Supporting Information

Section S1: Information about the dataset used in this work

Table S1: Information Contained in the Electric Vehicle (EV) Dataset

Feature Category	Feature
Model Specifications	Name
	Year
	Market (e.g., China, North America, Europe)
	Price
	Trim (if applicable)
	Alias (if applicable)
Battery Pack	Voltage
	Number of Battery Modules
	Weight
	Volume
	Cooling System/Mechanism (air convection, water-based cooling, and active thermal management) *
	Battery Type as Advertised (lithium-ion, lithium polymer)
	Energy Density
Performance Specifications	Top Speed
	Acceleration
	Range
	Curb Weight
	Payload
	Gross Vehicle Weight Rating (GVWR)

* It is important to note that, while air- and water-based cooling are considered active cooling mechanisms, the dataset source might have classified the cooling mechanism as “active thermal management” in cases of insufficient or unclear information.

Section S2: The steps for linear regression parameter estimation using bootstrapping

1. We define $X = [x_1,\dots,x_n]^T$ and $y = [y_1,\dots,y_n]^T$ as the vector of regressor and response variables, respectively, where n is the total number of observations. In our case, x is the vector of EV range observations according to the WLTP (or NEDC), and y is the vector of EV range observations according to the EPA. For the observations where both regressor and response variables are present (where the corresponding *(k x 1)* vectors are denoted by X_{comp} and y_{comp}), the linear regression parameter (θ) in $y_{comp} = X_{comp}\,\theta + e$ is calculated using:

$$\theta = \left(X_{comp}^{T} X_{comp}\right)^{-1} X_{comp} y_{comp} \qquad \text{S1}$$

2. The values of y_{comp} were predicted using the relationship $y_{comp}=X_{comp}\theta$ and are denoted by $\hat{y}_{comp}$. Subsequently, the residual, $e = y_{com} - y_{comp}$, was computed for *k* observations. It should be noted that there are *(n-k)* observations, where the response variable (EPA range) is missing. Each of the observations with a missing response variable was assigned a random calculated residual with a replacement.

3. From the assigned residual and the predicted value, the actual values of the response variables of the missing data were calculated. The bootstrapped linear regression parameter (θ_b) was calculated for all the observations using:

$$\theta_b = (X^T X)^{-1} X y \qquad \text{S2}$$

4. Steps 2-3 were repeated 2000 times, and the average bootstrapped linear regression parameters was calculated for each iteration.

Section S3: Relationship between ranges

$$EPA\ range = -34.721 + 1.049 * WLTP\ range \quad \text{S3}$$

$$EPA\ range = -38.9802 + 0.9703 * NEDC\ range \quad \text{S4}$$

Note: All ranges are in miles.

Section S4: Plots outlining trends between different variables

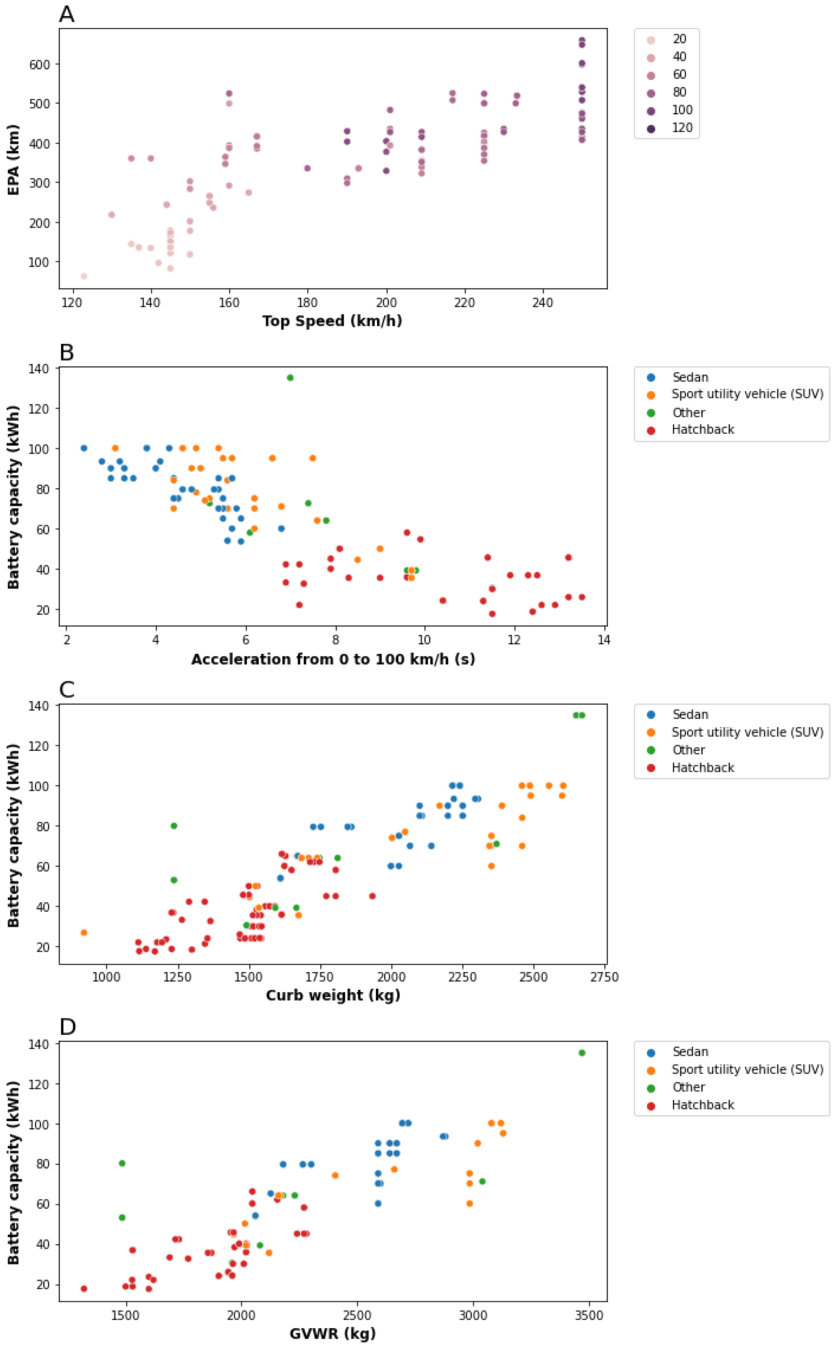

Figure S1: Scatter plots exploring different relations in the dataset. (A) top speed against its EPA range with different battery capacities, (B) acceleration against battery capacity with different body styles, (C) curb weight against battery capacity with different body styles, and (D) GVWR against battery capacity with different body styles.

Section S5: Relationship between the EPA range and the independent variables

$$
\begin{aligned}
EPA\ [miles] = {} & -37562 + 18.6 * Model\ Year + 3.8 \\
& * \ Battery\ capacity\ [kWh] + 0.08 \\
& * \ Top\ Speed\ [mph] + 10.2 \\
& * Acceleration\ (0\ to\ 100\ kmh^{-1})[s] + 0.22 \\
& * Curb\ Weight\ [lb] - 0.19 * GVWR\ [lb] + \mathrm{a} \\
& * \mathrm{Battery\ Type} + \mathrm{b} * \mathrm{Cooling\ Mehanism} + \ \mathrm{c} \\
& * \mathrm{Body\ Style}
\end{aligned}
\tag{S5}
$$

where a is equal to 12.6 when lithium polymer batteries are used and is equal to -12.6 when lithium-ion batteries are used. Similarly, b = 28.9 or -28.p when water-based cooling or another cooling mechanism are used, respectively. Moreover, c = -5.1, -60.2, 51.8, or 13.5 when the body styles hatchback, other, sedan, or SUV are considered, respectively.

Section S6: Model performance as calculated by various performance metrics

The following equations express the absolute mean error (MAE), mean absolute percent error (MAPE), and R-squared error (R2) used to evaluate the values in Table S2.

$$
MAE = \frac{1}{m}\sum_{i=1}^{m} |\hat{y}_i - y_i| \tag{S6}
$$

$$
MAPE = \frac{1}{m}\sum_{i=1}^{m} \left|\frac{\hat{y}_i - y_i}{\hat{y}_i}\right| \tag{S7}
$$

$$
R^2 = 1 - \frac{\sum_{i=1}^{m} \hat{y}_i - y_i}{\sum_{i=1}^{m} \bar{y} - y_i} \tag{S8}
$$

where $\hat{y}_i$ and y_i denote the model predicted and actual values, respectively, and m is the total number of observations. Moreover, $\bar{y}$, used in Equation S8, refers to the mean values of the actual values.

Table S2: Model evaluations from MAE, MAPE, and R^2

Model	**MAE**	**MAPE**	**R^2**
Normal Linear Regression	23.379	0.067	0.932
Ridge Regression	15.542	0.072	0.930
Lasso Regression	23.570	0.067	0.932
Elastic Net	29.674	0.088	0.913
SVM Regression	22.367	0.062	0.939