Data Wrangling with MongoDB

DASHBOARD

CLASSROOM

MATERIALS

DISCUSSION

OVERVIEW

Lesson 5: Problem Set	>	Average Population

```python
18  Your code will be run against a MongoDB instance that we have provided. If you want to run this code
19  locally on your machine, you have to install MongoDB, download and insert the dataset.
20  For instructions related to MongoDB setup and datasets please see Course Materials.
21
22  Please note that the dataset you are using here is a different version of the cities collection
23  provided in the course materials. If you attempt some of the same queries that we look at in the
24  problem set, your results may be different.
25  """
26
27  def get_db(db_name):
28      from pymongo import MongoClient
29      client = MongoClient('localhost:27017')
30      db = client[db_name]
31      return db
32
33  def make_pipeline():
34      # complete the aggregation pipeline
35      pipeline = [{"$match" : {"country" : {"$exists" : "true"}}},
36          {"$unwind" : "$isPartOf"},
37          {"$group" : {"_id" : "$isPartOf",
38                      "avgCityPopulation" : { "$avg" : "$population"}}},
39          #{"$group" :{ "_id" : "$_id",
40          #           "avgRegionalPopulation" : {"$avg" : "$avgCityPopulation"}}},
41          #{"$group" : {"country" :{"$_id" : "$country"},
42          #            "avgRegionalPopulation" : {"$avg" :"$avgRegionalPopulation"}}}
43          ]
44      return pipeline
```

Reset	Test Run		Submit

```
Traceback (most recent call last):
  File "vm_main.py", line 33, in <module>
    import main
  File "/tmp/vmuser_rtpqfrnjro/main.py", line 2, in <module>
    import studentMain
  File "/tmp/vmuser_rtpqfrnjro/studentMain.py", line 54, in <module>
    assert {'_id': 'Lithuania', 'avgRegionalPopulation': key_pop} in result
AssertionError
```

```
[{u'_id': u'Champhai district', u'avgCityPopulation': 2408.0},
 {u'_id': u'Kabirdham district', u'avgCityPopulation': 31788.0},
 {u'_id': u'Kanker district', u'avgCityPopulation': 24485.0},
 {u'_id': u'Atsimo-Atsinanana', u'avgCityPopulation': 24252.0},
 {u'_id': u'Bhopal district', u'avgCityPopulation': 24289.0},
 {u'_id': u'Bhadrak district', u'avgCityPopulation': 29998.0},
 {u'_id': u'Balod_district', u'avgCityPopulation': 35829.5},
 {u'_id': u'Kishanganj district', u'avgCityPopulation': 28224.0},
 {u'_id': u'Chikka_balapur_district', u'avgCityPopulation': 20120.0},
 {u'_id': u'Panchsheel_nagar_India', u'avgCityPopulation': 5938.0},
 {u'_id': u'Ramanagara', u'avgCityPopulation': 71912.0},
 {u'_id': u'Adzop\xe9 Department', u'avgCityPopulation': 61884.0},
 {u'_id': u'West Garo Hills district', u'avgCityPopulation': 58391.0},
 {u'_id': u'Anantnag district', u'avgCityPopulation': 64104.0},
 {u'_id': u'Debagarh district', u'avgCityPopulation': 20085.0},
 {u'_id': u'Meerut division', u'avgCityPopulation': 102733.0},
 {u'_id': u'Santiago Papasquiaro Municipality', u'avgCityPopulation': 47360.0},
 {u'_id': u'Garissa County', u'avgCityPopulation': 65881.0},
 {u'_id': u'Francisco I. Madero Municipality Coahuila',
  u'avgCityPopulation': 50084.0},
 {u'_id': u'Alto Guadalquivir', u'avgCityPopulation': 9915.0},
 {u'_id': u'Puerto Plata Province', u'avgCityPopulation': 286558.0},
 {u'_id': u'Mon district', u'avgCityPopulation': 16119.0},
 {u'_id': u'Mafraq Governorate', u'avgCityPopulation': 56340.0},
```

Get Help

Discussion Topics

Report an Issue

{u'_id': u'Ash Sharqiyah Region', u'avgCityPopulation': 0.0},
{u'_id': u'Southeastern Statistical Region', u'avgCityPopulation': 0.0},
{u'_id': u'Baran Rajasthan', u'avgCityPopulation': 78372.0},
{u'_id': u'Puerto L\xf3pez Canton', u'avgCityPopulation': 0.0},
{u'_id': u'Banyumas Regency', u'avgCityPopulation': 249705.0},
{u'_id': u'Jafarabad District', u'avgCityPopulation': 0.0},
{u'_id': u'Awaran Tehsil', u'avgCityPopulation': 0.0},
{u'_id': u'Awaran District', u'avgCityPopulation': 0.0},
{u'_id': u'Vakinankaratra', u'avgCityPopulation': 903450.0},
{u'_id': u'Upper Assam', u'avgCityPopulation': 97782.0},
{u'_id': u'Golaghat district', u'avgCityPopulation': 55573.0},
{u'_id': u'Bankura district', u'avgCityPopulation': 138036.0},
{u'_id': u'Godda district', u'avgCityPopulation': 37507.0},
{u'_id': u'Saidpur', u'avgCityPopulation': 264209.0},
{u'_id': u'Nilphamari District', u'avgCityPopulation': 264209.0},
{u'_id': u'Yobe State', u'avgCityPopulation': 88014.0},
{u'_id': u'Santa Catarina (state)', u'avgCityPopulation': 130700.5},
{u'_id': u'Curuz\xfa Cuati\xe1 Department', u'avgCityPopulation': 33089.0},
{u'_id': u'San Juan del R\xedo Municipality Quer\xe9taro',
 u'avgCityPopulation': 120984.0},
{u'_id': u'Basque Country', u'avgCityPopulation': 17.0},
{u'_id': u'Kostiantynivka Raion', u'avgCityPopulation': 94886.0},
{u'_id': u'Ti\u1ec1n Giang Province', u'avgCityPopulation': 220000.0},
{u'_id': u'Sharur District', u'avgCityPopulation': 6964.0},
{u'_id': u'Shahbuz District', u'avgCityPopulation': 3127.0},
{u'_id': u'Qazakh District', u'avgCityPopulation': 20900.0},
{u'_id': u'Kurdamir District', u'avgCityPopulation': 19088.0},
{u'_id': u'Shabran District', u'avgCityPopulation': 23248.0},
{u'_id': u'Balakan District', u'avgCityPopulation': 10400.0},
{u'_id': u'Moravian–Silesian Region', u'avgCityPopulation': 63045.0},
{u'_id': u'Karvin\xe1 District', u'avgCityPopulation': 63045.0},
{u'_id': u'Jhelum District', u'avgCityPopulation': 37000.0},
{u'_id': u'Kaduna State', u'avgCityPopulation': 408198.0},
{u'_id': u'Siwan district', u'avgCityPopulation': 108172.0},
{u'_id': u'Torres Municipality', u'avgCityPopulation': 105400.0},
{u'_id': u'Lara Venezuela', u'avgCityPopulation': 105400.0},
{u'_id': u'Barisal Division', u'avgCityPopulation': 272169.0},
{u'_id': u'Targovishte Municipality', u'avgCityPopulation': 37341.0},
{u'_id': u'Southcentral_Paran%C3%A1', u'avgCityPopulation': 38011.0},
{u'_id': u'Paran\xe1 (state)', u'avgCityPopulation': 38011.0},
{u'_id': u'Palma_(microrregi%C3%A3o)', u'avgCityPopulation': 38011.0},
{u'_id': u'Dakshin Dinajpur district', u'avgCityPopulation': 151183.0},
{u'_id': u'Western Province Zambia', u'avgCityPopulation': 44310.0},
{u'_id': u'Mongu District', u'avgCityPopulation': 44310.0},
{u'_id': u'Yenangyaung Township', u'avgCityPopulation': 0.0},
{u'_id': u'Mbeya Urban District', u'avgCityPopulation': 0.0},
{u'_id': u'Shinyanga Region', u'avgCityPopulation': 107360.0},
{u'_id': u'Arismendi_Municipality_Nueva_Esparta',
 u'avgCityPopulation': 28500.0},
{u'_id': u'Nnewi North', u'avgCityPopulation': 391227.0},
{u'_id': u'Clarendon Parish', u'avgCityPopulation': 0.0},
{u'_id': u'Mymensingh District', u'avgCityPopulation': 407798.0},
{u'_id': u'Dhaka', u'avgCityPopulation': 407798.0},
{u'_id': u'Kremenets Raion', u'avgCityPopulation': 21880.0},
{u'_id': u'Catemaco Veracruz', u'avgCityPopulation': 22965.0},
{u'_id': u'Pando Department', u'avgCityPopulation': 55692.0},
{u'_id': u'Legislative district of Southern Leyte',
 u'avgCityPopulation': 81737.0},
{u'_id': u'Manokwari Regency', u'avgCityPopulation': 136302.0},
{u'_id': u'Legislative districts of Misamis Occidental',
 u'avgCityPopulation': 31326.0},
{u'_id': u'Pyay Township', u'avgCityPopulation': 123800.0},
{u'_id': u'Narayanganj District', u'avgCityPopulation': 1572386.0},
{u'_id': u'Afyonkarahisar Province', u'avgCityPopulation': 0.0},
{u'_id': u'San Miguel Department (El Salvador)',
 u'avgCityPopulation': 183733.0},

    {u'_id': u'Anseba Region', u'avgCityPopulation': 82198.0},
    {u'_id': u'Barmer district', u'avgCityPopulation': 83517.0},
    {u'_id': u'Hvar', u'avgCityPopulation': 4138.0},
    {u'_id': u'Capital_Department_Misiones', u'avgCityPopulation': 344833.0},
    {u'_id': u'Srinagar district', u'avgCityPopulation': 13477.0},
    {u'_id': u'Camargo Municipality Chihuahua', u'avgCityPopulation': 39149.0},
    {u'_id': u'Claiborne County Tennessee', u'avgCityPopulation': 4389.0},
    {u'_id': u'T\xe9miscouata (electoral district)',
     u'avgCityPopulation': 3199.0},
    {u'_id': u'Xiangxi Tujia and Miao Autonomous Prefecture',
     u'avgCityPopulation': 291200.0},
    {u'_id': u"Ma'an Governorate", u'avgCityPopulation': 50350.0},
    {u'_id': u'Linfen', u'avgCityPopulation': 290000.0},
    {u'_id': u'Ad Dakhiliyah Governorate', u'avgCityPopulation': 0.0},
    {u'_id': u'Biskra Province', u'avgCityPopulation': 307987.0},
    {u'_id': u'Biskra District', u'avgCityPopulation': 307987.0},
    {u'_id': u'Sakarya Province', u'avgCityPopulation': 0.0},
    {u'_id': u'Patna division', u'avgCityPopulation': 41507.0}]

Instructor Notes

This is an example document:

```
{
    "_id" : ObjectId("52fe1d364b5ab856eea75ebc"),
    "elevation" : 1855,
    "name" : "Kud",
    "country" : "India",
    "lon" : 75.28,
    "lat" : 33.08,
    "isPartOf" : [
        "Jammu and Kashmir",
        "Udhampur district"
    ],
    "timeZone" : [
        "Indian Standard Time"
    ],
    "population" : 1140
}
```

Downloadables

There are no relevant downloads for this part of the course.