Data Analysis Project on Car_sales.csv DataSet

Ajayi David

Saint Louis University

AA 5200 Data Visualization, Feedback and Dissemination

Dr Keith Hacke

Octoberr 11, 2022

Executive Summary:

The following the analysis was carried out on a 'Car_sales.csv' containing various car models by manufacturers and factors which could influence their sales. The auto industry sometimes experiences very rough patches in terms of sales which could be attributed to inflation. This makes individuals rethink expensive purchases. Additionally, supply-chain snags are making it difficult to even find new cars which further causes a plummet in auto sales. Therefore, this Car_sales.csv dataset was analyzed for companies to determine which factors could boost their market sales.

From the Car_sales.csv dataset, 157 observations were found concerning 16 different variables relating to car sales some of which were manufacturers, model, resale value and different car features. For analysis, visualization and tables are employed to clearly depict how prices and sales influence each other.

DATA DEFINITION

The first step taken in the data mining process was to define this project and summarize the issues we would consider as they relate to car sales. The objective of this project is to analyze the Car_sale.csv dataset and various meaningful insights using the variables. Also, we discovered were no licensing restrictions to the obtaining the dataset.

What overarching goal set towards analyzing this dataset?

The dataset showed different factors such as engine size, fuel efficiency and horse power rating and how they impact overall car sales. The overall aim for analyzing the Car_sale.csv dataset is to accurately represent the data contained and depict trends from which clearer insights can be drawn as to why the data was gathered.

DATA PREPARATION

The dataset used was derived from Kaggle.com, a world renowned database repository (Kaggle, 2022) and comprised of 157 rows of data with 16 variables. Data cleaning and transformation were executed using the Tableau prep builder and the Tableau desktop filer function in the visualization interface.

Table 1: Description of the variables in the test.csv data set

Field Name	Data type
Manufacturer	String
Model	String
Sales in thousands	Numeric
__Year_Resale_Value	Numeric
Vehicle type	String
Price in thousands	Numeric
Engine size	Numeric
Horsepower	Numeric
Wheelbase	Numeric
Width	Numeric
Length	Numeric
Curb weight	Numeric
Fuel capacity	Numeric
Fuel efficiency	Numeric
Latest Launch	Date
Power perf factor	Numeric

ANALYSIS

For this analysis, both tabular and chart visualizations were used given various insights required to be drawn from the dataset. Tables were used for visualization due to their verbal interaction and bar chart were used also to provide more visual insights (Nussbaumer, 2020)

DEPLOYMENT

For the first inference we draw in this project, we considered the question**, 'Which feature has the most effect on car sales?** The variables considered towards creation of the line chart were Sales in thousands, Fuel efficiency, Engine size and Horsepower.

From the Line chart obtained using the Car_sales.csv dataset (Figure 1), it can be deduced that Fuel efficiency, Engine size and Horsepower had similar patterns on sales which depicts that they are all factors which influence a buyer's decision to buy a car. However, we could see the highest sales at Fuel Efficiency peaked at 1,118 while Engine size resulted in 959 and Horsepower 635. Therefore it can be said Fuel efficiency had the highest impact on car sales and is a factor largely considered by the public towards car purchase.

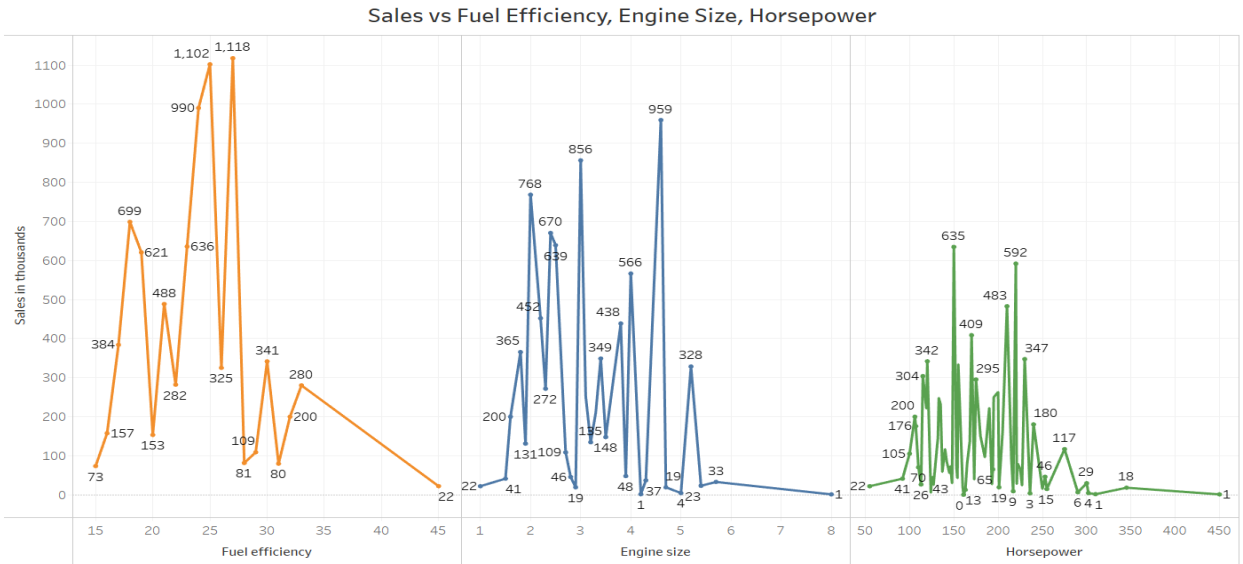

Figure 1: Depiction of Sales vs Fuel Efficiency, Engine Size, Horsepower

For the second inference we draw in this project we considered the question**, 'How do Car prices impact car sales?** The variables considered towards creation of the bar chart were Sales in thousands and Price in Thousands. From the bar chart obtained using the Car_sales.csv dataset (Figure 2), it can be deduced that car prices fairly impact car sales since the lower priced cars can be seen to be purchased more than the higher priced ones. However, the highest car sales comes from medium-priced cars of 26.94 thousand which sold about 540.6 thousand units which tells us that price is not the only factor buyers consider when making a purchase.

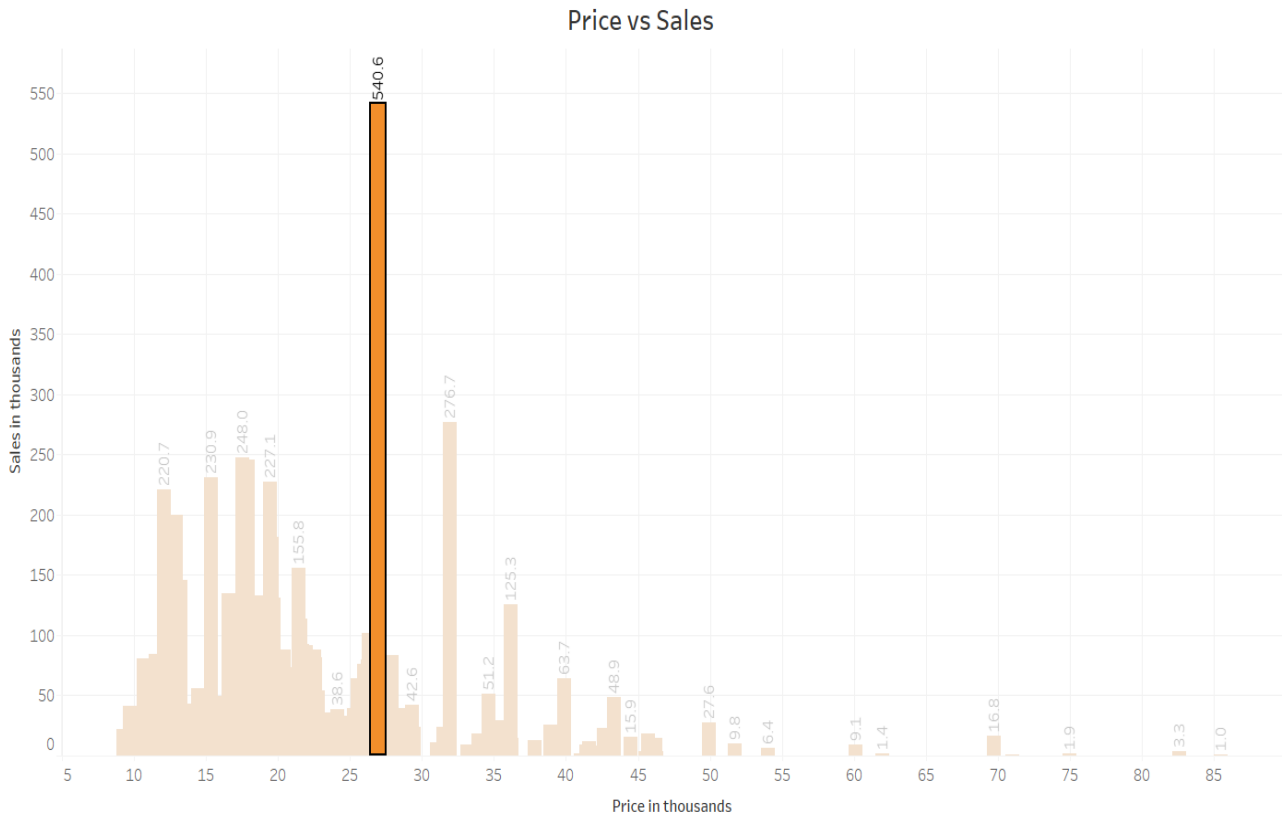

Figure 2: Depiction of Car Prices and Sales in Thousands

For the third inference in this project, we considered the question **'Which are the most expensive car models?'** The variables considered towards creation of the scatter plot were Manufacturer and Price in Thousands.

From the scatter plot obtained in (Figure 3), it can be deduced that top 3 most expensive cars were '**Mercedes-B'** at 476.3 followed by '**Dodge**' at 266.4 and '**Lexus**' at 264.3 Thousands. This implies that these top models may contain features or qualities which the less expensive models may not.

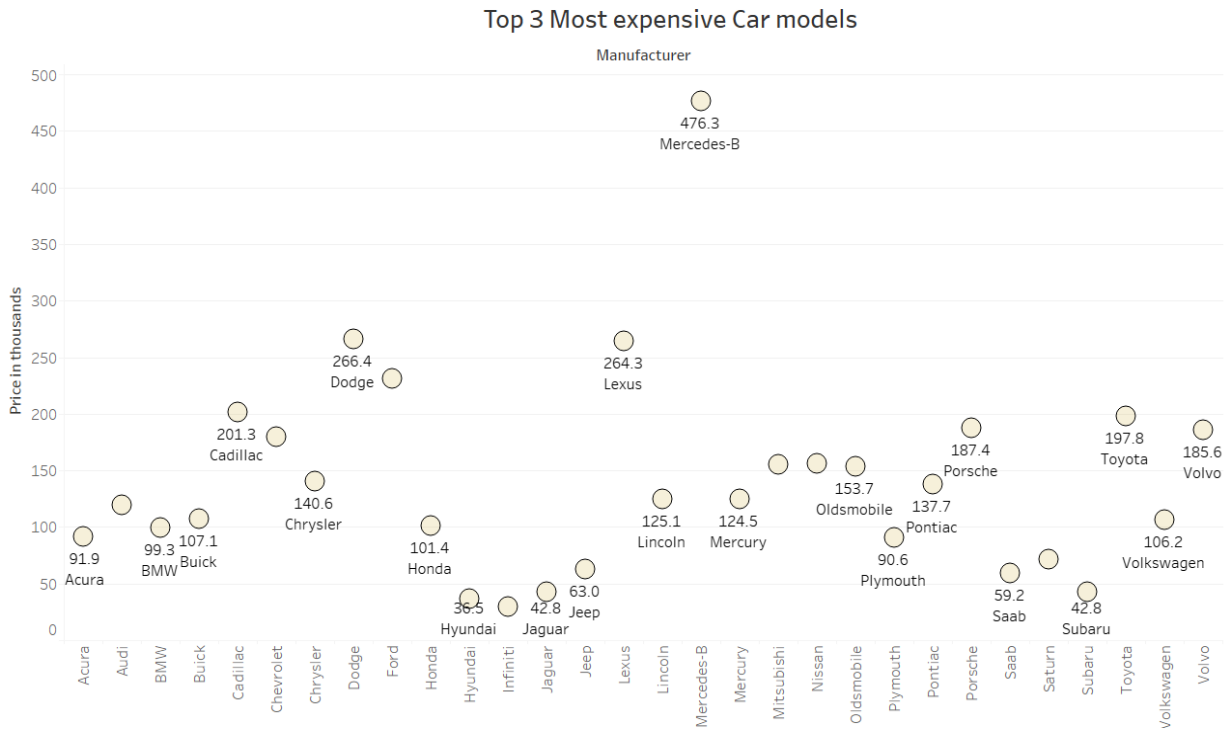

Figure 3: Depiction of the 3 most expensive car models

For the forth inference in this project we considered the question**, 'What car had the best resale value?'** The variables considered towards creation of the scatter plot were Year_Resale_Value and Price in Thousands. Resale value of cars is an important factor most buyers consider as it allows for them to make some money back whenever they plan on selling their vehicles. Additionally, cars without good resale values could pose a form of liability as the owners will struggle to profit or even sell them after use,

From the scatter plot obtained (Figure 4), it can be deduced that the '**Carrera Coupe'** had the highest resale value at 60.63 thousand and an initial cost price of 71.02 thousand. This shows

that the seller can still get 85% of their initial expense in purchasing the car back hence why the model is a popular choice among buyers.

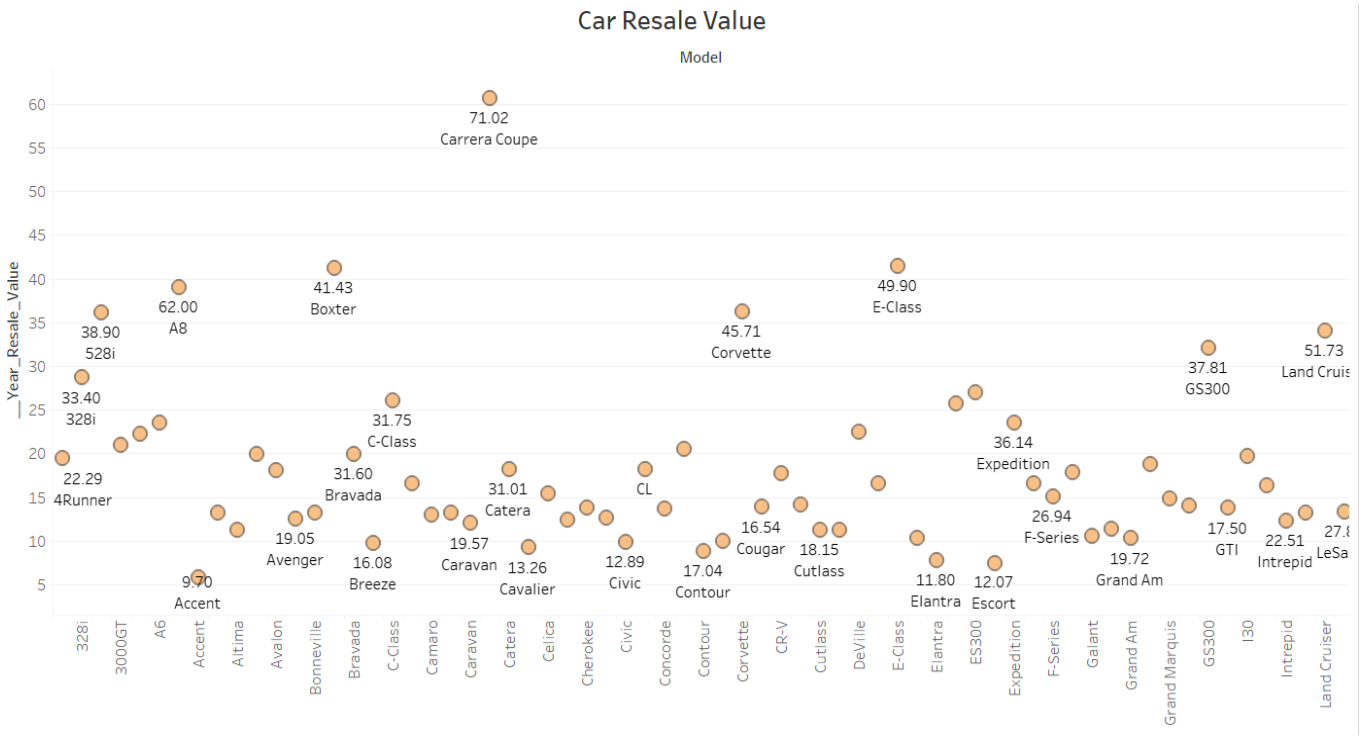

Figure 4: Depiction of car resale values.

For the final inference in this project we considered the question**, 'Latest Launch Impact on Car Sales'.** The variables considered towards creation of the table were Latest launch, Manufacturer and Sales in Thousands. Many car companies choose to launch cars to drive sales of their products since customers are always on the lookout for new car features (Kalia, 2020). The initiative take by these companies works sometimes and does not at other times given the insatiable nature of humans.

From the table obtained (Table 2), it can be seen that the launch of new models percentage increase in sales from the years 2008 to 2012. This could imply that other manufacturers were driven to launch new models given the growth in sales by their competitor manufacturers.

Table 2: Table of the percentages of children surviving in each passenger class

Latest lauch impact on Car Sales

	Latest Launch			
Manufactur..	2008	2009	2011	2012
Acura			1.274%	1.274%
Audi			1.274%	0.637%
BMW			1.274%	0.637%
Buick			1.911%	0.637%
Cadillac			1.911%	1.274%
Chevrolet			3.822%	1.911%
Chrysler			1.911%	2.548%
Dodge			3.185%	3.822%
Ford			1.274%	5.732%
Honda			0.637%	2.548%
Hyundai			0.637%	1.274%
Infiniti				0.637%
Jaguar				0.637%
Jeep			0.637%	1.274%
Lexus				3.822%
Lincoln				1.911%
Mercedes-B			5.732%	
Mercury	0.637%	0.637%		2.548%
Mitsubishi				4.459%
Nissan			4.459%	
Oldsmobile		0.637%	3.185%	
Plymouth			1.911%	0.637%
Pontiac			0.637%	3.185%
Porsche			0.637%	1.274%
Saab			0.637%	0.637%
Saturn			1.911%	1.274%
Subaru			0.637%	0.637%
Toyota			4.459%	1.274%
Volkswagen			3.185%	0.637%
Volvo			3.185%	0.637%

CONCLUSION

From analysis carried out on the Car_sales.csv dataset bar charts and tables were able to clearly show different insights that impact car sales with different companies For the first analysis, 'Fuel efficiency' had the highest and is a factor largely considered by public towards car purchase. For second analysis, we saw that car prices fairly impact car sales since the lower priced cars are usually purchased more than the higher priced ones. For the third inference, we saw top models may contain features or qualities which the less expensive models may not possess which drive sales. For the fourth analysis, we saw that car sellers can still get 85% of their initial expense back from selling the Carrera Coupe hence why the model is a popular choice among buyers. Finally, we saw manufacturers were driven to launch new models given the growth in sales by their competitors. Therefore, clearer insights on car sales incident have been drawn from the visualizations and tables shown in this report.

References

Myatt, Glenn J. “Making Sense of Data II: A Practical Guide to Data Visualization, Advanced Data Mining Methods, and Applications.” *Google Books*, John Wiley & Sons, 4 Mar. 2009, https://books.google.com/books/about/Making_Sense_of_Data_II.html?id=lFBqIpM-vuQC.

Kaila, S. (2020). How can Businesses Leverage Data Analytics to Influence Consumer

Purchase Journey at Various Digital Touchpoints? *Journal of Psychosocial Research, 15(2),* 699–714. https://doi.org/10.32381/JPR.2020.15.02.30

Nussbaumer, C. (2020, January 11). *Table vs Graph - The Visual Battle*. Storytelling With Data.

Retrieved September 25, 2022, from https://www.storytellingwithdata.com/blog/2011/11/visual-battle-table-vs-graph

Introduction to Data Mining. (n.d.). Retrieved September 25, 2022, from https://www-users.cse.umn.edu/%7Ekumar001/dmbook/index.php

https://www.kaggle.com/search?q=titanic+dataset+in%3Adatasets