



# Telecom Churn Case Study Doubts Session

**Course :** Data Science  
**Lecture On :** Assignment  
**Instructor :** Sumit Shukla

## What we will cover in this session?

- 1 Case Study Walkthrough
- 2 QnA

# Assignment

No ready made target column

upGrad

6	7	8	9
x   y   z	x   y   z	x   y   z	x   y   z

## Problem Statement

6 } Good phase  
7 }  
8 → Action phase  
9 → Churn phase

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.

**Predict which customers are at high risk of churn.**

## Data cleaning

### What you need to do?

#### 1. Handling Missing data.

- Impute with zero when you are very sure that a missing is a zero.
- For categorical, what to do?
- Remove those with high missing percentage.

<https://www.kaggle.com/athi94/investigating-imputation-methods>

date of recharge	amt of recharge	# of recharge
→ 01/01/2011	26	NA
→ NA	NA	NA

← actual missing (pointing to NA in # of recharge)  
 ← no recharge was made (pointing to NA in # of recharge)  
 = 0 (pointing to NA in # of recharge)

- identify & impute rows and columns with zero where a missing represents a zero
- Drop all columns with high %age of missing (50%.)
- Impute columns with less %age of missing  
mean / median / mode.

# Assignment

(using month 6 & 7 data)

## Filter high-value customers(HVC)

Pareto rule (80-20)  
80% of revenue → top 20% customers

### Good Phase

top 30%.

1. Calculate average recharge done by customer in June and July (total\_rech\_amt)
2. Look at the 70th percentile recharge amount
3. Retain only those customers who have recharged their mobiles with more than or equal to 70th percentile amount

total rech-amt-6	total- rech-amt-7	avg-rech amt-6-7
a	b	$(a+b)/2$

} 70th percentile value

around 27.1k rows

- i) Create avg-rech-amt-6-7
- ii) find the 70th percentile value of avg-rech-amt-6-7 column
- iii) filter those customers for which  $\text{avg-rech-amt-6-7} \geq$  70th percentile value.

## Derive Churn

use 9<sup>th</sup> month data

### 9th Month is our Churn Phase. Usage-based churn

1. Calculate total incoming and outgoing minutes of usage
2. Calculate 2g and 3g data consumption
3. Create churn variable: those who have not used either calls or internet in the month of September are customers who have churned
4. Check Churn percentage.
5. Delete columns that belong to the churn month

$y = 1$   
 else  
 $y = 0$

Zf for month 9<sup>th</sup>  
 total-incoming-min == 0  
 and  
 total-outgoing-min == 0  
 and  
 total-2g-data == 0  
 and  
 total-3g-data == 0

Before moving to the next step  
drop all 9<sup>th</sup> month data

## Data preparation

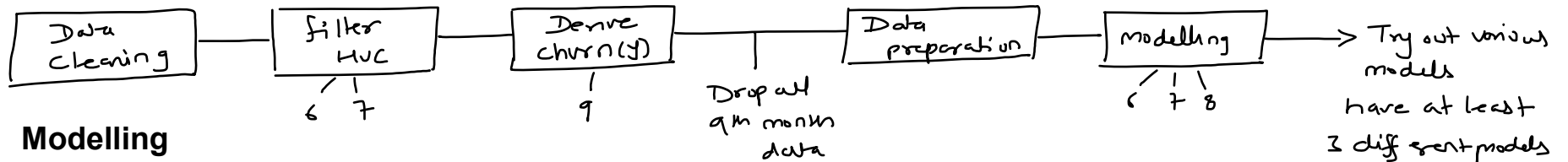
### Some more steps:

- ✓ 1. Derived variable. (at least 3)
- ✓ 2. EDA
- ✓ 3. Outlier treatment
4. \* Split train-test
5. scaling



# Assignment

upGrad



## Modelling

1. Try out various models to choose the one with great performance.
2. You need to handle the imbalance class for all the models.

<https://towardsdatascience.com/methods-for-dealing-with-imbalanced-data-5b761be45a18>

- Logistic Reg.
- Random Forest
- Decision Tree

↓  
for each of them  
handle class imbalance

↓  
~~do~~ Hyperparameter tuning

→ various evaluation metrics

using logistic model  
→ add recommendations to  
the company  
within your  
Jupyter notebook





Thank You!