Telecom data Churn

UPGRAD PROJECT

By- Priyanka

Introduction and 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 more important than customer acquisition.

For many incumbent operators, *retaining highly profitable customers is the number one business goal*.

To reduce customer churn, telecom companies need to **predict which highly profitable customers are at risk of churn.**

In this project, you 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.

Understanding the business objective and the data

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.

Data preparation

The following data preparation steps are crucial for this problem:

1. **Filter high-value customers**

As mentioned above, you need to predict churn only for high-value customers. Define high-value customers as follows: Those who have recharged with an amount more than or equal to X, where X is the **70th percentile** of the average recharge amount in the first two months (the good phase).

After filtering the high-value customers, you should get about 30k rows.

2. **Tag churners and remove attributes of the churn phase**

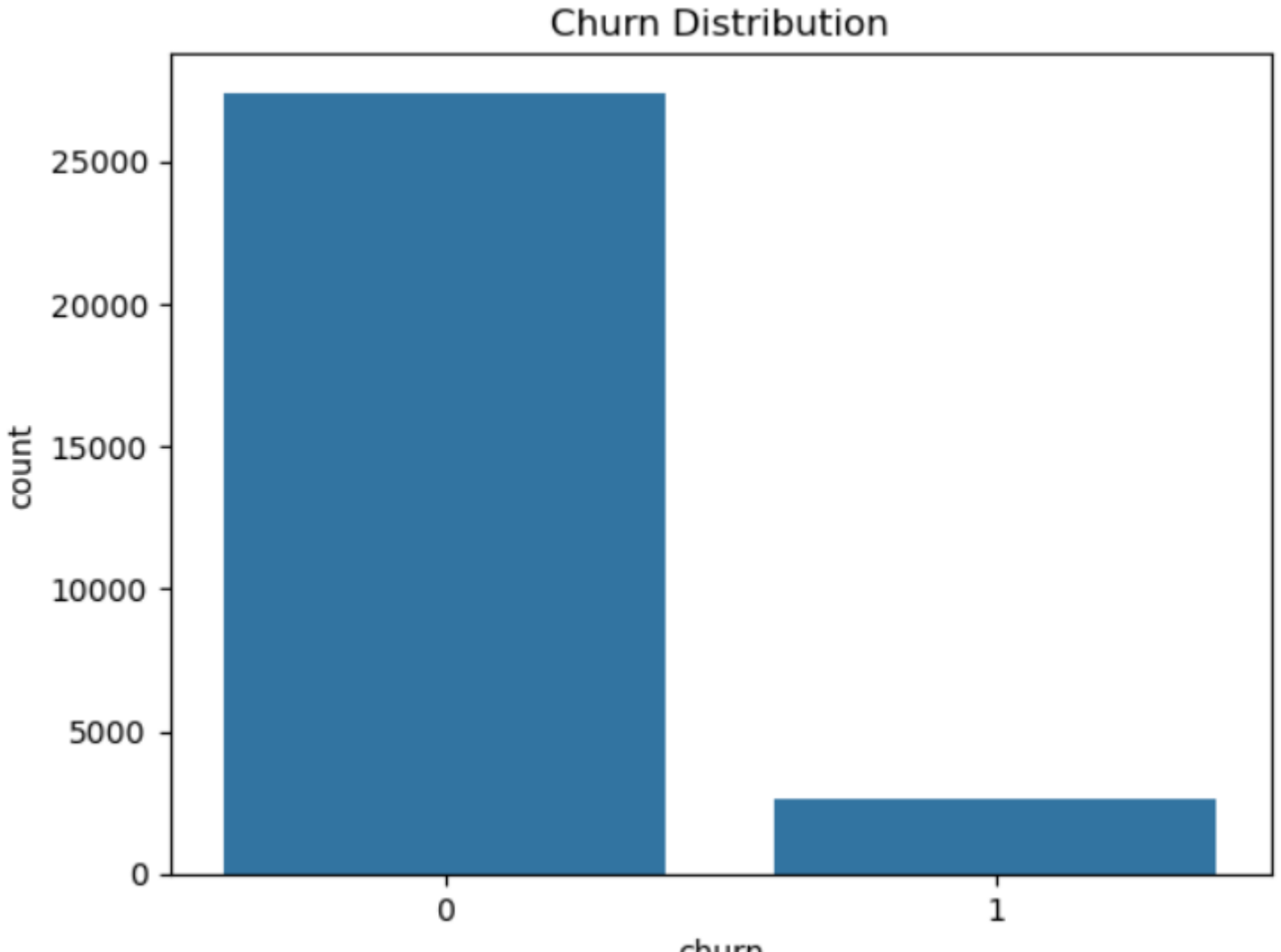
Now tag the churned customers (churn=1, else 0) based on the fourth month as follows: Those who have not made any calls (either incoming or outgoing) AND have not used mobile internet even once in the churn phase. The attributes you need to use to tag churners are:

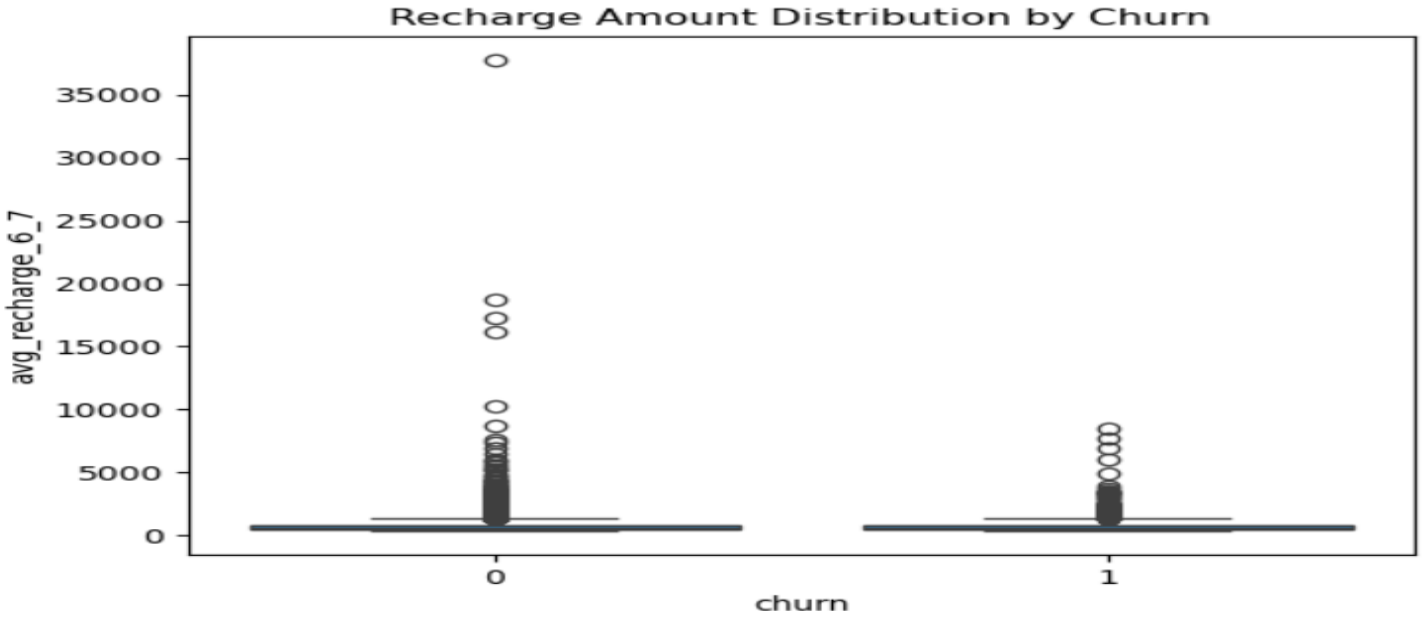
total_ic_mou_9
total_og_mou_9
vol_2g_mb_9
vol_3g_mb_9

After tagging churners, **remove all the attributes corresponding to the churn phase** (all attributes having ' _9', etc. in their names).

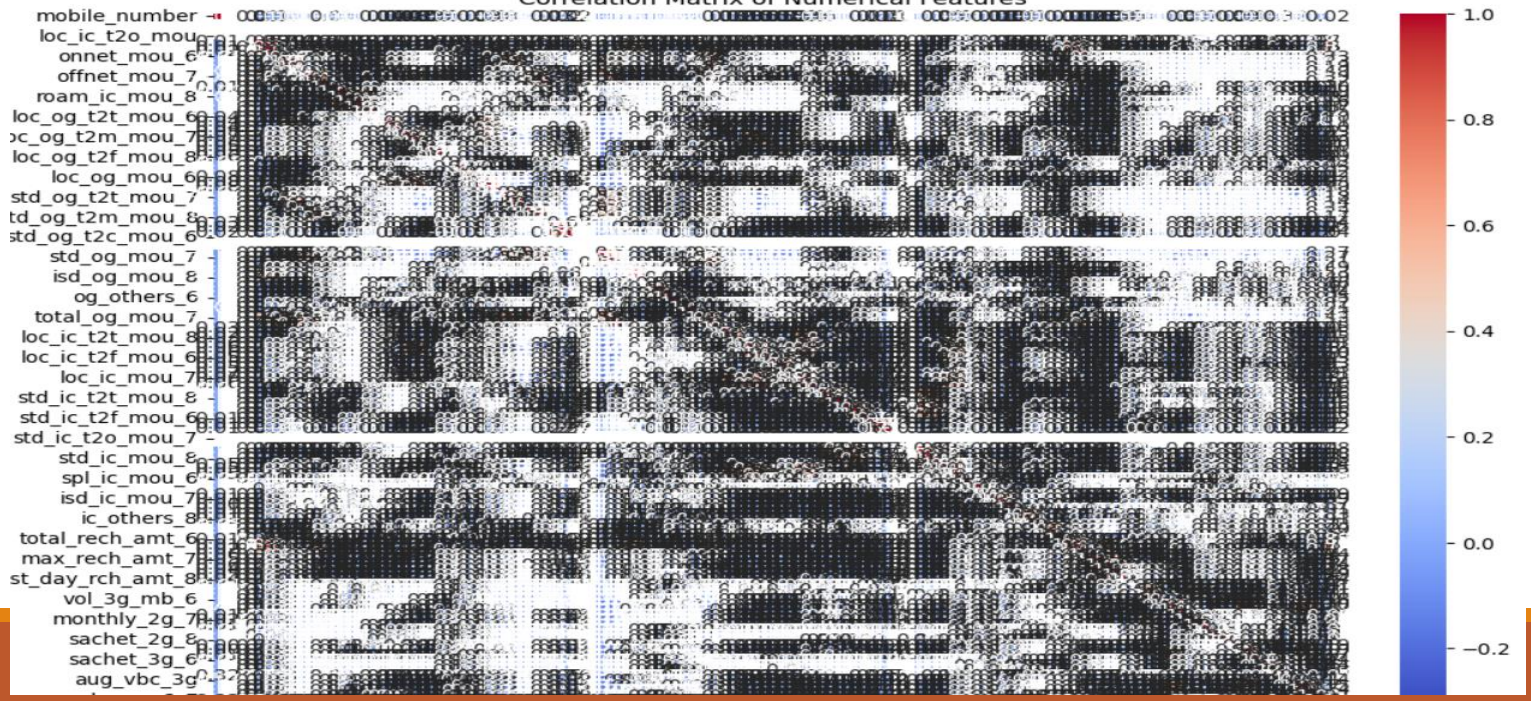
Modelling

Build models to predict churn. The predictive model that you're going to build will serve two purposes:
It will be used to predict whether a high-value customer will churn or not, in near future (i.e. churn phase). By knowing this, the company can take action steps such as providing special plans, discounts on recharge etc.





Corelation matrix of numerical features



Thank you