# Web Scraper Titoli Borsa Italiana / Italian Stock Market Scraper

Script Python per scaricare dati storici dei titoli della Borsa Italiana tramite Yahoo Finance, con gestione di batch, retry automatici e barra di avanzamento.

Python script to download historical data of Italian Stock Market tickers via Yahoo Finance, with batch management, automatic retries, and a progress bar.

## 🔹 Caratteristiche principali / Key Features

* Scarica dati storici per **singoli titoli o interi indici** (default FTSE MIB).
* Download historical data for **single tickers or entire indices** (default FTSE MIB).
* Gestisce **rate limit** e retry automatici (tenacity).
* Handles **rate limits** and automatic retries (tenacity).
* Salva dati in file CSV **per ticker**, oltre a un file aggregato e statistiche riassuntive.
* Saves data in CSV files **per ticker**, plus an aggregated file and summary statistics.
* Barra di avanzamento visuale grazie a tqdm.
* Visual progress bar using tqdm.
* Modularità e leggibilità del codice per facilitare estensioni future.
* Modular and readable code for easy future extensions.

## 🔹 Requisiti / Requirements

Python 3.8+ e le seguenti librerie:

Python 3.8+ and the following libraries:

pip install pandas yfinance requests tenacity tqdm

Oppure con un requirements.txt / Or with a requirements.txt:

pandas  
yfinance  
tqdm  
tenacity  
requests

## 🔹 Uso / Usage

Esegui lo script direttamente nella cartella dove vuoi salvare i CSV:

Run the script directly in the folder where you want to save the CSVs:

python scraper.py

Opzioni disponibili / Available options:

| Argomento / Argument | Default | Descrizione / Description |
| --- | --- | --- |
| --index | FTSEMIB.MI | Ticker indice / Index ticker (FTSEMIB.MI, FTSESTAR.MI, etc.) |
| --period | 1y | Periodo storico / Historical period (1mo, 3mo, 6mo, 1y, 5y, max) |
| --interval | 1d | Intervallo dati / Data interval (1d, 1wk, 1mo) |
| --batch-size | 5 | Numero di ticker per batch / Number of tickers per batch |

Esempio / Example:

python scraper.py --index FTSEMIB.MI --period 5y --interval 1wk --batch-size 10

## 🔹 Output

Tutti i file vengono salvati nella **stessa cartella dello script**:

All files are saved in the **same folder as the script**:

* TICKER\_PERIOD\_INTERVAL.csv → dati storici per ogni ticker / historical data per ticker
* all\_history.csv → dati aggregati di tutti i ticker / aggregated data of all tickers
* summary\_stats.csv → statistiche riassuntive (min, max, mean, std) / summary statistics (min, max, mean, std) per ogni ticker / per ticker
* errors.log → eventuali errori durante il download / any errors during download

## 🔹 Struttura del codice / Code Structure

* parse\_args() → gestisce gli argomenti da linea di comando / handle command line arguments
* get\_index\_constituents() → ottiene i ticker di un indice, con fallback manuale / get tickers of an index, with manual fallback
* download\_batch() → scarica dati con retry automatici e gestione MultiIndex di yfinance / download data with automatic retry and MultiIndex handling
* main() → orchestratore: batching, download, salvataggio e statistiche / orchestrator: batching, download, saving, statistics

## 🔹 Best Practices

* Evita scraping aggressivo di siti con ToS restrittivi / Avoid aggressive scraping of sites with restrictive ToS
* Usa sempre batching e retry per limitare richieste e errori / Always use batching and retry to limit requests and errors
* Nome dei file sicuro per filesystem (. → \_) / Safe file names for filesystem (. → \_)

## 🔹 Estensioni possibili / Possible Extensions

* Integrazione con **notebook Jupyter** per visualizzare grafici dei prezzi / Jupyter notebook integration for price charts
* Supporto per altri indici e borse / Support for other indices and markets
* Salvataggio in **database SQL o NoSQL** invece che CSV / Save data in SQL or NoSQL databases instead of CSV
* Analisi avanzate dei dati (volatilità, correlazioni, grafici interattivi) / Advanced data analysis (volatility, correlations, interactive charts)

## 🔹 Licenza / License

MIT License – libero utilizzo e modifica / free to use and modify