**Marketing Analysis Final Project**

**Prepared by Fatih Ayaz**

Table of Contents

[About Data Set 2](#_Toc165277467)

[Explatory Data Analysis 2](#_Toc165277468)

[Analysis Questions 3](#_Toc165277469)

[Insight: What are the top brands customer trust to buy? 4](#_Toc165277470)

[Insight: Does it really matter automatic or manuel transmission? 5](#_Toc165277471)

[Insight: What are the most trusted cars by consumers? 6](#_Toc165277472)

[Insight: What are the mostly sold years of the cars? 6](#_Toc165277473)

[Insight: What are the most selected colors of the cars? 7](#_Toc165277474)

[Insight: What is the seasonal trend? 9](#_Toc165277475)

[Insight: Which kilometers are mostly selected to buy by the customers? 9](#_Toc165277476)

[Insight: Based on top 5 sold brand which colors mostly selected by the customers? 11](#_Toc165277477)

[Insight: What are the most selected car body types? 11](#_Toc165277478)

[Insight: What are the most car buying states? 12](#_Toc165277479)

[Insight: What are the customers prefer interior car colors based on car color? 14](#_Toc165277480)

[Prediction: How car color and body type affect the sale price? 16](#_Toc165277481)

[Summary 20](#_Toc165277482)

# About Data Set

Vehicle data set had been downloaded from the below link. Basically this data set 16 features and 558837 records. After cleaning some of the null values records are getting less but still there are more than 500K records exist

<https://www.kaggle.com/datasets/syedanwarafridi/vehicle-sales-data>

# Explatory Data Analysis

This data set downloaded from kaggle and all together 16 features we have. What are we going to analyze here that the customer behavior on buying a car based on many reasons such as odometer, brand, model, year, color, interior, and demographic information as well

Current Dataset without wrangling:

Shape: (558837, 16)

A screenshot of a computer

Description automatically generated

After celaning null Values:

Shape: (472325, 16)

Feature descriptions as seen below:

* Year: Car build year
* Make: Car Brand
* Model: Car model
* Trim: submodel of the car. Such as Kia Sorento LX. Mostly related with the engine
* Body: Body of the car
* Transmission: Transmission information either automatic or manual
* State: State information
* Odometer: Kilometer
* Color: Color of the car
* Interior: Color of the interior
* Seller: Seller
* Sellingprice: Selling POrice
* Saledate: Sale Date

A screenshot of a computer

Description automatically generated

Total counts before and after removing null balues:

A screen shot of a computer

Description automatically generated

# Analysis Questions

In this section we are going to find some answers upon customers’ behavior buying cars based on many features.

## Insight: What are the top brands customer trust to buy?

As we can see that the top 10 selected car brands are Nissan, Ford, Toyota, Honda, Chevrolet, and BMW. Nissan is the lead with the Altima model. Followed by Ford Fusion and F-150 truck.

This report say us top-10 model are belong to sedan models but only F-150 except because it is the legend truck in the USD when it was first introduced. It had sold 249 count and getting the 3rd place in the list.

A screen shot of a computer

Description automatically generated

A graph of sales

Description automatically generated

## Insight: Does it really matter automatic or manuel transmission?

As per our report seems that it is not a big issue to select manuel or automatic transmission when buying a car. There might be several reasons fort hat:

* Except a few metropolitan cities traffic load is less than %50
* Manuel tramission are better on fuel economy
* Manuel transmission less expensive than automatic transmission

A screenshot of a computer

Description automatically generated

A blue and pink rectangular bars

Description automatically generated

## Insight: What are the most trusted cars by consumers?

Top 3 brands of the buying frequency fort he carsa re Ford, Chevrolet, Nissan, Toyota, and Dodge. This result will Show us consumers are selecting either Japanese car sor American cars. Mostly Japanese cars has severe advatages based on America carsa re:

* They are fuel economic
* They have less engine based on American cars
* Parts are less expensive than American cars

A graph of average cost of cars

Description automatically generated with medium confidence

## Insight: What are the mostly sold years of the cars?

Based on the dataset most carsa re build in 2012 and 2013. Based on the below CNN news it was related with the market. Market was suddenly down after 2007 and started to sales back again end of 2012 and beginning 0f 2013. And we can see that it was back to drop again 2015.

<https://money.cnn.com/2014/01/03/news/companies/car-sales/>

A white background with black text

Description automatically generated

A graph showing the distribution of car years

Description automatically generated

## Insight: What are the most selected colors of the cars?

Consumers rae mostly select the main colors like black, White, silver, and gray car colors. Also I used to find another reference link abot this as well

<https://www.germaincars.com/most-popular-car-colors/>

A screenshot of a computer program

Description automatically generated

A graph of blue and white lines

Description automatically generated

## Insight: What is the seasonal trend?

According to graph carsa re mostly bought in december, jabuary, february and march. These are the new models are mostly announced and old models are sold cheaper then the new ones. Also summer time hits the sale and I think this might be related with holiday season

A graph of blue bars

Description automatically generated

## Insight: Which kilometers are mostly selected to buy by the customers?

People mostly interested to buy less than 50K Km on the car. 230K people selects less than 50K, and 130K people slect between 50K – 100K KMs.

A screenshot of a computer

Description automatically generated

A graph of blue bars

Description automatically generated with medium confidence

## Insight: Based on top 5 sold brand which colors mostly selected by the customers?

As per reports we can see that top 2 colors on mostly sold brands are black and White colors.

A graph of different colors

Description automatically generated

## Insight: What are the most selected car body types?

As we can see from the graph that mostly US car market consumers selects the sedan cars. And it has been followed by SUV. Hatchback and VANs are rarely selected. This will give an idea about the culture I guess. Middle east people mostly selects the SUV models.

A screenshot of a computer

Description automatically generated

A graph of a car body type

Description automatically generated

## Insight: What are the most car buying states?

First 2 states are California and Florida. This will give an idea that these state consumers are very offenly change their cars. And also financially in good situation.

A screenshot of a computer

Description automatically generated

A graph with different colored lines

Description automatically generated

## Insight: What are the customers prefer interior car colors based on car color?

As per dataset if the car color is black then interior color is selected as be,ge or black mostly. If the car color is White then beige and black is selected. This gives us an idea that interior color selection mostly dark colors and not related with car colors

A screenshot of a computer program

Description automatically generated

A graph of different colors

Description automatically generated

## Prediction: How car color and body type affect the sale price?

As per dataset I have taken the average car sale price and use color and body features as an input. Result is compared the cars below the mean of the sale or above the sale of the price.

A screenshot of a computer program

Description automatically generated

A diagram of a confusion matrix

Description automatically generated with medium confidence

After having 0.62 accuracy model reevaluated with CV. Results are as seen below

A screenshot of a computer code

Description automatically generated

A graph of a graph

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated

**Summary of Results:**

Threshold: Classifies selling prices into "Above Threshold" and "Below Threshold" categories is the mean selling price of all cars in the dataset.

**Model Evaluation Metrics:**

Accuracy: Represents the proportion of correctly classified instances out of all instances. In this model, the accuracy is 63.25%

Precision: Measures the accuracy of positive predictions.

* "Below Threshold" class: 67%
* "Above Threshold" class: 57%

Recall: Measures the ability of the model to find all the relevant cases within a dataset.

* "Below Threshold" class: 72%
* "Above Threshold" class: 51%

F1 Score: The harmonic mean of precision and recall, providing a balance between precision and recall. A higher F1 score indicates better performance. F1 scores of 69% for "Below Threshold" and 54% for "Above Threshold" suggest moderate performance for both classes

* "Below Threshold" class: 69%
* "Above Threshold" class: 54%

Confusion Matrix:

* True Positives (TP): 45,431 instances correctly predicted as "Below Threshold"
* False Positives (FP): 17,769 instances incorrectly predicted as "Below Threshold"
* True Negatives (TN): 23,496 instances correctly predicted as "Above Threshold"
* False Negatives (FN): 22,285 instances incorrectly predicted as "Above Threshold”

Overall:

* The model's accuracy is approximately 63.25%.
* It performs better in predicting instances classified as "Below Threshold" compared to "Above Threshold," as indicated by higher precision, recall, and F1 score for the former.
* The weighted average of accuracy, precision, recall, and F1-score across both classes is approximately 63%, indicating moderate overall performance

Therefore, based on the provided evaluation results, we cannot confidently conclude that vehicle color and body type do not affect the sale price.

# Summary

As a summary after checking dataset consumer behavior as seen below:

* People are having more trust in Japanese cars. Durability, price, price of the parts might be affecting on this.
* Consumers doesnt care the car either manual or automatic tramission. This is not a marketing related item to discuss
* Country economic situations are directly affecting the car market
* Consumers selects the dark car color mostly like black, silver, nad gray. White is also in second. Seems that shiny colors are not selected mostly.
* Mostly selecting end of the, beginning of the year or summer time to sell and buy a car
* Less than 100K KMs carsa re selected in the market
* Sedan and SUV models are the two body types fort he interest
* California and Florida is the top states and followed by Texas as well for shanging cars regularly. These are the rich states in the US
* Interior colors is not related with the color of the car to be selected

From the marketing perspective I would suggest my clients to do this business:

* Sell less tahn 100K KM car and only stick to sedan and SUV models. Try to stick to Japanese cars except Ford and regular engine could be better.
* Target to richest states or consumers where they regularly change their cars to open the business.
* Do not buy light interior color cars, you can sell hardly