Project Analysis:

The following document contains the results after a simple linear regression performance on each iPhone feature for CO2E prediction. Before cross validation, RMSEs for each feature are calculated and sorted in **ascending** order like so:

Feature	RMSE
Storage (gb)	10.380059805094406
Weight	13.39788601517441
Number of Cameras (back)	13.4206271312757
Display (inches)	15.579679667283553
Battery (mAh)	15.816080939176855
Front Camera (mp)	16.064142543464158
Sum of megapixels	17.7954501391763
Memory (gb)	18.02367338211221
GPU (cores)	18.081814242108614
CPU (cores)	18.8057573550041
Main Camera (mp)	18.815161919243145
Neural Engine (cores)	18.84035137830503

From this table, we can observe that the top 5 important features are the following with the lowest RMSE values:

1. Storage
2. Weight
3. Number of Cameras
4. Display
5. Batter

We perform a 5-fold cross validation on these features and get average RMSE values for each fold:

Feature	**RMSE** (after cross-validation)
Storage (gb)	9.148848335471216
Number of Cameras (back)	12.893443971052417
Weight (gm)	14.690773550211768
Display (inch)	16.289728782537217
Battery (mAh)	16.935820895193373

We perform multiple linear regression on **all** features and get an average RMSE value of 3.111501504256903.

After 5-fold cross validation the average RMSE value is 12.436819812132.

Summary:

The analysis reveals that before cross-validation, Storage (gb) had the lowest RMSE at 10.38, indicating high predictive accuracy for CO2E. Post 5-fold cross-validation, Storage maintained its position as the most significant predictor with an RMSE of 9.15. Multiple linear regression across all features initially showed that the error rate was very low at 3.11. But after checking more carefully with cross-validation, the error rate went up to 12.44.

Github Repo:

https://github.com/naseebafaiza/Linear-Regression-iPhones