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

Complete this assignment using SQLiteStudio.

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

class	age	menopause	tumor_size	node_caps	deg_malig	breast	breast_quad	irradiat
NO	35	premeno	31	no	3	left	left_low	no
NO	42	premeno	22	no	2	right	right_up	no
NO	30	premeno	23	no	2	left	left_low	no
NO	61	ge40	16	no	2	right	left_up	no
NO	45	premeno	2	no	2	right	right_low	no
NO	64	ge40	17	no	2	left	left_low	no
NO	52	premeno	27	no	2	left	left_low	no
NO	67	ge40	21	no	1	left	left_low	no
YES	41	premeno	52	no	2	left	left_low	no
YES	43	premeno	22	no	2	right	left_up	no
YES	41	premeno	1	no	3	left	central	no
YES	44	ge40	27	no	2	left	left_low	no
YES	61	lt40	14	no	1	left	right_up	no
YES	55	ge40	26	no	3	left	right_up	no
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”.
    b. Select all rows where age is less than 41.
    c. Select all rows where age is less than 41 and menopause column has the value “ge40”.
    d. Compute the average age across all rows.
    e. Compute average age across rows where deg_malig value is equal to 3.

Reference: Full dataset available at GroupLens

End of CIND119 Assignment 2