**MongoDB -Aggregation Exercises**



Import the zips.json file into your MongoDB. Database name is "population" and collection name is "zipcodes".

mongoimport --db population --collection zipcodes --file zips.json

**Atlanta Population**



1. use db.zipcodes.find() to filter results to only the results where city is ATLANTA and state is GA.

Atlas atlas-pjnto1-shard-0 [primary] Population> db.zipcodes.find({$and:[{city:'ATLANTA'},{state: 'GA'}]})

[

{

\_id: '30303',

city: 'ATLANTA',

loc: [ -84.388846, 33.752504 ],

pop: 1845,

state: 'GA'

},

{

\_id: '30305',

city: 'ATLANTA',

loc: [ -84.385145, 33.831963 ],

pop: 19122,

state: 'GA'

},

{

\_id: '30306',

city: 'ATLANTA',

loc: [ -84.351418, 33.786027 ],

pop: 20081,

state: 'GA'

}

.

.

.

]

Type "it" for more.

1. use db.zipcodes.aggregate with $match to do the same as above.

Atlas atlas-pjnto1-shard-0 [primary] Population> db.zipcodes.aggregate([{$match:{$and:[{city:'ATLANTA'},{state:'GA'}]}}])

[

{

\_id: '30303',

city: 'ATLANTA',

loc: [ -84.388846, 33.752504 ],

pop: 1845,

state: 'GA'

},

{

\_id: '30305',

city: 'ATLANTA',

loc: [ -84.385145, 33.831963 ],

pop: 19122,

state: 'GA'

},

{

\_id: '30306',

city: 'ATLANTA',

loc: [ -84.351418, 33.786027 ],

pop: 20081,

state: 'GA'

}

.

.

.

]

Type "it" for more.

1. use $group to count the number of zip codes in Atlanta.

Atlas atlas-pjnto1-shard-0 [primary] Population> db.zipcodes.aggregate([{$group:{\_id: '$city'}},{$count:"ATLANTA"}])

[ { ATLANTA: 16584 } ]

1. use $group to find the total population in Atlanta.

Atlas atlas-pjnto1-shard-0 [primary] Population> db.zipcodes.aggregate([{$group:{\_id:'$city', totalpop: {$sum:'$pop'}}},{$match:{\_id:"ATLANTA"}}])

[ { \_id: 'ATLANTA', totalpop: 630046 } ]

**Populations By State**



1. use aggregate to calculate the total population for each state

Atlas atlas-pjnto1-shard-0 [primary] Population> db.zipcodes.aggregate([{$group:{\_id:"$state", totalpop:{ $sum: "$pop"}}}])

[

{ \_id: 'SD', totalpop: 695397 },

{ \_id: 'NM', totalpop: 1515069 },

{ \_id: 'VA', totalpop: 6181479 },

{ \_id: 'MS', totalpop: 2573216 },

{ \_id: 'NC', totalpop: 6628637 },

{ \_id: 'TX', totalpop: 16984601 },

{ \_id: 'LA', totalpop: 4217595 },

{ \_id: 'VT', totalpop: 562758 },

{ \_id: 'IA', totalpop: 2776420 },

{ \_id: 'NH', totalpop: 1109252 },

{ \_id: 'FL', totalpop: 12686644 },

{ \_id: 'IL', totalpop: 11427576 },

{ \_id: 'MN', totalpop: 4372982 },

{ \_id: 'AR', totalpop: 2350725 },

{ \_id: 'ID', totalpop: 1006749 },

{ \_id: 'OH', totalpop: 10846517 },

{ \_id: 'MA', totalpop: 6016425 },

{ \_id: 'PA', totalpop: 11881643 },

{ \_id: 'SC', totalpop: 3486703 },

{ \_id: 'NY', totalpop: 17990402 }

]

Type "it" for more

1. sort the results by population, highest first

db.zipcodes.aggregate([{$group:{\_id:"$state", totalpop:{ $sum: "$pop"}}},{$sort: {totalpop:-1}}])

[

{ \_id: 'CA', totalpop: 29754890 },

{ \_id: 'NY', totalpop: 17990402 },

{ \_id: 'TX', totalpop: 16984601 },

{ \_id: 'FL', totalpop: 12686644 },

{ \_id: 'PA', totalpop: 11881643 },

{ \_id: 'IL', totalpop: 11427576 },

{ \_id: 'OH', totalpop: 10846517 },

{ \_id: 'MI', totalpop: 9295297 },

{ \_id: 'NJ', totalpop: 7730188 },

{ \_id: 'NC', totalpop: 6628637 },

{ \_id: 'GA', totalpop: 6478216 },

{ \_id: 'VA', totalpop: 6181479 },

{ \_id: 'MA', totalpop: 6016425 },

{ \_id: 'IN', totalpop: 5544136 },

{ \_id: 'MO', totalpop: 5110648 },

{ \_id: 'WI', totalpop: 4891769 },

{ \_id: 'TN', totalpop: 4876457 },

{ \_id: 'WA', totalpop: 4866692 },

{ \_id: 'MD', totalpop: 4781379 },

{ \_id: 'MN', totalpop: 4372982 }

]

Type "it" for more

1. limit the results to just the first 3 results. What are the top 3 states in population?

db.zipcodes.aggregate([{$group:{\_id:"$state", totalpop:{ $sum: "$pop"}}},{$sort: {totalpop:-1}},{$limit:3}])

[

{ \_id: 'CA', totalpop: 29754890 },

{ \_id: 'NY', totalpop: 17990402 },

{ \_id: 'TX', totalpop: 16984601 }

]

**Populations by City**



1. use aggregate to calculate the total population for each city (you have to use city/state combination). You can use a combination for the \_id of the $group: { city: '$city', state: '$state' }

Atlas atlas-pjnto1-shard-0 [primary] Population> db.zipcodes.aggregate([{$group: {\_id:{state:"$state",city:"$city"},

... pop:{$sum:"$pop"}}}])

[

{ \_id: { state: 'CA', city: 'ONYX' }, pop: 380 },

{ \_id: { state: 'IL', city: 'MEDORA' }, pop: 531 },

{ \_id: { state: 'AR', city: 'DARDANELLE' }, pop: 8281 },

{ \_id: { state: 'ME', city: 'FRANKLIN' }, pop: 1433 },

{ \_id: { state: 'ND', city: 'BERWICK' }, pop: 1251 },

{ \_id: { state: 'PA', city: 'THOMASVILLE' }, pop: 3435 },

{ \_id: { state: 'NY', city: 'OGDENSBURG' }, pop: 19659 },

{ \_id: { state: 'OH', city: 'MAUMEE' }, pop: 22993 },

{ \_id: { state: 'GA', city: 'BROOKLET' }, pop: 4850 },

{ \_id: { state: 'CA', city: 'CEDARVILLE' }, pop: 991 },

{ \_id: { state: 'WV', city: 'ULER' }, pop: 724 },

{ \_id: { state: 'NC', city: 'DUDLEY' }, pop: 8450 },

{ \_id: { state: 'IN', city: 'MADISON' }, pop: 20596 },

{ \_id: { state: 'MN', city: 'GOODVIEW' }, pop: 34518 },

{ \_id: { state: 'OR', city: 'KENO' }, pop: 2696 },

{ \_id: { state: 'OR', city: 'MEDICAL SPRINGS' }, pop: 10024 },

{ \_id: { state: 'IL', city: 'ROYAL LAKES' }, pop: 844 },

{ \_id: { state: 'ND', city: 'CUMMINGS' }, pop: 286 },

{ \_id: { state: 'AL', city: 'MARION' }, pop: 7284 },

{ \_id: { state: 'PA', city: 'MINERAL POINT' }, pop: 420 }

]

Type "it" for more

1. sort the results by population, highest first

Atlas atlas-pjnto1-shard-0 [primary] Population> db.zipcodes.aggregate([{ $group: { \_id: { state: "$state", city: "$city" },

... pop: { $sum: "$pop" } } }, { $sort:{pop:-1} }])

[

{ \_id: { state: 'IL', city: 'CHICAGO' }, pop: 2452177 },

{ \_id: { state: 'NY', city: 'BROOKLYN' }, pop: 2300504 },

{ \_id: { state: 'CA', city: 'LOS ANGELES' }, pop: 2102295 },

{ \_id: { state: 'TX', city: 'HOUSTON' }, pop: 2095918 },

{ \_id: { state: 'PA', city: 'PHILADELPHIA' }, pop: 1610956 },

{ \_id: { state: 'NY', city: 'NEW YORK' }, pop: 1476790 },

{ \_id: { state: 'NY', city: 'BRONX' }, pop: 1209548 },

{ \_id: { state: 'CA', city: 'SAN DIEGO' }, pop: 1049298 },

{ \_id: { state: 'MI', city: 'DETROIT' }, pop: 963243 },

{ \_id: { state: 'TX', city: 'DALLAS' }, pop: 940191 },

{ \_id: { state: 'AZ', city: 'PHOENIX' }, pop: 890853 },

{ \_id: { state: 'FL', city: 'MIAMI' }, pop: 825232 },

{ \_id: { state: 'CA', city: 'SAN JOSE' }, pop: 816653 },

{ \_id: { state: 'TX', city: 'SAN ANTONIO' }, pop: 811792 },

{ \_id: { state: 'MD', city: 'BALTIMORE' }, pop: 733081 },

{ \_id: { state: 'CA', city: 'SAN FRANCISCO' }, pop: 723993 },

{ \_id: { state: 'TN', city: 'MEMPHIS' }, pop: 632837 },

{ \_id: { state: 'CA', city: 'SACRAMENTO' }, pop: 628279 },

{ \_id: { state: 'FL', city: 'JACKSONVILLE' }, pop: 610160 },

{ \_id: { state: 'GA', city: 'ATLANTA' }, pop: 609591 }

]

Type "it" for more

1. limit the results to just the first 3 results. What are the top 3 cities in population?

Atlas atlas-pjnto1-shard-0 [primary] Population> db.zipcodes.aggregate([{ $group: { \_id: { state: "$state", city: "$city" },

... pop: { $sum: "$pop" } } }, { $sort:{pop:-1} },{$limit:3}])

[

{ \_id: { state: 'IL', city: 'CHICAGO' }, pop: 2452177 },

{ \_id: { state: 'NY', city: 'BROOKLYN' }, pop: 2300504 },

{ \_id: { state: 'CA', city: 'LOS ANGELES' }, pop: 2102295 }

]

1. What are the top 3 cities in population in Texas?

**Bonus**



1. Write a query to get the average city population for each state.

Atlas atlas-pjnto1-shard-0 [primary] Population> db.zipcodes.aggregate( [

... { $group: { \_id: { state: "$state", city: "$city" }, pop: { $sum: "$pop" } } },

... { $group: { \_id: "$\_id.state", avgCityPop: { $avg: "$pop" } } }

... ] )

[

{ \_id: 'NV', avgCityPop: 18209.590909090908 },

{ \_id: 'KS', avgCityPop: 3819.884259259259 },

{ \_id: 'VT', avgCityPop: 2315.8765432098767 },

{ \_id: 'WI', avgCityPop: 7323.00748502994 },

{ \_id: 'TN', avgCityPop: 9656.350495049504 },

{ \_id: 'KY', avgCityPop: 4767.164721141375 },

{ \_id: 'AZ', avgCityPop: 20591.16853932584 },

{ \_id: 'UT', avgCityPop: 9518.508287292818 },

{ \_id: 'DE', avgCityPop: 14481.91304347826 },

{ \_id: 'DC', avgCityPop: 303450 },

{ \_id: 'SD', avgCityPop: 1839.6746031746031 },

{ \_id: 'IN', avgCityPop: 9271.130434782608 },

{ \_id: 'MI', avgCityPop: 12087.512353706112 },

{ \_id: 'OK', avgCityPop: 6155.743639921722 },

{ \_id: 'MD', avgCityPop: 12615.775725593667 },

{ \_id: 'CT', avgCityPop: 14674.625 },

{ \_id: 'RI', avgCityPop: 19292.653846153848 },

{ \_id: 'WA', avgCityPop: 12258.670025188916 },

{ \_id: 'MT', avgCityPop: 2593.987012987013 },

{ \_id: 'AK', avgCityPop: 2976.4918032786886 }

]

Type "it" for more

1. What are the top 3 states in terms of average city population?

Atlas atlas-pjnto1-shard-0 [primary] Population> db.zipcodes.aggregate([

... { $group: { \_id: { state: "$state", city: "$city" }, pop: { $sum: "$pop" } } },

... { $group: { \_id: "$\_id.state", avgCityPop: { $avg: "$pop" } } },

... {$sort:{avgCityPop:-1}}, {$limit:3}])

[

{ \_id: 'DC', avgCityPop: 303450 },

{ \_id: 'CA', avgCityPop: 27756.42723880597 },

{ \_id: 'FL', avgCityPop: 27400.958963282937 }

]