**Data Documentation Sheet**

**Data Process for EG Assessment**

­Step 1

**Defining the question**

* Define the questions(already provided).

Step 2

**Data Collection**

* Collect the data(already provided).
* Create a copy of the original dataset to be used for analysis.

Step 3:

**Data Cleaning for Data Validation**

* Empty spaces in all columns are removed using the find & replace tool.
* To ensure data integrity, records that display engagement rates higher than 1(100%) are deleted. To prevent the loss of data it is usually advisable to update and correct the record instead of deleting it. However, there are no indications of what the true values for total impressions are for those records so therefore they are deleted using the filter tool.

Step 4:

**Data Analysis**

* Determining engagement rates and likelihood of engagement rates of 15%
  + See assessment answers file
* Determining how day of week and time of day affect engagement rates
  + See assessment answers file
* Determining best performing game titles
  + See assessment answers file
* Determining best performing media type
  + See assessment answers file
* Determining best performing campaign
  + See assessment answers file
* Defining a posting strategy
  + See assessment answers file
* Making suggestions about where to expand
  + See assessment answers file
* Outlier Detection
  + After creating a separate tab, I used quartile functions to determine records that deviate from the rest in terms of engagements.
  + I used QUARTILE.INC function to determine the 1st and 3rd quartile.
  + Inter Quartile Range(IQR) is determined by finding the difference between the 1st and 3rd quartile.
  + Lower Limit is found using 1st Quartile – 1.5 \* IQR.
  + Upper Limit is found using 3rd Quartile + 1.5 \* IQR.
  + I defined an outlier as any value that is outside of the lower and upper limit
* Sensitivity Analysis
  + In order to determine the impact of outliers, I will be performing the same analysis with and without outliers.
  + After making a copy of the pivot table, I will remove outliers and perform the same analysis.
  + The second analysis indeed shows that deleting outliers made a large difference in values and thus the analysis.
  + After much thought, I determined that outliers are indeed necessary to include due to the fact that there are large variations in total engagement and engagement rates for days and time; therefore, outliers should be expected.

Step 5

**Reporting Results**

* See Github repo