

Proiect "Administrarea bazelor de date"

1. Overview

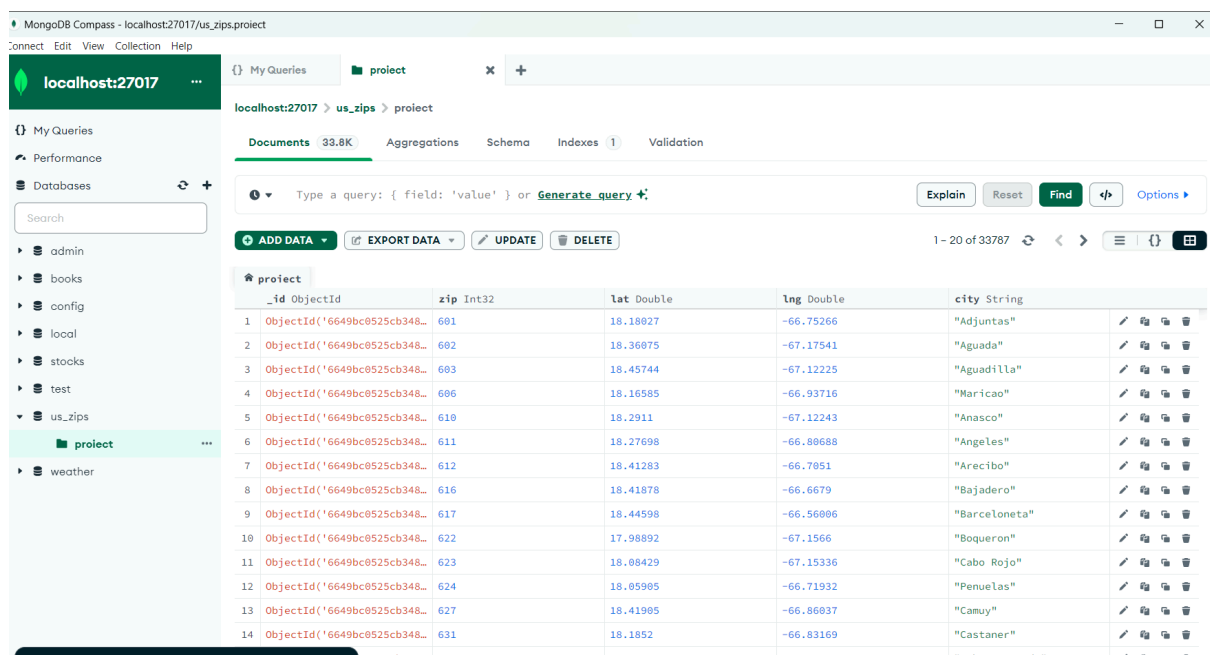
The goal of the project is to assess the MongoDB database skills of the student. The project has a medium difficulty level and is relevant to industry employers of today.

2. Project description

You are required to make several statistical computations on some US Zips dataset using MongoDB as the database platform.

Prerequisites:

- Download the latest US Zips dataset from <https://simplemaps.com/data/us-zips> (choose the free tier). The dataset has approximately 33k entries.
- Create a MongoDB instance. You may use your own MongoDB Atlas instance in cloud or use a local instance. For local instances Docker is preferred, but you may also choose to install MongoDB as a standalone server on your OS.
- Import the dataset into the MongoDB instance.



Requirements:

a) Get the states with a total population of over 10 million.

```
us_zips> db.project.aggregate([ { $group: { _id: "$state_name", totalPopulation: { $sum: { $toInt: "$population" } } }, { $match: { totalPopulation: { $gt: 10000000 } } }, { $sort: { totalPopulation: -1 } } ] )
```

```

us_zips> db.proiect.aggregate([ { $group: { _id: "$state_name", totalPopulation: { $sum: { $toInt: "$population" } } } }
{ $match: { totalPopulation: { $gt: 10000000 } } }, { $sort: { totalPopulation: -1 } } ] )
[
  { _id: 'California', totalPopulation: 39354820 },
  { _id: 'Texas', totalPopulation: 29242696 },
  { _id: 'Florida', totalPopulation: 21632200 },
  { _id: 'New York', totalPopulation: 19994379 },
  { _id: 'Pennsylvania', totalPopulation: 12989208 },
  { _id: 'Illinois', totalPopulation: 12757583 },
  { _id: 'Ohio', totalPopulation: 11774683 },
  { _id: 'Georgia', totalPopulation: 10722352 },
  { _id: 'North Carolina', totalPopulation: 10470214 },
  { _id: 'Michigan', totalPopulation: 10057902 }
]
us_zips>

```

b) Get the average city population by state.

```

us_zips> db.proiect.aggregate([ { $group: { _id: "$state_name", avgCityPopulation: { $avg: {
$toInt: "$population" } } } }, { $sort: { _id: 1 } } ] )

```

```

us_zips> db.proiect.aggregate([ { $group: { _id: "$state_name", avgCityPopulation: { $avg: { $toInt: "$population" } } }
}, { $sort: { _id: 1 } } ] )
[
  { _id: 'Alabama', avgCityPopulation: 7664.599085365854 },
  { _id: 'Alaska', avgCityPopulation: 2996.995918367347 },
  { _id: 'American Samoa', avgCityPopulation: null },
  { _id: 'Arizona', avgCityPopulation: 17203.745803357313 },
  { _id: 'Arkansas', avgCityPopulation: 4908.718699186992 },
  { _id: 'California', avgCityPopulation: 21839.522752497225 },
  { _id: 'Colorado', avgCityPopulation: 10950.96394686907 },
  { _id: 'Connecticut', avgCityPopulation: 12539.295138888889 },
  { _id: 'Delaware', avgCityPopulation: 14618.882352941177 },
  {
    _id: 'District of Columbia',
    avgCityPopulation: 11764.684210526315
  },
  { _id: 'Florida', avgCityPopulation: 21396.83481701286 },
  { _id: 'Georgia', avgCityPopulation: 14277.4327563249 },
  { _id: 'Guam', avgCityPopulation: null },
  { _id: 'Hawaii', avgCityPopulation: 14954.525773195875 },
  { _id: 'Idaho', avgCityPopulation: 6644.727598566308 },
  { _id: 'Illinois', avgCityPopulation: 9138.669770773638 },
  { _id: 'Indiana', avgCityPopulation: 8406.942998760842 },
  { _id: 'Iowa', avgCityPopulation: 3287.955670103093 },
  { _id: 'Kansas', avgCityPopulation: 4170.129261363636 },
  { _id: 'Kentucky', avgCityPopulation: 5782.792307692308 }
]
Type "it" for more
us_zips>

```

c) Get the largest and the smallest city in each state.

```

us_zips> db.proiect.aggregate([ { $group: { _id: "$state_name", largestCity: { $max: { city:
"$city", population: { $toInt: "$population" } } }, smallestCity: { $min: { city: "$city", population:
{ $toInt: "$population" } } } } }, { $project: { _id: 1, largestCity: 1, smallestCity: 1 } } ] )

```

```
us_zips> db.proiect.aggregate([ { $group: { _id: "$state_name", largestCity: { $max: { city: "$city", population: { $toInt: "$population" } } }, smallestCity: { $min: { city: "$city", population: { $toInt: "$population" } } } } }, { $project: { _id: 1, largestCity: 1, smallestCity: 1 } } ] )
```

```
[
  {
    _id: 'Alaska',
    largestCity: { city: 'Yakutat', population: 564 },
    smallestCity: { city: 'Adak', population: 248 }
  },
  {
    _id: 'Alabama',
    largestCity: { city: 'York', population: 2961 },
    smallestCity: { city: 'Abbeville', population: 6071 }
  },
  {
    _id: 'Vermont',
    largestCity: { city: 'Worcester', population: 1238 },
    smallestCity: { city: 'Adamant', population: 251 }
  },
  {
    _id: 'South Dakota',
    largestCity: { city: 'Yankton', population: 20266 },
    smallestCity: { city: 'Aberdeen', population: 32217 }
  },
  {
    _id: 'Washington',
    largestCity: { city: 'Zillah', population: 7043 },
    smallestCity: { city: 'Aberdeen', population: 24989 }
  },
  {
    _id: 'Nebraska',
    largestCity: { city: 'Yutan', population: 2525 },
    smallestCity: { city: 'Abie', population: 59 }
  },
  {
    _id: 'Massachusetts',
    largestCity: { city: 'Yarmouth Port', population: 7185 },
    smallestCity: { city: 'Abington', population: 17003 }
  },
  {
    _id: 'Maine',
    largestCity: { city: 'Calais', population: 1000 },
    smallestCity: { city: 'Calais', population: 1000 }
  },
  {
    _id: 'New Hampshire',
    largestCity: { city: 'Durham', population: 18000 },
    smallestCity: { city: 'Durham', population: 18000 }
  },
  {
    _id: 'New Jersey',
    largestCity: { city: 'Newark', population: 28000 },
    smallestCity: { city: 'Newark', population: 28000 }
  },
  {
    _id: 'New York',
    largestCity: { city: 'New York City', population: 18000000 },
    smallestCity: { city: 'New York City', population: 18000000 }
  },
  {
    _id: 'Pennsylvania',
    largestCity: { city: 'Philadelphia', population: 1500000 },
    smallestCity: { city: 'Philadelphia', population: 1500000 }
  },
  {
    _id: 'Rhode Island',
    largestCity: { city: 'Providence', population: 60000 },
    smallestCity: { city: 'Providence', population: 60000 }
  },
  {
    _id: 'South Carolina',
    largestCity: { city: 'Columbia', population: 130000 },
    smallestCity: { city: 'Columbia', population: 130000 }
  },
  {
    _id: 'Texas',
    largestCity: { city: 'Houston', population: 2300000 },
    smallestCity: { city: 'Houston', population: 2300000 }
  },
  {
    _id: 'Virginia',
    largestCity: { city: 'Richmond', population: 20000 },
    smallestCity: { city: 'Richmond', population: 20000 }
  },
  {
    _id: 'Washington',
    largestCity: { city: 'Seattle', population: 700000 },
    smallestCity: { city: 'Seattle', population: 700000 }
  },
  {
    _id: 'West Virginia',
    largestCity: { city: 'Martinsburg', population: 10000 },
    smallestCity: { city: 'Martinsburg', population: 10000 }
  },
  {
    _id: 'Wisconsin',
    largestCity: { city: 'Milwaukee', population: 580000 },
    smallestCity: { city: 'Milwaukee', population: 580000 }
  },
  {
    _id: 'Wyoming',
    largestCity: { city: 'Cheyenne', population: 60000 },
    smallestCity: { city: 'Cheyenne', population: 60000 }
  }
]
```

```
{
  _id: 'Florida',
  largestCity: { city: 'Zolfo Springs', population: 4821 },
  smallestCity: { city: 'Alachua', population: 123 }
},
{
  _id: 'Iowa',
  largestCity: { city: 'Zwingle', population: 685 },
  smallestCity: { city: 'Ackley', population: 2468 }
},
{
  _id: 'North Carolina',
  largestCity: { city: 'Zirconia', population: 3358 },
  smallestCity: { city: 'Aberdeen', population: 13997 }
},
{
  _id: 'Virgin Islands',
  largestCity: { city: 'St Thomas', population: null },
  smallestCity: { city: 'Christiansted', population: null }
},
{
  _id: 'Oklahoma',
  largestCity: { city: 'Yukon', population: 83451 },
  smallestCity: { city: 'Achille', population: 425 }
},
{
  _id: 'Hawaii',
  largestCity: { city: 'Waipahu', population: 72605 },
  smallestCity: { city: 'Aiea', population: 41560 }
},
{
  _id: 'Delaware',
  largestCity: { city: 'Yorklyn', population: 17 },
  smallestCity: { city: 'Bear', population: 44553 }
},
{
  _id: 'Indiana',
  largestCity: { city: 'Zionsville', population: 34056 },
  smallestCity: { city: 'Advance', population: 412 }
},
}
```

```

    },
    {
      _id: 'Arkansas',
      largestCity: { city: 'Yorktown', population: 0 },
      smallestCity: { city: 'Adona', population: 598 }
    },
    {
      _id: 'American Samoa',
      largestCity: { city: 'Pago Pago', population: null },
      smallestCity: { city: 'Pago Pago', population: null }
    },
    {
      _id: 'Rhode Island',
      largestCity: { city: 'Wyoming', population: 1857 },
      smallestCity: { city: 'Albion', population: 702 }
    },
    {
      _id: 'New York',
      largestCity: { city: 'Yulan', population: 167 },
      smallestCity: { city: 'Accord', population: 3864 }
    },
    {
      _id: 'New Jersey',
      largestCity: { city: 'Zarephath', population: 74 },
      smallestCity: { city: 'Absecon', population: 10078 }
    }
  ]
}

```

d) Get the largest and the smallest counties in each state.

```

db.proiect.aggregate([
  {
    $group: {
      _id: { state_name: "$state_name", county_name: "$county_name" },
      totalPopulation: { $sum: { $toInt: "$population" } }
    }
  },
  {
    $group: {
      _id: "$_id.state_name",
      largestCounty: { $max: { county: "$_id.county_name", population: "$totalPopulation" } },
      smallestCounty: { $min: { county: "$_id.county_name", population: "$totalPopulation" } }
    }
  },
  {
    $project: {
      _id: 1,
      largestCounty: 1,
      smallestCounty: 1
    }
  }
],

```

```

    {
      $sort: { _id: 1 }
    }
  ]
})

```

```

us_zips> db.proiect.aggregate([
...   {
...     $group: {
...       _id: { state_name: "$state_name", county_name: "$county_name" },
...       totalPopulation: { $sum: { $toInt: "$population" } }
...     },
...     {
...       $group: {
...         _id: "$_id.state_name",
...         largestCounty: { $max: { county: "$_id.county_name", population: "$totalPopulation" } },
...         smallestCounty: { $min: { county: "$_id.county_name", population: "$totalPopulation" } }
...       }
...     },
...     {
...       $project: {
...         _id: 1,
...         largestCounty: 1,
...         smallestCounty: 1
...       }
...     },
...     {
...       $sort: { _id: 1 }
...     }
...   ]
... ])
[
  {
    _id: 'Alabama',
    largestCounty: { county: 'Winston', population: 24708 },
    smallestCounty: { county: 'Autauga', population: 57307 }
  },
  {
    _id: 'Alaska',
    largestCounty: { county: 'Yukon-Koyukuk', population: 23437 },
    smallestCounty: { county: 'Aleutians East', population: 3407 }
  },
]

```

```
{
  _id: 'American Samoa',
  largestCounty: { county: 'Western', population: 0 },
  smallestCounty: { county: 'Western', population: 0 }
},
{
  _id: 'Arizona',
  largestCounty: { county: 'Yuma', population: 203831 },
  smallestCounty: { county: 'Apache', population: 66596 }
},
{
  _id: 'Arkansas',
  largestCounty: { county: 'Yell', population: 20817 },
  smallestCounty: { county: 'Arkansas', population: 17387 }
},
{
  _id: 'California',
  largestCounty: { county: 'Yuba', population: 82873 },
  smallestCounty: { county: 'Alameda', population: 1681259 }
},
{
  _id: 'Colorado',
  largestCounty: { county: 'Yuma', population: 9810 },
  smallestCounty: { county: 'Adams', population: 540349 }
},
{
  _id: 'Connecticut',
  largestCounty: { county: 'Western Connecticut', population: 620748 },
  smallestCounty: { county: 'Capitol', population: 977741 }
},
{
  _id: 'Delaware',
  largestCounty: { county: 'Sussex', population: 250743 },
  smallestCounty: { county: 'Kent', population: 177092 }
},
{
  _id: 'District of Columbia',
  largestCounty: { county: 'District of Columbia', population: 670587 },
  smallestCounty: { county: 'District of Columbia', population: 670587 }
},
}
```

```

{
  _id: 'Florida',
  largestCounty: { county: 'Washington', population: 23107 },
  smallestCounty: { county: 'Alachua', population: 291490 }
},
{
  _id: 'Georgia',
  largestCounty: { county: 'Worth', population: 21176 },
  smallestCounty: { county: 'Appling', population: 18143 }
},
{
  _id: 'Guam',
  largestCounty: { county: 'Guam', population: 0 },
  smallestCounty: { county: 'Guam', population: 0 }
},
{
  _id: 'Hawaii',
  largestCounty: { county: 'Maui', population: 164765 },
  smallestCounty: { county: 'Hawaii', population: 202163 }
},
{
  _id: 'Idaho',
  largestCounty: { county: 'Washington', population: 10612 },
  smallestCounty: { county: 'Ada', population: 480112 }
},
{
  _id: 'Illinois',
  largestCounty: { county: 'Woodford', population: 36245 },
  smallestCounty: { county: 'Adams', population: 65354 }
},
{
  _id: 'Indiana',
  largestCounty: { county: 'Whitley', population: 36575 },
  smallestCounty: { county: 'Adams', population: 35423 }
},
{
  _id: 'Iowa',
  largestCounty: { county: 'Wright', population: 12857 },
  smallestCounty: { county: 'Adair', population: 7162 }
},

```

```

{
  _id: 'Kansas',
  largestCounty: { county: 'Wyandotte', population: 160176 },
  smallestCounty: { county: 'Allen', population: 12246 }
},
{
  _id: 'Kentucky',
  largestCounty: { county: 'Woodford', population: 28086 },
  smallestCounty: { county: 'Adair', population: 17540 }
}
]

```


e) Get the nearest 10 zips from one of Chicago's landmarks, the Willis Tower situated at coordinates 41.878876, -87.635918.

```
db.proiect.updateMany({}, {
  $set: {
    location: {
      type: "Point",
      coordinates: [{ $toDouble: "$lng" }, { $toDouble: "$lat" }]
    }
  }
});
```

```
us_zips> db.proiect.updateMany({}, {
...   $set: {
...     location: {
...       type: "Point",
...       coordinates: [{ $toDouble: "$lng" }, { $toDouble: "$lat" }]
...     }
...   }
... });
{
  acknowledged: true,
  insertedId: null,
  matchedCount: 33787,
  modifiedCount: 33787,
  upsertedCount: 0
}
us_zips>
```

```
db.proiect.createIndex({ location: "2dsphere" });
```

```
us_zips> db.proiect.createIndex({ location: "2dsphere" });
location_2dsphere
us_zips>
```

```
db.proiect.aggregate([
  {
    $geoNear: {
      near: { type: "Point", coordinates: [-87.635918, 41.878876] },
      distanceField: "distance",
      spherical: true
    }
  },
  {
    $limit: 10
  },
  {
    $project: {
      _id: 0,
      zip: 1,
      distance: 1
    }
  }
]);
```

```

us_zips> db.proiect.aggregate([
...   {
...     $geoNear: {
...       near: { type: "Point", coordinates: [-87.635918, 41.878876]
...     },
...     distanceField: "distance",
...     spherical: true
...   },
...   {
...     $limit: 10
...   },
...   {
...     $project: {
...       _id: 0,
...       zip: 1,
...       distance: 1
...     }
...   }
... ]);
[
  { zip: 60606, distance: 341.33394676341226 },
  { zip: 60604, distance: 558.9535975361117 },
  { zip: 60603, distance: 720.7615244718804 },
  { zip: 60602, distance: 733.440314995619 },
  { zip: 60661, distance: 812.6448854556236 },
  { zip: 60601, distance: 1357.5540104509246 },
  { zip: 60607, distance: 1393.2775295915972 },
  { zip: 60654, distance: 1472.5057139025093 },
  { zip: 60605, distance: 1937.7876476806439 },
  { zip: 60611, distance: 2240.2942493381133 }
]
us_zips>

```

f) Get the total population situated between 50 and 200 kms around New York's landmark, the Statue of Liberty at coordinates 40.689247, -74.044502.

```
us_zips> db.proiect.createIndex({ location: "2dsphere" });
```

```

us_zips> db.proiect.createIndex({ location: "2dsphere" });
location_2dsphere
us_zips>

```

```

db.proiect.aggregate([
  {
    $geoNear: {
      near: { type: "Point", coordinates: [-74.044502, 40.689247] },
      distanceField: "distance",
      spherical: true
    }
  },
  {
    $match: {
      distance: { $gte: 50000, $lte: 200000 } // 50 km to 200 km in meters
    }
  }
])

```

```

    },
    {
      $group: {
        _id: null,
        totalPopulation: { $sum: { $toInt: "$population" } }
      }
    },
    {
      $project: {
        _id: 0,
        totalPopulation: 1
      }
    }
  ]
});
location_usphere
us_zips> db.project.aggregate([
...   {
...     $geoNear: {
...       near: { type: "Point", coordinates: [-74.044502, 40.689247]
...     },
...     distanceField: "distance",
...     spherical: true
...   },
...   {
...     $match: {
...       distance: { $gte: 50000, $lte: 200000 } // 50 km to 200 km
...     }
...   },
...   {
...     $group: {
...       _id: null,
...       totalPopulation: { $sum: { $toInt: "$population" } }
...     }
...   },
...   {
...     $project: {
...       _id: 0,
...       totalPopulation: 1
...     }
...   }
... ]);
[ { totalPopulation: 18089630 } ]
us_zips>

```

Notes:

- Create the indexes you deem relevant for your collection. You will be asked on the performance of your indexes so be prepared to defend your choice, preferably by analyzing the execution statistics.
- For requirements e) and f), you may add a geo field to your collection in order to leverage geospatial query operators.
- Your solution must be original so please don't rely on cheating.

3. Scoring:

- Requirement a) - 1 point
- Requirement b) - 1.5 points
- Requirement c) - 1.5 points
- Requirement d) - 1.5 points
- Requirement e) - 2 points
- Requirement f) - 2.5 points

In order to pass, you must earn 5 points or more.

4. Solution Delivery

Upload your final project solution to the designated area in the virtual campus.

Valer Bocan, PhD, CSSLP