**Extracting information from invoice from Amazon**

**Code:**

import pdfplumber

import re

from word2number import w2n

import csv

pdf\_files = ['invoice.pdf', 'invoice1.pdf', 'invoice2.pdf', 'invoice3.pdf']

csv\_file = "invoice\_data.csv"

fieldnames = [

    "Invoice Number", "Invoice Details", "Invoice Date",

    "Order Number", "Order Date", "PAN No",

    "GST Registration Number", "Amount in Words", "Total Amount"

]

with open(csv\_file, mode='w', newline='') as file:

    writer = csv.DictWriter(file, fieldnames=fieldnames)

    writer.writeheader()

    for loc in pdf\_files:

        with pdfplumber.open(loc) as pdf:

            page = pdf.pages[0]

            text = page.extract\_text()

        invoice\_number = re.search(r'Invoice Number :\s\*([A-Z0-9-]+)', text)

        invoice\_details = re.search(r'Invoice Details :\s\*(.\*?)\s+Invoice Date', text)

        invoice\_date = re.search(r'Invoice Date :\s\*(\b\d{2}\.\d{2}\.\d{4}\b)', text)

        order\_number = re.search(r'Order Number:\s\*([A-Z0-9-]+)', text)

        order\_date = re.search(r'Order Date:\s\*(\b\d{2}\.\d{2}\.\d{4}\b)', text)

        PAN\_No = re.search(r'PAN No:\s\*([A-Z0-9-]+)', text)

        gst\_registration\_number = re.search(r'GST Registration No:\s\*([A-Z0-9-]+)', text)

        amount\_in\_words = re.search(r'Amount in Words:\s\*(.+?)\s+only', text, re.IGNORECASE)

        total\_amount = re.search(r'Total Amount:\s\*([A-Z0-9\s,.-]+)', text)

        if total\_amount:

            try:

                amount\_number = float(total\_amount.group(1).replace(',', '').strip())

            except:

                amount\_number = "Could not convert"

        elif amount\_in\_words:

            try:

                amount\_number = w2n.word\_to\_num(amount\_in\_words.group(1))

            except:

                amount\_number = "Conversion failed"

        else:

            amount\_number = "Not found"

        data = {

            "Invoice Number": invoice\_number.group(1) if invoice\_number else "Not found",

            "Invoice Details": invoice\_details.group(1) if invoice\_details else "Not found",

            "Invoice Date": invoice\_date.group(1) if invoice\_date else "Not found",

            "Order Number": order\_number.group(1) if order\_number else "Not found",

            "Order Date": order\_date.group(1) if order\_date else "Not found",

            "PAN No": PAN\_No.group(1) if PAN\_No else "Not found",

            "GST Registration Number": gst\_registration\_number.group(1) if gst\_registration\_number else "Not found",

            "Amount in Words": amount\_in\_words.group(1) if amount\_in\_words else "Not found",

            "Total Amount": amount\_number

        }

        print(f"Processed {loc}: {data}")

        writer.writerow(data)

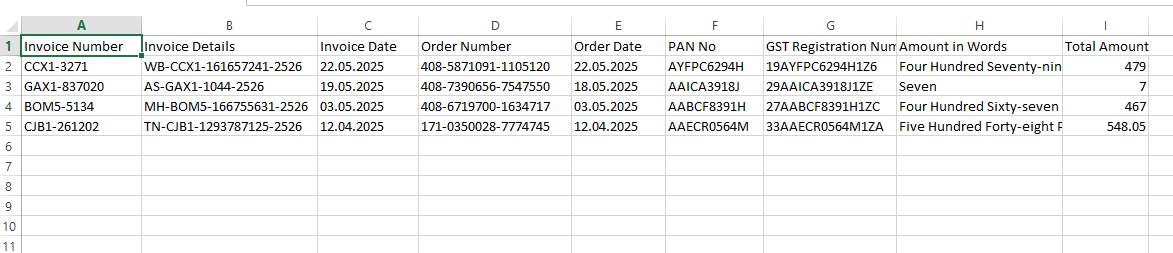
print(f"Data from all invoices saved to {csv\_file}")

**Output:**

One instance of all the output for clear understanding:



**The saved CSV file:**

****