UNIVERSITY OF WESTMINSTER⌗

INFORMATICS INSTITUTE OF TECHNOLOGY

5BUIS006C Data Visualisation and Communication

COURSEWORK

Module Leader:

Mr. Fouzul Hassan

Name: Kaushalya Jayakody

Blackboard Name: Jayakody Jayakody

IIT ID: 20232440

UOW ID: w2054878

Table of Contents

Research Question and Data Sourcing

Research Question

"How can the company maximize revenue by identifying top-performing product categories and regions?"

Knowing which product categories and regions generate more of the top line is crucial for business improvement. This research question focuses on drivers of revenue and offers valuable recommendations for improving the sales model. Using the sales data for individual products and regions, it is possible to optimise the marketing strategies, give more attention to the profitable areas, and consider ways to increase the sales of the less popular categories. These will enhance the decision-making process from the gathered information and data to result in improved overall profitability.

Data Source

The dataset "Analyzing Customer Spending Habits to Improve Sales Performance" is obtained from Kaggle. It contains transaction-level details like product type, regions, revenue and customer details. This comprehensive dataset makes it easier to analyse the sales pattern and identify the drivers of revenue in order to make strategic decisions.

Reference;

The Devastator (2022). *Analyzing Customer Spending Habits*. [online] Kaggle.com. Available at: https://www.kaggle.com/datasets/thedevastator/analyzing-customer-spending-habits-to-improve-sa [Accessed 31 Dec. 2024].

Data Preparation

First, the cleanliness and wellness of the dataset were checked. First, it was full of irrelevant columns (Column1 and index) and rows with missing values. Tidy data is measured by organising the data so that each variable is a column, each observation is a row, and each type of observational unit is a table. When examined, it can be observed that the dataset was not fully clean; it contained additional columns not used in the analysis.

Important variables like country, product categories, revenue, and sales are included in the dataset. These dimensions (country, categories) concerning a company are crucial for responding to the research question of the bestselling products and countries as well as measures (revenue, sales). In the process of the analysis, Revenue is used as the dependent variable because it captures the result impacted by other independent variables. The independent variables consist of Countries, Product Categories, Quantity, and year. Measurements in the current dataset refer to either single selling operations or average selling, depending on the level of data given in the current dataset.

Subsequently, the data was read and cleaned to increase its tidiness in the analysis using the R software. First, noise was eliminated by using the subset() function to omit unnecessary columns to conserve space and only select important fields. Any row containing null values was removed by this using the na.omit() function for cleaner datasets. This cleaned dataset was also saved for analysis with the same structure.

Finally, after applying transformations to the dataset above, the dataset was checked to ensure that it was tidy. Measured variables included all necessary information and did not seem to have any missing or duplicate values.

This process prepares the data for exploratory analysis, which is crucial for defining revenue trends and insights and enhancing the business's performance across regions and product categories.

Screenshots of the data preparation results

File Edit Code View Plots Session Build Debug Profile Tools Help

Coursework.R

```
1   # Load the dataset
2   file_path <- "SalesForCourse_quizz_table.csv"
3   data <- read.csv(file_path)
4
5   # View the first few rows of the dataset
6   head(data)
7
8   # View column names
9   colnames(data)
10
11  # Remove unnecessary columns
12  data_cleaned <- subset(data, select = -c(Column1, index))
13
14  # View the cleaned dataset
15  head(data_cleaned)
16
17  # Remove rows with null values
18  data_cleaned_new <- na.omit(data_cleaned)
19
20  # View the dataset after removing null values
21  head(data_cleaned_new)
22
23  #Save the new Cleaned Data
24  write.csv(data_cleaned_new, "SalesForCourse_quizz_table_cleaned.csv", row.names = FALSE)
25
```

Console   Terminal   Background Jobs

R R 4.4.1 · ~/UNI-2nd year/Data Visualisation and Dashboarding/Coursework/

```
> # Load the dataset
> file_path <- "SalesForCourse_quizz_table.csv"
> data <- read.csv(file_path)
> # View the first few rows of the dataset
> head(data)
  index    Date Year    Month Customer.Age Customer.Gender       Country      State Product.Category
1     0 02/19/16 2016 February           29                   F United States Washington       Accessories
2     1 02/20/16 2016 February           29                   F United States Washington          Clothing
3     2 02/27/16 2016 February           29                   F United States Washington       Accessories
4     3 03/12/16 2016    March           29                   F United States Washington       Accessories
5     4 03/12/16 2016    March           29                   F United States Washington       Accessories

     Sub.Category Quantity Unit.Cost Unit.Price Cost Revenue Column1
1 Tires and Tubes        1     80.00  109.00000   80     109      NA
2          Gloves        2     24.50   28.50000   49      57      NA
3 Tires and Tubes        3      3.67    5.00000   11      15      NA
4 Tires and Tubes        2     87.50  116.50000  175     233      NA
5 Tires and Tubes        3     35.00   41.66667  105     125      NA
6 Tires and Tubes        1     66.00   78.00000   66      78      NA
> # View column names
> colnames(data)
 [1] "index"           "Date"            "Year"            "Month"           "Customer.Age"
 [6] "Customer.Gender" "Country"         "State"           "Product.Category" "Sub.Category"
[11] "Quantity"        "Unit.Cost"       "Unit.Price"      "Cost"            "Revenue"
[16] "Column1"
> # Remove unnecessary columns
> data_cleaned <- subset(data, select = -c(Column1, index))
>
> # View the cleaned dataset
> head(data_cleaned)
      Date Year    Month Customer.Age Customer.Gender       Country      State Product.Category    Sub.Category
1 02/19/16 2016 February           29                   F United States Washington       Accessories Tires and Tubes
2 02/20/16 2016 February           29                   F United States Washington          Clothing          Gloves
3 02/27/16 2016 February           29                   F United States Washington       Accessories Tires and Tubes
4 03/12/16 2016    March           29                   F United States Washington       Accessories Tires and Tubes
5 03/12/16 2016    March           29                   F United States Washington       Accessories Tires and Tubes
6 04/08/16 2016    April           29                   F United States Washington       Accessories Tires and Tubes
  Quantity Unit.Cost Unit.Price Cost Revenue
1        1     80.00  109.00000   80     109
2        2     24.50   28.50000   49      57
3        3      3.67    5.00000   11      15
4        2     87.50  116.50000  175     233
5        3     35.00   41.66667  105     125
6        1     66.00   78.00000   66      78
```

```
> # Remove rows with null values
> data_cleaned_new <- na.omit(data_cleaned)
> # View the dataset after removing null values
> head(data_cleaned_new)
     Date Year    Month Customer.Age Customer.Gender       Country      State Product.Category    Sub.Category
1 02/19/16 2016 February           29               F United States Washington       Accessories Tires and Tubes
2 02/20/16 2016 February           29               F United States Washington          Clothing          Gloves
3 02/27/16 2016 February           29               F United States Washington       Accessories Tires and Tubes
4 03/12/16 2016    March           29               F United States Washington       Accessories Tires and Tubes
5 03/12/16 2016    March           29               F United States Washington       Accessories Tires and Tubes
6 04/08/16 2016    April           29               F United States Washington       Accessories Tires and Tubes
  Quantity Unit.Cost Unit.Price Cost Revenue
1        1     80.00  109.00000   80     109
2        2     24.50   28.50000   49      57
3        3      3.67    5.00000   11      15
4        2     87.50  116.50000  175     233

> #Save the new Cleaned Data
> write.csv(data_cleaned_new, "SalesForCourse_quizz_table_cleaned.csv", row.names = FALSE)
>
```

Environment	History	Connections	Tutorial

Import Dataset ▾ 146 MiB ▾ List ▾

R ▾ Global Environment ▾

Data

● data	34867 obs. of 16 variables
● data_cleaned	34867 obs. of 14 variables
● data_cleaned_new	34866 obs. of 14 variables

Values

file_path	"SalesForCourse_quizz_table.csv"

Exploratory Data Analysis

Univariate Analysis

- **Revenue Distribution**

  The analysis of the total revenue distribution shows that revenue is extremely concentrated, which means that significant sources of revenue highly depend on some particular product category or region. It is clear that many areas, or products, contribute a relatively small amount of revenue and that most of the revenue is generated from a few sectors.

  - **Insight:** This would indicate that companies have surged in revenue, and a single product category or region might dominate it.

Revenue Distribution
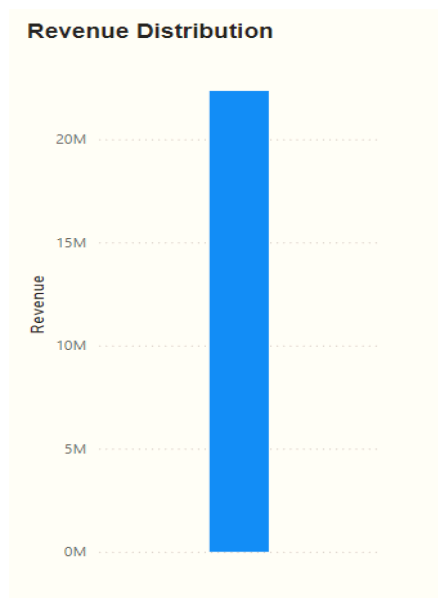
- **Revenue by Country**

  The United States contributes the highest revenue, substantially higher than other countries such as the United Kingdom, Germany, and France.

  - **Insight:** For instance, the United States could be the biggest market, emphasising regional dominance.

Revenue by Country
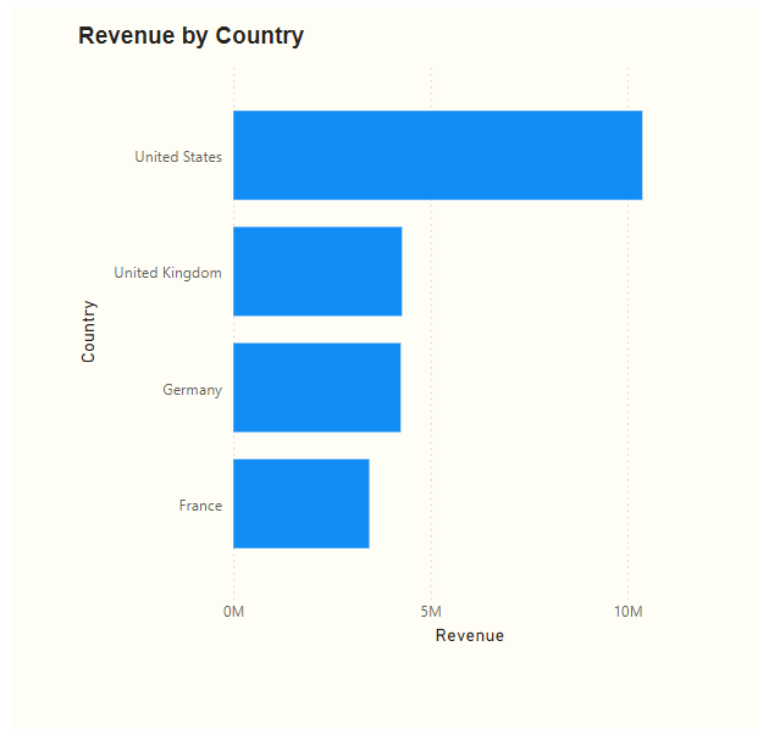
- **Revenue by Product Category**

  Among all product categories, "Bikes" contribute the highest revenue outperforming "Accessories" and "Clothing" significantly.

  - **Insight:** This could be attributed to higher unit pricing or greater demand for bikes, compared to other product categories.

Revenue by Product.Category
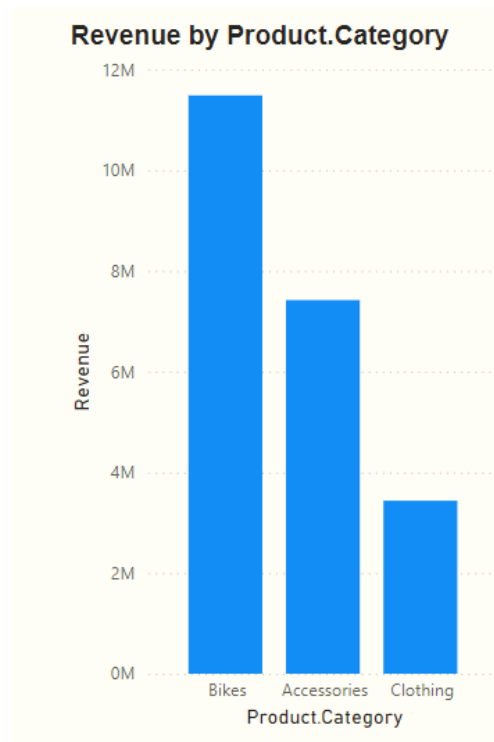
Some more Univariate Analysis visualisations;

Relationship Between Unit Price and Unit Cost
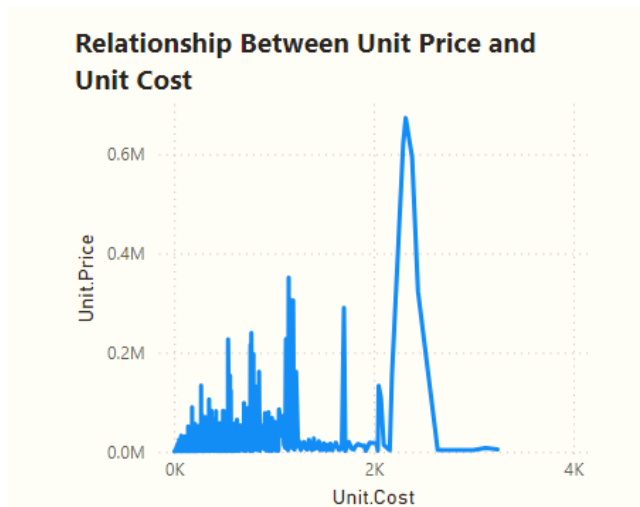
- This scatter plot shows a concentrated relationship where unit prices are significantly higher than unit costs, suggesting substantial profit margins.

Revenue Growth by Quantity Sold
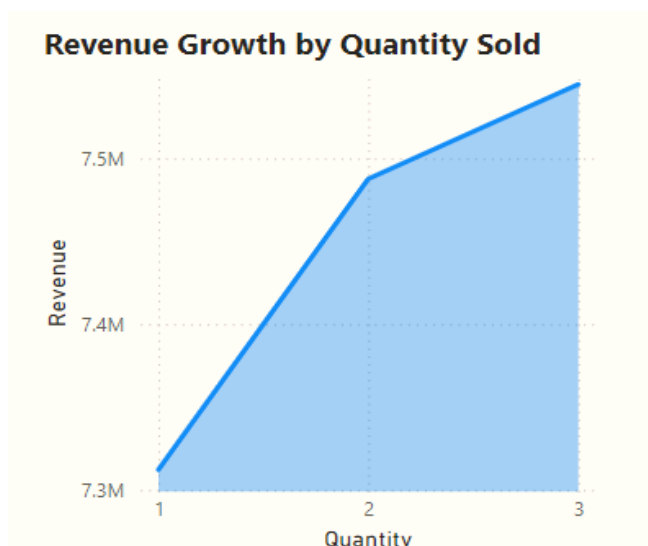
- The area chart highlights a positive correlation between the quantity sold and revenue, demonstrating increased revenue generation with higher sales volumes.

Customer Age Distribution by Gender
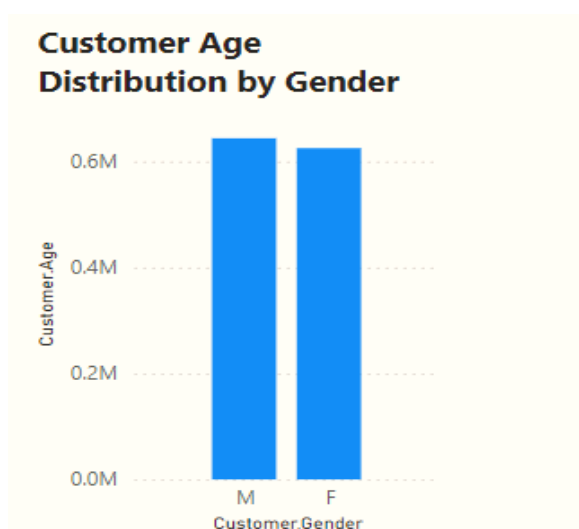
- The bar chart shows a nearly equal distribution of customer age between male and female demographics, indicating balanced engagement across genders.

Multivariate Analysis

- **Revenue by Product Category and Country**

  When analysing revenue by product category across countries, the United States consistently ranks highest in all categories, particularly for "Bikes" and "Accessories". Other countries show steady, albeit less remarkable, contributions, while "Clothing" consistently lags.

  - **Insight:** The U.S. market is vital for high-revenue categories like "Bikes" and "Accessories"; nevertheless, overall revenue from "Clothing" remains minimal across all countries.
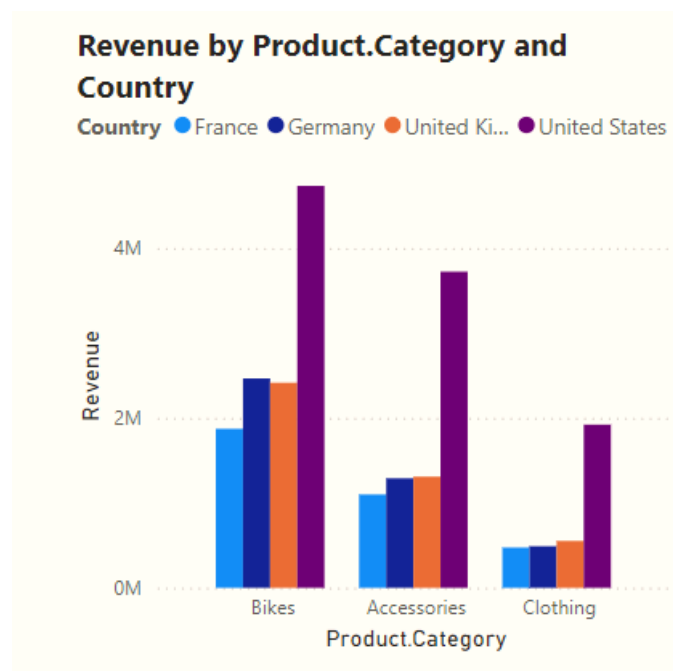
- **Revenue Vs Quantity**

  The scatter plot reveals that the product "Bikes" has high total revenue even if it has only a moderate quantity of sales, which means it is an expensive product. On the other hand, "Accessories" has reasonable revenue corresponding to the quantity of the same.

  - **Insights:** "Bikes" has a good pricing strategy, while "Accessories" contribute revenues through higher unit sales.

  - Clothing has the lowest revenue value and quantity which may refer to problems such as low consumption or market orientation.

Revenue Vs Quantity
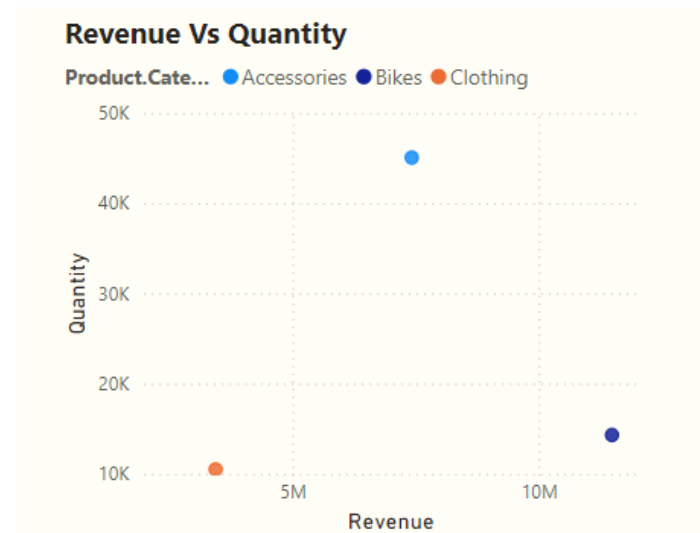
- **Revenue by Product Categories Across Countries**

  This stacked bar chart further validates the United States' top spot in terms of the revenue it generates across all products. A sample global stacked bar chart shows that the U.S retains the highest global marginal revenue for "Bikes" in every country

  - **Insight:** Market size and trends affect revenues in diverse territories based on customers' desired products and services.

Revenue by Product.Categories Across Countries
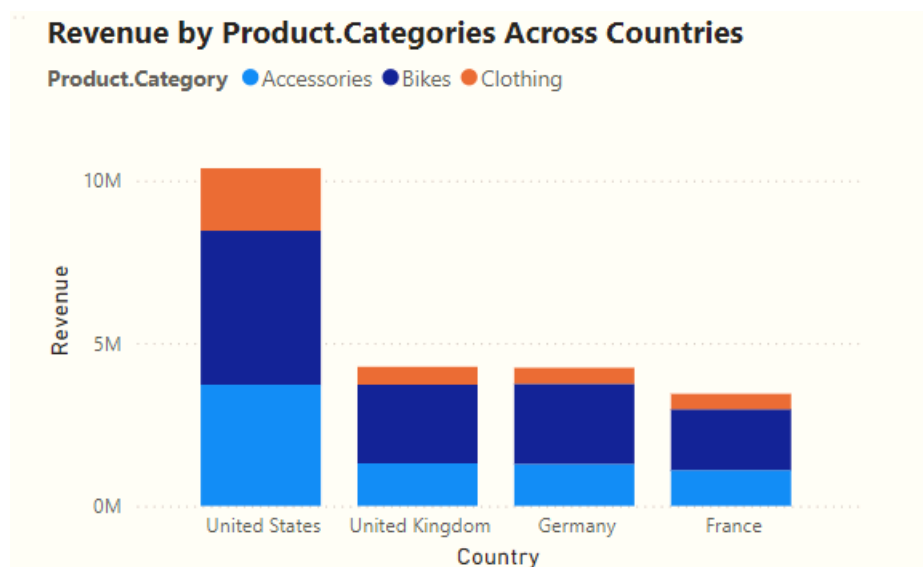
Some more Multivariate Analysis visualisations;

Unit Price vs. Quantity Sold by Product Category

Product.C... ●Accessories ●Bikes ●Clothing
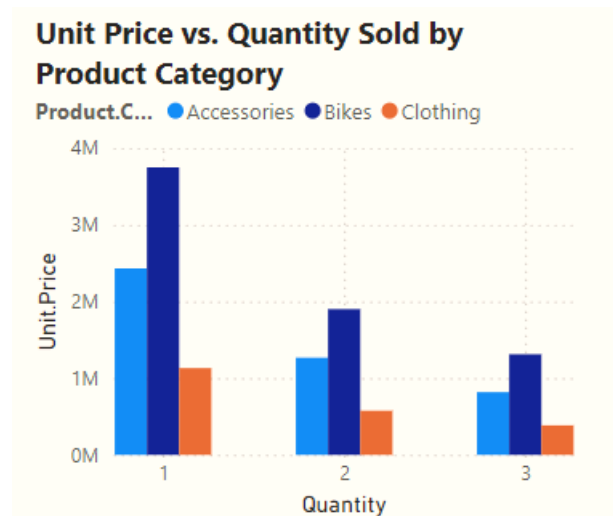
- This grouped bar chart reveals that Accessories generate the highest unit prices across different quantities sold, followed by Bikes and Clothing.

Revenue Share by Product Category

2.09M (15.42%)

Product.Category
●Bikes
●Accessories
●Clothing

4.52M (33.29%)

6.96M (51.3%)
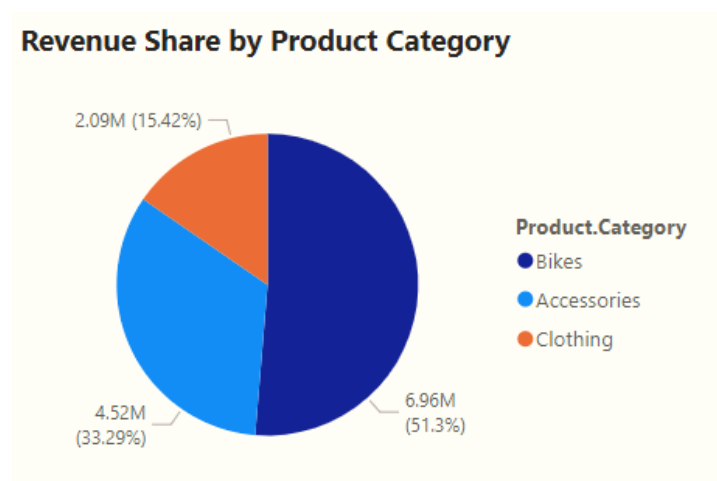
- The pie chart illustrates that Bikes contribute the highest share of total revenue (51.3%), followed by Accessories (33.29%) and Clothing (15.42%).

<u>Key Insights and Recommendations</u>

Key insights

- The United States is the most valuable market, generating the highest revenue across all product categories.
- Bikes are the leading revenue driver due to higher prices or demand, whereas "Clothing" struggles across all regions, contributing the least revenue.
- "Accessories" achieve balanced revenue through higher sales volume, suggesting broader appeal or affordability.

Recommendations

- **Expand "Bikes" Strategy:** Increase marketing and sales to "Bikes" targeting those countries where its market performance emulates the effectiveness of that in the United States.
- **Revamp "Clothing" Strategy:** Address poor sales of "Clothing" by exploring pricing adjustments, market repositioning, or promotional campaigns.
- **Diversify Revenue Streams:** Reduce over-reliance on dominant product categories or regions by investing in underperforming areas and exploring new markets.

Data Storytelling

The scene was set in a busy boardroom, where Elena Johnson, the chief executive officer of an international retail company, was meeting with her leadership team. However, she was always looking for ways to grow further and diversify the company even with these kinds of sustainable increases. "How can we maximise revenue by focusing on our best-performing products and regions?" she asked. The analytics team embarked on strategy formulation by using sales data from the United States, France, the United Kingdom, and Germany.

1. **Global Revenue Landscape**

The first action taken by the team was therefore to access overall performance by region. In a bar chart and a pie chart presenting the revenue, the differences between the four countries were evident.
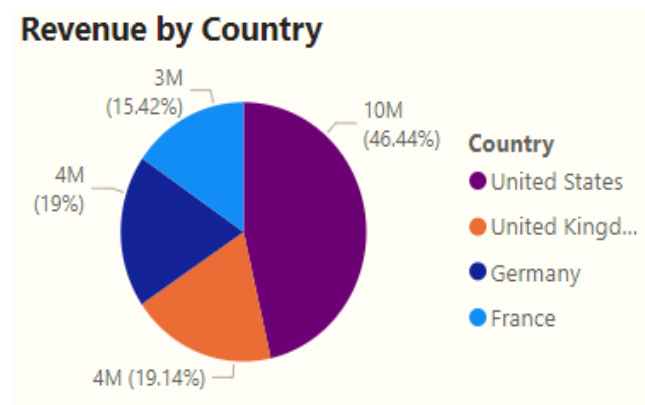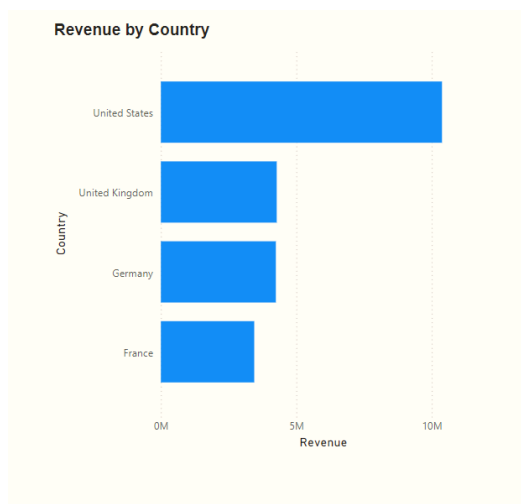
✓ **Key Insight:** America was most involved with 46.44% of the overall revenues, the Germans and British comprised 19% and 19.14% respectively. This shows France's potential because the country only provided 15.42% to the market while the rest of the continent was hurrying to catch up with its neighbour.

The implications of the results showed that logistics and marketing power were important factors important to the United States' success. France's poor results suggested that there may be a lack of market fit or consumer interest in the gas supply service.

Recommendation:

Enhance revenue, increase advertising efforts towards targeted markets, and work on partnerships with specific distributors in France. At the same time, strengthened its position in the USA due to already existing demand and well-developed logistics.

2. Regional Insights and Opportunities

Perhaps, the breakdown of how Accessories, Clothing, and Bikes brought revenue out of the total sales across countries offered additional insights.
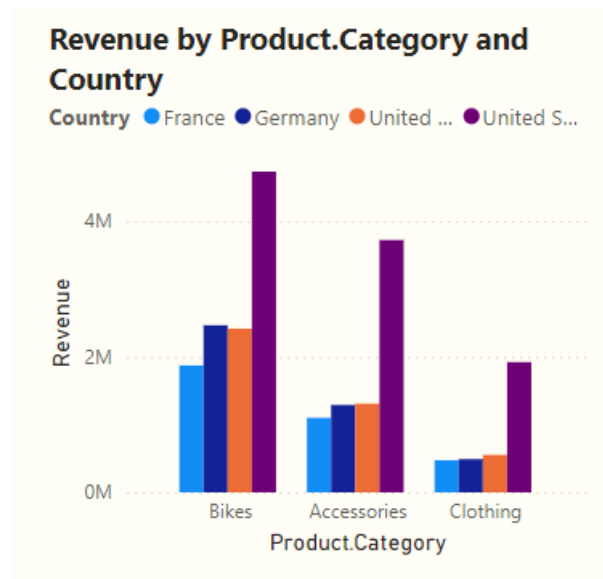
- ✓ **Key Insight:** Bikes were seen to be on top of the list in terms of revenue for all countries since bikes are popular all over the world and can be used in various terrains. The sales achieved in the United States reflected high demand for recreational and lifestyle products where Bikes are ranked the highest. In France for instance, they can be attributed to the escalating number of people who use environmentally friendly means of transport. As adequately pointed out by Siau and Zhang, Germany with much emphasis on sustainability and outdoor activities offered a very conducive market for Bikes. In the same year, UK urban commuting requirements and increasing volumes in the fitness category fine-tuned Bikes to be the highest revenue generator for the company in terms of product type.

Accessories and Clothing had relatively moderate growth across all the countries and were bought along Bikes.

Bikes have been performing consistently, so this should be capitalised by ensuring they have adequate stock and specific sales promotions. Promote accessories and clothing accessories by linking the sales of bikes with the sales of Accessories and Clothing.

3. Balancing Price and Quantity

Revenue optimisation is not only contingent on the quantity sold but also on the price at which the product is sold and the demand for such products.
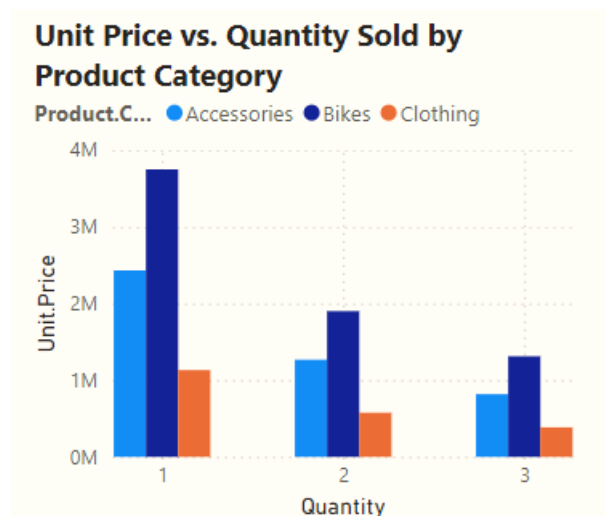

✓ **Key Insight:** Bikes had the highest unit price and contributed significantly to the revenue even though they were fewer in number, and this supported their premium market positioning. Accessories and Clothing proved to be constant in growth through moderate pricing and a higher number of pieces sold.

Recommendation:

Sustain the high-end pricing policy concerning Bikes and focus on increasing volume by promotions targeting valuable clients. Experiment with increasing the price of Accessories a little to enhance profitability and make use of sales promotions offers on Clothing, particularly in France and U.K.

4. Regional Growth Potential

Geographic revenue density allowed to identify areas of high potential, namely the countries and regions with low overall sales.


Revenue by State and Country
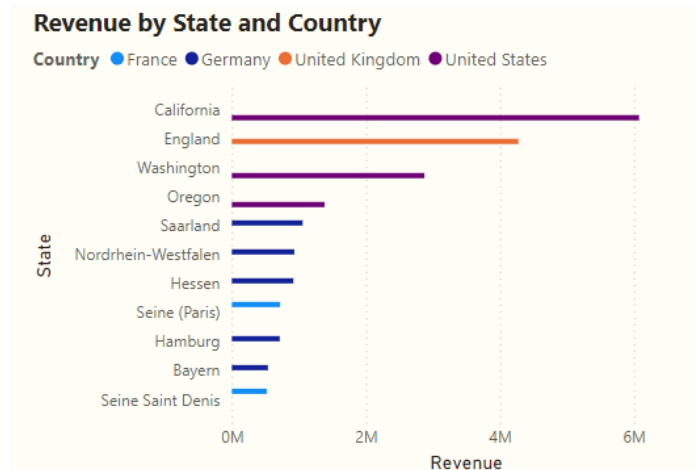
✓ **Key Insight:** In particular, the focus was on the strongest market areas, which were metropolitan areas, especially in California, due to the great density of the population and well-developed transport infrastructure. On the other hand, less urbanised areas such as Wyoming and Montana provided below-par results. Likewise, in Germany, the core city areas of Hessen stand out, while Saarland has potential for improvement.

Recommendation:

The distribution network needs expansion, particularly to areas like Montana and Saarland, to enhance its reach. For urban centres, there is potential to improve customer experiences and conveniences like speed and targeting.

5. Revenue Concentration Across Categories

A pie chart showing general revenue concerning the products indicated that certain products dominated others.

Revenue Share by Product Category
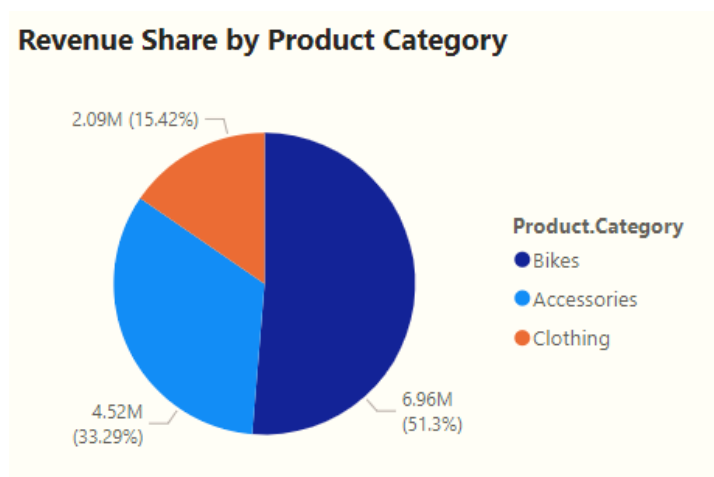
- ✓ **Key Insight:** The top-performing product category was Bikes which contributed 51% of overall revenue, while the second most significant source was Accessories which contributed 33% of revenue and finally Clothing products which contributed 15% of the revenue. The fact that Bikes made up the biggest sales portion demonstrated their importance to the company's operations, while Accessories and Clothing offered consistent sources of increased revenue.

Recommendation:

Invest in the continuation of growth for Bikes and find strategies to expand the market share for Accessories and Clothing, such as through combinations and promotions.

Conclusion:

Elena and her team distilled their findings into a clear strategy:

1. Expand Regional Presence: Enhancing logistics and using localised campaigns can address the underperformance in France, Wyoming, and Saarland regions.

2. Leverage Product Strengths: Take Bikes as the key product while making Accessories and Clothing complementary products, which will attract more customers.

3. Optimize Pricing Strategies: Keep Bikes priced at a premium level and use flexible pricing for Accessories and clothing depending on volume exposure.

By implementing and executing these kinds of analytics, the firm is well-positioned to turn market returns.