Akshay. A. R

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

```
C:\Windows\System32\cmd.exe                                        —    □    ×
(c) Microsoft Corporation. All rights reserved.

C:\Users\LENOVO\Downloads\mongosh-1.1.9-win32-x64\mongosh-1.1.9-win32-x64\bin>mongoimport --db populatio
n --collection zipcodes --drop --file D:\zip.json
2022-02-10T11:24:12.367+0530    connected to: mongodb://localhost/
2022-02-10T11:24:12.370+0530    dropping: population.zipcodes
2022-02-10T11:24:13.530+0530    29353 document(s) imported successfully. 0 document(s) failed to import.

C:\Users\LENOVO\Downloads\mongosh-1.1.9-win32-x64\mongosh-1.1.9-win32-x64\bin>
```

Atlanta Population

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

```
Command Prompt - mongo                                             —    □    ×
> db.zipcodes.find({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"
}
{ "_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"
}
{ "_id" : "30311", "city" : "ATLANTA", "loc" : [ -84.470219, 33.722957 ], "pop" : 34880, "state" : "GA"
}
{ "_id" : "30312", "city" : "ATLANTA", "loc" : [ -84.378125, 33.746749 ], "pop" : 17683, "state" : "GA"
}
{ "_id" : "30313", "city" : "ATLANTA", "loc" : [ -84.39352, 33.76825 ], "pop" : 8038, "state" : "GA" }
{ "_id" : "30314", "city" : "ATLANTA", "loc" : [ -84.425546, 33.756103 ], "pop" : 26649, "state" : "GA"
}
{ "_id" : "30315", "city" : "ATLANTA", "loc" : [ -84.380771, 33.705062 ], "pop" : 41061, "state" : "GA"
}
{ "_id" : "30316", "city" : "ATLANTA", "loc" : [ -84.333913, 33.721686 ], "pop" : 34668, "state" : "GA"
}
{ "_id" : "30317", "city" : "ATLANTA", "loc" : [ -84.31685, 33.749788 ], "pop" : 16395, "state" : "GA" }
{ "_id" : "30318", "city" : "ATLANTA", "loc" : [ -84.445432, 33.786454 ], "pop" : 53894, "state" : "GA"
}
{ "_id" : "30319", "city" : "ATLANTA", "loc" : [ -84.335091, 33.868728 ], "pop" : 32138, "state" : "GA"
}
{ "_id" : "30324", "city" : "ATLANTA", "loc" : [ -84.354867, 33.820609 ], "pop" : 15044, "state" : "GA"
}
{ "_id" : "30326", "city" : "ATLANTA", "loc" : [ -84.358232, 33.848168 ], "pop" : 125, "state" : "GA" }
{ "_id" : "30327", "city" : "ATLANTA", "loc" : [ -84.419966, 33.862723 ], "pop" : 18467, "state" : "GA"
}
{ "_id" : "30329", "city" : "ATLANTA", "loc" : [ -84.321402, 33.823555 ], "pop" : 17013, "state" : "GA"
}
Type "it" for more
> it
{ "_id" : "30330", "city" : "ATLANTA", "loc" : [ -84.434735, 33.70645 ], "pop" : 643, "state" : "GA" }
{ "_id" : "30331", "city" : "ATLANTA", "loc" : [ -84.520468, 33.72241 ], "pop" : 38185, "state" : "GA" }
{ "_id" : "30334", "city" : "ATLANTA", "loc" : [ -84.388188, 33.74715 ], "pop" : 0, "state" : "GA" }
{ "_id" : "30336", "city" : "ATLANTA", "loc" : [ -84.510028, 33.78534 ], "pop" : 1228, "state" : "GA" }
{ "_id" : "30339", "city" : "ATLANTA", "loc" : [ -84.462879, 33.87125 ], "pop" : 11158, "state" : "GA" }
{ "_id" : "30342", "city" : "ATLANTA", "loc" : [ -84.376091, 33.884245 ], "pop" : 19057, "state" : "GA"
}
```

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

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

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

```
Command Prompt - mongo                                             —    □    ×
> db.zipcodes.aggregate([
... {$match:{city:"ATLANTA"}},
... {$group:{_id:"$_id"}},
... {$count:"Totalzips"}])
{ "Totalzips" : 41 }
> 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

2. sort the results by population, highest first

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

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' }

2. sort the results by population, highest first

```
Command Prompt - mongo                                          —   □   ×
> db.zipcodes.aggregate([ {$group: { _id: {city:"$city", state:"$state"},totalpop:{$sum:"$pop"}}},{$sort
:{totalpop:-1}}])
{ "_id" : { "city" : "CHICAGO", "state" : "IL" }, "totalpop" : 2452177 }
{ "_id" : { "city" : "BROOKLYN", "state" : "NY" }, "totalpop" : 2300504 }
{ "_id" : { "city" : "LOS ANGELES", "state" : "CA" }, "totalpop" : 2102295 }
{ "_id" : { "city" : "HOUSTON", "state" : "TX" }, "totalpop" : 2095918 }
{ "_id" : { "city" : "PHILADELPHIA", "state" : "PA" }, "totalpop" : 1610956 }
{ "_id" : { "city" : "NEW YORK", "state" : "NY" }, "totalpop" : 1476790 }
{ "_id" : { "city" : "BRONX", "state" : "NY" }, "totalpop" : 1209548 }
{ "_id" : { "city" : "SAN DIEGO", "state" : "CA" }, "totalpop" : 1049298 }
{ "_id" : { "city" : "DETROIT", "state" : "MI" }, "totalpop" : 963243 }
{ "_id" : { "city" : "DALLAS", "state" : "TX" }, "totalpop" : 940191 }
{ "_id" : { "city" : "PHOENIX", "state" : "AZ" }, "totalpop" : 890853 }
{ "_id" : { "city" : "MIAMI", "state" : "FL" }, "totalpop" : 825232 }
{ "_id" : { "city" : "SAN JOSE", "state" : "CA" }, "totalpop" : 816653 }
{ "_id" : { "city" : "SAN ANTONIO", "state" : "TX" }, "totalpop" : 811792 }
{ "_id" : { "city" : "BALTIMORE", "state" : "MD" }, "totalpop" : 733081 }
{ "_id" : { "city" : "SAN FRANCISCO", "state" : "CA" }, "totalpop" : 723993 }
{ "_id" : { "city" : "MEMPHIS", "state" : "TN" }, "totalpop" : 632837 }
{ "_id" : { "city" : "SACRAMENTO", "state" : "CA" }, "totalpop" : 628279 }
{ "_id" : { "city" : "JACKSONVILLE", "state" : "FL" }, "totalpop" : 610160 }
{ "_id" : { "city" : "ATLANTA", "state" : "GA" }, "totalpop" : 609591 }
Type "it" for more
> it
{ "_id" : { "city" : "WASHINGTON", "state" : "DC" }, "totalpop" : 606879 }
{ "_id" : { "city" : "LAS VEGAS", "state" : "NV" }, "totalpop" : 597557 }
{ "_id" : { "city" : "MILWAUKEE", "state" : "WI" }, "totalpop" : 597324 }
{ "_id" : { "city" : "TUCSON", "state" : "AZ" }, "totalpop" : 586451 }
{ "_id" : { "city" : "CLEVELAND", "state" : "OH" }, "totalpop" : 536759 }
{ "_id" : { "city" : "AUSTIN", "state" : "TX" }, "totalpop" : 530049 }
{ "_id" : { "city" : "SEATTLE", "state" : "WA" }, "totalpop" : 520096 }
{ "_id" : { "city" : "COLUMBUS", "state" : "OH" }, "totalpop" : 520096 }
{ "_id" : { "city" : "PORTLAND", "state" : "OR" }, "totalpop" : 518543 }
{ "_id" : { "city" : "EL PASO", "state" : "TX" }, "totalpop" : 511351 }
{ "_id" : { "city" : "TAMPA", "state" : "FL" }, "totalpop" : 498997 }
{ "_id" : { "city" : "NEW ORLEANS", "state" : "LA" }, "totalpop" : 496937 }
{ "_id" : { "city" : "CHARLOTTE", "state" : "NC" }, "totalpop" : 465833 }
{ "_id" : { "city" : "FORT WORTH", "state" : "TX" }, "totalpop" : 451415 }
{ "_id" : { "city" : "DENVER", "state" : "CO" }, "totalpop" : 451182 }
{ "_id" : { "city" : "ALBUQUERQUE", "state" : "NM" }, "totalpop" : 449584 }
{ "_id" : { "city" : "SAINT LOUIS", "state" : "MO" }, "totalpop" : 397802 }
{ "_id" : { "city" : "HONOLULU", "state" : "HI" }, "totalpop" : 396643 }
{ "_id" : { "city" : "ROCHESTER", "state" : "NY" }, "totalpop" : 396013 }
{ "_id" : { "city" : "TULSA", "state" : "OK" }, "totalpop" : 389072 }
Type "it" for more
>
```

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

```
Command Prompt - mongo                                          —   □   ×
> db.zipcodes.aggregate([ {$group: { _id: {city:"$city", state:"$state"},totalpop:{$sum:"$pop"}}},{$sort
:{totalpop:-1}},{$limit:3}])
{ "_id" : { "city" : "CHICAGO", "state" : "IL" }, "totalpop" : 2452177 }
{ "_id" : { "city" : "BROOKLYN", "state" : "NY" }, "totalpop" : 2300504 }
{ "_id" : { "city" : "LOS ANGELES", "state" : "CA" }, "totalpop" : 2102295 }
>
```

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

```
Command Prompt - mongo                                          —   □   ×
> db.zipcodes.aggregate([ {$group: { _id: {city:"$city", state:"$state"},totalpop:{$sum:"$pop"}}},{$sort
:{totalpop:-1}},{$match:{"_id.state":"TX"}},{$limit:3}])
{ "_id" : { "city" : "HOUSTON", "state" : "TX" }, "totalpop" : 2095918 }
{ "_id" : { "city" : "DALLAS", "state" : "TX" }, "totalpop" : 940191 }
{ "_id" : { "city" : "SAN ANTONIO", "state" : "TX" }, "totalpop" : 811792 }
> ▪
```

Bonus

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

```
Command Prompt - mongo                                                    —   □   ×
> db.zipcodes.aggregate([
... {$group:{_id:{state:"$state",city:"$city"},pop:{$sum:"$pop"}}},
... {$group:{_id:"$_id.state",avgCityPop:{$avg:"$pop"}}}
... ])
{ "_id" : "GA", "avgCityPop" : 11547.62210338681 }
{ "_id" : "AL", "avgCityPop" : 7907.2152641878665 }
{ "_id" : "SD", "avgCityPop" : 1839.6746031746031 }
{ "_id" : "NC", "avgCityPop" : 10622.815705128205 }
{ "_id" : "LA", "avgCityPop" : 10465.496277915632 }
{ "_id" : "RI", "avgCityPop" : 19292.653846153848 }
{ "_id" : "MT", "avgCityPop" : 2593.987012987013 }
{ "_id" : "IA", "avgCityPop" : 3123.0821147356583 }
{ "_id" : "VT", "avgCityPop" : 2315.8765432098767 }
{ "_id" : "WI", "avgCityPop" : 7323.00748502994 }
{ "_id" : "WA", "avgCityPop" : 12258.670025188916 }
{ "_id" : "MS", "avgCityPop" : 7524.023391812865 }
{ "_id" : "HI", "avgCityPop" : 15831.842857142858 }
{ "_id" : "AK", "avgCityPop" : 2976.4918032786886 }
{ "_id" : "KS", "avgCityPop" : 3819.884259259259 }
{ "_id" : "AR", "avgCityPop" : 4175.355239786856 }
{ "_id" : "ME", "avgCityPop" : 3006.4901960784314 }
{ "_id" : "NY", "avgCityPop" : 13131.680291970803 }
{ "_id" : "FL", "avgCityPop" : 27400.958963282937 }
{ "_id" : "IN", "avgCityPop" : 9271.130434782608 }
Type "it" for more
> it
{ "_id" : "PA", "avgCityPop" : 8679.067202337472 }
{ "_id" : "ND", "avgCityPop" : 1645.0309278350514 }
{ "_id" : "NE", "avgCityPop" : 3034.882692307692 }
{ "_id" : "MN", "avgCityPop" : 5372.21375921376 }
{ "_id" : "NV", "avgCityPop" : 18209.590909090908 }
{ "_id" : "OK", "avgCityPop" : 6155.743639921722 }
{ "_id" : "OR", "avgCityPop" : 8262.561046511628 }
{ "_id" : "NH", "avgCityPop" : 5232.320754716981 }
{ "_id" : "CO", "avgCityPop" : 9981.075757575758 }
{ "_id" : "MD", "avgCityPop" : 12615.775725593667 }
{ "_id" : "AZ", "avgCityPop" : 20591.16853932584 }
{ "_id" : "MO", "avgCityPop" : 5672.195338512764 }
{ "_id" : "CT", "avgCityPop" : 14674.625 }
{ "_id" : "TX", "avgCityPop" : 13775.02108678021 }
{ "_id" : "NJ", "avgCityPop" : 15775.89387755102 }
{ "_id" : "CA", "avgCityPop" : 27756.42723880597 }
{ "_id" : "WY", "avgCityPop" : 3384.5373134328356 }
{ "_id" : "DE", "avgCityPop" : 14481.91304347826 }
{ "_id" : "SC", "avgCityPop" : 11139.626198083068 }
{ "_id" : "MA", "avgCityPop" : 14855.37037037037 }
Type "it" for more
> ∎
```

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

```
Command Prompt - mongo                                                    —   □   ×
> 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 }
> ∎
```