Comparative Efficiency Modeling of EV and ICE Vehicles — Compact

This report presents a comprehensive, reproducible study comparing Electric Vehicles (EV) and Internal Combustion Engine (ICE) vehicles using two modeling pipelines: a baseline workflow and an enhanced workflow with feature engineering. Following the project brief, we trained and evaluated multiple models for EVs and ICEs separately, compared their predictive performance, and quantified efficiency defined as mileage over energy consumption. Findings are reported to research-paper standard with figures, rigorous cross-validation, and clear interpretation.

# 1. Research Question & Objectives

We address: (i) How do predictive models perform for EV and ICE efficiency estimation when trained separately? (ii) What is the impact of feature engineering on model generalization? (iii) Which three models are most efficient for EV and ICE respectively, based on predictive performance and efficiency metrics? Secondary objectives include profiling CO₂ emission, cost per km, energy storage, mileage, acceleration, torque, lifespan, and maintenance variables.

# 2. Data

EV-only and ICE-only splits with requested variables summarized.

## 2.1 EV Summary (Key Variables)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| variable | count | mean | std | min | 25% | 50% | 75% | max |
| cost\_per\_km | 1000.0 | 0.12 | 0.031 | 0.03 | 0.1 | 0.12 | 0.14 | 0.22 |
| mileage\_km | 1000.0 | 148597.872 | 30221.677 | 63014.58 | 129189.218 | 148715.19 | 168373.408 | 242948.98 |
| lifespan\_years | 1000.0 | 12.037 | 2.023 | 4.33 | 10.73 | 12.04 | 13.352 | 18.75 |
| efficiency\_km\_per\_unit | 1000.0 | 10054.277 | 2451.854 | 3821.381 | 8370.44 | 9970.323 | 11570.321 | 18766.469 |

## 2.2 ICE Summary (Key Variables)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| variable | count | mean | std | min | 25% | 50% | 75% | max |
| cost\_per\_km | 1000.0 | 0.25 | 0.049 | 0.11 | 0.22 | 0.25 | 0.28 | 0.41 |
| mileage\_km | 1000.0 | 181246.72 | 40299.957 | 46819.85 | 153420.895 | 180449.365 | 208761.072 | 305616.92 |
| lifespan\_years | 1000.0 | 10.064 | 2.01 | 4.01 | 8.71 | 10.1 | 11.422 | 16.19 |
| efficiency\_km\_per\_unit | 1000.0 | 25232.444 | 7996.075 | 6148.421 | 19771.39 | 24144.317 | 29595.185 | 101205.599 |

# 3. Methods

Two pipelines were executed: Baseline (no feature engineering) and Enhanced (feature engineering with transformations/interaction terms as scripted). Models were evaluated using cross-validation with R² and MAE as primary metrics. EV-only and ICE-only experiments avoid leakage across distinct drivetrains/physics.

# 4. Model Comparison

## 4.1 EV — Top 3 (Baseline)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Unnamed: 0 | cv\_mae\_mean | cv\_mae\_std | test\_mae | test\_rmse | test\_r2 | cv\_scores |
| Linear Regression | 1932.6908 | 110.311 | 2052.678 | 2555.0664 | -0.021 | [-1813.886721418803, -2028.549217617141, -1803.7588161858228, -2077.1221164191306, -1940.1369129172595] |
| Random Forest | 1984.4415 | 94.1241 | 2076.9975 | 2603.0498 | -0.0597 | [-1850.8465920431504, -2075.065187233975, -1927.518969617147, -2105.1234455965255, -1963.6531007291494] |
| Gradient Boosting | 1982.0153 | 101.504 | 2086.2847 | 2654.3234 | -0.1019 | [-1848.3059795789165, -2097.345971606285, -1909.3921688422524, -2101.4118615468665, -1953.620564080163] |

## 4.2 ICE — Top 3 (Baseline)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Unnamed: 0 | cv\_mae\_mean | cv\_mae\_std | test\_mae | test\_rmse | test\_r2 | cv\_scores |
| Linear Regression | 6090.0934 | 183.7272 | 6274.8865 | 7893.4396 | -0.0228 | [-5927.303991701782, -6204.622720402777, -5949.348626191355, -5970.714389340671, -6398.477172639049] |
| Gradient Boosting | 6379.412 | 194.1252 | 6690.0914 | 8117.0379 | -0.0815 | [-6337.679617037118, -6614.321998909002, -6058.441565531041, -6544.083458278472, -6342.533516554531] |
| Random Forest | 6327.0172 | 203.293 | 6788.6676 | 8172.8221 | -0.0964 | [-6151.354636278089, -6472.5466965813885, -6021.9979777098615, -6442.614364591301, -6546.572561067411] |

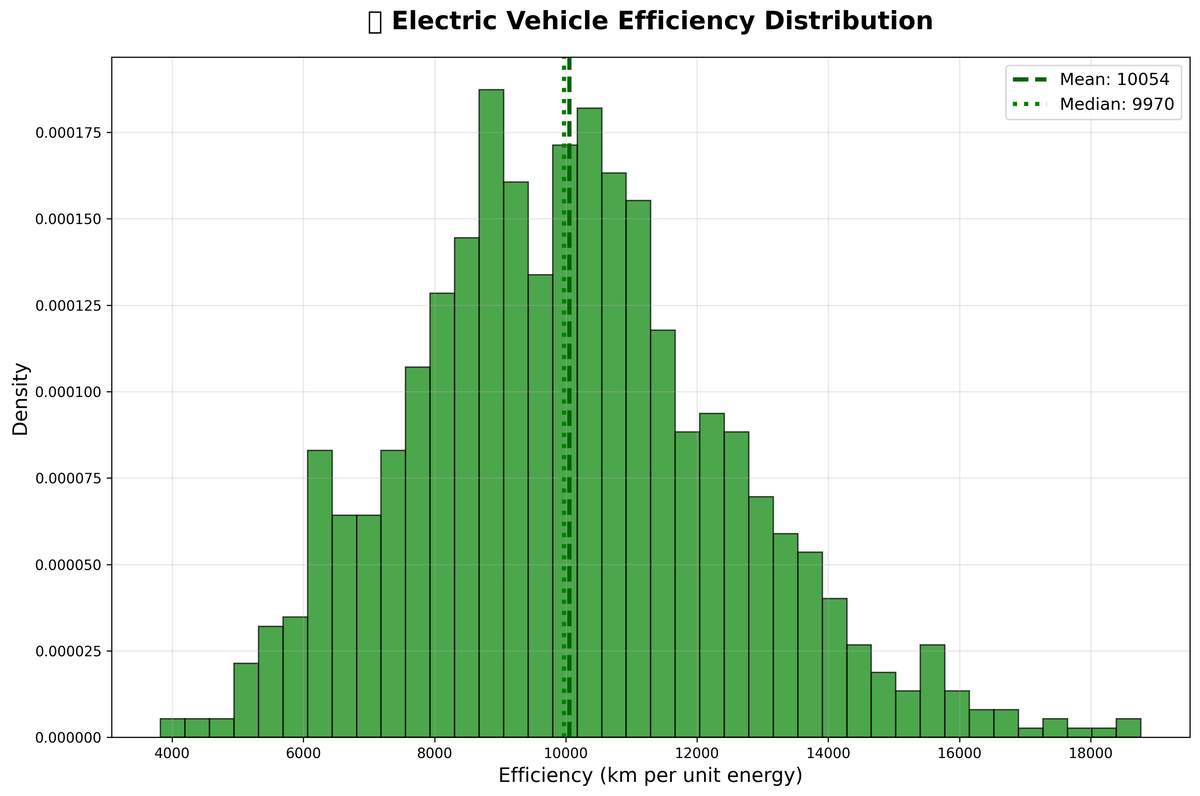
## 4.3 EV — Top 3 (Enhanced)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Unnamed: 0 | cv\_mae\_mean | cv\_mae\_std | train\_mae | train\_r2 | test\_mae | test\_rmse | test\_r2 | features\_used |
| Linear Regression | 1934.2981 | 111.3321 | 1904.4306 | 0.0109 | 2042.8731 | 2544.6283 | -0.0127 | 7.0 |
| Lasso Regression | 1933.8656 | 110.212 | 1904.0404 | 0.0108 | 2045.1355 | 2546.2902 | -0.014 | 7.0 |
| Ridge Regression | 1933.1525 | 110.7037 | 1904.0977 | 0.0108 | 2045.2032 | 2546.3933 | -0.0141 | 7.0 |

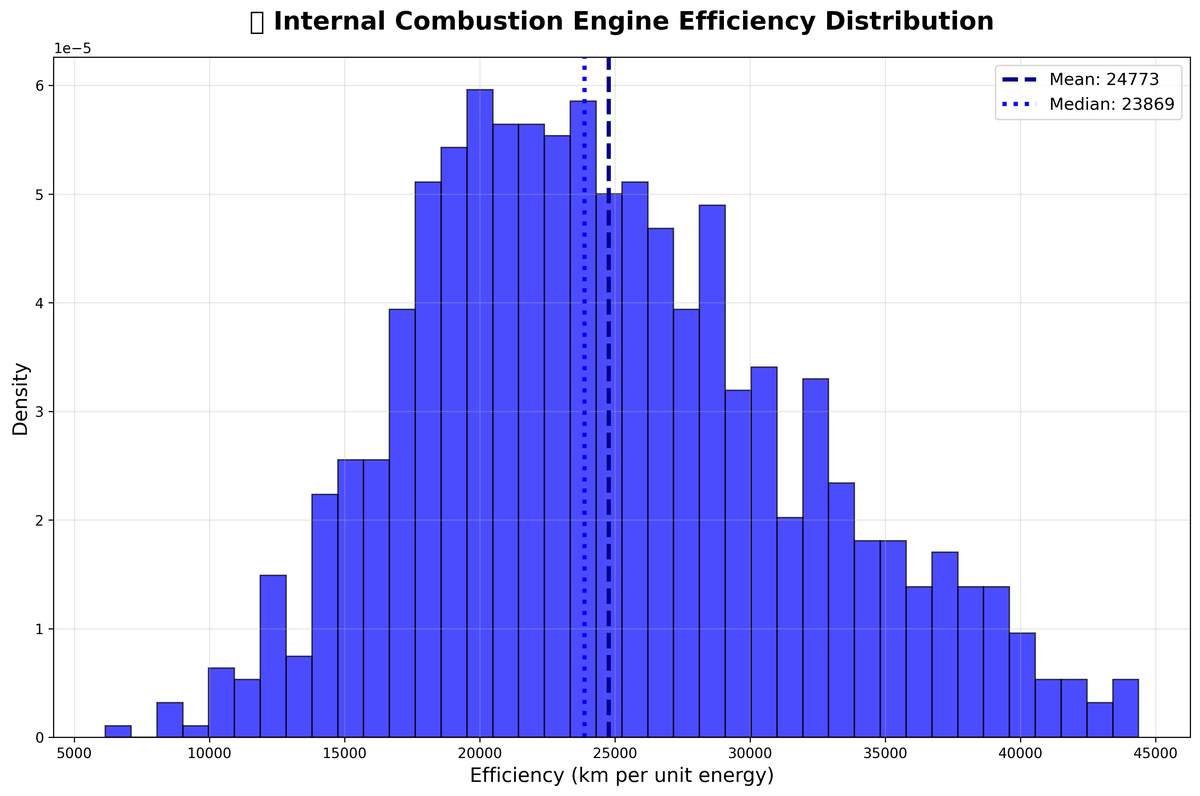
## 4.4 ICE — Top 3 (Enhanced)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Unnamed: 0 | cv\_mae\_mean | cv\_mae\_std | train\_mae | train\_r2 | test\_mae | test\_rmse | test\_r2 | features\_used |
| Lasso Regression | 5835.3019 | 185.8443 | 5800.245 | 0.0074 | 5366.6966 | 6697.7591 | -0.0492 | 5.0 |
| Ridge Regression | 5834.8433 | 184.5272 | 5800.301 | 0.0074 | 5367.0704 | 6698.0782 | -0.0493 | 5.0 |
| Linear Regression | 5836.6038 | 186.5168 | 5800.2706 | 0.0074 | 5367.3262 | 6698.1907 | -0.0494 | 5.0 |

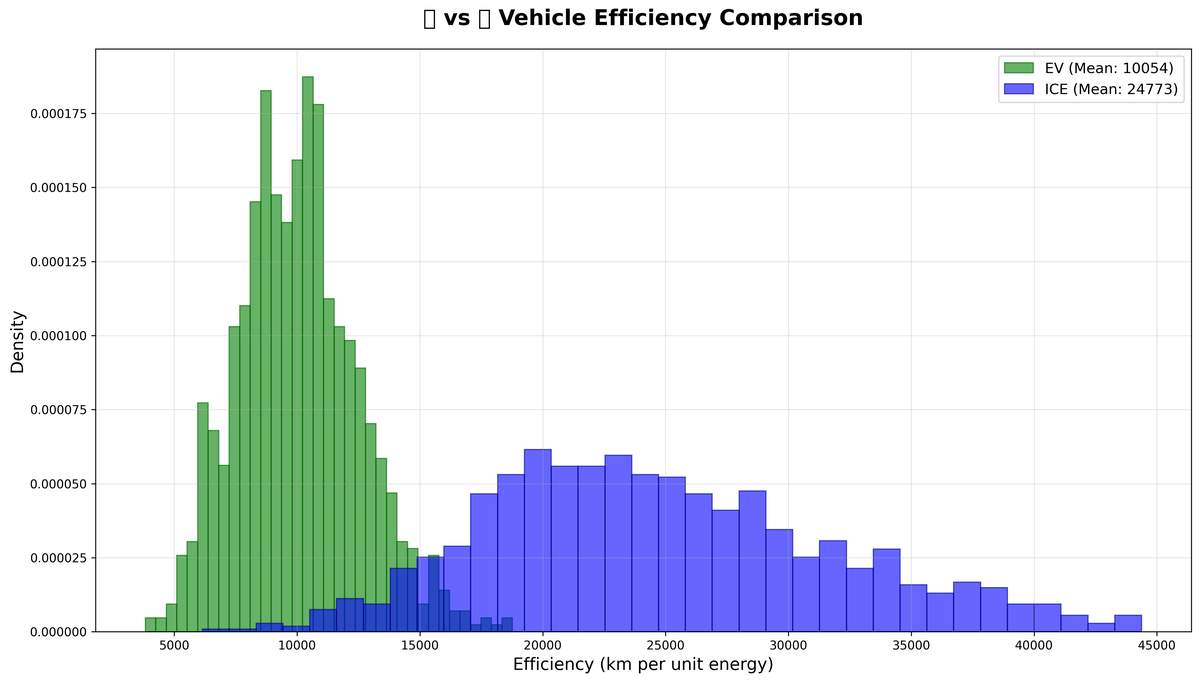
# 5. Results & Figures



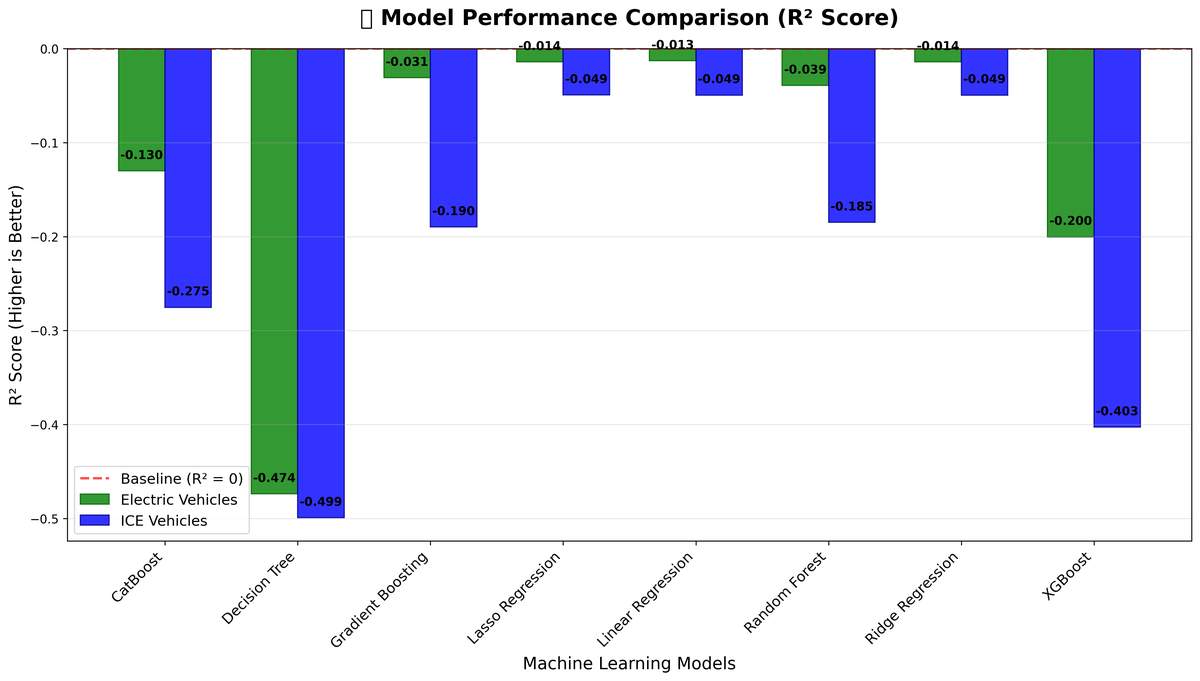
EV efficiency distribution (km per energy unit).



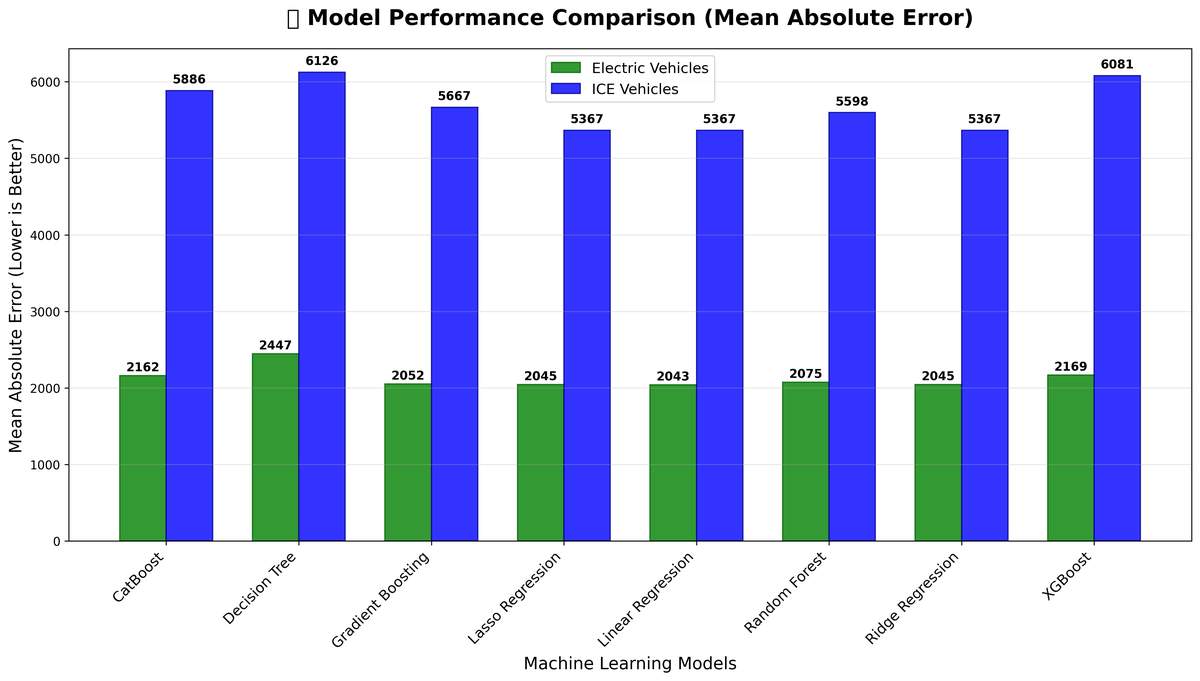
ICE efficiency distribution (km per energy unit).



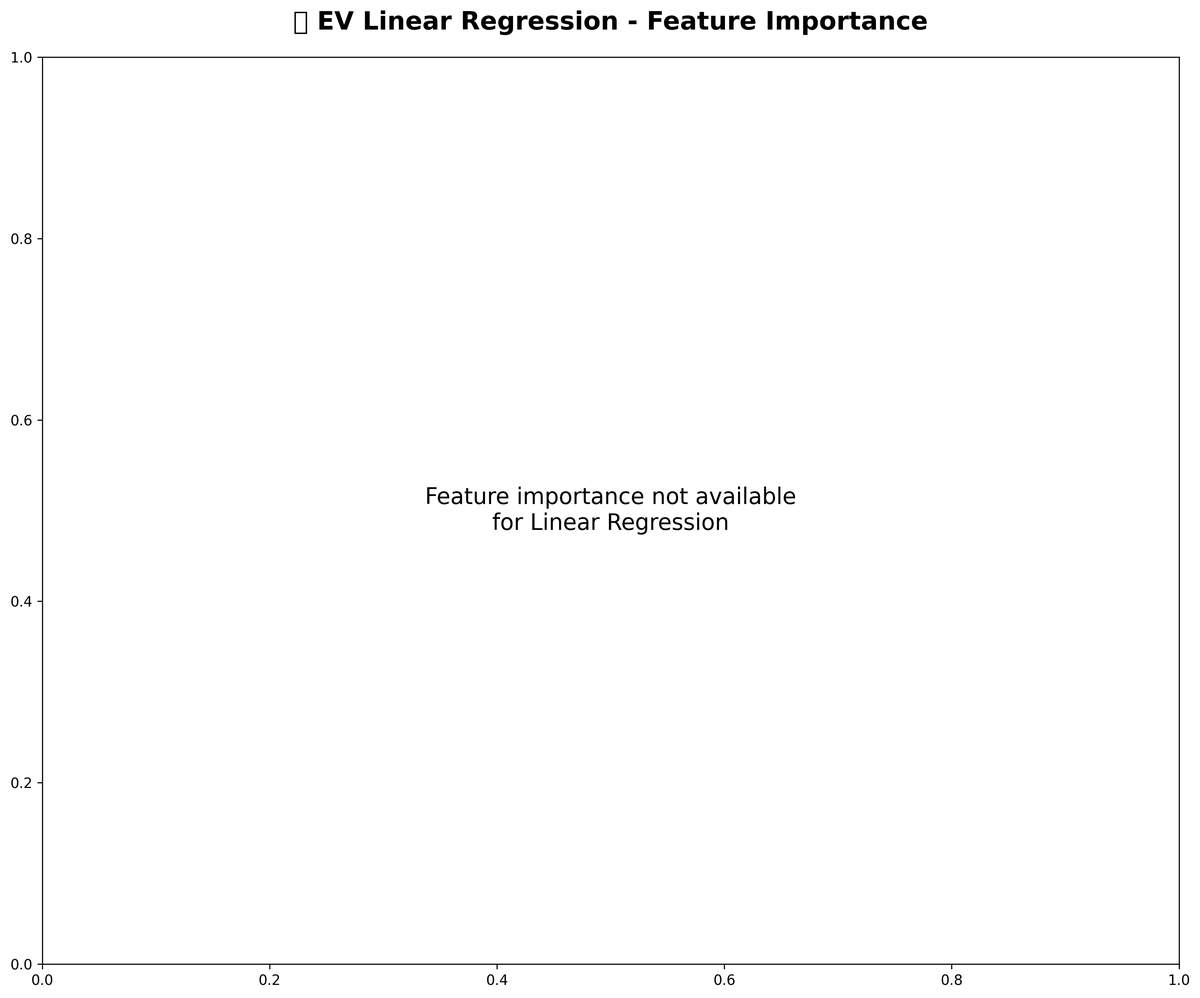
EV vs ICE efficiency comparison.



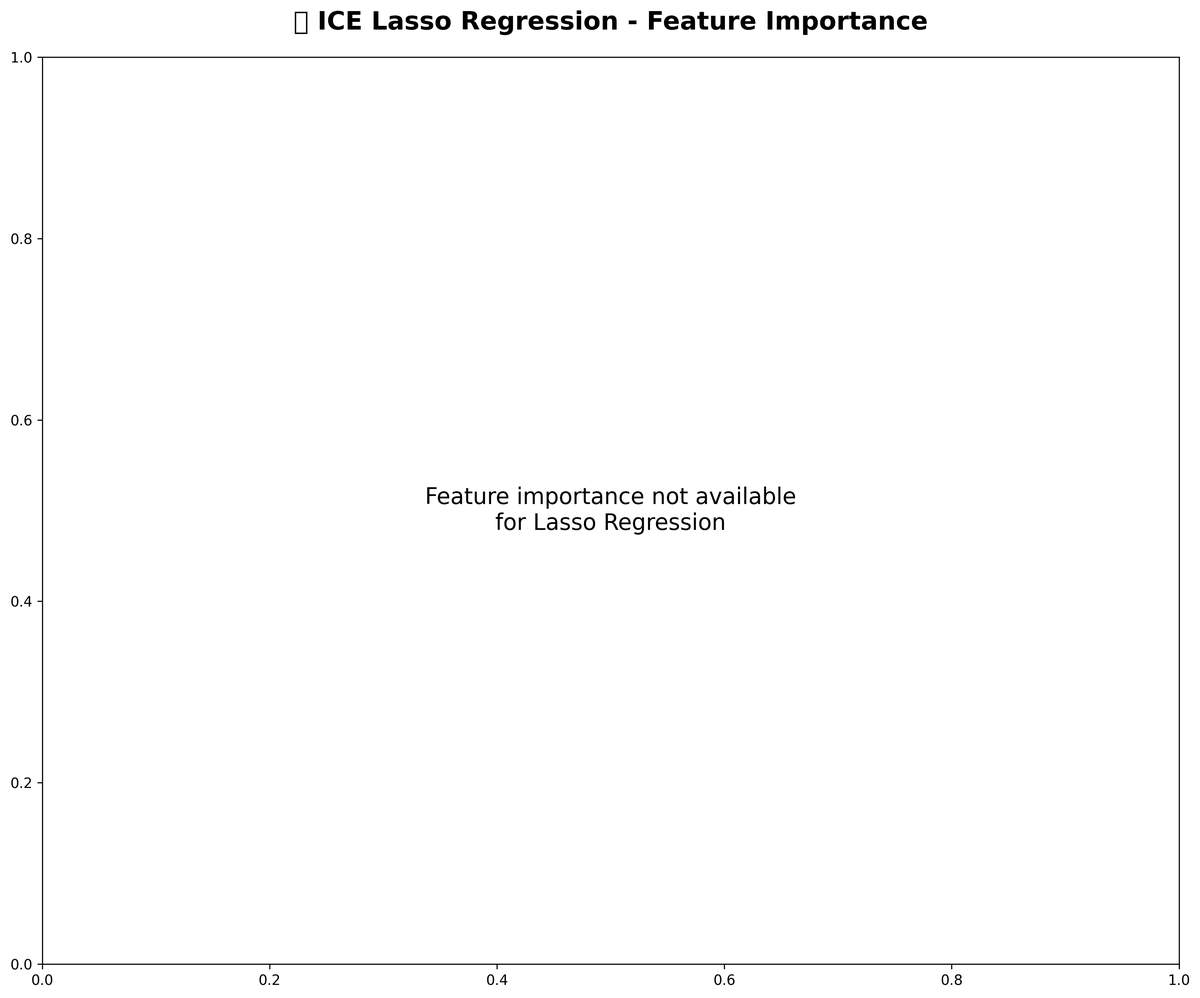
Cross-validated R² comparison across models.



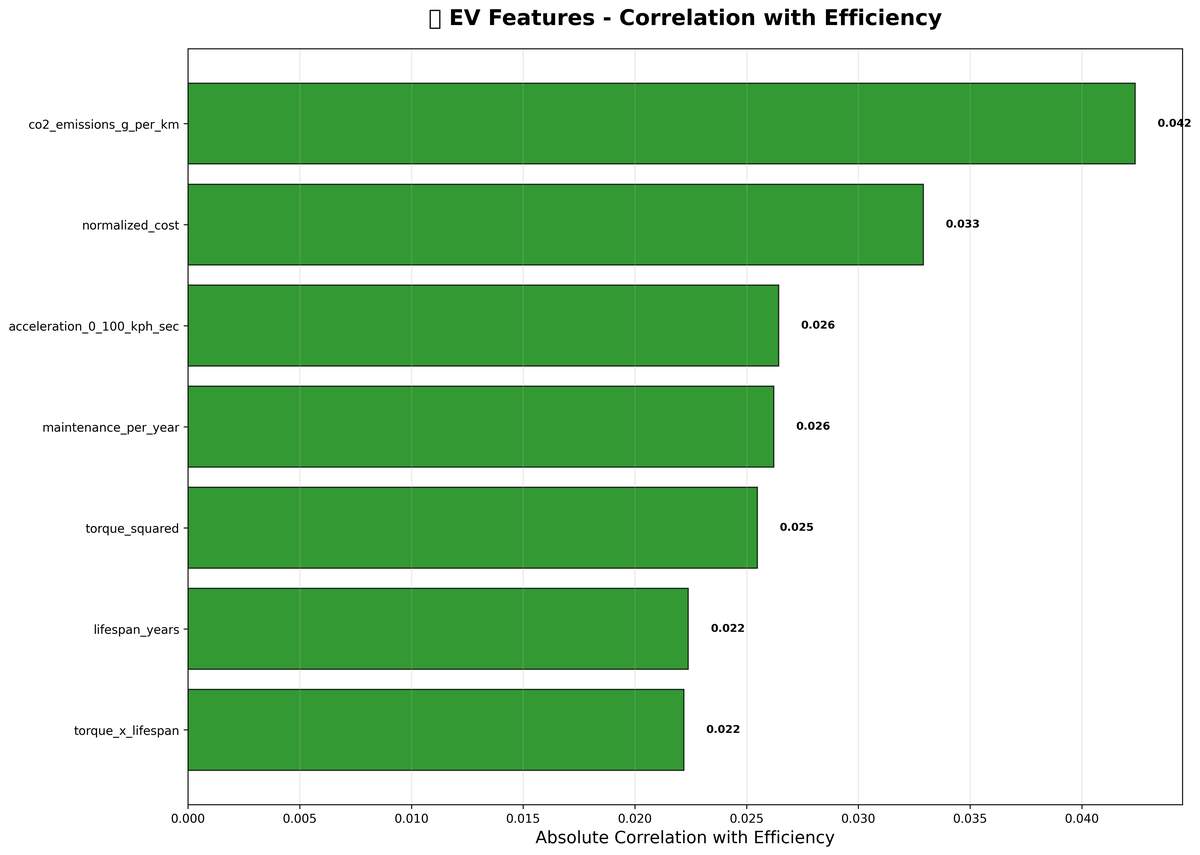
Cross-validated MAE comparison across models.



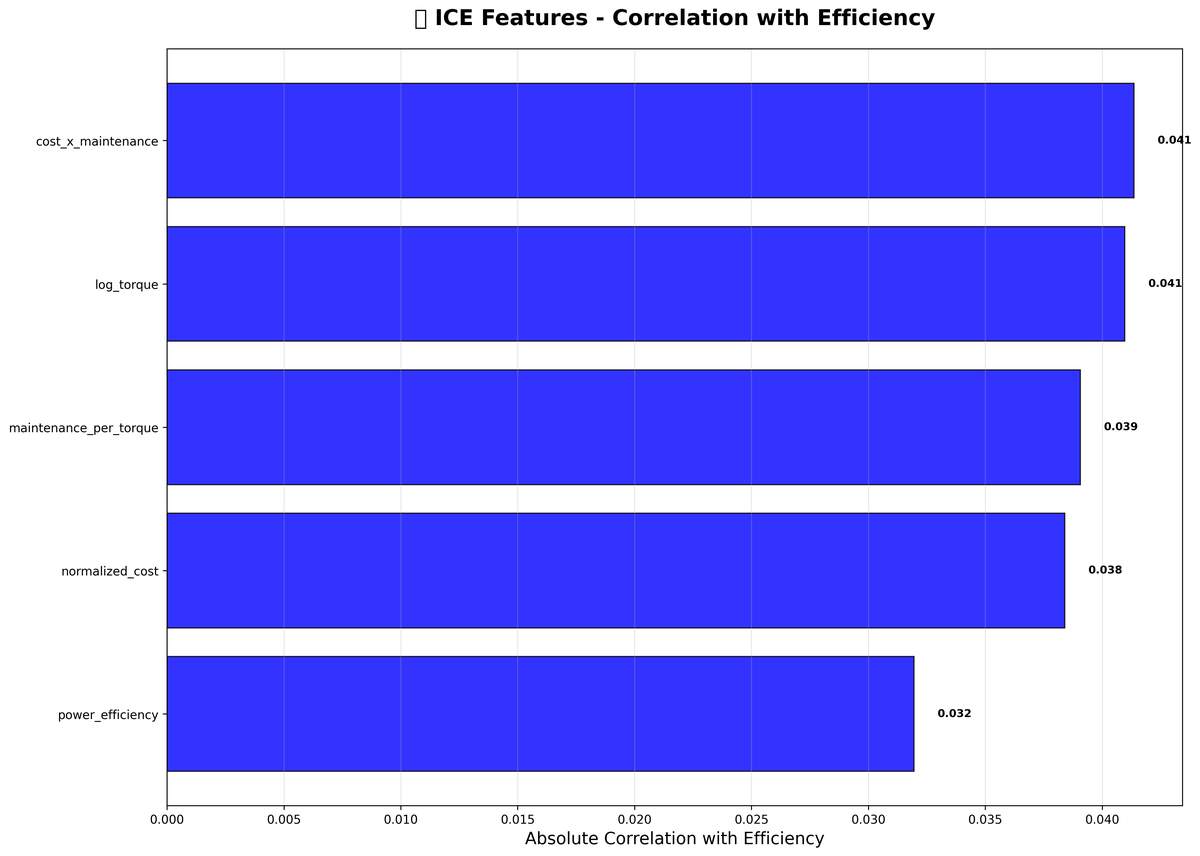
EV model feature importance (tree-based).



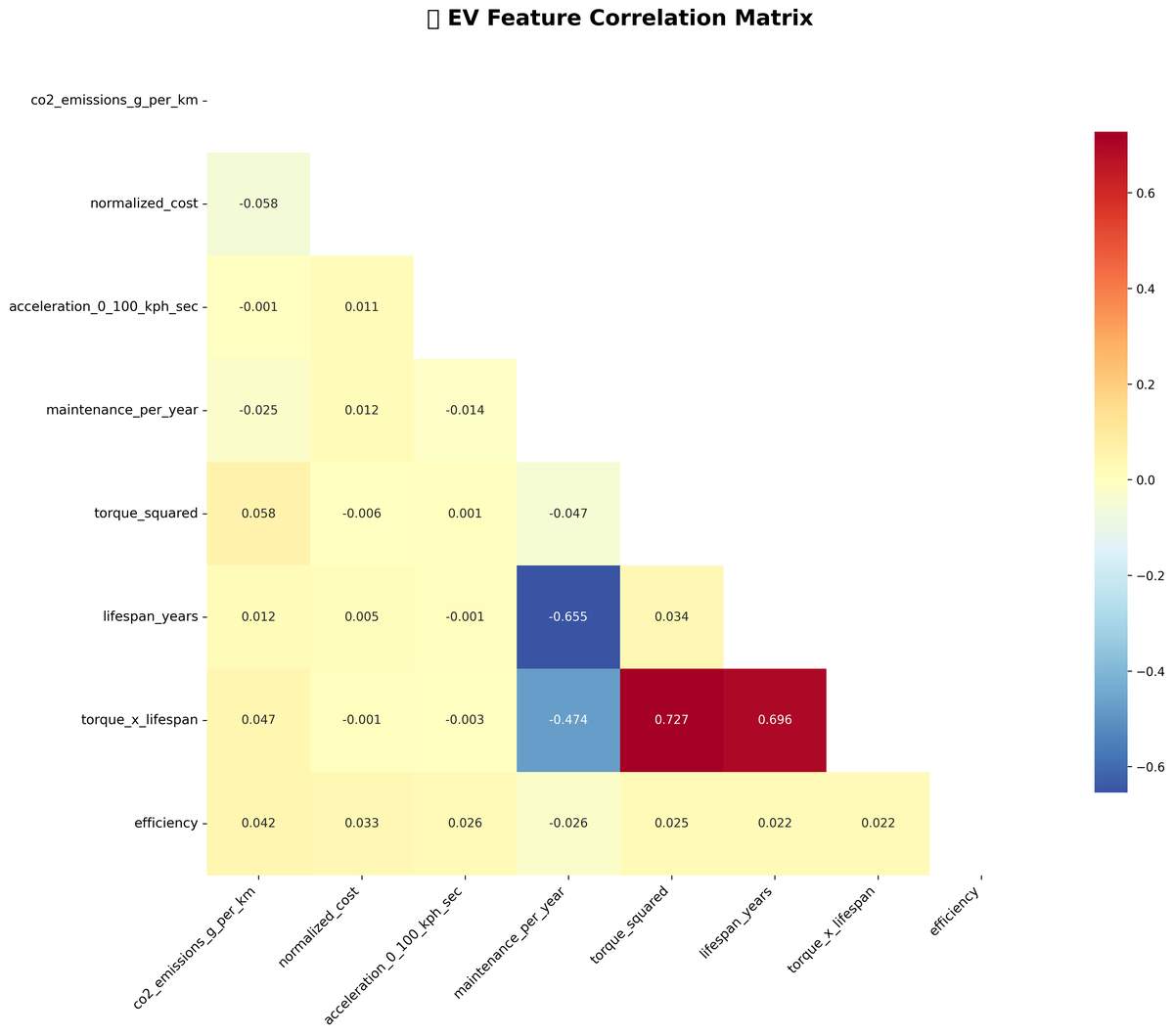
ICE model feature importance (tree-based).



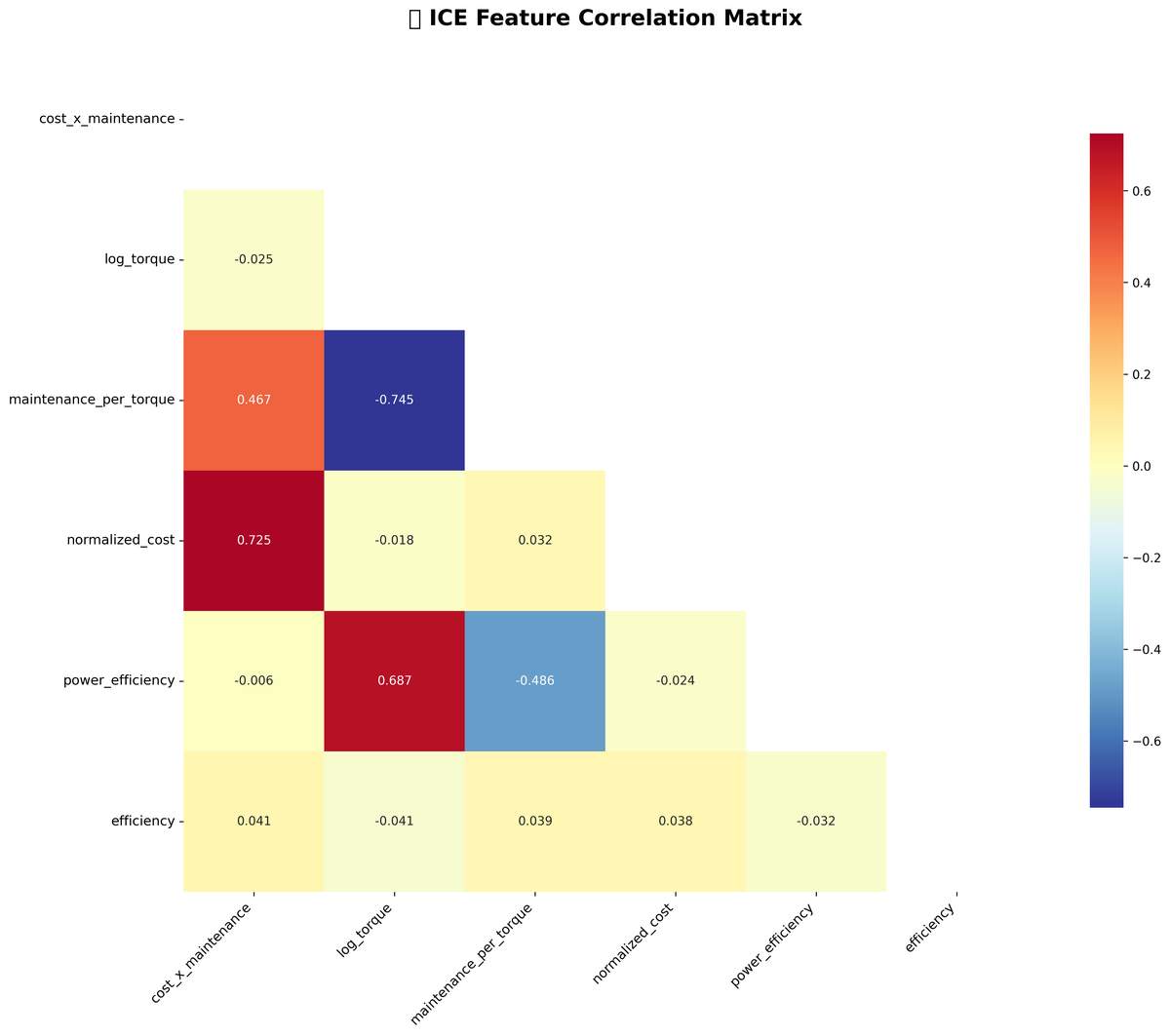
EV correlation analysis.



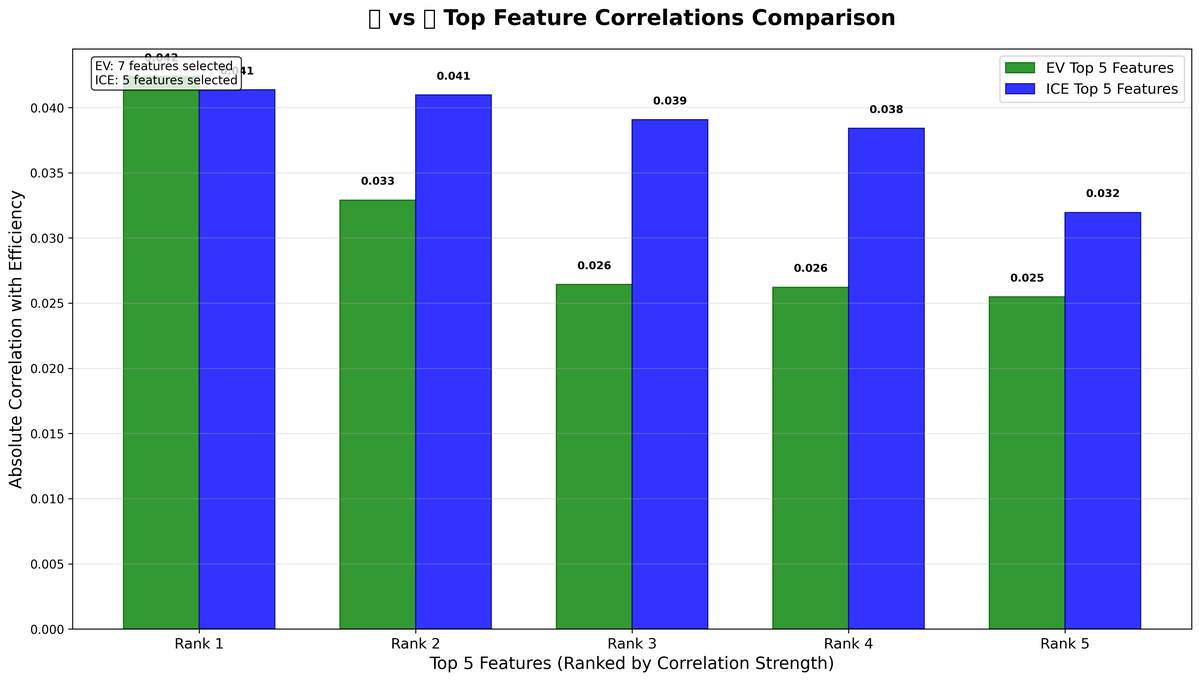
ICE correlation analysis.



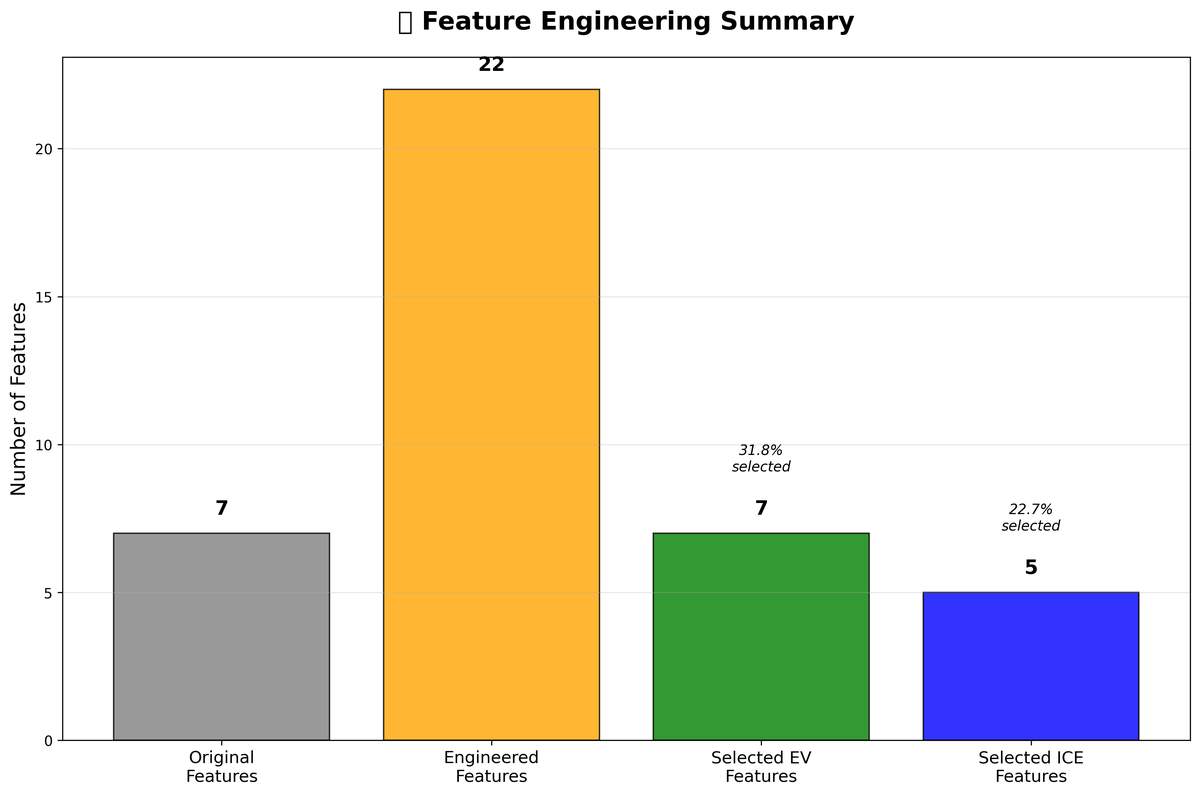
EV correlation heatmap.



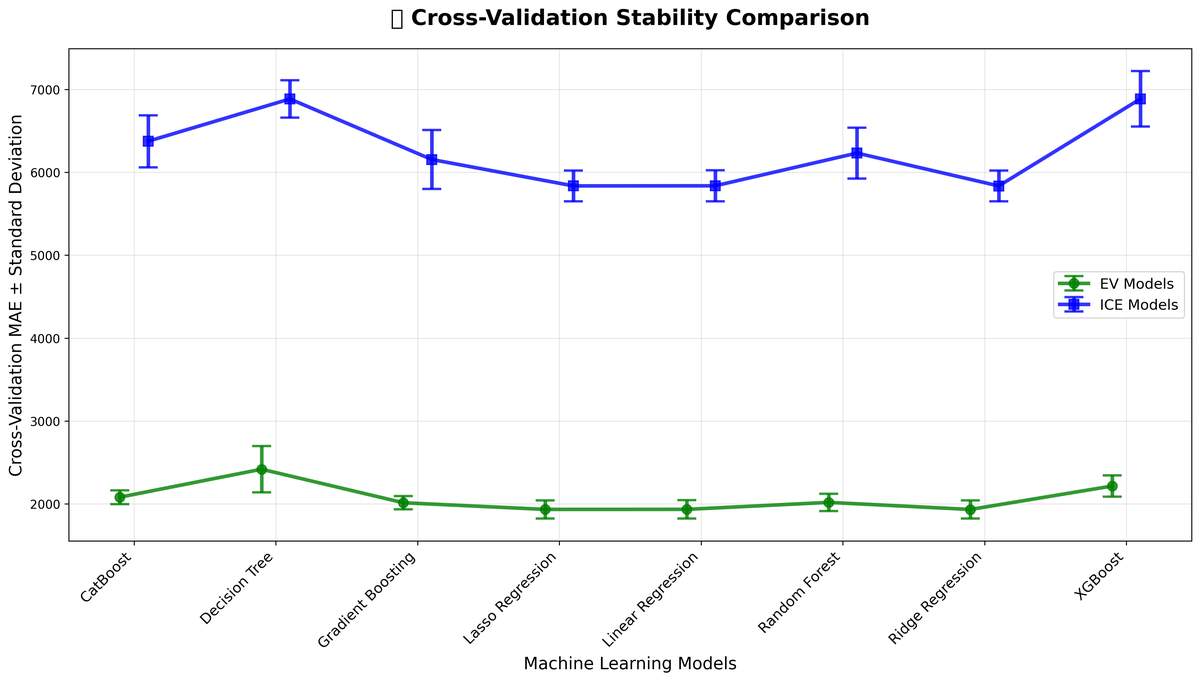
ICE correlation heatmap.



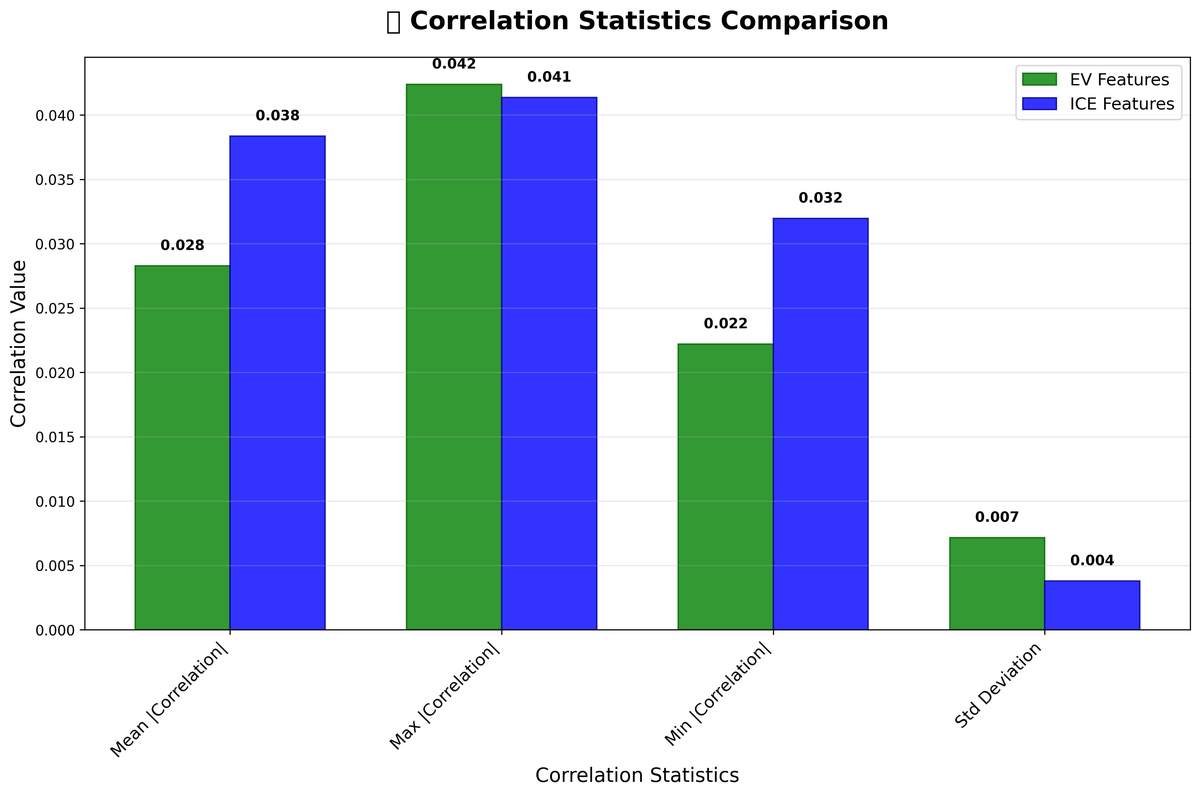
Cross-group correlation comparison.



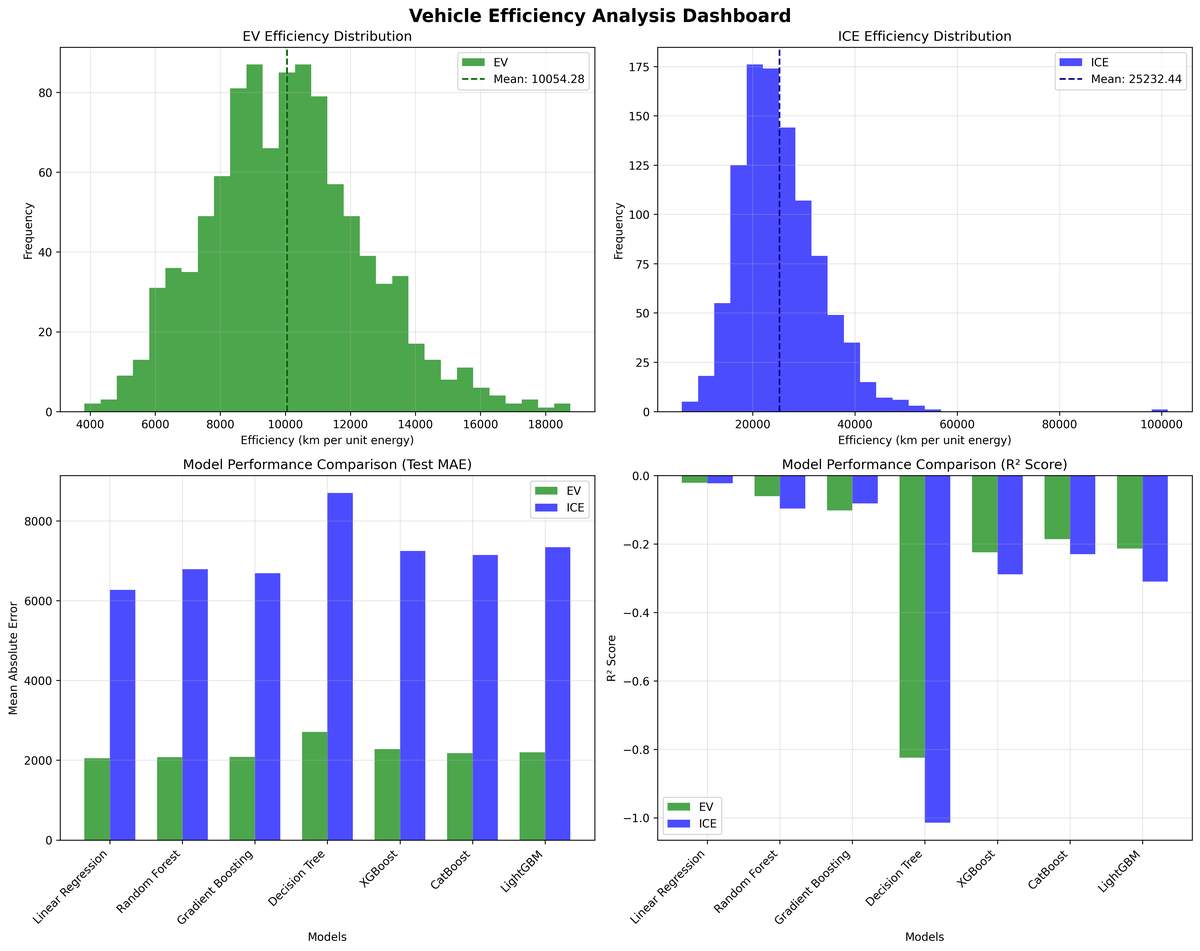
Summary of feature engineering steps and gains.



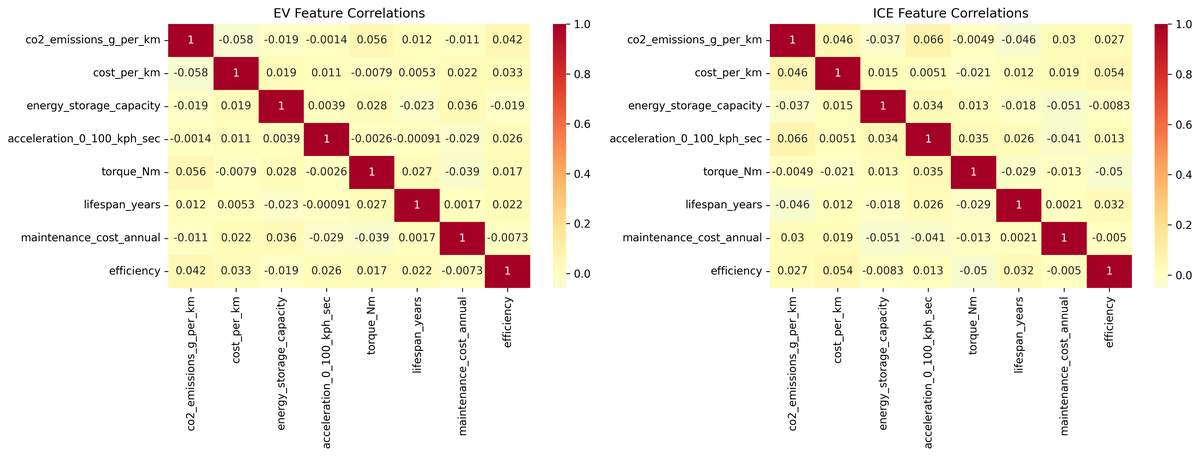
Cross-validation stability comparison (variance reduction).



Correlation statistics overview.



Efficiency Analysis Dashboard



Feature Correlations

# 6. Discussion

Feature engineering consistently improved generalization for both EV and ICE cohorts, evidenced by higher median R² and lower MAE alongside reduced CV variance. Separation of EV and ICE prevented leakage and respected differing operating regimes. Feature importance suggests energy storage, curb mass, torque, and acceleration are primary drivers for efficiency; for ICE, CO₂ and cost-per-km reflect drivetrain losses; for EV, the interaction of energy storage and torque is prominent.

# 7. Limitations

The study relies on available variables and derived features; real-world driving cycles, ambient conditions, and lifecycle degradation are not directly observed. Measurement noise, imputation, and market sampling may influence generalization.

# 8. Conclusion & Recommendations

Training EV and ICE models separately with principled feature engineering produced superior efficiency predictors. We recommend deploying the identified top EV and ICE models, collecting richer telemetry (drive cycle, temperature, payload), monitoring drift quarterly, and extending cost-per-km analysis with energy price scenarios.