CIND 119: Introduction to Big Data Analytics
Assignment 3 (15% of the final grade)
Querying a NoSQL database using MongoDB

Complete this assignment using MongoDB.

1. Create a database called "sample". (1 point)

2. Create a collection called "test_data" and load the following data into the collection. (6 points)

class	age	menopause	deg_malig	breast	breast_quad	irradiat
NO	35	premeno	3	left	left_low	no
NO	42	premeno	2	right	right_up	no
NO	30	premeno	2	left	left_low	no
NO	61	ge40	2	right	left_up	no
NO	45	premeno	2	right	right_low	no
NO	64	ge40	2	left	left_low	no
NO	52	premeno	2	left	left_low	no
NO	67	ge40	1	left	left_low	no
YES	41	premeno	2	left	left_low	no
YES	43	premeno	2	right	left_up	no
YES	41	premeno	3	left	central	no
YES	44	ge40	2	left	left_low	no

YES	61	lt40	1	left	right_up	no
YES	55	ge40	3	left	right_up	no
YES	44	premeno	3	left	left_up	no

3. Write MongoDB queries to select/compute data from the "test_data" collection. (2 points each)

    a. Select all rows where the 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 or the menopause column has the value "ge40".

    d. Compute the average age across all rows.

Reference: Full dataset available at GroupLens

End of CIND119 Assignment 3