Introduction to Framing ML Problem

1. Business Problem to ML Problem

   Netflix (To grow business)

   - Charge more from existing customer
   - Bring more customers
   - Reduce churn rate

2. Type of problem

   Reduce churn rate
   - identify the potential leaving customer
   - identify the reason of leaving Netflix
   - offer instant discount

3. Current solution

Probability of person leaving Netflix		
30 %	50 %	90 %
offer less discount	offer better discount	offer more discount

3.4. Getting Data

- Watch time
- Search but did not find
- content left in the middle
- Clicked on recommendations

5. Metrics to measure
6. Online vs Batch learning
7. Check assumptions
   will model work same as USA [In India]

__/__/__

* Data Gathering
  - CSV
  - JSON / SQL
  - Fetch API
  - Web scrapping

* Understanding your data

1. What How big the dataset is ? df.shape()
2. How does the data look alike ? df.sample(5)
3. What is the data type of the columns ? df.info()
4. Is there missings values ? df.isnull().sum()
5. How does data look mathematically ? describe
6. Are there any duplicate values in the data ? df.duplicated().sum()
7. correlations among the data ? df.corr()

* EDA for Univariate Analysis