README:

I have submitted 4 codes:

1. **image_to_text_extractor.py**- this code is for extracting a text from the pdf given. Iin this case only use the pdf with one page for testing (test PDFann.pdf).
2. **multiple files.py**- this code is used to extract the dwg no in several files and save the data in a .txt file. If the dwg no is not found the text file becomes None.txt.
3. **Data To Excel.py**- this code prepared extracted data for processing. It saves all the .txt files data into a simple readable excel file
4. **Final code.py**- this is the final code that combines all the aspects required for the project as stated by the project objectives. This is the *CODE TO SUBMIT.*

Note:

- Do not forget to change the file name in the main function depending on the location of your file in the computer.

```python
def main():
    # Replace 'pdf_directory' with the directory containing the PDF files
    pdf_directory = 'C:/Users/User/Documents/PROGRAMMING/Python projects'
```

- These are the libraries you will need if you want to test your code:

```python
import os
import fitz
import pandas as pd
import numpy as np
from PIL import Image
from skimage.measure import find_contours
import re
import string
```

So you can do a pip install in your terminal/cmd to install those libraries if you have not already installed them: