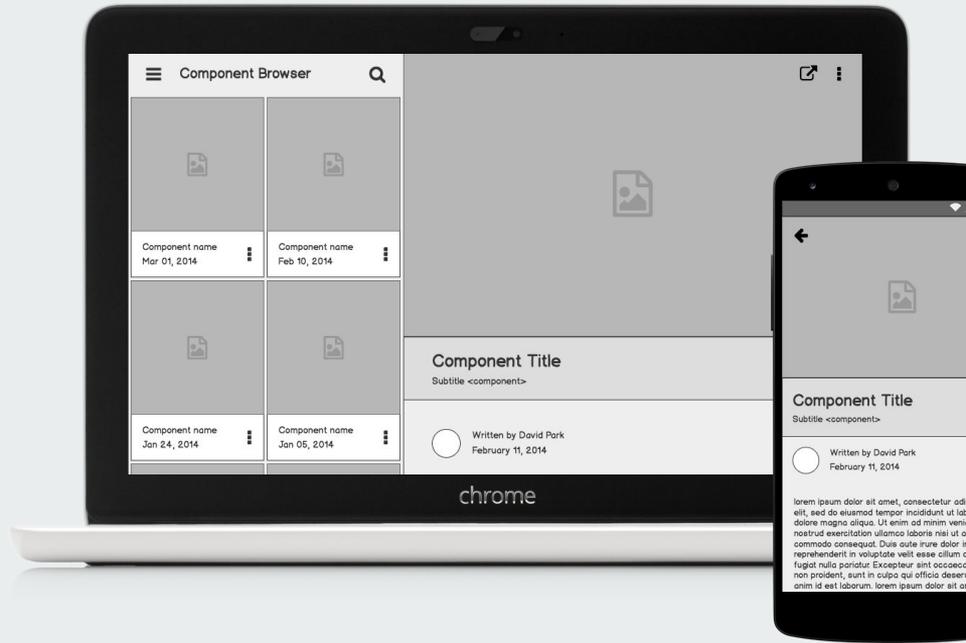




A/B TESTING Case Study

Anggita Ghozali



Deskripsi Data

Data yang digunakan berasal dari situs Kaggle.

Keterangan:

1. Campaign Name: Nama campaign/kampanya
2. Date: Tanggal pencatatan
3. Spend [USD]: Jumlah yang dihabiskan untuk kampanye dalam dolar
4. # of Impressions: Jumlah tayangan yang dilewati iklan melalui kampanye
5. Reach: Jumlah tayangan unik yang diterima dalam iklan
6. # of Website Clicks: Jumlah klik situs web yang diterima melalui iklan
7. # of Searches: Jumlah pengguna yang melakukan penelusuran di situs web
8. # of View Content: Jumlah pengguna yang melihat konten dan produk di situs web
9. # of Add to Cart: Jumlah pengguna yang menambahkan produk ke keranjang
10. # of Purchase: Jumlah pembelian



Dua kampanye dilakukan oleh perusahaan:

Control Campaign

Test Campaign

A/B Testing digunakan untuk menemukan kampanye terbaik bagi perusahaan untuk mendapatkan lebih banyak pelanggan.

Data Preparation

```
In [2]: import pandas as pd
```

```
In [3]: control_data = pd.read_csv("control_group.csv", sep = ";")  
test_data = pd.read_csv("test_group.csv", sep = ";")
```

```
In [4]: control_data.head()
```

```
Out[4]:
```

	Campaign Name	Date	Spend [USD]	# of Impressions	Reach	# of Website Clicks	# of Searches	# of View Content	# of Add to Cart	# of Purchase
0	Control Campaign	1.08.2019	2280	82702.0	56930.0	7016.0	2290.0	2159.0	1819.0	618.0
1	Control Campaign	2.08.2019	1757	121040.0	102513.0	8110.0	2033.0	1841.0	1219.0	511.0
2	Control Campaign	3.08.2019	2343	131711.0	110862.0	6508.0	1737.0	1549.0	1134.0	372.0
3	Control Campaign	4.08.2019	1940	72878.0	61235.0	3065.0	1042.0	982.0	1183.0	340.0
4	Control Campaign	5.08.2019	1835	NaN	NaN	NaN	NaN	NaN	NaN	NaN

```
In [5]: test_data.head()
```

```
Out[5]:
```

	Campaign Name	Date	Spend [USD]	# of Impressions	Reach	# of Website Clicks	# of Searches	# of View Content	# of Add to Cart	# of Purchase
0	Test Campaign	1.08.2019	3008	39550	35820	3038	1946	1069	894	255
1	Test Campaign	2.08.2019	2542	100719	91236	4657	2359	1548	879	677
2	Test Campaign	3.08.2019	2365	70263	45198	7885	2572	2367	1268	578
3	Test Campaign	4.08.2019	2710	78451	25937	4216	2216	1437	566	340
4	Test Campaign	5.08.2019	2297	114295	95138	5863	2106	858	956	768

Cek Missing Value

```
In [12]: #MISSING VALUE  
control_data.isnull().sum()
```

```
Out[12]: Campaign Name      0  
Date                      0  
Spend [USD]              0  
# of Impressions         1  
Reach                    1  
# of Website Clicks     1  
# of Searches            1  
# of View Content       1  
# of Add to Cart        1  
# of Purchase           1  
dtype: int64
```

```
In [13]: test_data.isnull().sum()
```

```
Out[13]: Campaign Name      0  
Date                      0  
Spend [USD]              0  
# of Impressions         0  
Reach                    0  
# of Website Clicks     0  
# of Searches            0  
# of View Content       0  
# of Add to Cart        0  
# of Purchase           0  
dtype: int64
```

Dapat diketahui bahwa Control Data memiliki missing value, sehingga perlu diimputasi

Imputasi Data

```
#imputasi data
control_data["# of Impressions"].fillna(value=control_data["# of Impressions"].mean(),
                                         inplace=True)
control_data["Reach"].fillna(value=control_data["Reach"].mean(),
                              inplace=True)
control_data["# of Website Clicks"].fillna(value=control_data["# of Website Clicks"].mean(),
                                             inplace=True)
control_data["# of Searches"].fillna(value=control_data["# of Searches"].mean(),
                                      inplace=True)
control_data["# of View Content"].fillna(value=control_data["# of View Content"].mean(),
                                          inplace=True)
control_data["# of Add to Cart"].fillna(value=control_data["# of Add to Cart"].mean(),
                                         inplace=True)
control_data["# of Purchase"].fillna(value=control_data["# of Purchase"].mean(),
                                      inplace=True)
```

Selanjutnya, buat dataset baru dengan menggabungkan kedua dataset:

```
ab_data = control_data.merge(test_data,  
                             how="outer").sort_values(["Date"])  
ab_data = ab_data.reset_index(drop=True)  
print(ab_data.head())
```

	Campaign Name	Date	Spend [USD]	# of Impressions	Reach \
0	Control Campaign	1.08.2019	2280	82702.0	56930.0
1	Test Campaign	1.08.2019	3008	39550.0	35820.0
2	Test Campaign	10.08.2019	2790	95054.0	79632.0
3	Control Campaign	10.08.2019	2149	117624.0	91257.0
4	Test Campaign	11.08.2019	2420	83633.0	71286.0

	# of Website Clicks	# of Searches	# of View Content	# of Add to Cart \
0	7016.0	2290.0	2159.0	1819.0
1	3038.0	1946.0	1069.0	894.0
2	8125.0	2312.0	1804.0	424.0
3	2277.0	2475.0	1984.0	1629.0
4	3750.0	2893.0	2617.0	1075.0

	# of Purchase
0	618.0
1	255.0
2	275.0
3	734.0
4	668.0



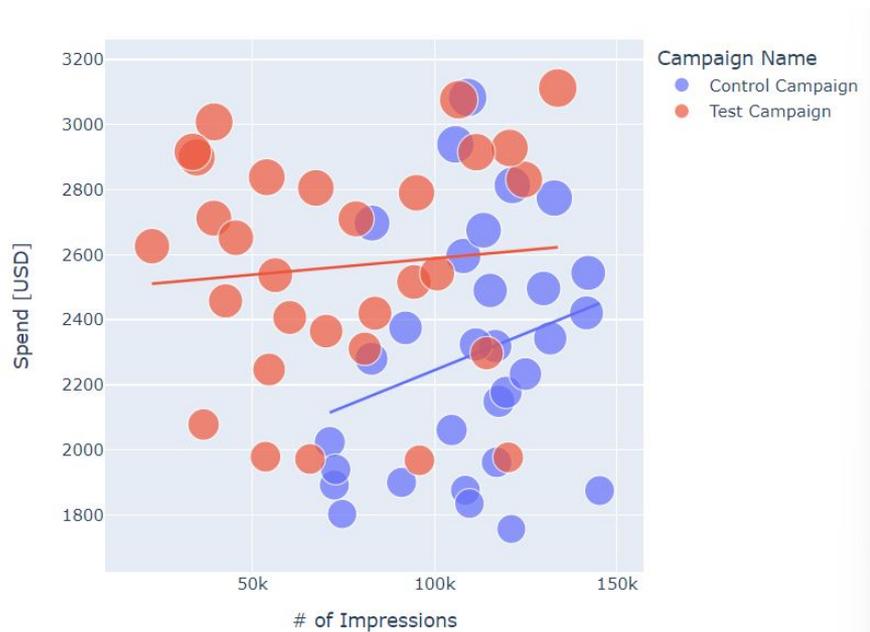
```
print(ab_data["Campaign Name"].value_counts())
```

```
Control Campaign    30  
Test Campaign      30  
Name: Campaign Name, dtype: int64
```

Control Campaign dan Test Campaign sudah memiliki jumlah sampel yang sama

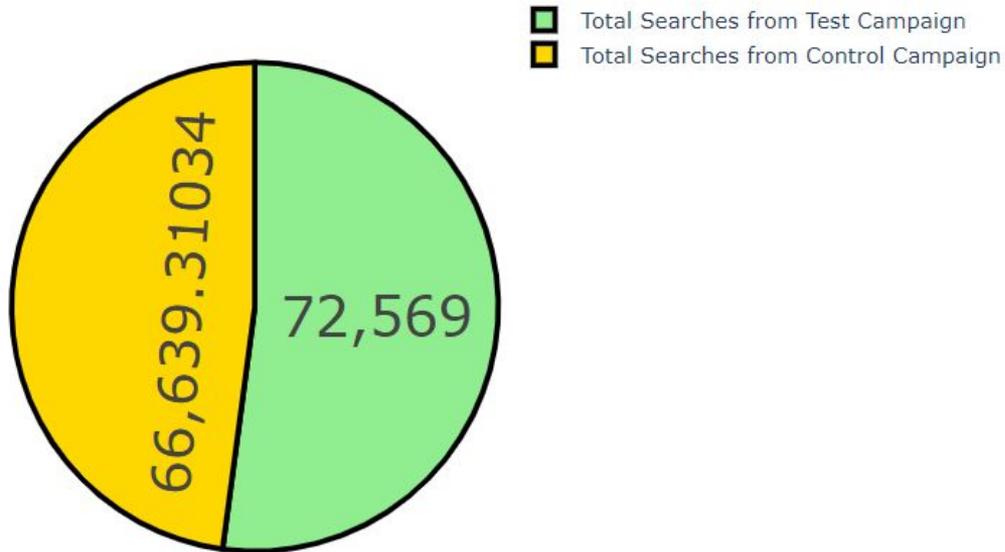
A/B TESTING

Pertama, disini akan dilihat hubungan antara jumlah tayangan yang didapatkan dari kedua kampanye dan jumlah yang dibelanjakan untuk kedua kampanye:



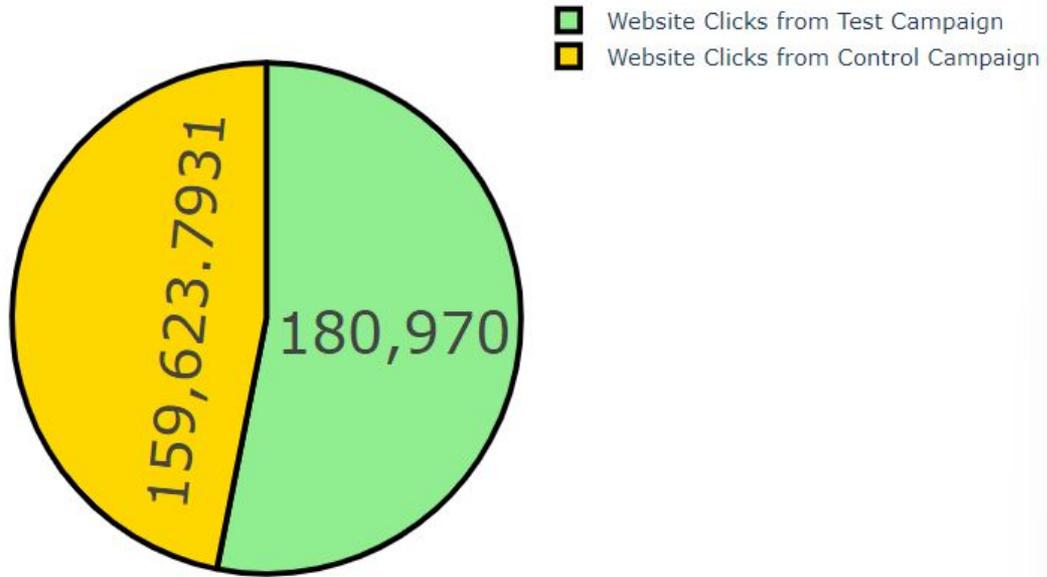
Control Campaign menghasilkan lebih banyak tayangan sesuai dengan jumlah yang dibelanjakan pada dua kampanye

Control Vs Test: Searches



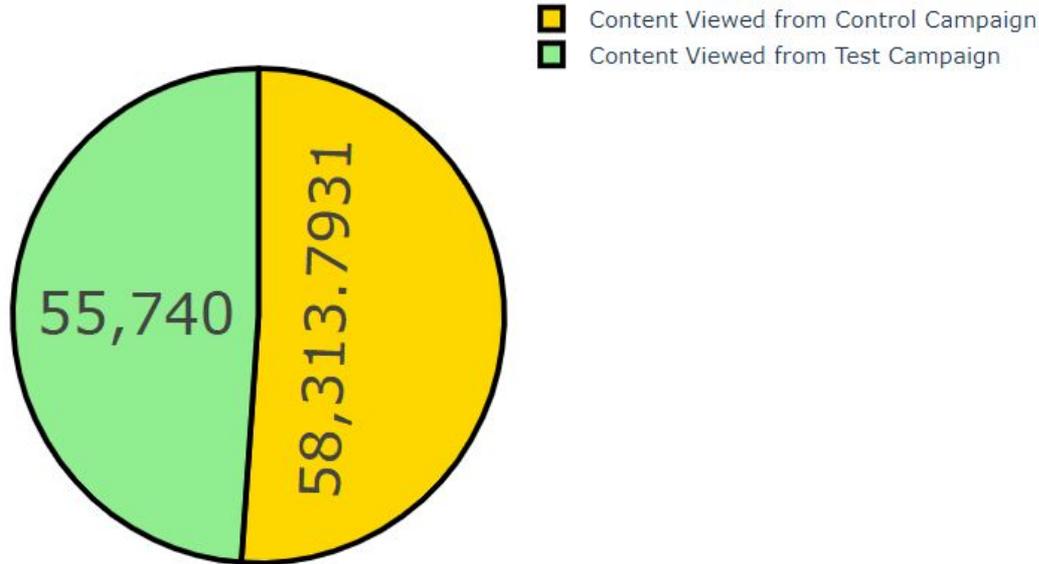
Pada jumlah penelusuran yang dilakukan di situs web kedua kampanye, didapatkan bahwa Test Campaign menghasilkan lebih banyak penelusuran dibandingkan Control Campaign

Control Vs Test: Website Clicks



Pada jumlah klik situs web dari kedua kampanye, didapatkan bahwa Test Campaign menghasilkan lebih banyak klik situs web dibandingkan Control Campaign

Control Vs Test: Content Viewed



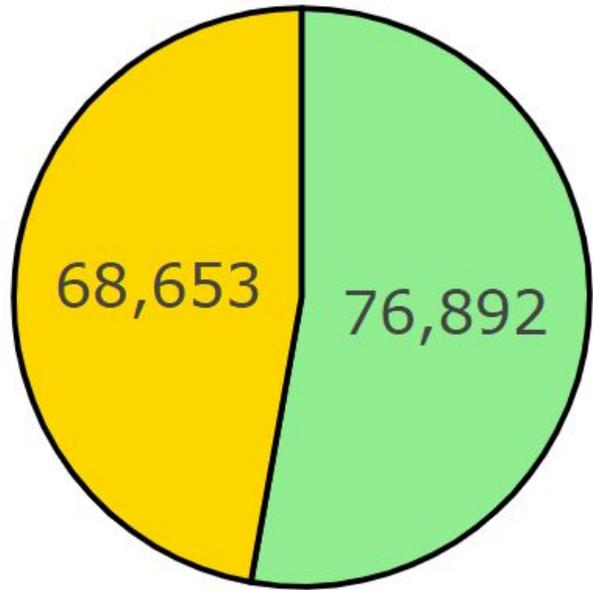
Penonton Control Campaign melihat lebih banyak konten daripada Test Campaign meskipun tidak terlalu jauh berbeda.

Control Vs Test: Added to Cart



Meskipun klik situs web rendah, namun Control Campaign menghasilkan lebih banyak produk yang ditambahkan daripada Test Campaign.

Control Vs Test: Spend [USD]



- Spend [USD] in Test Campaign
- Spend [USD] in Control Campaign

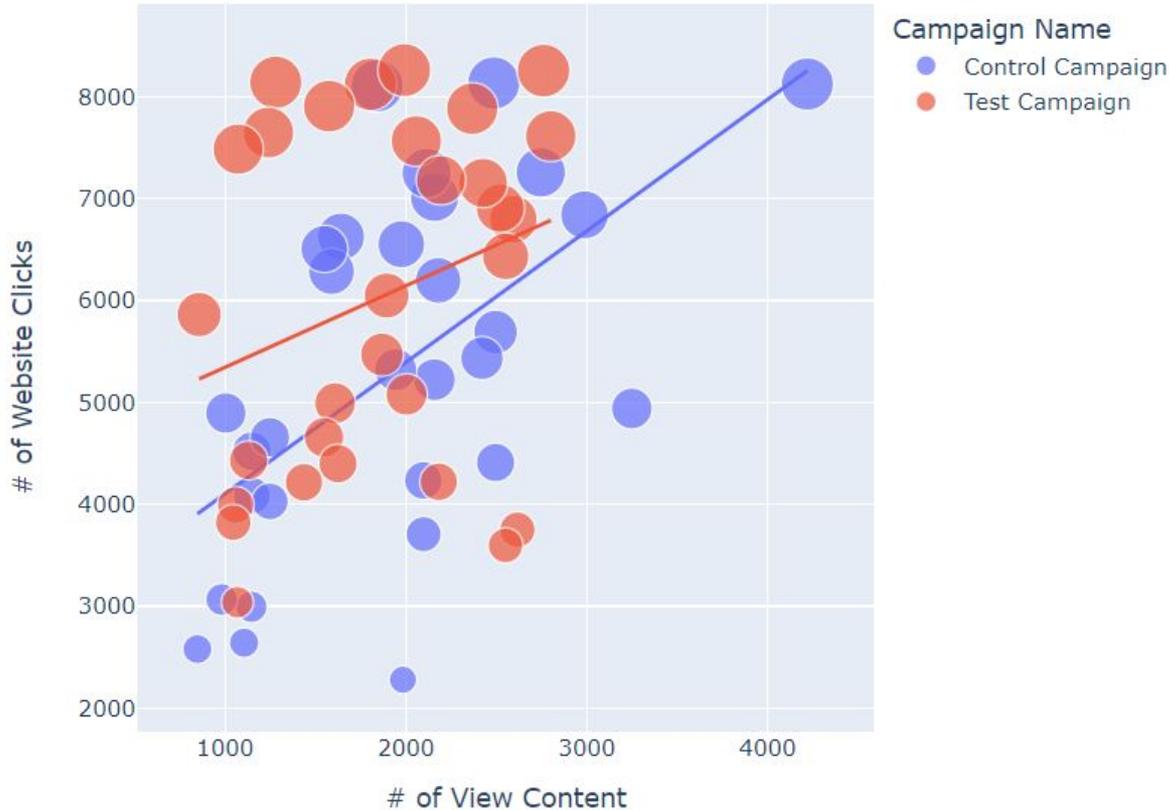
Jumlah yang dihabiskan untuk Test Campaign lebih banyak dibandingkan dengan Control Campaign.

Akan tetapi, Campaign Control menghasilkan lebih banyak content views dan produk yang ada di keranjang.

Control Vs Test: Purchase



Dari gambar disamping dapat diketahui bahwa perbedaan kedua kampanye menghasilkan pembelian yang tidak jauh berbeda. Akan tetapi, Control Campaign memang menghasilkan lebih banyak penjualan.



Selanjutnya, coba lihat korelasi lagi untuk mempertimbangkan antara Control Campaign dan Test Campaign. Dari Scatter Plot dapat diketahui bahwa Klik situs web lebih tinggi di Control Campaign daripada di Test Campaign

KESIMPULAN



Dari A/B Testing yang telah dilakukan, didapatkan kesimpulan bahwa:

Control Campaign menghasilkan lebih banyak penjualan dan keterlibatan dari pengunjung. Lebih banyak produk yang dilihat dari Control Campaign menghasilkan lebih banyak produk yang ditambahkan di keranjang, dan tentunya ke penjualan. Selain itu, Campaign Control lebih tepat untuk memasarkan produk ke audiensi yang lebih banyak.