

CIND 119: Introduction to Big Data Analytics
Assignment 2 (15% of the final grade)
Querying an RDBMS database using SQLiteStudio

Name: Ashwin David

Section: DK0

Student Number: 500830814

Due Date: March 28, 2021

1. Create a database called “sample”.
2. Create a table called “test_data” and load the following data into the table. (5 points)

Code:

```
/*Create a Table called "test_data" */
create table test_data(
ID integer Not null,
Class varchar(25),
age integer,
menopause varchar (25),
tumor_size integer,
node_caps varchar(3),
deg_malign integer,
breast varchar (20),
breast_quad varchar(25),
irradiat varchar(15),
Primary key (ID));

/*Load the following data into the table*/
insert into test_data(ID, Class, age, menopause, tumor_size, node_caps, deg_malign,
breast, breast_quad, irradiat)
Values (1, 'NO', 35, 'premeno', 31, 'no', 3, 'left', 'left_low', 'no');
insert into test_data(ID, Class, age, menopause, tumor_size, node_caps, deg_malign,
breast, breast_quad, irradiat)
Values (2, 'NO', 42, 'premeno', 22, 'no', 2, 'right', 'right_up', 'no');
insert into test_data(ID, Class, age, menopause, tumor_size, node_caps, deg_malign,
breast, breast_quad, irradiat)
Values (3, 'NO', 30, 'premeno', 23, 'no', 2, 'left', 'left_low', 'no');
insert into test_data(ID, Class, age, menopause, tumor_size, node_caps, deg_malign,
breast, breast_quad, irradiat)
Values (4, 'NO', 61, 'ge40', 16, 'no', 2, 'right', 'left_up', 'no');
insert into test_data(ID, Class, age, menopause, tumor_size, node_caps, deg_malign,
```

```

breast, breast_quad, irradiat)
Values (5, 'NO', 45, 'premeno', 2, 'no', 2, 'right', 'right_low', 'no');
insert into test_data(ID, Class, age, menopause, tumor_size, node_caps, deg_malign,
breast, breast_quad, irradiat)
Values (6, 'NO', 64, 'ge40', 17, 'no', 2, 'left', 'left_low', 'no');
insert into test_data(ID, Class, age, menopause, tumor_size, node_caps, deg_malign,
breast, breast_quad, irradiat)
Values (7, 'NO', 52, 'premeno', 27, 'no', 2, 'left', 'left_low', 'no');
insert into test_data(ID, Class, age, menopause, tumor_size, node_caps, deg_malign,
breast, breast_quad, irradiat)
Values (8, 'NO', 67, 'ge40', 21, 'no', 1, 'left', 'left_low', 'no');
insert into test_data(ID, Class, age, menopause, tumor_size, node_caps, deg_malign,
breast, breast_quad, irradiat)
Values (9, 'YES', 41, 'premeno', 52, 'no', 2, 'left', 'left_low', 'no');
insert into test_data(ID, Class, age, menopause, tumor_size, node_caps, deg_malign,
breast, breast_quad, irradiat)
Values (10, 'YES', 43, 'premeno', 22, 'no', 2, 'right', 'left_up', 'no');
insert into test_data(ID, Class, age, menopause, tumor_size, node_caps, deg_malign,
breast, breast_quad, irradiat)
Values (11, 'YES', 41, 'premeno', 1, 'no', 3, 'left', 'central', 'no');
insert into test_data(ID, Class, age, menopause, tumor_size, node_caps, deg_malign,
breast, breast_quad, irradiat)
Values (12, 'YES', 44, 'ge40', 27, 'no', 2, 'left', 'left_low', 'no');
insert into test_data(ID, Class, age, menopause, tumor_size, node_caps, deg_malign,
breast, breast_quad, irradiat)
Values (13, 'YES', 61, 'It40', 14, 'no', 1, 'left', 'right_up', 'no');
insert into test_data(ID, Class, age, menopause, tumor_size, node_caps, deg_malign,
breast, breast_quad, irradiat)
Values (14, 'YES', 55, 'ge40', 26, 'no', 3, 'left', 'right_up', 'no');
insert into test_data(ID, Class, age, menopause, tumor_size, node_caps, deg_malign,
breast, breast_quad, irradiat)
Values (15, 'YES', 44, 'premeno', 32, 'no', 3, 'left', 'left_up', 'no');

```

Answer:

ID	Class	Age	Menopause	Tumor_size	Node_caps	deg_malign	breast	breast_quad	irradiat
1	NO	35	premeno	31	no	3	left	left_low	no

2	NO	42	premeno	22	no	2	right	right_up	no
3	NO	30	premeno	23	no	2	left	left_low	no
4	NO	61	ge40	16	no	2	right	left_up	no
5	NO	45	premeno	2	no	2	right	right_low	no
6	NO	64	ge40	17	no	2	left	left_low	no
7	NO	52	premeno	27	no	2	left	left_low	no
8	NO	67	ge40	21	no	1	left	left_low	no
9	YES	41	premeno	52	no	2	left	left_low	no
10	YES	43	premeno	22	no	2	right	left_up	no
11	YES	41	premeno	1	no	3	left	central	no
12	YES	44	ge40	27	no	2	left	left_low	no
13	YES	61	It40	14	no	1	left	right_up	no

14	YES	55	ge40	26	no	3	left	right_up	no
15	YES	44	premeno	32	no	3	left	left_up	no

3. Write SQL queries to select/compute data from the “test_data” table. (2 points each)
- a. Select all rows where menopause column has the value “ge40”.

Code:

```
/*Visualize the table*/
select * from test_data;

/* a. Select all rows where menopause column has the value "ge40" */
select * from test_data
where menopause = "ge40";
```

Answer:

ID	Class	Age	Menopause	Tumor_size	Node_caps	deg_m alig	breast	breast_q uad	irradiat
4	NO	61	ge40	16	no	2	right	left_up	no
6	NO	64	ge40	17	no	2	left	left_low	no
8	NO	67	ge40	21	no	1	left	left_low	no
12	YES	44	ge40	27	no	2	left	left_low	no
14	YES	55	ge40	26	no	3	left	right_up	no

- b. Select all rows where age is less than 41.

Code:

```
/* b. Select all rows where age is less than 41*/
select * from test_data
where age < 41;
```

Answer:

ID	Class	Age	Menopause	Tumor_size	Node_caps	Deg-malig	Breast	Breast_quad	irradiat
1	NO	35	premeno	31	no	3	left	left_low	no
3	NO	30	premeno	23	no	2	left	left_low	no

- c. Select all rows where age is less than 41 and menopause column has the value "ge40".

Code:

```
/* c. select all rows where age is less than 41 and menopause column has the
value "ge40"*/
select * from test_data
where age < 41 AND menopause = "ge40";
```

Answer:

N/A

- d. Compute the average age across all rows.

Code:

```
/* d. Compute the average across all rows */
select AVG(age) from test_data;
```

Answer:

AVG(age):

48.3333334

- e. Compute average age across rows where deg_malig value is equal to 3.

Code:

```
/* e. Compute the average age across rows where deg_malig value is equal to 3
```

```
*/  
select deg_malig, AVG(age) from test_data  
where deg_malig = 3 group by deg_malig;
```

Answer:

Deg_malig	AVG(age)
3	43.75