CIND 119: Introduction to Big Data Analytics
Assignment 3 (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”. (1 point)**
   Answer:
   > show dbs
   Cind119 0.000GB
   Sample   0.000GB
   admin   0.000GB
   config   0.000GB
   local   0.000GB
   > use Sample
   switched to db Sample

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

   #Sample Code used to create database:

```
{
   "_id": {
      "$oid": "6074d04c319ac9494c22bb02"
   },
   "Class": "NO",
   "Age": 35,
   "Menopause": "premeno",
   "Deg_Malig": 3,
   "Breast": "left",
   "Breast_quad": "left_low",
   "Irradiated": "no"
}
```

```
#Display the collection of documents
> db.test_data.find()
```

Answer:

```
{ "_id" : ObjectId("6074d04c319ac9494c22bb02"), "Class" : "NO", "Age" : 35,
"Menopause" : "premeno", "Deg_Malig" : 3, "Breast" : "left", "Breast_quad" :
"left_low", "Irradiat" : "no" }
{ "_id" : ObjectId("6074d073319ac9494c22bb03"), "Class" : "NO", "Age" : 42,
"Menopause" : "premeno", "Deg_Malig" : 2, "Breast" : "right", "Breast_quad" :
"right_up", "Irradiat" : "no" }
{ "_id" : ObjectId("6074d088319ac9494c22bb04"), "Class" : "NO", "Age" : 30,
"Menopause" : "premeno", "Deg_Malig" : 2, "Breast" : "left", "Breast_quad" :
"left_low", "Irradiat" : "no" }
{ "_id" : ObjectId("6074d0b9319ac9494c22bb05"), "Class" : "NO", "Age" : 61,
"Menopause" : "ge40", "Deg_Malig" : 2, "Breast" : "right", "Breast_quad" : "left_up",
"Irradiat" : "no" }
{ "_id" : ObjectId("6074d0e2319ac9494c22bb06"), "Class" : "NO", "Age" : 45,
"Menopause" : "premeno", "Deg_Malig" : 2, "Breast" : "right", "Breast_quad" :
"right_low", "Irradiat" : "no" }
{ "_id" : ObjectId("6074d0fc319ac9494c22bb07"), "Class" : "NO", "Age" : 64,
"Menopause" : "ge40", "Deg_Malig" : 2, "Breast" : "left", "Breast_quad" : "left_low",
"Irradiat" : "no" }
{ "_id" : ObjectId("6074d112319ac9494c22bb08"), "Class" : "NO", "Age" : 52,
"Menopause" : "premeno", "Deg_Malig" : 2, "Breast" : "left", "Breast_quad" :
"left_low", "Irradiat" : "no" }
{ "_id" : ObjectId("6074d128319ac9494c22bb09"), "Class" : "NO", "Age" : 67,
"Menopause" : "ge40", "Deg_Malig" : 1, "Breast" : "left", "Breast_quad" : "left_low",
"Irradiat" : "no" }
{ "_id" : ObjectId("6074d140319ac9494c22bb0a"), "Class" : "YES", "Age" : 41,
"Menopause" : "premeno", "Deg_Malig" : 2, "Breast" : "left", "Breast_quad" :
"left_low", "Irradiat" : "no" }
{ "_id" : ObjectId("6074d162319ac9494c22bb0b"), "Class" : "YES", "Age" : 43,
"Menopause" : "premeno", "Deg_Malig" : 2, "Breast" : "right", "Breast_quad" :
"left_up", "Irradiat" : "no" }
{ "_id" : ObjectId("6074d176319ac9494c22bb0c"), "Class" : "YES", "Age" : 41,
"Menopause" : "premeno", "Deg_Malig" : 3, "Breast" : "left", "Breast_quad" :
"central", "Irradiat" : "no" }
```

{ "_id" : ObjectId("6074d189319ac9494c22bb0d"), "Class" : "YES", "Age" : 44, "Menopause" : "ge40", "Deg_Malig" : 2, "Breast" : "left", "Breast_quad" : "left_low", "Irradiat" : "no" }
{ "_id" : ObjectId("6074d1b5319ac9494c22bb0e"), "Class" : "YES", "Age" : 61, "Menopause" : "it40", "Deg_Malig" : 1, "Breast" : "left", "Breast_quad" : "right_up", "Irradiat" : "no" }
{ "_id" : ObjectId("6074d21b319ac9494c22bb0f"), "Class" : "YES", "Age" : 55, "Menopause" : "ge40", "Deg_Malig" : 3, "Breast" : "left", "Breast_quad" : "right_up", "Irradiat" : "no" }
{ "_id" : ObjectId("6074d22e319ac9494c22bb10"), "Class" : "YES", "Age" : 44, "Menopause" : "premeno", "Deg_Malig" : 3, "Breast" : "left", "Breast_quad" : "left_up", "Irradiat" : "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".**
   Code:

```
#Select all rows that have ge40
> db.test_data.find({Menopause: "ge40"})
```

   Answer:

{ "_id" : ObjectId("6074d0b9319ac9494c22bb05"), "Class" : "NO", "Age" : 61, "Menopause" : "ge40", "Deg_Malig" : 2, "Breast" : "right", "Breast_quad" : "left_up", "Irradiat" : "no" }
{ "_id" : ObjectId("6074d0fc319ac9494c22bb07"), "Class" : "NO", "Age" : 64, "Menopause" : "ge40", "Deg_Malig" : 2, "Breast" : "left", "Breast_quad" : "left_low", "Irradiat" : "no" }
{ "_id" : ObjectId("6074d128319ac9494c22bb09"), "Class" : "NO", "Age" : 67, "Menopause" : "ge40", "Deg_Malig" : 1, "Breast" : "left", "Breast_quad" : "left_low", "Irradiat" : "no" }
{ "_id" : ObjectId("6074d189319ac9494c22bb0d"), "Class" : "YES", "Age" : 44, "Menopause" : "ge40", "Deg_Malig" : 2, "Breast" : "left", "Breast_quad" : "left_low", "Irradiat" : "no" }

{ "_id" : ObjectId("6074d21b319ac9494c22bb0f"), "Class" : "YES", "Age" : 55, "Menopause" : "ge40", "Deg_Malig" : 3, "Breast" : "left", "Breast_quad" : "right_up", "Irradiat" : "no" }

b. Select all rows where age is less than 41.
Code:

```
#Select all rows where age is less than 40
> db.test_data.find({Age: {$lt: 40}})
```

Answer:

{ "_id" : ObjectId("6074d04c319ac9494c22bb02"), "Class" : "NO", "Age" : 35, "Menopause" : "premeno", "Deg_Malig" : 3, "Breast" : "left", "Breast_quad" : "left_low", "Irradiat" : "no" }
{ "_id" : ObjectId("6074d088319ac9494c22bb04"), "Class" : "NO", "Age" : 30, "Menopause" : "premeno", "Deg_Malig" : 2, "Breast" : "left", "Breast_quad" : "left_low", "Irradiat" : "no" }

c. Select all rows where age is less than 41 or the menopause column has the value "ge40".
Code:

```
#Select all rows where age is less than 41 or the menopause column has the value "ge40"
> db.test_data.find({$or:[{Age:{$lt:40}}, {Menopause: "ge40"}]})
```

Answer:

{ "_id" : ObjectId("6074d04c319ac9494c22bb02"), "Class" : "NO", "Age" : 35, "Menopause" : "premeno", "Deg_Malig" : 3, "Breast" : "left", "Breast_quad" : "left_low", "Irradiat" : "no" }
{ "_id" : ObjectId("6074d088319ac9494c22bb04"), "Class" : "NO", "Age" : 30, "Menopause" : "premeno", "Deg_Malig" : 2, "Breast" : "left", "Breast_quad" : "left_low", "Irradiat" : "no" }
{ "_id" : ObjectId("6074d0b9319ac9494c22bb05"), "Class" : "NO", "Age" : 61, "Menopause" : "ge40", "Deg_Malig" : 2, "Breast" : "right", "Breast_quad" : "left_up", "Irradiat" : "no" }
{ "_id" : ObjectId("6074d0fc319ac9494c22bb07"), "Class" : "NO", "Age" : 64, "Menopause" : "ge40", "Deg_Malig" : 2, "Breast" : "left", "Breast_quad" : "left_low", "Irradiat" : "no" }
{ "_id" : ObjectId("6074d128319ac9494c22bb09"), "Class" : "NO", "Age" : 67, "Menopause" : "ge40", "Deg_Malig" : 1, "Breast" : "left", "Breast_quad" : "left_low", "Irradiat" : "no" }
{ "_id" : ObjectId("6074d189319ac9494c22bb0d"), "Class" : "YES", "Age" : 44, "Menopause" : "ge40", "Deg_Malig" : 2, "Breast" : "left", "Breast_quad" : "left_low", "Irradiat" : "no" }
{ "_id" : ObjectId("6074d21b319ac9494c22bb0f"), "Class" : "YES", "Age" : 55, "Menopause" : "ge40", "Deg_Malig" : 3, "Breast" : "left", "Breast_quad" : "right_up", "Irradiat" : "no" }

d. Compute the average age across all rows.

Code:

```
#Compute the average age across all rows.
> db.test_data.aggregate([{$group:{_id:null, average_age: {$avg: "$Age"}}}])
```

Answer:

{ "_id" : null, "average_age" : 48.333333333333336 }