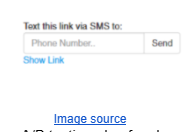
**Project Stages**

**Stage#1 – Choose Cases**

Case 1 : Website of the grocery store chain

You are an analyst for a large grocery chain. One of your company’s goals is to drive more customers to download our mobile app and register for the loyalty program. Your manager is curious if changing the link to a button of the app store will improve the user’s ability to download the app. Here is the existing link button of the app store.

Your manager asks you to create an A/B testing plan for changing the link to a button of the app store with the expectation it will the user’s interest to download the app.

**Stage#2 – Setting Up Problem**

***Website of the grocery store chain***

**Mendorong lebih banyak kostumer untuk mendownload aplikasi seluler dan mendaftar program loyalitas**

1. **Experiment goal**

Based on the case you selected, describe the purpose of your experiments

->*Melihat apakah perubahan pada link to a button of the app store dapat meningkatkan ketertarikan user dalam mendownload mobile app*

1. **Choosing Metrics**

You should define driver metrics and guardrail metrics in you’re a/B testing plan. Also write down how you formulated the metrics and the reasons why they were chosen.

*-> Goal Metrics : User Counts*

*- Merupakan representasi dari tujuan atau core bisnis perusahaan*

*Objective : Drive more customers to download our mobile app and register for the loyalty program*

*Reason :**User count bisa mengukur berapa banyak customer yang tertarik dan telah mengunduh mobile app tersebut di perangkat mereka*. ***Addition for reason***

*- Simple untuk dikomunikasikan dengan stakeholder*

*Stakeholder : Internal Team, Manager, Investor, Executives*

*Reason : Ketika pakai metrik ini, stakeholder bisa paham sejauh apa misi/tujuan sudah tercapai*

*-> Driver Metrics : CTR = Click Through Rate*

*- Metric ini mengukur berapa banyak customer yang mengunduh aplikasi dan mendaftar program loyalitas pada tautan*

*- Saat CTR naik, potensi mendownload mobile app semakin besar*

*- Untuk menhitung CTR maka : People who click to download mobile app from website/total number of website visitor\*100 = CTR*

*-> Guardrail Metrics : Mobile app loading time*

*- Mobile app load time can be characterized as the measure of time taken by the application to totally introduce before the interface opens and the application gets significant or interactive for the user.*

*- Mobile app loading time increase a few ms -> penurunan kepuasan -> meninggalkan/uninstall mobile app -> kehilangan user -> berpotensi rugi*

1. **Define Variants**

*Control : Existing Link*

*Treatment : New link, such as in picture not text.*

1. **Define hypothesis**

Write down the hypothesis of your experiment.

H0 (*Null Hypothesis*) : *CTR New link, such as in picture not text sama atau kurang dari existing link*

H1 (*Alternative Hypothesis*) : *CTR New link, such as in picture not text lebih dari existing link*

**Stage#3 – Designing Experiments**

In designing experiments, you are asked to determine:

1. Randomization unit -> *user*
2. Target of randomization unit -> *semua user yang mengunjungi halaman website of the grocery store chain yang berisikan link tautan*
3. Sample Size

For calculating sample size, you need to define :

1. *Significance level (peluang menerima H1 padahal H0 yang benar) ini adalah kesalahan dan harus dikecilkan. Maka dari itu, semakin kecil alpha semakin baik untuk eksperimen namun sampel yang dibutuhkan semakin besar -> 5% atau 0.05. Nilai tersebut merupakan tingkat signifikansi yang paling umum digunakan/standar industri yang mengindikasikan bahwa risiko kesalahan yang dapat diterima adalah 5% atau 0.05. Hal ini dipilih karena memberikan keseimbangan antara resiko type I error dan type II error.*

*Tingkat signifikansi 0,05 berarti bahwa kita menerima peluang 5% untuk membuat kesalahan Tipe I yang menolak hipotesis nol yang sebenarnya, dengan imbalan kepercayaan 95% bahwa hasilnya bukan karena kebetulan.*

*Type I error adalah kesalahan yang terjadi ketika menolak H0 (null hypothesis) padahal H0 benar, sedangkan type II error terjadi ketika gagal menolak H0 (null hypothesis) padahal H0 salah*

*Di sisi lain, tingkat signifikansi 0,01 (atau alfa = 0,01) lebih konservatif daripada tingkat signifikansi 0,05, konservatif maksudnya ketika hasil eksperimen yang diinginkan seakurat mungkin. Selain itu nilai tersebut digunakan dalam situasi di mana membuat kesalahan Tipe I memiliki konsekuensi yang signifikan atau di mana ukuran sampelnya kecil. Tingkat signifikansi 0,01 berarti bahwa eksperimen menerima peluang 1% untuk membuat kesalahan Tipe I, yang menghasilkan tingkat kepercayaan yang lebih tinggi (99%) pada hasilnya.*

1. *Power level (peluang menolak H0 padahal H0 yang salah) ini adalah kebenaran yang harus dibesarkan. Maka dari itu semakin besar power level semakin baik untuk eksperiman namun sampel yang dibutuhkan semakin besar -> 80% atau 0.8 dimana nilai tersebut adalah nilai standar yang umum digunakan. Hal ini disebabkan karena nilai power level yang tinggi menunjukkan bahwa eksperimen memiliki kemampuan yang kuat untuk mendeteksi perbedaan yang signifikan antara kelompok yang diuji jika perbedaan tersebut memang ada. Semakin tinggi nilai power level, semakin besar kemungkinan bahwa eksperimen akan dapat mendeteksi perbedaan yang sebenarnya antara kelompok. Jika kesalahan tipe I lebih penting, maka nilai power level yang lebih tinggi dapat dipilih untuk meminimalkan risiko kesalahan tersebut.*
2. Standard deviation of population ->
3. MDE = 2% (ingin bisa mendeteksi perbedaan 2%)

*Minimum detectable effect (MDE) adalah ukuran kecilnya perbedaan antara dua kelompok yang dapat dideteksi dengan tingkat kepercayaan tertentu dalam sebuah eksperimen A/B testing. Dalam A/B testing, kita membandingkan dua versi dari suatu elemen, seperti halaman web atau iklan, dengan tujuan untuk menentukan versi mana yang lebih efektif dalam mencapai tujuan bisnis kita*

*MDE umumnya dinyatakan sebagai persentase atau proporsi perbedaan antara kelompok yang dibandingkan. Misalnya, jika kita ingin mendeteksi perbedaan konversi antara dua kelompok dengan tingkat signifikansi 95%, kita mungkin memutuskan bahwa MDE yang dapat dideteksi adalah 5%. Artinya, jika perbedaan antara kedua kelompok kurang dari 5%, kita mungkin tidak dapat membedakan secara signifikan antara keduanya dalam eksperimen.*

*MDE umumnya dinyatakan sebagai persentase atau proporsi perbedaan antara kelompok yang dibandingkan. Misalnya, jika kita ingin mendeteksi perbedaan konversi antara dua kelompok dengan tingkat signifikansi 95%, kita mungkin memutuskan bahwa MDE yang dapat dideteksi adalah 5%. Artinya, jika perbedaan antara kedua kelompok kurang dari 5%, kita mungkin tidak dapat membedakan secara signifikan antara keduanya dalam eksperimen.*

1. Difference between control and treatment ->

**Note :**

For point a-c, you should define it by yourself. **Give any reason why you choose that number.**

For point d, it will be more profitable if the conversion rate increases 1%

1. How long run experiment

Note : you can make your own scenarios or assumptions in determining these points as prior information in designing experiments.

**Stage#4 – Running Experiment and Obtaining Data**

Since you can’t collect data directly, assume the following dataset is your experiment data. You can take a sample according to the experimental design you made in step #3

E.g from calculation of sample size, suppose that you get 50 users in each variant, so you can sample using Simple Random Sampling in this dataset to get sample size like your design

You can download the dataset at the following link

Case#1 : xxxx

**Stage#5 – Analyzing and Interpreting Data**

1. Ensure the trustworthiness
2. Check the data quality (missing value, duplicate data, distribution of data)
3. Data exploration (how many users in each group, and other insight from dataset you choose)
4. Perform SRM test with chi-square test
5. Conduct hypothesis testing and analyze the result
6. Calculate confidence interval of difference between treatment and control

**Stage#6 – Make conclusion and recommendation**

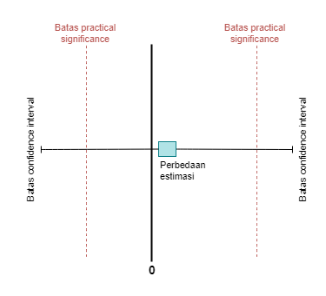
In this step, you must write down :

- The conclusions from the previous analysis

Statistically significant or not?

- Recommendation for the business?

When making recommendations, consider other things in the business such as costs etc, whether thery are practically significant or not. You can compare the confidence intercal of statistics with the confidence interval in practical

Example of comparison statistically significant with practically significant. You can make figure like this comparison :

Recommendation for the next experiment

From the result of the current experiment, if you want to do a similar experiment but with a different variant, what do you recommend?

**Stage#7 – Make easy report and presentation**

After you finish making an A/B testing plan and analyzing the dataset, you are asked to make an easy report and a presentation in ppt to present the analysis results obtained