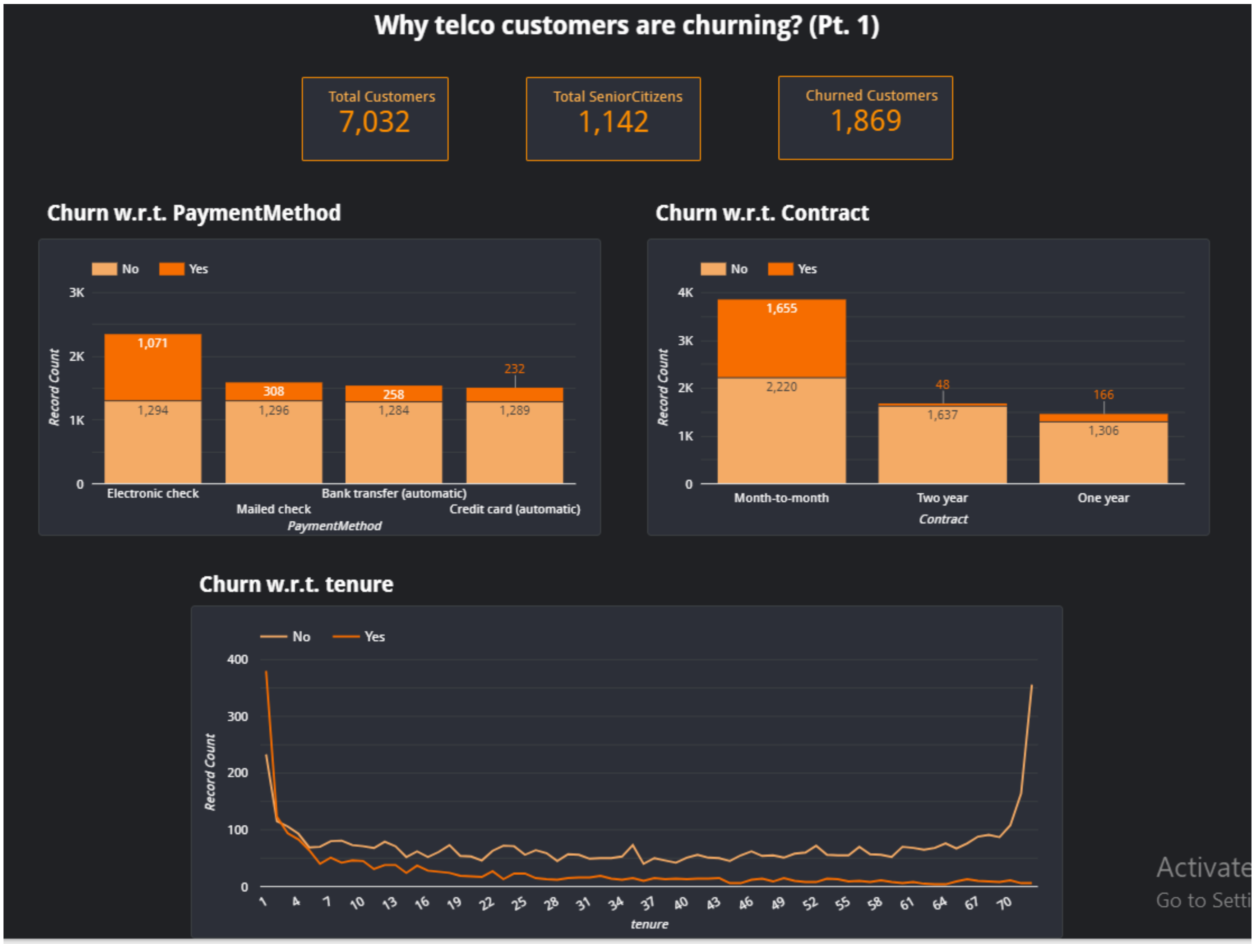
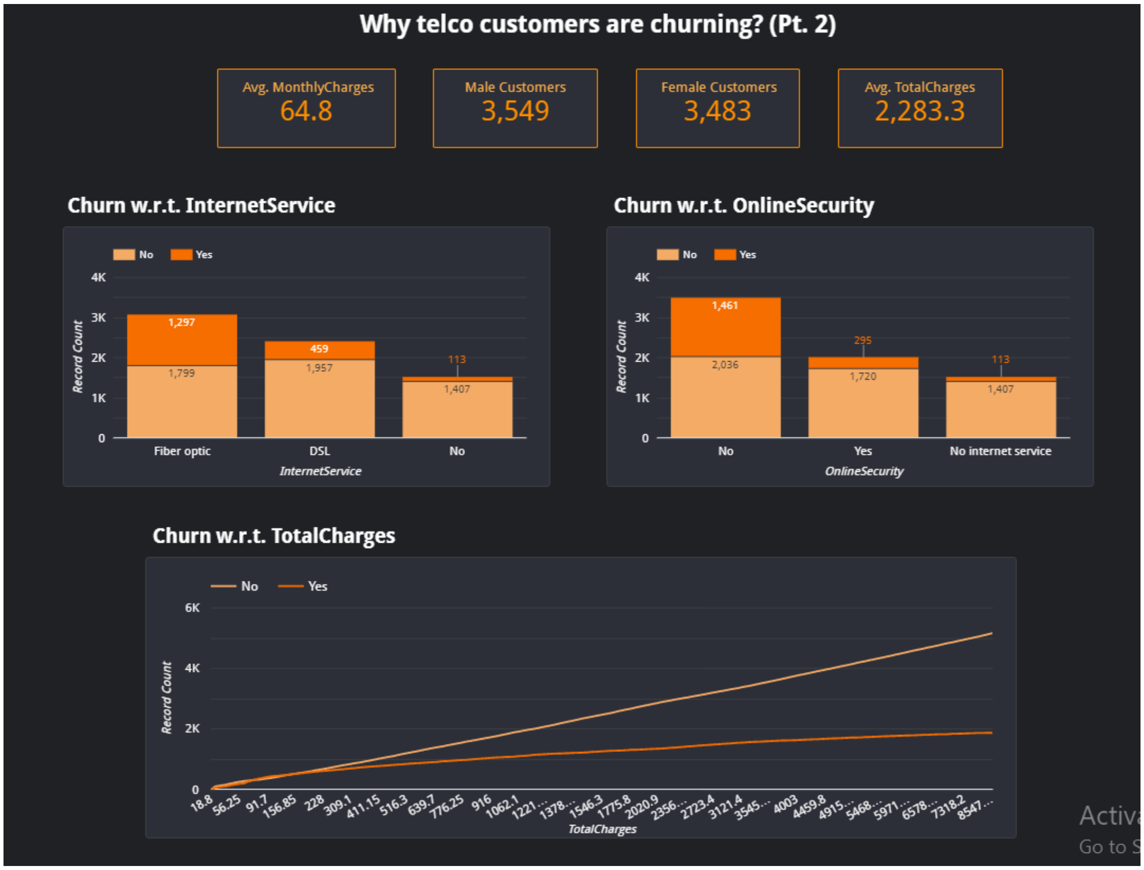
Business Intelligence – TelcoChurn Assignment

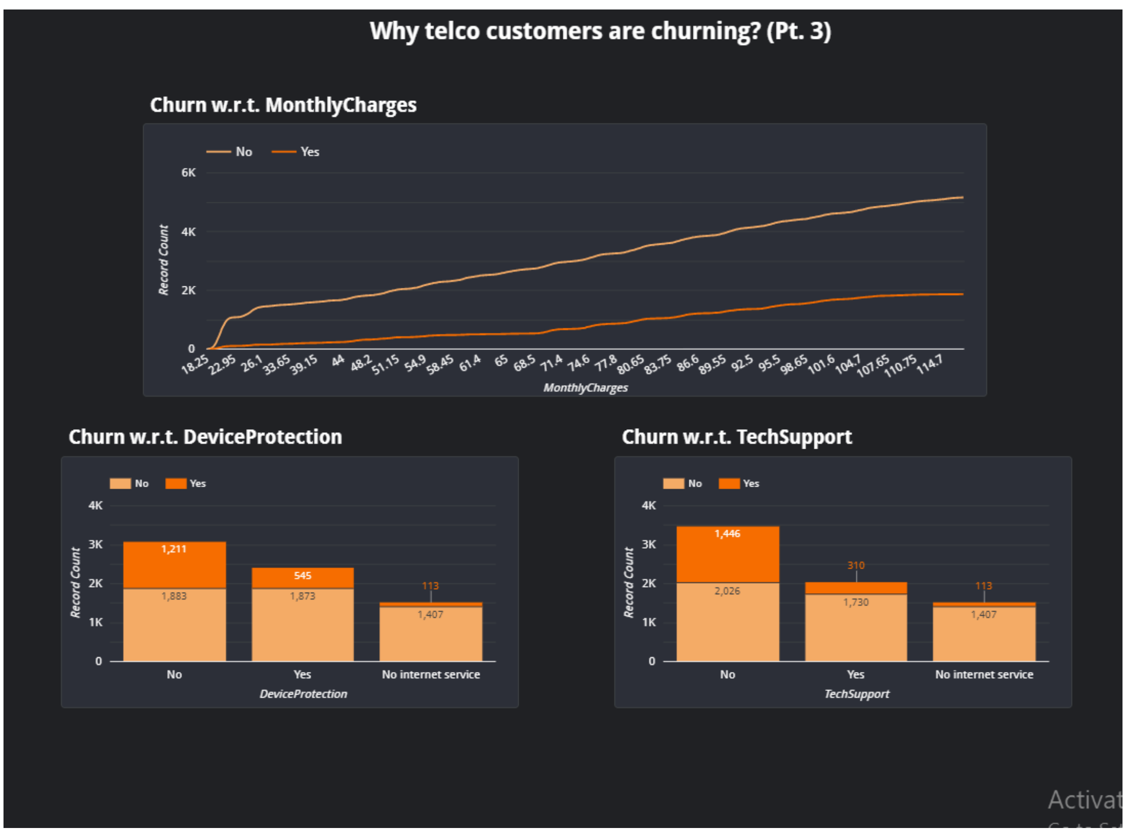
Marium Jamal – 14881

Google Data Studio Link:
https://datastudio.google.com/reporting/e7d4426e-ff28-465d-a4be-745309969129

Report:




Why telco customers are churning? (Pt. 2)
Avg. MonthlyCharges 64.8
Male Customers 3,549
Female Customers 3,483
Avg. TotalCharges 2,283.3
Churn w.r.t. InternetService
No
Yes
Record Count
1,297
1,799
459
1,957
113
1,407
Fiber optic
DSL
No
InternetService
Churn w.r.t. OnlineSecurity
No
Yes
Record Count
1,461
2,036
295
1,720
113
1,407
No
Yes
No internet service
OnlineSecurity
Churn w.r.t. TotalCharges
No
Yes
Record Count
TotalCharges


Why telco customers are churning? (Pt. 3)
Churn w.r.t. MonthlyCharges
No
Yes
Record Count
MonthlyCharges
Churn w.r.t. DeviceProtection
No
Yes
Record Count
1,211
1,883
545
1,873
113
1,407
No
Yes
No internet service
DeviceProtection
Churn w.r.t. TechSupport
No
Yes
Record Count
1,446
2,026
310
1,730
113
1,407
No
Yes
No internet service
TechSupport

Explanation:

1) **Churn w.r.t PaymentMethod:**
Graph indicates that although electronic check brings the large number of customers but it also causes the large amount of customer churn among all other categories. Hence, organisation must introduce offers that attract the customers who are availing electronic check facility.
2) **Churn w.r.t Contract:**
Although, Month-to-Month has the largest number of customers but the churn rate is also high for this category; company must focus on identifying the problems that are causing customers in Month-to-Month category to churn.
3) **Churn w.r.t Tenure:**
Graph indicates that as the number of tenures increases, churn decreases. This indicates that old customers are likely to stay. Therefore, company must involve its customers in activities that will keep them in a contract for longer time, leading to increase in their tenure which will eventually cause a decline in churn. It also indicates that new customers are not staying and are churning, company must introduce offers which can persuade the customers to stay.
4) **Churn w.r.t InternetService:**
Large number of customers are availing fiber optic service, but the rate of churn is also high in this category. Therefore, company must focus on working on the weak areas of fiber optic service because large number of company's customers avail this service.
5) **Churn w.r.t OnlineSecurity:**
Customers who are not provided with OnlineSecurity are churning more because customers' first priority is to ensure security and failure to do so causes them to churn. Hence, company must provide OnlineSecurity to large number of its customers.
6) **Churn w.r.t TotalCharges:**
Graph indicates that as the total charges increase, churn decreases.
7) **Churn w.r.t MonthlyCharges:**
Graph indicates that as the MonthlyCharges increase, churn decreases. However, this was not observed in Python and there might be some error which I'm unable to identify.
8) **Churn w.r.t DeviceProtection:**
Customers who are not provided with DeviceProtection are tend to churn more because there might be a competitor who is providing them with this facility.
9) **Churn w.r.t TechSupport:**
Customers who are not provided with TechSupport are tend to churn more because there might be a competitor who is providing them with this facility.

Wrangling:

In [1]:
```python
#importing basic libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
```

In [2]:
```python
#importing dataset
df = pd.read_csv("churndata.csv")
```

In [3]:
```python
#reading data
df.head()
```

Out[3]:

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

5 rows × 21 columns

In [4]:
```python
#exploring the datatypes of all columns of dataset
df.dtypes
```

Out[4]:
```
customerID          object
gender              object
SeniorCitizen        int64
Partner             object
Dependents          object
tenure               int64
PhoneService        object
MultipleLines       object
InternetService     object
OnlineSecurity      object
OnlineBackup        object
```

In [5]:
```python
#Total charges is an object type, but it should be numeric; hence, I am converting it to numerical type
df.TotalCharges = pd.to_numeric(df.TotalCharges, errors='coerce')
```

In [6]:
```python
#exploring the count of null values in each column
df.isnull().sum()
```

Out[6]:
```
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        11
Churn                0
dtype: int64
```

In [7]:
```python
#only TotalCharges column has 11 values which is approx. 0.16% of the total data. Hence, we can remove rows with null values in
df.dropna(inplace = True)
```

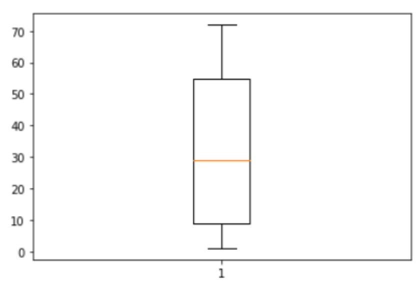
In [8]:
```python
df.isnull().sum()
```

Out[8]:
```
customerID          0
gender              0
SeniorCitizen       0
Partner             0
Dependents          0
tenure              0
```


In [9]:
```python
#dropping CustomerID column because it is of no use in analysis
df.drop(columns= ['customerID'], axis=1, inplace=True)
```

In [10]:
```python
#Plotting box plot for numerical values to identify outliers
plt.boxplot(df.tenure) #tenure
```

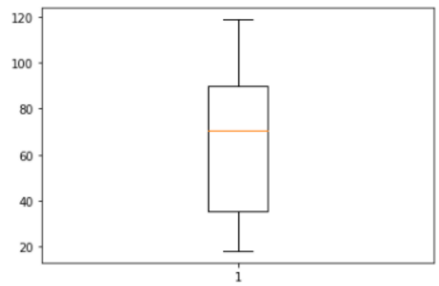
Out[10]:
```
{'whiskers': [<matplotlib.lines.Line2D at 0x1c23a8ede48>,
  <matplotlib.lines.Line2D at 0x1c23a8edf08>],
 'caps': [<matplotlib.lines.Line2D at 0x1c23a8f7f48>,
  <matplotlib.lines.Line2D at 0x1c23a8f7ec8>],
 'boxes': [<matplotlib.lines.Line2D at 0x1c23a8ed288>],
 'medians': [<matplotlib.lines.Line2D at 0x1c23a8fbec8>],
 'fliers': [<matplotlib.lines.Line2D at 0x1c23a8fbfc8>],
 'means': []}
```



In [11]:
```python
plt.boxplot(df.MonthlyCharges) #MonthlyCharges
```

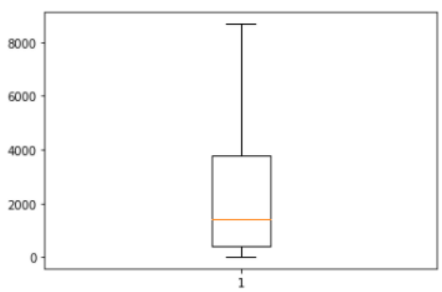
In [11]:
```
plt.boxplot(df.MonthlyCharges) #MonthlyCharges
```

Out[11]:
```
{'whiskers': [<matplotlib.lines.Line2D at 0x1c23a99fec8>,
  <matplotlib.lines.Line2D at 0x1c23a99ffc8>],
 'caps': [<matplotlib.lines.Line2D at 0x1c23a9a4f08>,
  <matplotlib.lines.Line2D at 0x1c23a9a4fc8>],
 'boxes': [<matplotlib.lines.Line2D at 0x1c23a99f648>],
 'medians': [<matplotlib.lines.Line2D at 0x1c23a9a7fc8>],
 'fliers': [<matplotlib.lines.Line2D at 0x1c23a9a7f48>],
 'means': []}
```



In [12]:
```
plt.boxplot(df.TotalCharges) #TotalCharges
```

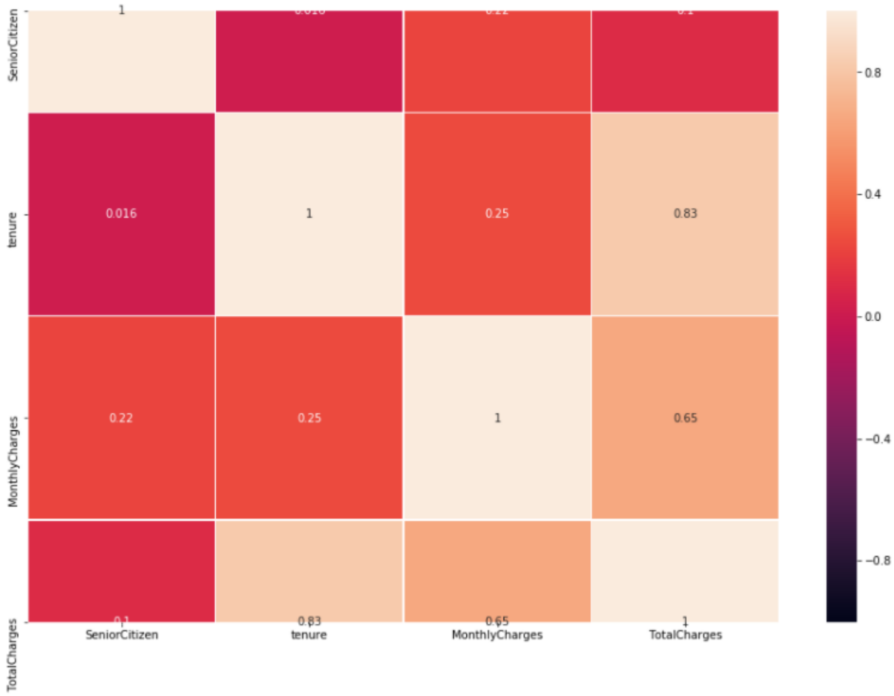
Out[12]:
```
{'whiskers': [<matplotlib.lines.Line2D at 0x1c23aa0df08>,
  <matplotlib.lines.Line2D at 0x1c23aa0dfc8>],
 'caps': [<matplotlib.lines.Line2D at 0x1c23aa12ec8>,
  <matplotlib.lines.Line2D at 0x1c23aa12f88>],
 'boxes': [<matplotlib.lines.Line2D at 0x1c23aa0d4c8>],
 'medians': [<matplotlib.lines.Line2D at 0x1c23aa15e48>],
 'fliers': [<matplotlib.lines.Line2D at 0x1c23aa15f08>],
 'means': []}
```



In [13]:
```
#No outliers are found in any of the numerical column
```

In [14]:
```
#Time to identify the corelation b/w quantitaive data
import seaborn as sns
corrmatrix = df.corr()
f, axis = plt.subplots(figsize =(15, 10))
sns.heatmap(corrmatrix, ax = axis, linewidths = 0.2, vmin=-1, vmax=1, annot=True)
```

Out[14]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1c23c87d888>
```

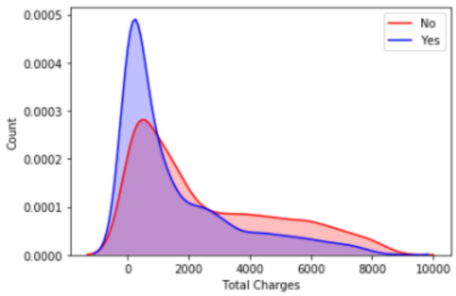


In [15]:
```
#Heatmap shows the strong correlation between tenure & TotaCharges and TotalCharges & MonthlyCharges
```

In [15]:
```
#Heatmap shows the strong correlation between tenure & TotaCharges and TotalCharges & MonthlyCharges
```

In [16]:
```
#Plotting graphs to evaluate the affect of quantitative variables on churn
TC_Churn = sns.kdeplot(df.TotalCharges[(df["Churn"] == "No") ], color="Red", shade = True, label="No")
TC_Churn = sns.kdeplot(df.TotalCharges[(df["Churn"] == "Yes") ], color="Blue", shade= True, label="Yes")
TC_Churn.set_ylabel('Count')
TC_Churn.set_xlabel('Total Charges')
```

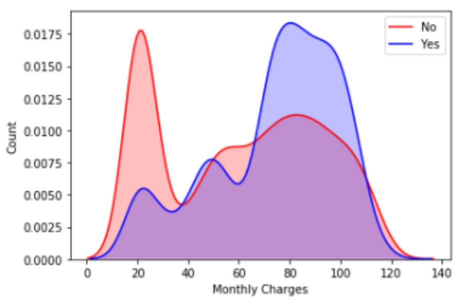
Out[16]:
```
Text(0.5, 0, 'Total Charges')
```



In [17]:
```
#Above line plot shows that highest churn (YES) is observed at low total charges.
```

In [18]:
```
MC_Churn = sns.kdeplot(df.MonthlyCharges[(df["Churn"] == "No") ], color="Red", shade = True, label="No")
MC_Churn = sns.kdeplot(df.MonthlyCharges[(df["Churn"] == "Yes") ], color="Blue", shade= True, label="Yes")
MC_Churn.set_ylabel('Count')
MC_Churn.set_xlabel('Monthly Charges')
```

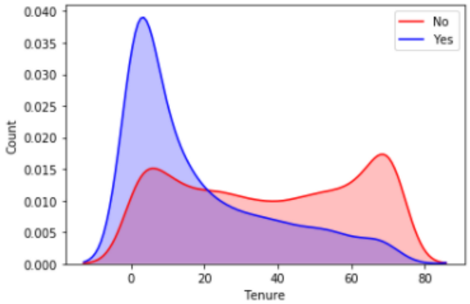
Out[18]:
```
Text(0.5, 0, 'Monthly Charges')
```



In [19]:
```
#Above line plot shows that as the monthly charges increase, churn also increases(YES).
```

In [20]:
```
Tenure_Churn = sns.kdeplot(df.tenure[(df["Churn"] == "No") ], color="Red", shade = True, label="No")
Tenure_Churn = sns.kdeplot(df.tenure[(df["Churn"] == "Yes") ], color="Blue", shade= True, label="Yes")
Tenure_Churn.set_ylabel('Count')
Tenure_Churn.set_xlabel('Tenure')
```

Out[20]: Text(0.5, 0, 'Tenure')

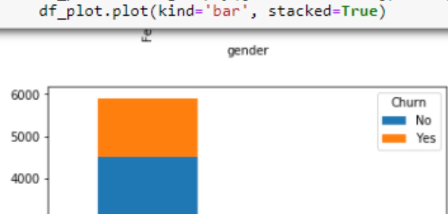


In [21]:
```
#Above line plot indicates that churn(YES) is high at less number of tenures
```

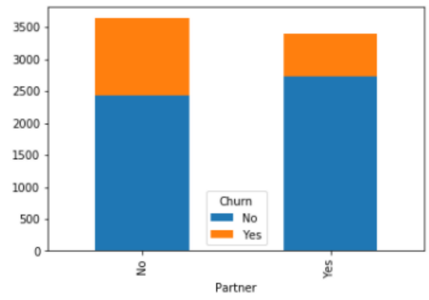
In [41]:
```
#Replacing values of SeniorCitizen with Yes and No
df['SeniorCitizen'].replace(1,'Yes', inplace=True)
df['SeniorCitizen'].replace(0,'No', inplace=True)


#keeping categorical columns only
df2 = df.drop(columns=[ 'tenure', 'MonthlyCharges', 'TotalCharges'])

#plotting graphs to understand the relation b/w categorical columns and churn
for i, value in enumerate(df2):
    df_plot = df2.groupby(['Churn', value]).size().reset_index().pivot(columns='Churn', index=value, values=0)
    df_plot.plot(kind='bar', stacked=True)
```



```
for i, value in enumerate(df2):
    df_plot = df2.groupby(['Churn', value]).size().reset_index().pivot(columns='Churn', index=value, values=0)
    df_plot.plot(kind='bar', stacked=True)
```



In [42]:
```
#Important insights that can be obtained from above bar plots are :
# 1) young people tends to churn more than senior citizen because they tend to avail different oragnization's offer
# 2) ppl without dependents & partners tend to churn more because they are on their own only and can get involved with other oofe
# 3) ppl with fiber optic cable tend to churn more
# 4) little to no onlinesecurity, OnlineBackup, DeviceProtection & TechSupport results in customer churning
# 5) Churn is same with & without StreamingTV and StreamingMovies (neglecting no internet service for now)
# 6) Churn is high for month-to-month contracts
# 7) Churn is little high for paperlessBilling
# 8) ppl who avail Electronic check payment method tend to churn more than ppl with any other payment method
# But if we look on each category individually then customer churn(YES) is still less than the customers who are not churning(NO
#the following attributes where churn is significant: TechSupport, DeviceProtection, OnlineSecurity, InternetService, Contract, I
```

In [45]:
```
df.to_csv("TelcoChurn.csv")
```

Note: Refer to jupyter file for further reference.