1 1

Project .docx

trabajos -- no repository 046

Trabajos de grado finales 2024A

Trabajos de Grado

Document Details

Submission ID

trn:oid:::1:3105835303

Submission Date

Dec 7, 2024, 5:12 AM GMT-5

Download Date

Dec 7, 2024, 5:13 AM GMT-5

File Name

Project_.docx

File Size

410.2 KB

7 Pages

1,663 Words

8,709 Characters

4% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Match Groups

- **4** Not Cited or Quoted 4%
  Matches with neither in-text citation nor quotation marks
- **0** Missing Quotations 0%
  Matches that are still very similar to source material
- **0** Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

- 0% Internet sources
- 0% Publications
- 4% Submitted works (Student Papers)

Match Groups

- **4** Not Cited or Quoted 4%
  Matches with neither in-text citation nor quotation marks
- **0** Missing Quotations 0%
  Matches that are still very similar to source material
- **0** Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

0% Internet sources
0% Publications
4% Submitted works (Student Papers)

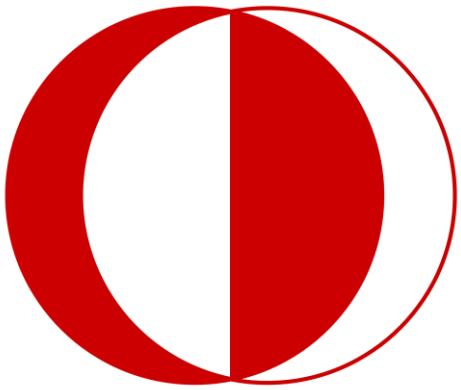
Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1 Student papers

Middle East Technical University **4%**

AUTO SALES DYNAMICS AND THEIR ROLE IN SHAPING ECONOMIC CHARACTERISTICS AMONG G10 COUNTRIES



A PROJECT REPORT SUBMITTED

IN FULFILMENT OF THE REQUIREMENTS FOR THE COURSE

STAT 112 - INTRODUCTION TO DATA PROCESSING AND VISUALIZATION

DEPARTEMENT OF STATISTICS OF

MIDDLE EAST TECHNICAL UNIVERISTY

BY

Inas Sekar Nilamsari

December 2024

1.ABSTRACT:

This project aims to analyze the relationship between automobile sales, economics performance, and environmental behavior that focus on the Group of 10 (G10) countries i.e. countries that are leading global economic growth, include the United States, United Kingdom, Switzerland, Spain, Japan, Germany, Belgium, Canada, France, and Italy. To achieve this, two datasets (auto sales and global country data) have been provided which serve as the basis for addressing research questions and drawing meaningful conclusions. In this analysis, with the sales of vehicle information and fundamental data from each country in G10 Countries, the relationship between vehicle sales, economic and environmental patterns are interpreted using descriptive statistics and exploratory data analysis (EDA) techniques. From the analysis, the following were found:

- Classic car is the most common vehicle type on sales.
- The USA is leading the sales and almost all economic factors in the world.
- There are detected widespread on medium size from quantity ordered.
- There is a linear relationship between tax revenue and sales.
- There is an increasing trend between the relationship of each price and sales across the product line.
- There is a strong positive linear relationship between Co2-emission and gasoline price with sales.

2.INTRODUCTION:

The dynamics in sales of vehicles has become a major influence of both economic and environmental changes, particularly in the G10 countries. The Automobile industry plays a vital role in growing economic, yet it also may bring with it concerns, especially in terms of environmental sustainability. Therefore, this project aims to conduct the economic and environmental effects of automobile sales in the G10 countries, firstly with a focus on how vehicle sales pattern in G10, includes the common product line sales, relationship of the prices and quantity ordered and finally indicating how the dynamic sales is influencing on the tax revenue, gasoline price and CO2 emissions.

In this research the datasets that have been prepared and analyzed contain information of Sales data of an Automobile company for 3 years between 2018 until 2020 and Global Country information which contributes a wealth of information about all countries worldwide, encompasses demographic statistics, economic indicators, and many more. By the end of the project, it was able to answer 5 research questions regarding the impact of auto sales on the G10's economy and environment by following the conclusions through the data visualizations.

2.1 Data Description:

The automobile sales dataset there exists of 20 variables with 2747 observations. The global country information there exists of 14 variables with 195 observations. Subsequently, after the left joined to the auto sales data that correspond to country variable, it was generated 35 variables (2 country variables from each dataset) with 2747 observations in total. That occurs in 15 dimensions, qualitative variables and 19 measurements, quantitative variables. Moreover, the sales variable is applied for 4 over 5 research questions to demonstrate visualizations followed by product line, each price of product, tax revenue, Co2 emissions and much more.

2.2 Research Questions:

Through the analysis of the datasets, there are 5 research questions that have been established by looking at the correlation between the sales of automobile with other variables among G10 countries.

- Which type of vehicle is the most common in sales?
- How does each product's price affect the sales of vehicles across the product line?
- How does the relationship between each product's size level and the quantity ordered?
- How does the sales of automobile affect the tax revenue?
- How does the gasoline prices and CO2 emissions are affected by the sales of automobiles?

3.DATA TIDYNG AND CLEANING STEPS:

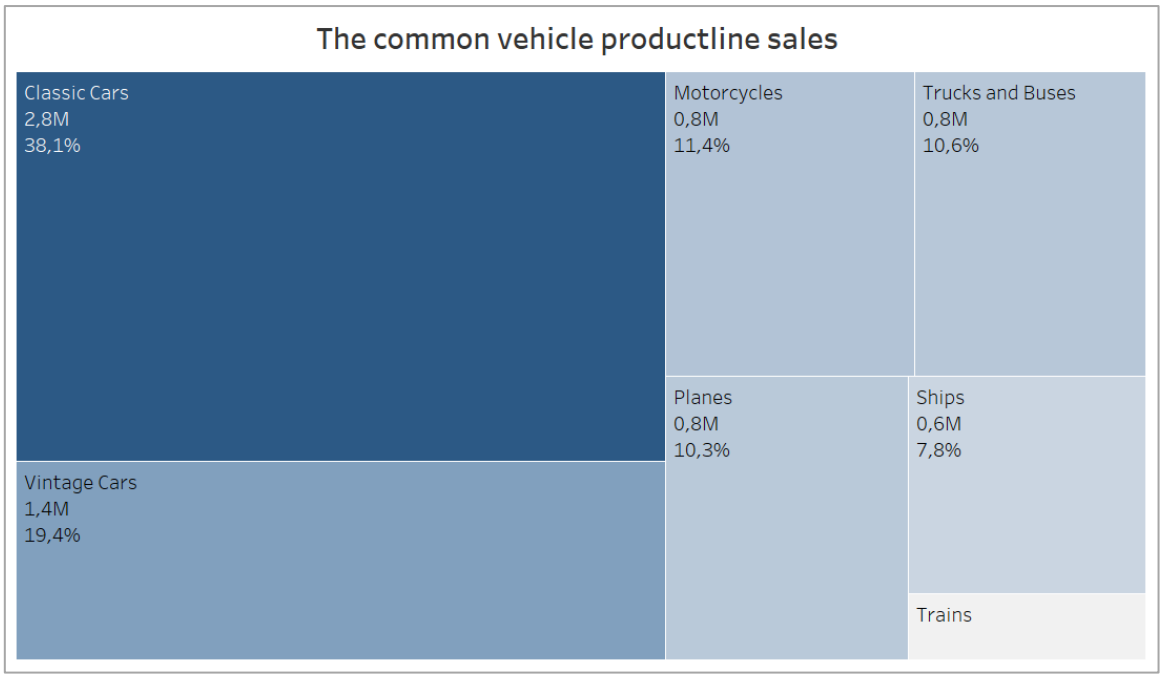
Two data sets are left joined to the auto sales data that are examined based on the country variable by both datasets. The data cleaning process only includes identifying null values. However, there are no detected null values on the data. The data tidying that had been done while the project was under analyzing is creating the new set that is called G10 countries to help on filtering the data while visualizing to focus on G10 countries effectively. In addition, formatting

of variables, uniformity of the units, detecting the outliers etc. were not applied on this project since the quality of datasets has been adequate.

4.EXPLORATORY DATA ANALYSIS:

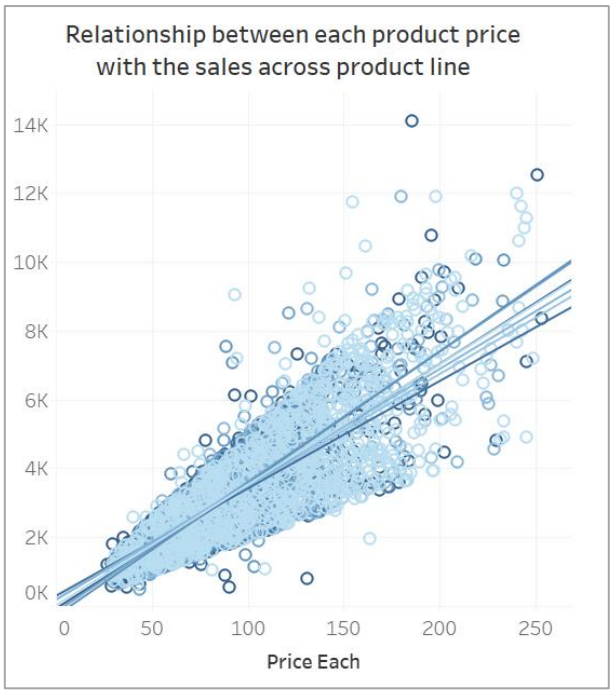
This project is established 5 research questions.

1) Which type of vehicle is the most common in sales?



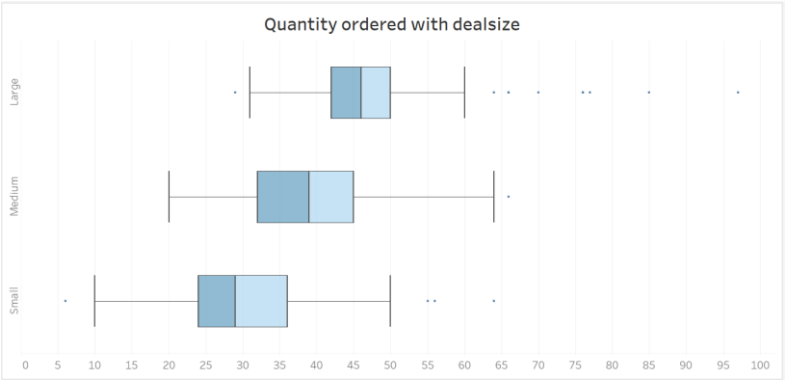
Product line or vehicle type is categorical variable, and sales is numerical variable. To present these binary variables, tree-map is one of the best visualizations to utilize. As presented on the tree-map, classic cars are leading to be the most sales in G10 countries that cover more than 35% of the total sales, followed by vintage cars 19%, motorcycles 11% and more. The classic car sales trend line seems likely to remain the same by looking at the line chart that is provided on the dashboard which means it may lead the top of most in sales in the following years. It is also can be interpreted that top seller country for almost all the product line types is the USA, in here the readers may get the idea easily on dashboard because tooltip is used to provide a bar-chart of sum sales by G10 countries.

2) How does each product's price affect the sales across the product line?



The variables used to compile this research question are sales, product line, and each price of the product line. The scatter plot provides the most appropriate graph to conclude whether there is a correlation between each product price and sales across product types. From the graph between two numerical variables of sales and each product's price, there is a weak positive relationship. The reference lines that correspond for each product line indicate a linear relationship between sales and each product's price. It can be interpreted that the higher price of vehicle follows the higher of sales for each type of vehicle.

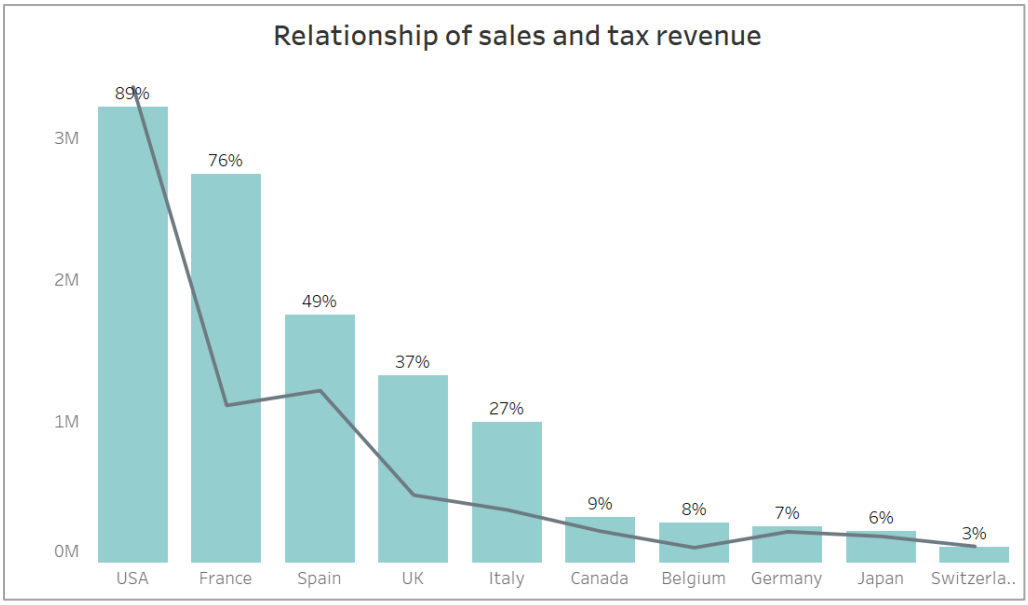
3) How does each product's price affect the sales across the product line?



A box plot was considered as the best way to display the quantity of the size of the product and the quantity of the ordered since the variables were used are categorical and numerical. The medians of the sales increase while the sizes are bigger {median according to size (small = 29, medium = 37, and large = 46)}, yet their median values are significantly different from each other. In addition, according to the distribution of spread, despite large size being the biggest median among all the sizes, medium size has the widest range of quantity ordered, followed by small size. For this reason, as expected from the most common vehicle in sales, medium size types of vehicles are leading the sales. The

distribution of medium and small size shows a strong positive skewed as the range of whiskers are longer on the right side. However, the small size box plot is slightly symmetrical distribution despite having a huge number of upper outliers.

4) How does the sales of automobile affect the tax revenue?

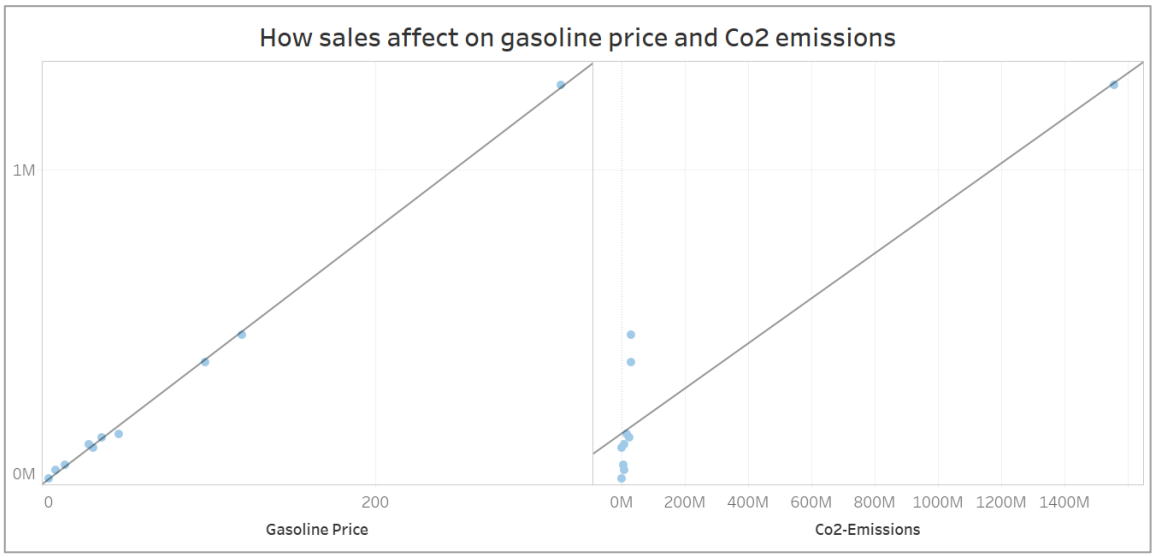


To illustrate the correlation between tax revenue, sales and G10 countries a comparative of bar chart and line chart may conduct applicable visualization to conclude efficiently as it includes two quantitative variables and a qualitative variable. The graph shows that the highest tax revenue among G10 is the USA for 89% that also being top seller of automobile. Despite the second highest sales is Spain with the third place of tax revenue which is 49%, the remaining line charts show the linear relationship between the tax revenue and sales, it may be concluded that the sales may be one of parameters that affect the tax revenue of countries, which same as what it's expected.

5) How does the gasoline prices and CO2 emissions are affected by the sales of automobiles?

Scatter plots are beneficial in presenting multiple numerical variables to show the increasing or decreasing correlation. Thus, comparative scatter plots are used to visualize and answer this question. According to the graph below, gasoline price and sales show a strong positive relationship. It clarifies that higher sales are followed by rising gasoline

prices in G10 countries. Therefore, it is also expected that Co2 emission may increase while the sales of vehicle are escalated. As a result, the increasing sales of vehicles in such



countries can result in rising gasoline prices and CO2 emissions. The USA is being the top plots for both comparison with sales of more than 1 million with 311 price of gasoline fuels and 1500 million Co2 emissions. Thus, the gasoline fuel demands get higher since the vehicles that are gasoline-powered are widely abundant. This increases Co2 emissions because the consumption of gasoline fuels is continuously high.

5. CONCLUSION:

To conclude, this study that focus on G10 countries emphasized on 3 main subjects: product line sales and impact of price, the tax revenue trend which result from the sales pattern and the Co2 emissions which effects from the fossil fuel and automobile sales. In general, auto sales effect the economic pattern since it plays a big role of income that impacts on the tax revenue in such countries. Moreover, high product price and size seems likely to sold more which might be one of the factors that affect the country's income. Thus, the environmental factors such as Co2 emission also effected by the huge amount of auto sales. Lastly, the USA is leading the sales of vehicles that has a relation on the high tax revenue and Co2 emissions because of the high population of the USA, and for the rest of G10 countries seems have a stable of economic growth since their big role in sales of automobile.

https://public.tableau.com/app/profile/inas.nilamsari/viz/project_17328269690390/Autosales?publish=yes and https://github.com/inassekar/STAT112-Project-1