Per evitare la replica dei dati dovuti alla JOIN ho effettuato un altro FOREACH

IN PARTICOLARE:

**naz\_reg = JOIN nazione1 by data, regione1 by data**

ho definito la nuova variabile:

**clean\_naz\_reg = FOREACH naz\_reg GENERATE $0, $1, $2, $3, $4, $5, $7**

per evitare la duplicazione.

$ mi permette di riferirmi alle colonne di interesse poiché la join oltre a replicare la colonna ‘data’ mi fa perdere i nomi delle colonne.

Sto lavorando per capire il perché di questa scomparsa dei nomi delle colonne.