

## Car Data Analysis Report

### Basic Data Inspection

```
=====
Unnamed: 0    model year price ... tax mpg engineSize tax(£)
0      0      Fiesta 2015  6998 ... 125.0 54.3    1.2   NaN
1      1      Astra 2017 10998 ... 150.0 72.4    1.6   NaN
2      2  Grandland X 2019 22995 ... 145.0 57.7    1.5   NaN
3      3      Yaris 2018  9990 ... 150.0 58.9    1.5   NaN
4      4      5 Series 2019 33900 ... 145.0 56.5    3.0   NaN
```

[5 rows x 11 columns]

None

```

      Unnamed: 0    year ... engineSize    tax(£)
count 108540.000000 108540.000000 ... 108540.000000 4860.000000
mean   54269.500000 2017.098028 ...    1.661644 121.147119
std    31332.943446  2.130057 ...    0.557058  58.003289
min      0.000000 1970.000000 ...    0.000000  0.000000
25%    27134.750000 2016.000000 ...    1.200000 125.000000
50%    54269.500000 2017.000000 ...    1.600000 145.000000
75%    81404.250000 2019.000000 ...    2.000000 145.000000
max    108539.000000 2060.000000 ...    6.600000 555.000000
```

[8 rows x 8 columns]

### Missing Data Check

```
=====
```

No missing data.

### K-Means Clustering with Tax-Centric Features

```
=====
```

K-Means Silhouette Score: 0.4473094344377048

### Linear Regression Model Performance for Price Prediction

```
=====
```

R-squared Score: 0.6958108636850833

Mean Absolute Error (MAE): 3811.5700675406815

Mean Squared Error (MSE): 29690394.004996125

### Grid Search Results

```
=====
```

Best Parameters: {'max\_depth': 5, 'max\_features': 'sqrt', 'min\_samples\_leaf': 1, 'min\_samples\_split': 3, 'n\_estimators': 200}

Best R-squared Score from Grid Search: 0.7756487079029595

### Model Evaluation

```
=====
```

R-squared Score: 0.7803643356110432

Mean Absolute Error (MAE): 3208.3648349087157

Mean Squared Error (MSE): 21437548.665466417

# K-Means Clustering Plot

