Data Spark: Illuminating Insights for Global Electronics

Project Deliverables:

Data Cleaning and Preparation: Ensure all datasets are clean, integrated, and ready for analysis. Exploratory Data Analysis (EDA): Perform EDA to uncover trends, patterns, and insights.

Visualizations: Create visualizations to effectively communicate key findings.

Report: Summarize the analysis and provide actionable recommendations for Global Electronics.

Approach:

1. Data Cleaning and Preparation
2. Load Data
3. Power BI Visualization
4. Develop 10 SQL Queries.

Datasets Description:

Customers: Demographic and location data.

Sales: Transactional data including order and delivery details. Products: Product attributes such as cost, price, and category. Stores: Store locations, sizes, and operational dates.

Exchange Rates: Historical exchange rate data.

Approach:

1. Data Cleaning and Preparation
2. Load Data
3. Power BI Visualization
4. Develop 10 SQL Queries.

Datasets Description:

Table	Field	Description
Sales	Order Number	Unique ID for each order
Sales	Line Item	Identifies individual products purchased as part of an order
Sales	Order Date	Date the order was placed
Sales	Delivery Date	Date the order was delivered
Sales	Customer Key	Unique key identifying which customer placed the order
Sales	Store Key	Unique key identifying which store processed the order
Sales	Product Key	Unique key identifying which product was purchased
Sales	Quantity	Number of items purchased
Sales	Currency Code	Currency used to process the order
Customers	Customer Key	Primary key to identify customers

Customers	Gender	Customer gender
Customers	Name	Customer full name
Customers	City	Customer city
Customers	State Code	Customer state (abbreviated)
Customers	State	Customer state (full)
Customers	Zip Code	Customer zip code
Customers	Country	Customer country
Customers	Continent	Customer continent
Customers	Birthday	Customer date of birth
Products	Product Key	Primary key to identify products
Products	Product Name	Product name
Products	Brand	Product brand
Products	Color	Product color
Products	Unit Cost USD	Cost to produce the product in USD
Products	Unit Price USD	Product list price in USD
Products	Subcategory Key	Key to identify product subcategories
Products	Subcategory	Product subcategory name
Products	Category Key	Key to identify product categories
Products	Category	Product category name
Stores	Store Key	Primary key to identify stores

Customers	Gender	Customer gender
Customers	Name	Customer full name
Customers	City	Customer city
Customers	State Code	Customer state (abbreviated)
Customers	State	Customer state (full)
Customers	Zip Code	Customer zip code
Customers	Country	Customer country
Customers	Continent	Customer continent
Customers	Birthday	Customer date of birth
Products	Product Key	Primary key to identify products
Products	Product Name	Product name
Products	Brand	Product brand
Products	Color	Product color
Products	Unit Cost USD	Cost to produce the product in USD
Products	Unit Price USD	Product list price in USD
Products	Subcategory Key	Key to identify product subcategories
Products	Subcategory	Product subcategory name
Products	Category Key	Key to identify product categories
Products	Category	Product category name
Stores	Store Key	Primary key to identify stores

EDA:

Exploratory Data Analysis (EDA) is a critical step in the data analysis process, involving the investigation and summarization of datasets to uncover patterns, spot anomalies, test hypotheses, and check assumptions with the aid of summary statistics and graphical representations.

Understand the Data:

- Data types: Identify if the data is numerical, categorical, or textual.
- Structure: Check dimensions, column names, and data formats.
- Missing Values: Locate missing or incomplete data.

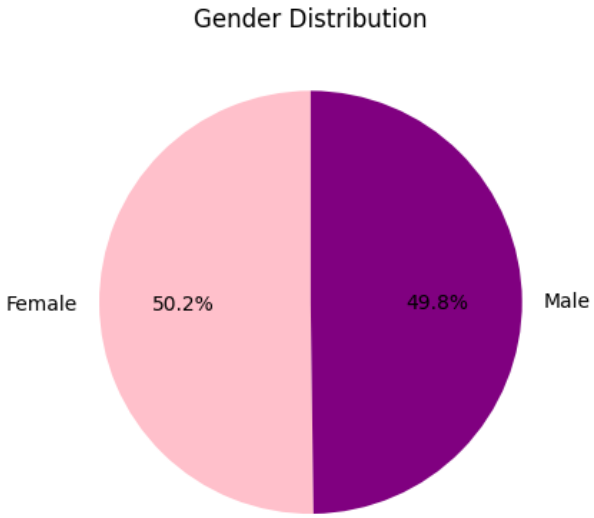
Data Cleaning:

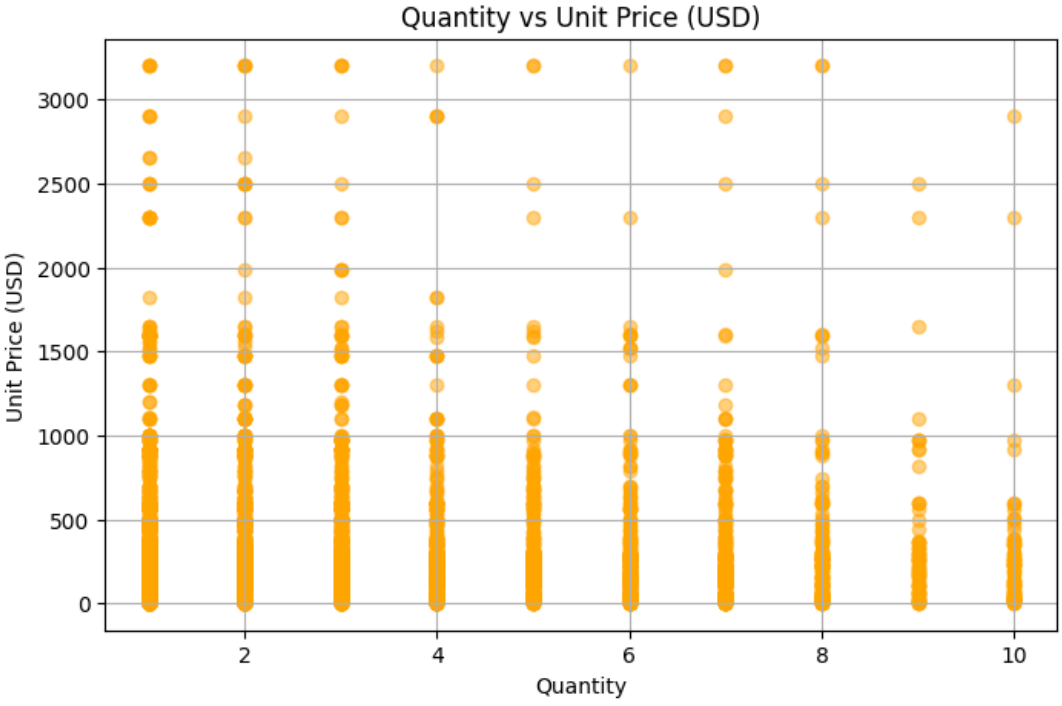
- Handle missing values (imputation, removal, or substitution).
- Address duplicate data.
- Correct data types or encoding issues.

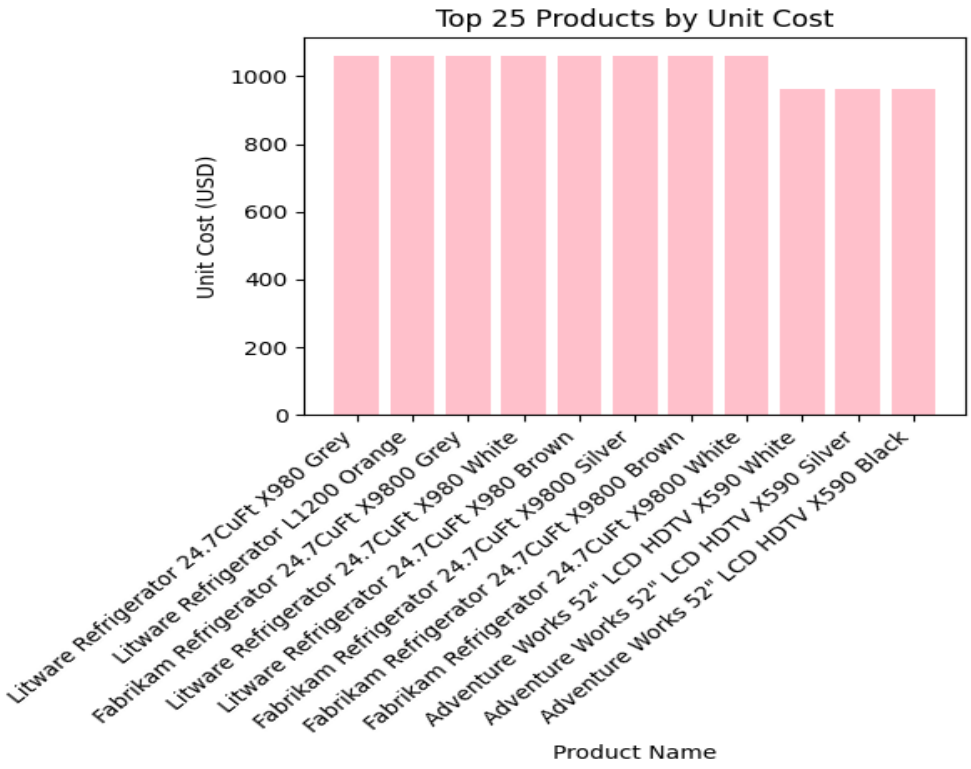
Data Visualization:

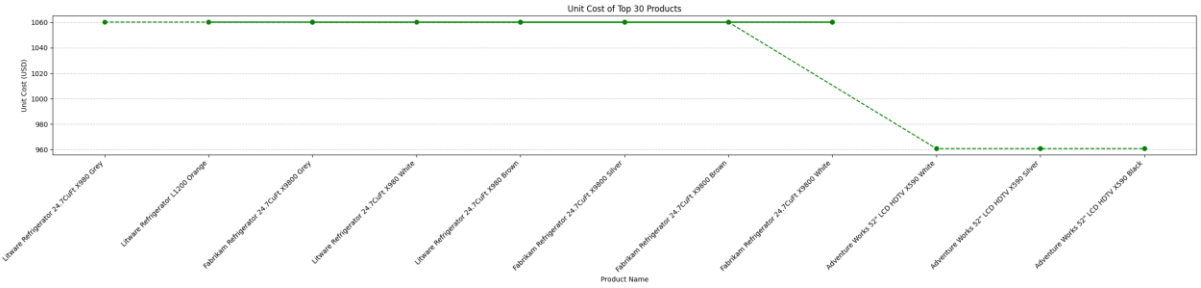
- Key tool for EDA.

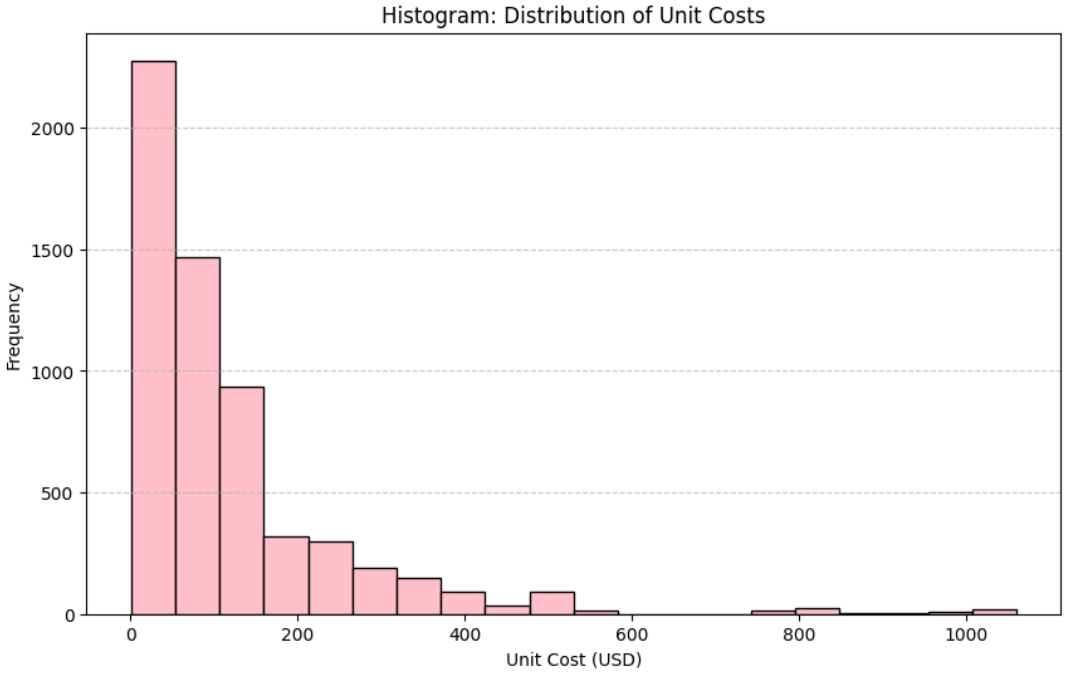
- Libraries like Matplotlib, Seaborn, Plotly, and others are commonly used for creating effective visuals.

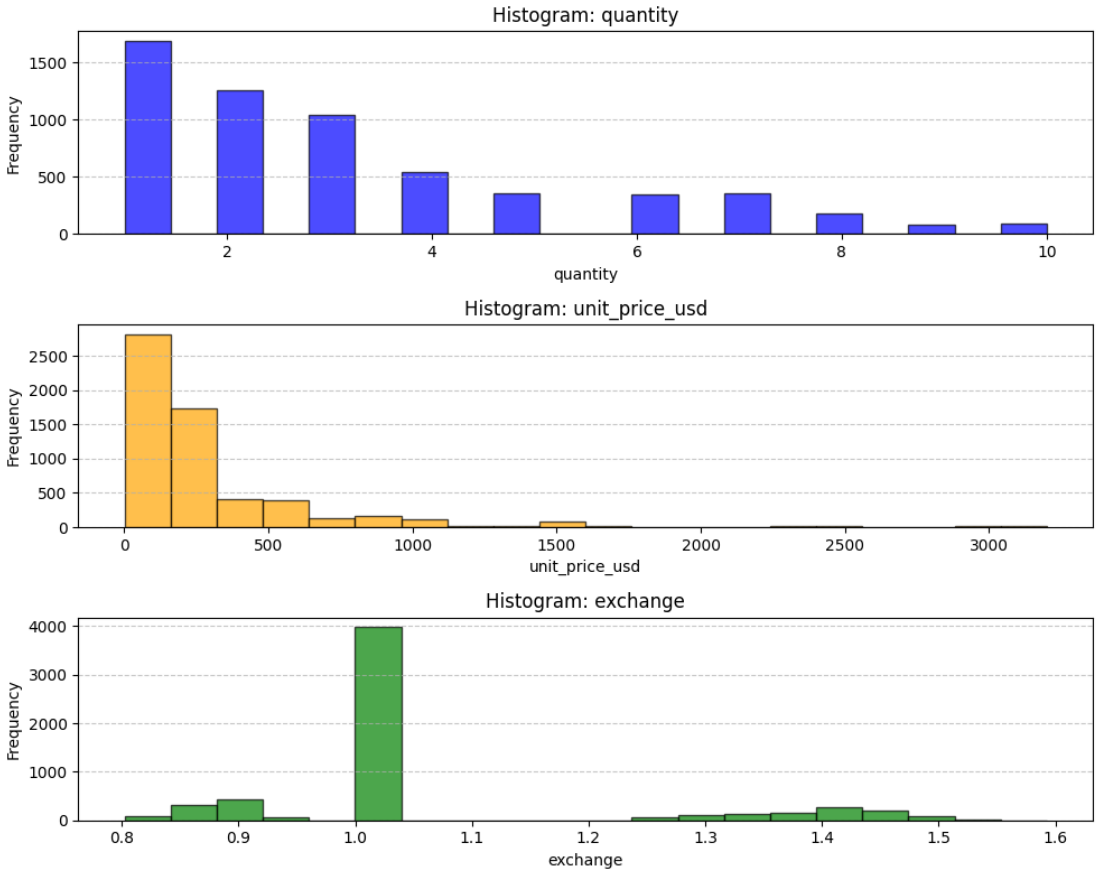





Top 25 Products by Unit Cost
Unit Cost (USD)
0
200
400
600
800
1000
Litware Refrigerator 24.7CuFt X980 Grey
Litware Refrigerator L1200 Orange
Fabrikam Refrigerator 24.7CuFt X9800 Grey
Litware Refrigerator 24.7CuFt X980 White
Litware Refrigerator 24.7CuFt X980 Brown
Litware Refrigerator 24.7CuFt X9800 Silver
Fabrikam Refrigerator 24.7CuFt X9800 Brown
Fabrikam Refrigerator 24.7CuFt X9800 White
Adventure Works 52" LCD HDTV X590 White
Adventure Works 52" LCD HDTV X590 Silver
Adventure Works 52" LCD HDTV X590 Black
Product Name


Unit Cost of Top 30 Products
Unit Cost (USD)
Product Name


Histogram: Distribution of Unit Costs
Frequency
0
500
1000
1500
2000
0
200
400
600
800
1000
Unit Cost (USD)



CUSTOMER ANALYSIS:

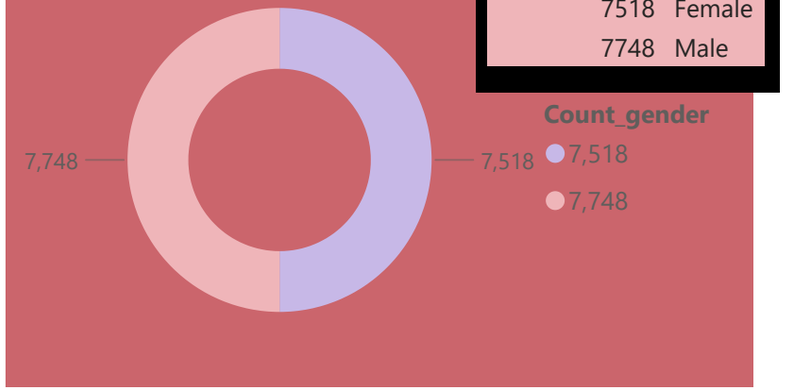
PRODUCT ANALYSIS:

SALES ANALYSIS:

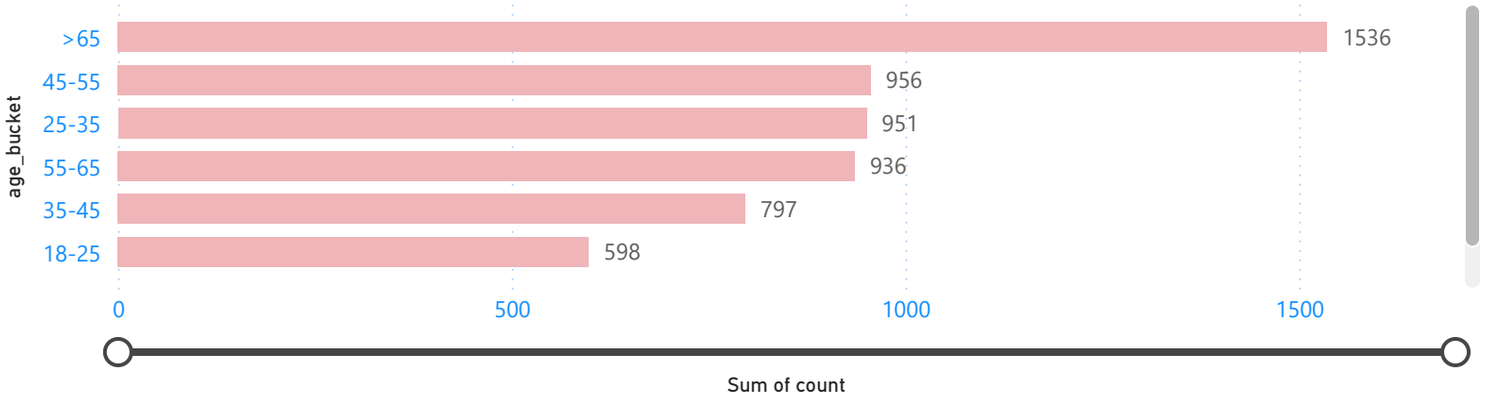
STORE ANALYSIS:

Count of gender by Count_gender

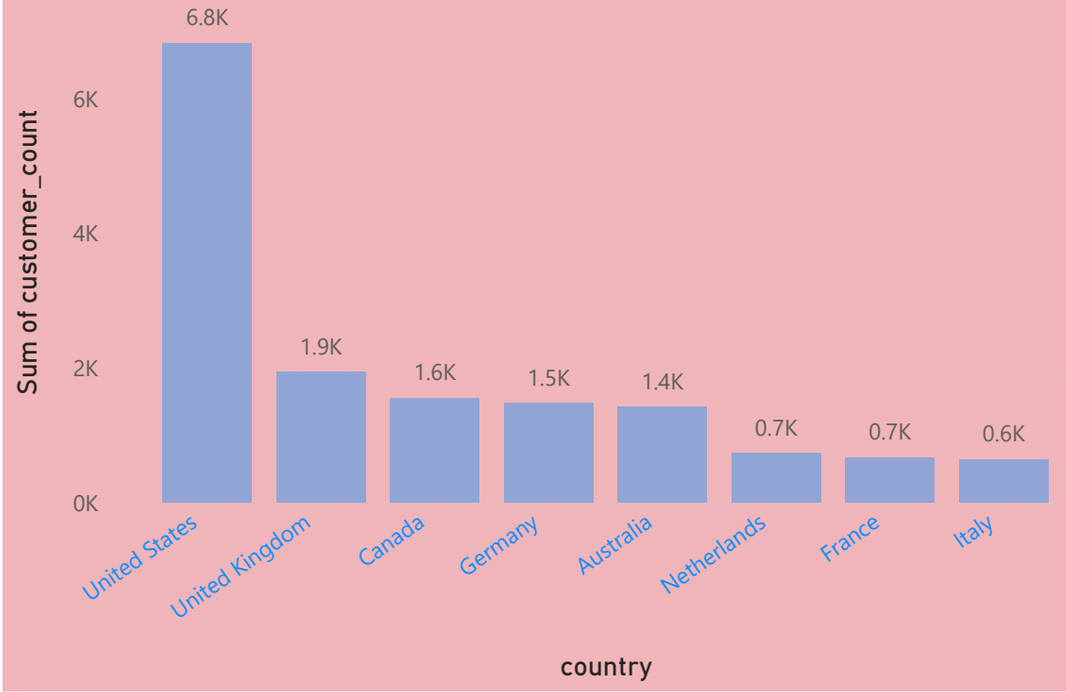
Count_gender	gender
7518	Female
7748	Male



Sum of count by age_bucket

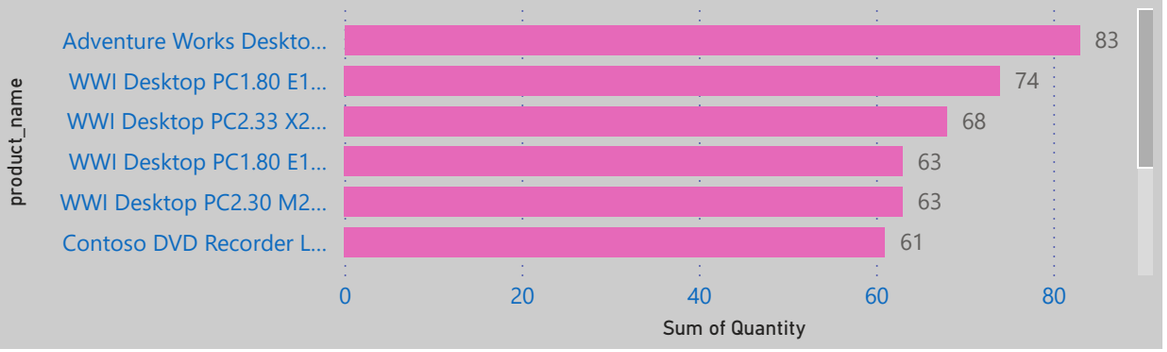


Sum of customer_count by country

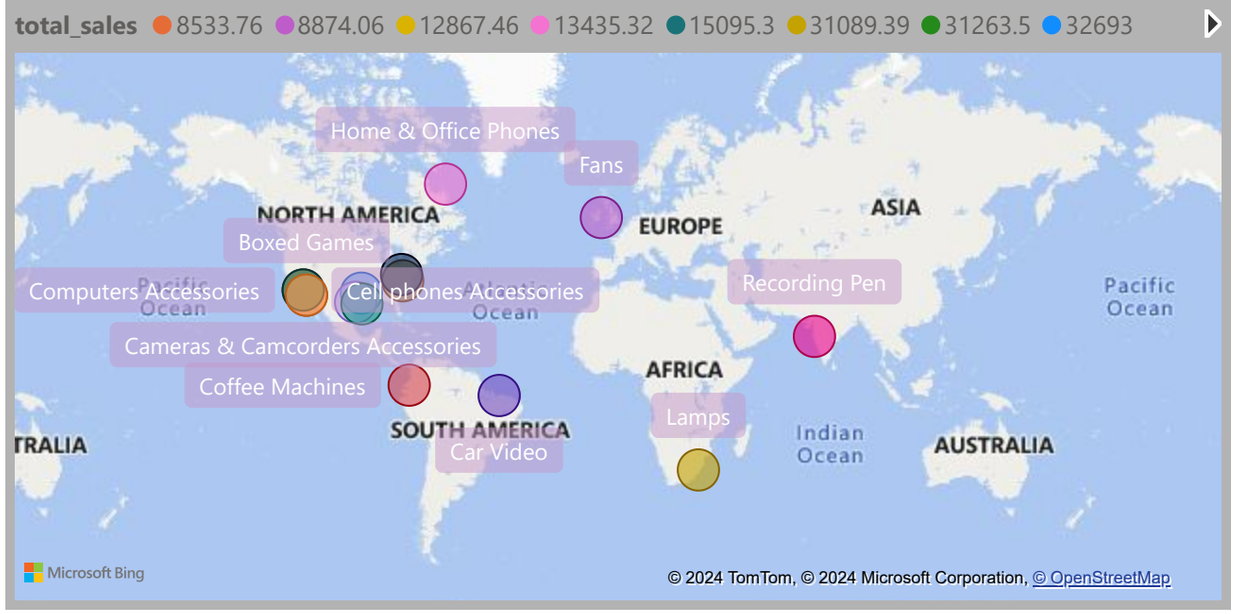


Sum of customer_count by continent

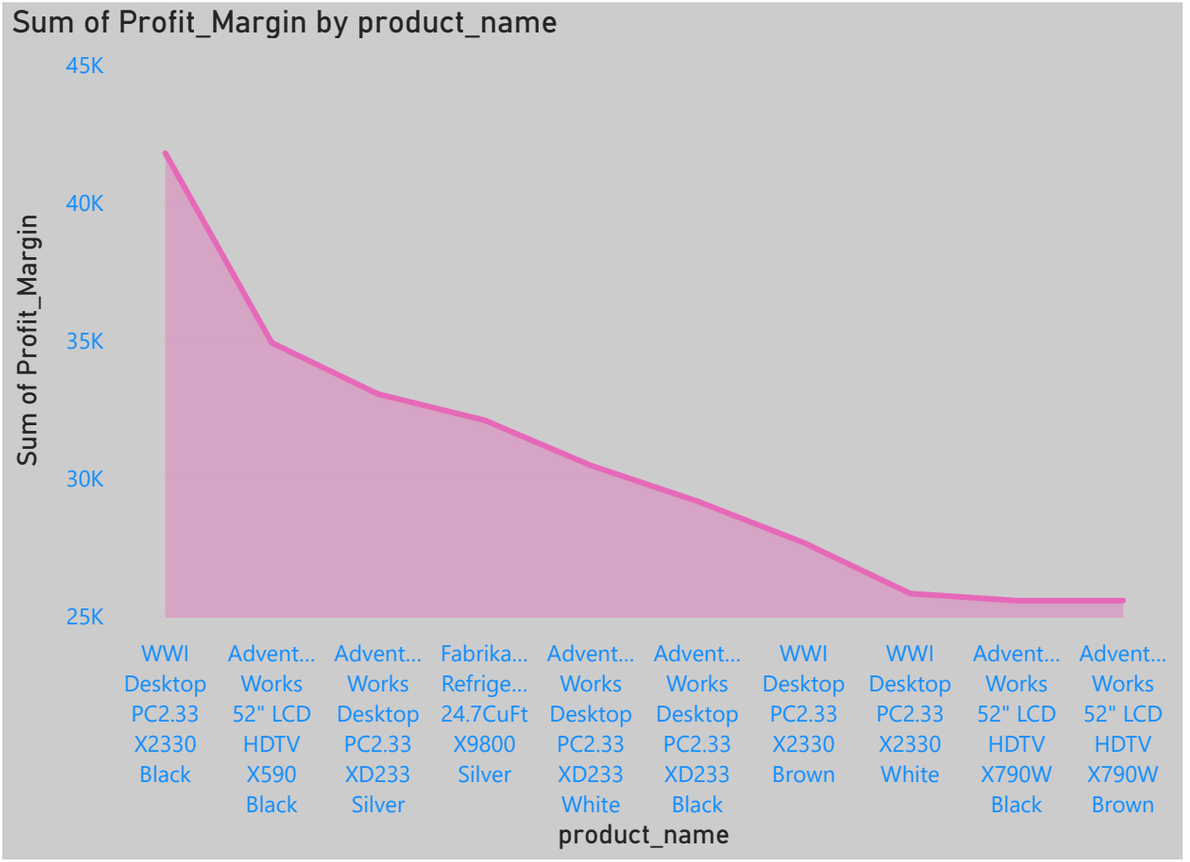
Sum of Quantity by product_name



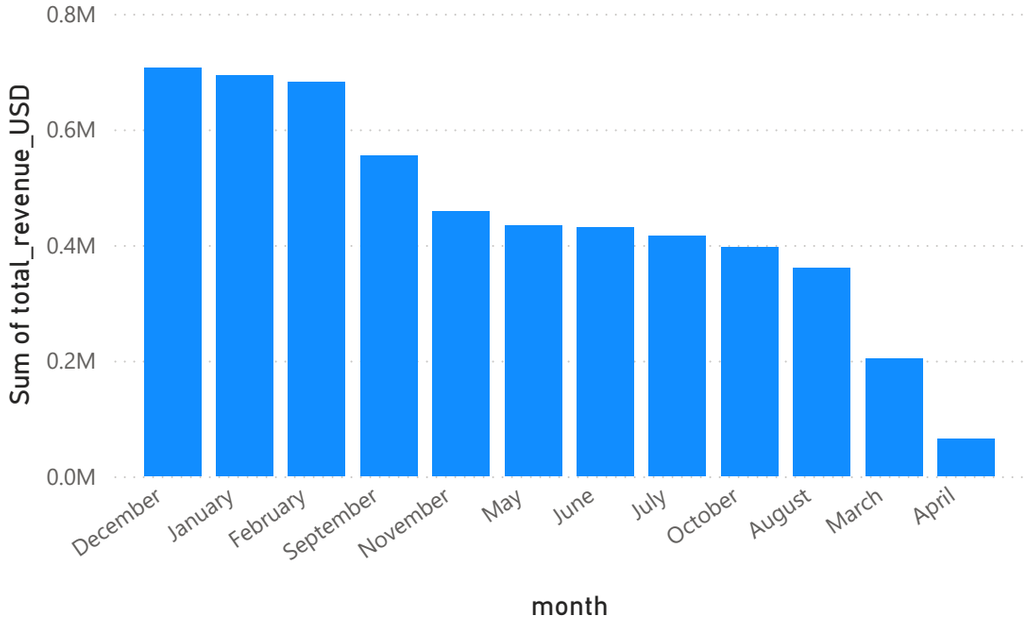
First category by subcategory and total_sales



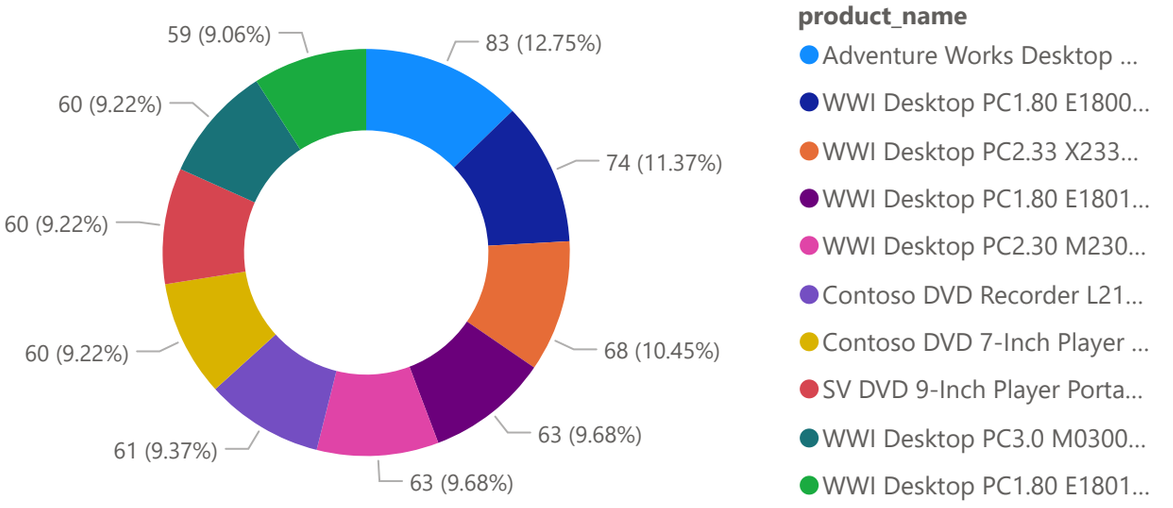
Sum of Profit_Margin by product_name



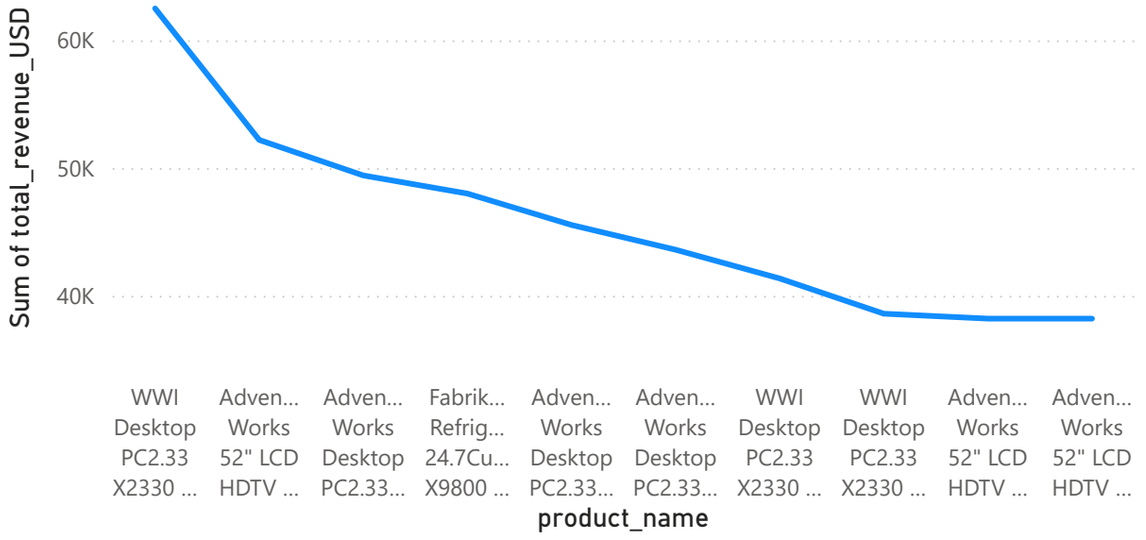
Sum of total_revenue_USD by month



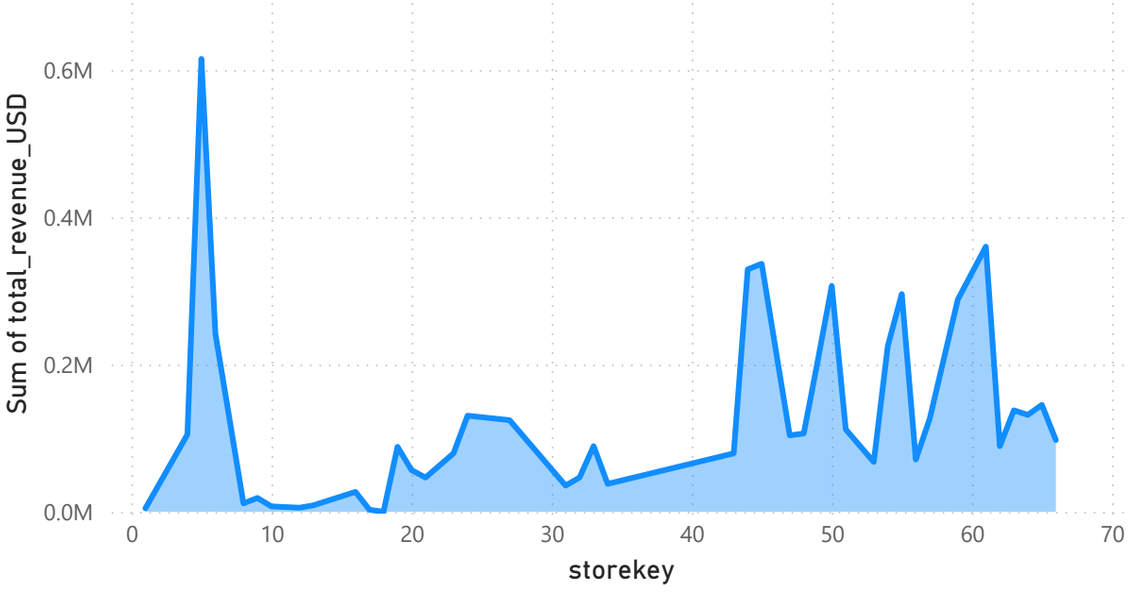
Sum of total_quantity by product_name



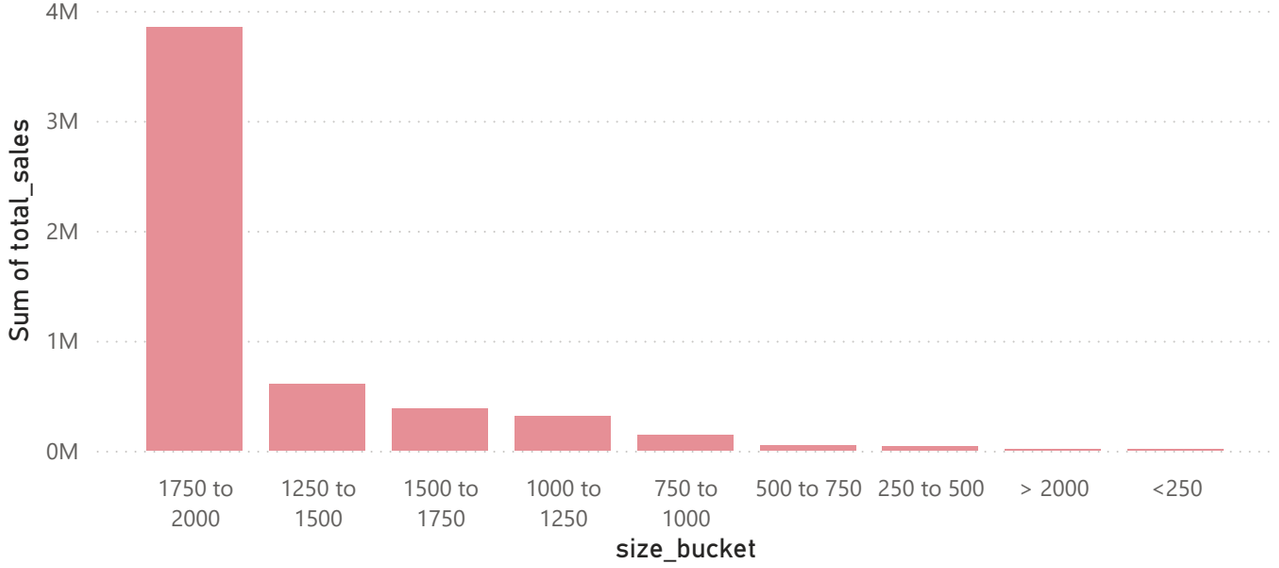
Sum of total_revenue_USD by product_name



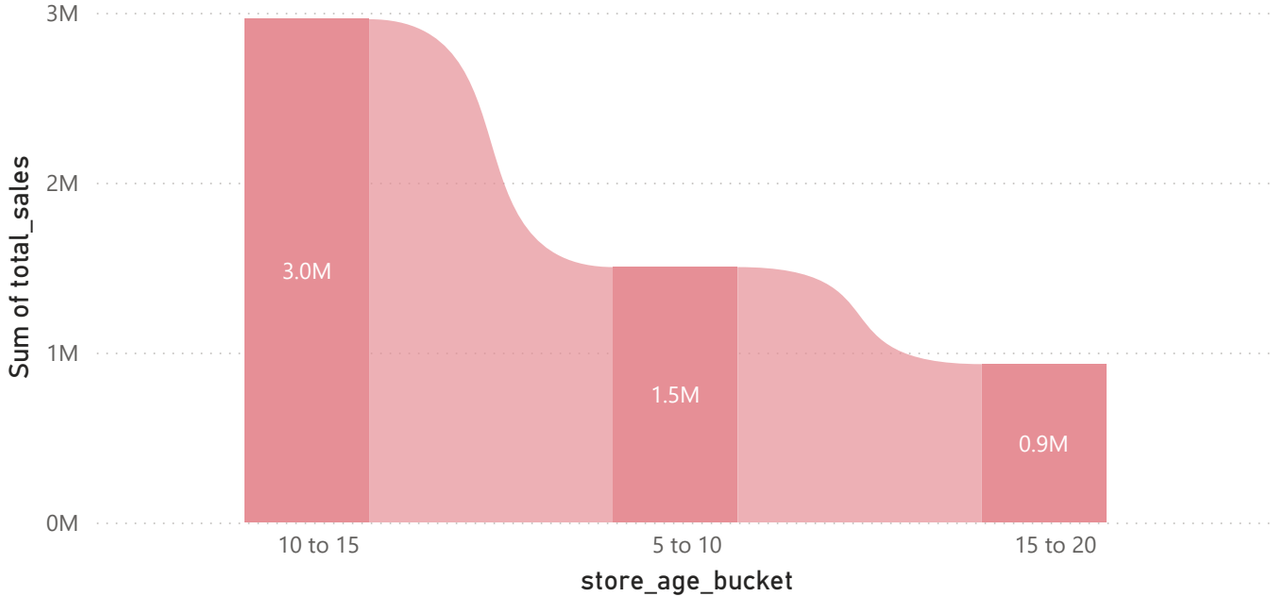
Sum of total_revenue_USD by storekey



Sum of total_sales by size_bucket



Sum of total_sales by store_age_bucket



Sum of storekey and Sum of total_revenue_USD by Continent and Country