Low Level Design (LLD)

Consumer Goods Ad-Hoc Project

By

Pranit Patil

Contents

Title	Page Number
Abstract	1
1.Introduction 1.1. Why this Low-Level Design Document? 1.2. Scope 1.3. Problem Statement and Task	2
2.Architecture Description 2.1. Data Description 2.2. SQL Queries	3 to 13
3.Reporting	13
4.Deployment	14

Abstract

Atliq Hardwares (imaginary company) is one of the leading computer hardware producers in India and well expanded in other countries too.

However, the management noticed that they do not get enough insights to make quick and smart data-informed decisions. They want to expand their data analytics team by adding several junior data analysts. Tony Sharma, their data analytics director wanted to hire someone who is good at both tech and soft skills. Hence, he decided to conduct a SQL challenge which will help him understand both the skills.

1 Introduction

1.1 Why this Low-Level Design Document?

The goal of the LDD or Low-level design document (LLDD) is to give the internal logic design of the actual program code for the Heart Disease Diagnostic-Analysis dashboard. LLDD describes the class diagrams with the methods and relations between classes and programs specs. It describes the modules so that the programmer can directly code the program from the document.

1.2 Scope

Low-level design (LLD) is a component-level design process that follows a step-by-step refinement process. The process can be used for designing data structures, required software architecture, source code and ultimately, performance algorithms. Overall, the data organization may be defined during requirement analysis and then refined during data design work.

1.3 Problem Statement and Task

Atliq Hardwares (imaginary company) is one of the leading computer hardware producers in India and well expanded in other countries too.

However, the management noticed that they do not get enough insights to make quick and smart data-informed decisions. They want to expand their data analytics team by adding several junior data analysts. Tony Sharma, their data analytics director wanted to hire someone who is good at both tech and soft skills. Hence, he decided to conduct a SQL challenge which will help him understand both the skills.

Task:

Imagine yourself as the applicant for this role and perform the following task:

1. Check ‘ad-hoc-requests.pdf’ - there are 10 ad hoc requests for which the business needs insights.
2. You need to run a SQL query to answer these requests.
3. The target audience of this dashboard is top-level management - hence you need to create a presentation to show the insights.

2 Architecture Description

2.1. Data Description

Database Link: Consumer-Goods-Ad_Hoc-Project/atliq_hardware_db.zip at main · Pp11112000/Consumer-Goods-Ad_Hoc-Project (github.com)

Below provided a comprehensive overview of the tables found in the 'gdb023' (atliq_hardware_db) database (Link is Provided). It includes information for six main tables:

1. dim_customer: contains customer-related data.
2. dim_product: contains product-related data.
3. fact_gross_price: contains gross price information for each product.
4. fact_manufacturing_cost: contains the cost incurred in the production of each product.
5. fact_pre_invoice_deductions: contains pre-invoice deductions information for each product.
6. fact_sales_monthly: contains monthly sales data for each product.

- **Column Description for dim_customer table:**

1. customer_code: The 'customer_code' column features unique identification codes for every customer in the dataset. These codes can be used to track a customer's sales history, demographic information, and other relevant details. For example, the codes could look like '70002017', '90005160', and '80007195' respectively.

2. customer: The 'customer' column lists the names of customers, for example 'Atliq Exclusive', 'Flipkart', and 'Surface Stores' etc.

3. platform: The 'platform' column identifies the means by which a company's products or services are sold. "Brick & Mortar" represents the physical store/location, and "E-Commerce" represents online platforms.

4. channel: The 'channel' column reflects the distribution methods used to sell a product. These methods include "Retailers", "Direct", and "Distributors". Retailers refer to physical or online stores that sell products to consumers. Direct sales refer to sales made directly to consumers through a company's website or other direct means, and distributors refer to intermediaries or middlemen between the

manufacturer and retailer or end consumers.

5. market: The 'market' column lists the countries in which the customer is located.

6. region: The 'region' column categorizes countries according to their geographic location, including "APAC" (Asia Pacific), "EU" (Europe), "NA" (North America), and "LATAM" (Latin America).

7. sub_zone: "The 'sub_zone' column further breaks down the regions into sub-regions, such as "India", "ROA" (Rest of Asia), "ANZ" (Australia and New Zealand), "SE" (Southeast Asia), "NE" (Northeast Asia), "NA" (North America), and "LATAM" (Latin America)."

- **Column Description for dim_product table:**

1. product_code: The 'product_code' column features unique identification codes for each product, serving as a means to track and distinguish individual products within a database or system.

2. division: The 'division' column categorizes products into groups such as "P & A" (Peripherals and Accessories), "N & S" (Network and Storage) and "PC" (Personal Computer).

3. segment: The 'segment' column categorizes products further within the division, such as "Peripherals" (keyboard, mouse, monitor, etc.), "Accessories" (cases, cooling solutions, power supplies), "Notebook" (laptops), "Desktop" (desktops, all-in-one PCs, etc), "Storage" (hard disks, SSDs, external storage), and "Networking" (routers, switches, modems, etc.).

4. category: The 'category' column classifies products into specific subcategories within the segment.

5. product: The 'product' column lists the names of individual products, corresponding to the unique identification codes found in the 'product_code' column.

6. variant: The "variant" column classifies products according to their features, prices, and other characteristics. The column includes variants such as "Standard","Plus", "Premium" that represent different versions of the same product.

- **Column Description for fact_gross_price table:**

1. product_code: The 'product_code' column features unique identification codes for each product.

2. fiscal_year: The 'fiscal_year' column contains the fiscal period in which the product sale was recorded. A fiscal year is a 12-month period that is used for accounting purposes and often differs from the calendar year. For Atliq Hardware, the fiscal year starts in September. The data available in this column covers the fiscal years 2020 and 2021.

3. gross_price: The 'gross_price' column holds the initial price of a product, prior to any reductions or taxes. It is the original selling price of the product.

- **Column Description for fact_manufacturing_cost:**

1. product_code: The 'product_code' column features unique identification codes for each product

2. cost_year: The "cost_year" column contains the fiscal year in which the product was manufactured.

3. manufacturing_cost: The "manufacturing_cost" column contains the total cost incurred for the production of a product. This cost includes direct costs like raw materials, labor, and overhead expenses that are directly associated with the production process.

- **Column Description for fact_pre_invoice_deductions:**

1. customer_code: The 'customer_code' column features unique identification codes for every customer in the dataset. These codes can be used to track a customer's sales history, demographic information, and other relevant details. For example, the codes could look like '70002017', '90005160', and '80007195' respectively.

2. fiscal_year: The "fiscal_year" column holds the fiscal period when the sale of a product occurred.

3. pre_invoice_discount_pct: The "pre_invoice_discount_pct" column contains the percentage of pre-invoice deductions for each product. Pre-invoice deductions are discounts that are applied to the gross price of a product before

the invoice is generated, and typically applied to large orders or long-term contracts.

- **Column Description for fact_sales_monthly:**

1. date: The "date" column contains the date when the sale of a product was made, in a monthly format for 2020 and 2021 fiscal years. This information can be used to understand the sales performance of products over time.

2. product_code: The "product_code" column contains a unique identification code for each product. This code is used to track and differentiate individual products within a database or system.

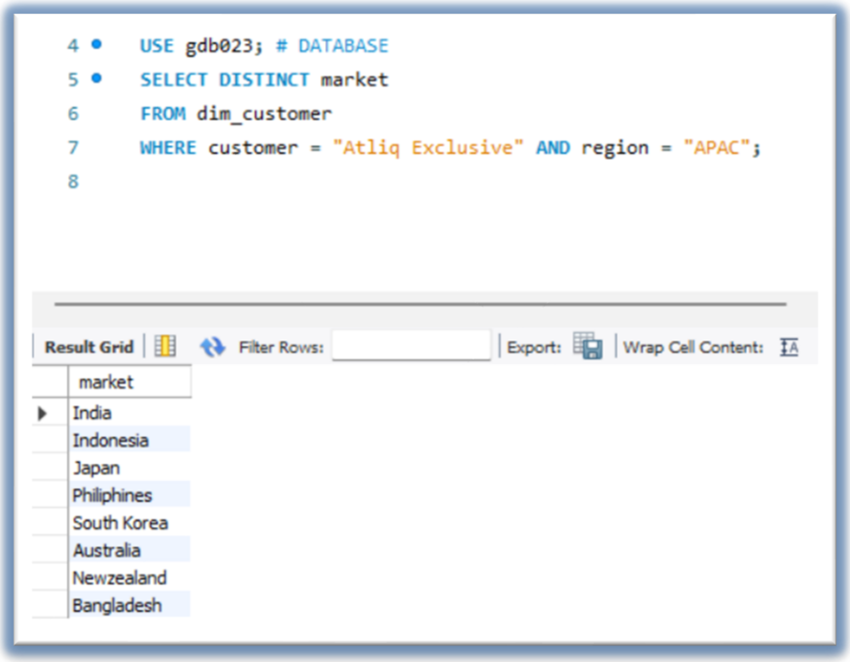
3. customer_code: The 'customer_code' column features unique identification codes for every customer in the dataset. These codes can be used to track a customer's sales history, demographic information, and other relevant details. For example, the codes could look like '70002017', '90005160', and '80007195' respectively.

4. sold_quantity: The "sold_quantity" column contains the number of units of a product that were sold. This information can be used to understand the sales volume ofproducts and to compare the sales volume of different products or variants.

5. fiscal_year: The "fiscal_year" column holds the fiscal period when the sale of a product occurred.

2.2. SQL Queries

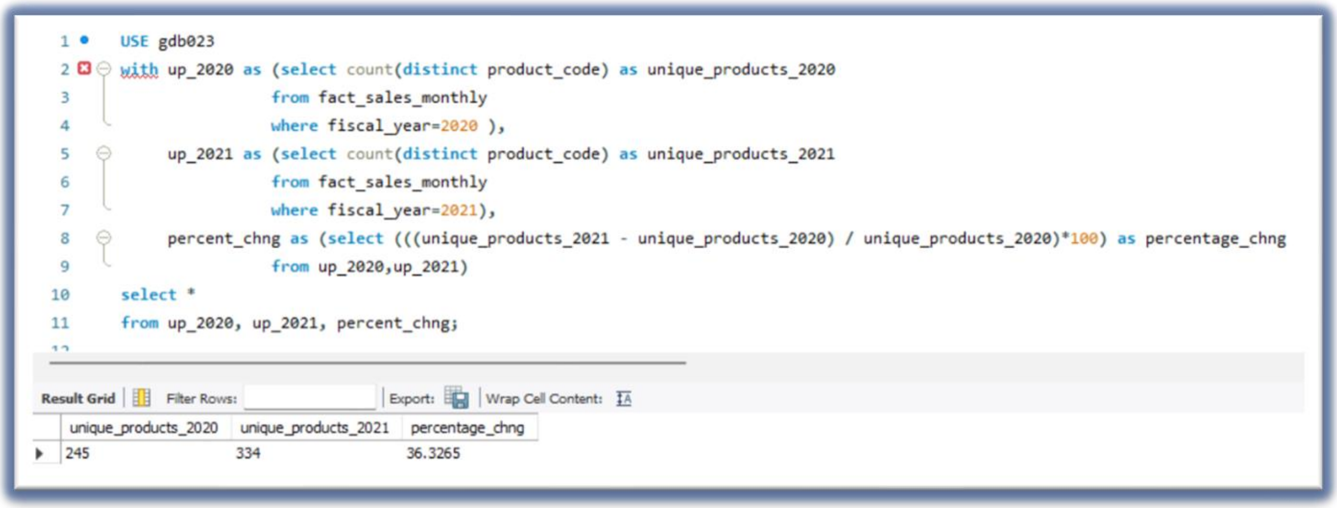
1 Provide the list of markets in which customer "Atliq Exclusive" operates its business in the APAC region.



```
USE gdb023; # DATABASE
SELECT DISTINCT market
FROM dim_customer
WHERE customer = "Atliq Exclusive" AND region = "APAC";
```

market
India
Indonesia
Japan
Philiphines
South Korea
Australia
Newzealand
Bangladesh

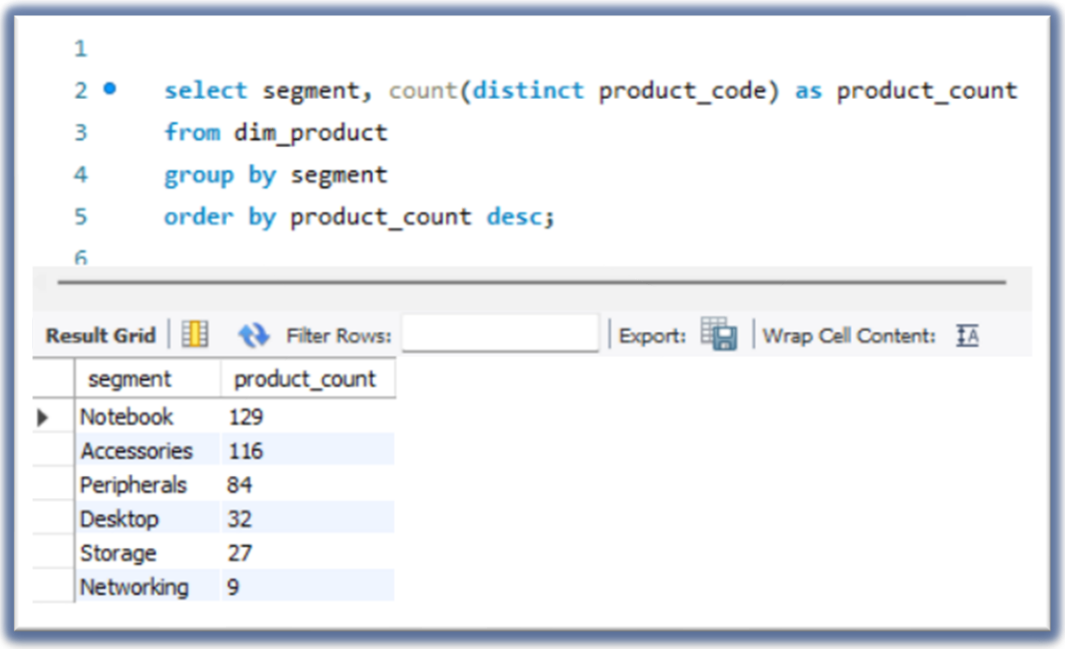
2 What is the percentage of unique product increase in 2021 vs. 2020? The final output contains below fields:
unique_products_2020, unique_products_2021, percentage_chg.



```
USE gdb023
with up_2020 as (select count(distinct product_code) as unique_products_2020
                from fact_sales_monthly
                where fiscal_year=2020 ),
     up_2021 as (select count(distinct product_code) as unique_products_2021
                from fact_sales_monthly
                where fiscal_year=2021),
     percent_chng as (select (((unique_products_2021 - unique_products_2020) / unique_products_2020)*100) as percentage_chng
                from up_2020,up_2021)
select *
from up_2020, up_2021, percent_chng;
```

unique_products_2020	unique_products_2021	percentage_chng
245	334	36.3265

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.



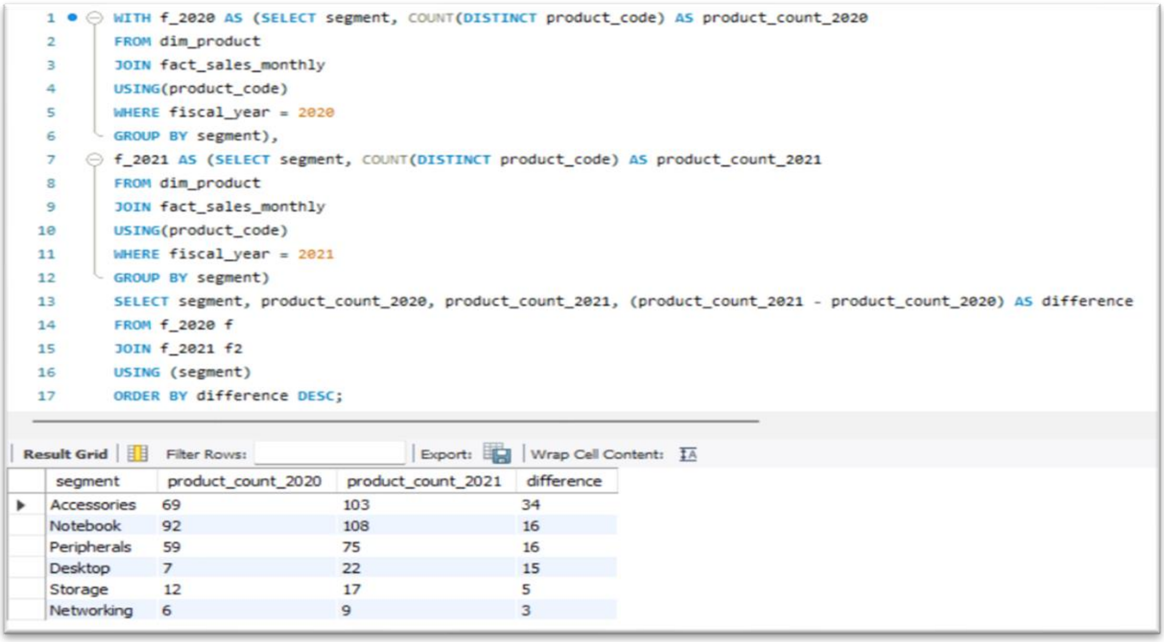
```sql

select segment, count(distinct product_code) as product_count
from dim_product
group by segment
order by product_count desc;

```

segment	product_count
Notebook	129
Accessories	116
Peripherals	84
Desktop	32
Storage	27
Networking	9

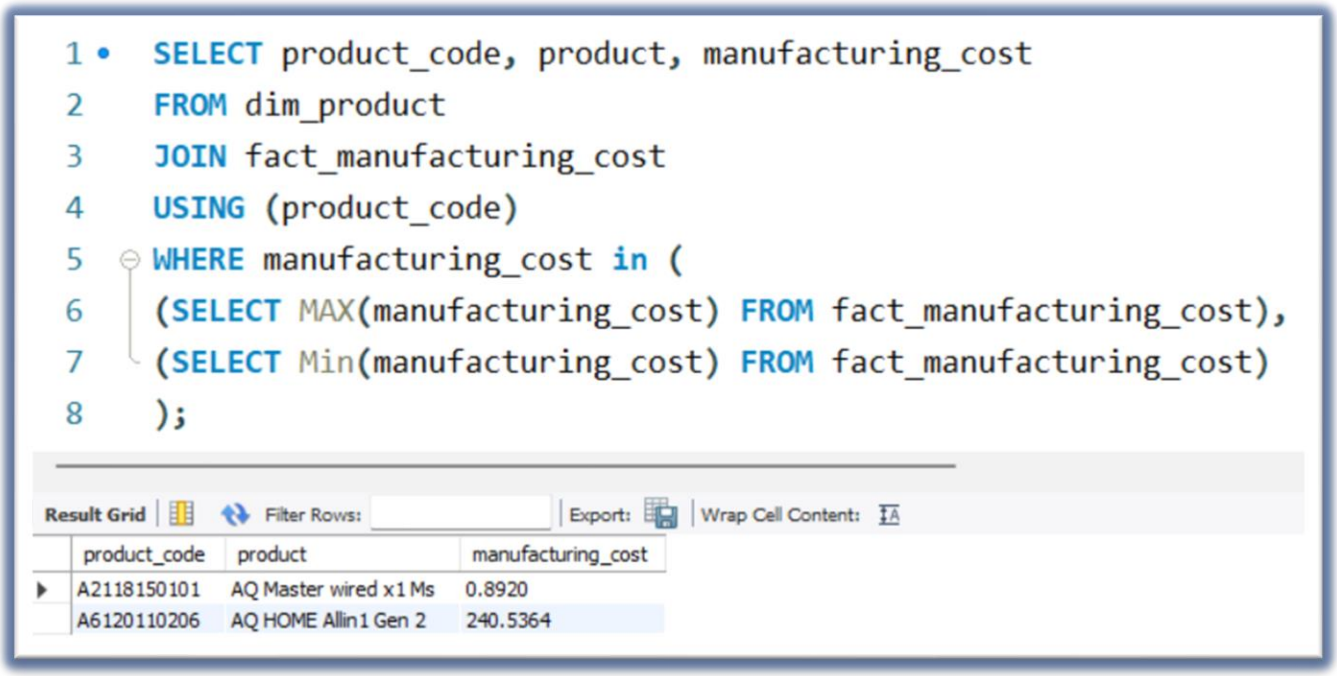
4 Follow-up: Which segment had the most increase in unique products in 2021 vs 2020? The final output contains these fields:
Segment, product_count_2020, product_count_2021, difference.



```sql
WITH f_2020 AS (SELECT segment, COUNT(DISTINCT product_code) AS product_count_2020
FROM dim_product
JOIN fact_sales_monthly
USING(product_code)
WHERE fiscal_year = 2020
GROUP BY segment),
f_2021 AS (SELECT segment, COUNT(DISTINCT product_code) AS product_count_2021
FROM dim_product
JOIN fact_sales_monthly
USING(product_code)
WHERE fiscal_year = 2021
GROUP BY segment)
SELECT segment, product_count_2020, product_count_2021, (product_count_2021 - product_count_2020) AS difference
FROM f_2020 f
JOIN f_2021 f2
USING (segment)
ORDER BY difference DESC;
```

segment	product_count_2020	product_count_2021	difference
Accessories	69	103	34
Notebook	92	108	16
Peripherals	59	75	16
Desktop	7	22	15
Storage	12	17	5
Networking	6	9	3

5 Get the products that have the highest and lowest manufacturing costs. The final output should contain these fields: product_code, product, manufacturing_cost.

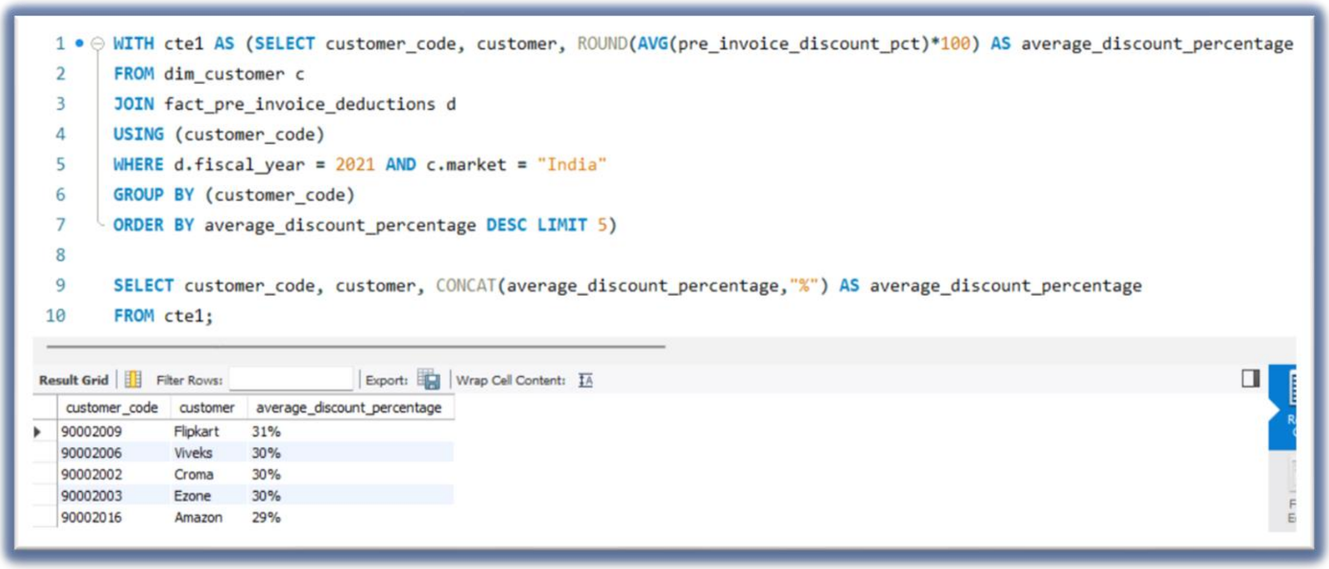


```sql
SELECT product_code, product, manufacturing_cost
FROM dim_product
JOIN fact_manufacturing_cost
USING (product_code)
WHERE manufacturing_cost in (
(SELECT MAX(manufacturing_cost) FROM fact_manufacturing_cost),
(SELECT Min(manufacturing_cost) FROM fact_manufacturing_cost)
);
```

product_code	product	manufacturing_cost
A2118150101	AQ Master wired x1 Ms	0.8920
A6120110206	AQ HOME Allin1 Gen 2	240.5364

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.



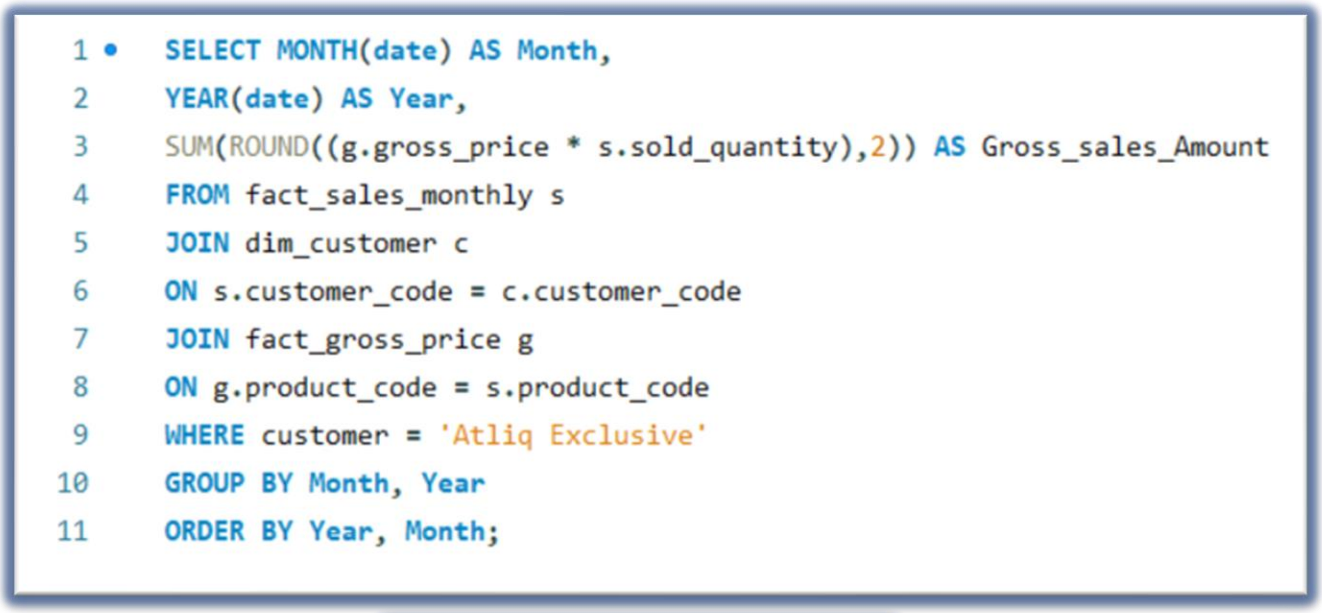
```sql
WITH cte1 AS (SELECT customer_code, customer, ROUND(AVG(pre_invoice_discount_pct)*100) AS average_discount_percentage
FROM dim_customer c
JOIN fact_pre_invoice_deductions d
USING (customer_code)
WHERE d.fiscal_year = 2021 AND c.market = "India"
GROUP BY (customer_code)
ORDER BY average_discount_percentage DESC LIMIT 5)

SELECT customer_code, customer, CONCAT(average_discount_percentage,"%") AS average_discount_percentage
FROM cte1;
```

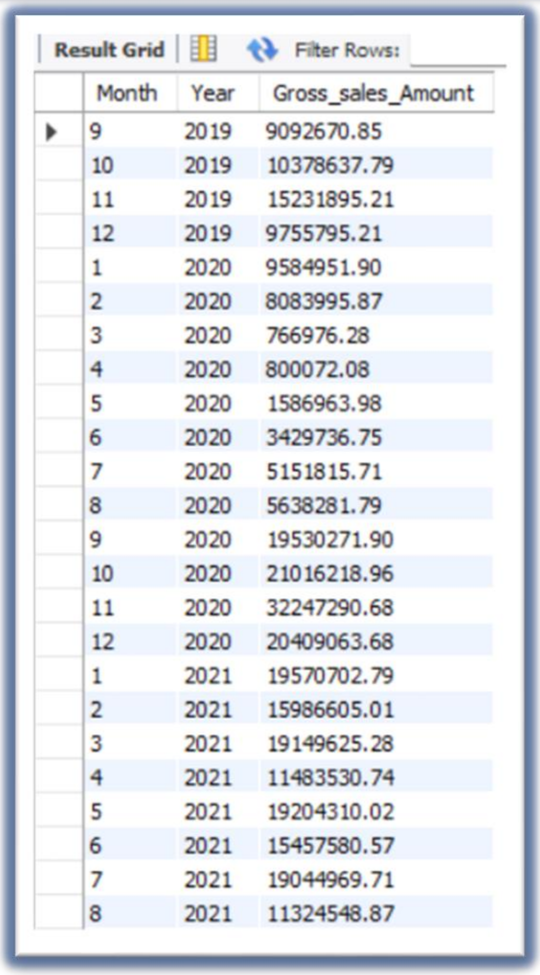
customer_code	customer	average_discount_percentage
90002009	Flipkart	31%
90002006	Viveks	30%
90002002	Croma	30%
90002003	Ezone	30%
90002016	Amazon	29%

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, Gross sales Amount.



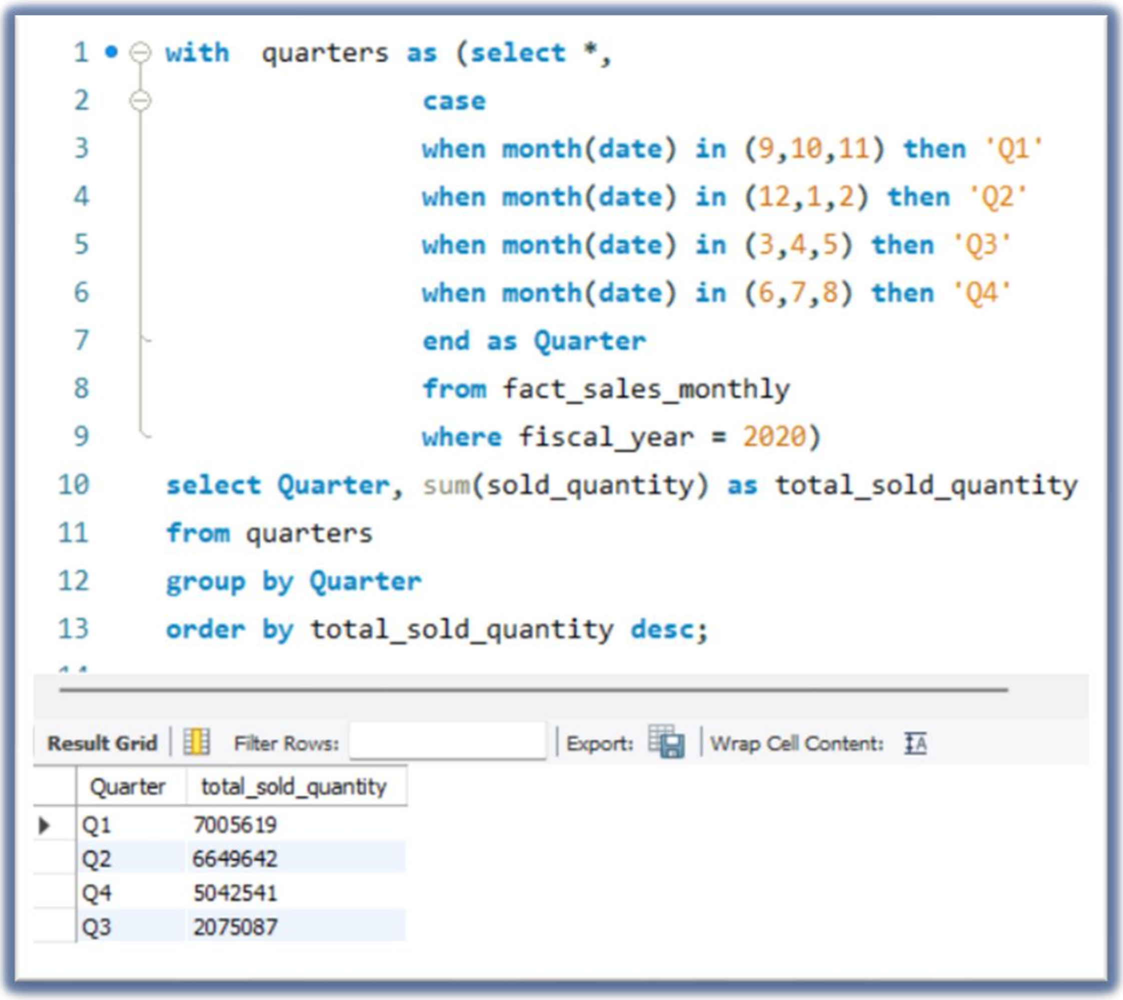
```sql
SELECT MONTH(date) AS Month,
YEAR(date) AS Year,
SUM(ROUND((g.gross_price * s.sold_quantity),2)) AS Gross_sales_Amount
FROM fact_sales_monthly s
JOIN dim_customer c
ON s.customer_code = c.customer_code
JOIN fact_gross_price g
ON g.product_code = s.product_code
WHERE customer = 'Atliq Exclusive'
GROUP BY Month, Year
ORDER BY Year, Month;
```



Month	Year	Gross_sales_Amount
9	2019	9092670.85
10	2019	10378637.79
11	2019	15231895.21
12	2019	9755795.21
1	2020	9584951.90
2	2020	8083995.87
3	2020	766976.28
4	2020	800072.08
5	2020	1586963.98
6	2020	3429736.75
7	2020	5151815.71
8	2020	5638281.79
9	2020	19530271.90
10	2020	21016218.96
11	2020	32247290.68
12	2020	20409063.68
1	2021	19570702.79
2	2021	15986605.01
3	2021	19149625.28
4	2021	11483530.74
5	2021	19204310.02
6	2021	15457580.57
7	2021	19044969.71
8	2021	11324548.87

8 In which quarter of 2020, got the maximum total_sold_quantity? The final output contains these fields sorted by the total_sold_quantity:

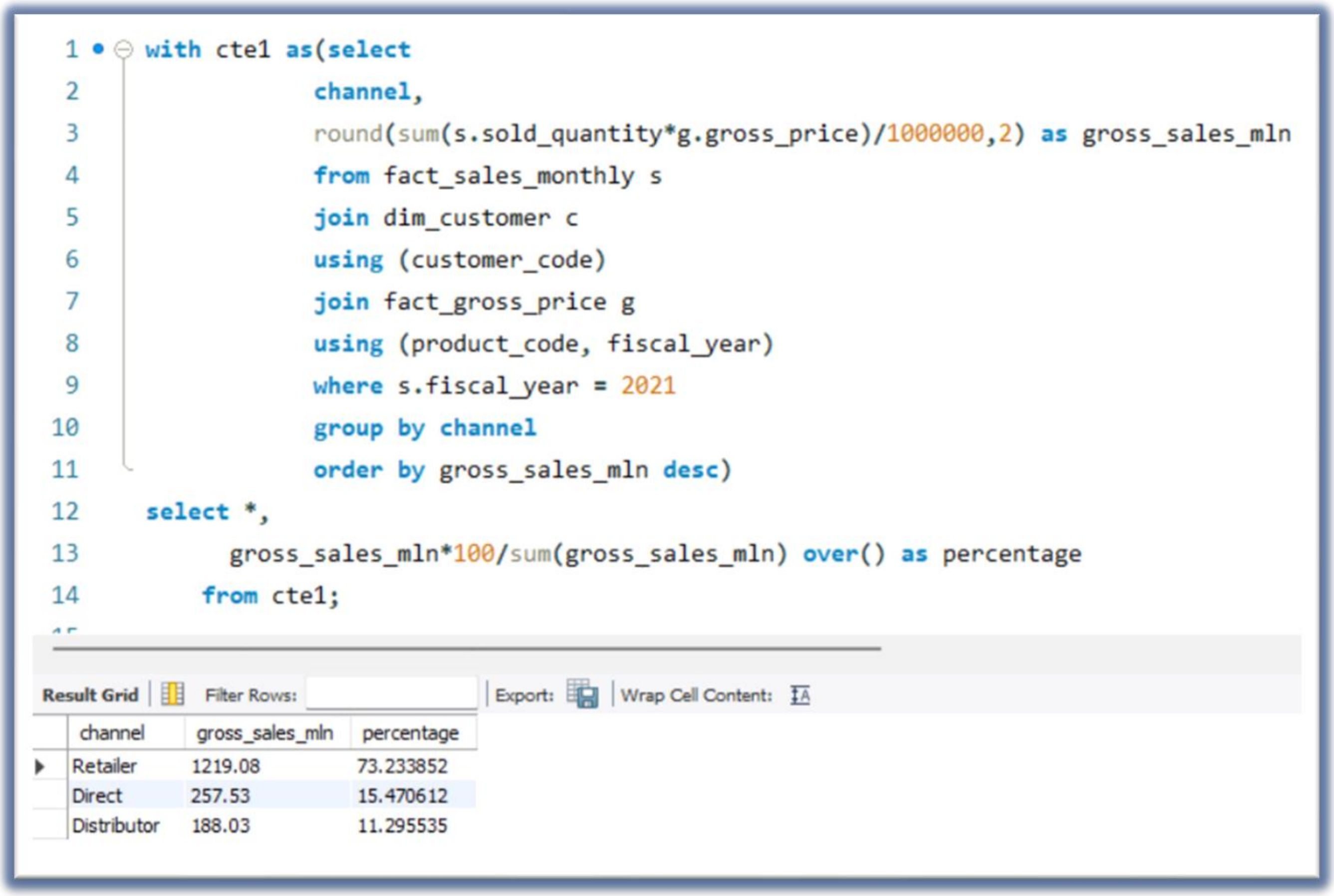
Quarter, total_sold_quantity.



```
with  quarters as (select *,
                case
                when month(date) in (9,10,11) then 'Q1'
                when month(date) in (12,1,2) then 'Q2'
                when month(date) in (3,4,5) then 'Q3'
                when month(date) in (6,7,8) then 'Q4'
                end as Quarter
                from fact_sales_monthly
                where fiscal_year = 2020)
select Quarter, sum(sold_quantity) as total_sold_quantity
from quarters
group by Quarter
order by total_sold_quantity desc;
```

Quarter	total_sold_quantity
Q1	7005619
Q2	6649642
Q4	5042541
Q3	2075087

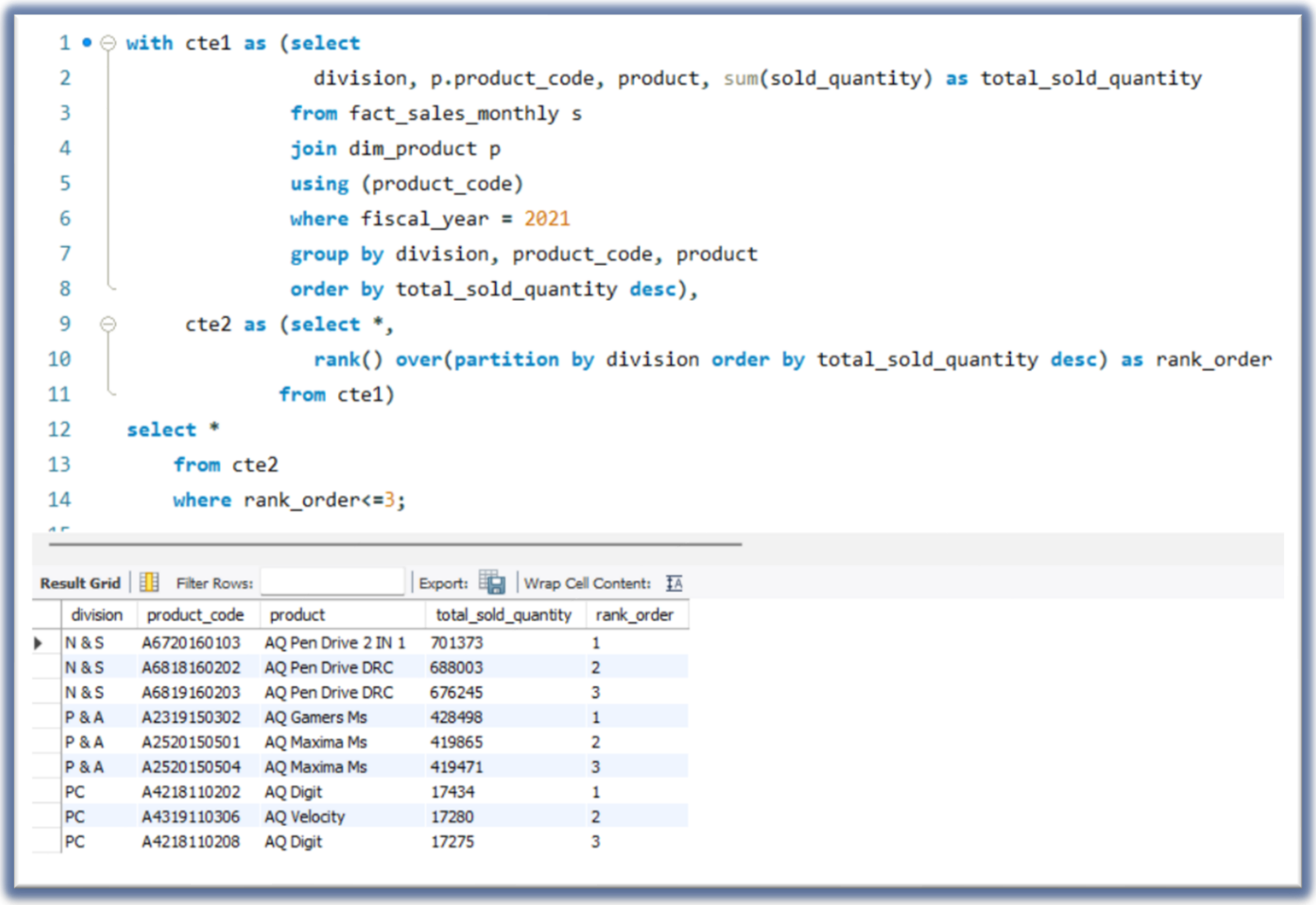
9 Which channel helped to bring more gross sales in the fiscal year 2021 and the percentage of contribution? The final output contains these fields:
Channel, gross_sales_mln, percentage.



```
with cte1 as(select
                channel,
                round(sum(s.sold_quantity*g.gross_price)/1000000,2) as gross_sales_mln
                from fact_sales_monthly s
                join dim_customer c
                using (customer_code)
                join fact_gross_price g
                using (product_code, fiscal_year)
                where s.fiscal_year = 2021
                group by channel
                order by gross_sales_mln desc)
select *,
        gross_sales_mln*100/sum(gross_sales_mln) over() as percentage
    from cte1;
```

channel	gross_sales_mln	percentage
Retailer	1219.08	73.233852
Direct	257.53	15.470612
Distributor	188.03	11.295535

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, product, total_sold_quantity, rank_order.



```
with cte1 as (select
                division, p.product_code, product, sum(sold_quantity) as total_sold_quantity
              from fact_sales_monthly s
              join dim_product p
              using (product_code)
              where fiscal_year = 2021
              group by division, product_code, product
              order by total_sold_quantity desc),
     cte2 as (select *,
                rank() over(partition by division order by total_sold_quantity desc) as rank_order
             from cte1)
select *
    from cte2
    where rank_order<=3;
```

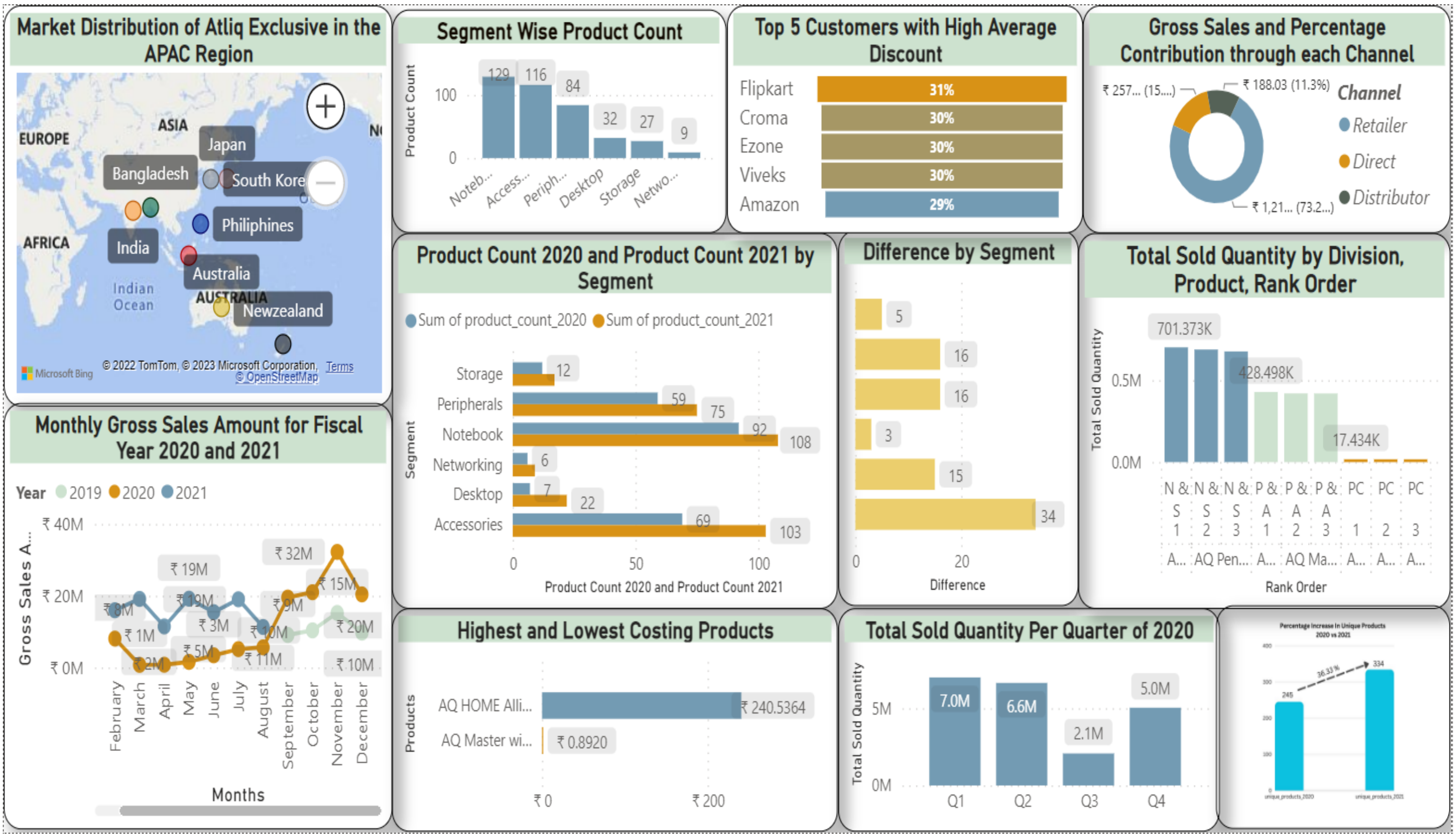
division	product_code	product	total_sold_quantity	rank_order
N & S	A6720160103	AQ Pen Drive 2 IN 1	701373	1
N & S	A6818160202	AQ Pen Drive DRC	688003	2
N & S	A6819160203	AQ Pen Drive DRC	676245	3
P & A	A2319150302	AQ Gamers Ms	428498	1
P & A	A2520150501	AQ Maxima Ms	419865	2
P & A	A2520150504	AQ Maxima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

3 Reporting

Reporting is a most important and underrated skill of a data analytics field. Because of being a Data Analyst, you should be good in easy and self-explanatory report because your model will beused by many stakeholders who are not from technical background.

- High-Level Design Document (HLD)
- Low-Level Design Document (LLD)
- PowerPoint Presentation

4 Deployment


Market Distribution of Atliq Exclusive in the APAC Region
EUROPE
ASIA
AFRICA
Japan
Bangladesh
South Kore
Philiphines
India
Australia
AUSTRALIA
Newzealand
Indian Ocean
Microsoft Bing
© 2022 TomTom, © 2023 Microsoft Corporation, Terms
© OpenStreetMap
Segment Wise Product Count
Product Count
100
0
129
116
84
32
27
9
Noteb...
Access...
Periph...
Desktop
Storage
Netwo...
Top 5 Customers with High Average Discount
Flipkart 31%
Croma 30%
Ezone 30%
Viveks 30%
Amazon 29%
Gross Sales and Percentage Contribution through each Channel
₹ 257... (15....)
₹ 188.03 (11.3%)
₹ 1,21... (73.2...)
Channel
Retailer
Direct
Distributor
Monthly Gross Sales Amount for Fiscal Year 2020 and 2021
Year ● 2019 ● 2020 ● 2021
Gross Sales A...
₹ 40M
₹ 20M
₹ 0M
₹ 19M
₹ 32M
₹ 19M
₹ 9M
₹ 15M
₹ 1M
₹ 3M
₹ 20M
₹ 5M
₹ 11M
₹ 10M
February
March
April
May
June
July
August
September
October
November
December
Months
Product Count 2020 and Product Count 2021 by Segment
● Sum of product_count_2020 ● Sum of product_count_2021
Segment
Storage 12
Peripherals 59 75
Notebook 92 108
Networking 6
Desktop 7 22
Accessories 69 103
0
50
100
Product Count 2020 and Product Count 2021
Difference by Segment
5
16
16
3
15
34
0
20
Difference
Total Sold Quantity by Division, Product, Rank Order
Total Sold Quantity
0.5M
0.0M
701.373K
428.498K
17.434K
N & S 1
N & S 2
N & S 3
P & A 1
P & A 2
P & A 3
PC 1
PC 2
PC 3
A...
AQ Pen...
A...
AQ Ma...
A...
A...
A...
Rank Order
Highest and Lowest Costing Products
Products
AQ HOME Alli... ₹ 240.5364
AQ Master wi... ₹ 0.8920
₹ 0
₹ 200
Total Sold Quantity Per Quarter of 2020
Total Sold Quantity
5M
0M
7.0M
6.6M
2.1M
5.0M
Q1
Q2
Q3
Q4
Percentage Increase In Unique Products 2020 vs 2021
36.33 %
245
334
unique_products_2020
unique_products_2021