

# AUTOMOBILE SPECIFICATION DATA ANALYSIS REPORT

---



Submitted By:

Name : ANJANA P NAIR

Registration Number : 12209538

Section :K22UN

Roll Number : 10

# Table of Contents

---

1. Introduction
2. Problem statement
3. Solution approach and Abstract
4. Methodology
5. Result and analysis
6. Conclusion
7. Reference

# 1.Introduction

---

The Exploratory Data Analysis (EDA) project on Automobile Specification Data Analysis involves analyzing a dataset containing various technical specifications and performance metrics of automobiles. The objective is to identify patterns, trends, and relationships among key features such as engine size, horsepower, weight, fuel efficiency, and pricing. Through statistical analysis and visualization techniques, this EDA aims to uncover insights that can assist stakeholders, including manufacturers, consumers, and insurers, in making informed decisions related to vehicle performance, pricing, and market strategies. The process highlights the importance of data preprocessing, feature analysis, and pattern recognition in deriving meaningful conclusions from complex datasets.

## 2.Problem Statement

---

The problem statement for the Exploratory Data Analysis (EDA) on Automobile Specification Data Analysis is to investigate the relationships between various automobile specifications (such as engine size, horsepower, weight, and fuel efficiency) and their impact on pricing and performance. The goal is to uncover patterns, identify key factors influencing vehicle pricing, fuel efficiency, and market trends, and provide actionable insights for stakeholders like manufacturers, consumers, and insurers to make data-driven decisions in the automotive industry.

# 3.Abstract

---

This EDA project focuses on analyzing a dataset of automobile specifications to uncover insights into the relationships between vehicle attributes and key outcomes like pricing, fuel efficiency, and performance. By exploring features such as engine size, horsepower, and weight, the analysis identifies patterns and trends that can guide decisions for manufacturers, consumers, and insurance companies. The project utilizes statistical and visualization techniques to provide a deeper understanding of how automobile specifications influence market dynamics and performance metrics.

# 4. Methodology

---

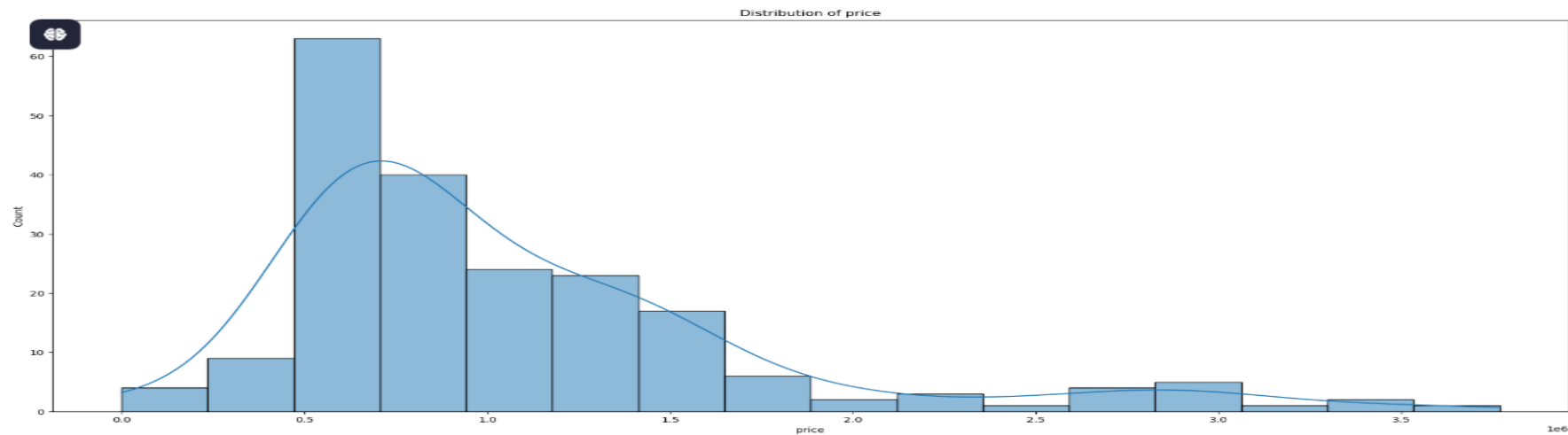
The methodology for the EDA project on automobile specification data involves the following steps:

- 1.Data Collection and Preprocessing:** Obtain the dataset and handle missing values, data transformations, and outlier detection to clean the data.
- 2.Descriptive Analysis:** Summarize the data using descriptive statistics like mean, median, and standard deviation for numerical features, and frequency distributions for categorical variables.
- 3.Visualization:** Use histograms, box plots, scatter plots, and correlation matrices to visually explore relationships between vehicle specifications and key performance metrics.
- 4.Bivariate and Multivariate Analysis:** Examine pairwise relationships between variables and perform correlation analysis to identify key interactions.
- 5.Outlier Detection and Feature Engineering:** Identify and treat outliers, and create new features to enhance the dataset.
- 6.Interpretation and Insights:** Summarize the findings and derive actionable insights to inform decision-making in the automotive industry.

# 5. Result and Analysis

## 1. Hist Plot

```
if p_value < 0.05:  
    print(f'The distribution of {column_name} is not normal.')  
else:  
    print(f'The distribution of {column_name} appears to be normal.')
```



Shapiro-Wilk Test - Statistic: 0.82834792137146, P-value: 2.7323571871870275e-14  
The distribution of price is not normal.

# Result and Analysis

```
In [31]: corr = df_automobile.corr()
df_automobile['price'] = pd.to_numeric(df_automobile['price'], errors='coerce')
df_automobile['engine-size'] = pd.to_numeric(df_automobile['engine-size'], errors='coerce')
df_automobile['horsepower'] = pd.to_numeric(df_automobile['horsepower'], errors='coerce')
df_automobile['normalized-losses'] = pd.to_numeric(df_automobile['normalized-losses'], errors='coerce')
plt.figure(figsize=(20,9))
a = sns.heatmap(corr, annot=True, fmt='.2f')
```



Figure - 2



# Result and Analysis

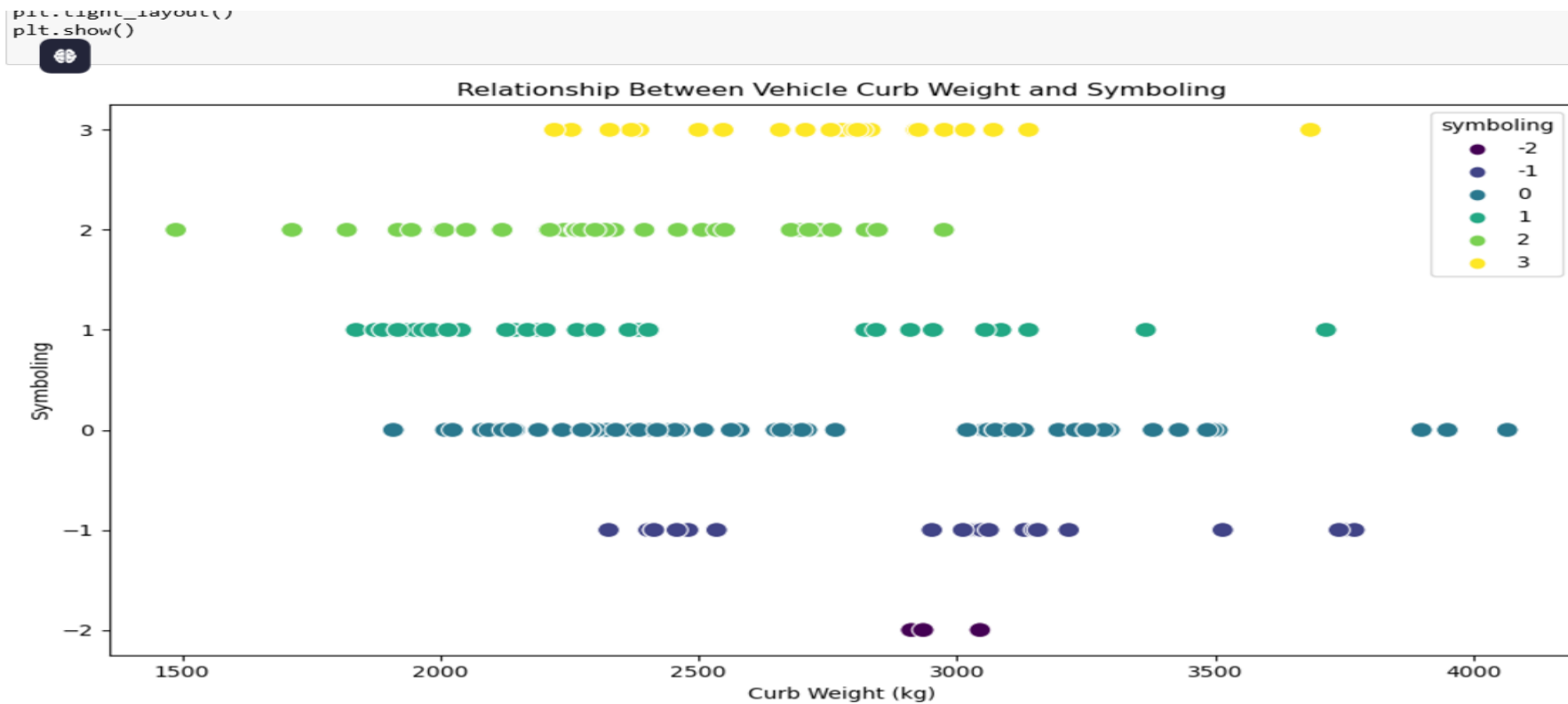
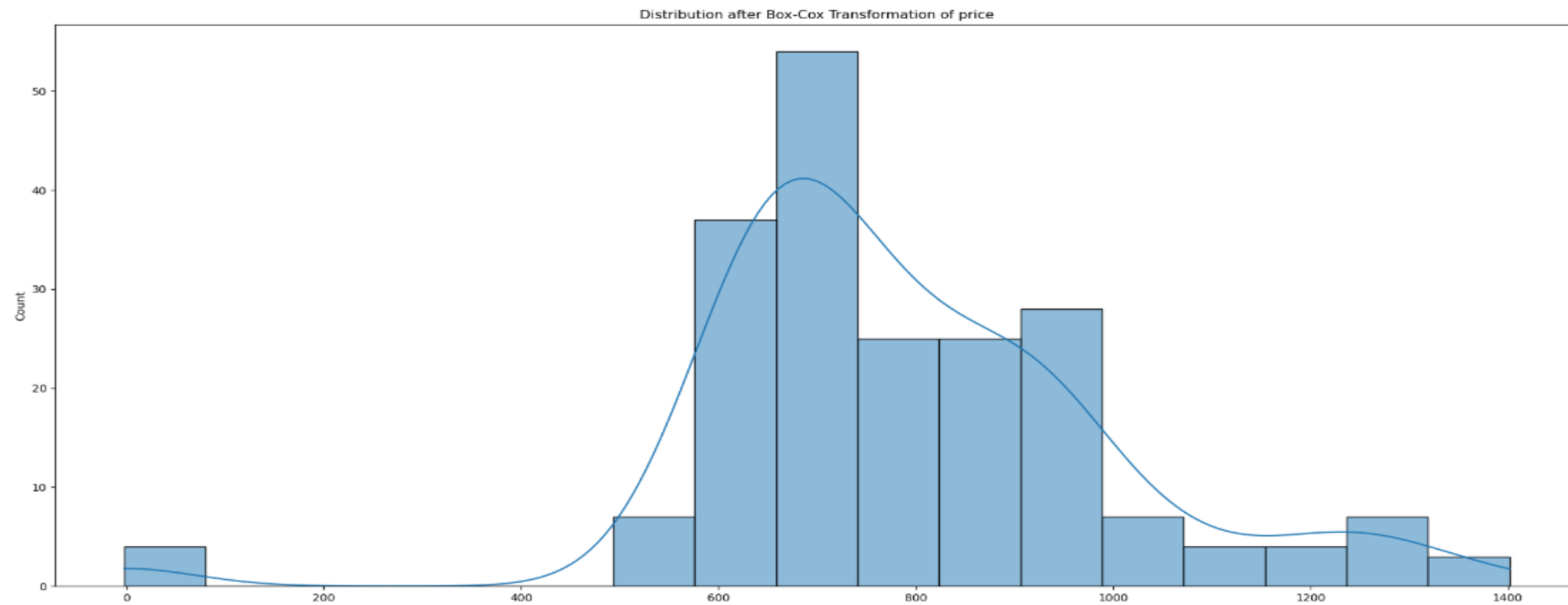


Figure - 3

# Result and Analysis

```
print(f'Optimal lambda value for Box-Cox transformation: {lambda_value}')
```



Optimal lambda value for Box-Cox transformation: 0.4215576851126361

Figure 4

# Result and Analysis

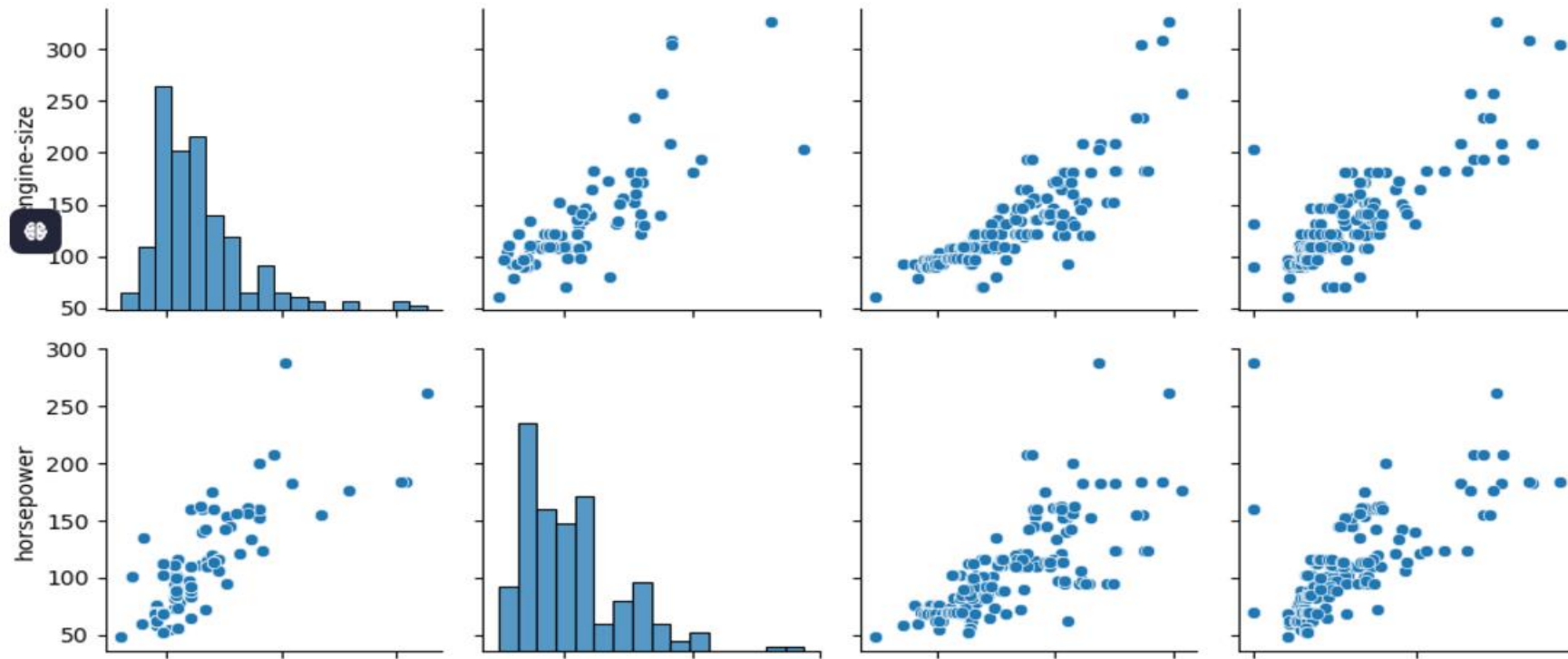


Figure - 5

## 6.Conculsion

---

In conclusion, the Exploratory Data Analysis (EDA) on automobile specification data provides valuable insights into the relationships between various vehicle attributes and their impact on pricing, fuel efficiency, and performance. Through the application of statistical methods and visualization techniques, the analysis uncovers patterns, identifies key factors influencing market trends, and highlights areas for further investigation. The findings from this EDA can assist stakeholders such as automobile manufacturers, consumers, and insurers in making data-driven decisions, ultimately optimizing pricing strategies, improving vehicle design, and enhancing customer satisfaction in the automotive industry.

# 7.Reference

---

To Discover the insights and experiences of the consumers of the automobile industry. This EDA project was made with lot of effort. Accessed a lot of resources for bringing useful insights.

- Kaggle.com – For the Dataset

Kaggle is an online community of data scientists and machine learning engineering that allows users to find datasets.

- Wikipedia- open source

For information of the terms and few examples to bring out the insights.

A wooden desk with a pen, paper clips, and a ruler. A white card with the text "THANK YOU" is placed on the desk.

**THANK  
YOU**