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 “Sex”
19. Insert sex=female if code=2
20. Insert sex=male if code=1
21. Insert sex=other if code =167410000
22. Insert sex=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...

select \* into blu.demographics from Demographics

--1.Rename all the columns

EXEC sp\_rename 'blu.demographics.tri\_age', 'Age', 'COLUMN'

EXEC sp\_rename 'blu.demographics.gendercode', 'Gender', 'COLUMN'

EXEC sp\_rename 'blu.demographics.contactid', 'ID', 'COLUMN'

EXEC sp\_rename 'blu.demographics.address1\_stateorprovince', 'State', 'COLUMN'

EXEC sp\_rename 'blu.demographics.tri\_imaginecareenrollmentemailsentdate', 'EmailSentdate', 'COLUMN'

EXEC sp\_rename 'blu.demographics.tri\_enrollmentcompletedate', 'Completedate', 'COLUMN'

alter table blu.demographics add datedifference nvarchar(100)

update blu.demographics set datedifference = datediff(day, TRY\_CONVERT(date, [EmailSentdate]),

TRY\_CONVERT(date,[Completedate])) from blu.demographics

--2.Create a new column “Enrollment Status”

alter table blu.demographics add EnrollmentStatus nvarchar(100)

update blu.demographics set EnrollmentStatus = 'Complete'

where [tri\_imaginecareenrollmentstatus] = 167410011

update blu.demographics set EnrollmentStatus = 'Email Sent'

where [tri\_imaginecareenrollmentstatus] = 167410001

update blu.demographics set EnrollmentStatus = 'Non Responder'

where [tri\_imaginecareenrollmentstatus] = 167410004

update blu.demographics set EnrollmentStatus = 'Facilitated Enrollment'

where [tri\_imaginecareenrollmentstatus] = 167410005

update blu.demographics set EnrollmentStatus = 'Incomplete Enrollments'

where [tri\_imaginecareenrollmentstatus] = 167410002

update blu.demographics set EnrollmentStatus = 'Opted Out'

where [tri\_imaginecareenrollmentstatus] = 167410003

update blu.demographics set EnrollmentStatus = 'Unprocessed'

where [tri\_imaginecareenrollmentstatus] = 167410000

update blu.demographics set EnrollmentStatus = 'Second Email Sent'

where [tri\_imaginecareenrollmentstatus] = 167410006

--3.Create a new Column “Sex”

alter table blu.demographics add Sex nvarchar(100)

update blu.demographics set Sex = 'female'

where [Gender] = '2'

update blu.demographics set Sex = 'male'

where [Gender] = '1'

update blu.demographics set Sex = 'other'

where [Gender] = '167410000'

update blu.demographics set Sex = 'Unknown'

where [Gender] = 'NULL'

--4.Create a new column “Age group” and create age groups with an interval of 25 yrs.

alter table blu.demographics add Agegroup nvarchar(100)

update blu.demographics set Agegroup = '0-25'

where [Age] >= 0 and [Age] <= 25

update blu.demographics set Agegroup = '26-50'

where [Age] >= 26 and [Age] <= 50

update blu.demographics set Agegroup = '51-75'

where [Age] >= 51 and [Age] <= 75

update blu.demographics set Agegroup = '76-100'

where [Age] >= 76 and [Age] <= 100

update blu.demographics set Agegroup = '101-125'

where [Age] >= 101 and [Age] <= 125

SELECT TOP 10 \* from blu.demographics

ORDER BY NEWID()

