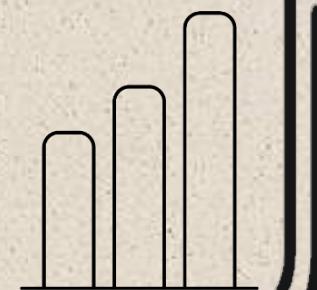




# PORTOFOLIO



MUZNI ZAKI RIYADI



# Muzni Zaki Riyadi

*Fresh graduate Statistika dengan keahlian dalam Data Analytics, Data Mining, dan Machine Learning yang dipelajari selama kuliah dan bootcamp data analyst. Memiliki pengalaman dalam analisis data menggunakan SQL, Python, dan R, serta mampu mengimplementasikan metode Descriptive, Diagnostic, Predictive, dan Prescriptive Analytics untuk mendukung pengambilan keputusan. Menguasai konsep forecasting dan statistik inferensial untuk analisis data yang lebih mendalam. Berpengalaman untuk membuat visualisasi interaktif dan dashboard. Memiliki keterampilan komunikasi yang baik dan mampu berkolaborasi dalam tim untuk mengoptimalkan solusi berbasis data.*



# Pendidikan \*

# Organisasi



Universitas Syiah Kuala, 2020-2024  
FMIPA, Statistika  
IPK: 3.45/4.0

Penelitian Tugas Akhir: Implementasi Metode *Structural Equation Modeling* untuk Menganalisis Faktor-Faktor yang Memengaruhi Kepuasan dan Loyalitas Nasabah Bank Syariah Indonesia (Studi Kasus: Mahasiswa USK)



HIMPUNAN MAHASISWA STATISTIKA, Feb 2023 – Feb 2024  
Anggota Departemen *Organization and Human Resource*



Computational and Applied Statistics Research Group, Jan 2023 – Sekarang  
Anggota Divisi Monitoring dan Evaluasi



Computational and Applied Statistics Research Group, Jan 2023 – Des 2023  
Wakil Ketua



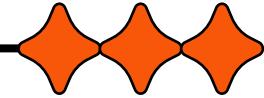
Asisten Laboratorium FMIPA USK, Mar 2022 – Mei 2023  
Mata kuliah Pengantar Statistika



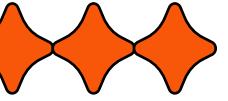
Tendor Pengantar Statistika, Mar 2022 – Mei 2023  
Mata kuliah Pengantar Statistika



Unit Kegiatan Mahasiswa, Apr 2021 – Nov 2024  
Anggota Bulutangkis



# Pengalaman \*



REMOTE, Desember 2024 – Sekarang

## Pusat Olimpiade Sains Indonesia

- Menyusun dan menulis soal ujian berkualitas tinggi untuk Pusat Olimpiade Sains Indonesia (POSI).
- Memastikan soal sesuai dengan standar akademik dan relevan untuk tingkat kompetisi nasional.
- Bekerja sama dengan ahli materi untuk memastikan konsistensi dan kualitas konten.



MAGANG, Januari 2023 – Juni 2023

## Dinas Komunikasi, Informatika dan Statistik Kota Banda Aceh

- Mengumpulkan dan menganalisis data sektoral untuk mendukung transformasi digital dan meningkatkan layanan informasi publik.
- Memperbarui data sektoral guna meningkatkan aksesibilitas dan keakuratan informasi.
- Menyusun laporan dan dokumen resmi secara sistematis menggunakan Microsoft Excel dan Word untuk mendukung transformasi digital.



MAGANG, Juni 2022 – Juli 2022

## Dinas Pemberdayaan Perempuan dan Perlindungan Anak Aceh

- Membuat grafik dan laporan data kasus menggunakan Microsoft Excel, memastikan presentasi yang informatif dan terstruktur.
- Menganalisis dan mengolah data kasus kekerasan dan indeks pemberdayaan gender di Aceh.
- Melakukan pendataan dan kegiatan lapangan bersama tim.

# Sertifikat



**Design Principles**  
208983/GFD/LM/12/2024, 13 Desember 2024



**Frontend in Website Development**  
208840/WEB/LM/12/2024, 12 Desember 2024



**Compensation & Benefits**  
207226/HRC/LM/12/2024, 10 Desember 2024



**Index Match, Vlookup and Hlookup in Ms. Excel**  
205135/EXL/LM/12/2024, 4 Desember 2024



**Introduction to Data Analytics**  
7636109, 6 Desember 2024



**Data Analyst 101: Excel Formulas & Functions**  
7634228, 5 Desember 2024



**Data Analysis : Fullstack Intensive Bootcamp**  
MS-7/3/2025-RgqUbGN1ETswMh4Oq5U4, 7 Maret 2025



**Program Basiq SQL 1 Kejuruan TIK**  
2.8/1767/LP.00.04/IV/2025, 28 April 2025



# Keahlian



## Analisis Statistik & Data



## Survei & Pengumpulan Data



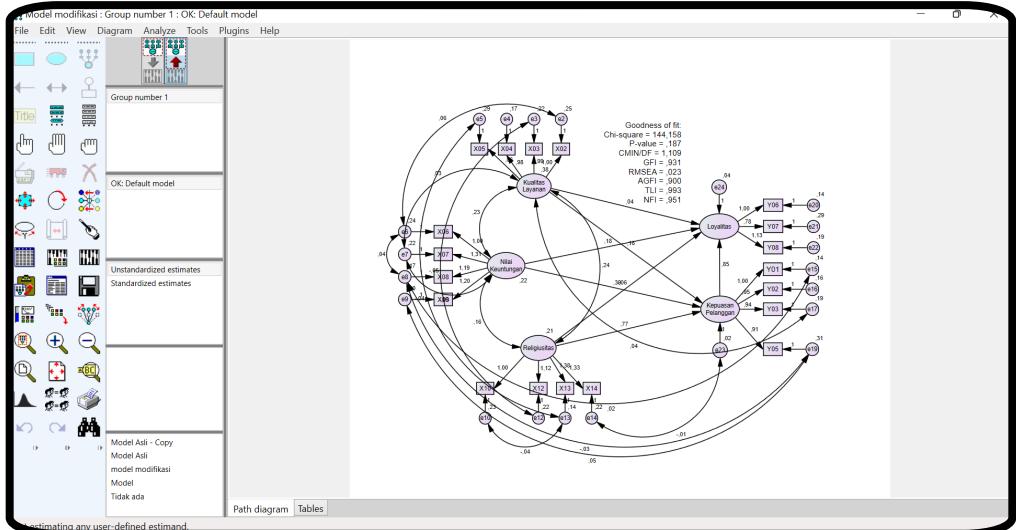
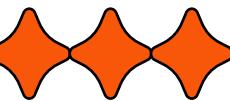
## Pemrograman & Ilmu Data



## Desain

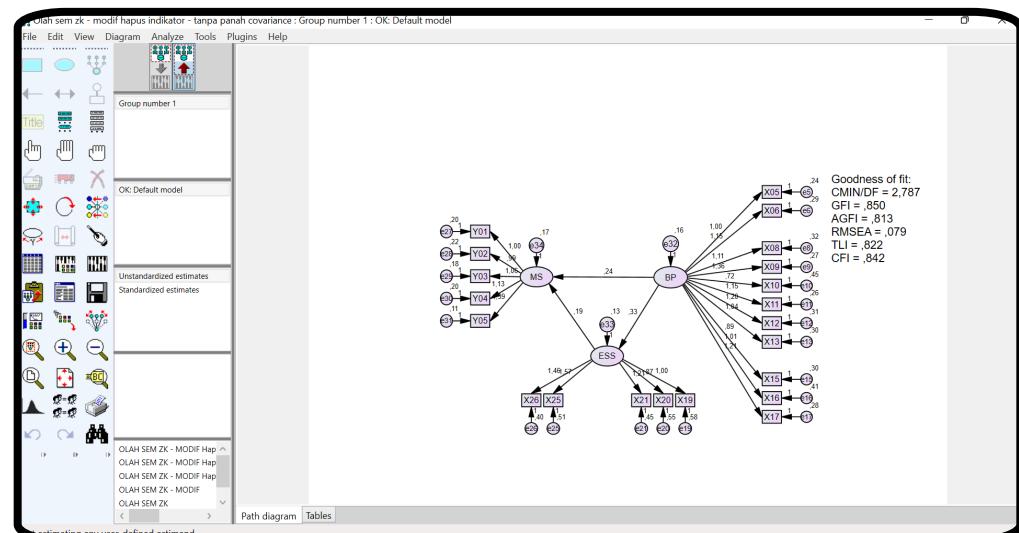


# Analisis Statistik & Data



## AMOS

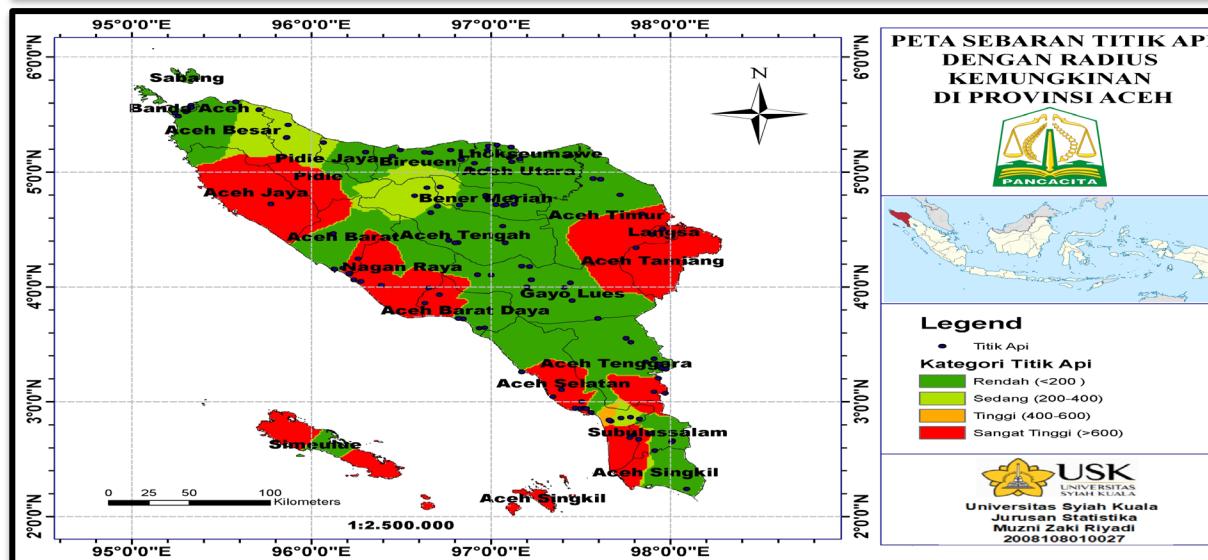
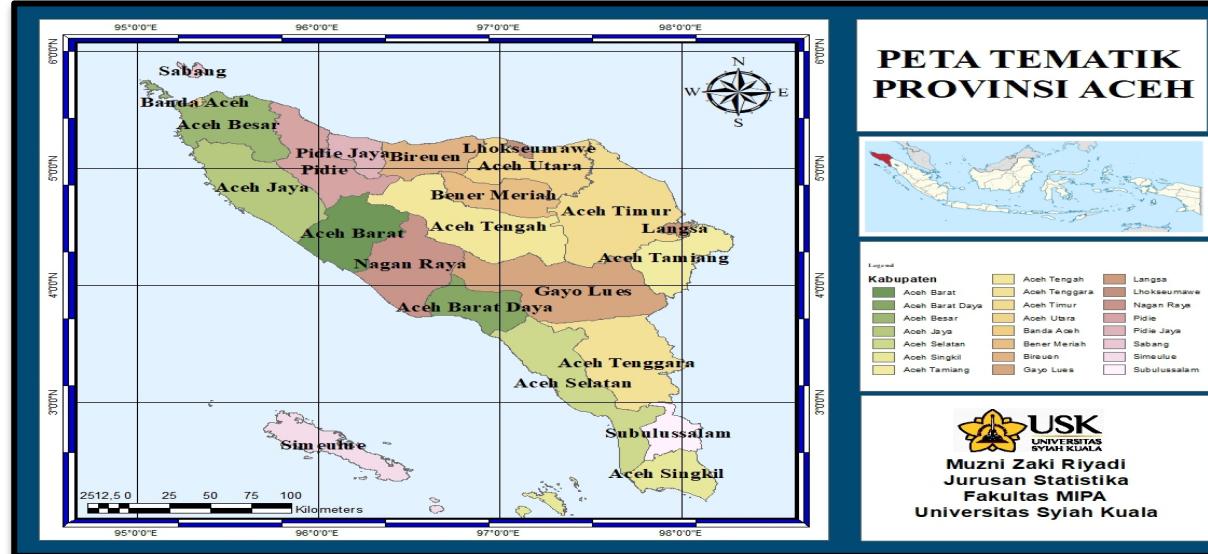
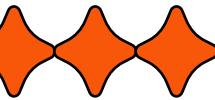
Proyek ini merupakan bagian dari skripsi mahasiswa S2 Teknik Sipil, di mana saya yang melakukan analisis data menggunakan *software AMOS* dengan pendekatan *Structural Equation Modeling* (SEM). Penelitian bertujuan untuk melihat hubungan antara variabel *Boredom Proneness* (BP), *Epworth Sleepiness Scale* (ESS), dan *Microsleep* (MS) dalam sebuah model konseptual serta mengevaluasi kesesuaian model tersebut.



## AMOS

Proyek ini merupakan bagian dari skripsi saya, di mana saya menggunakan *software AMOS* untuk melakukan analisis statistik berbasis *Structural Equation Modeling* (SEM). Penelitian ini bertujuan untuk menganalisis faktor-faktor yang memengaruhi kepuasan dan loyalitas nasabah BSI.

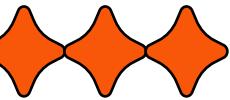
# Analisis Statistik & Data



## QGIS & ArcGIS

Pada proyek ini, saya menganalisis data geospasial menggunakan perangkat lunak QGIS dan ArcGIS. Analisis tersebut bertujuan untuk memvisualisasikan peta tematik Provinsi Aceh serta menyusun peta sebaran titik api berdasarkan radius kemungkinan

# Analisis Statistik & Data



## Microsoft Excel

Pada proyek ini, saya menggunakan Microsoft Excel untuk menganalisis *dataset* dengan menerapkan formula seperti INDEX MATCH, VLOOKUP, dan HLOOKUP. Formula ini digunakan untuk pencarian data secara dinamis dan efisien.

MySkill

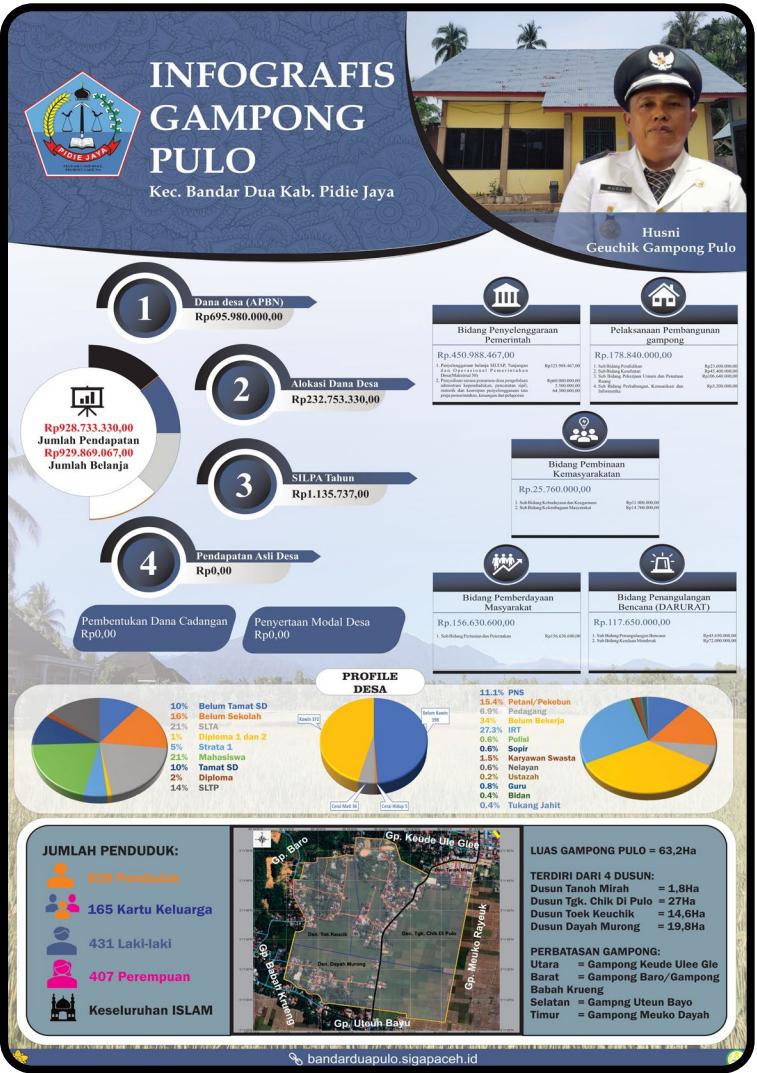
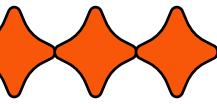
## Mini Task Index Match, VLOOKUP, HLOOKUP

sft_no	created_at	product_id	product_name	item_id	item_name	route_id	route_name	weight	price	total_price	insurance_type
LP1	2022-01-01	2	REGULAR	56	DOKUMEN	20	JAKARTA-SOLO	1,0	10000	10000	BASIC
LP2	2022-01-01	2	REGULAR	55	SEPATU	13	JAKARTA-BOGOR	1,0	7000	7000	BASIC
LP3	2022-01-01	2	REGULAR	57	BARANG ELEKTRONIK	24	JAKARTA-MADURA	3,0	14000	42000	PREMIUM
LP4	2022-01-01	1	EXPRESS	53	PAKAIAN	24	JAKARTA-MADURA	1,0	14000	14000	FREE
LP5	2022-01-01	2	REGULAR	57	BARANG ELEKTRONIK	19	JAKARTA-SALATIGA	3,0	9500	28500	BASIC
LP6	2022-01-01	1	EXPRESS	53	PAKAIAN	16	JAKARTA-GARUT	1,0	8500	8500	FREE
LP7	2022-01-01	1	EXPRESS	55	SEPATU	12	JAKARTA-TANGERANG	1,0	7000	7000	BASIC
LP8	2022-01-01	1	EXPRESS	55	SEPATU	16	JAKARTA-GARUT	1,0	8500	8500	FREE
LP9	2022-01-01	2	REGULAR	53	PAKAIAN	15	JAKARTA-BANDUNG	1,0	8000	8000	PREMIUM
LP10	2022-01-01	2	REGULAR	57	BARANG ELEKTRONIK	21	JAKARTA-JOGJAKARTA	3,0	9750	29250	BASIC
LP11	2022-01-01	1	EXPRESS	56	DOKUMEN	21	JAKARTA-JOGJAKARTA	1,0	9750	9750	FREE
LP12	2022-01-01	1	EXPRESS	54	TAS	21	JAKARTA-JOGJAKARTA	1,5	9750	14625	BASIC
LP13	2022-01-01	1	EXPRESS	55	SEPATU	22	JAKARTA-MALANG	1,0	11000	11000	PREMIUM
LP14	2022-01-01	1	EXPRESS	54	TAS	25	JAKARTA-LAMPUNG	1,5	15000	22500	FREE
LP15	2022-01-01	1	EXPRESS	53	PAKAIAN	25	JAKARTA-LAMPUNG	1,0	15000	15000	PREMIUM
LP16	2022-01-01	2	REGULAR	56	DOKUMEN	20	JAKARTA-SOLO	1,0	10000	10000	BASIC
LP17	2022-01-01	2	REGULAR	53	PAKAIAN	19	JAKARTA-SALATIGA	1,0	9500	9500	FREE
LP18	2022-01-01	1	EXPRESS	53	PAKAIAN	19	JAKARTA-SALATIGA	1,0	9500	9500	BASIC
LP19	2022-01-01	2	REGULAR	54	TAS	23	JAKARTA-SURABAYA	1,5	12000	18000	PREMIUM

#RintikKarirImpian

MySkill

# Desain



**DOKUMENTASI PENGADAAN BARANG DAN KEGIATAN COVID TAHUN 2016 S/D 2022 GAMPOONG PULO KEC. BANDAR DUA KAB. PIDIE JAYA**

**PENGADAAN BARANG**

Bidang Penyelegaran Pemerintah  
Rp 450.988.467,00  
1. Pengadaan Operasional Pemerintahan  
2. Bidang Kesehatan  
3. Bidang Pendidikan, Kebudayaan, dan Olahraga  
4. Bidang Pertahanan, Komunikasi dan Infrastruktur

Pelaksanaan Pembangunan gampong  
Rp 178.840.000,00  
1. Bidang Pembangunan dan Operasional Pemerintahan  
2. Bidang Kesehatan  
3. Bidang Pendidikan, Kebudayaan, dan Olahraga  
4. Bidang Pertahanan, Komunikasi dan Infrastruktur

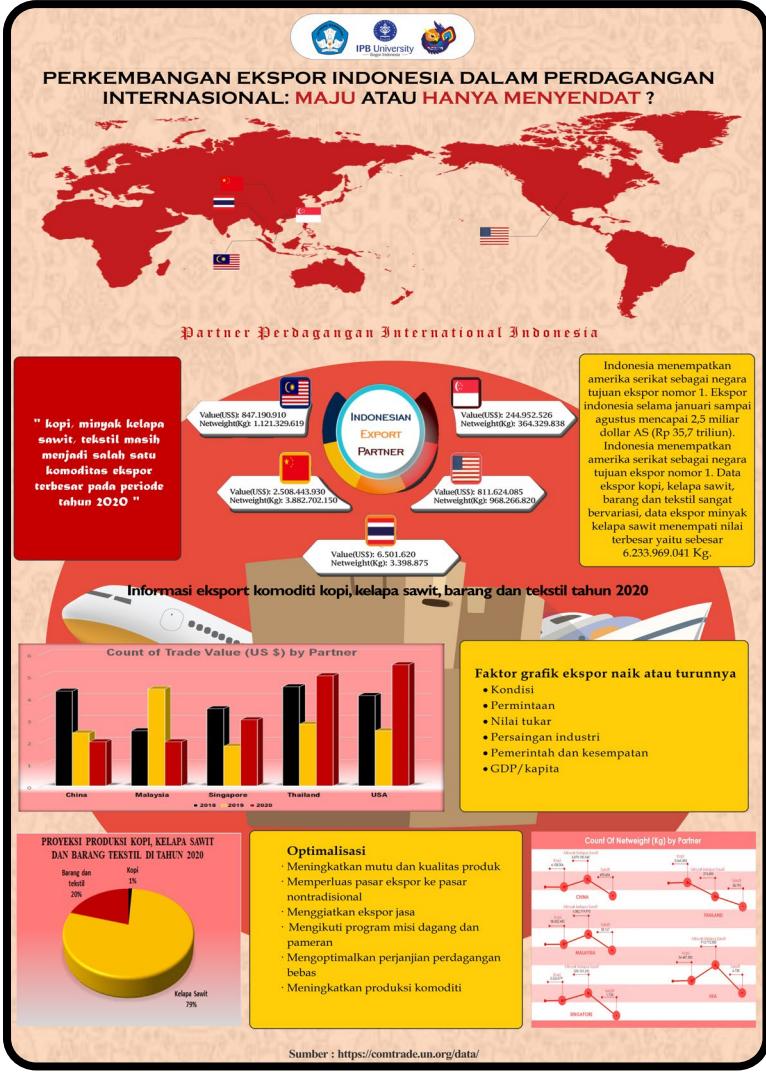
Bidang Binaan Kemasyarakatan  
Rp 25.760.000,00  
1. Bidang Kesehatan dan Kependidikan  
2. Bidang Kebudayaan dan Olahraga

**KEGIATAN SELAMA COVID**

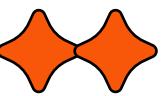
Dokumentasi Kegiatan Fisik Dari Tahun 2016 s/d. 2018  
Gampong Pulo  
Kecamatan Bandar Dua Kabupaten Pidie Jaya

**Pembukaan Jalan Baru, Rabat Beton, Pembuatan Saluran Air dan Sumur Bor Umum**

Tahun 2017 Sumber Dana DDS  
Tahun 2018 Sumber Dana DDS  
Tahun 2019 Sumber Dana DDS  
Tahun 2020 Sumber Dana DDS  
Tahun 2021 Sumber Dana DDS



# Let's Collaboration With Me !



+62-8225-8157-135

muznizakiriyadi@gmail.com

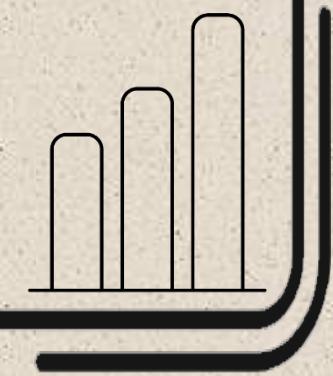
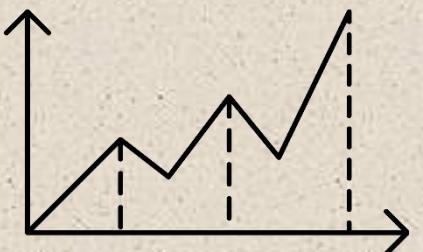
Muzni Zaki Riyadi

@muzni.riyadi

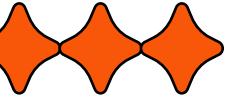
# PROJECT

MySkill

SQL



# Project SQL \*



## Introduction to SQL

SQL (Structured Query Language) merupakan bahasa yang digunakan untuk mengelola serta mengakses basis data relasional. SQL sendiri digunakan dalam melakukan berbagai tugas seperti membuat, memodifikasi serta menghapus struktur tabel dan skema data, mengambil, memperbarui, dan menghapus data di dalam sebuah basis data.

SQL memiliki fungsi sebagai interface antara pengguna dan sistem manajemen basis data (DBMS). Melalui perintah (Command) SQL, pengguna dapat berinteraksi dengan DBMS untuk membuat, mengolah dan mengambil informasi dari basis data. SQL juga merupakan alat penting pada sebuah program database serta analisis data, dan juga digunakan di berbagai platform maupun sistem manajemen basis data seperti MySQL, Microsoft SQL Server, PostgreSQL, Oracle dan lainnya.

Tools yang digunakan:



BigQuery

# Project SQL \*

## Overview

Dataset bermanfaat untuk memperoleh wawasan mandalam tentang pola perilaku pelanggan, efisiensi pelanggan, serta mengoptimalkan strategi penjualan produk.

Berikut dataset yang akan digunakan:

1. order\_detail
2. sku\_detail
3. customer\_detail
4. payment\_detail

Data yang digunakan adalah data yang berasal dari Tokopedia (bukan data sesungguhnya). Mengenai penjelasan dataset adalah sebagai berikut.

### **order\_detail:**

1. id → angka unik dari order / *id\_order*
2. customer\_id → angka unik dari pelanggan
3. 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
6. qty\_ordered → jumlah barang yang dibeli oleh pelanggan
7. 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

### **sku\_detail:**

1. id → angka unik dari produk (dapat digunakan untuk *key* saat join)
2. sku\_name → nama dari produk
3. 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

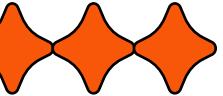
### **customer\_detail:**

1. id → angka unik dari pelanggan
2. registered\_date → tanggal pelanggan mulai mendaftarkan diri sebagai anggota

### **Payment\_detail:**

1. id → angka unik dari metode pembayaran
2. payment\_method → metode pembayaran yang digunakan

# Project SQL \*



## Dataset

Sehingga, dataset memiliki skema sebagai berikut:

- Tabel `customer_detail` berisi ID dan tanggal registrasi.
- Tabel `order_detail` menyimpan data transaksi pesanan, menghubungkan pelanggan dengan produk yang dibeli, serta detail harga dan jumlah pesanan.
- Tabel `sku_detail` mencakup informasi tentang produk yang dijual, termasuk nama produk, harga dasar dan kategori.
- Tabel `payment_detail` mencatat metode pembayaran yang digunakan dalam transaksi.

Data dari `order_detail` berfungsi sebagai data utama (primary\_key) yang terhubung dengan data lainnya, seperti `customer_detail`, `sku_detail`, dan `payment_detail`.

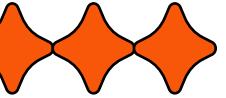
## Skema Tabel

order_detail
id
customer_id
order_date
sku_id
price
qty_ordered
before_discount
discount_amount
after_discount
is_gross
is_valid
is_net
payment_id

customer_detail
id
registered_date

sku_detail
id
sku_name
base_price
cogs
category

payment_detail
id
payment_method

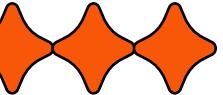


## Question

1. Selama transaksi yang terjadi selama 2021, pada bulan apa total nilai transaksi (after\_discount) paling besar? Gunakan is\_valid = 1 untuk memfilter data transaksi. Source table: order\_detail.
2. Selama transaksi pada tahun 2022, kategori apa yang menghasilkan nilai transaksi paling besar? Gunakan is\_valid = 1 untuk memfilter nilai transaksi. Source table: order\_detail, sku\_detail.
3. Bandingkan nilai transaksi dari masing-masing kategori pada tahun 2021 dengan 2022. sebutkan kategori apa saja yang sedang mengalami peningkatan dan kategori apa yang mengalami penurunan nilai transaksi dari tahun 2021 ke 2022. Gunakan is\_valid = 1 untuk memfilter data transaksi. Source table: order\_detail, sku\_detail.
4. Tampilkan top 5 metode pembayaran yang paling popular digunakan selama 2022 (berdasarkan total unique order). Gunakan is\_valid = 1 untuk memfilter data transaksi. Source table: order\_detail, payment\_method.
5. Urutkan dari ke-5 produk ini berdasarkan nilai transaksinya.
  - Samsung
  - Apple
  - Sony
  - Huawei
  - Lenovo

Gunakan is\_valid = 1 untuk memfilter data transaksi. Source table: order\_detail, sku\_detail.

# Project SQL \*



## Question 1

Selama transaksi yang terjadi selama 2021, pada bulan apa total nilai transaksi (after\_discount) paling besar?

Gunakan is\_valid = 1 untuk memfilter data transaksi. Source table: order\_detail.

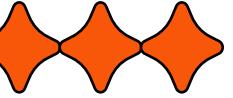
### Query

```
select
extract (year from order_date) as tahun,
extract (month from order_date) as bulan,
round (sum(after_discount), 1) as total
from `tokopedia.order_detail`
where is_valid = 1
and extract (year from order_date) = 2021
group by 1,2
order by 3 desc
limit 5
```

### Result

Query results

JOB INFORMATION		RESULTS	CHART	JSON
Row	tahun	bulan	total	
1	2021	8	227862744.0	
2	2021	12	217309963.0	
3	2021	10	207603259.9	
4	2021	11	180396009.7	
5	2021	7	148007735.0	



## Question 1

### Penjelasan Query

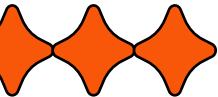
Query ini berfungsi untuk menghitung total transaksi bulanan sepanjang tahun 2021 dari tabel order\_detail. Penjelasannya sebagai berikut.

- Extract : Mengambil tahun dari tanggal order\_date.
- Round (Sum (after\_discount),1) : Membulatkan hasil penjumlahan nilai transaksi pada kolom after\_discount ke 1 digit di belakang koma.
- is\_valid = 1 artinya transaksi tersebut dianggap valid atau sah, dengan kata lain pembayaran sudah berhasil dilakukan untuk pemesanan tersebut.
- Group by 1,2 : Mengelompokkan tahun dan bulan pada table order\_date.
- Order by 3 desc : Mengurutkan hasil berdasarkan total nilai transaksi terbesar.
- Limit 5 : Membatasi hanya 5 baris hasil yang ditampilkan.

### Insight

Data yang diperoleh yaitu bulan ke-8 (Agustus) tahun 2021 yang memiliki total transaksi tertinggi sebesar 227.862.744.

# Project SQL \*



## Question 2

Selama transaksi pada tahun 2022, kategori apa yang menghasilkan nilai transaksi paling besar?

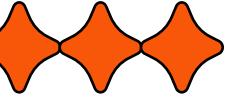
Gunakan is\_valid = 1 untuk memfilter nilai transaksi. Source table: order\_detail, sku\_detail.

### Query

```
select
extract (year from a.order_date) as tahun,
b.category,
sum (a.after_discount) as transaksi_terbesar,
from `tokopedia.order_detail` as a
left join `tokopedia.sku_detail` as b on a.sku_id = b.id
where extract (year from a.order_date) = 2022 and is_valid = 1
group by 1,2
order by 3 desc
limit 5
```

### Result

Query results				
JOB INFORMATION		RESULTS		CHART
Row	tahun	category	transaksi_terbesar	
1	2022	Mobiles & Tablets	918451576.0	
2	2022	Entertainment	365344148.9	
3	2022	Appliances	316358100.0	
4	2022	Computing	214028543.4	
5	2022	Men Fashion	135588253.0	



## Question 2

### Penjelasan Query

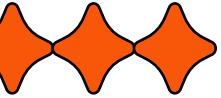
Query ini berfungsi untuk menghitung total transaksi bulanan sepanjang tahun 2021 dari tabel order\_detail. Penjelasannya sebagai berikut.

- **Select extract (year from a.order\_date) as tahun, b.category,** : Memilih tahun pada kolom order\_date pada tabel a dan kolom category pada tabel b, dan memberikan alias sebagai "tahun".
- **Sum (a.after\_discount) as transaksi\_terbesar** : Menghitung jumlah total dari kolom after\_discount dan memberi alias sebagai "transaksi terbesar".
- **Left join** : Menggabungkan tabel order\_detail dengan sku\_detail (dialiaskan sebagai b) melalui kolom sku\_id dari order\_detail dan id dari sku\_detail.
- **Where extract (year from a.order\_date) = 2022** : Memfilter data berdasarkan tahun 2022 pada table order\_date.
- **Is\_valid = 1** artinya transaksi tersebut dianggap valid atau sah, dengan kata lain pembayaran sudah berhasil dilakukan untuk pemesanan tersebut.
- **Limit 5** : Membatasi hanya 5 baris hasil yang ditampilkan.

### Insight

Berdasarkan data yang diperoleh, kategori produk Mobiles & Tablets merupakan nilai transaksi tertinggi sepanjang tahun 2022 dengan total nilai transaksi mencapai 918.451.576.

# Project SQL \*



## Question 3

Bandingkan nilai transaksi dari masing-masing kategori pada tahun 2021 dengan 2022. sebutkan kategori apa saja yang sedang mengalami peningkatan dan kategori apa yang mengalami penurunan nilai transaksi dari tahun 2021 ke 2022.

Gunakan is\_valid = 1 untuk memfilter data transaksi. Source table: order\_detail, sku\_detail.

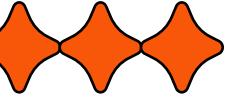
### Query

```
with trans2021
as
(select
extract(year from order_date) as tahun,
tod.category,
round(sum(after_discount),2) as transaksi_2021
from
tokopedia_order_detail as tod
left join tokopedia_sku_detail as tod on tod.sku_id = tod.id
where
extract(year from order_date) = 2021
and tod.is_valid = 1
group by 1,2
),
trans2022
as
(select
extract(year from order_date) as tahun,
tod.category,
round(sum(after_discount),2) as transaksi_2022
from
tokopedia_order_detail as tod
left join tokopedia_sku_detail as tod on tod.sku_id = tod.id
where
extract(year from order_date) = 2022
and tod.is_valid = 1
group by 1,2
)

select
coalesce(t1.category, t2.category) as category,
coalesce(t1.transaksi_2021,0) as transaksi_2021,
coalesce(t2.transaksi_2022,0) as transaksi_2022,
coalesce(round(coalesce(t2.transaksi_2022,0)-coalesce(t1.transaksi_2021,0),2),0) as selisih_transaksi,
case
when coalesce(t2.transaksi_2022, 0) > coalesce(t1.transaksi_2021,0) then 'Peningkatan'
when coalesce(t2.transaksi_2022, 0) < coalesce(t1.transaksi_2021,0) then 'Penurunan'
else 'Stabil'
end as status_perubahan
from
trans2021 as t1
full outer join
trans2022 as t2 on t1.category = t2.category
```

### Result

category	transaksi_2021	transaksi_2022	selisih_transaksi	status_perubahan
Appliances	218550177.0	316358100.0	97807923.0	Peningkatan
Beauty & Grooming	46047360.0	46211019.18	163659.18	Peningkatan
Books	10124596.0	6792519.2	-3332076.8	Penurunan
Computing	172878860.0	214028543.4	41149683.4	Peningkatan
Entertainment	162326357.4	365344148.9	203017791.5	Peningkatan
Health & Sports	33837965.6	54235579.6	20397614.0	Peningkatan
Home & Living	45797873.0	79483716.2	33685843.2	Peningkatan
Kids & Baby	23971057.8	25931276.84	1960219.04	Peningkatan



## Question 3

### Penjelasan Query

Query ini bertujuan untuk membandingkan nilai transaksi berdasarkan kategori antara tahun 2021 dan 2022.

#### 1. Pengambilan Data

- Menggunakan tabel order\_detail dan sku\_detail untuk mendapatkan total nilai transaksi setelah diskon (after\_discount).
- Data disaring dengan is\_valid = 1 agar hanya transaksi valid yang dihitung.
- Transaksi dari tahun 2021 dan 2022 diekstrak secara terpisah dan kemudian digabungkan.

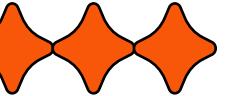
#### 2. Pengolahan Data

- Menggunakan COALESCE untuk menangani kategori yang mungkin hanya muncul di salah satu tahun.
- Status perubahan transaksi (Peningkatan atau Penurunan) ditentukan dengan perbandingan nilai transaksi antara tahun 2021 dan 2022.

### Insight

Kategori produk yang mengalami penurunan nilai transaksi pada tahun 2022 adalah Books. Sedangkan kategori lainnya mengalami peningkatan seperti Appliances, Beauty & Grooming, Computing dan produk lainnya seperti pada tabel.

# Project SQL \*



## Question 4

Tampilkan top 5 metode pembayaran yang paling popular digunakan selama 2022 (berdasarkan total unique order).

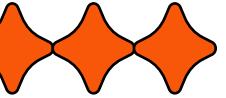
Gunakan is\_valid = 1 untuk memfilter data transaksi. Source table: order\_detail, payment\_method.

### Query

```
select
    extract(year from order_date) as tahun,
    tpd.payment_method,
    count(distinct tod.customer_id) as total_transaksi
from
    `Tokopedia.order_detail` as tod
    left join `Tokopedia.payment_detail`as tpd on tod.payment_id = tpd.id
where
    extract(year from order_date) = 2022
    and
    tod.is_valid =
group by 1,2
order by 3 desc
limit 5
```

### Result

JOB INFORMATION		RESULTS	CHART	JSON	EXECUT
Row	tahun	payment_method		total_transaksi	
1	2022	cod		1385	
2	2022	Payaxis		147	
3	2022	customercredit		66	
4	2022	Easypay		63	
5	2022	jazzwallet		25	



## Question 4

### Penjelasan Query

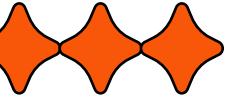
Query ini berfungsi untuk menghitung jumlah metode pembayaran yang paling popular digunakan selama tahun 2022 dari tabel order\_detail dan payment\_detail yang telah digabungkan. Penjelasannya sebagai berikut.

- **Count (distinct o.consumer\_id) as jumlah\_transaksi** : Menghitung jumlah transaksi unik dari kolom consumer\_id pada table order\_detail. Fungsi DISTINCT digunakan untuk memastikan hanya transaksi unik yang dihitung, menghindari duplikasi, yang diberi alias "jumlah\_transaksi".
- **Left join `tokopedia.payment\_detail` as p on o.payment\_id = p.id**: Menggabungkan tabel order\_detail dengan payment\_detail melalui kolom payment.id dari order\_detail yang dihubungkan dengan id dari payment\_detail.
- **is\_valid = 1** artinya transaksi tersebut dianggap valid atau sah, dengan kata lain pembayaran sudah berhasil dilakukan untuk pemesanan tersebut.
- **Group by 1** : Mengelompokkan hasil berdasarkan metode pembayaran yang dipilih pada kolom pertama (p.payment\_method).
- **Order by 2 desc** : Mengurutkan hasil berdasarkan jumlah\_transaksi dari yang terbesar hingga terkecil.
- **Limit 5** : Membatasi hasil query untuk menampilkan 5 metode pembayaran terpopuler berdasarkan total transaksi unik.

### Insight

Metode pembayaran yang paling sering digunakan oleh konsumen selama tahun 2022 adalah COD, Payaxis, Customer Credit, Easypay dan Jazzwallet.

# Project SQL \*



## Question 5

Urutkan dari ke-5 produk ini berdasarkan nilai transaksinya: 1.Samsung, 2. Apple, 3.Sony, 4.Huawei, 5.Lenovo.

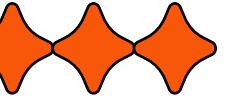
Gunakan is\_valid = 1 untuk memfilter data transaksi. Source table: order\_detail, sku\_detail.

### Query

```
with barang as(
  select
    case
      when lower (b.sku_name) like '%samsung%' then 'Samsung'
      when lower (b.sku_name) like '%apple%' then 'Apple'
      when lower (b.sku_name) like '%sony%' then 'Sony'
      when lower (b.sku_name) like '%huawei%' then 'Huawei'
      when lower (b.sku_name) like '%lenovo%' then 'Lenovo'
    end as Merek,
    sum (a.after_discount) as nilai_transaksi
  from
    `tokopedia.order_detail` as a
  left join
    `tokopedia.sku_detail` as b on a.sku_id = b.id
  where
    a.is_valid = 1
    and extract (year from order_date) = 2022
    group by 1
    order by 2 desc
)
select * from barang
where Merek is not null
```

### Result

JOB INFORMATION		RESULTS	CHART	JSO
Row	Merek	nilai_transaksi		
1	Samsung	412357844.0		
2	Sony	32343178.0		
3	Huawei	30942826.0		
4	Apple	24447000.0		
5	Lenovo	23621590.4		

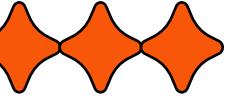


## Question 5

### Penjelasan Query

Query ini berfungsi untuk mengelompokkan barang/produk berdasarkan merek dan menghitung total nilai transaksi untuk masing-masing merek dan hasilnya diurutkan berdasarkan nilai transaksi. Penjelasannya sebagai berikut.

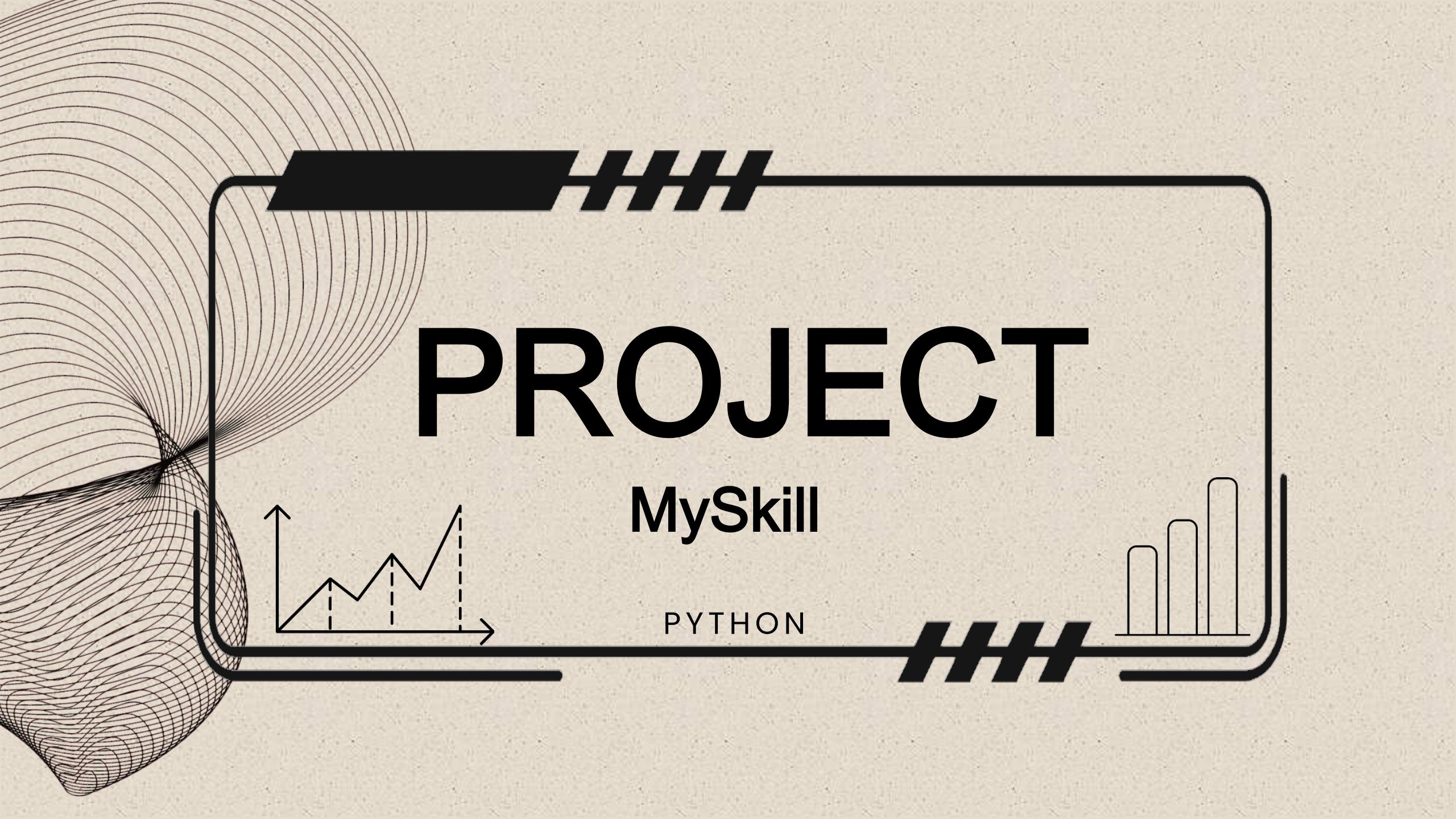
- **With barang as (...)** : Mendefinisikan CTE (Common Table Expression) yang dinamakan barang, yang akan menyimpan hasil perhitungan dari query didalamnya.
- **Case when lower (b.sku\_name) like `%samsung%` (...)** : Menggunakan struktur case untuk mengelompokkan barang/produk kedalam merek tertentu berdasarkan nama barang/produk (sku\_name). Contohnya, jika nama barang/produk mengandung kata "Samsung", maka akan diberi alias Merek sebagai "Samsung".
- **Sum (a.after\_discount) as nilai\_transaksi** : Menghitung total nilai transaksi dari kolom after\_discount dan memberi alias "nilai\_transaksi".
- **From `tokopedia.order\_detail` as a** : Memilih tabel order\_detail sebagai sumber data utama dan memberi alias sebagai "a"
- **Left join `okopedia.sku\_detail` as b on a.sku\_id = b.id** : Menggabungkan tabel order\_detail dengan sku\_detail melalui kolom sku\_id dari order\_detail dan kolom id dari sku\_detail.
- **Extract (year from order\_date) = 2022** : Memilih tahun 2022 pada table order\_date.
- **Group by 1** : Mengelompokkan hasil berdasarkan Merek yang didefinisikan di dalam case.
- **Order by 2 desc** : Mengurutkan hasil berdasarkan nilai\_transaksi dari yang terbesar hingga terkecil.
- **Select \* from barang** : Mengambil semua hasil dari CTE barang/produk yang telah didefinisikan.
- **Where Merek is not null** : Memfilter hasil untuk hanya menyertakan merek yang valid (tidak bernilai null).



## Question 5

### Insight

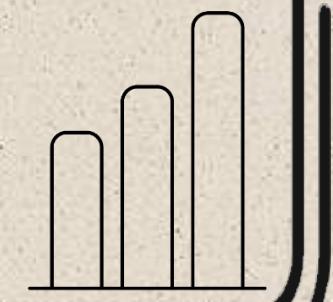
- Samsung menduduki peringkat pertama dengan total nilai transaksi tertinggi, diikuti oleh Apple dan Sony.
- Huawei dan Lenovo masih berada dalam lima besar, tetapi dengan selisih nilai transaksi yang lebih rendah dibanding merek lainnya.
- Ini mengindikasikan bahwa Samsung dan Apple memiliki daya tarik paling kuat di pasar, kemungkinan karena branding, inovasi, dan strategi pemasaran yang efektif.



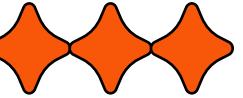
# PROJECT

MySkill

PYTHON



# Project Python \*



## Introduction to Python

Python adalah bahasa pemrograman tingkat tinggi yang dapat digunakan untuk berbagai macam keperluan, seperti pengembangan website, data analytics, dan pengembangan perangkat lunak.

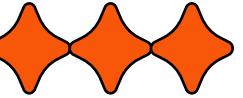
Salah satu cara untuk menggunakan python adalah dengan bantuan google collab yang menjalankan fungsi dari python. Library yang digunakan antara lainnya:

- Pandas untuk memanipulasi data
- Numpy untuk menjalankan fungsi matematika
- Matplotlib dan Seaborn untuk memvisualisasikan data

Tools yang digunakan:



# Project Python \*



## Overview

Dataset bermanfaat untuk memperoleh wawasan mendalam tentang pola perilaku pelanggan, efisiensi pelanggan, serta mengoptimalkan strategi penjualan produk.

Berikut dataset yang akan digunakan:

1. order\_detail
2. sku\_detail
3. customer\_detail
4. payment\_detail

Data yang digunakan adalah data yang berasal dari Tokopedia (bukan data sesungguhnya). Mengenai penjelasan dataset adalah sebagai berikut.

### **order\_detail:**

1. id → angka unik dari order / *id\_order*
2. customer\_id → angka unik dari pelanggan
3. 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
6. qty\_ordered → jumlah barang yang dibeli oleh pelanggan
7. 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

### **sku\_detail:**

1. id → angka unik dari produk (dapat digunakan untuk *key* saat join)
2. sku\_name → nama dari produk
3. 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

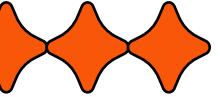
### **customer\_detail:**

1. id → angka unik dari pelanggan
2. registered\_date → tanggal pelanggan mulai mendaftarkan diri sebagai anggota

### **Payment\_detail:**

1. id → angka unik dari metode pembayaran
2. payment\_method → metode pembayaran yang digunakan

# Project Python \*



## Dataset

Sehingga, dataset memiliki skema sebagai berikut:

- Tabel `customer_detail` berisi ID dan tanggal registrasi.
- Tabel `order_detail` menyimpan data transaksi pesanan, menghubungkan pelanggan dengan produk yang dibeli, serta detail harga dan jumlah pesanan.
- Tabel `sku_detail` mencakup informasi tentang produk yang dijual, termasuk nama produk, harga dasar dan kategori.
- Tabel `payment_detail` mencatat metode pembayaran yang digunakan dalam transaksi.

Data dari `order_detail` berfungsi sebagai data utama (primary\_key) yang terhubung dengan data lainnya, seperti `customer_detail`, `sku_detail`, dan `payment_detail`.

## Skema Tabel

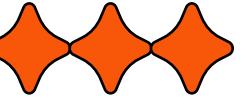
order_detail
id
customer_id
order_date
sku_id
price
qty_ordered
before_discount
discount_amount
after_discount
is_gross
is_valid
is_net
payment_id

customer_detail
id
registered_date

sku_detail
id
sku_name
base_price
cogs
category

payment_detail
id
payment_method

# Project Python \*



## Dataset Preparation

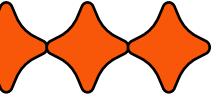
Menggunakan Import library Pandas, Numpy, Matplotlib, and Seaborn into Gogle Colab

```
✓ [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 data kedalam Google Colab

```
✓ [2] #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)
```

# Project Python \*



## Dataset Preparation

Menampilkan data yang telah di import untuk dilakukan pengecekan

```
#Mengampilkan 5 baris pertama
df_od.head()
```

	id	customer_id	order_date	sku_id	price	qty_ordered	before_discount	discount_amount	after_discount	is_gross	is_valid	is_net	payment_id
0	ODR9939707760w	C713589L	2021-11-19	P858068	26100	200	5220000.0	2610000.00	2610000.00	1	1	0	5
1	ODR7448356649d	C551551L	2021-11-19	P886455	1971942	5	9859710.0	2464927.50	7394782.50	1	0	0	5
2	ODR4011281866z	C685596L	2021-11-25	P678648	7482000	1	7482000.0	2065344.62	5416655.38	1	0	0	4
3	ODR3378927994s	C830683L	2021-11-22	P540013	3593680	1	3593680.0	1455440.40	2138239.60	1	1	1	5
4	ODR4904430099k	C191766L	2021-11-21	P491032	4413220	1	4413220.0	1059172.80	3354047.20	1	1	1	4

Import data kedalam Google Colab

```
#Mengampilkan 5 baris pertama
df_pd.head()
```

```
#Mengampilkan 5 baris pertama
df_cd.head()
```

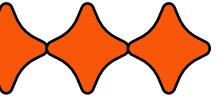
```
#Mengampilkan 5 baris pertama
df_sd.head()
```

	id	payment_method
0	1	cod
1	2	jazzvoucher
2	3	customercredit
3	4	Payaxis
4	5	jazzwallet

	id	registered_date
0	C996508L	2021-07-10
1	C180415L	2021-07-18
2	C535451L	2021-07-23
3	C177843L	2021-07-12
4	C951682L	2021-07-27

	id	sku_name	base_price	cogs	category
0	P798444	AT-FSM-35	57631.70	46052	Kids & Baby
1	P938347	AYS_Haier-18HNF	3931789.26	3499256	Appliances
2	P826364	Atalian_DV206A-Brown-41	324597.00	243426	Men Fashion
3	P467533	Darul_Sakoon_Food_Bundle	2870.42	2378	Superstore
4	P229955	HP_15AY-15-Ay072NIA-ci3	2265625.00	1631250	Computing

# Project Python \*



## Dataset Preparation

Melakukan pengecekan tipe data setiap kolom dan mengubah beberapa tipe kolom untuk memudahkan

```
#Menampikan tipe data tiap kolom
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
is_gross	int64
is_valid	int64
is_net	int64
payment_id	int64
payment_method	object
sku_name	object
base_price	float64
cogs	int64
category	object
registered_date	object
dtype:	object

```
#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
```

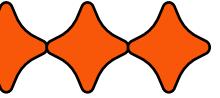
id	object
customer_id	object
order_date	object
sku_id	object
price	int64
qty_ordered	int64
before_discount	int64
discount_amount	int64
after_discount	int64
is_gross	int64
is_valid	int64
is_net	int64
payment_id	int64
payment_method	object
sku_name	object
base_price	int64
cogs	int64
category	object
registered_date	object
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
order_date	datetime64[ns]
sku_id	object
price	int64
qty_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
category	object
registered_date	datetime64[ns]

# Project Python \*



## Question 1

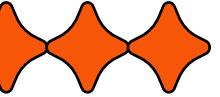
Dear Data Analyst,

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

# Project Python \*



## Question 1

### Syntax

```
# No 1

# Menggabungkan tabel df_od dan df_sd
df_mg = pd.merge(df_od, df_sd, left_on='sku_id', right_on='id', how='left')

# order_datanya diubah type datanya jadi datetime
df_mg['order_date'] = pd.to_datetime(df_mg['order_date'])

# Filter Datanya
df_filter = df_mg[(df_mg['category'] == 'Mobiles & Tablets') & (df_mg['order_date'].dt.year == 2022)]
df_filter = df_filter[df_filter['is_valid'] == 1]

# Melakukan groupby pada df_sd berdasarkan kolom "category" dan menjumlahkan "after_discount"
df_groupby = df_filter.groupby(by=['sku_id', 'sku_name', 'category'], as_index=False)[['qty_ordered']].sum()
df_top5 = df_groupby.sort_values(by = 'qty_ordered', ascending = False)

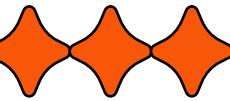
# Menampilkan top 5 produk
df_top5.head()
```

```
## Menampilkan top 5 produk
df_top5.head()
df_top5 = df_top5.head()

df_top5.sort_values(
    by = ['qty_ordered'],
    ascending = True,
    inplace= True)

##Menampilkan grafik
df_top5.plot(x='sku_name',
              y=[ 'qty_ordered'],
              kind='barh',
              grid =True,
              xlabel = 'Quantity',
              ylabel = 'Product Name',
              figsize =(12,7),
              rot=0,
              title = 'TOP 5 Product',
              table = False,
              secondary_y = False
)
```

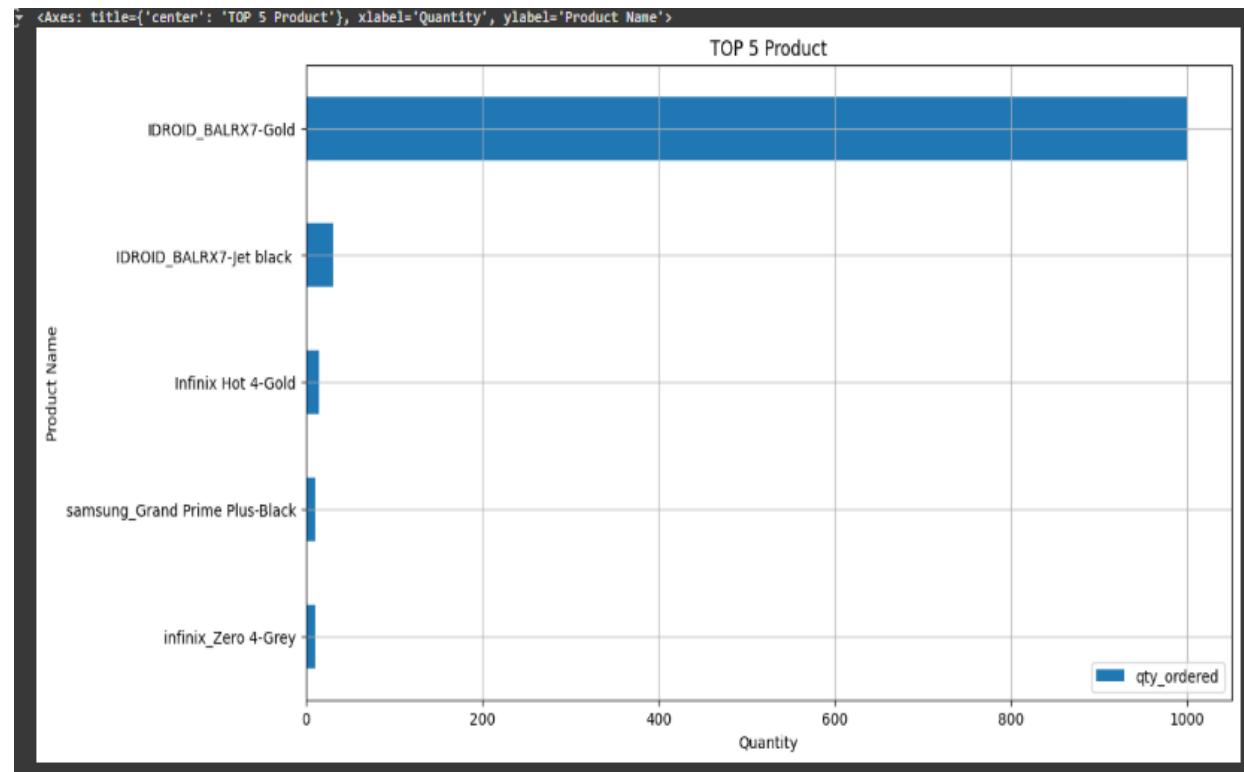
# Project Python \*

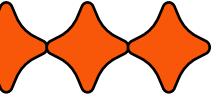


## Question 1

### Result

sku_id	sku_name	category	qty_ordered	
11	P313303	IDROID_BALRX7-Gold	Mobiles & Tablets	1000
36	P754831	IDROID_BALRX7-Jet black	Mobiles & Tablets	31
37	P774890	Infinix Hot 4-Gold	Mobiles & Tablets	15
31	P661542	samsung_Grand Prime Plus-Black	Mobiles & Tablets	11
43	P926387	infinix_Zero 4-Grey	Mobiles & Tablets	10





## Question 1

### Insight

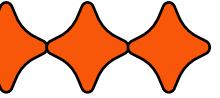
TOP 5 Produk dari Kategori Mobiles & Tablets selama tahun 2022 :

- |                                    |                |
|------------------------------------|----------------|
| 1. IDROID_BALRX7-Gold,             | kuantitas 1000 |
| 2. IDROID_BALRX7-Jet black,        | kuantitas 31   |
| 3. Infinix Hot 4_Gold              | kuantitas 15   |
| 4. Samsung_Grand Prime Plus-Black, | kuantitas 11   |
| 5. Infinix Zero 4-Grey             | kuantitas 10   |

### Analisis Lanjutan

Dominasi penjualan IDROID\_BBALRX7\_Gold kemungkinan besar disebabkan oleh fitur dan inovasi yang ditawarkannya. Berdasarkan data penjualan, produk dengan peringkat 4 atau 5 dapat digunakan sebagai hadiah untuk pemenang utama. Strategi ini bertujuan untuk mempromosikan produk secara gratis melalui pemenang kompetisi dengan meminta mereka melakukan review, yang diharapkan dapat meningkatkan penjualan produk lain di masa depan.

# Project Python \*



## Question 2

Dear Data Analyst,

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

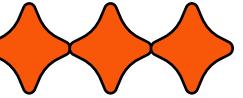
- Kami mohon bantuan untuk melakukan pengecekan data penjualan kategori tersebut dengan tahun 2021 secara kuantitas penjualan. Dugaan sementara kami, telah terjadi penurunan kuantitas penjualan pada 2022 dibandingkan 2021. (Mohon juga menampilkan data ke-15 kategori).
- Apabila memang terjadi penurunan 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

# Project Python \*



## Question 2

### Syntax

```
[ ] # No 2.1
# Menggabungkan tabel df_od dan df_sd
df_mg = pd.merge(df_od, df_sd, left_on='sku_id', right_on='id', how='left')

# ubah tipe data order_date menjadi datetime
df_mg['order_date'] = pd.to_datetime(df_mg['order_date'])

# Filter data 2021
df2021_filter = df_mg[(df_mg['order_date'].dt.year == 2021)]
df2021_filter = df2021_filter[df2021_filter['is_valid'] == 1]

# Filter data 2022
df2022_filter = df_mg[(df_mg['order_date'].dt.year == 2022)]
df2022_filter = df2022_filter[df2022_filter['is_valid'] == 1]

# Total kuantitas 2021
df2021_groupby = df2021_filter.groupby(by = ['category'], as_index=False)[['qty_ordered']].sum()
df2021_groupby = df2021_groupby.rename(columns = {'qty_ordered' : 'qty_2021'})

# Total kuantitas 2022
df2022_groupby = df2022_filter.groupby(by = ['category'], as_index=False)[['qty_ordered']].sum()
df2022_groupby = df2022_groupby.rename(columns = {'qty_ordered' : 'qty_2022'})

# Gabungkan tabel kuantitas 2021 & 2022
df_gabung = pd.merge(df2021_groupby, df2022_groupby, on='category', how='left')

# Hitung selisih kuantitas 2021 & 2022
df_gabung['qty_different'] = df_gabung['qty_2022'] - df_gabung['qty_2021']

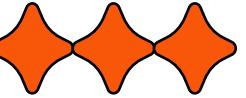
# Tambahkan kategori perubahan
df_gabung['status'] = np.where(df_gabung['qty_different'] > 0, 'Peningkatan',
                               np.where(df_gabung['qty_different']<0, 'Penurunan', 'Setara'))

# Hasil
df_15 = df_gabung.sort_values(by='qty_different', ascending=True)
df_15.head(15)
```

### Result

	category	qty_2021	qty_2022	qty_different	status
10	Others	426	263	-163	Penurunan
12	Soghaat	759	812	-147	Penurunan
8	Men Fashion	237	175	-62	Penurunan
1	Beauty & Grooming	168	153	-15	Penurunan
0	Appliances	124	148	24	Peningkatan
2	Books	171	195	24	Peningkatan
5	Health & Sports	173	200	27	Peningkatan
3	Computing	109	153	44	Peningkatan
11	School & Education	184	237	53	Peningkatan
6	Home & Living	193	250	57	Peningkatan
7	Kids & Baby	170	227	57	Peningkatan
4	Entertainment	77	150	73	Peningkatan
13	Superstore	327	536	209	Peningkatan
14	Women Fashion	140	489	349	Peningkatan
9	Mobiles & Tablets	107	1154	1047	Peningkatan

# Project Python \*



## Question 2

### Syntax

```
# 2.2
# Menggabungkan tabel df_od dan df_sd
df_mg = pd.merge(df_od, df_sd, left_on='sku_id', right_on='id', how='left')

# ubah tipe data order_date menjadi datetime
df_mg['order_date'] = pd.to_datetime(df_mg['order_date'])

# Filter data 2021
df2021_filter = df_mg[(df_mg['category'] == 'Others') & (df_mg['order_date'].dt.year == 2021)]
df2021_filter = df2021_filter[df2021_filter['is_valid'] == 1]

# Filter data 2022
df2022_filter = df_mg[(df_mg['category'] == 'Others') & (df_mg['order_date'].dt.year == 2022)]
df2022_filter = df2022_filter[df2022_filter['is_valid'] == 1]

# Total kuantitas 2021
df2021_groupby = df2021_filter.groupby(by = ['sku_id', 'sku_name'], as_index=False)[['qty_ordered']].sum()
df2021_groupby = df2021_groupby.rename(columns = {'qty_ordered' : 'qty_2021'})

# Total kuantitas 2022
df2022_groupby = df2022_filter.groupby(by = ['sku_id'], as_index=False)[['qty_ordered']].sum()
df2022_groupby = df2022_groupby.rename(columns = {'qty_ordered' : 'qty_2022'})

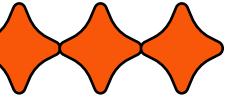
# Gabungkan tabel kuantitas 2021 & 2022
df_gabung = pd.merge(df2021_groupby, df2022_groupby, on='sku_id', how='outer').fillna(0)
# Hitung selisih kuantitas 2021 & 2022
df_gabung['qty_different'] = df_gabung['qty_2021'] - df_gabung['qty_2022']

df_20 = df_gabung.sort_values(by='qty_different', ascending=False)
df_20.head(20)
```

### Result

	sku_id	sku_name	qty_2021	qty_2022	qty_different
137	P858068	RB_Dettol Germ Busting Kit-bf	200.0	45.0	155.0
95	P664665	Dawlance _MD 10 + DWB 600	23.0	0.0	23.0
114	P738924	Telemall_MM-DR-HB-L	23.0	2.0	21.0
6	P139556	iu_Tickets General Enclosure-Islamabad	20.0	0.0	20.0
83	P564043	RS_Rehmat-e-Shereen Mix Mithai	13.0	0.0	13.0
62	P468366	kansai_NeverWet	10.0	1.0	9.0
75	P531621	sindbad_Sindbad Gift Card-3	7.0	0.0	7.0
153	P948769	emart_00-1	7.0	1.0	6.0
130	P830104	Am-PTV_ATS-004-M	5.0	0.0	5.0
37	P324617	Vouch 365 2016	5.0	0.0	5.0
44	P376314	sockeye_QG in Quarter Grey	4.0	0.0	4.0
98	P669089	The Vitamin Company Kojic Acid Whitening Cream...	4.0	0.0	4.0
93	P653370	duma_4561253300294	4.0	0.0	4.0
43	P373014	aw_Octane Booster-12oz./354ml	3.0	0.0	3.0
86	P587474	Trans2_LW 999	3.0	0.0	3.0
2	P121796	MEGUIAR_G12711	4.0	1.0	3.0
45	P390266	aw_Super Concentrated Fuel Injector (Black)-5....	2.0	0.0	2.0
154	P950153	aw_CRC DUST FREE -400ml	2.0	0.0	2.0
41	P356515	MEGUIAR_G13616	2.0	0.0	2.0
129	P819607	sstop_Universallensclipkit	2.0	0.0	2.0

# Project Python \*



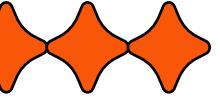
## Question 2

### Insight

Hasil dari 2.1 menunjukkan bahwa memang benar produk dengan kategori Others mengalami penurunan kuantitas penjualan pada 2022 sebesar 163 dibandingkan tahun 2021.

Hasil dari 2.2 menunjukkan Top 20 Produk dari kategori Others adalah sebagai berikut :

1. RB\_Dettol Germ Busting Kit-bf
2. Dawlance\_MD 10 + DWB 600
3. Telemall\_MM-DR-HB-L
4. iu\_Tickets General Enclosure-Islamabad20.0
5. RS\_Rehmat-e-Shereen Mix Mithai
6. kansai\_NeverWet
7. sindbad\_Sindbad Gift Card-3
8. emart\_00-1
9. Vouch 365 2016
10. Am-PTV\_ATS-004-M
11. Vouch 365 2016
12. sockeye\_QG in Quarter Grey
13. The Vitamin Company Kojic Acid Whitening Cream
14. duma\_4561253300294
15. MEGUIAR\_G12711
16. Trans2\_LW 999
17. aw\_Super Concentrated Fuel Injector (Black)
18. aw\_CRC DUST FREE -400ml
19. MEGUIAR\_G13616
20. sstop\_Universallensclipkit

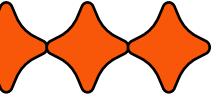


## Question 2

### Analisis Lanjutan

Berdasarkan data penjualan, terdapat beberapa produk dari kategori beauty dan kesehatan yang mengalami penurunan penjualan. Kemungkinan besar, hal ini disebabkan oleh munculnya produk serupa yang menawarkan promo dan packaging yang lebih menarik. Untuk meningkatkan penjualan, kita dapat melakukan inovasi pada formulasi dan packaging produk, serta melakukan promo atau penjualan bundle untuk beberapa jenis produk. Dengan demikian, kita dapat meningkatkan daya saing produk dan meningkatkan penjualan.

# Project Python \*



## Question 3

Dear Data Analyst,

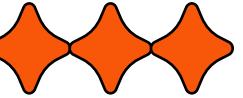
Terkait ulang tahun perusahaan pada 2 bulan mendatang, Tim Digital Marketing akan memberikan informasi promo bagi pelanggan pada akhir bulan ini. Kriteria pelanggan yang akan kami butuhkan adalah mereka yang sudah melakukan check-out namun belum melakukan pembayaran (`is_gross = 1`) selama tahun 2022.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

# Project Python \*



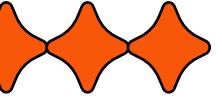
## Question 3

### Syntax

```
# 3  
# Menggabungkan Tabel df_od dan df_cd  
df_mg = pd.merge(df_od, df_cd, left_on='customer_id', right_on='id', how='left')  
  
# Mengubah type data register_date menjadi datetime  
df_mg['register_date'] = pd.to_datetime(df_mg['register_date'])  
  
# Filter data hanya tahun 2022 dan is_gross = 1, valid = 0, dan net = 0  
df_filter = df_mg[(df_mg['register_date'].dt.year == 2022)]  
df_filter = df_filter[df_filter['is_gross'] == 1]  
df_filter = df_filter[df_filter['is_valid'] == 0]  
df_filter = df_filter[df_filter['is_net'] == 0]  
  
# Menampilkan data yang dibutuhkan adalah ID Customer dan Registered Date  
dfneed_filter = df_filter[['customer_id','register_date']]  
dfneed_filter = dfneed_filter.sort_values(by= 'customer_id', ascending=True)  
  
# Drop data duplikat  
dfneed_filter = dfneed_filter[['customer_id','register_date']].drop_duplicates()  
dfneed_filter
```

### Result

	customer_id	registered_date	grid icon	row icon	edit icon
652	C107850L	2022-08-03			
4254	C110122L	2022-08-14			
378	C114786L	2022-01-28			
616	C115342L	2022-06-17			
3667	C115470L	2022-06-11			
...	...	...			
426	C994467L	2022-03-15			
891	C995583L	2022-03-02			
4663	C995774L	2022-02-18			
3480	C995819L	2022-07-21			
4586	C998847L	2022-07-03			
625 rows × 2 columns					



## Question 3

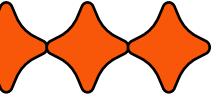
### Insight

Pada data tersebut kelompokkan pelanggan berdasarkan durasi mereka telah menjadi pelanggan (misalnya, pelanggan yang telah bergabung selama lebih dari 1 tahun, 2 tahun, dst.). Berikan diskon eksklusif, voucher, atau penawaran khusus lainnya untuk mendorong pelanggan menyelesaikan transaksi mereka. Pertimbangkan untuk memberikan bonus atau keuntungan tambahan selama periode ulang tahun perusahaan untuk meningkatkan daya tarik kampanye.

### Analisis Lanjutan

Kita bisa memberikan diskon eksklusif, voucher, atau penawaran khusus lainnya kepada pelanggan – pelanggan yang sudah bergabung dalam jangka waktu tertentu sehingga akan mendorong mereka untuk menyelesaikan transaksi mereka. Campaign ini bisa dilakukan saat ulang tahun perusahaan selain itu hal seperti ini diharapkan dapat menarik pelanggan baru menjadi pelanggan tetap kedepannya dan pelanggan lainnya untuk melakukan transaksi lebih banyak lagi.

# Project Python \*



## Question 4

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 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:

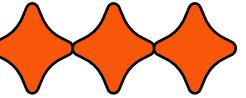
1. Rata-rata harian penjualan weekends (Sabtu dan Minggu) vs rata-rata harian penjualan weekdays (Senin-Jumat) per bulan tersebut. Apakah ada peningkatan penjualan pada masing-masing bulan tersebut.
2. Rata-rata harian penjualan weekends (Sabtu dan Minggu) vs rata-rata harian penjualan weekdays (Senin-Jumat)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

# Project Python \*



## Question 4

### Syntax

```
# 4.1
# Membuat kolom tambahan untuk day, month, month_num
df['order_date'] = pd.to_datetime(df['order_date'], errors='coerce')
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)

# Filter weekday sales
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()
    .reset_index(name='avg_sales_weekday')
).copy()

# Filter weekend sales
monthly_weekend2 = 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()
    .reset_index(name='avg_sales_weekend')
).copy()

# Gabungkan weekday & weekend sales
monthly2 = monthly_weekday2.merge(monthly_weekend2, on='month')

# Urutkan berdasarkan bulan (descending)
monthly2.sort_values(by=['month'], ascending=False, inplace=True)

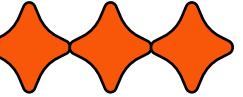
# Pilih kolom yang akan ditampilkan
df_groupby_weekend = monthly2[['month', 'avg_sales_weekend', 'avg_sales_weekday']]
df_groupby_weekend['diff_value'] = df_groupby_weekend['avg_sales_weekend'] - df_groupby_weekend['avg_sales_weekday']
df_groupby_weekend['diff_percent'] = round((df_groupby_weekend['diff_value'] / df_groupby_weekend['avg_sales_weekend']) * 100, 1)

# Tampilkan hasil
df_groupby_weekend
```

### Result

	month	avg_sales_weekend	avg_sales_weekday	diff_value	diff_percent
2	October	634260.074074	874690.266667	-240430.192593	-37.9
1	November	607794.210526	641862.000000	-34067.789474	-5.6
0	December	410599.400000	813574.285714	-402974.885714	-98.1

# Project Python \*

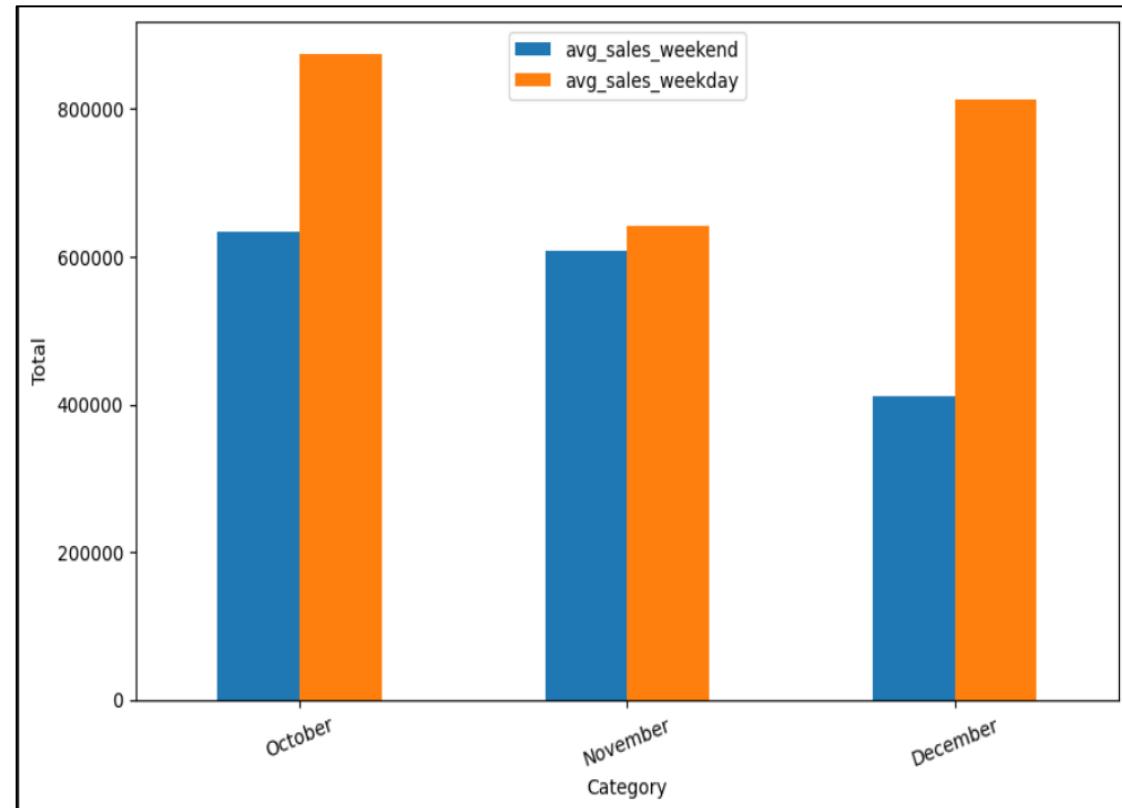


## Question 4

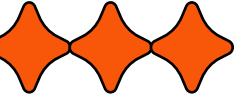
### Syntax

```
monthly2.plot(x='month',
               y=['avg_sales_weekend', 'avg_sales_weekday'],
               kind='bar',
               grid=False,
               xlabel='Category',
               ylabel='Total',
               figsize=(10, 6),
               rot=20,
               table=False,
               secondary_y=False)
```

### Result



# Project Python \*



## Question 4

### Syntax

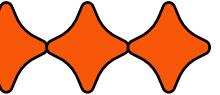
```
#4.2
data21 = {
    'Periode': ['Total 3 months'],
    'Avg Weekend Sales': round(monthly_weekend2['avg_sales_weekend'].mean(), 2),
    'Avg Weekdays Sales': round(monthly_weekday2['avg_sales_weekday'].mean(), 2),
    'Diff (Value)': round(monthly_weekend2['avg_sales_weekend'].mean() - monthly_weekday2['avg_sales_weekday'].mean(), 2),
    '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) + '%'
}

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

### Result

	Periode	Avg Weekend Sales	Avg Weekdays Sales	Diff (Value)	Diff (%)
0	Total 3 months	550884.56	776708.85	-225824.29	-40.99%

# Project Python \*

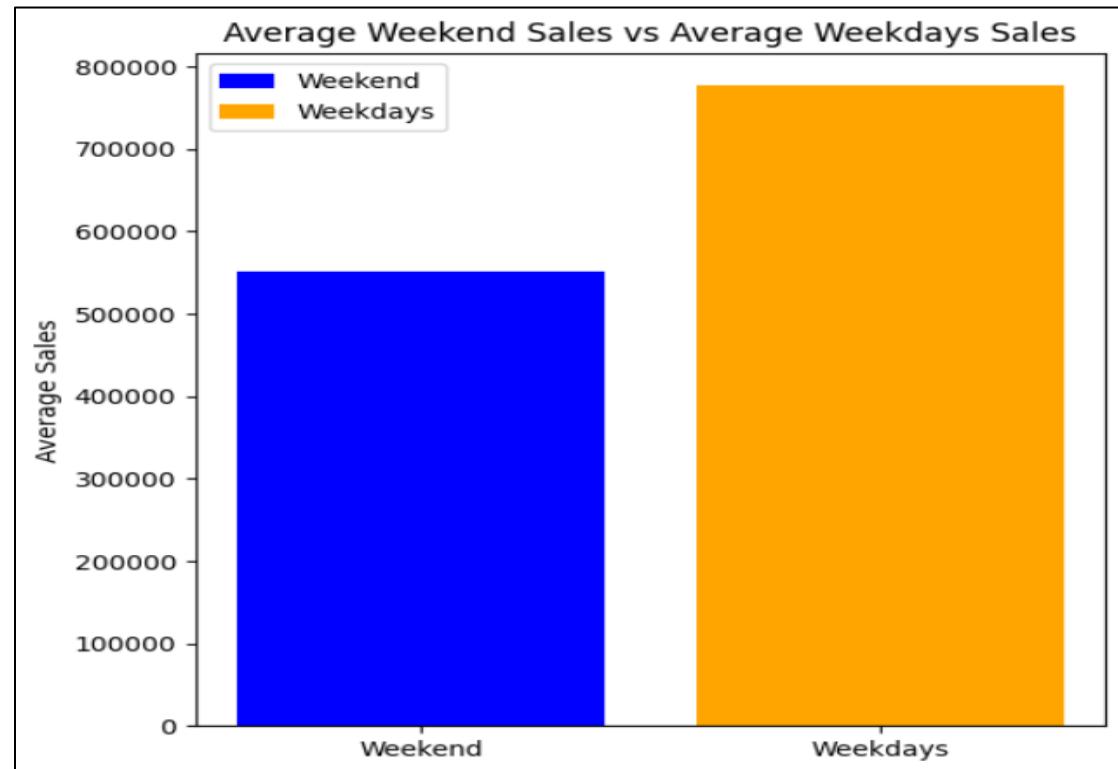


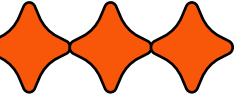
## Question 4

### Syntax

```
avg_weekend_sales = df.groupby_weekend['avg_sales_weekend'].mean()  
avg_weekday_sales = df.groupby_weekend['avg_sales_weekday'].mean()  
  
fig, ax = plt.subplots(figsize=(6, 6))  
ax.bar('Weekend', avg_weekend_sales, label='Weekend', color='blue')  
ax.bar('Weekdays', avg_weekday_sales, label='Weekdays', color='orange')  
ax.set_ylabel('Average Sales')  
ax.set_title('Average Weekend Sales vs Average Weekdays Sales')  
ax.legend()  
plt.show()
```

### Result





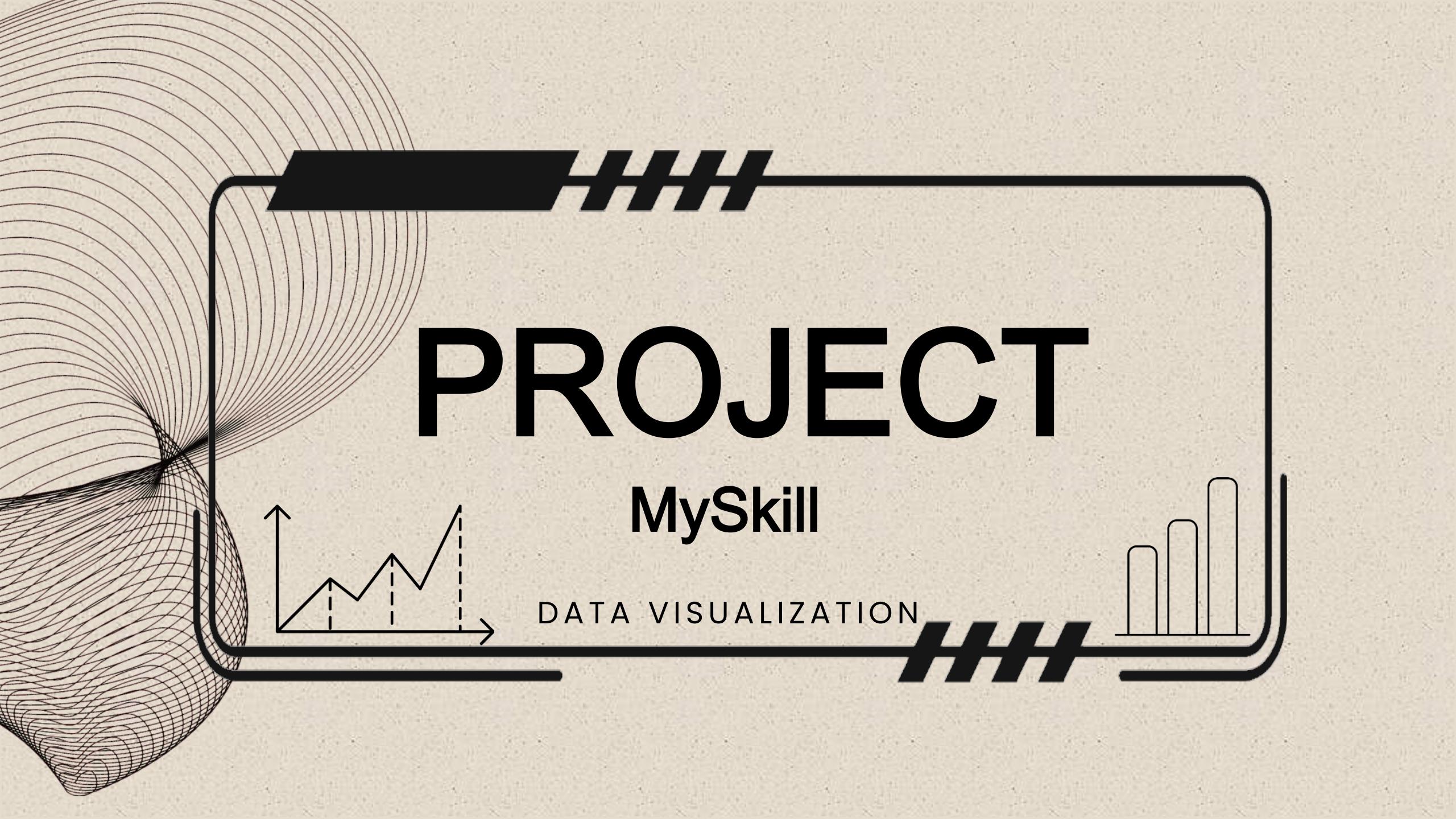
## Question 4

### Insight

Pada data tersebut dapat dilihat rata - rata penjualan selama 3 bulan mengalami penurunan, tetapi pada bulan ke 12 pada saat weekday penjualan mengalami peningkatan sementara dibulan yang sama saat weekend penjualan mengalami penurunan yang cukup drastic.

### Analisis Lanjutan

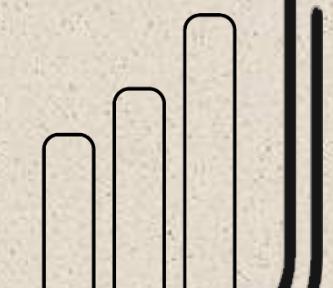
Penurunan penjualan saat weekend bisa dipengaruhi karena saat weekend adalah saatnya orang - orang beristirahat sehingga malas untuk berpergian keluar rumah. Untuk mengatasinya kita bisa melakukan beberapa promo seperti pemberian voucher, atau promo bundle product disaat weekend sehingga diharapkan dapat menarik minat konsumen untuk berbelanja.



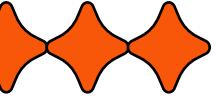
# PROJECT

## MySkill

DATA VISUALIZATION



# Data Visualization \*



## Introduction to Data Visualization

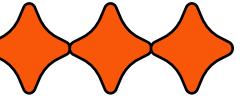
Data visualization adalah proses penyajian data dalam bentuk visual seperti grafik, diagram, atau peta, untuk mempermudah pemahaman dan analisis informasi. Tujuan utamanya untuk mengubah data mentah yang kompleks menjadi bentuk yang lebih mudah dipahami, sehingga pengguna dapat dengan cepat menangkap pola, tren, dan hubungan yang ada dalam data.

Data visualization membantu dalam menyederhanakan informasi yang besar dan rumit, membuatnya lebih intuitif bagi berbagai kalangan, baik itu pengambil keputusan, analis data, atau pembaca yang tidak memiliki background teknik yang kuat. Dengan data visualization, cerita yang terkandung dalam data dapat disampaikan secara lebih jelas dan menarik.

Tools yang digunakan:



# Data Visualization \*



## Overview

Dataset bermanfaat untuk memperoleh wawasan mendalam tentang pola perilaku pelanggan, efisiensi pelanggan, serta mengoptimalkan strategi penjualan produk.

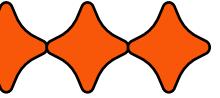
Berikut dataset yang akan digunakan [df](#)

Data yang digunakan adalah data yang berasal dari Tokopedia (bukan data sesungguhnya). Mengenai penjelasan dataset adalah sebagai berikut.

### order\_detail:

- |                     |  |
|---------------------|--|
| 1. id               | → angka unik dari order / <i>id_order</i>                        |
| 2. customer_id      | → angka unik dari pelanggan                                      |
| 3. order_date       | → tanggal saat dilakukan transaksi                               |
| 4. sku_id           | → angka unik dari produk (sku adalah <i>stock keeping unit</i> ) |
| 5. price            | → harga yang tertera pada tagging harga                          |
| 6. qty_ordered      | → jumlah barang yang dibeli oleh pelanggan                       |
| 7. before_discount  | → nilai harga total dari produk ( <i>price * qty_ordered</i> )   |
| 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                              |
| 14. sku_name        | → nama dari produk   |
| 15. base_price      | → harga barang yang tertera pada tagging harga / <i>price</i>    |
| 16. cogs            | → <i>cost of goods sold</i> / total biaya untuk menjual 1 produk |
| 17. category        | → kategori produk  |
| 18. registered_date | → tanggal pelanggan mulai mendaftarkan diri sebagai anggota      |
| 19. payment_method  | → metode pembayaran yang digunakan                               |

# Data Visualization \*



## Question 1

Dear Data Analyst,

Menindaklanjuti meeting gabungan kemarin, kami akan membuat suatu dashboard untuk memantau dan mengevaluasi pencapaian penjualan tiap bulannya. Dalam dashboard tersebut akan terdiri dari 2 halaman dan dibuat dengan menggunakan lookerstudio.google.com.

Halaman 1:

1. Tim marketing ingin melihat perkembangan campaign pada tahun 2022, trend nya seperti apa? Diberikan penjelasan dan bagaimana call to action-nya agar tim dapat mengambil sebuah keputusan ke depannya.

Dashboard yang berupa:

- a. Hubungan antara Value Sales (before discount), Net Profit, dan AOV (average order value).

Note:

Net profit = Value Sales (before discount) - (cogs \* qty)

AOV = Value Sales (before discount) / Total Unique Order.

- b. Terdapat slicer Order Date, Category, Sales Value, Value Transaction, Payment.

Note:

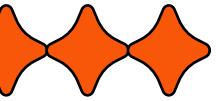
Value Transaction:

Valid → is\_valid = 1

Not Valid → is\_valid = 0

Payment → payment\_method

# Data Visualization \*



## Question 1

Menambahkan kolom baru

Data

Search

df - df

after\_discount

base\_price

before\_discount

category

cogs

customer\_id

discount\_amount

Properti

Batang filter

Tambahkan kolom kalkulasi

Tambahkan grup

Tambahkan bin

+ Tambahkan kolom

+ Tambahkan parameter

+ Tambahkan Data

Menambahkan kolom AOV

df - df

← SEMUA KOLOM

Kolom yang Tersedia

misalnya, Kolom Kalkulasi Baru — AOV

Formula

1 `SUM(before_discount)/COUNT_DISTINCT(id)`

Nama Kolom

ID kolom

Cakupan: Disemakkan | Kredensial data: Mila Aulia N

Menambahkan kolom Net Profit

df - df

← SEMUA KOLOM

Kolom yang Tersedia

misalnya, Kolom Kalkulasi Baru — Net Profit

Formula

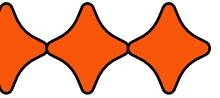
1 `before_discount - (cogs * qty_ordered)`

Nama Kolom

ID kolom

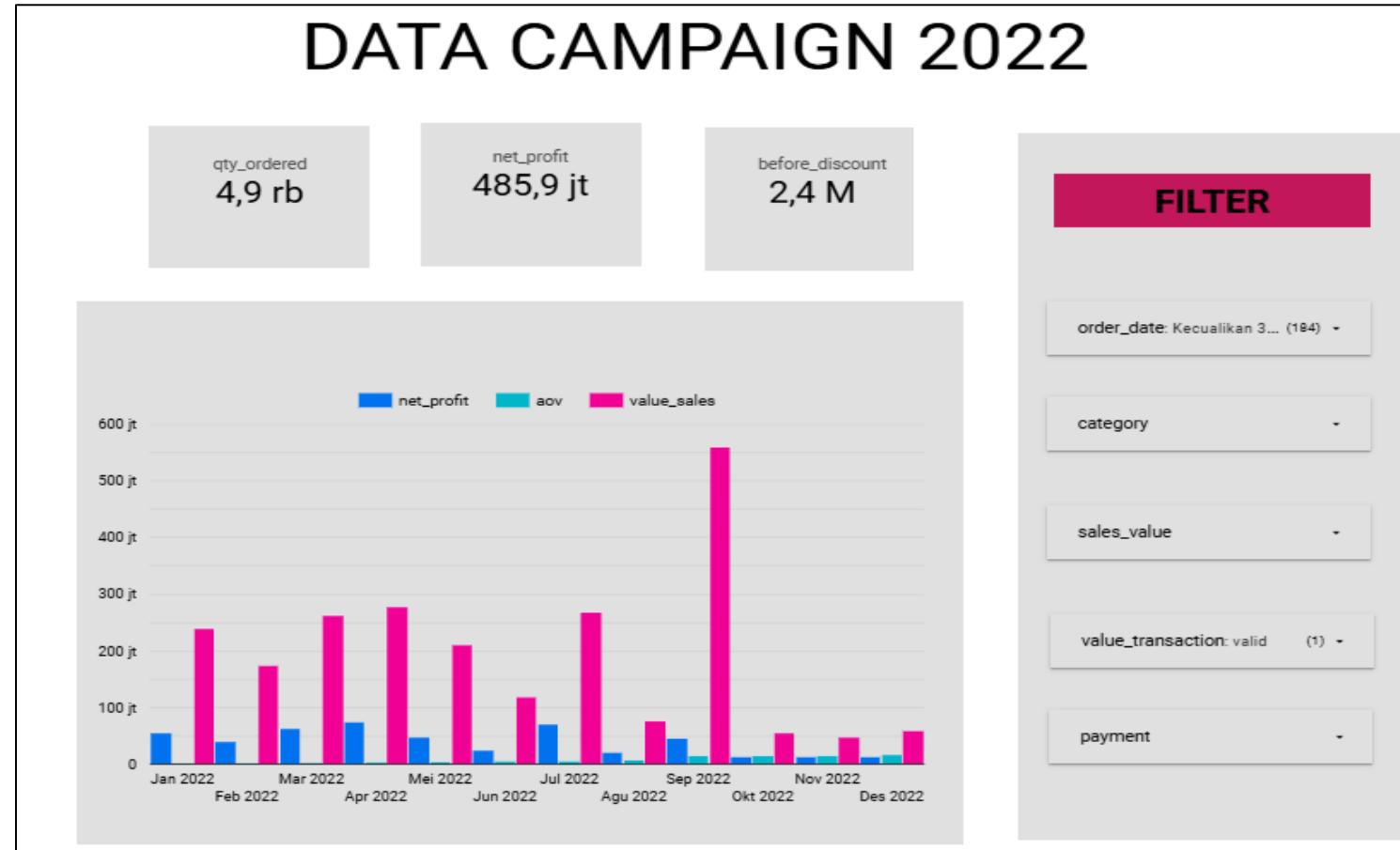
Cakupan: Disemakkan | Kredensial data: Mila Aulia N

# Data Visualization \*

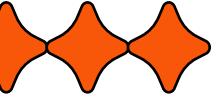


## Question 1

### Chart



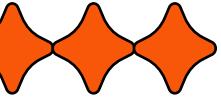
# Data Visualization \*



## Question 1

1. Membuat slicer Order Date, Category, Value Sales, Value Transaction dan Payment
  - slicer: digunakan untuk melakukan filter data yang akan ditampilkan
  - order date: menampilkan data berdasarkan date range tertentu (tahun 2022)
  - category: menampilkan data penjualan pada kategori yang ingin ditampilkan
  - value sales: menampilkan transaksi berdasarkan nilai `is_valid`, dimana `is_valid = 1` artinya transaksi 'valid' dan `is_valid = 0` artinya transaksi 'not valid'
  - payment: menampilkan data penjualan berdasarkan metode transaksi yang digunakan
2. Menampilkan data penjualan selama tahun 2022  
data yang ditampilkan merupakan data yang telah difilter yaitu data penjualan pada tahun 2022
3. menampilkan data 'valid'  
menampilkan transaksi dengan ketentuan 'valid'
4. Melihat besaran nilai pada sumbu 'Y'  
pada sumbu Y menunjukkan net profit, AOV dan value sales

# Data Visualization \*



## Question 1

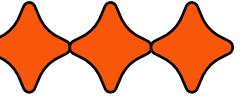
### Insight

1. Penjualan tertinggi terjadi pada bulan September 2022 berbanding terbalik dengan Net Profit yang sedikit.
2. Pada Q4 tahun 2022 (bulan Oktober-Desember), penjualan mengalami penurunan. Hal ini menunjukkan bahwa kampanye yang telah dilakukan tidak lagi efektif dalam menarik perhatian pelanggan. Oleh karena itu, perusahaan dapat melihat kesempatan ini sebagai peluang untuk melakukan evaluasi dan merancang strategi baru untuk meningkatkan penjualan di masa depan.

Berikut beberapa Call to Action yang kami sarankan kepada Tim Marketing:

1. Tinjau kembali strategi diskon dan penawaran harga. Pastikan bahwa strategi ini tidak mengorbankan margin keuntungan yang optimal.
2. Fokus pada efisiensi operasional untuk mengurangi biaya tambahan yang mungkin telah meningkat.
3. Lakukan analisis mendalam terhadap biaya-biaya yang muncul pada bulan September untuk mengidentifikasi area di mana penghematan dapat dicapai.
4. Tingkatkan kolaborasi antara tim marketing dan tim keuangan untuk memahami dampak strategi pemasaran terhadap keuntungan bersih.
5. Tetapkan target spesifik dan terukur untuk meningkatkan Net Profit di bulan-bulan mendatang.

# Data Visualization \*



## Question 2

Halaman 2

1. Dapat memperlihatkan:

a. Tabel berisi:

- Product Name
- Category
- Before Discount
- After Discount
- Net Profit
- Qty
- Customer (unique value)

b. Terdapat slicer: Order Date, Category, Value Transaction, Payment

c. Scorecard:

- Before Discount
- After Discount
- Net Profit
- Quantity
- Customer (unique value)
- AOV

2. Selama 2022, tampilkan category mobile & tablet yang sudah terjadi pembayaran secara jazzwallet. Berapa quantity dan customer nya?

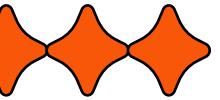
3. Buatlah chart berdasarkan dashboard nomor 2.

Terima kasih.

Regard,

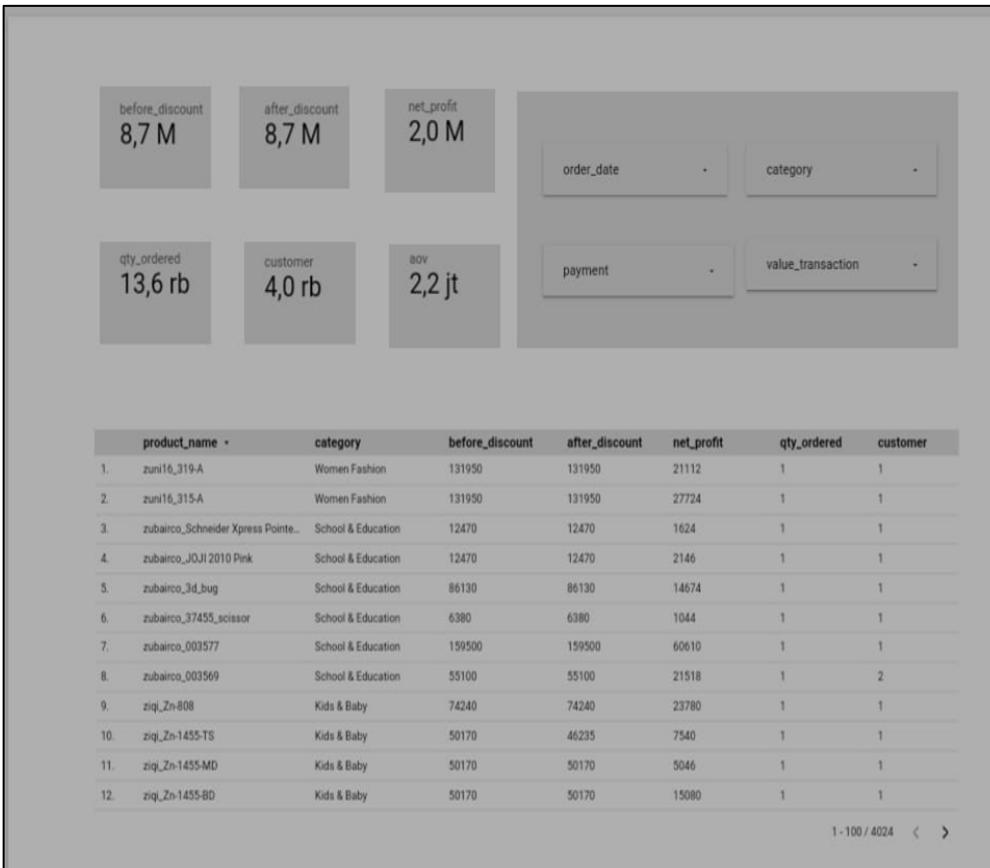
Tim Marketing

# Data Visualization \*

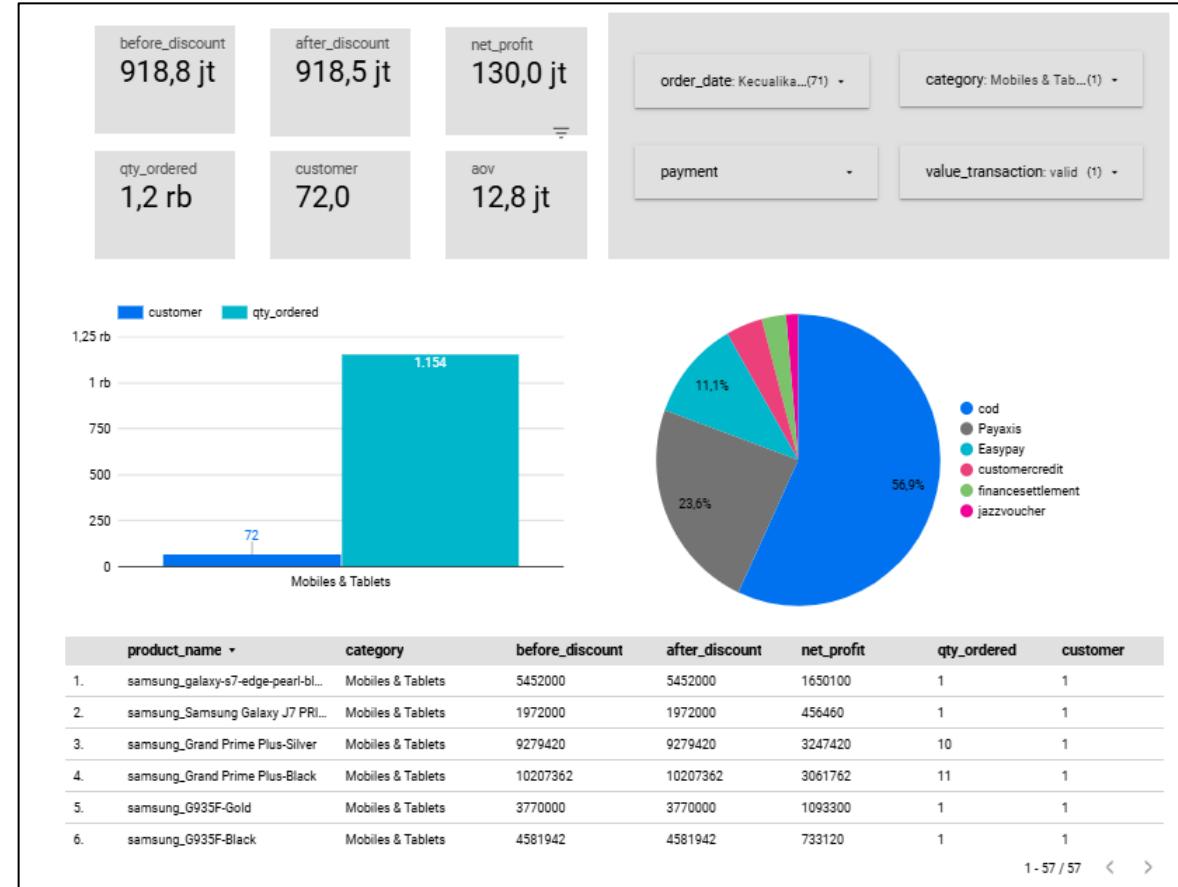


## Question 2

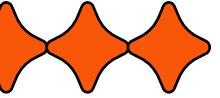
### Dashboard



### Dashboard



# Data Visualization \*

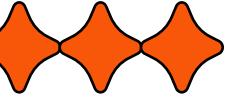


## Question 2

### Penjelasan

1. Membuat slicer Order Date, Category, Value Sales, Value Transaction dan Payment
  - slicer: digunakan untuk melakukan filter data yang akan ditampilkan.
  - order date: menampilkan data berdasarkan date range tertentu (tahun 2022).
  - category: menampilkan data penjualan pada kategori 'Mobile & Tablets'.
  - value transaction: menyajikan nilai transaksi 'valid'.
  - payment: menampilkan semua metode pembayaran (Pada hal ini, tidak ditemukan pembayaran menggunakan jazzwallet oleh customer di kategori 'Mobile & Tablets'. Kemungkinan penggunaan jazzwallet kurang diminati dikalangan pelanggan 'Mobile & Tablets' karena kurangnya kepercayaan pelanggan hal ini disebabkan budaya cash yang masih banyak dilakukan oleh pelanggan. Selain itu karena 'Mobile & Tablets' tergolong barang mahal sehingga pelanggan masih memiliki kekhawatiran dalam bertransaksi menggunakan jazzwallet dan kemungkinan lainnya adalah customer memanfaatkan promo di metode pembayaran lainnya).
2. Menampilkan data penjualan selama tahun 2022
  - data yang ditampilkan merupakan data yang telah difilter yaitu data penjualan pada tahun 2022.

# Data Visualization \*

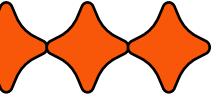


## Question 2

### Insight

1. Perusahaan mencapai Net Profit sebesar Rp 130 juta dari 72 Customer, menunjukkan kontribusi signifikan dari setiap pelanggan.
2. Penjualan 1.200 unit produk menghasilkan Net Profit sebesar Rp 130 juta, menunjukkan bahwa strategi penjualan produk sangat efektif.
3. Average Order Value (AOV) sebesar 12,8jt, menyatakan bahwa pelanggan memiliki daya beli yang tinggi.
4. Penjualan produk didominasi oleh produk premium di kategori 'Mobile & Tablet', hal ini menunjukkan kekuatan merek dan produk Perusahaan di pasaran.
5. Payment didominasi oleh transaksi COD, hal ini menunjukkan masih adanya keraguan costumer terhadap transaksi online.

# Data Visualization \*



## Storyline

Current Situation	Findings	Recommendation
<p>Secara keseluruhan, total qty_ordered mencapai 4,9 rb dengan net_profit sebesar 485,9 jt dan juga nilai sebelum diskon sebesar 2,4 M. Angka tersebut memperlihatkan dampak yang cukup besar dari strategi pemasaran yang digunakan.</p>	<p>Melalui grafik dapat dilihat bahwa values_sales mengalami fluktuasi sepanjang tahun. Terjadi lonjakan secara signifikan pada bulan Maret, Mei, Agustus dan puncaknya yaitu September 2022. Hal tersebut kemungkinan disebabkan oleh campaign besar-besaran. net_profit serta avg_value juga terlihat mengikuti pola yang sama walaupun skalanya lebih kecil dibanding dengan value_sales.</p>	<p>Puncak pada bulan September 2022 memperlihatkan bahwa campaign promosi yang efektif dapat meningkatkan angka penjualan secara signifikan. Sedangkan penurunan setelah September perlu dievaluasi lebih lanjut untuk menjaga kestabilan penjualan. Penggunaan filter data juga membantu mengidentifikasi pelanggan untuk menyesuaikan strategi pemasaran di tahun berikutnya</p>