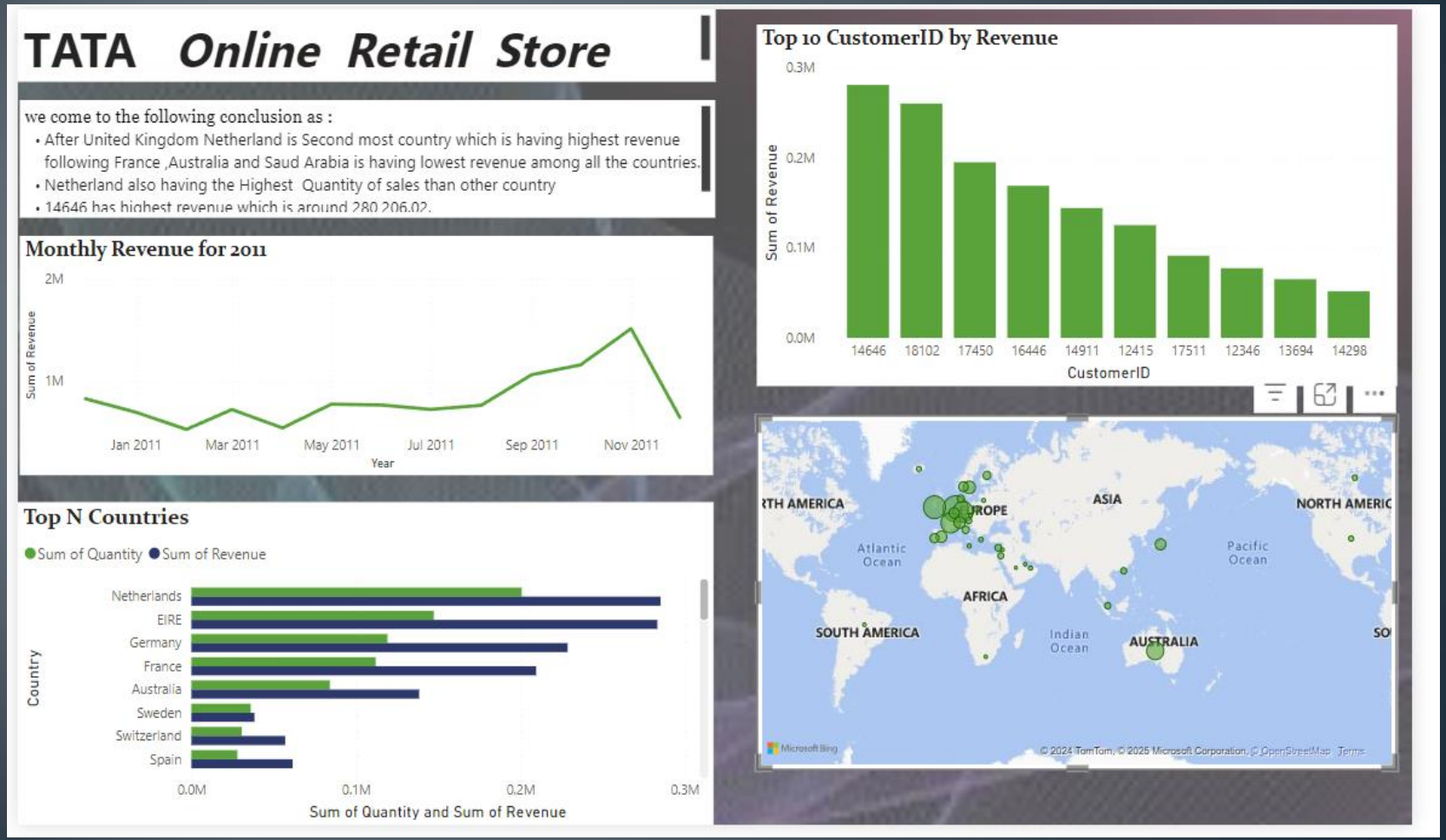
TATA SALES REPORT

BY PRIYANKA JOSHI



In [4]:
```
df.describe()
```

Out[4]:

	Quantity	UnitPrice	CustomerID
count	541909.000000	541909.000000	406829.000000
mean	9.552250	4.611114	15287.690570
std	218.081158	96.759853	1713.600303
min	-80995.000000	-11062.060000	12346.000000
25%	1.000000	1.250000	13953.000000
50%	3.000000	2.080000	15152.000000
75%	10.000000	4.130000	16791.000000
max	80995.000000	38970.000000	18287.000000

In [5]:
```
cleaned_data = df[(df['Quantity']>= 1)& (df['UnitPrice']>= 0)]
```

In [6]:
```
cleaned_data['Revenue']=cleaned_data['Quantity']*cleaned_data['UnitPrice']
```

```
C:\Users\hp\AppData\Local\Temp\ipykernel_2336\3494526920.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-v
rsus-a-copy
  cleaned_data['Revenue']=cleaned_data['Quantity']*cleaned_data['UnitPrice']
```

In [7]:
```
cleaned_data.head()
```

In [1]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
C:\Users\hp\anaconda3\lib\site-packages\scipy\__init__.py:146: UserWarning: A NumPy version >=1.16.5 and <1.23.0 is re
r this version of SciPy (detected version 1.23.5
  warnings.warn(f"A NumPy version >={np_minversion} and <{np_maxversion}"
```

In [2]:
```
df = pd.read_excel('Online Retail Data Set.xlsx')
```

In [3]:
```
df.head(5)
```

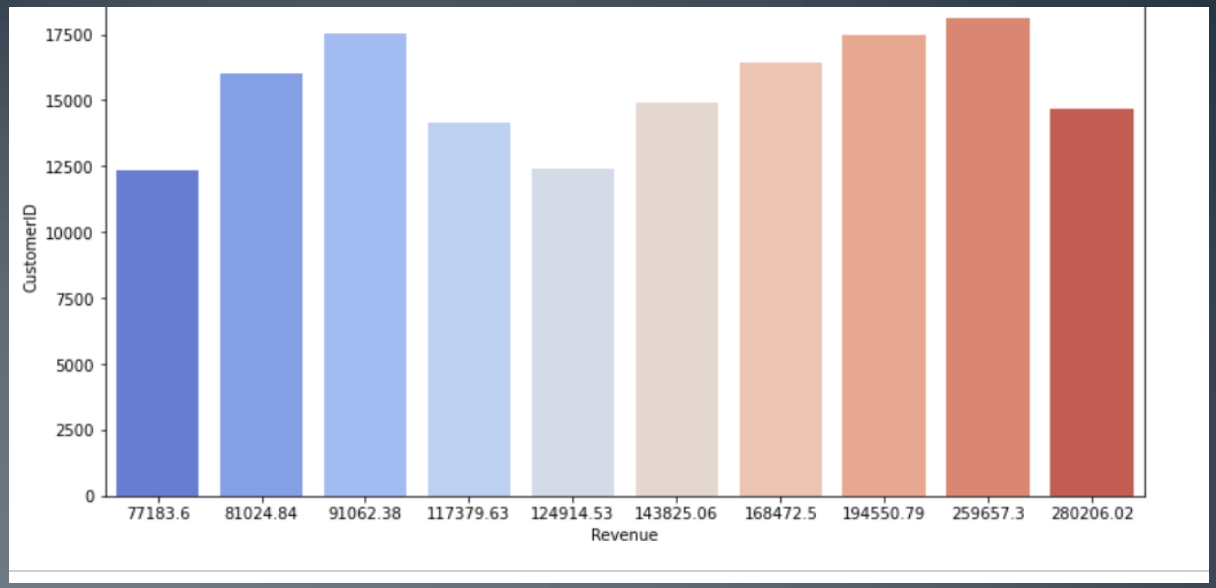
Out[3]:

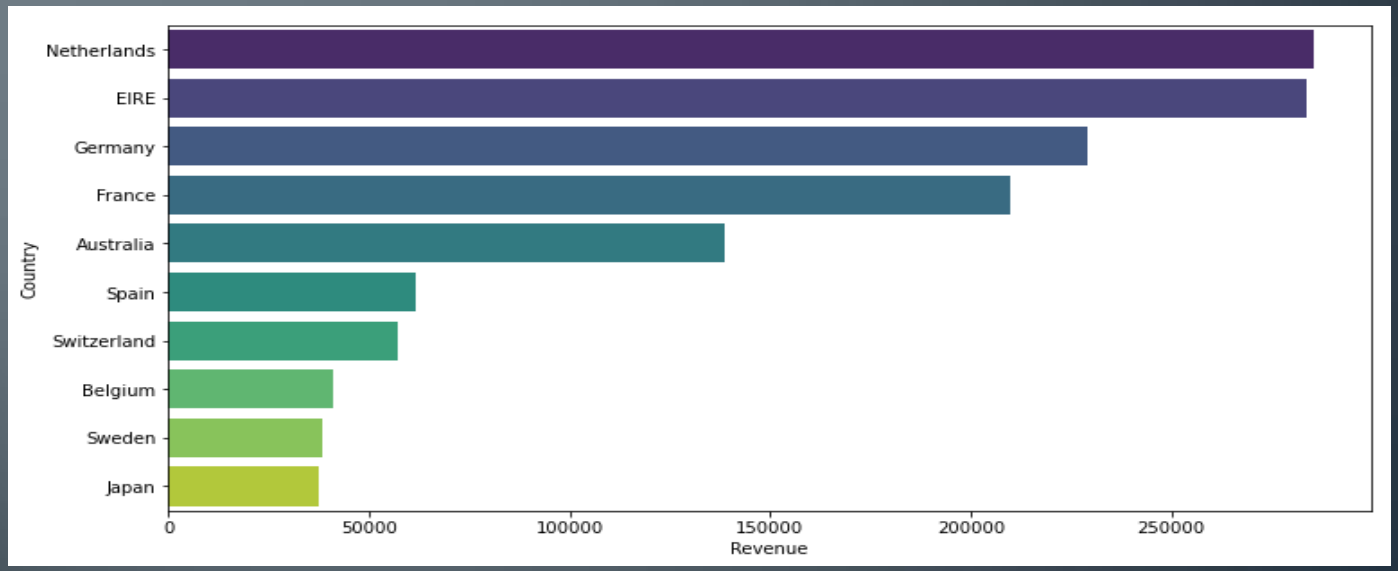
	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
0	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	2010-12-01 08:26:00	2.55	17850.0	United Kingdom
1	536365	71053	WHITE METAL LANTERN	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
2	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	2010-12-01 08:26:00	2.75	17850.0	United Kingdom
3	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom
4	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom

In [4]:
```
df.describe()
```

Out[4]:

	Quantity	UnitPrice	CustomerID
count	541909.000000	541909.000000	406829.000000
mean	9.552250	4.611114	15287.690570
std	218.081158	96.759853	1713.600303
min	-80995.000000	-11062.060000	12346.000000
25%	1.000000	1.250000	13953.000000
50%	3.000000	2.080000	15152.000000
75%	10.000000	4.130000	16791.000000
max	80995.000000	38970.000000	18287.000000


17500
15000
12500
10000
7500
5000
2500
0
CustomerID
77183.6
81024.84
91062.38
117379.63
124914.53
143825.06
168472.5
194550.79
259657.3
280206.02
Revenue


Netherlands
EIRE
Germany
France
Australia
Spain
Switzerland
Belgium
Sweden
Japan
Country
0
50000
100000
150000
200000
250000
Revenue

Presentation Script for CEO and CMO

Data Cleanup: Before diving into the analysis, we ensured that the data was clean and accurate. During the initial assessment, we identified two major issues:

Negative quantities, representing returns, were excluded to avoid skewing the results.

Unit prices below $0, likely due to input errors, were also removed.

We applied conditional formulas and data transformation methods to ensure these anomalies were corrected. This step was crucial for providing you with reliable insights.

Question 1: Revenue Trends for 2011 The CEO's interest in revenue trends by month was analyzed using a time series visualization for 2011. Key findings include:

Seasonal peaks were observed in [specific months].

[Mention any trends like holiday season impact or promotional periods if applicable].

These insights provide a basis for forecasting revenue trends and planning seasonal strategies for the upcoming year.

Question 2: Top 10 Countries by Revenue (Excluding the UK) The CMO's focus was on identifying the top 10 revenue-generating countries, excluding the United Kingdom. The analysis revealed:

Countries like [list top countries] contributed significantly to revenue.

A parallel comparison of quantities sold highlighted [specific patterns, e.g., higher volume vs. higher value markets].

These insights can guide marketing efforts and resource allocation to optimize returns from these high-performing regions

Question 3: Top 10 Customers by Revenue Understanding our highest revenue-generating customers is essential. The analysis highlighted:
The top 10 customers contributed to [X]% of total revenue.
The visualized revenue gradient from the top to the 10th customer emphasizes the importance of customer retention and satisfaction strategies for these key contributors.
This information equips us to tailor customer loyalty programs and personalized engagements for our most valuable clients.

Question 4: Product Demand by Region (Excluding the UK) For the CEO's interest in identifying regions with the highest demand, we analyzed product quantities sold across all countries, excluding the UK. The findings revealed:
The highest demand was observed in regions such as [list top-demand regions].
This indicates potential expansion opportunities in [specific regions].
This insight provides a roadmap for targeting untapped markets and driving growth in demand-heavy areas Thank You

Thank You