



(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 4, April 2025

⊕ www.ijirset.com ⊠ ijirset@gmail.com 🖄 +91-9940572462 🕓 +91 63819 07438





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404654|

Vehicle Sales Analysis and Data Visualization

Ankita Mane, Shruti Pandharkar, Harshada Thite, Aboli Rathod, Sanika Mahajan

Dept. of Computer Engineering, DY Patil University, Ambi, Pune, India

Prof.Pankaj Shinde

Associate Professor, Dept. of Computer Engineering, DY Patil University, Ambi, Pune, India

ABSTRACT: In today's competitive automotive market, understanding sales trends and performance metrics is crucial for strategic decision-making. This project focuses on the analysis and visualization of vehicle sales data to provide comprehensive insights into sales performance. Utilizing advanced data analytics and visualization tools, we aim to uncover patterns and trends that can drive business growth.

The analysis covers key performance indicators (KPIs) such as total sales, average prices, and the number of vehicles sold over various time periods. By leveraging interactive dashboards, stakeholders can explore data through dynamic charts and graphs, facilitating a deeper understanding of sales dynamics. The visualizations highlight year-to-date (YTD) and month-to-date (MTD) sales figures, year-over-year (YOY) growth, and regional sales distributions. This project not only aids in identifying successful sales strategies but also pinpoints areas needing improvement. The insights gained from this analysis empower automotive businesses to make data-driven decisions, optimize inventory management, and enhance customer satisfaction. Ultimately, this comprehensive approach to vehicle sales analysis and data visualization serves as a valuable tool for achieving sustained business success.

KEYWORDS: Pandas, Matplotlib, Seaborn, Folium, Requests

I. INTRODUCTION

The prediction of car sales is an important and rewarding problem in current times. According to data obtained from the Government of India [1], the revenue generated from cars registered between 2009-2010 and 2015-16 has witnessed a spectacular increase of 34%. The number of cars in 2016 has reached 25,634,824. With a rise in new technologies and advancements, it is likely that sale of cars and the scope of this study will see a growth.Kuiper used a multivariate regression model to predict the price of 2005 General Motor (GM) cars. The results were satisfactory and can be repeated with accuracy. Support Vector Machines (SVM) were used by Listiani to predict the sale of leased cars. It is concluded thatSVM is The prediction of car sales is an important and rewarding problem in current times. According to data obtained from the Government of India [1], the revenue generated from cars registered between 2009-2010 and 2015-16 has witnessed a spectacular increase of 34%. The number of cars in 2016 has reached 25,634,824. With a rise in new technologies and advancements, it is likely that sale of cars and the scope of this study will see a growth. Kuiper used a multivariate regression model to predict the price of 2005. The number of cars in 2016 has reached 25,634,824. With a rise in new technologies and advancements, it is likely that sale of cars and the scope of this study will see a growth. Kuiper used a multivariate regression model to predict the price of 2005 General Motor (GM) cars. The results were satisfactory and can be repeated with accuracy. Support Vector Machines (SVM) were used by Listiani to predict the sale of leased cars. It is concluded that SVM.

II. METHODOLOGY

Data was collected from Kaggle. We have tried to make sure that the data collected was as recent as possible. In most countries, seasonal patterns are somewhat of a minor problem only, as this does not affect the purchase or selling of cars [1]. Now-a-days, with the technological advancement, Techniques like Machine Learning, etc are being used on a large scale in many organisations. These models usually work with a set of predefined data-points available in the form of datasets. These datasets contain the past/previous information on a specific domain.

Organising these datapoints before it is fed to the model is very important. This is where we use Data Analysis. If the data fed to the machine learning model is not well organised, it gives out false or undesired output. This can cause major losses to the organisation. Hence making use of proper data analysis is very important.





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404654|

1. Data Sources Sales-

- Data: Transaction records, including date, vehicle type, price, customer information, etc.
- Inventory Data: Information on available vehicles, specifications, and stock levels.
- Market Data: Industry trends, competitor pricing, and economic indicators.

2. Visualization-

- Dashboards: Interactive dashboards displaying key metrics and KPIs (e.g., total sales, inventory levels).
- **Reports:** Regular reports for stakeholders summarizing findings and recommendations.
- 3. Tools and Technologies-
 - Python (Pandas, NumPy), SQL, Matplotlib/Seaborn for Python.

4. Analysis and Insights-

- Descriptive Analytics: Sales trends, best-selling models, and seasonal patterns.
- Predictive Analytics: Sales forecasting using machine learning algorithms.
- Customer Segmentation: Analysis of customer demographics and preferences.

III. DATA VISUALIZATION TYPES

The most frequently used data presenting types will be bar graph or pie charts. This is most integral part of visualizing data and most common baseline for many graphical data representation. There is much more types which can be used by pairing to present the right set of data in the right format. Some of the examples include and not limited to Charts

- Tables.
- Maps.
- Dashboards.
- Pie chart.
- Bar chart.
- Levelled bar chart.
- Stacked bar chart.

We will use some of them to interpret the data which is use for this research of car selection process [2].

Proposed Block Diagram Quality Lead New Car Sales **Discuss Price Rang** dentify nd Selection of Ca Pre Owned Car Sales if lead wants a Mainte Update Price Generate Invoice Add Quotation Paper other Work and ervice Arrange delivery

IV. BLOCK DIAGRAM

Fig:1: Block Diagram Of Proposed System

ALGORITHM

Machine learning algorithms are used in data visualization to process and analyze datasets. These algorithms can perform tasks such as: classification, regression, clustering, and anomaly detection. Some tools used in data visualization include: Seaborn.

A Python data visualization library that provides a high-level interface for drawing statistical graphics Algorithm

IJIRSET©2025





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404654|

visualization is the graphical representation of algorithms and their execution. It's an educational tool that helps students, educators, and professionals understand how algorithms operate [8].

In our study, we studied and implemented three supervised machine learning algorithms to predict and estimate prices of cars. Data was into test and train and after necessary implementation steps, we found out that Linear Regression is able to most accurately predict the sales for this particular type of data-set. Any such data set that consists of industry-based most impactful features and a general mix of manufacturers will be accepted by the algorithm and a similar prediction can be replicated [4].

Machine Learning Algorithm-

For vehicle sales and data visualization, various machine learning Algorithms can be used, depending on the task you aim to accomplish [6].

1. Regression Algorithms-

• Linear Regression: Predict future vehicle sales based on past trends (e.g., sales vs. time).

• **Multiple Regression:** Evaluate the impact of multiple factors (e.g., price, marketing spend, and features) on sales [1].

2. Classification Algorithms-

- Logistic Regression: Classify vehicles into categories like "high demand" or "low demand."
- **Decision Trees/Random Forest**: Identify key factors affecting vehicle sales [2].
- 3. Clustering Algorithms-
- **K-Means Clustering**: Group vehicles into clusters based on sales trends or customer preferences.
- **DBSCAN**: Identify outlier vehicles with abnormal sales performance [7].

V. DATASET DESCRIPTION

The dataset used for this visualization contains historical_automobile_sales data representing automobile sales and related variables during the recession and non- recession periods. You can download here. The dataset includes the following variables:

- 1. **Date:** The date of the observation.
- 2. Recession: A binary variable indicating a recession period; means a recession, and 0 means it was normal.
- 3. Automobile Sales: The number of vehicles sold during the period.
- 4. **GDP:** The per capita GDP value in USD.
- 5. Unemployment_Rate: The monthly unemployment rate.
- 6. **Consumer_Confidence:** A synthetic index representing consumer confidence, which can impact consumer spending and automobile purchases.
- 7. Seasonality_Weight: The weight representing the seasonality effect on automobile sales during the period.
- 8. **Price:** The average vehicle price during the period.
- 9. Advertising_Expenditure: The advertising expenditure of the company.
- 10. Vehicle_Type: The type of vehicles sold; Supper mini car, Small family car, Medium family car, Executive car, Sports.
- 11. **Competition:** The measure of competition in the market, such as the number of competitors or market share of major manufacturers.





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404654|

- 12. Month: Month of the observation extracted from Date.
- 13. Year: Year of the observation extracted from Date.

By examining various factors mentioned above from the dataset, I aim to gain insights into how recessions impacted automobile sales for this company.

VI. CONCLUSION

Data visualization is the process of representing data in a graphical way in a clear and effective manner. It has emerged as a powerful and widely applicable tool for analyzing and interpreting large and complex data. It has become a quick, easy means of conveying concepts in a universal format. It must communicate complex ideas with clarity, accuracy, and efficiency. These benefits have allowed data visualization to be useful in many fields of study [1].

The proposed work Is based on the concept of classification and grouping the data, based on the applied classificationrules. On the basis of comparison rules the customers data is append into the associated cluster or can say group. On the basis of this analysis, the market strategy can be planed for specific location and for specific age group [3].

VII. RESULTS



Fig:1 From the image above we can see the disparity in sales in the two periods, recession and non-recession. The average total sales per year is 28,232. Even with an eye test, you could guess it was about \sim 30,000.



Fig:2 From the graph, we can see that the sale of Sports cars is extremely low compared to others. People mainly bought family cars during this period. Executive cars were totally absent.





www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

|DOI: 10.15680/IJIRSET.2025.1404654|



Fig:3 Drastic decline in sales of automobiles from the normal season compared to the recession period, but Sports cars and executive cars have the lowest purchases and are the most affected.



Fig:4 From this plot, it is evident that seasonality has not affected the overall sales. However, there is some difference in the sales in the first half of the year compared to the second half of the year.







[www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 4, April 2025

DOI: 10.15680/IJIRSET.2025.1404654

Fig:5 From the graphs above we can see that there is a direct correlation between the prices and the sales volumes, given that as the price increased, the volumes increased. Also, there were more sales when consumer confidence was higher. We can deduce a direct.



Fig:6 We can see that more money was spent during the recession. Almost 80% of the total advertising expenses were during the recession periods, to enable more customer reach which is a reasonable thing to do.

REFERENCES

TMzMTBhYW1zX3ZvbF8yMDNfamFudWFyeV8yMDIwX2E0X3AzNjctMzc1X3ByYXNoYW50X2d1cHRhX2FuZF9 uaWRoaV9zaW5naC 5wZGY&ntb=1

- [2] https://www.researchgate.net/publication/343156538_Data_Visualizatio n_for_Vehicle_Selection_Process
- [3] CarSalesData-
- set, retrieved from: https://www.kaggle.com/gagandeep16/caraleshttps://www.researchgate.net/publication/343156538
- [4] https://deepblue.lib.umich.edu/bitstream/handle/2027.42/112123/lycumi ch.pdf
- [5] https://www.sciencedirect.com/science/article/pii/S187705091730412X
- [6] https://www.coursehero.com/file/104413950/4-2-Scenario-Analysis-Cars-Solddocx/
- [7] https://www.researchgate.net/publication/352689200 A Descriptive A nalysis on Automobile Sales
- [8] https://www.mdpi.com/2071-1050/15/1/399
- [9] https://www.mililink.com/upload/article/1776193310aams_vol_203_january_2020_a4_p36775_prashant_gupta_and __nidhi_singh.pdfhttps://www.researchgate.net/publication/311597028_DATA_VISUALI_ZATION

[10] https://www.researchgate.net/publication/362841150_Intelligent_Vehicle_Sales_Prediction_Based_on_Online_Publi c Opinion and Online Se arch Index

- [11] https://medium.com/@adeleke.joe/eda-on-automobile-sales- 248a58e86597
- [12] Sample Car Data obtained from Kelly Blue Book. Available online at https://www.kbb.com









INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY





www.ijirset.com