TELECOM CHURN CASE STUDY

YOGESH BHARDWAJ
AMRIN MOHAMMED
UJJWAL KUMAR

THE 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.**

BUSINESS PROBLEM OVERVIEW

In this project, 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.

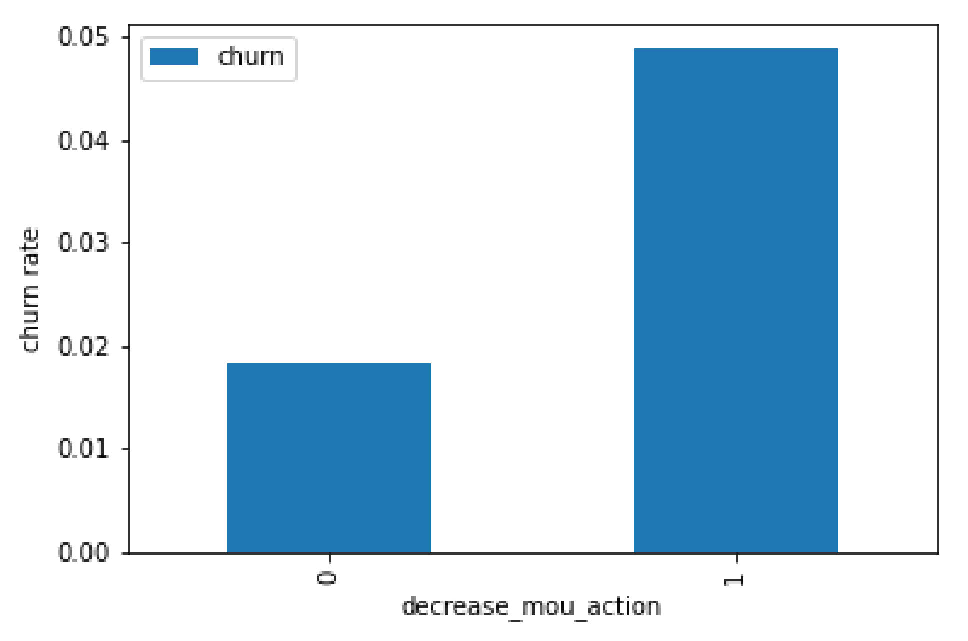
ANALYSIS APPROACH

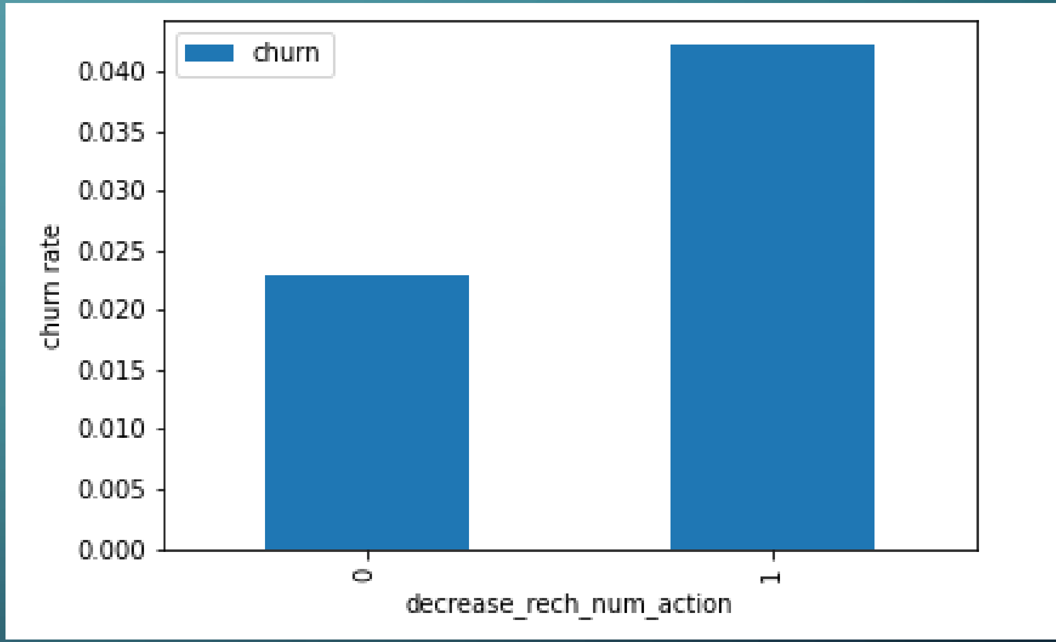
- We are provided with 4 months of data related to customer usage. In this project we will analyse the data, build predictive models to identify customers at high risk of churn and to know the main indicators of churn.
- The dataset contains customer-level information for a span of four consecutive months - 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. To do this task well, understanding the typical customer behaviour during churn will be helpful.

ANALYSIS STEPS

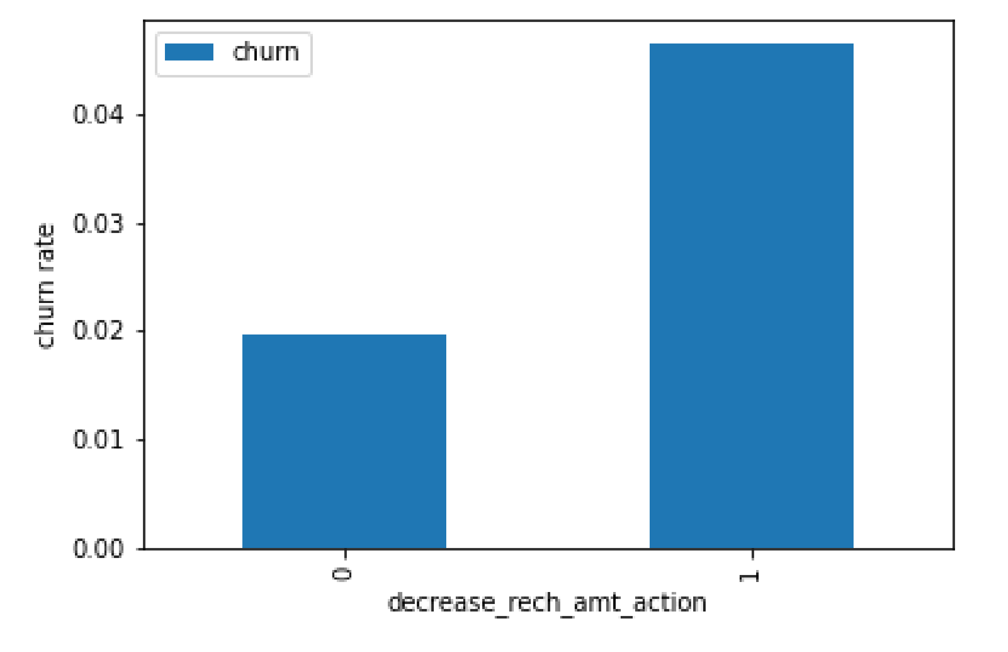
1. Reading, understanding, cleaning and visualising the data. Here we import the necessary libraries, load the data in a data frame and get to know about it's overview including the missing values.
2. Preparing the data for modelling. Here we do EDA including Univariate Analysis, Outlier treatment and get the churn rate.
3. Building different predictive models
4. Evaluating the models.

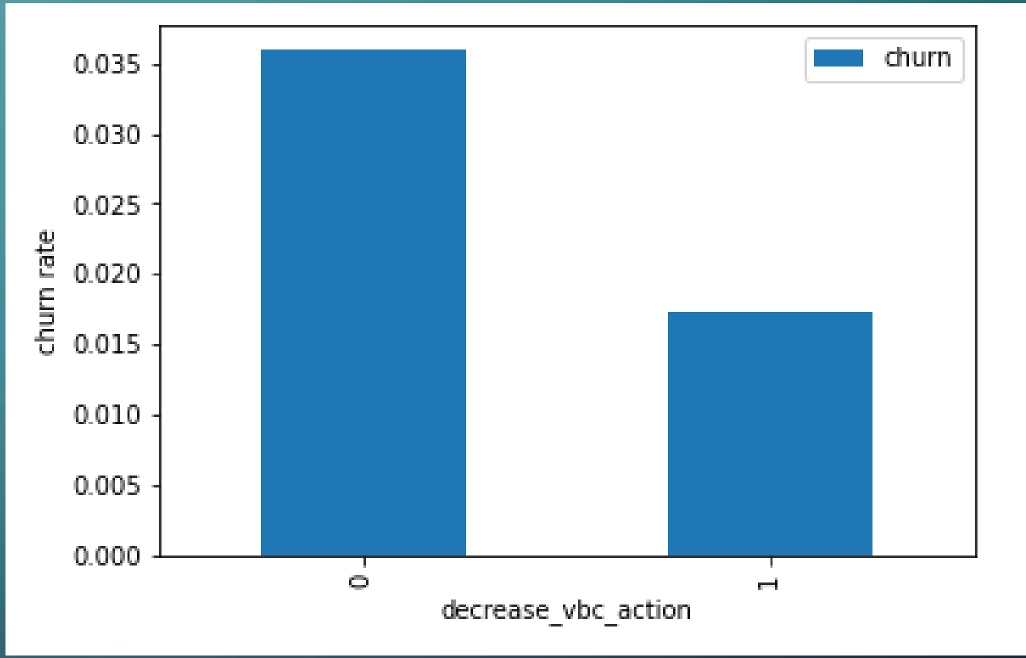
UNIVARIATE ANALYSIS



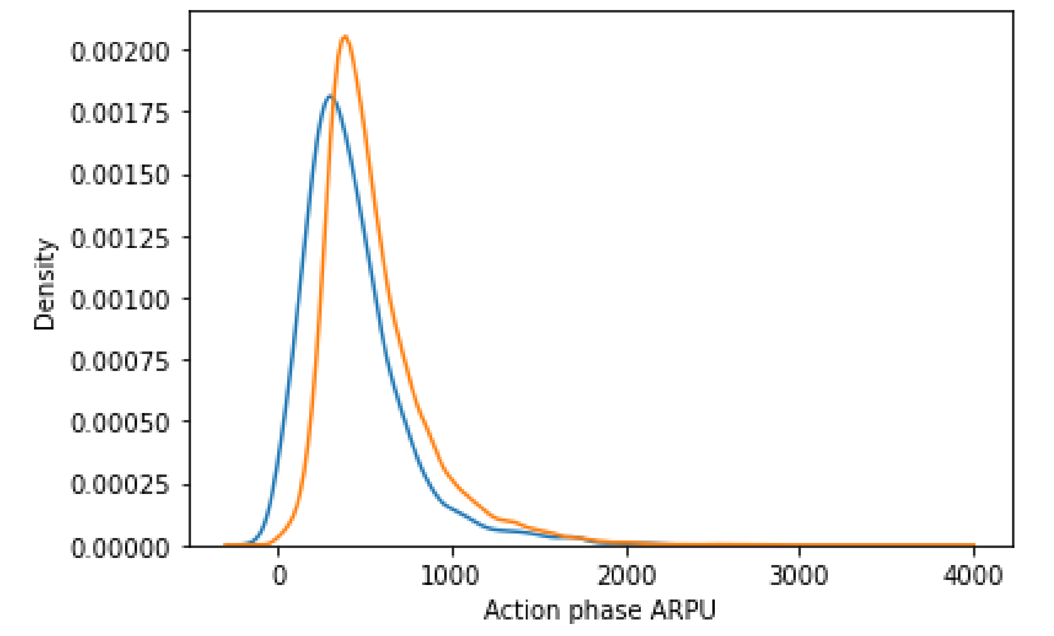


UNIVARIATE ANALYSIS





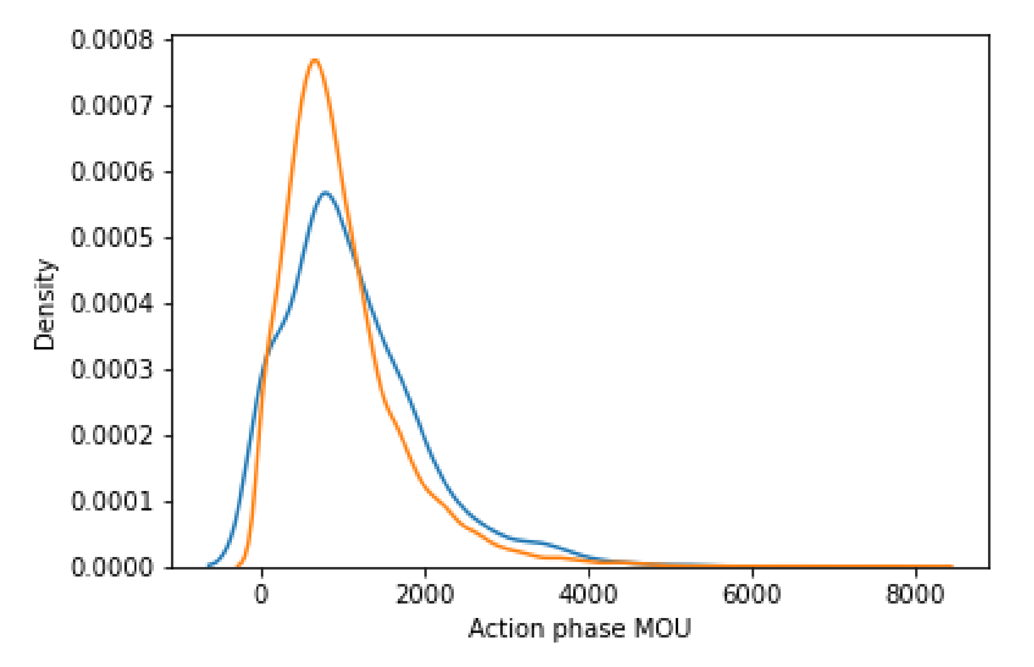
DISTRIBUTION PLOTS



- 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.

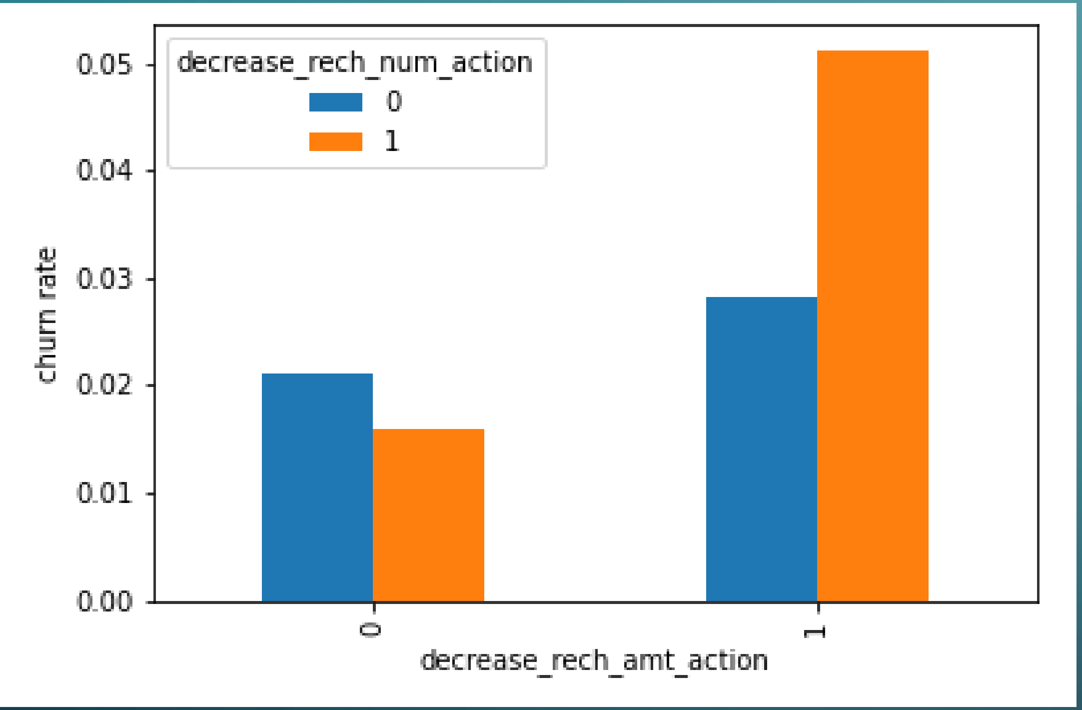
- ARPU for the not churned customers is mostly densed on the 0 to 1000.

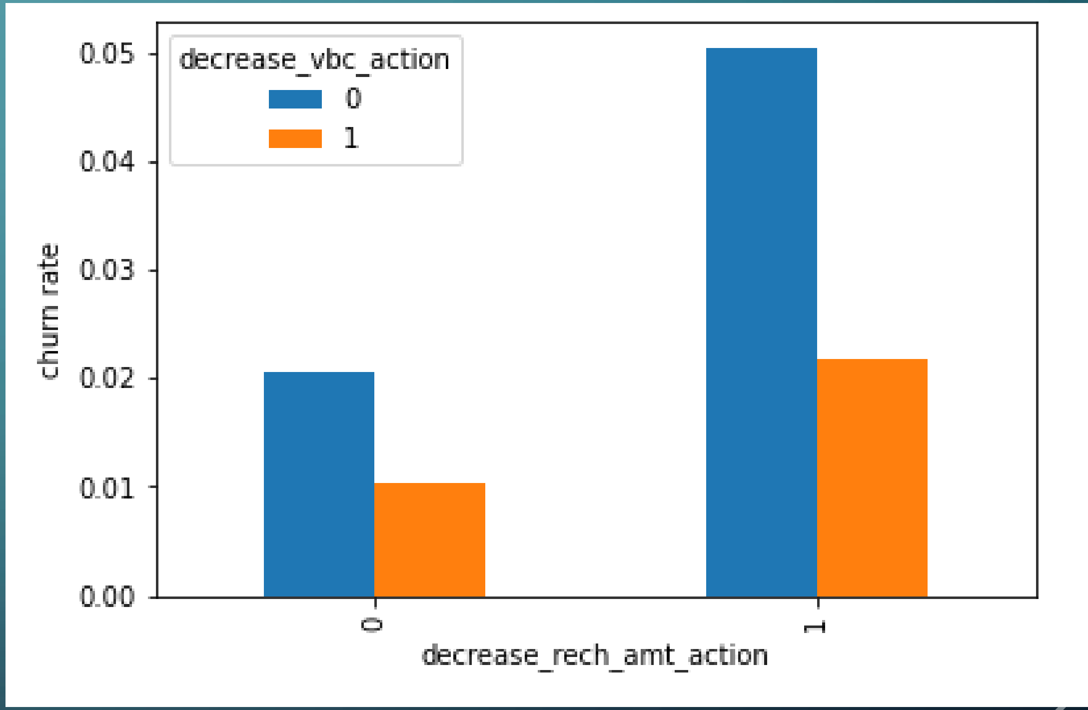
DISTRIBUTION PLOTS



- 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





PREPROCESSING

- Train-Test Split has been performed.
- The data has high class imbalance with a ratio of 0.095 (Class 1 : Class 0)
- SMOTE technique has been used to overcome class imbalance
- Predictor columns have been standardized to mean 0 and standard deviation 1.

MODELLING

Model Summary

- Train Set:
  - Accuracy = 0.84
  - Sensitivity = 0.81
  - Specificity = 0.83
- Test Set:
  - Accuracy = 0.84
  - Sensitivity = 0.76
  - Specificity = 0.85

MODELLING

Final conclusion with no PCA

- 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 explains the important predictor variables as well as the significance of each variable. The model also helps 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.

RECOMMENDATIONS

1. 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).
2. Target the customers, whose outgoing others charge in July and incoming others on August are less.
3. 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.
4. Customers whose monthly 3G recharge in August is more, are likely to be churned.
5. 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.
6. Customers decreasing monthly 2g usage for August are most probable to churn.
7. Customers having decreasing incoming minutes of usage for operators T to fixed lines of T for August are more likely to churn.
8. 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