Homework 6
Simply & Multiple Linear Regression Analysis

Christopher Williams

October 30, 2025

Console Outputs. For every question that requires an output in the console of the program, the output will be seen in the program, but I will also include answers in this document for easier grading (I hope). Thank you for your hard work!

1 EDA

1.1 Shelve Location vs Sales Analysis

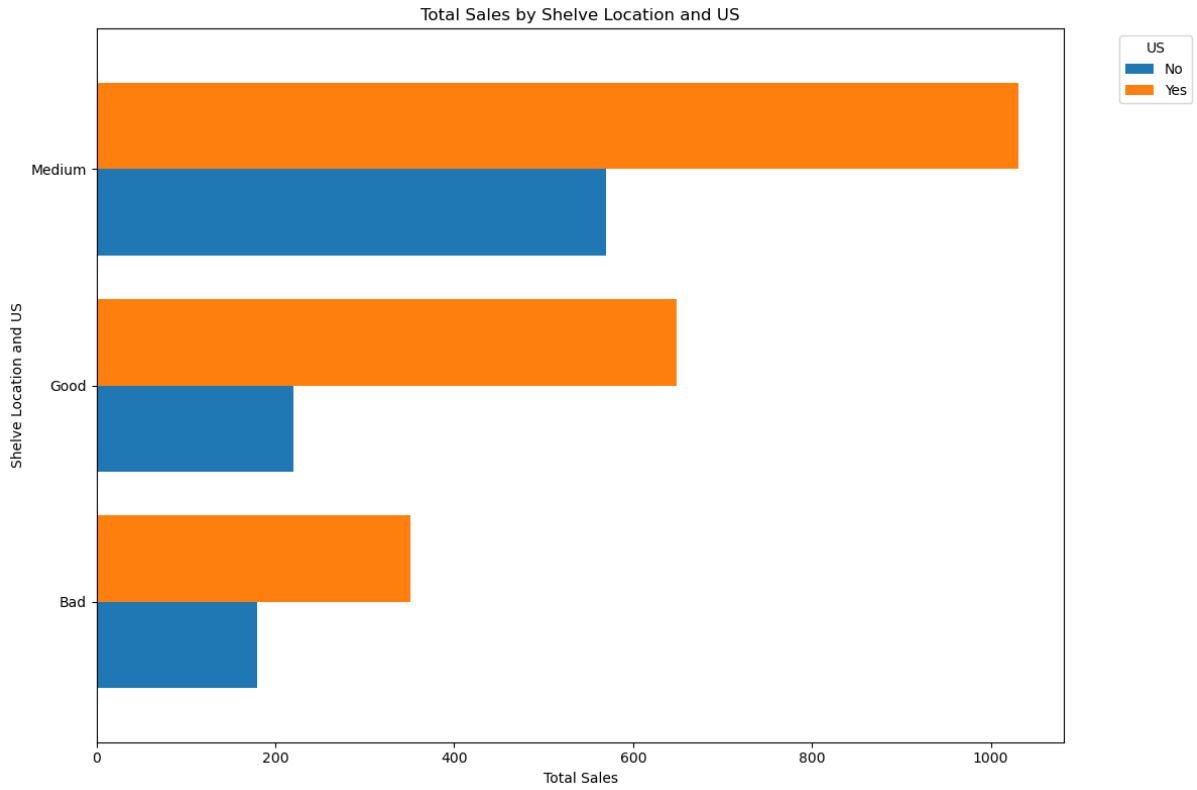

Figure 1: ShelveLoc vs Sales grouped by US location

1.2 One-Hot Encoding

The dataset was successfully encoded using one-hot encoding. The qualitative features (ShelveLoc, Urban, US) were converted to binary features while avoiding the dummy trap by dropping one category from each.

1.3 Train-Test Split and Standardization

The dataset was split with 80% training (320 samples) and 20% testing (80 samples) with shuffle=True and random_state=5805. The continuous features were standardized while encoded features were kept binary.

Then, for ease of use later in on in the file, I kept the **Sales scaling parameters** so that reverse transforms would be easier. **Sales scaling parameters:**

- Mean: 7.5120
- Scale (std): 2.8262

Feature Selection & Prediction

2 Backward Stepwise Regression

2.1 Elimination Process Table

The backward stepwise regression started with 11 features and eliminated 4 features based on p-value threshold of 0.01.

Table 1: Backward Stepwise Regression Elimination Process

Iteration	Features Count	Feature Eliminated	P-value	AIC	BIC	Adj. R^2
1	11	Population	0.962	-633.132	-587.913	0.867
2	10	Education	0.599	-635.130	-593.679	0.867
3	9	US_Yes	0.338	-636.843	-599.160	0.867
4	8	Urban_Yes	0.213	-637.893	-603.978	0.867
5	7	None	-	-638.298	-608.151	0.867

Eliminated Features (4):

1. Population (p-value: 0.962)
2. Education (p-value: 0.599)
3. US_Yes (p-value: 0.338)
4. Urban_Yes (p-value: 0.213)

Final Selected Features (7):

1. CompPrice
2. Income
3. Advertising
4. Price

5. Age
6. ShelveLoc_Good
7. ShelveLoc_Medium

2.2 OLS Regression Summary

Five OLS regression summaries are provided showing the elimination process:

```
ITERATION 1 - OLS SUMMARY (11 features)
                            OLS Regression Results
==============================================================================
Dep. Variable:                  Sales   R-squared:                       0.872
Model:                            OLS   Adj. R-squared:                  0.867
Method:                 Least Squares   F-statistic:                     190.3
Date:                Thu, 30 Oct 2025   Prob (F-statistic):          3.54e-130
Time:                        15:24:05   Log-Likelihood:                -125.49
No. Observations:                 320   AIC:                             275.0
Df Residuals:                     308   BIC:                             320.2
Df Model:                          11
Covariance Type:            nonrobust
================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
const               -0.7494      0.070    -10.697      0.000      -0.887      -0.612
CompPrice            0.5061      0.025     19.967      0.000       0.456       0.556
Income               0.1674      0.021      8.079      0.000       0.127       0.208
Advertising          0.2733      0.030      9.162      0.000       0.215       0.332
Population           0.0010      0.022      0.048      0.962      -0.042       0.044
Price               -0.7904      0.025    -31.204      0.000      -0.840      -0.741
Age                 -0.2720      0.021    -13.214      0.000      -0.312      -0.231
Education           -0.0107      0.021     -0.520      0.603      -0.051       0.030
ShelveLoc_Good       1.7526      0.061     28.758      0.000       1.633       1.873
ShelveLoc_Medium     0.7007      0.050     13.976      0.000       0.602       0.799
Urban_Yes            0.0560      0.046      1.226      0.221      -0.034       0.146
US_Yes              -0.0604      0.061     -0.985      0.325      -0.181       0.060
==============================================================================
Omnibus:                        0.705   Durbin-Watson:                   2.057
Prob(Omnibus):                  0.703   Jarque-Bera (JB):                0.804
Skew:                           0.103   Prob(JB):                        0.669
Kurtosis:                       2.866   Cond. No.                         7.28
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Iteration: 1
  Eliminating: Population (p-value: 0.962
  AIC: -633.132, BIC: -587.913, Adj_R_squared: 0.867
```

Figure 2: OLS Regression Summary - Iteration 1 (11 features)

```
ITERATION 2 - OLS SUMMARY (10 features)
                            OLS Regression Results
==============================================================================
Dep. Variable:                  Sales   R-squared:                       0.872
Model:                            OLS   Adj. R-squared:                  0.868
Method:                 Least Squares   F-statistic:                     210.0
Date:                Thu, 30 Oct 2025   Prob (F-statistic):          2.39e-131
Time:                        15:24:05   Log-Likelihood:                -125.50
No. Observations:                 320   AIC:                             273.0
Df Residuals:                     309   BIC:                             314.4
Df Model:                          10
Covariance Type:            nonrobust
================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
const               -0.7489      0.069    -10.840      0.000      -0.885      -0.613
CompPrice            0.5060      0.025     20.025      0.000       0.456       0.556
Income               0.1674      0.021      8.105      0.000       0.127       0.208
Advertising          0.2737      0.028      9.621      0.000       0.218       0.330
Price               -0.7904      0.025    -31.264      0.000      -0.840      -0.741
Age                 -0.2720      0.021    -13.241      0.000      -0.312      -0.232
Education           -0.0108      0.021     -0.527      0.599      -0.051       0.030
ShelveLoc_Good       1.7524      0.061     28.840      0.000       1.633       1.872
ShelveLoc_Medium     0.7006      0.050     14.031      0.000       0.602       0.799
Urban_Yes            0.0559      0.046      1.227      0.221      -0.034       0.145
US_Yes              -0.0608      0.060     -1.005      0.316      -0.180       0.058
==============================================================================
Omnibus:                        0.711   Durbin-Watson:                   2.056
Prob(Omnibus):                  0.701   Jarque-Bera (JB):                0.810
Skew:                           0.104   Prob(JB):                        0.667
Kurtosis:                       2.867   Cond. No.                         7.17
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Iteration: 2
  Eliminating: Education (p-value: 0.599
  AIC: -635.130, BIC: -593.679, Adj_R_squared: 0.867
```

Figure 3: OLS Regression Summary - Iteration 2 (10 features)

```
ITERATION 3 - OLS SUMMARY (9 features)
                            OLS Regression Results
==============================================================================
Dep. Variable:                  Sales   R-squared:                       0.872
Model:                            OLS   Adj. R-squared:                  0.868
Method:                 Least Squares   F-statistic:                     233.8
Date:                Thu, 30 Oct 2025   Prob (F-statistic):          1.74e-132
Time:                        15:24:05   Log-Likelihood:                -125.64
No. Observations:                 320   AIC:                             271.3
Df Residuals:                     310   BIC:                             309.0
Df Model:                           9
Covariance Type:            nonrobust
================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
const             -0.7517      0.069    -10.929      0.000      -0.887      -0.616
CompPrice          0.5061      0.025     20.052      0.000       0.456       0.556
Income             0.1679      0.021      8.150      0.000       0.127       0.208
Advertising        0.2728      0.028      9.618      0.000       0.217       0.329
Price             -0.7911      0.025    -31.376      0.000      -0.841      -0.741
Age               -0.2720      0.021    -13.257      0.000      -0.312      -0.232
ShelveLoc_Good     1.7537      0.061     28.915      0.000       1.634       1.873
ShelveLoc_Medium   0.7009      0.050     14.055      0.000       0.603       0.799
Urban_Yes          0.0565      0.045      1.242      0.215      -0.033       0.146
US_Yes            -0.0577      0.060     -0.960      0.338      -0.176       0.061
==============================================================================
Omnibus:                        0.760   Durbin-Watson:                   2.054
Prob(Omnibus):                  0.684   Jarque-Bera (JB):                0.877
Skew:                           0.092   Prob(JB):                        0.645
Kurtosis:                       2.821   Cond. No.                         7.15
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Iteration: 3
  Eliminating: US_Yes (p-value: 0.338
  AIC: -636.843, BIC: -599.160, Adj_R_squared: 0.867
```

Figure 4: OLS Regression Summary - Iteration 3 (9 features)

```
ITERATION 4 - OLS SUMMARY (8 features)
                            OLS Regression Results
==============================================================================
Dep. Variable:                  Sales   R-squared:                       0.871
Model:                            OLS   Adj. R-squared:                  0.868
Method:                 Least Squares   F-statistic:                     263.0
Date:                Thu, 30 Oct 2025   Prob (F-statistic):          1.64e-133
Time:                        15:24:05   Log-Likelihood:                -126.11
No. Observations:                 320   AIC:                             270.2
Df Residuals:                     311   BIC:                             304.1
Df Model:                           8
Covariance Type:            nonrobust
================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
const             -0.7922      0.054    -14.585      0.000      -0.899      -0.685
CompPrice          0.5055      0.025     20.038      0.000       0.456       0.555
Income             0.1669      0.021      8.114      0.000       0.126       0.207
Advertising        0.2539      0.020     12.405      0.000       0.214       0.294
Price             -0.7920      0.025    -31.443      0.000      -0.842      -0.742
Age               -0.2723      0.021    -13.276      0.000      -0.313      -0.232
ShelveLoc_Good     1.7540      0.061     28.924      0.000       1.635       1.873
ShelveLoc_Medium   0.7050      0.050     14.193      0.000       0.607       0.803
Urban_Yes          0.0567      0.045      1.247      0.213      -0.033       0.146
==============================================================================
Omnibus:                        0.758   Durbin-Watson:                   2.047
Prob(Omnibus):                  0.685   Jarque-Bera (JB):                0.867
Skew:                           0.102   Prob(JB):                        0.648
Kurtosis:                       2.848   Cond. No.                         5.69
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Iteration: 4
  Eliminating: Urban_Yes (p-value: 0.213
  AIC: -637.893, BIC: -603.978, Adj_R_squared: 0.867
```

Figure 5: OLS Regression Summary - Iteration 4 (8 features)

```
ITERATION 5 - OLS SUMMARY (7 features)
                            OLS Regression Results
==============================================================================
Dep. Variable:                  Sales   R-squared:                       0.871
Model:                            OLS   Adj. R-squared:                  0.868
Method:                 Least Squares   F-statistic:                     299.8
Date:                Thu, 30 Oct 2025   Prob (F-statistic):          1.98e-134
Time:                        15:24:05   Log-Likelihood:                -126.91
No. Observations:                 320   AIC:                             269.8
Df Residuals:                     312   BIC:                             300.0
Df Model:                           7
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.7477      0.041    -18.270      0.000      -0.828      -0.667
CompPrice      0.5075      0.025     20.136      0.000       0.458       0.557
Income         0.1681      0.021      8.173      0.000       0.128       0.209
Advertising    0.2538      0.020     12.387      0.000       0.213       0.294
Price         -0.7925      0.025    -31.434      0.000      -0.842      -0.743
Age           -0.2714      0.021    -13.228      0.000      -0.312      -0.231
ShelveLoc_Good  1.7487     0.061     28.882      0.000       1.630       1.868
ShelveLoc_Medium 0.6997    0.050     14.126      0.000       0.602       0.797
==============================================================================
Omnibus:                        0.862   Durbin-Watson:                   2.053
Prob(Omnibus):                  0.650   Jarque-Bera (JB):                0.976
Skew:                           0.099   Prob(JB):                        0.614
Kurtosis:                       2.815   Cond. No.                         4.89
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Stopping: All features have p-value <= 0.01
 Iteration  Features_Count Feature_Eliminated      P_value         AIC         BIC  Adj_R_squared
         1             11         Population 9.617507e-01 -633.132493 -587.912641       0.866706
         2             10          Education 5.987210e-01 -635.130100 -593.678569       0.867137
         3              9             US_Yes 3.378304e-01 -636.842848 -599.159638       0.867449
         4              8          Urban_Yes 2.134418e-01 -637.893037 -603.978148       0.867483
         5              7               None 7.619271e-15 -638.297753 -608.151185       0.867249
```

Figure 6: OLS Regression Summary - Iteration 5 (7 features)

Final Model (Iteration 5) - 7 Features:

- R-squared: 0.871
- Adjusted R-squared: 0.868
- F-statistic: 299.8
- AIC: 269.8
- BIC: 300.0
- All features have p-value ≤ 0.01

2.3 Final Regression Model

The final regression equation with 7 significant features is:

$$\begin{aligned}\text{Sales} = &-0.748 + 0.507 \times \text{CompPrice} + 0.168 \times \text{Income} \\ &+ 0.254 \times \text{Advertising} - 0.792 \times \text{Price} \\ &- 0.271 \times \text{Age} + 1.749 \times \text{ShelveLoc\_Good} \\ &+ 0.700 \times \text{ShelveLoc\_Medium}\end{aligned} \tag{1}$$

2.4 Prediction vs Test Set

	Actual_Sales	Predicted_Sales	Difference
0	2.263812	2.228757	0.035056
1	0.101893	-0.146335	0.248228
2	-0.998528	-0.471935	-0.526593
3	-0.351013	-0.279924	-0.071089
4	-0.602235	-0.469375	-0.132860
5	-0.800381	-1.127828	0.327447
6	0.289424	0.396014	-0.106589
7	-1.391282	-0.397780	-0.993502
8	1.050165	1.321531	-0.271366
9	-0.358090	-0.413283	0.055193

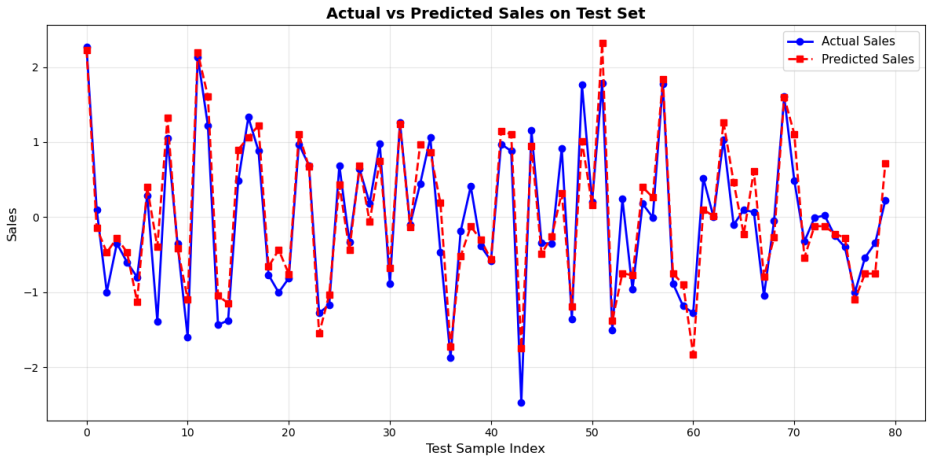

Figure 7: Original Test Set vs Predicted Sales (Backward Stepwise)

2.5 Mean Squared Error

- **Mean Squared Error (MSE): 0.984**
- Root Mean Squared Error (RMSE): 0.992
- Mean Absolute Error (MAE): 0.273

3 PCA

3.1 95% Variance Explained

8 principal components are needed to explain more than 95% of the variance.

- Total number of features: 11
- Number of components for 95% variance: 8
- Cumulative variance explained: 0.9512 (95.12%)

Table 2: Principal Component Analysis - Variance Explained

Component	Variance Explained	Cumulative Variance
PC1	0.208	0.208
PC2	0.172	0.381
PC3	0.134	0.515
PC4	0.125	0.640
PC5	0.122	0.762
PC6	0.095	0.857
PC7	0.053	0.910
PC8	0.042	**0.951**
PC9	0.026	0.977
PC10	0.013	0.990
PC11	0.010	1.000

3.2 Cumulative Variance Plot & 95% Variance Threshold

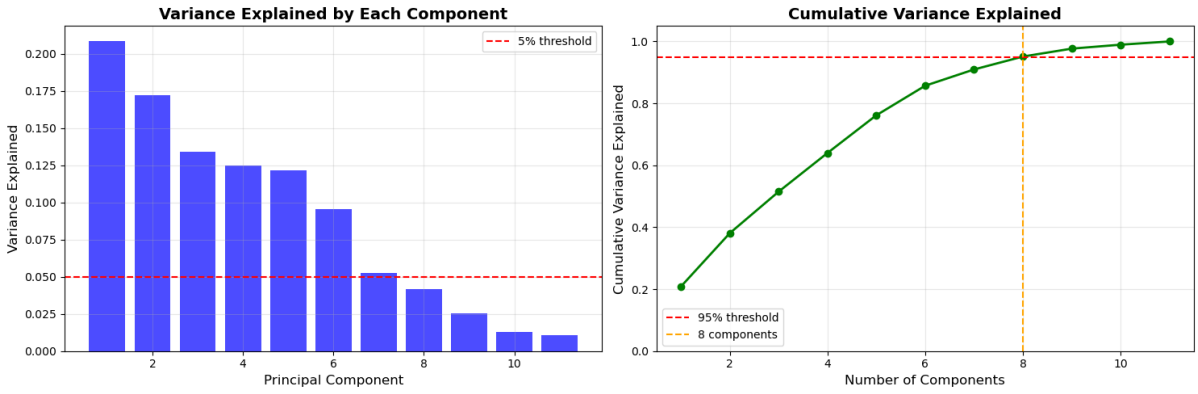

Figure 8: Cumulative Explained Variance vs Number of Components & Horizontal & Vertical Lines Displaying 95% Threshold

3.3 What Does PCA Say?

PCA has told us that of the 11 original features, we need 8 principal components to explain 95% of the variance. This is indicating that our data has some redundancy/correlation, so after we

transform the data into a new coordinate system, we can sufficiently use 8 dimensions instead of 11. However, since PCA gives us a weighted combination of all original features, and not the individual features, it doesn't tell us explicitly which features to remove, but has created new features for us to use. To learn which features we should remove, we should use backward/forward stepwise regression, or use a Random Forest model to find which features to split on.

4 Random Forest Analysis

4.1 Feature Importance Bar Plot

The Random Forest analysis identified the following feature importances:

Table 3: Random Forest Feature Importance

Feature	Importance
Price	0.281
ShelveLoc_Good	0.251
CompPrice	0.108
Age	0.106
Advertising	0.066
Income	0.057
ShelveLoc_Medium	0.055
Population	0.037
Education	0.029
Urban_Yes	0.005
US_Yes	0.004

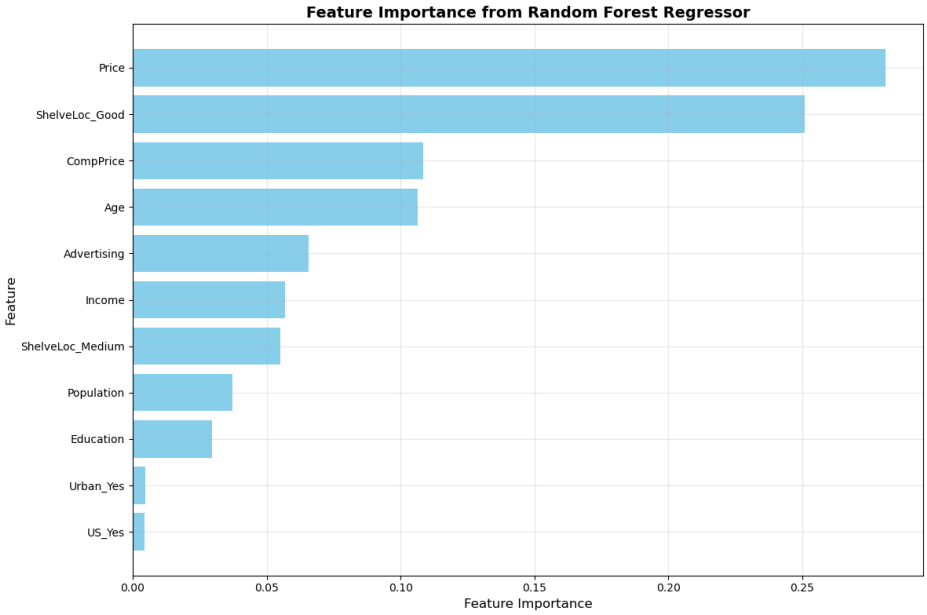

Figure 9: Random Forest Feature Importance (Descending Order)

4.2 Feature Selection Comparison

Random Forest Selected Features (7) - Threshold: 0.05:

1. Price
2. ShelveLoc_Good
3. CompPrice
4. Age
5. Advertising
6. Income
7. ShelveLoc_Medium

Random Forest Eliminated Features (4):

1. Population
2. Education
3. Urban_Yes
4. US_Yes

Yes, the selected features from Random Forest and Backward Stepwise Regression are **identical**. Both methods selected the same 7 features and eliminated the same 4 features (Population, Education, Urban_Yes, US_Yes).

4.3 OLS Regression on Random Forest Selected Features

The features selected to be remove by Random Forest are the exact same as the features selected by the Backward Stepwise Regression, the same as displayed in Figures 2 - 6. This means that the OLS summaries are identical. That being said, I will still include a screenshot!

```
OLS REGRESSION SUMMARY (Random Forest Selected Features)
                            OLS Regression Results
==============================================================================
Dep. Variable:                  Sales   R-squared:                       0.871
Model:                            OLS   Adj. R-squared:                  0.868
Method:                 Least Squares   F-statistic:                     299.8
Date:                Thu, 30 Oct 2025   Prob (F-statistic):          1.98e-134
Time:                        15:39:39   Log-Likelihood:                -126.91
No. Observations:                 320   AIC:                             269.8
Df Residuals:                     312   BIC:                             300.0
Df Model:                           7
Covariance Type:            nonrobust
==============================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const             -0.7477      0.041    -18.270      0.000      -0.828      -0.667
Price             -0.7925      0.025    -31.434      0.000      -0.842      -0.743
ShelveLoc_Good     1.7487      0.061     28.882      0.000       1.630       1.868
CompPrice          0.5075      0.025     20.136      0.000       0.458       0.557
Age               -0.2714      0.021    -13.228      0.000      -0.312      -0.231
Advertising        0.2538      0.020     12.387      0.000       0.213       0.294
Income             0.1681      0.021      8.173      0.000       0.128       0.209
ShelveLoc_Medium   0.6997      0.050     14.126      0.000       0.602       0.797
==============================================================================
Omnibus:                        0.862   Durbin-Watson:                   2.053
Prob(Omnibus):                  0.650   Jarque-Bera (JB):                0.976
Skew:                           0.099   Prob(JB):                        0.614
Kurtosis:                       2.815   Cond. No.                         4.89
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

PREDICTION PERFORMANCE (Random Forest Selected Features)
Mean Squared Error (MSE): 0.9841
Root Mean Squared Error (RMSE): 0.9920
```

Figure 10: OLS Summary (Random Forest Features)

4.4 Prediction vs Test Set

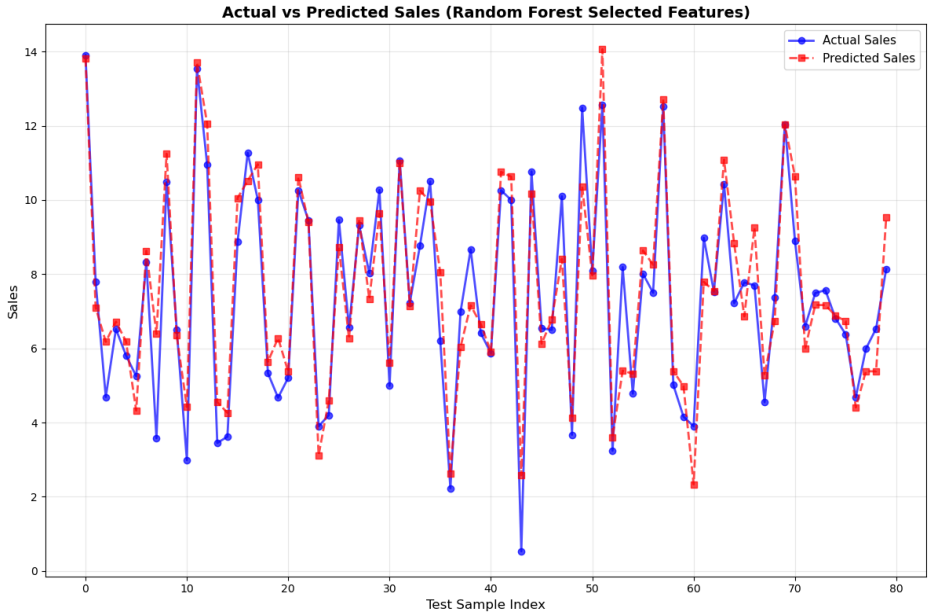

Figure 11: Original Test Set vs Predicted Sales (Random Forest Features)

4.5 Mean Squared Error

- **Mean Squared Error (MSE): 0.9841**
- Root Mean Squared Error (RMSE): 0.9920

5 Comparison of Feature Selection Methods

5.1 Metrics Comparison Table

There is a *PrettyTable* output in the console of the file, by I have also included a formatted table here.

Table 4: Model Comparison Summary

Model	R^2	Adj. R^2	AIC	BIC	MSE
Backward Stepwise	0.8706	0.8677	269.82	299.97	0.9841
Random Forest	0.8706	0.8677	269.82	299.97	0.9841

5.2 Recommended Method and Features

Both methods produced **identical results** in terms of all performance metrics:

- Same R^2 and Adjusted R^2 (0.8706 and 0.8677)
- Same AIC and BIC (269.82 and 299.97)
- Same MSE (0.9841)
- Same 7 selected features

Recommended Method: Either method can be recommended, but:

- **Backward Stepwise Regression** is preferred for interpretability as it provides statistical significance (p-values) for each feature
- **Random Forest** is preferred for capturing non-linear relationships and feature interactions

Recommended Features for Elimination:

1. Population
2. Education
3. Urban_Yes
4. US_Yes

6 Prediction Interval

6.1 95% Prediction Intervals

Prediction Summary (First 10 rows):

Table 5: Prediction Intervals (Standardized Scale)

Index	Mean	Mean SE	CI Lower	CI Upper	PI Lower	PI Upper
18	2.229	0.075	2.081	2.376	1.497	2.961
372	-0.146	0.042	-0.228	-0.065	-0.868	0.575
9	-0.472	0.059	-0.587	-0.357	-1.198	0.254
127	-0.280	0.033	-0.345	-0.215	-1.000	0.440
379	-0.469	0.055	-0.577	-0.362	-1.194	0.256
362	-1.128	0.059	-1.244	-1.012	-1.854	-0.402
26	0.396	0.069	0.260	0.532	-0.334	1.126
356	-0.398	0.079	-0.552	-0.243	-1.131	0.336
177	1.322	0.062	1.199	1.444	0.594	2.049
131	-0.413	0.047	-0.506	-0.321	-1.136	0.310

Prediction Intervals (Original Scale - First 10 samples):

Actual_Sales	Predicted_Sales	Lower_95%_PI	Upper_95%_PI	Within_Interval
13.91	13.810926	11.742477	15.879375	True
7.80	7.098459	5.059313	9.137605	True
4.69	6.178252	4.126199	8.230305	True
6.52	6.720913	4.686524	8.755301	True

5.81	6.185488	4.136832	8.234144	True
5.25	4.324572	2.272089	6.377055	True
8.33	8.631242	6.569331	10.693152	True
3.58	6.387829	4.315299	8.460358	False
10.48	11.246932	9.191514	13.302351	True
6.50	6.344013	4.301129	8.386898	True

Coverage Statistics:

- Coverage: 96.250% of actual values fall within 95% prediction intervals
- Expected: ∼95%
- Average prediction interval width: 4.103

6.2 Prediction Interval Plot

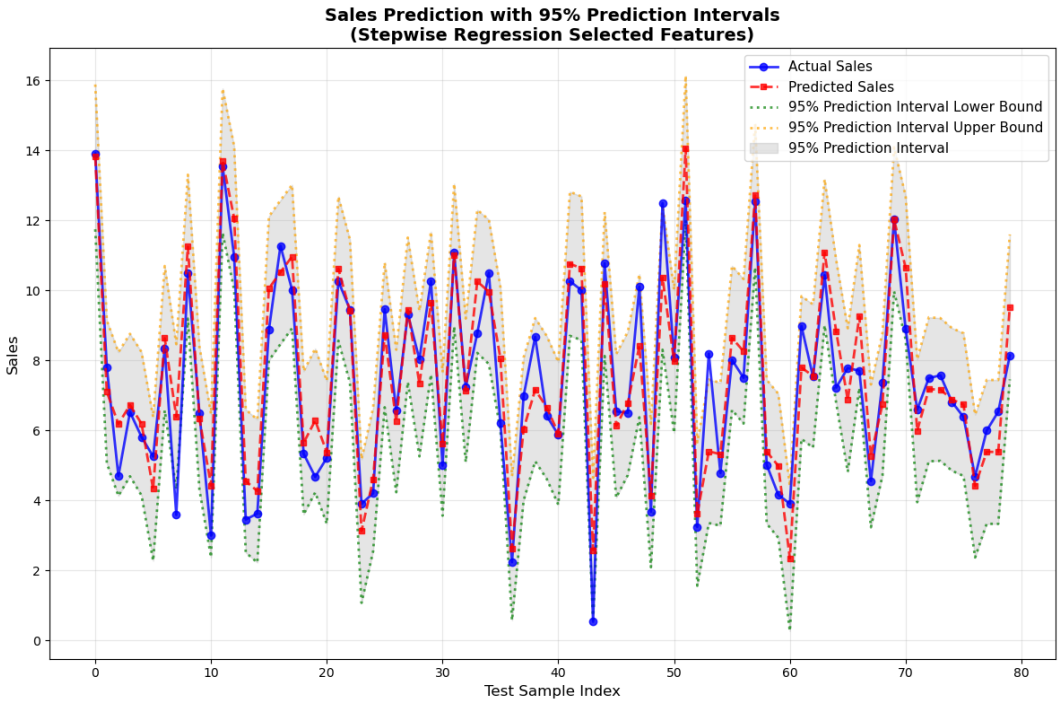

Figure 12: Predicted Sales with 95% Prediction Intervals

7 Polynomial Regression and Grid Search

7.1 Grid Search with RMSE Minimization

Grid search was performed with 5-fold cross-validation for polynomial degrees 1 through 15.

- Training samples: 320
- Testing samples: 80
- Cross-validation folds: 5
- Total fits: 75 (15 candidates × 5 folds)

7.2 Optimum Polynomial Order

$\boxed{\text{The optimum polynomial degree is n} = 4}$ with a cross-validation RMSE of 2.565.

7.3 RMSE vs Polynomial Order Plot

Table 6: RMSE for Each Polynomial Degree

Degree	RMSE
1	2.578
2	2.586
3	2.574
4	**2.565**
5	2.570
6	2.577
7	2.586
8	2.605
9	2.646
10	2.717
11	2.816
12	2.939
13	3.075
14	3.206
15	3.312

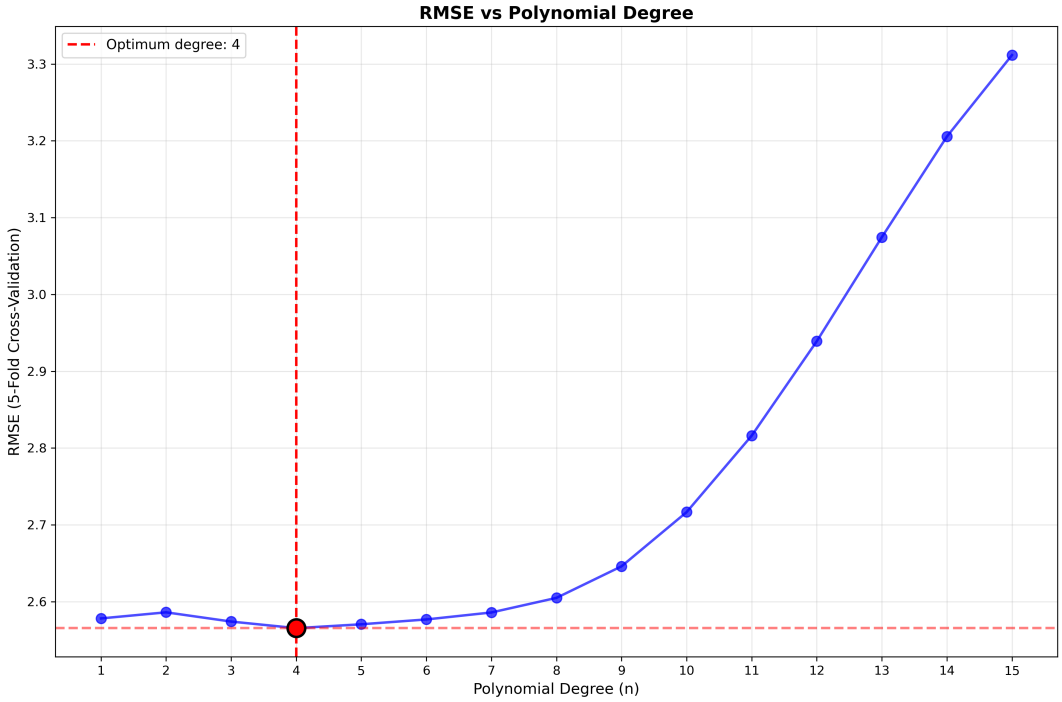

Figure 13: RMSE vs Polynomial Degree (n)

7.4 Training and Prediction

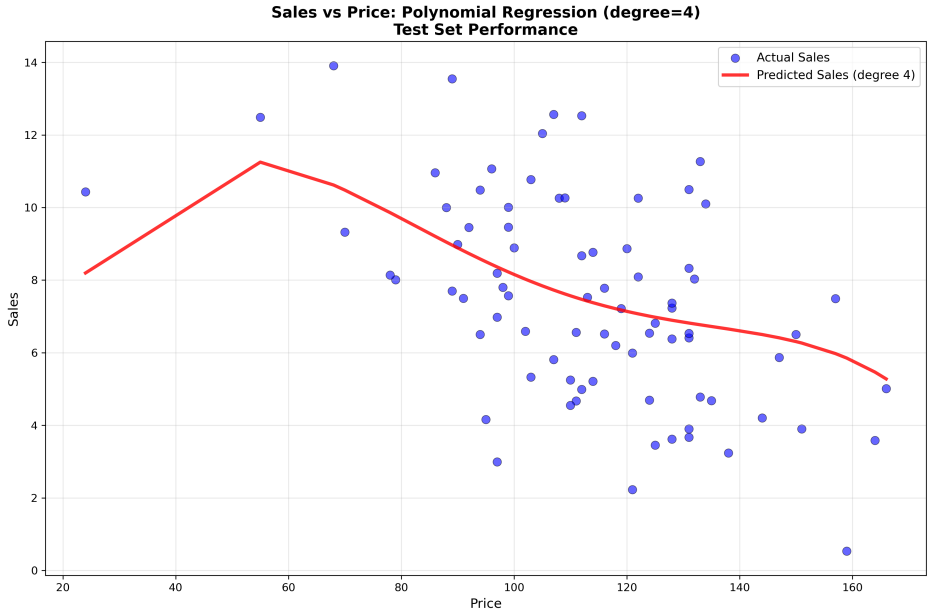

Figure 14: Test Set vs Predicted Sales (4th Order Polynomial)

7.5 Mean Squared Error

- Polynomial degree: 4
- **Test Set MSE: 5.826**

- Test Set RMSE: 2.414
- Test Set R^2: 0.255

8 Simple Linear Regression Proof

In a simple linear regression with n observations:

$$y_i \approx \hat{\beta}_0 + \hat{\beta}_1 x_i \tag{2}$$

Prove the following:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{3}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \tag{4}$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ are the sample mean.

First, we have to begin with the Residual Sum of Squares (RSS) equation as given in class:

$$\text{RSS} = \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \tag{5}$$

and then take the derivative of this equation with respect to $\hat{\beta}_0$ and $\hat{\beta}_1$.

First, we will start with $\hat{\beta}_0$.

$$\frac{\partial \text{RSS}}{\partial \hat{\beta}_0} = \frac{\partial}{\partial \hat{\beta}_0}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \tag{6}$$

$$\tag{7}$$

A key part of this derivation is using substitution. We can make $u = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_n$ and then use the rule $\frac{\partial}{\partial \hat{\beta}_0} u^2 = 2u\frac{\partial u}{\partial \hat{\beta}_0}$. Going from there, we can get

$$\frac{\partial}{\partial \hat{\beta}_0}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = -2(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) \tag{8}$$

$$\frac{\partial \text{RSS}}{\partial \hat{\beta}_0} = \sum_{i=1}^{n} -2(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) \tag{9}$$

$$\tag{10}$$

Setting the derivation equal to zero, we can continue onward.

$$\frac{\partial \text{RSS}}{\partial \hat{\beta}_0} = -2\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \tag{11}$$

$$\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \tag{12}$$

$$\sum_{i=1}^{n} y_i - \sum_{i=1}^{n}\hat{\beta}_0 - \sum_{i=1}^{n}\hat{\beta}_1 x_i = 0 \tag{13}$$

$$\tag{14}$$

Because $\hat{\beta}_0$ and $\hat{\beta}_1$ are constants, they can come outside the summations. For $\hat{\beta}_0$ specifically, the summation will just be $\hat{\beta}_0$ multiplied n times by itself. So, continuing on

$$\sum_{i=1}^{n} y_i - n\hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^{n} x_i = 0 \tag{15}$$

$$-n\hat{\beta}_0 = \hat{\beta}_1 \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i \tag{16}$$

$$\hat{\beta}_0 = \frac{1}{n}\sum_{i=1}^{n} y_i - \hat{\beta}_1 \frac{1}{n}\sum_{i=1}^{n} x_i \tag{17}$$

$$\tag{18}$$

Which then becomes

$$\boxed{\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}} \tag{19}$$

Wonderful! Now, to $\hat{\beta}_1$.

$$\frac{\partial \text{RSS}}{\partial \hat{\beta}_1} = \frac{\partial}{\partial \hat{\beta}_1} \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \tag{20}$$

With the same substitution and rule from the first derivation, we get

$$\frac{\partial}{\partial \hat{\beta}_1} \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = -2x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) \tag{21}$$

We can then continue on

$$\frac{\partial \text{RSS}}{\partial \hat{\beta}_1} = -2\sum_{i=1}^{n} x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \tag{22}$$

$$\sum_{i=1}^{n} x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \tag{23}$$

$$\sum_{i=1}^{n} x_i(y_i - (\bar{y} - \hat{\beta}_1 \bar{x}) - \hat{\beta}_1 x_i) = 0 \tag{24}$$

$$\sum_{i=1}^{n} x_i(y_i - \bar{y} - \hat{\beta}_1(x_i - \bar{x})) = 0 \tag{25}$$

$$\sum_{i=1}^{n} x_i(y_i - \bar{y}) - \sum_{i=1}^{n} x_i \hat{\beta}_1 x_i(x_i - \bar{x}) = 0 \tag{26}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i(y_i - \bar{y})}{\sum_{i=1}^{n} x_i(x_i - \bar{x})} \tag{27}$$

$$\tag{28}$$