**———Project Report———**

**Auto Sales Performance And**

**Product Line Preferences Across Countries**

Data Visualization Report for STAT 112

**————————**

By Ahmet Furkan Köşedaşı

November 2024

METU/STAT

**1. Introduction:**  In this research, a dataset that combines two separate datasets has been interpreted and analyzed. The data spans from the sales records of an automobile company to extensive information about countries worldwide, covering the distribution of product line preferences and customer analysis. The main objective of this report is to effectively visualize the data to assess the current automobile sales performance of our company and provide insights into identifying the product types and customers that should be prioritized to enhance profitability. Furthermore, this report aims to serve as a guideline for the company to determine which countries or products to focus on, based on the observations and analysis presented.

**1.1 Data Description:** The original dataset contains 2,747 observations and 33 different variables. However, for this project, a total of 12 variables will be utilized, including 6 numerical and 6 categorical ones. Raw Data:

|  |  |
| --- | --- |
| Numerical Variables | Categorical Variables |
| Sales (Total Sales Amount of Each Order) | **Country** (Geographical) |
| CO2 Emissions (Continuous) | **Product Line** (Nominal**)** |
| Gasoline Prices (Continuous) | **Customer Name** (Nominal) |
| Price for Each (Continuous) | **Contact First Name** (Nominal) |
| MSRP (Continuous) | **Contact Last Name** (Nominal) |
| # of Sales (Continuous) | **Order Date** (Date) |

**2. Data Preprocessing:** One of the most crucial processes in data analysis is Data Cleaning. First, after loading the data into Tableau, an inner join was applied to combine the datasets, allowing us to focus on working with a single dataset. This step was carefully checked multiple times to ensure that an appropriate join type was used, resulting in no gaps or null data. Then, “Data Interpreter” feature of Tableau was utilized to verify that no issues were overlooked during the data scanning process. After checking for null values, variable names were edited by removing unnecessary marks and setting more readable names for clarity. Since the dataset included two variables which are named “Country,” one of them was hidden to avoid confusion. Additionally, two errors were identified in the variable types for Gasoline Price and GDP. These variables were intended to be numeric but were mistakenly assigned as strings due to their formats (dollar signs). To prevent potential issues during data visualization and analysis, “Calculated Field” feature of Tableau was used to remove the dollar signs, converting these variables into numeric format.

**3. Exploratory Data Analysis:** Before addressing the questions and explanations, it is important to clarify certain aspects related to the variable names. Two variables used in the dashboard and other visualizations require explanation to ensure a better understanding of the concepts. The first variable, MSRP, stands for Manufacturer's Suggested Retail Price and represents the recommended selling price for each item. The second variable, Price for Each, specifies the actual price of each item in the order. The remaining variables in the dataset do not require detailed explanation due to the simplification steps applied during the data processing stage.

**3.1 Research Questions:** We will analyze the company's sales, customer behaviour, product lines, and the influence of global indicators on product line preferences through at least six research questions. Additionally, we will aim to identify potential strategies for the company to enhance profitability.

**3.1.1 How Do Auto Sales Performance Vary Across Different Countries?**

It is important to know the contributions of countries on global auto sales. We used bubbles that represent countries and their total auto sales contribution which is transformed from symbol map. Bubbles will be good assets for our interactive dashboard.

A group of circles with text

Description automatically generated

Visualization: Auto Sales Performance Across Countries

**Descriptive Statistics:**

1) The USA ($3,355,576) represents the largest share of global auto sales in the dataset, while Spain ($1,215,687) although having considerably lower sales than the USA, ranks as the second-highest contributor.

2) Countries like France ($1,110,917), Australia ($630,623), and the UK ($478.880) fall in the mid-range of sales, reflecting strong markets, however not as dominant as the USA.

3) Ireland ($57,756), Belgium ($108,413), and Philippines ($94,016) represent the smaller contributors to global auto sales.

**Interpreting The Results:**

1) The Dominance of USA: The USA is the largest market for auto sales in the dataset by far. This may be due to high population, a more car-centric culture, or greater power of purchasing compared to other countries in data set.

2) Spain’s Surprising Performance: Spain is unexpectedly the second-largest market for auto sales. Considering the similar population and economic conditions with other European countries, we might make a consumption our company has strong demand customers in Spain.

**3.1.2 Is There a Relationship Between Product Line Preferences and Their Average CO2 Emissions?**

One of the essential global indicators is CO2 emissions of product types. We used scatter plot to see if there is a connection between product line and average CO2 emission. Because of the outcomes, this scatter plot will not be represented in our interactive dashboard.

A white grid with black text

Description automatically generated

Visualization: Relationship Between Product Line Preference and CO2 Emissions

**Descriptive Statistics:**

1) Classic Cars have the highest number of sales and have moderate average CO2 emission levels.

2) Ships have the lowest average CO2 emission levels, while Motorcycles have the highest average CO2 emission levels.

**Interpreting The Results:**

1) Although motorcycles have the highest average CO2 emission levels, number of sales may not be change due to transportation preferences.

2) There is no clear, linear correlation between average CO2 emission levels and sales numbers of product type.

3) The outcomes show number of Classic Cars sales is an outlier, but it is not related with CO2 emissions.

**3.1.3 How Do the Average Gasoline Prices Differ Across Different Product Lines?**

Another one of the essential global indicators is Gasoline Price. We used bar chart to compare the average gasoline price of product types. Because of similar reasons with our first global indicator, this chart will not be represented in the interactive dashboard.

A graph of blue rectangular bars with white text

Description automatically generated

Visualization: Average Gasoline Price for Each Product Class

**Descriptive Statistics:**

1) Product line with highest average gasoline price is Trains at $1.156 per gallon.

2) Product line with lowest average gasoline price is Motorcycles at $1.056 per gallon

3) The range in average gasoline prices across all product lines is $0.10 per gallon.

**Interpreting the Results:**

1) The data suggests that gasoline prices do not have a significant change across the product class.

2) The range of average gasoline prices, difference between highest prices and lowest prices, is negligible and due to findings, it is not an important factor in determining product line preference.

**3.1.4 How Do the Distribution of Sales Vary Across Different Product Line and Countries?**

Distribution of sales is an important factor for companies to expand their business. In this visualization, we used box plots, five-number-summary, with excluding the USA which is a huge outlier that complicates to see the distribution.

A screenshot of a computer

Description automatically generated

Visualization: Sale Distribution of Countries

**Descriptive Statistics:**

1) Several product lines have outliers (Spain, France) that are significantly higher than main body of the data.

2) The price range for each product line vary significantly, some have wider spread of prices (Vintage Cars and Classic Cars), others have more narrow range (Trains, Ships).

**Interpreting the Results:**

1) Presence of outliers and the wide range of prices for certain product lines, such as Vintage Cars and Classic Cars, suggest potential opportunities for targeted price optimization, expand into niche markets, or develop new products to cater to understand customer segments.

2) Total train sales are not satisfactory compared to other product lines, company may consider changing the price policy or the investments in train manufacturing.

**3.1.5 How Do the Differences Between Suggested Retail Price (MSRP) and the Current Market Price Vary Across Different Product Lines and Order Years?**

In this visualization, we used clustered bar charts with support of lines that represent yearly suggested price for each product class. Comparison between median suggested price and median current price can allow company detect price anomalies.

**A graph of different colored bars

Description automatically generated with medium confidence**

Visualization: Comparison Between Suggested Price vs Current Price for Each

**Descriptive Statistics:** (Suggested and Current Price values are median in this visualization)

1)Suggested price for each product line and current price values are very similar in 2018.

2) The price gap between the two variables has narrowed for Planes but widened for the rest of the product types in 2020 while Trucks and Buses and Vintage car remain steady.

**Interpreting the Results:**

1) Some comments can be made with the help of information from this and the previous visualization.

- Company has a significant number of sales in Classic Cars, to increase the profitability prices of this category can be increased to the suggested price level.

- Also, company has lower number of sales in Trains, Ships and Motorcycles compared to other product lines. Therefore, price levels of these product lines can be decreased to the suggested price level to increase the sales.

**3.1.6 How the Number of Sales Distribute Across Different Customers?**

Detecting favourite and loyal customers of company is an important part of the work. In this visualization, we used tabular data visualization including the information of customer and number of sales which is indexed with a colour palette. The data have filtered to the customers who is in top 20 at number of sales.

**A screenshot of a computer

Description automatically generated**

Visualization: Customer Ranking in Number of Sales

**Descriptive Statistics:**

1) The wide range of index values suggests significant differences in sales performance across the customer base.

2) The highest index value is 9,327, is associated with the customer “Euro Shopping Chanel” based in Spain.

**Interpreting the Results:**

1) It is possible to identify the key factors driving the outstanding sales performance of top customers and find ways to replicate these successful practices across the broader customer base.

2) Company can allocate additional resources and implement personalized account management strategies to strengthen and grow relationships with high-performing customers.

**4. Conclusion and Discussion:**

Many questions were answered in these six carefully selected questions. This report explored contributions of countries on global auto sales, relationship between product line preferences and CO2 emissions, average gasoline price for each product class, comparison of sale distribution of each product class across countries, comparison between suggested price and current price for each product class, and also, customer rankings in number of sales.

The insights gained from this these analyses can guide the company in applying suitable market strategies to less dominant markets, arranging price levels by observing suggested prices by manufactory according to the sales proficiency of product type, and planning about price policy and investment plans.

The findings also highlight opportunities for expanding into niche markets to explore new opportunities, targeted price optimization for countries and customers, developing new products to cater to understand customer segment, and allocating additional resources and implementing personalized account management strategies to strengthen and grow relationships with high-performing customers.

Overall, this report provides actionable insights for the company to enhance customer satisfaction, increase sales, and stay competitive. The detailed visualization and observations offer a roadmap for strategic decision-making in the dynamic retail landscape.

Tableau Public Link: <https://public.tableau.com/views/TableauMidtermVisualizationAhmetFurkanKeda/AnalysisDashboard?:language=en-US&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link>