Documentation for data cleaning in Excel/SQL

1. Deleted all identifying information (names, location, billing info, etc)
2. Trimming/cleaning columns to be more recognizable
   1. Changing Emergency Department AZ in the category problem to the correct Mens/Ladies label
   2. Text to columns on product\_id to have only the item id that is used (changing to just the 4 letter/number combo)
   3. Removing names from department and cleaning the names so each department is spell and presented the same
   4. Removing the “choose color;” and “choose size;” from their respective columns and removes “extra pricing” on the size columns for bigger sizes since the extra dollar amounts are included in the price column.
   5. Removed “not available” labels and logo color labels from color column
3. Made the product\_name column all caps to keep consistent with how most of the data was pulled
4. Deleted all cancelled order and special “non group orders”
5. Created a primary key column with line\_item because some orders have more than one item apart of it and they are counted as separate rows
6. Created a month\_submitted columns to identify the different months at a glance and sort them easier