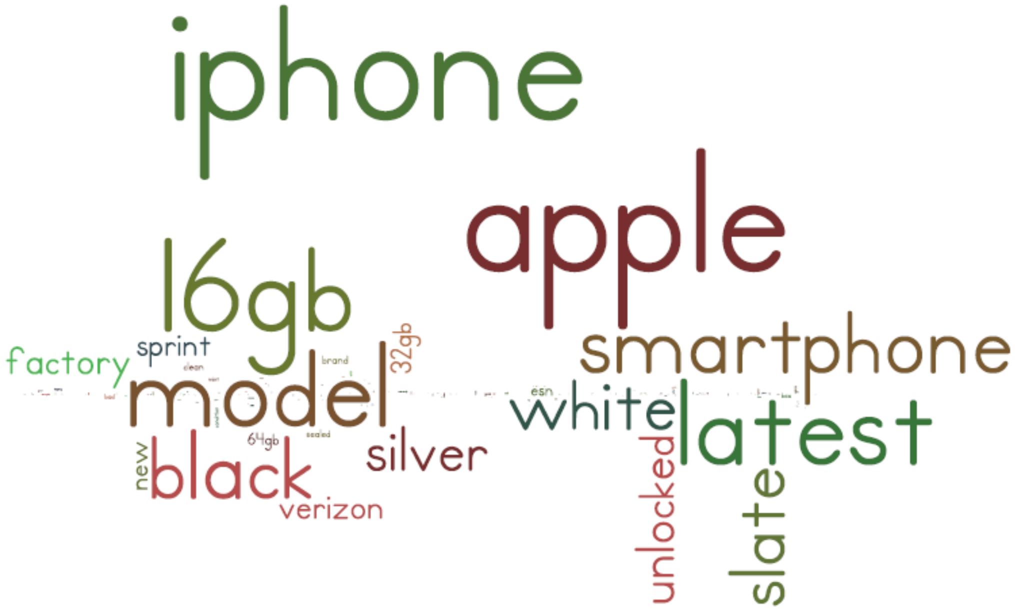
# Data Analysis of Trends in iPhone 5 Sales on eBay



By Wenyu Zhang

Mentor: Professor David Aldous

# Contents

# 1. INTRODUCTION

eBay Inc. is an American multinational internet corporation founded in 1995. It provides a variety of online marketplaces for the sale of goods and services, online payment services, and online communication offerings to individuals and businesses in the United States and internationally. The corporation manages eBay.com, a popular online platform on which people buy and sell a wide range of goods and services, such as clothes, electronics, and even real property. The website supports two methods of sale, that is, auction-style and "Buy-It-Now" standard shopping. eBay.com has an estimated 19 million users worldwide, with over 2 million visitors each day.

The buying and selling patterns on eBay has been an object of interest for the many people using the website. How do prices vary across time within a day? Are items priced higher on weekends? How differently are new and used products priced? This report will attempt to answer some of these questions, with respect to a particular product: iPhone 5.

The iPhone 5 was released on September 21, 2012. Following the success of its predecessors, iPhone 5 is received with overwhelming popularity. At the online Apple store, the product is priced as follows:

| Carrier & Contract | Storage | Price |
|---|---|---|
| Unlocked, Contract-free | 16 GB | $649 |
| | 32 GB | $749 |
| | 64 GB | $849 |
| Network carrier, with Contract | 16 GB | From $199 |
| | 32 GB | From $299 |
| | 64 GB | From $399 |

Table 1.1 iPhone 5 prices at Apple store

From Table 1.1 above, price differences for new iPhones at the store arise mainly from the type of carrier and contract, as well as the storage capacity. While the store prices are fixed, prices on online marketplaces like eBay tend to spread over a much wider range. One difference to note is that the iPhones on eBay are sold without contract, though many are still carrier-specific. Also, the condition of item, the format of sale, and the time of the listing all introduce additional fluctuation to the final price at with the item is sold.

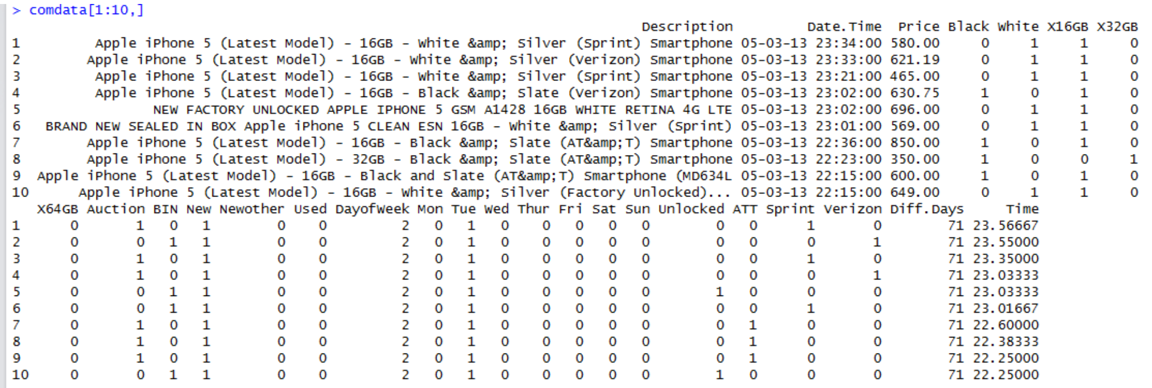
The goal of the project is to build a model to predict the price of an iPhone 5 with a particular condition and carrier, that is sold on eBay through a particular format of sale, at a specific day of the week and time period of the day. Due to the nature of the data collected, I will focus mainly on the use of regression trees and random forests. Tree methods are nonlinear, nonparametric, and can simultaneously handle numerical and categorical data.

The project hopes to provide prospective buyers with more information and viable strategies they can employ, so as to maximize utility. In general, the project attains more insights and perspectives on merchant and consumer behaviors on online shopping sites.

## 2. DATA AND ANALYSIS

### 2.1 DESCRIPTION OF DATA

The dataset consists of approximately 2 months' worth of data, from December 25, 2012 to March 5, 2013. After the removal of faulty data, there are a total of 36062 entries. Fig 2.1.1 below shows a portion of this dataset. 3000 entries are randomly selected as the test set, the remaining 33062 entries make up the training set.



```
> comdata[1:10,]
                                                                            Description         Date.Time  Price Black White X16GB X32GB
1         Apple iPhone 5 (Latest Model) - 16GB - White &amp; Silver (Sprint) Smartphone 05-03-13 23:34:00 580.00     0     1     1     0
2        Apple iPhone 5 (Latest Model) - 16GB - White &amp; Silver (Verizon) Smartphone 05-03-13 23:33:00 621.19     0     1     1     0
3         Apple iPhone 5 (Latest Model) - 16GB - White &amp; Silver (Sprint) Smartphone 05-03-13 23:21:00 465.00     0     1     1     0
4         Apple iPhone 5 (Latest Model) - 16GB - Black &amp; Slate (Verizon) Smartphone 05-03-13 23:02:00 630.75     1     0     1     0
5                NEW FACTORY UNLOCKED APPLE IPHONE 5 GSM A1428 16GB WHITE RETINA 4G LTE 05-03-13 23:02:00 696.00     0     1     1     0
6   BRAND NEW SEALED IN BOX Apple iPhone 5 CLEAN ESN 16GB - White &amp; Silver (Sprint) 05-03-13 23:01:00 569.00     0     1     1     0
7        Apple iPhone 5 (Latest Model) - 16GB - Black &amp; Slate (AT&amp;T) Smartphone 05-03-13 22:36:00 850.00     1     0     1     0
8        Apple iPhone 5 (Latest Model) - 32GB - Black &amp; Slate (AT&amp;T) Smartphone 05-03-13 22:23:00 350.00     1     0     0     1
9  Apple iPhone 5 (Latest Model) - 16GB - Black and Slate (AT&amp;T) Smartphone (MD634L 05-03-13 22:15:00 600.00     1     0     1     0
10      Apple iPhone 5 (Latest Model) - 16GB - White &amp; Silver (Factory Unlocked)... 05-03-13 22:15:00 649.00     0     1     1     0
   X64GB Auction BIN New Newother Used DayofWeek Mon Tue Wed Thur Fri Sat Sun Unlocked ATT Sprint Verizon Diff.Days     Time
1      0       1   0   1        0    0         2   0   1   0    0   0   0   0        0   0      1       0        71 23.56667
2      0       0   1   1        0    0         2   0   1   0    0   0   0   0        0   0      0       1        71 23.55000
3      0       1   0   1        0    0         2   0   1   0    0   0   0   0        0   0      1       0        71 23.35000
4      0       1   0   1        0    0         2   0   1   0    0   0   0   0        0   0      0       1        71 23.03333
5      0       0   1   1        0    0         2   0   1   0    0   0   0   0        1   0      0       0        71 23.03333
6      0       0   1   1        0    0         2   0   1   0    0   0   0   0        0   0      1       0        71 23.01667
7      0       1   0   1        0    0         2   0   1   0    0   0   0   0        0   1      0       0        71 22.60000
8      0       1   0   1        0    0         2   0   1   0    0   0   0   0        0   1      0       0        71 22.38333
9      0       1   0   1        0    0         2   0   1   0    0   0   0   0        0   1      0       0        71 22.25000
10     0       0   1   1        0    0         2   0   1   0    0   0   0   0        1   0      0       0        71 22.25000
```

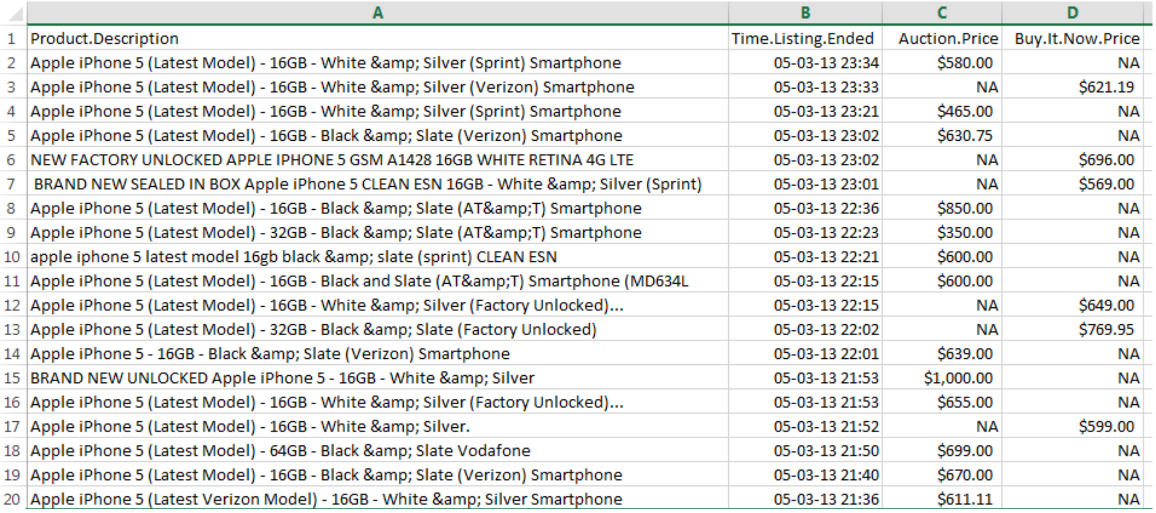
Fig 2.1.1 Portion of complete dataset

The response variable is the Price column, and the predictor variables are on the columns thereafter. The first two columns of Description and Date.Time are used to extract the predictors, and more details can be found on the following subsection. The full set of predictors consist of 8 variables:

- Color (Black, White)
- Storage capacity (16GB, 32GB, 64 GB)
- Format of sale (Auction, Buy-It-Now)
- Condition of item (New, New: other, Used)
- Day of the week (Mon, Tue, …, Sun)
- Carrier (Unlocked, AT&T, Sprint, Verizon)
- Diff.Days i.e. number of days from the first day of available data (Dec 25, 2012)
- Time i.e. time at which the listing ended

## 2.2 RETRIEVAL OF DATA

Data is retrieved directly from eBay.com, under the Completed Listing tab. Fig 2.2.1 shows a portion of the raw data obtained from the page source of the website.



| | A | B | C | D |
|---|---|---|---|---|
| 1 | Product.Description | Time.Listing.Ended | Auction.Price | Buy.It.Now.Price |
| 2 | Apple iPhone 5 (Latest Model) - 16GB - White &amp; Silver (Sprint) Smartphone | 05-03-13 23:34 | $580.00 | NA |
| 3 | Apple iPhone 5 (Latest Model) - 16GB - White &amp; Silver (Verizon) Smartphone | 05-03-13 23:33 | NA | $621.19 |
| 4 | Apple iPhone 5 (Latest Model) - 16GB - White &amp; Silver (Sprint) Smartphone | 05-03-13 23:21 | $465.00 | NA |
| 5 | Apple iPhone 5 (Latest Model) - 16GB - Black &amp; Slate (Verizon) Smartphone | 05-03-13 23:02 | $630.75 | NA |
| 6 | NEW FACTORY UNLOCKED APPLE IPHONE 5 GSM A1428 16GB WHITE RETINA 4G LTE | 05-03-13 23:02 | NA | $696.00 |
| 7 | BRAND NEW SEALED IN BOX Apple iPhone 5 CLEAN ESN 16GB - White &amp; Silver (Sprint) | 05-03-13 23:01 | NA | $569.00 |
| 8 | Apple iPhone 5 (Latest Model) - 16GB - Black &amp; Slate (AT&amp;T) Smartphone | 05-03-13 22:36 | $850.00 | NA |
| 9 | Apple iPhone 5 (Latest Model) - 32GB - Black &amp; Slate (AT&amp;T) Smartphone | 05-03-13 22:23 | $350.00 | NA |
| 10 | apple iphone 5 latest model 16gb black &amp; slate (sprint) CLEAN ESN | 05-03-13 22:21 | $600.00 | NA |
| 11 | Apple iPhone 5 (Latest Model) - 16GB - Black and Slate (AT&amp;T) Smartphone (MD634L | 05-03-13 22:15 | $600.00 | NA |
| 12 | Apple iPhone 5 (Latest Model) - 16GB - White &amp; Silver (Factory Unlocked)... | 05-03-13 22:15 | NA | $649.00 |
| 13 | Apple iPhone 5 (Latest Model) - 32GB - Black &amp; Slate (Factory Unlocked) | 05-03-13 22:02 | NA | $769.95 |
| 14 | Apple iPhone 5 - 16GB - Black &amp; Slate (Verizon) Smartphone | 05-03-13 22:01 | $639.00 | NA |
| 15 | BRAND NEW UNLOCKED Apple iPhone 5 - 16GB - White &amp; Silver | 05-03-13 21:53 | $1,000.00 | NA |
| 16 | Apple iPhone 5 (Latest Model) - 16GB - White &amp; Silver (Factory Unlocked)... | 05-03-13 21:53 | $655.00 | NA |
| 17 | Apple iPhone 5 (Latest Model) - 16GB - White &amp; Silver. | 05-03-13 21:52 | NA | $599.00 |
| 18 | Apple iPhone 5 (Latest Model) - 64GB - Black &amp; Slate Vodafone | 05-03-13 21:50 | $699.00 | NA |
| 19 | Apple iPhone 5 (Latest Model) - 16GB - Black &amp; Slate (Verizon) Smartphone | 05-03-13 21:40 | $670.00 | NA |
| 20 | Apple iPhone 5 (Latest Verizon Model) - 16GB - White &amp; Silver Smartphone | 05-03-13 21:36 | $611.11 | NA |

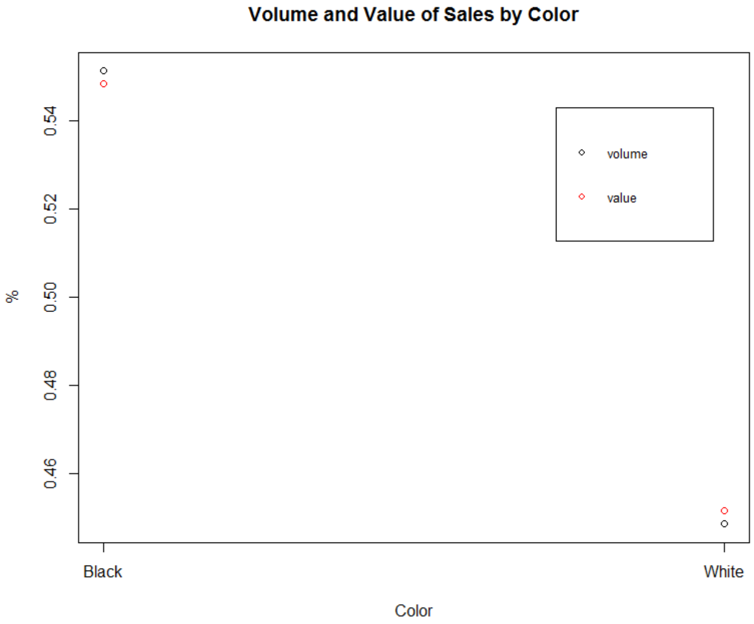
Fig 2.2.1 Portion of raw data

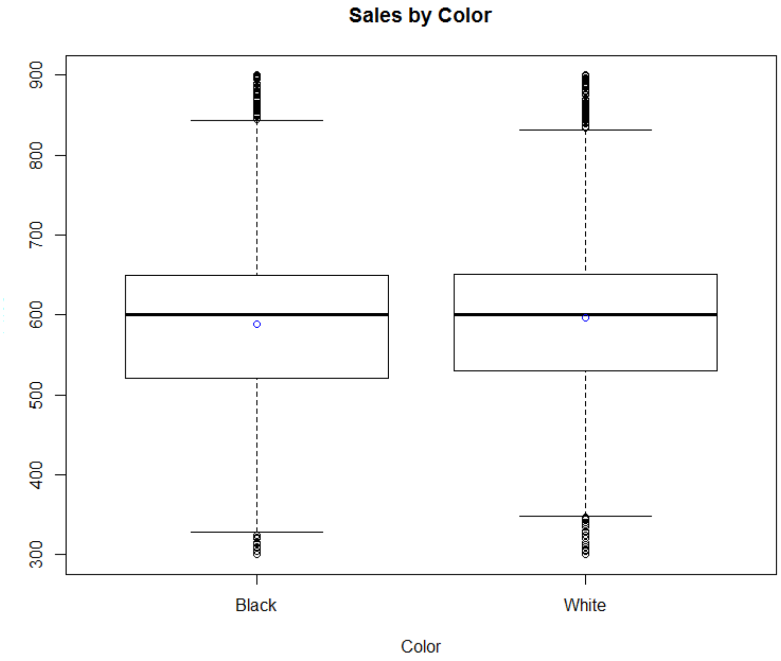
The predictor features of Color, Storage capacity, Condition of item, and Carrier are extracted from the Product Description column through word matching. Day of the week, Diff.Days and Time are extracted from the Time.Listing.Ended column. Format of sale is extracted based on whether the item has a value under Auction.Price or Buy.It.Now.Price.

## 2.3 PRELIMINARY ANALYSIS

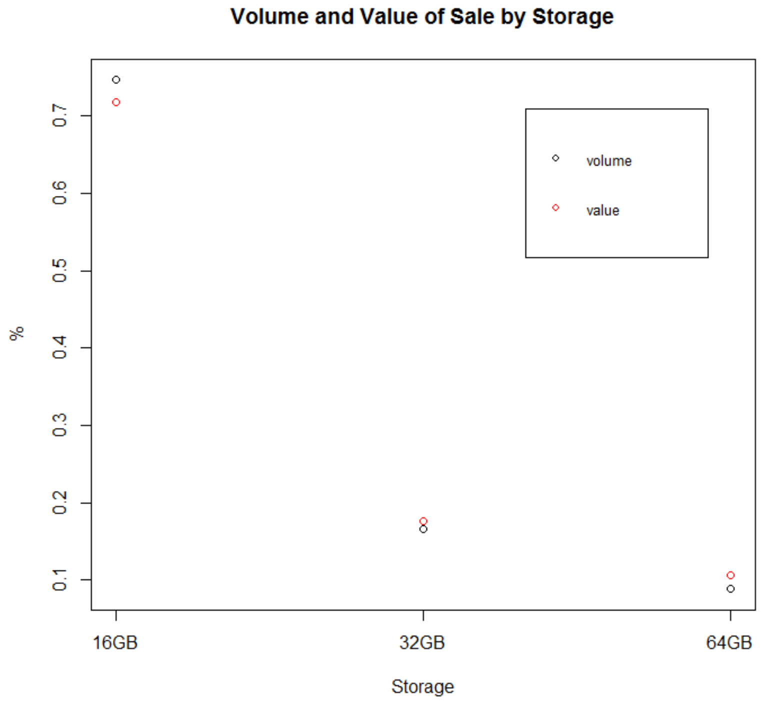
Some preliminary analysis is conducted to explore the general relationship between price and each predictor variable. For each categorical variable, the entries are separated into the corresponding categories. The percentage volume and value of sale for each category are calculated and plotted for comparison. For instance, if percentage volume is higher than percentage value, then it is possible that iPhones in that category tend to be sold at a price lower than those in other categories. The mean price for each category is added as a point on the boxplot. T-tests or one-way ANOVA tests are used to check if the samples are from populations with equal mean.

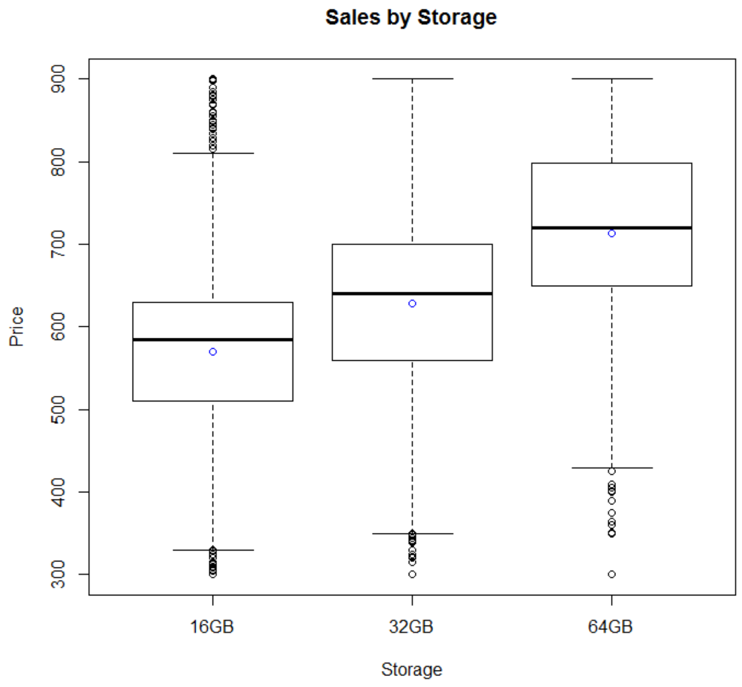
(Differences between points in plots may appear visually exaggerated due to the scale.)



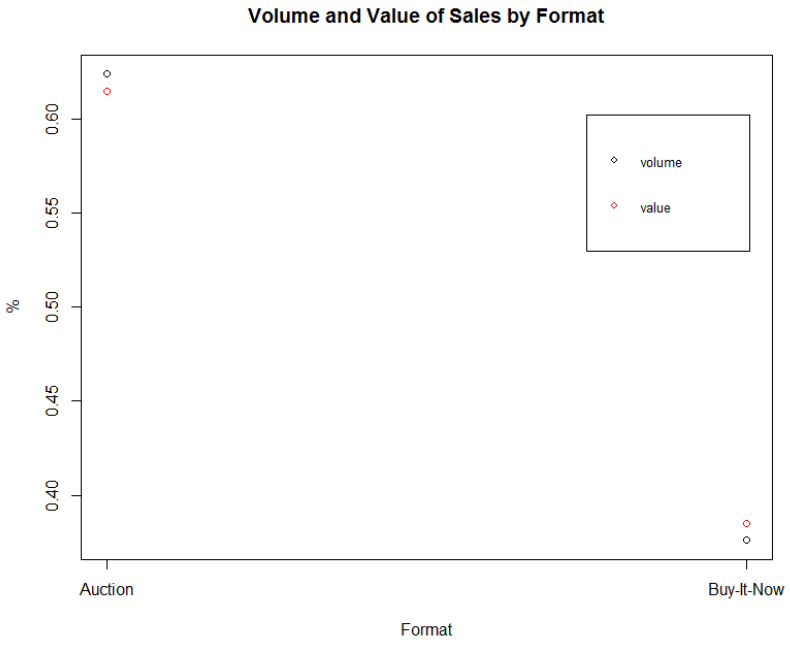


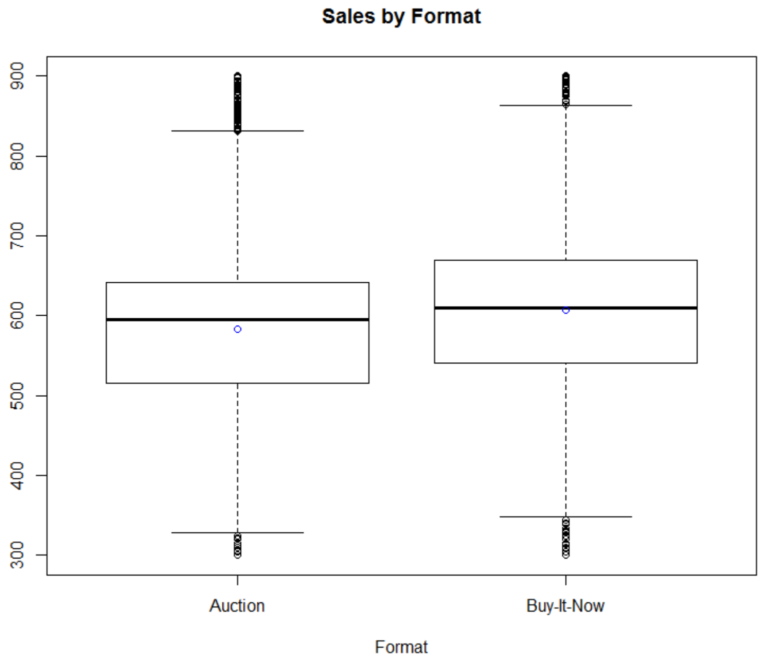
The mean prices of black and white iPhones differ by $7.15. The t-test returns a p-value of 8.038e-11, such that the null hypothesis of equal mean can be rejected.



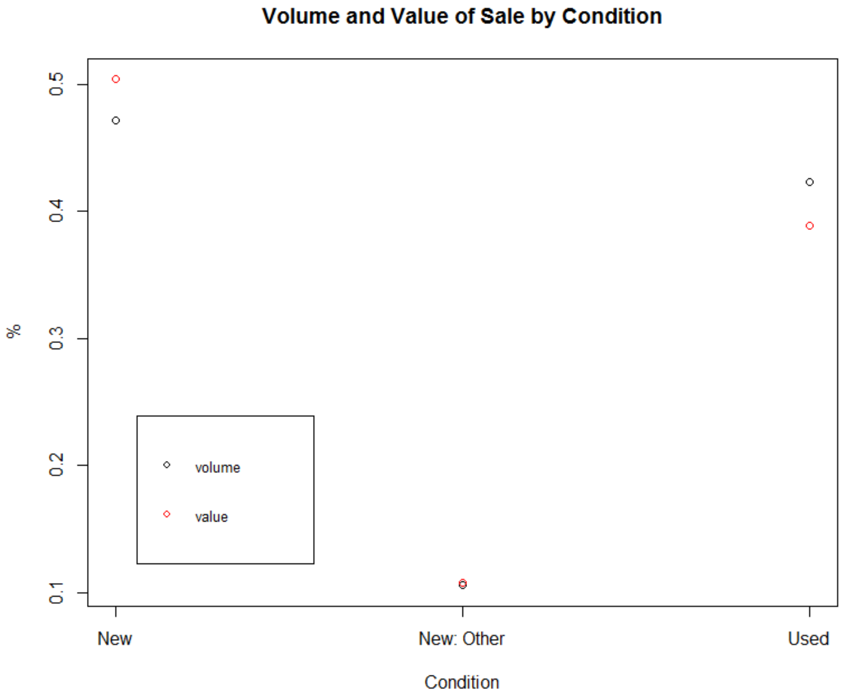


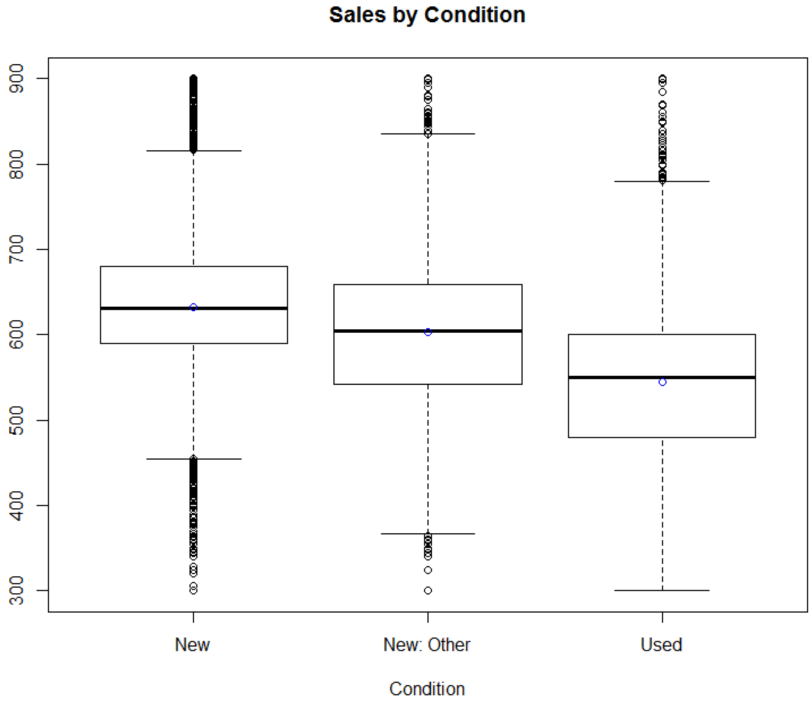
The mean prices of 16GB and 64GB iPhones differ by $142.60. The one-way ANOVA test returns a p-value of < 2.2e-16, such that the null hypothesis of equal mean can be rejected.



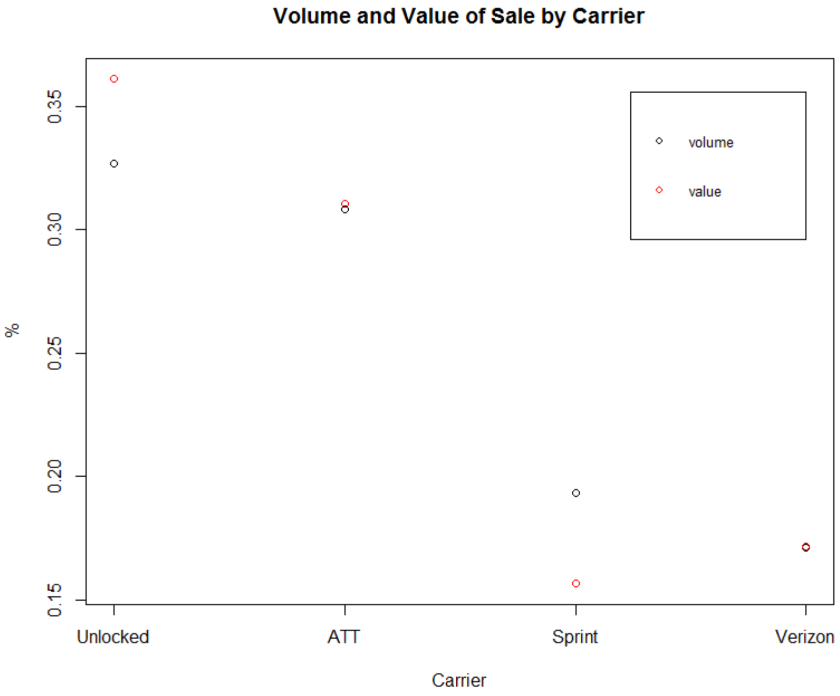


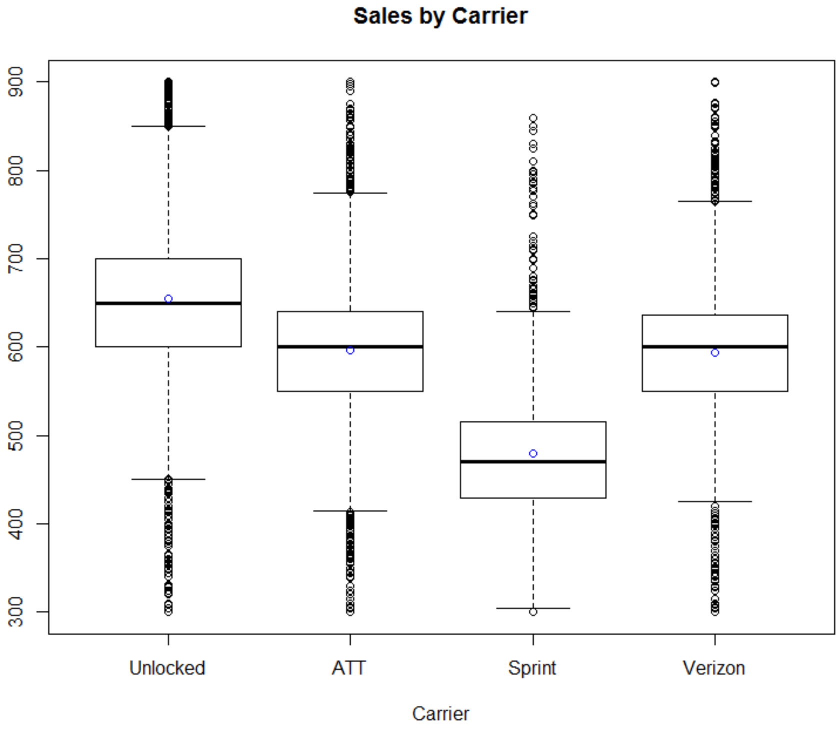
The mean prices of iPhones sold in auction and Buy-It-Now style differ by $23.16. The t-test returns a p-value of < 2.2e-16, such that the null hypothesis of equal mean can be rejected.



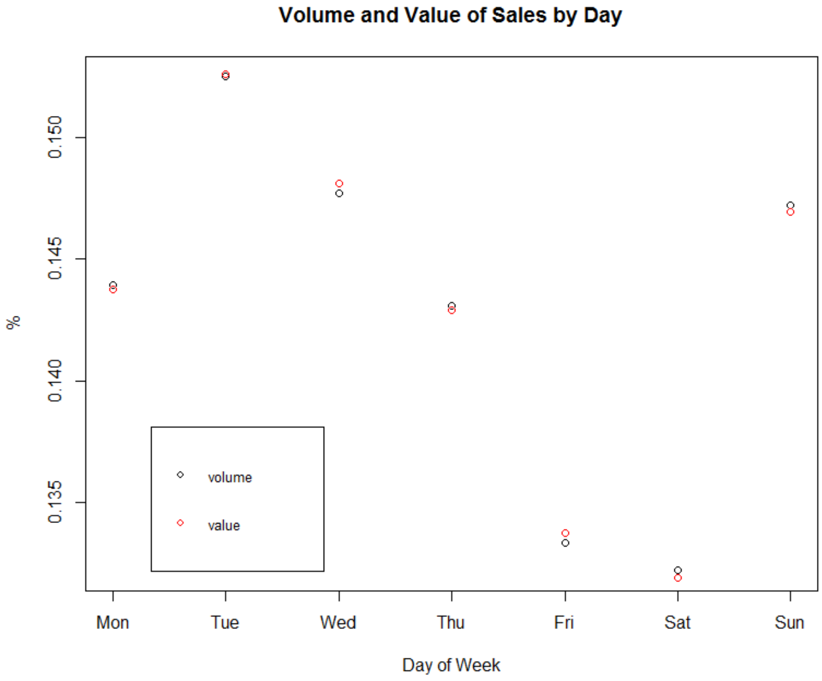


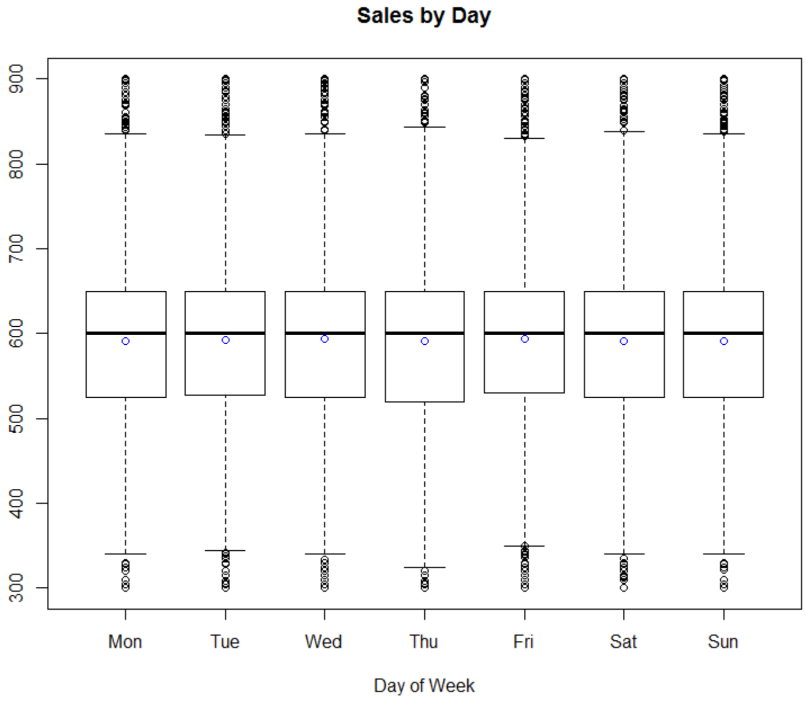
The mean prices of new and used iPhones differ by $88.26. The one-way ANOVA test returns a p-value of < 2.2e-16, such that the null hypothesis of equal mean can be rejected.



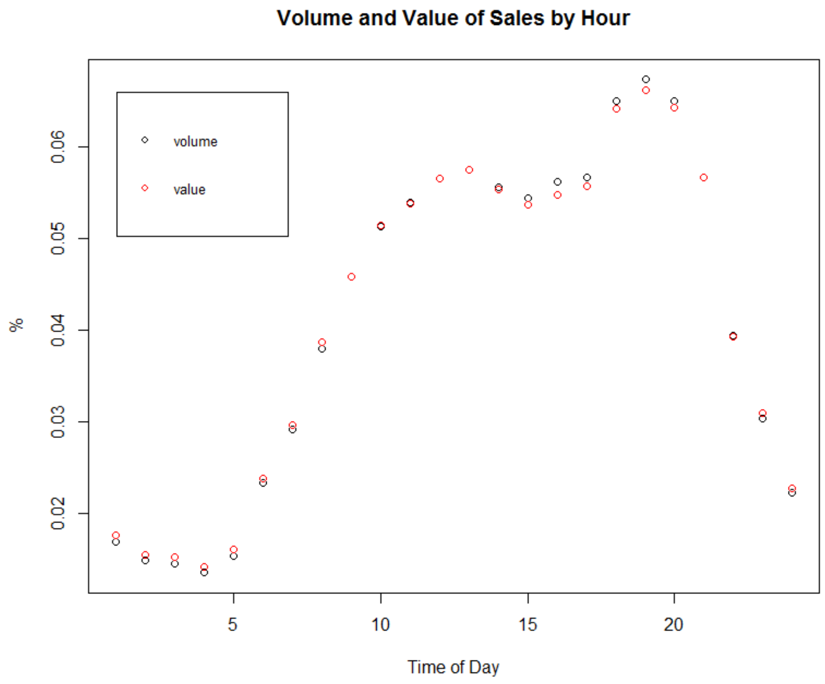


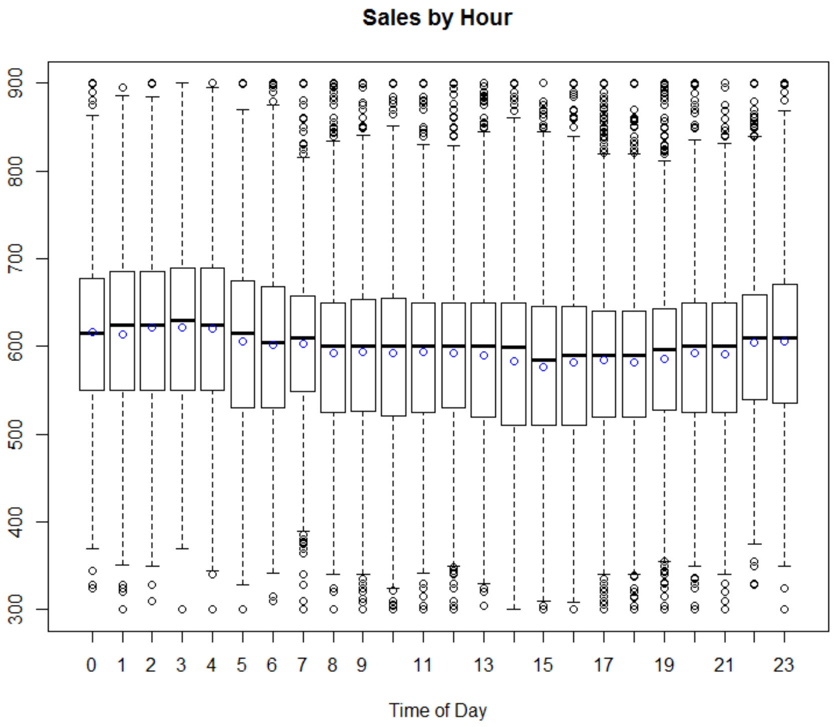
The mean prices of unlocked and Sprint iPhones differ by $175.26. The one-way ANOVA test returns a p-value of < 2.2e-16, such that the null hypothesis of equal mean can be rejected.





The maximum difference between mean prices of iPhones sold on each day of the week is $2.97. The one-way ANOVA test returns a p-value of 0.6292, such that the null hypothesis of equal mean cannot be rejected.





The maximum difference between mean prices of iPhones sold within one-hour blocks is $44.88. The one-way ANOVA test returns a p-value of < 2.2e-16, such that the null hypothesis of equal mean can be rejected.

# 3. MODELS

## 3.1 REGRESSION TREE

The *rpart* function in R grows the tree shown in Fig 3.1.1. Since the first split is made at whether the carrier is Sprint, it seems that the type of carrier plays the most important role on the price of the iPhone 5 and that Sprint phones are generally priced lower than those of the other carriers. As observed from the other splits made, storage capacity and product condition are also important determinants of price.

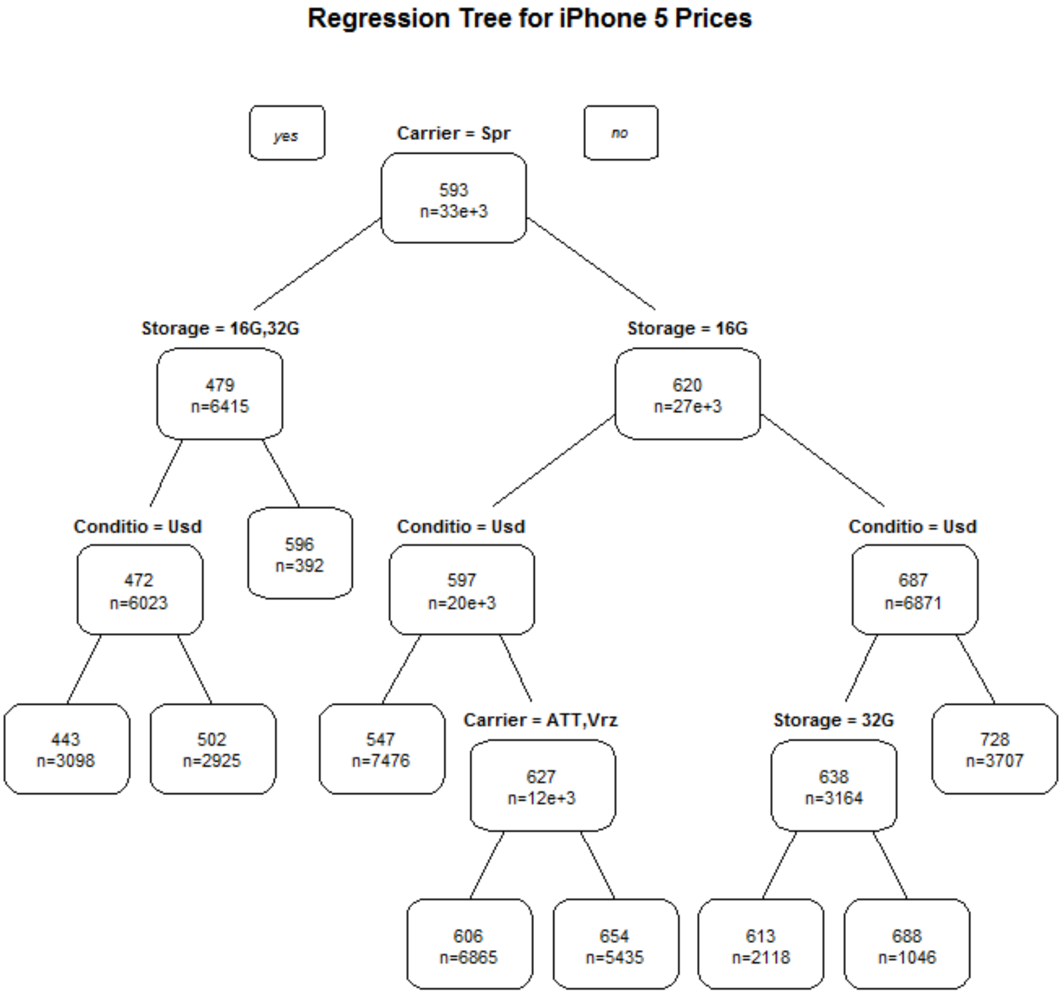


Fig 3.1.1 Regression tree

By a plot of the cross-validation results below, a tree size of 9 gives a minimum validation error of 0.42321. This is exactly the original tree grown, so pruning produces the same tree in this case.

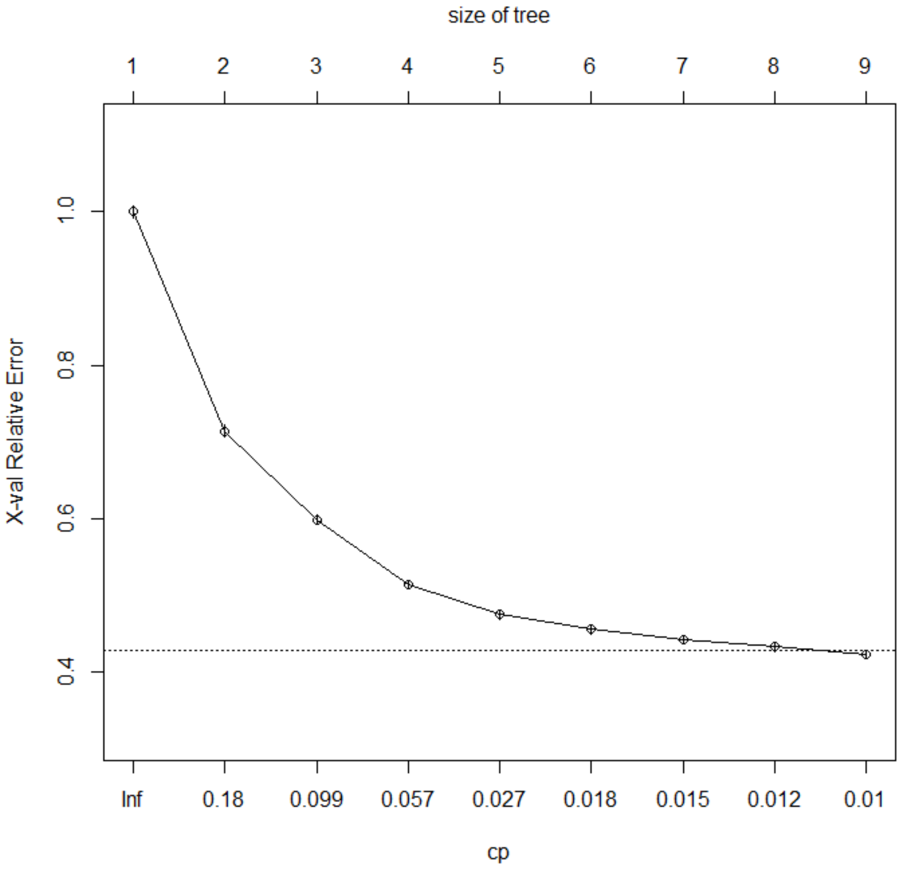


Fig 3.1.1 Cross-validation results

Fitting the model to the test data and using mean squared error as the risk function, a test error of 4598.462 is obtained. I attempt to improve the results using random forests in the following section.

## 3.2 RANDOM FOREST

Random forests improve predictive accuracy by generating a large number of bootstrapped trees, and deciding a final predicted outcome by averaging the results across all the trees. Using the *randomForest* function in R, a forest of 200 trees are created with bootstrapped samples. The parameter *ntree = 200* is chosen by cross validation using the options 10, 100, 200, 300, 400, 500, and choosing the value that minimizes the validation error. The test error is reduced slightly to 3890.83.

Fig 3.2.1 shows the importance of each variable as measured by the random forest. For regression, importance is calculated as how much the split reduces node impurity in terms of residual sum of squares. Similar to the observations from the previous regression tree, the most important variable is the type of carrier, followed by storage capacity and product condition. Time and date related features come next. Sale format and color are the least important determinants of iPhone 5 prices.

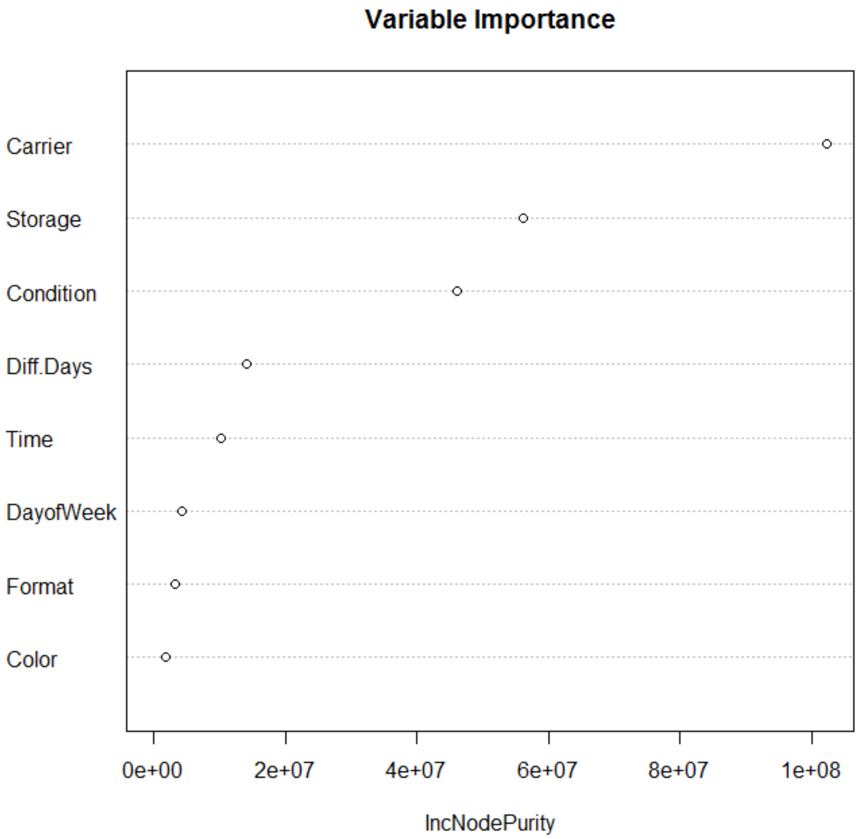


Fig 3.2.1 Plot of importance of variables

## 4. CONCLUSION

Looking at the 3000 test set values predicted by the random forest model and the three most important variables identified (i.e. Carrier, Storage, and Condition), I took the average price in each category and obtained the following summary:

| Carrier | ATT | Verizon | Sprint | Unlocked |
|---|---|---|---|---|
| Price | 598.34 | 593.96 | 486.43 | 644.72 |

| Storage | 16GB | 32GB | 64GB |
|---|---|---|---|
| Price | 570.66 | 630.57 | 697.95 |

| Condition | New | New: other | Used |
|---|---|---|---|
| Price | 626.85 | 606.12 | 550.17 |

Table 4.1 Average price by category

The variables Storage and Condition reflect physical differences between the products sold, and the price differences are as expected with new, 64GB phones being more expensive than used, 16GB ones. While phones with specific carriers are expectedly sold at prices lower than their unlocked counterparts, it is interesting that Sprint phones are on average more than $100 cheaper than ATT and Verizon phones. This may be due to differences in the kind of service plans available at each carrier.

The average prices of new, unlocked iPhones of each storage capacity sold on eBay are calculated and tabulated in Table 4.2. Since phones with these attributes are sold at the official Apple stores as well, a comparison can be made between the prices of the same products sold via two different channels. As the table shows, 16GB phones are priced similarly on eBay and at Apple stores, while 32GB and 64GB phones are generally cheaper on eBay. On average, a 32GB iPhone 5 can be bought for $20 less on eBay than at Apple stores, and a 64GB for $75 less.

| Storage | 16GB | 32GB | 64GB |
|---|---|---|---|
| Price on eBay | 649.82 | 729.11 | 773.33 |
| Price on Apple | 649 | 749 | 849 |

Table 4.2 Price comparison between eBay and Apple store

## 5. FUTURE RESEARCH

Further work can be done in optimizing the regression tree and random forest models, and in trying to fit other models to the dataset, so that more accurate predictions of price can be attained.

More in-depth study can also be done with the data available through the Completed Listing tab on eBay. For instance, since ending time of a Buy-It-Now listing marks exactly the time a product is purchased, the data reflect appropriately consumer behavior across time and can help to identify peak buying periods on eBay. Time series analysis can be incorporated in this case. The information obtained would be particularly useful for sellers in determining the best times to list a product in order to maximize profits. A similar analysis can be conducted for Auction listings, although more coding work would be necessary to extract relevant bidding history.