*How to Use the Python Script*

Download the file using this link:

<https://github.com/cartershields/KVC-Data-Cleaning>

Download the data using this link:

<https://www.anecdata.org/projects/view/477>

It is recommended to use Jupyter Notebook to run the script as the .ipynb file. In Jupyter Notebook, the script can be run block-by-block. This allows you to only run the parts that you want and leave the rest. All of the cleaning blocks must be run no matter what, but if you do not want the summaries throughout, or to continue with the pivot tables, you are able to do so. If you need to run the script as a .py file in the Idle Shell or another application, it still works, it just will give you more information than needed, or be a bit confusing to a beginner Python user.

If you are an advanced enough Python user to run it as an entire file in the shell, once you press run, press close on the pop-up. Paste your file path name and press enter. Ignore the red warning text. Before the red chunk of words, you will see the quantity of observations and columns before cleaning, and after the red chunk of words you will see the quantity of observations and columns after cleaning. You will then be prompted to enter CSV or JSON depending what file type you want the cleaned data exported as. I advise to use CSV as you can then import it to QGIS. Type CSV in all caps and press enter. Now you will see all the summaries from the script printed in a difficult to read format. Hence why I recommend using Jupyter Notebook. You will now also be able to see the 3 other useful files created by the script in addition to the cleaned dataset: parent pivot table, child pivot table, and volume summary.

Once you have the .ipynb file opened in a Jupyter Notebook, you are ready to go. The file has a lot of comments in it, signified with a ‘#’ symbol at the beginning of the line. You should read all comments as you go to understand what each block is doing. To execute each block press run at the top of the screen. If you read the comments as you go, it should be relatively self-explanatory.