Table of Content

* Introduction
* Data collection
* Data Cleaning and pre-processing
* Data Visualizations
* Limitations and challenges
* Conclusions
* Recommendations
* References

**Introduction**

This is a detailed report of the analysis of car sales by Chukwumaeze Henry. This report contains data source used, techniques and key findings from this analysis.

**Data Collection**

The dataset was gotten through a secondary source which is a website called Kaggle which is a popularly known as an online community where data engineers, scientists, analysts etc. can find datasets they want to use to build or perform any form of operation they want.

The dataset used for this analysis is a car sales report dataset gotten from the sales report section on Kaggle, created by Vasu Avashi.

**Data Cleaning and Pre-processing**

This dataset was fairly accurate but had some inconsistencies in various fields in the data set.

Firstly, I started off by correcting the irregular column name fields and updating each column with the accurate data type to ensure consistency.

However, the Engine column was one of those corrected and some fields with errors were corrected. The price field was formatted as a currency using dollar ($) as the official purchasing currency.

Also, the date column was also corrected and formatted appropriately. The model column had a data type issue which was corrected at last. Overall the data set was not void of errors but it was also fairly consistent.

Furthermore, I proceeded to subjecting this now refined dataset to some research questions which included What is the total number of unique customers that purchased at the company, what is the total revenue generated from car sales, visualize the all-time total revenue generated from each company, what is the highest selling models by transmission etc.

With the help of pivot tables, sort and filters, and lots of other functionalities in the excel tool, I answered and discovered some patterns, solutions and areas which needed looking into from the dataset.

**Data Visualization**

The following key findings were discovered from the analysis of the car sales report data set.

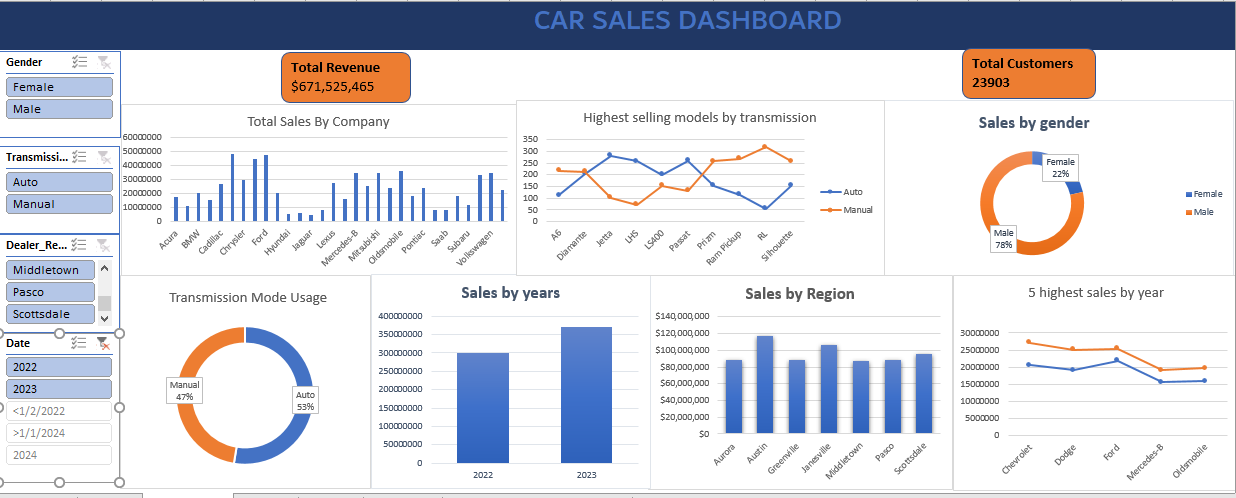
Firstly, after my analysis of the sales report, I discovered that a total of $671,525,465 was generated from the car sales between 2022 and 2023 from a total of 23906 customers.

The highest selling company was the Chevrolet with an estimated sale of $476, 552,265, closely followed by Ford. A total of 5108 females purchased cars from the company while 18798 males purchased from the company.

Males spent a total of $527,085,194 which is 78% of the total sales while females spent a total of $144,440,271 which takes up the remaining 22% of the total sales.

More Auto transmission mode were purchased by the customers than manual transmission mode cars. Auto had 53% of the total sales while manual had 47% of that. Between then 7 regions, Austin had the most sales of out of the 7 regions. The highest selling company was Chevrolet and closely followed by Ford while the worst was Hyundai. The total sales of 2022 were $300,340,345 which increased in 2023 to a total of 371,185,120 a 19% increase from the previous year.

Using various charts to visualize these tables helped in showing us a better and retrospective view of our dataset. Which is depicted below.



**Fig 1.0**

**Limitations and Challenges**

The main stumbling block from this dataset was the absence of age, reviews, quantity in stock and quantity sold. These missing data fields makes it hard to implement inventory management, group sales by age in order to ascertain the age range of customers also the missing reviews will leave the company and its shareholders unaware of the level of customer satisfaction.

**Conclusion**

In conclusion, after cleaning, calculating and analyzing the car sales dataset, I noticed that an increase in purchase of manual cars and also an increase in purchase by the females spiked up the sales in 2023 which increased by 19% from the 2022 sales.

**Recommendations**

My recommendation to the car sales company would be to increase the production and sales of manual transmission cars because they are more affordable to the customers. From the dataset, the average annual income of those that buy auto mode cars is higher than that of the manual modes. More production of manual and affordable cars will increase purchase. Also, regions like Austin, Janesville and Scottsdale should get considered as locations for any potential expansion in future due to the high concentration of potential customers which can in turn bring more customers.

**Reference**

* Car sales report, 2023 February, Kaggle.com.
* Microsoft Excel 2022