Preferenze alimentari in SQL

Assignment week 5

Riprendiamo l’assignment sulle preferenze alimentari e svolgiamo una serie di task in SQL.

# Prerequisiti

Aver creato la tabella con le preferenze alimentari su DBeaver. ([ddl\_e\_dati](https://drive.google.com/file/d/1sHqSdYp-J_EGT4tKwXYbia72wqajH8y4/view?usp=drive_link))

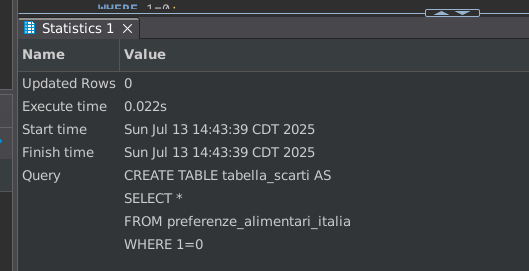
# Task

1. Creare una nuova tabella chiamata “tabella\_scarti” e inserirgli tutte le righe che:
   * Hanno ID NULL
   * Hanno ID Duplicati: conservare solo il MAX ROWID
   * Hanno Eta: NULL o >120, <10
   * Hanno Citta: NULL

– creo nuova tabella  
**CREATE** **TABLE** tabella\_scarti **AS**

**SELECT** \*   
 **FROM** preferenze\_alimentari\_italia

**WHERE** 1=0;



WHERE scritto cosi’ consente di prndere tutte le colonne e copiarle

**CREATE** **TABLE** tabella\_scarti **AS**

**SELECT** \*

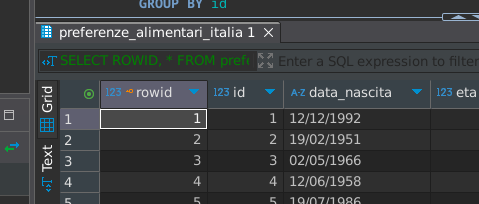
**FROM** preferenze\_alimentari\_italia

**WHERE** 1=0;

--trovo i dati da scartare nella tabella scarti con ROWID

**SELECT** ROWID, \* **FROM** preferenze\_alimentari\_italia **LIMIT** 5;

-- verifico che esista ROWID



**SELECT** \*

**FROM** preferenze\_alimentari\_italia

**WHERE** id **IN** (

**SELECT** id

**FROM** preferenze\_alimentari\_italia

**GROUP** **BY** id

**HAVING** **COUNT**(\*) > 1

-- questo trova gli ID con con piu' un count. almeno spero :)

)

**AND** ROWID **NOT** **IN** (

-- ho visto che IN ha altre funzioni e qui gli dico di cercare tutti gli ID che NON SONO il massimo

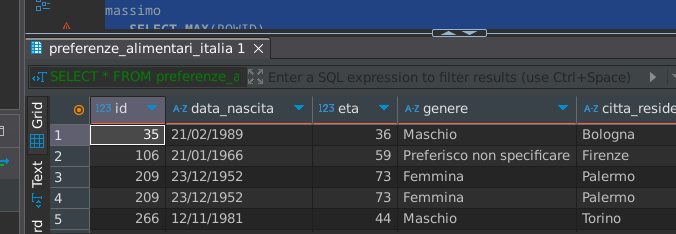
**SELECT** **MAX**(ROWID)

**FROM** preferenze\_alimentari\_italia

**GROUP** **BY** id

**HAVING** **COUNT**(\*) > 1

);



-- ora inserisco i dati

**INSERT** **INTO** tabella\_scarti

**SELECT** \*

**FROM** preferenze\_alimentari\_italia

**WHERE** id **IS** **NULL** **OR** (id **IN** (

**SELECT** id **FROM** preferenze\_alimentari\_italia

**GROUP** **BY** id

**HAVING** **COUNT**(\*) > 1

) -- stavo impazzendo con le parentesi ed ho scperto adesso che ti fa vedere la paratensi prima se ci clicchi sopra

**AND** ROWID **NOT** **IN** ( **SELECT**

**MAX**(ROWID)

**FROM** preferenze\_alimentari\_italia

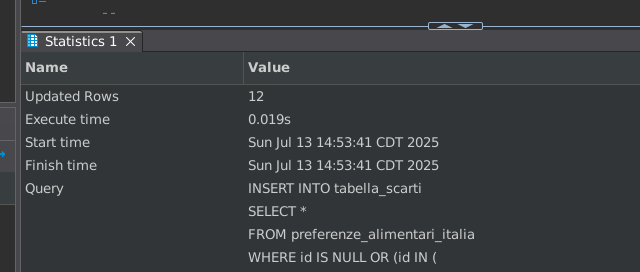
**GROUP** **BY** id

**HAVING** **COUNT**(\*) > 1 ))

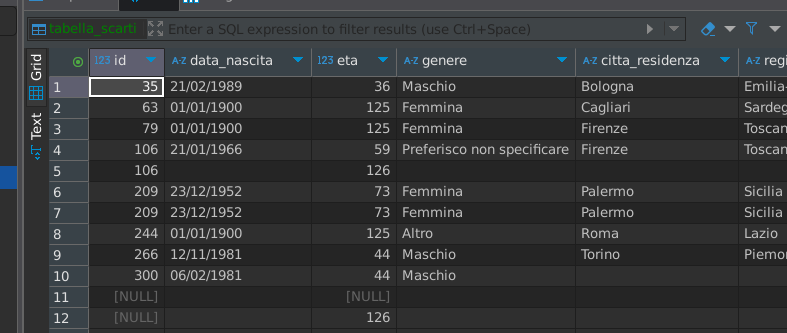
**OR** eta **IS** **NULL** **OR** eta > 120 **OR** eta < 10

**OR** citta\_residenza **IS** **NULL** **OR** citta\_residenza = **''**;

–



--



1. Contare il numero di righe della tabella “preferenze\_alimentari” e della tabella degli scarti “tabella\_scarti”

**SELECT**

'preferenze\_alimenatari\_italia' **AS** *tabella*, --qui creo il dato preferenza dentro una tabella temporanea tipo

**COUNT**(\*) **AS** *conteggio\_totale*

**FROM** preferenze\_alimentari\_italia

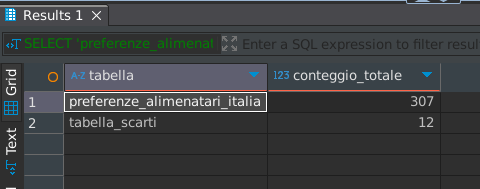
**UNION** **ALL** --ho deciso di usare UNION ALL perhce' avevo capito di dobessero unire due select e count perche' lo conoscevo

**SELECT**

'tabella\_scarti' **AS** *tabella*, -- qui inserisco direttamente la seconda riga nella colonna

**COUNT**(\*)

**FROM** tabella\_scarti;



1. Cancellare le righe del punto 1 dalla tabella “preferenze\_alimentari” del dataset delle preferenze alimentari  
   Qui ho avuto molta difficolta’ ed ho cercato online non sapendo che ROWID esiste ed e’un valore che biosnga bloccare per creare la tabella nuova e quindi ho rifattp

--siccome qui non ho capito ninete mi sono rguardato tutti i dati ma non ho capito se mantiene il numero delle rowid

**SELECT** ROWID, \* **from** preferenze\_alimentari\_italia pai

**WHERE** ID=35

**OR**

ID=63

**OR**

id=79

**or**

ID=79

**OR**

id=106

**OR**

ID=209

**OR**

ID=244

**OR**

ID=266

**OR**

ID=300

**OR** ID **IS** **NULL**

--siccome qui non ne uscivo ho deciso di cercare online ed ho scoperto che si puo mantenere il raw id

-- quindi droppo la tabella e rifaccio tutto

**DROP** **TABLE** **IF** **EXISTS** tabella\_scarti;

**CREATE** **TABLE** tabella\_scarti **AS**

**SELECT** ROWID **AS** rowid\_originale, \*

**FROM** preferenze\_alimentari\_italia

**WHERE**

--uno

id **IS** **NULL**

-- due

**OR** eta **IS** **NULL** **OR** eta > 120 **OR** eta < 10

--tre

**OR** citta\_residenza **IS** **NULL** **OR** citta\_residenza = **''**

--quattro

**OR** (id **IN** (

**SELECT**

id **FROM** preferenze\_alimentari\_italia

**GROUP** **BY**

id **HAVING** **COUNT**(\*) > 1

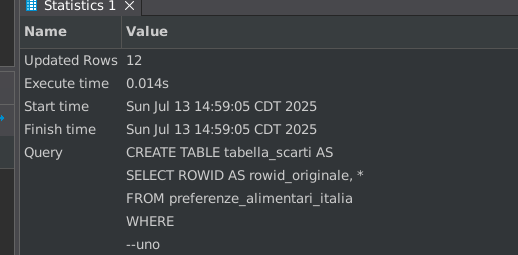
) --qui gli dico di cercare id che hanno conto superiore a 1

**AND** ROWID **NOT** **IN** (

**SELECT** **MAX**(ROWID) **FROM** preferenze\_alimentari\_italia

**GROUP** **BY** id **HAVING** **COUNT**(\*) > 1

));

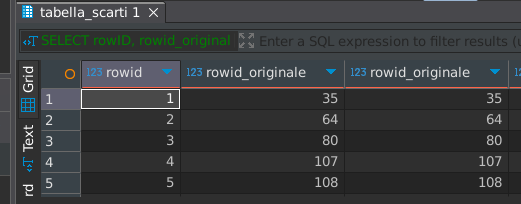


--riprovo a vedere rowid

**SELECT** rowID, rowid\_originale, \*

**FROM** tabella\_scarti ts

–



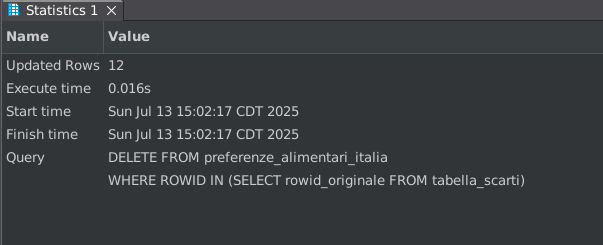
--ora posso cancellare tranquillo spero

--

**DELETE** **FROM** preferenze\_alimentari\_italia

**WHERE** ROWID **IN** (**SELECT** rowid\_originale **FROM** tabella\_scarti);

–



1. Verificare che il numero totale di righe dalla tabella “preferenze\_alimentari” dopo il punto 3 sia uguale alla differenza tra il conteggio ottenuto al punto 2 e il conteggio delle righe della tabella degli scarti

--

--Verificare che il numero totale di righe dalla tabella “preferenze\_alimentari” dopo il punto 3 sia uguale alla differenza tra il conteggio ottenuto al punto 2 e il conteggio delle righe della tabella degli scarti

**SELECT**

**CASE WHEN**

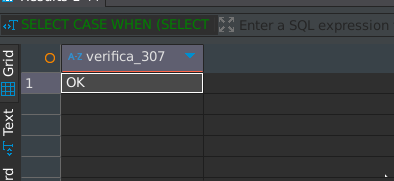
**(SELECT COUNT(\*) FROM preferenze\_alimentari\_italia) +**

**(SELECT COUNT(\*) FROM tabella\_scarti) = 307**

**THEN 'OK'**

**ELSE 'ERRORE'**

**END AS *verifica\_307*;**



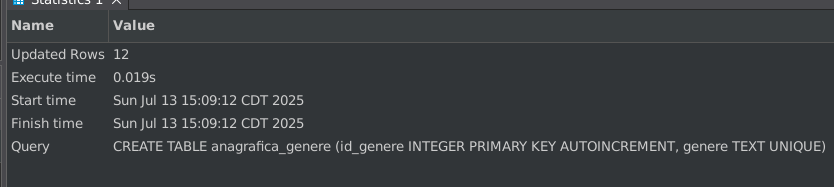
1. Creare le seguenti anagrafiche esterne (ricorda che una tabella di anagrafica contiene i valori distinti della colonna selezionata e per ogni valore viene associato un id univoco):
   * Genere
   * Regione
   * Titolo studio
   * Città collegata a regione:
     + Per creare questa anagrafica dovete raggruppare il dataset originale “preferenze\_alimentari” per città e regione e considerare per ogni città la sola regione che compare più volte (MAX)

Non capisco perche’ si fa questa cosa ma ho eseguito creando delle tabelle con degli id univoci

Usando distinct

**SELECT** **DISTINCT** genere **FROM** preferenze\_alimentari\_italia *pai*;

**CREATE** **TABLE** anagrafica\_genere (id\_genere **INTEGER** **PRIMARY** **KEY** **AUTOINCREMENT**, genere **TEXT** **UNIQUE**);



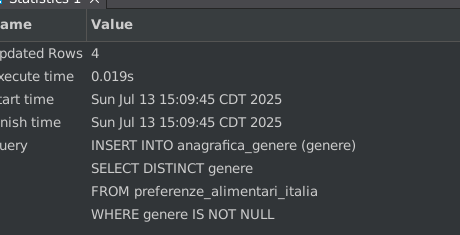
--popolo la tabella

**INSERT** **INTO** anagrafica\_genere (genere)

**SELECT** **DISTINCT** genere

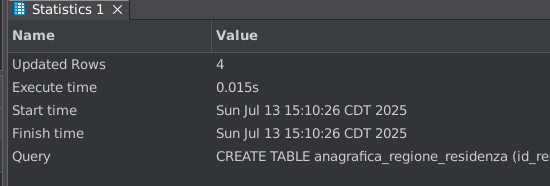
**FROM** preferenze\_alimentari\_italia

**WHERE** genere **IS** **NOT** **NULL**;



**SELECT** **distinct** regione\_residenza **FROM** preferenze\_alimentari\_italia pai;

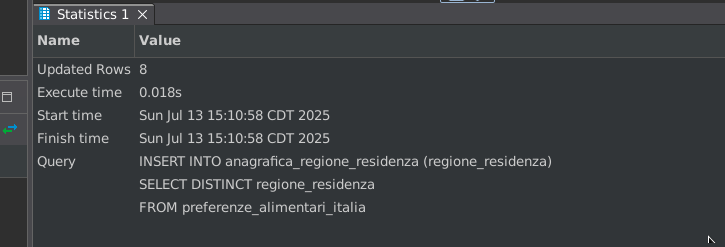
**CREATE** **TABLE** anagrafica\_regione\_residenza (id\_residenza **INTEGER** **PRIMARY** **KEY** **AUTOINCREMENT**, regione\_residenza **TEXT** **UNIQUE**) --ho avuto un idea qui... ma sto creando il dominio? non so perche' dovrei farlo



**INSERT** **INTO** anagrafica\_regione\_residenza (regione\_residenza)

**SELECT** **DISTINCT** regione\_residenza

**FROM** preferenze\_alimentari\_italia;



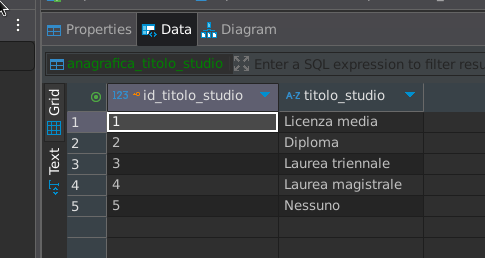
**SELECT** **distinct** titolo\_studio **FROM** preferenze\_alimentari\_italia pai;

**CREATE** **TABLE** anagrafica\_titolo\_studio (id\_titolo\_studio **INTEGER** **PRIMARY** **KEY** **AUTOINCREMENT**, titolo\_studio **TEXT** **UNIQUE**);

**INSERT** **INTO** anagrafica\_titolo\_studio (titolo\_studio)

**SELECT** **DISTINCT** titolo\_studio

**FROM** preferenze\_alimentari\_italia;



* + Città collegata a regione:
    - Per creare questa anagrafica dovete raggruppare il dataset originale “preferenze\_alimentari” per città e regione e considerare per ogni città la sola regione che compare più volte (MAX)

**CREATE** **TABLE** anagrafica\_citta (

id\_citta **INTEGER** **PRIMARY** **KEY** **AUTOINCREMENT**,

citta **TEXT** **UNIQUE**,

regione ); --qui ripeto tutto quello che ho fatto sopra

**INSERT** **INTO** anagrafica\_citta (citta, regione)

**SELECT**

citta\_residenza,

regione\_residenza

**FROM** (

**SELECT**

citta\_residenza,

regione\_residenza,

**COUNT**(\*) **AS** *conteggio*

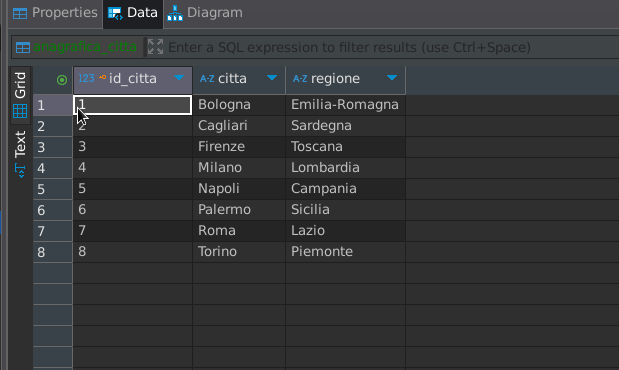
**FROM** preferenze\_alimentari\_italia

**GROUP** **BY** citta\_residenza, regione\_residenza --qui prendo solo la prima riga visto che e' ordinato in maniera desc

**ORDER** **BY** citta\_residenza, *conteggio* **DESC**

)

**GROUP** **BY** citta\_residenza;



1. Creare anagrafica classe età.   
   Per farlo create una nuova colonna classe di età nella tabella “preferenze\_alimentari” con i seguenti valori e criteri:

| **Valore** | **Criterio** |
| --- | --- |
| “<=20” | Età minore uguale di 20 |
| “21-30” | Età compresa tra 21 e 30 anni |
| “31-40” | Età compresa tra 31 e 40 anni |
| “41-50” | Età compresa tra 41 e 50 anni |
| …(continuate la sequenza) | … |
| >80 | Età maggiore di 80 |

Dopo aver creato l’anagrafica arricchirla aggiungendo le seguenti colonne

* + Età Inizio
  + Età Fine
  + Descrizione fascia

**SELECT** ID, eta,

**CASE**

**WHEN** eta < 20 **THEN** **'<20'**

**WHEN** eta >= 20 **AND** eta < 30 **THEN** **'20-29'**

**WHEN** eta >= 30 **AND** eta < 40 **THEN** **'30-39'**

**WHEN** eta >= 40 **AND** eta < 50 **THEN** **'40-49'**

**WHEN** eta >= 50 **AND** eta < 60 **THEN** **'50-59'**

**WHEN** eta >= 60 **AND** eta < 70 **THEN** **'60-69'**

**WHEN** eta >= 70 **AND** eta < 80 **THEN** **'70-79'**

**WHEN** eta >= 80 **THEN** **'80+'**

**END** **AS** classe\_eta

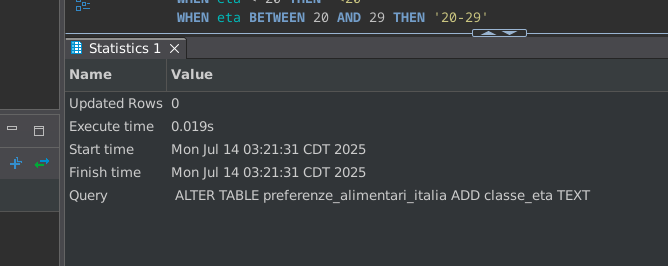
**from** preferenze\_alimentari\_italia pai

**WHERE**

eta <=100;

--inserisco tabella

**ALTER** **TABLE** preferenze\_alimentari\_italia **ADD** classe\_eta **TEXT**;



--ci metto le fasce usando SET invece di insert

**UPDATE** preferenze\_alimentari\_italia

**SET** classe\_eta = **CASE**

**WHEN** eta < 20 **THEN** '<20'

**WHEN** eta **BETWEEN** 20 **AND** 29 **THEN** '20-29'

**WHEN** eta **BETWEEN** 30 **AND** 39 **THEN** '30-39'

**WHEN** eta **BETWEEN** 40 **AND** 49 **THEN** '40-49'

**WHEN** eta **BETWEEN** 50 **AND** 59 **THEN** '50-59'

**WHEN** eta **BETWEEN** 60 **AND** 69 **THEN** '60-69'

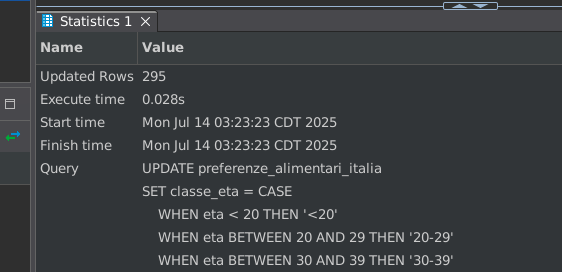
**WHEN** eta **BETWEEN** 70 **AND** 79 **THEN** '70-79'

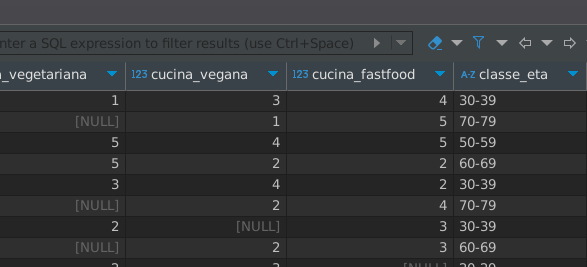
**WHEN** eta >= 80 **THEN** '80+'

**ELSE** **NULL**

**END**

**WHERE** eta <= 100;





**CREATE** **TABLE** anagrafica\_fasce\_eta (id\_fasce\_eta **INTEGER** **PRIMARY** **KEY** **AUTOINCREMENT**, titolo\_studio **TEXT** **UNIQUE**)

**INSERT** **INTO** anagrafica\_titolo\_studio (titolo\_studio)

**SELECT** **DISTINCT** titolo\_studio

**FROM** preferenze\_alimentari\_italia;

**INSERT** **INTO** preferenze\_alimentari\_italia (classe\_eta)

**SELECT** ID, eta,

**CASE**

**WHEN** eta < 20 **THEN** '<20'

**WHEN** eta >= 20 **AND** eta < 30 **THEN** '20-29'

**WHEN** eta >= 30 **AND** eta < 40 **THEN** '30-39'

**WHEN** eta >= 40 **AND** eta < 50 **THEN** '40-49'

**WHEN** eta >= 50 **AND** eta < 60 **THEN** '50-59'

**WHEN** eta >= 60 **AND** eta < 70 **THEN** '60-69'

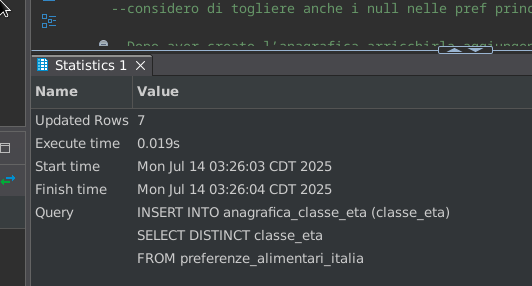
**WHEN** eta >= 70 **AND** eta < 80 **THEN** '70-79'

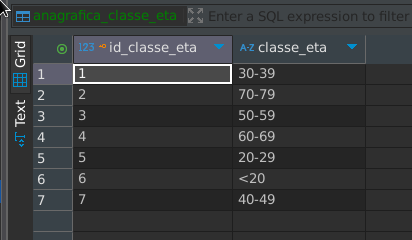
**WHEN** eta >= 80 **THEN** '>80'

**WHEN** eta **IS** **NULL** **THEN**

**END** **AS** *classe\_eta*

--poi creo anagrafica come ho fatto sopra





**SELECT** **distinct** classe\_eta **FROM** preferenze\_alimentari\_italia *pai* **WHERE** eta **is** **not** **null**;

**CREATE** **TABLE** anagrafica\_classe\_eta (id\_classe\_eta **INTEGER** **PRIMARY** **KEY** **AUTOINCREMENT**, classe\_eta **TEXT** **UNIQUE**);

**INSERT** **INTO** anagrafica\_classe\_eta (classe\_eta)

**SELECT** **DISTINCT** classe\_eta

**FROM** preferenze\_alimentari\_italia;

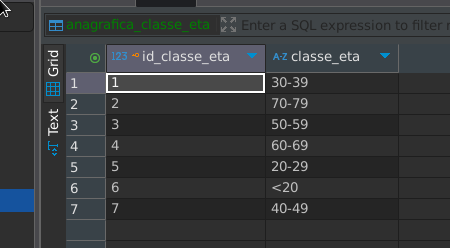
-- qui ho notato che mi ha preso un NULL e droppato la tabella e che non ci sono persone con piu' di 80 anni cosi'ho aggiunto a mano il dato. quindi qui ho sbagliato qualcosa quando sopra ho selezionato le eta nulle? Ho visto che rifacendo tutto per assignment con screenshot non mi prende piu’ null. Ma come si fa se una persona viene inserita ed ha 80 anni? Si fa di nuovo tutto? Quindi devo settare l’ultima riga come >80?

**UPDATE** anagrafica\_classe\_eta

**SET** classe\_eta = '>80'

**WHERE** classe\_eta **IS** **NULL**;

--considero di togliere anche i null nelle pref principali?



--Dopo aver creato l’anagrafica arricchirla aggiungendo le seguenti colonne

--Età Inizio

--Età Fine

--Descrizione fascia

**ALTER** **TABLE** anagrafica\_classe\_eta

**ADD** eta\_inizio **integer**;

**ALTER** **TABLE** anagrafica\_classe\_eta

**ADD** eta\_fine **integer**;

**ALTER** **TABLE** anagrafica\_classe\_eta

**ADD** descrizione **varchar**;

**INSERT** **INTO** anagrafica\_classe\_eta (eta\_inizio, eta\_fine, descrizione)

**SELECT**

**CASE**

**WHEN** classe\_eta = '<20' **THEN** 0

**WHEN** classe\_eta **LIKE** '20-%' **THEN** 20

**WHEN** classe\_eta **LIKE** '30-%' **THEN** 30

**WHEN** classe\_eta **LIKE** '40-%' **THEN** 40

**WHEN** classe\_eta **LIKE** '50-%' **THEN** 50

**WHEN** classe\_eta **LIKE** '60-%' **THEN** 60

**WHEN** classe\_eta **LIKE** '70-%' **THEN** 70

**WHEN** classe\_eta = '>80' **THEN** 80

**END** **AS** *eta\_inizio*,

**CASE**

**WHEN** classe\_eta = **'<20'** **THEN** 19

**WHEN** classe\_eta **LIKE** **'20-%'** **THEN** 29

**WHEN** classe\_eta **LIKE** **'30-%'** **THEN** 39

**WHEN** classe\_eta **LIKE** **'40-%'** **THEN** 49

**WHEN** classe\_eta **LIKE** **'50-%'** **THEN** 59

**WHEN** classe\_eta **LIKE** **'60-%'** **THEN** 69

**WHEN** classe\_eta **LIKE** **'70-%'** **THEN** 79

**WHEN** classe\_eta = **'>80'** **THEN** 999

**END** **AS** eta\_fine,

**CASE**

**WHEN** classe\_eta = **'<20'** **THEN** **'minorie di 20 anni'**

**WHEN** classe\_eta **LIKE** **'20-%'** **THEN** **'ventenne'**

**WHEN** classe\_eta **LIKE** **'30-%'** **THEN** **'trentenne'**

**WHEN** classe\_eta **LIKE** **'40-%'** **THEN** **'quarantenne'**

**WHEN** classe\_eta **LIKE** **'50-%'** **THEN** **'cinquantenne'**

**WHEN** classe\_eta **LIKE** **'60-%'** **THEN** **'sessantenne'**

**WHEN** classe\_eta **LIKE** **'70-%'** **THEN** **'settantenne'**

**WHEN** classe\_eta = **'>80'** **THEN** **'Over 80'**

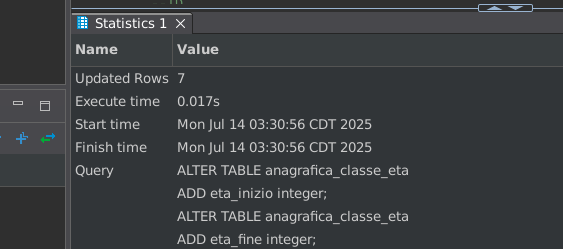
**END** **AS** descrizione

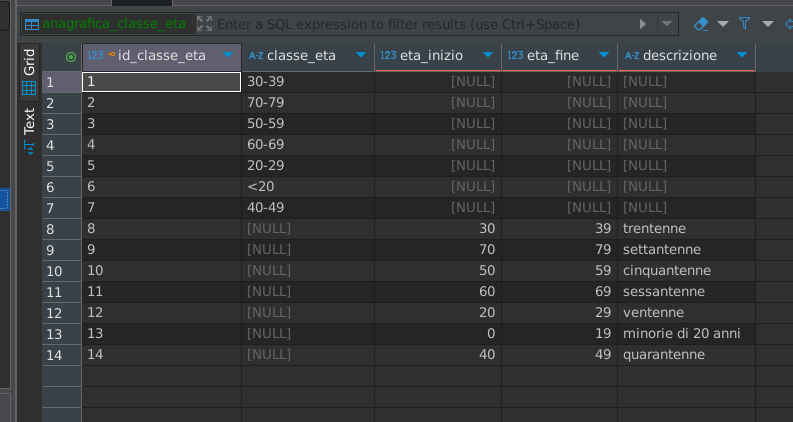
**FROM** (

**SELECT** **DISTINCT** classe\_eta

**FROM** preferenze\_alimentari\_italia

**WHERE** classe\_eta **IS** **NOT** **NULL**);





1. (task **opzionale**):
   * Creare un anagrafica con le tipologie di cucine
   * Per ogni persona creare una riga che abbia come colonne ID, Tipo Cucina e valutazione.   
     Per ogni persona ci devono essere N righe dove N è il numero di cucine su cui hanno espresso la preferenza. Ad esempio per ID 1 l’output dovrebbe essere così:

| ID | Tipo Cucina | Valutazione |
| --- | --- | --- |
| 1 | Italiana | 1 |
| 1 | Cinese | 4 |
| 1 | Giapponese | 1 |
| 1 | Indiana |  |
| 1 | … | … |

* + Sulla tabella ottenuta al punto precedente aggiungere una nuova colonna per indicare il cibo preferito (una soluzione potrebbe prevedere comandi che non abbiamo ancora visto in classe)

| ID | Tipo Cucina | Valutazione | Cibo Preferito |
| --- | --- | --- | --- |
| 1 | Italiana | 1 | Africana |
| 1 | Cinese | 4 | Africana |
| 1 | Giapponese | 1 | Africana |
| 1 | Indiana |  | Africana |
| 1 | … | … | Africana |

* + Cosa succede per le persone che hanno un punteggio massimo su più di una cucina?

# Vincolo

* E’ **necessario tracciare le query utilizzate e l’output ottenuto** in un google doc da salvare nelle propria cartelle personale

# Data di consegna (per gli amici DEADLINE)

14/07/2025 entro le 13:00 UTC+1