Documentation

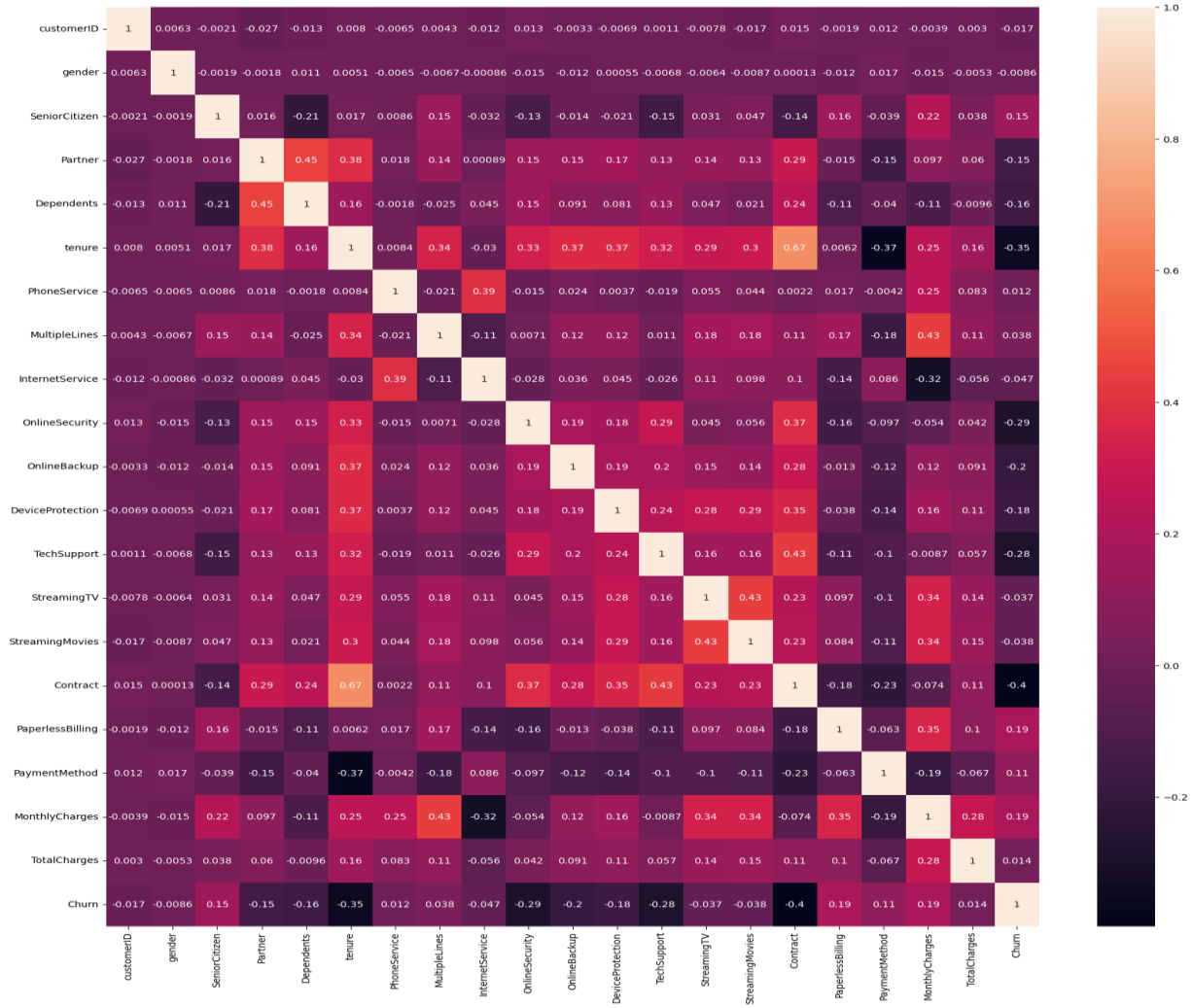
Objective:
The primary objective of this project is to develop a predictive model that can identify customers at risk of churning, enabling the company to take proactive measures to retain them.

Data Collection and Preprocessing:

The dataset we have used here is downloaded from Kaggle named as telco-customer-churn. Here we have implemented label Encoder for smooth analysis of the dataset, fortunately no null values are present in the dataset so simple imputer is not used here. Our dataset consists of 7043 rows and 21 column.

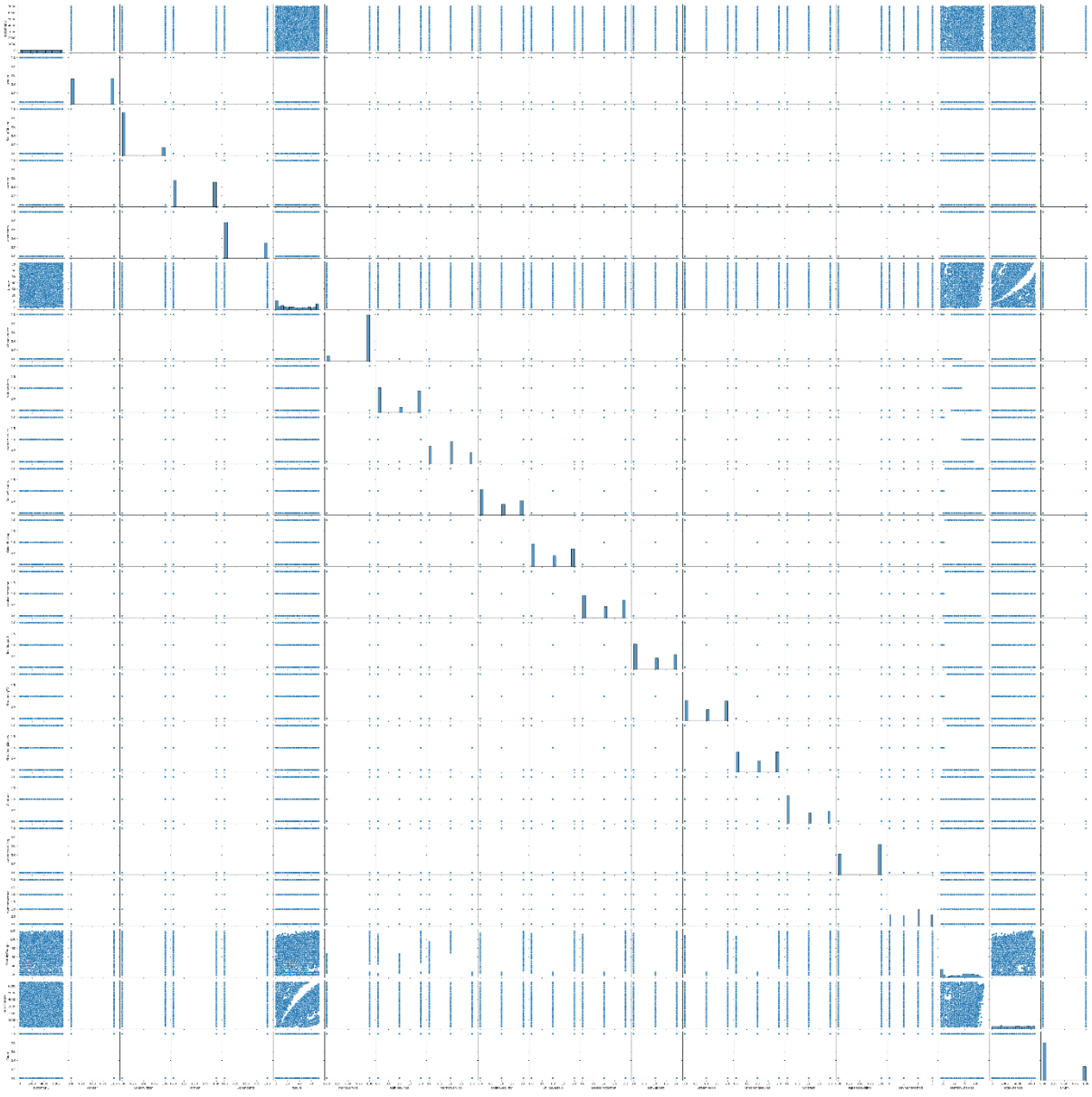
EDA:

In this step we have visualize the Data . using seaborn library. Our first plot is heatmap which has provided us the correlation in between the attributes present in the dataset . here only by observing the values we can find which attributes are actually affecting the churn process.

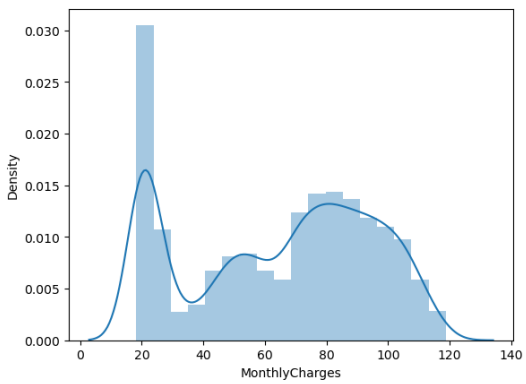


According to the observation it has been found that gender,Partner,tenure,Internet service,online security,online backup etc has negative relation with the churn attribute.

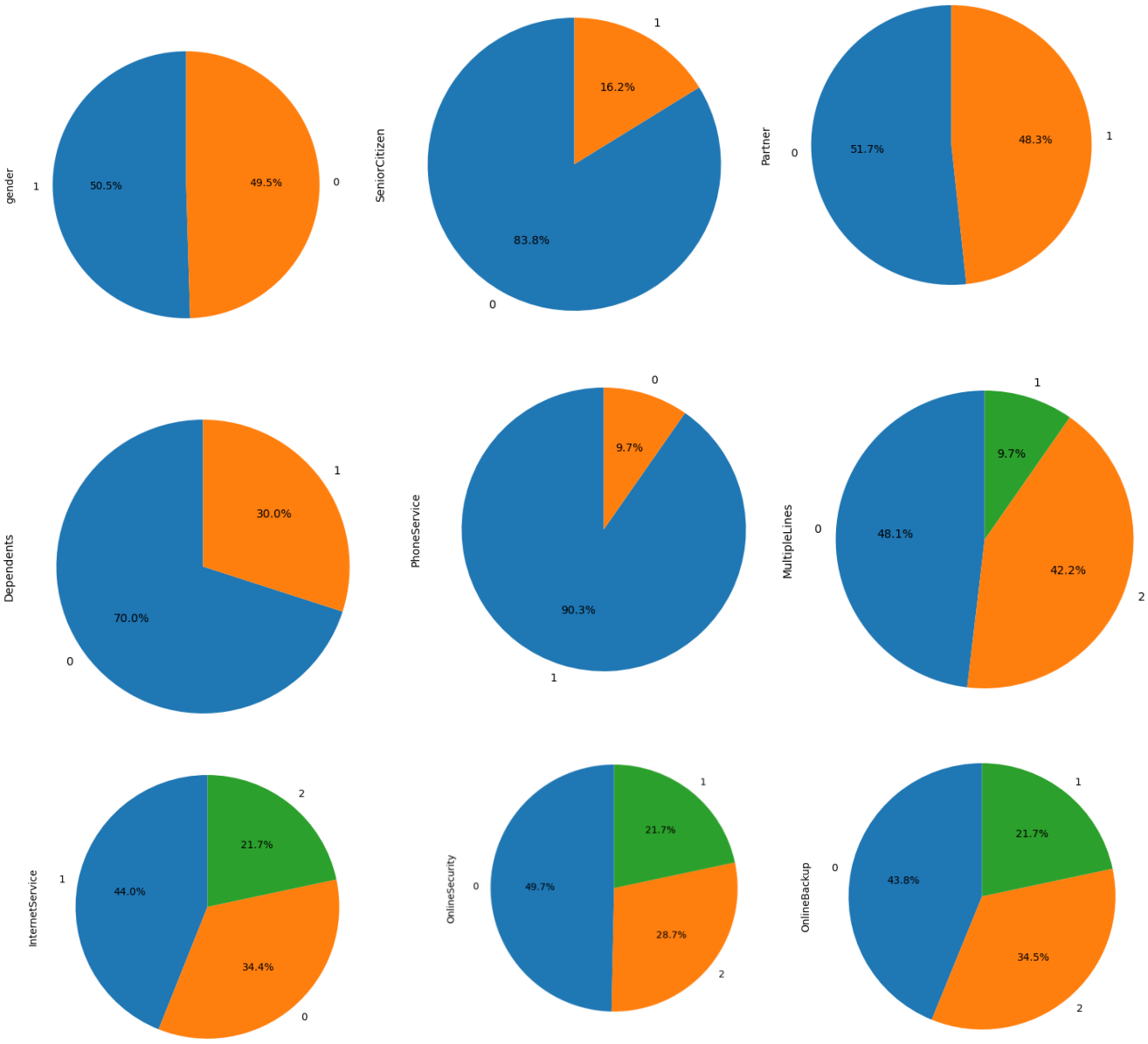
The next is the pairplot here is the graphical representation of the relation between the attributes are given



The next plot is the distplot which has been used to represent mothly charges since it has the highest relational value with churn attribute in the heat map



Next we have pie chart to determine the number of percentage unique values each attribute has



Models implemented are:

Decision tree- Accuracy=76.47%

Classification report=Classification Report:

	precision	recall	f1-score	support
0	0.78	0.94	0.85	1539
1	0.66	0.28	0.39	574
accuracy			0.76	2113
macro avg	0.72	0.61	0.62	2113
weighted avg	0.75	0.76	0.73	2113

Logistic Regression- Accuracy=76.85%

Classification report = Classification Report:

	precision	recall	f1-score	support
0	0.82	0.87	0.85	1539
1	0.59	0.49	0.53	574
accuracy			0.77	2113
macro avg	0.71	0.68	0.69	2113
weighted avg	0.76	0.77	0.76	2113

Random Forest – Accuracy= 76.38%

Classification report=Classification Report:

	precision	recall	f1-score	support
0	0.81	0.89	0.85	1539
1	0.59	0.43	0.50	574
accuracy			0.76	2113
macro avg	0.70	0.66	0.67	2113
weighted avg	0.75	0.76	0.75	2113

xgboost – Accuracy=77.47%

Classificaton report=Classification Report:

	precision	recall	f1-score	support
0	0.82	0.88	0.85	1539
1	0.61	0.49	0.54	574
accuracy			0.77	2113
macro avg	0.71	0.69	0.70	2113
weighted avg	0.76	0.77	0.77	2113

Hence we can see xgboost has provided us with the highest accuracy.