Presentation

Telecom Churn Case Study

Business Goal and Objective:

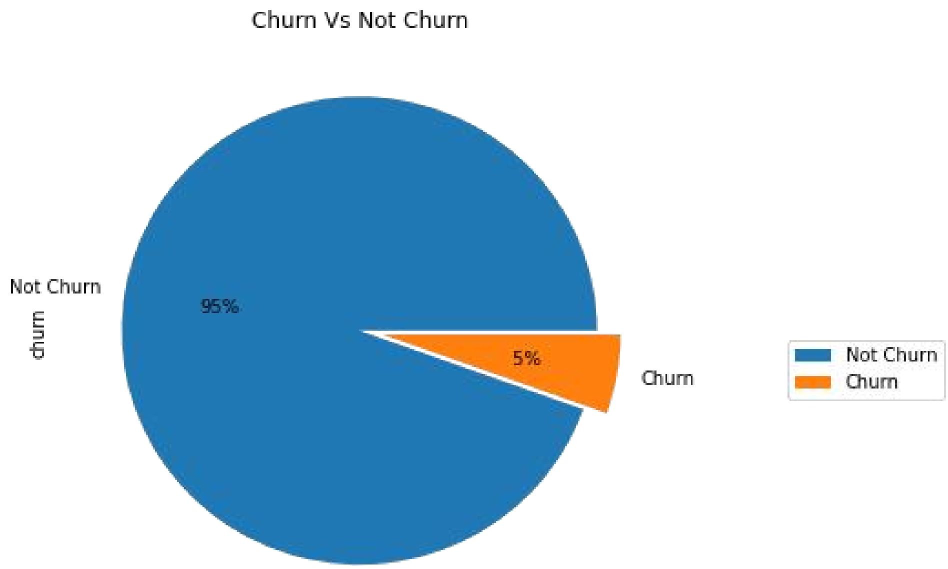
Analyse the customer-level data to identify customers which are at a high risk of churn (usage-based churn) and identify the main indicators of it. Retaining highly profitable customers is the main objective of the company hence correctly predicting customers who are likely to churn will be the focus of this case study to reduce the revenue leakage.

The business objective is to predict the churn in the last (i.e. the ninth) month using the data (features) from the first three months. To do this task well, understanding the typical customer behaviour during churn will be helpful.

We will be using crisp-dm framework

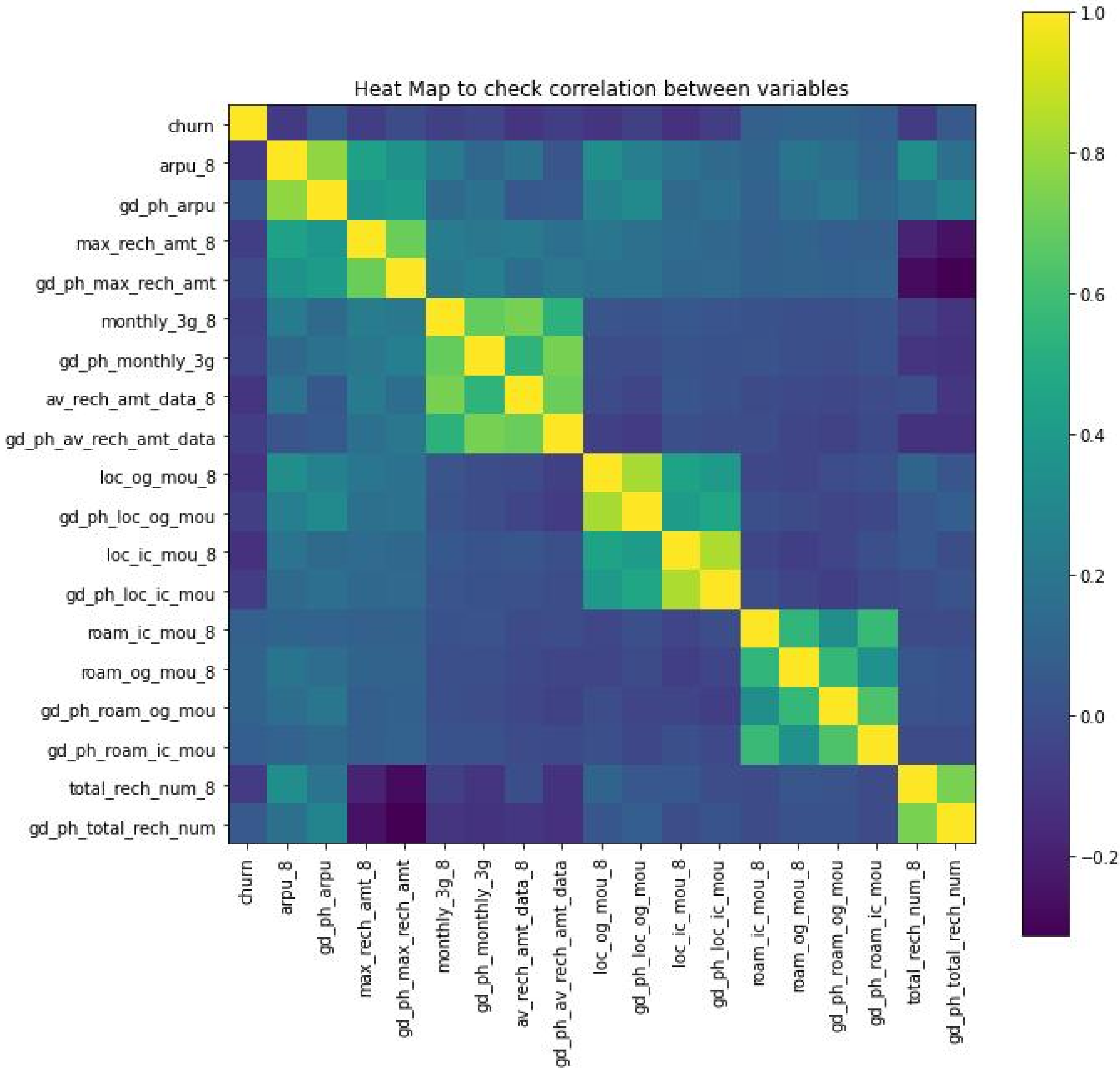
- Business Understanding
- Data Understanding
- Data Preparation
- Model Building
- Model Evaluation

List of features to be analyzed based on the data understanding

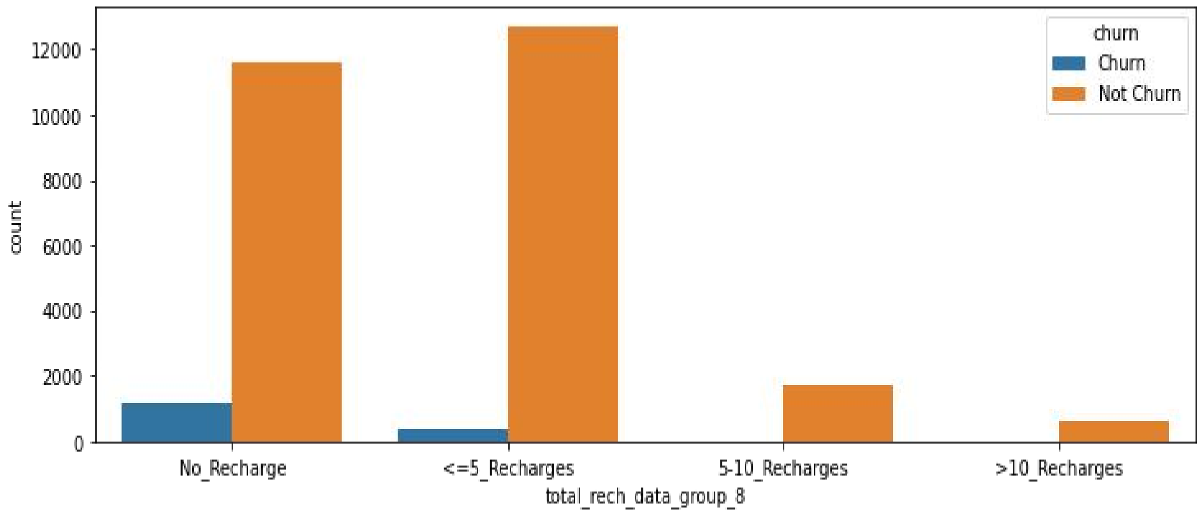


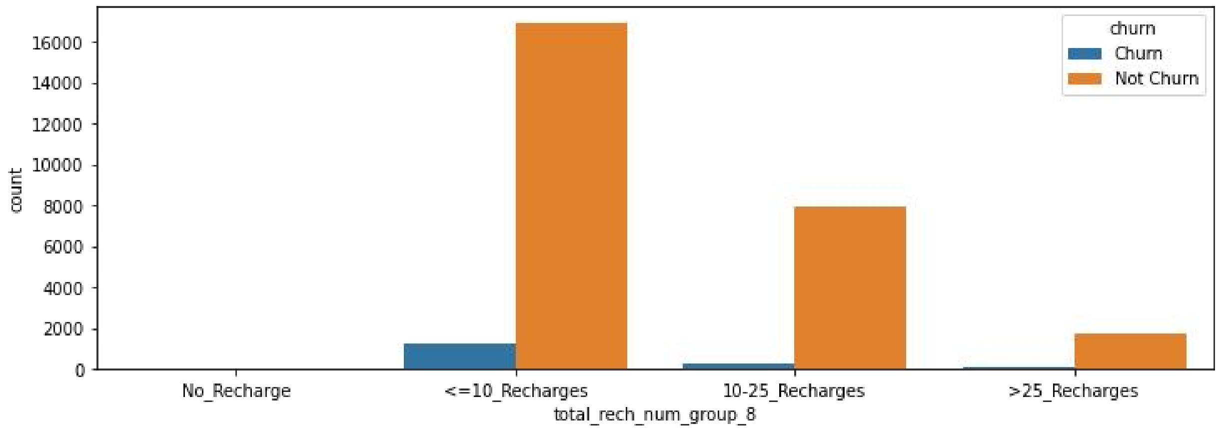
It is quite evitable that there is class imbalance in the data.

We are find correlation between variables

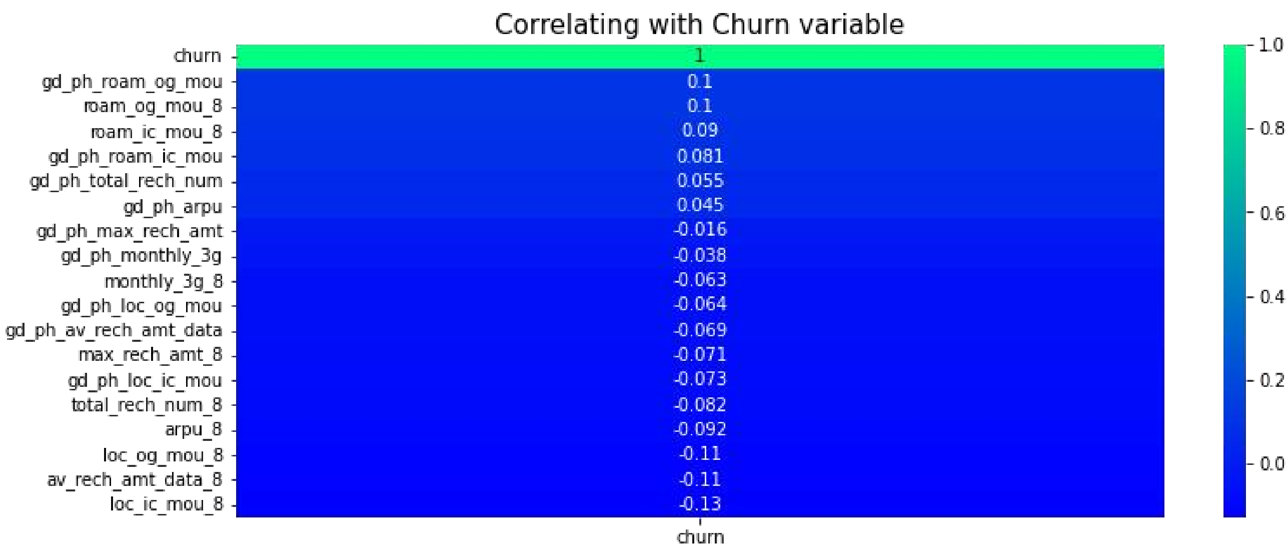


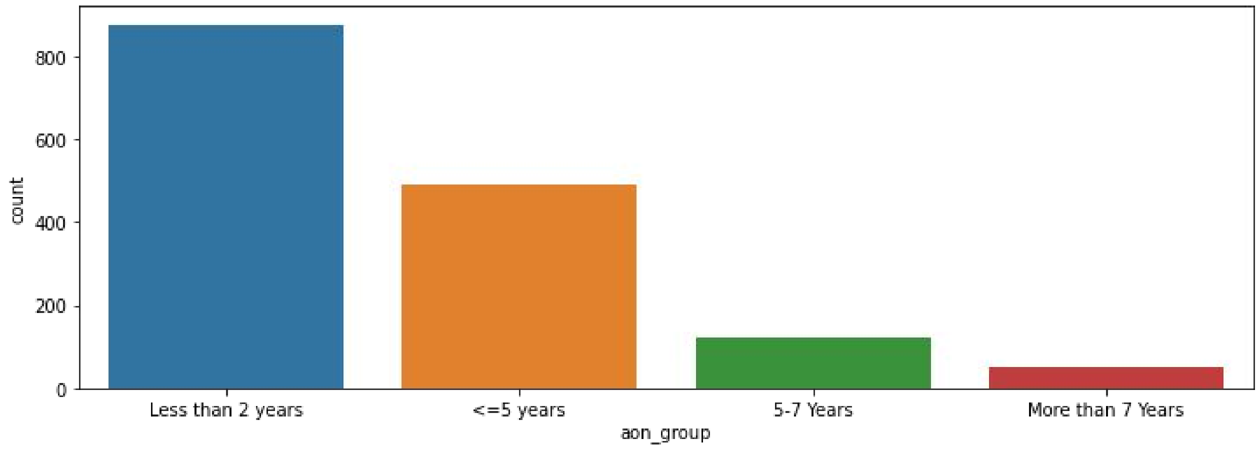
There are good correlation which means people using between the variables which means most people having same type of behaviour in good phase and Action phase.



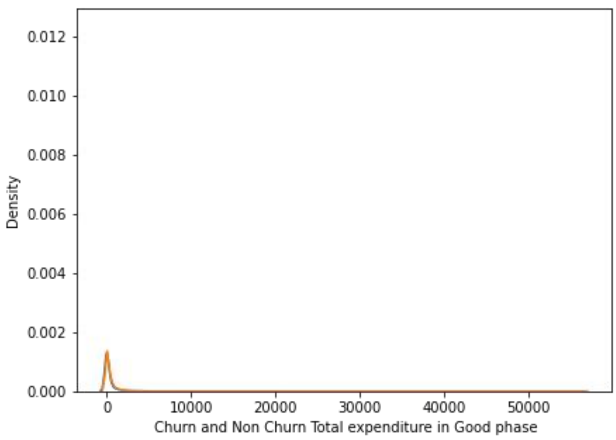


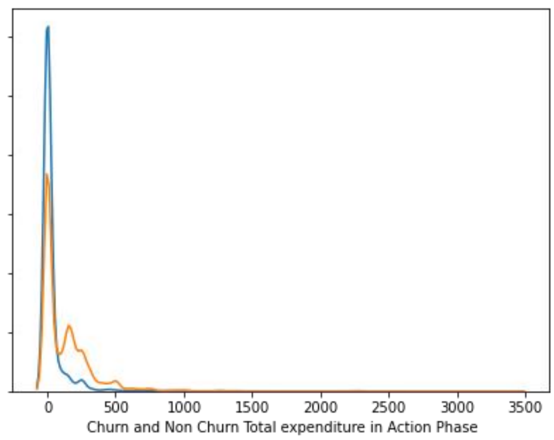
By referring both the graph we can say that users who are doing less number of recharges or not doing recharges are less likely to churn.



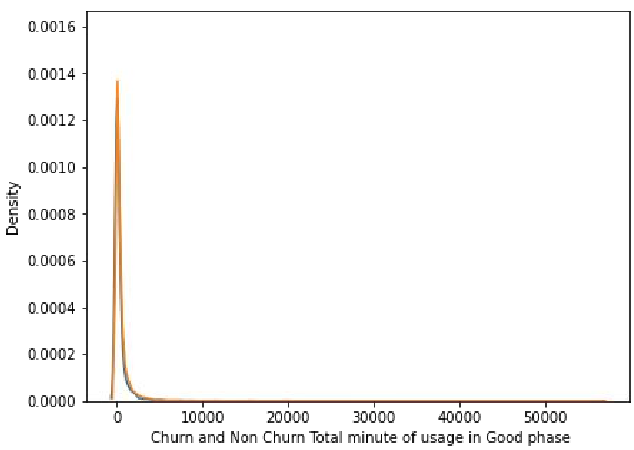


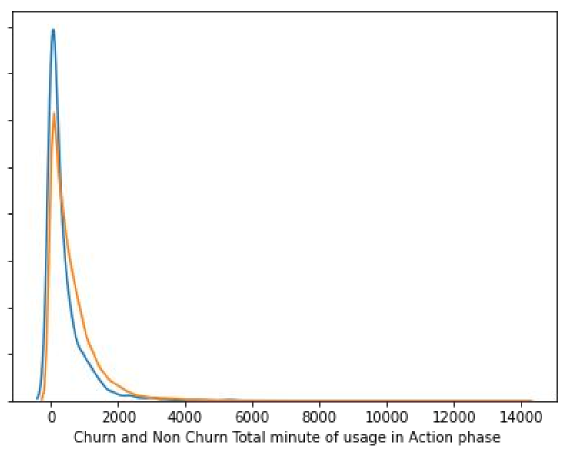
we got proper idea that the age of network is inversely proportional to churn, means user are less likely to churn if they spend a more number of years on same telecom provider.



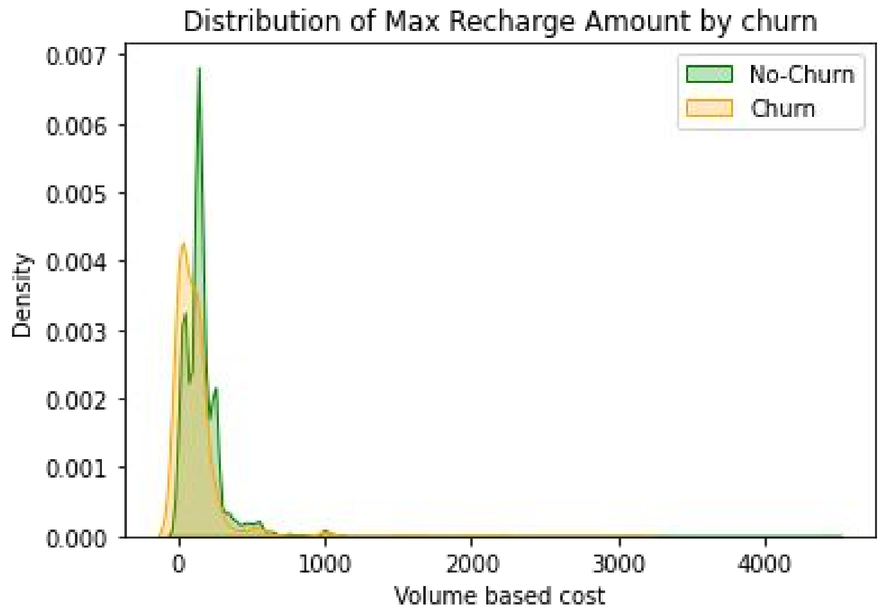


After seeing First graph we can say that there is no difference in good phase for Churn and Non Churn but from second graph we can say people reduces their expenditure on the network before getting churned.

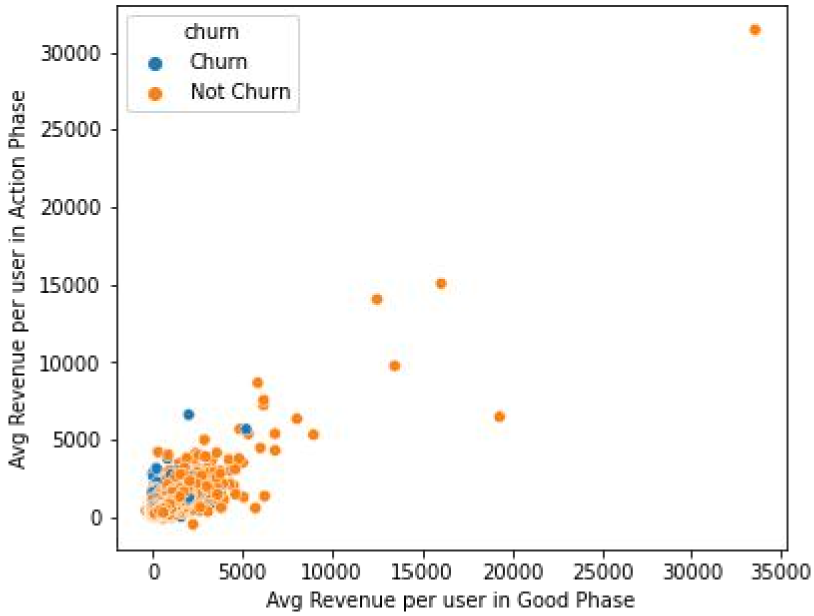




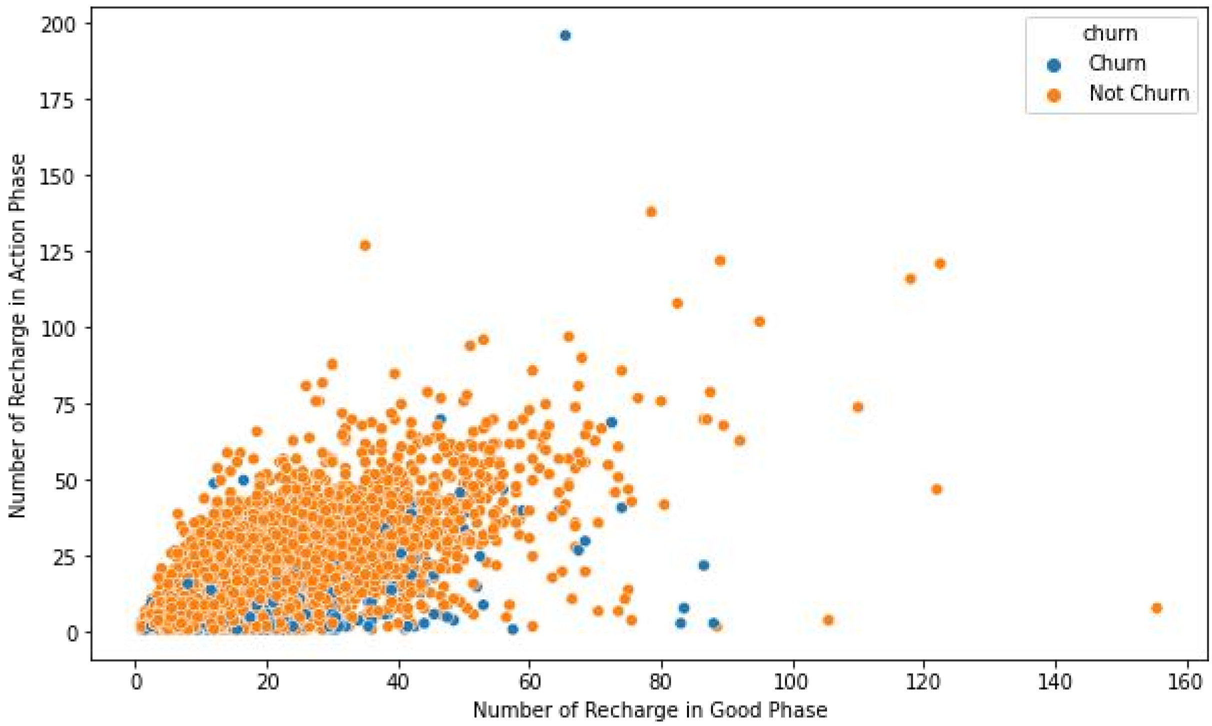
Same behaviour as ealier graph, user usage decreases in Action phase for Churn and Not Churn whereas it is same in good phase.



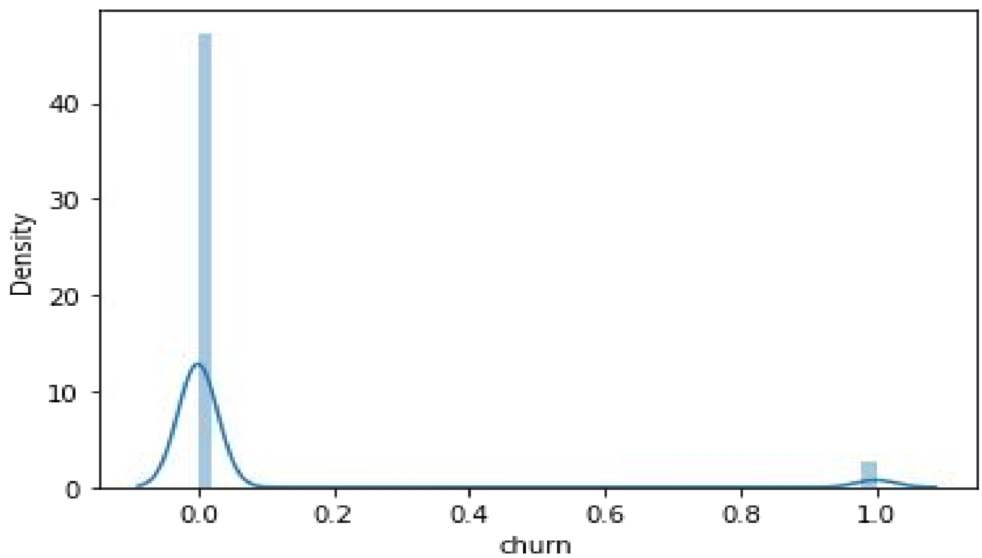
People who are tend to churn are not spending on long term recharges like prepaid recharge of 6 months or 1 years which can bound them for longer period.



We didnot get any clear picture from this but we can say that ARPU is less for the person who are going to churn.



We can say that people who are getting churn are tend to do less number of recharges respect to Non Churn(except some outliers)



Though the target varible is not skewed, it is highly imbalanced.

Handling Class Imbalance

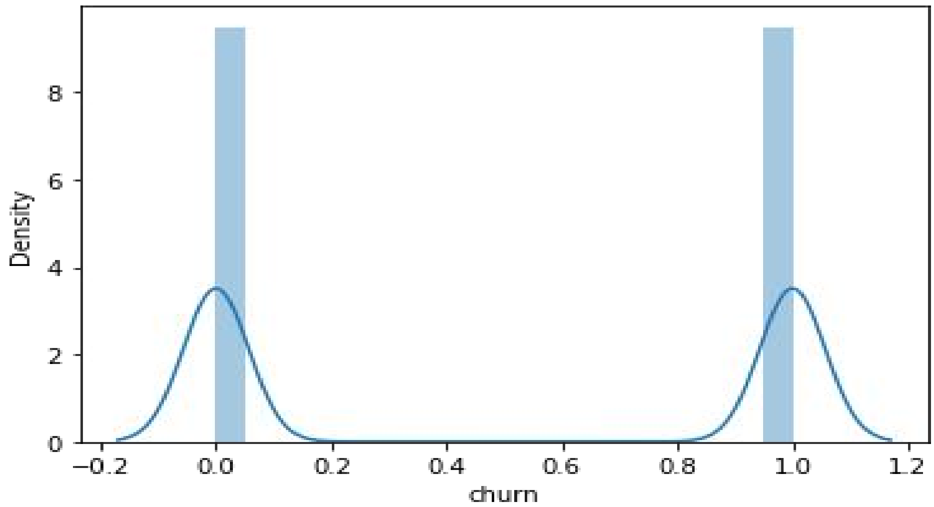
```
# Using SMOTE for class imbalance
from imblearn.over_sampling import SMOTE

smote = SMOTE(random_state=42)
X_res, y_res = smote.fit_resample(data_X, data_y)
```

```
y_res.value_counts()
```

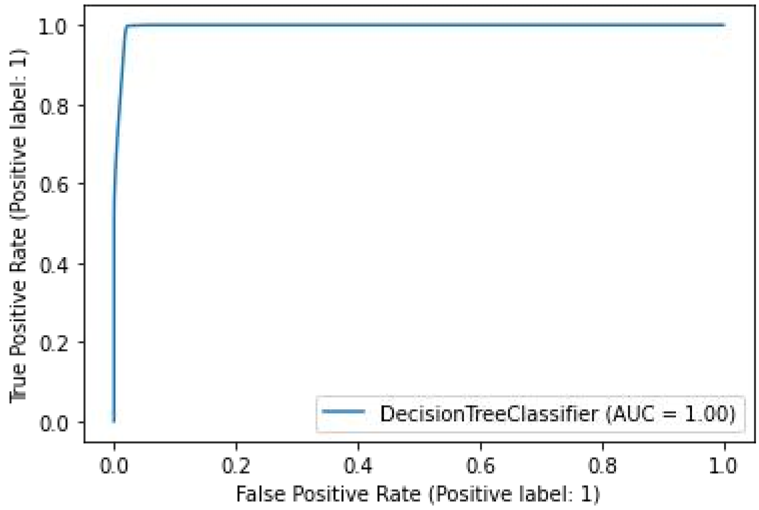
66]:
```
1    26625
0    26625
Name: churn, dtype: int64
```

```
sns.distplot(y_res)
plt.show()
```



As we can see that the target class is balanced

- So using the Logistic Regression Model we are geting an accuracy of 80.9% on train data and 80.73% on test data.
- We can clearly see most of the important features are form the action phase, which is inline with our business understanding that action phase needs more attention.



We are getting an accuracy of 88% on the test data with the decision tree model.

Random Forrest (After Hyperparameter Tuning)

Parameter	Train	Test
Accuracy	98.08%	94%
Recall	-	92%
Precision	-	95%

> Therefore we chose the Random Forest Model as our final model and got an **Accuracy** of `94%` on the test data set with `92%` **Recall** which suggest that it has a very high probability of correctly predicting the customer who are likely to churn.

The top churn predictors are :

Features	Rank
total_rech_amt_data_8	1
arpu_8	2
std_og_mou_8	3
total_rech_num_8	4
max_rech_amt_8	5
std_ic_mou_8	6
total_rech_data_8	7
gd_ph_std_og_mou	8
spl_ic_mou_8	9
std_ic_t2t_mou_8	10
loc_og_mou_8	11

We can see from the above table that most features that affect the churn are related to the Recharge Amount and Minutes of Usage which are basically the two features that generates the most revenue for the telecom operator as well as decrease in these usage features suggest that the customers have started to invest less on the network. From this we can also infer from our business understanding that as we move into the Action phase which is the 8th month, the impact of Usage-Based Churn becomes more and more evident.

As now we can see the impact we can formulate strategies that the telecom operater can use in order to reduce this churn and prevent the revenue leakage as much as possible.

> Average Revenue Per User is one of the most imporant feature in churn prediction which suggest that customers who are willing to invest less on the network have a high tendency to churn. The company should focus on deals for such customers which can encourage them to invest in the network and in turn reduce churn propability along with bringing in revenue.
>
> Age On Network is also one of the influential features as we could see during our analysis that customers whom age is less than 44 years on the network are more likely to churn. This behaviour also suggests about those customers who have churned even after being with the network for more than 4 years is that either, they are temporarily switching to a different network and will be back or there is something that has gone quite wrong with the services provided to them due to which they churned. Such customers can be brought back to the network by genuinely investing some time to fix the issue these customers are facing. As all these customers are High-Value customers, the revenue brought in by each of them would not be small.
>
> Minutes of Usage in a Local or STD capacity as well as Data Recharge Amount are also strong predictors of churn. It suggests the customers who are willing to invest less on the outgoing calls or on the data services are either not happy with the network or don't have that much network usage. As we have build our model based on the High-value customers, the second possibility seems to be far reaching hence assuming that these customers are not happy with the services and the network provider, the telecom company should focus on this segment of customers and make sure that the network coverage is optimal and data availability is there for them to use the services of the company to the fullest and not churn