ISSN: 2230-9926

Available online at http://www.journalijdr.com

IJDR

International Journal of Development Research
Vol. 08, Issue, 07, pp.21724-21732, July, 2018

ORIGINAL RESEARCH ARTICLE

OPEN ACCESS

THE PRICE OF MEMORIES

[1]Yulia Tertytsia and *[2]Paul D. Berger

[1]Dell EMC, Hopkinton, MA 01748, USA
[2]Bentley University, Waltham, MA 02452, USA

ARTICLE INFO

Article History:
Received 11th April, 2018
Received in revised form
19th May, 2018
Accepted 26th June, 2018
Published online 30th July, 2018

Key Words:
Digital-camera pricing;
Multiple regression,
Multilayer Perceptron Neural Networks;
Decision trees.

ABSTRACT

Digital cameras play a significant role in the dynamics of the camera universe. The goal of this study is to predict the price of a digital camera based on a set of attributes, as well as to identify the attributes/ variables that significantly contribute to its price estimation. We obtained our data set from Perso Telecom Paristech. The dependent variable is Price (US dollars); independent variables include Weight, Model (Maker), Zoom Tele, Zoom Wide, Macro Focus Range, Dimensions, Low Resolution, and Effective Pixels. We used Multiple Regression, Multilayer Perceptron Neural Network, and Decision Tree models for analyzing the data. The study revealed that camera Weight, Maker, and Zoom Tele are the most significant attributes in estimating the price of a digital camera.

Citation: Yulia Tertytsia and Paul D. Berger, 2018. "The price of memories", *International Journal of Development Research*, 8, (07), 21724-21732.

INTRODUCTION

It is impossible to imagine the current Western world without a camera. No matter where you go, you always can see people taking pictures. We have cameras in our phones, watches, pens, cars, computers, etc. People love to capture special moments. Who has not spent at least one evening looking through an album remembering the "good-old days," a family reunion two years ago, or the first ultrasound of your very own baby? Images provide us with memories, whether good or bad. Some people spend thousands of dollars to produce such images. In the past 25 years, Sir Elton John has purchased 8000 images for his private photography collection (Nick, 2017). It seems that our society would not be able to function without cameras. But how did it start? Who was that genius that first thought about capturing an image? Johann Zahn designed the first camera in 1685, while Joseph Nicephore clicked the first photograph in 1814. We find the first record of a device that could capture images as recorded by Iraqi scientist, Ibn-al-Haytham, in his Book of Optics in 1021 (Who invented the camera, 2010).

Corresponding author: **Paul D. Berger**
Bentley University, Waltham, MA 02452, USA

Color photography first began in 1940. Eastman-Kodak engineer, Steven Sasson, developed the first recognized digital camera in 1975, being the size of a toaster and weighting nearly 4 kg (almost 9 lbs) (Richard Trenholm, 2007). Compare it to the smallest camera in the world, developed by researchers at the Fraunhofer Institute in Germany (being no larger than a grain of salt with dimensions of 1x1x1 millimeters), and you can begin to understand the dramatic improvement that has transpired in camera development (Peter Parchal, 2011). This paper studies digital cameras; specifically, estimating the price of digital cameras based on their attributes. We found the camera data set at Telecom-Paristech (https://perso.telecom-paristech.fr/eagan/class/igr204/datasets). We wish to thank Petra Isenberg, Pierre Dragicevic, and Yvonne Jansen for aid in collecting and cleaning the dataset.

Objectives of the study

- The objectives of this study are:
- To estimate digital-camera prices based on selected independent variables;
- To investigate specific the role of independent variables in estimating digital camera prices; and

- To analyze the performance of the Multilayer Perceptron Neural Network model and Decision Tree Model in achieving our objectives.

Literature Review

"A Few Thoughts About the Camera Market" written by Thomas Stirr for *Photography Life* highpoints the shifts and trends in the camera market, including digital cameras. Mr. Stirrconducted a statistical analysis of the camera market for the time period of 1965–2015, using data provided by the Camera & Imaging Association (CIPA). In his study, Mr. Stirrdiscusses the fluctuation in camera sales based on its type and certain attributes (Thomas Stirr, 2017). "Your Guide to Digital Photography" not only offers an introduction into the world of digital cameras and their usage, but also provides insights on consumer prices of digital cameras and their shift over the years (https://books.google.com). "Does sensor size matter? Yes!" (2016) unveils reasons why the number of megapixels is actually not a primary factor in determining the quality of an image (Martin Thoma, 2018). Often, consumers pay close attention to the number of megapixels in determining whether to make a purchase. The 'more pixels, the better' reflects a common thought of someone who does not go beyond pressing the shutter button. Additional references are cited throughout the rest of the paper.

Data source and analysis

Variable Description: The camera data set contains n = 1038 observations and 13 variables. The camera models comprise: Agfa, Canon, Casio, Contax, Epson, Fujifilm, HP Photosmart, JVC, Kodak, Kyocera, Leica, Nikon, Olympus, Panasonic, Pentax, Ricoh Caplio, Samsung, Sanyo, Sigma, Sony, and Toshiba. The years of release range from 1994 to 2007. The variables comprise: Maker, Release Date, Max Resolution, Low Resolution, Effective Pixels, Zoom Wide, Zoom Tele, Normal Focus Range, Macro Focus Range, Storage Included, Weight (including batteries), Dimensions, and Price (in US dollars). Appendix 1 provides attribute definitions and measurements. All the variables except Maker are numerical; Maker is a categorical variable. We selected Price as the dependent variable; the rest of the attributes were viewed as independent variables. Initially, we included all the variables in estimating the price of a digital camera. Our final model contains only those predictor variables that provide the best model. Entering the study, we had no specific hypotheses about which independent variables would be significant, although the direction of the impact of the many of the variables, if significant, was clear. There was one exception - we hypothesized that Zoom Wide would not be significant, since in practice, the Zoom function in digital cameras crops the image, resulting in a reduction in the quality of the photos.

Data Analysis Techniques: In order to analyze the camera data set, we used the model building principles as described in, *A Second Course in Statistics Regression Analysis, 7th edition*, by William Mendenhall and Terry Sincich; and also, *Data Mining: Concepts, Models, Methods, and Algorithms, second edition,* by Mehmed Kantardzik (William, 2012; Mehmed, 2011). We used multiple regression analysis, multilayer perceptron neural networks, and decision tree models to analyze the data. Analytical software packages included R Studio and SPSS. We utilized the IBM SPSS Neural Networks package description for the final model interpretation (Output, 2012).

RESULTS AND DISCUSSION

The first model we considered was multiple linear regression

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki} + \epsilon_i.$$

However, when we examine the relationships among the variables, we note two issues that suggest that the linear-regression-analysis approach is not appropriate. One is that the majority of the bivariate relationships between the independent variables and the dependent variable, Price, are not linear. While in certain circumstances it may be true that redefining each independent variable into a set of categorical (dummy) variables might overcome the problem of a lack of linearity, in our study, several of the variables exhibit a degree of non-linearity that was not able to be "rescued" by such a redefinition.

Outlier Analysis: The second issue is that there are a large number of outliers. We used the boxplot command in SPSS to identify outliers. Figures 1 - 11 show the outlier situation for the 11 variables (excluding "Maker" and "Release date.") SPSS uses different markers for "out" values (marked with a small circle) and "far out" or extreme values (marked with an asterisk). Based on Figures 1-11, the camera data has a large number of outliers (Eugene Horber, 2017).

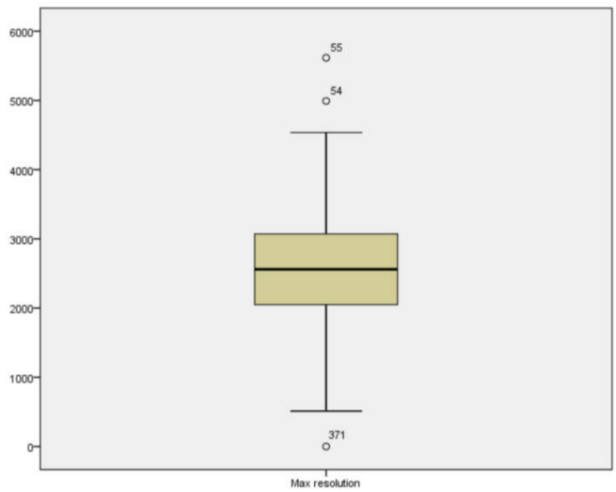

Figure 1. Max resolution outlier analysis

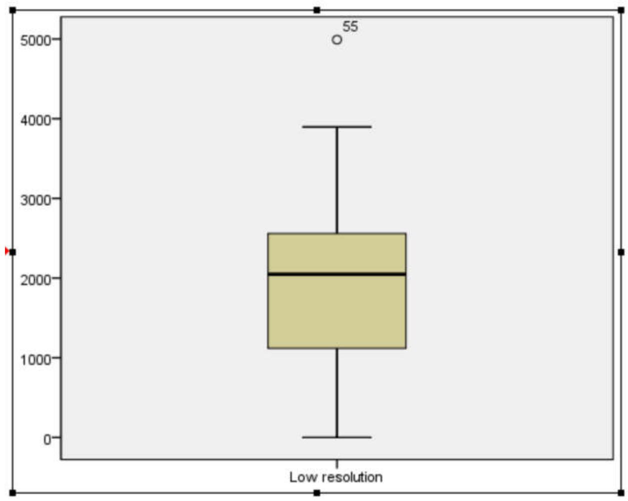

Figure 2. Low resolution outlier analysis

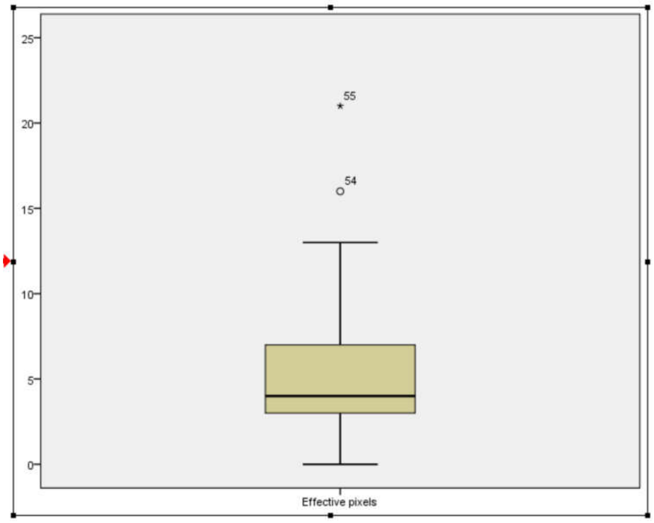

Figure 3. Effective pixels outlier analysis

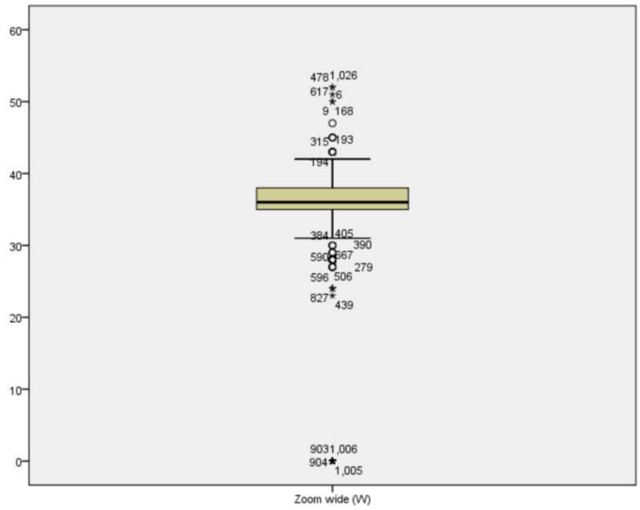

Figure 4. Zoom wide outlier analysis

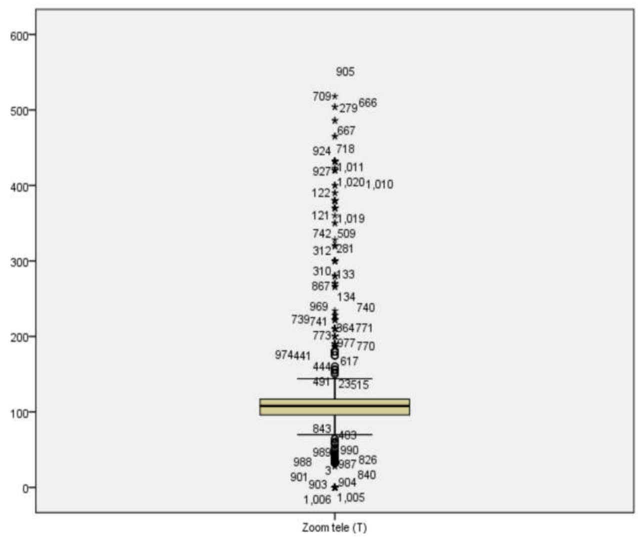

Figure 5. Zoom tele outlier analysis

Non-linearity Analysis: In Figures 12-19, we illustrate the large degree of non-linearity in the relationship between Price and selected independent variables. From the above figures, the relationships between the response variable and several predictors are clearly not linear.

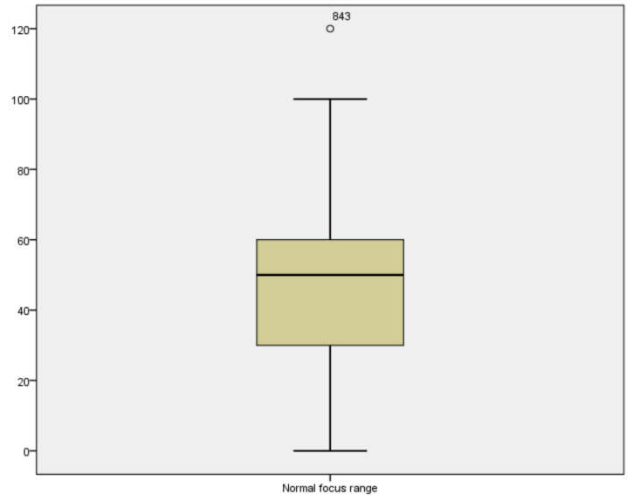

Figure 6. Normal focus range outlier analysis

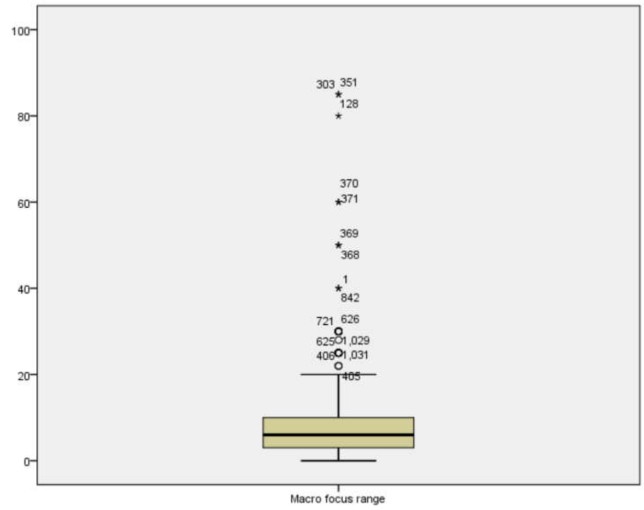

Figure 7. Macro focus range outlier analysis

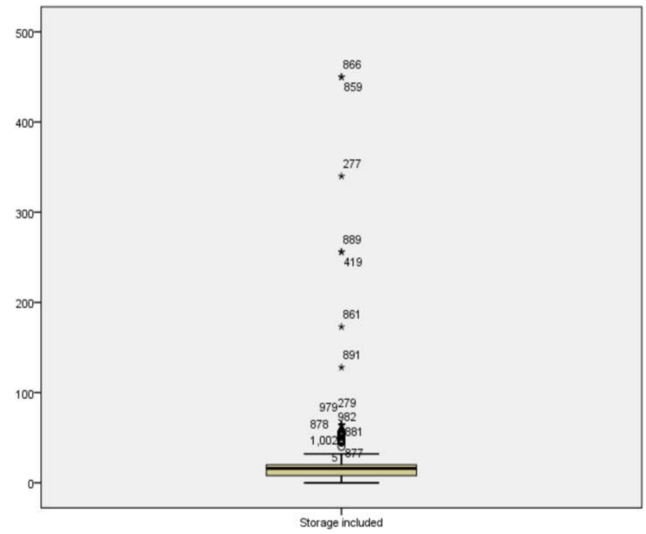

Figure 8: Storage included outlier analysis

Multi-Collinearity Analysis: In addition, we checked for multi-collinearity using the *vifstep* function in R. Results are in Figure 20. Apparently, the predictor variable, Max. resolution, presents multi-collinearity issues. We thus abandoned multiple linear regression as our analysis technique.

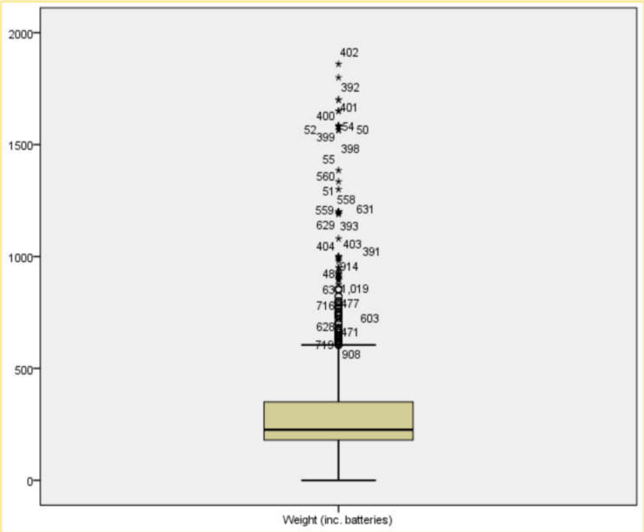

Figure 9. Weight (including batteries) outlier analysis

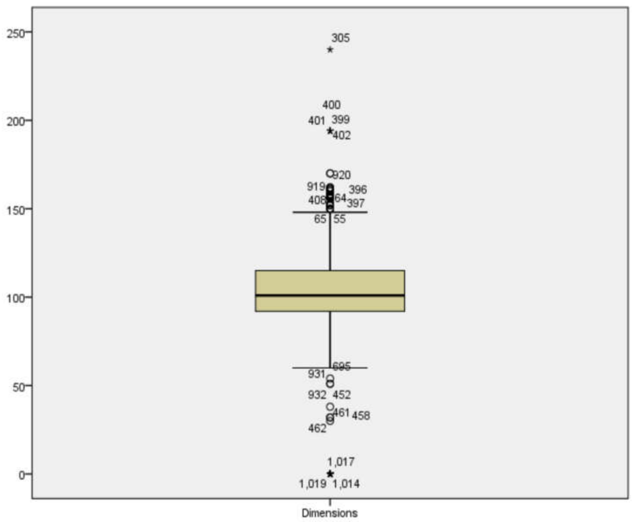

Figure 10. Dimensions outlier analysis

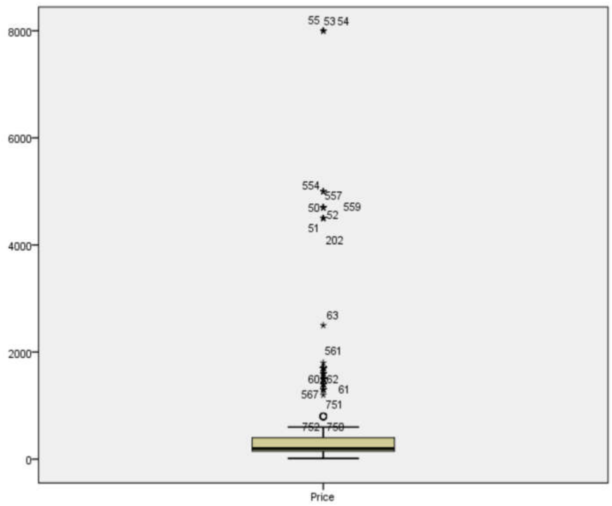

Figure 11. Price outlier analysis

We next examined the relationship between price and release date, as shown in Figure 21. Looking at Figure 21, it can be seen that the mean price of digital cameras spiked in 1996 and again in 2001, and then slowly (and irregularly) decreased after 2001. This result might be related to the change in price per pixel.

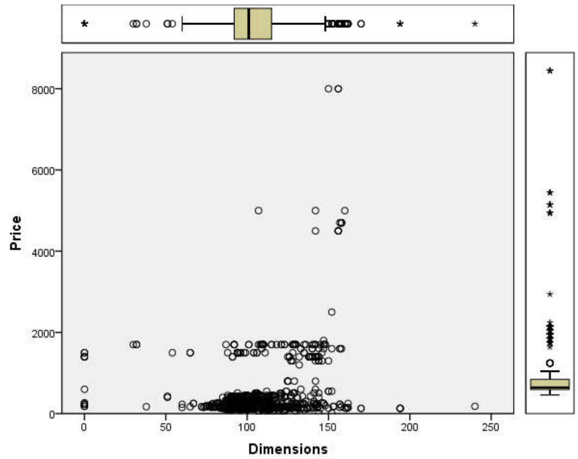

Figure 12. Price vs. dimensions

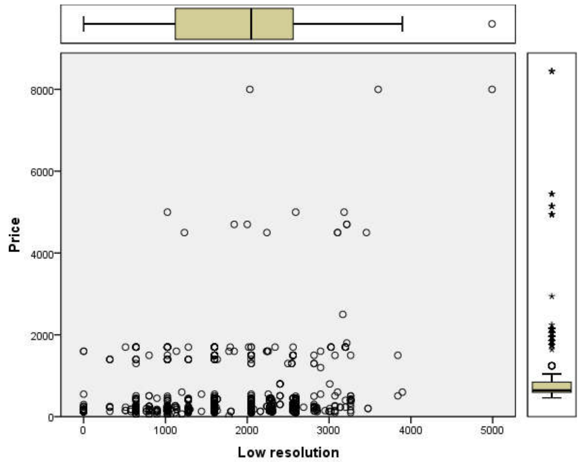

Figure 13. Price vs. low resolution

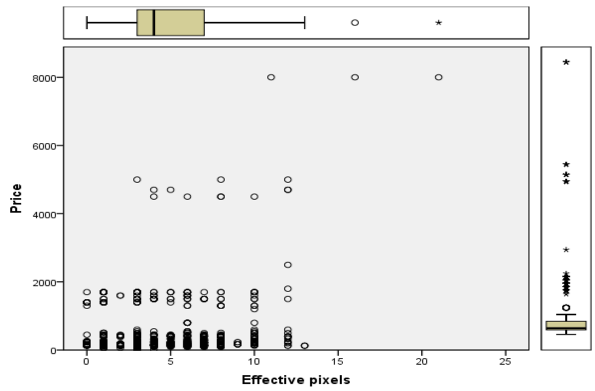

Figure 14. Price vs. effective pixels

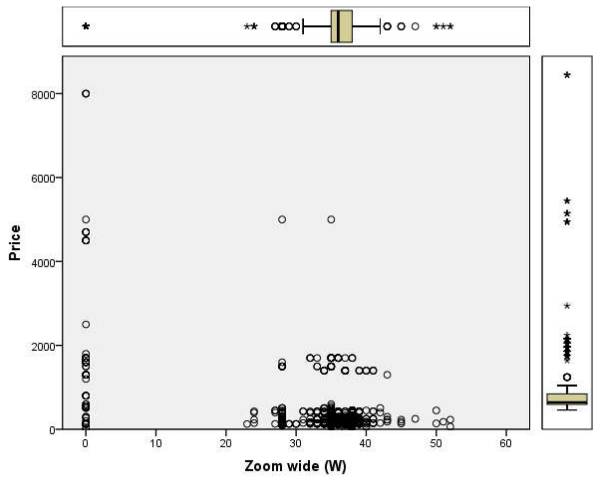

Figure 15. Price vs. zoom wide

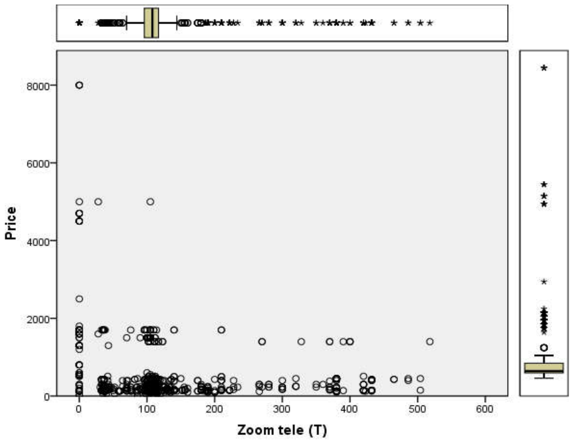

Figure 16. Price vs. zoom tele

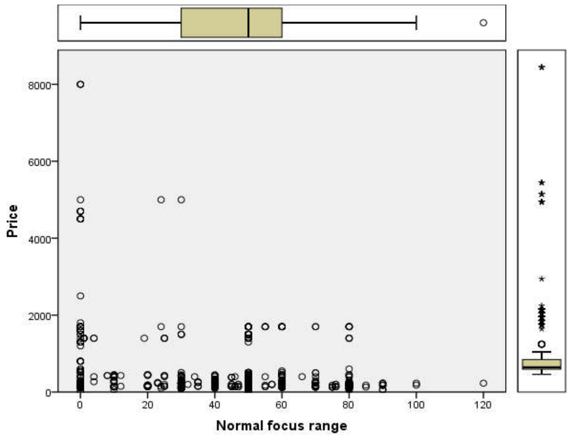

Figure 17. Price vs. normal focus range

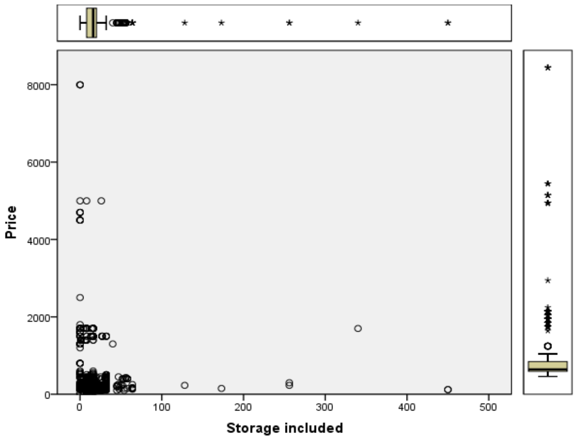

Figure 18: Price vs. storage included

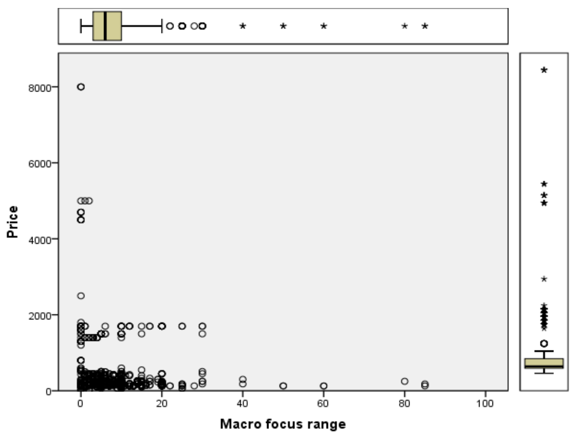

Figure 19. Price vs. macro focus range

```
camera=camera [,-1]
install.packages("usdm")
library ("usdm")
x=cbind (camera[,-1])
vifstep (x,th=10)
```

```
1 variables from the 11 input
variables have collinearity
problem:

Max.resolution
```

Figure 20. Analysis of multi-collinearity

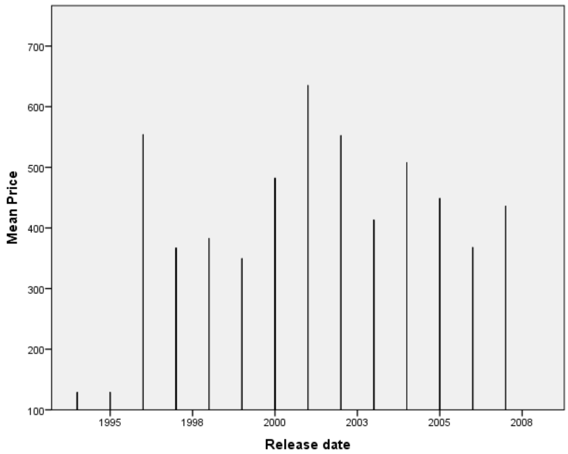

Figure 21. Mean price of digital cameras by release year

Model Summary

Training	Sum of Squares Error	65.888
	Relative Error	.184
	Stopping Rule Used	1 consecutive step(s) with no decrease in error[a]
	Training Time	0:00:00.90
Testing	Sum of Squares Error	38.126
	Relative Error	.170

Dependent Variable: Price

a. Error computations are based on the testing sample.

Figure 22. Model output

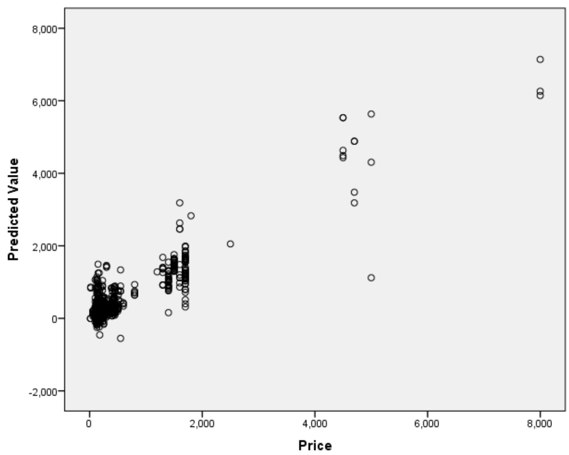

Figure 23. Predicted-by-observed graph

As "Your guide to digital photography" states, in 2006, one could buy a nice 8–10 megapixel digital camera for half the price (about $1000) of a 6.3 megapixel camera just two years prior (https://books.google.com; Mark, 2012).

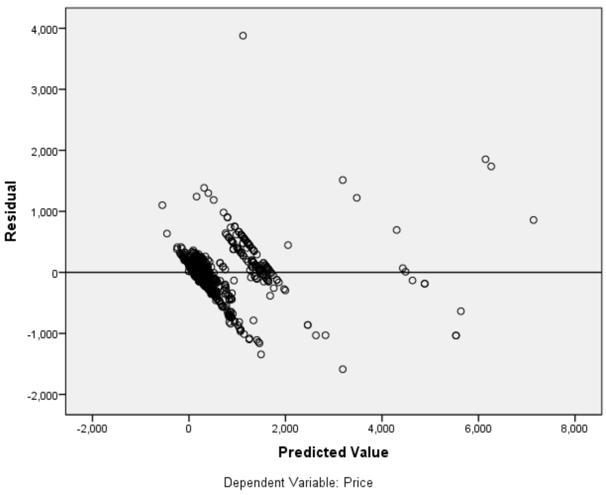

Figure 24. Residual-by-predicted graph

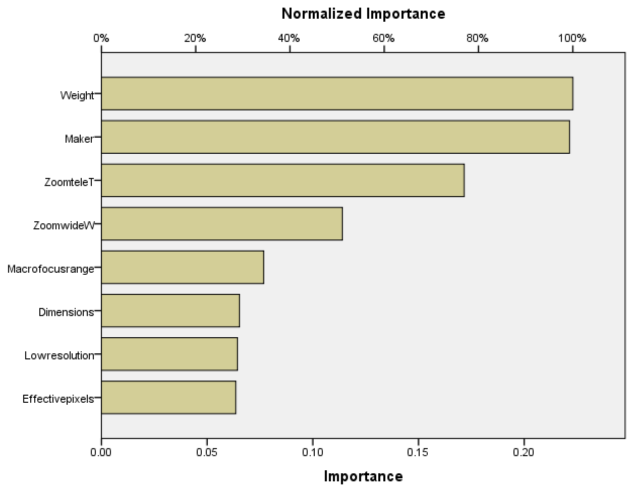

Figure 25. Normalized importance of predicted variables

Table 1. Independent variable importance

Independent Variable Importance

	Importance	Normalized Importance
Maker	.221	99.3%
Zoom wide (W)	.114	51.1%
Zoom tele (T)	.172	76.9%
Macro focus range	.077	34.4%
Weight	.223	100.0%
Dim	.065	29.3%
Low resolution	.064	28.9%
Effective pixels	.064	28.5%

Neural Network Analysis: All of the above suggested the idea of using Neural Networks for the data analysis; first and foremost were the nonlinear relationships and patterns that could not easily be algebraically described. One strength of an artificial neural network (ANN) is accommodating non-linearities. Several neural network products are available for running ANN's; some of the most well-known are R, SAS eMiner, SPSS, and STATISTICA. These packages use Predictive Model Markup Language (PMML), allowing the neural network models to be shared by different applications (Neural Network Software, 2017). In this study, we used SPSS software (for its relative simplicity) as the primary package for running neural nets. In contrast, fitting an ANN in R increases complexity, requiring significant tuning that involves a significant amount of coding. After fitting the model using different sets of predictor variables and model setting parameters, we achieved the best model by using eight specific predictor variables: Maker, Zoom Wide, Zoom Tele, Macro Focus Range, Weight, Dimensions, Low Resolution, and Effective Pixels.

The model set up parameters that included:

- Splitting the data set into 70% Training and 30% Testing sets;
- Re-scaling of numeric variables using standardization;
- Setting the number of hidden layers to 1;
- Setting the number of units in the hidden layer to 9;
- Activation function: hyperbolic tangent;
- Activation function in output layer: identity;
- Error function: sum of squares;
- One dependent variable: Price.

Figure 22 presents the output of the ANN analysis. The sum of squares error (SSE) values represent the cross-entropy error, considerably lower for the Testing set. As we mentioned earlier, the camera data set had a multi-collinearity problem with the independent variable, Max.resolution. Collinearity impacts the model performance by increasing the difficulty of identifying the true relationships between response and predictor.[15, 16] So, we removed Max.resolution from the model, and the SSE for the Training set reduced from about 115 to a bit below 66. Multiple online articles and data mining blogs touch on the problem of multicollinearity while fitting ANN. The relative error for the Training set registered 0.184, while for the Testing set it registered 0.170. A value closer to 0 indicates that the model has a lower random-error component, thus serving as a more useful fit for prediction. Figure 23 displays a predicted-by-observed value for each data value. It almost seems as if some of the cheaper cameras might have a negative price in the future, which, in a strange way, makes conceptual, if not actual, sense. Some of the cameras released in the 1990s can today be bought, used, on eBay for $30 or less. Cameras in the $1000–$2000 range vary in their predicted prices, from nearly $0 up to about $3000. The majority of cameras fall in the range of $0–$2000; yet several cameras fall in the $4000–$6000 range, and a few cameras approach $8000. Notably, and as expected, as observed camera prices increased, *on average* the predicted value of a camera also increased, with one clear exception being an observed value of $5000 for which the predicted value dropped to below $2000. Figure 24 displays a residual-by-predicted value for each scale-dependent data value.

Importance of Variables in Predicting Camera Price: Figure 25 and Table 1 show the importance of each independent variable in predicting the dependent variable (Camera Price). We performed the independent variable importance analysis based on the combined Training and Testing samples. Earlier, we expressed the belief that the variable, Zoom Wide, might lack significance in predicting the price of a camera. Based on the output, the predicted variables Weight, Maker, and Zoom Tele held the highest importance in predicting the price of a camera; meanwhile, Zoom Wide, Macro Focus Range, Dimensions, Low Resolution, and Effective Pixels all held less importance in the prediction process.

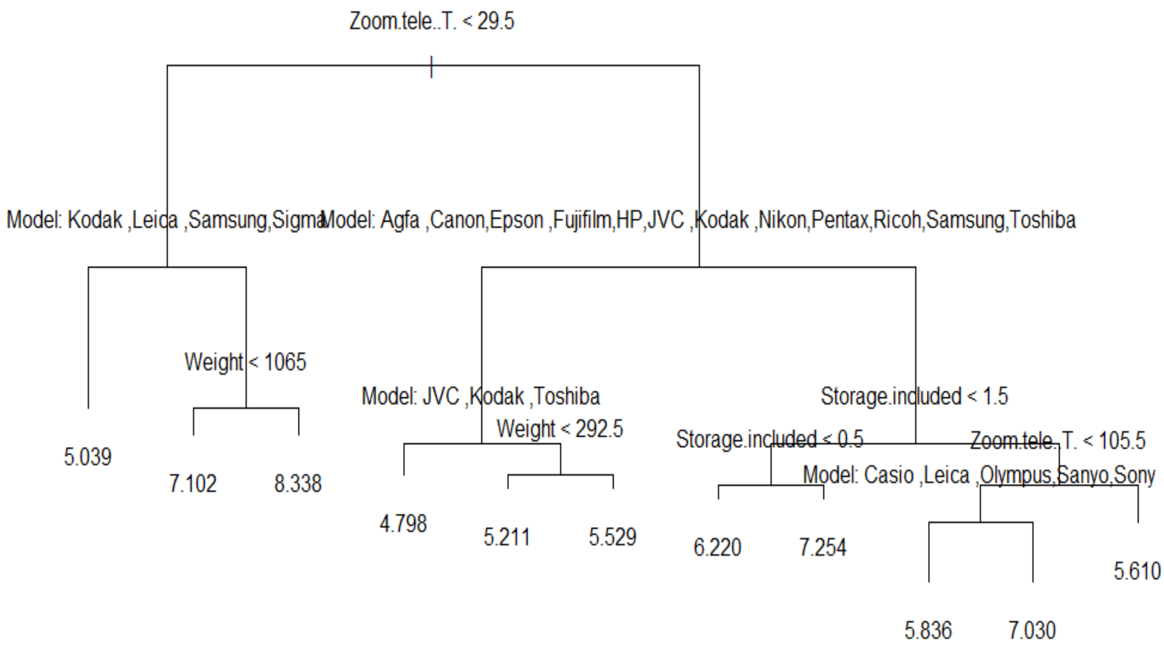

Figure 26. Regression Tree

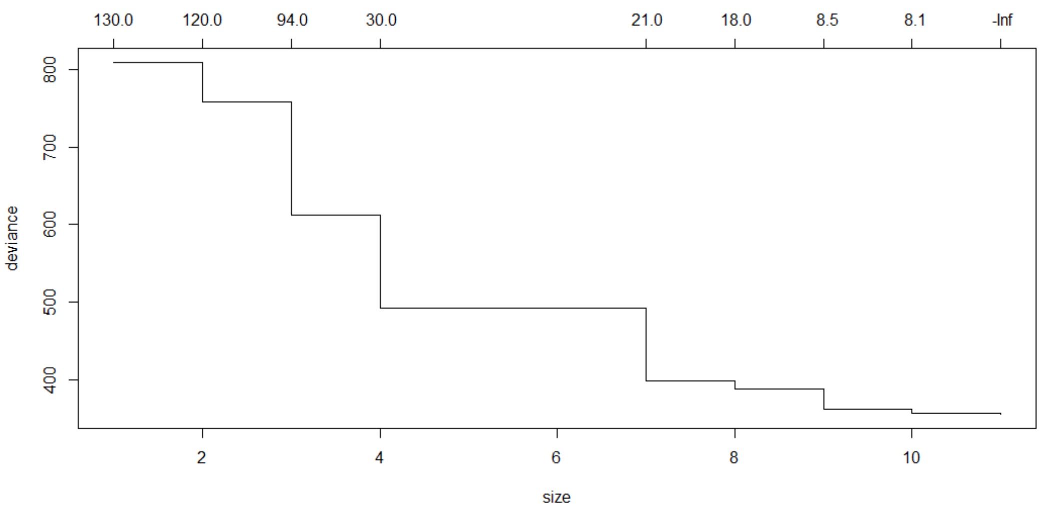

Figure 27. Cross-validation

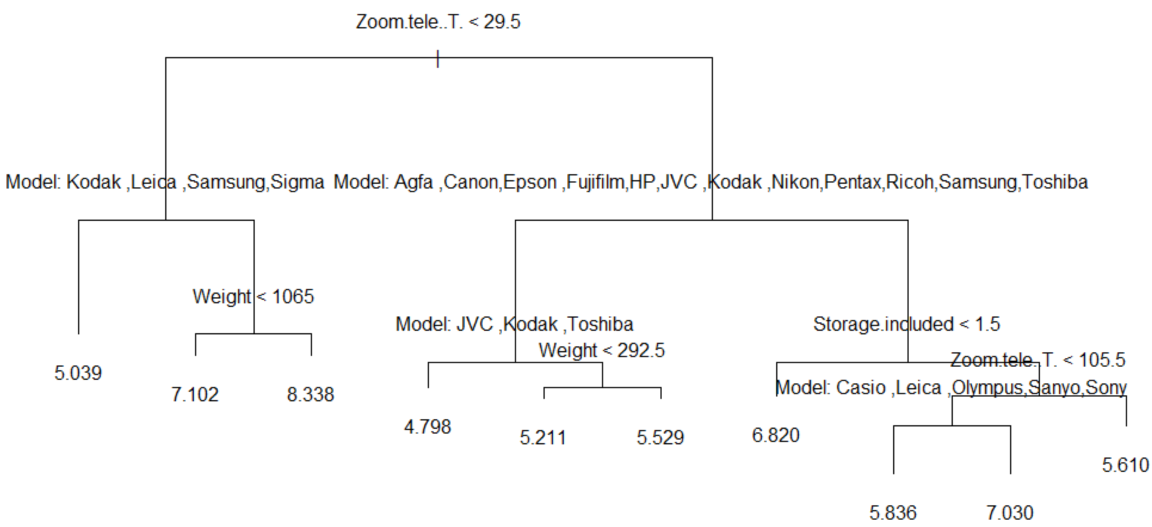

Figure 28. Pruned tree

In Table 2, the R code used for fitting the Regression Tree is provided.

Table 2. Code in R for regression tree

```
>camera<-read.csv(file.choose(), header =
TRUE)
> attach (camera)
library("tree")
> camera.tree=tree(log(Price)~.,data=camera)
> plot(camera.tree)
> text(camera.tree, pretty=0)
> result=cv.tree(camera.tree, K=10,
FUN=prune.tree)
> plot(result)
> tree.new=prune.tree(camera.tree, best=10)
> plot(tree.new)
> text(tree.new, pretty=0)
> exp(5.610)
[1] 273.1442
> exp(5.836)
[1] 342.407
```

It turned out that Zoom Wide's importance score, while not among the "top group," did exceed the importance score of the latter four variables. Table 1 provides the percentage distribution of variable importance in the price estimation of a digital camera, with Weight, the variable with highest importance, normalized to take on a value of 100.

Regression Trees: While the multilayer perceptron neural network model provided insight about the attributes that play a significant role in estimating the price of a digital camera, a regression tree helped to identify an actual predictive value for the price. In order to fit the Regression Tree in R, we use the “tree” function in a “tree” package. For this study, the response of interest represented Price. Predictor variables consisted of the set of attributes used in fitting the neural network model. Based on the Regression-Tree output in Figure 26, if the Zoom Tele value is less than 29.5 mm, and the Maker is among Kodak, Leica, Samsung, or Sigma, then the Price is predicted to be in between exp(5.039) or $154 and exp(7.102) or $1214. But, if Weight is more than 1065 grams, in addition to the Zoom Tele value being less than 29.5 mm, and the Maker is among Kodak, Leica, Samsung or Sigma, then the Price is predicted to be exp(8.338) or $4180; and so on. In general, a smaller tree with fewer splits usually leads to less variance and a better interpretation value, possibly at the cost of a limited bias. One possible way to prune the tree involves using cross-validation. Cross-validation enables selection of the number of terminal nodes for a given data set. Based on Figure 27, we posited that a regression tree with ten terminal nodes might yield the best performance. At the (horizontal axis) value of "10," it can be seen that the variability, called "deviance" (vertical axis), is not reduced materially, but rather, stays virtually unchanged. Figure 28 provides the results for the pruned tree. The pruned tree indicates a similar outcome, except when the Zoom Tele value is more than 29.5 mm, and the Maker is among Agfa, Canon, Epson, Fujifilm, HP, JVC, Kodak, Nikon, Pentax, Ricoh, Samsung, or Toshiba, and the Storage Included value is less than 1.5 GB; then the Price was exp(6.820) or $916.

Comments and Conclusions

Our results from the Multilayer Perceptron Neural Network demonstrate that the predictor variables Maker, Weight, and Zoom Tele played a significant role in predicting the price of a digital camera. The Regression Tree analysis has shown the following:

- If the Zoom Tele value is less than 29.5 mm, and the Model is among Kodak, Leica, Samsung, or Sigma, and the Weight is less than 1065 grams then the predicted price is in between exp(5.039) or $154 and exp(7.102) or $1214;
- If the Zoom Tele value is less than 29.5 mm, and Model is among Kodak, Leica, Samsung, or Sigma, and Weight is more than 1065 grams, then the predicted price is exp(8.338) or $4179;
- If the Zoom Tele value is more than 29.5 mm, and the Maker is among Agfa, Canon, Epson, Fujifilm, HP, JVC, Kodak, Nikon, Pentax, Ricoh, Samsung, or Toshiba, and the Weight is less than 292.5 grams then the predicted price is in between exp(4.798) or $121 and exp(5.211) or $183;
- If the Zoom Tele value is more than 29.5 mm, and the Maker is among Agfa, Canon, Epson, Fujifilm, HP, JVC, Kodak, Nikon, Pentax, Ricoh, Samsung, or Toshiba, and Weight is more than 292.5 grams, then the predicted price is exp(5.529) or $252;
- If the Zoom Tele value is more than 29.5 mm, and the Maker is among Agfa, Canon, Epson, Fujifilm, HP, JVC, Kodak, Nikon, Pentax, Ricoh, Samsung, or Toshiba, and Storage Included is less than 1.5 GB, then the predicted price is exp(6.820) or $916;
- If the Maker is among Agfa, Canon, Epson, Fujifilm, HP, JVC, Kodak, Nikon, Pentax, Ricoh, Samsung, or Toshiba, and Storage Included is less than 1.5 GB, and the Zoom Tele value is more than 105.5 mm, then the predicted price is exp(5.610) or $273;
- If the Maker is among Agfa, Canon, Epson, Fujifilm, HP, JVC, Kodak, Nikon, Pentax, Ricoh, Samsung, or Toshiba, and Storage Included is more than 1.5 GB, and the Zoom Tele value is less than 105.5 mm, then the predicted price can waver between exp(5.836) or $342 and exp(7.030) or $1130.

APPENDIX 1

Camera Data Set Attributes

- Model - maker
- Release Date – year of release
- Max Resolution – the amount of detail that the camera can capture. Measured in pixels. More pixels give the flexibility to crop a picture and still print the remaining image at a decent size (David Peterson, 2014).
- Low Resolution - the amount of detail that the camera can capture. Measured in pixels. Allows one to save memory on the camera’s memory card. Modern digital cameras meet basic resolution requirements, so there is no need to consider image resolution as much as in the past (David Peterson, 2014).
- Effective Pixels – pixels capturing incoming light. Measured in pixels. Effective pixels end up in the final image (Photographer, 2012; Resolution of digital images, 2017).
- Zoom Wide – short focal length. Takes in a large area, making objects appear small. Measured in millimeters (Kirk Wool, 2011).
- Zoom Tele – long focal length. Has a narrow field of view, making objects in front appear larger in the photograph. Measured in millimeters (Kirk Wool, 2011).

- Normal Focus Range – produces an image that roughly matches the same angle of view as the human eye. Measured in millimeters (www.photographymad.com).
- Macro Focus Range – on a fixed lens camera the "macro" mode allows the camera to focus faster on close objects. Measured in millimeters (www.photographymad.com).
- Storage Included – internal memories are often small; as a result camera purchasers typically end up buying a high-capacity card. Measured in GB (www.photoreview.com.au).
- Weight – measured in grams (Stephen Holt, 2011).
- Dimensions – camera size. Measured in millimeters.
- Price – US $.

REFERENCES

Nick Glass 2017. Inside Elton John's spectacular private photography collection. Retrieved from: www.cnn.com

Richard Trenholm 2007. Photos: The history of the digital camera. Retrieved from: https://www.cnet.com

Peter Parchal, 2011. Researchers create world's smallest camera, no bigger than a grain of salt. Retrieved from: http://www.pcmag.com

Project Datasets https://perso.telecom-paristech. fr/eagan/class/igr204/datasets

Thomas Stirr, 2017. A few thoughts about the camera market. Retrieved from: https://photographylife.com

Your guide to Digital Photography. Retrieved from: https://books.google.com

Martin Thoma 2018. How does multicollinearity affect neural networks? [Online forum comment]. Message posted to: https://datascience.stackexchange.com

William, M. and Terry, S. 2012. A second course in statistics: regression analysis. Boston, MA: Pearson.

Mehmed, K. 2011. Data Mining: concepts, methods, and algorithms. Hoboken, NJ: John Wiley & Sons, Inc.

Output (Multilayer Perceptron) 2012. Retrieved from: https://www.ibm.com

Eugene Horber 2017. Outliers in SPSS. Retrieved from: http://www.unige.ch

Mark J. Perry 2012. Christmas shopping for a digital camera: 2000 vs. 2012. Retrieved from: www.aei.org

Neural Network Software 2017. Retrieved from: en.wikipedia.org

Horace T. 2016. Does collinearity largely impact the performance of neural networks? [Online forum comment]. Message posted to https://stats.stackexchange.com/questions/231311

Martin Thoma 2018. How does multicollinearity affect neural networks? [Online forum comment]. Message posted to https://datascience.stackexchange.com

David Peterson, 2014. What is image resolution? Why does it matter? Retrieved from: http://www.digital-photo-secrets.com.

Photographer 2012. Total and effective pixels. Retrieved from: www.whatdigitalcamera.com

Kirk Wool 2011. Photography [Online forum comment]. Message posted to https://photo.stackexchange.com/questions/11164

Standard Lenses. Retrieved from: www.photographymad.com

How much memory do you require? Retrieved from: www.photoreview.com.au

Stephen Holt, 2011. What makes a SLR camera relatively heavy? What is the heaviest part of a SLR camera? Retrieved from: https://www.quora.com

Who invented the camera? (2010, December 8). Retrieved from: http://www.whoinvented.org

Resolution of digital images (2017, November 12). Retrieved from: https://en.wikipedia.org

****