Lufthansa Industry Solutions

AI Automation Task

Date: 21.07.2025

Lufthansa Industry Solutions

Table of Content

1 Introduction

This exercise focuses on building a basic automation pipeline, by creating a web scraper, that saves information about each book from a fake online bookstore. (url: https://books.toscrape.com/), into a CSV file. The scraper runs every 5 minutes and collects as much information as possible from each book product and adds the information into the CSV file. Ideally the program creates a new sheet for each product page on the website.

2 Data Sets

The following URL contains the necessary data: https://books.toscrape.com/. Once you open it on a browser, there are paginations with products per page. You will be using the paginations in order to save each batch of books in the CSV file.

3 Requirements

You will need to create a script that utilizes the modules listed on the “Necessary Python Packages” part in order to create an automation script, focused on web scraping with the following general requirements.

3.1 Scraping Logic

- Able to scrape the following information:

  I) Image URL (can be created by combining base URL + image source)
  II) Rating (1, 2, 3, 4, or 5 out of 5)
  III) Title
  IV) Price (only the number, not the currency)

3.2 CSV saving logic

- Able to save the scraped information on a CSV file.

  I) Excel sheet for each product page.

  II) Each Sheet should have a table with columns where each column corresponds to the

4 Necessary Python Packages

- BeautifulSoup4, pandas, requests, openpyxl

5 Deliverables

- Upload your program package on your GitHub profile, by creating a public repository and sharing the link.