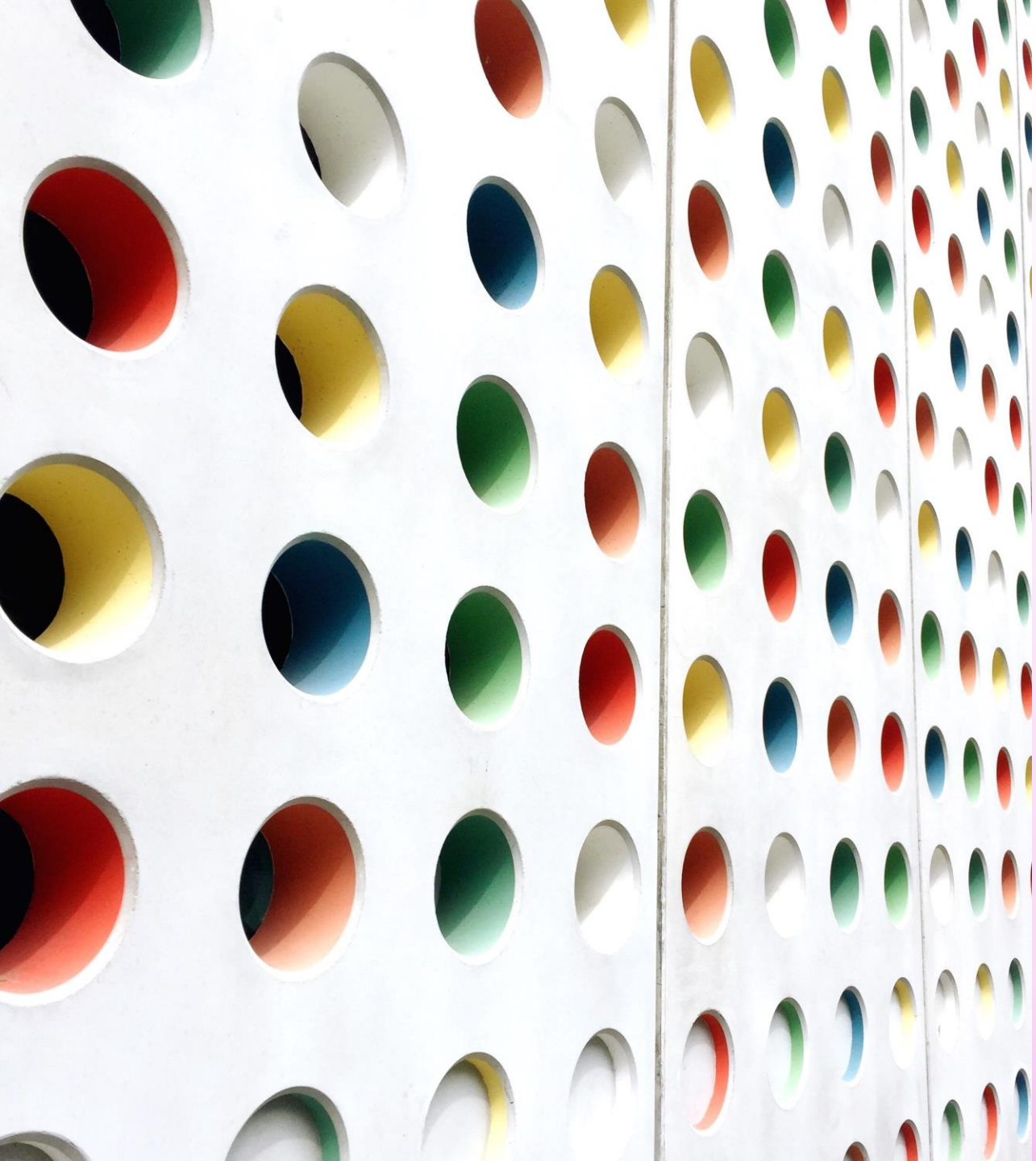


Online Retail Analysis

ANALYZING SALES DATA TO DRIVE BUSINESS INSIGHTS

Project By:
Priyanka Sahu
PGA 38

Objective

The objective of this project is to analyze an Online Retail dataset using SQL for data extraction and Power BI for visualization. The project aims to uncover key insights related to sales performance, customer behavior, and geographical demand. By leveraging SQL queries, the data is cleaned, aggregated, and prepared for analysis. Power BI is then used to create dynamic visuals that help in identifying trends, top-performing products, and potential markets for expansion. The final goal is to provide actionable insights that can guide strategic business decisions.

About the Dataset

•**Invoice No:** A unique identifier for each transaction. Each invoice number corresponds to a specific order placed by a customer.
•**Stock Code:** A unique code assigned to each product. This code is used for inventory management and product identification.
•**Description:** A brief description of the product purchased. It provides insight into the types of products being sold and their characteristics.
•**Quantity:** The number of units of each product purchased in a specific transaction. This column helps in analyzing sales volume and product demand.
•**Invoice Date:** The date and time when the transaction occurred.
•. This column is crucial for time-series analysis, trend identification, and understanding purchasing behavior over time.
•**Unit Price:** The price per unit of the product at the time of the transaction. This column allows for revenue calculations and price analysis.
•**Customer ID:** A unique identifier assigned to each customer. This column is used to track customer behavior, segment customers, and analyze repeat purchases.
•**Country:** The country where the customer is located. This information is used for geographic analysis and helps in understanding the distribution of sales across different regions.

Procedure

Step I: Importing the Dataset

1.Open Power BI:

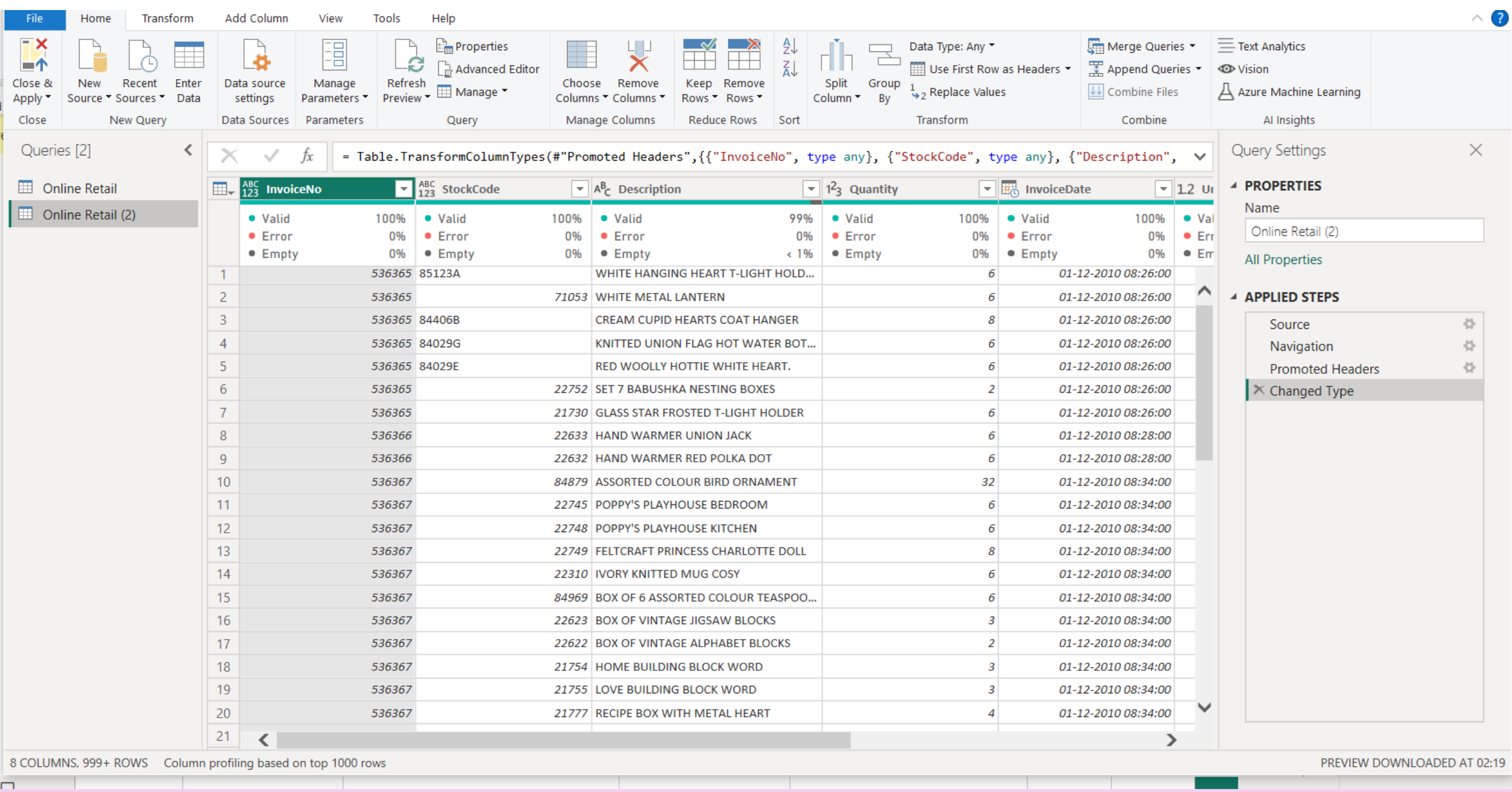
1. Launch Power BI Desktop.

2.Load Data:

1. Go to the **Home** tab and click on **Get Data**.
2. Select the data source Excel that contains your dataset.
3. Locate and select your dataset file, then click **Load.**

Step II :

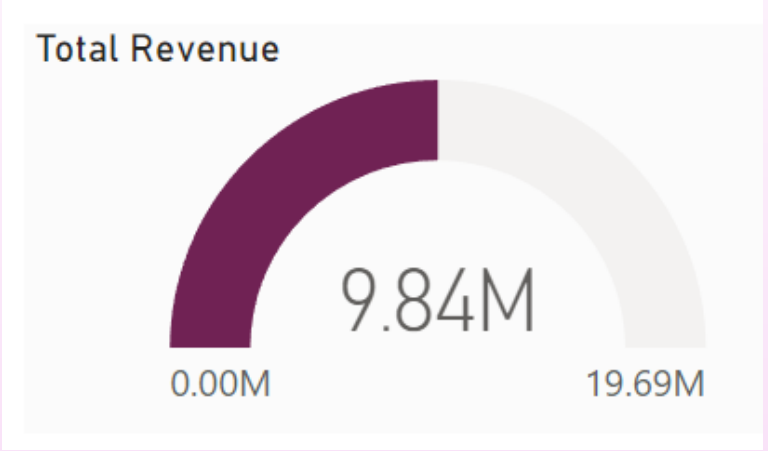
1. Review all the Columns : InvoiceNo. , StockCode , Description , Quantity , InvoiceDate , UnitePrice , CustomerID , Country.
2. Ensure that the data types are correct :
    - InvoiceDate should be **DateTime** type .
    - Quantity and UnitPrice should be **Number** types.
    - Customer ID is in both **number** and **text** (can be a Text or Number type depending on your needs).
3. Check for any missing values. If there are nulls, decide whether to fill them, drop them, or replace them with default values.


File
Home
Transform
Add Column
View
Tools
Help
Close & Apply
New Source
Recent Sources
Enter Data
Data source settings
Manage Parameters
Refresh Preview
Properties
Advanced Editor
Manage
Choose Columns
Remove Columns
Keep Rows
Remove Rows
Split Column
Group By
Data Type: Any
Use First Row as Headers
Replace Values
Merge Queries
Append Queries
Combine Files
Text Analytics
Vision
Azure Machine Learning
Close
New Query
Data Sources
Parameters
Query
Manage Columns
Reduce Rows
Sort
Transform
Combine
AI Insights
Queries [2]
Online Retail
Online Retail (2)
= Table.TransformColumnTypes(#"Promoted Headers",{{"InvoiceNo", type any}, {"StockCode", type any}, {"Description",
InvoiceNo
StockCode
Description
Quantity
InvoiceDate
Valid 100%
Error 0%
Empty 0%
Valid 100%
Error 0%
Empty 0%
Valid 99%
Error 0%
Empty < 1%
Valid 100%
Error 0%
Empty 0%
Valid 100%
Error 0%
Empty 0%
1 536365 85123A WHITE HANGING HEART T-LIGHT HOLD... 6 01-12-2010 08:26:00
2 536365 71053 WHITE METAL LANTERN 6 01-12-2010 08:26:00
3 536365 84406B CREAM CUPID HEARTS COAT HANGER 8 01-12-2010 08:26:00
4 536365 84029G KNITTED UNION FLAG HOT WATER BOT... 6 01-12-2010 08:26:00
5 536365 84029E RED WOOLLY HOTTIE WHITE HEART. 6 01-12-2010 08:26:00
6 536365 22752 SET 7 BABUSHKA NESTING BOXES 2 01-12-2010 08:26:00
7 536365 21730 GLASS STAR FROSTED T-LIGHT HOLDER 6 01-12-2010 08:26:00
8 536366 22633 HAND WARMER UNION JACK 6 01-12-2010 08:28:00
9 536366 22632 HAND WARMER RED POLKA DOT 6 01-12-2010 08:28:00
10 536367 84879 ASSORTED COLOUR BIRD ORNAMENT 32 01-12-2010 08:34:00
11 536367 22745 POPPY'S PLAYHOUSE BEDROOM 6 01-12-2010 08:34:00
12 536367 22748 POPPY'S PLAYHOUSE KITCHEN 6 01-12-2010 08:34:00
13 536367 22749 FELTCRAFT PRINCESS CHARLOTTE DOLL 8 01-12-2010 08:34:00
14 536367 22310 IVORY KNITTED MUG COSY 6 01-12-2010 08:34:00
15 536367 84969 BOX OF 6 ASSORTED COLOUR TEASPOO... 6 01-12-2010 08:34:00
16 536367 22623 BOX OF VINTAGE JIGSAW BLOCKS 3 01-12-2010 08:34:00
17 536367 22622 BOX OF VINTAGE ALPHABET BLOCKS 2 01-12-2010 08:34:00
18 536367 21754 HOME BUILDING BLOCK WORD 3 01-12-2010 08:34:00
19 536367 21755 LOVE BUILDING BLOCK WORD 3 01-12-2010 08:34:00
20 536367 21777 RECIPE BOX WITH METAL HEART 4 01-12-2010 08:34:00
21
Query Settings
PROPERTIES
Name
Online Retail (2)
All Properties
APPLIED STEPS
Source
Navigation
Promoted Headers
Changed Type
8 COLUMNS, 999+ ROWS
Column profiling based on top 1000 rows
PREVIEW DOWNLOADED AT 02:19

Visual Creation Process and Key Insights

Step III: Creating Visuals

1.Donut Chart: Total Revenue
- Insert a Donut Chart.
- Add the Revenue field to Values
- Optionally, add Country or CustomerID to legend for breakdown.



Offers a clear visual representation of total revenue is 9.84M, allowing quick assessment of overall sales performance.

2. Dropdown Filters: Country, Month, Customer ID
- Add Slicer for Country, InvoiceDate(format to Month), and CustomerID
- These filters will refine the visuals.

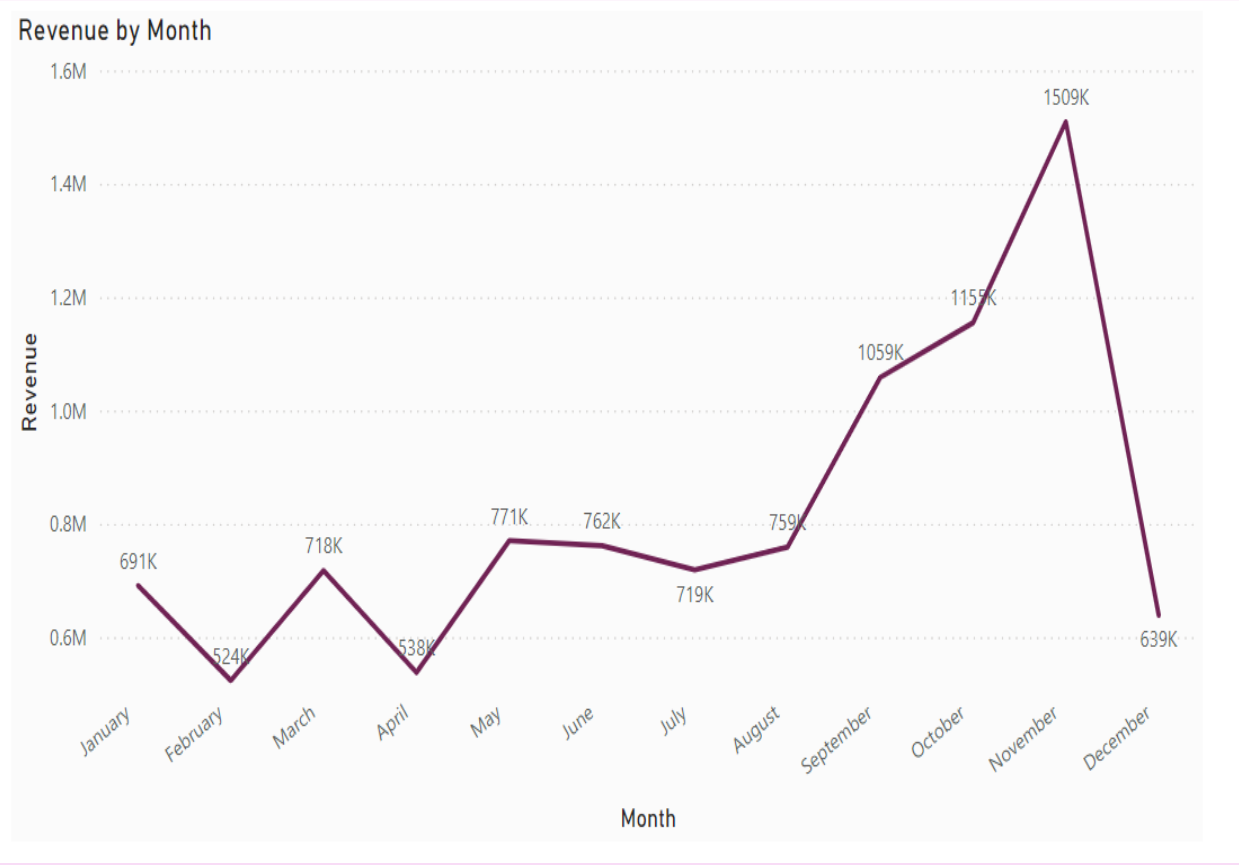
Allows dynamic filtering of data, enabling users to drill down into specific regions, periods, or customer segments. This feature helps in identifying trends and outliers within specific contexts.

Visual Creation Process and Key Insights

Step III: Creating Visuals

3. Line Chart: Revenue by Month
- Insert a Line Chart.
- Add InvoiceDate to Axis (set to Month).
- Add Revenue to Values.

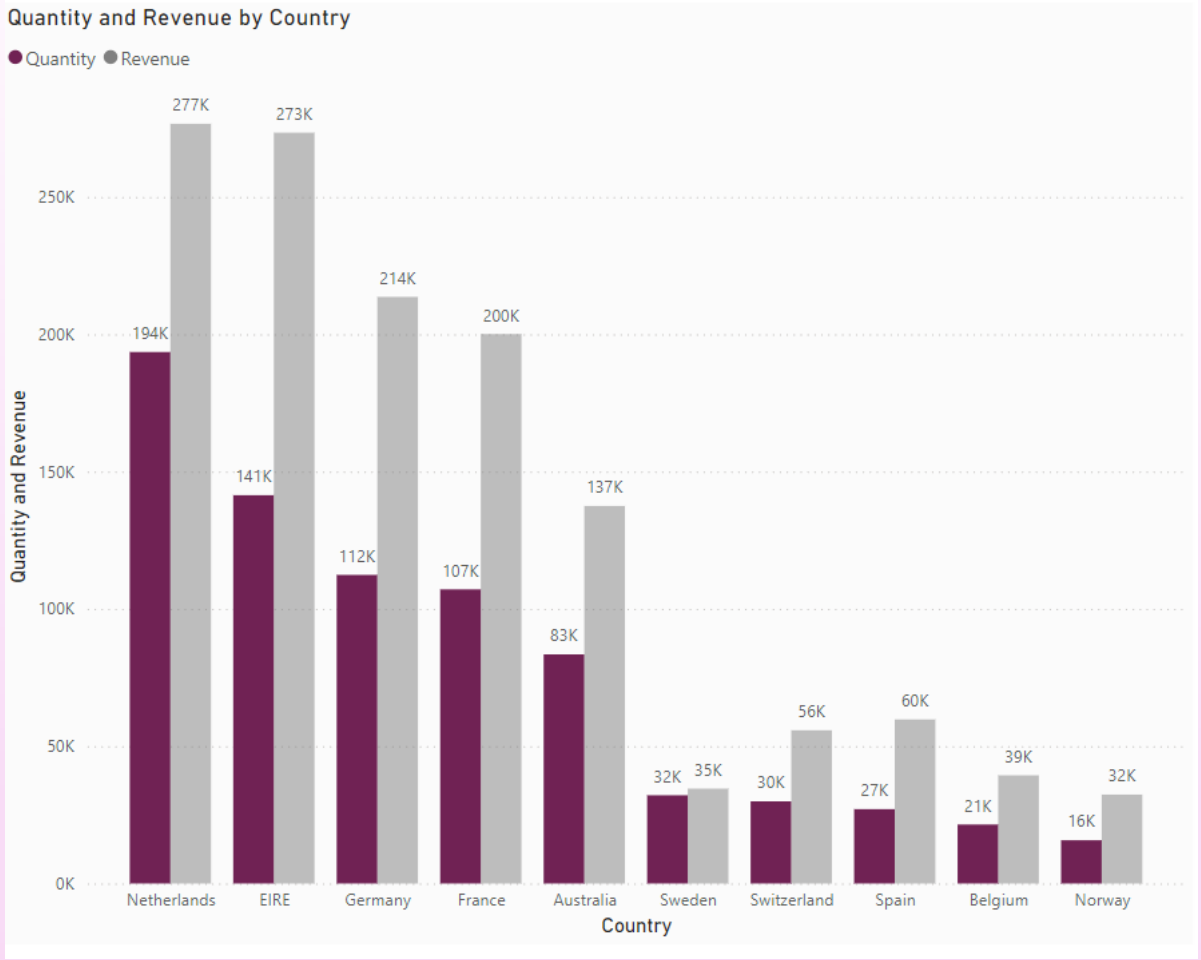
According to the data, there are several months of the year that have significant development. According to the statistics, the first eight months of sales from January to August were very stable, with an average of 759K in revenue per month. The increase in revenue begins in September, when it grows by 40% over the previous month. This pattern persisted up until November, when it rose to 1509K (thousand), the largest amount of the whole year. Unfortunately, since the data for December is insufficient, no inferences can be made from it



Visual Creation Process and Key Insights

4. Bar Chart: Quantity and Revenue by Country
- Insert a Clustered Bar Chart.
- Add Country to Axis
- Add Revenue and Quantity to Values
- Consider dual-axis for separate value display.

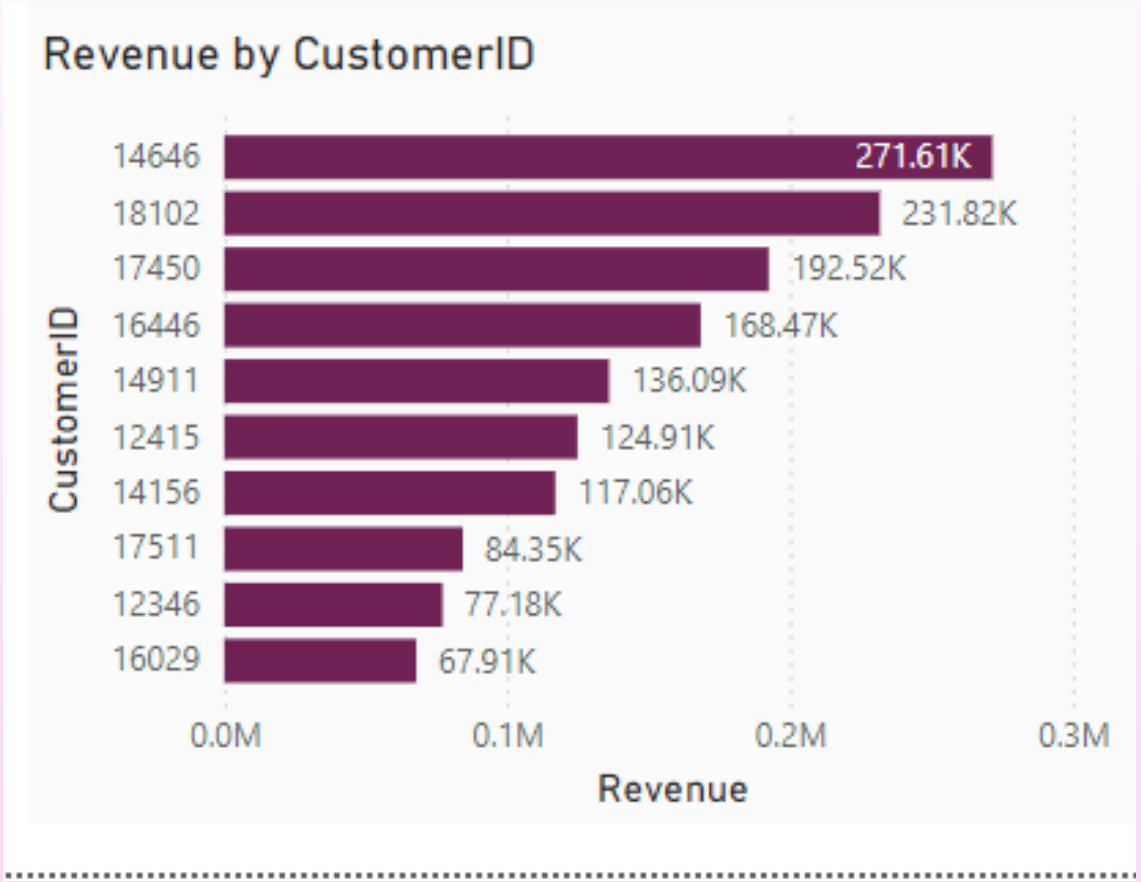
The top 10 countries with the most potential for growth are represented in the second graph. Since the UK already has a large demand and I understand you are more interested in nations where demand may be boosted, the UK is not included in these statistics. According to the data, sales of units and income are quite high in nations like the Netherlands, Ireland, Germany, and France. To guarantee that steps are taken to further seize these markets, I would propose concentrating on these nations.



Visual Creation Process and Key Insights

5. Bar Chart: Revenue by Customer ID
- Insert a Clustered Bar Chart.
- Add CustomerID to Axis
- Add Revenue to Values

The top 10 consumers who have made the most purchases from the business have been the subject of the third study. According to the statistics, there are not many differences between the top 10 consumer purchases. The fact that the highest revenue-producing consumer only spent 17% more than the second highest demonstrates that the company does not rely solely on a small number of consumers to generate income. This demonstrates that consumers' ability to negotiate is limited and that the state of business is positive.



Visual Creation Process and Key Insights

6. Map: Quantity by Country
- Insert a Map.
- Add Country to Location
- Add Quantity to Size.

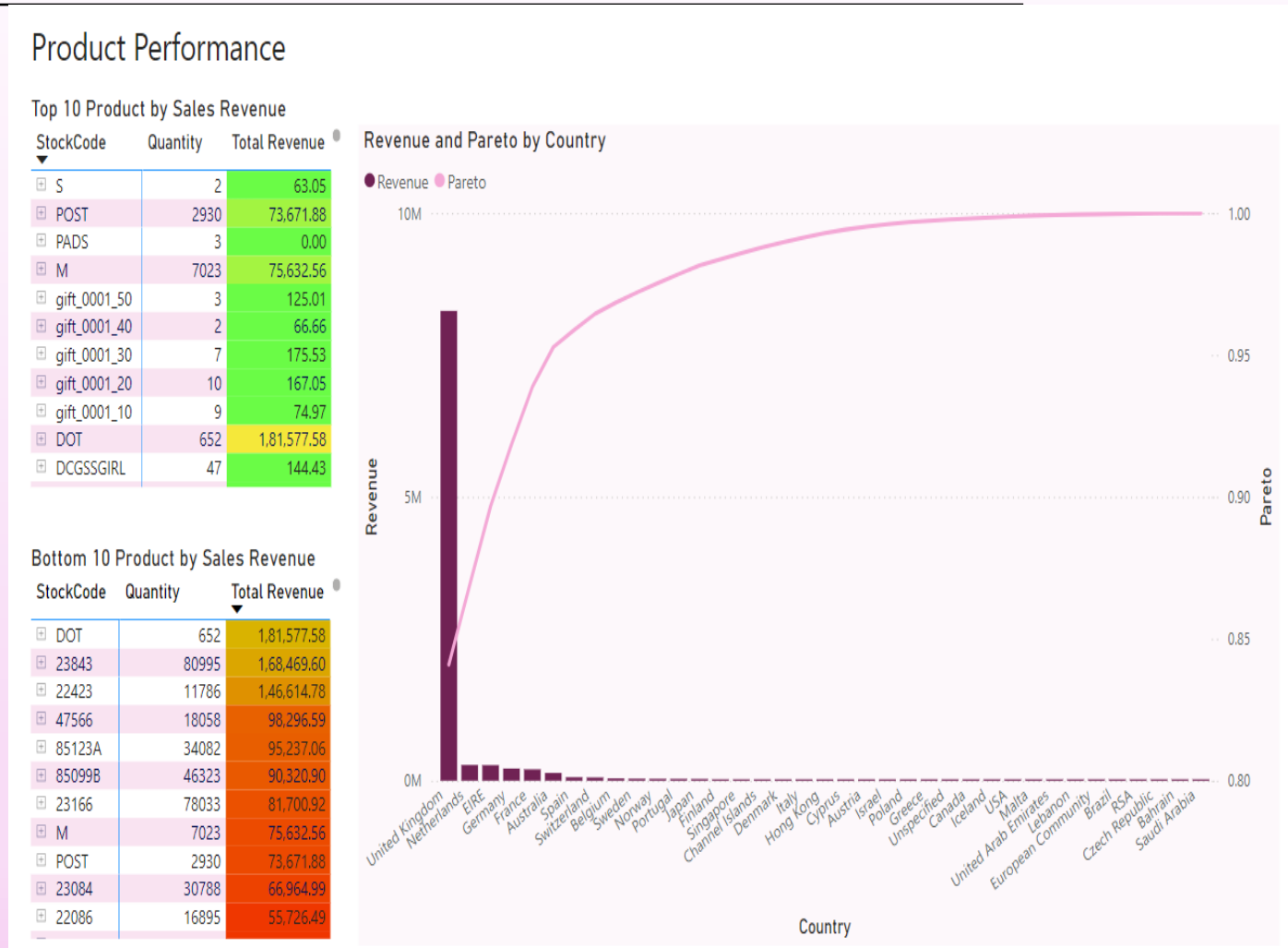
The map chart concludes by comparing the places that have produced the greatest revenue to those that have not. Apart from the UK, it is clear that nations like the Netherlands, Ireland, Germany, France, and Australia generate large profits, and the company should invest more in these nations to boost product demand. The map also reveals that the majority of sales occur only in the European zone, with only a small number in the American region. Along with Russia, there is no market for the items in Africa or Asia. Sales revenues and profitability might increase with the implementation of a fresh strategy focused on these areas.

Visual Creation Process and Key Insights

7. Table: Top 10 Products by Sales Revenue
- Insert a Table.
- Add Description and Revenue to Values
- Sort by Revenue in descending order.
- Apply Top N filter for top 10 products.

8. Line and Bar Chart: Total Revenue & Revenue % by Country
- Insert a Line and Clustered Column Chart.
- Add Country to Shared Axis
- Add Revenue to Column Values.
- Add Revenue Percentage to Line Values (calculated DAX measure).

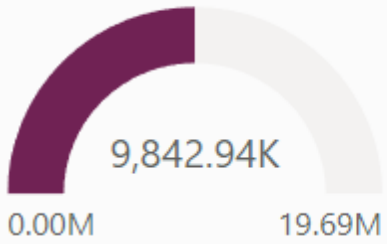
9. Scatter Chart : Product Wise Sales Revenue vs Quantity
- Insert Scatter Chart.
- Add Quantity to X-axis.
- Add Total Revenue to Y-axis.
- Set legend Stock Code.



Online Retail Dashboard

Month: All | Country: All | CustomerID: All

Total Revenue



4220
TotalCustomers

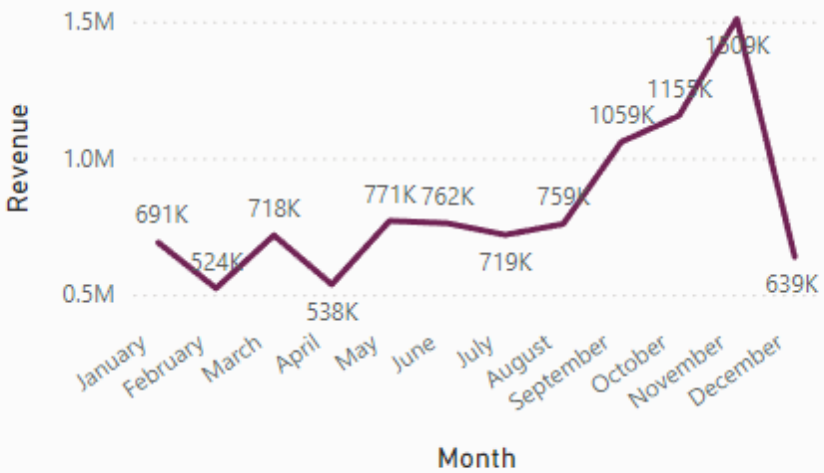
37
Total Country

11.1K
Highest UnitPrice

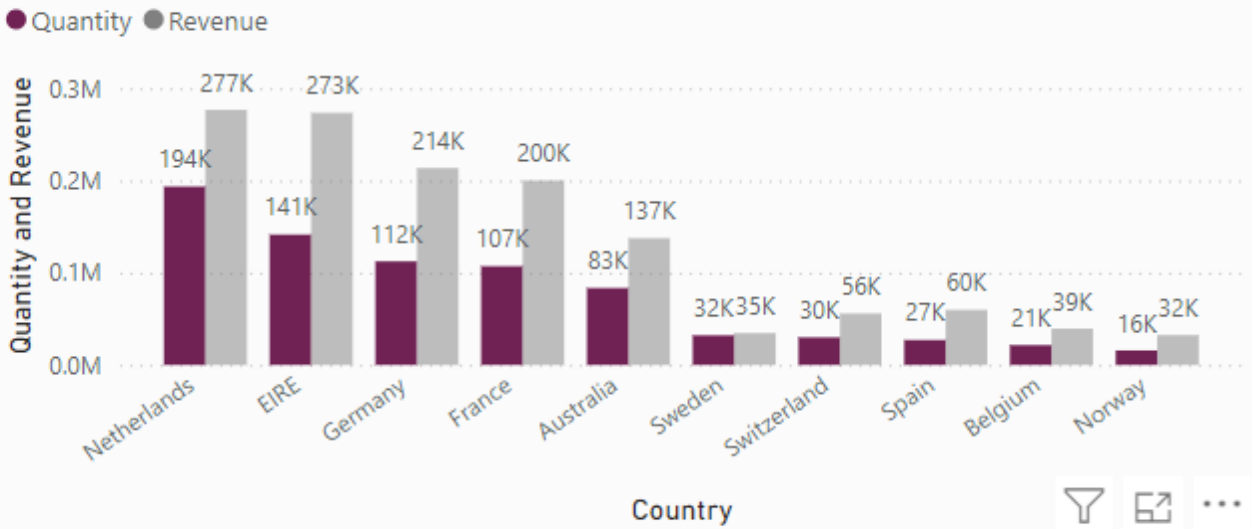
18.4K
Total Invoice

3730
Total StockCode

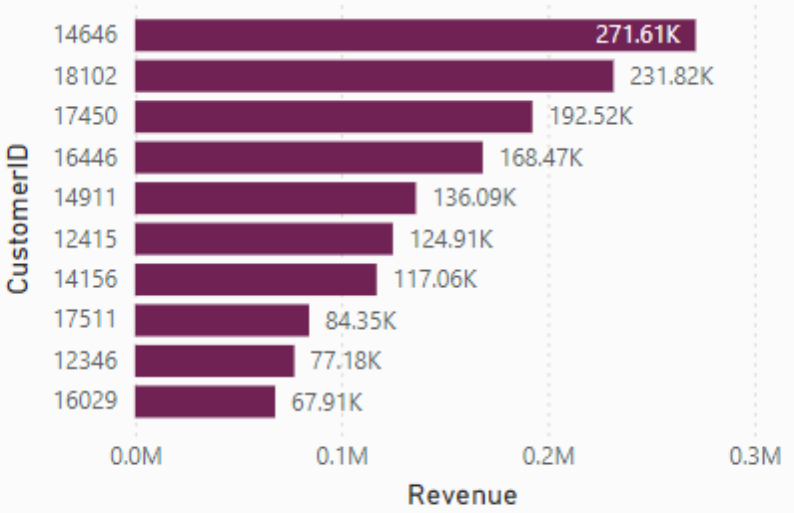
Revenue by Month



Quantity and Revenue by Country

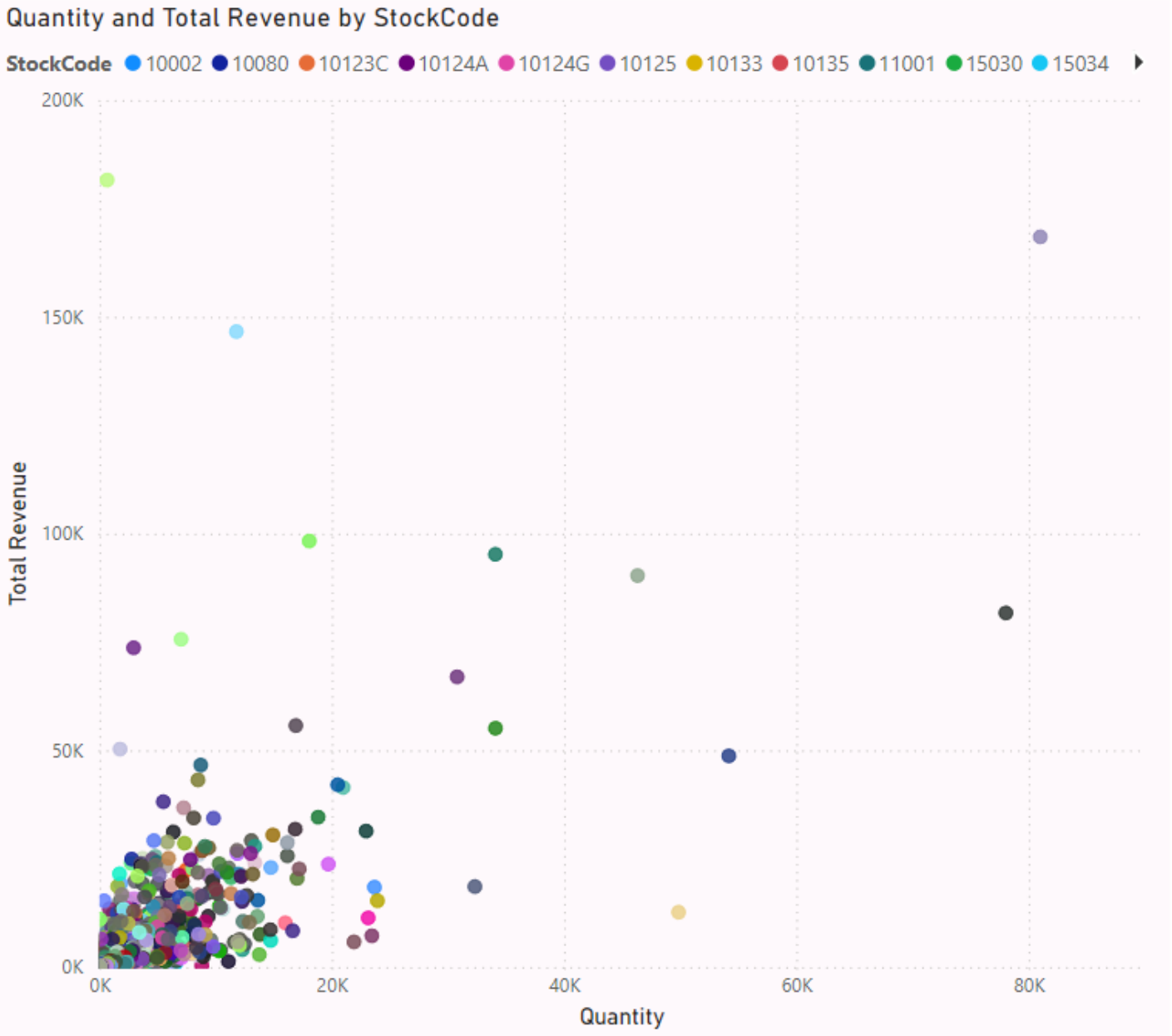


Revenue by CustomerID



Quantity by Country

Product Wise Sales Revenue vs Quantity



StockCode	Description	Quantity	Total Revenue
10002	INFLATABLE POLITICAL GLOBE	609	525.48
10080	GROOVY CACTUS INFLATABLE	303	119.09
10120	DOGGY RUBBER	177	37.17
10123C	HEARTS WRAPPING TAPE	4	2.60
10124A	SPOTS ON RED BOOKCOVER TAPE	12	5.04
10124G	ARMY CAMO BOOKCOVER TAPE	12	5.04
10125	MINI FUNKY DESIGN TAPES	1142	903.98
10133	COLOURING PENCILS BROWN TUBE	2737	1,458.95
10135	COLOURING PENCILS BROWN TUBE	1820	1,928.37
11001	ASSTD DESIGN RACING CAR PEN	1541	2,251.02
15030	FAN BLACK FRAME	143	41.47
15034	PAPER POCKET TRAVELING FAN	6546	828.75
15036	ASSORTED COLOURS SILK FAN	23665	18,433.72
15039	SANDALWOOD FAN	2071	1,960.06
15044A	PINK PAPER PARASOL	458	1,439.02
15044B	BLUE PAPER PARASOL	318	935.02
15044C	PURPLE PAPER PARASOL	304	987.56
15044D	RED PAPER PARASOL	644	1,816.80
15056BL	EDWARDIAN PARASOL BLACK	2726	15,591.34
15056N	EDWARDIAN PARASOL NATURAL	3929	22,803.48
15056P	EDWARDIAN PARASOL PINK	733	4,332.61
15058A	BLUE POLKADOT GARDEN PARASOL	199	1,660.54
15058B	PINK POLKADOT GARDEN PARASOL	227	1,877.32
15058C	ICE CREAM DESIGN GARDEN PARASOL	240	1,429.82
15060B	FAIRY CAKE DESIGN UMBRELLA	398	1,575.22
16008	SMALL FOLDING SCISSOR(POINTED EDGE)	2878	357.21
16010	FOLDING CAMPING SCISSOR W/KNIF & S	18	2.16
16011	ANIMAL STICKERS	1679	352.59
Total		**5229137**	**98,42,938.40**

Overall Dashboard Insights

•Strategic Decision-Making: The dashboard provides a comprehensive view of the business's performance, helping leadership to make data-driven decisions regarding marketing, sales strategies, and regional expansion

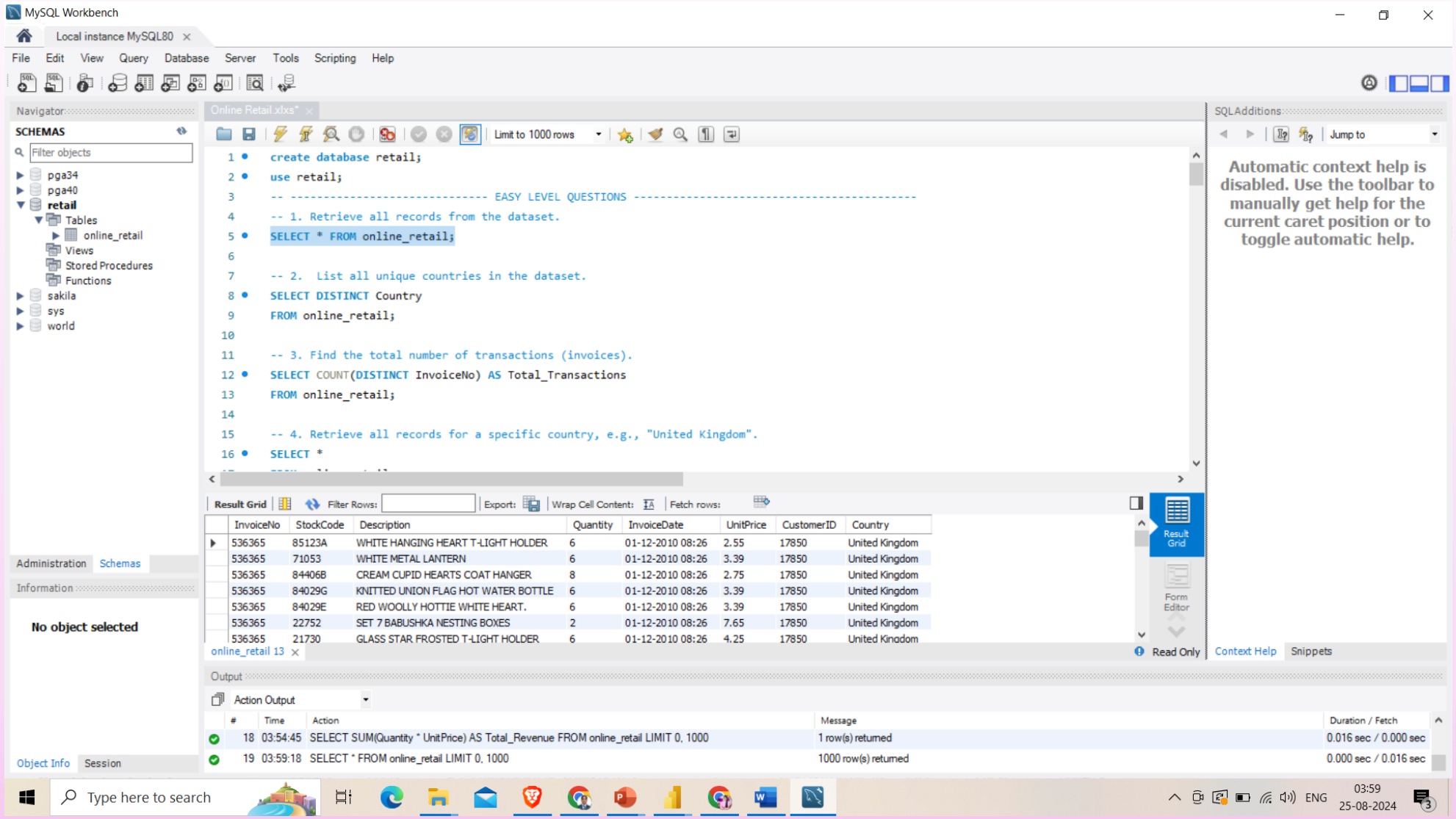
•Market Opportunities: By analyzing revenue and quantity across countries, the company can identify underperforming regions that may benefit from targeted marketing efforts, or high-performing regions that could support expanded operations.

•Customer Targeting: Insights into the top revenue-generating customers allow the business to focus on customer retention strategies, ensuring long-term loyalty and higher customer lifetime value.

•Product Performance: The data on top-selling products can inform inventory management, pricing strategies, and promotional campaigns, ensuring that the most popular products are always available and competitively priced.

SQL ANALYSICS

Import the Online Retail dataset into the MySQL server. After importing, perform data cleansing and analysis to ensure the data is accurate and ready for further insights and decision making.



SQL Questions

1. List all unique countries in the dataset.
2. Find all records where the quantity is greater than 10.
3. List the top 5 countries by the number of transactions.
4. Retrieve the most frequently purchased item (StockCode).
5. Find the top 3 customers by total purchase amount.
6. Find all invoices where more than 10 unique items (StockCodes) were purchased.
7. Find the total number of items sold per country.
8. Retrieve the top 5 products with stockcode by revenue generated.
9. Find the average quantity of items purchased per transaction for each customer.
10. Determine the top 3 products that contributed to the highest revenue in 'United Kingdom’.
11. Identify the customers who purchased more than 100 items in a single transaction.
12. Find the average revenue per customer and compare it between two specific countries, e.g., 'Germany' and 'France'.

```sql
-- 1.  List all unique countries in the dataset.
SELECT DISTINCT Country
FROM online_retail;
```

Country
United Kingdom
France
Australia
Netherlands
Germany
Norway
EIRE

online_retail 35

```sql
-- 2. Find all records where the quantity is greater than 10.
SELECT *
FROM online_retail
WHERE Quantity > 10;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
536367	84879	ASSORTED COLOUR BIRD ORNAMENT	32	01-12-2010 08:34	1.69	13047	United Kingdom
536370	22728	ALARM CLOCK BAKELIKE PINK	24	01-12-2010 08:45	3.75	12583	France
536370	22727	ALARM CLOCK BAKELIKE RED	24	01-12-2010 08:45	3.75	12583	France
536370	22726	ALARM CLOCK BAKELIKE GREEN	12	01-12-2010 08:45	3.75	12583	France
536370	21724	PANDA AND BUNNIES STICKER SHEET	12	01-12-2010 08:45	0.85	12583	France
536370	21883	STARS GIFT TAPE	24	01-12-2010 08:45	0.65	12583	France
536370	10002	INFLATABLE POLITICAL GLOBE	48	01-12-2010 08:45	0.85	12583	France
536370	21791	VINTAGE HEADS AND TAILS CARD GAME	24	01-12-2010 08:45	1.25	12583	France
536370	21035	SET/2 RED RETROSPOT TEA TOWELS	18	01-12-2010 08:45	2.95	12583	France
536370	22326	ROUND SNACK BOXES SET OF4 WOODLAND	24	01-12-2010 08:45	2.95	12583	France

online_retail 36

```sql
-- 3. List the top 5 countries by the number of transactions.
SELECT Country, COUNT(DISTINCT InvoiceNo) AS Transactions
FROM online_retail
GROUP BY Country
ORDER BY Transactions DESC
LIMIT 5;
```

Country	Transactions
United Kingdom	702
Germany	15
France	11
EIRE	9
Lithuania	4

```sql
-- 4 Retrieve the most frequently purchased item (StockCode).
SELECT StockCode, COUNT(*) AS Purchase_Count
FROM online_retail
GROUP BY StockCode
ORDER BY Purchase_Count DESC
LIMIT 10;
```

StockCode	Purchase_Count
85123A	104
22867	91
22086	87
22633	86
22632	77
22866	77
84029E	74

Result 38

```sql
-- 5.ind the top 3 customers by total purchase amount.
SELECT CustomerID, SUM(Quantity * UnitPrice) AS Total_Purchase
FROM online_retail
GROUP BY CustomerID
ORDER BY Total_Purchase DESC
LIMIT 3;
```

Result Grid | Filter Rows:

CustomerID	Total_Purchase
18102	27834.61
15061	11429.499999999996
13777	6840.16

```sql
-- 6.Find all invoices where more than 10 unique items (StockCodes) were purchased.
SELECT InvoiceNo
FROM online_retail
GROUP BY InvoiceNo
HAVING COUNT(DISTINCT StockCode) > 10;
```

InvoiceNo
536367
536370
536373
536375
536378
536381
536382

online_retail 41

```sql
-- 7.Find the total number of items sold per country.
SELECT Country, SUM(Quantity) AS Total_Items_Sold
FROM online_retail
GROUP BY Country
ORDER BY Total_Items_Sold DESC;
```

Country	Total_Items_Sold
United Kingdom	138060
Norway	3582
EIRE	3525
France	3189
Germany	2616
Japan	1684
Portugal	660

Result 42

```sql
-- 8.Retrieve the top 5 products with stockcode by revenue generated.
SELECT Description, Stockcode, SUM(Quantity * UnitPrice) AS Revenue
FROM online_retail
GROUP BY Description, stockcode
ORDER BY Revenue DESC
LIMIT 5;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Conten

Description	Stockcode	Revenue
REGENCY CAKESTAND 3 TIER	22423	8234.099999999999
VINTAGE UNION JACK MEMOBOARD	21623	6639.989999999999
WOOD BLACK BOARD ANT WHITE FINISH	82484	5898.630000000001
CREAM HEART CARD HOLDER	22189	5165.6900000000005
BLACK HEART CARD HOLDER	22188	5051.14

```sql
-- 9.Find the average quantity of items purchased per transaction for each customer.
SELECT CustomerID, AVG(Quantity) AS Avg_Quantity
FROM online_retail
GROUP BY CustomerID;
```

Result Grid | Filter Rows:

CustomerID	Avg_Quantity
17850	5.8350
13047	5.9412
12583	22.4500
13748	80.0000
15100	32.0000
15291	56.0000
14688	23.8947

```sql
-- 10. Determine the top 3 products that contributed to the highest revenue in 'United Kingdom'.
SELECT StockCode, SUM(Quantity * UnitPrice) AS Revenue
FROM online_retail
WHERE Country = 'United Kingdom'
GROUP BY StockCode
ORDER BY Revenue DESC
LIMIT 3;
```

StockCode	Revenue
22423	7778.399999999998
21623	6639.989999999999
82484	5898.630000000001

```sql
-- 11. identify the customers who purchased more than 100 items in a single transaction
SELECT CustomerID, InvoiceNo, SUM(Quantity) AS Total_Items
FROM online_retail
GROUP BY CustomerID, InvoiceNo
HAVING Total_Items > 100;
```

CustomerID	InvoiceNo	Total_Items
12583	536370	449
15291	536376	112
14688	536378	454
15311	536381	198
16098	536382	134
18074	536384	190
16029	536386	236

```sql
-- 12. Find the average revenue per customer and compare it between two specific countries, e.g., 'Germany' and 'France'.
SELECT Country, AVG(Revenue) AS Avg_Revenue_Per_Customer
FROM (
    SELECT Country, CustomerID, SUM(Quantity * UnitPrice) AS Revenue
    FROM online_retail
    WHERE Country IN ('Germany', 'France')
    GROUP BY Country, CustomerID
) AS RevenueData
GROUP BY Country;
```

Result Grid | Filter Rows:

Country	Avg_Revenue_Per_Customer
France	786.3025000000001
Germany	473.6733333333334

Conclusion

The project successfully demonstrated the process of importing, cleansing, and analyzing an Online Retail dataset using SQL and visualizing key insights through Power BI. The analysis provided valuable information on sales trends, customer behavior, and geographic demand, which can be leveraged for strategic decision-making. By utilizing SQL for data preparation and Power BI for dynamic visualizations, the project offers actionable insights that can guide business strategies, improve customer satisfaction, and identify potential areas for market expansion.

THANK YOU

Priyanka Sahu
priyankasahu6226@gmail.com
PGA 38