

FINAL PROJECT

PYTHON

DYON SETIO UTOMO

Tools

Dalam final project ini, **Google Colab** digunakan untuk menjalankan dan mengeksekusi bahasa pemrograman Python. Library yang digunakan dalam proyek ini meliputi **Pandas**, **Numpy**, **Matplotlib**, dan **Seaborn**.



Dataset

```
#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)
```

Mengambil Data dari Github

Data yang digunakan merupakan data dari Tokopedia (**bukan data asli**). Terdapat empat jenis data utama yang digunakan, yaitu: **order_detail**, **sku_detail**, **customer_detail**, dan **payment_detail**.

Dataset

variable	class	description
order_detail:		
id	object	angka unik dari order / id_order
customer_id	object	angka unik dari pelanggan
order_date	object	tanggal saat dilakukan transaksi
sku_id	object	angka unik dari produk (sku adalah stock keeping unit)
price	int64	harga yang tertera pada tagging harga
qty_ordered	int64	jumlah barang yang dibeli oleh pelanggan
before_discount	float64	nilai harga total dari produk (price * qty_ordered)
discount_amount	float64	nilai diskon product total
after_discount	float64	nilai harga total produk ketika sudah dikurangi dengan diskon
is_gross	int64	menunjukkan pelanggan belum membayar pesanan
is_valid	int64	menunjukkan pelanggan sudah melakukan pembayaran
is_net	int64	menunjukkan transaksi sudah selesai
payment_id	int64	angka unik dari metode pembayaran

sku_detail:		
id	object	angka unik dari produk (dapat digunakan untuk key saat join)
sku_name	object	nama dari produk
base_price	float64	harga barang yang tertera pada tagging harga / price
cogs	int64	cost of goods sold / total biaya untuk menjual 1 produk
category	object	kategori produk

customer_detail:		
id	object	angka unik dari pelanggan
registered_date	object	tanggal pelanggan mulai mendaftarkan diri sebagai anggota

payment_detail:		
id	int64	angka unik dari metode pembayaran
payment_method	object	metode pembayaran yang digunakan

Data Preparation

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

Mengimpor modul **Pandas**, **Numpy**, **Matplotlib**, dan **Seaborn** agar dapat memanfaatkan fitur-fitur dalam *library* tersebut serta mempermudah penggunaan *syntax*.

```
#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')
```

Menjalankan fungsi SQL di Google Colab

```
#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)
```

Query untuk menggabungkan kolom **payment_detail**, **sku_detail**, dan **customer_detail** ke dalam tabel **order_detail**

Data Preparation

```
#Mengubah tipe data agar mudah dilakukan pengolahan data  
df = df.astype({"before_discount":'int', "discount_amount":'int', "after_discount":'int',"base_price":'int'})  
df.dtypes
```

id	object
customer_id	object
order_date	object
sku_id	object
price	int64
qty_ordered	int64
before_discount	float64
discount_amount	float64
after_discount	float64



id	object
customer_id	object
order_date	object
sku_id	object
price	int64
qty_ordered	int64
before_discount	int64
discount_amount	int64
after_discount	int64

Data Preparation

```
#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
order_date	object
sku_id	object
price	int64
qty_ordered	int64
before_discount	float64
discount_amount	float64
after_discount	float64



id	object
customer_id	object
order_date	datetime64[ns]
sku_id	object
price	int64
qty_ordered	int64
before_discount	int64
discount_amount	int64
after_discount	int64

Kasus 1

Dear Data Analyst,

Akhir tahun ini, perusahaan akan memberikan hadiah bagi pelanggan yang memenangkan kompetisi **Festival Akhir Tahun**. Tim Marketing membutuhkan bantuan untuk menentukan perkiraan hadiah yang akan diberikan pada pemenang kompetisi nantinya. Hadiah tersebut akan diambil dari **TOP 5** Produk dari Kategori **Mobiles & Tablets** selama tahun 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

Jawaban Kasus 1

Syntax untuk memfilter data

```
# df_filtered = df[(kondisi1) & (kondisi2) | (kondisi3) ]
# kondisi = df[df['is_valid']==1]

df_filtered = df[(df['is_valid']==1) &\
                  (df['category']=='Mobiles & Tablets') &\
                  ((df['order_date'] >= '2022-01-01') & (df['order_date'] <= '2022-12-31'))]

df_filtered.head()
```

Syntax untuk menjawab kasus 1

```
answer_1 = df_filtered.groupby(by=["sku_name"])[["qty_ordered"]].sum()\
              .sort_values(ascending=False)\
              .head(5)\
              .reset_index(name='qty_2022')

answer_1
```

Hasil dari Kasus 1

	sku_name	qty_2022
0	IDROID_BALRX7-Gold	1000
1	IDROID_BALRX7-Jet black	31
2	Infinix Hot 4-Gold	15
3	samsung_Grand Prime Plus-Black	11
4	infinix_Zero 4-Grey	10

Berdasarkan data yang diperoleh, produk **"IDROID_BALRX7-Gold"** menunjukkan potensi besar sebagai hadiah utama dalam Kompetisi Festival Akhir Tahun. Hadiah ini diharapkan dapat menarik minat peserta dan meningkatkan keterlibatan dengan merek.

Kasus 2

Dear Data Analyst,

Menindaklanjuti meeting gabungan Tim Warehouse 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 Warehouse

Jawaban Kasus 2.1

Memfilter data

```
df_2021 = df[(df['is_valid']==1) &\
              ((df['order_date'] >= '2021-01-01') & (df['order_date'] <= '2021-12-31'))]\
              .groupby(by=["category"])[ "qty_ordered" ].sum()\
              .sort_values(ascending=False)\
              .reset_index(name='qty_2021')
```

df_2021

```
df_2022 = df[(df['is_valid']==1) &\
              ((df['order_date'] >= '2022-01-01') & (df['order_date'] <= '2022-12-31'))]\
              .groupby(by=["category"])[ "qty_ordered" ].sum()\
              .sort_values(ascending=False)\
              .reset_index(name='qty_2022')
```

df_2022

Jawaban Kasus 2.1

Menggabungkan data

```
#menggabungkan data
answer_2_1 = df_2021.merge(df_2022, left_on = 'category', right_on = 'category') #setara inner join
answer_2_1.head()
```

Mendefinisikan kolom perubahan penjualan

```
#define kolom perubahan penjualan
answer_2_1['changes'] = answer_2_1['qty_2022'] - answer_2_1['qty_2021']
answer_2_1.sort_values(by=['changes'],ascending=True,inplace=True)

answer_2_1
```

Hasil dari Kasus 2.1

	category	qty_2021	qty_2022	changes
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

Berdasarkan analisis data, dapat disimpulkan bahwa **terjadi ketidakseimbangan antara penawaran dan permintaan dalam kategori "Others"**. Untuk mengatasi masalah ini, perlu dilakukan analisis lebih mendalam serta langkah-langkah yang tepat untuk memperbaiki strategi penjualan dan pengelolaan inventaris. Selain itu, perusahaan juga harus memanfaatkan peluang yang ada di kategori yang menunjukkan pertumbuhan tinggi.

Jawaban Kasus 2.2

Memfilter data

```
df_others_2021 = df[(df['is_valid']==1) &\
                    (df['category']=='Others') &\
                    ((df['order_date'] >= '2021-01-01') & (df['order_date'] <= '2021-12-31'))]\
                    .groupby(by=["sku_name"])[["qty_ordered"]].sum()\
                    .sort_values(ascending=False)\
                    .reset_index(name='qty_2021')

df_others_2021.head()
```

```
df_others_2022 = df[(df['is_valid']==1) &\
                    (df['category']=='Others') &\
                    ((df['order_date'] >= '2022-01-01') & (df['order_date'] <= '2022-12-31'))]\
                    .groupby(by=["sku_name"])[["qty_ordered"]].sum()\
                    .sort_values(ascending=False)\
                    .reset_index(name='qty_2022')

df_others_2022.head()
```

Jawaban Kasus 2.2

Menggabungkan data

```
#Menggabungkan data
answer_2_2 = df_others_2021.merge(df_others_2022, how='outer') #sama kaya full outer join
answer_2_2.head()
```

Menampilkan 20 produk teratas yang mengalami penurunan terbesar.

```
answer_2_2.fillna(0, inplace=True)
answer_2_2['changes'] = answer_2_2['qty_2022'] - answer_2_2['qty_2021']
answer_2_2.sort_values(by=['changes'], ascending=True, inplace=True)

answer_2_2.head(20)
```


Hasil dari Kasus 2.2

	sku_name	qty_2021	qty_2022	changes
0	RB_Dettol Germ Busting Kit-bf	200.0	45.0	-155.0
2	Dawlance_MD 10 + DWB 600	23.0	0.0	-23.0
1	Telemall_MM-DR-HB-L	23.0	2.0	-21.0
3	iu_Tickets General Enclosure-Islamabad	20.0	0.0	-20.0
4	RS_Rehmat-e-Shereen Mix Mithai	13.0	0.0	-13.0
5	kansai_NeverWet	10.0	1.0	-9.0
6	sindbad_Sindbad Gift Card-3	7.0	0.0	-7.0
7	emart_00-1	7.0	1.0	-6.0
8	Vouch 365 2016	5.0	0.0	-5.0
10	Am-PTV_ATS-004-M	5.0	0.0	-5.0
11	The Vitamin Company Kojic Acid Whitening Cream...	4.0	0.0	-4.0
12	duma_4561253300294	4.0	0.0	-4.0
14	sockoye_QG in Quarter Grey	4.0	0.0	-4.0
17	Trans2_LW 999	3.0	0.0	-3.0
18	aw_Octane Booster-12oz./354ml	3.0	0.0	-3.0
13	MEGUIAR_G12711	4.0	1.0	-3.0
19	aw_Super Concentrated Fuel Injector (Black)-5....	2.0	0.0	-2.0
20	MEGUIAR_X1030EU	2.0	0.0	-2.0
21	MEGUIAR_G19216	2.0	0.0	-2.0
22	seagate_SG E-2TB	2.0	0.0	-2.0

Insight

- Penurunan kuantitas penjualan produk kategori **"Others"** signifikan di tahun 2022 dibandingkan 2021.
- Produk dengan penurunan terbesar termasuk **"RB_Dettol Germ Busting Kit-bf"** dan **"Dawlance_MD 10 + DWB 600"**.
- Penurunan ini menunjukkan potensi penumpukan stok pada akhir tahun 2022.

Rekomendasi:

- Lakukan promosi khusus untuk produk dengan penurunan penjualan.
- Evaluasi dan sesuaikan harga jual produk.
- Gabungkan produk kurang diminati dengan produk populer.
- Ubah posisi produk di platform penjualan online.

Kasus 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 butuhkan adalah 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

Jawaban Kasus 3

Memfilter data

```
df_promo = df[(df['is_gross']==1) & (df['is_valid']==0) & (df['is_net']==0) & ((df['order_date'] >= '2022-01-01') & (df['order_date'] <= '2022-12-31'))]
answer_3 = df_promo[['customer_id', 'registered_date']]

answer_3.sort_values(by=['customer_id'])
```

Menampilkan data pelanggan yang telah melakukan checkout tetapi belum menyelesaikan pembayaran

```
unique_answer_3 = answer_3.drop_duplicates(subset=['customer_id'])
unique_answer_3
```

Mengunduh hasil data

```
#Jalankan kode ini untuk mendownload file
from google.colab import files
unique_answer_3.to_csv('audience_list.csv', encoding = 'utf-8-sig', index=False)
files.download('audience_list.csv')
```

Hasil dari Kasus 3

	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
...
5855	C653797L	2022-04-03
5856	C394076L	2021-10-12
5859	C248585L	2022-07-10
5865	C471304L	2022-05-13
5881	C265450L	2022-02-17

820 rows × 2 columns

Berdasarkan data yang telah didapatkan, terdapat **820 pelanggan yang memenuhi syarat** untuk menerima informasi mengenai promosi ulang tahun perusahaan.

Kasus 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 diberikan, kami mengucapkan terima kasih.

Regards

Tim Campaign

Jawaban Kasus 4.1

Membuat kolom tambahan untuk day, month, dan month_num

```
#Membuat kolom tambahan untuk day, month, month_num
df['day']=df['order_date'].dt.day_name()
df['month']=df['order_date'].dt.month_name()
df['month_num']=df['order_date'].dt.month
df.head(5)
```

Memfilter data untuk penjualan weekend

```
daily_weekend = df[(df['is_valid']==1) &\
                    (df['day'].isin(['Saturday','Sunday'])) &\
                    (df['order_date'] >= '2022-10-01') & (df['order_date'] <= '2022-12-31')]\
                    .groupby(by=["month_num","month", "order_date"])["before_discount"].sum()\
                    .round()\
                    .reset_index(name='daily_sales_weekend')

daily_weekend
```

Menghitung rata-rata penjualan selama weekend

```
df_weekend = daily_weekend.groupby(by=["month_num", "month"])["daily_sales_weekend"].mean().round().reset_index(name='avg_sales_weekend')
df_weekend
```

Jawaban Kasus 4.1

Memfilter data untuk penjualan weekday

```
daily_weekday = df[(df['is_valid']==1) &\
    (df['day'].isin(['Monday', 'Tuesday', 'Wednesday', 'Thursday', 'Friday'])) &\
    (df['order_date'] >= '2022-10-01') & (df['order_date'] <= '2022-12-31')]\
    .groupby(by=["month_num", "month", "order_date"])["before_discount"].sum()\
    .round()\
    .reset_index(name='daily_sales_weekday')
```

```
daily_weekday
```

Menghitung rata-rata penjualan selama weekday

```
df_weekday = daily_weekday.groupby(by=["month_num", "month"])["daily_sales_weekday"].mean().round().reset_index(name='avg_sales_weekday')
df_weekday
```

Jawaban Kasus 4.1

Menampilkan rata-rata penjualan serta perbedaan dalam penjualannya.

```
answer_4_1 = df_weekend.merge(df_weekday, left_on = 'month', right_on = 'month')
answer_4_1 = answer_4_1[['month', 'avg_sales_weekend', 'avg_sales_weekday']]
answer_4_1['diff'] = answer_4_1['avg_sales_weekend'] - answer_4_1['avg_sales_weekday']
answer_4_1['diff_percent'] = (answer_4_1['diff']/answer_4_1['avg_sales_weekend'])*100

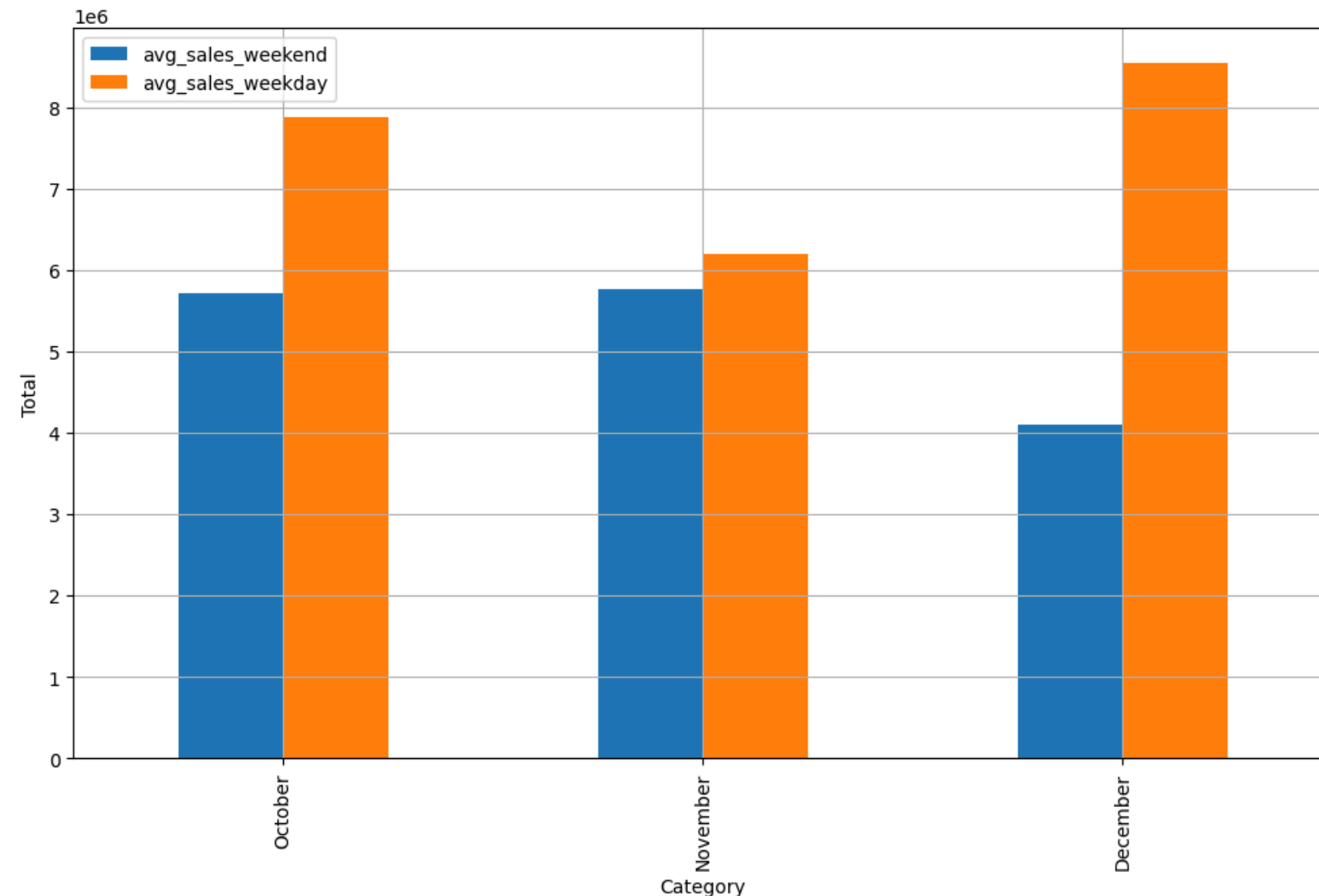
answer_4_1
```

Menampilkan grafik rata-rata penjualan

```
answer_4_1.plot(x='month',
                y=['avg_sales_weekend', 'avg_sales_weekday'],
                kind='bar',
                grid = True,
                xlabel = 'Category',
                ylabel = 'Total',
                figsize=(12,7),
                rot = 90,
                table = False,
                secondary_y = False)
```


Hasil dari Kasus 4.1

	month	avg_sales_weekend	avg_sales_weekday	diff	diff_percent
0	October	5708341.0	7872212.0	-2163871.0	-37.907178
1	November	5774045.0	6204666.0	-430621.0	-7.457874
2	December	4105994.0	8542530.0	-4436536.0	-108.050231



Berdasarkan data yang ada, **kampanye yang dilakukan pada weekend (Sabtu dan Minggu) selama bulan Oktober hingga Desember 2022 tidak memberikan dampak positif pada peningkatan penjualan.** Justru, data menunjukkan adanya penurunan penjualan yang signifikan pada akhir pekan dibandingkan dengan hari kerja.

Rekomendasi

- Evaluasi strategi kampanye, produk, harga promo, dan saluran distribusi.
- Analisis mendalam data penjualan, bandingkan kinerja di segmen pelanggan berbeda.
- Uji coba strategi baru: produk bundling, diskon tambahan, event akhir pekan.
- Pertimbangkan faktor eksternal yang mempengaruhi penjualan dan sesuaikan strategi kampanye.

Jawaban Kasus 4.2

Menampilkan rata-rata penjualan weekend dan weekday

```
answer_4_2 = {\n    'Periode': 'Total 3 months',\n    'Avg Weekend Sales': round(daily_weekend['daily_sales_weekend'].mean(),2), \n    'Avg Weekdays Sales': round(daily_weekday['daily_sales_weekday'].mean(),2),\n    'Diff (Value)': round(daily_weekend['daily_sales_weekend'].mean() - daily_weekday['daily_sales_weekday'].mean(),2),\n    'Diff (%)': pd.Series(round(((daily_weekend['daily_sales_weekend'].mean() - daily_weekday['daily_sales_weekday'].mean())/daily_weekend['daily_sales_weekend'].mean()*100,2), dtype=str)+'%\n}\n\npd.DataFrame(data=answer_4_2, index=[0])
```

Hasil dari Kasus 4.2

	Periode	Avg Weekend Sales	Avg Weekdays Sales	Diff (Value)	Diff (%)
0	Total 3 months	5269300.0	7520249.29	-2250949.29	-42.72%

Dari data yang diperoleh, dapat disimpulkan bahwa:

- Penjualan Hari Kerja Lebih Tinggi: Rata-rata penjualan harian pada hari kerja jauh lebih tinggi dibandingkan dengan penjualan harian pada akhir pekan.
- Penurunan Signifikan: Terdapat penurunan penjualan yang cukup signifikan pada akhir pekan, yaitu sebesar 42,72% jika dibandingkan dengan hari kerja.

Rekomendasi Selanjutnya

- Evaluasi kembali strategi kampanye yang telah dilakukan.
- Lakukan analisis data penjualan lebih mendalam.
- Uji coba strategi kampanye baru.
- Pertimbangkan faktor eksternal yang mempengaruhi penjualan.

TERIMA
KASIH