Analysis Of Used Cars Information Using Data Visualization

By

Sumanya Bairagoni(11507719)
Sudhamshu Boddu(11548202)
Prathyusha Nemani(11548205)
Nikhitha Doragari(11513291)

1. **Goals and Objectives**:
   - **Motivation**: The idea of developing a project based on used cars has emerged from the need of cars for people who are on constant move to carry out their daily businesses. People who travel places very often find it difficult and expensive to buy and sell new cars every now and then. So, the concept of buying and selling used cars would help and serve a greater purpose in such circumstances. And having a analyzed data with respect to the cars under consideration would help greatly. The initial step is developing a model that analyzes whether the initial offer for a certain automobile is acceptable considering the facts supplied in the listing. Aside from that, the code will provide suggestions for similar automobiles that are less costly, have less miles, and seem to be more expensive.

   - **Significance**: This data collection is focused on used autos. We can display this data set for the user's convenience so that he can select a used automobile from among the variouscategories accessible. For instance, the user should select a vehicle by defining thenumber of kilometers he or she want to go in connection to the price. As a result, if the number of miles driven increases, the cost may decrease. As a result, the visualization is focused on the number of cars sold based on the vehicle's value, mileage, and brand, as well as the year it was registered. Because the data collection's owner is a German, the words automatic and automatic are spelled differently manually are not the same thing, and adhere refers to something else entirely.

   - **Objective**: Our project's purpose is to create an algorithm that can estimate the price of aused car based on a variety of factors (data) about that car. According to popular perception, an automobile loses 10% of its market value the moment it is driven off a lot. As a result, I would predict since one of the most important indicators is the maximum number of miles driven in the vehicle, as more driving wears the vehicle down. Furthermore, I believe that the automobile's brand (make) will play a factor in determining the price ofa used car, as certain brands are more expensive and may be of greater quality. Because some automotive features are really valuable, Consequently, I expected to run into some issues with multi-

collinearity. Larger autos, for example, are more likely to have larger engines and doors than smaller automobiles. The number of cylinders in an engine is inversely proportionate to its size

- **Features**: The data contains a complicated visualization relating to the sales of used automobiles and their significance, with manual automobiles accounting for more than three times the number of sales as automated cars, according with data visualization. Automobile providers' inventories must include a bigger quantity of manual automobiles than automated vehicles to accommodate such high demand. While many vehicle manufacturers use benzene, diesel, and LPG- based fuels, because none of the manufacturers have used LPG-based fuel in recent years, with the exception of Dacia and Ford. As a result, auto shops will need to stock more automobiles that run on benzene and diesel-based fuels. Gas stations must keep benzene and diesel fuel on hand to meet client demands. One of the most desired automobiles Audi, BMW, Ford, Mercedes-Benz, Opel, Volkswagen, Renault, and Peugeot are the brands. Car dealers with a larger inventory of vehicles from these manufacturers will notice a boost in sales. Volkswagen's Golf, Passat, and Polo models have outsold Opel's Astra and Corsa, as well as Mercedes-C- Benz's and E-Class models, which had been the most popular used vehicles for the previous decade. These models'/brands' marketing strategies might be tweaked to attract a larger number of customers by giving better deals, higher discounts, a longer warranty, and more free vehicle inspections, among other things.

Domain : The data type used by a column in a database is a simple definition of a database domain. This data type can be either a built-in type (such an integer or a string) or a custom type that defines data constraints.

  Here we used integer type of data to show pricing, year of manufacture, miles driven and model names by string datatype. The data type, such as integer, character, or decimal, determines domain integrity. The maximum data length that can be used. The upper and bottom limits of the range are defined.Any restrictions or limitations on the values that can be used. A entire ZIP+4 code or a full nine-digit code, for example, can be required in a US ZIP code field. The type of NULL support (whether or not an attribute can have a value of unknown or NULL).

Data Abstraction:

Dataset: This project's data was sourced from data world. All the information is saved in a comma-separated file with the extension ".csv.". The dataset gives overview of used cars sales. For this analysis I have used variables like Name, Price, Year of registration, model, miles, brand.

This dataset contains the following Columns:

- Name - The brand and model of the car.
- Year of registration - The year or edition of the model.
- Miles Driven - The total miles are driven in the car by the previous owner(s) in KM.
- Price - The price of the car (target).
- Brand- Brand of vehicle.

Task abstraction : Visualization of histogram: count of vehicle on y axis corresponding to a brand is plotted on x axis representing distribution of vehicle by brand.

Visualization of histogram: count of vehicle on y axis corresponding to a brand is plotted on x axis representing distribution of vehicle by brand

**Implementation using tools:** The tool used is Tableau 2020.4 version
  ➢ Description: Tableau is a data visualization and business intelligence application that can be used to report and analyze large amounts of data. It allows users to build various charts, graphs, maps, dashboards, and stories for the purpose of displaying and analyzing data and making business choices. Tableau is one of the most popular business intelligence applications because of its many unique and intriguing features. Tableau delivers robust data discovery and exploration, allowing users to quickly answer critical questions.

Preliminary Results for Analysis:

The Visualization methods that we used for the desired analysis are:
- Gantt chart
- Bubble chart
- Bar chart
- Text table

- ## Bar Graph:

This the dashboard of the 3 different worksheets which is combined, and the final visualization has end result comparing the price, brand, miles and year of registration where all are interlinked and visualized.
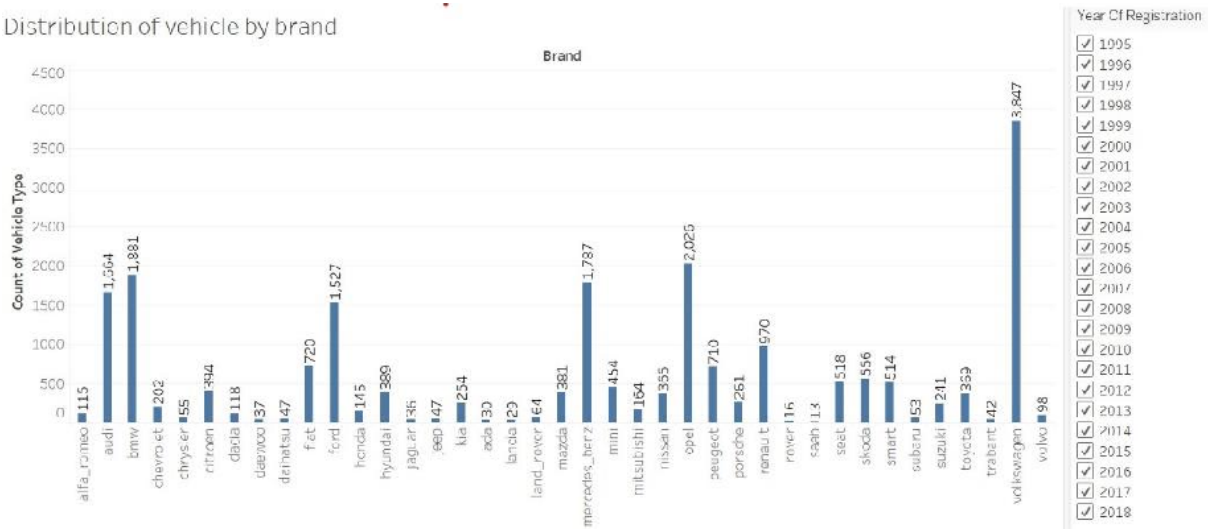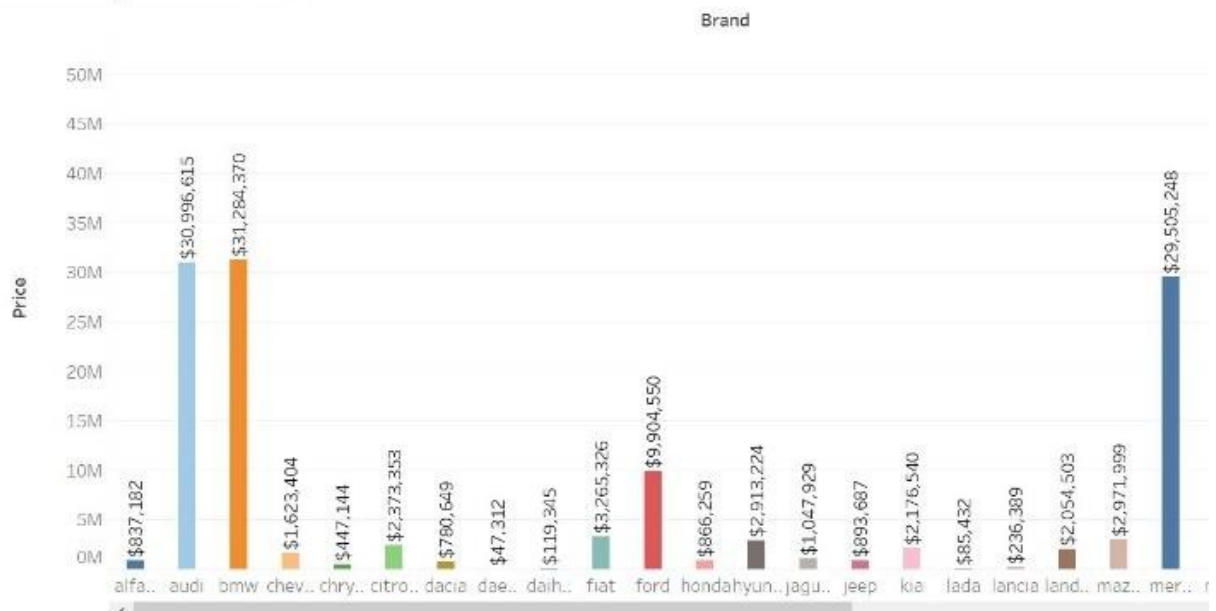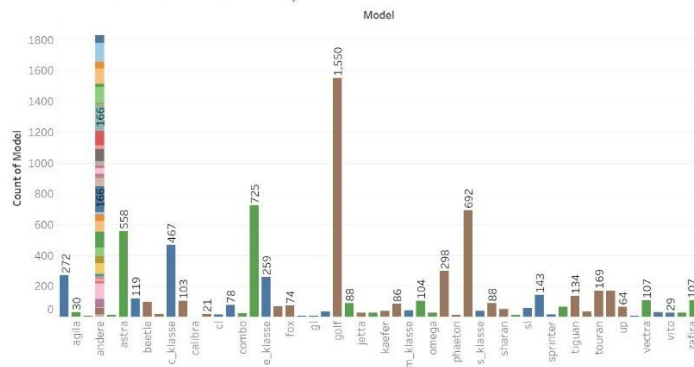
Visualization graphs

This is how we can find out the price of a car with respect to miles, brand, and year of registration. In this visualization the price of the car is estimated and visualized based on number of miles, brand, and year of registration. So, the price of the car may decrease if the number of miles is high, and the year of registration is too old.

Models sold in all brand across all years

Data depicting how many numbers of car models are sold on the total. In this visualization the model has many models that has been visualized and the count of golf that is Volkswagen has high number of counts.

Project Management:
Implementation status report
Work completed:

- Description: Developing this project involved a number of tasks and immense contribution towards collecting and analyzing the data. The initial step is developing a model that analyzes whether the initial offer for a certain automobile is acceptable considering the facts supplied in the listing. Aside from that, the code will provide suggestions for similar automobiles that are less costly, have less miles, and seem to be more expensive. The later stages involved making the data available in both text and visualized formats in a desired manner for the users.

- Responsibility:

Task	Person
Project vision, collecting, analyzing and presenting the data	Sumanya Bairagoni
Story creation, task distribution, allotting responsibilities, team meetings and project follow-up, project schedule	Nikitha Doragari
Dashboard overview, documentation and concept expertise, setbacks	Prathyusha Nemani
Code implementation and corresponding data visualization tasks	Sudhamshu Boddu

Contributions:

Team Member	Overall contribution to the project
Sumanya Bairagoni	25%
Nikitha Doragari	25%
Prathyusha Nemani	25%
Sudhamshu Boddu	25%

Work to be completed:

Description: Developing a software is an eternal process. We ought to add other features to the data such as including info ration related to vehicle type, fuel type, number of transmissions for the vehicle, power of engine, date crawled, seller type, moth of registration, whether the car is repaired or not, postal code of the car availability location, last seen details of the vehicle etc. Enhancing visualizations by more detailed and simplified data is among the other things that we are working on.

Responsibility:

Task	Person
Collecting more data with respect to fuel type of vehicle, gear model etc	Sumanya Bairagoni
Conducting phase wise reviews and development follow ups, story updations	Nikitha Doragari
Updating dashboard with new worksheets created, modeling setbacks if needed	Prathyusha Nemani
Enhancing visualizations and simplifying with more detailed data view and analysis	Sudhamshu Boddu

Issues/Concerns:

- New Data accumulation as large amounts of data is involved.
- Data cleaning if invalid data or bill values are present in such accumulated data
- Creating new visualizations in a more detailed manner without disturbing the old ones

References/Bibliography:

➢ Gareth, J., Daniela, W., Trevor, H., & Tibshirani, R. (2013). *An Introduction to Statistical.*

➢ Hurwitz, E., & Marwala, T. (2012). Common mistakes when applying computational intelligence and machine learning tostock market modelling. *arXiv preprint arXiv:1208.4429.*