PROJECT REPORT

Regularisation Regression -Car Price Prediction Assignment

IIIT Bangalore & upGrad — Data Science Program
Course 4: Machine Learning
Assignment: Regularisation in Regression
Submission Date: 28 October 2025

Dataset Source: AutoScout (Germany)

Submitted by: Akash Singh

This report presents a step-by-step analysis and predictive modelling of used car prices using **Linear Regression** and **Regularisation techniques** (Ridge and Lasso).
The primary objective is to **build an accurate price prediction model**, minimise overfitting, and identify the **most influential factors** affecting car prices.
The dataset comprises **15,915 car listings** sourced from **AutoScout (Germany)**, providing a comprehensive basis for model development and evaluation.

1.2.1. **1.1 Data Loading**

Importing Necessary Libraries

For this project, we imported essential Python libraries for data analysis and modelling:

- **pandas** and **numpy** for data handling and numerical operations.
- **matplotlib** and **seaborn** for data visualisation.
  These libraries help in efficient data exploration, cleaning, and building regression models.

1.1.1 Load the Data
The dataset Car_Price_data.csv was loaded using **pandas**.
It contains **15,915 rows** and **23 columns** with various car attributes.
No missing values were found. The target variable is **price**.
The dataset has both **numerical** and **categorical** features, suitable for regression modelling.

1.3. **2 Analysis and Feature Engineering**

1.3.1. **2.1 Preliminary Analysis and Frequency Distributions**

1.3.1.1. **2.1.1**

Check and fix missing values.

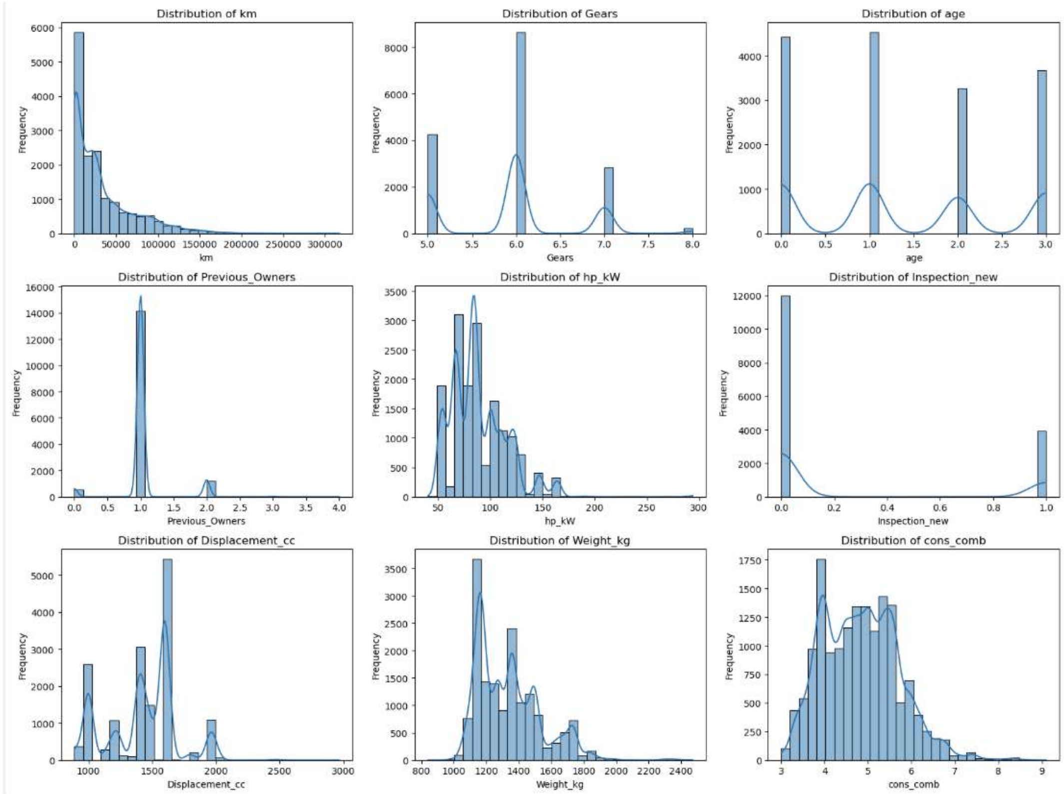
- All columns were checked for missing values.
- **No missing values** were found in the dataset.
- Target variable identified: **price**.
- **Numerical features (9)** include variables like km, Gears, age, hp_kW, Weight_kg, etc.
- **Categorical features (13)** include variables like make_model, Fuel, Gearing_Type, Drive_chain, etc.
- Since no missing data was found, no imputation or removal was required.

1.3.1.2. **2.1.2**

Identify numerical predictors and plot their frequency distributions.

Below is the histogram plot of all numerical features.

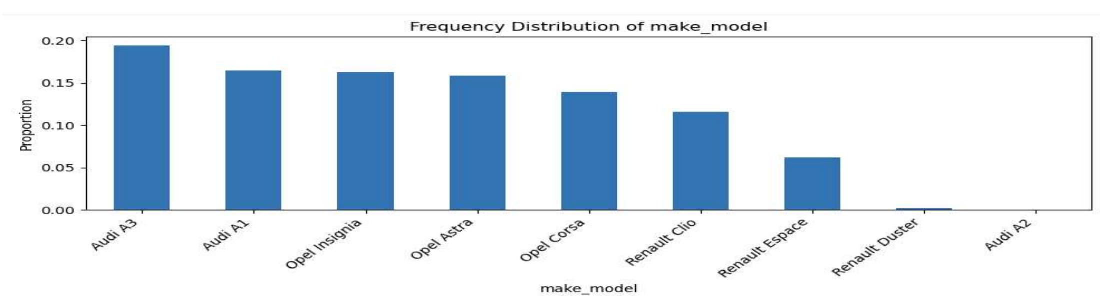
- Most numerical features such as km, Previous_Owners, and Inspection_new are **right-skewed** with values concentrated at lower ranges.
- hp_kW, Displacement_cc, and Weight_kg show **multiple peaks**, indicating different car categories or engine power groups.
- cons_comb (fuel consumption) is **close to normal** with slight skewness.
- These distributions help in identifying **outliers** and deciding on **scaling techniques** later.

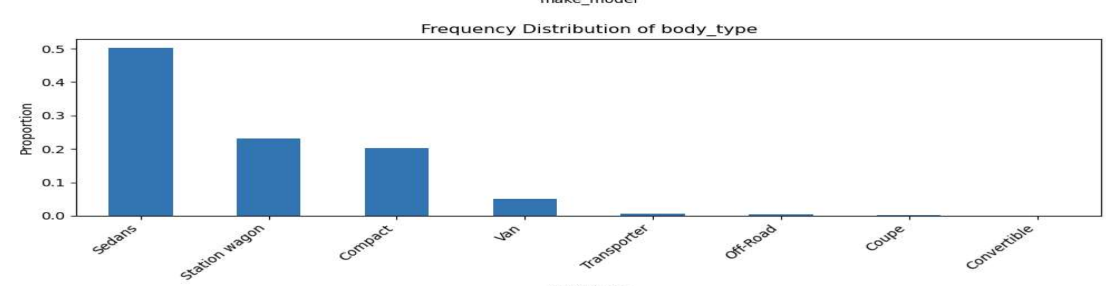


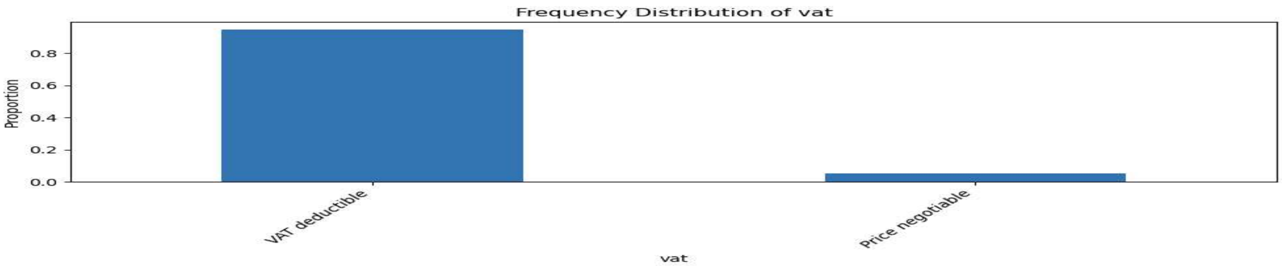
1.3.1.3. **2.1.3**
Identify categorical predictors and plot their frequency distributions.

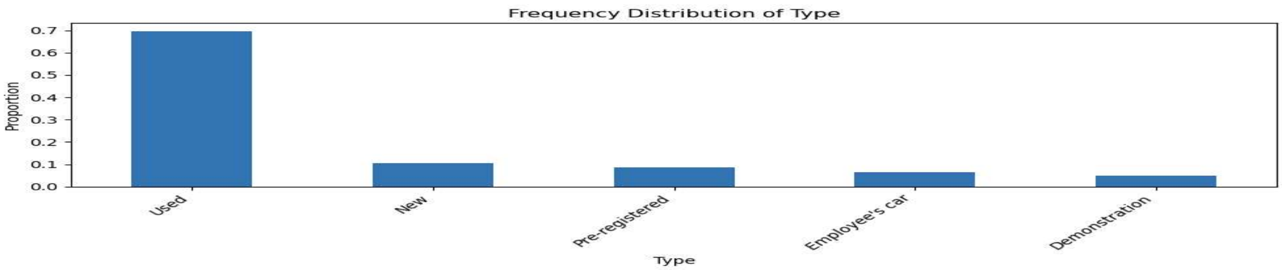
Below is the bar plot of all categorical features.

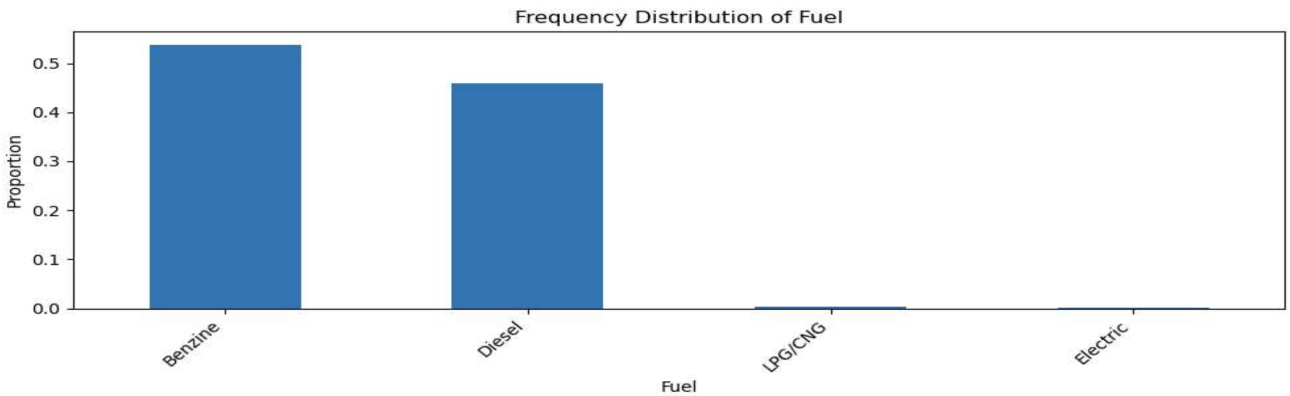
- Most categorical variables show **a few dominant categories** with many less frequent ones.
- Columns like make_model and Fuel have **clear top categories** (e.g., common car models and popular fuel types).
- Some columns, such as Extras and Comfort_Convenience, contain **multiple bundled values**, which may need further feature engineering.
- This step helps identify **low-frequency categories** and **class imbalance**, which can affect model performance.

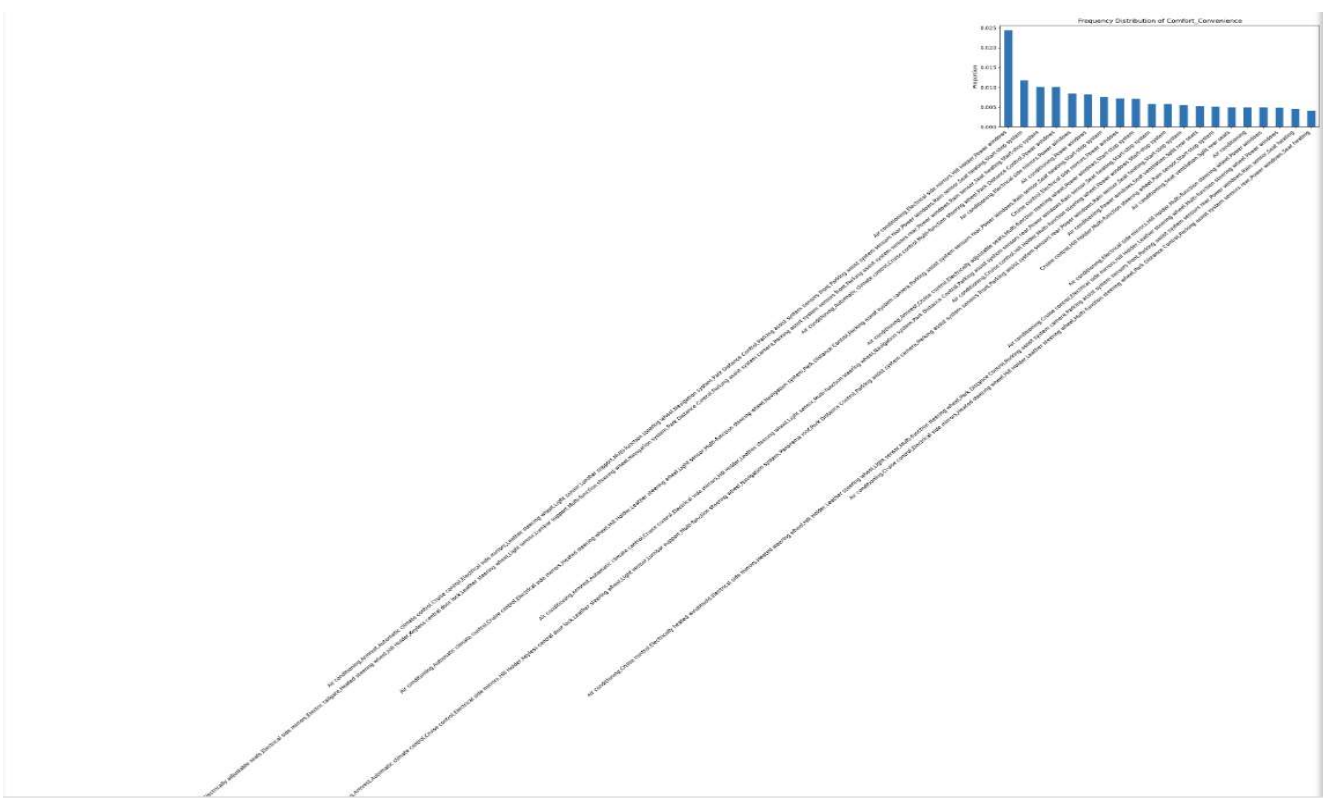


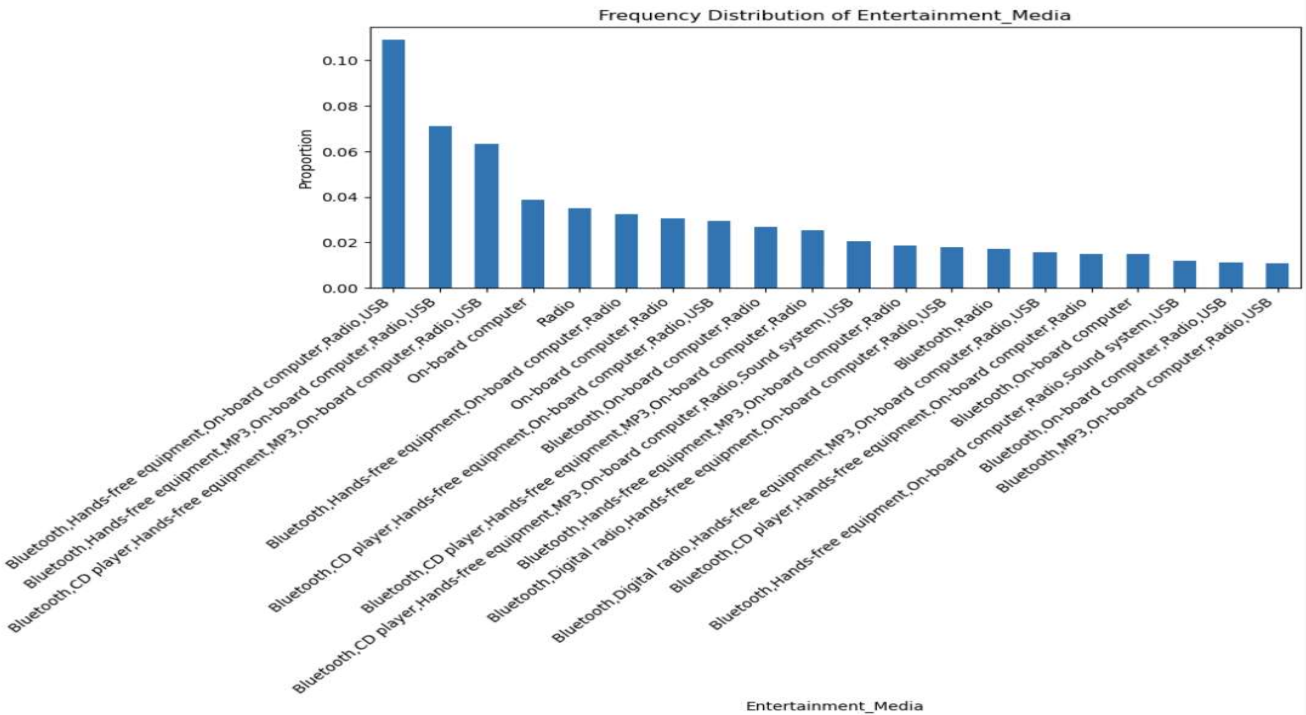



Frequency Distribution of vat
Proportion
0.0
0.2
0.4
0.6
0.8
VAT deductible
Price negotiable
vat

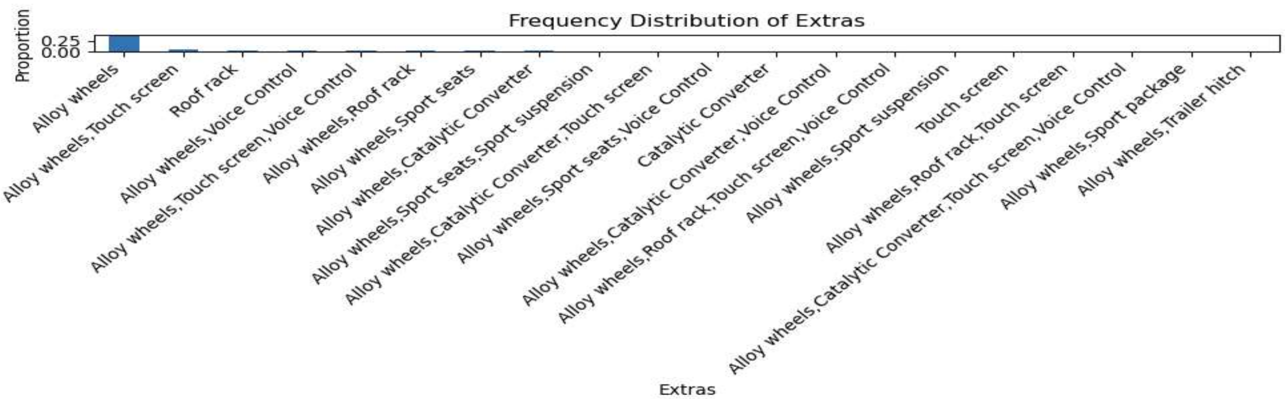

Frequency Distribution of Type
Proportion
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
Used
New
Pre-registered
Employee's car
Demonstration
Type


Frequency Distribution of Fuel
Proportion
0.0
0.1
0.2
0.3
0.4
0.5
Benzine
Diesel
LPG/CNG
Electric
Fuel

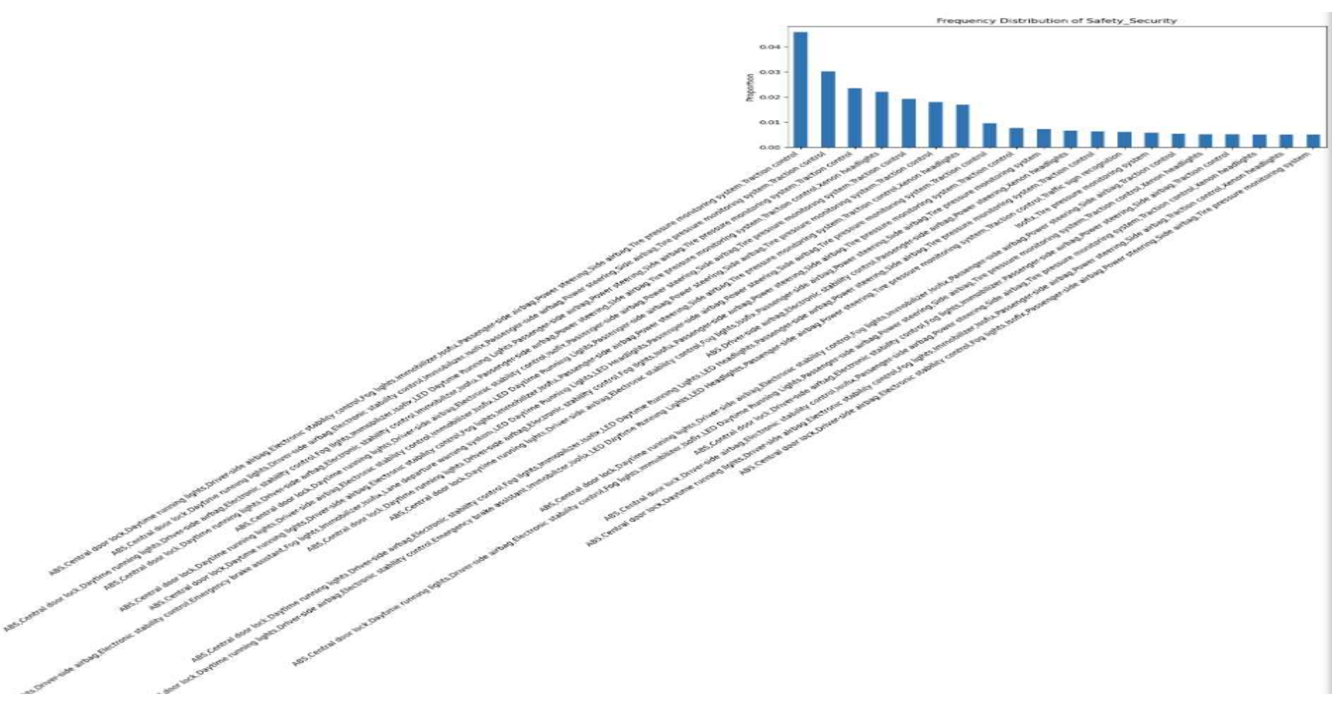

Frequency Distribution of Comfort_Convenience



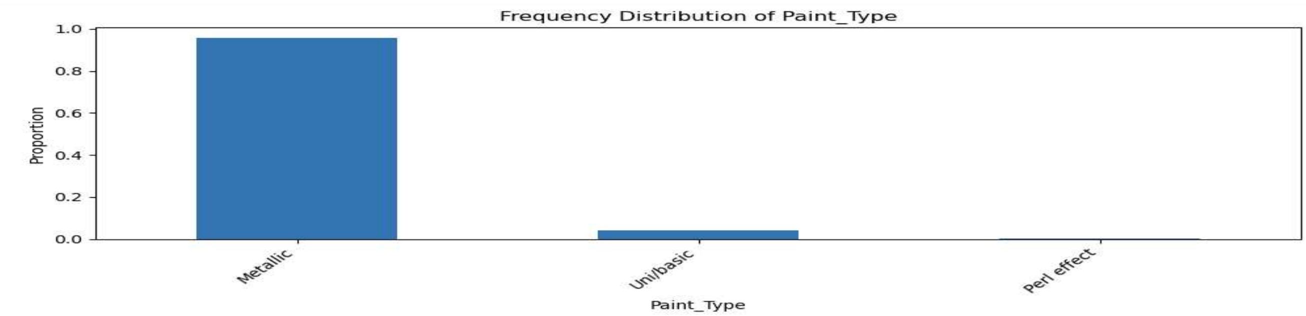
Frequency Distribution of Entertainment_Media
Proportion
0.00
0.02
0.04
0.06
0.08
0.10
Bluetooth,Hands-free equipment,On-board computer,Radio,USB
Bluetooth,Hands-free equipment,MP3,On-board computer,Radio,USB
Bluetooth,CD player,Hands-free equipment,MP3,On-board computer,Radio,USB
On-board computer
Radio
Bluetooth,Hands-free equipment,On-board computer,Radio
On-board computer,Radio
Bluetooth,CD player,Hands-free equipment,On-board computer,Radio,USB
Bluetooth,On-board computer,Radio
Bluetooth,CD player,Hands-free equipment,MP3,On-board computer,Radio
Bluetooth,CD player,Hands-free equipment,MP3,On-board computer,Radio,Sound system,USB
Bluetooth,Hands-free equipment,MP3,On-board computer,Radio
Bluetooth,Digital radio,Hands-free equipment,On-board computer,Radio,USB
Bluetooth,Radio
Bluetooth,Digital radio,Hands-free equipment,MP3,On-board computer,Radio,USB
Bluetooth,CD player,Hands-free equipment,On-board computer,Radio
Bluetooth,On-board computer
Bluetooth,Hands-free equipment,On-board computer,Radio,Sound system,USB
Bluetooth,On-board computer,Radio,USB
Bluetooth,MP3,On-board computer,Radio,USB
Entertainment_Media

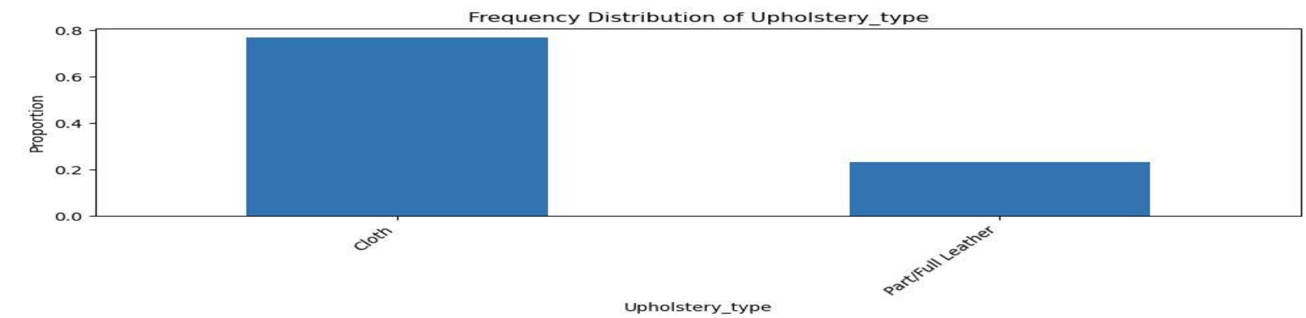


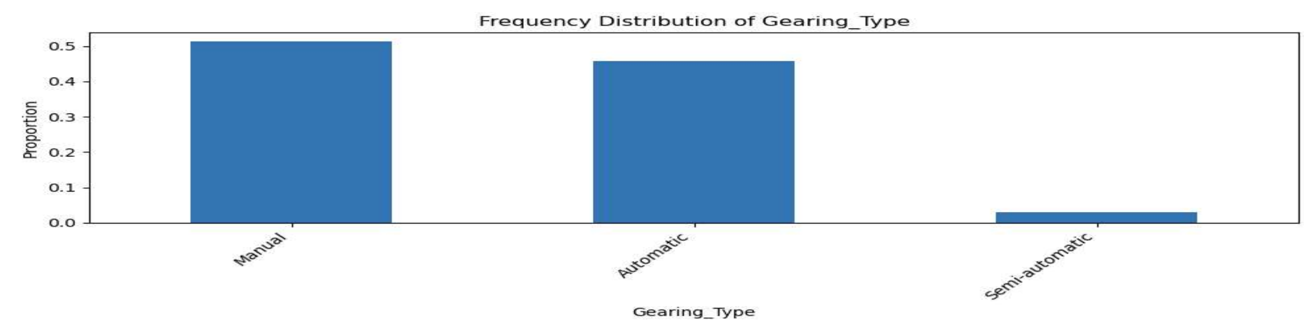
Frequency Distribution of Extras
Proportion
0.25
0.00
Alloy wheels
Alloy wheels,Touch screen
Roof rack
Alloy wheels,Voice Control
Alloy wheels,Touch screen,Voice Control
Alloy wheels,Roof rack
Alloy wheels,Sport seats
Alloy wheels,Catalytic Converter
Alloy wheels,Sport seats,Sport suspension
Alloy wheels,Catalytic Converter,Touch screen
Alloy wheels,Sport seats,Voice Control
Catalytic Converter
Alloy wheels,Catalytic Converter,Voice Control
Alloy wheels,Roof rack,Touch screen,Voice Control
Alloy wheels,Sport suspension
Touch screen
Alloy wheels,Roof rack,Touch screen
Alloy wheels,Catalytic Converter,Touch screen,Voice Control
Alloy wheels,Sport package
Alloy wheels,Trailer hitch
Extras

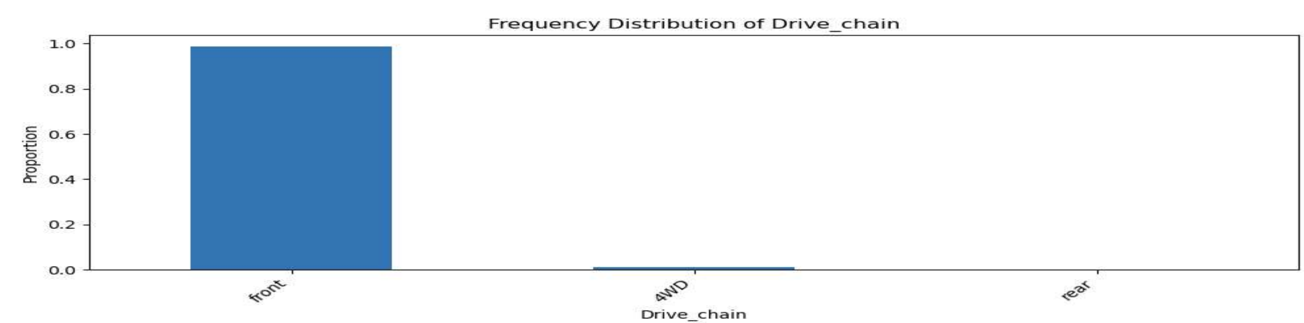


Frequency Distribution of Safety_Security









1.3.1.4. **2.1.4**
Fix columns with low frequency values and class imbalances.

- The **Type** column was simplified using business rules: all categories were grouped under 'Used' since they represent similar conditions.
- In other categorical columns, categories with less than **1% frequency** were grouped into **'Other'**.
- This step reduces class imbalance and helps the model generalise better by avoiding overfitting on rare categories.

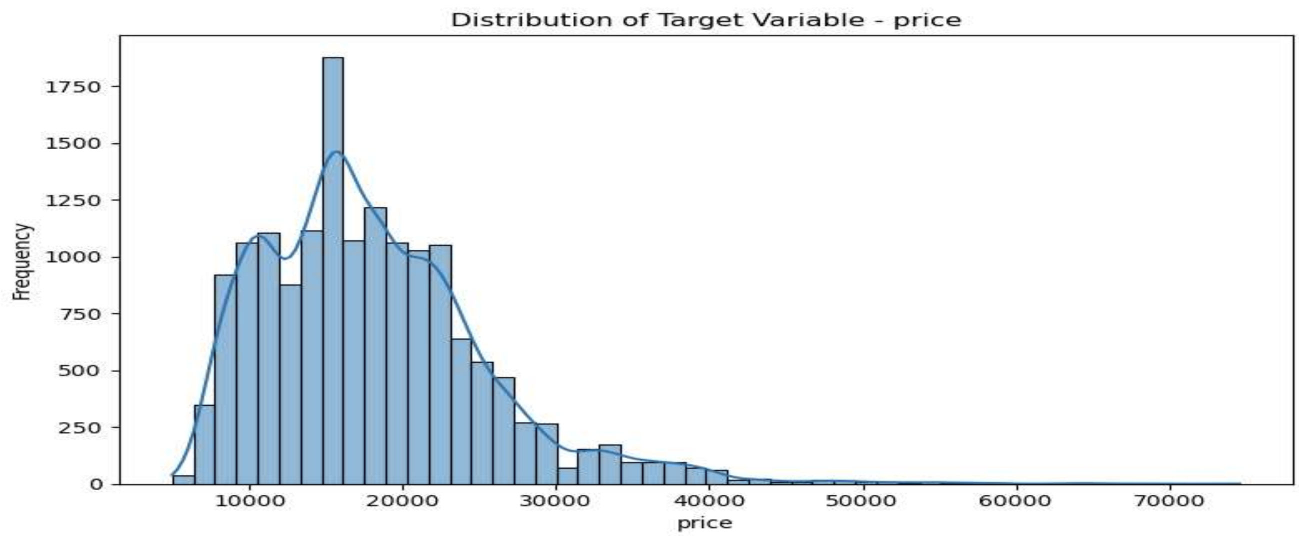
1.3.1.5. **2.1.5**

Identify target variable and plot the frequency distributions. Apply necessary transformations.

(a) Target Variable Distribution (Before Transformation)

Below is the histogram of the target variable price before transformation.

- The distribution is **right-skewed** (skewness ≈ 1.24).
- Most of the prices are concentrated between 10,000 and 25,000, with a long tail towards higher prices.
- A transformation is required to normalise the distribution and reduce the impact of extreme values.

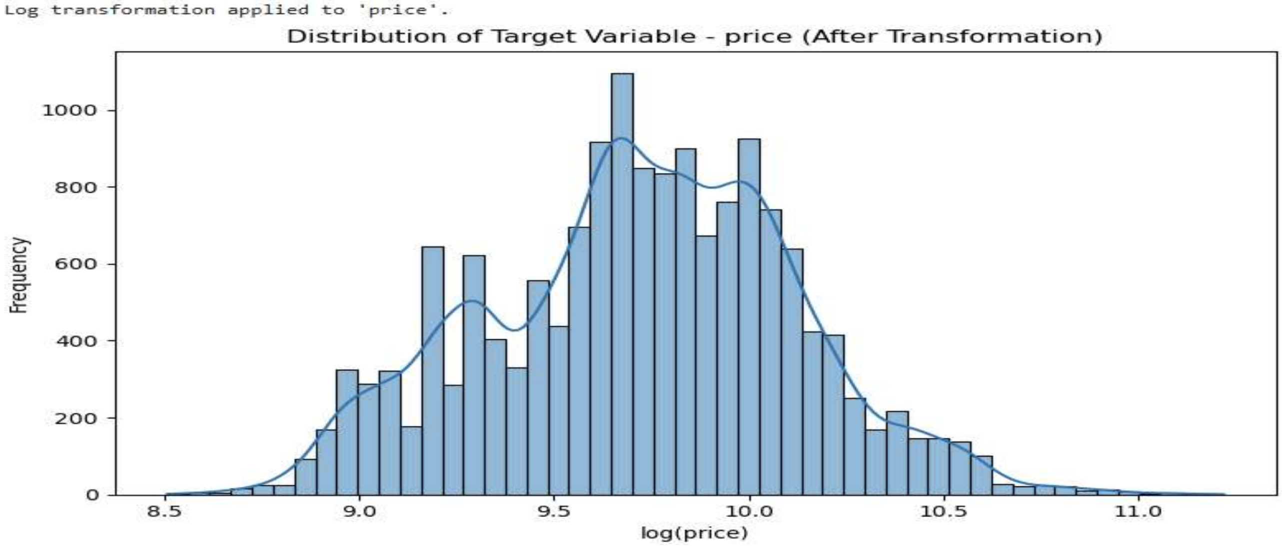



```
Skewness of price: 1.24
```

(b) Target Variable Transformation

- A **log transformation** (log1p) was applied to the price column.
- After transformation, the distribution became **more symmetric** and closer to a normal shape.
- This will help improve **model stability** and **reduce bias** from extreme high-price values**.**

```
Log transformation applied to 'price'.
```



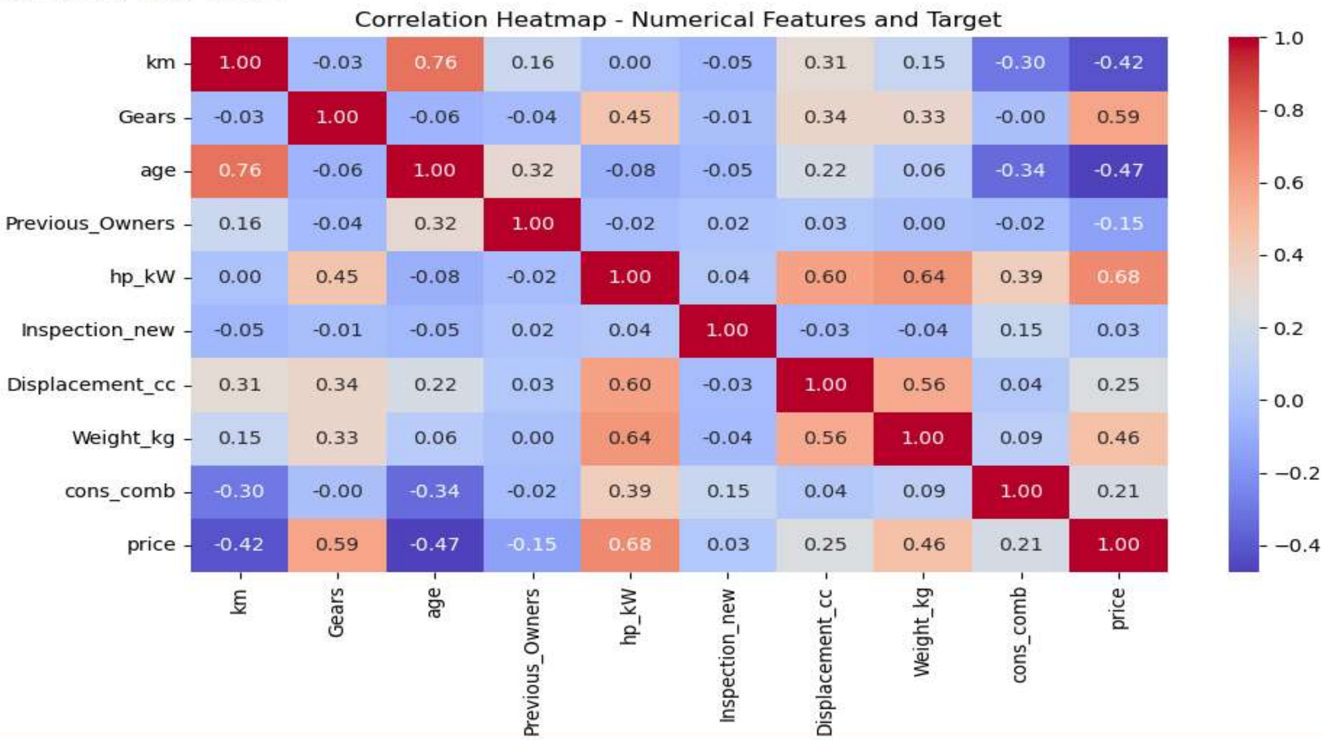
1.3.2. 2.2 Correlation analysis

1.3.2.1. 2.2.1
Correlation map between features and target variable.

Below is the heatmap showing the correlation between numerical features and the target variable price.

- **hp_kW (0.678)** and **Gears (0.590)** have the **strongest positive correlation** with price.
- Weight_kg and Displacement_cc show moderate positive correlations.
- **km (-0.419)** and **age (-0.475)** have **negative correlations**, indicating older and more driven cars tend to be cheaper.
- Other variables like Previous_Owners and Inspection_new have weak correlations.

```
Correlation of numerical features with target variable (price):
price              1.000
hp_kW              0.678
Gears              0.588
Weight_kg          0.465
Displacement_cc    0.255
cons_comb          0.211
Inspection_new     0.031
Previous_Owners   -0.152
km                -0.419
age               -0.475
Name: price, dtype: float64
```

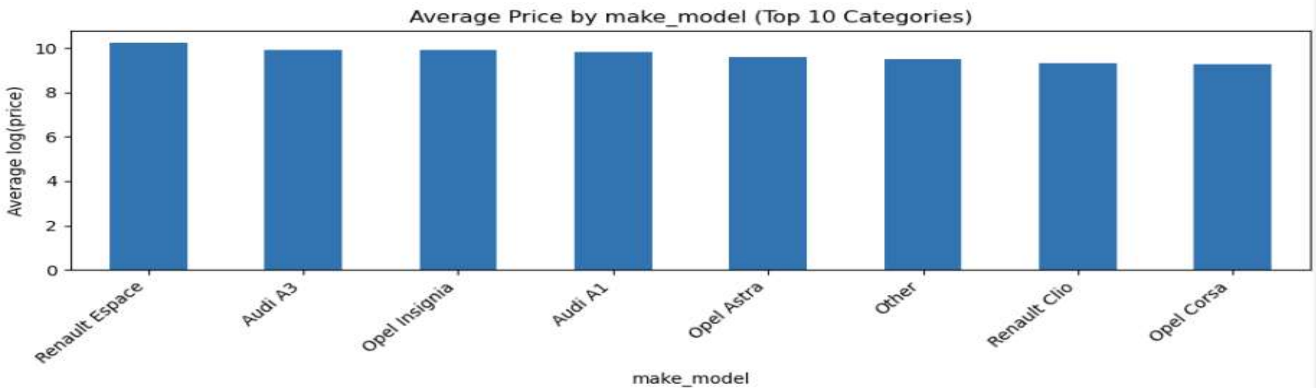


1.3.2.2. 2.2.2
Correlation between Categorical Features and Target Variable

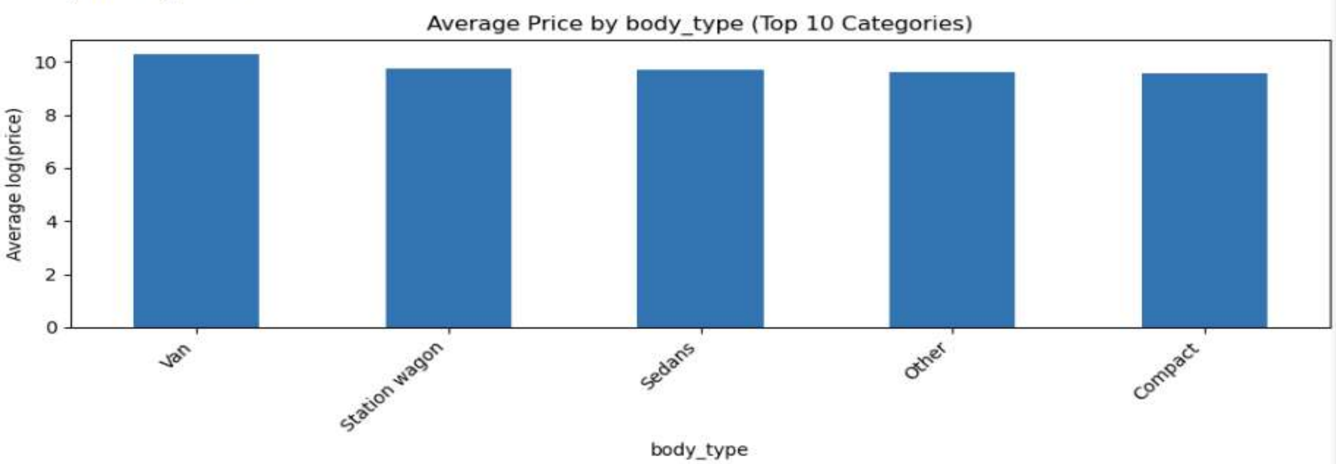
- The bar plots below show the **average target value (log price)** for the top categories of each categorical feature.
- Some categories, such as specific **make_model**, **Fuel**, and **Gearing_Type**, are associated with **higher average prices**, while others show lower values.
- Features like Extras, Comfort_Convenience, and Safety_Security also show variations in price based on additional car features.

- This indicates that **categorical attributes influence pricing significantly** and should be properly encoded during model building.

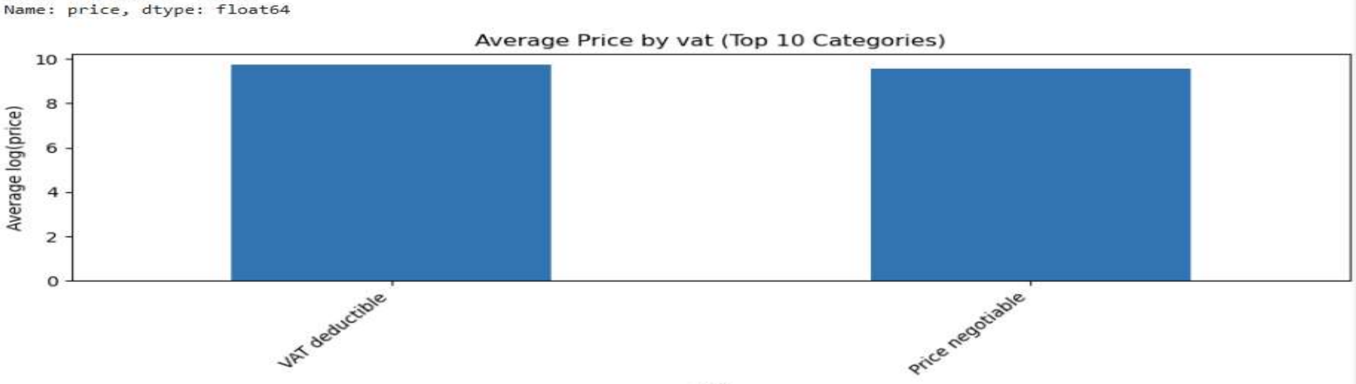
```
Average target (price) for top categories in: make_model
make_model
Renault Espace   10.272
Audi A3           9.928
Opel Insignia     9.913
Audi A1           9.818
Opel Astra        9.626
Other             9.505
Renault Clio      9.336
Opel Corsa        9.276
Name: price, dtype: float64
```



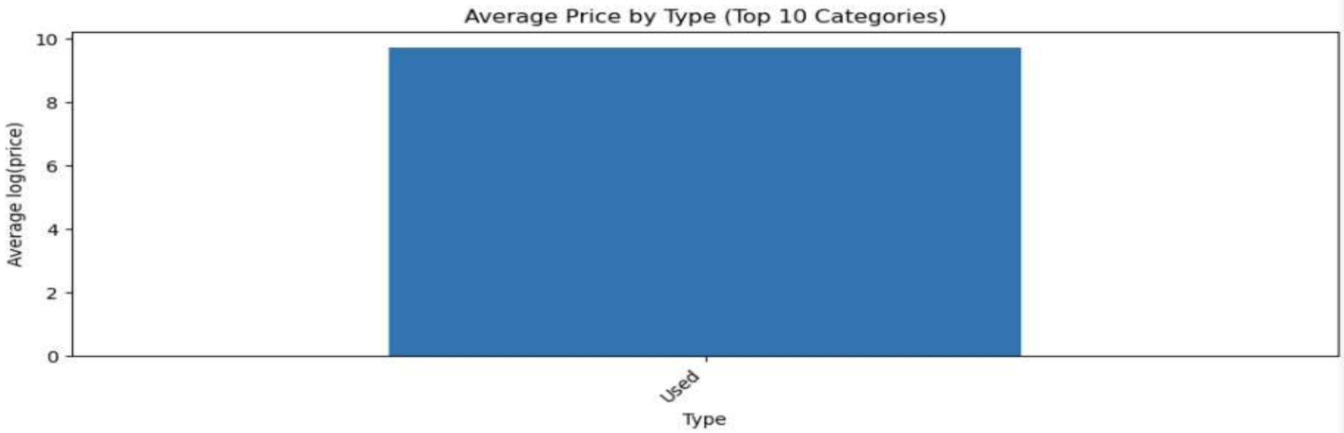

```
Average target (price) for top categories in: body_type
body_type
Van             10.295
Station wagon    9.753
Sedans           9.716
Other            9.623
Compact          9.556
Name: price, dtype: float64
```



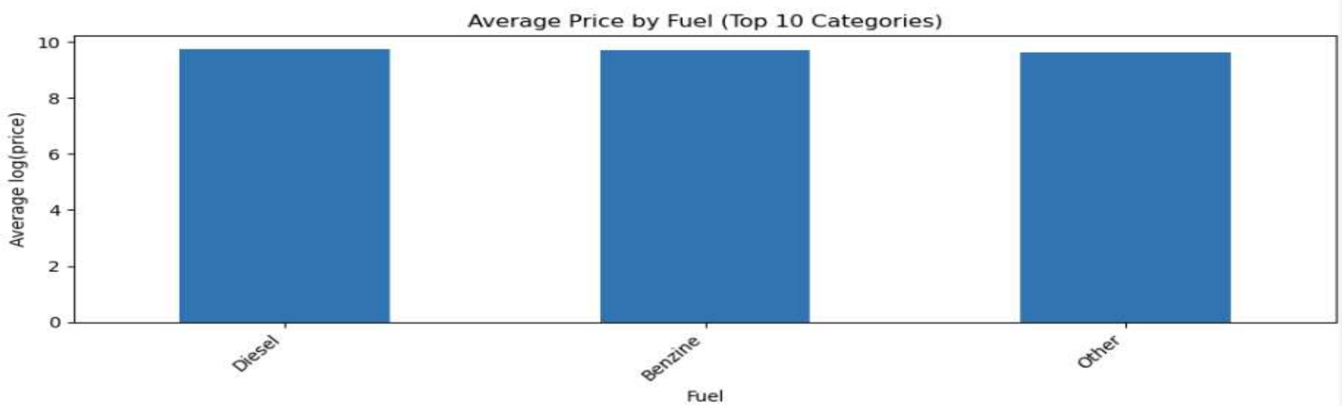

```
Average target (price) for top categories in: vat
vat
VAT deductible     9.730
Price negotiable   9.556
Name: price, dtype: float64
```



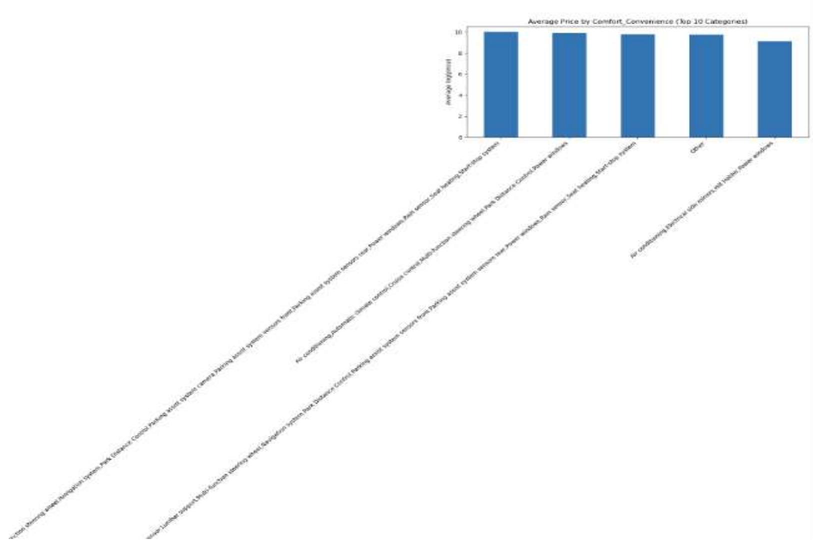
```
Average target (price) for top categories in: Type
Type
Used    9.721
Name: price, dtype: float64
```



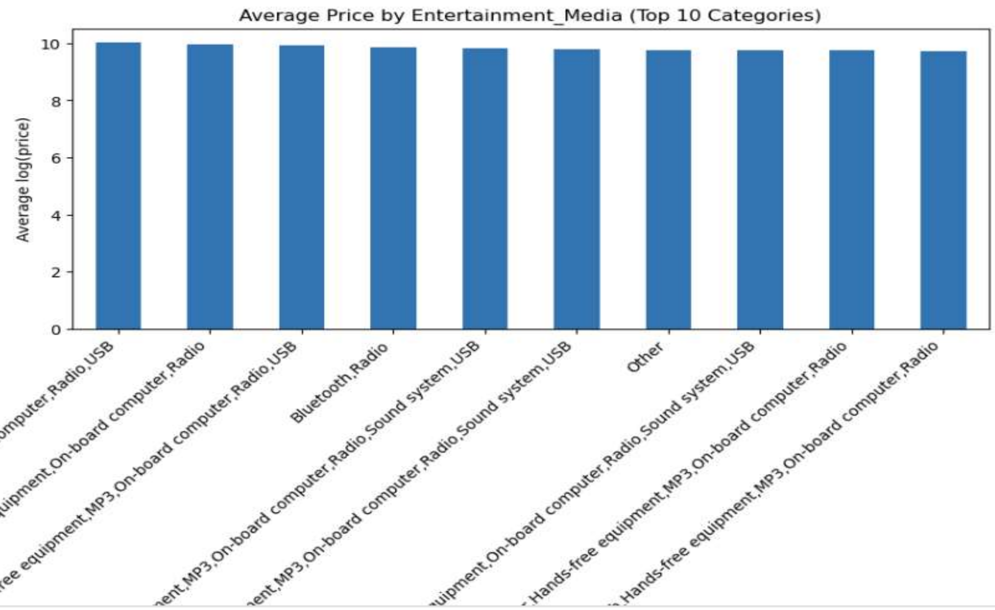

```
Average target (price) for top categories in: Fuel
Fuel
Diesel     9.735
Benzine    9.709
Other      9.639
Name: price, dtype: float64
```



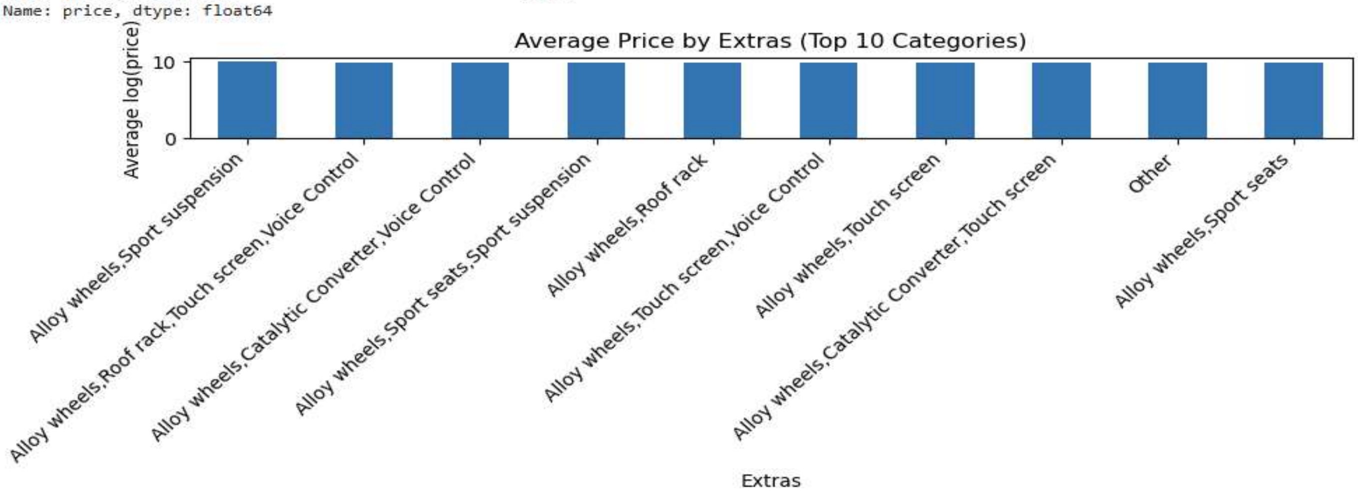

```
Average target (price) for top categories in: Comfort_Convenience
Comfort_Convenience
Air conditioning,Armrest,Automatic climate control,Cruise control,Electrically adjustable seats,Electrical side mirrors,Electric tailgate,Heated steeri
ng wheel,Hill Holder,Keyless central door lock,Leather steering wheel,Light sensor,Lumbar support,Multi-function steering wheel,Navigation system,Park
Distance Control,Parking assist system camera,Parking assist system sensors front,Parking assist system sensors rear,Power windows,Rain sensor,Seat hea
ting,Start-stop system   10.017
Air conditioning,Automatic climate control,Cruise control,Multi-function steering wheel,Park Distance Control,Power windows
9.899
Air conditioning,Armrest,Automatic climate control,Cruise control,Electrical side mirrors,Leather steering wheel,Light sensor,Lumbar support,Multi-func
tion steering wheel,Navigation system,Park Distance Control,Parking assist system sensors front,Parking assist system sensors rear,Power windows,Rain s
ensor,Seat heating,Start-stop system
9.762
Other
9.730
Air conditioning,Electrical side mirrors,Hill Holder,Power windows
9.131
Name: price, dtype: float64
```



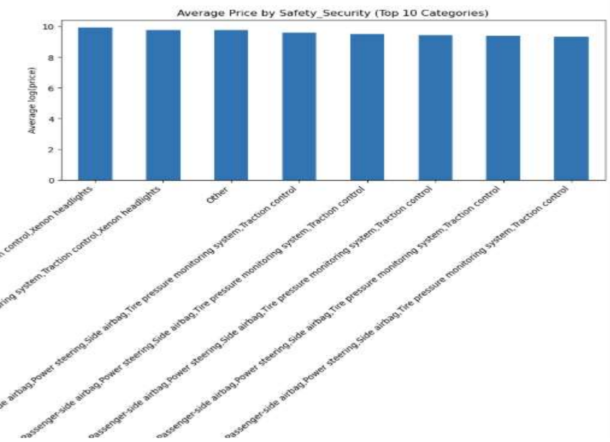
```
Average target (price) for top categories in: Entertainment_Media
Entertainment_Media
Bluetooth,Digital radio,Hands-free equipment,On-board computer,Radio,USB                              10.018
Bluetooth,Hands-free equipment,On-board computer,Radio                                                9.964
Bluetooth,Digital radio,Hands-free equipment,MP3,On-board computer,Radio,USB                          9.911
Bluetooth,Radio                                                                                       9.858
Bluetooth,CD player,Hands-free equipment,MP3,On-board computer,Radio,Sound system,USB                 9.819
Bluetooth,CD player,Digital radio,Hands-free equipment,MP3,On-board computer,Radio,Sound system,USB   9.784
Other                                                                                                 9.771
Bluetooth,Hands-free equipment,On-board computer,Radio,Sound system,USB                               9.744
Bluetooth,CD player,Hands-free equipment,MP3,On-board computer,Radio                                  9.743
Bluetooth,Hands-free equipment,MP3,On-board computer,Radio                                            9.713
Name: price, dtype: float64
```



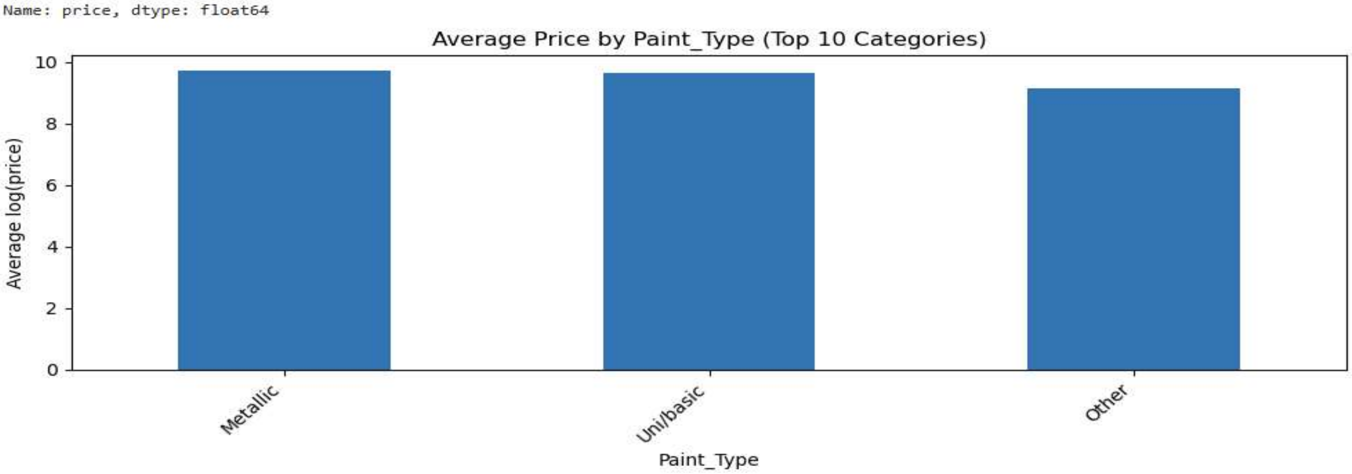

```
Average target (price) for top categories in: Extras
Extras
Alloy wheels,Sport suspension                          9.943
Alloy wheels,Roof rack,Touch screen,Voice Control      9.896
Alloy wheels,Catalytic Converter,Voice Control         9.892
Alloy wheels,Sport seats,Sport suspension              9.871
Alloy wheels,Roof rack                                 9.837
Alloy wheels,Touch screen,Voice Control                9.831
Alloy wheels,Touch screen                              9.826
Alloy wheels,Catalytic Converter,Touch screen          9.813
Other                                                  9.808
Alloy wheels,Sport seats                               9.794
Name: price, dtype: float64
```



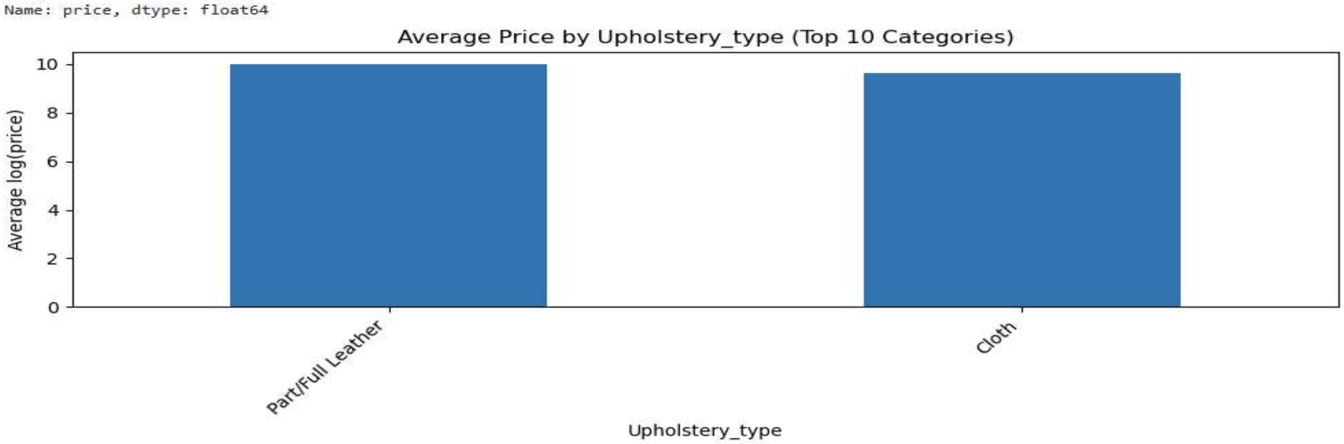
```
Average target (price) for top categories in: Safety_Security
Safety_Security
ABS,Central door lock,Daytime running lights,Driver-side airbag,Electronic stability control,Immobilizer,Isofix,Passenger-side airbag,Power steering,Si
de airbag,Tire pressure monitoring system,Traction control,Xenon headlights                                  9.930
ABS,Central door lock,Daytime running lights,Driver-side airbag,Electronic stability control,Fog lights,Immobilizer,Isofix,Passenger-side airbag,Power
steering,Side airbag,Tire pressure monitoring system,Traction control,Xenon headlights                       9.775
Other
9.756
ABS,Central door lock,Daytime running lights,Driver-side airbag,Electronic stability control,Fog lights,Immobilizer,Isofix,LED Daytime Running Lights,P
assenger-side airbag,Power steering,Side airbag,Tire pressure monitoring system,Traction control   9.597
ABS,Central door lock,Daytime running lights,Driver-side airbag,Electronic stability control,Fog lights,Immobilizer,Isofix,Passenger-side airbag,Power
steering,Side airbag,Tire pressure monitoring system,Traction control                                        9.510
ABS,Central door lock,Daytime running lights,Driver-side airbag,Electronic stability control,Isofix,Passenger-side airbag,Power steering,Side airbag,Ti
re pressure monitoring system,Traction control                                                               9.448
ABS,Central door lock,Daytime running lights,Driver-side airbag,Electronic stability control,Immobilizer,Isofix,Passenger-side airbag,Power steering,Si
de airbag,Tire pressure monitoring system,Traction control                                                   9.394
ABS,Central door lock,Daytime running lights,Driver-side airbag,Electronic stability control,Immobilizer,Isofix,LED Daytime Running Lights,Passenger-si
de airbag,Power steering,Side airbag,Tire pressure monitoring system,Traction control                        9.344
Name: price, dtype: float64
```



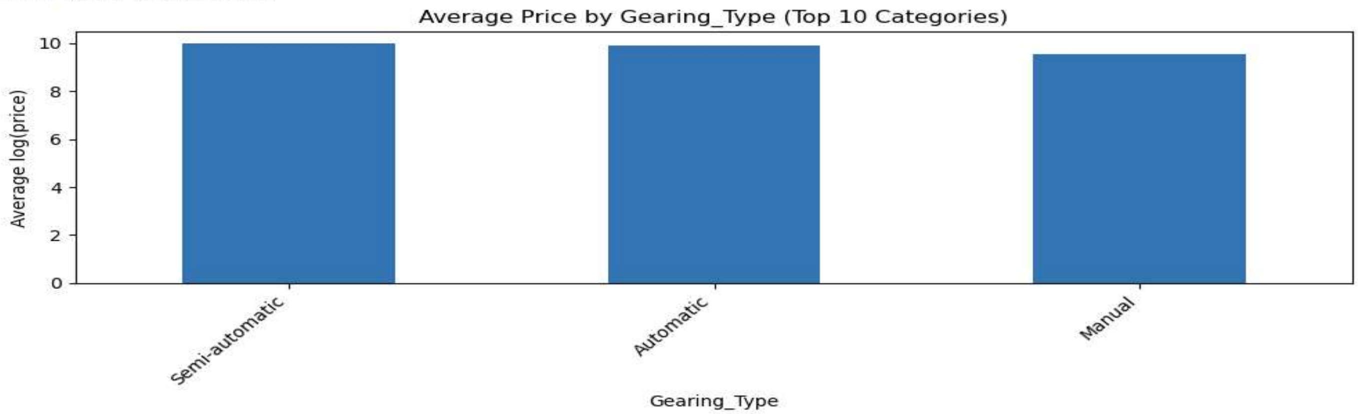

```
Average target (price) for top categories in: Paint_Type
Paint_Type
Metallic    9.725
Uni/basic   9.639
Other       9.145
Name: price, dtype: float64
```



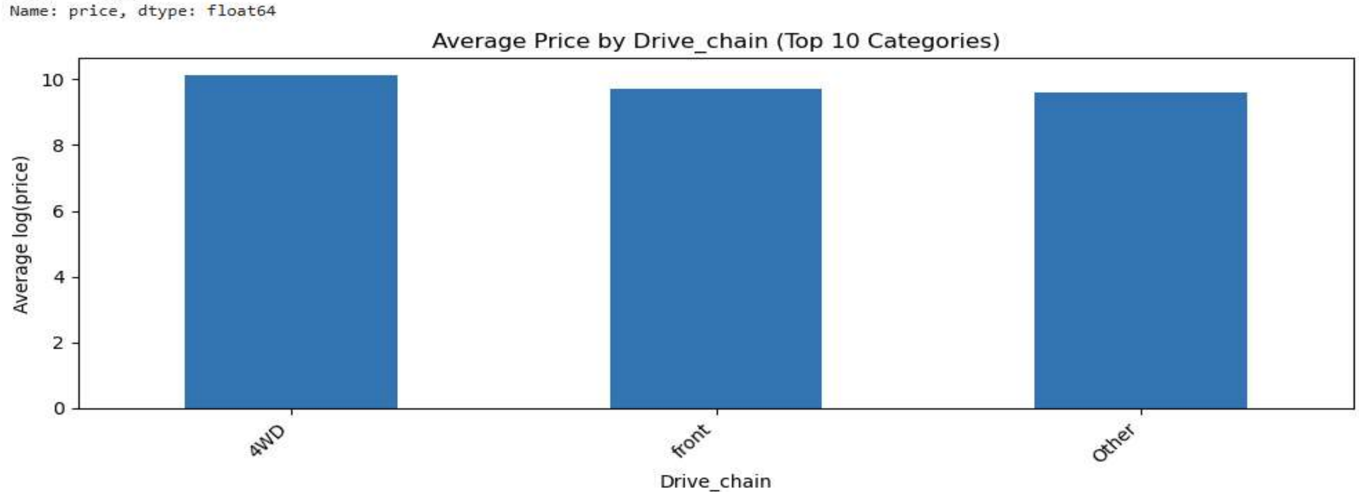

```
Average target (price) for top categories in: Upholstery_type
Upholstery_type
Part/Full Leather    9.981
Cloth                9.642
Name: price, dtype: float64
```



```
Average target (price) for top categories in: Gearing_Type
Gearing_Type
Semi-automatic   9.972
Automatic        9.906
Manual           9.540
Name: price, dtype: float64
```




```
Average target (price) for top categories in: Drive_chain
Drive_chain
4WD      10.133
front     9.715
Other     9.606
Name: price, dtype: float64
```



1.3.3. 2.3 Outlier analysis

1.3.3.1. **2.3.1**

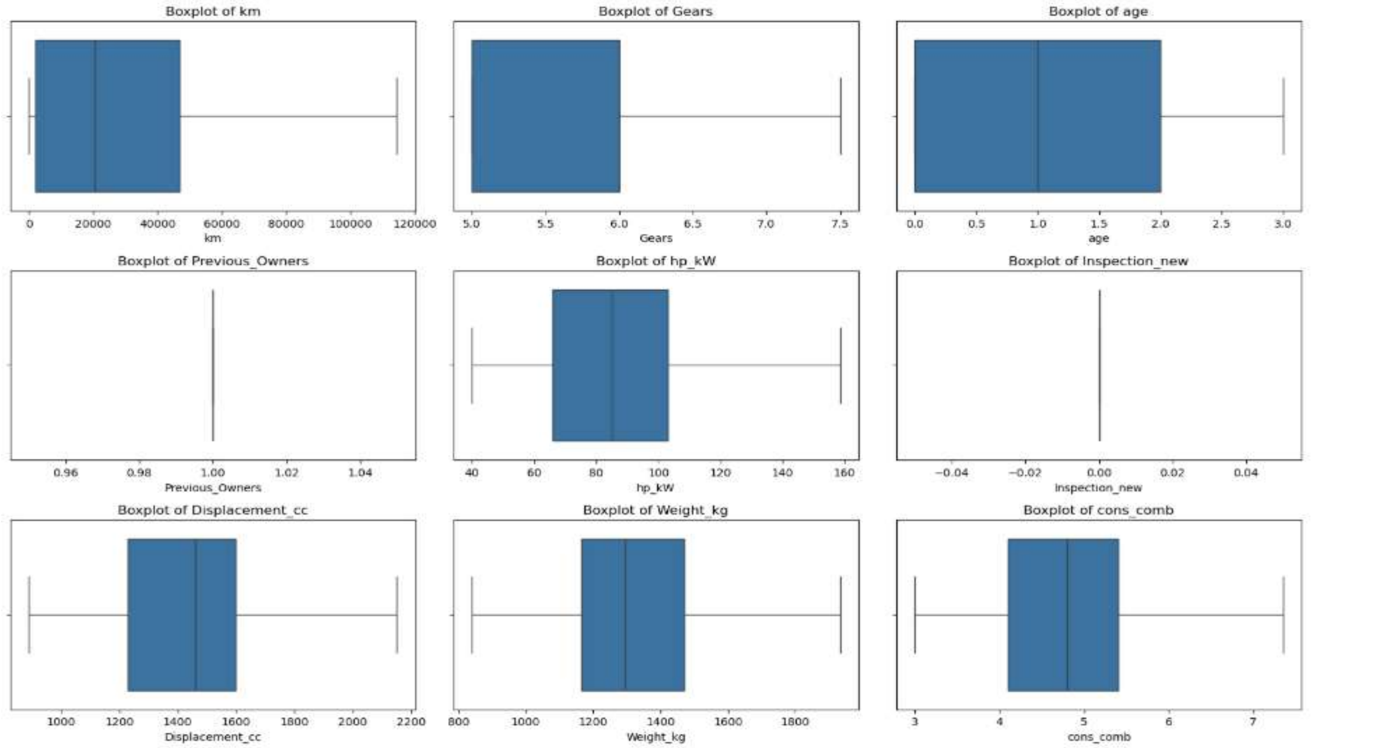
Identify Potential Outliers

- Outlier detection was performed using the **IQR (Interquartile Range)** method for all numerical features.
- Columns such as Inspection_new (3932), Previous_Owners (1757), and km (689) showed **a high number of outliers**.
- Moderate outliers were observed in Gears and hp_kW, while age had no outliers.
- Identifying these outliers is important to prevent them from **negatively affecting the regression model**.

1.3.3.2. **2.3.2**

Handle Outliers

- Outliers in numerical features were treated using the **IQR capping (winsorization)** method.
- After capping, boxplots show a more **balanced distribution** with extreme values capped at boundary points.
- Features like km, hp_kW, and Inspection_new show a **significant reduction in outliers**.
- This helps improve model stability and reduces the risk of overfitting due to extreme values.



1.3.4. 2.4 Feature Engineering

1.3.4.1. **2.4.1**
Fix/Create Columns

- The column **vat** was dropped as it does not contribute directly to predicting car prices.
- A new feature **power_to_weight** was created by dividing hp_kW by Weight_kg, representing engine performance relative to car weight.
- An additional derived feature **age_category** was created to capture non-linear effects of car age.
- These engineered features are expected to improve model performance and interpretability.

1.3.4.2. **2.4.2**
Analysis and Feature Engineering on Specification Columns

- The specification columns (Comfort_Convenience, Entertainment_Media, Extras, Safety_Security) contained multiple comma-separated features.
- Their unique values were checked to understand the type and spread of features.
- Since these columns contain text-based lists and can increase model complexity, they were **dropped** from the dataset after analysis.
- This step helped in **reducing dimensionality** and simplifying the dataset for modelling.

1.3.4.3. **2.4.3**
Feature encoding

- No categorical columns remained after feature engineering.
- No encoding was required.
- The dataset is now fully numerical and ready for modelling.

1.3.4.4. 2.4.4
Train-Test Split

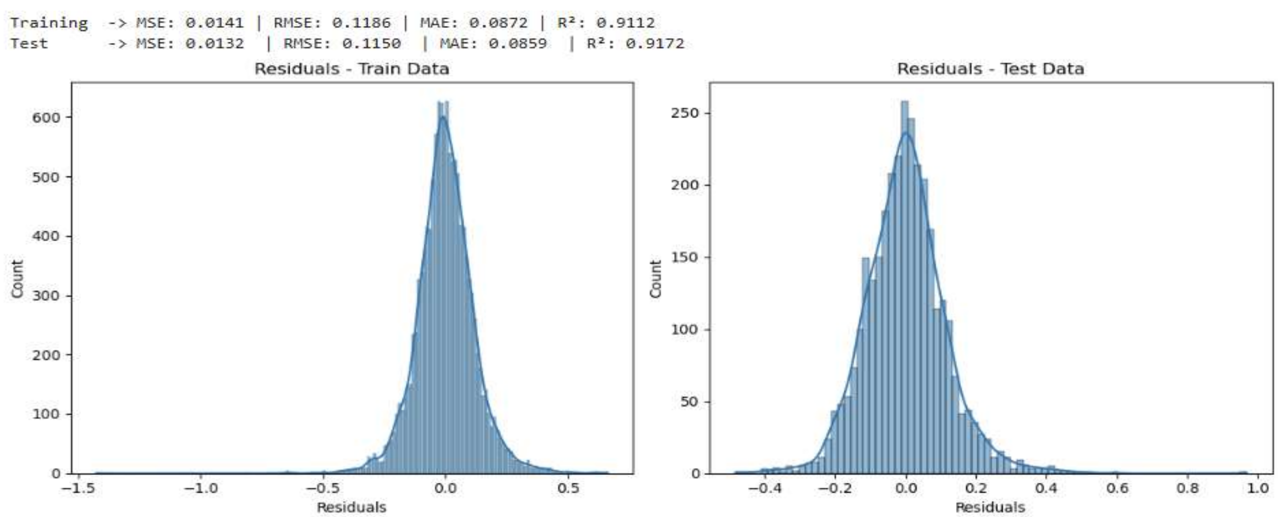
- The dataset was split into training and testing sets using an 80:20 ratio.
- X_train shape: (12732, 31) and X_test shape: (3183, 31).
- This ensures enough data for training while keeping a portion aside for unbiased model evaluation.

1.3.4.5. **2.4.5**
Feature Scaling
- Features were scaled using **StandardScaler** on numeric columns.
- The scaler was fit on the training set and applied to the test set.
- This standardisation improves model stability and performance.

1.4. 3 Linear Regression Models

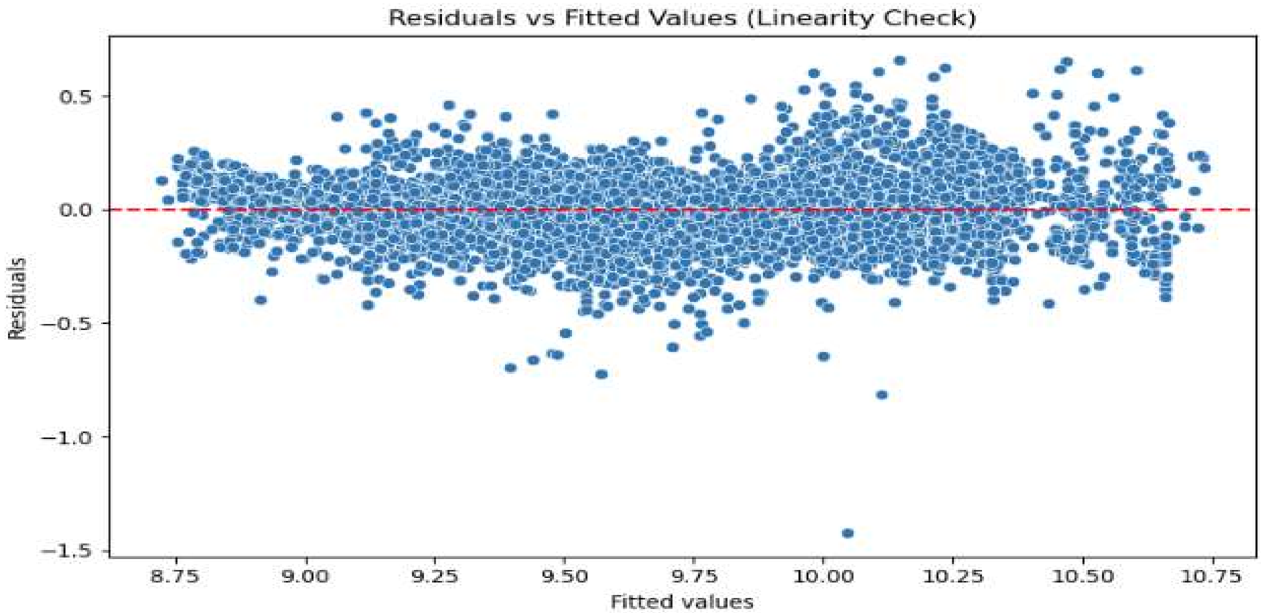
1.4.1. 3.1 Baseline Linear Regression Model

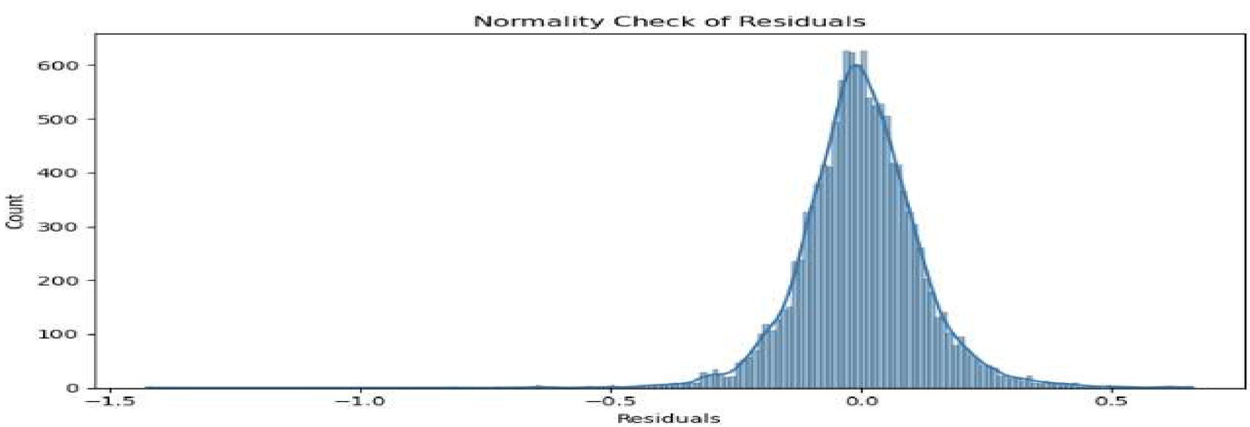
1.4.1.1. **3.1.1**
Build and Evaluate Basic Linear Regression Model
- A basic Linear Regression model was trained on scaled and encoded features.
- Model achieved $\mathbf{R^2 = 0.9112}$ (Train) and $\mathbf{R^2 = 0.9172}$ (Test) with low MSE and RMSE.
- Residuals for both training and test data are approximately normal and centered around 0.
- This indicates a good baseline fit and no major signs of overfitting.

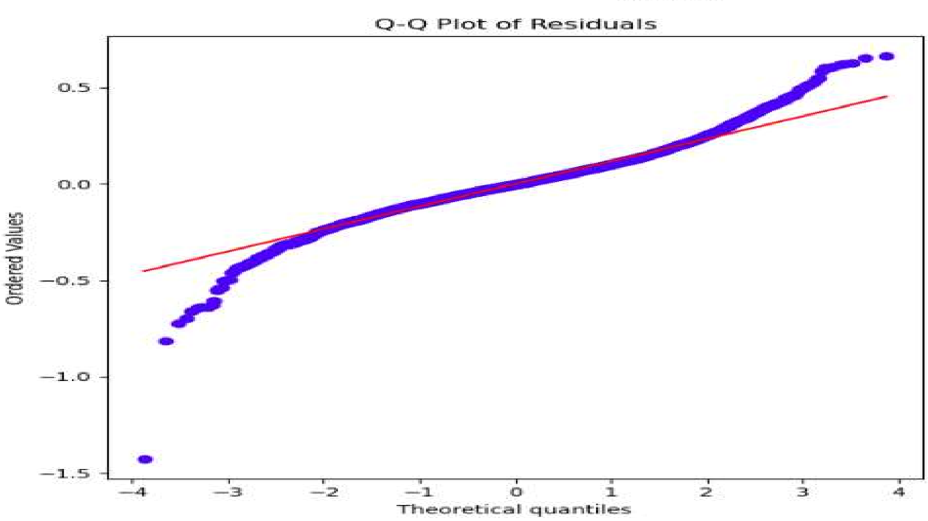


1.4.1.2. **3.1.2**
Residual and Assumption Analysis
- **Linearity Check:**
  Residuals vs fitted values plot showed points randomly scattered around zero, indicating **linearity assumption is satisfied**.
- **Normality Check:**
  Residuals are approximately **normally distributed** with no major skewness. Q-Q plot follows a near-straight line.
- **Multicollinearity Check (VIF):**
  All VIF values are **below 5** (km: 2.91, age: 2.79, others < 2), indicating **no multicollinearity**

problem.
No feature removal was required.







	Feature	VIF
0	km	2.919
2	age	2.788
3	Displacement_cc	1.301
4	cons_comb	1.173
1	Gears	1.156

1.4.2. 3.2 Ridge Regression Implementation

1.4.2.1. **3.2.1**

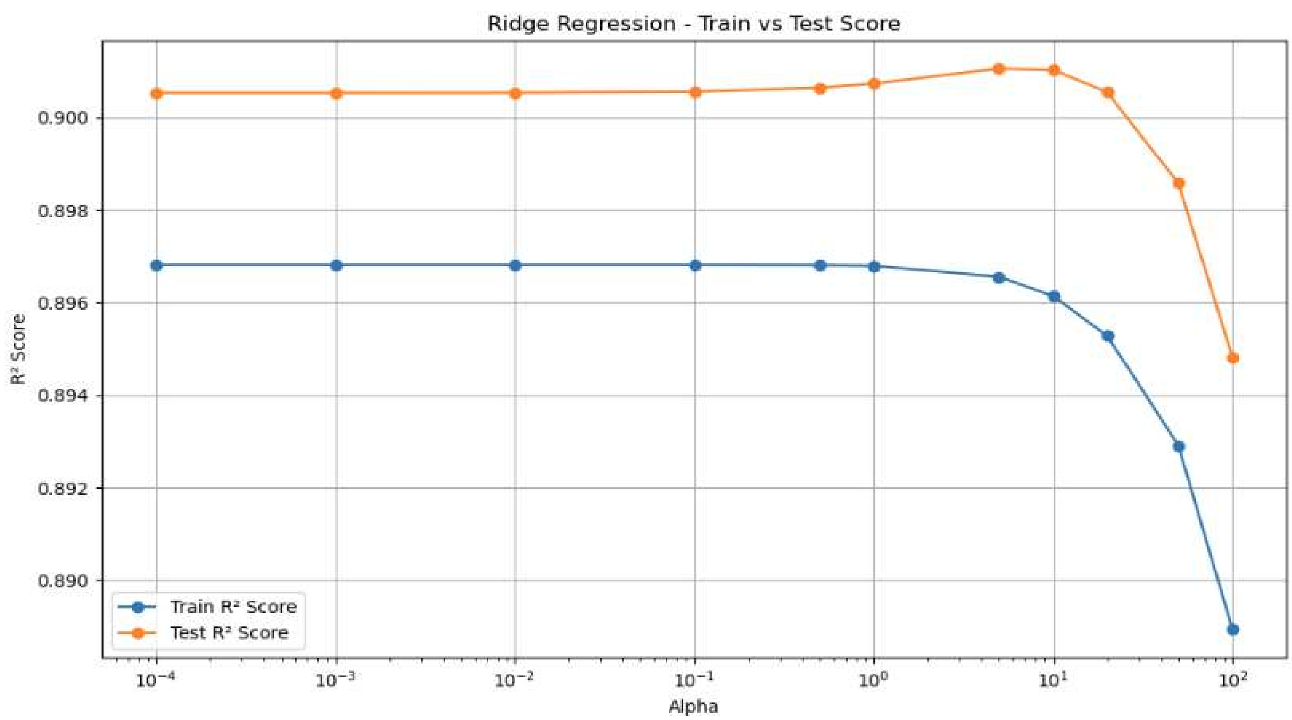
Define Alpha Values

- A list of **alpha values** was defined to tune the Ridge Regression model.
- These values range from **very small (0.0001)** to **large (100)**, covering weak to strong regularisation levels.
- The optimal alpha will be selected through tuning to balance **bias–variance trade-off**.

1.4.2.2. **3.2.2**

Ridge Regression — Alpha Tuning

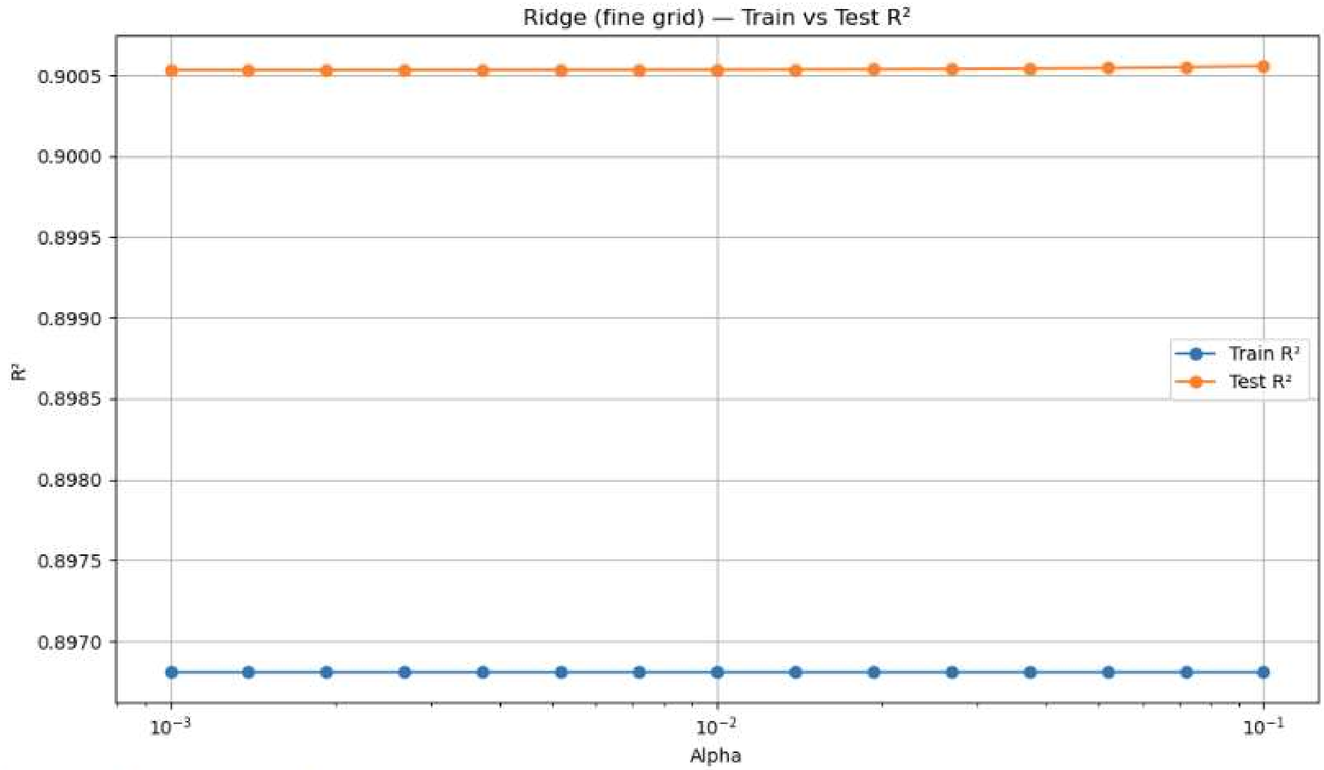
- Ridge Regression was applied using multiple alpha values from the predefined list.
- The train–test R^2 score plot showed stable performance across most alpha values.
- **Best alpha value:** 0.01
- **Best negative MAE:** -0.0953
- Regularisation helped control overfitting while maintaining high test performance.



1.4.2.3. **3.2.3**

Ridge Regression — Fine Tuning

- A smaller range of alpha values was used to fine-tune the Ridge model.
- **Best alpha (fine grid): 0.051795** with **Best negative MAE ≈ -0.0953**.
- The **Train R^2 = 0.8968** and **Test R^2 = 0.9005**, showing good generalisation.
- Feature coefficient analysis shows:
  - Top positive/negative coefficients are mostly related to **car model** and **drive/gearing type**.
  - Least impact from columns like Inspection_new, Previous_Owners, age_category_Old.
- Regularisation slightly reduced variance without major loss in accuracy.




```
Best alpha (fine grid): 0.051795
Best CV score (negative MAE): -0.095317
```

1.4.3. 3.3 Lasso Regression Implementation

1.4.3.1. **3.3.1**

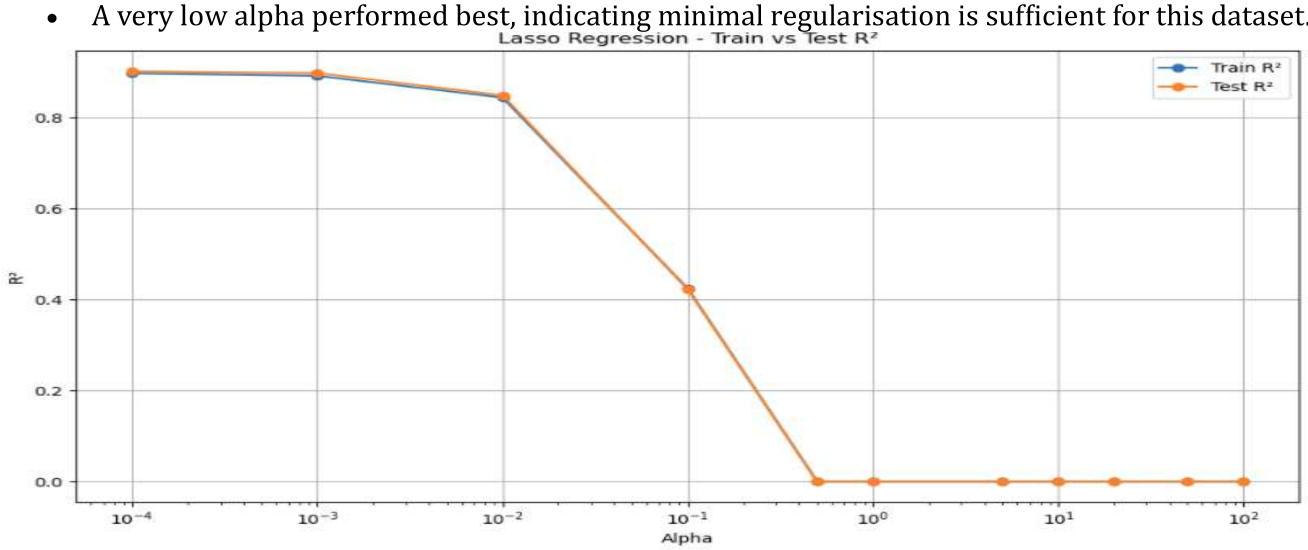
Define Alpha Values — Lasso Regression

- A list of alpha values was defined for Lasso regularisation ranging from **0.0001 to 100**.
- This range allows tuning across weak to strong regularisation strengths.
- Lasso can perform **feature selection** by shrinking some coefficients to zero.
- The optimal alpha will be selected through model evaluation and tuning.

1.4.3.2. **3.3.2**

Lasso Regression — Alpha Tuning

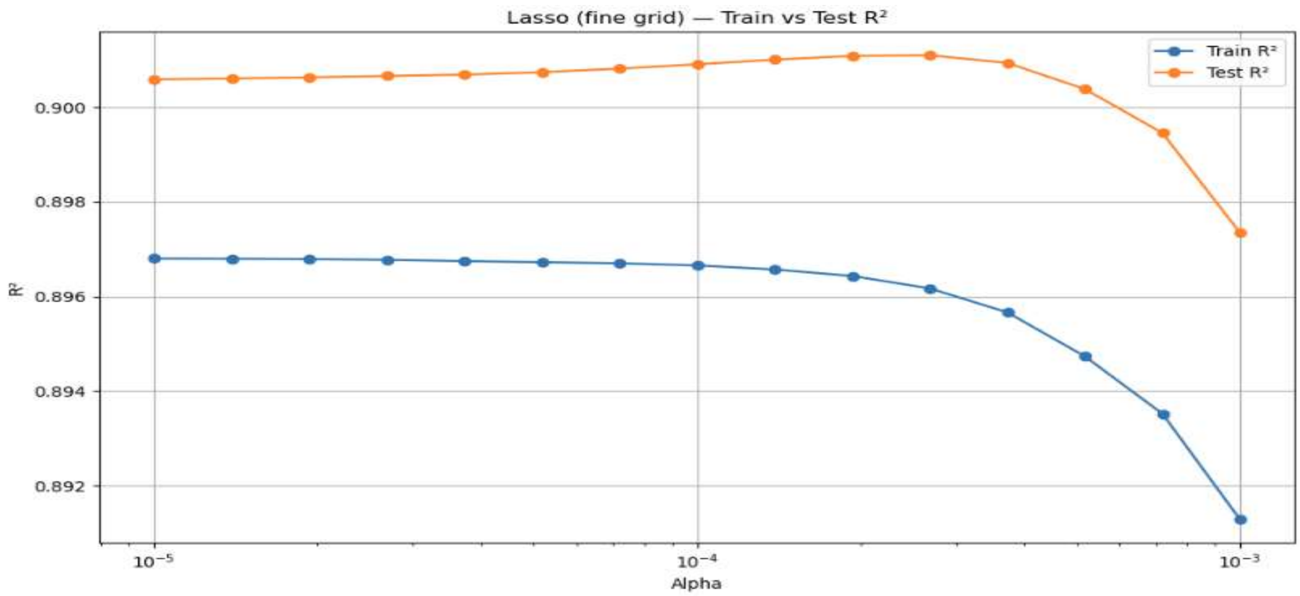
- Lasso Regression was applied across multiple alpha values.
- The R^2 scores remained stable for smaller alphas, and performance dropped at higher alphas.
- **Best alpha value:** 0.0001
- **Best negative MAE:** -0.0953
- A very low alpha performed best, indicating minimal regularisation is sufficient for this dataset.



1.4.3.3. **3.3.3**

Lasso Regression — Fine Tuning

- A fine grid of smaller alpha values (1e-5 to 1e-3) was tested.
- **Best alpha:** 0.000010 with **Best negative MAE:** -0.0953.
- Model achieved **Train R^2 = 0.8968** and **Test R^2 = 0.9006**, showing stable performance.
- Lasso performed **feature selection**, shrinking 3 coefficients to zero.
- Top influencing features are mainly related to make_model and Gearing_Type.
- Very small alpha was optimal, confirming low regularisation was sufficient.



1.4.4. 3.4 Regularisation Comparison & Analysis

1.4.4.1. **3.4.1**

Regularisation Comparison — Evaluation Metrics

- Linear Regression achieved the **highest R^2**, but may be more sensitive to noise.
- Ridge and Lasso had slightly lower scores, but **better regularisation stability**.
- Lasso additionally performed **feature selection** (3 coefficients reduced to zero).
- Ridge and Lasso results are close because **multicollinearity was already low**.

Model Performance Comparison:

	Model	Train_MSE	Test_MSE	Train_RMSE	Test_RMSE	Train_MAE	Test_MAE	Train_R2	Test_R2
0	Linear Regression	0.014	0.013	0.119	0.115	0.087	0.086	0.911	0.917
1	Ridge Regression	0.016	0.016	0.128	0.126	0.095	0.094	0.897	0.901
2	Lasso Regression	0.016	0.016	0.128	0.126	0.095	0.094	0.897	0.901

1.4.4.2. **3.4.2**

Coefficient Comparison

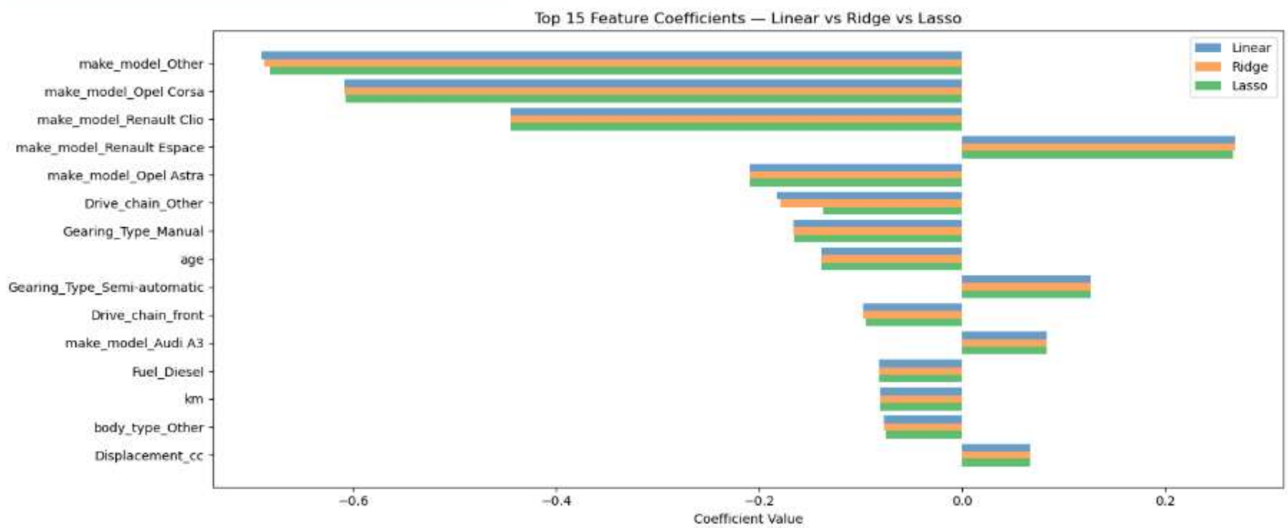
- Top 15 most influential features were compared across **Linear**, **Ridge**, and **Lasso** models.
- Ridge reduced the magnitude of coefficients but did not eliminate any features.
- Lasso not only shrunk coefficients but also set some of them to zero.
- **Number of features eliminated by Lasso:** 3
- Eliminated features: ['Previous_Owners', 'Inspection_new', 'age_category_Old'].

The plot shows how regularisation impacts feature weights:

- Linear model retains full weight.
- Ridge smoothens coefficients.
- Lasso drops less important features entirely.

```
Top 15 coefficients by absolute value (Linear):
```

	Feature	Linear	Ridge	Lasso	abs_Linear
11	make_model_Other	-0.690	-0.688	-0.683	0.690
9	make_model_Opel Corsa	-0.609	-0.608	-0.608	0.609
12	make_model_Renault Clio	-0.445	-0.445	-0.445	0.445
13	make_model_Renault Espace	0.269	0.269	0.267	0.269
8	make_model_Opel Astra	-0.209	-0.209	-0.209	0.209
25	Drive_chain_Other	-0.182	-0.179	-0.137	0.182
23	Gearing_Type_Manual	-0.166	-0.166	-0.166	0.166
2	age	-0.139	-0.139	-0.139	0.139
24	Gearing_Type_Semi-automatic	0.127	0.127	0.126	0.127
26	Drive_chain_front	-0.097	-0.097	-0.094	0.097
7	make_model_Audi A3	0.083	0.083	0.083	0.083
18	Fuel_Diesel	-0.082	-0.082	-0.081	0.082
0	km	-0.081	-0.081	-0.081	0.081
14	body_type_Other	-0.077	-0.077	-0.075	0.077
5	Displacement_cc	0.067	0.067	0.067	0.067




```
Number of features eliminated by Lasso: 3
Eliminated features: ['Previous_Owners', 'Inspection_new', 'age_category_Old']
```

4.1 Conclusion & Key Takeaways

The baseline Linear Regression model achieved the best R^2 score of **0.917** on the test data, showing that a simple linear model can perform very well on this dataset. Ridge and Lasso regularisation methods gave slightly lower R^2 scores (around **0.900**) but improved the model's **stability** and **robustness**.

Regularisation techniques helped to control the model complexity. Ridge shrunk the magnitude of the coefficients, while Lasso additionally **eliminated three less important features** (Previous_Owners, Inspection_new, and age_category_Old). This made the model more interpretable without a major drop in performance.

No significant overfitting was observed, as training and testing scores were very close. The dataset was sufficiently large and clean, which contributed to the strong performance of linear models. Although regularisation did not significantly increase accuracy, it improved **generalisation** and **feature selection**.

Overall, the linear model was sufficient for this prediction task. Regularisation improved interpretability and reduced the influence of less important features, making the model more reliable for deployment. In future, more advanced or non-linear models can be tested if more complex relationships need to be captured.