Laboratorio finale "Data Analysis modulo 1"

Usando i dati nel file csv "infarto" allegato a questo testo svolgere i seguenti punti:

1. Caricare tutti i dati in un unico dataframe detto "TUTTI_0".
2. Creare un nuovo dataframe detto TUTTI con le sole colonne: "gender","age","bmi","avg_glucose_level","smoking status","stroke" ed elimando tutti i record che presentino qualche valore non definito in tali colonne.
3. Far stampare a console il sommario statistico di TUTTI
4. creare tre figure, ciascuna con tre subplot (uno per ciascun dataframe) che mostrino rispettivamente gli istogrammi dell'"age", del "bmi" e dell'"avg_glucose_level".
5. Usando TUTTI calcolare gli indici di correlazione di PEARSON tra le variabili numeriche
6. Formare un nuovo dataframe chiamato Z con due colonne: bmi e glucosio estratti dal dataframe TUTTI ma trasformati con il metodo Z-score.
7. Far stampare a console il sommario statistico di Z.
8. Produrre uno scatterplot di Z
9. Calcolare un modello di regressione lineare che leghi tra loro le variabili di Z
10. Formare due nuovi dataframe Z1 e Z0 estratti da Z. Z1 contiene i record dei pazienti con stroke=1, e Z1 i record dei pazienti con stroke=0.
11. Produrre scatterplot di Z0 e Z1
12. Calcolare un modello di regressione lineare che leghi tra loro le variabili di Z0
13. Calcolare un modello di regressione lineare che leghi tra loro le variabili di Z1