Online Retail Data Case Study: Customer, Sales, and Expansion Strategy

During my case study internship with Tata, I analyzed an online retail dataset to better understand customer behavior, sales trends, and product performance. The goal was to generate actionable insights that could help business leaders identify growth opportunities, strengthen customer relationships, and optimize expansion strategies.

Dataset Overview

Rows: 541,910 transactions

Columns:

- InvoiceNo → Unique identifier for each invoice/order
- StockCode → Product code
- Description → Product description
- Quantity → Number of items purchased
- InvoiceDate → Date and time of transaction
- UnitPrice → Price per item
- CustomerID → Unique identifier for customer (anonymized)
- Country → Customer's country of residence

Derived Metric:

- Total amount = Quantity × UnitPrice

	A	B	C	D	E	F	G	H	I	J	K
1	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	Month	time	UnitPrice	CustomerID	Country	Total amount
2	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	01-12-2010	12-2010	08:26:00	2.55	17850	United Kingdom	15.3
3	536365	71053	WHITE METAL LANTERN	6	01-12-2010	12-2010	08:26:00	3.39	17850	United Kingdom	20.34
4	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	01-12-2010	12-2010	08:26:00	2.75	17850	United Kingdom	22
5	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	01-12-2010	12-2010	08:26:00	3.39	17850	United Kingdom	20.34
6	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	01-12-2010	12-2010	08:26:00	3.39	17850	United Kingdom	20.34
7	536365	22752	SET 7 BABUSHKA NESTING BOXES	2	01-12-2010	12-2010	08:26:00	7.65	17850	United Kingdom	15.3
8	536365	21730	GLASS STAR FROSTED T-LIGHT HOLDER	6	01-12-2010	12-2010	08:26:00	4.25	17850	United Kingdom	25.5
9	536366	22633	HAND WARMER UNION JACK	6	01-12-2010	12-2010	08:28:00	1.85	17850	United Kingdom	11.1
10	536366	22632	HAND WARMER RED POLKA DOT	6	01-12-2010	12-2010	08:28:00	1.85	17850	United Kingdom	11.1
11	536367	84879	ASSORTED COLOUR BIRD ORNAMENT	32	01-12-2010	12-2010	08:34:00	1.69	13047	United Kingdom	54.08
12	536367	22745	POPPY'S PLAYHOUSE BEDROOM	6	01-12-2010	12-2010	08:34:00	2.1	13047	United Kingdom	12.6
13	536367	22748	POPPY'S PLAYHOUSE KITCHEN	6	01-12-2010	12-2010	08:34:00	2.1	13047	United Kingdom	12.6
14	536367	22749	FELTCRAFT PRINCESS CHARLOTTE DOLL	8	01-12-2010	12-2010	08:34:00	3.75	13047	United Kingdom	30
15	536367	22310	IVORY KNITTED MUG COSY	6	01-12-2010	12-2010	08:34:00	1.65	13047	United Kingdom	9.9

The first step I took after receiving the dataset was to create a copy and name it **"Working Sheet."** This ensured that the original dataset remained untouched, providing a backup in case any errors occurred during the tasks.

Next, I validated the dataset to confirm that the data doesn't have Incorrect values, Duplicates records, Empty rows, Columns, or blocks

TASK 1: Framing questions that would be most relevant for the CEO and CMO

Before diving into the datasets, I first focused on understanding what matters most to the CEO and CMO.

CEO (Chief Executive Officer) cares about:

- Revenue trends
- Profitability
- Operational efficiency
- Market expansion opportunities
- Top-performing products/countries

CMO (Chief Marketing Officer) cares about:

- Customer demographics
- Customer retention
- Purchase behavior
- Marketing ROI
- Popular product categories and timing

Next, I utilized the working sheet to build three pivot tables for deeper analysis:

1. Top 5 performing countries along with their Top 6 products
2. Bottom 5 performing countries along with their Top 6 products
3. Monthly sale

	Top 5 performing Countries					Bottom 5 performing Countries					Monthwise sale	
Sr no.	Stock code (United kingdom)	Description	Total Sale		Sr no.	Stock code (saudi)	Description	Total sale		Sr no.	Months	Sale
1	DOT	DOTCOM POSTAGE	206245		1	22555	PLASTERS IN TIN CIRCUS PARADE	20		1	11-2011	1461756
2	22423	REGENCY CAKESTAND 3 TIER	134406		2	22556	PLASTERS IN TIN STRONGMAN	20		2	10-2011	1070705
3	85123A	WHITE HANGING HEART T-LIGHT HOLDER	94132		3	22553	PLASTERS IN TIN SKULLS	20		3	9-2011	1019688
4	47566	PARTY BUNTING	92502		4	22362	GLASS JAR MARMALADE	18		4	12-2010	748957
5	85099B	JUMBO BAG RED RETROSPOT	84516		5	22363	GLASS JAR PEACOCK BATH SALTS	18		5	5-2011	723334
6	22086	PAPER CHAIN KIT 50'S CHRISTMAS	61888		6	22969	HOMEMADE JAM SCENTED CANDLES	17		6	6-2011	691123
Sr no.	Stock code (Netherland)	Description	Total sale		Sr no.	Stock code (Bahrain)	Description	Total sale		7	3-2011	683267
1	23084	RABBIT NIGHT LIGHT	9568		1	23076	ICE CREAM SUNDAE LIP GLOSS	120		8	8-2011	682681
2	22326	ROUND SNACK BOXES SET OF4 WOODLAND	7991		2	23077	DOUGHNUT LIP GLOSS	75		9	7-2011	681300
3	22629	SPACEBOY LUNCH BOX	7486		3	22890	NOVELTY BISCUITS CAKE STAND 3 TIER	60		10	1-2011	560000
4	22630	DOLLY GIRL LUNCH BOX	6829		4	22649	STRAWBERRY FAIRY CAKE TEAPOT	40		11	2-2011	498063
5	22328	ROUND SNACK BOXES SET OF 4 FRUITS	4039		5	22693	GROW A FLYTRAP OR SUNFLOWER IN TIN	30		12	4-2011	493207
6	21731	RED TOADSTOOL LED NIGHT LIGHT	3479		6	72802A	VANILLA SCENT CANDLE JEWELLED BOX	26		13	12-2011	433668
Sr no.	Stock Code (EiRE)	Description	Total sale		Sr no.	Stock code (Czech)	Description	Total sale				
1	22423	REGENCY CAKESTAND 3 TIER	7443		1	22326	ROUND SNACK BOXES SET OF4 WOODLAND	71				
2	C2	CARRIAGE	5175		2	84347	ROTATING SILVER ANGELS T-LIGHT HLDR	61				
3	22960	JAM MAKING SET WITH JARS	3089		3	21428	SET3 BOOK BOX GREEN GINGHAM FLOWER	51				
4	22838	3 TIER CAKE TIN RED AND CREAM	3042		4	47594B	SCOTTIES DESIGN WASHBAG	47				
5	85123A	WHITE HANGING HEART T-LIGHT HOLDER	2858		5	22930	BAKING MOULD HEART MILK CHOCOLATE	46				
6	21843	RED RETROSPOT CAKE STAND	2705		6	23271	CHRISTMAS TABLE SILVER CANDLE SPIKE	40				
Sr no.	Stock code (Germany)	Description	Total sale		Sr no.	Stock code (RSA)	Description	Total sale				
1	POST	POSTAGE	20821		1	21340	CLASSIC METAL BIRDCAGE PLANT HOLDER	38				
2	22423	REGENCY CAKESTAND 3 TIER	8257		2	22605	WOODEN CROQUET GARDEN SET	30				
3	22326	ROUND SNACK BOXES SET OF4 WOODLAND	3555		3	23298	SPOTTY BUNTING	30				
4	22328	ROUND SNACK BOXES SET OF 4 FRUITS	1950		4	22423	REGENCY CAKESTAND 3 TIER	26				
5	22629	SPACEBOY LUNCH BOX	1629		5	85066	JAM MAKING SET WITH JARS	26				
6	22554	PLASTERS IN TIN WOODLAND ANIMALS	1389		6	22526	CREAM SWEETHEART MINI CHEST	26				
Sr no.	Stock code (France)	Description	Total sale		Sr no.	Stock code (Brazil)	Description	Total sale				
1	POST	POSTAGE	15065		1	22423	REGENCY CAKESTAND 3 TIER	175				
2	23084	RABBIT NIGHT LIGHT	7275		2	22722	SET OF 6 SPICE TINS PANTRY DESIGN	83				
3	22423	REGENCY CAKESTAND 3 TIER	2582		3	21430	SET/3 RED GINGHAM ROSE STORAGE BOX	81				
4	21731	RED TOADSTOOL LED NIGHT LIGHT	2168		4	22366	DOORMAT AIRMAIL	68				
5	22554	PLASTERS IN TIN WOODLAND ANIMALS	1868		5	22699	ROSES REGENCY TEACUP AND SAUCER	61				
6	84879	ASSORTED COLOUR BIRD ORNAMENT	1843		6	22698	GREEN REGENCY TEACUP AND SAUCER	61				

The purpose of this analysis was to gain a clear understanding of the countries, products, sales trends, and overall performance. **This understanding helped me frame the relevant questions.**

<u>**CEO**</u>

1. Can launching our best selling products of the top 5 countries which are the UK, Netherlands, EIRE, Germany, France in the low performing countries help boost their sales and reduce the performance gap?
2. Should we look to expand in countries that have similar buying patterns to the UK, based on what products are doing well?
3. Are we possibly missing out on revenue in the bottom 5 countries because the right products aren't available or aligned with customer needs?
4. Would it make sense to adjust our inventory and introduce the top UK products in low-performing countries to get better results?

<u>**CMO**</u>

1. Have our bestselling top 5 countries products even been promoted or made visible in the bottom 5 countries?
2. Can targeted marketing campaigns around our best selling UK products help create demand in countries like Finland or Bahrain?
3. Is the low performance in bottom countries because of cultural differences, or are those products just not available there yet?
4. Which customer segments in low performing countries are similar to our UK audience and can we target them with the same products?
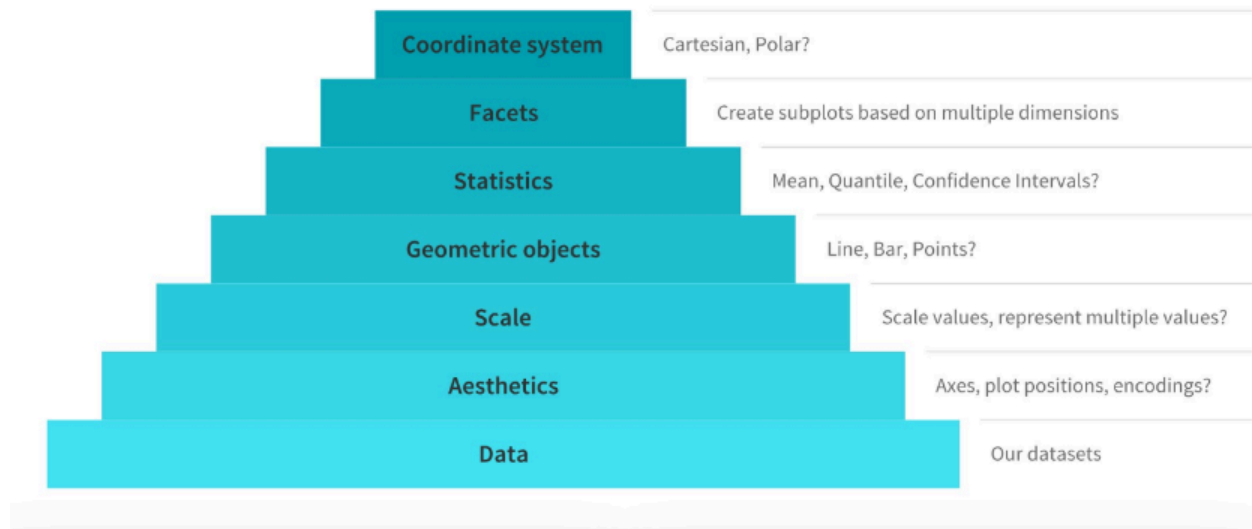
TASK 2 : Identifying the visual which would explain the data and insights in the simplest possible manner

To gain a clear understanding of the data, I explored different types of visualizations. Each chart type highlights different aspects of the data, and choosing the right chart makes insights easier to understand

So from my understanding there are 5 main charts

1. **Comparison chart** (Column chart, bar chart, line chart, scatter chart)
2. **Distribution chart** (Histogram chart, boxplot chart, KDE plot)
3. **Whole chart** (Pie chart, Donut chart, Stacked column, Stacked bar chart)
4. **Relationship chart** (Scatter chart, line chart)
5. **Trend chart** (Line chart, area chart, column chart)

Major Components of the Grammar of Graphics

Coordinate system	Cartesian, Polar?
Facets	Create subplots based on multiple dimensions
Statistics	Mean, Quantile, Confidence Intervals?
Geometric objects	Line, Bar, Points?
Scale	Scale values, represent multiple values?
Aesthetics	Axes, plot positions, encodings?
Data	Our datasets

TASK 3: Creating the visuals around the questions the CEO and CMO requested

Apart from general cleaning (removal of duplicates, missing values, and invalid records), two specific validation checks were implemented:

- Ensured that **Quantity ≥ 1** (no transactions with zero or negative units).
- Ensured that **UnitPrice > 0** (no free or invalid-priced items).

When uploading the data into Tableau, we apply filters so that the dataset is immediately transformed to meet our requirements.

1. Click **Connect → Microsoft Excel**.
2. Select the dataset (the online retail Excel file we've been using).

In the **Data Source** tab:

- Drag the main sheet
- Apply filters: Quantity ≥ 1, Unit Price ≥ 0
- This ensures you only include valid transactions.

As part of the case study, the Chief Officers outlined four specific questions for analysis:

Question 1

The CEO of the retail store is interested to view the time series of the revenue data for the year 2011 only. He would like to view granular data by looking into revenue for each month. The CEO is interested in viewing the seasonal trends and wants to dig deeper into why these trends occur. This analysis will be helpful for the CEO to forecast for the next year.

A **line chart** was chosen because it's the best way to show revenue trends over time. It connects each month's data, making seasonal peaks and dips easy to see, which helps the CEO quickly spot patterns and use them for forecasting.
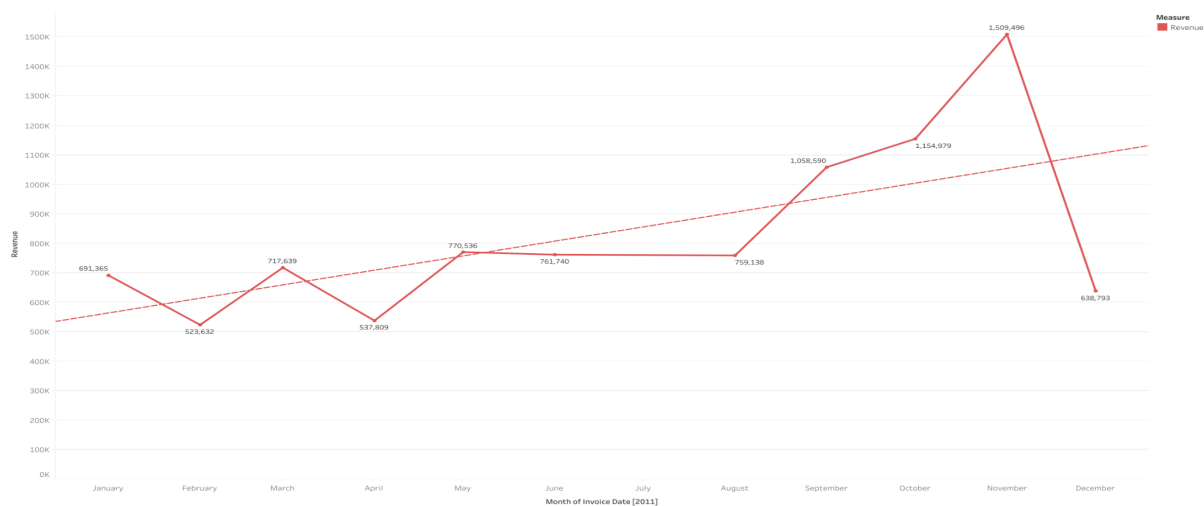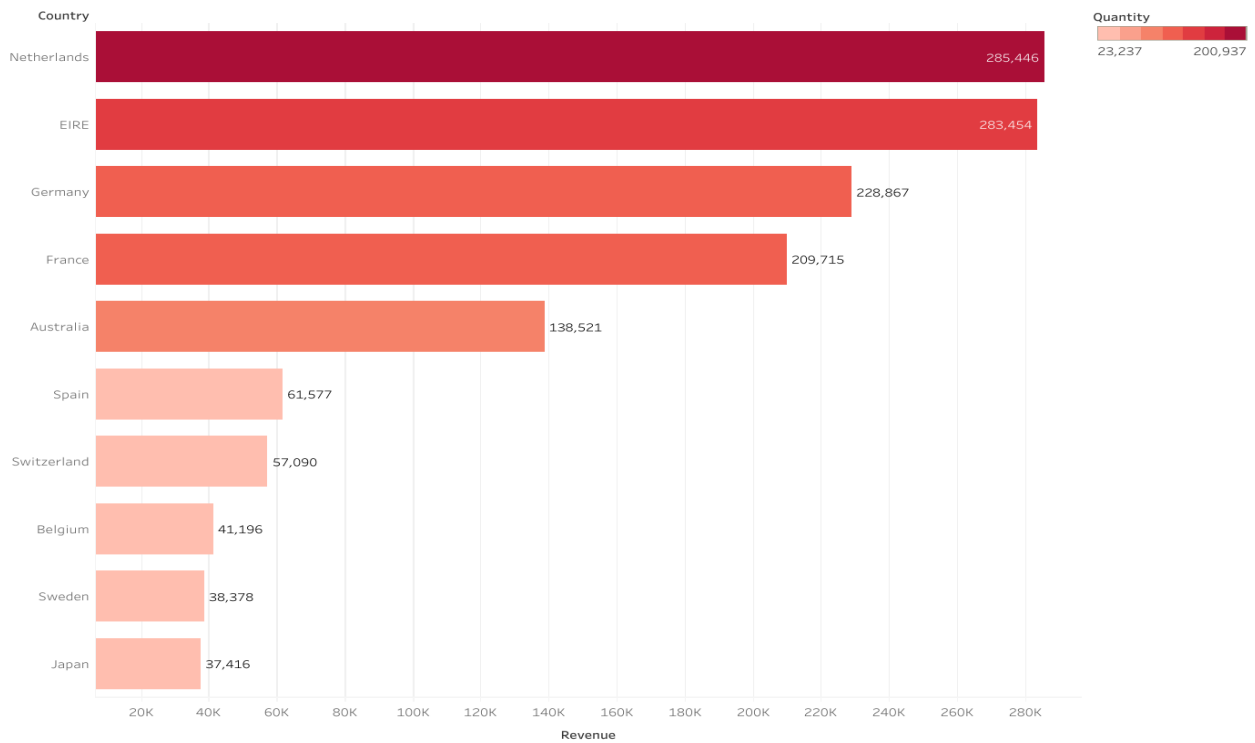

Question 2

The CMO is interested in viewing the top 10 countries which are generating the highest revenue. Additionally, the CMO is also interested in viewing the quantity sold along with the revenue generated. The CMO does not want to have the United Kingdom in this visual.

A **Stacked bar chart** was chosen because it allows the CMO to see both revenue and quantity sold in a single visual. The bars make it easy to compare countries side by side, while the color intensity adds another layer showing the relative quantities. This way, the CMO doesn't just see who brings the highest revenue, but also understands how volume contributes to it making it a much richer comparison across the top 10 countries.
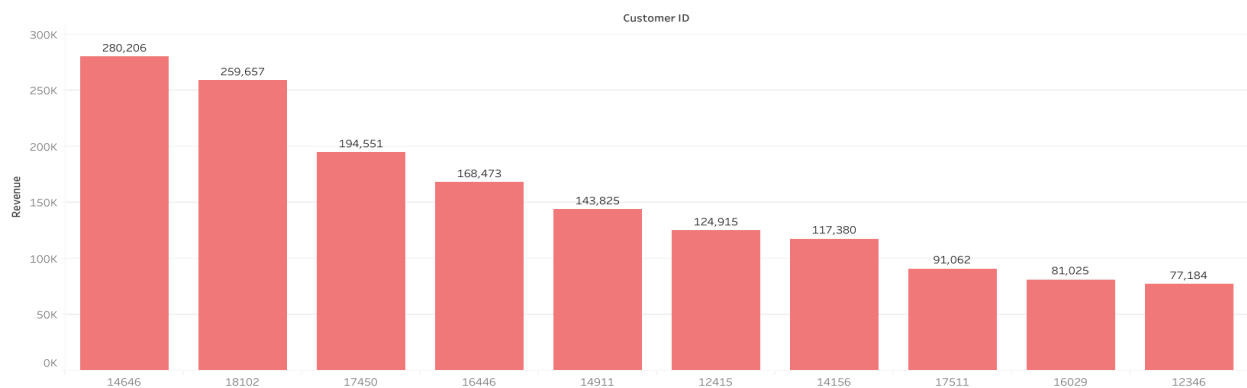
Top 10 Countries (Excluding UK)

Country	Revenue	Quantity

Quantity: 23,237 — 200,937

Netherlands — 285,446
EIRE — 283,454
Germany — 228,867
France — 209,715
Australia — 138,521
Spain — 61,577
Switzerland — 57,090
Belgium — 41,196
Sweden — 38,378
Japan — 37,416

Revenue axis: 20K 40K 60K 80K 100K 120K 140K 160K 180K 200K 220K 240K 260K 280K

Revenue

Question 3

The CMO wants a visual of the top 10 customers ranked by revenue, starting with the highest and declining to the lowest. This will help identify and prioritize high-value customers to ensure their continued satisfaction.

A **Stacked column chart** was chosen since it best highlights the top 10 customers by revenue in descending order. This makes it easy for the CMO to quickly spot high-value customers and see how revenue drops across the top segment

Top 10 customers by revenue(Excluding unknown customer)

Customer ID

Customer ID	Revenue
14646	280,206
18102	259,657
17450	194,551
16446	168,473
14911	143,825
12415	124,915
14156	117,380
17511	91,062
16029	81,025
12346	77,184

Revenue axis: 0K 50K 100K 150K 200K 250K 300K

Question 4

The CEO wants a single-view analysis of product demand across all countries (excluding the UK) to identify regions with the highest potential. These insights will guide expansion strategies aimed at targeting high-demand areas and generating more business.

A **Filled map chart** was chosen since it best highlights a clear,single-view analysis of product demand across countries. The geographic layout makes it easy for the CEO to spot high-demand regions at a glance, which directly supports expansion planning.

Product's demand in regions (Excluding UK)
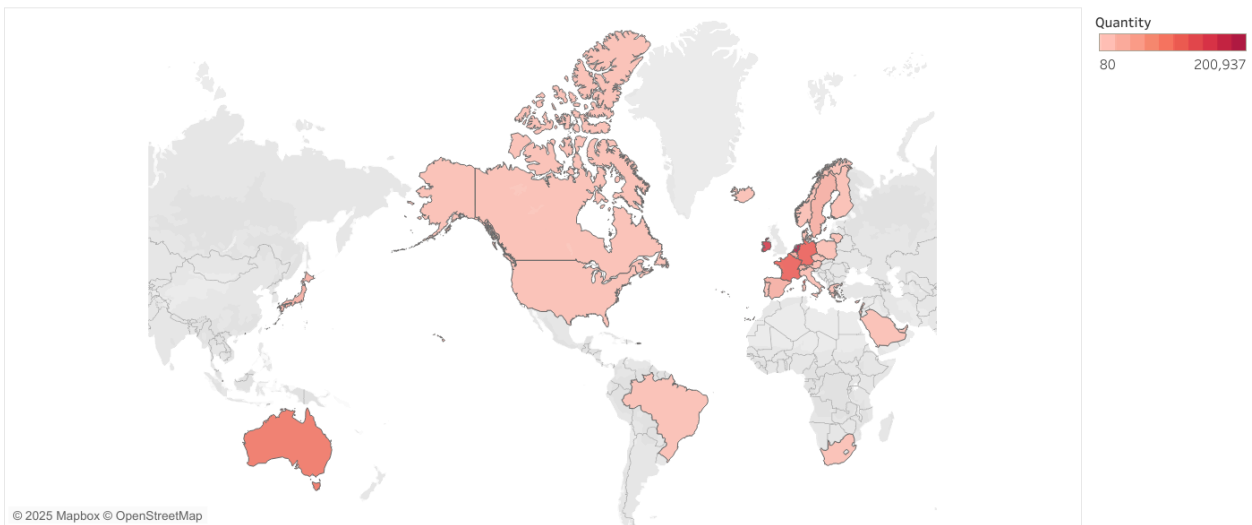
For each question, we excluded the UK since it is an outlier its revenue is significantly higher than the other countries, which would otherwise distort the comparison

TASK 4: Presenting the findings to the CEO and CMO

While working with this dataset, a few interesting patterns stood out.

Revenue Trends

- Strong seasonality observed: November 2011 peaked at **1.5M**, followed by October and September (both above 1M).
- Quarter 4 is the biggest sales window, likely driven by holiday purchases.
- A smaller rise in May–June suggests an opportunity for summer-focused campaigns.

Top Markets (Excluding UK)

- Leading countries: Netherlands, EIRE, Germany, France, Australia.
- Each market shows different preferences (e.g., Germany/EIRE - tiered cake stands, Netherlands/France - decorative kitchenware).
- Insight: Localized marketing strategies will outperform one size fits all campaigns.

High-Value Customers

- Top 10 customers contribute significantly, with the leader generating 280K+ in revenue.
- Most are wholesale/bulk buyers making large seasonal purchases.
- Strategy: Build loyalty via early access, bulk purchase discounts, and exclusive offers.

Expansion Potential

- Beyond Europe, Australia, Canada, Asia, and the Middle East show strong demand.
- Smart entry strategy: Double down on products already doing well in each region e.g., novelty/gift items in the Middle East, kitchenware in Australia.

.

Final Takeaways

Working on this case study gave me the chance to move from raw data to clear business insights by applying proper cleaning, filtering, and visualization techniques. More than just answering the CEO and CMO's questions, the process showed me how important it is to structure data in a way that directly connects to decision-making.

This experience reinforces my belief that good analytics is not only about finding patterns in the data but also about presenting them in a way that drives strategy. As a next step, combining these visuals with predictive models could help anticipate future demand and guide expansion with even more confidence.