Step-by-Step Excel Data Manipulation

Data Manipulation

1. Downloaded political advertisement data for 2019 from the Snapchat website
2. Filtered for USD only and U.S. only (U.S., Argentina, Canada, Denmark, Germany, India, Kuwait, Norway, Puerto Rico, South Africa, Sweden, Turkey, UAE originally included)
3. Used Text to Columns tool to separate the date and time for both Start Date and End Date
4. Subtracted the difference between Start Date and End Date in Days_Ad_Up column
5. Filtered out the advertisements that were still running by the end of 2019 with no end date
6. Formatted cells to Numbers (to change date format to number of days)

Multiple Linear Regression Model

1. Used Data Analysis tool to perform a multiple linear regression
2. Selected “Impressions” as known Y with “Spend” and “Days Ad Up” as known X

Simple Linear Regression Graph and Finding Outliers

1. “Impressions” was dependent variable and “Spend” was independent variable
2. Created a scatter graph that included labeled best fit line and R2 value
3. Wrote predicted impressions equation to include predicted impressions column based on slope and intercept values
4. Created Error column by subtracting Impressions - Predicted Impressions
5. Found outliers through the use of an IF statement
   =IF(ABS(errorcolumn)>2*STEYX,"Outlier","Not Outlier")
6. Found total percent of outliers using a COUNTIF statement to find total number of "Outlier" and "Not Outlier"

Simple Histogram

1. Calculated maximum (=MAX), minimum (=MIN)
2. Created bins for “Spend” and “Impressions” to calculate frequency (=FREQUENCY)
3. Inserted bar graphs for both “Spend” and “Impression”
4. Changed gap width to 0 to resemble traditional histograms and changed the horizontal x-axis of histogram graph to reflect bins amount