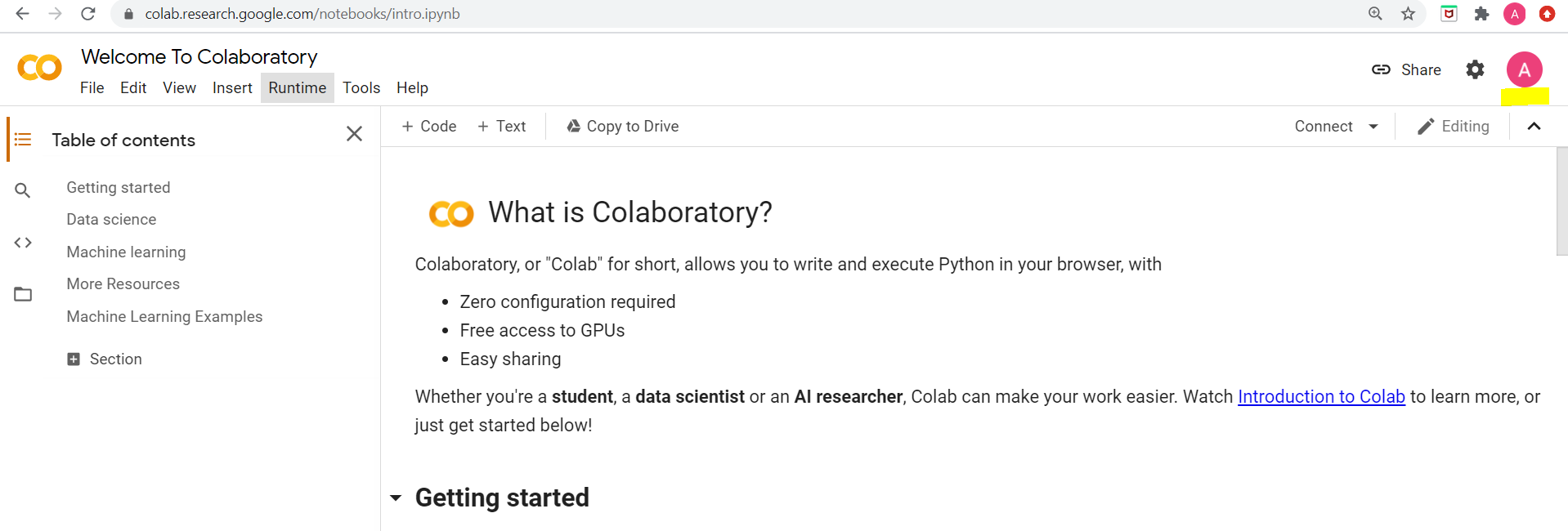
**User Steps Manual**

Team 3 | 2020

**Using Data Analytics Extracts Top Keywords and Trends in Information Technologies**

V1

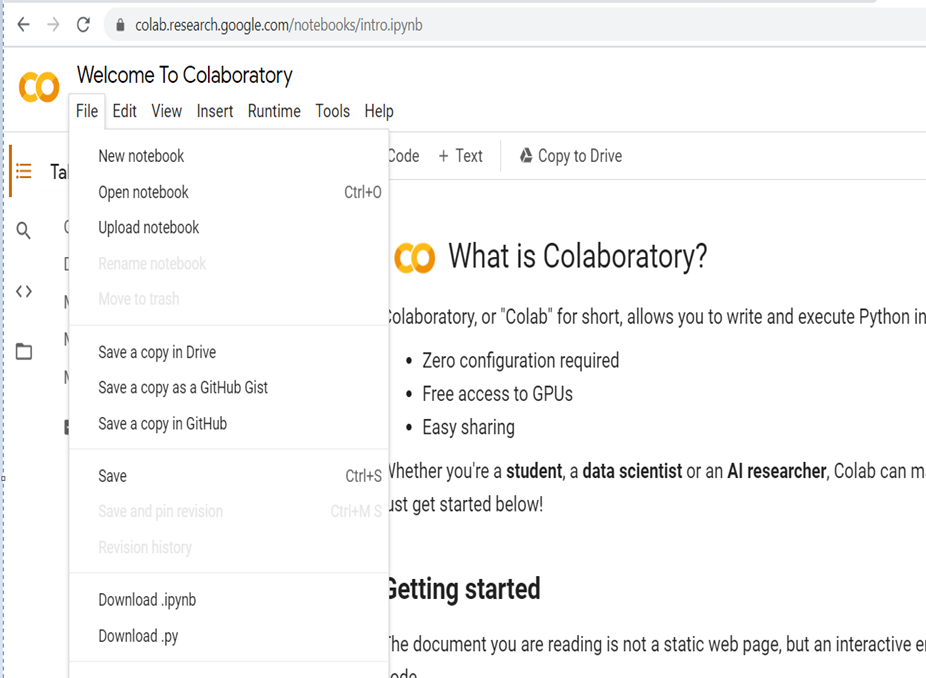
**STEP 1:** Log in to your Google Colaboratory account. *(*[*https://colab.research.google.com/notebooks/intro.ipynb*](https://colab.research.google.com/notebooks/intro.ipynb)*)*



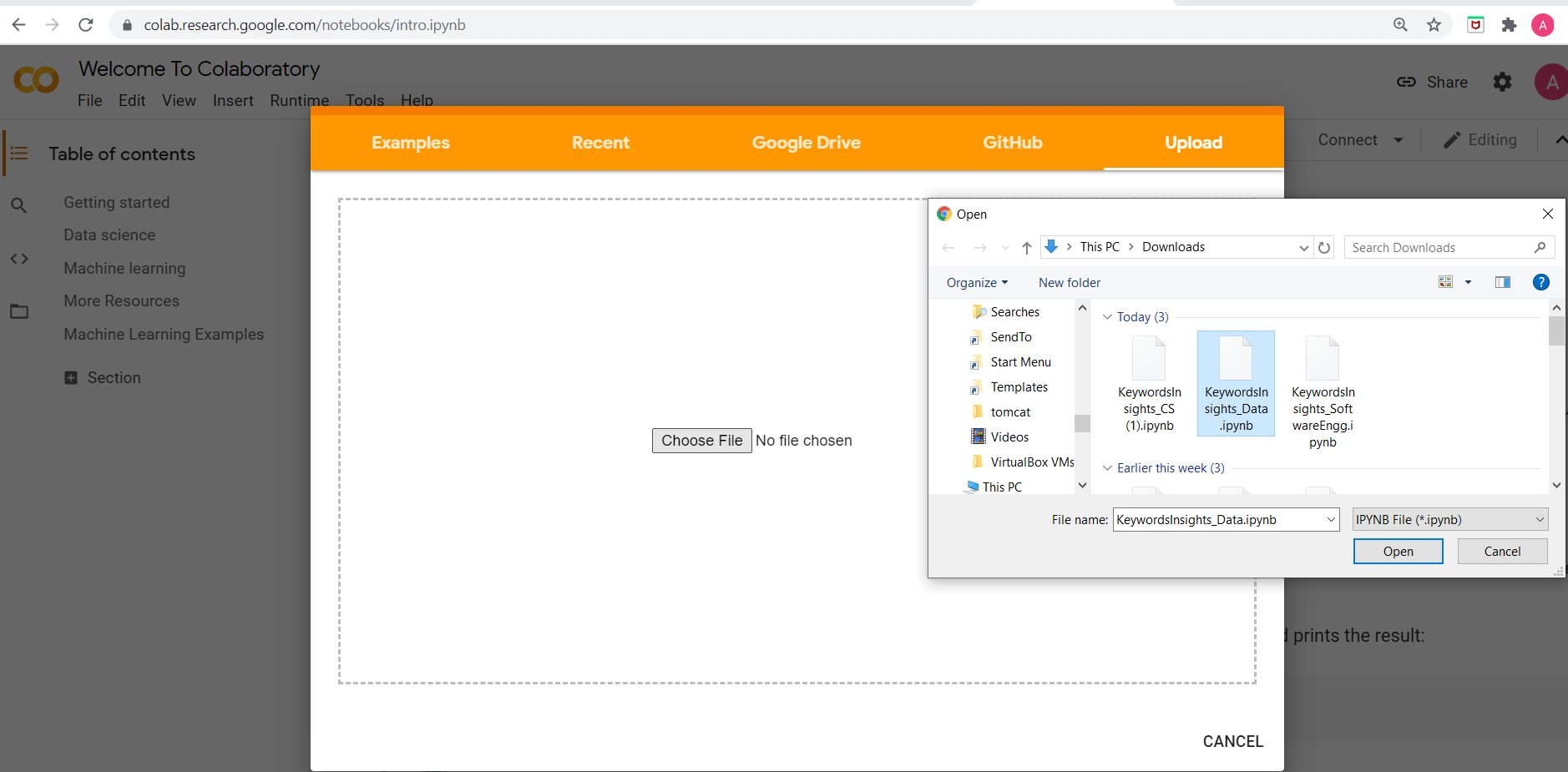
**Figure 1**

**STEP 2:** Upload the Python Utility Notebook as shown below in Figures 2 and 3.

*Please note- When downloaded, most files are placed in the Local or Google Drive location.*

**

**Figure 2**



**Figure 3**

*If you are going to test for Cybersecurity – Open “KeywordsInsights\_CS.ipynb”. If you are testing for software engineering – Open “KeywordsInsights\_SoftwareEngg.ipynb’. If you are testing for Data Science – Open “KeywordsInsights\_Data.ipynb”.*

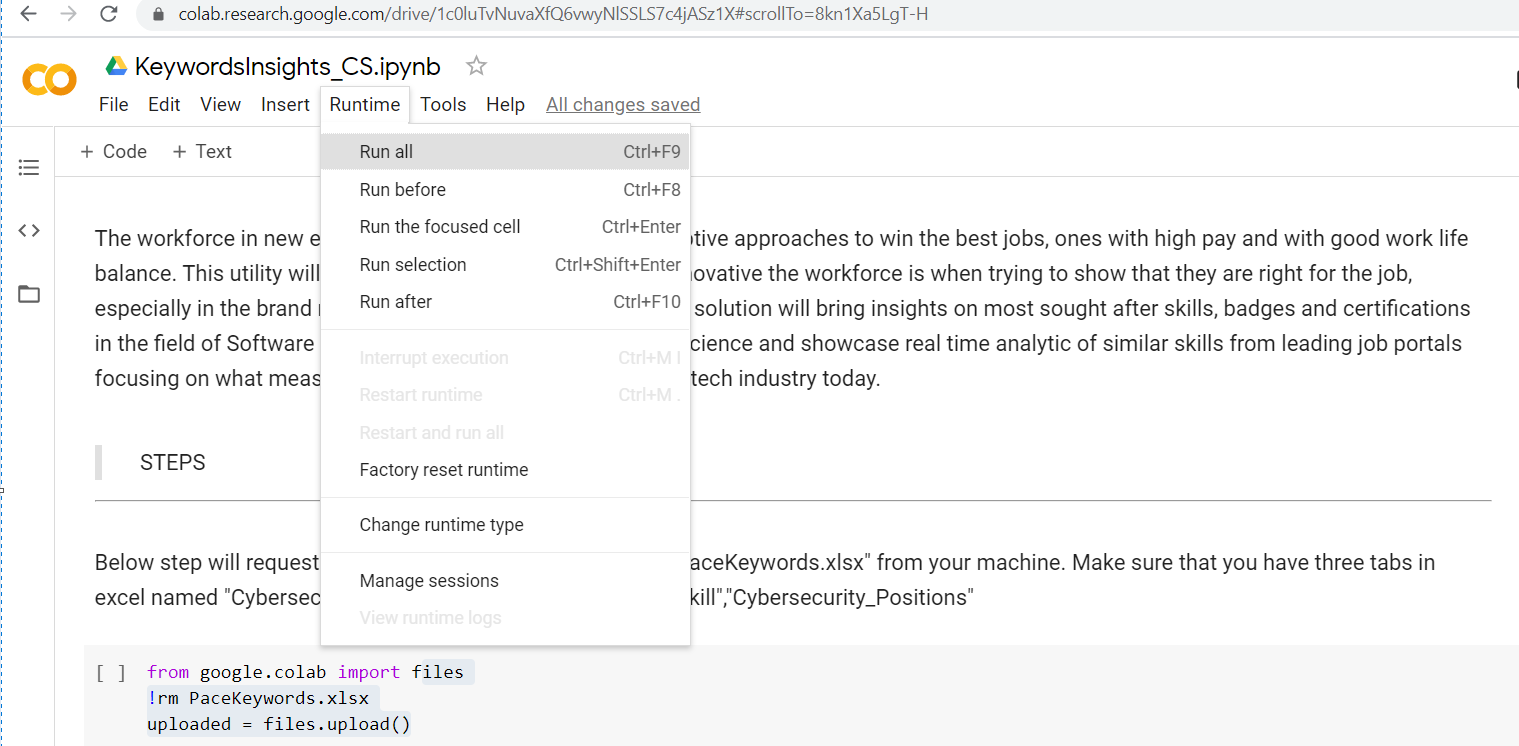
**VALIDATION:** It is important to preform validations prior to running the Utility. Start by validating the keywords in the input file. Then ensure the following:

* Make sure that there are no spelling mistakes.
* Make sure that there are no spaces at the end of the keywords.
* Make sure that the selected file does not have duplicate keywords.
* Make sure file name is “pacekeywords.xlsx”. The extension must be “.xlsx”

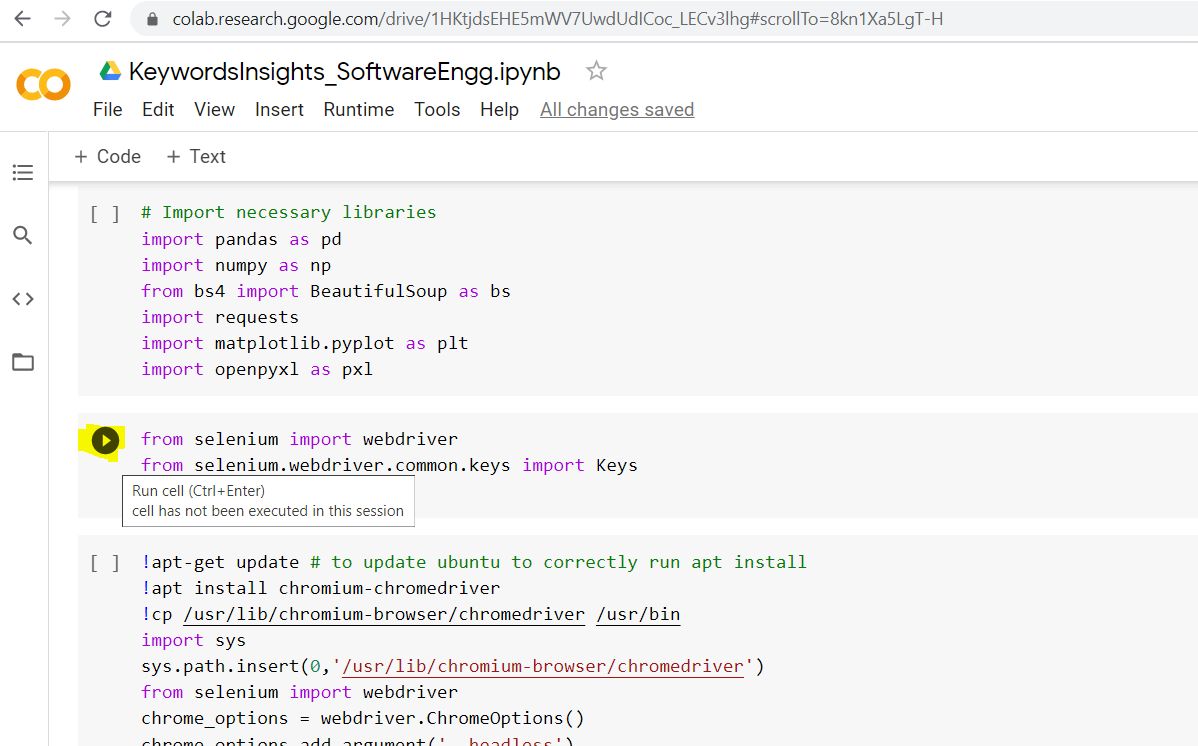
*Please note- The Utility tool is case sensitive. It searches for keywords strictly as if in quotes therefor it is important that there are no mistakes, spaces or other characters when conducting a search to avoid incorrect counts.*

**STEP 3:** Click on the *Runtime* drop-down list and select *Run all* (Figure 4) **-OR-** Click on the *Play Button* highlighted in Figure 5 - you must click on the *Play Button* for every step, one by one.

*Please note- The steps below are demonstrated for the Cybersecurity file, but they must be followed the exact same way for the other streams (Software Engineering and Data Science).*

**

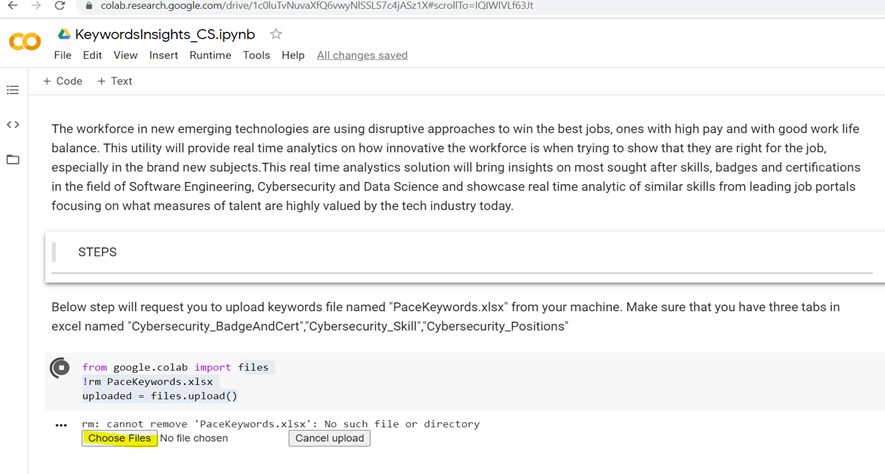
**Figure 4**

**

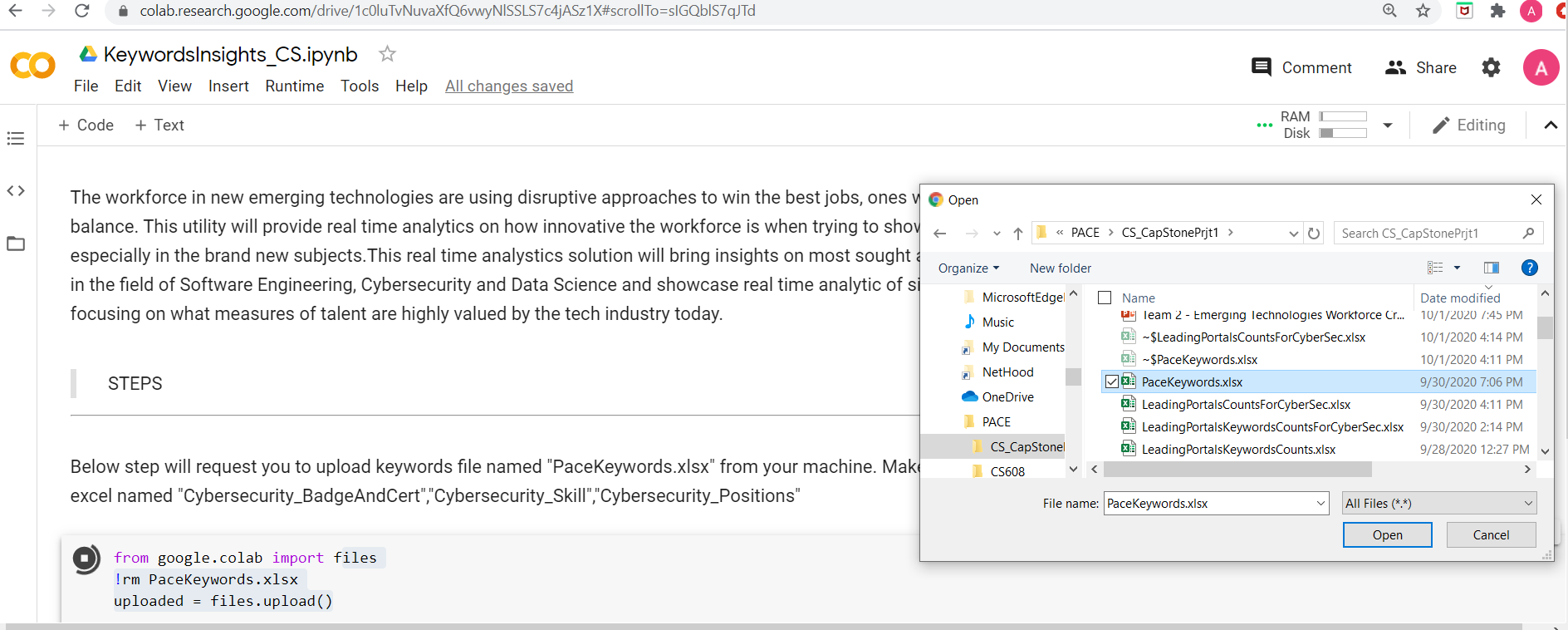
**Figure 5**

**STEP 4:** Click *Choose Files* as shown in Figure 6 and upload the Keywords file from your Local or Google Drive Location named *PaceKeywords.xlsx* as shown in Figure 7.

*Please note- The steps below are demonstrated for the Cybersecurity file, but they must be followed the exact same way for the other streams (Software Engineering and Data Science).*



**Figure 6**

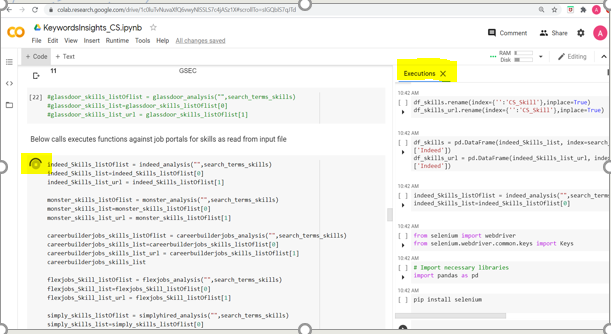


**Figure 7**

*When uploading the PaceKeywords.xlsx file, please make sure that the extension is xlsx.*

**STEP 5:** The Utility will finish executing the rest of the steps on its own, this will take some time depending on the file size and your connectivity. While the execution steps are in progress you may check on that progress in *Executed Code* under *View Menu* **-OR-** simply scroll-down to see in progress run for some steps as highlighted (Figure 8).

*Please note- The steps below are demonstrated for the Cybersecurity file, but they must be followed the exact same way for the other streams (Software Engineering and Data Science).*

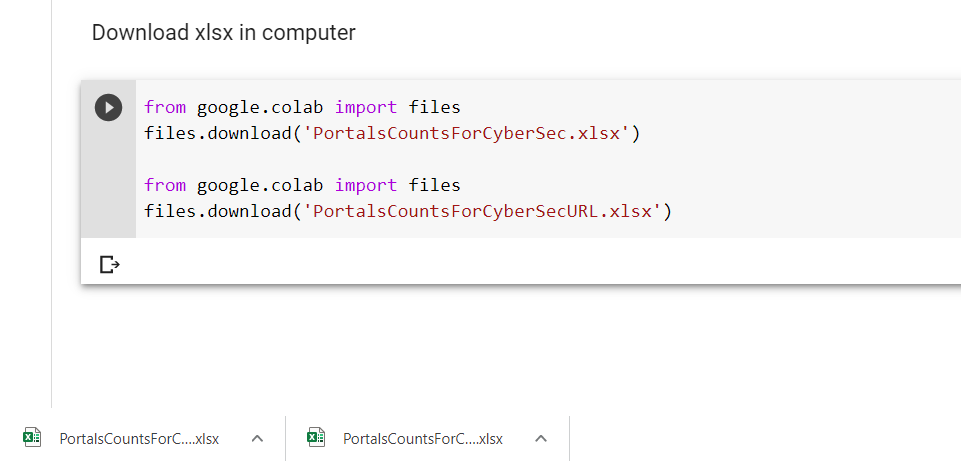


**Figure 8**

**STEP 6:** Once the executions are completed, download the Output Dataset file (Figure 9). If the Output Dataset file is not already in a common location, please move it to your Local or Google Drive.

*Please note- When downloaded, most files are placed in the Local or Google Drive location.*

The Output Dataset file will contain the dataset with their corresponding counts for visuals and Adhoc analysis and it will also contain the dataset with respective URL’s for unit testing.



**Figure 9**

**TROUBLESHOOTING TIPS:**

* Your Internet must be stable and consistent to avoid any interruptions to the Utility which may skew errors.
* Be aware of the steps that take time to execute. If the step stops or fails for any reason, you can restart that individual step without having to restart the entire Utility from the beginning.
* The Secondary file that contains the URL’s can be used to perform unit testing if the counts are doubtful or if the respective site changes its underlying structure.
* Please note that the Utility will need tech maintenance from the code modification standpoint.
* Currently, the Utility takes about an hour to execute per stream (Cybersecurity, Data Software Engineering and Data Science) but may take additional depending on the length of the Keywords list and may result in the connection timing-out at some steps. If the execution times-out, you can restart that individual step only.
* If any step happens to be executed twice, then it is possible to see two tabs with same name in the excel file. Please cross check dataset before refreshing the visuals and remove any additional tabs as needed or re-run the Utility from the beginning.
* The counts could be incorrect if the keywords in the input file have any leading or trailing spaces, spelling mistakes or other characters.

***PLEASE ENHANCE THIS USER MANUAL AS NEEDED***