MS3109 - Advanced Database Technologies

MongoDB Assignment: Analysing UN SDG Data for the Public Good

Table of Contents

Project Summary

This project focuses on using NoSQL databases and data visualization to address real-world business, policy, and societal issues. We have been challenged with generating thirty MongoDB queries based on data sets taken from the United Nations Sustainable Development Goals (SDG) Global Database. In this project, our team focused on SDG 11 (Sustainable Cities and Communities), this is important because it focuses on making cities and human settlements inclusive, safe, resilient, and sustainable with over half of the world's population living in urban areas. Our group selected five indicators from SDG 11 investigating – Inadequate housing, Convenient access to public transport, Number of people affected by disaster, National urban policies and Air quality.

Each dataset from the selected indicators will be carefully analysed to develop relevant queries, allowing decision-makers and policymakers, such as those within the UN, to gain insights and implement strategies that enhance urban sustainability, resilience, and inclusivity, ultimately contributing to the creation of safe and sustainable cities for all.

Throughout this assignment, AI tools like ChatGPT guided us in generating ideas and crafting MongoDB queries that provided the most useful and valuable data. We combined this with demonstrations of various operators for educational purposes and conducted our own testing to ensure accuracy and effectiveness.

Matrix for MongoDB Queries

MongoDB NoSQLBooster – Queries 1 - 15															
	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15
Aggregate															
Find															
Distinct															
$project															
$match															
$exists															
$count															
$sort															
$gt															
$unwind															
$limit															
$in															
$group															
$eq															
$sum															
$toInt															
$toDouble															
$ne															
$avg															
$out															
$lookup															
$expr															
$lt															
$ifNull															
$convert															
$max															
$push															
$slice															
$cond															
$gte															

MongoDB NoSQLBooster – Queries 16 - 30

	Q16	Q17	Q18	Q19	Q20	Q21	Q22	Q23	Q24	Q25	Q26	Q27	Q28	Q29	Q30
Aggregate	■	■	■		■			■					■	■	
Find				■		■	■		■	■	■	■			■
Distinct															
$project	■		■		■			■					■	■	
$match	■				■								■	■	
$exists															
$count															
$sort	■		■		■			■					■	■	
$gt				■		■						■			
$unwind	■													■	
$limit					■			■					■		
$in										■				■	
$group	■	■											■	■	
$eq															
$sum		■											■		
$toInt															
$toDouble	■		■		■			■							
$ne					■								■		
$avg	■														
$out	■														
$lookup	■														
$expr						■	■					■			
$lt							■								
$ifNull								■					■		
$convert													■	■	
$max															
$push															
$slice															
$cond															
$gte					■										

Section 1:

Introduction:

Our group has chosen Sustainable Development Goal 11, to make cities and human settlements inclusive, safe, resilient, and sustainable. We selected SDG 11 as it aligns closely with current university initiatives regarding sustainability, urban development, and upholding community values. Additionally, SDG 11 significantly impacts and interacts with numerous other Sustainable Development Goals. Progress in achieving sustainable urban environments directly supports health and well-being (SDG 3), economic growth (SDG 8), reduced inequalities (SDG 10), and climate action (SDG 13). As we delved deeper into our research, we recognized that SDG 11 indicators provide measurable, actionable insights that are critical to shaping effective policies and sustainable practices for policy and decision makers to implement.

The Five Indicators we chose:

1. ***11.1 (Proportion of urban population living in slums or inadequate housing):*** Target 11.1 aims to ensure access for all to adequate, safe, and affordable housing and basic services and upgrade slums by 2030.
2. ***11.2 (Proportion of population with convenient access to public transport, by city)*** Target 11.2 seeks to provide access to safe, affordable, accessible, and sustainable transport systems for all by 2030, with special attention to vulnerable groups.
3. ***11.6.2 (Annual mean levels of fine particulate matter (PM2.5) in cities):*** Target 11.6 aims to reduce the adverse environmental impact of cities by improving air quality and municipal waste management.
4. ***11.5.1 (Number of deaths, missing persons, and directly affected persons attributed to disasters):*** Target 11.5 focuses on significantly reducing the impact of disasters.
5. ***11.A* (Regions that have national urban policies or regional development plans that respond to population dynamics; ensure balanced territorial development; and increase local fiscal space):** Target 11.A supports positive economic, social, and environmental links between urban and rural areas by strengthening national and regional development planning.

Countries/Regions:

11.1: Ireland and LDCS

11.2: Ireland and LDCS

11.5.1: Global (All country groupings)

11.6.2: Ireland and World Total (Regions)

11.A: Global (Regions)

Time frames:

11.1: 2003 - 2023

11.2: 2020

11.5.1: 2005 - 2022

11.6.2: 2010 - 2019

11.A: 2020

Collection	**Indicator**	**Location**	**Timeframe**
Inadequatehousing Ireland	11.1	Ireland	2003 - 2023
Slumsldc	11.1	Least Developed Countries	2003 - 2023
Publictransport Ireand	11.2	Ireland	2020
Publictransportldc	11.2	Least Developed Countries	2020
Naturaldisasters	11.5.1	World (All Countries)	2005 - 2022
Nationalurban policies	11.A	World (In Regions)	2020
Fineparticulate matterIreland	11.6.2	Ireland	2010 - 2019
Fineparticulate matterworld	11.6.2	World (In Regions)	2010 - 2019

Section 2:

Data Sources:

1. Indicator 11.1: Proportion of urban population living in inadequate housing/slums.

2. Indicator 11.2: Proportion of population that has convenient access to public transport

3. Indicator 11.5.1: Number of people affected by disaster

4. Indicator 11.6.2: Annual mean levels of fine particulate matter (PM2.5) in urban areas.

5. Indicator 11.A: National urban policies or regional development plans incorporating population projections

To ensure the highest level of accuracy in our analysis, we opted to collect all our data from a single, reliable source to avoid potential discrepancies, inaccuracies, or gaps that can arise when combining datasets from multiple sources. Given our focus on SDG 11, a Sustainable Development Goal established by the United Nations, we utilised the United Nations Statistics & SDG Indicator Database as our primary data source. This approach ensures consistency and reliability in our findings. The link to the following databases is: https://unstats.un.org/sdgs/dataportal/database.

Our data on SDG 11: Sustainable Cities and Communities demonstrate high quality, with well-documented records from multiple years, ensuring consistency and reliability. The United Nations Statistics & SDG Indicator Database provides a wealth of information, often available in yearly or five-year intervals, making trend analysis straightforward. Indicators such as 11.1.1 (Proportion of urban population living in slums) and 11.6.2 (Annual mean levels of fine particulate matter in cities) offered extensive datasets, categorized by factors like region, population group, and measurement units. This structured approach to data collection provides confidence, enabling well-informed decisions for building more resilient, inclusive, and sustainable urban environments.

We decided to select and compare data from the Least Developed Countries when generating our queries, as this approach is valuable and highlights the unique challenges these nations face in achieving SDG 11. Least Developed Countries often experience higher rates of urban poverty, weaker infrastructure, and greater environmental risks, making their data essential for understanding inequalities.

A challenge we faced with the datasets in Excel was ensuring the data was in the correct format for easier integration into MongoDB. Specifically, we had to select the Excel sheet where the data was structured in “Record Format” rather than “Table Format”. This was crucial because when saving the data as a CSV file and importing it into MongoDB Compass, the “Record Format” preserved the data structure, enabling us to perform queries effectively. In contrast, the “Table Format” caused issues during testing and was unsuitable for our needs. This approach proved particularly beneficial when using tools like https://beautifytools.com/ to convert the CSV file to JSON and https://transform.tools/json-to-json-schema to convert the JSON file into a JSON Schema. This was helpful as it allowed ChatGPT to better understand the datasets and generate queries from them. We also struggled to find datasets for a specific time with our data ranging from 2003-2022.

Section 3: MongoDB queries:

Query 1

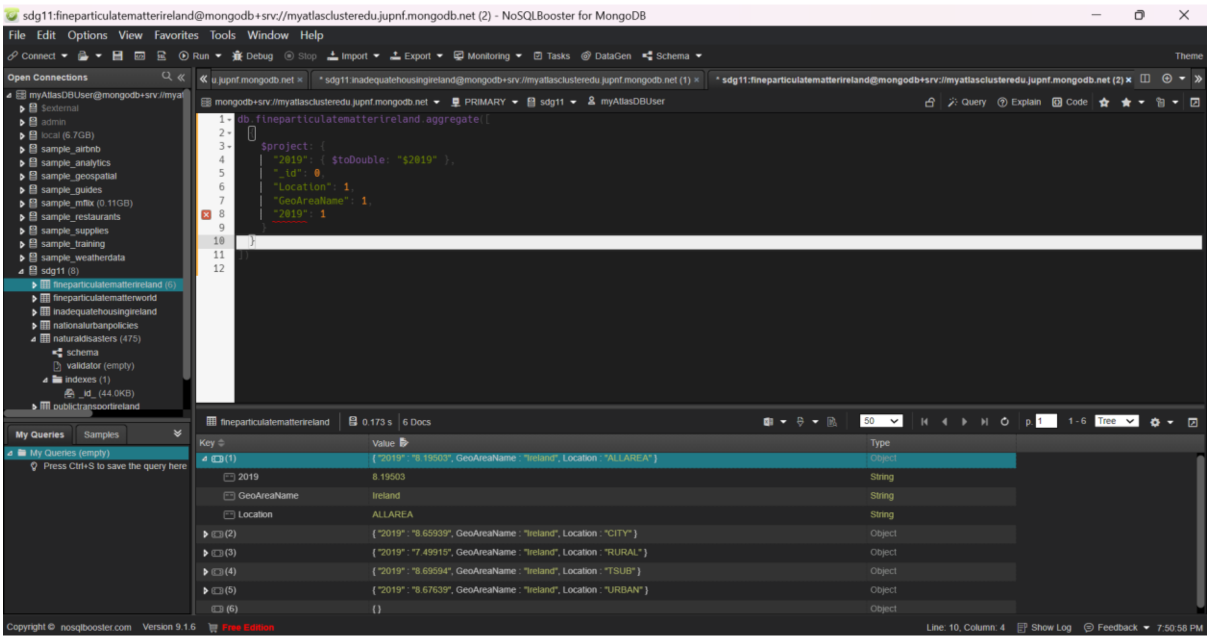
Fine particulate matter Ireland

Retrieves all records from the year 2019

Query Reasoning: We decided to start off with a simple query that retrieves all the records from 2019. This first step serves multiple purposes: it allows us to quickly verify that our dataset contains the expected data for that year, and it ensures that the 2019 field is correctly formatted for subsequent numerical analyses. By confirming the completeness and consistency of the records for 2019, we set a reliable foundation for any further data manipulation.

Code:

```
db.fineparticulatematterireland.aggregate([
 {
  $project: {
   "2019": { $toDouble: "$2019" },
   "_id": 0,
   "Location": 1,
   "GeoAreaName": 1,
   "2019": 1
  }
 }
])
```



Explanation of Results: The results show that urban areas in Ireland had higher levels of fine particulate matter compared to rural areas in 2019. This confirms that air pollution is more concentrated in populated areas, while rural regions experience lower levels.

Query 2

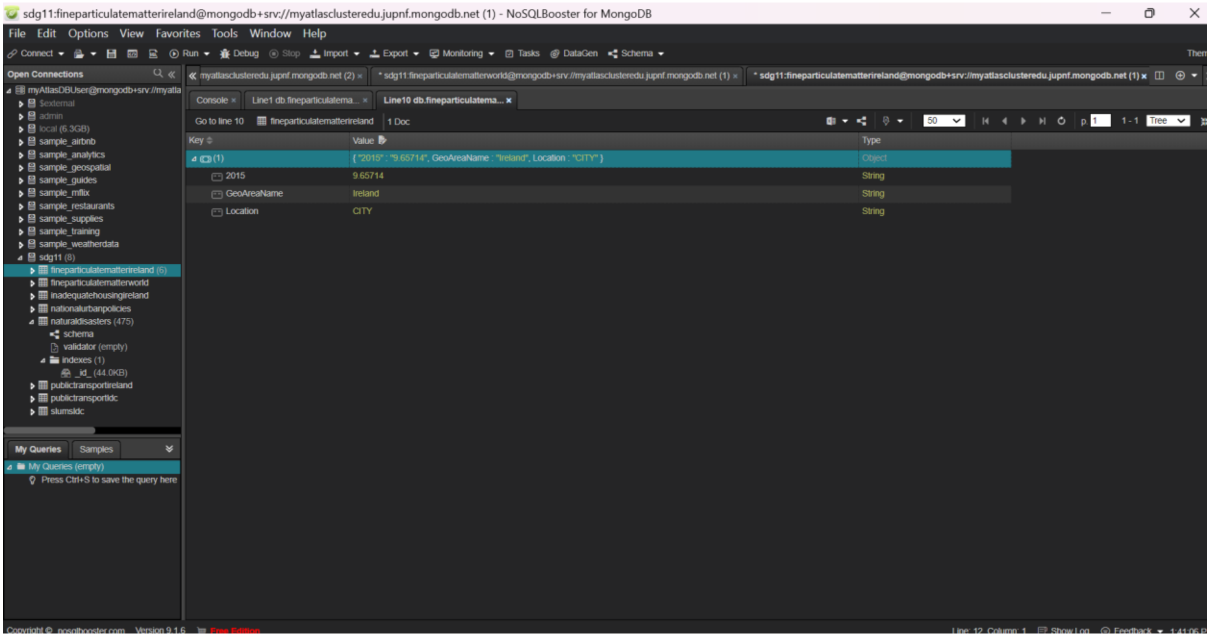
Fine particulate matter Ireland

Shows the particulate matter level in 2015 for cities.

Query Reasoning: This query retrieves PM2.5 levels for Irish cities in 2015, allowing for a direct comparison with global urban air pollution levels. By filtering for "GeoAreaName": "Ireland" and "Location": "CITY", the query ensures that only city-specific data from Ireland is shown.

<u>Code:</u>

```
db.fineparticulatematterireland.find({
  "GeoAreaName": "Ireland",
  "Location": "CITY"
}, {
  "GeoAreaName": 1,
  "Location": 1,
  "2015": 1,
  "_id": 0
})
```



Explanation of Results: This query retrieves PM2.5 levels in 2015 specifically for city areas in Ireland, making it comparable to the global city average for the next query.

Query 3

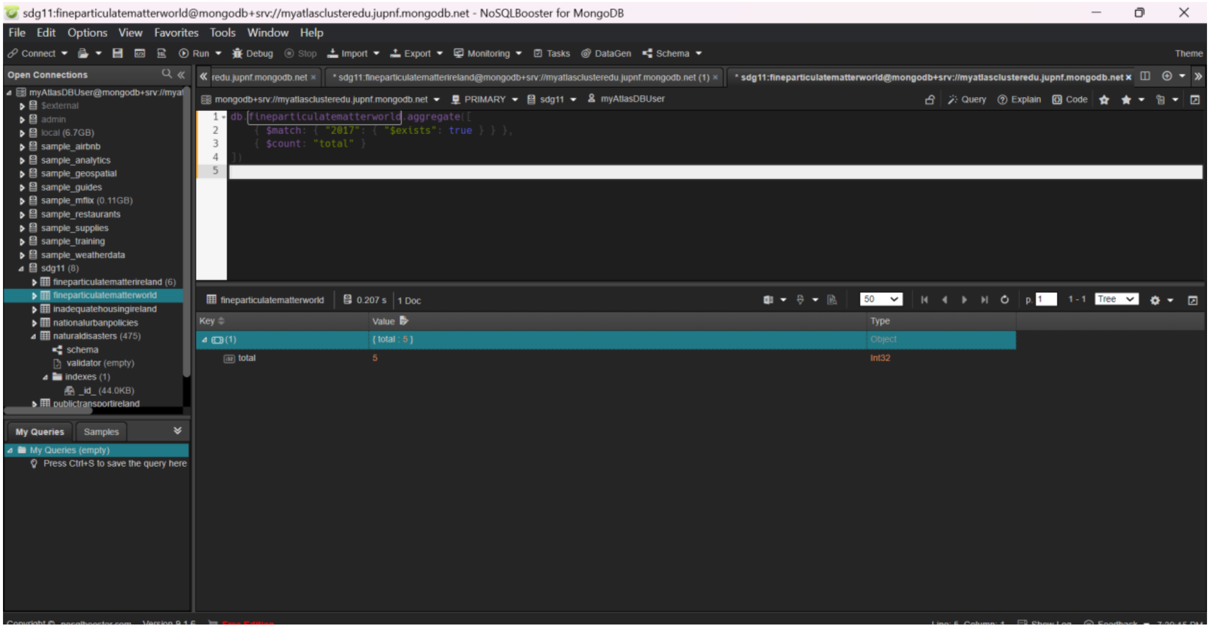
Fine particulate matter world

Counts entries for a particular year (e.g., 2017)

Query Reasoning: We decided to implement another simple query to quickly assess the data coverage for a specific year—in this case, 2017. By filtering on the existence of the "2017" field, the query verifies that the dataset includes the relevant records for that year. The use of $count then provides a straightforward metric, showing us, exactly how many entries are available for analysis.

Code:

```
db.fineparticulatematterworld.aggregate([
  { $match: { "2017": { "$exists": true } } },
  { $count: "total" }
])
```



Explanation of Results: This shows that dataset contains 5 records for fine particulate matter levels in the world for 2017. This is because this dataset is split into regions rather than by country.

Query 4

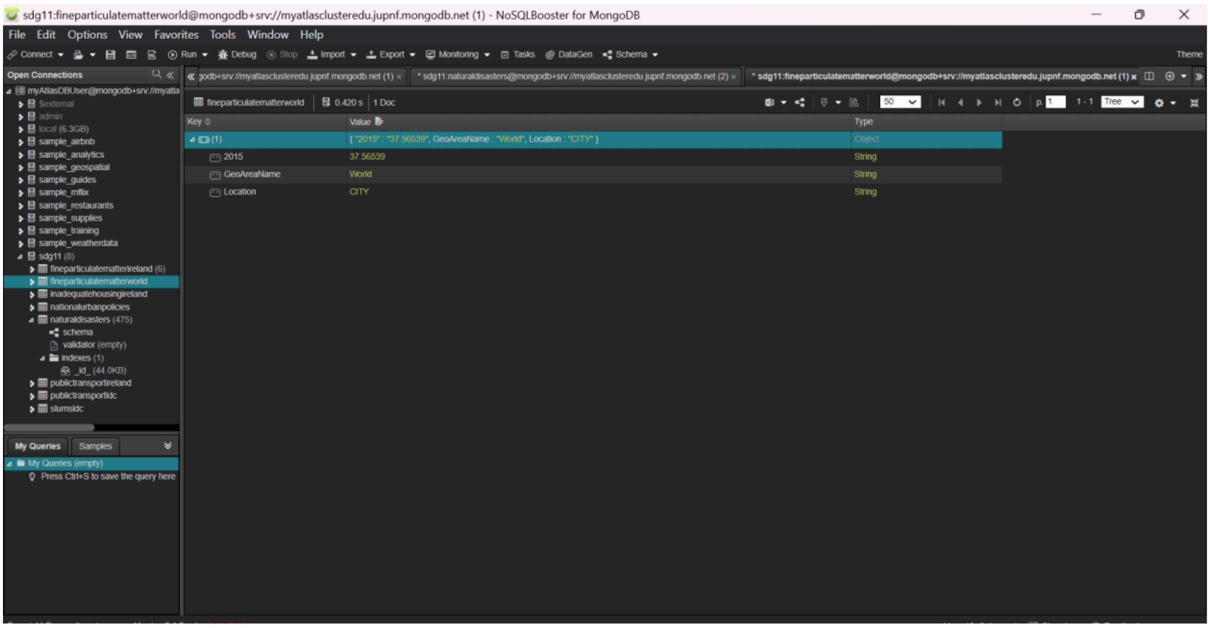
Fine particulate matter world

Retrieves PM2.5 levels in 2015 for CITY areas globally

Query reasoning: This query retrieves PM2.5 levels in 2015 for city areas globally, allowing for a direct comparison with Ireland's air quality. Since Ireland's urban PM2.5 levels looked low, we wanted to analyse how pollution levels in global cities compare to Ireland's urban environment.

<u>Code:</u>

```
db.fineparticulatematterworld.find({
  "GeoAreaName": "World",
  "Location": "CITY"
}, {
  "GeoAreaName": 1,
  "Location": 1,
  "2015": 1,
  "_id": 0
})
```



Explanation of results: In 2015, the PM2.5 level for city areas globally was much higher than, Ireland's. This suggests that global cities generally experienced higher air pollution than Ireland, likely due to higher population density, industrial activity, and traffic congestion.

Query 5

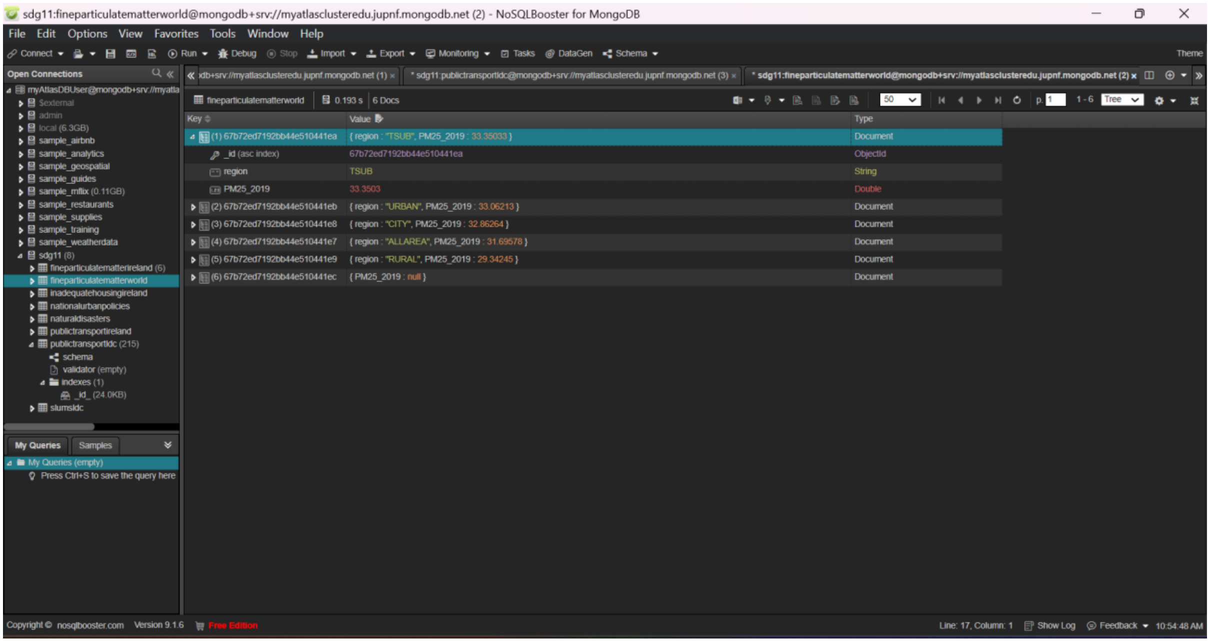
Fine particulate matter world

Comparing PM2.5 Levels Across Regions in 2019

Query reasoning: Given the limitations of our dataset, we cannot directly match city-level public transport data with air pollution levels. There is also a one-year gap between the datasets. However, we can still examine how air pollution levels differ between urban, suburban, and rural regions globally. By retrieving PM2.5 levels for different regions in 2019, we can explore whether urban areas—where public transport is often more widely available—tend to have lower or higher pollution levels compared to suburban and rural areas, where reliance on private vehicles may be greater. This can provide insights into how transportation infrastructure impacts pollution on a broader scale.

Code:

```
db.fineparticulatematterworld.aggregate([
  {
    $project: {
      region: "$Location",
      PM25_2019: { $toDouble: "$2019" }
    }
  },
  {
    $match: {
      PM25_2019: { $ne: NaN } // Ensure only numeric values remain
    }
  },
  {
    $sort: { PM25_2019: -1 } // Sorting in descending order to see the most polluted areas
  }
])
```



Explanation of Results: The results show that suburban (33.35 m) and urban (33.06 m) areas have the highest PM2.5 levels, while rural areas (29.34 m) have the lowest. Cities (32.86 m) fall slightly below suburban and urban pollution levels, and the global average (31.69 m) sits in between. This suggests that higher urbanization is linked to increased pollution, likely due to traffic, industry, and energy use. The fact that suburban areas have the highest pollution may indicate greater reliance on private vehicles due to limited public transport, reinforcing our earlier hypothesis that poor transit access can contribute to environmental burdens.

Query 6

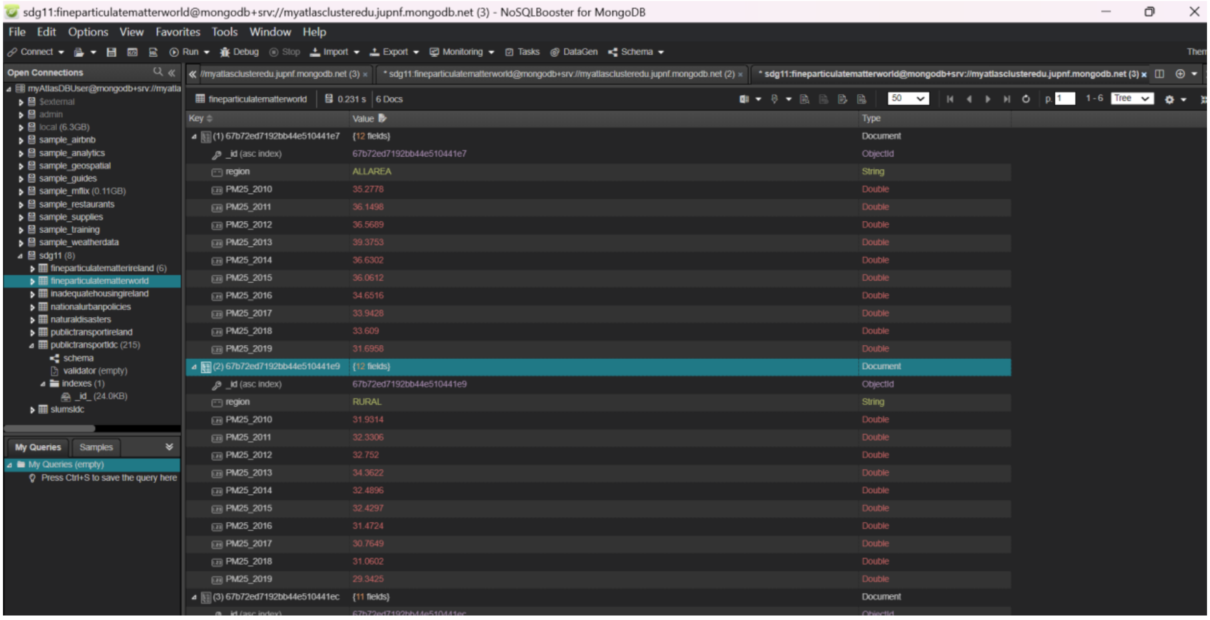
Fine particulate matter world

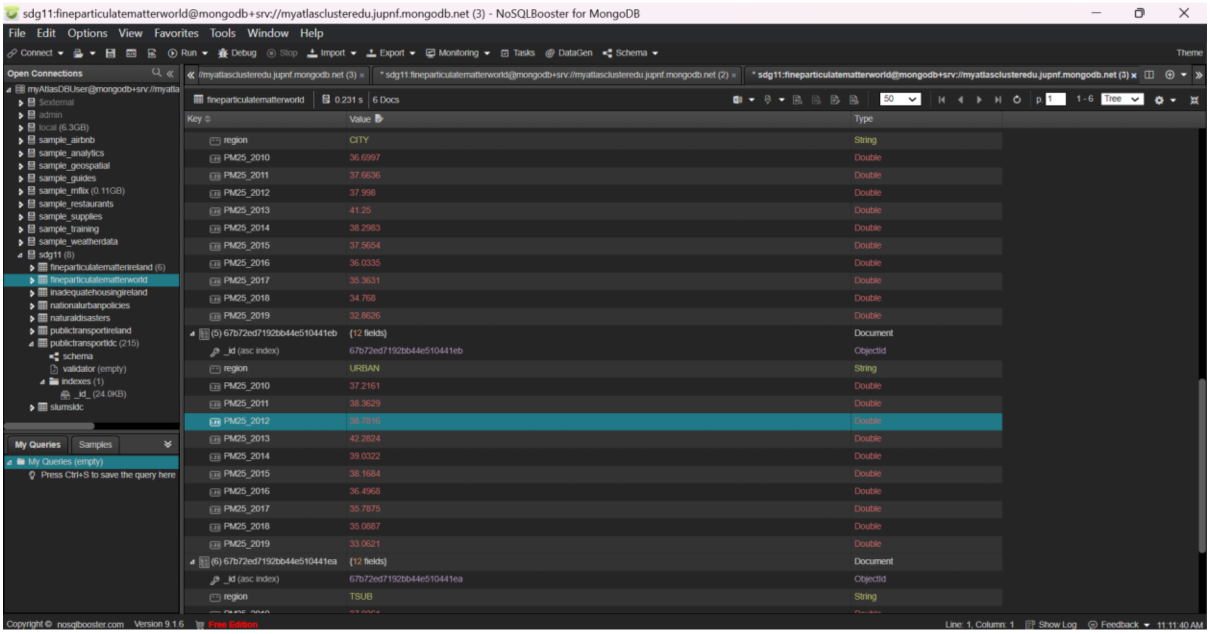
Air Pollution Trends Over Time

Query Reasoning: This query examines how PM2.5 levels have changed over time in different regions. By analysing yearly trends from 2010 to 2019, we can determine whether air quality has improved or worsened over the past decade. This insight helps contextualise whether shifts in transportation infrastructure have had a measurable impact on pollution levels. While we cannot directly link this data to public transport access, identifying regions with worsening or improving pollution can help inform discussions on sustainable urban development.

Code:

```
db.fineparticulatematterworld.aggregate([
  {
    $project: {
      region: "$Location",
      PM25_2010: { $toDouble: "$2010" },
      PM25_2011: { $toDouble: "$2011" },
      PM25_2012: { $toDouble: "$2012" },
      PM25_2013: { $toDouble: "$2013" },
      PM25_2014: { $toDouble: "$2014" },
      PM25_2015: { $toDouble: "$2015" },
      PM25_2016: { $toDouble: "$2016" },
      PM25_2017: { $toDouble: "$2017" },
      PM25_2018: { $toDouble: "$2018" },
      PM25_2019: { $toDouble: "$2019" }
    }
  },
  {
    $unwind: {
      path: "$region",
      preserveNullAndEmptyArrays: true
    }
  }
])
```





Explanation of results: The results show PM2.5 trends from 2010 to 2019 across different regions, revealing that air pollution levels have generally decreased over the decade, showing improvement for SDG 11. The persistently higher levels in urban and suburban areas suggest that transportation and industrial activity continue to be major pollution sources, aligning with our earlier analysis that public transport access may influence pollution levels.

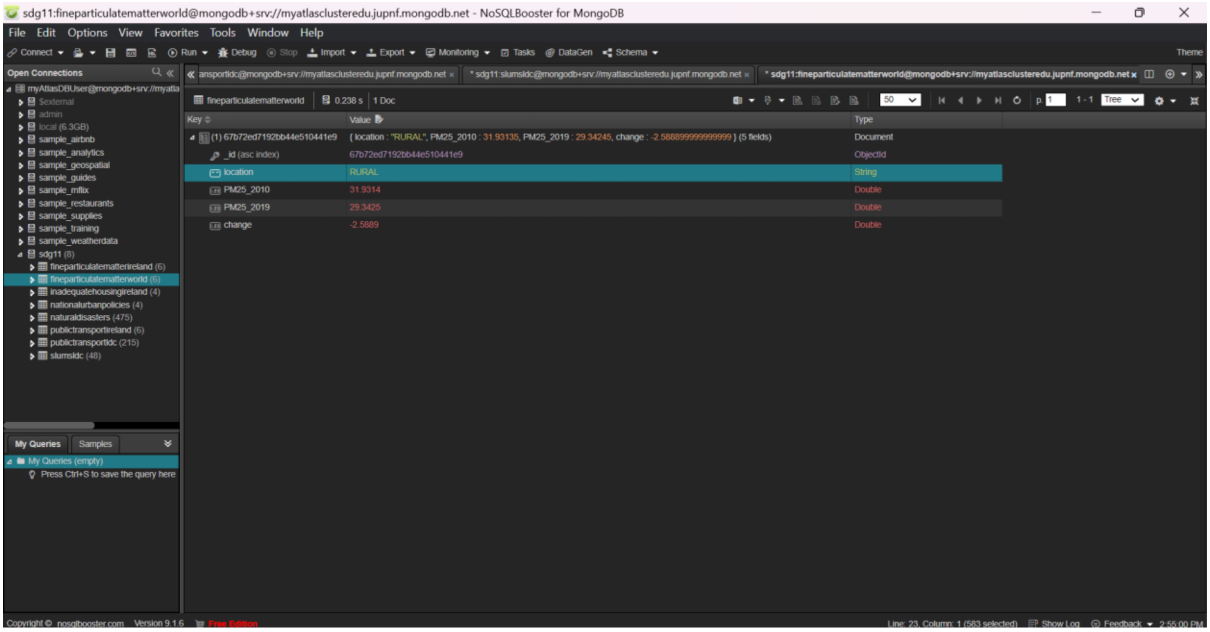
Query 7

Fine particulate matter world

Identifying the Region with the Highest Increase in PM2.5 Levels

Query Reasoning: This query analyses how PM2.5 pollution levels changed from 2010 to 2019 across different regions, identifying the area with the highest increase or decrease in air pollution. By calculating the difference in pollution levels over the decade, we can determine whether certain areas have improved or worsened in air quality. This connects to our findings on public transport and slum populations , as both factors can influence pollution levels. Additionally, understanding pollution trends helps policymakers assess the effectiveness of environmental regulations and urban planning.

Code:
```
db.fineparticulatematterworld.aggregate([
  {
    $project: {
      location: "$Location",
      PM25_2010: { $toDouble: "$2010" },
      PM25_2019: { $toDouble: "$2019" },
      change: { $subtract: [{ $toDouble: "$2019" }, { $toDouble: "$2010" }] }
    }
  },
  {
    $match: {
      PM25_2010: { $ne: NaN },
      PM25_2019: { $ne: NaN }
    }
  },
  {
    $sort: { "change": -1 }
  },
  {
    $limit: 1
  }
]);
```



Explanation of results: The results show that rural areas actually experienced a decrease in PM2.5 levels from 31.93 m in 2010 to 29.34 m in 2019, a reduction of 2.59 m. This suggests that air quality in rural regions improved over the decade, possibly due to reduced deforestation, lower industrial activity, or improved emission regulations. In contrast, previous queries have shown that urban and suburban areas tend to experience higher pollution, likely due to vehicle emissions, industrialization, and population density. This result reinforces the idea that urbanisation and poor transport infrastructure contribute significantly to worsening air quality, emphasizing the need for sustainable urban planning and expanded public transport systems.

Query 8

National urban policies

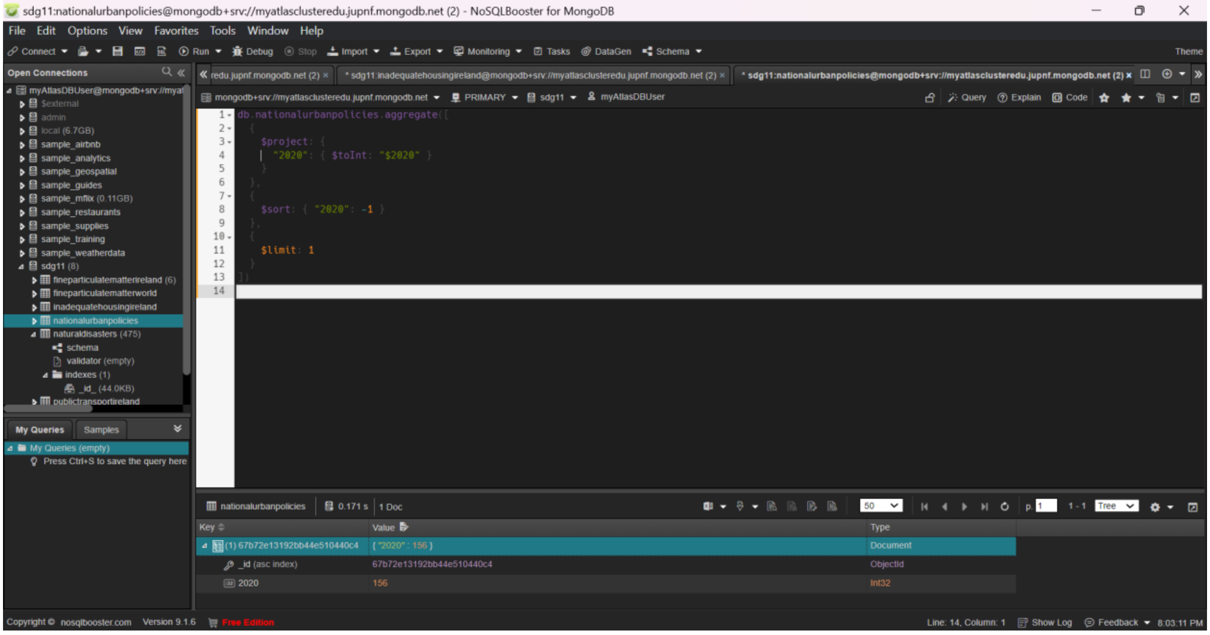
Find the maximum value for "2020" across all documents.

Query Reasoning: This process helps us quickly pinpoint the maximum policy measure or indicator for 2020, laying the groundwork for further analysis or comparisons, applying a limit of 1 extract only the top record.

Code:

```
db.nationalurbanpolicies.aggregate([
  {
    $project: {
      "2020": { $toInt: "$2020" }
    }
  },
  {
    $sort: { "2020": -1 }
  },
```

```
  {
    $limit: 1
  }
])
```



Explanation of Results: In 2020, the highest recorded value for national urban policies was 156. This suggests that at least one area had a significant number of policies or measures in place to address urban development, territorial balance, and population dynamics.

Query 9

National urban policies

Comparison of National Urban Policies: North America & Europe vs. Sub-Saharan Africa

Query Reasoning: The goal of this query is to compare the total number of national urban policies reported in 2020 between two major regions: North America & Europe and Sub-Saharan Africa. This allows us to analyse differences in urban policy development and assess regional progress in sustainable urbanization. North America & Europe generally have well-established urban policies, reflecting long-standing investments in infrastructure, housing, and environmental planning. In contrast, Sub-Saharan Africa faces rapid urbanization, where national policies play a crucial role in managing informal settlements, infrastructure gaps, and climate resilience.

Code:

```
db.nationalurbanpolicies.aggregate([ { $match: { "GeoAreaName": { $in: ["Europe and Northern America", "Sub-Saharan Africa"] } } }, { $group: { _id: { $cond: { if: { $eq: ["$GeoAreaName", "Sub-Saharan Africa"] }, then: "Sub-Saharan Africa", else: "North America & Europe" } }, totalPolicies: { $sum: { $toInt: "$2020" } } } }
])
```

(1) North America & Europe	{ totalPolicies : 37 }	Document
_id (asc index)	North America & Europe	String
totalPolicies	37	Int32
(2) Sub-Saharan Africa	{ totalPolicies : 40 }	Document
_id (asc index)	Sub-Saharan Africa	String
totalPolicies	40	Int32

Explanation of Results: The output presents two records, one for North America and Europe and another for Sub-Saharan Africa. Each one displays the total number of urban policies implemented in 2020. Based on the example data, the North America and European group has a total of 37 policies, while Sub-Saharan Africa has 40 policies. This result provides a comparative view of national urban policy adoption between these two broad regions. This insight can be useful for policymakers and urban planners trying to analyse regional trends in urban policy development.

Query 10

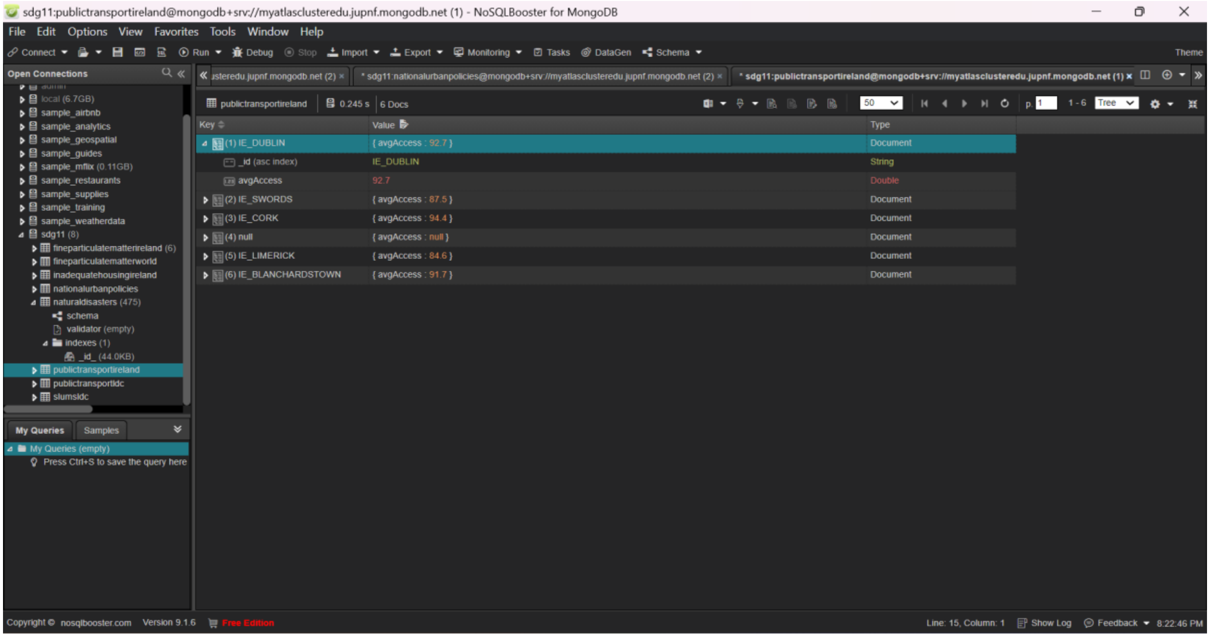
Public transport, Ireland

Finds the average proportion of the population with access to public transport for each city

Query Reasoning: This query computes the average proportion of the population with access to public transport for each city. By converting the "2020" field to a numeric format and grouping by city, we can benchmark and compare urban transport accessibility, helping identify areas that may need policy intervention.

Code:

```
db.publictransportireland.aggregate([
  {
   $project: {
     Cities: 1,
     "2020": { $toDouble: "$2020" }
   }
  },
  {
   $group: {
     _id: "$Cities",
     avgAccess: { $avg: "$2020" }
   }
  }
])
```



Explanation of Results: This result highlights that public transport access in Ireland varies by city, with Cork having the highest average access at 94.4%, followed by Dublin at 92.7% and Blanchardstown at 91.7%. Limerick has the lowest among recorded cities at 84.6%. The data suggests strong public transport availability in major urban areas, with slight variations between cities.

Query 11

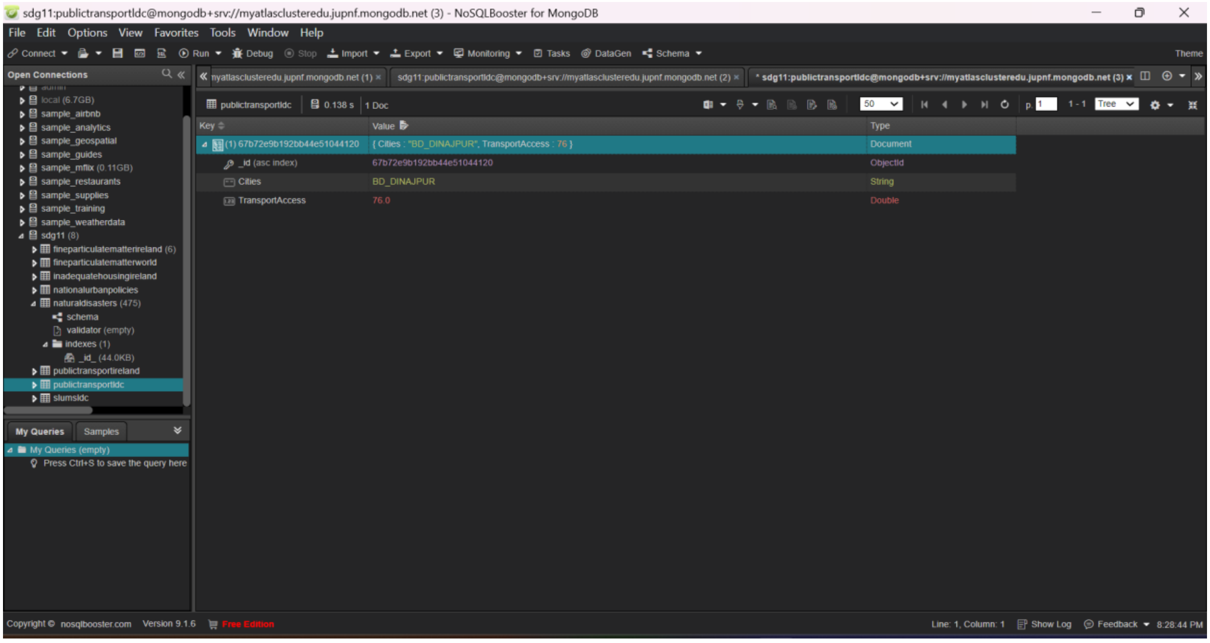
Public transport, least developed countries

Find the city in Bangladesh with the highest public transport access

Query Reasoning: This query identifies the city in Bangladesh with the highest public transport access by filtering out empty 2020 values, converting them to numeric, and sorting in descending order. This result can be compared with the average public transport access in Ireland (from Query 10) to benchmark differences between regions.

Code:

```
db.publictransportldc.aggregate([
  { $match: { "GeoAreaName": "Bangladesh", "2020": { $ne: "" } } },
  { $project: { Cities: 1, TransportAccess: { $toDouble: "$2020" } } },
  { $sort: { TransportAccess: -1 } },
  { $limit: 1 }
])
```



Explanation of Results: Dinajpur (population = 3,315,236) is the city in Bangladesh with the highest public transport access and has a value of 76%. While this shows relatively good accessibility within the country, it still falls short compared to Irish cities, where access ranges from 84.6% in Limerick to 94.4% in Cork. This highlights a gap in urban transport infrastructure, emphasizing the need for improved accessibility in developing regions to align with SDG 11's goal of sustainable cities and communities.

Query 12

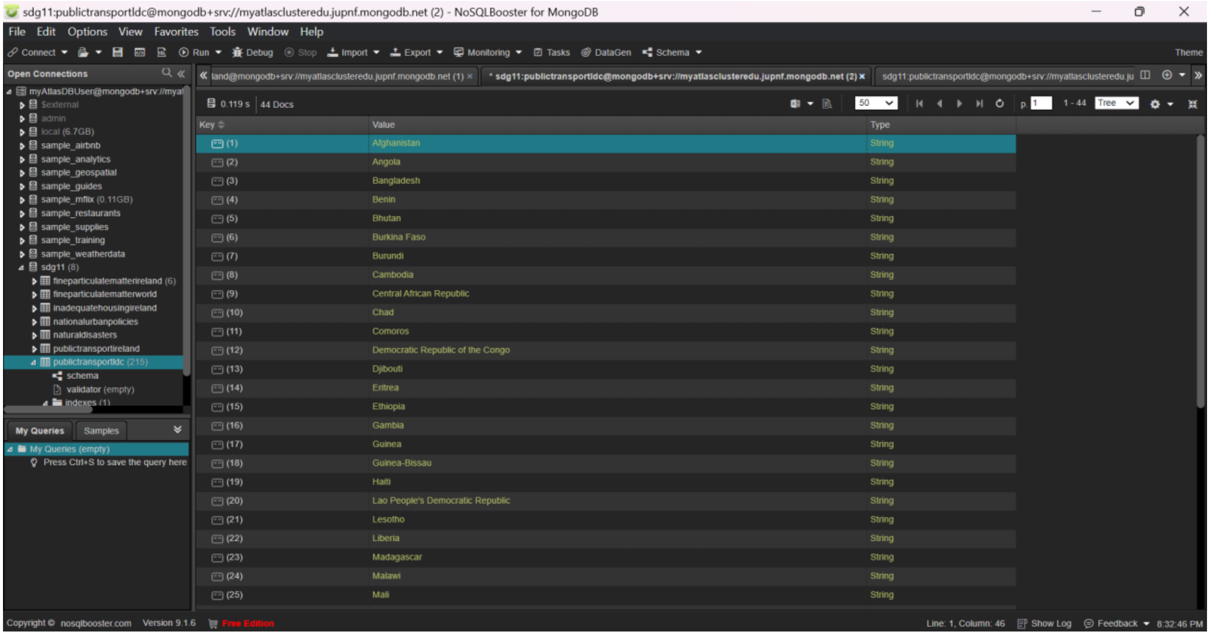
Public transport, least developed countries

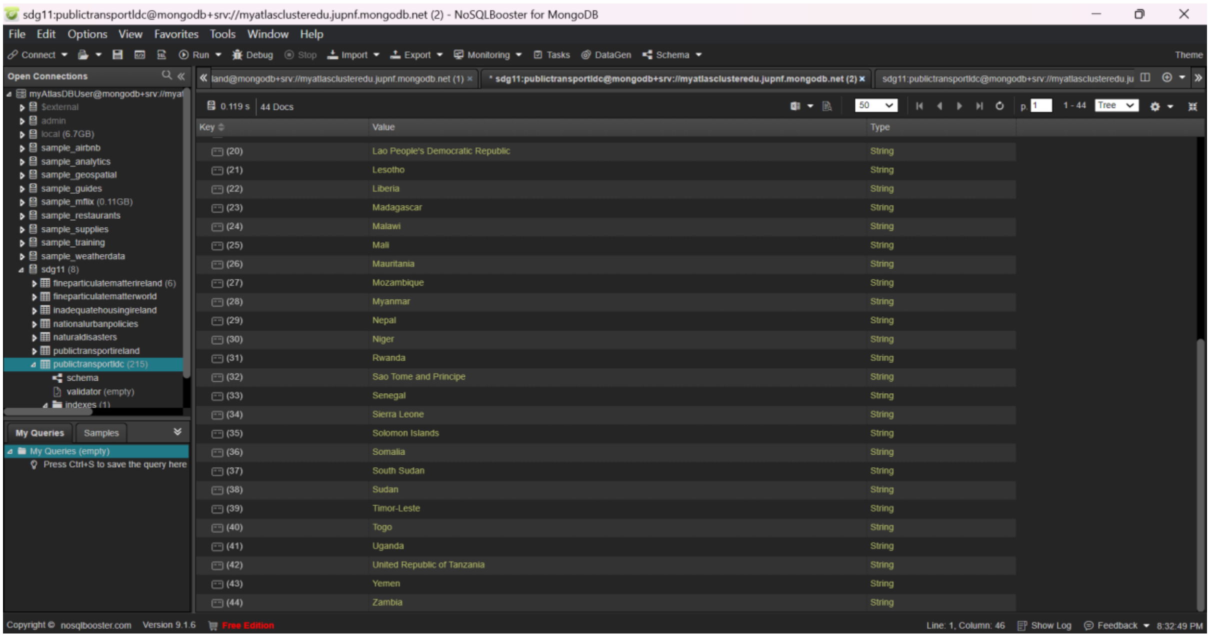
This will return a list of all the countries available in the dataset.

Query Reasoning: This query extracts a distinct list of countries from the dataset by retrieving unique values of the "GeoAreaName" field. It provides a clear view of the dataset's geographic coverage, which is essential for understanding and comparing regional differences in public transport access.

<u>Code:</u>

```
db.publictransportldc.distinct("GeoAreaName")
```





Explanation of Results: This dataset includes public transport access data for the 43 least developed countries according to UN.

Query 13

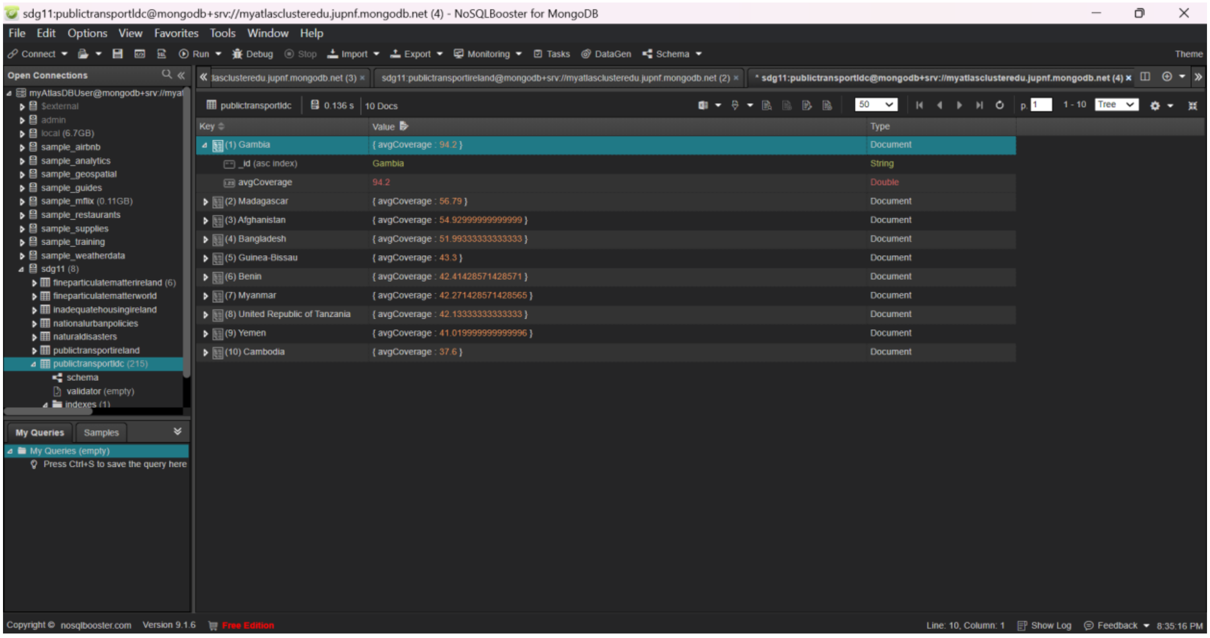
Public transport, least developed countries

Calculates the average public transport access and returns the top 10 countries with the highest access.

Query Reasoning: This query calculates the average public transport access for each country by converting the 2020 values to a numeric format, then groups by country and sorts them in descending order. Limiting the output to the top 10 highlights the countries with the best public transport accessibility, offering insights into regional performance and potential areas for policy benchmarking.

<u>Code:</u>

```
db.publictransportldc.aggregate([
  { $match: { "2020": { $ne: "" } } },
  { $group: {
    _id: "$GeoAreaName",
    avgCoverage: { $avg: { $toDouble: "$2020" } }
  } },
  { $sort: { avgCoverage: -1 } },
  { $limit: 10 }
])
```



Explanation of Results: Gambia leads with the highest average public transport access among least developed countries at 94.2%, significantly ahead of others. Madagascar (56.79%), Afghanistan (54.93%), and Bangladesh (51.99%) also show relatively strong access, while Cambodia ranks lowest in the top ten at 37.6%. These figures highlight disparities in transport infrastructure, reinforcing the importance of SDG 11 in ensuring equitable urban mobility.

Query 14

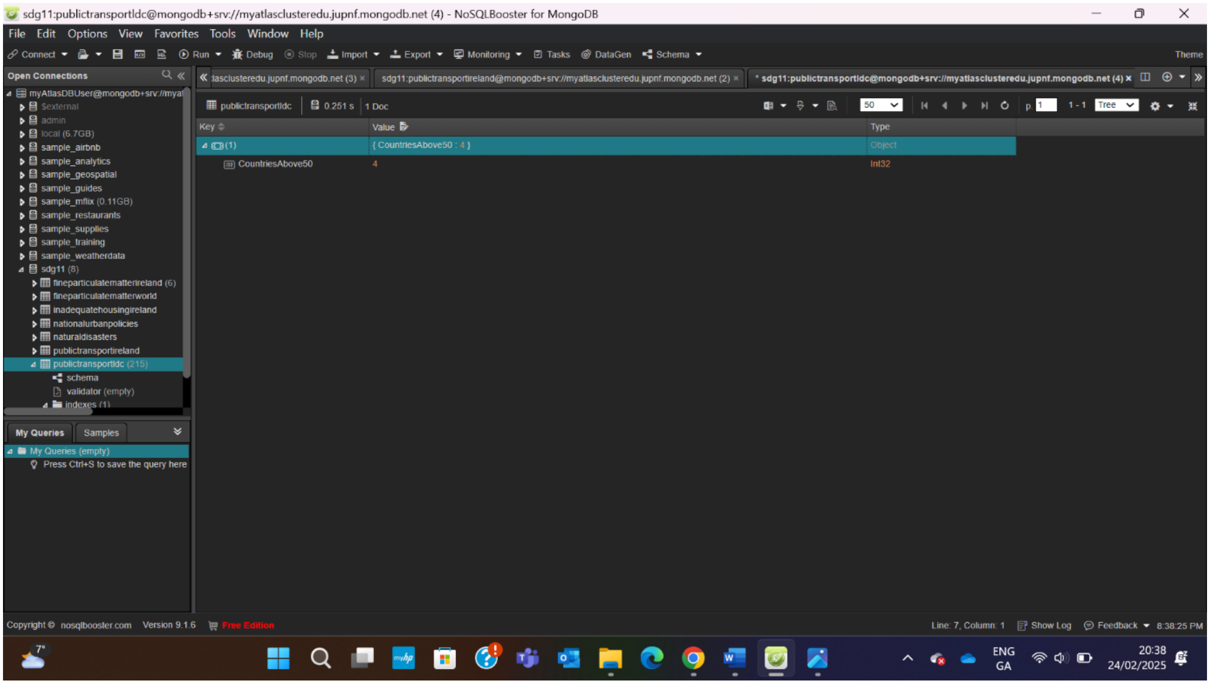
Public transport, least developed countries

This counts the number of countries where the average public transport access is greater than 50%.

Query Reasoning: This query calculates the average public transport access for each country and then filters for those with at least 50% access. Counting these countries helps us gauge how many regions meet a benchmark for adequate public transport, providing a metric for comparing performance and guiding policy decisions.

<u>Code:</u>

```
db.publictransportldc.aggregate([
   { $match: { "2020": { $ne: "" } } },
   { $group: { _id: "$GeoAreaName", avgCoverage: { $avg: { $toDouble: "$2020" } } } },
   { $match: { avgCoverage: { $gte: 50 } } },
   { $count: "CountriesAbove50" }
])
```



Explanation of Results: This shows that only four of the least developed countries have an average public transport access above 50%, highlighting significant gaps in urban mobility and the need for improved infrastructure to meet SDG 11 goals. While also showing that Query 13's findings were correct.

Query 15

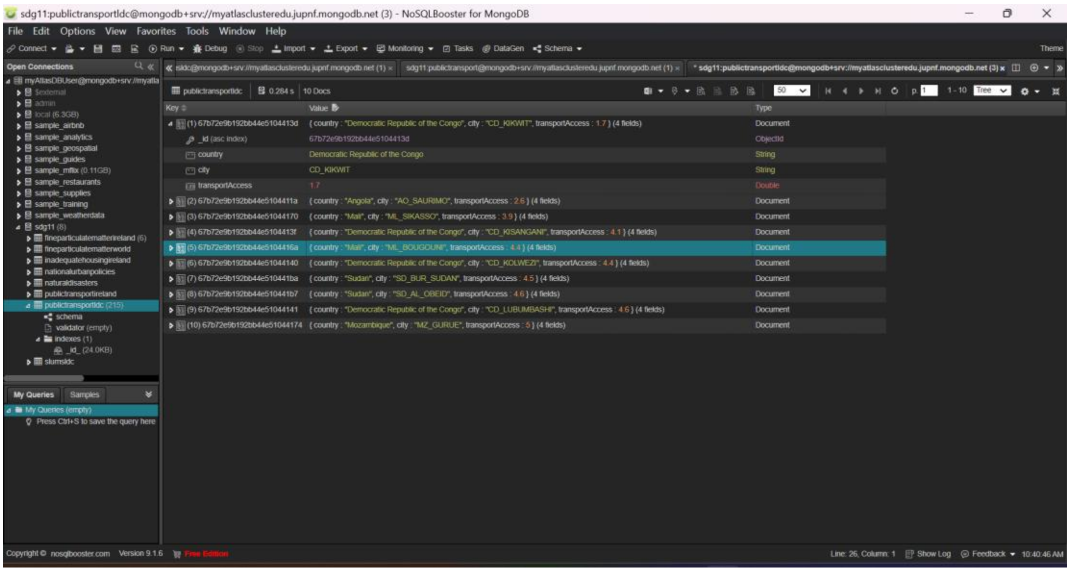
Public transport, least developed countries

Finds 10 Cities with Lowest Public Transport Access

Query Reasoning: This query aims to identify the 10 cities with the lowest proportion of their population having convenient access to public transport. Limited public transport access can indicate infrastructure deficiencies, urban planning challenges, and a higher reliance on private vehicles, which we believe can contribute to higher air pollution levels (PM2.5).

Code:

```
db.publictransportldc.aggregate([
  {
    $match: {
      "2020": { $ne: "", $ne: null }
    }
  },
  {
    $project: {
      country: "$GeoAreaName",
      city: "$Cities",
      transportAccess: { $toDouble: "$2020" }
    }
  },
  {
    $match: {
      transportAccess: { $ne: NaN }
    }
  },
  {
    $sort: { transportAccess: 1 }
  },
  {
    $limit: 10
  }
])
```



Explanation of Results: The results highlight significant disparities in public transport access, with the Democratic Republic of Congo, Mali, Angola, and Sudan having some of the lowest-ranked cities. Kikwit, DRC, has the worst access at just 1.7%, suggesting severe infrastructure deficiencies. Other cities, like Saurimo (Angola) and Sikasso (Mali), also face major transportation challenges, likely due to inadequate investment or reliance on informal transit systems. The dominance of these countries in the bottom rankings may indicate systemic issues in urban planning and government policies. Limited public transport could force greater reliance on private vehicles, potentially leading to higher air pollution levels, a relationship that can be further explored by analysing PM2.5 data for these locations.

Query 16

Public transport, least developed countries

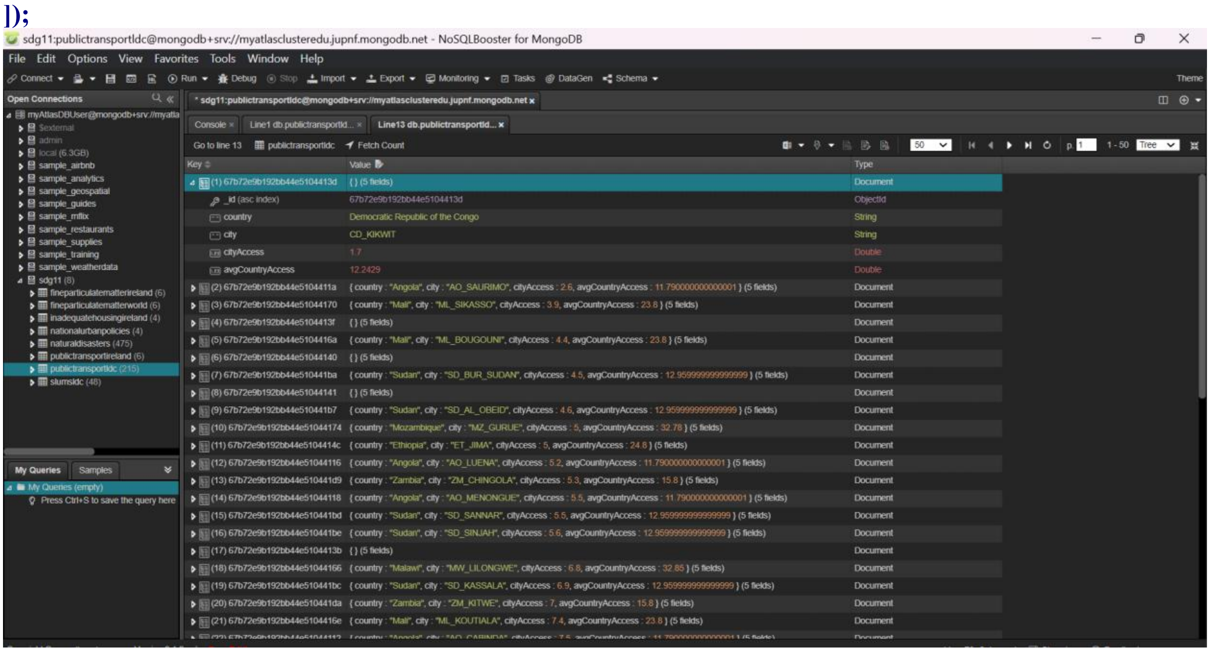
Identifying Cities in Least Developed Countries (LDCs) with Public Transport Access Below Their National Average

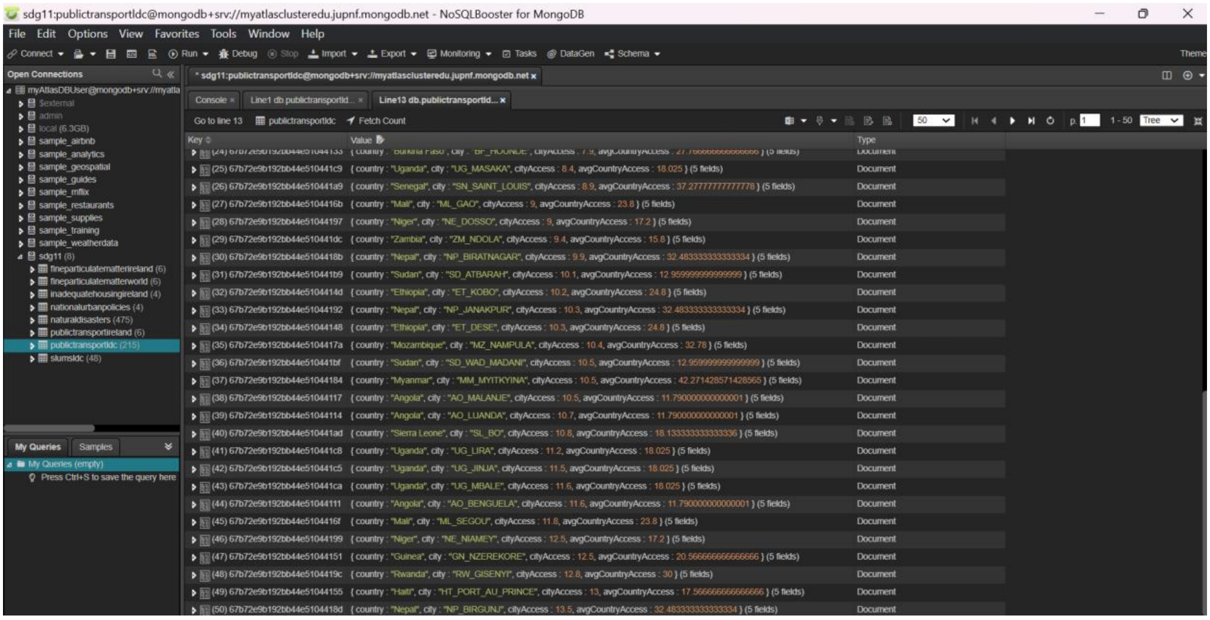
Query Reasoning: This query identifies cities in Least Developed Countries (LDCs) where public transport access is below their national average. It first calculates the average public transport accessibility for each country and stores it in a temporary collection. Then, it compares each city's access level against its country's average and filters out those that fall below the benchmark. This helps pinpoint urban areas with insufficient public transport infrastructure, guiding us toward cities and countries needing investment and improvement.

Code:
```
db.publictransportldc.aggregate([
  {
    $group: {
      _id: { country: "$GeoAreaName" },
      avgAccess: { $avg: { $toDouble: "$2020" } }
    }
  },
  {
    $out: "temp_avg_access"
  }
]);

db.publictransportldc.aggregate([
  {
    $lookup: {
      from: "temp_avg_access",
      localField: "GeoAreaName",
      foreignField: "_id.country",
      as: "countryAvg"
    }
  },
  {
    $unwind: "$countryAvg"
  },
  {
    $project: {
      country: "$GeoAreaName",
      city: "$Cities",
      cityAccess: { $toDouble: "$2020" },
      avgCountryAccess: "$countryAvg.avgAccess"
```

```
      }
    },
    {
      $match: {
        $expr: { $lt: ["$cityAccess", "$avgCountryAccess"] }
      }
    },
    {
      $sort: { "cityAccess": 1 }
    }
]);
```





Explanation of result: The results highlight Zambia, the Democratic Republic of the Congo (DRC), and Angola as having some of the lowest public transport accessibility relative to their national averages. DRC appears multiple times in the top-ranked cities with the lowest access, reinforcing its systemic transport infrastructure challenges, as previously observed in Query 14. Cities like Kikwit (1.7%) and Kisangani (4.1%) in DRC fall well below the country's average of 12.24%, showing a significant disparity in urban mobility. Similarly, Angola, Mali, Sudan, and Mozambique also have cities where public transport coverage is far lower than their national averages, indicating regional inequalities in transit accessibility. The presence of multiple cities from the same countries suggests that certain nations struggle with consistent public transport development across different urban areas, pointing to potential policy gaps or resource allocation issues.

Query 17

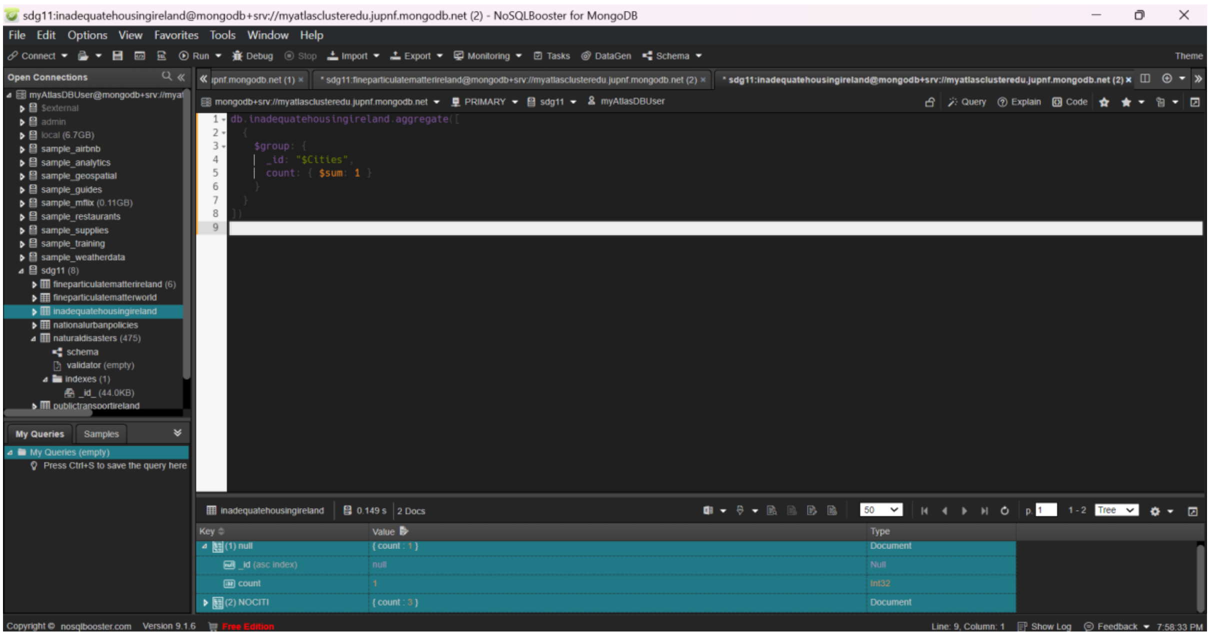
Inadequate housing Ireland

Counts the number of documents for each "Cities" field

Query Reasoning: This is also a pretty simple query as we decided to use the first few queries to flesh out our knowledge of our data and NOSQLbooster. By grouping data based on the "Cities" field and summing the occurrences, this query provides an overview of the dataset's structure.

Code:

```
db.inadequatehousingireland.aggregate([
  {
    $group: {
      _id: "$Cities",
      count: { $sum: 1 }
    }
  }
])
```



Explanation of Results: This shows that the dataset contains data for two categories: "NOCITI" appears three times, while one entry has no city specified.

Query 18

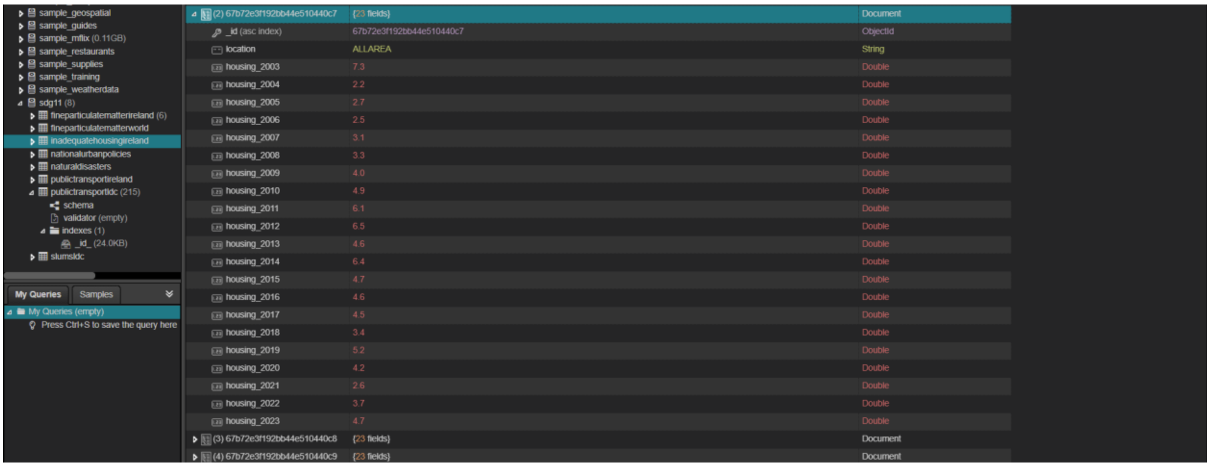
Inadequate housing Ireland

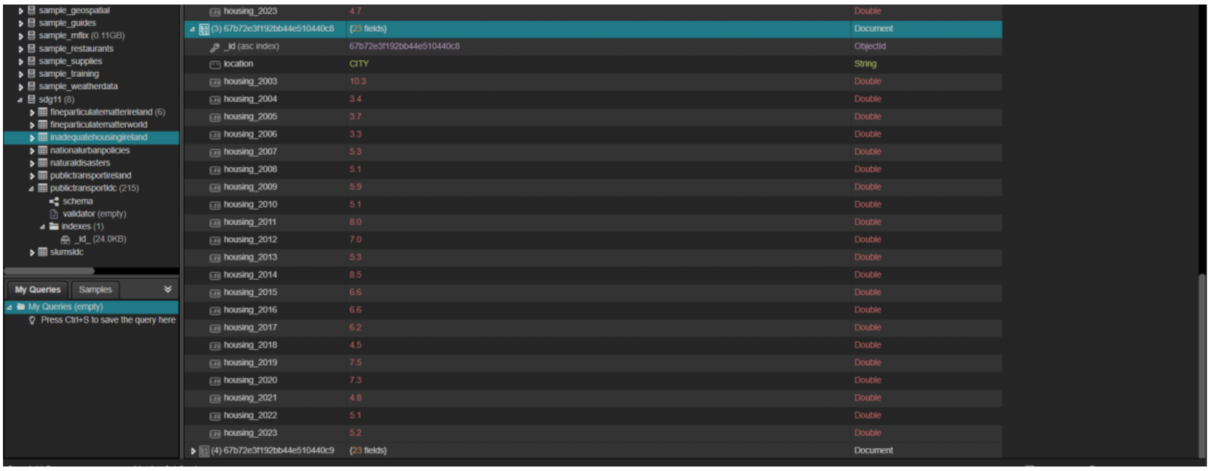
Analysing Trends in Inadequate Housing in Ireland

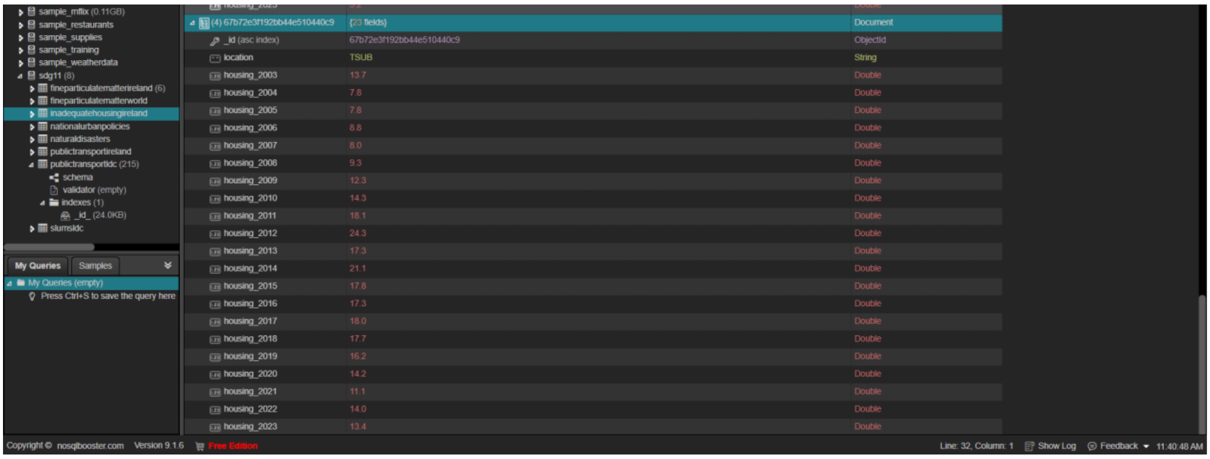
Query Reasoning: This query examines how inadequate housing in Ireland has changed over time (2003–2023). By tracking these we can explore whether housing conditions have improved or worsened. This also allows us to analyse whether external factors have influenced housing adequacy in Ireland.

Code:

```
db.inadequatehousingireland.aggregate([
  {
    $project: {
      location: "$Location",
      housing_2003: { $toDouble: "$2003" },
      housing_2004: { $toDouble: "$2004" },
      housing_2005: { $toDouble: "$2005" },
      housing_2006: { $toDouble: "$2006" },
      housing_2007: { $toDouble: "$2007" },
      housing_2008: { $toDouble: "$2008" },
      housing_2009: { $toDouble: "$2009" },
      housing_2010: { $toDouble: "$2010" },
      housing_2011: { $toDouble: "$2011" },
      housing_2012: { $toDouble: "$2012" },
      housing_2013: { $toDouble: "$2013" },
      housing_2014: { $toDouble: "$2014" },
      housing_2015: { $toDouble: "$2015" },
      housing_2016: { $toDouble: "$2016" },
      housing_2017: { $toDouble: "$2017" },
      housing_2018: { $toDouble: "$2018" },
      housing_2019: { $toDouble: "$2019" },
      housing_2020: { $toDouble: "$2020" },
      housing_2021: { $toDouble: "$2021" },
      housing_2022: { $toDouble: "$2022" },
      housing_2023: { $toDouble: "$2023" }
    }
  },
  {
    $sort: { location: 1 } // Sorting by location for readability
  }
])
```







Explanation of results: The suburban (TSUB) areas consistently had the highest percentage of inadequate housing, peaking at 24.3% in 2012 before gradually declining to 13.4% in 2023. City areas (CITY) also saw fluctuations, with a peak of 8.5% in 2014, followed by a gradual decline, though it remained above 5% in 2023. The overall national trend (ALLAREA) remained relatively low, staying below 7% for most of the period, with the lowest point at 2.2% in 2004. The significant increase in suburban inadequate housing between 2010 and 2013 may indicate the impact of the 2008 financial crisis, while the gradual decline after 2015 suggests possible housing policy interventions or economic recovery.

Query 19

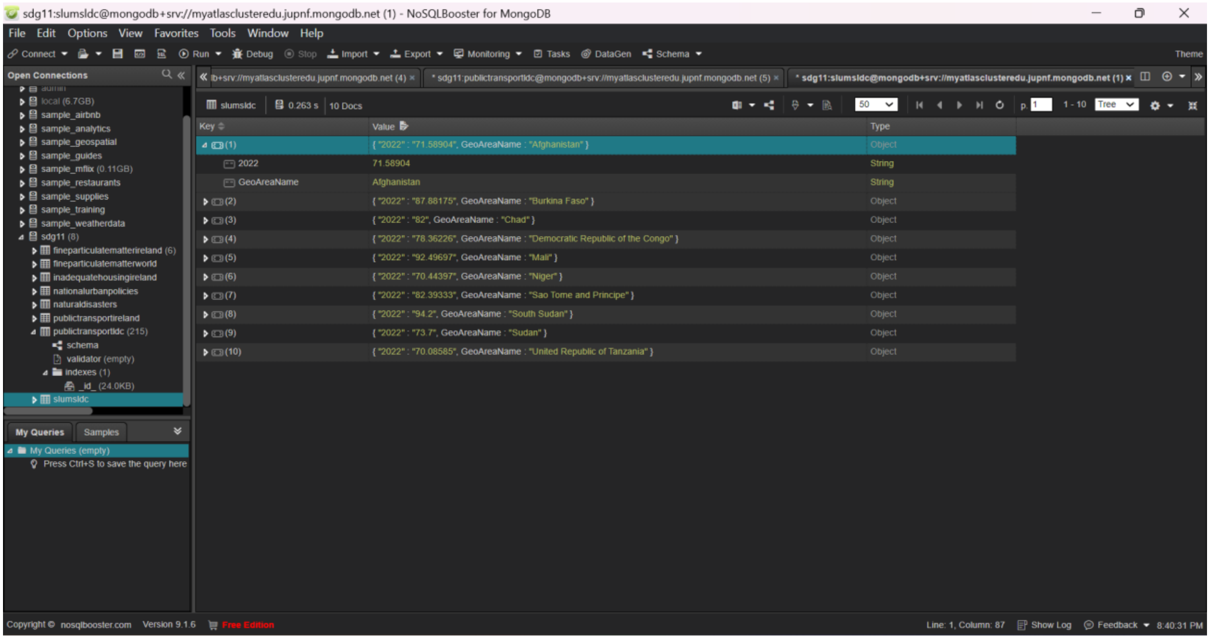
Slums, Least Developed Countries

countries where more than 70% of the urban population lived in slums in 2022.

Query Reasoning: This query retrieves records where the 2022 slum percentage exceeds 70%, identifying countries with extremely high levels of urban slum populations. Analysing this data highlights critical regions needing urgent housing and infrastructure interventions.

Code:

```
db.slumsldc.find({ "2022": { $gt: "70" } }, { "GeoAreaName": 1, "2022": 1, "_id": 0 })
```



Explanation of Results: In 2022, ten least developed countries had over 70% of their urban population living in slums, with South Sudan (94.2%) and Mali (92.5%) among the highest. These figures highlight severe housing challenges, emphasizing the urgency of SDG 11 in promoting sustainable urban development and improving living conditions.

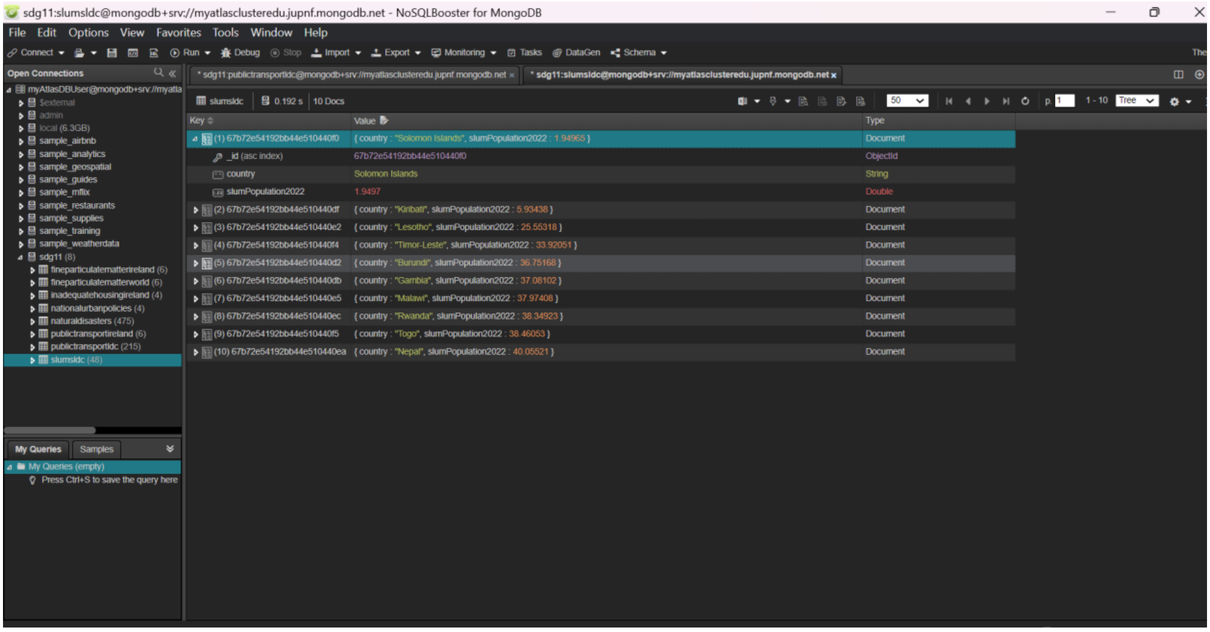
Query 20

Slums, Least Developed Countries

Find Countries with the lowest slum populations in 2022

Query Reasoning: This query identifies the ten least developed countries (LDCs) with the lowest proportion of their urban population living in slums in 2022. We wanted to analyse disparities between countries, highlighting those with better urban housing conditions while recognizing that even the most developed LDCs still lag far behind countries like Ireland. This provides insight into which nations have made progress in reducing slum populations and where further improvements are needed to meet SDG 11.1 goals.

Code:
```
db.slumsldc.aggregate([
  {
    $project: {
      country: "$GeoAreaName",
      slumPopulation2022: { $toDouble: "$2022" }
    }
  },
  {
    $match: {
      slumPopulation2022: { $ne: null, $ne: "", $gte: 0 }
    }
  },
  {
    $sort: { slumPopulation2022: 1 }
  },
  {
    $limit: 10
  }
]);
```



Explanation of Results: The results show that while Solomon Islands has the lowest proportion of its urban population living in slums at 1.95%, the remaining nine countries still have relatively high percentages, even among the least affected LDCs. Nepal (40.05%), Togo (38.46%), and Rwanda (38.34%) highlight how even the more developed LDCs struggle with significant urban slum populations. In comparison, Ireland's worst-affected urban areas typically have around 13% of the population in inadequate housing, showing a stark disparity between developed nations and LDCs. This reinforces the ongoing challenge of urban infrastructure and housing quality in LDCs, where even the most improved regions still fall far behind higher-income countries.

Query 21

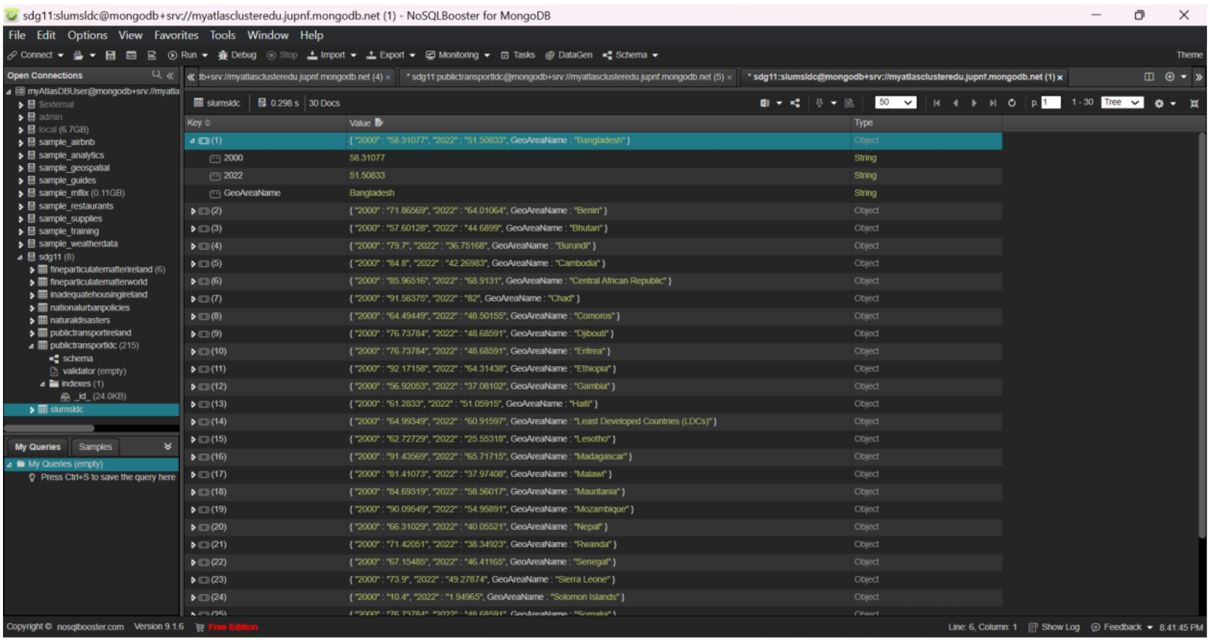
Slums, Least Developed Countries

Countries where slum population percentage decreased between 2000 and 2022.

Query Reasoning: This query uses $expr with a $gt operator to compare the 2000 and 2022 slum percentages. It returns countries where the slum rate in 2000 was higher than in 2022, indicating an improvement in urban living conditions.

Code:

```
db.slumsldc.find({
   $expr: {
      $gt: [ "$2000", "$2022" ]
   }
}, { "GeoAreaName": 1, "2000": 1, "2022": 1, "_id": 0 })
```



Explanation of Results: Several of the least developed countries have significantly reduced their urban slum populations between 2000 and 2022, with notable improvements in Lesotho (62.7% to

25.5%), Malawi (81.4% to 37.9%), and Rwanda (71.4% to 38.3%). While progress has been made, many countries still have high slum populations, emphasising the ongoing need for policies aligned with SDG 11 to improve urban living conditions.

Query 22

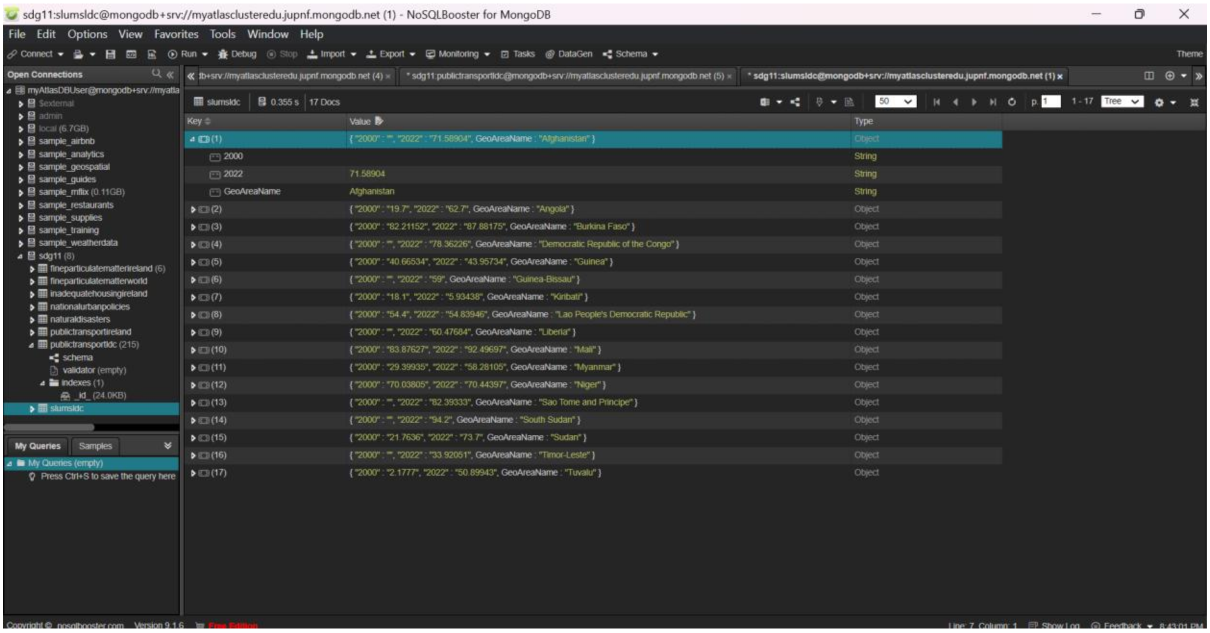
Slums, Least Developed Countries

Countries where slum percentage increased from 2000 to 2022.

Query Reasoning: This query is essentially the reverse of Query 21. Instead of identifying countries where the slum percentage decreased from 2000 to 2022, it uses $expr with a $lt operator to return countries where the 2022 value is higher than the 2000 value, highlighting regions where urban slum conditions have worsened over time.

<u>Code:</u>

```
db.slumsldc.find({
   $expr: {
      $lt: [ "$2000", "$2022" ]
   }
}, { "GeoAreaName": 1, "2000": 1, "2022": 1, "_id": 0 })
```



Explanation of Results: Several countries have seen an increase in their urban slum populations between 2000 and 2022, with Tuvalu (2.1% to 50.9%), Sudan (21.7% to 73.7%), and Angola (19.7% to 62.7%) experiencing significant rises. Mali (83.9% to 92.5%) remain among the worst affected. These trends highlight ongoing urban housing challenges.

Query 23

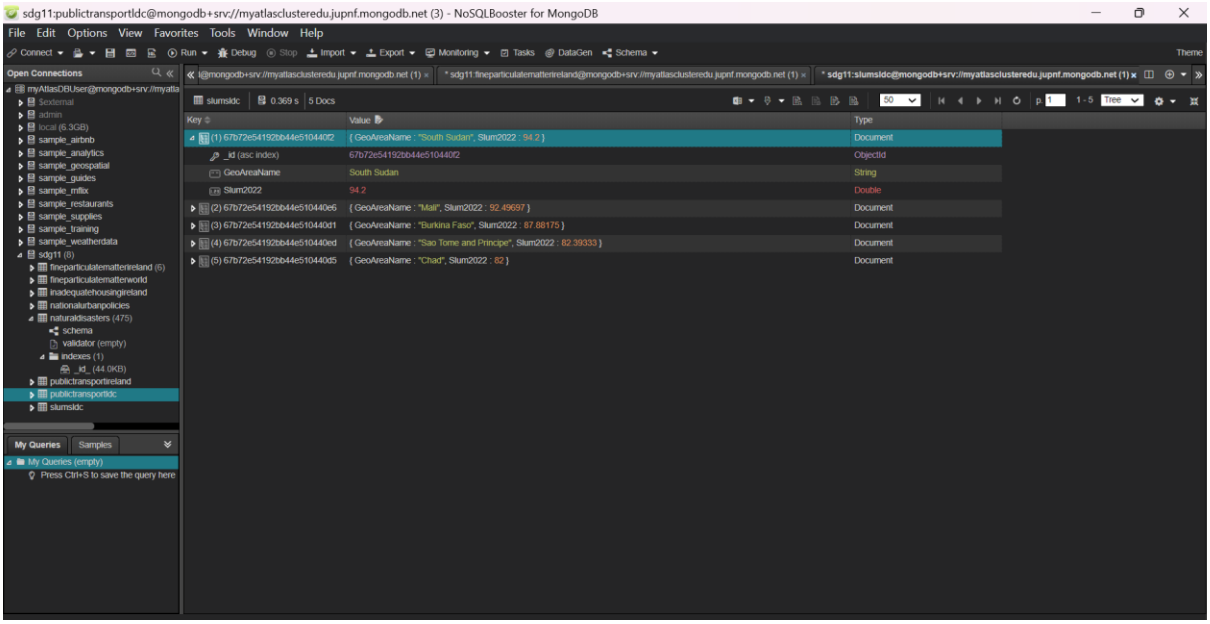
Slums, Least Developed Countries

Countries with the highest Slum populations in 2022

Query Reasoning: The purpose of this query is to identify the top 5 countries where the highest percentage of the urban population lived in slums in 2022. This helps in understanding which regions are experiencing severe urban housing challenges. Governments and international organisations like the United Nations and World Bank can then use this data to prioritize housing development programs in countries in most in need of them.

Code:

```
db.slumsldc.aggregate([
  { $project: { GeoAreaName: 1, Slum2022: { $toDouble: { $ifNull: ["$2022", "0"] } } } },
  { $sort: { Slum2022: -1 } },
  { $limit: 5 }
]);
```



Explanation of Results: This query highlights how these countries have over 70-90% of their urban populations living in slums. It signals a major challenge in achieving sustainable urbanization. These nations face rapid urban growth, weak infrastructure, and economic constraints, making SDG 11's goal of inclusive and resilient cities more urgent than ever. This query provides critical data to accelerate progress toward sustainable urban development.

Query 24

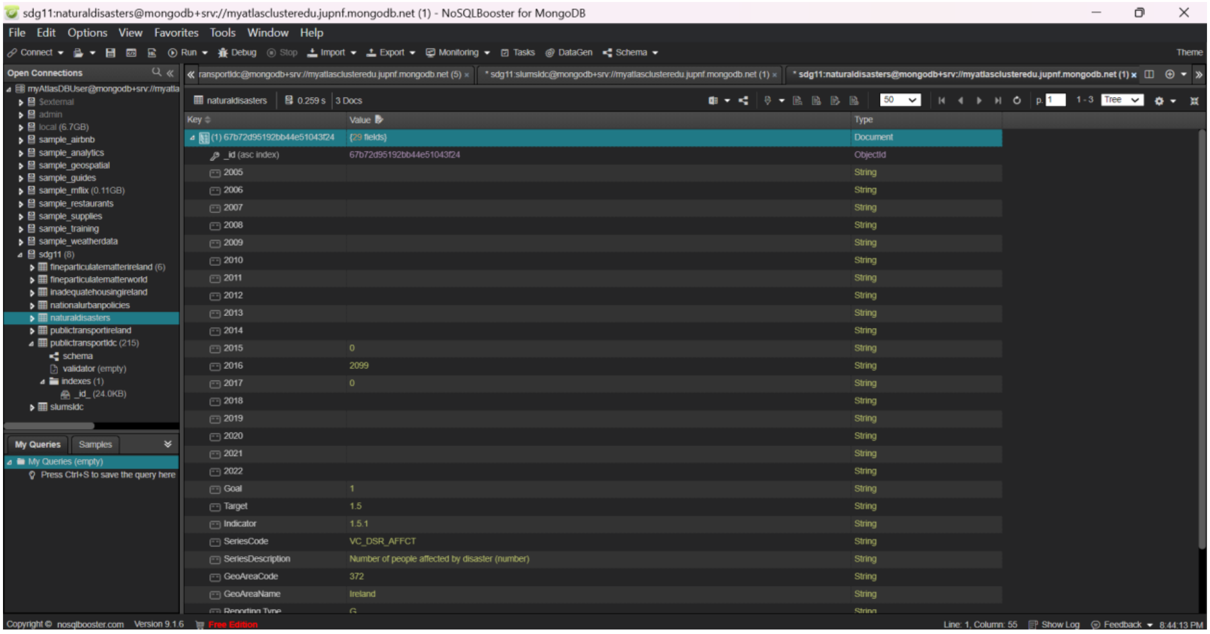
Natural disasters

Searching for Ireland.

Query Reasoning: This query retrieves records for Ireland from the natural disasters dataset using a simple find command. Since Ireland has its own dedicated datasets for indicators 11.1, 11.2, and 11.6, for indicator 11.5 we wanted to isolate Ireland's data to focus on its specific disaster impact profile.

Code:

```
db.naturaldisasters.find({ "GeoAreaName": "Ireland" })
```



Explanation of Results: We were unable to retrieve data specifically for Ireland alone in this dataset. However, the available records indicate that Ireland has been largely unaffected by natural disasters, with zero recorded impacts in most years except for 2016, when 2,099 people were affected. This suggests that Ireland experiences minimal disaster-related disruptions, aligning with its relatively stable climate and infrastructure resilience.

Query 25

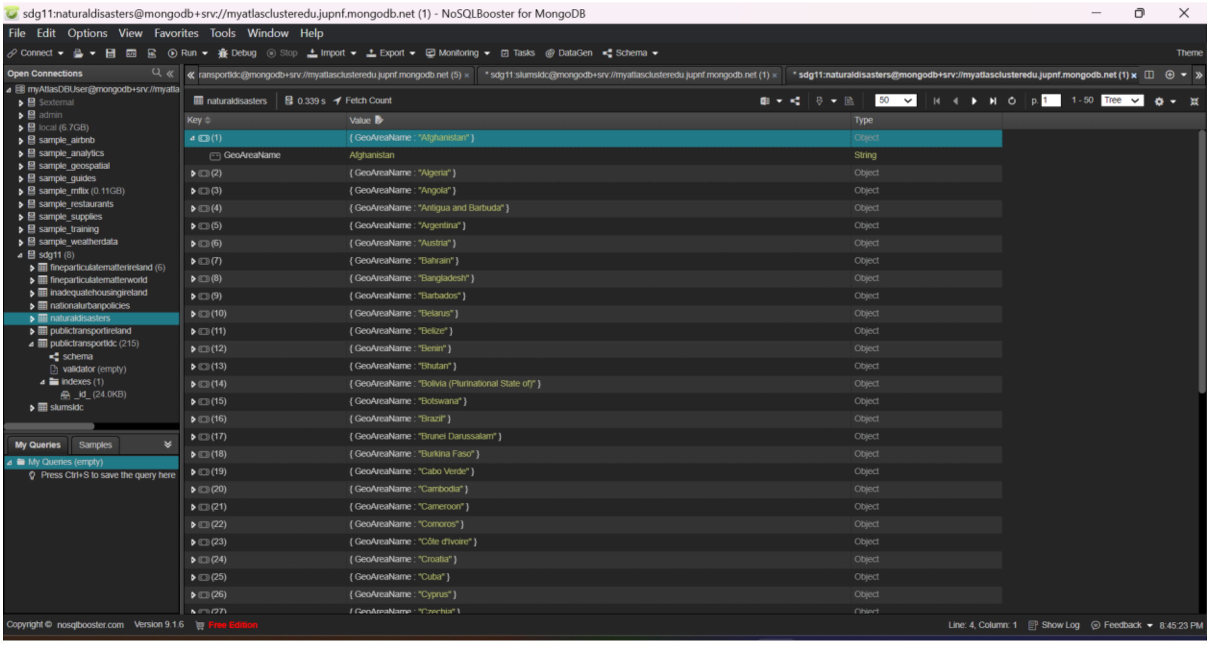
Natural disasters

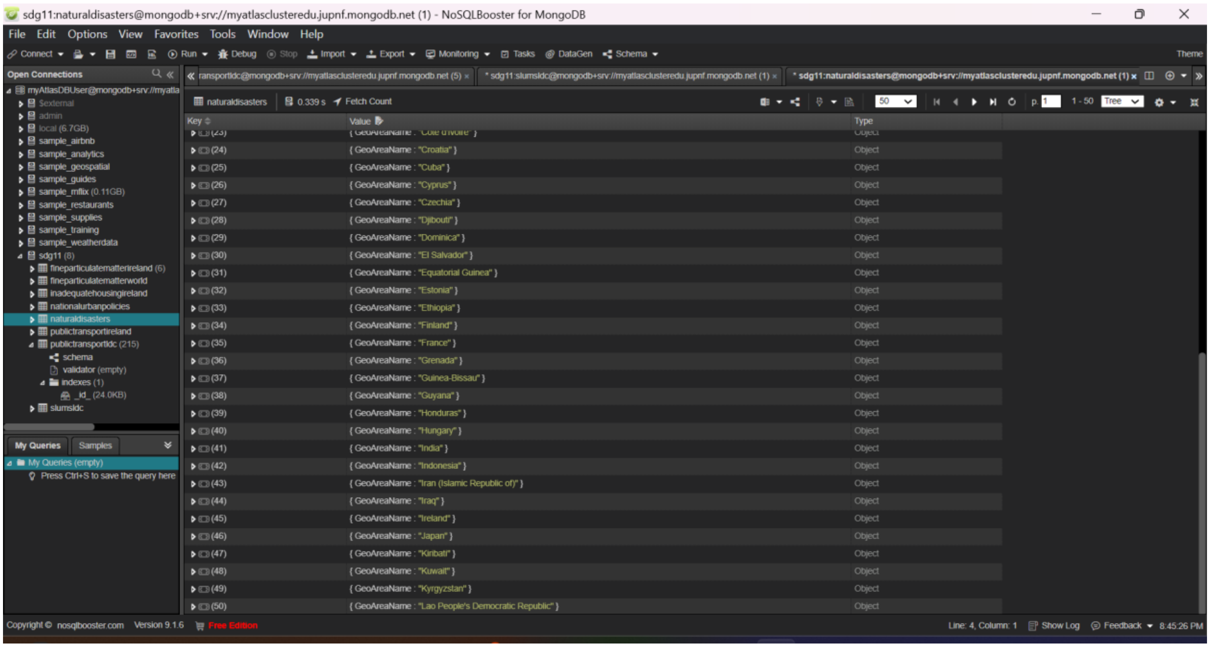
Countries where no disaster impact was recorded in 2022.

Query Reasoning: This query retrieves countries where the 2022 field is either an empty string, "0", or null, indicating no recorded disaster impact. This approach helps isolate regions that either experienced no impact or failed to record it.

<u>Code:</u>

```
db.naturaldisasters.find({
  "2022": { $in: ["", "0", null] }
}, { "GeoAreaName": 1, "_id": 0 })
```





Explanation of Results: The dataset includes a list of countries where no disaster impact was recorded in 2022. Ireland is among them as expected, with many countries remaining unaffected.

Query 26

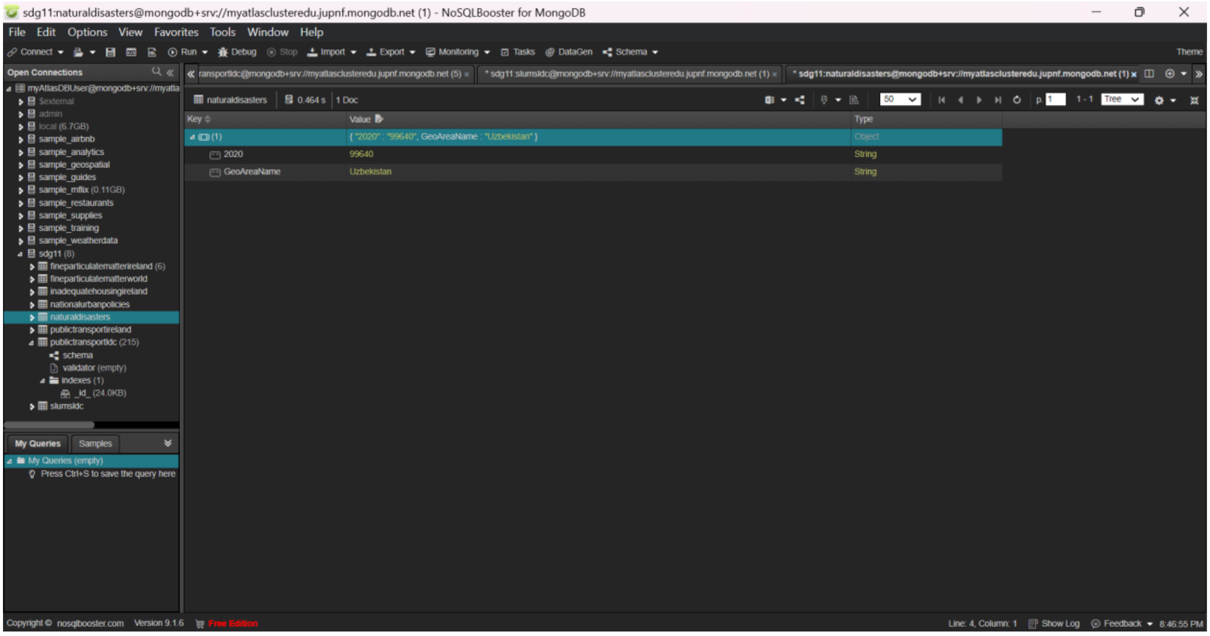
Natural Disasters

Country with the highest number of people affected by disasters in 2020.

Query Reasoning: This query retrieves all records with the 2020 disaster impact data, sorts them in descending order by the 2020 value, and limits the output to the top record—effectively isolating the country with the highest impact. It functions as the inverse of Query 25, which identifies countries with little to no reported impact.

Code:

```
db.naturaldisasters.find({}, { "GeoAreaName": 1, "2020": 1, "_id": 0 })
   .sort({ "2020": -1 })
   .limit(1)
```



Explanation of Results: 2020 was a different year because of Covid, in 2020, Uzbekistan recorded the highest number of people affected by disasters, with 99,640 impacted. This highlights significant vulnerability to natural events, emphasizing the need for disaster resilience measures in line with SDG 11.

Query 27

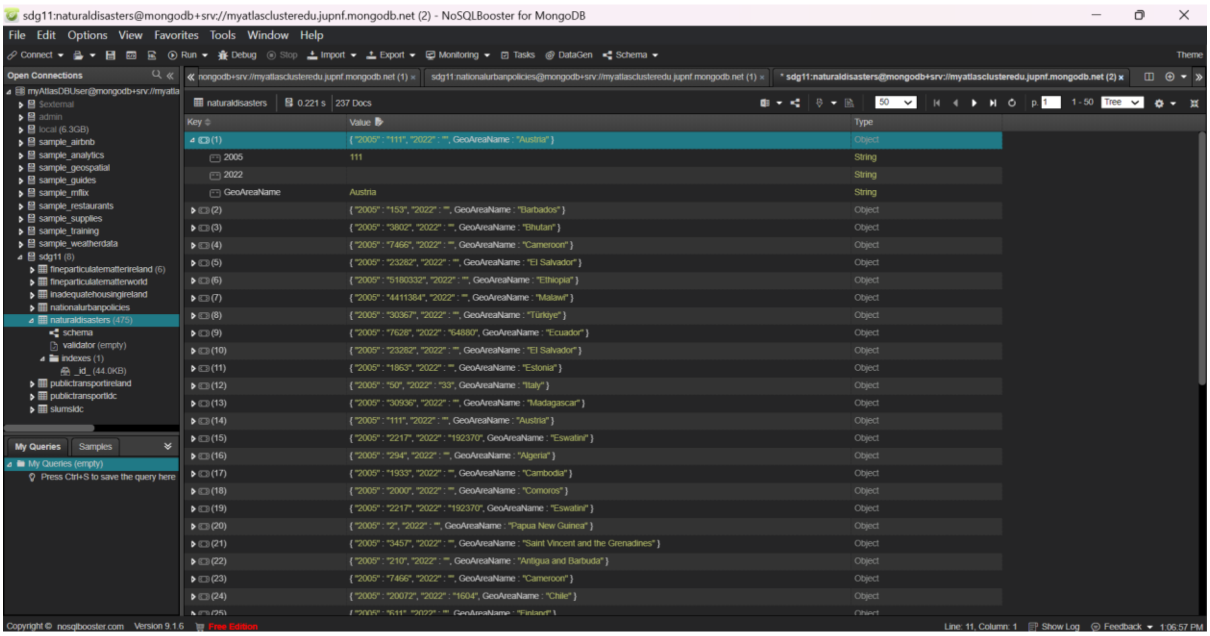
Natural disasters

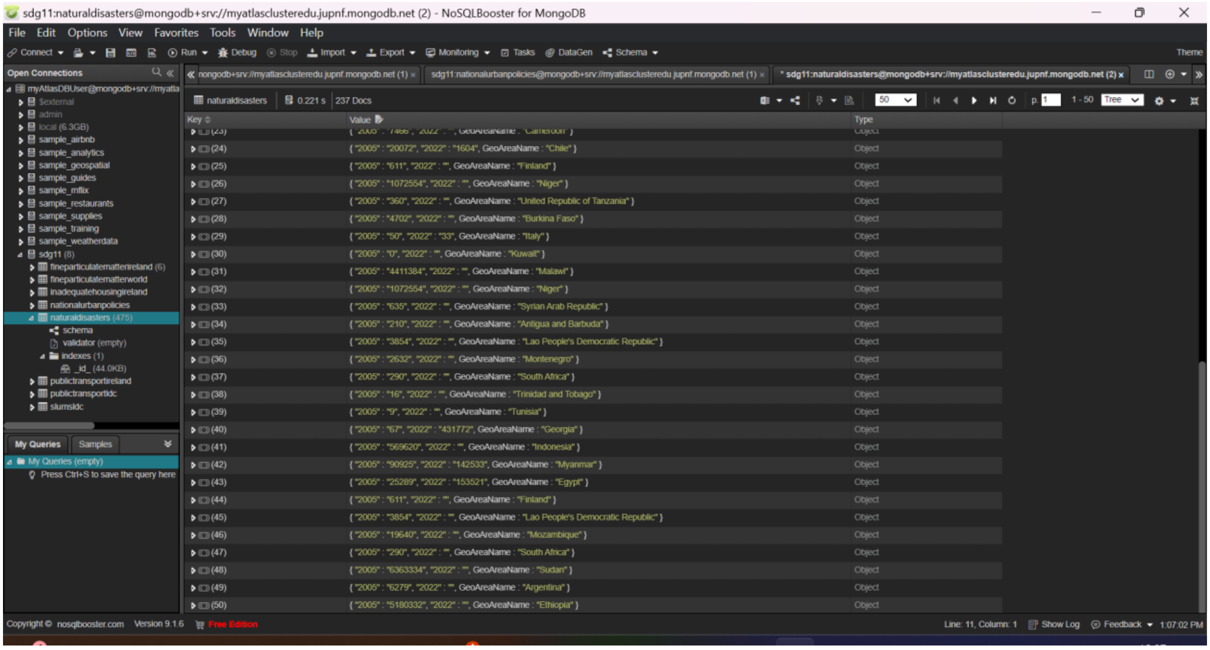
Countries Where Disaster Impact Has Decreased Over Time

Query Reasoning: Find the countries where the number of people affected by disasters has decreased from 2005 to 2022. Showing where disaster management has improved over time, highlighting progress toward SDG 11.

Code:

```
db.naturaldisasters.find({
  $expr: {
    $gt: [ "$2005", "$2022" ]
  }
}, {
  GeoAreaName: 1,
  "2005": 1,
  "2022": 1,
  _id: 0
})
```





Explanation of results: The query identifies countries where disaster impact in 2005 was higher than in 2022. Some, like Italy and Chile, saw genuine reductions, while others have missing 2022 data, making trends unclear.

Query 28

Natural disasters

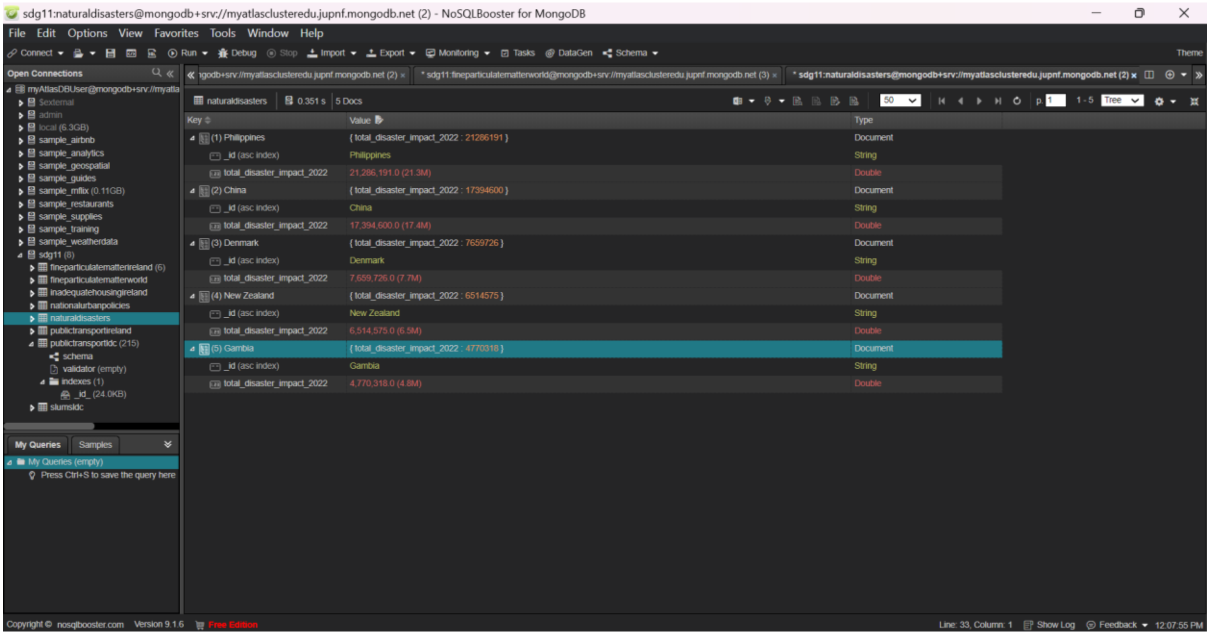
Countries with the Highest Disaster Impact in 2022

Query Reasoning: Since we already identified the top five countries with the highest slum populations in 2022 (query 23), this query will identify the top five countries most affected by disasters in 2022 from the naturaldisasters dataset. This allows us to compare whether there is an overlap between high slum populations and high disaster impact, reinforcing the link between poor housing conditions and disaster vulnerability.

<u>Code:</u>

```
db.naturaldisasters.aggregate([
  {
    $match: {
      "2022": { $ne: "", $ne: null }
    }
  },
  {
    $project: {
      country: "$GeoAreaName",
      disaster_impact_2022: {
        $convert: {
```

```
            input: { $ifNull: ["$2022", "0"] },
            to: "double",
            onError: 0,
            onNull: 0
          }
        }
      }
    },
    {
      $group: {
        _id: "$country",
        total_disaster_impact_2022: { $sum: "$disaster_impact_2022" }
      }
    },
    {
      $sort: { total_disaster_impact_2022: -1 }
    },
    {
      $limit: 5
    }
])
```



Explanation of results: The results show that the top five countries most affected by natural disasters in 2022 were the Philippines, China, Denmark, New Zealand, and Gambia. There is no clear correlation between countries with the highest slum populations and those most impacted by disasters, as none of the Least Developed Countries (LDCs) with the highest slum populations—South Sudan, Burkina Faso, Mali, São Tomé and Príncipe, and Chad—appear in this dataset. A key limitation in this analysis is the lack of recorded disaster impact data for LDCs in 2022, which suggests that these countries may struggle with data collection and reporting. This aligns with broader challenges in sustainable development, where LDCs often lack the infrastructure and resources needed for accurate disaster tracking and response planning. While this does not necessarily mean that these countries were unaffected by disasters, it highlights the data availability gap in vulnerable regions, making it harder to assess risks.

Query 29

Natural disasters

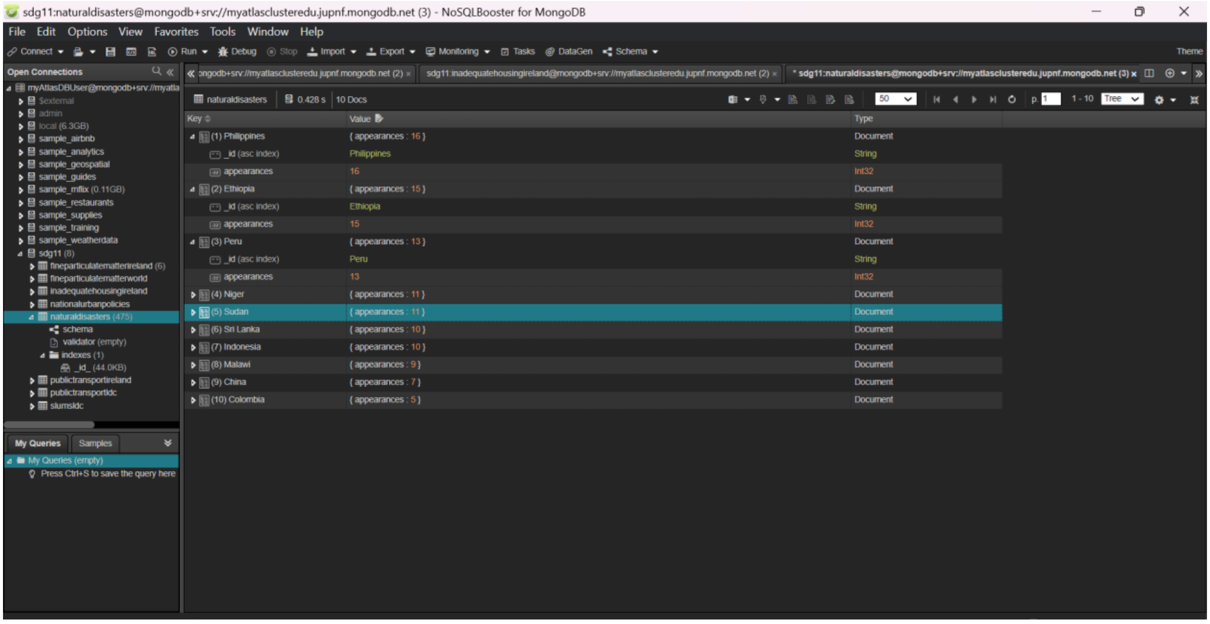
Identifies the 10 most disaster-prone countries

Query Reasoning: The purpose of this query is to determine which countries have been the most consistently affected by disasters from 2005 to 2022. Instead of just identifying the worst-affected country in a single year, this query provides a long-term perspective by counting how many times a country appeared in the top 10 most affected countries per year. This allows us to recognize patterns in disaster vulnerability and identify regions where resilience measures may be insufficient.

Code:

```
db.naturaldisasters.aggregate([
  {
    $project: {
      GeoAreaName: 1,
      years: [
        { year: "2005", value: { $convert: { input: "$2005", to: "double", onError: 0, onNull: 0 } } },
        { year: "2006", value: { $convert: { input: "$2006", to: "double", onError: 0, onNull: 0 } } },
        { year: "2007", value: { $convert: { input: "$2007", to: "double", onError: 0, onNull: 0 } } },
        { year: "2008", value: { $convert: { input: "$2008", to: "double", onError: 0, onNull: 0 } } },
        { year: "2009", value: { $convert: { input: "$2009", to: "double", onError: 0, onNull: 0 } } },
        { year: "2010", value: { $convert: { input: "$2010", to: "double", onError: 0, onNull: 0 } } },
        { year: "2011", value: { $convert: { input: "$2011", to: "double", onError: 0, onNull: 0 } } },
        { year: "2012", value: { $convert: { input: "$2012", to: "double", onError: 0, onNull: 0 } } },
        { year: "2013", value: { $convert: { input: "$2013", to: "double", onError: 0, onNull: 0 } } },
        { year: "2014", value: { $convert: { input: "$2014", to: "double", onError: 0, onNull: 0 } } },
        { year: "2015", value: { $convert: { input: "$2015", to: "double", onError: 0, onNull: 0 } } },
        { year: "2016", value: { $convert: { input: "$2016", to: "double", onError: 0, onNull: 0 } } },
        { year: "2017", value: { $convert: { input: "$2017", to: "double", onError: 0, onNull: 0 } } },
        { year: "2018", value: { $convert: { input: "$2018", to: "double", onError: 0, onNull: 0 } } },
        { year: "2019", value: { $convert: { input: "$2019", to: "double", onError: 0, onNull: 0 } } },
        { year: "2020", value: { $convert: { input: "$2020", to: "double", onError: 0, onNull: 0 } } },
        { year: "2021", value: { $convert: { input: "$2021", to: "double", onError: 0, onNull: 0 } } },
        { year: "2022", value: { $convert: { input: "$2022", to: "double", onError: 0, onNull: 0 } } }
      ]
    }
  },
  { $unwind: "$years" },
  { $match: { "years.value": { $gt: 0 } } },
  { $sort: { "years.year": 1, "years.value": -1 } },
  {
    $group: {
      _id: { year: "$years.year", country: "$GeoAreaName" },
      max_impact: { $max: "$years.value" }
    }
  },
  { $sort: { "_id.year": 1, "max_impact": -1 } },  // Ensure correct sorting
  {
    $group: {
      _id: "$_id.year",
      top_countries: { $push: "$_id.country" }
    }
  },
  { $project: { _id: 1, top_10_countries: { $slice: ["$top_countries", 10] } } },
  { $unwind: "$top_10_countries" },
```

```
  {
    $group: {
      _id: "$top_10_countries",
      appearances: { $sum: 1 }
    }
  },
  { $sort: { appearances: -1 } },
  { $limit: 10 }
])
```



Explanation of results: The results show that the Philippines (16) is the most disaster-prone country, which makes sense as it is frequently affected by typhoons, earthquakes, and floods due to its location in the Pacific Ring of Fire. Ethiopia (15) also faces severe droughts and floods, which contribute to food insecurity. These results suggest that coastal and tectonic regions face recurring natural disasters, while drought-prone countries struggle with environmental instability. The data could be linked with national urban policies to assess disaster preparedness however our urban policies data was insufficient as it is divided by region rather than individual countries.

Query 30

Natural Disasters

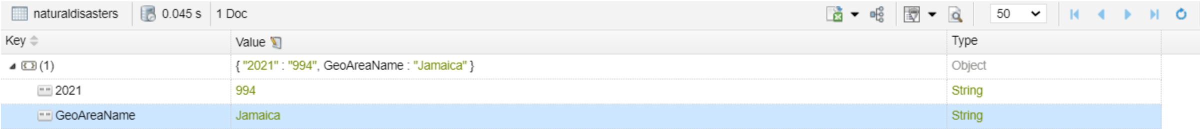
Country with the Highest Number of People Affected by Disasters in 2021

Query Reasoning: The purpose of this query is to identify the country that experienced the highest number of people affected by disasters in 2021. This query focuses on a single-year snapshot to determine the most impacted region during that specific year. This is useful for assessing extreme disaster events in 2021, understanding immediate vulnerabilities, and developing optimal emergency response efforts.

<u>Code:</u>

```
db.naturaldisasters.find({}, { "GeoAreaName": 1, "2021": 1, "_id": 0 })
```

```
.sort({ "2021": -1 })
.limit(1)
```



naturaldisasters | 0.045 s | 1 Doc | 50

Key	Value	Type
(1)	{ "2021" : "994", GeoAreaName : "Jamaica" }	Object
2021	994	String
GeoAreaName	Jamaica	String

Explanation of Results: The results show that Jamaica had the highest number of people affected by disasters in 2021. This is unsurprising given Jamaica's location in the Caribbean hurricane belt, where it is frequently impacted by tropical storms, hurricanes, and flooding. Jamaica's disaster impact is often linked to hurricanes and tropical storms, such as Hurricane Elsa in 2021, which caused severe flooding, infrastructure damage, and displacement of communities. These findings highlight the need for stronger climate adaptation strategies, including improved disaster preparedness and resilient urban infrastructure. We wanted to ask this query to see if Covid (Query 26) had any effect on natural disaster impact. The results were inconclusive.

SECTION 4: CONCLUSION:

Throughout this project, we explored Sustainable Development Goal 11 (SDG 11) using MongoDB to analyse real-world urban challenges related to housing, public transport, air quality, natural disasters, and urban policies. The experience provided valuable insights into global disparities, particularly between Least Developed Countries (LDCs) and Ireland, where the gap in infrastructure, living conditions, and disaster preparedness was stark.

For example, we found that over 70% of urban populations live in slums in some LDCs, whereas even the worst-affected areas in Ireland typically have only around 13% of people in inadequate housing. Public transport accessibility also highlighted disparities, with cities like Kikwit (Democratic Republic of the Congo) having just 1.7% access compared to over 90% in Irish cities like Cork and Dublin.

Although our findings revealed some improvements in SDG 11, they did not align with the rate at which technology and global infrastructure have advanced. Air pollution data showed a general decline in PM2.5 levels between 2010 and 2019, but urban and suburban areas still had the highest pollution levels, reinforcing the link between urbanisation, transport infrastructure, and environmental impact.

Challenges and Limitations

While AI tools like ChatGPT were instrumental in generating and refining queries, they struggled with certain aspects, such as handling null values, misinterpreting data at times, and requiring manual adjustments to avoid hallucinated results. The AI also had difficulty fully understanding our dataset structure, especially when values were formatted inconsistently.

A key limitation in our analysis was data availability.

- Public transport and slum data were useful, but often from different time frames, making direct comparisons difficult.
- National urban policies data was lacking in scope, making it harder to link findings to other SDG 11 indicators.
- Natural disaster data was missing for many LDCs, and air quality data was divided by regions instead of specific countries, limiting granular insights.

Final Thoughts

Despite these challenges, this project was an engaging and insightful experience, allowing us to apply NoSQL database techniques to real-world global issues. We gained technical skills in MongoDB, learned how to handle complex datasets, and developed a deeper understanding of sustainable urban development. While the world has made progress toward SDG 11, our findings suggest that more rapid improvements is necessary, particularly in transport access, housing conditions, and environmental policies for LDCs.

Overall, this project reinforced the importance of data-driven decision-making for global development and how databases can be leveraged to track, analyze, and influence policy changes to create safer, more sustainable cities for all.

SECTION 5: REFERENCES

ChatGPT (2025) *AI Language Model*, OpenAI. Available at: https://openai.com/chatgpt

Code Beautify *Online JSON Beautifier and Converter*. Available at: https://codebeautify.org (Accessed: 05/02/2025).

Lang, M. *NoSQL and Special Purpose Databases*. Business Information Systems Group, J.E. Cairnes School of Business & Economics, National University of Ireland, Galway.

MongoDB (2025) *MongoDB Documentation*. Available at: https://www.mongodb.com/docs (Accessed: 04/03/2025).

NoSQL Booster (n.d.) *NoSQL Booster for MongoDB*. Available at: https://nosqlbooster.com (Accessed: 04/03/2025).

United Nations (2025) *Sustainable Development Goals (SDG) Global Database*. Available at: https://unstats.un.org/sdgs/dataportal/database (Accessed: 01/02/2025). SDG 11

SECTION 6: APPENDIX

APPENDIX A: Layout & Formatting:

As a group we made sure that the document is as well formatted as possible, the screenshots are as big as possible, and we ensured to articulate on our query reasonings and results. We also used the same colour for headings to give the document a professional and uniform look. We used appropriate headings where relevant, which allows for seamless transition through the document as the viewer pleases. Lastly, when transferring code from NoSQLBooster, we first transferred it to Notepad++ to allow us to copy and paste it with the MongoDB Syntax intact within the Word file, as direct copying from NoSQLBooster does not format the code correctly and it appears messy and disorganised. Using tools like Outlook & OneDrive, we could easily access and edit one anothers work as we pleased thanks to the online sharing facilities these tools provide.

APPENDIX B: Collaboration:

Throughout this project, we collaborated consistently and effectively. Sharing constructive feedback with one another was essential as we learned off each other what work was best and which steps were the right ones to take after. Meeting frequently ensured that none of us forgot the tasks at hand and provide us with constant opportunities to discuss and improve on our existing work.