E-Commerce Data Analysis: Sales Performance & Profitability

SQL-based Analysis using Google BigQuery and Tableau

Author: Vishnu M S
Data Source: https://www.kaggle.com/datasets/benroshan/ecommerce-data
Date: October 2025

Table of Contents

1. Introduction
2. Methodology
3. Data Cleaning
4. Analysis & Results
5. Conclusion
6. References

Introduction

This project presents an end-to-end data analysis of an e-commerce dataset covering January 2018 to December 2019. The dataset includes sales transactions, customer information, product details, and geographic locations. The primary goal is to evaluate sales performance, profitability, and customer value using SQL queries on Google BigQuery.
The dataset used in this analysis was sourced from Kaggle (E-Commerce Data by benroshan)

Problem Statement

The objective of this project is to understand the sales performance and profitability of the e-commerce business using transactional data. The analysis is guided by the following key business questions:

1. **Sales Trend:** What is the overall sales trend over time?
2. **Sales Targets:** How are we performing against sales targets?
3. **Product Performance:** Which product categories and sub-categories drive the most sales and profit?
4. **Customer Value:** Who are our most valuable customers?
5. **Geographic Insights:** What is the geographic distribution of our sales?

These questions form the foundation of the analysis and help identify growth opportunities, inefficiencies, and customer behavior patterns.

Methodology

The project followed a structured workflow:
1. Data Ingestion into Google BigQuery.
2. Data Cleaning using SQL (handling missing values, duplicates, and formatting).
3. Data Analysis with queries to answer specific business questions.
4. Data Visualization using Tableau Public. The complete, interactive dashboard summarizing the analysis is available online at: https://public.tableau.com/shared/T5Z2TWP9C?:display_count=n&:origin=viz_share_link

Data Cleaning

The following SQL queries were used for cleaning the dataset, ensuring quality and consistency.

1. List of Order Table

```
## To check the number of rows

SELECT
  count(*)
FROM
  `helpful-kingdom-469519-f5.e_commerce_data.list_of_orders`;

## Check for null values

SELECT
  countif (`Order ID`is null) AS order_id_null,
  countif (`Order Date` is null) AS order_date_null,
  countif (`CustomerName` is null) AS customer_name_null,
  countif (`State` is null) AS state_null,
```

```
  countif(`City` is null) AS city_null
FROM
  `helpful-kingdom-469519-f5.e_commerce_data.list_of_orders`;

## check for rows with all null values

SELECT
  count (*)
FROM
  `helpful-kingdom-469519-f5.e_commerce_data.list_of_orders`
WHERE
  `Order ID`is null AND
  `Order Date` is null AND
  `CustomerName` is null AND
  `State` is null AND
  `City` is null;

## Create a clean table

CREATE OR REPLACE TABLE
  `helpful-kingdom-469519-f5.e_commerce_data.list_of_orders_CLEAN` AS
SELECT
  *
FROM
  `helpful-kingdom-469519-f5.e_commerce_data.list_of_orders`
WHERE
  `Order ID` IS NOT NULL;

## Preview all the rows in the cleaned table

SELECT
 *
FROM
  `helpful-kingdom-469519-f5.e_commerce_data.list_of_orders_CLEAN`;

## Check for any null values in the cleaned table

SELECT
  count (*)
FROM
  `helpful-kingdom-469519-f5.e_commerce_data.list_of_orders_CLEAN`
WHERE
  `Order ID`is null or
  `Order Date` is null or
  `CustomerName` is null or
  `State` is null or
  `City` is null;

## Check for duplicates

SELECT
  `Order ID`,
  `Order Date`,
```

```
  CustomerName,
  State,
  City,
  COUNT(*) AS record_count
FROM
  `helpful-kingdom-469519-f5.e_commerce_data.list_of_orders_CLEAN`
GROUP BY
  `Order ID`,
  `Order Date`,
  CustomerName,
  State,
  City
HAVING
  record_count > 1
ORDER BY
  record_count DESC;

## Standardize the column and create a new cleaned table

SELECT
  TRIM(upper(CustomerName))AS customer_name_clean,
  TRIM(upper(State)) AS state_clean,
  TRIM(upper(City)) AS city_clean
FROM
 `helpful-kingdom-469519-f5.e_commerce_data.list_of_orders_CLEAN`;
CREATE OR REPLACE TABLE
  `helpful-kingdom-469519-f5.e_commerce_data.list_of_orders_FINAL` AS
SELECT
  TRIM(UPPER(CustomerName)) AS CustomerName,
  TRIM(UPPER(State)) AS State,
  TRIM(UPPER(City)) AS City,
  * EXCEPT (CustomerName, State, City)
FROM
  `helpful-kingdom-469519-f5.e_commerce_data.list_of_orders_CLEAN`;

## Check for any duplicates in the new table

SELECT
  TRIM(upper(CustomerName))AS customer_name_clean,
  TRIM(upper(State)) AS state_clean,
  TRIM(upper(City)) AS city_clean
FROM
 `helpful-kingdom-469519-f5.e_commerce_data.list_of_orders_CLEAN`;
CREATE OR REPLACE TABLE
  `helpful-kingdom-469519-f5.e_commerce_data.list_of_orders_FINAL` AS
SELECT
  TRIM(UPPER(CustomerName)) AS CustomerName,
  TRIM(UPPER(State)) AS State,
  TRIM(UPPER(City)) AS City,
  * EXCEPT (CustomerName, State, City)
FROM
  `helpful-kingdom-469519-f5.e_commerce_data.list_of_orders_CLEAN`;
```

2. Order Details Table

```
##Check for any null values in the columns

SELECT
  count(*)
FROM
  `helpful-kingdom-469519-f5.e_commerce_data.order_details`
WHERE
  `Order ID` is null or
  `Amount` is null or
  `Profit` is null or
  `Quantity` is null or
  `Category` is null or
  `Sub-Category` is null;

## Check for duplicates

SELECT
  `Order ID`,
  `Amount`,
  `Profit`,
  `Quantity`,
  `Category`,
  `Sub-Category`,
  count(*) AS record_count
FROM
  `helpful-kingdom-469519-f5.e_commerce_data.order_details`
GROUP BY
  `Order ID`,
  `Amount`,
  `Profit`,
  `Quantity`,
  `Category`,
  `Sub-Category`
HAVING
  record_count>1;

## New table after cleaning

CREATE OR REPLACE TABLE
  `helpful-kingdom-469519-f5.e_commerce_data.order_details_final`
AS
SELECT
  TRIM(UPPER(Category)) AS category_CLEAN,
  TRIM(UPPER(`Sub-Category`)) AS sub_category_CLEAN,
  * EXCEPT (`Category`, `Sub-Category`)
FROM
  `helpful-kingdom-469519-f5.e_commerce_data.order_details`;
```

Check for duplicates in the new table

```
SELECT
  category_CLEAN,
  sub_category_CLEAN,
  `Order ID`,
  Amount,
  Profit,
  Quantity,
  COUNT(*) AS record_count
FROM
 `helpful-kingdom-469519-f5.e_commerce_data.order_details_final`
GROUP BY
  category_CLEAN,
  sub_category_CLEAN,
  `Order ID`,
  Amount,
  Profit,
  Quantity
HAVING
  record_count > 1;
```

3. Sales Target Table

Check for null values

```
SELECT
  count(*)
FROM
  `helpful-kingdom-469519-f5.e_commerce_data.sales_target`
WHERE
  `Month of Order Date` is null or
  `Category` is null or
  `Target` is null;
```

Check for duplicates

```
SELECT
  `Month of Order Date`,
  `Category`,
  `Target`,
  count(*) AS record_count
FROM
  `helpful-kingdom-469519-f5.e_commerce_data.sales_target`
GROUP BY
  `Month of Order Date`,
  `Category`,
  `Target`
HAVING
  record_count > 1;
```

```
##  Inspect the original values in the Month of Order Date column

SELECT
  DISTINCT `Month of Order Date`
FROM
  `helpful-kingdom-469519-f5.e_commerce_data.sales_target`
  LIMIT 50;

## To  convert a non-standard date string

SELECT
  parse_date('%b-%y',`Month of Order Date`) As order_month
FROM
  `helpful-kingdom-469519-f5.e_commerce_data.sales_target`;

## clean and standardize date column by converting a text string like "Jan-21" into string format like "2021-01".

SELECT
  FORMAT_DATE('%Y-%m',PARSE_DATE('%b-%y', `Month of Order Date`)) AS clean_order_month
FROM
  `helpful-kingdom-469519-f5.e_commerce_data.sales_target`;

## Create a new table with cleaned data

CREATE OR REPLACE TABLE
  `helpful-kingdom-469519-f5.e_commerce_data.sales_target_final`
AS
SELECT
  TRIM(UPPER(Category)) AS category_final,
  FORMAT_DATE('%Y-%m', PARSE_DATE('%b-%y', `Month of Order Date`)) AS clean_order_month,
  * EXCEPT (Category, `Month of Order Date`)
FROM
  `helpful-kingdom-469519-f5.e_commerce_data.sales_target`;

## Check for duplicates in the new table

SELECT
  category_final,
  clean_order_month,
  `Target`,
  COUNT(*) AS record_count
FROM
 `helpful-kingdom-469519-f5.e_commerce_data.sales_target_final`
GROUP BY
  category_final,
  clean_order_month,
  `Target`
HAVING
  record_count > 1;
```

Data Analysis

1. Sales Trend

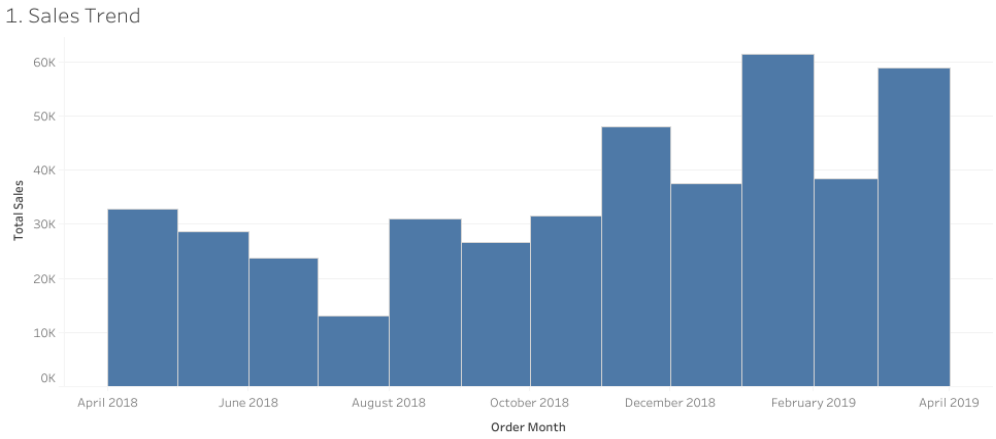
##Create a view or derived table with month-year

```
CREATE OR REPLACE TABLE
 `helpful-kingdom-469519-f5.e_commerce_data.list_of_order_finalwithmonth`
AS
SELECT
  *,
  FORMAT_DATE('%Y-%m', DATE(`Order Date`)) AS order_month
FROM
  `helpful-kingdom-469519-f5.e_commerce_data.list_of_orders_FINAL`;
```

calculate total sales per month

```
SELECT
  o.order_month,
  SUM(od.Amount) AS total_sales
FROM
  `helpful-kingdom-469519-f5.e_commerce_data.list_of_order_finalwithmonth` AS o
JOIN
  `helpful-kingdom-469519-f5.e_commerce_data.order_details_final`
AS
  od
ON
  o.`Order ID` = od.`Order ID`
GROUP BY
  o.order_month
ORDER BY
  o.order_month;
```

Visualization



Insight

- Sales show a clear upward trend, increasing significantly from a low of $12,966 in July 2018 to a peak of $61,439 in January 2019, and remaining high in March 2019. This indicates business expansion.
- The trend is highly volatile, with sharp drops (e.g., a 38% drop from January to February 2019) and rises (e.g., a 63% jump from December 2018 to January 2019).

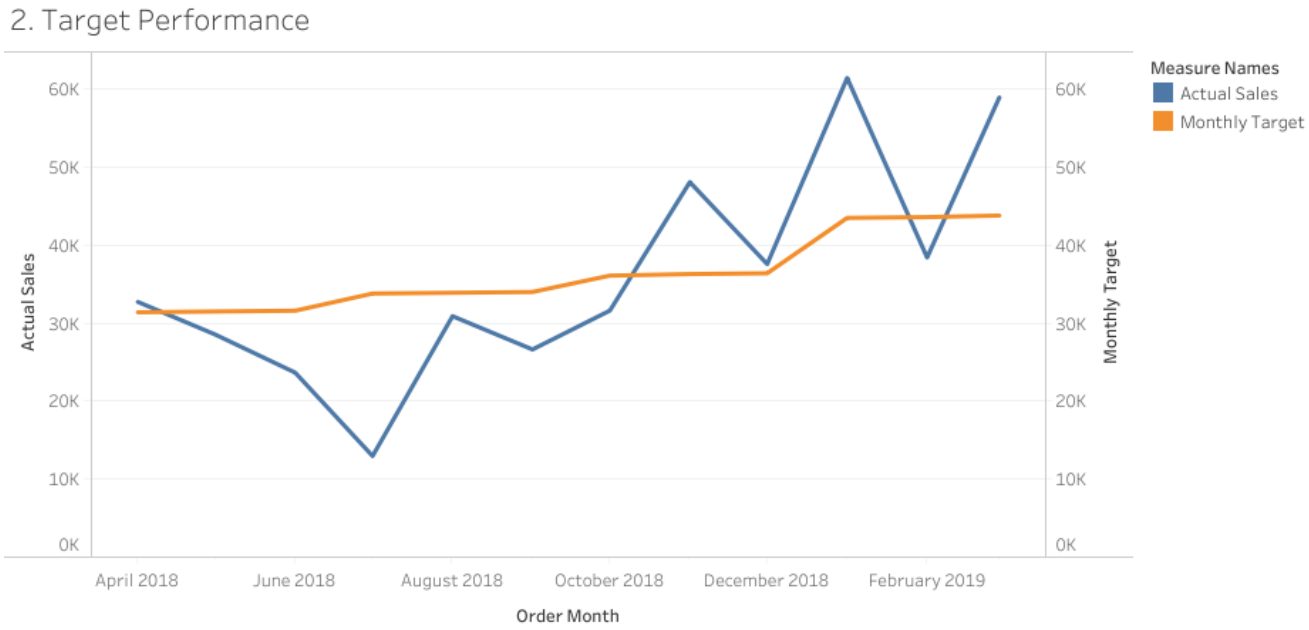
Actionable Next Steps: The business must investigate the cause of the volatility to stabilize demand and ensure resource

2. Target Performance

Calculating actual sales and montly target

```
SELECT
  t3.clean_order_month,
  t3.category_final,
  SUM(t2.Amount) AS Actual_Sales,
  t3.Target AS Monthly_Target,
  (SUM(t2.Amount) - t3.Target) AS Performance_Delta,
  SAFE_DIVIDE(SUM(t2.Amount), t3.Target)  AS Target_Attainment_Ratio
FROM
  `helpful-kingdom-469519-f5.e_commerce_data.list_of_order_finalwithmonth` AS t1
INNER JOIN
  `helpful-kingdom-469519-f5.e_commerce_data.order_details_final` AS t2
ON
  t1.`Order ID` = t2.`Order ID`
INNER JOIN
  `helpful-kingdom-469519-f5.e_commerce_data.sales_target_final`
AS
  t3
ON
  t2.category_CLEAN = t3.category_final
AND
  t1.order_month = t3.clean_order_month
GROUP BY
  t3.clean_order_month,
  t3.category_final,
  t3.Target
ORDER BY
  t3.clean_order_month,
  t3.category_final;
```

Visualization



Insight

- Overall targets are missed due to consistent underperformance in Clothing and Furniture.
- Electronics is the key revenue driver, hitting its target in almost every month.
- The only period of collective success was January 2019, when all categories simultaneously exceeded targets.

Actionable Next Steps: Immediately address target failure in the Clothing and Furniture categories to stabilize overall performance

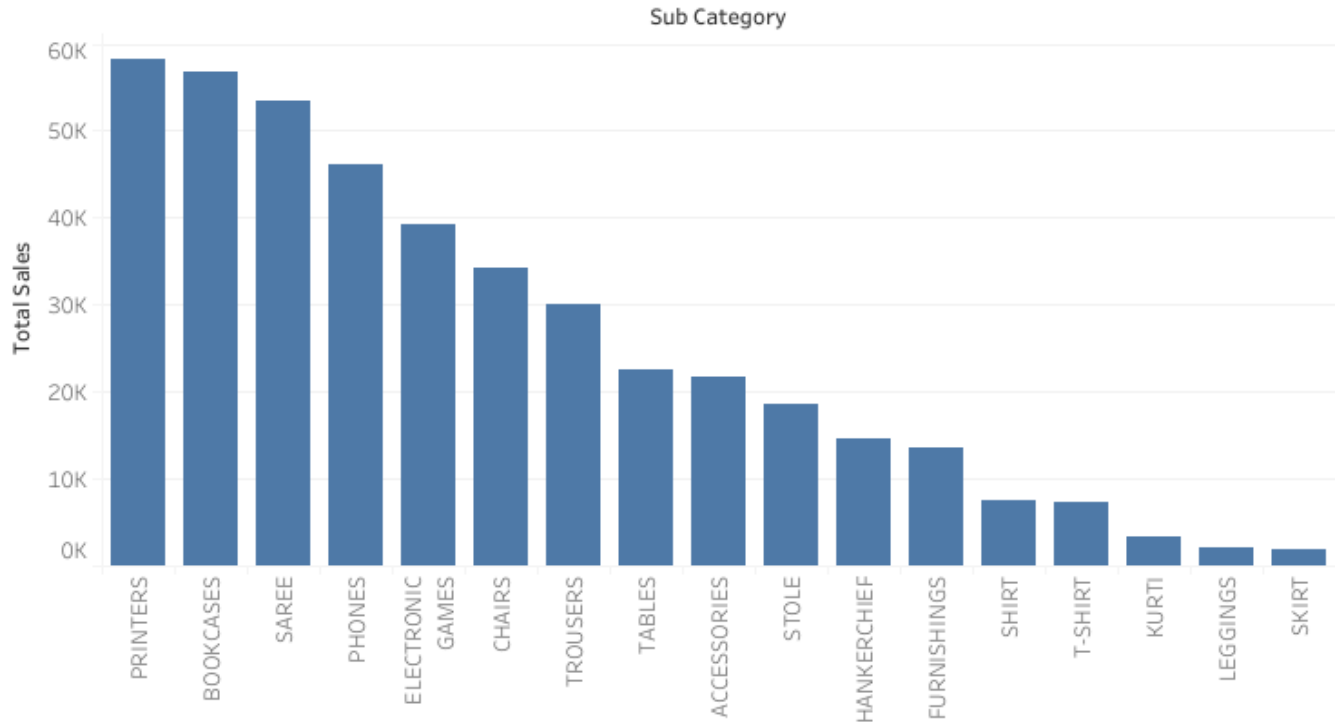
3. Sales by Sub Categories and Profit Margin

Calculate Total Sales, Profit, and Margin by Category and Sub-Category

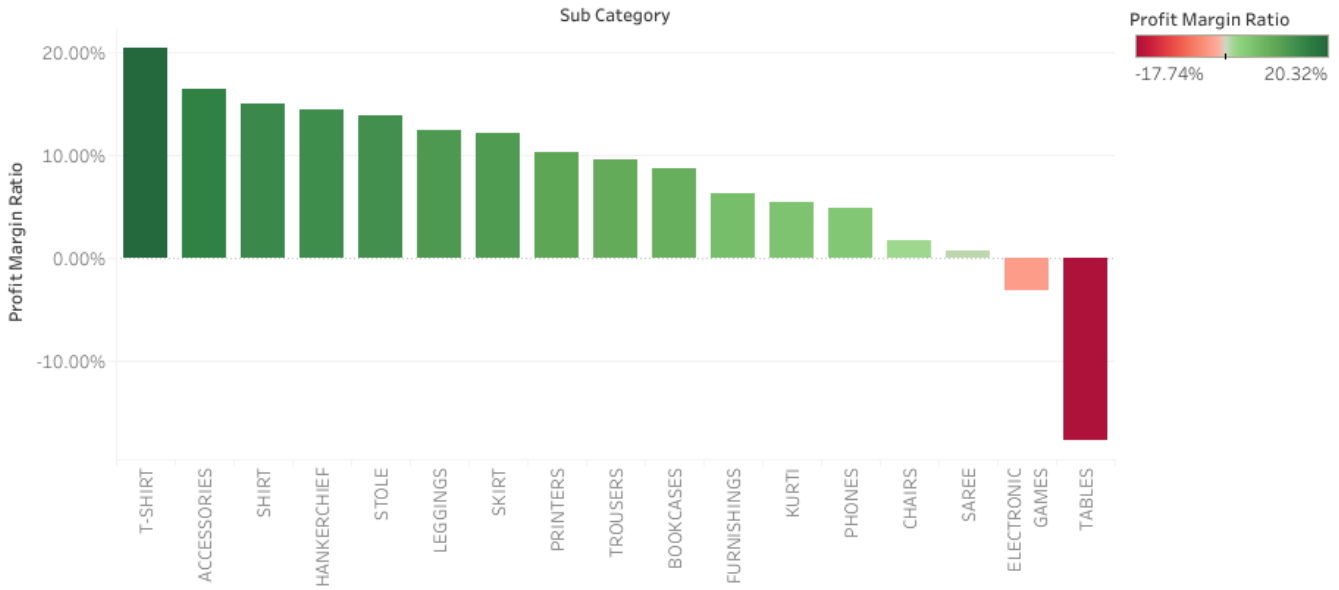
```
SELECT
  category_CLEAN,
  sub_category_CLEAN,
  SUM(Amount) AS Total_Sales,
  SUM(Profit) AS Total_Profit,
  SAFE_DIVIDE(SUM(Profit), SUM(Amount)) AS Profit_Margin_Ratio
FROM
  `helpful-kingdom-469519-f5.e_commerce_data.order_details_final`
GROUP BY
  category_CLEAN,
  sub_category_CLEAN
ORDER BY
  Total_Profit DESC,
  Total_Sales DESC;
```

Visualization

3A. Sales by Sub Category



3B. Profit Margin



Insight

- The top profit-driving products are PRINTERS ($5,964 profit) and BOOKCASES ($4,888 profit).
- The highest-margin product is the T-SHIRT (20.32% profit margin), which should be prioritized for increased sales volume due to its efficiency.

Actionable Next Steps: The most urgent issue is the loss-making status of TABLES (-17.74% margin) and ELECTRONIC GAMES (-3.16% margin), which must be immediately investigated for pricing or cost issues.

4. Top Customers

Identify Most Valuable Customers by Total Profit and Sales

```
SELECT
  t1.CustomerName,
  SUM(t2.Amount) AS Customer_Total_Sales,
  SUM(t2.Profit) AS Customer_Total_Profit,
  COUNT(t1.`Order ID`) AS Total_Orders_Placed
FROM
  `helpful-kingdom-469519-f5.e_commerce_data.list_of_orders_FINAL` AS t1
INNER JOIN
  `helpful-kingdom-469519-f5.e_commerce_data.order_details_final`
AS
  t2
ON
  t1.`Order ID` = t2.`Order ID`
GROUP BY
  t1.CustomerName
ORDER BY
  Customer_Total_Profit DESC,
  Customer_Total_Sales DESC;
```

Visualization

Insight

- SEEMA is the highest-value customer, generating the most profit (nearly $2,000), followed by ABHIJEET and PRIYANKA.
- The most frequent buyer (ABHISHEK, 25 orders) is only the third most profitable. The highest-profit customer (SEEMA) does not appear in the top 10 for order count, indicating she places fewer, higher-value orders.

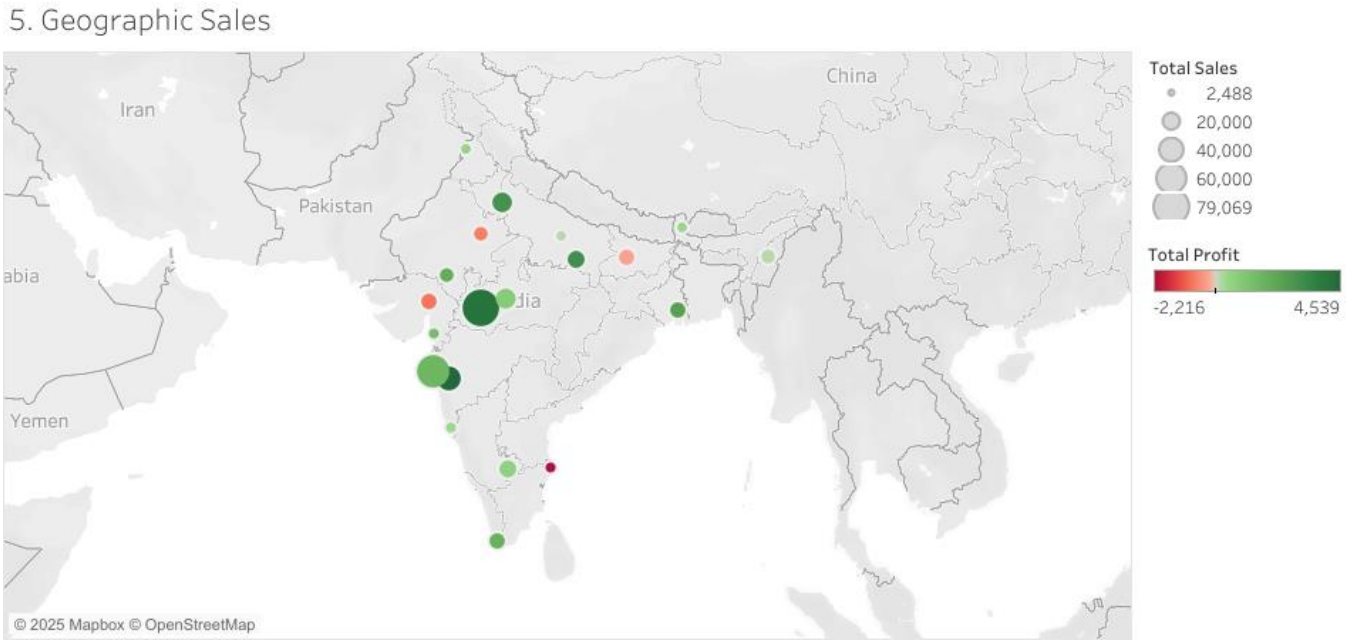
Actionable Next Steps: Prioritize retention and targeted offers for high-profit customers (SEEMA, ABHIJEET). Analyze high-frequency buyers like ABHISHEK to encourage purchases of higher-margin items.

5. Geographic Sales

Calculate Sales, Profit, and Orders by Geographic Location (City/State)

```
SELECT
  t1.State,
  t1.City,
  SUM(t2.Amount) AS Total_Sales,
  SUM(t2.Profit) AS Total_Profit,
  COUNT(t1.`Order ID`) AS Total_Orders
FROM
  `helpful-kingdom-469519-f5.e_commerce_data.list_of_order_finalwithmonth`
AS
  t1
INNER JOIN
  `helpful-kingdom-469519-f5.e_commerce_data.order_details_final`
AS
  t2
ON
  t1.`Order ID` = t2.`Order ID`
GROUP BY
  t1.State,
  t1.City
ORDER BY
  Total_Sales DESC;
```

Visualization



Insight

- INDORE (highest sales and highest profit, $4,159) and PUNE (high profit, $4,539, despite lower sales than Mumbai) are the most valuable and efficient cities.
- MUMBAI ($61,867 sales, low $1,637 profit) and BHOPAL are high-volume markets with significant efficiency issues.
- The business is actively losing money in several cities, most notably CHENNAI (-$2,216), PUNJAB/CHANDIGARH (-$1,153), and AHMEDABAD (-$880).

Actionable Next Steps: Replicate the successful strategies of Indore and Pune, and urgently address the loss-making operations in Chennai and other negative-profit cities.

Conclusion

The analysis highlighted clear sales growth with high volatility, identified top-performing products (Printers and Bookcases), and revealed loss-making categories (Tables, Electronic Games). Seema emerged as the top profit-generating customer, while Indore and Pune were the most profitable cities. These findings can help management set realistic sales targets, optimize product mix, and refine geographic strategy.

References

- Kaggle. *E-Commerce Data by benroshan*. Available at: https://www.kaggle.com/datasets/benroshan/ecommerce-data
- Google Cloud. *BigQuery Documentation*
- Tableau Public. *E-Commerce Sales Analysis Dashboard*. Available at: https://public.tableau.com/shared/T5Z2TWP9C?:display_count=n&:origin=viz_share_link