

A glowing, semi-transparent globe is centered in the image. It features a network of white lines connecting various points, with some points highlighted in yellow and orange. The globe is positioned over a blurred background of a laptop keyboard. The overall color palette is light blue and white, with a soft, ethereal glow.

# Project

Data Analyst

Achmad Fani F I

# PYTHON



**LINK**

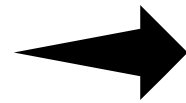
<https://colab.research.google.com/drive/1-aUE9xvHjk4uWd2FmJJOoo2M6UGeAPns?usp=sharing>



# DATA PREPARATION

## Step 1

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

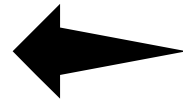


Import Library

## Step 2

```
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)
```

Import Dataset dari GitHub



## Step 3

```
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')
```



Connect Dataset dari SQL

# DATA PREPARATION

## Step 4

```
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)
```



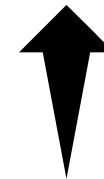
Menggabungkan Dataset dengan SQL

```
df['order_date'] = pd.to_datetime(df['order_date'])
df['registered_date'] = pd.to_datetime(df['registered_date'])
df.dtypes
```



## Step 5

Mengubah tipe data Float menjadi Int dan tipe kolom Date



```
df = df.astype({"before_discount": 'int', "discount_amount": 'int', "after_discount": 'int', "base_price": 'int'})
df.dtypes
```

# Pertanyaan 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 1

## Query

```
df1 = df[(df['is_valid'] == 1) & (df['category'] == 'Mobiles & Tablets') & (df['order_date'].dt.year == 2022)]\
        .groupby('sku_name')['qty_ordered'].sum()\
        .sort_values(ascending=False).head(5).reset_index(name = 'qty_ordered')\
df1
```

## Result

	sku_name	qty_ordered
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

## Insight

Penjualan produk IDROID\_BALRX7-Gold mendominasi dengan 1.000 unit, jauh di atas produk lain.

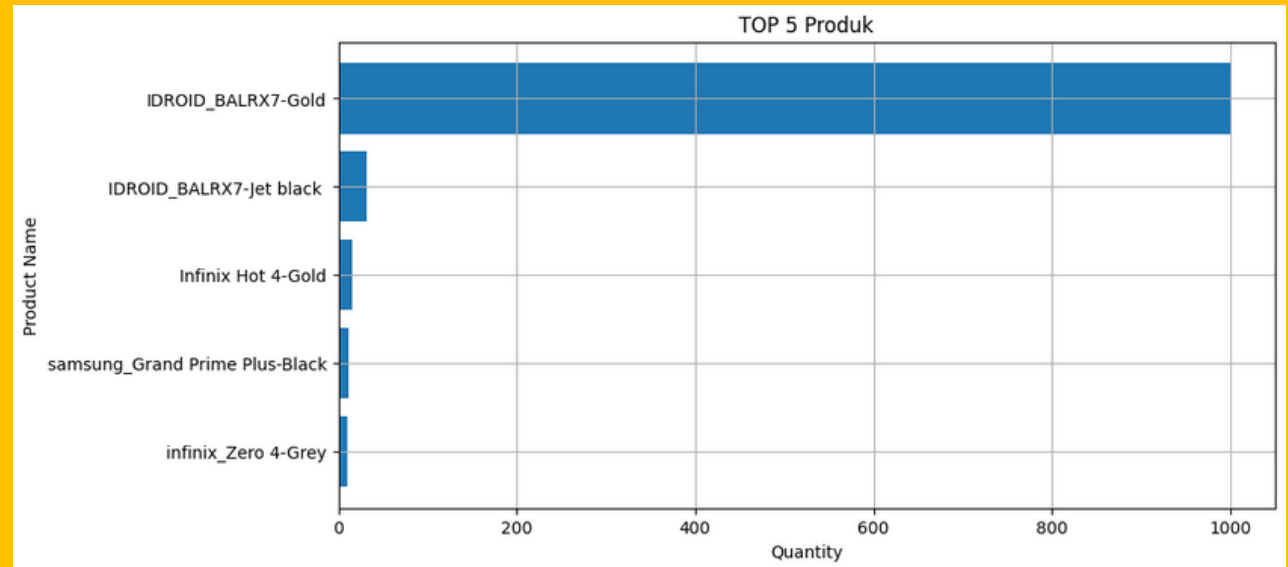
# Jawaban 1

## Query

```
df1.sort_values(by='qty_ordered', ascending=True, inplace=True)

# Grafik
x=df1['sku_name']
y=df1['qty_ordered']
plt.figure(figsize=(10,5))
plt.title('TOP 5 Produk')
plt.xlabel('Quantity')
plt.ylabel('Product Name')
plt.grid(True)
plt.barh(x,y)
plt.show()
```

## Result



Dari TOP 5 produk teratas, produk tersebut sangat cocok digunakan sebagai hadiah dalam kompetisi **Festival Akhir Tahun**. Terkhusus untuk produk IDROID\_BALRX7-Gold bisa dikoordinasikan terlebih dahulu dengan Tim Warehouse terkait ketersediaan produk, dikarena produk IDROID\_BALRX7-Gold terjual paling banyak hingga mencaai 1000 unit.

# Pertanyaan 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 2.1

## Query

## Result

```
qty2022 = df[
    (df['is_valid'] == 1) & (df['order_date'].dt.year == 2022)]\
    .groupby('category')['qty_ordered'].sum()\
    .reset_index(name='qty_2022')

qty2021 = df[
    (df['is_valid'] == 1) & (df['order_date'].dt.year == 2021)]\
    .groupby('category')['qty_ordered'].sum()\
    .reset_index(name='qty_2021')

qty15 = qty2021.merge(qty2022, on='category', how='outer')
qty15.fillna(0, inplace=True)
qty15['qty_difference'] = qty15['qty_2022'] - qty15['qty_2021']
qty15.sort_values(by='qty_difference', ascending=True, inplace=True)

percentage_change = round(((qty15['qty_2022'] - qty15['qty_2021']) / qty15['qty_2021']) * 100, 2)
qty15['information'] = np.where(qty15['qty_2022'] > qty15['qty_2021'], '↑ ' + percentage_change.astype(str) + '%',
                               np.where(qty15['qty_2022'] < qty15['qty_2021'], '↓ ' + percentage_change.astype(str) + '%', '→ 0.0%'))
qty15 = qty15.head(15)

qty15
```

	category	qty_2021	qty_2022	qty_difference	information
10	Others	426	263	-163	↓ -38.26%
12	Soghaat	759	612	-147	↓ -19.37%
8	Men Fashion	237	175	-62	↓ -26.16%
1	Beauty & Grooming	168	153	-15	↓ -8.93%
0	Appliances	124	148	24	↑ 19.35%
2	Books	171	195	24	↑ 14.04%
5	Health & Sports	173	200	27	↑ 15.61%
3	Computing	109	153	44	↑ 40.37%
11	School & Education	184	237	53	↑ 28.8%
6	Home & Living	193	250	57	↑ 29.53%
7	Kids & Baby	170	227	57	↑ 33.53%
4	Entertainment	77	150	73	↑ 94.81%
13	Superstore	327	536	209	↑ 63.91%
14	Women Fashion	140	489	349	↑ 249.29%
9	Mobiles & Tablets	107	1154	1047	↑ 978.5%

## Insight

Dugaan ialah benar, terjadi penurunan kuantitas penjualan dari kategori Others pada tahun 2022 dibanding tahun 2021. Selisih penurunan tersebut mencapai -163 atau -38,26%.

# Jawaban 2.2

## Query

```
# Data kuantitas penjualan 2022 berdasarkan sku_name
sku2022 = df[
    (df['is_valid'] == 1) & (df['category'] == 'Others') & (df['order_date'].dt.year == 2022)]\
    .groupby('sku_name')['qty_ordered'].sum()\
    .reset_index(name='qty_2022')

# Data kuantitas penjualan 2021 berdasarkan sku_name
sku2021 = df[
    (df['is_valid'] == 1) & (df['category'] == 'Others') & (df['order_date'].dt.year == 2021)]\
    .groupby('sku_name')['qty_ordered'].sum()\
    .reset_index(name='qty_2021')

# Menggabungkan data dari sku2021 dengan sku2022
sku20 = sku2021.merge(sku2022, on='sku_name', how='outer')
sku20.fillna(0, inplace=True)

# Menghitung selisih kuantitas penjualan 2022 dengan 2021
sku20['difference'] = sku20['qty_2022'] - sku20['qty_2021']
sku20.sort_values(by='difference', ascending=True, inplace=True)

# Menghitung persentase selisih kuantitas penjualan
percentage_change = round(((sku20['qty_2022'] - sku20['qty_2021']) / sku20['qty_2021']) * 100, 2)
sku20['information'] = np.where(sku20['qty_2022'] > sku20['qty_2021'], '↑ ' + percentage_change.astype(str) + '%',
                                np.where(sku20['qty_2022'] < sku20['qty_2021'], '↓ ' + percentage_change.astype(str) + '%', '→ 0.0%'))

sku20 = sku20.head(20)
sku20
```

## Result

	sku_name	qty_2021	qty_2022	difference	information
69	RB_Dettol Germ Busting Kit-bf	200.0	45.0	-155.0	↓ -77.5%
18	Dawlance_MD 10 + DWB 600	23.0	0.0	-23.0	↓ -100.0%
84	Telemall_MM-DR-HB-L	23.0	2.0	-21.0	↓ -91.3%
137	iu_Tickets General Enclosure-Islamabad	20.0	0.0	-20.0	↓ -100.0%
70	RS_Rehmat-e-Shereen Mix Mithai	13.0	0.0	-13.0	↓ -100.0%
141	kansai_NeverWet	10.0	1.0	-9.0	↓ -90.0%
154	sindbad_Sindbad Gift Card-3	7.0	0.0	-7.0	↓ -100.0%
133	emart_00-1	7.0	1.0	-6.0	↓ -85.71%
93	Vouch 365 2016	5.0	0.0	-5.0	↓ -100.0%
4	Am-PTV_ATS-004-M	5.0	0.0	-5.0	↓ -100.0%
129	duma_4561253300294	4.0	0.0	-4.0	↓ -100.0%
155	sockoye_QG in Quarter Grey	4.0	0.0	-4.0	↓ -100.0%
85	The Vitamin Company Kojic Acid Whitening Cream...	4.0	0.0	-4.0	↓ -100.0%
115	aw_Octane Booster-12oz./354ml	3.0	0.0	-3.0	↓ -100.0%
43	MEGUIAR_G12711	4.0	1.0	-3.0	↓ -75.0%
88	Trans2_LW 999	3.0	0.0	-3.0	↓ -100.0%
50	MEGUIAR_G19216	2.0	0.0	-2.0	↓ -100.0%
26	JBS_IFAM-009	2.0	0.0	-2.0	↓ -100.0%
60	MEGUIAR_X1030EU	2.0	0.0	-2.0	↓ -100.0%
156	sstop_Universallensclipkit	2.0	0.0	-2.0	↓ -100.0%

Salah satu produk yang mengalami penurunan terbanyak yaitu RB\_Dettol Germ Busting Kit-bf dengan selisih -155 produk yang terjual dan persentase penurunan sekitar 77.5%. Oleh karena itu, perlu adanya tindak lanjut untuk melakukan feedback setelah pembelian, dan evaluasi penjualan dengan rebranding produk dan repositioning pelanggan agar dapat menghasilkan include peningkatan profit penjualan.

# Pertanyaan 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 3

## Query

```
promo = df[(df['is_gross'] == 1) & (df['is_valid'] == 0) & (df['is_net'] == 0) & (df['order_date'].dt.year == 2022)]

df_promo = promo[['customer_id', 'registered_date']]
df_promo.drop_duplicates(inplace=True)
df_promo
```

## Insight

Berdasarkan data yang telah didapat menunjukkan ada sekitar 820 pelanggan yang memenuhi kriteria yaitu pelanggan yang sudah melakukan check-out namun belum melakukan pembayaran.

## Rekomendasi

Tim Digital Marketing bisa memperluas informasi promo ulang tahun tidak hanya untuk pelanggan yang sudah melakukan check-out namun belum membayar, tetapi informasi tersebut bisa disampaikan kepada pelanggan yang belum melakukan check-out, guna menarik pelanggan baru.

## Result

	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		

# Pertanyaan 4

Dear Data Analyst,

Pada bulan Oktober 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**

## Query

## Result

[illegible]

# Jawaban 4.1

Mencari rata-rata penjualan Weekend vs Weekday

Query

```
monthly_weekday2 = pd.DataFrame(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"])[["before_discount"].mean()
    .round()
    .reset_index(name='avg_sales_weekday'))
monthly_weekend2 = pd.DataFrame(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"])[["before_discount"].mean()
    .round()
    .reset_index(name='avg_sales_weekend'))
monthly2 = monthly_weekend2.merge(monthly_weekday2, left_on = 'month', right_on = 'month')
monthly2.sort_values(by='month',ascending=False, inplace=True)
monthly2 = monthly2[["month","avg_sales_weekend","avg_sales_weekday"]]

df_groupby_weekend['diff_value'] = df_groupby_weekend['avg_sales_weekend'] - df_groupby_weekend['avg_sales_weekday']
df_groupby_weekend['diff_percent (%)'] = (round(df_groupby_weekend['diff_value']/df_groupby_weekend['avg_sales_weekend']*100)).astype(str) + '%'

df_groupby_weekend
```

Result

	month	avg_sales_weekend	avg_sales_weekday	diff_value	diff_percent (%)
2	October	634260.0	874690.0	-240430.0	-38.0%
1	November	607794.0	641862.0	-34068.0	-6.0%
0	December	410599.0	813574.0	-402975.0	-98.0%



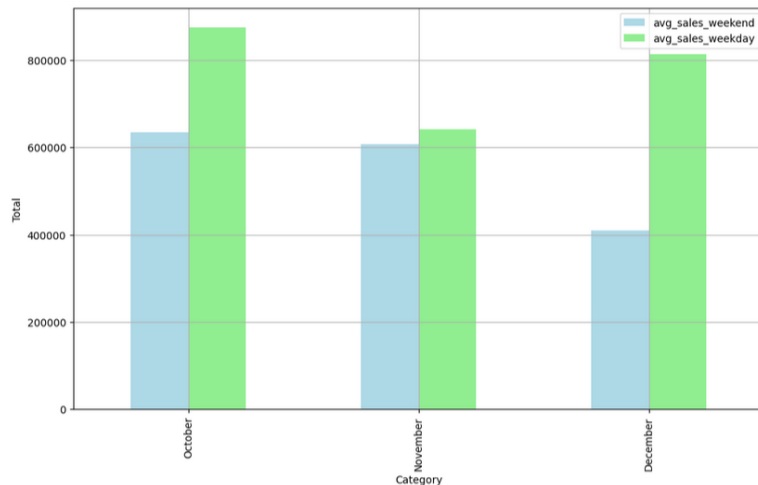
# Jawaban 4.1

## Visualisasi diagram batang hasil rata-rata penjualan kuartal 4 2022

### Query

```
monthly2.plot(x='month',
              y=['avg_sales_weekend', 'avg_sales_weekday'],
              kind='bar',
              color=['lightblue', 'lightgreen'],
              grid = True,
              xlabel = 'Category',
              ylabel = 'Total',
              figsize=(12,7),
              rot = 90,
              table = False,
              secondary_y = False)
```

### Result



### Insight

Diperoleh, rata - rata penjualan bulan Oktober - Desember 2022 adalah

- Berdasarkan data yang didapat menunjukkan bahwa rata-rata penjualan weekend mencapai nilai 634.260 dan untuk rata-rata penjualan weekday mencapai nilai 874.690. Rata-rata penjualan weekend dan weekday pada bulan oktober menjadi yang tertinggi diantara bulan November dan Desember.
- Peningkatan nilai rata-rata penjualan hanya terjadi pada weekday bulan Desember. Untuk nilai rata-rata penjualan weekend selama bulan Oktober sampai Desember mengalami penurunan.
- Rata-rata penjualan weekend pada bulan Desember menjadi menjadi yang terendah dan memiliki selisih yang sangat jauh dengan rata-rata penjualan weekday bulan Desember, selisih nilai tersebut mencapai 402.975 atau sekitar 98%.



# Jawaban 4.2

Menghitung rata-rata penjualan Weekends vs Weekdays secara keseluruhan

Query

```
data21 = {\n    'Periode': 'Total 3 months',\n    'Avg Weekend Sales': round(monthly_weekend2['avg_sales_weekend'].mean(),2), \n    'Avg Weekdays Sales': round(monthly_weekday2['avg_sales_weekday'].mean(),2),\n    'Diff (Value)': round(monthly_weekend2['avg_sales_weekend'].mean() - monthly_weekday2['avg_sales_weekday'].mean(),2),\n    'Diff (%)': pd.Series(round(((monthly_weekend2['avg_sales_weekend'].mean() - monthly_weekday2['avg_sales_weekday'].mean())/monthly_weekend2['avg_sales_weekend'].mean())*100,2), dtype=str)+'%\n}\n\npd.DataFrame(data=data21, index=[0])
```

Result

	Periode	Avg Weekend Sales	Avg Weekdays Sales	Diff (Value)	Diff (%)
0	Total 3 months	550884.33	776708.67	-225824.33	-40.99%

Insight

- Rata-rata penjualan weekdays lebih tinggi daripada weekends dengan selisih nilai rata-rata mencapai 225.824,33 atau sekitar 40,99%. Dari hasil tersebut menunjukkan campaign weekends belum efektif dalam meningkatkan penjualan di hari tersebut.
- Rekomendasi: Perlu peningkatan efektivitas campaign weekends melalui evaluasi strategi, promosi, atau insentif. Pendekatan dan target audiens juga perlu ditinjau agar penjualan weekends dapat mendekati atau melampaui weekdays.

*Terima Kasih*