*Analyzes the “Adidas” dataset in R with barplots and dataset manipulation.*

**Assignment**

**5**

A5

ALY6110 Data Management & Big Data

Assignment 5 – Individual Lab #2

**PREPERATION:**

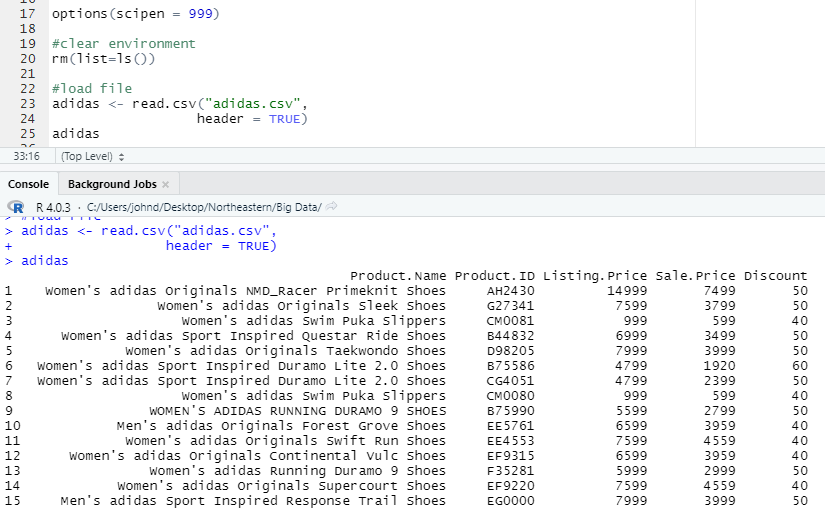
By: John DiSessa

For: Professor Mako

On: October 23rd, 2022

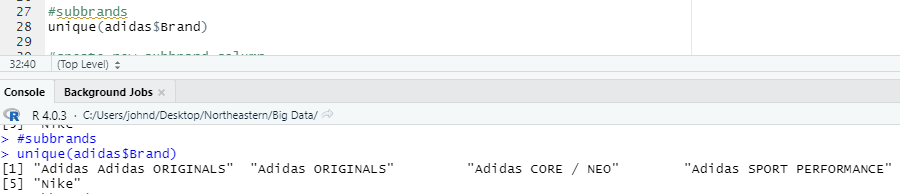
Load the Dataset

After clearing my environment from all previous work, I loaded in the Adidas dataset into R and did a quick check for glaring errors or issues with the column names.



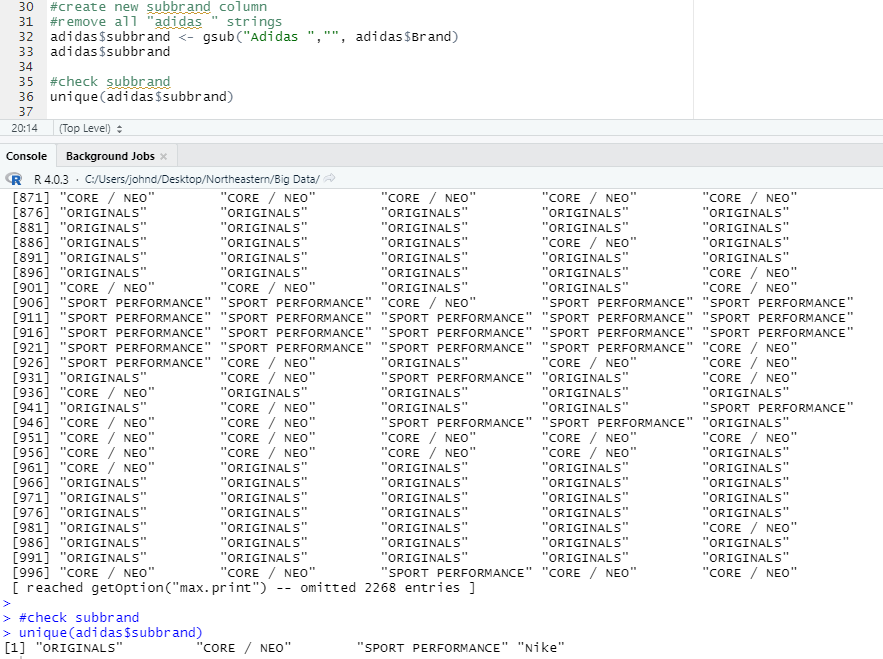
Identify Sub Brands

In order to identify all of the sub brands in the dataset, I listed all of the unique values from the Brand column. Despite having over 3,000 different products, they are all classified within these 5 sub brands.



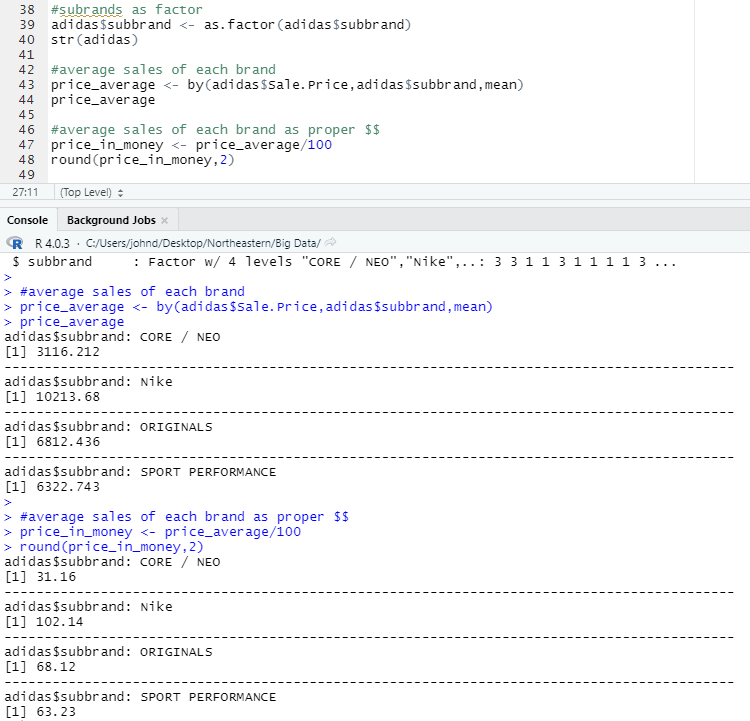
Brand Name Errors

As seen from the previous image, there are two sub brands that need to be combined into one sub brand: Adidas Adidas Originals and Adidas Originals. There is clearly a typo with “Adidas” being listed twice. For this step, I removed both the extra “Adidas” as well as reduced the sub brand names by removing “Adidas” from all of them to clean up and create a new Sub Brand column. Additionally, I noticed one of the sub brands is “Nike” which is a different company than “Adidas”. However, given the lack of details in the assignment about project goals, and considering that the Nike sub brand made up about 20% of the dataset, I decided to keep the Nike data points in the dataset.



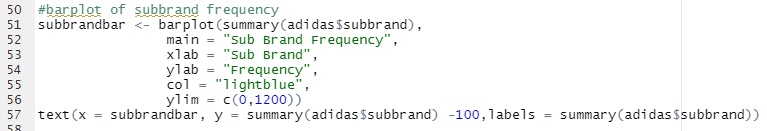
Average Brand Sales Price

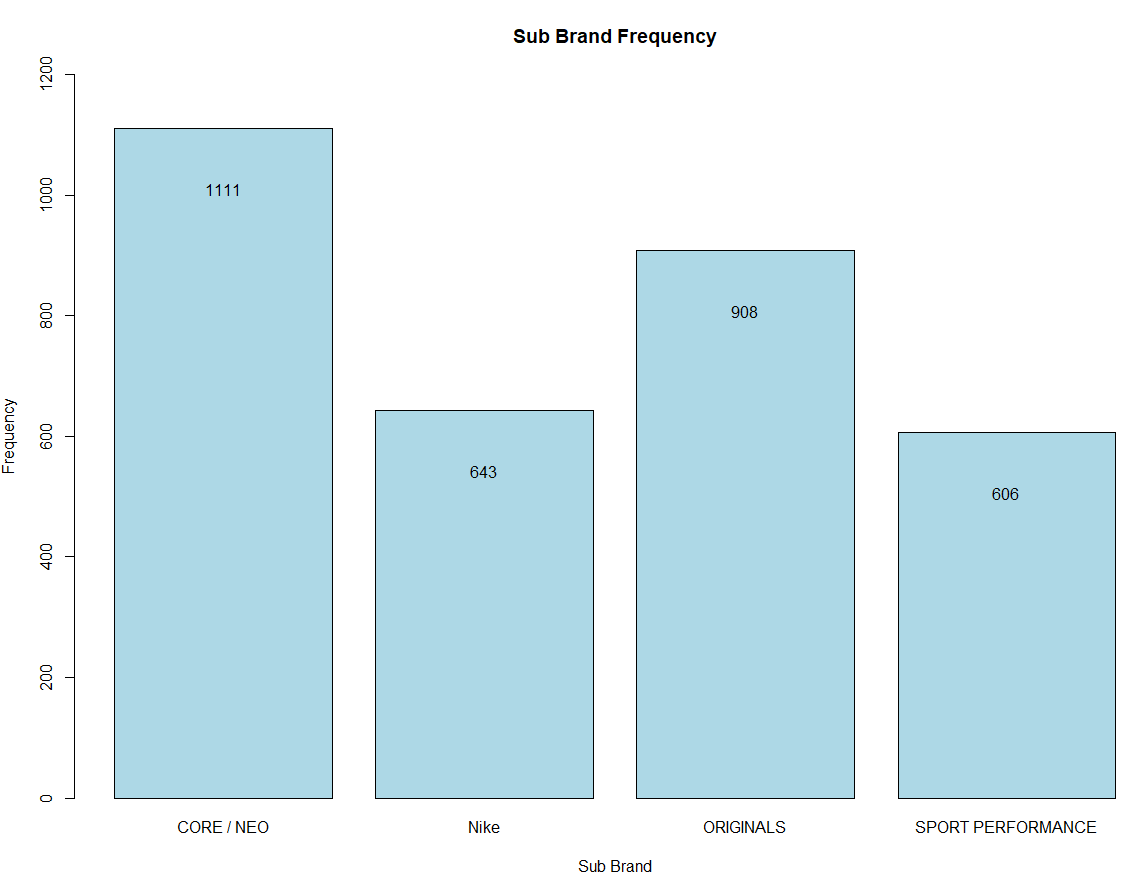
Even though the dataset had both the original list price and the sale price of each product, I focused only on the sale price in order to be conservative with our calculations. I used the “by” formula to find the mean sale price per sub brand. However, the sale prices were listed as whole numbers instead of including the cents as decimals. Most of the sale prices ended with 99, so it was fair to assume the sale prices should have been read as “27.99” instead of “2799”, for example. I divided the results by 100 and rounded to two decimal places to get the average sale prices of each sub brand. Nike had the largest average sale price of $102.14 while the Adidas Core / NEO brand had the cheapest average sale price of only $13.16.

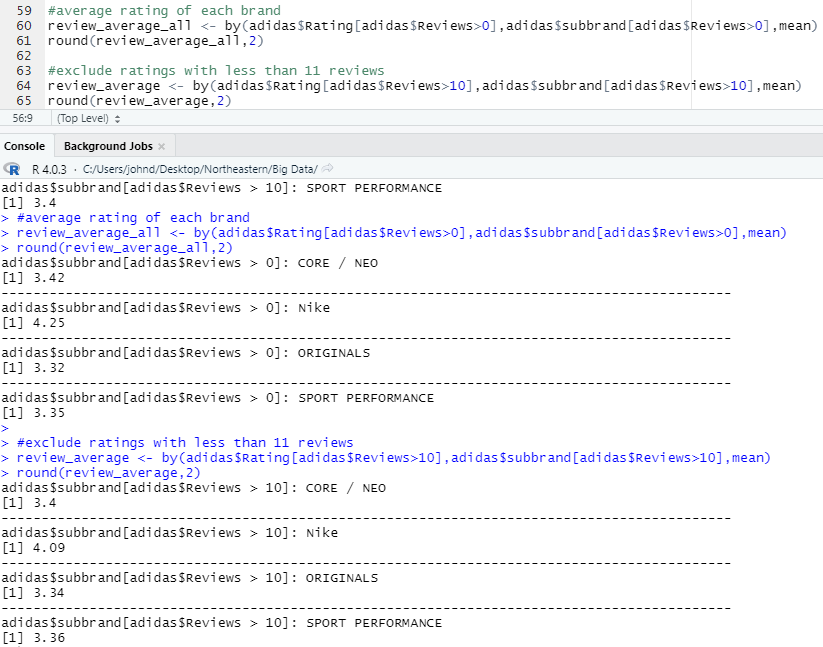


Sub Brand Popularity

I first created a bar plot to see which sub brand had the most products available for sale. The Core / NEO sub brand had the most products while the Sport Performance sub brand had the least. In some ways, that could indicate popularity since Adidas would sell and create more products for their best-selling brands. However, I also looked at the ratings and reviews of each brand to learn about customer opinions since that has a direct impact on popularity as well. I first calculated the average customer rating per sub brand, by only factoring in products with at least 1 customer review. Since many products had no reviews, their ratings would count as 0 and skew the averages. However, many products only had 1 or 2 reviews. I then calculated the average customer rating by only factoring in products with more than 10 reviews.







Conclusion

Without data regarding the sales of each product or the cost to make and market each product, it is very difficult to make suggestions about how Adidas should allocate their resources. However, all 3 Adidas brands have inferior customer ratings compared to Nike. The 3 Adidas brands are all similarly rated, but Nike reigns supreme. If I were an executive or consultant for Adidas, I would recommend conducting more market research to figure out how to catch up to Nike in terms of customer satisfaction. This dataset is a great start for Adidas market research, but it only sheds light on how far behind they are compared to Nike. Future market research could include customer segmentation to learn who to better target products for or even a specific product by product analysis of Nike sales to find their weaknesses and take advantage.

Citations

“Bar Plot in R Using Barplot() Function.” *DataMentor*, 8 Oct. 2018, https://www.datamentor.io/r-programming/bar-plot/.

wesleysc352, et al. “How to Add Labels above the Bar of ‘Barplot’ Graphics?” *Stack Overflow*, 1 Mar. 1968, https://stackoverflow.com/questions/65057352/how-to-add-labels-above-the-bar-of-barplot-graphics.