Telecom Churn Case Study

By

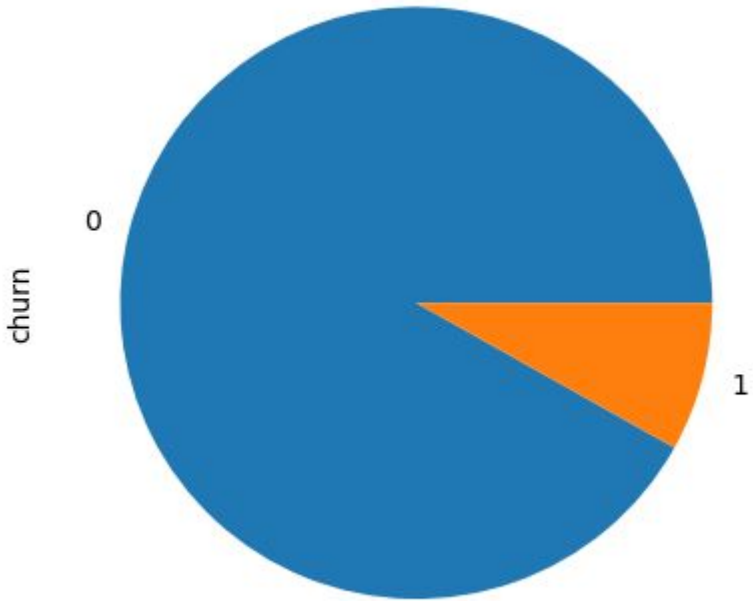
Ankit Kumar
Harmeet Singh
Jafar Imam

Business Problem

Overview

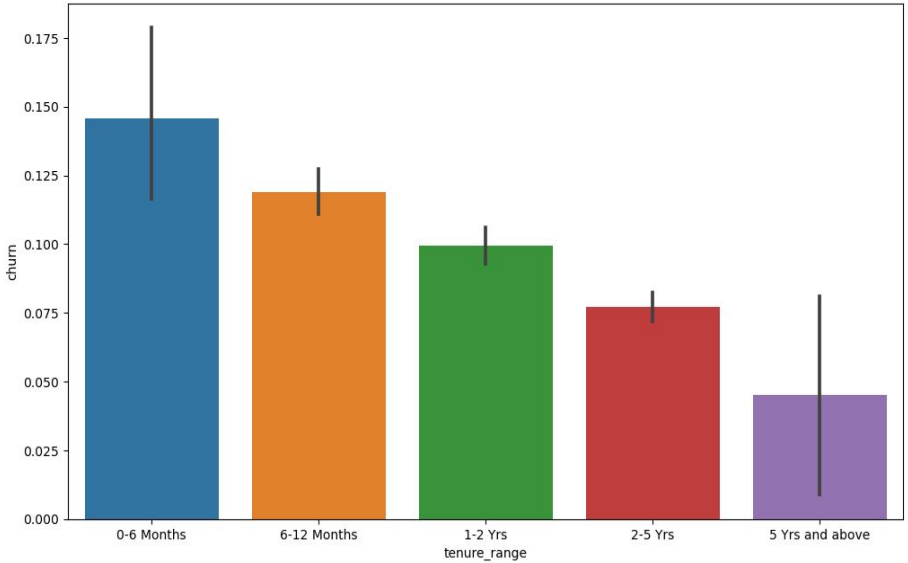
- In the telecom industry, customers are able to choose from multiple service providers and actively switch from one operator to another. In this highly competitive market, the telecommunications industry experiences an average of 15-25% annual churn rate.
- Given the fact that it costs 5-10 times more 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.
- In this project, we will 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 Percentage



As we can see that 91% of the customers do not churn, there is a possibility of class imbalance.

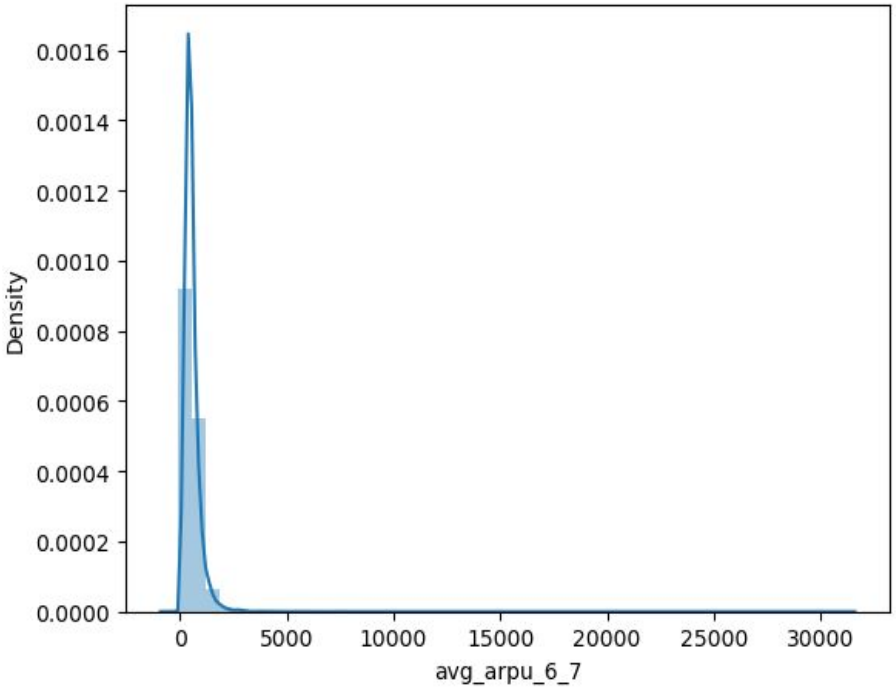
Tenure Range vs Churn



Derived New Variable Tenure Range using binning technique

It can be seen that the maximum churn rate happens within 0-6 month, but it gradually decreases as the customer retains in the network.

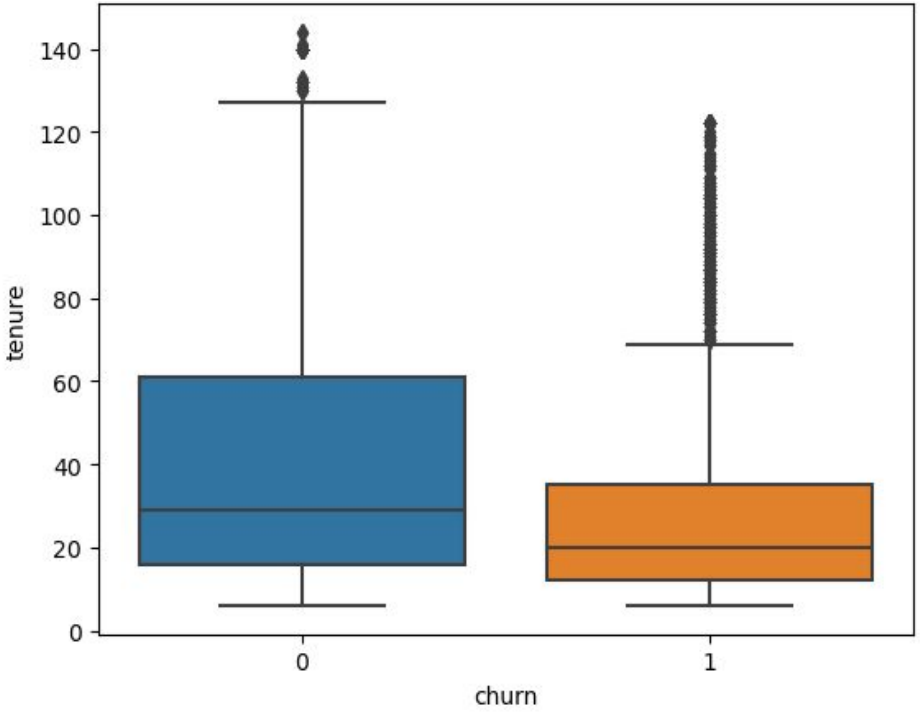
Avg. of arpu_6 & arpu_7



The average revenue per user for the good phase of customers is given by arpu_6 and arpu_7. Since we have two seperate averages, let's take an average to these two and drop the other columns.

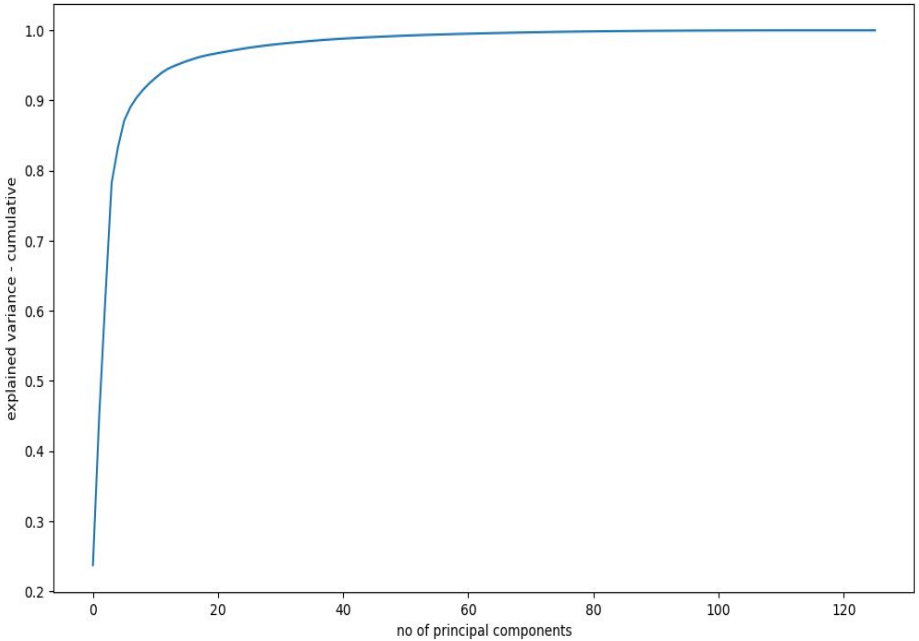
After, aggregation the graph looks positively skewed.

Tenure vs Churn



It is evident by the plot that tenured customers do not churn and continue the telecom services.

Logistic Regression Using PCA



Model Summary

Accuracy of the logistic regression model with PCA for training set: 0.82 approx.

Accuracy of the logistic regression model with PCA for test set: 0.75 approx.

Approximately 90% of the data can be explained using 8 PCA components

Logistic Regression Equation

Below are few top variables selected in the logistic regression model.

- Variables Coefficients - loc_ic_mou_8 -3.3287 og_others_7 -2.4711 ic_others_8 -1.5131 isd_og_mou_8 -1.3811 decrease_vbc_action -1.3293 monthly_3g_8 -1.0943 std_ic_t2f_mou_8 -0.9503 monthly_2g_8 -0.9279 loc_ic_t2f_mou_8 -0.7102 roam_og_mou_8 0.7135 We can see most of the top variables have negative coefficients. That means, the variables are inversely correlated with the churn probability.

E.g.:-

- If the local incoming minutes of usage (loc_ic_mou_8) is lesser in the month of August than any other month, then there is a higher chance that the customer is likely to churn.

Business Recommendations

- 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.
- 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.