Birla Institute of Technology & Science, Pilani

Work Integrated Learning Programmes

Division First Semester 2022-2023

Assignment I

Course No.: DSECL ZG522

Course Title: Big Data Systems

Group 93

Sr.No.	Name	BITS ID	Contribution (%)
1	Sherwin Philip	2021FC04106	100
2	Avinash Kumar	2021FC04746	100
3	Sidharth vij	2021fc04507	100
4	Saurabh Arunrao Dhande	2021fc04700	100

Sr.No.	Software Used	Version	Comment
1	MongoDBCompass	MongoDB 3.6	MongoDB Compass is a powerful GUI for querying, aggregating, and analyzing your MongoDB data in a visual environment
2	Studio 3T	Studio 3T 2022.10.0	Studio 3T is made for growing professional teams, offering a variety of ways to view and interrogate data collections, including sophisticated aggregations, native Mongo JSON extensions, traditional SQL queries, and a drag and drop query builder.

1]The number of new cases, new deaths and new recovered

Answer - 228693, 5415, 174623

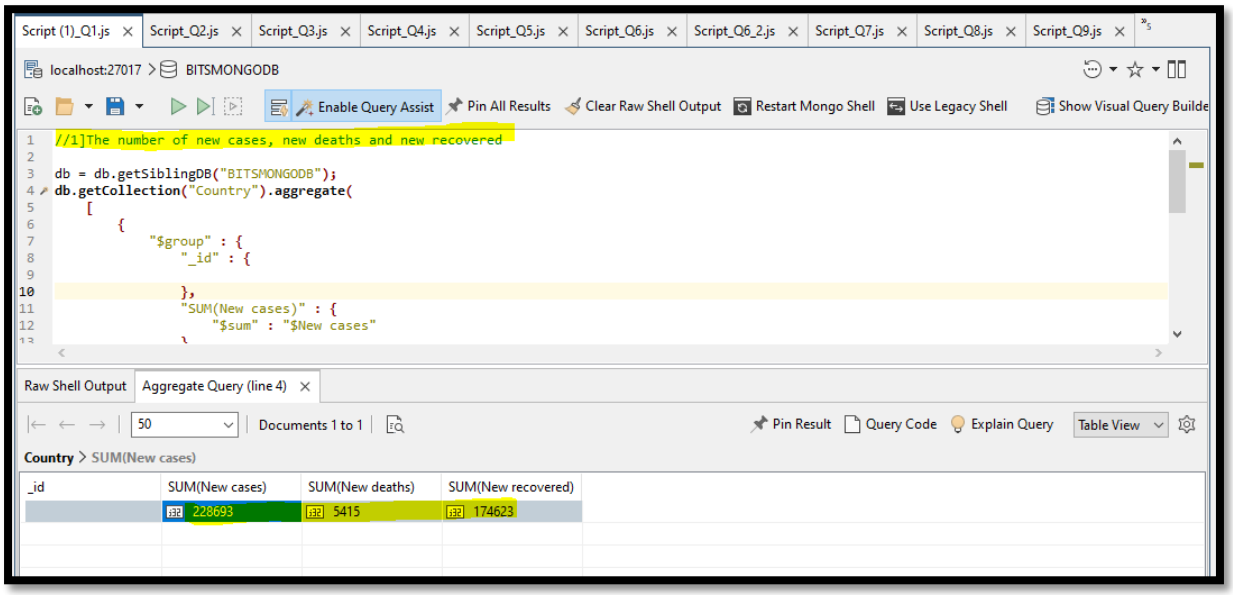
```
db = db.getSiblingDB("BITSMONGODB");
db.getCollection("Country").aggregate(
   [
```

```
    {
        "$group" : {
            "_id" : {

            },
            "SUM(New cases)" : {
                "$sum" : "$New cases"
            },
            "SUM(New deaths)" : {
                "$sum" : "$New deaths"
            },
            "SUM(New recovered)" : {
                "$sum" : "$New recovered"
            }
        }
    },
    {
        "$project" : {
            "SUM(New cases)" : "$SUM(New cases)",
            "SUM(New deaths)" : "$SUM(New deaths)",
            "SUM(New recovered)" : "$SUM(New recovered)",
            "_id" : NumberInt(0)
        }
    }
],
{
    "allowDiskUse" : true
```

}

);

Screenshot:



2]The number of death cases in each country of continent Asia and also the corresponding WHO regions

Answere:

db = db.getSiblingDB("BITSMONGODB");

db.getCollection("World").find(

{

"Continent" : "Asia"

},

{

"Country/Region" : "$Country/Region",

"WHO Region" : "$WHO Region",

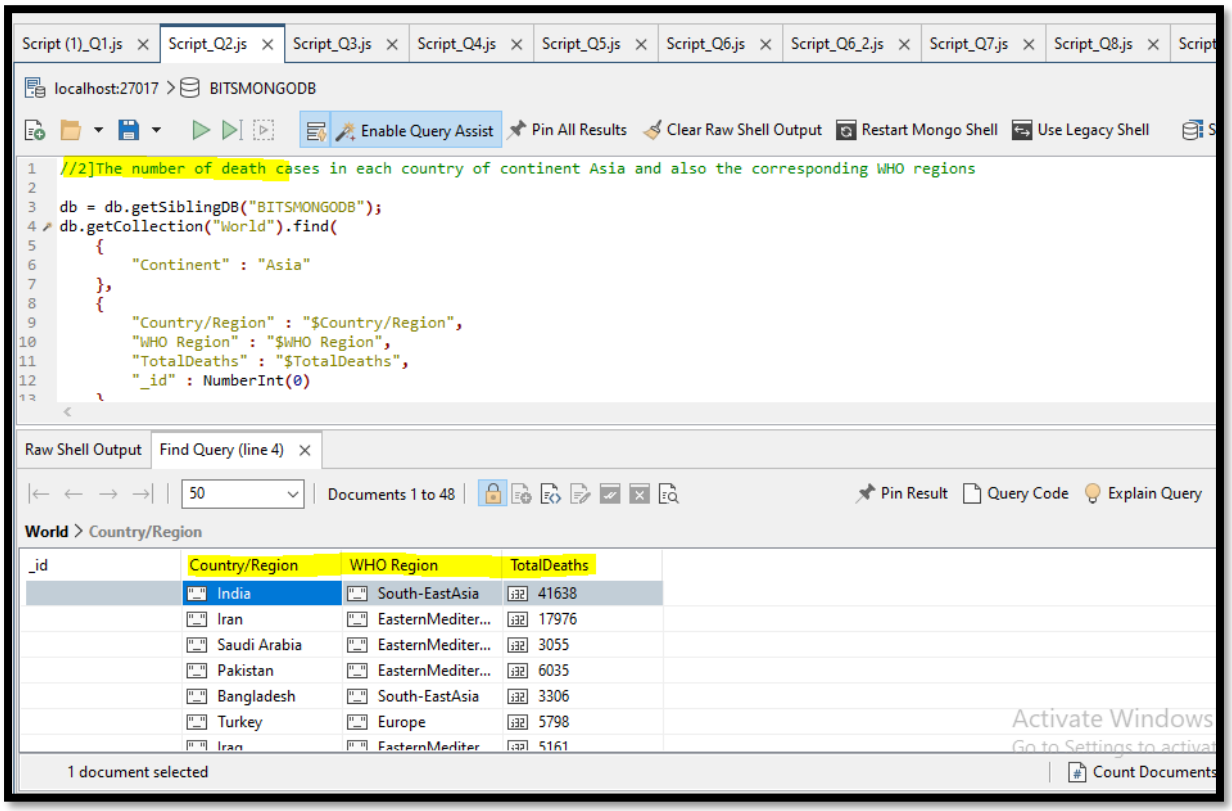
"TotalDeaths" : "$TotalDeaths",

"_id" : NumberInt(0)

}

);

Screenhot:



3]The number of deaths that occurred on 12-02-2020

Answer - 1118

db = db.getSiblingDB("BITSMONGODB");

db.getCollection("Day").find(

{

"Date" : ISODate("2020-02-12")

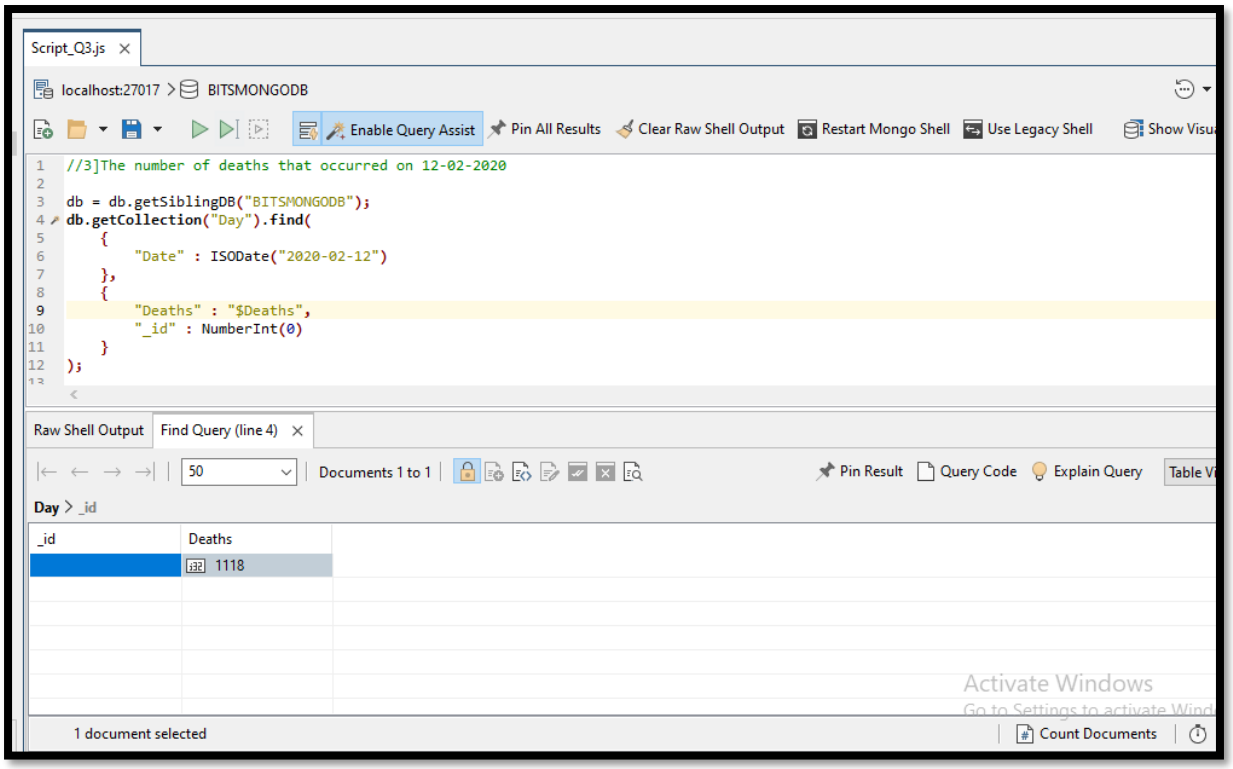
},

{

"Deaths" : "$Deaths",

```
        "_id" : NumberInt(0)
    }
);
```

Screenshot:



4]The number of active new cases (new cases-(new death+new recovered)) in a reverse sorted order based on the country name

Answere:

```
db = db.getSiblingDB("BITSMONGODB");

db.getCollection("Country").aggregate(
    [
        {
            "$group" : {
                "_id" : {
```

```
            "Country/Region" : "$Country/Region"
         },
         "SUM(New deaths)" : {
            "$sum" : "$New deaths"
         },
         "SUM(New recovered)" : {
            "$sum" : "$New recovered"
         },
         "SUM(New cases)" : {
            "$sum" : "$New cases"
         }
      }
   },
   {
      "$project" : {
         "Country/Region" : "$_id.Country/Region",
         "SUM(New deaths)" : "$SUM(New deaths)",
         "SUM(New recovered)" : "$SUM(New recovered)",
         "SUM(New cases)" : "$SUM(New cases)",
         'ActiveNewCases': {'$subtract': ['$SUM(New cases)', { '$add' : [ '$SUM(New deaths)', '$SUM(New recovered)' ] }]},
         "_id" : NumberInt(0)
      }
   },

 {  "$sort":
{
```

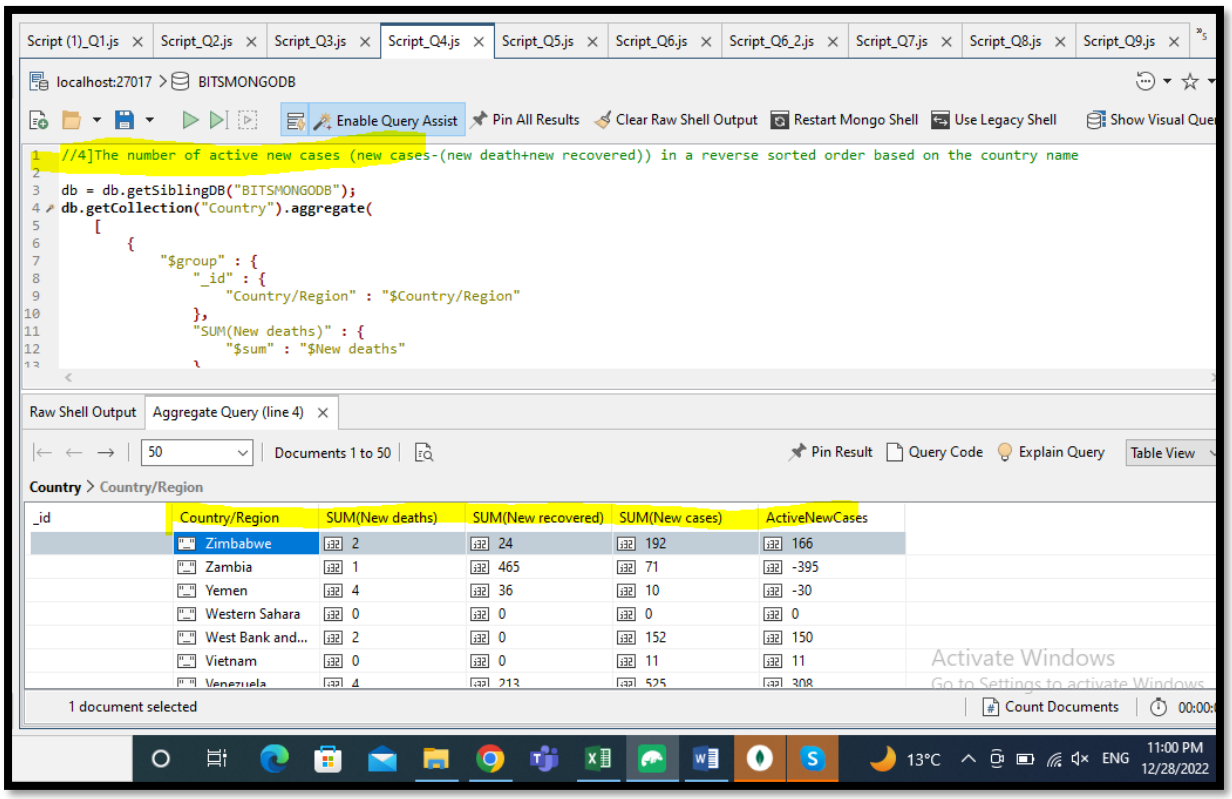
```
        "Country/Region" : NumberInt(-1)
    }
     }


    ],
    {
        "allowDiskUse" : true
    }
);
```

Explanation:

The number of active new cases (new cases-(new death+new recovered)) in a reverse sorted order based on the country name

Observation - The number of active new cases are showing up as negative for few countries. This is because the sum of "new death" and "new recovered" is more than "new cases".

Screenshot:

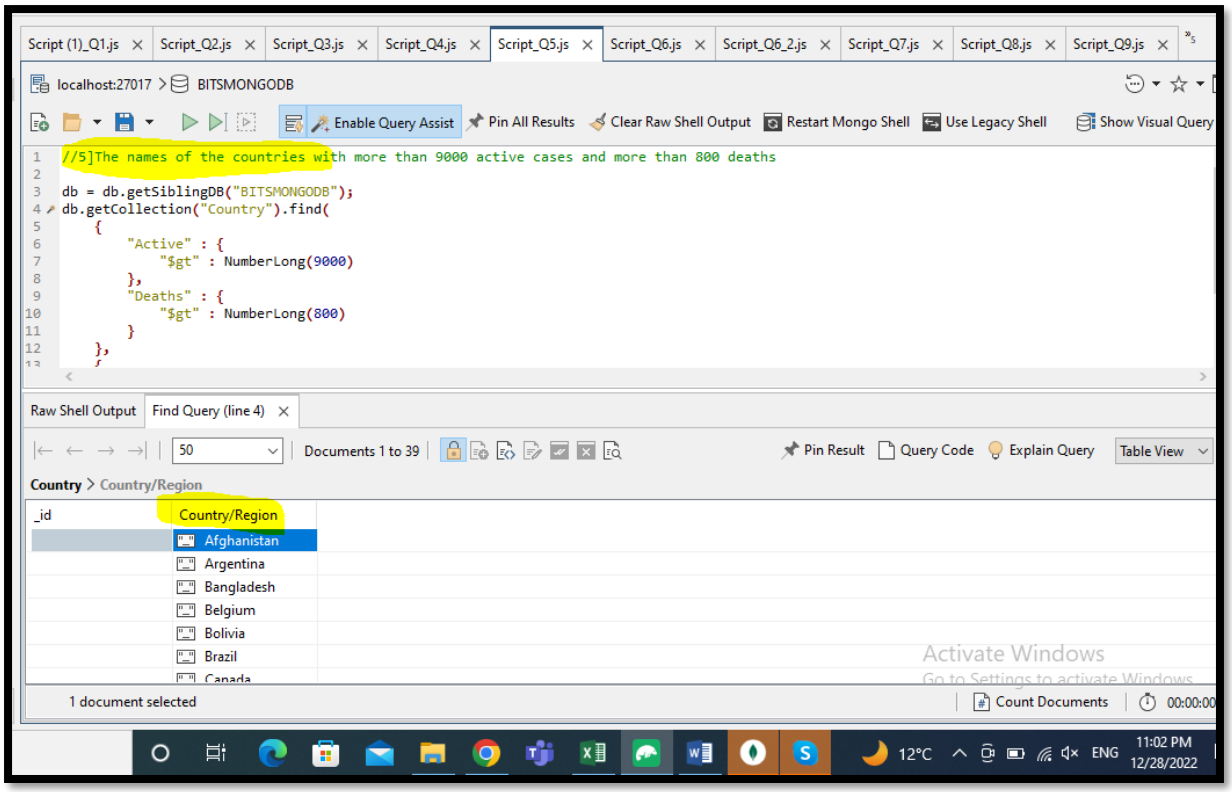


5]The names of the countries with more than 9000 active cases and more than 800 deaths

Answere:

```
db = db.getSiblingDB("BITSMONGODB");

db.getCollection("Country").find(
    {
        "Active" : {
            "$gt" : NumberLong(9000)
        },
        "Deaths" : {
            "$gt" : NumberLong(800)
        }
    },
    {
        "Country/Region" : "$Country/Region",
        "_id" : NumberInt(0)
    }
);
```

Screenshot :



6] The country with the highest number of active cases and also with second highest death rate

Answere: - US, UNITED KINGDOM

HIGHEST ACTIVE

```
db = db.getSiblingDB("BITSMONGODB");
db.getCollection("Country").find({},
    {
        "Country/Region" : "$Country/Region",
        "_id" : NumberInt(0)
    }
).sort(
    {
```

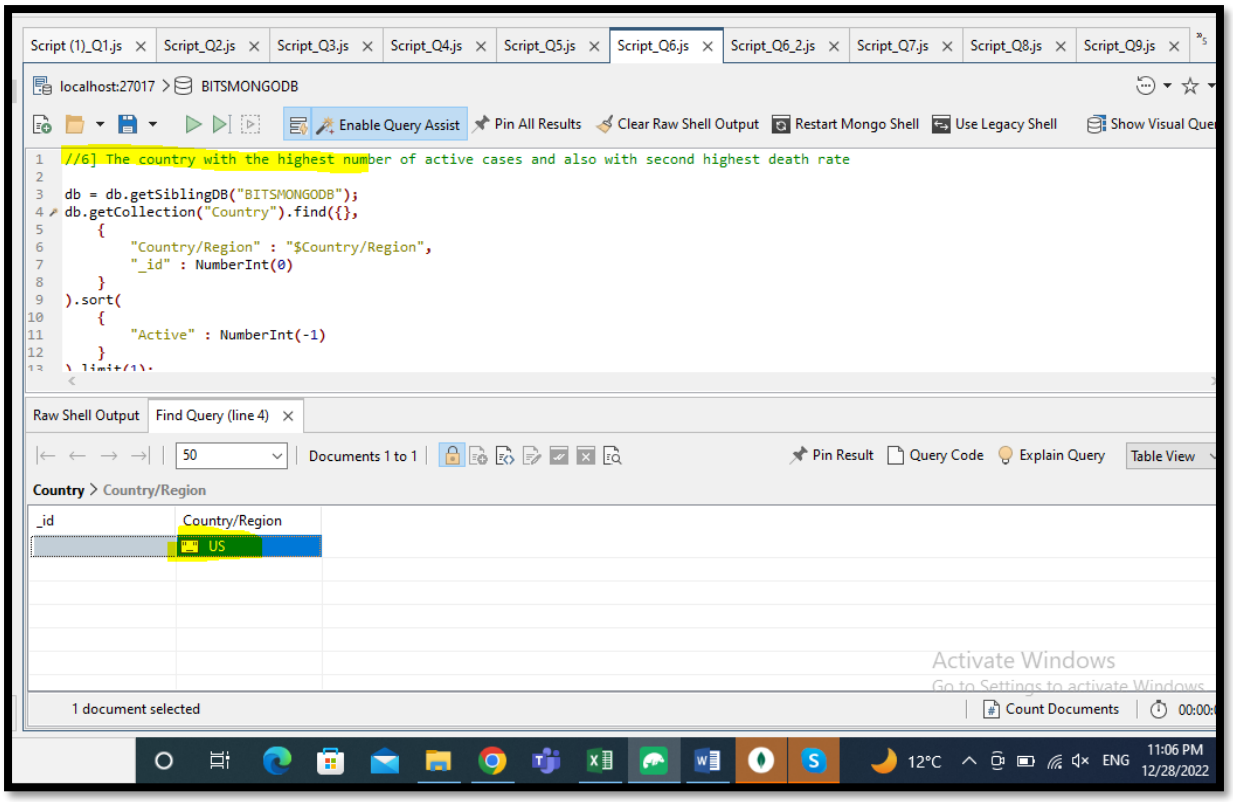
```
        "Active" : NumberInt(-1)
    }
).limit(1);
```

SECOND HIGHEST DEATH RATE

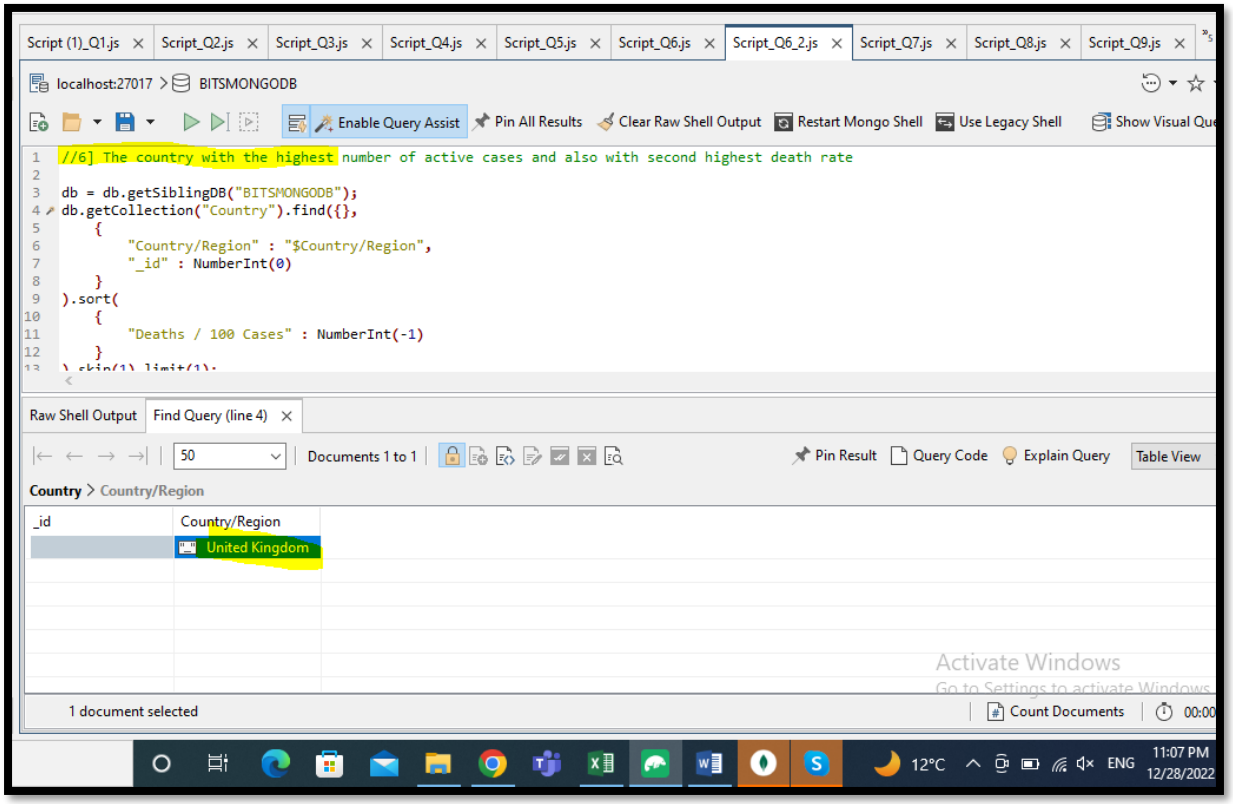
```
db = db.getSiblingDB("BITSMONGODB");
db.getCollection("Country").find({},
    {
        "Country/Region" : "$Country/Region",
        "_id" : NumberInt(0)
    }
).sort(
    {
        "Deaths / 100 Cases" : NumberInt(-1)
    }
).skip(1).limit(1);
```

Screenshot:

HIGHEST ACTIVE



SECOND HIGHEST DEATH RATE



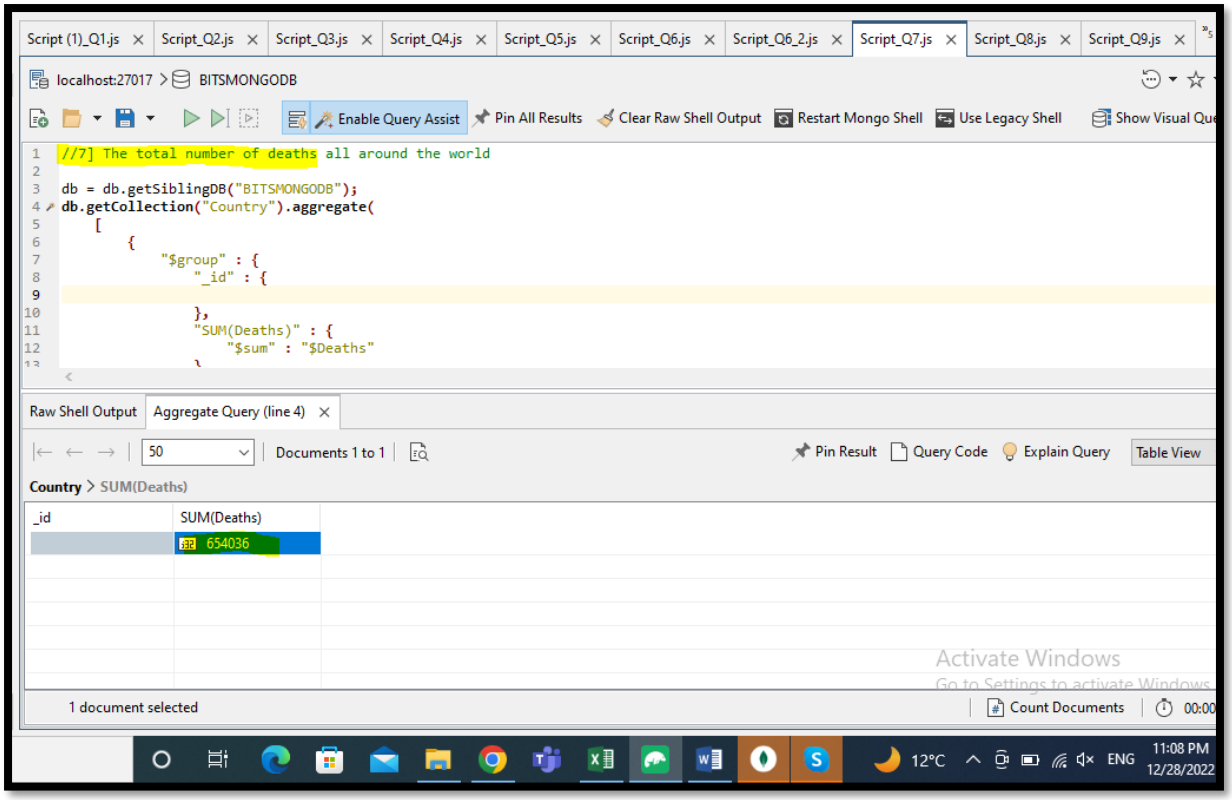
7] The total number of deaths all around the world

Answer - 654036

```
db = db.getSiblingDB("BITSMONGODB");
db.getCollection("Country").aggregate(
   [
      {
         "$group" : {
            "_id" : {

            },
            "SUM(Deaths)" : {
               "$sum" : "$Deaths"
            }
         }
      },
      {
         "$project" : {
            "SUM(Deaths)" : "$SUM(Deaths)",
            "_id" : NumberInt(0)
         }
      }
   ],
   {
      "allowDiskUse" : true
   }
```

);

Screenshot:



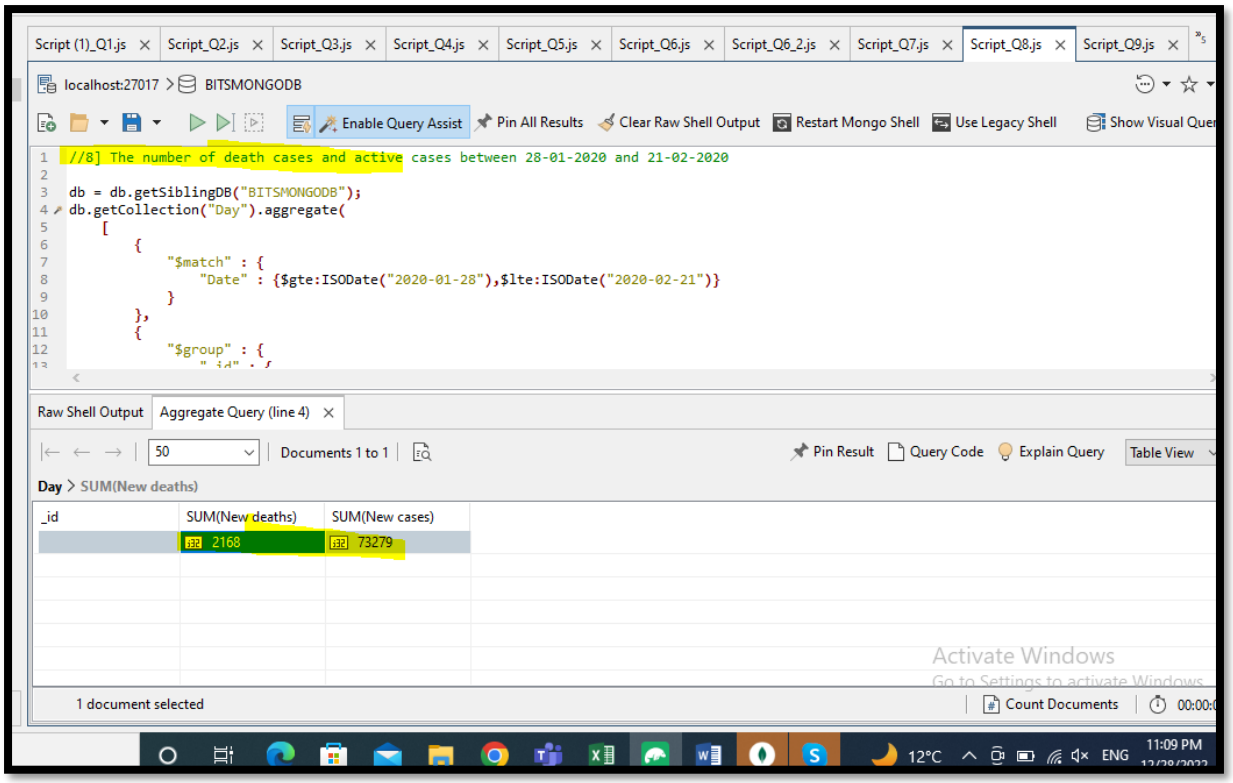
8] The number of death cases and active cases between 28-01-2020 and 21-02-2020

Answere- 2168, 73279

```
db = db.getSiblingDB("BITSMONGODB");
db.getCollection("Day").aggregate(
  [
    {
      "$match" : {
        "Date" : {$gte:ISODate("2020-01-28"),$lte:ISODate("2020-02-21")}
      }
    },
```

```
        {
            "$group" : {
                "_id" : {


                },
                "SUM(New deaths)" : {
                    "$sum" : "$New deaths"
                },
                "SUM(New cases)" : {
                    "$sum" : "$New cases"
                }
            }
        },
        {
            "$project" : {
                "SUM(New deaths)" : "$SUM(New deaths)",
                "SUM(New cases)" : "$SUM(New cases)",
                "_id" : NumberInt(0)
            }
        }
    ],
    {
        "allowDiskUse" : true
    }
);
```

Screenshot:



9] The latitude and longitude of countries ending with “ia” and the number of countries

Answere:

```
db = db.getSiblingDB("BITSMONGODB");
db.getCollection("Covid19").find(
   {
      "Country/Region" : /^.*ia$/i
   },
   {
      "Country/Region" : "$Country/Region",
      "Lat" : "$Lat",
      "Long" : "$Long",
      "_id" : NumberInt(0)
   }
```

);

Answer part 2 - 38

```
db = db.getSiblingDB("BITSMONGODB");
db.getCollection("Covid19").aggregate(
   [
      {
         "$match" : {
            "Country/Region" : /^.*ia$/i
         }
      },
      {
         "$group" : {
            "_id" : {

            },
            "COUNT_DISTINCT(Country/Region)" : {
               "$addToSet" : "$Country/Region"
            }
         }
      },
      {
         "$project" : {
            "COUNT(Country/Region)" : {$size: "$COUNT_DISTINCT(Country/Region)"},
            "_id" : NumberInt(0)
```

```
        }
      }
   ],
   {
      "allowDiskUse" : true
   }
);
```

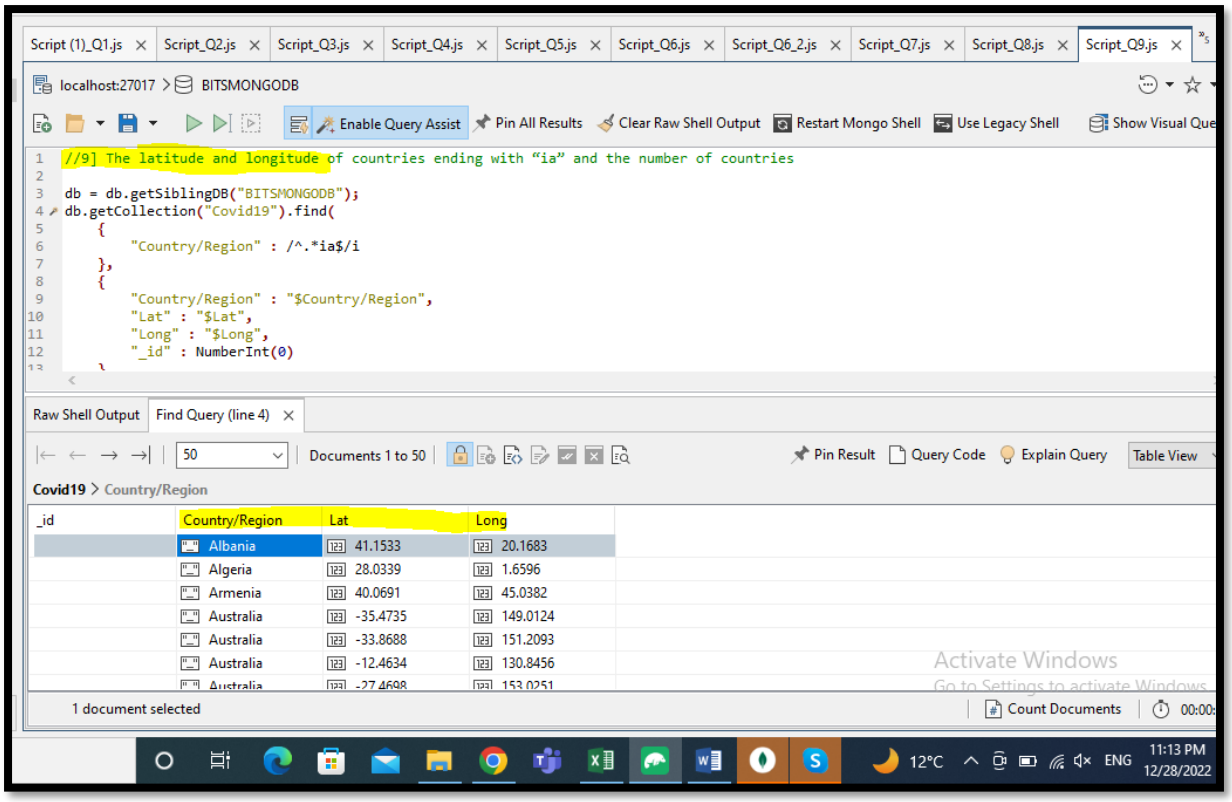
Explanation:

Please note that "country" is not the lowest level of heirarchy in this data. We have "province/state" which is one level lower in heirarchy compared to "country". Hence, you will see there are multiple records for countries such as "Australia" that has many "Lat" and"Long" values. This is because "Australia" has many data "province/state" each with its own "Lat" and "Long" values.

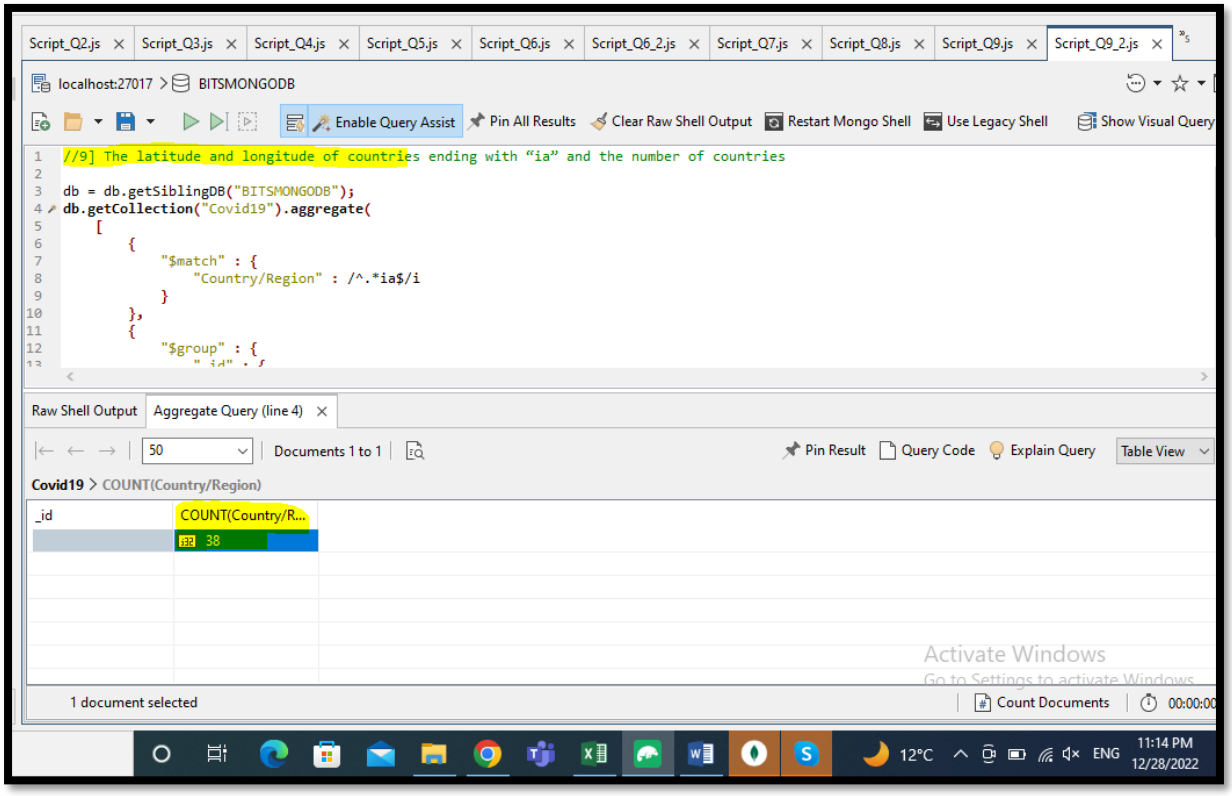
Another point to note here is that when we count such number of "countries" ending with "ia", we should do a distinct count of such countries and not take in to account "province/state" column.

Screenshot:

Part 1:


Script (1)_Q1.js
Script_Q2.js
Script_Q3.js
Script_Q4.js
Script_Q5.js
Script_Q6.js
Script_Q6_2.js
Script_Q7.js
Script_Q8.js
Script_Q9.js
localhost:27017 > BITSMONGODB
Enable Query Assist
Pin All Results
Clear Raw Shell Output
Restart Mongo Shell
Use Legacy Shell
Show Visual Que
//9] The latitude and longitude of countries ending with "ia" and the number of countries
db = db.getSiblingDB("BITSMONGODB");
db.getCollection("Covid19").find(
{
"Country/Region" : /^.*ia$/i
},
{
"Country/Region" : "$Country/Region",
"Lat" : "$Lat",
"Long" : "$Long",
"_id" : NumberInt(0)
Raw Shell Output
Find Query (line 4)
50
Documents 1 to 50
Pin Result
Query Code
Explain Query
Table View
Covid19 > Country/Region
_id
Country/Region
Lat
Long
Albania 41.1533 20.1683
Algeria 28.0339 1.6596
Armenia 40.0691 45.0382
Australia -35.4735 149.0124
Australia -33.8688 151.2093
Australia -12.4634 130.8456
1 document selected
Count Documents
Activate Windows
12°C
ENG
11:13 PM
12/28/2022

Part 2:

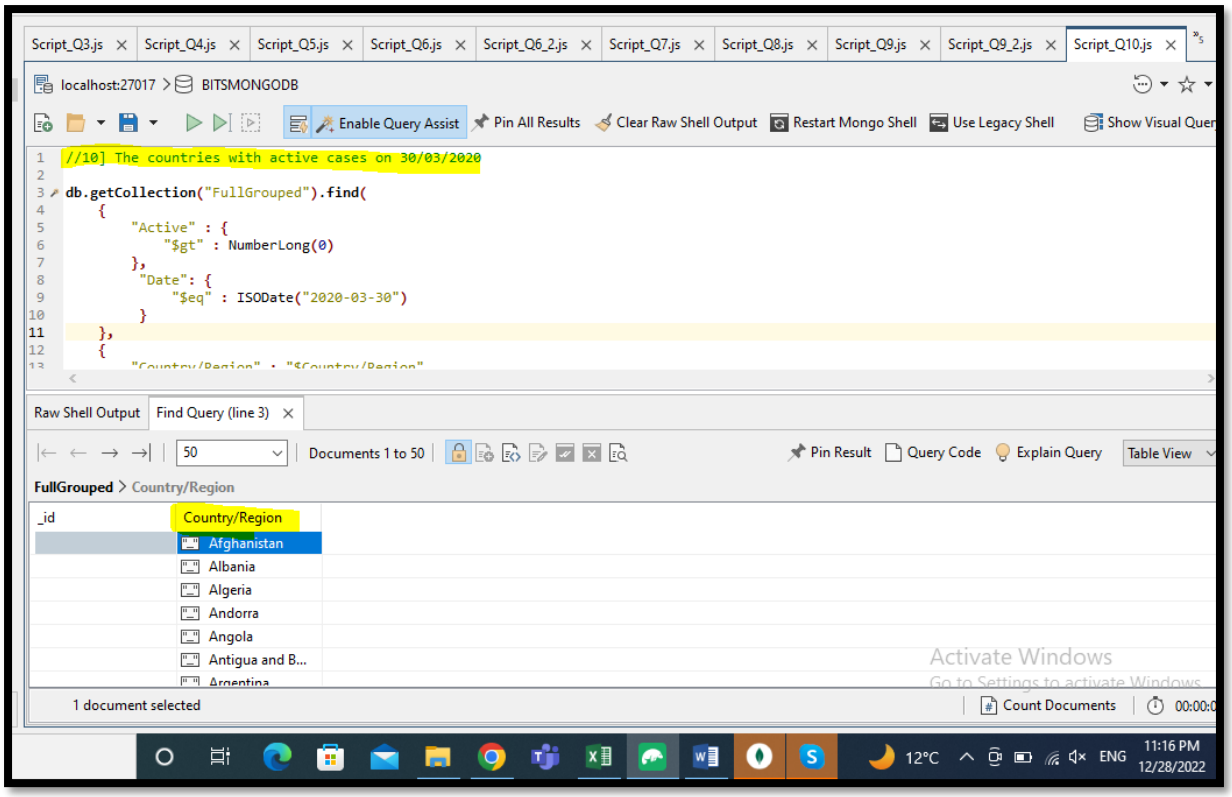

Script_Q2.js
Script_Q3.js
Script_Q4.js
Script_Q5.js
Script_Q6.js
Script_Q6_2.js
Script_Q7.js
Script_Q8.js
Script_Q9.js
Script_Q9_2.js
localhost:27017 > BITSMONGODB
Enable Query Assist
Pin All Results
Clear Raw Shell Output
Restart Mongo Shell
Use Legacy Shell
Show Visual Query
//9] The latitude and longitude of countries ending with "ia" and the number of countries
db = db.getSiblingDB("BITSMONGODB");
db.getCollection("Covid19").aggregate(
[
{
"$match" : {
"Country/Region" : /^.*ia$/i
}
},
{
"$group" : {
Raw Shell Output
Aggregate Query (line 4)
50
Documents 1 to 1
Pin Result
Query Code
Explain Query
Table View
Covid19 > COUNT(Country/Region)
_id
COUNT(Country/R...
38
1 document selected
Count Documents
Activate Windows
12°C
ENG
11:14 PM
12/28/2022

10] The countries with active cases on 30/03/2020

Answere:

```
db.getCollection("FullGrouped").find(
    {
        "Active" : {
            "$gt" : NumberLong(0)
        },
         "Date": {
             "$eq" : ISODate("2020-03-30")
         }
    },
    {
        "Country/Region" : "$Country/Region",
        "_id" : NumberInt(0)
    }
);
```

Screenshot:



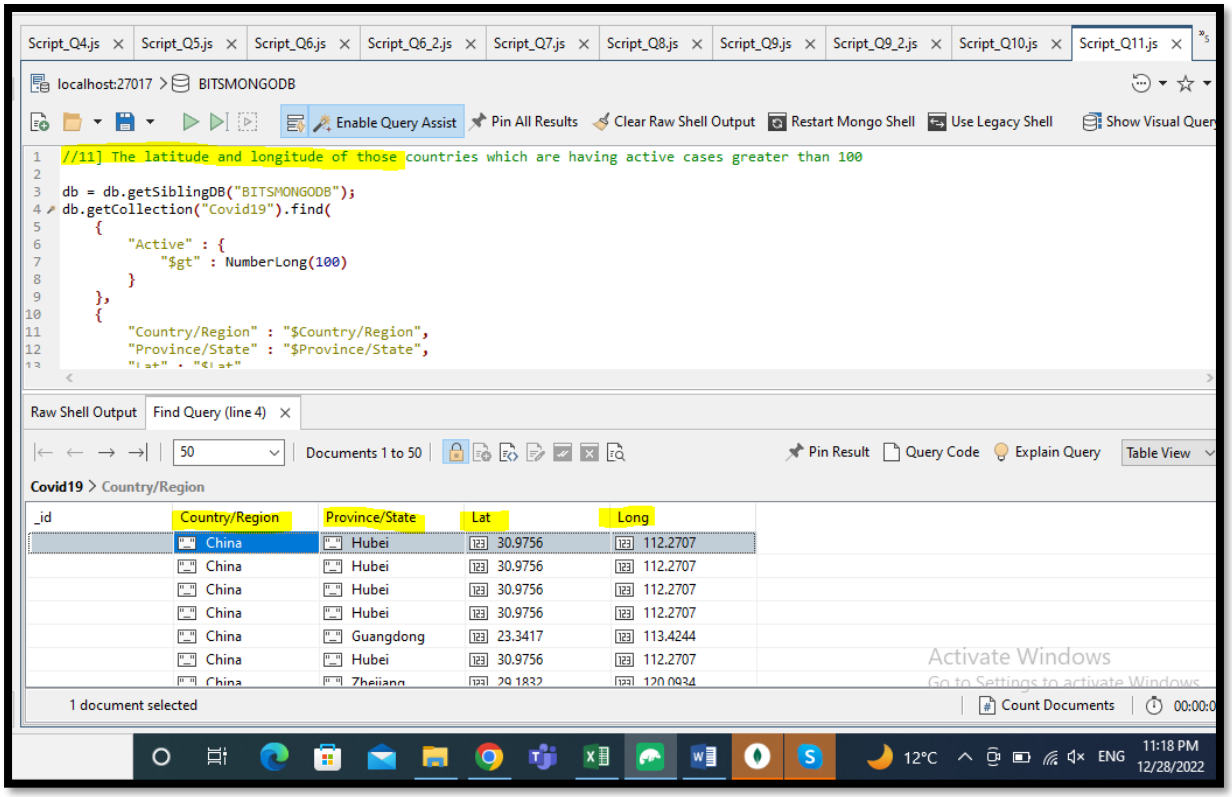
11] The latitude and longitude of those countries which are having active cases greater than 100

Answere:

```
db = db.getSiblingDB("BITSMONGODB");

db.getCollection("Covid19").find(

  {

    "Active" : {

      "$gt" : NumberLong(100)

    }

  },

  {

    "Country/Region" : "$Country/Region",

    "Province/State" : "$Province/State",
```

```
        "Lat" : "$Lat",
        "Long" : "$Long",
        "_id" : NumberInt(0)
    }
);
```

Screenshot:



12]The countries and respective dates in which maximum increase of active cases occurred.

Answere:

```
db = db.getSiblingDB("BITSMONGODB");
db.getCollection("FullGrouped").aggregate(
    [
```

```
{
  $setWindowFields: {
     partitionBy: "$Country/Region",
     sortBy: { Date: 1 },
     output: {
       lag: {
          $shift: {
            output: "$Active",
            by: -1,
            default: NumberInt(0)
          }
        },

     }
  }
}



 ,
    {
       "$project" : {
          "Country/Region" : "$Country/Region",
          "Date" : "$Date",
          'ActiveNewCases': {'$subtract': ['$Active', '$lag']},
          "_id" : NumberInt(0)
       }
    },
```

```
  {
 $group: {
   "_id": "$Country/Region",
   "max_active_new_cases": {
     $max: "$ActiveNewCases"
   },
   "records": {
     $push: "$$ROOT"
   }
 }
},
{
 "$project": {
   items: {
     "$filter": {
       "input": "$records",
       "as": "records",
       "cond": {
         $eq: [
           "$$records.ActiveNewCases",
           "$max_active_new_cases"
         ]
       }
     }
   }
```

```
    }
  },
  {
    $unwind: "$items"
  },
  {
    $replaceWith: "$items"
  }




      ],
    {
      "allowDiskUse" : true
    }
);
```

Explanation:

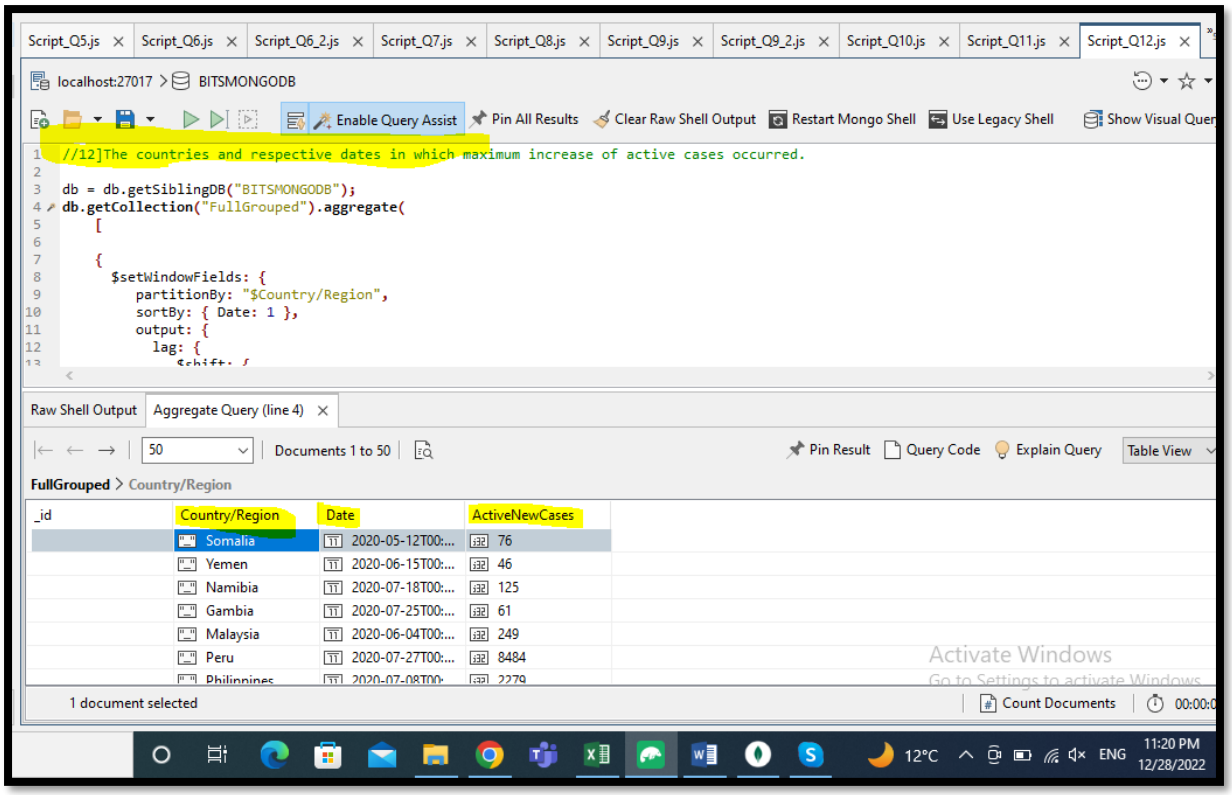
This is a complex question. To solve this question, we need to find the increase of cases on a each date with respect to the previous date. Once we have this increase, we should find the max value of this increase for each country and show the result.

Step 1 - first we do a "lag" on "active" cases.

Step 2 - subtract "active" cases and cases on previous date - "lag" value.

Step 3 - group the data based country and find the max value of the subtracted value we recieved in step 2

Screenshot:


Script_Q5.js
Script_Q6.js
Script_Q6_2.js
Script_Q7.js
Script_Q8.js
Script_Q9.js
Script_Q9_2.js
Script_Q10.js
Script_Q11.js
Script_Q12.js
localhost:27017 > BITSMONGODB
Enable Query Assist
Pin All Results
Clear Raw Shell Output
Restart Mongo Shell
Use Legacy Shell
Show Visual Quer
//12]The countries and respective dates in which maximum increase of active cases occurred.
db = db.getSiblingDB("BITSMONGODB");
db.getCollection("FullGrouped").aggregate(
[
{
$setWindowFields: {
partitionBy: "$Country/Region",
sortBy: { Date: 1 },
output: {
lag: {
Raw Shell Output
Aggregate Query (line 4)
50
Documents 1 to 50
Pin Result
Query Code
Explain Query
Table View
FullGrouped > Country/Region
_id
Country/Region
Date
ActiveNewCases
Somalia
2020-05-12T00:...
76
Yemen
2020-06-15T00:...
46
Namibia
2020-07-18T00:...
125
Gambia
2020-07-25T00:...
61
Malaysia
2020-06-04T00:...
249
Peru
2020-07-27T00:...
8484
Philippines
2020-07-08T00:...
2279
Activate Windows
Go to Settings to activate Windows.
1 document selected
Count Documents
00:00:0
12°C
ENG
11:20 PM
12/28/2022