Project Details

The project goal is the generation of a point-plot line graph based on a formatted data set. This data set contains shock/vibration samples taken by an electronic logging tool working in the oil-and-gas industry. The tool generates this data in a medium to large Excel spreadsheet.

The following image depicts an example of the spreadsheet data as it appears directly after the user downloads it from the electronic tool (figure 1).

Figure 1. – Raw dataset example

As shown in the example, the dataset contains a date/time field associated with multiple shock/vibration values. The data relevant to this project is the date/time (Excel column B), the XYVibe (Excel column H), and the Zvib (Excel column K). The XYVibe and Zvib columns contain vibration force samples measured along three axis of the logging tool using the tool's internal accelerometers. The image below highlights the relevant fields in yellow (figure 2).

Figure 2. – Highlighted relevant data columns

The final graph produced by the script will not contain data from the entire dataset but will be limited by a start time and end time. These times should be enterable by the user before processing of the data begins. The time period entered by the user will vary in length between 30 minutes and 150 minutes. All data outside of this designated period will be irrelevant to this project. It is also important to note that the date will be relevant in some cases. The designated time period may crossover from one day to the next so the user will need to specify a date for both the start time and end time.

The following example shows a highlighted dataset in which the user may be interested in graphing. The sampling data starts at 23:52:10 on Feb 26 and ends at 00:04:10 on Feb 27 (figure 3).

242	2021/02/26 23:51:10	19940	3	0	0	25	0.571812	2.269231	0	0.421000	0	0	0	0	0	0
243	2021/02/26 23:51:40	49537	3	0	0	25	0.781835	2.192308	0	0.762119	2.730769	0	0	0	0	0
244	2021/02/26 23:52:10	7120	3	0	0	25	0.977363	1.846154	0	1.012283	2.846154	0	0	0	0	0
245	2021/02/26 23:52:40	31566	3	0	0	25	0.970583	1.807692	0	0.996472	2.538462	0	0	0	0	0
246	2021/02/26 23:53:10	9306	3	0	0	25	0.94571	1.615385	0	0.988927	1.730769	0	0	0	0	0
247	2021/02/26 23:53:40	59309	3	0	0	25	0.575179	2	0	0.644296	2.346154	0	0	0	0	0
248	2021/02/26 23:54:10	38231	3	0	0	25	0.273415	0	0	0.281336	1.5	0	0	0	0	0
249	2021/02/26 23:54:40	42808	3	0	0	25	0.483892	2.653846	0	0.508286	1.923077	0	0	0	0	0
250	2021/02/26 23:55:10	9542	3	0	0	25	0.270657	0	0	0.290997	1.5	0	0	0	0	0
251	2021/02/26 23:55:40	44594	3	0	0	25	0.273692	0	0	0.28823	0	0	0	0	0	0
252	2021/02/26 23:56:10	56848	3	0	0	25	NaN	NaN	65535	NaN	NaN	65535	255	255	255	255
253	2021/02/26 23:56:40	13932	3	0	0	25	0.273133	0	0	0.283785	0	0	0	0	0	0
254	2021/02/26 23:57:10	35016	3	0	0	25	0.265015	0	0	0.291687	0	0	0	0	0	0
255	2021/02/26 23:57:40	8744	3	0	0	25	0.270784	1.538462	0	0.289314	0	0	0	0	0	0
256	2021/02/26 23:58:10	6977	3	0	0	25	0.268659	0	0	0.290401	1.730769	0	0	0	0	0
257	2021/02/26 23:58:40	58148	3	0	0	25	0.26387	0	0	0.296748	1.615385	0	0	0	0	0
258	2021/02/26 23:59:10	24855	3	0	0	25	0.266445	0	0	0.292507	1.538462	0	0	0	0	0
259	2021/02/27 00:00:10	63937	3	0	0	25	0.264193	2.076923	0	0.293047	1.923077	0	0	0	0	0
260	2021/02/27 00:00:40	60955	3	0	0	25	0.26651	0	0	0.292173	0	0	0	0	0	0
261	2021/02/27 00:01:10	51634	3	0	0	25	0.262429	1.653846	0	0.292639	1.884615	0	0	0	0	0
262	2021/02/27 00:01:40	29154	3	0	0	25	0.260808	0	0	0.294407	0	0	0	0	0	0
263	2021/02/27 00:02:10	10079	3	0	0	25	0.258114	0	0	0.29619	1.884615	0	0	0	0	0
264	2021/02/27 00:02:40	12975	3	0	0	25	0.265539	0	0	0.286686	0	0	0	0	0	0
265	2021/02/27 00:03:10	39346	3	0	0	25	0.267091	0	0	0.284938	0	0	0	0	0	0
266	2021/02/27 00:03:40	20308	3	0	0	25	0.261725	0	0	0.293007	1.576923	0	0	0	0	0
267	2021/02/27 00:04:10	35113	3	0	0	25	0.258058	0	0	0.296034	1.807692	0	0	0	0	0
268	2021/02/27 00:04:40	1689	3	0	0	25	0.252741	0	0	0.302291	0	0	0	0	0	0

Figure 3. – Time period designated by user that crosses dates

When the user enters time data, they should only be required to enter hour and minute. The script should start and end the graph from the sampling times nearest to what the user inputs. For example: the user may input a start time of 23:52 and an end time of 00:04 in the above example.

The dataset will sometimes contain shock/vibe values that are null. These lines are generated as an artifact of system resets performed by the tool for various reasons. These lines of data should be removed from the data used to create the final graph output. Leaving these null values in will create inconsistencies in the graph which complicates any meaningful analysis. These null lines will always have the same formatting. The shock/vibe value will be replaced by a string "NaN".

The below example demonstrates data containing these null values (figure 4.)

32	24	2021/01/	35129	3	0	0	25	0.69933	0	0	0.744103	2.115385	0	0	0	0	0	0
33	25	2021/01/	7499	3	0	0	25	0.698134	1.576923	0	0.74941	2	0	0	0	0	0	0
34	26	2021/01/	28406	3	0	0	25	0.696691	1.692308	0	0.747472	2.038462	0	0	0	0	0	0
35	27	2021/01/	43554	3	0	0	25	0.708387	2.346154	0	0.737373	1.769231	0	0	0	0	0	0
36	28	2021/01/	36352	3	0	0	25	NaN	NaN	65535	NaN	NaN	65535	255	255	255	255	255
37	29	2021/01/	55463	3	0	0	25	0.697855	2.076923	0	0.745113	2.153846	0	0	0	0	0	0
38	30	2021/01/	547	3	0	0	25	0.697123	0	0	0.744427	0	0	0	0	0	0	0
39	31	2021/01/	35130	3	0	0	25	0.700274	1.576923	0	0.739087	1.769231	0	0	0	0	0	0
40	32	2021/01/	33401	3	0	0	25	0.696271	0	0	0.745062	1.692308	0	0	0	0	0	0
41	33	2021/01/	63218	3	0	0	25	0.700095	0	0	0.739485	2.076923	0	0	0	0	0	0
42	34	2021/01/	20037	3	0	0	25	0.692634	1.5	0	0.747968	2	0	0	0	0	0	0
43	35	2021/01/	63357	3	0	0	25	0.698714	2.115385	0	0.741031	2.576923	0	0	0	0	0	0
44	36	2021/01/	19153	3	0	0	25	0.696115	2.269231	0	0.743478	1.807692	0	0	0	0	0	0
45	37	2021/01/	9198	3	0	0	25	0.696972	0	0	0.745351	2	0	0	0	0	0	0
46	38	2021/01/	50930	3	0	0	25	0.691348	1.692308	0	0.748128	1.884615	0	0	0	0	0	0
47	39	2021/01/	90	3	0	0	25	NaN	NaN	65535	NaN	NaN	65535	255	255	255	255	255
48	40	2021/01/	15198	3	0	0	25	0.694981	1.769231	0	0.743175	1.769231	0	0	0	0	0	0
49	41	2021/01/	39528	3	0	0	25	0.689481	0	0	0.744087	2.423077	0	0	0	0	0	0
50	42	2021/01/	22741	3	0	0	25	0.68815	0	0	0.748089	2.153846	0	0	0	0	0	0
51	43	2021/01/	61925	3	0	0	25	0.690507	2.115385	0	0.745935	0	0	0	0	0	0	0
52	44	2021/01/	57171	3	0	0	25	0.692605	0	0	0.74764	0	0	0	0	0	0	0

Figure 4. – Example highlighting null values in dataset

The script must output a graph with the Y-axis representing shock value and the X-axis representing time in minutes. Each column of shock/vibe data (column H and column K) should be an independent line plot. The data curves should be different colors (preferably orange and blue) and should be labeled somewhere on the graph.

The following example shows a graph created in Excel (figure 5).
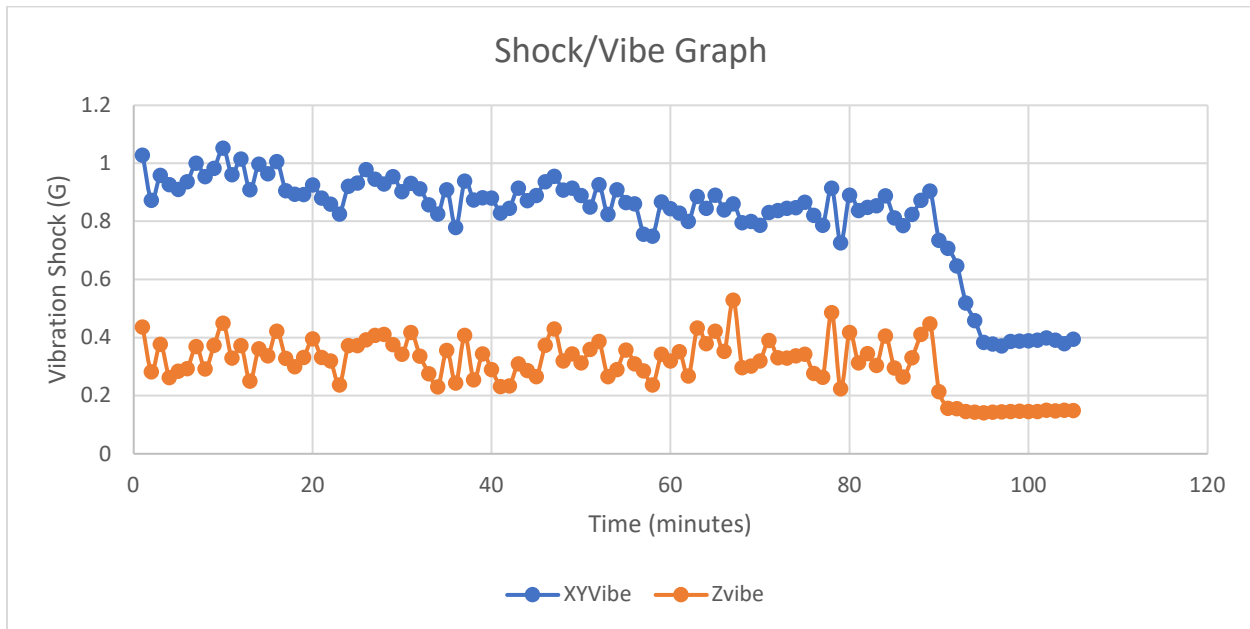
Shock/Vibe Graph

Figure 5. – Example of possible graph output

The start time and end time of the final graph should be controlled by the user input times. The y-axis limits should be reasonable per the data represented.

The ultimate user experience should be as follows:

1. User downloads the Excel spreadsheet from the logging tool and saves it to their computer.
2. User initiates the script.
3. Script requests a file location, and the user selects the Excel spreadsheet in file explorer.
4. Script requests user input start date, start time, end date, end time.
5. Script generates graph of shock/vibe data versus time.