

Proiect ABD

- Import the dataset into MongoDB instance

- `mongoimport --db=abdproject --collection=uszip --type=csv --headerline --uri="mongodb://mongoadmin:secret@localhost:27888/?authSource=admin" --file=C:/Users/WIN10/proiect_abd/uszip.csv --ignoreBlanks`

```
C:\Users\WIN10>mongoimport --db=abdproject --collection=uszip --type=csv --headerline --uri="mongodb://mongoadmin:secret@localhost:27888/?authSource=admin" --file=C:/Users/WIN10/proiect_abd/uszip.csv --ignoreBlanks
2024-05-20T22:03:37.902+0200    connected to: mongodb://[**REDACTED**]@localhost:27888/?authSource=admin
2024-05-20T22:03:39.263+0200    33787 document(s) imported successfully. 0 document(s) failed to import.
```

- Connect to MongoDB instance:

- `mongosh "mongodb://mongoadmin:secret@localhost:27888/?authSource=admin"`

```
C:\Users\WIN10>mongosh "mongodb://mongoadmin:secret@localhost:27888/?authSource=admin"
Current Mongosh Log ID: 664bad7e84fd0ddfed14a0d
Connecting to:  mongodb://<credentials>@localhost:27888/?authSource=admin&directConnection=true&serverSelectionTimeoutMS=2000&appName=mongosh+2.2.1
Using MongoDB:  7.0.6
Using Mongosh:  2.2.1
mongosh 2.2.2 is available for download: https://www.mongodb.com/try/download/shell

For mongosh info see: https://docs.mongodb.com/mongosh-shell/

-----
The server generated these startup warnings when booting
2024-05-27T19:57:29.999+00:00: Using the XFS filesystem is strongly recommended with the WiredTiger storage engine. See http://dochub.mongodb.org/core/production-file-system
2024-05-27T19:57:31.355+00:00: /sys/kernel/mm/transparent_hugepage/enabled is 'always'. We suggest setting it to 'never'
2024-05-27T19:57:31.355+00:00: vm.max_map_count is too low
-----
```

```
test> use abdproject
switched to db abdproject
abdproject> db.uszip.countDocuments({})
33787
abdproject>
```

Requirements:

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

`db.uszip.aggregate([{"$group": {"_id": "$state_id", "state_population": {"$sum": "$population"} }}, {"$match": {"state_population": {"$gt": 10000000 } }])`

```
abdproject> db.uszip.aggregate([{"$group": {"_id": "$state_id", "state_population": {"$sum": "$population"} }}, {"$match": {"state_population": {"$gt": 10000000 } }])
[
  { "_id": "PA", state_population: 12989208 },
  { "_id": "NC", state_population: 10470214 },
  { "_id": "GA", state_population: 10722352 },
  { "_id": "NY", state_population: 19994379 },
  { "_id": "OH", state_population: 11774683 },
  { "_id": "MI", state_population: 10057902 },
  { "_id": "TX", state_population: 29242696 },
  { "_id": "IL", state_population: 12757583 },
  { "_id": "CA", state_population: 39354820 },
  { "_id": "FL", state_population: 21632200 }
]
```

b) Get the average city population by state.

```
db.uszips.aggregate([ { "$group": { "_id": { "state_id": "$state_id", "city": "$city" },  
"population": { "$sum": "$population" } } }, { "$group": { "_id": "$_id.state_id",  
"average_city_population": { "$avg": "$population" } } } ] )
```

```
abdproject> db.uszips.aggregate([ { "$group": { "_id": { "state_id": "$state_id", "city": "$city" }, "population": { "$sum": "$population" } } }, { "$group":  
: { "_id": "$_id.state_id", "average_city_population": { "$avg": "$population" } } } ] )  
[  
  { "_id": "DE", "average_city_population": 18488.962962962964 },  
  { "_id": "CO", "average_city_population": 14835.881748871979 },  
  { "_id": "WV", "average_city_population": 2360.2313933488372 },  
  { "_id": "PA", "average_city_population": 7915.422383473491 },  
  { "_id": "MD", "average_city_population": 15488.547738693467 },  
  { "_id": "GA", "average_city_population": 18112.08108108108 },  
  { "_id": "NC", "average_city_population": 14862.470762889137 },  
  { "_id": "WV", "average_city_population": 2514.178120617111 },  
  { "_id": "IN", "average_city_population": 9789.90318983318 },  
  { "_id": "NY", "average_city_population": 13111.068196721311 },  
  { "_id": "VA", "average_city_population": 11912.862406187846 },  
  { "_id": "KS", "average_city_population": 4712.313881685136 },  
  { "_id": "WA", "average_city_population": 16685.714982887777 },  
  { "_id": "MT", "average_city_population": 3985.864480779661 },  
  { "_id": "NE", "average_city_population": 2239.854592369668 },  
  { "_id": "NH", "average_city_population": 5821.42616837553 },  
  { "_id": "IA", "average_city_population": 3444.1868250539956 },  
  { "_id": "OK", "average_city_population": 6965.724561483589 },  
  { "_id": "AL", "average_city_population": 8468.882239121275 },  
  { "_id": "RI", "average_city_population": 16332.587462686568 }  
]  
Type "it" for more
```

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

```
db.uszips.aggregate([ { "$group": { "_id": { "state_id": "$state_id", "city": "$city" },  
"population": { "$sum": "$population" } } }, { "$sort": { "population": 1 } }, { "$group": {  
"_id": "$_id.state_id", "smallest_city": { "$first": "$_id.city" },  
"population_of_the_smallest_city": { "$first": "$population" }, "largest_city": { "$last":  
"$_id.city" }, "population_of_the_largest_city": { "$last": "$population" } } } ] )
```

```
abdproject> db.uszips.aggregate([ { "$group": { "_id": { "state_id": "$state_id", "city": "$city" }, "population": { "$sum": "$population" } } }, { "$sort": { "population": 1 } }, { "$group": { "_id": "$_id.state_id", "smallest_city":  
: { "$first": "$_id.city" }, "population_of_the_smallest_city": { "$first": "$population" }, "largest_city": { "$last": "$_id.city" }, "population_of_the_largest_city": { "$last": "$population" } } } ] )  
[  
  {  
    _id: 'TN',  
    smallest_city: 'Campaign',  
    population_of_the_smallest_city: 0,  
    largest_city: 'Memphis',  
    population_of_the_largest_city: 680293  
  },  
  {  
    _id: 'UT',  
    smallest_city: 'Bonzan',  
    population_of_the_smallest_city: 0,  
    largest_city: 'Salt Lake City',  
    population_of_the_largest_city: 500818  
  },  
  {  
    _id: 'HI',  
    smallest_city: 'Tripler Army Medical Center',  
    population_of_the_smallest_city: 0,  
    largest_city: 'Honolulu',  
    population_of_the_largest_city: 416884  
  },  
  {  
    _id: 'PR',  
    smallest_city: 'Fort Buchanan',  
    population_of_the_smallest_city: 128,  
    largest_city: 'San Juan',  
    population_of_the_largest_city: 345339  
  },  
  {  
    _id: 'MT',  
    smallest_city: 'Sumatra',  
    population_of_the_smallest_city: 3,  
    largest_city: 'Billings',  
    population_of_the_largest_city: 143280  
  },  
  {  
    _id: 'AK',  
    smallest_city: 'Meyers Chuck',  
    population_of_the_smallest_city: 1,  
    largest_city: 'Anchorage',  
    population_of_the_largest_city: 242198  
  },  
  {  
    _id: 'VT',  
    smallest_city: 'Saint Albans Bay',  
    population_of_the_smallest_city: 0,  
    largest_city: 'Burlington',  
    population_of_the_largest_city: 44661  
  },  
  {  
    _id: 'PA',  
    smallest_city: 'Refton',  
    population_of_the_smallest_city: 0,  
    largest_city: 'Philadelphia',  
    population_of_the_largest_city: 1593611  
  },  
]
```

```
{
  _id: 'GA',
  smallest_city: 'Perkins',
  population_of_the_smallest_city: 0,
  largest_city: 'Atlanta',
  population_of_the_largest_city: 1029695
},
{
  _id: 'MD',
  smallest_city: 'Spring Gap',
  population_of_the_smallest_city: 0,
  largest_city: 'Baltimore',
  population_of_the_largest_city: 598831
},
{
  _id: 'AZ',
  smallest_city: 'Rillito',
  population_of_the_smallest_city: 0,
  largest_city: 'Phoenix',
  population_of_the_largest_city: 1447391
},
{
  _id: 'WV',
  smallest_city: 'Bartley',
  population_of_the_smallest_city: 0,
  largest_city: 'Morgantown',
  population_of_the_largest_city: 97154
},
{
  _id: 'FL',
  smallest_city: 'Mango',
  population_of_the_smallest_city: 0,
  largest_city: 'Miami',
  population_of_the_largest_city: 1855275
},
{
  _id: 'NH',
  smallest_city: 'Bretton Woods',
  population_of_the_smallest_city: 0,
  largest_city: 'Manchester',
  population_of_the_largest_city: 119416
},
{
  _id: 'OR',
  smallest_city: 'Post',
  population_of_the_smallest_city: 0,
  largest_city: 'Portland',
  population_of_the_largest_city: 934313
},
{
  _id: 'WA',
  smallest_city: 'Buena',
  population_of_the_smallest_city: 0,
  largest_city: 'Seattle',
  population_of_the_largest_city: 988217
}
]
Type "it" for more
abdproject> |
```

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

```
db.uszips.aggregate([ { "$group": { "_id": { "state_id": "$state_id", "county_name":  
"$county_name" }, "population": { "$sum": "$population" } } }, { "$sort": { "population": 1 } }, {  
"$group": { "_id": "$_id.state_id", "smallest_county": { "$first": "$_id.county_name" },  
"population_of_the_smallest_county": { "$first": "$population" }, "largest_county": { "$last":  
"$_id.county_name" }, "population_of_the_largest_county": { "$last": "$population" } } } ] )
```

```
abdproject> db.uszips.aggregate([ { "$group": { "_id": { "state_id": "$state_id", "county_name": "$county_name" }, "population": { "$sum": "$population" } } }, { "$sort": { "population": 1 } }, { "$group": { "_id": "$_id.state_id", "smallest_county": { "$first": "$_id.county_name" }, "population_of_the_smallest_county": { "$first": "$population" }, "largest_county": { "$last": "$_id.county_name" }, "population_of_the_largest_county": { "$last": "$population" } } } ] )  
{  
  "_id": "OK",  
  "smallest_county": "Harper",  
  "population_of_the_smallest_county": 1783,  
  "largest_county": "Oklahoma",  
  "population_of_the_largest_county": 825347  
},  
{  
  "_id": "ND",  
  "smallest_county": "Oliver",  
  "population_of_the_smallest_county": 1113,  
  "largest_county": "Cass",  
  "population_of_the_largest_county": 186640  
},  
{  
  "_id": "WV",  
  "smallest_county": "Mallory",  
  "population_of_the_smallest_county": 363,  
  "largest_county": "Clarksburg",  
  "population_of_the_largest_county": 2265930  
},  
{  
  "_id": "NC",  
  "smallest_county": "Tyrrell",  
  "population_of_the_smallest_county": 3117,  
  "largest_county": "Mecklenburg",  
  "population_of_the_largest_county": 1131993  
},  
{  
  "_id": "MN",  
  "smallest_county": "Roberts",  
  "population_of_the_smallest_county": 981,  
  "largest_county": "Hennepin",  
  "population_of_the_largest_county": 1279974  
},  
{  
  "_id": "GA",  
  "smallest_county": "Webster",  
  "population_of_the_smallest_county": 1363,  
  "largest_county": "Fulton",  
  "population_of_the_largest_county": 1130551  
},  
{  
  "_id": "AZ",  
  "smallest_county": "McKinley",  
  "population_of_the_smallest_county": 2883,  
  "largest_county": "Maricopa",  
  "population_of_the_largest_county": 1293421  
}
```

->Adding a geo field to collection

```
db.uszips.updateMany( {}, [ { "$set": { "location2dsphere": { "type": "Point", "coordinates": [ "$lng",  
"$lat" ] } } } ] )
```

```
abdproject> db.uszips.updateMany( {}, [ { "$set": { "location2dsphere": { "type": "Point", "coordinates": [ "$lng", "$lat" ] } } } ] )  
{  
  acknowledged: true,  
  insertedId: null,  
  matchedCount: 33787,  
  modifiedCount: 0,  
  upsertedCount: 0  
}
```

-> Creating index for geo field:

```
db.uszips.createIndex( { location2dsphere : "2dsphere" } )
```

```
abdproject> db.uszips.createIndex( { location2dsphere : "2dsphere" } )  
location2dsphere_2dsphere
```

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

```
db.uszips.aggregate([ { $geoNear: { near: { type: "Point", coordinates: [-87.635918, 41.878876] }, key: "location2dsphere", distanceField: "dist.calculated" } }, { $limit: 10 } ] )
```

```

dbproj> db.uszip.aggregate([ { $geoNear: { near: { type: 'Point', coordinates: [-87.635916, 41.878976] }, key: 'location2dsphere', distanceField: '$dist.calculated' } }, { $limit: 10 } ] )
[
  {
    _id: ObjectId('665980181288ccla1a8ffa8'),
    zip: 60606,
    lat: 41.88178,
    lng: -87.63724,
    city: 'Chicago',
    state_id: 'IL',
    state_name: 'Illinois',
    zcta: 'TRUE',
    population: 3513,
    density: 5044.8,
    county_fips: 17031,
    county_name: 'Cook',
    county_weights: {'17031': 100},
    county_name_all: 'Cook',
    county_fips_all: 17031,
    imprecise: 'FALSE',
    military: 'FALSE',
    timezone: 'America/Chicago',
    location2dsphere: { type: 'Point', coordinates: [ -87.63724, 41.88178 ] },
    dist: { calculated: 241.33394676341226 }
  },
  {
    _id: ObjectId('665980181288ccla1a8ffa9'),
    zip: 60604,
    lat: 41.87799,
    lng: -87.62928,
    city: 'Chicago',
    state_id: 'IL',
    state_name: 'Illinois',
    zcta: 'TRUE',
    population: 737,
    density: 3684.3,
    county_fips: 17031,
    county_name: 'Cook',
    county_weights: {'17031': 100},
    county_name_all: 'Cook',
    county_fips_all: 17031,
    imprecise: 'FALSE',
    military: 'FALSE',
    timezone: 'America/Chicago',
    location2dsphere: { type: 'Point', coordinates: [ -87.62928, 41.87799 ] },
    dist: { calculated: 558.9535973361117 }
  },
  {
    _id: ObjectId('665980181288ccla1a8ffa3'),
    zip: 60603,
    lat: 41.88018,
    lng: -87.6274,
    city: 'Chicago',
    state_id: 'IL',
    state_name: 'Illinois',
    zcta: 'TRUE',
    population: 1318,
    density: 3612.6,
    county_fips: 17031,
    county_name: 'Cook',
    county_weights: {'17031': 100},
    county_name_all: 'Cook',
    county_fips_all: 17031,
    imprecise: 'FALSE',
    military: 'FALSE',
    timezone: 'America/Chicago',
    location2dsphere: { type: 'Point', coordinates: [ -87.6274, 41.88018 ] },
    dist: { calculated: 726.7615244718894 }
  },
]

```

```
{
  _id: ObjectId('665980181288cc1aala0ff9d'),
  zip: 60602,
  lat: 41.88304,
  lng: -87.62906,
  city: 'Chicago',
  state_id: 'IL',
  state_name: 'Illinois',
  zcta: 'TRUE',
  population: 938,
  density: 4394.9,
  county_fips: 17031,
  county_name: 'Cook',
  county_weights: '{"17031": 100}',
  county_names_all: 'Cook',
  county_fips_all: '17031',
  imprecise: 'FALSE',
  military: 'FALSE',
  timezone: 'America/Chicago',
  location2dsphere: { type: 'Point', coordinates: [ -87.62906, 41.88304 ] },
  dist: { calculated: 733.4403149960813 }
},
{
  _id: ObjectId('665980181288cc1aala0ff9d'),
  zip: 60601,
  lat: 41.88291,
  lng: -87.64409,
  city: 'Chicago',
  state_id: 'IL',
  state_name: 'Illinois',
  zcta: 'TRUE',
  population: 11513,
  density: 14946.6,
  county_fips: 17031,
  county_name: 'Cook',
  county_weights: '{"17031": 100}',
  county_names_all: 'Cook',
  county_fips_all: '17031',
  imprecise: 'FALSE',
  military: 'FALSE',
  timezone: 'America/Chicago',
  location2dsphere: { type: 'Point', coordinates: [ -87.64409, 41.88291 ] },
  dist: { calculated: 812.6448854594362 }
},
{
  _id: ObjectId('665980181288cc1aala0ff9c'),
  zip: 60601,
  lat: 41.88527,
  lng: -87.62197,
  city: 'Chicago',
  state_id: 'IL',
  state_name: 'Illinois',
  zcta: 'TRUE',
  population: 16398,
  density: 16290.5,
  county_fips: 17031,
  county_name: 'Cook',
  county_weights: '{"17031": 100}',
  county_names_all: 'Cook',
  county_fips_all: '17031',
  imprecise: 'FALSE',
  military: 'FALSE',
  timezone: 'America/Chicago',
  location2dsphere: { type: 'Point', coordinates: [ -87.62197, 41.88527 ] },
  dist: { calculated: 1397.5540104589246 }
}
```

```
{
  _id: ObjectId('665980181288cc1aala0ffa8'),
  zip: 60607,
  lat: 41.87467,
  lng: -87.65175,
  city: 'Chicago',
  state_id: 'IL',
  state_name: 'Illinois',
  zcta: 'TRUE',
  population: 28797,
  density: 4806.3,
  county_fips: 17031,
  county_name: 'Cook',
  county_weights: '{"17031": 100}',
  county_names_all: 'Cook',
  county_fips_all: 17031,
  imprecise: 'FALSE',
  military: 'FALSE',
  timezone: 'America/Chicago',
  location2dsphere: { type: 'Point', coordinates: [ -87.65175, 41.87467 ] },
  dist: { calculated: 1393.2775295915972 }
},
{
  _id: ObjectId('665980181288cc1aala0ffd6'),
  zip: 60654,
  lat: 41.89209,
  lng: -87.63673,
  city: 'Chicago',
  state_id: 'IL',
  state_name: 'Illinois',
  zcta: 'TRUE',
  population: 23475,
  density: 16729.1,
  county_fips: 17031,
  county_name: 'Cook',
  county_weights: '{"17031": 100}',
  county_names_all: 'Cook',
  county_fips_all: 17031,
  imprecise: 'FALSE',
  military: 'FALSE',
  timezone: 'America/Chicago',
  location2dsphere: { type: 'Point', coordinates: [ -87.63673, 41.89209 ] },
  dist: { calculated: 1472.5057139025093 }
},
{
  _id: ObjectId('665980181288cc1aala0ffa6'),
  zip: 60605,
  lat: 41.86644,
  lng: -87.61956,
  city: 'Chicago',
  state_id: 'IL',
  state_name: 'Illinois',
  zcta: 'TRUE',
  population: 34314,
  density: 9262.1,
  county_fips: 17031,
  county_name: 'Cook',
  county_weights: '{"17031": 100}',
  county_names_all: 'Cook',
  county_fips_all: 17031,
  imprecise: 'FALSE',
  military: 'FALSE',
  timezone: 'America/Chicago',
  location2dsphere: { type: 'Point', coordinates: [ -87.61956, 41.86644 ] },
  dist: { calculated: 1937.7876476801412 }
},
{
  _id: ObjectId('665980181288cc1aala0ffad'),
  zip: 60611,
  lat: 41.89483,
  lng: -87.61944,
  city: 'Chicago',
  state_id: 'IL',
  state_name: 'Illinois',
  zcta: 'TRUE',
  population: 35247,
  density: 17163,
  county_fips: 17031,
  county_name: 'Cook',
  county_weights: '{"17031": 100}',
  county_names_all: 'Cook',
  county_fips_all: 17031,
  imprecise: 'FALSE',
  military: 'FALSE',
  timezone: 'America/Chicago',
  location2dsphere: { type: 'Point', coordinates: [ -87.61944, 41.89483 ] },
  dist: { calculated: 2240.2942493381133 }
}
]
```

- 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.**

```
db.uszip.aggregate([ { "$match": { "location2dsphere": { "$geoWithin": { "$centerSphere":  
[[-74.044502, 40.689247], 200 / 6378.1] } } } }, { "$match": { "location2dsphere": { "$not": {  
"$geoWithin": { "$centerSphere": [[-74.044502, 40.689247], 50 / 6378.1] } } } } }, { "$group":  
{ "_id": "result", "count": { "$sum": "$population" } } } ] )
```

```
abdproject> db.uszip.aggregate([ { "$match": { "location2dsphere": { "$geoWithin": { "$centerSphere": [[-74.044502, 40.689247], 200 / 6378.1] } } } }, { "$match": { "location2dsphere": {  
"$not": { "$geoWithin": { "$centerSphere": [[-74.044502, 40.689247], 50 / 6378.1] } } } } }, { "$group": { "_id": "result", "count": { "$sum": "$population" } } } ] ).pretty();  
{ "_id": "result", count: 18089630 }
```