# Model Performance Report

## Linear Regression Model Metrics:

* Mean Squared Error (MSE) on Testing Data: 3.174
* R-Squared (R2) on Testing Data: 0.899
* Mean Squared Error (MSE) using Cross-Validation: 2.943
* Average R-Squared (R2) using Cross-Validation: 0.859

## Polynomial Regression Model Metrics:

* Mean Squared Error (MSE) on Testing Data: 0.413
* R-Squared (R2) on Testing Data: 0.987
* Mean Squared Error (MSE) using Cross-Validation: 0.488
* Average R-Squared (R2) using Cross-Validation: 0.977

## Ridge Regression Model Metrics:

* Mean Squared Error (MSE) on Testing Data: 3.194
* R-Squared (R2) on Testing Data: 0.899
* Mean Squared Error (MSE) using Cross-Validation: 2.946
* Average R-Squared (R2) using Cross-Validation: 0.860

## Random Forest Regressor Model Metrics:

* Mean Squared Error (MSE) on Testing Data: 0.587
* R-Squared (R2) on Testing Data: 0.981
* Mean Squared Error (MSE) using Cross-Validation: 0.683
* Average R-Squared (R2) using Cross-Validation: 0.969

In this report, the performance of each model is assessed using various evaluation metrics. These metrics provide insights into how well the models predict product sales based on advertising budgets. The metrics include Mean Squared Error (MSE), R-Squared (R2), Mean Squared Error using Cross-Validation (MSE CV), and Average R-Squared using Cross-Validation (Avg R2 CV).

The **Polynomial Regression model** demonstrates superior performance across multiple metrics, with the lowest MSE on testing data, the highest R2 on testing data, and consistently strong performance in cross-validation. Its ability to capture both linear and non-linear relationships contribute to its excellent predictive capabilities.

The **Random Forest Regressor model** also performs well, with competitive metrics in terms of both accuracy and generalization. Its ensemble nature allows it to capture complex interactions in the data.

While the Linear Regression and Ridge Regression models provide reasonable performance, the Polynomial Regression and Random Forest Regressor models stand out as top performers in this context.

# Model Interpretation

## Polynomial Regression Model Coefficients Interpretation:

The Polynomial Regression model includes both linear and non-linear terms, capturing complex relationships between advertising budgets and product sales.

* **Coefficient for TV (3.598):** For every unit increase in the TV advertising budget, the predicted sales are expected to increase by approximately 3.598 units. This suggests that TV advertising has a strong positive influence on product sales.
* **Coefficient for Radio (2.857):** A unit increase in the radio advertising budget is associated with an approximate increase of 2.857 units in predicted sales. Radio advertising also has a significant positive impact on sales.
* **Coefficient for Newspaper (0.082):** An increase in the newspaper advertising budget by one unit is associated with a relatively smaller increase of 0.082 units in predicted sales. Newspaper advertising has a weaker positive impact on sales compared to TV and radio.
* **Interaction Terms:** The interaction terms (e.g., TV \* Radio, TV \* Newspaper) reveal combined effects of advertising media. For instance, the positive coefficient for TV \* Radio interaction indicates that TV and radio advertising together can have an amplifying effect on sales.
* **Squared Terms:** The positive coefficients for squared terms (e.g., TV^2, Radio^2) suggest that the relationships between advertising budgets and sales may exhibit some degree of non-linearity. This means that the impact of increasing budgets may not be constant, and it might change based on the existing budget levels.

## Random Forest Regressor Feature Importances Interpretation:

Random Forest Regressor models don't have traditional coefficients, but feature importances reveal the relative influence of each feature on predictions.

* **TV Feature Importance (0.622):** TV advertising has the highest feature importance, indicating that changes in the TV advertising budget strongly influence the model's predictions. It plays a pivotal role in determining product sales.
* **Radio Feature Importance (0.365):** Radio advertising also holds significant importance, suggesting that variations in radio advertising budgets have a substantial impact on sales predictions.
* **Newspaper Feature Importance (0.012):** Newspaper advertising has the lowest importance, implying that changes in the newspaper advertising budget have a relatively smaller impact on sales predictions compared to TV and radio.

## Linear Regression Model Coefficients Interpretation:

The Linear Regression model estimates the linear relationships between advertising budgets and product sales.

* **Coefficient for TV (3.764):** A unit increase in the TV advertising budget is associated with an approximate increase of 3.764 units in predicted sales. TV advertising has a strong positive impact on sales predictions.
* **Coefficient for Radio (2.792):** An increase in the radio advertising budget by one unit is expected to lead to an approximate increase of 2.792 units in predicted sales. Radio advertising also positively influences sales.
* **Coefficient for Newspaper (0.056):** A unit increase in the newspaper advertising budget corresponds to a smaller increase of 0.056 units in predicted sales. Newspaper advertising has the weakest positive impact on sales compared to TV and radio.

In summary, each model's coefficients or feature importances offer insights into how different advertising media contribute to predicting product sales. The Polynomial Regression model captures complex relationships and non-linearities, while the Random Forest Regressor model considers interactions and non-linear effects. The Linear Regression model provides linear insights. Interpretation helps businesses make informed decisions on how to allocate advertising budgets effectively for maximizing product sales.