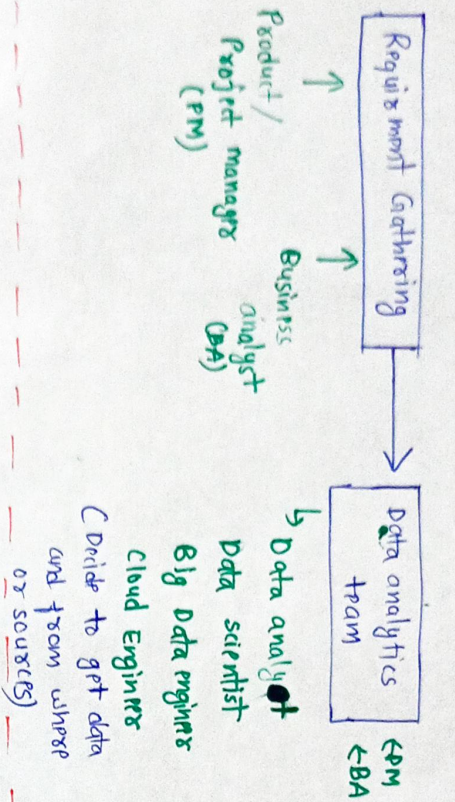


# Life Cycle of Data Science Project



**Statistics:** is the science of collecting, organizing and analyzing the data.

**Data:** facts or pieces of information

## Statistics

### Descriptive

[EDA+FE]

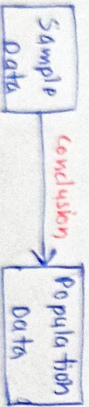
① It consists of organizing and summarizing data, using all different kind of plots.

### Inferential

It consists of collecting sample data and making conclusion about population data using some experiments.

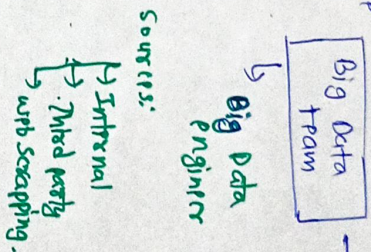
Hypothesis Testing

→ Entire data is population data  
→ Sample from population is sample data.



→ Relationships b/w ps / Comparisons

Data Source



## Life cycle of DS project

- ① EDA (Exploratory data analysis)
- ② FE (Feature extraction)
- ③ FS (Feature selection)
- ④ Model Training
- ⑤ Hyperparameters tuning
- ⑥ Deployment

## Sampling Techniques

Select technique which will give good results/data

- Simple random sampling:- Every member of population (N) has equal chance of being selected for your sample(n)
- stratified sampling:- strata → layers → clusters → subgroup.
- systematic sampling:- select every nth (1st, 2nd...) individual out of population (N).
- Convenience sampling:- only those who are interested in survey will only participate.

**Variables:** a property that can take values.

- Quantitative: measured numerically (mathematical operations)
- Qualitative/Categorical: categories based on characteristics
- Descriptive: whole numbers, many repetitions.
- Continuous: Fractions/decimals.