Version One

**Automated-Labeling-Tool**

This automated labeling tool was created to streamline the data scrubbing process which looks for common words used in an excel file and replaces empty or N/A labels cells with the correct information.

**Introduction**

This repository displays the automated labeling tool that I created to help me in my day job as a Data Analyst for The Disneyland Resort. **The data used for this repository is public data of IMDB’s Disney+ shows’ information. The reason is I’m not able to use the actual excel file due to its sensitive nature.** This tool was created to help deal with my ongoing process of data scrubbing by updating cells with the correct labels and filling in empty cells in an automated way. I have included the original excel file with the python file of the automated labeling tool.

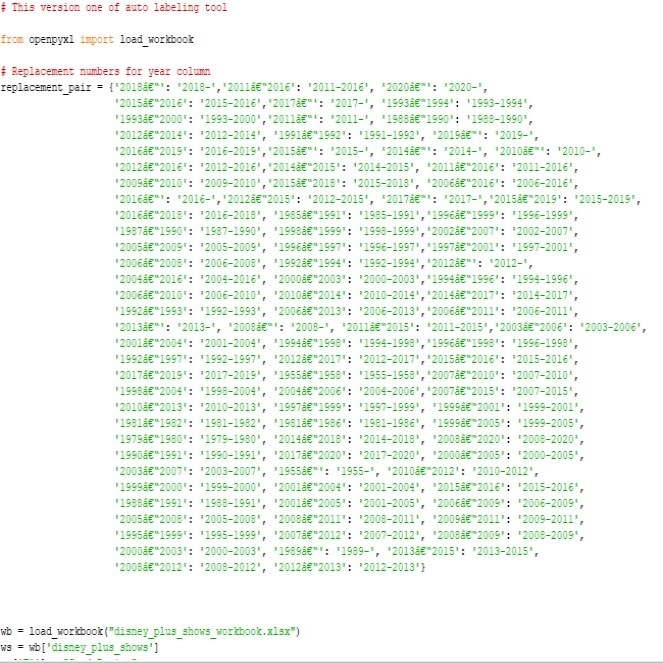
**Abstract**

For anyone who deals with data, it is understood that most of your time will be spent cleaning data (e.g., data scrubbing) and arranging the data to find trends. In my day as a Data Analyst, the original process of data scrubbing took about five hours to complete. After researching, a solution was found to use python to work with Microsoft Excel and help streamline and cut down the time spent on data scrubbing. This tool has saved my department hours per week, which would be spent on data scrubbing.

**Summary of Skills**

This tool was created in the python programming language and uses the Openpyxl python module to work with Microsoft Excel. I have python 3.10 as my environment and it can work with any python 3 environment.

**Preview**

[](https://github.com/micgonzalez/Automated-Labeling-Tool/blob/main/Automated_Labeling_Tool_Images/preview%20of%20version%2001%20tool.jpg)

This is a preview of the code for version 1 of Automated Labeling tool.

**Findings**

It was amazing to find out about python modules that allow you to work with excel files. The use of the Openpyxl module made a world of difference for me. It allowed me to create this automated labeling tool to help in data scrubbing. There is still more room for features that I want to include for this tool.

**Challenges**

The main challenge for creating this automated labeling tool was the automation part. It took me a while to figure out the code to have the tool look at the file and correctly label the N/A or empty cells. Another challenge was to gather all the needed information to create lists and references for the tool to compare and update the cells with the correct information.

**Conclusion**

This automated labeling tool is a great start that has real world implications in my day job. I do see other features that could be implemented in the future. One feature would allow the tool to compare multiple line string with a list and/or reference word to update cells or create new rows with the correct labels with minimal supervision.