Sales Data Analysis Using SQL

Insights for Business Growth

AtliQ Hardware

COMPUTER HARDWARE PRODUCERS

Presentation Roadmap

Data Analysis Using SQL

Business Introduction - About AtliQ Hardware Business

AtliQ Hardware - Database Structure

Insights from Key Tables - Products, Customers and Sales
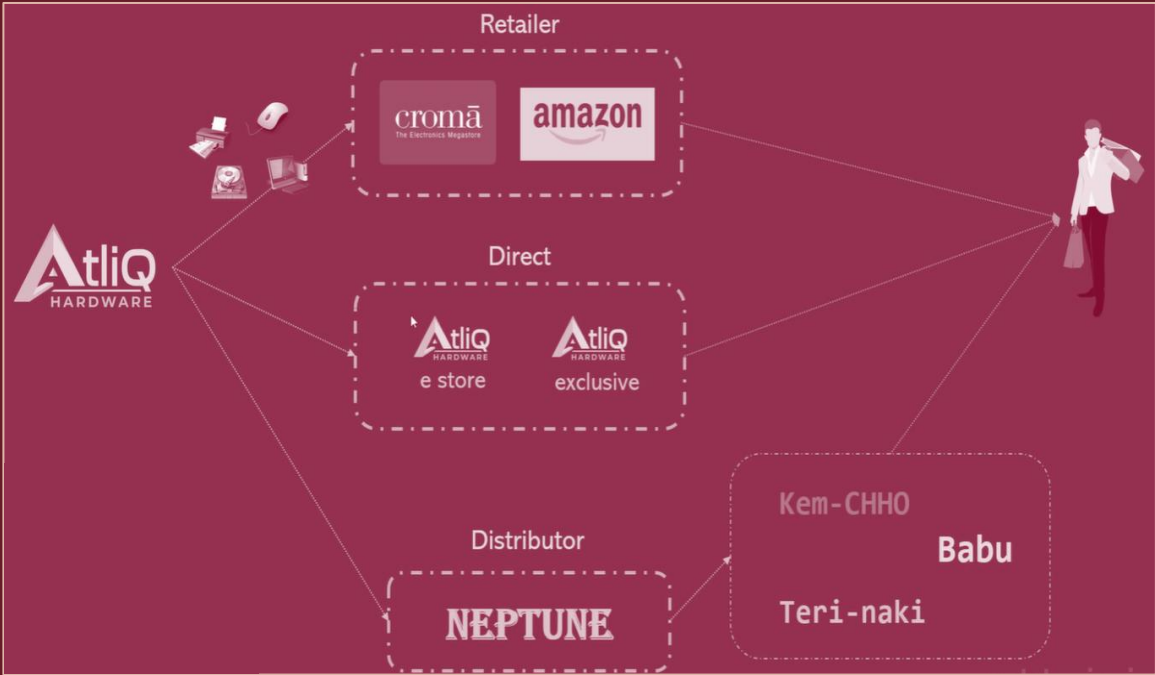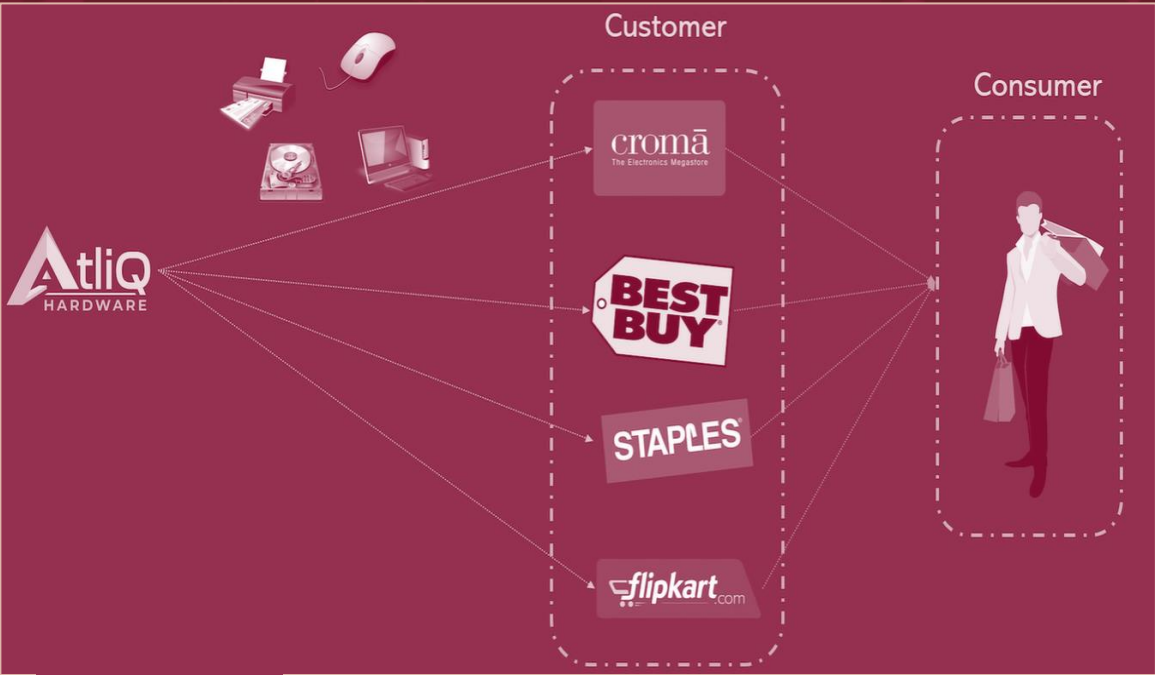
SQL Flow and Structure

- SQL Execution Order
- Key SQL Techniques (JOINs, Window Functions, Conditional Statements)
- Stored Procedures

Data-Driven Insights and Analysis - Insightful Reports

Business Introduction

AtliQ Hardware's (imaginary company) is one of the leading computer hardware producers in India and well expanded in other countries too.

AtliQ Hardware - Database

Dimension Tables

customer_code	customer	platform	channel	market	sub_zone	region
70002017	Atliq Exclusive	Brick & Mortar	Direct	India	India	APAC
70002018	Atliq e Store	E-Commerce	Direct	India	India	APAC
70003181	Atliq Exclusive	Brick & Mortar	Direct	Indonesia	ROA	APAC
70003182	Atliq e Store	E-Commerce	Direct	Indonesia	ROA	APAC
70004069	Atliq Exclusive	Brick & Mortar	Direct	Japan	ROA	APAC

dim_customer

product_code	division	segment	category	product	variant
A0118150101	P & A	Peripherals	Internal HDD	AQ Dracula HDD – 3.5 Inch SATA 6 Gb/s 5400 RPM 256 MB Cache	Standard
A0118150102	P & A	Peripherals	Internal HDD	AQ Dracula HDD – 3.5 Inch SATA 6 Gb/s 5400 RPM 256 MB Cache	Plus
A0118150103	P & A	Peripherals	Internal HDD	AQ Dracula HDD – 3.5 Inch SATA 6 Gb/s 5400 RPM 256 MB Cache	Premium
A0118150104	P & A	Peripherals	Internal HDD	AQ Dracula HDD – 3.5 Inch SATA 6 Gb/s 5400 RPM 256 MB Cache	Premium Plus
A0219150201	P & A	Peripherals	Internal HDD	AQ WereWolf NAS Internal Hard Drive HDD – 8.89 cm	Standard

dim_product

Fact Tables

product_code	fiscal_year	gross_price
A0118150101	2020	16.2323
A0118150101	2021	19.0573
A0118150102	2020	19.8577
A0118150102	2021	21.4565
A0118150103	2020	22.1317

fact_gross_price

date	product_code	customer_code	sold_quantity	fiscal_year
2019-09-01	A0118150101	70002017	137	2020
2019-09-01	A0118150101	70002018	47	2020
2019-09-01	A0118150101	70003181	57	2020
2019-09-01	A0118150101	70003182	63	2020
2019-09-01	A0118150101	70004069	9	2020

fact_sales_monthly

product_code	cost_year	manufacturing_cost
A0118150101	2020	5.0207
A0118150101	2021	5.5172
A0118150102	2020	5.7180
A0118150102	2021	6.2835
A0118150103	2020	6.3264

fact_manufacturing_cost

customer_code	fiscal_year	pre_invoice_discount_pct
70002017	2020	0.0735
70002017	2021	0.0703
70002018	2020	0.2255
70002018	2021	0.2061
70003181	2020	0.0531

fact_pre-invoice_deductions

Customer Data

Platforms	2	(B&M, E-Commerce)
Channel	3	(Retailer, Direct, Distributor)
Market	27	
Region	4	(APAC, EU, NA, LATAM)
Sub Zone	7	
Total Customers	74	

Platforms

Channels

Market

Product Data

Division → Segment → Category → Product → Variant (SKU)
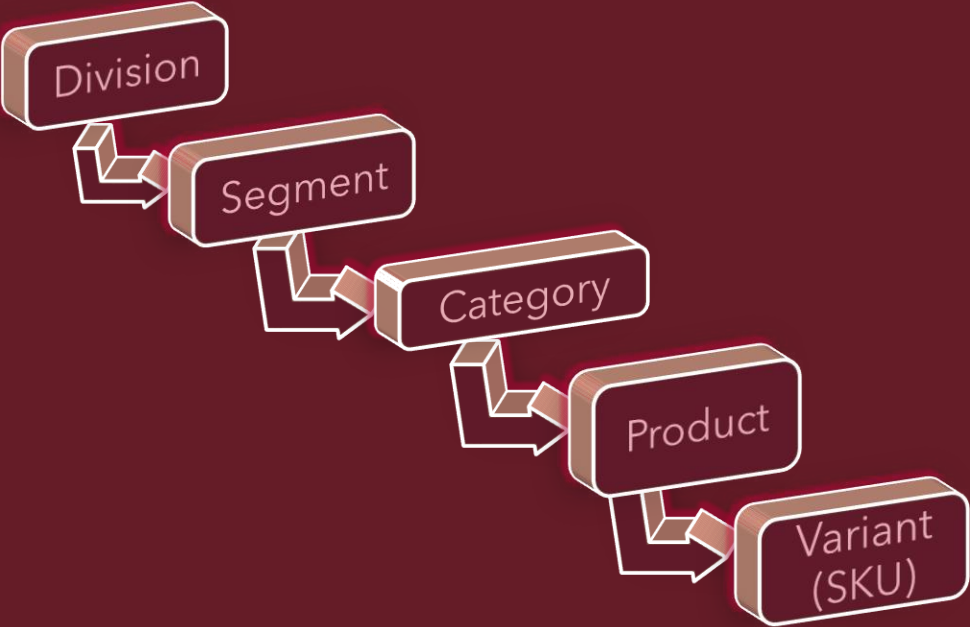
Division	3
Segments	6
Category	14
Total Products	73
Variant	26

SQL Execution Order

FROM → HAVING → WINDOW FUNCTION

JOIN → GROUP BY → SELECT → LIMIT

USING → WHERE → DISTINCT → ORDER BY
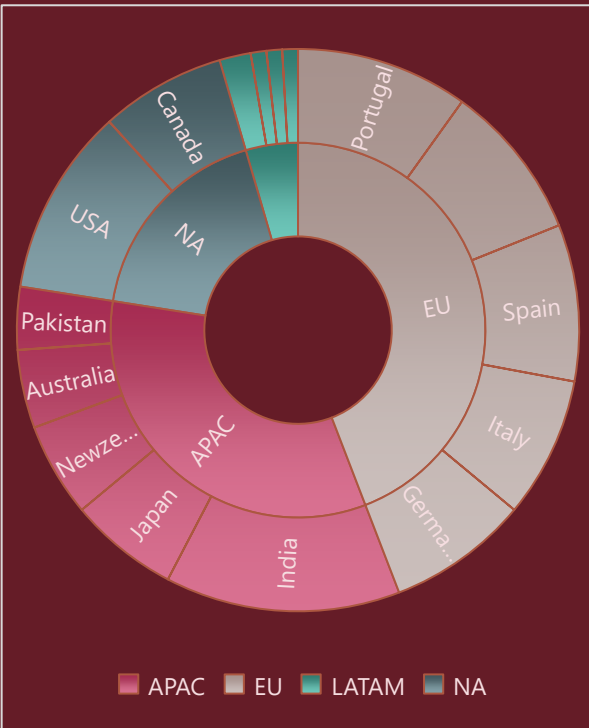
region	market	customer_count	top5_cust_by_region
APAC	India	15	1
APAC	Japan	7	2
APAC	Newzealand	6	3
APAC	Australia	5	4
APAC	Pakistan	4	5
EU	Portugal	11	1
EU	Spain	10	2
EU	UK	10	3
EU	Germany	9	4
EU	Italy	9	5
LATAM	Chile	2	1
LATAM	Columbia	1	2
LATAM	Mexico	1	3
LATAM	Brazil	1	4
NA	USA	12	1
NA	Canada	8	2

Legend: APAC EU LATAM NA

Report Request –

List down Top-5 market in each region based on customer count

```
WITH cust_count_by_region AS (
     SELECT
       region,
       market,
       COUNT(DISTINCT customer) AS customer_count
     FROM dim_customer
     WHERE channel = "retailer"
     GROUP BY market
     ORDER BY region, customer_count DESC),

   top_customer_by_region AS (
     SELECT
       region,
       market,
       customer_count,
       ROW_NUMBER() OVER(partition by region) as
       top5_cust_by_region
     FROM cust_count_by_region)

SELECT
  *
FROM customer_row_by_region
WHERE top5_cust_by_region  < 6
```
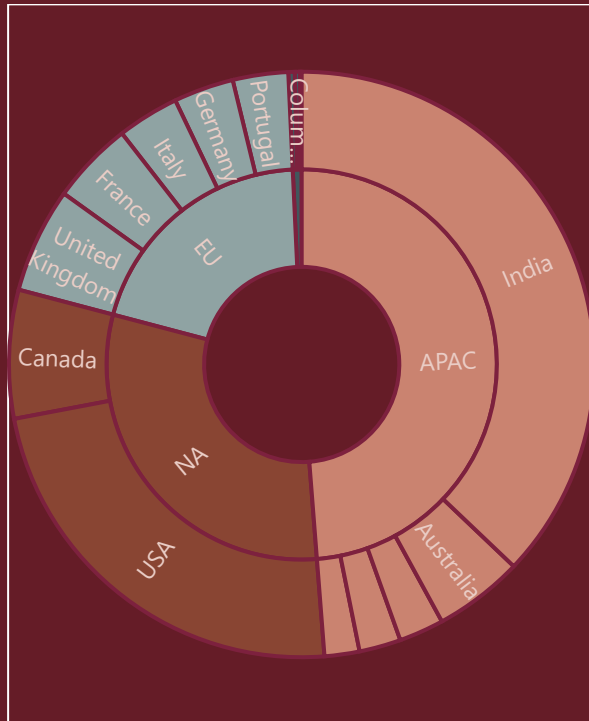
...extended report

region	market	sold_qty	top5_by_region
APAC	India	4771530	1
APAC	Australia	635656	2
APAC	Pakistan	319099	3
APAC	Indonesia	295466	4
APAC	South Korea	248259	5
EU	United Kingdom	746239	1
EU	France	586451	2
EU	Italy	439379	3
EU	Germany	426564	4
EU	Portugal	393110	5
LATAM	Brazil	45324	1
LATAM	Mexico	37019	2
LATAM	Chile	10143	3
LATAM	Columbia	1811	4
NA	USA	2991497	1
NA	Canada	911852	2

```sql
WITH sold_qty_by_market AS (

    SELECT
        c.region,
        c.market,
        count (distinct c.customer) as cust_count,
        sum (s.sold_quantity) as sold_qty
    FROM fact_sales_monthly s
    JOIN dim_customer c
            USING(customer_code)
    WHERE c.channel = "retailer" and s.fiscal_year = 2020
    GROUP BY c.market
    ORDER BY c.region, sold_qty desc),

    top5_markets_by_sold_qty AS (
    SELECT
    *,
    row_number() OVER(PARTITION BY region ORDER   BY sold_qty
    DESC) as top5_by_region
    FROM sold_qty_by_market)

SELECT
    *
FROM top5_markets_by_sold_qty
WHERE top5_by_region < 6;
```
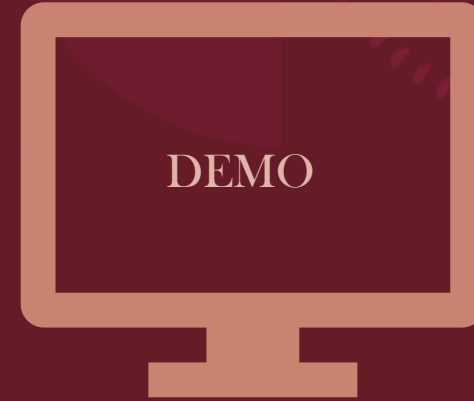
Stored Procedure

Request :

Get top N customers list for a market who got more than X% discount in the given FY

DEMO

```
CREATE PROCEDURE `get_top_N_disc_cust`(
        IN in_market VARCHAR(45),
        IN in_discount_pct DECIMAL,
        IN in_fiscal_year YEAR,
        IN in_topN INT)
BEGIN
        SELECT
            customer,
            pre_invoice_discount_pct
        FROM fact_pre_invoice_deductions
        JOIN dim_customer
            USING(customer_code)
        WHERE fiscal_year = in_fiscal_year
                    AND pre_invoice_discount_pct > in_discount_pct
                    AND market = in_market
        ORDER BY pre_invoice_discount_pct DESC
        LIMIT in_topN;
END
```

Importance of Stored Procedures

• Efficient Data Processing
• Reusable Data Transformation Logic
• Data Security and Integrity
• Automated Data Preparation
• Improved Query Performance for Analysis
• Consistent Data Transformation and Business Logic
• Reduced Network Load for Analytical Applications
• Error Handling for Reliable Analysis Pipelines

Other important reports

Which Category has more/less Product list?

```
44 •    SELECT
45          category,
46          COUNT(DISTINCT product) as product_count
47      FROM dim_product
48      GROUP BY category
49      ORDER BY product_count DESC
50
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

category	product_count
Keyboard	8
Mouse	8
Personal Laptop	8
Business Laptop	6
Graphic Card	6
Processors	6
External Solid State Drives	5
Gaming Laptop	5
MotherBoard	5
Batteries	4
USB Flash Drives	4
Internal HDD	3
Wi fi extender	3
Personal Desktop	2

Product Count by Category
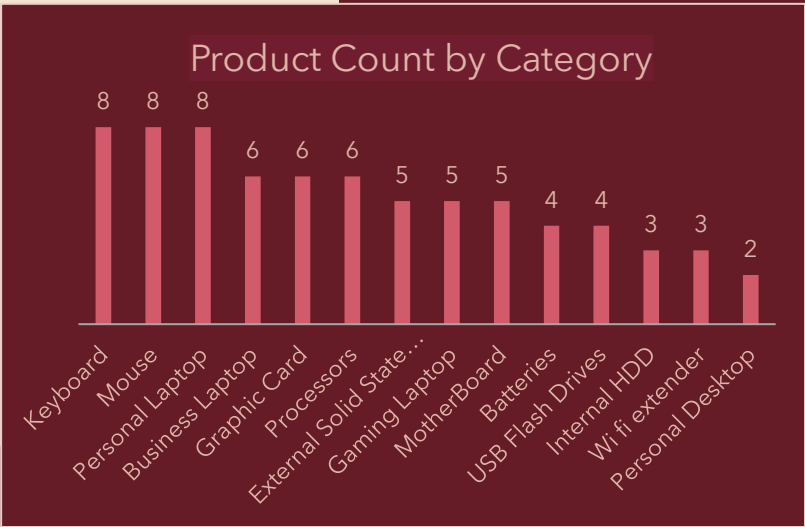
Report how many products has maximum variant along with variant count?

```
45 •  ⊖  WITH cte1 as (
46              SELECT
47                  category,
48                  COUNT(DISTINCT product) as product_count
49              FROM dim_product
50              GROUP BY category
51              ORDER BY product_count DESC
52              )
53      SELECT
54      DISTINCT p.product,
55      COUNT( DISTINCT p.variant) as product_variant_count
56      FROM cte1 c
57      JOIN dim_product p
58          USING(category)
59      GROUP BY p.product
60      ORDER BY product_variant_count DESC
```

Product Variants Details

Business Report Requests

For how many products, manufacturing cost increased from 2020 to 2021 and by how much change in % ?

List down Top/Bottom 5 products whose manufacturing cost is max/min respectively for a given FY

```sql
64  SELECT
65      p.category,
66      p.product,
67      m.product_code,
68      p.variant,
69      m.cost_year,
70      MAX(m.manufacturing_cost) as max_manufacturing_cost
71  FROM fact_manufacturing_cost m
72  JOIN dim_product p
73  USING(product_code)
74  WHERE m.cost_year = 2021
75  GROUP BY product
76  ORDER BY max_manufacturing_cost DESC
77  LIMIT 5;
78
```

| category	product	product_code	variant	cost_year	max_manufacturing_cost
Personal Desktop	AQ HOME Allin1 Gen 2	A6119110201	Standard 1	2021	240.5364
Personal Desktop	AQ Home Allin1	A6018110101	Standard 1	2021	234.1762
Business Laptop	AQ BZ Allin1	A5820110101	Standard 1	2021	219.1571
Gaming Laptop	AQ Smash 2	A5721110501	Standard Firey Red	2021	207.2797
Gaming Laptop	AQ Smash 1	A5620110401	Standard Firey Red	2021	190.4981

```sql
79  --  For how many products, manufacturing cost increased YOY(from 2020 to 2021) and by how much % change in cost
80  WITH manufacturing_cost_2020_vs_2021 AS(
81          SELECT
82              p.category,
83              p.product,
84              m.product_code,
85              p.variant,
86              MAX(CASE WHEN m.cost_year = 2020 THEN m.manufacturing_cost ELSE 0 END) AS cost_2020, -- use MAX()
87              MAX(CASE WHEN m.cost_year = 2021 THEN m.manufacturing_cost ELSE 0 END) AS cost_2021 -- Else # rec
88          FROM fact_manufacturing_cost m
89          JOIN dim_product p
90              USING(product_code)
91          GROUP BY product_code -- Returns cost by each variant / If we group by product then 2020 and 2021 cost
92          ORDER BY cost_2020,cost_2021)
93  SELECT
94      *,
95      ROUND((cost_2021-cost_2020)*100/cost_2020,2) as cost_change_pct -- Calculate change in cost(%) compare to previous
96  FROM manufacturing_cost_2020_vs_2021
97  ORDER BY cost_change_pct DESC;
98
99  -- Average manufacturing cost for given category/segment/division for given FY.
100 SELECT
101     p.category,
```

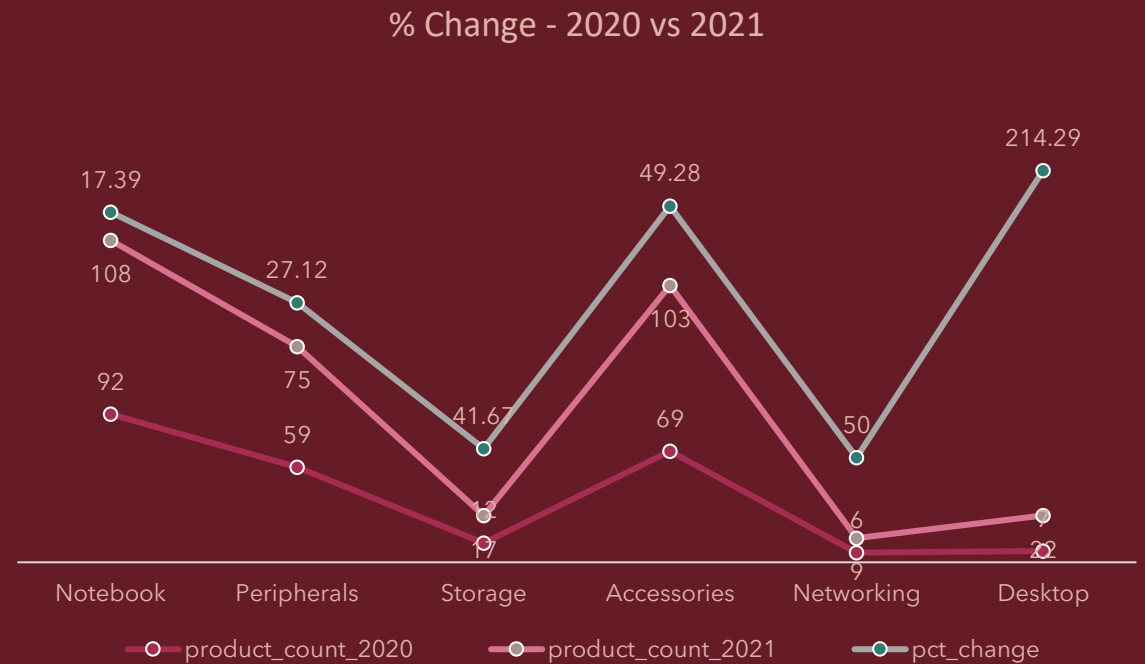| category	product	product_code	variant	cost_2020	cost_2021	cost_change_pct
Mouse	AQ Master wireless x1 Ms	A2219150205	Premium 1	2.1379	2.3755	11.11
Mouse	AQ Lite Ms	A2419150404	Plus 2	3.2414	3.6015	11.11
Keyboard	AQ Lite	A3220150401	Standard 1	5.0345	5.5939	11.11
Batteries	AQ LION x2	A3819150205	Premium	6.0000	6.6667	11.11
Graphic Card	AQ Mforce Gen Y	A0519150206	Plus 3	8.2069	9.1188	11.11
Wi fi extender	AQ Wi Power Dx2	A7219160201	Standard	8.5517	9.5019	11.11
MotherBoard	AQ MB Crossx 2	A1819150304	Premium	8.8276	9.8084	11.11
Personal Laptop	AQ Digit	A4218110208	Premium Misty Green	67.1724	74.6360	11.11
Personal Laptop	AQ Velocity	A4319110302	Standard Blue	70.1379	77.9310	11.11
Personal Laptop	AQ Velocity	A4319110303	Standard Red	71.0345	78.9272	11.11
Personal Laptop	AQ Velocity	A4319110304	Plus Grey	71.2414	79.1571	11.11

Ad-hock Requests

What is the percentage of unique product increase in 2021 vs. 2020? The final output contains these fields, unique_products_2020 unique_products_2021 percentage_chg
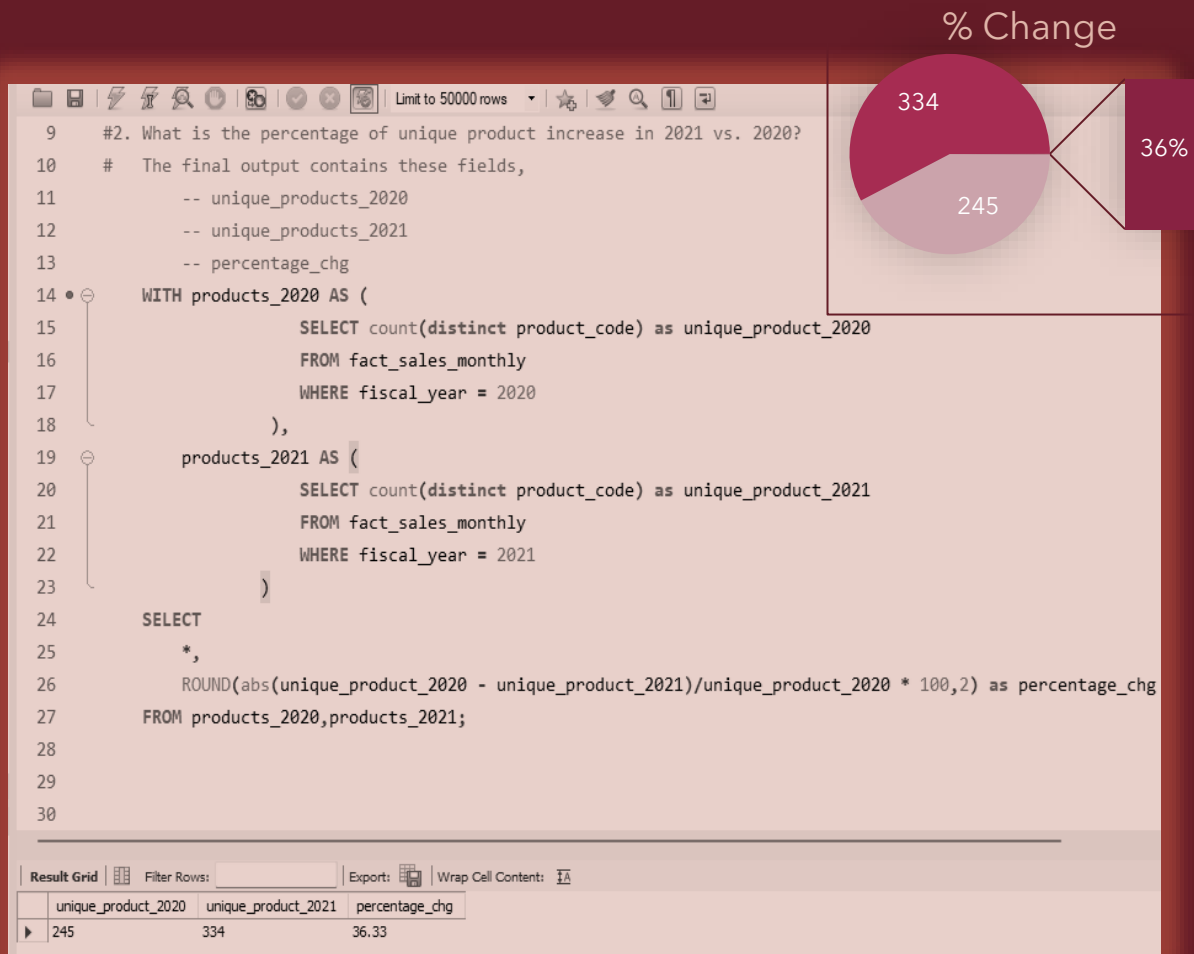
Follow-up: Which segment had the most increase in unique products in 2021 vs 2020? The final output contains these fields, segment product_count_2020 product_count_2021 difference

segment	product_count_2020	product_count_2021	pct_change
Desktop	7	22	214.29
Networking	6	9	50.00
Accessories	69	103	49.28
Storage	12	17	41.67
Peripherals	59	75	27.12
Notebook	92	108	17.39

% Change

334
245
36%

```
 9   #2. What is the percentage of unique product increase in 2021 vs. 2020?
10   #   The final output contains these fields,
11          -- unique_products_2020
12          -- unique_products_2021
13          -- percentage_chg
14   WITH products_2020 AS (
15               SELECT count(distinct product_code) as unique_product_2020
16               FROM fact_sales_monthly
17               WHERE fiscal_year = 2020
18            ),
19        products_2021 AS (
20               SELECT count(distinct product_code) as unique_product_2021
21               FROM fact_sales_monthly
22               WHERE fiscal_year = 2021
23            )
24   SELECT
25        *,
26        ROUND(abs(unique_product_2020 - unique_product_2021)/unique_product_2020 * 100,2) as percentage_chg
27   FROM products_2020,products_2021;
28
29
30
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

unique_product_2020	unique_product_2021	percentage_chg
245	334	36.33

% Change - 2020 vs 2021

product_count_2020   product_count_2021   pct_change

...requests continues

In which quarter of 2020, got the maximum total_sold_quantity?
The final output contains these fields sorted by the total_sold_quantity, Quarter total_sold_quantity
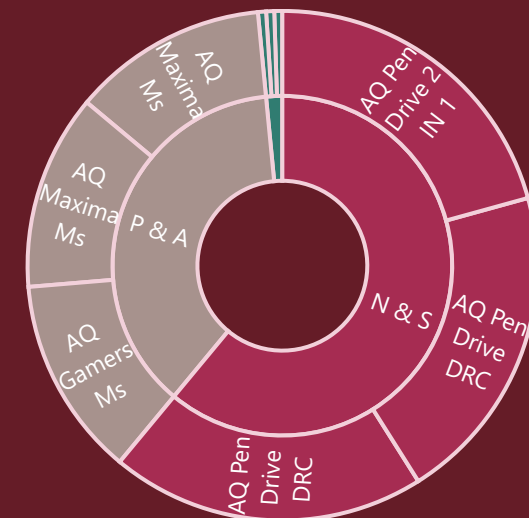
Get the Top 3 products in each division that have a high total_sold_quantity in the fiscal_year 2021?
The final output contains these fields, division product_code product total_sold_quantity rank_order

```sql
83  ●  SELECT
84         month(date) as month,
85     ⊖   CASE
86             WHEN month(date) BETWEEN 9 and 11 THEN "Q1"
87             WHEN TRIM(month(date))BETWEEN 12 and 2 THEN "Q2"
88             WHEN month(date) BETWEEN 3 and 5 THEN "Q3"
89             WHEN month(date) BETWEEN 6 and 8 THEN "Q4"
90         END as Quarters,
91         SUM(sold_quantity) as total_sold_quantity
92     FROM fact_sales_monthly
93     WHERE fiscal_year = 2020
94     GROUP BY Quarters
95     ORDER BY total_sold_quantity DESC;
```

division	product_code	product	total_sold_qty	rnk
N & S	A6720160103	AQ Pen Drive 2 IN 1	701373	1
N & S	A6818160202	AQ Pen Drive DRC	688003	2
N & S	A6819160203	AQ Pen Drive DRC	676245	3
P & A	A2319150302	AQ Gamers Ms	428498	1
P & A	A2520150501	AQ Maxima Ms	419865	2
P & A	A2520150504	AQ Maxima Ms	419471	3
PC	A4218110202	AQ Digit	17434	1
PC	A4319110306	AQ Velocity	17280	2
PC	A4218110208	AQ Digit	17275	3

Result Grid | Filter Rows: | Exp

month	Quarters	total_sold_quantity
9	Q1	7005619
12	Q2	6649642
6	Q4	5042541
3	Q3	2075087

Quarterly Sold Qty

- Q3: 2075087
- Q4: 5042541
- Q2: 6649642
- Q1: 7005619

■ total_sold_quantity

High Total Sold Qty in 2021

Thanks for watching

I hope you found the presentation insightful

If you have any questions or would like further discussion, feel free to ask

Praveen Vanahalli