Module 12 NoSQL MongoDB Lab Assignment

June 10, 2024

```python
[1]: import pymongo, pandas as pd
```

```python
[2]: # Connect to MongoDB
     client = pymongo.MongoClient("mongodb://localhost:27017/")
     client
```

```
[2]: MongoClient(host=['localhost:27017'], document_class=dict, tz_aware=False,
     connect=True)
```

```python
[3]: db = client.housing
     db
```

```
[3]: Database(MongoClient(host=['localhost:27017'], document_class=dict,
     tz_aware=False, connect=True), 'housing')
```

```python
[4]: collection = db.house
     collection
```

```
[4]: Collection(Database(MongoClient(host=['localhost:27017'], document_class=dict,
     tz_aware=False, connect=True), 'housing'), 'house')
```

```python
[5]: # Retrieve the first three records
     records = collection.find({}).limit(3)

     # Print the records
     for rec in records:
         print("{")
         for key in rec.keys():
             print(f"{key}: {rec[key]},")
         print("},")
```

```
{
_id: 6666a3fdb1a4744f2a70a99b,
id: 7129300520,
date: 20141013T000000,
price: 221900,
bedrooms: 3,
bathrooms: 1,
sqft_living: 1180,
```

sqft_lot: 5650,
floors: 1,
waterfront: 0,
view: 0,
condition: 3,
grade: 7,
sqft_above: 1180,
sqft_basement: 0,
yr_built: 1955,
yr_renovated: 0,
zipcode: 98178,
lat: 47.5112,
long: -122.257,
sqft_living15: 1340,
sqft_lot15: 5650,
},
{
_id: 6666a3fdb1a4744f2a70a99c,
id: 6414100192,
date: 20141209T000000,
price: 538000,
bedrooms: 3,
bathrooms: 2.25,
sqft_living: 2570,
sqft_lot: 7242,
floors: 2,
waterfront: 0,
view: 0,
condition: 3,
grade: 7,
sqft_above: 2170,
sqft_basement: 400,
yr_built: 1951,
yr_renovated: 1991,
zipcode: 98125,
lat: 47.721,
long: -122.319,
sqft_living15: 1690,
sqft_lot15: 7639,
},
{
_id: 6666a3fdb1a4744f2a70a99d,
id: 5631500400,
date: 20150225T000000,
price: 180000,
bedrooms: 2,
bathrooms: 1,
sqft_living: 770,

```
sqft_lot: 10000,
floors: 1,
waterfront: 0,
view: 0,
condition: 3,
grade: 6,
sqft_above: 770,
sqft_basement: 0,
yr_built: 1933,
yr_renovated: 0,
zipcode: 98028,
lat: 47.7379,
long: -122.233,
sqft_living15: 2720,
sqft_lot15: 8062,
},
```

Find one document with invoice number is equals to 536365

```python
[6]: # Find one document
     result = collection.find({"price" : 221900}).limit(1)

     for rec in result:
         print("{")
         for key in rec.keys():
             print(f"{key}: {rec[key]},")
         print("},")
```

```
{
_id: 6666a3fdb1a4744f2a70a99b,
id: 7129300520,
date: 20141013T000000,
price: 221900,
bedrooms: 3,
bathrooms: 1,
sqft_living: 1180,
sqft_lot: 5650,
floors: 1,
waterfront: 0,
view: 0,
condition: 3,
grade: 7,
sqft_above: 1180,
sqft_basement: 0,
yr_built: 1955,
yr_renovated: 0,
zipcode: 98178,
lat: 47.5112,
long: -122.257,
```

```
    sqft_living15: 1340,
    sqft_lot15: 5650,
    },
```

[7]:
```python
query = {'sqft_living15': 1340}

result = collection.find(query)

# Convert the result to a pandas DataFrame
df = pd.DataFrame(result)

# Display the DataFrame
df.head()
```

[7]:
```
                        _id          id             date      price  bedrooms  \
0  6666a3fdb1a4744f2a70a99b  7129300520  20141013T000000   221900.0         3
1  6666a3fdb1a4744f2a70a9c7  1432701230  20140729T000000   309000.0         3
2  6666a3fdb1a4744f2a70ab01   325059171  20140505T000000   900000.0         3
3  6666a3fdb1a4744f2a70abc4  3294700310  20140902T000000   261000.0         2
4  6666a3fdb1a4744f2a70ace3  2843200070  20141215T000000   282000.0         4

   bathrooms  sqft_living  sqft_lot  floors  waterfront  …  grade  \
0       1.00         1180      5650     1.0           0  …      7
1       1.00         1280      9656     1.0           0  …      6
2       1.00         1330     77972     1.0           0  …      7
3       1.00          750      8125     1.0           0  …      6
4       1.75         1660     10725     1.0           0  …      7

   sqft_above  sqft_basement  yr_built  yr_renovated  zipcode      lat  \
0        1180              0      1955             0    98178  47.5112
1         920            360      1959             0    98058  47.4485
2        1330              0      1928          1954    98033  47.6891
3         750              0      1943             0    98055  47.4727
4         960            700      1956             0    98168  47.5033

      long  sqft_living15  sqft_lot15
0 -122.257           1340        5650
1 -122.175           1340        8808
2 -122.159           1340       17689
3 -122.198           1340        8750
4 -122.300           1340        9023

[5 rows x 22 columns]
```

[8]:

```
[12]: query = {'bedrooms': 1}
      result = collection.find(query)

      # Convert the result to a pandas DataFrame
      df = pd.DataFrame(result)
```

```
[11]: df.head()
```

```
[11]:                       _id          id             date      price  bedrooms  \
      0  6666a3fdb1a4744f2a70a99b  7129300520  20141013T000000  221900.0         3
      1  6666a3fdb1a4744f2a70a99c  6414100192  20141209T000000  538000.0         3
      2  6666a3fdb1a4744f2a70a99f  1954400510  20150218T000000  510000.0         3
      3  6666a3fdb1a4744f2a70a9a1  1321400060  20140627T000000  257500.0         3
      4  6666a3fdb1a4744f2a70a9a2  2008000270  20150115T000000  291850.0         3

         bathrooms  sqft_living  sqft_lot  floors  waterfront  …  grade  \
      0       1.00         1180      5650     1.0           0  …      7
      1       2.25         2570      7242     2.0           0  …      7
      2       2.00         1680      8080     1.0           0  …      8
      3       2.25         1715      6819     2.0           0  …      7
      4       1.50         1060      9711     1.0           0  …      7

         sqft_above  sqft_basement  yr_built  yr_renovated  zipcode      lat  \
      0        1180              0      1955             0    98178  47.5112
      1        2170            400      1951          1991    98125  47.7210
      2        1680              0      1987             0    98074  47.6168
      3        1715              0      1995             0    98003  47.3097
      4        1060              0      1963             0    98198  47.4095

             long  sqft_living15  sqft_lot15
      0  -122.257           1340        5650
      1  -122.319           1690        7639
      2  -122.045           1800        7503
      3  -122.327           2238        6819
      4  -122.315           1650        9711

      [5 rows x 22 columns]
```

```
[13]: df.tail(5)
```

```
[13]:                        _id          id             date      price  \
      194  6666a3ffb1a4744f2a70fce9  6003000851  20140522T000000  353000.0
      195  6666a3ffb1a4744f2a70fcef  9266700190  20150511T000000  245000.0
      196  6666a3ffb1a4744f2a70fd5e  3278600750  20150407T000000  250000.0
      197  6666a3ffb1a4744f2a70fdaa  1702901618  20150407T000000  420000.0
      198  6666a3ffb1a4744f2a70fdcb   688000017  20140627T000000  516500.0
```

```
        bedrooms  bathrooms  sqft_living  sqft_lot  floors  waterfront  …  \
194            1       1.00          550      1279     2.0           0  …
195            1       1.00          390      2000     1.0           0  …
196            1       1.50         1180      1688     2.0           0  …
197            1       2.00         1070       675     2.0           0  …
198            1       1.25         1100       638     3.0           0  …


        grade  sqft_above  sqft_basement  yr_built  yr_renovated  zipcode  \
194         7         550              0      2008             0    98122
195         6         390              0      1920             0    98103
196         8        1070            110      2007             0    98126
197         8         880            190      2007             0    98118
198         9        1100              0      2014             0    98112


          lat     long  sqft_living15  sqft_lot15
194   47.6160 -122.314           1460        1385
195   47.6938 -122.347           1340        5100
196   47.5490 -122.372           1380        2059
197   47.5574 -122.284           1220         788
198   47.6228 -122.307           1110        1933

[5 rows x 22 columns]
```

[14]: `list(df.columns)`

[14]:
```
['_id',
 'id',
 'date',
 'price',
 'bedrooms',
 'bathrooms',
 'sqft_living',
 'sqft_lot',
 'floors',
 'waterfront',
 'view',
 'condition',
 'grade',
 'sqft_above',
 'sqft_basement',
 'yr_built',
 'yr_renovated',
 'zipcode',
 'lat',
 'long',
 'sqft_living15',
 'sqft_lot15']
```

```
[19]: query = {'bedrooms': {"$gt" : 6}, "price" : {"$lt" : 1000000}}
      result = collection.find(query)
      df = pd.DataFrame(result)
```

```
[20]: df
```

```
[20]:                          _id          id            date    price  bedrooms  \
      0   6666a3feb1a4744f2a70ae0a  4024100951  20150105T000000   420000         7
      1   6666a3feb1a4744f2a70ae74  7227802030  20140623T000000   350000         7
      2   6666a3feb1a4744f2a70b017  9126101740  20141204T000000   490000         8
      3   6666a3feb1a4744f2a70b824  5451100490  20150115T000000   884900         7
      4   6666a3feb1a4744f2a70b953  9175600025  20141007T000000   800000         7
      5   6666a3feb1a4744f2a70b97e  7226500100  20150219T000000   373000         8
      6   6666a3feb1a4744f2a70b99b  1997200215  20140507T000000   599999         9
      7   6666a3feb1a4744f2a70ba26  2902200015  20150106T000000   700000         9
      8   6666a3feb1a4744f2a70bc08  2025049006  20141112T000000   750000         7
      9   6666a3feb1a4744f2a70bc33  4178300130  20150413T000000   950000         7
      10  6666a3feb1a4744f2a70bc9e  5561300480  20150408T000000   600000         7
      11  6666a3feb1a4744f2a70c1b9  1873400020  20140703T000000   340000         8
      12  6666a3feb1a4744f2a70c50a   686400380  20141002T000000   770000         7
      13  6666a3feb1a4744f2a70c953  4006000183  20140909T000000   450000         7
      14  6666a3feb1a4744f2a70cadf  3425059141  20140528T000000   999000         7
      15  6666a3feb1a4744f2a70cafd   424049043  20140811T000000   450000         9
      16  6666a3feb1a4744f2a70cb50  2926049086  20141028T000000   575000         7
      17  6666a3feb1a4744f2a70cbd0  1773100755  20140821T000000   520000        11
      18  6666a3feb1a4744f2a70cd10  6746700615  20150318T000000   700000         8
      19  6666a3feb1a4744f2a70cd47  1471700410  20150506T000000   310000         7
      20  6666a3feb1a4744f2a70ce87  2769600590  20141016T000000   900000         8
      21  6666a3feb1a4744f2a70cf56  1683400165  20150430T000000   853800         7
      22  6666a3feb1a4744f2a70cf9f   263000324  20140513T000000   550000         7
      23  6666a3feb1a4744f2a70d1be  1420700030  20140922T000000   597157         7
      24  6666a3ffb1a4744f2a70d3be  6411600370  20140515T000000   475000         7
      25  6666a3ffb1a4744f2a70d88c   424069271  20141024T000000   936000         7
      26  6666a3ffb1a4744f2a70dad8  2862100366  20141015T000000   730000         7
      27  6666a3ffb1a4744f2a70dadb  3812400455  20141104T000000   291000         7
      28  6666a3ffb1a4744f2a70db78  7663700968  20140528T000000   565000         7
      29  6666a3ffb1a4744f2a70dbf0  8813400155  20141219T000000   808000         8
      30  6666a3ffb1a4744f2a70dcdb  6453300306  20140910T000000   419000         7
      31  6666a3ffb1a4744f2a70e114  6190701484  20150311T000000   450000         7
      32  6666a3ffb1a4744f2a70e3de  9187200345  20140709T000000   599000         7
      33  6666a3ffb1a4744f2a70e479  7418000130  20141211T000000   430000         8
      34  6666a3ffb1a4744f2a70e4d4  5566100170  20141029T000000   650000        10
      35  6666a3ffb1a4744f2a70e63c  3052700225  20140814T000000   727160         7
      36  6666a3ffb1a4744f2a70e6d1  8141200080  20140814T000000   680000         8
      37  6666a3ffb1a4744f2a70e722  6744700427  20140507T000000   540000         7
      38  6666a3ffb1a4744f2a70e799  2402100895  20140625T000000   640000        33
      39  6666a3ffb1a4744f2a70e954  2887950020  20140625T000000   280000         7
```

```
40  6666a3ffb1a4744f2a70ed52  2856100381  20140711T000000  580000          7
41  6666a3ffb1a4744f2a70f1a6  8823901445  20150313T000000  934000          9
42  6666a3ffb1a4744f2a70f1b7  2616800600  20140530T000000  840000          7
43  6666a3ffb1a4744f2a70f322  4040500100  20141020T000000  539000          7
44  6666a3ffb1a4744f2a70f4d1  8812401450  20141229T000000  660000         10
45  6666a3ffb1a4744f2a70f501  3756900027  20141125T000000  575000          8
46  6666a3ffb1a4744f2a70f51a  2771604190  20140617T000000  824000          7
```

	bathrooms	sqft_living	sqft_lot	floors	waterfront	…	grade \
0	3.00	2940	8624	1.0	0	…	8
1	3.00	2800	9569	1.0	0	…	7
2	5.00	2800	2580	2.0	0	…	8
3	4.75	5370	10800	1.5	0	…	8
4	6.75	7480	41664	2.0	0	…	11
5	3.00	2850	12714	1.0	0	…	7
6	4.50	3830	6988	2.5	0	…	7
7	3.00	3680	4400	2.0	0	…	7
8	2.75	3410	4056	1.5	0	…	8
9	3.50	3470	16264	2.0	0	…	9
10	2.25	3170	36384	2.0	0	…	8
11	2.75	2790	6695	1.0	0	…	7
12	2.25	3260	8145	2.0	0	…	8
13	4.00	3150	7800	2.0	0	…	8
14	4.00	3150	34830	1.0	0	…	9
15	7.50	4050	6504	2.0	0	…	7
16	1.50	2670	11250	1.5	0	…	8
17	3.00	3000	4960	2.0	0	…	7
18	2.50	2280	3000	1.5	0	…	7
19	1.50	2660	15111	1.5	0	…	7
20	4.00	4020	7500	1.0	0	…	8
21	4.00	2960	2665	2.0	0	…	9
22	4.00	3440	8100	2.0	0	…	7
23	4.00	2690	10880	1.0	0	…	8
24	3.50	2870	29699	1.0	0	…	7
25	3.75	5100	21802	2.0	0	…	10
26	2.75	3110	4400	1.5	0	…	7
27	1.00	2350	8636	1.0	0	…	7
28	4.50	4140	9066	1.0	0	…	7
29	3.75	3460	4600	2.0	0	…	7
30	3.25	4340	8521	2.0	0	…	7
31	3.50	2830	8625	1.0	0	…	8
32	2.50	2580	5750	1.0	0	…	7
33	3.25	4300	10441	2.0	0	…	8
34	2.00	3610	11914	2.0	0	…	7
35	3.75	2310	5000	2.0	0	…	8
36	2.75	2530	4800	2.0	0	…	7
37	5.75	3700	7647	2.0	0	…	8

38	1.75	1620	6000	1.0	0	…	7
39	2.50	1940	5458	2.0	0	…	7
40	2.75	2310	2400	1.5	0	…	6
41	3.00	2820	4480	2.0	0	…	7
42	4.50	4290	37607	1.5	0	…	10
43	2.25	2620	6890	2.0	0	…	7
44	3.00	2920	3745	2.0	0	…	7
45	3.00	3840	15990	1.0	0	…	7
46	4.25	3670	4000	2.0	0	…	8

	sqft_above	sqft_basement	yr_built	yr_renovated	zipcode	lat \
0	1690	1250	1977	0	98155	47.7555
1	1400	1400	1963	0	98056	47.5102
2	1880	920	1997	0	98122	47.6086
3	5370	0	1967	0	98040	47.5380
4	5080	2400	1953	0	98166	47.4643
5	2850	0	1959	0	98055	47.4859
6	2450	1380	1938	0	98103	47.6927
7	2830	850	1908	0	98102	47.6374
8	2130	1280	1906	0	98102	47.6454
9	3470	0	1980	0	98007	47.6203
10	3170	0	1969	0	98027	47.4654
11	1470	1320	1977	0	98133	47.7565
12	3260	0	1967	0	98008	47.6336
13	3150	0	2013	0	98118	47.5259
14	3150	0	1957	2005	98007	47.6029
15	4050	0	1996	0	98144	47.5923
16	2320	350	1948	0	98133	47.7121
17	2400	600	1918	1999	98106	47.5560
18	1210	1070	1911	0	98105	47.6675
19	2660	0	1962	0	98059	47.4644
20	2010	2010	1968	0	98107	47.6732
21	1950	1010	1927	2013	98144	47.5835
22	3440	0	1970	0	98103	47.6981
23	2690	0	1960	0	98033	47.6787
24	1520	1350	1961	0	98125	47.7153
25	3640	1460	2001	0	98075	47.5950
26	2010	1100	1914	0	98105	47.6684
27	1550	800	1962	0	98118	47.5432
28	2070	2070	1978	0	98125	47.7302
29	2860	600	1987	0	98105	47.6617
30	2550	1790	1986	0	98106	47.5200
31	1830	1000	1957	0	98133	47.7493
32	1880	700	1901	0	98122	47.6025
33	2800	1500	1979	0	98059	47.4786
34	3010	600	1958	0	98006	47.5705
35	2310	0	1984	0	98117	47.6781

36	1390	1140	1901	0	98112	47.6241
37	3700	0	1948	1984	98155	47.7393
38	1040	580	1947	0	98103	47.6878
39	1940	0	1994	0	98092	47.3191
40	2310	0	1915	0	98117	47.6775
41	1880	940	1918	0	98105	47.6654
42	4290	0	1982	0	98027	47.4812
43	2620	0	1961	0	98007	47.6123
44	1860	1060	1913	0	98105	47.6635
45	2530	1310	1961	0	98034	47.7111
46	2800	870	1964	0	98199	47.6375

	long	sqft_living15	sqft_lot15
0	-122.307	1850	8031
1	-122.183	2150	7333
2	-122.303	1800	2580
3	-122.223	2310	10910
4	-122.368	2810	33190
5	-122.205	1480	4942
6	-122.338	1460	6291
7	-122.324	1960	2450
8	-122.316	2510	4056
9	-122.149	3040	13500
10	-122.003	2460	38370
11	-122.331	1760	7624
12	-122.115	2340	8145
13	-122.279	1880	6000
14	-122.147	2390	12054
15	-122.301	1448	3866
16	-122.332	2030	9000
17	-122.363	1420	4960
18	-122.316	1610	3000
19	-122.066	1710	15429
20	-122.363	1560	3737
21	-122.313	1970	4410
22	-122.349	1420	1560
23	-122.168	1840	10836
24	-122.327	1380	7555
25	-122.040	3350	10005
26	-122.319	1240	4280
27	-122.277	1500	7366
28	-122.291	1440	1865
29	-122.289	2170	3750
30	-122.338	1890	8951
31	-122.355	1570	8400
32	-122.294	2280	5750
33	-122.131	1780	10457

```
34  -122.175              2040          11914
35  -122.376              1360           1552
36  -122.305              1540           4800
37  -122.289              2510           7479
38  -122.331              1330           4700
39  -122.177              1710           5688
40  -122.390              1340           3825
41  -122.307              2460           4400
42  -122.033              2810          40510
43  -122.134              2070           7910
44  -122.320              1810           3745
45  -122.211              1380           8172
46  -122.388              2010           4000

[47 rows x 22 columns]
```

```
[21]:  query = {'bedrooms': {"$gt" : 6}, "price" : {"$gt" : 1000000}}
       result = collection.find(query)
       df = pd.DataFrame(result)
```

```
[22]:  df
```

```
[22]:                            _id           id              date        price  \
       0   6666a3fdb1a4744f2a70abc7   5486800070  20140620T000000   1950000.0
       1   6666a3feb1a4744f2a70b95e    685000115  20141007T000000   2150000.0
       2   6666a3feb1a4744f2a70c15a   9822700190  20140808T000000   1280000.0
       3   6666a3feb1a4744f2a70c5b2   7159200005  20140507T000000   3200000.0
       4   6666a3ffb1a4744f2a70d469   1652500060  20140711T000000   1650000.0
       5   6666a3ffb1a4744f2a70db84   1225069038  20140505T000000   2280000.0
       6   6666a3ffb1a4744f2a70dd9d    627300145  20140814T000000   1148000.0
       7   6666a3ffb1a4744f2a70e069   5316101075  20140926T000000   2885000.0
       8   6666a3ffb1a4744f2a70e449   2210500010  20140930T000000   2450000.0
       9   6666a3ffb1a4744f2a70e529   5536100005  20140808T000000   2300000.0
       10  6666a3ffb1a4744f2a70eafb   2818600010  20150314T000000   1185000.0
       11  6666a3ffb1a4744f2a70eb67   8823900290  20150317T000000   1400000.0
       12  6666a3ffb1a4744f2a70ecee   1338600225  20140528T000000   1970000.0
       13  6666a3ffb1a4744f2a70f1c8   1346300150  20141020T000000   3300000.0
       14  6666a3ffb1a4744f2a70f3ba   1778360150  20140620T000000   1240000.0

           bedrooms  bathrooms  sqft_living  sqft_lot  floors  waterfront  … \
       0          7       3.50         4640     15235     2.0           0  …
       1          8       6.00         4340      9415     2.0           0  …
       2          9       4.50         3650      5000     2.0           0  …
       3          7       4.50         6210      8856     2.5           0  …
       4          8       2.75         4040     20666     1.0           0  …
       5          7       8.00        13540    307752     3.0           0  …
       6         10       5.25         4590     10920     1.0           0  …
```

```
7             7        3.00       5350      14400      2.5          0  …
8             7        4.25       4670      23115      2.0          0  …
9             7        4.75       5310       8816      2.0          0  …
10            7        3.50       3890       8342      2.0          0  …
11            9        4.00       4620       5508      2.5          0  …
12            8        3.50       4440       6480      2.0          0  …
13            8        4.00       7710      11750      3.5          0  …
14            7        5.50       6630      13782      2.0          0  …

      grade  sqft_above  sqft_basement  yr_built  yr_renovated  zipcode  \
0        11        2860           1780      1965          2003    98040
1         8        4340              0      1967             0    98004
2         8        2530           1120      1915          2010    98105
3        11        4760           1450      1910             0    98109
4         9        2020           2020      1962             0    98004
5        12        9410           4130      1999             0    98053
6         9        2500           2090      2008             0    98004
7        10        5020            330      1910             0    98112
8        11        4670              0      1992             0    98039
9        10        3650           1660      2013             0    98004
10        9        2840           1050      1968             0    98117
11       11        3870            750      1915             0    98105
12       10        3140           1300      1959             0    98112
13       12        6090           1620      1904             0    98112
14       10        4930           1700      2004             0    98006

        lat     long  sqft_living15  sqft_lot15
0   47.5666 -122.231           3230       20697
1   47.6316 -122.202           2050        9100
2   47.6604 -122.289           2510        5000
3   47.6307 -122.354           2940        5400
4   47.6340 -122.221           3670       20500
5   47.6675 -121.986           4850      217800
6   47.5861 -122.113           2730       10400
7   47.6295 -122.285           3050        7469
8   47.6183 -122.227           3240       13912
9   47.6221 -122.208           2920       10610
10  47.7011 -122.392           2870        8342
11  47.6684 -122.309           2710        4320
12  47.6310 -122.303           4440        8640
13  47.6263 -122.314           4210        8325
14  47.5399 -122.118           4470        8639

[15 rows x 22 columns]
```

[ ]: