

RPA PROJECT

PDF EXTRACTION

Name: Y.LIKITHA REDDY

ID no : 2300030763

Branch: CSE-HTE , 2nd year

INTRODUCTION:

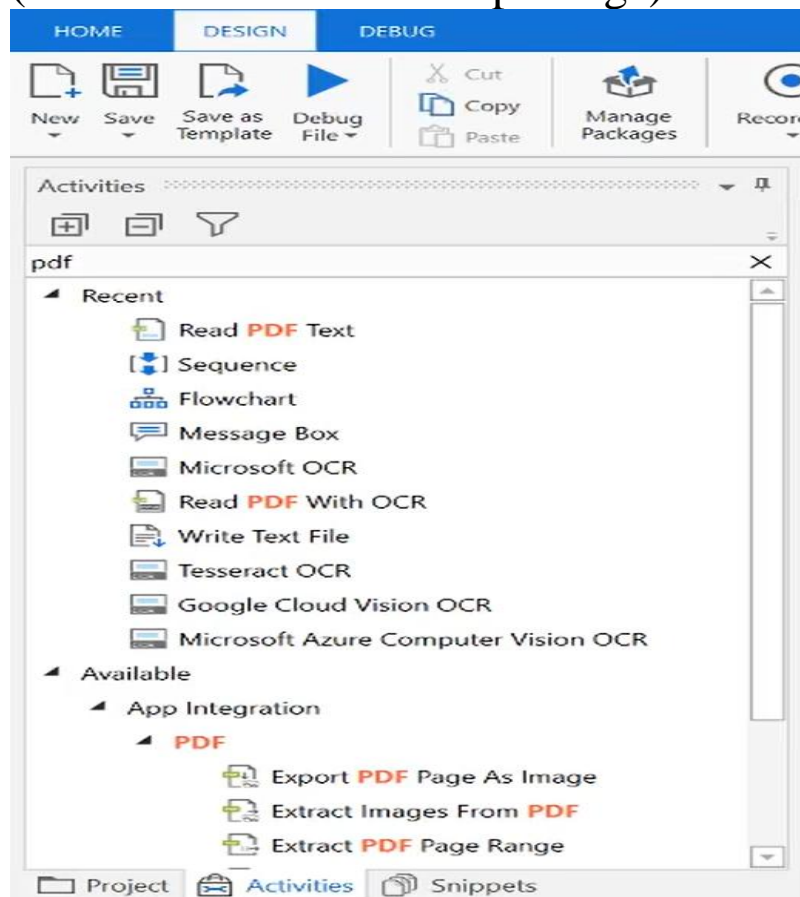
PDF extraction in **UiPath** is done using **Read PDF Text** for searchable PDFs and **Read PDF with OCR** for scanned PDFs.

PROJECT DESCRIPTION:

Project Steps: PDF Data Extraction Using UiPath:

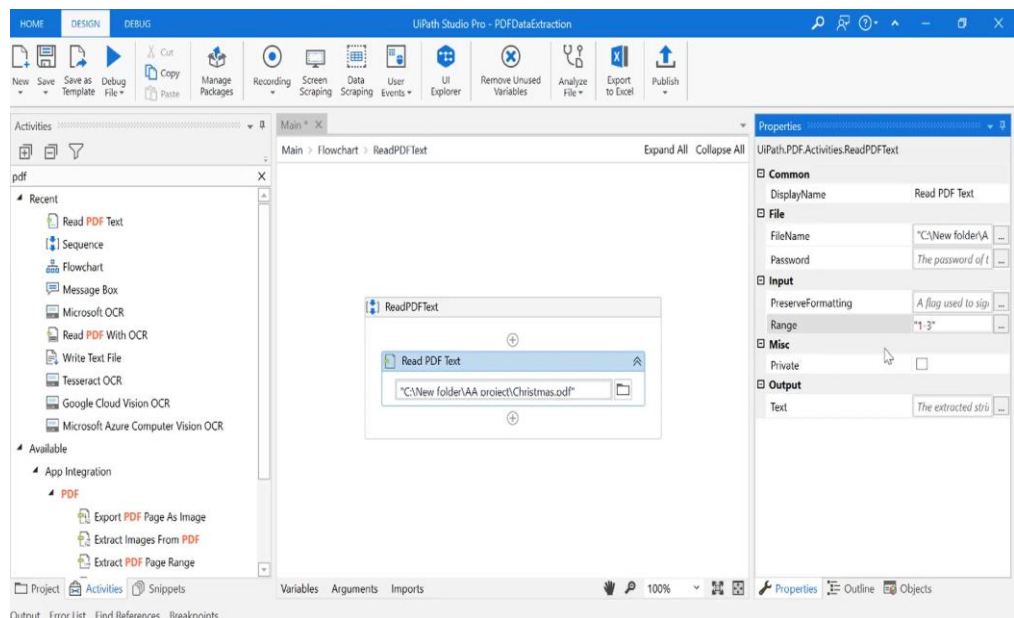
1. Project Setup:

- Open **UiPath Studio** and create a new project
- Install the necessary dependencies (**UiPath.PDF.Activities** package).



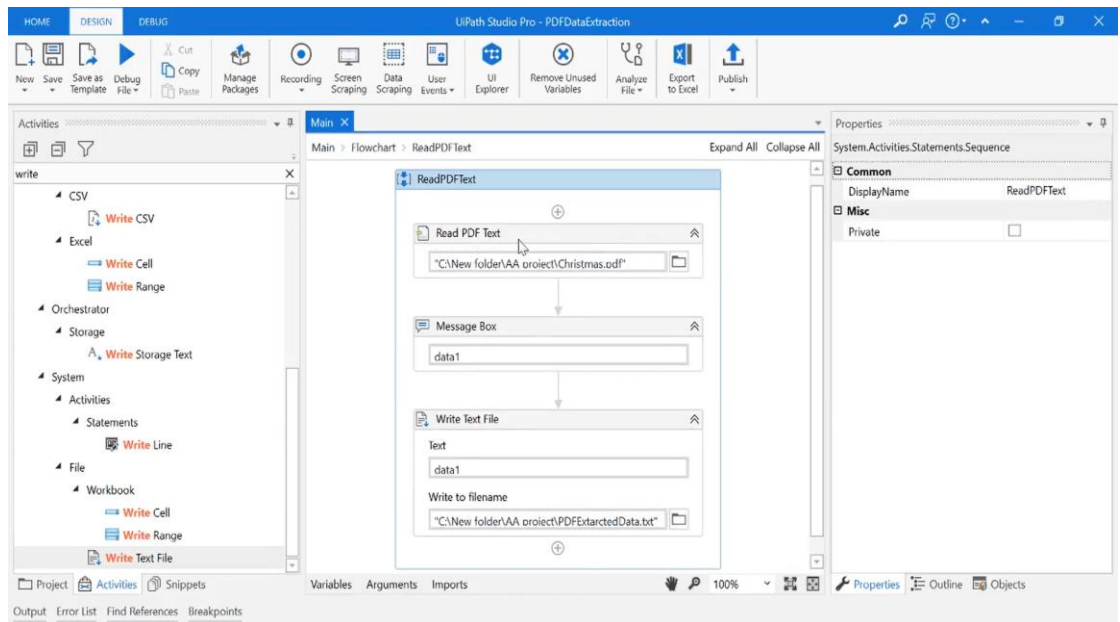
2. Input PDF Selection

- Define the **PDF file path** from which data needs to be extracted.
- Use the "**Read PDF Text**" activity for text-based PDFs.
- Use the "**Read PDF with OCR**" activity for scanned PDFs.



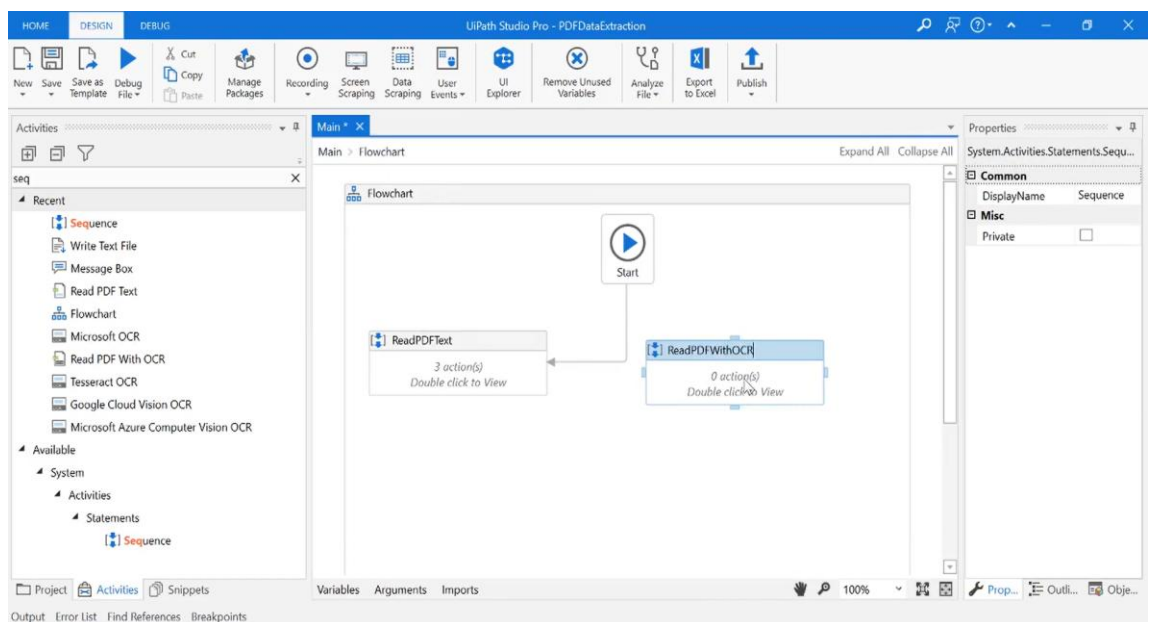
3. Extract Required Data

- For plain text, use **string manipulations** (e.g., Regex, Split).
- For specific elements, use the "**Get Text**" activity.
- For tables or structured data, use **Data Scraping**.



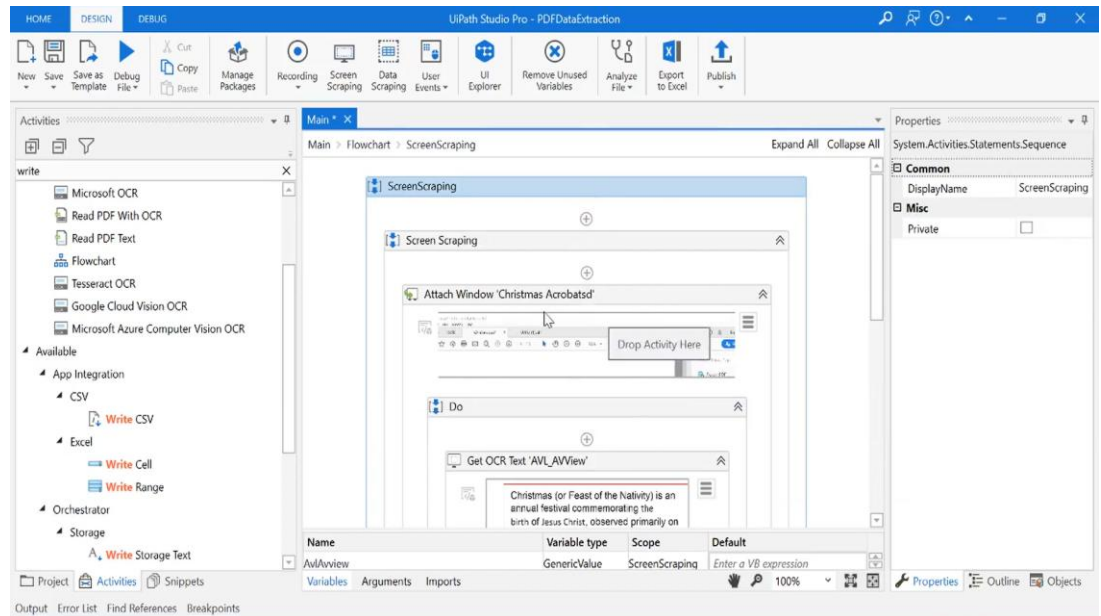
4. Process and Store Data

- Store extracted data in **Excel, CSV, or a database.**
- Apply **data validation** or transformations if required



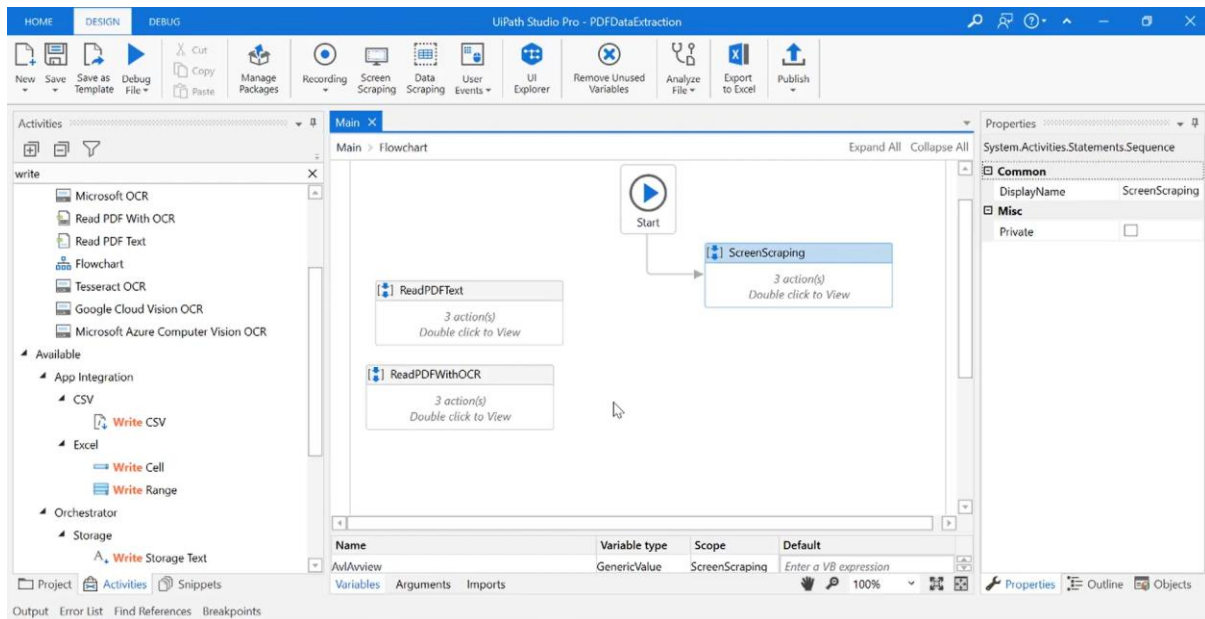
5. Automation Execution and Testing

- Run the **workflow** to verify extraction accuracy.
- Handle errors using **Try-Catch** for exceptions like missing text.



6. Deployment and Integration

- Deploy the bot for scheduled or triggered execution.
- Integrate with other applications if necessary (e.g., ERP, CRM).



CONCLUSION:

UiPath provides an efficient way to automate **PDF data extraction**, reducing manual effort and improving accuracy. By leveraging activities like **Read PDF Text**, **OCR**, and **Data Scraping**, extracted data can be processed and integrated seamlessly into various applications. This automation enhances productivity, ensuring fast and reliable data handling.