TELECOM CHURN CASE STUDY

* GROUP MEMBERS

- ABHISHEK CHENDVANKAR
- PRIYA NARAYAN
- SHIVANGI SHARMA

PROBLEM STATEMENT

- In the telecom industry, customers are able to choose from multiple service providers and actively switch from one operator to another.
- In this highly competitive market, the telecommunications industry experiences an average of 15-25% annual churn rate.
- Given the fact that it costs 5-10 times more to acquire a new customer than to retain an existing one, **customer retention** has now become even more important than customer acquisition
- For many incumbent operators, *retaining high profitable customers is the number one business goal.*
- To reduce customer churn, telecom companies need to **predict which customers are at high risk of churn**.

* Here we will analyze customer-level data of a leading telecom firm, build predictive models to identify customers at high risk of churn and identify the main indicators of churn.

BUSINESS GOAL

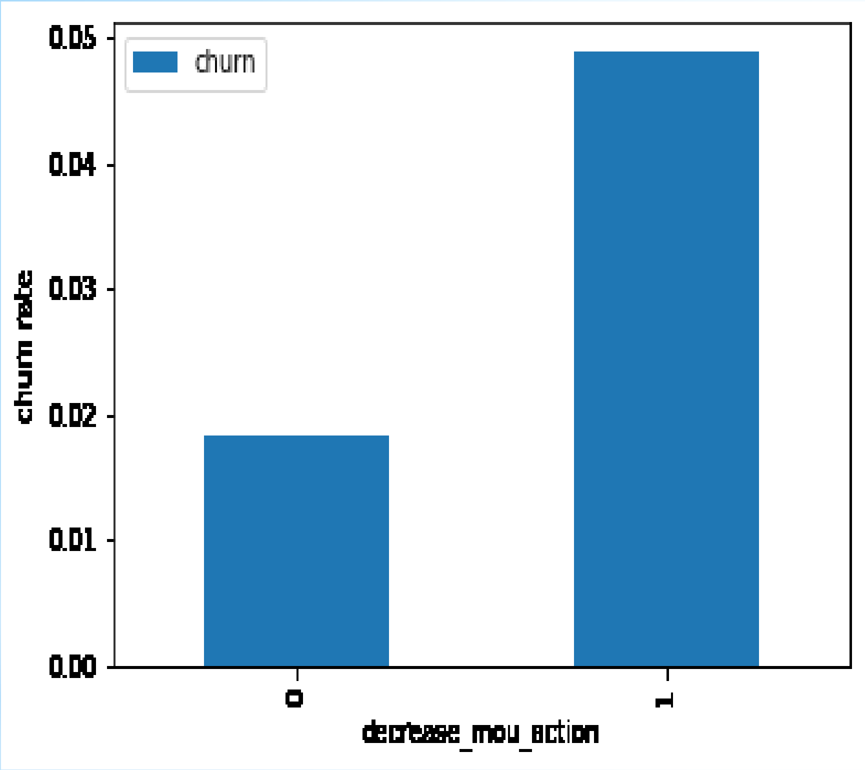
The dataset provided in this dataset contains customer-level information for a span of four consecutive months - June, July, August and September. The months are encoded as 6, 7, 8 and 9, respectively. The **business goal:**

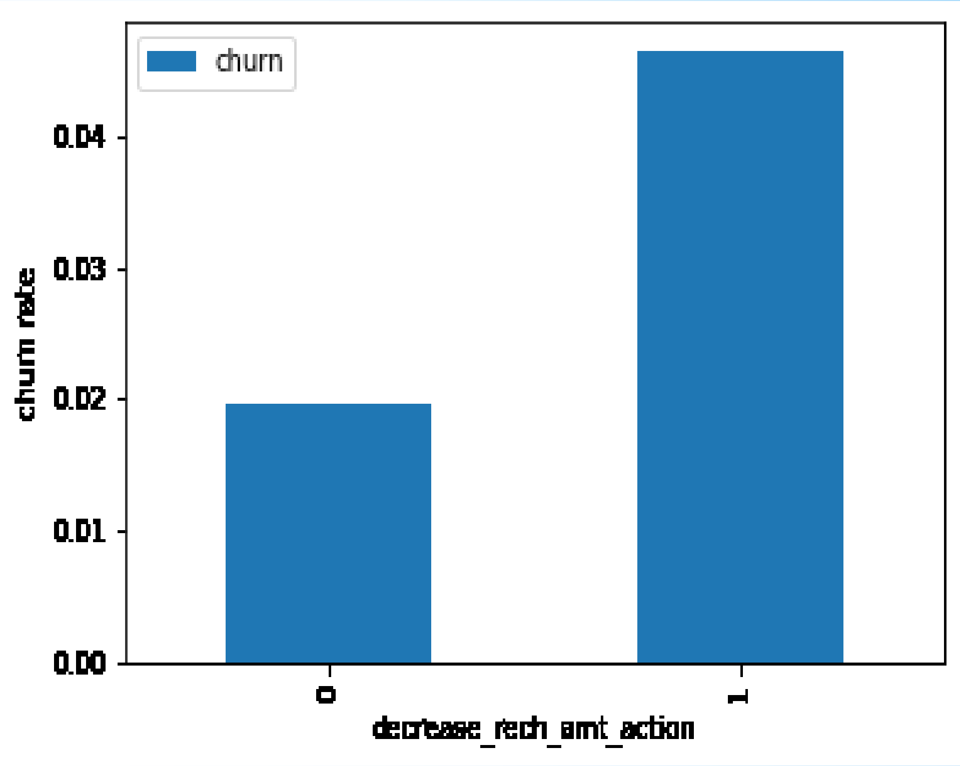
- **To predict** the churn in the last (i.e. the ninth) month using the data (features) from the first three months.
- Understand the typical customer behavior during churn .
- As, we are working over a four-month window, the first two months are the 'good' phase, the third month is the 'action' phase, and the fourth month is the 'churn' phase.
- Build model with the main objective of identifying important predictor attributes which help the business understand indicators of churn.
- A good choice to identify important variables is a **logistic regression** model or a model from the **tree family**.

Steps Followed:

Steps 1:Reading and Importing the dataset
Step 2:Inspecting the Dataset
Step 3: Cleaning the null values
Step 4:EDA
i) Univariate Analysis
ii)Bivariate Analysis
Step 5:Train-Test Split
i)Data Imbalance
ii)Feature Scaling
Step 6:Model with PCA
Step 7: Sensitivity/Recall than Accuracy
Step 8:Tuning Hyperparameter C
Step 9:Support Vector Machine(SVM) with C
Step 10:Decision Tree with PCA
Step 11.Feature Selection Using RFE
Step12.Checking VIFs
Step 13:Plotting ROC curve

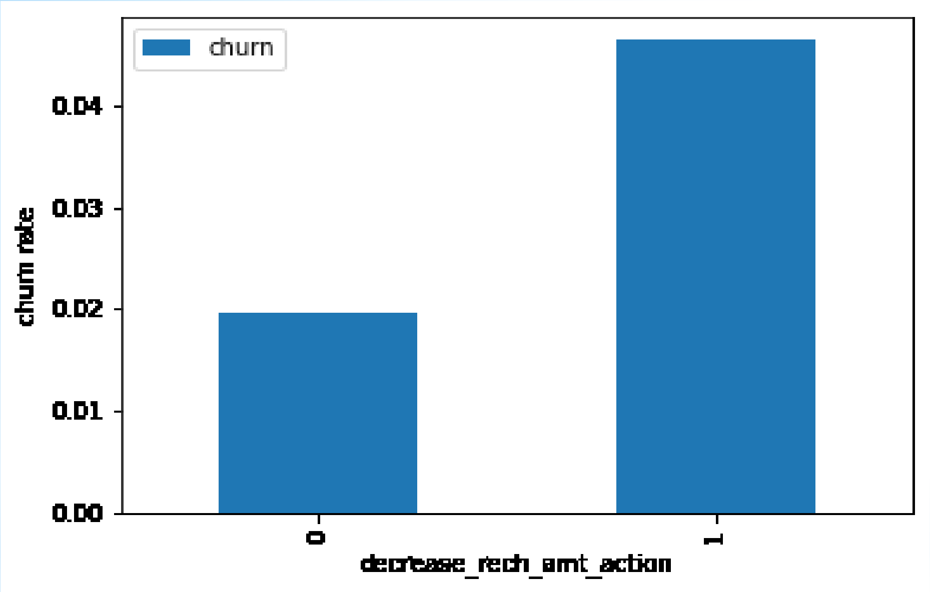
UNIVARIATE ANALYSIS

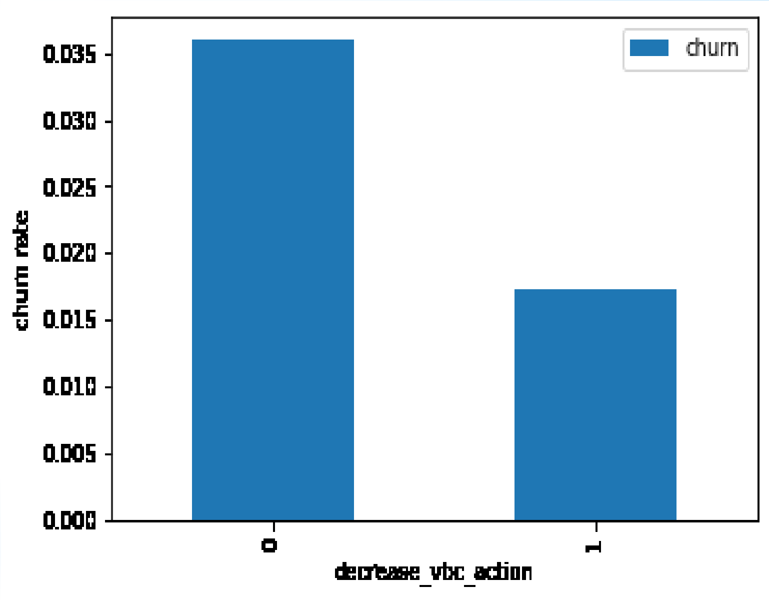




Analysis for decrease_mou_action
We can see that the churn rate is more for the customers, whose minutes of usage(mou) decreased in the action phase than the good phase.

Analysis for decrease_rech_amt_action
As expected, the churn rate is more for the customers, whose number of recharge in the action phase is lesser than the number in good phase.



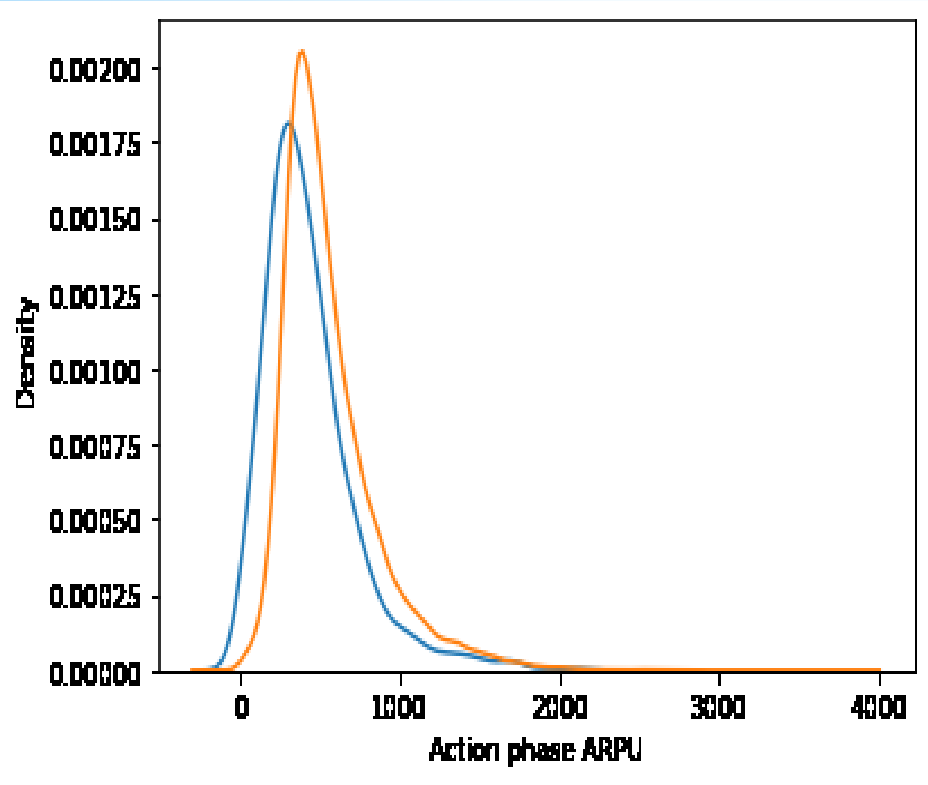


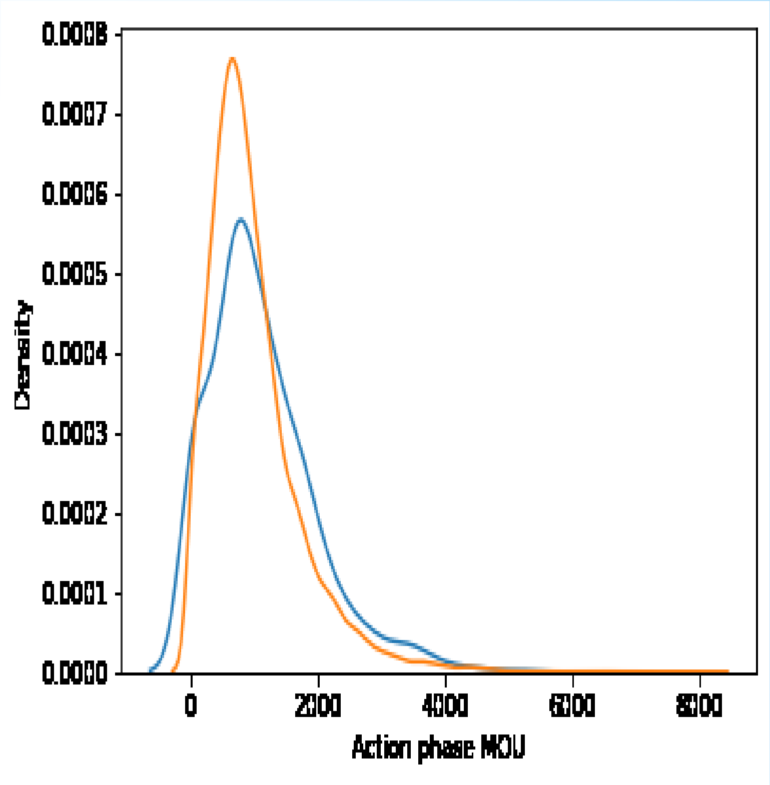
Analysis for decrease_rech_amt_action

Here also we see the same behavior. The churn rate is more for the customers, whose amount of recharge in the action phase is lesser than the amount in good phase.

Analysis for decrease_vbc_action

Here we see the expected result. The churn rate is more for the customers, whose volume based cost in action month is increased. That means the customers do not do the monthly recharge more when they are in the action phase.





Analysis for Action phase ARPU

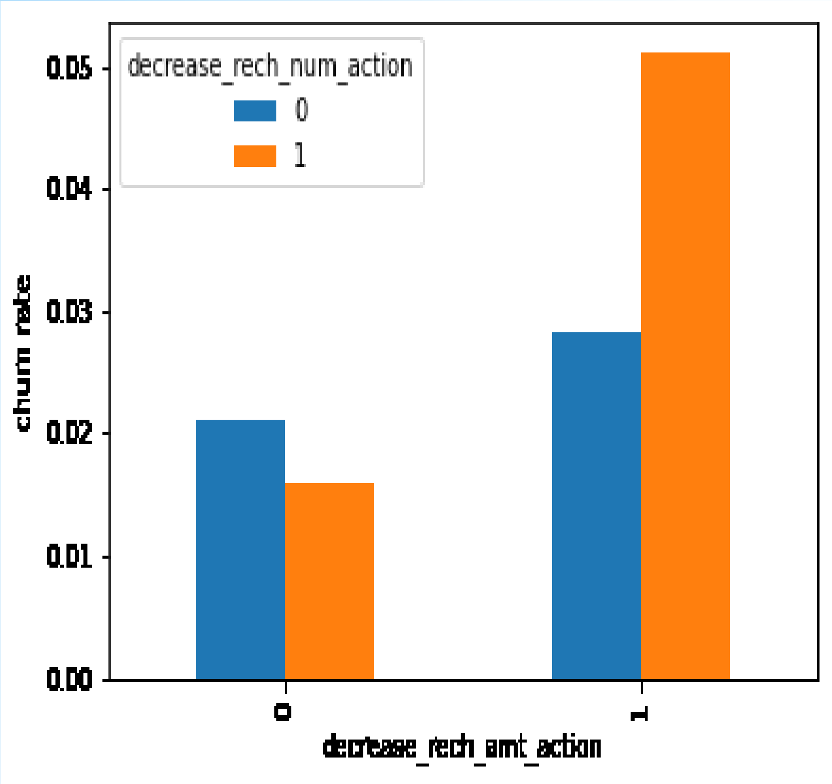
Average revenue per user (ARPU) for the churned customers is mostly densed on the 0 to 900.The higher customers are less likely to be churned.

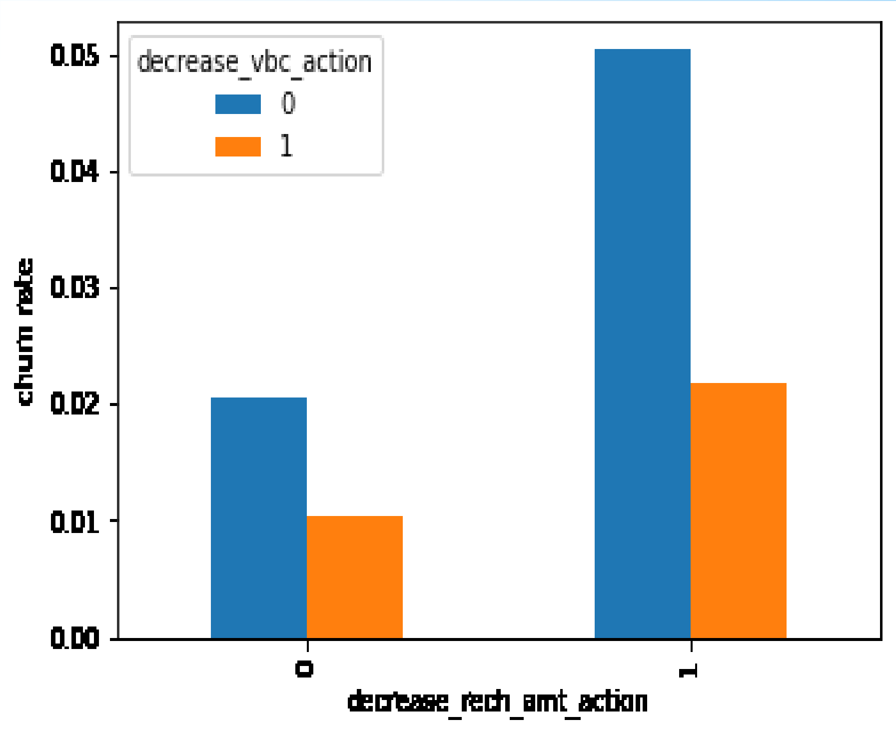
ARPU for the not churned customers is mostly densed on the 0 to 1000.

Analysis for Action phase MOU

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

BIVARIATE ANALYSIS



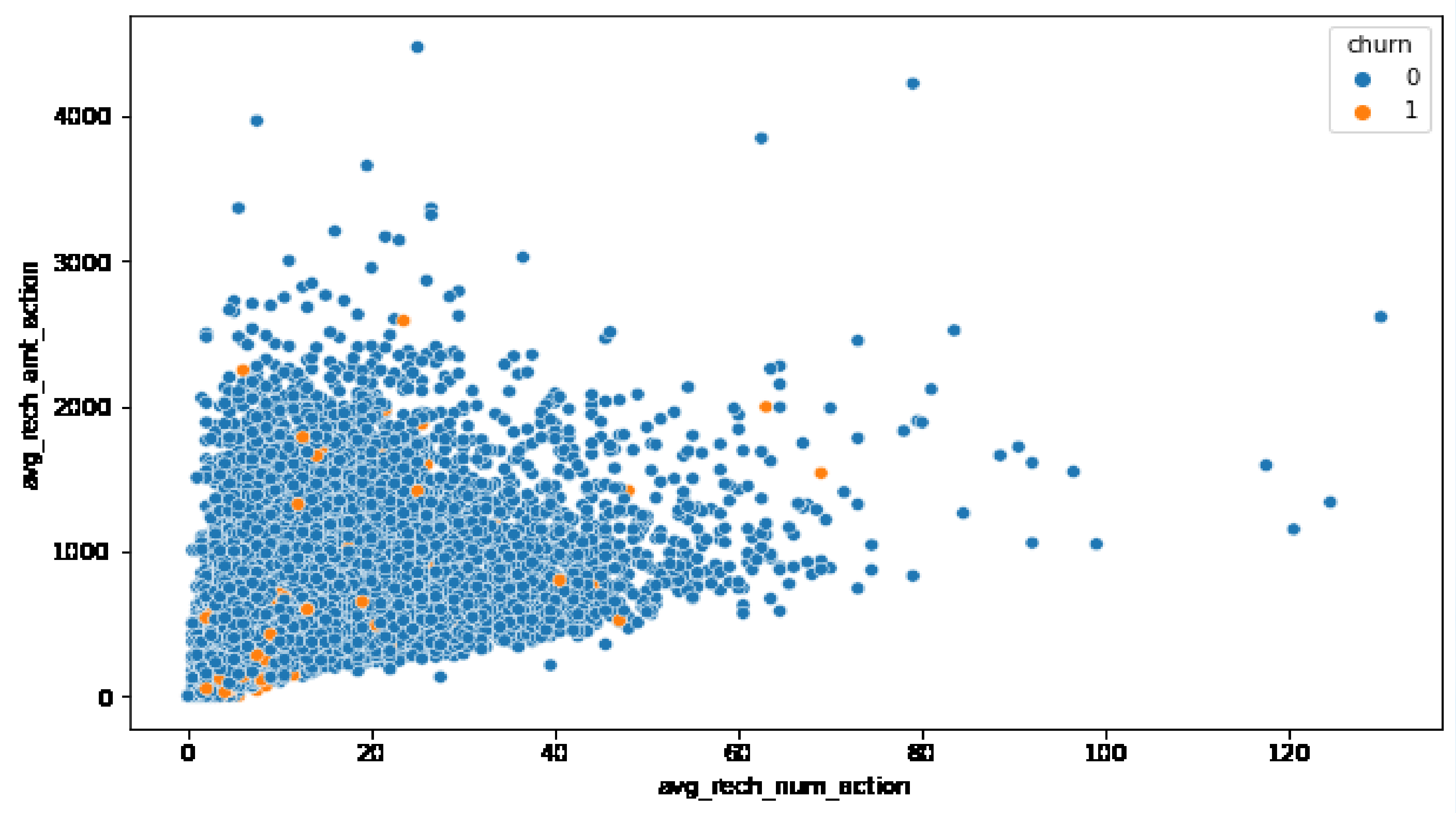


Analysis for decrease_rech_amt_action

We can see 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.

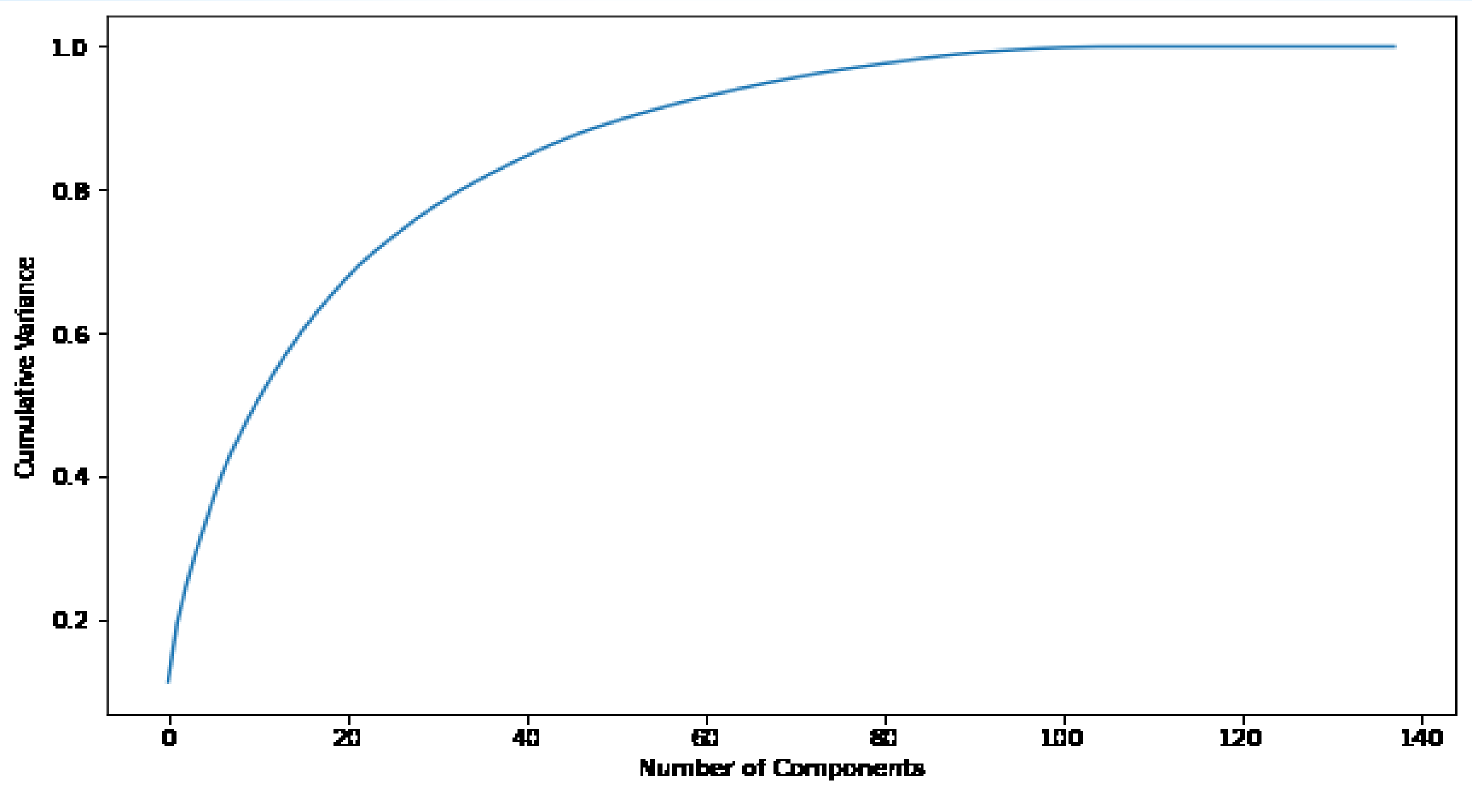
Analysis for decrease_vbc_action

Here, also we can see that the churn rate is more for the customers, whose recharge amount is decreased along with the volume based cost is increased in the action month.



Analysis:

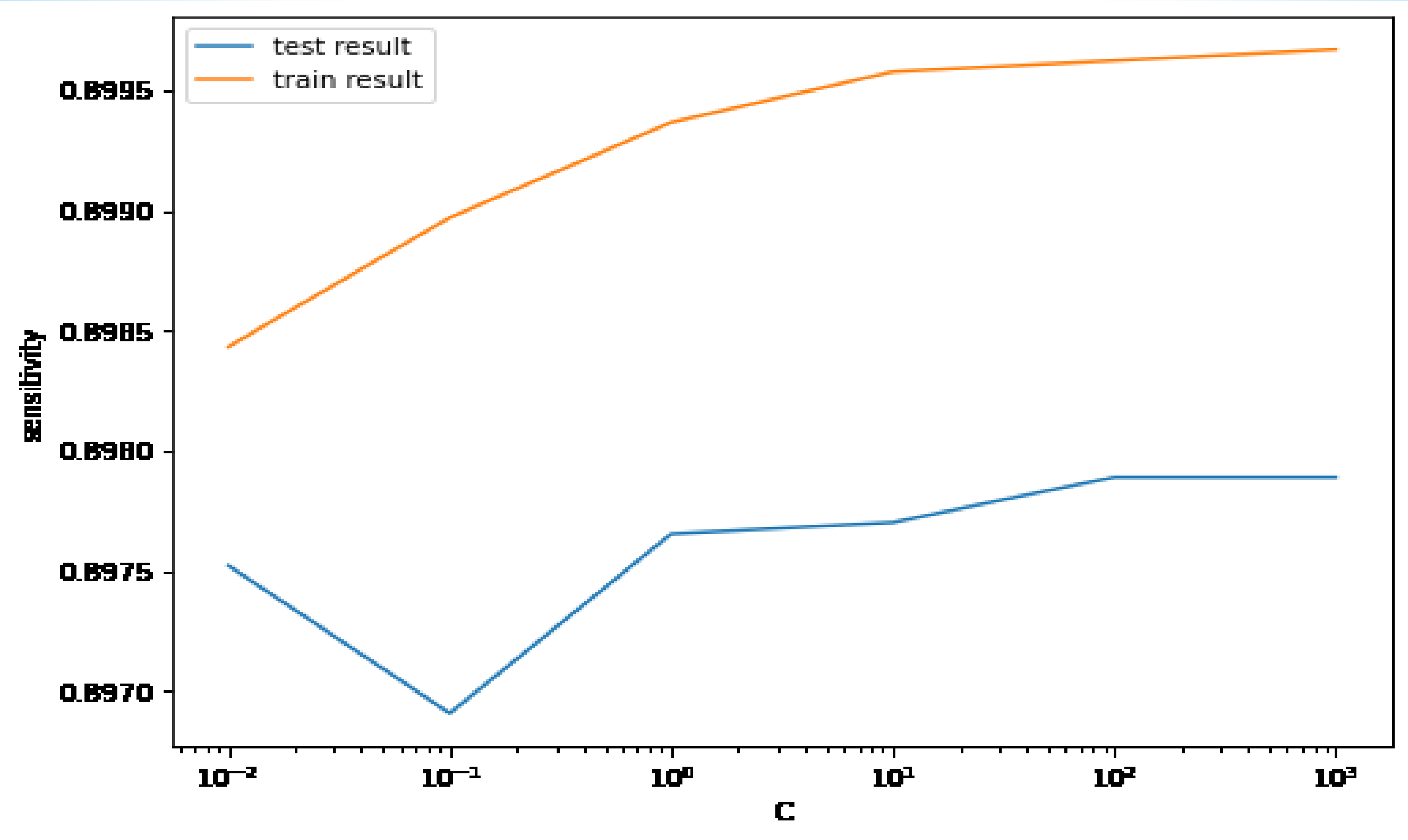
We can see from the above pattern that the recharge number and the recharge amount are mostly proportional. More the number of recharge, more the amount of the recharge.



Analyse:

We can see that 60 components explain almost more than 90% variance of the data. So, we will perform PCA with 60 components.

Plotting Sensitivity Graph



Summary

- Target the customers, whose minutes of usage of the incoming local calls and outgoing ISD calls are less in the action phase (mostly in the month of August).
- Target the customers, whose outgoing others charge in July and incoming others on August are less.
- 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.
- Customers, whose monthly 3G recharge in August is more, are likely to be churned.
- Customers having decreasing STD incoming minutes of usage for operators T to fixed lines of T for the month of August are more likely to churn.
- Customers decreasing monthly 2g usage for August are most probable to churn. Customers having decreasing incoming minutes of usage for operators T to fixed lines of T for August are more likely to churn. roam_og_mou_8 variables have positive coefficients (0.7135). That means for the customers, whose roaming outgoing minutes of usage is increasing are more likely to churn.

THANK YOU