**Pulizia dei dati**

Per eseguire l’esercizio ho avuto come prima difficoltà la circostanza che i numeri riferiti alla popolazione non erano in formato numero interi, forzando il tipo di dato l’editor di power query mi dava errore su tutta la colonna, probabilmente per come questi numeri erano scritti; per ovviare a ciò ho usato una colonna ad esempi ed ho riscritto i numeri per poi convertirli.

Avuto poi un problema simile con il dataset relativo agli omicidi ho risolto allo stesso modo.

Arrivato a fare il merge della query mi sono accorto della presenza di alcuni campi nulli, dovuti al fatto che i nomi degli ststi americani nei due dataset non coincidono. Ho rifatto da capo la merge assicurandomi che i nomi degli stati coincidessero con “sostituisci valori”.

Il dataset relativo agli omicidi non presenta dati per alcuni stati come Porto Rico o le Samoa Americane, gli stati che non sono presenti nel suddetto dataset verranno eliminati dalla mia analisi.

Adesso la mia merge di query si è formata con 51 collegamenti su 51 e non abbiamo campi vuoti.

Al punto 5) mi sono importato il GPO che secondo me tra le varie abbreviazioni presenti nella tabella questa era la più significativa; insieme a questa colonna mi sono importato anche quelle relative allo stato ed alla sigla (dopo averle rinominate), nel caso mi servissero più avanti per eventuali merge e relazioni.

Nella suddetta query ho dovuto anche eliminare ben 26 righe di modo da avere solo i 51 stati che avevo cominciato ad analizzare.

Creando la tabella al punto 7) mi sono accorto che i nomi degli stati non corrispondevano tra il dataset Popolazione e Abbreviazioni, ho messo in luce i campi da sostituire facendo una merge fra i suddetti dataset anche se non veniva chiesta dalla traccia.

Sempre su questo punto dell’esercizio ho notato la stessa cosa andando ad inserire le capitali nella mia tabella, accertatomi che esistesse la relazione tra il dataset “capitali” e “popolazione” passo a controllare che il nome degli stati (chiave primaria) sia scritto ugualmente nei due dataset.

Ho fatto una nuova merge non richiesta.

Nella pulizia di “capitali” ho dovuto anche utilizzare il comando “taglio” per eliminare spazi superflui.

Mi accorgo che in “capitali” non abbiamo la capitale del Distretto di Columbia che provvedo a cercare su internet e ad aggiungere manualmente.

Ho scoperto che in america esiste lo stato di washington con capitale olympia e la città di washington dc capitale del distretto di columbia e di tutti gli stati uniti 😊 sta cosa mi ha fatto perdere tempo.

Non avendo le informazioni relative al distretto di columbia su ”capitali” per far trovare l’oggetto vsivo table ho preso le capitali dalla merge di query in popolazione.