**Amazon scrapping process**

**Only Countries (UK\_CA\_EG\_UAE)**

This process is to scrap Amazon product form many Amazon websites Countries.

The process starts by reading ConfigFile which contains each country and its website and build a **dictionary from this config file where key is the country and value is the webUrl** and then in the same workflow it is reading the input which will be provided by the user and this input contains 2 rows one is the product name item and second is the countries where user want these products to be scraped and build a **dictionary where key is the product name and value is the countries required to get data from its Amazon**.

Next workflow is to open browser to the Amazon Web.

Next workflow is process items in the input dictionary (key is product name and value is one or many countries separated by comma) will make a Main loop through this dictionary and using regex will get the value into match collection of the countries of the key (productName) in the main loop and then will make nested loop through these values and there is a State Machine to handle these countries as per each value in the match collection will go it execute its State to scrap data from its website using the config dictionary value (states UK,EG,UAE,CA) and in the final state the scrapped data be written into the output datatable in

This is Fully **Unattened** Robot. User inserts the products in the input and bot could handle it.

**Data Type used**:

Strings : many strings used in the process like the get text for product name and product price this in wokflows: AmazonScrappingLogicOne and AmazonScrappingLogicTwo

Array: I used array to get all files from the output folder to send them by mail as attachments this in work flow : mainProcess

Dictionary: I used dictionary to get the config file and input file as key value pairs this in first workflow : ReadConfigAndInput

Boolean :I used Boolean in the workflow : Open browser as when element appears retryscope finish when the Boolean became true

DataTable : used to read the config file and input file and iterate through both of them this is in the workflow : ReadConfigAndInput also used in the WriteExcel workflow to write the output.

Integer: used anytime when I tried to fine-tuning the Selectors by increasing the index through each iteration and use this integer in the selector for the indexing elements this is workflow : AmazonScrappingLogicOne and AmazonScrappingLogicTWO.

**Applications Used**: Excel in ReadConfingAndInput workflow also in WriteExcel Workflow – Chrome Web in OpenBrowser workflow also in Mainprocess workflow when scarping the products

No Asset used otherwise I can convert the config (amazon country and its url) file into assets if required.

**Workflows**: In Main workflow there are below workflows:

ReadConfigAndInput

OpenAmazonWeb

mainProcess and in this workflow there is nested below workflows :

AmazonScrappingLogicOne

AmazonScrappingLogicTwo

WriteExcelFile

Not used queues because I have got all input as a dictionary key value pairs and then make a separate performer for each key element

**Data Manipulation**: In mainProcess Workflow inside the main loop for each key in dictionary I used regex to get the value in the dictionary key into Match Collection list and made a loop through this regex output.

**Dependencies**:

Microsoft.Activities.extensions

Uipath.Mail.Activities

Uipath.Excel.Activities

**To run the bot and get output mails**:

- Please change the value of mail to your mail; it exists in the config file in row "BusinessUnitEmail"

Future Modifications to the robot: improve the logic to handle many other countries except those mentioned at the beginning of this description.