

PapyrusNebulae 2023 Document Cloud Hackathon: Round 1

1. The problem is **Extracting information from PDF invoices using Adobe PDF Services Extract API** <https://developer.adobe.com/document-services/apis/pdf-extract/>
2. [InvoicesData](#) provided here has two folders
 - a. SampleInvoicesAndData : This folder has two invoices, and the Extracted Data folder contains the ExtractedData.csv file which contains the data extracted from these invoices. This CSV represents the format in which data needs to be extracted from the invoices.
 - b. TestDataSet folder contains 100 sample invoices and participants are expected to extract data from these 100 invoices in the same form as ExtractedData.csv file
3. For this exercise, you'll need to sign up for an Adobe PDF Services account if you haven't already. <https://developer.adobe.com/document-services/apis/pdf-extract/> has instructions on how to get started with the free trial. Please note that you will get 1000 free transactions with free credentials, which shall be enough for this exercise. But in case the credentials expire, please create one more credential using another personal email address.
4. There are multiple SDKs available for Extract API in Java, Python, NodeJS and C#. You are highly recommended to write your program in one of these languages using the SDK. But in case you want to use a different runtime for which SDK does not exist, feel free to do so but in this case, you will need to create the API request on your own following the documentation available here <https://developer.adobe.com/document-services/docs/apis/#tag/Extract-PDF>
5. Please make sure to write good quality modularized code that not just machines but humans can understand as well.
6. Submission link will be shared with you by 14th June, you can submit the same, any day post that before 26th June 2023, 11:59:59 pm
7. Submission would be in the form of a single link (Git Repo or Shared folder) with your submission code.