

## Homework #2 – InfluxDB

**Introduzione** - L'esercizio richiede di caricare in InfluxDB un dataset di dati meteo (temperature, pressioni, umidità) per diverse città (per ogni città sono riportate anche la nazione di appartenenza e le coordinate geografiche), raccolti con cadenza oraria dalle ore 12 del 01 ottobre 2012 alla mezzanotte del 30 novembre 2017. Eseguire il compito e rispondere a 5 domande analitiche utilizzando Python. Il dataset è quello disponibile a questo indirizzo: <https://www.kaggle.com/datasets/selfishgene/historical-hourly-weather-data>

**Descrizione del compito** - Scaricare il dataset e simulare in Python il funzionamento di N sensori (uno per ogni città e per ogni grandezza misurata) per caricare i dati dentro un database InfluxDB. Il periodo di rilevazione è rappresentato dalla colonna temporale "datetime". Durante il caricamento, aggiungere un *tag* per specificare la stagione meteorologica (Inverno, Primavera, Estate, Autunno), basata sulla data di registrazione.

Durante il salvataggio, impostare delle soglie di warning e di allarme sui valori di umidità e temperatura, secondo i criteri seguenti: (a) se l'umidità è superiore al 90%, emettere un segnale di pre-allerta (warning); se invece l'umidità supera il 98%, emettere un segnale di allerta; (b) se la temperatura è superiore a 45°C, emettere un segnale di allerta; (c) escludere direttamente durante la fase di caricamento i valori di temperatura non compresi tra -50 e 50°C; (d) escludere direttamente durante la fase di caricamento i valori di umidità non compresi tra 0 e 100%.

Eseguire le seguenti interrogazioni sul database appena popolato:

1. Calcolare la temperatura media per ogni città, raggruppata per stagione e visualizzarla in un grafico a barre.
2. Calcolare la variazione (massimo - minimo) di umidità giornaliera per ogni città, per un mese a scelta, e visualizzarla in un grafico a linee.
3. Trovare le giornate con il picco massimo di precipitazioni per ogni città nel corso dell'ultimo anno disponibile.
4. Calcolare la temperatura media mensile per ciascuna città, creare un grafico che mostra l'andamento delle temperature medie nel tempo e indicare le città con i trend più marcati (in crescita o in calo).
5. Calcolare per ogni città la media annuale dell'umidità e trovare quelle con i valori medi più alti e più bassi.

Ripetere le interrogazioni precedenti, però questa volta considerando solo i dati relativi a ciascuna stagione (Inverno, Primavera, Estate, Autunno). Documentare e confrontare le tempistiche richieste da InfluxDB in ciascun caso. Come si potrebbe migliorare l'efficienza delle query, mantenendo la possibilità di distinguere per stagione?

Applicando un algoritmo di clustering incrementale, verificare se in ogni nazione esistono dei valori di temperatura attorno a dei valori ricorrenti, differenziati per stagione, e se questi valori sono cambiati dal 2012 al 2017. Motivare la risposta.

**Dettagli sulla consegna** – Predisporre un notebook Python (o un file .py) in cui è riportato lo svolgimento di tutti gli esercizi e un file PDF in cui sono riportati i commenti sui risultati laddove esplicitamente richiesto. Includere tutti i file in uno zip e caricare l'archivio su Moodle.