**MongoDB -Aggregation Exercises**

**Jitendra Singh Barod**

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

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'})

{ "\_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

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

=> db.zipcodes.aggregate([{$match:{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

{ "\_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" }

{ "\_id" : "30345", "city" : "ATLANTA", "loc" : [ -84.286961, 33.851347 ], "pop" : 19825, "state" : "GA" }

{ "\_id" : "30346", "city" : "ATLANTA", "loc" : [ -84.333354, 33.926717 ], "pop" : 18, "state" : "GA" }

{ "\_id" : "30349", "city" : "ATLANTA", "loc" : [ -84.481258, 33.605331 ], "pop" : 48116, "state" : "GA" }

{ "\_id" : "30350", "city" : "ATLANTA", "loc" : [ -84.341146, 33.979471 ], "pop" : 24573, "state" : "GA" }

{ "\_id" : "30360", "city" : "ATLANTA", "loc" : [ -84.271645, 33.937772 ], "pop" : 16023, "state" : "GA" }

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

=> db.zipcodes.aggregate([{$group:{\_id:'$state'}},{$count:'Number\_of\_states'}])

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

=> db.zipcodes.aggregate([{$match:{city:'ATLANTA'}},{$group:{\_id:'$city',Totalpopulation:{$sum:'$pop'}}}])

**Populations By State:**

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

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

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

{ "\_id" : { "state" : "TX" }, "Totalpopulation" : 16984601 }

{ "\_id" : { "state" : "WY" }, "Totalpopulation" : 453528 }

{ "\_id" : { "state" : "ID" }, "Totalpopulation" : 1006749 }

{ "\_id" : { "state" : "MD" }, "Totalpopulation" : 4781379 }

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

{ "\_id" : { "state" : "NJ" }, "Totalpopulation" : 7730188 }

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

{ "\_id" : { "state" : "CT" }, "Totalpopulation" : 3287116 }

{ "\_id" : { "state" : "CO" }, "Totalpopulation" : 3293755 }

{ "\_id" : { "state" : "MO" }, "Totalpopulation" : 5110648 }

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

{ "\_id" : { "state" : "MI" }, "Totalpopulation" : 9295297 }

{ "\_id" : { "state" : "SC" }, "Totalpopulation" : 3486703 }

{ "\_id" : { "state" : "OH" }, "Totalpopulation" : 10846517 }

{ "\_id" : { "state" : "WV" }, "Totalpopulation" : 1793146 }

{ "\_id" : { "state" : "TN" }, "Totalpopulation" : 4876457 }

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

{ "\_id" : { "state" : "KY" }, "Totalpopulation" : 3675484 }

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

Type "it" for more

2. sort the results by population, highest first

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

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

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

{ "\_id" : { "state" : "TX" }, "Totalpopulation" : 16984601 }

{ "\_id" : { "state" : "FL" }, "Totalpopulation" : 12686644 }

{ "\_id" : { "state" : "PA" }, "Totalpopulation" : 11881643 }

{ "\_id" : { "state" : "IL" }, "Totalpopulation" : 11427576 }

{ "\_id" : { "state" : "OH" }, "Totalpopulation" : 10846517 }

{ "\_id" : { "state" : "MI" }, "Totalpopulation" : 9295297 }

{ "\_id" : { "state" : "NJ" }, "Totalpopulation" : 7730188 }

{ "\_id" : { "state" : "NC" }, "Totalpopulation" : 6628637 }

{ "\_id" : { "state" : "GA" }, "Totalpopulation" : 6478216 }

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

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

{ "\_id" : { "state" : "IN" }, "Totalpopulation" : 5544136 }

{ "\_id" : { "state" : "MO" }, "Totalpopulation" : 5110648 }

{ "\_id" : { "state" : "WI" }, "Totalpopulation" : 4891769 }

{ "\_id" : { "state" : "TN" }, "Totalpopulation" : 4876457 }

{ "\_id" : { "state" : "WA" }, "Totalpopulation" : 4866692 }

{ "\_id" : { "state" : "MD" }, "Totalpopulation" : 4781379 }

{ "\_id" : { "state" : "MN" }, "Totalpopulation" : 4372982 }

Type "it" for more

3. 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:'$state'},Totalpopulation:{$sum:'$pop'}}}])

{ "\_id" : { "city" : "ANNAPOLIS", "state" : "IL" }, "Totalpopulation" : 692 }

{ "\_id" : { "city" : "WYOMING", "state" : "NY" }, "Totalpopulation" : 2074 }

{ "\_id" : { "city" : "POMEROYTON", "state" : "KY" }, "Totalpopulation" : 480 }

{ "\_id" : { "city" : "MOUNT PENN", "state" : "PA" }, "Totalpopulation" : 27732 }

{ "\_id" : { "city" : "LONETREE", "state" : "WY" }, "Totalpopulation" : 24 }

{ "\_id" : { "city" : "WEST BURLINGTON", "state" : "NY" }, "Totalpopulation" : 22 }

{ "\_id" : { "city" : "RODMAN", "state" : "NY" }, "Totalpopulation" : 1097 }

{ "\_id" : { "city" : "TRENT", "state" : "SD" }, "Totalpopulation" : 865 }

{ "\_id" : { "city" : "VERNON", "state" : "CO" }, "Totalpopulation" : 318 }

{ "\_id" : { "city" : "LANSING", "state" : "KS" }, "Totalpopulation" : 8145 }

{ "\_id" : { "city" : "WEST YELLOWSTONE", "state" : "MT" }, "Totalpopulation" : 1987 }

{ "\_id" : { "city" : "ETNA", "state" : "WY" }, "Totalpopulation" : 524 }

{ "\_id" : { "city" : "COS COB", "state" : "CT" }, "Totalpopulation" : 7000 }

{ "\_id" : { "city" : "FLORIN", "state" : "PA" }, "Totalpopulation" : 12282 }

{ "\_id" : { "city" : "APALACHICOLA", "state" : "FL" }, "Totalpopulation" : 3859 }

{ "\_id" : { "city" : "GALVESTON", "state" : "KY" }, "Totalpopulation" : 1030 }

{ "\_id" : { "city" : "KEOKUK", "state" : "IA" }, "Totalpopulation" : 13995 }

{ "\_id" : { "city" : "PRIMROSE", "state" : "NE" }, "Totalpopulation" : 255 }

{ "\_id" : { "city" : "LOSTANT", "state" : "IL" }, "Totalpopulation" : 747 }

{ "\_id" : { "city" : "BADEN", "state" : "PA" }, "Totalpopulation" : 10068 }

Type "it" for more

2. sort the results by population, highest first

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Type "it" for more

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:'$state'},Totalpopulation:{$sum:'$pop'}}},{$sort:{Totalpopulation:-1}},{$limit:3} ])

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

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

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

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

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

{ "\_id" : "HOUSTON", "population" : 2095918 }

{ "\_id" : "DALLAS", "population" : 940191 }

{ "\_id" : "SAN ANTONIO", "population" : 811792 }

Bonus:

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

=> db.zipcodes.aggregate([{$group:{\_id:{city:'$city'},Average\_population:{$avg:'$pop'}}}])

{ "\_id" : { "city" : "KENNEDY SPACE CE" }, "Average\_population" : 1 }

{ "\_id" : { "city" : "RIPPEY" }, "Average\_population" : 563 }

{ "\_id" : { "city" : "OLD ORCHARD BEAC" }, "Average\_population" : 8451 }

{ "\_id" : { "city" : "PAOLI" }, "Average\_population" : 4916 }

{ "\_id" : { "city" : "NICASIO" }, "Average\_population" : 665 }

{ "\_id" : { "city" : "HYDES" }, "Average\_population" : 883 }

{ "\_id" : { "city" : "WAPITI" }, "Average\_population" : 214 }

{ "\_id" : { "city" : "ERSKINE" }, "Average\_population" : 993 }

{ "\_id" : { "city" : "BENS RUN" }, "Average\_population" : 282 }

{ "\_id" : { "city" : "LEVERING" }, "Average\_population" : 1251 }

{ "\_id" : { "city" : "CARRIZO SPRINGS" }, "Average\_population" : 7682 }

{ "\_id" : { "city" : "OREGON" }, "Average\_population" : 11677 }

{ "\_id" : { "city" : "BLACK ROCK" }, "Average\_population" : 1190 }

{ "\_id" : { "city" : "WAUPUN" }, "Average\_population" : 11222 }

{ "\_id" : { "city" : "CONTINENTAL" }, "Average\_population" : 3518 }

{ "\_id" : { "city" : "HOLDREGE" }, "Average\_population" : 6744 }

{ "\_id" : { "city" : "CLAIRE CITY" }, "Average\_population" : 394 }

{ "\_id" : { "city" : "LE SUEUR" }, "Average\_population" : 5491 }

{ "\_id" : { "city" : "KRESS" }, "Average\_population" : 1643 }

{ "\_id" : { "city" : "COLLINSTON" }, "Average\_population" : 1040 }

Type "it" for more

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

=> db.zipcodes.aggregate([{$group:{\_id:{city:'$city'},Average\_pop:{$avg:'$pop'}}},{$sort:{Average\_pop:-1}},{$limit:3}])

{ "\_id" : { "city" : "BELL GARDENS" }, "Average\_population" : 99568 }

{ "\_id" : { "city" : "ARLETA" }, "Average\_population" : 88114 }

{ "\_id" : { "city" : "SOUTH GATE" }, "Average\_population" : 87026 }