Summary

Increase in the volume of laptops in recent years has given way to an increased competition among the major laptop manufacturing brands. American and Asian brands have been the leaders in manufacturing laptops. In order to have a competitive edge over one another it is important to analyze and determine the factors which are responsible for the pricing of the laptop. Understanding these specifications and their relations will help in understanding the bargain powers, which not only the brand but also the local electronic retailers have, while displaying their laptops on a market place such as skroutz or Amazon.

Research Questions

- What are the factors affecting the price of a laptop?
- Is there a price difference for laptops of Asian / American brands?
- Is there a difference in revenue of Asian / American Laptop brands?
- How do laptop specifications vary across Asian / American Laptops?
- Can brand size dramatically influence the price of a laptop?

Note: All the results are in terms of **log of price** as this transformed variable is the dependent variable for the entire analysis.

Approach

The steps followed for the analysis are listed below in the order of execution.

Data Extractions: Data for 1303 models of laptop for different brands was extracted from Kaggle. The other dataset for brand size was extracted from Kenoma.com.

Data Exploration: Both the datasets were thoroughly explored for important variable selection and for merging. The exploration part mostly required eyeballing the dataset to find out important information for the analysis.

Data merging: The two datasets were merged using 'Brand Name' as a primary key. This resulted in addition two new columns namely 'Brand Size and 'Origin' in the original dataset.

Data Cleaning: The data I obtained was clean and did not have any NULL or NA values. Some variables required data type conversion which has been explained in the 'Data Cleaning' section of the report with a detailed explanation.

Data Wrangling: One of the important steps in the project was the data wrangling part. This was done in order to simplify the variables and convert them into a smaller number of levels as some variables had more than 20 levels. This transformation helped us in a better decision-making process by analyzing the simplified data.
Formulation of research Question: After obtaining the cleaned data I devised a set of research questions which have been provided in the above section. These research questions were aligned with the business problems to provide useful insights for the same.

Variable Selection: This step involved selecting the important variables for univariate and bivariate analysis and for selecting the dependent variables for the modelling (Price was selected as the independent variable for regressing modeling).

Data analysis and Hypothesis testing: After selecting the important variables, univariate and bivariate analysis was done for some of the important variables. Hypothesis regarding the specifications, origin and brand sizes were used to analyze the data further. Liner and logistic regressions were also implemented as a part of this step.

Gathering Results: The above processes resulted in some useful insights which I used to provide recommendation to the retailers and the laptop manufacturing brands.

Identifying challenges and Limitations: This process involves identifying the challenges I faced while implementing the result and the limitation I had with respect to the dataset and the analysis.

Data Characteristics

The primary dataset for the laptops was obtained from Kaggle.com. The dataset has 1303 rows and 13 variables. This dataset was originally scrapped from a website skroutz.gr. The data was scraped in 2018 and hence all the information regarding the laptops specs which were on the catalogue for the website in 2018. The important variables in the dataset have been listed below along with their description:

Variable Name	**Description**	**Dependent/ Independent Variable**
Company_ID	Unique identifier for rows	Not Used
Company	Laptop manufacturing company name	Not Used
Product	Model of the laptop	Not Used
TypeName	Type of laptop	Independent
Inches	Screen size in inches	Independent
ScreenResolution	Pixel resolution of screen	Independent
CPU	Processor information	Independent
Ram	RAM information (in GB)	Independent
Memory	Amount of memory (in GB)	Independent
GPU	Graphic card information	Independent
OpSys	Operating system installed	Independent
Weight	Weight of the laptop	Independent
Price_euros	Price of a laptop	**Dependent**

Table-1.0

The second dataset has been obtained from Knoema.com and comprises of data for top 100 brands in the year 2018 along with their rank and brand size. This data was merged with the primary dataset to obtain the revenues of the brands and the manufacturing origin for the laptop, increasing the number of variables to 15.

Link to Laptop Data:
https://www.kaggle.com/ionaskel/laptop-prices

Link to Brand size Data:
https://knoema.com/ynmtjnc/the-world-s-most-valuablebrands

Link to the marketplace website:
https://www.skroutz.gr/c/25/laptop.html?page=1

Data Cleaning and Data Wrangling

Some variables were converted from character to a factor data type for the analysis.

Variables converted from character to Factor are given below:

Company Name
Type of Laptop
Screen Resolution
Memory
RAM
CPU
Memory Type
Operating System

Variable converted from character to numeric are given below:

Weights
CPU clock
Price

Data wrangling was done in order to simplify the data and use it for the analysis. The variables which were either changed in term of reduction in the number of levels or truncated, keeping only the brand information and removing all the technical specifications.

Variable	Conversion	Reason
Company	Converted to Asian and American brands	Had too many factors levels but very few rows in some categories.
Type Name	Combined Desktops and netbooks	Very few entries for these two categories. Moreover, these are not laptops.
Screen Resolution	Divided the data into 2 parts: 1920*1080 and other	The dataset had most entries for 1920*1080 as most laptop come up with this as a preferred resolution.
CPU	Kept information regarding the brand name	This can be useful in determining the preferred brand for laptop companies. Different models of the CPU made it difficult to compare the laptop.
RAM	Converted the values to more than 8 GB and less than 8 GB	Most laptop now a days have 8 GB of RAM, so I took that as a standard.
GPU	Kept information regarding the brand name	This can be useful in determining the preferred brand for Graphic Cards.
Memory	Converted the values to more than 500 GB and less than 500 GB	I took 500GB as a standard and divided the variable into 2 levels.
Operating System	Combined all the windows OS into one category and all the Macs into one	One unified category of all the windows and mac OS can give better insights about the brand being used.

Table-1.1

The data wrangling and data cleaning resulted in the final data frame with 15 variables which was used for further analysis.

Univariate Analysis

Univariate analysis was done on few of the important variables in the dataset such as Price of a laptop (Dependent Variable), Type name and laptop manufacturing company's origin.

Distribution of Price (DV):

Price had a right skewed distribution initially. The skewness is mostly because of the increased number of notebooks in the dataset which are priced at a lower price. I obtained a normalized graph for the price distribution after log transformation, which would be used as a dependent variable for further analysis.

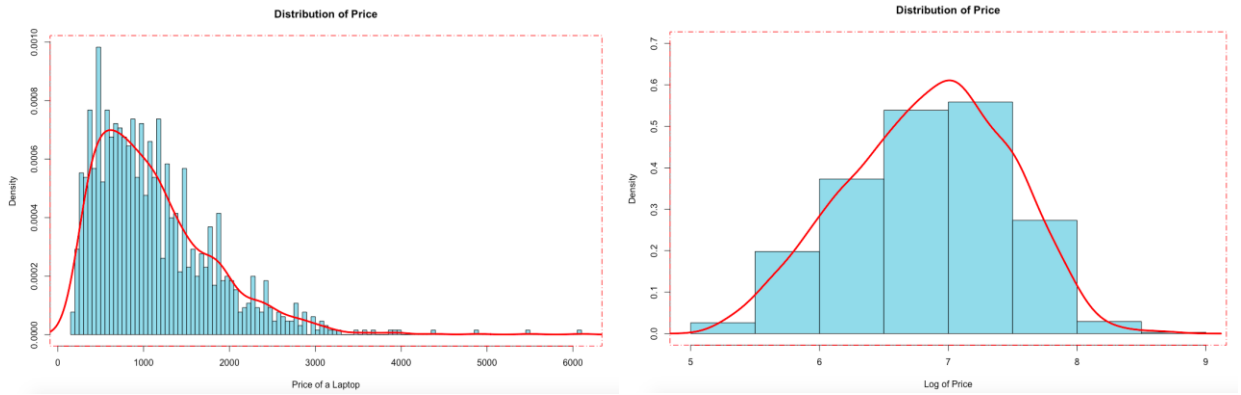


Fig 1.0 Price in Euros and Log(Price in Euros)

Distribution of type of laptops in the market:

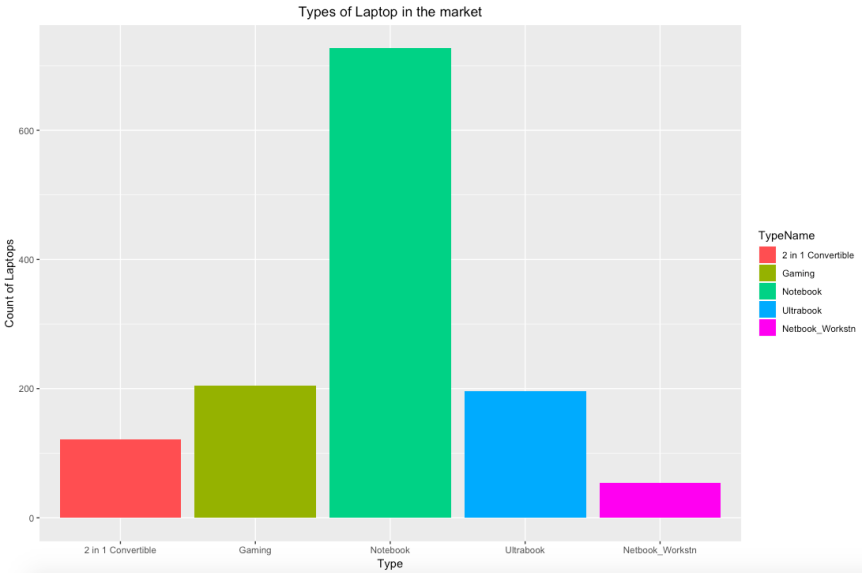


Fig 1.1 Distribution of laptop types

The laptops are categorized in to 5 categories. 2 in 1 convertibles, Gaming Laptops, Notebooks, Ultra Books and Net Books. I have combined workstations and Netbooks as they had very few entries and because workstation and net book are a class of electronics which are not exactly laptops and hence, I decided to drop these levels from the analysis. Notebook are more in number

compared to all the others may be because they are priced at a lower price and serve the same purpose as that of the laptops. Gaming and Ultrabook are almost same in number with a smaller number of 2 in 1 laptop.

Distribution of Operating system across Laptops:

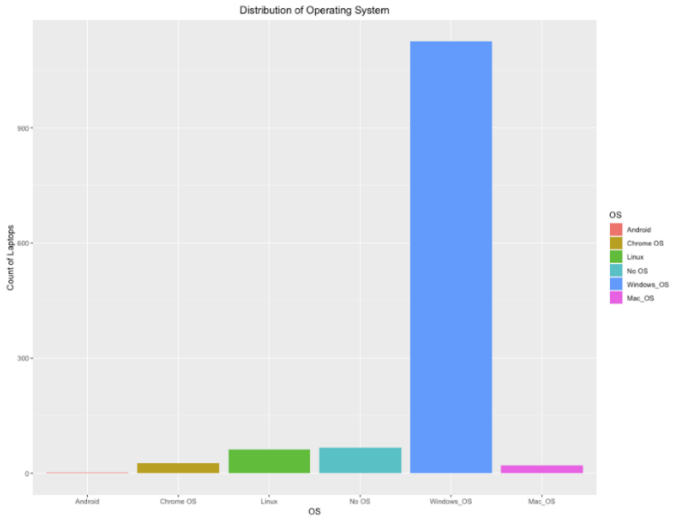


Fig 1.2 Distribution of Operating Systems

More than 90% of the laptop brands have windows as their preferred operating systems with only a few records for MAC OS, Linux. Few laptops also come with no operating system and ultimately installing operating system on these laptops would increase the overall price for the customer.

Distribution of laptops on the website's catalogue based on origin:

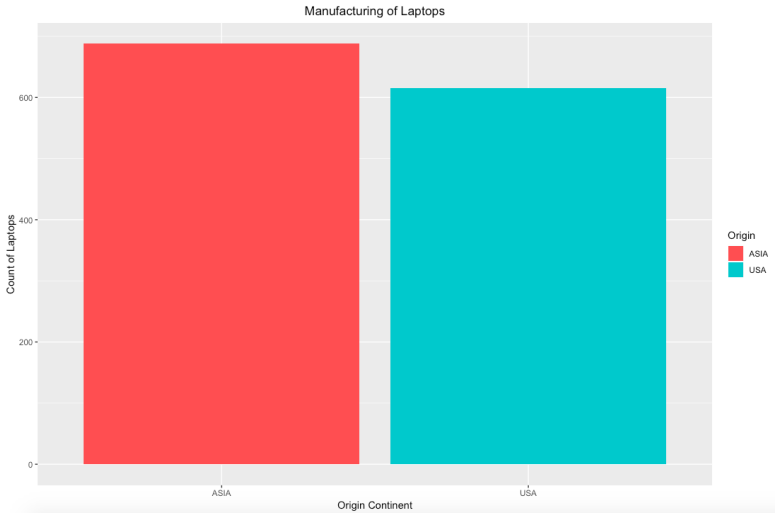


Fig 1.3 Laptop brands from Asia vs the USA

Laptops manufactured by both Asian and USA are similar in terms of the quantities displayed on the website will a slightly greater number of Laptops for Asian brand. The Asian brands have 688 models of the laptop whereas USA brands have 615 models of laptops on the catalogue.

Highlights of Univariate analysis:

- Price was right skewed for which a log transformation is done.
- Slightly a greater number of laptops for Asian brands than USA brands
- More number of Notebook than any other type of laptops.
- Almost all laptops have windows as their preferred operating system

Bivariate analysis

Once the variables were analyzed at a univariate level, I analyzed the relationship across important variables in the dataset and obtained the following results.

Log of price across brand origin:

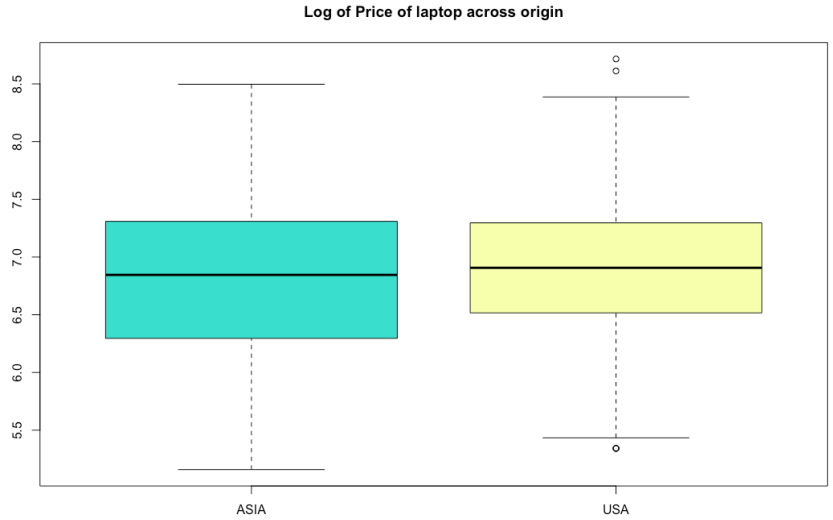


Fig 1.4 Log Price in Euros of Laptop in Asia vs USA

The log of price across US and Asian brands is almost similar with a slightly higher median for USA laptops although the variation is slightly higher for the log of price of Asian brands as compared to US brands. The high variation could be a result of huge amount of gaming laptops manufactured by Asian brands.

The type of laptops manufactured in each origin are given below:

Brand/Type	2 in 1 convertible	Gaming	Notebook	Ultrabook
ASIA	71	148	372	79
USA	50	57	355	117

Table 1.2

Distribution of memory across category of laptop:

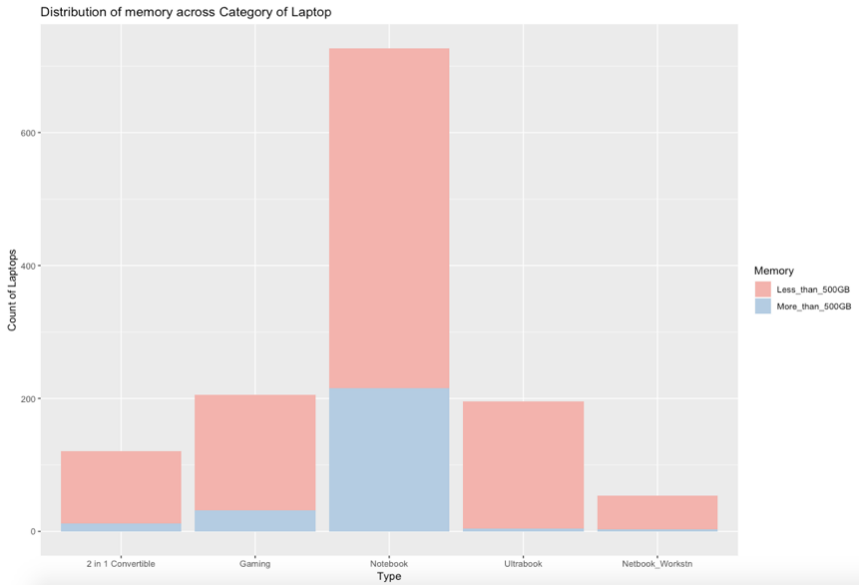


Fig 1.5 Distribution of memory across laptop types

Most of the laptops on the catalog are less than 500 GB. More than 95% of the Ultrabook and Net books have low power processors and hence their memory capacity is also less. Notebook, Gaming and 2 in 1 convertible category have some proportion of laptops which have memory capacity of more than 500 GB. The distribution of memory across type of laptops is given below:

Memory/Type	2 in 1 convertible	Gaming	Notebook	Ultrabook
More than 500 GB	8.4%	13.3%	39.3%	14.7%
Less than 500 GB	0.9%	2.46%	16.5%	0.23%

Table 1.3

Brand comparison for CPU and CPU:

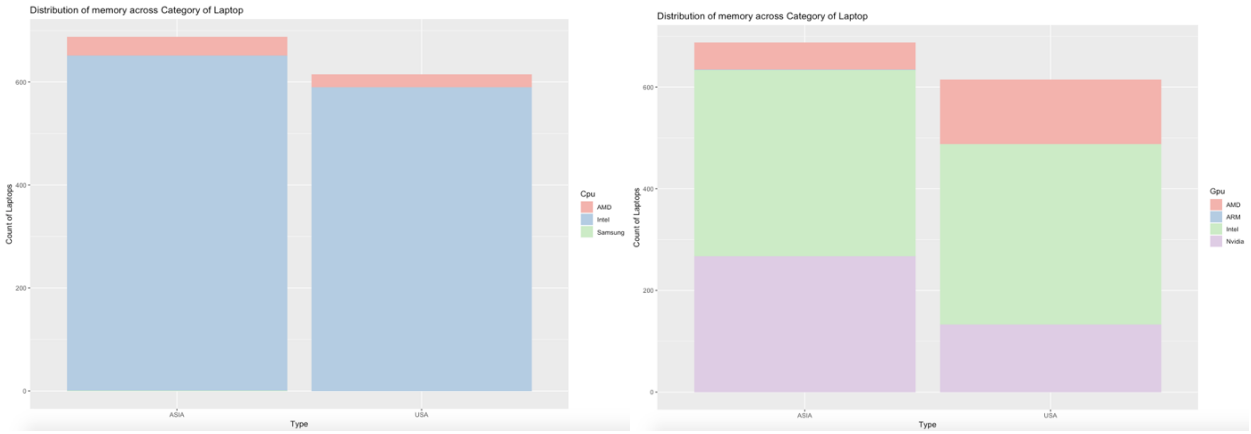


Fig 1.6 Distribution of memory across laptop types

The above graphs indicate that in terms of processor and memory, Intel is the most preferred brand irrespective of the origin of the laptop brand. A very small number of laptops have AMD processors. In terms of graphic card, Asian-brand laptops have Nvidia and Intel as their preferred brand. The American brand still have Intel as their major graphic card partners.
Processor companies for laptops:

Origin/Processor	AMD	ARM	Intel	Nvidia
Asia	53	1	367	267
USA	127	0	355	133

Table 1.4

Graphic card companies for laptops:

Origin/Graphic card	AMD	Intel	Samsung
Asia	37	650	1
USA	25	590	0

Table 1.5

Highlights of Bivariate analysis:

- Log of price is almost similar for US and Asian brand laptops with only a slight difference in median.
- More than 50% of the laptops on the catalogue have less than 500 GB of memory. Notebooks have the highest number of laptops with more than 500 GB of memory.
- Intel is the industry leader for CPU as well as GPU.
- In Asia, Nvidia holds high market share.
- In USA, Nvidia and AMD hold similar market share.

Predicting Laptop Prices (OLS Regression)

After performing the bivariate analyses on different pairs of possibly related variables, the next step was to build a Linear Regression model that predicts the price of a laptop, given the numerous characteristics of the laptop. The variables which were statistically significant are listed below.

The following conclusions can be drawn based on the results of regression for the (log) price of a laptop. On an **average,**

- Compared to Notebooks, the log price of Two-in-one convertible model laptop was 0.27 times lower.
- Compared to Flash memory, HDD memory increased the log of laptop price by 0.222 times and SSD increased the same by 0.476 times.
- Compared to AMD chip, Intel laptops were priced 0.411 times higher.
- Compared to the laptops with 8GB RAM, the laptops with a RAM capacity higher or lower than 8GB were priced approximately 0.3 times higher or lower, respectively.
- For every 1 GHz increase in the CPU Clock Speed, the laptop price went up 0.3 times.

```
> summary(model2)

Call:
lm(formula = log(Price_euros) ~ TypeName + Cpu + CpuClockSpeed +
    Ram + Weight + MemoryType, data = laptop)

Residuals:
     Min       1Q   Median       3Q      Max
-1.35876 -0.21344 -0.01396  0.19916  1.09880

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)               5.524494   0.082215  67.196  < 2e-16 ***
TypeNameGaming           -0.186859   0.045884  -4.072 4.94e-05 ***
TypeNameNetbook_Workstn -0.002389   0.053207  -0.045 0.964189
TypeNameNotebook         -0.297731   0.034038  -8.747  < 2e-16 ***
TypeNameUltrabook         0.127529   0.037167   3.431 0.000620 ***
CpuIntel                  0.418119   0.043300   9.656  < 2e-16 ***
CpuSamsung                0.682120   0.322080   2.118 0.034379 *
CpuClockSpeed             0.303246   0.021308  14.232  < 2e-16 ***
RamLess_than_8GB         -0.397278   0.023693 -16.768  < 2e-16 ***
RamMore_than_8GB          0.339404   0.027054  12.545  < 2e-16 ***
Weight                    0.065127   0.020425   3.189 0.001464 **
MemoryTypeHDD             0.170652   0.047813   3.569 0.000371 ***
MemoryTypeHybrid          0.351636   0.110845   3.172 0.001548 **
MemoryTypeSSD             0.435335   0.046335   9.395  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3158 on 1282 degrees of freedom
Multiple R-squared:  0.7411,    Adjusted R-squared:  0.7385
F-statistic: 282.3 on 13 and 1282 DF,  p-value: < 2.2e-16
```

Fig 1.7 Summary of OLS model for Log(Price_euros)

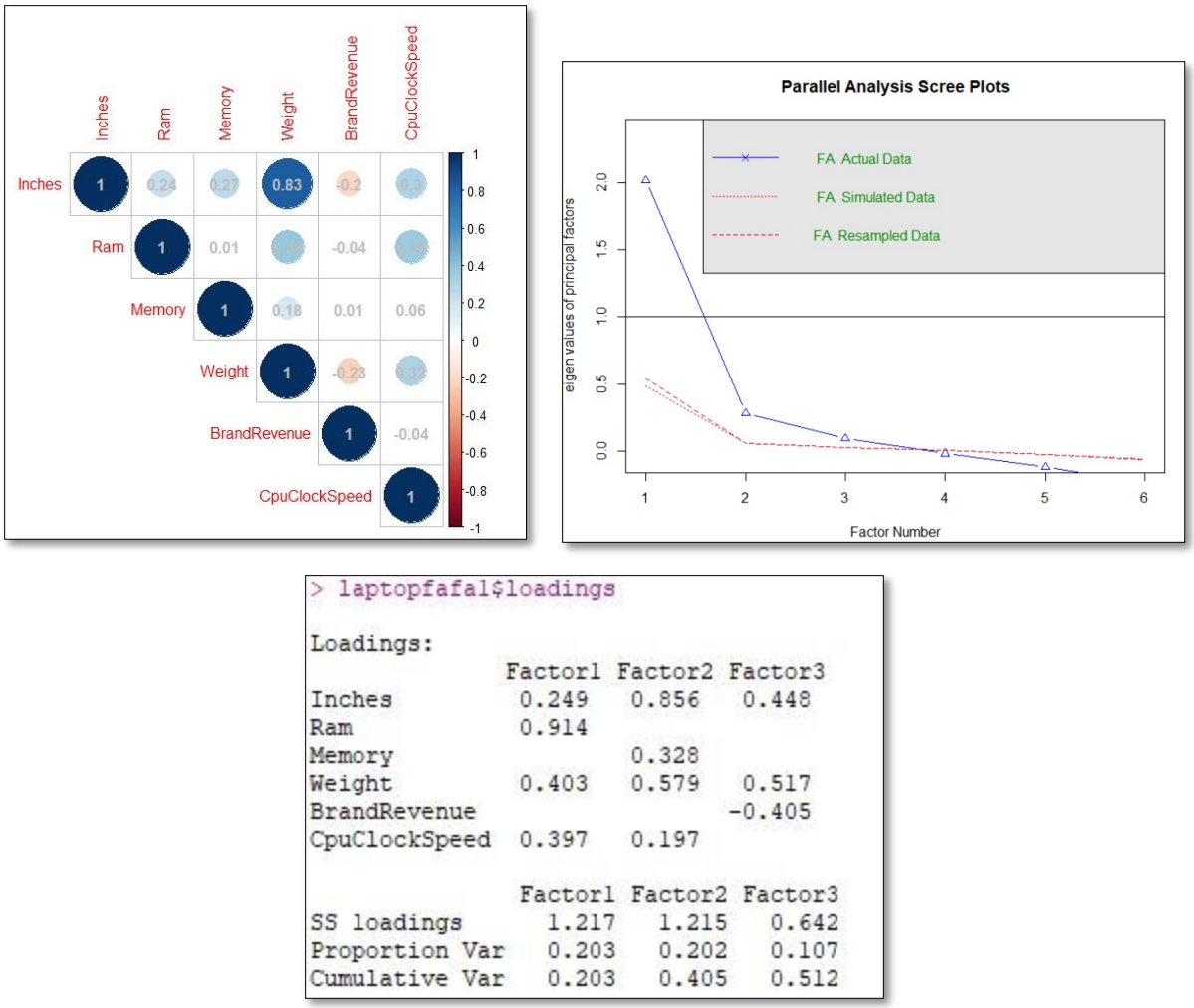
The model had an adjusted R squared value of 0.74. There were no significant interactions found in the model. The VIF values of the model coefficients were within the limits and the diagnostic plots were normal as well. Further, the robustness of the model was verified by predicting the results on the dataset. The Mean Squared Error value was found to be 0.098 which indicated a good model.

Performing Factor Analysis

For the numerical variables in the data, there was a possibility of obtaining meaningful factors from the dataset. For this reason, Factor Analysis was performed on the set of following numerical variables.

1. Inches
2. RAM
3. Memory
4. Weight
5. CPU Clock speed

The correlation plot showed a high correlation of 0.83 between the Inches and Weight of a laptop. While performing the factor analysis, the ideal number of factors was obtained as 3, as shown below. However, the factor-1 did not make sense, as it was a combination of RAM, Weight and CPU clock speed. The factor-2 was meaningful, as it indicated a factor that represents the Physical Dimensions of the laptop, namely Inches and Weight. The factor-3 was a repetition of factor-2 and was ignored.




```
> laptopfafa1$loadings

Loadings:
             Factor1 Factor2 Factor3
Inches         0.249   0.856   0.448
Ram            0.914
Memory                 0.328
Weight         0.403   0.579   0.517
BrandRevenue                  -0.405
CpuClockSpeed  0.397   0.197

               Factor1 Factor2 Factor3
SS loadings      1.217   1.215   0.642
Proportion Var   0.203   0.202   0.107
Cumulative Var   0.203   0.405   0.512
```

Fig 1.8 Correlation plot, Scree plot and Factor Analysis summary

Analyzing difference in Laptop Prices by Origin

The hypothesis assumed at the beginning of the analysis was that the laptops belonging to American brands are priced higher than the laptops manufactured by their Asian Counterparts. The Place of origin of the laptops was set to either America or Asia. This was done by pooling together the laptop brands scattered over China, Japan, Korea and so on into one Asian category.

The ANOVA test between the Origin variable and the Price of a laptop yielded a marginally significant relationship. In order to reinforce the validity of the finding, a Logistic Regression model was considered across the same variables. The results were statistically significant and made it clear that on an average, the (log) price of a laptop belonging to an American origin was 1.10 times more than that of an Asian origin. The coefficient corresponding to the odds was exponentiated to find out the relation between the two variables.

```
> summary(price_origin.aov)
                Df    Sum Sq Mean Sq F value Pr(>F)
laptop$Origin    1   1772884 1772884   3.647 0.0564 .
Residuals     1294 629093472  486162
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> model4<-lm(log(Price_euros) ~ Origin, data=laptop)
> summary(model4)

Call:
lm(formula = log(Price_euros) ~ Origin, data = laptop)

Residuals:
     Min       1Q   Median       3Q      Max
-1.64176 -0.43473  0.03105  0.44897  1.81920

Coefficients:
            Estimate Std. Error t value             Pr(>|t|)
(Intercept)  6.80082    0.02360 288.117 < 0.0000000000000002 ***
OriginUSA    0.09586    0.03427   2.798              0.00522 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.616 on 1294 degrees of freedom
Multiple R-squared:  0.006012,  Adjusted R-squared:  0.005244
F-statistic: 7.827 on 1 and 1294 DF,  p-value: 0.005223

> exp(coef(model4))
(Intercept)   OriginUSA
  898.58046     1.10061
```

Fig 1.9 ANOVA test and OLS model for Log(Price_euros) vs Origin

Analyzing laptop specification by origin (Logistic Regression)

It was posited that the laptops belonging to Asian countries had a certain set of characteristic specifications, which were different from their American counterparts. This led to the Research question – How do the laptop specification trends differ across Asian and American manufactured laptops? A detailed bivariate analysis was carried out between the Origin location variable and each of the specification variables to verify this trend. Chi Squared test was carried out between the numeric variables and Origin, while ANOVA test was performed for the categorical variables.

```
> chisq.test(laptop$Origin,laptop$MemoryType)

        Pearson's Chi-squared test

data:  laptop$Origin and laptop$MemoryType
X-squared = 8.9644, df = 3, p-value = 0.02977

Warning message:
In chisq.test(laptop$Origin, laptop$MemoryType) :
  Chi-squared approximation may be incorrect
> chisq.test(laptop$Origin,laptop$TypeName)

        Pearson's Chi-squared test

data:  laptop$Origin and laptop$TypeName
X-squared = 54.326, df = 4, p-value = 4.498e-11

> chisq.test(laptop$Origin,laptop$Ram)

        Pearson's Chi-squared test

data:  laptop$Origin and laptop$Ram
X-squared = 13.918, df = 2, p-value = 0.0009502

> ##Memory is not significant
> chisq.test(laptop$Origin,laptop$Memory)

        Pearson's Chi-squared test with Yates' continuity correction

data:  laptop$Origin and laptop$Memory
X-squared = 0.52749, df = 1, p-value = 0.4677
```

```
> aov_price <- aov(log(laptop$Price_euros)~laptop$Origin, data=laptop)
> summary(aov_price)
                Df Sum Sq Mean Sq F value  Pr(>F)
laptop$Origin    1      3  2.9698   7.827 0.00522 **
Residuals     1294    491  0.3794
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> aov_clockspeed <- aov(laptop$CpuClockSpeed~laptop$Origin, data=laptop)
> summary(aov_clockspeed)
                Df Sum Sq Mean Sq F value Pr(>F)
laptop$Origin    1    1.1  1.0638    4.23 0.0399 *
Residuals     1294  325.4  0.2515
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> aov_weight <- aov(laptop$Weight~laptop$Origin, data=laptop)
> summary(aov_weight)
                Df Sum Sq Mean Sq F value  Pr(>F)
laptop$Origin    1    3.0  2.9683   6.725 0.00961 **
Residuals     1294  571.1  0.4414
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> aov_inches <- aov(laptop$Inches~laptop$Origin, data=laptop)
> summary(aov_inches)
                Df Sum Sq Mean Sq F value Pr(>F)
laptop$Origin    1     10  10.009   4.935 0.0265 *
Residuals     1294   2625   2.028
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Fig 1.10 Chi Square test and ANOVA test performed

With the statistically significant variables, a Logistic regression was also done in order to reinforce the findings of the statistical tests. The exponentiated results of the variables are provided in the below table. From the value of coefficients, it could be said that the Asian origin brands manufactured predominantly Workstations and Netbook category of laptops and clearly lacked products in the Notebook and Two-in-one convertible sections.

Specs	**Coefficients**
TypeNameGaming	0.4697715
TypeNameNetbook_Workstn	3.42398394
RamMore_than_8GB	0.61097106
log(Price_euros)	2.42277741
MemoryTypeHDD	2.03941661

Table 1.6

```
> origin_mod2_1 <- glm(Origin~TypeName+Ram+log(Price_euros)+MemoryType, data=laptop, family=binomial)
> summary(origin_mod2_1)

Call:
glm(formula = Origin ~ TypeName + Ram + log(Price_euros) + MemoryType,
    family = binomial, data = laptop)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.8013  -1.1168  -0.6923   1.1202   1.8768

Coefficients:
                        Estimate Std. Error z value Pr(>|z|)
(Intercept)             -6.60296    1.17344  -5.627 1.83e-08 ***
TypeNameGaming          -0.75551    0.25472  -2.966  0.00302 **
TypeNameNetbook_Workstn  1.23080    0.37427   3.289  0.00101 **
TypeNameNotebook         0.45635    0.21735   2.100  0.03576 *
TypeNameUltrabook        0.56407    0.24098   2.341  0.01925 *
RamLess_than_8GB        -0.02791    0.16938  -0.165  0.86910
RamMore_than_8GB        -0.49271    0.18856  -2.613  0.00898 **
log(Price_euros)         0.88491    0.17166   5.155 2.54e-07 ***
MemoryTypeHDD            0.71266    0.31174   2.286  0.02225 *
MemoryTypeHybrid         0.54636    0.71424   0.765  0.44430
MemoryTypeSSD            0.08228    0.31788   0.259  0.79576
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1793.3  on 1295  degrees of freedom
Residual deviance: 1685.8  on 1285  degrees of freedom
AIC: 1707.8

Number of Fisher Scoring iterations: 4
```

Fig 1.11 Summary of Logistic Regression model between Origin and all the spec

Analyzing Brand Size and Origin

After The next analysis intended to find out any significant relation between the Origin location of the laptop brand and the size of the company or Brand Size. The hypothesis was that the companies which had larger Brand Size have had their origin in the USA. The companies were divided based on the median value of 50 Billion USD as those less than and greater than the median. The graph comparing the two Origin Locations plotted across the two categories of Brand Size is shown below. It is conspicuous that the companies with a size greater than 50 billion USD are exponentially higher in USA than in Asia.

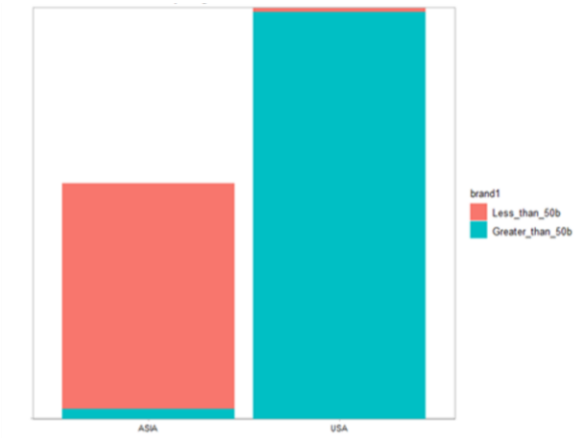


Fig 1.12 Distribution of Brand Size for Origins Asia and USA

The results of a Chi Square Test between the laptop brand size and origin also presented statistically significant results, as shown in Fig 11. The Logistic model, finally, also conformed the same result with statistically significant coefficient corresponding to the Origin variable. On an average, the chances of a company with greater than 50 Billion USD as the brand size to be an American company is 2045 times that of it being an Asian company.

```
> table(laptop$brand1,laptop$Origin)

                ASIA USA
Less_than_50b    667  14
Greater_than_50b  14 601
> prop.table(table(laptop$brand1,laptop$Origin)*100)

                       ASIA        USA
Less_than_50b     0.51466049 0.01080247
Greater_than_50b 0.01080247 0.46373457

> cl <- chisq.test(laptop$brand1,laptop$Origin)
> cl

        Pearson's Chi-squared test with Yates' continuity correction

data:  laptop$brand1 and laptop$Origin
X-squared = 1182.3, df = 1, p-value < 0.0000000000000022
```

```
glm(formula = laptop$brand1 ~ Origin, family = "binomial", data = laptop)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.7505  -0.2038  -0.2038   0.2146   2.7873

Coefficients:
            Estimate Std. Error z value             Pr(>|z|)
(Intercept)  -3.8637     0.2700  -14.31 <0.0000000000000002 ***
OriginUSA     7.6233     0.3821   19.95 <0.0000000000000002 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1793.27  on 1295  degrees of freedom
Residual deviance:  270.07  on 1294  degrees of freedom
AIC: 274.07

Number of Fisher Scoring iterations: 6

> exp(coef(mod1))
  (Intercept)     OriginUSA
   0.02098951 2045.23978301
> mod_fit_3$stats["R2"]
       R2
0.9225031
```

Fig 1.13 Summary of Chi Square Test and Logistic Regression model for Brand Size vs Origin

Analyzing Laptop Price and Brand Size

From the previous analyses, it was found that there was a clear relationship between the Place of Origin of a laptop manufacturer and its Brand Size. The Origin had a humongous influence over the Brand Size. Thus, it was inevitable to quantify the relationship, if any, between the Brand Size and the Price of a laptop. The hypothesis in this case was that there is a relationship between laptop price & Brand Size. To make the analysis more accurate, the companies were segregated into three categories, Low, Mid and High (Less than 30 Billion USD, 30 to 90 30 Billion USD and Greater than 90 30 Billion USD) The comparative boxplot between the Price in Euros of the laptop and the Brand Size is shown below. It is evident that the median price of High Brand Size company laptops is much higher than that of the Low and Mid companies.

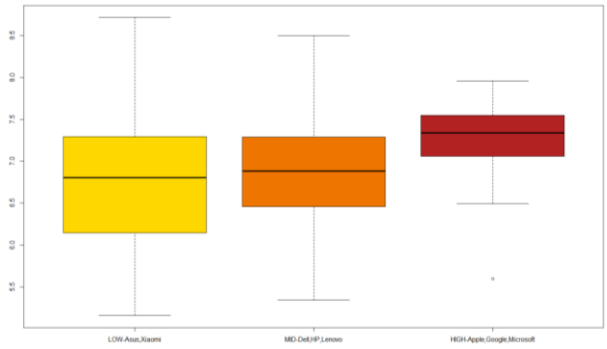


Fig 1.14 Comparative boxplot between log(Price_euros) and Brand Size

To reinforce the findings, ANOVA test was performed between the Brand size and the (log) Price of the laptop. The results were statistically significant. The simple OLS model between the two variables indicated that on an average, the Price of a laptop in Euros is 1.47 times higher for the High-end companies compared to the Low-end companies. The disparity was evident, and one could say that the High-end brands such as Apple, Google, Microsoft price their laptops at a much higher level than the Mid and Low tier companies.

```
> #Performing the ANOVA Test to reinforce the finding
> expense_catalog_model <- aov(log(Price_euros)~ brand2, data=laptop)
> summary(expense_catalog_model)
             Df Sum Sq Mean Sq F value   Pr(>F)
brand2        2   11.8   5.879   15.77 1.72e-07 ***
Residuals   1293  482.2   0.373
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
```

```
> #Performing OLS Regression to reinforce the finding
> model_1 <- lm(log(Price_euros) ~ brand2, data= laptop)
> summary(model_1)

Call:
lm(formula = log(Price_euros) ~ brand2, data = laptop)

Residuals:
     Min       1Q   Median       3Q      Max
-1.70220 -0.42542  0.02617  0.43893  1.96527

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.96929    0.03400 204.979  < 2e-16 ***
brand2.L     0.38629    0.07071   5.463  5.6e-08 ***
brand2.Q     0.13342    0.04401   3.032  0.00248 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6107 on 1293 degrees of freedom
Multiple R-squared:  0.02381,   Adjusted R-squared:  0.0223
F-statistic: 15.77 on 2 and 1293 DF,  p-value: 1.718e-07
```

Fig 1.15 Summary of ANOVA Test and OLS model for log(Price_euros) and Brand Size

Results

- On an average, the (log) price of a laptop belonging to an American brand is 1.10 times more than that of an Asian Brand.
- On an average, the odds of an American brand having revenue >50 billion USD is 2045 times higher than that of Asian brand.
- On an average, the laptops belonging to covetable brands price their laptops 1.47 times more than the lower valued brands.
- The price of a laptop surges when it includes an Intel chip, higher than 8GB RAM, HDD and SSD memory, belongs to the Two-in-one convertible category and has higher CPU Clock speeds.