Data Wrangling QBS 181

Homework -1

Instructions:

* Print random 10 rows to show these changes. Please also include the code
* Insert the answers directly into this document.
* You may collaborate on the homework but you must write it up yourselves.
* Please save the files in this format:
* Name of the file should be username(database)\_HW1\_QBS181
* Submit it via Canvas.

­­

1. Rename all the columns for eg.,­
2. TriAge to Age
3. GenderCode to Gender
4. ContactID to ID
5. Address1Stateorprovince to State
6. TriImagineCareenrollmentemailsentdate to EmailSentdate
7. Trienrollmentcompletedate to Completedate
8. Calculate the time (in days) to complete enrollment and create a new column to have this data
9. Create a new column “Enrollment Status”
10. Insert Status=Complete :code is 167410011
11. Insert Status=Email sent :code is 167410001
12. Insert Status=Non responder: Code is 167410004
13. Insert Status=Facilitated Enrollment: Code is 167410005
14. Insert Status= Incomplete Enrollments: Code is 167410002
15. Insert Status= Opted Out: Code is 167410003
16. Insert Status= Unprocessed: Code is 167410000
17. Insert Status= Second email sent : Code is 167410006
18. Create a new Column “Gender”
19. Insert Gender=female if code=2
20. Insert Gender=male if code=1
21. Insert Gender=other if code =167410000
22. Insert Gender=Unknown if code =’NULL’
23. Create a new column “Age group” and create age groups with an interval of 25 yrs. for example 0-25 years as ‘0-25’, 26-50 as “26-50” and so on...

--Question1 :

-- delete all null values, calculate the different

-- store in the new column

--Rename

select \* from Demographics

select\*into yzhao.Demographics from dbo.Demographics

select contactid as ID, gendercode as Gender, tri\_age as Age, parentcustomeridname, tri\_imaginecareenrollmentstatus, address1\_stateorprovince, tri\_imaginecareenrollmentemailsentdate as EmailSentdate, tri\_enrollmentcompletedate as completedate from yzhao.Demographics

alter table yzhao.Demographics

add timecompleteenrollment int

select\* from yzhao.Demographics

delete from yzhao.Demographics where tri\_enrollmentcompletedate like 'NULL'

--Calculate the time (in days) to complete enrollment and create a new column to have this data

select\* from yzhao.Demographics

-- Try to convert the data type in two columns

-- Try to calculate the difference between two columns and store in the new column

update yzhao.Demographics

set timecompleteenrollment = datediff(day, try\_convert(datetime,tri\_imaginecareenrollmentemailsentdate),try\_convert(datetime,tri\_enrollmentcompletedate))

select timecompleteenrollment from yzhao.Demographics

-- Question 2

-- select different code

-- make a new column

--

select \* from yzhao.Demographics

-- Create a new column:

alter table yzhao.Demographics

add Enrollmentstatus nvarchar(max) -- why if I didn't write (max), the error would show "String or binary data would be truncated" what does this error mean?

-- based on different number, define different status in the new column:

select \* from yzhao.Demographics

update yzhao.Demographics

set Enrollmentstatus = 'Complete'

where tri\_imaginecareenrollmentstatus = 167410011

update yzhao.Demographics

set Enrollmentstatus = 'Email sent'

where tri\_imaginecareenrollmentstatus = 167410001

update yzhao.Demographics

set Enrollmentstatus = 'Non responder'

where tri\_imaginecareenrollmentstatus = 167410004

update yzhao.Demographics

set Enrollmentstatus = 'Facilitated Enrollment'

where tri\_imaginecareenrollmentstatus = 167410005

update yzhao.Demographics

set Enrollmentstatus = 'Incomplete Enrollments'

where tri\_imaginecareenrollmentstatus = 167410002

update yzhao.Demographics

set Enrollmentstatus = 'Opted Out'

where tri\_imaginecareenrollmentstatus = 167410003

update yzhao.Demographics

set Enrollmentstatus = 'Unprocessed'

where tri\_imaginecareenrollmentstatus = 167410000

update yzhao.Demographics

set Enrollmentstatus = 'Second email sent'

where tri\_imaginecareenrollmentstatus = 167410006

--Question 3

select\* from yzhao.Demographics

alter table yzhao.Demographics

add Gender1 nvarchar(MAX)

update yzhao.Demographics

set Gender1 = 'female'

where try\_convert(int,gender) = 2

update yzhao.Demographics

set Gender1 = 'male'

where try\_convert(int,gender) = 1

update yzhao.Demographics

set Gender1 = 'other'

where try\_convert(int,gender) = 167410000

update yzhao.Demographics

set Gender1 = 'Unknown'

where gender is NULL

select\* from yzhao.Demographics

--Question 4

select\* from yzhao.Demographics

alter table yzhao.Demographics

add AgeGroup nvarchar(max)

update yzhao.Demographics

set AgeGroup = '0-25'

where tri\_age<25

update yzhao.Demographics

set AgeGroup = '26-50'

where tri\_age<=50 and tri\_age>=25

update yzhao.Demographics

set AgeGroup = '51-75'

where tri\_age<=75 and tri\_age>=51

update yzhao.Demographics

set AgeGroup = '76-100'

where tri\_age<=100 and tri\_age>=76

--check whether people are older than 100

select tri\_age from yzhao.Demographics

where tri\_age >100

Table

Description automatically generated