E-Commerce

Data Analysis using MySQL

Presented by Swati Pandit

2024

-

Analytics Project

Project Overview

This project involves analyzing an e-commerce database using MySQL. The goal is to derive actionable insights from customer, product, and order data. The analysis focuses on understanding customer distribution, product performance, and market trends, which can be used for targeted marketing, inventory management, and business strategy optimization.

Problem Statement and Objectives

As a Data Analyst at a dynamic e-commerce company, you're tasked with leveraging our extensive databases to extract insights that drive our business strategies forward. Your analysis will inform various departments, from marketing to supply chain, providing them with actionable data to optimize our operations, enhance customer satisfaction, and boost our sales performance.

01. Analyze customer distribution to identify key markets and customer segments.

02. Evaluate product performance based on sales data to identify high-revenue items and trends.

03. Segment customers based on purchasing behavior to tailor marketing strategies.

04. Understand market trends to optimize logistics and business operations.

Data-set Tables

Customers Dataset

- customer_id
- name
- location

Products Dataset

- product_id
- name
- category
- price

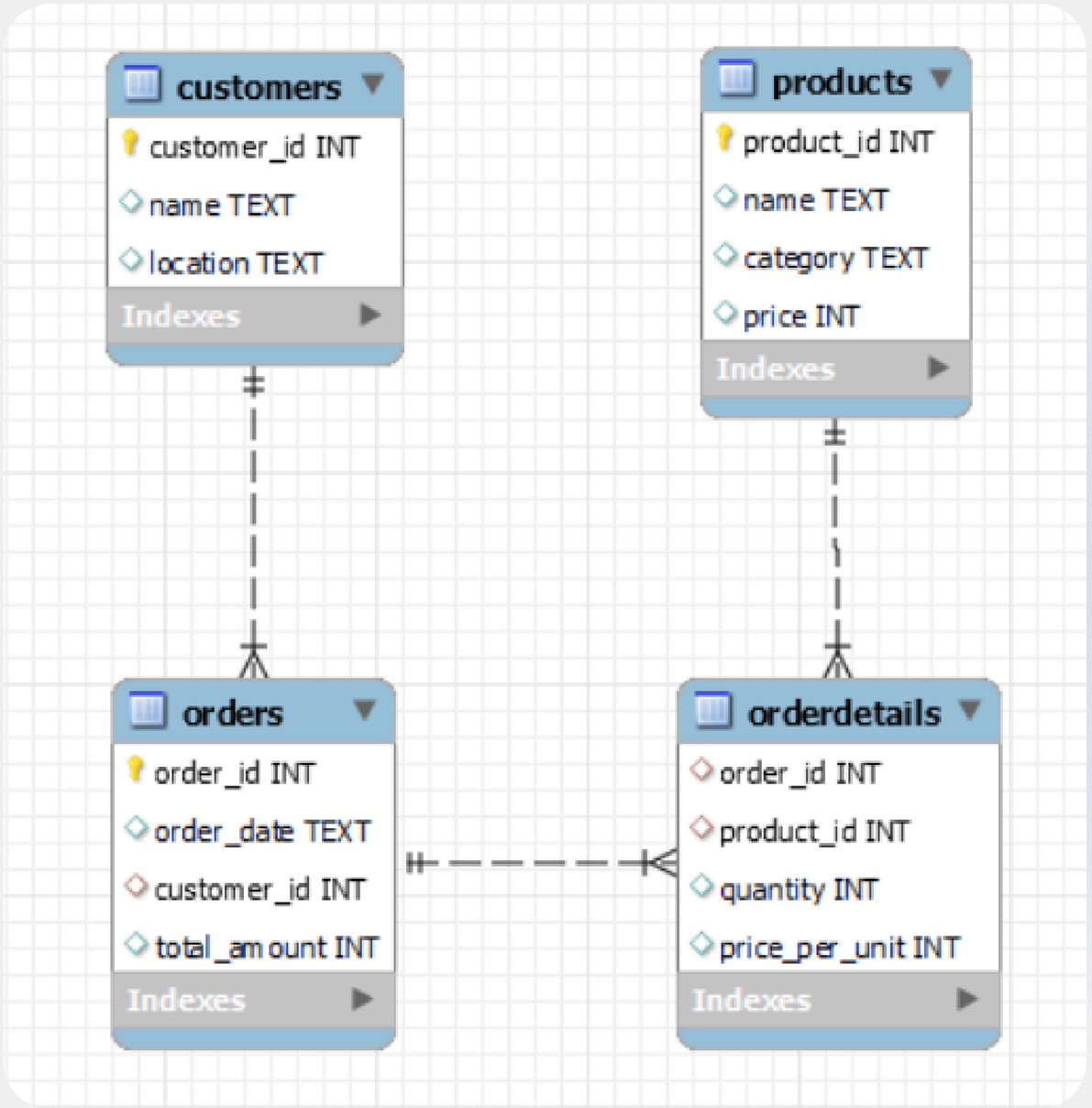
Orders Dataset

- order_id
- order_date
- customer_id
- total_amount

OrderDetails Dataset

- order_id
- product_id
- quantity
- price_per_unit

08 ER DIAGRAM



Data Analysis

Identifying the top 3 cities with the highest number of customers

Purpose : To determine key markets for targeted marketing and logistic optimization

```
SELECT
    location, COUNT(*) AS number_of_customers
FROM
    Customers
GROUP BY location
ORDER BY number_of_customers DESC
LIMIT 3;
```

Output

location	number_of_customers
Delhi	16
Chennai	15
Jaipur	11

01

Determining the distribution of customers by the number of orders placed.

02

Purpose : to get insight like segmenting customers into one-time buyers, occasional shoppers, and regular customers for tailored marketing strategies

```
SELECT
    NumberOfOrders, COUNT(*) AS CustomerCount
FROM
    (SELECT
        COUNT(*) AS NumberOfOrders
    FROM
        Orders
    GROUP BY customer_id) AS t
GROUP BY NumberOfOrders
ORDER BY NumberOfOrders;
```

Output

NumberOfOrders	CustomerCount
1	26
2	26
3	18
4	6
5	6
6	1
8	1

Identifying products where the average purchase quantity per order is 2 but with a high total revenue

Purpose : For suggesting premium product trends

```
SELECT
    product_id AS Product_Id,
    AVG(quantity) AS AvgQuantity,
    SUM(quantity * price_per_unit) AS TotalRevenue
FROM
    OrderDetails
GROUP BY product_id
HAVING AVG(quantity) = 2
ORDER BY TotalRevenue DESC;
```

Output

Product_Id	AvgQuantity	TotalRevenue
1	2.0000	1620000
8	2.0000	390000

Q3

For each product category, calculated the unique number of customers purchasing from it

Purpose : This will help understand which categories have wider appeal across the customer base

```
SELECT
    category, COUNT(DISTINCT customer_id) AS unique_customers
FROM
    Products pt
        JOIN
    OrderDetails od ON pt.product_id = od.product_id
        JOIN
    Orders os ON od.order_id = os.order_id
GROUP BY category
ORDER BY unique_customers DESC;
```

Output

category	unique_customers
Electronics	79
Wearable Tech	61
Photography	45

04

Analyzed the month-on-month percentage change in total sales to identify growth trends

```sql
WITH helper_table AS (
    SELECT
        DATE_FORMAT(order_date, '%Y-%m') AS Month,
        SUM(total_amount) AS TotalSales
    FROM
        Orders
    GROUP BY Month
)

SELECT
    Month,
    TotalSales,
    ROUND(
        (
         (TotalSales - LAG(TotalSales) OVER (ORDER BY Month)) /
         LAG(TotalSales) OVER (ORDER BY Month)
        ) * 100,
        2
    ) AS PercentChange
FROM helper_table;
```

Output

Month	TotalSales	PercentChange
2023-03	789000	NULL
2023-04	1704000	115.97
2023-05	1582000	-7.16
2023-06	1040000	-34.26
2023-07	2568000	146.92
2023-08	1800000	-29.91
2023-09	2927000	62.61
2023-10	1497000	-48.86
2023-11	1151000	-23.11
2023-12	2774000	141.01
2024-01	1555000	-43.94
2024-02	396000	-74.53

05

Examine how the average order value changes month-on-month

06

Purpose : Insights can guide pricing and promotional strategies to enhance order value

```
WITH helper_table AS
(
   SELECT
     DATE_FORMAT(order_date,'%Y-%m') AS Month,
     AVG(total_amount) AS AvgOrderValue
   FROM Orders
   GROUP BY Month
)

SELECT
    Month,
    AvgOrderValue,
      ROUND((AvgOrderValue - (LAG(AvgOrderValue) OVER
(order by Month))),2) AS ChangeInValue
FROM helper_table;
```

Output

Month	AvgOrderValue	ChangeInValue
2023-03	60692.3077	NULL
2023-04	81142.8571	20450.55
2023-05	87888.8889	6746.03
2023-06	104000.0000	16111.11
2023-07	98769.2308	-5230.77
2023-08	112500.0000	13730.77
2023-09	121958.3333	9458.33
2023-10	83166.6667	-38791.67
2023-11	95916.6667	12750.00
2023-12	132095.2381	36178.57
2024-01	129583.3333	-2511.90
2024-02	44000.0000	-85583.33

07

Based on sales data, identified products with the fastest turnover rates

Purpose : For suggesting high demand and the need for frequent restocking

```
SELECT
    product_id, COUNT(*) SalesFrequency
FROM
    OrderDetails
GROUP BY product_id
ORDER BY SalesFrequency DESC
LIMIT 5;
```

Output

product_id	SalesFrequency
7	78
3	68
4	68
2	67
8	65

List products purchased by less than 40% of the customer base

08

```
set @total_count = (SELECT COUNT(DISTINCT customer_id) FROM customers);

-- Total number of unique customers

WITH product_detail AS
(
   SELECT
      products.product_id,
      COUNT(DISTINCT orders.customer_id)  AS UniqueCustomerCount
   FROM products
   JOIN OrderDetails
      ON products.product_id = OrderDetails.product_id
         JOIN orders
            ON OrderDetails.order_id = orders.order_id
   GROUP BY products.product_id
),
helper_table AS (
   SELECT
      pd.product_id,name,UniqueCustomerCount
   From products p
   JOIN product_detail pd ON p.product_id = pd.product_id
)
SELECT *
FROM helper_table
WHERE UniqueCustomerCount/@total_count < 0.4;
```

Purpose : Indicating potential mismatches between inventory and customer interest.

Output

product_id	name	UniqueCustomerCount
1	Smartphone 6"	36
8	Wireless Earbuds	38

Evaluate the month-on-month growth rate in the customer base

09

Purpose : To understand the effectiveness of marketing campaigns and market expansion efforts

```
WITH helper_table AS
(
   SELECT
      customer_id,
      MIN(order_date) AS firstpurchasedate
   FROM orders
   GROUP BY customer_id
)
SELECT
   DATE_FORMAT(firstpurchasedate,'%Y-%m') firstpurchasemonth,
   COUNT(*) TotalNewCustomers
FROM helper_table
GROUP BY firstpurchasemonth
ORDER BY firstpurchasemonth;
```

Output

firstpurchasemonth	TotalNewCustomers
2023-03	11
2023-04	18
2023-05	11
2023-06	8
2023-07	11
2023-08	9
2023-09	5
2023-10	3
2023-11	1
2023-12	4
2024-01	2
2024-02	1

Identify the months with the highest sales volume, aiding in planning

Purpose : For stock levels, marketing efforts, and staffing in anticipation of peak demand periods

```
SELECT
    DATE_FORMAT(order_date, '%Y-%m') AS Month,
    SUM(total_amount) TotalSales
FROM
    Orders
GROUP BY Month
ORDER BY TotalSales DESC
LIMIT 3;
```

Output

Month	TotalSales
2023-09	2927000
2023-12	2774000
2023-07	2568000

06 Insights

- A **downward** growth trend in the customer base suggests the current marketing campaigns are not as effective as intended.
- **Sales trends** fluctuated between March and August with no clear pattern.
- The largest **decline** in sales was seen in February 2024.
- **Product turnover** is highest for Product ID 7, indicating it requires frequent restocking.
- December saw the highest change in the **average order value**.

Insights

- **Electronics products** are in high demand and need more attention.
- Among products with an average purchase quantity of two, **Product 1** exhibit s the **highest** total revenue.
- The **Engagement Depth Analysis** shows that as the number of orders increases, the customer count decreases. Most of the customers are **Occasional** shoppers.
- Certain products purchased by less than 40% of the customer base have lower-than-expected purchase rates, likely due to **poor visibility** on the platform.

2 Recommendations

- **Implement targeted marketing campaigns** to address the decline in the customer base, specifically in underperforming areas like February.
- **Restock** for September and December in advance, and hire additional staff to handle increased demand.
- **Improve** the visibility of underperforming products (those purchased by less than 40% of customers) to boost sales.
- Focus marketing strategies on Delhi, Chennai, and Jaipur to tap into potential **high-performing regions**.
- Maintain **frequent stock checks** for Product ID 7 due to its high turnover rate.
- **Optimize marketing** efforts toward Electronics, as they are in high demand.
- Develop strategies to **convert Occasional shoppers** into regular customers, perhaps through loyalty programs or personalized offers.

Thank You

Checkout my GitHub Profile