



**TUNIS BUSINESS SCHOOL**  
**UNIVERSITY OF TUNIS**

---

# Cyber Security Project

---

Final deliverable

## Web Scraping for Assessing Customer Satisfaction on an E-Commerce Site

### Participants:

**Chammaa Chaima**

*ID: \*\*\*\*6736*

*Email: chaimachammaa9@gmail.com*

**Jerbi Siwar**

*ID: \*\*\*\*2645*

*Email: siwar.jerbi.tbs@gmail.com*

**Marco Elaa**

*ID: \*\*\*\*7089*

*Email: ella.marco.tn@gmail.com*

**Smiri Lina**

*ID: \*\*\*\*6262*

*Email: lina01sm@gmail.com*

**IT-360 Information Assurance and Security**

Pr. Manel Abdelkader

**[GitHub Repository: Security Project](#)**

# Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Project Overview</b>	<b>2</b>
2.1	Objective . . . . .	2
2.2	Tools and Libraries . . . . .	2
<b>3</b>	<b>Methodology</b>	<b>2</b>
3.1	Step 1: Web Scraping . . . . .	2
3.2	Step 2: Data Cleaning . . . . .	2
3.3	Step 3: Data Visualization . . . . .	3
<b>4</b>	<b>Conclusion</b>	<b>3</b>

# 1 Introduction

Customer reviews are a vital indicator of product performance and brand perception. In this project, we developed a Python-based web scraper using Selenium to extract reviews from Decathlon's international website. The extracted data was cleaned using pandas and visualized with Power BI to derive meaningful insights.

## 2 Project Overview

### 2.1 Objective

To automate the process of collecting customer feedback from Decathlon and analyze satisfaction levels through graphical representation.

### 2.2 Tools and Libraries

- **Selenium:** for automated web browsing and scraping.
- **pandas:** for data cleaning and manipulation.
- **Power BI:** for visualizing customer feedback.
- **VS Code:** for running scripts.

## 3 Methodology

### 3.1 Step 1: Web Scraping

We used Selenium to locate and extract the following data from product pages:

- **Rating (stars):** Numeric values ranging mostly between 1 and 5 stars.
- **Date:** The date when the review was submitted.
- **Title:**
- **Content:** The main body of the review, containing customer opinions, feedback, and use cases. Many reviews are detailed and multilingual (e.g., Spanish, French, Turkish).
- **Reviewer:** The name of the reviewer, followed by their country.
- **Verified:** A boolean field (Yes/No) indicating whether the reviewer is a verified purchaser.
- **Brand Response:** Indicates whether the brand responded to the review.

### 3.2 Step 2: Data Cleaning

The raw dataset was cleaned using pandas:

- Drop rows with missing essential data
- Convert Date to datetime format
- Drop rows with invalid dates

- Fill missing values
- Create Year-Month column
- Extract Country
- Reset index

### 3.3 Step 3: Data Visualization

Using Power BI, we created:

- Overview Page: Summarizes overall review trends, including average rating, total reviews, and rating distribution
- Geographic Insights Page: Displays review volume and satisfaction levels by country.
- Product Comparison Page: Compares customer ratings and feedback across different Decathlon products.
- Review Content Analysis Page
- Review Insights Page

## 4 Conclusion

This project demonstrates the effectiveness of web scraping and data visualization in capturing and understanding customer sentiment. By automating data collection from Decathlon and interpreting it visually, we gained actionable insights into customer satisfaction trends, helping businesses refine their customer engagement strategies.