/\* In this project, I have used most of the SQL commands. For this SQL data Cleaning Project I am using data set given BY Ian K.

It has 3 table such as Aliens,Location and Details\*/

--1.TOP, The SELECT TOP clause is used to specify the number of records to return

SELECT TOP (5) [id]

,[first\_name]

,[last\_name]

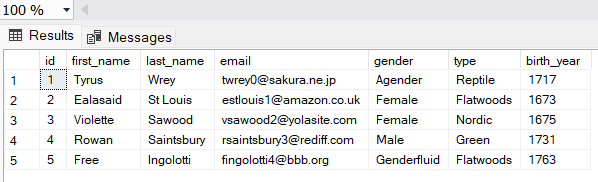
,[email]

,[gender]

,[type]

,[birth\_year]

FROM [test].[dbo].[Aliens$] ORDER BY id



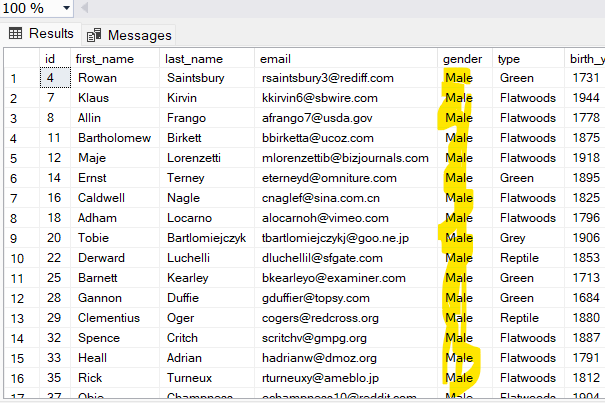
--2.WHERE CLAUSE.The WHERE clause is used to filter records. Below I try to find only male aliens

SELECT [id]

,[first\_name],[last\_name] ,[email],[gender],[type],[birth\_year]

FROM [test].[dbo].[Aliens$]

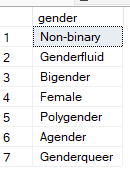
WHERE gender= 'Male'



-- 3.NOT EQUAL(!=). Below query to use not equal to find other gender

SELECT distinct [gender]

FROM [test].[dbo].[Aliens$]

WHERE gender != 'Male'  


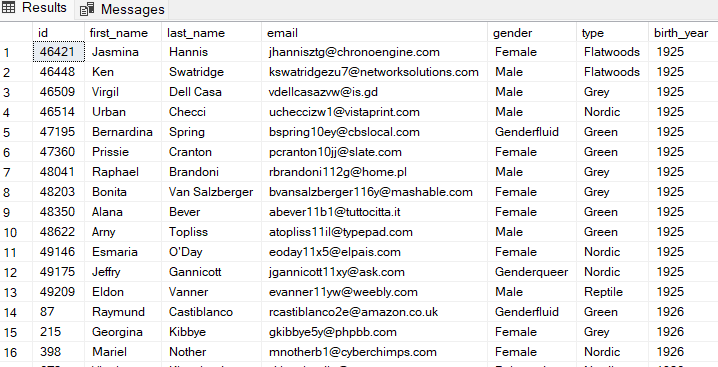
/\*4.BETWEEN The BETWEEN operator selects values within a given range. The values can be numbers, text, or dates.

--The BETWEEN operator is inclusive: begin and end values are included. \*/

SELECT [id]

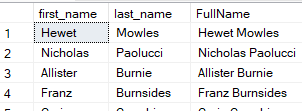
,[first\_name],[last\_name] ,[email],[gender],[type],[birth\_year]

FROM [test].[dbo].[Aliens$]

WHERE birth\_year between 1900 and 2000  


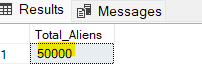
--5.The CONCAT() function adds two or more strings together.

SELECT [first\_name],[last\_name] ,CONCAT(first\_name,' ',last\_name) as FullName

FROM [test].[dbo].[Aliens$]   


--6.COUNT. To count how many ALIENS in this dataset.

SELECT count(id) as Total\_Aliens

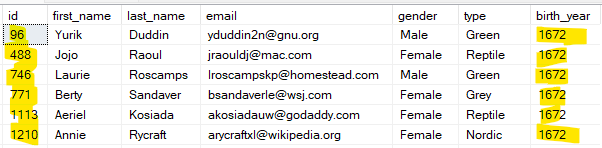
FROM [test].[dbo].[Aliens$]   


--7.ORDER BY. Below query to order by Birth year

SELECT [id]

,[first\_name],[last\_name] ,[email],[gender],[type],[birth\_year]

FROM [test].[dbo].[Aliens$]

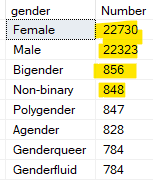
ORDER BY birth\_year  


-- 8.Group By . below query to group by genders

SELECT [gender], COUNT(gender) as Number

FROM [test].[dbo].[Aliens$]

GROUP BY gender

ORDER BY Number DESC  


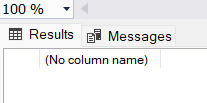
--9. HAVING Below query used HAVING clause to find the duplicate email address

SELECT count(email)

FROM [test].[dbo].[Aliens$]

GROUP BY email

HAVING count(email)>1



--10. LIKE Operator used in below query to find last name start with A

SELECT [last\_name]

FROM [test].[dbo].[Aliens$]

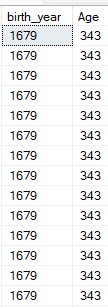
WHERE last\_name LIKE 'A%'



--11.YEAR() Function used below query to calculate the age of alien.

SELECT [birth\_year], [Age]=YEAR(getdate())-birth\_year

FROM [test].[dbo].[Aliens$]



--12.MAX() Function used below query to find the last alien birth year.

SELECT MAX(birth\_year) as Last\_Alien\_birth\_year

FROM [test].[dbo].[Aliens$]



--13.MIN() Function used below query to find the last alien birth year.

SELECT MIN(birth\_year) as First\_Alien\_birth\_year

FROM [test].[dbo].[Aliens$]



--14.VIEW() statement to create virtual table calculate age

CREATE VIEW [Age] as

SELECT [id] ,[first\_name],[last\_name] ,[email],[gender],[type],[birth\_year],[Age]=YEAR(getdate())-birth\_year

FROM [test].[dbo].[Aliens$]

SELECT \* FROM Age



--15.AVG() funtion use to find the agerage age of aliens

SELECT avg(Age) FROM Age



--16. CASE()expression use to group by them every century

SELECT Century,COUNT(Century) as Total\_Aliens

FROM

(SELECT

[Century]=

CASE

WHEN birth\_year between 0 and 1699 THEN 'Before 1700'

WHEN birth\_year between 1700 and 1799 THEN '18TH century'

WHEN birth\_year between 1800 and 1899 THEN '19TH centrury'

WHEN birth\_year >= 1900 THEN '20TH centrury'

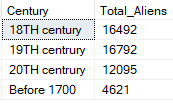
else 'other'

end

FROM [test].[dbo].[Aliens$] ) as cent

GROUP BY Century

ORDER BY Century



--17. JOIN clause use to joing aliens table with details table AND FIND total aggressive aliens

SELECT count( \*) as Total\_Agrresive\_aliens

FROM [test].[dbo].[Aliens$] A

INNER JOIN [test].[dbo].[Details$] D

ON A.id= D.detail\_id

WHERE aggressive= 1



--18.Below query used RIGHT() function to find the domain name and total users

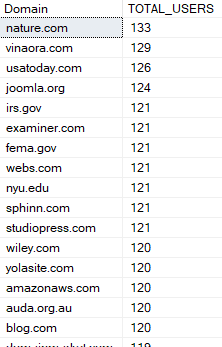
SELECT \*,COUNT(\*) AS TOTAL\_USERS from(

SELECT RIGHT(email,CHARINDEX('@',REVERSE(email))-1) as Domain

FROM [test].[dbo].[Aliens$] ) as Domian

GROUP BY Domain

ORDER BY COUNT(\*) DESC



--19. Replace() function use to include all gender together except male and female

SELECT [gender],REPLACE(gender,'Bigender','other')

FROM [test].[dbo].[Aliens$]

