**MongoDB -Aggregation** by K Jyothi Gangarsha Id:259921

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

C:\Users\HARSHA>mongoimport --db population --collection zipcodes --file F:\zip.json

**Atlanta Population**

1. **use db.zipcodes.find() to filter results to only the results where city is ATLANTA and state is GA. (limited to 6 to avoid clumsiness here)**

**> db.zipcodes.find({city:"ATLANTA",state:"GA"}).pretty().limit(6)**

{

"\_id" : "30306",

"city" : "ATLANTA",

"loc" : [

-84.351418,

33.786027

],

"pop" : 20081,

"state" : "GA"

}

{

"\_id" : "30305",

"city" : "ATLANTA",

"loc" : [

-84.385145,

33.831963

],

"pop" : 19122,

"state" : "GA"

}

{

"\_id" : "30307",

"city" : "ATLANTA",

"loc" : [

-84.335957,

33.769138

],

"pop" : 16330,

"state" : "GA"

}

{

"\_id" : "30308",

"city" : "ATLANTA",

"loc" : [

-84.375744,

33.771839

],

"pop" : 8549,

"state" : "GA"

}

{

"\_id" : "30309",

"city" : "ATLANTA",

"loc" : [

-84.388338,

33.798407

],

"pop" : 14766,

"state" : "GA"

}

{

"\_id" : "30310",

"city" : "ATLANTA",

"loc" : [

-84.423173,

33.727849

],

"pop" : 34017,

"state" : "GA"

}

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

**>**db.zipcodes.aggregate([{$match:{$and:[{city:"ATLANTA"},{state:"GA"}]}},{$limit:6}])

{ "\_id" : "30306", "city" : "ATLANTA", "loc" : [ -84.351418, 33.786027 ], "pop" : 20081, "state" : "GA" }

{ "\_id" : "30305", "city" : "ATLANTA", "loc" : [ -84.385145, 33.831963 ], "pop" : 19122, "state" : "GA" }

{ "\_id" : "30307", "city" : "ATLANTA", "loc" : [ -84.335957, 33.769138 ], "pop" : 16330, "state" : "GA" }

{ "\_id" : "30308", "city" : "ATLANTA", "loc" : [ -84.375744, 33.771839 ], "pop" : 8549, "state" : "GA" }

{ "\_id" : "30309", "city" : "ATLANTA", "loc" : [ -84.388338, 33.798407 ], "pop" : 14766, "state" : "GA" }

{ "\_id" : "30310", "city" : "ATLANTA", "loc" : [ -84.423173, 33.727849 ], "pop" : 34017, "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:"$city",totalpopulation:{$sum:"$pop"}}} ])

{ "\_id" : "ATLANTA", "totalpopulation" : 630046 }

**Populations By State**

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

**> db.zipcodes.aggregate([ {$group:{\_id:"$state",population:{$sum:"$pop"}}},{$limit:15} ])**

{ "\_id" : "WI", "population" : 4891769 }

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

{ "\_id" : "NH", "population" : 1109252 }

{ "\_id" : "MN", "population" : 4372982 }

{ "\_id" : "WA", "population" : 4866692 }

{ "\_id" : "AK", "population" : 544698 }

{ "\_id" : "IN", "population" : 5544136 }

{ "\_id" : "ME", "population" : 1226648 }

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

{ "\_id" : "RI", "population" : 1003218 }

{ "\_id" : "CT", "population" : 3287116 }

{ "\_id" : "MT", "population" : 798948 }

{ "\_id" : "MO", "population" : 5110648 }

{ "\_id" : "NE", "population" : 1578139 }

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

1. **sort the results by population, highest first**

**> db.zipcodes.aggregate([ {$group:{\_id:"$state",population:{$sum:"$pop"}}},{$sort:{population:-1}},{$limit:15} ])**

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

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

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

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

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

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

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

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

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

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

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

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

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

{ "\_id" : "IN", "population" : 5544136 }

{ "\_id" : "MO", "population" : 5110648 }

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

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

{ "\_id" : { "state" : "CA" }, "Totalpopulation" : 29754890 }

{ "\_id" : { "state" : "NY" }, "Totalpopulation" : 17990402 }

{ "\_id" : { "state" : "TX" }, "Totalpopulation" : 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:"$state"},Population:{$sum:"$pop"}}}])**

{ "\_id" : { "city" : "FARINA", "state" : "IL" }, "Population" : 925 }

{ "\_id" : { "city" : "PIKE ROAD", "state" : "AL" }, "Population" : 3398 }

{ "\_id" : { "city" : "SOMERSET", "state" : "WI" }, "Population" : 2937 }

{ "\_id" : { "city" : "PINEHURST", "state" : "TX" }, "Population" : 2498 }

{ "\_id" : { "city" : "TEXLINE", "state" : "TX" }, "Population" : 672 }

{ "\_id" : { "city" : "LITCHFIELD", "state" : "NE" }, "Population" : 629 }

{ "\_id" : { "city" : "WORDEN", "state" : "IL" }, "Population" : 1818 }

{ "\_id" : { "city" : "DENTON", "state" : "GA" }, "Population" : 1457 }

{ "\_id" : { "city" : "MUSKOGEE", "state" : "OK" }, "Population" : 46600 }

{ "\_id" : { "city" : "KELSO", "state" : "CA" }, "Population" : 30 }

{ "\_id" : { "city" : "DORCHESTER", "state" : "IA" }, "Population" : 757 }

{ "\_id" : { "city" : "BRADFORDSVILLE", "state" : "KY" }, "Population" : 1136 }

{ "\_id" : { "city" : "ALPINE", "state" : "NY" }, "Population" : 1098 }

{ "\_id" : { "city" : "CLINTON", "state" : "MD" }, "Population" : 26417 }

{ "\_id" : { "city" : "PAWLEYS ISLAND", "state" : "SC" }, "Population" : 5445 }

{ "\_id" : { "city" : "NEW SALEM", "state" : "PA" }, "Population" : 4509 }

{ "\_id" : { "city" : "OSWEGATCHIE", "state" : "NY" }, "Population" : 287 }

{ "\_id" : { "city" : "FRANKLIN FURNACE", "state" : "OH" }, "Population" : 3956 }

{ "\_id" : { "city" : "SARDIS", "state" : "AL" }, "Population" : 1683 }

{ "\_id" : { "city" : "ROMULUS", "state" : "MI" }, "Population" : 23471 }

Type "it" for more

> it

{ "\_id" : { "city" : "COURTLAND", "state" : "MN" }, "Population" : 1122 }

{ "\_id" : { "city" : "WOODMERE", "state" : "NY" }, "Population" : 13084 }

{ "\_id" : { "city" : "ROSEBUD", "state" : "MT" }, "Population" : 3867 }

{ "\_id" : { "city" : "RUSSELLVILLE", "state" : "MO" }, "Population" : 2043 }

{ "\_id" : { "city" : "CASTLE HILLS", "state" : "TX" }, "Population" : 36060 }

{ "\_id" : { "city" : "MAPLECREST", "state" : "NY" }, "Population" : 342 }

{ "\_id" : { "city" : "PLINY", "state" : "WV" }, "Population" : 352 }

{ "\_id" : { "city" : "SAINT LANDRY", "state" : "LA" }, "Population" : 5679 }

{ "\_id" : { "city" : "FARMERSVILLE", "state" : "TX" }, "Population" : 4777 }

{ "\_id" : { "city" : "TROY", "state" : "WV" }, "Population" : 115 }

{ "\_id" : { "city" : "PIGEON FORGE", "state" : "TN" }, "Population" : 3063 }

{ "\_id" : { "city" : "LOVELADY", "state" : "TX" }, "Population" : 3649 }

{ "\_id" : { "city" : "LOCHBUI", "state" : "CO" }, "Population" : 20533 }

{ "\_id" : { "city" : "ALHAMBRA", "state" : "IL" }, "Population" : 1543 }

{ "\_id" : { "city" : "PORTLAND", "state" : "TN" }, "Population" : 12790 }

{ "\_id" : { "city" : "ALEXANDRIA", "state" : "MO" }, "Population" : 753 }

{ "\_id" : { "city" : "BENSON", "state" : "AZ" }, "Population" : 6141 }

{ "\_id" : { "city" : "MOUNTAIN MESA", "state" : "CA" }, "Population" : 6285 }

{ "\_id" : { "city" : "MILPITAS", "state" : "CA" }, "Population" : 50907 }

{ "\_id" : { "city" : "POINT HOPE", "state" : "AK" }, "Population" : 640 }

Type "it" for more

1. **sort the results by population, highest first**

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

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

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

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

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

{ "\_id" : { "city" : "PHILADELPHIA", "state" : "PA" }, "Population" : 1610956 }

{ "\_id" : { "city" : "NEW YORK", "state" : "NY" }, "Population" : 1476790 }

{ "\_id" : { "city" : "BRONX", "state" : "NY" }, "Population" : 1209548 }

{ "\_id" : { "city" : "SAN DIEGO", "state" : "CA" }, "Population" : 1049298 }

{ "\_id" : { "city" : "DETROIT", "state" : "MI" }, "Population" : 963243 }

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

{ "\_id" : { "city" : "PHOENIX", "state" : "AZ" }, "Population" : 890853 }

{ "\_id" : { "city" : "MIAMI", "state" : "FL" }, "Population" : 825232 }

{ "\_id" : { "city" : "SAN JOSE", "state" : "CA" }, "Population" : 816653 }

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

{ "\_id" : { "city" : "BALTIMORE", "state" : "MD" }, "Population" : 733081 }

{ "\_id" : { "city" : "SAN FRANCISCO", "state" : "CA" }, "Population" : 723993 }

{ "\_id" : { "city" : "MEMPHIS", "state" : "TN" }, "Population" : 632837 }

{ "\_id" : { "city" : "SACRAMENTO", "state" : "CA" }, "Population" : 628279 }

{ "\_id" : { "city" : "JACKSONVILLE", "state" : "FL" }, "Population" : 610160 }

{ "\_id" : { "city" : "ATLANTA", "state" : "GA" }, "Population" : 609591 }

Type "it" for more

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

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

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

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

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

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

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

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

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

**BONUS**

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

**> db.zipcodes.aggregate([ {$group:{\_id:"$state",avgpopulation:{$avg:"$pop"}}} ])**

{ "\_id" : "WI", "avgpopulation" : 6832.079608938548 }

{ "\_id" : "IL", "avgpopulation" : 9238.137429264349 }

{ "\_id" : "NH", "avgpopulation" : 5088.311926605505 }

{ "\_id" : "MN", "avgpopulation" : 4958.02947845805 }

{ "\_id" : "WA", "avgpopulation" : 10055.148760330578 }

{ "\_id" : "AK", "avgpopulation" : 2793.3230769230768 }

{ "\_id" : "IN", "avgpopulation" : 8201.384615384615 }

{ "\_id" : "ME", "avgpopulation" : 2991.8243902439026 }

{ "\_id" : "NY", "avgpopulation" : 11279.248902821317 }

{ "\_id" : "RI", "avgpopulation" : 14539.391304347826 }

{ "\_id" : "CT", "avgpopulation" : 12498.539923954373 }

{ "\_id" : "MT", "avgpopulation" : 2544.420382165605 }

{ "\_id" : "MO", "avgpopulation" : 5141.496981891348 }

{ "\_id" : "NE", "avgpopulation" : 2749.3710801393727 }

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

{ "\_id" : "WY", "avgpopulation" : 3239.4857142857145 }

{ "\_id" : "MD", "avgpopulation" : 11384.235714285714 }

{ "\_id" : "NJ", "avgpopulation" : 14315.162962962962 }

{ "\_id" : "CO", "avgpopulation" : 7955.929951690821 }

{ "\_id" : "ID", "avgpopulation" : 4126.020491803279 }

Type "it" for more

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

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

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

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

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

**---THANKYOU---**