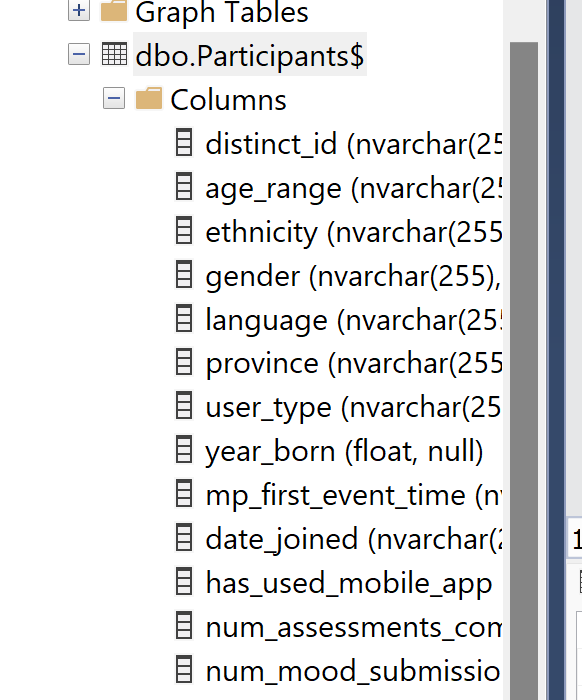
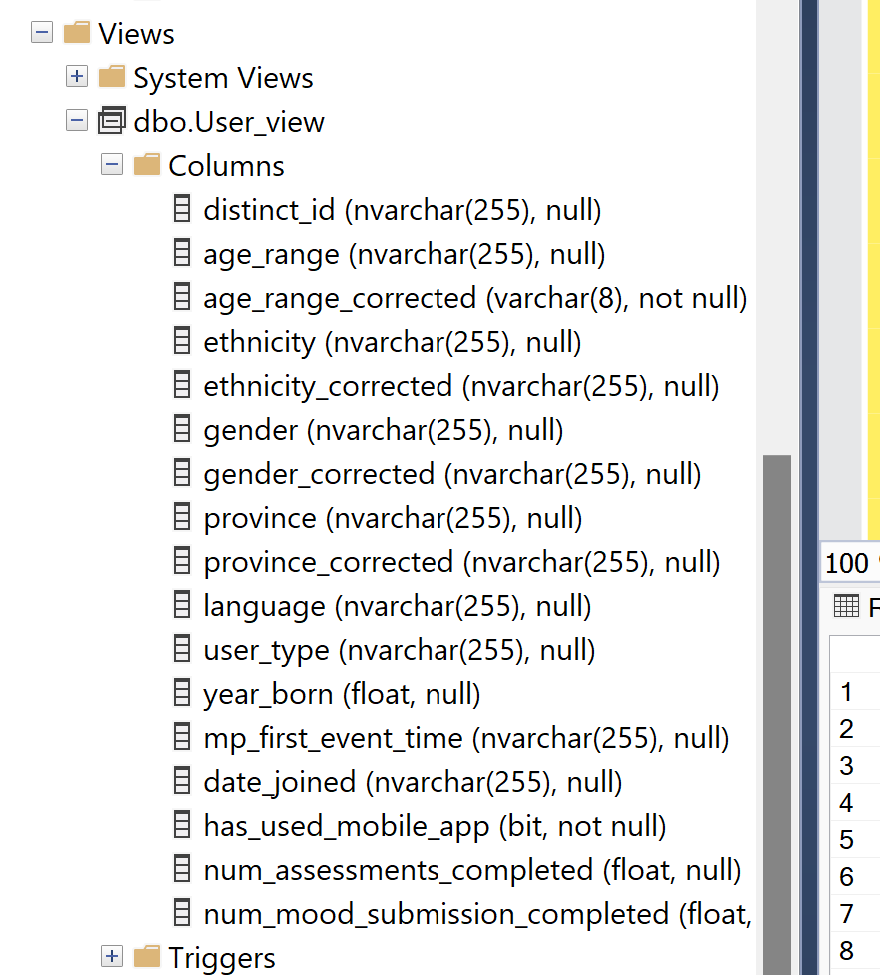
**Steps:**

1) create a table DBO.Participants in Microsoft SQL server managemnet studio



2) create a view “User\_view” from DBO.Participants



**Clean the data:-**

⦁ Cleaning done on gender to get the first item within ""

⦁ Cleaning done on Ethnicity to get the first item within ""

⦁ Cleaning done on Province to get the first item within "" as well as replace 'QuÃ©bec' with 'Quebec'

⦁ Cleaning done on age range because of too many Nulls to get the range based of year born

**Query used to create view :-**

CREATE VIEW User\_view AS

SELECT

distinct\_id,

age\_range,

case

when YEAR(GETDATE()) - year\_born >=10and YEAR(GETDATE()) - year\_born <=18 then '10 to 18'

when YEAR(GETDATE()) - year\_born >=19 and YEAR(GETDATE()) - year\_born <=29 then '19 to 29'

when YEAR(GETDATE()) - year\_born >=30 and YEAR(GETDATE()) - year\_born <=39 then '30 to 39'

when YEAR(GETDATE()) - year\_born >=40 and YEAR(GETDATE()) - year\_born <=49 then '40 to 49'

when YEAR(GETDATE()) - year\_born >=50 and YEAR(GETDATE()) - year\_born <=59 then '50 to 59'

when YEAR(GETDATE()) - year\_born >=60 and YEAR(GETDATE()) - year\_born <=69 then '60 to 69'

when YEAR(GETDATE()) - year\_born >=70 then '70+'

else 'NA' end AS age\_range\_corrected, -- Calculates corrected age range

ethnicity,

CASE

WHEN CHARINDEX('"', ethnicity, CHARINDEX('"', ethnicity) + 1) > CHARINDEX('"', ethnicity) THEN

SUBSTRING(

ethnicity,

CHARINDEX('"', ethnicity) + 1,

CHARINDEX('"', ethnicity, CHARINDEX('"', ethnicity) + 1) - CHARINDEX('"', ethnicity) - 1

)

ELSE

NULL

END AS ethnicity\_corrected, -- Extracts first ethnicity

gender,

CASE

WHEN CHARINDEX('"', gender, CHARINDEX('"', gender) + 1) > CHARINDEX('"', gender) THEN

SUBSTRING(

gender,

CHARINDEX('"', gender) + 1,

CHARINDEX('"', gender, CHARINDEX('"', gender) + 1) - CHARINDEX('"', gender) - 1

)

ELSE

NULL

END AS gender\_corrected, -- Extracts first gender

province,

CASE

WHEN CHARINDEX('"', province, CHARINDEX('"', province) + 1) > CHARINDEX('"', province) THEN

CASE

WHEN SUBSTRING(

province,

CHARINDEX('"', province) + 1,

CHARINDEX('"', province, CHARINDEX('"', province) + 1) - CHARINDEX('"', province) - 1

) = 'QuÃ©bec' THEN 'Quebec'

ELSE SUBSTRING(

province,

CHARINDEX('"', province) + 1,

CHARINDEX('"', province, CHARINDEX('"', province) + 1) - CHARINDEX('"', province) - 1

)

END

ELSE

NULL

END AS province\_corrected, -- Extracts first province and corrects Québec

-- Include other existing columns from the customer table

language,

user\_type,

year\_born,

mp\_first\_event\_time,

date\_joined,

has\_used\_mobile\_app,

num\_assessments\_completed,

num\_mood\_submission\_completed

FROM

dbo.Participants$;

**Query for Percentage :-**

SELECT

'Gender' AS Category,

gender\_corrected AS Types,

COUNT(\*) \* 100.0 / (SELECT COUNT(\*) FROM dbo.User\_view WHERE gender\_corrected IS NOT NULL) AS Percentage

FROM

dbo.User\_view

GROUP BY

gender\_corrected

UNION ALL

SELECT

'Age Range' AS Category,

age\_range\_corrected AS Types,

COUNT(\*) \* 100.0 / (SELECT COUNT(\*) FROM dbo.User\_view WHERE age\_range\_corrected IS NOT NULL) AS Percentage

FROM

dbo.User\_view

GROUP BY

age\_range\_corrected

UNION ALL

SELECT

'Ethnicity' AS Category,

ethnicity\_corrected AS Types,

COUNT(\*) \* 100.0 / (SELECT COUNT(\*) FROM dbo.User\_view WHERE ethnicity\_corrected IS NOT NULL) AS Percentage

FROM

dbo.User\_view

GROUP BY

ethnicity\_corrected

UNION ALL

SELECT

'Province' AS Category,

province\_corrected AS Types,

COUNT(\*) \* 100.0 / (SELECT COUNT(\*) FROM dbo.User\_view WHERE province\_corrected IS NOT NULL) AS Percentage

FROM

dbo.User\_view

GROUP BY

province\_corrected

UNION ALL

SELECT

'Language' AS Category,

language AS Types,

COUNT(\*) \* 100.0 / (SELECT COUNT(\*) FROM dbo.User\_view WHERE language IS NOT NULL) AS Percentage

FROM

dbo.User\_view

GROUP BY

language;