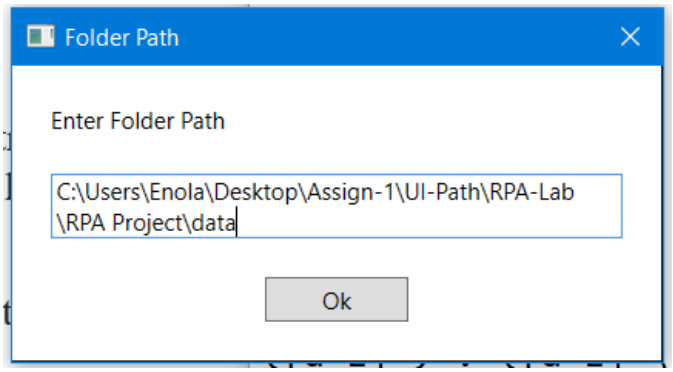
5 Modules

5.1 File Module

This module prompts the user to select a particular location within the system, a location from which the user wants to extract the data from the files present in that particular directory.

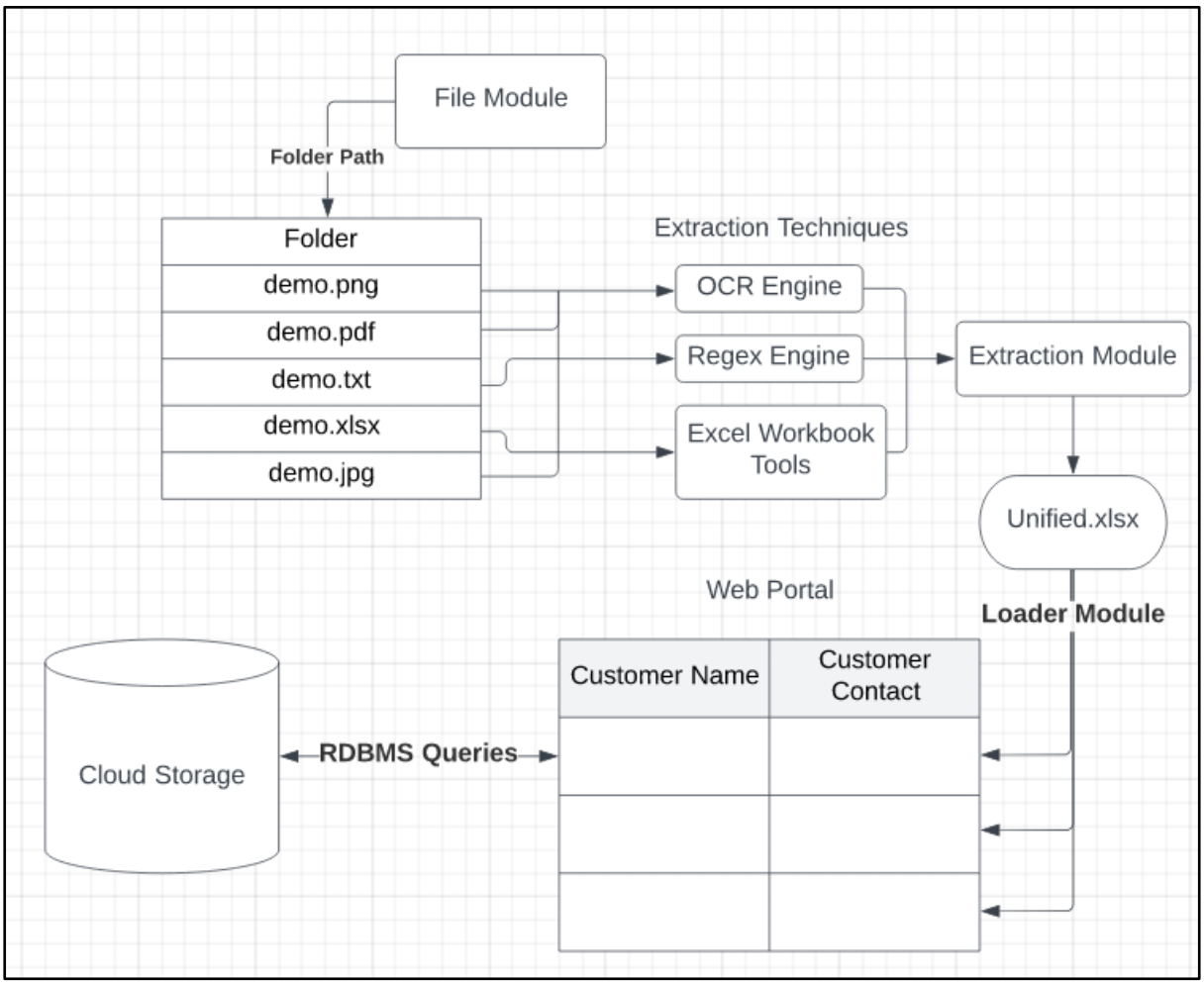


5.2 Extraction Module

This module is responsible for extracting data from the files of different types from the location mentioned by the user through the File Module. The module checks the type of each file using a for each method. After confirming the file type, the module applies the specified technique for data extraction as mentioned in its workflow. For e.g., OCR Engine method for extracting data from PDF and Image files, REGEX for extracting data from text file or xml files or json files and using Excel methods (present in UI Path) to extract data from excel files. After extracting the data from the file using the file specific technique for data extraction the workflow/module stores all the data extracted in a unified excel workbook for further data loading.

5.3 Loader Module. This module is responsible for migration data from the unified excel workbook to the cloud storage via a web portal interface which is hostel online/locally. Using Excel workbook read activities (UI Path) the module reads the data row wise and stores it inside a data table variable. Using Web Recording activities (UI Path) the module loads the data present in the data table variable under the specified tag into the web portal which uses MySQL database as its backend. The web portal uses HTML5, CSS3 and ExpressJS for its frontend and backend consists of MySQL database applied using NodeJS. The module first opens the web portal in the browser application and first identifies the label and enters the appropriate data from the data table variable in the input form under each label and once all input forms are filled the submit button clicked. On clicking the button, the inputted data is converted into an RDBMS insert query by ExpressJS and is sent to the database for insertion, similar steps are followed for record searching process. Hence the module iterates through all the excel data ranges and transfers the data to cloud storage via the web portal with the help of Web Recording functionality of UI Path.

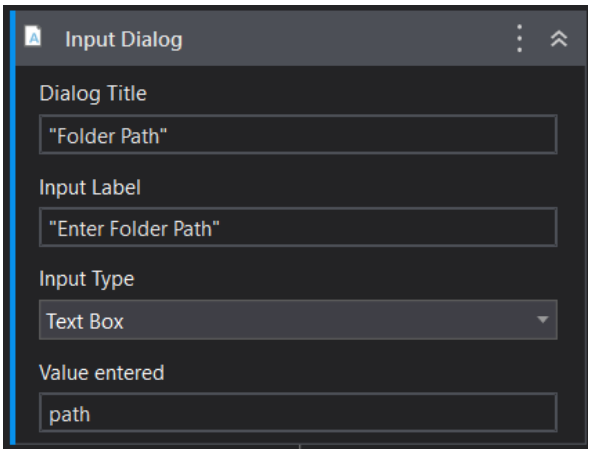
Proposed Model



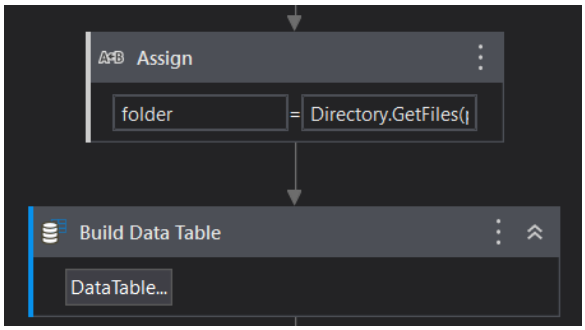
6 Result

Workflow

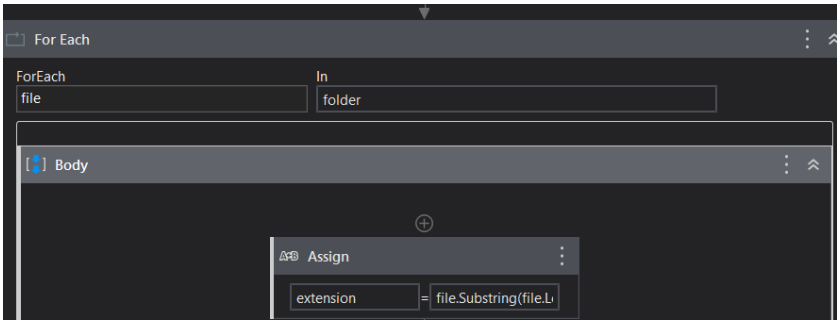
Taking the file path where all the customer data in different file format is stored



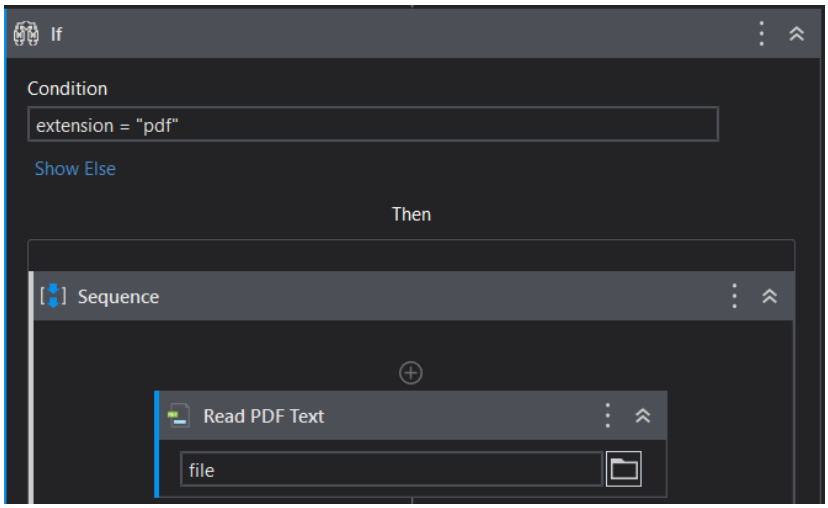
Extracting the list of files present in that directories and building a datatable to store all the customer data present in different file formats.



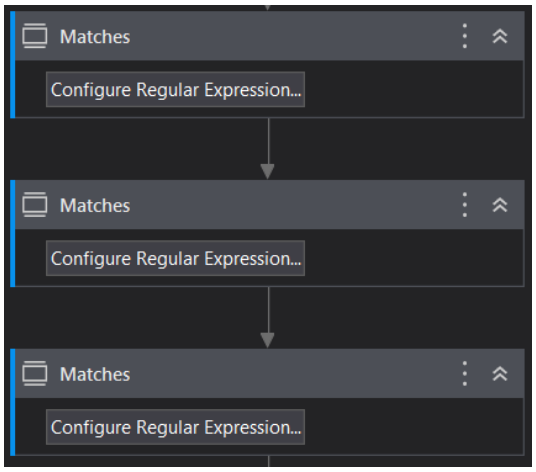
Looping through all the files present in the directory and extracting their file extension to identify the file formats



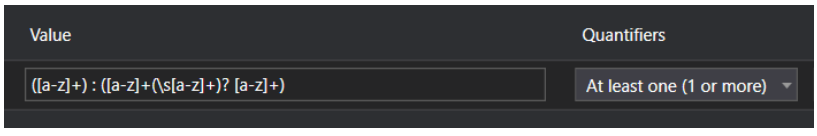
If current file type is of PDF format, extracting the text present in the pdf and applying regex modules to extract customer name, email and phone numbers



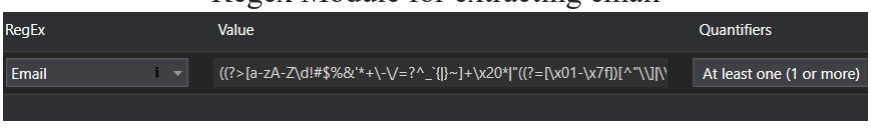
Regex Module



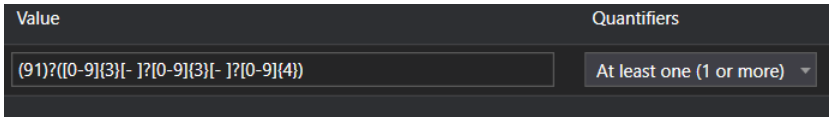
Regex Module for extracting name



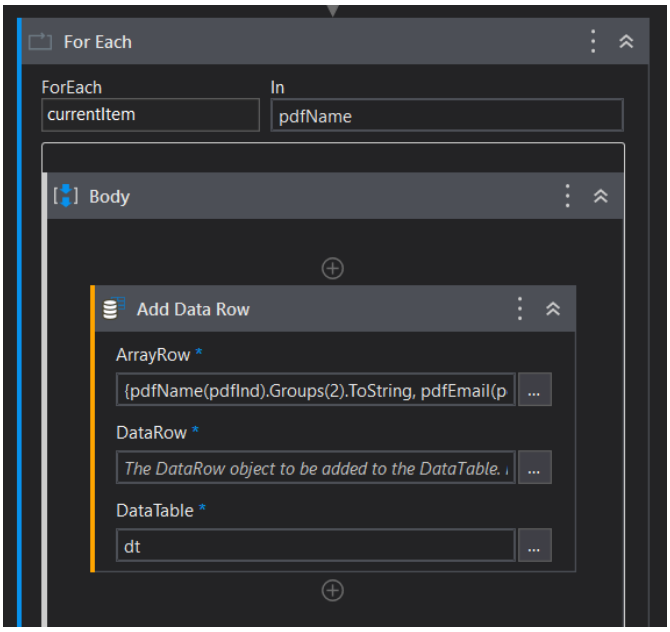
Regex Module for extracting email



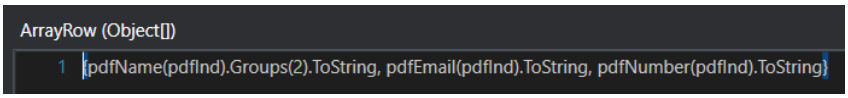
Regex Module for extracting phone number



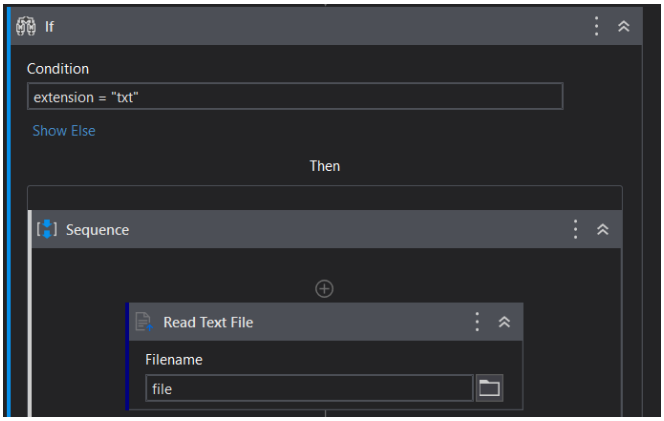
Data Row Append Module for adding the extracted customer data using regex



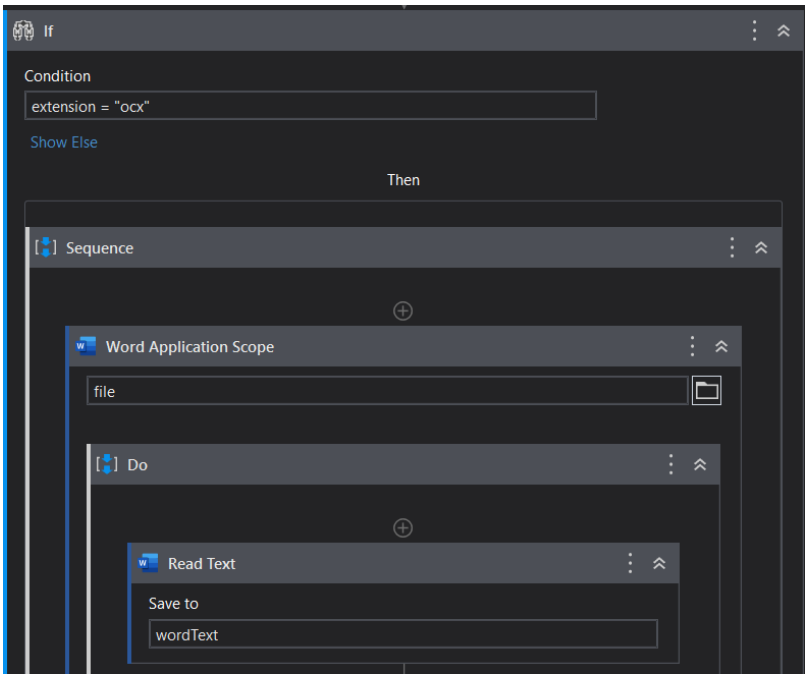
Array Object loaded into the Datatable



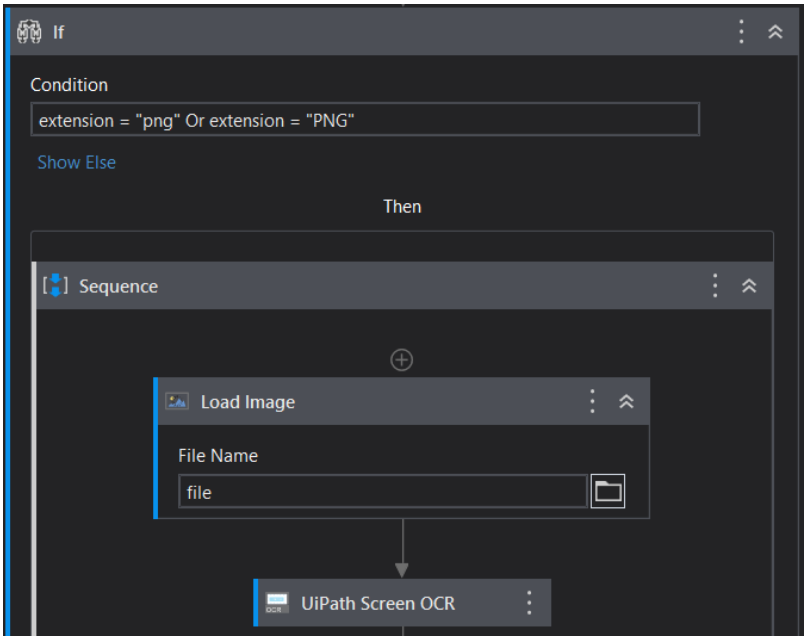
If the file is of type Text (.txt) then we extract the data in text format and then apply the regex module followed by Data row append module to add it to the common datatable



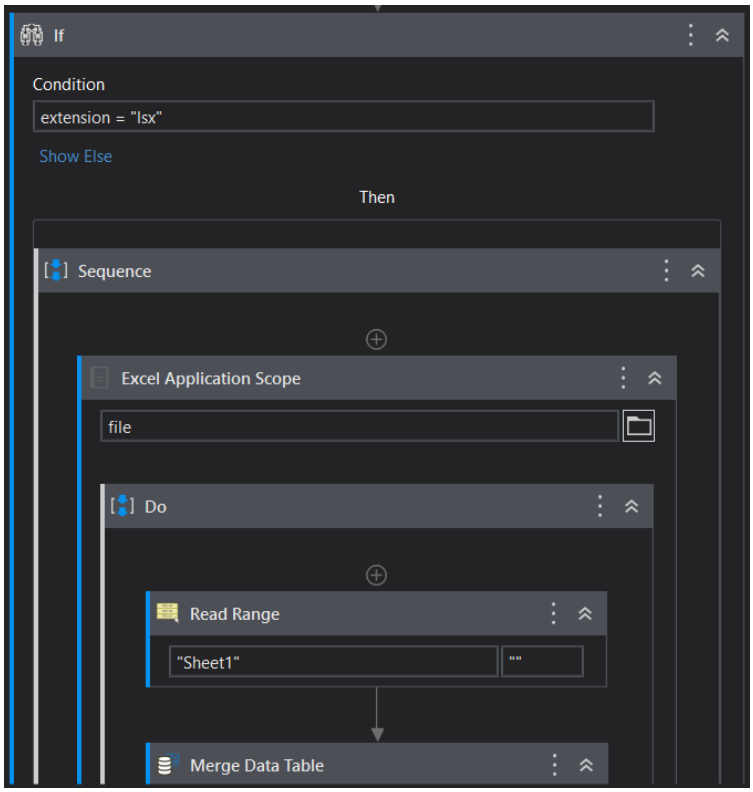
If file is a docx type file then we extract text using the following activities and later we apply the regex module to extract the customer data followed by data row append module to add the data to the common datatable



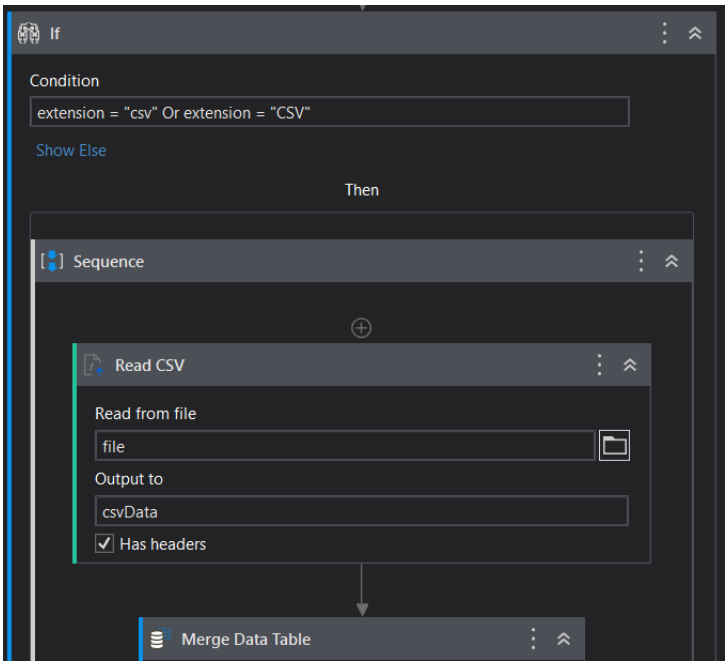
If file is an image file, then we apply OCR activities to extract the data from the image file and later apply regex module and data row append module for further data processing.



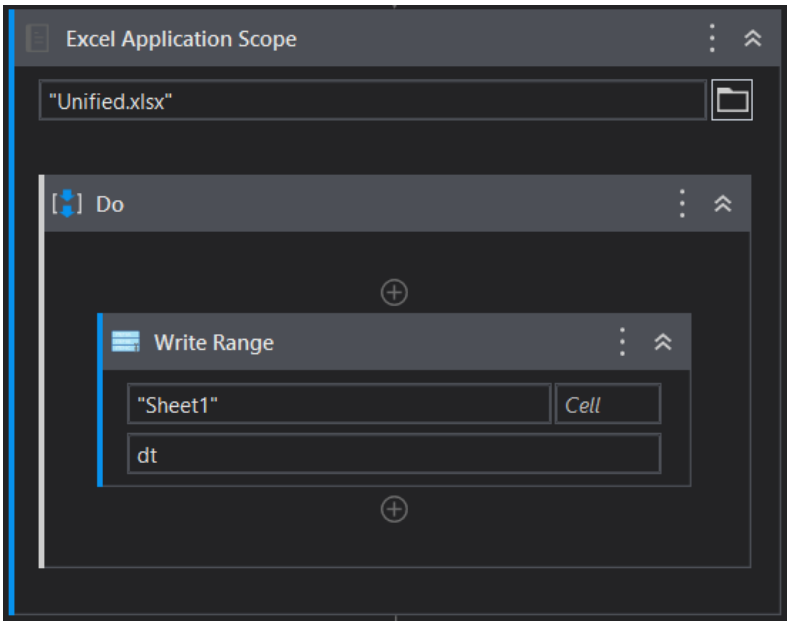
If file is of excel format, then we read the data using read range activity and later we simply merge the data table produced by read range to parent data table where we are storing the unified customer data from all file formats



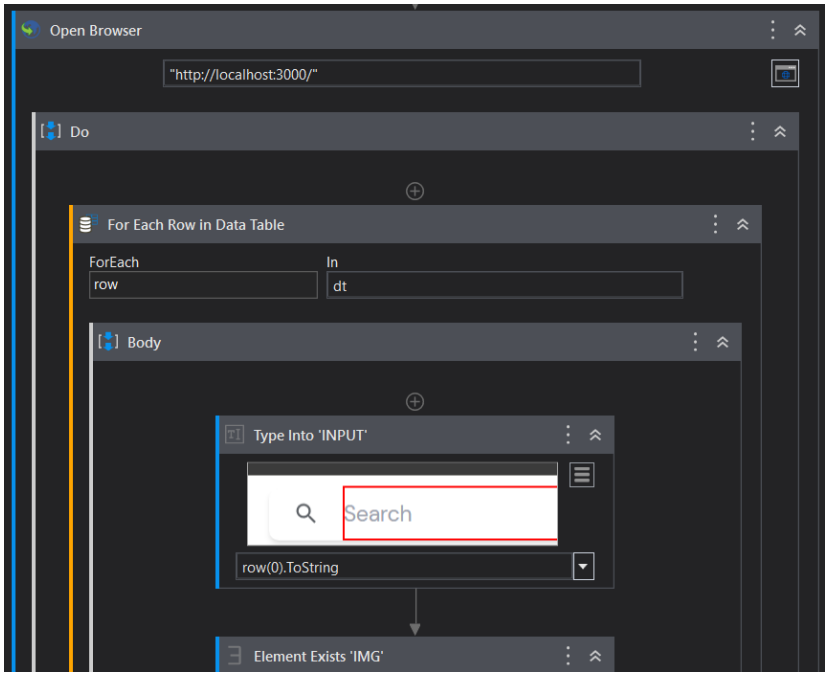
If file of CSV format then we extract the data into a datatable using read CSV activity and merge this extracted data table to the parent unified datatable



After data is extracted from all the files present in the path provided, the datatable holding all the customer information is converted to an excel file

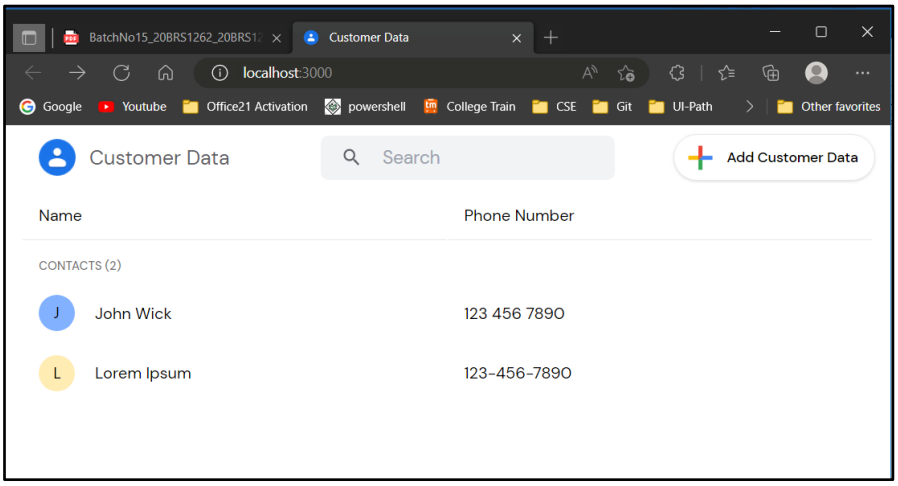


We open a browser and append the data present in the unified excel file to the company's website. Firstly, customer data is extracted row-wise and appending before appending the activities checks whether the customer is already registered in the website by searching in the name in the website search bar and identifying if it has found any match using Find Element activity, if not found then it creates a new customer contact.

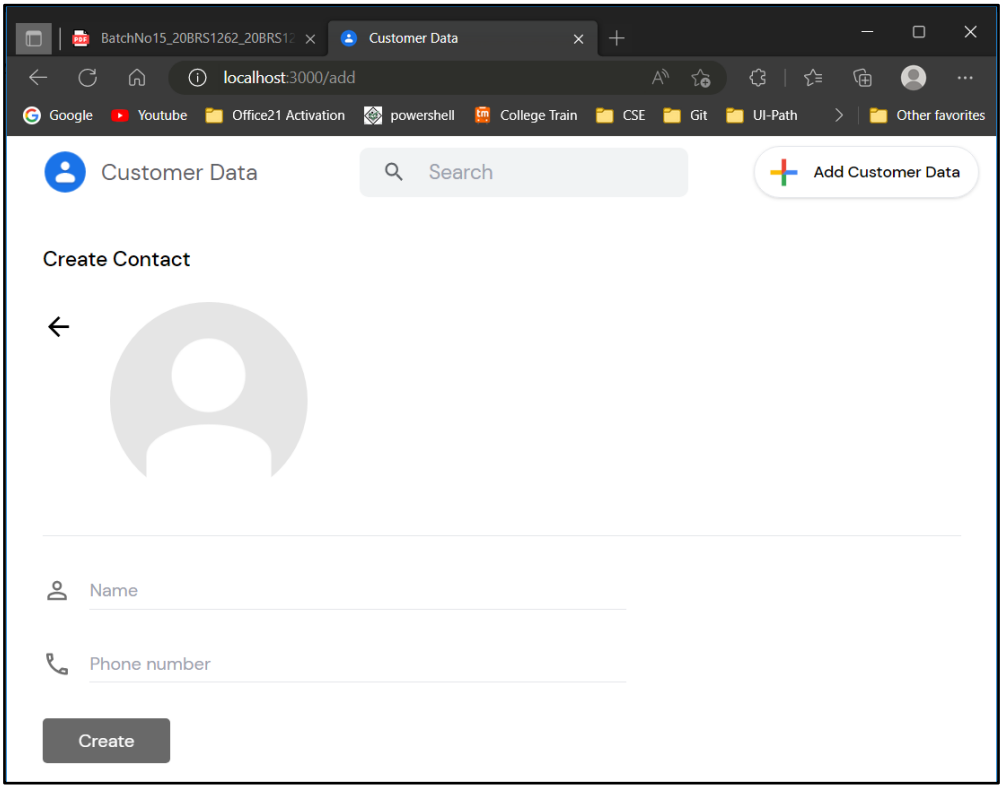


Website

Home Page (Initial State)

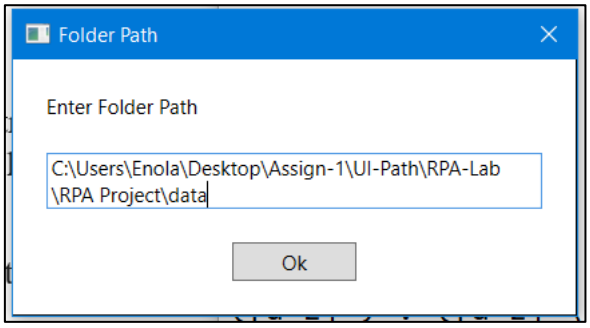


Create Customer Page



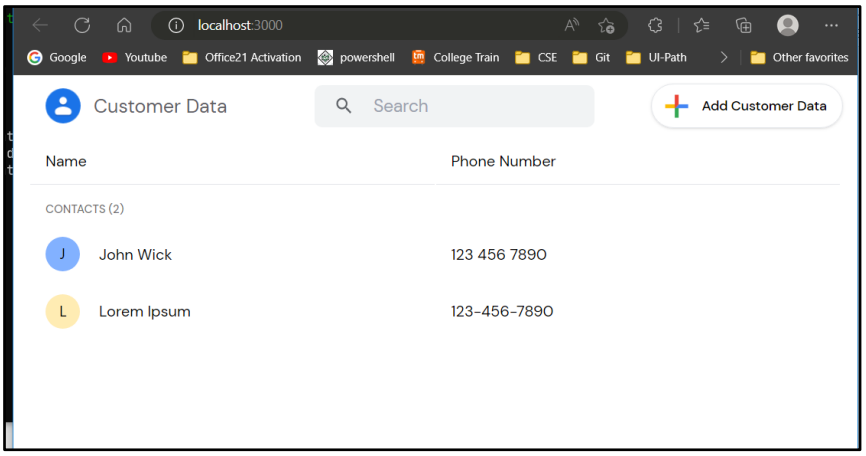
Workflow in action

Step 1: Taking the folder path where the customer data is stored

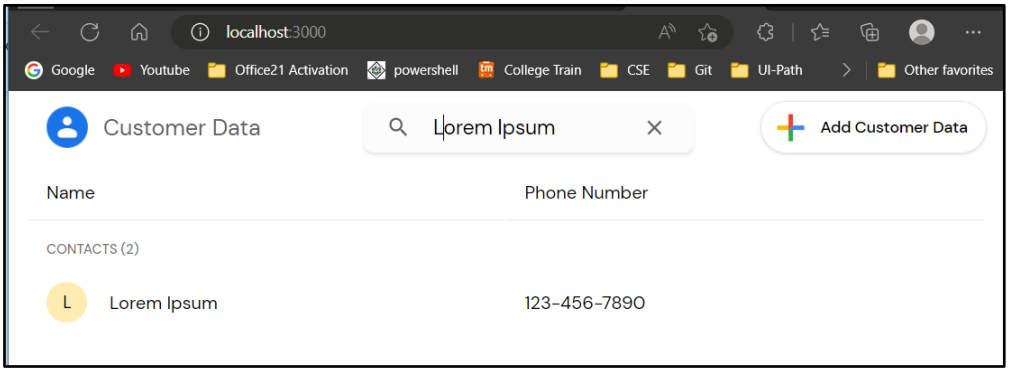


Step 2: The workflow loops through all the files present in the path and extracts data from all the files and stores it a common unified excel file

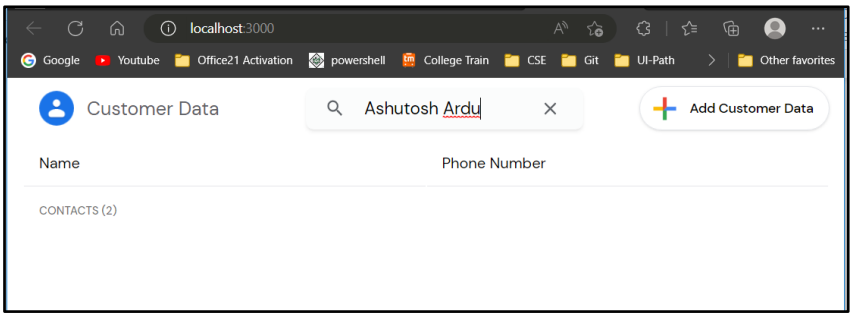
Step 3: The workflow now opens the web portal and loads the customer data onto the server/database.

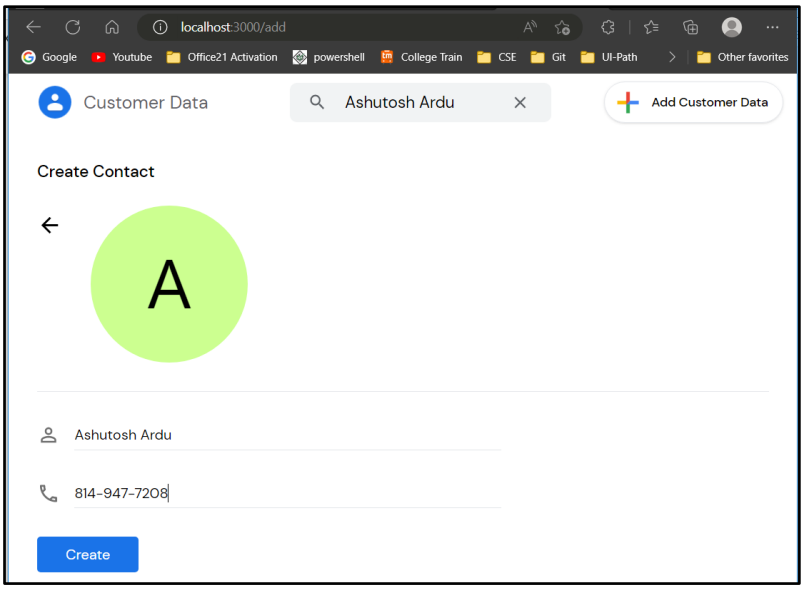


If the customer entry is already present it continues on to the next data row in the unified excel file

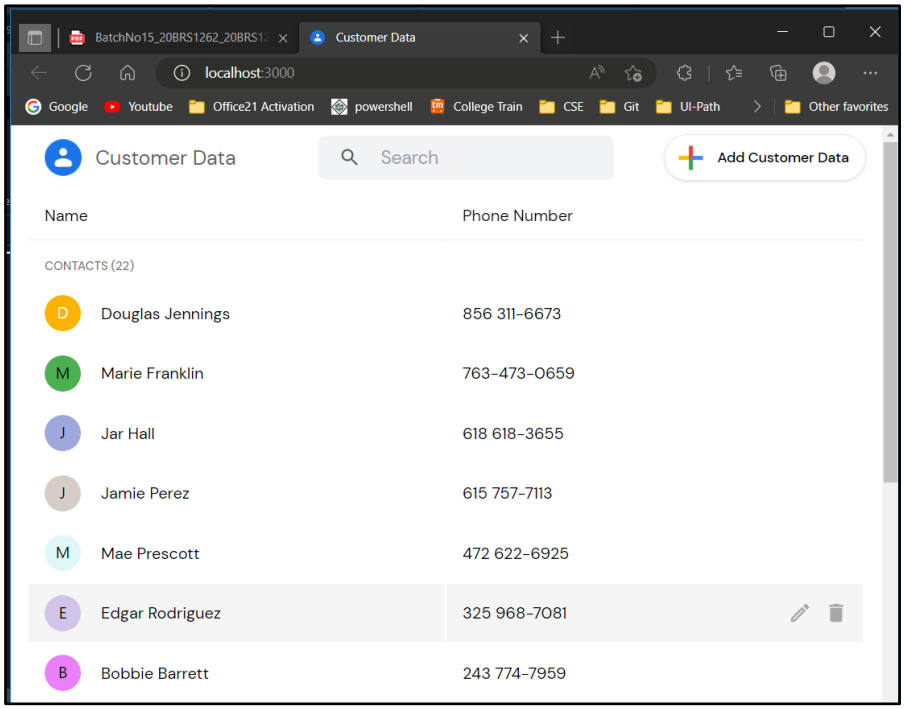


If customer data is not yet entered then it creates a new customer contact





Website (Final State after running the workflow)



7 Conclusion

Data migrations have been present ever since IT solutions were introduced and extensive research has been performed in order to further optimize the migration approach. Nevertheless, the data migration phase remains an activity with many challenges which, to be avoided, require appropriate attention. By means of this project we have attempted to provide insight into the main areas of attention that need to be managed by the project in order to avoid a failing migration. The journey to the successful Data Migration is not an easy one, but if done well, it can have many benefits. However, without the correct precautions and visibility, you can easily end up with suboptimal performance or infrastructure spend. Data migration is a necessary process, but not an impossible one to conquer. The key is to prepare for it very early on, and continue to monitor data migration throughout the life of the project. Project timelines tend to become more rigid as time passes, so it really makes sense to meet migration head on. A devoted team with a clearly defined project plan from the inception of the project, armed with automated tools where applicable, is indeed the formula to success.

8 References

https://wiki.grooper.com/index.php?title=OCR_Engine_-_2021

https://docs.uipath.com/activities/docs/google-ocr

https://www.uipath.com/blog/rpa/the-solution-to-data-migration-troubles
https://www.researchgate.net/publication/336084389_Data_Migration

https://www.researchgate.net/publication/259119355_Data_migration_A_theoretical_perspective

Extraction data from JSON reference

Image text extraction Reference

Website using Reactjs reference

Cloud Data Migration Research Paper

Research Paper on Future Scope of Data Migration

Image Processing Using UIPath

Extracting Data from PDF

Backend for React JS using Express