

**CROSS SECTIONAL vs TIME SERIE vs Panel / Longitudinal Data**

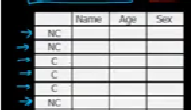
**Cross sectiona**l is a data that doesn’t depends with time and date, order of data is also not important here. We will have multiple variable here.

EX: Assume we have dataset that has information about who have

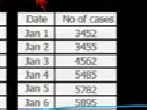
Covid, who are not having covid, with the date and time they affected, to predict

If a new person have covid then the date and time info is not required from the dataset.

So this data is not depends on time and date it is called cross sectional data



In **time series** data order is important, all the affected peoples date time should be order to know is the case is increasing or decreasing. We will have single variable that depends on date and time



**Panel / Longitudinal Data** is the combination of cross sectional and time series data. Here also we have multiple variable feature.

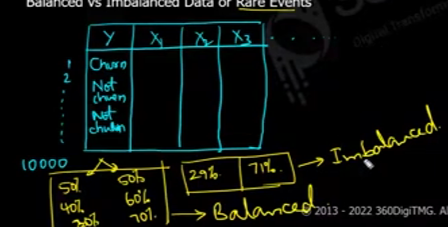


**BALANCED vs IMBALANCED DATA or Rare Events**

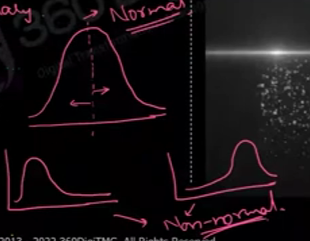
If the result of output variable is 50% 50% equal than it is called balanced

EX: In binary classification 50% images has some animal 50% images has not animals

One of the class in binary classification is less than 30% it is called imbalanced, in multi classification problem also if one of the class is having very less percentage compare to other class than it is called imbalanced



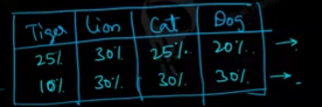
In below normal graph are balanced, non normal graph are imbalanced



**Imbalanced in multi classification**

**1st row -** balanced

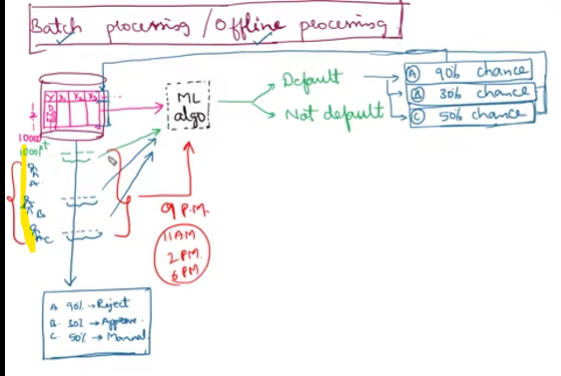
**2ndrow –** imbalanced



**BATCH DATA(OFFLINE) vs LIVE STREAMING DATA (ONLINE)**

Instead of predicting result for one input, we will predict the output for bunch of input or batch of input.

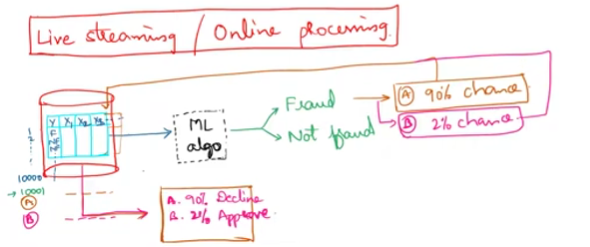
Highlighted persons are input in below, here we take batch of loan application from specific time and all the persons who applied loan on time 9 pm is the batch that will predicted, we store all the loan application each day, at the end of the day the prediction will be made here.



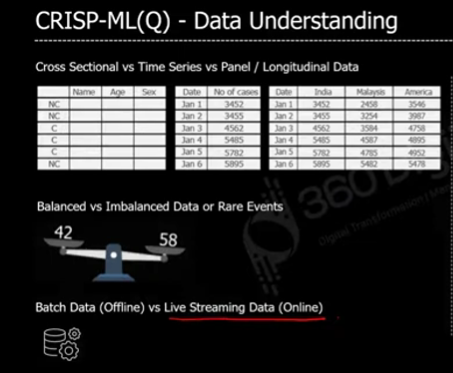
In **live streaming data** we have to predict a data that needs immediate result

EX: Bank credit card transaction

We have to predict is the transaction is fraud or not on the time of the user making the transaction.



**SUMMARY:**

****