### **Project: Capstone Project 1 - Data Wrangling**

For this capstone project I obtained the Google Analytics keyword data from MELA Handbags. I was able to export this data using a csv file. The first step I took in data wrangling was to change the actual company name to MELA to mask the data. I did this within pandas by replacing all strings of the actual company name with MELA then wrote this data back to a csv file.

Next, I used the masked data to create a pandas dataframe. I then cleaned the data by removing all commas from Clicks and Sessions variables, removing all commas and dollar signs from the Cost, Cost per Click (CPC), and Revenue variables, and removing the percent sign from Bounce Rate and Ecommerce Conversion Rate variables. This allowed me to then convert these columns along with the Transactions variable to floats.

Then, I computed several new variables using the ones existing in the dataset. I computed the cost of each keyword per revenue the keyword produced. I also computed a Return on Investment (ROI) variable which is the difference between the revenue and cost divided by the cost of the ad. Additionally, I computed the sessions the keyword produced per the cost of the keyword. In creating these new variables I created values that were infinite or not a number. I decided to leave these value as is and use them for further analysis.

Finally, I created a categorical variable for the Bounce Rate. Upon researching bounce rates I found that any bounce rate under 25% is considered very excellent, anything between 26 and 40% bounce rate is considered excellent, between 41 and 55% bounce rates are considered average, between 56 and 70% bounce rate would be considered higher than average, and bounce rates above 70% are considered high and unideal. In creating this variable I created 447 null values since these keywords had a bounce rate of 0% due to no clicks on the keyword.

I did not exclude any data for being an outlier.