Page No.:
Date:

Week 2

for a DAst, you have to avoid conflict, overcome challenges, answer the questions

So, let's check data...
How?

- Analyze "Bias" & "Credibility".
- Is data "Good" or "Bad"?
- Data ethics, privacy & access.

---

(1) "Bias" & "Credibility"

Bias:

A preference in favor or against a person, group or a thing.
(conscious or sub-conscious)

Page No.:
Date:

Data Bias:

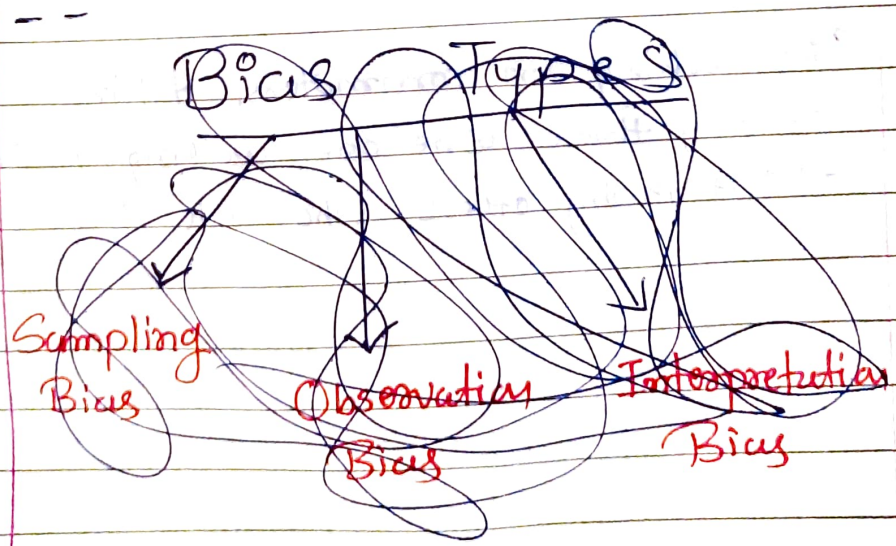
A type of error that systematically skews result in a certain direction.

- A survey has some influence to certain answers,
  or
- sample group is not truly representative of population

DAst starts looks for Bias & Fairness from the moment they start collecting data.

Sometimes we also need biased data to narrow focus

--



Page No.:
Date:

Bias Types

- Sampling Bias
- Observer Bias (જોવામાં)
- Interpretation Bias [redacted] (અર્થ નહીં, પણ અર્થ)
- Confirmation Bias (ખાતરીપૂર્વક કર્યું)

1] Sampling Bias:

It is when a sample isn't representative of the population as a whole

(Researching on commuters, taking surve
of only those who are walking in sidew
- But leaving others who ride bicycles, dr

Page No.:
Date:

2| Observer Bias
(experimenter bias / research bias)

A tendency for different people to observe things differently.

(Blood pressure 125.7 → 126 (Person A)
125.7 → 124 (Person B))

3| Interpretation Bias

A tendency to always interpret ambiguous situations in a positive or negative way.

(You think your dad has scold you and you get angry on him, but the same recorded if you listen after a day - you find nothing like scolding!
- It is when, which time with which mindset (bias) you are listening or interpreting things)

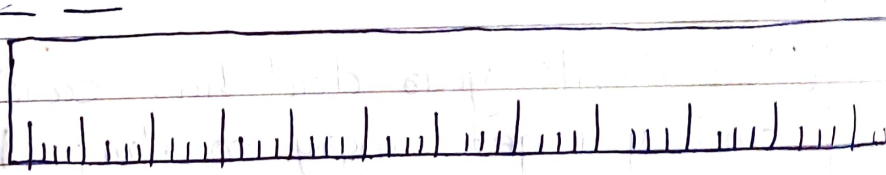
4] Confirmation Bias:

People see, what they want to see.

A tendency to search for, or interpret information in a way that confirms preexisting beliefs.

Real vs.

(Eager to confirm gut feeling – only notice those which support feeling — ignoring & other signals)



Measures of Good Data.

R - reliable —— Accurate, Complete, Unbiased
T R U S T

O - original —— Mostly SP, TP ... Make sure

C - comprehensive —— All critical information to answer.

C - current —— Self-explanatory

C - cited. —— Credible.

Page No.:
Date:

Simply ask these to start...

- Who created the data?
- From credible organization?
- When was last updated?

(Data.gov - good site for data...)

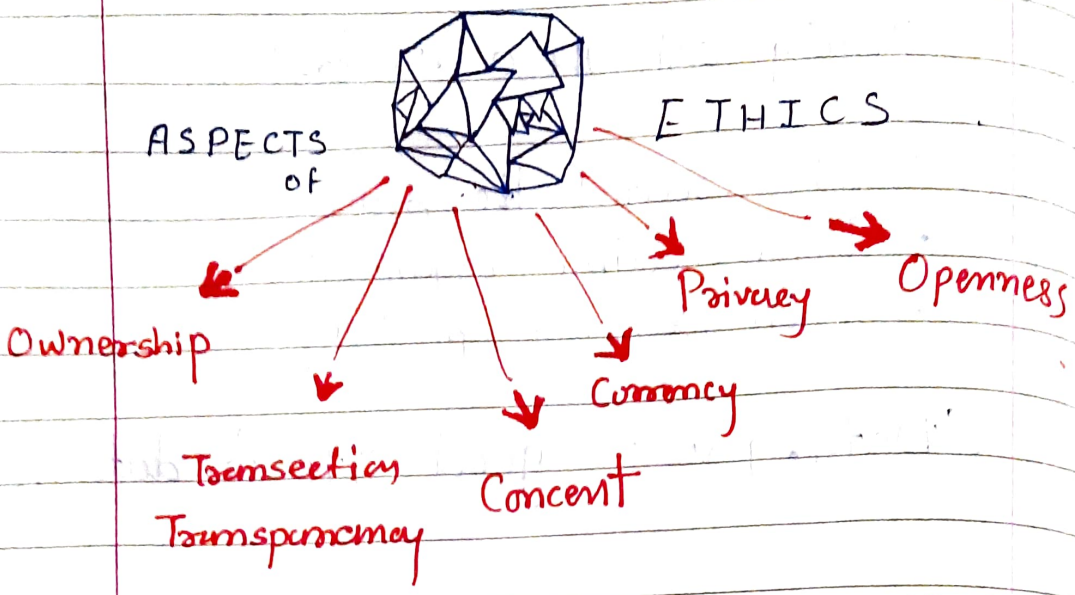
"Every good solution is found by avoiding bad Data"

Ethics: (human)

Well founded standards of right or wrong that prescribe what humans ought to do, usually in terms of rights, obligations, benefits to society, fairness. [redacted]

Ethics: (data)

Well founded standards of right or wrong that dictate how data is collected, shared and used.



Ownership: Who owns data?

Individuals own the raw data they provide and they have primary control over its usage - How processed
- How shared

Transaction
Transparency: All data being processed and activities & algorithms should be completely explainable and understood by the individual who provides the data

Page No.:
Date:

Concent: This is the individual's right to know explicit details about how and why their data will be used before agreeing to provide it.

- How will it be used?
- How long will be stored?
- Why is collected?

(Terms & Conditions)

Currency: Individuals should be aware of financial transactions resulting from the use of their personal data and the scale of these transections.

Privacy: Preserving a data subject's information and activity any time a data transection occurs.

All about access, use and collection of data.

"Person's legal right to the data".

Openness: Free access, usage and sharing

Here we still be transparent respect privacy and make sure we have concent for data that's owned by others.

- re-use and redistribution

Data Anonymysation:

PII - Personally Identifiable Information

type of data must be anonmyse

Done by - Blanking, Hashing, Masking