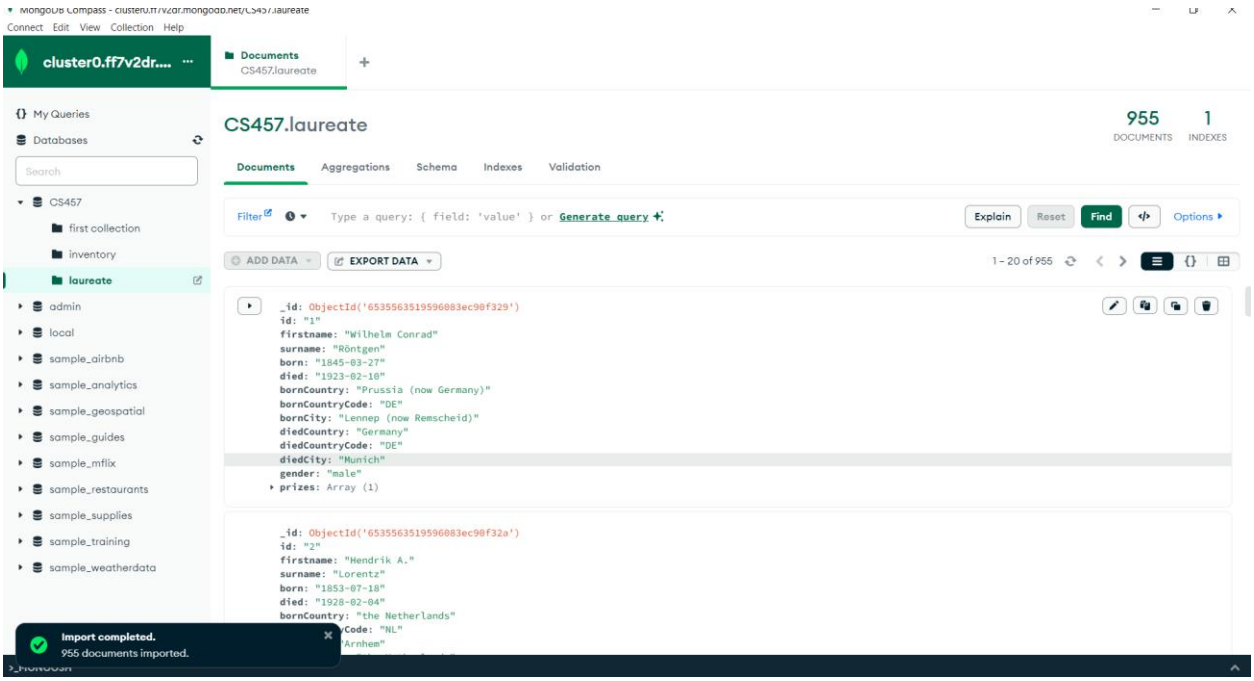
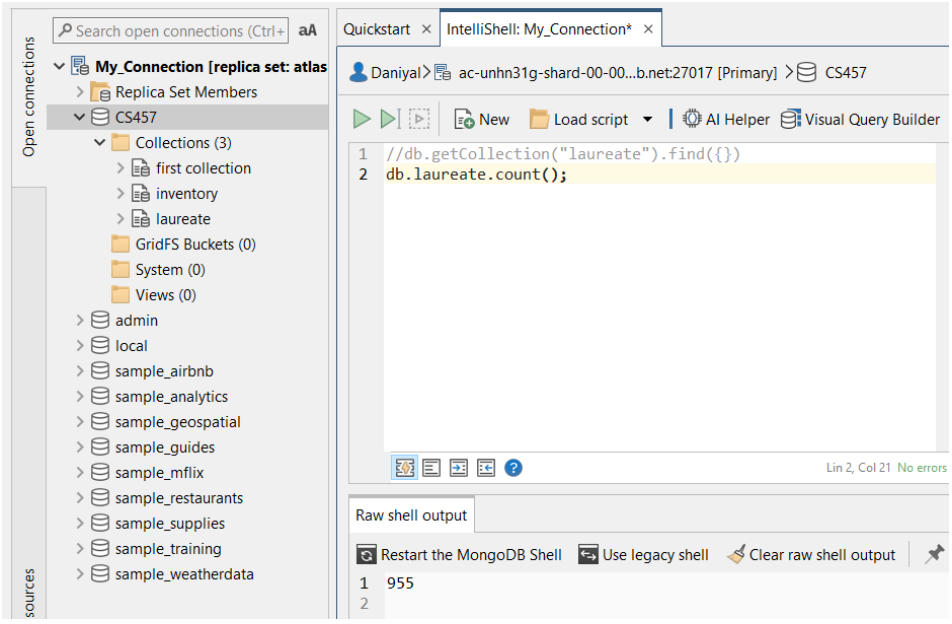
Imported the Dataset into collection using dataset laureate.json



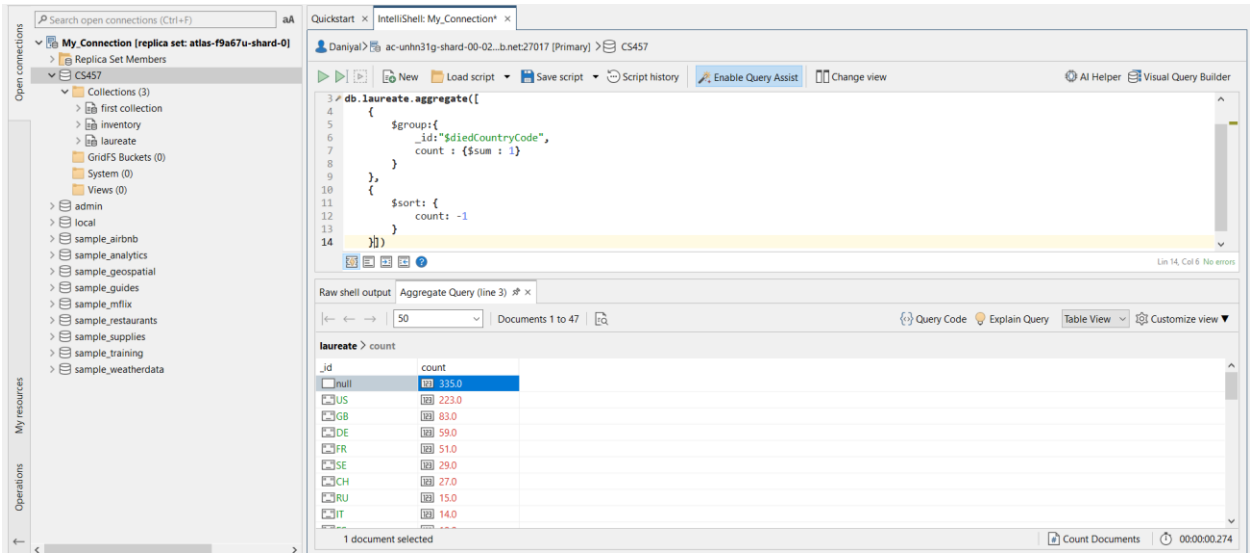
Querying the collection and interpreting the results:

- **The count of total number of records in the collection**



It means that our collection "laureate" has 955 records. A lot of laureates has been passed out. Since this record looks apparently old it may happen that till date over 1000 would have been passed out.

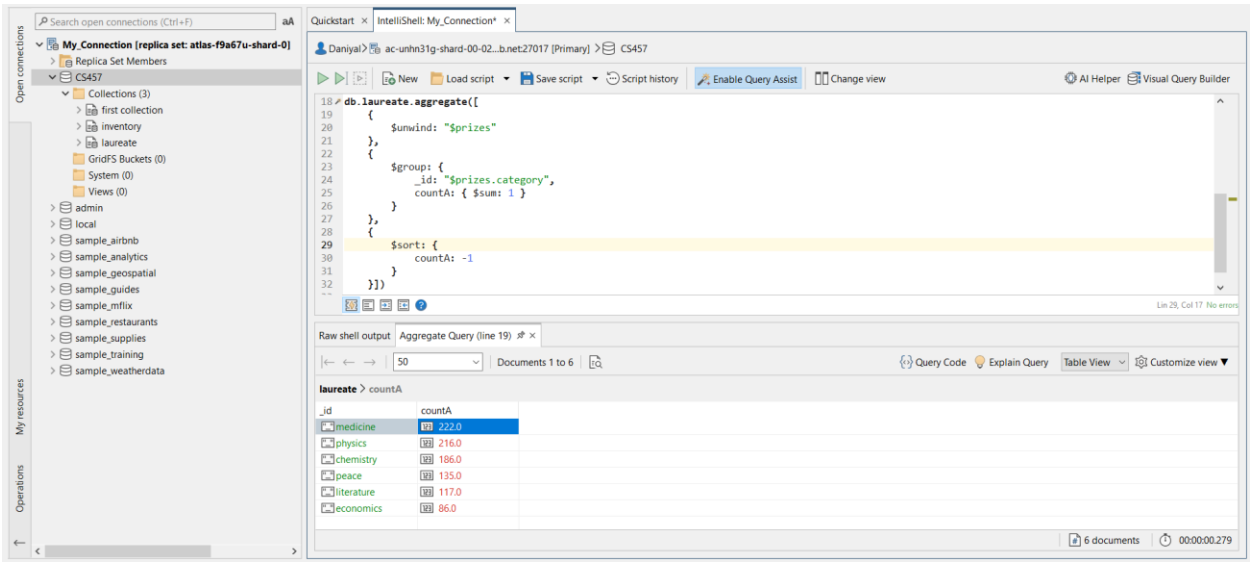
- **the count of records for each diedCountryCode in descending order of count.**



This query has counted the documents according to diedCountryCode and sorted them in descending order. There are a total of 47 unique values for diedCountryCode.

It comes out that 335 laureate holders had no registered diedCountryCode. It means that we have lost the information of the death place for many laureate holders. Below that, we can see that 223 of them died in US, 83 died in GB and so on.

- **the count of records for each prizes.category in descending order of count**

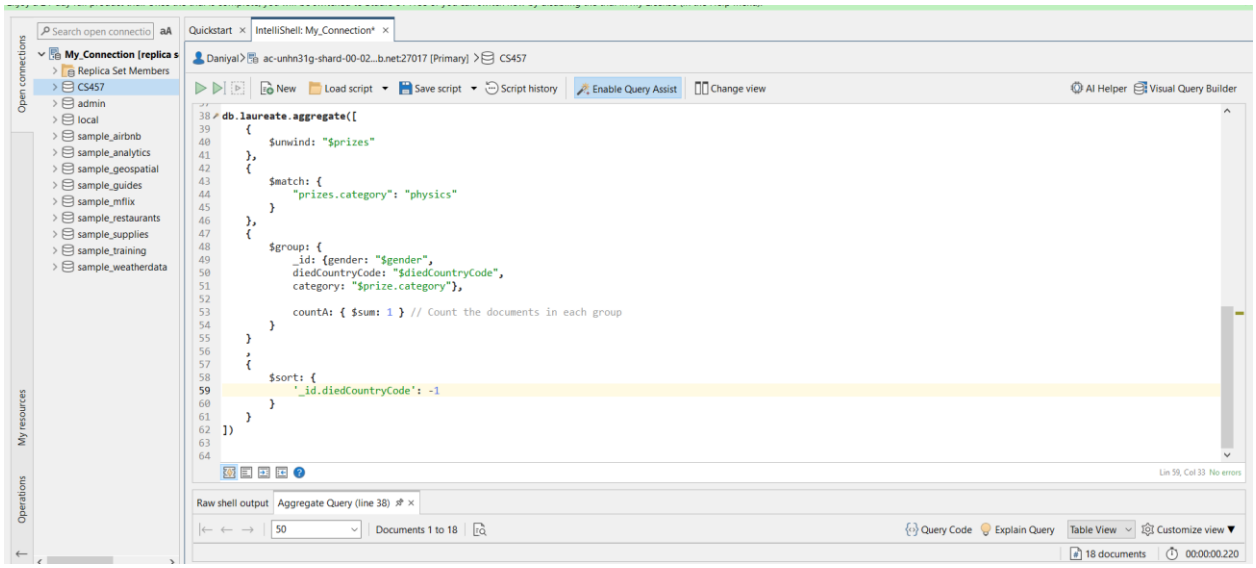


This query has counted the documents for each unique prizes.category. It has then sorted them in descending order. There are a total of 6 unique categories in prizes.

"$unwind" command is used to break every new value in prizes array into separate documents while keeping all other attributes same. It is important to perform action on each element of the array.

One can observe that highest number of prizes have been earned in field of medicine. Therefore we can infer that a large amount of researches are have been conducted in medicine in past. While the least number of awards have been earned in economics. Hence not much scholarship and literature has been written in economics.

- **the count of records for each gender, diedCountryCode, prize.category when prize.category is "physics". Order the output by diedCountryCode.**



laureate

_id	countA
{ 2 fields }	1.0
{ 2 fields }	60.0
{ 2 fields }	4.0
{ 2 fields }	8.0
{ 2 fields }	6.0
{ 2 fields }	4.0
{ 2 fields }	2.0
{ 2 fields }	1.0
{ 2 fields }	18.0
{ 2 fields }	8.0
{ 2 fields }	1.0
{ 2 fields }	2.0
{ 2 fields }	15.0
{ 2 fields }	5.0
{ 2 fields }	2.0
{ 2 fields }	1.0
{ 1 fields }	76.0
{ 1 fields }	2.0

laureate > _id > gender

{Document id}	gender	diedCountryCode
{ 2 fields }	female	US
{ 2 fields }	male	US
{ 2 fields }	male	SE
{ 2 fields }	male	RU
{ 2 fields }	male	NL
{ 2 fields }	male	JP
{ 2 fields }	male	IT
{ 2 fields }	male	IN
{ 2 fields }	male	GB
{ 2 fields }	male	FR
{ 2 fields }	female	FR
{ 2 fields }	male	DK
{ 2 fields }	male	DE
{ 2 fields }	male	CH
{ 2 fields }	male	CA
{ 2 fields }	male	AT
{ 1 fields }	male	
{ 1 fields }	female	

This query has counted the documents for each unique each gender, diedCountryCode, prize.category when prize.category is “physics”. It has then sorted them with respect to diedCountryCode. There are a total of 18 documents.

We can see that no. of laureate females buried in US are much less then no. of laureate males buried in US. We can thus interpret that the no. of women who chose to opt for higher studies and researches in physics were much less than that of men. Moreover, US and FR are only two known countries where both males and females were buried. However, in one country, whose code is unknown, a female was buried. In all other places only males were buried. This again shows that revolutionary physics works done in the past were majorly by males.

- **Come up with your own query to show any interesting insight. Use atleast two fields for match and two fields for group.**

```
db.laureate.aggregate([
    {
        $unwind: "$prizes"
    },
    {
        $match: {
            "prizes.category": "medicine",
            "gender": "male"
        }
    },
    {
        $group: {
            _id: {
                BornCountry: "$bornCountry",
                PrizeShare: "$prizes.share"
            },
            countA: { $sum: 1 }
        }
    },
    {
        $sort: {
                countA: -1
        }
    }
])
```

Raw shell output | Aggregate Query (line 65) ×

50 | Documents 1 to 50 | Query Code | Explain Query | Table View

1 document selected | 74 documents

laureate

_id	countA
{ 2 fields }	39.0
{ 2 fields }	27.0
{ 2 fields }	14.0
{ 2 fields }	10.0
{ 2 fields }	7.0
{ 2 fields }	6.0
{ 2 fields }	5.0
{ 2 fields }	4.0
{ 2 fields }	4.0
{ 2 fields }	4.0
{ 2 fields }	4.0
{ 2 fields }	3.0
{ 2 fields }	3.0
{ 2 fields }	3.0
{ 2 fields }	3.0
{ 2 fields }	3.0
{ 2 fields }	2.0
{ 2 fields }	2.0
{ 2 fields }	2.0
{ 2 fields }	2.0

{Document id}	BornCountry	PrizeShare
{ 2 fields }	USA	3
{ 2 fields }	USA	2
{ 2 fields }	United Kingdom	3
{ 2 fields }	United Kingdom	2
{ 2 fields }	Germany	2
{ 2 fields }	Germany	3
{ 2 fields }	France	3
{ 2 fields }	France	1
{ 2 fields }	Australia	2
{ 2 fields }	Sweden	3
{ 2 fields }	USA	1
{ 2 fields }	Switzerland	1
{ 2 fields }	Italy	3
{ 2 fields }	Austria	3
{ 2 fields }	Canada	2
{ 2 fields }	USA	4
{ 2 fields }	Germany (now ...	1
{ 2 fields }	Sweden	1
{ 2 fields }	Germany	1
{ 2 fields }	Denmark	1

```
{
    "_id" : {
        "BornCountry" : "USA",
        "PrizeShare" : "3"
    },
    "countA" : 39.0
}
{
    "_id" : {
        "BornCountry" : "USA",
        "PrizeShare" : "2"
    },
    "countA" : 27.0
}
{
    "_id" : {
        "BornCountry" : "United Kingdom",
        "PrizeShare" : "3"
    },
    "countA" : 14.0
}
{
    "_id" : {
        "BornCountry" : "United Kingdom",
        "PrizeShare" : "2"
    },
    "countA" : 10.0
}
```

This query has counted the documents for each BornCountry and PrizeShare when prize.category is "medicine" and gender is "male". It has then sorted them with respect to the count. There are a total of 74 documents.

We initially knew that total numbers of laureates are highest in medicine. Moreover we also knew that more laureate holders are among males than females. Keeping these two initial queries In mind I set them as the matching condition. Then I grouped by bornCountry and PrizeShare. And at the end I counted them. This shows that even the highest number of males laureate holder in medicine are from US. 39 of them have contributed in a group of 3 inclusively while 27 of the US laureates has contributed in a group of 2 inclusively. Majorly, US has produced highest bulk of laureates in the past.