# Car Prediction Model

Mini-Project

# Introduction

Our project involves implementing a Neural Network to suggest consumers a car based on their preferences such as price, mileage, transmission type and fuel type. We have developed a model using a rich dataset.

# Scope

01 Dataset

02 Data Cleaning and Visualisation

03 A.I. Model

04 Testing

05 Conclusion

# Dataset

## Dataset Overview

A comprehensive dataset of car listings which includes features like make, model, year, mileage, engine size, and price.

## Data Source

The data was sourced from kaggle.

## Objective

To develop a model capable of suggesting cars based on a buyers needs.

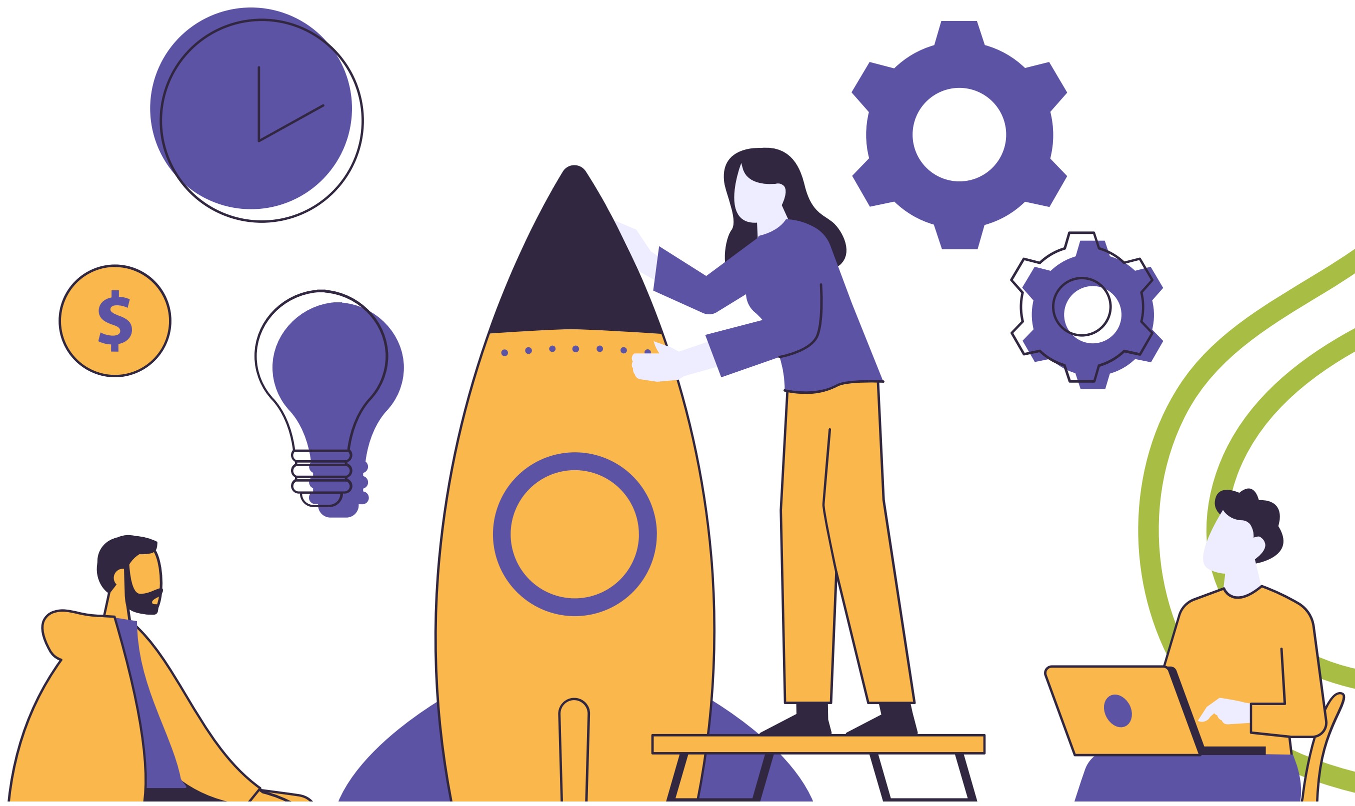
# Data Cleaning

### Common Issues

Initial data inspection showed missing values, duplicate entries, and inconsistent formats, especially in the 'year' and 'price' fields.

### Data Quality

Significant outliers were found in 'price' and 'mileage', which skewed the data distribution and needed addressing.



### Handling Missing Values

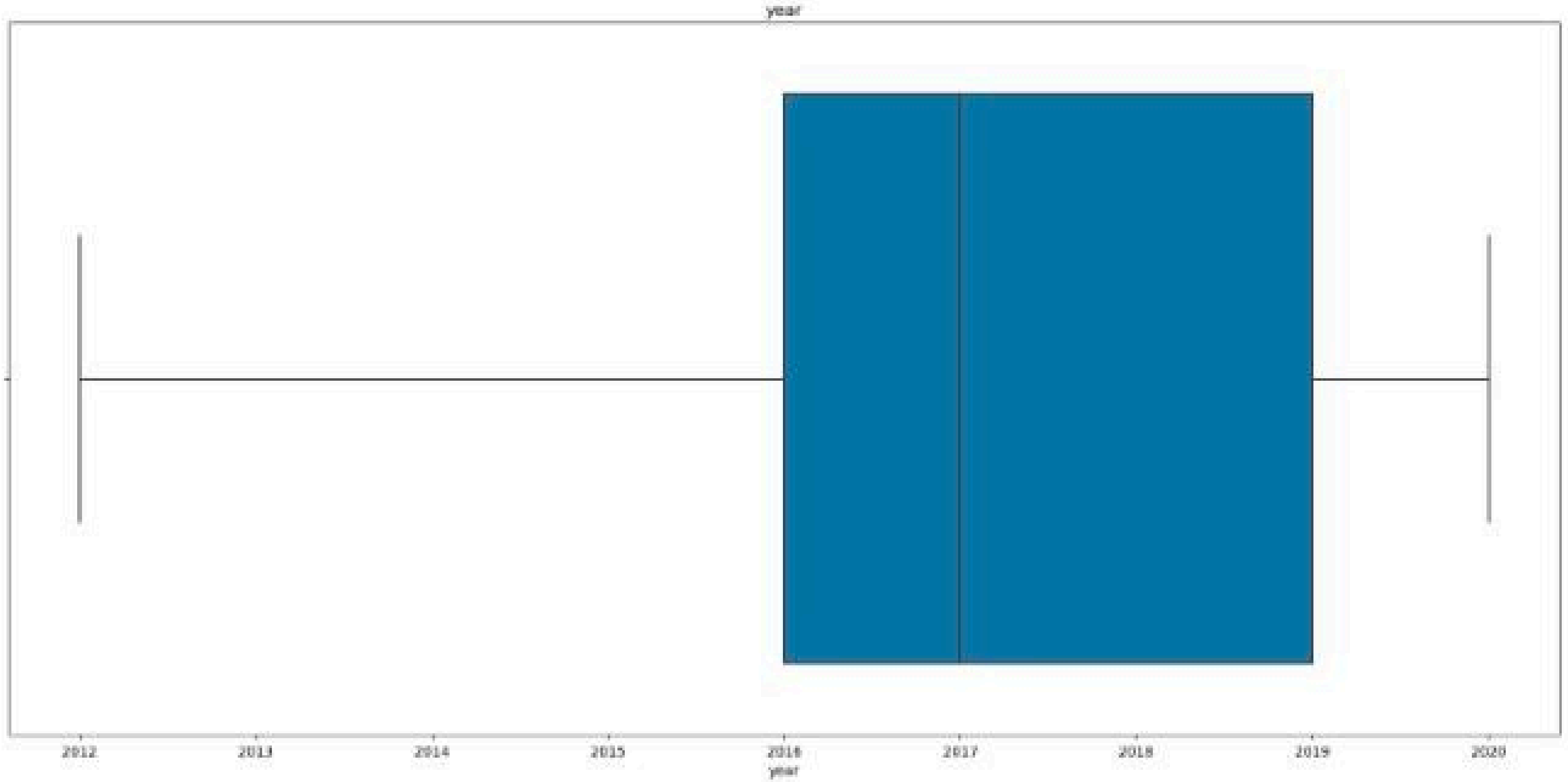
Missing values in critical fields like 'engine size' and 'mileage' were added using median values based on the car's make and model.

### Standardising Formats

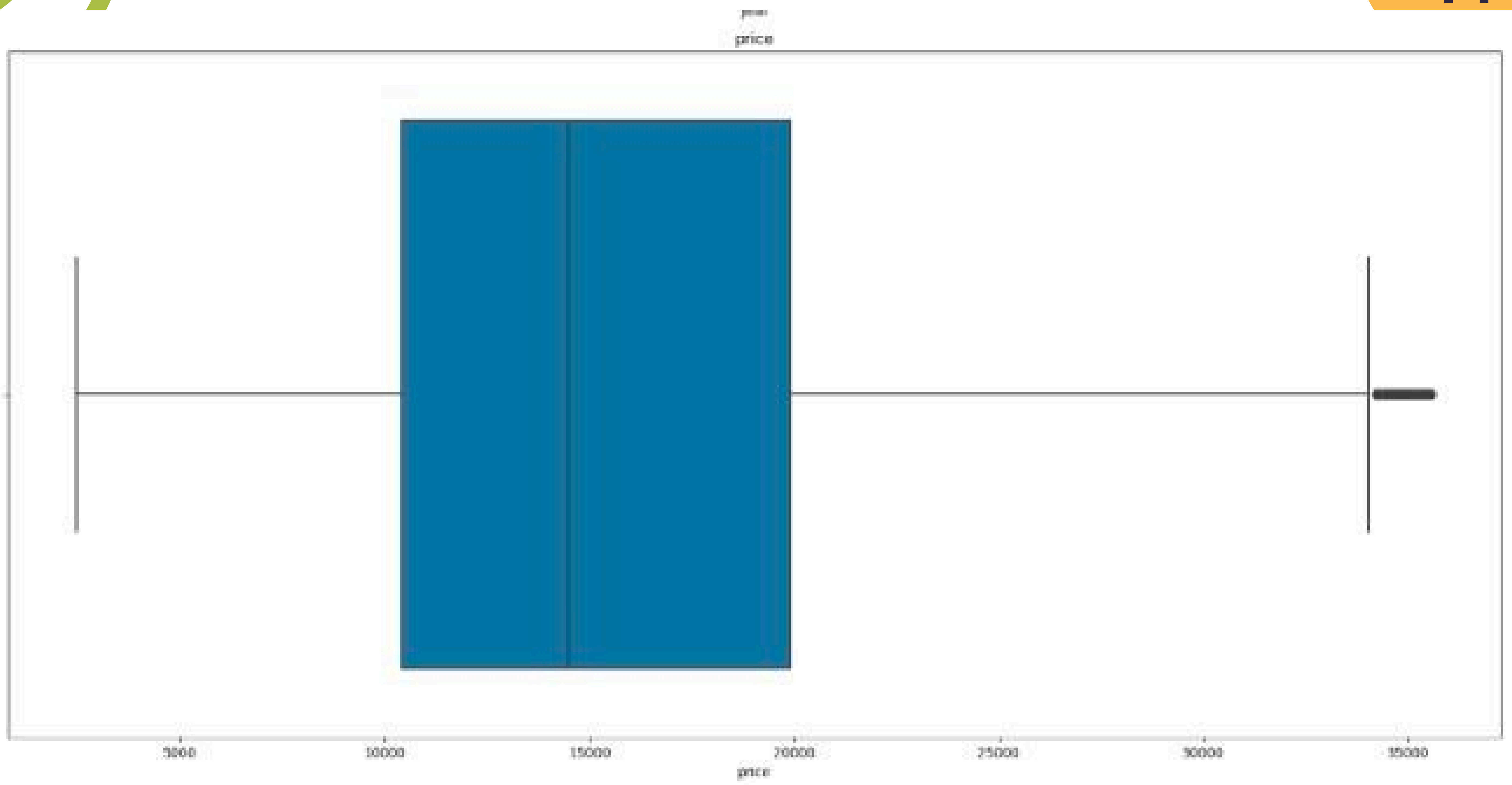
All date fields were standardised to YYYY format, and categorical data was encoded for model compatibility.

### Removing Outliers

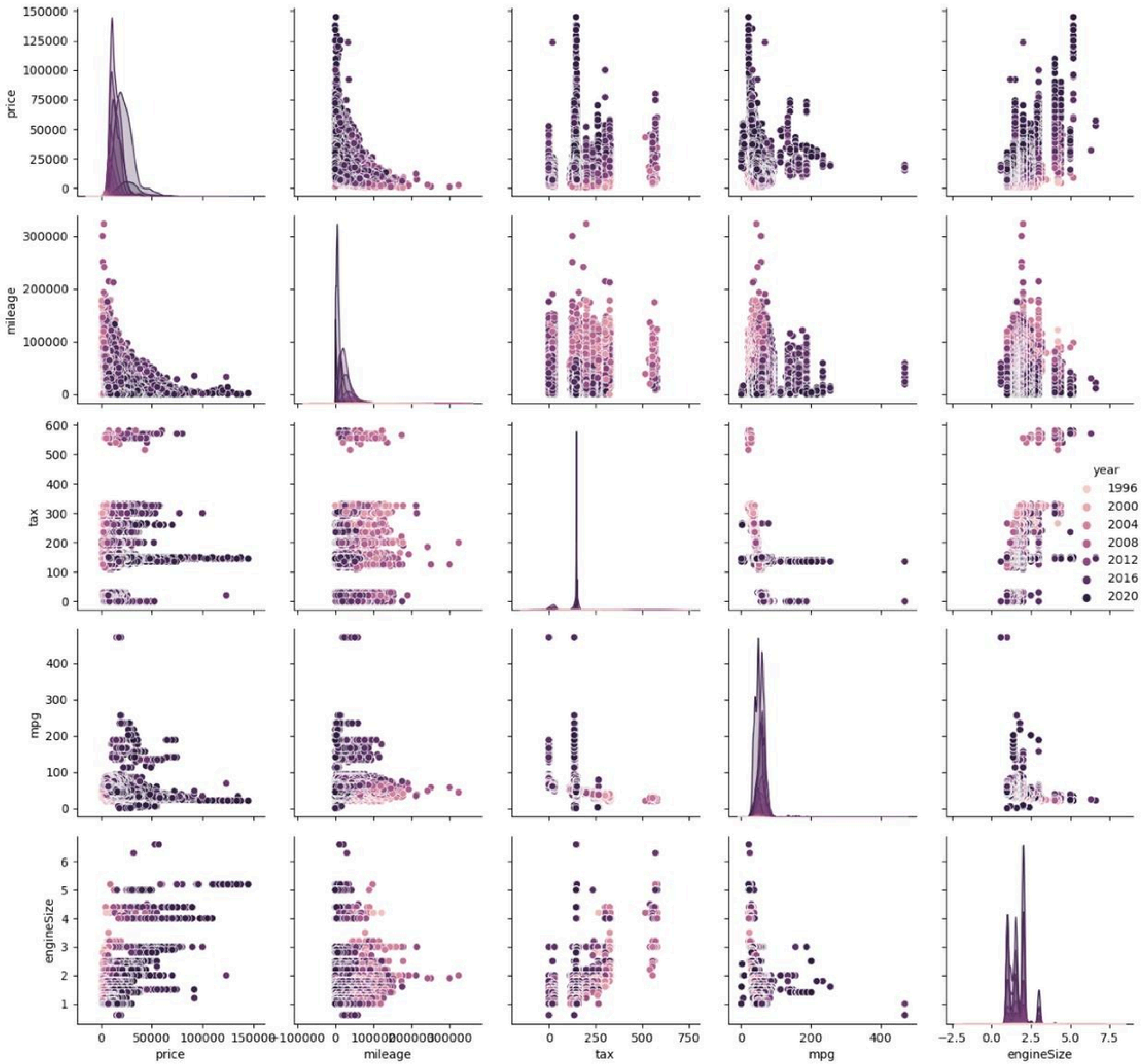
Price and mileage outliers were identified and removed using IQR scores to normalize the data distribution.




# Year Box Plot




# Price Box Plot


price
mileage
tax
mpg
engineSize
year
1996
2000
2004
2008
2012
2016
2020

# Pair Plot

# Visualizing Data Distribution (EDA)

### Box Plots

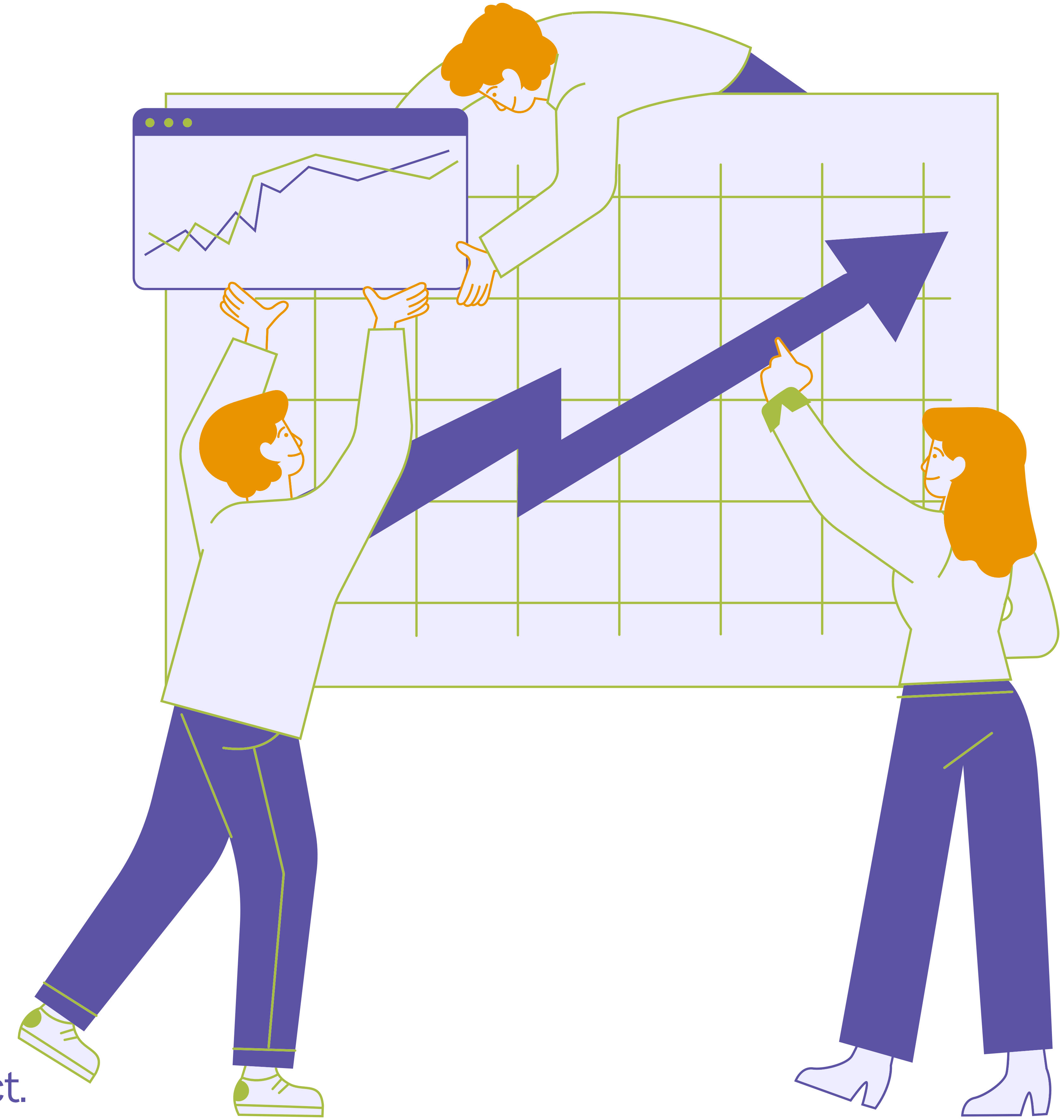
Used to assess the distribution of prices and mileage, identifying skewness and the impact of outlier removal.

### Correlation Heatmap

A heatmap was employed to visualise the correlation between numerical features, highlighting the strong predictors for car pricing.

### Pair Plot

Plotting engine size against price showed patterns and trends, aiding in hypothesis generation about feature impact.



# A.I. Model

## Autoencoder

Autoencoders are a type of neural network used to learn efficient codings of unlabeled data. They work by compressing the input into a latent-space representation and then reconstructing the output from this representation.

Autoencoder is used to reduce the dimensionality of the data, enhancing the neural network's ability to capture essential features without overfitting.

Autoencoder is used to reduce the dimensionality of the data, enhancing the neural network's ability to capture essential features without overfitting.

# Testing & Optimisation

## Methods

01

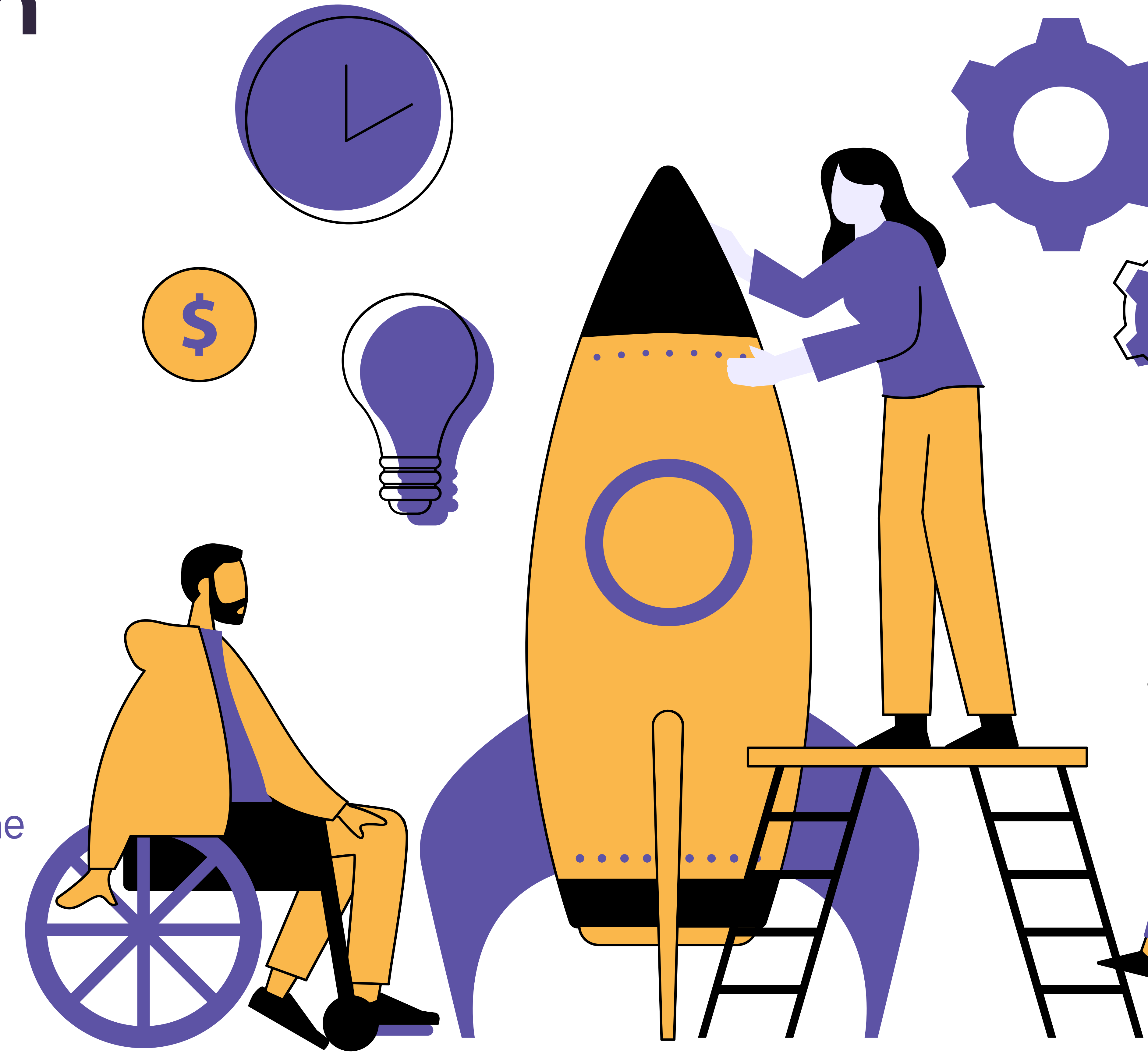
### Loss Function

Mean Squared Error (MSE) was used as the loss function to measure how well the autoencoder reconstructed the data, with the aim being to minimize the difference between the original input and the reconstructed output

02

### Reconstruction Accuracy

Additionally, reconstruction accuracy helps in identifying potential shortcomings or limitations of the model, guiding further refinement or adaptation strategies.



# Testing & Optimisation

## Methods

### Optimizer

03 The Adam optimizer was utilised for its efficient computation and adaptive learning rate properties, which help in converging to the minimum loss faster.

### Training Process

04 The autoencoder was trained on the entire dataset without labels, focusing on learning the most salient features. We utilised early stopping to prevent overtraining and ensure the model generalized well.

In [24]:

```python
# Define the function that will be called when the button is clicked
def on_button_clicked(b):
    with output:
        output.clear_output()
        recommend_similar_cars()
button = widgets.Button(description="Recommend Cars")
button.on_click(on_button_clicked)

# Display widgets for user input
output = widgets.Output()
display(price_input, mileage_input, transmission_input, fuel_type_input, button, output)
```

Price: 25000

Mileage: 30

Transmission: Automatic

Fuel Type: Petrol

Recommend Cars

```
Based on your input, here are suggested cars:
          model  year  price transmission  mileage fuelType    tax   mpg  \
45550        A5  2019  25000    Automatic      100   Petrol  145.0  53.3
15271  2 Series  2020  25000    Automatic      101   Petrol  145.0  47.9
15267  2 Series  2019  25000    Automatic     4000   Petrol  145.0  47.9
37631        X1  2019  25000    Automatic     4534   Petrol  145.0  46.3
10340        A5  2017  25000    Automatic    19000   Petrol  145.0  45.6

       engineSize  Make
45550         1.4  audi
15271         2.0   BMW
15267         2.0   BMW
37631         1.5   BMW
10340         2.0  audi
```

# Pair Plot

# Conclusion

Our project successfully uses a neural network to address a real-world need in the automotive industry. By integrating data cleaning, visualisations, and innovative AI modelling, we've developed a tool that not only predicts but also understands consumer preferences.

# Thanks