S7\_L5\_Luca\_Silvi\_BB Relazione su Dataset Climatico

Dalla relazione tra le colonne standardizzate con metodo ZScore e comparazione tra istogrammi e boxplot emerge che effettivamente il dato si può disporre in una distribuzione della variazione rispetto al periodo considerato.

Per ricercare inoltre la presenza di correlazione tra le variabili metereologiche comprese nel data\_set, ovvero temp\_media, precipitazioni o indice pluviometrico, umidità o tasso di umidità e velocità del vento, si osserva una heatmap della matrice di correlazione delle stesse.

Emerge che togliendo la diagonale centrale, che rappresenta la correlazione di una variabile con se stessa, si ottengono lievi correlazioni tra le variabili, sia negative che positive.

Ad esempio:

1. leggera correlazione + tra umidità e precipitazioni;
2. leggera correlazione – tra temperatura e umidità;
3. leggera correlazione + tra umidità e velocità del vento.

Prendendo poi quest’ultima correlazione emerge come il dato non sia affidabile, in quanto si pensa una correlazione positiva tra assenza di vento ed umidità, in quanto il vento spazza via o diminuisce il tasso locale di umidità.

Perciò infine si potrebbero ritenere poco significativi e prevedibili le correlazioni anche per le altre variabili.

Al chè suggerirei di costruire il campione di analisi su un periodo di tempo maggiore in anni, o comunque periodi aventi date legate tra loro da variabili che le assegnino maggiore significatività per le correlazioni suddette.

Ad esempio considerando un periodo di tempo quale una stagione.

Luca Silvi