READ\_ME PPAR Date Lessons (Databricks)

**Overview:**  
  
This code is designed to extract lessons from Project Performance Assessment Report (PPAR) documents using natural language processing techniques. PPAR documents contain valuable insights and lessons learned from various projects. By analyzing the document content and applying predefined patterns, the code identifies and extracts specific sections related to lessons. The extracted lessons are then organized and stored in an Excel file for further analysis and reference. This code aims to automate the process of extracting lessons from PPAR documents, saving time and effort for project evaluation and knowledge sharing purposes.

**Running the Code:**

First of all please make sure that lending.xlsx, bulkDownload.py and APIDocInfo.py is present in the directory of the code.

To execute the provided code, follow these steps:

1. Set Up Databricks Environment:

Log in to your Databricks workspace.

Create a new notebook or open an existing one where you want to run the script.

1. Install Required Libraries:

Ensure libraries like pandas and any custom libraries (APIDocInfo, bulkDownload) are installed and accessible in your environment.

1. Upload the Python Script:

Copy the Python script you have.

Paste the script into a cell (or multiple cells) in the Databricks notebook.

1. Modify the variables as per your requirements:

* Set the data\_folder variable to the path of the data folder.
* Set the file\_type variable to specify the file type(s) to download (e.g., ['ppar']).
* Set the download\_version variable to specify the download version ('txt' or 'pdf').
* Set the start\_date and end\_date variables to specify the date range for downloading files.
* Set the file\_number variable to specify the number of files to download.

1. Run the Script:

Execute the notebook cells sequentially.

Monitor the execution for any errors and debug as necessary.

1. Access and Review Output:

Once the script completes, the output will be Excel files containing the extracted data and reports. These files will be saved to the specified output paths.