Homework 4

Professor: Jeremy Harper **TAs**: Bhavana Dhonvan, Latha Sree Rayala

Student Name: Deepak Rajput **Program**: MS Applied Data Science

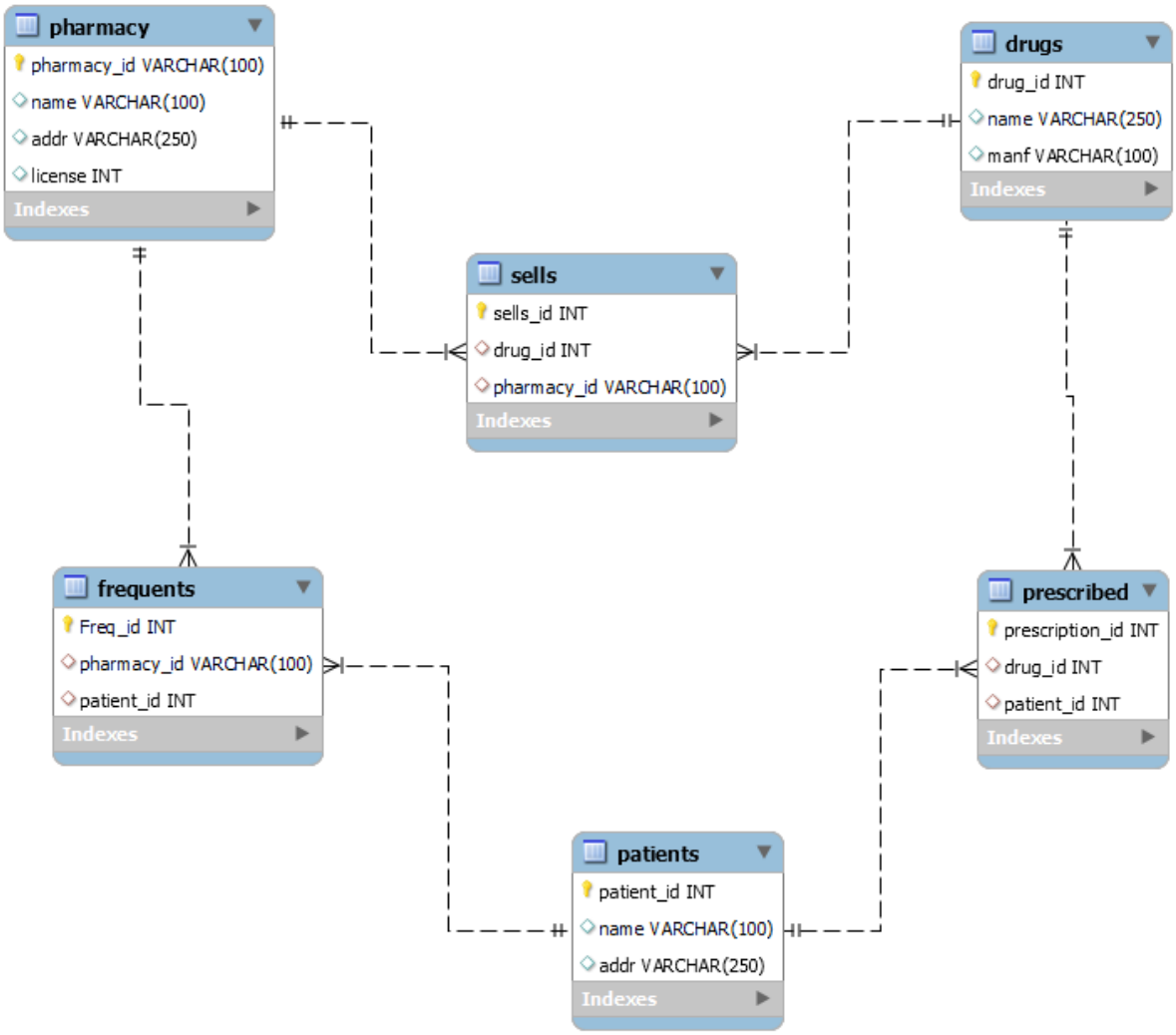
Question 1) Write the create table statements for the following ERD. Make sure to have the primary key and foreign key constraints

Answer: Create table code:

```
create schema Homework4;
use Homework4;
# first data table add id for primary key in every table
create table Drugs(
drug_id int not null primary key,
name varchar(250),
manf varchar(100));
# Second data table
create table Pharmacy(
pharmacy_id varchar(100) not null primary key,
name varchar(100),
addr varchar(250),
license int);
# Third data table
create table patients(
patient_id int not null primary key,
name varchar(100),
addr varchar(250));
# Fourth data table
create table sells(
sells_id int not null primary key,
drug_id int,
pharmacy_id varchar(100),
foreign key (drug_id) references Drugs(drug_id),
foreign key (pharmacy_id) references pharmacy(pharmacy_id));
# Fifth data table
create table Frequents(
Freq_id int not null primary key,
pharmacy_id varchar(100),
patient_id int,
foreign key (pharmacy_id) references Pharmacy(pharmacy_id),
foreign key (patient_id) references Patients(patient_id));
# Sixth data table

create table Prescribed(
prescription_id int not null primary key,
drug_id int,
patient_id int,
foreign key (drug_id) references Drugs(drug_id),
foreign key (patient_id) references Patients(patient_id));
```

Generated ERD diagram:



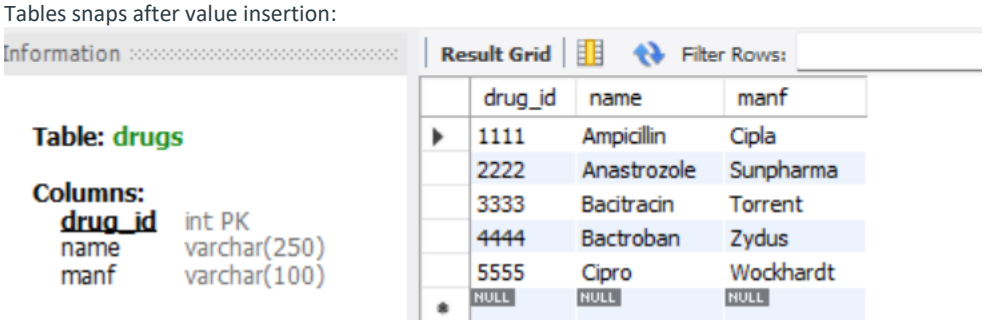
Question 2) Insert 5 records into each table of question.

Answer: MYSQL code for insertion:

```
insert into Drugs (drug_id , name, manf)
values
(1111, 'Ampicillin', 'Cipla'),
(2222, 'Anastrozole', 'Sunpharma'),
(3333, 'Bacitracin', 'Torrent'),
(4444, 'Bactroban', 'Zydus'),
(5555, 'Cipro', 'Wockhardt');
select * from drugs;
insert into Pharmacy (pharmacy_id , name, addr, license)
values
('PH11', 'CVC', '24 ST, INDIANAPOLIS 46222', 2302),
('PH22', 'APOLO', '56 ST, INDIANAPOLIS 46256', 9989),
('PH33', 'MEDMART', '16 ST, INDIANAPOLIS 46224', 1220),
('PH44', 'NICO', 'CANAL ST, INDIANAPOLIS 46205', 2304),
('PH55', 'MEXICA_MEDICA', '34 ST, INDIANAPOLIS 46234', 2456);
select * from Patients;
```

```
insert into Patients (patient_id , name, addr)
values
(3301, 'Mark', '15 ST, INDIANAPOLIS 46202'),
(3302, 'Miya', '12 ST, INDIANAPOLIS 46256'),
(3303, 'Monika', '16 ST, INDIANAPOLIS 46224'),
(3304, 'Deepak', 'CANAL ST, INDIANAPOLIS 46205'),
(3305, 'Seera', '21 ST, INDIANAPOLIS 46221');
select * from Patients;
insert into sells (sells_id, drug_id, pharmacy_id)
values
(2201,(select drug_id from drugs where drug_id= '1111'),
(select pharmacy_id from pharmacy where pharmacy_id= 'PH11')),
(2202,(select drug_id from drugs where drug_id= '2222'),
(select pharmacy_id from pharmacy where pharmacy_id= 'PH22')),
(2203,(select drug_id from drugs where drug_id= '3333'),
(select pharmacy_id from pharmacy where pharmacy_id= 'PH33')),
(2204,(select drug_id from drugs where drug_id= '4444'),
(select pharmacy_id from pharmacy where pharmacy_id= 'PH44')),
(2205,(select drug_id from drugs where drug_id= '5555'),
(select pharmacy_id from pharmacy where pharmacy_id= 'PH55'));
select * from sells;
insert into Frequents (Freq_id, pharmacy_id, patient_id)
values
(1101,(select pharmacy_id from pharmacy where pharmacy_id= 'PH11'),
(select patient_id from patients where patient_id= '3301')),
(1102,(select pharmacy_id from pharmacy where pharmacy_id= 'PH22'),
(select patient_id from patients where patient_id= '3302')),
(1103,(select pharmacy_id from pharmacy where pharmacy_id= 'PH33'),
(select patient_id from patients where patient_id= '3303')),
(1104,(select pharmacy_id from pharmacy where pharmacy_id= 'PH44'),
(select patient_id from patients where patient_id= '3304')),
(1105,(select pharmacy_id from pharmacy where pharmacy_id= 'PH55'),
(select patient_id from patients where patient_id= '3305'));
select * from Frequents;
insert into Prescribed (prescription_id, drug_id, patient_id)
values
(4401,(select drug_id from drugs where drug_id= '1111'),
(select patient_id from patients where patient_id= '3301')),
(4402,(select drug_id from drugs where drug_id= '2222'),
(select patient_id from patients where patient_id= '3302')),
(4403,(select drug_id from drugs where drug_id= '3333'),
(select patient_id from patients where patient_id= '3303')),
(4404,(select drug_id from drugs where drug_id= '4444'),
(select patient_id from patients where patient_id= '3304')),
(4405,(select drug_id from drugs where drug_id= '5555'),
(select patient_id from patients where patient_id= '3305'));
select * from Prescribed;
```

Tables snaps after value insertion:



Information

Table: pharmacy

Columns:

pharmacy_id	varchar(100) PK
name	varchar(100)
addr	varchar(250)
license	int

Result Grid | Filter Rows: | Edit: | Export/Imp

pharmacy_id	name	addr	license
PH11	CVC	24 ST, INDIANAPOLIS 46222	2302
PH22	APOLO	56 ST, INDIANAPOLIS 46256	9989
PH33	MEDMART	16 ST, INDIANAPOLIS 46224	1220
PH44	NICO	CANAL ST, INDIANAPOLIS 46205	2304
PH55	MEXICA_MEDICA	34 ST, INDIANAPOLIS 46234	2456
NULL	NULL	NULL	NULL

Information

Table: patients

Columns:

patient_id	int PK
name	varchar(100)
addr	varchar(250)

Result Grid | Filter Rows: | Edit:

patient_id	name	addr
3301	Mark	15 ST, INDIANAPOLIS 46202
3302	Miya	12 ST, INDIANAPOLIS 46256
3303	Monika	16 ST, INDIANAPOLIS 46224
3304	Deepak	CANAL ST, INDIANAPOLIS 46205
3305	Seera	21 ST, INDIANAPOLIS 46221
NULL	NULL	NULL

Administration Schemas

Information

Table: sells

Columns:

sells_id	int PK
drug_id	int
pharmacy_id	varchar(100)

Result Grid | Filter Rows:

sells_id	drug_id	pharmacy_id
2201	1111	PH11
2202	2222	PH22
2203	3333	PH33
2204	4444	PH44
2205	5555	PH55
NULL	NULL	NULL

Information

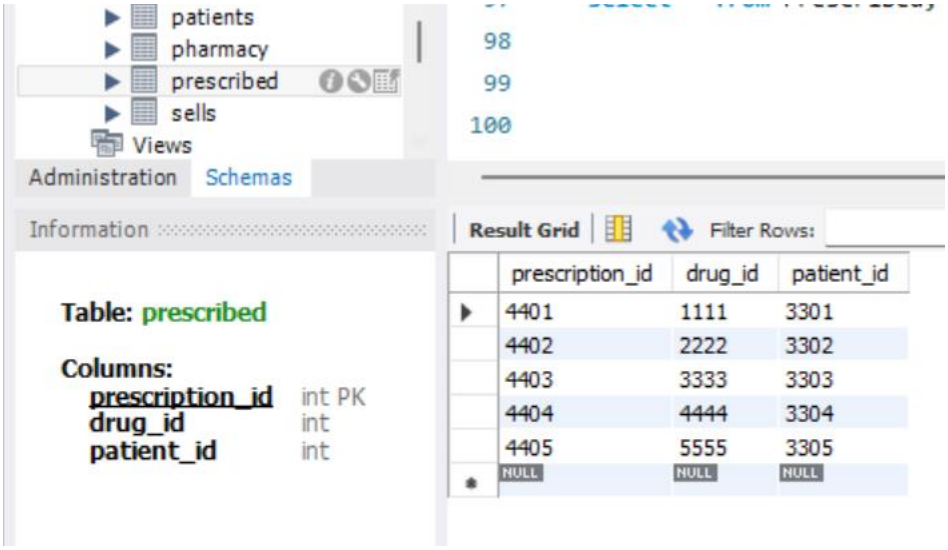
Table: frequents

Columns:

Freq_id	int PK
pharmacy_id	varchar(100)
patient_id	int

Result Grid | Filter Rows:

Freq_id	pharmacy_id	patient_id
1101	PH11	3301
1102	PH22	3302
1103	PH33	3303
1104	PH44	3304
1105	PH55	3305
NULL	NULL	NULL



Question 3) You may want to import some or all the following OMOP tables, please note you do not need to link the tables with the appropriate keys to be able to query.

We imported all data tables small one by import wizard and large data tables by below commands after creation of these data tables.
Sample commands:

```
load data infile 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/OMOP Data/drug_exposure.csv'
into table drug_exposure
fields terminated by ','
ignore 1 rows;
load data infile 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/OMOP Data/location.csv'
into table location
fields terminated by ','
ignore 1 rows;
```

Import snaps:

```
select * from cdm_source;
select * from condition_era;
select * from condition_occurrence;
select * from death;
select * from drug_era;
select * from drug_exposure;
select * from location;
select * from measurement;
select * from observation;
select * from observation_period;
select * from person;
select * from procedure_occurrence;
select * from visit_occurrence;
```

```
58 • select * from procedure_occurrence;
```

Output

Action Output

#	Time	Action	Message
123	22:49:29	Apply changes to measurement	Changes applied
124	22:49:36	load data infile 'C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/OMOP Data...	1048575 row(s) affected Records: 1048575 Deleted: 0 Skipped: 0 Warnings: 0
125	22:56:09	select * from cdm_source	1 row(s) returned
126	22:56:09	select * from condition_era	102155 row(s) returned
127	22:56:10	select * from condition_occurrence	103950 row(s) returned
128	22:56:10	select * from death	2359 row(s) returned
129	22:56:10	select * from drug_era	90903 row(s) returned
130	22:56:11	select * from drug_exposure	634992 row(s) returned
131	22:56:16	select * from location	12359 row(s) returned
132	22:56:16	select * from measurement	1048575 row(s) returned
133	22:56:24	select * from observation	109665 row(s) returned
134	22:56:25	select * from observation_period	12359 row(s) returned
135	22:56:25	select * from person	12359 row(s) returned
136	22:56:25	select * from procedure_occurrence	382314 row(s) returned
137	22:56:27	select * from visit_occurrence	516189 row(s) returned

Question 3.a) Write a SQL statement to find all patient records where the patient's birthdate is more than 50 years ago and the patient has a diagnosis of diabetes. Diabetes concept codes can be found [https://atlas-demo.ohdsi.org/#/conceptset/1872607/expressionLinks to an external site.](https://atlas-demo.ohdsi.org/#/conceptset/1872607/expression)

Answer:

Get all concept id from atlas for diabetes

CODE: select * from person p
join condition_occurrence co on p.person_id = co.person_id
where condition_concept_id in
(201254,201530,201820,201826,201826,376065,376114,380097,442793,443729,443731,443732,443733,443734,602345,604741,
608884,609095,609096,609099,609101,609103,609104,609105,609106,609109,609112,609114,609116,609117,609119,760989,
761053,761063,765375,766253,1326491,1326492,1409150,1409151,1409152,1567956,1567957,1567967,1567970,1567971,3105075,
3121806,3142220,3193274,3194082,3329005,3522470,3522662,3522803,4063043,4099216,4099651,4129519,4130162,4131908,
4140466,4142579,4177050,4193704,4196141,4200875,4215719,4221487,4221495,4222415,4222876,4223463,4223739,4226121,
4228443,4230254,4266637,4290822,4304377,4321756,4334340,35206881,35626070,36676219,36684827,36712670,36712686,36712687,
36714116,36717156,37016349,37016354,37016768,37017432,37018728,37018912,37200250,37200251,37200252,37200253,37200254,37309630,
37312202,37312203,37312204,37312205,40320749,40350832,40386778,40482801,40483315,40485020,42485992,42485993,43530656,43530685,
43530689,43530690,43531010,43531559,43531562,43531563,43531564,43531566,43531577,43531578,43531588,43531597,43531608,
43531616,43531651,43531653,44824072,44826461,44827617,44829878,44829882,44831047,45420114,45420119,45453110,45486691,45493234,
45493235,45496536,45513201,45525872,45533022,45533023,45542738,45566731,45571656,45576441,45581355,45591031,45595797,45595799,
45600641,45605405,45611690,45757075,45757079,45757277,45757278,45757280,45757363,45757392,45757435,45757444,45757445,45757446,
45757447,45757449,45757450,45757474,45757499,45757508,45763582,45766052,45769828,45769835,45769836,45769872,45769875,45769888,
45769890,45769894,45769905,45769906,45770830,45770831,45770832,45770880,45770881,45770883,45770928,45771064,45771072,45772019,
45772060,45773064,46274058) and p.year_of_birth < DATE_SUB(CURDATE(), INTERVAL 50 YEAR);

OUTPUT

```
45493235,45496536,45513201,45525872,45533022,45533023,45542738,45566731,45571656,45576441,45581355,455916
45600641,45605405,45611690,45757075,45757079,45757277,45757278,45757280,45757363,45757392,45757435,457574
45757447,45757449,45757450,45757474,45757499,45757508,45763582,45766052,45769828,45769835,45769836,457698
45769890,45769894,45769905,45769906,45770830,45770831,45770832,45770880,45770881,45770883,45770928,457716
45772060,45773064,46274058) and p.year_of_birth < DATE_SUB(CURDATE(), INTERVAL 50 YEAR);
```

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

person_id	gender_concept_id	year_of_birth	month_of_birth	day_of_birth	birth_datetime	death_datetime	race_concept_id
44	8507	1914	4	19	19-04-1914 00:00	05-01-1969 00:00	8527
44	8507	1914	4	19	19-04-1914 00:00	05-01-1969 00:00	8527
47	8507	1962	1	18	18-01-1962 00:00	07-02-2016 00:00	8527
59	8507	1960	5	1	01-05-1960 00:00	0	8527
59	8507	1960	5	1	01-05-1960 00:00	0	8527
77	8532	1972	3	9	09-03-1972 00:00	0	8515
85	8507	1945	10	12	12-10-1945 00:00	14-02-2020 00:00	8527
85	8507	1945	10	12	12-10-1945 00:00	14-02-2020 00:00	8527

Result Grid | Form Editor | Field Types

Result 33 | Read Only

Output

Action Output

#	Time	Action	Message
149	23:33:42	select * from person p join condition_occurrence co on p.person_id = co.person_id...	1498 row(s) returned

Question 3 b) Write a SQL statement to calculate the average blood pressure for each patient in the most recent 30 days. (Note the data won't allow you to see information from the last 30 days, you'd have to create your own data if you want to test... or you could test between dates and verify)

Answer: code

select measurement.person_id, avg(measurement.value_as_number) as avg_blood_pressure
from measurement
where measurement.measurement_concept_id IN (3004249,3012888)
and measurement.measurement_date BETWEEN '2016-05-24' AND '2016-06-23'
GROUP BY measurement.person_id;

```
45772060,45773064,46274058) and p.year_of_birth < DATE_SUB(CURDATE(), INTERVAL 50 YEAR);
select measurement.person_id, avg(measurement.value_as_number) as avg_blood_pressure
from measurement
where measurement.measurement_concept_id IN (3004249,3012888)
and measurement.measurement_date BETWEEN '2016-05-24' AND '2016-06-23'
GROUP BY measurement.person_id;
```

person_id	avg_blood_press..
8	104
21	106.5
25	99
38	103
58	96.5
74	103.5

Question 3.c1)
Write a SQL statement to find the number of male and female patients in the database.
Answer:
Code: select gender_source_value, count(*) patient_count from person where gender_source_value in ('M','F') group by gender_source_value;

```
82 •  select gender_source_value, count(*) patient_count from person where
83    gender_source_value in ('M','F') group by gender_source_value;
84
```

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

gender_source_value	patient_count
F	6087
M	6272

3.c2) Breakdown the Male/Female by 10 year age group and count how many people fit that category.
Answer: code

```
SELECT
  CASE
    WHEN FLOOR((YEAR(CURRENT_DATE) - year_of_birth) / 10) * 10 BETWEEN 0 AND 9 THEN '0-9'
    WHEN FLOOR((YEAR(CURRENT_DATE) - year_of_birth) / 10) * 10 BETWEEN 10 AND 19 THEN '10-19'
    WHEN FLOOR((YEAR(CURRENT_DATE) - year_of_birth) / 10) * 10 BETWEEN 20 AND 29 THEN '20-29'
    WHEN FLOOR((YEAR(CURRENT_DATE) - year_of_birth) / 10) * 10 BETWEEN 30 AND 39 THEN '30-39'
    WHEN FLOOR((YEAR(CURRENT_DATE) - year_of_birth) / 10) * 10 BETWEEN 40 AND 49 THEN '40-49'
    WHEN FLOOR((YEAR(CURRENT_DATE) - year_of_birth) / 10) * 10 BETWEEN 50 AND 59 THEN '50-59'
    WHEN FLOOR((YEAR(CURRENT_DATE) - year_of_birth) / 10) * 10 BETWEEN 60 AND 69 THEN '60-69'
    WHEN FLOOR((YEAR(CURRENT_DATE) - year_of_birth) / 10) * 10 BETWEEN 70 AND 79 THEN '70-79'
    WHEN FLOOR((YEAR(CURRENT_DATE) - year_of_birth) / 10) * 10 BETWEEN 80 AND 89 THEN '80-89'
    WHEN FLOOR((YEAR(CURRENT_DATE) - year_of_birth) / 10) * 10 BETWEEN 90 AND 99 THEN '90-99'
    ELSE '100+'
  END AS age_group,
  SUM(CASE WHEN gender_concept_id = 8507 THEN 1 ELSE 0 END) AS male_count,
  SUM(CASE WHEN gender_concept_id = 8532 THEN 1 ELSE 0 END) AS female_count
FROM person
GROUP BY age_group;
```

```
97    WHEN FLOOR((YEAR(CURRENT_DATE) - year_of_birth) / 10) * 10 BETWEEN 60 AND 69 THEN '60-69'
98    WHEN FLOOR((YEAR(CURRENT_DATE) - year_of_birth) / 10) * 10 BETWEEN 70 AND 79 THEN '70-79'
99    WHEN FLOOR((YEAR(CURRENT_DATE) - year_of_birth) / 10) * 10 BETWEEN 80 AND 89 THEN '80-89'
100   WHEN FLOOR((YEAR(CURRENT_DATE) - year_of_birth) / 10) * 10 BETWEEN 90 AND 99 THEN '90-99'
101   ELSE '100+'
102 END AS age_group,
```

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

age_group	male_count	female_count
20-29	776	794
40-49	647	678
0-9	351	358
100+	323	298
80-89	541	332
30-39	668	745
50-59	720	772
90-99	225	161
60-69	789	747

Result 51

3.d Write a SQL statement to find the average number of appointments per patient in the last year. Output the following fields. person_id, gender_concept_id, year_of_birth, race_concept_id, AppointmentCount

Code:

```
SELECT p.person_id, p.gender_concept_id, p.year_of_birth, p.race_concept_id, COUNT(DISTINCT vo.visit_occurrence_id)
AS Appointment_Count
FROM person p
JOIN visit_occurrence vo ON p.person_id = vo.person_id
WHERE vo.visit_start_date BETWEEN '2016-06-23' AND '2016-07-22'
GROUP BY p.person_id, p.gender_concept_id, p.year_of_birth, p.race_concept_id;
```

```
118 • SELECT p.person_id, p.gender_concept_id, p.year_of_birth, p.race_concept_id, COUNT(DISTINCT vo.visit_occu
119   FROM person p
120   JOIN visit_occurrence vo ON p.person_id = vo.person_id
121   WHERE vo.visit_start_date BETWEEN '2016-05-24' AND '2016-06-23'
122   GROUP BY p.person_id, p.gender_concept_id, p.year_of_birth, p.race_concept_id;
```

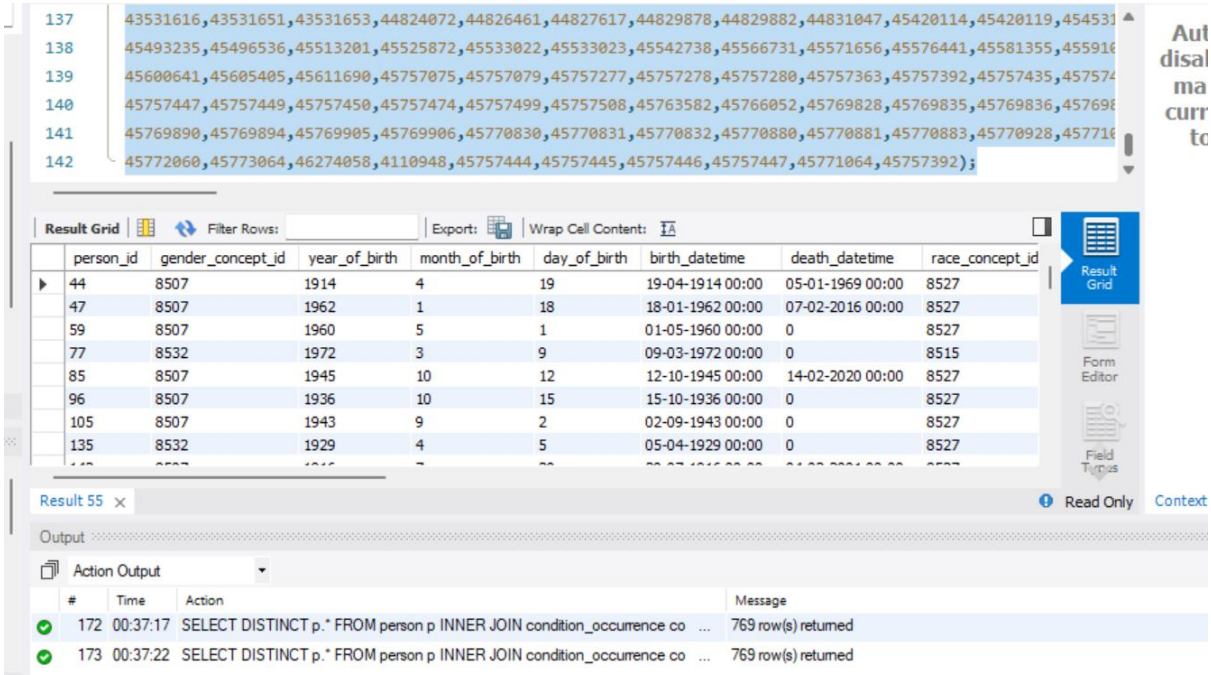
Result Grid | Filter Rows: | Export: | Wrap Cell Content:

person_id	gender_concept_id	year_of_birth	race_concept_id	Appointment_Count

3.e Write a SQL statement to find all patients who have been diagnosed with both hypertension and diabetes. (You will need to find concept list yourself from atlas)

Answer:

Get all concept id from atlas for both hypertension and diabetes

```
SELECT DISTINCT p.*
FROM person p
INNER JOIN condition_occurrence co
  ON p.person_id = co.person_id
WHERE co.condition_concept_id IN
(201254,201530,201820,201826,201826,376065,376114,380097,442793,443729,443731,443732,443733,443734,602345,
604741,
```

608884,609095,609096,609099,609101,609103,609104,609105,609106,609109,609112,609114,609116,609117,609119,760989,
761053,761063,765375,766253,1326491,1326492,1409150,1409151,1409152,1567956,1567957,1567967,1567970,1567971,3105075,
3121806,3142220,3193274,3194082,3329005,3522470,3522662,3522803,4063043,4099216,4099651,4129519,4130162,4131908,
4140466,4142579,4177050,4193704,4196141,4200875,4215719,4221487,4221495,4222415,4222876,4223463,4223739,4226121,
4228443,4230254,4266637,4290822,4304377,4321756,4334340,35206881,35626070,36676219,36684827,36712670,36712686,36712687,
36714116,36717156,37016349,37016354,37016768,37017432,37018728,37018912,37200250,37200251,37200252,37200253,37200254,37309630,
37312202,37312203,37312204,37312205,40320749,40350832,40386778,40482801,40483315,40485020,42485992,42485993,43530656,43530685,
43530689,43530690,43531010,43531559,43531562,43531563,43531564,43531566,43531577,43531578,43531588,43531597,43531608,
43531616,43531651,43531653,44824072,44826461,44827617,44829878,44829882,44831047,45420114,45420119,45453110,45486691,45493234,
45493235,45496536,45513201,45525872,45533022,45533023,45542738,45566731,45571656,45576441,45581355,45591031,45595797,45595799,
45600641,45605405,45611690,45757075,45757079,45757277,45757278,45757280,45757363,45757392,45757435,45757444,45757445,45757446,
45757447,45757449,45757450,45757474,45757499,45757508,45763582,45766052,45769828,45769835,45769836,45769872,45769875,45769888,
45769890,45769894,45769905,45769906,45770830,45770831,45770832,45770880,45770881,45770883,45770928,45771064,45771072,45772019,
45772060,45773064,46274058,4110948,45757444,45757445,45757446,45757447,45771064,45757392);