STAT 112

INTRODUCTION TO DATA PROCESSING AND VISUALIZATION PROJECT

***GLOBAL AUTOMOTIVE INDUSTRY ANALYSIS***

By

*Bedirhan Salhan*

*2561496*

***INTRODUCTION***

This project analyzes the global automotive industry. AutoSales and World Data 2023 datasets were combined to answer some research questions. The project focused on automobile sales volumes in general and the economic and environmental factors that affect and are affected by them.

|  |  |
| --- | --- |
| ***VARIABLE NAME*** | ***DATA SCALES*** |
| Sales | Numerical (ratio) |
| GDP | Numerical (ratio) |
| Country | Categorical (nominal) |
| Product Line | Categorical (nominal) |
| Deal Size | Categorical (ordinal) |
| Quantity ordered | Numerical (ratio) |
| CO2 Emission | Numerical (ratio) |

***DATA DESCRIPTION AND DATA TIDYING***

AutoSales and World Data 2023 datasets contained many variables. Since I wanted to explain the economic and environmental impacts of car sales here, I preferred to use only some of these variables. I used the inner join method in Tableau to combine these two datasets. Because there were countries common in both datasets. In this way, I prevented null values. In addition, I noticed that these two datasets had some points where they could be related to each other. For example, numeric data such as sales and auto sales GDP contained a positive correlation. However, I noticed that the GDP variable in the raw data was not classified as numeric and I set it as numeric from the data source section. When I realized that the string format could not be changed completely due to the dollar sign and commas, I eliminated this problem by using the code INT(REPLACE(REPLACE([GDP], '$', ''), ',', '')) from the creating calculated fields section. I checked whether duplicate data was using unique variables such as product code, order number, contact first name, and contact second name. USA was not used in most of the graphs because it was an outlier. We can say that it had disproportionately large values ​​for GDP, car sales and many other variables. This caused distortions in the visualizations and made it challenging to identify global patterns.

**EXPLORATORY DATA ANALYSIS**

**A map of the world

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**1- Which countries have the highest car sales, and how does GDP impact these figures?**

The symbol map includes three variables. These are country, GDP and Sales. This map compares car sales by country and their GDP. Each dot corresponds to a country, while the size of the dots represents the sales volume and the colors of the dots represent their GDP (darker colors represent higher GDP). We can say that the highest sales volume and GDP are in France. European countries such as Spain and Germany also have a similar approach to sales volume and GDP. We can say that countries with lower GDP, such as Indonesia, also have lower sales volumes.

A screenshot of a computer screen

Description automatically generated

**2- How are product line sales distributed globally across countries?**

A screenshot of a graph

Description automatically generatedThe treemap includes three variables. These are product line, country, and sales. I showed the product line as a label, countries as color, and sales as size. When we look at the treemap, we see that classic cars have the highest sales volume in every country. We see vintage cars in second place. Other products are mostly similar but vary from country to country. France and Spain also seem to have an effective market share in this chart.

**3- How do sales volumes by deal size vary across countries?**

There are three variables in this box plot. These are deal size, sales, and country. Deal size in both color and columns. Sales are shown as rows and countries as detail. Deal size is divided into Large, medium, and small. Large Deal Size Sales values mostly are concentrated at lower levels but include outliers reaching up to 800K. The graph shows relatively low variability in comparison with other deal sizes. Medium Deal Size Sales distribution is wider; the average is around 200K. This category includes more outliers than other categories, and hence higher variability. Small Deal Size Most sales are concentrated around 100K with minimal variation. The smallest range among the sales of the three deal sizes.

**4- How do unit price and quantity sold impact total sales across different product lines?**

A graph with numbers and circles

Description automatically generatedThere are four variables were used in this chart. These are Unit price, total sales, product line, and quantity ordered. Unit price is specified in columns, sales rows, product line color, and quantity order size. There is a connection between Unit Price and Sales. There is a positive linear relationship between unit price and total sales. Higher-priced products tend to generate more revenue. Classic Cars (blue) and Vintage Cars (purple) generate the highest sales volumes, likely due to their high unit prices and massive quantities sold. Categories like Ships and Planes generate less revenue, suggesting lower market demand or fewer units sold despite higher prices. Larger circles, for example, Vintage Cars, indicate higher quantities sold and thus contribute a lot to the total sales.

A group of colorful circles with names

Description automatically generated **Comparison of CO2 Emission and Car Sales by Country**

A group of colorful circles with text

Description automatically generated

CO2 Emission & Country Car Sales & Country

**5- How do CO2 emissions relate to car sales across countries?**

I used 2 packed bubbles to answer this question. There are three variables in these packed bubble displays. These are country, CO2 emissions, and sales. While the packed bubble on the left shows CO2 emissions by country, the one on the right shows’ car sales amounts. When we examine both charts separately, we can say that similar countries have similar sizes in the charts. For example, while France and Spain are the largest bubbles in the CO2 chart, they are also the largest bubbles in the sales chart. These charts highlight the impact of car usage on emissions, with significant disparities across countries.

***CONCLUSION***

This project examined the economic and environmental impacts of car sales. Using AutoSales and World Data 2023 data together helped to determine the relationship between car sales and GDP, CO2 emissions, country and other variables. Some results emerged from this project. For example, we can say that countries with higher GDP have higher sales volume and CO2 emissions. Moreover, we can conclude that classic and vintage cars have higher sales volume, and the unit price of these vehicles is positively linked to their total sales. These insights helped to better understand the economic and environmental dynamics of the automotive industry.

***DASHBOARD LINK***

https://public.tableau.com/app/profile/bedirhan.salhan/viz/112project\_1/Dashboard1?publish=yes