**STAT 112 - Introduction to Data Processing and Visualization Project**

By

**Şeyha Baran Erdoğan**

**December, 2024**

**Abstract**

This research project examines the relationships between a number of variables and correlations with the global sales of automobiles. Of course, we need to clean, organize, and prepare the available data to make it suitable for our research questions beforehand.

1. **Introduction:** If you are selling anything, your first goal is naturally to make more sales and to follow an appropriate policy for this. Of course, these are not random things, so it would be the most rational move to determine some methods and methods by using the data we have properly**.**

**Data Description:**

* 1. **Data Description**

**Automobile Sales Dataset**

The dataset contains Sales data of an Automobile company. In the Automobile dataset, there exist 20 variables and 2747 observations. The dataset is saved as **“autosale.xlsx”**. The variable information is as follows:

|  |  |
| --- | --- |
| **Column Name** | **Description** |
| **ORDERNUMBER** | This column represents the unique identification number assigned to each order. |
| **QUANTITYORDERED** | It indicates the number of items ordered in each order. |
| **PRICEEACH** | This column specifies the price of each item in the order. |
| **ORDERLINENUMBER** | It represents the line number of each item within an order. |
| **SALES** | This column denotes the total sales amount for each order, which is calculated by multiplying the quantity ordered by the price of each item. |
| **ORDERDATE** | It denotes the date on which the order was placed. |
| **DAYS\_SINCE\_LASTORDER** | This column represents the number of days that have passed since the last order for each customer. It can be used to analyze customer purchasing patterns. |
| **STATUS** | It indicates the status of the order, such as "Shipped," "In Process," "Cancelled," "Disputed," "On Hold," or "Resolved." |
| **PRODUCTLINE** | This column 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. |
| **PRODUCTCODE** | This column represents the unique code assigned to each product. |
| **CUSTOMERNAME** | It denotes the name of the customer who placed the order. |
| **PHONE** | This column contains the contact phone number for the customer. |
| **ADDRESSLINE1** | It represents the first line of the customer's address. |
| **CITY** | This column specifies the city where the customer is located. |
| **POSTALCODE** | It denotes the postal code or ZIP code associated with the customer's address. |
| **COUNTRY** | This column indicates the country where the customer is located. |
| **CONTACTLASTNAME** | It represents the last name of the contact person associated with the customer. |
| **CONTACTFIRSTNAME** | This column denotes the first name of the contact person associated with the customer. |
| **DEALSIZE** | It indicates the size of the deal or order, which are the categories "Small," "Medium," or "Large." |

**Global Country Information**

This comprehensive dataset provides a wealth of information about **all countries worldwide**, covering a wide range of indicators and attributes. It encompasses demographic statistics, economic indicators, and much more. In the dataset, there exist 14 variables and 195 observations. The dataset is saved as **“world-data-2023.xlsx”**. The variable information is as follows:

|  |  |
| --- | --- |
| **Column Name** | **Description** |
| **Country** | Name of the country. |
| **Latitude** | Latitude coordinate of the country's location |
| **Longitude** | Longitude coordinate of the country's location |
| **Birth Rate** | Number of births per 1,000 population per year |
| **CO2-Emission** | 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** | 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 |

* 1. **Data Preprocessing:** If doing business with clean data is not a necessity, irrelevant results may occur. So, first of all, before visualizing, we must make sure that the data we have is stable and clean. Before sending the file to the table, I cleared null and invalid or empty data points through Excel. Then I made the format of some data, such as date, the same. After adding it to the table, while checking to make sure that the variable types were correct, I made sure that the numeric values ​​were not blue and the categorical data were not green, and I made some minor changes (such as converting monetary values ​​containing dollars to numerical data by deleting the unit).

**2. Exploratory Data Analysis:** After compiling and cleaning the data, we had 35 groups and 2747 rows of data. Among these groups, some of the characteristics of the groups I used for my research questions are as follows:

**A screenshot of a green and white table

Description automatically generated**

**3.Research Questions**

**3.1 Is there a correlation between countries' gross domestic product (GDP) and number of vehicle sales?**

Gross domestic product is one of the leading indicators of a country's general economic status on an international scale. To grow when selling any product, it is necessary to pay attention to how much of the sales are well distributed with the GDP proportion so that more sales volume can be achieved**.**

**A screenshot of a graph

Description automatically generated**

**Interpreting the Visualization 3.1 :**

In the visualization here, it is obvious that the countries with high GDP also have high sales, and they seem to be very compatible in terms of multiplier compared to each other. In fact, this shows us how well the distribution of the distributor companies through which vehicle sales are made is organized. Care should be taken to maintain this distribution ratio in future steps.

**Interpreting the Results 3.1 :**

Concluding, it is useful to constantly update the GDP of the countries and to monitor and maintain the correlation between these sales figures from quarter to quarter, because the developments in the world indirectly affect the economic situation of the countries. As can be understood, it would be misleading to always do business with the same data.

**3.2 *What kind of relationship is there between gasoline prices and sales?***

Gasoline prices in countries vary greatly depending on geopolitical situations. The question here is whether the differences in gasoline prices influence sales and in which way.

A graph of gas prices

Description automatically generated

**Interpreting the Visualization 3.2:**

This chart tells us that the reflection of gasoline prices on sales in countries is like bimodal distribution. As in most examples, America is the outlier here due to the difference in population and consumer profile. That two top countries are UK and Spain especially.

**Interpreting the Results 3.2:**

These observations help us draw some conclusions. While countries with cheap gasoline prices, such as Canada and the Philippines, are expected to be more likely to buy autos considering normal consumer habits, we see that sales are below expectations. Therefore, it makes sense that it can help us capture a global market share by increasing sales to these countries through campaigns and government incentives.

**3.3 *What is the impact of the classic cars (most selling auto) on countries' carbon emissions? Is there a correlation between them?***

We all know that the impact and importance of carbon emissions have been talked about more frequently in recent years. It is an important concept for a world that is so industrialized and modernized. This question was about how much emissions this amount of circulation causes in countries while focusing on sales.

A map of the world

Description automatically generated

**Interpreting the Visualization 3.3:**

This visual appears as proof of how much the vehicle industry pollutes the world. It's a sad picture, but this is also a negative effect of industrialization. As can be seen on the map, the country most affected by it is America. The reason for choosing classic cars is that they are the most preferred vehicle type in daily use and have the most impact on traffic.

**Interpreting the Results 3.4:**

To be honest, it is not possible to change this situation in a short period of time, but the attitude of companies and authorities can improve these in the long term and carbon emissions can be reduced with developing technology. Of course, there are things that companies can do personally, the first of which is exhaust filter and engine power balance. These effects can be reduced with good engineering.

**3.4 What is the relationship between tax rates and vehicle sales prices?**

We know from our daily lives that taxes are a problem for the people. It is necessary to observe how much or whether this problem affects their vehicle purchasing habits.

A screenshot of a computer

Description automatically generated

**Interpreting the Visualization 3.4:**

It is difficult to draw a clear conclusion from here because tax rates are very affected by the personal policies of countries and the human profile varies a lot, but good comments can be made. For example, if we accept values ​​around 0.4 as average, there is a sharp distinction between countries below and above it. For example, what is obvious in countries below 0.4 tax rate is that the tax rate and vehicle prices are in a positive relationship. Likewise, at tax rates above this value, there is a full the opposite situation is that as tax rates increase, vehicle prices decrease.

**Interpreting the Results 3.4:**

Tax rates above a certain level actually show the profile of the country. In countries with high tax rates, they are uncomfortable with this, so they reflect it on demand and therefore price their vehicles low. In this case, a step can actually be taken for countries with low tax rates. Most obviously, in these countries, a middle ground can be found with more frequent campaigns and incentives on prices, which will have a positive impact on the other side.

**3.5 *Does the mismatch of MSRP with sold prices affect sales?***

Prices suggested by manufacturers who are more knowledgeable in the business should be taken into consideration. This issue, which is of great importance, may produce unexpected results in the future.

People's reaction to this is changing noticeably.

**A graph showing a graph of green dots

Description automatically generated with medium confidence**

**Interpreting the Results 3.5:**

To get a clear picture of this issue, I based it on the 5 countries with the most sales. As seen in the image here, Italy and Spain have acted very appropriately with MSRP pricing. But other countries are not so careful. This causes the attitude of the consumer who knows this situation, and the sales increase and progress in a parallel manner. This seems to be a situation that renews confidence in the long run.

**3.6 *Could the state's income from taxes (by government) have a relationship with sales?***

Observing the impact of the country's tax revenues on sales helps to observe many things. Consumers' awareness on this issue changes many habits, and this research question is based on this.

**A green bar graph with white text

Description automatically generated**

**Interpreting the Visualization and Results 3.6:**

As seen in the histogram graph, there is a very clear rightly skewed (positively skewed) distribution. We can easily see that sales are much more dominant in countries where tax revenues are low. How the institutions or people who buy vehicles are shaped according to the tax policy of the country. The tax revenue rate here changes people's attitudes a lot. It's quite normal because it's not a situation that would be in his own interest.

We see how effective tax policies are on a global scale. It wouldn't be wrong if we summarized it as countries that are more liberal and operate with lower tax revenues and practice liberal economy. It is a result of the fact that sales are higher in these countries. The only thing that can be done in this regard is to put pressure on countries and demand that they be included in the common free market. If this happens, both the company, the country and the residents will be more profitable.

**4. Conclusions:**

Based on many observations and data, we have seen how many situations can be improved by taking steps. We have also accessed a lot of curious summary information. We have seen that in some cases, the country's politicians have a lot of influence, and in some cases, the citizen profile and living conditions have an impact. It is obvious that this information can be evaluated from the perspective of the people in the analysis and R&D team and very good results can be achieved. In the same way, the impact of vehicles on the world ecosystem can be determined and action can be taken. We have also seen that some extreme values ​​can always happen and that it is not always a pure truth. The biggest example of this has always been America. 😊

Submission Date: 08.12.2024

Tableu Publish link:

https://public.tableau.com/app/profile/baran.erdogan/viz/Stat112ProjecteyhaBaranErdoan/Dashboard1?publish=yes