MYSCRAPPER

Python Screen Scrapping Project using Playwright

Automating Data Extraction from Websites with GUI Interface

INTRODUCTION

- This project demonstrates automated web data extraction using Python and the Playwright library.
- A GUI which allows users to interact with multiple scraping modes.
- Data is extracted from public websites and saved locally.
- Focuses on both open-access and login-required scraping workflows.

TECHNOLOGIES USED

- Python 3.11+
- Playwright – for browser automation (Chromium/Microsoft Edge)
- Tkinter – to display data in text box
- OpenPyXL – for saving data to Excel (optional)
- ConfigParser – for securely handling login credentials
- Webbrowser module – for opening scraped data in Edge
- HTML, CSS - to build the GUI

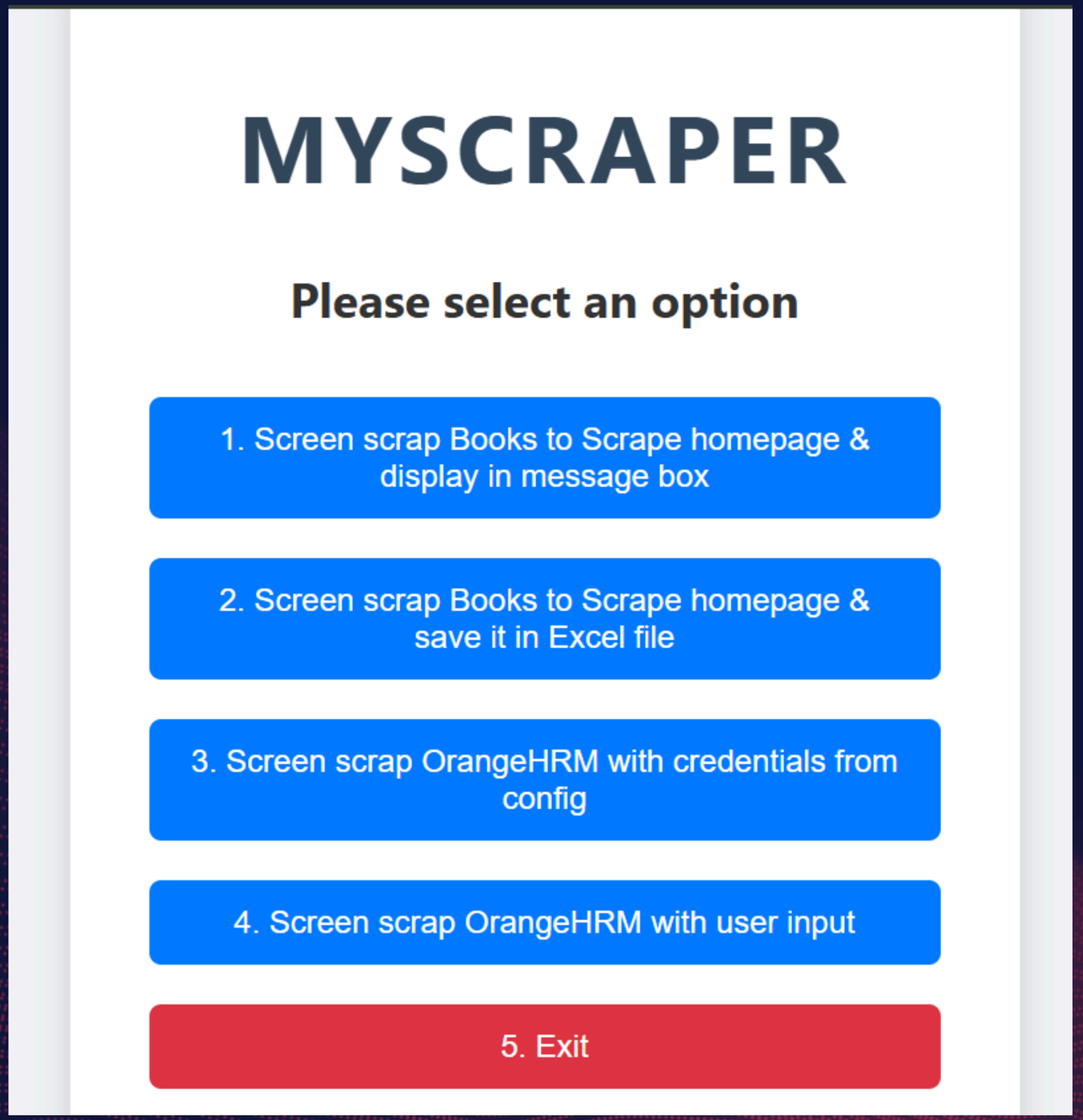
Problem Statement

- Manual data collection from websites is time-consuming, repetitive and error-prone.
- Many websites require user login or have dynamic JavaScript content, making traditional scraping ineffective.
- Need a reusable tool that:
  - Handles different scraping scenarios.
  - Supports user input and secure logins.
  - Provides output in a user-friendly format.

Proposed OBJECTIVES

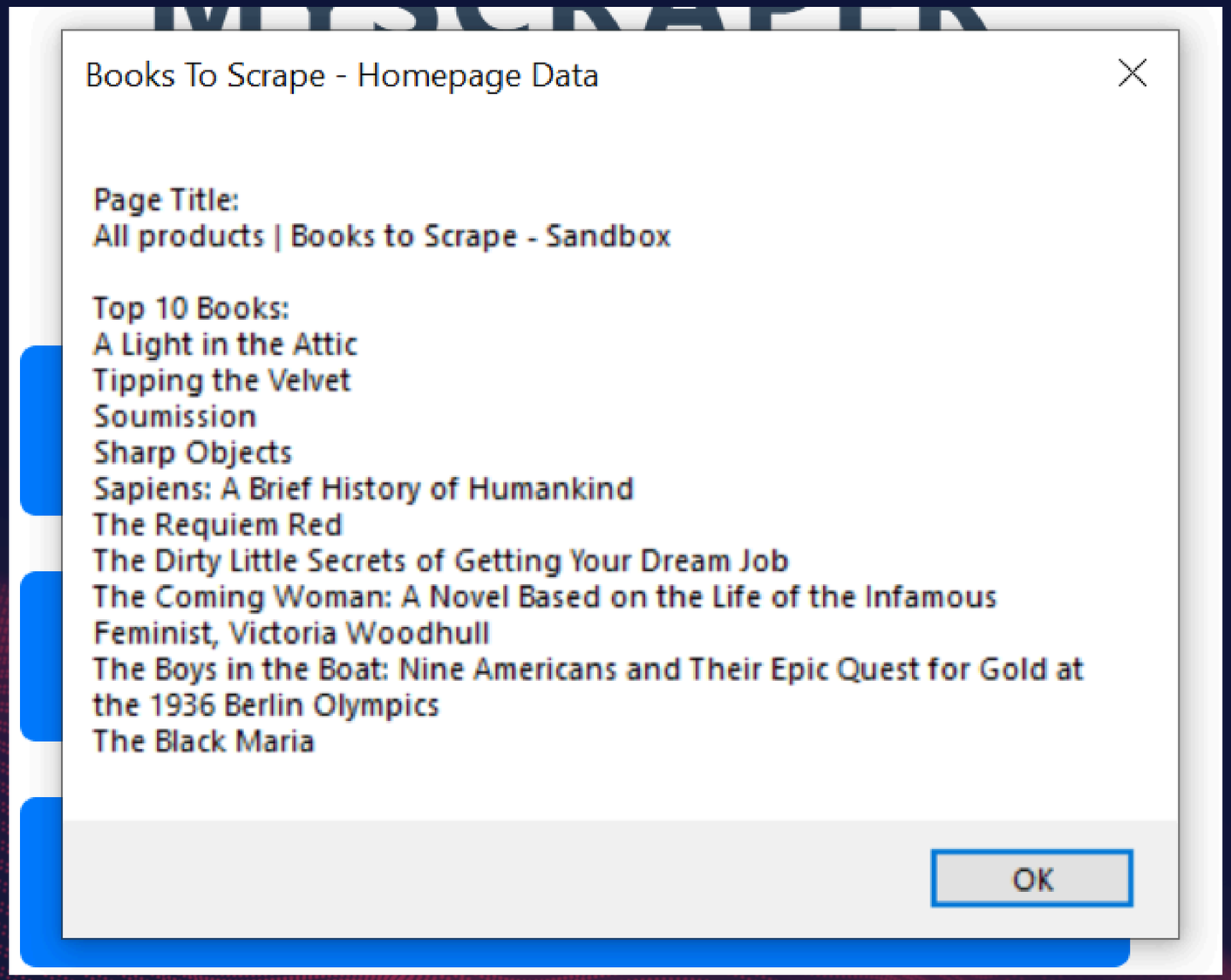
- Automate scraping of structured data from websites.
- Handle both public and login-based scraping flows.
- Provide users with an interactive GUI interface.
- Display scraped data in a text box and excel sheet.
- Ensure compatibility with the web browser.

GUI


MYSCRAPER
Please select an option
1. Screen scrap Books to Scrape homepage & display in message box
2. Screen scrap Books to Scrape homepage & save it in Excel file
3. Screen scrap OrangeHRM with credentials from config
4. Screen scrap OrangeHRM with user input
5. Exit

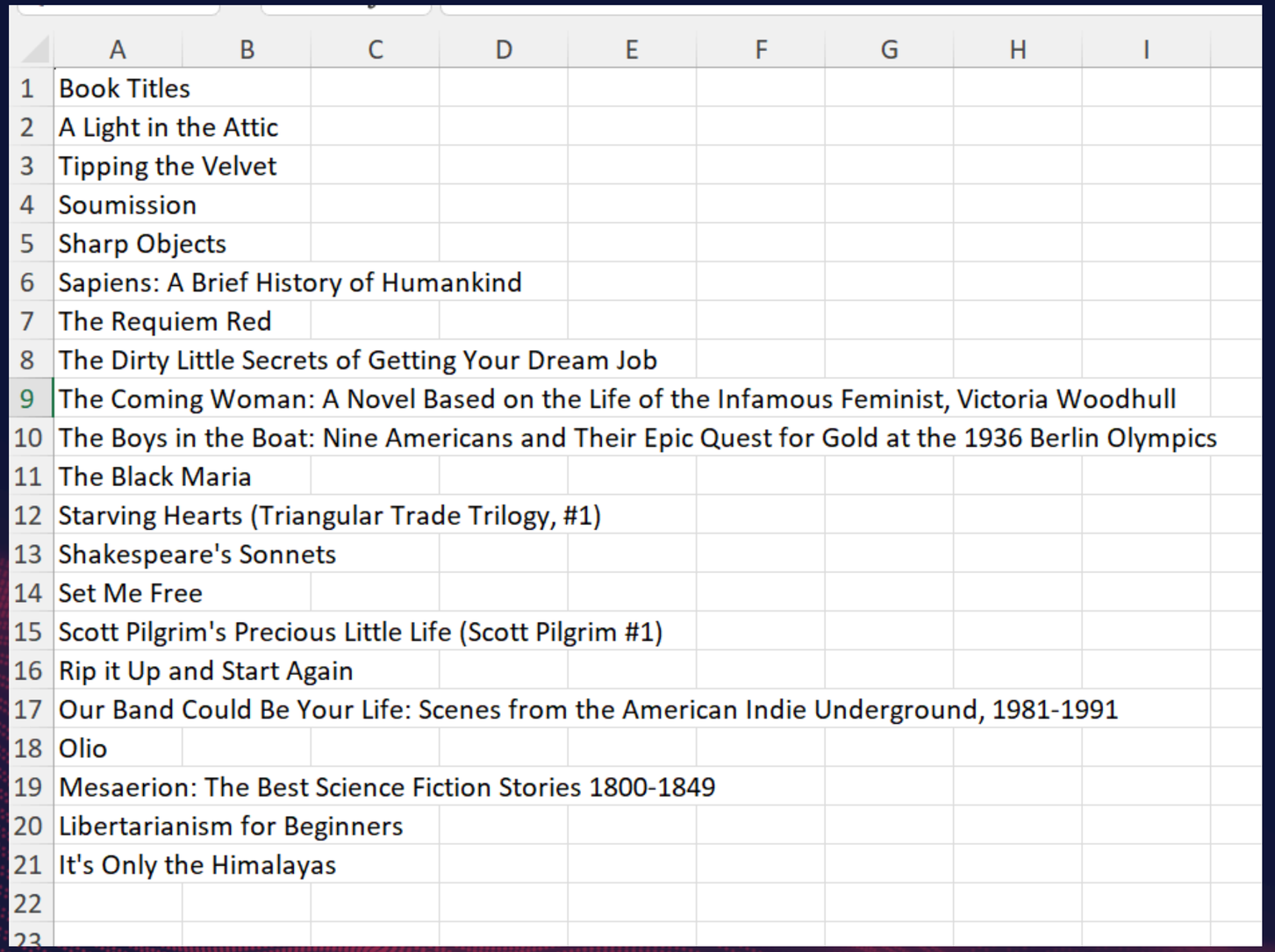
OUTPUT FOR

Option 1



OUTPUT FOR

Option 2



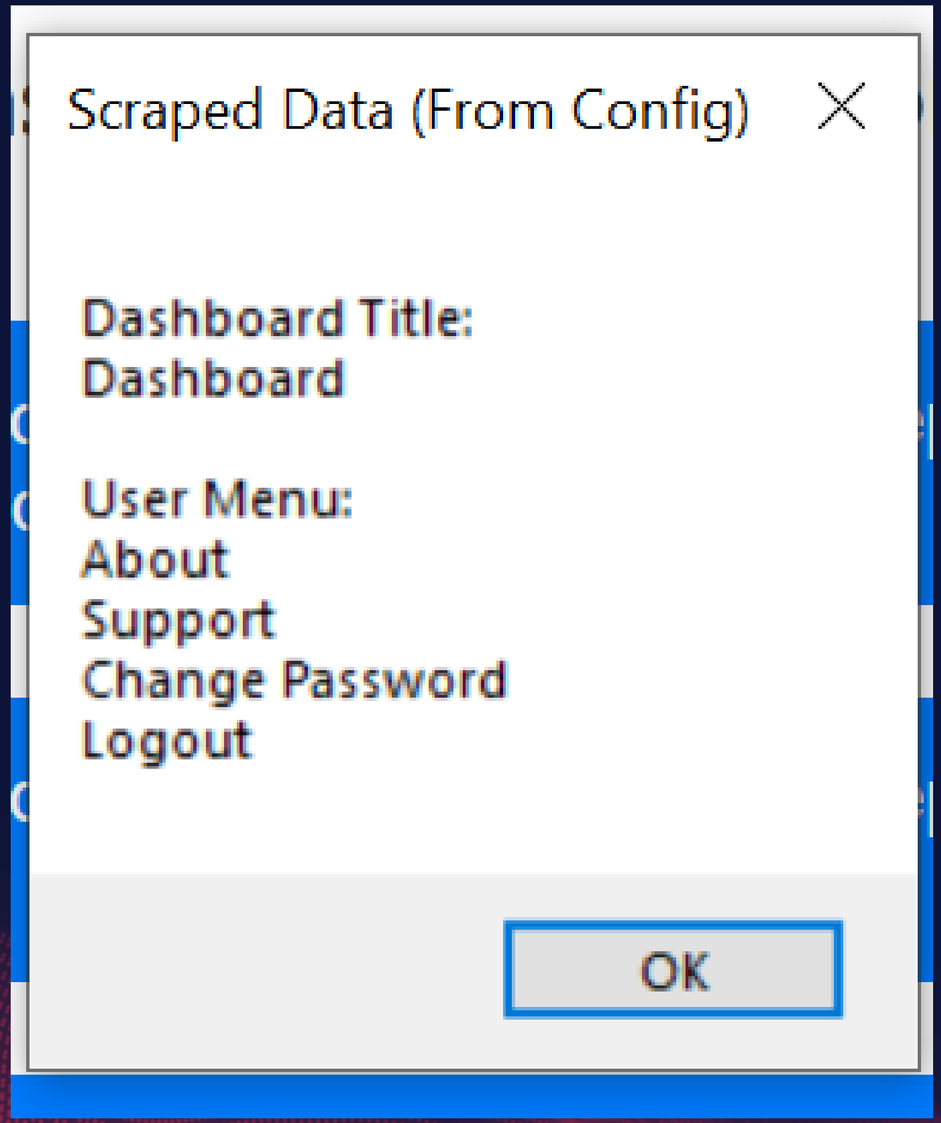
	A
1	Book Titles
2	A Light in the Attic
3	Tipping the Velvet
4	Soumission
5	Sharp Objects
6	Sapiens: A Brief History of Humankind
7	The Requiem Red
8	The Dirty Little Secrets of Getting Your Dream Job
9	The Coming Woman: A Novel Based on the Life of the Infamous Feminist, Victoria Woodhull
10	The Boys in the Boat: Nine Americans and Their Epic Quest for Gold at the 1936 Berlin Olympics
11	The Black Maria
12	Starving Hearts (Triangular Trade Trilogy, #1)
13	Shakespeare's Sonnets
14	Set Me Free
15	Scott Pilgrim's Precious Little Life (Scott Pilgrim #1)
16	Rip it Up and Start Again
17	Our Band Could Be Your Life: Scenes from the American Indie Underground, 1981-1991
18	Olio
19	Mesaerion: The Best Science Fiction Stories 1800-1849
20	Libertarianism for Beginners
21	It's Only the Himalayas
22	
23	

OUTPUT FOR

Option 3

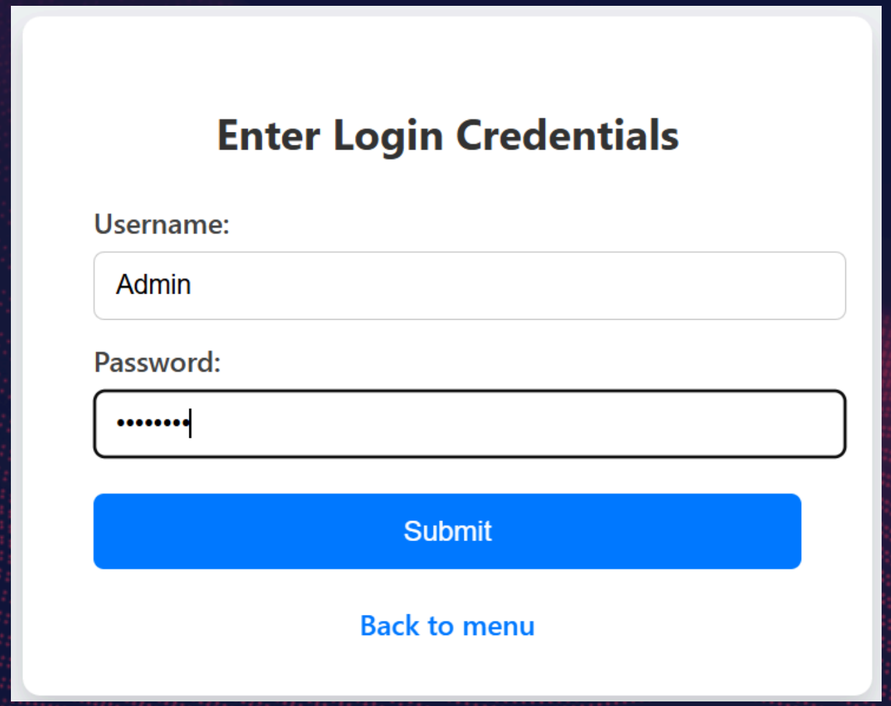
{} config.json > ...

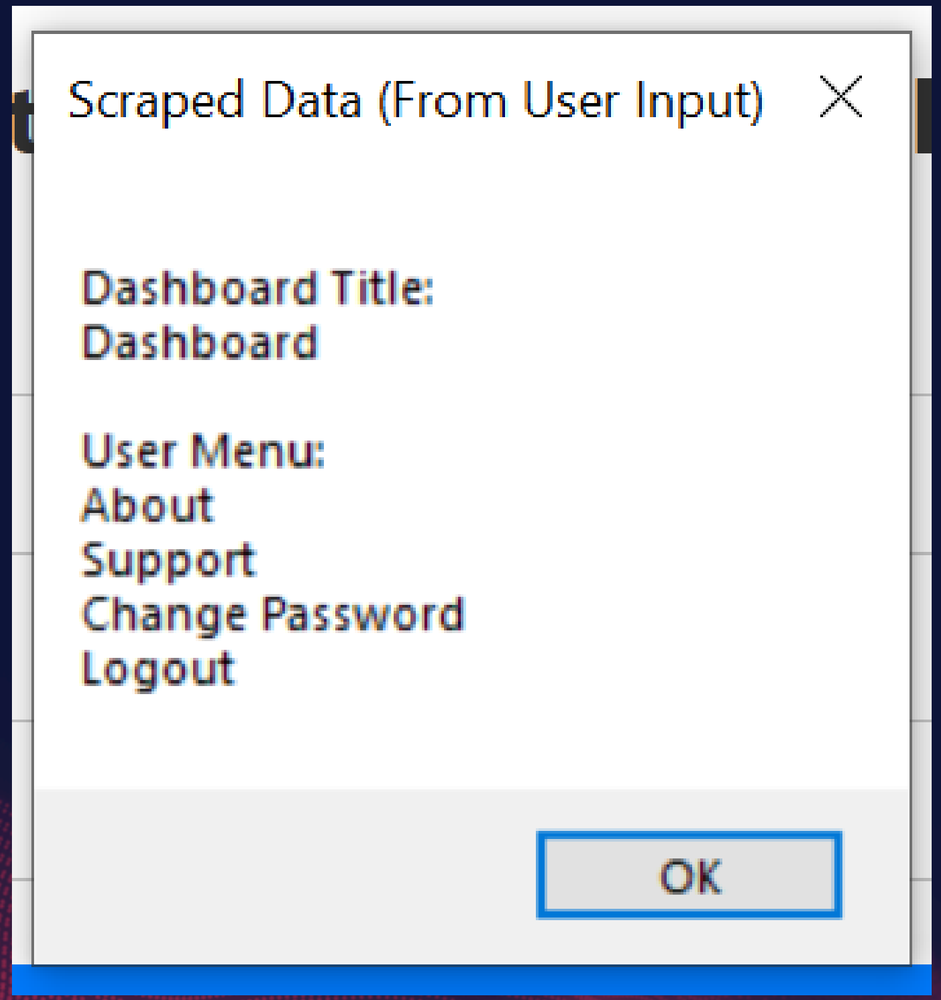
```
1   {
2       "username": "Admin",
3       "password": "admin123"
4   }
5
```



OUTPUT FOR

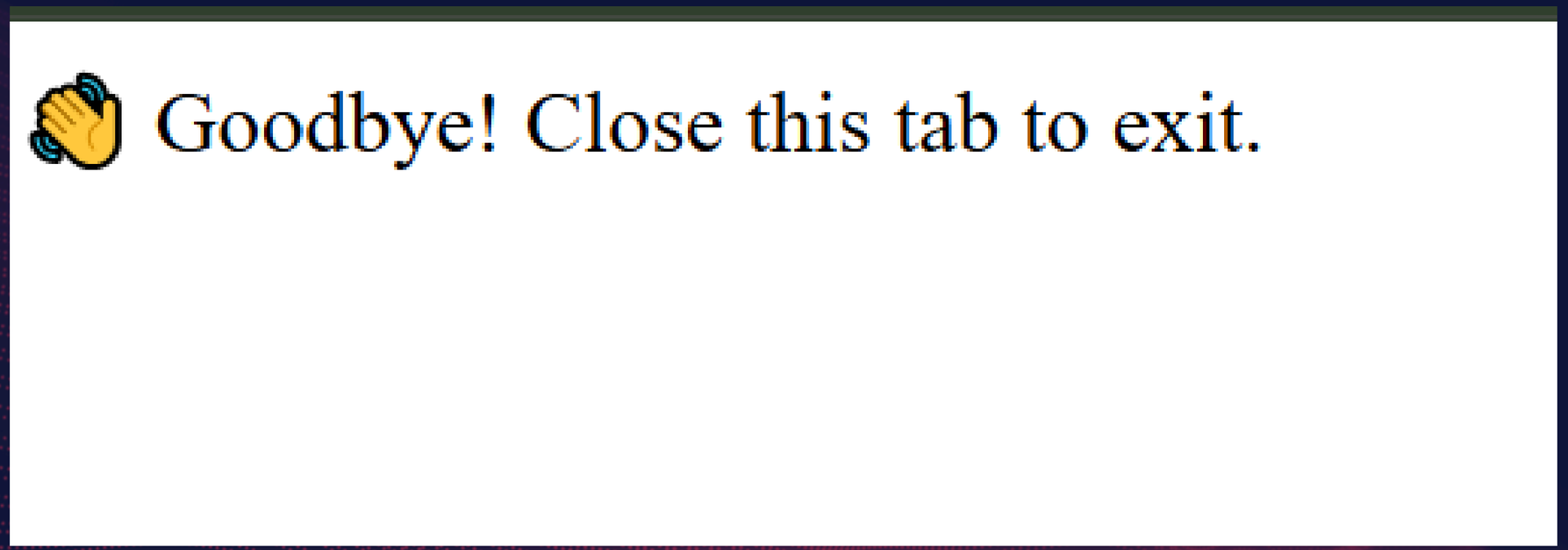
Option 4





OUTPUT FOR

Option 5



KEY FEATURES

- User-friendly GUI with 5 options
- Headless and headful browser modes supported
- Compatible with modern JavaScript sites
- Dynamic login support
- HTML report generation with auto-open
- Reusable and scalable architecture

CONCLUSION

- The project achieves its goal of automating data collection using Playwright.
- It provides a simple, secure, and visual way to handle scraping tasks.
- Easily extensible for academic or business applications.

THANK YOU