Churn Case Study

To reduce customer churn, telecom companies need to predict which customers are at high risk of churn

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.

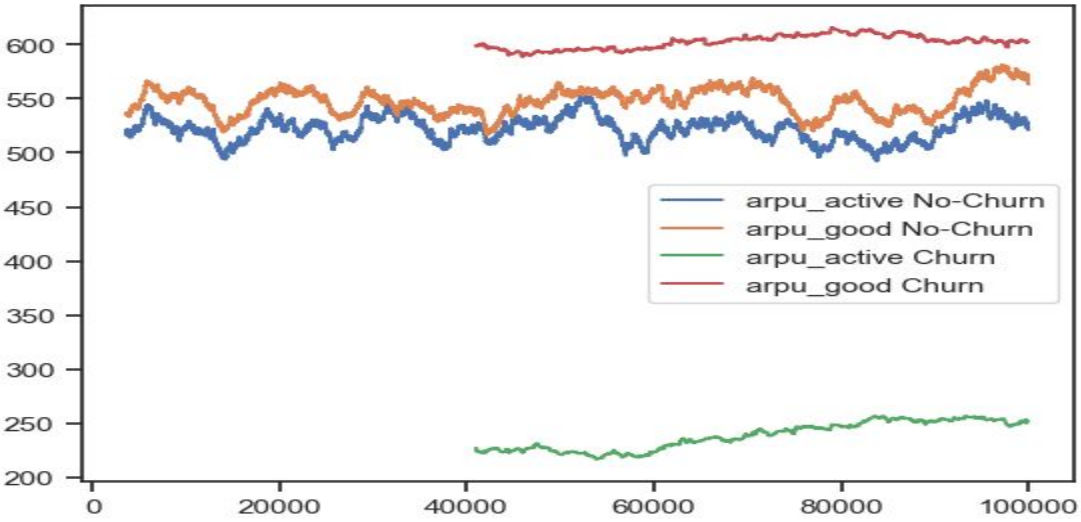
In this project, you will analyse customer-level data of a leading telecom firm, build predictive models to identify customers at high risk of churn and identify the main indicators of churn.

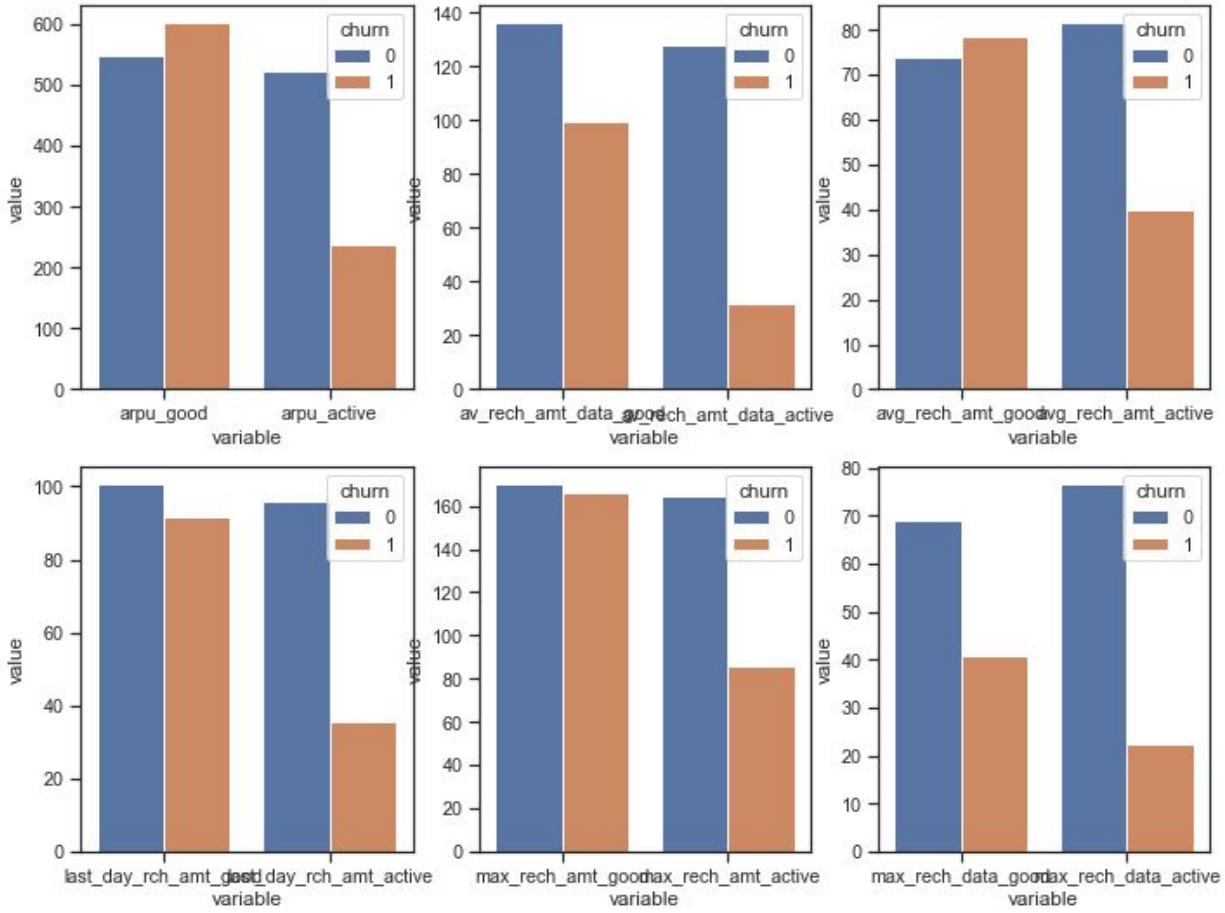
Data Preparation

- Derive new features
- Filter high-value customers
- Tag churners and remove attributes of the churn phase

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

Exploratory Data Analysis




churn
0
1
value
variable
arpu_good
arpu_active
av_rech_amt_data_good
av_rech_amt_data_active
avg_rech_amt_good
avg_rech_amt_active
last_day_rch_amt_good
last_day_rch_amt_active
max_rech_amt_good
max_rech_amt_active
max_rech_data_good
max_rech_data_active

Modelling

1. Preprocess data (convert columns to appropriate formats, handle missing values, etc.)
2. Conduct appropriate exploratory analysis to extract useful insights (whether directly useful for business or for eventual modelling/feature engineering).
3. Derive new features.
4. Reduce the number of variables using PCA.
5. Train a variety of models, tune model hyperparameters, etc. (handle class imbalance using appropriate techniques).
6. Evaluate the models using appropriate evaluation metrics. Note that is is more important to identify churners than the non-churners accurately - choose an appropriate evaluation metric which reflects this business goal.
7. Finally, choose a model based on some evaluation metric.

Conclusion

- The sudden surge in the special incoming and out going calls in the Active months
- Considerable reduction in the std incoming and out going calls
- Sudden reduction in the usage of 3g data services
- Sudden drop in the local incoming calls. (The contact preferences of the customer changing)
- Sudden reduction in the usage of services with validity less than a month - for example 3g Sachets It is also evident from Exploratory Data Analysis, that where the data is not collected (NaNs), the percentage of Churn is higher.The KPI data collection methods will need to be reviewed for a better prediction.