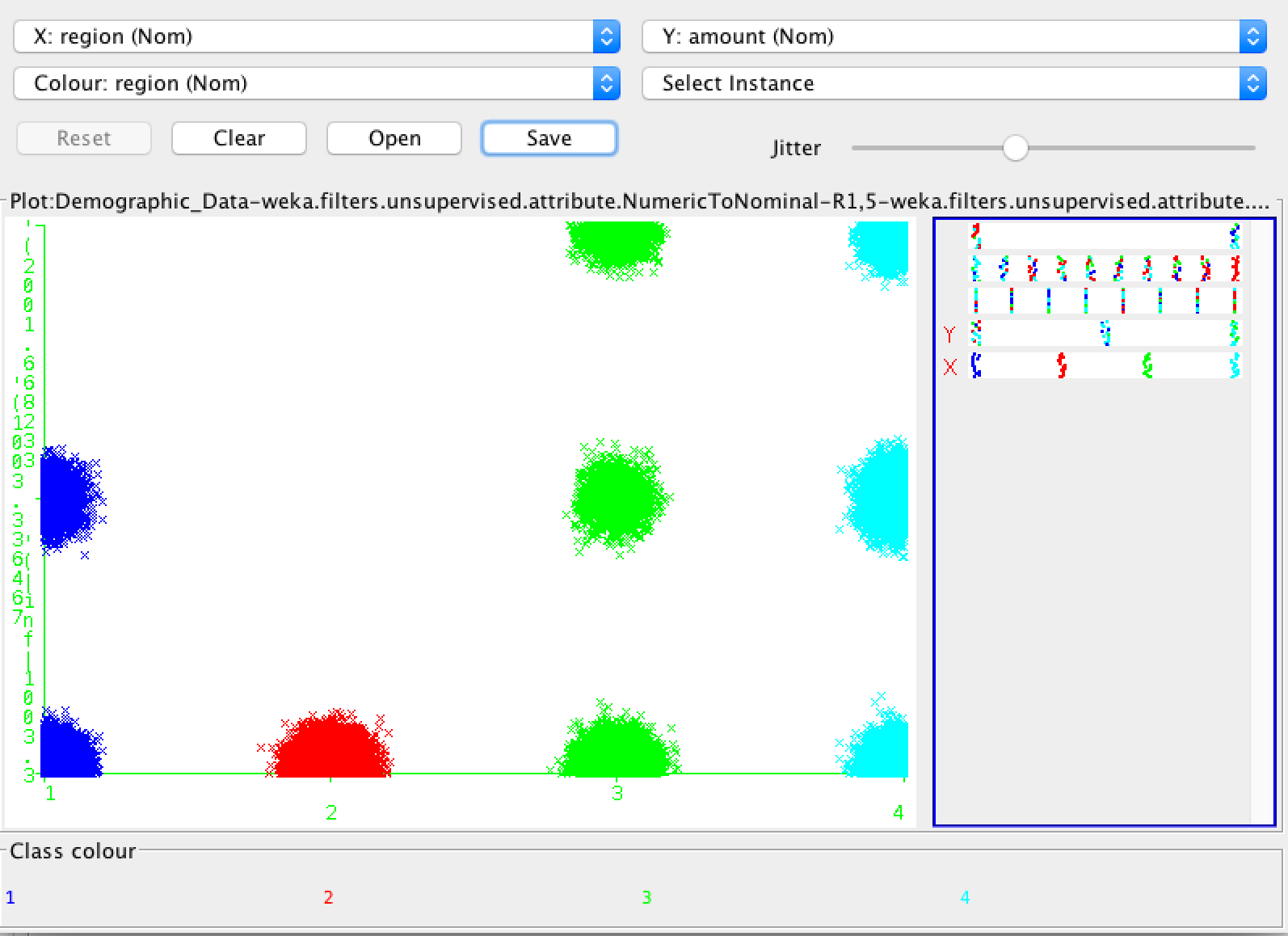
Matthew Reed

Blackwell's eCommerce Team

Danielle Sherman

April 18, 2016

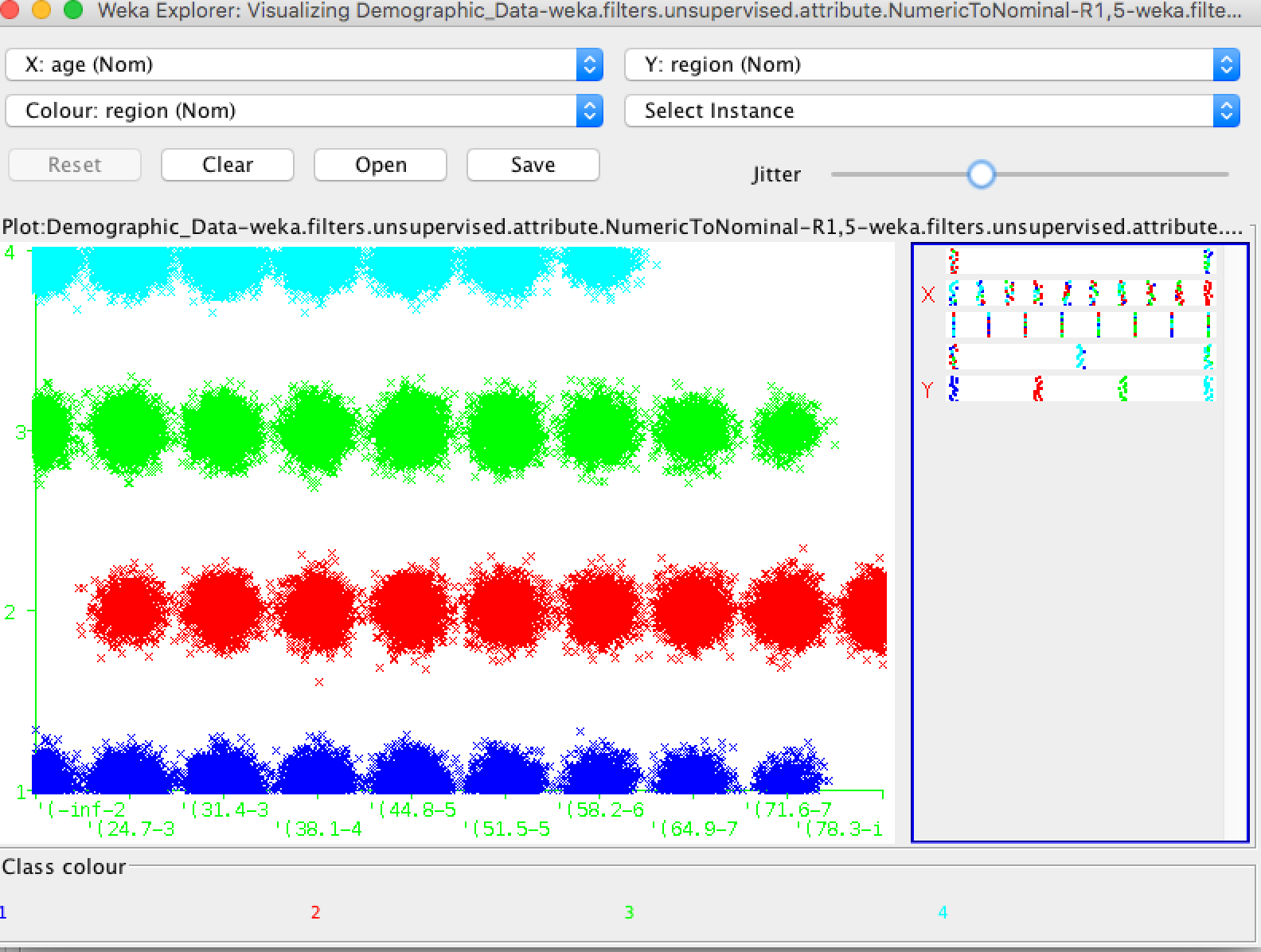
Customer Buying Patterns

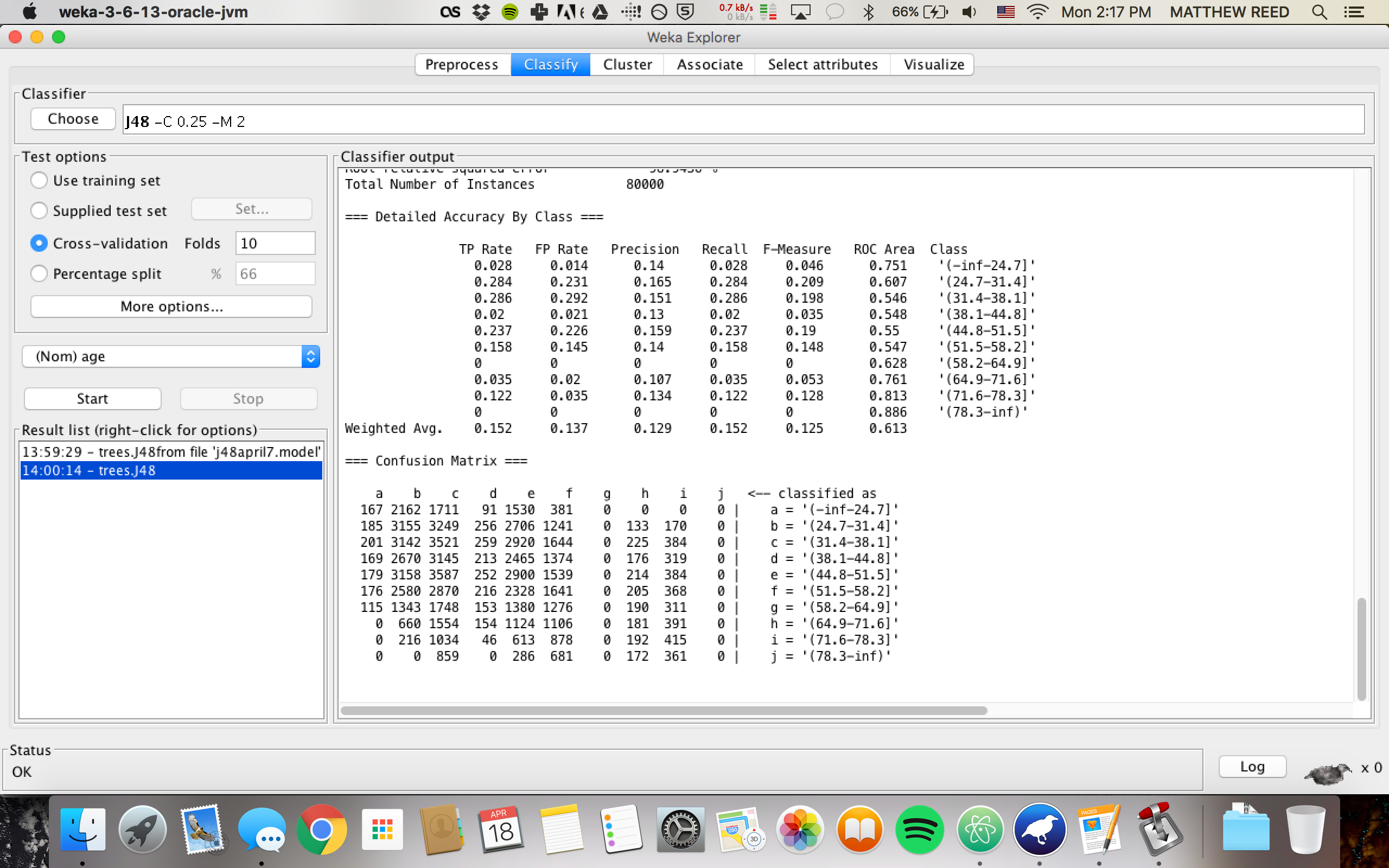
Data mining is the application of computational methods, including machine learning, for pattern recognition to data sets in order to extract information from the data that can then be transformed or interpreted for use in decision-making or additional analyses.

REGION

(1 = East, 2 = West, 3 = South, 4 = Central)

**Q: Do customers in different regions spend more per transaction? Which regions spend the most/least?**

A: The South and Central regions tend to spend more per transaction while West spends the least. I found this by pre-processing the data, converting the “in-store” and “region” purchase data from a numeric range to nominal classes using the “NumericToNominal" filter. I also organized the “age” and “amount” purchase data into bins. The data is then visualized on the chart above.

**Q: Are there differences in the age of customers between regions? If so, can we predict the age of a customer in a region based on other demographic data?**

A: There are slight differences in age based on region, and we can predict the age of a customer based on reading the confusion matrix. This metric will provide insight into how well the classifier was able to predict for each instance (transaction) the age class of the customer

15% of the time I was able to properly classify the instance, with the confusion matrix stating a better classification of younger people than old people.

**Q: Can we predict the amount a customer will spend per transaction based on other data we have collected about that customer?**

We can successfully predict the amount a customer will spend based on region and age. The M5P Algorithm predicted amount spent in regions 1,3,4 narrowed down by age.

**Q: Is there any correlation between age of a customer and if the transaction was made online or in the store? Or do other factors correlate to an online or in-store transaction?**

Using a J48 Classification Algorithm, I was able to correctly classify instances 87% of the time. Region 1 is primarily in-store while Region 2 is primarily online. In Region 3, people under 24.7 years buy online. In region 4, ages 51-65 bought online.

Overall, we discovered that through different data mining methods, we are able to come up with a variety of insight concerning customer buying patterns. Although not every task was a complete success, much can still learned from the data. Understanding spending habits by age, region, and method will help Blackwell understand their customer base and in turn help set up the corporation for success. The data mining methods used in this report will definitely aid in the continued success of Blackwell and their eCommerce Team now and into the future.

**Feedback Questions: Was there anything you didn't find in the data that you thought you might have seen? Knowing the accuracy and correlation coefficients of your models are there any of your results that should not be presented as being valid? Why or why not?**