了解しました。以下に **Windows環境でPython・C#のどちらでも動作させるための具体手順** を示します。環境構築から実行確認まで一通り解説します。

**✅ 1. 共通事項：PDF → OCR → 伝票番号抽出 → PDF分割 → CSV → PostgreSQL**

**🐍 Python版：Windowsでの実行手順**

**🔧 ① 必須ソフトのインストール**

| **ソフト名** | **用途** | **備考** |
| --- | --- | --- |
| Python 3.x | 実行環境 | [公式サイト](https://www.python.org/downloads/windows/) |
| Tesseract OCR | 画像から文字を読み取る | [公式配布 (tesseract-ocr)](https://github.com/tesseract-ocr/tesseract/wiki) の Windows版 |
| Poppler for Windows | PDFページ画像化用 pdf2image用 | [Popplerバイナリ](https://github.com/oschwartz10612/poppler-windows/releases) |

**✅ 環境変数設定**

* Tesseract をインストールしたら、例：
* C:\Program Files\Tesseract-OCR\

を **環境変数 PATH に追加**。

* Poppler は展開後の bin フォルダを PATH に追加（例：C:\tools\poppler\bin）

**🔧 ② Pythonライブラリのインストール**

pip install pytesseract pdf2image PyMuPDF psycopg2

**🔧 ③ サンプルPDFとスクリプトの配置**

プロジェクト/

├─ input/orders.pdf

├─ output/（自動作成）

├─ pdf\_process.py（前述コード）

**▶ 実行コマンド（PowerShell）**

cd プロジェクト

python pdf\_process.py

**→ PDF分割 + CSV出力 + DB登録が output/ に保存されます。**

**💠 C#版：Windowsでの実行手順（Visual Studio使用）**

**🔧 ① 必須ソフト**

| **ソフト** | **用途** | **備考** |
| --- | --- | --- |
| Visual Studio 2022以降 | C# IDE | [.NET Desktop開発ワークロード] が必要 |
| Tesseract OCR | OCR本体 | Python版と同じインストーラ可 |
| Ghostscript（必要に応じて） | PDF→画像変換に使う場合あり | 無料、インストーラ形式あり |

**🔧 ② NuGetパッケージを追加**

Visual Studio → 「プロジェクト」→「NuGetパッケージ管理」

* Tesseract
* PdfPig（PDF読み取り＆分割用）
* Npgsql（PostgreSQL用）

**📁 プロジェクト構成例**

Solution/

├─ Program.cs（前述コード）

├─ input/orders.pdf

├─ output/

**📌 Tesseractの言語データ指定**

var ocrEngine = new TesseractEngine(@"./tessdata", "jpn", EngineMode.Default);

* tessdata フォルダを実行ファイルの隣に置き、そこに jpn.traineddata を配置する必要があります。

📥 言語ファイルDL:  
<https://github.com/tesseract-ocr/tessdata>

**▶ 実行方法**

Visual Studio → デバッグ → スタート または Ctrl+F5

**🗃 PostgreSQL 側の準備（共通）**

**SQL例（テーブル作成）**

CREATE TABLE orders (

id SERIAL PRIMARY KEY,

denno TEXT NOT NULL,

content TEXT NOT NULL

);

* DB：your\_db
* ユーザー：your\_user
* パスワード：your\_password
* 接続確認には pgAdmin, psql, DBeaver などを使用

**✅ トラブル時のチェックポイント**

| **問題** | **原因候補** |
| --- | --- |
| Tesseractが動かない | PATHが通っていない or tessdata/langが無い |
| pdf2imageが画像化できない | Popplerが未インストール or PATHが未設定 |
| PostgreSQL接続エラー | psycopg2またはNpgsqlの接続文字列の間違い |
| PDFの日本語が文字化けする | PDFが画像形式 → OCR前提 / テキスト抽出は困難 |

**✅ 推奨構成（自動実行にも対応）**

* バッチやスケジューラーから python / exe を実行可能
* 出力先を日付フォルダで分けるなどの運用も簡単に実装可能
* エラーログを log.txt に残す処理を追加すると安心

ご希望があれば、以下も提供可能です：

* Pythonスクリプトを .exe に変換（pyinstaller）
* C#アプリをGUI（WinForms/WPF）化
* VBAから呼び出すテンプレート（Shell or WScript.Shell）

必要があればお気軽にお知らせください。