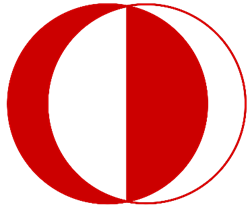
**Distribution of Purchased Vehicles of Countries and Its Order Quantity**

**by Ahmet DOĞAN**

December 6, 2024



**STAT 112**

**Introduction to Data Processing and Visualization**

Tableau Project 1

**Department of Statistics**

**Middle East Technical University**

**1. Introduction**

This report provides a comprehensive analysis of the spending, order volume, and impact of various countries on various vehicles by using 2 different datasets (“autosale"), with different features detailed below. Data visualizations include data on sales by vehicle type, country spending by vehicle, population, total orders, and CO2 emissions by country. The purpose of this report is to derive meaningful analysis and interpret the relationships between these variables. The original datasets include 20 features 2747 observations and 14 features and 195 observations.

**The features in the first dataset (“autosale”):**

* **Order Number**: It represents the unique identification number assigned to each order.
* **Quantity Ordered**: It indicates the number of items ordered in each order.
* **Price of Each**: It specifies the price of each item in the order.
* **Order Line Number**: It represents the line number of each item within an order.
* **Sales**: It denotes the total sales amount for each order, which is calculated by multiplying the quantity ordered by the price of each item.
* **Order Date**: It denotes the date on which the order was placed.
* **Day Since Last Order**: It represents the number of days that have passed since the last order for each customer.
* **Status**: It indicates the status of the order, such as "Shipped," "In Process," "Cancelled," "Disputed," "On Hold," or "Resolved."
* **Product Line:** It specifies the product line categories to which each item belongs.
* **MSRP:** It stands for Manufacturer's Suggested Retail Price and represents the suggested selling price for each item.
* **Product Code:** It represents the unique code assigned to each product.
* **Customer Name:** It denotes the name of the customer who placed the order.
* **Phone:** It contains the contact phone number for the customer.
* **Address Line:** It represents the first line of the customer's address.
* **City:** It specifies the city where the customer is located.
* **Postal Code:** It denotes the postal code or ZIP code associated with the customer's address.
* **Country:** It indicates the country where the customer is located.
* **Contact Last Name:** It represents the last name of the contact person associated with the customer.
* **Contact First Name:** It denotes the first name of the contact person associated with the customer.
* **Dela Size:** It indicates the size of the deal or order, which are the categories "Small," "Medium," or "Large."

**The features in the second dataset (world-data-2023):**

* **Country:** Name of the country.
* **Latitude:** Latitude coordinates of the country's location.
* **Longitude:** Longitude coordinates of the country's location.
* **Birth Rate:** Number of births per 1,000 population per year.
* **CO2 Emissions:** Carbon dioxide emissions in tons.
* **CPI: Consumer** Price Index, a measure of inflation and purchasing power.
* **Gasoline Price:** Price of gasoline per liter in local currency.
* **GDP:** Gross Domestic Product, the total value of goods and services produced in the country.
* **Life Expectancy:** The average number of years a newborn is expected to live.
* **Population:** Total population of the country.
* **Tax Revenue (%):** Tax revenue as a percentage of GDP.
* **Total Tax Rate:** Overall tax burden as a percentage of commercial profits.
* **Unemployment Rate:** Percentage of the labor force that is unemployed.
* **Urban Population:** Percentage of the population living in urban areas.

**2. Data Preprocessing**

Data preprocessing is an essential step to ensure accuracy and reliability in data analysis. The datasets used in this dashboard were assumed to be clean and structured, but basic data cleaning procedures should be applied when handling similar datasets.

* **Editing Column Names:** Especially in the first dataset, the columns where the data was separated were difficult to read and separate in terms of font size and rules. I made them lowercase except for the first one, separated the compound words, changed them to be more readable and understandable.
* **Handling Missing Data:** Missing or null data can distort results, so the dataset should be checked for any incomplete rows or columns. In my case, there were too many null values ​​on Tableau. When I looked at the Excel table containing the data, I noticed that the data was there but because the formats were different, it was shown as null in Tableau. I edited the formats and reloaded my data sets.
* **Correction of Data Types:** In gasoline prices and GDP data, they appeared as string values ​​due to the dollar sign, I converted them to numeric values ​​by creating a calculated field via Tableau.

**3. Exploratory Data Analysis (EDA)**

**Research Questions**

The following key questions were explored during the analysis:

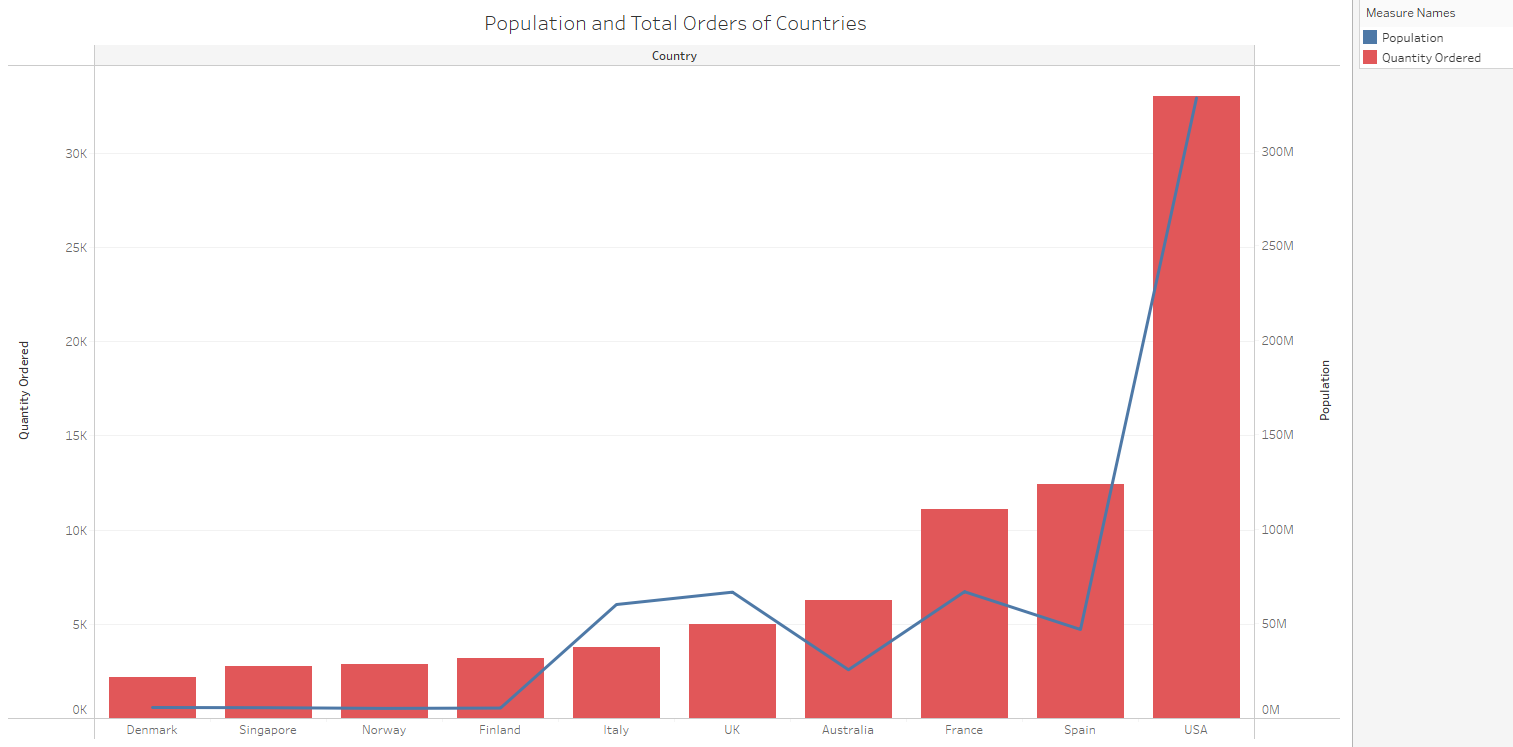
1. **How is the population related to total order volume?** (Figure 1)

As we can see in the graph, since we do not have population information for some countries in our dataset, we cannot perform analysis on these countries.

Population Impact: The US has both the largest population and the highest number of orders. However, in other countries, population and order volume are not always proportional (e.g. France and Spain).

Economic and Cultural Factors: Population is not the only factor that determines order volume. Economic status, e-commerce infrastructure, consumer habits are also important. Even countries with small populations, such as Singapore, have high order volumes.

US Dominance: The US's large market share can be explained by its high internet usage and advanced e-commerce system.

As a result, when examining the relationship between population and order, cultural, economic, technological factors will also need to be considered.

(Figure 1)

1. **Which product groups do different countries spend more on, and what are the factors that determine this spending?** (Figure 2)

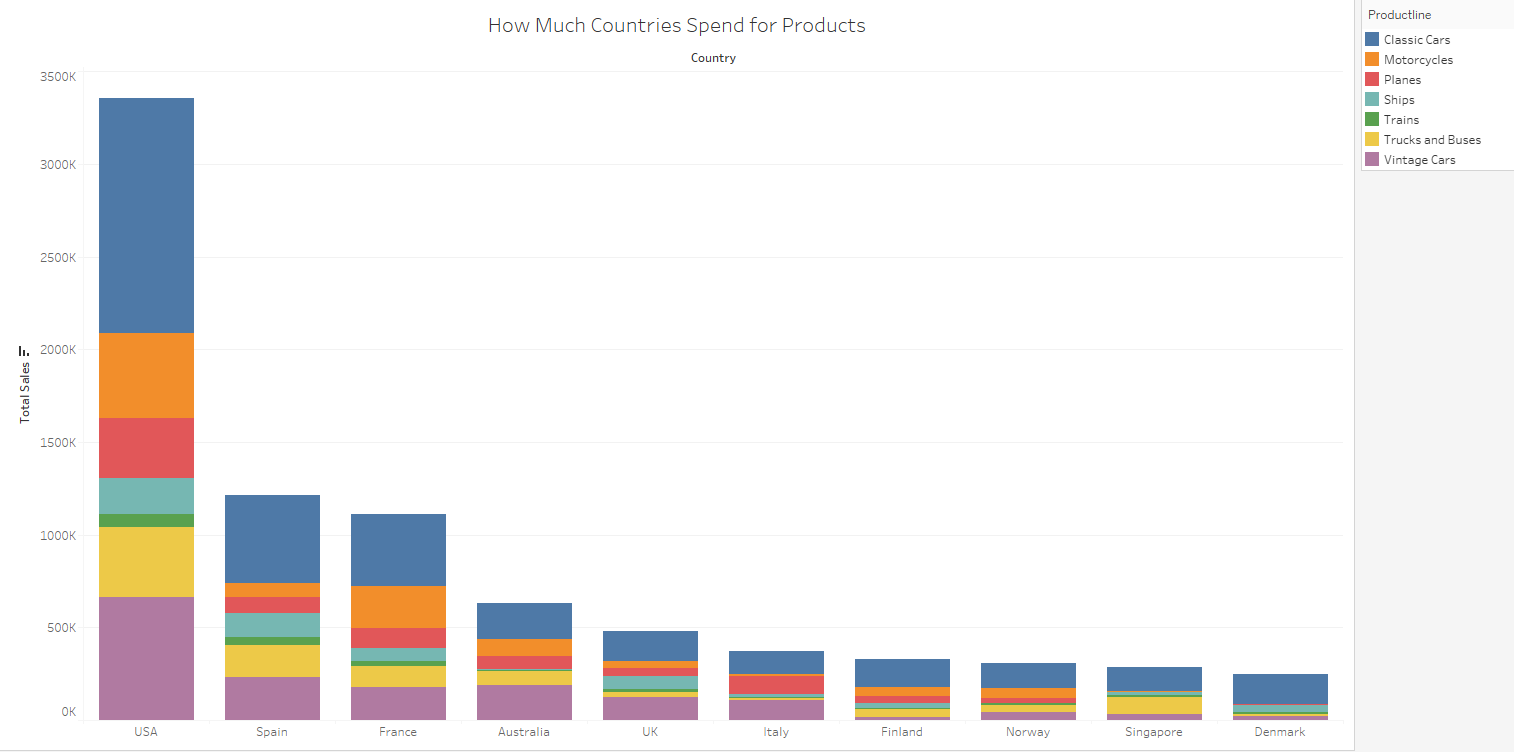
US Outperforms Spending: The graph clearly shows that the US spends much more than all other countries. It has a large market share in product categories such as classic cars (blue), motorcycles (orange), and vintage cars (purple).

Spain and France Spending Distribution: Spain and France focus on classic cars and motorcycles. However, France’s spending distribution is more diverse.

Other Countries’ Spending Structure: Australia, the UK, Italy, and other small countries have similar spending amounts, but there are differences in their product distributions. For example, Australia spends more on airplanes (red), while Italy invests more on ships (green) and motorcycles (orange).

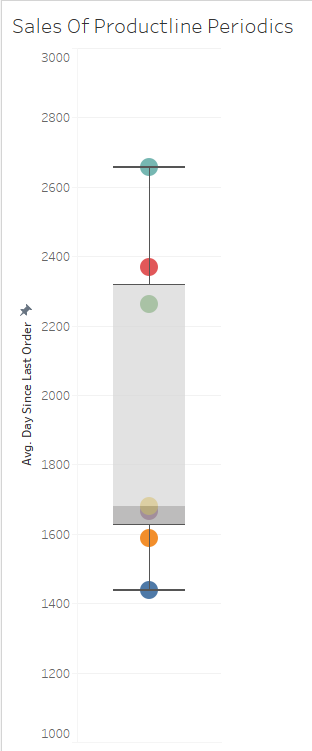
Small Countries (Norway, Singapore, Denmark): Although these countries have lower total spending, there is a balanced distribution across product categories. The focus is mostly on classic cars and motorcycles.

Conclusion: Research based on this graph aims to examine which products countries spend more on and the possible cultural, economic, and technological factors behind these expenditures. In addition, differences in spending habits between countries can provide important insights into international market strategies.



(Figure 2)

**3.** **How do sales frequencies differ by product category and which products are ordered more frequently?** (Figure 3)

****The chart shows the average number of days since the last order for each product category. The order frequency of the products is evaluated according to the average number of days in the chart.

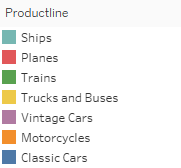
Frequently Ordered Products: Classic cars (blue): It stands out as the most frequently ordered product because the number of days between orders is the lowest (around 1400). This shows that classic cars are popular and in constant demand.

Motorcycles (orange): The most frequently ordered product category after classic cars. The average number of days is around 1600.

Less Frequently Ordered Products: Ships (light blue), planes (red), and trains (green): These categories stand out as the products with the longest intervals between orders. The average period for ships and trains is between 2200-2600 days. This may indicate that larger or more expensive products are ordered less frequently.

Lower priced, every day, or more common products (e.g. motorcycles and classic cars) have higher order frequencies. These products may be ordered more frequently because they appeal to a wider range of consumers.

On the other hand, larger, industrial, or special-use products such as ships, planes, and trains are ordered less frequently because the market for these products is narrower, and the purchasing process may be longer.

****Order frequencies by product category vary depending on the product's price, intended use and market demand. More common consumer products are ordered more frequently, while larger and more expensive products are ordered less frequently. This can be considered when creating market strategies.

(Figure 3)

**4. In which periods of the year do demand for which product groups increase and in which products is this demand increase more noticeable?** (Figure 4)

When the graph is examined, it is observed that demand for different product groups varies throughout the year and there are noticeable increases in some periods.

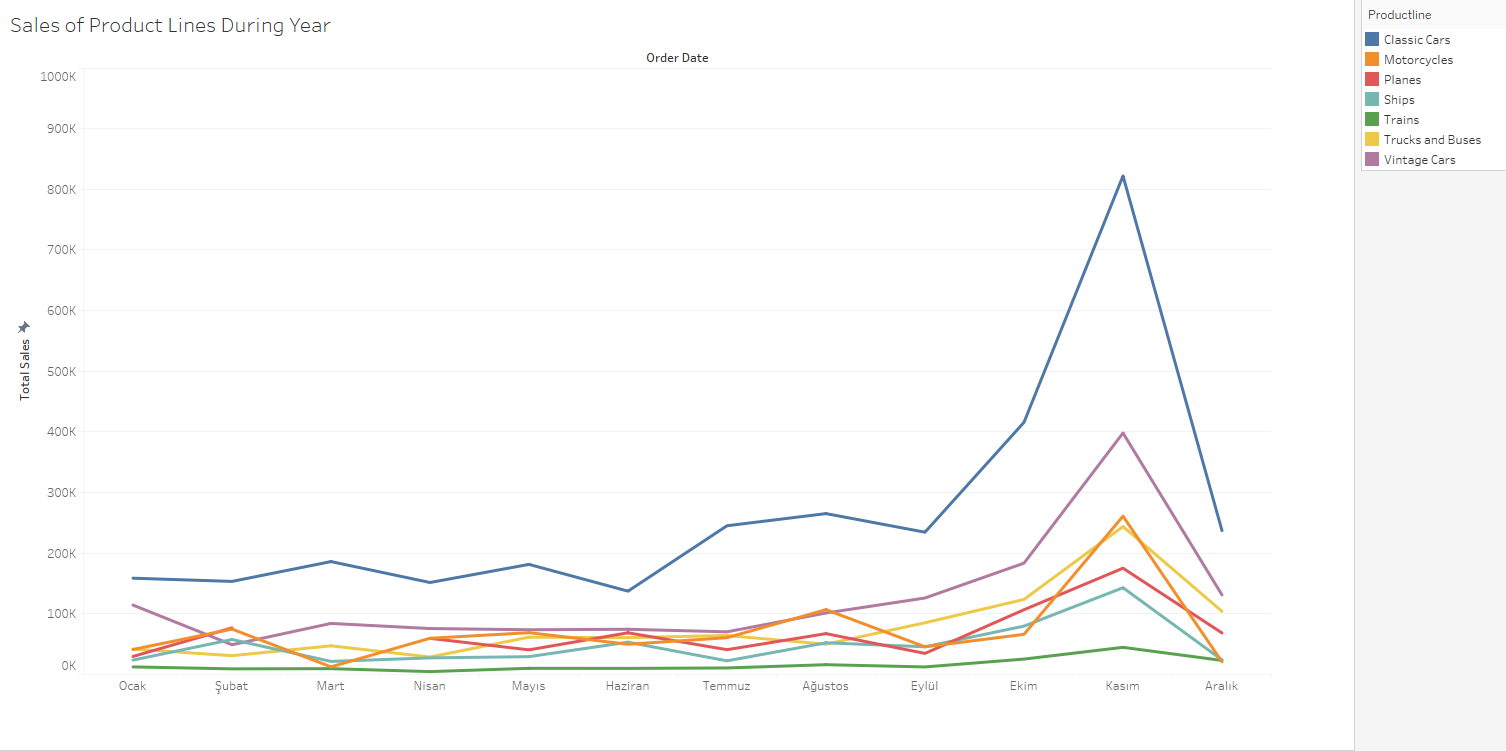
Classic Cars (Blue Line): Classic cars are the product group that sells the most throughout the year. A dramatic increase is seen especially in November and sales approach 1 million units. This increase may be due to discounts, promotions or year-end events held before the holiday season.

Vintage Cars (Purple Line): Vintage cars also show a significant increase in November, but at a lower level compared to classic cars. It can be said that these products appeal to a more niche market and are in demand during the year-end period.

Motorcycles (Orange Line): Motorcycles also peak in November, starting from the summer months. The increase in motorcycle use in the summer and the sales campaigns held at the end of the year may trigger this increase.

Other Product Groups: Airplanes (Red), Ships (Light Blue), Trains (Green): These groups do not see a significant change in sales throughout the year, but there is a slight increase at the end of the year. These large and expensive products may be dependent on longer-term purchasing decisions.

Year-End Sales Increase: There is a significant increase in classic cars, vintage cars, and motorcycles in November. This may be due to the holiday season and year-end sales.

****This analysis can be used to understand which product groups are in higher demand during which periods when creating market strategies. Especially year-end promotions and the holiday season stand out as important factors that increase sales for certain product groups.

(Figure 4)

**5. How are CO2 emissions distributed worldwide, and which countries have the highest CO2 emissions?** (Figure 5)

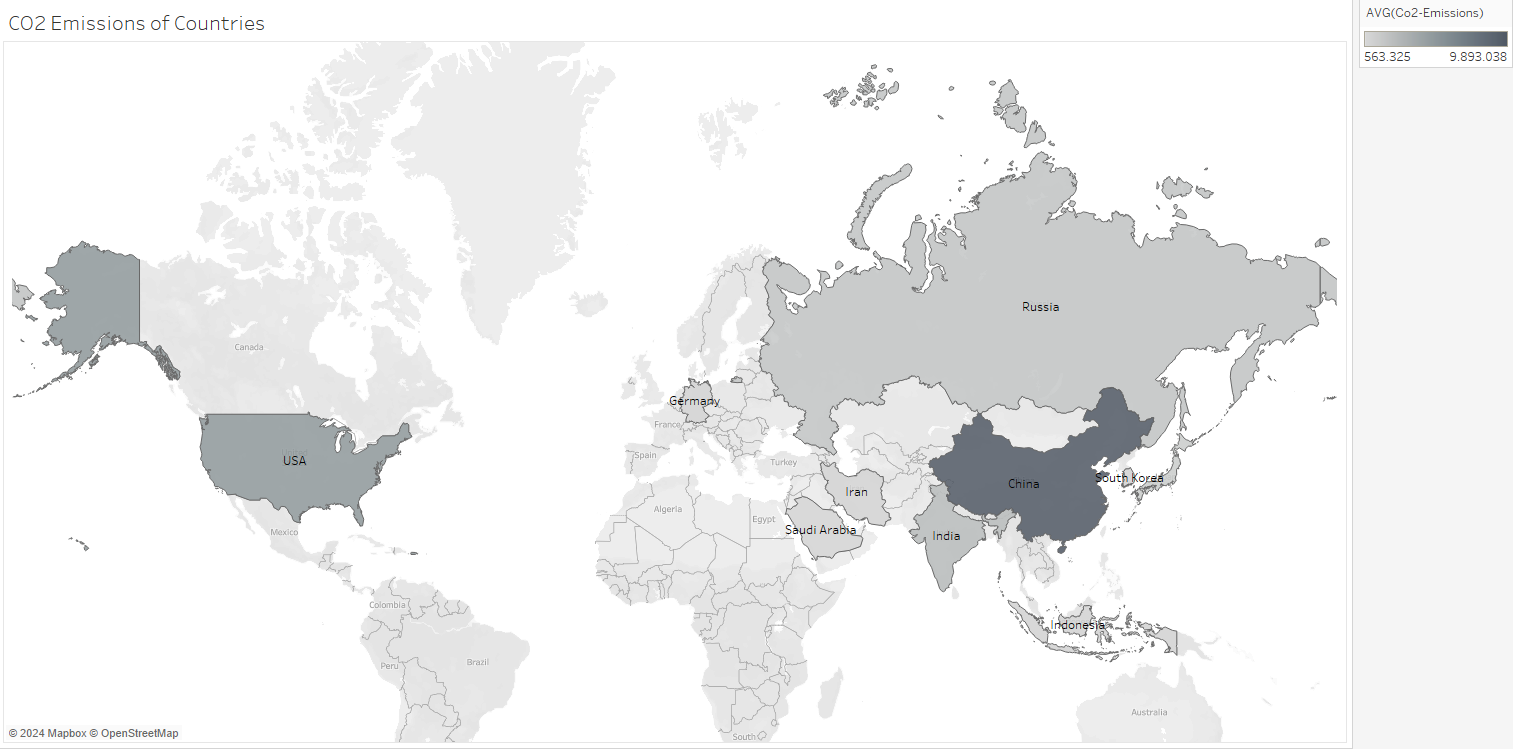
The graph shows the intensity of CO2 emissions worldwide by country. The following important observations can be made:

Countries with the Highest CO2 Emissions: China is the country shown in the darkest color on the map and is understood to have the highest CO2 emissions.

The United States also has significant CO2 emissions. India and Russia stand out as other important countries in CO2 emissions.

Other Important Countries: Countries such as Germany, Iran, Saudi Arabia and South Korea are also among the countries that contribute significantly in terms of emissions.

Distribution and Clustering: It is seen that the countries that make the largest contributions in terms of CO2 emissions are generally large economies and countries with large populations. This can be a direct indicator of industrial production, energy use and fossil fuel consumption.

****The countries that make the largest contributions to CO2 emissions are also countries with large industrial and manufacturing centers. China, the USA, India and Russia stand out as the world's largest CO2 emitters. These countries have a critical role to play in reducing global emissions by shifting towards sustainability and renewable energy use in their industrial and energy policies.

(Figure 5)

**4. Conclusion and Discussion**

The exploratory data analysis has revealed several significant insights. Firstly, larger populations and wealthier nations, such as the USA, have a direct impact on total orders and product spending. Additionally, certain product lines see seasonal sales peaks, which is valuable information for inventory and marketing strategies.

From an environmental perspective, the relationship between sales and CO2 emissions can be indicative of a country's industrial impact. While the USA appears prominently in both sales and emissions, countries like China and India, though major CO2 emitters, are not as dominant in product sales, hinting at their focus on industrial production rather than consumption.

**Key Findings:**

* **Population Impact:** Larger populations drive higher total orders.
* **Seasonal Sales Trends:** Product sales, particularly in Classic Cars, peak towards the end of the year.
* **Spending Disparity:** The USA significantly outspends other countries on products.
* **Environmental Impact:** There is a potential link between high industrial emissions and economic activity reflected in sales.

While many of these findings align with known economic principles, the clear visualization and detailed breakdown of sales trends across product lines provide valuable insights for business strategy and environmental considerations.

**GitHub Link:**

https://github.com/Scrraty/STAT-112-PROJECT-1