

MongoDB

Assignment-2

Assignment 2

#import

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

>Problem faced during the import due of absence of mongodb database tools.

>Installed the database tools then performed the import operation by the following command

```
mongoimport --host localhost:27017 --db population --collection zipcodes <
D:\Capgemini\MongoDB\zip.json
```

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()
```

```
db.zipcodes.find({city:"ATLANTA", state:"GA"}).count()
```

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

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

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

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

```
Out put:{ "_id" : "ATLANTA", "count" : 41 }
```

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

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

```
Out put:{ "_id" : "ATLANTA", "Totalpopulation" : 41 }
```

population by state

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

```
db.zipcodes.aggregate([{$group:{_id:"$state", populationbystate:{$sum:"$pop"}}}])
```

2. sort the results by population, highest first

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

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

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

Output: { "_id" : "CA", "populationbystate" : 29754890 }

{ "_id" : "NY", "populationbystate" : 17990402 }

{ "_id" : "TX", "populationbystate" : 16984601 }

population 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", Totalpopulation:{$sum:"$pop"}}}])
```

2. sort the results by population, highest first

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

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

```
db.zipcodes.aggregate([{$group:{_id:"$city", populationofstate:{$sum:"$pop"}},{sort:{populationofcity:-1}},{limit:3}])
```

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

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

Output: { "_id" : "HOUSTON", "populationofcity" : 2095918 }

{ "_id" : "DALLAS", "populationofcity" : 940191 }

```
{ "_id" : "SAN ANTONIO", "populationofcity" : 811792 }
```

Bonus

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

```
db.zipcodes.aggregate([{$group:{_id:"$state", averagepopulation:{$avg:"$pop"}}}])
```

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

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

Output: { "_id" : "DC", "averagepopulation" : 25287.5 }

```
{ "_id" : "CA", "averagepopulation" : 19627.236147757256 }
```

```
{ "_id" : "FL", "averagepopulation" : 15779.407960199005 }
```