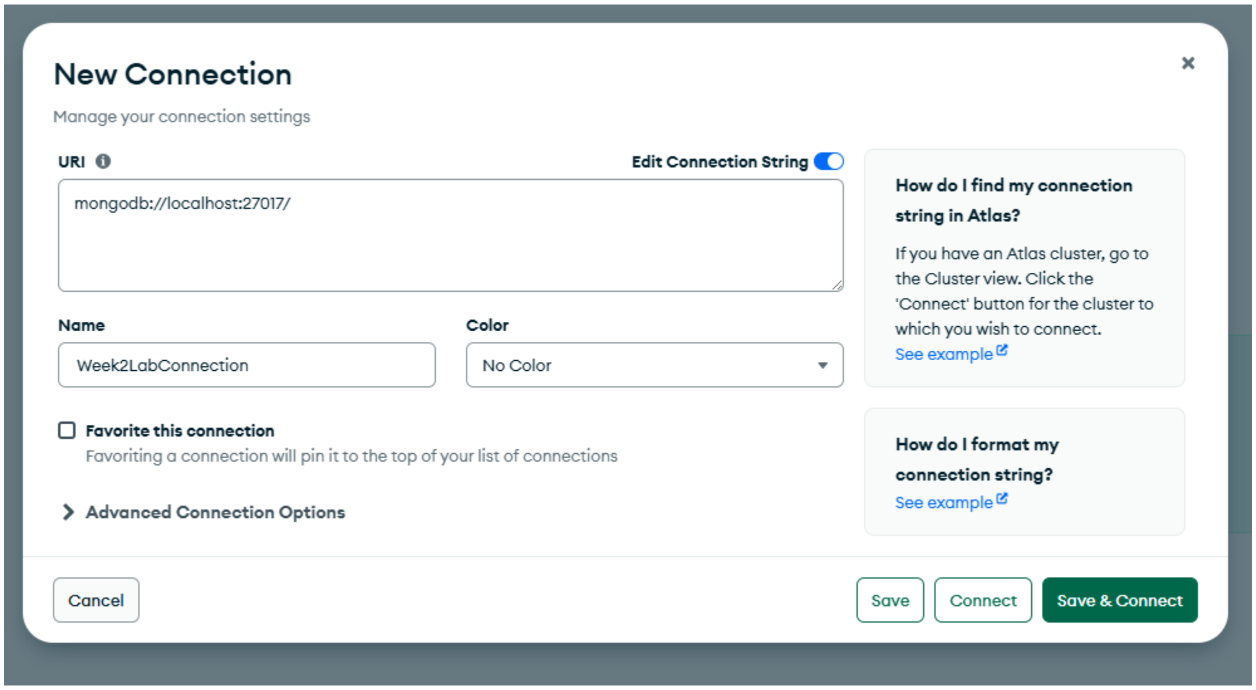
Name : Gaurav kumar Parsaila

Student id : 2833748

Creating the new connection


New Connection

Manage your connection settings

URI

Edit Connection String

mongodb://localhost:27017/

Name: Week2LabConnection

Color: No Color

Favorite this connection

Favoriting a connection will pin it to the top of your list of connections

Advanced Connection Options

How do I find my connection string in Atlas?

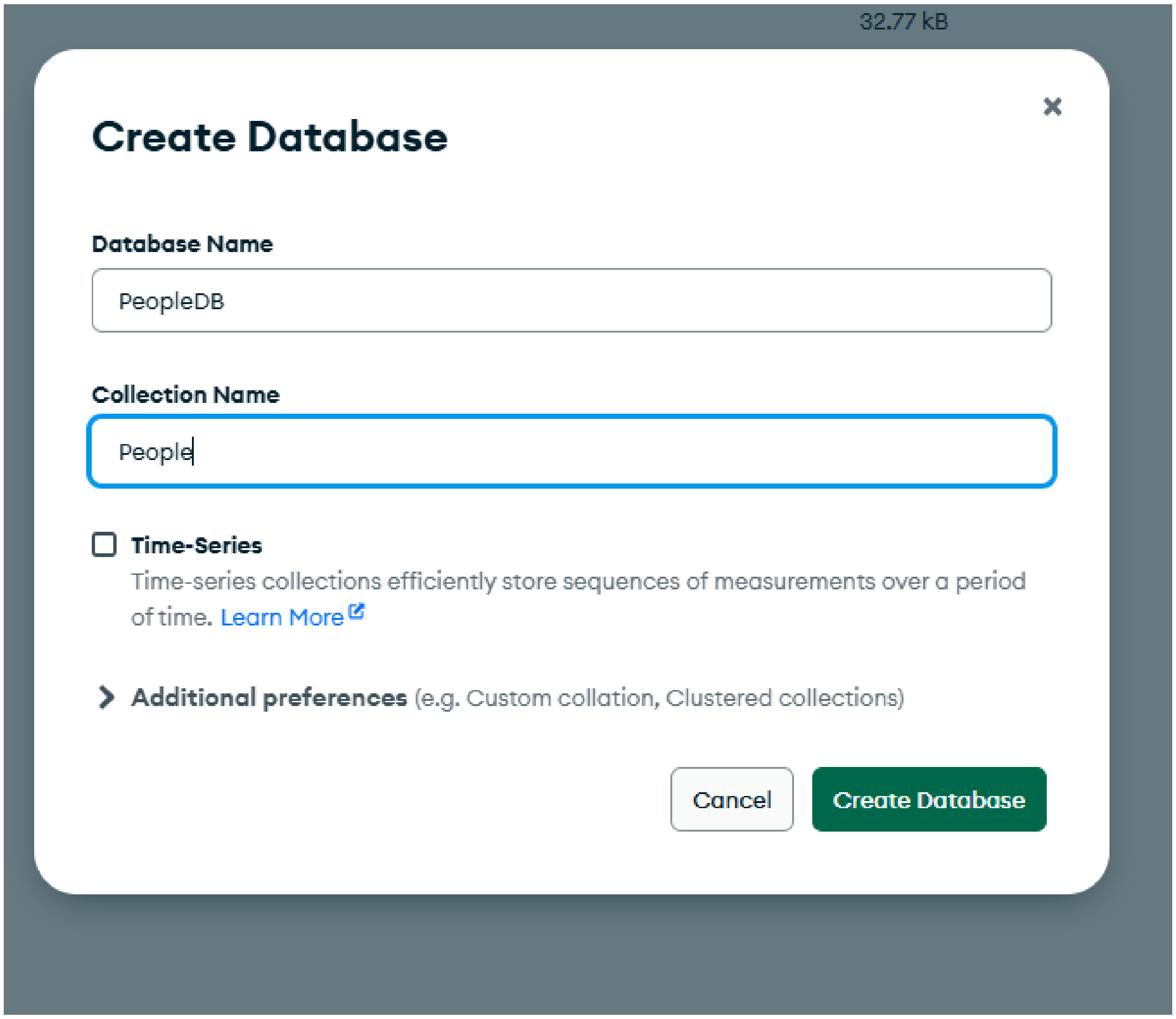
If you have an Atlas cluster, go to the Cluster view. Click the 'Connect' button for the cluster to which you wish to connect.

See example

How do I format my connection string?

See example

Creating the database


32.77 kB

Create Database

Database Name

PeopleDB

Collection Name

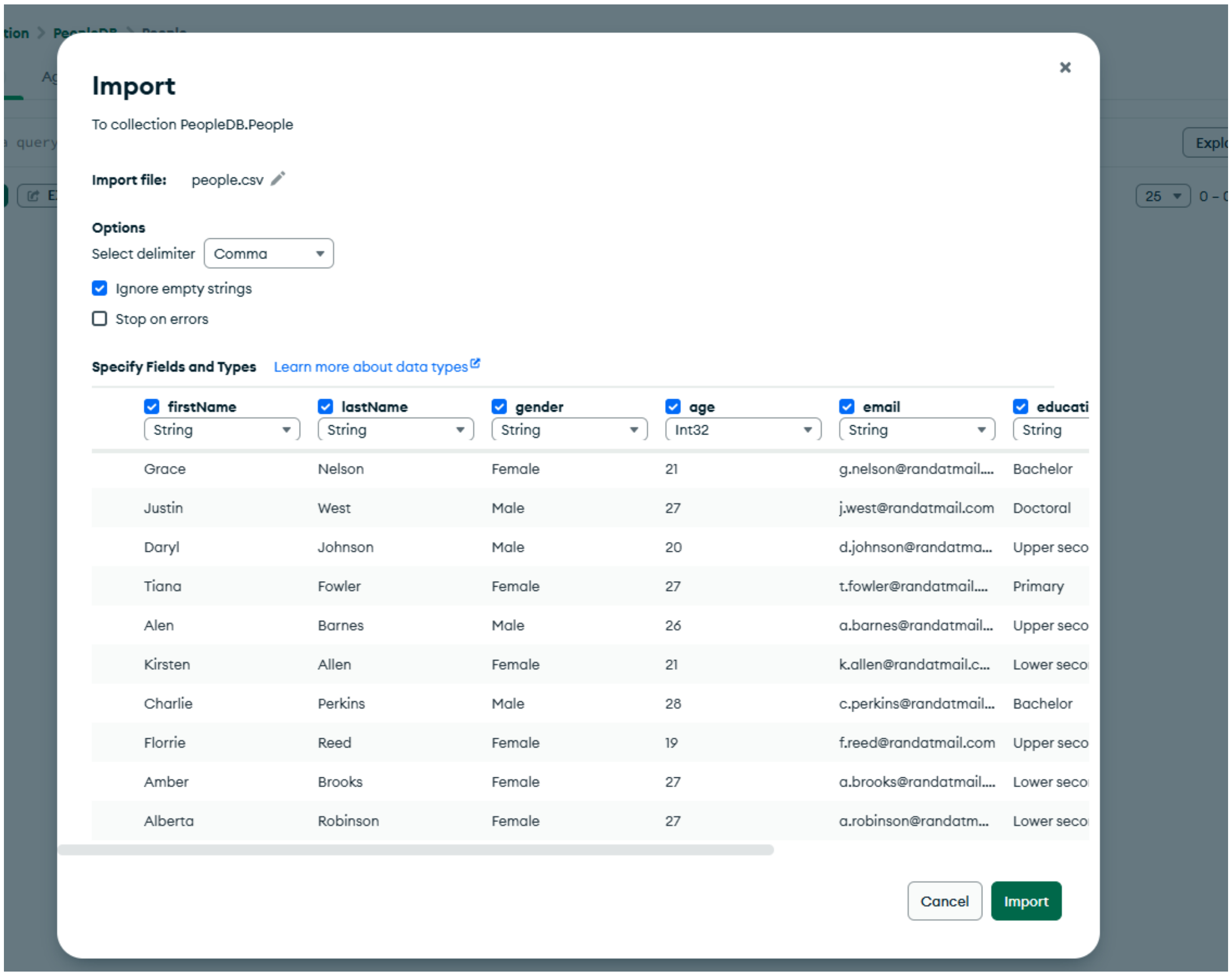
People

☐ Time-Series

Time-series collections efficiently store sequences of measurements over a period of time. Learn More

Additional preferences (e.g. Custom collation, Clustered collections)

Importing the csv file



Import

To collection PeopleDB.People

Import file: people.csv

Options

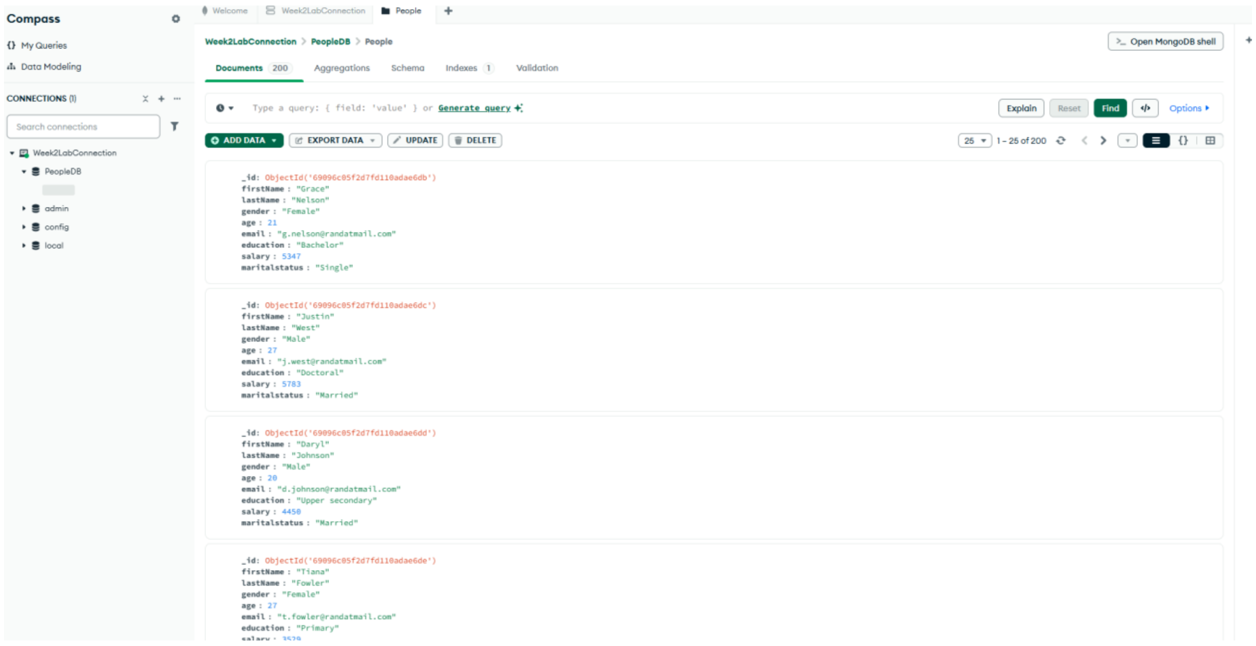
Select delimiter: Comma

☑ Ignore empty strings

☐ Stop on errors

Specify Fields and Types Learn more about data types

☑ firstName String	☑ lastName String	☑ gender String	☑ age Int32	☑ email String	☑ educati String
Grace	Nelson	Female	21	g.nelson@randatmail....	Bachelor
Justin	West	Male	27	j.west@randatmail.com	Doctoral
Daryl	Johnson	Male	20	d.johnson@randatma...	Upper seco
Tiana	Fowler	Female	27	t.fowler@randatmail....	Primary
Alen	Barnes	Male	26	a.barnes@randatmail...	Upper seco
Kirsten	Allen	Female	21	k.allen@randatmail.c...	Lower seco
Charlie	Perkins	Male	28	c.perkins@randatmail...	Bachelor
Florrie	Reed	Female	19	f.reed@randatmail.com	Upper seco
Amber	Brooks	Female	27	a.brooks@randatmail....	Lower seco
Alberta	Robinson	Female	27	a.robinson@randatm...	Lower seco



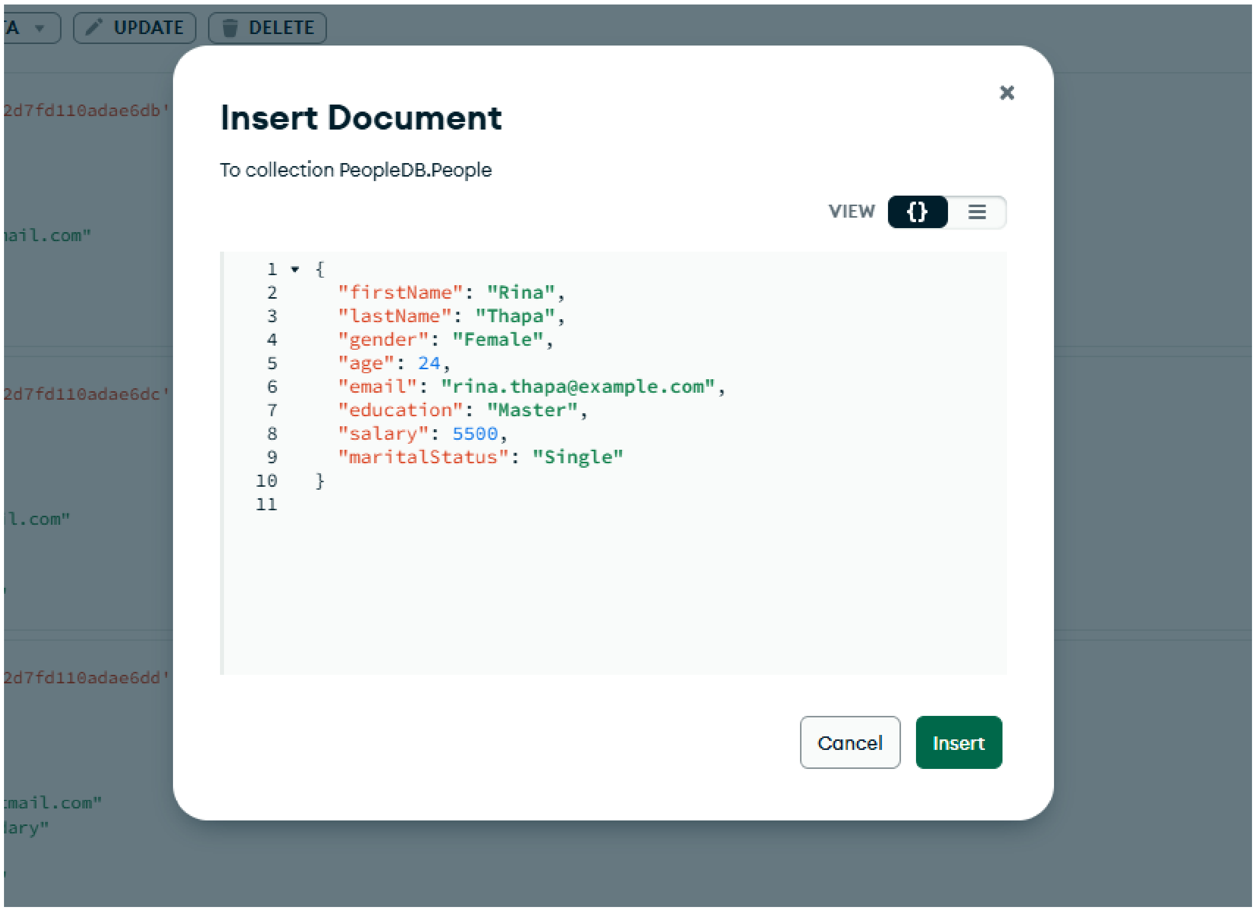
```
_id: ObjectId('69096c05f2d7fd110adae6db')
firstName : "Grace"
lastName : "Nelson"
gender : "Female"
age : 21
email : "g.nelson@randatmail.com"
education : "Bachelor"
salary : 5347
maritalstatus : "Single"
```

```
_id: ObjectId('69096c05f2d7fd110adae6dc')
firstName : "Justin"
lastName : "West"
gender : "Male"
age : 27
email : "j.west@randatmail.com"
education : "Doctoral"
salary : 5783
maritalstatus : "Married"
```

```
_id: ObjectId('69096c05f2d7fd110adae6dd')
firstName : "Daryl"
lastName : "Johnson"
gender : "Male"
age : 20
email : "d.johnson@randatmail.com"
education : "Upper secondary"
salary : 4450
maritalstatus : "Married"
```

```
_id: ObjectId('69096c05f2d7fd110adae6de')
firstName : "Tiana"
lastName : "Fowler"
gender : "Female"
age : 27
email : "t.fowler@randatmail.com"
education : "Primary"
salary : 3529
```

Inserting a new record

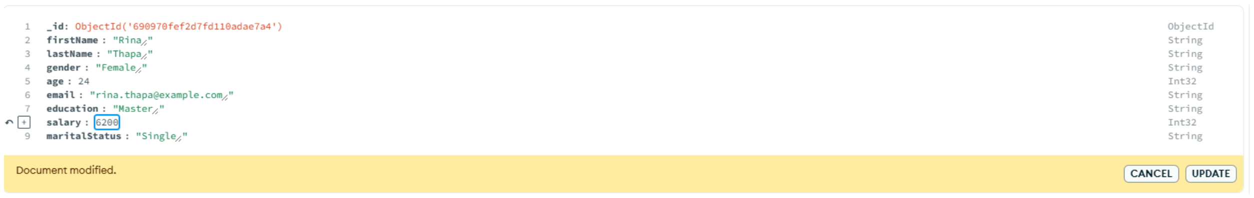


Insert Document

To collection PeopleDB.People

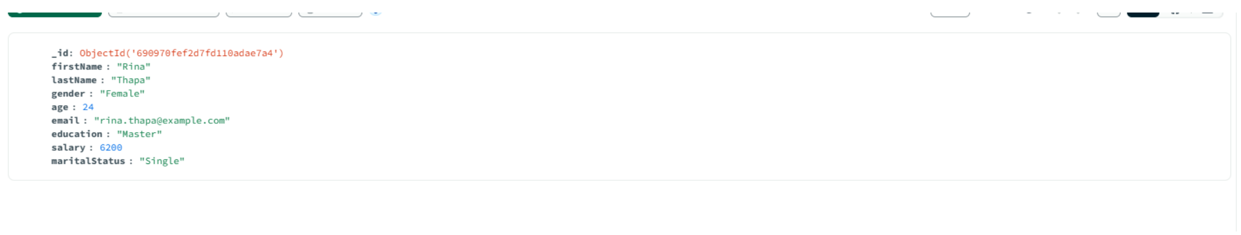
```
{
    "firstName": "Rina",
    "lastName": "Thapa",
    "gender": "Female",
    "age": 24,
    "email": "rina.thapa@example.com",
    "education": "Master",
    "salary": 5500,
    "maritalStatus": "Single"
}
```

Updating the document



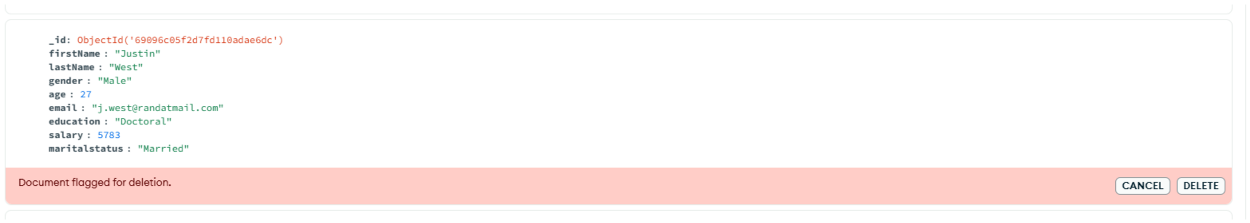
```
_id: ObjectId('690970fef2d7fd110adae7a4')      ObjectId
firstName : "Rina"                             String
lastName : "Thapa"                             String
gender : "Female"                              String
age : 24                                       Int32
email : "rina.thapa@example.com"               String
education : "Master"                           String
salary : 6200                                  Int32
maritalStatus : "Single"                       String
```

Document modified.



```
_id: ObjectId('690970fef2d7fd110adae7a4')
firstName : "Rina"
lastName : "Thapa"
gender : "Female"
age : 24
email : "rina.thapa@example.com"
education : "Master"
salary : 6200
maritalStatus : "Single"
```

Deleting the document

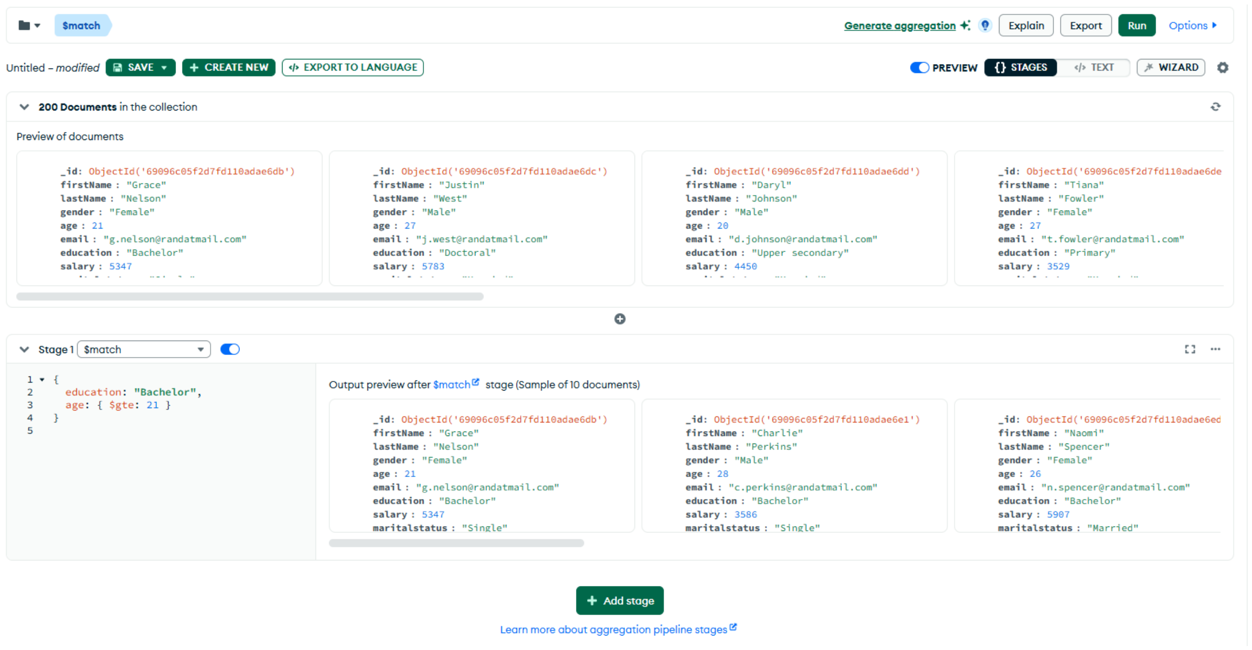


```
_id: ObjectId('69096c05f2d7fd110adae6dc')
firstName : "Justin"
lastName : "West"
gender : "Male"
age : 27
email : "j.west@randatmail.com"
education : "Doctoral"
salary : 5783
maritalstatus : "Married"
```

Document flagged for deletion.

Aggregation tab

Query 1 – Match Bachelor & Age > 21



200 Documents in the collection

Preview of documents

```
_id: ObjectId('69096c05f2d7fd110adae6db')
firstName : "Grace"
lastName : "Nelson"
gender : "Female"
age : 21
email : "g.nelson@randatmail.com"
education : "Bachelor"
salary : 5347
```

```
_id: ObjectId('69096c05f2d7fd110adae6dc')
firstName : "Justin"
lastName : "West"
gender : "Male"
age : 27
email : "j.west@randatmail.com"
education : "Doctoral"
salary : 5783
```

```
_id: ObjectId('69096c05f2d7fd110adae6dd')
firstName : "Daryl"
lastName : "Johnson"
gender : "Male"
age : 20
email : "d.johnson@randatmail.com"
education : "Upper secondary"
salary : 4450
```

```
_id: ObjectId('69096c05f2d7fd110adae6de
firstName : "Tiana"
lastName : "Fowler"
gender : "Female"
age : 27
email : "t.fowler@randatmail.com"
education : "Primary"
salary : 3529
```

Stage 1 $match

```
{
  education: "Bachelor",
  age: { $gte: 21 }
}
```

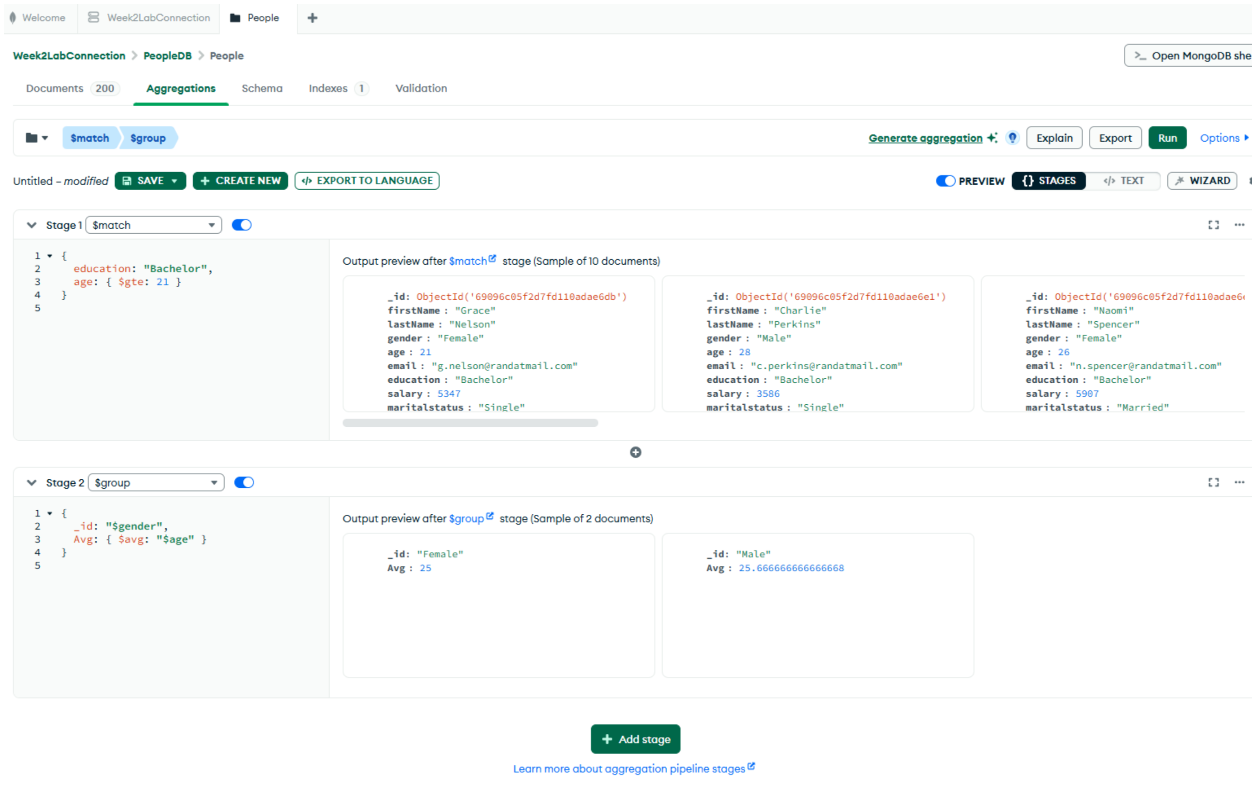
Output preview after $match stage (Sample of 10 documents)

```
_id: ObjectId('69096c05f2d7fd110adae6db')
firstName : "Grace"
lastName : "Nelson"
gender : "Female"
age : 21
email : "g.nelson@randatmail.com"
education : "Bachelor"
salary : 5347
maritalstatus : "Single"
```

```
_id: ObjectId('69096c05f2d7fd110adae6e1')
firstName : "Charlie"
lastName : "Perkins"
gender : "Male"
age : 28
email : "c.perkins@randatmail.com"
education : "Bachelor"
salary : 3586
maritalstatus : "Single"
```

```
_id: ObjectId('69096c05f2d7fd110adae6ed
firstName : "Naomi"
lastName : "Spencer"
gender : "Female"
age : 26
email : "n.spencer@randatmail.com"
education : "Bachelor"
salary : 5907
maritalstatus : "Married"
```

Query 2 – Group by Gender (Avg)



Stage 1 $match

```
{
  education: "Bachelor",
  age: { $gte: 21 }
}
```

Output preview after $match stage (Sample of 10 documents)

```
_id: ObjectId('69096c05f2d7fd110adae6db')
firstName : "Grace"
lastName : "Nelson"
gender : "Female"
age : 21
email : "g.nelson@randatmail.com"
education : "Bachelor"
salary : 5347
maritalstatus : "Single"
```

```
_id: ObjectId('69096c05f2d7fd110adae6e1')
firstName : "Charlie"
lastName : "Perkins"
gender : "Male"
age : 28
email : "c.perkins@randatmail.com"
education : "Bachelor"
salary : 3586
maritalstatus : "Single"
```

```
_id: ObjectId('69096c05f2d7fd110adae6
firstName : "Naomi"
lastName : "Spencer"
gender : "Female"
age : 26
email : "n.spencer@randatmail.com"
education : "Bachelor"
salary : 5907
maritalstatus : "Married"
```

Stage 2 $group

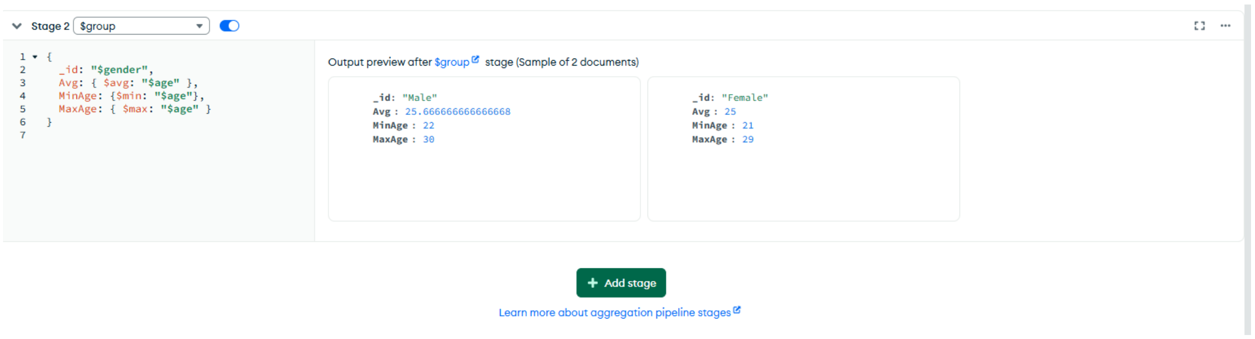
```
{
  _id: "$gender",
  Avg: { $avg: "$age" }
}
```

Output preview after $group stage (Sample of 2 documents)

```
_id: "Female"
Avg : 25
```

```
_id: "Male"
Avg : 25.666666666666668
```

Query 3 – Group by Gender (Min & Max Age)



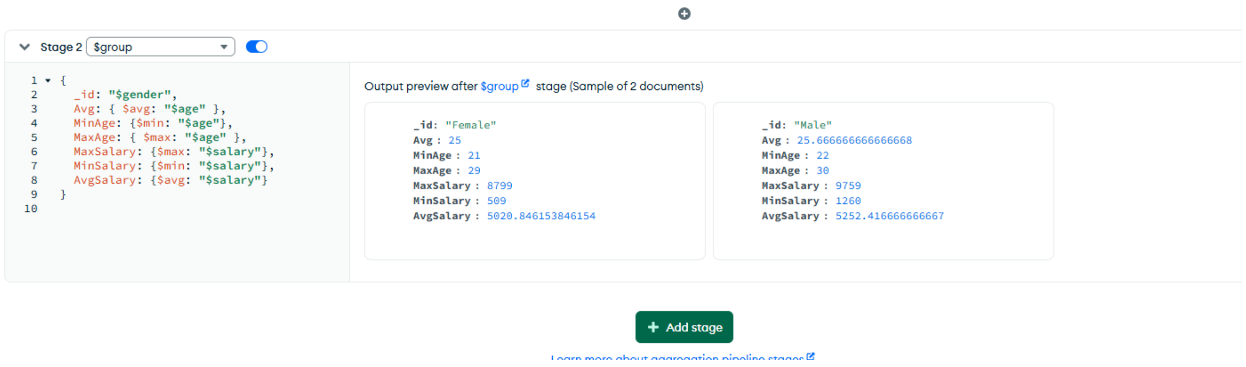
```
{
  _id: "$gender",
  Avg: { $avg: "$age" },
  MinAge: {$min: "$age"},
  MaxAge: { $max: "$age" }
}
```

Output preview after $group stage (Sample of 2 documents)

```
_id: "Male"
Avg : 25.666666666666668
MinAge : 22
MaxAge : 30
```

```
_id: "Female"
Avg : 25
MinAge : 21
MaxAge : 29
```

Query 4 – Group by Gender (Salary Stats)



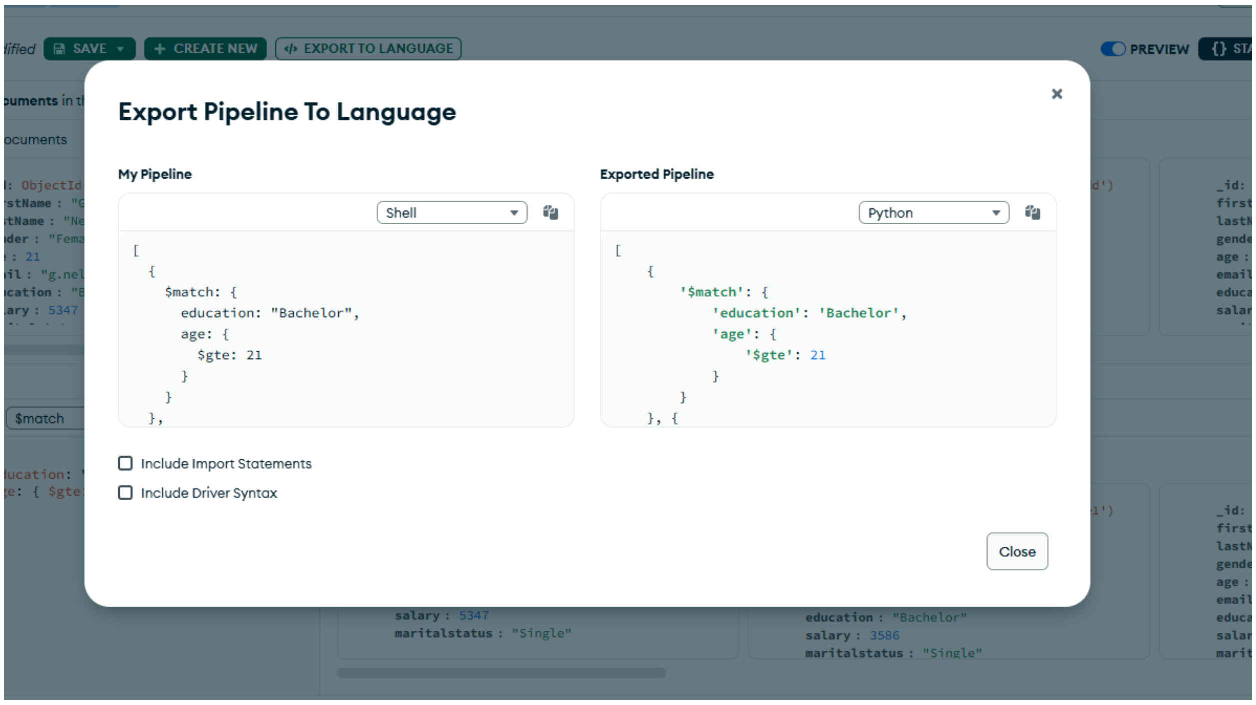
```
{
  _id: "$gender",
  Avg: { $avg: "$age" },
  MinAge: {$min: "$age"},
  MaxAge: { $max: "$age" },
  MaxSalary: {$max: "$salary"},
  MinSalary: {$min: "$salary"},
  AvgSalary: {$avg: "$salary"}
}
```

Output preview after $group stage (Sample of 2 documents)

```
_id: "Female"
Avg : 25
MinAge : 21
MaxAge : 29
MaxSalary : 8799
MinSalary : 509
AvgSalary : 5020.846153846154
```

```
_id: "Male"
Avg : 25.666666666666668
MinAge : 22
MaxAge : 30
MaxSalary : 9759
MinSalary : 1260
AvgSalary : 5252.416666666667
```

Exporting to the language:



Export Pipeline To Language

My Pipeline

Shell

```
[
  {
    $match: {
      education: "Bachelor",
      age: {
        $gte: 21
      }
    }
  },
```

Exported Pipeline

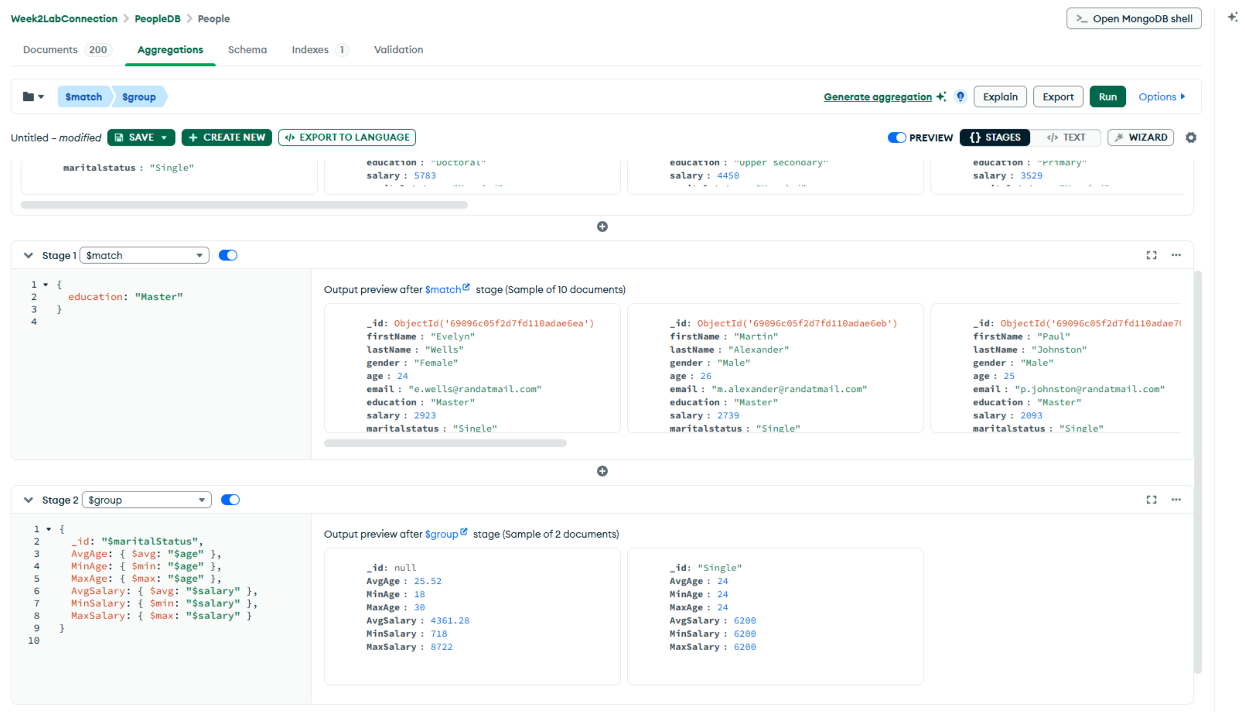
Python

```
[
    {
        '$match': {
            'education': 'Bachelor',
            'age': {
                '$gte': 21
            }
        }
    }, {
```

- [ ] Include Import Statements
- [ ] Include Driver Syntax

Lab Task Queries

Master’s Education -> Group by Marital Status



```
maritalstatus : "Single"

education : "Doctoral"
salary : 5783

education : "Upper secondary"
salary : 4450

education : "Primary"
salary : 3529
```

Stage 1 $match

```
{
  education: "Master"
}
```

Output preview after $match stage (Sample of 10 documents)

```
_id: ObjectId('69096c05f2d7fd110adae6ea')
firstName : "Evelyn"
lastName : "Wells"
gender : "Female"
age : 24
email : "e.wells@randatmail.com"
education : "Master"
salary : 2923
maritalstatus : "Single"

_id: ObjectId('69096c05f2d7fd110adae6eb')
firstName : "Martin"
lastName : "Alexander"
gender : "Male"
age : 26
email : "m.alexander@randatmail.com"
education : "Master"
salary : 2739
maritalstatus : "Single"

_id: ObjectId('69096c05f2d7fd110adae7(
firstName : "Paul"
lastName : "Johnston"
gender : "Male"
age : 25
email : "p.johnston@randatmail.com"
education : "Master"
salary : 2093
maritalstatus : "Single"
```

Stage 2 $group

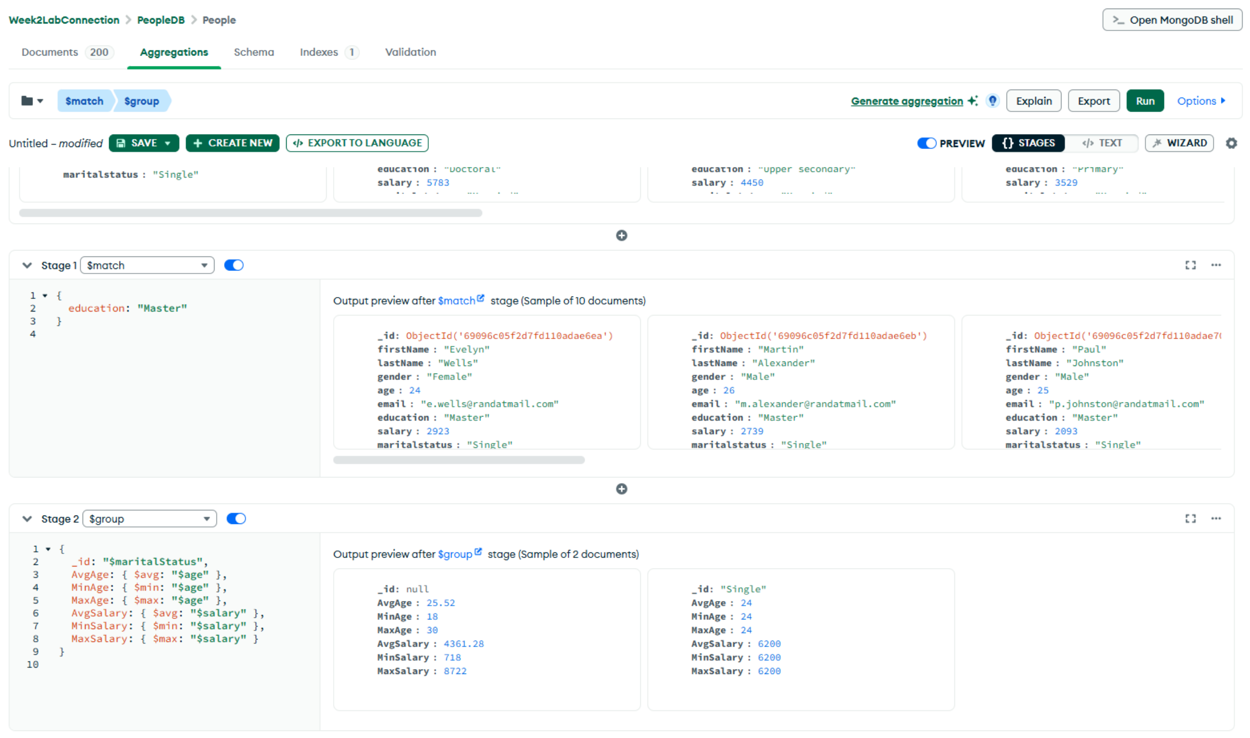
```
{
  _id: "$maritalStatus",
  AvgAge: { $avg: "$age" },
  MinAge: { $min: "$age" },
  MaxAge: { $max: "$age" },
  AvgSalary: { $avg: "$salary" },
  MinSalary: { $min: "$salary" },
  MaxSalary: { $max: "$salary" }
}
```

Output preview after $group stage (Sample of 2 documents)

```
_id: null
AvgAge : 25.52
MinAge : 18
MaxAge : 30
AvgSalary : 4361.28
MinSalary : 718
MaxSalary : 8722

_id: "Single"
AvgAge : 24
MinAge : 24
MaxAge : 24
AvgSalary : 6200
MinSalary : 6200
MaxSalary : 6200
```

Female Salary by Age Group



Week2LabConnection > PeopleDB > People

Stage 1 $match

```
{
  education: "Master"
}
```

Output preview after $match stage (Sample of 10 documents)

```
_id: ObjectId('69096c05f2d7fd110adae6ea')
firstName : "Evelyn"
lastName : "Wells"
gender : "Female"
age : 24
email : "e.wells@randatmail.com"
education : "Master"
salary : 2923
maritalstatus : "Single"
```

```
_id: ObjectId('69096c05f2d7fd110adae6eb')
firstName : "Martin"
lastName : "Alexander"
gender : "Male"
age : 26
email : "m.alexander@randatmail.com"
education : "Master"
salary : 2739
maritalstatus : "Single"
```

```
_id: ObjectId('69096c05f2d7fd110adae7
firstName : "Paul"
lastName : "Johnston"
gender : "Male"
age : 25
email : "p.johnston@randatmail.com"
education : "Master"
salary : 2093
maritalstatus : "Single"
```

Stage 2 $group

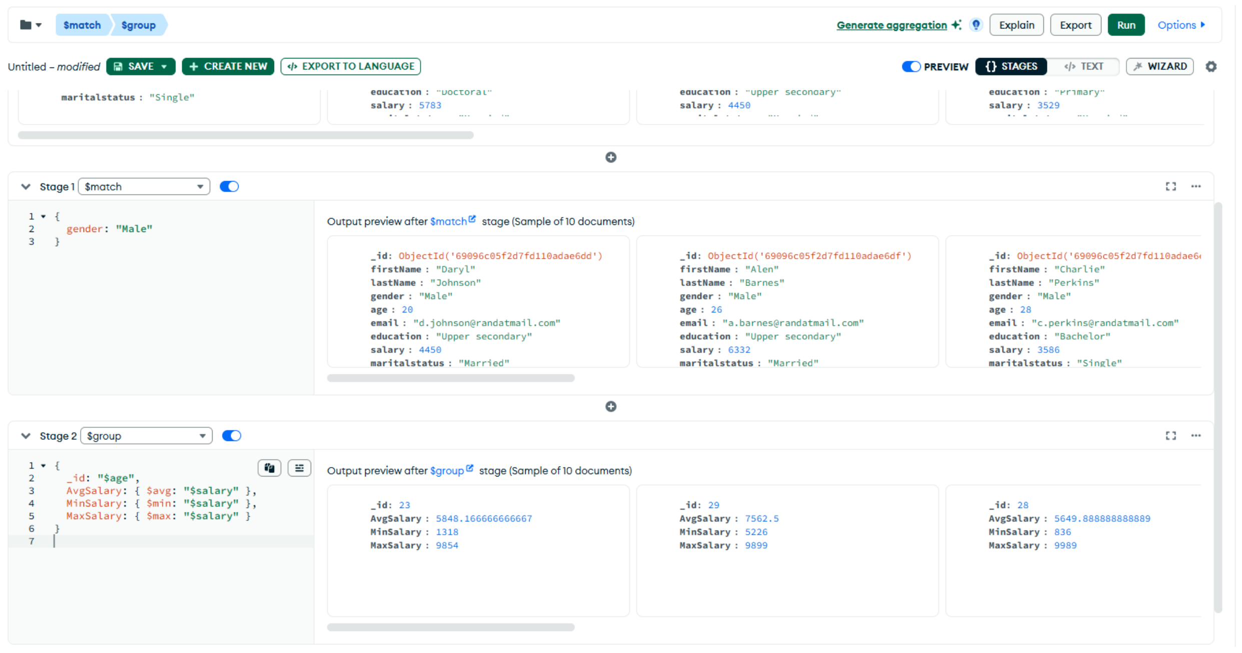
```
{
  _id: "$maritalStatus",
  AvgAge: { $avg: "$age" },
  MinAge: { $min: "$age" },
  MaxAge: { $max: "$age" },
  AvgSalary: { $avg: "$salary" },
  MinSalary: { $min: "$salary" },
  MaxSalary: { $max: "$salary" }
}
```

Output preview after $group stage (Sample of 2 documents)

```
_id: null
AvgAge : 25.52
MinAge : 18
MaxAge : 30
AvgSalary : 4361.28
MinSalary : 718
MaxSalary : 8722
```

```
_id: "Single"
AvgAge : 24
MinAge : 24
MaxAge : 24
AvgSalary : 6200
MinSalary : 6200
MaxSalary : 6200
```

Male Salary by Age Group



Stage 1 $match

```
{
  gender: "Male"
}
```

Output preview after $match stage (Sample of 10 documents)

```
_id: ObjectId('69096c05f2d7fd110adae6dd')
firstName : "Daryl"
lastName : "Johnson"
gender : "Male"
age : 20
email : "d.johnson@randatmail.com"
education : "Upper secondary"
salary : 4450
maritalstatus : "Married"
```

```
_id: ObjectId('69096c05f2d7fd110adae6df')
firstName : "Alen"
lastName : "Barnes"
gender : "Male"
age : 26
email : "a.barnes@randatmail.com"
education : "Upper secondary"
salary : 6332
maritalstatus : "Married"
```

```
_id: ObjectId('69096c05f2d7fd110adae6
firstName : "Charlie"
lastName : "Perkins"
gender : "Male"
age : 28
email : "c.perkins@randatmail.com"
education : "Bachelor"
salary : 3586
maritalstatus : "Single"
```

Stage 2 $group

```
{
  _id: "$age",
  AvgSalary: { $avg: "$salary" },
  MinSalary: { $min: "$salary" },
  MaxSalary: { $max: "$salary" }
}
```

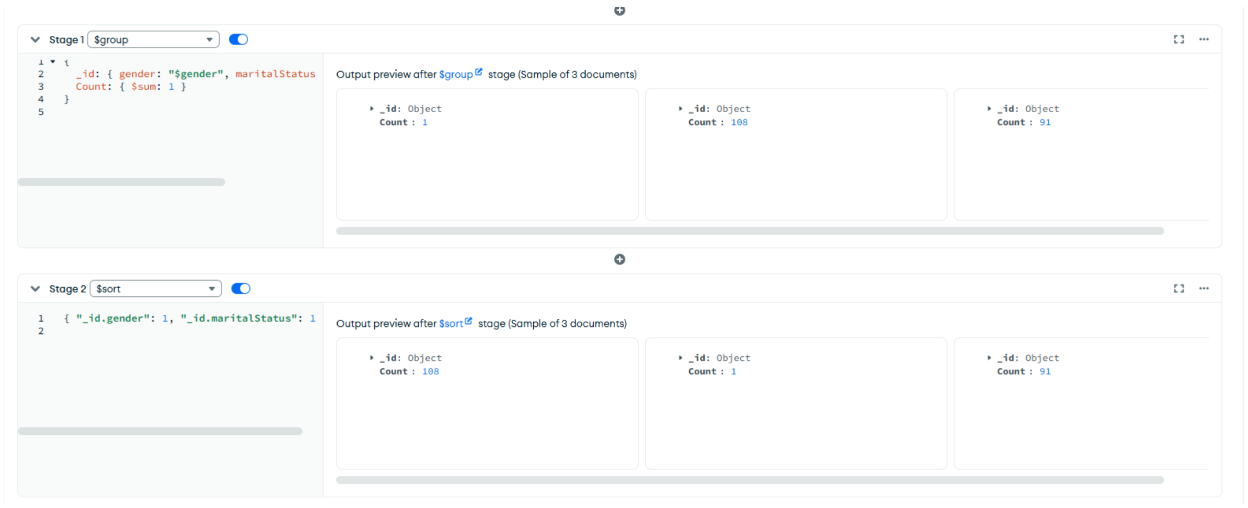
Output preview after $group stage (Sample of 10 documents)

```
_id: 23
AvgSalary : 5848.166666666667
MinSalary : 1318
MaxSalary : 9854
```

```
_id: 29
AvgSalary : 7562.5
MinSalary : 5226
MaxSalary : 9899
```

```
_id: 28
AvgSalary : 5649.888888888889
MinSalary : 836
MaxSalary : 9989
```

Count married and unmarried females and males.



```
{
  _id: { gender: "$gender", maritalStatus
  Count: { $sum: 1 }
}
```

Output preview after $group stage (Sample of 3 documents)

```
_id: Object
Count : 1
```

```
_id: Object
Count : 108
```

```
_id: Object
Count : 91
```

```
{ "_id.gender": 1, "_id.maritalStatus": 1
```

Output preview after $sort stage (Sample of 3 documents)

```
_id: Object
Count : 108
```

```
_id: Object
Count : 1
```

```
_id: Object
Count : 91
```

Reflection Report

The lab this week further equipped me with insight into the functionality and working of MongoDB as a NoSQL database system and how data can be edited graphically using MongoDB Compass. The practical began with the creation of a new database called PeopleDB and a collection called People. The people.csv file was importated and included the records of 200 people with first and last names, age, gender, education, wage, and marital status. In the process, I came to understand the way that MongoDB stores data in flexible documents in the form of JSON as opposed to tables that are used in SQL. This offers unstructured and semi structured storage of data and with this it will be extremely efficient in web and mobile applications of today.

After the successful importation of data, I used the basic CRUD functions, which are Create, Read, Update and Delete. I added one record manually and also edited the salary field of an already existing user and removed another record to see how the changes were going to appear in the Compass interface in real-time. Such operations further improved my knowledge of the way MongoDB handles individual documents without interfering with the remaining part of the collection. It was pretty to observe that each document is automatically assigned a unique ObjectID which is a primary key.

The second part of the lab was concerned with Aggregation Pipelines. I also learned to use functions like $match, $group and $sort to carry out a higher level of analysis of the data directly in MongoDB Compass. As an example, I applied the $match to find those people that have a certain level of education (e.g., “Bachelor or Master) and the $group to obtain aggregate operations, such as an average, minimum, and maximum age or salary. Moreover, I categorized the data according to gender and marital status in order to determine the number of people that were married or single. The visualization features of these computations at a glance in Compass allowed me to gain more insight into the stepwise processing of data in pipelines.

In general, the lab of this week enhanced my technical capabilities of working with NoSQL databases and creating analytical queries. In detail, I have grasped how the document-based model of the MongoDB is different than the relational databases, how the CRUD operations change collections, and how the aggregation pipelines simplify the complex analysis of the data. This will prove to be highly useful in the creation of data-driven web or mobile applications, in which performance, flexibility, and scalability are paramount.