

FIUME ARNO

1- plot 01 serve #per avere una visione iniziale del dataframe

MISSING

2- 02River_Arno_cut_missing seconda visualizzazione per avere una visione del dataframe dal 01-01-2004, ho tagliato i primi missing. Si notano notevoli valori mancanti sui valori di rainfall di alcune località' e sulla temperatura su Firenze.

Run del codice, si vede la tabella dei missing:

```
# A tibble: 17 x 3
  variable      n_miss pct_miss
  <chr>        <int>   <dbl>
1 Rainfall_Stia      4743  78.7
2 Rainfall_Consuma    4743  78.7
3 Rainfall_S_Savino   4743  78.7
4 Rainfall_Laterina   4743  78.7
5 Rainfall_Camaldoli   4743  78.7
6 Rainfall_Montevarchi 4379  72.7
7 Rainfall_Bibbiena   3648  60.5
8 Rainfall_Vernio     1743  28.9
9 Rainfall_Incisa     1458  24.2
10 Temperature_Firenze 1082  18.0
11 Hydrometry_Nave_di_Rosano 3  0.0498
12 Date              0  0
13 Rainfall_Le_Croci    0  0
14 Rainfall_Cavallina    0  0
15 Rainfall_S_Agata      0  0
16 Rainfall_Mangona      0  0
17 Rainfall_S_Piero      0  0
```

3- Grafici per visualizzare i valori mancanti. A titolo esemplificativo possiamo fare vedere i grafici 03 e 04 per sottolineare la pesante mancanza di dati di rainfall su molte località'. In totale le colonne con queste mancanze di rainfall sono 9, grafico 11_1. (Si puo' confrontare con il grafico 11_2 dove ci sono le località' con rainfall senza missing).

4- Aggiungo il grafico dei valori mancanti della temperatura di Firenze, grafico 12.

Per tutte queste variabili spiego che raccoglierò i dati reali da 3bmeteo.

5- Aggiungo il grafico dei valori mancanti della variabile target 13Hydrometry_Nave_di_Rosano_missing (grafico 13)

scarico i dati dal 2011 al 2020 su 3bmeteo

FILLING GAPS WITH METEO

6- A titolo esemplificativo posso riportare il codice di riempimento della temperatura di Firenze e della località' Bibbiena.

7-Grafico 14, riempiti i missing, rimangono delle date antecedenti al 2011 dove non e' possibile recuperare i dati.

8-Decido un nuovo cut sul mio dataframe. cutdata_2 <- as.Date("2011-01-01")
Il nuovo data frame e' River_Arno_cut1

9- Interpolo l'unico valore mancante del data frame, nella variabile target.

```
####interpolo il valore mancante####
```

```
# interpolation
```

```
#River_Arno_cut1$Hydrometry_Nave_di_Rosano
```

10- Grafico 15 ho il data frame pulito River_Arno_cut1

11- Grafico 16, osservo l'andamento dell'idrometria. Commento

```
#TARGET: Hydrometry_Nave_di_Rosano
```

```
#Il livello delle acque sotterranee rilevato dalla stazione idrometrica
```

```
#è stagionale, più elevato durante novembre-maggio.
```

```
#Ci sono dei picchi improvvisi verso lo zero, subito ripristinati
```

12- Matrice di correlazione, Grafico 17

Commento:

```
#### Correlation Matrix ####
```

```
#plot 17Correlation_Matrix.jpg
```

```
#La matrice di correlazione mostra una serie di correlazioni
```

```
#positive tra precipitazioni e correlazioni negative tra
```

```
# la temperatura e precipitazioni.
```

```
#L'idrometria e' inversamente correlata con la temperatura.
```

```
#Le precipitazioni possono essere suddivise in due categorie:
```

```
#la prima per le prime 6 variabili che sono le zone, attraverso cui
```

```
#passa il principale affluente dell'arno, il fiume Sive.(localita':
```

```
#Le_Croci, Cavallina, S_Agata, Mangona, S_Piero, Vernio).
```

```
#La seconda categoria e' rappresentata dalle restanti 8 localita'
```

```
#che si trovano lungo l'Arno a partire dalla fonte (localita':
```

```
#Stia, Camaldoli, Bibbiena, Laterina, S_Savino, Montevarchi,
```

```
#Incisa, Consuma).
```

```
#Le variabili sono fortemente correlate tra loro all'interno dei gruppi,
```

```
#ma molto debolmente tra i gruppi diversi, quindi per il modello finale,
```

```
#studieremo esempi di gruppi diversi.
```

13-#### visualizzo la pioggia per localita'#### da inserire il grafico numero 19

Commento:

```
#abbiamo due gruppi (affluente Sieve e sorgente dell'arno) in cui le precipitazioni sono correlate tra loro,
```

```
#ma non sono correlate tra i gruppi diversi.
```

```
#nella maggior parte dei casi non superano i 100 mm di pioggia
```

```
#e a volte, in certe localita', raramente superano anche i 50 mm.
```

```
#Decidiamo di scegliere due variabili la somma delle piogge tra sorgente e affluente Sieve
```

```
#
```

```
#Faro il confronto con il test, guardando la correlazione tra i gruppi pioggia
```

```
#Prendo una localita per gruppo per un test:
```

```
#dal primo gruppo, dell'affluente Sieve, scegliamo le precipitazioni da "Le Croci"
```

```
#e dal secondo gruppo, della sorgente dell'Arno, scelgo "Stia" per la posizione geografica centrale, e per
```

```
#alta correlazione tra le altre componenti del proprio gruppo.
```

```
#poi, per fare un terzo test di forecast, possiamo prendere
```

```
# due diverse localita' come esempio Cavallina (dal gruppo dell'Affluente Sieve)
```

```
# e Bibbiena per il gruppo della sorgente dell'Arno.
```

14- Prendo il grafico 20 per la temperatura:

Commento

```
#La temperatura della regione di Firenze è l'unica variabile
```

```
#sulla temperatura del dataset Arno.
```

```
#Ha una distribuzione stagionale e la maggior parte dei valori è
```

```
#compresa tra 0 e 30 gradi, in linea per una regione del centro Italia.
```

15- Grafico 21 analizzo la variabile target con le stagioni, outliers

Commento:

#valori superiori a 5, che rappresentano piogge intense,

#si vedono solo in autunno e in inverno.

#le stagioni spiegano poco la variabile target: la mediana e' leggerment

piu' alta in primavera e in inverno.

16- Grafico 16 ### andamento della variabile target####

Commento

Livello idrometrico: indica l'altezza d'acqua del fiume rispetto a un riferimento fisso,

#denominato zero idrometrico (m).

#La variabile target esplicita il livello del fiume espresso in metri

#misurato dalla stazione idrometrica "Nave_di_Rosano".

#I valori di questa variabile sono compresi tra 0 e 7 metri.

#La maggioranza dei valori misurati è compresa tra 1 e 4 metri.

#Dopo qualche stagione irregolare, ci sono diversi cali

#fino a 0 nel 2013, 2014 e 2019, che tornano rapidamente alla normalità.

17- Oltre ai lag previsti come giorni e nuove variabili di pioggia, per il fiume, creo nuovi lag, in base alla distanza dal punto dell'idrometria Nave di ROsano.

Commento

nuova analisi Arno + LAG +####

#divisione delle zone delle rainfall in base alla distanza da Rosano (FI)

#

Studiando la geografia e la disposizione idrogeologica delle localita"

divido in gruppi le zone:

Gruppo 1: Incisa(fiume arno), Consuma (fiume arno), Montevarchi (fiume arno),

#S. Plero (affluente Sieve)

non imposto lag perche' hanno una distanza inferiore ai 50km dal

#punto idrometrico Nave di ROsano

Gruppo 2: Laterina (fiume arno), S. Agata (affluente Sieve),

Le Croci (affluente Sieve), Cavallina (affluente Sieve), Mangona inserisco un lag di un giorno

perche' hanno una distanza fino a circa 100 km dal punto idrometrico

Per le localita' di distanza superiore ai 100km: Stia, Camaldoli, Bibbiena, S. Avinro (per il fiume arno),

e Vernio (affluente Sieve) sinserisco un lag di 3 giorni

Possiamo fare vedere una parte del codice:

```
arno_LAG <- River_Arno_cut1 %>%
```

```
  mutate(lag_Laterina = lag(Rainfall_Laterina, +1),
```

```
         lag_Cavallina = lag(Rainfall_Cavallina, +1),
```

```
         lag_Le_Croci = lag(Rainfall_Le_Croci, +1),
```

```
         lag_S_Agata = lag(Rainfall_S_Agata, +1),
```

```
         lag_Mangona = lag(Rainfall_Mangona, +1),
```

```
         lag_Stia = lag(Rainfall_Stia, +2),
```

```
         lag_Camaldoli = lag(Rainfall_Camaldoli, +2),
```

```
         lag_Bibbiena= lag(Rainfall_Bibbiena, +2),
```

```
         lag_S_Savino = lag(Rainfall_S_Savino, +2),
```

```
         lag_Vernio= lag(Rainfall_Vernio, +2),
```

```
  )
```

18- #### RANdom Forest test ####

COMMENT: SPIEGO DI aver fatto vari test:

test 1 con localita Le_Croci e Stia #### RMSE Test: 0.48 le croci e Stia RF

test 2 con localita Cavallina e Bibbiena #### RMSE Test : 0.48 Cavallina e Bibbiena RF

test con nuovi lag ####con la disatnza####RMSE Test: 0.37 con arno_LAG

Scelgo questo ultimo plot 25

19- Gboost ##### GRADIENT BOOST MACHINE ##### anche qui sono stati fatti vari test, i migliori sono##### RMSE 0.3385 Test con LAG PERSONALIZZATI #####
test GB con LAG personalizzati #####
Grafici 26 e 27

Diverso test gb

testgb 2 con localita Le_Croci Cavallina per il primo gruppo e Bibbiena Stia per il secondo gruppo ##### RMSE 0.3423 Test2 con Le_Croci, Cavallina, Bibbiena, Stia #####
Grafici 28 e 29