**Used Car Selling Price - Linear Regression**

**Artificial Intelligence CS-617-A**

**Avalons**

****

**Sacred Heart University**

School of Computer Science & Engineering

The Jack Welch College of Business & Technology

Submitted To:

**Dr. Reza Sadeghi**

**Fall 2022**

**Project Report of Used Car Selling Price - Linear Regression**

**Team Name**

Name of the Team **Avalons**

**Team Members**

1. Sambasiva Rao Chennamsetty [chennamsettys@mail.sacredheart.edu](mailto:chennamsettys@mail.sacredheart.edu)(Team Head)
2. Arif Pasha Shaik [shaiks11@mail.sacredheart.edu](mailto:shaiks11@mail.sacredheart.edu)(TM)
3. Jagadishwar Reddy Velma [velmaj@mail.sacredheart.edu](mailto:velmaj@sacredheart.edu) (TM)
4. Sai Hrithik Peddi [peddis4@mail.sacredheart.edu](mailto:peddis4@mail.sacredheart.edu) (TM)
5. Vamsi Kiran Kakkera [kakkerav@mail.sacredheart.edu](mailto:ravulat2@mail.sacredheart.edu) (TM)

**Description of Team Members**

1. **Sambasiva Rao Chennamsetty**

I completed my Bachelor’s in Information Technology. I had 3+ years of experience as a full-stack developer with Java programming as a backend. I like to work with a team with more commitment to work.

1. **Arif Pasha Shaik**

I have completed my Bachelor’s in Information Technology, I have done a couple of internships on Visual Basic .net, and Business Analytics: Data mining and Data warehousing.And I love working in a team that has its full dedication.

1. **Jagadishwar Reddy Velma**

I hold 7+ years of experience in SQL Database Administration. I am here to learn and improve better development skills which help me to become an extensive experienced Core Developer.

1. **Sai Hrithik Peddi**

I am a graduate student at sacred Heart University. I have completed my Undergraduate in Computer Science. After, I worked as an Android Developer at Sensorise Digital services for 6 months. I’m very passionate about my work role.

1. **Vamsi Kiran Kakkera**

I have done my Bachelor's degree in the stream of computer science. I'm having work Experience of 2.5 years in the AWS cloud as an Associate Developer. I've chosen this team as they are very coordinative and discuss everything with the team members.

**Table of Contents**

[**1 Introduction 3**](#_Toc119087586)

[**1.1 Research Question 3**](#_Toc119087587)

[**1.2 GitHub Repository 4**](#_Toc119087588)

[**2 Dataset Description 4**](#_Toc119087589)

[**2.1 URL of Dataset 4**](#_Toc119087590)

[**2.2 Dataset Explanation 4**](#_Toc119087591)

[**2.3 Features of Dataset 4**](#_Toc119087592)

[**3 Related Work 4**](#_Toc119087593)

[**3.1 Pro’s 5**](#_Toc119087594)

[**3.2 Con’s 5**](#_Toc119087595)

[**4 Project Plan 5**](#_Toc119087596)

[**4.1 Data Preprocessing 5**](#_Toc119087597)

[**Data Cleaning 6**](#_Toc119087598)

[**5 GitHub Repository 12**](#_Toc119087599)

[**6 References 13**](#_Toc119087600)

# Introduction

As the world evolving in all directions significantly, the economic gaps between the people are still exist. The livelihood of different people from different financial backgrounds are changing a lot. When it comes to the comfortable travel the cars are playing a vital role. Also, considering the COVID pandemic, most of the lower- and middle-income group of people also attracting to travel in a safe environment and not willing to choose public transport.

* At the same time the car manufacturers also increased the price of the new cars, which is directly affecting the buying capability of low-income group people.
* Hence, most of the people are looking at the used cars now.
* There are few people who cannot afford to buy new luxury car, but they wish to travel in it. For those, this used cars are the sunlight in dark. [1]
* This used cars has become an opportunity for the business. And it's going to generate a decent revenue for business as well.

## Research Question

* Which variables are significant in predicting the price of a used car?
* How well those variables describe the price of a car?

## GitHub Repository

[**https://github.com/samba-chennamsetty/used-car-selling-price-linear-regression**](https://github.com/samba-chennamsetty/used-car-selling-price-linear-regression)

# Dataset Description

## URL of Dataset

[Old Car Selling Price with Linear Regression | Kaggle](https://www.kaggle.com/code/gauravduttakiit/old-car-selling-price-with-linear-regression/data?select=car+data.csv) [2]

## Dataset Explanation

* This dataset contains information about used cars listed on [www.cardekho.com](http://www.cardekho.com) [3]
* This data can be used for a lot of purposes such as price prediction to exemplify the use of linear regression in Machine Learning.

## Features of Dataset

The columns are in the given dataset is as follows:

* + - 1. **Car\_Name:** This column should be filled with the name of the car.
      2. **Year:** This column should be filled with the year in which the car was bought.
      3. **Selling\_Price:** This column should be filled with the price the owner wants to sell the car at.
      4. **Present\_Price:** This is the current ex-showroom price of the car.
      5. **Kms\_Driven:** This is the distance completed by the car in km.
      6. **Fuel\_Type:** Fuel type of the car.
      7. **Seller\_Type:** Defines whether the seller is a dealer or an individual.
      8. **Transmission:** Defines whether the car is manual or automatic.
      9. **Owner:** Defines the number of owners the car has previously had.

# Related Work

## Pro’s

The advantages we have over the other related works are

* We are following Linear Regression to make it to the point for easy analysis.
* Considering the best prediction relational fields.
* Portraying the many visualities of impact with each feature.
* Plan to build multiple models based on companies.
* We use better data-cleaning techniques.

## Con’s

The disadvantages we have over the other related works are

* We don’t have multiple regression i.e, based on many features compared to the source project we referred.

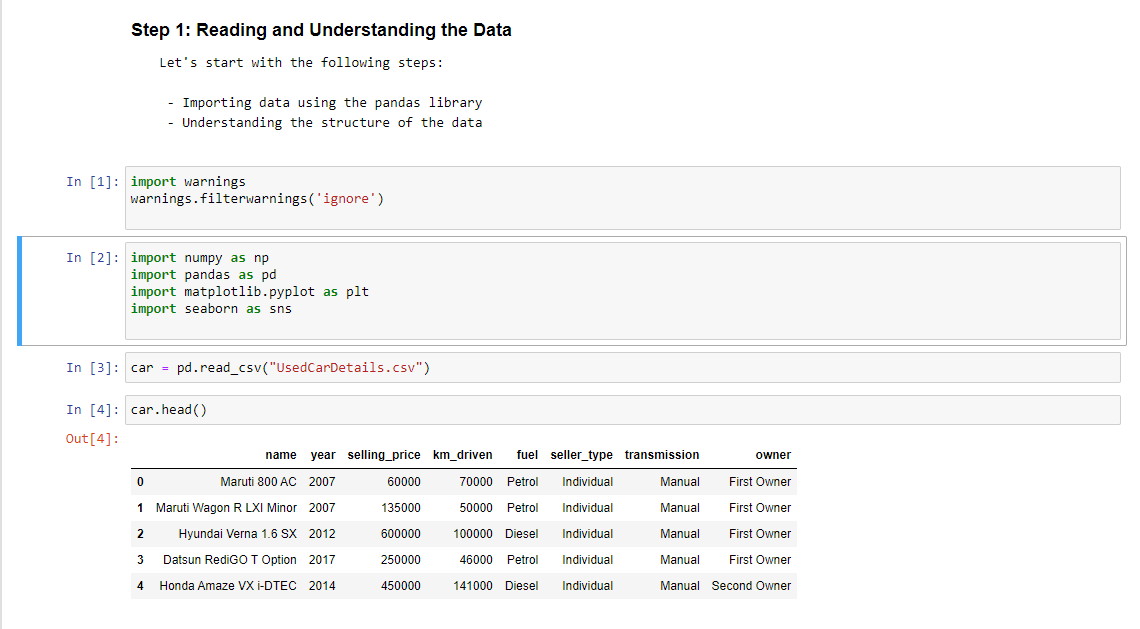
# Project Plan

The project plan has the below steps in it.

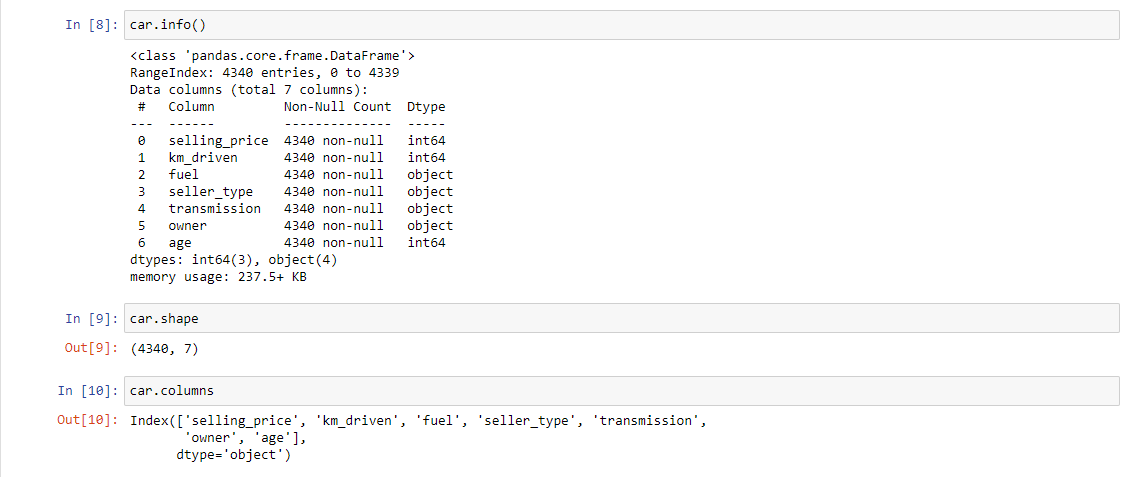
1. Data-preprocessing
2. Model building
3. Optimizing Model
4. Model Evaluation

## Data Preprocessing

We import the dataset initially as below and look for the head rows in the dataset



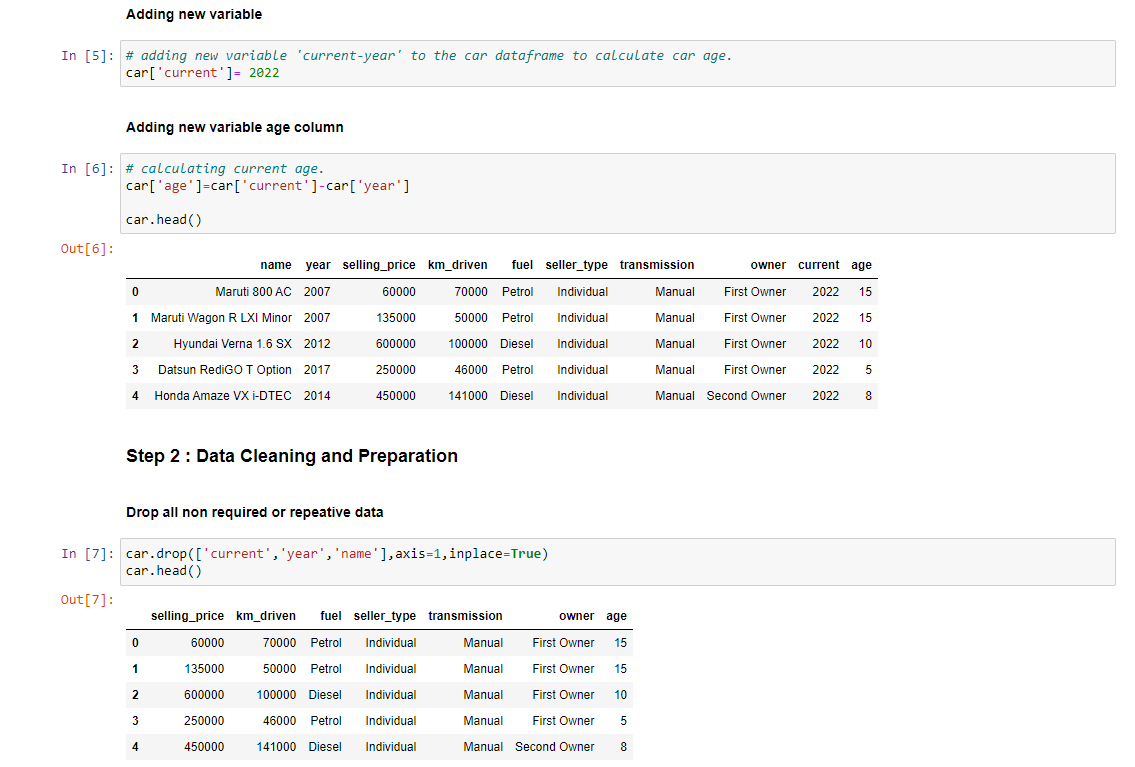
* Using the shape function, we find the number of rows and columns in the dataset.
* We use columns function to view the columns in the function.
* We use info function to know all the details of the car data set with their datatype.



## Data Cleaning

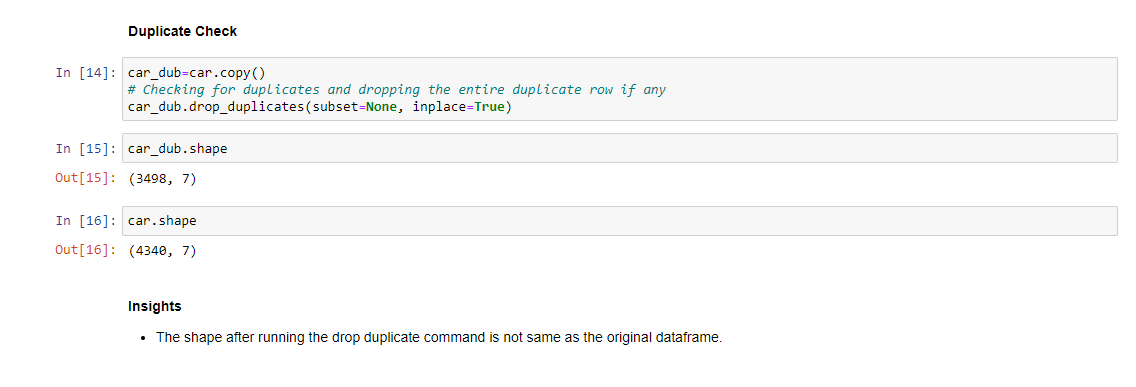
Adding a new variable for calculating the age of the car.

As part of this we clean the unwanted data and make the data right and good for the model with error free.

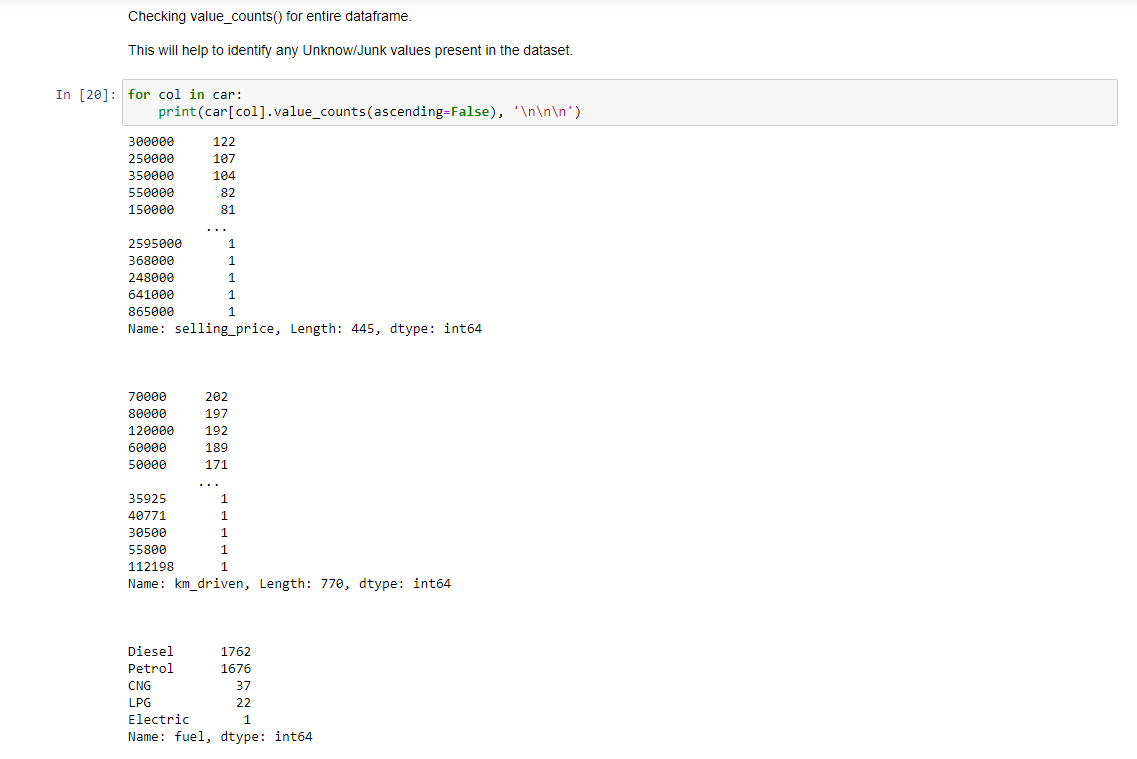


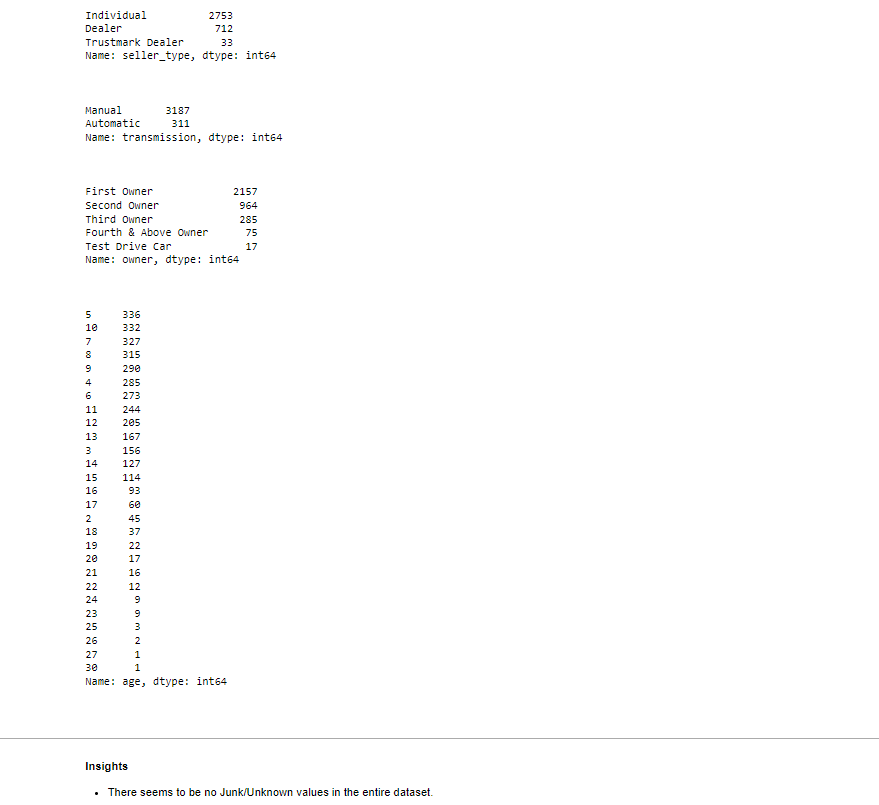
**Duplicate Data Check**

Checking if there is any duplicate data and dropping the entire duplicate row if any



**Identifying junk values**

****

****

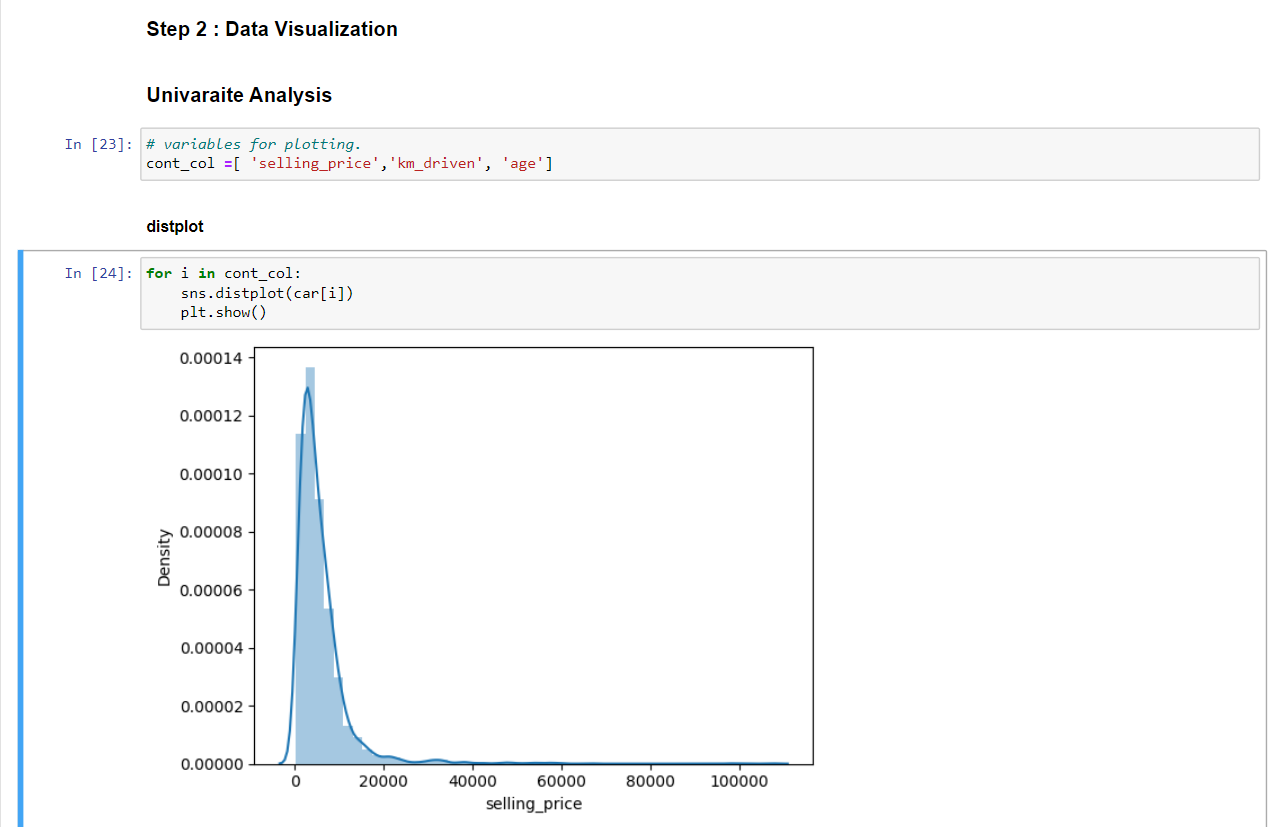
\* We found that there is no Junk or Unknown values exists in the data set.

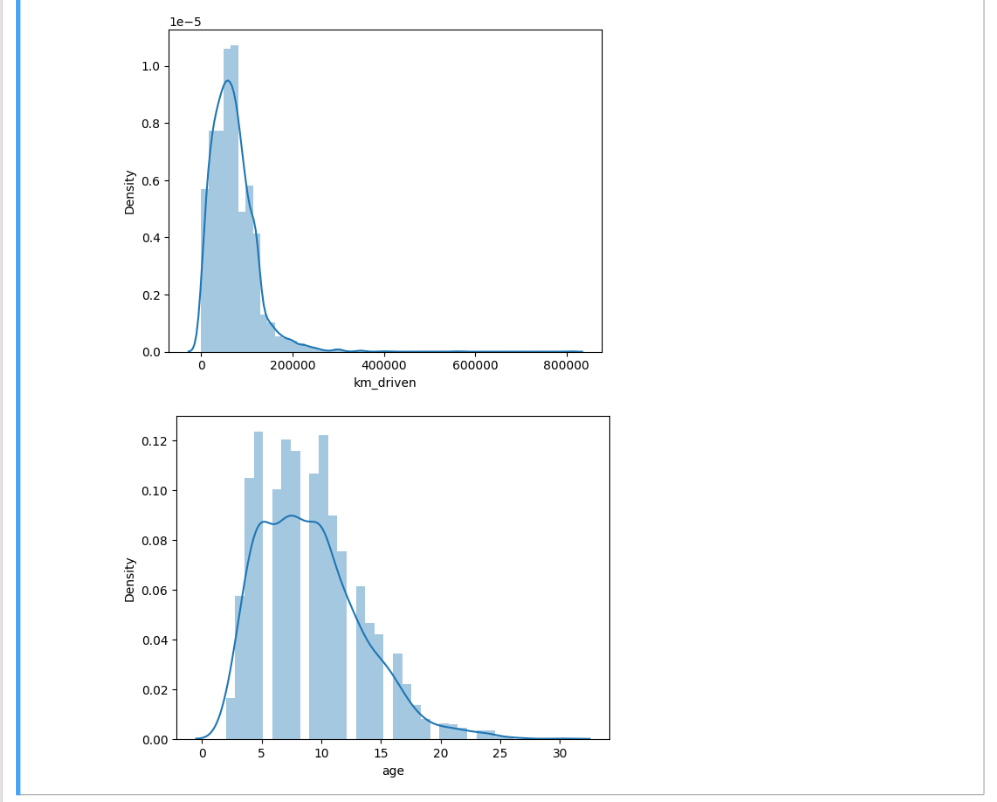
**Data Exploration**

**Univariate Analysis:**

Univariate analyses are used extensively in quality-of-life research. Univariate analysis is defined as analysis carried out on only one (“uni”) variable (“variate”) to summarize or describe the variable. However, another use of the term “univariate analysis” exists and refers to statistical analyses that involve only one dependent variable and which are used to test hypotheses and draw inferences about populations based on samples, also referred to as univariate.

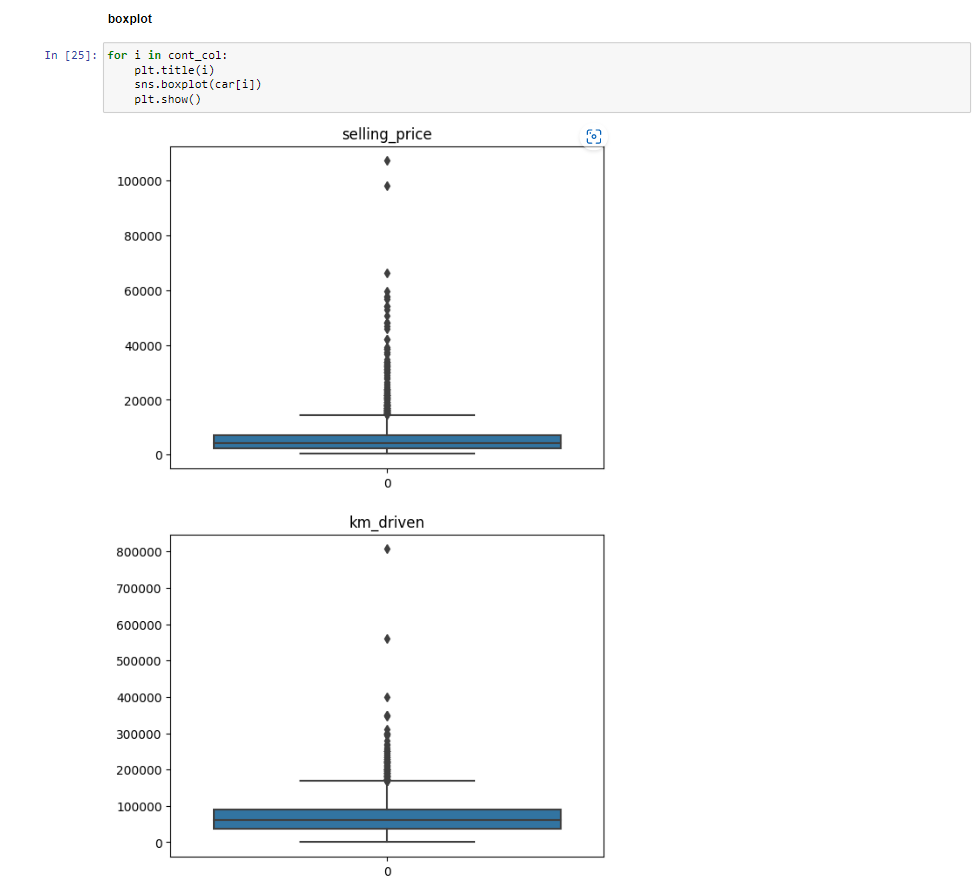
We find the univariate using distplot and boxplot graphs with below code. Here we’re using only uni one feature for the analysis.

****

****

Using distplot function Univariate Analysis has been made which gives a similar kind of distribution, some features are showing nearby normal distribution while some are skewed.

**Boxplot**

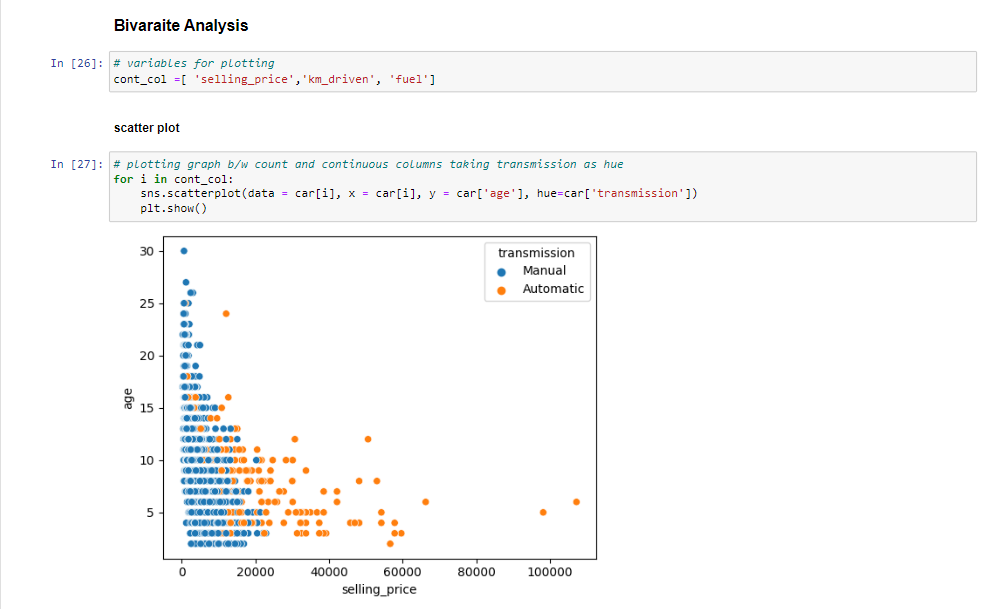


**Bivariate Analysis:**

Bivariate analysis refers to the analysis of two variables to determine relationships between them. Bivariate analyses are often reported in quality-of-life research. For an excellent example of research that utilizes bivariate analyses and demonstrates how the results of bivariate analyses can be used to inform furthermore complex analyses.

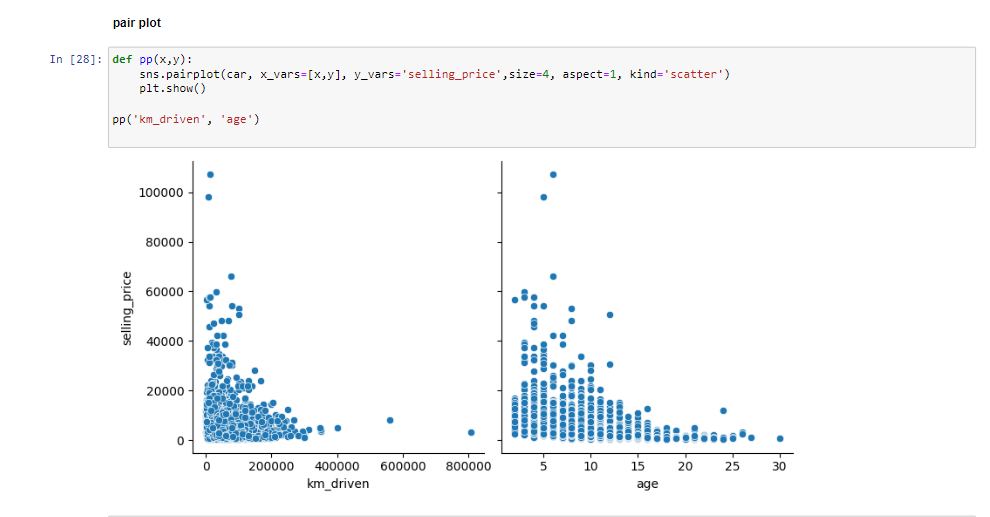
We find the relation between Selling Price and Car age which is bi with scatter plotting

Scatter Plot:



**PairPlot:**

Here we took selling price and compare it with km driven and age of car.



# GitHub Repository

**<https://github.com/samba-chennamsetty/used-car-selling-price-linear-regression>**

# References

[1] <https://www.kaggle.com/code/gauravduttakiit/old-car-selling-price-with-linear-regression>

[2] <https://www.kaggle.com/code/gauravduttakiit/old-car-selling-price-with-linear-regression/data?select=car+data.csv>

[3] <www.cardekho.com>