

Technical Test – Data Mining Internship

Nama: Yusuf Budi Kusuma

Email & No. HP: yusufbudikusuma7@gmail.com / +62858-9475-3213

Tanggal Pengumpulan: 22 Mei 2025

Introduction

Pada technical test ini diberikan data dengan total 75970 baris data dan 13 kolom, data data tersebut dibarengi dengan diberikannya beberapa pertanyaan yang mencakup pada technical test yang diberikan. Total ada 4 pertanyaan yang akan coba dijawab pada lembar pengerjaan technical test ini.

Alat yang digunakan antara lain :

- Microsoft Excel: Untuk pengelolaan data awal.
- XAMPP : Untuk membuka akses ke MySQL.
- Google Colab : Untuk menjalankan python dalam proses persiapan dan pembersihan data
- PHP My Admin : Untuk menjalankan MySQL dan melakukan query dan analisis data.

Data Preparation

a. Source Data

Data yang akan diolah diberikan dalam bentuk Excel dengan total 75970 data dan 13 kolom, data tersebut untuk diolah dalam MySQL akan diubah tipe nya menjadi CSV terlebih dahulu dengan tujuan mempermudah saat akan di import ke database.

b. Struktur tabel

Kolom	Contoh Data	Tipe Data (MySQL)	Penjelasan
tgl_trs	1/2/2024	DATE	Tanggal transaksi, dapat dikonversi ke format YYYY-MM-DD.
shipment_id	JJ03/0124/90001	VARCHAR(20)	Kombinasi huruf dan angka, bukan numerik murni.
kode_toko	AB03	VARCHAR(10)	Kode toko, huruf dan angka.
members	11003443935	BIGINT	Nomor anggota (numeric panjang).
SKU	125127	INT	Kode SKU (barang), berupa angka.
kategori	2413 - SKM (...)	VARCHAR(100)	Kombinasi kode dan nama kategori, lebih cocok disimpan sebagai teks.
sub_dept	241 - CIGARETTE	VARCHAR(100)	Kombinasi kode dan nama sub-department.
departemen	24 - CIGARETTE & LIGHTER	VARCHAR(100)	Sama seperti di atas, gabungan kode dan nama.
tag	A	CHAR(1)	Label kategori, satu huruf (A, B, dst).
qty	10	INT	Kuantitas produk.
net_sales	216250	INT	Penjualan bersih (tanpa simbol mata uang).
gm	4250	INT	Gross margin dalam satuan nilai.
%gm	1.97%	DECIMAL(5,2)	Persentase gross margin, perlu diubah dari string '1.97%' menjadi 1.97.

c. Data Cleaning & Check

Pada bagian ini, saya akan menggunakan python melalui google colab untuk menyesuaikan data yang diberikan agar siap untuk diolah dengan query SQL. Berikut langkah langkah yang dilakukan

1. Mempersiapkan data excel
2. Membuka Google Colab dan membuat projek baru.
3. Melakukan instalasi library yang diperlukan
4. Menampilkan ringkasan data dan melakukan EDA sederhana
5. Cleansing data mencakup mencari data yang kosong, data yang duplikat atau data dengan nilai NaN.
6. Karena akan dilakukan upload data ke SQL, maka saya melakukan analisis pada data dan menemukan beberapa penyesuaian yang perlu dilakukan, antara lain sebagai berikut.

Kolom	Tipe Saat Ini	Masalah / Catatan	Solusi (Langkah Preparation)
tgl_trs	String/Tanggal	Format campuran, perlu diseragamkan	Konversi ke datetime
members	Numeric (int-like)	Pastikan bertipe str atau bigint di SQL	Konversi ke str jika perlu
SKU	Numeric (int-like)	Harus disesuaikan sebagai int di SQL	Konversi ke int
qty	Numeric	Pastikan sebagai int	Konversi ke int
net_sales	Numeric	Pastikan sebagai int	Konversi ke int
gm	Numeric	Pastikan sebagai int	Konversi ke int
%gm	String (mengandung %)	Harus dihapuskan % dan dikonversi ke float	Bersihkan dan konversi

7. Setelah proses selesai, data disimpan dalam format CSV dan dapat mulai untuk proses mengunggahnya ke SQL.

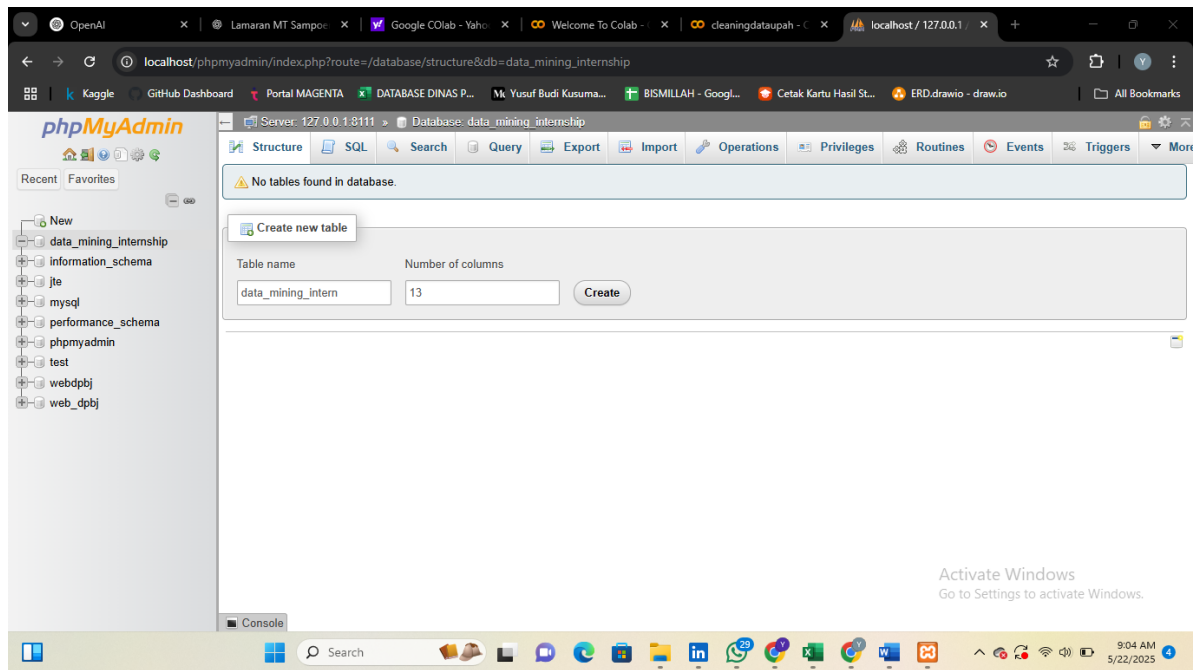
Note:

Berikut ini link worksheet dari google colab yang saya buat: [google colab link](#)

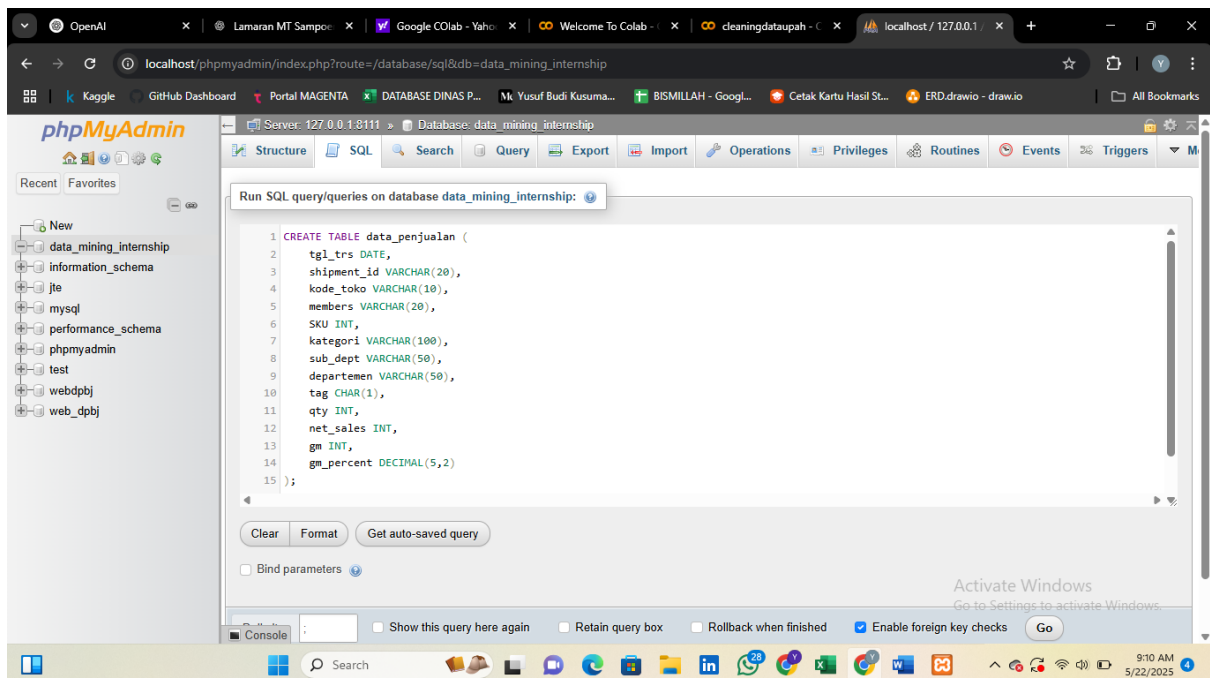
Berikut ini file excel yang sudah melalui proses preparation data: [data.csv](#)

d. Import data ke MySQL

Pada proses ini saya akan mengunggah data ke SQL, menggunakan XAMPP untuk akses MySQL. Langkah Langkah yang dilakukan secara singkat adalah dengan memulai program XAMPP kemudian mengaktifkan Apache dan MySQL sehingga bisa lanjut ke browser untuk menjalankan local host dan memilih PHPMyAdmin untuk mengakses SQL. Kemudian menambahkan database baru dengan nama data_mining_internship dan didalamnya ditambahkan tabel data_mining_intern sehingga gambarannya sebagai berikut.



Selanjutnya, dilakukan dengan membuat query untuk menambahkan tabel sebagai berikut.



Setelah tabel dibuat, selanjutnya import data ke tabel pada database, dan setelah berhasil di import sebagai berikut

Server: 127.0.0.1:31111 Database: data_mining_internship Table: data_penjualan

Extra options

tgl_trs	shipment_id	kode_toko	members	SKU	kategori	sub_dept	departemen	tag	qty	net_sales	gm	gm_percent
2024-01-02	JJ03/0124/90001	AB03	11003443935	125127	2413 - SKM (SIGARET KRETEK MESIN) FF	241 - CIGARETTE	24 - CIGARETTE & LIGHTER	A	10	216250	4250	0.02
2024-01-03	JJ12/0124/90013	AB12	11001476143	115882	2411 - SKT (SIGARET KRETEK TANGAN)	241 - CIGARETTE	24 - CIGARETTE & LIGHTER	A	2	27350	550	0.02
2024-01-04	JJ15/0124/90008	AB15	11003751915	8802	1912 - INSTANT NOODLE CUP	191 - INSTANT NOODLES & PASTA	19 - INSTANT FOODS	B	1	3964	262	0.07
2024-01-05	JJ39/0124/90033	AB39	11002963768	7861	1212 - BAG TEA	121 - TEA	12 - TEA, COFFEE, CREAMER & TRADITIONAL	A	9	51892	15018	0.29
2024-01-06	JJ27/0124/90031	AB27	10001271801	807718	3231 - WOMEN DEO	323 - BODY FRAGRANCE	32 - PERSONAL CARE	J	1	21982	1055	0.05
2024-01-07	JJ15/0124/90024	AB15	11001929190	438869	2411 - SKT (SIGARET KRETEK TANGAN)	241 - CIGARETTE	24 - CIGARETTE & LIGHTER	A	50	635000	5000	0.01
2024-01-08	JJ36/0124/90009	AB36	11003319135	438000	1723 - SOY SAUCE	172 - SAUCE & CONDIMENTS	17 - SEASONING	J	6	48108	1749	0.04
2024-01-09	JJ03/0124/90011	AB03	11002443571	429049	1714 - STOCK SOUP	171 - MSG, SPICES, SALT & OTHER INGREDIENTS	17 - SEASONING	J	2	8468	410	0.05
2024-01-10	JJ26/0124/90003	AB26	11002341411	429048	1714 - STOCK SOUP	171 - MSG, SPICES, SALT & OTHER INGREDIENTS	17 - SEASONING	J	2	8468	410	0.05
2024-01-11	JJ19/0124/90006	AB19	11004141041	427879	1113 - MINERAL WATER	111 - SOFT DRINK	11 - BEVERAGES	J	5	65315	6903	0.11
2024-01-12	JJ20/0124/90031	AB20	11002337211	406368	2322 - CAN & BOTTLE CANDY	232 - CANDY	23 - CONFECTIONERY	J	1	20631	2461	0.12
2024-01-13	JJ71/0124/90005	AB71	11001871442	244994	3711 - POWDER DETERGENT	371 - MAIN WASH	37 - LAUNDRY	F	1	4414	171	0.04
2024-01-14	JJ02/0124/90015	AB02	11004092749	240018	2412 - SKM (SIGARET KRETEK MESIN) LTLN	241 - CIGARETTE	24 - CIGARETTE & LIGHTER	O	50	1737500	22394	0.01

Proses preparation data telah selesai dan siap memasuki sesi jawaban dari pertanyaan pertanyaan yang diberikan.

Jawaban Pertanyaan

1. Buatkan Query menggunakan MySQL untuk mencari sales per bulan per toko, lalu sort dari sales tertinggi.

Jawaban:

Query SQL yang digunakan

```
SELECT
    kode_toko,
    DATE_FORMAT(tgl_trs, '%Y-%m') AS bulan,
    SUM(net_sales) AS total_sales

FROM
    data_penjualan

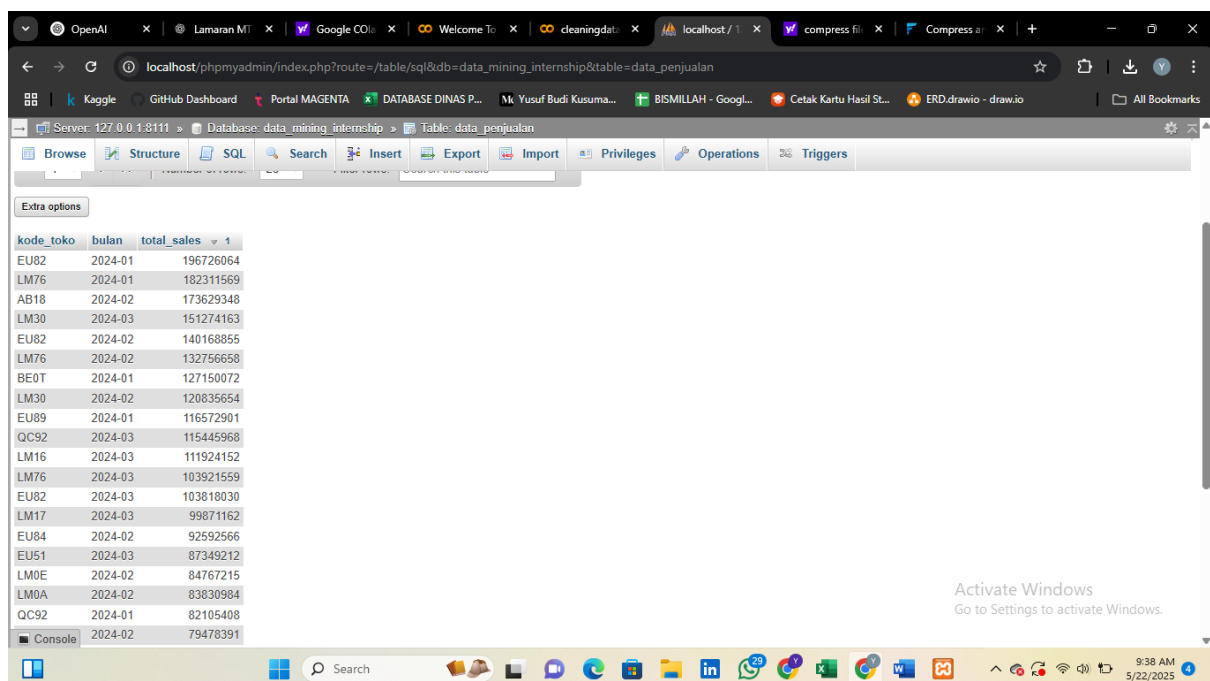
GROUP BY
    kode_toko, bulan

ORDER BY
    total_sales DESC;
```

Penjelasan:

Query ini akan mengambil data penjualan bersih (net_sales) dari setiap toko (kode_toko) yang dikelompokkan per bulan (tgl_trs diformat ke 'YYYY-MM') menggunakan fungsi DATE_FORMAT. Total penjualan per toko per bulan dihitung dengan SUM(net_sales), dan hasil akhirnya diurutkan berdasarkan total_sales dari yang tertinggi.

Output yang dihasilkan sebagai berikut:



kode_toko	bulan	total_sales
EU82	2024-01	196726064
LM76	2024-01	182311569
AB18	2024-02	173629348
LM30	2024-03	151274163
EU82	2024-02	140168855
LM76	2024-02	132756658
BE0T	2024-01	127150072
LM30	2024-02	120835654
EU89	2024-01	116572901
QC92	2024-03	115445968
LM16	2024-03	111924152
LM76	2024-03	103921559
EU82	2024-03	103818030
LM17	2024-03	99871162
EU84	2024-02	92592566
EU51	2024-03	87349212
LM0E	2024-02	84767215
LM0A	2024-02	83830984
QC92	2024-01	82105408
	2024-02	79478391

The screenshot shows the phpMyAdmin interface for a database named 'data_mining_internship'. The selected table is 'data_penjualan'. A SQL query is entered in the query box:

```
SELECT kode_toko, DATE_FORMAT(tgl_trs, '%Y-%m') AS bulan, SUM(net_sales) AS total_sales FROM data_penjualan GROUP BY kode_toko, bulan ORDER BY total_sales DESC;
```

The query results are displayed in a table with 43 rows. The first few rows are:

kode_toko	bulan	total_sales
BE81	2024-02	1282873
QC94	2024-03	1196847
EU94	2024-02	1188152
BE39	2024-02	746128
BE0J	2024-01	595181
BE0J	2024-03	450046

Dari data yang didapatkan tersebut, dapat ditarik kesimpulan bahwa toko dengan **kode toko EU82 memiliki total penjualan tertinggi** pada bulan February 2024 atau (2024-02) dengan total penjualan 196726064 atau Rp. 196.726.064 dan toko dengan **penjualan terendah adalah toko dengan kode toko BE0J** pada bulan Maret 2024 atau (2024-03) dengan total penjualan 450046 atau sebesar Rp. 450.046.

2. Buatkan Query menggunakan MySQL untuk mencari top 20 member dengan %GM tertinggi

Jawaban:

Query SQL yang digunakan

```
SELECT
    members,
    AVG(`gm_percent`) AS avg_gm

FROM
    data_penjualan

GROUP BY
    Members

ORDER BY
    avg_gm DESC

LIMIT 20;
```

Penjelasan:

- Query ini mengambil rata-rata gross margin (**%gm**) untuk setiap **members** menggunakan **AVG()**.
- Dikelompokkan berdasarkan **members** (karena satu member bisa melakukan lebih dari satu transaksi).
- Hasilnya diurutkan dari rata-rata **%GM** tertinggi dan dibatasi hanya **20 data teratas** menggunakan **LIMIT 20**.

Output yang dihasilkan:

members	avg_gm
11003928059	0.443333
11003319721	0.420000
11001930230	0.420000
11000445016	0.420000
11001297036	0.416667
11004043163	0.416667
11000966342	0.416667
11003969665	0.413333
10001509911	0.413333
11004052198	0.413333
11004054319	0.413333
11003935799	0.413333
11001684449	0.413333
11003699130	0.413333
11003101095	0.400000
11003314649	0.400000
11003234102	0.390000
11003182897	0.390000
11003765527	0.390000
11002214584	0.380000

Dari hasil yang didapatkan, kesimpulan yang bisa diambil adalah members 11003928059 sampai dengan members 11002214584 adalah 20 members teratas dengan rata rata %gm tertinggi.

3. Buatkan Query menggunakan MySQL untuk mencari kategori dengan qty terbanyak berdasarkan untuk tag OFMBJ saja

Jawaban:

Query yang digunakan

```
SELECT
    tag, kategori,
    SUM(qty) AS total_qty

FROM
    data_penjualan

WHERE
    tag IN ('O', 'F', 'M', 'B', 'J')

GROUP BY
```

tag, kategori
ORDER BY
total_qty DESC;

Penjelasan:

Query ini menampilkan kategori dengan jumlah qty terbanyak untuk masing-masing tag (O, F, M, B, J), dan menampilkannya.

Output yang dihasilkan:

tag	kategori	total_qty ▼ 1
J	1714 - STOCK SOUP	243023
O	2023 - TETRAPACK LIQUID MILK	70167
J	1723 - SOY SAUCE	65914
B	1911 - INSTANT NOODLE BAG	46772
F	2412 - SKM (SIGARET KRETEK MESIN) LTLN	34584
O	2412 - SKM (SIGARET KRETEK MESIN) LTLN	23635
F	1111 - ENERGY DRINK	22625
O	1911 - INSTANT NOODLE BAG	15966
F	1911 - INSTANT NOODLE BAG	15679
F	3221 - BAR SOAP	12836
J	3711 - POWDER DETERGENT	12201
F	3272 - TOOTH PASTE	9165
M	1512 - SUGAR	8445
J	3714 - LIQUID DETERGENT	8109
B	1912 - INSTANT NOODLE CUP	7714
M	1911 - INSTANT NOODLE BAG	7583
J	3211 - SHAMPOO	5724
J	1154 - RTD TEA	5684
M	3272 - TOOTH PASTE	5274

Kategori 1714- STOCK SOUP dengan tag "J" memiliki total qty tertinggi sebesar 243023, jika ingin ditampilkan 1 total qty tertinggi dari per jenis tag maka bisa menggunakan SQL berikut.

```
SELECT tag, kategori, total_qty
FROM (
  SELECT
    tag, kategori,
    SUM(qty) AS total_qty,
    ROW_NUMBER() OVER (PARTITION BY tag ORDER BY SUM(qty) DESC) AS rn
  FROM
    data_penjualan
```



```

WHERE
    tag IN ('O', 'F', 'M', 'B', 'J')
GROUP BY
    tag, kategori
) AS ranked
WHERE rn = 1;

```

Maka akan menampilkan hanya 5 data sesuai kategori tag dengan jumlah qty tertinggi per masing masing tag sebagai berikut

tag	kategori	total_qty
B	1911 - INSTANT NOODLE BAG	46772
F	2412 - SKM (SIGARET KRETEK MESIN) LTLN	34584
J	1714 - STOCK SOUP	243023
M	1512 - SUGAR	8445
O	2023 - TETRAPACK LIQUID MILK	70167

4. Buatkan Query menggunakan MySQL untuk mencari toko dengan penjualan department non CIGARETTE & LIGHTER.

Jawaban:

Query yang digunakan

```

SELECT
    kode_toko,
    SUM(net_sales) AS total_penjualan

FROM
    data_penjualan

WHERE
    departemen != '24 - CIGARETTE & LIGHTER'

GROUP BY
    kode_toko

ORDER BY
    total_penjualan DESC;

```

Penjelasan:

- Departemen != 'CIGARETTE & LIGHTER' memastikan hanya data dari department non-CIGARETTE & LIGHTER yang diambil.
- SUM(net_sales) menghitung total penjualan per kode_toko.
- GROUP BY kode_toko mengelompokkan hasil berdasarkan toko.
- ORDER BY total_penjualan DESC untuk mengurutkan toko berdasarkan penjualan tertinggi.

Output yang dihasilkan:

kode_toko	total_penjualan ▾ 1
QC66	102800014
LM30	101405461
LM0E	100626506
LM76	90649690
EU84	67152386
QC96	60078706
LM16	58091004
AB40	57775325
QC0H	57084587
QC18	57005908
BE85	50080027
BE34	41884519
QC76	40311814
QC28	38871729
QC68	37862029
QC0B	37762852
QC0D	34778953
EU67	34183196
QC36	32182942

Dari cuplikan data yang dihasilkan tersebut dapat ditarik kesimpulan bahwa kode toko QC66 merupakan toko dengan penjualan department non-cigarette & lighter tertinggi dengan total penjualan Rp 102.800.014.