

Blackwell Electronics Brand Preference Report

By: Jessica Krueger

Executive Summary

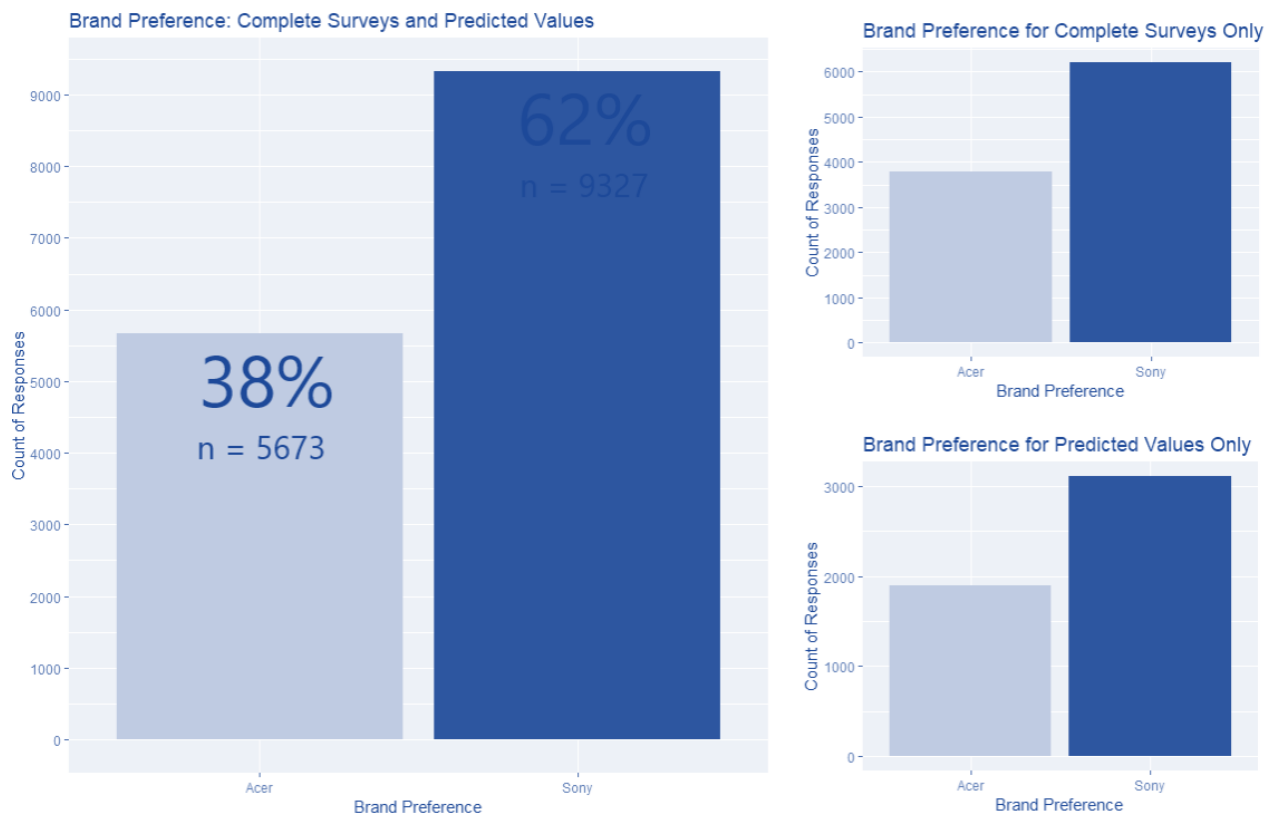
Goal: This report contains an analysis of marketing surveys to understand our customers' preference for Acer or Sony computers. We were asked to predict the brand preference for 5,000 surveys where this information was not provided. For all surveys where brand preference was left blank, the goal was to predict whether each respondent prefers Acer or Sony computers based on their other survey answers.

Process: A Random Forest classification model based on 10,000 examples and 2 attributes (salary and age) was used to predict the brand preference of 5,000 examples where brand preference was not given. The random forest model resulted in an accuracy rate of .901 and a kappa of .789.

Conclusion: Based on the complete surveys and adding our predictions, 62% of customers prefer Sony computers over Acer computers. The chart below on the left shows the brand preference for all 15,000 examples. This chart includes the 10,000 examples for which we had complete surveys as well as the 5,000 examples for which we predicted the brand preference.

The chart on the top right shows the breakdown of brand preference for the 10,000 examples for which we had complete data. The chart on the bottom right shows the values we predicted for the 5,000 examples for which brand preference was not given on the survey.

Based on this information, we should focus on strengthening our relationship with Sony, because that is the brand that the majority of our customers prefer.



Blackwell Electronics Brand Preference Report

By: Jessica Krueger

Model Selection

Model Selected:

Random Forest:

<u>Accuracy</u>	<u>Kappa</u>
0.9007603	0.7886384

7501 samples
2 predictors (Salary & Age)

No pre-processing
Resampling: Cross-Validated (10 fold, repeated 10 times)
Summary of sample sizes: 6751, 6751, 6751, 6751, 6752, 6750, ...

Tuning parameter 'mtry' was held constant at a value of 2

Model Tested but **Not** Selected:

k-Nearest Neighbor:

<u>Accuracy</u>	<u>Kappa</u>
0.7022809	0.3645954

7501 samples
2 predictors (Salary & Age)

No pre-processing
Resampling: Cross-Validated (10 fold, repeated 10 times)
Summary of sample sizes: 6751, 6751, 6751, 6751, 6752, 6750, ...

Resampling results across tuning parameters:

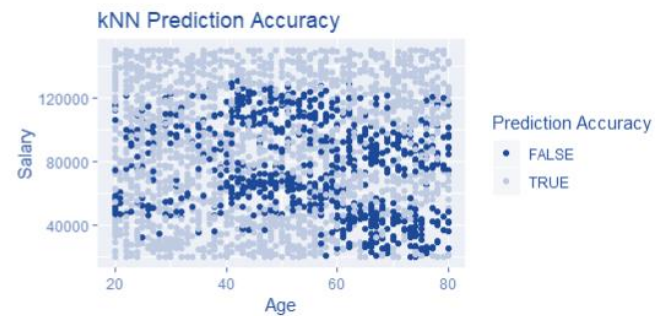
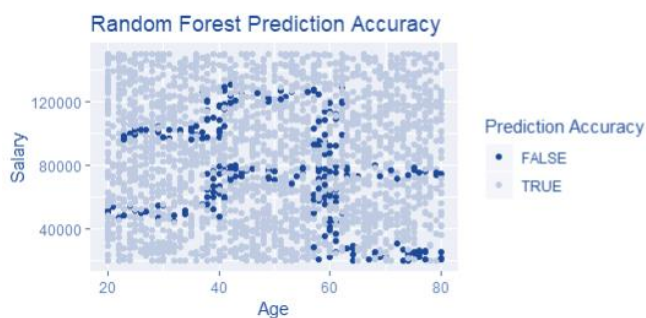
<u>k</u>	<u>Accuracy</u>	<u>Kappa</u>
5	0.7017715	0.3625212
7	0.7054122	0.3701203
9	0.7105051	0.3797124

Accuracy was used to select the optimal model using the largest value.

The final value used for the model was k = 9

Model Selection Rationale:

Based simply on the accuracy and kappa statistics, we can see that the Random Forest model is performing better than the k-NN model. By also visualizing the errors each model made, it's clear that the Random Forest model is more accurately able to understand the relationship that salary and age have on brand preference. The plots below display the errors made by each model and clearly show how much better the Random Forest model performs.

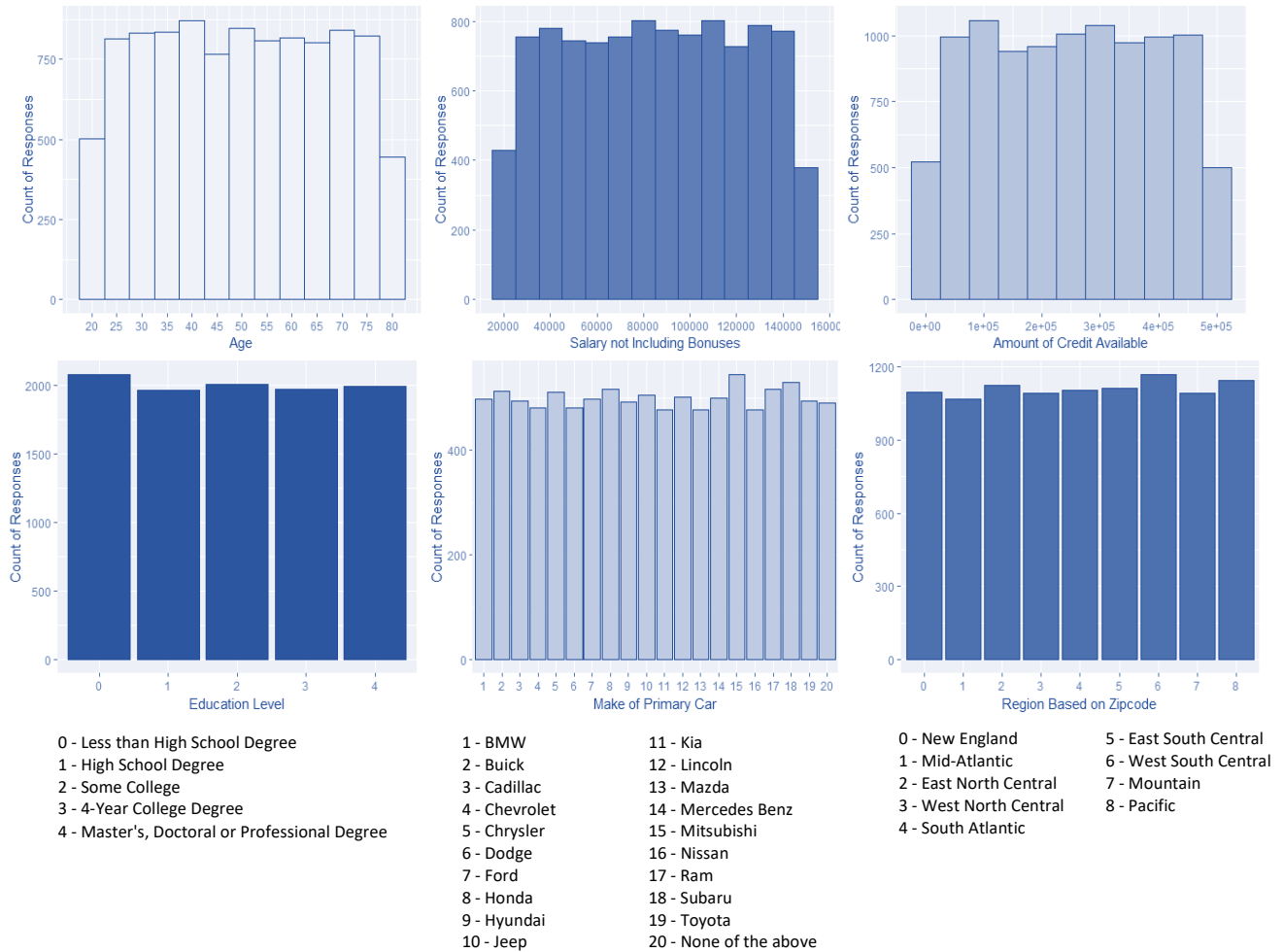


Blackwell Electronics Brand Preference Report

By: Jessica Krueger

Data Cleaning Performed

This data set did not include any missing values, and it did not appear to include any outliers or inconsistent data. Therefore, no data cleaning was performed. Below are plots of each attribute to show the relatively equal distributions for each variable as well as the lack of outliers.



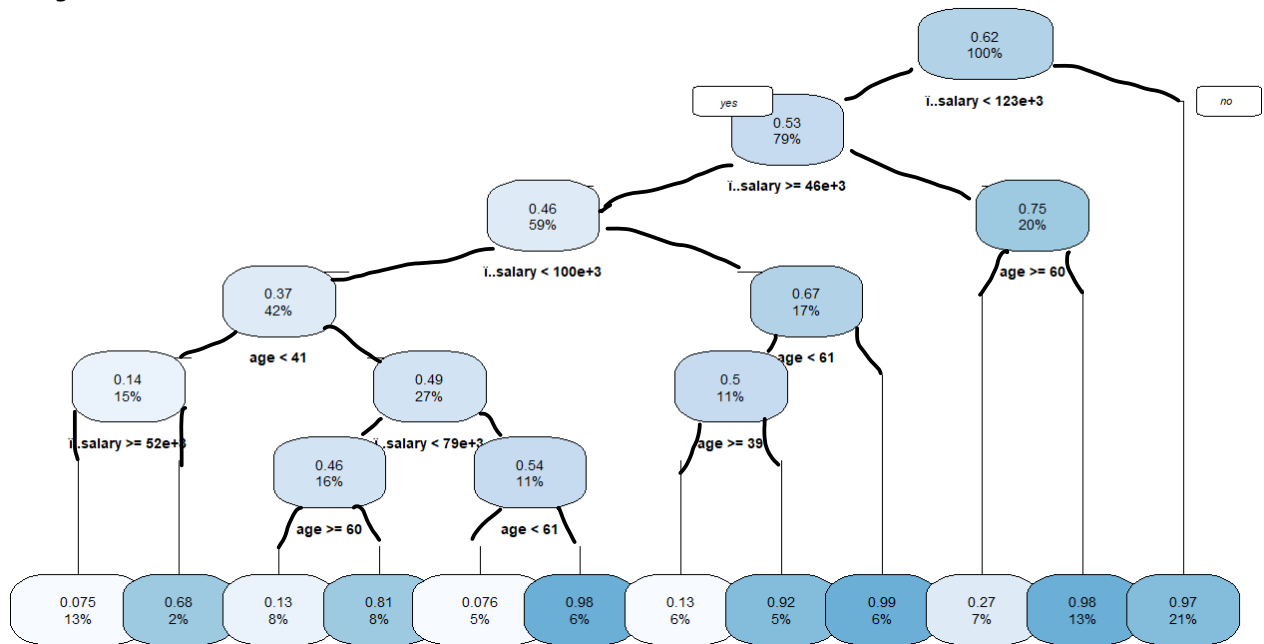
Feature Selection

Features Selected: *Salary* and *Age* were the two attributes selected to create each model. From the initial data exploration, it was clear that salary seemed to have the biggest impact on brand preference. Using a basic decision tree, we learned that age is the only other attribute that has a meaningful effect on brand preference. Figure 1 shows the decision tree which only includes salary and age as the factors that influence brand preference. In Figure 2, we explore this relationship further with a plot of each respondent's age, salary and what brand they prefer. From this plot, we can see that there is a clear pattern in brand preference based on age and salary.

Blackwell Electronics Brand Preference Report

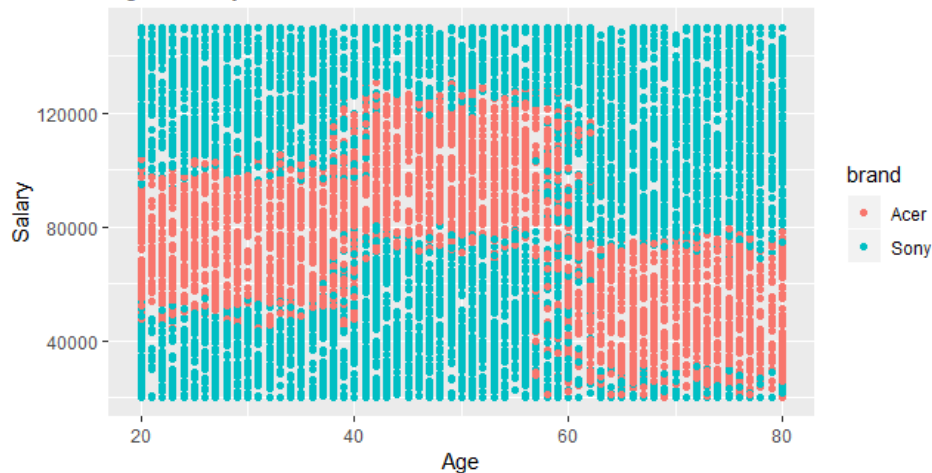
By: Jessica Krueger

Figure 1



Age, Salary, and Brand Preference

Figure 2



Features Excluded: *Education Level, Make of Primary Car, Region, and Available Credit* were not selected to create the model because they were deemed irrelevant to brand preference. From the charts in Figure 3, we can see that the *Education Level, Make of Primary Car, and Region* attributes do not have an impact on brand preference. For each attribute, there is no clear preference for Acer or Sony based on the value of the attribute. For example, among people who prefer Sony, there is roughly an equal distribution among education levels. Similarly, for people who prefer Acer, there is a relatively equal distribution

Blackwell Electronics Brand Preference Report

By: Jessica Krueger

among education levels. If education level had an impact on brand preference, we would expect to see some pattern where respondent's with a certain education level prefer a certain brand.

The attribute *Amount of Credit Available* was also excluded from the analysis due to irrelevancy. Because *Available Credit* is a continuous variable, we have to create a different kind of chart to explore its relationship with brand preference. Figure 4 shows for each level of *Available Credit*, the distribution of brand preference doesn't change in a meaningful way. From this chart, we can see that credit is not impacting brand preference.

Figure 3

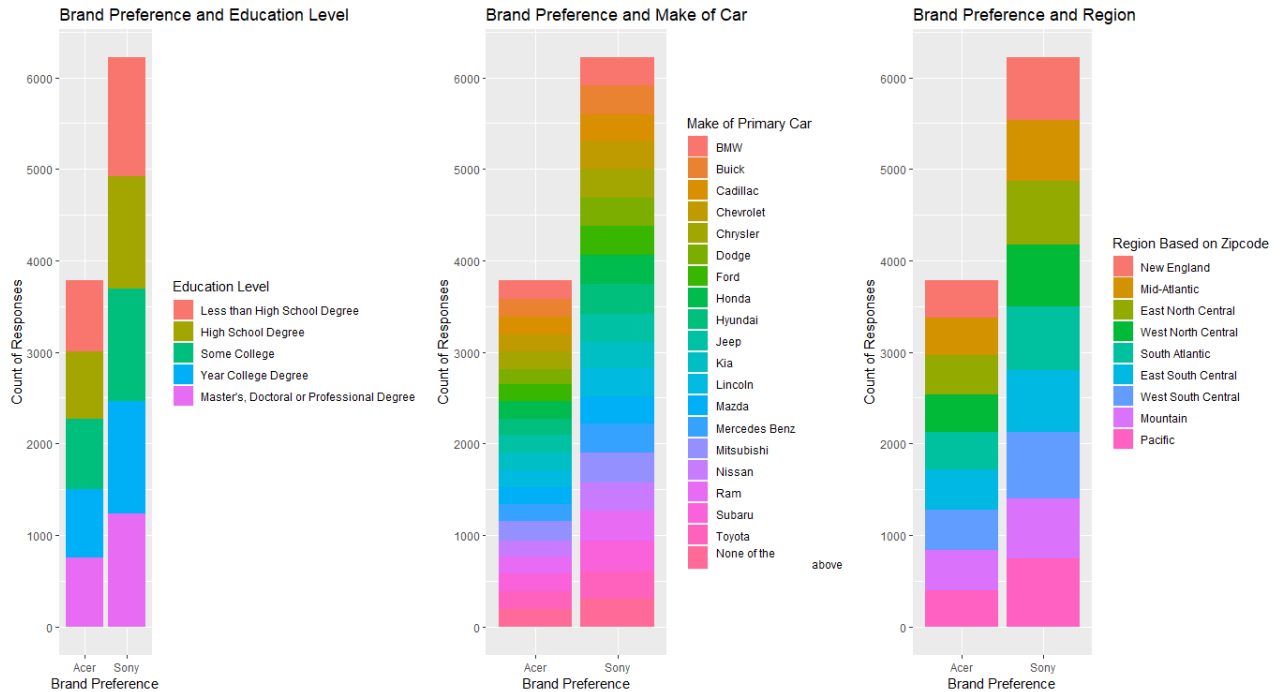


Figure 4

