Selecting the Right Candidate for Automation for Effective RPA Implementation

Himadeep Movva

movvahimadeep@gmail.com
Independent Researcher

Abstract

Selecting the right candidate for automation is often the core and first step of the automation process. This arena has some standard frameworks and best practices that must be followed. Usually, across various industries where RPA is being implemented, automation candidate selection is fraught with challenges and prone to failures, leading to the selectionof automation that isn't a quick win. In many cases, this situation compounds intomany issues, including but not limited to ineffective resource management, lackluster client partnership, and strained budgets. This literature, through a series of impactful criteria, assesses and points out what factors, which are variables or elements that are believed to influence the outcome of the study, should be considered, in conjunction with any applicable assessments, during the discovery phase of any process, providing an effective, well rounded and all-encompassing approach in selecting the right candidate for automation.

Keywords: Robotic Process Automation, Process Assessment, Unstructured data, automation candidate, semi-structured data, Citrix, rule-based, thin client, thick client, feasibility, the volume of transactions, OCR (Optical character recognition), and structured data

I. Introduction:

Many automation opportunities exist in the pipeline, but not every opportunity can be an RPA use case.When taking up automation, it is crucial to analyze various factors,including but not limitedto complexity, volume, potential for change in logic, frequent layout changes, and ROI.After carefully analyzing every factor that helps in the assessment, automation can be grouped into different quadrants: quick-win orlow effort, medium effort, and high effort. While automation that is quick wins can be immediately picked up for development, automation that requires high effortmay not be the suitable candidate with respect to ROI. Automation with medium effort can be picked up as a strategic initiative.It is recommended not to prioritize low ROIorcomplex processes.This research paper presents comprehensive steps to assess the suitability of automation, helping project managersrestructurethe automation pipeline.Through this process assessment, every factor that is important for assessing the automation feasibility, including factors related to time, complexity, dynamic UI,and image-based data, will be analyzed and scored, giving an automation score that can help decide to pick an automation candidatethat can reap quick and high gains for the team instead of picking a process that may be complex or has low ROI.

II. Detailed analysis of process:

Identifying the right candidate for automation is the core step of process automation. Various factors mustbe analyzed and assessed before deeming whether a candidateis the right fit for automation.These parameters[1]include, but are not limited to, process maturity, standardized inputs, the volume of transactions, and the complexity of the process.[2]Assessing these factors during discovery sessions helpsto categorizethe complexity of an automation candidate[3]

1. Process Maturity:

While assessing process maturity, several important factors need to be assessed. [4]Some of these factors include process stability, standardization of process across teams, documentation of Process, and measurement of KPIs.

Process Stability:

An ideal process for automation must neither have dynamic UI elements nor change in business process steps in the future. Along with frequent UI changes, if a process is more prone to varying Input/Output, unpredictable human intervention, dependency on external systems,and complex exception handling, it could be inferred that the process isn't stable enough and all the other factors including the factors mentioned above must be reevaluated carefully before marking a process as an automation candidate. One of the main reasons forthe failure of RPA projects is a [5]A process that is not suitable for automation is picked. There are no set standards specifically to rule out these candidates. However, an informed decision can be made by following and assessing the above-discussed process features.

Frequent UI Changes:

If a system is bound to change frequently at the UI level, it would be almost impossible for the Bot to run without interruption. In most cases, the code should be such that the new UI change should be reconfiguredand accounted for. But each time the UI changes, the code must be changed accordingly, inflicting strain on the maintenance and related costs. The change includes updating the selectors or reconfiguring the workflow alongside efforts towards testing. If the team is relatively new and in the process of achieving maturity in terms of RPA, this could mean that the team could hop off from the ongoing development work and work on the change on an immediate basis owing to the criticality of the operations.

Even if the team is mature, this situation could pose serious challenges in terms of maintenance. As a thumb rule, an ideal candidate for automation can be asystem whose UI doesn't get changed frequently. That kind of system can often be categorized as a legacy system, whose UI is very unlikely to change, or an enterprise application with a stable interface. Additionally, the update history of the application can be gauged to determine if it is prone to frequent UI updates, and those kinds of applications should be deprioritized.The graph below depicts, in 2-D coordinates form, how automation suitability will vary based on UI Changes. As it can be deduced, the more UI changes, the less suitable the process is for automation.

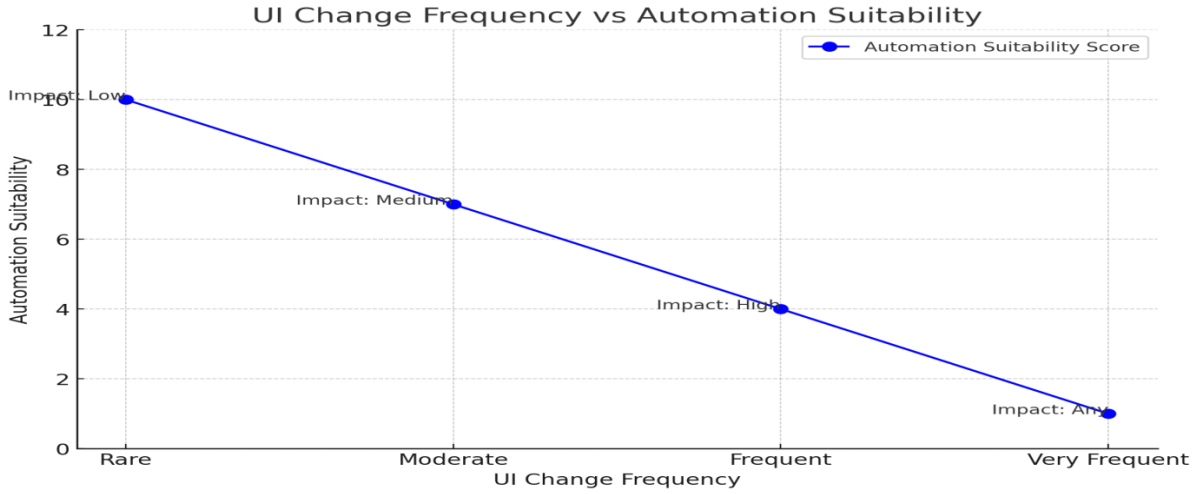

Figure 1- Change in Automation Suitability vs change in UI

Varying Input/Output:

The Input Data should be structured and in digital format.The inputmust be consistent across all the input files. Some of the input formats/types may include API calls, JSON, text files, spreadsheets, and emails. For example, column names in the input file across any type of input file or predictable patterns in input documents shouldn't vary. Overall, inputs that do not require complex interpretation or decision-making and maintain a stable data structure are preferred.

On the other hand, data that lacks a given structure aren't suitable as standardized Inputs, reducing the potential for automation to be defined as an automation candidate. Somedocuments include unstructured PDF documents, varying headers or positions of data in spreadsheets, scanned documents, or images.Also, data that requires contextual understanding isn't a suitable input for automation. For instance, some legal documents may require contextual understanding and subjective information to be processed and entered as input.Moreover, handwritten documents are also not suitable inputs for automation. Sometimes, data must be filled out based on general or domain knowledge. That kind of input is also not as effective as automation inputs.

The report should have a structured and consistent format. It can be in any digital format, including, but not limited to, Excel, text file, JSON, CSV, or email. Output cannot be unstructured or free text—this may require subjective description and interpretation. Output requiring extensive domain knowledge or expertise to comprehend the report can be avoided. Also, an output that may require human validation or oversight can be avoided.

2. Standardized Inputs:

As mentioned above, there could be various cases of what can be a standardized input. In this research article, we will present some of the very important types of those inputs and mention theircharacteristics.[6]

Spreadsheets:

Data can be effectively retrieved from spreadsheets such as Excel or CSV.However, the columns/headers of those spreadsheets are expected to remain constant.Varying columns lead to varying inputs, making the input highly unreliable for the bot. Also, even in case of variability, it should be configurable so that input data is processed accurately without any errors.

Standardized PDFs or forms:

In the case of standardized PDFs or digital forms in any format,the fieldsare expected to be consistent across all documents.If there is variability, it should be configurable initially during development so that the content can be read and processed further using OCR-based activities.

Data from Applications:

The input data can also be scraped or retrieved from different applications. In many cases, ERP tools or applications present a robust framework for obtaining data as input and using it for further processing. Reports/data can also be retrieved from different applications, including but not limited to SAP, CRM, Salesforce, desktop applications,or Citrix apps, effectively providing a case for reliable input data.

Data from Databases and cloud storage:

Also, data from on-prem or cloud databases can be retrieved and used as reliable input data.

Data from APIs:

Data can be retrieved for input by making API calls.This is one of the most effective ways to get input data. As there is no involvement of the front end, there's no need to interact with UI elements of theapplication, making the process more reliable. Not always an application will have a readily available end point for the Bot to consume data. These endpoints need to be created so that once an HTTP request is made from the Bot, the data is retrieved.

Email Reports:

The data from the email can be used as input. The only requirement is that the data should not vary. Even if it varies, the email body or subject variations should be configurable based on certain rules. Some of the use cases, if not all, where emails can be used as input are accounts receivable and payable processes, where invoices are sent as attachments, Candidate screening using ATS, IT support incident creation, order processing in e-commerce platforms, appointment scheduling,and loan application processing.

3. Volume of Transactions:

High Volume Processes:

Process with high volume, repetitive, and time-consuming factors is a good automation candidate. Usually,the process's high volume and high error rate translate to high automation ROI.Often, a high-volume process requires scalability, which is difficult for humans to provide but can easily be accomplished by a bot.Also, organizations tend to prioritize high-volume, rule-based tasks for automation due to immediate time savings and quick gains.The cost-benefit calculation also involves volume as one of the important factors in determining ROI.While assessing volume, it is equally important to evaluate the complexity, as high volume and low complexity should be picked up first for automation to receive immediate gains.[7]

Low Volume Processes:

While low-volume processes may not be an immediate gain for the organization, other factors should be carefully considered and vetted before deeming the process an automation candidate. Firstly, error rate and complexity must be assessed. When a process is more prone to errors due to various factors, including requiring compliance and thorough validation, and is complex, it can be considered a crucial factor to determine whether the automation candidate can be pushed to a current pipeline of opportunities. Even if a process is of low volume, sometimes the process can be so complex that it may require many hours of dedicated work to finish the process. RPA Bot can free up those manual hours, making it more reliable and less prone to errors while saving much manual effort. Scalability must also be considered in the case of processes withlow volume. If a process is scalable, it can be developed for other functions within or across the department that does similar functions.

Even if the current volume is low, if the process can be scaled and standardized across various functions, it can still give good benefits after implementation. Moreover, in the case of strict service level agreements, automation can reduce turnaround time. It can meet the deadline instead of assigning multiple employees to complete the task on time and meet the deadline.Thirdly, automation can be beneficial when dependent on other processes. For example,when the output of the current automation is used by a current bot as input or when the data prepared by the current process is used by the current bot in production, automating the process can free up the dependency and streamline the operations, potentially eliminating any manual intervention.

4. Process Complexity (too many logics/decisions):

Process complexity is crucial in determining whether a candidate is a good fit for automation. This process may require multiple decision points, various paths in case of exceptions, and domain knowledge for human judgment, which, at some points, might be challenging to automate in terms of development time and costs.Processes with high variability and frequent exceptions increase the complexity of automation.Also, complexity increases if the process is one that has a very high number of decision points or branches. In addition, processes that are bound to change frequently may require extensive time towards change management, increasing costs. [8]

III. Process Assessment Tool

From UiPath Academy, a process assessment sheet is taken and used for analysis in this paper. In this section, detailed analysis and assessment of each factor in determining whether an automation is a candidate for development are analyzed, in a detailed way, helping to consider the factors in the process assessment phase. These factors are grouped into feasibility, suitability, process details, and input data.After we fill in the required information, the output sections- estimated benefit and estimated implementation effort– are auto-populated based on the underlying formulae.Those factors, according to each section, are discussed below.

Feasibility:

Decision type is defined by the answer to the question: Are the decisions rule-based or subjective? The options under this question would be exclusively rule-based, mostly rule-based, with Some level of subjectivity, mostly subjective, and exclusively subjective. Based on the selection, the output score will be calculated as feasible or low feasibility. Any steps that are categorized as mostly subjective or exclusively subjective will be categorized as low feasibility. The following calculation is used to score the feasibility. VLOOKUP function uses the table below, available in the scorecard sheet from D7 to E11, to score the feasibility.

=IF (ISERROR (VLOOKUP($F7, Scorecard!$D$7:$E$11,2,0)),"",VLOOKUP($F7,Scorecard!$D$7:$E$11,2,0))

Rule Based vs Ad-hoc/Judgmental	
Possible Answers	**Scoring**
Exclusively Rule Based	Feasible
Mostly Rule Based	Feasible
Some level of subjectivity involved	Feasible
Mostly subjective	Low Feasibility
Exclusively Subjective	Low Feasibility

Figure 2- Scoring based on how rule-based a process is

Input Type should be selected based on the question of what most of the data input looks like. The answer options would be non-digital and unstructured, non-digital and structured, digital and unstructured, and digital and structured. Based on the formula below and the table mentioned in Figure 2 and Figure 3- Scoring based on the Type Of Input, the result will be categorized as feasible or low feasibility.After selecting the type of input that was shown according to a range of values from D16 to D19, VLOOKUP will calculate the scoring based on the values from E16 to E19 correspondingly. All in all, VLOOKUP looks up the value in cell H7 within the range D16:D19 on the scorecard sheet, and if a match is found, it returns the corresponding value from Column E from the range E16:E19.

=IF (ISERROR (VLOOKUP ($H7,Scorecard!
$D$16:$E$19,2,0)),"",VLOOKUP($H7,Scorecard!$D$16:$E$19,2,0))

Type of Input	
Possible Answers	**Scoring**
Not digital and Unstructured	Low Feasibility
Non-Digital and Structured	Digitize first
Digital and Unstructured	Feasible
Digital and Structured	Feasible

Figure 3- Scoring based on the Type Of Input

Process Stability is determined by the question of how frequently the process will change in the next 6 months. The answers would be no change expected, very small change, some change, medium change, and significant change. The score will be calculated based on the below formula

=IF (ISERROR (VLOOKUP ($J7, Scorecard! $D$26: $E$30,2,0)),"“, VLOOKUP ($J7, Scorecard!
$D$26:$E$30,2,0))

The calculation is performed and scored according to the table below in Figure 3.If an error is encountered while calculating, the result will be an empty string, but if a match is found based on the answers from D26:D30, the score column will be updated based on the values from E20:E30.

Process Stability	
Possible Answers	**Scoring**
No change expected	0
Very Small Change	0.2
Some change	0.4
Medium Change	0.8
Significant Change	Postpone

Figure 4- Scoring based on Process Stability

Application suitability is assessed by the question how will your application change in the next 6 months? The possible answers would be no change expected, very small change, some change, medium change, and significant change.Based on the formula below, the score will be calculated

=IF (ISERROR (VLOOKUP ($L7,Scorecard! $D$35: $E$39,2,0)),"“,
VLOOKUP($L7,Scorecard!$D$35:$E$39,2,0))

The same calculation that was used for process stability is used for application suitability.

Suitability:

There are three categories of questions. In category 1, the questions are as follows.

What is the frequency of the process? The answer options will be daily, weekly, bi-weekly, monthly, quarterly,orannually.

What is the volume of transactions? The answer will be in the number of transactions. Ex: 100,900 etc.

What is the average time taken to complete the process (Average Handling Time of one transaction in minutes)? The answer will be in minutes. Ex: 90, 220 etc

Using the answers to the above questions, the Equivalent FTEvalue is calculated using the formula below

=IF (ISERROR (VLOOKUP ($P7,Scorecard! $D$46:
$F$51,3,0)*Q7*R7/8/60/260),"",VLOOKUP($P7,Scorecard!$D$46:$F$51,3,0)*Q7*R7/8/60/260)

The calculation is used to arrive at the calculated value of Equivalent FTEs is 1 FTE = 260 days

Frequency of the Process		
Possible Answers	Scoring	
Daily		260
Weekly		52
Biweekly		26
Monthly		12
Quarterly	These are control questions that are not scored. The frequency x AHT x Volume/frequency will be used to estimate the	4
Annually	number of FTEs	1

Figure 5- Frequency of the process

Figure 5- Frequency of the process

By multiplying the Average Handling Time of one transaction by the number of transactions, the number of total human hours required to complete the process will b achieved. After offsetting holiday hours, an employee's annual work would be 2080 hours(40 hours per week*52 weeks).

$$\text{Equivalent FTE} = \frac{\text{Total Transactions* Average handling time of one Transaction}}{\text{Annual work hour per FTE}}$$

What is the average number of human errors? The answer will be in percentage ex: 20%

How would you categorize the peak of the process? The answer options are Regular lasting several days or weeks; rare but predictable lasting several days or weeks in a row; rare and hard to predict; andthe process doesn't have any peaks

Based on the answer, the score will be calculated based on the below formula. The calculation and result will be based on the table in Figure 5. Based on the option selected, scoring will be assigned using the formula.

=IF(ISERROR(VLOOKUP ($U7,Scorecard!
$D$56:$E$59,2,0)),"",VLOOKUP($U7,Scorecard!$D$56:$E$59,2,0))

Process Peaks	
Possible Answers	**Scoring**
Regular (e.g. month closing), lasting for several days or weeks in a row and increasing the utilization of the team capacity by > 20%	1
Rare but predictable event (e.g. winter holidays/ yearly closing), lasting for the several days or weeks in a row and increasing the utilization of the team capacity by > 20%	2
Rare event, hard to predict, of short duration	3
The process does not have peaks	N/A

Figure 6- Scoring based on process peaks

Process Details:

1. The following are the questions in the section on process details
2. How many steps does the process have? The options are <=10,10-15,15-25,25-40 and >40.

Using the formula here

=IF(ISERROR(VLOOKUP($W7,Scorecard!$D$66:$E$70,2,0)),"",VLOOKUP($W7,Scorecard!$D$66:$E$70,2,0)) , which in turn uses the table below in figure6 for scoring, scoring will be assigned. For ex: a process with <=10 steps will have scoring of 0.1

Number of Steps - complexity proxy 1	
Possible Answers	**Scoring**
<=10 steps	0.1
10-15 steps	0.2
15-25 steps	0.4
25-40 steps	0.6
>40 steps	1

Figure 7- Scoring based on the number of steps in a process

3. How difficult are the decisions to complete the process?

The answer options are that the process is linear and involves simple and complex decisions. Using the formula below and the table in Figure 7, the scoring will be Made

=IF(ISERROR(VLOOKUP($Y7,Scorecard!$D$73:$E$75,2,0)),"",VLOOKUP($Y7,Scorecard!$D$73:$E$75,2,0))

Difficulty of decisions	
Possible Answers	**Scoring**
The process is linear - there are no decisions to be taken	0.1
The process involves simple decisions (yes/no type)	0.2
The process involves complex decisions	0.7

Figure 8- Scoring based on the linearity of a process

4. What is the average number of exceptions where you cannot complete the entire process? Ex: 5%
5. What is the number of applications used for the process? The answer options are 1 application, 2-3 applications, 4-5 applications, and >5 applications.

A score is calculated using the below formula based on the responses to questions 4 and 5. Also, the table in Figure 8is used for scoring

=IF(ISERROR(VLOOKUP($AB7,Scorecard!$D$80:$E$83,2,0)),"",VLOOKUP($AB7,Scorecard!$D$80:$E$83,2,0))

Number of Applications	
Possible Answers	**Scoring**
1 application	0.1
2-3 applications	0.3
4-5 applications	0.6
> 5 applications	1

Figure 9- scoring based on the number of applications involved

6. Any application uses Citrix/VDI? The answer would be Yes/No

Using the formula below and the table in Figure 9, scoring is calculated

=IF(ISERROR(VLOOKUP($AD7,Scorecard!$D$88:$E$89,2,0)),"",VLOOKUP($AD7,Scorecard!$D$88:$E$89,2,0))

Thin Client?	
Possible Answers	**Scoring**
Yes	1.6
No	1

Figure 10- Scoring based on Thin client or Thick client application

Input Data:

The following are the questions asked in this section

1. What % of your input is digital? Ex: 100%
2. Is any of your input scanned? Yes/No

Based on the answers to these two questions and based on the formula below and table in Figure 10, scoring is calculated

=IF(ISERROR(VLOOKUP($AG7,Scorecard!$D$94:$E$95,2,0)),"",VLOOKUP($AG7,Scorecard!$D$94:$E$95,2,0))

OCR?	
Possible Answers	**Scoring**
Yes	1.2
No	1

Figure 11- Scoring based on the need of OCR

3. What percentage of your input is structured? The answer options will be >=80%, 60-80%,40-60%, and <40%

Based on the option selected and based on the formula below and information in the table below in figure 11, scoring will be calculated

=IF(ISERROR(VLOOKUP($AI7,Scorecard!$D$100:$E$104,2,0)),"",VLOOKUP($AI7,Scorecard!$D$100:$E$104,2,0))

Structured Data vs Un-Structured	
Possible Answers	**Scoring**
>= 80%	0
60%-80%	0.4
40%-60%	0.7
<40%	1

Figure 12- Scoring based on the structure of the data

Based on the scores calculated until now, the final score %Automation will be calculated using the below formula

=IF(ISERROR(IF($AF7*(1-$AA7-50%*$AJ7-10%*$Z7)<0%,0%,$AF7*(1-$AA7-50%*$AJ7-10%*$Z7))),"",IF($AF7*(1-$AA7-50%*$AJ7-10%*$Z7)<0%,0%,$AF7*(1-$AA7-50%*$AJ7-10%*$Z7)))

IV. Conclusion

Hence, selecting the right automation candidate to prioritize the pipeline of automation opportunities involves a comprehensive analysis of the process and of various factors that impact the process metrics. RPA software vendors, with their own tools, provide the most advanced indicator systems. These are based on a step-by-step subjective interviewing procedure of the client to identify the complexity of each process and to use a simple classification algorithm based on existing experiences to determine whether it is suitable to be automated using RPA [9]Organizations can quantify and qualify the automation potential by leveraging process assessment tools.Based on the input parameters, the output will be any of the four quadrants: quick win, low-hanging fruit, must-doimprovement, or long-term improvement. Processes with high transaction volumes, standardized inputs, low to moderate complexity processes, and a mature process that is stable would be a typical automation candidate.Considering these attributes during assessment and evaluation, a cost-effective, scalable solution with good ROI can be attained.

References:

[1] N. ostdick, "UiPath Blog," [Online]. Available: https://www.uipath.com/blog/rpa/5-factors-in-choosing-which-processes-to-automate[Accessed Jul,2021].

[2] K. Casey, "Enterprises Project," 4 June 2019. [Online]. Available: https://enterprisersproject.com/article/2019/6/rpa-robotic-process-automation-find-use-cases [Accessed Aug,2021].

[3] M. Desiderio, "How To Know When RPA Is Right For Your Organization," 15 April 2019. [Online]. Available: https://www.huronconsultinggroup.com/insights/robotic-process-automation[Accessed Aug,2021].

[4] P. Stuardo, "https://www.linkedin.com/pulse/how-evaluate-which-processes-automate-rpa-pablo-stuardo-pmp," 10 March 2019. [Online]. Available: https://www.linkedin.com/pulse/how-evaluate-which-processes-automate-rpa-pablo-stuardo-pmp[Accessed Aug,2021].

[5] Goris, "An Assessment of Process Discovery Techniques With The Purpose," p. 68, 2019[Accessed Aug,2021].

[6] "Great Tips for Evaluating Whether a Process is Suitable for Robotic Process Automation (RPA)," April 2017. [Online]. Available: https://www.cfb-bots.com/single-post/2017/04/19/great-tips-for-evaluating-whether-a-process-is-suitable-for-robotic-process-automation[Accessed Jul,2021].

[7] Y. Riedl and K. R. Beetz, "Robotic Process Automation: Selection of Automatable Business Processes," p. 10, 2019[Accessed Jul,2021].

[8] J. Geyer-Klingeberg, J. Nakladal, F. Baldauf and F. Veit, "Process Mining and Robotic Process Automation:," CEUR, p. 8, 2018[Accessed Jul,2021].

[9] Wanner et al., "Process Selection in RPA Projects - Towards a Quantifiable Method of Decision Making," 2019[Accessed Aug,2021].