**Overview**

Our mission is to enable all people to do the best work of their lives—the first act in achieving that mission is to help companies automate tedious but critical business processes. This RPA challenge should showcase your ability to build a bot for purposes of process automation.

**🟢 The Challenge**

Your challenge is to automate the process of extracting data from a news site.

You should push your code to a **public** Github repo, and then use that repo to [create a Robocorp Control Room process](https://robocorp.com/docs/courses/beginners-course-python/12-running-in-robocorp-cloud). The process should have a completed successful run before submission. Make sure to [write your files to the /output directory](https://robocorp.com/docs/courses/beginners-course-python/9-collecting-the-results#saving-the-file-to-the-output-directory) so that they are visible in the artifacts list.

**The Source**

You are free to choose from any general news website, feel free to select from one of the following examples.

* <https://apnews.com/>
* <https://www.aljazeera.com/>
* <https://www.reuters.com/>
* <https://gothamist.com/>
* <https://www.latimes.com/>
* <https://news.yahoo.com/>

**Parameters**

The process must process three parameters via the robocluod work item

* search phrase
* news category/section/topic
* number of months for which you need to receive news (if applicable)

Example of how this should work: 0 or 1 - only the current month, 2 - current and previous month, 3 - current and two previous months, and so on

These may be defined within a configuration file, but we’d prefer they be provided via a [Robocloud workitem](https://rpaframework.org/libraries/robocorp_workitems/)

**The Process**

The main steps:

1. Open the site by following the link
2. Enter a phrase in the search field
3. On the result page

If possible select a news category or section from the

Choose the latest (i.e., newest) news

1. Get the values: title, date, and description.
2. Store in an Excel file:
   * title
   * date
   * description (if available)
   * picture filename
   * count of search phrases in the title and description
   * True or False, depending on whether the title or description contains any amount of money

Possible formats: $11.1 | $111,111.11 | 11 dollars | 11 USD

1. Download the news picture and specify the file name in the Excel file
2. Follow steps 4-6 for all news that falls within the required time period

Specifically, we will be looking for the following in your submission:

1. Quality code Your code is clean, maintainable, and well-architected. The use of an object-oriented model is preferred.

We would advise you ensure your work is [PEP8 compliant](https://peps.python.org/pep-0008/)

Employ [OOP](https://peps.python.org/pep-0008/)

1. Resiliency Your architecture is fault-tolerant and can handle failures both at the application level and website level.

Such as using [explicit waits](https://selenium-python.readthedocs.io/waits.html) even when using the [robocorp wrapper browser for selenium](https://rpaframework.org/libraries/browser_selenium/python.html)

1. Best practices Your implementation follows best RPA practices.

Use proper [logging](https://docs.python.org/3/library/logging.html) or a suitable third party library

Use appropriate [string formatting](https://www.digitalocean.com/community/tutorials/python-string-concatenation) in your logs (note we use python 3.8+)