Used Car Price Prediction

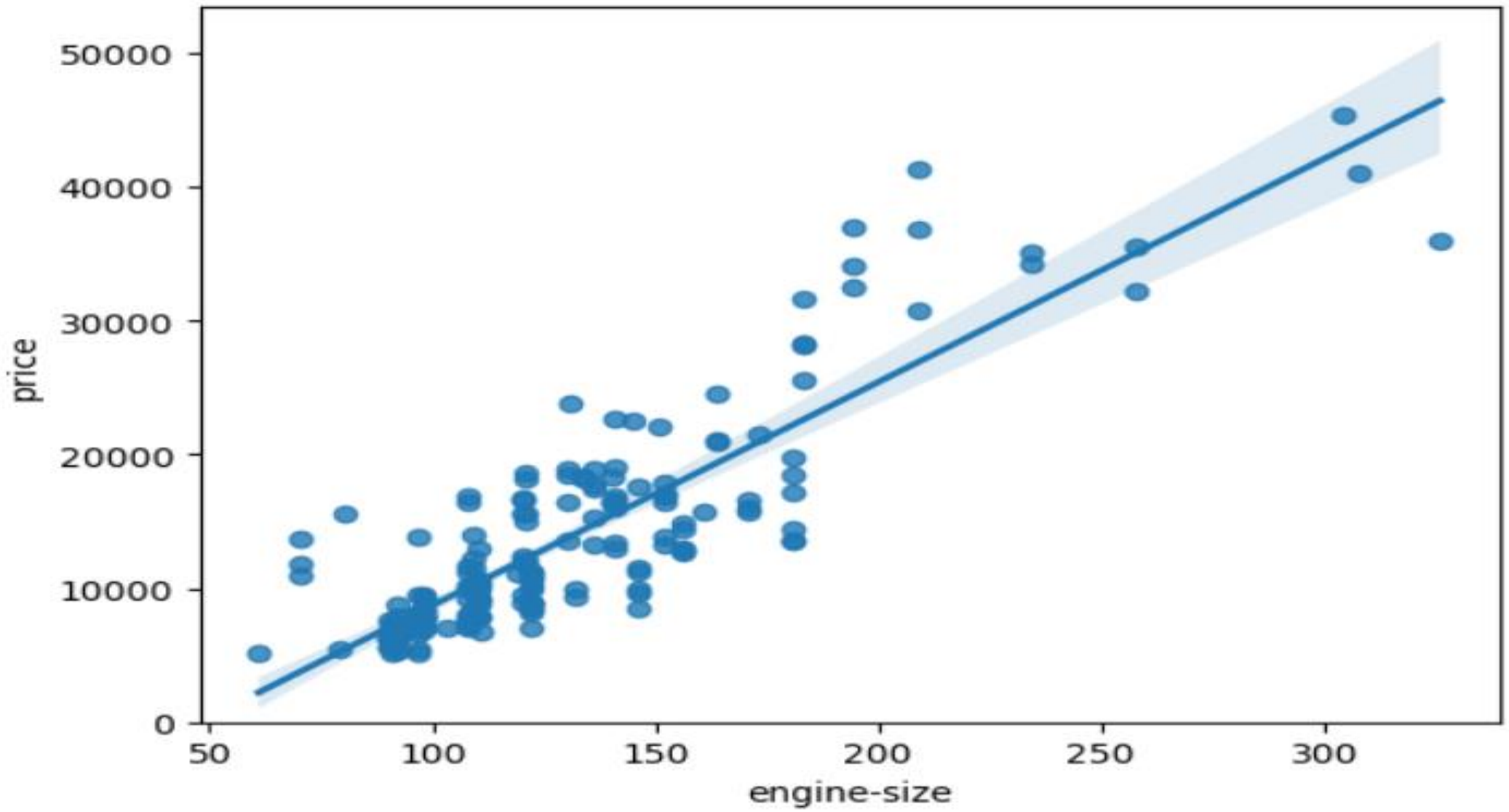
A DATA SCIENCE PROJECT TO ESTIMATE FAIR MARKET PRICE FOR USED CARS

Business Problem

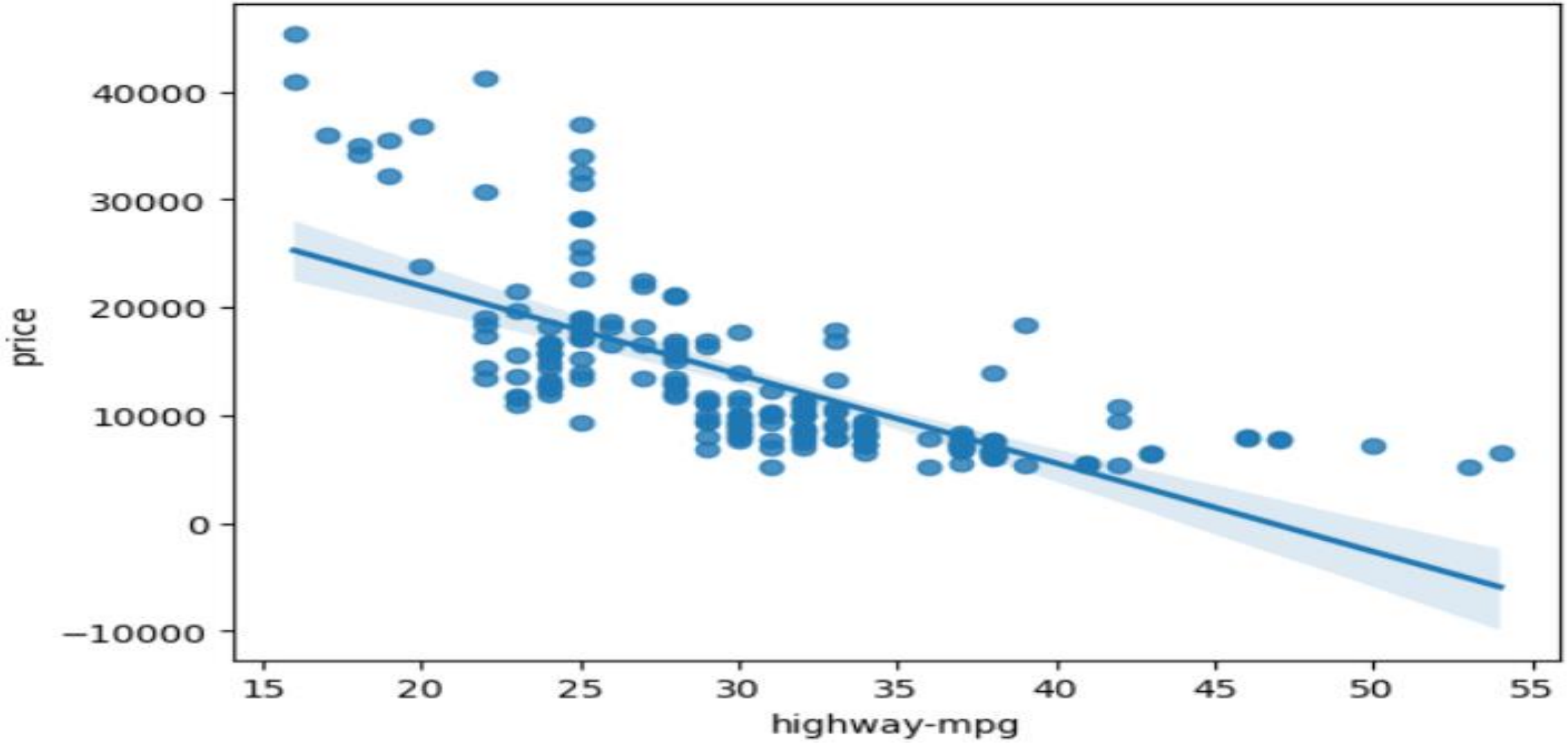
Our friend Tom wants to sell his used car but doesn't know the right price to ask. He wants to maximize the price without scaring away potential buyers. The goal is to help Tom estimate a fair and competitive price using data analysis and machine learning techniques.

EDA (Exploratory Data Analysis)

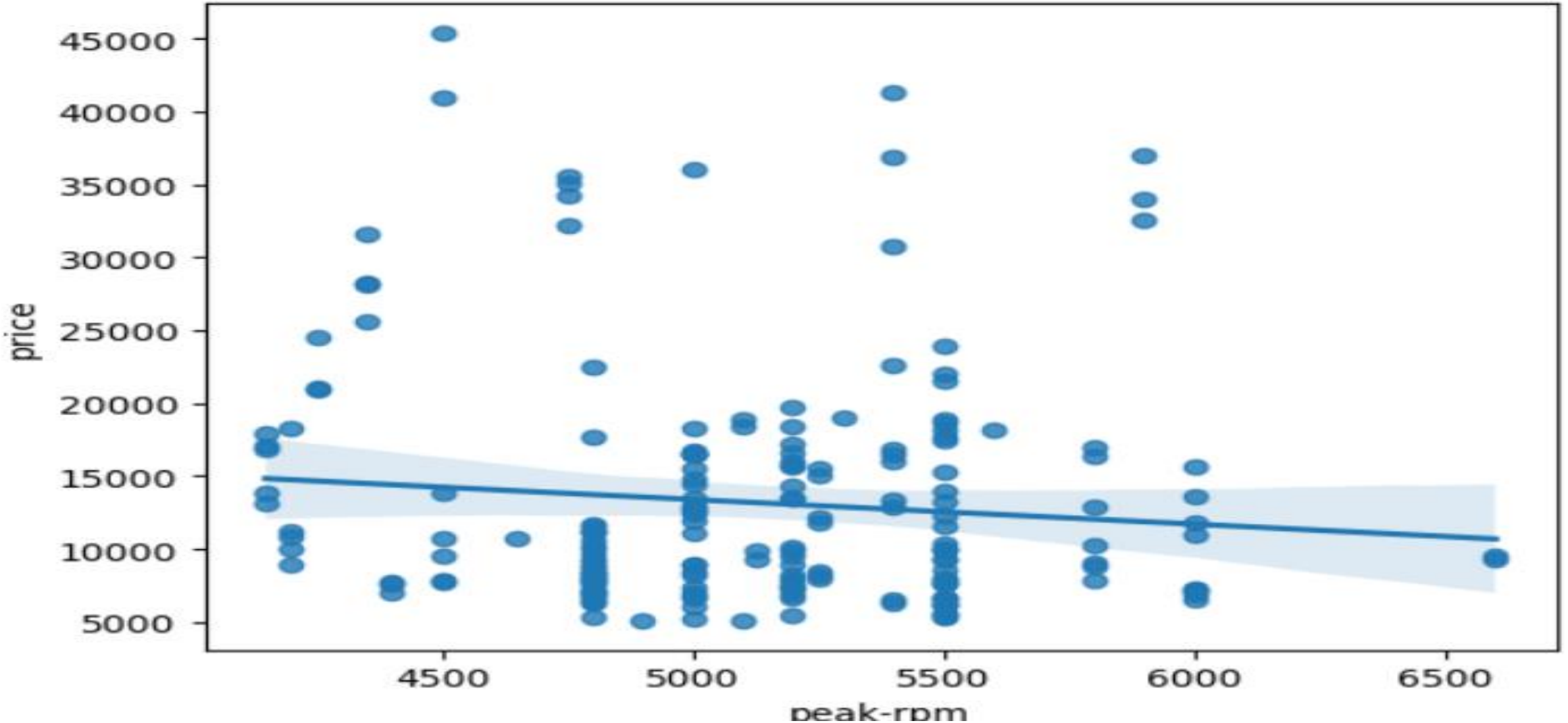
Scatterplot of 'engine-size' and 'price'



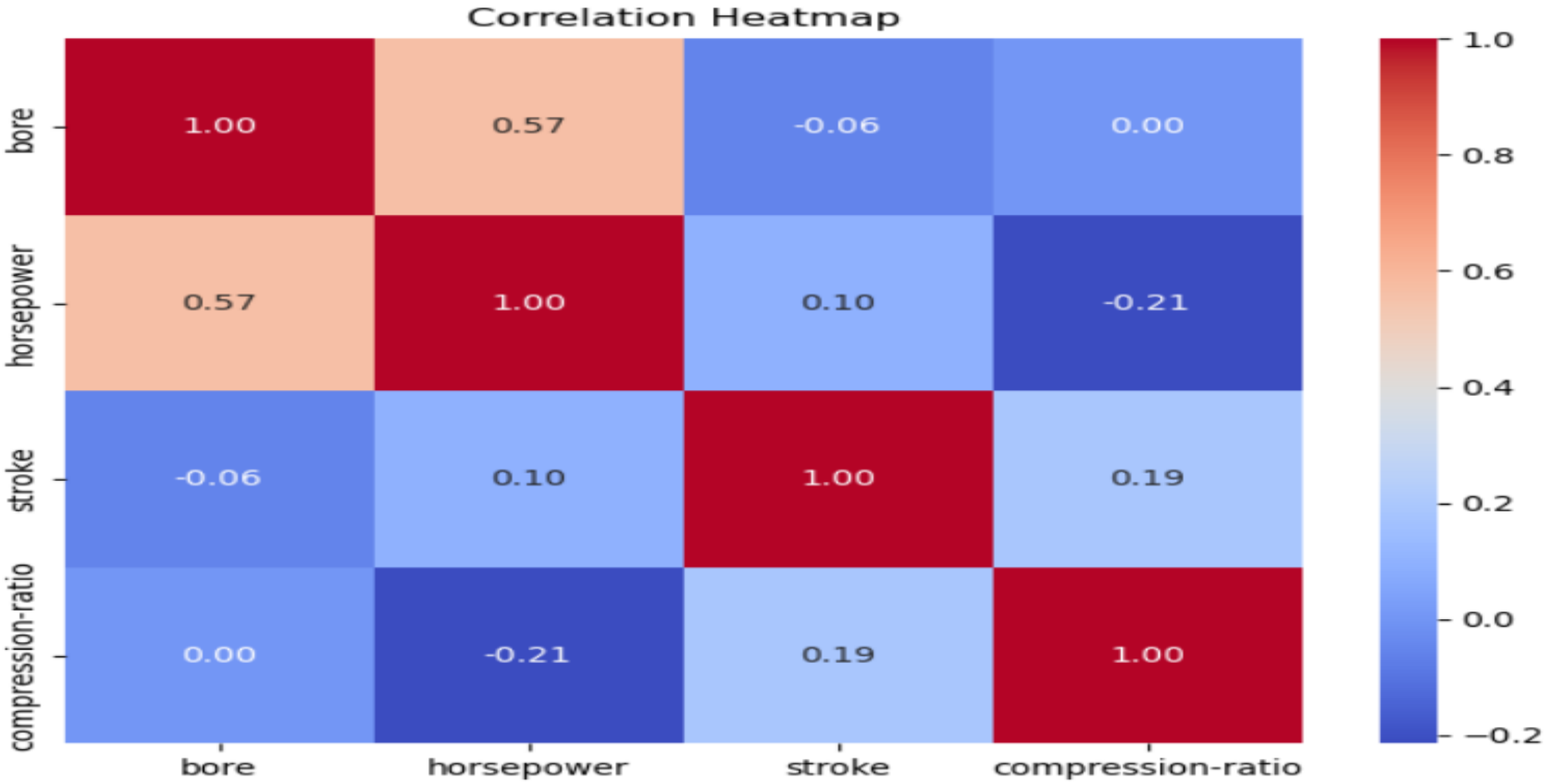
Scatterplot of ‘highway-mpg' and 'price'



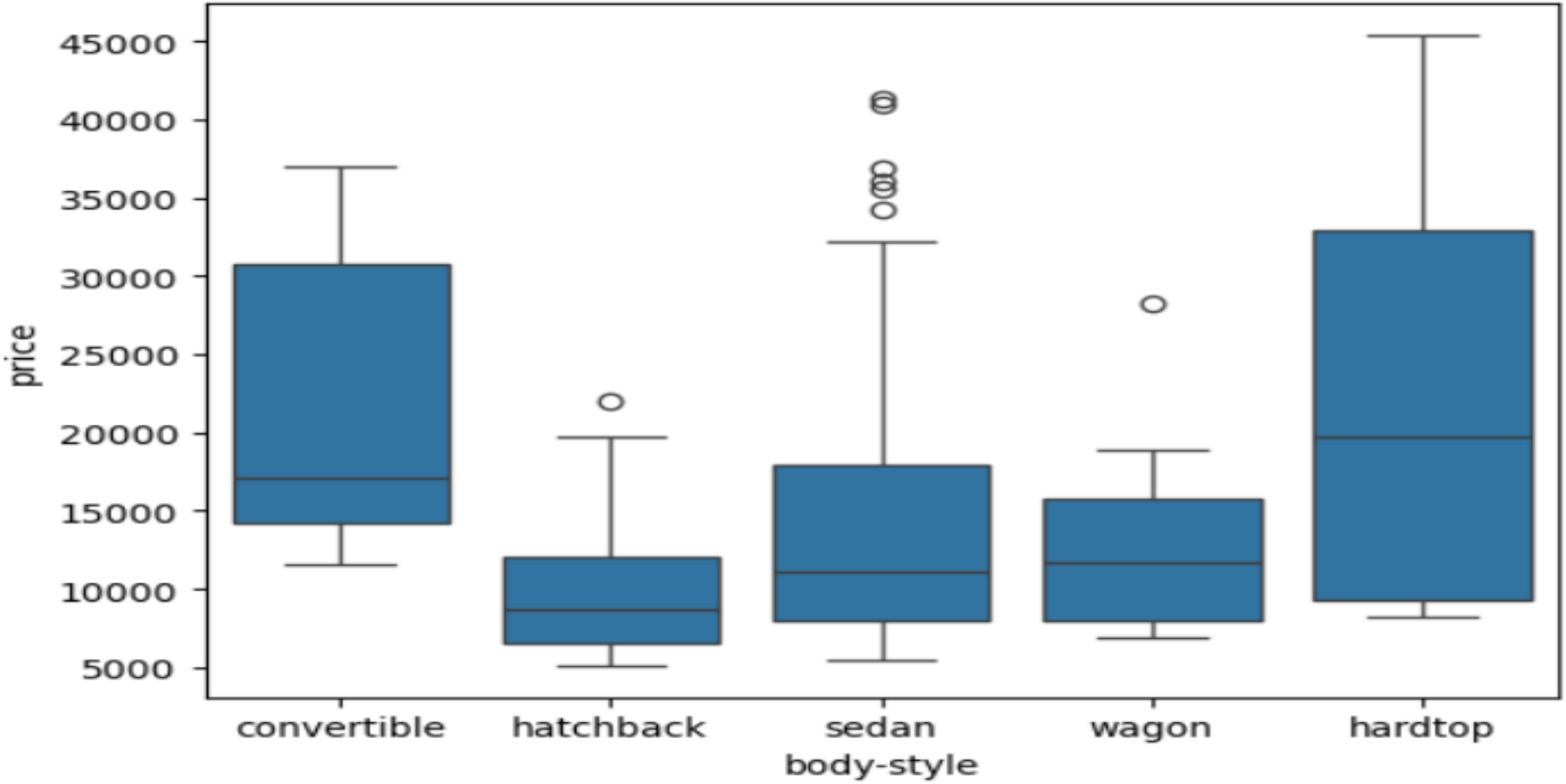
Scatterplot of 'peak-rpm' and 'price'



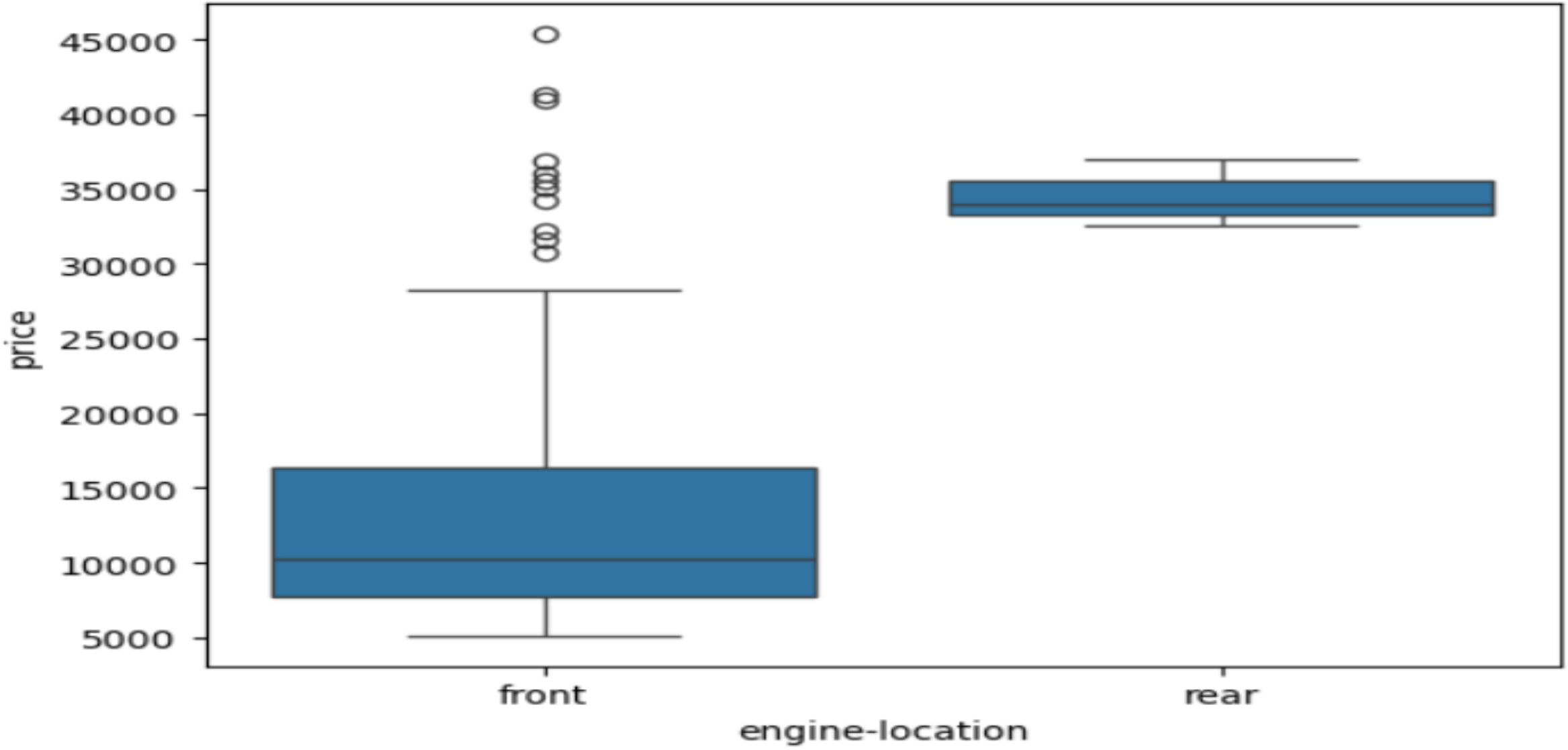
Heatmap Plot



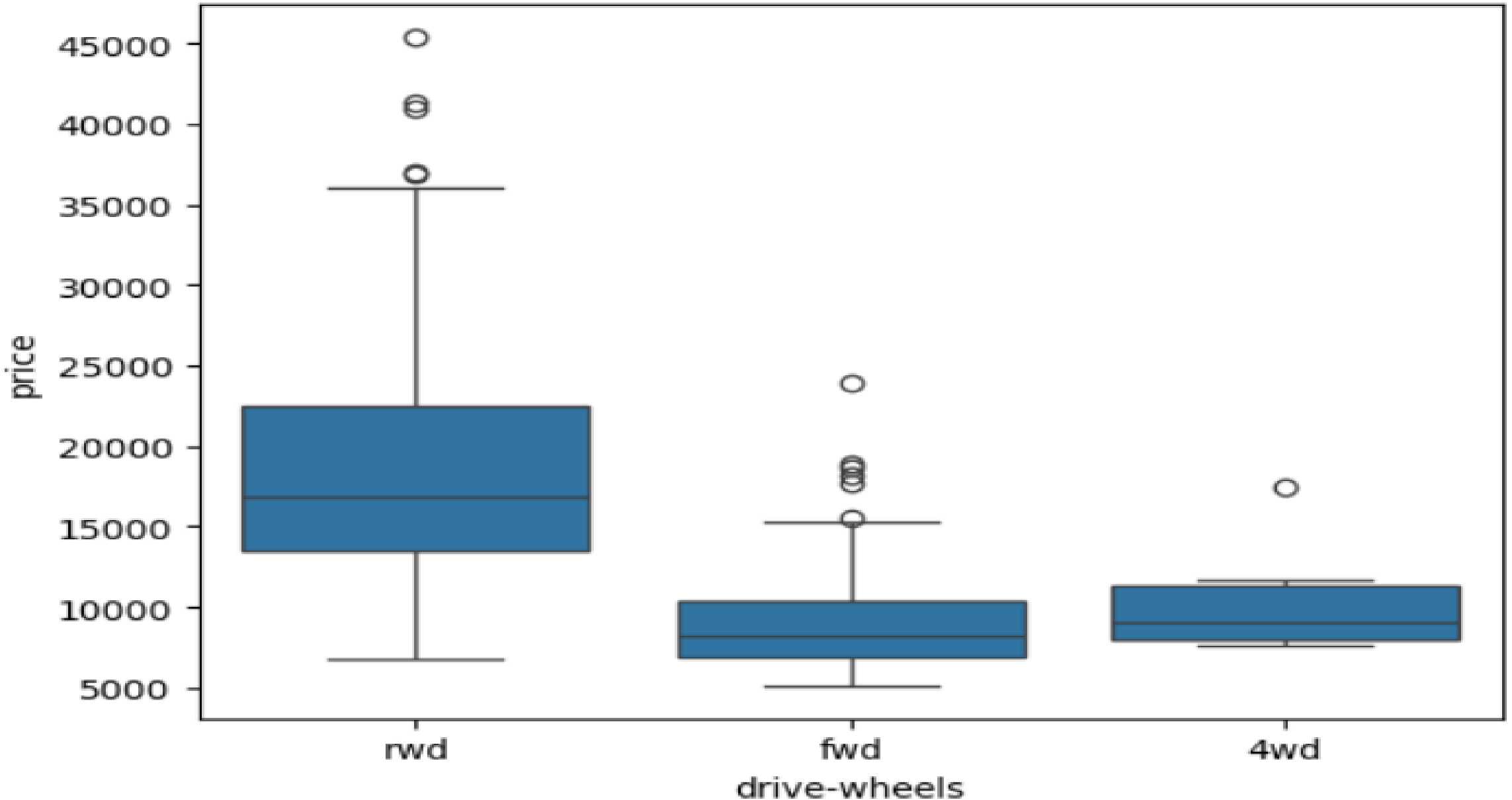
BoxPlot of ‘Body-style’ and ‘Price’

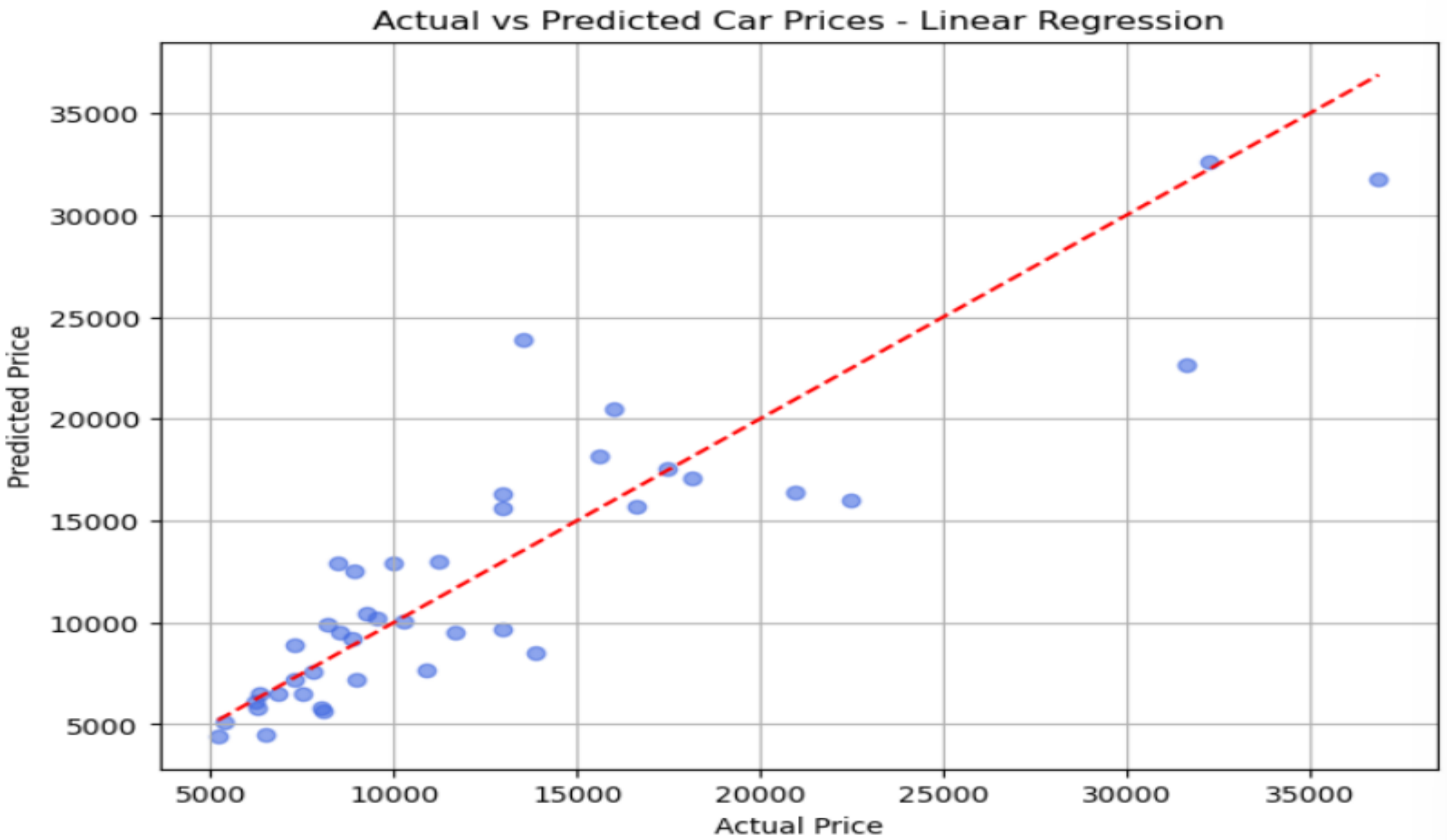


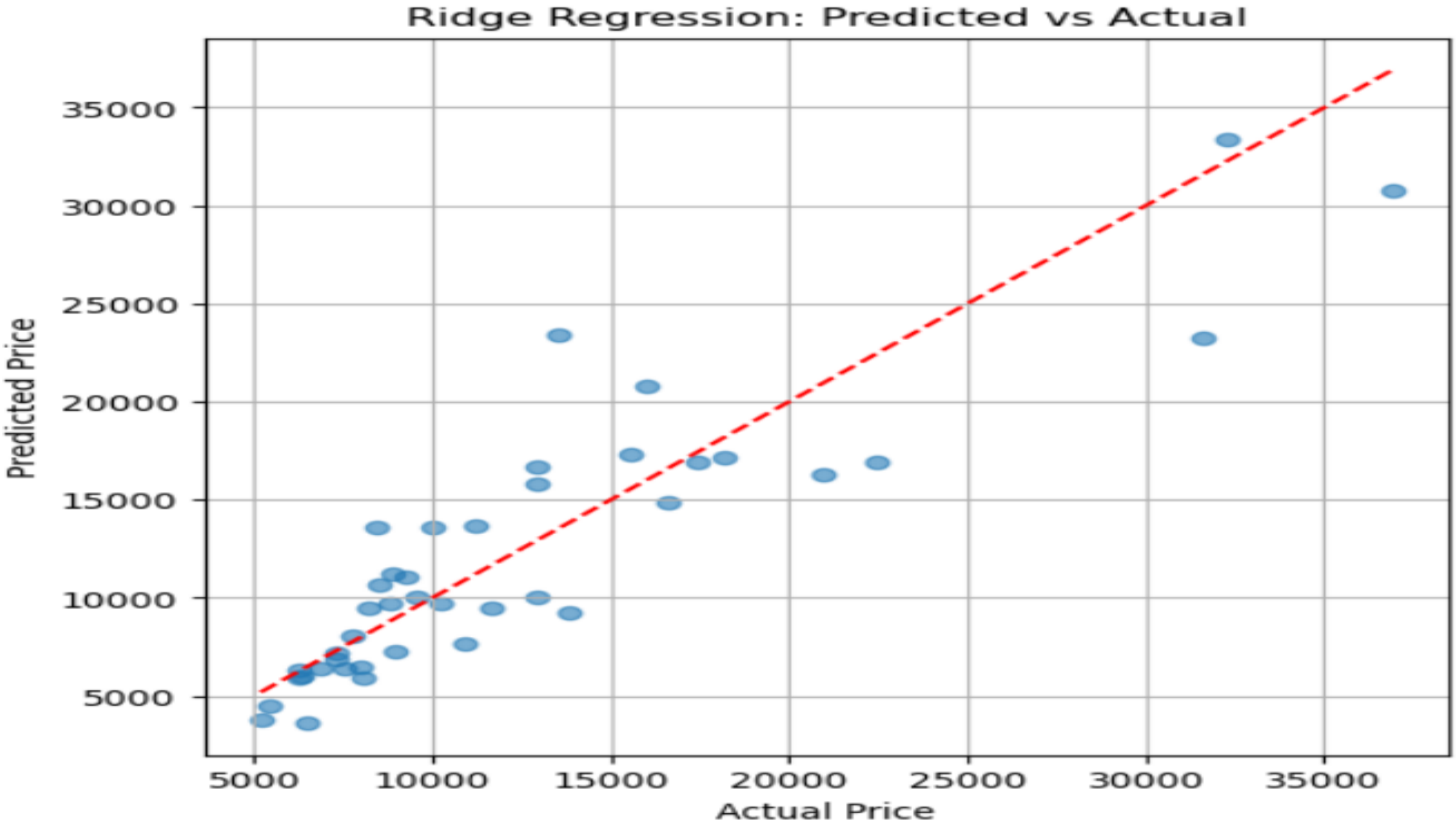
BoxPlot of ‘engine-location’ and ‘Price’

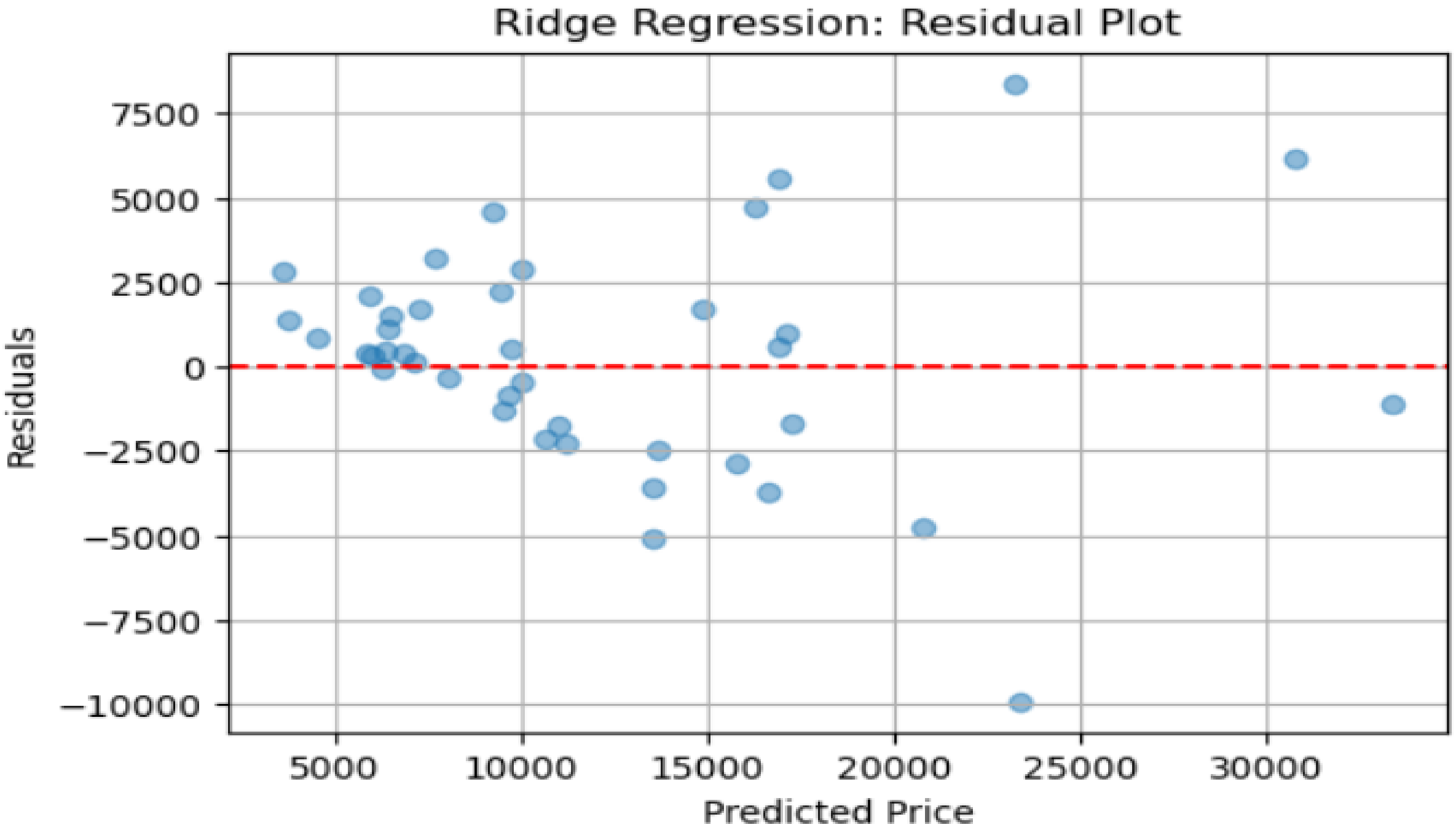


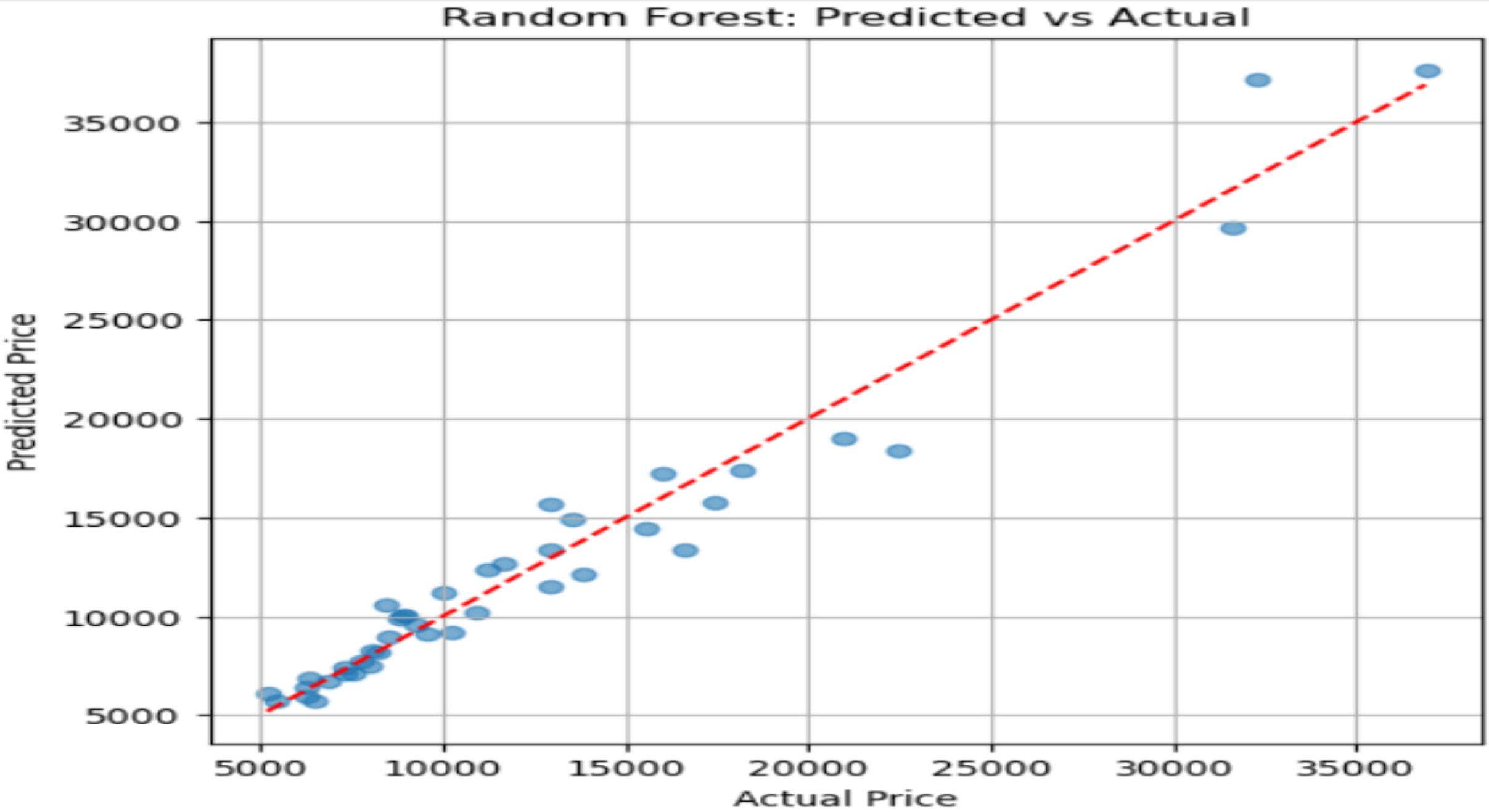
BoxPlot of ‘drive-wheels’ and ‘Price’

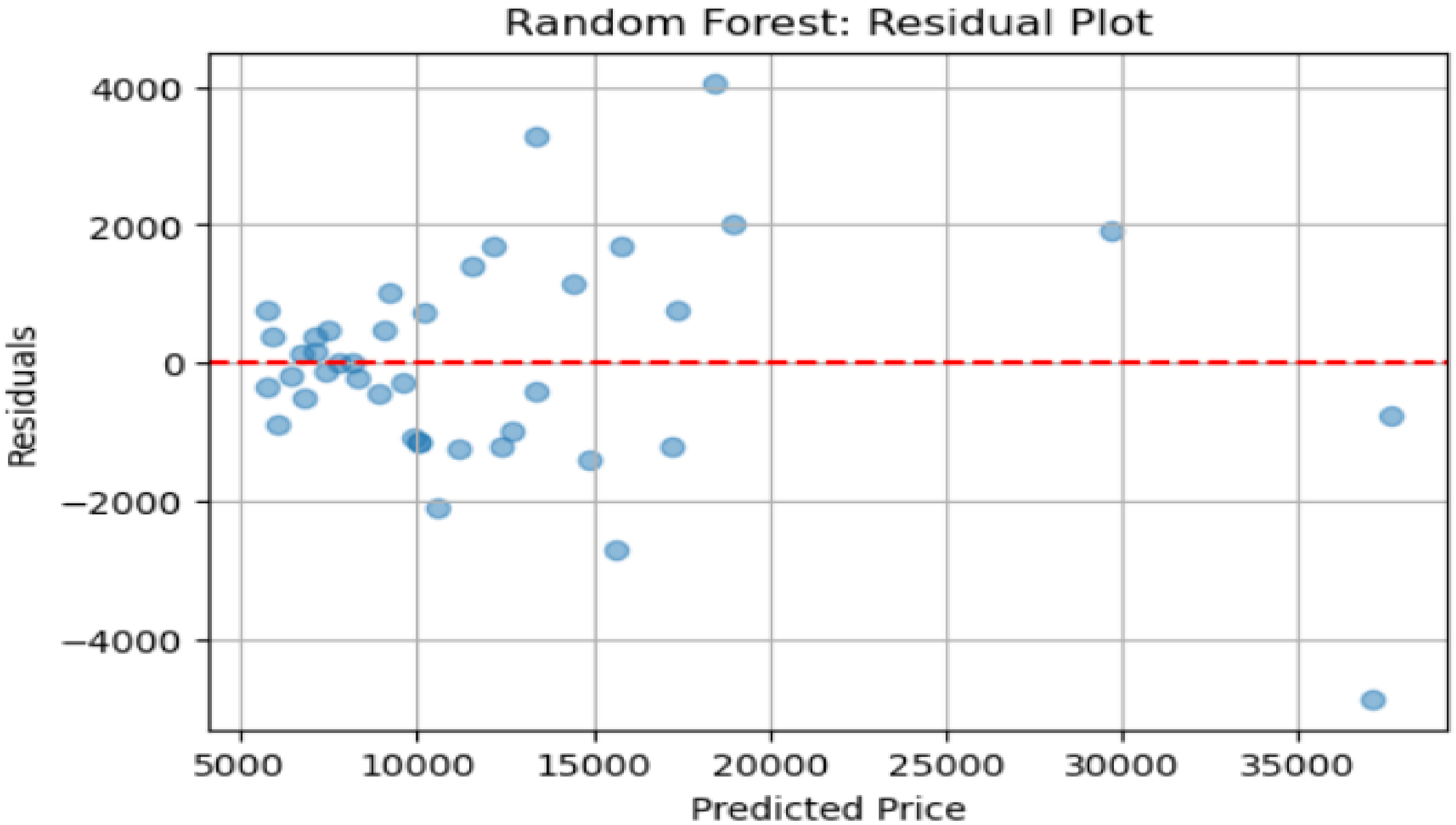


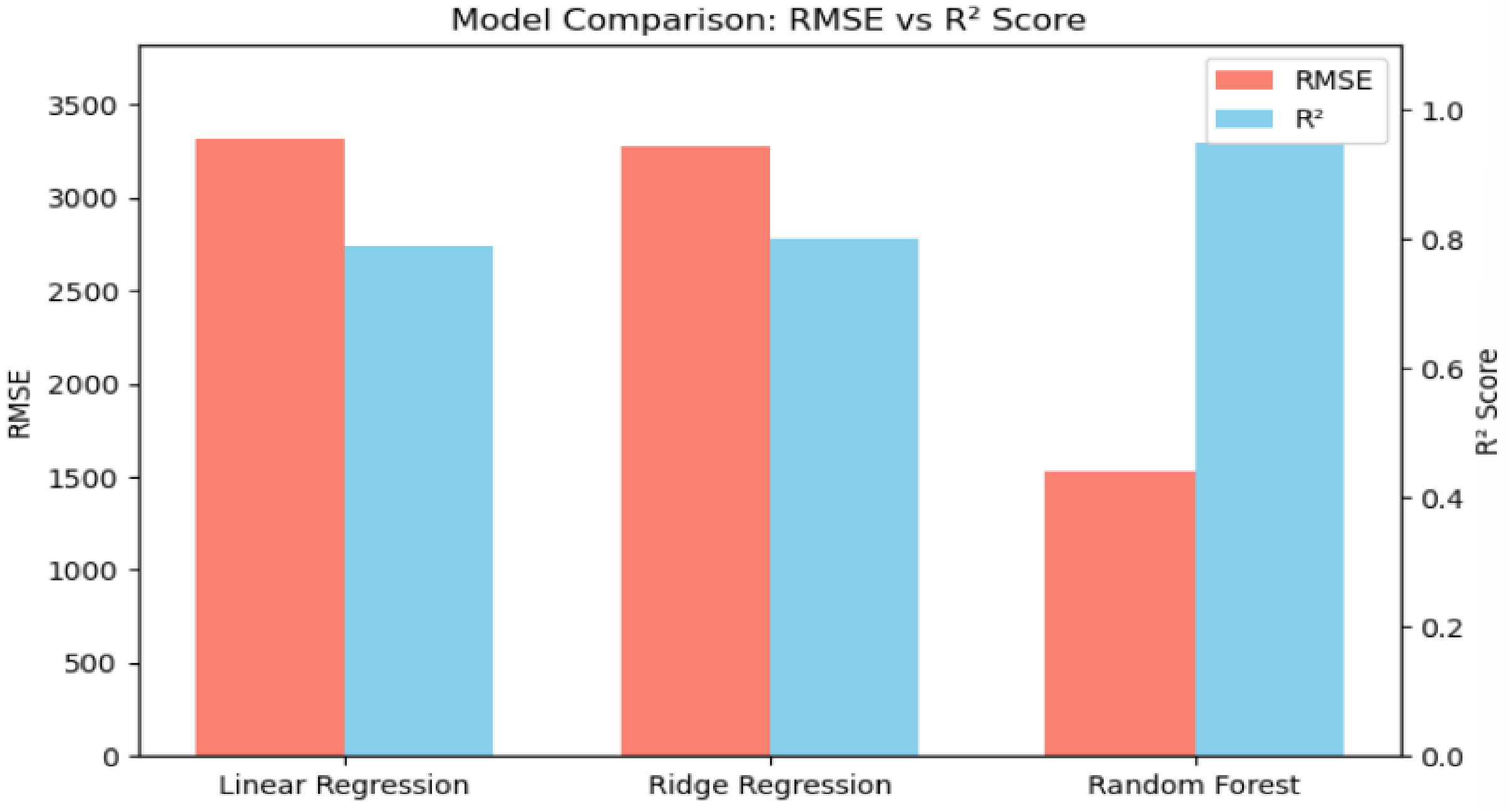

Actual vs Predicted Car Prices - Linear Regression
Predicted Price
Actual Price
5000
10000
15000
20000
25000
30000
35000


Ridge Regression: Predicted vs Actual
Predicted Price
Actual Price
5000
10000
15000
20000
25000
30000
35000


Ridge Regression: Residual Plot
Residuals
Predicted Price
7500
5000
2500
0
−2500
−5000
−7500
−10000
5000
10000
15000
20000
25000
30000


Random Forest: Predicted vs Actual
Predicted Price
Actual Price
5000
10000
15000
20000
25000
30000
35000


Random Forest: Residual Plot
Residuals
Predicted Price
4000
2000
0
−2000
−4000
5000
10000
15000
20000
25000
30000
35000


Model Comparison: RMSE vs R² Score
RMSE
R²
RMSE
R² Score
3500
3000
2500
2000
1500
1000
500
0
1.0
0.8
0.6
0.4
0.2
0.0
Linear Regression
Ridge Regression
Random Forest

Conclusion

- **Key Takeaways:**
- **Random Forest Regressor** delivered the best performance
  - Train R^2: 0.988
  - Test R^2: 0.956
  - RMSE: 1533.12
- **Ridge Regression** and **Linear Regression** performed well, but less accurate than Random Forest.
- The model captures **key price-driving features** such as engine size, horsepower, curb weight, and fuel type.
- **Cross-validation confirms** model reliability:
  - Random Forest CV Mean R^2: **0.907** (lowest std dev = most stable)
- **Business Insight:**
- The final model can be reliably used to **predict car prices** based on car attributes.
- This can help sellers like "Tom" set **fair and data-driven prices**.