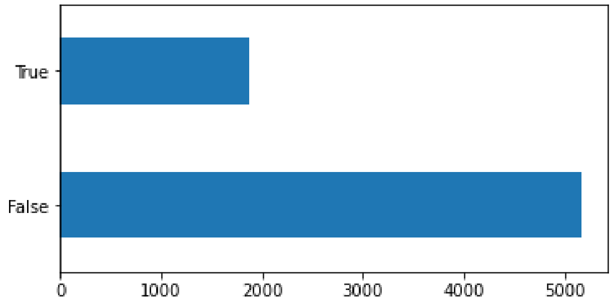
Assignment #2

Introduction:

The customer *churn rate* describes the rate at which customers leave a company or a service. For many organizations, this is a very important factor or behavior to model and understand. It is often more expensive to obtain a new customer base rather than to keep the existing one on board. Therefore it is worthwhile predicting which customers may want to leave and trying to keep them on board as customers.

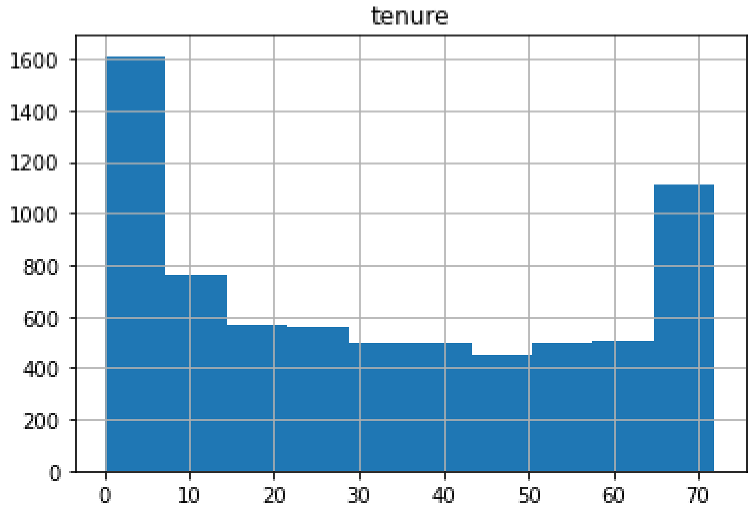
Plots:

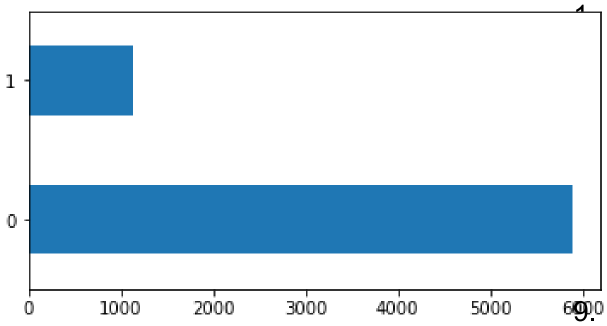


Churn:

Customers who do not churn: 5174: 73.46%

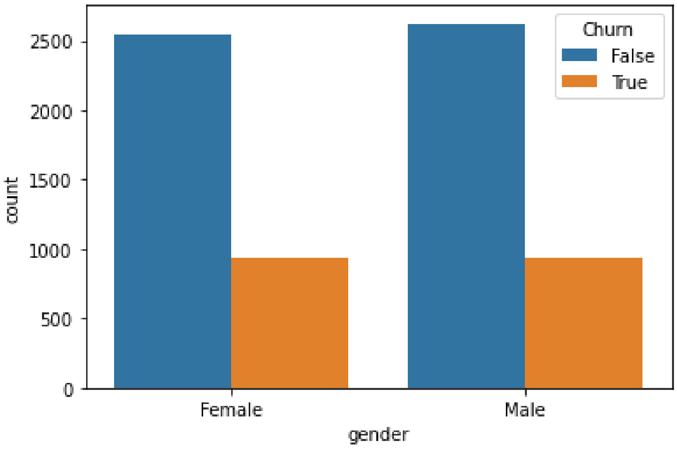
Customers who churn: 1869 : 26.54%





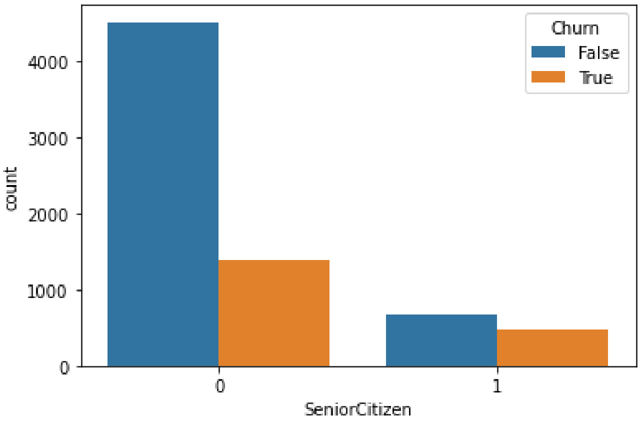
- Not senior citizen: 5901 = 83.785319%
- Senior citizen: 1142 =16.214681%

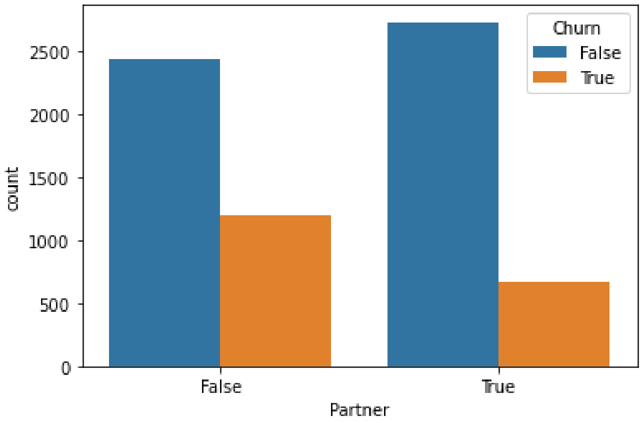
Name: SeniorCitizen, dtype: int64

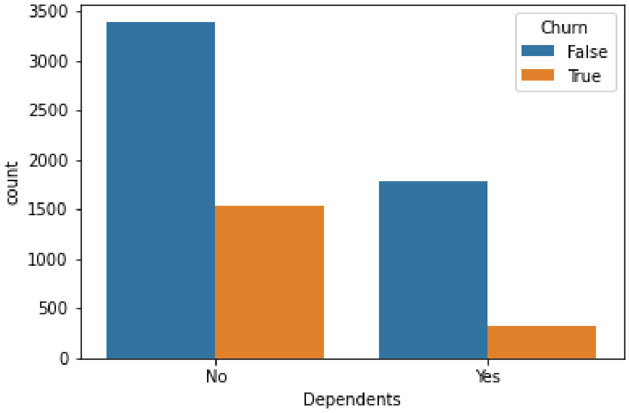


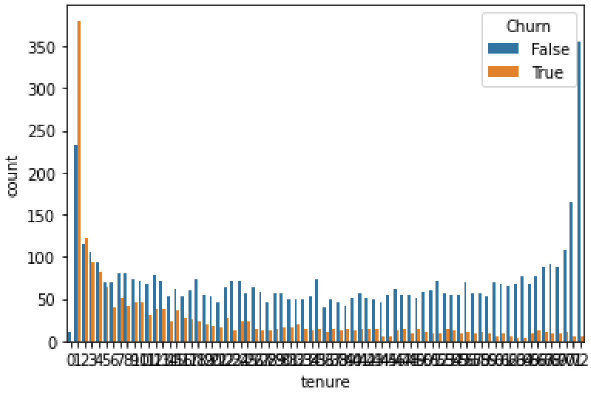
Gender wise churn rate:
Approximately 50-50

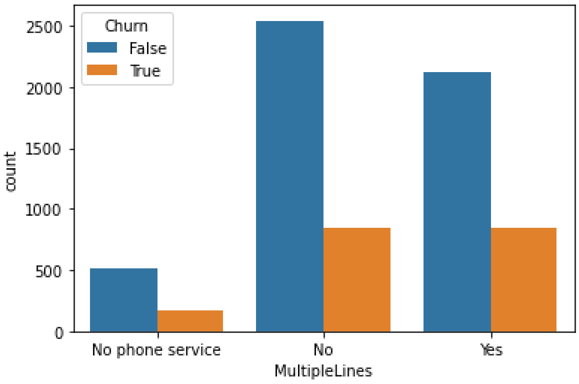
Other similar plots:

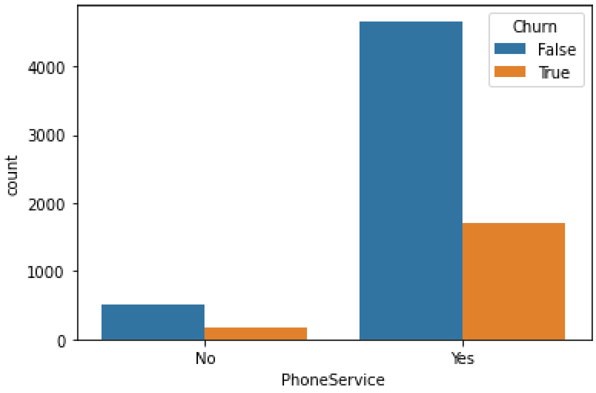


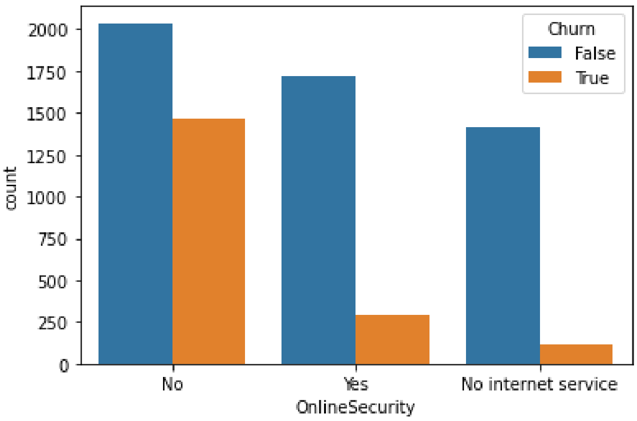


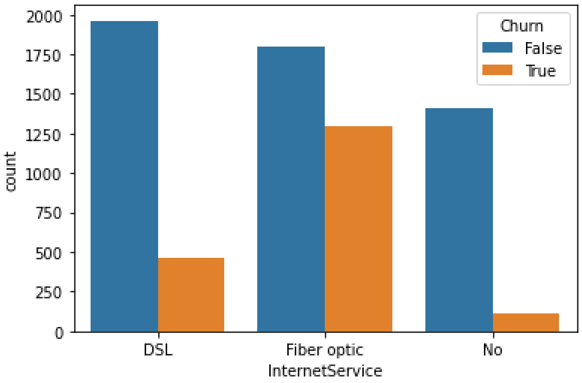


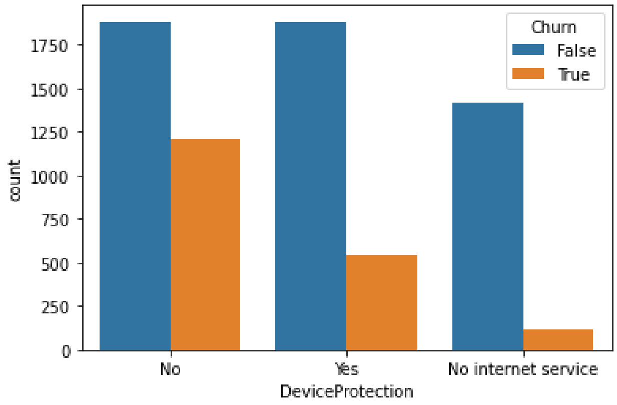


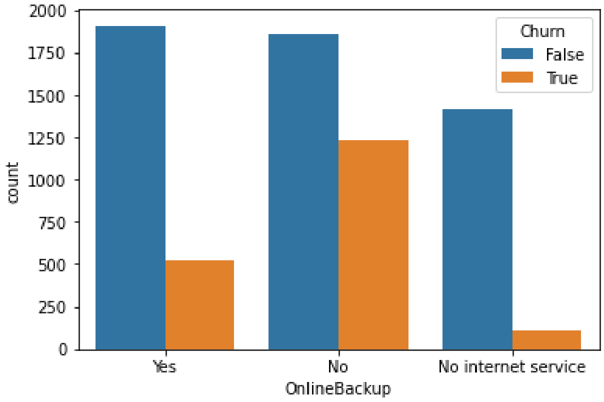


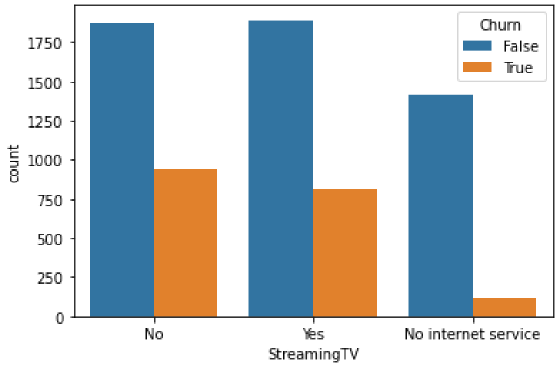


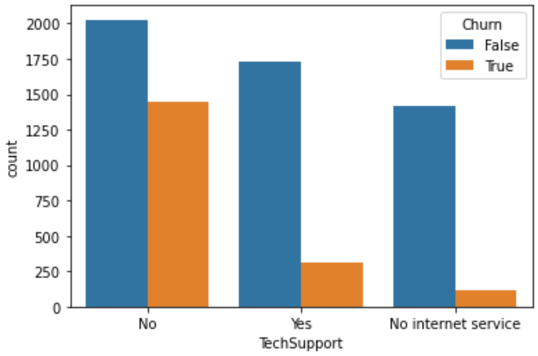


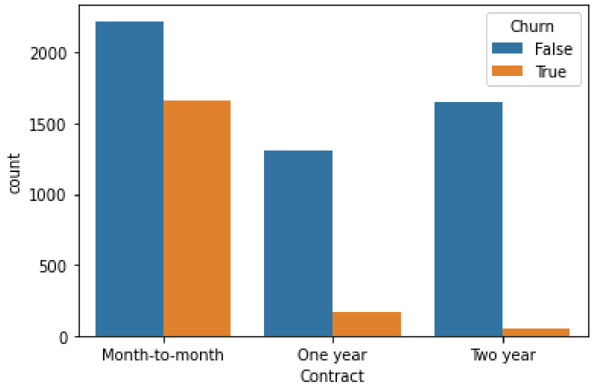


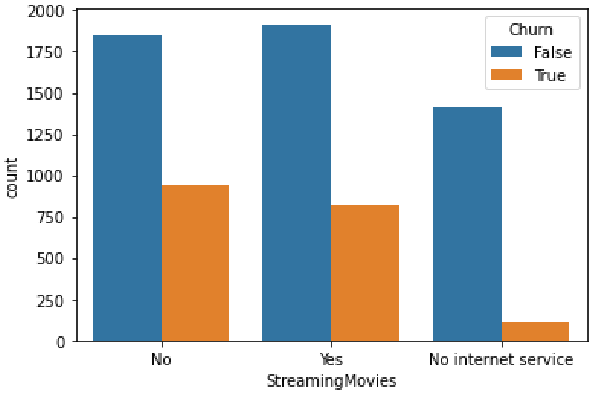

count
0
250
500
750
1000
1250
1500
1750
No
Yes
No internet service
DeviceProtection
Churn
False
True

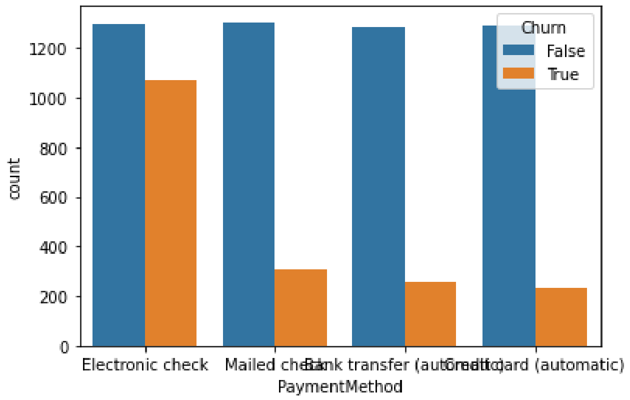

count
0
250
500
750
1000
1250
1500
1750
2000
Yes
No
No internet service
OnlineBackup
Churn
False
True

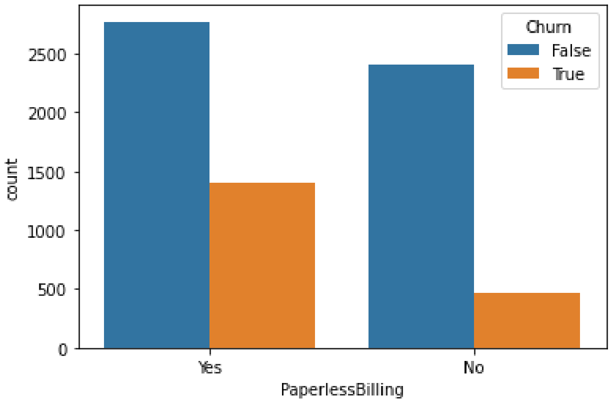

count
0
250
500
750
1000
1250
1500
1750
No
Yes
No internet service
StreamingTV
Churn
False
True

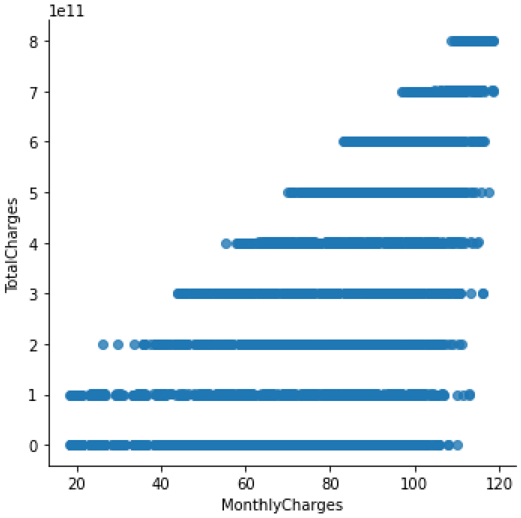

count
0
250
500
750
1000
1250
1500
1750
2000
No
Yes
No internet service
TechSupport
Churn
False
True

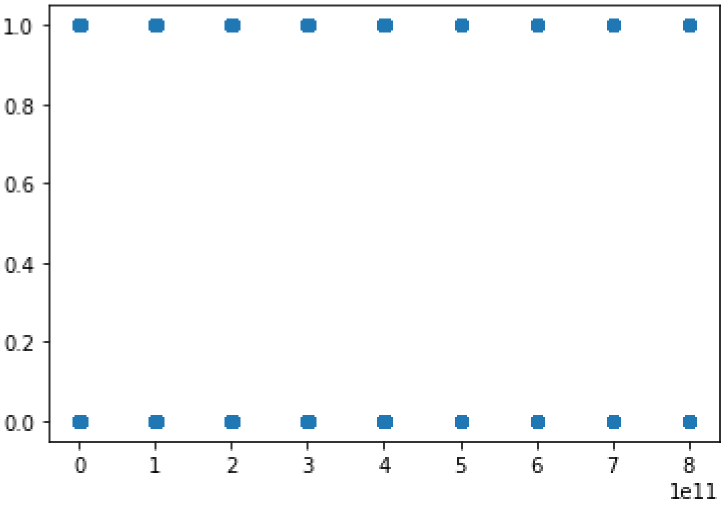

count
0
500
1000
1500
2000
Month-to-month
One year
Two year
Contract
Churn
False
True


count
0
250
500
750
1000
1250
1500
1750
2000
No
Yes
No internet service
StreamingMovies
Churn
False
True


count
0
200
400
600
800
1000
1200
Electronic check
Mailed check
Bank transfer (automatic)
Credit card (automatic)
PaymentMethod
Churn
False
True


count
0
500
1000
1500
2000
2500
Yes
No
PaperlessBilling
Churn
False
True

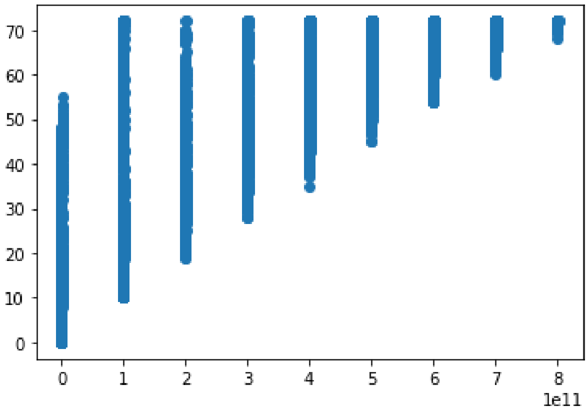


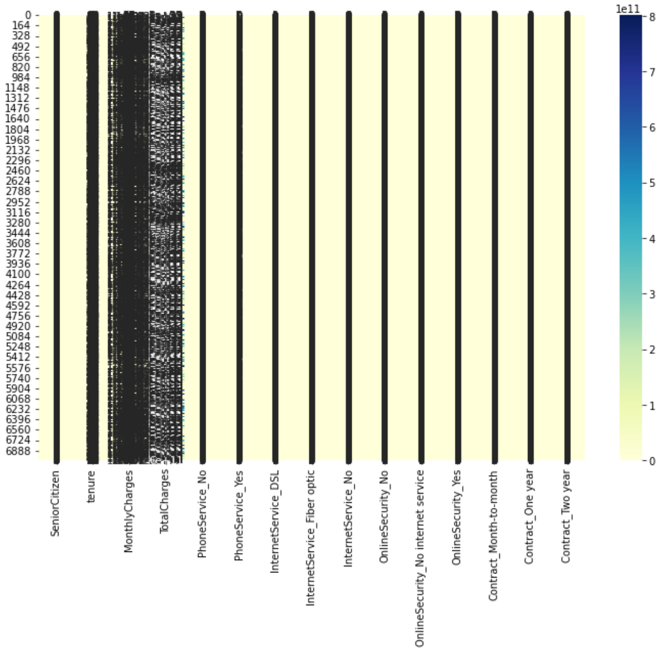


.

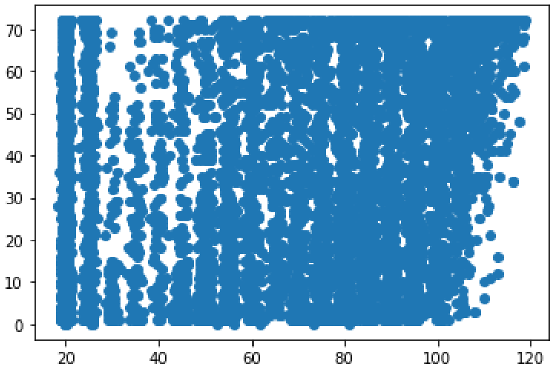
Monthly charges are increasing the total charges also increase, which shows a positive correlation.

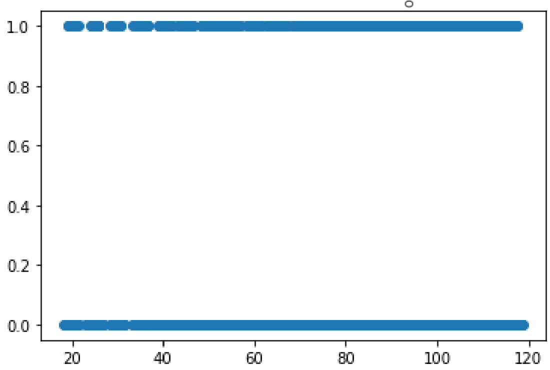
TotalCharges vs. 'SeniorCitizen



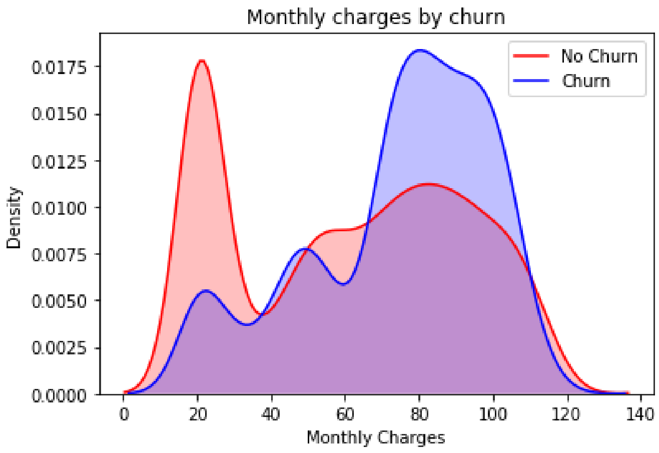


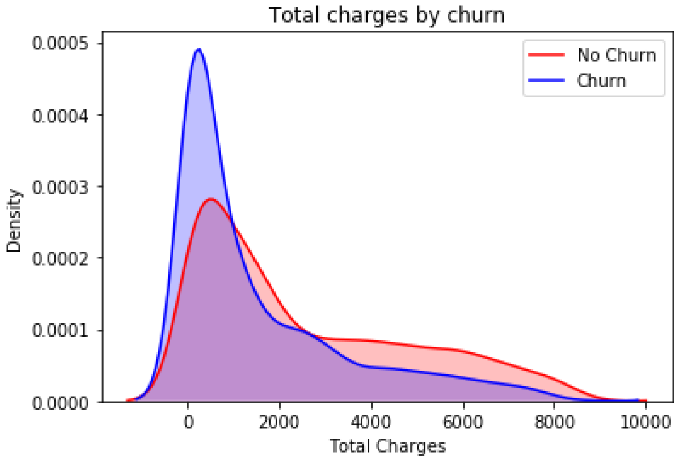
TotalCharges vs. tenure





MonthlyCharges vs tenure

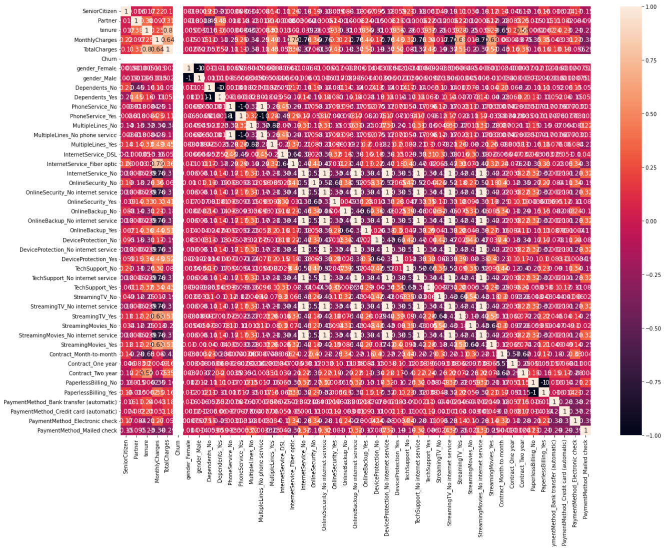


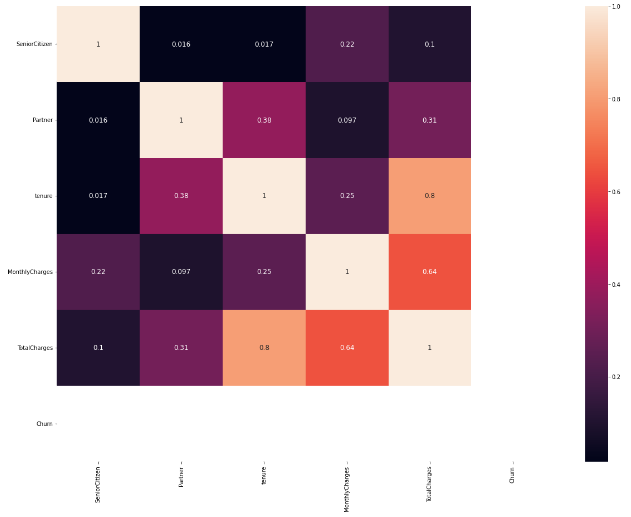


More churn with high monthly charges than low.

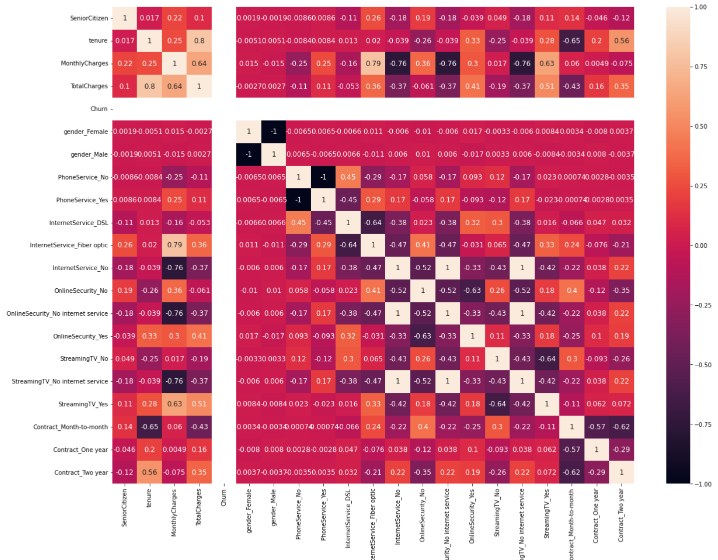
More churn is there with a lower value of total charge

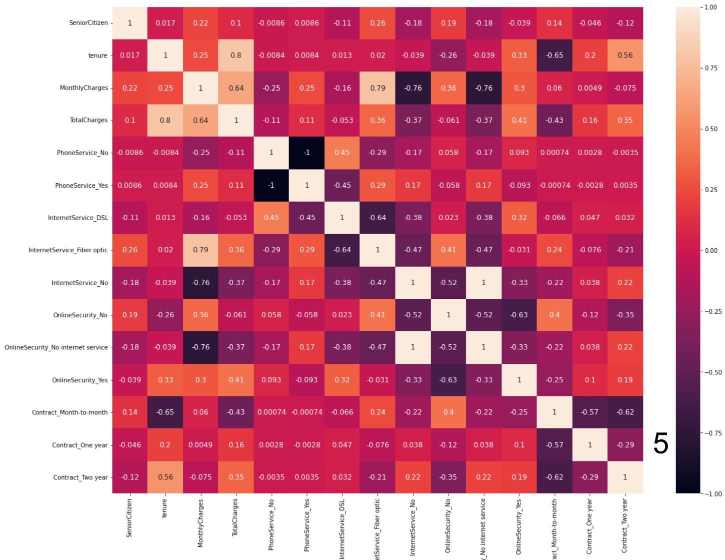
Correlation plots:



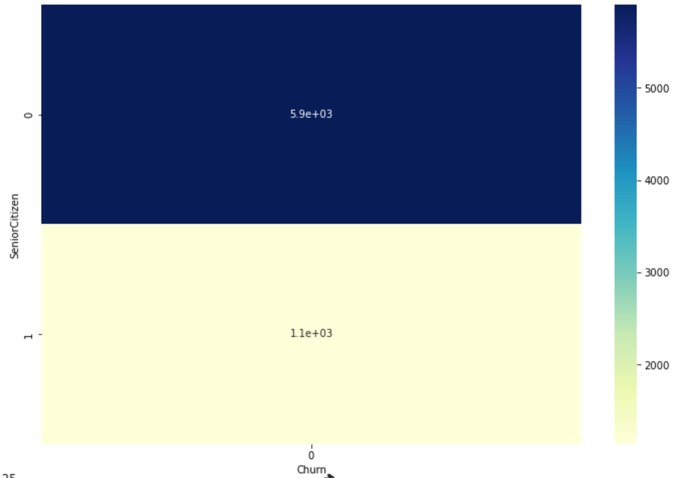


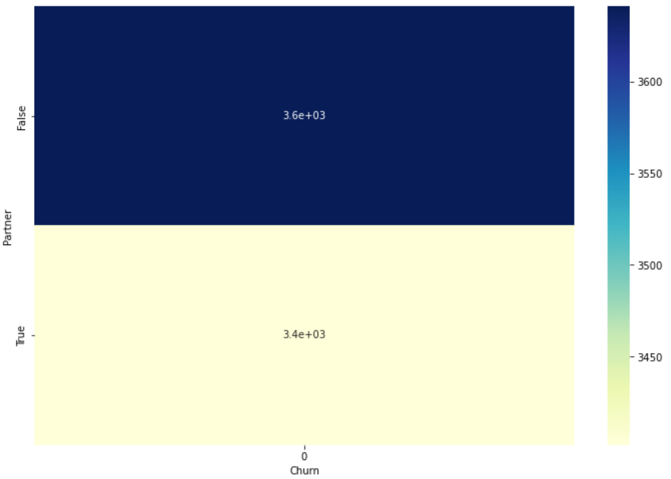
Reducing the attributes to get a better view of the correlations and get filtered data

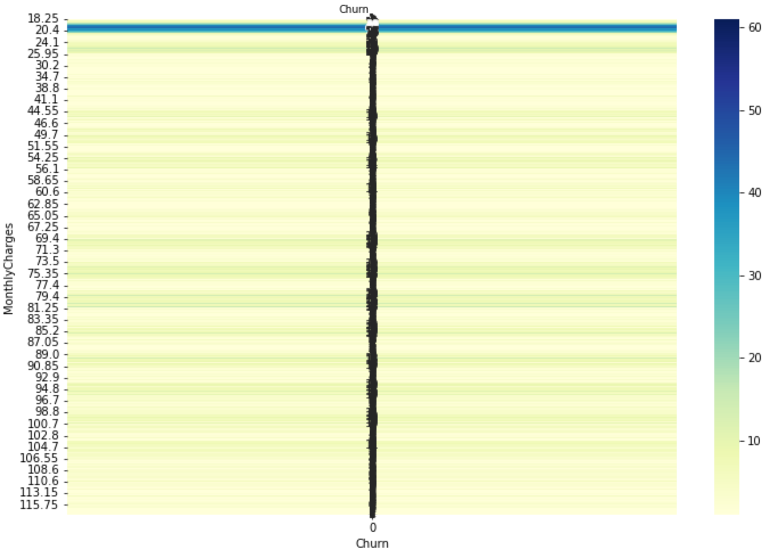


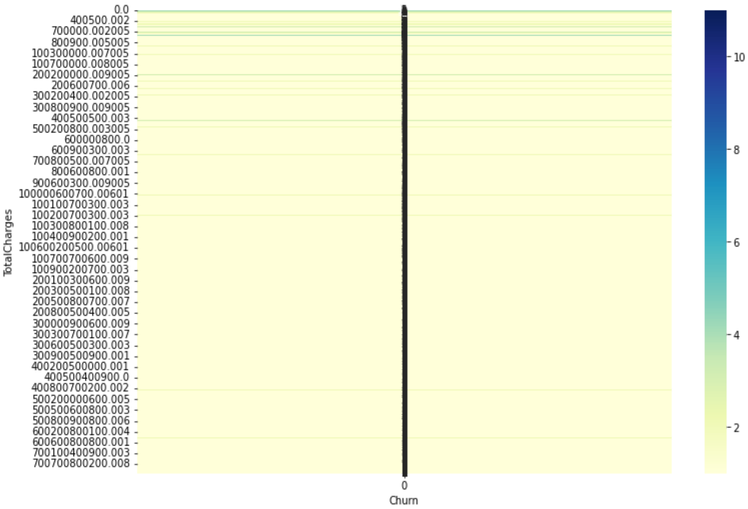


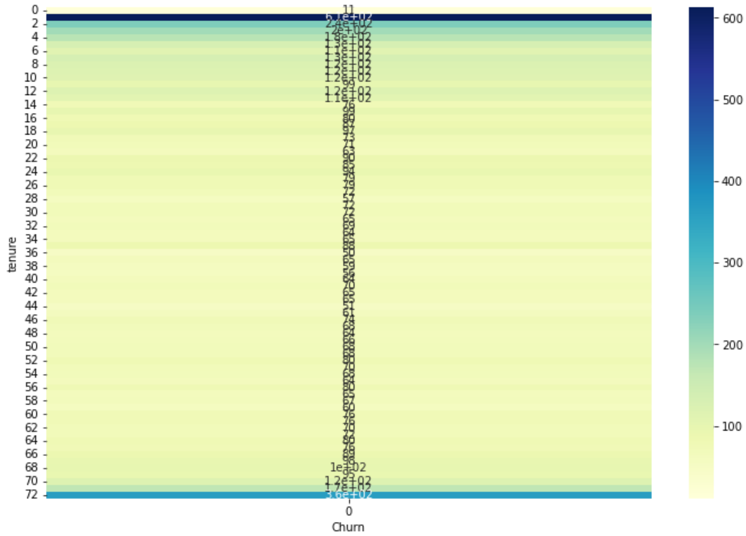
Chi-square tests:

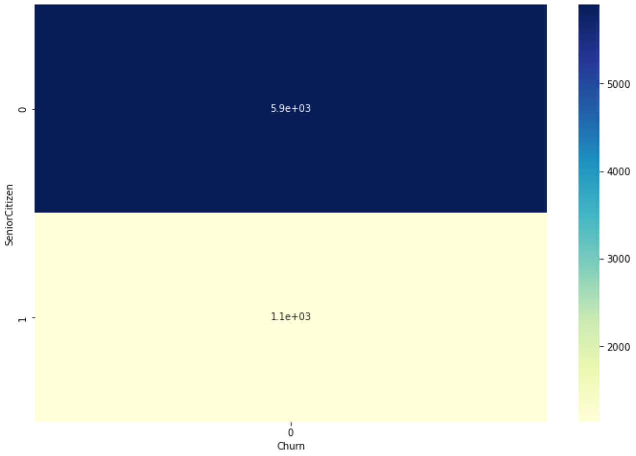












Observations based on plots and coded data:

- Gender distribution is relatively even.
- Roughly 2000 customers churned, and 5000 customers retained.
- Senior citizens are three times less likely to churn
- Partners are almost two times less likely to leave
- Customers without dependents are four times more likely to churn
- The dataset is imbalanced, with the majority of customers being active.
- There is multicollinearity between Monthly Charges and Total Charges.
- Most of the customers in the dataset are younger people.
- There are lots of customers with long and short tenures.
- Monthly charges have a lot of low charges and, aside from that, are reasonably normally distributed.
- The majority of customers that cancel their subscription have Phone Service-enabled.
- Customers with Internet Service as *Fiber-Optic* are more likely to cancel than those who have *DSL*
- Customers that do not have Online Security, Device Protection, Online Backup, and Tech Support services enabled are more likely to leave
- The strongest positive correlation with the target features is Monthly Charges and Age, while the negative correlation is with Partners, Dependents, and Tenure.
- There are a lot of new customers in the organization, followed by a loyal customer base above 70 months old.
- Customers with a month-to-month connection have a high probability of churning if they have subscribed to pay via electronic checks.