**Executive summary**

For this experiment, we were required to find 100 SKU associations for the company, Dillard’s, such that they can select 20 “moves” around the store to maximize customer purchases. Whether they find that it is beneficial to keep SKUs that are associated with each other close to each other or further away from each other, having this information on hand will be useful for business decisions later on.

After understanding the data, I subsetted the dataset into the state of California and selected the largest store in the state to conduct the analysis on. Subsetting the data was required given the size of the dataset. California was chosen as the state of choice as after some research, California tends to have the largest departmental stores and focusing in this area would lead to the greatest probability of an increase in customer revenue. Additionally, exploring the data allowed me to understand key insights within the data, including the range in prices within the store; this allowed me to see how many of the more expensive items were purchased and assisted me in choosing my minimum support value.

Prior to the association experiment, I selected 1000 randomly selected distinct SKUs for the analysis. Running more than 1000 SKUs resulted in computation failures and hence, the random selection of 1000 SKUs yielded a fair way to subset the SKUs. I separated this further into training and testing sets in order to validate whether the associations I selected appeared in the test group. After conducting the analysis and checking against the test group, it was noticed that two associations appeared in both groups. Hence, in the final data frame, these two associations were included along with every other association.

The top 100 SKU associations (ordered by lift, confidence and support) were added in the final data frame and I suggest looking into moving these SKUs in the planogram to maximize customer purchase. The top 5 SKU associations are seen below:

