# Report Dati Climatici

## GRUPPO COMICO:

MICHELA VENIER, LEANDRO TARANTINO, LUCA SILVA, FEDERICO VACIRCA

Come prima cosa abbiamo letto il file csv usando la libreria pandas, generando un dataframe con i dati ivi contenuti, chiamato “dati\_climatici”. Il dataframe contiene 1000 righe e 6 colonne:” data\_osservazione”,” temperatura\_media”, “precipitazioni”, “umidita”, “velocita\_vento” e “stazione\_metereologica”.

Dopodiché, abbiamo “pulito” il dataset rimuovendo i valori “Not a number” (Nan) ed i duplicati e rinominandolo “dati\_puliti”. Quindi abbiamo calcolato la media, la deviazione standard, il valore minimo e massimo, i quartili e lo Z-score delle colonne numeriche “temperatura\_media”, “precipitazioni”, “umidita” e “velocita\_vento”.

Dunque, per facilità di manipolazione dei dati normalizzati come Z-score, abbiamo creato un nuovo DataFrame chiamato “Z\_score\_df” contenente solamente i dati numerici, normalizzati sotto forma di Z-score.

Per facilitare la valutazione, abbiamo prodotto una serie di grafici dai dati.

In primo luogo abbiamo osservato la distribuzione delle variabili normalizzate usando degli istogrammi e dei box plot.

Dai grafici si osserva che la distribuzione, in termini di Z-score, è simile tra temperatura\_media, precipitazioni, umidita e velocita\_vento, e che i dati si distribuiscono omogeneamente tra -1.5 e 2.0 punti di deviazioni standard dalla media.

Per capire se vi fosse una correlazione tra le variabili numeriche, abbiamo calcolato la matrice di correlazione tra “temperatura\_media”, “precipitazioni”, “umidita” e “velocita\_vento”. I risultati indicano che non vi è alcuna correlazione significativa tra nessuna di queste variabili.

Queste analisi inducono a ipotizzare che il dataset non sia composto da dati ottenuti dal rilevamento di condizioni metereologiche realmente sussistite ma che invece sia stato realizzato utilizzando un algoritmo di generazione di numeri casuali, in quanto tra alcune delle variabili metereologiche sopra menzionate, come per precipitazioni e umidità, si osserva generalmente una correlazione marcata.