Backend Developer Assignment

Hi! Here below goes the assignment.

To start off, evaluation will be on the following criteria:

- Writing a clean, well-commented, tested, error-handled code
- Storing and querying data from a SQL DB
- Consistency using Git commits
- Correctness around edge-cases

Step 1: Python Script

Input:
PDF Document

Output:
A Python function that takes the path to the input .pdf and converts it to a .csv file.

Task Description:
Your first task is to convert the .pdf to a .csv. We have also attached a sample BalSheet.csv to show you what the output should be from BalSheet.pdf.

Cheats:
1. Try using Tabula to convert the .pdf file to .csv and then working to clean the .csv.
2. You can also try using pdftotext (command line tool) to convert the .pdf to a text file and then use regex
3. Try using the Pandas library in Python to clean up the data or add data to a dataframe and then write to .csv.

Supporting Files:
- BalSheet.csv
- BalSheet.pdf

Step 2: Server (Python - Django/Flask)

Input:

- Facility to upload a balance sheet, which will be of the same format.
- Facility to enter a year as well as a variable column name to query from balance sheet.

Output:

- Value of that variable for that year
- Facility to download entire CSV, similar in format generated in Task 1

Task:

- Create a server, which allows a user to upload a balance sheet as well as the variable which he/she is interested to query and the year.
- Once the user hits 'Submit' button, a post request is generated which calls the parsing module.
- The parsing module parses the balance sheet, store the variable in SQL database, and then queries the required variable for the required year as per the text fields
- Also, save the csv generated from parsing module and allow the user to download it

Cheats:

- In Django, file Downloading facility can be done through Response object modification.

Supporting Documents:

- Screen_1.png
- Screen_2.png
- BalSheet.csv
- BalSheet.pdf

Step 3: Final Deliverables

- Start a new repository on your GitHub account and use that for this assignment.
- We will clone that repository and will execute both task on our systems.