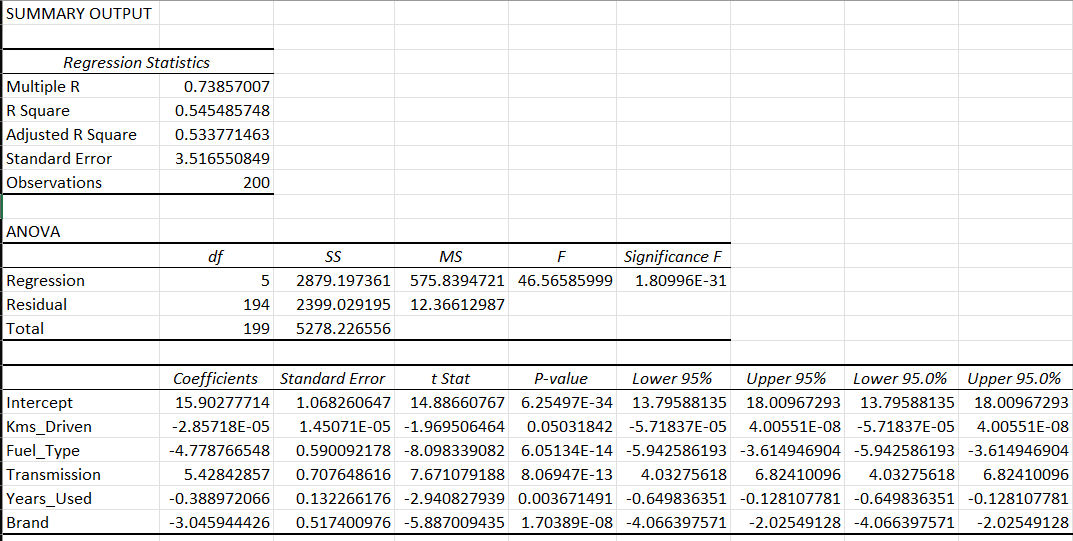
**CaseProject3**

**From the CaseProject3.csv**

**Model1\_Summary after performing regression analysis:**

****

**Interpretation of the Regression Output:**

* The R Square (R²) value is 0.5455, meaning 54.55% of the variation in the target variable is explained by the independent variables in the model.
* The Adjusted R Square value is 0.5338, which is more reliable because it accounts for the number of predictors and reduces the effect of overfitting.
* The Standard Error is 3.5166, representing the average prediction error. This means predictions from the model deviate from the actual values by about 3.52 units on average.

**Individual Predictor Analysis (Based on P-values):**

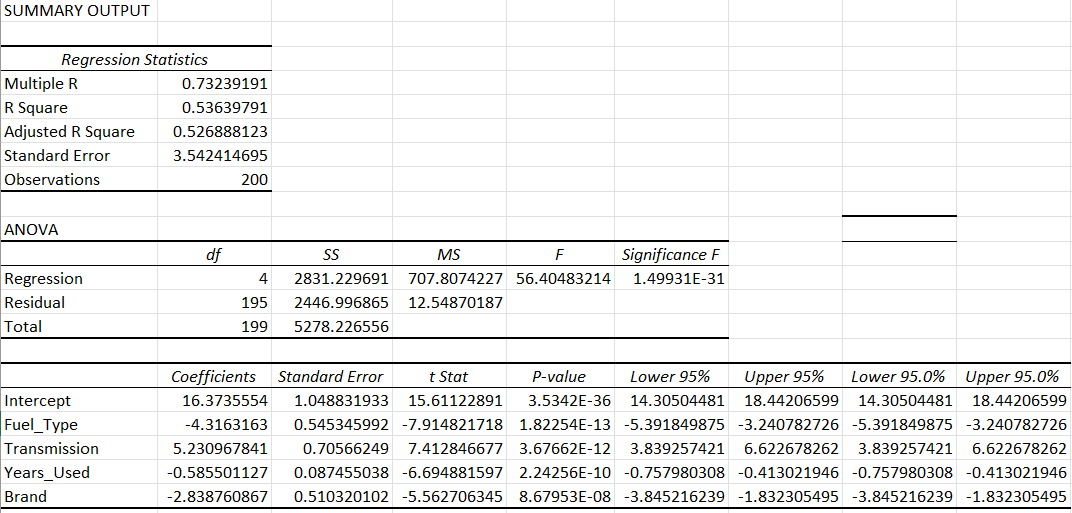
| Variable | Coefficient | P-value | Interpretation |
| --- | --- | --- | --- |
| Kms\_Driven | -2.85718E-05 | 0.0503 | Just on the edge of significance (slightly above 0.05). More kilometers driven reduces the predicted value slightly. |
| Fuel\_Type | -4.7788 | 6.0513E-14 | Highly significant. Changing the fuel type results in a decrease of ~4.78 units. |
| Transmission | 5.4248 | 6.8947E-13 | Highly significant. Switching transmission type leads to an increase of ~5.42 units. |
| Years\_Used | -0.38897 | 0.0037 | Statistically significant. Each additional year of use reduces value by ~0.39. |
| Brand | -3.0459 | 1.7039E-08 | Highly significant. Different brand reduces predicted value by ~3.05 units. |

All predictors except Kms\_Driven have p-values < 0.05, meaning they are statistically significant.  
Kms\_Driven is borderline (p ≈ 0.0503), so its significance is weak and may be considered marginal depending on context.

**Conclusion:**

The regression model is statistically significant overall (Significance F = 1.80996E-31) and explains 54.55% of the variance in the dependent variable.  
The Adjusted R² = 0.5338 reflects a reliable model with moderate explanatory power.  
Fuel type, transmission, years used, and brand are strong, significant predictors, while Kms driven is only marginally significant.

**Model2\_Summary after excluding the Kms-Driven from the sheet:**

****

**Interpretation of the Regression Output:**

* The R Square (R²) value is 0.5364, meaning 53.64% of the variation in the target variable (e.g., car price) is explained by the independent variables used in the model.
* The Adjusted R Square value is 0.5269, which is more reliable than R² because it adjusts for the number of predictors and penalizes the inclusion of irrelevant variables.
* The Standard Error is 3.5424, representing the average prediction error. On average, the predicted values deviate from the actual values by about 3.54 units.

**Individual Predictor Analysis (Based on P-values):**

| Variable | Coefficient | P-value | Interpretation |
| --- | --- | --- | --- |
| Fuel\_Type | -4.3163 | 1.8225E-13 | Highly significant. A change in fuel type is associated with a decrease of ~4.32 in the predicted value. |
| Transmission | 5.2309 | 3.6766E-12 | Highly significant. Changing transmission type leads to an increase of ~5.23 in the predicted value. |
| Years\_Used | -0.5855 | 2.2426E-10 | Highly significant. Each additional year reduces the value by ~0.59. |
| Brand | -2.8388 | 8.6795E-08 | Highly significant. Different brands reduce the predicted value by ~2.84 on average. |

All predictors have p-values less than 0.05, indicating they are statistically significant contributors to the model.

**Conclusion:**

The regression model is statistically significant overall (Significance F = 1.49931E-31) and has a moderate predictive power with R² ≈ 0.536.  
The Adjusted R² = 0.5269 confirms the reliability of the model after adjusting for predictor count.  
All included variables are meaningful and contribute significantly to the prediction.

**Comparison of both models:**

1. Overall Model Performance

| Metric | Model 1 | Model 2 | Which is Better? |
| --- | --- | --- | --- |
| R Square (R²) | 0.5455 (54.55%) | 0.5364 (53.64%) | Model 1 (Slightly higher) |
| Adjusted R Square | 0.5338 (53.38%) | 0.5269 (52.69%) | Model 1 (More reliable) |
| Standard Error | 3.5166 | 3.5424 | Model 1 (Lower error) |
| Significance F | 1.80996E-31 | 1.49931E-31 | Both are highly significant |

**Conclusion:** Model 1 has slightly better explanatory power and prediction accuracy.

**Predictor Significance:**

| Predictor | Model 1 | Model 2 |
| --- | --- | --- |
| Kms\_Driven | -2.86E-05, p = 0.0503 (marginal) | ❌ Not included |
| Fuel\_Type | -4.78, p ≈ 6.05E-14 (very significant) | -4.32, p ≈ 1.82E-13 (very significant) |
| Transmission | 5.42, p ≈ 6.89E-13 (very significant) | 5.23, p ≈ 3.68E-12 (very significant) |
| Years\_Used | -0.39, p ≈ 0.0037 (significant) | -0.59, p ≈ 2.24E-10 (very significant) |
| Brand | -3.05, p ≈ 1.70E-08 (very significant) | -2.84, p ≈ 8.67E-08 (very significant) |

**Conclusion:**

* Both models have the same strong predictors: Fuel\_Type, Transmission, Years\_Used, Brand.
* Model 1 additionally includes Kms\_Driven, but it is only marginally significant.

**Final Verdict: Which Model is Better?**

| **Criteria** | **Winner** | | **Why?** | |
| --- | --- | --- | --- | --- |
| Explanatory Power | | Model 1 | | Higher R² and Adjusted R² |
| Prediction Accuracy | | Model 1 | | Lower standard error |
| Simplicity | | Model 2 | | Fewer variables (no Kms\_Driven) |
| Significant Predictors | | Tie | | Same strong variables |

**Model 1 is overall more reliable and performs slightly better, but Model 2 is simpler and still statistically strong.**

Why Insignificant Variables (p-value > 0.05) Were Removed?

"Insignificant variables (p-value > 0.05) were removed from the model because they do not contribute meaningfully to explaining the variation in the target variable. Keeping them would increase complexity without improving accuracy, and may even reduce the model’s reliability and interpretability."

-----------------