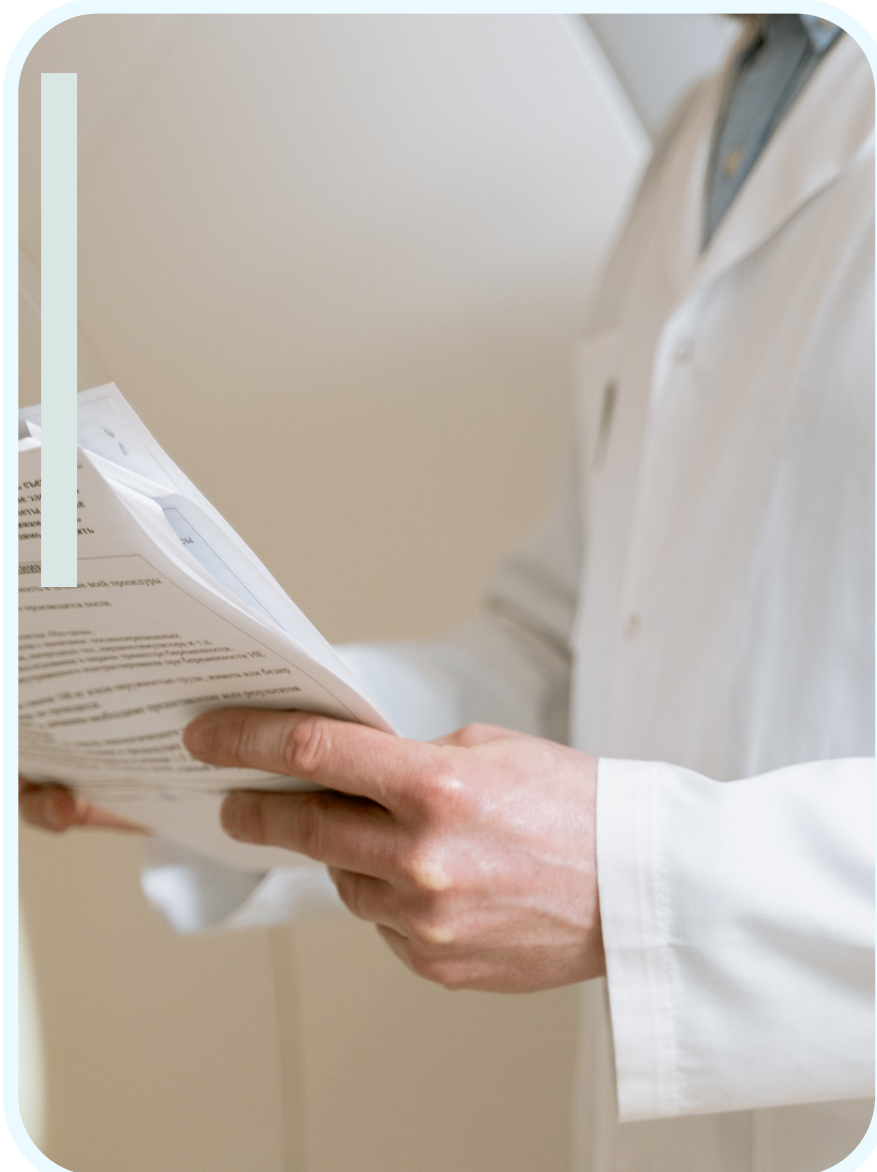




EXERCISE PYTHON

Rina Raihana Zakiya

Pertanyaan



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

1

INPUT

```
#Menyimpan data dalam bentuk Pandas DataFrame
data1 = pd.DataFrame(\
    #Memfilter data dengan valid=1
    df[(df['is_valid']==1) &\
        #Memfilter data dengan Category = Mobiles & Tablets
        (df['category']=='Mobiles & Tablets') &\
        #Memfilter data pada transaksi selama 2022
        ((df['order_date'] >= '2022-01-01') & (df['order_date'] <= '2022-12-31'))]\
    #Mengelompokkan data berdasarkan nama produk
    .groupby(by=["sku_name"])[["qty_ordered"]]\
    #Pengelompokan berdasarkan penjumlahan
    .sum()\
    #Mengurutkan data berdasarkan nilai kuantitas paling banyak
    .sort_values(ascending=False)\
    #Mengambil TOP 5 Produk
    .head(5)\
    #Merest nama header
    .reset_index(name='qty_2022'))
```

data1

OUTPUT

	sku_name	qty_2022
0	IDROID_BALRX7-Gold	1632
1	IDROID_BALRX7-Jet black	1014
2	IDROID_BALRX7-Jet black	542
3	Mobo_ar1	524
4	Infinix Hot 4-Black	396

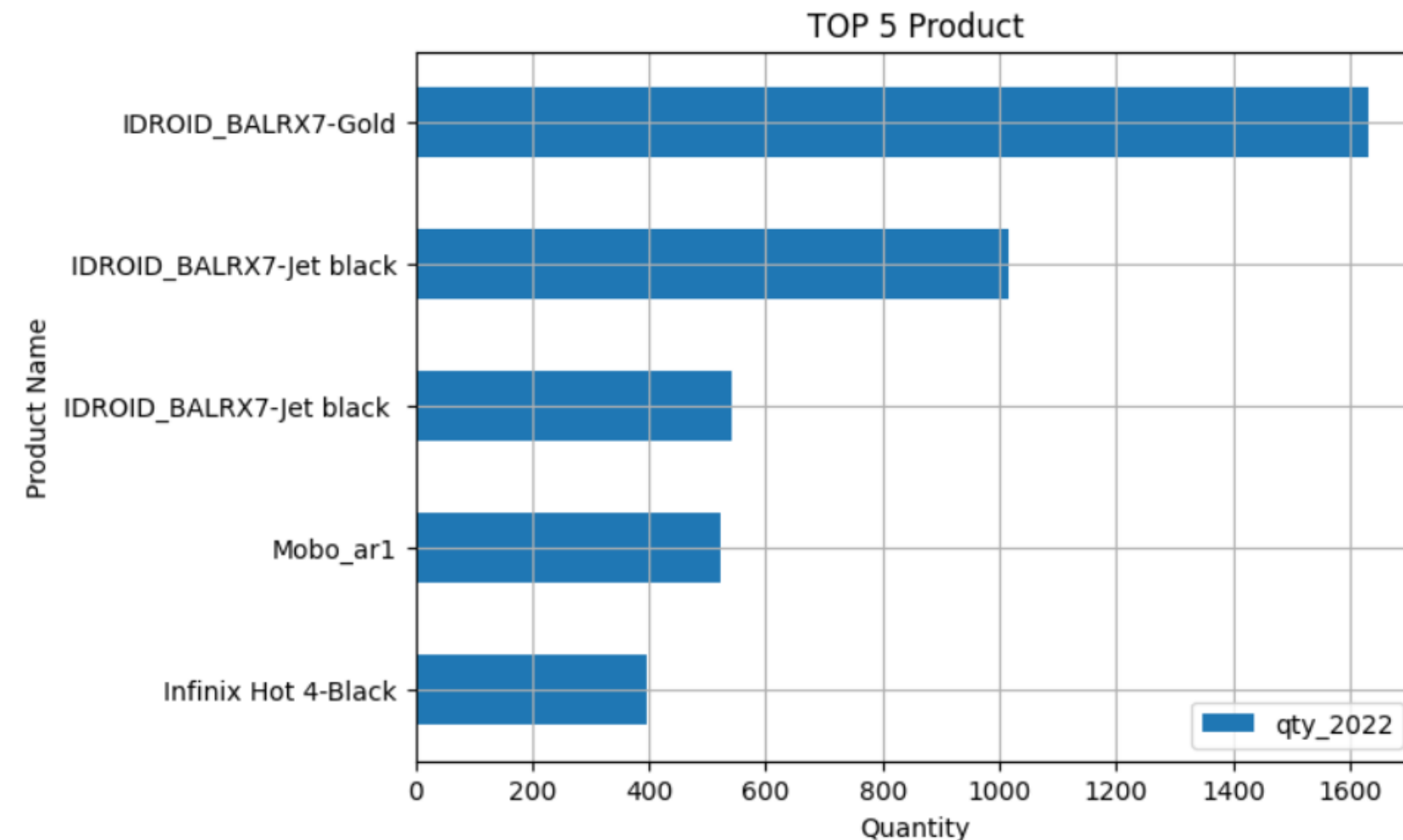
INPUT

```
#Sort data untuk dapat digunakan sebagai grafik
data1.sort_values(\
    by=['qty_2022'], \
    ascending=True,\
    inplace=True)

#Membuat grafik
data1.plot(x='sku_name',
           y=['qty_2022'],
           kind='barh',
           grid = True,
           xlabel = 'Quantity',
           ylabel = 'Product Name',
           figsize=(7,5),
           rot = 0,
           title = 'TOP 5 Product',
           table = False,
           sort_columns = False,
           secondary_y = False)
```

OUTPUT

```
<ipython-input-17-be0a6630da14>:8: FutureWarning: `sort_columns` is deprecated and will be removed in a future version.
data1.plot(x='sku_name',
<Axes: title={'center': 'TOP 5 Product'}, xlabel='Quantity', ylabel='Product Name'>
```



Kesimpulan :

TOP 5 Produk dari Kategori Mobiles & Tablets selama tahun 2022 dengan jumlah kuantitas penjualan paling tinggi secara berurutan adalah IDROID_BALRX7-Gold, IDROID_BALRX7-jet black, IDROID_BALLRX7-Jet black, Mobo_ar1 dan Infinix Hot 4-Black. Oleh karena itu, Tim Marketing dapat memberikan ke-5 produk tersebut sebagai hadiah kepada pemenang kompetisi Festival Akhir Tahun.

Pertanyaan

Dear Data Analyst,

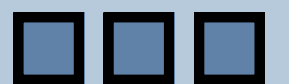
Menindaklanjuti meeting gabungan Tim Warehouse dan Tim Marketing, kami menemukan bahwa ketersediaan stock produk dengan Kategori Beauty & Grooming pada akhir 2022 kemarin masih banyak.

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 Beauty & Grooming, 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

2

INPUT

```
#Menyimpan data dalam bentuk Pandas DataFrame
data2 = pd.DataFrame(\
    #Memfilter data dengan valid=1
    df[(df['is_valid']==1) &\
        #Memfilter data pada transaksi selama 2021
        ((df['order_date'] >= '2021-01-01') & (df['order_date'] <= '2021-12-31'))]\
    #Mengelompokkan data berdasarkan nama produk
    .groupby(by=["category"])[["qty_ordered"]]\
    #Pengelompokan berdasarkan penjumlahan
    .sum()\
    #Mengurutkan data berdasarkan nilai kuantitas paling banyak
    .sort_values(ascending=False)\
    #Mereset nama header
    .reset_index(name='qty_2021'))

data2
```

OUTPUT

	category	qty_2021
0	Men Fashion	22304
1	Soghaat	16523
2	Mobiles & Tablets	14431
3	Beauty & Grooming	11987
4	Women Fashion	7240
5	Appliances	5947
6	Superstore	5274
7	Home & Living	4408
8	Others	4017
9	Health & Sports	3607
10	Kids & Baby	2971
11	Computing	2418
12	Entertainment	1502
13	School & Education	1101
14	Books	454

INPUT

```
#Menyimpan data dalam bentuk Pandas DataFrame
data3 = pd.DataFrame(\
    #Memfilter data dengan valid=1
    df[(df['is_valid']==1) &\
        #Memfilter data pada transaksi selama 2022
        ((df['order_date'] >= '2022-01-01') & (df['order_date'] <= '2022-12-31'))]\
    #Mengelompokkan data berdasarkan nama produk
    .groupby(by=["category"])[["qty_ordered"]]\
    #Pengelompokan berdasarkan penjumlahan
    .sum()\
    #Mengurutkan data berdasarkan nilai kuantitas paling banyak
    .sort_values(ascending=False)\
    #Mereset nama header
    .reset_index(name='qty_2022'))

data3
```

OUTPUT

	category	qty_2022
0	Superstore	27905
1	Men Fashion	21904
2	Mobiles & Tablets	20890
3	Women Fashion	17325
4	Soghaat	11491
5	Beauty & Grooming	8005
6	Home & Living	5986
7	Appliances	5651
8	Kids & Baby	5437
9	Health & Sports	4039
10	Others	2594
11	Entertainment	2336
12	Computing	2251
13	School & Education	1165
14	Books	691

INPUT

```
data4 = data2.merge(data3, left_on = 'category', right_on = 'category')
data4
```

INPUT

```
#Growth = tahun terakhir - tahun sebelumnya
data4['qty_growth']=data4['qty_2022']-data4['qty_2021']
data4.sort_values(by=['qty_growth'],ascending=True,inplace=True)
data4
```

Kesimpulan :

Dari 15 kategori, 6 kategori mengalami penurunan kuantitas penjualan pada tahun 2022 dibandingkan dengan tahun 2021. Kategori yang mengalami penurunan diantaranya adalah Soghaat, Beauty & Grooming, Others, Men Fashion , Appliances dan Computing. Penurunan kuantitas terbesar adalah kategori Soghaat dengan angka penurunan mencapai 5.032 kuantitas dari tahun 2021. Sebaliknya kategori Superstore mengalami peningkatan kuantitas penjualan terbesar yaitu mencapai 22.631 kuantitas dari tahun 2021.

OUTPUT

	category	qty_2021	qty_2022	qty_growth
1	Soghaat	16523	11491	-5032
3	Beauty & Grooming	11987	8005	-3982
8	Others	4017	2594	-1423
0	Men Fashion	22304	21904	-400
5	Appliances	5947	5651	-296
11	Computing	2418	2251	-167
13	School & Education	1101	1165	64
14	Books	454	691	237
9	Health & Sports	3607	4039	432
12	Entertainment	1502	2336	834
7	Home & Living	4408	5986	1578
10	Kids & Baby	2971	5437	2466
2	Mobiles & Tablets	14431	20890	6459
4	Women Fashion	7240	17325	10085
6	Superstore	5274	27905	22631

3

Jawaban

INPUT

```
#Menyimpan data dalam bentuk Pandas DataFrame
data5 = pd.DataFrame(\
    #Memfilter data dengan valid=1
    df[(df['is_valid']==1) &\
        (df['category']=='Beauty & Grooming') &\
        #Memfilter data pada transaksi selama 2021
        ((df['order_date'] >= '2021-01-01') & (df['order_date'] <= '2021-12-31'))]\
    #Mengelompokkan data berdasarkan nama produk
    .groupby(by=["sku_name"])[["qty_ordered"]]\
    #Pengelompokan berdasarkan penjumlahan
    .sum()\
    #Mengurutkan data berdasarkan nilai kuantitas paling banyak
    .sort_values(ascending=False)\
    #Mereset nama header
    .reset_index(name='qty_bg_2021'))
```

data5

OUTPUT

	sku_name	qty_bg_2021
0	kcc_krone deal	1583
1	kcc_glamour deal	535
2	kcc_Buy 2 Frey Air Freshener & Get 1 Kasual Bo...	268
3	emart_Nova Straightener + Curler	245
4	kcc_blaze	239
...
1304	TO_066-2 IN 1 MAKE UP BRUSH	1
1305	TO_073-The Balm Cosmetics Mary-Lou Manizer	1
1306	TO_080-BUNDLE DEAL CONTOUR & BLENDER	1
1307	TO_083-PACK OF 2 ADS EYELINER GEL	1
1308	Saloni_VINCE4	1

1309 rows × 2 columns

INPUT

```
#Menyimpan data dalam bentuk Pandas DataFrame
data6 = pd.DataFrame(\
    #Memfilter data dengan valid=1
    df[(df['is_valid']==1) &\
        (df['category']=='Beauty & Grooming') &\
        #Memfilter data pada transaksi selama 2021
        ((df['order_date'] >= '2022-01-01') & (df['order_date'] <= '2022-12-31'))]\
    #Mengelompokkan data berdasarkan nama produk
    .groupby(by=["sku_name"])[["qty_ordered"]]\
    #Pengelompokan berdasarkan penjumlahan
    .sum()\
    #Mengurutkan data berdasarkan nilai kuantitas paling banyak
    .sort_values(ascending=False)\
    #Mereset nama header
    .reset_index(name='qty_bg_2022'))
```

data6

OUTPUT

	sku_name	qty_bg_2022
0	kcc_krone deal	448
1	kcc_glamour deal	175
2	un_Beauty Care Bundle_BC	103
3	BO_All-In-One Manicure Set	100
4	nabila_Nchant FOR Olive	90
...
1807	SFEVER_CH369HB0EHEHSNAFAMZ	1
1808	SFEVER_CH813HB0FO4AWNAFAMZ	1
1809	SFEVER_CH813HB0G7EO4NAFAMZ	1
1810	SFEVER_CH813HB0U0NZCNAFAMZ	1
1811	zaffiro_Z-P-069	1

1812 rows × 2 columns

INPUT

```
#Menggabungkan data
data7 = data5.merge(data6, left_on = 'sku_name', right_on = 'sku_name')
data7['qty_bg_growth']=data7['qty_bg_2022']-data7['qty_bg_2021']
data7.sort_values(by=['qty_bg_growth'],ascending=True,inplace=True)
data7 = data7.head(20)
data7
```

Kesimpulan :

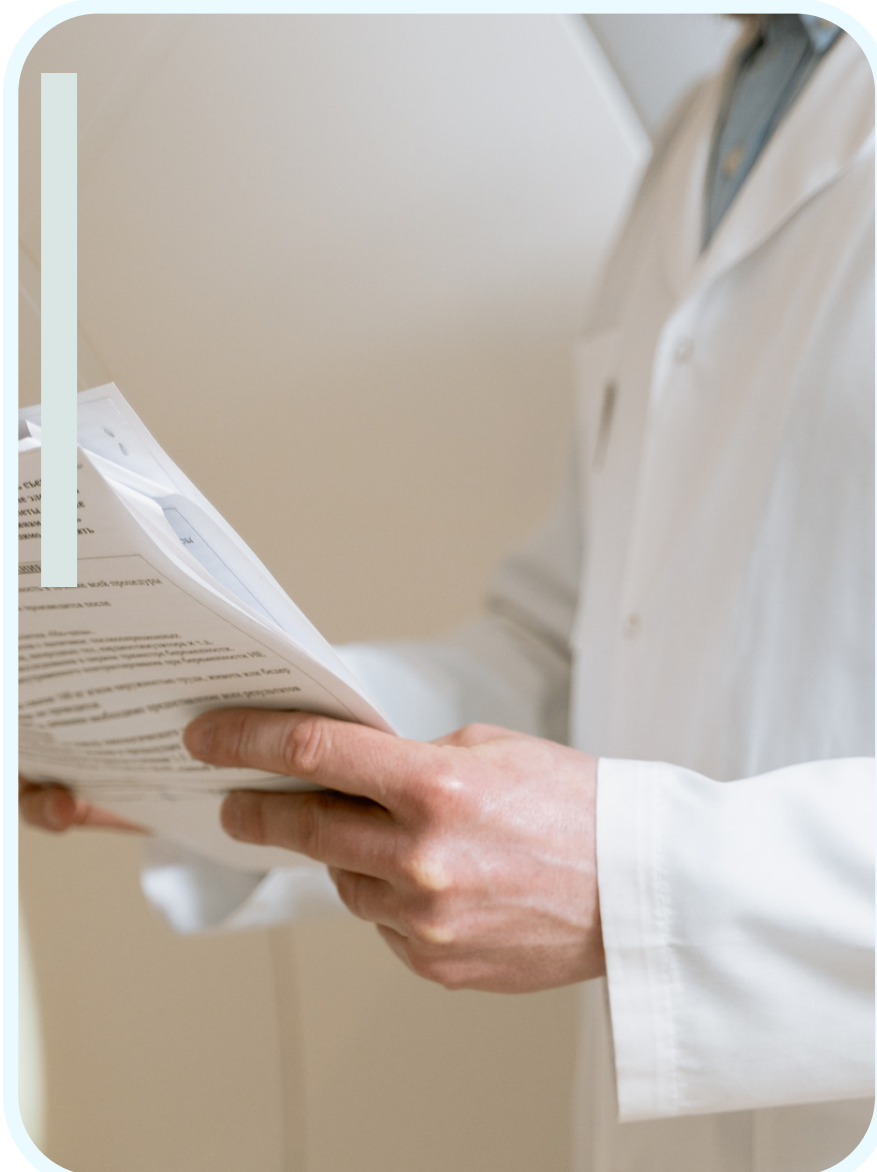
TOP 20 nama produk yang mengalami penurunan paling tinggi pada 2022 jika dibanding dengan 2021 diantaranya adalah kcc_krone deal, kcc_glamour deal, kcc_blaze, emart_Ding Ling Trimmer rf 608, emart_Nova Straightener + Curler dll.

Produk yang mengalami penurunan paling tinggi adalah kcc_krone deal yaitu berkurang 1.135 kuantitas dari tahun 2021. Sedangkan di posisi kedua yaitu kcc_glamour deal dengan angka penurunan sebesar 360 kuantitas. Di posisi ke-20 yaitu dengan angka penurunan sebesar 85 kuantitas dimiliki oleh produk Bold_Bold-REVIVE.

OUTPUT

	sku_name	qty_bg_2021	qty_bg_2022	qty_bg_growth
0	kcc_krone deal	1583	448	-1135
1	kcc_glamour deal	535	175	-360
2	kcc_Buy 2 Frey Air Freshener & Get 1 Kasual Bo...	268	2	-266
4	kcc_blaze	239	4	-235
5	emart_Ding Ling Trimmer rf 608	214	2	-212
3	emart_Nova Straightener + Curler	245	35	-210
7	kcc_jazzy	208	47	-161
6	itter_AB 1211	213	77	-136
9	Bold_Fresh	135	9	-126
8	Bold_Active	135	11	-124
15	kcc_effect	109	8	-101
16	kcc_force	106	6	-100
10	Bold_Noir	127	27	-100
11	Bold_Spice	116	18	-98
14	emart_Mozar Hair Dryer	113	18	-95
13	kcc_Bold Pocket Perfume	114	28	-86
25	Bold_Classic	89	3	-86
19	Bold_Bold-AQUA	100	15	-85
24	Bold_Bold-EPIC	90	5	-85

Pertanyaan



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

4

Jawaban

INPUT

```
#Menyimpan data dalam bentuk Pandas DataFrame
data8 = df[\
    #Memfilter data dengan gross=1
    (df['is_gross']==1) &\
    #Memfilter data dengan valid=0
    (df['is_valid']==0) &\
    #Memfilter data dengan net=0
    (df['is_net']==0) &\
    #Memfilter data pada transaksi selama 2022
    ((df['order_date'] >= '2022-01-01') & (df['order_date'] <= '2022-12-31'))]

data9 = data8[['customer_id','registered_date']]
data9
```

INPUT

```
from google.colab import files
data9.to_csv('audience_list.csv', encoding = 'utf-8-sig',index=False)
files.download('audience_list.csv')
```



audience_list.csv

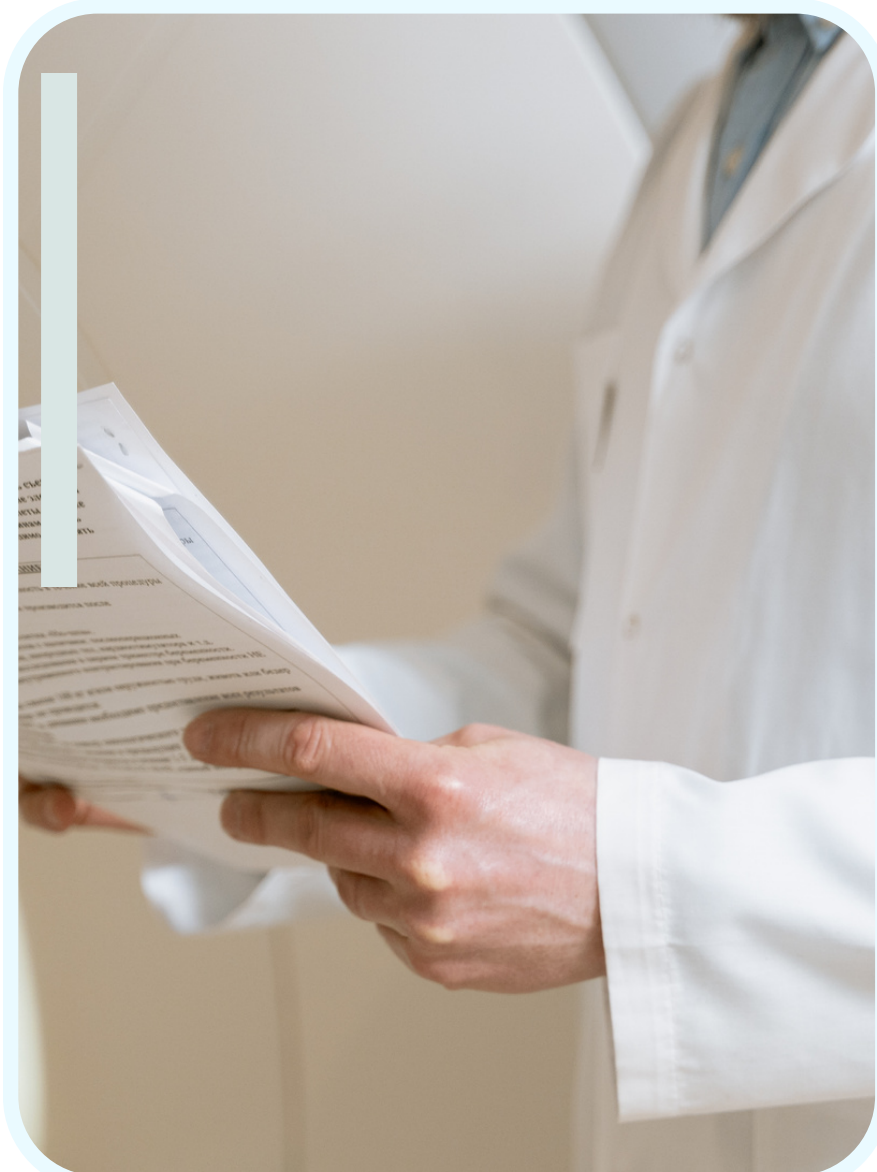
Completed — 906 KB

Berikut data ID Customer dan Registered Date customer yang sudah melakukan check-out namun belum melakukan pembayaran selama tahun 2022.

OUTPUT

	customer_id	registered_date
276	C996508L	2021-07-10
279	C996508L	2021-07-10
280	C996508L	2021-07-10
286	C996508L	2021-07-10
294	C996508L	2021-07-10
...
287243	C125653L	2022-08-09
287245	C173391L	2022-08-26
287246	C173393L	2022-08-03
287247	C173393L	2022-08-03
287248	C173395L	2022-08-19
46385 rows × 2 columns		

Pertanyaan



Dear Data Analyst,

Bulan depan kami akan membuat Laporan Tahunan yang akan kami sampaikan ke Investor, terkait hal tersebut mohon sediakan data berikut:

1. Profit Growth (%) secara keseluruhan tahun 2021 vs 2022 sebagai hasil kinerja tahunan penjualan.
2. Profit Growth (%) berdasarkan Kategori produk tahun 2021 vs 2022.

Mohon bantuan untuk mengirimkan data tersebut paling lambat besok. Atas bantuan yang diberikan, kami mengucapkan terima kasih.

Regards
Board of Director (BOD)



Jawaban

INPUT

```
#Menghitung profit
df['profit'] = df['after_discount'] - df['cogs']

#Menyimpan data dalam bentuk Pandas DataFrame
data10 = df[\
    #Memfilter data dengan valid=1
    (df['is_valid']==1) &\
    #Memfilter data pada transaksi selama 2022
    ((df['order_date'] >= '2022-01-01') & (df['order_date'] <= '2022-12-31'))]

#Menyimpan data dalam bentuk Pandas DataFrame
data11 = df[\
    #Memfilter data dengan valid=1
    (df['is_valid']==1) &\
    #Memfilter data pada transaksi selama 2021
    ((df['order_date'] >= '2021-01-01') & (df['order_date'] <= '2021-12-31'))]

#Membuat Dataframe dan Summary
data12 = {\
    'Periode Profit':'Total',\
    '2021': data11['profit'].sum(), \
    '2022': data10['profit'].sum(),\
    'Growth (Value)': data10['profit'].sum() - data11['profit'].sum(),\
    'Growth': pd.Series(round(((data10['profit'].sum() - data11['profit'].sum())/data11['profit'].sum()*100,2), dtype=str)+'\
}

pd.DataFrame(data=data12, index=[0])
```

OUTPUT					
Periode	Profit	2021	2022	Growth (Value)	Growth
0	Total	3044341855	5153155179	2108813324	69.27%

Kesimpulan :

Hasil kinerja tahunan penjualan yang dilihat dari Profit Growth (%) secara keseluruhan tahun 2021 vs 2022 adalah sebesar 69,27% dengan Growth (value) sebesar 210881.

6

Jawaban

INPUT

```
#Menggunakan Groupby untuk menjumlahkan
data13 = pd.DataFrame(data10\
                        .groupby(by="category")["profit"].sum()\
                        .sort_values(ascending=False)\
                        .reset_index(name='profit_2022'))

data13

data14 = pd.DataFrame(data11\
                        .groupby(by="category")["profit"].sum()\
                        .sort_values(ascending=False)\
                        .reset_index(name='profit_2021'))

data14

#Menggabungkan data
data15 = data14.merge(data13, left_on = 'category', right_on = 'category')
data15
```

OUTPUT

	category	profit_2021	profit_2022
0	Mobiles & Tablets	1570556101	2972334731
1	Appliances	388136061	497790138
2	Men Fashion	245341028	234632719
3	Entertainment	200529969	316810200
4	Women Fashion	145004578	408306607
5	Computing	89620746	133779051
6	Beauty & Grooming	83353396	94244754
7	Home & Living	73542298	84116394
8	Soghaat	72174592	62756359
9	Superstore	50548595	225547431
10	Health & Sports	48849715	55628929
11	Others	32240675	22778462
12	Kids & Baby	26996299	30286302
13	School & Education	9134125	10049300
14	Books	8313677	4093802

INPUT

```
#Melakukan kalkulasi
data15['Growth (Value)'] = data15['profit_2022']-data15['profit_2021']
data15['Growth (%)'] = round(data15['Growth (Value)']/data15['profit_2021']*100,2)
data15.sort_values(by=['Growth (%)'], ascending = False, inplace = True)
data15
```

OUTPUT

	category	profit_2021	profit_2022	Growth (Value)	Growth (%)
9	Superstore	50548595	225547431	174998836	346.20
4	Women Fashion	145004578	408306607	263302029	181.58
0	Mobiles & Tablets	1570556101	2972334731	1401778630	89.25
3	Entertainment	200529969	316810200	116280231	57.99
5	Computing	89620746	133779051	44158305	49.27
1	Appliances	388136061	497790138	109654077	28.25
7	Home & Living	73542298	84116394	10574096	14.38
10	Health & Sports	48849715	55628929	6779214	13.88
6	Beauty & Grooming	83353396	94244754	10891358	13.07
12	Kids & Baby	26996299	30286302	3290003	12.19
13	School & Education	9134125	10049300	915175	10.02
2	Men Fashion	245341028	234632719	-10708309	-4.36
8	Soghaat	72174592	62756359	-9418233	-13.05
11	Others	32240675	22778462	-9462213	-29.35
14	Books	8313677	4093802	-4219875	-50.76

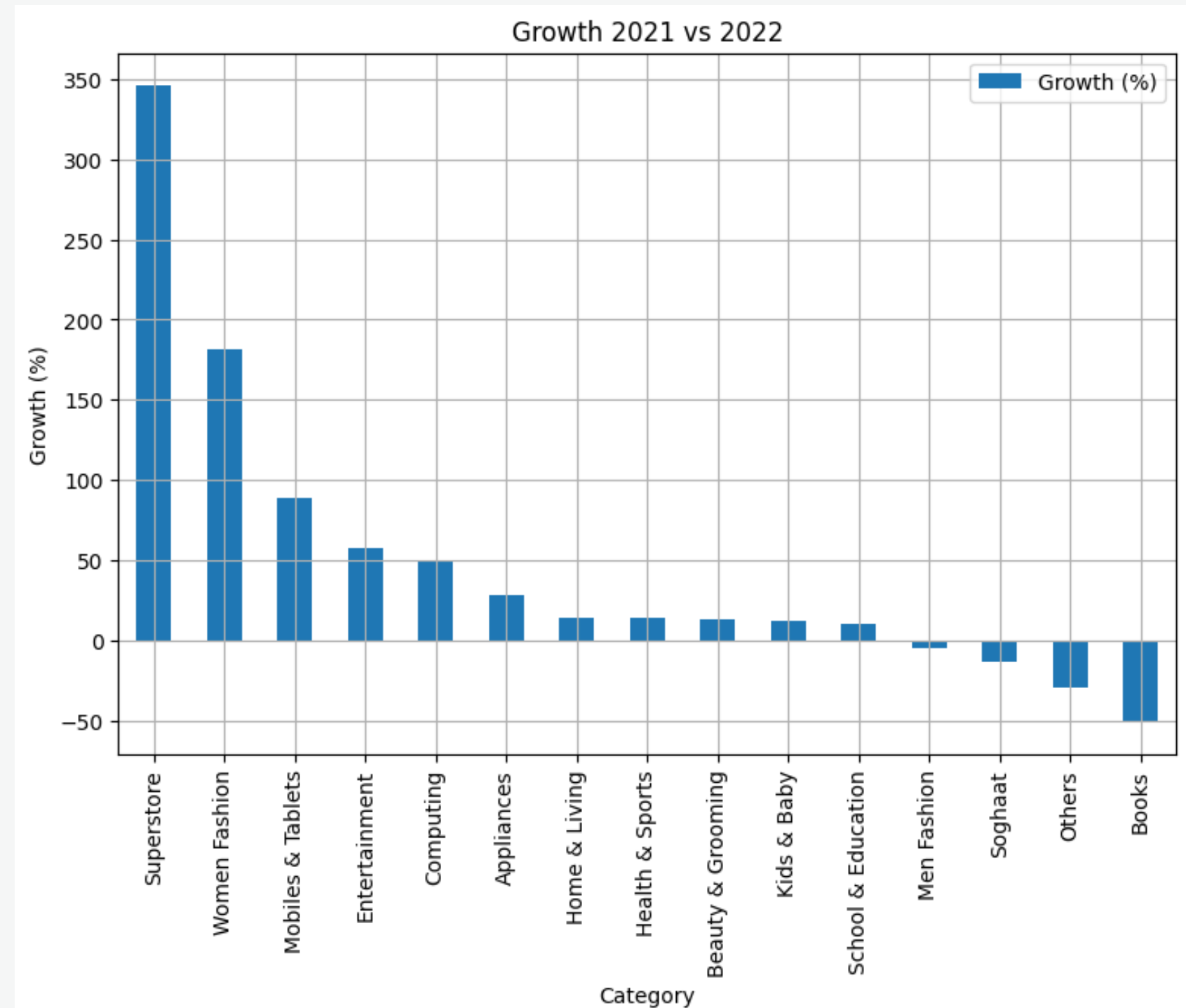
INPUT

```
#Membuat grafik
data15.plot(x='category',
            y=['Growth (%)'],
            kind='bar',
            grid = True,
            title = 'Growth 2021 vs 2022',
            xlabel = 'Category',
            ylabel = 'Growth (%)',
            figsize=(9,6),
            rot = 90,
            table = False,
            sort_columns = False,
            secondary_y = False)
```

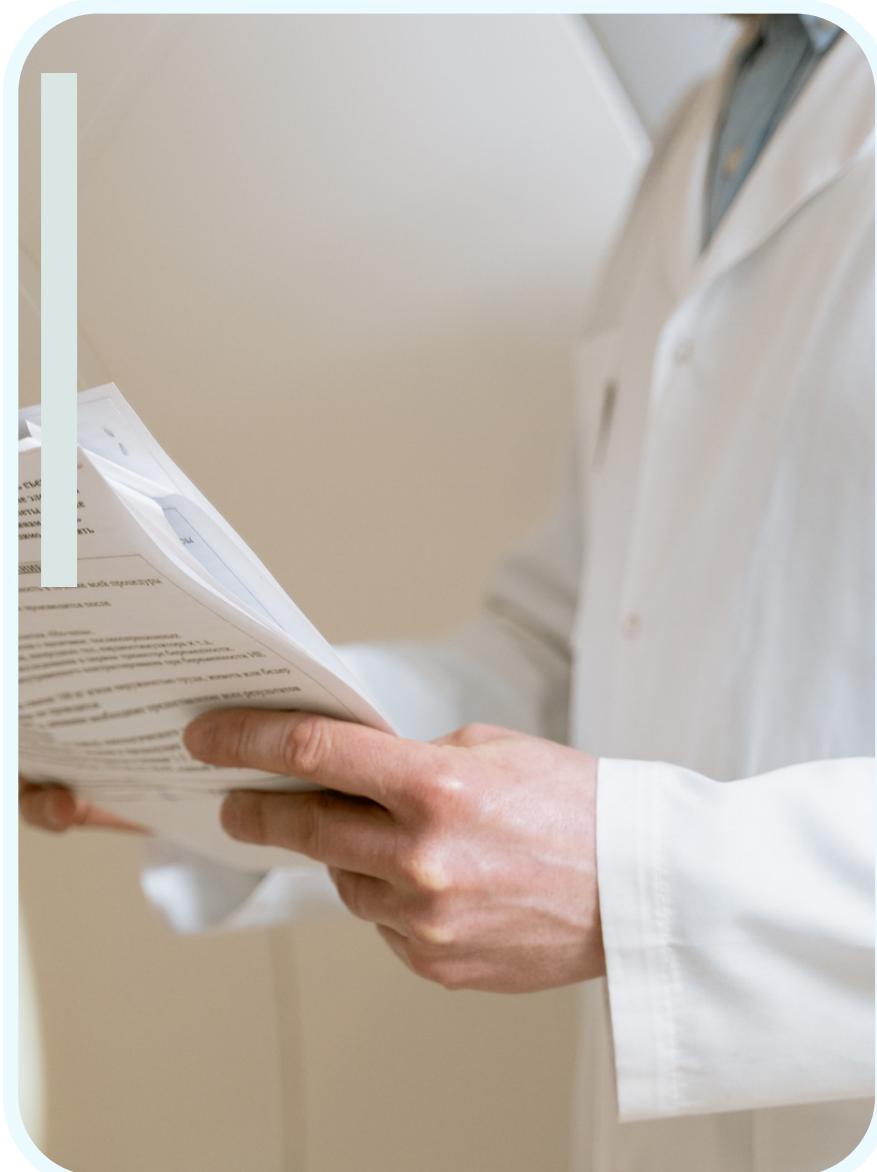
Kesimpulan :

Profit Growth (%) berdasarkan Kategori produk tahun 2021 vs 2022 ditunjukkan pada grafik disamping. Terlihat bahwa Profit growth (%) tertinggi adalah kategori Superstore, diikuti oleh Women Fashion dan Mobiles & Tablets yang menduduki peringkat kedua dan ketiga. Sedangkan 4 kategori dengan profit growth (%) rendah berada di posisi negatif. Dimana Books menjadi kategori dengan profit growth (%) paling rendah.

OUTPUT



Pertanyaan



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

1

Jawaban

INPUT

```
#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)

data16 = pd.DataFrame(df[(df['is_valid']==1) \
                        & (df['day'].isin(['Saturday','Sunday']))\
                        & (df['order_date'] >= '2022-01-01') & (df['order_date'] <= '2022-12-31'))\
                      .groupby(by=["month_num","month"])["before_discount"].mean()\
                      .round()\
                      .sort_values(ascending=False)\
                      .reset_index(name='avg_profit_weekend'))

data16

data17 = pd.DataFrame(df[(df['is_valid']==1) \
                        & (df['day'].isin(['Monday','Tuesday','Wednesday','Thursday','Friday']))\
                        & (df['order_date'] >= '2022-01-01') & (df['order_date'] <= '2022-12-31'))\
                      .groupby(by=["month_num","month"])["before_discount"].mean()\
                      .sort_values(ascending=False)\
                      .round()\
                      .reset_index(name='avg_profit_weekdays'))

data17

data18 = data16.merge(data17, left_on = 'month', right_on = 'month')
data18.sort_values(by='month_num_x',ascending=True, inplace=True)
data18 = data18[["month","avg_profit_weekend","avg_profit_weekdays"]]
data18
```

OUTPUT

	month	avg_profit_weekend	avg_profit_weekdays
2	January	197972.0	184474.0
0	February	214281.0	174799.0
6	March	162998.0	185331.0
7	April	156517.0	147262.0
1	May	208437.0	155492.0
11	June	94235.0	149632.0
3	July	169175.0	125668.0
10	August	113892.0	165806.0
5	September	164873.0	850539.0
9	October	124681.0	168393.0
8	November	138973.0	154603.0
4	December	168817.0	112494.0

INPUT

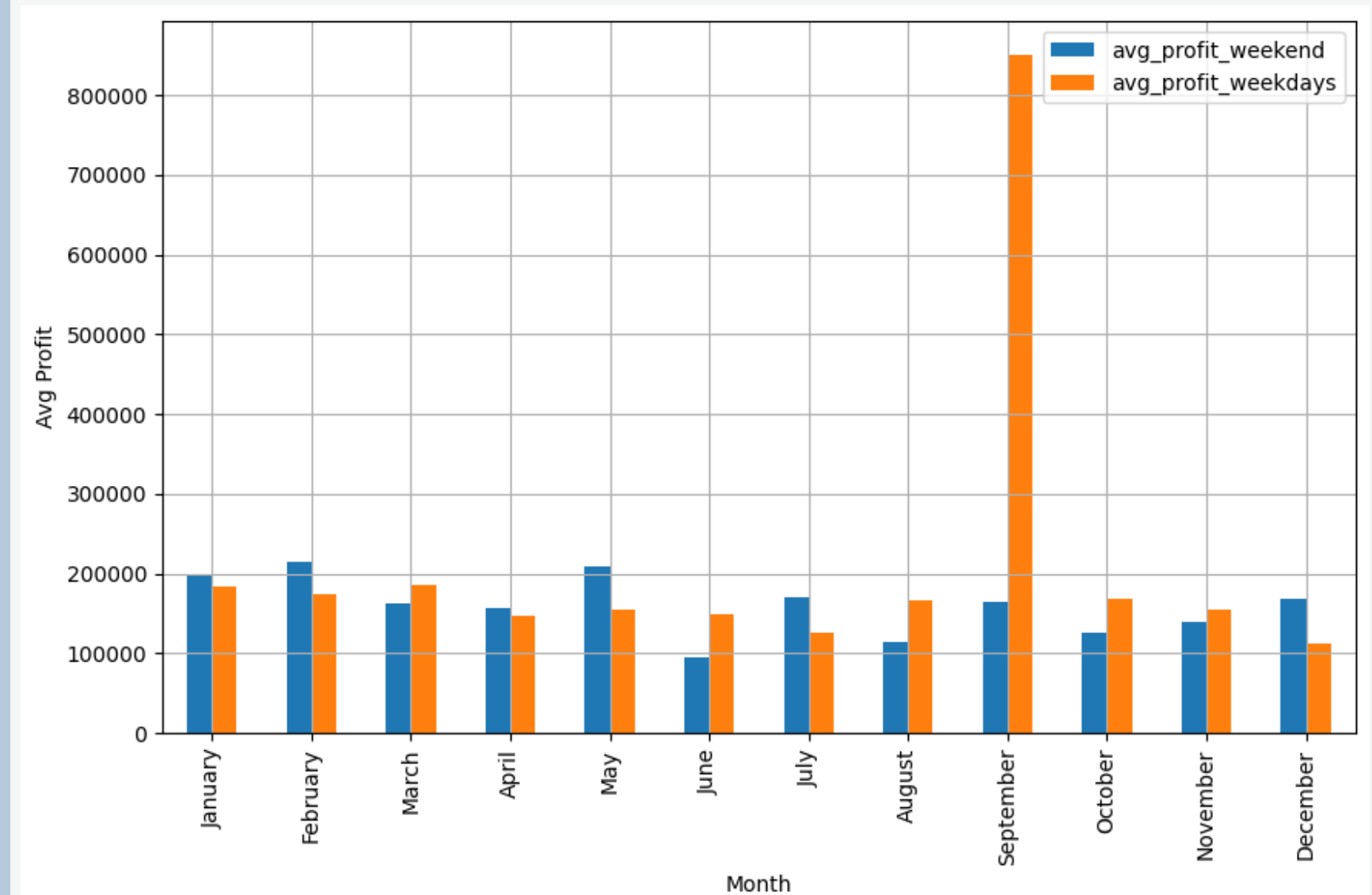
```
data18.plot(x='month',  
            y=['avg_profit_weekend','avg_profit_weekdays'],  
            kind='bar',  
            grid = True,  
            xlabel = 'Month',  
            ylabel = 'Avg Profit',  
            figsize=(12,7),  
            rot = 90,  
            table = False,  
            secondary_y = False)
```

Kesimpulan :

Berdasarkan grafik, pada Bulan Oktober sampai dengan Desember rata-rata harian penjualan weekends mengalami peningkatan penjualan dari bulan-bulan sebelumnya.

Sebaliknya, pada Bulan Oktober sampai dengan Desember rata-rata harian penjualan weekdays mengalami penurunan penjualan dari bulan-bulan sebelumnya.

OUTPUT



Dikarenakan campaign dilakukan setiap hari Sabtu dan Minggu (weekends), maka dapat disimpulkan bahwa campaign tersebut cukup berdampak pada kenaikan penjualan.

Jawaban

INPUT

```
#Menyimpan data dalam bentuk Pandas DataFrame
data19 = df[\
    #Memfilter data dengan valid=1
    (df['is_valid']==1) &\
    (df['day'].isin(['Saturday','Sunday'])) &\
    #Memfilter data pada transaksi selama 2022
    ((df['order_date'] >= '2022-10-01') & (df['order_date'] <= '2022-12-31'))]

#Menyimpan data dalam bentuk Pandas DataFrame
data20 = df[\
    #Memfilter data dengan valid=1
    (df['is_valid']==1) &\
    (df['day'].isin(['Monday','Tuesday','Wednesday','Thursday','Friday'])) &\
    #Memfilter data pada transaksi selama 2022
    ((df['order_date'] >= '2022-10-01') & (df['order_date'] <= '2022-12-31'))]

data21 = {\
    'Periode':'Total 3 months',\
    'Avg Weekend Sales': round(data19['before_discount'].mean(),2), \
    'Avg Weekdays Sales': round(data20['before_discount'].mean(),2),\
    'Diff (Value)': round(data19['before_discount'].mean() - data20['before_discount'].mean(),2),\
    'Diff (%)': pd.Series(round(((data19['before_discount'].mean() - data20['before_discount'].mean())/data19['before_discount'].mean()*100,2), dtype=str)+'%'
}

pd.DataFrame(data=data21, index=[0])
```

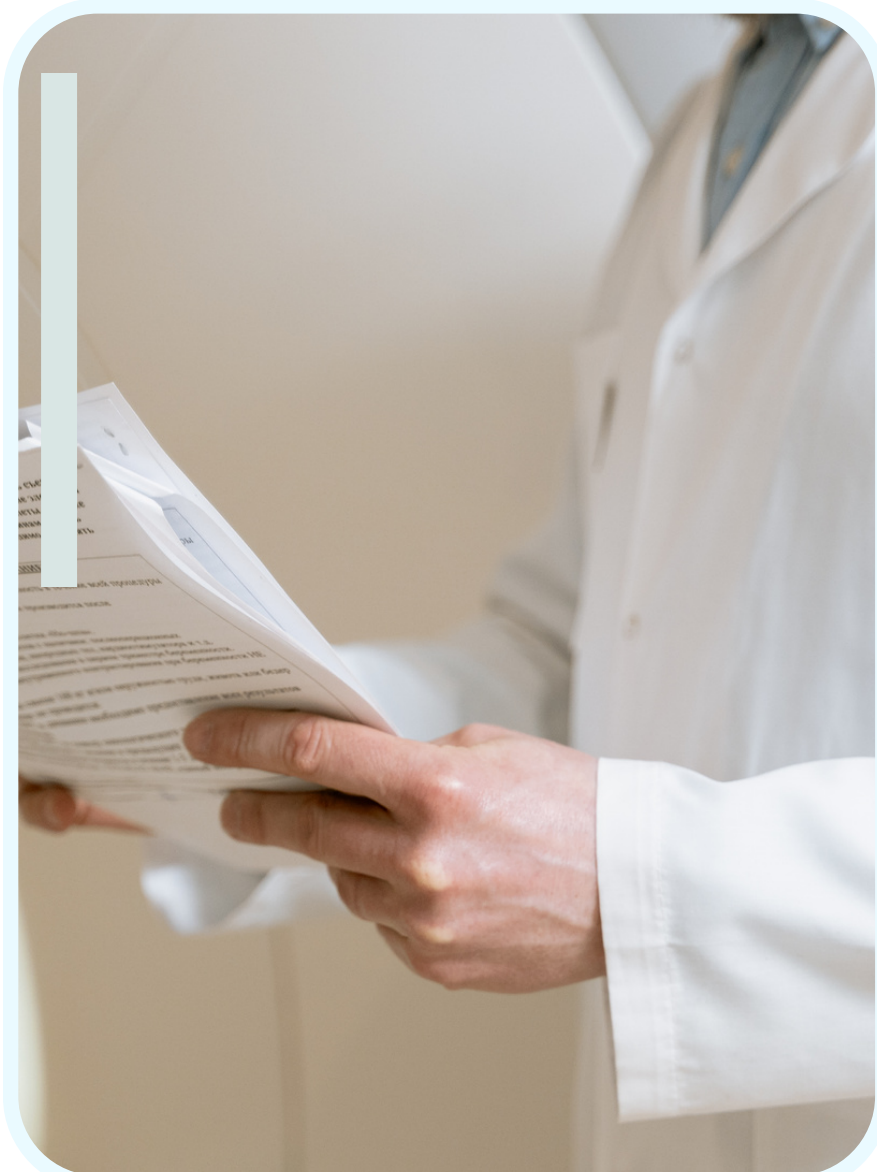
OUTPUT

	Periode	Avg Weekend Sales	Avg Weekdays Sales	Diff (Value)	Diff (%)
0	Total 3 months	136301.11	143687.84	-7386.73	-5.42%

Kesimpulan :

Secara keseluruhan dalam kurun waktu 3 bulan yaitu Bulan Oktober-Desember, rata-rata harian penjualan weekends sebesar 136301,11. Sedangkan rata-rata harian penjualan weekdays adalah sebesar 143687,84. Hal ini menunjukkan rata-rata harian weekdays lebih besar 5,42% dari rata-rata harian weekend.

Pertanyaan



Dear Data Analyst,

Guna mengetahui kemampuan beli pelanggan, kami membutuhkan data:

- 1.AOV (Average Order Value atau Total sales / Total Order) tiap bulan selama tahun 2021 vs 2022.
- 2.AOV tahun 2021 vs 2022.

Atas bantuan yang diberikan, kami mengucapkan terima kasih

Regards
Tim Campaign

```
data22 = pd.DataFrame(df[(df['is_valid']==1) & ((df['order_date'] >= '2022-01-01') & (df['order_date'] <= '2022-12-31'))]\
    .groupby(by=['month_num', 'month'])["before_discount"].sum()\
    .round()\
    .reset_index(name='sales_2022'))\
    .sort_values(by=['month_num'], ascending=True)\
    .head(12)
```

data22

```
data23 = pd.DataFrame(df[(df['is_valid']==1) & ((df['order_date'] >= '2022-01-01') & (df['order_date'] <= '2022-12-31'))]\
    .groupby(by=['month_num', 'month'])["id"].nunique()\
    .round()\
    .reset_index(name='order_2022'))\
    .sort_values(by=['month_num'], ascending=True)\
    # .head(12)
```

data23

```
data24 = pd.DataFrame(df[(df['is_valid']==1) & ((df['order_date'] >= '2021-01-01') & (df['order_date'] <= '2021-12-31'))]\
    .groupby(by=['month_num', 'month'])["before_discount"].sum()\
    .round()\
    .reset_index(name='sales_2021'))\
    .sort_values(by=['month_num'], ascending=True)\
    .head(12)
```

data24

```
data25 = pd.DataFrame(df[(df['is_valid']==1) & ((df['order_date'] >= '2021-01-01') & (df['order_date'] <= '2021-12-31'))]\
    .groupby(by=['month_num', 'month'])["id"].nunique()\
    .round()\
    .reset_index(name='order_2021'))\
    .sort_values(by=['month_num'], ascending=True)\
    .head(12)
```

data25

INPUT

```
from sqlite3 import connect
conn = connect(':memory:')
data22.to_sql('sales_2022', conn, index=False, if_exists='replace')
data23.to_sql('order_2022', conn, index=False, if_exists='replace')
data24.to_sql('sales_2021', conn, index=False, if_exists='replace')
data25.to_sql('order_2021', conn, index=False, if_exists='replace')
```

INPUT

```
data26 = pd.read_sql("""
SELECT
    sales_2021.*,
    order_2021.order_2021,
    sales_2022.sales_2022,
    order_2022.order_2022
FROM sales_2022
LEFT JOIN order_2022
    on order_2022.month_num = sales_2022.month_num
LEFT JOIN sales_2021
    on sales_2021.month_num = sales_2022.month_num
LEFT JOIN order_2021
    on order_2021.month_num = sales_2022.month_num
""", conn)
data26
```

OUTPUT

	month_num	month	sales_2021	order_2021	sales_2022	order_2022
0	1	January	393963352	1331	1577227762	4799
1	2	February	269498794	1119	1512559408	4790
2	3	March	277998494	1101	2314882193	6922
3	4	April	251696051	1080	2223917288	6348
4	5	May	275985904	992	3803696734	9037
5	6	June	294017782	974	1366975717	5219
6	7	July	1076526211	3132	1773954449	6278
7	8	August	1283552029	4019	740712197	2412
8	9	September	1729388535	5282	1548445294	1692
9	10	October	1441307055	4428	494201049	1772
10	11	November	5501400857	22414	622241676	2047
11	12	December	1281255946	4842	443946546	1747

INPUT

```
data26['AOV 2021'] = round(data26['sales_2021']/data26['order_2021'],2)
data26['AOV 2022'] = round(data26['sales_2022']/data26['order_2022'],2)
data26['Diff AOV (value)']= data26['AOV 2022']-data26['AOV 2021']
data26['Diff AOV (%)'] = round((data26['AOV 2022']-data26['AOV 2021'])/data26['AOV 2021']*100,2)
data26
```

OUTPUT

	month_num	month	sales_2021	order_2021	sales_2022	order_2022	AOV 2021	AOV 2022	Diff AOV (value)	Diff AOV (%)
0	1	January	393963352	1331	1577227762	4799	295990.50	328657.59	32667.09	11.04
1	2	February	269498794	1119	1512559408	4790	240838.96	315774.41	74935.45	31.11
2	3	March	277998494	1101	2314882193	6922	252496.36	334423.89	81927.53	32.45
3	4	April	251696051	1080	2223917288	6348	233051.90	350333.54	117281.64	50.32
4	5	May	275985904	992	3803696734	9037	278211.60	420902.59	142690.99	51.29
5	6	June	294017782	974	1366975717	5219	301866.31	261922.92	-39943.39	-13.23
6	7	July	1076526211	3132	1773954449	6278	343718.46	282566.81	-61151.65	-17.79
7	8	August	1283552029	4019	740712197	2412	319371.00	307094.61	-12276.39	-3.84
8	9	September	1729388535	5282	1548445294	1692	327411.69	915156.79	587745.10	179.51
9	10	October	1441307055	4428	494201049	1772	325498.43	278894.50	-46603.93	-14.32
10	11	November	5501400857	22414	622241676	2047	245444.85	303977.37	58532.52	23.85
11	12	December	1281255946	4842	443946546	1747	264612.96	254119.37	-10493.59	-3.97

INPUT

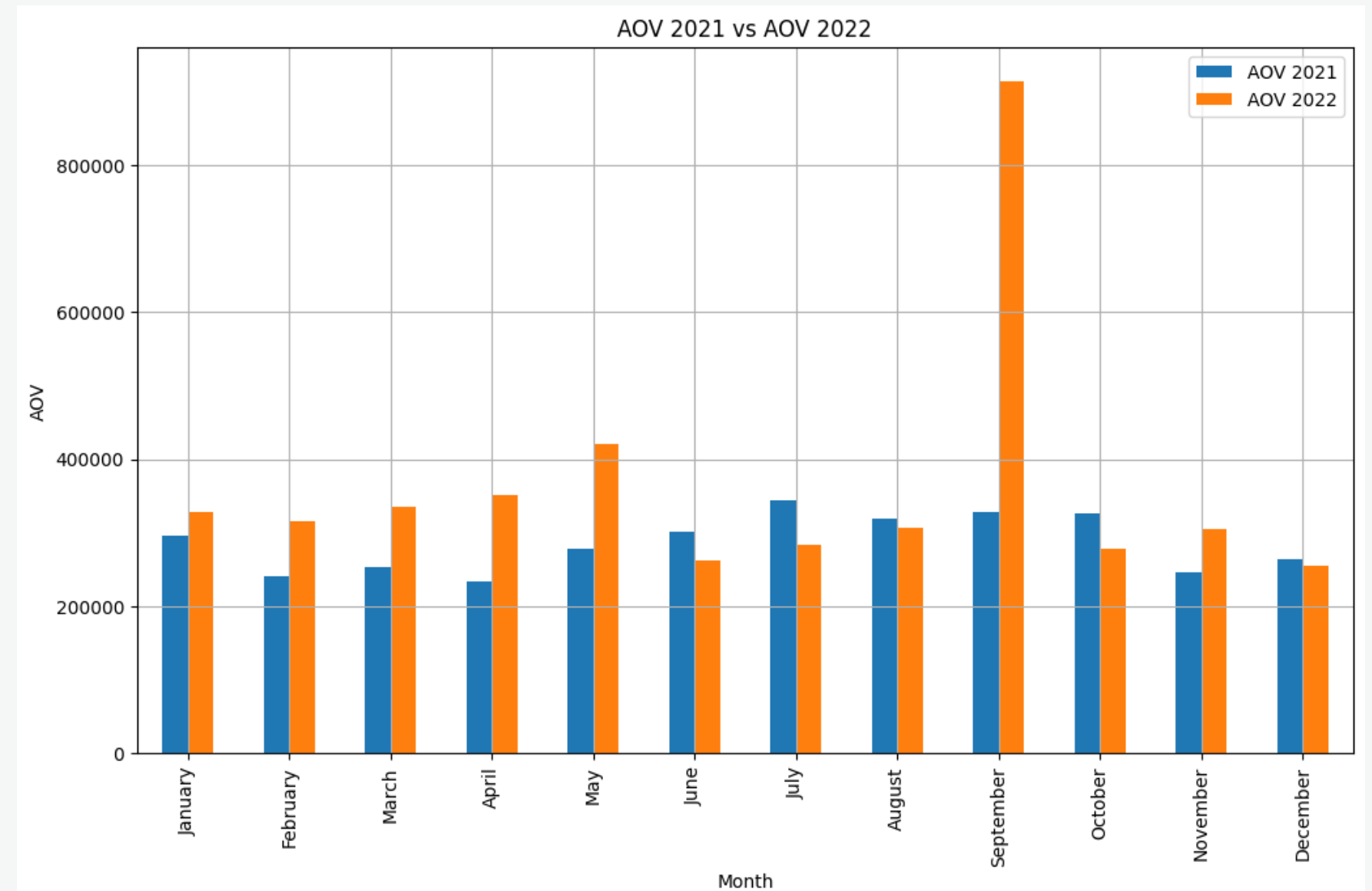
```
data26.plot(x='month',  
            y=['AOV 2021','AOV 2022'],  
            kind='bar',  
            grid = True,  
            title = 'AOV 2021 vs AOV 2022',  
            xlabel = 'Month',  
            ylabel = 'AOV',  
            figsize=(12,7),  
            rot = 90,  
            table = False,  
            secondary_y = False)
```

Kesimpulan :

Berdasarkan grafik angka AOV terbesar pada tahun 2021 yaitu terjadi di Bulan Juli yaitu sebesar 343718,46. Sedangkan angka AOV terendah pada tahun 2021 terjadi di Bulan April yaitu sebesar 233051,90.

Angka AOV terbesar pada tahun 2022 terjadi di Bulan September yaitu sebesar 915156.79. Sedangkan angka AOV terendah pada tahun 2022 terjadi di Bulan Desember yaitu sebesar .

OUTPUT



Jawaban

10

INPUT

```
aov_2021 = round(data26['sales_2021'].sum()/data26['order_2021'].sum(),2)
aov_2022 = round(data26['sales_2022'].sum()/data26['order_2022'].sum(),2)
data27 = {\
    'Periode':'Total',\
    'AOV 2021': aov_2021, \
    'AOV 2022': aov_2022, \
    'Growth (value)': aov_2022-aov_2021,\
    'Growth': pd.Series(round((aov_2022-aov_2021)/aov_2021*100,2), dtype=str)+'%\
}\
pd.DataFrame(data=data27, index=[0])
```

INPUT

	Periode	AOV 2021	AOV 2022	Growth (value)	Growth
0	Total	277568.15	347186.56	69618.41	25.08%

Kesimpulan :

AOV pada tahun 2021 adalah sebesar 277.568,15, sedangkan AOV pada tahun 2022 adalah sebesar 347.186,56. Hal ini menunjukkan terjadi peningkatan AOV sebesar 25,8% dari tahun 2021.

LINK

<https://colab.research.google.com/drive/16sFKtbgH5NeWipftvA9zEcOI6zQtqoti?usp=sharing>



Terima Kasih

