

# Assignment 4

## E-Commerce A/B Testing

*by Aditya Bagus Pratama - DS03003*

*[link collab]*

# TUJUAN

- Melakukan A/B Testing untuk menguji apakah pembaharuan website e-commerce menggunakan design yang baru akan menghasilkan conversion rate yang berbeda secara statistik (statistically different) dibandingkan jika menggunakan design website yang lama.
- Menggunakan A/B Testing untuk memberikan rekomendasi pada perusahaan apakah design website yang baru layak digunakan, atau perlu dilakukan revisi design, atau cukup menggunakan design yang lama.

# DATASET

Dataset yang digunakan berisi data testing terhadap user di tiap sesinya, diperoleh dari [kaggle](#), memiliki 5 kolom dan 294.478 baris.

	user_id	timestamp	group	landing_page	converted
0	851104	2017-01-21 22:11:48.556739	control	old_page	0
1	804228	2017-01-12 08:01:45.159739	control	old_page	0
2	661590	2017-01-11 16:55:06.154213	treatment	new_page	0
3	853541	2017-01-08 18:28:03.143765	treatment	new_page	0
4	864975	2017-01-21 01:52:26.210827	control	old_page	1

**user\_id** = id participant dalam satu sesi

**timestamp** = waktu kunjungan

**landing\_page** = tipe desain (desain lama atau baru)

**group** = klasifikasi participant, group control disajikan desain *lawas*, group treatment disajikan desain baru.

**converted** = apakah user tertarik untuk membeli (1) atau tidak (0)

## Data Understanding and Cleansing

Dari proses data understanding, diketahui terdapat **3894** user\_id yang duplikat. Duplikasi user\_id ini disebabkan karena adanya *anomali* pada data, dimana terdapat **1928** user control yang sempat mendapatkan design baru, dan **1965** user treatment yang sempat mendapatkan design lawas. Hal ini tidak sesuai dengan ketentuan fitur group yang telah ditetapkan.

Maka dari itu, data dilakukan imputasi untuk menghilangkan anomali.

landing_page	new_page	old_page
group		
control	0	145274
treatment	145310	0

Setelah dilakukan cleansing, baris pada data menjadi 290.584 baris. Data kini siap dilakukan sampling dan testing.

## Sampling & Summary

Akan diambil sampel dengan jumlah yang diperlukan untuk pengujian. Method `DataFrame.sample()` pada library pandas, akan digunakan untuk melakukan Simple Random Sampling.

```
control      4720
treatment    4720
Name: group, dtype: int64
```

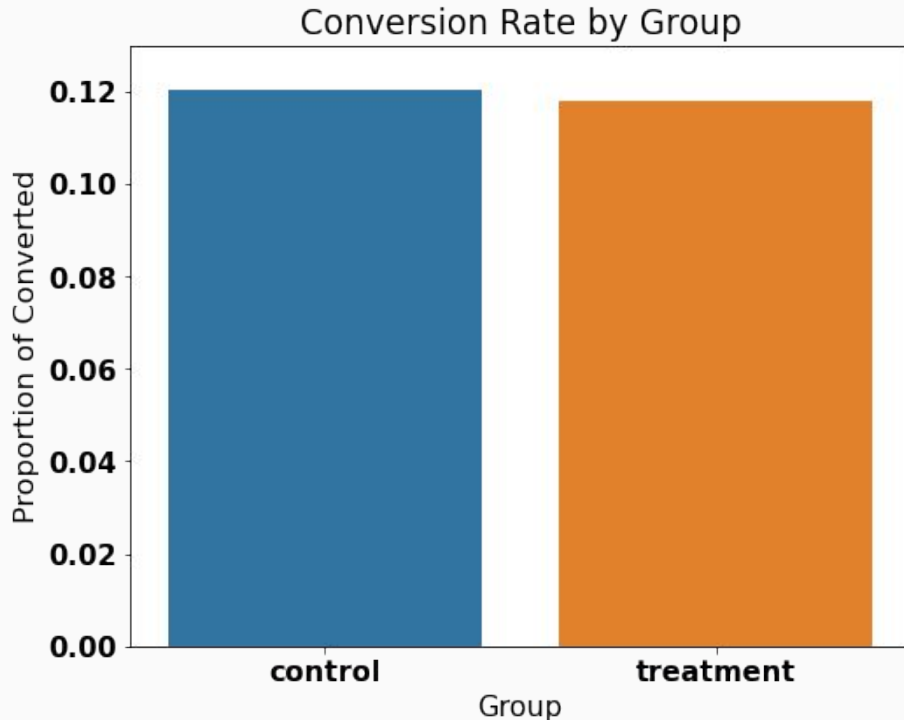
Dari hasil random sampling, akan digunakan 4720 data pada masing-masing group untuk selanjutnya dilakukan pengujian.

Menghitung beberapa statistik dasar untuk mengetahui bentuk sampel yang diambil.

	conversion_rate	std_deviation	std_error
group			
control	0.120	0.325	0.005
treatment	0.118	0.322	0.005

Dari statistic summary, nampak bahwa kedua desain memiliki kinerja yang mirip, desain website yang lama ternyata memiliki tingkat konversi rata-rata yang sedikit lebih baik.

## Visualization



## Hypothesis Testing

Akan dilakukan pengujian hipotesis menggunakan z-test. Nilai confidence interval 95% untuk masing” grup dan p-value juga akan diketahui. Berikut adalah hasilnya,

```
z statistic: 0.38  
p-value: 0.703  
ci 95% for control group: [0.111, 0.130]  
ci 95% for treatment group: [0.109, 0.127]
```

## KESIMPULAN

Berdasarkan pengujian, nilai p-value berada di atas ambang batas (0.05), maka Null Hipotesis  $H_0$  tidak dapat ditolak, yang berarti bahwa desain website yang baru tidak memiliki perbedaan kinerja yang statistically significant, (bahkan lebih buruk) daripada desain yang lama.

Dengan demikian, rekomendasi yang dihasilkan adalah untuk mempertahankan design lama, atau merevisi design baru, jika tetap ingin menyegarkan tampilan website.