IME672A: DATA MINING AND KNOWLEDGE DISCOVERY
Customer Churn Rate in Telecom Industry

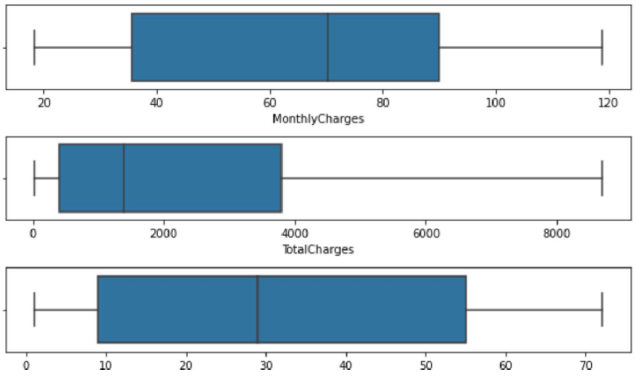
Submitted by: Ayushi Singh, 200258

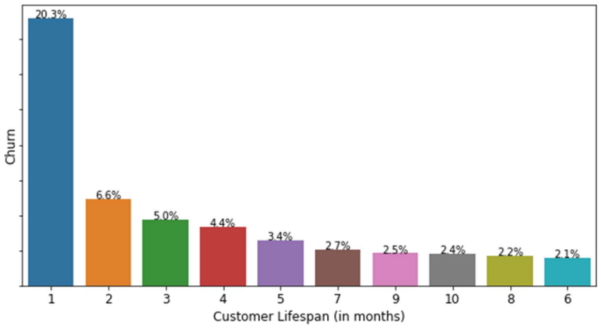
The provided data is from a Telecom Company which is used by them to predict which customers are most likely to switch/quit their services, i.e., to analyse the 'churn rate'. This dataset contains a total of 7,043 customers and 21 attributes, which includes some personal characteristics such as 'Gender', 'Senior Citizen'; services applied, and contract details. Out of the 21 attributes, 3 are numeric and 18 are categorical. I further converted 'TotalCharges' to numeric. 'customerID' does not provide any insights therefore I dropped it.

Out of the entries, 5,174 are active customers and 1,869 are churned, which demonstrates that the dataset is highly unbalanced. Our analysis would revolve around the feature 'Churn'. Also, 11 entries were missing for the feature 'TotalCharges', therefore we can either replace them with the median value or drop these rows since 11 is a very small number for our dataset. I have dropped these rows in my analysis.

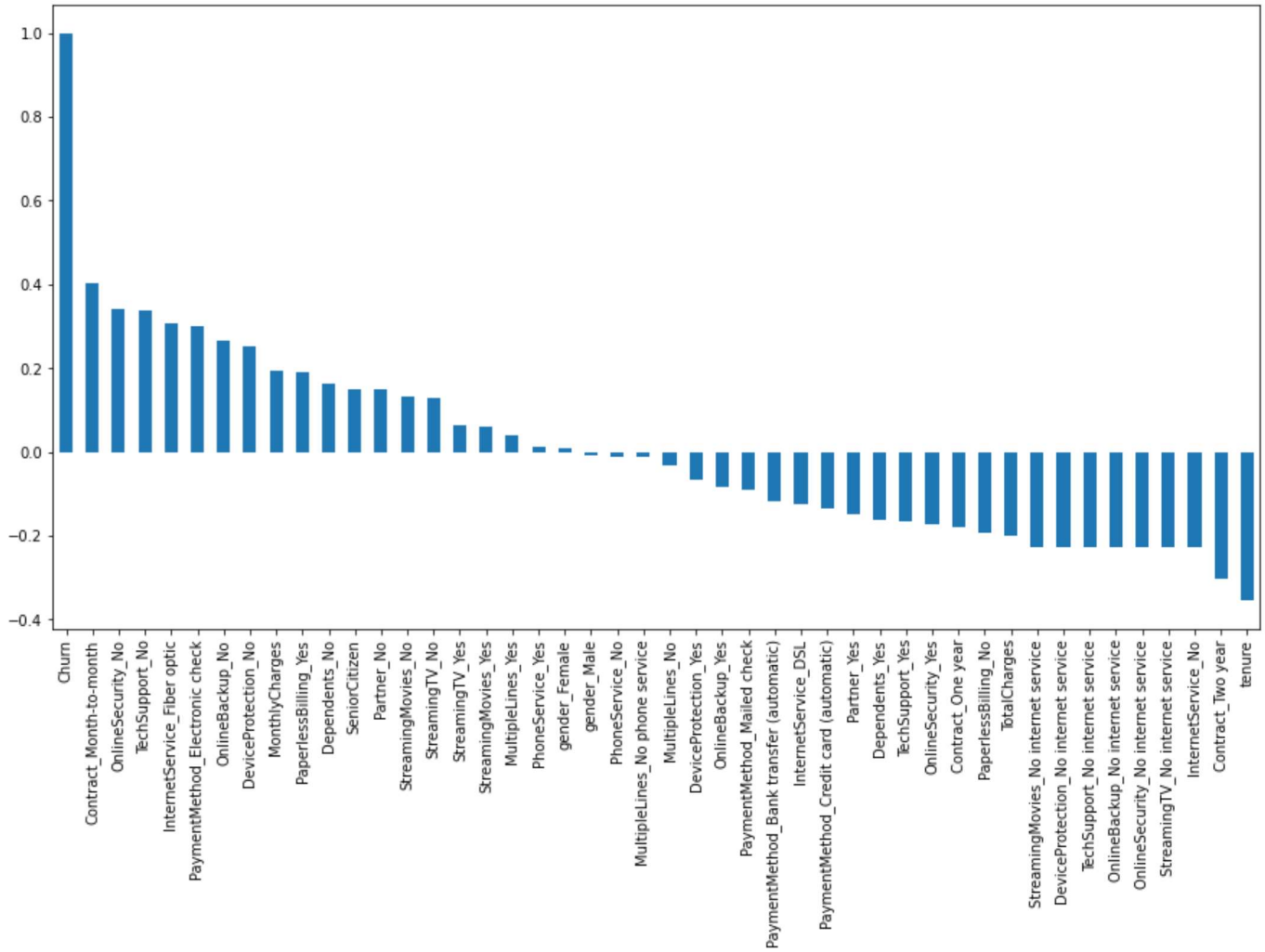
First, I analysed the numerical variables statistically, from which I got the following statistical measures.

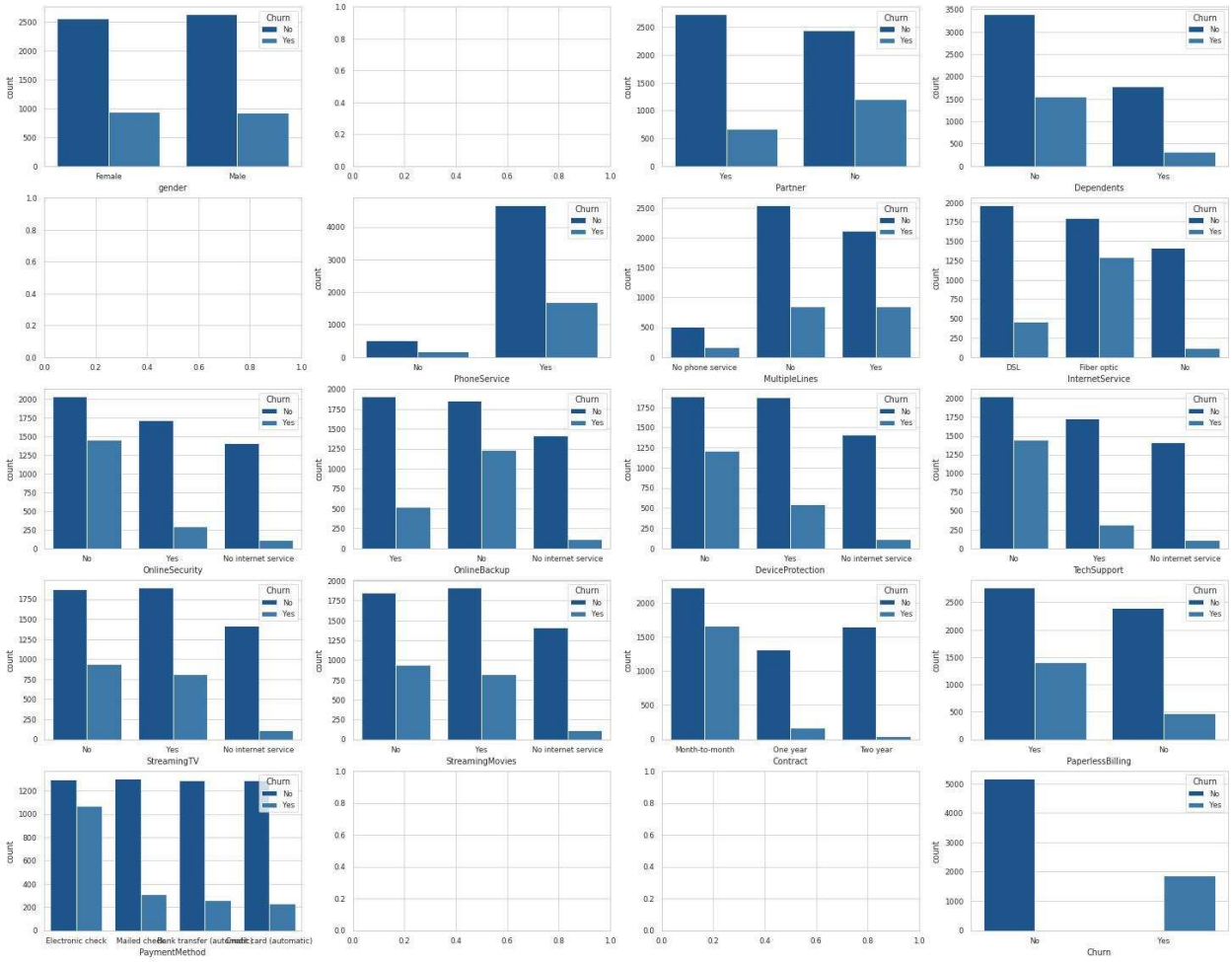
	SeniorCitizen	Tenure	MonthlyCharges
count	7043.000000	7043.00000	7043.000000
mean	0.162147	32.371149	64.761692
std	0.368612	24.559481	30.090047
min	0.000000	0.000000	18.250000
25%	0.000000	9.000000	35.500000
50%	0.000000	29.000000	70.350000
75%	0.000000	55.000000	89.850000
max	1.000000	72.000000	118.750000





Plotting the correlation of churn with other attributes we get the following plot from which we can observe that service and contract details affect churn rate the most, whereas personal characteristics such as gender does not play a role in determining.

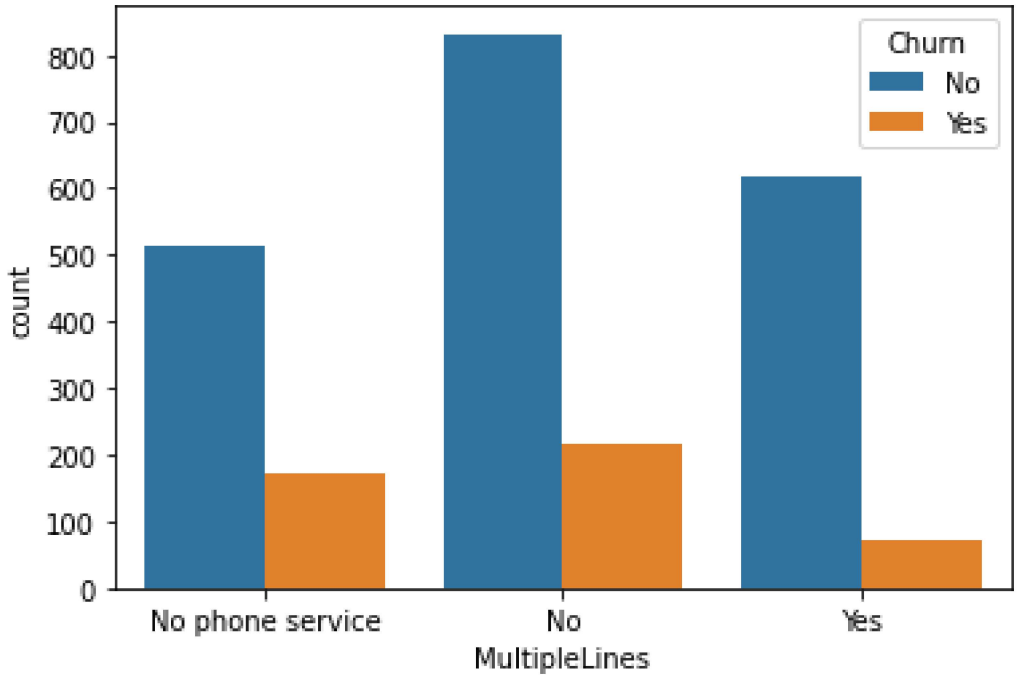


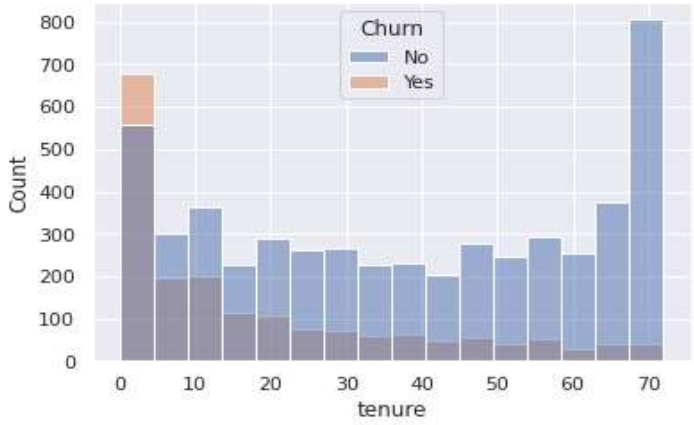


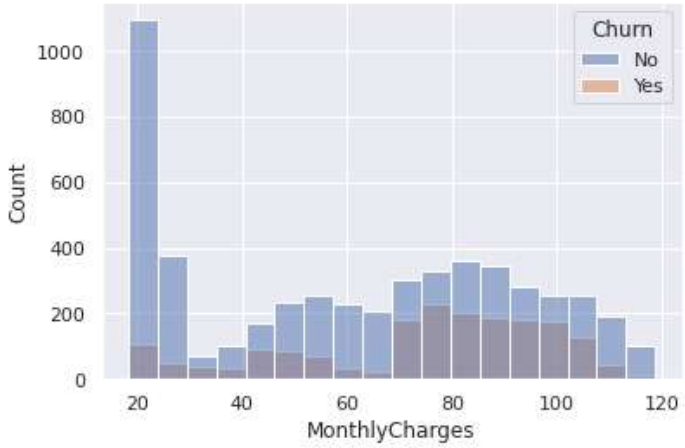
Observation from the above plot:

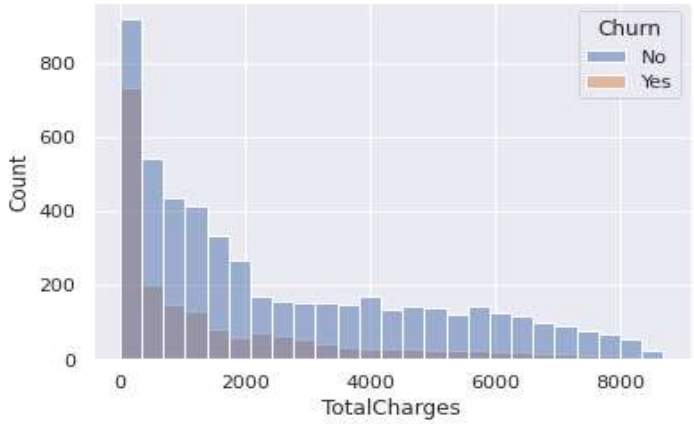
- In the given data, around 71% people **didn't churn**
- Churn distribution is almost similar for both genders
- Electronic payment method has more churn rate as compared to other payment methods
- Month-to-Month contract has higher churn rate than other contracts. It is logical too, cause once you have bought the service for an year or two, people generally do not prefer changing the service due to the hassle it might cause.
- Customer with no Internet service (wherever this value is present) have a very low churn rate. (Maybe because they are not used to the modern technology, and do not prefer changing services)

- People with device protection and Tech-Support (Yes) have less churn rate than people with no device protection or Tech-Support (No), indicating that people are satisfied with the services provided
- Around 650 people didn't take phone service, number is relatively small compared to the total size of the dataset and won't be analyzed.
- People with dependencies or partners have comparatively less churn rate (Around 12.5% for dependencies and 69.5% for partners), than people with no partners or dependencies.
- Customers with no internet services are generally from rural area or want to use their service for just calling or other purpose, thus leaving a very small margin for dissatisfaction and changing the service.
- Another interesting pattern in internet-services is customer with DSL have relatively very less churn rate as compared to Fiber-optics, indicating dissatisfaction with the latter.
- A general patter can be observed: Customers who didn't take services like Online-Security, Online-Backup, Device-Support, Tech-Support, have higher churn rate.



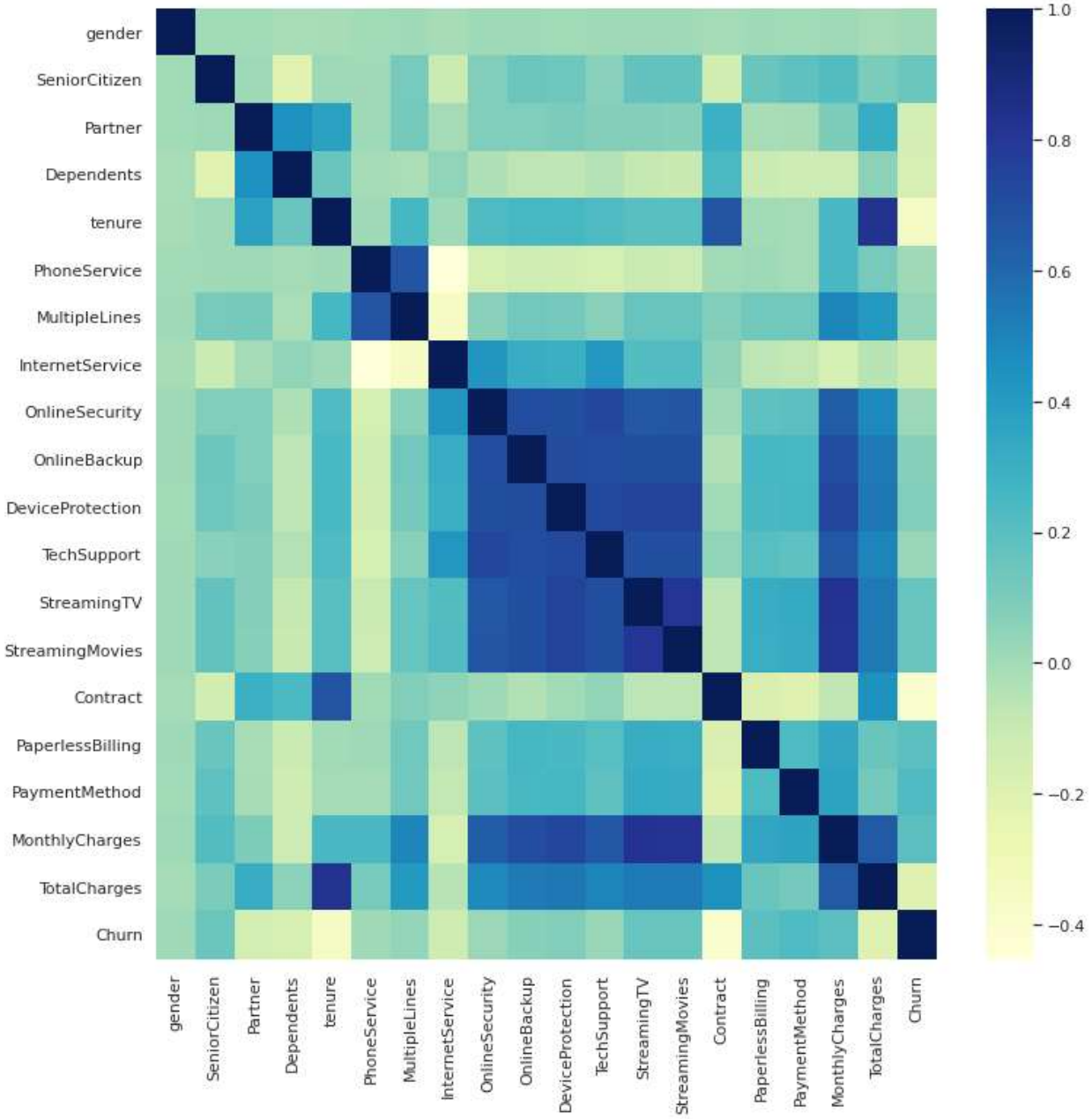






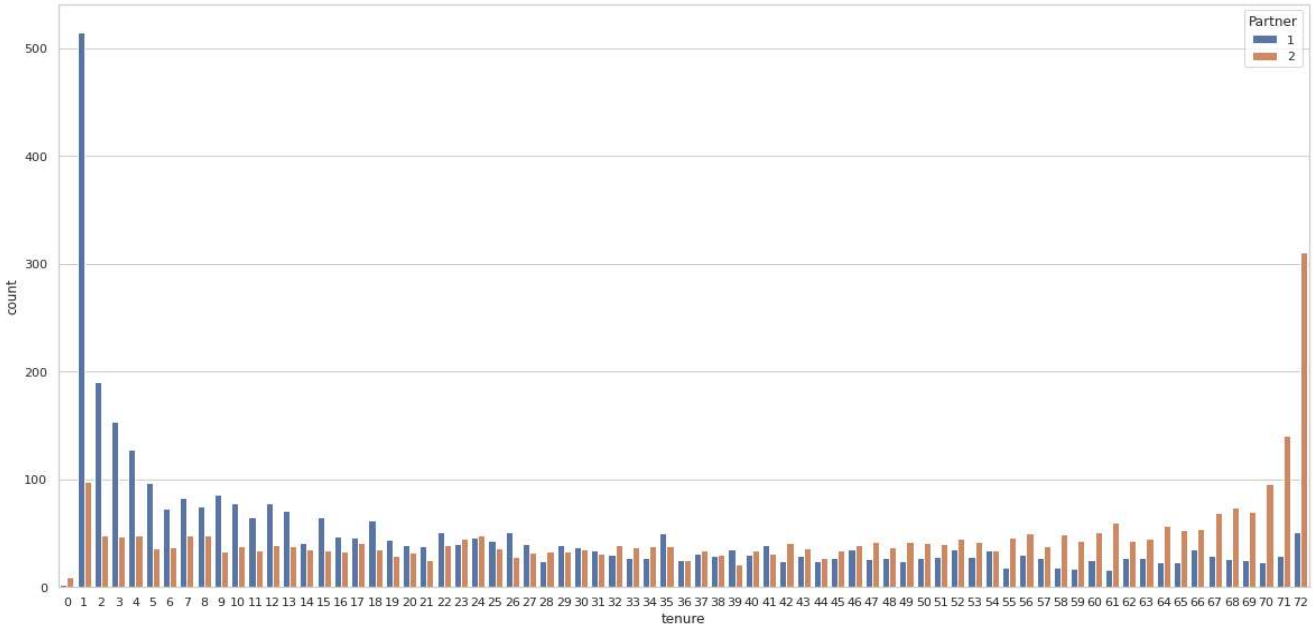
Observations from the above plots:

- People with less value of tenure are more likely to churn.
- Customer's whose monthly bill is less (<= 70) or among the greatest (>=110) are less likely to churn, than customers in the middle range.
- Senior citizens are more likely to churn.

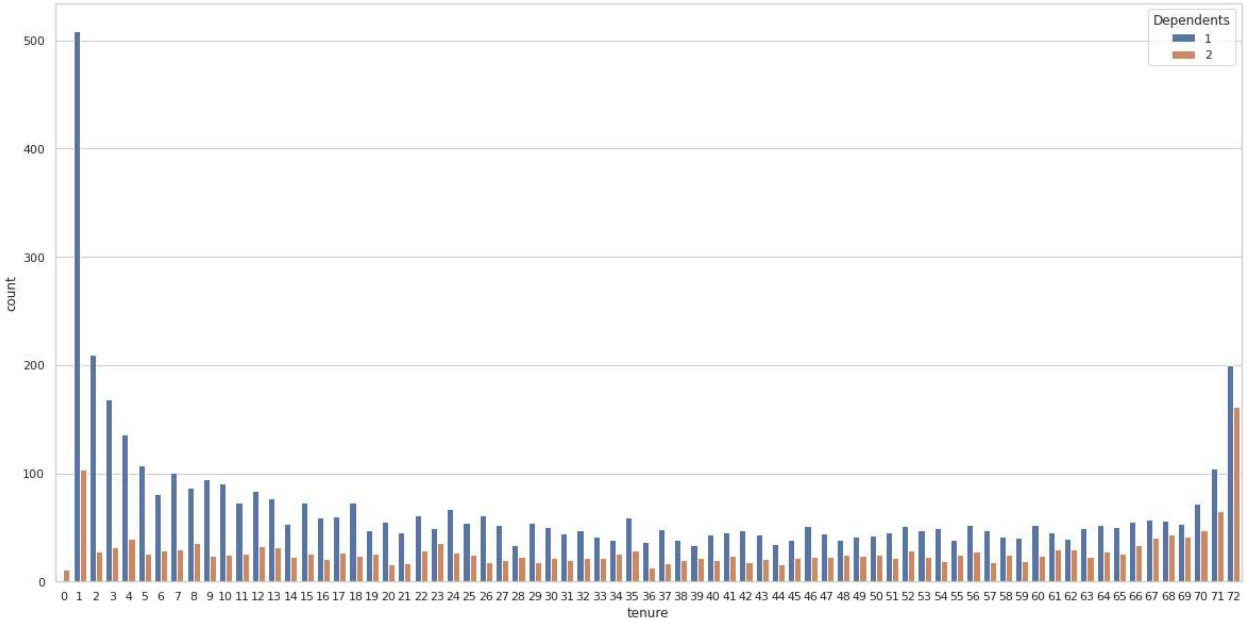


Observations from the above heatmap:

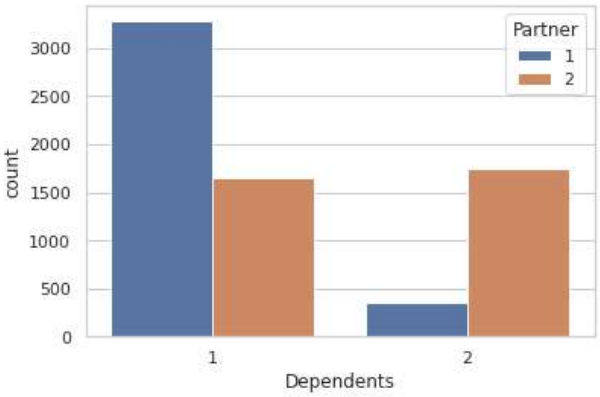
- Contract, Total Charges and Tenure are highly correlated
- Monthly charges increases when customers bought other services (Tech-Support, Online backup, Streaming, etc.)
- Most Senior citizens have no dependents
- All the offered services are correlated, we can just replace them with one attribute Services
- Tenure and Churn are negatively correlated as expected

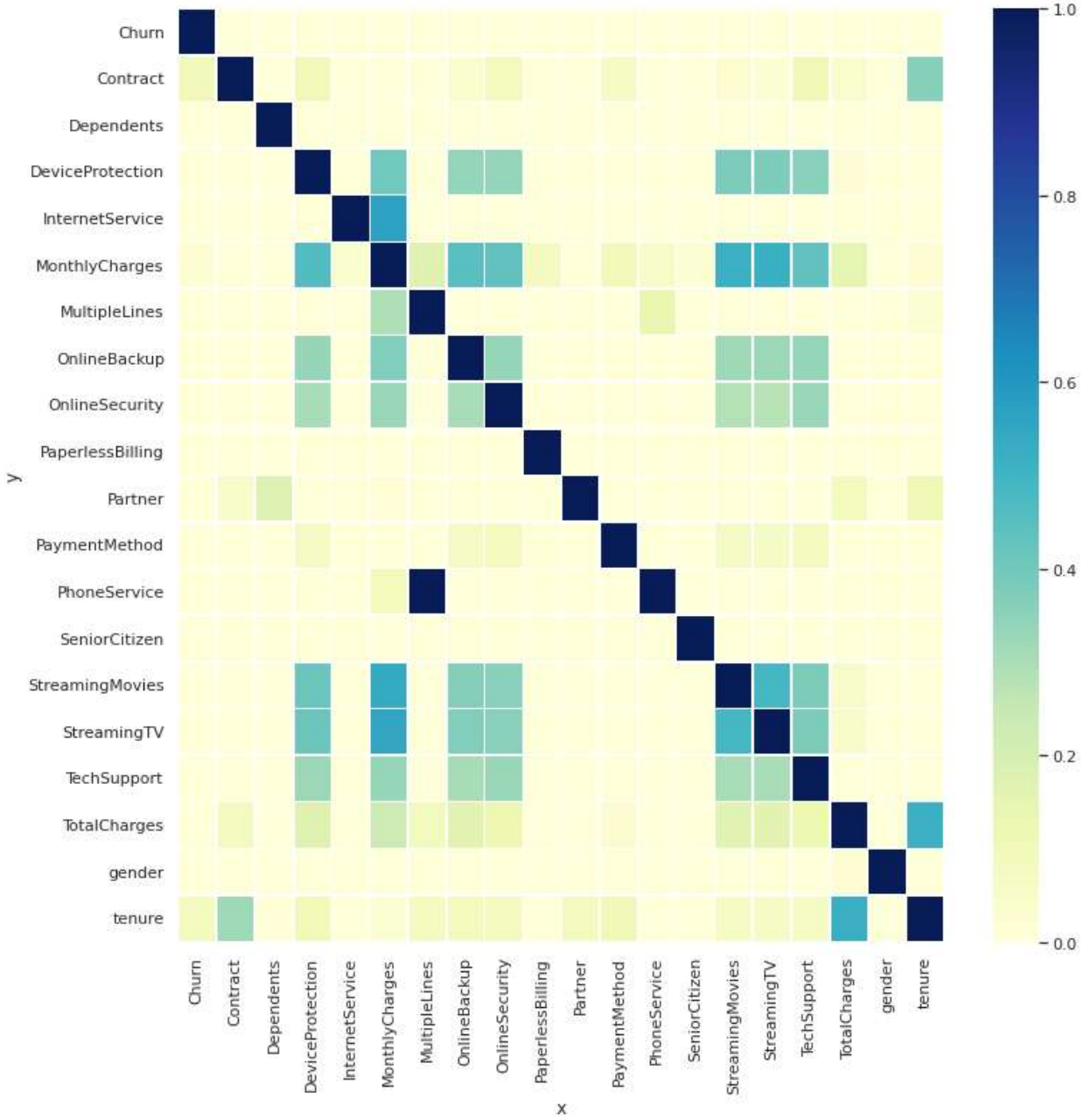


If customers have partner, then their tenure is more likely to be greater than to the ones without a partner.



Customers with no dependent are likely to have less tenure.


Partner
1
2
count
3000
2500
2000
1500
1000
500
0
1
2
Dependents


Churn
Contract
Dependents
DeviceProtection
InternetService
MonthlyCharges
MultipleLines
OnlineBackup
OnlineSecurity
PaperlessBilling
Partner
PaymentMethod
PhoneService
SeniorCitizen
StreamingMovies
StreamingTV
TechSupport
TotalCharges
gender
tenure
y
x
1.0
0.8
0.6
0.4
0.2
0.0

Observation from the above plot:

- Services are good predictor of monthly charges.
- Total charges and tenure are good predictors of each other
- Dependents can predict Partners, but the opposite is not True
- Most relation observed are among services provided, charges (monthly and tenure