Business Problem:

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.**

Objective:

- 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:

About 1 lakh users, 4 month's data are given. Data contains every month's minute of usage for incoming, outgoing to same operator or another operator includes local, std calls, mou for roaming, ISD, special services etc. in short grand total of incoming and outgoing minutes of usage for every month. Average revenue per user for every month.

Amount of recharge for talk time and data, data consumption 2g and 3g, special services like social media, special night hours etc.

Age on network shows users age on network.

target variable is churn variable, which we need to figure out using last month data, if users incoming and outgoing mou and also if data consumption of user for last month is zero, then we can consider user is churned.

Data Pre-processing:

Missing value:

There are some missing values where zero imputation makes sense and rest of the missing values filled by mean values as data was skewed. For e.g. If user did make a data recharge its data consumption will be zero, data arpu will be zero, number of and amount of data recharge will be zero.

Filter out high value customers:

put a filter based on good phase month's (6^{th} and 7^{th}) a user's whose average of total of talk time and data recharge amount resides above 70^{th} percentile, considered as high value customer.

Derive Churn (target variable):

target variable is churn variable, which we need to figure out using last month data, if users incoming and outgoing mou and also if data consumption of user for last month is zero, then we can consider user is churned.

Derived Variable (Feature engineering):

so, we need to take corrective measures before churn phase and within action phase. for that we need to check the behaviour of customers while transacting from good phase to action phase. if we able to find out customer with changed behaviour we can take corrective measures before it's too late. We derive change in behaviour variable.

Outlier Treatment:

There were some customers who spend very high amount, consume very high data, very high arpu. But this was expected. They provide useful information so cat just remove them. K-sigma imputation used to keep all the users within 3 std deviation distance from the mean.

EDA:

- we can clearly see that, change in user arpu behaviour reflects on the churn. not churned users arpu is near zero whereas churn user.
- total outgoing and incoming minute of usage also separate churned user from non-churned user.
- change in total recharge amount for non-churned user is around zero, where as it is negative for churned customer.
- total recharge data amount there is very less behavioural change in user of churn and non-churn.
- maximum recharge of talk time and data also helps to separate out churn from non-churn.
- onnett and offnet change separate's churn user very narrowly.
- last day recharge amount also separates churn and non-churn users.
- from above we can see that, very few users avail night pack services.
- large number of users uses social media platform such as Facebook.

Data Balancing:

Target variable was highly imbalanced, 92% in favour of non-churned. Our requirement was to create high recall model. So, used Imblearn random under sampling technique to make data balance.

Interpretable model:

- Used Logistic Regression to find out important features and for interpretability.
- from above results we can see that, when user's Social media (Facebook, etc.) usage reduces, odds of user's churning increases.
- when user make less number of times data recharges compare to previously, its odds increases to churn.
- when users minute of usage for incoming and outgoing calls changes negatively in action phase, odds of user getting churn out increases.
- when user is making less amount of recharges compare to previous months, users' odds of churning increases.
- age on network defines, customers loyalty towards telecom operator, old users' odds of churning out are less whereas new users are more prone to churn out.
- reduction in 2g and 3g data consumption (vbc, arpu) also makes users odds high of churning.

Prediction model:

	Model_Name	Recall_Train	Recall_Test	Accuracy_Train	Accuracy_Test	Precision_Train	Precision_Test
0	RandomForest	0.980031	0.979508	0.605223	0.280287	0.560140	0.099854
1	Logistic Regression	0.916027	0.930328	0.705325	0.518080	0.644452	0.137077
2	XGBoost	1.000000	0.961066	0.721198	0.406432	0.642012	0.116899
3	SVM	0.851510	0.842213	0.794163	0.728712	0.763895	0.209480

- from above predictive model we found Random Forest having a high recall of approximately 98%.
- XG boost also having recall score with 96% and having a better test accuracy. Also having AUC of 0.95.
- Logistic regression and SVM having recall respectively 93% and 84%. SVM having high Test accuracy of 72%. all the model has poor precision.
- we can select Random Forest model as it has high Recall 98% on test side.

Recommendations:

1. large number of users are social media like Facebook consumers, company can come up with attractive offers to make them stay.
2. users who are with company for longer time period, company can offer them special offers. because older users less likely to churn.
3. for new customers company can come up with discounted long term offers, since old users less likely to churn.
4. company can offer customized recharge schemes for different users according to their usage. for example, a user having very high outgoing mou, company can offer discounted talk time recharge. same goes for data usages.
5. company can enhance user experience by providing quality issue resolution.