

## Elaborato R

L'elaborato prevede la realizzazione di uno script in R con lo scopo di fare alcune elaborazioni di dati presenti su un dataset. Il dataset ha una struttura matriciale in cui ogni riga rappresenta un particolare distretto della città presa in considerazione e le colonne rappresentano rispettivamente i seguenti dati:

- Zip Code
- Popolazione totale
- Età media
- Popolazione totale maschile
- Popolazione totale femminile
- Nuclei familiari totali
- Dimensione media dei nuclei familiari

Inizialmente è stata letta la struttura del dataset dal file "2010\_Census\_Populations\_by\_Zip\_Code.csv" fornito.

Successivamente, è stata calcolata la popolazione totale della città sommando la popolazione totale di ciascun distretto. Tale valore è stato poi stampato in console.

In seguito, sono state calcolate le percentuali di uomini e donne per ciascun distretto e queste sono state inserite in una matrice con le seguenti colonne:

- ZipCode
- MalesPercent
- FemalesPercent

Per fare questo, è stata utilizzata la funzione "apply" come segue:

```
matrix = as.matrix(t(apply(cityCensus,1,percentMaleFemale))) #conversion from list to matrix
dimnames(matrix)[[2]]<-c("ZipCode","MalesPercent","FemalesPercent") #renaming of matrix columns
print(matrix)
```

Essa applica a ciascuna riga del dataframe cityCensus la funzione "percentMaleFemale". Tale funzione si occupa di calcolare le percentuali richieste e aggiungerle alla lista dei risultati forniti da "apply". Poiché il testo richiede che i risultati siano contenuti in una matrice, è stata utilizzata la funzione di conversione "as.matrix". Inoltre, è stata calcolata la trasposta mediante la funzione "t", per far sì che i risultati vengano visualizzati nella forma richiesta.

Infine, sono stati estratti due subset dal dataframe, secondo le condizioni richieste dal testo mediante l'utilizzo della funzione "subset". In tale funzione è necessario specificare tre campi:

- Dataframe
- Condizione di estrazione mediante la keyword **subset**
- Colonne (o righe) da estrarre mediante la keyword **select**

Il primo subset è stato salvato sul file "Output3.txt" mediante l'uso della funzione "write.table", mentre il secondo viene visualizzato in console.

Giuseppa Reitano

Giuseppe Pennisi