Telecom Churn – Domain Oriented Case Study

To predict the churn in the ninth month using the data (features) from the first three months

GROUP MEMBERS:

ABHISHEK CHAUHAN

ABHISHEK SHARMA

ANUJA PATIL

Problem Statement

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

- 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 become even more important than customer acquisition.
- Here we are given with 4 months of data related to customer usage. In this case study, we analyses 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.

Data Cleaning and EDA Analysis Steps

1. We have started with importing Necessary packages and libraries.
2. We have loaded the dataset into a data frame.
3. We have checked the number of columns, their data types, Null count and unique value_value_count to get some understanding about data and to check if the columns are under correct data-type.
4. Checking for duplicate records (rows) in the data. There were no duplicates.
5. Since 'mobile_number' is the unique identifier available, we have made it our index to retain the identity.
6. Have found some columns that do not follow the naming standard, we have renamed those columns to make sure all the variables follow the same naming convention.
7. Following with column renaming, we have dealt with converting the columns into their respective data types. Here, we have evaluated all the columns which are having less than or equal to 29 unique values as categorical columns and rest as continuous columns.
8. The date columns were having 'object' as their data type, we have converted to the proper datetime format.

9. Since, our analysis is focused on the HVC(High value customers), we have filtered for high value customers to carryout the further analysis. The metric of this filtering of HVC is such that all the customers whose 'Average_rech_amt' of months 6 and 7 greater than or equal to 70th percentile of the 'Average_rech_amt' are considered as High Value Customers.

10.Checked for missing values.

11.Dropped all the columns with missing values greater than 50%.

12.We have been given 4 months data. Since each month's revenue and usage data is not related to other, we did month-wise drill down on missing values.

13.Some columns had similar range of missing values. So, we have looked at their related columns and checked if these might be imputed with zero.

14.We have found that 'last_date_of_the_month' had some missing values, so this is very meaningful and we have imputed the last date based on the month.

15.We have found some columns with only one unique value, so it is of no use for the analysis, hence we have dropped those columns.

16.Once after checking all the data preparation tasks, tagged the Churn variable(which is our target variable).

17.After imputing, we have dropped churn phase columns (Columns belonging to month - 9).

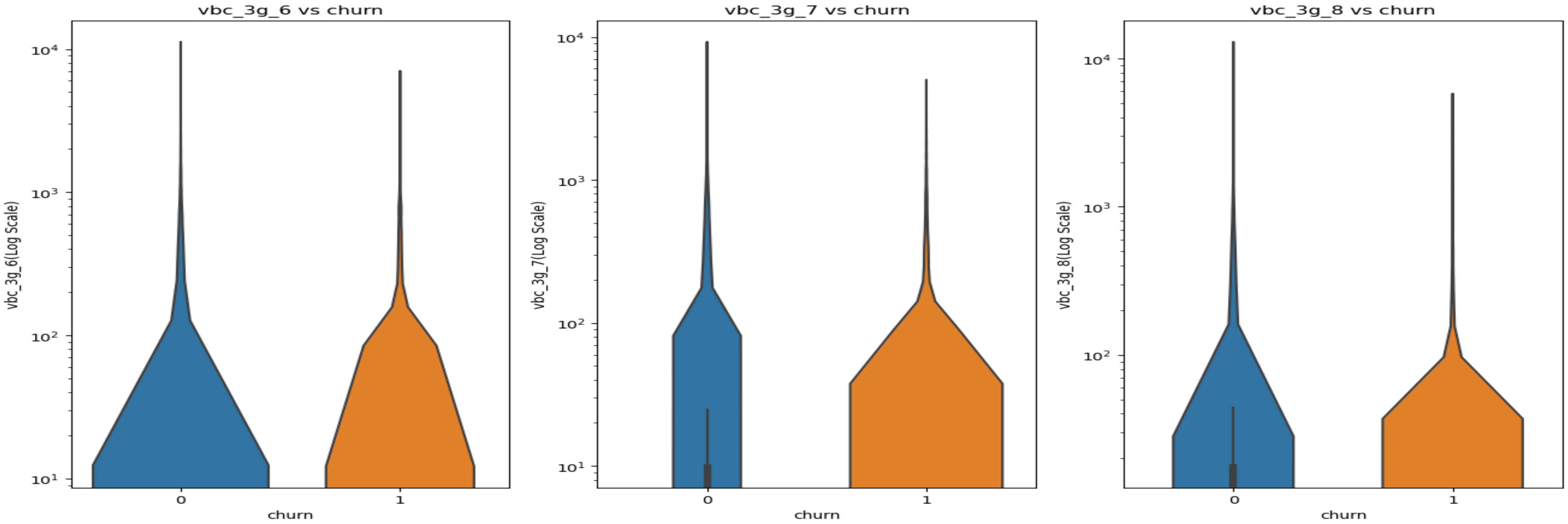
18.After all the above processing, we have retained 30,011 rows and 126 columns.

Prerequisites Steps

- Train-Test Split has been performed.
- The data has high class-imbalance with the 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.

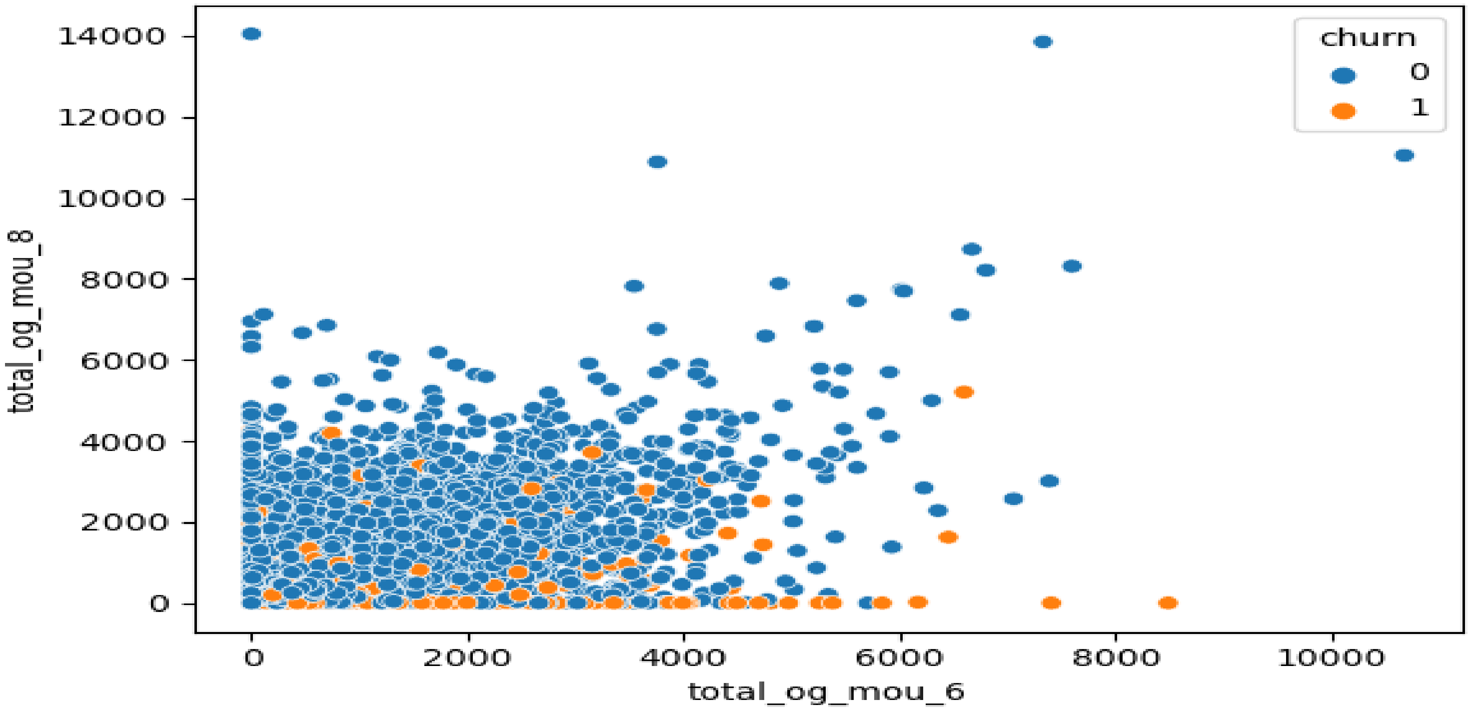
Churn rate based on whether the customer decreased her/his MOU in the action month

- Customers whose minutes of usage (MoU) decreased in the action phase have a higher churn rate than those whose minutes of usage increased in the good phase



Analyzing recharge amount and number of recharge in action month

- We can see from the pattern that the recharge number and the recharge amount are almost proportional. The higher the number of recharges, the Higher the amount of the recharge.



COMPLETE MODEL STATS

Sr. no.	Model	Accuracy	Sensitivity	Precision
1	Train Performance	0.89	0.83	0.43
2	Test Performance	0.87	0.43	0.39

Conclusion and Strategy

1. From EDA, we observed that there is a considerable drop in recharge, call usage and data usage in the 8th month which is the `Action Phase`.

2. Average revenue per user in the `7th month` plays a vital role in deciding churn. A sudden drop in it might indicate that the customer might be thinking about churning and appropriate actions should be taken.

3. Local Minutes of usage (outgoing) are the most affecting features on the customer churn.

4. Roaming Minutes of usage (incoming & outgoing) are also affecting features on the customer churn.

5. Total minutes of usage for outgoing is also an important factor affecting the churn

Following strategies can be incorporated

- A sudden drop in Local Minutes of usage might be because of unsatisfactory customer service because of poor network or unsuitable customer schemes/plans. Efforts shall be made to provide a better network and focus on customer satisfaction.
- Based on the usage / last recharge/ net usage, routine feedback calls should be made for customer satisfaction and services that can understand their grievances & expectations. Appropriate action should be taken to avoid them from churning.
- Various attractive offers can be introduced to customers showing a sudden drop in the total amount spent on calls & data recharge in the action phase to lure them.
- Customized plans should be provided to such customers to stop them from churning.
- Promotional offers can also be very helpful.

Thank you