Data Wrangling and Analysis Process –

Goal1– To determine if there is a cost benefit of purchasing various product categories at Amazon or in the physical store

Goal2 – To determine if sale items have better pricing on-line or in the store

Goal3 – Determine if there is bigger price gap between online and retail correlates to difference between the time the item was compared online and in store

Data Wrangling Approach –

* Import data original\_amazon\_compare.csv into R
* Use dplyr library to wrangle data set

The data contains many columns (variables) that are not needed for my analysis so I will first clean the data table and only select the columns I need...

price – price in store

price\_amazon – price in amazon

sale\_online – the item was on sale online

PRICETYPE – if the item was on sale in store

datediff – difference between time purchase online – in store

category – groupings of product type (electronics, office supply, etc.)

Next I need to create using mutate a column that has the price difference between on-line and in store pricing

Then I should graph the results to see outliner and fix errors ....

\*\*\*\* will R plots tell me what row to look for our should I do a filter???

Or should I do a select If delta row is > than...???