**Instructions for Filtering Data from Snapchat Political Ads**

1. Download 2018 data from Snapchat and extract the raw data.
2. Apply a filter to the column headers. Keep the **Impressions** column as the independent variable, and keep the **Spend**, **Start Date, End Date, Age Bracket**, and **Country Code** columns. Delete the rest.
3. Insert two columns to the left of the **Start Date** column.
4. Create a separate worksheet and call it “**reference\_age**.” List all the age ranges from the **Age Bracket** column and categorize it according to an age range. Ad age ranges including ages under 18 are assigned a 1. Ads age ranges 18+ are assigned a 2. Ads with no age range specified are assigned a 0. If an ad age range includes ages under and over 18, then take the average of 1+2 and assign that ad a score of 1.5.
5. In one of the empty columns from step 3, rename it as **Age Range**. Use the VLOOKUP command with the table from the **reference\_age** worksheet to assign a 0, 1, 1.5, or 2 to each ad.
6. In the other empty column from step 3, rename it as **Run Time (days)**. In the same row, find the corresponding values from the **End Date** and the **Start Date** columns. Subtract the **Start Date** value from the **End Date** value.
   1. Note: Some advertisers put their end date in the **Start Date** category and their start date in the **End Date** category. Filter the **Run Time** column and select only error values. Correct the error by reversing the calculation in step 6 – subtract the value in the **End Date** column from the value in the **Start Date** column.
7. If a row does not have an end date, create a column that’s called **No End Date**. Sort the **End Date** column and select to show only the rows that have a blank value. These are the rows that don’t specify an end date. In these rows, put a 1 in the **No End Date** column.
   1. Sort the **No End Date** column from high to low and delete all the rows that have a 1.
8. Select the **Run Time** column and assign it the “Accounting” number format.
9. Reorder the columns so that the **Impressions** column is on the far left and that the **Spend, Age Range,** and **Run Time** columns are next to each other.
10. Use the Data Analysis ToolPak and conduct a correlation analysis for the **Spend, Age Range,** and **Run Time** columns. Name the new worksheet as **“correlation”** and check to make sure that the three variables do not influence each other.
11. Use the Data Analysis ToolPak and conduct a multiple linear regression. The **Spend, Age Range,** and **Run Time** columns are the dependent variables. The **Impressions** column is the independent variable.
12. Create a new worksheet called **“predicting impressions”** and put together the formula for predicting impressions. The intercept and coefficients for each of the dependent variables are from the **multiple linear regression** worksheet.
    1. Note: The p-values for any of the dependent variables should be <0.001. If the p-value is higher than that, do not use that variable in the final formula to predict impressions.
13. Repeat steps 1-12 for the Snapchat 2020 Political Ads data.