PRESENTATION

TELECOM CHURN CASE STUDY

SHASHANK JAISWAL,
PIYUSH NARAYAN ,
ANJALI

TABLE OF CONTENTS

1. *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.

2. Analysis Approach

- *Here we are given with 4 months of data related to customer usage. In this case study, we analyse 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.*
- *Churn is predicted using two approaches. Usage based churn and Revenue based churn. Usage based churn:*
- *Customers who have zero usage, either incoming or outgoing - in terms of calls, internet etc. over a period of time.*
- *This case study only considers usage based churn.*
- *In the Indian and the southeast Asian market, approximately 80% of revenue comes from the top 20% customers (called high-value customers). Thus, if we can reduce churn of the high-value customers, we will be able to reduce significant revenue leakage. Hence, this case study focuses on high value customers only.*
- *The 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 objective is to predict the churn in the last (i.e. the ninth) month using the data (features) from the first three months.*
- *This is a classification problem, where we need to predict whether the customers is about to churn or not. We have carried out Baseline Logistic Regression, then Logistic Regression with PCA, PCA + Random Forest, PCA + XGBoost.*

3 Analysis Steps

- *Reading, understanding cleaning and visualising the data*

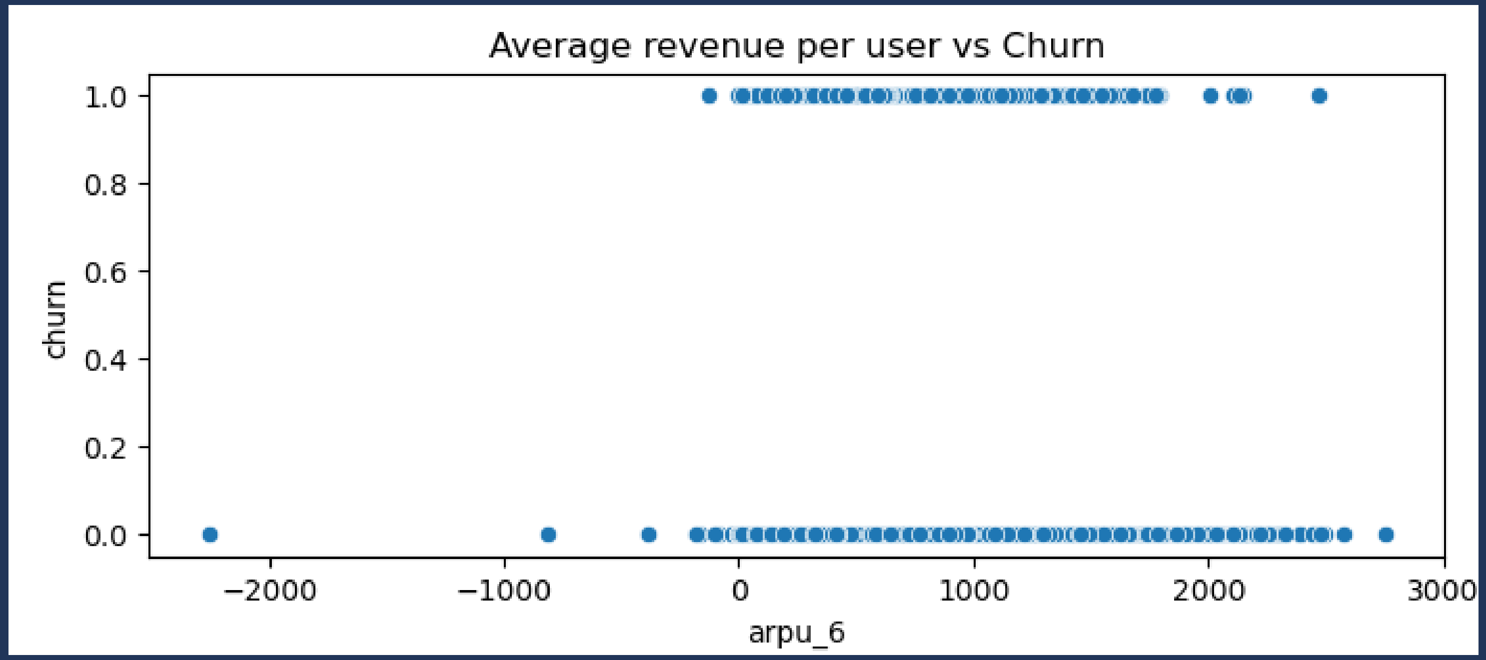
  In this step we loaded all the necessary libraries and loaded the data set in data frame.
  After this we identified shape of the data, number of rows columns, null values and unique values as well.

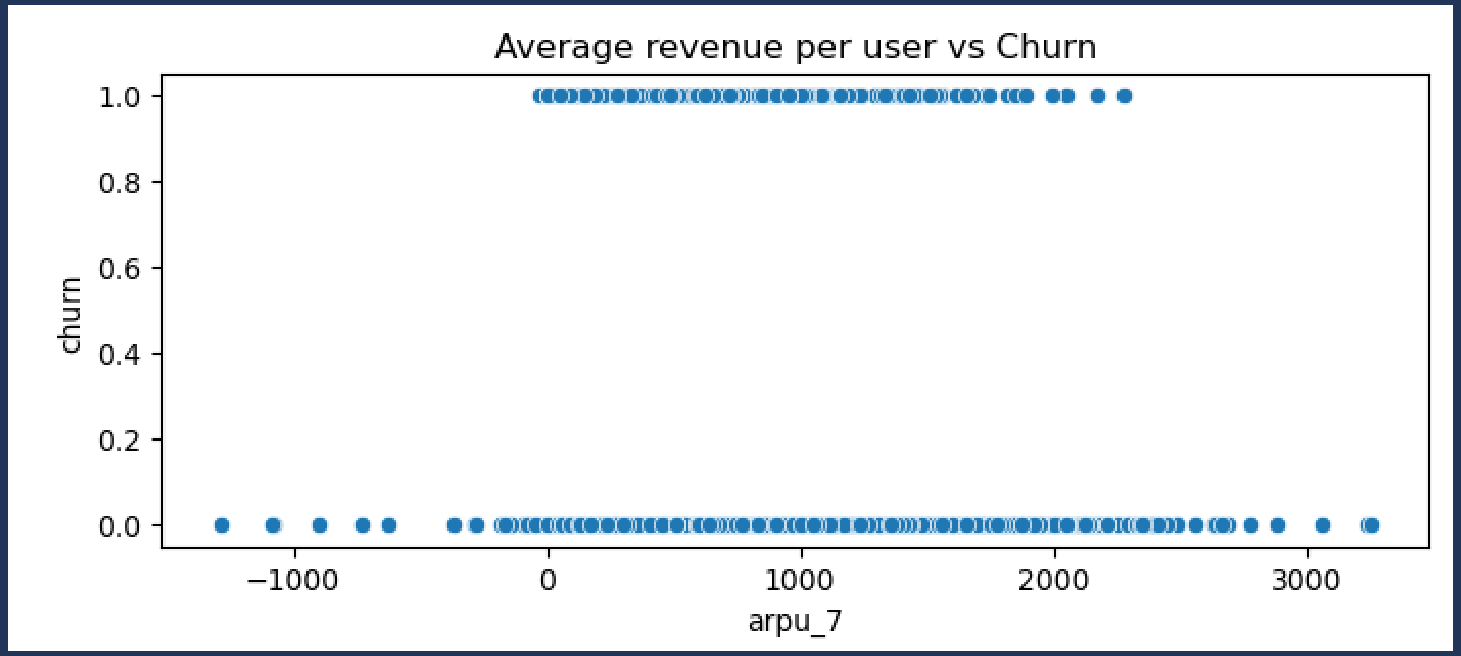
- *Preparing the data for modelling*

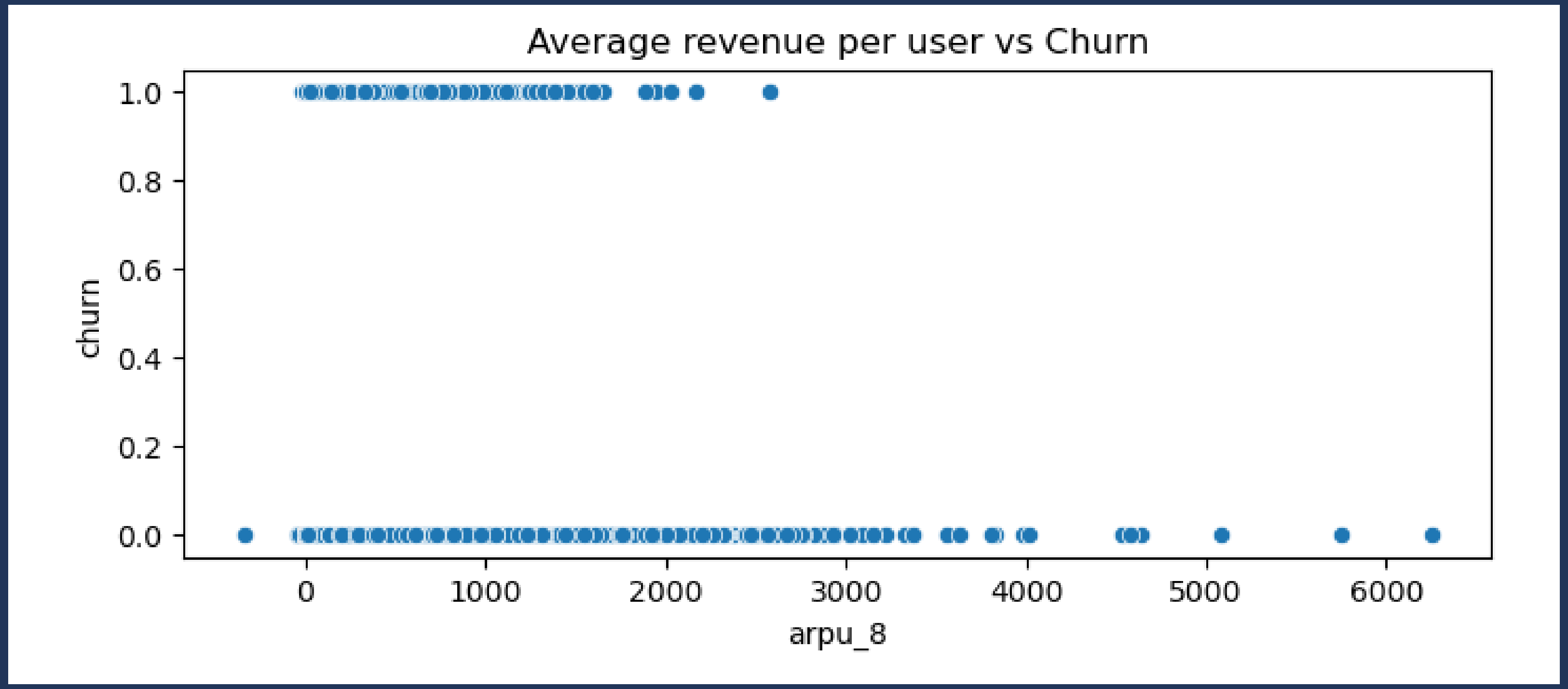
  In this step we do EDA. Carrying out steps like Univariate analysis and get the Churn rate on the basis whether the customer decreased her/his MOU in action month.

- *Building the model*
- *Evaluate the model*

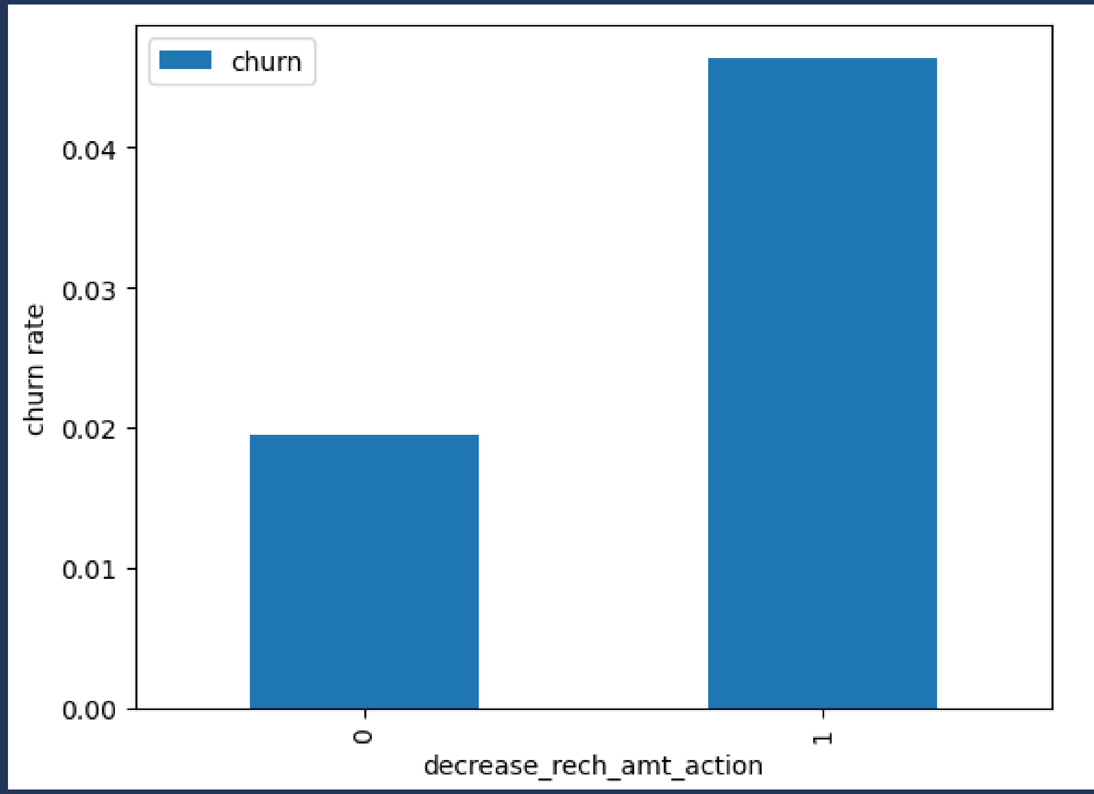
3 Results of Univariative Analysys

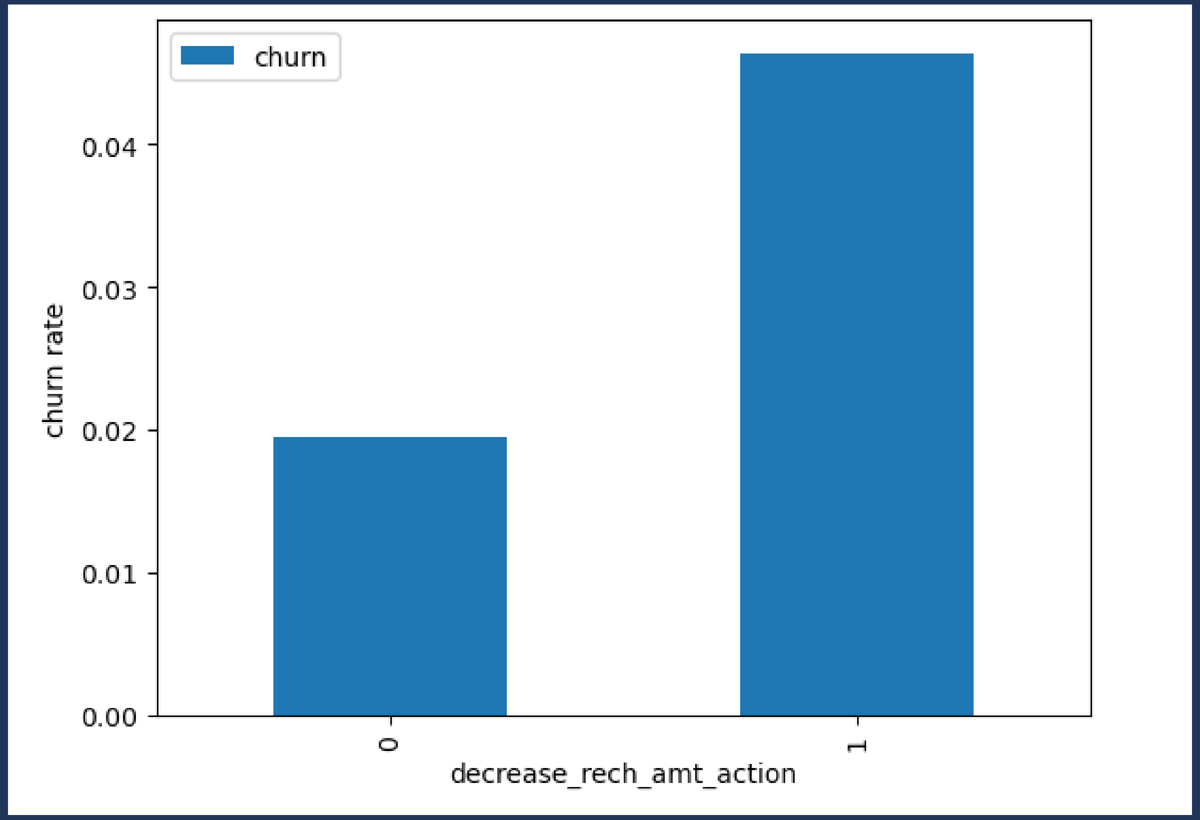


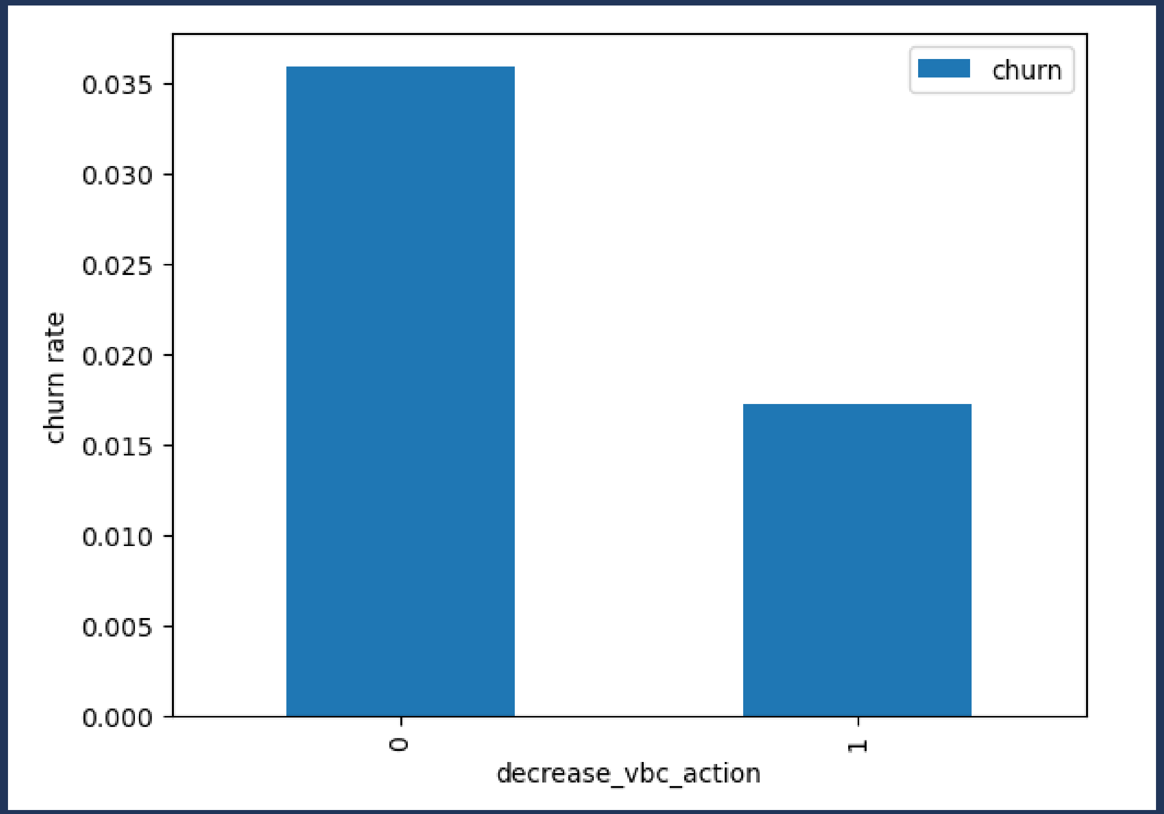




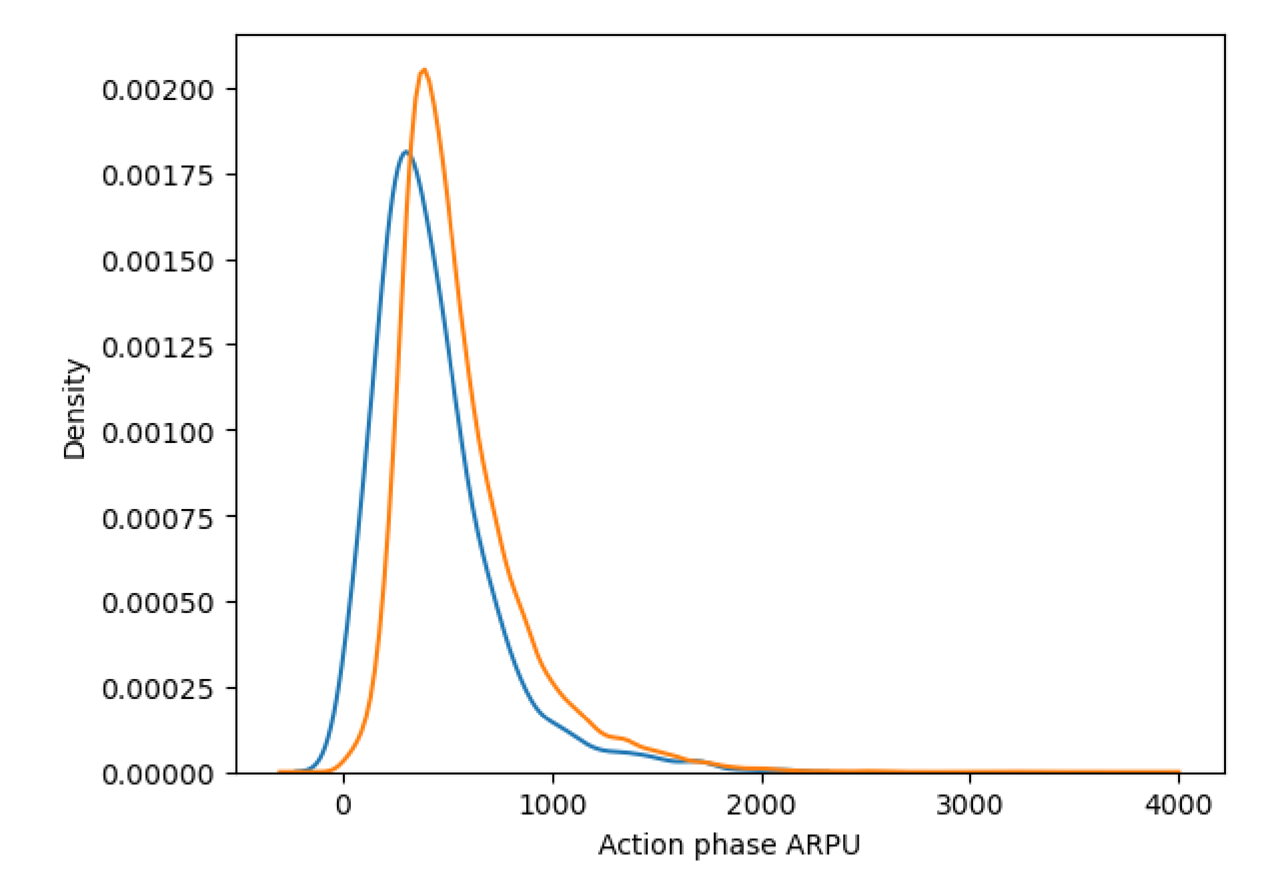
Pivot tables

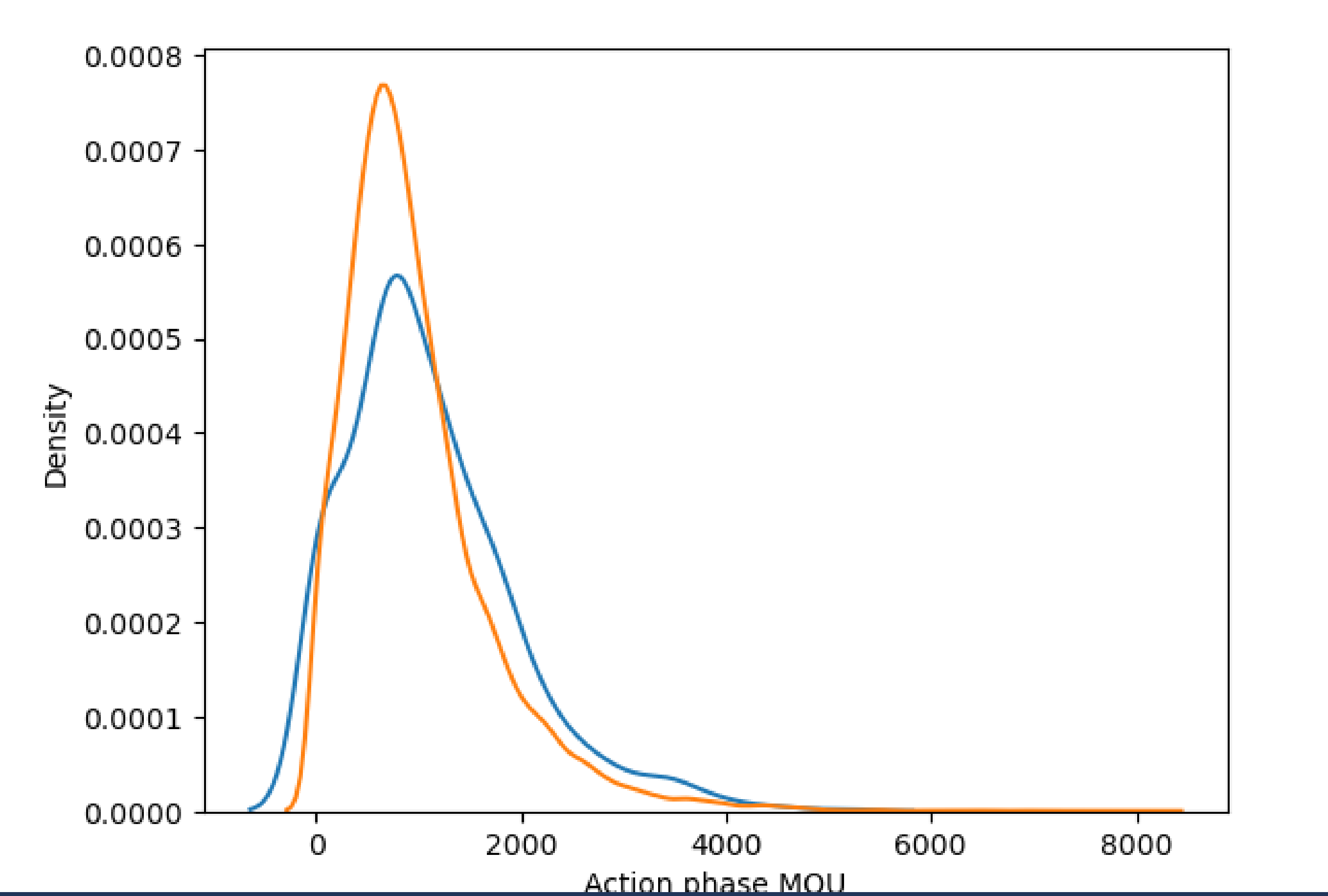






Creating Churn Database, using distribution plot





3 Preprocessing

1. Train-Test Split has been performed.
2. The data has high class-imbalance with the ratio of 0.095 (class 1 : class 0).
3. SMOTE technique has been used to overcome class-imbalance.
4. Predictor columns have been standardized to mean - 0 and standard_deviation- 1.

5 Modelling

Model summary*

Train set

1. Accuracy = 0.84
2. Sensitivity = 0.81
3. Specificity = 0.83
4. Test set
5. Accuracy = 0.78
6. Sensitivity = 0.82
7. Specificity = 0.78

Overall, the model is performing well in the test set, what it had learnt from the train set.

1. Final conclusion with no PCA

2. We can see that the logistic model with no PCA has good sensitivity and accuracy, which are comparable to the models with PCA. So, we can go for the more simplistic model such as logistic regression with PCA as it expliains the important predictor variables as well as the significance of each variable. The model also hels us to identify the variables which should be act upon for making the decision of the to be churned customers. Hence, the model is more relevant in terms of explaining to the business.

5. Recommendation

- 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.
- Also, 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.
- Cutomers, 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.
- Cutomers 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.