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...



drop table wwu.hw1

select \* into wwu.hw1 from Demographics

--Problem 1

EXEC sp\_rename 'wwu.hw1.tri\_age','Age';

EXEC sp\_rename 'wwu.hw1.gendercode','Gender';

EXEC sp\_rename 'wwu.hw1.contactid','ID';

EXEC sp\_rename 'wwu.hw1.[address1\_stateorprovince]','State';

EXEC sp\_rename 'wwu.hw1.[tri\_imaginecareenrollmentemailsentdate]','EmailSentdate';

EXEC sp\_rename 'wwu.hw1.[tri\_enrollmentcompletedate]','Completedate';

ALTER TABLE wwu.hw1 ADD TIMEINTERVAL int;

update wwu.hw1 set TIMEINTERVAL = DATEDIFF(day,try\_convert(date,EmailSentdate),try\_convert(date,Completedate));

select TIMEINTERVAL from wwu.hw1;

--Problem 2

alter table wwu.hw1 add Enrollment\_Status text;

update wwu.hw1 set Enrollment\_Status='Complete' where [tri\_imaginecareenrollmentstatus]='167410011';

update wwu.hw1 set Enrollment\_Status='Email' where [tri\_imaginecareenrollmentstatus]='167410001';

update wwu.hw1 set Enrollment\_Status='Non responder' where [tri\_imaginecareenrollmentstatus]='167410004';

update wwu.hw1 set Enrollment\_Status='Facilitated Enrollment' where [tri\_imaginecareenrollmentstatus]='167410005';

update wwu.hw1 set Enrollment\_Status='Incomplete Enrollments' where [tri\_imaginecareenrollmentstatus]='167410002';

update wwu.hw1 set Enrollment\_Status='Opted Out' where [tri\_imaginecareenrollmentstatus]='167410003';

update wwu.hw1 set Enrollment\_Status='Unprocessed' where [tri\_imaginecareenrollmentstatus]='167410000';

update wwu.hw1 set Enrollment\_Status='Second email sent' where [tri\_imaginecareenrollmentstatus]='167410006';

select Enrollment\_Status from wwu.hw1

alter table wwu.hw1 add SEX text;

update wwu.hw1 set SEX='male' where [Gender]='1';

update wwu.hw1 set SEX='female' where [Gender]='2';

update wwu.hw1 set SEX='other' where [Gender]='167410000';

update wwu.hw1 set SEX='Unknown' where [Gender]='NULL';

--Problem 3

select max(Age) from wwu.hw1

alter table wwu.hw1 add Age\_group text;

update wwu.hw1 set Age\_group='0-25' where [Age]<=25;

update wwu.hw1 set Age\_group='26-50' where [Age]>25 and [Age]<=50

update wwu.hw1 set Age\_group='51-75' where [Age]>50 and [Age]<=75

update wwu.hw1 set Age\_group='76-100' where [Age]>75 and [Age]<=100