

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.,

- a) TriAge to Age
- b) GenderCode to Gender
- c) ContactID to ID
- d) Address1Stateorprovince to State
- e) TriImagineCareenrollmentemailsentdate to EmailSentdate
- f) Trienrollmentcompletedate to Completedate
- g) Calculate the time (in days) to complete enrollment and create a new column to have this data

	Age	Gender	ID	State	EmailSentdate	Completedate	DaysToComplete
1	32	2	B05CD306-D547-E611-80E6-5065F38BA151	NH	7/12/2016	7/27/2016	15
2	35	NULL	5F9C2AA6-1624-E611-812A-C4346BB59854	NH	5/27/2016	NULL	NULL
3	31	NULL	9A92A7B0-7148-E611-80E7-5065F38B3241	NH	7/12/2016	NULL	NULL
4	54	NULL	805FD306-D547-E611-80E6-5065F38BA151	NH	7/12/2016	NULL	NULL
5	63	NULL	B65FD306-D547-E611-80E6-5065F38BA151	NH	7/12/2016	NULL	NULL
6	34	2	0F52413C-72A0-E611-80E9-5065F38A3B91	NH	11/1/2016	11/5/2016	4
7	40	2	D686A327-A108-E611-811D-C4346BACD1A8	NH	4/22/2016	4/22/2016	0
8	44	NULL	985AD306-D547-E611-80E6-5065F38BA151	NH	7/12/2016	NULL	NULL
9	32	NULL	4293A7B0-7148-E611-80E7-5065F38B3241	NH	7/12/2016	NULL	NULL
10	44	NULL	7A9AA7B0-7148-E611-80E7-5065F38B3241	NH	7/12/2016	NULL	NULL

```
-- 1 (a:f)
SELECT tri_age AS Age, gendercode AS Gender, contactid AS ID,
       address1_stateorprovince AS State, tri_imaginecareenrollmentemailsentdate AS
EmailSentdate,
       tri_enrollmentcompletedate AS Completedate
INTO qbs181.mtaylor.Q1
FROM qbs181.dbo.Demographics;
```

```
-- 1 (g)
ALTER TABLE qbs181.mtaylor.Q1
ADD DaysToComplete AS CASE WHEN TRY_CONVERT(DATETIME,Completedate) IS NULL THEN NULL
                                WHEN TRY_CONVERT(DATETIME,EmailSentdate) IS NULL THEN
NULL
                                ELSE DATEDIFF(day, CONVERT(DATETIME,EmailSentdate),
CONVERT(DATETIME,Completedate))
                                END;
```

```
SELECT TOP 10 * FROM qbs181.mtaylor.Q1
ORDER BY NEWID();
```

2. Create a new column "Enrollment Status"

- Insert Status=Complete :code is 167410011
- Insert Status=Email sent :code is 167410001
- Insert Status=Non responder: Code is 167410004
- Insert Status=Facilitated Enrollment: Code is 167410005
- Insert Status= Incomplete Enrollments: Code is 167410002
- Insert Status= Opted Out: Code is 167410003
- Insert Status= Unprocessed: Code is 167410000
- Insert Status= Second email sent : Code is 167410006

contactid	gendercode	tri_age	parentcustomeridname	tri_imaginecareenrollmentstatus	address1_stateorprovince	tri_imaginecareenrollmentemailsentdate	tri_enrollmentcompletedate	gender	Enrollment_Status
1	5C8AA7B0-7148-E611-80E7-5065F38B3241	NULL	55	Dartmouth-Hitchcock	167410004	VT	7/12/2016	NULL	Non responder
2	085BD306-D547-E611-80E6-5065F38BA151	NULL	41	Dartmouth-Hitchcock	167410004	NH	7/12/2016	NULL	Non responder
3	33298D77-EF80-E611-80EE-5065F38B3241	2	22	Dartmouth-Hitchcock	167410011	NH	9/25/2016	2	Complete
4	0B8701DC-7D38-E611-80E4-C4346BDC0341	NULL	44	Dartmouth-Hitchcock	100000001	NULL	6/22/2016	NULL	NULL
5	9889E3C8-5111-E611-811B-C4346BAC02E8	NULL	24	Dartmouth-Hitchcock	167410004	NULL	5/3/2016	NULL	Non responder
6	168AA7B0-7148-E611-80E7-5065F38B3241	NULL	58	Dartmouth-Hitchcock	167410004	VT	7/12/2016	NULL	Non responder
7	11794C35-2945-E611-80E6-C4346BDC9111	NULL	32	Dartmouth-Hitchcock	167410004	VT	7/8/2016	NULL	Non responder
8	C8534411-A7F7-E511-811C-FC15B42886E8	2	55	The ImagineCare Tes...	167410011	CT	4/1/2016	2	Complete
9	057C4C35-2945-E611-80E6-C4346BDC9111	NULL	48	Dartmouth-Hitchcock	167410004	NH	7/8/2016	NULL	Non responder
10	DB40A719-F3C8-E611-80F5-5065F38A4B01	1	33	ProdSquad Testers N...	167410011	NH	12/20/2016	1	Complete

```
-- 2
ALTER TABLE qbs181.mtaylor.HW1
ADD Enrollment_Status AS CASE WHEN tri_imaginecareenrollmentstatus = 167410011 THEN
'Complete'
                                WHEN tri_imaginecareenrollmentstatus =
167410001 THEN 'Email sent'
                                WHEN tri_imaginecareenrollmentstatus =
167410004 THEN 'Non responder'
                                WHEN tri_imaginecareenrollmentstatus =
167410005 THEN 'Facilitated Enrollment'
                                WHEN tri_imaginecareenrollmentstatus =
167410002 THEN 'Incomplete Enrollments'
                                WHEN tri_imaginecareenrollmentstatus =
167410003 THEN 'Opted Out'
                                WHEN tri_imaginecareenrollmentstatus =
167410000 THEN 'Unprocessed'
                                WHEN tri_imaginecareenrollmentstatus =
167410006 THEN 'Second email sent'
                                END;
```

```
SELECT TOP 10 * FROM qbs181.mtaylor.HW1
ORDER BY NEWID();
```

3. Create a new Column “Gender”

- Insert Gender=female if code=2
- Insert Gender=male if code=1
- Insert Gender=other if code =167410000
- Insert Gender=Unknownn if code ='NULL'

	contactid	gendercode	tri_age	parentcustomeridname	tri_imaginecareenrollmentstatus	address1_stateorprovince	tri_imaginecareenrollmentemailsentdate	tri_enrollmentcompleteddate	Gender
1	07794C35-2945-E611-80E6-C4346BDC9111	NULL	29	Dartmouth-Hitchcock	167410004	VT	7/8/2016	NULL	Unknown
2	DDD77666-0A0E-E611-811E-C4346BACD1A8	1	28	Dartmouth-Hitchcock	167410011	NH	4/29/2016	4/29/2016	male
3	4847C5F7-0C16-E611-8128-C4346BB59854	1	37	Dartmouth-Hitchcock	167410011	NH	5/9/2016	5/12/2016	male
4	F0D8C3D3-E9E6-E511-8116-C4346BAC02E8	2	19	The ImagineCare Tes...	167410011	CT	3/10/2016	3/15/2016	female
5	DB86A327-A108-E611-811D-C4346BACD1A8	2	48	Dartmouth-Hitchcock	167410003	NH	4/22/2016	NULL	female
6	09794C35-2945-E611-80E6-C4346BDC9111	NULL	35	Dartmouth-Hitchcock	167410004	NH	7/8/2016	NULL	Unknown
7	BC9CA7B0-7148-E611-80E7-5065F38B3241	NULL	45	Dartmouth-Hitchcock	167410004	VT	7/12/2016	NULL	Unknown
8	88BEF1E3-BE43-E611-80E6-5065F38BA151	NULL	64	Dartmouth-Hitchcock	167410004	NH	7/6/2016	NULL	Unknown
9	B486A7B0-7148-E611-80E7-5065F38B3241	NULL	59	Dartmouth-Hitchcock	167410004	NH	7/12/2016	NULL	Unknown
10	BABBF1E3-BE43-E611-80E6-5065F38BA151	NULL	43	Dartmouth-Hitchcock	167410004	NH	7/6/2016	NULL	Unknown

--3

```
ALTER TABLE qbs181.mtaylor.HW1
DROP COLUMN gender
```

```
ALTER TABLE qbs181.mtaylor.HW1
ADD Gender AS CASE WHEN gendercode = 'NULL' THEN 'Unknown'
                    WHEN gendercode = 2 THEN 'female'
                    WHEN gendercode = 1 THEN 'male'
                    WHEN tri_imaginecareenrollmentstatus = 167410000 THEN 'other'
END;
```

```
SELECT TOP 10 * FROM qbs181.mtaylor.HW1
ORDER BY NEWID();
```

4. 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...

	contactid	gendercode	tri_age	parentcustomeridname	tri_imaginecareenrollmentstatus	address1_stateorprovince	tri_imaginecareenrollmentemailsentdate	tri_enrollmentcompleteddate	gender	Age_group
1	AD784C35-2945-E611-80E6-C4346BDC9111	NULL	50	Dartmouth-Hitchcock	167410004	NH	7/8/2016	NULL	NULL	26-50
2	F1747ED3-02C3-E611-80F5-5065F38A4B01	NULL	26	Dartmouth-Hitchcock	100000001	NULL	12/15/2016	NULL	NULL	26-50
3	2687A7B0-7148-E611-80E7-5065F38B3241	NULL	28	Dartmouth-Hitchcock	167410004	NH	7/12/2016	NULL	NULL	26-50
4	DB784C35-2945-E611-80E6-C4346BDC9111	2	49	Dartmouth-Hitchcock	167410003	VT	7/8/2016	7/18/2016	2	26-50
5	47C234AD-C355-E611-80EC-5065F38B0171	1	63	Dartmouth-Hitchcock	167410011	NH	7/29/2016	7/29/2016	1	51-75
6	D499A7B0-7148-E611-80E7-5065F38B3241	NULL	40	Dartmouth-Hitchcock	167410004	NH	7/12/2016	NULL	NULL	26-50
7	CA90A7B0-7148-E611-80E7-5065F38B3241	NULL	60	Dartmouth-Hitchcock	167410004	NH	7/12/2016	NULL	NULL	51-75
8	3295A7B0-7148-E611-80E7-5065F38B3241	NULL	31	Dartmouth-Hitchcock	167410004	NH	7/12/2016	NULL	NULL	26-50
9	1887A327-A108-E611-811D-C4346BACD1A8	NULL	56	Dartmouth-Hitchcock	100000001	VT	4/22/2016	NULL	NULL	51-75
10	2752413C-72A0-E611-80E9-5065F38A3B91	NULL	60	Dartmouth-Hitchcock	167410004	NH	11/1/2016	NULL	NULL	51-75

--4

```
SELECT *
INTO qbs181.mtaylor.Q4
FROM qbs181.dbo.Demographics;

ALTER TABLE qbs181.mtaylor.Q4
ADD Age_group AS CASE WHEN tri_age < 25 THEN '0-25'
                      WHEN tri_age < 51 THEN '26-50'
                      WHEN tri_age < 76 THEN '51-75'
                      ELSE '75+'
END;
```

```
SELECT TOP 10 * FROM qbs181.mtaylor.Q4
ORDER BY NEWID();
```

