MongoDB -Aggregation Exercises:

Atlanta Population:

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

> > db.zipcodes.find({city:"ATLANTA",state:"GA"}).pretty()

{

"\_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"

}

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

> db.zipcodes.aggregate([{$match:{$and:[{city:"ATLANTA"},{state:"GA"}]}}]).pretty()

{

"\_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"

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

db.zipcodes.aggregate([{$match:{"city":"ATLANTA"}},{$group:{\_id:"$city",count:{$sum:1}}}])

{ "\_id" : "ATLANTA", "count" : 41 }

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

db.zipcodes.aggregate([{$match:{"city":"ATLANTA"}},{$group:{\_id:"$city",total:{$sum:"$pop"}}}])

{ "\_id" : "ATLANTA", "total population" : 630046 }

>

Populations By State:

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

> db.zipcodes.aggregate([ {$group:{\_id:"$state",Totalpopulation:{$sum:"$pop"}}} ]).pretty()

{ "\_id" : "VA", "Totalpopulation" : 6181479 }

{ "\_id" : "DC", "Totalpopulation" : 606900 }

{ "\_id" : "SD", "Totalpopulation" : 695397 }

{ "\_id" : "OR", "Totalpopulation" : 2842321 }

{ "\_id" : "AZ", "Totalpopulation" : 3665228 }

{ "\_id" : "MA", "Totalpopulation" : 6016425 }

1. sort the results by population, highest first

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

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

{ "\_id" : "CA", "total" : 29754890 }

{ "\_id" : "NY", "total" : 17990402 }

{ "\_id" : "TX", "total" : 16984601 }

{ "\_id" : "FL", "total" : 12686644 }

{ "\_id" : "PA", "total" : 11881643 }

{ "\_id" : "IL", "total" : 11427576 }

{ "\_id" : "OH", "total" : 10846517 }

{ "\_id" : "MI", "total" : 9295297 }

{ "\_id" : "NJ", "total" : 7730188 }

{ "\_id" : "NC", "total" : 6628637 }

{ "\_id" : "GA", "total" : 6478216 }

{ "\_id" : "VA", "total" : 6181479 }…………..

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

> db.zipcodes.aggregate([{$group:{\_id:"$state",total:{$sum

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

{ "\_id" : "CA", "total" : 29754890 }

{ "\_id" : "NY", "total" : 17990402 }

{ "\_id" : "TX", "total" : 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' }

>

> db.zipcodes.aggregate([{$group:{\_id:{city:"$city",state:"$state"},total:{$sum:"$pop"}}}])

{ "\_id" : { "city" : "BROWNDALE", "state" : "PA" }, "total" : 3728 }

{ "\_id" : { "city" : "MC INTIRE", "state" : "IA" }, "total" : 382 }

{ "\_id" : { "city" : "THORNTOWN", "state" : "IN" }, "total" : 4159 }

{ "\_id" : { "city" : "OXFORD", "state" : "IN" }, "total" : 1641 }

{ "\_id" : { "city" : "ALLENWOOD", "state" : "PA" }, "total" : 2666 }

1. sort the results by population, highest first

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

{ "\_id" : { "city" : "CHICAGO", "state" : "IL" }, "total" : 2452177 }

{ "\_id" : { "city" : "BROOKLYN", "state" : "NY" }, "total" : 2300504 }

{ "\_id" : { "city" : "LOS ANGELES", "state" : "CA" }, "total" : 2102295 }

{ "\_id" : { "city" : "HOUSTON", "state" : "TX" }, "total" : 2095918 }

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

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

{ "\_id" : { "city" : "CHICAGO", "state" : "IL" }, "total" : 2452177 }

{ "\_id" : { "city" : "BROOKLYN", "state" : "NY" }, "total" : 2300504 }

{ "\_id" : { "city" : "LOS ANGELES", "state" : "CA" }, "total" : 2102295 }

4.

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

{ "\_id" : { "city" : "HOUSTON", "state" : "TX" }, "total" : 2095918 }

{ "\_id" : { "city" : "DALLAS", "state" : "TX" }, "total" : 940191 }

{ "\_id" : { "city" : "SAN ANTONIO", "state" : "TX" }, "total" : 811792 }

Bonus

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

> db.zipcodes.aggregate([{$group:{\_id:"$city",total:{$avg:"$pop"}}},{$sort:{total:-1}}])

{ "\_id" : "BELL GARDENS", "total" : 99568 }

{ "\_id" : "ARLETA", "total" : 88114 }

{ "\_id" : "SOUTH GATE", "total" : 87026 }

{ "\_id" : "WESTLAND", "total" : 84712 }

{ "\_id" : "HOLLY PARK", "total" : 78511 }

{ "\_id" : "COAST GUARD ISLA", "total" : 76110 }

{ "\_id" : "GWYNN OAK", "total" : 76002 }

{ "\_id" : "GLEN BURNIE", "total" : 75692 }

{ "\_id" : "RIALTO", "total" : 75341 }

{ "\_id" : "JACKSON HEIGHTS", "total" : 72983.5 }

{ "\_id" : "HUNTINGTON PARK", "total" : 72139 }

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

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

{ "\_id" : "DC", "total" : 25287.5 }

{ "\_id" : "CA", "total" : 19627.236147757256 }

{ "\_id" : "FL", "total" : 15779.407960199005 }