TELECOM CHURN CASE STUDY

PRESENTED BY
SAKSHI BHASIN
RUCHI GUPTA
RUCHIRA DOLAI

DS C54
BUSINESS ANALYTICS

EDA Analysis

Step 3.1 : Univariate analysis

Churn rate on the basis whether the customer decreased his/her MOU in action month

```
In [74]: # Converting churn column to int in order to do aggfunc in the pivot table
         df_new['churn'] = df_new['churn'].astype('int64')

In [75]: df_new.pivot_table(values='churn', index='decrease_mou_action', aggfunc='mean').plot.bar()
         plt.ylabel('churn rate')
         plt.show()
```

Observations:

We can observe that the churn rate is more for the customers, whose minutes of usage(mou) decreased in the action phase compared to that when the customer is in the good phase

Churn rate on the basis whether the customer decreased her/his number of recharge in action month

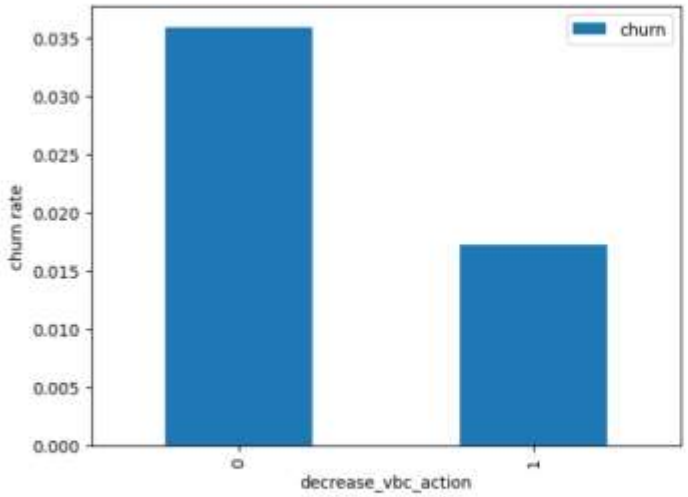
In [76]:
```
df_new.pivot_table(values='churn', index='decrease_rech_num_action', aggfunc='mean').plot.bar()
plt.ylabel('churn rate')
plt.show()
```

Observations:

We can observe that the churn rate is more for the customers, whose number of recharge in the action phase is lesser than the number in good phase.

Churn rate on the basis whether the customer decreased her/his amount of recharge in action month

In [78]:
```
df_new.pivot_table(values='churn', index='decrease_vbc_action', aggfunc='mean').plot.bar()
plt.ylabel('churn rate')
plt.show()
```



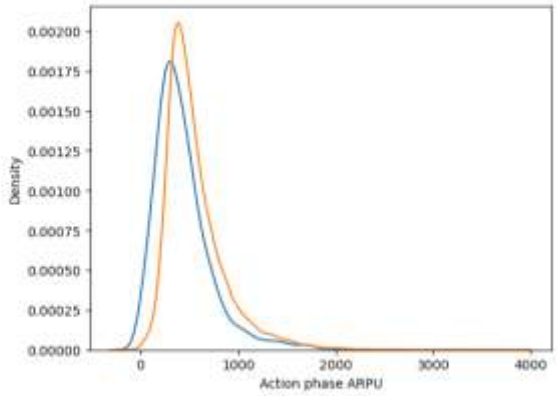
Observations:

We can observe that the churn rate is more for the customers, whose volume based cost in action month is increased which means that the customers do not do the monthly recharge more when they are in the action phase.

In [79]:
```
# Creating churn dataframe
churn_df = df_new[df_new['churn'] == 1]
# Creating not churn dataframe
non_churn_df = df_new[df_new['churn'] == 0]
```

In [80]:
```
# Distribution plot
ax = sns.distplot(churn_df['avg_arpu_action'],label='churn',hist=False)
ax = sns.distplot(non_churn_df['avg_arpu_action'],label='not churn',hist=False)
ax.set(xlabel='Action phase ARPU')
```

Out[80]:
```
[Text(0.5, 0, 'Action phase ARPU')]
```



Observations:

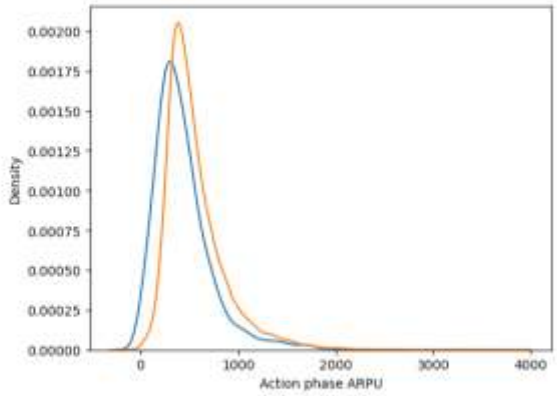
We can observe that the Average revenue per user (ARPU) for the churned customers is mostly densed on the 0 to 900. The higher ARPU customers are less likely to be churned

ARPU for the not churned customers is mostly densed on the 0 to 1000.

In [79]:
```
# Creating churn dataframe
churn_df = df_new[df_new['churn'] == 1]
# Creating not churn dataframe
non_churn_df = df_new[df_new['churn'] == 0]
```

In [80]:
```
# Distribution plot
ax = sns.distplot(churn_df['avg_arpu_action'],label='churn',hist=False)
ax = sns.distplot(non_churn_df['avg_arpu_action'],label='not churn',hist=False)
ax.set(xlabel='Action phase ARPU')
```

Out[80]:
```
[Text(0.5, 0, 'Action phase ARPU')]
```



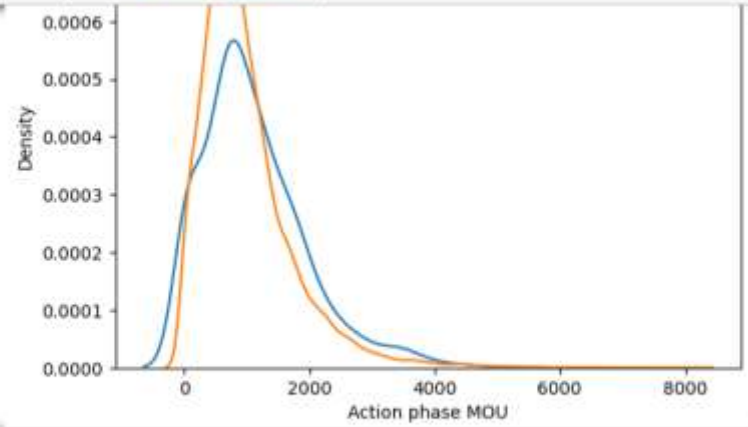
Observations:

We can observe that the Average revenue per user (ARPU) for the churned customers is mostly densed on the 0 to 900. The higher ARPU customers are less likely to be churned

ARPU for the not churned customers is mostly densed on the 0 to 1000.

Analysis of the minutes of usage MOU (churn and not churn) in the action phase

In [81]:
```
# Distribution plot
ax = sns.distplot(churn_df['total_mou_good'],label='churn',hist=False)
ax = sns.distplot(non_churn_df['total_mou_good'],label='non churn',hist=False)
ax.set(xlabel='Action phase MOU')
```



Observations:

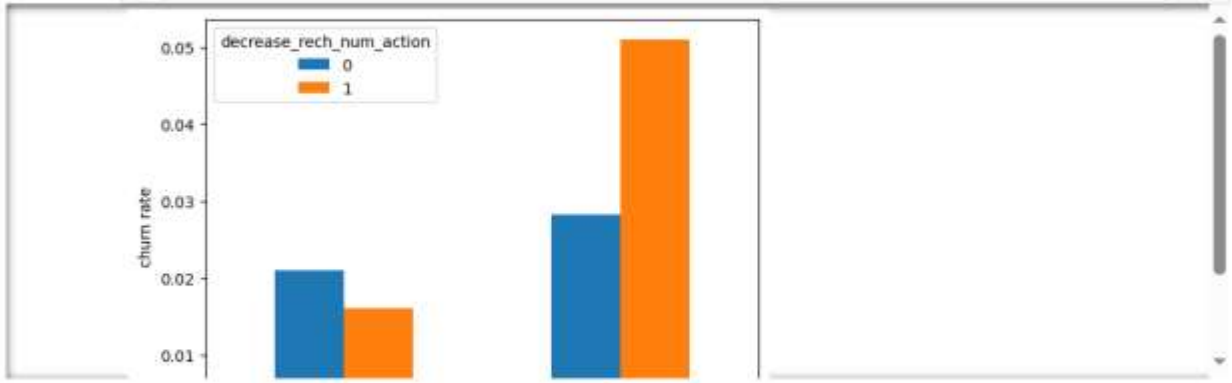
We can observe that the Minutes of usage(MOU) of the churn customers is mostly populated on the 0 to 2500 range. Higher the MOU, lesser the churn probability.

Step 3.2 : Bivariate analysis

Step 3.2 : Bivariate analysis

Analysis of churn rate by the decreasing recharge amount and number of recharge in the action phase

```
In [82]: df_new.pivot_table(values='churn', index='decrease_rech_amt_action', columns='decrease_rech_num_action', aggfunc='mean').plot.bar
         plt.ylabel('churn rate')
         plt.show()
```

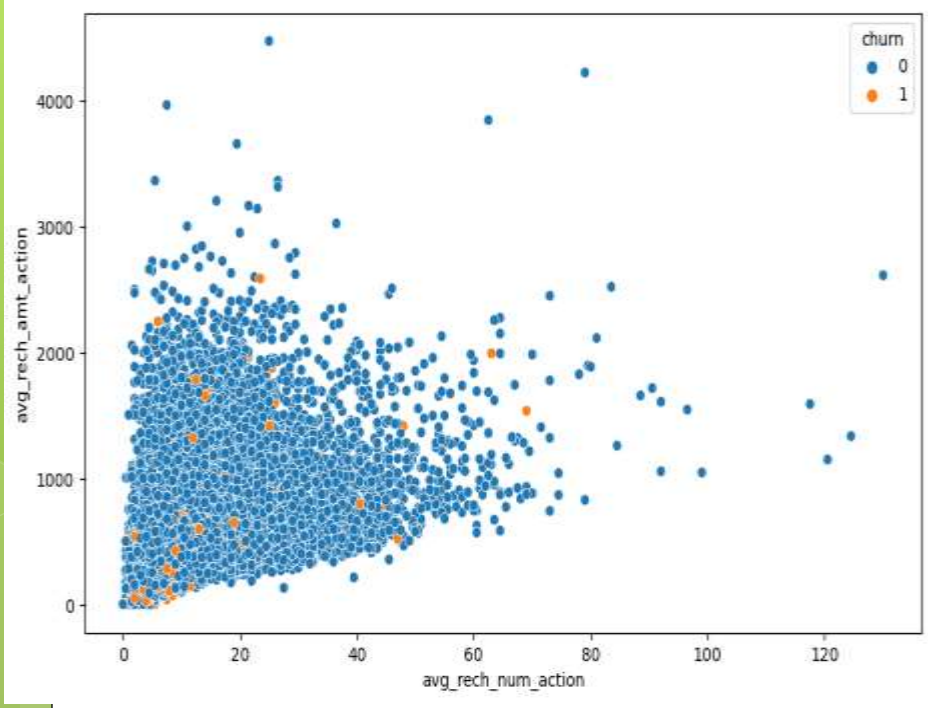


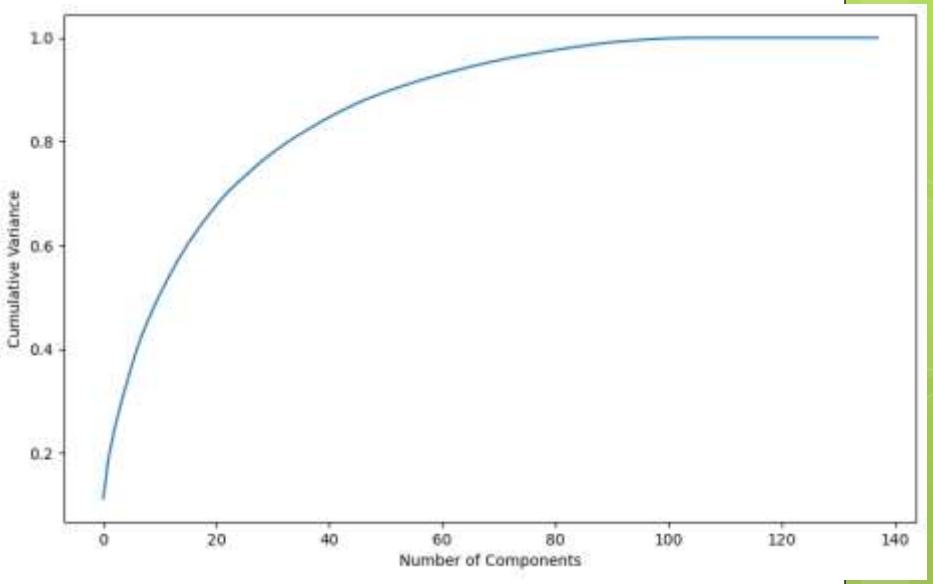
Observations:

We can observe from the above plot that the churn rate is more for the customers, whose recharge amount as well as number of recharge have decreased in the action phase than the good phase.

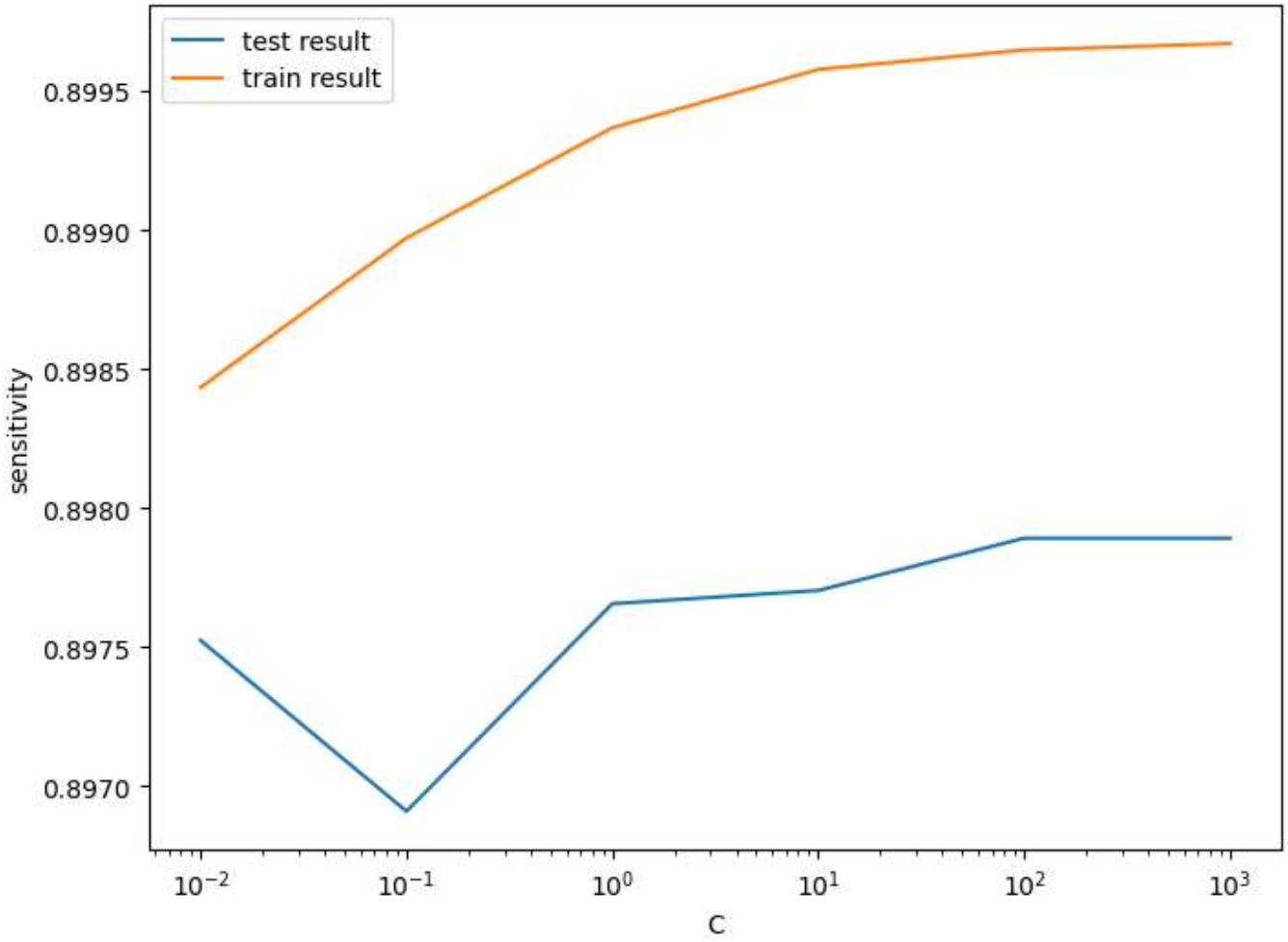
Analysis of churn rate by the decreasing recharge amount and volume based cost in the action phase

Step 7 : Model Building with PCA

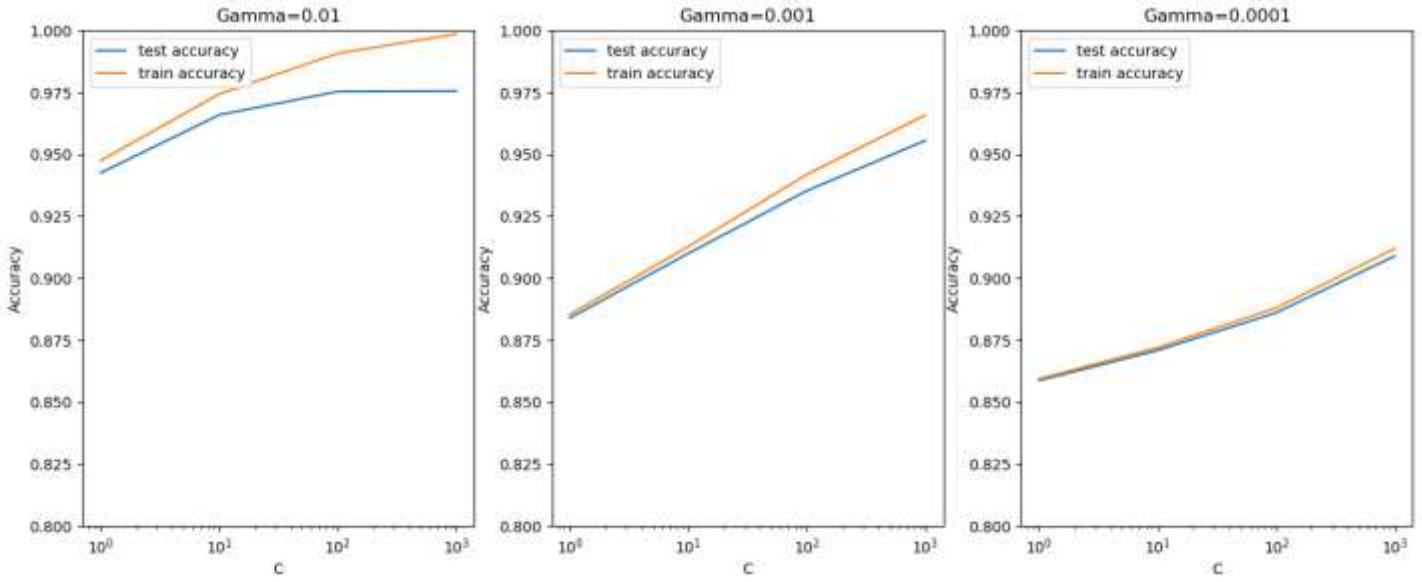




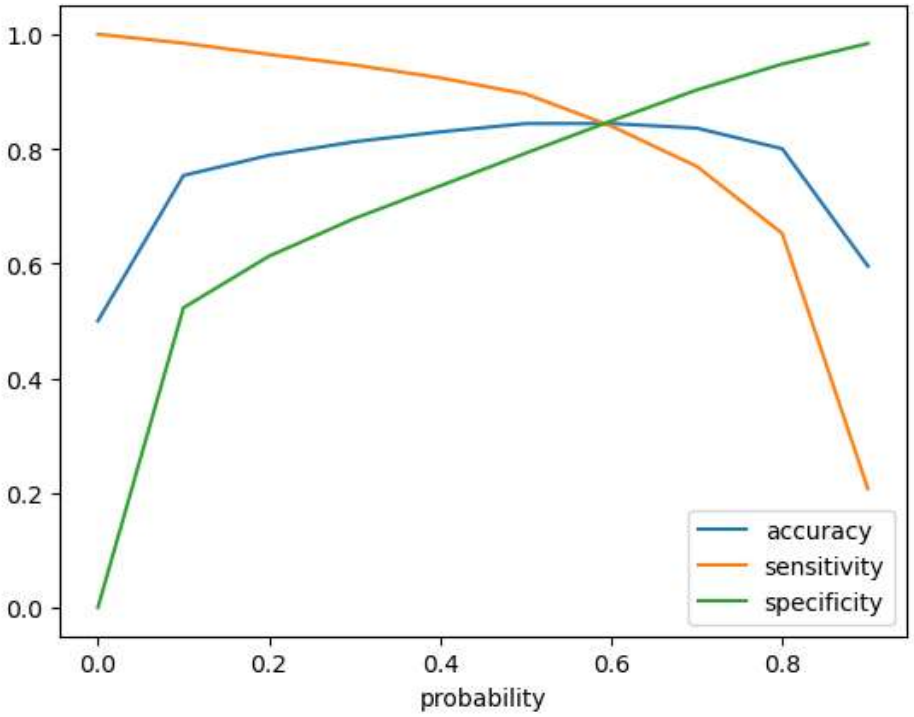
Step 8 : Logistic regression with PCA



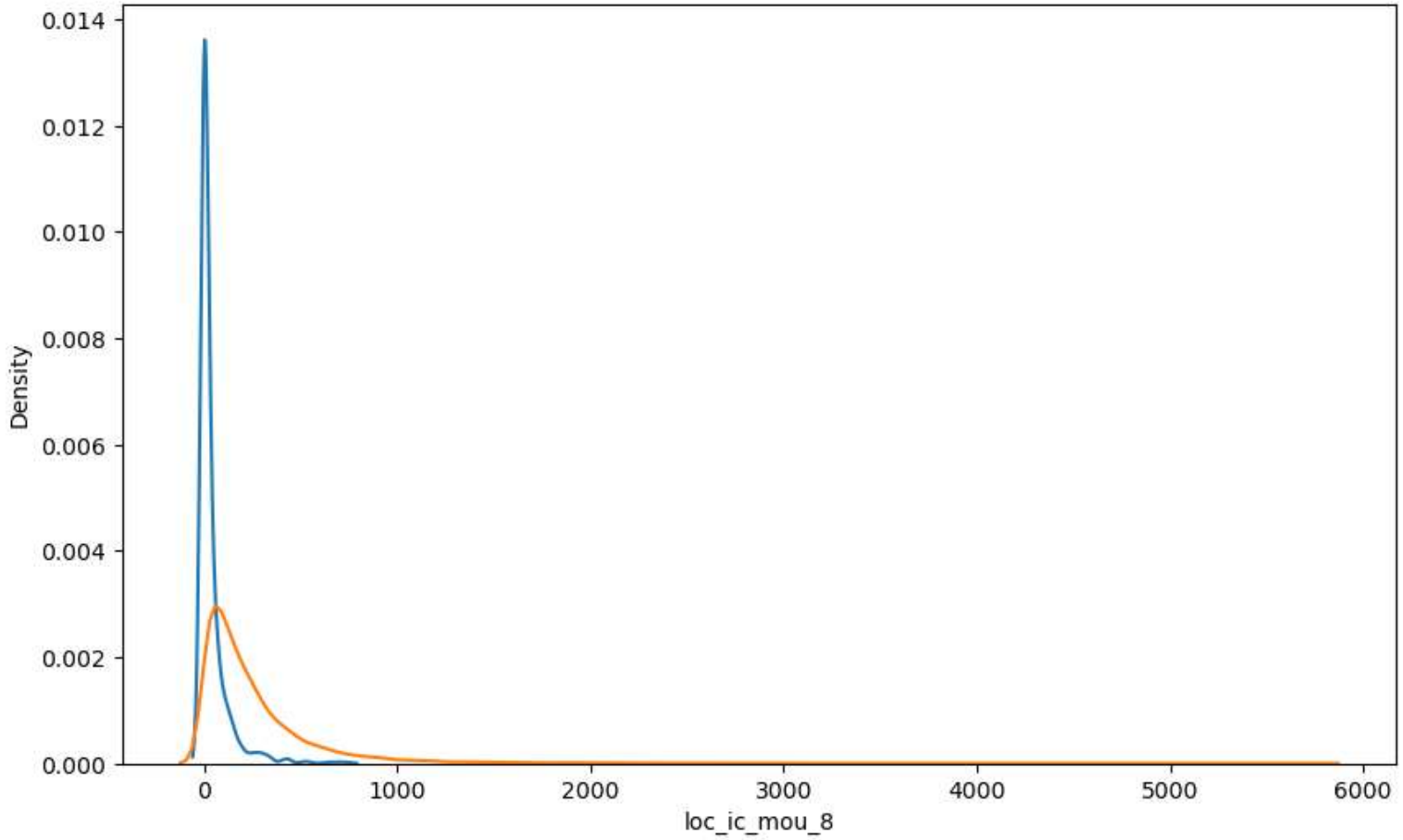
Step 9.2 : Plotting the accuracy with various C and gamma values

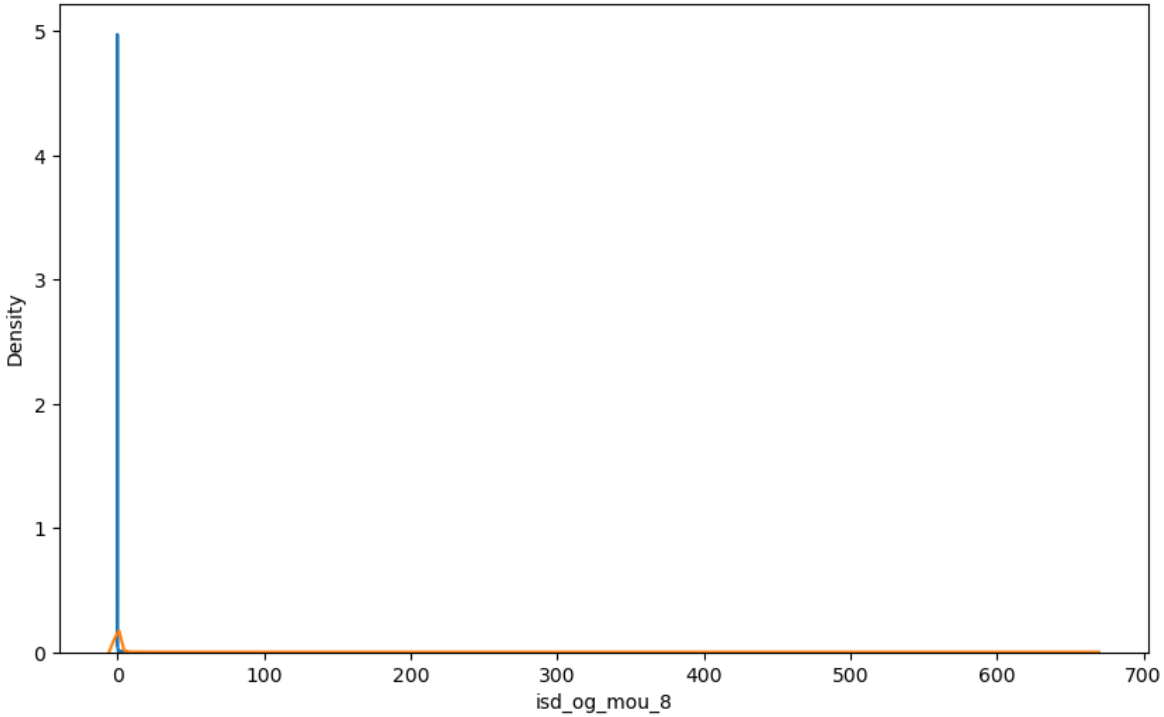


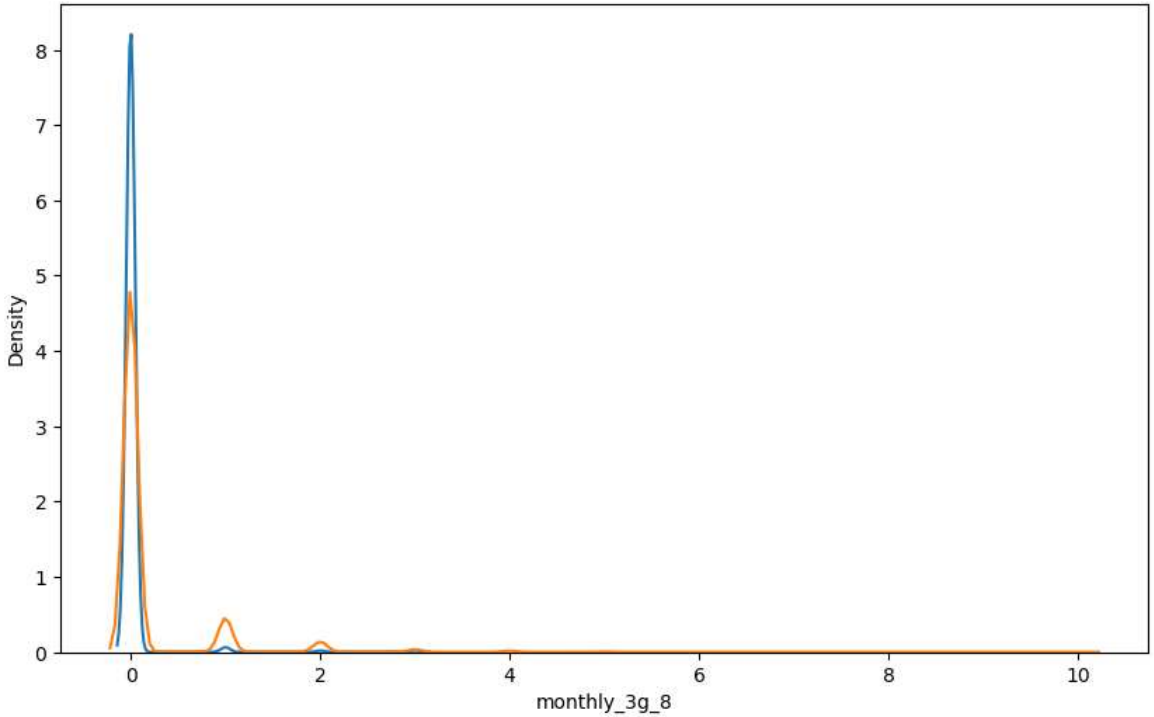
Step 15 : Calculate the accuracy sensitivity and specificity for various probability cutoffs.



Step 18 : Plots of important predictors for churn and non churn customers




Density
0
1
2
3
4
5
0
100
200
300
400
500
600
700
isd_og_mou_8


Density
0
1
2
3
4
5
6
7
8
0
2
4
6
8
10
monthly_3g_8

THANK YOU