

Week 2 :Lab

Important Note.

At the End of the lab manual, you have given a short task. Complete the task as mentioned and write a reflection for your portfolio.

What is MongoDB

MongoDB is an open-source database that uses a document-oriented data model. MongoDB is built on an architecture of collections. It is very useful when there is no specific database **structure** like RDBMS and varies data (i.e. columns) as per requirements. Terminologies used in MongoDB are Collection, Documents, Fields, Schema, and Models.

1. **Collections** in Mongo are equivalent to **tables** in relational databases. They can hold multiple JSON documents.
2. **Documents** are equivalent to **records or rows** of data in SQL. While a SQL row can reference data in other tables, Mongo documents usually combine that in a document.
3. **Fields** or attributes are similar to **columns** in a SQL table.
4. **Schema**: A Mongoose 'schema' is a document data structure (or shape of the document) that is enforced via the application layer.
5. **Models**: are higher-order constructors that take a schema and create an instance of a document equivalent to records in a relational database.

Week 2 Lab tutorial has following tasks

1. You are provided with a CSV file named Peoples.csv. Your task is to download the file and perform the following MongoDB operations using MongoDB Compass.
 - a. **Inserting a document**
 - b. **Updating a document**
 - c. **Deleting a document**
 - d. **Aggerate pipe line**

Managing Data With Compass (CRUD)

Creating a Database & Initial Collection

Step 1

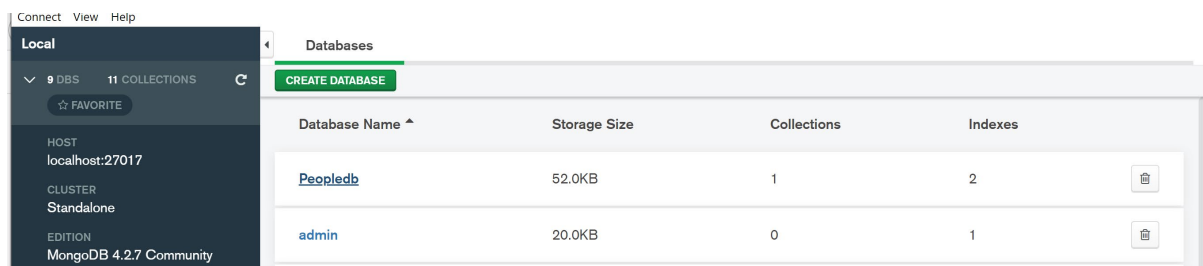
- Open **MongoDB Compass Community** that you installed earlier.
- Connect to the MongoDB server by entering the following parameters:

- **Hostname:** localhost
- **Port:** 27017

- Click **Connect**.
- From the dashboard, click the **Create Database** button.
- The **Create Database** helper will appear (as shown in Figure 7).
- Enter the following details:

- **Database Name:** peopledb
- **Collection Name:** people

1. • Click the **Create Database** button..



write Peopledb in Database Name

people in collection Name

More Information'. At the bottom are 'CANCEL' and 'CREATE DATABASE' buttons. A blue arrow points to the 'CREATE DATABASE' button."/>

Create Database

Database Name

Peopledb

Collection Name

people

☐ Capped Collection ⓘ

☐ Use Custom Collation ⓘ

Before MongoDB can save your new database, a collection name must also be specified at the time of creation. [More Information](#)

CANCEL CREATE DATABASE

Adding Data to the collection either through

1. Importing a CSV File or
2. Insert Document

Adding Data through Importing a CSV File

On the left-hand panel click Peopledb and click people collection then click **ADD DATA**, and then click Import JSON or CSV file .



The following screen will appear. Select CSV , Click browser to locate the file people.csv (provided to you for this week 's lab on Moodle) in your computer select the format CSV for the people.csv file and finally click import to import the selected file.

Import To Collection Peopledb,people

Select File

BROWSE

Select Input File Type

JSON CSV

Options

Select delimiter COMMA

☒ Ignore empty strings

☐ Stop on errors

CANCEL IMPORT

Changing the Data type of the Column Field , while importing CSV data

Click the field or column you want to change the field. In this case, we select the Age field. Change the type to number from string.

Documents Aggregations Schema

ADD DATA

Select File

C:\Users\mqazi\NQI\Cn5006_CourseContent\mongodb\people.csv BROWSE

Select Input File Type

JSON CSV

Options

Select delimiter COMMA

☒ Ignore empty strings

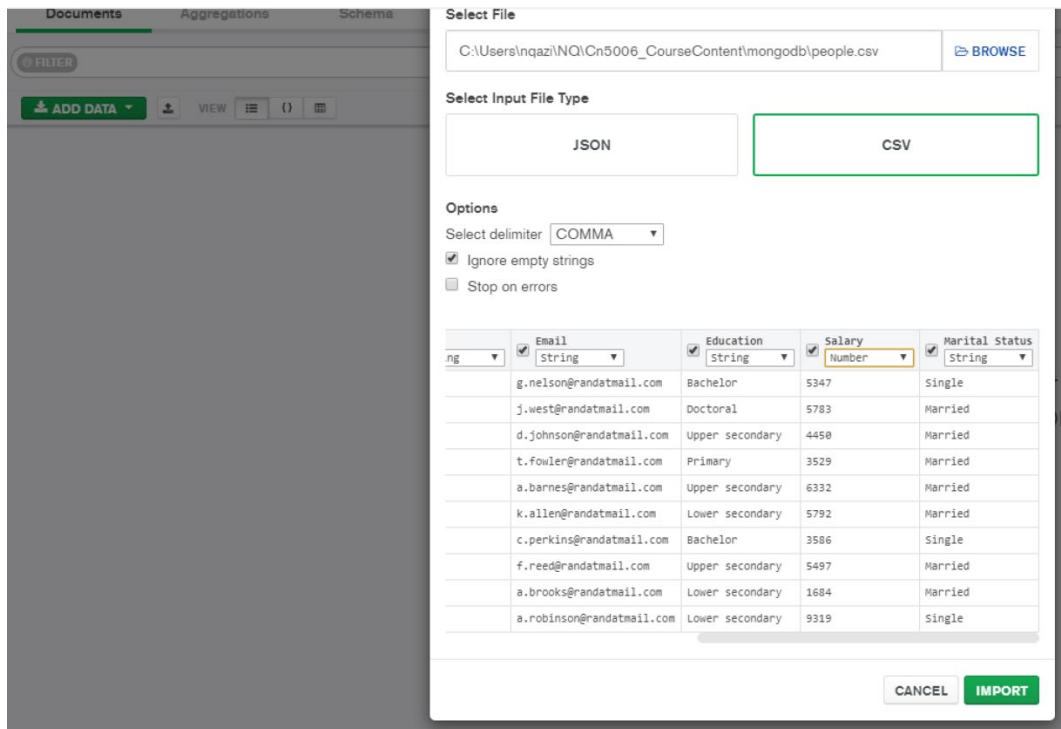
☐ Stop on errors

Specify Fields and Types

	First Name	Last Name	Gender	Age	Email
	String	String	String	Number	String
1	Grace	Nelson	Female	21	g.nelson@randi
2	Justin	West	Male	27	j.west@randat
3	Daryl	Johnson	Male	28	d.johnson@ran
4	Tiana	Fowler	Female	27	t.fowler@randi
5	Alen	Barnes	Male	26	a.barnes@randi
6	Kirsten	Allen	Female	21	k.allen@randat
7	Charlie	Perkins	Male	28	c.perkins@ran
8	Florrie	Reed	Female	19	f.reed@randat
9	Amber	Brooks	Female	27	a.brooks@randi
10	Alberta	Robinson	Female	27	a.robinson@ra

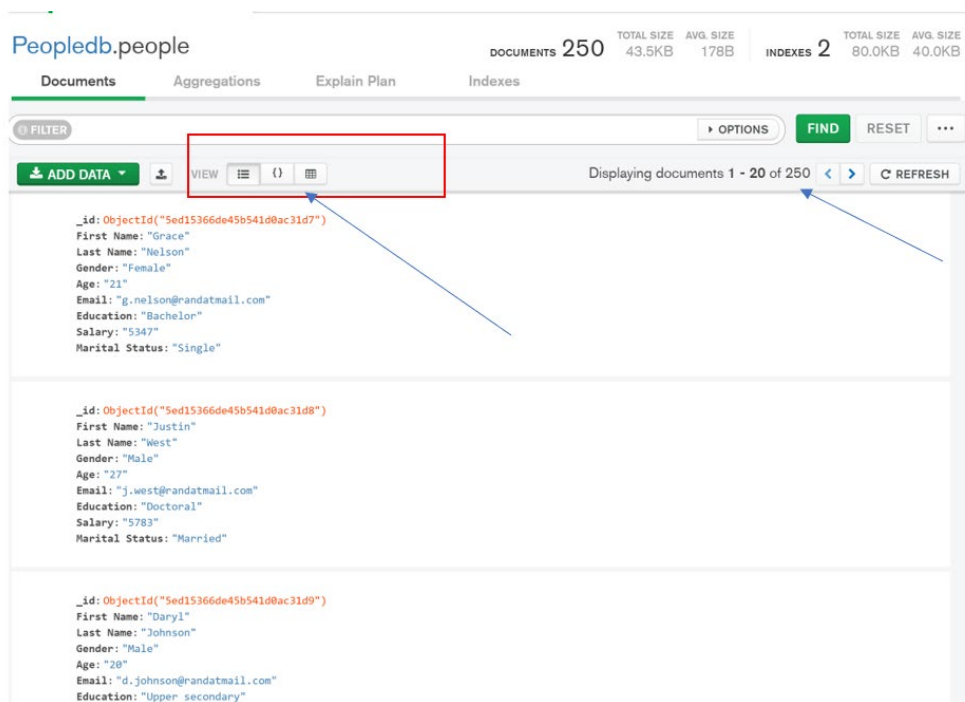
CANCEL IMPORT

Next change the type of the field Salary. Select Number and click the import button



This will import the data in the people.csv file to the people collection of your **peopledb** database. Note the number of documents.

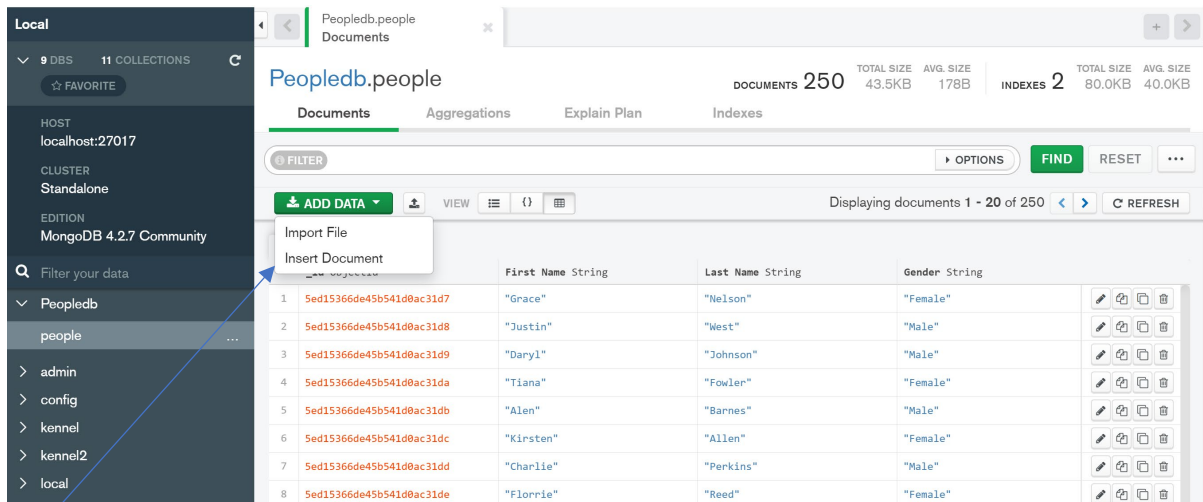
Once the data is imported it's time to perform some basic queries in MongoDB . Select Document Tab



Next to View, you will observe three view settings list, JSON and table view. Click each of them to observe how data can be displayed in the collection. Note, , for each document there is a `_id` field associate DO NOT change it.

Inserting a document in the collection

Now is the time to insert a new document to the people collection. To do this click Add data but this time select Insert document



A new window will be displayed , select the list view as shown in the figure below. Note there is a default `_id` field created for you do not change it

Insert to Collection Peopledb.people

VIEW ☐ ☒

1

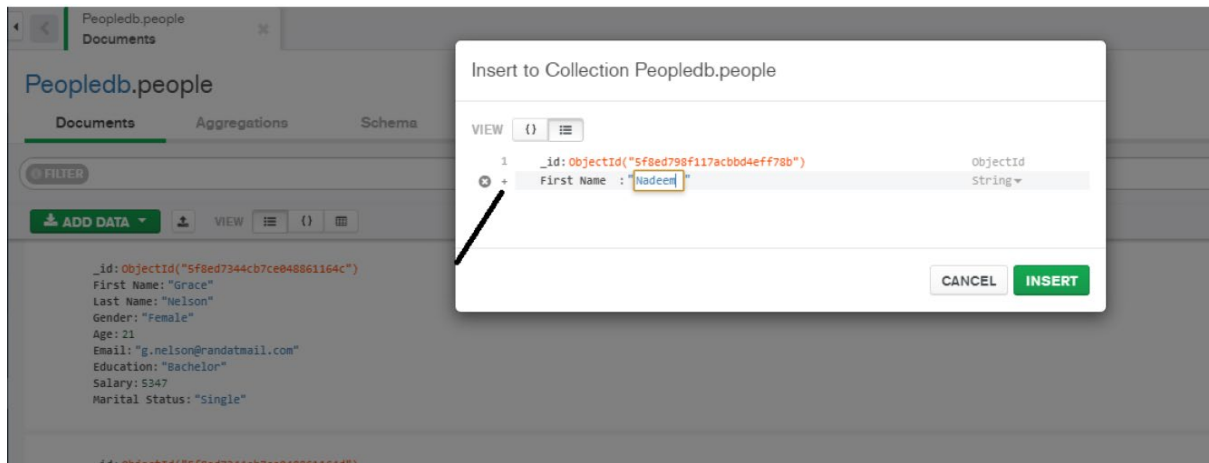
`_id: ObjectId("5ef5209a0216721de7152906")`

`: " "`

ObjectId
String

For Inserting a new record:

click the + sign , it will add one field in the document for you to give a name and value in that field.



Add more fields to this document using + sign.

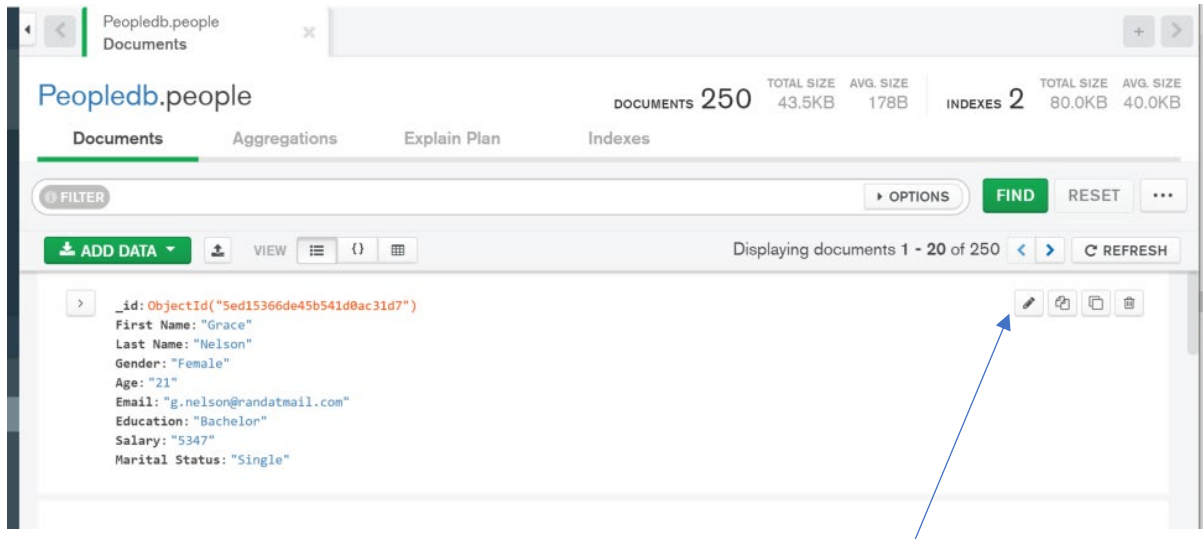
Remember for each of this fields you to click + sign on the left and select the type from the right.

Once you filled all the fields from 2 to 9 then click **Insert** to insert the document in the collection.

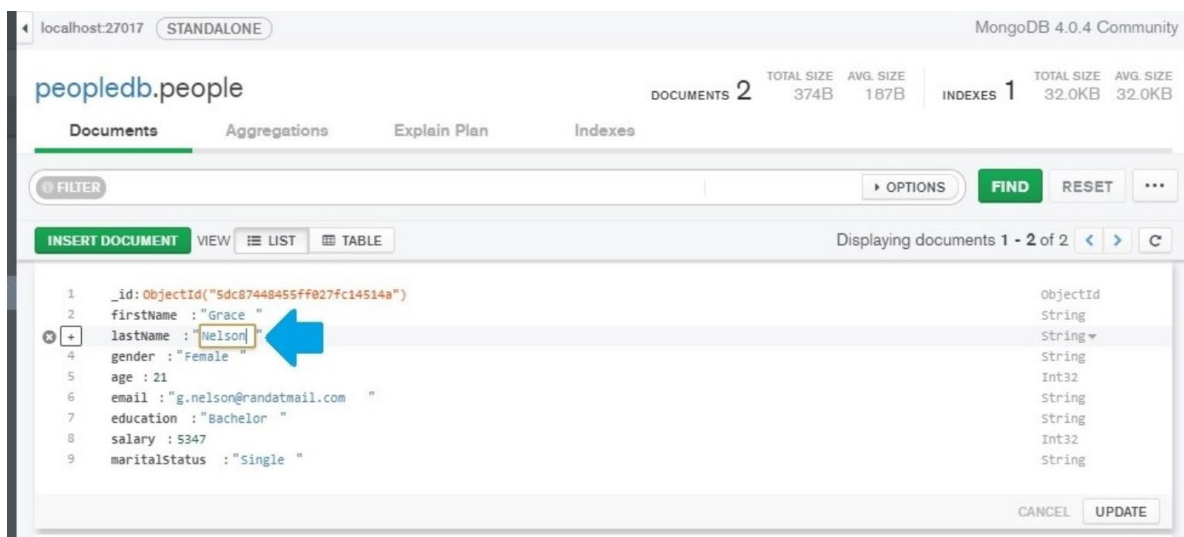


Updating the document

1. Locate the document you want to update and hover over it until the tooltip pops up.
2. Click the edit document Pencil icon as shown in Figure This will enable the full modification of the document within the viewer.



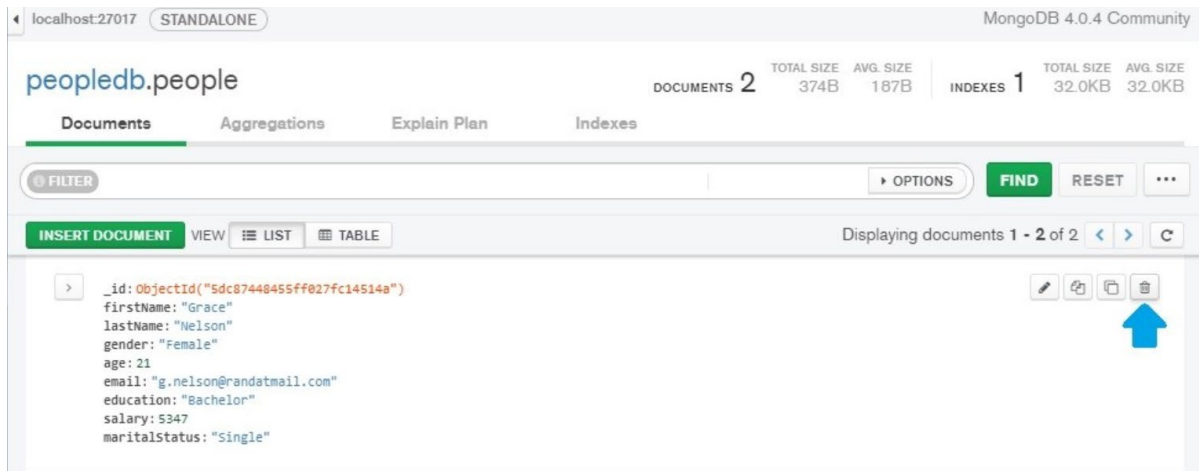
3. Double click the value field of the data you want to update, as shown in the Figure below
4. Change the data to your requirement and press enter.
5. You will notice a Document Modified notification will appear, click the update button on this notification to save changes. See Figure 17.



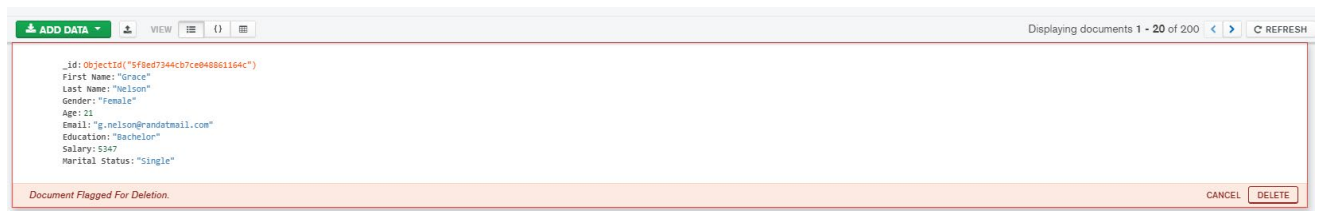
Deleting the documents

Deleting documents is another simple task with the MongoDB Compass software and can be done with two clicks of a button. Please follow the next steps to delete a document.

1. Locate the document you want to update and hover over it until the tooltip pops up.
2. Click the delete Trash Can icon as shown in the Figure below:



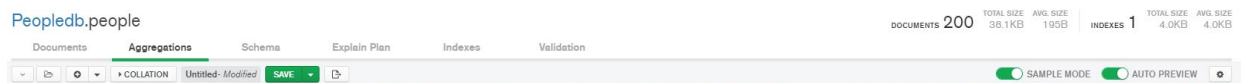
1. A notification will appear stating the document has been flagged for deletion in red, Click the delete button as shown in figure the document will be deleted.



Aggregations:

Aggregation operations process data records and return computed results. Aggregation operations group values from multiple documents together, and can perform a variety of operations on the grouped data to return a single result. MongoDB provides three ways to perform aggregation: the [aggregation pipeline](#), the [map-reduce function](#), and [single purpose aggregation methods](#). In this tutorial we will discuss the aggregation pipeline only.

You can find the aggregation tab just adjacent to the document tab as shown below:



Aggregation:

Aggregation is performed in multiple stages to produce aggregated results from the documents. Aggregation pipelines are a composition of various stages that transform and filter the data. For example suppose we in our collection people we want to see:

- How many people are there who has got the bachelor degree and are older than 21 years of age.
- what is average age of Female and male in this group
- what is the age salary of the Female and male in this group?
- What is max age of male and female in this group

records that match this criteria. Note as soon as you finished the write syntax of the query you will see the output in the adjacent window as shown in the figure below

```
{
  Education:"Bachelor",
  Age:{$gte:21}
}
```

Peopledb.people

Documents Aggregations Schema Explain Plan Indexes Validation

DOCUMENTS 200 TOTAL SIZE 38.1KB AVG. SIZE 195B INDEXES 1 TOTAL SIZE 4.0KB AVG. SIZE 4.0KB

COLLATION Untitled- Modified SAVE SAMPLE MODE AUTO PREVIEW

200 Documents in the Collection Preview of Documents in the Collection

\$match Output after \$match stage (Sample of 20 documents)

```
1 /**
2  * query: The query in MQL.
3  */
4 {
5   Education:"Bachelor",
6   Age:{$gte:21}
7 }
```

ADD STAGE

Click it to add another query

2nd Query :

- **Now lets find out the** what is average age of Female and male in this group. For this we will be using group aggregate.
- \$group: Groups input documents by the specified _id expression and for each distinct grouping, outputs a document. The _id field of each output document contains the unique group by value. The output documents can also contain computed fields that hold the values of some accumulator expression.
- In order to do group, click add stage button and it will create another window , click select and then choose group as shown below :

Peopledb.people

Documents Aggregations Schema Explain Plan Indexes Validation

DOCUMENTS 200 TOTAL SIZE 38.1KB AVG. SIZE 195B INDEXES 1 TOTAL SIZE 4.0KB AVG. SIZE 4.0KB

COLLATION Untitled- Modified SAVE SAMPLE MODE AUTO PREVIEW

200 Documents in the Collection Preview of Documents in the Collection

\$match Output after \$match stage (Sample of 25 documents)

```
1 /**
2  * query: The query in MQL.
3  */
4 {
5   Education:"Bachelor",
6   Age:{$gte:21}
7 }
```

\$group Output after \$group stage (Sample of 0 documents)

```
1 /**
2  * id: The id of the group.
3  * fieldN: The first field name.
4  */
5 {
6   _id: '$expression',
7   FieldN: {
8     accumulatorN: 'expressionN'
9   }
10 }
```

clear the text and write new query

Now write the following query in the space shown in red box in the above figure

```
{
  _id: "$Gender",
  Avg: {
    $avg: "$Age"
  }
}
```

The screenshot shows the MongoDB Compass interface. The top bar indicates 200 documents, 38.1KB total size, 195B avg. size, and 1 index. The 'Aggregations' tab is active. The query editor shows the following query:

```
1 /**
2  * query: The query in NQL.
3  */
4 {
5   $group: {
6     _id: "$Gender",
7     Avg: {
8       $avg: "$Age"
9     }
10  }
11 }
```

The output of the aggregation is shown in the right pane, displaying two documents:

```
{ "_id": "Male", "Avg": 25.666666666666668 }
{ "_id": "Female", "Avg": 25 }
```

Query Explanation: Note there is \$ sign before the field Gender and it is enclosed in double quote

\$avg is the aggregate function provided the mongo dB and in double quotes "\$Age" is the name of the field in the documents over which this aggregate function is applied. So this query returns average age of female and male .

it is analogy of the SQL statement `select avg(age) from people group by gender`

Add another query for the max and min Age of male and female in this group as shown in the figure below:

Query is :

```
MinAge: { $min: "$Age" },
MaxAge: { $max: "$Age" },
```

Your screen should look like this

Match stage (Sample of 20 documents)

```

1 /**
2  * query: The query in HQL.
3  */
4 {
5   "Education": "Bachelor",
6   Age: {$gte:21}
7 }

```

Output after Match stage (Sample of 20 documents)

```

1 {
2   "_id": "ObjectID('5f8f72e727eba786384dfc5e')",
3   "First Name": "Charlie",
4   "Last Name": "Perkins",
5   "Gender": "Male",
6   "Age": 28,
7   "Email": "c.perkins@randatmail.com",
8   "Education": "Bachelor",
9   "Salary": 3586,
10  "Marital Status": "Single"
11 }
12 {
13   "_id": "ObjectID('5f8f72e727eba786384dfc5c')",
14   "First Name": "Naomi",
15   "Last Name": "Spencer",
16   "Gender": "Female",
17   "Age": 26,
18   "Email": "n.spencer@randatmail.com",
19   "Education": "Bachelor",
20   "Salary": 5987,
21   "Marital Status": "Married"
22 }
23 {
24   "_id": "ObjectID('5f8f72e727eba786384dfc61')",
25   "First Name": "Nicholas",
26   "Last Name": "Barrett",
27   "Gender": "Male",
28   "Age": 24,
29   "Email": "n.barrett@randatmail.com",
30   "Education": "Bachelor",
31   "Salary": 5577,
32   "Marital Status": "Married"
33 }
34 {
35   "_id": "ObjectID('5f8f72e727eba786384dfc61')",
36   "First Name": "Nicholas",
37   "Last Name": "Barrett",
38   "Gender": "Male",
39   "Age": 24,
40   "Email": "n.barrett@randatmail.com",
41   "Education": "Bachelor",
42   "Salary": 5577,
43   "Marital Status": "Married"
44 }

```

Group stage (Sample of 2 documents)

```

1 /**
2  * _id: The id of the group.
3  * fieldN: The first field name.
4  */
5 {
6   "_id": "$Gender",
7   Avg: {$avg: "$Age"},
8   MinAge: {$min: "$Age"},
9   MaxAge: {$max: "$Age"}
10 }
11 }
12 }

```

Output after Group stage (Sample of 2 documents)

```

1 {
2   "_id": "Female",
3   "Avg": 25.333333333333332,
4   "MinAge": 21,
5   "MaxAge": 29
6 }
7 {
8   "_id": "Male",
9   "Avg": 25.666666666666668,
10  "MinAge": 22,
11  "MaxAge": 30
12 }

```

Next Query find the min, max and average salary of the male and female
Add the following query. Notice this time it is not age but salary in double quotes

MaxSalary: {\$max: "\$Salary"},

MinSalary: {\$min: "\$Salary"},

AvgSalary: {\$avg: "\$Salary"}

Your screen now should look like this

Match stage (Sample of 20 documents)

```

1 /**
2  * query: The query in HQL.
3  */
4 {
5   "Education": "Bachelor",
6   Age: {$gte:21}
7 }

```

Output after Match stage (Sample of 20 documents)

```

1 {
2   "_id": "ObjectID('5f8f72e727eba786384dfc5e')",
3   "First Name": "Charlie",
4   "Last Name": "Perkins",
5   "Gender": "Male",
6   "Age": 28,
7   "Email": "c.perkins@randatmail.com",
8   "Education": "Bachelor",
9   "Salary": 3586,
10  "Marital Status": "Single"
11 }
12 {
13   "_id": "ObjectID('5f8f72e727eba786384dfc5c')",
14   "First Name": "Naomi",
15   "Last Name": "Spencer",
16   "Gender": "Female",
17   "Age": 26,
18   "Email": "n.spencer@randatmail.com",
19   "Education": "Bachelor",
20   "Salary": 5987,
21   "Marital Status": "Married"
22 }
23 {
24   "_id": "ObjectID('5f8f72e727eba786384dfc61')",
25   "First Name": "Nicholas",
26   "Last Name": "Barrett",
27   "Gender": "Male",
28   "Age": 24,
29   "Email": "n.barrett@randatmail.com",
30   "Education": "Bachelor",
31   "Salary": 5577,
32   "Marital Status": "Married"
33 }
34 {
35   "_id": "ObjectID('5f8f72e727eba786384dfc61')",
36   "First Name": "Nicholas",
37   "Last Name": "Barrett",
38   "Gender": "Male",
39   "Age": 24,
40   "Email": "n.barrett@randatmail.com",
41   "Education": "Bachelor",
42   "Salary": 5577,
43   "Marital Status": "Married"
44 }

```

Group stage (Sample of 2 documents)

```

1 /**
2  * _id: The id of the group.
3  * fieldN: The first field name.
4  */
5 {
6   "_id": "$Gender",
7   Avg: {$avg: "$Age"},
8   MinAge: {$min: "$Age"},
9   MaxAge: {$max: "$Age"},
10  MaxSalary: {$max: "$Salary"},
11  MinSalary: {$min: "$Salary"},
12  AvgSalary: {$avg: "$Salary"}
13 }
14 }
15 }

```

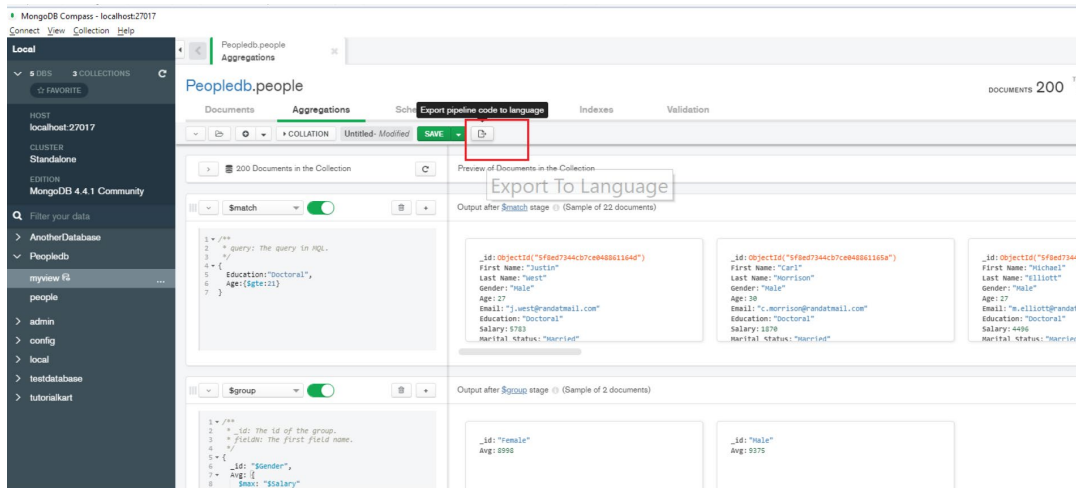
Output after Group stage (Sample of 2 documents)

```

1 {
2   "_id": "Female",
3   "Avg": 25.333333333333332,
4   "MinAge": 21,
5   "MaxAge": 29,
6   "MaxSalary": 8799,
7   "MinSalary": 509,
8   "AvgSalary": 4993.666666666667
9 }
10 {
11   "_id": "Male",
12   "Avg": 25.666666666666668,
13   "MinAge": 22,
14   "MaxAge": 30,
15   "MaxSalary": 9759,
16   "MinSalary": 1268,
17   "AvgSalary": 5252.416666666667
18 }

```

Lastly you can also save these queries that you have just written in to any language i.e. python or node. To do so click the -> button as shown in the figure



Then save the pipe line in Node language as shown in the figure :

Export Pipeline To Language



Note once imported , you can use these code on mongodb shell.

Lab tasks

Write MongoDB queries for the following using either command shell:

1. Repeat the same process to search Education for Master and .Find the avg,min,max age and avg min max Salary of the people group by **Marital status**.
2. find min,max average salary of each age group of female
3. find min,max average salary of each age group of male
4. Count married and unmarried females and males.

After completing this task, write a reflective report summarizing your lab work for today. Include screenshots of the MongoDB queries you performed. Add these to your portfolio for Week 2. Inform your Lab tutor once you have finished the lab. Submit your complete portfolio, covering all weeks (Week 1 to Week 11), by the end of Week 11/