Telco Customer Churn

Description:

Customer Churn means loss of customers/clients With the rapid development of the telecommunications industry, service providers tend to lean towards the expansion of subscriber base because they are the business target market. Telephone service companies, internet service providers, TV companies and insurance firms, often using customer churn analysis and customer churn rates as one of their key business metrics because maintaining existing customers cost is much lower than receiving a new one.

Therefore, it can help to estimate customer support by gathering knowledge from the telecommunications industry, whether they will leave the company or not. Telecommunication industries need to take necessary action to take initiatives to acquire their affiliate customers to stabilize their market value

Data Overview:

• Rows (represents customer) :7043

• Columns (represents customer's attributes) :21

Problem Statement: Predict behavior to retain customers i.e. we need to analyze all relevant customer data and develop focused customer retention program.

In [134]:

```python
#importing all the packages
import numpy as np
import pandas as pd
import os
import seaborn as sns
import matplotlib.pyplot as plt
```

In [135]:

```python
#Read the dataset and display the first five entries using head function.
telco = pd.read_csv('C:\\Users\\sshah\\Desktop\\MVA\\project\\Telco-Customer-Churn.csv')
telco.head()
```

Out[135]:

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	...	D
0	7590-VHVEG	Female	0	Yes	No	1	No	No phone service	DSL	No	...	
1	5575-GNVDE	Male	0	No	No	34	Yes	No	DSL	Yes	...	
2	3668-QPYBK	Male	0	No	No	2	Yes	No	DSL	Yes	...	
3	7795-CFOCW	Male	0	No	No	45	No	No phone service	DSL	Yes	...	
4	9237-HQITU	Female	0	No	No	2	Yes	No	Fiber optic	No	...	

5 rows × 21 columns

In [136]:

```python
#Display the number of dimensions in the array.
telco.shape
```

Out[136]:

(7043, 21)

In [137]:

```python
#nunique() method is used to get number of all unique values in all the columns.
print(telco.info())
telco.nunique()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7043 entries, 0 to 7042
Data columns (total 21 columns):
customerID          7043 non-null object
gender              7043 non-null object
SeniorCitizen       7043 non-null int64
Partner             7043 non-null object
Dependents          7043 non-null object
tenure              7043 non-null int64
PhoneService        7043 non-null object
MultipleLines       7043 non-null object
InternetService     7043 non-null object
OnlineSecurity      7043 non-null object
OnlineBackup        7043 non-null object
DeviceProtection    7043 non-null object
TechSupport         7043 non-null object
StreamingTV         7043 non-null object
StreamingMovies     7043 non-null object
Contract            7043 non-null object
PaperlessBilling    7043 non-null object
PaymentMethod       7043 non-null object
MonthlyCharges      7043 non-null float64
TotalCharges        7043 non-null object
Churn               7043 non-null object
dtypes: float64(1), int64(2), object(18)
memory usage: 1.1+ MB
None
```

Out[137]:

```
customerID          7043
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      1585
TotalCharges        6531
Churn                  2
dtype: int64
```

In [138]:

```python
#Replacing the blank spaces in Total Charges column with Nan.
telco['TotalCharges'] = telco["TotalCharges"].replace(" ",np.nan)

telco = telco[telco["TotalCharges"].notnull()]
telco = telco.reset_index()[telco.columns]

#Converting Total charges column values to float.
telco["TotalCharges"] = telco["TotalCharges"].astype(float)
```

In [139]:

```python
#Using the sum we get the count of NaN's for each column
telco.isna().sum()
```

```
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
dtype: int64
```

In [140]:

```python
#Replace Numeric 0's and 1's in SeniorCitizen column with 'No' and 'Yes'.
telco['SeniorCitizen'] = telco['SeniorCitizen'].replace({1:'Yes', 0:'No'})
telco
```

Out[140]:

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	...
0	7590-VHVEG	Female	No	Yes	No	1	No	No phone service	DSL	No	...
1	5575-GNVDE	Male	No	No	No	34	Yes	No	DSL	Yes	...
2	3668-QPYBK	Male	No	No	No	2	Yes	No	DSL	Yes	...
3	7795-CFOCW	Male	No	No	No	45	No	No phone service	DSL	Yes	...
4	9237-HQITU	Female	No	No	No	2	Yes	No	Fiber optic	No	...
5	9305-CDSKC	Female	No	No	No	8	Yes	Yes	Fiber optic	No	...
6	1452-KIOVK	Male	No	No	Yes	22	Yes	Yes	Fiber optic	No	...
7	6713-OKOMC	Female	No	No	No	10	No	No phone service	DSL	Yes	...
8	7892-POOKP	Female	No	Yes	No	28	Yes	Yes	Fiber optic	No	...
9	6388-TABGU	Male	No	No	Yes	62	Yes	No	DSL	Yes	...
10	9763-GRSKD	Male	No	Yes	Yes	13	Yes	No	DSL	Yes	...
11	7469-LKBCI	Male	No	No	No	16	Yes	No	No	No internet service	...
12	8091-TTVAX	Male	No	Yes	No	58	Yes	Yes	Fiber optic	No	...
13	0280-XJGEX	Male	No	No	No	49	Yes	Yes	Fiber optic	No	...
14	5129-JLPIS	Male	No	No	No	25	Yes	No	Fiber optic	Yes	...
15	3655-SNQYZ	Female	No	Yes	Yes	69	Yes	Yes	Fiber optic	Yes	...
16	8191-XWSZG	Female	No	No	No	52	Yes	No	No	No internet service	...

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	service ...
17	9959-WOFKT	Male	No	No	Yes	71	Yes	Yes	Fiber optic	Yes	...
18	4190-MFLUW	Female	No	Yes	Yes	10	Yes	No	DSL	No	...
19	4183-MYFRB	Female	No	No	No	21	Yes	No	Fiber optic	No	...
20	8779-QRDMV	Male	Yes	No	No	1	No	No phone service	DSL	No	...
21	1680-VDCWW	Male	No	Yes	No	12	Yes	No	No	No internet service	...
22	1066-JKSGK	Male	No	No	No	1	Yes	No	No	No internet service	...
23	3638-WEABW	Female	No	Yes	No	58	Yes	Yes	DSL	No	...
24	6322-HRPFA	Male	No	Yes	Yes	49	Yes	No	DSL	Yes	...
25	6865-JZNKO	Female	No	No	No	30	Yes	No	DSL	Yes	...
26	6467-CHFZW	Male	No	Yes	Yes	47	Yes	Yes	Fiber optic	No	...
27	8665-UTDHZ	Male	No	Yes	Yes	1	No	No phone service	DSL	No	...
28	5248-YGIJN	Male	No	Yes	No	72	Yes	Yes	DSL	Yes	...
29	8773-HHUOZ	Female	No	No	Yes	17	Yes	No	DSL	No	...
...	...	...	...	...	...	...	...	...	...	...	...
7002	1685-BQULA	Female	No	No	No	40	Yes	Yes	Fiber optic	No	...
7003	9053-EJUNL	Male	No	No	No	41	Yes	Yes	Fiber optic	No	...
7004	0666-UXTJO	Male	Yes	Yes	No	34	Yes	No	Fiber optic	No	...
7005	1471-GIQKQ	Female	No	No	No	1	Yes	No	DSL	No	...
7006	4807-IZYOZ	Female	No	No	No	51	Yes	No	No	No internet service	...
7007	1122-JWTJW	Male	No	Yes	Yes	1	Yes	No	Fiber optic	No	...
7008	9710-NJERN	Female	No	No	No	39	Yes	No	No	No internet service	...
7009	9837-FWLCH	Male	No	Yes	Yes	12	Yes	No	No	No internet service	...
7010	1699-HPSBG	Male	No	No	No	12	Yes	No	DSL	No	...
7011	7203-OYKCT	Male	No	No	No	72	Yes	Yes	Fiber optic	No	...
7012	1035-IPQPU	Female	Yes	Yes	No	63	Yes	Yes	Fiber optic	No	...
7013	7398-LXGYX	Male	No	Yes	No	44	Yes	Yes	Fiber optic	Yes	...
7014	2823-LKABH	Female	No	No	No	18	Yes	Yes	Fiber optic	No	...
7015	8775-CEBBJ	Female	No	No	No	9	Yes	No	DSL	No	...
7016	0550-DCXLH	Male	No	No	No	13	Yes	No	DSL	No	...
7017	9281-CEDRU	Female	No	Yes	No	68	Yes	No	DSL	No	...
7018	2235-DWLJU	Female	Yes	No	No	6	No	No phone service	DSL	No	...
7019	0871-OPBXW	Female	No	No	No	2	Yes	No	No	No internet service	...
7020	3605-JISKB	Male	Yes	Yes	No	55	Yes	Yes	DSL	Yes	...

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	...
7021	6894-LFHLY	Male	Yes	No	No	1	Yes	Yes	Fiber optic	No	...
7022	9767-FFLEM	Male	No	No	No	38	Yes	No	Fiber optic	No	...
7023	0639-TSIQW	Female	No	No	No	67	Yes	Yes	Fiber optic	Yes	...
7024	8456-QDAVC	Male	No	No	No	19	Yes	No	Fiber optic	No	...
7025	7750-EYXWZ	Female	No	No	No	12	No	No phone service	DSL	No	...
7026	2569-WGERO	Female	No	No	No	72	Yes	No	No	No internet service	...
7027	6840-RESVB	Male	No	Yes	Yes	24	Yes	Yes	DSL	Yes	...
7028	2234-XADUH	Female	No	Yes	Yes	72	Yes	Yes	Fiber optic	No	...
7029	4801-JZAZL	Female	No	Yes	Yes	11	No	No phone service	DSL	Yes	...
7030	8361-LTMKD	Male	Yes	Yes	No	4	Yes	Yes	Fiber optic	No	...
7031	3186-AJIEK	Male	No	No	No	66	Yes	No	Fiber optic	Yes	...

7032 rows × 21 columns

In [141]:

```
#Replace the column value in of different columns.
columns_to_replace = [ 'OnlineSecurity', 'OnlineBackup', 'DeviceProtection','TechSupport','StreamingTV', 'StreamingMovies']
for i in columns_to_replace:
    telco[i]  = telco[i].replace({'No internet service' : 'No'})
telco
```

Out[141]:

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	...
0	7590-VHVEG	Female	No	Yes	No	1	No	No phone service	DSL	No	...
1	5575-GNVDE	Male	No	No	No	34	Yes	No	DSL	Yes	...
2	3668-QPYBK	Male	No	No	No	2	Yes	No	DSL	Yes	...
3	7795-CFOCW	Male	No	No	No	45	No	No phone service	DSL	Yes	...
4	9237-HQITU	Female	No	No	No	2	Yes	No	Fiber optic	No	...
5	9305-CDSKC	Female	No	No	No	8	Yes	Yes	Fiber optic	No	...
6	1452-KIOVK	Male	No	No	Yes	22	Yes	Yes	Fiber optic	No	...
7	6713-OKOMC	Female	No	No	No	10	No	No phone service	DSL	Yes	...
8	7892-POOKP	Female	No	Yes	No	28	Yes	Yes	Fiber optic	No	...
9	6388-TABGU	Male	No	No	Yes	62	Yes	No	DSL	Yes	...
10	9763-GRSKD	Male	No	Yes	Yes	13	Yes	No	DSL	Yes	...
11	7469-LKBCI	Male	No	No	No	16	Yes	No	No	No	...
12	8091-TTVAX	Male	No	Yes	No	58	Yes	Yes	Fiber optic	No	...
13	0280-XJGEX	Male	No	No	No	49	Yes	Yes	Fiber optic	No	...

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	
14	5129-JLPIS	Male	No	No	No	25	Yes	No	Fiber optic	Yes	...
15	3655-SNQYZ	Female	No	Yes	Yes	69	Yes	Yes	Fiber optic	Yes	...
16	8191-XWSZG	Female	No	No	No	52	Yes	No	No	No	...
17	9959-WOFKT	Male	No	No	Yes	71	Yes	Yes	Fiber optic	Yes	...
18	4190-MFLUW	Female	No	Yes	Yes	10	Yes	No	DSL	No	...
19	4183-MYFRB	Female	No	No	No	21	Yes	No	Fiber optic	No	...
20	8779-QRDMV	Male	Yes	No	No	1	No	No phone service	DSL	No	...
21	1680-VDCWW	Male	No	Yes	No	12	Yes	No	No	No	...
22	1066-JKSGK	Male	No	No	No	1	Yes	No	No	No	...
23	3638-WEABW	Female	No	Yes	No	58	Yes	Yes	DSL	No	...
24	6322-HRPFA	Male	No	Yes	Yes	49	Yes	No	DSL	Yes	...
25	6865-JZNKO	Female	No	No	No	30	Yes	No	DSL	Yes	...
26	6467-CHFZW	Male	No	Yes	Yes	47	Yes	Yes	Fiber optic	No	...
27	8665-UTDHZ	Male	No	Yes	Yes	1	No	No phone service	DSL	No	...
28	5248-YGIJN	Male	No	Yes	No	72	Yes	Yes	DSL	Yes	...
29	8773-HHUOZ	Female	No	No	Yes	17	Yes	No	DSL	No	...
...	...	...	...	...	...	...	...	...	...	...	
7002	1685-BQULA	Female	No	No	No	40	Yes	Yes	Fiber optic	No	...
7003	9053-EJUNL	Male	No	No	No	41	Yes	Yes	Fiber optic	No	...
7004	0666-UXTJO	Male	Yes	Yes	No	34	Yes	No	Fiber optic	No	...
7005	1471-GIQKQ	Female	No	No	No	1	Yes	No	DSL	No	...
7006	4807-IZYOZ	Female	No	No	No	51	Yes	No	No	No	...
7007	1122-JWTJW	Male	No	Yes	Yes	1	Yes	No	Fiber optic	No	...
7008	9710-NJERN	Female	No	No	No	39	Yes	No	No	No	...
7009	9837-FWLCH	Male	No	Yes	Yes	12	Yes	No	No	No	...
7010	1699-HPSBG	Male	No	No	No	12	Yes	No	DSL	No	...
7011	7203-OYKCT	Male	No	No	No	72	Yes	Yes	Fiber optic	No	...
7012	1035-IPQPU	Female	Yes	Yes	No	63	Yes	Yes	Fiber optic	No	...
7013	7398-LXGYX	Male	No	Yes	No	44	Yes	Yes	Fiber optic	Yes	...
7014	2823-LKABH	Female	No	No	No	18	Yes	Yes	Fiber optic	No	...
7015	8775-CEBBJ	Female	No	No	No	9	Yes	No	DSL	No	...
7016	0550-DCXLH	Male	No	No	No	13	Yes	No	DSL	No	...
7017	9281-CEDRU	Female	No	Yes	No	68	Yes	No	DSL	No	...

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	...
7018	2235-DWLJU	Female	Yes	No	No	8	No	No phone service	DSL	No	...
7019	0871-OPBXW	Female	No	No	No	2	Yes	No	No	No	...
7020	3605-JISKB	Male	Yes	Yes	No	55	Yes	Yes	DSL	Yes	...
7021	6894-LFHLY	Male	Yes	No	No	1	Yes	Yes	Fiber optic	No	...
7022	9767-FFLEM	Male	No	No	No	38	Yes	No	Fiber optic	No	...
7023	0639-TSIQW	Female	No	No	No	67	Yes	Yes	Fiber optic	Yes	...
7024	8456-QDAVC	Male	No	No	No	19	Yes	No	Fiber optic	No	...
7025	7750-EYXWZ	Female	No	No	No	12	No	No phone service	DSL	No	...
7026	2569-WGERO	Female	No	No	No	72	Yes	No	No	No	...
7027	6840-RESVB	Male	No	Yes	Yes	24	Yes	Yes	DSL	Yes	...
7028	2234-XADUH	Female	No	Yes	Yes	72	Yes	Yes	Fiber optic	No	...
7029	4801-JZAZL	Female	No	Yes	Yes	11	No	No phone service	DSL	Yes	...
7030	8361-LTMKD	Male	Yes	Yes	No	4	Yes	Yes	Fiber optic	No	...
7031	3186-AJIEK	Male	No	No	No	66	Yes	No	Fiber optic	Yes	...

7032 rows × 21 columns

In [142]:

```python
#Returns the max value from the tenure column.
telco["tenure"].max()
```

Out[142]:

72

In [143]:

```python
def tenure_lab(telco) :

    if telco["tenure"] <= 12 :
        return "1 Year"

    elif (telco["tenure"] > 12) & (telco["tenure"] <= 24) :
        return "1-2 Year"

    elif (telco["tenure"] > 24) & (telco["tenure"] <= 48) :
        return "2-3 Year"

    elif (telco["tenure"] > 48) & (telco["tenure"] <= 60) :
        return "3-4 Year"

    elif telco["tenure"] > 60  & (telco["tenure"] <= 72) :
        return "4-5 Year"
```

In [144]:

```python
telco["tenure"] = telco.apply(lambda telco:tenure_lab(telco), axis = 1)
```

In [145]:

```python
telco.head()
```

Out[145]:

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	...	D
0	7590-VHVEG	Female	No	Yes	No	1 Year	No	No phone service	DSL	No	...	
1	5575-GNVDE	Male	No	No	No	2-3 Year	Yes	No	DSL	Yes	...	
2	3668-QPYBK	Male	No	No	No	1 Year	Yes	No	DSL	Yes	...	
3	7795-CFOCW	Male	No	No	No	2-3 Year	No	No phone service	DSL	Yes	...	
4	9237-HQITU	Female	No	No	No	1 Year	Yes	No	Fiber optic	No	...	

5 rows × 21 columns

In [146]:

```python
#Distinguish Churn and non-churn customers.
churn_yes = telco[telco["Churn"]=="Yes"]
churn_no = telco[telco["Churn"]=="No"]
```

In [147]:

```python
# Plot
plt.scatter(TotalCharges,MonthlyCharges, s=area, c=colors, alpha=0.5)
plt.title('Scatter plot pythonspot.com')
plt.xlabel('TotalCharges')
plt.ylabel('MonthlyCharges')
plt.show()
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call last)
<ipython-input-147-dc41ee031773> in <module>
      1 # Plot
----> 2 plt.scatter(TotalCharges,MonthlyCharges, s=area, c=colors, alpha=0.5)
      3 plt.title('Scatter plot pythonspot.com')
      4 plt.xlabel('TotalCharges')
      5 plt.ylabel('MonthlyCharges')

NameError: name 'TotalCharges' is not defined
```

In [148]:

```python
sns.distplot(telco['MonthlyCharges'],kde=False, rug=True)
```

Out[148]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x27dd3383cf8>
```
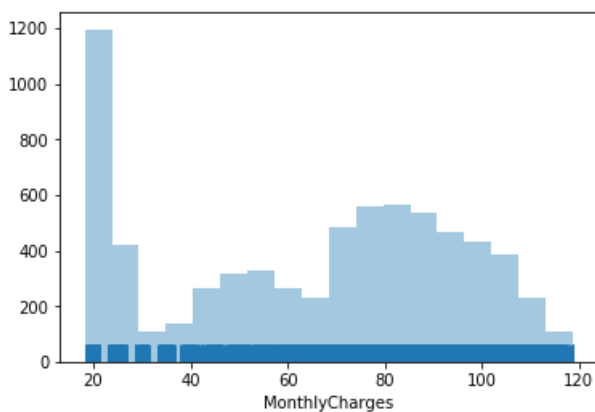

In [149]:

```python
sns.distplot(telco['TotalCharges'],kde=False, rug = True)
```

<matplotlib.axes._subplots.AxesSubplot at 0x27dd5143b70>
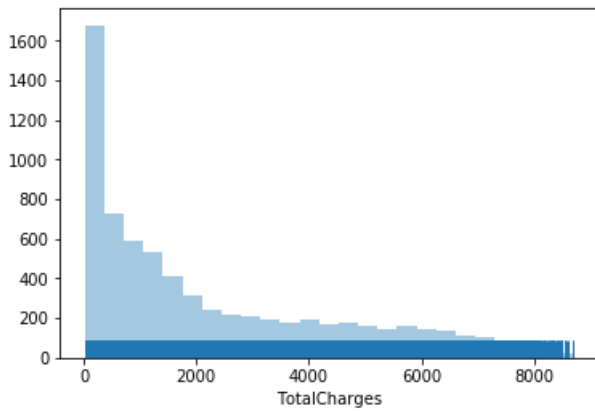
In [116]:

```
sns.kdeplot(telco['MonthlyCharges'])
```

Out[116]:

<matplotlib.axes._subplots.AxesSubplot at 0x27dd34fe438>
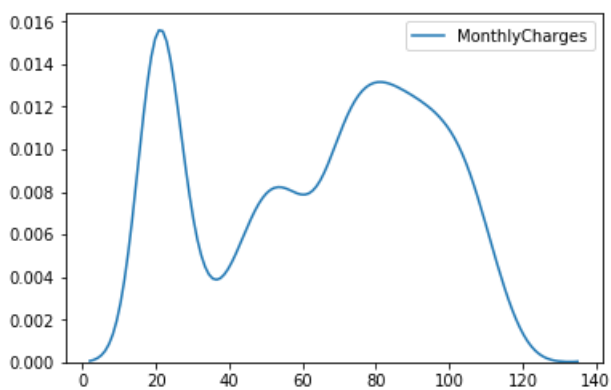
In [117]:

```
sns.kdeplot(telco['TotalCharges'])
```

Out[117]:

<matplotlib.axes._subplots.AxesSubplot at 0x27dd2eaf208>
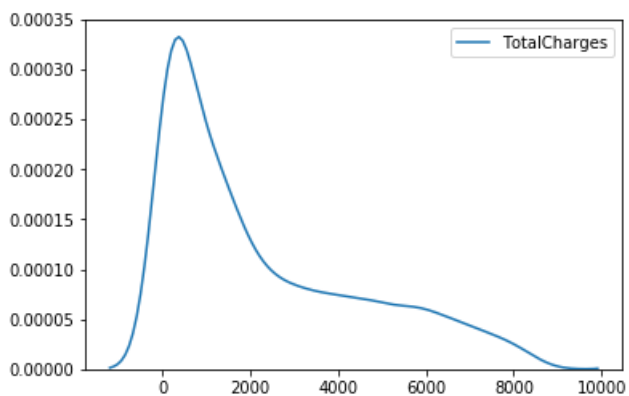
In [118]:

```
sns.jointplot(x='MonthlyCharges',y='TotalCharges',data=telco)
```
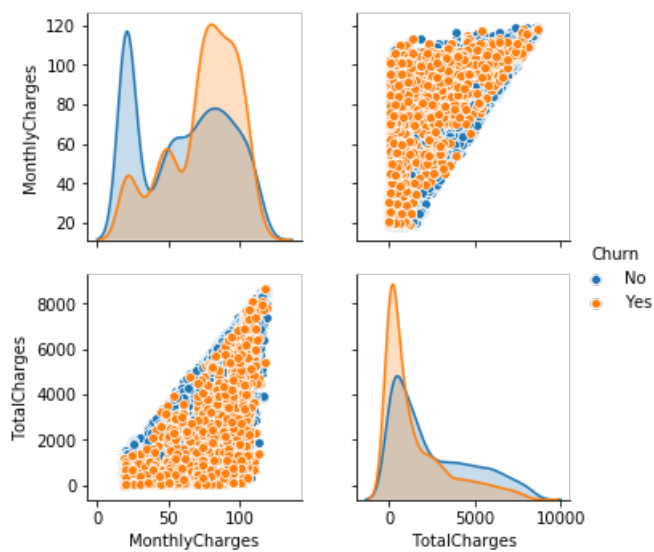
Out[118]:
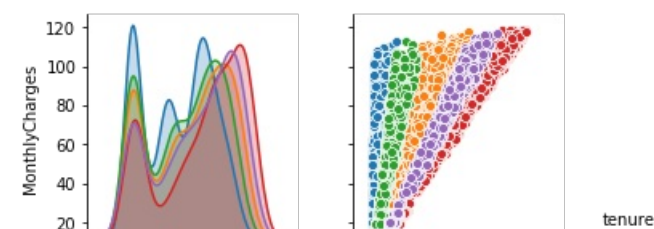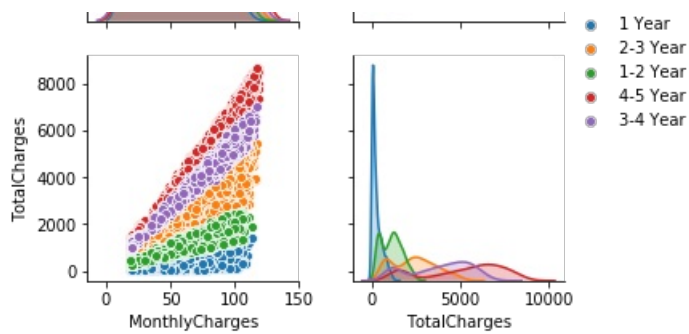
```
<seaborn.axisgrid.JointGrid at 0x27dd26824a8>
```
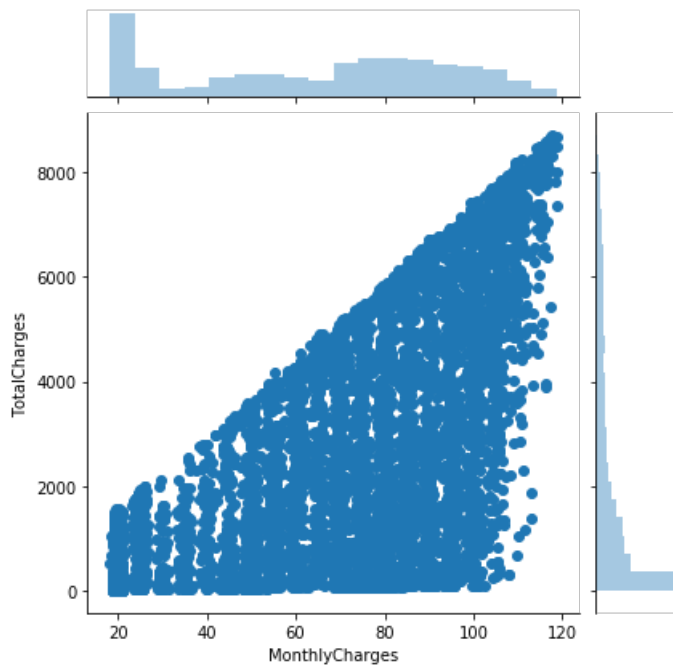
```
sns.pairplot(telco,hue ='Churn')
```

```
<seaborn.axisgrid.PairGrid at 0x27dd5570b38>
```

```
sns.pairplot(telco,hue='tenure')
```

```
<seaborn.axisgrid.PairGrid at 0x27dd2b298d0>
```

In [121]:

```
sns.countplot(x='gender',data=telco)
```

Out[121]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x27dd2cfccc0>
```
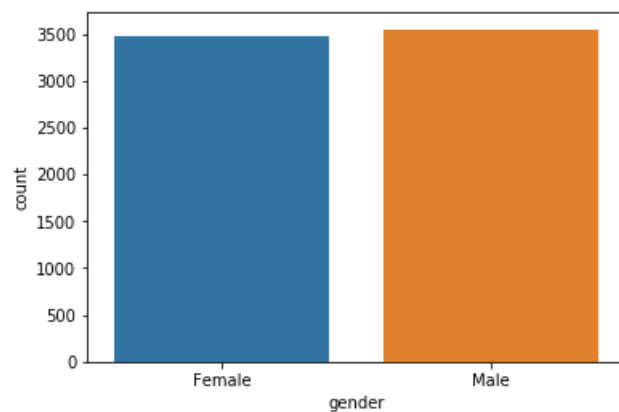

In [122]:

```
sns.countplot(x='tenure',data=telco,hue='Churn')
```

Out[122]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x27dd2d9c828>
```
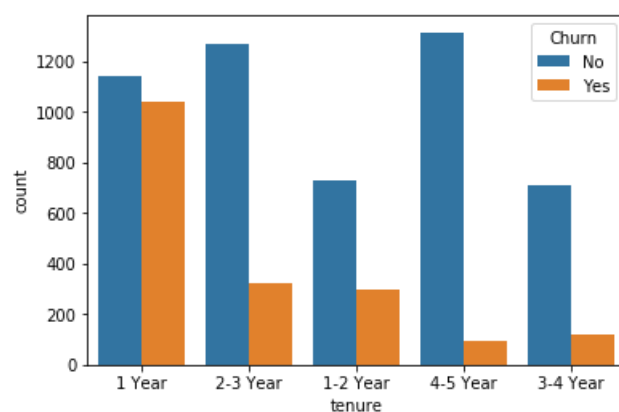

In [123]:

```
sns.barplot(x='gender',y='TotalCharges',data=telco)
```

Out[123]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x27dd2df0b38>
```
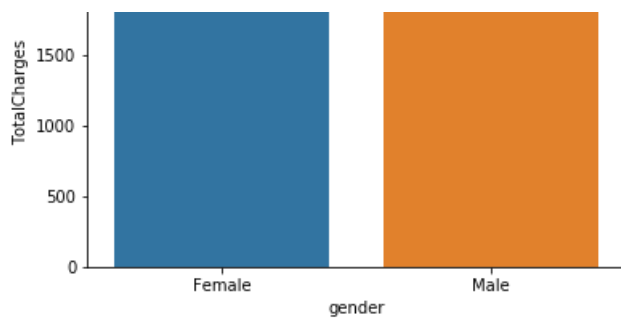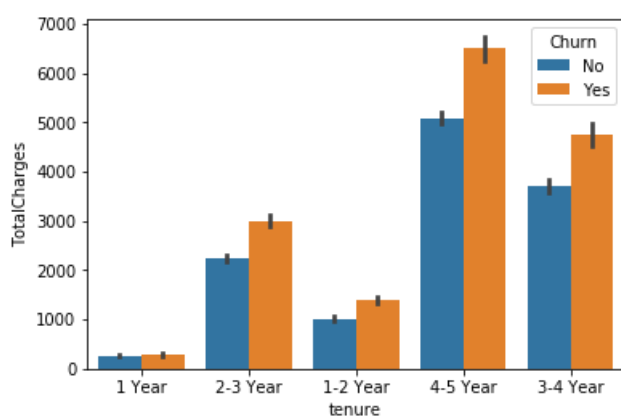
```
t = telco.corr()
sns.heatmap(t)
```

Out[124]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x27dd2fa08d0>
```
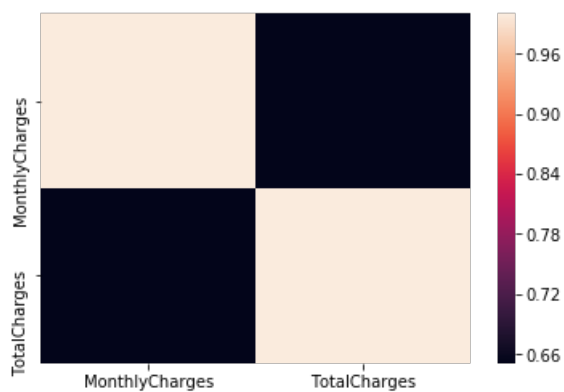

In [125]:

```
sns.barplot(x='tenure',y='TotalCharges',data=telco,hue='Churn')
```
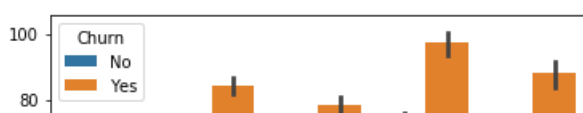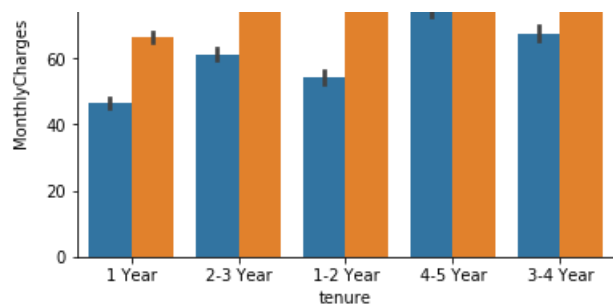
Out[125]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x27dd3019eb8>
```


In [126]:

```
sns.barplot(x='tenure',y='MonthlyCharges',data=telco,hue='Churn')
```

Out[126]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x27dd30aa048>
```

```
correlation = telco.corr()
matrix_cols = correlation.columns.tolist()
corr_array  = np.array(correlation)
corr_array
```

Out[127]:

```
array([[1.       , 0.6510648],
       [0.6510648, 1.       ]])
```