Data Wrangling QBS 181

Homework -1

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

Code in SQL:

**select** \* **into** agerutshang.demo2 **from** Demographics

sp\_RENAME 'agerutshang.demo2.tri\_age' , 'Age', 'COLUMN'

sp\_RENAME 'agerutshang.demo2.GenderCode', 'Gender', 'COLUMN'

sp\_RENAME 'agerutshang.demo2.ContactID', 'ID', 'COLUMN'

sp\_RENAME 'agerutshang.demo2.address1\_stateorprovince', 'State', 'COLUMN'

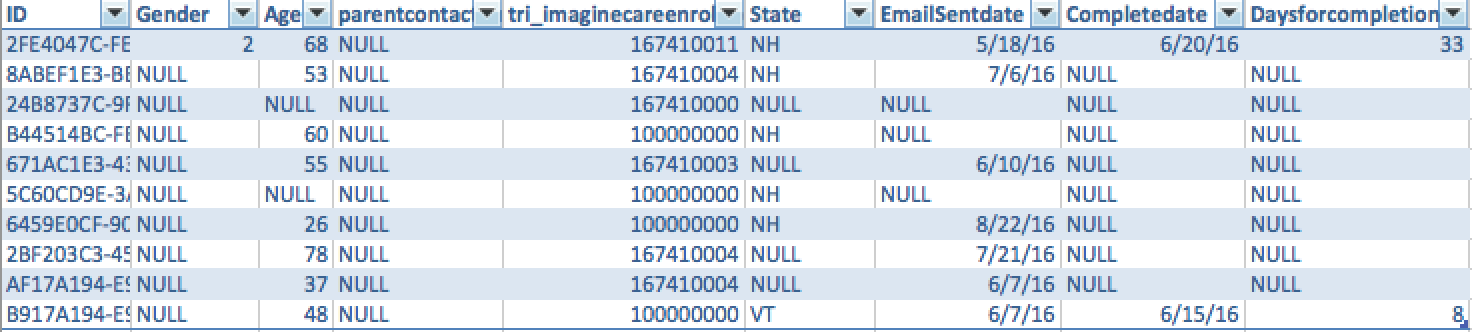
sp\_RENAME 'agerutshang.demo2.tri\_imaginecareenrollmentemailsentdate', 'EmailSentdate', 'COLUMN'

sp\_RENAME 'agerutshang.demo2.tri\_enrollmentcompletedate', 'Completedate', 'COLUMN'

**alter** **table** agerutshang.demo2

**add** Daysforcompletion **as** datediff(**day**, try\_convert(**date**,EmailSentdate),try\_convert(**date**,Completedate))

**select** **top** 10 \* **from** agerutshang.demo2

Table Result:

1. Create a new column “Enrollment Status”
2. Insert Status=Complete :code is 167410011
3. Insert Status=Email sent :code is 167410001
4. Insert Status=Non responder: Code is 167410004
5. Insert Status=Facilitated Enrollment: Code is 167410005
6. Insert Status= Incomplete Enrollments: Code is 167410002
7. Insert Status= Opted Out: Code is 167410003
8. Insert Status= Unprocessed: Code is 167410000
9. Insert Status= Second email sent : Code is 167410006

Code in SQL:

**alter** **table** agerutshang.demo2 **add** EnrollmentStatus **nvarchar** (225)

**update** agerutshang.demo2

**set** EnrollmentStatus='Complete' **where** tri\_imaginecareenrollmentstatus=167410011

**update** agerutshang.demo2

**set** EnrollmentStatus='Email Sent' **where** tri\_imaginecareenrollmentstatus=167410001

**update** agerutshang.demo2

**set** EnrollmentStatus='Non responder' **where** tri\_imaginecareenrollmentstatus=167410004

**update** agerutshang.demo2

**set** EnrollmentStatus='Facilitated Enrollment' **where** tri\_imaginecareenrollmentstatus=167410005

**update** agerutshang.demo2

**set** EnrollmentStatus='Incomplete Enrollments' **where** tri\_imaginecareenrollmentstatus=167410002

**update** agerutshang.demo2

**set** EnrollmentStatus='Opted out' **where** tri\_imaginecareenrollmentstatus=167410003

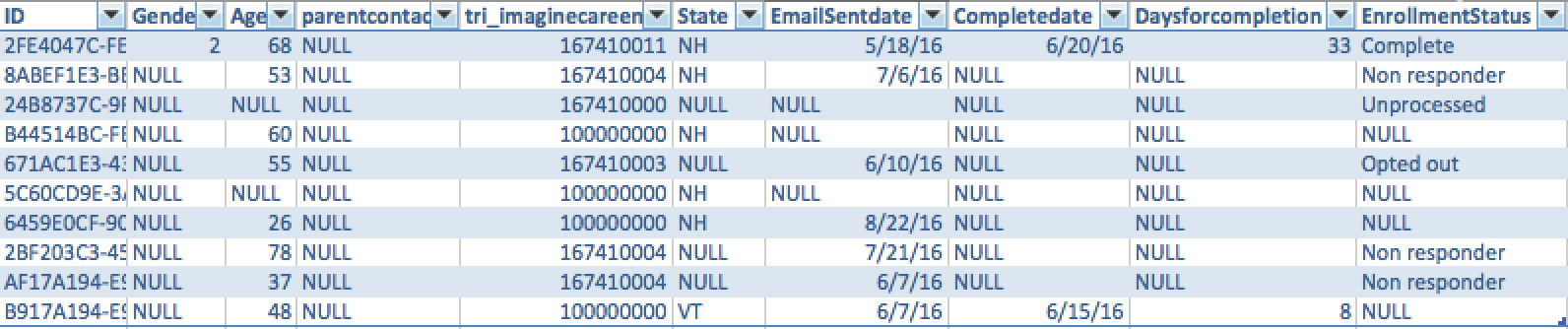
**update** agerutshang.demo2

**set** EnrollmentStatus='Unprocessed' **where** tri\_imaginecareenrollmentstatus=167410000

**update** agerutshang.demo2

**set** EnrollmentStatus='Second email' **where** tri\_imaginecareenrollmentstatus=167410006

**select** **top** 10 \* **from** agerutshang.demo2

Table Result:

1. Create a new Column “Sex”
2. Insert sex=female if code=2
3. Insert sex=male if code=1
4. Insert sex=other if code =167410000
5. Insert sex=Unknown if code =’NULL’

Code in SQL:

**alter** **table** agerutshang.demo2 **add** Sex **nvarchar**(20)

**update** agerutshang.demo2

**set** Sex='female' **where** Gender='2'

**update** agerutshang.demo2

**set** Sex='male' **where** Gender='1'

**update** agerutshang.demo2

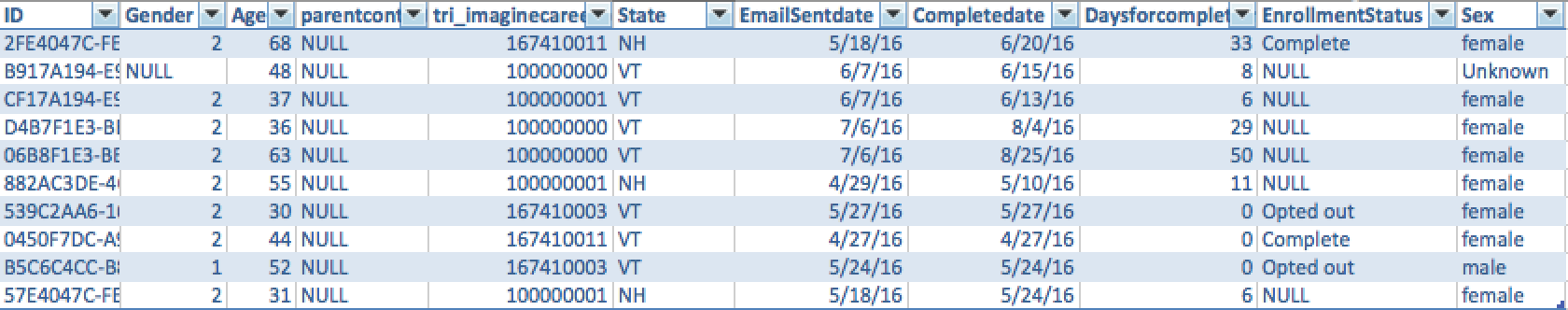
**set** Sex='other' **where** Gender='167410000'

**update** agerutshang.demo2

**set** Sex='Unknown' **where** Gender='NULL'

**select** **top** 10 \* **from** agerutshang.demo2 **where** Daysforcompletion **is** **not** **NULL**

The last line selecting the top 10 was so that most of the values in some of the columns wouldn’t just be NULL.

Table Result:

1. 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...

Code in SQL:

**alter** **table** agerutshang.demo2 **add** AgeGroup **nvarchar** (10)

**update** agerutshang.demo2

**set** AgeGroup='0-25' **where** Age **between** 0 **and** 25

**update** agerutshang.demo2

**set** AgeGroup='26-30' **where** Age **between** 26 **and** 30

**update** agerutshang.demo2

**set** AgeGroup='31-45' **where** Age **between** 31 **and** 45

**update** agerutshang.demo2

**set** AgeGroup='46-60' **where** Age **between** 46 **and** 60

**update** agerutshang.demo2

**set** AgeGroup='61-75' **where** Age **between** 61 **and** 75

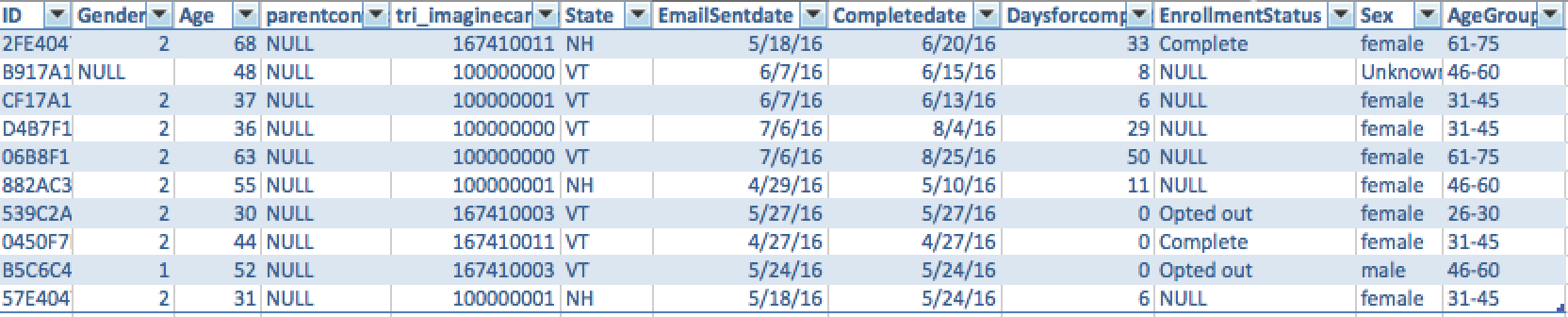
**update** agerutshang.demo2

**set** AgeGroup='76-90' **where** Age **between** 76 **and** 90

**update** agerutshang.demo2

**set** AgeGroup='>90' **where** Age>90

**select** **top** 10 \* **from** agerutshang.demo2 **where** Daysforcompletion **is** **not** **NULL**

Table Result: