

## Mini Project—2

### Data Spark: Illuminating insights for Global Electronics

#### Introduction

We have datasets related to customers, products, sales, stores and currency exchange rates. By analysing this data I was able to identify key factors that will help improve the sales, develop better products and better customer outreach.

#### Approach

Firstly I have loaded my data which is in .csv format into my jupyter notebook and checked the shape of the data, identified the null values and duplicates if any and cleaned the data.

Then I have created schemas in postgresSQL in a pythonic fashion and inserted my cleaned data into the respective schemas in my table Global\_Electronics

	CustomerKey	Gender	Name	City	State	Country	Continent	Birthday
0	301	Female	Lilly Harding	WANDEARAH EAST	South Australia	Australia	Australia	1939-07-03
1	325	Female	Madison Hull	MOUNT BUDD	Western Australia	Australia	Australia	1979-09-27
2	554	Female	Claire Ferres	WINJALLOK	Victoria	Australia	Australia	1947-05-26
3	786	Male	Jai Poltpalingada	MIDDLE RIVER	South Australia	Australia	Australia	1957-09-17
4	1042	Male	Aidan Pankhurst	TAWONGA SOUTH	Victoria	Australia	Australia	1965-11-19
...	...	...	...	...	...	...	...	...
15261	2099600	Female	Denisa Duřková	Houston	Texas	United States	North America	1936-03-25
15262	2099618	Male	Justin Solórzano	McLean	Virginia	United States	North America	1992-02-16
15263	2099758	Male	Svend Petrussen	Wilmington	North Carolina	United States	North America	1937-11-09
15264	2099862	Female	Lorenza Rush	Riverside	California	United States	North America	1937-10-12
15265	2099937	Male	Zygmunt Kaminski	Bloomfield Township	Michigan	United States	North America	1965-08-18

15266 rows x 8 columns

Fig. Customers.csv after data cleaning

	Date	Currency	Exchange
0	2015-01-01	USD	1.0000
1	2015-01-01	CAD	1.1583
2	2015-01-01	AUD	1.2214
3	2015-01-01	EUR	0.8237
4	2015-01-01	GBP	0.6415
...	...	...	...
11210	2021-02-20	USD	1.0000
11211	2021-02-20	CAD	1.2610
11212	2021-02-20	AUD	1.2723
11213	2021-02-20	EUR	0.8238
11214	2021-02-20	GBP	0.7126

11215 rows x 3 columns

Fig. Exchange\_Rates.csv after data cleaning

	ProductKey	Product Name	Brand	Color	Unit Cost USD	Unit Price USD	SubcategoryKey	Subcategory	CategoryKey	Category
0	1	Contoso 512MB MP3 Player E51 Silver	Contoso	Silver	\$6.62	\$12.99	101	MP4&MP3	1	Audio
1	2	Contoso 512MB MP3 Player E51 Blue	Contoso	Blue	\$6.62	\$12.99	101	MP4&MP3	1	Audio
2	3	Contoso 1G MP3 Player E100 White	Contoso	White	\$7.40	\$14.52	101	MP4&MP3	1	Audio
3	4	Contoso 2G MP3 Player E200 Silver	Contoso	Silver	\$11.00	\$21.57	101	MP4&MP3	1	Audio
4	5	Contoso 2G MP3 Player E200 Red	Contoso	Red	\$11.00	\$21.57	101	MP4&MP3	1	Audio

Fig.Products.csv after data cleaning

	Order Number	Line Item	Order Date	Delivery Date	CustomerKey	StoreKey	ProductKey	Quantity	Currency Code
0	366000	1	1/1/2016	NaN	265598	10	1304	1	CAD
1	366001	1	1/1/2016	1/13/2016	1269051	0	1048	2	USD
2	366001	2	1/1/2016	1/13/2016	1269051	0	2007	1	USD
3	366002	1	1/1/2016	1/12/2016	266019	0	1106	7	CAD
4	366002	2	1/1/2016	1/12/2016	266019	0	373	1	CAD

Fig.Sales.csv after data cleaning

	StoreKey	Country	State	Square Meters	Open Date
0	1	Australia	Australian Capital Territory	595.0	2008-01-01
1	2	Australia	Northern Territory	665.0	2008-01-12
2	3	Australia	South Australia	2000.0	2012-01-07
3	4	Australia	Tasmania	2000.0	2010-01-01
4	5	Australia	Victoria	2000.0	2015-12-09

Fig.Stores.csv after data cleaning

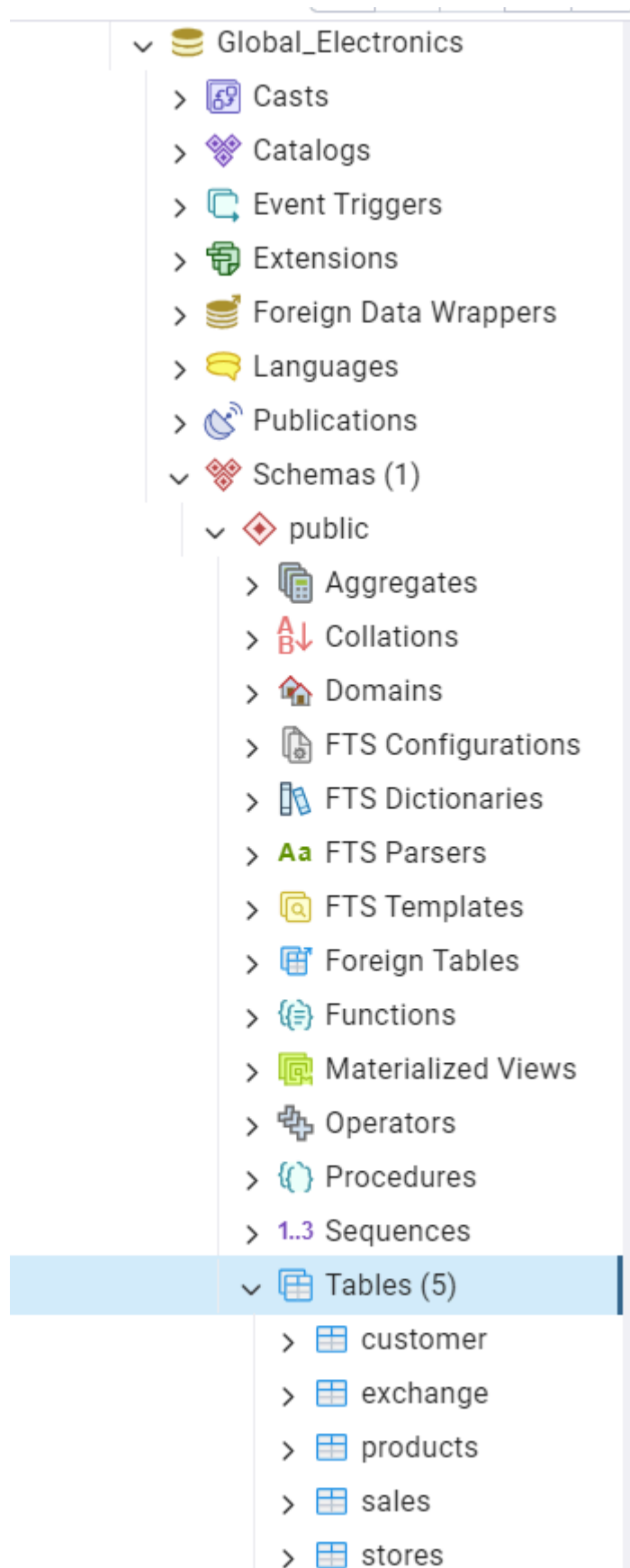
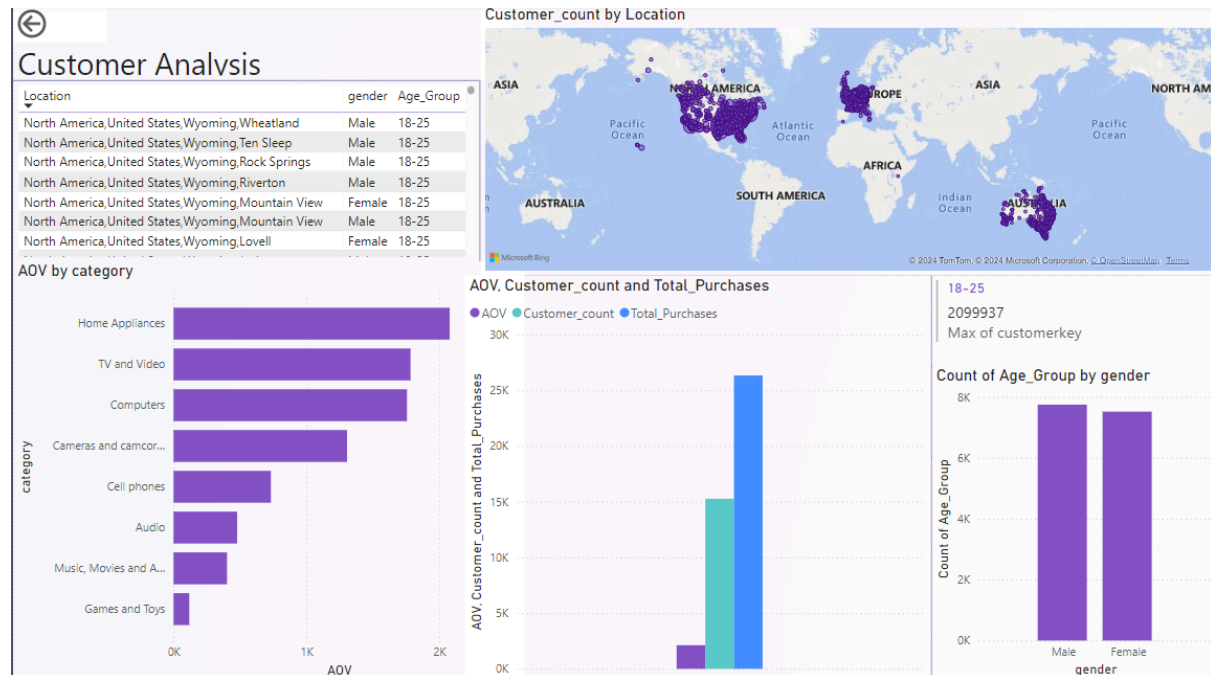


Fig.Global\_Electronics schemas

Then using Power BI I have created visualisations by following below approach:

- Data Preparation
- Importing Data
- Creating relevant Measures using DAX in Power Query
- Create relevant charts and corresponding slicers and tables

## 1.CUSTOMER ANALYSIS



1. From from the **CUSTOMER ANALYSIS** we derive the following

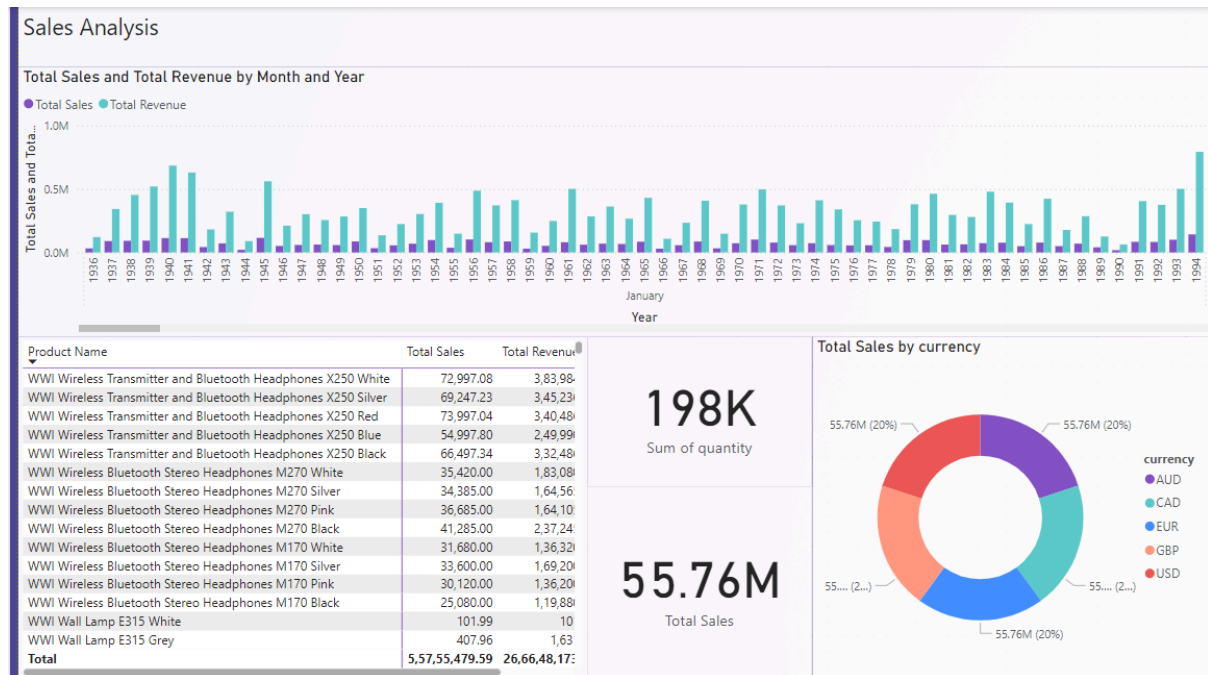
→ From the map which demonstrates Customer segmentation based on location Here we observe that there is high demand for products from North America,Europe and Australia and a relatively lower customer segment in regions like Asia, Africa, and South America. This distribution indicates that There is a good outreach in Western provinces.

→ The bar chart which demonstrates AOV by Category tells us that **Home Appliances** have the highest AOV,followed by TV and Video and Computers.**Games and Toys** have the lowest AOV

→ The clustered bar graph tells us that there is a correlation between customer count and total purchases. However, AOV varies independently, suggesting that higher customer counts do not necessarily equate to higher spending per order.

→ From the measure created called Age\_Group the top customer belong to an age group of 18-25.

## 2.SALES ANALYSIS



From the **SALES ANALYSIS** we derive the following:

### Total Sales and Revenue Trends by Month and Year:

- The bar chart shows monthly trends in **Total Sales** and **Total Revenue** across multiple years. There are noticeable spikes in certain years, which could indicate peak sales periods or successful promotions.

### Top-Selling Products:

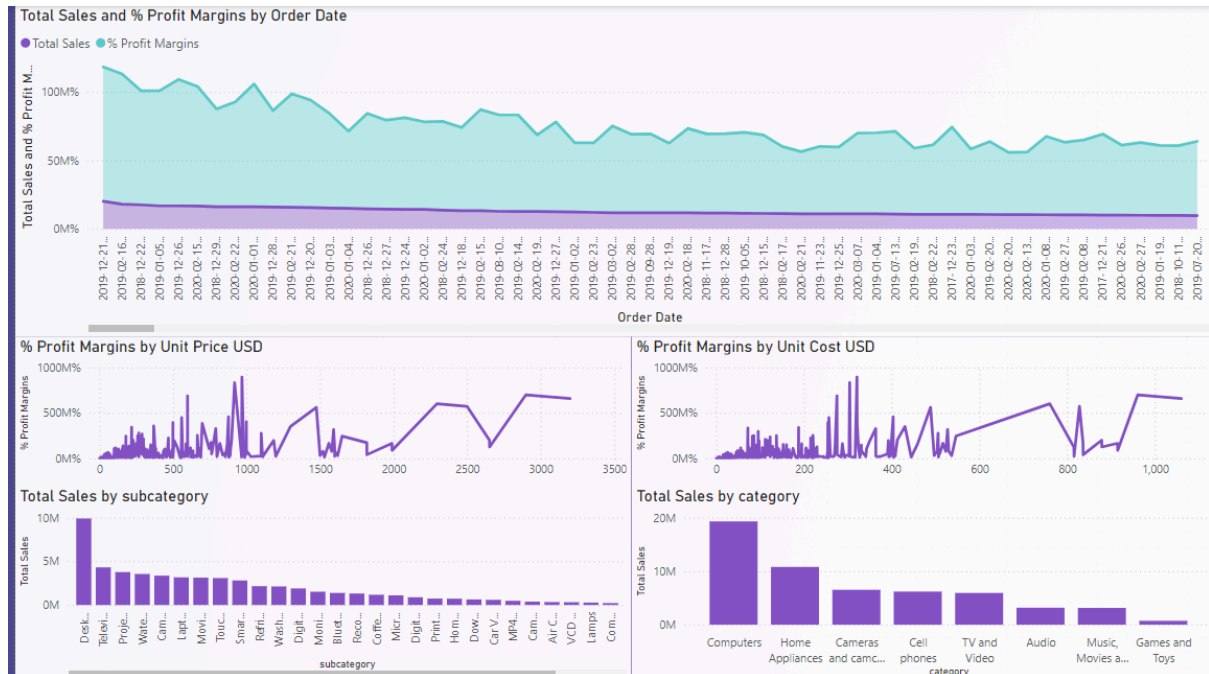
- The table displays individual product performance, showing the **Total Sales** and **Total Revenue** for each product.
- The "WWI Wireless Transmitter and Bluetooth Headphones X250" in various colours (White, Silver, Red, Blue, and Black) appear to be the top-selling items. This suggests that this product line is a significant revenue driver.

### Total Sales by Currency:

- The pie chart indicates that sales are distributed across multiple currencies (AUD, CAD, EUR, GBP, and USD), each contributing around 20% to the total sales. This

suggests a balanced international market reach, with each currency zone representing a significant portion of sales.

### 3.PRODUCT ANALYSIS



From the **PRODUCTS ANALYSIS** we derive the following:

#### Total Sales and Profit Margins over Time:

- The top area chart shows **Total Sales** and **% Profit Margins** over time. The **Total Sales** appears relatively steady, with minor fluctuations. However, the **Profit Margin** remains consistently low, indicating a potential issue with profitability, even during high sales periods.
- This trend suggests that while the business is generating revenue, it's not achieving substantial profit margins, which could imply high costs or low pricing strategies.

#### Profit Margins by Unit Price:

- The left chart displays **% Profit Margins** relative to the **Unit Price** in USD. There are spikes in profit margin percentages, particularly at certain price points around 1,000 and 3,000 USD.

- This could indicate specific products at these price levels that yield higher profit margins, potentially due to premium pricing or lower associated costs. These price points may represent an opportunity for profit maximisation.

### **Profit Margins by Unit Cost:**

- The right chart shows **% Profit Margins** based on **Unit Cost** in USD. There are notable spikes in profitability for certain unit costs, especially at lower costs around 200 and 800 USD.
- This suggests that products with lower production costs tend to yield higher profit margins, possibly due to cost efficiency. Focusing on cost-effective production or sourcing for these items could improve overall profitability.

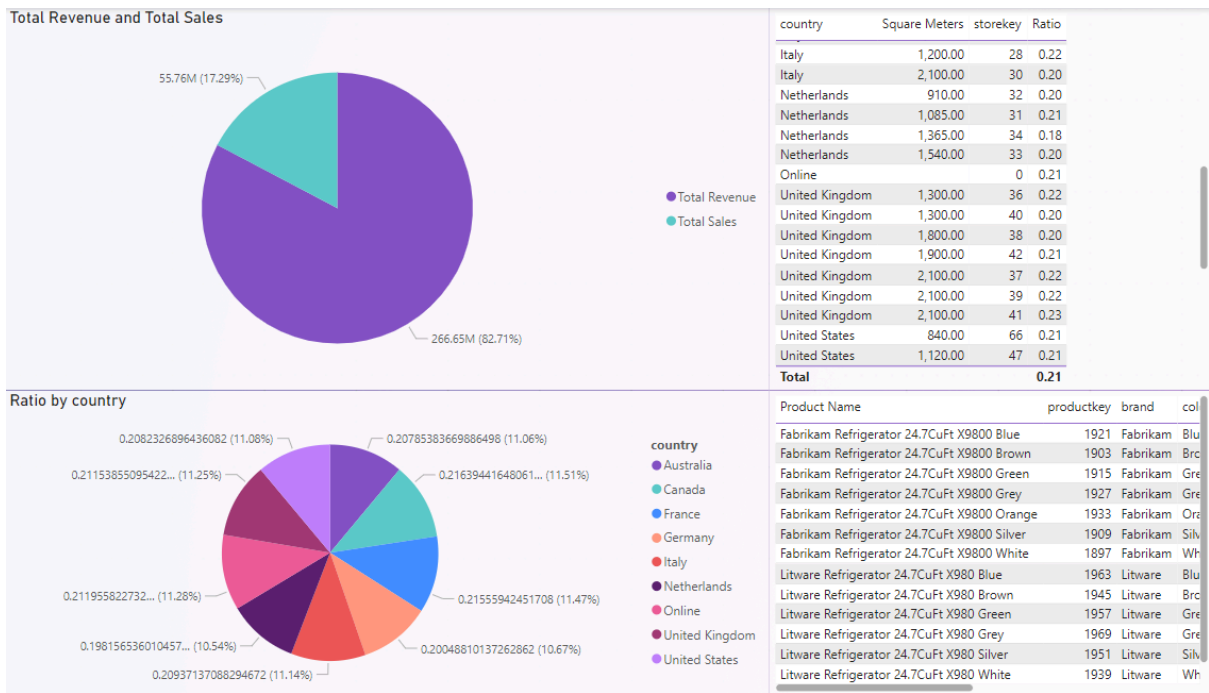
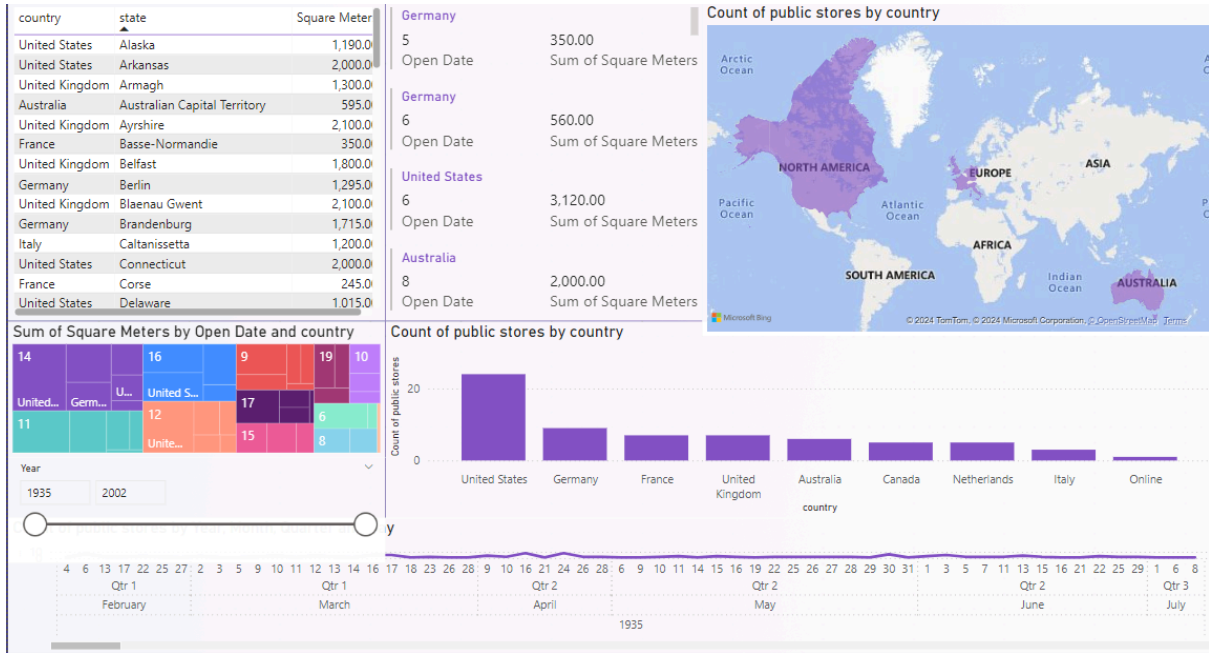
### **Total Sales by Subcategory:**

- The bar chart for **Total Sales by Subcategory** shows that **Desktops** have the highest sales, followed by other subcategories like **Projectors**, **Phones**, and **Cameras**.
- This insight indicates that desktops are a significant revenue driver, and inventory or marketing could be tailored to support this popular subcategory.

### **Total Sales by Category:**

- The bottom-right chart displays **Total Sales by Category**, with **Computers** having the highest sales, followed by **Home Appliances** and **Cameras**.
- Focusing on these top-selling categories can improve sales performance. Further, it might be beneficial to analyse whether products in other categories, like **Games and Toys**, have potential for growth or need repositioning.

4.STORE ANALYSIS



From STORE ANALYSIS we derive the following:

Overall Store Count and Square Meters:

- United States dominates in terms of the highest number of public stores and total square meters.



- Germany follows closely in terms of total square meters.
- France, the United Kingdom, and Australia have a significant presence as well.
- Smaller countries like Italy, Canada, and the Netherlands have a smaller but still notable footprint.

#### **Store Count by Country:**

- The bar chart highlights the top countries in terms of public store count.
- United States clearly leads, followed by Germany and France.
- The remaining countries show a gradual decline in store count.

#### **Store Count by Quarter and Year:**

- The line chart shows a general upward trend in store count over the years.
- There are peaks and troughs, indicating fluctuations in store openings or closures.
- The quarterly breakdown reveals seasonal patterns, with potential spikes in certain quarters.

#### **Overall Ratio:**

- The overall ratio is 0.21, indicating that for every dollar of revenue, 21 cents are generated in sales.

#### **Ratio by Country:**

- **Italy:** The highest ratio is in Italy (0.22), suggesting a relatively strong sales performance compared to revenue.
- **Netherlands:** The Netherlands also has a high ratio (0.21), indicating good sales efficiency.
- **United Kingdom:** The UK has a slightly lower ratio (0.20), suggesting a potential need to improve sales efficiency.
- **United States:** The US has the lowest ratio (0.18), indicating room for improvement in sales performance relative to revenue.

#### **Ratio by Product Name:**

- **Fabrikam Refrigerator 24.7CuFt X9800 Blue:** This product has the highest ratio (0.21), indicating strong sales performance.
- **Litware Refrigerator 24.7CuFt X9800 Brown:** This product has the lowest ratio (0.19), suggesting potential room for sales improvement.

#### **Additional Insights:**

- The total revenue is significantly higher than the total sales, indicating that there might be factors like returns, discounts, or other adjustments affecting the sales figure.
- The online sales channel has a relatively high ratio (0.21), which could be due to lower operational costs or higher online conversion rates.

