

# Case Study: Customer Behaviour Insights

## 1. Title:

**Leveraging Data to Boost Sales and Customer Retention at EcomSphere**

## 2. Introduction:

In the rapidly evolving e-commerce landscape, understanding customer purchasing behaviour and sales patterns is vital for sustained growth. EcomSphere aims to analyze its sales data to derive actionable insights that can enhance marketing strategies, optimize inventory management, and improve customer retention. This case study outlines a comprehensive analysis of e-commerce sales data, focusing on key metrics that drive business decisions.

## 3. Objective:

The goal of this project is to analyse e-commerce sales data to generate actionable insights for business decisions. Key objectives include:

- Evaluating total sales, average order value, and the number of transactions.
- Identifying customer behavior trends and product performance.
- Developing a dashboard to visualize critical sales metrics for informed decision-making.

## 4. Methodology:

The analysis employs the following methodology:

- **Data Sources:** The project utilizes publicly available e-commerce datasets, including the UCI Online Retail Dataset.
- **Tools:** Google BigQuery for data storage and SQL queries; Looker Studio for visualization.
- **Data Preparation:** Data loading, cleaning, and transformation (ETL) processes are implemented to ensure data quality.

## **5. Data Cleaning and Transformation (ETL):**

Data cleaning and transformation processes were performed to ensure data quality:

### **1. Identify Missing Values:**

```
SELECT *  
FROM `ecommerce-da.data_analysis.1`  
WHERE Description IS NULL  
AND CustomerID IS NULL;
```

QUERY:

[https://docs.google.com/spreadsheets/d/1h0Du6MhaWYqGBJVjFF5rouAfmFJAcKm2t4gC\\_JWWvE/edit?gid=501113392#gid=501113392](https://docs.google.com/spreadsheets/d/1h0Du6MhaWYqGBJVjFF5rouAfmFJAcKm2t4gC_JWWvE/edit?gid=501113392#gid=501113392)

### **2. Replace null values in certain columns:**

```
UPDATE `ecommerce-da.data_analysis.1`  
SET CustomerID = -1,  
    Description = '-1'  
WHERE CustomerID IS NULL  
AND Description IS NULL;
```

### **3. Create calculated fields:**

- Total Sales (total\_sales = quantity \* price)
- Average Order Value (AOV) per customer

```
SELECT  
    CustomerID, -- Select the CustomerID for grouping data per customer.  
    SUM(quantity * UnitPrice) AS total_sales, -- Calculates the total sales for each customer.  
    COUNT(DISTINCT InvoiceNo) AS total_orders, -- Counts the distinct number of orders (transactions) for each customer.  
    AVG(quantity * UnitPrice) AS avg_order_value -- Calculates the average value of an order (AOV) for each customer.  
FROM `ecommerce-da.data_analysis.1` -- Specifies the table where the data is located.  
GROUP BY CustomerID; -- Groups the results by CustomerID to aggregate the data per customer.
```

QUERY:

[https://docs.google.com/spreadsheets/d/1FpihJp3DazWk6MAViMa\\_ezmyDm0YQRYpP\\_m4RN1fC\\_I/edit?gid=898742244#gid=898742244](https://docs.google.com/spreadsheets/d/1FpihJp3DazWk6MAViMa_ezmyDm0YQRYpP_m4RN1fC_I/edit?gid=898742244#gid=898742244)

## 6. Questions Addressed:

Key business questions analyzed include:

1. What is the total sales revenue generated by each customer?
2. Which customers have made the most purchases?
3. What is the distribution of customers across different countries?
4. What are the peak sales periods during the day, week, or month?

## 7. Analysis & Insights: with SQL Queries

### 1. Total Sales per Product:

```
SELECT
  p.Description,
  SUM(t.quantity * t.UnitPrice) AS total_sales
FROM `ecommerce-da.data_analysis.1` t
JOIN `ecommerce-da.data_analysis.1` p
ON t.Description = p.Description
GROUP BY p.Description
ORDER BY total_sales DESC;
```

- *This analysis identifies the most popular products by revenue.*

QUERY:

[https://docs.google.com/spreadsheets/d/1\\_EAGmcrvcveJkZGSoROn2jcqwVLOzPvEne8dSCoBlf8/edit?gid=1762963505#gid=1762963505](https://docs.google.com/spreadsheets/d/1_EAGmcrvcveJkZGSoROn2jcqwVLOzPvEne8dSCoBlf8/edit?gid=1762963505#gid=1762963505)

### 2. Top Customers by Region:

```
SELECT
  c.Country,
  c.CustomerID,
  SUM(t.quantity * t.UnitPrice) AS total_spent
FROM `ecommerce-da.data_analysis.1` t
JOIN `ecommerce-da.data_analysis.1` c
ON t.CustomerID = c.CustomerID
GROUP BY c.Country, c.CustomerID
ORDER BY total_spent DESC;
```

- *This query reveals total spending by each customer, grouped by country, highlighting high-value customers and geographical spending trends.*

QUERY:

<https://docs.google.com/spreadsheets/d/1Qt8yILs3TKSTKupFMxMgA7Pelupe7ZApfCew7BJ0jA/edit?gid=1100669314#gid=1100669314>

### 3. Daily Sales Trend:

SELECT

```
InvoiceDate,  
SUM(quantity * UnitPrice) AS daily_sales  
FROM `ecommerce-da.data_analysis.1`  
GROUP BY InvoiceDate  
ORDER BY InvoiceDate;
```

- This query provides insights into sales trends over time, enabling better planning of promotions.

QUERY:

<https://docs.google.com/spreadsheets/d/1bn2vr0jEN8HRUYhjiZFUiHoquP3KGZ5xZl4tPPs7sc/edit?gid=828177003#gid=828177003>

### 8. Visualization with Looker Studio:

1. Connect Looker Studio to BigQuery or export your query results to CSV for visualization.
2. Create Visualizations using the insights derived from the analysis.

#### VISUALIZATION:

<https://lookerstudio.google.com/u/0/reporting/21a16dad-3bf3-4cfc-ba29d219f896b136/page/tEnnC>

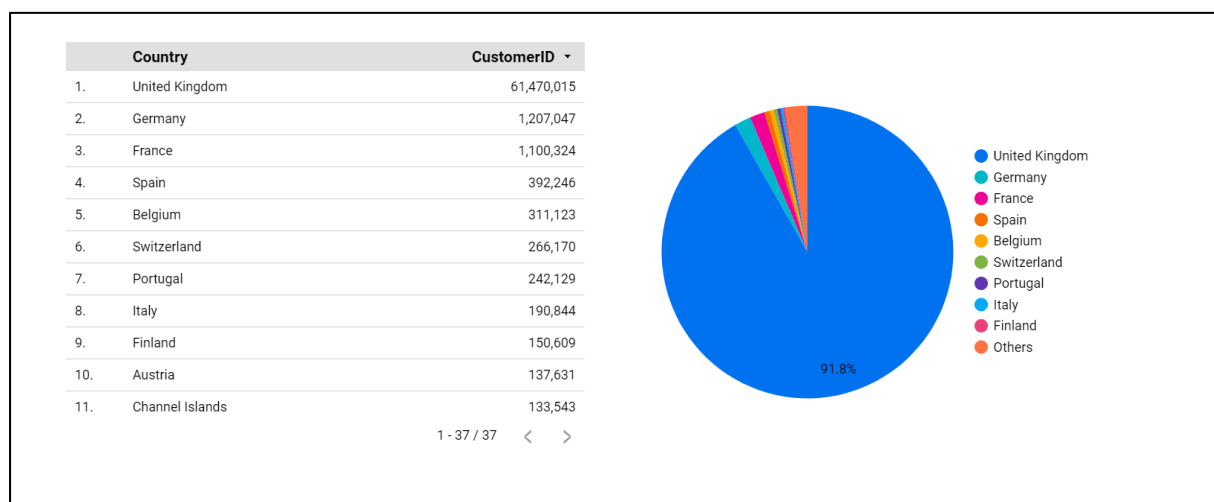


Figure 1: Showing the data of top customer by region.

## 9. Conclusion:

### Summary of Customer Distribution by Country:

- The United Kingdom dominates the customer base, with a staggering **61,470,015** customers, accounting for **91.8%** of all customers.
- The next largest customer bases are in Germany (**1,207,047**), France (**1,100,324**), and Spain (**392,246**). However, each of these countries has a much smaller share compared to the UK.
- Smaller but notable customer bases are present in Belgium, Switzerland, Portugal, Italy, Finland, and Austria, with customer counts ranging from around **137,000** to **311,000**.
- The "Others" category, which likely includes smaller countries or unlisted regions, totals **1,659,058** customers.

## 10. Final Thoughts:

The data indicates a highly UK-centric customer base, with the United Kingdom accounting for nearly 92% of the total customers. This suggests that the business is primarily focused or has most of its market presence in the UK. Expanding marketing efforts and customer acquisition in other countries—especially Germany, France, and Spain—could help diversify the customer base and reduce dependency on the UK market. Additionally, the small but meaningful presence in other European countries shows potential for growth in these regions. Tailoring strategies for specific markets like Germany, France, and Spain may yield higher engagement and sales outside of the UK.

Additionally, the small but meaningful presence in other European countries shows potential for growth in these regions. Tailoring strategies for specific markets like **Germany**, **France**, and **Spain** may yield higher engagement and sales outside of the UK.