

Foglio Esercizi n. 2 Itertools

Considerate nuovamente i dati delle partite dei campionati del mondo di calcio dal 1872 al 2020 disponibili gratuitamente sul sito Kaggle:

<https://www.kaggle.com/martj42/international-football-results-from-1872-to-2017>

Il file csv (con “,” come separatore) contiene le seguenti colonne:

date home_team away_team home_score away_score tournament city country neutral

Esempio:

1872-11-30,Scotland,England,0,0,Friendly,Glasgow,Scotland,FALSE

...

2020-12-09,United States,El Salvador,6,0,Friendly,Fort Lauderdale,US,FALSE

Usate generatori e librerie di itertools per realizzare una pipeline con valutazione lazy che realizza la seguente sequenza di trasformazioni e filtri:

- 1) Leggere i dati riga per riga (usando un generatore)
- 2) Convertire ogni record in un dizionario usando i nomi di colonna come nomi di attributi e unendo tutti i dizionari in una sola lista

Es. la riga sul match US-El Salvador del 2020-12-09 deve essere convertita in:

```
{'date': '2020-12-09',  
 'home_team': 'United States',  
 'away_team': 'El Salvador',  
 'home_score': '6',  
 'away_score': '0',  
 'tournament': 'Friendly',  
 'city': 'Fort Lauderdale', 'country': 'United States', 'neutral':  
 'FALSE'}
```

(hint: provate ad usare zip)

- 3) Elaborare i dizionari ottenuti al passo 2 per calcolare usando le librerie di itertools (e quindi ottimizzando l'uso della memoria) la somma totale di goal realizzati dall'Italia in tutte le partite della World Cup.
- 4) Usate memory_profiler (<https://pypi.org/project/memory-profiler/>) o strumenti simili per analizzare l'uso della memoria e tempi del programma risultante confrontandolo con una versione del programma senza generatori.

Come ulteriori esperimenti (non da consegnare) provare a selezionare la massima sequenze di vittorie, ecc.

Caricate su aulaweb nel link delle consegne:

- Sorgenti Python con lista di query e commenti nel codice in formato
 - .ipynb (notebook) oppure .py (codice sorgente)
 - PDF (versione leggibile in PDF)
- Eventuale file csv con i dati se caricati in locale dal notebook

La scadenza per la consegna è l'8 aprile 2021 alle ore 23:59