REVIEW OF THE TOP 5 COUNTRIES WITH THE HIGHEST TOTAL SALES AMOUNT

Mehmet Alperen Çalış

Department of Statistics, Middle East Technical University, Ankara, Turkiye

Email: [e266649@metu.edu.tr](mailto:e266649@metu.edu.tr)

**1.Introduction:**

First, it should be examined well how the total sales amount behaves and is distributed according to what factors. In order to find the behavior and distribution of the total sales amount, the data should be collected properly and interpreted together because interpreting the sales amount according to the behavior of a single variable can be misleading. This report examines the top 5 countries with the highest sales amount. The data is based on the given countries, years and different categories. The main purpose of this report is to examine how the total sales amount changes according to which variables for the top 5 countries with the highest sales amount and to try to draw conclusions.

**1.1 Data Description:**

We have two separate datasets. These datasets were combined with an inner join (based on country) and worked on a single dataset. As a result of this process, 36 fields were created and 2747 observations were obtained. Since we have too many unnecessary variables depending on the scale and purpose of the report, the data used in the report will be explained.

|  |  |
| --- | --- |
| Numerical Variables: | Categorical Variables: |
| Sales (Continuous) | Order Date (Ordinal) |
| Days Since Last Order (Continuous) | Status (Nominal) |
|  | Product Line (Nominal) |
|  | Country (Nominal) |
|  | Deal Size (Ordinal) |

**2.Data Preprocessing:**

Cleaning and reviewing data is one of our most important tasks. First of all, before loading the data into Tableau, it was reviewed in Excel, especially the Autosales dataset was reviewed via Excel to see if there were duplicate data, and no duplicate data was found. After loading the data into Tableau, it was combined appropriately as mentioned and reduced to a single dataset. Then, the data was reviewed to check if our process was correct. No missing data was found during this preliminary review. Then, the variable names were converted to a single common type (We separated the words and made them more readable). Since the data types of some fields were incorrectly matched, they were converted to the appropriate data type and made ready to be used (Although not used in this project, data such as GDP and Gasoline Price were made ready to use. By creating a Calculated Field, the dollar symbols were removed with the replace command and data converted in to numeric.). Variables that were not suitable for the purpose of our project were hidden ( such as Address Line, Postal Code, Old Gasoline Price, Old GDP and etc.)

**3. Exploratory Data Analysis:**

Before moving on to the questions, while reviewing the data we had to get to know it better, it was difficult to establish a connection between Birth Rate, Population, GDP, CO2 Emission, Gasoline Price, Life Expectancy etc. (numerical data in our second data set) and our main data set, the Autosales data set. The main reason for this is that the Autosales data set actually provides very limited data and the data mentioned above does not change over the years. The use of the mentioned data was avoided because the main goal of the project was to examine the behavior of the sales quantity, which is quite variable. These mentioned data were not used because they could be limited in explaining the behavior of the sales quantity and could cause misinterpretations.

**4.Research Questions:**

We will analyze the change in sales volume with 5 questions and try to make inferences. Although the amount of data we have is small, we will evaluate it according to the data we have.

**4.1: How do total sales amounts vary by country?**

This is our first question to find the top 5 countries with the highest total sales, which is one of the project's goals, because in the following questions and graphs we will make and try to interpret these top 5 countries. A bar chart was used to answer this question. Accordingly, in order to best read the visual, the total sales amount was placed in the column, and the countries were placed in the rows and ranked from most to least. In order to draw attention to the top 5 countries, which are our target, they were indicated with different colors.

A screenshot of a computer

Description automatically generated

Visualization1:Total Sales for Each Country

**Interpreting the Visualization1:**

The chart is quite simple and understandable, our main goal here was to find the top 5 countries with the highest total sales volume. These countries are USA, Spain, France, Australia and UK respectively.

**4.2: What is the distribution of total sales volumes of the top 5 countries with the highest total sales volume for product line and deal size?**

The main purpose of our second question is to examine and try to interpret the distribution of the total sales amount of the first 5 countries in product line and deal size. A stacked bar chart was used to answer this question. In order to improve the visualization, similar temperatures but different colors were assigned to the Deal Size category. The total sales amount of the bar was written on top of each bar with a reference line. In addition, the grand total for Deal Sizes was added to the far right of the chart. The total sales amount for different deal sizes of each separate column of the Product line was given in the tooltip

A chart showing the size of a product

Description automatically generated

Visualization2: Sales Distribution by Product Line and Deal Size

**Interpreting the Visualization2:**

Our main purpose in this question was to examine the distribution of the total sales amounts of the first 5 countries with the highest total sales amount according to Product Line and Deal Size. As we can easily see from the graph, the most Classic Car and Vintage Car were sold, the conclusion we can draw from this is that the most cars were sold, and the conclusion we can draw from the Grand Total section is that the most Sales Amount was obtained from Medium Size.

**4.3: What is the distribution of Total Sales Amount and Days Since Last Order by month for the top 5 countries with the highest Total Sales Amount?**

Our main purpose in this question is to examine the change in Total Sales and Days Since Last Order for the first 5 countries with the highest total sales amount by month and to examine and interpret whether there is a connection between them. Dual Combination was used for this. While the line represents the total sales amount, the bars represent the totals of Days Since Last Orders. Different colors were assigned to the lines and bars in order to improve the graph. Labels were added to the points we wanted to draw attention to and if it was desired to examine in more detail, the necessary information was added to the tooltips.

A graph of a line graph

Description automatically generated with medium confidence

Visualization3: Monthly Sales vs. Days Since Last Order

**Interpreting the Visualization3:**

When we examine the graph, it is possible to say that there is a connection between Days Since Last Order and Total Sales Quantity because when one increases, the other also increases. We can draw the following conclusion from here: If the total of Days Since Last Order increases, the Total Sales Quantity will also increase for that month. Also, if we examine the graph here, the total of Sales Quantity and Days Since Last Order is the highest for November 2018 and November 2019.

**4.4: Is there a relationship between Total Sales Amount and Shipment Status for the top 5 countries with the highest Total Sales Amount?**

Our main goal in this question is to see if there is a relationship between Total Sales Quantities and Shipment Status for the top 5 countries with the highest Total Sales Quantity. Tree Map was used for this question. To improve the graphic, different colors were assigned to Shipment Status and the label was activated.

A screenshot of a computer screen

Description automatically generated

Visualization4: Total Sales by Shipment Status

**Interpreting the Visualization4:**

The first conclusion we can draw from this graph is that for the top 5 countries with the highest Total Sales Amount, the highest Total Sales Amount was obtained from the Shipped Status, while the lowest was obtained from the Disputed Status. Another conclusion we can draw from here is that the status of most sales is shipped.

**4.5: How do the total sales amounts change between 2018 and 2020 for the top 5 countries with the highest Total Sales Amount?**

The main purpose of our last question is to examine the difference between 2018 and 2020 for the top 5 countries with the highest Total Sales Amount. A dumbbell chart was used for this question. To improve visualization, different colors were assigned to 2018 and 2020, the chart was sorted from low to high, and total sales amounts and years were labeled on the dots.

A diagram of sales between 2018 and 2020

Description automatically generatedVisualization5: Comparison of Total Sales Between 2018 and 2020

**Interpreting the Visualization5:**

The first conclusion we can draw from this graph is that the total sales volume of the top 5 countries with the highest total sales volume in 2020 decreased compared to 2018. Although the largest decrease in quantity was in the USA, the largest decrease in proportion was in the UK.

**5.Conclusion and Discussion:**

These 5 selected questions have been carefully interpreted and concluded. This report has examined the distribution and the variables by which the Total Sales Amounts of the first 5 countries with the highest Total Sales Amounts have changed. The analysis has examined the variables of the first 5 countries with the highest Total Sales Amounts, Total Sales Amounts, Product Line, Shipment Status, Years, Deal Size and Days Since Last Order. These analyses and inferences can be used to find the patterns of the Total Sales Amounts of the selected countries and from here to show how these total sales amounts can develop for which variables in the coming years.

**6.Dashboard Link:**

<https://public.tableau.com/views/Book1_17328287935370/Project1?:language=en-US&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link>