitizen	Partner	Dependents	Tenure	MultipleLines	InternetService	OnlineSecurity	OnlineBackup	DeviceProtection	TechSupport	StreamingTV	StreamingMovies	Contract	PaperlessBilling	PaymentMethod	MonthlyCharges	TotalCharges	Churn	Provider	Age
PhoneService																						
No	598	598	598	598	598	598	598	598	598	598	598	598	598	598	598	598	598	598	598	598	598	598
Yes	5452	5452	5452	5452	5452	5452	5452	5452	5452	5452	5452	5452	5452	5452	5452	5452	5452	5452	5452	5452	5452	5452

2 rows × 22 columns

```
In[:]: df.columns

Out[:]:Index(['customerID', 'gender', 'SeniorCitizen', 'Partner', 'Dependents', 'Tenure', 'PhoneService', 'MultipleLines', 'InternetService', 'OnlineSecurity', 'OnlineBackup',
             'DeviceProtection', 'TechSupport', 'StreamingTV', 'StreamingMovies', 'Contract', 'PaperlessBilling', 'PaymentMethod', 'MonthlyCharges', 'TotalCharges', 'Churn', 'Provider',
             'Age'],
            dtype='object')

In[:]: # Phone Service
df_gp_data = df.groupby(['gender', 'SeniorCitizen', 'PhoneService'])

In[:]: # Phone Service
phone_group = df_gp_data['PhoneService'].count()
print(phone_group)
```

gender	SeniorCitizen	PhoneService	
Female	Not Senior Citizen	No	254
		Yes	2224
	Senior Citizen	No	37
		Yes	454
Male	Not Senior Citizen	No	255
		Yes	2336
	Senior Citizen	No	52
		Yes	438

```
Name: PhoneService, dtype: int64

In[:]: phone_group.to_excel('phone groups.xlsx', index=False, sheet_name='Sheet1')
```

```
In[:]: df.columns

Out[:]:Index(['customerID', 'gender', 'SeniorCitizen', 'Partner', 'Dependents', 'Tenure', 'PhoneService', 'MultipleLines', 'InternetService', 'OnlineSecurity', 'OnlineBackup',
             'DeviceProtection', 'TechSupport', 'StreamingTV', 'StreamingMovies', 'Contract', 'PaperlessBilling', 'PaymentMethod', 'MonthlyCharges', 'TotalCharges', 'Churn', 'Provider',
             'Age'],
            dtype='object')
```

```
er',
    'Age'],
    dtype='object')

In[: # Internet Service
df_gp_data2 = df.groupby(['gender', 'SeniorCitizen', 'InternetService'])
internet_group = df_gp_data2['InternetService'].count()
print(internet_group)
```

gender	SeniorCitizen	InternetService	
Female	Not Senior Citizen	DSL	909
		Fiber optic	954
		No	615
	Senior Citizen	DSL	108
		Fiber optic	368
		No	15
Male	Not Senior Citizen	DSL	959
		Fiber optic	988
		No	644
	Senior Citizen	DSL	116
		Fiber optic	348
		No	26

Name: InternetService, dtype: int64

```
In[: # Device Protection
df_gp_data3 = df.groupby(['gender', 'SeniorCitizen', 'DeviceProtection'])
device_group = df_gp_data3['DeviceProtection'].count()
print(device_group)
```

gender	SeniorCitizen	DeviceProtection	
Female	Not Senior Citizen	No	1046
		No internet service	615
		Yes	817
	Senior Citizen	No	265
		No internet service	15
		Yes	211
Male	Not Senior Citizen	No	1089
		No internet service	644
		Yes	858
	Senior Citizen	No	270
		No internet service	26
		Yes	194

Name: DeviceProtection, dtype: int64

```
In[: # Streaming TV
df_gp_data4 = df.groupby(['gender', 'SeniorCitizen', 'StreamingTV'])
streaming_group = df_gp_data4['StreamingTV'].count()
print(streaming_group)
```

gender	SeniorCitizen	StreamingTV	
Female	Not Senior Citizen	No	956
		No internet service	615
		Yes	907
	Senior Citizen	No	223
		No internet service	15
		Yes	253
Male	Not Senior Citizen	No	1014
		No internet service	644
		Yes	933
	Senior Citizen	No	225

No internet service 26
Yes 239

Name: StreamingTV, dtype: int64

```
In[]: # Paperless Billing
df_gp_data5 = df.groupby(['gender', 'SeniorCitizen', 'PaperlessBilling'])
paperless_group = df_gp_data5['PaperlessBilling'].count()
print(paperless_group)
```

gender	SeniorCitizen	PaperlessBilling	
Female	Not Senior Citizen	No	1077
		Yes	1401
	Senior Citizen	No	116
		Yes	375
Male	Not Senior Citizen	No	1167
		Yes	1424
	Senior Citizen	No	117
		Yes	373

Name: PaperlessBilling, dtype: int64

```
In[]: # Payment Method
df_gp_data6 = df.groupby(['gender', 'SeniorCitizen', 'PaymentMethod'])
paymethod_group = df_gp_data6['PaymentMethod'].count()
print(paymethod_group)
```

gender	SeniorCitizen	PaymentMethod	
Female	Not Senior Citizen	Bank transfer (automatic)	571
		Credit card (automatic)	545
		Electronic check	734
		Mailed check	628
	Senior Citizen	Bank transfer (automatic)	100
		Credit card (automatic)	98
		Electronic check	253
		Mailed check	40
Male	Not Senior Citizen	Bank transfer (automatic)	547
		Credit card (automatic)	584
		Electronic check	790
		Mailed check	670
	Senior Citizen	Bank transfer (automatic)	103
		Credit card (automatic)	88
		Electronic check	256
		Mailed check	43

Name: PaymentMethod, dtype: int64

```
In[]: # Contract
df_gp_data7 = df.groupby(['gender', 'SeniorCitizen', 'Contract'])
contract_group = df_gp_data7['Contract'].count()
print(contract_group)
```

gender	SeniorCitizen	Contract	
Female	Not Senior Citizen	Month-to-month	1287
		One year	523
		Two year	668
	Senior Citizen	Month-to-month	346
		One year	87
		Two year	58
Male	Not Senior Citizen	Month-to-month	1325
		One year	594
		Two year	672
	Senior Citizen	Month-to-month	347

One year	84
Two year	59

Name: Contract, dtype: int64

```
In[: # Convert the file to html
```

```
%%shell
```

```
jupyter nbconvert --to html /content/Telecom_Users_Data_Analysis.ipynb
```

[NbConvertApp] Converting notebook /content/Telecom_Users_Data_Analysis.ipynb to html

[NbConvertApp] Writing 1556729 bytes to /content/Telecom_Users_Data_Analysis.html

Out[: