Telecom Churn Case Study

Assignment by

- Mayur Punamiya

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.

METHODOLOGY USED TO DERIVE CONCLUSION

Step-1:

Data Importing, Inspecting, Cleaning & Manipulation
   a) Filtering high-value customers
   b) Tagging Churn Customers
   c) Handling NA or Missing Values.
   d) Dropping of Unnecessary Columns (i.e., which are not taken for in Analysis)
   e) Dropping of Columns having large number of missing values.
   f) Imputation of Values where required.

Step-2:

Data Analysis - Exploration
   ▪ Univariate Analysis.
   ▪ Bivariate data analysis:

METHODOLOGY USED TO DERIVE CONCLUSION

Step-3:

Model Building Preparation & Validation
- Test-Train Split
- Scaling

Step-4:

Model Evaluation
- Creating a data frame with the actual conversion flag and predicted probabilities
- Creating new column 'Predicted'
- Finding the Optimal Cutoff

**Step-5:**

Making Predictions based on the Test Set

Step-6:

Deriving Conclusion & Recommendation based on Model.

Data Cleaning and Preparation

➢ Firstly, we dropped all the columns which had more than 30% values missing or NA.
➢ We Checked the other remaining columns and drop columns which are not required for our analysis

➢ Next the columns with null or missing values, we imputed them with 0.0
➢ We Checked the other remaining columns and drop columns which are not required for our analysis
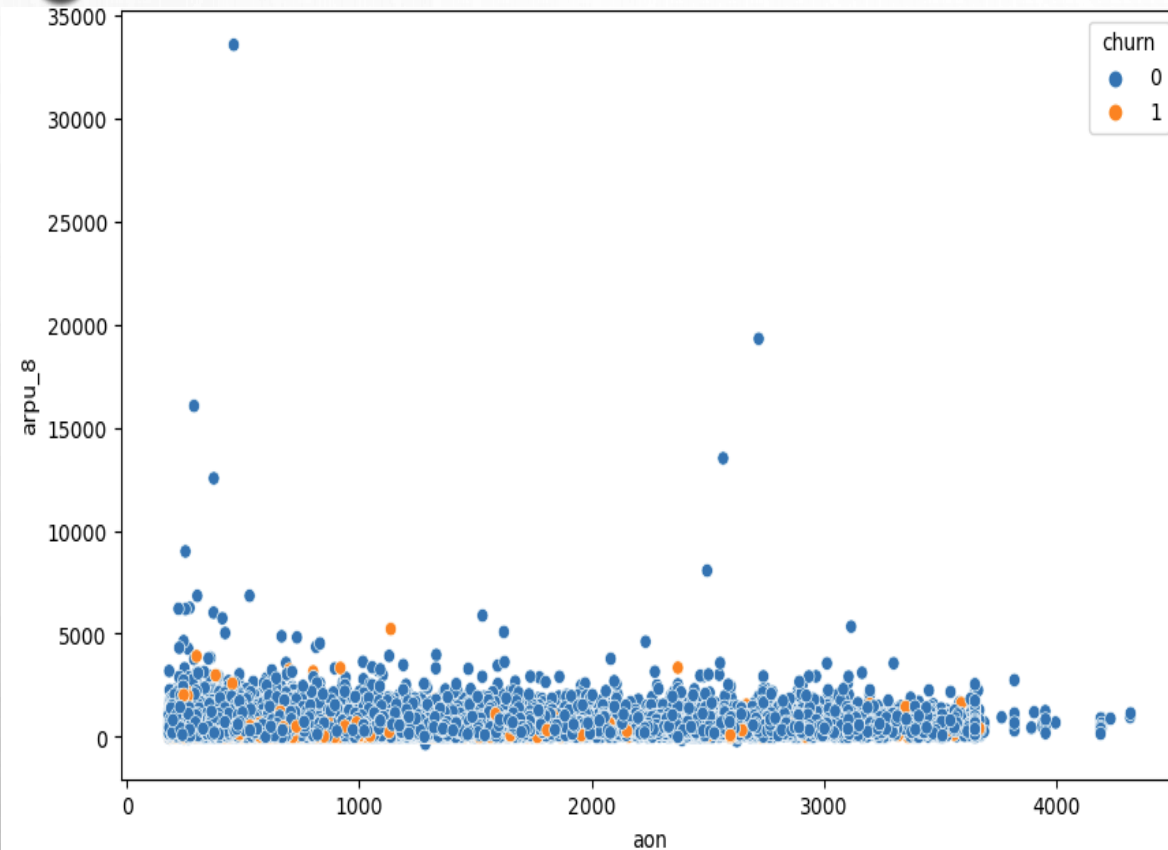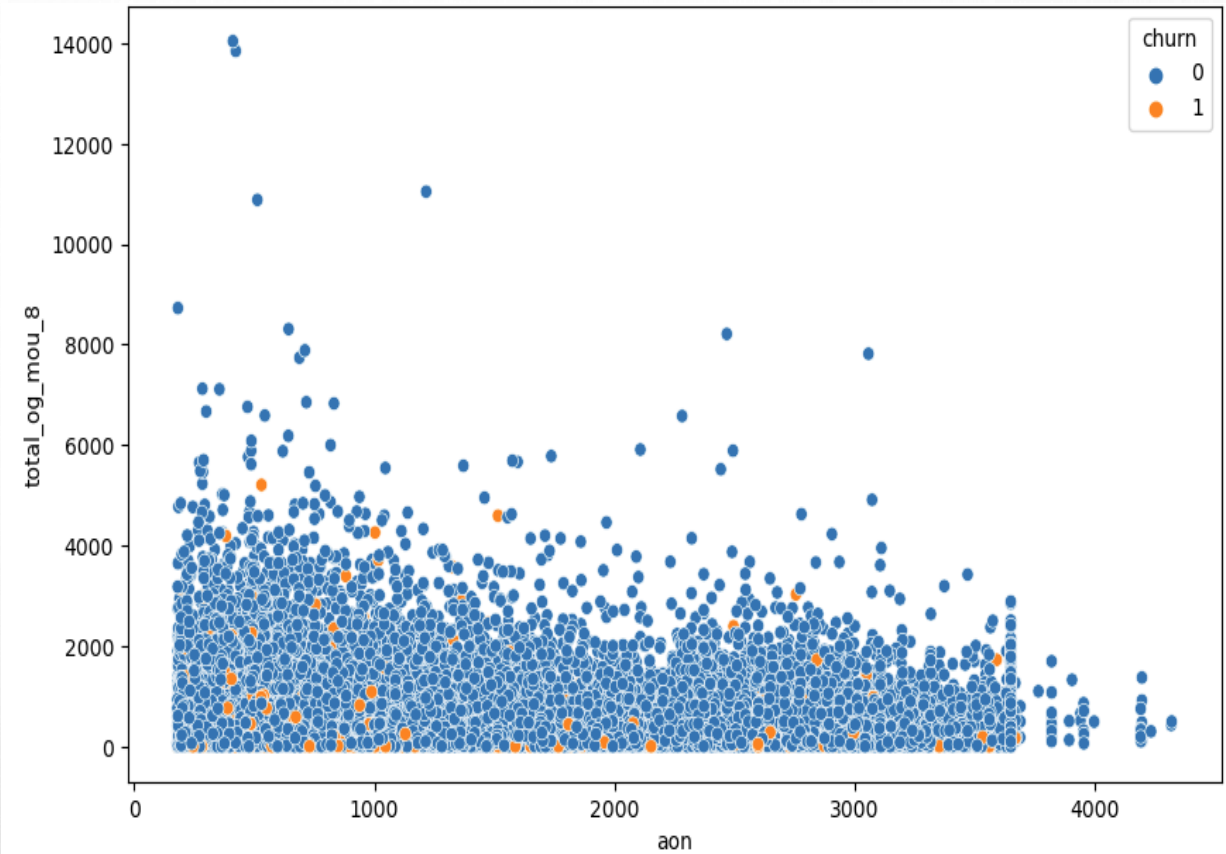
DATA EXPLORATION (EDA) FINDINGS
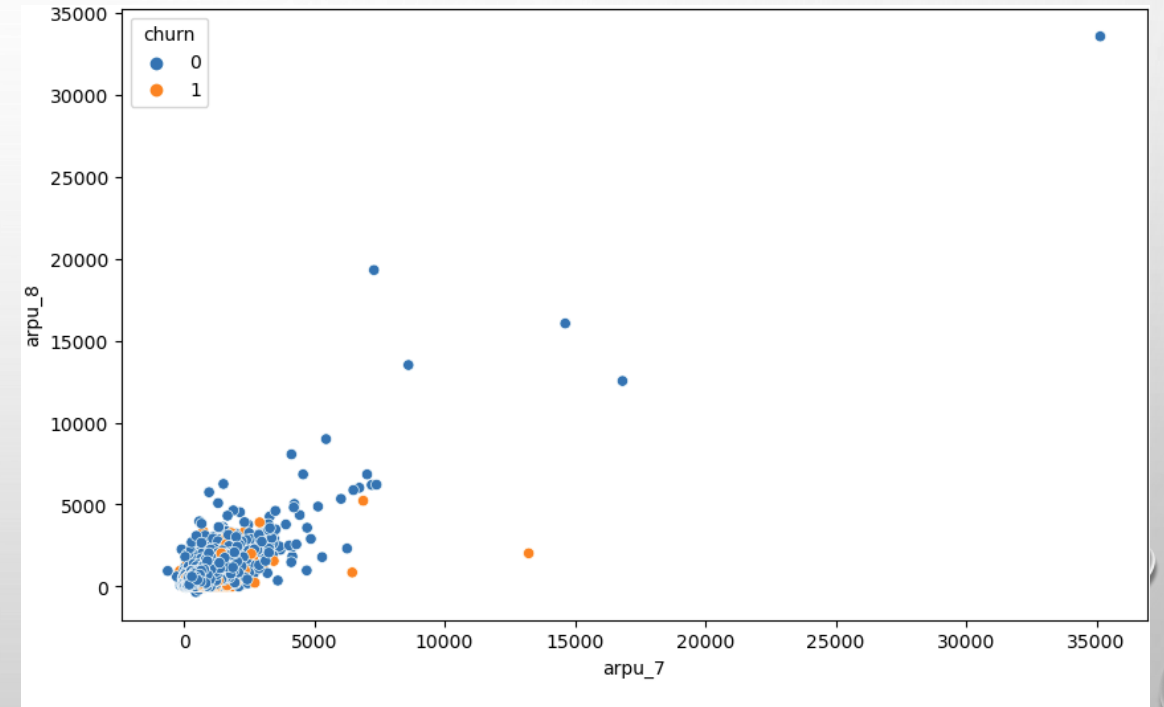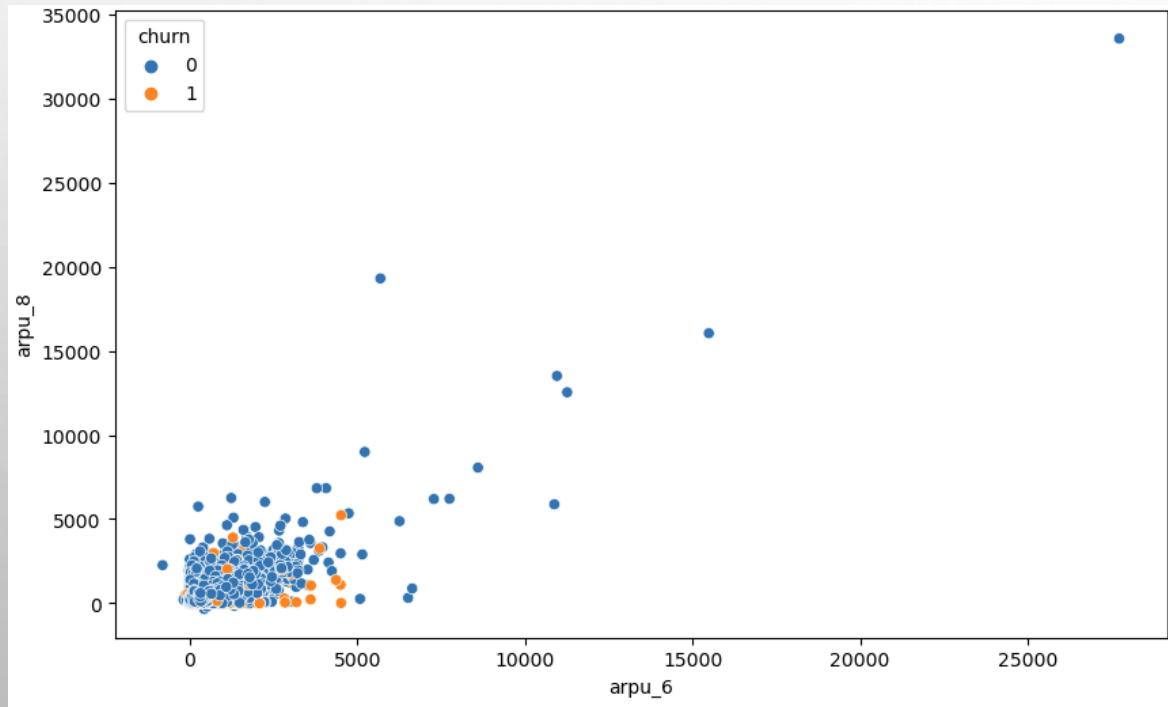
What Chart Shows:
Customers with lesser "arpu" in 6,7 are more lickely to churn'
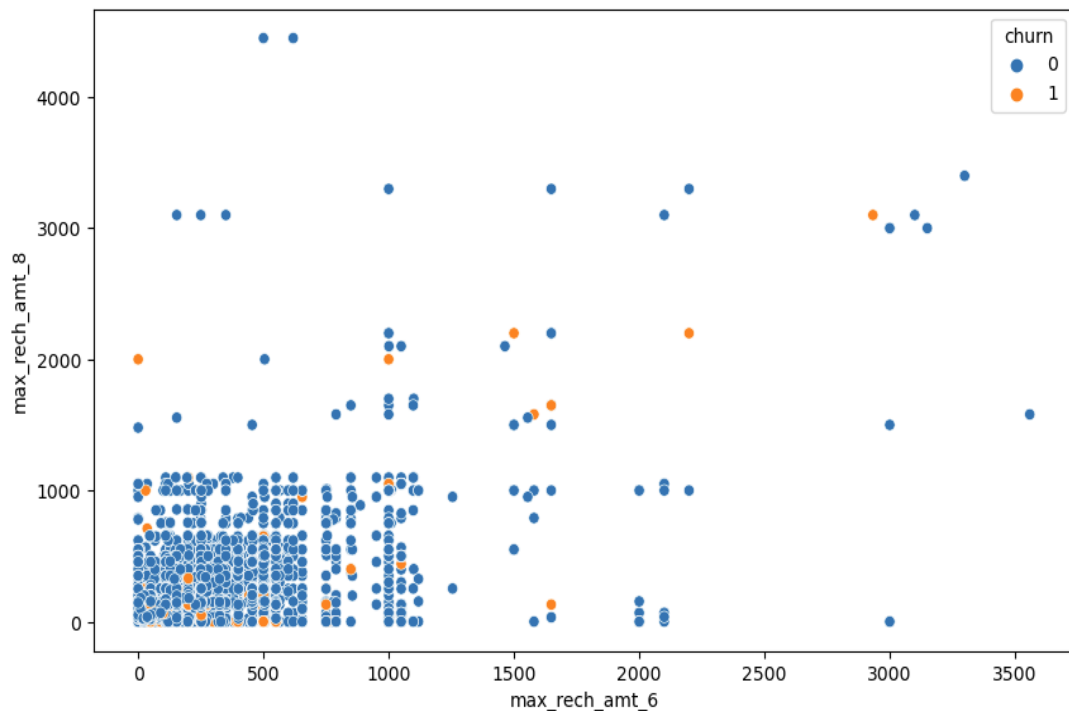
DATA EXPLORATION (EDA) FINDINGS

What Chart Shows:
Customers with lesser "arpu_8" and "aon" are more likely to churn.'
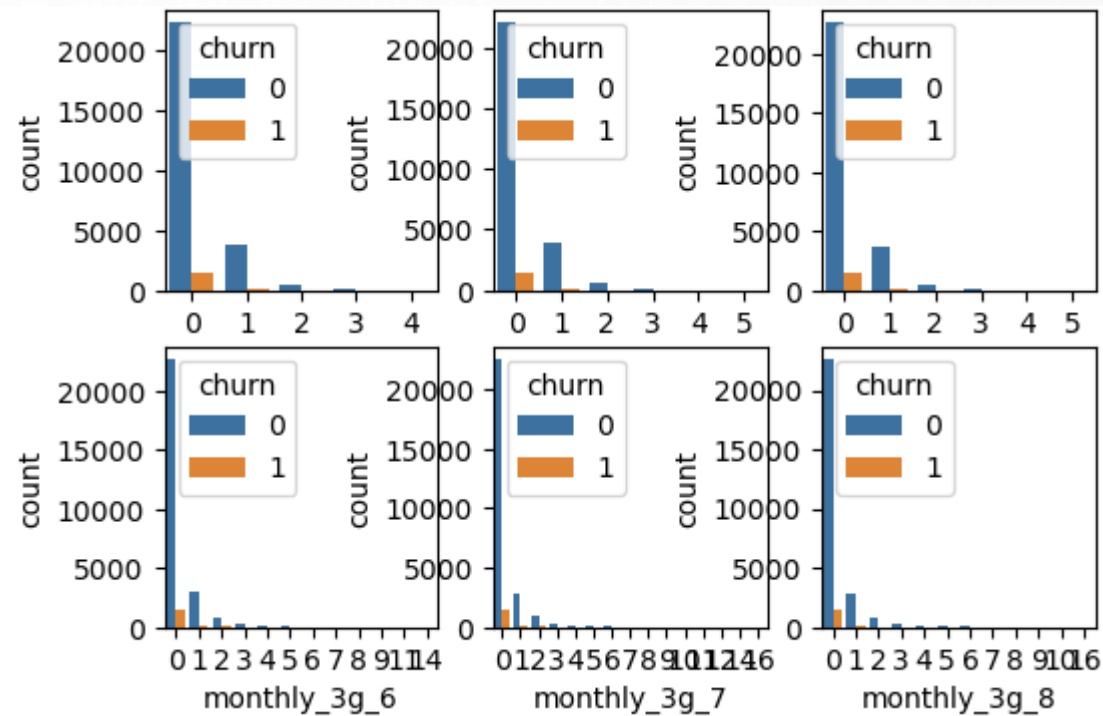
What Chart Shows:
Customers with lesser "total_og_mou_8" and "aon" are more likely to churn.

DATA EXPLORATION (EDA) FINDINGS

What Chart Shows:
Lesser "max_rech_amt_6" and "max_rech_amt_8" indicates higher churn chances

What Chart Shows:
Higher use of 2g or 3g indicates lesser chances of churn

MODEL BUILDING

- ✓ Splitting the Data into Training and Testing Sets.
- ✓ The first basic step for regression is performing a train-test split, we have chosen 70:30 ratio.
- ✓ Generalized Logistic Regression Results.
- ✓ Feature Selection Using RFE.
- ✓ Building Model.
- ✓ Assessing the model.
- ✓ Predictions made based on test data set.

PLOTTING ROC CURVE
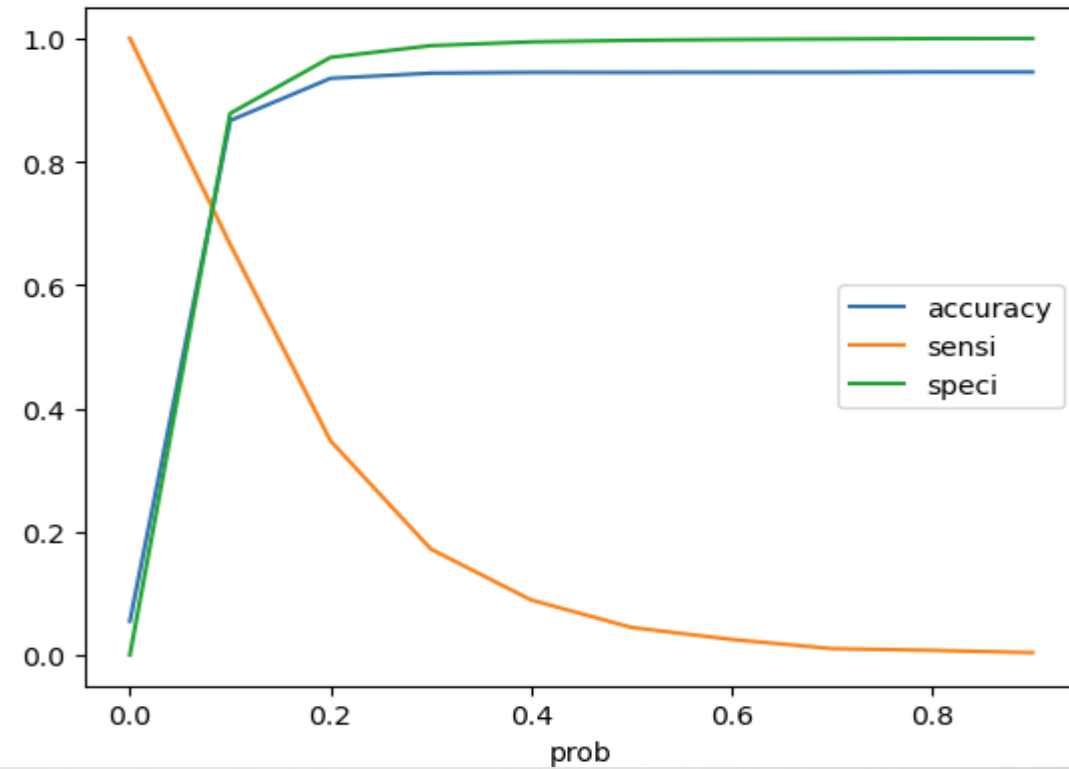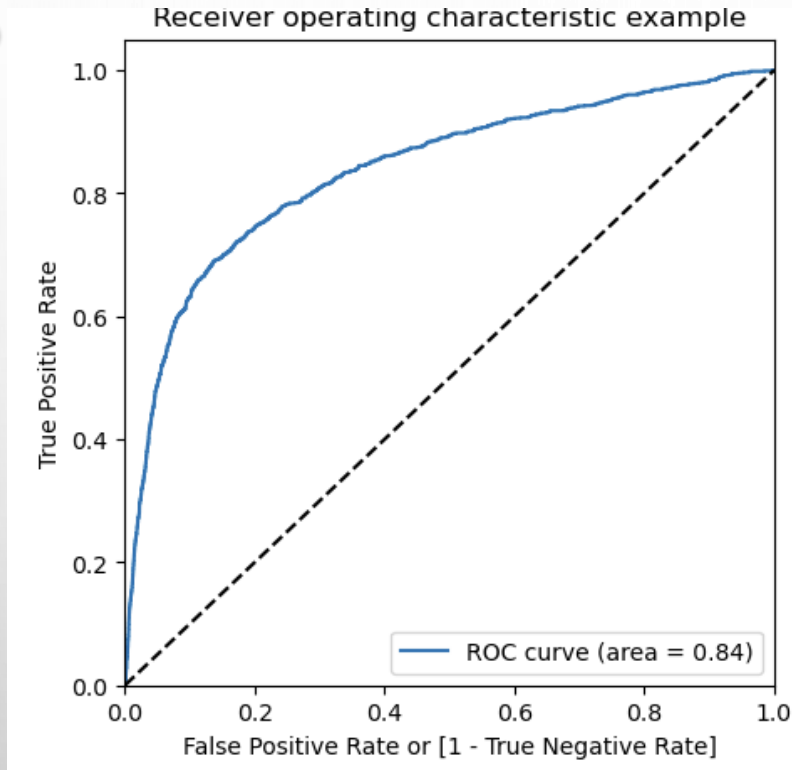
❖ Since we know that the perfect ROC Curve should be a value close to 1. We are getting a value of 0.84 indicating a good predictive model.

❖ From the curve above, we see that 0.15 is the optimum point to take it as a cutoff probability.

CONCLUSIONS & PREDICTIONS

➤ So using Logistic regression we are getting an accuracy of 94% on train data and 92% on test data.

➤ Roaming outgoing in month 7 and 8 are strong indicators, hence telecom company should reduce the costs for roaming. They need to provide good offers to the customers who are using services from a roaming zone.

➤ MOU is one of the major factors.

➤ Max recharge in the good phase are indicators of churn.

➤ We can clearly see most of the critical features are form the action phase, which is inline with the business understanding that action phase needs more attention.

```
roam_og_mou_7          0.002930
loc_ic_t2m_mou_7       0.002800
roam_og_mou_8          0.001733
total_og_mou_7         0.000665
aon                   -0.000230
vol_2g_mb_8           -0.001532
total_rech_amt_8      -0.002181
loc_ic_mou_8          -0.006970
const                 -1.554039
```

The image attached shows the important parameters for predicting churn.