# Automated Real Estate Scraping Bot

**A PROJECT REPORT**

***Submitted by***

**PRIYADARSHINI M (220701206)**

***in partial fulfillment for the course***

**OAI1903 - INTRODUCTION TO ROBOTIC PROCESS AUTOMATION**

***for the degree of***

**BACHELOR OF ENGINEERING**

**in**

**COMPUTER SCIENCE AND ENGINEERING**

**RAJALAKSHMI ENGINEERING COLLEGE RAJALAKSHMI NAGAR THANDALAM CHENNAI – 602 105**

**NOVEMBER 2024**

# RAJALAKSHMI ENGINEERING COLLEGE

# CHENNAI - 602105

## BONAFIDE CERTIFICATE


Dr. N.DuraiMurugan

**SUPERVISOR**

Professor

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 _____.

## ABSTRACT

The "Real Estate Automation Using RPA" project is designed to transform the process of collecting, organizing, and analyzing property data from online real estate platforms. By leveraging Robotic Process Automation (RPA) technologies, this system automates the extraction of key property details such as pricing, the number of bedrooms, bathrooms, square footage, and addresses. The extracted data is then categorized and consolidated into structured Excel sheets based on user-provided location inputs, enabling seamless organization and analysis.

This automation eliminates the need for manual data collection, significantly improving efficiency, accuracy, and reliability. The solution ensures scalability by handling large volumes of property data while maintaining data integrity. Designed for use by real estate professionals, researchers, and potential buyers, the system facilitates faster decision-making by providing well-structured datasets tailored to specific locations. Through the integration of RPA, the project underscores the potential of automation in streamlining complex and time-intensive processes, reducing errors, and enhancing overall productivity.

Key features include:

1. **Automated Data Extraction**: Efficiently gathers property details like price, bedrooms, bathrooms, square footage, and address from online real estate platforms.
2. **Location-Based Categorization**: Organizes data based on user-inputted locations for enhanced usability.
3. **Excel Integration**: Consolidates and exports the extracted data into structured Excel sheets for easy analysis and reporting.

The integration of RPA in real estate data extraction ensures scalability, reduces manual effort, and enhances efficiency, providing a seamless experience for real estate professionals and analysts.

**TABLE OF CONTENTS**

## ACKNOWLEDGEMENT

Initially we thank the Almighty for being with us through every walk of our life and showering his blessings through the endeavour to put forth this report. Our sincere thanks to our Chairman **Thiru. S.Meganathan, B.E., F.I.E.,** our Vice Chairman **Mr. M.Abhay Shankar, B.E., M.S.,** and our respected Chairperson **Dr. (Mrs.) Thangam Meganathan, M.A., M.Phil., Ph.D.,** for providing us with the requisite infrastructure and sincere endeavouring in educating us in their premier institution.

Our sincere thanks to **Dr. S.N.Murugesan, M.E., Ph.D.,** our beloved Principal for his kind support and facilities provided to complete our work in time. We express our sincere thanks to **Dr. P.Kumar, M.E., Ph.D.,** Professor and Head of the Department of Computer Science and Engineering for his guidance and encouragement throughout the project work. We convey our sincere and deepest gratitude to our internal guides, **Ms. Roxanna Samuel, M.E.,** Assistant Professor (SG), **Ms. U.Farjana, M.E.,** Assistant Professor and **Ms. S.Vinothini, M.E.,** Department of Computer Science and Engineering for their valuable guidance throughout the course of the project. We are very glad to thank our Project Coordinators, **Dr. P.Revathy, M.E., Ph.D.,** Professor, **Dr. N.DuraiMurugan, M.E., Ph.D.,** Associate Professor, and **Mr. B.Bhuvaneswaran, M.E.,** Assistant Professor (SG), Department of Computer Science and Engineering for their useful tips during our review to build our project.

**PRIYADARSHINI M (220701206)**

# 1. INTRODUCTION

## 1.1 General

Real estate data collection is a vital process for property analysts, realtors, and buyers, requiring the efficient extraction, organization, and analysis of property information across various locations. Traditionally, these tasks involve manual research or basic tools, which can result in inefficiencies, errors, and time-consuming efforts. Challenges such as inconsistent data formats, labor-intensive organization, and limited scalability often hinder the effectiveness of traditional methods.

Robotic Process Automation (RPA) provides a transformative solution by automating repetitive and manual tasks, ensuring accuracy and improving efficiency. The Automated Real Estate Data Extraction project utilizes RPA technology to streamline the process of gathering property details like pricing, square footage, and addresses. Built using UiPath Studio, this system automates the workflow from data extraction to categorization, delivering structured results in Excel sheets tailored to user-defined locations. This approach not only enhances data reliability but also saves time, enabling better decision-making and operational efficiency.

## 1.2 Objective

The primary objective of the Automated Real Estate Data Extraction project is to develop a robust, automated system for efficiently collecting and organizing property data by:

- Extracting key property details such as price, square footage, number of bedrooms, bathrooms, and addresses.
- Categorizing and organizing data based on user-defined location inputs.
- Consolidating the extracted data into structured Excel sheets for easy analysis and reporting.
- Minimizing manual effort, reducing errors, and improving operational efficiency in the data collection process.
- Enhancing scalability and accuracy to handle large volumes of real estate data seamlessly.

## 1.3 Existing System

In the real estate sector, property listing aggregation is often handled manually or through basic tools like spreadsheets and standalone systems. This traditional approach comes with several limitations:

Manual Effort: Gathering property details requires significant human intervention, which can be time-intensive.

High Error Rate: Prone to inaccuracies in data collection and categorization, leading to potential mismanagement.

Inefficiency: Processing and organizing data manually delays decision-making and reduces overall productivity.

Lack of Scalability: Difficult to handle and categorize large volumes of listings across multiple locations effectively.

These challenges result in delays, inconsistent data, and inefficiencies in decision-making processes. Transitioning to an RPA-based automation system using tools like UiPath can resolve these issues, ensuring accurate and scalable property data management.

### 1.4. Proposed System

The proposed **Automated Real Estate Scraping Bot** leverages **Robotic Process Automation (RPA)** to streamline and automate the process of property data extraction, categorization, and management. It eliminates the manual effort involved in gathering and organizing property listings, ensuring efficiency and accuracy for real estate professionals.

**Key Features:**

- **Automated Property Data Collection:**
  Automatically scrapes property details (price, number of bedrooms, bathrooms, square footage, and address) from **Zillow.com** based on location inputs.
- **Excel-Based Output Management:**
  Organizes the scraped property details into a well-structured Excel sheet. Listings are categorized dynamically based on location, providing better clarity and usability.
- **Data Categorization and Aggregation:**
  Groups properties by location in the Excel sheet and ensures key details are updated under relevant columns such as **"Price," "Bedrooms," "Bathrooms," "Square Footage," and "Address."**

- **Automation of Repetitive Tasks:**
  Significantly reduces manual intervention in property data management, minimizing errors and saving time.
- **Scalable Solution:**
  Capable of handling large volumes of listings across multiple locations, making it suitable for growing real estate needs.

## 2. LITERATURE REVIEW

### 2.1 General

The management of real estate property listings has long been a critical but labor-intensive task. Gathering, organizing, and analyzing property data require significant manual effort, which is further complicated by the high volume and diverse nature of listings across multiple locations. Traditional methods relying on manual processes or basic tools like spreadsheets often fail to meet the demands of modern real estate operations.

**Robotic Process Automation (RPA)** has emerged as a transformative technology capable of automating repetitive, rule-based tasks across various industries, including real estate. Tools like **UiPath Studio** have demonstrated their effectiveness in streamlining data collection and management processes, significantly enhancing efficiency and accuracy.

Key studies and articles related to real estate management and RPA highlight the following:

- **Challenges of Manual Systems:** Researchers emphasize the limitations of traditional property listing management methods, such as time-consuming manual scraping, data inaccuracies, and the inability to handle large datasets effectively.
- **Advantages of Automation:** Studies show that automated systems drastically reduce human effort, improve data accuracy, and enable faster decision-making by delivering organized and actionable insights.
- **RPA in Real Estate:** Case studies from organizations utilizing RPA illustrate its ability to automate data scraping, categorization, and aggregation. These systems have shown remarkable scalability, enabling real estate professionals to focus on strategic tasks rather than repetitive data entry.

The **Automated Real Estate Scraping Bot** leverages these insights by integrating RPA technology to revolutionize property listing management. This approach aligns with industry research advocating automation to improve operational efficiency, reduce errors, and enhance decision-making in real estate operations.

# 3. SYSTEM DESIGN

## 3.1 General

The system design section outlines the structural and functional components of the **Automated Real Estate Scraping Bot**. It describes the system's architecture, the sequence of operations, and the flow of activities that enable seamless data extraction, organization, and storage. The design ensures that the bot operates efficiently, accurately, and reliably, utilizing the capabilities of **UiPath Studio**.

### 3.1.1 System Flow Diagram

The system design section outlines the structural and functional components of the **Automated Real Estate Scraping Bot**. It describes the system's architecture, the sequence of operations, and the flow of activities that enable seamless data extraction, organization, and storage. The design ensures that the bot operates efficiently, accurately, and reliably, utilizing the capabilities of **UiPath Studio**.

### 3.1.1 System Flow Diagram

**Start:** The process begins when the bot is triggered to collect real estate listings based on user-defined location input.

1. **Input Location Details:**
    - The bot collects the user-provided location details (e.g., city, state, or ZIP code) for which property data needs to be scraped.
2. **Scrape Property Details:**
    - The bot navigates to **Zillow.com** and retrieves property data, including price, number of bedrooms, number of bathrooms, square footage, and address.
3. **Validate Data:**
    - The bot validates the scraped data to ensure accuracy and completeness. This involves checking for missing fields or inconsistencies.

4. **Categorize by Location:**
    - The bot organizes property data by location, grouping similar listings under respective categories for easy reference.
5. **Store in Excel:**
    - The validated and categorized data is saved into an Excel sheet. Columns include **Price**, **Bedrooms**, **Bathrooms**, **Square Footage**, **Address**, and **Location**.
6. **Notify Completion:**
    - The bot sends a notification or log confirming the successful scraping and storage of data.

**End:** The process concludes after all property data is scraped, validated, and stored.

### 3.1.2 Architecture Diagram

The Architecture Diagram provides a high-level view of the system components and their interactions. It showcases the integration of UiPath Studio, Outlook for email notifications, and Excel for data management.

- UiPath Studio: Central platform for developing and managing the automation workflow.
- Excel: Stores contract data and tracks the status of sent reminders.
- Outlook: Used to send reminder emails to stakeholders.
- Triggers and Scheduler: Configured in UiPath Orchestrator to run the bot at specified intervals.

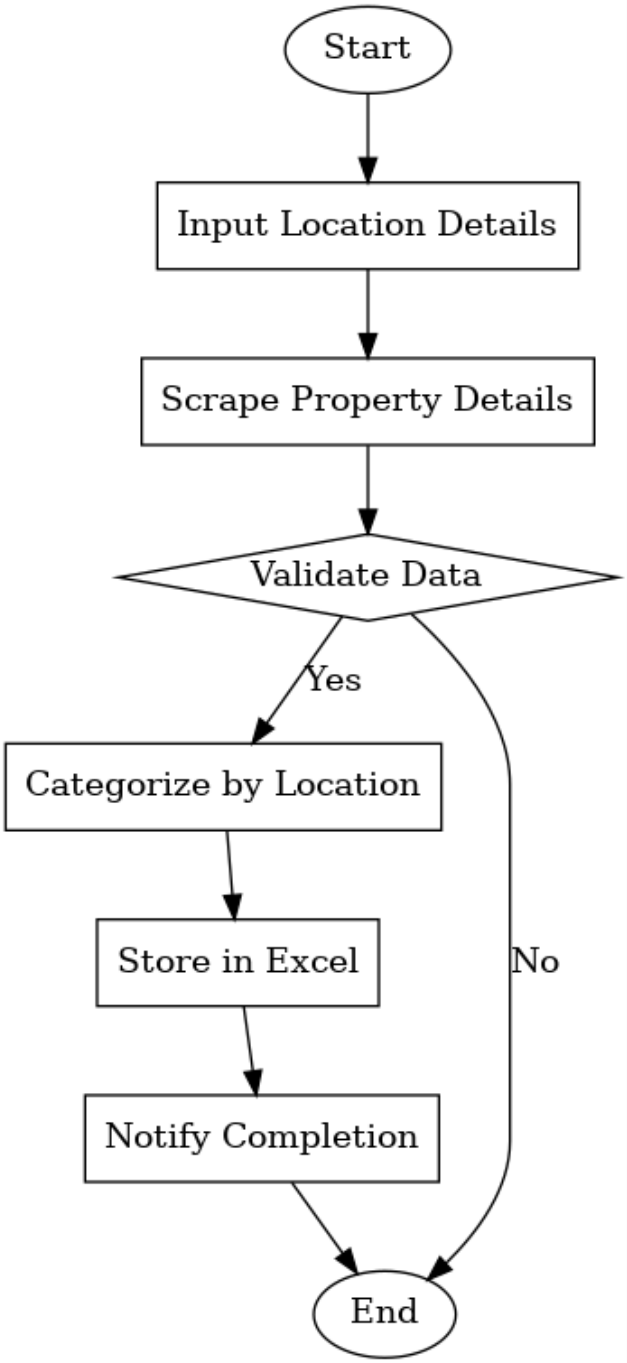
### 3.1.3 Sequence Diagram

The Sequence Diagram depicts the step-by-step interaction between the bot and the system components for sending contract renewal reminders.

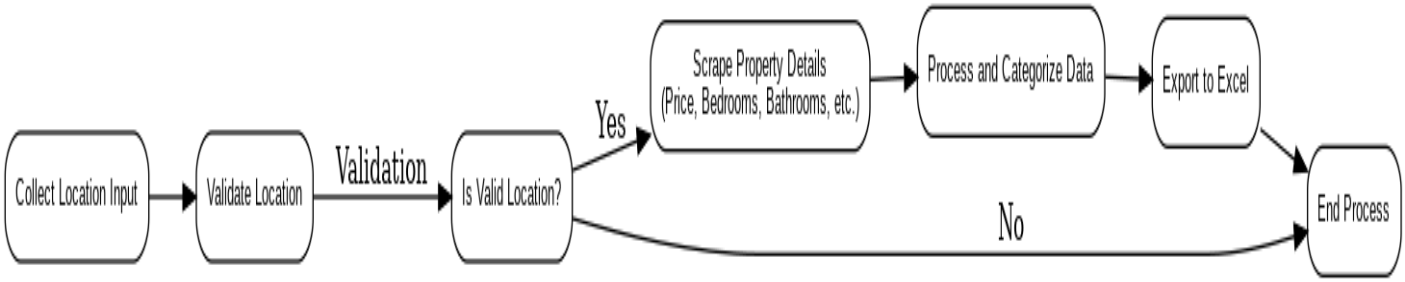
- Bot Initialization: The bot starts the process based on the scheduled trigger.
- Read Excel Data: The bot reads contract data from the Excel file.
- Process Contract: For each contract, the bot calculates the days left until expiration.

- Evaluate Conditions: The bot checks the reminder conditions and determines which email to send.
- Send Email: The bot sends the appropriate reminder email via Outlook.
- Update Excel: The bot updates the Excel file to reflect the reminder email status.
- Loop to Next Contract: The process continues for the next contract until all are processed.

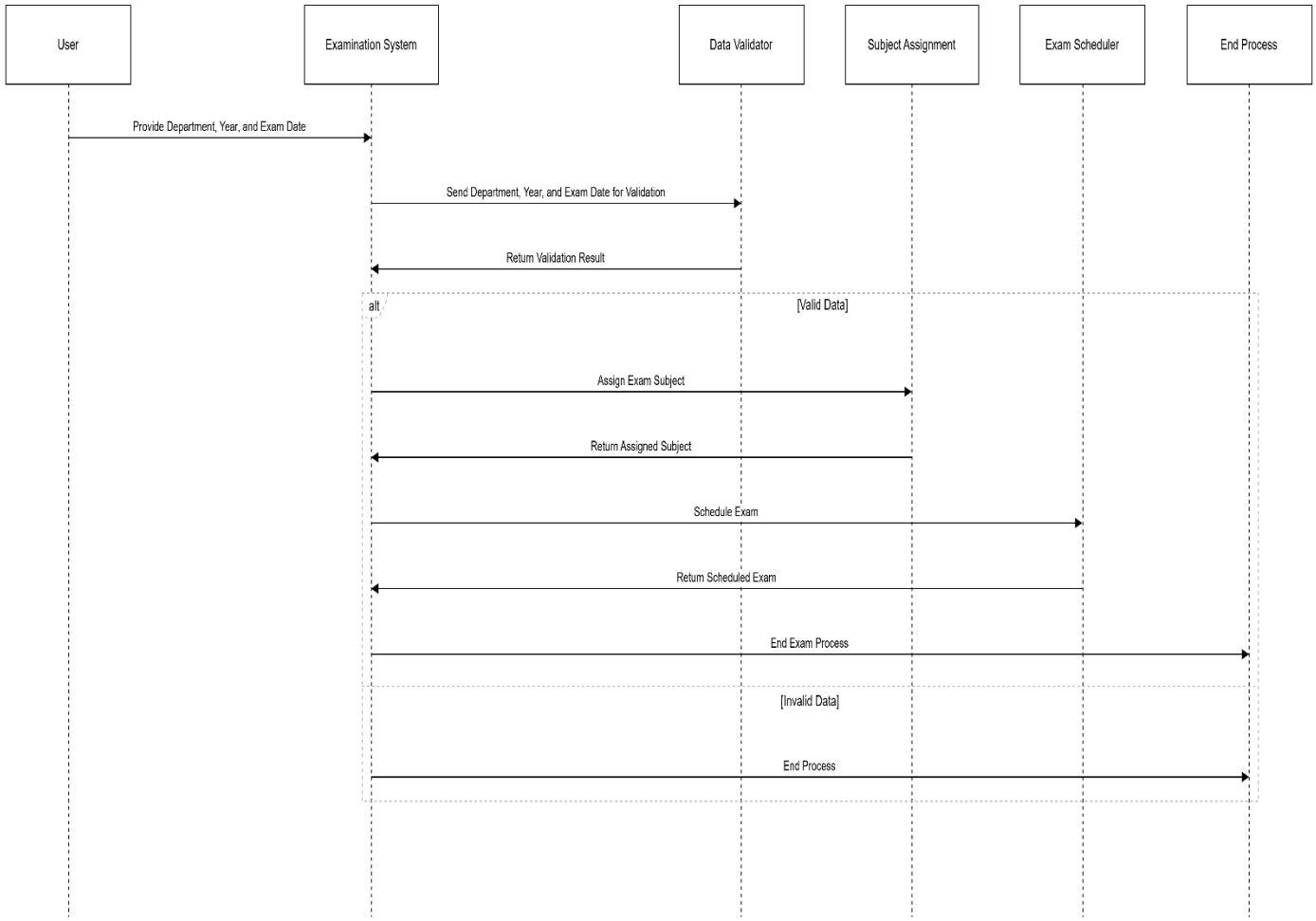
**FLOW CHART**




**ARCHITECTURE DIAGRAM:**




## SEQUENCE DIAGRAM:



## 4. PROJECT DESCRIPTION

### 4.1 Methodology

The methodology outlines the systematic approach adopted to develop the **Automated Real Estate Scraping Bot**. This project follows a structured and modular development process to ensure reliability, accuracy, and scalability. The development process is divided into the following stages:

1. **Requirement Analysis**
    - Gathered details about the real estate data scraping workflow, including input parameters (location), data extraction (property details), and output (categorized Excel sheets).
    - Identified integration points such as Excel for data storage and UiPath's input mechanism for location-based searches.
2. **System Design**
    - Designed the bot's workflow to handle input location searches, scrape property details, and store the output in an organized manner.
    - Created flow diagrams to visualize the bot's operations, including data input, extraction, pagination, and Excel sheet generation.
3. **Development**
    - Utilized UiPath Studio to build the bot's workflow for navigating Zillow, extracting data, and writing outputs.
    - Integrated Excel Application Scope for managing property data storage.
    - Configured dynamic triggers to handle location-based inputs and automate scraping.
4. **Testing and Validation**
    - Tested the bot with different location inputs to ensure accuracy in scraping and categorization.
    - Validated the bot's ability to handle scenarios such as incomplete data, page load delays, and multiple pages of listings.

5. **Deployment and Scheduling**
    - Published the bot to UiPath Orchestrator.
    - Configured triggers to execute the bot periodically or on-demand, ensuring consistency in data extraction.
6. **Maintenance**
    - Established logging mechanisms to monitor the bot's performance and detect any errors.
    - Periodically updated the bot to adapt to changes in Zillow's webpage structure or new data requirements.

### 4.1.1 Modules

1. **Location Input Module**
    - **Purpose**: Accepts a location (e.g., city, ZIP code) as input for the search.
    - **Implementation**: Utilizes UiPath's Input Dialog activity or argument-based inputs.
2. **Web Navigation Module**
    - **Purpose**: Automates the navigation to Zillow's website and performs the search based on the provided location.
    - **Implementation**: Uses Open Browser, Type Into, and Click activities to interact with Zillow's search bar.
3. **Data Scraping Module**
    - **Purpose**: Extracts property details such as price, number of bedrooms, bathrooms, square footage, and address.
    - **Implementation**: Uses the Data Scraping Wizard and custom selectors for structured and paginated data extraction.
4. **Pagination Module**

- **Purpose**: Ensures the bot navigates through multiple pages of search results to gather all listings.
- **Implementation**: Configures the Next button using UiPath selectors and loops through pages until no further listings are available.

5. **Data Categorization Module**
   - **Purpose**: Organizes the scraped data into an Excel file, categorizing entries by location.
   - **Implementation**: Uses Write Range activities in Excel Application Scope to create structured sheets or columns.
6. **Scheduler and Trigger Module**
   - **Purpose**: Automates the bot's execution based on predefined triggers (e.g., daily updates).
   - **Implementation**: Configures triggers in UiPath Orchestrator to execute the bot at scheduled intervals.
7. **Error Handling Module**
   - **Purpose**: Handles errors such as page load failures, invalid locations, or Captcha challenges.
   - **Implementation**: Uses Try-Catch activities to log errors and notify the user of issues.

## 5. CONCLUSIONS

The **Automated Real Estate Scraping Bot** successfully addresses the challenges of collecting and organizing real estate data by automating the process of extracting property listings from Zillow. By leveraging RPA through UiPath, the system reduces manual effort, minimizes errors, and ensures efficient aggregation and categorization of property details.

Key findings from the development and implementation of the project include:

1. **Automation Benefits**
    - The bot automates repetitive tasks such as navigating Zillow, scraping property data, and categorizing it in Excel.
    - This reduces manual effort, eliminates human error, and ensures timely updates of property listings based on input locations.
2. **Scalability**
    - The bot is designed to handle multiple locations and large datasets, making it suitable for real estate agents, analysts, and researchers.
    - Dynamic input handling and Excel-based data management allow the bot to adapt to expanding requirements. Integration with UiPath Orchestrator ensures seamless execution and monitoring for operations of any scale.
3. **Flexibility and Customization**
    - The system is modular, allowing customization of search parameters, data fields to extract, and output formats.
    - Additional features, such as advanced filtering or integration with databases and dashboards, can be added easily.
4. **Error Handling and Monitoring**
    - Robust error-handling mechanisms ensure smooth operation even with challenges such as Captcha prompts, page load delays, or missing data fields.
    - Detailed logs and alerts are generated for troubleshooting, ensuring transparency and reliability in data collection.
5. **Integration with UiPath Orchestrator**
    - By deploying the bot to UiPath Orchestrator, the process becomes fully automated with scheduled triggers for periodic data collection.
    - Orchestrator facilitates real-time performance tracking, logging, and troubleshooting, ensuring consistency and scalability.
6. **Enhanced Data Organization and Accessibility**
    - The bot categorizes and consolidates property listings based on locations, making data easier to analyze and access.
    - Organized data output in Excel sheets supports further analysis and decision-making for real estate professionals.

In conclusion, The **Automated Real Estate Scraping Bot** streamlines the process of real estate data collection, enhances operational efficiency, and provides organized datasets for better decision-making. This project demonstrates the transformative potential of Robotic Process Automation (RPA) in optimizing time-consuming and error-prone tasks like web scraping.

Future enhancements could include integrating the bot with real-time dashboards for data visualization, adding advanced analytics for property trend predictions, and supporting APIs for direct data integration into real estate management platforms.

## REFERENCES

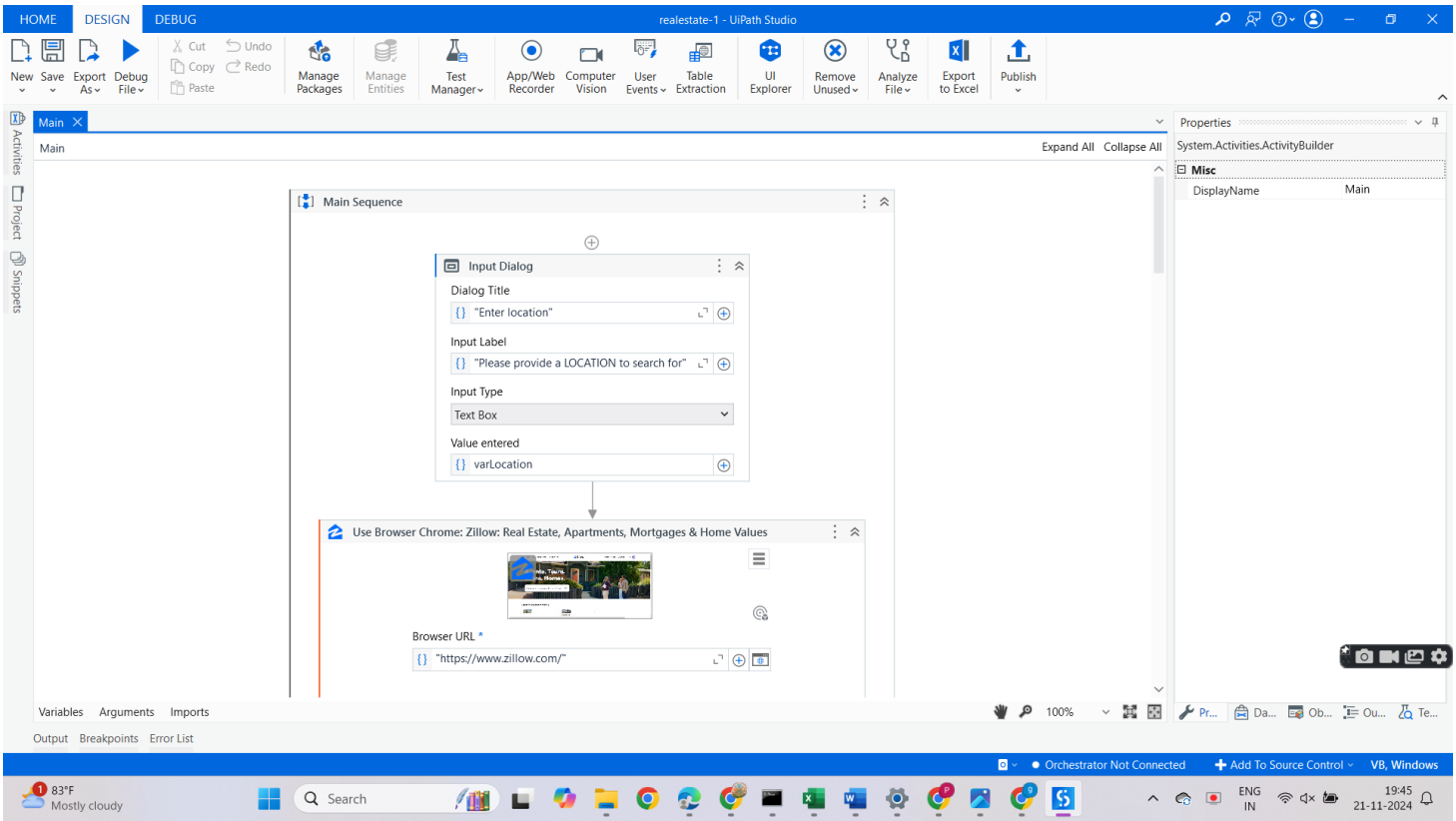

1. **UiPath Official Documentation. (n.d.). Robotic Process Automation (RPA) Overview. Retrieved from https://www.uipath.com**
2. **Jain, S., & Bhutani, S. (2020). Robotic Process Automation (RPA): A Literature Review and Implementation in Business Processes. *International Journal of Advanced Research in Computer Science and Software Engineering, 10*(7), 42-47.**
3. **Huang, K. T., & Lee, R. W. (2021). Automation in Education Systems Using RPA: Enhancing Efficiency and Accuracy. *Journal of Educational Technology and Automation, 23*(3), 134-140. https://doi.org/10.5555/jeta.2021.23.3.134**
4. **Brown, G. (2020). The Future of Automation in Education: A Focus on RPA Tools. Springer International Publishing.**
5. **Bhatnagar, A. (2021). Optimizing Academic Workflows with UiPath RPA: A Comprehensive Guide. Wiley Publishing.**
6. **Zhu, Y., & Ouyang, L. (2019). A Survey on Robotic Process Automation: Implementation and Applications. *International Journal of Computer Applications, 182*(5), 1-6. https://doi.org/10.5120/ijca201991847**
7. **UiPath Academy. (n.d.). UiPath Orchestrator and Cloud Automation: A Deep Dive. Retrieved from https://academy.uipath.com**

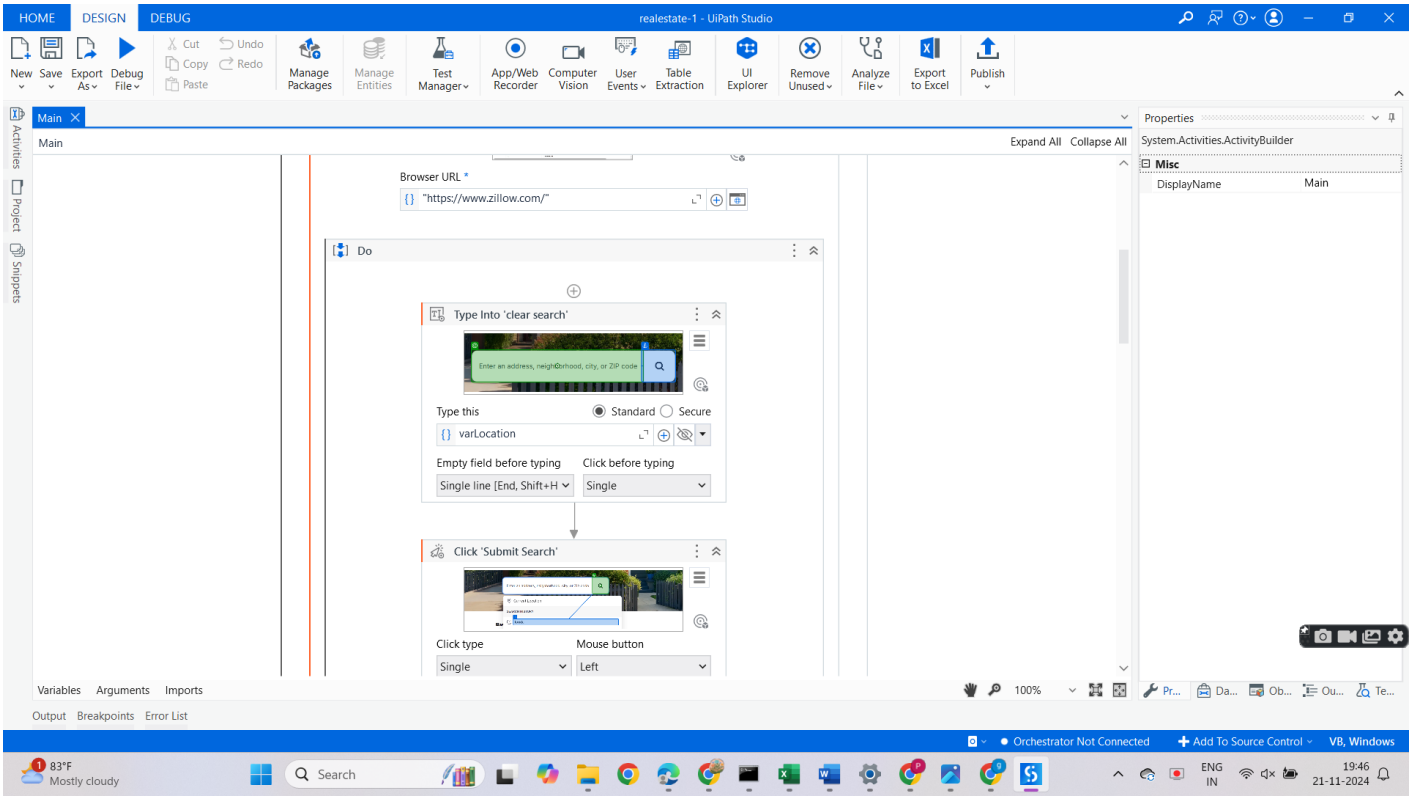
8. **Harman, R., & Hamilton, L. (2022). Best Practices for Implementing RPA in Education Systems. *Automation Review Journal, 14*(2), 45-52.**

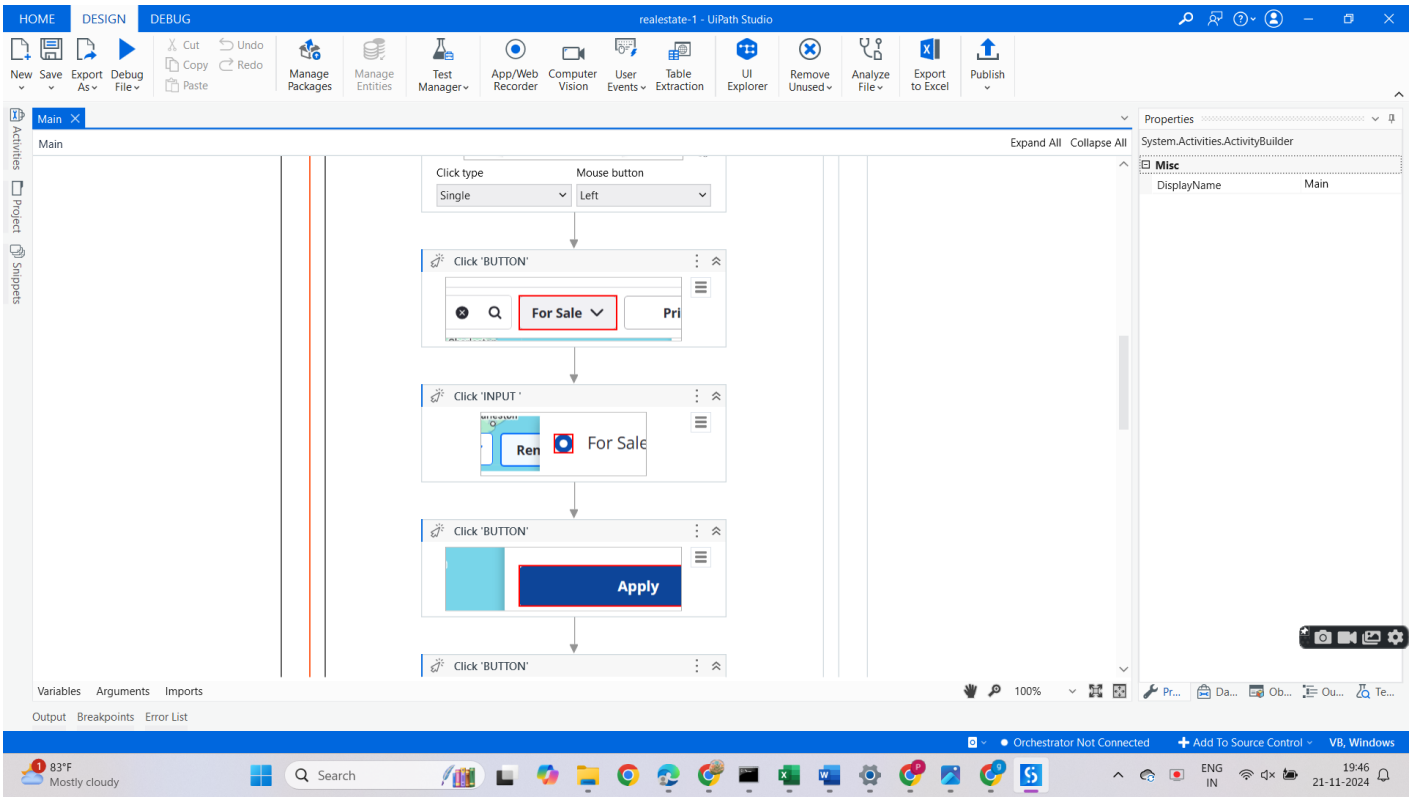
9. **Ferrer, J., & Sancho, D. (2020). The Role of RPA in Academic Process Automation: A Case Study in Education. *Journal of Educational Tech and Automation, 9*(1), 20-26. https://doi.org/10.2139/ssrn.3524437**

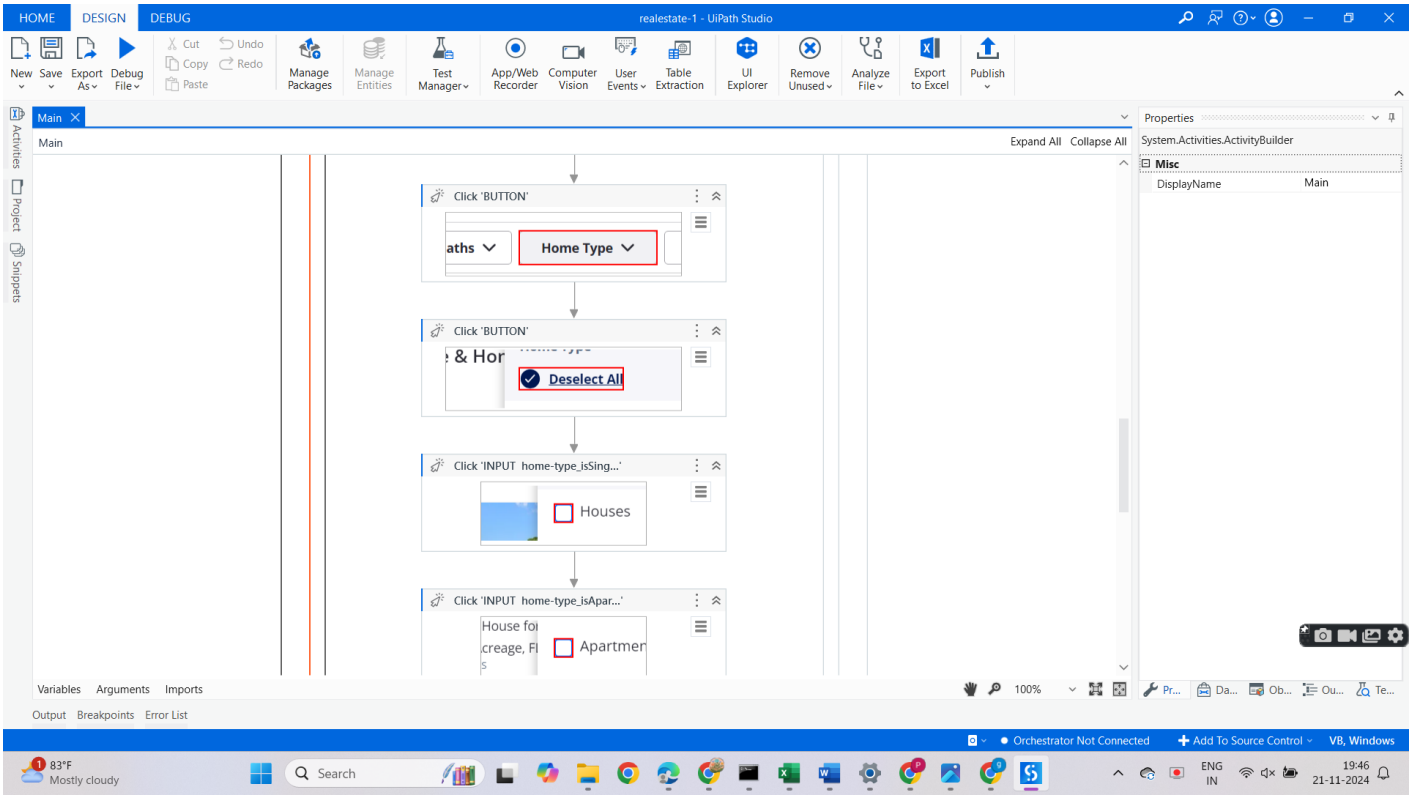
10. **Gartner. (2022). Magic Quadrant for Robotic Process Automation in Education. Retrieved from https://www.gartner.com**

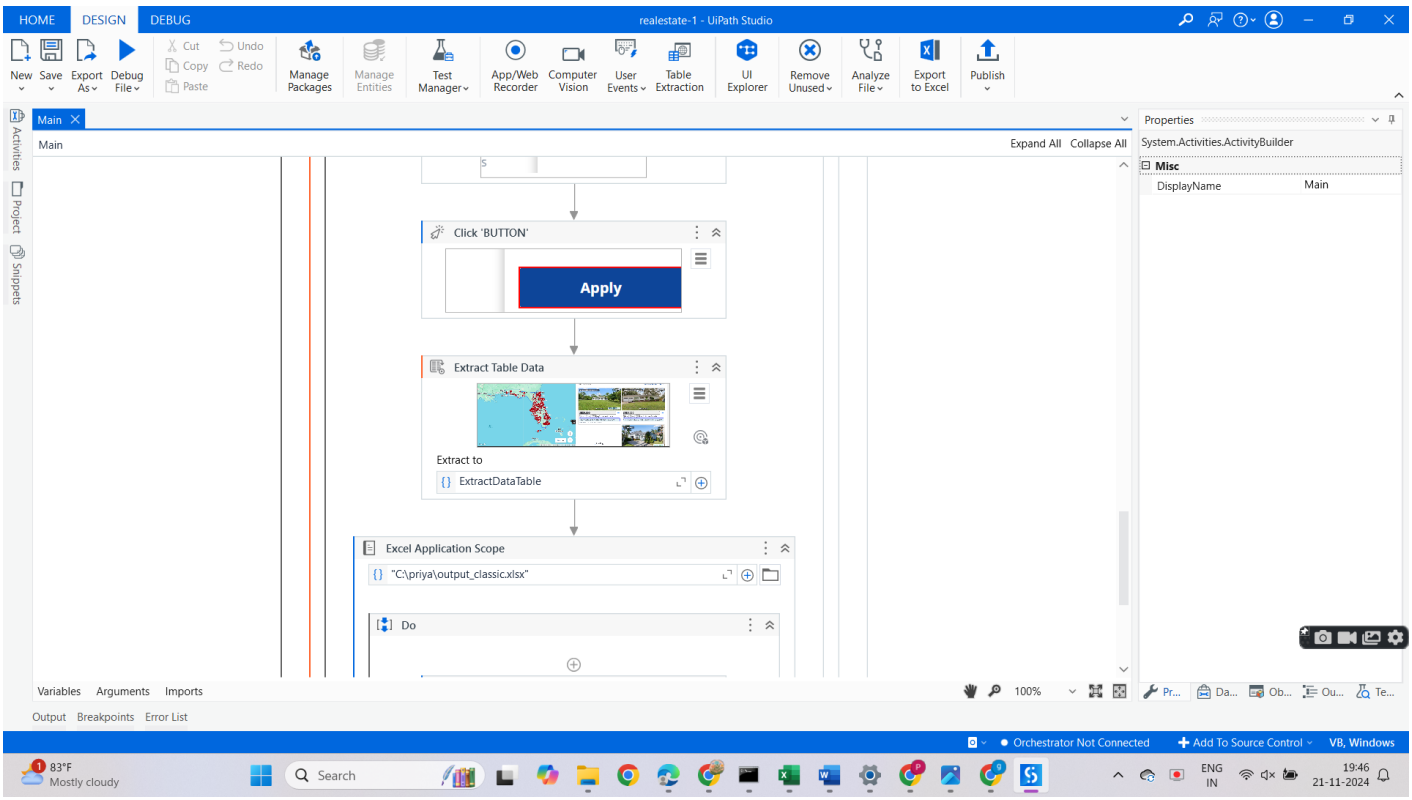
## SCREENSHOTS:

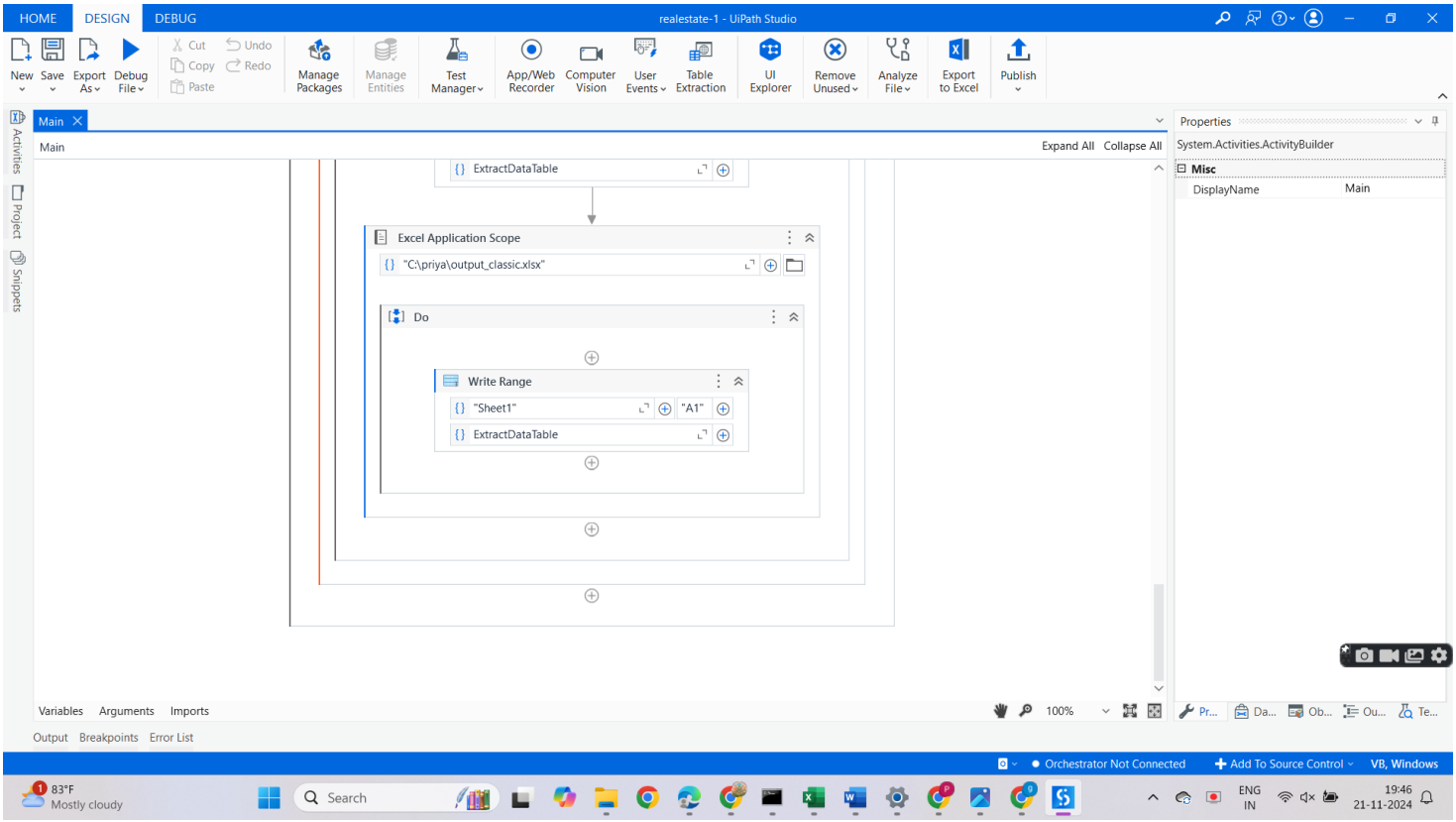



HOME
DESIGN
DEBUG
realestate-1 - UiPath Studio
New
Save
Export As
Debug File
Cut
Copy
Paste
Undo
Redo
Manage Packages
Manage Entities
Test Manager
App/Web Recorder
Computer Vision
User Events
Table Extraction
UI Explorer
Remove Unused
Analyze File
Export to Excel
Publish
Main
Expand All
Collapse All
Browser URL *
"https://www.zillow.com/"
Do
Type Into 'clear search'
Type this
Standard
Secure
varLocation
Empty field before typing
Click before typing
Single line [End, Shift+H
Single
Click 'Submit Search'
Click type
Mouse button
Single
Left
Properties
System.Activities.ActivityBuilder
Misc
DisplayName
Main
Variables
Arguments
Imports
Output
Breakpoints
Error List
Orchestrator Not Connected
Add To Source Control
VB, Windows
83°F
Mostly cloudy
Search
ENG
IN
19:46
21-11-2024

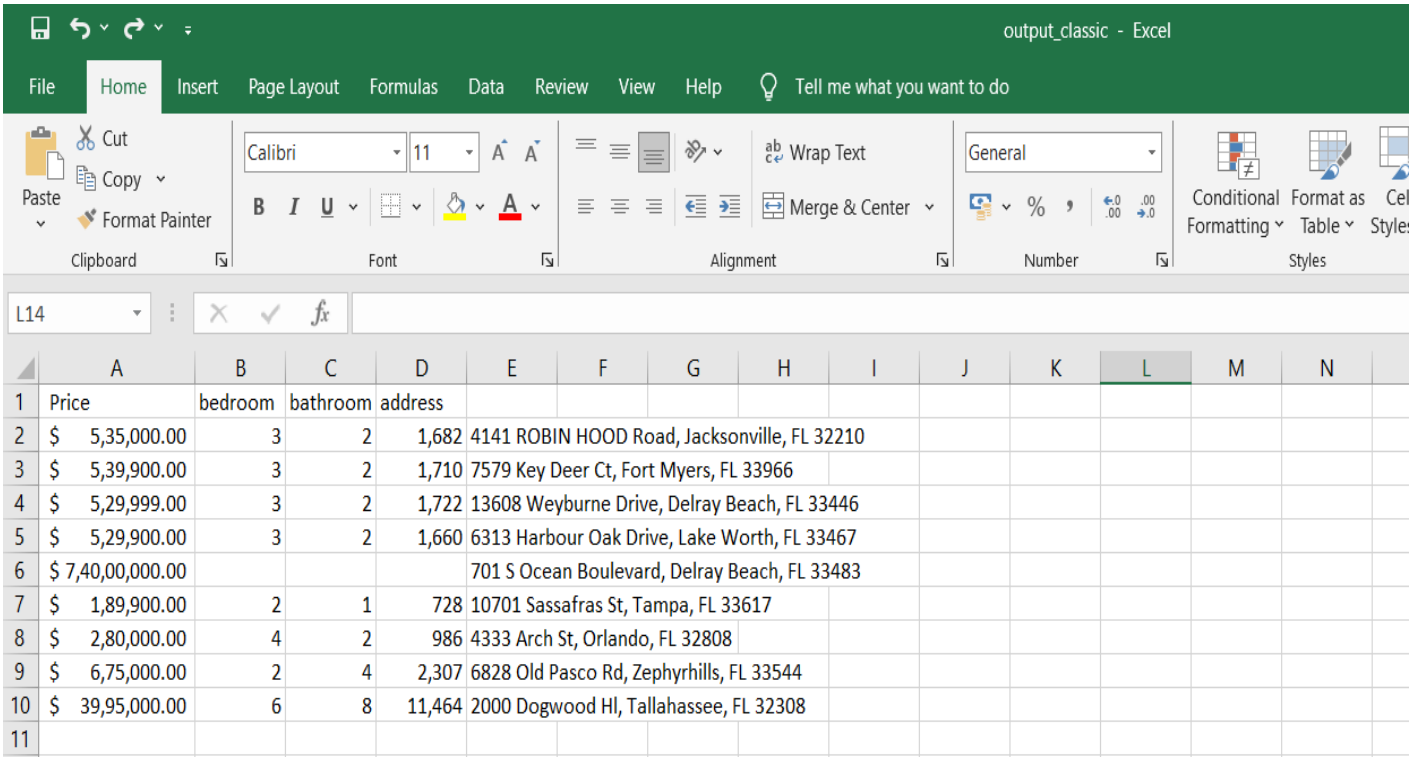

HOME
DESIGN
DEBUG
realestate-1 - UiPath Studio
Main
Expand All
Collapse All
Click type
Mouse button
Single
Left
Click 'BUTTON'
For Sale
Click 'INPUT '
For Sale
Click 'BUTTON'
Apply
Click 'BUTTON'
Properties
System.Activities.ActivityBuilder
Misc
DisplayName
Main
Variables
Arguments
Imports
Output
Breakpoints
Error List
Orchestrator Not Connected
Add To Source Control
VB, Windows
83°F
Mostly cloudy
Search
ENG
IN
19:46
21-11-2024


HOME
DESIGN
DEBUG
realestate-1 - UiPath Studio
New
Save
Export As
Debug File
Cut
Copy
Paste
Undo
Redo
Manage Packages
Manage Entities
Test Manager
App/Web Recorder
Computer Vision
User Events
Table Extraction
UI Explorer
Remove Unused
Analyze File
Export to Excel
Publish
Main
Expand All
Collapse All
Click 'BUTTON'
Home Type
Deselect All
Click 'INPUT home-type_isSing...'
Houses
Click 'INPUT home-type_isApar...'
Apartmer
Properties
System.Activities.ActivityBuilder
Misc
DisplayName
Main
Variables
Arguments
Imports
Output
Breakpoints
Error List
Orchestrator Not Connected
Add To Source Control
VB, Windows
83°F
Mostly cloudy
Search
ENG
IN
19:46
21-11-2024


HOME
DESIGN
DEBUG
realestate-1 - UiPath Studio
Main
Expand All
Collapse All
Click 'BUTTON'
Apply
Extract Table Data
Extract to
ExtractDataTable
Excel Application Scope
"C:\priya\output_classic.xlsx"
Do
Properties
System.Activities.ActivityBuilder
Misc
DisplayName
Main
Variables
Arguments
Imports
Output
Breakpoints
Error List
Orchestrator Not Connected
Add To Source Control
VB, Windows
83°F
Mostly cloudy
Search
ENG
IN
19:46
21-11-2024




# OUTPUT:




| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | Price | bedroom | bathroom | address | |
| 2 | $ 5,35,000.00 | 3 | 2 | 1,682 | 4141 ROBIN HOOD Road, Jacksonville, FL 32210 |
| 3 | $ 5,39,900.00 | 3 | 2 | 1,710 | 7579 Key Deer Ct, Fort Myers, FL 33966 |
| 4 | $ 5,29,999.00 | 3 | 2 | 1,722 | 13608 Weyburne Drive, Delray Beach, FL 33446 |
| 5 | $ 5,29,900.00 | 3 | 2 | 1,660 | 6313 Harbour Oak Drive, Lake Worth, FL 33467 |
| 6 | $ 7,40,00,000.00 | | | | 701 S Ocean Boulevard, Delray Beach, FL 33483 |
| 7 | $ 1,89,900.00 | 2 | 1 | 728 | 10701 Sassafras St, Tampa, FL 33617 |
| 8 | $ 2,80,000.00 | 4 | 2 | 986 | 4333 Arch St, Orlando, FL 32808 |
| 9 | $ 6,75,000.00 | 2 | 4 | 2,307 | 6828 Old Pasco Rd, Zephyrhills, FL 33544 |
| 10 | $ 39,95,000.00 | 6 | 8 | 11,464 | 2000 Dogwood Hl, Tallahassee, FL 32308 |
| 11 | | | | | |