Name : Suraj Prasad
Email : surajprasadoffcial@gmail.com

DIABETES PREDICTION ASSESSMENT

Q1. Retrieve the Patient_id and ages of all patients.

```
1   -- Q1--
2 • SELECT patient_id,age FROM patients;
3
```

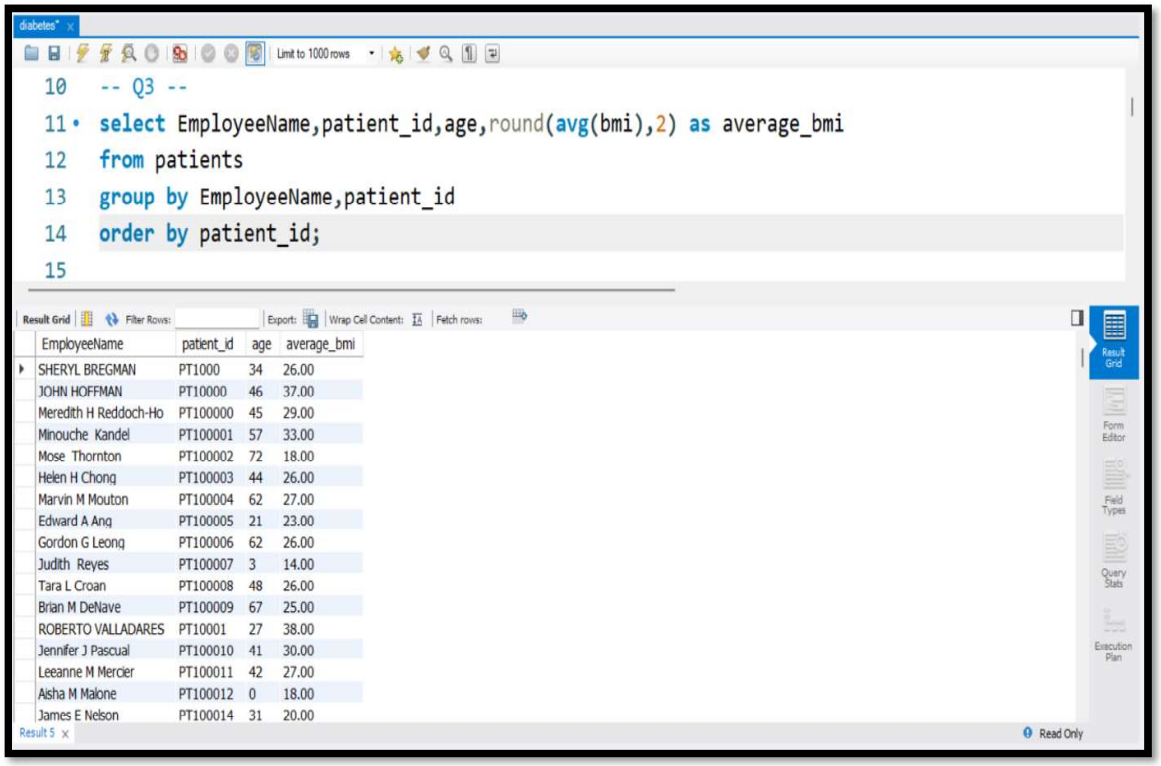
patient_id	age
PT1000	34
PT10000	46
PT100000	45
PT100001	57
PT100002	72
PT100003	44
PT100004	62
PT100005	21
PT100006	62
PT100007	3
PT100008	48
PT100009	67
PT10001	27
PT100010	41
PT100011	42
PT100012	0
PT100014	31
PT100015	47
PT100016	46
PT100017	1
PT100018	20

Q2. Select all female patients who are older than 40.

```
5   -- Q2--
6 • select * from patients
7   where gender = 'Female' and age > 40;
```

EmployeeName	Patient_id	gender	age	hypertension	heart_disease	smoking_history	bmi	HbA1c_level	blood_glucose_level	diabetes
JOHN HOFFMAN	PT10000	Female	46	No	No	No Info	37	6	200	No
Mose Thornton	PT100002	Female	72	No	No	current	18	6	145	No
Helen H Chong	PT100003	Female	44	No	No	No Info	26	6	90	No
Marvin M Mouton	PT100004	Female	62	No	No	No Info	27	5	200	No
Leeanne M Mercier	PT100011	Female	42	No	No	never	27	4	140	No
Joshua R Mcdonald	PT100015	Female	47	Yes	No	current	25	5	80	No
Loretta G Mild	PT100016	Female	46	No	No	No Info	27	7	100	No
Kenny Nguyen	PT100019	Female	78	No	No	never	37	6	155	No
Kristen W Peterson	PT100020	Female	74	No	No	current	30	8	130	Yes
Dorothy Chan	PT100021	Female	52	No	No	No Info	27	6	90	No
Roy Johnson Jr	PT100023	Female	80	No	No	not current	27	7	85	No
Jennifer M Acha	PT100026	Female	65	No	No	never	24	6	200	No
Tommy McGowan	PT100027	Female	50	No	No	former	34	6	160	No
MICHAEL STEZ	PT10003	Female	60	No	No	never	38	6	80	No
Jensa Woo	PT100031	Female	54	No	No	not current	27	7	80	No
Elena Guslkov	PT100032	Female	50	No	No	No Info	27	7	160	No
Stephanie Chang	PT100036	Female	65	Yes	Yes	never	34	8	140	Yes
Sandy C.H. Lui	PT100038	Female	47	No	No	No Info	27	4	200	No
JAMES GIAMBRUNO	PT10004	Female	56	No	No	never	24	6	200	No
Marcel C Barnes	PT100043	Female	51	No	No	current	30	7	160	No

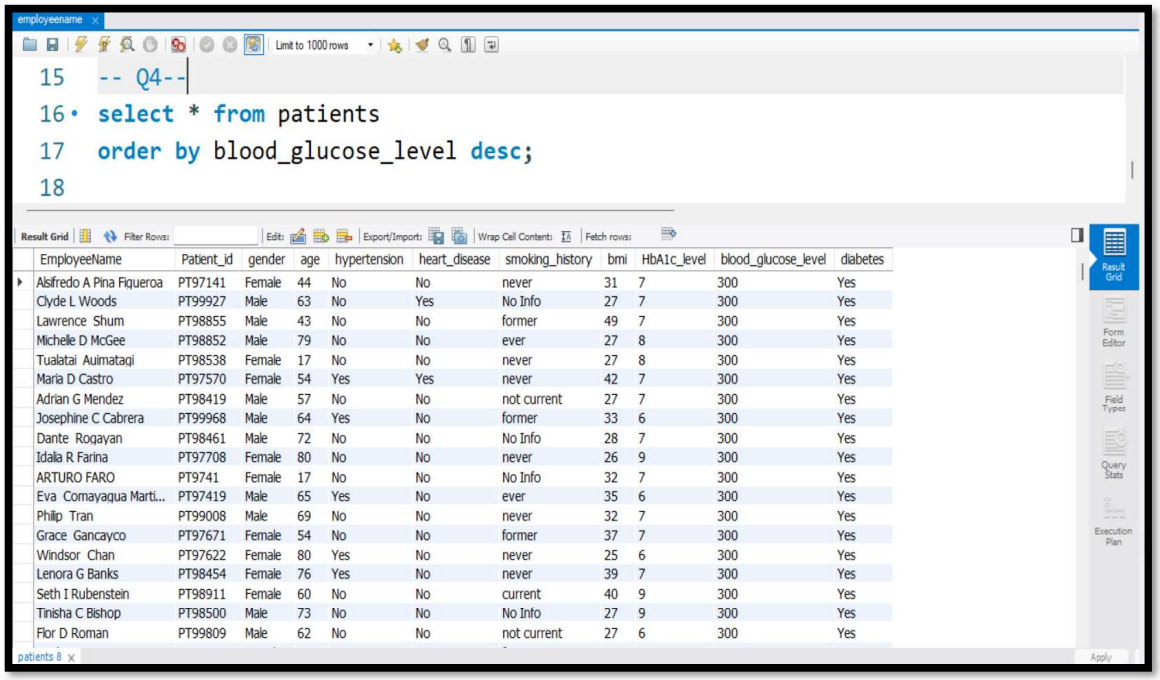
Q3. Calculate the average BMI of patients.



```
10  -- Q3 --
11  select EmployeeName,patient_id,age,round(avg(bmi),2) as average_bmi
12  from patients
13  group by EmployeeName,patient_id
14  order by patient_id;
15
```

EmployeeName	patient_id	age	average_bmi
SHERYL BREGMAN	PT1000	34	26.00
JOHN HOFFMAN	PT10000	46	37.00
Meredith H Reddoch-Ho	PT100000	45	29.00
Minouche Kandel	PT100001	57	33.00
Mose Thornton	PT100002	72	18.00
Helen H Chong	PT100003	44	26.00
Marvin M Mouton	PT100004	62	27.00
Edward A Ang	PT100005	21	23.00
Gordon G Leong	PT100006	62	26.00
Judith Reyes	PT100007	3	14.00
Tara L Croan	PT100008	48	26.00
Brian M DeNave	PT100009	67	25.00
ROBERTO VALLADARES	PT10001	27	38.00
Jennifer J Pascual	PT100010	41	30.00
Leeanne M Mercier	PT100011	42	27.00
Aisha M Malone	PT100012	0	18.00
James E Nelson	PT100014	31	20.00

Q4. List patients in descending order of blood glucose levels.



```
15  -- Q4--
16  select * from patients
17  order by blood_glucose_level desc;
18
```

EmployeeName	Patient_id	gender	age	hypertension	heart_disease	smoking_history	bmi	HbA1c_level	blood_glucose_level	diabetes
Alsifredo A Pina Figueroa	PT97141	Female	44	No	No	never	31	7	300	Yes
Clyde L Woods	PT99927	Male	63	No	Yes	No Info	27	7	300	Yes
Lawrence Shum	PT98855	Male	43	No	No	former	49	7	300	Yes
Michelle D McGee	PT98852	Male	79	No	No	ever	27	8	300	Yes
Tualatai Auimatagi	PT98538	Female	17	No	No	never	27	8	300	Yes
Maria D Castro	PT97570	Female	54	Yes	Yes	never	42	7	300	Yes
Adrian G Mendez	PT98419	Male	57	No	No	not current	27	7	300	Yes
Josephine C Cabrera	PT99968	Male	64	Yes	No	former	33	6	300	Yes
Dante Roqayan	PT98461	Male	72	No	No	No Info	28	7	300	Yes
Idalia R Farina	PT97708	Female	80	No	No	never	26	9	300	Yes
ARTURO FARO	PT9741	Female	17	No	No	No Info	32	7	300	Yes
Eva Comayaqua Marti...	PT97419	Male	65	Yes	No	ever	35	6	300	Yes
Philp Tran	PT99008	Male	69	No	No	never	32	7	300	Yes
Grace Gancayco	PT97671	Female	54	No	No	former	37	7	300	Yes
Windsor Chan	PT97622	Female	80	Yes	No	never	25	6	300	Yes
Lenora G Banks	PT98454	Female	76	Yes	No	never	39	7	300	Yes
Seth I Rubenstein	PT98911	Female	60	No	No	current	40	9	300	Yes
Tinisha C Bishop	PT98500	Male	73	No	No	No Info	27	9	300	Yes
Flor D Roman	PT99809	Male	62	No	No	not current	27	6	300	Yes

Q5. Find patients who have hypertension and diabetes.

```
18  -- Q5 --
19  select * from patients
20  where hypertension = 'Yes' and diabetes = 'Yes';
21
```

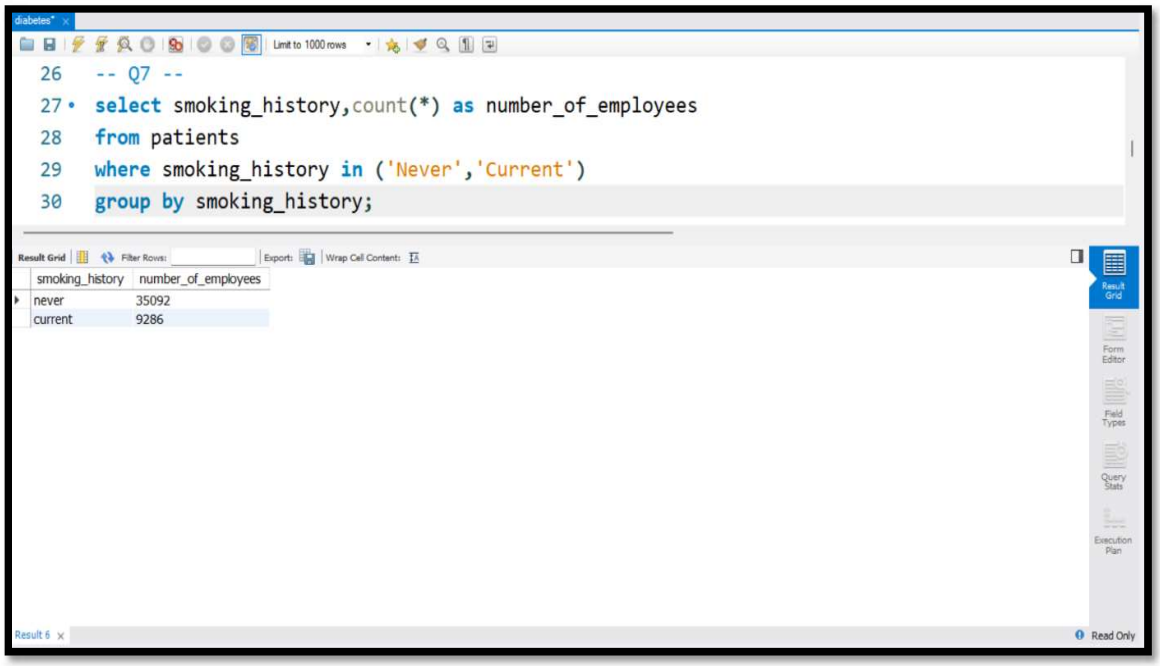
EmployeeName	Patient_id	gender	age	hypertension	heart_disease	smoking_history	bmi	HbA1c_level	blood_glucose_level	diabetes
Stephanie Chang	PT100036	Female	65	Yes	Yes	never	34	8	140	Yes
Katherine J Hoeber	PT100063	Female	58	Yes	No	never	38	7	200	Yes
LELA RUSSO	PT10007	Female	67	Yes	No	ever	25	7	159	Yes
PEDRO SANDOVAL	PT10083	Female	65	Yes	No	never	28	7	130	Yes
OMAR DAPIAOEN	PT10159	Female	59	Yes	No	current	50	7	140	Yes
EBENEZER ESPINOZA	PT10311	Male	64	Yes	No	not current	23	6	145	Yes
KENNETH KWONG	PT10315	Female	64	Yes	Yes	former	32	8	145	Yes
BRIAN LOUIE	PT10318	Male	56	Yes	No	never	49	7	155	Yes
LILLIAN LOUIE	PT10476	Male	80	Yes	No	never	27	6	145	Yes
THU-YEN PHAN	PT10498	Female	35	Yes	No	ever	56	6	145	Yes
JUAN GARCIA	PT10537	Female	75	Yes	No	No Info	23	9	140	Yes
BENJAMIN MELLOTT	PT10558	Male	47	Yes	No	No Info	27	9	200	Yes
AVELINA PACHECO	PT10674	Male	73	Yes	Yes	current	25	9	155	Yes
DANIEL SMITH	PT10694	Female	54	Yes	No	No Info	35	6	200	Yes
LARRY CAMILLERI	PT1075	Female	44	Yes	No	former	37	7	126	Yes
JACK WU	PT10773	Male	51	Yes	No	never	27	7	240	Yes
JESSICA RANGE	PT10854	Male	76	Yes	No	never	47	6	140	Yes
RONALD CRIVELLO JR	PT10973	Female	71	Yes	Yes	never	31	9	130	Yes
RAMON VELASQUEZ	PT10974	Female	47	Yes	No	never	43	6	145	Yes

Q6. Determine the number of patients with heart disease.

```
22  -- Q6 --
23  select * from patients
24  where heart_disease = 'Yes';
```

EmployeeName	Patient_id	gender	age	hypertension	heart_disease	smoking_history	bmi	HbA1c_level	blood_glucose_level	diabetes
Estelle Yancey	PT100013	Male	80	No	Yes	former	27	5	140	No
Stephanie Chang	PT100036	Female	65	Yes	Yes	never	34	8	140	Yes
Marquis D Walker	PT100039	Male	55	No	Yes	former	30	6	300	Yes
PATRICK CORTESE	PT10006	Male	68	No	Yes	never	31	8	160	Yes
NATHAN NG	PT10017	Male	80	No	Yes	former	27	6	126	No
JOAN MCNAMARA	PT10022	Male	78	No	Yes	No Info	27	6	280	Yes
ASHELEY EPPERSON	PT10048	Male	71	No	Yes	ever	28	7	130	No
NATHANIEL FORD	PT101	Female	80	No	Yes	never	25	7	140	No
ENOCH CHU	PT10149	Female	70	No	Yes	former	38	5	155	No
VAGN PETERSEN	PT1015	Female	62	No	Yes	never	42	6	159	No
ETHAN BANFORD	PT1027	Male	65	No	Yes	former	30	5	158	No
ENG ENG CHAN	PT10295	Male	80	No	Yes	former	24	6	130	No
KENNETH KWONG	PT10315	Female	64	Yes	Yes	former	32	8	145	Yes
OMID TALAI	PT10316	Male	70	No	Yes	never	27	6	200	Yes
CAROL RANEY	PT10360	Male	76	Yes	Yes	No Info	30	6	85	No
HECTOR RODRIGUEZ	PT10366	Female	66	No	Yes	not current	27	4	160	No
MICHAEL FONG	PT10373	Female	46	No	Yes	No Info	26	7	90	No
HENRY YEE	PT1038	Female	80	No	Yes	never	15	7	140	No
LOUIE BENAVIDEZ	PT10394	Female	57	No	Yes	current	29	5	200	No

Q7. Group patients by smoking history and count how many smokers and non-smokers there are.



```
26  -- Q7 --
27  select smoking_history,count(*) as number_of_employees
28  from patients
29  where smoking_history in ('Never','Current')
30  group by smoking_history;
```

smoking_history	number_of_employees
never	35092
current	9286

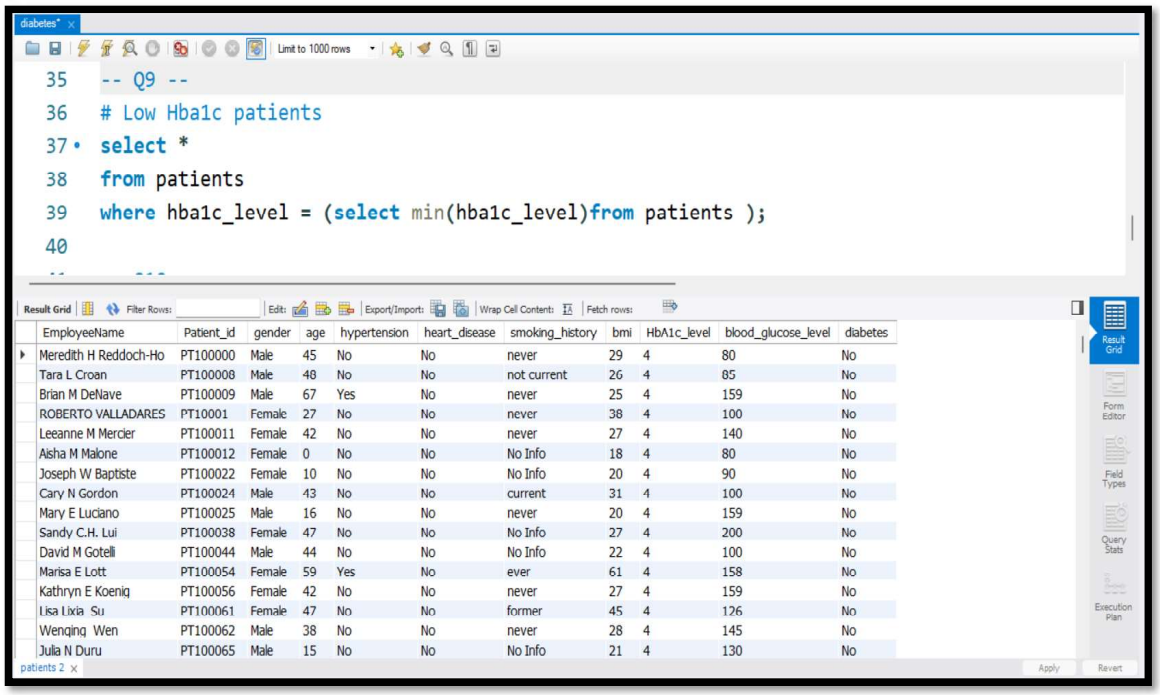
Q8. Retrieve the Patient_ids of patients who have a BMI greater than the average BMI.

```
32  -- Q8 --
33  select * from patients
34  where bmi >(select avg(bmi) from patients);
35
```

EmployeeName	Patient_id	gender	age	hypertension	heart_disease	smoking_history	bmi	HbA1c_level	blood_glucose_level	diabetes
JOHN HOFFMAN	PT10000	Female	46	No	No	No Info	37	6	200	No
Meredith H Reddoch-Ho	PT100000	Male	45	No	No	never	29	4	80	No
Minouche Kandel	PT100001	Male	57	No	No	never	33	6	126	No
ROBERTO VALLADARES	PT10001	Female	27	No	No	never	38	4	100	No
Jennifer J Pascual	PT100010	Male	41	Yes	No	former	30	6	160	No
Kenny Nguyen	PT100019	Female	78	No	No	never	37	6	155	No
Kristen W Peterson	PT100020	Female	74	No	No	current	30	8	130	Yes
Cary N Gordon	PT100024	Male	43	No	No	current	31	4	100	No
Tommy McGowan	PT100027	Female	50	No	No	former	34	6	160	No
MICHAEL STEZ	PT10003	Female	60	No	No	never	38	6	80	No
Lorae C Rose	PT100030	Female	37	No	No	No Info	37	9	160	Yes
Stephanie Chang	PT100036	Female	65	Yes	Yes	never	34	8	140	Yes
Marquis D Walker	PT100039	Male	55	No	Yes	former	30	6	300	Yes
Lily Y Yep	PT100042	Female	38	No	No	No Info	33	5	200	No
Marcel C Barnes	PT100043	Female	51	No	No	current	30	7	160	No
Sharon S Young	PT100046	Female	57	No	No	never	34	6	80	No
Mimi Su	PT100049	Female	56	Yes	No	former	40	5	126	No
CARRIE WONG	PT10005	Male	53	No	No	former	30	6	155	No
Anna E Boyd	PT100051	Male	20	No	No	never	35	5	145	No

Q9. Find the patient with the highest HbA1c level and the patient with the lowest HbA1clevel.

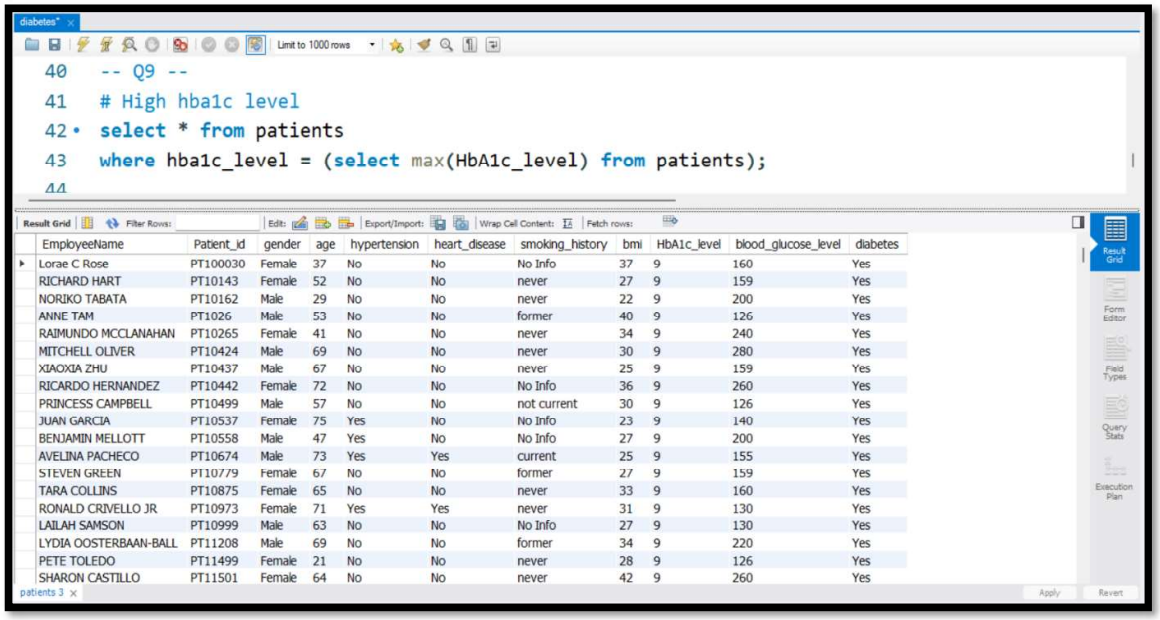
For Low Hba1c_level Patients



```
35  -- Q9 --
36  # Low Hba1c patients
37  select *
38  from patients
39  where hba1c_level = (select min(hba1c_level)from patients );
40
```

EmployeeName	Patient_id	gender	age	hypertension	heart_disease	smoking_history	bmi	HbA1c_level	blood_glucose_level	diabetes
Meredith H Reddoch-Ho	PT100000	Male	45	No	No	never	29	4	80	No
Tara L Croan	PT100008	Male	48	No	No	not current	26	4	85	No
Brian M DeNave	PT100009	Male	67	Yes	No	never	25	4	159	No
ROBERTO VALLADARES	PT10001	Female	27	No	No	never	38	4	100	No
Leeanne M Mercier	PT100011	Female	42	No	No	never	27	4	140	No
Aisha M Malone	PT100012	Female	0	No	No	No Info	18	4	80	No
Joseph W Baptiste	PT100022	Female	10	No	No	No Info	20	4	90	No
Cary N Gordon	PT100024	Male	43	No	No	current	31	4	100	No
Mary E Luciano	PT100025	Male	16	No	No	never	20	4	159	No
Sandy C.H. Lui	PT100038	Female	47	No	No	No Info	27	4	200	No
David M Gotelli	PT100044	Male	44	No	No	No Info	22	4	100	No
Marisa E Lott	PT100054	Female	59	Yes	No	ever	61	4	158	No
Kathryn E Koenig	PT100056	Female	42	No	No	never	27	4	159	No
Lisa Lixia Su	PT100061	Female	47	No	No	former	45	4	126	No
Wenqing Wen	PT100062	Male	38	No	No	never	28	4	145	No
Julia N Duru	PT100065	Male	15	No	No	No Info	21	4	130	No

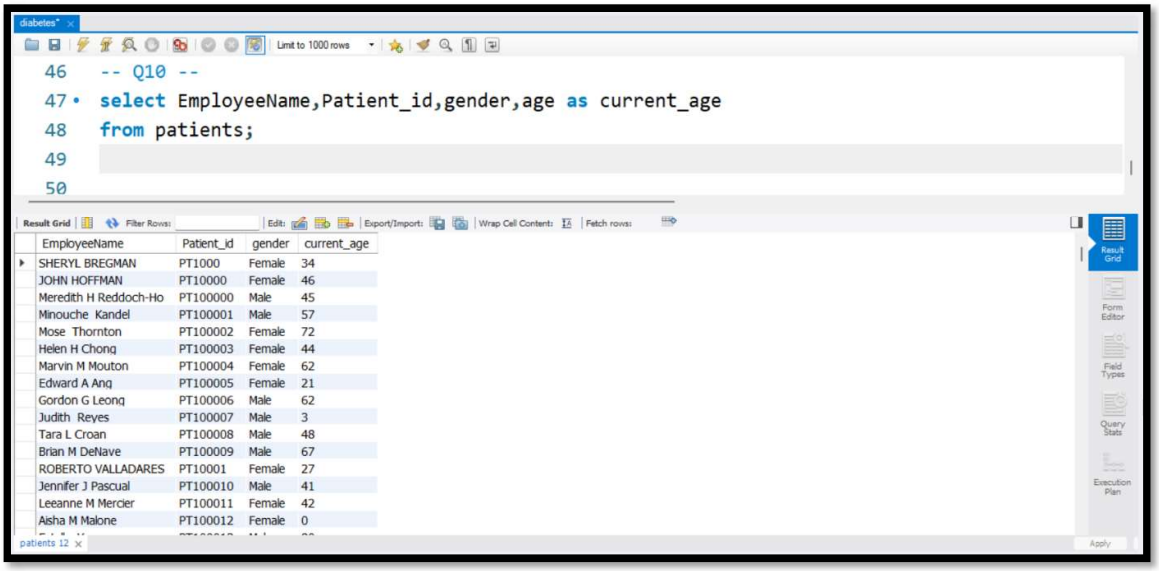
For High Hba1c_level Patients



```
40  -- Q9 --
41  # High hba1c level
42  select * from patients
43  where hba1c_level = (select max(HbA1c_level) from patients);
44
```

EmployeeName	Patient_id	gender	age	hypertension	heart_disease	smoking_history	bmi	HbA1c_level	blood_glucose_level	diabetes
Lorae C Rose	PT100030	Female	37	No	No	No Info	37	9	160	Yes
RICHARD HART	PT10143	Female	52	No	No	never	27	9	159	Yes
NORIKO TABATA	PT10162	Male	29	No	No	never	22	9	200	Yes
ANNE TAM	PT1026	Male	53	No	No	former	40	9	126	Yes
RAIMUNDO MCCLANAHAN	PT10265	Female	41	No	No	never	34	9	240	Yes
MITCHELL OLIVER	PT10424	Male	69	No	No	never	30	9	280	Yes
XIAOXIA ZHU	PT10437	Male	67	No	No	never	25	9	159	Yes
RICARDO HERNANDEZ	PT10442	Female	72	No	No	No Info	36	9	260	Yes
PRINCESS CAMPBELL	PT10499	Male	57	No	No	not current	30	9	126	Yes
JUAN GARCIA	PT10537	Female	75	Yes	No	No Info	23	9	140	Yes
BENJAMIN MELLOTT	PT10558	Male	47	Yes	No	No Info	27	9	200	Yes
AVELINA PACHECO	PT10674	Male	73	Yes	Yes	current	25	9	155	Yes
STEVEN GREEN	PT10779	Female	67	No	No	former	27	9	159	Yes
TARA COLLINS	PT10875	Female	65	No	No	never	33	9	160	Yes
RONALD CRIVELLO JR	PT10973	Female	71	Yes	Yes	never	31	9	130	Yes
LAILAH SAMSON	PT10999	Male	63	No	No	No Info	27	9	130	Yes
LYDIA OOSTERBAAN-BALL	PT11208	Male	69	No	No	former	34	9	220	Yes
PETE TOLEDO	PT11499	Female	21	No	No	never	28	9	126	Yes
SHARON CASTILLO	PT11501	Female	64	No	No	never	42	9	260	Yes

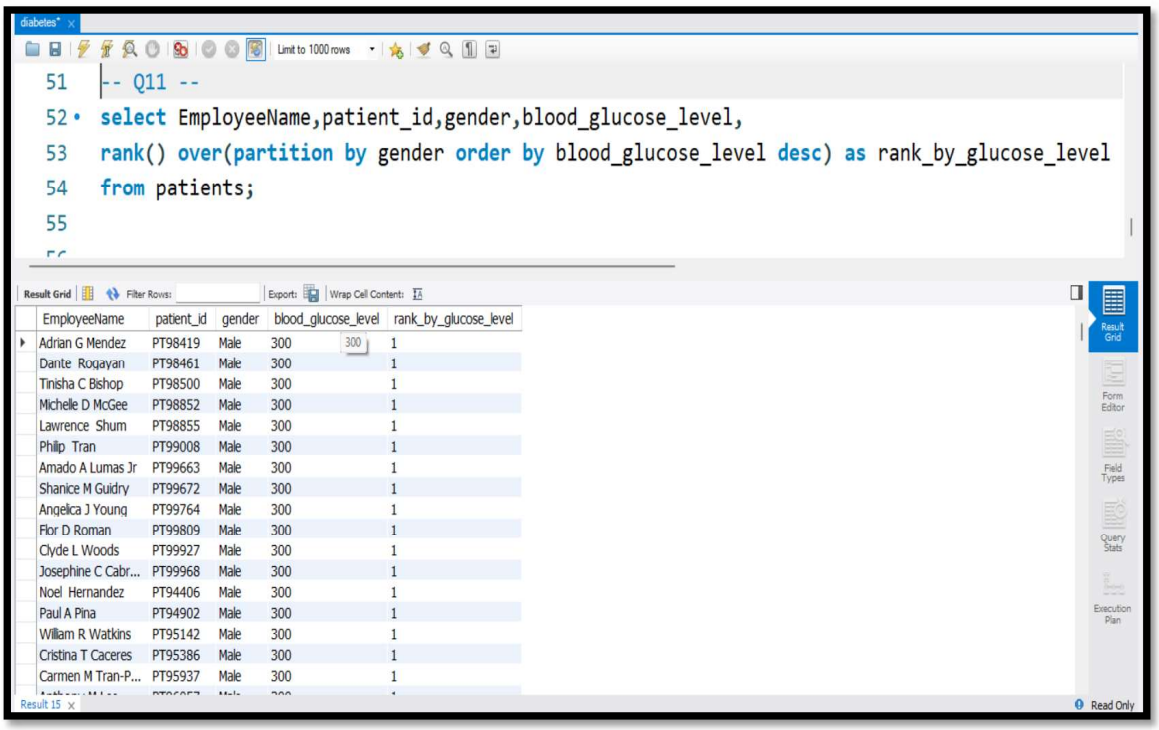
Q10. Calculate the age of patients in years (assuming the current date as of now).



```
46  -- Q10 --
47  select EmployeeName,Patient_id,gender,age as current_age
48  from patients;
49
50
```

EmployeeName	Patient_id	gender	current_age
SHERYL BREGMAN	PT1000	Female	34
JOHN HOFFMAN	PT10000	Female	46
Meredith H Reddoch-Ho	PT100000	Male	45
Minouche Kandel	PT100001	Male	57
Mose Thornton	PT100002	Female	72
Helen H Chong	PT100003	Female	44
Marvin M Mouton	PT100004	Female	62
Edward A Ang	PT100005	Female	21
Gordon G Leong	PT100006	Male	62
Judith Reyes	PT100007	Male	3
Tara L Croan	PT100008	Male	48
Brian M DeNave	PT100009	Male	67
ROBERTO VALLADARES	PT10001	Female	27
Jennifer J Pascual	PT100010	Male	41
Leeanne M Mercier	PT100011	Female	42
Aisha M Malone	PT100012	Female	0

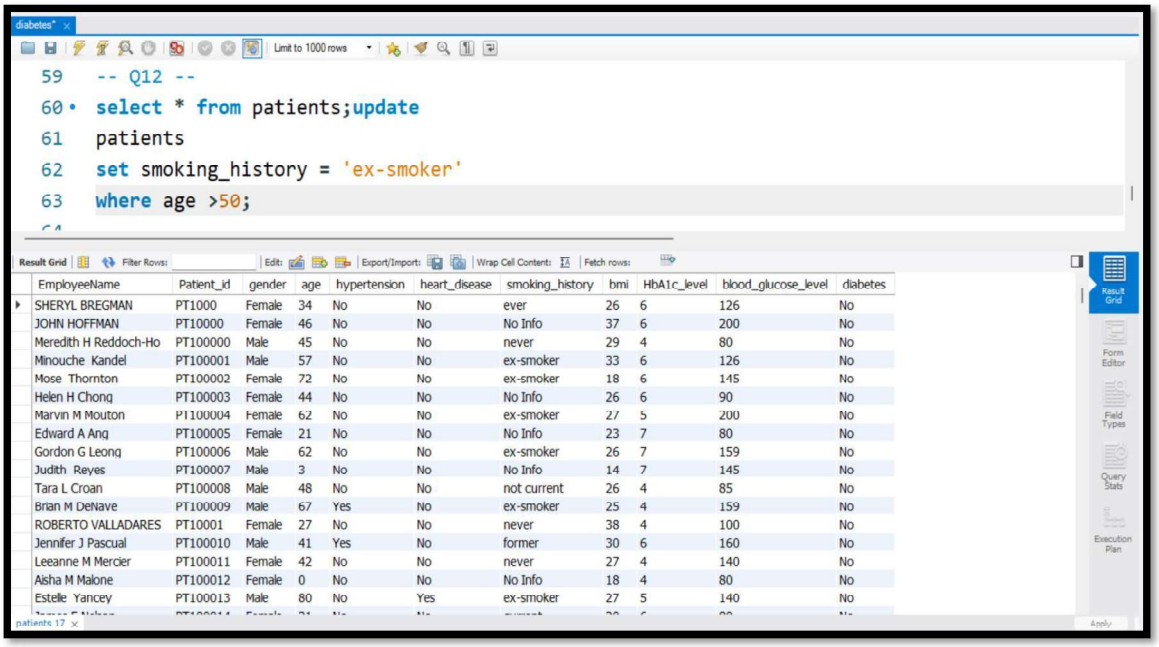
Q11.Rank patients by blood glucose level within each gender group.



```
51  -- Q11 --
52  select EmployeeName,patient_id,gender,blood_glucose_level,
53  rank() over(partition by gender order by blood_glucose_level desc) as rank_by_glucose_level
54  from patients;
55
```

EmployeeName	patient_id	gender	blood_glucose_level	rank_by_glucose_level
Adrian G Mendez	PT98419	Male	300	1
Dante Rogayan	PT98461	Male	300	1
Tinisha C Bishop	PT98500	Male	300	1
Michelle D McGee	PT98852	Male	300	1
Lawrence Shum	PT98855	Male	300	1
Philip Tran	PT99008	Male	300	1
Amado A Lumas Jr	PT99663	Male	300	1
Shanice M Guidry	PT99672	Male	300	1
Angelica J Young	PT99764	Male	300	1
Flor D Roman	PT99809	Male	300	1
Clyde L Woods	PT99927	Male	300	1
Josephine C Cabr...	PT99968	Male	300	1
Noel Hernandez	PT94406	Male	300	1
Paul A Pina	PT94902	Male	300	1
William R Watkins	PT95142	Male	300	1
Cristina T Caceres	PT95386	Male	300	1
Carmen M Tran-P...	PT95937	Male	300	1

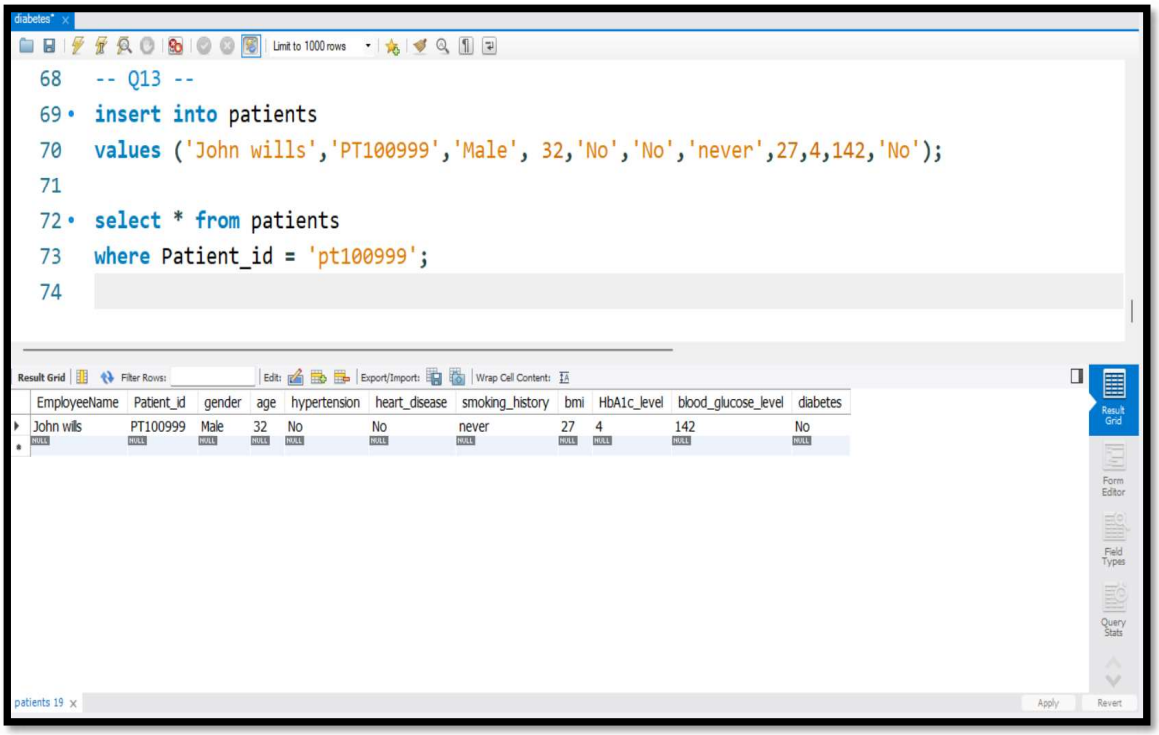
Q12. Update the smoking history of patients who are older than 50 to "Ex-smoker."



```
-- Q12 --
select * from patients;update
patients
set smoking_history = 'ex-smoker'
where age >50;
```

EmployeeName	Patient_id	gender	age	hypertension	heart_disease	smoking_history	bmi	HbA1c_level	blood_glucose_level	diabetes
SHERYL BREGMAN	PT1000	Female	34	No	No	ever	26	6	126	No
JOHN HOFFMAN	PT10000	Female	46	No	No	No Info	37	6	200	No
Meredith H Reddoch-Ho	PT100000	Male	45	No	No	never	29	4	80	No
Minouche Kandel	PT100001	Male	57	No	No	ex-smoker	33	6	126	No
Mose Thornton	PT100002	Female	72	No	No	ex-smoker	18	6	145	No
Helen H Chong	PT100003	Female	44	No	No	No Info	26	6	90	No
Marvin M Mouton	PT100004	Female	62	No	No	ex-smoker	27	5	200	No
Edward A Ang	PT100005	Female	21	No	No	No Info	23	7	80	No
Gordon G Leong	PT100006	Male	62	No	No	ex-smoker	26	7	159	No
Judith Reyes	PT100007	Male	3	No	No	No Info	14	7	145	No
Tara L Croan	PT100008	Male	48	No	No	not current	26	4	85	No
Brian M DeNave	PT100009	Male	67	Yes	No	ex-smoker	25	4	159	No
ROBERTO VALLADARES	PT10001	Female	27	No	No	never	38	4	100	No
Jennifer J Pascual	PT100010	Male	41	Yes	No	former	30	6	160	No
Leeanne M Mercier	PT100011	Female	42	No	No	never	27	4	140	No
Aisha M Malone	PT100012	Female	0	No	No	No Info	18	4	80	No
Estelle Yancey	PT100013	Male	80	No	Yes	ex-smoker	27	5	140	No

Q13. Insert a new patient into the database with sample data.

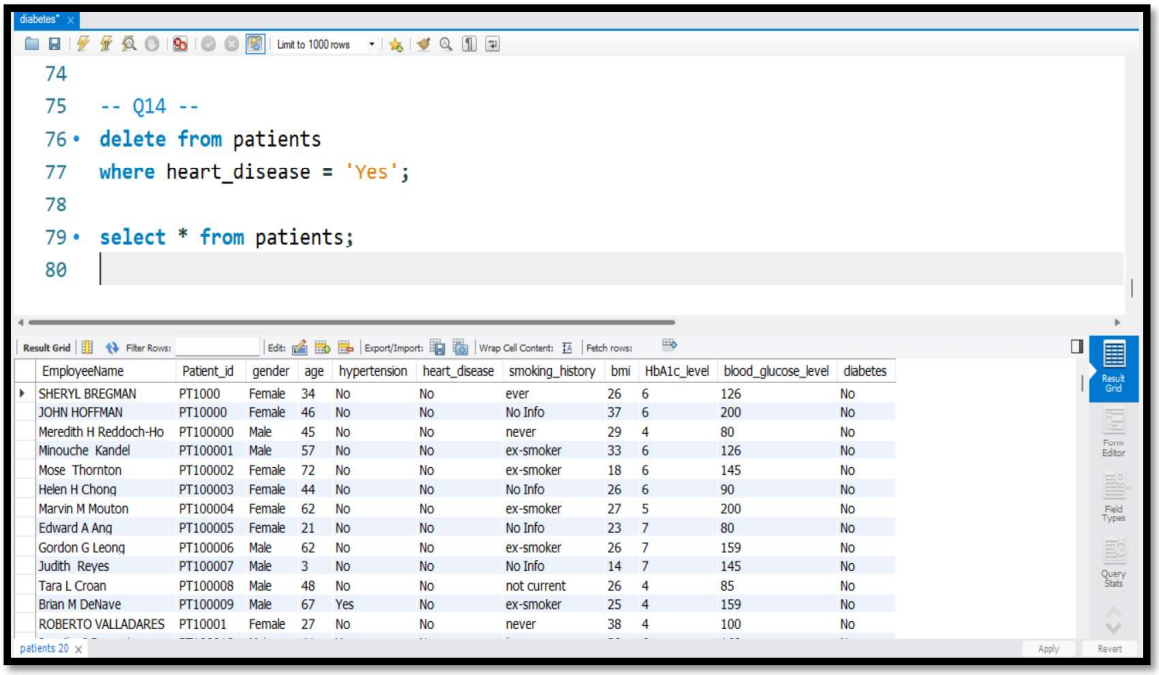


```
-- Q13 --
insert into patients
values ('John wills','PT100999','Male', 32,'No','No','never',27,4,142,'No');

select * from patients
where Patient_id = 'pt100999';
```

EmployeeName	Patient_id	gender	age	hypertension	heart_disease	smoking_history	bmi	HbA1c_level	blood_glucose_level	diabetes
John wills	PT100999	Male	32	No	No	never	27	4	142	No

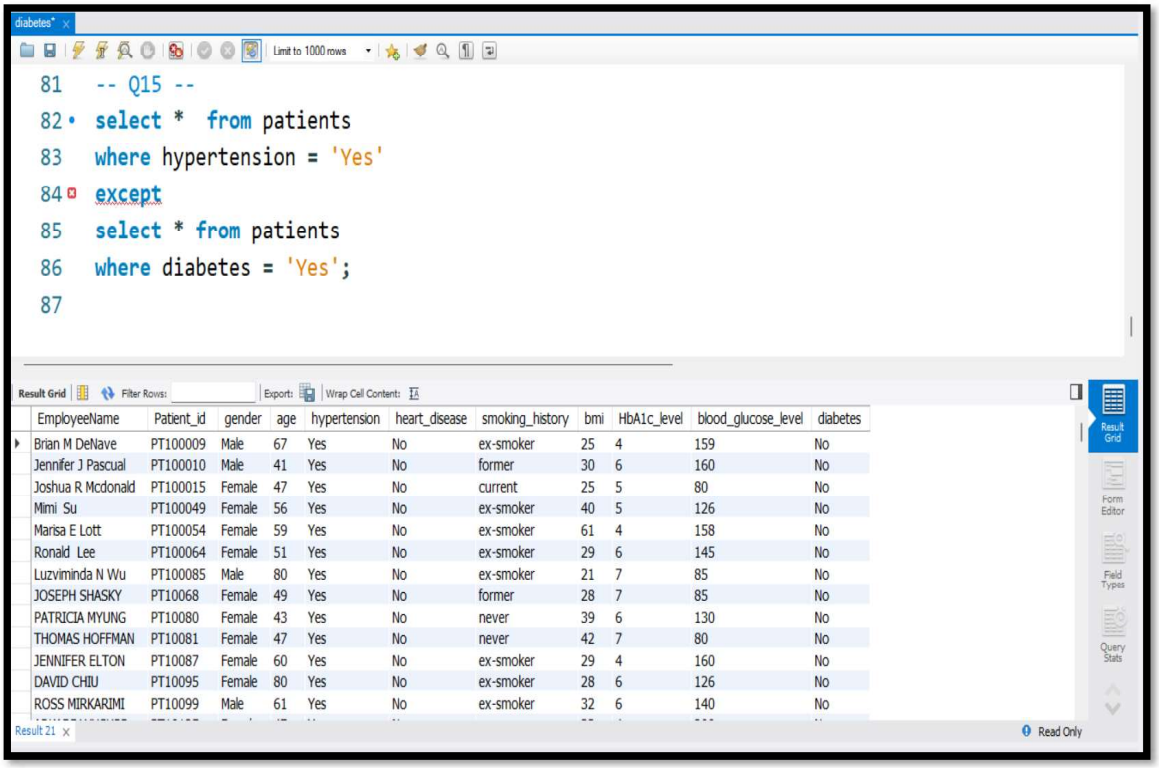
Q14.Delete all patients with heart disease from the database.



```
74
75   -- Q14 --
76 • delete from patients
77   where heart_disease = 'Yes';
78
79 • select * from patients;
80
```

EmployeeName	Patient_id	gender	age	hypertension	heart_disease	smoking_history	bmi	HbA1c_level	blood_glucose_level	diabetes
SHERYL BREGMAN	PT1000	Female	34	No	No	ever	26	6	126	No
JOHN HOFFMAN	PT10000	Female	46	No	No	No Info	37	6	200	No
Meredith H Reddoch-Ho	PT100000	Male	45	No	No	never	29	4	80	No
Minouche Kandel	PT100001	Male	57	No	No	ex-smoker	33	6	126	No
Mose Thornton	PT100002	Female	72	No	No	ex-smoker	18	6	145	No
Helen H Chong	PT100003	Female	44	No	No	No Info	26	6	90	No
Marvin M Mouton	PT100004	Female	62	No	No	ex-smoker	27	5	200	No
Edward A Ang	PT100005	Female	21	No	No	No Info	23	7	80	No
Gordon G Leong	PT100006	Male	62	No	No	ex-smoker	26	7	159	No
Judith Reyes	PT100007	Male	3	No	No	No Info	14	7	145	No
Tara L Croan	PT100008	Male	48	No	No	not current	26	4	85	No
Brian M DeNave	PT100009	Male	67	Yes	No	ex-smoker	25	4	159	No
ROBERTO VALLADARES	PT10001	Female	27	No	No	never	38	4	100	No

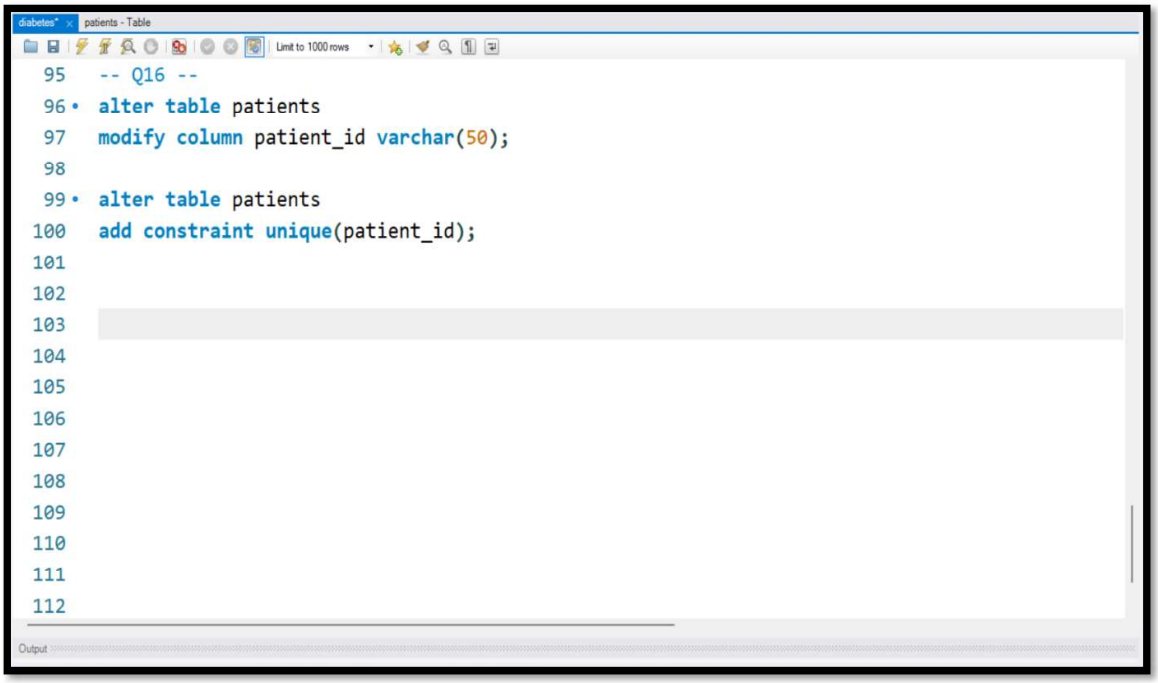
Q15.Find patients who have hypertension but not diabetes using the EXCEPT operator.



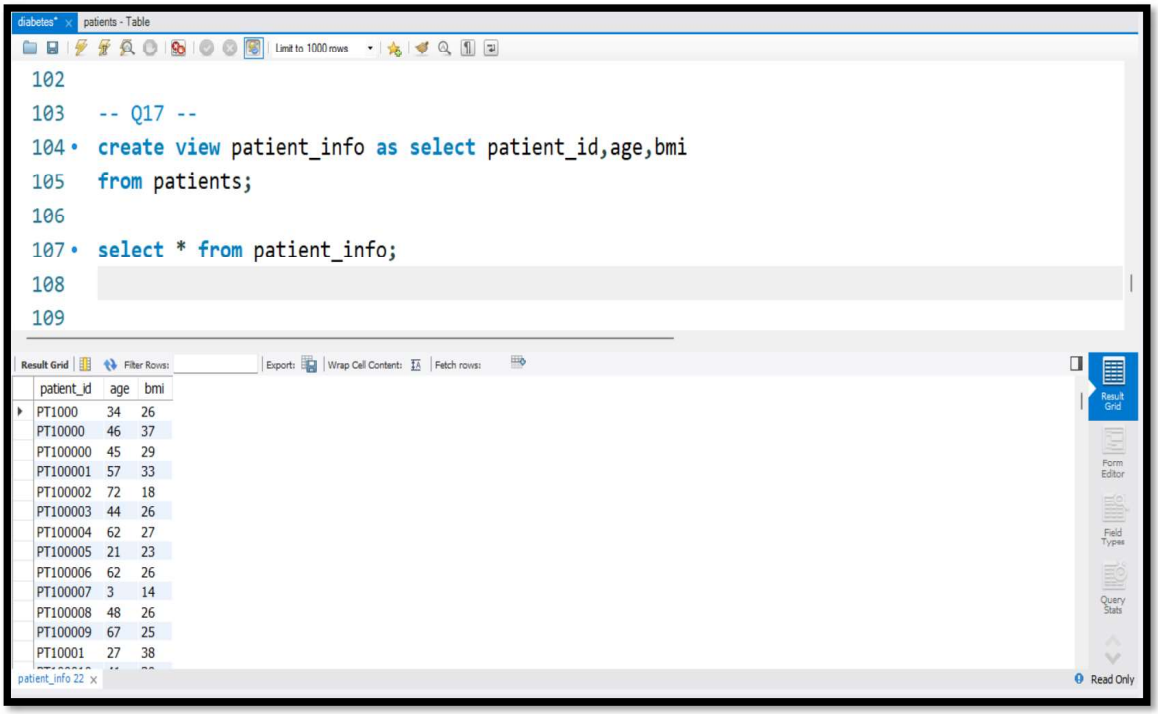
```
81   -- Q15 --
82 • select *  from patients
83   where hypertension = 'Yes'
84   except
85   select * from patients
86   where diabetes = 'Yes';
87
```

EmployeeName	Patient_id	gender	age	hypertension	heart_disease	smoking_history	bmi	HbA1c_level	blood_glucose_level	diabetes
Brian M DeNave	PT100009	Male	67	Yes	No	ex-smoker	25	4	159	No
Jennifer J Pascual	PT100010	Male	41	Yes	No	former	30	6	160	No
Joshua R Mcdonald	PT100015	Female	47	Yes	No	current	25	5	80	No
Mimi Su	PT100049	Female	56	Yes	No	ex-smoker	40	5	126	No
Marisa E Lott	PT100054	Female	59	Yes	No	ex-smoker	61	4	158	No
Ronald Lee	PT100064	Female	51	Yes	No	ex-smoker	29	6	145	No
Luzviminda N Wu	PT100085	Male	80	Yes	No	ex-smoker	21	7	85	No
JOSEPH SHASKY	PT10068	Female	49	Yes	No	former	28	7	85	No
PATRICIA MYUNG	PT10080	Female	43	Yes	No	never	39	6	130	No
THOMAS HOFFMAN	PT10081	Female	47	Yes	No	never	42	7	80	No
JENNIFER ELTON	PT10087	Female	60	Yes	No	ex-smoker	29	4	160	No
DAVID CHIU	PT10095	Female	80	Yes	No	ex-smoker	28	6	126	No
ROSS MIRKARIMI	PT10099	Male	61	Yes	No	ex-smoker	32	6	140	No

Q16.Define a unique constraint on the "patient_id" column to ensure its values are unique.



Q17.Create a view that displays the Patient_ids, ages, and BMI of patients.



Q18.Suggest improvements in the database schema to reduce data redundancy and improve data integrity.

To reduce the Data redundancy and improve data integrity in a database Schema, consider the following Improvements:

- **Normalize the database**: Normalization is the process of organizing data in a database to minimize redundancy and improve data integrity. By normalizing the database, you can reduce the amount of duplicate data and ensure that each piece of data is stored in only one place. This can help to improve data consistency and accuracy.

- **Use primary and foreign keys**: Primary keys should be unique identifiers for each record in a table, while foreign keys should reference the primary key of another table. This can help to ensure that related data is linked correctly and that there are no orphaned records.

- **Eliminate transitive dependencies**: Transitive dependencies occur when a piece of data is dependent on another piece of data that is itself dependent on a third piece of data. This can lead to redundancy and inconsistencies in the data. By eliminating transitive dependencies, you can reduce redundancy and improve data integrity.

- **Use views**: Views are virtual tables that are derived from existing tables in the database. They can be used to provide a more organized and normalized view of the data, reducing redundancy, and improving data integrity by presenting a single source of truth for related information.

- **Implement referential integrity**: Referential integrity ensures that foreign keys reference valid primary keys in other tables, preventing orphaned records and ensuring that related data is linked correctly. By implementing referential integrity, you can improve data consistency and accuracy, reducing redundancy by ensuring that related information is stored in only one place.

- **Use triggers**: Triggers are automated procedures that are executed when certain events occur in the database, such as inserting or updating a record. They can be used to enforce business rules and ensure data integrity, reducing redundancy by ensuring that related information is updated automatically when changes are made to other records.

- **Use indexes**: Indexes can be used to speed up database queries by providing a quick way to locate specific records based on certain criteria. By using indexes, you can reduce redundancy by ensuring that related information is easily accessible without having to scan through large amounts of unnecessary data.

- **Implement data validation**: Data validation ensures that the data entered the database meets certain criteria, such as being within a certain range or having a certain format. By implementing data validation, you can reduce redundancy by ensuring that related information is entered consistently and accurately, reducing the need for duplicate records or corrections due to errors in inputting the data.

- **Document the Schema**: Maintain comprehensive documentation explaining the database schema, relationships, constraints, and any considerations for developers and users.

- **Regularly Review and Update**: Periodically review the database schema for optimizations and improvements. Adapt the schema based on evolving business needs and performance considerations.

- By implementing these improvements, you can enhance data integrity, reduce redundancy, improve query performance, and ensure better overall database management and scalability.

Q19.Explain how you can optimize the performance of SQL queries on this dataset.

Optimizing the performance of SQL queries involves various strategies aimed at improving query execution speed, reducing resource consumption, and enhancing overall database performance. Here are several ways to optimize SQL queries on a dataset:

- **Create Indexes**: Identify columns frequently used in **WHERE** clauses, joins, or sorting operations, and create indexes on these columns. Proper indexing can significantly speed up query execution by reducing the number of rows that need to be scanned.

- **Use Efficient Joins**: Choose appropriate join types (e.g., INNER, LEFT, RIGHT) based on the relationship between tables. Consider using explicit join syntax rather than implicit joins.

- Instead of selecting all columns using SELECT *, specify only the required columns. This reduces the amount of data retrieved and improves query performance.

- Optimize **WHERE** Clauses: Construct efficient WHERE clauses by using indexed columns, avoiding unnecessary comparisons, and ensuring the use of appropriate comparison operators.

- **Partition Large Tables**: For very large tables, consider partitioning based on specific criteria (e.g., by date ranges) to improve manageability and query performance.
- **Cluster Tables**: Cluster tables based on commonly accessed columns to physically group related data together, reducing disk I/O and enhancing retrieval speed.

Proper Data Modelling and Normalization:

- **Normalize Data**: Organize data into normalized structures (e.g., 3rd normal form) to minimize redundancy and improve query efficiency.

- **De-normalize for Performance**: In some cases, denormalization can improve performance by reducing complex joins. However, it should be carefully implemented to avoid data integrity issues.

Consider Query Caching:

- **Ensure Sufficient Resources**: Ensure that the database server has adequate memory, CPU, and storage to handle query processing efficiently.

- **Optimize Server Settings**: Adjust database server settings such as buffer sizes, memory allocation, and query timeouts to optimize performance.

Regular Database Maintenance:

- Perform Regular Maintenance: Regularly update statistics, reorganize indexes, and perform database maintenance tasks to keep the database in optimal condition.