Codebasics Challenge #4

DATA ANALYSIS FOR BUSINESS

Consumer Goods Ad-Hoc Insights

created by Teodor Cristia

Challenge #4: Provide Insights to Management in Consumer Goods Domain

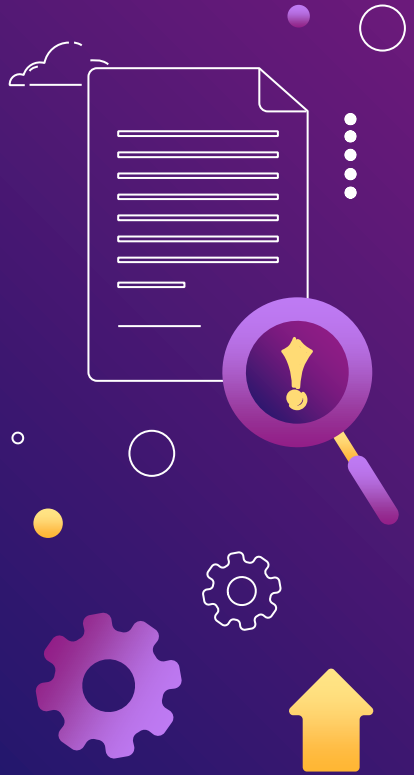
Domain: Consumer Goods | **Function**: Executive Management

Problem Statement:

- Atliq Hardware is a leading computer hardware producer in India that requires quick and informed decisions based on data insights.
- They need a junior data analyst with strong technical and soft skills to expand their Data Analytics team.
- The junior data analyst is tasked with generating insights and answering 10 Ad-Hoc requests.
- The insights generated will be presented to top-level management.

Task

- Analyze the dataset provided using SQL to answer 10 Ad-Hoc requests.
- Generate insights based on the data and present them in a creative and engaging format.
- Use effective presentation skills to convey the insights to top-level management.
- Highlight key takeaways and insights from the analysis.
- Demonstrate strong technical and soft skills as a junior data analyst.



Important Considerations

- Queries and visualizations are based on the fiscal calendar.
- The fiscal calendar at Atliq Hardware is starting in September.
- The Database provided covers the history of two fiscal years (2020 and 2021), with almost 1 million rows of data.

Data Model

fact_sales_monthly
customer_code
date
fiscal_year
product_code
sold_quantity
Collapse
dim_product
category
division
product
product_code
segment
variant
Collapse
fact_gross_price
fiscal_year
gross_price
product_code
Collapse
dim_customer
channel
customer
customer_code
market
platform
region
sub_zone
Collapse
fact_pre_invoice_deductions
customer_code
fiscal_year
pre_invoice_discount_pct
Collapse
fiscal_year
fiscal_year
Collapse
fact_manufacturing_cost
cost_year
manufacturing_cost
product_code
Collapse

Data Analysis Process

01

Requests

Introduce the questions we aim to answer in this presentation.

02

SQL Queries

Show the technical process used to retrieve the data.

03

SQL Outputs

Present the results of the data retrieval process.

04

Results Visualizations

Highlight key insights from the data with a clear and easy-to-understand format.

05

Insights

Key takeaways and observations from data analysis for decision-making and useful insights.

Request #1

Provide the list of markets in which customer "Atliq Exclusive" operates its business in the APAC region.

SQL Query

```sql
SELECT DISTINCT
    market
FROM dim_customer
WHERE customer = 'Atliq Exclusive'
    AND region = 'APAC'
ORDER BY market;
```

SQL Output

market
Australia
Bangladesh
India
Indonesia
Japan
Newzealand
Philiphines
South Korea

Results Visualization

Insights

• Atliq Exclusive customer operates its business in 8 markets across the APAC region, including Australia, Bangladesh, India, Indonesia, Japan, New Zealand, Philippines, and South Korea.

• Analyzing the **quantity sold** and **gross sales** in each market can reveal its importance to customer Atliq Exclusive in terms of revenue and growth potential in the APAC region.

➥ Now, let's take the next step forward and explore it…

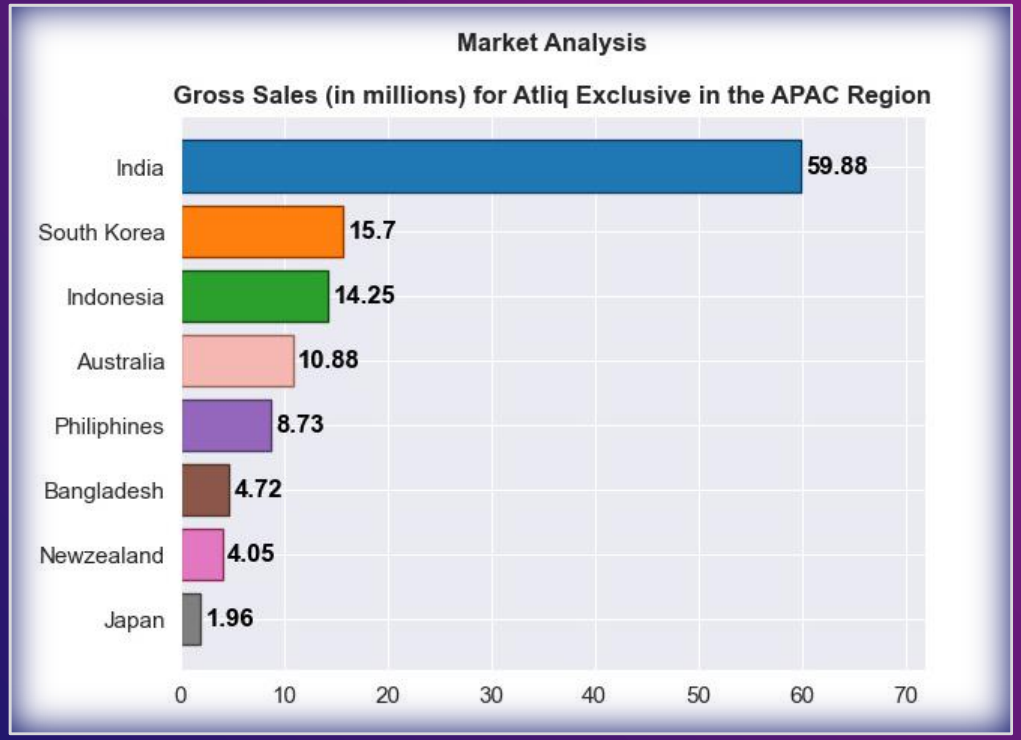
SQL Query

```sql
WITH atliq_exclusive_apac AS (
    SELECT
        c.market,
        SUM(s.sold_quantity) AS total_sold_quantity,
        ROUND(
                SUM(s.sold_quantity * gp.gross_price) / 1000000, 2
            ) AS gross_sales_mln
    FROM fact_sales_monthly s
    JOIN fact_gross_price gp
        ON s.product_code = gp.product_code
        AND s.fiscal_year = gp.fiscal_year
    JOIN dim_customer c
        ON s.customer_code = c.customer_code
    WHERE customer = 'Atliq Exclusive'
        AND region = 'APAC'
    GROUP BY market
)
SELECT
    market,
    ROUND(
        total_sold_quantity / SUM(total_sold_quantity) OVER() * 100, 2
    ) AS sold_qty_percentage,
    gross_sales_mln
FROM atliq_exclusive_apac
ORDER BY gross_sales_mln DESC;
```

SQL Output

market	sold_qty_percentage	gross_sales_mln
India	50.04	59.88
South Korea	13.14	15.70
Indonesia	11.65	14.25
Australia	9.05	10.88
Philiphines	7.19	8.73
Bangladesh	3.91	4.72
Newzealand	3.36	4.05
Japan	1.66	1.96

Results Visualization



Insights

- India is by far the most important market for customer Atliq Exclusive, accounting for more than half of the sold quantity and gross sales in the APAC region.

- South Korea and Indonesia are the next two largest markets for Atliq Exclusive, with each accounting for over 10% of the sold quantity and gross sales.

- Japan has the smallest sold quantity percentage and gross sales figure among the markets.

Request #2

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

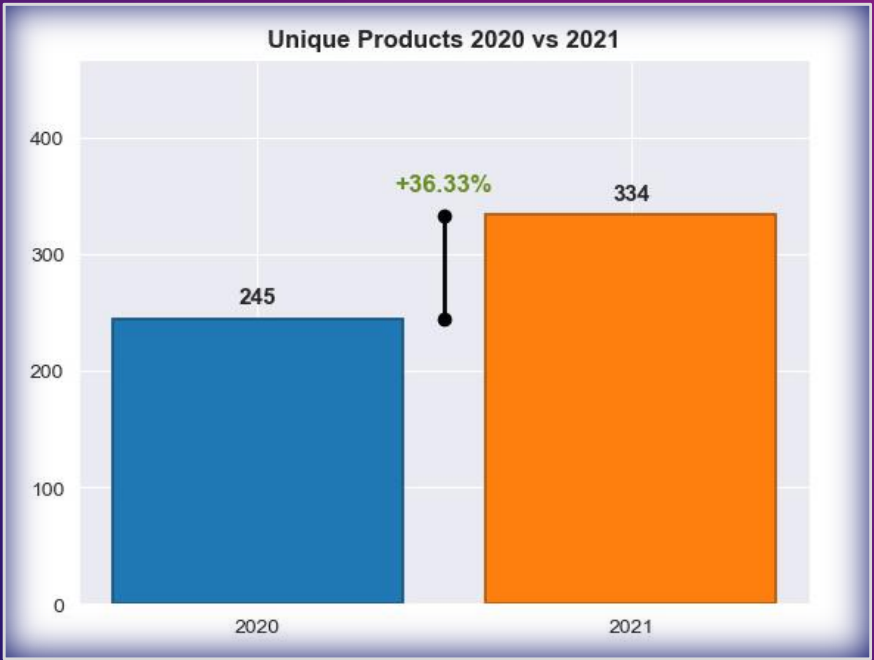
SQL Query

```sql
WITH unique_products AS (
    SELECT
        COUNT(DISTINCT CASE WHEN fiscal_year = 2020 THEN product_code END) AS unique_products_2020,
        COUNT(DISTINCT CASE WHEN fiscal_year = 2021 THEN product_code END) AS unique_products_2021
    FROM fact_sales_monthly
)
SELECT
    unique_products_2020,
    unique_products_2021,
    ROUND(
        (unique_products_2021 - unique_products_2020) * 100.0 / unique_products_2020, 2
    ) AS percentage_chg
FROM unique_products;
```

SQL Output

	unique_products_2020	unique_products_2021	percentage_chg
▶	245	334	36.33

Results Visualization



Insights

- The number of unique products increased from **245** in 2020 to **334** in 2021, which represents a significant increase of **36.33%**.

- The increase in the number of unique products from 2020 to 2021 indicates that the business is expanding its product offerings, which may lead to increased sales and revenue.

Request #3

Provide a report with all the unique product counts for each segment and sort them in descending order of product counts.
The final output contains 2 fields:

- segment
- product_count

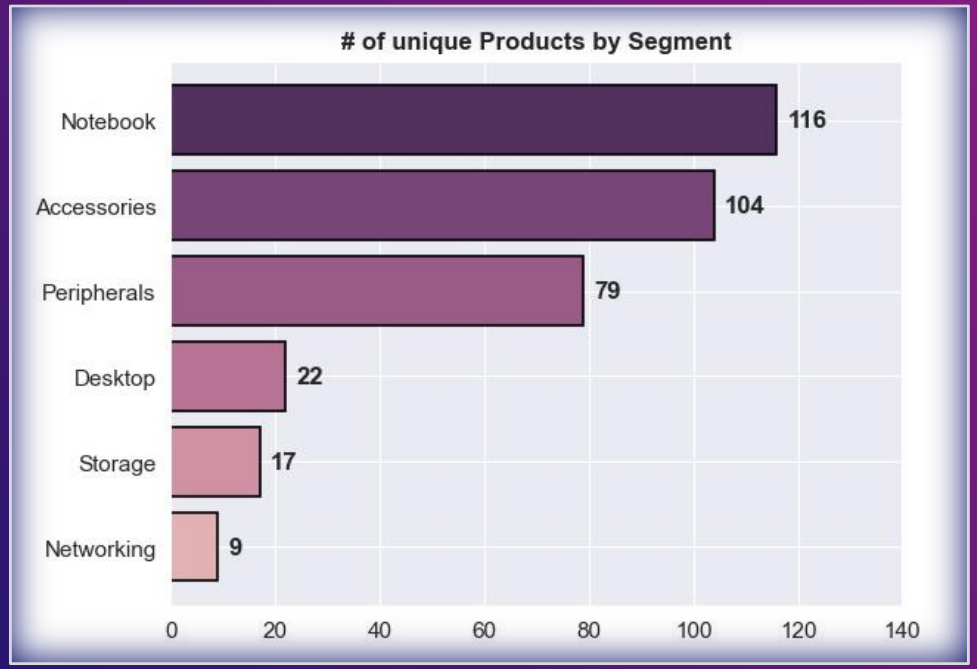
SQL Query

```sql
SELECT
    p.segment,
    COUNT(DISTINCT s.product_code) AS product_count
FROM fact_sales_monthly s
JOIN dim_product p
    ON s.product_code = p.product_code
GROUP BY p.segment
ORDER BY product_count DESC;
```

SQL Output

segment	product_count
Notebook	116
Accessories	104
Peripherals	79
Desktop	22
Storage	17
Networking	9

Results Visualization



Insights

- The Notebook segment has the highest number of products, with a count of **116**. This suggests that this is a key area of focus for the business.
- Accessories is the second largest segment with **104** products.
- Networking has the lowest product count with only **9** products.

Request #4

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

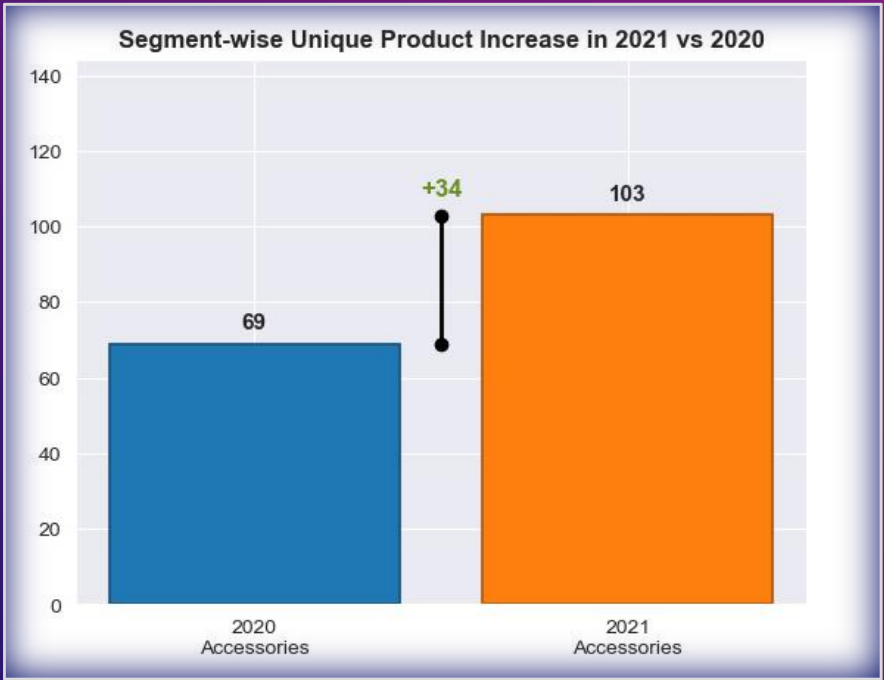
SQL Query

```sql
WITH segment_products AS (
        SELECT
            p.segment,
            COUNT(DISTINCT CASE WHEN fiscal_year = 2020 THEN s.product_code END) AS product_count_2020,
            COUNT(DISTINCT CASE WHEN fiscal_year = 2021 THEN s.product_code END) AS product_count_2021,
            COUNT(DISTINCT CASE WHEN fiscal_year = 2021 THEN s.product_code END) -
            COUNT(DISTINCT CASE WHEN fiscal_year = 2020 THEN s.product_code END) AS difference
        FROM fact_sales_monthly s
        JOIN dim_product p
            ON p.product_code = s.product_code
        WHERE fiscal_year IN (2020, 2021)
        GROUP BY p.segment
)
SELECT
    *
FROM segment_products
WHERE difference = (
        SELECT MAX(difference)
        FROM segment_products
    );
```

SQL Output

	segment	product_count_2020	product_count_2021	difference
▶	Accessories	69	103	34

Results Visualization



Insights

• Accessories segment had the most increase in unique products in 2021 compared to 2020, with a difference of 34 products. In 2020, the segment had 69 unique products, while in 2021, it had 103 unique products.

Request #5

Get the products that have the highest and lowest manufacturing costs.
The final output should contain these fields:

- product_code
- product
- manufacturing_cost

SQL Query

```sql
SELECT
    m.product_code,
    p.category,
    CONCAT(p.product, " - ", p.variant) AS product_name,
    m.manufacturing_cost
FROM fact_manufacturing_cost m
JOIN dim_product p
    ON m.product_code = p.product_code
WHERE m.manufacturing_cost IN (
        (SELECT MIN(manufacturing_cost) FROM fact_manufacturing_cost),
        (SELECT MAX(manufacturing_cost) FROM fact_manufacturing_cost)
);
```

SQL Output

product_code	category	product_name	manufacturing_cost
A2118150101	Mouse	AQ Master wired x1 Ms - Standard 1	0.8920
A6120110206	Personal Desktop	AQ HOME Allin1 Gen 2 - Plus 3	240.5364

Results Visualization

Insights

• The minimum manufacturing cost is ₹ 0.8920 and it's a Mouse product, which is likely a relatively simple or low-cost device.

• The maximum manufacturing cost is ₹ 240.5364, it is a Personal Desktop product, which is likely more complex or higher-end.

Request #6

Generate a report which contains the top 5 customers who received an average high pre_invoice_discount_pct for the fiscal year 2021 and in the Indian market.

The final output contains these fields:

- customer_code
- customer
- average_discount_percentage

SQL Query

```sql
WITH ranked_india_customers_2021 AS (
    SELECT
        c.customer_code,
        c.customer,
        pre_invoice_discount_pct,
        DENSE_RANK() OVER(ORDER BY pre_invoice_discount_pct DESC) AS drnk
    FROM dim_customer c
    JOIN fact_pre_invoice_deductions pid
        ON c.customer_code = pid.customer_code
    WHERE market = "India"
        AND fiscal_year = 2021
)
SELECT
    customer_code,
    customer,
    pre_invoice_discount_pct
FROM ranked_india_customers_2021
WHERE drnk <= 5
    AND pre_invoice_discount_pct > (
            SELECT AVG(pre_invoice_discount_pct)
            FROM ranked_india_customers_2021
    );
```

SQL Output

customer_code	customer	pre_invoice_discount_pct
90002009	Flipkart	0.3083
90002006	Viveks	0.3038
90002003	Ezone	0.3028
90002002	Croma	0.3025
90002016	Amazon	0.2933

Results Visualization

Insights

• These customers are likely to be important to the business as they are receiving a high average discount percentage for the fiscal year 2021 and in the Indian market, which could indicate that they are making large or frequent purchases.

Request #7

Get the complete report of the Gross sales amount for the customer "Atliq Exclusive" for each month. This analysis helps to get an idea of low and high-performing months and take strategic decisions.

The final report contains these columns:

- Month
- Year
- GrossSalesAmount

SQL Query

```sql
SELECT
    s.fiscal_year,
    MONTHNAME( DATE_ADD(date, INTERVAL 4 MONTH) ) AS fiscal_month,
    ROUND( SUM(s.sold_quantity * gp.gross_price) / 1000000, 2 ) AS gross_sales_amount_mln
FROM fact_sales_monthly s
JOIN fact_gross_price gp
    ON s.product_code = gp.product_code
    AND s.fiscal_year = gp.fiscal_year
JOIN dim_customer c
    ON s.customer_code = c.customer_code
WHERE c.customer = "Atliq Exclusive"
GROUP BY fiscal_year, fiscal_month
ORDER BY gross_sales_amount_mln DESC;
```

SQL Output

fiscal_year	fiscal_month	gross_sales_amount_mln
2021	March	20.46
2021	February	13.22
2021	April	12.94
2021	May	12.40
2021	January	12.35
2021	September	12.15
2021	July	12.14
2021	November	12.09
2021	June	10.13
2021	October	9.82
2020	March	7.52
2021	August	7.31
2021	December	7.18
2020	February	5.14
2020	April	4.83
2020	May	4.74
2020	January	4.50
2020	June	4.00
2020	December	2.79
2020	November	2.55
2020	October	1.70
2020	September	0.78
2020	August	0.40
2020	July	0.38

Results Visualization

Insights

• The gross sales amount has been increasing for all months in the fiscal year 2021 compared to the same months in the previous year 2020.

• The highest gross sales amount was recorded in March 2021, followed by February and April 2021.

• The lowest gross sales amount was recorded in July and August 2020.

Request #8

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

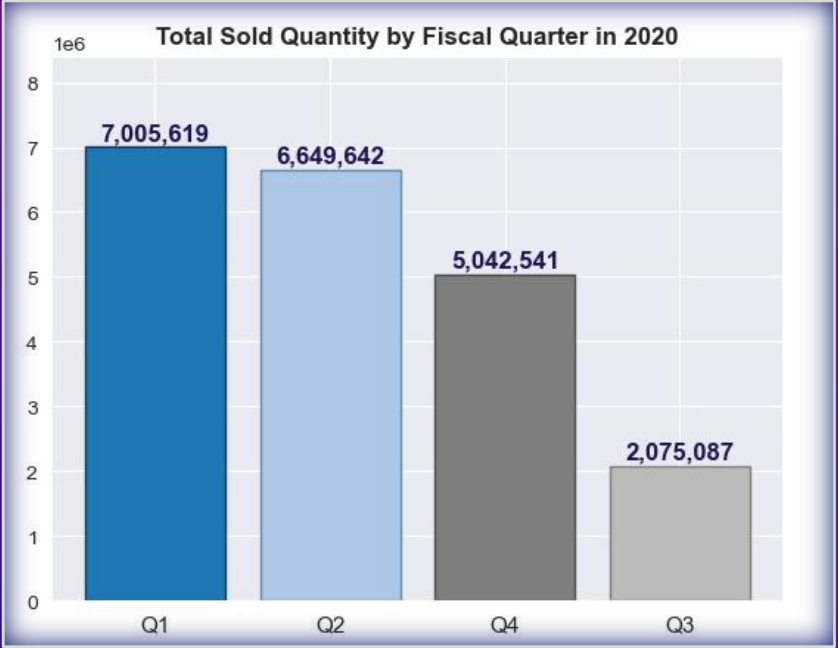
SQL Query

```sql
SELECT
    CONCAT("Q",
            QUARTER( DATE_ADD(date, INTERVAL 4 MONTH) )
        ) AS fiscal_quarter,
    FORMAT( SUM(sold_quantity), 0) AS total_sold_quantity
FROM fact_sales_monthly
WHERE fiscal_year = 2020
GROUP BY fiscal_quarter
ORDER BY SUM(sold_quantity) DESC;
```

SQL Output

fiscal_quarter	total_sold_quantity
Q1	7,005,619
Q2	6,649,642
Q4	5,042,541
Q3	2,075,087

Results Visualization



Insights

• The maximum total sold quantity in 2020 was in Q1 with 7,005,619 units sold, followed by Q2 with 6,649,642 units sold, and then Q4 with 5,042,541 units sold.

• The lowest total sold quantity was in Q3 with 2,075,087 units sold.

Request #9

Which channel helped to bring more gross sales in the fiscal year 2021 and the percentage of contribution?

The final output contains these fields:

- channel
- gross_sales_mln
- percentage

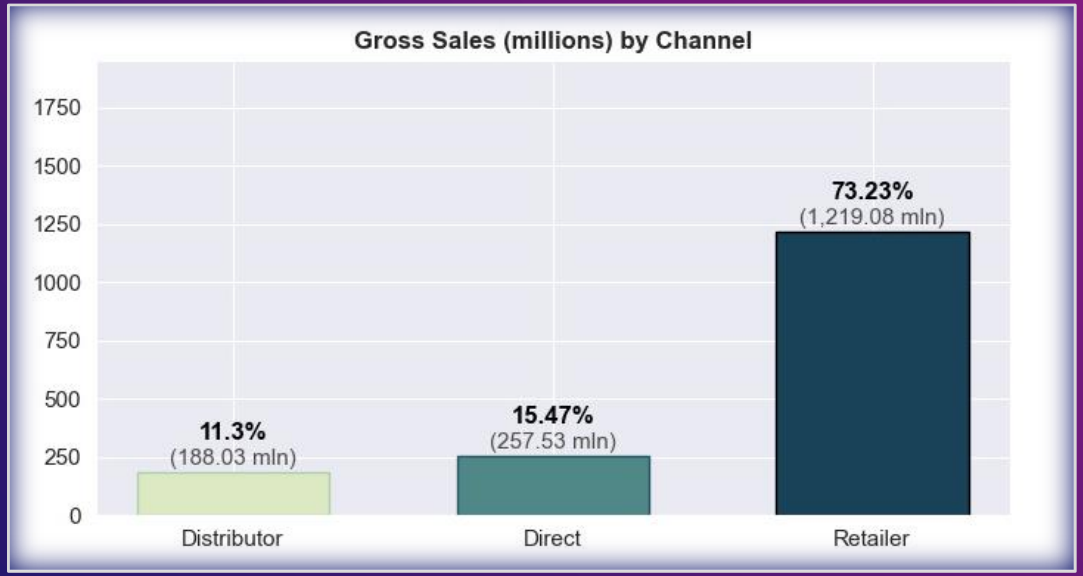
SQL Query

```sql
WITH channel_gross_sales AS (
    SELECT
        c.channel,
        ROUND(
            SUM(s.sold_quantity * gp.gross_price) / 1000000, 2
        ) AS gross_sales_mln
    FROM fact_sales_monthly s
    JOIN dim_customer c
        ON s.customer_code = c.customer_code
    JOIN fact_gross_price gp
        ON s.product_code = gp.product_code
        AND s.fiscal_year = gp.fiscal_year
    WHERE s.fiscal_year = 2021
    GROUP BY c.channel
)
SELECT
    channel,
    gross_sales_mln,
    ROUND(
        gross_sales_mln / SUM(gross_sales_mln) OVER() * 100, 2
    ) AS percentage
FROM channel_gross_sales
ORDER BY percentage DESC;
```

SQL Output

channel	gross_sales_mln	percentage
Retailer	1219.08	73.23
Direct	257.53	15.47
Distributor	188.03	11.30

Results Visualization



Insights

- The retailer channel was the main contributor to the gross sales in the fiscal year 2021, accounting for 73.23% of the total gross sales.

- The direct channel accounted for 15.47% of the gross sales, which is significantly lower than the retailer channel.

- The distributor channel had the smallest contribution to the gross sales, accounting for only 11.30% of the total gross sales.

Request #10

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

SQL Query

```sql
WITH division_sold_qty AS (
    SELECT
        p.division,
        p.product_code,
        FORMAT(SUM(s.sold_quantity), 0) AS total_sold_quantity,
        DENSE_RANK() OVER(
                PARTITION BY division
                ORDER BY SUM(s.sold_quantity) DESC
            ) AS drnk
    FROM fact_sales_monthly s
    JOIN dim_product p
        ON s.product_code = p.product_code
    WHERE fiscal_year = 2021
    GROUP BY p.division, p.product_code
)
SELECT
    d.division,
    d.product_code,
    CONCAT(p.product, ' - ', p.variant) AS product_name,
    d.total_sold_quantity
FROM division_sold_qty d
JOIN dim_product p
    ON d.product_code = p.product_code
WHERE drnk <= 3
ORDER BY division ASC, total_sold_quantity DESC;
```

SQL Output

division	product_code	product_name	total_sold_quantity
N & S	A6720160103	AQ Pen Drive 2 IN 1 - Premium	701,373
N & S	A6818160202	AQ Pen Drive DRC - Plus	688,003
N & S	A6819160203	AQ Pen Drive DRC - Premium	676,245
P & A	A2319150302	AQ Gamers Ms - Standard 2	428,498
P & A	A2520150501	AQ Maxima Ms - Standard 1	419,865
P & A	A2520150504	AQ Maxima Ms - Plus 2	419,471
PC	A4218110202	AQ Digit - Standard Blue	17,434
PC	A4319110306	AQ Velocity - Plus Red	17,280
PC	A4218110208	AQ Digit - Premium Misty Green	17,275

Results Visualization

Insights

- The top-selling products in each division have significantly high total sold quantity in the fiscal year 2021.
- The N & S division has three products that stand out in terms of total sold quantity, with each product selling over 670,000 units in the given fiscal year.
- The P & A division's top-selling products have a relatively lower total sold quantity than the top-selling products of the N & S division, with the highest-selling product selling around 428,000 units.
- The PC division has a significantly lower total sold quantity than the other two divisions, with the highest-selling product selling only around 17,000 units.

Tools used in my project

used to create the presentation and present information in a visually engaging way

PowerPoint

used to store, manage, and retrieve data, and solve all data requests in the presentation

MySQL

used for editing and refining the visuals in the presentation, which includes the ability to crop images

GIMP

used to create visualizations for a few of the requests in the presentation, providing a great view of the data

Power BI

Python

Pandas

Matplotlib

Seaborn

used to create the majority of the visualizations, displaying the data in a visually impactful and intuitive way

THANK YOU!

linkedin.com/in/teodor-cristia

github.com/Teodor-CTI/PortfolioProjects

codebasics.io/event/codebasics-resume-project-challenge

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon** and infographics & images by **Freepik**