TELECOM CHURN CASE STUDY ASSIGMENT

Submitted by:

Kalpana Banjare

Kiran Jain

Jyoti Maruti Kadam

Problem Statement

- customers are able to choose from multiple service providers and actively switch from one operator to another.
- the telecommunications industry experiences an average of 15-25% annual churn rate.
- build predictive models to identify customers at high risk of churn and identify the main indicators of churn.
- Retaining high profitable customers is the main business goal here.

STEPS

- Reading and understanding the data set
- Cleaning and visualizing the data
- Preparing the data for modelling
- Building the model
- Evaluate the model

Filter high-value customers

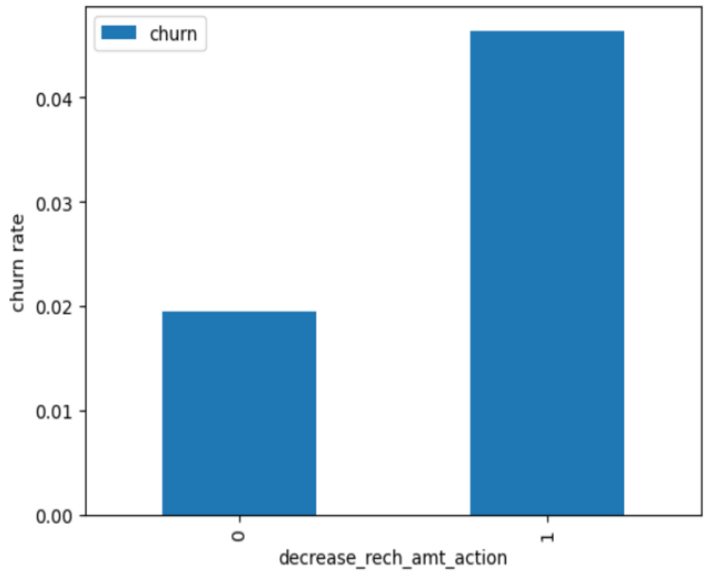
- Creating column avg of first two month by summing up total recharge amount of month 6 and 7. Then taking the average of the sum.
- who have recharged with an amount more than or equal to X, where X is the 70th percentile of the average recharge amount in the first two months (the good phase).
- After this we get almost 30K rows.

Tag churners and remove attributes of the churn phase

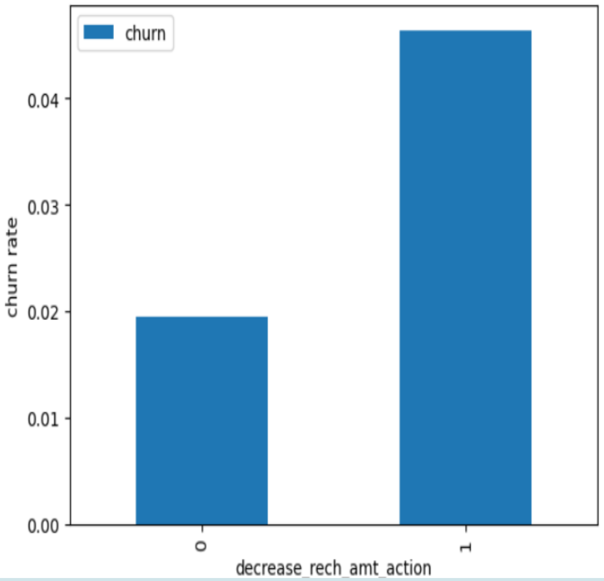
- Tag the churned customers (churn=1, else 0)
- Charn based on the fourth month from June, July, August, September.
- Who have not made any calls AND have not used mobile internet even once in the churn phase.
- Remove all the attributes corresponding to the churn phase after tagging churn

Analysis

```
In [91]: # recharge amount duging action time
data.pivot_table(values='churn', index='decrease_rech_amt_action', aggfunc='mean').plot.bar()
plt.ylabel('churn rate')
plt.show()
```



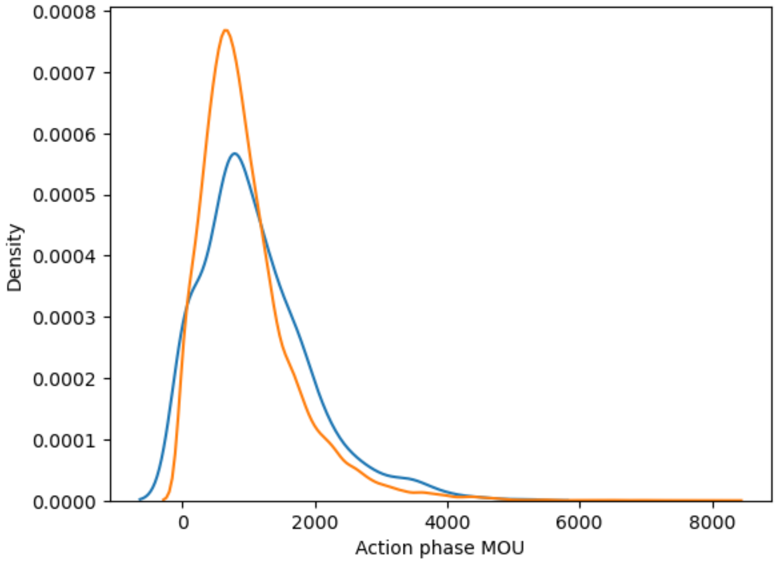

```
In [91]: # recharge amount duging action time
data.pivot_table(values='churn', index='decrease_rech_amt_action', aggfunc='mean').plot.bar()
plt.ylabel('churn rate')
plt.show()
```

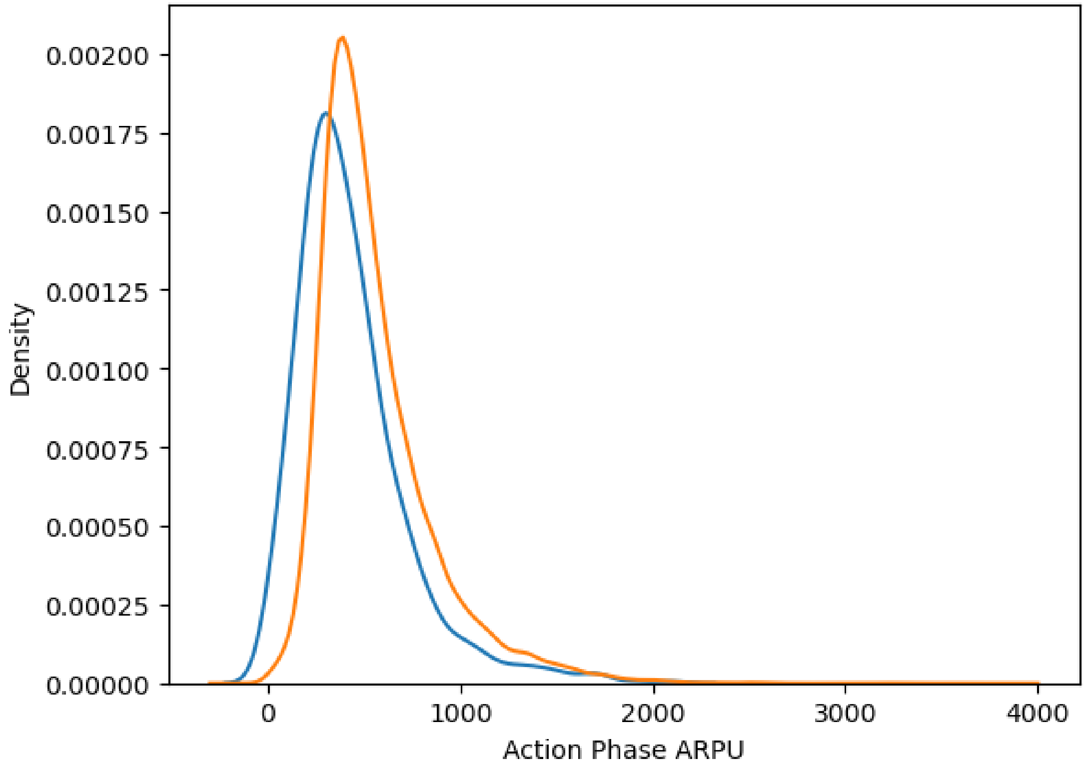


- IN Action phase number of recharge decrees.
- recharge amount is less.
- and churn rate is high

Univariate Analysis

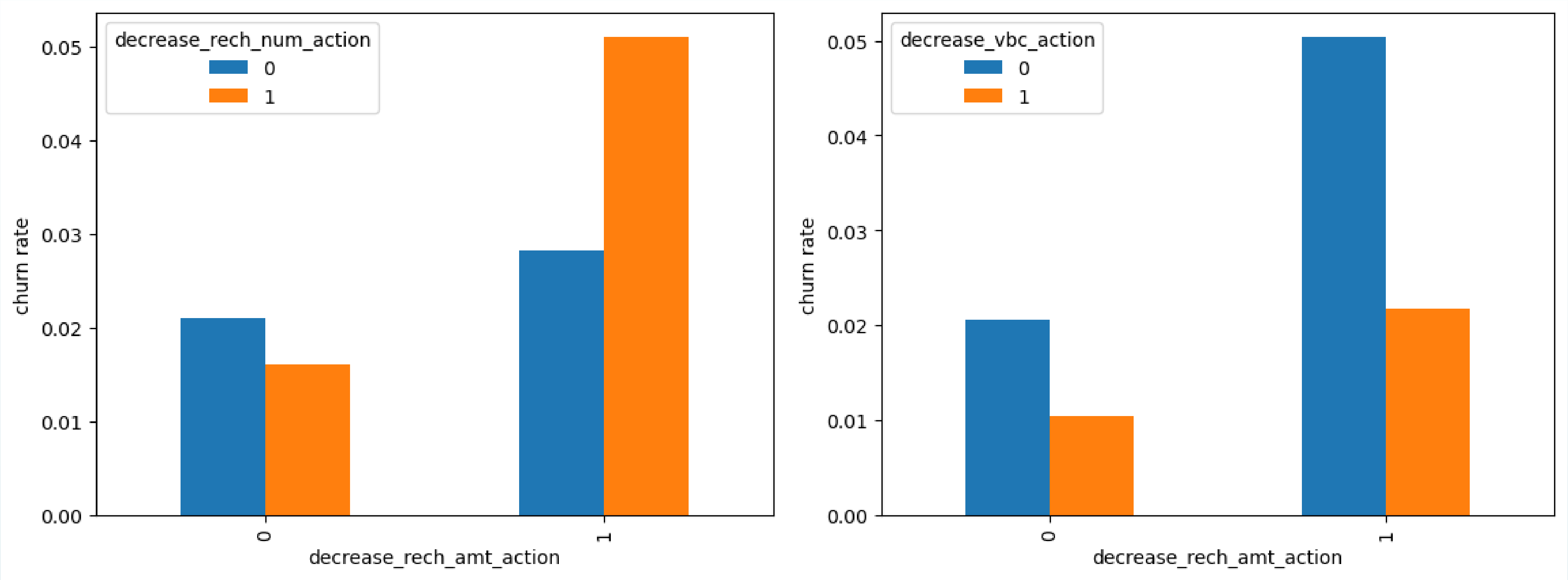
In [126]:
```
# Distribution plot mou in good phase
ax = sns.distplot(data_churn['total_mou_good'], hist=False)
ax = sns.distplot(data_non_churn['total_mou_good'], hist=False)
ax.set(xlabel='Action phase MOU')
```
Out[126]: [Text(0.5, 0, 'Action phase MOU')]





- Higher average rate revenue par customer the lesser the churn.
- Higher the mou lesser the churn.

Bivariate Analysis



- Amount and number of recharge decrees in action phase and churn rate is high.
- Recharge amount is less in action phase

Modelling

- P-value should be less then 0.5 .
- VIF should be less then 5.
- Taking 15 variable during RFE.
- After final modelling we have 12 value

Logistic Regression

- Tread-off between Accuracy, Sensitivity and Specificity meeting at 0.6 that is good.
- ROC Curve is .92 almost 1

TP = confusion[1,1] # ture positive

TN = confusion[0,0] # ture négatives

FP = confusion[0,1] # false positives

FN = confusion[1,0] # false négatives

- Performance on Train and test set is good
- Testing on train dataset

  Accuracy 84%

  Sensitivity 89%

  Specificity 79%
- Testing on test dataset

  Accuracy 78%

  Sensitivity 82%

  Specificity 78%

Decision Tree

- Testing on train dataset

  Accuracy 95%

  Sensitivity 97%

  Specificity 94%

- Testing on test dataset

  Accuracy 92%

  Sensitivity 77%

  Specificity 93%

- Performance for train and test set are good.
- Decision tree is better then Logistic regression.

Random Forest

- Testing on train dataset

  Accuracy 95%

  Sensitivity 98%

  Specificity 91%

- Testing on test dataset

  Accuracy 96%

  Sensitivity 16%

  Specificity 99%

- Performance is good in train dataset but not good in test dataset.

Conclusion

- The best model among logistic regression, Decision tree and Random forest is Decision is the best model
- Decision tree all value for accuracy, Sensitivity and Specificity are accurate in both train and test dataset
- Customer who decrees uses of service in the month of august mostly likely to churn.
- Customer who decrees data pack are likely to churn.
- Target on high revenue generate customer.

Recommendations

- Focus on the customer of August month based on MOU(minute of uses)
- Gives offer to customer so they do not churn.
- Focus on Customers, whose monthly 3G recharge in August is more, are likely to be churned.
- Focus on Customers, whose monthly 2G recharge in August is more, are likely to be churned.
- The customers having value based cost in the action phase increased are more likely to churn than the other customers. Hence, these customers may be a good target to provide offer.

THANKYOU!