1. Project Title:
Regional Pricing Impact Analysis

Name:
Gundeti Sailikith – Aspiring Data Analyst

Tools Used:
SQL, Python, Excel

2. Executive Summary:

This project investigates regional pricing inefficiencies for a global electronics retailer, aiming to optimize profitability without compromising customer trust. The business challenge stemmed from inconsistent gross profits and margins across different countries and store locations. The goal was to identify where pricing was misaligned and recommend data-driven adjustments.

Using a dataset of over 62,000 records across six relational tables—including customer details, product pricing, sales transactions, store attributes, and currency exchange rates—I built a comprehensive analysis pipeline. SQL was used to ingest and aggregate data within Jupyter Notebook, followed by Python-based cleaning, exploratory analysis, and hypothesis testing. Excel dashboards were developed to present insights in a business-friendly format.

Key outcomes include:

- Identification of low-performing regions with inflated selling prices relative to manufacturing costs

- Identification of Top contribution customers according to sales, to know from which regions we have trusted customers most

- Statistically significant margin gaps between product categories across countries

- Actionable pricing recommendations tailored to regional cost structures and customer sensitivity

3. Business Objective:

The central business question guiding this analysis is: **How can we adjust pricing to improve margins?** To address this, the project focused on identifying which products and product categories were performing well across different regions, and where pricing strategies were misaligned with customer behaviour and cost structures.

This analysis is designed to benefit key stakeholders including the **pricing team** and **retail managers**, who are responsible for setting and executing store-level pricing strategies. By uncovering regional performance trends and margin disparities, the findings offer targeted insights to support smarter pricing decisions.

The ideal outcome is to **implement an optimized pricing strategy across stores** that improves profitability **without losing customer trust**. This means aligning prices with

regional cost dynamics and customer expectations, ensuring competitive positioning while maintaining brand integrity and loyalty.

4. Data Overview:

Source: Global Electronics Retailer-Maven Analytics

Link: https://mavenanalytics.io/data-playground/global-electronics-retailer

Volume:

It contains 6 tables including Data dictionary which is Database Schema that contains 5 tables information and their relationships

Tables:

Customers

Products

Sales

Stores

Exchange_Rates

Data_Dictionary(Schema)

Data types: Categorical, numerical, geographic

5. Methodology:

Tools Used:

- SQL (via Jupyter Notebook) for data ingestion and aggregation
- Python (Pandas, Seaborn, Matplotlib, SciPy) for cleaning, analysis, visualization and Hypothesis Testing
- Excel for dashboard creation and stakeholder presentation

Workflow:

1. Ingested and joined datasets using SQL queries
2. Created summary table and added new columns (e.g., Total Cost, Gross Profit, etc)
3. Performed exploratory data analysis (EDA) numerical and categorical columns characteristics
4. Conducted regional and product comparisons and hypothesis testing
5. Built Excel dashboards to visualize pricing inefficiencies and recommendations

Key Techniques:

- Margin calculation based on selling price vs. manufacturing cost

- Regional performance clustering
- Category-level profitability analysis
- Sensitivity checks to ensure pricing changes wouldn't erode trust

Challenges & Solutions:

- Ensured reproducibility across tools by documenting SQL and Python workflows

- Balanced analytical depth with business clarity in dashboard design

6. Analysis & Insights:

Exploratory Data Analysis:

Statistical Summary:

|	Customer_ID	storekey	Total_Quantity	Unit_Cost_USD	Unit_Price_USD	Total_Cost	Total_Sales	Gross_Profit	Gross_Margin
count	6.277200e+04	62772.000000	62772.000000	62772.000000	62772.000000	62772.000000	62772.000000	62772.000000	62772.000000
mean	1.180822e+06	31.800819	3.150401	7.471155	12.505510	23.515849	39.424250	15.908402	0.394379
std	5.859164e+05	22.979233	2.262465	1.458739	1.439215	17.767456	28.888168	14.014507	0.139011
min	3.010000e+02	0.000000	1.000000	5.000000	10.000000	5.000000	10.000000	0.080000	0.010000
25%	6.808580e+05	8.000000	1.000000	6.160000	11.300000	9.320000	14.300000	5.930000	0.300000
50%	1.261289e+06	37.000000	2.000000	7.445000	12.480000	17.980000	29.720000	11.280000	0.400000
75%	1.686269e+06	53.000000	4.000000	8.750000	13.780000	31.600000	54.085000	21.320000	0.510000
max	2.099937e+06	66.000000	15.000000	10.000000	15.000000	144.300000	187.200000	100.000000	0.670000

Numerical Characteristics:
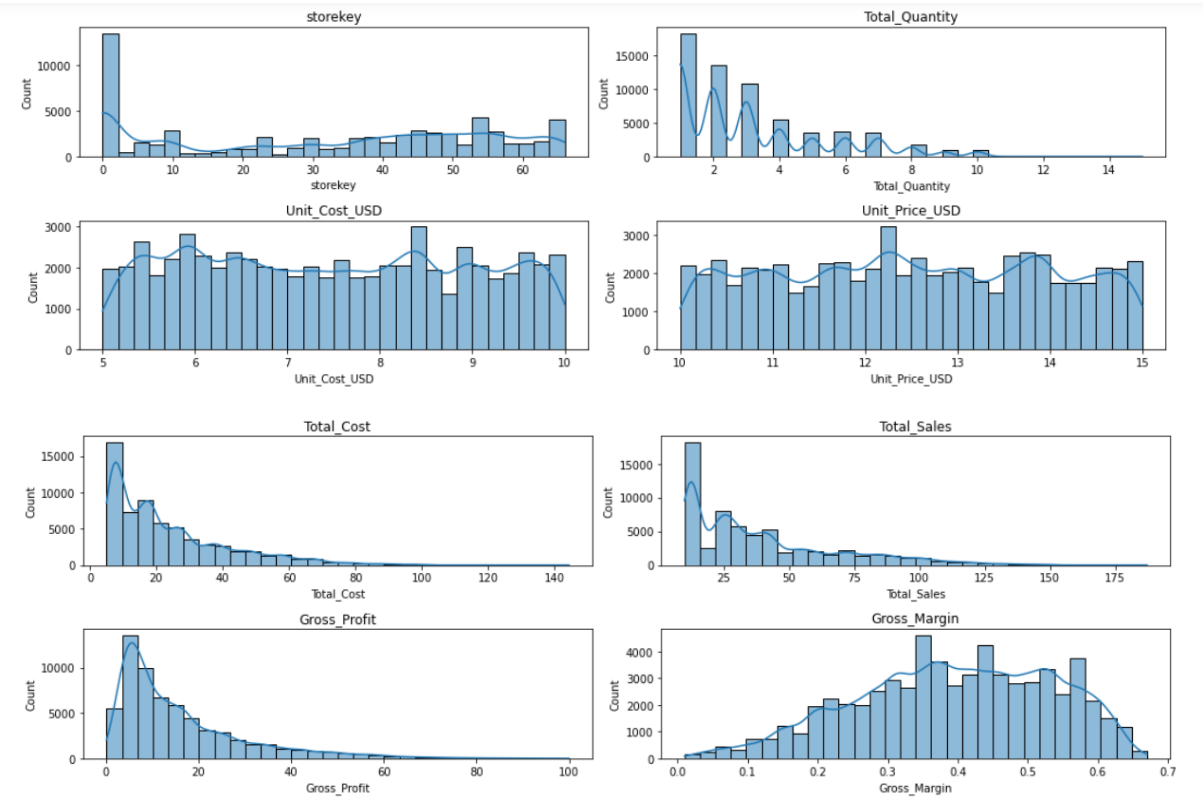
Summary Statistics Insights:

- Total Quantity: Total quantity count is more for 1, 2 and 3 quantities, which tells us most of the customers are interested in placing single quantity orders, less bulk orders

- Total cost and Total sales : The max values(144.3,187.2) which are higher than the mean(23.51,39.42) this may be due to bulk ordering of some products and this is also reason for max Gross_Profit in some orders

Categorical Columns characteristics:
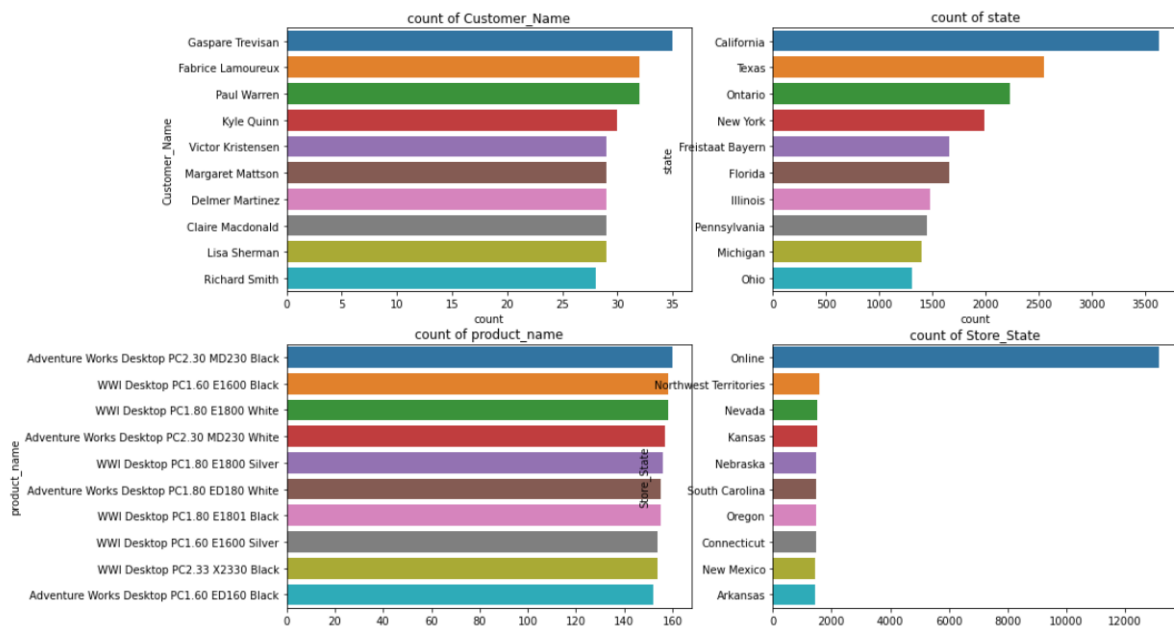

Correlation Matrix:
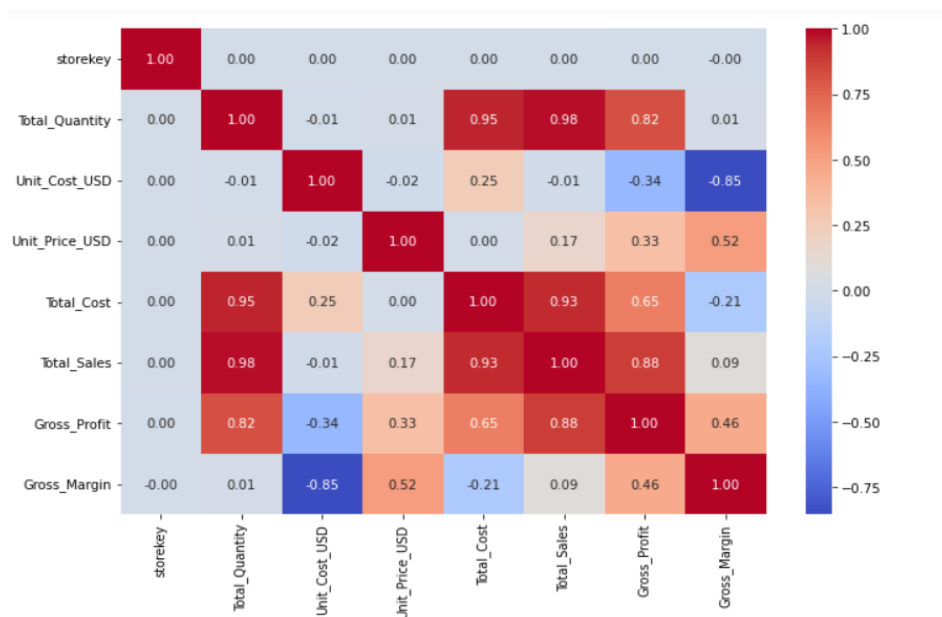
Correlation Insights:

- Unit_Cost_USD has strong negative correlation with Gross_Margin(-0.85) and Gross_Profit(-0.34) suggests costprice per unit increses, Gross_Profit and Gross_Margin decreses, possibly due to more discoutns on the product

- Gross_Profit has strong positive correlation with Total_Quantity, confirming bulk orders reults high profits

- Unit_Price_USD has positve correlation with Gross_Margin(0.52), suggests Unit_Price_USD increses, Gross_Margin increses, if we change prices in an optimal way results in increse in Gross_Margin

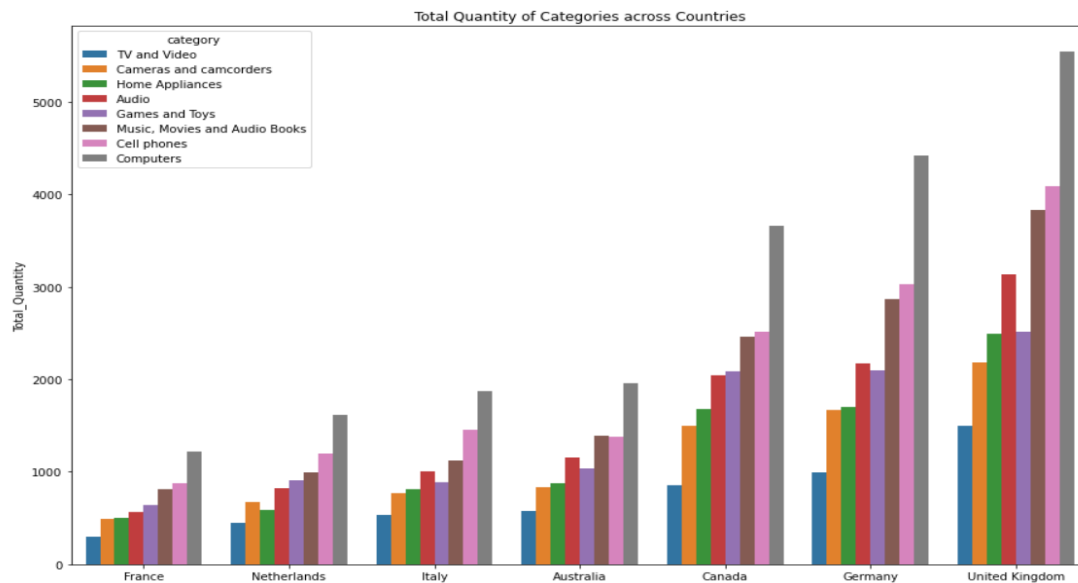**Data Analysis:**

States that generate the highest and lowest gross profits
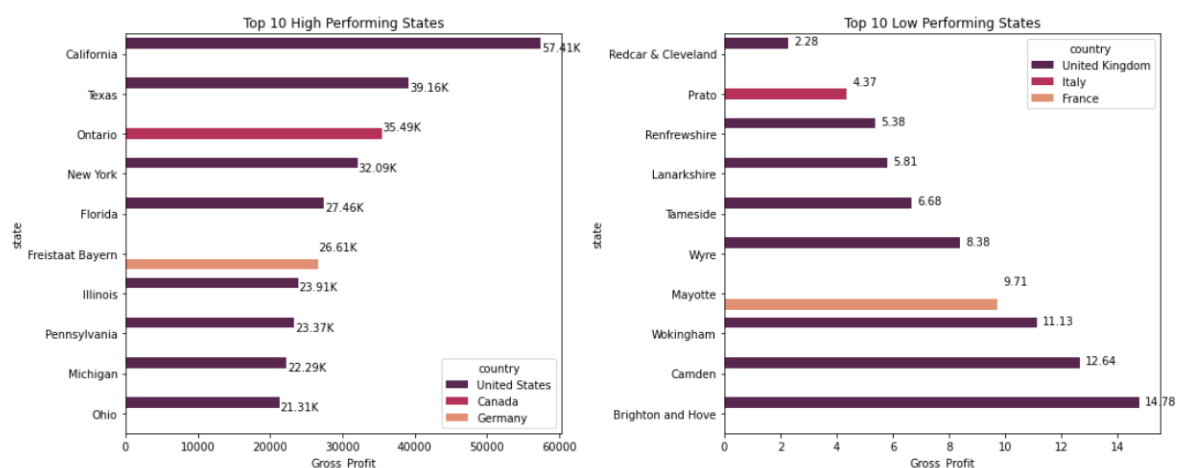
Product categories which show the highest price sensitivity across regions

Total Quantity of Categories across Countries
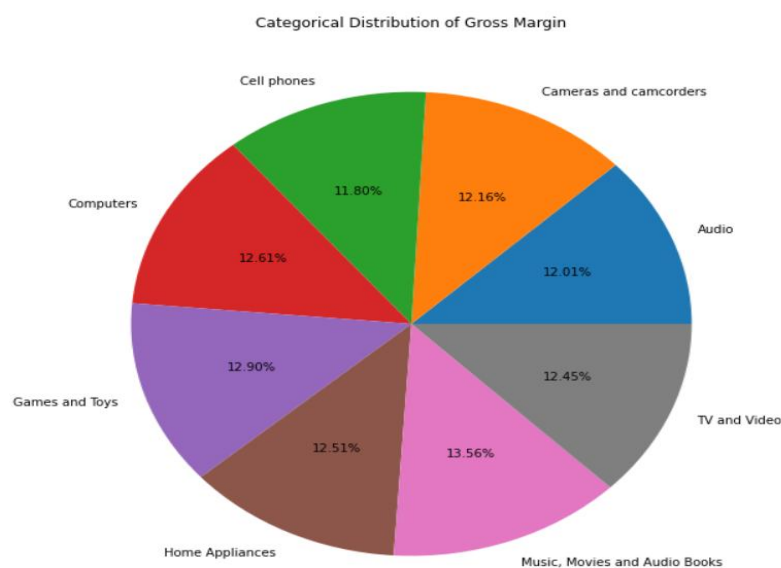
We observe product categories performing differently in each country

- Most of the countries perofroms bad in 3 categories TV and Video, cameras and camcoders,home appliances

- Top performing categories across all countries Computers, cell phones,music movies and audio books

- Audio ,Games and Toys performs avg in all countries

- According to the bar plot price sensitive categories are TV and Video, cameras and camcoders,home appliances

Product categories that consistently underperform in certain regions despite strong performance elsewhere
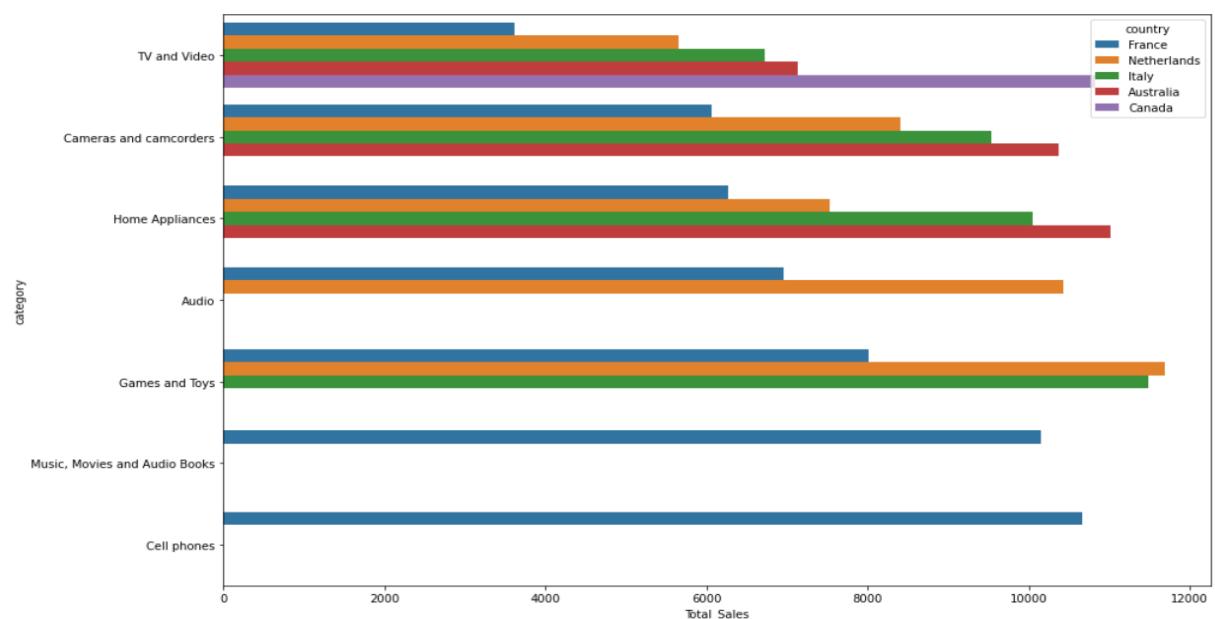
As we observe low performing product categories are performing differently in each country

- TV and Video category is performing bad in France and good in Canada

- Cameras and Camcoders are performing bad in France and good in Italy and Australia

- Home appliances are bad in France and good in Italy and Australia

According to the performance in different countries we can update the price of products

Top performing products in sales but low profits

Products with high sales and low profits
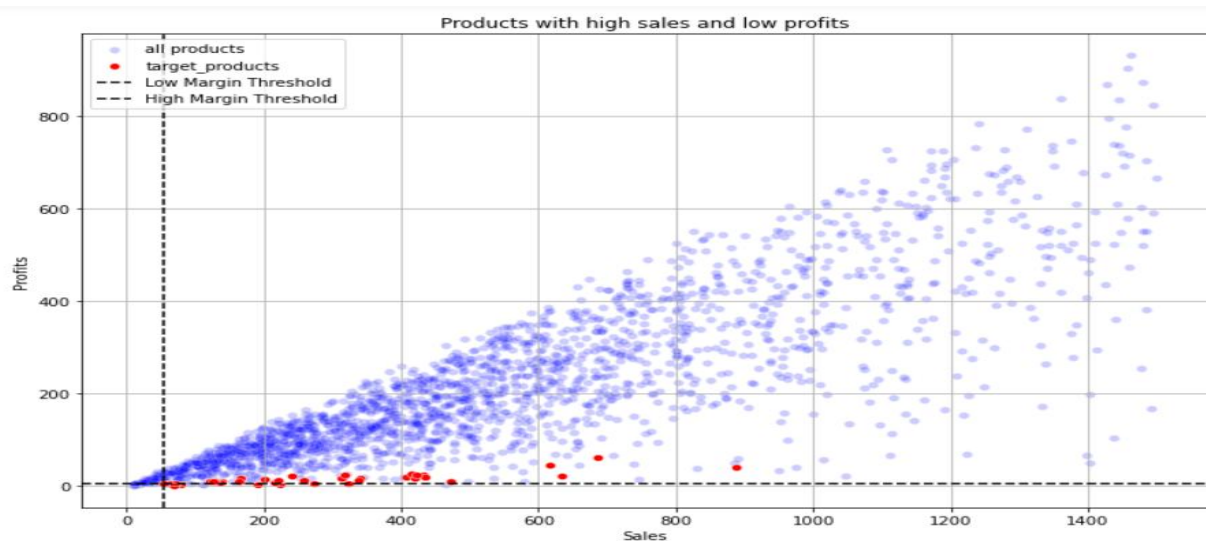
- **39** target products that performed good in US(59.77%), UK(11.77%) and Germany(8.29%) and bad in other countries

- The target products are the products with high demand but with low profits, if we increase price as the products is in demand they won't get saturated even if prices were increased

- But we have to do update prices according to regions, we have to observe these target products were performing good in all regions or in some particular regions

Sales distribution acorss different countries for top performing products
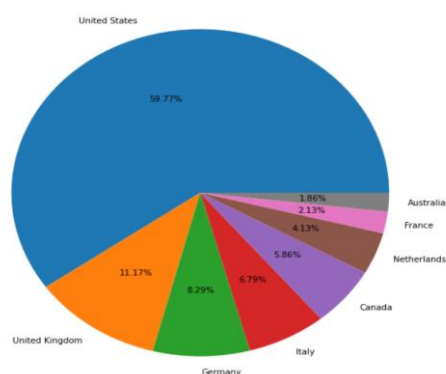
- According to the target products distribution we have to focus more on US, UK, Germany and Italy, this places give more profits to us if we update our price optimally, especially US (59.77%)

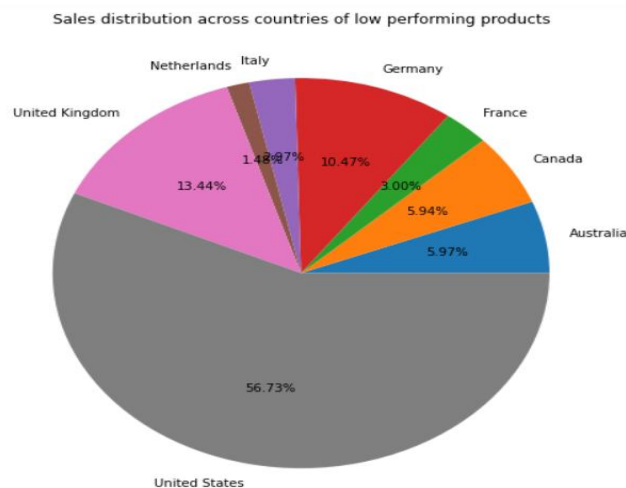- Since other countries yield lower profits, we should first consider increasing the price. If this adjustment negatively impacts the predicted profits, we can either apply a small incremental increase or keep the price unchanged.

Top performing customers according to sales and how much they contribute towards total sales

Total Sales Contribution by Top 10% of Customers

- Netherlands
- Italy
- United Kingdom 13.23%
- 1.72% 4.04%
- Germany 9.37%
- France 1.29%
- Canada 5.73%
- Australia 1.82%
- **Total Contribution : 28.23%**
- 62.79%
- United States
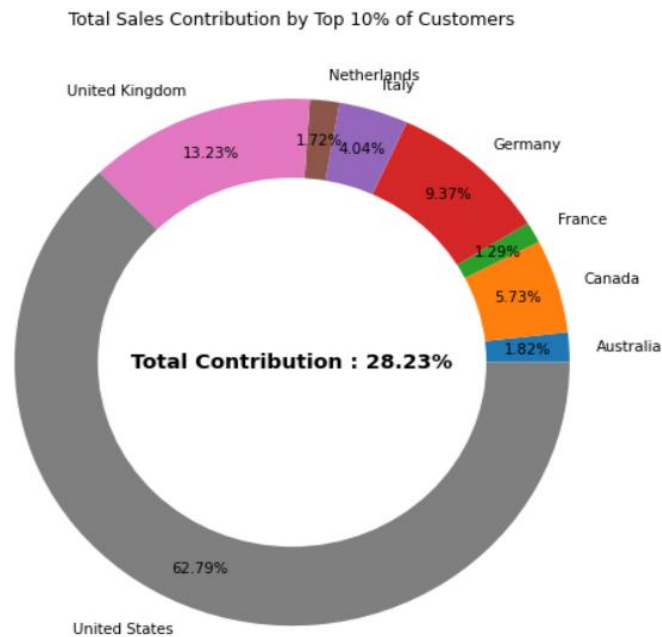
- Target customers are more from US, UK, Germany ,Canada here we can focus on updating to best price to get more profit margin

Products with low sales and high profits

Sales distribution across countries of low performing products

- Netherlands
- Italy
- Germany 10.47%
- United Kingdom 13.44%
- France 3.00%
- 1.48% ... 7%
- Canada 5.94%
- Australia 5.97%
- 56.73%
- United States

- **27** unique products that performed low in multiple countries

- Netherlands, Italy, France, Canada and Australia are the countries who were not showing interest to buy low performing products, possibly due to high prices, suggests us that reduce of price may increase the sales in that countries

- Even though prices are high US, UK, Germany were giving more sales, possibly due to there interest towards quality of the products irrespective of its price, suggest to keep the price constant, may increase further there will be no change in the sales from these countries

Top and low performing states in each country according to sales

```python
high_performing_states = (df.groupby("country", group_keys=False).apply(lambda x: x.nlargest(10, "Total_Sales")))

high_performing_states=high_performing_states[['state','Store_Country','Total_Sales',
                                               'Total_Quantity','Gross_Profit','Gross_Margin']]

high_performing_states.head()
```

	state	Store_Country	Total_Sales	Total_Quantity	Gross_Profit	Gross_Margin
61600	New South Wales	Australia	149.3	10.0	64.8	0.43
59771	Victoria	Australia	147.8	10.0	92.6	0.63
56620	Queensland	Australia	145.4	10.0	76.4	0.53
54164	New South Wales	Australia	143.2	10.0	46.5	0.32
53583	New South Wales	Australia	142.9	10.0	54.0	0.38

```python
low_performing_states = (df.groupby("country", group_keys=False).apply(lambda x: x.nsmallest(10, "Total_Sales")))

low_performing_states=low_performing_states[['state','Store_Country','Total_Sales',
                                             'Total_Quantity','Gross_Profit','Gross_Margin']]

low_performing_states.head()
```

	state	Store_Country	Total_Sales	Total_Quantity	Gross_Profit	Gross_Margin
2	Queensland	Australia	10.00	1.0	0.94	0.09
100	New South Wales	Australia	10.01	1.0	1.12	0.11
163	New South Wales	Australia	10.02	1.0	2.09	0.21
344	New South Wales	Australia	10.04	1.0	3.64	0.36
390	New South Wales	Online	10.05	1.0	1.25	0.12

Is there a significant difference between top performing and low performing states?

Hypothesis:

$H0$ (Null Hypothesis): There is no significant difference in Gross Profits of top performing and low performing states.

$H1$ (Alternative Hypothesis): The Gross Profits of top performing and low performing states are significantly different.

```python
top_states=high_performing_states['Gross_Profit']
low_states=low_performing_states['Gross_Profit']


t_stat, p_value=ttest_ind(top_states,low_states,equal_var=False)

print(f'T_Statistics:T-stat: {t_stat:.4f}, P_value: {p_value:.4f}')

if p_value<0.05:
    print("Reject H0: The Gross Profits of top performing and low performing states are significantly different.")
else:
    print('Accept H1: There is no significant difference in Gross Profits of top performing and low performing states.')
```

```
T_Statistics:T-stat: 36.5294, P_value: 0.0000
Reject H0: The Gross Profits of top performing and low performing states are significantly different.
```

7. Recommendations:

Pricing Strategy:

- Increase prices in the US, UK, and Germany to boost margins without affecting sales.

- Raise prices on high-selling, low-profit products in top-performing states of the US, UK, Germany, and Canada.

- Lower prices on low-selling, high-profit products in underperforming regions like the Netherlands, Italy, France, and Australia.

- Increase prices for strong categories: computers, cell phones, music, movies, and audiobooks across all regions

- Reduce prices in price-sensitive categories in low-performing countries; keep prices stable in top-performing ones like US,UK and Germany

Regional Strategy:

- Expand store presence in the Netherlands, Italy, France, and Australia.

- Launch targeted marketing and discount campaigns in these regions to drive sales.

Customer Strategy:

- Offer premium services and enhanced security to top-contributing customers.

- Introduce seasonal offers to attract new customers and boost engagement.

8. Price Adjustment Strategy

- To validate my strategy, I conducted a Shadow Price Back-test on identified **66** products to ensure our price changes are safe and profitable. I simulated a **5%** price adjustment and used Elasticity to predict how much the sales volume would naturally drop. I then checked if this new volume stayed within the Confidence Interval (2 Standard Deviations) of historical store performance. This confirms that our plan is realistic and won't cause an extreme sales crash.
- Through back-testing and elasticity modeling, this analysis identified a dual-path pricing strategy.
- Out of 39 products only 1 product is not between Confidence Interval(Outlier) and other 38 products are safe for price adjustments. For **38** high-volume products, a 5% price optimization was validated, and for 1 outlier product used old profit(Gross Profit), yielding a projected profit increase of $1365.
- For **27** high-margin products, back-testing across multiple price-drop scenarios (1%-5%) confirmed that these items are price-insensitive. Consequently, the recommendation is to hold current pricing to protect margins, as volume stimulation via price-cuts is not mathematically viable.
- Total Project Impact: **$1365** in immediate bottom-line growth and prevention of margin erosion on low sales high profit products.

9. Appendix:

Dashboard Screenshot: