

DATA ANALYSIS BASED ON HYPOTHESIS

1. Peningkatan penjualan produk

Metric: Rata-rata perkembangan profit dari tahun ke tahun (*average gain rate per year*)

1. Pseudocode SQL

```
with ProfitPerYear as (select EXTRACT(YEAR FROM Order_Date) AS Year,
                           SUM(Profit) AS Total_Profit
                        FROM
                           gleaming-advice-372312.KM_SEC_PALEMBANG_2024.Cleaned_table
                        GROUP BY
                           Year),
Prev AS (SELECT Year, Total_Profit AS Total_Profit_Prev,
                LAG(Total_Profit) OVER (ORDER BY Year) AS Prev_Profit
        FROM
            ProfitPerYear),
ProfitGrowth AS (SELECT Year,      Total_Profit_Prev AS Profit,
                        Prev_Profit, (Total_Profit_Prev - Prev_Profit) / Prev_Profit AS Growth
        FROM
            Prev)
SELECT Year, Profit, Prev_Profit, AVG(Growth) AS Avg_Profit_Growth
FROM
    ProfitGrowth
GROUP BY
    Year, Profit, Prev_Profit;
```

2. SQL Query

```
WITH ProfitPerYear AS (
    SELECT
        EXTRACT(YEAR FROM Order_Date) AS Year,
        SUM(Profit) AS Total_Profit
    FROM
        gleaming-advice-372312.KM_SEC_PALEMBANG_2024.Cleaned_table
    GROUP BY
        Year
),
Prev AS (
    SELECT
        Year,
        Total_Profit AS Total_Profit_Prev,
        LAG(Total_Profit) OVER (ORDER BY Year) AS Prev_Profit
    FROM

```

```

        ProfitPerYear
    ),
    ProfitGrowth AS (
        SELECT
            Year,
            Total_Profit_Prev AS Profit,
            Prev_Profit,
            (Total_Profit_Prev - Prev_Profit) / Prev_Profit AS Growth
        FROM
            Prev
    )
    SELECT
        Year,
        Profit,
        Prev_Profit,
        AVG(Growth) AS Avg_Profit_Growth
    FROM
        ProfitGrowth
    GROUP BY
        Year, Profit, Prev_Profit;

```

3. Hasil SQL

Row	Year	Profit	Prev_Profit	Avg_Profit_Growth
1	2017	93439.269600000174	81795.174300000086	0.14235675147892038
2	2016	81795.174300000086	61618.60370000002	0.32744284012394487
3	2015	61618.60370000002	49543.974100000065	0.24371540271736342
4	2014	49543.974100000065	<i>null</i>	<i>null</i>

Menurut data di atas, terdapat peningkatan profit secara signifikan dari tahun ke tahun, dengan rata-rata pertumbuhan profit selama periode sebesar 24%.

4. Analisis tambahan

- Pertumbuhan profit tampak melambat di tahun 2016 ke 2017, meskipun profit masih meningkat setiap tahun nya.
- Dengan rata-rata pertumbuhan profit sekitar 24%, menunjukan bahwa perusahaan telah berhasil meningkatkan profit secara konsisten.
- Ada hubungan antara profit tahun sebelumnya dengan profit tahun berikutnya. Semakin tinggi profit tahun sebelumnya, semakin besar kemungkinan profit tahun berikutnya juga akan tinggi.
- Statistik Deskriptif

Mean	Median	Max	Min
73148.25	71706.38	93439.26	49543.97

2. Pemberian diskon pada beberapa produk

Metric: ROI (Return on Investment)

1. Pseudocode SQL

```
WITH aggregated_data AS
(SELECT nama produk, SUM(Profit) AS Total profit, SUM(Sales * Discount)
AS diskon produk,
FROM `strong-province-416412.capstone.data_neww`
WHERE Discount > 0 GROUP BY nama produk)
SELECT nama produk, Total Profit / diskon produk AS ROI,
FROM aggregated_data, ORDER BY ROI DESC;
```

2. Query

```
WITH aggregated_data AS (
SELECT
Product_Name,
SUM(Profit) AS Total_Profit,
SUM(Sales*Discount) AS product_discount,
FROM `strong-province-416412.capstone.data_neww`
WHERE Discount > 0
GROUP BY Product_Name
)
```

```
SELECT
Product_Name,
Total_Profit / product_discount AS ROI,
FROM aggregated_data
ORDER BY ROI DESC;
```

3. Hasil SQL

Row	Product_Name	ROI
1	I.R.I.S IRISCard Anywhere 5 Card Scanner	4.333333333333333
2	Honeywell Enviracaire Portable HEPA Air Cleaner for up to 10 x 16 Room	2.0275229357798166
3	Honeywell Quietcare HEPA Air Cleaner	2.0192307692307692
4	Avery 490	1.8750000000000002
5	Alphabetical Labels for Top Tab Filing	1.8750000000000002
6	Avery 505	1.875
7	Memo Book, 100 Message Capacity, 5 3/8" x 11"	1.875
8	netTALK DUO VoIP Telephone Service	1.875
9	Recycled Interoffice Envelopes with String and Button Closure, 10 x 13	1.875
10	REDIFORM Incoming/Outgoing Call Register 11" X 8 1/2" 100 Messages	1.8749999999999998

Berdasarkan output, diperoleh bahwa ROI tertinggi mencapai 4.33333 dengan produk “I.R.I.S IRISCard Anywhere 5 Card Scanner” dari total 1657 produk. Adapun produk dengan ROI terendah merupakan produk “Eureka Disposable Bags for Sanitaire Vibra Groomer I Upright Vac” dengan ROI -3.4375.

4. Analisis tambahan

- Kategori dengan ROI tertinggi pada data *superstore* merupakan kategori “Technology” dengan ROI mencapai 0.107

Row	Category	ROI
1	Technology	0.10685193182573062
2	Office Supplies	-0.10490559297679759
3	Furniture	-0.32126712151717457

- Sub kategori dengan ROI tertinggi pada data *superstore* berupa “Labels” dengan ROI mencapai 1.731

Row	Sub_Category	ROI
1	Labels	1.7312039847823313
2	Paper	1.7226779819200713
3	Envelopes	1.6926818992727706
4	Fasteners	1.2373193834065788
5	Copiers	1.1145550268908466
6	Accessories	0.6781657534110177
7	Art	0.63003473579966562
8	Phones	0.21106053131391941
9	Chairs	0.0934876821878485...
10	Binders	-0.20785623310094248

- Barang berupa kategori “Technology” merupakan barang dengan ROI tertinggi. Hal ini dapat dilihat berdasarkan produk dengan ROI tertinggi merupakan produk dengan kategori “Technology”.
- Berdasarkan output, diketahui bahwa 2 dari 3 kategori memiliki ROI negatif, yaitu kategori “Furniture” dan “Office Supplies”. Kemudian, total subkategori yang memiliki ROI negatif yaitu sebanyak 8 dengan ROI sub kategori terendah adalah “Binders”. Selain itu, dari total 1657 produk, 542 diantaranya memiliki ROI negatif.

Row	Category	ROI
1	Furniture	-0.32126712151717457
2	Office Supplies	-0.10490559297679759

Row	Sub_Category	ROI
1	Supplies	-0.96183359921790745
2	Tables	-0.70152159961288785
3	Appliances	-0.67363535533831742
4	Machines	-0.50791333012010009
5	Bookcases	-0.45903828795919482
6	Furnishings	-0.43504947001205496
7	Storage	-0.32196960811305431
8	Binders	-0.20785623310094248

Row	Product_Name	ROI
1	Kensington 6 Outlet SmartSocket Surge Protector	-3.4375
2	Eureka Disposable Bags for Sanitaire Vibra Groomer I Upright Vac	-3.4375
3	Commercial WindTunnel Clean Air Upright Vacuum, Replacement Belts, Filtration Bags	-3.375
4	Hoover Commercial Lightweight Upright Vacuum with E-Z Empty Dirt Cup	-3.3749999999999996
5	Acco 6 Outlet Guardian Basic Surge Suppressor	-3.25
6	Acco 6 Outlet Guardian Premium Plus Surge Suppressor	-3.1875
7	Belkin F9G930V10-GRY 9 Outlet Surge	-3.1874999999999996
8	Belkin F9G930V10-GRY 9 Outlet Surge	-3.1874999999999996

- Statistik deskriptif

Mean	Median	Max.	Min.
0.395	0.437	4.333	-3.438

3. Peningkatan kuantitas produk

Metric: Clustering modus produk jenis apa yang sering terjual dan mengelompokkan produk dengan penjualan yang tinggi, menengah, dan rendah.

1. Pseudocode SQL

```
SELECT distinct subkategori,SUM(kuantitas), CASE
WHEN jumlah kuantitas ≥3158, THEN 'Tinggi'
WHEN jumlah kuantitas di rentang 906-3157, THEN 'Menengah'
ELSE 'Rendah'
END AS Kelompok
FROM `strong-province-416412.capstone.data_neww`
GROUP BY subkategori
ORDER BY kuantitas DESC
```

2. Query

```
SELECT
    distinct Sub_Category,
    SUM(Quantity) AS Quantity,
    CASE
        WHEN SUM(Quantity) >= 3158 THEN 'Tinggi'
        WHEN SUM(Quantity) BETWEEN 906 AND 3157 THEN 'Menengah'
        ELSE 'Rendah'
    END AS Kelompok
FROM `strong-province-416412.capstone.data_neww`
GROUP BY 1
ORDER BY 2 DESC
```

3. Hasil SQL

Row	Sub_Category	Quantity	Kelompok
1	Binders	5974	Tinggi
2	Paper	5178	Tinggi
3	Furnishings	3563	Tinggi
4	Phones	3289	Tinggi
5	Storage	3158	Tinggi
6	Art	3000	Menengah
7	Accessories	2976	Menengah
8	Chairs	2356	Menengah
9	Appliances	1729	Menengah
10	Labels	1400	Menengah

Berdasarkan output, dapat diketahui bahwa subkategori “Binders” merupakan produk yang paling banyak terjual sehingga dikelompokkan sebagai produk yang penjualannya tinggi.

4. Analisis Tambahan

- Terdapat sebanyak 5 sub kategori yang dikelompokkan sebagai produk yang penjualannya tinggi, yaitu “Binders”, “Paper”, “Furnishings”, “Phones”, dan “Storage”; 8 sub kategori dengan penjualan menengah, yaitu “Art”, “Accessories”, “Chairs”, “Appliances”, “Labels”, “Tabels”, “Fasteners”, dan “Envelopes”; dan 4 sub kategori dengan penjualan rendah, yaitu “Bookcases”, “Supplies”, “Machines”, dan “Copiers”.
- Berdasarkan data *superstore* yang telah dikelompokkan, diperoleh sebagai berikut.

Kategori	Profit	Sales
Tinggi	143129.033	927447.765
Menengah	88927.062	869432.900
Rendah	54340.926	500320.195

Berdasarkan tabel, dapat diketahui bahwa subkategori yang dikelompokkan menjadi penjualan tinggi ini memang menguntungkan bagi *store*. Hal ini juga dapat dibuktikan dengan margin untuk setiap pengelompokan sebagai berikut.

Margin Tinggi	Margin Menengah	Margin Rendah
0.154	0.102	0.109

- Statistik deskriptif

Mean	Median	Max	Min
2227.82	1729	5974	234

4. Durasi pengiriman produk di Kota Los Angeles pada tiap kategori pengiriman

Metric: Rata-rata durasi pengiriman berdasarkan mode pengiriman dan segmen di Kota Los Angeles

1. Pseudocode SQL

```
SELECT Segmen AS segmen_customer, Ship_Mode,
AVG(DATE_DIFF(Ship_Date, Order_Date, DAY)) AS rataaan_hari
FROM `strong-province-416412.capstone.data_neww` WHERE City='Los Angeles'
GROUP BY Segmen, Ship_Mode
ORDER BY 1;
```

2. Query

```

SELECT
    Segment AS segmen_customer,
    Ship_Mode,
    AVG(DATE_DIFF(Ship_Date, Order_Date, DAY)) AS rataan_hari
FROM `strong-province-416412.capstone.data_neww`
WHERE City='Los Angeles'
GROUP BY Segment, Ship_Mode
ORDER BY 1;

```

3. Hasil SQL

Row	segmen_customer	Ship_Mode	rataan_hari
1	Consumer	Same Day	0.0
2	Consumer	First Class	2.1428571428571423
3	Consumer	Second Class	3.1898734177215191
4	Consumer	Standard Class	5.0083682008368227
5	Corporate	Same Day	0.3571428571428571
6	Corporate	First Class	2.2571428571428576
7	Corporate	Second Class	3.3396226415094343
8	Corporate	Standard Class	5.2814814814814817
9	Home Office	Same Day	0.0
10	Home Office	First Class	1.7999999999999998

Berdasarkan output, diketahui bahwa untuk semua segmen customer, pengiriman produk tercepat menggunakan pengiriman mode *Same Day*. Adapun urutan rata-rata lama pengiriman berdasarkan mode pengiriman terlama ke tercepat yaitu *Standard Class*, *Second Class*, *First Class*, dan *Same Day*.

4. Analisis Tambahan

- Untuk semua segmen customer dan kategori barang, mode pengiriman tercepat menggunakan mode *Same Day* sebagai tercepat pertama dan disusul oleh *Standard Class*, *Second Class*, *First Class*, dan *Same Day*.
- Dalam rangka menaikkan kepuasan pelanggan dalam berbelanja di *superstore*, maka untuk semua mode pengiriman hendaknya dilakukan pengurangan hari dalam masa pengirimannya. Semakin cepat mode pengiriman, maka semakin meningkat pula kepuasan pelanggan dalam berbelanja di *superstore*. Melalui tindakan ini, maka pembelian produk oleh pelanggan diperkirakan akan meningkat. Hal ini berdampak secara tidak langsung dalam menaikkan profit *store*.
- Statistik deskriptif

Mean	Median	Max	Min
2.589	2.536	5.281	0.000

5. Segmentasi pelanggan pembelian kategori teknologi di Kota Los Angeles

Metric: Kebutuhan dan preferensi produk

Penalaran *Metric* : Analisis jumlah produk kategori teknologi yang kurang diminati, frekuensi pembelian, sub kategori, dan preferensi merk yang kurang diinginkan di kota Los Angeles. Sehingga hal ini dapat memungkinkan perusahaan untuk lebih memahami kebutuhan konsumen dan dapat menargetkan strategi pemasaran yang efektif dan sesuai dengan setiap kelompok pelanggan

1. Pseudocode SQL

```
WITH Technology_Products AS (
  SELECT
    Product_ID,
    Product_Name,
    Sub_Category,
    COUNT(*) AS Total_Purchases
  FROM
    km-revou-sql-class-418307.new.baru
  WHERE
    Category = 'Technology'
  GROUP BY
    Product_ID, Product_Name, Sub_Category
),
```

```
-- Menghitung frekuensi pembelian produk kategori teknologi
Purchase_Frequency AS (
  SELECT
    Customer_ID,
    COUNT(*) AS Total_Purchases_Technology
  FROM
    km-revou-sql-class-418307.new.baru
  WHERE
    Category = 'Technology'
  GROUP BY
    Customer_ID
),
```

-- Menghitung preferensi merk yang kurang diinginkan

Unwanted_Brands AS (

SELECT

Product_ID,

COUNT(*) AS Total_Purchases_Unwanted

FROM

km-revou-sql-class-418307.new.baru

WHERE

Category = 'Technology'

AND City = 'Los Angeles'

-- tambahkan syarat lain jika diperlukan

GROUP BY

Product_ID

),

-- Menggabungkan semua data untuk analisis

Customer_Segmentation AS (

SELECT

o.Customer_ID,

o.City,

o.Region,

tp.Product_ID,

tp.Product_Name,

tp.Sub_Category,

tp.Total_Purchases,

pf.Total_Purchases_Technology,

COALESCE(ub.Total_Purchases_Unwanted, 0) AS

Total_Purchases_Unwanted

FROM

km-revou-sql-class-418307.new.baru o

LEFT JOIN

Technology_Products tp ON o.Product_ID =

tp.Product_ID

LEFT JOIN

Purchase_Frequency pf ON o.Customer_ID =

pf.Customer_ID

LEFT JOIN

Unwanted_Brands ub ON o.Product_ID = ub.Product_ID

WHERE

o.Category = 'Technology'

AND o.City = 'Los Angeles'

)

```

-- Outputting hasil analisis
SELECT
    Customer_ID,
    City,
    Region,
    Product_ID,
    Product_Name,
    Sub_Category,
    Total_Purchases,
    Total_Purchases_Technology,
    Total_Purchases_Unwanted
FROM
    Customer_Segmentation
ORDER BY
    Total_Purchases DESC;

```

2. Query

```

WITH Technology_Products AS (
    SELECT
        Product_ID,
        Product_Name,
        Sub_Category,
        COUNT(*) AS Total_Purchases
    FROM
        `direct-terminal-417415.capstone.cleaned_table`
    WHERE
        Category = 'Technology'
    GROUP BY
        Product_ID, Product_Name, Sub_Category
),

-- Menghitung frekuensi pembelian produk kategori teknologi
Purchase_Frequency AS (
    SELECT
        Customer_ID,
        COUNT(*) AS Total_Purchases_Technology
    FROM
        `direct-terminal-417415.capstone.cleaned_table`

```

```

WHERE
    Category = 'Technology'
GROUP BY
    Customer_ID
),

-- Menghitung preferensi merk yang kurang diinginkan
Unwanted_Brands AS (
    SELECT
        Product_ID,
        COUNT(*) AS Total_Purchases_Unwanted
    FROM
        `direct-terminal-417415.capstone.cleaned_table`
    WHERE
        Category = 'Technology'
        AND City = 'Los Angeles'
        -- tambahkan syarat lain jika diperlukan
    GROUP BY
        Product_ID
),

```

```

-- Menggabungkan semua data untuk analisis
Customer_Segmentation AS (
    SELECT
        o.Customer_ID,
        o.City,
        o.Region,
        o.Product_ID,
        o.Product_Name,
        o.Sub_Category,
        tp.Total_Purchases,
        pf.Total_Purchases_Technology,
        COALESCE(ub.Total_Purchases_Unwanted, 0) AS Total_Purchases_Unwanted
    FROM
        `direct-terminal-417415.capstone.cleaned_table` o
    LEFT JOIN
        Technology_Products tp ON o.Product_ID = tp.Product_ID
    LEFT JOIN
        Purchase_Frequency pf ON o.Customer_ID = pf.Customer_ID
    LEFT JOIN
        Unwanted_Brands ub ON o.Product_ID = ub.Product_ID
    WHERE

```

```

        o.Category = 'Technology'
        AND o.City = 'Los Angeles'
    )

```

```
-- Outputting hasil analisis
```

```
SELECT
```

```

    Customer_ID,
    City,
    Region,
    Product_ID,
    Product_Name,
    Sub_Category,
    Total_Purchases,
    Total_Purchases_Technology,
    Total_Purchases_Unwanted

```

```
FROM
```

```
    Customer_Segmentation
```

```
ORDER BY
```

```
    Total_Purchases DESC;
```

3. Hasil SQL

Row	Product_ID	Product_Name	Sub_Category	Total_Purchases	Total_Purchases_Tech	Total_Purchases_Unwanted
1	:C-AC-10003038	Kingston Digital DataTraveler 16GB USB 2.0	Accessories	13	4	2
2	:C-AC-10003038	Kingston Digital DataTraveler 16GB USB 2.0	Accessories	13	5	2
3	:C-AC-10002323	SanDisk Ultra 32 GB MicroSDHC Class 10 Memory Card	Accessories	11	4	2
4	:C-AC-10002323	SanDisk Ultra 32 GB MicroSDHC Class 10 Memory Card	Accessories	11	4	2
5	:C-PH-10002923	Logitech B530 USB Headset - headset - Full size, Binaural	Phones	10	2	1
6	:C-PH-10001817	Wilson Electronics DB Pro Signal Booster	Phones	10	7	1
7	:C-AC-10003033	Plantronics CS510 - Over-the-Head monaural Wireless Headset System	Accessories	10	7	1
8	:C-PH-10002496	Cisco SPA301	Phones	9	4	2
9	:C-AC-10003832	Logitech P710e Mobile Speakerphone	Accessories	9	2	4

Berdasarkan output, dari produk “Kingston Digital DataTraveler” adalah produk yang paling diminati dari product Plantronics CS510. Selain itu, dapat dilihat dari kolom *total_purchases* menentukan produk apa yang paling diminati dan kolom *total_purchases_unwanted* menampilkan produk yang kurang diminati.

4. Analisis Tambahan

- Dari kolom *total_purchases*, dapat diketahui produk apa saja yang paling diminati dan kurang diminati oleh pelanggan di kota Los Angeles. Dari data terlihat mana Produk dengan jumlah pembelian yang tinggi dan penjualan rendah.
- Dengan menggabungkan semua informasi dari data yang dihasilkan, segmentasi pelanggan berdasarkan kebutuhan, preferensi, dan perilaku pembelian dapat dilakukan.

- Untuk produk yang kurang diminati atau produk dengan penjualan terendah dapat mengalokasikan sumber daya dengan efisien, seperti meningkatkan stok produk agar dapat dijadikan target untuk strategi pemasaran yang lebih agresif atau promosi khusus.
- Statistik deskriptif

Mean	Median	Max	Min
56.3	40.7	50	7.4