

E-commerce Data Analysis with SQL: Sales & Product Performance Insights

By: Restiani Mutoharoh

Project Goals for SQL Analysis



The primary goal of this project is to perform a data analysis on e-commerce transactions using SQL to generate relevant business insights. This analysis will include evaluating sales performance over time, identifying high-performing products and categories, and understanding customer behavior. More specifically, the objectives of this project are as follows:

- **Analyze Monthly Sales Performance:** To identify the month with the highest total discounted transaction value in 2021 to pinpoint the most effective sales period in terms of discounts.
- **Evaluate Top-Performing Product Categories:** To find the product categories with the largest transaction value in 2022 to determine which market segments are the most profitable.
- **Understand Long-Term Sales Trends:** To compare transaction values across categories between 2021 and 2022 to identify which categories are experiencing sales growth and decline, providing insights into market trends.
- **Identify Payment Method Preferences:** To display the top five most popular payment methods in 2022 based on the total unique transaction value, which is useful for optimizing payment strategies.
- **Analyze Specific Product Performance:** To sort and compare five specific products (Samsung, Apple, Sony, Huawei, Lenovo) based on their total transaction value to understand the competitive position of each product.

This project will produce a series of data-driven reports and conclusions that can be used for strategic decision-making, such as promotional strategies, inventory management, and future product development focus.





Data Source



For this analysis, we leverage a simulated e-commerce dataset that mirrors two years of operations from a fictional company, Tokopedia. The data is specifically designed for analytical purposes, structured into four relational tables: order_detail for capturing every transaction, sku_detail for product information, customer_detail for customer data, and payment_detail for payment methods. This data architecture enables us to efficiently extract valuable business insights.

Problem Statements



Q: During the transactions that occurred in 2021, which month had the highest total transaction value (after_discount)? Use `is_valid = 1` to filter the transaction data.



Q: During transactions in 2022, which category generated the highest transaction value? Use `is_valid = 1` to filter the transaction data.



Q: Compare the transaction values of each category in 2021 with 2022. State which categories experienced an increase and which categories experienced a decrease in transaction value from 2021 to 2022. Use `is_valid = 1` to filter the transaction data.



Q: Display the top 5 most popular payment methods used during 2022 (based on the total number of unique orders). Use `is_valid = 1` to filter the transaction data.

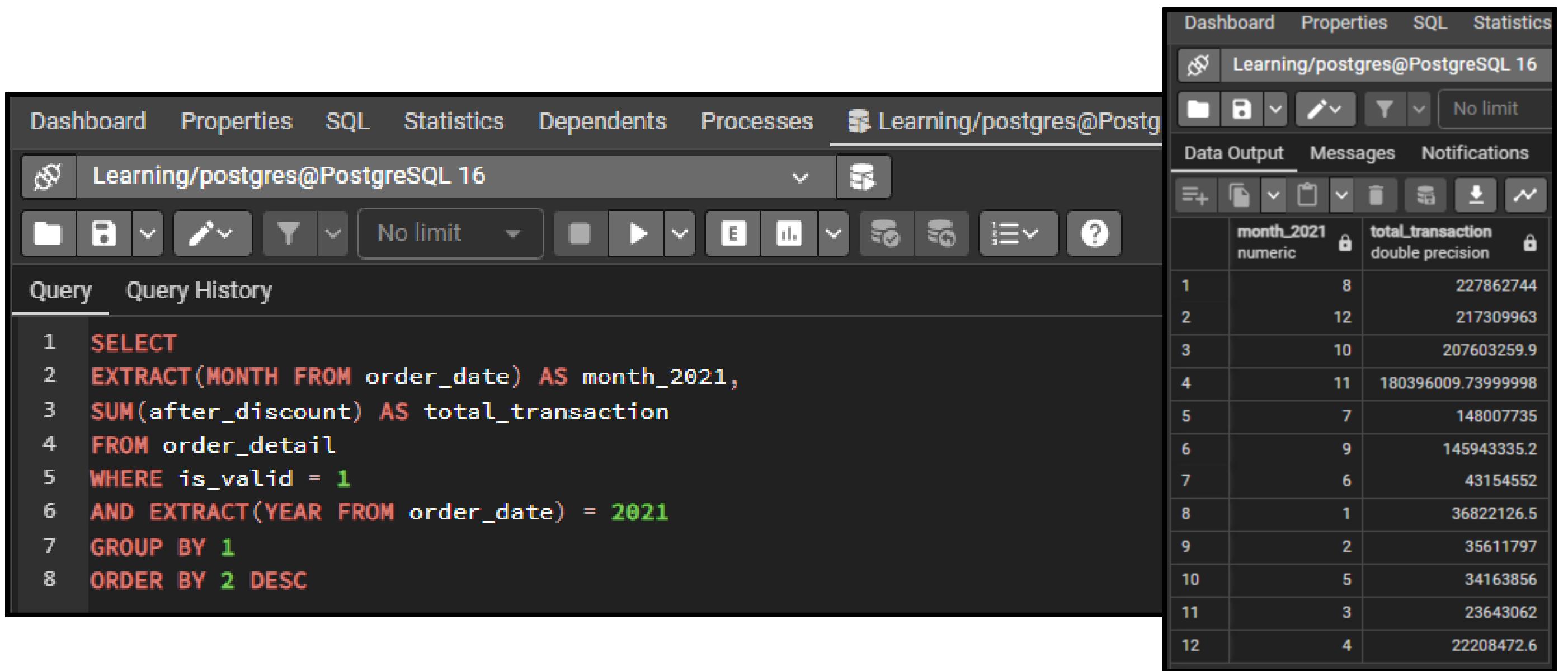


Q: Rank these 5 products based on their transaction value:

1. Samsung
2. Apple
3. Sony
4. Huawei
5. Lenovo

Use `is_valid = 1` to filter the transaction data.

Monthly Performance Analysis for 2021



The image shows a screenshot of a PostgreSQL client interface. On the left, there is a dark-themed window containing a SQL query and its execution results. The SQL query is:

```
1 SELECT
2 EXTRACT(MONTH FROM order_date) AS month_2021,
3 SUM(after_discount) AS total_transaction
4 FROM order_detail
5 WHERE is_valid = 1
6 AND EXTRACT(YEAR FROM order_date) = 2021
7 GROUP BY 1
8 ORDER BY 2 DESC
```

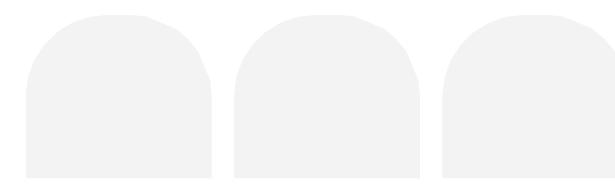
On the right, there is a light-themed window showing the results of the query. The results are presented in a table with two columns: 'month_2021' and 'total_transaction'. The data is as follows:

month_2021	total_transaction
8	227862744
12	217309963
10	207603259.9
11	180396009.73999998
7	148007735
9	145943335.2
6	43154552
1	36822126.5
2	35611797
5	34163856
3	23643062
4	22208472.6



Insight

1. November Mega-Sale Dominance (11.11): November (11) is a very dominant peak sales period, recording a total of nearly 1.8 billion transactions. Sales in November are 8 times higher than the second highest month (August), and far exceed the peak month for year-end shopping (December).
2. Effectiveness of Major Promotional Strategies: The massive surge in November shows that the discount strategies, marketing campaigns, and major promotions implemented during the 11.11 (Mega-Sale) period were highly effective in driving overall transaction volume, even at discounted prices (after_discount).
3. Extreme Performance Gap: There is a huge gap between the big promotion month (November) and normal months. This indicates that e-commerce sales are highly dependent on event periods.
4. Year-End Sales (December): Sales in December (12) were relatively good, but ranked third, below November and August, indicating that the 11.11 event may have “stolen” most of the customer spending budget that should have been allocated for December.





Recommendations

1. Maximize November Preparation

- Inventory: Allocate the majority of your stock and supply chain management budget to ensure the availability of leading products before November, especially for the best-selling categories (which we will see in Query 2). Avoid stock-outs so as not to lose potential revenue of 1.8 billion.
- Marketing Budget: Allocate the largest marketing budget in October and November to ensure maximum campaign effectiveness and gain customer mindshare before competitors.

2. Optimize Cross-Selling and Upselling in November:

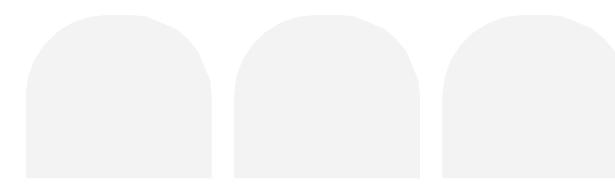
- Focus on increasing AOV (Average Order Value) during November. Since customers are already in shopping mode, encourage them to buy more with bundling strategies or “buy X, get Y” offers to increase margins, rather than just offering pure discounts.

3. Analyzing and Strengthening ‘Non-Event’ Months:

- Conduct an in-depth analysis of August (8) and October (10), which ranked second and fourth. Identify what factors (other than mega-sales) drove their sales. If there are no major events, replicate the small initiatives (such as weekly flash sales or thematic campaigns) that were successful in those months to months with very low performance (such as April/4 or March/3).

4. December Sales Security Strategy.

- Consider shifting the focus of your December campaign from aggressive discounts to last-minute offers and gifts (e.g., free shipping on all purchases, free gift wrapping). The goal is to capture shoppers who are looking for Christmas/end-of-year gifts and may have already spent their discount budget in November.



Analysis of the Best-Selling Product Categories in 2022

The image shows a PostgreSQL database interface with two main panes. The left pane displays a SQL query and its execution results. The right pane shows a table of data output.

Query:

```
1 SELECT
2     sku_detail.category,
3     SUM(order_detail.after_discount) AS total_transaction
4 FROM order_detail
5 INNER JOIN sku_detail
6     ON order_detail.sku_id = sku_detail.id
7 WHERE order_detail.is_valid = 1
8 AND EXTRACT(YEAR FROM order_detail.order_date) = 2022
9 GROUP BY 1
10 ORDER BY 2 DESC
```

Data Output:

	category	total_transaction
1	Mobiles & Tablets	918451576
2	Entertainment	365344148.9
3	Appliances	316358100
4	Computing	214028543.4
5	Men Fashion	135588253
6	Women Fashion	93014970.62
7	Home & Living	79483716.2
8	Health & Sports	54235579.6
9	Beauty & Grooming	46211019.18
10	Superstore	32643266.52
11	Kids & Baby	25931276.84
12	Others	21744646.02
13	Soghaat	17658332
14	School & Education	17362465.3
15	Books	6792519.2



Insight

1. “Mobiles & Tablets” is the Main Cash Cow: The Mobiles & Tablets category is an overwhelmingly dominant revenue driver. With a transaction value of 918 million, this category far exceeds the combined total of all other categories, making it the core of this e-commerce business model.
2. The Power of Necessity and Lifestyle Categories: After Mobile & Tablets, there are two other strong sales pillars: Women's Fashion and Home & Living. This shows that your platform is not only strong in the High-Value Item (Electronics) sector, but also in the Daily Necessities/Lifestyle sector, which tends to have repeat purchase frequency.
3. Computing Sector Potential: Although Computing ranks sixth (21.4 Million), this category often has a high Average Order Value (AOV) and healthy margins, indicating a strong technology customer base with room for sales growth.
4. Other Categories (Entertainment, Appliances): These categories are in the middle tier. They contribute significantly, but do not have the appeal of the top three categories.





Recommendations

1. Protection and Growth of Cash Cows (Mobiles & Tablets):

- Vendor Partnership Strategy: Use these sales volumes as leverage to negotiate better purchase prices (lower COGS) or larger marketing budget support from major manufacturers (Samsung, Apple, etc.).
- Exclusive Supply Chain: Ensure the most efficient and exclusive supply chain for products in this category, as they are the largest source of revenue.

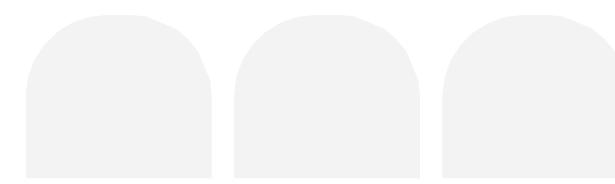
2. Frequency Optimization (Women's Fashion & Home & Living):

- Loyalty Program: Since this category tends to have shorter purchase cycles, focus on loyalty programs and retargeting. Encourage repeat purchases by offering small discounts after the first purchase or seasonal promotions (e.g., Rainy Season collection for Women's Fashion).
- Lifestyle Bundles: For Home & Living, suggest product bundling (e.g., Buy Bed Sheets + Get 50% Off Pillows) to increase AOV in this category.

3. Computing Category Improvement Strategy:

- Specialization Campaign: Launch campaigns specifically targeted at gamers or Work From Home (WFH) workers. Offer computing products with bundled services (e.g., free software installation) to differentiate yourself from competitors.

4. Next Steps: Use the results from Query 3 (2021 vs. 2022 Trends) to see if these top categories are experiencing growth or stagnation/decline. Dominant but declining categories are an alarm that must be addressed immediately.



Comparison of Transaction Values by Category 2021 vs. 2022

The screenshot shows a PostgreSQL database interface with two queries. The first query creates two temporary tables, year_2021 and year_2022, each summing the total transaction value for each category in 2021 and 2022 respectively. The second query then joins these two tables to compare the total transaction values for each category, using a CASE statement to determine if the value increased ('Naik'), decreased ('Turun'), or remained the same ('Sama atau tidak diketahui').

```
1 WITH year_2021 AS (
2     SELECT
3         sku_detail.category,
4         SUM(order_detail.after_discount) AS total_transaction_2021
5     FROM order_detail
6     INNER JOIN sku_detail
7         ON order_detail.sku_id = sku_detail.id
8     WHERE order_detail.is_valid = 1
9         AND EXTRACT(YEAR FROM order_detail.order_date) = 2021
10    GROUP BY 1
11 ), year_2022 AS (
12     SELECT
13         sku_detail.category,
14         SUM(order_detail.after_discount) AS total_transaction_2022
15     FROM order_detail
16     INNER JOIN sku_detail
17         ON order_detail.sku_id = sku_detail.id
18     WHERE order_detail.is_valid = 1
19         AND EXTRACT(YEAR FROM order_detail.order_date) = 2022
20    GROUP BY 1
21 )
22 SELECT
23     year_2021.category,
24     year_2021.total_transaction_2021,
25     year_2022.total_transaction_2022,
26     CASE
27         WHEN year_2021.total_transaction_2021 > year_2022.total_transaction_2022 THEN 'Naik'
28         WHEN year_2021.total_transaction_2021 < year_2022.total_transaction_2022 THEN 'Turun'
29         ELSE 'Sama atau tidak diketahui'
30     END AS category_status
31 FROM year_2021
32 JOIN year_2022
33     ON year_2021.category = year_2022.category;
```

Dashboard Properties SQL Statistics Dependencies Dependents Processes

Learning/postgres@PostgreSQL 16 No limit

Data Output Messages Notifications

	category text	total_transaction_2021 double precision	total_transaction_2022 double precision	category_status text
1	Appliances	218550177	316358100	Turun
2	Beauty & Grooming	46047360	46211019.18	Turun
3	Books	10124596	6792519.2	Naik
4	Computing	172878860	214028543.4	Turun
5	Entertainment	162326357.4	365344148.9	Turun
6	Health & Sports	33837965.6	54235579.6	Turun
7	Home & Living	45797873	79483716.2	Turun
8	Kids & Baby	23971057.8	25931276.84	Turun
9	Men Fashion	58628198	135588253	Turun
10	Mobiles & Tablets	370606718	918451576	Turun
11	Others	40468515.74	21744646.02	Naik
12	School & Education	11558982.4	17362465.3	Turun
13	Soghaat	15056202.6	17658332	Turun
14	Superstore	28828088	32643266.52	Turun
15	Women Fashion	84045961.4	93014970.62	Turun



Insight

1. Exponential Growth in Major Categories:

- Mobiles & Tablets (No. 1 category in 2022) not only dominated, but also showed remarkable growth of +147.8% (from 370 million to 918 million). This shows that investment and focus on this category in 2022 has yielded fantastic results.
- Entertainment, Men's Fashion, and Home & Living also showed very strong growth, at +125.1%, +130.5%, and +73.6% respectively. This proves that sales diversification beyond the core Mobile & Tablets business is going well.

2. Category with the Highest Percentage Growth: Books recorded the highest percentage growth (+580%), albeit from a very small transaction base (0.1 million to 0.68 million). This indicates new demand in this segment that needs to be recognized.

3. Alarm for the “Others” and “Beauty & Grooming” categories:

- Others experienced a drastic decline of -46.3% (-187 million). This is a significant loss and must be investigated. Since this is the “Others” category, it is likely that products that were not properly classified in 2021 have now been moved to the appropriate category in 2022, causing a decline in Others (this is data cleaning insight).
- Beauty & Grooming experienced an extreme revenue loss of -90%. The decline from IDR 46 million to IDR 4.6 million is catastrophic. This indicates a strategic failure, a shift in customer interest, or a loss of competitive advantage in this market.





Recommendations

1. High Growth Category Strengthening Strategy:

- **Mobiles & Tablets:** Maintain momentum! Immediately negotiate the addition of exclusive stock keeping units (SKUs) or the latest models with vendors to ensure this platform remains the top destination for gadgets.
- **Entertainment, Men's Fashion, Home & Living:** These categories are the new stars. Shift part of the marketing budget from general campaigns to highly targeted campaigns for these three categories to accelerate market penetration.

2. Emergency Intervention for Beauty & Grooming:

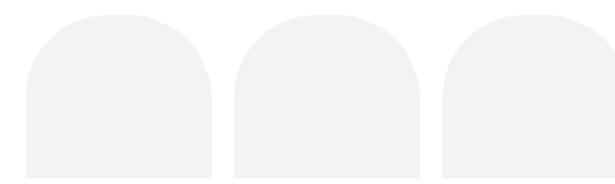
- **In-depth Audit:** Conduct an immediate audit for Beauty & Grooming. Compare prices, SKUs, and brands offered in 2022 with competitors. Have popular brands stopped selling? Has there been an increase in COGS that makes prices uncompetitive?
- **Reactivation:** Launch flash sales or aggressive Beauty & Grooming product bundles to attract back old customers who may have switched to other platforms.

3. Opportunities for Scaling Books:

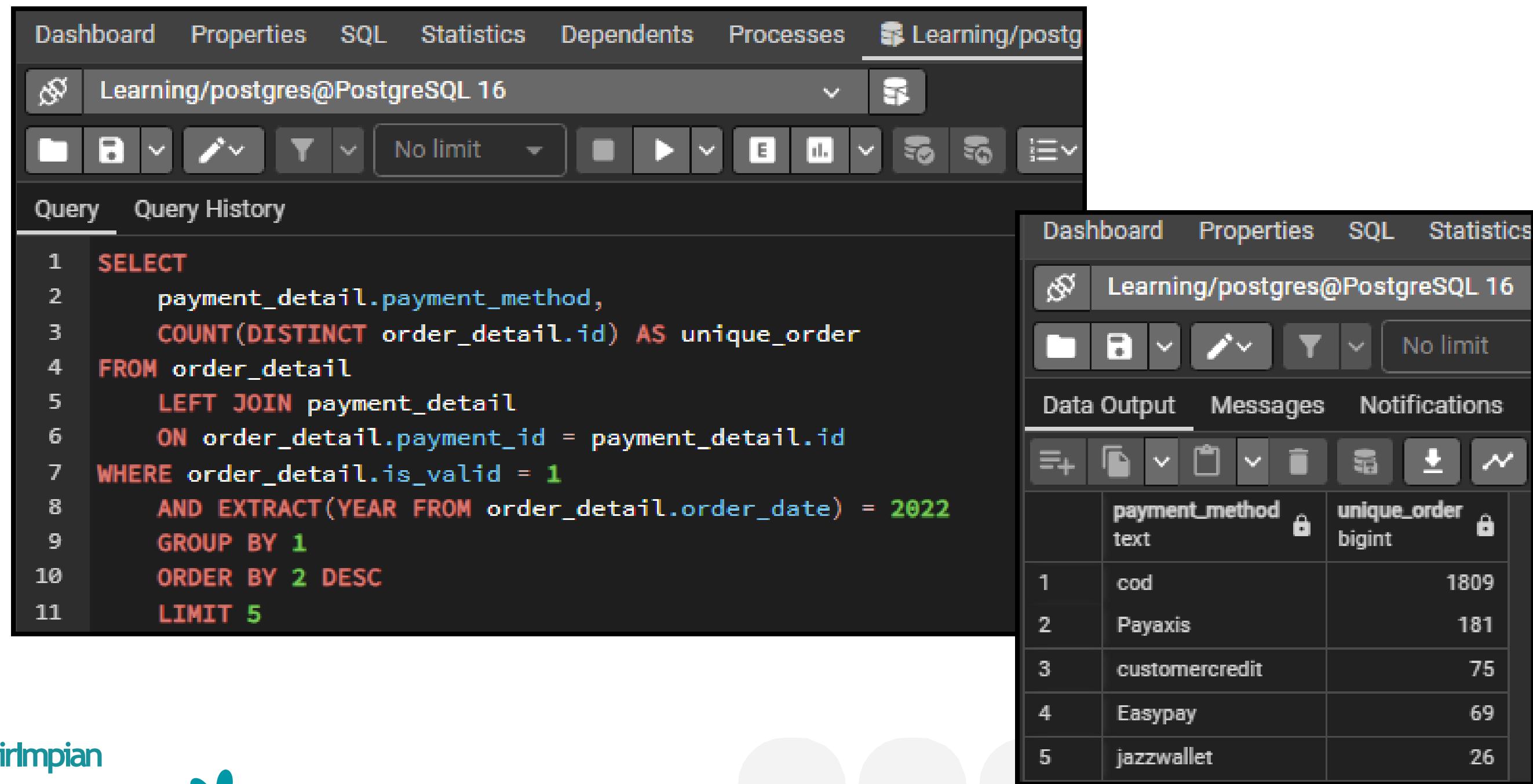
- **Smart Expansion:** Despite its small base, Books' 580% growth indicates strong demand. Instead of large investments, explore partnerships with major publishers to offer digital catalogs (e-books) or exclusive physical books without having to bear the risk of large inventories.

4. Clarification of the “Others” Category:

- Recommend that the Data/BI team create more specific categories for products in “Others” 2021 remaining in 2022. This will make future reporting and trend analysis cleaner and more accurate.



Analysis of the Most Popular Payment Methods in 2022



The image shows a PostgreSQL database interface with two main panes. The left pane displays a SQL query for analyzing payment methods. The right pane shows the resulting data output.

Query:

```
1 SELECT
2     payment_detail.payment_method,
3     COUNT(DISTINCT order_detail.id) AS unique_order
4 FROM order_detail
5     LEFT JOIN payment_detail
6         ON order_detail.payment_id = payment_detail.id
7 WHERE order_detail.is_valid = 1
8     AND EXTRACT(YEAR FROM order_detail.order_date) = 2022
9 GROUP BY 1
10 ORDER BY 2 DESC
11 LIMIT 5
```

Data Output:

	payment_method	unique_order
1	cod	1809
2	Payaxis	181
3	customercredit	75
4	Easypay	69
5	jazzwallet	26



Insight

1. Cash on Delivery (COD) Dominance: COD is not only the most popular payment method, but also dominates by a huge margin. With 1809 unique orders, COD is almost 10 times more popular than Payaxis (ranked 2nd).
2. Indications of Customer Trust: The massive reliance on COD suggests that:
 - Most customers may be in areas where digital banking services are not yet widespread (unbanked/underbanked).
 - There is a level of doubt or lack of trust in the platform (or delivery service), so they prefer to pay after the goods have been received and inspected.
3. Digital Methods Still Minority: Although there are several digital/credit payment methods (Payaxis, customercredit, Easypay, jazzwallet), their combined contribution is still very small compared to COD. This indicates that companies have not fully adopted or promoted digital payments effectively, or that the main market still requires COD.
4. Operational Challenges: High dependence on COD increases operational risks (unpaid returns, double logistics costs, longer payment settlement times).





Recommendations

1. Managing and Optimizing COD (Short Term):

- Cost Audit: Conduct a comprehensive cost analysis for each COD order. Ensure that operational, logistics, and unpaid risk costs are taken into account in the pricing strategy.
- Delivery Quality Tracking: Since COD is closely tied to delivery, increase oversight of the service quality of logistics partners. Poor delivery quality can directly impact COD payment rejections.

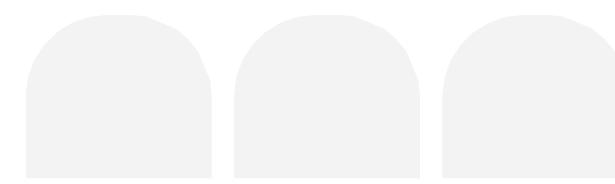
2. Encouraging Migration to Digital Payments (Medium Term):

- Digital Payment Incentives: Launch an aggressive incentive campaign for non-COD payments. Example: Free Shipping or an Additional 5% Discount for transactions using Payaxis or other e-wallet methods.
- Customer Credit Promotions: Since customer credit is in the Top 5, collaborate with credit providers to offer 0% installment promotions or pay-later schemes for high-value purchases (such as mobiles and tablets), which are likely to increase AOV and reduce the burden of COD.

3. Improving Trust (Long Term):

- Transparent Return Policy: Promote a very easy and transparent return policy for goods and money. If customers feel secure that their money will be returned if there are any problems, they may be more willing to pay digitally in advance.

4. Next Steps: Combine this data with Query 5. What payment methods are most commonly used for expensive products (Samsung, Apple, etc.)? If COD dominates for expensive products, the company's risk of loss will be higher.



Product Brand Performance Analysis

The screenshot shows the pgAdmin interface for PostgreSQL 16. The left panel displays a SQL query for analyzing product brand performance:

```
1 WITH transaction_table AS (
2     SELECT
3         CASE
4             WHEN lower(sku_detail.sku_name) LIKE '%samsung%' THEN 'Samsung'
5             WHEN lower(sku_detail.sku_name) LIKE '%apple%' THEN 'Apple'
6             WHEN lower(sku_detail.sku_name) LIKE '%sony%' THEN 'Sony'
7             WHEN lower(sku_detail.sku_name) LIKE '%huawei%' THEN 'Huawei'
8             WHEN lower(sku_detail.sku_name) LIKE '%lenovo%' THEN 'Lenovo'
9             ELSE 'Others'
10        END AS product_brand,
11        SUM(order_detail.after_discount) AS total_transaction
12    FROM order_detail
13    LEFT JOIN sku_detail
14        ON order_detail.sku_id = sku_detail.id
15    WHERE is_valid = 1
16    GROUP BY 1
17    ORDER BY 2 DESC
18 )
19     SELECT *
20     FROM transaction_table
```

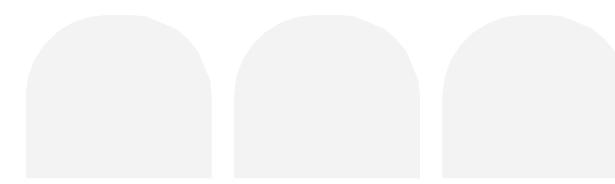
The right panel shows the results of the query, which is a table with two columns: `product_brand` (text) and `total_transaction` (double precision). The data is as follows:

	product_brand	total_transaction
1	Samsung	588764148
2	Apple	213310080
3	Sony	63960718
4	Huawei	63160260
5	Lenovo	62379800.4



Insight

1. Samsung is the Market Leader (Brand): Samsung is the best performing brand among the five brands tested. Its transaction value is nearly three times that of its closest competitor, Apple (588 million vs. 213 million). This demonstrates Samsung's strong dominance in the customer base of this platform.
2. Class Gap: There is a huge gap between the two market leaders (Samsung and Apple) and the next three brands (Sony, Huawei, Lenovo). The transaction values of Sony, Huawei, and Lenovo are in the range of 60 million, which means that the combined performance of the three is still far below that of Apple alone.
3. Hidden Potential in ‘Others’: The Others category recorded total transactions that far exceeded the combined total of all brands analyzed (2.6 billion). This indicates that:
 - Most sales came from unclassified brands/products or small/local brands not covered in the CASE statement.
 - This query focuses only on five brands and ignores other brands that may be equally important (such as Xiaomi, Oppo, Vivo, etc., which may fall under ‘Others’).





Recommendations

1. Strengthen Samsung Partnership (No. 1 Brand):

- Exclusive Program: Leverage this dominance to negotiate exclusive product launches or pre-orders for Samsung only on your platform. This will attract a larger volume of customers.
- Co-Marketing: Allocate a larger co-marketing budget with Samsung to maintain and increase their market share, especially during peak months (November, such as Query 1).

2. Different Strategy for Apple (Brand No. 2):

- Focus on Margins: Apple products tend to have high customer loyalty and stable selling prices. Focus your strategy on margins and value-added services (e.g., bundling accessories, extended warranties, or trade-in services) rather than deep discounts.

3. Intervention in Lower-Ranking Brands:

- Cost vs. Return Audit: For brands such as Sony, Huawei, and Lenovo, evaluate whether the profit margin is worth the effort of inventory and marketing management.
- Niche Strategy: Instead of competing directly, encourage these brands to target specific niche markets. For example, market Lenovo specifically to students or work-from-home (WFH) professionals looking for affordable products.

4. Critical Step: In-Depth Analysis of the “Others” Category:

- Advanced Query: Run an advanced query to identify the top brands hidden within the “Others” category. The goal is to reclassify high-transaction-value brands into their own column. These brands, which contribute 2.6 billion, are likely to be unmeasured growth engines at this time.

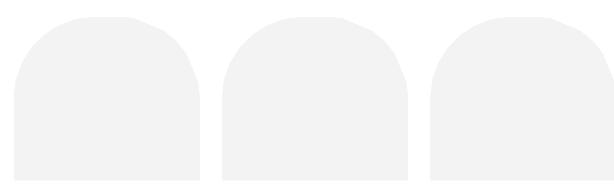


Executive Summary



Key Business Insights

1. Extreme Dependence on Mega-Sales (November): Annual sales performance is heavily dependent on November (11.11), which dominated 2021 sales by 8 times that of a normal month.
2. Electronics is a Double Cash Cow: The Mobiles & Tablets category is the largest category (Query 2) and the fastest growing category (+147.8% from 2021 to 2022, Query 3). Within it, Samsung is the most dominant brand among major brands.
3. Customer Trust Challenge: The majority of customers (1,809 unique orders) still prefer COD over digital payment methods (Query 4), indicating the need for increased trust or stronger digital incentives.





Priority Strategic Recommendations

1. Cash Cow Risk Mitigation & Strengthening Strategy:

- Action: Allocate R&D/Data resources to identify major brands hidden in the ‘Others’ category (2.6 billion) and classify them immediately. Negotiate exclusive partnerships with Samsung and other major vendors.

2. Emergency Intervention for Loss-Making Categories:

- Action: Conduct an immediate audit and reactivation campaign for the Beauty & Grooming category, which experienced a 90% decline in transactions from 2021 to 2022.

3. Aggressive Digital Payment Migration:

- Action: Introduce strong digital payment incentives (such as a 5% discount or exclusive free shipping) to reduce dependence on COD. This will reduce operational costs, the risk of returns, and accelerate the company's cash flow.

End.

Thank You