Review of Shopping Choices in Europe

Mehmet Toprak Özen  
METU Statistics  
STAT 112: Introduction to Data Processing and VisualizationAnkara, Turkey  
e266673@metu.edu.tr

Introduction

Shopping has an important place in modern life. In this project, the shopping behavior of people in Europe will be interpreted. The project will use the datasets “autosale” and “world-data-2023” where the aim is to learn about Europeans shopping behaviors.

data descrıptıon

*Data and their explanation: Automobile Sales Dataset*

ORDERNUMBER: This represents the unique identification number assigned to each order.

QUANTITYORDERED: Indicates the number of items ordered in each order.

PRICEEACH: 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.".

*Data and their explanation: Global Country Information*

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.

DATA PREPROCESSING

In the datasets given, data cleaning is not required because it is already cleaned to analyze. For the joining part, we must choose the option “Inner Join” because we are trying to get common values to explore. After the joining part, some variable categories should be changed. Gasoline Price and GDP are considered as strings so we should change them to numbers.

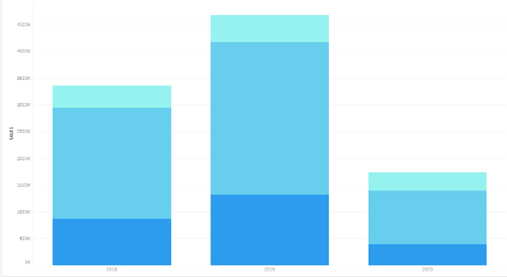
EXPLORATORY DATA ANALYSIS

*Research questions*

Now we will dive into the questions and answer them by visualizing some charts, maps, and plots.

*Question 1: How do product sales including their size vary between the years 2018 and 2020?*

This question is very important to answer because in the datasets sales variable is often correlated with other variables and just knowing how many sales were done in three years might help us to make some forecasts about next year's sales. Moreover, we can observe how “Small”, “Medium”, and “Large” sized sales change across 2018 and 2020 by observing their color changes. From small to large, colors change with respect to shades of blue, from dark blue to lighter blue.



*Interpreting the 1. visualization:*

In 2018, small-sized products had 430.035$, medium-sized products had 1.044.426$, and large-sized products had 185.426. $ in sales, which made a total of 1.660.315$ in sales. In 2019, small-sized sales increased around %56 to 671.987$, medium-sized sales increased around %39 to 1.448.900$, and large-sized sales increased around % 45 to 268.242$ which made total sales increase around %39 to 2.389.129$. In 2020 small-sized product sales decreased around %67 to 225.150$, medium-sized product sales decreased around %64 to 524.565$, and large-sized product sales decreased around %32 to 183.114$ which made total sales decrease around %61 to 929.829.

*Question 2: What is the distribution of shipping status among products?*

Shipping is crucial in our age. Most people do online shopping and they want to see if their product is shipped or not.

ekran görüntüsü, metin, çizgi, öykü gelişim çizgisi; kumpas; grafiğini çıkarma içeren bir resim

Açıklama otomatik olarak oluşturuldu*Interpreting the 2. Visualization*

The categories are “Cancelled”, “Disputed”, “In Process”, “On Hold”, and “Shipped”. By looking at the chart, we can observe that most of the products are shipped or almost shipped.

*Question 3: What are the sales between different product types?*

metin, ekran görüntüsü, dikdörtgen, diyagram içeren bir resim

Açıklama otomatik olarak oluşturuldu

*Interpreting the 3. visualization*

When we look at the treemap, we can see that the areas are related to the total sales of products. The rankings would look like this: Classical Cars with 3.842.869$, Vintage Cars with 1.806.676$, Trucks and Buses with 1.111.559$, Motorcycles with 1.103.512$, Planes with 969.323$, Ships with 700.039$, and Trains with 226.243$ worth of sales.

*Question 4: What is the ranking of product sales across 2018 and 2020?*

Knowing the sale ranks of products would be extremely helpful for companies. This will indicate the most sold product and the least sold one, or changing ranks toward years, etc.

çizgi, diyagram, ekran görüntüsü, öykü gelişim çizgisi; kumpas; grafiğini çıkarma içeren bir resim

Açıklama otomatik olarak oluşturuldu

*Interpreting the 4. visualization*

The bump chart shows the changes, the least and most popular products, and the general rankings of products. When we look at the chart we see that the leading product is classic cars and never changed, the last product does not change rank too. However, other categories of products tend to change.

*Question 5: How do sales differ with product types among their size?*

*daire, ekran görüntüsü, metin, diyagram içeren bir resim

Açıklama otomatik olarak oluşturuldu*

*Interpreting the 5. visualization*

In this bubble plot, we can see sales changing with respect to product size and product type. For instance, we can observe that the most sales were in medium-sized classic cars and the least sales were in large-sized trains.

*Question 6: Which country in Europe has the top sales over the years among product types, and product size?*

Knowing the top country for sales is important for the companies because knowing this can lead to a new line of sight, which would be whether the company focuses on other countries this may change because of the results. metin, yazı tipi, beyaz, tasarım içeren bir resim

Açıklama otomatik olarak oluşturuldu

*Interpreting the 6. visualization*

No matter what the product type is or what year we are looking for sales values, Spain is always the leading country.

*Question 7: What are the percentages of product types according to sales?*

*metin, ekran görüntüsü, yazı tipi, tasarım içeren bir resim

Açıklama otomatik olarak oluşturuldu*

*Interpreting the 7. visualization*

The leading category is classic cars at 41,91% and it goes on like this: Vintage cars with 17.05%, Motorcycles at 10.10%, Trucks and Buses at 10.01, Planes at 9.56%, Ships at 8.58%, and Trains at 2.78%.

CONCLUSION

In the exploratory data analysis, we answered some questions and answered them with some visualizations. In those questions, we used “autosale” and “world-data-2023” datasets, which were uploaded for us to analyze. The variables that were used in the dashboard are ORDERDATE, SALES, Country, DEALSIZE, Latitude, Longitude, PRODUCTLINE, STATUS, and QUANTITYORDERED.

For question one, total sales over years can be observed. If we rank the sales by year, it will look like this: 2019>2018>2020, and by interpreting the data, some forecasts can be made.

In question two, we want to see the shipping distribution, and with the bar chart, we can see the number of products shipped, resolved, on hold, in process, disputed, and canceled. By interpreting the chart, it can be observed that almost %97 of the products are shipped or almost shipped.

In question three, in the treemap we can see the exact value of sales for every product type.

In question four, we added time to our worksheet after question three, and by adding ranks to the treemap, it transformed into a bump chart where we can observe the ranks of products changing over years.

In question five, we answered the question with a bubble plot. The size of the bubbles indicates the sales, and the colors indicate the type of product while dividing the product bubble into three pieces according to the size of the product.

In question six, we showed what country is the leading one according to their sales. Across Europe, the leading country was Spain.

In question seven, the percentages of product types were shown by the visualization.

In addition, to have an easier experience for the dashboard, there is a part where you can choose the countries you want to see. If a country is chosen it will automatically change the other visualizations for the country chosen.

harita, metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

The link to the dashboard: https://public.tableau.com/shared/X6QTSXTF6?:display\_count=n&:origin=viz\_share\_link