

SALES PREDICTION REPORT

BLACKWELL ELECTRONICS

Nastaran Mahmoodzadeh

Fall 2020

Historical Data Analysis

INDIVIDUAL VALUE PLOTS

We see the value distribution of each feature, separately.

- Most of Prices are under \$1000
- Left Skewed distribution in RecommendedProduct



Figure 2

As we see in Figure 3, most of the prices we have are under \$1000. The highest price average belongs to Laptop and PC. Also, the least price distribution is for Netbook. Finally, PC has the most outliers in price distribution in compare to other products.

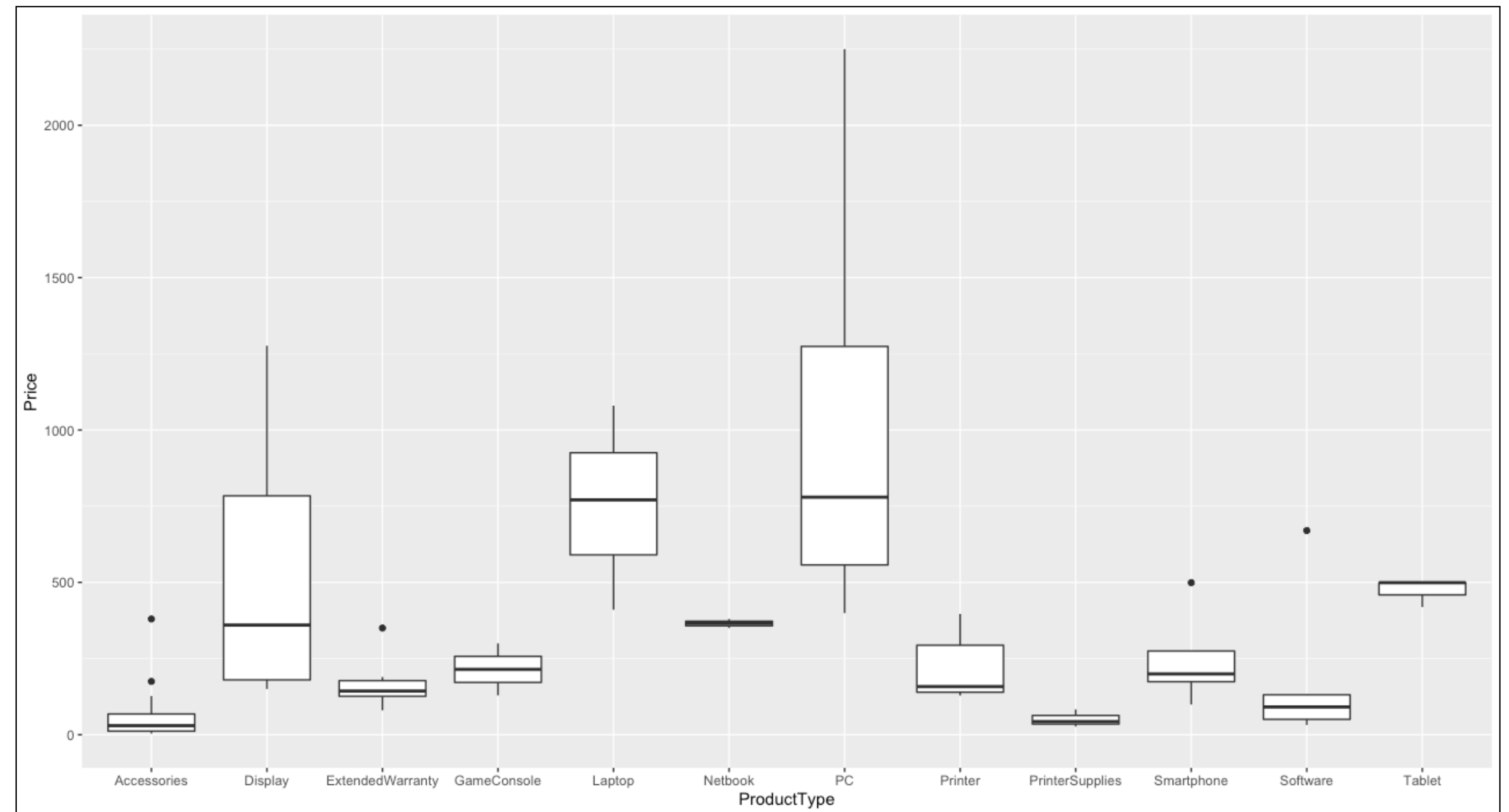


Figure 3

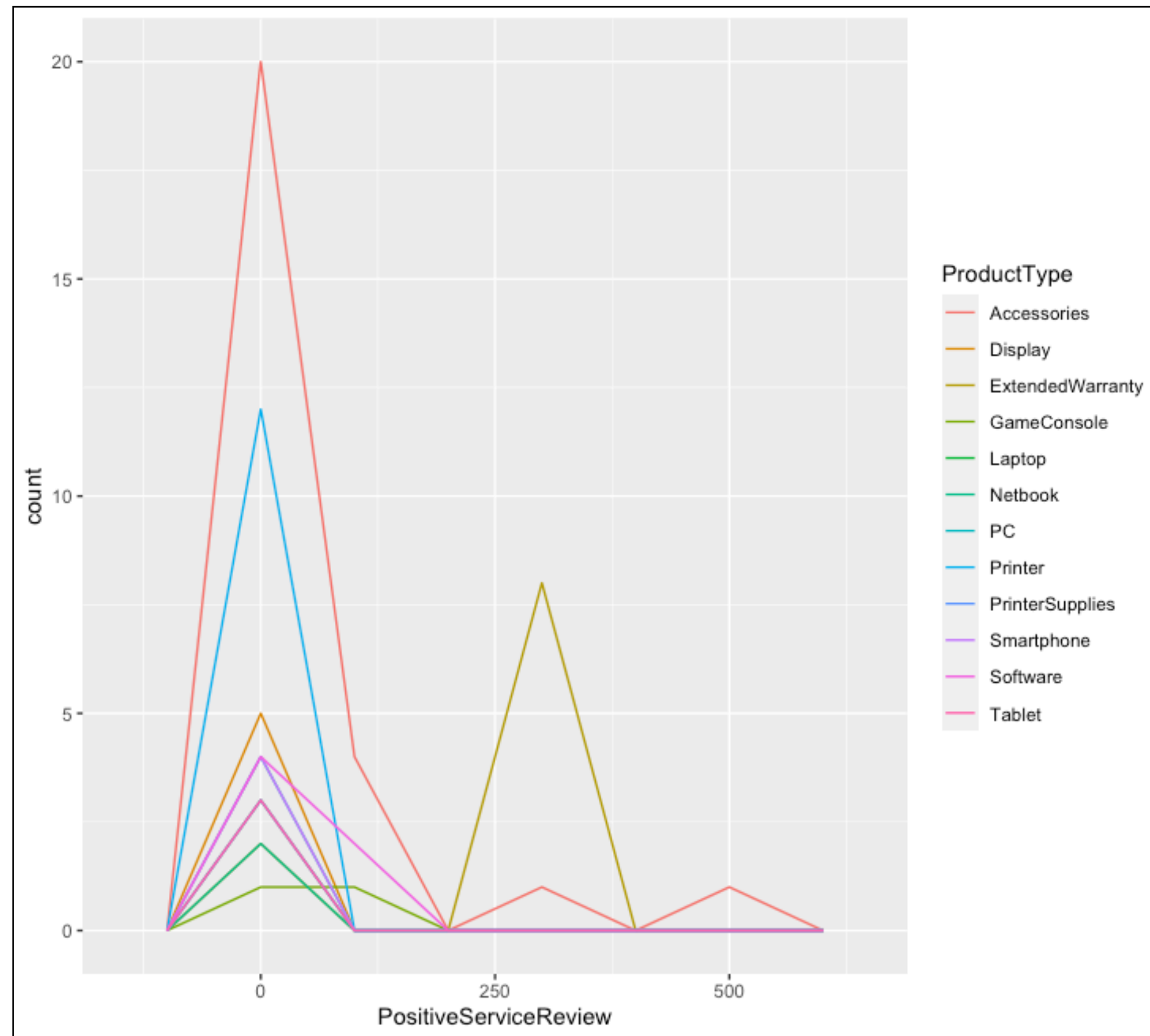


Figure 4

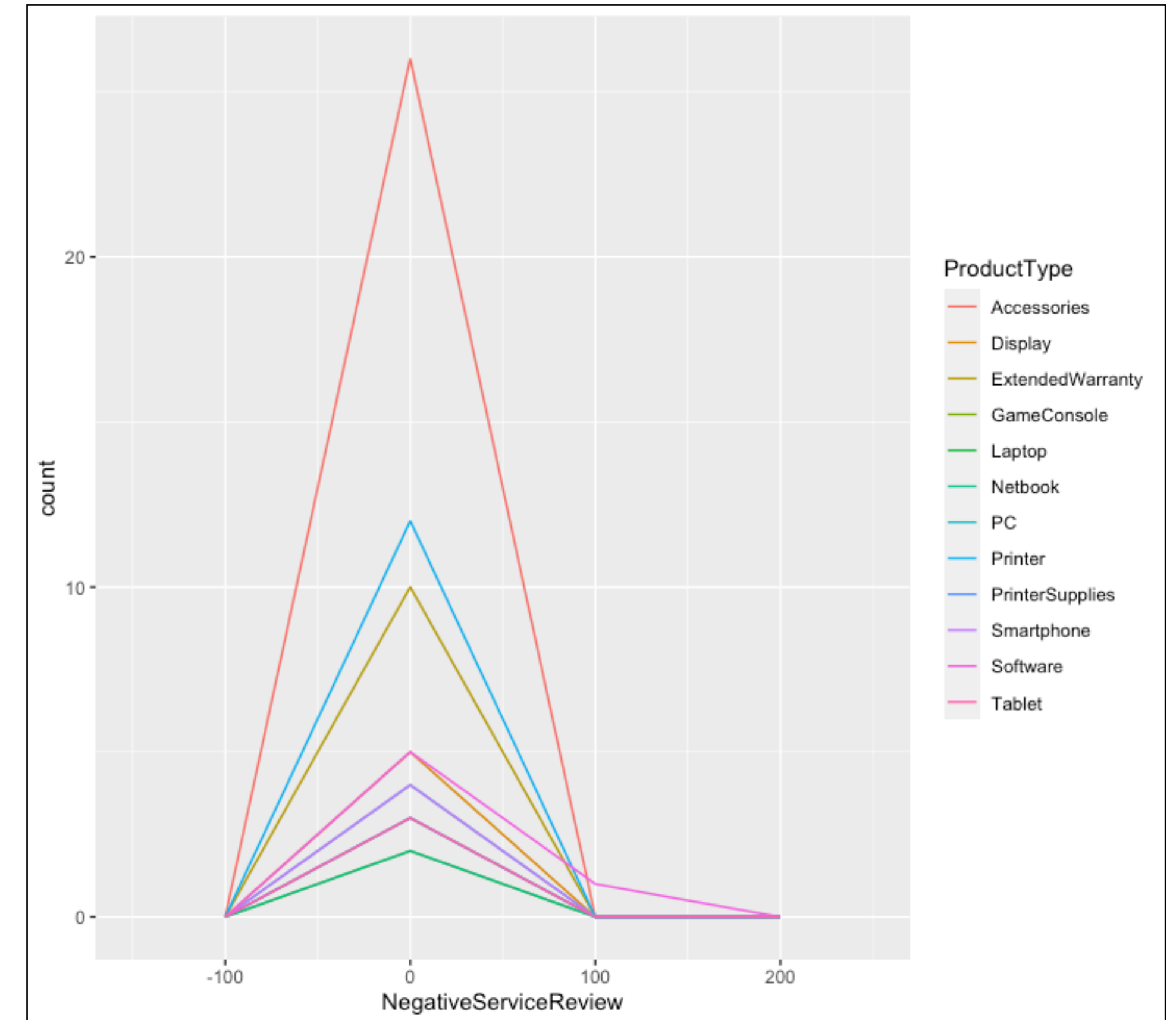


Figure 5

Looking at Figure 4, Accessory, Extended Warranty, Tablet, and Software have gotten the highest positive reviews from customers. However, in Figure 5, we see that they mostly got under 100 negative reviews.

In the correlation matrix, we can see the relationship between predictors (features) themselves, and the relationship they have with the ground truth which is Volume.

Multicollinearity: 5StarReviews has a really high correlation with Volume which is concerning. We tried *Combolnfo* function, and it too warned us about 5StarReviews variable. Hence, we decided to omit it from the analysis process.

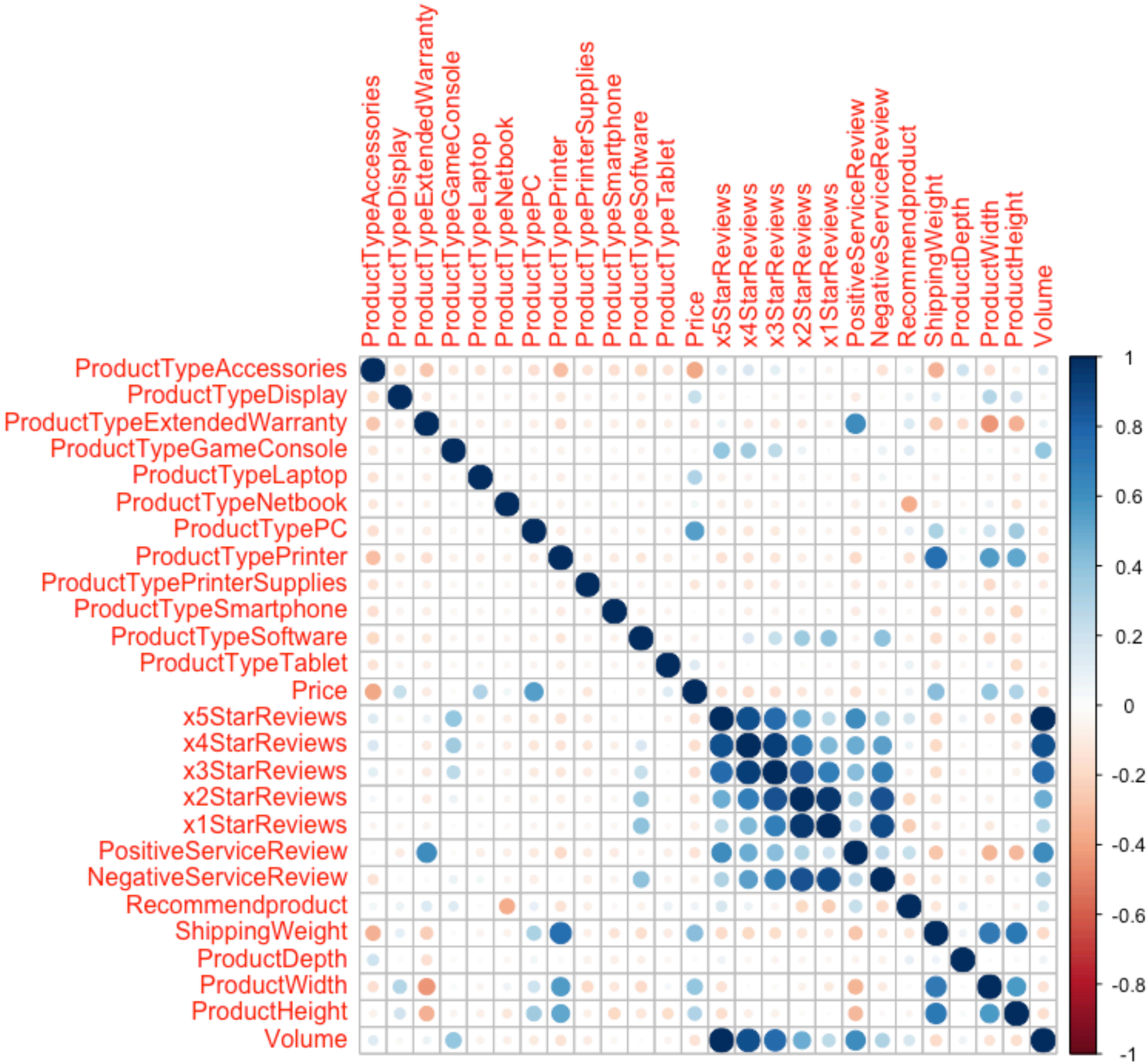


Figure 6. Correlation Heatmap

RANDOM FOREST CLASSIFIER

By running *postResample* on the test set, we got 297 for the RMSE and 0.76 for the R-Squared for this model.

