



ZARA

Zara US fashion products

Indice

01

Descrizione del dataset

Download da Kaggle
del set di dati: zara_data

02

Fase di pre-processing

Esplorazione, pulizia e
trasformazione dei dati

03

Funzione di inserimento

Implementata per inserire
nuovi articoli nel database

04

Funzione di cancellazione

Usata per eliminare
un record dal database

05

Funzione di modifica

Implementata per modificare
un record nel database

06

Funzione di ricerca

Usata per cercare
uno o più record nel database

07

Query statistiche

Funzione usata per fare query
statistiche sui dati



01

**Descrizione
del dataset**

ZARA US fashion products dataset

ZARA US retails products dataset è un file in formato .json con una dimensione di 1.68MB che contiene diversi attributi riguardanti i prodotti venduti da ZARA negli Stati Uniti. Gli attributi includono informazioni dettagliate sui prodotti come nome, categoria, prezzo, descrizione, materiale, taglie disponibili e altre caratteristiche rilevanti.

```
_id: ObjectId('64b96af8bf078435d557f8f7')
url: "https://www.zara.com/us/en/floral-print-mini-dress-p02157021.html"
language: "en-US"
name: "FLORAL PRINT MINI DRESS"
sku: "118515176-084-1"
brand: "ZARA"
description: "V-neck dress with short sleeves. A-line silhouette. Front closure with..."
price: 45.9
availability: "InStock"
color: "Color Black / White | 2157/021"
size_list: "XS/S/M/L"
scraped_at: "2021-10-13 01:21:34"
```

Attributi del dataset



Id

L'id è un numero univoco assegnato ad ogni record



Url

L'url è un link che indirizza all'articolo presente sul sito



Language

Language indica la lingua associata al set di dati che nel dataset di riferimento è en-US



Name

Nella colonna name sono presenti i nomi degli articoli



Sku

Sku è un numero univoco utilizzato per identificare un articolo nell'inventario



Brand

Brand indica la marca di quell'articolo che nel caso specifico è ZARA o ZARAHOME

Attributi del dataset



Description

La colonna contiene la descrizione per ogni articolo



Price

Price corrisponde al prezzo di ogni articolo nel dataset



Available

Indica se l'articolo è presente o meno nel magazzino



Color

La colonna indica il colore associato ad ogni articolo



Size

Size corrisponde alla taglia degli articoli presenti nel database



Scraped

Viene utilizzato per registrare la data e l'ora in cui ciascun articolo è stato estratto o raccolto dal sito web di Zara

Il database Nosql utilizzato è **MongoDB** nel quale è stato inserito un database chiamato **ZARA** e all'interno è stata aggiunta una collection chiamata: **zata_data**

localhost:27017

Collections

+ Create collection

Refresh

View

Sort by

Collection Name

My Queries

Databases

ZARA

zara_data

zara_data

Storage size: 339.97 kB

Documents: 1.5 K

Avg. document size: 496.00 B

Indexes: 1

Total index size: 61.44 kB

zara_data						
	_id ObjectId	url String	language String	name String	sku String	brand String
1	ObjectId('64b96af8bf078435d55...	"https://www.zara.com/us/en/s...	"en-US"	"SATIN EFFECT CORSET BODYSUIT"	"128666521-966-2"	"ZARA"
2	ObjectId('64b96af8bf078435d55...	"https://www.zara.com/us/en/l...	"en-US"	"LONG PRINT DRESS"	"142448983-701-1"	"ZARA"
3	ObjectId('64b96af8bf078435d55...	"https://www.zara.com/us/en/s...	"en-US"	"STEEL KITCHEN SCISSORS"	"125067259-220-99"	"ZARAHOME"
4	ObjectId('64b96af8bf078435d55...	"https://www.zara.com/us/en/l...	"en-US"	"LINEN BLEND STRIPED ORGANIZE...	"135325938-737-99"	"ZARAHOME"
5	ObjectId('64b96af8bf078435d55...	"https://www.zara.com/us/en/c...	"en-US"	"CHIFFON ORGANIZER WITH MOONS...	"138551470-727-99"	"ZARAHOME"
6	ObjectId('64b96af8bf078435d55...	"https://www.zara.com/us/en/p...	"en-US"	"PLAID SHIRT DRESS"	"138550324-800-18"	"ZARA"
7	ObjectId('64b96af8bf078435d55...	"https://www.zara.com/us/en/c...	"en-US"	"CORDUROY CAP"	"131213214-807-3"	"ZARA"
8	ObjectId('64b96af8bf078435d55...	"https://www.zara.com/us/en/c...	"en-US"	"CLASSIC STRIPED SWIMSUIT"	"120689725-401-3"	"ZARA"
9	ObjectId('64b96af8bf078435d55...	"https://www.zara.com/us/en/f...	"en-US"	"Faux fur slippers"	"116014050-002-35"	"ZARAHOME"
10	ObjectId('64b96af8bf078435d55...	"https://www.zara.com/us/en/c...	"en-US"	"CROPPED T-SHIRT"	"128032853-800-1"	"ZARA"
11	ObjectId('64b96af8bf078435d55...	"https://www.zara.com/us/en/vi...	"en-US"	"W/END TILL 8.00 PM W/END TIL...	"135326155-999-99"	"ZARA"
12	ObjectId('64b96af8bf078435d55...	"https://www.zara.com/us/en/g...	"en-US"	"GEOMETRIC PRINT OVERSHIRT"	"142882821-427-2"	"ZARA"

A low-angle, black and white photograph of several skyscrapers reaching towards a cloudy sky. The perspective creates a sense of height and scale. The buildings are dark, with some windows reflecting light. The sky is a mix of dark and light grey clouds.

02

Fase di
pre-processing

Pre-processing

Query che converte i record presenti nella colonna prezzi da stringa a **Double**

```
pipeline = [{ "$set": { "price": { "$toDouble": "$price" }}}]
```

```
mycol.update_many({}, pipeline)
```

Query che converte le righe vuote presenti nella colonna availability con la stringa **"NotStock"**

```
query = { "availability": { "$eq": " " } }
```

```
new_values = { "$set": { "availability": "NotStock" } }
```

```
result = mycol.update_many(query, new_values)
```

In corrispondenza delle righe con attributo **"NotStock"** il campo size_list viene sostituito con **"Unavailable"**

```
query = { "availability": { "$regex": "NotStock", "$options": "i" }}
```

```
new_values = { "$set": { "size_list": "Unavailable" }}
```

```
result = mycol.update_many(query, new_values)
```

A low-angle, black and white photograph of several modern skyscrapers reaching towards a dark, cloudy sky. The perspective creates a sense of height and architectural scale. The buildings are framed by a thin white rectangular border.

03

Funzione
di inserimento

Funzione Insert

La **funzione di inserimento** ha lo scopo di aggiungere nuovi record nel database. È possibile compilare tutti i campi degli attributi o lasciarli vuoti, tranne per il campo "prezzo", che è obbligatorio per poter inserire un nuovo articolo.

```
def insert():
    global search_results, global_results
    try:
        r = messagebox.askyesno("Insert?", "Desideri inserire un record?")
        if r:
            new_data = {
                "url": curl.get(),
                "language": clanguage.get(),
                "name": cname.get(),
                "sku": csku.get(),
                "brand": cbrand.get(),
                "description": cdescription.get(),
                "price": float(cprice.get()),
                "availability": cavailability.get(),
                "color": ccolor.get(),
                "size_list": csize_list.get(),
                "scraped_at": cscraped_at.get()
            }
            new_id = mycol.insert_one(new_data).inserted_id
            new_record = mycol.find_one({"_id": new_id})
            global_results.append(new_record)
            if search_results:
                search_results.append(new_record)
            populate_table()

    except ValueError:
        messagebox.showerror("Errore", f"non puoi inserire un articolo senza inserire il prezzo")
```

Esempio pratico

ZARA

Id <input type="text"/>	Url <input type="text"/>	Language <input type="text"/>	Name <input type="text"/>
Sku <input type="text"/>	Brand <input type="text"/>	Description <input type="text"/>	Price <input type="text"/>
Availability <input type="text"/>	Color <input type="text"/>	Size_list <input type="text"/>	Scraped_at <input type="text"/>

InsertDeleteUpdateFindStatsClear

Index	_id	url	language	name	sku	brand	description	price	availability	Color
1	64b96af8bf078435	https://www.zara	en-US	SATIN EFFECT CO	128666521-966-2	ZARA	Bodysuit with swi	29.9	InStock	Color
2	64b96af8bf078435	https://www.zara	en-US	LONG PRINT DRE	142448983-701-1	ZARA	Long dress with n	49.9	NotStock	Color
3	64b96af8bf078435	https://www.zara	en-US	STEEL KITCHEN S	125067259-220-95	ZARAHOME	Forged stainless s	49.9	InStock	Color
4	64b96af8bf078435	https://www.zara	en-US	LINEN BLEND STF	135325938-737-95	ZARAHOME	Hanging organize	49.9	InStock	Color
5	64b96af8bf078435	https://www.zara	en-US	CHIFFON ORGAN	138551470-727-95	ZARAHOME	Hanging organize	35.9	InStock	Color
6	64b96af8bf078435	https://www.zara	en-US	PLAID SHIRT DRE	138550324-800-15	ZARA	Shirtdress with la	27.9	NotStock	Color



04

**Funzione di
cancellazione**

Funzione delete

La **funzione di cancellazione**

ha il preciso scopo di
rimuovere articoli dal
magazzino.

```
def delete():  
    r = messagebox.askyesno("Delete?", "Desideri cancellare un record?")  
    if r:  
        myquery = {"_id": ObjectId(cid.get())}  
        mycol.delete_one(myquery)  
  
        for i, result in enumerate(search_results):  
            if "_id" in result and result["_id"] == ObjectId(cid.get()):  
                del search_results[i]  
                break  
    populate_table()
```

Esempio pratico

ZARA

Id <input type="text"/>	Url <input type="text"/>	Language <input type="text"/>	Name <input type="text"/>
Sku <input type="text"/>	Brand <input type="text"/>	Description <input type="text"/>	Price <input type="text"/>
Availability <input type="text"/>	Color <input type="text"/>	Size_list <input type="text"/>	Scraped_at <input type="text"/>

Index	_id	url	language	name	sku	brand	description	price	availability	Color
1553	64c4ce46017e43d	https://www.zara	en-US	JACQUARD T-SHI	126358056-401-2	ZARA	Round neck top v	25.9	InStock	Colo
1554	64c4ce48017e43d	https://www.zara	en-US	JACQUARD T-SHI	126358056-401-2	ZARA	Round neck top v	25.9	InStock	Colo
1555	64c4ce4b017e43d	https://www.zara	en-US	JACQUARD T-SHI	126358056-401-2	ZARA	Round neck top v	25.9	InStock	Colo
1556	64c4e5a0acc6330	https://www.zara	en-US	JACQUARD T-SHI	126358056-401-2	ZARA	Round neck top v	25.9	InStock	Colo
1557	64c4e5a4acc6330	https://www.zara	en-US	JACQUARD T-SHI	126358056-401-2	ZARA	Round neck top v	25.9	InStock	Colo
1558	64c6222e59fe12e0			Dress				45.0		Mr



05

Funzione
di modifica

Funzione update

La **funzione di aggiornamento** permette di apportare modifiche a ciascun campo di input dei record esistenti. In altre parole, consente di modificare i valori degli attributi di articoli già presenti nel database.

```
def update():
    r = messagebox.askyesno("Update?", "Desideri modificare un record?")
    if r:
        myquery = {"_id": ObjectId(custid.get())}
        newvalues = {
            "$set": {
                "url": curl.get(),
                "language": clanguage.get(),
                "name": cname.get(),
                "sku": csku.get(),
                "brand": cbrand.get(),
                "description": cdescription.get(),
                "price": float(cprice.get()),
                "availability": cavailability.get(),
                "color": ccolor.get(),
                "size_list": csize_list.get(),
                "scraped_at": cscraped_at.get()
            }
        }
        mycol.update_one(myquery, newvalues)
    for i, result in enumerate(search_results):
        if "_id" in result and result["_id"] == ObjectId(custid.get()):
            search_results[i] = mycol.find_one({"_id": ObjectId(custid.get())})
            break
    populate_table()
```

Esempio pratico

ZARA

Id	Url	Language	Name
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Sku	Brand	Description	Price
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Availability	Color	Size_list	Scraped_at
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

InsertDeleteUpdateFindStatsClear

Index	_id	url	language	name	sku	brand	description	price	availability	Color
1553	64c4ce46017e43d	https://www.zara	en-US	JACQUARD T-SHI	126358056-401-2	ZARA	Round neck top v	25.9	InStock	Colo
1554	64c4ce48017e43d	https://www.zara	en-US	JACQUARD T-SHI	126358056-401-2	ZARA	Round neck top v	25.9	InStock	Colo
1555	64c4ce4b017e43d	https://www.zara	en-US	JACQUARD T-SHI	126358056-401-2	ZARA	Round neck top v	25.9	InStock	Colo
1556	64c4e5a0acc6330	https://www.zara	en-US	JACQUARD T-SHI	126358056-401-2	ZARA	Round neck top v	25.9	InStock	Colo
1557	64c4e5a4acc6330	https://www.zara	en-US	JACQUARD T-SHI	126358056-401-2	ZARA	Round neck top v	25.9	InStock	Colo
1558	64c625b82b86ff5k			Dress				45.0		Mi



06

Funzione
di ricerca

```

def find():
    global search_results
    r = messagebox.askyesno("Find?", "Desideri effettuare una ricerca?")
    if r:
        search_results.clear()
        query = {}
        url_term = curl.get().strip()
        if url_term:
            query["url"] = {"$regex": re.compile(r"\b" + re.escape(url_term) + r"\b", re.IGNORECASE)}
        language_term = clanguage.get().strip()
        if language_term:
            query["language"] = {"$regex": re.compile(r"\b" + re.escape(language_term) + r"\b", re.IGNORECASE)}
        name_term = cname.get().strip()
        if name_term:
            query["name"] = {"$regex": re.compile(r"\b" + re.escape(name_term) + r"\b", re.IGNORECASE)}
        ...
        price_input = cprice.get().strip()
        if price_input:
            query["price"] = {"$lt": float(price_input)}
        availability_term = cavailability.get().strip()
        if availability_term:
            query["availability"] = {"$regex": re.compile(r"\b" + re.escape(availability_term) + r"\b", re.IGNORECASE)}
        color_term = ccolor.get().strip()
        if color_term:
            query["color"] = {"$regex": re.compile(r"\b" + re.escape(color_term) + r"\b", re.IGNORECASE)}
        size_list_term = csize_list.get().strip()
        if size_list_term:
            query["size_list"] = {"$regex": re.compile(r"\b" + re.escape(size_list_term) + r"\b", re.IGNORECASE)}
        scraped_at_term = cscraped_at.get().strip()
        if scraped_at_term:
            query["scraped_at"] = {"$regex": re.compile(r"\b" + re.escape(scraped_at_term) + r"\b", re.IGNORECASE)}
        if not any(query.values()): # Se tutti i valori della query sono vuoti, mostra l'intera tabella
            populate_table()
        else:
            cursor = mycol.find(query).sort("price", pymongo.ASCENDING)
            for document in cursor:
                search_results.append(document)
            if not search_results:
                messagebox.showinfo("Ricerca completata", "Mi dispiace, ma la tua ricerca non ha prodotto risultati.")
            populate_table()

```

Funzione find

La **funzione find** ha lo scopo di eseguire ricerche nel database, consentendo di specificare qualsiasi valore per ciascun campo di input come filtro. In altre parole, la funzione **offre flessibilità** nell'effettuare ricerche dettagliate o di visualizzare l'intero contenuto del magazzino nel caso in cui non vengano specificati filtri. Se nessun articolo corrispondente ai criteri di ricerca viene trovato, verrà visualizzato un messaggio per informare l'utente che la ricerca non ha prodotto risultati.

Esempio pratico

ZARA

Id <input type="text"/>	Url <input type="text"/>	Language <input type="text"/>	Name <input type="text"/>
Sku <input type="text"/>	Brand <input type="text"/>	Description <input type="text"/>	Price <input type="text"/>
Availability <input type="text"/>	Color <input type="text"/>	Size_list <input type="text"/>	Scraped_at <input type="text"/>

Index	_id	url	language	name	sku	brand	description	price	availability	Color
1553	64c4ce46017e43d	https://www.zara	en-US	JACQUARD T-SHI	126358056-401-2	ZARA	Round neck top v	25.9	InStock	Colo
1554	64c4ce48017e43d	https://www.zara	en-US	JACQUARD T-SHI	126358056-401-2	ZARA	Round neck top v	25.9	InStock	Colo
1555	64c4ce4b017e43d	https://www.zara	en-US	JACQUARD T-SHI	126358056-401-2	ZARA	Round neck top v	25.9	InStock	Colo
1556	64c4e5a0acc6330	https://www.zara	en-US	JACQUARD T-SHI	126358056-401-2	ZARA	Round neck top v	25.9	InStock	Colo
1557	64c4e5a4acc6330	https://www.zara	en-US	JACQUARD T-SHI	126358056-401-2	ZARA	Round neck top v	25.9	InStock	Colo
1558	64c625b82b86ff5t			t-shirt				78.0		Ta



07

Query
statistiche

```
def find_complex():
    name_list_term = cname.get().strip()
    color_list_term = ccolor.get().strip()
    size_list_term = csize_list.get().strip()
    price_list_term = cprice.get().strip()
    query = {}

    if name_list_term:
        query["name"] = {"$regex": rf"\b{re.escape(name_list_term)}\b", "$options": "i"}
    if color_list_term:
        query["color"] = {"$regex": rf"\b{re.escape(color_list_term)}\b", "$options": "i"}
    if size_list_term:
        query["size_list"] = {"$regex": rf"\b{re.escape(size_list_term)}\b", "$options": "i"}
    if price_list_term:
        query["price"] = {"$lt": float(price_list_term)}

    aggregate_pipeline = [{"$match": query}, {"$group": {"_id": None, "sum_occurrences": {"$sum": 1},
                                                         "avg_price": {"$avg": "$price"}}}]

    size_list_aggregate = list(mycol.aggregate(aggregate_pipeline))
    tree.delete(*tree.get_children())
    tree.heading("#1", text="Name")
    tree.heading("#2", text="Color")
    tree.heading("#3", text="Size List")
    tree.heading("#4", text="Occurrences")
    tree.heading("#5", text="Average Price")
    for col_number in range(6, len(tree["columns"]) + 1):
        tree.heading("#{}".format(col_number), text="")
    for i, item in enumerate(size_list_aggregate, start=1):
        row_data = [
            str(name_list_term) if name_list_term else "N/A",
            str(color_list_term) if color_list_term else "N/A",
            str(size_list_term) if size_list_term else "N/A",
            str(item["sum_occurrences"]),
            f"${round(item['avg_price'], 2)}" if "avg_price" in item else "N/A"
        ]
    tree.insert(parent="", index="end", iid=i, text=i, values=row_data)
```

Query statistiche

Questa funzione esegue **query complesse** su un database e restituisce i risultati in una struttura tabellare, utilizzando i criteri di ricerca specificati dall'utente. I risultati ottenuti dalla query vengono scritti come righe nella tabella. Per ogni elemento trovato, vengono visualizzati il numero totale di occorrenze e il prezzo medio di tali elementi. Se la taglia, il nome o il colore non sono stati specificati, verrà mostrato "N/A" come valore corrispondente nella tabella.

Esempio pratico

ZARA

Id

Url

Language

Name

Sku

Brand

Description

Price

Availability

Color

Size_list

Scraped_at

Insert

Delete

Update

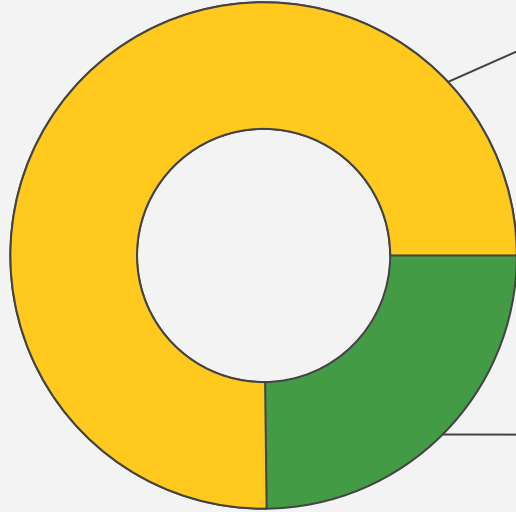
Find

Stats

Clear

language	name	sku	brand	description	price	availability	color	size_list	scraped_at
en-US	DRAPED LINEN BI	135326164-800-1	ZARA	Short dress made	25.99	InStock	Color Black 4437	XS/S/M/L/XL	2021-10-13 01:
en-US	MIDI DRESS WITH	112907088-500-1	ZARA	Sleeveless round i	29.9	InStock	Color Green 096	XS/S/M/L/XL	2021-10-13 01:
en-US	VOLUMINOUS SH-	122558330-401-2	ZARA	Short dress with r	35.9	InStock	Color Navy blue	S/M/L	2021-10-13 01:
en-US	FITTED MINI DRES	121397801-800-2	ZARA	Fitted mini dress	35.9	InStock	Color Black 1131	S/M/L	2021-10-13 01:
en-US	PLEATED DRESS	116368915-250-2	ZARA	Round neck short	35.9	InStock	Color White 026	S/M/L	2021-10-13 01:
en-US	DRAPED MIDI DRI	123657687-402-2	ZARA	Sleeveless round i	35.9	InStock	Color Turquoise	S/M/L	2021-10-13 01:

Framework



Team



Rosaria Leone



Giuseppe Genito

Grazie a tutti per
l'attenzione! ♥