



DATA SCIENCE, BUSINESS ANALYTICS E INNOVAZIONE
METODI DI APPRENDIMENTO STATISTICO PER IL DATA SCIENCE

Homework 3

Si utilizzi il seguente formato, seguendo l'ordine delle istruzioni, per produrre uno script che svolga le funzionalità richieste.

```
##### Homework 3 #####  
# Nome:  
# Cognome:  
# Matricola:  
# CdL:  
#####  
##### 1 #####  
...your code here...  
  
##### 2 #####  
...your code here...  
  
##### Homework 3 #####
```

N.B. Il file dovrà essere rinominato nel seguente modo: **00000_Nome_Cognome.R**

1. Si scarichi il pacchetto **alr3** e lo si carichi in memoria
2. Si visualizzi l'help del dataset contenuto nel pacchetto alr3: **BigMac2003**
3. **Eccezion fatta per la prima, la quarta e la nona**, si rinominino le colonne in funzione delle informazioni ottenute sulle stesse (dall'help, da altre fonti o da conoscenze pregresse)
4. Si producano le statistiche generali per le singole colonne del dataset

5. Si analizzi l'output del punto precedente e si cerchino **eventuali valori anomali** provvedendo, eventualmente, ad eliminare dal dataset la/le osservazione/i anomala/e e dunque a **produrre nuovamente** le statistiche generali per le singole colonne del dataset
6. Si producano i grafici per ciascuna colonna, **scegliendo la tipologia più opportuna** per il tipo di dato presente (**l'uso di ggplot2 è un plus**)
7. Si trasformi la colonna **TaxRate** da quantitativa a qualitativa in funzione della media della stessa in un factor a due livelli: "**Basso**" (se inferiore), "**Alto**" (se pari o superiore)
8. Si produca un **nuovo grafico per TaxRate**, scegliendo la **tipologia più opportuna** per i dati qualitativi che TaxRate ora contiene
9. Si produca la matrice di correlazione lineare del dataset (si suggerisce l'uso di **corrplot**)
10. Si produca una **selezione di grafici** che mettano a confronto **BigMac** ed un'altra **variabile**, associando a ciascuno di essi una piccola **giustificazione per la scelta** effettuata.