Miguel Sanchez

Project Report

The data set selected was collected from Kaggle. The user that shared this data set explained that the original data set comes from Linio, a well-known eCommerce platform in Latin America, they released this data set in 2016. Linio was interested in knowing the average price in US dollars of 14 popular electronic devices across 72 different countries. In the original data set, countries were ranked by the sum of the average costs of all 14 products.

The original data set has 72 rows for 72 countries, and, 16 columns that include the country and the rank for each, the other 14 columns are the prices of each device. For this project, the data set was modified by deleting one row, leaving the data set with 71 rows, this row represented Venezuela. The reason this row was deleted is because due to inflation, the prices for all the devices in this country were very high and keeping it would have an unexpected impact in the results of the project.

For this project, we want to know which two products will help predicting the cost of the iPhone the best. Since this data set is from 2016, we are also interested in knowing if the linear model we come up with will help predicting the value of an iPhone in the current year, 2019.

Ten different linear models were created using Microsoft Excel and then one of them was selected based on the results. All the linear models have the iPhone as the response variable and two different predictor variables. The process to make a model is to first shuffle all the rows for the purpose of dividing the data into test and training sets. We want to split the data in this way because we want to know how well our linear model adapts to new data, so we create the linear model with the training set, and we test it with the test set. The training set is about 80% of our data, in this case, that is roughly 56 rows. The test set is the remaining 20% of our data, in this case, that is the remaining 15 rows. Then, with the help of an add-in called XLMiner Analysis ToolPak we make the linear model using the training set. By using this add-in, we get some information, but we do not use all of it. The information used was the coefficients *a*, *b*, and *c* for the equation of the linear model, and the residuals *r*.

With the residuals we calculated the square residuals since we want to work with only positive numbers, and it was not given by the Analysis ToolPak. The square residual is equal to *r2*. We also calculated the MSE (mean squared error), this is, the average of all the square residuals. The MSE we calculated is the training set MSE, we now want to know how our linear model behaves with the test set. We want to calculate the average of the square residuals, but we do not have the square residuals for the test set. To calculate this, we need the predicted value of the iPhone price. If we let *a* be the intercept coefficient, *b* be the first predictor variable coefficient, *c* be the second predictor variable coefficient, *X* be the first predictor variable, and *Y* be the second predictor variable, then the predicted value, *Z*, is:

*Z = a + bX + cY.*

Then, we calculate the residuals, this is calculated by taking the difference of the actual value of the iPhone price and the predicted value *Z*. If we let *z* be the actual value of the iPhone price, the residual *r* is *r = z – Z*. After this we calculate the square residuals *r2* and calculate the test set MSE by taking the average of all the square residuals.

To determine which is the best linear model we select the one that the training set MSE is not very large which means that it is a good linear model for the training set. We also want that the test MSE is a number as close as possible to the training MSE, this would mean that our linear model adapts well to new data in general. Out of the 10 models, the one that fits this definition of best the most is when we let the MacBook and the Apple Watch values be our predictor variables. Here we get a training MSE of 12605.1 and a test MSE of 14905.2. It was interesting to see that these two Apple products predict the value of an iPhone good even though they are not smartphones. Some linear models where we used other smartphones like Android were not very good models.

Consequently, we can say that we can predict most accurately the value of an iPhone based on the prices for a MacBook and an Apple Watch. To know if the linear model selected will help predicting the cost of the iPhone in 2019, we must look for prices in US dollars of a MacBook and an Apple Watch. Although the research done by Linio was done with many countries and physical and online stores, I am going to limit myself to only look at stores online and in the United States. Another limitation that this will have is that there are several models and variations of MacBook as well as for Apple Watch. To solve this, I will only consider prices of the newest MacBook Pro and MacBook Air, both 13’’ and 16’’, and the newest Apple Watch both 40mm and 44mm. Also, we could have the same model of MacBook, for example, but with different specifications, this could make the MacBooks have a significant price difference, for this we will only consider the most popular variations, we will do the same with the Apple watch. In addition, I will only consider not pre-owned products.

With all these limitations, we get that the average cost of a MacBook is 1702.9254 USD and the average cost of an Apple Watch is 484.1161 USD. We got these average prices by going to the Apple, Amazon, Costco, and Best Buy online stores.

Now, we know the coefficients *a*, *b*, and *c* from the chosen linear model, they are a = 433.587119496407, b = 0.0466410630887031, and c = 0.43049893402328. We will also let *X* be the average cost of a MacBook, and *Y* be the average cost of an Apple Watch. Then we have that our predicted value of and iPhone, *Z*, is:

*Z* = 433.587119496407 + (0.0466410630887031)\*(1702.9254) + (0.43049893402328)\*(484.1161)

= 721.4248

Therefore, the predicted average iPhone price in 2019 based on the average price of a MacBook and an Apple Watch is 721.42 UDS. To see how accurate this is, we calculate the actual average value of the newest iPhone by going to the same online stores as we did for the other two products and only considering the most common varieties, the newest iPhone models, and only not pre-owned iPhones. The calculated average iPhone cost is 898.15 USD, this is 176.73 USD more than the prediction, an underestimate. This is interesting considering that when we analyze our linear model, we get that the change of getting an underestimate is 46.48%, while we have 53.52% chance of getting an overestimate.

References

Data set collected from:

<https://www.kaggle.com/ikalats/TechnologyPriceIndex>

<https://www.linio.com.mx/sp/technology-price-index-2016>

MacBook Prices taken from:

<https://www.amazon.com/s?k=macbook+pro&language=en_US&ref=nb_sb_noss_1>

<https://www.amazon.com/s?k=macbook+air&ref=nb_sb_noss_2>

<https://www.apple.com/mac/>

<https://www.bestbuy.com/site/searchpage.jsp?_dyncharset=UTF-8&id=pcat17071&iht=y&keys=keys&ks=960&list=n&qp=condition_facet%3DCondition~New&sc=Global&st=macbook%20pro&type=page&usc=All%20Categories>

<https://www.bestbuy.com/site/searchpage.jsp?st=macbook+air&_dyncharset=UTF-8&_dynSessConf=&id=pcat17071&type=page&sc=Global&cp=1&nrp=&sp=&qp=&list=n&af=true&iht=y&usc=All+Categories&ks=960&keys=keys>

<https://www.costco.com/CatalogSearch?dept=All&keyword=macbook+pro>

<https://www.costco.com/CatalogSearch?dept=All&keyword=macbook+air>

Apple Watch prices taken from:

<https://www.amazon.com/s?k=apple+watch+series+5&ref=nb_sb_noss>

<https://www.apple.com/us-hed/shop/buy-watch/apple-watch>

<https://www.costco.com/CatalogSearch?dept=All&keyword=apple+watch+series+5>

<https://www.bestbuy.com/site/searchpage.jsp?st=apple+watch+series+5&_dyncharset=UTF-8&_dynSessConf=&id=pcat17071&type=page&sc=Global&cp=1&nrp=&sp=&qp=&list=n&af=true&iht=y&usc=All+Categories&ks=960&keys=keys>

iPhone prices taken from:

<https://www.amazon.com/s?k=iPhone+11&ref=nb_sb_noss>

<https://www.bestbuy.com/site/iphone/shop-iphone-11-iphone-11-pro-and-iphone-11-pro-max/pcmcat1568147641595.c?id=pcmcat1568147641595>

<https://www.costco.com/iphones.html>

<https://www.apple.com/iphone/>