What changed from Test Phase 1 to Test Phase 2?

1. Included code to process all files in a single directory, previously you could only pass one file at a time. Now we can process any number of files present inside a directory.
2. Invoices were being handled by analyze_expenses.py under the assumption that they would be jpeg or png images. As such analyze_expenses.py was not capable of handling pdfs (as shown in the demo and mentioned in the limitations section of the previous document), instead PDFs were being handled in the extract_text.py which was supposed to be for text-based documents such as contracts and NDAs which are generally in PDF files. As such when an invoice (which was supposed to be in jpeg or png) was present inside a pdf file, the performance wasn't that great. To solve this issue we added pdf file type support in analyze_expense.py.
3. In the first phase we were not splitting larger pdfs (multiple pages) into smaller pdfs (single page). Now we are doing this as that is leading to better results. The output structure of the output JSON has also been changed to incorporate this change in the code. Now for a pdf with multiple pages the JSON format will be as follows:

```
[
        {
                Page 1 details
        },
        {
                Page 2 details
        },
        …
        {
                Page n details
        }
]
```