PROJECT REPORT

CONTENTS:

* PROBLEM STATEMENT
* SALES AND DELIVERY DATASET
  + DATA DESCRIPTION
  + OBJECTIVES
  + ANALYSIS
    - QUERY
    - OUTPUT
    - DESCRIPTION
    - INFERENCE
  + CONCLUSION

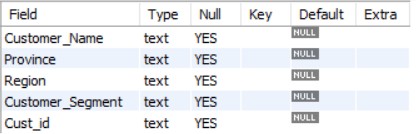
PROBLEM STATEMENT:

To analyze and seek insights for different scenarios from the composite data of a business organization, confined to the ‘sales and delivery’ domain is given for the period of the last decade.

* DATASET: SALES AND DELIVERY
* DATA DESCRIPTION List Of Tables:

## Cust\_dimen:

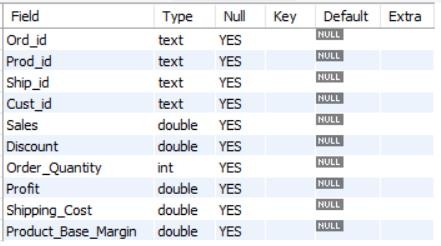
|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data type** | **description** |
| Customer\_name | Varchar | Name of the customer |
| Province | Varchar | ID for Province |
| Region | Varchar | ID for Region |
| Customer\_Segment | Varchar | Types of the customer segments |
| Cust\_id | Varchar | Id to the customers |



There are 5 columns in the table and all of them can contain null values.

## Market Fact:

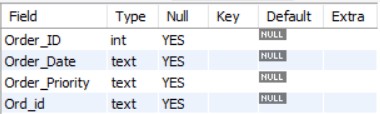
|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Types** | **description** |
| Ord id | Varchar | Id for the Order |
| Prod id | Varchar | Id for the product |
| Ship id | Varchar | ID for the shipping |
| Cust id | Varchar | ID for the customer |
| Sales | Float | The sales price for the product |
| Discount | Float | Discount for the product |
| Order Quantity | Float | Number of products have been ordered |
| Profit | Float | Profit that has been gained from the product |
| Shipping cost | Float | Shipping cost for the product |
| Product Base Margin | Float | Base margin value for the product |



This table contains 10 columns. The table contains columns with datatypes varchar, float, and int.

## Orders\_Dimen:

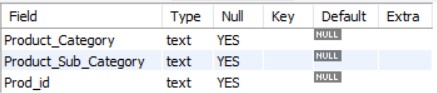
|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data type** | **Description** |
| Order id | integer | Id for the order |
| Order Date | Varchar | The order date for that order has been ordered |
| Order Priority | varchar | Priority for the orders |
| Ord id | varchar | Order id as a varchar |



This table consists of 4 columns. Three of the columns have text datatype and one column has integer datatype. All the columns can contain null values.

## Prod\_Dimen:

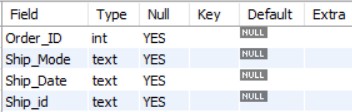
|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data type** | **Description** |
| Product\_Category | Varchar | Type of the product |
| Product\_Sub\_category | Varchar | Name of the sub- category |
| Prod\_id | Varchar | Id of Product |



This table contains 3 columns. All three have char datatype and contain null values.

## Shipping Dimen:

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Data Type** | **Description** |
| Order\_ID | Varchar | Id for the orders |
| Ship\_Mode | Varchar | Type of the shipping |
| Ship\_Date | Varchar | Shipping date |
| Ship\_ID | Varchar | ID for the shipping |



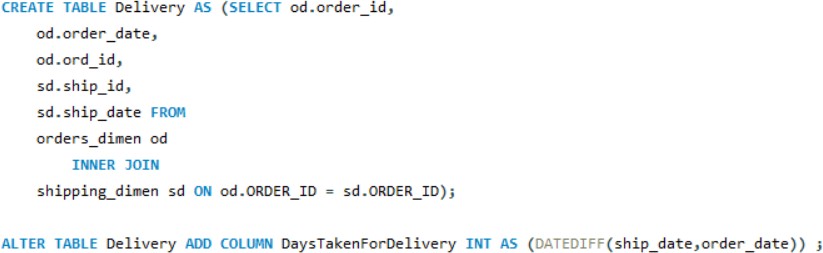
This table contains 4 columns. Three out of four columns have text values and one column has integer datatype.

Changes made in the tables:

* In the table order\_dimen, Changed the datatype of column, order\_date from text to date.
* In the table order\_dimen, Changed the datatype of column, order\_date from text to date.

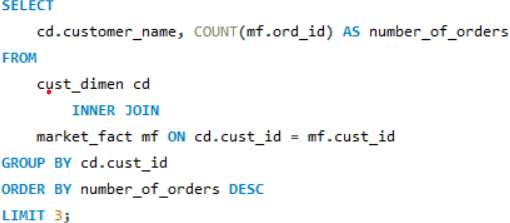


* Created a new table called ‘Delivery’ and populated it with data from the orders\_dimen and shipping\_dimen. In this table, we created a new column, ‘DaysTakenForDelivery’ that contains the date difference between Order\_Date and Ship\_Date.

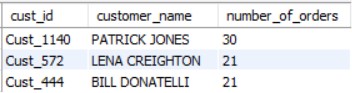


1. Objective:

To find the top 3 customers who have the maximum number of orders. Query**:**



Output:



Explanation:

In this query, we used an inner join between the customer dimension table and the market fact table on the common field of the customer ID. The COUNT function is used to count the number of order IDs in the market\_fact table that match each customer ID. GROUP BY groups the result set by the customer ID to ensure that (COUNT) is applied to each customer separately.

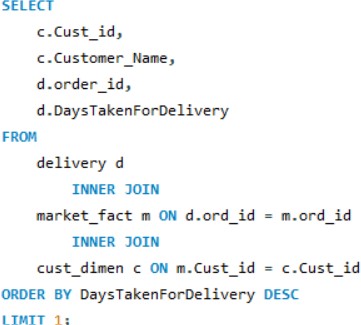
Inference:

The top three customers with maximum number of orders are Patrick, Lena, and Bill with 30,21, and 21 orders respectively.

This SQL query is helpful in providing valuable insights for sales and marketing strategies.

1. Objective:

Find the customer whose order took the maximum time to get delivered. Query:



Output:

Graphical user interface, application  Description automatically generated

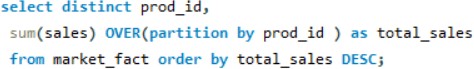
Explanation:

We used an inner join between the delivery table and the market fact table on the common field of the order ID (ord\_id). It then uses another inner join with the customer dimension table. The ORDER BY clause sorts the result set in descending order of the number of days taken for delivery. 'LIMIT' clause limits the result set to only the first row.

Inference:

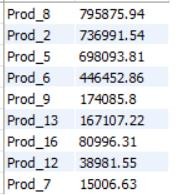
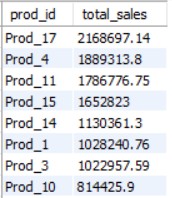
The customer whose order took the maximum time to get delivered is Dean Percer, who had to wait for 92 days to get his order delivered. This analysis can be used for further analysis to identify any issues in the delivery process and improve customer satisfaction.

1. Objective**:**

To retrieve total sales made by each product from the data. Query**:**

Output:

Number of rows: 17



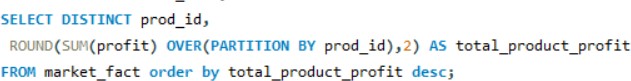
Explanation:

Here, DISTINCT keyword is used to retrieve only unique product IDs from the market fact table. SUM is used to calculate the total sales for each product ID by partitioning the result set by product ID using the OVER clause. This creates a separate group of rows for each product ID, and the SUM function is applied to each group to calculate the total sales.

Inference:

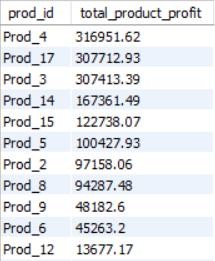
* From the output the product with product id 17 has made the maximum sales of more than 21 lakhs. And product id 7 made least sales of only 15 thousand.
* This query helps in analyzing the sales performance of different products and identifying the products with the highest total sales.

1. Objective**:**

To retrieve the total profit made from each product from the data. Query:

Output:

Number of rows: 17

Explanation:

We used SUM with the OVER clause to calculate the total profit for each product ID by partitioning the result set by product ID. This creates a separate group of rows for each product ID, and the SUM function is applied to each group to calculate the total profit.

Inference:

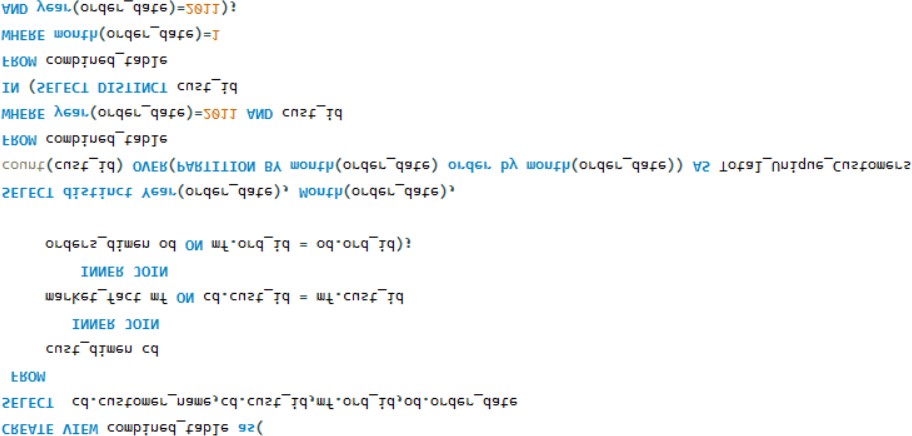
It can be seen that product 4 did the highest profit of more than 3 lakhs whereas product 7, product 16, product 10, and product 11 led to losses (product 11 was seen to have the highest loss of more than 1 lakh. It can also be observed that product 10 has fairly high sales but that is not turning into profits.

We should discuss this with the organization and try to find the reasons for it.

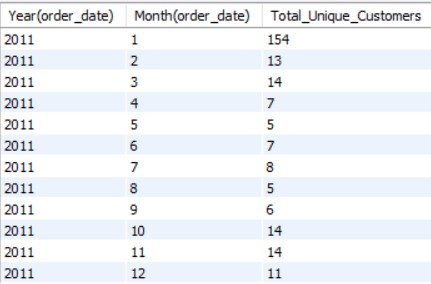
1. Objective**:**

To find the total number of unique customers in January and how many of them came back every month over the entire year in 2011.

Query**:**



Output:



Explanation:

The first query creates a **view** called **combined table** which combines data from three tables: 1. cust\_dimen

2. market\_fact, and

3. orders\_dimen.

It joins cust\_dimen and market\_fact on the cust\_id column and market\_fact and orders\_dimen on the ord\_id column.

The second query retrieves the **total number of unique customers for each month** in the year 2011, using a window function and a subquery.

The SELECT statement selects distinct Year and Month values from the order\_date column of the combined\_table and calculates the Total\_Unique\_Customers for each month. The count(cust\_id) OVER(PARTITION BY month(order\_date) order by month(order\_date)) calculates the count of unique cust\_id values for each month using a window function, and assigns the result to the Total\_Unique\_Customers column.

The WHERE clause filters the result to only include records from the year 2011 and where the cust\_id values are present in the subquery.

Takeaway and Conclusion:

It’s seen that a total of 154 customers were in January. Out of these customers, only minimum o f 5 customers visit every month.

* CONCLUSION:

From the above analysis, the important insights which can be derived are:

* The top three customers in the business organization were recognized. These customers must be retained by maintaining the same service and product quality.
* Maximum time taken for the delivery was observed to be 92 days. The organization needs to work on its delivery services. Also, they must try to give the required benefits to such customers who experience delays in delivery to compensate them.
* Certain products have high sales but still lead to losses. Since this could be due to various reasons, the organization must investigate them.
* According to the data, there are 154 customers at the start of the year, that is, in the month of January. But, only a few of them visit every month or even for half of the year. We should try to find out ways to retain our customers and make sure they utilize our services more often.