

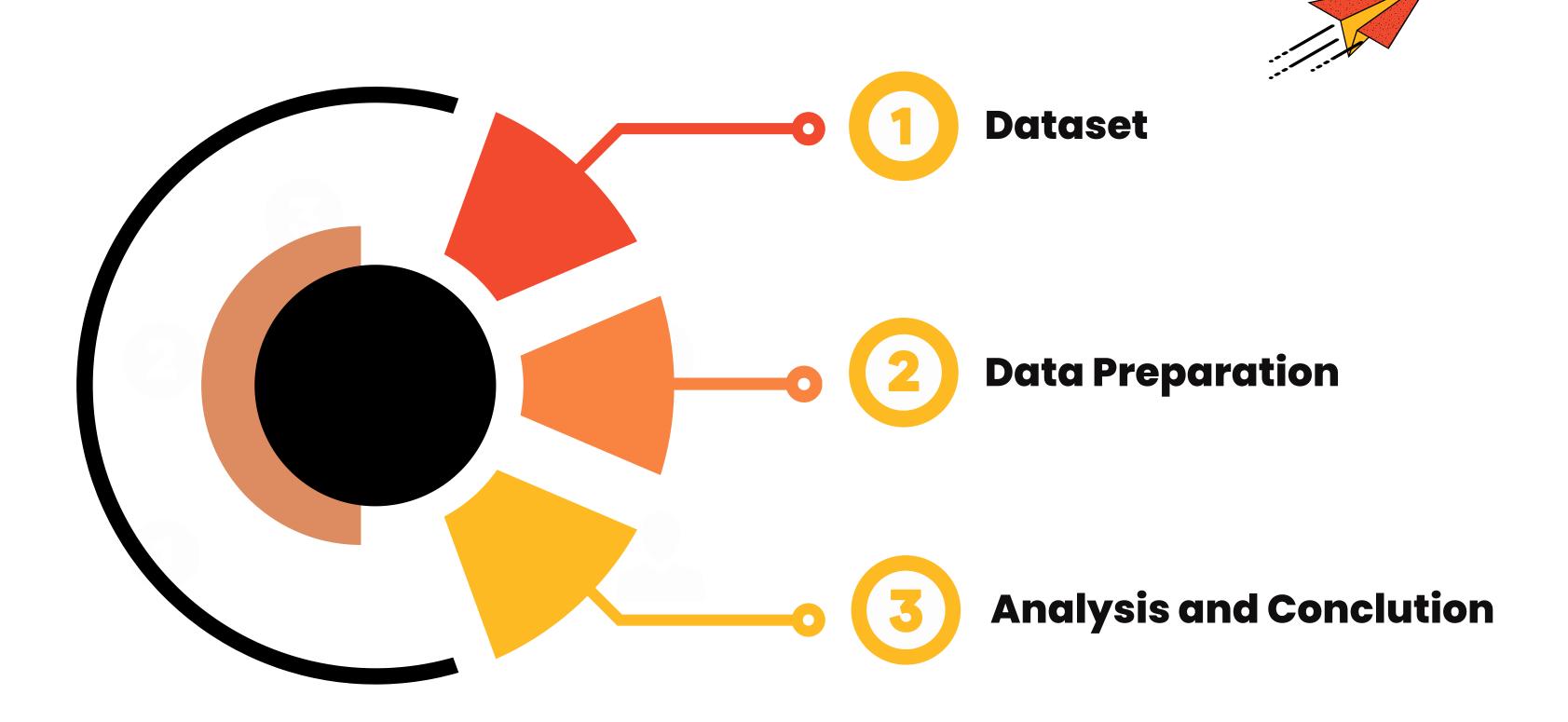
FINAL PROJECT PHYTON

Bootcamp Data Analysis by @myskill.id





Table Of Content



Dataset



Terdapat 4 dataset yang digunakan dalam menganalisis dan menjawab pertanyaan di atas yaitu order_detail, sku_detail, customer_detail, payment_detail. Data tersebut diambil dari salah satu Perusahaan e-commerce yang memungkinkannya mengumpulkan transaksi dari pelanggannya, melakukan pembelian selama 2 tahun terakhir (2021 dan 2022). Mengenaipenjelasan dataset adalah sebagai berikut:

order_detail:

1	ī.	d		
ı	ı	u		

→ angka unik dari order / id_order

customer_id

→ angka unik dari pelanggan

order_date

→ tanggal saat dilakukan transaksi

4. sku_id

→ angka unik dari produk (sku adalah stock keeping unit)

5. price

→ harga yang tertera pada tagging harga

qty_ordered

→ jumlah barang yang dibeli oleh pelanggan

before_discount

→ nilai harga total dari produk (price * qty_ordered)

8. discount_amount

→ nilai diskon product total

9. after_discount

→ nilai harga total produk ketika sudah dikurangi dengan diskon

10. is_gross

→ menunjukkan pelanggan belum membayar pesanan

11. is_valid

→ menunjukkan pelanggan sudah melakukan pembayaran

12. is_net

→ menunjukkan transaksi sudah selesai

13. payment_id

→ angka unik dari metode pembayaran

Dataset



sku_detail:

1. id

→ angka unik dari produk (dapat digunakan untuk key saat join)

sku_name

→ nama dari produk

base_price

→ harga barang yang tertera pada tagging harga / price

4. cogs

→ cost of goods sold / total biaya untuk menjual 1 produk

5. category

→ kategori produk

customer_detail:

1. id

- → angka unik dari pelanggan
- registered_date
- → tanggal pelanggan mulai mendaftarkan diri sebagai anggota

Payment_detail:

1. id

- → angka unik dari metode pembayaran
- payment_method
- → metode pembayaran yang digunakan

DATA PREPARATION



Mengimport library yang akan digunakan di phyton

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from pandas.tseries.offsets import BDay
```

Mengambil sumber data dari github

```
#Sumber data yang digunakan
path_od = "https://raw.githubusercontent.com/dataskillsboost/FinalProjectDA11/main/order_detail.csv"
path_pd = "https://raw.githubusercontent.com/dataskillsboost/FinalProjectDA11/main/payment_detail.csv"
path_cd = "https://raw.githubusercontent.com/dataskillsboost/FinalProjectDA11/main/customer_detail.csv"
path_sd = "https://raw.githubusercontent.com/dataskillsboost/FinalProjectDA11/main/sku_detail.csv"
df_od = pd.read_csv(path_od)
df_pd = pd.read_csv(path_pd)
df_cd = pd.read_csv(path_cd)
df_sd = pd.read_csv(path_sd)
```

DATA PREPARATION



Menjalankan library SQL di Google colab

```
#Menjalankan SQL di Colab
from sqlite3 import connect
conn = connect(':memory:')
df_od.to_sql('order_detail',conn, index=False, if_exists='replace')
df_pd.to_sql('payment_detail', conn, index=False, if_exists='replace')
df_sd.to_sql('sku_detail', conn, index=False, if_exists='replace')
df_cd.to_sql('customer_detail', conn, index=False, if_exists='replace')

3998
```

Menggabungkan data menggunakan SQL Query

```
#Query SQL untuk menggabungkan data
df = pd.read_sql("""
SELECT
    order detail.*,
    payment detail.payment method,
    sku detail.sku name,
    sku detail.base price,
    sku detail.cogs,
    sku detail.category,
    customer detail.registered date
FROM order detail
LEFT JOIN payment detail
    on payment detail.id = order detail.payment id
LEFT JOIN sku detail
    on sku_detail.id = order_detail.sku_id
LEFT JOIN customer detail
    on customer detail.id = order detail.customer id
   conn)
```

DATA PREPARATION





Mengubah Type data agar data mudah dikelola

```
#Mengubah tipe data agar mudah dilakukan pengolahan data
df = df.astype({"before_discount":'int', "discount_amount":'int',\
              "after discount": 'int', "base price": 'int'})
df['order_date']= pd.to_datetime(df['order_date'])
df['registered date']= pd.to datetime(df['registered date'])
df.dtypes
                           object
id
customer id
                           object
order date
                   datetime64[ns]
sku id
                           object
price
                            int64
gty ordered
                            int64
before discount
                            int64
discount amount
                            int64
after discount
                            int64
is gross
                            int64
is_valid
                            int64
is net
                            int64
payment_id
                            int64
payment method
                           object
sku name
                           object
base price
                            int64
cogs
                            int64
                           object
category
registered date
                   datetime64[ns]
dtype: object
```



```
#Mengubah tipe kolom Date menjadi Datetime
df['order_date'] = pd.to_datetime(df['order_date'])
df['registered_date']= pd.to_datetime(df['registered_date']
df.dtypes
id
                           object
customer id
                           object
                   datetime64[ns]
order date
sku id
                           object
price
                            int64
gty ordered
                            int64
before discount
                            int64
discount amount
                            int64
after discount
                            int64
is gross
                            int64
is_valid
                            int64
                            int64
is net
                            int64
payment id
payment method
                           object
sku name
                           object
                            int64
base price
                            int64
cogs
                           object
category
registered date
                   datetime64[ns]
dtype: object
```

PROBLEM STATEMENT



Question 1

Dear Data Analyst,

Akhir tahun ini, perusahaan akan memberikan hadiah bagi pelanggan yang memenangkan kompetisi Festival Akhir Tahun. Tim Marketing membutuhkanbantuan untuk menentukan perkiraan hadiah yang akan diberikan pada pemenang kompetisi nantinya. Hadiah tersebut akan diambil dari TOP 5 Produk dari Kategori Mobiles & Tablets selamatahun 2022, dengan jumlah kuantitas penjualan (valid = 1) paling tinggi.

Mohon bantuan, untuk mengirimkan data tersebut sebelum akhir bulan ini ke Tim Marketing. Atas bantuan yang diberikan, kami mengucapkan terima kasih.

Regards Tim Marketing





INPUT

```
# Memfilter Is_valid = 1
condition_1 = df['is_valid']==1
# Memfilter Category = Mobiles & Tables
condition_2 = df['category']=='Mobiles & Tablets'
# Memfilter tahun = 2022
condition 3 = df['year']==2022
        # Memamnggil seluruh Variable filter condition
df no1 = df[(condition 1) & (condition 2) & (condition 3)]\
        # Menggroup by berdasarkan Sku_name
        .groupby(by=["sku_name"])["qty_ordered"]\
        # Menggunakan fungsi agregat Sum pada kolom Qty ordered
        .sum()\
        # Mereset Nama Header
        .reset_index(name='total_qty_2022')\
        # Mengurutkan data berdasarkan kolom total qty 2022
        .sort_values(by = 'total_qty_2022', ascending=False)
# Menggunakan fungsi head untuk menampilakan Top 5 Produk
df_no1.head()
```

OUTPUT

	sku_name	total_qty_2022
1	IDROID_BALRX7-Gold	1000
2	IDROID_BALRX7-Jet black	31
3	Infinix Hot 4-Gold	15
43	samsung_Grand Prime Plus-Black	11
34	infinix_Zero 4-Grey	10

Berdasarkanhasil output diatas menampilkan bahwa Top 5 produk dalam category Mobiles dan Tablets yang berhakmendapatkan hadiah dari Tim Marketinng pada kompetisi Festival Akhir Tahun adalah IDROID_BALRX7-Gold, IDROID_BALRX7-Jet black, Infinix Hot 4-Gold, Samsung_Grand prime Plus-Black, dan infinix_Zero 4-Grey.

PROBLEM STATEMENT



Question 2

Dear Data Analyst,

Menindak lanjuti meeting gabungan Tim Werehouse dan Tim Marketing, kami menemukan bahwa ketersediaan stock produk dengan Kategori Others pada akhir 2022 kemarin masih banyak.

- 1.Kami mohon bantuan untuk melakukan pengecekan data penjualan kategori tersebut dengan tahun 2021 secara kuantitas penjualan. Dugaan sementara kami, telah terjadi penurunan kuantitas penjualan pada 2022 dibandingkan 2021. (Mohon juga menampilkan data ke-15 kategori)
- 2. Apabila memang terjadi penurunan kuantitas penjualan pada kategori Others, kami mohon bantuan untuk menyediakan data TOP 20 nama produk yang mengalami penurunan paling tinggi pada 2022 jika dibanding dengan 2021. Hal ini kami gunakan sebagai bahan diskusi pada meeting selanjutnya.

Mohon bantuan untuk mengirimkan data tersebut paling lambat 4 hari dari hari ini. Atas bantuan yang diberikan, kami mengucapkan terima kasih.

Regards Tim Werehouse



2.1

Melakukan pengecekan jumlah penjualan **percategory** di tahun 2021 dan 2022 serta menggabungkan kedua tabel



INPUT

```
# Memfilter Is valid =1
df_qty21 = df[(df['is_valid']==1) &\
        # Memfilter tahun = 2021
        (df['year']==2021)]\
        # Mengroupby berdasarkan kolom category
        .groupby(by=["category"])["qty_ordered"]\
        # Melakukan fungsi Agregat sum
        .sum()\
        # Mereset nama header
        .reset_index(name='total_qty_2021')\
        # Mengurutkan berdasarkan kolom Total_qty_2021
        .sort_values(by = 'total_qty_2021' ,ascending=False)
df_qty21
df_qty22 = df[(df['is_valid']==1) &\
        # Memfilter tahun = 2022
        (df['year']==2022)]\
        .groupby(by=["category"])["qty_ordered"]\
        .sum()\
        .reset index(name='total qty 2022')\
        .sort_values(by = 'total_qty_2022' ,ascending=False)
df_qty22
# Melakukan fungsi merge untuk kedua data penjualan
df no2 = df qty21.merge(df qty22, left on = 'category', right on = 'category')
# Growth = penjualan di tahun terakhir - penjualan di tahun sebelumnya
df_no2['qty_growth']=df_no2['total_qty_2022']-df_no2['total_qty_2021']
df no2.sort values(by = ['qty growth'],ascending=True,inplace=True)
df no2
```



2.1

Berdasarkan hasil output menampilkan bahwa dari 15 kategori, ada 4 kategori yang mengalami penurunan kuantitas penjualan di tahun 2022. Kategori tersebut diantaranya Others, Soghaat, Men Fashion, dan Beauty & Grooming.

Penurunan kuantitas penjualan terbesar adalah kategori **Others** dengan angka penurunan mencapai **163 dari tahun 2021**, sedangkan peningkatan kuantitas penjualan terbesar adalah **Mobiles dan Tablets** denganangkapeningkatan mencapai **1047 dari tahun 2021**

OUTPUT

	category	total_qty_2021	total_qty_2022	qty_growth	E
1	Others	426	263	-163	(
0	Soghaat	759	612	-147	
3	Men Fashion	237	175	-62	
9	Beauty & Grooming	168	153	-15	
7	Books	171	195	24	
11	Appliances	124	148	24	
6	Health & Sports	173	200	27	
12	Computing	109	153	44	
5	School & Education	184	237	53	
4	Home & Living	193	250	57	
8	Kids & Baby	170	227	57	
14	Entertainment	77	150	73	
2	Superstore	327	536	209	
10	Women Fashion	140	489	349	
13	Mobiles & Tablets	107	1154	1047	



2.2

Melakukan pengecekan jumlah penjualan pada kategori "**Others**" di tahun 2021 dan 2022 serta menggabungkan kedua tabel

INPUT

```
others_21 = df[(df['is_valid']==1) &\
        # Memfilter tahun = 2021
        (df['year']==2021)&\
        (df['category']=='Others')]\
        .groupby(by=["sku_name"])["qty_ordered"]\
        .sum()\
        .reset_index(name='others_qty_2021')\
        .sort_values(by = 'others_qty_2021' ,ascending=False)
others 21
others_22 = df[(df['is_valid']==1) &\
        # Memfilter tahun = 2022
        (df['year']==2022)&\
        (df['category']=='Others')]\
        .groupby(by=["sku_name"])["qty_ordered"]\
        .sum()\
        .reset index(name='others qty 2022')\
        .sort_values(by = 'others qty 2022' ,ascending=False)
others_22
# menggabungkan data
df_no22 = others_21.merge(others_22, left_on='sku_name', right_on='sku_name')
df_no22['Other_qty_growth']=df_no22['others_qty_2021']-df_no22['others_qty_2022']
df_no22.sort_values(by=['Other_qty_growth'], ascending=True, inplace=True)
df_{no22} = df_{no22.head(20)}
df_no22
```

2.2

Berdasarkan hasil output menampilkan bahwa dari 20 produk pada kategori **Others**, ada 9 produk yang mengalami penurunan kuantitas penjualan di tahun 2022. Penurunan kuantitas terbesar pada kategori Others adalah **Produk RB_ Dettol Germ Busting Kit-bf** dengan angka penurunan mencapai **155 dari tahun 2021**,

sedangkan peningkatan kuantitas terbesar pada kategori Others adalah **Produk Emart_Tyre Shape Air Compressor** dengan angka peningkatan mencapai **29 daritahun 2021**

OUTPUT

	sku_name	others_2021	others_2022	Other_qty_growth
0	RB_Dettol Germ Busting Kit-bf	200	45	-155
1	Telemall_MM-DR-HB-L	23	2	-21
2	kansai_NeverWet	10	1	-9
3	emart_00-1	7	1	-6
5	MEGUIAR_G12711	4	1	-3
10	aw_Ultra Shine Wash & Wax-64oz./1893ml	2	1	-1
8	Entertainer Asia_Vouch 365-2017 Mobile App Lahore	2	1	-1
9	Tmall_MM-DR-PAD	2	1	-1
6	Aladdin_bike_cover	3	2	-1
12	electro_Humidifier	1	1	0
15	Lacie_9000146	1	1	0
16	MEGUIAR_A1214	1	1	0
17	MEGUIAR_G18211	1	2	1
11	aw_CONSTRUCTION FOAM-700ml	1	2	1
14	Entertainer Asia_Vouch 365 - 2017 Book Karachi	1	2	1
18	vitamin_265	1	2	1
7	kansai_Undercoating Aerosol	3	6	3
13	emart_0-37	1	7	6
4	emart_Tyre Shape Air Compressor	5	34	29

PROBLEM STATEMENT



Question 3

Dear Data Analyst,

Terkait ulang tahun perusahaan pada 2 bulan mendatang, Tim Digital Marketing akan memberikan informasi promo bagi pelanggan pada akhir bulan ini. Kriteria pelanggan yang akan kami butuhkan adalah mereka yang sudah melakukan check-out namun belum melakukan pembayaran (is_gross = 1) selama tahun 2022. Data yang kami butuhkanadalah ID Customer dan Registered Date.

Mohon bantuan, untuk mengirimkan data tersebut sebelum akhir bulan ini ke Tim Digital Marketing. Atas bantuan yang diberikan, kami mengucapkan terima kasih.

Regards

Tim Digital Marketing



INPUT

INPUT

```
#Jalankan kode ini untuk mendownload file
from google.colab import files

df_no3.to_csv('audience_list.csv', encoding = 'utf-8-sig',index=False)
files.download('audience_list.csv')

audience_list(1).csv
```

20.6 KB • Done



OUTPUT

	customer_id	registered_date			
9	C246762L	2022-05-08			
18	C848774L	2021-11-07			
19	C693415L	2022-04-12			
21	C180595L	2022-04-22			
22	C587425L	2022-03-22			
		•••			
5856	C394076L	2021-10-12			
5859	C248585L	2022-07-10			
5865	C471304L	2022-05-13			
5881	C265450L	2022-02-17			
5883	C676393L	2021-07-27			
1052 rows × 2 columns					

Berikut adalah data **customer_id** dan **Registered_date** customer yang mana telah melakukan check-out namun belum melakukan pembayaran selama tahun 2022, Serta hasil download dengan format **.csv**



OUTPUT

	customer_id	registered_date
9	C246762L	2022-05-08
18	C848774L	2021-11-07
19	C693415L	2022-04-12
21	C180595L	2022-04-22
22	C587425L	2022-03-22
5856	C394076L	2021-10-12
5859	C248585L	2022-07-10
5865	C471304L	2022-05-13
5881	C265450L	2022-02-17
5883	C676393L	2021-07-27
1052 ro	ws × 2 columns	

Pada Case question 3 ada beberapa hal yang dapat dilakukan tim digital marketing agar dapat melakukan promosi akhir bulan yang efektif sebagai berikut:

- 1. Analisis Data dan Penelusuran: Lacak pola perilaku pelanggan yang meninggalkan transaksi tanpa menyelesaikan pembayaran. Analisis data ini dapat membantu dalam memahami alasan di balik keluar dari proses pembayaran dan merancang solusi yang tepat.
- 2. Pengingat Pembayaran: Kirimkan pengingat atau notifikasi kepada pelanggan melalui email atau pesan teks untuk mengingatkan mereka bahwa ada transaksi yang belum diselesaikan. Berikan informasi tentang langkah-langkah yang perlu dilakukan untuk menyelesaikan pembayaran.
- 3. Penawaran Khusus dengan Batas Waktu: Berikan penawaran khusus kepada pelanggan untuk menyelesaikan pembayaran dengan batas waktu tertentu. Diskon khusus atau bonus tambahan dapat menjadi insentif yang kuat untuk mendorong mereka menyelesaikan transaksi.
- 4. Pilihan Pembayaran yang Fleksibel: Sediakan beberapa pilihan pembayaran yang lebih fleksibel bagi pelanggan. Ini dapat mencakup opsi cicilan, pembayaran tunai di tempat, atau opsi pembayaran dengan metode lain yang lebih sesuai bagi mereka.

PROBLEM STATEMENT



Question 4

Dear Data Analyst,

Pada bulan October hingga Desember 2022, kami melakukan campaign setiap hari Sabtu dan Minggu. Kami hendak menilai, apakah campaign tersebut cukup berdampak pada kenaikan penjualan (before_discount). Mohon bantuan untuk menampilkan data:

- 1.Rata-rata harian penjualan weekends (Sabtu dan Minggu) vs rata-rata harian penjualan weekdays (Senin-Jumat) per bulan tersebut. Apakah ada peningkatan penjualan pada masing-masing bulan tersebut.
- 2.Rata-rata harian penjualan weekends (Sabtu dan Minggu) vs rata-rata harian penjualan weekdays (Senin-Jumat) keseluruhan 3 bulan tersebut.

Mohon bantuan untuk mengirimkan data tersebut paling lambat minggu depan. Atas bantuan yang diber<mark>ikan,</mark> kami mengucapkan terima kasih.

Regards Tim Campaign





4.1

Melakukan pengecekan jumlah penjualan pada hari **Weekdays** dan **Weekend** di tahun 2022 serta menggabungkan kedua tabel

INPUT

```
df_weekend = pd.DataFrame(df[\
            # Memfilter data is_valid =1
             (df['is valid']==1) &\
            # Memfilter data tahun =2022
             (df['year']==2022) &\
             (df['month'].isin(['October','November','December']))&\
            # Melakukan fungsi isin pada kolom day "saturday dan sunday"
             (df['day'].isin(['Saturday','Sunday'])) ]\
            # Melakukan Group By dan fungsi agregat mean
             .groupby(by = ["month", 'month_num'])["before_discount"].mean()\
            # Melakukan Reset pada Header
             .reset_index(name='profit_avg_weekend').round()\
            # Mengurutkan data berdasarkan kolom profit avg weekend
             .sort_values(by = 'profit_avg_weekend' ,ascending=False))
df weekend
df_weekdays = pd.DataFrame(df[(df['is_valid']==1) &\
             (df['month'].isin(['October','November','December']))&\
             (df['day'].isin(['Monday', 'Tuesday', 'Wednesday', 'Thusday', 'Friday'])) &\
             (df['year']==2022)]\
             .groupby(by = ['month','month_num'])['before_discount'].mean()\
             .reset_index(name='profit_avg_weekdays')\
             .round()\
             .sort_values(by = 'profit_avg_weekdays' ,ascending=False))
df weekdays
# Melakukan merge terhadap dua kolom
df_no41 = df_weekend.merge(df_weekdays, left_on='month', right_on='month')
df_no41.sort_values(by = ['month_num_x'],ascending=True,inplace=True)
df_no41 = df_no41[['month','profit_avg_weekend', 'profit_avg_weekdays']]
df_no41
```

4.1

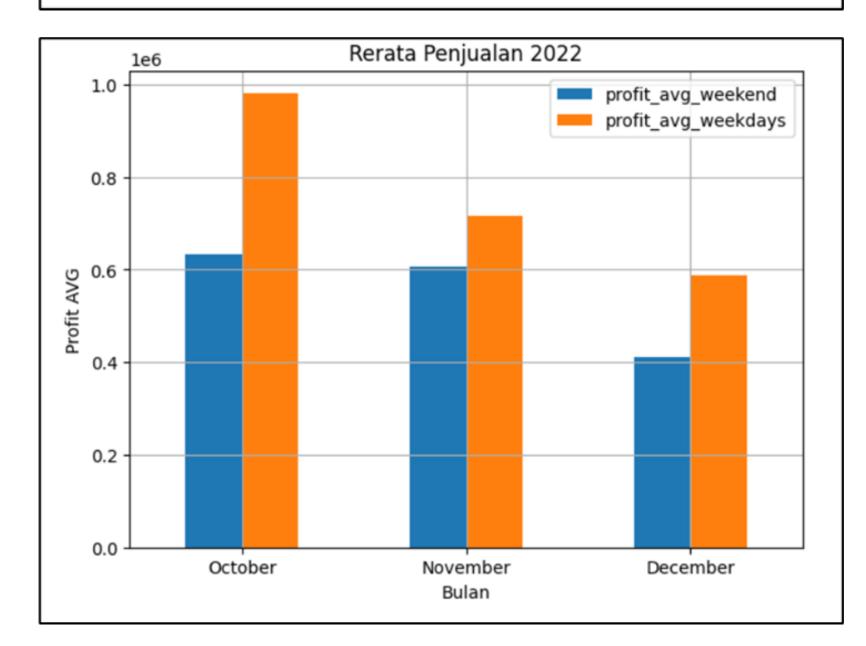
Berdasarkan grafik disamping menunjukan bahwa rata rata profit penjualan 3 bulan terakhir pada **Weekend** dan **Weekdays** mengalami penurunan secara berturut turut,

namun dibandingkan weekend rata rata penjualan pada weekdays lebihtinggi. Dapat disimpulkan bahwa campaign yang dilakukan setiap weekend kurang efektif.



OUTPUT

	month	<pre>profit_avg_weekend</pre>	<pre>profit_avg_weekdays</pre>
0	October	634260.0	980851.0
1	November	607794.0	715893.0
2	December	410599.0	587475.0



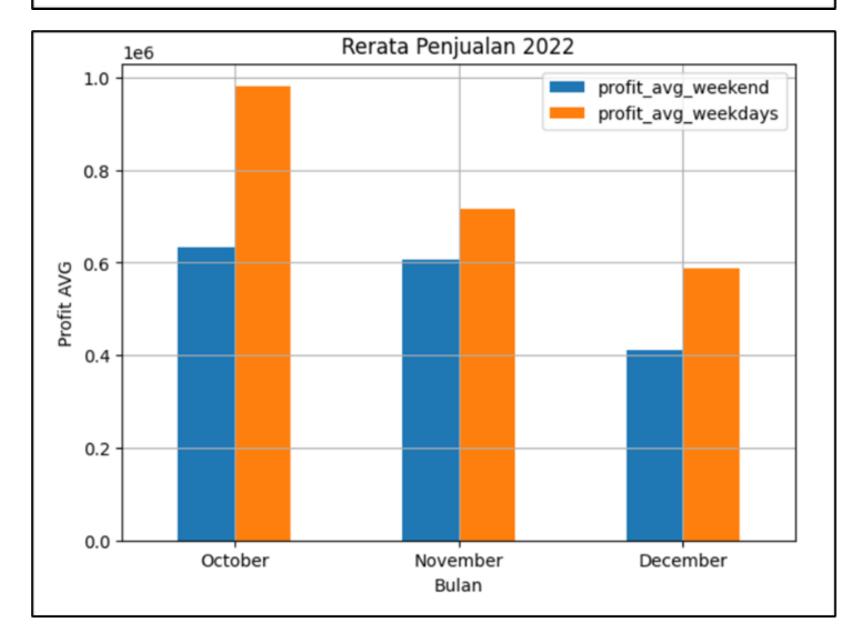


OUTPUT

Pada Case question 4.1 ada beberapa hal yang dapat dilakukan tim Campaign agar dapat meningkatkan penjualan yang efektif sebagai berikut:

- 1. A**nalisis Data**: Lakukan analisis mendalam terhadap data penjualan produk selama 3 bulan terakhir tahun 2022. Identifikasi produk apa saja yang mengalami penurunan dan temukan penyebabnya.
- 2. **Promosi dan Diskon Khusus:** Tawarkan promosi terbatas atau diskon spesial menjelang akhir tahun untuk mendorong pembelian. Penawaran eksklusif ini dapat menjadi dorongan bagi pelanggan untuk segera bertransaksi.
- 3. Paket Bundel atau Penawaran Spesial: Rancang paket bundel yang menarik atau penawaran khusus yang mencakup beberapa produk atau layanan dengan harga yang menarik. Ini bisa menjadi cara efektif untuk meningkatkan nilai pesanan per transaksi.

	month	profit_avg_weekend	profit_avg_weekdays
0	October	634260.0	980851.0
1	November	607794.0	715893.0
2	December	410599.0	587475.0





4.2

Melakukan pengecekan jumlah penjualan secara keseluruhan pada hari **Weekdays** dan **Weekend** di tahun 2022 serta menggabungkan kedua tabel

INPUT



4.2

Berdasarkanhasil output dibawah menunjukan bahwa secara keseluruhan dalam kurun waktu 3 bulan terakhir yaitu **Oktober hingga Desember 2022** rata rata penjualan Weekend sebanyak 558865,15 dan pada Weekdays sebanyak 751972,85, sehingga perbandingan kedua parameter tersebut **Weekdays** memiliki nilai rata rata **lebih besar 34,55%** dibandingkan penjualan pada **Weekend**

OUTPUT

	periode	Weekend Avg Sales	Weekdays Avg Sales	Diff value	Diff (%)	
0	Total 3 Month	558865.15	751972.85	-193107.7	-34.55%	





THANKYOU

- in linkedin.com/in/tfqhidayat
- 082251404245

- taufiqhidayat1807@gmail.com
- @fiq.hl

