TAREA DE BASES DE DATOS NO SQL





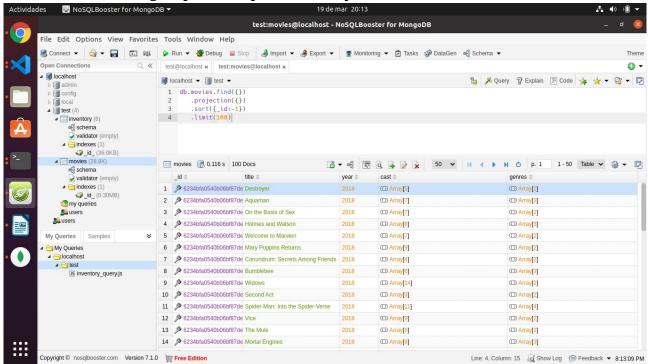
MASTER BIG DATA Y BUSINESS ANALYTICS

ALUMNA: ESTHER FERNÁNDEZ BARCHÍN

FECHA DE ENTREGA: 29/03/2022

EJERCICIOS

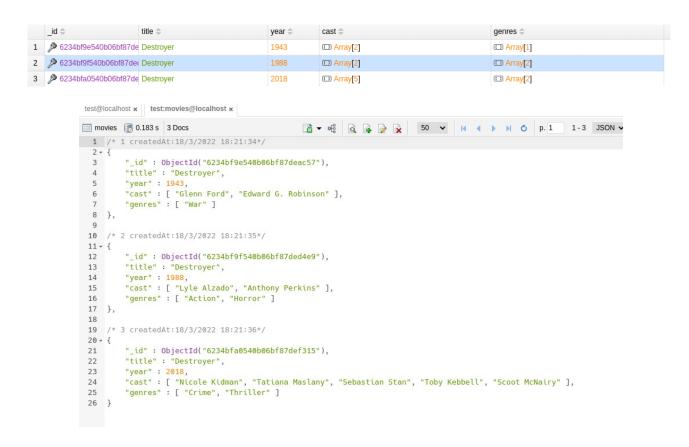
0. Realizar con mongoimport la importación del json en una colección llamada "movies".



1. Analizar con find la colección.

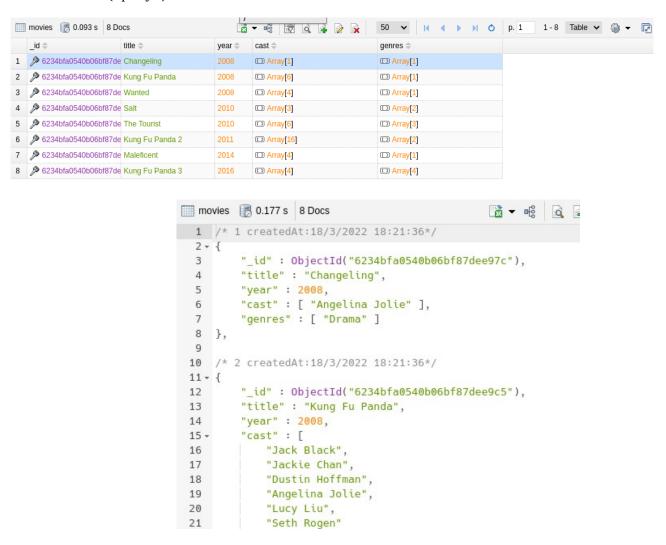
- En el primer ejemplo queremos ver qué información hay relacionada con la película Destroyer.

var query= { "title": "Destroyer" }
db.movies.find(query)



- Ahora queremos investigar sobre el cast alguien que se llame Angelina Jolie a partir del año 2008.

var query1 = { "cast":"Angelina Jolie", year: {\$gte : 2008}}
db.movies.find(query1)



Y por último, queremos ver qué películas salieron en 1900.

var query2 = {"year": 1900}
db.movies.find(query2)

2. Contar cuántos documentos (películas) tiene cargado.

db.movies.find().count()



3. Insertar una película.

```
item = {title :"Nueva peli"}
db.movies.insert(item)
```

db.movies.find(item)



4. Borrar la película insertada en el punto anterior (en el 3).

db.movies.remove({'title':'Nueva peli'})
db.movies.find(item)



5. Contar cuantas películas tienen actores (cast) que se llaman "and. Estos nombres de actores están por ERROR.

var actores = {"cast": "and"}
db.movies.find(actores).count()



6. Actualizar los documentos cuyo actor (cast) tenga por error el valor "and" como si realmente fuera un actor. Para ello, se debe sacar únicamente ese valor del array cast. Por lo tanto, no se debe eliminar ni el documento (película) ni su array cast con el resto de actores. db.movies.updateMany({"cast": "and"},{ \$pull: { "cast": "and" }})



7. Contar cuantos documentos (películas) no tienen actores (array cast).

```
var query = {"cast": {$size: 0}}
db.movies.find(query).count()
```

```
© 0.119 s
1 986
```

8. Actualizar TODOS los documentos (películas) que no tengan actores (array cast), añadiendo un nuevo elemento dentro del array con valor Undefined. Cuidado! El tipo de cast debe seguir siendo un array. El array debe ser así -> ["Undefined"].

db.movies.updateMany({"cast": {\$size: 0}},{ \$push: { "cast": {\$each:["Undefined"] }})
db.movies.find({"cast": "Undefined"})

```
0.112 s

1 - {

2     "acknowledged" : true,

3     "matchedCount" : 986,

4     "modifiedCount" : 986

5 }
```

9. Contar cuantos documentos (películas) no tienen Género (array genres).

```
var query = {"genres"{$size: 0}}
db.movies.find(query).count()
```

```
0.184 s
```

10. Actualizar TODOS los documentos (películas) que no tengan géneros (array genres), añadiendo un nuevo elemento dentro del array con valor Undefined. Cuidado! El tipo de genres debe seguir siendo un array. El array debe ser así -> ["Undefined"].

```
db.movies.updateMany({"genres"{$size: 0}},{ $push:{ "genres":{$each:["Undefined"] }})
db.movies.find({"genres": "Undefined"})
```

```
0.109 s
1 - {
        "acknowledged" : true,
2
        "matchedCount" : 901,
3
        "modifiedCount" : 901
4
5
   }
1 /* 1 createdAt:22/3/2022 17:34:49*/
       "_id" : ObjectId("6239faa9b5836f9d40c913be"),
3
4
       "title" : "Caught",
5
       "year" : 1900,
       "cast" : [ "Undefined" ],
       "genres" : [ "Undefined" ]
7
  },
8
```

11. Mostrar el año más reciente / actual que tenemos sobre todas las películas.

```
db.movies.aggregate([
  {
                                                   movies 3 0.088 s 1 Doc
    $project:{_id:null,"year":-1}
                                                    1 - {
  },
                                                             "year" : 2018,
                                                    2
                                                             "_id" : null
                                                    3
    $sort:{"year":-1}
                                                    4
                                                        }
  },
  {
    $limit:1
])
```

12. Contar cuántas películas han salido en los últimos 20 años. Debe hacerse desde el último año que se tienen registradas películas en la colección, mostrando el resultado total de esos años. Se debe hacer con el Framework de Agregación.

El periodo que comprende es entre 1998-2008

```
movies 0.066 s 1 Doc

1 - {
2     "_id" : null,
3     "count" : 4787
4 }
```

13. Contar cuántas películas han salido en la década de los 60 (del 60 al 69 incluidos). Se debe hacer con el Framework de Agregación.

```
movies 0.054 s 1 Doc

1 - {
2     "_id" : null,
3     "count" : 1414
4 }
```

14. Mostrar el año u años con más películas mostrando el número de películas de ese año. Revisar si varios años pueden compartir tener el mayor número de películas.

Para revisar si hay varios años que comparten el mayor número de películas, se ha filtrado mediante sort y al no haber ninguna repetida, se ha hecho un limit(5) para ver las 5 primeras.

```
📆 ▼ 🛕 🔒 🉀 200 ∨ 🖟 4 🕨 🐧 p. 1 1-5 JSON ∨
                                       movies 0.057 s 5 Docs
db.movies.aggregate([
      $unwind: "$year"
                                               " id" : 1919.
                                         3
                                               "pelis" : 634
                                         4
   },
                                         5 },
  {
                                         6
                                         7 /* 2 */
      $group: {
         _id: "$year",
                                               "_id" : 1925,
                                         9
                                               "pelis" : 572
                                        10
         "pelis": {$sum: 1}
                                        11 },
                                        12
                                        13 /* 3 */
   }
                                               "_id" : 1936,
                                        15
                                               "pelis" : 504
                                        16
      $sort:{"pelis":-1}
                                        17 },
                                        18
                                        19 /* 4 */
      $limit:5
                                               "_id" : 1926,
                                        21
                                               "pelis" : 491
                                        22
                                        23 },
                                        24
                                        25 /* 5 */
])
                                        27
                                              "_id" : 1924,
                                        28
                                               "pelis" : 480
                                        29 }
```

15. Mostrar el año u años con menos películas mostrando el número de películas de ese año. Revisar si varios años pueden compartir tener el menor número de películas.

```
db.movies.aggregate([
                                    movies 0.064 s 5 Docs
                                                                               200 ∨ | ( ) p. 1 1-5 JSON ∨ ( ) →
   {
                                    1 /* 1 */
      $unwind: "$year"
                                            "_id" : 1906,
                                            "pelis" : 7
   },
                                      5 },
      $group: {
                                           "_id" : 1907,
"pelis" : 7
         _id: "$year",
                                     10
         "pelis": {$sum: 1}
                                     12
                                           "_id" : 1902,
                                     16
17 },
                                            "pelis" : 7
                                     19 /* 4 */
      $sort:{"pelis":1}
                                            "_id" : 1900,
                                     21
   }
                                     23 }.
                                     25 /* 5 */
      $limit:5
                                     26 - {
                                           "_id" : 1908,
"pelis" : 18
                                     28
])
```

16. Guardar en nueva colección llamada "actors" realizando la fase \$unwind por actor. Después, contar cuantos documentos existen en la nueva colección.

```
■ Iocalhost

db.movies.aggregate([

    □ config (2)

                                      $unwind: "$cast"
                                      △ [ test (5)
  },

■ actors (83.2K)

                                           o schema
                                           validator (empty)
    $project: {"_id": 0}
                                          },
                                             id_ (0.82MB)
                                        inventory (6)
  {
                                        Sout: "actors"
                                          my queries
                                         Susers
                                        users
1)
db.actors.find().count()
 0.026 s
  1 83224
```

17. Sobre actors (nueva colección), mostrar la lista con los 5 actores que han participado en más películas mostrando el número de películas en las que ha participado. Importante! Se necesita previamente filtrar para descartar aquellos actores llamados "Undefined". Aclarar que no se eliminan de la colección, sólo que filtramos para que no aparezcan.

```
db.actors.aggregate([ {
```

```
$unwind: "$cast"
                                             1 /* 1 */
                                              2 - {
},
                                                      "_id" : "Harold Lloyd",
                                              3
{
                                                      "cuenta" : 190
                                              4
  $match: {
                                              5
                                                  },
    "cast":
                                              6
                                              7 /* 2 */
       $ne:"Undefined"
                                              8 - {
                                                      "_id" : "Hoot Gibson",
                                              9
                                                      "cuenta" : 142
                                              10
                                                  },
                                              11
}},
                                              12
                                              13 /* 3 */
  $group: {
                                             14 ▼ {
    _id: "$cast",
                                                      "_id" : "John Wayne",
                                             15
    "cuenta": {$sum: 1}
                                                      "cuenta" : 136
                                              16
                                              17
                                                  },
                                             18
                                             19 /* 4 */
                                             20 - {
  $sort:{"cuenta":- 1}
                                                      "_id" : "Charles Starrett",
                                             21
}
                                                      "cuenta" : 116
                                              22
                                              23
  $limit:5
                                              24
                                              25 /* 5 */
                                              26 + {
                                                      "_id" : "Bebe Daniels",
                                              27
                                                      "cuenta" : 103
                                              28
                                              29 }
```

18. Sobre actors (nueva colección), agrupar por película y año mostrando las 5 en las que más actores hayan participado, mostrando el número total de actores.

])

```
actors 📳 0.338 s 5 Docs
                                                 13 - 2 2 2
1 /* 1 */
            "title" : "The Twilight Saga: Breaking Dawn - Part 2",
            "year" : 2012
  6
         "actores" : 35
    },
  8
  9
 10 /* 2 */
 11 - {
         "_id" : {
 12 -
            "title" : "Anchorman 2: The Legend Continues",
 13
            "year" : 2013
 14
 15
 16
         "actores" : 33
 17 },
 18
 19 /* 3 */
         "_id" : {
          "title" : "Cars 2",
"year" : 2011
 23
 8 /* 4 */
 9 + {
          "_id" : {
 0 -
           "title" : "Avengers: Infinity War",
  2
              "year" : 2018
  3
          },
  4
          "actores" : 29
  5 },
  6
  7 /* 5 */
  8 + {
          "_id" : {
  9 +
  0
             "title" : "Grown Ups 2",
              "year" : 2013
  1
  2
          "actores" : 28
  4
```

19. Sobre actors (nueva colección), mostrar los 5 actores cuya carrera haya sido la más larga. Para ello, se debe mostrar cuándo comenzó su carrera, cuándo finalizó y cuántos años han pasado. Importante! Se necesita previamente filtrar para descartar aquellos actores llamados "Undefined". Aclarar que no se eliminan de la colección, sólo que filtramos para que no aparezcan.

```
db.actors.aggregate([
  {
     $match: {
       "cast":
       {
          $ne:"Undefined"
  },
     $group: {
       _id: {actor:"$cast"},
       comienza: {$min:"$year"},
       termina: {$max:"$year"}
     }
  },
     $project: {
       "comienza": 1,
       "termina":1,
       experiencia: {$subtract:
["$termina","$comienza"]
     }
  }
  {
     $sort: {"experiencia": -1}
  },
     $limit: 5
1)
```

```
1 /* 1 */
2 - {
3 +
       "_id" : {
       "actor" : "Harrison Ford"
4
5
6
       "comienza" : 1919,
7
       "termina" : 2017,
       "experiencia": 98
8
9
   },
.0
.1
  /* 2 */
2 - {
       "_id" : {
3 -
       "actor" : "Gloria Stuart"
.4
.5
.6
       "comienza" : 1932,
7
       "termina" : 2012,
8.
       "experiencia": 80
9
  },
0
1 /* 3 */
2 - {
       "_id" : {
3 -
       "actor" : "Lillian Gish"
4
5
6
       "comienza" : 1912,
       "termina" : 1987,
7
       "experiencia": 75
8
9 },
30
31 /* 4 */
32 - {
        "_id" : {
33 +
        "actor" : "Kenny Baker"
34
35
        },
36
        "comienza" : 1937,
        "termina" : 2012,
37
        "experiencia" : 75
38
39 },
40
41 /* 5 */
42 - {
        "_id" : {
43 -
           "actor" : "Angela Lansbury"
44
45
        },
        "comienza" : 1944,
46
47
        "termina" : 2018,
        "experiencia" : 74
48
49 }
```

20. Sobre actors (nueva colección), Guardar en nueva colección llamada "genres" realizando la fase \$unwind por genres. Después, contar cuantos documentos existen en la nueva colección.

```
■ test (6)

db.actors.aggregate([
                                                  $unwind: "$genres"
                                                      o schema
  },
                                                      validator (empty)
  {
                                                    indexes (1)
    $project: {"_id": 0}
                                                        id (4.0KB)
  },
  {
    $out: "genres"
                                                       0.017 s
1)
db.genres.find().count()
                                                       1 104950
```

21. Sobre genres (nueva colección), mostrar los 5 documentos agrupados por "Año y Género" que más número de películas diferentes tienen mostrando el número total de películas.

```
db.genres.aggregate([
                                                                 1 /* 1 */
                                                                 2 - {
     $group: {
                                                                 3 →
                                                                           "_id" : {
        _id: {año:"$year", generos:"$genres"},
                                                                                "año" : 2012,
                                                                 4
        "total": {$sum: 1}}
                                                                                "generos" : "Comedy"
                                                                 5
   },
                                                                           "total" : 790
                                                                 7
     $sort: {"total":-1}
                                                                     },
   },
   {
     $limit:5
])
                                                                                        total 🛊
   año 🖨
                            generos $
 1 2012 (2.0K)
                            Comedy
 2 2013 (2.0K)
                                                                                        694
                            Comedy
                                                                                        677
 3 2017 (2.0K)
                            Drama
 4 2012 (2.0K)
                                                                                        657
```

5 1919 (1.9K)

Drama

618

22. Sobre genres (nueva colección), mostrar los 5 actores y los géneros en los que han participado con más número de géneros diferentes, se debe mostrar el número de géneros diferentes que ha interpretado. Importante! Se necesita previamente filtrar para descartar aquellos actores llamados "Undefined". Aclarar que no se eliminan de la colección, sólo que filtramos para que no aparezcan.

```
db.genres.aggregate([
  {
     $match: {
       "cast":
       {
          $ne:"Undefined"
       "genres":
          $ne:"Undefined"
  },
     $group: {
       _id: {actor:"$cast"},
        generos: {$addToSet: "$genres"}
     }
  },
     $project: {
       _id:1
       generos:1
       numgeneros:{$size:"$generos"}
     }
  },
     $sort:{numgeneros:-1}
  },
     $limit:5
  }
```

```
1 /* 1 */
2 + {
3 +
        "_id" : {
            "actor" : "Dennis Quaid"
5
        },
5 +
        "generos" : [
7
            "Drama",
8
            "Horror"
9
            "Comedy",
0
            "Crime",
1
            "Dance",
2
            "Animated",
3
            "Family",
4
            "Disaster",
5
            "Suspense",
            "Western",
7
            "Satire",
8
            "Thriller",
9
            "Musical",
            "Fantasy",
1
            "Adventure",
2
            "Biography",
3
            "Sports",
4
            "Romance",
5
            "Action",
6
            "Science Fiction"
7
        ],
8
        "numgeneros": 20
9
   },
```

])

genres 📳 0.482 s 5 Docs		2 ▼ 1	1-5 Table ∨ 🎡 ▼
	_id	generos \$	numgeneros \$
1	{ actor : "Dennis Quaid" }	☐ Array[20]	20
2		□ Array[18]	18
3		□ Array[18]	18
4	☑ { actor : "Helen Mirren" }	☐ Array[18]	18
5	(actor : "Danny Glover" }	☐ Array[18]	18

23. Sobre genres (nueva colección), mostrar las 5 películas y su año correspondiente en los que más géneros diferentes han sido catalogados, mostrando esos géneros y el número de géneros

que contiene.

```
db.genres.aggregate([
  {
     $match: {
       "genres":
          $ne:"Undefined"
     }
  },
     $group: {
       _id: {titulo:"$title", año:"$year"},
        generos: {$addToSet: "$genres"}
  },
  {
     $project: {
       _id:1
       generos:1
       numgeneros:{$size:"$generos"}
     }
  },
     $sort:{numgeneros:-1}
  },
  {
     $limit:5
  }
])
```

```
2 - {
 3 →
         "_id" : {
             "titulo" : "American Made",
 4
 5
             "año" : 2017
 6
         },
         "generos" : [
 7 -
 8
             "Crime",
 9
             "Action",
             "Thriller",
10
             "Drama",
11
             "Historical",
12
             "Biography",
13
14
             "Comedy"
15
         "numgeneros": 7
16
17 },
```

```
      Key $
      Value ¶
      Type

      ▷ ☒ (1) { titulo : "American Made", "año" } { generos : [ "Crime", "Action", "Thriller", "Drama", "Historical", "Biography", "Comedy"], numge Document

      ▷ ☒ (2) { titulo : "Wonder Woman", "año" } { generos : [ "Superhero", "Action", "Fantasy", "Adventure", "Drama", "War"], numgeneros : 6 } Document

      ▷ ☒ (3) { titulo : "The Dark Tower", "año" } { generos : [ "Fantasy", "Adventure", "Horror", "Action", "Science Fiction", "Western"], numgeneros Document

      ▷ ☒ (4) { titulo : "My Little Pony: The Mov } { generos : [ "Family", "Adventure", "Fantasy", "Animated", "Comedy", "Musical"], numgeneros Document

      ▷ ☒ (5) { titulo : "Dunkirk", "año" : 2017 } { generos : [ "Action", "Thriller", "Drama", "Adventure", "Historical", "War"], numgeneros : 6 } Document
```

24. Sobre genres buscar el nombre de Tom y ver qué Tom ha sido el que mayor numero de géneros ha participado y en cuales.

```
db.actors.aggregate([
  {
                                        actors 3 0.144 s 1 Doc
                                                                                               1 ▼ 3 3
    $match: {
                                           1 - {
       "cast":{$regex: "Tom"},
                                                  "_id" : {
                                           2 -
       "genres":{$ne:"Undefined"}
                                           3
                                                      "actor" : "Tom Hanks"
                                           4
                                                  },
  }
                                          5 +
                                                  "generos" : [
  {
                                                      [ "Comedy", "Drama" ],
                                          7
                                                      [ "Family" ],
    $group: {
                                                      [ "Adventure" ],
                                          8
       _id: { actor:"$cast"}
                                                      [ "Romance", "Comedy" ],
                                          9
       generos: {$addToSet: "$genres"}
                                                      [ "Adventure", "Mystery" ],
                                          10
     }
                                                      [ "Adventure", "Drama" ],
                                          11
                                                      [ "Biography", "Drama", "Historical", "Thriller" ],
  },
                                          12
                                          13
                                                      [ "Thriller" ],
                                                      [ "Action", "Drama", "Thriller" ],
                                          14
    $project: {
                                                      [ "Drama", "Musical" ],
                                          15
       _id: 1
                                                      [ "Crime", "Drama" ],
                                          16
       generos:1
                                          17
                                                      [ "War", "Drama" ],
       numgeneros:{$size :"$generos"}
                                                      [ "Animated", "Family" ],
                                          18
                                                      [ "Drama", "Comedy" ],
     }
                                          19
                                                      [ "Drama" ],
                                          20
  },
                                                      [ "Biography", "Drama" ],
                                          21
  {
                                          22
                                                      [ "Comedy" ],
    $sort: {numgeneros:-1}
                                                      [ "Drama", "Biography" ],
                                          23
  },
                                          24
                                                      [ "Animated" ],
  {
                                          25
                                                      [ "Drama", "War" ],
                                                      [ "Animated", "Comedy", "Family" ],
    $limit: 1
                                          26
                                          27
                                                      [ "Science Fiction", "Drama", "Thriller" ]
                                          28
                                                  "numgeneros" : 22
                                          29
                                         30 }
```

])

25. Query sobre actors en el que buscamos el Tom que ha hecho mayor número de películas.

```
actors 🕝 0.141 s 5 Docs
db.actors.aggregate([
                                     1 /* 1 */
  {
    $match: {
                                     2 - {
      "cast":{$regex: "Tom"},
                                     3 +
                                            "_id" : {
                                     4
                                                "actor" : "Tom and Jerry"
  }
                                     5
                                            },
                                     6 +
                                             "pelis" : [
    $group: {
                                     7
                                                 "Salt Water Tabby",
                                                 "Springtime for Thomas",
      _id: { actor:"$cast"}
                                     8
                                                 "The Framed Cat",
                                     9
       pelis: {$addToSet: "$title"}
                                    10
                                                 "Sleepy-Time Tom",
                                                 "Love That Pup",
                                    11
  },
                                    12
                                                 "Little Quacker",
                                                 "Pup on a Picnic",
                                    13
    $project: {
                                                 "Fine Feathered Friend",
                                    14
      _id: 1
                                    15
                                                 "The Flying Sorceress",
      pelis:1
                                                 "Tee for Two",
                                    16
      numpelis:{$size:"$pelis"}
                                    17
                                                 "Timid Tabby",
    }
                                                 "The Dog House",
                                    18
  },
                                                 "The Missing Mouse",
                                    19
                                    20
                                                 "Muscle Beach Tom",
    $sort:{numpelis:-1}
                                    21
                                                 "The Milky Waif",
  }
                                                 "Trap Happy",
                                    22
                                    23
                                                 "The Lonesome Mouse",
    $limit:1
                                                 "Tops with Pops",
                                    24
  }
                                    25
                                                 "Robin Hoodwinked",
])
                                                 "Solid Serenade",
                                    26
                                                 "A Mouse in the House",
                                    27
                                    28
                                                 "Saturday Evening Puss",
                                    29
                                                 "Little Runaway",
                                    30
                                                 "Southbound Duckling",
                                                 "Happy Go Ducky",
                                    31
                                   actors 🕝 0.141 s 5 Docs
                                      78
                                                   "Busy Buddies",
                                      79
                                                   "Tom's Photo Finish",
                                                   "Switchin' Kitten",
                                      80
                                      81
                                                   "Barbecue Brawl",
                                                   "Quiet Please!",
                                      82
                                                   "Carmen Get It!",
                                      83
                                                   "His Mouse Friday",
                                      84
                                      85
                                                   "Cue Ball Cat",
                                                   "Dr. Jekyll and Mr. Mouse",
                                      86
                                      87
                                                   "Jerry and Jumbo",
                                      88
                                                   "The Vanishing Duck",
                                                   "Jerry and the Goldfish",
                                      89
                                      90
                                                   "The Mouse Comes to Dinner",
                                                   "Mouse Into Space"
                                      91
                                      92
                                               ],
                                              "numpelis" : 85
                                      93
                                     94 },
```

26.¿Qué película ha hecho Tom Hardy a partir de 2010 y de género de Comedia?