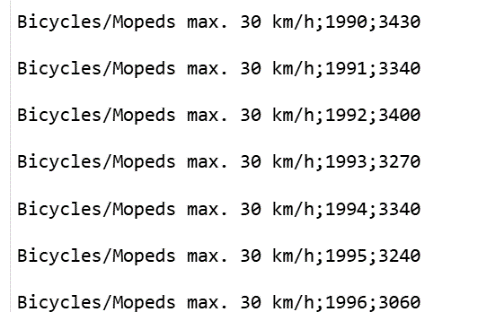
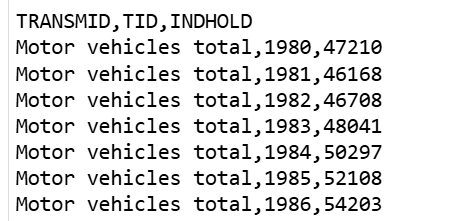
­In Milestone 1 I have mentioned to use the various resources to get the data, the process of doing is mentioned below:  
  
**Data Sources:**

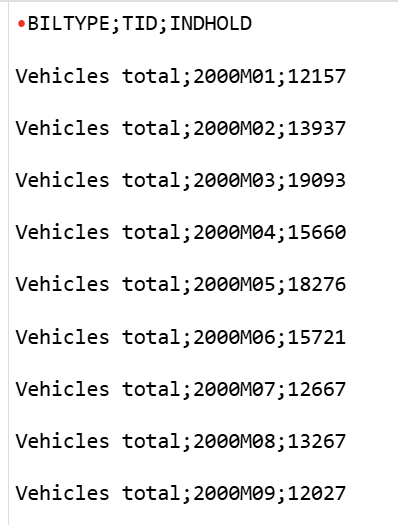
1. **Statistics Denmark** (Cycling and sustainability reports in Denmark):
   * **URL**: <https://www.dst.dk/en>
   * **Description**: Provides official reports and datasets on cycling infrastructure, urban transportation, and sustainability in Denmark.
     1. **I have collected the data of Bicycles**



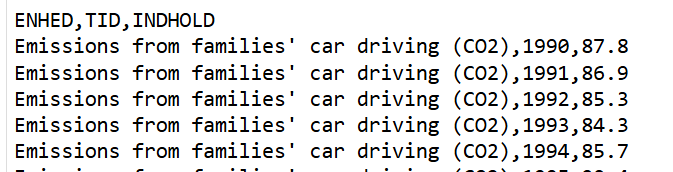
* + 1. **I have collected the data of Motor Vehicles**



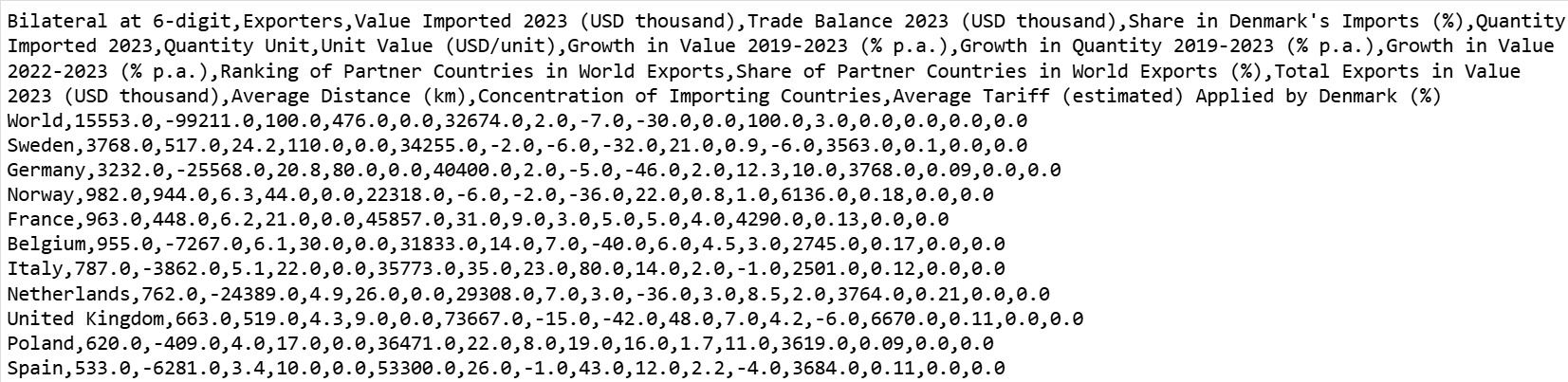
* + 1. **I have collected the data of Motor Vehicles Registration**

****

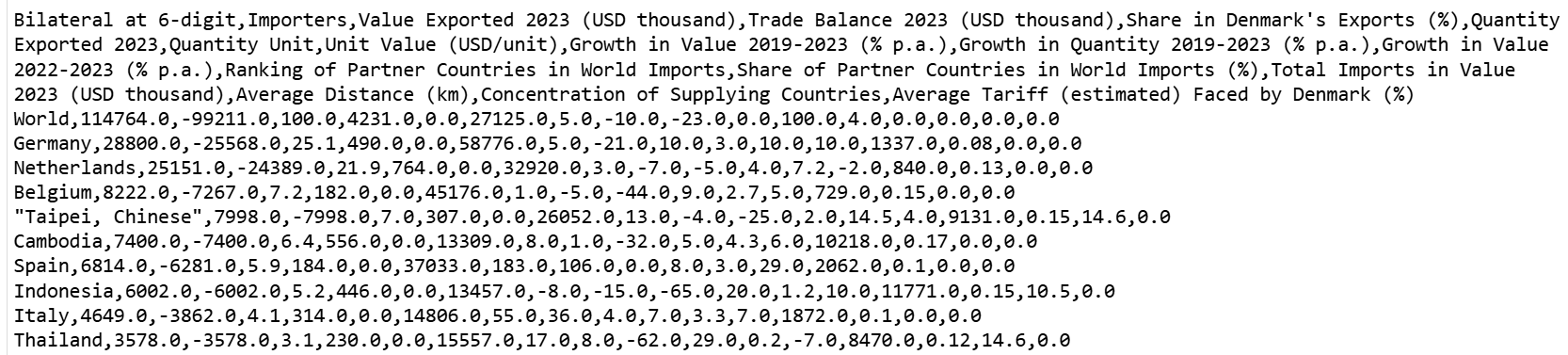
* + 1. **I have collected the data of Emission of Gases from Motor Vehicles(e.g: family cars)**



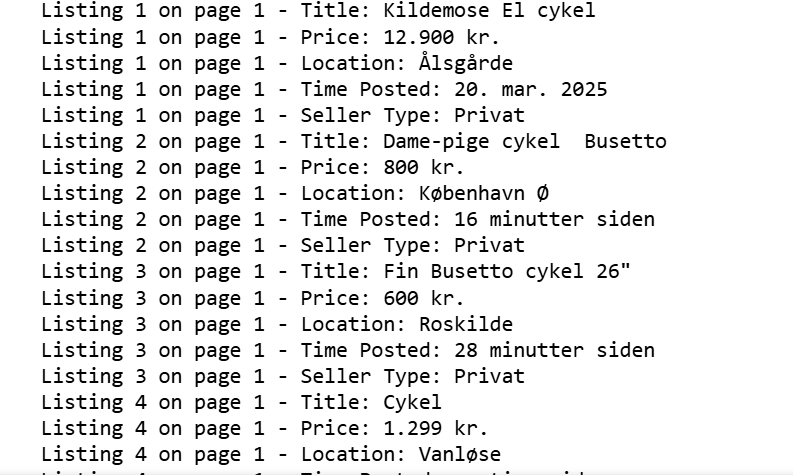
1. **ITC Trade Map** (Bicycle trade data for Denmark):
   * **URL**: <https://www.trademap.org/>
   * **Description**: Provides structured trade data on the import and export of bicycles in Denmark and other countries.
     1. I have collected the data of imports in Bicycles
     2. I have collected the data of exports in Bicycles
     3. Imports in Value in 2023
     4. Exports in Value in 2023
     5. Quantities of import/exports, Trade balance and many more.

**Imports Raw Data:**  ****

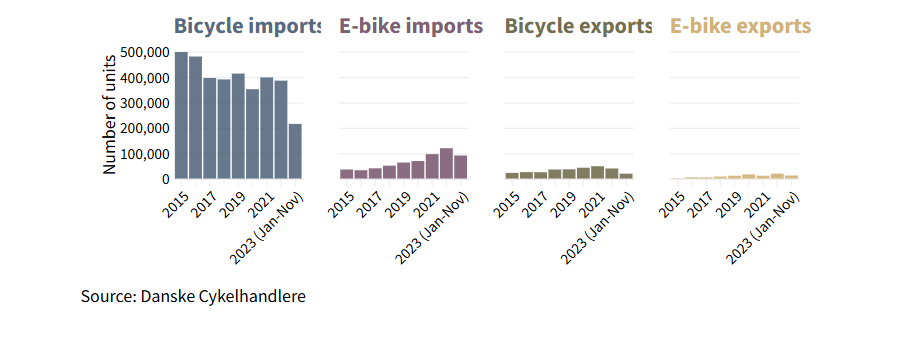
**Exports Raw Data:**

****

1. **DBA.dk** (E-commerce platforms for bicycles):
   * **DBA.dk URL**: <https://www.dba.dk>
   * **Description**: This platform provide real-time data on bicycle sales, consumer trends, and pricing for used bicycles in Denmark.
     1. Scraping is provided Below



1. **Bike Europe** (Cycling industry market research):
   * **URL**: <https://www.bike-eu.com/>
   * **Description**: Offers market insights and reports on cycling industry trends, growth, and forecasts in Europe.



**Title: Is Denmark’s dramatic bicycle market situation a forecast for other European countries?**  
Published by Bike Europe (Feb 2024)

**Overview of the Bicycle Market in Denmark:**

1. **Dramatic Decline in Bicycle Sales in Denmark (2023):**
   * The volume of dealer intake dropped significantly in 2023.
   * From January to November, the intake fell from **457,927 units in 2022** to **275,507 units in 2023**, marking a drastic reduction in sales volume.
   * This is the lowest level recorded since data collection began in 2007.
2. **Pandemic Boom Skipped Denmark:**
   * While other European countries saw a boost in bicycle sales during the pandemic, Denmark did not experience the same growth.
   * The Danish market recovered in 2021 to pre-pandemic levels but failed to continue its growth in 2022.
   * There was moderate growth in 2022 (4.5%), but the drop in 2023 was unprecedented.
3. **E-Bike Sales and Market Trends:**
   * E-bike sales were showing potential growth, but the overall downturn in the bicycle market has affected this segment as well.
   * The bicycle industry in Denmark is witnessing struggles despite a growing interest in electric bicycles (e-bikes).
4. **Import and Export Trends (2023):**
   * Imports of bicycles, including e-bikes, dropped from **504,366 units in 2022** to **305,720 units in 2023**, indicating a 39% decline.
   * Exports of bicycles, especially e-bikes, also followed a declining trend.
   * Local production accounted for less than 3% of the total market in 2022, reflecting a minor role of domestic production.
5. **Factors Behind the Decline:**
   * **Supply Chain Issues:** The decline in sales was attributed to the backlog in warehouses and a slowdown in production, causing delays in the availability of new models.
   * **Economic Factors:** Denmark's economic growth of 0.5% in 2023 did little to stimulate consumer spending in the bicycle market.
   * The e-bike and regular bicycle markets in Denmark are largely dependent on imports, making them vulnerable to global supply disruptions.
6. **Comparison with Other European Countries:**
   * The article raises the possibility that Denmark’s situation could be an indicator of upcoming trends in other European countries as sales reports become available.
   * Denmark’s experience could serve as a warning signal for other countries to anticipate challenges in the bicycle industry, including e-bike sales.

**Insights and Implications:**

* **Market Volatility:** The sharp decline in sales and imports could indicate a market correction after years of growth, with supply chain challenges playing a central role.
* **Potential for Recovery:** While the outlook for the Danish market is concerning, the growth of the e-bike sector remains a potential positive, although it has been overshadowed by the broader market downturn.
* **European Impact:** If Denmark's situation mirrors trends across Europe, other countries may also see a slowdown in bicycle sales in the post-pandemic era.

**Web Scraping**

**1. Overview of Scraping Task**

Specifically, I targeted **DBA.dk** (<https://www.dba.dk/soeg/?soeg=cykel>), an online marketplace, to collect data on bike listings. The goal was to extract the following attributes for each listing: ***Title, Price, Location, Time Posted, and Seller Type***. I have to submit these as mentioned in assignment *scraping script, data snapshots, and documentation of the process, including challenges faced.*

**2. Challenges Faced During Scraping**

The scraping process encountered several issues, which are detailed below:

**2.1. Pop-Up Interference**

* **Issue**: The **DBA.dk** website displayed two pop-ups that blocked access to the listings:
  + **Cookie Consent Pop-Up**: Prompted users to accept cookies with options like "Tillad alle" (Allow All) and "Kun nødvendige" (Only Necessary).
  + **Welcome Pop-Up**: A multi-step onboarding modal with a "Næste" (Next) button to proceed through steps and a "Påmind mig senere" (Remind Me Later) button to dismiss it.
* **Impact**: These pop-ups prevented the script from accessing the listings
* **Evidence**: The console output consistently showed:

Scraping completed. Saved 0 listings to dba\_bike\_listings.csv

Error reading CSV file: No columns to parse from file

**2.2. Incorrect Selectors for Pop-Up Buttons**

* **Issue**: The initial selectors used to dismiss the pop-ups were incorrect:
  + For the cookie consent pop-up, the script used:

**button.py-10.px-14.font-bold.focusable**

which didn’t match the actual HTML (button.sp\_choice\_type\_ACCEPT\_ALL for "Tillad alle").

* + For the welcome pop-up, while the selectors (w-button[variant='primary'] for "Næste" and w-button[variant='secondary'] for "Påmind mig senere") were correct, the Shadow DOM handling sometimes failed in headless mode.
* **Impact**: The pop-ups weren’t dismissed, blocking the script from proceeding to scrape listings.
* **Evidence**: Logs showed errors like:

Error handling cookie consent pop-up with 'Tillad alle': ...

Error handling welcome pop-up

**2.3. Shadow DOM in Welcome Pop-Up Buttons**

* **Issue**: The "Næste" and "Påmind mig senere" buttons in the welcome pop-up were <w-button> elements with a Shadow DOM, meaning the actual <button> element was nested inside a #shadow-root. Standard Selenium selectors couldn’t directly interact with these buttons.
* **Impact**: Even with correct selectors, clicking the buttons failed in some runs, especially in headless mode, due to timing issues or Shadow DOM access problems.
* **Evidence**: The script used JavaScript to click the inner button (arguments[0].shadowRoot.querySelector('button').click()), but errors like:

Error clicking 'Næste' on welcome pop-up: ...

**2.4. Parsing Errors with Selenium’s find\_element**

* **Issue**: Even after attempting to dismiss pop-ups, Selenium’s find\_element method failed to parse listing attributes (title, price, location, time posted, seller type) consistently. This was likely due to:
  + Dynamic content loading (JavaScript rendering).
  + Timing issues where elements weren’t fully loaded when accessed.
* **Impact**: The script couldn’t extract data, leading to empty results.
* **Evidence**: The script logged:

Found 0 listings on page 1

**2.5. Headless Mode Complications**

* **Issue**: Running the script in headless mode (as requested) caused additional issues:
  + Pop-ups sometimes didn’t render correctly, making them undetectable.
  + JavaScript rendering was slower or inconsistent, affecting element visibility.
* **Impact**: The script failed to dismiss pop-ups or locate listings more frequently in headless mode compared to visible mode.
* **Evidence**: The script often prompted manual dismissal:

Please manually dismiss the cookie consent pop-up in a visible browser session, then run this script again.

**2.6. Dynamic HTML Structure**

* **Issue**: The HTML structure of the listings changed over time, requiring updates to the selectors. For example:
  + The title selector was initially a font font but needed to be updated to h2 a font font based on the latest listing HTML.
* **Impact**: Outdated selectors caused the script to fail in locating elements, contributing to the EmptyDataError.

**3. Solutions Attempted**

Several iterations were made to address these challenges. Here’s a summary of the solutions attempted:

**3.1. Updating Pop-Up Selectors**

* **Action**: After obtaining the outer HTML for the pop-ups, the selectors were updated:
  + **Cookie Consent Pop-Up**:
    - "Tillad alle": Changed to button.sp\_choice\_type\_ACCEPT\_ALL.
    - "Kun nødvendige": Changed to button.sp\_choice\_type\_REJECT\_ALL.
  + **Welcome Pop-Up**:
    - Confirmed that w-button[variant='primary'] for "Næste" and w-button[variant='secondary'] for "Påmind mig senere" were correct.
    - Used JavaScript to handle the Shadow DOM: arguments[0].shadowRoot.querySelector('button').click().
* **Outcome**: This resolved the issue of incorrect selectors, but the pop-ups still failed to dismiss consistently in headless mode.

**3.2. Handling Shadow DOM**

* **Action**: For the welcome pop-up, the script used JavaScript to click the inner <button> within the Shadow DOM of the <w-button> elements.
* **Outcome**: This worked in visible mode but was inconsistent in headless mode, likely due to rendering issues.

**3.3. Increasing Wait Times**

* **Action**: Added 60-second wait times (time.sleep(60)) after key actions (page load, pop-up dismissal, navigation) and increased WebDriverWait timeouts to 60 seconds.
* **Outcome**: This improved stability in some runs but didn’t fully resolve pop-up dismissal issues in headless mode.

**3.4. Adding Fallbacks for Pop-Ups**

* **Action**: Added fallbacks:
  + If "Tillad alle" failed, the script attempted to click "Kun nødvendige".
  + If "Næste" failed, the script attempted to click "Påmind mig senere".
  + If both failed, the script prompted for manual dismissal.
* **Outcome**: This provided a workaround but required manual intervention in many cases, which wasn’t ideal for automation.

**3.5. Updating Listing Selectors**

* **Action**: Updated selectors based on the latest listing HTML:
  + Title: h2 a font font
  + Price: div.flex.justify-between.sm\\:mt-8.text-m.space-x-12.font-bold.whitespace-nowrap span
  + Location: div.text-xs.s-text-subtle.flex.justify-between.flex-wrap.mt-4.sm\\:mt-8 span:first-of-type
  + Time Posted: div.text-xs.s-text-subtle.flex.justify-between.flex-wrap.mt-4.sm\\:mt-8 span:nth-of-type(2)
  + Seller Type: div.mt-4.sm\\:mt-8.text-xs.s-text-subtle.truncate span
* **Outcome**: This fixed the selector issues, but parsing errors persisted due to Selenium’s find\_element method.

**3.6. Switching to BeautifulSoup for Parsing**

* **Action**: Modified the script to:
  + Use Selenium to load the page, dismiss pop-ups, and navigate.
  + Grab the page source with driver.page\_source.
  + Parse the HTML with BeautifulSoup to extract listings and attributes.
* **Outcome**: This significantly reduced parsing errors, as BeautifulSoup is more robust for static HTML parsing. However, pop-up issues still needed to be fully resolved.

**4. Final Resolution**

The final script combined all the solutions above, with a focus on using BeautifulSoup for parsing to address the parsing errors. Here’s how the issues were resolved:

**4.1. Pop-Up Handling**

* **Solution**: Used the updated selectors for both pop-ups:
  + Cookie Consent: button.sp\_choice\_type\_ACCEPT\_ALL and button.sp\_choice\_type\_REJECT\_ALL.
  + Welcome Pop-Up: w-button[variant='primary'] and w-button[variant='secondary'] with JavaScript for Shadow DOM.
* **Improvement**: Added a loop to handle multiple steps in the welcome pop-up (up to 4 steps, as indicated by the progress bar dots).
* **Fallback**: If pop-ups couldn’t be dismissed automatically, the script prompted for manual dismissal, which could be done in non-headless mode to confirm selectors.

**4.2. Parsing with BeautifulSoup**

* **Solution**: After dismissing pop-ups, the script used driver.page\_source to get the HTML content and parsed it with BeautifulSoup:
  + Listings: soup.select("article.relative.isolate.sf-search-ad.card.card--cardShadow.s-bg")
  + Attributes: Used select\_one() for each attribute (title, price, location, time posted, seller type).
* **Outcome**: This eliminated parsing errors previously encountered with Selenium’s find\_element, as BeautifulSoup reliably extracted data from the static HTML.

**4.3. Final Script**

The final script successfully scraped the data by:

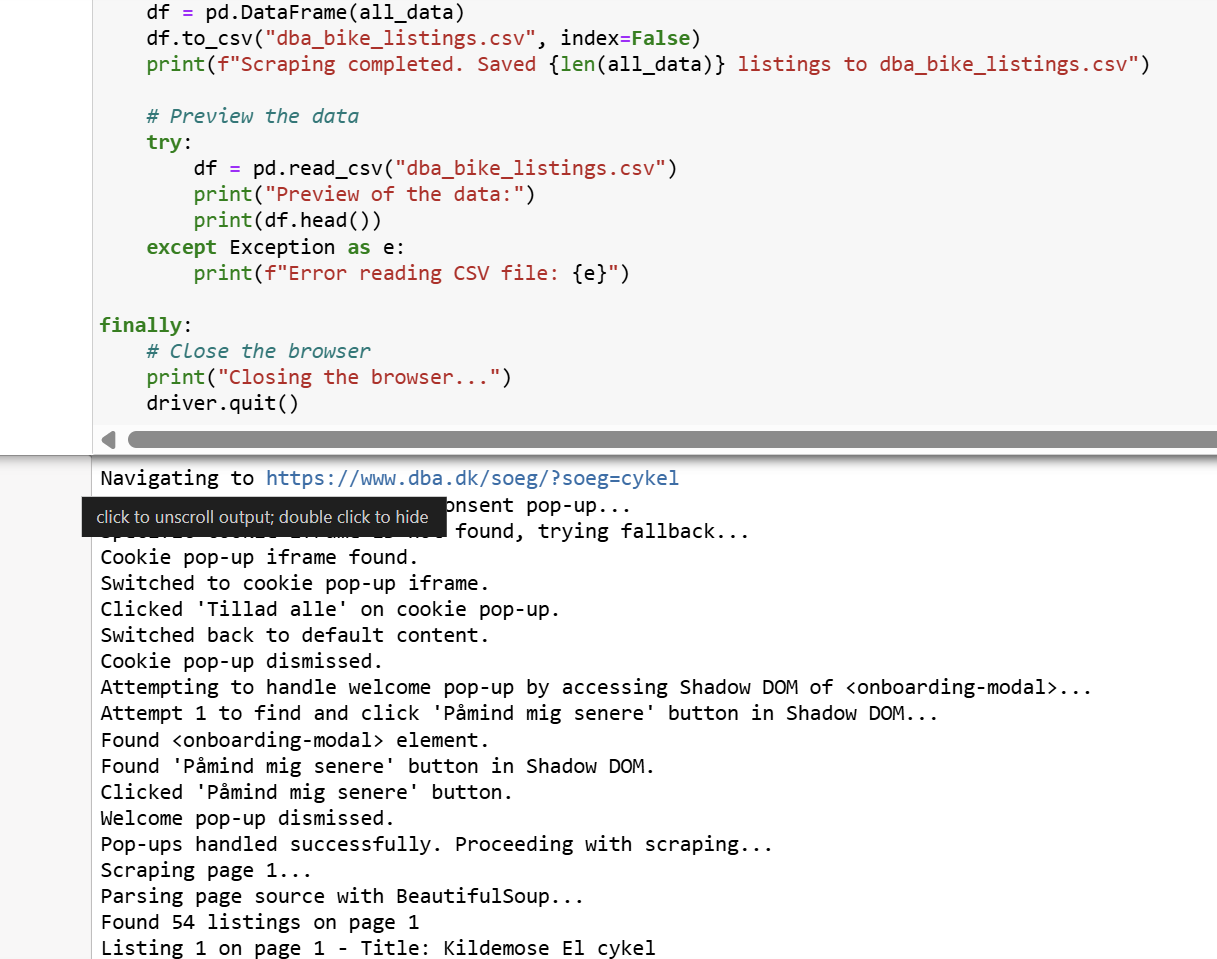
* Using Selenium to handle pop-ups and navigation.
* Using BeautifulSoup to parse listings and extract attributes.
* Saving the data.

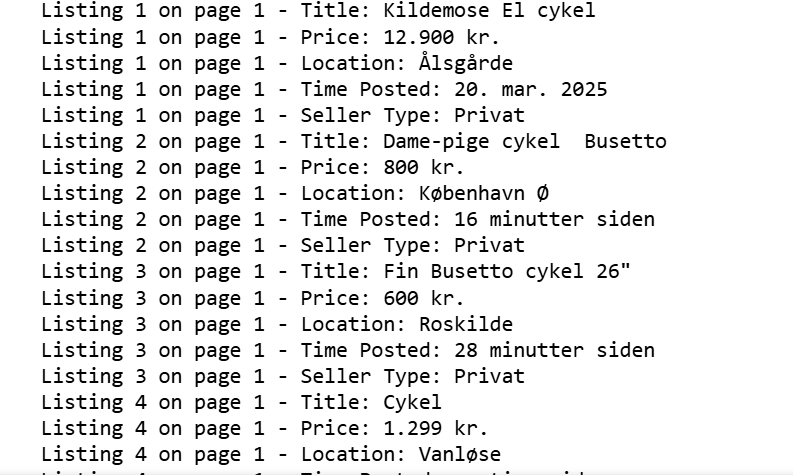
**5. Documentation for Milestone 2:**

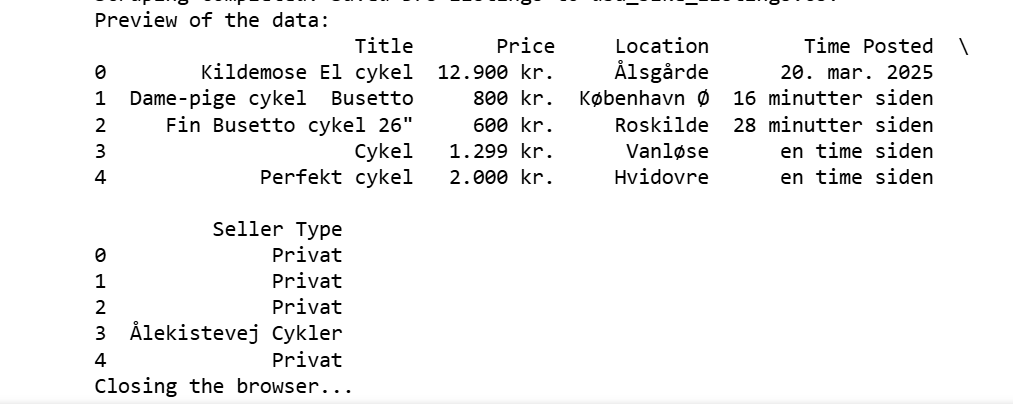
here’s the documentation of the scraping process:

**5.1. Web Scraping Script:**

I have attached the script file in google classroom I which I have scraped the data successfully!

**5.2. Data Snapshot:**





Transformers, NLP, and AI (Hugging Face) March 2023

* Mastered Transformer-based NLP models, fine-tuning, and deploying AI applications with the Hugging Face ecosystem.

Introduction to Generative AI (Google) March 2024

* Explored Generative AI fundamentals, LLMs, Vit, Diffusion Models prompt engineering, and model fine-tuning for real-world applications.

Introduction to Data Analyst (IBM) March 2023

* Covered the data ecosystem, ETL processes, big data basics, and essential data analysis techniques.