

Clothing website: Scraping & Analysis

Michael Tavoni

6 gennaio 2026

Indice

1	Introduzione	2
2	Website scraping	2
3	Data analysis	2
3.1	Dataset	2
3.2	Analisi esplorativa	2

1 Introduzione

Il progetto è stato sviluppato nel rispetto dei vincoli contenuti all'interno del file `Robots.txt` del sito in analisi.

2 Website scraping

Per affrontare la sfida dello scraping del sito web, che per motivi di privacy è stato oscurato e mantenuto all'interno del file `.env` ho deciso di preparare un apposito pacchetto python con il quale recuperare in modo semplice e veloce il contenuto delle diverse pagine html che volevo indagare.

Il pacchetto è stato denominato: ‘htmlGrabber’ proprio perché ha come obiettivo quello di raccogliere gli html che contengono l'intero catalogo di abbigliamento del sito.

3 Data analysis

3.1 Dataset

Il dataset contiene 1266 righe e le seguenti 4 colonne:

1. **Brand** (categorica): variabile che contiene il brand del capo d'abbigliamento;
2. **item** (string): variabile che contiene una breve descrizione del capo d'abbigliamento;
3. **Category** (categorica): variabile che contiene la categoria del capo d'abbigliamento;
4. **Price** (quantitativa): variabile che contiene il prezzo del capo d'abbigliamento.

3.2 Analisi esplorativa

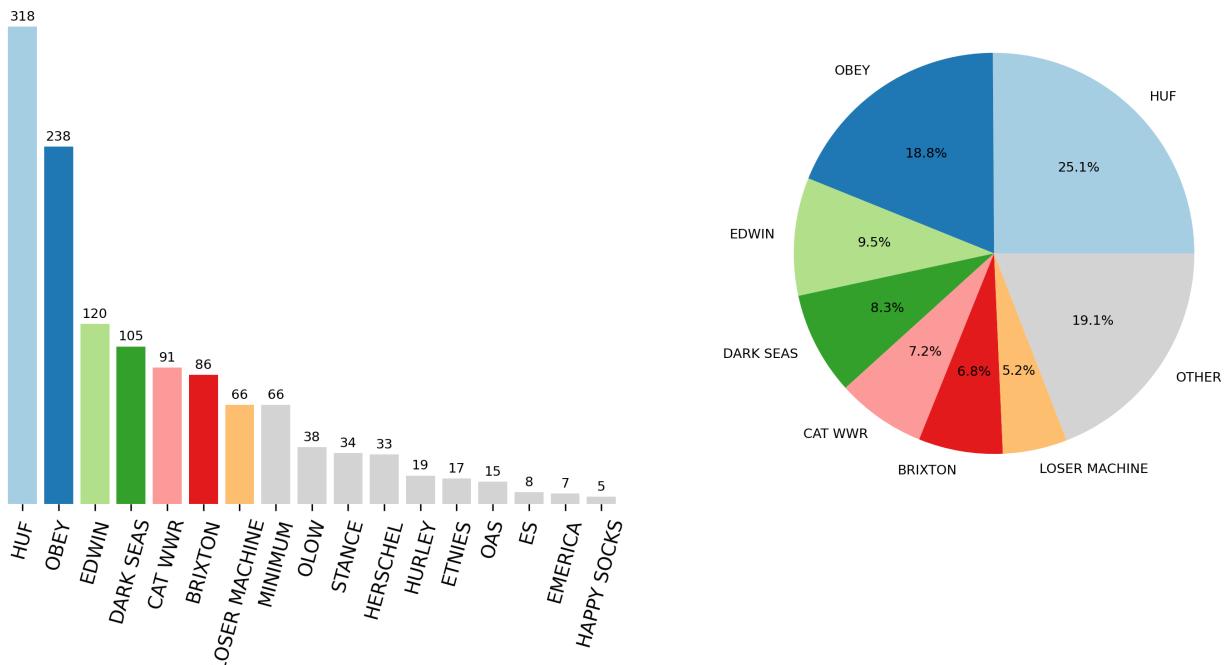


Figura 1: Brand

I 1266 prodotti presenti nel sito web sono solamente di 12 brand. In particolare, l'80% degli articoli appartiene a 7 brand, ovvero:

1. HUF
2. OBEY
3. EDWIN
4. DARK SEAS

5. CAT WWR
6. BRIXTON
7. LOSER MACHINE