1.5 Discussion and interpretation

(1) Best Model Performance

Based on the test set metrics, the Polynomial Regression (Degree 4) technically performed the best, achieving the highest Test R^2 of 0.9138 and the lowest Test MSE. However, the Linear Regression model was extremely close behind with a Test R^2 of 0.9125.

- Since the improvement from Degree 1 to Degree 4 is less than 0.2%, we conclude that the underlying relationship between Fuel Economy and Horsepower is predominantly linear. In a real-world engineering context, we would likely choose the Linear model for its simplicity and interpretability, despite the marginal statistical gain of the Degree 4 model.

(2) Impact of Polynomial Degree

Increasing the polynomial degree did not consistently yield significant improvements.

- Observation: Moving from Linear ($R^2 \approx 0.91$) to Degree 2 actually caused a slight drop in performance on the test set (Test MSE increased from 318 to 331).

- Conclusion: This indicates that simply adding complexity does not guarantee better generalization. The data does not exhibit strong non-linear curvature that requires higher-order terms.

(3) Residual Error Analysis

Although the models performed well ($R^2 > 0.9$), there is still unexplained variance (approx. 9%). Plausible reasons for this remaining error include:

- Insufficient Feature Information: We are predicting Horsepower only based on MPG. In reality, HP is also determined by engine displacement, turbocharging, and even vehicle weight, which are missing from this dataset.

- Inherent Noise: Automotive data often contains variation due to different testing conditions or driving behaviors that cannot be captured by a deterministic curve.

2.5 Discussion and interpretation

(1) Best Generalization

The Linear Regression model generalized best, achieving the highest Test R^2 of 0.299. While this score is low objectively, it significantly outperformed the polynomial models.

- Implication: This suggests the relationship between weather and electricity usage in this dataset contains a linear component, but the signal is weak compared to the noise.

(2) Polynomial Model Failure

Polynomial models did not improve the fit compared to linear regression.

- Observation: The Test R^2 dropped from 0.299 (Linear) to 0.279 (Degree 2) and collapsed to -33.31 (Degree 4).
- Reason: While we physically expect a non-linear "U-shaped" dependency (high AC usage in summer, heating in winter), the polynomial models likely failed because the dataset size (1,433 samples) was not sufficient to constrain the high-variance behavior of high-degree polynomials, leading them to fit random noise instead of the physical curve.

(3) Evidence of Overfitting (Degree 4)

The Degree 4 model exhibits a textbook case of overfitting:

- Metric Evidence: It achieved the lowest Training MSE (251,909), indicating it memorized the training data better than any other model. However, its Test MSE exploded to 1.21e+07, causing the negative R^2.
- Explanation: The model became too complex, creating wild oscillations between data points to satisfy the training set, which resulted in massive errors when predicting on unseen test data.

(4) Reasons for Low Overall Performance

None of the models achieved a "good" Test R^2 (all were below 0.3). Two plausible reasons supported by this outcome are:

- Unmodeled Drivers (Occupancy/Behavior): Electricity consumption is heavily driven by human behavior (e.g., weekends vs. weekdays, holidays, work-from-home patterns). This dataset *only* provides weather features, completely missing the "human factor" which likely accounts for the majority of the variance.
- Limited Feature Set: We only have simple weather metrics (Wind, Precip, Temp). We are missing critical variables like "Previous Day's Consumption" (lag features) or "Time of Year" (seasonality), which are standard requirements for accurate time-series forecasting.