

FULLSTACK INTENSIVE BOOTCAMP

Data Analyst

MySkill

FINAL PROJECT SQL

By : Abid Muhammad Taufiq

MINI PORTFOLIO



INTRODUCTION



- This analysis is the final project of the Fullstack Intensive Bootcamp Data Analyst held by Myskill.
- This analysis was carried out using MySQL with the DBeaver application
- The dataset contains products sales transaction data on Tokopedia e-commerce for 2 years (2021-2022).
- All datasets are provided in TXT format and have been imported into the local database.
- The dataset used is a dataset provided by Myskill as a learning medium. This dataset is a dummy dataset, not a real dataset.



TOOLS



Database Management System

+



Graphical User Interface



DATASET



There are 4 datasets used, the following is an explanation of each dataset.

A. order_detail :

1. id_order : unique code for each transaction
2. id_customer : unique code for each customer
3. order_date : transaction date
4. id_sku : unique code for each product
5. price : selling price
6. qty_ordered : quantity of product ordered
7. before_discount : total price value of the product (price * qty_ordered)
8. discount_amount : total product discount value
9. after_discount : the total price value of the product when it has been reduced by discounts
10. is_gross : shows the customer has not paid for the order
11. is_valid : shows the customer has made a payment
12. is_net : shows the transaction is complete
13. id_payment : unique code for each payment method

B. sku_detail :

1. id_sku : unique code for each product
2. sku_name : name of product
3. cogs : cost of goods sold / total cost to sell 1 product
4. category : product category

C. customer_detail :

1. id_customer : unique code for each customer
2. registered_date : the date customers start registering as members

D. payment_detail :

1. id_payment : unique code for each payment method
2. payment_method : payment method used

DATA PREPROCESSING

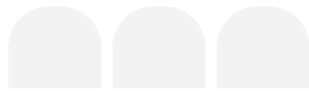


```
CREATE TABLE IF NOT EXISTS customer_detail (  
  id VARCHAR(10) PRIMARY KEY,  
  registered_date DATE  
);  
INSERT INTO customer_detail VALUES  
( 'C996508L', '2021-07-10 00:00:00',  
( 'C180415L', '2021-07-18 00:00:00',  
( 'C535451L', '2021-07-23 00:00:00',  
( 'C177843L', '2021-07-12 00:00:00',  
( 'C951682L', '2021-07-27 00:00:00',  
( 'C483469L', '2021-07-26 00:00:00',  
( 'C346553L', '2021-07-11 00:00:00',  
( 'C631361L', '2021-07-15 00:00:00',  
( 'C534453L', '2021-07-18 00:00:00',  
( 'C404192L', '2021-07-06 00:00:00',  
( 'C720927L', '2021-07-17 00:00:00',  
( 'C525405L', '2021-07-24 00:00:00',  
( 'C614407L', '2021-07-02 00:00:00',  
( 'C955726L', '2021-07-23 00:00:00',  
( 'C818990L', '2021-07-05 00:00:00',  
( 'C143684L', '2021-07-05 00:00:00',  
( 'C665132L', '2021-07-27 00:00:00',  
( 'C946451L', '2021-07-18 00:00:00',  
( 'C113061L', '2021-07-20 00:00:00',  
( 'C663962L', '2021-07-13 00:00:00',
```

customer_detail

```
CREATE TABLE IF NOT EXISTS sku_detail (  
  id_sku VARCHAR(10) PRIMARY KEY,  
  sku_name VARCHAR(100),  
  base_price FLOAT,  
  cogs FLOAT,  
  category VARCHAR(100)  
);  
INSERT INTO sku_detail VALUES  
( 'P798444', 'AT-FSM-35', 57631.7, 46052, 'Kids & Baby',  
( 'P938347', 'AYS_Haier-18HNF', 3931789.26, 3499256, 'Appliances',  
( 'P826364', 'Atalian_DV206A-Brown-41', 324597, 243426, 'Men Fashion',  
( 'P467533', 'Darul_Sakoon_Food_Bundle', 2870.42, 2378, 'Superstore',  
( 'P229955', 'HP_15AY-15-Ay072NIA-ci3', 2265625, 1631250, 'Computing',  
( 'P985828', 'UnzeLondon_GS4666-9', 202855, 172376, 'Men Fashion',  
( 'P381119', 'UnzeLondon_GS5348-10', 202855, 139954, 'Men Fashion',  
( 'P499423', 'UnzeLondon_GS5348-9', 202855, 158224, 'Men Fashion',  
( 'P672852', 'Paramount_9781107677364', 182856.6, 138968, 'School & Education',  
( 'P447157', 'Paramount_9781107697690', 182856.6, 153584, 'School & Education',  
( 'P346967', 'AKL_A131128712_SS-48_Yellow', 173275, 103936, 'Women Fashion',  
( 'P129115', 'Jaza_2Premium5kg+F0C', 99302.38, 75458, 'Superstore',  
( 'P176799', 'Paramount_9781107657823', 96552.6, 71398, 'School & Education',  
( 'P834446', 'emo_Emo-VB-06 Navy Blue Blazer', 93745.4, 84332, 'Women Fashion',
```

sku_detail



DATA PREPROCESSING

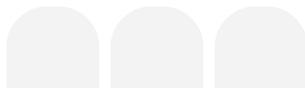


```
● CREATE TABLE IF NOT EXISTS payment_detail (  
    id_payment INT PRIMARY KEY,  
    payment_method VARCHAR(50)  
);  
● INSERT INTO payment_detail VALUES  
(1,'cod'),  
(2,'jazzvoucher'),  
(3,'customercredit'),  
(4,'Payaxis'),  
(5,'jazzwallet'),  
(6,'easypay_voucher'),  
(7,'Easypay'),  
(8,'ublcrcard'),  
(9,'mygateway'),  
(10,'mcbllite'),  
(11,'cashatdoorstep'),  
(12,'internetbanking'),  
(13,'Easypay_MA'),  
(14,'productcredit'),  
(15,'marketingexpense'),  
(16,'financesettlement');
```

payment_detail

```
● CREATE TABLE IF NOT EXISTS order_detail (  
    id_order VARCHAR(50) PRIMARY KEY,  
    id_customer VARCHAR(10),  
    order_date DATE,  
    id_sku VARCHAR(10),  
    price INT,  
    qty_ordered INT,  
    before_discount FLOAT,  
    discount_amount FLOAT,  
    after_discount FLOAT,  
    is_gross INT,  
    is_valid INT,  
    is_net INT,  
    id_payment INT,  
    FOREIGN KEY (id_customer) REFERENCES customer_detail(id_customer),  
    FOREIGN KEY (id_sku) REFERENCES sku_detail(id_sku),  
    FOREIGN KEY (id_payment) REFERENCES payment_detail(id_payment)  
);  
● INSERT INTO order_detail VALUES  
(('ODR9939707760w','C713589L','2021-11-19 00:00:00','P858068',26100,200,5220000,2610000,2610000,1,1,0,5),  
(('ODR7448356649d','C551551L','2021-11-19 00:00:00','P886455',1971942,5,9859710,2464927.5,7394782.5,1,0,0,5),  
(('ODR4011281866z','C685596L','2021-11-25 00:00:00','P678648',7482000,1,7482000,2065344.62,5416655.38,1,0,0,4),  
(('ODR3378927994s','C830683L','2021-11-22 00:00:00','P540013',3593680,1,3593680,1455440.4,2138239.6,1,1,1,5),  
(('ODR4904430099k','C191766L','2021-11-21 00:00:00','P491032',4413220,1,4413220,1059172.8,3354047.2,1,1,1,4),  
(('ODR7618778722h','C299859L','2021-11-20 00:00:00','P886455',1971942,2,3943884,985971,2957913,1,0,0,5),  
(('ODR7610732813d','C313534L','2022-12-01 00:00:00','P849301',2697000,1,2697000,809100,1887900,1,1,1,1),  
(('ODR4415476736l','C215074L','2022-12-01 00:00:00','P918122',2533672,1,2533672,760101.6,1773570.4,1,1,0,3),
```

order_detail



DATABASE VALUE

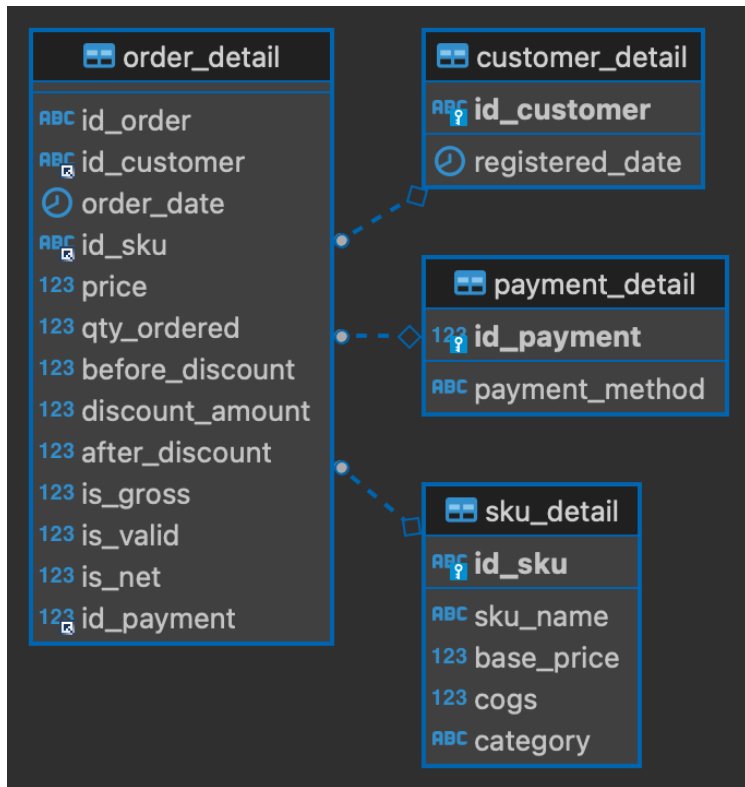
order_detail												
id_order	id_customer	order_date	id_sku	price	qty_ordered	before_discount	discount_amount	after_discount	is_gross	is_valid	is_net	id_payment
ODR9939707760w	C713589L	11/19/2021	P858068	26,100	200	5,220,000	2,610,000	2,610,000	1	1	0	5
ODR7448356649d	C551551L	11/19/2021	P886455	1,971,942	5	9,859,710	2,464,930	7,394,780	1	0	0	5
ODR4011281866z	C685596L	11/25/2021	P678648	7,482,000	1	7,482,000	2,065,340	5,416,660	1	0	0	4

customer_detail	
id_customer	registered_date
C107850L	8/3/2022
C110122L	8/14/2022
C111666L	8/3/2021

payment_detail	
id_payment	payment_method
1	cod
2	jazzvoucher
3	customercredit

sku_detail				
id_sku	sku_name	base_price	cogs	category
P100134	aki-GFSU1265-M	347,710	302,470	Women Fashion
P100314	Surmawala_PEL-PRAS-2500-Refrigerator-Grey	1,747,830	1,223,450	Appliances
P100667	Progressive_9781742488103	8,410	7,192	Books

ENTITY RELATIONSHIP DIAGRAM



In my database, there is a table called 'order_detail' which functions to record each customer order. So that we can view this order data more comprehensively, we have added three special tables that act as foreign keys. The first is 'customer_id,' which connects orders with customer data stored in the 'customer_detail' table. Then we have 'sku_id,' which is used to associate orders with product details in the 'sku_detail' table. Finally, there is 'id_payment,' which associates the order with the payment method available in the 'payment_detail' table. Through the use of foreign keys, we can combine data from three different tables and perform deeper analysis.

QUESTIONS



1. During transactions that occurred in 2021, in which month was the total transaction value (after_discount) the highest? Use is_valid = 1 to filter transaction data.
2. During transactions in 2022, which category generated the highest transaction value? Use is_valid = 1 to filter transaction data.
3. Compare the transaction values of each category in 2021 with 2022. Reflect on which categories experienced an increase and which categories saw a decrease in transaction value from 2021 to 2022. Use is_valid = 1 to filter transaction data.
4. Show the top 5 most popular payment methods used during 2022 (based on total unique orders). Use is_valid = 1 to filter transaction data.
5. Sort the transaction data for Samsung, Apple, Sony, Huawei, and Lenovo products by transaction value. Use is_valid = 1 to filter transaction data.





EXPLORATORY DATA ANALYSIS



NUMBER 1



Query :

```
⊖ -- number 1
SELECT
    MONTHNAME(order_date) month_name
    , ROUND(SUM(after_discount)) total_transaction_value
FROM
    order_detail od
WHERE
    EXTRACT(YEAR FROM order_date) = 2021
    and is_valid = 1
group by
    1
order by
    2 desc;
```



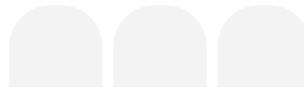
NUMBER 1



Result :

ABC month_name ▼	123 total_transaction_value ▼
August	227,862,744
December	217,309,963
October	207,603,260
November	180,396,010
July	148,007,735
September	145,943,335
June	43,154,552
January	36,822,126
February	35,611,797
May	34,163,856
March	23,643,062
April	22,208,473

The highest total transaction value for transactions occurring in 2021 was recorded in August.

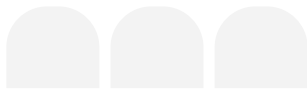


NUMBER 2



Query :

```
⊖ -- number 2
SELECT
    sd.category
    , ROUND(SUM(after_discount), 0) total_transaction_value
FROM
    order_detail od
LEFT JOIN
    sku_detail sd ON od.id_sku = sd.id_sku
WHERE
    EXTRACT(YEAR FROM order_date) = 2022
    AND is_valid = 1
GROUP BY
    1
ORDER BY
    2 DESC;
```



NUMBER 2

Result :

ABC category ▼	123 total_transaction_value ▼
Mobiles & Tablets	918,451,576
Entertainment	365,344,151
Appliances	316,358,100
Computing	214,028,543
Men Fashion	135,588,253
Women Fashion	93,014,971
Home & Living	79,483,716
Health & Sports	54,235,580
Beauty & Grooming	46,211,019
Superstore	32,643,267
Kids & Baby	25,931,277
Others	21,744,646
Soghaat	17,658,332
School & Education	17,362,465
Books	6,792,519

The category with the highest total transaction value for transactions carried out in 2022 is the Mobiles & Tablets category.

NUMBER 3



Query :

```
-- number 3
SELECT
    category,
    ROUND(SUM(CASE WHEN EXTRACT(YEAR FROM order_date) = 2021 THEN after_discount ELSE 0 END), 0) AS total_value_2021,
    ROUND(SUM(CASE WHEN EXTRACT(YEAR FROM order_date) = 2022 THEN after_discount ELSE 0 END), 0) AS total_value_2022,
    CASE WHEN
        ROUND(SUM(CASE WHEN EXTRACT(YEAR FROM order_date) = 2021 THEN after_discount ELSE 0 END), 0) <
        ROUND(SUM(CASE WHEN EXTRACT(YEAR FROM order_date) = 2022 THEN after_discount ELSE 0 END), 0)
    THEN 'Increase' ELSE 'Decrease' END AS progres
FROM
    order_detail od
JOIN
    sku_detail sd ON od.id_sku = sd.id_sku
WHERE
    od.is_valid = 1
GROUP BY
    1
HAVING
    total_value_2022 > total_value_2021
    OR total_value_2022 < total_value_2021;
```



NUMBER 3

Result :

ABC category ▼	123 total_value_2021 ▼	123 total_value_2022 ▼	ABC progres ▼
Others	40,468,516	21,744,646	Decrease
Appliances	218,550,177	316,358,100	Increase
Computing	172,878,860	214,028,543	Increase
Superstore	28,828,088	32,643,267	Increase
Mobiles & Tablets	370,606,718	918,451,576	Increase
Health & Sports	33,837,966	54,235,580	Increase
Women Fashion	84,045,961	93,014,971	Increase
Entertainment	162,326,357	365,344,151	Increase
Home & Living	45,797,873	79,483,716	Increase
Men Fashion	58,628,198	135,588,253	Increase
Beauty & Grooming	46,047,360	46,211,019	Increase
School & Education	11,558,982	17,362,465	Increase
Books	10,124,596	6,792,519	Decrease
Soghaat	15,056,203	17,658,332	Increase
Kids & Baby	23,971,058	25,931,277	Increase

There are two categories that experienced a decrease in total transactions from 2021 to 2022: the "Others" category and the "Books" category. Apart from these two categories, the transaction value increased.

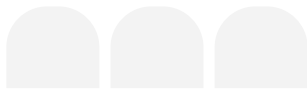
NUMBER 4



Query :

⊖ -- number 4

```
SELECT
    pd.payment_method,
    COUNT(DISTINCT od.id_order) AS total_unique_orders
FROM
    order_detail od
LEFT JOIN
    payment_detail pd ON od.id_payment = pd.id_payment
WHERE
    od.is_valid = 1
    AND YEAR(od.order_date) = 2022
GROUP BY
    pd.payment_method
ORDER BY
    2 DESC
LIMIT 5;
```



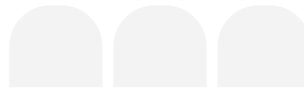
NUMBER 4



Result :

ABC payment_method ▼	123 total_unique_orders ▼
cod	1,809
Payaxis	181
customercredit	75
Easypay	69
jazzwallet	26

The 5 most popular payment methods used in transactions carried out in 2022 include: **COD, Payaxis, Customer Credit, Easypay and Jazzwallet.**



NUMBER 5

Query :

```
-- number 5
with pb as(
SELECT
CASE
WHEN LOWER(sd.sku_name) LIKE '%samsung%' THEN 'Samsung'
WHEN LOWER(sd.sku_name) LIKE '%apple%' OR LOWER(sd.sku_name) LIKE '%iphone%'
OR LOWER(sd.sku_name) LIKE '%macbook%' THEN 'Apple'
WHEN LOWER(sd.sku_name) LIKE '%sony%' THEN 'Sony'
WHEN LOWER(sd.sku_name) LIKE '%huawei%' THEN 'Huawei'
WHEN LOWER(sd.sku_name) LIKE '%lenovo%' THEN 'Lenovo'
ELSE 'Others'
END AS product_brand,
SUM(od.after_discount) AS total_transaction_value
FROM
order_detail od
LEFT JOIN
sku_detail sd ON od.id_sku = sd.id_sku
WHERE
od.is_valid = 1
GROUP BY
1
ORDER BY
2 DESC
)
SELECT * FROM pb WHERE product_brand != 'Others';
```

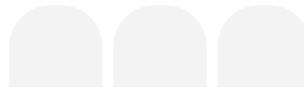
NUMBER 5



Result :

ABC product_brand ▼	123 total_transaction_value ▼
Samsung	588,764,150
Apple	444,855,360
Sony	63,960,718
Huawei	63,160,260
Lenovo	62,379,800.375

**Among the 5 product brands,
the Samsung product brand has
the highest total transaction value.**



Follow me!

Instagram : @abideee_

LinkedIn : [https://www.linkedin.com/in/abid-](https://www.linkedin.com/in/abid-Muhammad-taufiq/)

[Muhammad-taufiq/](#)

Bootcamp Data Analysis

by @myskill.id

