December 2023

ESILV
ENGINEERING SCHOOL
DE VINCI PARIS

SUSTAINABLE NIKE SNEAKER MARKETPLACE

WEBSCRAPPING & DATA PROCESSING
FINAL PROJECT

Github Link
KARUNATHASAN Nilany
SAMBATH Sïndoumady
DIA2

TABLE OF CONTENT

Introduction

In the evolving landscape of e-commerce, our project aims to address the growing demand for sustainable options within niche markets, specifically focusing on second-hand Nike sneakers.

The project involves the development of a dedicated interface that allows users to explore pre-owned Nike sneakers. How can we empower users to make sustainable and economical choices when purchasing second-hand Nike sneakers ?

Through web scraping, we will gather data from popular platforms like Vinted and Vestiaire Collective to create a comprehensive database. Features such as profit calculation and the estimation of the carbon footprint associated with shipping will be proposed.

The price comparison feature entails evaluating the average original price scraped directly from Nike's data against the second-hand price. Additionally, the carbon footprint comparison provides users with insights into the environmental impact associated with shipping second-hand items compared to ordering directly from Nike.

Github Link

Project Details

Web Scraping Marketplace Data

To ensure a diverse and up-to-date inventory, we implemented web scraping functionalities for Vinted and Vestiaire Collective. The scraping scripts collect various information, including titles, prices, brands, sizes, links, image sources, localizations, and calculate approximate carbon footprints and profit made, providing users with a comprehensive selection of second-hand Nike sneakers given a query. We will use the Selenium library to scrape comprehensive data from all three websites, ensuring a robust and effective extraction process.

1.1. Legal and ethical considerations

Before embarking on the data gathering process, we conducted research to ensure that web scraping activities for Vinted, Vestiaire Collective, and Nike adhere to legal and ethical standards. It is essential to emphasize that our project respects the terms of service of these platforms, and the scraping process is carried out responsibly, avoiding any disruptive impact on the websites. Furthermore, our aim is not to compromise user privacy or violate any regulations. With a commitment to ethical data practices, we have crafted our scraping scripts to access only publicly available information, contributing to a positive and lawful online environment.

1.2. Vinted Data Scraping

Our **get_vinted_data** function allows users to specify parameters such as gender, color, and size to tailor their search. These parameters are intelligently mapped to corresponding Vinted category IDs, enabling more precise search results.

For gender-specific searches, the function utilizes mappings for men, women, girls, and boys, each associated with a unique Vinted category ID. Similarly, color and size parameters are mapped to their respective IDs based on commonly used variations. The constructed URL incorporates these parameters to generate a refined search query, ensuring that the scraped data aligns with the user's preferences. The function then navigates to the constructed URL page, accepts cookies, and then iterates through the paginated search results. For each article, it extracts relevant details such as title, price, brand, size, link, image source, and localization. The localization information is used to calculate an approximate carbon footprint with the function **calculate_approximate_carbon_footprint** using geopy and predefined emission rates. The resulting data is stored in a pandas DataFrame.

1.3. Vestiaire Collective Data Scraping

Similarly, our **get_vestiaire_data** function navigates to a constructed URL page for the specified user characteristics, accepts cookies, and iterates through the product containers on the page. It extracts details such as title, price, brand, size, link, image source, localization, and approximate carbon footprint. The localization information is once again used to estimate the carbon footprint. The extracted data is stored in another pandas DataFrame.

1.4. Nike Price Scraping for Average Price Calculation

To provide users with a benchmark for pricing comparison, we incorporated a scraping functionality to obtain the average price of Nike sneakers directly from the official Nike website. Our **get_nike_prices** function navigates to the Nike website, accept cookies, and search for the specified Nike product. It then extracts the prices of the products listed on the search results page. The obtained prices are stored in a list, and the average price is calculated.

Data Processing and Concatenation

After scraping data from both sources, the process_dataframe function is applied to refine the DataFrame structure for easier analysis.

The next step involved calculating the profit made for each item by subtracting the second-hand item's price from the average Nike price. This is done using the **calculate_profit_made** function.
Finally, the data frames from Vinted and Vestiaire Collective are concatenated into a single DataFrame named **merged_data**.

Here is a snippet of the resulting dataframe generated based on the following user specifications :

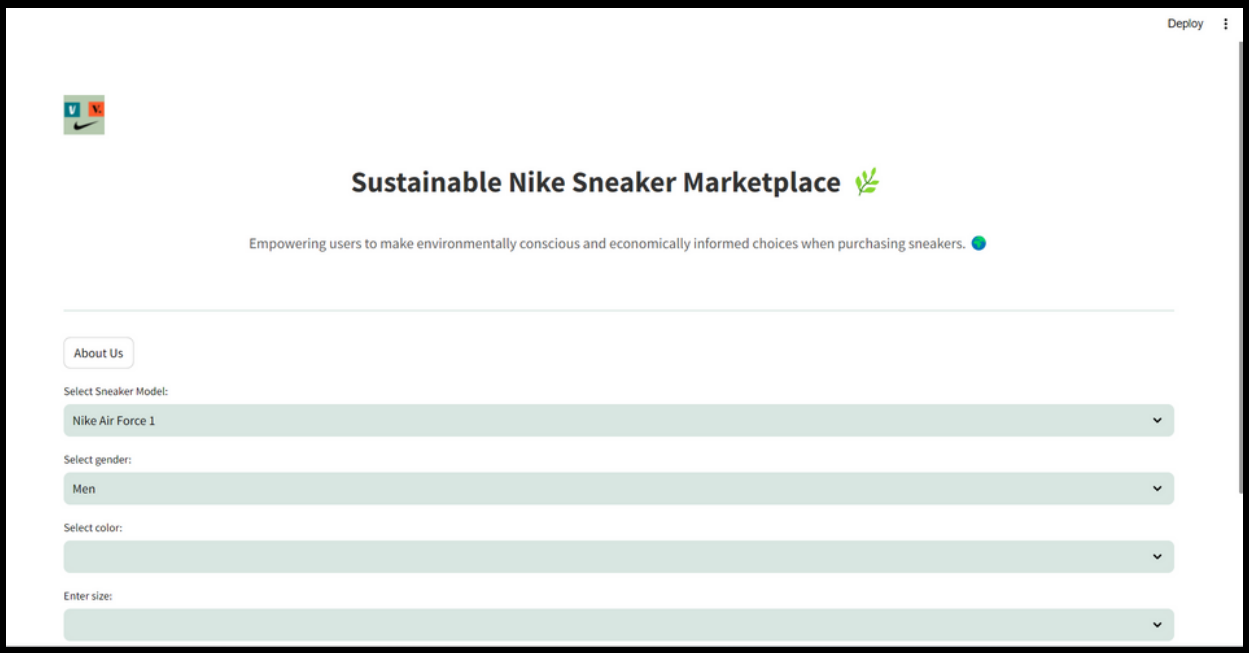
- Query: Nike Cortez
- Gender: Women
- Color: Multicolore
- Size: 40

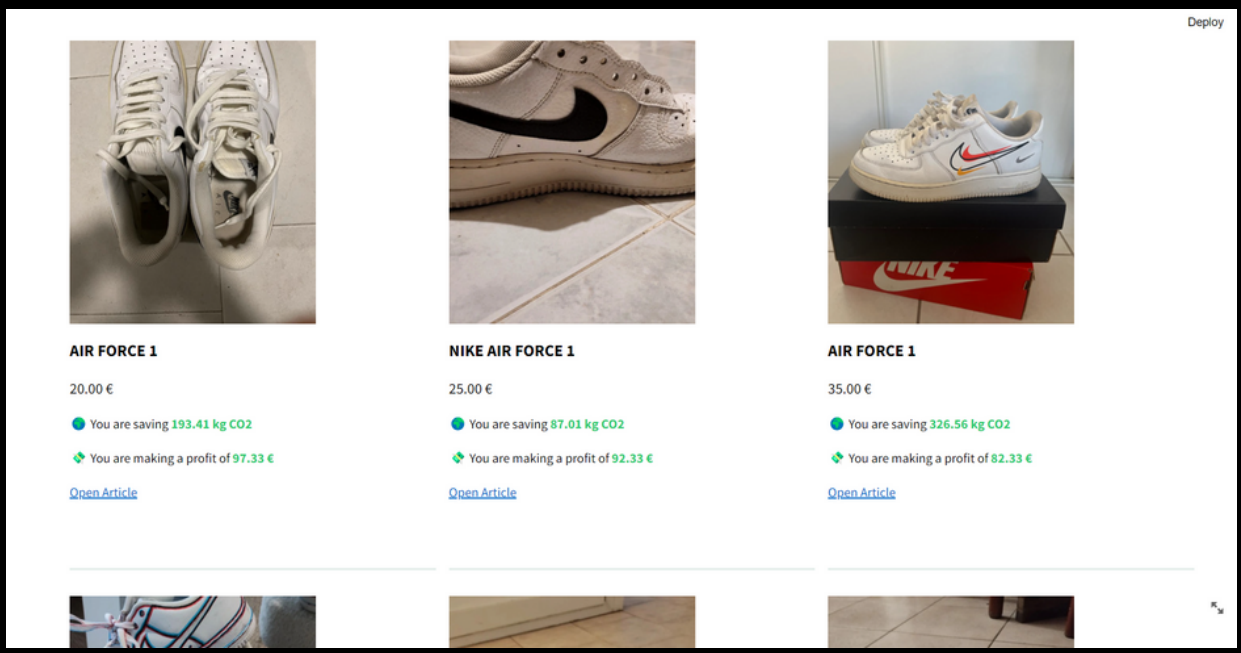
Title	Price	Brand	Size	Link	Image Source	Localisation	Approximate Carbon Footprint (kg CO2)	Approximative Carbon Print from Ordering on Nike (kg CO2/ton)	Carbon Profit (kg CO2)	Average price of the product on Nike	Profit Made
NIKE Cortez Baskets	141.75	NIKE	40	https://fr.vestiairecollective.com/chaussures-...	https://images.vestiairecollective.com/cdn-cgi...	France	50.586158	405	354.413842	107.99	-33.76
NIKE Cortez Baskets	178.50	NIKE	40	https://fr.vestiairecollective.com/chaussures-...	https://images.vestiairecollective.com/cdn-cgi...	France	50.586158	405	354.413842	107.99	-70.51
Nike Cortez Granates	5.00	NIKE	40	https://www.vinted.fr/items/1057462789-nike-co...	https://images1.vinted.net/t/02_00dd8_L3visLio...	ESPAGNE	240.819687	405	164.180313	107.99	102.99
Nike cortez	15.00	NIKE	40	https://www.vinted.fr/items/2095600086-nike-co...	https://images1.vinted.net/t/02_00305_HQCSd3Lb...	ESPAGNE	240.819687	405	164.180313	107.99	92.99
Nike cortez blanche	15.00	NIKE	40	https://www.vinted.fr/items/3176601437-nike-co...	https://images1.vinted.net/t/02_020c9_9NrFZnjK...	VENDAT, FRANCE	61.694651	405	343.305349	107.99	92.99
nike cortez	15.00	NIKE	40	https://www.vinted.fr/items/499423376-nike-cor...	https://images1.vinted.net/t/01_01221_hMbAP9Yk...	SAINT-QUENTIN, FRANCE	25.876746	405	379.123254	107.99	92.99
Nike cortez	19.00	NIKE	40	https://www.vinted.fr/items/2967850216-nike-co...	https://images1.vinted.net/t/02_007d1_Bu3HUfko...	FRANCE	50.586158	405	354.413842	107.99	88.99
Basket CORTEZ de Forest Gump	20.00	NIKE	40	https://www.vinted.fr/items/3693845712-basket-...	https://images1.vinted.net/t/03_00965_vXo2qfwy...	AUBERVILLIERS, FRANCE	1.363107	405	403.636893	107.99	87.99
Basket Nike cortez	20.00	NIKE	40	https://www.vinted.fr/items/3713657388-basket-...	https://images1.vinted.net/t/03_0217e_n2BbfRqv...	SAINT-MARTIN-DU-VAR, FRANCE	134.401491	405	270.598509	107.99	87.99
Nike Cortez	20.00	NIKE	40	https://www.vinted.fr/items/2499159015-	https://images1.vinted.net/t/02_020c0_5k2cKKa...	FRANCE	50.586158	405	354.413842	107.99	87.99

Fig 1 : Snipet of the result dataframe for the specified query

The next phase of our project involves the development of a user-friendly Streamlit application. It aims to provide an intuitive and visually appealing interface for users to effortlessly navigate and make informed decisions when shopping for sustainable and second-hand Nike sneakers.

Here are some screenshots of the first interface we have made :





Challenges

2.1 Lack of Nike Transparency

A significant hurdle in our project stems from Nike's limited transparency regarding the carbon footprint associated with their products shipping. The absence of comprehensive data acts as a roadblock, hindering our ability to furnish users with detailed environmental information concerning second-hand Nike sneakers.

2.2 Uncertainty in Manufacturing Locations

Compounding the challenge is the undisclosed manufacturing locations of Nike sneakers. While anecdotal evidence suggests production concentrations in Vietnam, the lack of official data introduces uncertainty into our efforts to provide users with precise details about the origin of the products. This opacity in manufacturing locations affects overall transparency and traceability.


VIETNAMNET GLOBAL

08/01/2022 08:26 (GMT+07:00)

Vietnam becomes major manufacturer of Nike sports shoes

Follow VietNamnet on Google News

Vietnam has surpassed China to become the major manufacturer of Nike sports shoes.

Vietnam surpasses China to become the main manufacturer of Nike sports shoes.

Fig 2 : Screenshot of an article from vietnamnet stating Vietnam as the major manufacturer of Nike sneakers

2.3. Calculation of Carbon Footprint

In the absence of precise manufacturing location information, we resort to estimating the carbon footprint associated with shipping. For instance, considering the transportation from Vietnam to Laakdal (Nike's logistics hub for Europe as per their official website) and then from Laakdal to Paris.

Fig 3 : Screenshot from Nike official website

Shipping from Vietnam to Laakdal (by container ship) :
Carbon Emissions (maritime) = 9,000 km × 40 g CO2/ton-km = 360,000 g CO2/ton

Shipping from Laakdal to Paris (by truck) :
Carbon Emissions (truck) = 300 km × 150 g CO2/ton-km = 45,000 g CO2/ton

Now, adding both values for the total estimate:
Total Carbon Emissions ≈ 360,000 g CO2/ton + 45,000 g CO2/ton
≈ 405,000 g CO2/ton

Converting this to kilograms (1 ton = 1,000 kg) :

Total Carbon Emissions ≈ 405,000 g CO2/ton × (1 kg / 1,000 g) = **405 kg CO2/ton**

This simplified approximation provides a rough estimate of the total carbon emissions for shipping from Vietnam to Laakdal and then from Laakdal to Paris, amounting to approximately 405 kilograms of CO2 per ton of goods shipped. It's essential to note that this is a basic calculation, and the actual carbon footprint may be influenced by various factors not considered in this simplified model.

2.4 Difficulty of Real-time Processing

The delay introduced by Selenium scraping for both platforms made the real-time extraction of user-defined queries impractical due to extended waiting periods.
To address this challenge, we made a strategic decision to focus on a curated selection of the five most sold Nike shoe models, namely Nike Air Force 1, Nike Blazer, Nike Dunk, Nike Mag and Nike Air Max 1.

This narrowed down our scope to specific categories (women, men, kids), encompassing all available colors and sizes on both Vinted and Vestiaire Collective.
The resulting concatened dataset includes the same informations such as title, price, brand, size, link, image source, localization, approximate carbon footprint, and profit made. This curated and pre-processed dataset will be exported to a CSV file and utilized in our upcoming Streamlit app.

Conclusion

In conclusion, our project stands at the intersection of sustainability, e-commerce, and technology, aiming to address the increasing demand for second-hand Nike sneakers. By leveraging web scraping techniques, we will curate a diverse and up-to-date inventory from popular platforms like Vinted and Vestiaire Collective.

The integration of a robust data processing pipeline, which includes price comparison, profit calculation, and an estimation of carbon footprint, adds value to users seeking sustainable and economically sound options.

The project's ethical foundation ensures responsible web scraping practices, respecting the terms of service of each platform. Our commitment to transparency in the absence of official information from Nike itself is reflected in the calculated carbon footprint and profit metrics. The development of a forthcoming Streamlit application promises to offer users a user-friendly interface for exploring Nike sneakers with tailored search options.

In an era where conscious consumerism is on the rise, our project aligns with the growing desire for sustainable alternatives in the e-commerce landscape. We envision our application becoming a valuable tool for individuals seeking not only affordable second-hand options but also a deeper understanding of the environmental impact associated with their choices.

BIBLIOGRAPHY

https://github.com/nilanyK/Webscrapping-DataProcessing

https://www.vinted.fr/

https://fr.vestiairecollective.com/

https://www.nike.com/fr/

https://www.pivr.fr/blog/les-francais-et-la-seconde-main-ou-lart-de-consommer-mieux-autrement

https://www.lemonde.fr/l-occasion-une-valeur-durable/article/2023/04/26/la-seconde-main-nouvelle-source-de-fierte_6171055_347.html

https://vietnamnet.vn/en/vietnam-becomes-major-manufacturer-of-nike-sports-shoes-807282.html

https://www.suffolkgazette.com/top-5-best-selling-nike-shoes-and-sneakers-of-all-time/

https://jobs.nike.com/fr/elc-2