CORRELATION AND CAUSATION

Correlation → It is a term that is a measure of strength of a linear relationship between two quantative variable (eg height, weight).

Causation → The action of causing something.

Correlation implies Causation / Correlation does not imply causation.

Implies Causation → When action A causes outcome B. Causation explicitly applies to cases where action A causes outcome B.

Does not implies Causation → When action A does not causes outcome B.

Example → Correlation vs Causation in Product Analytics.

→ **Problem statement** : We just launched a new version of mobile app. We make the key bet that user retention for our product is link to in-app social behaviours. You ask your team to develop a new feature that allow users to join "communities".

→ **Time frame** → A month after we release & announce new community features, adoption sits about 20% of all users.

→ **Step 1** → We make **2 cohort** . equally size randomly selected users.
Cohort 1 → Only users who joined the community . **Cohort 2** → Who did not join community.

→ **Analysis report** → Users who joined atleast one community are being retained far greater than the average user.

So can we say that, we have enough information to conclude whether joining communities causes better retention. All we know is that two are correlated.

→ **Causal relationship don't happen by accident** (It seems to exist, but actually isn't there).

→ **Run robust experiments to determine causation** → i) Hypothesis testing ii) A/B experiment

i) **Hypothesis testing** — In this H_0 is there is no relationship.
H_1 should identify the relationship we expect between dependent and independent variable.

H_0 - There is no relationship between joining an in-app community & user retention
H_1 - If a user join a community, then they will remain a customer for more than one year.
Start with onboarding flow. For next 1,000 users who sign up, split into two groups.
Half will be force to join community when thy sign up and other half won't be.
Run experiment for next 30 days and then compare retention rate between two groups.

ii) **A/B testing** — A/B/n testing is ideal when we are comparing impact of different versions.
- A split test of your product onboarding flow, for example might compare how different strategies performed based on characteristics including i) Copy variation ii) Using third party app to automatically recognise name & company of user iii) Different graphics
After running multiple product variation, use community join feedback page and then take a look at the results to compare metrics such as drop off rate, conversion and even retention.