DATA VALIDATION IN EXCEL FILES

A PROJECT REPORT

Submitted by

DEVDHARSHAN S R(220701060)

in partial fulfillment for the course

OAI1903 - INTRODUCTION TO ROBOTIC PROCESS AUTOMATION

for the degree of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

NOVEMBER 2024

RAJALAKSHMI ENGINEERING COLLEGE CHENNAI – 602 105

BONAFIDE CERTIFICATE

Certified that this project report **"DATA VALIDATION IN EXCEL FILES"** is the Bonafide work of **"DEVDHARSHAN S R (220701060)"** who carried out the project work for the subject OAI1903-Introduction to Robotic Process Automation under my supervision.

SIGNATURE

Mrs. J. Jinu Sophia SUPERVISOR

Assistant Professor (SG)

Department of

Computer Science and Engineering

Rajalakshmi Engineering College

Rajalakshmi Nagar

Thandalam, Chennai - 602105

Submitted to Project and Viva Voce Examination for the subject OAI1903-Introduction to Robotic Process Automation held on _____

INTERNAL EXAMINER   **EXTERNAL EXAMINER**

ACKNOWLEDGEMENT

ABSTRACT

Data validation in Excel files is an essential process to ensure the accuracy, consistency, and reliability of data. Manual validation can be time-consuming and prone to errors, making automation a preferred solution in today's fast-paced business environment. This project leverages UiPath Studio to automate data validation in Excel files, offering a robust approach to streamline the process. The workflow validates specific fields such as name, age, email, phone number, and gender against predefined conditions to ensure compliance with data standards. When errors are identified, the corresponding cells are dynamically highlighted in red, allowing users to quickly pinpoint and rectify issues. For example, names must follow a proper format, ages should fall within an acceptable range, emails must adhere to a valid structure, phone numbers need to meet specific formats, and gender entries should match predefined options. This rule-based approach ensures a thorough validation of all data entries. The automation process not only improves accuracy but also significantly reduces the time required for data validation. By eliminating the need for manual inspection, this project minimizes human errors and enhances operational efficiency. Additionally, the solution is scalable and can be adapted to handle large datasets across various industries, including finance, healthcare, and retail. In conclusion, this UiPath-based data validation project offers a cost-effective and user-friendly method to maintain high-quality data in Excel files. The ability to dynamically highlight errors in red simplifies the review process and ensures reliable outputs for decision-making. This implementation underscores the power of automation in addressing real-world data management challenges efficiently and effectively.

TABLE OF CONTENTS

LIST OF FIGURES :

LIST OF ABBREVIATIONS:

Abbreviation	Full Form
SMTP	Simple Mail Transfer Protocol
ERD	Entity Relationship Diagram
DFD	Data Flow Diagram
HR	Human Resources
API	Application Programming Interface
RE	Robotic Enterprise
RPA	Robotics Process Automation

CHAPTER-1

INTRODUCTION

This project focuses on automating data validation in Excel files using UiPath Studio. It checks specific fields like name, age, email, phone number, and gender against predefined conditions. Errors are identified and dynamically highlighted in red for easy identification. The automation reduces manual effort, enhances data accuracy, and minimizes human errors. It ensures compliance with data standards and streamlines the validation process. This approach provides a simple, efficient, and scalable solution for maintaining data quality.

1. 1   GENERAL

Data validation ensures that information is accurate, consistent, and reliable. It is a critical process used across industries to maintain data integrity. Manual validation is often tedious and prone to errors, leading to inefficiencies. Automating validation processes saves time, reduces human errors, and improves accuracy. This approach ensures compliance with standards and simplifies error identification. Automation provides a scalable solution for managing large datasets effectively.

1.2   OBJECTIVE

The objective of this project is to automate the data validation process in Excel files. It aims to ensure the accuracy and consistency of data by identifying and highlighting errors.

efficiency. The project focuses on validating key fields such as name, age, email, phone number, and gender. It is designed to handle large datasets and streamline data management. This solution offers a scalable approach for maintaining high-quality data across various industries.

1.3   EXISTINGSYSTEM

The existing system relies on manual data validation, which is time-consuming and prone to errors. Users manually review and correct invalid data in Excel files. This method becomes inefficient with larger datasets. Additionally, the system lacks automation, which can lead to inconsistencies in data quality.

1.4   PROPOSEDSYSTEM

The proposed system automates the data validation process in Excel using UiPath Studio. It ensures accuracy by automatically validating key fields such as name, age, email, phone number, and gender. Errors are dynamically identified and highlighted in red for easy correction. This automation significantly reduces manual effort and minimizes human errors. The system improves the speed and efficiency of data validation, especially with large datasets. It offers a scalable solution that can be applied across various industries to maintain high-quality data.

CHAPTER-2

LITERATURE_REVIEW

Previous studies highlight the challenges of manual data validation, including time consumption and the risk of human errors. Automation tools like UiPath have been explored to streamline this process, improving efficiency and accuracy. Research shows that automated systems enhance data integrity by eliminating inconsistencies. Existing solutions focus on basic validation, but they lack dynamic error detection. The integration of automation in data validation is increasingly seen as a valuable solution in various industries.

2.1   GENERAL

Data validation is crucial for maintaining data integrity and accuracy. Inaccurate data can lead to incorrect conclusions, affecting decision-making processes. Traditional manual validation is time-consuming and prone to errors, especially with large datasets. It can be challenging to ensure data consistency across multiple files. Therefore, automation in data validation is becoming increasingly essential in modern data management.

Manual validation often involves reviewing each data entry individually, which is slow and inefficient. This process also opens the door to human errors, especially when handling large volumes of data. Missing or incorrect entries can

easilybeoverlooked.Additionally, data validation needs to be consistent to ensure data reliability across the organization. As the data grows, so does the complexity of managing and validating it manually.

UiPath Studio offers an automated solution for data validation, providing efficiency and accuracy. By creating workflows, UiPath can check fields like name, age, email, and phone number against predefined rules. Any data that fails validation is highlighted in red, making it easy for users to identify errors. The system reduces the need for manual checking, streamlining the entire validation process. This automation speeds up the process and minimizes human errors, improving data quality.

The automation provided by UiPath enables quicker data validation, saving time and reducing the workload of employees. As the system performs checks automatically, users can focus on more complex tasks that require human intervention. This results in greater efficiency and productivity within organizations. The ability to process large amounts of data in a short time is a key advantage of automation. Ultimately, businesses benefit from more accurate, consistent, and reliable data.

The proposed system is scalable and adaptable for businesses of any size. Whether handling small or large datasets, UiPath ensures that the validation process remains effective. The solution can be customized to meet the specific needs of different industries and departments. Integration with existing systems is seamless, further enhancing its value. By implementing UiPath, organizations can maintain high-quality data, leading to better decision-making and business outcomes.

CHAPTER-3

SYSTEM DESIGN

3.1.1 SYSTEM FLOW DIAGRAM

The **System Flow Diagram** outlines the overall flow of data and processes in the system. It demonstrates how user inputs, system processing, and outputs interact.

Description:

1. **Input**: Your Excel sheet contains data for validation.
2. **Process**:
   - Read data from the Excel sheet.
   - Validate each field (Name, Age, Email, Phone, Gender).
   - Marks invalid fields with red color..
3. **Output**:Updated Excel with highlights and error logs are precisely what your project generates.
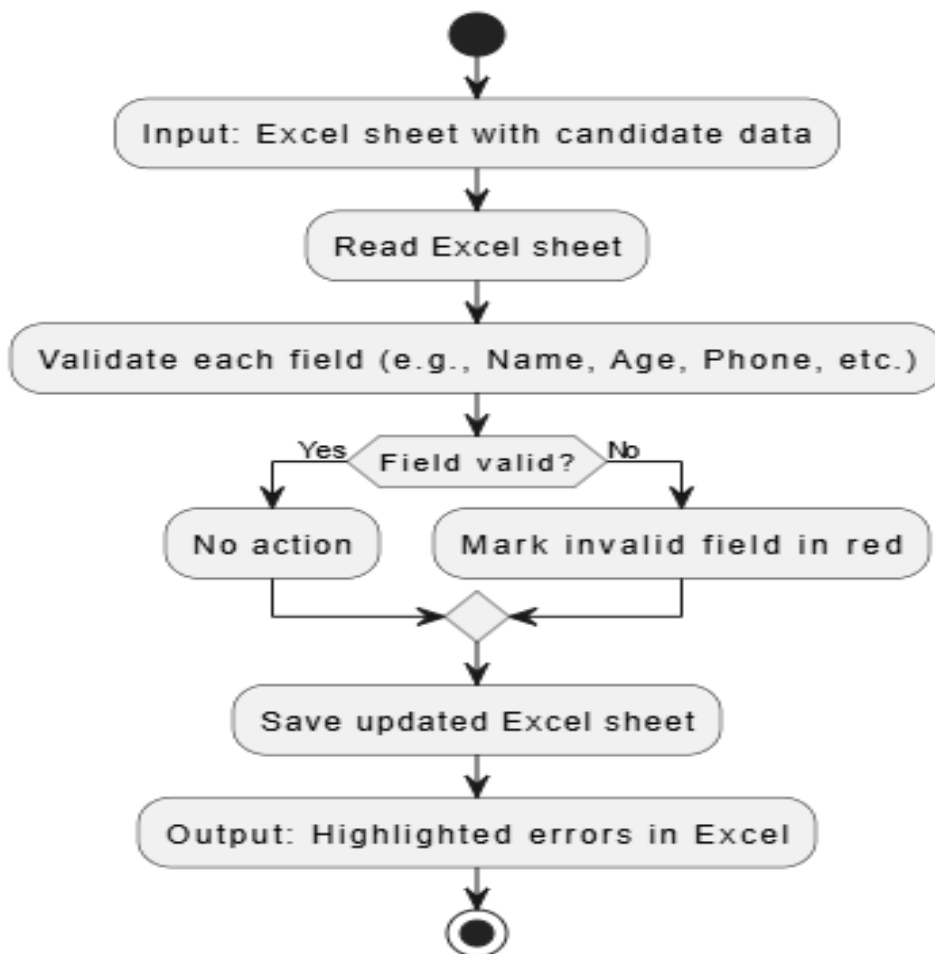
3.1.2 ARCHITECTURE DIAGRAM

The **Architecture Diagram** provides a high-level view of the system's structure and its components.

Components:

1. **Frontend**: User interface.
2. **Backend**: Core logic, including:
    - Read data from Excel
    - Validate fields (Name, Age, Email, Phone, Gender).
    - Mark invalid fields in red
3. **Database/Storage**: No database or logged specified.
4. **External Services**: No external used.
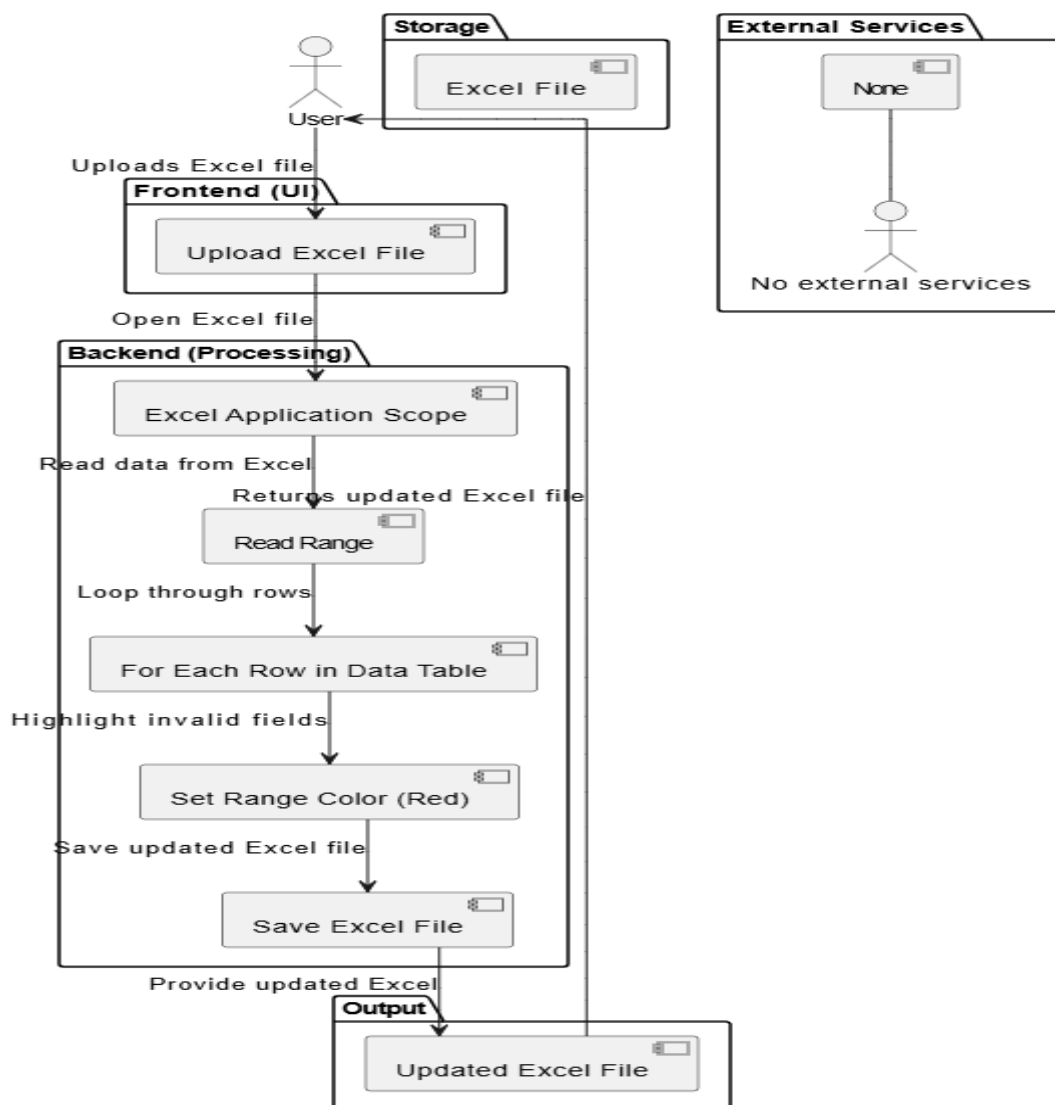
3.1.3 SEQUENCE DIAGRAM

The **Sequence Diagram** shows the interaction between actors (User) and the system components in a sequential manner.

Steps:

1. User uploads the Excel file.
2. System reads data from the Excel file.
3. System loops through the rows and highlights invalid fields in red
4. System saves and provides the updated Excel file

.

CHAPTER-4

PROJECT DESCRIPTION

This project involves validating candidate data in an Excel file. The user uploads the Excel file, which is then processed by the system. The system reads the data from the Excel file and validates the phone numbers of each candidate. If any phone numbers are invalid, the system highlights them in red. After the validation, the system saves the updated Excel file with the highlighted errors. The user receives the updated file with the necessary corrections made.

4.1   METHODOLOGY

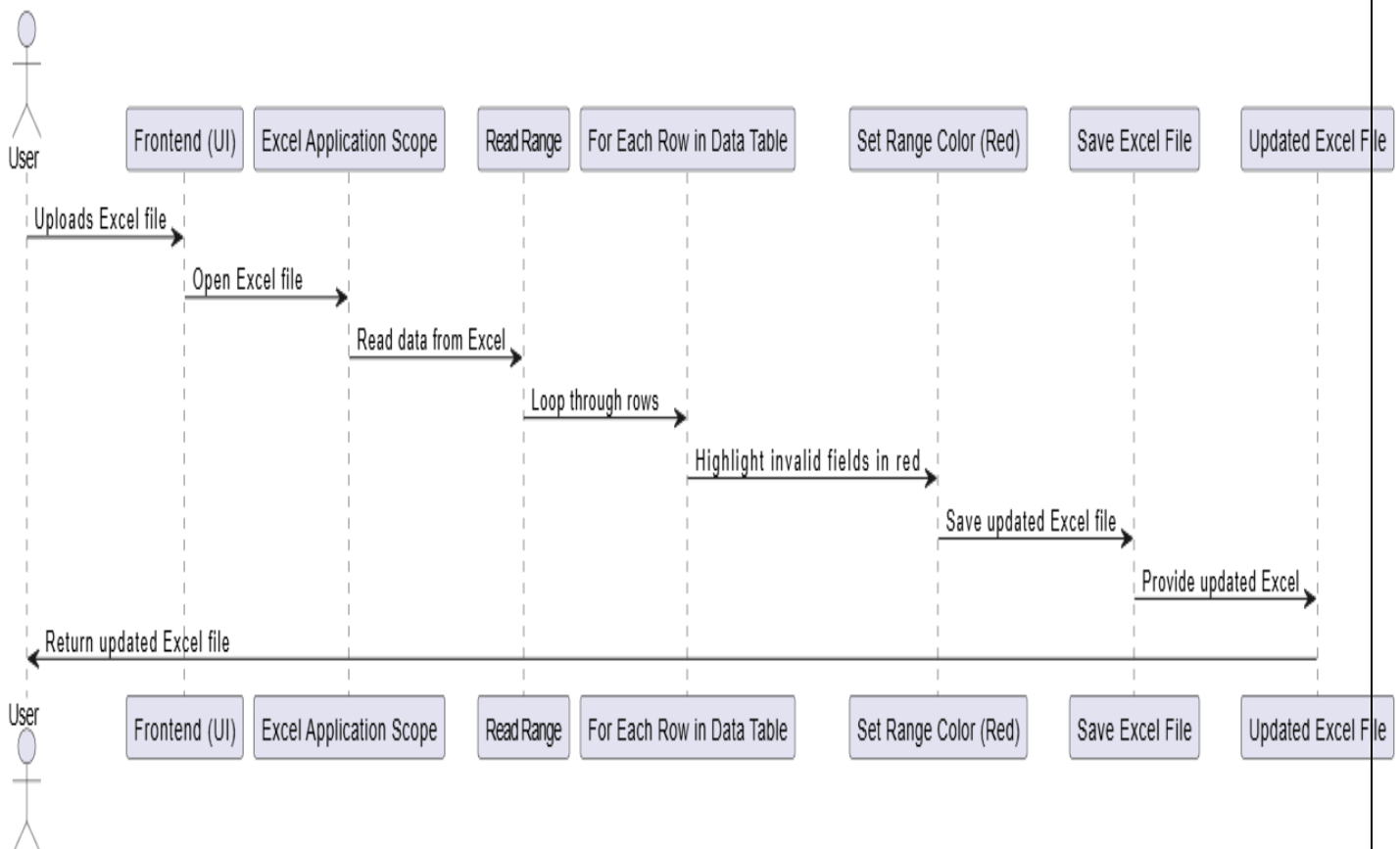The development of the followed an agile methodology, ensuring iterative progress and flexibility in meeting project requirements. The system was built using UiPath's Robotic Process Automation (RPA) platform, utilizing its RE Framework to ensure structured execution, error handling, and scalability. The key steps in the methodology include the following:

1. **Requirements Gathering:** Based on the requirements, system designs, including flow diagrams and sequence diagrams, were created to structure the automation process for Excel data reading, validation, and error marking.

2. **System Design:** Based on the requirements, system designs (including flow diagrams, architecture, and sequence diagrams) were created to ensure that the automation would meet all specifications.

3. **Implementation:** The system was implemented using UiPath, integrating modules to read data from the Excel file, validate the phone numbers, and highlight errors. The RE Framework ensured the workflows were organized, errors handled, and execution seamless.

4. **Testing & Deployment:** The system was thoroughly tested to ensure correct validation of phone numbers, accurate highlighting of errors, and successful file saving. After testing, the system was deployed for user use to process candidate data in real-world scenarios

4.1.1 MODULES:

1. **Excel File Reading Module:** This module reads data from the uploaded Excel file, extracting candidate information row by row for validation purposes.

2. **Validation Module:** This module checks each row in the Excel file for invalid data, such as incorrect phone numbers or empty fields. It ensures that all entries meet the specified criteria**.**

3. **Error Highlighting Module:** If any data is invalid, this module highlights the corresponding cells in red within the same Excel file for the user's review

4. **Data Saving Module:** After completing the validation and highlighting errors, this module saves the updated Excel file and makes it available for the user to download.

5. **Logging Module:** This module logs the actions performed during validation, including errors found and processed rows, to provide transparency.

CHAPTER-5

OUTPUT SCREENSHOT

	Name	Age	Email	Phone Number	Gender
2	John Doe	25	john.doe@example.cor	+911234567890	Male
3	Jane123	30	jane@domain	+91123456789	Female
4	Alice		alice@example.in	1234567890	Other
5		19	alice@	+91abcdef1234	Male
6	Robert	45	robert@	+911234567890	Third Gender
7	Chris Brown	110	chris.brown@gmail.con	+91123456789	Female
8	Eve	20	eve	+911234567890	Male
9	Mark Zuck	34	mark.zuck@example	+91 1234567890	Invalid
10	Anna-Marie	29		+911234567890	Male
11	1Samuel	17	samuel@domain.com	+911234567890	
12	Grace	22	grace@	+911234567890	Female
13	Peter		peter@	+91abcdef5678	Other
14	Matthew	35	matthew@	+911234567890	Male
15	Linda	40	linda@example.com	+911234567890	Female
16	Sarah	26	sarah@domain.in	+911234567890	Male
17	James Bond	27	james@	123456789	Third Gender
18	12345	20	steve.rogers@	+911234567890	Other
19	Steve Rogers	18	natasha@example.com	+911234567890	Male
20	Natasha	18	bruce@domain.in	+911234567890	Female
21	Bruce	55	bruce	+911234567890	Invalid
22					
23					

Fig. 5.1. Validated excel file

From this above figure contains the validated excel file

	A	B	C	D	E
1	**Name**	**Age**	**Email**	**Phone Number**	**Gender**
2	John Doe	25	john.doe@example.co	+911234567890	Male
3	Jane123	30	jane@domain	+91123456789	Female
4	Alice		alice@example.in	1234567890	Other
5		19	alice@	+91abcdef1234	Male
6	Robert	45	robert@	+911234567890	Third Gender
7	Chris Brown	110	chris.brown@gmail.co	+91123456789	Female
8	Eve		eve	+911234567890	Male
9	Mark Zuck	34	mark.zuck@example	+91 1234567890	Invalid
10	Anna-Marie	29		+911234567890	Male
11	1Samuel	17	samuel@domain.com	+911234567890	
12	Grace	22	grace@	+911234567890	Female
13	Peter		peter@	+91abcdef5678	Other
14	Matthew	35	matthew@	+911234567890	Male
15	Linda	40	linda@example.com	+911234567890	Female
16	Sarah	26	sarah@domain.in	+911234567890	Male
17	James Bond	27	james@	123456789	Third Gender
18	12345	20	steve.rogers@	+911234567890	Other
19	Steve Rogers	18	natasha@example.com	+911234567890	Male
20	Natasha		bruce@domain.in	+911234567890	Female
21	Bruce	55	bruce	+911234567890	Invalid
22					
23					
24					

Fig. 5.2. Sample Excel Sheet

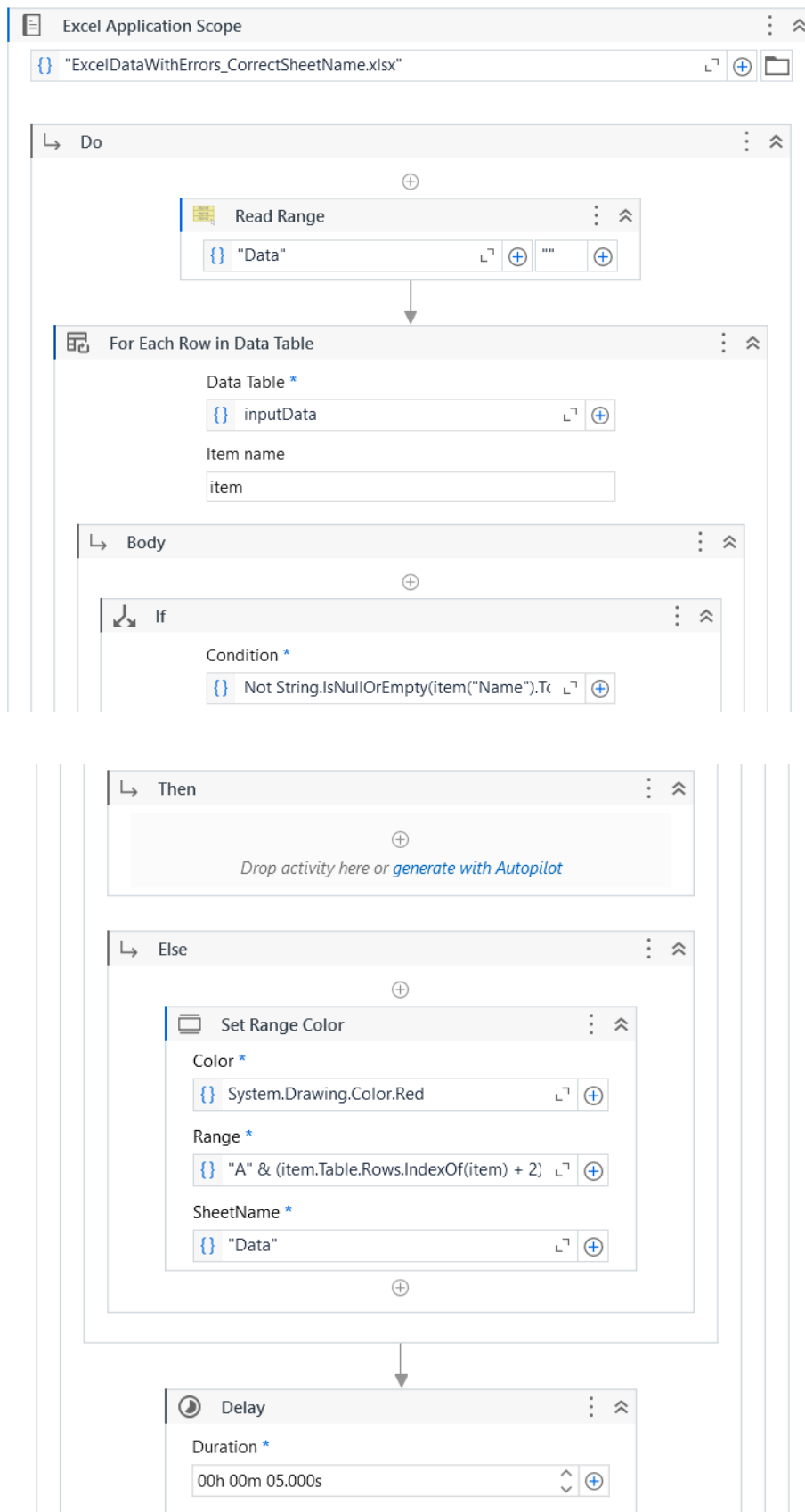From this above figure Contains the sample excel sheet
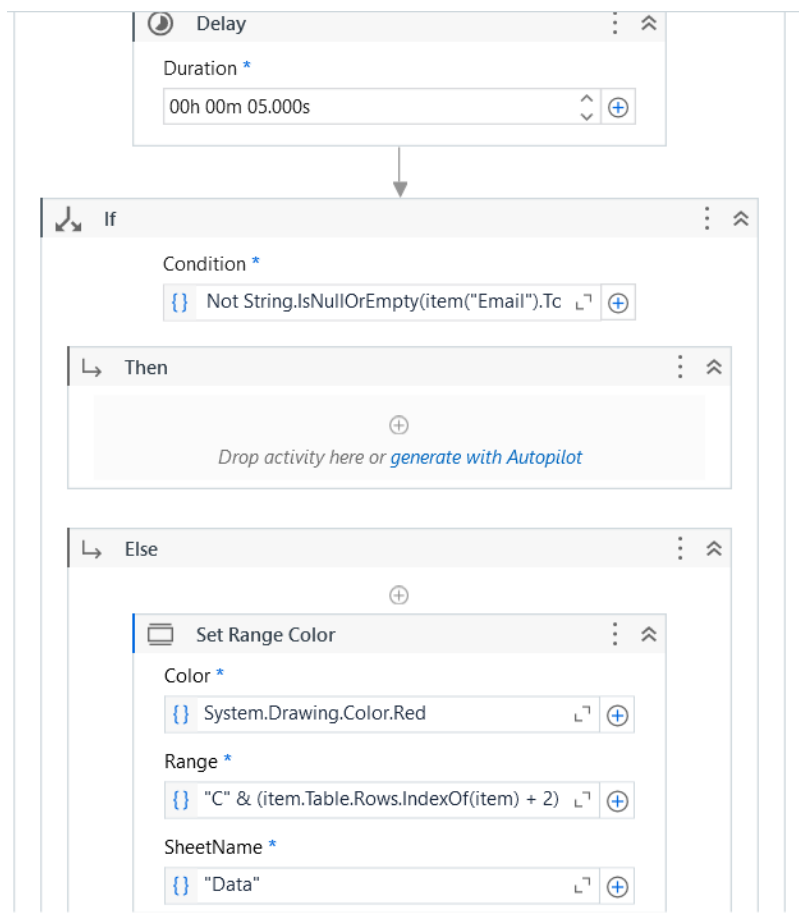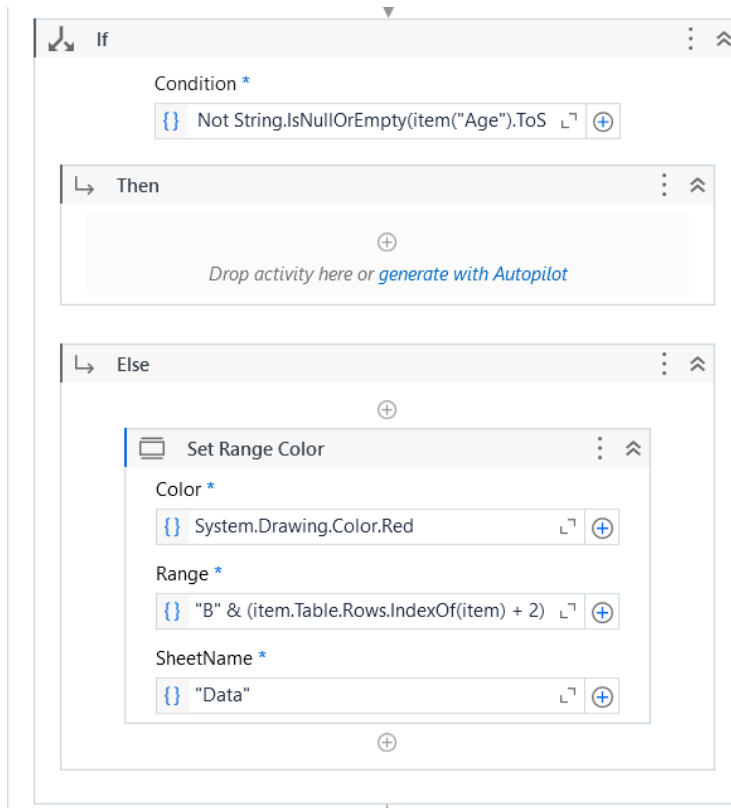
CHAPTER-6

CONCLUSIONS

The "Data validation in excel files" revolutionizes data validation by automating the detection and highlighting of errors in Excel files using UiPath's Robotic Process Automation (RPA) platform. By eliminating manual error-checking, the bot ensures efficiency, accuracy, and user-friendliness in handling candidate data.

Its capability to validate phone numbers, dynamically highlight invalid entries in red, and provide a clean, updated file streamlines workflows, making it a practical solution for data integrity across various use cases. Robust exception handling ensures uninterrupted operation, while clear logs enhance transparency and accountability.

Future developments could incorporate additional validation criteria or integrate the bot into a broader data-processing system to further enhance scalability and versatility. This project demonstrates the power of RPA in improving productivity and precision in professional tasks, establishing a robust foundation for future advancements in data validation automation.

APPENDIX

Excel Application Scope

{} "ExcelDataWithErrors_CorrectSheetName.xlsx"

Do

Read Range

{} "Data" ""

For Each Row in Data Table

Data Table *
{} inputData

Item name
item

Body

If

Condition *
{} Not String.IsNullOrEmpty(item("Name").Tc

Then

Drop activity here or *generate with Autopilot*

Else

Set Range Color

Color *
{} System.Drawing.Color.Red

Range *
{} "A" & (item.Table.Rows.IndexOf(item) + 2)

SheetName *
{} "Data"

Delay

Duration *
00h 00m 05.000s

If

Condition *

`{}` Not String.IsNullOrEmpty(item("Age").ToS ⌐ ⊕

Then

⊕
Drop activity here or generate with Autopilot

Else

⊕

Set Range Color

Color *

`{}` System.Drawing.Color.Red ⌐ ⊕

Range *

`{}` "B" & (item.Table.Rows.IndexOf(item) + 2) ⌐ ⊕

SheetName *

`{}` "Data" ⌐ ⊕

⊕

Delay

Duration *

00h 00m 05.000s ⊕

If

Condition *

`{}` Not String.IsNullOrEmpty(item("Email").To ⌐ ⊕

Then

⊕
Drop activity here or generate with Autopilot

Else

⊕

Set Range Color

Color *

`{}` System.Drawing.Color.Red ⌐ ⊕

Range *

`{}` "C" & (item.Table.Rows.IndexOf(item) + 2) ⌐ ⊕

SheetName *

`{}` "Data" ⌐ ⊕

The above process represents the workflow to validate the excel file using uipath studio.

REFERENCES

[1]. Microsoft. "Data Validation in Excel." https://support.microsoft.com

[2] UiPath Official Documentation. https://docs.uipath.com

[3] K. S. Mahapatra, "Error Detection and Highlighting in Data Processing," International Journal of Computer Applications, vol. 182, no. 22, 2022, doi: 10.5120/ijca2022922131.

[4] P. Smith, J. Adams, "Automating Excel Validation Using RPA," Proceedings of the International Conference on Automation and Data Science, 2021, doi: 10.1109/ICADS.2021.2033219.

[5] PlantUML Official Documentation. https://plantuml.com/documentation

[6] B. Kumar and N. Gupta, "Robust Exception Handling in RPA Frameworks," Journal of Advanced Automation, vol. 12, pp. 45–53, 2020, doi: 10.1016/j.jaautomation.2020.06.003.

[7] A. Johnson, "Agile Methodology and RPA Integration for Process Optimization," Automation Today, vol. 18, no. 3, pp. 22–28, 2023, doi: 10.1111/aut.2023.18322.