# Laptop Price Prediction

**By**

**Vamshi Krishna Vallam**

**Advanced Data Analytics**

**11625644**

**University of North Texas**

**Final Project**

**ADTA 5130 - Data Analytics I**

**Dr. Denise Philpot**

# Contents

# Problem Statement and Hypothesis

The purpose of this project is to use different statistical approaches and compare different variables present in the data.

**Problem Statement 1**

To determine whether there is any significant difference between the average price of laptops based on the presence of Bundled Applications present in it at 95% Confidence Level.

**Hypothesis**

**Null Hypothesis (Ho) –** There is no difference between the means of laptop prices based on the presence of Bundled Applications.

**Alternate Hypothesis (Ha) –** There is a difference between the means of laptop prices based on the presence of Bundled Applications.

**Problem Statement 2**

To Predict the Laptop price based on Screen Size, Battery Life, RAM, Processor Speed, Integrated Wireless, Hard Disk Size, and Bundled Applications.

# Dataset

The dataset consists of 9 variables which are ‘Configuration’, ‘Screen Size (Inches)’, ‘Battery Life (Hours)’, ‘RAM (GB)’, ‘Processor Speed (GHz)’, ‘Integrated Wireless?’, ‘HD Size (GB)’, ‘Bundled Applications?’ and ‘Price’. The ‘Price’ is the dependent variable in the dataset.

Configuration: It is the model of the laptop.

Screen Size (Inches): It is the size of the screen of laptop.

Battery Life (Hours): It is the number of hours the laptop works on a single charge.

RAM (GB): It is the short-term memory, where the data is stored temporarily.

Processor Speed (GHz): It determines the speed of the laptop.

Integrated Wireless?: It refers to the inbuilt hardware for connecting wirelessly.

HD Size (GB): It refers to the space in the hard drive in Gigabytes.

Bundled Applications?: It is the set of Software that is sold along with the laptop.

Price: It is the cost at which the laptop is sold.

The dataset consists of 99999-row values. The dataset consists of 2 Categorical variables which are ‘Integrated Wireless? And ‘Bundled Applications?

The ‘Integrated Wireless’ variable consists of 2 values which are ‘Yes’ if it is Integrated Wireless and ‘No’ if it is not Integrated Wireless.

The ‘Bundled Application’ variable also consists of 2 values which are ‘Yes’ if it has Bundled Applications and ‘No’ if it does not have Bundled Applications.

Below is a sample of the dataset.

| | Configuration | Screen Size (Inches) | Battery Life (Hours) | RAM (GB) | Processor Speeds (GHz) | Integrated Wireless? | HD Size (GB) | Bundled Applications? | Price |
|---|---|---|---|---|---|---|---|---|---|
| **0** | 163 | 15 | 5 | 4 | 2.0 | Yes | 80 | Yes | 1455 |
| **1** | 320 | 15 | 6 | 4 | 2.0 | No | 300 | No | 1545 |
| **2** | 23 | 15 | 4 | 4 | 2.0 | Yes | 300 | Yes | 1515 |
| **3** | 169 | 15 | 5 | 4 | 2.0 | No | 40 | Yes | 1395 |
| **4** | 365 | 15 | 6 | 8 | 2.0 | No | 120 | Yes | 1585 |

# Data Preparation

It is the method of data manipulation that is required if there are any irregularities in the data. Data is prepared so that it is suitable for model building and analysis.

## Null Values

We need to check the presence of Null values in the dataset. These Null values are the blank values that are not present in the dataset due to many reasons and can be a problem for the Analysts. The Null values must be handled by removing them or replacing them in the dataset.

Let's check for the Null values using Python by running the '**isnull().sum()**' function.

```
sales.isnull().sum()
```

```
Configuration             0
Screen Size (Inches)      0
Battery Life (Hours)      0
RAM (GB)                  0
Processor Speeds (GHz)    0
Integrated Wireless?      0
HD Size (GB)              0
Bundled Applications?     0
Price                     0
dtype: int64
```

We can see that none of the variables have Null values and is good to proceed to the next step.

## Dummies

Since we have 2 categorical variables 'Integrated Wireless?' and 'Bundled Applications?' with 'Yes' or 'No' as their values, we need to change those values into dummies which will help in better analyzing. We need to change these categorical values into numerical dummies which will help in developing a proper regression model.

We will create dummy variables for 'Integrated Wireless?' and 'Bundled applications?' using Excel.

We have used the formula "**=IF(F2="Yes",1,0)"** for 'Integrated Wireless' and **"=IF(H2="Yes",1,0)"** for 'Bundled Application?' in new columns and got the required dummies.

| Configu | Screen | Battery | RAM (G | Process | Integra | HD Size | Bundle | Price | Integrated Wireless | Bundled Applications |
|---|---|---|---|---|---|---|---|---|---|---|
| 163 | 15 | 5 | 4 | 2 | Yes | 80 | Yes | 1455 | 1 | 1 |
| 320 | 15 | 6 | 4 | 2 | No | 300 | No | 1545 | 0 | 0 |
| 23 | 15 | 4 | 4 | 2 | Yes | 300 | Yes | 1515 | 1 | 1 |
| 169 | 15 | 5 | 4 | 2 | No | 40 | Yes | 1395 | 0 | 1 |
| 365 | 15 | 6 | 8 | 2 | No | 120 | Yes | 1585 | 0 | 1 |

We can see that the new columns ‘Integrated Wireless’ and ‘Bundled Applications’ are created with dummies.

We will properly arrange the variables and remove the categorical variables and keep only the new dummy variables created.

Below is the final prepared data.

| Configura | Screen Siz | Battery Lif | RAM (GB) | Processor | Integrated Wireless | HD Size (G | Bundled Applications | Price |
|---|---|---|---|---|---|---|---|---|
| 163 | 15 | 5 | 4 | 2 | 1 | 80 | 1 | 1455 |
| 320 | 15 | 6 | 4 | 2 | 0 | 300 | 0 | 1545 |
| 23 | 15 | 4 | 4 | 2 | 1 | 300 | 1 | 1515 |
| 169 | 15 | 5 | 4 | 2 | 0 | 40 | 1 | 1395 |
| 365 | 15 | 6 | 8 | 2 | 0 | 120 | 1 | 1585 |
| 309 | 15 | 6 | 4 | 2 | 1 | 120 | 1 | 1555 |
| 75 | 15 | 4 | 8 | 2 | 0 | 80 | 1 | 1465 |
| 346 | 15 | 6 | 8 | 1.5 | 0 | 40 | 0 | 1450 |

# Basic Statistics and Visualizations

Now, we’ll be looking at different descriptive statistics of various variables present in the dataset. I am using Python for statistical values and visualizations.

## Configuration

```
sales['Configuration'].describe()
```

```
count    99999.000000
mean       328.490615
std        219.346758
min          1.000000
25%        163.000000
50%        304.000000
75%        470.000000
max        864.000000
Name: Configuration, dtype: float64
```

```
sales['Configuration'].nunique()
```

```
864
```

```
sales['Configuration'].mode()
```

```
0    61
dtype: int64
```

There are a total of 864 unique values for ‘Configuration’ in the dataset starting from 1. These 864 different ‘Configuration’ values are divided between 99999 values based on the laptop.

The mode of the ‘Configuration’ variable in 61, means that the ‘Configuration’ type 61 has occurred mostly in the dataset.

## Screen Size (Inches)

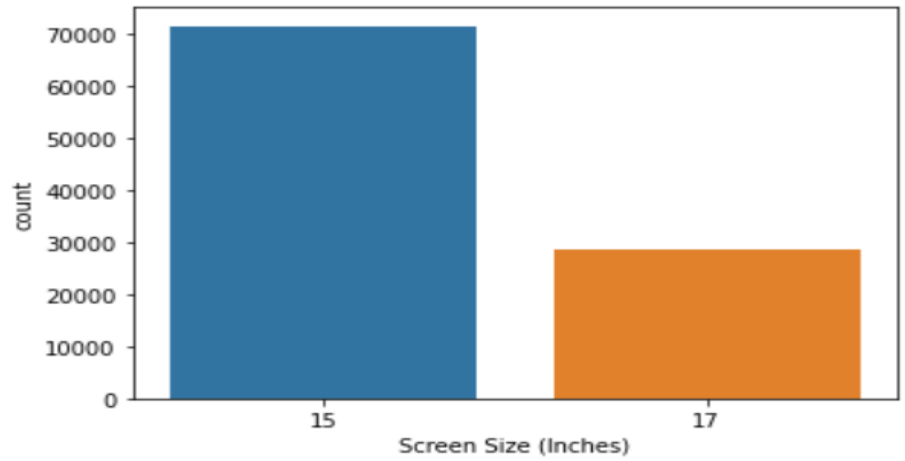
```
sales['Screen Size (Inches)'].describe()
```

```
count    99999.000000
mean        15.569966
std          0.902817
min         15.000000
25%         15.000000
50%         15.000000
75%         17.000000
max         17.000000
Name: Screen Size (Inches), dtype: float64
```

We can see that the minimum value of ‘Screen Size’ is 15 inches and the maximum value of ‘Screen Size’ is 17 inches.

```
import matplotlib.pyplot as plt
import seaborn as sns
```

```
sns.countplot(x = "Screen Size (Inches)",data = sales)
plt.show()
```



We can see that around 70,000 laptops have 15 inches of display, and the remaining 30,000 laptops are with 17 inches of display.

## Battery Life

```
sales['Battery Life (Hours)'].describe()
```

```
count    99999.000000
mean         5.022310
std          0.815111
min          4.000000
25%          4.000000
50%          5.000000
75%          6.000000
max          6.000000
Name: Battery Life (Hours), dtype: float64
```

```
sales['Battery Life (Hours)'].mode()
```

```
0    6
dtype: int64
```

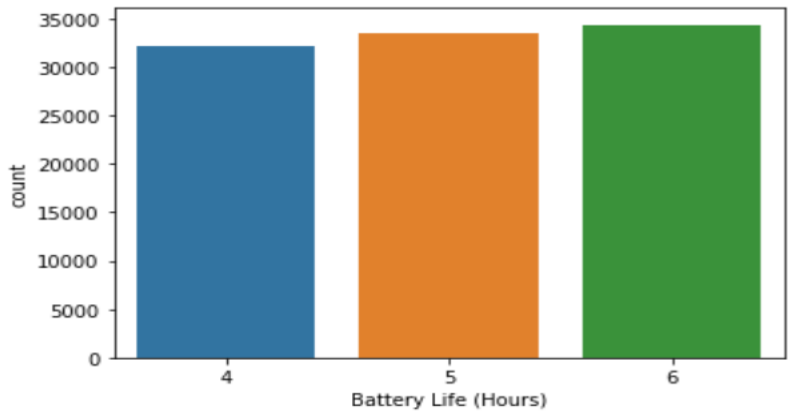
```
sales['Battery Life (Hours)'].nunique()
```

```
3
```

There are 3 different values present in the 'Battery Life' variable. 6 hours is the most occurred value in the 'Battery Life'.

The count plot for the 'Battery Life (Hours)' is shown below.

```
sns.countplot(x = "Battery Life (Hours)",data = sales)
plt.show()
```



## RAM (GB)

```
sales['RAM (GB)'].describe()
```

```
count    99999.000000
mean         7.738157
std          4.120614
min          4.000000
25%          4.000000
50%          8.000000
75%          8.000000
max         16.000000
Name: RAM (GB), dtype: float64
```

```
sales['RAM (GB)'].mode()
```

```
0    8
dtype: int64
```

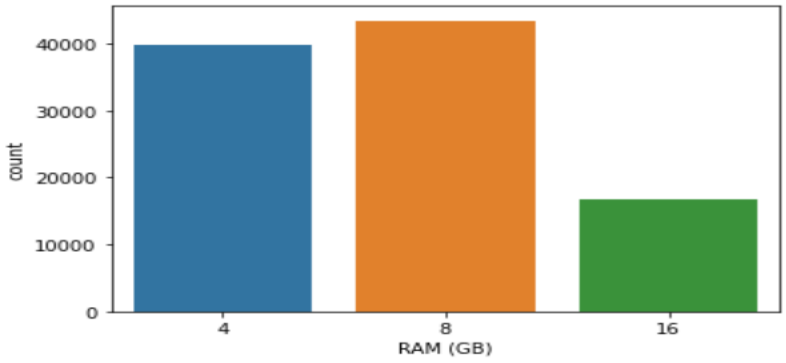
```
sales['RAM (GB)'].nunique()
```

```
3
```

There are 3 unique values for 'RAM (GB)' which are 4, 8, and 16. 8 GB of RAM has the maximum count in the dataset.

The count plot for 'RAM' is shown below.

```
sns.countplot(x = "RAM (GB)",data = sales)
plt.show()
```



## Processor Speeds (GHz)

```
sales['Processor Speeds (GHz)'].describe()
```

```
count    99999.000000
mean         1.879778
std          0.340255
min          1.500000
25%          1.500000
50%          2.000000
75%          2.000000
max          2.400000
Name: Processor Speeds (GHz), dtype: float64
```

```
sales['Processor Speeds (GHz)'].mode()
```

```
0    2.0
dtype: float64
```

```
sales['Processor Speeds (GHz)'].nunique()
```

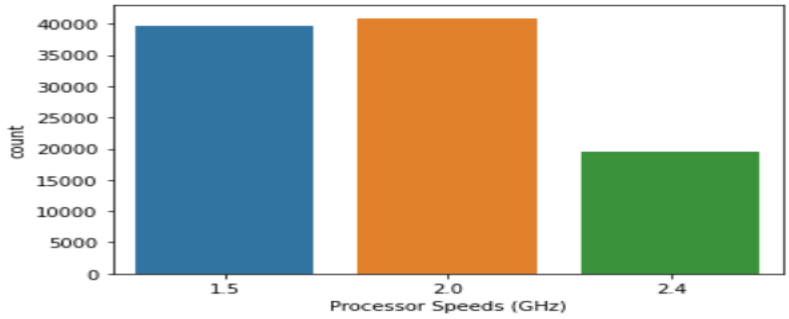
```
3
```

There are 3 unique values for 'Processor Speeds (GHz)' which are 1.5, 2, and 2.4 GHz.

Processor Speed of 2 GHz has the highest count in the dataset.

The count plot for the 'Processor Speeds (GHz) is shown below.

```
sns.countplot(x = "Processor Speeds (GHz)",data = sales)
plt.show()
```



## HD Size (GB)

```
sales['HD Size (GB)'].describe()
```

```
count    99999.000000
mean       137.455375
std         99.528363
min         40.000000
25%         80.000000
50%        120.000000
75%        300.000000
max        300.000000
Name: HD Size (GB), dtype: float64
```

```
sales['HD Size (GB)'].mode()
```

```
0    120
dtype: int64
```

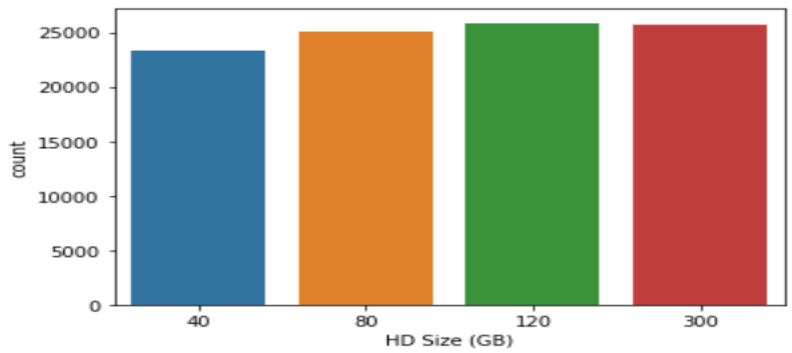
```
sales['HD Size (GB)'].nunique()
```

```
4
```

There are 4 unique values present in the 'HD Size (GB)' variable and they are 40, 80, 120, and 320. The 120 GB HD size has the highest count in the dataset.

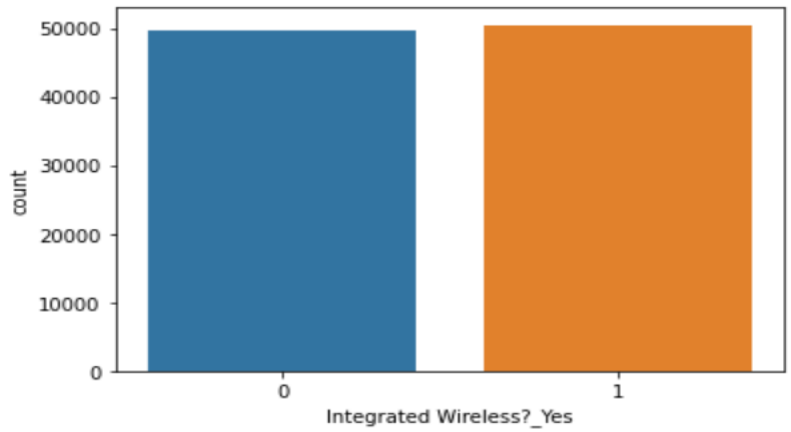
The count plot for the 'HD Size (GB)' is shown below.

```
sns.countplot(x = "HD Size (GB)",data = sales)
plt.show()
```



## Integrated Wireless?

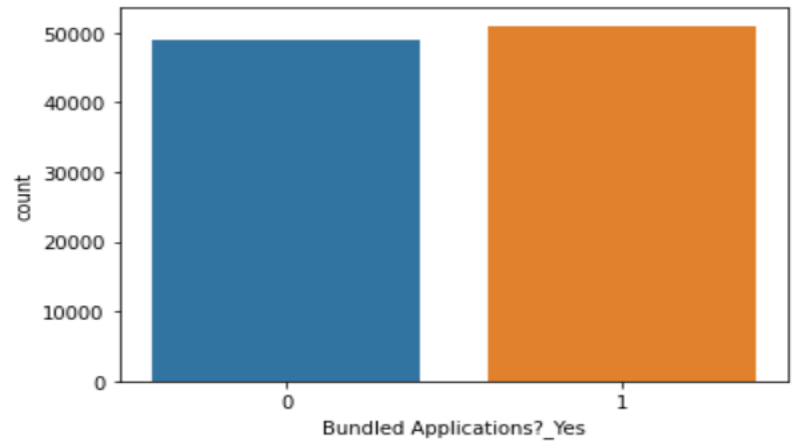
```
sns.countplot(x = "Integrated Wireless?_Yes",data = sales)
plt.show()
```




The newly created variable using dummies 'Integrated Wireless?' has 2 values 1 and 0, in which '1' shows it is Integrated Wireless and '0' shows not Integrated Wireless. The value count of '1' is higher which tells that there are more Integrated Wireless laptops in the dataset.

## Bundled Applications?

```
sns.countplot(x = "Bundled Applications?_Yes",data = sales)
plt.show()
```



The newly created variable using dummies ‘Bundled Applications?’ has 2 values 1 and 0, in which ‘1’ shows it has Bundled Applications and ‘0’ shows do not have Bundled Applications. The value count of ‘1’ is higher which means that the count of laptops with Bundled Applications is higher in the dataset.

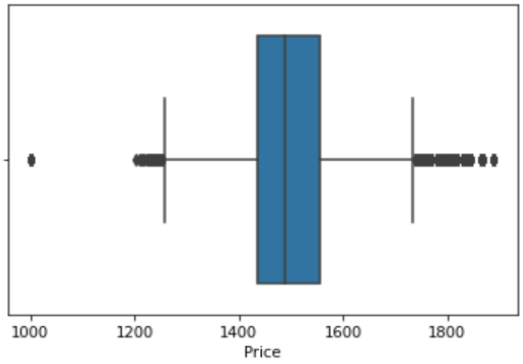
## Price

```
sales['Price'].describe()
```

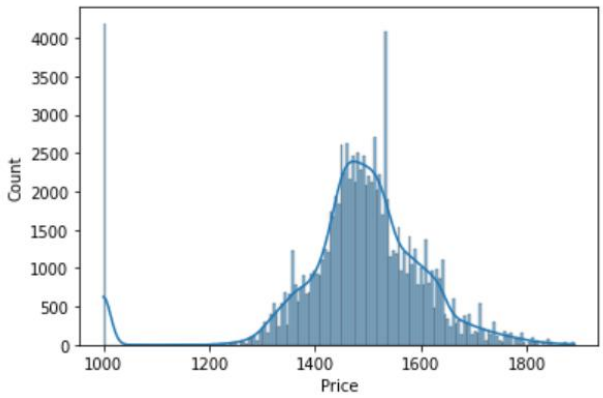
```
count    99999.000000
mean      1481.551246
std        137.789266
min       1000.000000
25%       1435.000000
50%       1490.000000
75%       1555.000000
max       1890.000000
Name: Price, dtype: float64
```

The dependent value Price has a range of values. The minimum value of the laptop Price is 1000 and the maximum value is 1890.

```
sns.boxplot(x= 'Price', data = sales)
plt.show()
```



```
sns.histplot(x = 'Price', data = sales, kde = True)
plt.show()
```




The above box plot shows the median Price is around 1500. The box plot also shows the lower quartile and upper quartile.

The histogram plot shows the distribution of the Price values.

# Approach

Now we will be working on different statistical techniques to work on our problem statements and find the best statistical answer.

## Problem Statement 1

To find whether there is any significant difference between the mean prices of the laptops based on whether they have any Bundled Applications in them, we will collect a sample of data from the dataset which will tell the laptop price based on the Bundled Application variable.

We can add filters and collect sample data.

I have taken 60 samples of laptop prices for each for Bundled Applications -Yes and Bundled Applications - No.

Below is the sample data.

| Bundled Applications? | |
|---|---|
| Yes | No |
| 1455 | 1545 |
| 1515 | 1450 |
| 1395 | 1455 |
| 1585 | 1440 |
| 1555 | 1405 |
| 1465 | 1505 |
| 1620 | 1425 |
| 1465 | 1545 |
| 1665 | 1535 |
| 1565 | 1595 |
| 1480 | 1570 |
| 1490 | 1510 |
| 1480 | 1590 |
| 1460 | 1510 |
| 1465 | 1460 |
| 1530 | 1510 |
| 1585 | 1420 |
| 1540 | 1495 |
| 1510 | 1545 |
| 1520 | 1395 |
| 1530 | 1455 |
| 1525 | 1320 |

| | |
|---|---|
| 1500 | 1465 |
| 1350 | 1495 |
| 1445 | 1370 |
| 1455 | 1490 |
| 1435 | 1465 |
| 1585 | 1480 |
| 1535 | 1460 |
| 1510 | 1445 |
| 1500 | 1360 |
| 1615 | 1490 |
| 1515 | 1460 |
| 1370 | 1510 |
| 1560 | 1540 |
| 1490 | 1485 |
| 1510 | 1615 |
| 1490 | 1435 |
| 1535 | 1570 |
| 1525 | 1505 |
| 1455 | 1475 |
| 1605 | 1490 |
| 1470 | 1320 |
| 1525 | 1345 |
| 1460 | 1420 |
| 1560 | 1455 |

| | |
|---|---|
| 1515 | 1460 |
| 1480 | 1475 |
| 1475 | 1510 |
| 1615 | 1475 |
| 1400 | 1470 |
| 1565 | 1465 |
| 1500 | 1545 |
| 1490 | 1460 |
| 1535 | 1465 |
| 1470 | 1360 |
| 1520 | 1460 |
| 1435 | 1495 |
| 1390 | 1370 |
| 1540 | 1460 |

Now we will use **ANOVA- One Factor Analysis** to find out whether there is any significant difference between mean laptop prices.

## ANOVA- Single Factor

Analysis of Variance or ANOVA is a method that is used to compare the means of two or more group values. In a Single Factor ANOVA there is only one independent group present for comparison.

Different statistical comparisons can be done using an ANOVA table using Microsoft Excel or R Studio.

## Steps for ANOVA: Single Factor

- Add Analysis Tool Pack in the Microsoft Excel.
- Click on ‘Data’.

- Select ‘Data Analysis’ present in the top right corner.
- Select ANOVA: Single Factor.
- Enter the input range.
- Check the box if the text labels are present.
- Enter the significance level.
- Select the cell for the output.
- Click OK

For the above sample data for Problem Statement 1, the ANOVA table is obtained using Microsoft Excel.

Anova: Single Factor

SUMMARY

| *Groups* | *Count* | *Sum* | *Average* | *Variance* |
|---|---|---|---|---|
| Yes | 60 | 90330 | 1505.5 | 3914.153 |
| No | 60 | 88295 | 1471.583 | 4114.823 |

ANOVA

| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
|---|---|---|---|---|---|---|
| Between Groups | 34510.21 | 1 | 34510.21 | 8.596416 | 0.004046 | 3.921478 |
| Within Groups | 473709.6 | 118 | 4014.488 | | | |
| Total | 508219.8 | 119 | | | | |

We have obtained different statistics from our sample.

But for the problem statement, we will be focusing on the F-statistics value, P-value, and the F-Critical value.

The **F- Statistical value** is **~8.6** which is greater than the **F- Critical value** which is **3.92**.

Since our **significance leve**l is **95%** our **Alpha value** is **0.05**

The P-value obtained is **0.0040** which is less than the Alpha value which is **0.05**.

Since **F- Statistic > F-Critical and P-Value < Alpha**, we can reject our Null Hypothesis and accept the Alternate Hypothesis.

**We can conclude that there is a difference between the means of laptop prices based on the presence of Bundled Applications. This means that the mean laptop prices for in-built Bundled Applications are comparatively higher than the laptops that don't have any Bundled Applications. This conclusion cannot have been drawn by just looking at the data, as the prices were almost similar. So, ANOVA helped in getting a proper conclusion on Mean Laptop prices for Bundled Applications.**

## Problem Statement 2

To predict the laptop price which is a dependent variable based on different independent variables such as Screen Size, Battery Life, RAM, Processor Speed, Integrated Wireless, Hard Disk Size, and Bundled Applications.

We will be using the Multiple Regression technique to predict laptop prices. This can be done using Microsoft Excel.

### Regression

Regression is a statistical method that gives the relationship between one dependent variable and one or more independent variables. A Regression model tells us that the change in the dependent variable is due to the change in independent variables.

The dependent variable is denoted by 'Y' and the independent variables are denoted by X1, X2, X3,…….

The prediction of the dependent variable 'Y' is given by the below equation

**Y = Intercept + (Coefficient of X1) *X1 + (Coefficient of X2) *X2 + (Coefficient of X2) *X2……**

The values of Intercept, X1, X2 can be achieved by Regression Table using Analysis tool pack in Microsoft Excel.

For our problem statement,

**Steps for Regression**

- Add Analysis Tool Pack in the Microsoft Excel.
- Click on 'Data'.
- Select 'Data Analysis' present in the top right corner.
- Select Regression
- Select the input range.
- Check the labels.
- Enter the significance level.
- Select the cell for Output.
- Click OK.

For our problem statement, we are considering Screen Size, Battery Life, RAM, Processor Speed, Integrated Wireless, Hard Disk Size, and Bundled Applications as our independent variables. We are supposed to enter the respective cell values of these independent variables in the Regression analysis.

Below is the regression table for our data.

| *Regression Statistics* | |
|---|---|
| Multiple R | 0.55945826 |
| R Square | 0.31299354 |
| Adjusted R Square | 0.31294545 |
| Standard Error | 114.211806 |
| Observations | 99999 |

ANOVA

| | *df* | *SS* | *MS* | *F* | *Significance F* |
|---|---|---|---|---|---|
| Regression | 7 | 594233955.4 | 84890565 | 6507.848 | 0 |
| Residual | 99991 | 1304316266 | 13044.34 | | |
| Total | 99998 | 1898550221 | | | |

| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
|---|---|---|---|---|---|---|---|---|
| Intercept | 259.032486 | 7.326958387 | 35.35334 | 4.3E-272 | 244.671738 | 273.39323 | 244.671738 | 273.393234 |
| Screen Size (Inches) | 46.7305305 | 0.408533816 | 114.386 | 0 | 45.9298093 | 47.531252 | 45.9298093 | 47.5312518 |
| Battery Life (Hours) | 46.5809978 | 0.451632137 | 103.1392 | 0 | 45.6958044 | 47.466191 | 45.6958044 | 47.4661912 |
| RAM (GB) | 11.2232119 | 0.088971056 | 126.1445 | 0 | 11.0488298 | 11.397594 | 11.0488298 | 11.3975941 |
| Processor Speeds (GHz) | 46.5574375 | 1.072457195 | 43.41193 | 0 | 44.4554345 | 48.65944 | 44.4554345 | 48.6594404 |
| Integrated Wireless? | 19.0594862 | 0.723337049 | 26.34938 | 1.7E-152 | 17.6417545 | 20.477218 | 17.6417545 | 20.4772179 |
| HD Size (GB) | 0.38387042 | 0.003692193 | 103.9681 | 0 | 0.37663376 | 0.3911071 | 0.37663376 | 0.39110707 |
| Bundled Applications? | 47.5439051 | 0.727595098 | 65.34391 | 0 | 46.1178277 | 48.969983 | 46.1178277 | 48.9699826 |

We will be focusing mainly on the R Square value, Intercept, and the Coefficients of independent variables for our prediction.

**R Square** value of **0.31** which is also known as the coefficient of determination, as it explains how good the regression model is. 0.31 R Square value is moderate and tells that there is a slight effect on the dependent variable.

31% of the R Square value means 31% of Laptop Prices can be predicted by our independent variables.

So, the general **Regression equation** for our data is,

**Price=259.032 + 46.730*(Screen Size (Inches)) + 46.581*(Battery Life (Hours)) + 11.223*(RAM GB) + 46.557*(Processor Speeds (GHz) +19.06*(Integrated Wireless?) + 0.39*(HD Size (GB)) + 47.544*(Bundled Applications?)**

From the Regression Equation, we can interpret that,

- Price is determined by Screen Size (Inches) with an increase in the factor of 46.730
- Price is determined by Battery Life (Hours) with an increase in the factor of 46.581
- Price is determined by RAM (GB) with an increase in the factor of 11.223
- Price is determined by Processor Speeds (GHz) with an increase in the factor of 46.557
- Price is determined by Integrated Wireless with an increase in the factor of 19.06
- Price is determined by HD Size (GB) with an increase in the factor of 0.39
- Price is determined by Bundled Applications with an increase in the factor of 47.544

  Since we have Categorical Variables in our dataset, this regression equation will change according to the values of these categorical variables. Since, Yes =1, No =0 based on the dummy values present.

If the value of Integrated Wireless is Yes and Bundled applications is Yes. Then the Regression equation will change to

**Price=259.032 + 46.730*(Screen Size (Inches)) + 46.581*(Battery Life (Hours)) + 11.223*(RAM GB) + 46.557*(Processor Speeds (GHz) +19.06 + 0.39*(HD Size (GB)) + 47.544**

If the value of Integrated Wireless is Yes and Bundled applications is No. Then the Regression equation will change to

**Price=259.032 + 46.730*(Screen Size (Inches)) + 46.581*(Battery Life (Hours)) + 11.223*(RAM GB) + 46.557*(Processor Speeds (GHz) +19.06 + 0.39*(HD Size (GB))**

If Integrated Wireless is No and Bundled applications is Yes. Then the Regression equation is given by

**Price=259.032 + 46.730*(Screen Size (Inches)) + 46.581*(Battery Life (Hours)) + 11.223*(RAM GB) + 46.557*(Processor Speeds (GHz) + 0.39*(HD Size (GB)) + 47.544**

If Integrated Wireless is No and Bundled applications is No. Then the Regression equation is given by

**Price=259.032 + 46.730*(Screen Size (Inches)) + 46.581*(Battery Life (Hours)) + 11.223*(RAM GB) + 46.557*(Processor Speeds (GHz) + 0.39*(HD Size (GB))**

Now we will check how close is the predicted price from the regression equation to the value from the dataset price.

For a laptop, Screen Size (Inches) = 15, Battery Life (Hours) = 5, RAM (GB) = 8, Processor Speeds (GHz) = 2.4, Integrated Wireless? = Yes, HD Size (GB) = 120, Bundled Applications? = No

The Price of the dataset is **1490.**

| Configu | Screen | Battery | RAM (G | Process | Integra | HD Size | Bundle | Price |
|---|---|---|---|---|---|---|---|---|
| 230 | 15 | 5 | 8 | 2.4 | Yes | 120 | No | 1490 |

The Price predicted using the Regression Equation is **1459.543**

**Price=259.032 + 46.730*(15) + 46.581*(5) + 11.223*(8) + 46.557*(2.4) +19.06(1) + 0.39*(120) = 1459.543**

We can see that the price from the dataset and the predicted price are slightly different but do not have a big difference. Because of our low R Square value, we are getting the difference in actual and predicted prices.

Now, let's predict the laptop price for Screen Size (Inches) = 18, Battery Life (Hours) = 8, RAM (GB) = 16, Processor Speeds (GHz) = 2.4, Integrated Wireless? = Yes, HD Size (GB) = 360, Bundled Applications? = Yes

**Price=259.032 + 46.730*(18) + 46.581*(8) + 11.223*(16) + 46.557*(2.4) +19.06*(1) + 0.39*(360) + 47.544*(1) = 1968.936**

We have predicted that the laptop price for the above specifications would be around **1968.936**

## Discussion

Based on the ANOVA results, we can clearly see that there was a difference between the laptop prices based on the presence of Bundle Applications present in it. It clearly tells us that the Bundled Applications play an important role in determining the laptop price.

Based on the Regression results, we are easily able to predict the laptop prices from the model we built based on the configuration and specifications we need in our laptop. The Linear Regression model was moderate in predicting the laptop prices. There was a certain relationship between the independent variables and dependent variables by which we were able to predict the laptop prices. More data preprocessing could help in making a better model which would help in predicting laptop prices more accurately.

## Limitations

There are certain limitations to this dataset. The variables used are very less to determine the laptop price. There are other variables like weight, graphics, brand, operating system used, and touchscreen options that might also help in considering the laptop price. Therefore, only some basic variables might not alone help in predicting the laptop price.