Diabetes Prediction Dataset Analysis SQL Project

The diabetes_prediction_dataset file contains medical and demographic data of patients along with their diabetes status, whether positive or negative.
It consists of various features such as age, gender, body mass index (BMI), hypertension, heart disease, smoking history, HbA1c level, and blood glucose level.

Retrieve the Patient_id and ages of all patients.

```
1 • use cv;
2 • SELECT * FROM diabetes_prediction;
3
4 • select Patient_id , age from diabetes_prediction ;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

Patient_id	age
PT102	54
PT103	28
PT104	36
PT106	20
PT107	44
PT108	79
PT109	42
PT110	32
PT111	53
PT112	54
PT113	78
PT114	67

diabetes_prediction 11 ×

Result Grid | Form Editor | Field Types | Read Only

Output

Action Output

#	Time	Action	Message
50	11:01:17	SELECT * FROM diabetes_prediction LIMIT 0, 1000	1000 row(s) returned
51	11:01:22	Insert into diabetes_prediction values ('Michael','PT101679','Male',52,0,1, 'never',27.89,6.7,89,1)	1 row(s) affected
52	11:30:33	select Patient_id , age from diabetes_prediction LIMIT 0, 1000	1000 row(s) returned

Select all female patients who are older than 40

```
1 • use cv;
2 • select * from diabetes_prediction;
3
4
5 • select * from diabetes_prediction where gender = 'Female' and age > 40;
6
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

EmployeeName	Patient_id	gender	age	hypertension	heart_disease	smoking_history	bmi	HbA1c_level	blood_glucose_level	diabetes
NATHANIEL FORD	PT101	Female	80	0	1	Ex-smoker	25.19	6.6	140	0
GARY JIMENEZ	PT102	Female	54	0	0	Ex-smoker	27.32	6.6	80	0
ALSON LEE	PT107	Female	44	0	0	never	19.31	6.5	200	1
DAVID KUSHNER	PT108	Female	79	0	0	Ex-smoker	23.86	5.7	85	0
ARTHUR KENNEY	PT111	Female	53	0	0	Ex-smoker	27.32	6.1	85	0
PATRICIA JACKSON	PT112	Female	54	0	0	Ex-smoker	54.7	6	100	0
EDWARD HARRINGTON	PT113	Female	78	0	0	Ex-smoker	36.05	5	130	0
JOHN MARTIN	PT114	Female	67	0	0	Ex-smoker	25.69	5.8	200	0
DAVID FRANKLIN	PT115	Female	76	0	0	Ex-smoker	27.32	5	160	0
SEBASTIAN WONG	PT118	Female	42	0	0	never	24.48	5.7	158	0
MARTY ROSS	PT119	Female	42	0	0	No Info	27.32	5.7	80	0
GEORGE GARCIA	PT123	Female	69	0	0	Ex-smoker	21.24	4.8	85	0

Result Grid | Form Editor | Field Types | Read Only

diabetes_prediction 8

Output

Action Output

#	Time	Action	Message
8	22:26:58	select * from diabetes_prediction where gender = 'Female' and age > 40 LIMIT 0, 1000	1000 row(s) returned
9	22:26:59	select * from diabetes_prediction where gender = 'Female' and age > 40 LIMIT 0, 1000	1000 row(s) returned
10	22:27:01	select * from diabetes_prediction where gender = 'Female' and age > 40 LIMIT 0, 1000	1000 row(s) returned

Calculate the average BMI of patients.

```
1 • use cv;
2 • select * from diabetes_prediction;
3
4
5 • select avg(bmi) from diabetes_prediction;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

avg(bmi)
27.32076709999422

Result Grid | Form Editor | Field Types | Read Only

Result 11

Output

Action Output

#	Time	Action	Message
11	22:34:07	select avg(bmi) from diabetes_prediction LIMIT 0, 1000	1 row(s) returned
12	22:34:56	select * from diabetes_prediction LIMIT 0, 1000	1000 row(s) returned
13	22:35:24	select avg(bmi) from diabetes_prediction LIMIT 0, 1000	1 row(s) returned

List patients in descending order of blood glucose levels

```sql
use cv;
select * from diabetes_prediction;


select EmployeeName,Patient_id , blood_glucose_level from diabetes_prediction
order by blood_glucose_level desc;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

EmployeeName	Patient_id	blood_glucose_level
Gilbert J Fragoso	PT99638	300
Amado A Lumas Jr	PT99663	300
Shanice M Guidry	PT99672	300
Angelica J Young	PT99764	300
Flor D Roman	PT99809	300
Clyde L Woods	PT99927	300
Josephine C Cabrera	PT99968	300
Marquis D Walker	PT100039	300
Brenda G Velasquez	PT89459	300
Haroon Ahmad	PT89934	300
Marc S Slavin	PT93637	300
Silvia Woo	PT91896	300

Result Grid | Form Editor | Field Types | Read Only

diabetes_prediction 13

Output

Action Output

#	Time	Action	Message
13	22:35:24	select avg(bmi) from diabetes_prediction LIMIT 0, 1000	1 row(s) returned
14	22:48:43	select * from diabetes_prediction order by blood_glucose_level desc LIMIT 0, 1000	1000 row(s) returned
15	22:50:30	select EmployeeName,Patient_id , blood_glucose_level from diabetes_prediction order by blood_glucose_level desc ...	1000 row(s) returned

Find patients who have hypertension and diabetes

```
1 • use cv;
2 • select * from diabetes_prediction;
3
4 • select EmployeeName , Patient_id from diabetes_prediction
5   where hypertension = 1 and diabetes= 1 ;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

EmployeeName	Patient_id
JONES WONG	PT139
PATRIC STEELE	PT205
ARTHUR STELLINI	PT343
CHAD LAW	PT355
CATHERINE JAMES	PT451
JOHN HART	PT565
JOHN BARKER	PT567
ROBERT BONNET	PT632
VITANI BENJAMIN	PT727
LANNIE ADELMAN	PT828
JOEL DELIZONNA	PT852
KAREN KUBICK	PT861

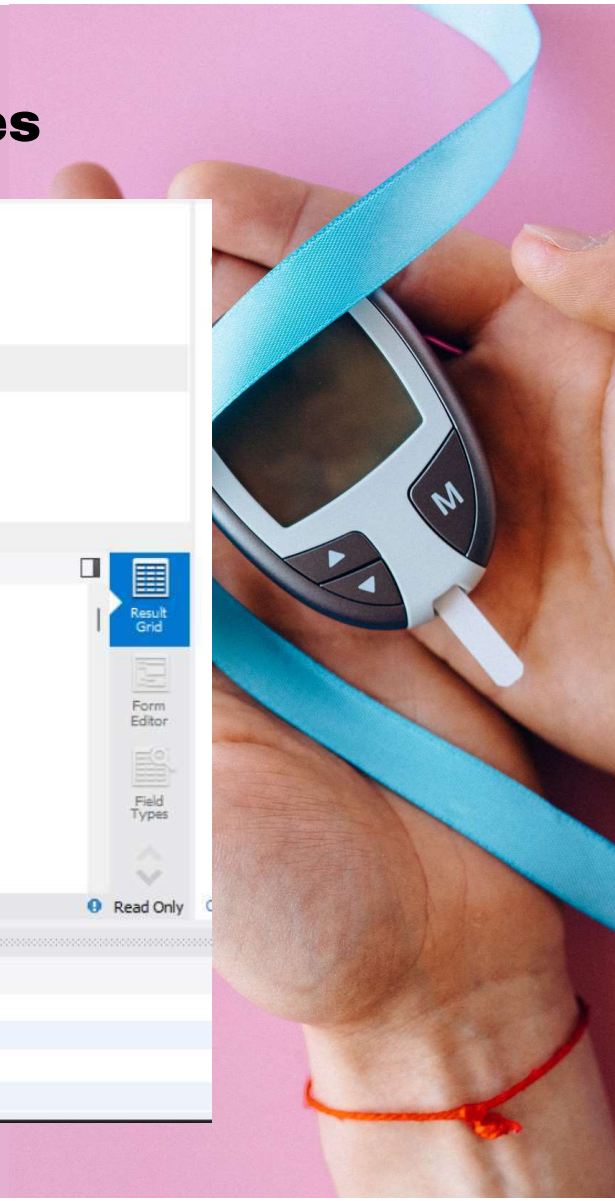
Result Grid | Form Editor | Field Types | Read Only

diabetes_prediction 18

Output

Action Output

#	Time	Action	Message
18	23:08:33	select EmployeeName , Patient_id from diabetes_prediction where hypertension = 1 and diabetes= 1 LIMIT 0, 1000	1000 row(s) returned
19	23:08:59	select EmployeeName , Patient_id from diabetes_prediction where hypertension = 1 and diabetes= 1 LIMIT 0, 1000	1000 row(s) returned
20	23:09:21	select EmployeeName , Patient_id from diabetes_prediction where hypertension = 1 and diabetes= 1 LIMIT 0, 1000	1000 row(s) returned



Determine the number of patients with heart disease

```
1 • use cv;
2 • select * from diabetes_prediction;
3
4 • select count(*) as Patient_with_heart_disease from diabetes_prediction
5   where heart_disease = 1;
6
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Patient_with_heart_disease
3942

Result Grid | Form Editor | Field Types | Read Only

Result 20

Output

Action Output

#	Time	Action	Message
21	23:10:16	select count(Patient_id) from diabetes_prediction where heart_disease = 1 LIMIT 0, 1000	1 row(s) returned
22	23:11:29	select count(*) as Patient_with_heart_disease from diabetes_prediction where heart_disease = 1 LIMIT 0, 1000	1 row(s) returned
23	23:11:40	select count(*) as Patient_with_heart_disease from diabetes_prediction where heart_disease = 1 LIMIT 0, 1000	1 row(s) returned

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

```
1 • use cv;
2 • select * from diabetes_prediction;
3
4 • select smoking_history ,count(smoking_history) as count
5   from diabetes_prediction where smoking_history in ("current","never")
6   group by smoking_history ;
7
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

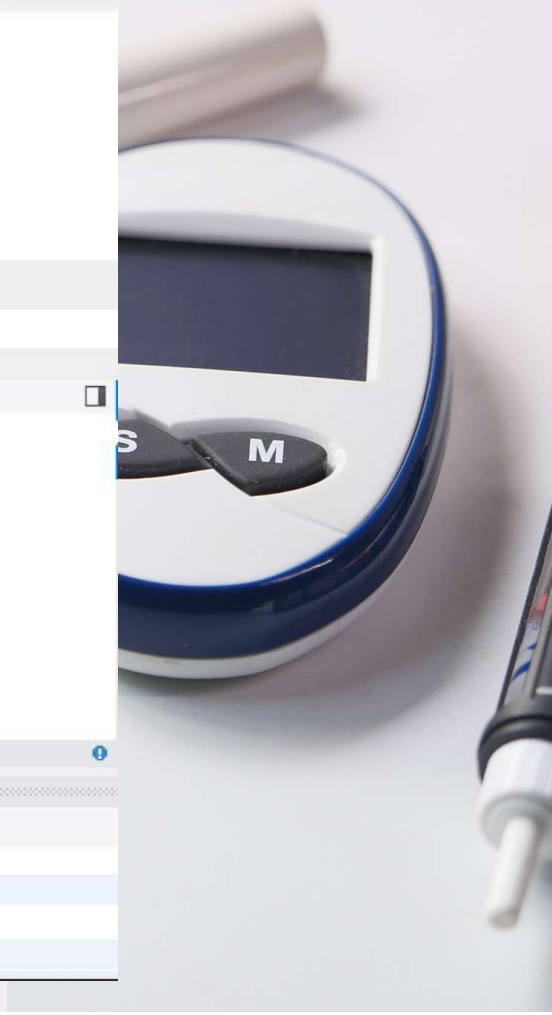
smoking_history	count
never	21469
current	6080

Result 22 ×

Output

Action Output

#	Time	Action	Message
24	23:17:52	select employeeName,Patient_id from diabetes_prediction where bmi > (select avg(bmi) from diabetes_prediction) LI...	1000 row(s) returned
25	23:19:22	select smoking_history ,count(smoking_history) as count from diabetes_prediction group by smoking_history LIMIT 0,...	7 row(s) returned
26	23:20:51	select smoking_history ,count(smoking_history) as count from diabetes_prediction where smoking_history in ("current...	2 row(s) returned



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

```
1 • use cv;
2 • select * from diabetes_prediction;
3
4 • select Patient_id from diabetes_prediction
5   where bmi > (select avg(bmi) from diabetes_prediction);
6
7
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

Patient_id
PT109
PT112
PT113
PT117
PT121
PT124
PT126
PT128
PT131
PT140
PT143
PT144

diabetes_prediction 24

Output

Action Output

#	Time	Action	Message
26	23:20:51	select smoking_history ,count(smoking_history) as count from diabetes_prediction where smoking_history in ("current...	2 row(s) returned
27	23:23:38	select Patient_id from diabetes_prediction where bmi > (select avg(bmi) from diabetes_prediction) LIMIT 0, 1000	1000 row(s) returned
28	23:23:54	select Patient_id from diabetes_prediction where bmi > (select avg(bmi) from diabetes_prediction) LIMIT 0, 1000	1000 row(s) returned

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

```
4 • SELECT * FROM diabetes_prediction ORDER BY hba1c_level DESC;
5   -- Highest HbA1c level
6
7 • SELECT * FROM diabetes_prediction ORDER BY hba1c_level;
8   -- Lowest HbA1c level
9
10
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

EmployeeName	Patient_id	gender	age	hypertension	heart_disease	smoking_history	bmi	HbA1c_level	blood_glucose_level	diabetes
Seth I Rubenstein	PT98911	Female	60	0	0	Ex-smoker	40.18	9	300	1
Michelle K Shinn	PT97928	Female	55	0	0	Ex-smoker	46	9	145	1
George B Peng	PT99841	Male	80	0	1	Ex-smoker	25.76	9	220	1
Sidney Green	PT99807	Male	54	1	0	Ex-smoker	23.06	9	126	1
Angelica J Young	PT99764	Male	80	0	0	Ex-smoker	34	9	300	1
Miao Ling Huang	PT99613	Male	60	0	0	Ex-smoker	27.32	9	280	1
James Portlock	PT98276	Male	48	1	0	former	28.35	9	130	1
Idalia R Farina	PT97708	Female	80	0	0	Ex-smoker	25.86	9	300	1
Medrine A Baltazar	PT99442	Female	80	0	0	Ex-smoker	20.99	9	126	1
Frank A Johnson-III	PT98061	Male	67	0	0	Ex-smoker	31.07	9	160	1
Yolanda Broussard	PT99298	Female	61	0	0	Ex-smoker	22.95	9	240	1
Ralph R Garcia	PT97789	Female	73	0	0	Ex-smoker	27.32	9	220	1

diabetes_prediction 32 ×

Result Grid | Form Editor | Field Types | Read Only

Output

Action Output

#	Time	Action	Message
35	23:28:31	SELECT * FROM diabetes_prediction ORDER BY hba1c_level DESC LIMIT 0, 1000	1000 row(s) returned
36	23:32:21	SELECT * FROM diabetes_prediction ORDER BY hba1c_level LIMIT 0, 1000	1000 row(s) returned
37	23:32:29	SELECT * FROM diabetes_prediction ORDER BY hba1c_level DESC LIMIT 0, 1000	1000 row(s) returned

Rank patients by blood glucose level within each gender group

```
1 • use cv;
2 • select * from diabetes_prediction;
3
4 • SELECT EmployeeName , Patient_id , Gender , age , blood_glucose_level,
5   RANK() OVER (partition by gender order by  blood_glucose_level)
6   as glucose_rank FROM diabetes_prediction ;
7
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

EmployeeName	Patient_id	Gender	age	blood_glucose_level	glucose_rank
Judi Soto	PT99994	Female	20	80	1
Benny M Choi	PT99774	Female	57	80	1
Tania Garcia	PT99716	Female	3	80	1
Trina R Pellette	PT99702	Female	19	80	1
Jensa Woo	PT100031	Female	54	80	1
James E Nelson	PT100014	Female	31	80	1
Francis C Cheung	PT99751	Female	55	80	1
Eve Bekker	PT100070	Female	30	80	1
Edward A Ang	PT100005	Female	21	80	1
Joshua R Mcdonald	PT100015	Female	47	80	1
Michael T Feist	PT99910	Female	34	80	1
Aisha M Malone	PT100012	Female	0	80	1

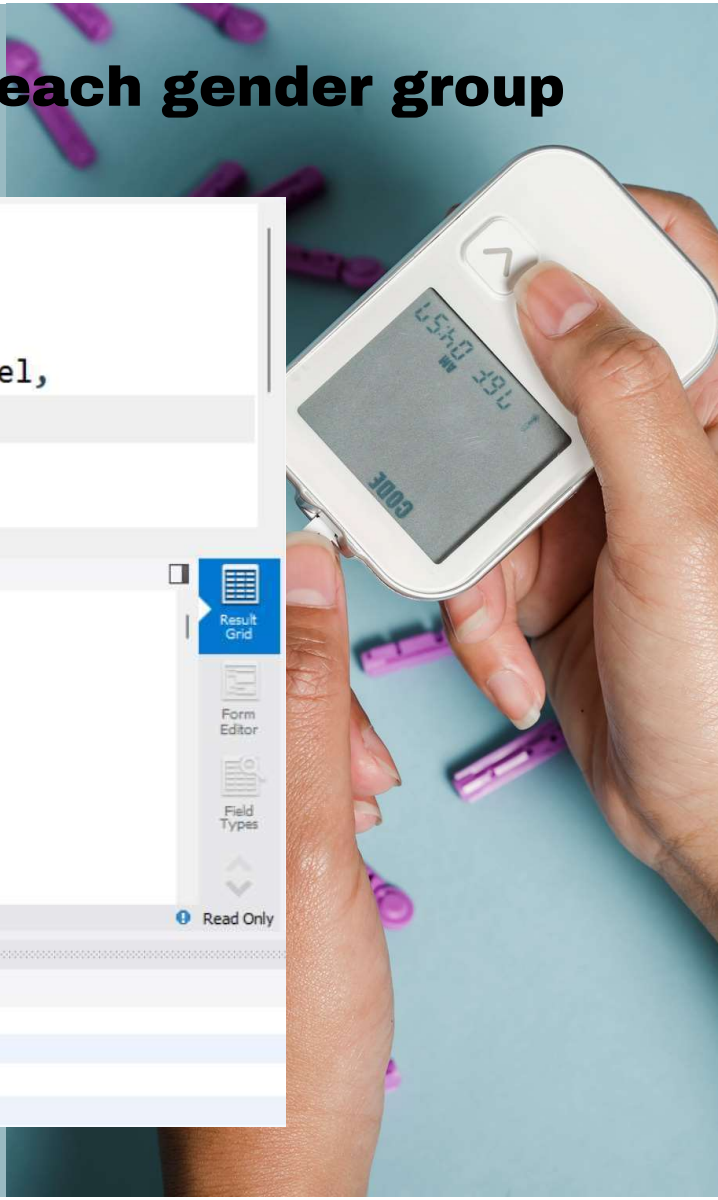
Result 33 | Read Only

Result Grid | Form Editor | Field Types

Output

Action Output

#	Time	Action	Message
40	23:33:53	SELECT Patient_id FROM diabetes_prediction WHERE hba1c_level = (SELECT MAX(hba1c_level) FROM diabetes...	1000 row(s) returned
41	23:34:45	SELECT patient_id, employeeName, gender, blood_glucose_level, RANK() OVER (PARTITION BY gen...	100000 row(s) returned
42	23:38:56	SELECT EmployeeName , Patient_id , Gender , age , blood_glucose_level, RANK() OVER (partition by gender order ...	100000 row(s) returned



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

```
2 •  select * from diabetes_prediction;
3
4 •  update diabetes_prediction set smoking_history = 'Ex-Smoker'
5    where age > 50;
6
7
8
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

EmployeeName	Patient_id	gender	age	hypertension	heart_disease	smoking_history	bmi	HbA1c_level	blood_glucose_level	diabetes
NATHANIEL FORD	PT101	Female	80	0	1	Ex-Smoker	25.19	6.6	140	0
GARY JIMENEZ	PT102	Female	54	0	0	Ex-Smoker	27.32	6.6	80	0
ALBERT PARDINI	PT103	Male	28	0	0	never	27.32	5.7	158	0
CHRISTOPHER CHONG	PT104	Female	36	0	0	current	23.45	5	155	0
PATRICK GARDNER	PT105	Male	76	1	1	Ex-Smoker	20.14	4.8	155	0
DAVID SULLIVAN	PT106	Female	20	0	0	never	27.32	6.6	85	0
ALSON LEE	PT107	Female	44	0	0	never	19.31	6.5	200	1
DAVID KUSHNER	PT108	Female	79	0	0	Ex-Smoker	23.86	5.7	85	0
MICHAEL MORRIS	PT109	Male	42	0	0	never	33.64	4.8	145	0
JOANNE HAYES-WHITE	PT110	Female	32	0	0	never	27.32	5	100	0
ARTHUR KENNEY	PT111	Female	53	0	0	Ex-Smoker	27.32	6.1	85	0
PATRICIA JACKSON	PT112	Female	54	0	0	Ex-Smoker	54.7	6	100	0

diabetes_prediction 35 ×

Output

Action Output

#	Time	Action	Message
45	23:44:25	SET SQL_SAFE_UPDATES = 0	0 row(s) affected
46	23:44:28	update diabetes_prediction set smoking_history = 'Ex-Smoker' where age > 50	38463 row(s) affected Rows matched: 38463 Changed: 38463 Warnings: 0
47	23:44:42	select * from diabetes_prediction LIMIT 0, 1000	1000 row(s) returned

Insert a new patient into the database with sample data

```
1 • use cv;
2 • SELECT * FROM diabetes_prediction;
3
4 • Insert into diabetes_prediction values ('Michael','PT101679','Male',52,0,1,
5   'never',27.89,6.7,89,1);
```

Output

Action Output

#	Time	Action	Message
49	11:01:15	use cv	0 row(s) affected
50	11:01:17	SELECT * FROM diabetes_prediction LIMIT 0, 1000	1000 row(s) returned
51	11:01:22	Insert into diabetes_prediction values ('Michael','PT101679','Male',52,0,1, 'never',27.89,6.7,89,1)	1 row(s) affected

Delete all patients with heart disease from the database

```
1 • use cv;
2 • select * from diabetes_prediction;
3 • delete from diabetes_prediction where heart_disease = 1;
4
5
6
7
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

EmployeeName	Patient_id	gender	age	hypertension	heart_disease	smoking_history	bmi	HbA1c_level	blood_glucose_level	diabetes
GARY JIMENEZ	PT102	Female	54	0	0	Ex-Smoker	27.32	6.6	80	0
ALBERT PARDINI	PT103	Male	28	0	0	never	27.32	5.7	158	0
CHRISTOPHER CHONG	PT104	Female	36	0	0	current	23.45	5	155	0
DAVID SULLIVAN	PT106	Female	20	0	0	never	27.32	6.6	85	0
ALSON LEE	PT107	Female	44	0	0	never	19.31	6.5	200	1
DAVID KUSHNER	PT108	Female	79	0	0	Ex-Smoker	23.86	5.7	85	0
MICHAEL MORRIS	PT109	Male	42	0	0	never	33.64	4.8	145	0
JOANNE HAYES-WHITE	PT110	Female	32	0	0	never	27.32	5	100	0
ARTHUR KENNEY	PT111	Female	53	0	0	Ex-Smoker	27.32	6.1	85	0
PATRICIA JACKSON	PT112	Female	54	0	0	Ex-Smoker	54.7	6	100	0
EDWARD HARRINGTON	PT113	Female	78	0	0	Ex-Smoker	36.05	5	130	0
JOHN MARTIN	PT114	Female	67	0	0	Ex-Smoker	25.69	5.8	200	0

Result Grid | Form Editor | Field Types | Read Only

diabetes_prediction 37

Output

Action Output

#	Time	Action	Message
48	23:55:53	select * from diabetes_prediction LIMIT 0, 1000	1000 row(s) returned
49	23:56:41	delete from diabetes_prediction where heart_disease = 1	3942 row(s) affected
50	23:56:45	select * from diabetes_prediction LIMIT 0, 1000	1000 row(s) returned

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

```
1 • use cv;
2 • SELECT * FROM diabetes_prediction WHERE hypertension = 1
3 EXCEPT
4   SELECT * FROM diabetes_prediction WHERE diabetes = 1;
5
6
7
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

EmployeeName	Patient_id	gender	age	hypertension	heart_disease	smoking_history	bmi	HbA1c_level	blood_glucose_level	diabetes
DENISE SCHMITT	PT129	Male	45	1	0	never	26.47	4	158	0
RAY CRAWFORD	PT155	Female	45	1	0	never	23.05	4.8	130	0
KENNETH SMITH	PT161	Male	44	1	0	current	27.86	6.6	145	0
CHARLES SCOTT	PT215	Female	55	1	0	Ex-Smoker	34.2	5.7	140	0
SHANNON SAKOWSKI	PT227	Male	79	1	0	Ex-Smoker	28.73	6.6	160	0
MARISA MORET	PT241	Female	80	1	0	Ex-Smoker	44.06	6.5	160	0
STEPHEN TACCHINI	PT326	Female	48	1	0	never	36.73	6.6	126	0
ANDREW LOGAN	PT339	Male	59	1	0	Ex-Smoker	25.31	6	130	0
HAGOP HAJIAN	PT357	Female	52	1	0	Ex-Smoker	21.46	4	80	0
PERRY LEONG	PT377	Female	48	1	0	No Info	24.29	3.5	90	0
MELISSA LERMA	PT379	Female	59	1	0	Ex-Smoker	27.4	5.7	140	0
JOHN KOSTA	PT446	Female	52	1	0	Ex-Smoker	22.48	5	158	0

Result 39 | Read Only

Result Grid | Form Editor | Field Types

Output

Action Output

#	Time	Action	Message
50	23:56:45	select * from diabetes_prediction LIMIT 0, 1000	1000 row(s) returned
51	00:01:34	select * from diabetes_prediction where hypertension = 1 Except select * from diabetes_prediction where diabetes = 1	4839 row(s) returned
52	00:02:29	SELECT * FROM diabetes_prediction WHERE hypertension = 1 EXCEPT SELECT * FROM diabetes_prediction WH...	4839 row(s) returned

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

```sql
use cv;
SELECT * FROM diabetes_prediction;

ALTER TABLE diabetes_prediction
MODIFY patient_id VARCHAR(100);

ALTER TABLE diabetes_prediction
ADD CONSTRAINT unique_patient_id UNIQUE(patient_id);
```

Output

Action Output

#	Time	Action	Message
34	10:49:24	use cv	0 row(s) affected
35	10:49:26	SELECT * FROM diabetes_prediction LIMIT 0, 1000	1000 row(s) returned
36	10:49:29	ALTER TABLE diabetes_prediction MODIFY patient_id VARCHAR(100)	0 row(s) affected Records: 0 Duplicates: 0 Warnings: 0

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

```
1 • use cv;
2 • SELECT * FROM diabetes_prediction;
3
4 • create view Bmi_details as
5   select Patient_id, age , bmi from diabetes_prediction;
6 • SELECT * FROM Bmi_details;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

Patient_id	age	bmi
PT102	54	27.32
PT103	28	27.32
PT104	36	23.45
PT106	20	27.32
PT107	44	19.31
PT108	79	23.86
PT109	42	33.64
PT110	32	27.32
PT111	53	27.32
PT112	54	54.7
PT113	78	36.05
PT114	67	25.69

Bmi_details 40 ×

Result Grid | Form Editor | Field Types | Read Only