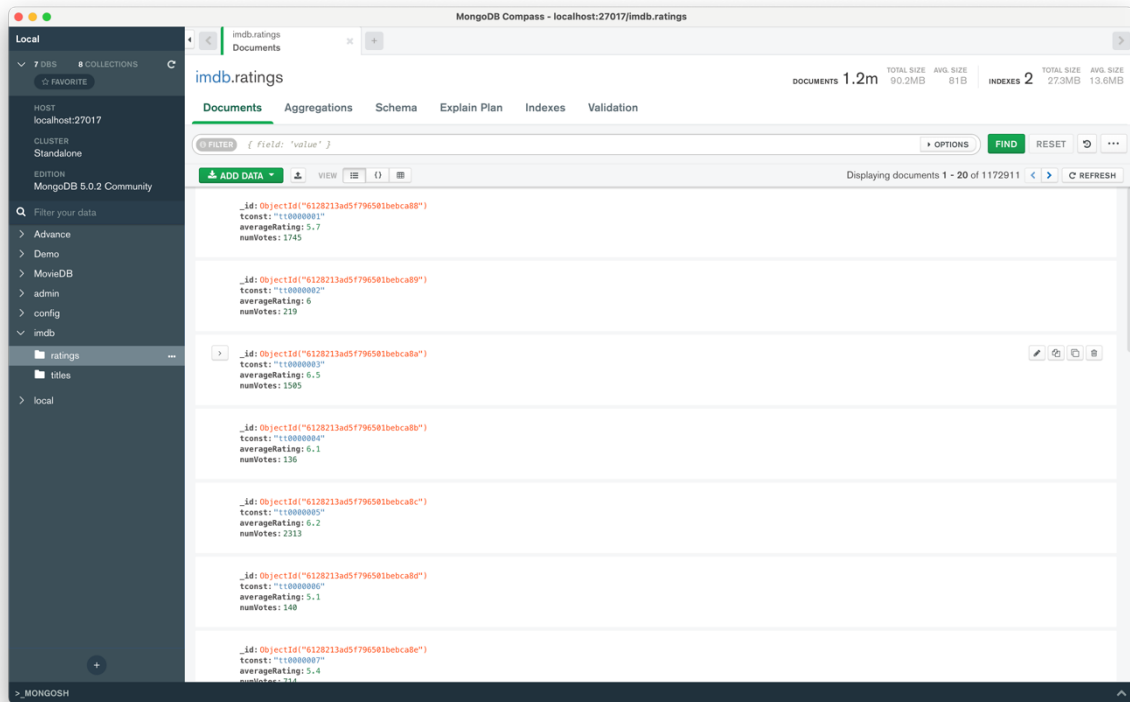
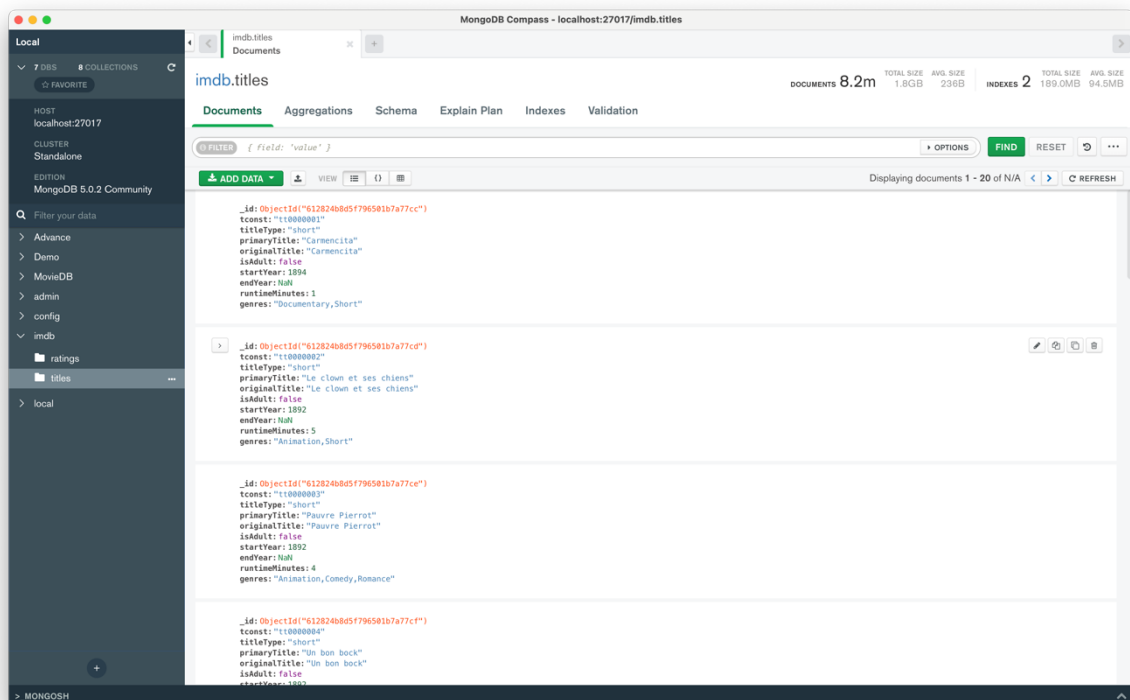


S5196431  
Duwon HA

Task 1 & 2 and 4



Collection ratings



Collection titles

## Task 3

### Schema Validation for titles

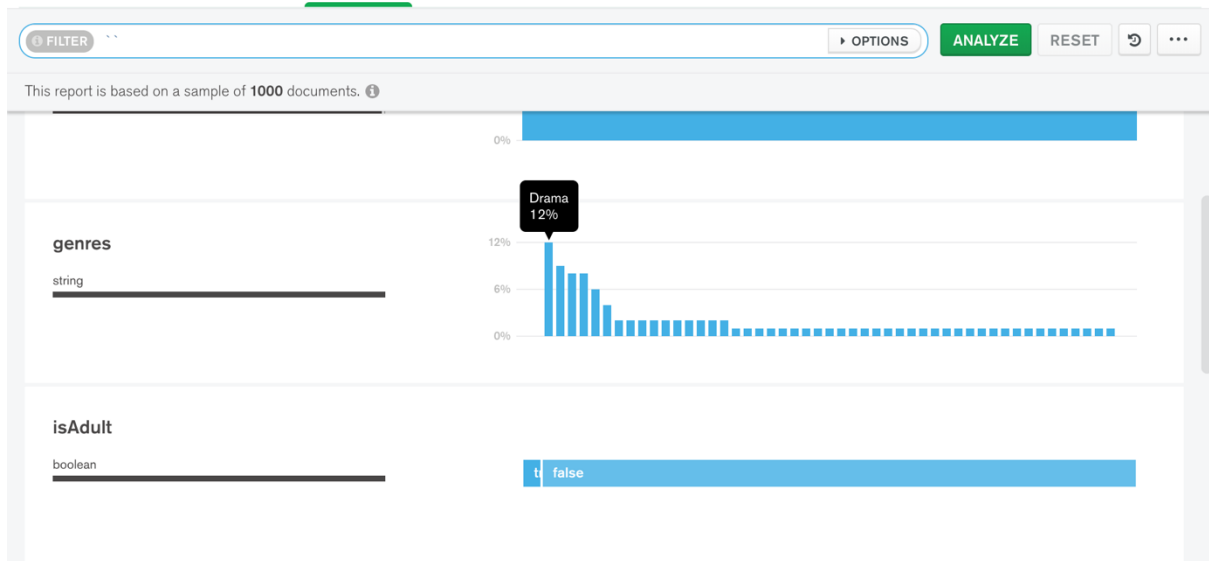
```
1 {
2   $jsonSchema: {
3     bsonType: 'object',
4     required: [
5       'tconst',
6       'primaryTitle'
7     ],
8     properties: {
9       tconst: {
10        bsonType: 'string',
11        description: 'alphanumeric unique identifier of the title'
12      },
13       titleType: {
14        bsonType: 'string',
15        description: 'the type/format of the title (e.g., movie, short, tvseries, tvepisode, video, etc)'
16      },
17       primaryTitle: {
18        bsonType: 'string',
19        description: 'the more popular title / the title used by the filmmakers on promotional materials at the point of release'
20      },
21       originalTitle: {
22        bsonType: 'string',
23        description: 'original title, in the original language'
24      },
25       isAdult: {
26        bsonType: 'bool',
27        description: '0: non-adult title; 1: adult title'
28      },
29       startYear: {
30        bsonType: 'number',
31        description: 'represents the release year of a title. In the case of TV Series, it is the series start year'
32      },
33       endYear: {
34        bsonType: 'number',
35        description: 'TV Series end year. \'\N\' for all other title types'
36      },
37       runtimeMinutes: {
38        bsonType: 'number',
39        description: 'primary runtime of the title, in minutes'
40      },
41       genres: {
42        bsonType: 'string',
43        description: 'includes up to three genres associated with the title'
44      }
45     }
46   }
47 }
```

## Schema Validation for ratings

```
1 {  
2   $jsonSchema: {  
3     bsonType: 'object',  
4     required: [  
5       'tconst',  
6       'averageRating'  
7     ],  
8     properties: {  
9       tconst: {  
10        bsonType: 'string',  
11        description: 'alphanumeric unique identifier of the title'  
12      },  
13      averageRating: {  
14        bsonType: 'number',  
15        description: 'weighted average of all the individual user ratings'  
16      },  
17      numVotes: {  
18        bsonType: 'number',  
19        description: 'number of votes the title has received'  
20      }  
21    }  
22  }  
23 }
```

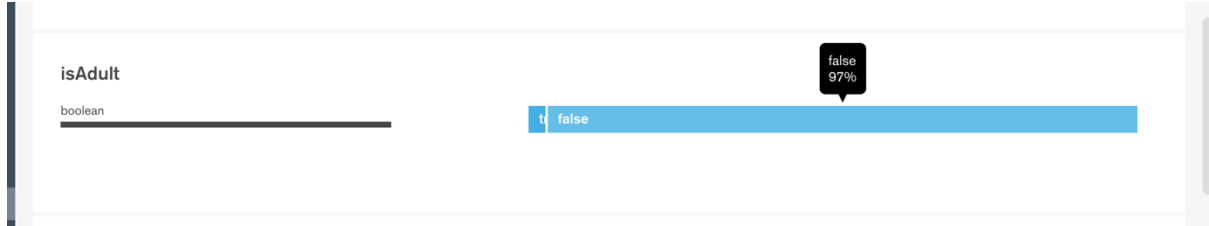
## Task 5

1)



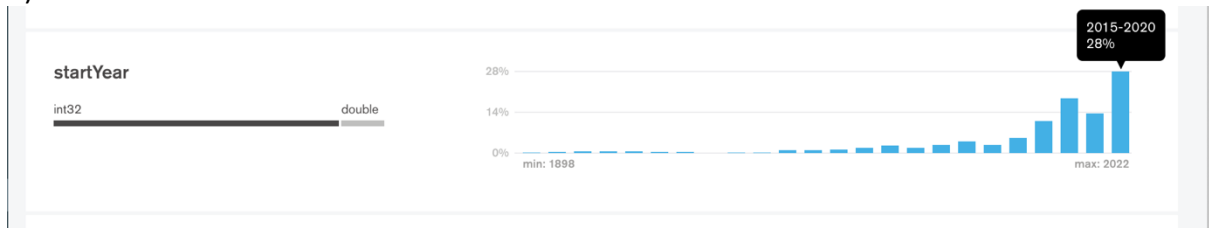
Genre of Drama covers the highest number of movies which is 12%, and followed by Documentary 9%, Talk show and Comedy are 8%.

2)



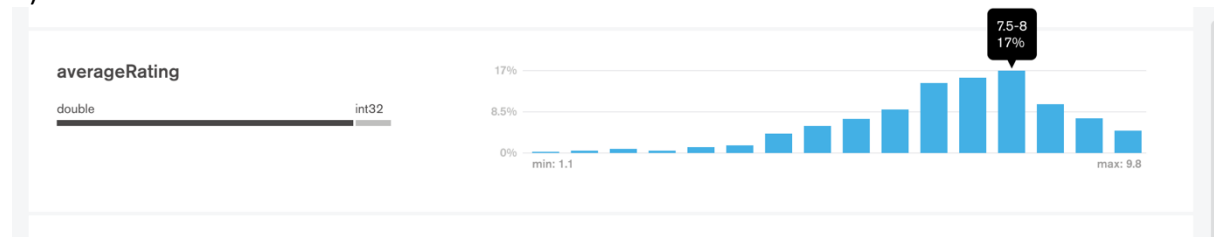
97% of movies are not an adult movie. Only 3% of movies are adult movie which is interesting

3)



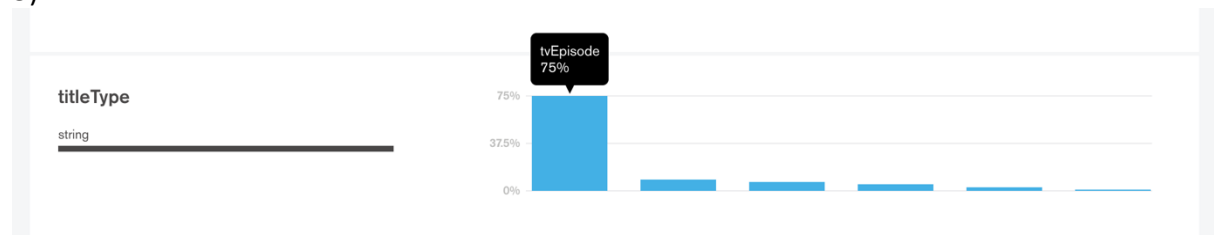
We can see that the film industry has continued to develop, however it decreased in 2010 to 2015 but reached to peak in 2015 to 2020 which is 28%.

4)



The average movie rating of 7.5 to 8 covers the highest number, followed by 6.5-7 and 7-7.5. Rating of 2.5 – 3 covers the lowest number. Most of the movies got over average movie rating 5 out of 10.

5)



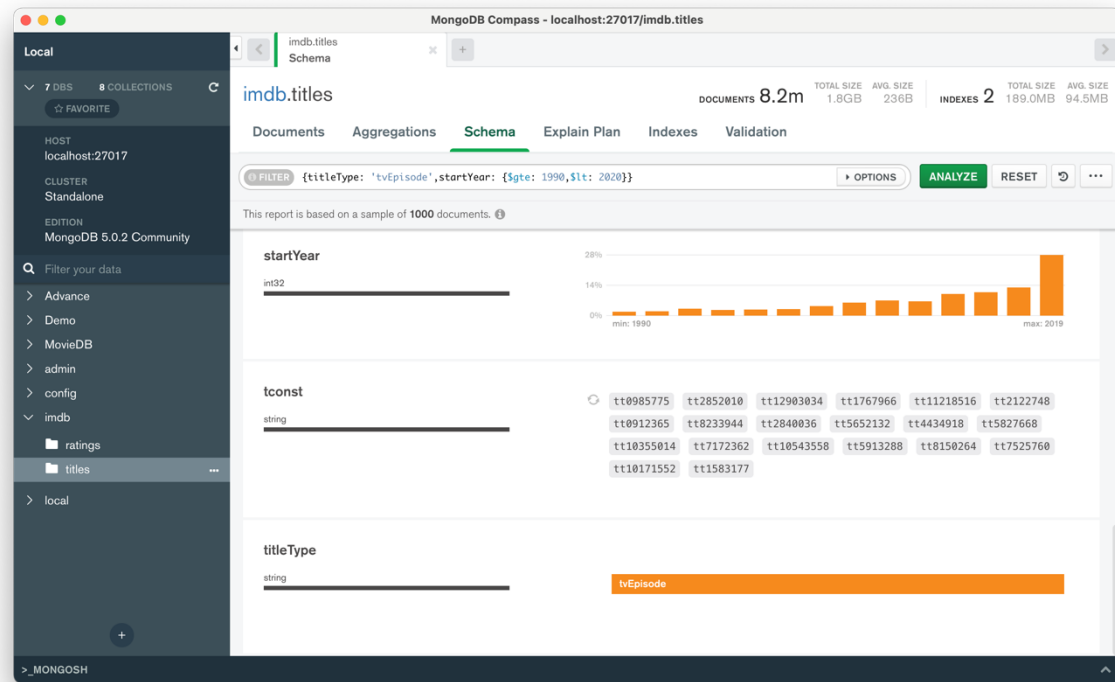
Tv episode covers the highest number of title type, followed by short and movie.

6)



15% of the movies got greater or equal to 10500 and less than 11000 votes and which is the highest rate. We can know that generally voting rate is low.

7)



We can see that tv Episode became super popular in 2016 to 2018. It seems to be caused by the popularisation of streaming services such as Netflix worldwide.

## Task 6

A)

MongoDB Compass - localhost:27017/imdb.movies

imdb.movies

DOCUMENTS 8.2m TOTAL SIZE 1.8GB AVG. SIZE 236B INDEXES 1 TOTAL SIZE 86.7MB AVG. SIZE 86.7MB

Documents Aggregations Schema Explain Plan Indexes Validation

COLLATION: Untitled - Modified SAVE SAMPLE MODE: ON AUTO PREVIEW: ON

8177542 Documents in the Collection Preview of Documents in the Collection

Select an operator to construct expressions used in the aggregation pipeline stages. [Learn more](#)

1. \$sortByCount

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

```
1 // **
2 * expression: Grouping expression or string.
3 //
4
5 "$startYear"
6
```

ADD STAGE

Sample documents:

- `{ "_id": ObjectId("612824b8d5f796501b7a77cc"), "tconst": "t10000001", "titleType": "short", "primaryTitle": "Carmencita", "isAdult": false, "startYear": 1894, "endYear": null, "runtimeMinutes": 1 }`
- `{ "_id": ObjectId("612824b8d5f796501b7a77cd"), "tconst": "t10000002", "titleType": "short", "primaryTitle": "Le clown et ses chiens", "originalTitle": "Le clown et ses chiens", "isAdult": false, "startYear": 1892, "endYear": null, "runtimeMinutes": 5 }`
- `{ "_id": 2018, "count": 480654 }`

B)

MongoDB Compass - localhost:27017/imdb.movies

imdb.movies

DOCUMENTS 8.2m TOTAL SIZE 1.8GB AVG. SIZE 236B INDEXES 1 TOTAL SIZE 86.7MB AVG. SIZE 86.7MB

Documents Aggregations Schema Explain Plan Indexes Validation

COLLATION: Assignment1 - Modified SAVE SAMPLE MODE: ON AUTO PREVIEW: ON

8177542 Documents in the Collection Preview of Documents in the Collection

Select an operator to construct expressions used in the aggregation pipeline stages. [Learn more](#)

1. \$lookup

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

```
1 {
2   $from: 'ratings',
3   $localField: 'tconst',
4   $foreignField: 'tconst',
5   $as: 'rating'
6 }
```

Sample documents:

- `{ "titleType": "short", "primaryTitle": "Carmencita", "originalTitle": "Carmencita", "isAdult": false, "startYear": 1894, "endYear": null, "runtimeMinutes": 1, "genres": "Documentary,Short", "rating": Array }`
- `{ "_id": ObjectId("612824b8d5f796501b7a77cd"), "tconst": "t10000002", "titleType": "short", "primaryTitle": "Le clown et ses chiens", "originalTitle": "Le clown et ses chiens", "isAdult": false, "startYear": 1892, "endYear": null, "runtimeMinutes": 5 }`
- `{ "_id": 0, "tconst": "t10000002", "titleType": "short", "primaryTitle": "Le clown et ses chiens", "originalTitle": "Le clown et ses chiens", "isAdult": false, "startYear": 1892, "endYear": null, "runtimeMinutes": 5 }`

2. \$unwind

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

```
1 '$rating'
```

Sample documents:

- `{ "_id": ObjectId("612824b8d5f796501b7a77cc"), "tconst": "t10000001", "titleType": "short", "primaryTitle": "Carmencita", "originalTitle": "Carmencita", "isAdult": false, "startYear": 1894, "endYear": null, "runtimeMinutes": 1 }`
- `{ "_id": ObjectId("612824b8d5f796501b7a77cd"), "tconst": "t10000002", "titleType": "short", "primaryTitle": "Le clown et ses chiens", "originalTitle": "Le clown et ses chiens", "isAdult": false, "startYear": 1892, "endYear": null, "runtimeMinutes": 5 }`
- `{ "_id": 0, "tconst": "t10000002", "titleType": "short", "primaryTitle": "Le clown et ses chiens", "originalTitle": "Le clown et ses chiens", "isAdult": false, "startYear": 1892, "endYear": null, "runtimeMinutes": 5 }`

1- {  
2 Rating: '\$rating.averageRating',  
3 primaryTitle: 1,  
4 genres: 1,  
5 startYear: 1,  
6 \_id: 0  
7 }  
8

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

primaryTitle: "Carmencita"  
startYear: 1894  
genres: "Documentary, Short"  
Rating: 5.7

primaryTitle: "Le clown et ses chiens"  
startYear: 1892  
genres: "Animation, Short"  
Rating: 6

primaryTitle: "Pauvre Pierrot"  
startYear: 1892  
genres: "Animation, Comedy, Romance"  
Rating: 6.5

1- {  
2 \$and:  
3 [  
4 {genres: {\$regex: "Fantasy"}},  
5 {genres: {\$regex: "Adventure"}},  
6 {startYear: 2021}  
7 }  
8 ]  
9 }  
10  
11  
12  
13  
14  
15

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

primaryTitle: "Mortal Kombat"  
startYear: 2021  
genres: "Action, Adventure, Fantasy"  
Rating: 6.1

primaryTitle: "Realm of Terracotta"  
startYear: 2021  
genres: "Adventure, Animation, Fantasy"  
Rating: 7.6

primaryTitle: "Glorious Purpose"  
startYear: 2021  
genres: "Action, Adventure, Fantasy"  
Rating: 8.8

1- {  
2 Rating: -1  
3 }  
4

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

primaryTitle: "Long May He Reign"  
startYear: 2021  
genres: "Adventure, Fantasy"  
Rating: 9.4

primaryTitle: "The Nexus Event"  
startYear: 2021  
genres: "Action, Adventure, Fantasy"  
Rating: 9.2

primaryTitle: "Journey Into Mystery"  
startYear: 2021  
genres: "Action, Adventure, Fantasy"  
Rating: 9.1

1- {  
2 Rating: -1  
3 }  
4

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

primaryTitle: "Long May He Reign"  
startYear: 2021  
genres: "Adventure, Fantasy"  
Rating: 9.4

primaryTitle: "The Nexus Event"  
startYear: 2021  
genres: "Action, Adventure, Fantasy"  
Rating: 9.2

primaryTitle: "Journey Into Mystery"  
startYear: 2021  
genres: "Action, Adventure, Fantasy"  
Rating: 9.1

1 5

Output after \$limit stage (Sample of 5 documents)

primaryTitle: "Long May He Reign"  
startYear: 2021  
genres: "Adventure, Fantasy"  
Rating: 9.4

primaryTitle: "The Nexus Event"  
startYear: 2021  
genres: "Action, Adventure, Fantasy"  
Rating: 9.2

primaryTitle: "Journey Into Mystery"  
startYear: 2021  
genres: "Action, Adventure, Fantasy"  
Rating: 9.1

ADD STAGE



C)

MongoDB Compass - localhost:27017/imdb.title

imdb.title

Documents Aggregations Schema Explain Plan Indexes Validation

8177542 Documents in the Collection

Preview of Documents in the Collection

Select an operator to construct expressions used in the aggregation pipeline stages. [Learn more](#)

1- \$lookup

2- \$unwind

3- \$sum

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

1- {

2- from: 'ratings',

3- localField: 'tconst',

4- foreignField: 'tconst',

5- as: 'rating'

6- }

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

1- "\$rating"

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

1- {

2- primaryTitle: 1,

3- startYear: 1,

4- genres: 1,

5- votesNum: '\$rating.numVotes',

6- \_id: 0

7- }

primaryTitle: "Carmencita"

startYear: 1894

genres: "Documentary,Short"

votesNum: 1745

primaryTitle: "Le clown et ses chiens"

startYear: 1892

genres: "Animation,Short"

votesNum: 219

primaryTitle: "Pauvre Pierrot"

startYear: 1892

genres: "Animation,Comedy,Romance"

votesNum: 1505

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

1- {

2- \$and:

3- [

4- {genres: {\$regex: "Fantasy"}},

5- {genres: {\$regex: "Adventure"}},

6- {startYear: 2021}

7- ]

8- }

9- }

10- }

primaryTitle: "Mortal Kombat"

startYear: 2021

genres: "Action,Adventure,Fantasy"

votesNum: 126530

primaryTitle: "Realm of Terracotta"

startYear: 2021

genres: "Adventure,Animation,Fantasy"

votesNum: 21

primaryTitle: "Glorious Purpose"

startYear: 2021

genres: "Action,Adventure,Fantasy"

votesNum: 24510

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

1- /\*\*

2- \* Provide any number of field/order pairs.

3- \*/

4- {

5- votesNum: -1

6- }

primaryTitle: "Zack Snyder's Justice League"

startYear: 2021

genres: "Action,Adventure,Fantasy"

votesNum: 316048

primaryTitle: "Loki"

startYear: 2021

genres: "Action,Adventure,Fantasy"

votesNum: 135726

primaryTitle: "Mortal Kombat"

startYear: 2021



genres: "Action,Adventure,Fantasy"

votesNum: 126530

|||

\$sort

☒

```
1 //**
2 * Provide any number of field/order pairs.
3 */
4 {
5   votesNum : -1
6 }
```

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

primaryTitle: "Zack Snyder's Justice League"

startYear: 2021

genres: "Action,Adventure,Fantasy"

votesNum: 316048

primaryTitle: "Loki"

startYear: 2021

genres: "Action,Adventure,Fantasy"

votesNum: 135726

primaryTitle: "Mortal Kombat"

startYear: 2021



genres: "Action,Adventure,Fantasy"

votesNum: 126530

|||

\$limit

☒

```
1 5
```

Output after \$limit stage (Sample of 5 documents)

primaryTitle: "Zack Snyder's Justice League"

startYear: 2021

genres: "Action,Adventure,Fantasy"

votesNum: 316048

primaryTitle: "Loki"

startYear: 2021

genres: "Action,Adventure,Fantasy"

votesNum: 135726

primaryTitle: "Mortal Kombat"

startYear: 2021

genres: "Action,Adventure,Fantasy"

votesNum: 126530

ADD STAGE