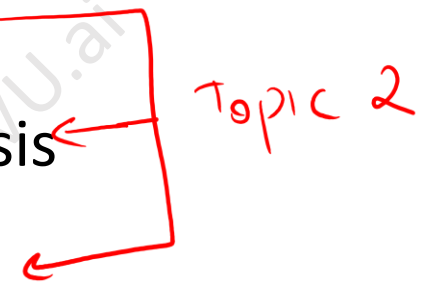
A/B Testing and Apriori

Author: Rohit Mande
Founder and CEO, inttrvu.ai

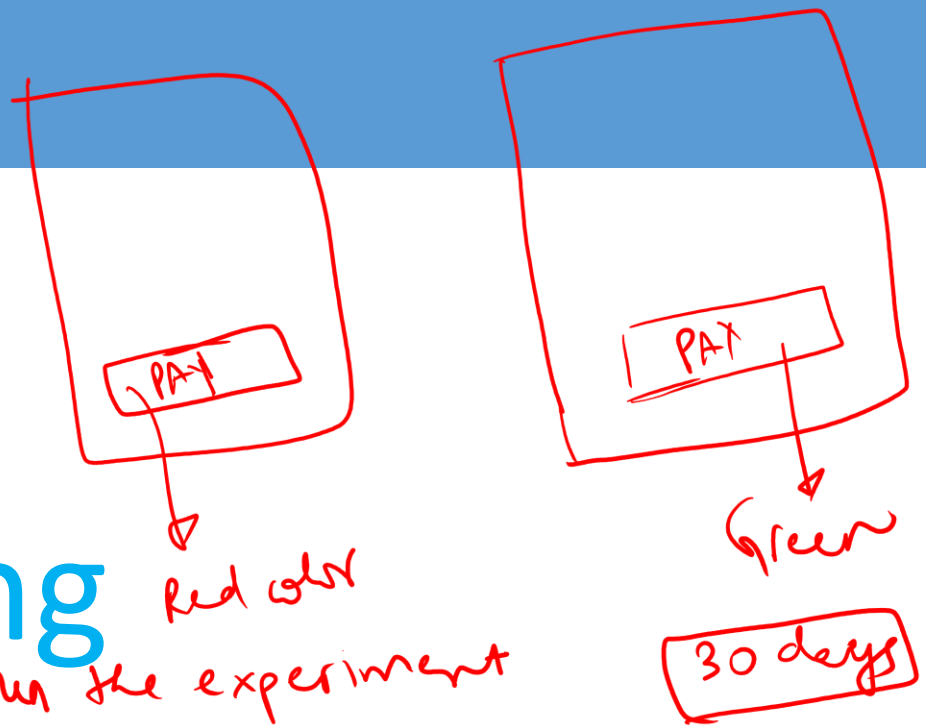
Index

- What is A/B testing?
- Hypothesis testing for A/B test
- What is Market Basket Analysis
- Applications of Market Basket Analysis
- Apriori algorithm



Item (A)
Item (B)

Two products / versions / items / design / entities, and we want to compare the performance



A/B testing

Assumption → Changing the button color will increase the revenue

Design, perform and collect the data

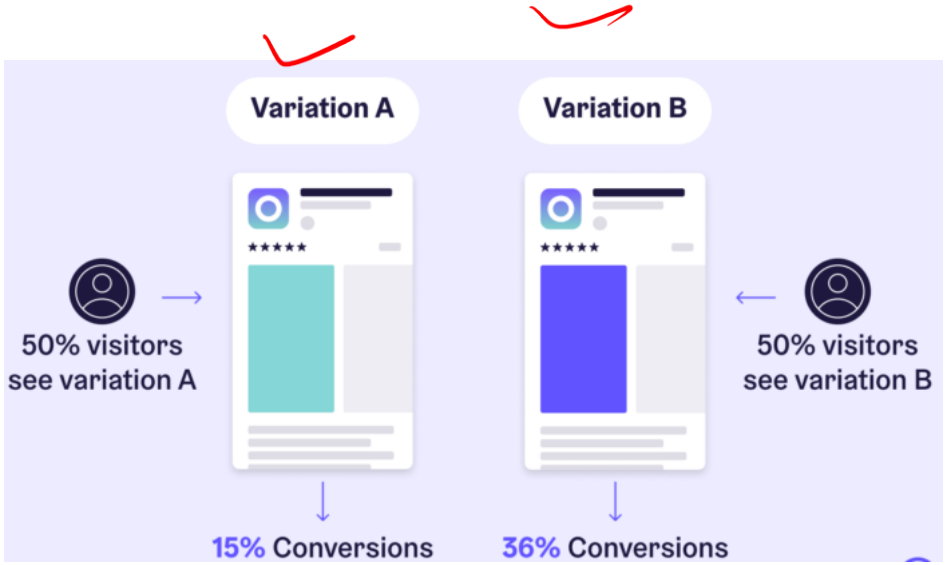
① Decide for the number of days you want to run the experiment

② Decide the number of participants 2000 daily

What is A/B testing?

- A/B testing (or split testing) is a user experience research methodology.
- It consists of a randomized experiment that usually involves two variants (A and B)

- A/B tests are useful for understanding user engagement and satisfaction for online features like a new feature or product
- Large social media sites like LinkedIn, Facebook, and Instagram use A/B testing



INTTRVU

Hypothesis Testing for A/B Testing

Hypothesis testing is used for A/B testing

Samples A: Users visiting older version of website

Samples B: Users visiting new version of website

Null Hypothesis (H_0) **:**

- The null hypothesis states that the difference in sample observations (A and B) result purely from chance
- There is no difference between the control (A) and variant (B) groups

Alternate Hypothesis (H_1) **:**

- It indicates that observed difference is not by chance – there is relationship /pattern
- The conversion rate of newer version of website (B) is higher than older version (A)

Hypothesis Testing for A/B Testing

Samples A:

- Daily 1000 random users sampled from users visiting older version of website
- Daily conversion rate calculated for 1000 users
- Experiment continued for 5 weeks – 35 data points

Samples B:

- Daily 1000 random users sampled from users visiting Newer version of website
- Daily conversion rate calculated for 1000 users
- Experiment continued for 5 weeks – 35 data points

Day	Conversion rate A	Conversion rate B
1	0.14	0.16
2	0.12	0.11
.		
35	0.18	0.19

Hypothesis Testing for A/B Testing

Should we simply calculate average conversion rate for A and B to conclude?

- No
- Higher average doesn't necessarily indicate B is better version than A

Solution:

- Two sample t-test
- It is needed to prove that higher average is not just by chance
- For rejecting null hypothesis, we should prove the statistical significance (p-value <0.05)

Day	Conversion rate A	Conversion rate B
1	0.14	0.16
2	0.12	0.11
….	….	….
35	0.18	0.19

For supermarket

For warehouse management

Idea is to understand the purchasing pattern/behavior to know the affinity b/w multiple items

Market Basket Analysis

What is market basket analysis?

① Item 1 reinforces the purchase of item 2
② Item 1 & 2 are independent
③ Item 1 is substitute of item 2

- Market basket analysis is used by retailers to understand customer purchasing patterns
- It involves analyzing large data sets, such as purchase history, to identify products that are likely to be purchased together
- It helps large retailers to uncover known/unknown associations between items

Basket Number	Items Purchased
1	Apples, Bread, Eggs, Yoghurt
2	Bananas, Sugar, Onion
3	Bread, Eggs, Ketchup
4	Beans, Bread, Eggs, Onion

Applications of market basket analysis

- Market Basket analysis can be used for recommending products

  E.g. Amazon recommends people who bought X also bought Y

- It can also help in product placement in retail store | warehouses

  E.g. Customer who buy new technology book also buy Marker – place them nearby

Apriori Algorithm

is used for market basket analysis or association rule mining

- It is used for association rule mining (finding items which are frequently bought together)
- Name of the algorithm is Apriori because it uses prior knowledge (e.g., purchase history) of frequent itemset properties.
- **Apriori Property:**
  - All subsets of a frequent itemset must be frequent
  - If an itemset is infrequent, all its supersets will be infrequent (If any item is bought infrequently then its associations with other items will also be infrequent)

Property 1

All item subsets of a frequent itemset must be frequent

A frequent itemset is one occurring atleast 20% times.

[A, B, C, D] ⇒ frequent itemset

(5) [A], [B], [C], [D], [A, B], [A, C], [A, D]

[B, D], [C, D], [A, B, C] [A, B, D]

[A, C, D] [B, C, D]

Property 3 →

If an itemset is infrequent, then all its supersets will also be infrequent.

[A, B, C, D, E] → infrequent ⇒ occuring only once.

[A, B, C, D, E, F] → assume this is frequent ✗ infrequent

[A, B, C, D, E, G]

[A, B, C, D, E, H]

Apriori Algorithm

Support, confidence, ~~lift~~ Lift

evaluation metric

Support

Support is an indication of how frequently the itemset appears in the dataset

Support(A) = 6/10 = 0.6 = 60%.

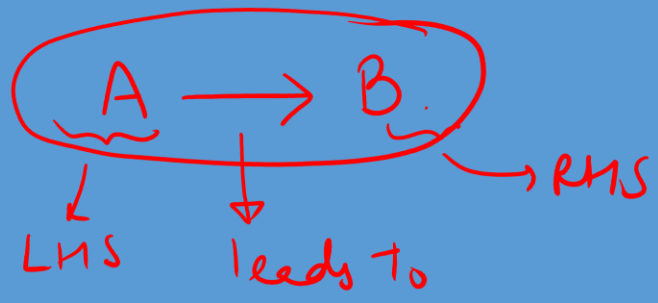
Support(A,B) = 5/10 = 0.5 = 50%.

A → B support(A,B)

A,B → C support(A,B,C)

Basket	Items (sorted list)
1	A, B, F, H, K
2	B, T, Z
3	A, M, N, P, R
4	A, B, C, D, M, P, Q
5	D, E, F
6	A, B, F, H, Z
7	A, B, M, N, O, P
8	C, M, P, Q, R
9	D, E, G, I
10	A, B, C, M, N, P, T

Apriori Algorithm



A, B → C

A → B, C

A, B → C, D

Confidence

Confidence is the percentage of all baskets with A that also have B

confidence (A → A1) = supp(A, A1) / supp(A) = 6/6 = 1

Confidence(A,B) = Support(A,B) / Support(A)

= 0.5/0.6

= 0.83

Lift (A → A1) = supp(A, A1) / supp(A) supp(A1) = 6/6x10 = 0.1

Every time customer buys A -> they buy B ~83% of the time

confidence (B → A) = ?

= supp(B, A) / supp(B) = 0.5/0.6 = 0.83

Basket	Items (sorted list)
1	A, B, F, H, K
2	B, T, Z
3	A, M, N, P, R
4	A, B, C, D, M, P, Q
5	D, E, F
6	A, B, F, H, Z
7	A, B, M, N, O, P
8	C, M, P, Q, R
9	D, E, G, I
10	A, B, C, M, N, P, T

Apriori Algorithm

$Lift(A \rightarrow B) = \frac{supp(A,B)}{supp(A)\ supp(B)}$

Lift > 1 → meaningful rule

Lift = 1 → No association b/w LHS & RHS

Lift

It is useful to extract actual associations and remove noise

E.g., If B is present in all baskets Confidence metric will be 1 for (A,B) but its simply due to high occurrence of B and not due to co-occurrence of A and B

Lift (A,B)

= Support(A,B) / (Support(A)*Support(B))

= 0.5 /(0.6*0.6)

= 1.38

Basket	Items (sorted list)
1	A, B, F, H, K
2	B, T, Z
3	A, M, N, P, R
4	A, B, C, D, M, P, Q
5	D, E, F
6	A, B, F, H, Z
7	A, B, M, N, O, P
8	C, M, P, Q, R
9	D, E, G, I
10	A, B, C, M, N, P, T

Lift < 1 → Items have negative impact

Apriori Algorithm

Lift < 1

- Items are substitute to each other.
- Presence of one item has negative effect on another item

Lift = 1

- Probability of occurrence of A and B (any two items under consideration) are independent of each other

Lift > 1

- Items are complementary to each other.
- Presence of one item has positive effect on another item