SQL – Ecommerce Analysis

Final Project Myskill | Luthfi Mahfuzh

Table Of Contents

MySkill

Overview

This SQL Final Project is one of practice to answer any business question with doing query with "Real data" in Industry.
With using SQL we can analyze any data to be insight to facing any question or business task with fast and efficient

In this case, we already provided sales dataset from MySkill team,
The dataset will be analyzed to be answered business question using SQL for handling any complex query and big database
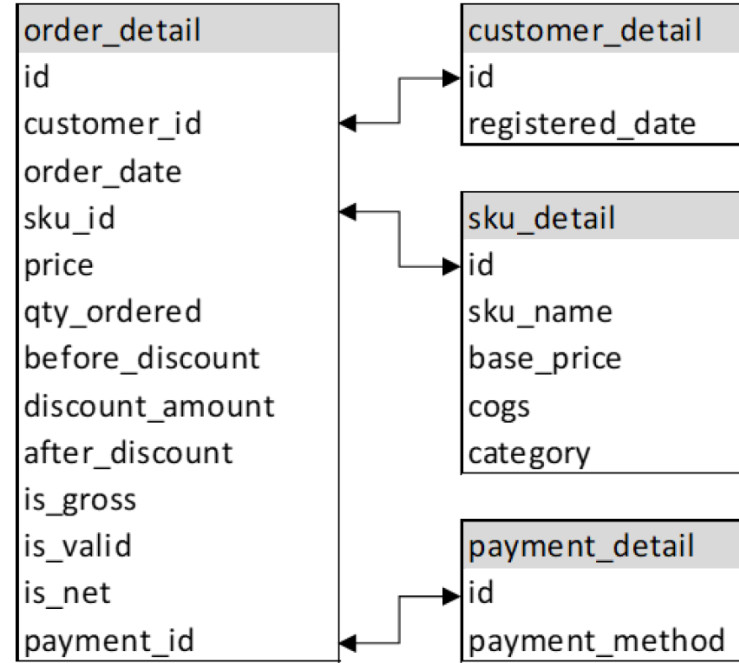
MySkill

Dataset

Data source that we used is sales data from one of e-commerce platform which already modified for exercise using SQL

The Dataset is consist of 4 datas :
1. Order Detail
2. Customer Detail
3. SKU Detail
4. Payment Detail

Table sceme

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

MySkill

Tools & Preparation

What is SQL ?

- SQL (Structured Query Language) is a standard programming language designed for managing relational databases
- It is used for storing, manipulating, and retrieving data stored in a relational database management system (RDBMS)
- SQL is widely used by database administrators, data analysts, and software developers for various tasks such as creating and modifying database schemas, inserting, updating, and deleting records, and querying the database to extract specific information

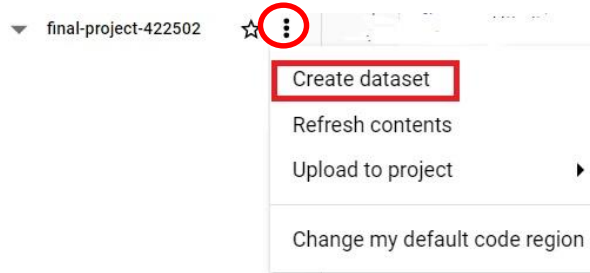Google Big Query

One of tools that we use is google Big Query

a fully managed, serverless data warehouse provided by Google Cloud Platform (GCP). It's designed to handle and analyze large-scale datasets using SQL-like queries
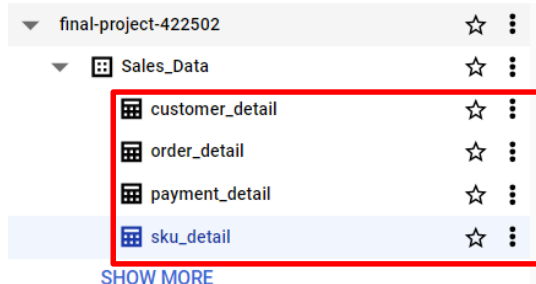
MySkill

Tools & Preparation

Import Data in BigQuery

1. Create Dataset

And then all table already import and ready to query

2. Import table

Question 1


```sql
1  -- Pertanyaan 1 Selama transaksi yang terjadi selama 2021, pada bulan apa total nilai transaksi(after_discount)
   paling besar? --
2
3  SELECT
4    EXTRACT (MONTH from order_date) Month,
5    SUM (after_discount) Transaction_Total
6  FROM
7    `final-project-422502.Sales_Data.order_detail`
8  WHERE
9      EXTRACT (YEAR from order_date) = 2021
10     AND is_valid = 1
11 GROUP BY
12     EXTRACT (MONTH from order_date)
13 ORDER BY
14     SUM (after_discount) DESC;
```

Query results

JOB INFORMATION · RESULTS · CHART · JSON · EXECUTION DETAILS · EXECUTION GRAPH

Row	Month	Transaction_Total
1	8	227862744.0
2	12	217309963.0
3	10	207603259.9
4	11	180396009.74
5	7	148007735.0
6	9	145943335.2
7	6	43154552.0
8	1	36822126.5
9	2	35611797.0
10	5	34163856.0

The highest transaction in 2021 is in august (8th month 227.862.744

Question 2


```
16  -- Pertanyaan 2 : Selama transaksi pada tahun 2022, kategori apa yang menghasilkan nilai transaksi paling besar? --
17
18  SELECT
19    sd.category,
20    SUM (od.after_discount) Transaction_Total -- aggregate after discount column in every rows
21  FROM
22    `final-project-422502.Sales_Data.order_detail` od
23  JOIN
24    `final-project-422502.Sales_Data.sku_detail` sd
25    ON
26    od.sku_id = sd.id -- JOIN data from table order detail and SKU detail
27  WHERE
28    EXTRACT (YEAR from order_date) = 2022 -- filter only for transaction in year 2022
29    and is_valid = 1
30  GROUP BY
31    1 -- Grouping for category
32  ORDER BY
33    2 DESC -- Sorting Descendants for Transaction Total
34  LIMIT
35    1 -- Displaying for the highest Transaction Total
```

Query results

JOB INFORMATION	RESULTS	CHART	JSON

Row	category	Transaction_Total
1	Mobiles & Tablets	918451576.0

The highest total transaction in 2022 is Mobiles & Tablets Category with value 918.451.576

MySkill

Question 3

```sql
1   -- Pertanyaan 3 : Bandingkan nilai transaksi dari masing-masing kategori pada tahun 2021 dengan 2022.Sebutkan kategori apa
    saja yang mengalami peningkatan dan kategori apa yang mengalami penurunan nilai transaksi dari tahun 2021 ke 2022 --
2
3   SELECT
4     category Category,
5     Transaction_Total_2021,
6     Transaction_Total_2022,
7     CASE
8       WHEN Transaction_Total_2021 < Transaction_Total_2022 THEN 'Increase'
9       WHEN Transaction_Total_2021 > Transaction_Total_2022 THEN 'Decrease'
10    END
11      as Trend_Status, -- for create trend status increase or decrease
12    ROUND (((Transaction_Total_2022 - Transaction_Total_2021)/ Transaction_Total_2021) * 100, 2) as Percent_Gap
13  FROM (
14    SELECT
15      sd.category,
16      round (SUM(CASE
17        WHEN EXTRACT (YEAR from od.order_date) = 2021 THEN od.after_discount
18        ELSE 0
19        END ), 2) Transaction_Total_2021, -- Aggregate for SUM transaction in 2021
20      round (SUM(CASE
21        WHEN EXTRACT (YEAR from od.order_date) = 2022 THEN od.after_discount
22        ELSE 0
23        END ), 2) Transaction_Total_2022  -- Aggregate for SUM transaction in 2022
24    FROM `final-project-422502.Sales_Data.order_detail` od
25    JOIN
26      `final-project-422502.Sales_Data.sku_detail` sd
27    ON
28      od.sku_id = sd.id
29    WHERE
30      od.is_valid = 1
31    GROUP BY
32      sd.category ) yearly_totals -- subquery for transaction total 2021 & 2022 with using JOIN
33  WHERE
34    Transaction_Total_2021 != Transaction_Total_2022; -- filter only for Transaction Total 2021
```

> **The highest increment from 2021 to 2022 is in category "Mobiles & Tablets" with 147 % increase in 2022 compare to 2021**

Query results

⬇ SAVE RESULTS ▾   📊 EXPLORE

JOB INFORMATION   RESULTS   CHART   JSON   EXECUTION DETAILS   EXECUTION GRAPH

Row	Category ▾	Transaction_Total_2	Transaction_Total_2	Trend_Status ▾	Percent_Gap ▾
1	Appliances	218550177.0	316358100.0	Increase	44.75
2	Computing	172878860.0	214028543.4	Increase	23.8
3	Superstore	28828088.0	32643266.52	Increase	13.23
4	Health & Sports	33837965.6	54235579.6	Increase	60.28
5	Mobiles & Tablets	370606718.0	918451576.0	Increase	147.82
6	Entertainment	162326357.4	365344148.9	Increase	125.07
7	Home & Living	45797873.0	79483716.2	Increase	73.55
8	Men Fashion	58628198.0	135588253.0	Increase	131.27
9	Others	40468515.74	21744646.02	Decrease	-46.27
10	Women Fashion	84045961.4	93014970.62	Increase	10.67
11	Beauty & Grooming	46047360.0	46211019.18	Increase	0.36
12	School & Education	11558982.4	17362465.3	Increase	50.21
13	Books	10124596.0	6792519.2	Decrease	-32.91
14	Soghaat	15056202.6	17658332.0	Increase	17.28
15	Kids & Baby	23971057.8	25931276.84	Increase	8.18

MySkill

Question 4

```sql
1  -- Pertanyaan 4 : Tampilkan top 5 metode pembayaran yang paling populer digunakan selama 2022, berdasarkan total unique
   order
2
3  SELECT
4    pd.payment_method Payment_Method,
5    count (DISTINCT od.id) Total_Unique_Orders -- Count total Unique Order
6  FROM
7    `Sales_Data.order_detail` od
8  JOIN
9    `Sales_Data.payment_detail` pd -- JOIN table order detail & payment detail based on payment id
10 ON
11   od.payment_id = pd.id
12 WHERE
13 EXTRACT (YEAR FROM od.order_date) = 2022 -- Filter for only transaction in year 2022
14 AND is_valid = 1
15 GROUP BY 1 -- grouping payment method
16 ORDER BY 2 DESC -- sorting total unique orders in descending
17 LIMIT 5 -- displaying top 5 payment method in 2022
```

Query results

Row	Payment_Method	Total_Unique_Orders
1	cod	1809
2	Payaxis	181
3	customercredit	75
4	Easypay	69
5	jazzwallet	26

The highest payment method in 2022 is "cod" with total orders is 1809 orders

MySkill

Question 5

```
91  -- Pertanyaan 5 : Urutkan dari ke-5 produk ini berdasarkan nilai transaksinya 1.Samsung, 2.Apple, 3.Sony, 4.Huawei, 5.Lenovo
92
93  SELECT
94    CASE
95      WHEN LOWER (sku_name) LIKE '%samsung%' THEN 'Samsung'
96      WHEN LOWER (sku_name) LIKE '%iphone%' OR LOWER (sku_name) LIKE '%macbook%' OR LOWER (sku_name) LIKE '%apple%' THEN 'Apple'
97      WHEN LOWER (sku_name) LIKE '%sony%' THEN 'Sony'
98      WHEN LOWER (sku_name) LIKE '%huawei%' THEN 'Huawei'
99      WHEN LOWER (sku_name) LIKE '%lenovo%' THEN 'Lenovo'
100     END as Brand,-- searching specific sku name with CASE WHEN Function, LOWER casing and LIKE THEN Function for create same format
101     SUM (od.after_discount) Total_Transaction -- Transaction Total after discount
102  FROM
103    `Sales_Data.order_detail` od
104  JOIN
105    `Sales_Data.sku_detail`sd -- JOIN Function for combine SKU detail and order detail based on sku id
106  ON
107  od.sku_id = sd.id
108  WHERE
109    EXTRACT (YEAR from order_date) = 2022 -- filter for only Transaction in year 2022 only
110    AND is_valid = 1
111    AND
112    (LOWER (sku_name) LIKE '%samsung%'
113    OR LOWER (sku_name) LIKE '%iphone%'
114    OR LOWER (sku_name) LIKE '%macbook%'
115    OR LOWER (sku_name) LIKE '%apple%'
116    OR LOWER (sku_name) LIKE '%sonv%'
117    OR LOWER (sku_name) LIKE '%huawei%'
118    OR LOWER (sku_name) LIKE '%lenovo%'
119    ) --filter for any value match with sku name with CASE WHEN Function, LOWER casing and LIKE Function
120    GROUP BY 1 -- group by Brand
121    ORDER BY 2 DESC; -- sort by Total Transaction
122
```

Query results

	JOB INFORMATION	RESULTS	CHART	JSON

Row	Brand ▾	Total_Transaction ↗
1	Samsung	412357844.0
2	Apple	187654592.0
3	Sony	32343178.0
4	Huawei	30942826.0
5	Lenovo	23621590.4

The highest transaction in 2022 is Samsung Brand with total transaction 412.357.844

MySkill

Please give me feedback to :
Email : luthfimahfuzh03@gmail.com
Telegram : @luthfimahfuzh
Linkedin : https://www.linkedin.com/in/luthfim/

Link for Query SQL : https://github.com/luthfimahfuzh27/portofolio

Thank You!

MySkill