

## Tutorial 2 – Scraping Data from the Web

v003

### Contents

<b>1. Scraping from a Website and Storing the Data in a Text Editor .....</b>	<b>2</b>
1.1. Create a new project .....	2
1.2. Add Open Browser Activity .....	3
1.3. Scrape Textual Data from the Web .....	4
1.4. Create a String Variable .....	6
1.5. Add Open Application Activity (Word) .....	6
<b>2. Scrape Stock Market Data from the Web and Putting Them Into Excel File .....</b>	<b>8</b>
2.1. Create a new UiPath Project .....	8
2.2. Add Open Browser Activity .....	9
2.3. Data Scraping .....	9

# 1. Scraping from a Website and Storing the Data in a Text Editor

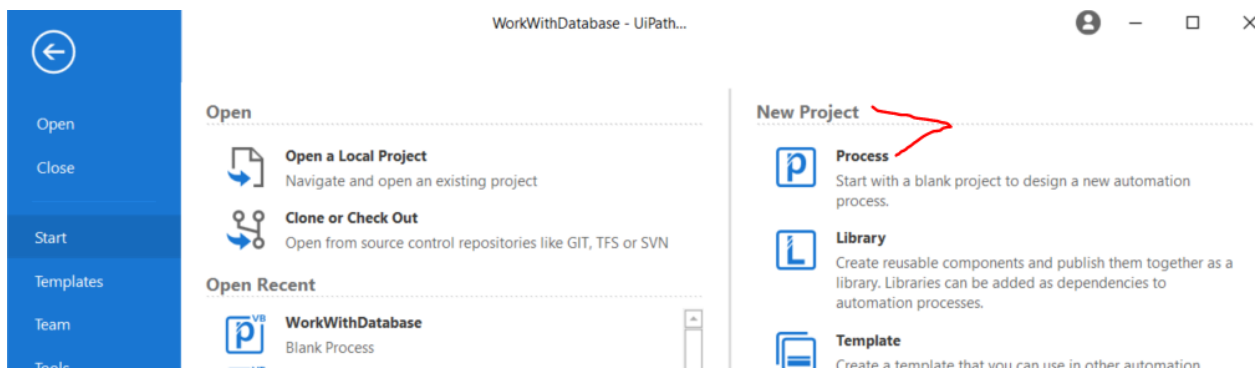
**Prerequisite:** - You need your browser extension installed for scraping data from a browser. However, if you want to scrape from Internet Explorer, you do not need to do anything.

What you will do:

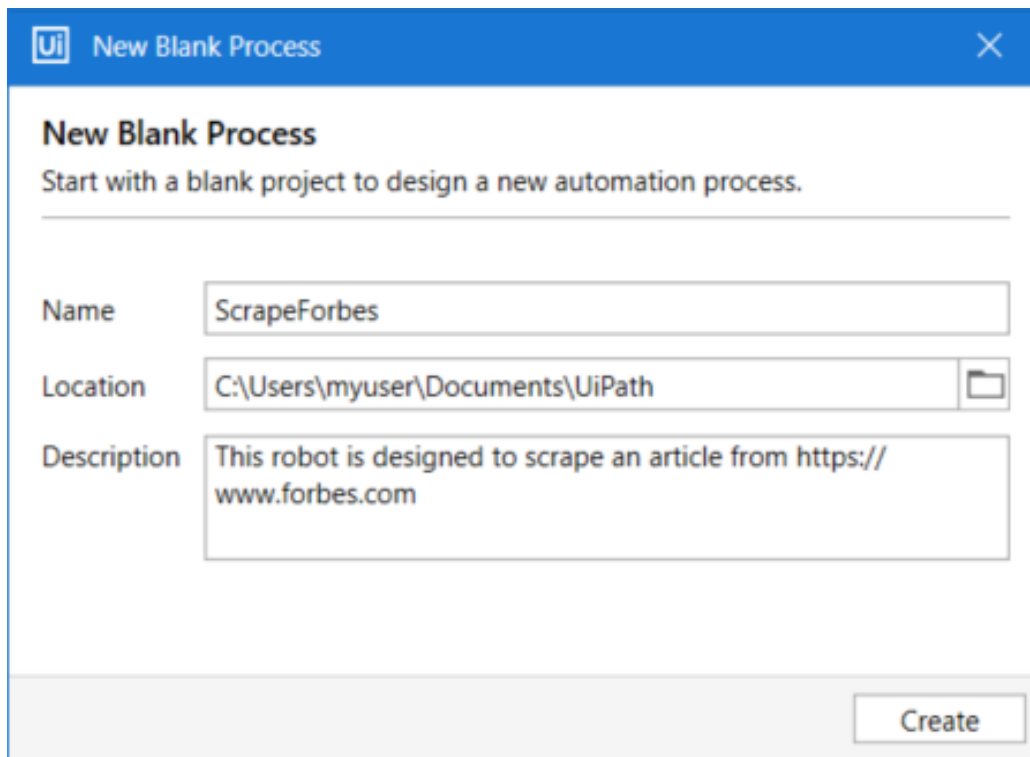
- Open a browser
- Load the article “The Four Most Important Digital Automation Trends Circa 2021” from [www.forbes.com](https://www.forbes.com/sites/sap/2021/02/16/the-four-most-important-digital-automation-trends-circa-2021/). The full address of the article is: <https://www.forbes.com/sites/sap/2021/02/16/the-four-most-important-digital-automation-trends-circa-2021/>
- Scrape the full article from this page
- Open a text editor program such as Microsoft Word, paste the data there.

## 1.1. Create a new project

- Start UiPath if you have not started it and if UiPath is already started,
- Select Process under **New Project** from the Start tab as illustrated below.

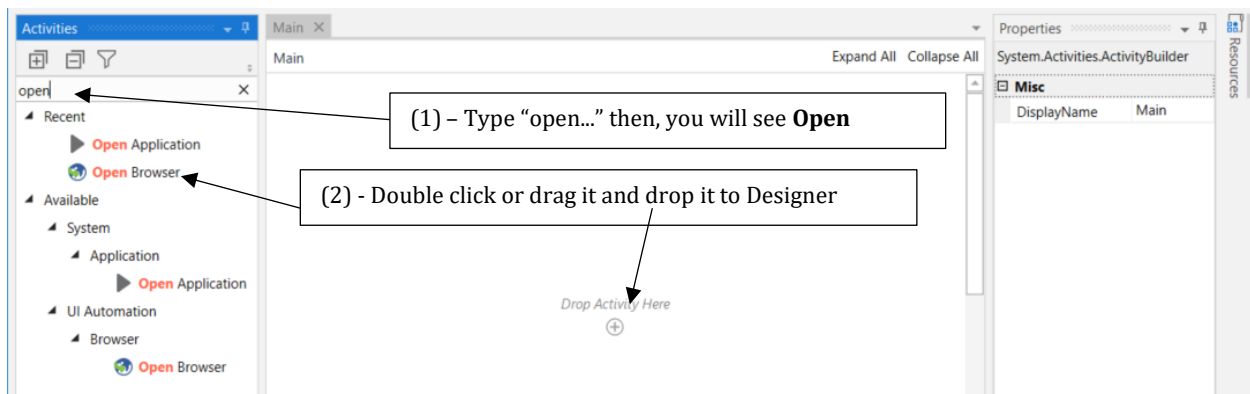


- In the **New Blank Process** dialog box, illustrated below, enter *ScrapeForbes* under **Name**.
- Specify the location of your file under **Location**.
- Specify what your robot will do under **Description**.
- Click on **Create**.



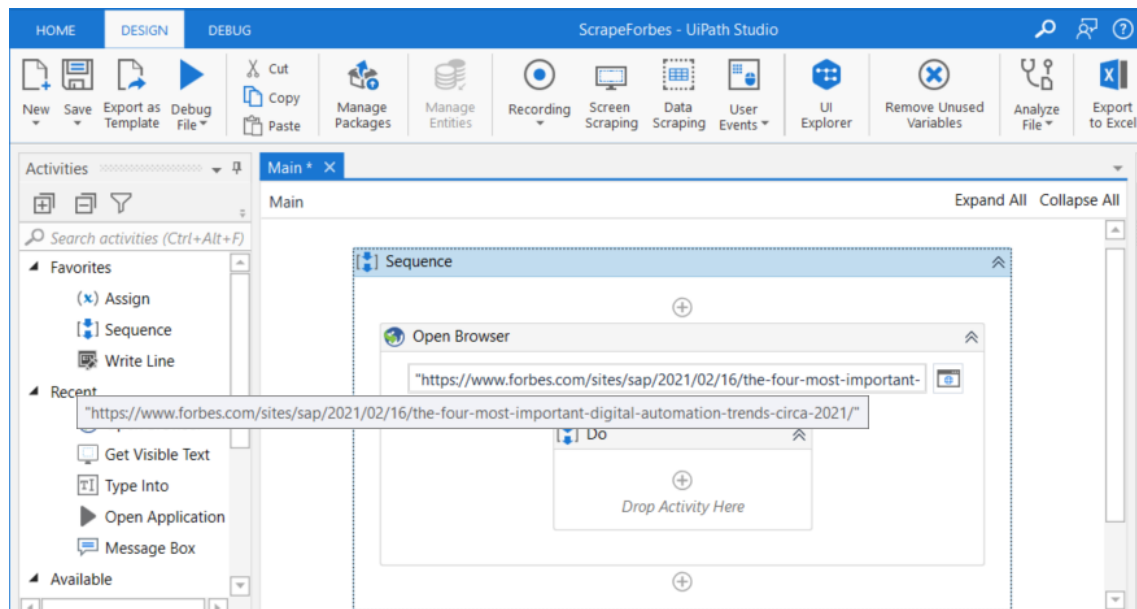
## 1.2. Add Open Browser Activity

- Click on **Open Main Workflow** from the Designer Panel
- Search for **Open Browser** under **Activities** panel using search option as illustrated below



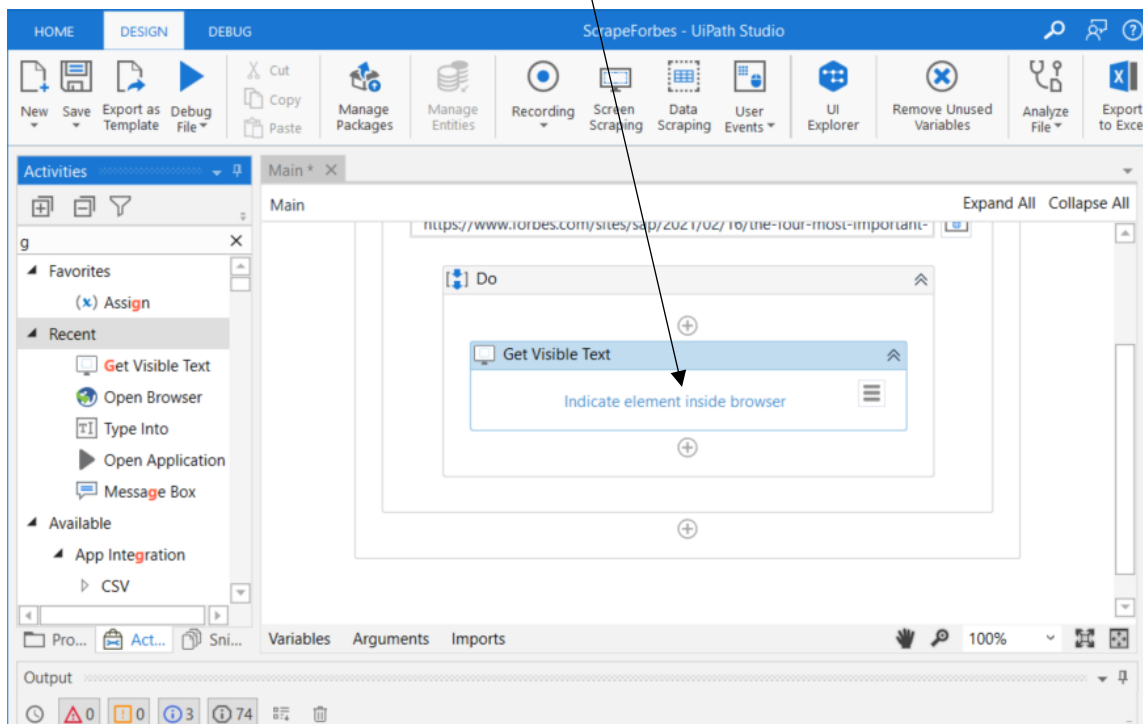
- A sequence will be added with the activity “Open Browser,” as illustrated below.
- Add the URL of the article in double quotes.  
[“https://www.forbes.com/sites/sap/2021/02/16/the-four-most-important-digital-automation-trends-circa-2021/”](https://www.forbes.com/sites/sap/2021/02/16/the-four-most-important-digital-automation-trends-circa-2021/)

**Sequence** executes a set of child activities according to a single defined order (Included in default activities)



### 1.3. Scrape Textual Data from the Web

- The next task is to scrape the visible text of the article, find **Get Visible Text** from the **Activities** panel > and drop it under **Do** Container of the Open Browser activity > the **Do** container will change allow you to **indicate element inside browser** as illustrated below.



- Now you need to open your browser and locate the article before you indicate the text that you are going to scrape.
- Click on **indicate element inside browser**.

- Hover over your pointer on the browser where the article is opened until everything that you need to scrape changes color (selected). It could be a paragraph, as illustrated below, or whatever you want to select from the website > when the selected text is highlighted as illustrated below **Click**

Feb 16, 2021, 09:00am EST

## The Four Most Important Digital Automation Trends Circa 2021

**Susan Galer** Brand Contributor  
SAP BRANDVOICE | Paid Program  
Innovation

The pandemic turned decades of accepted business principles upside down. Leaders now know everything has to be digital because it's the only way to future-proof the business. These are some of the most important predictions about digital transformation as we head into a post-pandemic world.

Gartner researchers believe that everything that can and should be automated will be automated. Everything else must be augmented. GETTY

### Pragmatism rules the day

Innovation for its own sake is so early aughts. While the pandemic is a force factor for sweeping digital transformation, make no mistake, this is about pragmatic innovation from edge to core. IDC predicted that by 2022, 70 percent of all organizations will have accelerated use of digital technologies, transforming existing business processes to drive customer engagement, employee productivity and business resiliency. Two years later, IDC predicted that 75 percent of organizations will have comprehensive digital transformation implementation roadmaps, compared to 27 percent today, resulting in true transformation across all facets of business and society. In just three years, Gartner predicted organizations will lower operational costs by 30 percent by combining hyperautomation technologies (see the next paragraph) with redesigned operational processes.

### No tech innovation is an island

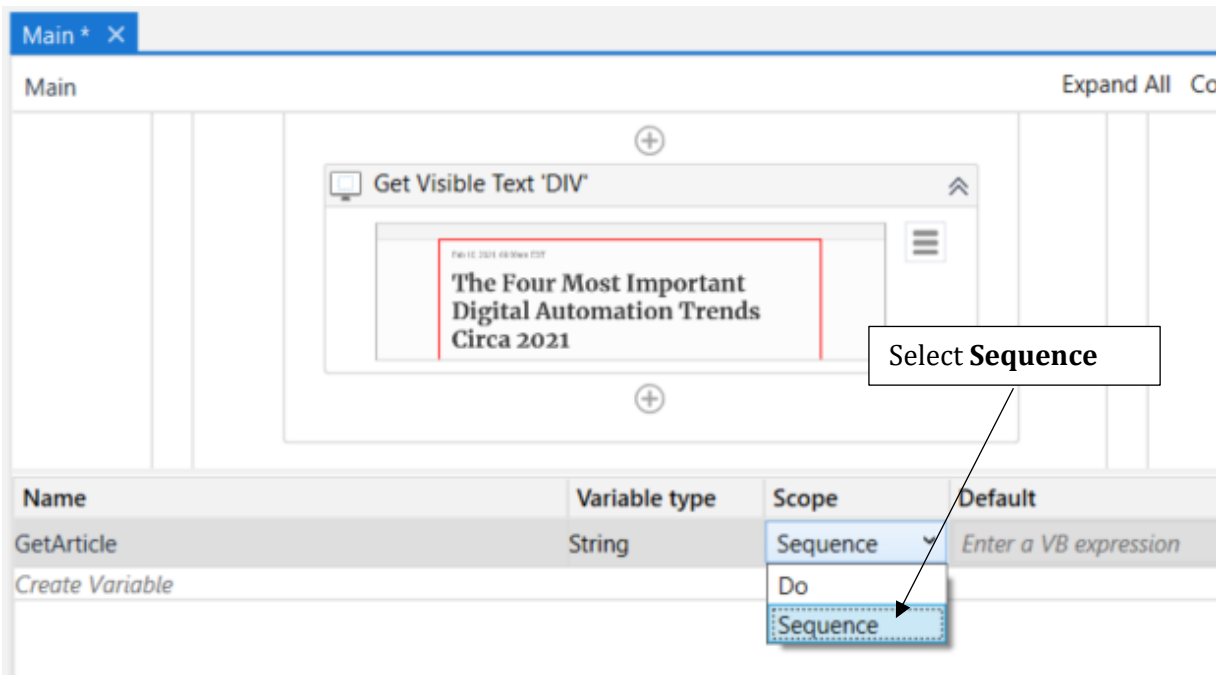
Market interest is high in AI, IoT, machine learning, digital twins, and robotic process automation (RPA). But these capabilities cannot deliver business results if they're kept in silos. Companies getting the most value are combining the right capabilities together to solve specific problems and uncover opportunities. Gartner called this hyperautomation.

In fact, Gartner researchers said that "hyperautomation is not about technology. It's a process that never ends...a disciplined approach

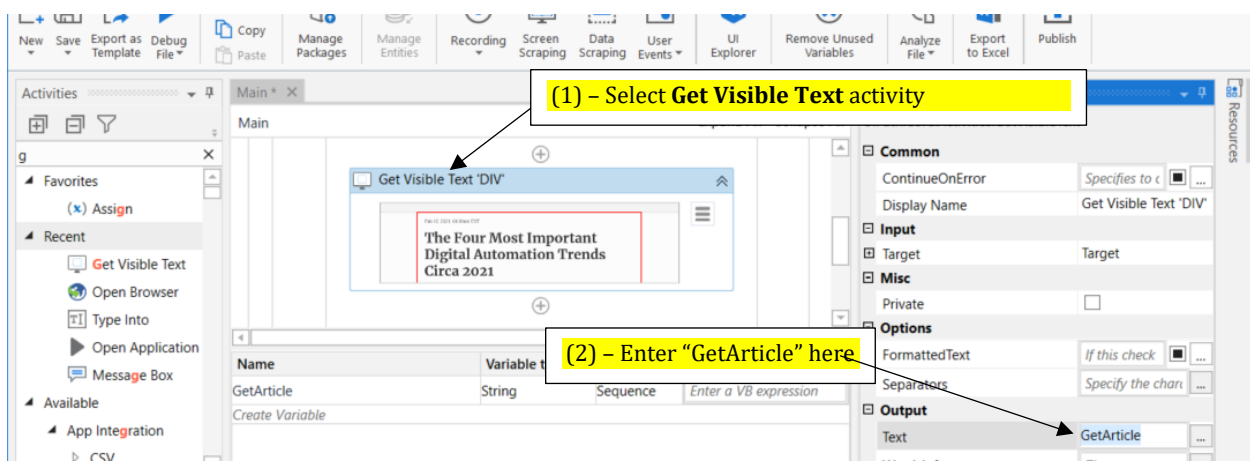
- Now you need to define a variable to hold the text data that is will be scraped from [www.forbes.com](http://www.forbes.com).

#### 1.4. Create a String Variable

- Locate the **Variable** panel and define a variable with the name **GetArticle**, as illustrated below. The scope of the variable should be changed to **Sequence** because you will use it for further activities under the bigger scope (Sequence).

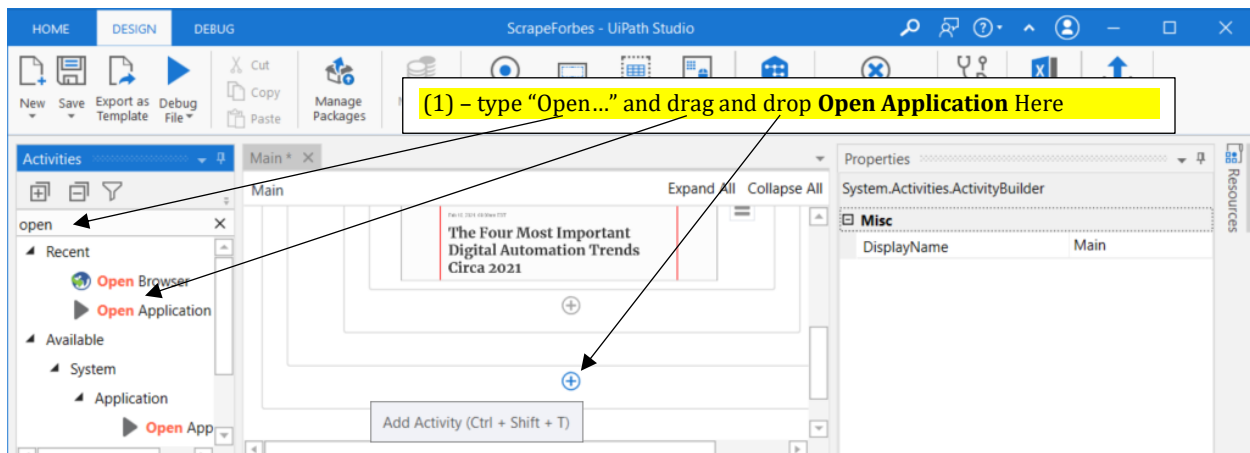


- The variable is used to hold the text scraped from [www.forebs.com](http://www.forebs.com), and you need to assign the text obtained from the **Get Visible Text** by following the steps indicated in the illustration below.

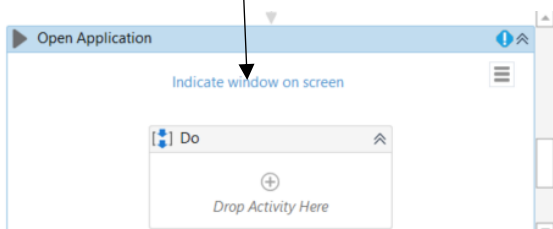


#### 1.5. Add Open Application Activity (Word)

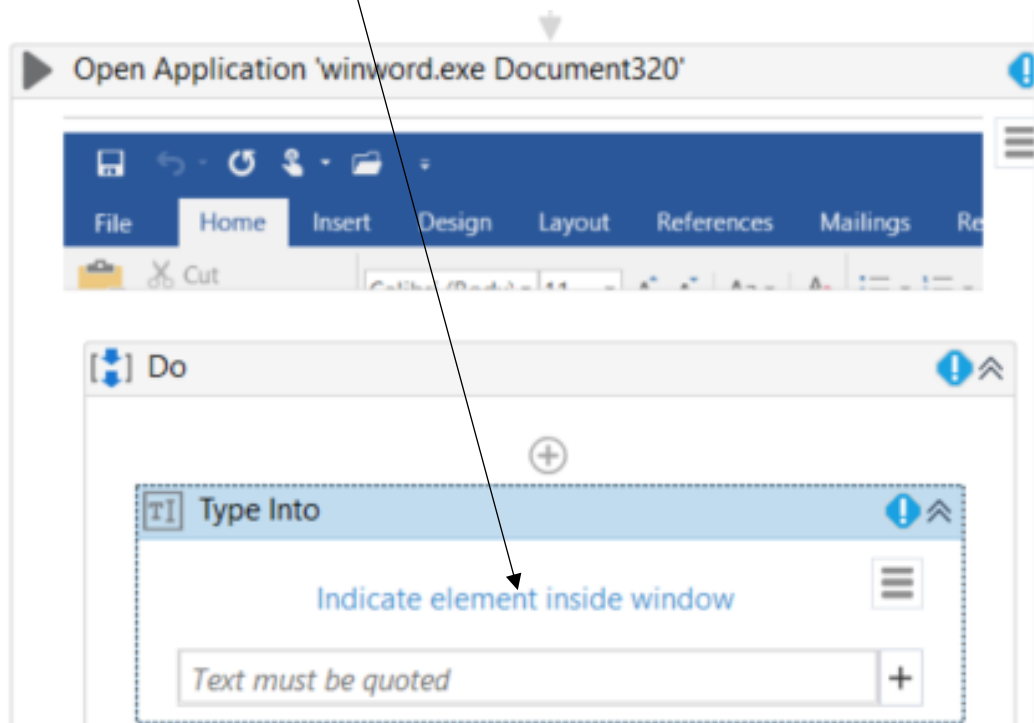
- Add **Open Application** activity under the main sequence as illustrated below.



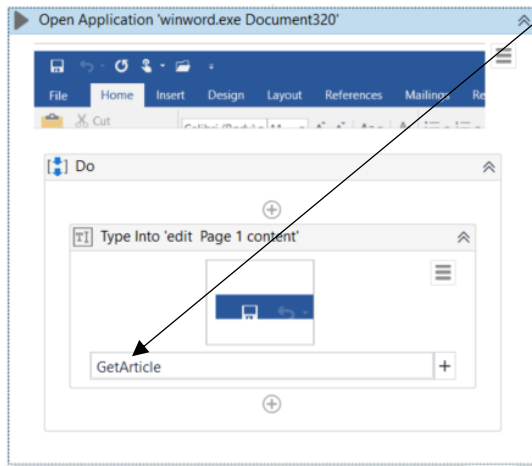
- Now open Microsoft Word from your start menu
- Select **Indicate windows on screen**.



- Select your recently opened **Microsoft Word**
- Under **Do** find and add **Type into** activity
- Select **Indicate element inside window** then indicate the insertion cursor inside the **Microsoft Word** window



- Under Type Into activity add or enter **GetArticle** variable as illustrated below



- Test the project

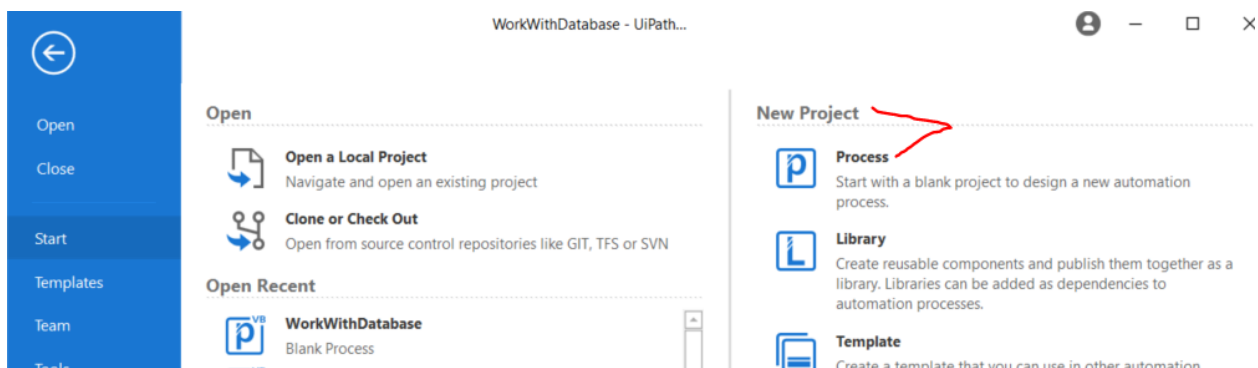
#### Note:

YOUR APPLICATION MIGHT NOT WORK NOW. IF IT DOES NOT WORK, THEN THE REASON COULD BE THAT YOUR NEWLY OPENED MICROSOFT WORD FILE NAME IS SLIGHTLY DIFFERENT FROM THE ONE YOU USED TO INDICATE THE APPLICATION, FIGURE OUT WHAT WENT WRONG, AND FIX IT!

## 2. Scrape Stock Market Data from the Web and Putting Them Into Excel File

### 2.1. Create a new UiPath Project

- Start UiPath if you have not started it and if UiPath is already started.
- Select Process under **New Project** from the Start tab as illustrated below.

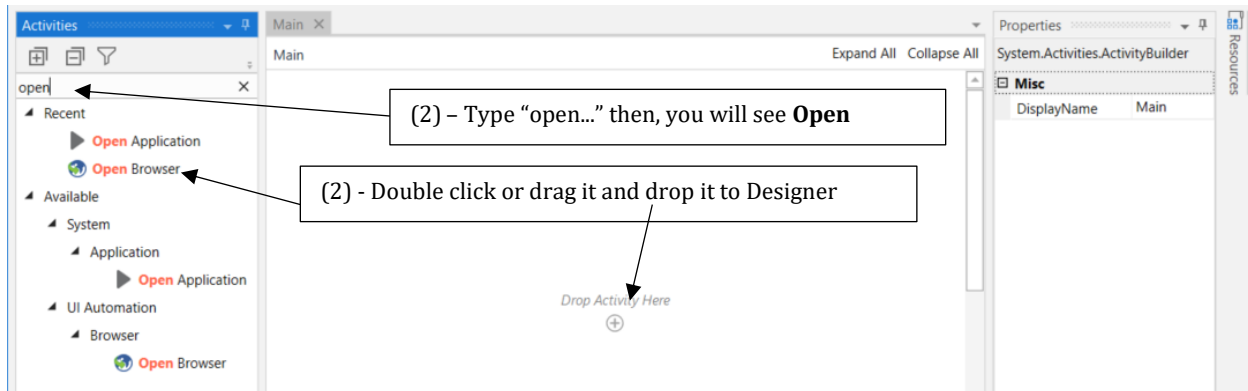


- In the **New Blank Process** dialog box, illustrated below, enter *ScrapeStockmarket* under **Name**.
- Specify the location of your file under **Location**.
- Specify what your robot will do under **Description**.
- Click on **Create**.

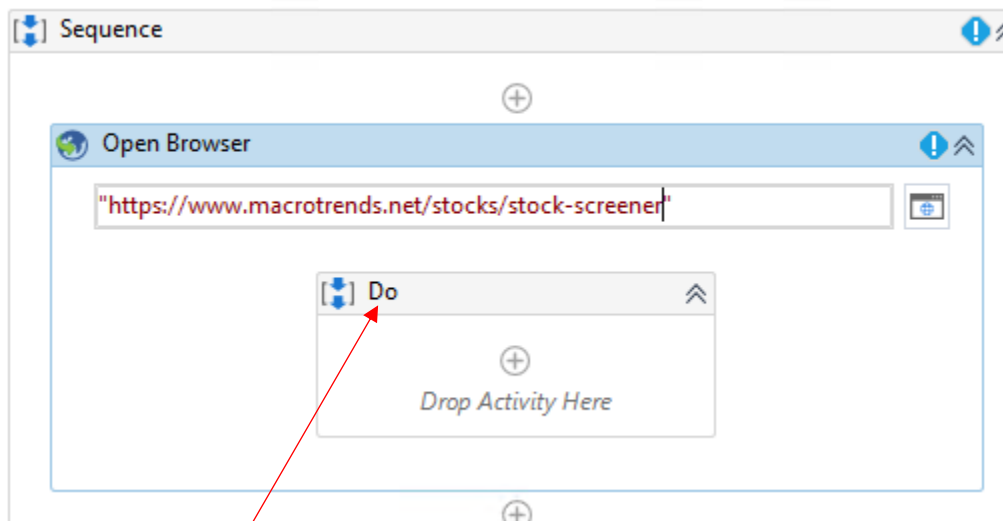


## 2.2. Add Open Browser Activity

- Click on **Open Main Workflow** from the Designer Panel
- Search for **Open Browser** under **Activities** panel using search option as illustrated below

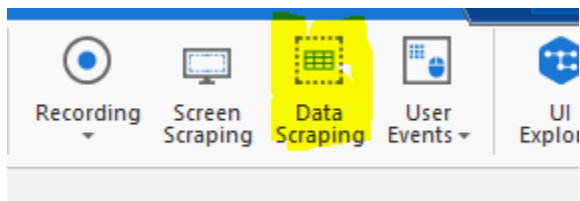


- A sequence will be added with an activity "Open Browser," as illustrated below.
- Add the URL of the stock market in double quotes.  
["https://www.macrotrends.net/stocks/stock-screener"](https://www.macrotrends.net/stocks/stock-screener) as illustrated below.

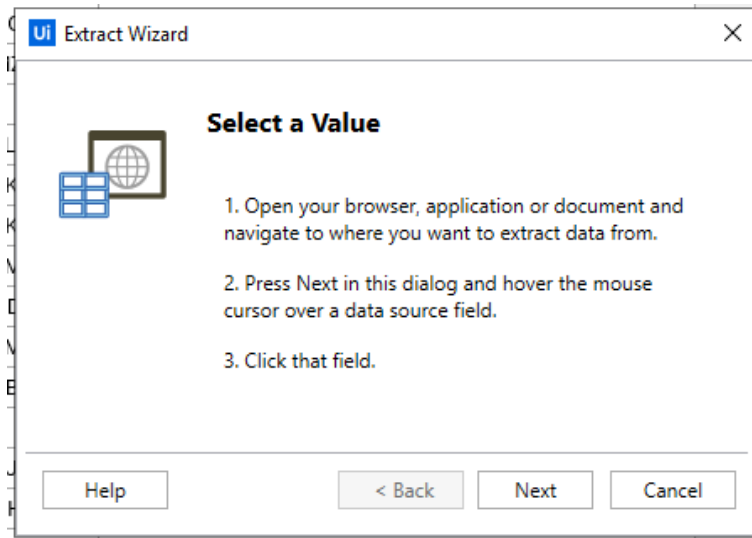


## 2.3. Data Scraping

- Select **Do**, under the **Design** panel, select **Data Scraping**.



- Now the **Extract Wizard** dialog will pop up, as illustrated below. Click on **Next**.



- Select the First Cell in the column, as illustrated below – **NOTE – the Cell should be thoroughly highlighted, which means if you select only the text, then it will not work.**

Stock Name (Hover for T1 Chart)	
Apple	AA
Microsoft	MS
Alphabet	GC
Alphabet	GC
Amazon	AM
Facebook	FB
Tesla	TS
Berkshire Hathaway	BR
Berkshire Hathaway	BR
Taiwan Semiconductor Manufacturing	TS
NVIDIA	NV
JPMorgan Chase	JP
Alibaba	BA
Visa	V
Johnson & Johnson	JN
UnitedHealth Group	UH
Walmart	WM
Bank Of America	BA
Home Depot	HD
Mastercard	MC


- Click on **Next**
- Select the second element – **NOTE – the Cell should be thoroughly highlighted, which means if you select only the text, then it will not work.**
- Specify the column name – It should be **Stock Name** and uncheck Extract URL > Click **Next**.
- Click on **Extract Correlated Data**, then do the previous two steps to specify the first and last cell of the next column.
- Continue to add columns until all columns are selected.

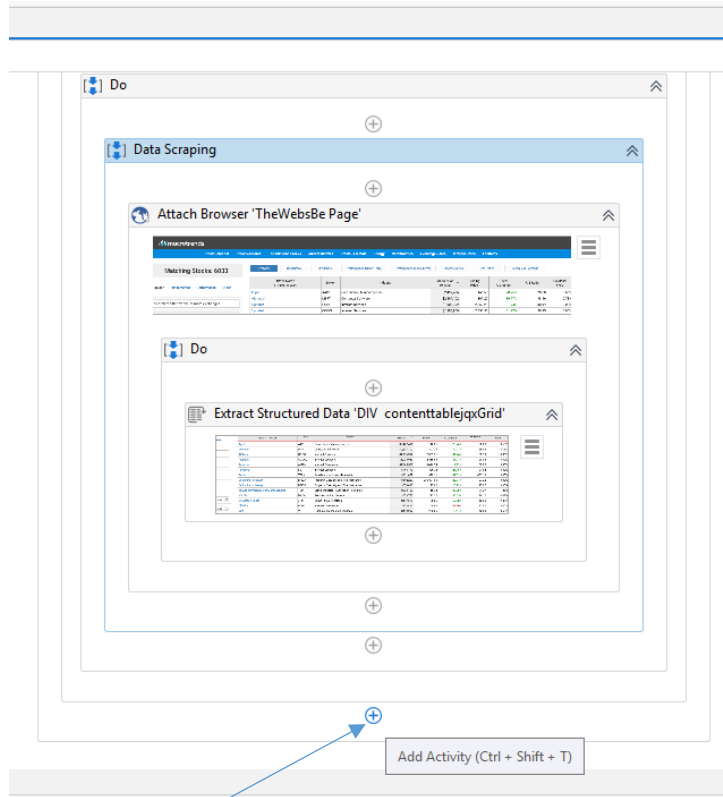
StockName	Tricker	Industry	Market Ca	Closing Pri	First % Yea	P/E Ratio	Dividend Y
Apple	AAPL	Computers -	\$2,422,496	146.55	20.76%	28.68	0.60%
Apple							
Apple							
AAPL	AAPL						
Computers -		Computers -					
\$2,422,496			\$2,422,496				
146.55				146.55			
20.76%					20.76%		
28.68						28.68	
0.60%							0.60%
Microsoft	MSFT	Computer S	\$2,307,102	307.29	39.77%	38.56	0.73%
Microsoft							
Microsoft							
MSFT	MSFT						
Computer S		Computer S					
\$2,307,102			\$2,307,102				
307.29				307.29			

100

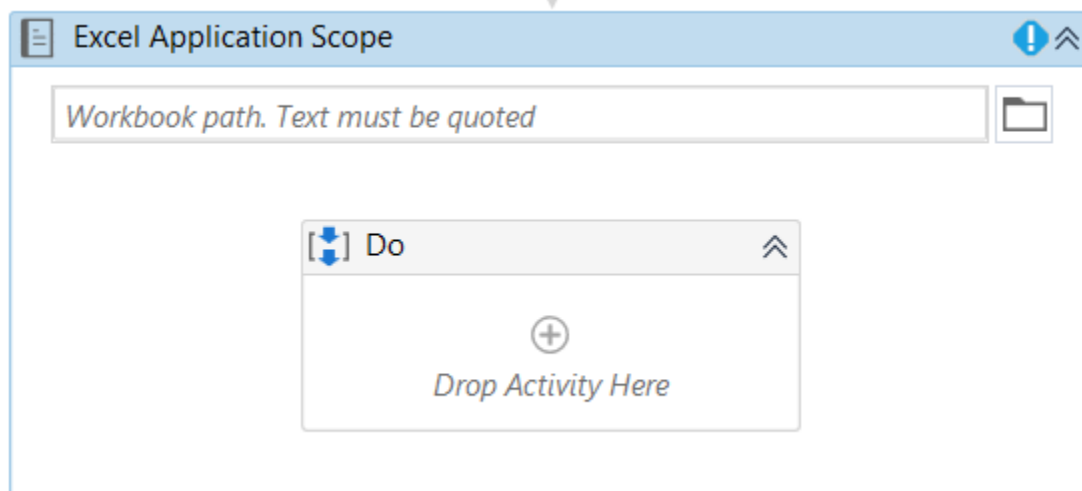
Cancel

-

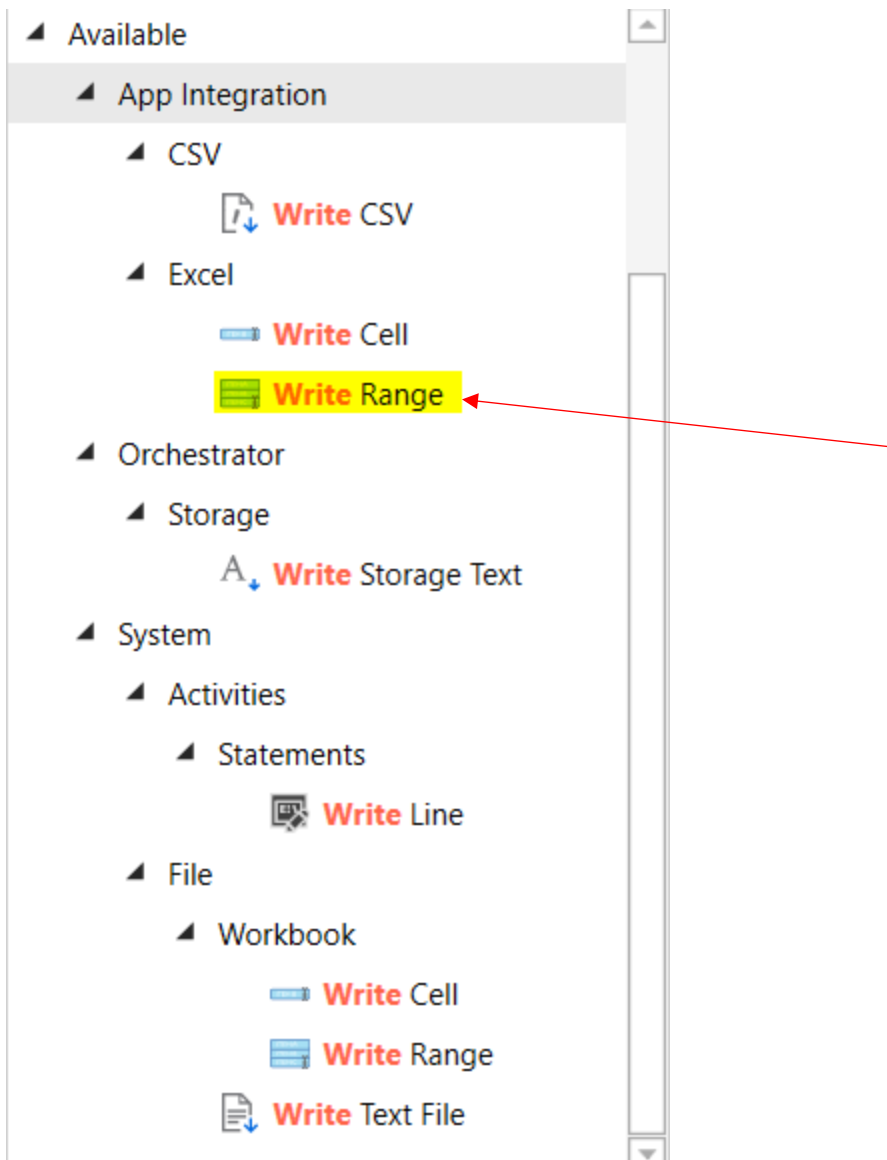
- Select **NEXT** navigator, indicated in yellow in the illustration above. 
- Now a variable under the Data Scrapping sequence is automatically created. Change the scope of this variable to the Sequence to assign it to a higher scope.

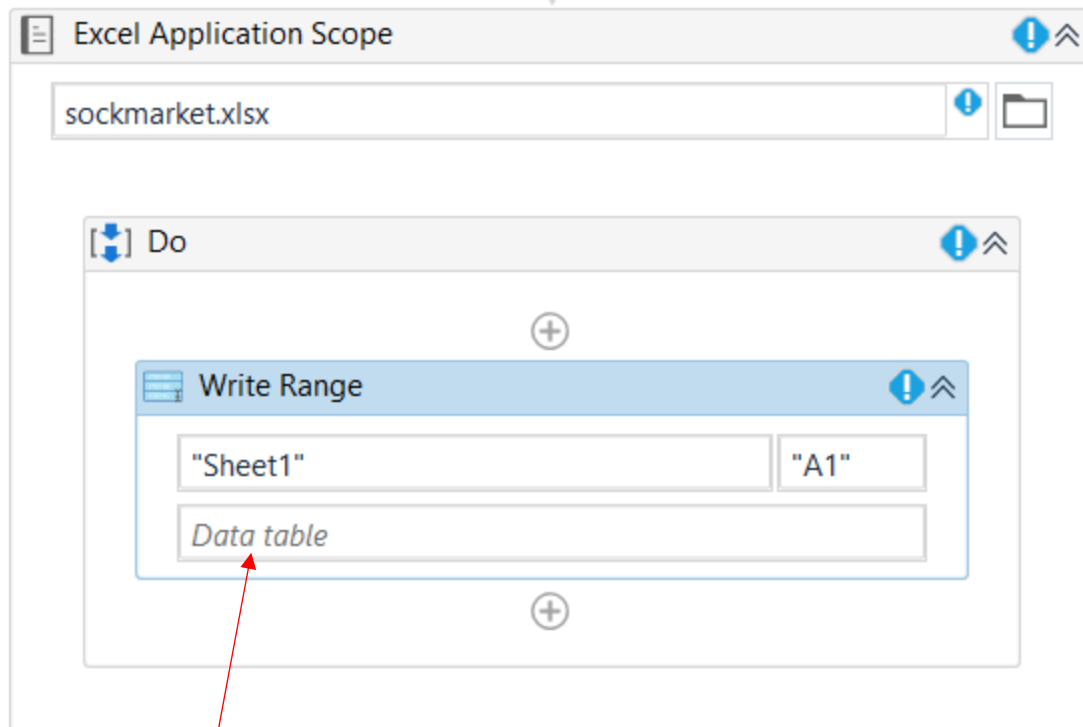


- Add Excel application scope Activity here.



- Enter - "**sockmarket.xlsx**" under Workbook path box as illustrated below.
- And Enter Write Range from **Write Range** that is found under Available > App Integration > Excel





- Add *ExtractDataTable* to **Data table**.
- Test your project.

If your project works appropriately, it will create an excel file under the folder in which your project is saved with all the data from the stock market website.