A Study of the Heating Temperature in the Major Cities in the World through Data Mining using MongoDB

Jose Miguel Agustin R. Yabut

Department of Information Systems and Computer Science

Ateneo de Manila University

Quezon City, Philippines

+639165267640

[jmaryabut@gmail.com](mailto:jmaryabut@gmail.com)

Daniel S. Jabonete

Department of Information Systems and Computer Science

Ateneo de Manila University

Quezon City, Philippines

+639202115067

[djabonete@gmail.com](mailto:djabonete@gmail.com)

ABSTRACT

Climate change in our Earth has been an endless issue to everyone. “Global warming”, which is one of the direct effect of climate change, points out the scorching temperature that is affecting our major resources such as water, food and shelter. Another catalyst to this societal progeny is the greenhouse effect. Where warming of the Earth’s surface receives additional heat from the greenhouse gases trapped inside the atmosphere that is on top of the uncontrollable intensity coming from the Sun. The facts gathered from the dataset used for the research which was obtained from Kaggle, identified the average temperature of the major cities in the world and how it soared to what it is in year 2016. That would help prevent further upsurge or damage to what we are now experiencing on a global scale. The file was uploaded by the account, Berkeley Earth, with an open permission for the community to freely analyze the datasets published on the website. The file which was last updated only 7 months ago, includes a historical information of the average temperatures of most of the major cities in the world. The results obtained from the study can be used by concerned government and non-government agencies around the world. They may include it in their campaigns, contingency plans and possible courses of action in the future, to improve our way of living.

KEYWORDS

Climate Change, Global Warming, Greenhouse Effect

MongoDB, Database, Averaging Temperatures

1 INTRODUCTION

There is constant evolution in our lands and oceans and as cliché as it may sound, there is the inexhaustible climate change. Whether we live in a blissful tropical country or in an endless snowing plateau, the comfort of our locales has continuous subtle changes in temperature. This change can either be positive or more often than not, a negative effect.

In Japan, housewives would still visit nearby shops multiple times a week to buy food in small amounts. This is due to several factors they consider such as the size, shape and quality of the product, which are not pre-packed and does not have constant pricing and quantity. Those aforementioned facts, aside from other conditions, mainly depend on the weather that may mean getting less than expected from the previous days caused by climate change [1]. Global warming heats up not only the temperature but also the intensity, occurrence and frequency of heat waves. This in turn increases incidents of non-communicable diseases such as heat stress that eventually leads to fatality. Which is the case during recent summers in Europe where thousands died because of the said phenomena [2].

Whenever climate change and any other weather-related situation or global warming is discussed, it is accompanied by immeasurable data. For us to cope and be able to analyze all this information, we use several data mining techniques accompanied by a reliable database and algorithms to design a model that can narrow down things, events and factors to consider getting the data that we need to further improve ongoing actions to keep our planet a livable land to humankind. MongoDB allows us to do this [3] and in addition gives us capability to do other data mining practices such as getting frequent pattern mining, sequence recognition, clustering and regression analysis, outlier detection and so many more. To make sure the enormous amount of data is used and relevant information is extracted [8].

2 SIGNIFICANCE OF THE STUDY

Climate change is a very broad topic to be tackled at any given time. A very significant piece of that subject matter is the reality of the limitless increase and decrease in temperature in all parts of the world. The study focuses on that point to help concerned organizations around the world work on a more effective approach for the betterment of the society.

3 RESEARCH OBJECTIVES

The research intends to identify how critical the climate change is in the major cities of the world, specifically the rising temperature in our lands and some of its effects that has already taken place in our nations. Data analysis, aggregation and modeling is done by using MongoDB as primary tool and repository. The specific question the study wants to answer is what is the average increase and/or decrease of temperature in the world. The averages will have two categories: 1) Year, the research will analyze data from 2006 to 2016 and 2) Country, which will have sub-categories analyzing which cities have the most significant increase and decrease in temperature by the year 2016.

4 SCOPE AND LIMITATIONS

The dataset used for the study is taken from the open-resource-data science-concentrated website of Kaggle. The analysis covers temperature volatility in relation to global warming caused by climate change. Taking into consideration the past ten years from the file, where 2016 marks the tenth year for the study. On account of the latest available recorded facts as per NASA on their latest annual average Global Temperature data posted on their [website](https://climate.nasa.gov/vital-signs/global-temperature/). The study includes MongoDB as the sole database technology used for the research. Pre-processing for the dataset is incorporated. MapReduce modeling which included sharding, replication, data aggregation and server backup are all done using the MongoDB database.

5 RELATED STUDIES

5.1 Climate Change

Historically, humans did not start the climate change in our planet. Evidences recorded in the past showed origin from acts of God which includes natural calamities and disasters. Such are the volcanoes that erupted in the past and other tectonic activities prompting unexpected movement in the Earth. Although people think that we ourselves are the reason behind it. That may not be entirely true but surely with the growth of technology and development of new products that require burning of forests, fossil fuels and other resources it increased the greenhouse gases trapped within the Earth’s atmosphere [4].

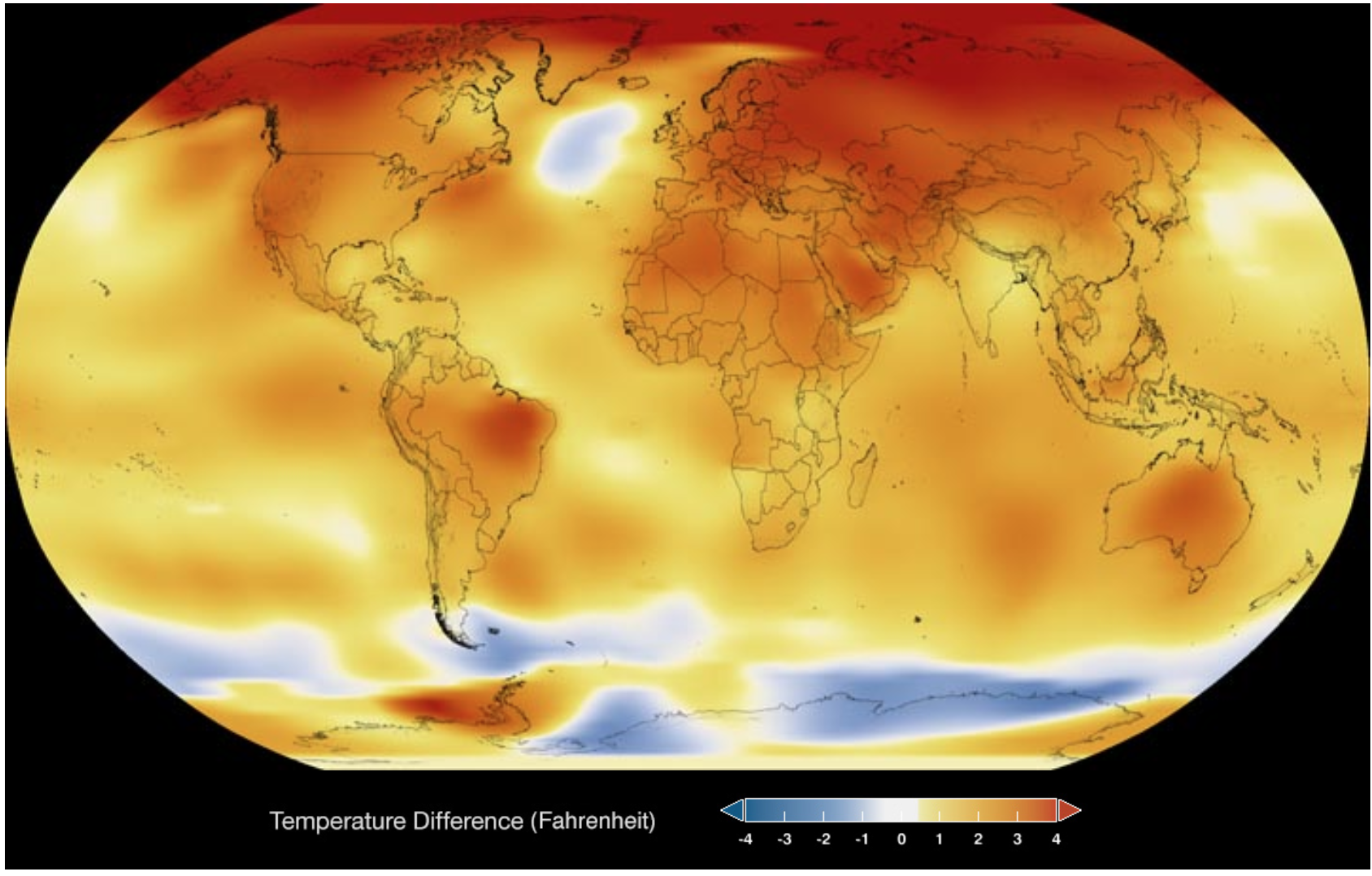


**Fig 1: This color-coded map shows the global surface temperatures in 1986. Dark blue indicates areas cooler than average. Dark red indicates areas warmer than average.**

Data source: NASA/GISS   
Credit: NASA Scientific Visualization Studio

According to scientists, these gases increased due to the increase of carbon dioxide levels and temperature. They also said that we humans have been fostering this greenhouse effect for over 250 years now and its only swelling more every year [5]. Evidences of global warming can be seen numerically and visually as is on Figure 2 for the year 2016, if you compare it to the temperature around the globe 30 years before that which can be seen on Figure 1 for the year 1986 [4].

Even with the Arctic sea ice, which holds the ice pack that cover the Arctic Ocean that prevents sea-level rising, reached the level where scientists from NOAA, National Oceanic and Atmospheric Administration, said there was a 17.7% average drop which is the lowest recorded since they began documenting in 1979 [[4-](https://earthobservatory.nasa.gov/Features/GlobalWarming/global_warming_2007.pdf)[5]](https://www.cbsnews.com/news/climate-change-global-warming-transformed-the-earth-in-2016/).



**Fig 2: This color-coded map shows the global surface temperatures in 2016. Dark blue indicates areas cooler than average. Dark red indicates areas warmer than average.**

Data source: NASA/GISS   
Credit: NASA Scientific Visualization Studio

5.2 Averaging Temperature

There are a lot of variables to consider that make our air, land and sea hotter and colder than its usual state. Although what we always see, is different temperature values. This information may be coming from our mobile devices, the internet or more often than not, the nightly news in our televisions.

During the news hour is also when we hear of hospitalizations and fatalities that happen due to stroke. In a study conducted last year, June of 2016, they were able to determine that there has been a significant increase in stroke hospitalization rate that was caused by an increase in temperature values in all regions and seasons.

The average temperature was identified to be a major factor due to the area of the Northeast where they have colder land, has remarkably less cases of the heat stroke. In an investigation back in 2009, hospitalizations were also studied. In this research, they observed the effect of heat waves for a span of 10 years where they started during the year of 1999. Although they did not have very strong evidences at the time, the cases of stroke are still noticeable associated with higher temperature caused by heat waves [11].

In Table 1 it can be seen that there has been a significant 31% increase of cardiovascular hospitalizations and visits in California during the period of May to October 2009 where they reported an increase heat wave effect [10].

Table 1. Statewide effect estimates of temperature and heat wave on hospitalizations by diagnosis.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Diagnosis** | **ICD-9 Codes** | **No. of Visits (% of total****[a](http://www.sciencedirect.com/science/article/pii/S0013935117310009" \l "tbl2fna))** | **Rel. Risk for Temp** **Inc**[b](http://www.sciencedirect.com/science/article/pii/S0013935117310009#tbl2fnb)**(95% CI)** | **Rel. Risk Assoc. w/ Heat Wave (95% CI)** | **I2Statistic for Temperature** | **I2Statistic for Heat Wave** |
| All cardiovascular diseases | 390–459 | 2,128,178 (31%) | **0.99 (0.98–0.99)** | 0.99 (0.97–1.01) | 11 | 0 |

5.3 MongoDB database

There is numerous database technology to choose from but considering massive volume of datasets for research, the scalability and the ability to use containers to deduce a schema aside from being an open-source system, MongoDB is really one of the best option. A feature to consider from this NoSQL type of database is the capability of server replication and sharding of enormous datasets. Believed to be one of the fastest-growing database in the field of information technology, MongoDB also features a rapid handling of load when it comes to big data. The typical operations of our databases such as inserting, updating and deleting rows and tables has a very consistent high-speed processing of data when it comes to large amount of records [6,7].

A study showed an Oracle database and MongoDB database performance comparison. Syntaxes aside, as seen on Table 2, you will notice the consistency, reliability and fast-paced operation of MongoDB when it comes to processing big data [7]. MongoDB as a widely spread document-oriented database management system provides allowance of time for its users to do more compared to an Oracle database or any other SQL database when extracting or processing big data and database server replication. On replication mode, we conduct primary-secondary and primary-backup structure where each secondary replica or backup has exact copies of the main database [9]. Another feature of MongoDB that is excellent for aggregating results for reports and records is the map-reduce function [7]. It gives you flexibility where you can easily merge fields and table features to get summed up results that you intend to extract from your dataset [7,9].

Table 2. Operation times comparison(msec).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **No. of Records** | **Oracle Database** | | | **MongoDB** | | |
|  | Insert | Update | Delete | Insert | Update | Delete |
| 10 | 31 | 453 | 94 | 800 | 1 | 1 |
| 100 | 47 | 47 | 47 | 4 | 1 | 1 |
| 1000 | 1563 | 47 | 62 | 40 | 1 | 1 |
| 10000 | 8750 | 94 | 94 | 681 | 1 | 1 |
| 100000 | 83287 | 1343 | 1234 | 4350 | 2 | 1 |
| 1000000 | 882078 | 27782 | 38079 | 57871 | 3 | 1 |

6 METHODOLOGY

6.1 Data Collection

This study obtained the dataset from the website of Kaggle. Under their datasets section, the researchers looked for climate change related files that would fit the criteria needed. There were different column metadata grouped together, as loads of CSV files were searched, the account of Berkeley Earth was seen where the most suitable dataset for the research was found. It included the top 3 priority data that the researchers were looking for, which includes the dates, average temperature on the given dates, and cities.

6.2 Pre-processing

This paper has used combined three ways for dataset pre-processing. First, a manual cleaning of the downloaded CSV file with a native label of GlobalLandTemperaturesByMajorCity, from Kaggle was done.

The fields from the dataset that was not going to be used for the research was taken out. As seen on Table 3, the column metadata for the CSV file includes dt, which was changed to Date, averagetemperature which was switched to AvgTemp, and averagetemperatureuncertainty was omitted. The CSV file volume was initially 14.1MB which included 239,177 rows, it was then reduced to 11.9MB and 9,108 rows.

Table 3. Column Metadata of GlobalLandTemperaturesByMajorCity.csv

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **dt** | **averagetemperature** | **averagetemperatureuncertainty** | **city** | **country** | **latitude** | **longitude** |
| 1849-01-01 | 1849-01-01 | 1849-01-01 | 1849-01-01 | 1849-01-01 | 1849-01-01 | 1849-01-01 |
| 26.704 | 26.704 | 26.704 | 26.704 | 26.704 | 26.704 | 26.704 |
| 1.435 | 1.435 | 1.435 | 1.435 | 1.435 | 1.435 | 1.435 |
| Abidjan | Abidjan | Abidjan | Abidjan | Abidjan | Abidjan | Abidjan |
| Côte D'Ivoire | Côte D'Ivoire | Côte D'Ivoire | Côte D'Ivoire | Côte D'Ivoire | Côte D'Ivoire | Côte D'Ivoire |
| 5.63N | 5.63N | 5.63N | 5.63N | 5.63N | 5.63N | 5.63N |

Next is the automated pre-processing with data analysis. After the manual pre-process of the dataset, it is then loaded to a java designed program that utilizes available API interfaces connecting to MongoDB, it then formats the rows according to MongoDB document style. The third pre-processing step includes supplementing information that adds the Region and Sub-Region field to categorize each country as to which continent it belongs to and then converts it to a JSON file.

The process took thirty seconds which reduced the file volume from the manual pre-process output down to 782KB and compressed the rows into JSON content format which minimized it to 2,638 documents. A sample converted CSV row to json format, can be seen on Figure 1. This is to ensure and highlight the non-relational database structure of MongoDB.



**Fig 3: Sample document from the dataset after converting it to json format**

6.3 MongoDB database setup and modeling

The machine used for the research is a MAC computer. Similar to every repository setup, installation was the first phase. The database program was obtained from the download center of the MongoDB website. At the time of the research, it was on version 3.4. The database program was then extracted to a specific location in the computer, where we created a folder for the whole program which completes the installation process of the MongoDB database. Afterwards, we began the setup of MongoDB as the database for the study.

Initially, the researchers went through the MongoDB website guides to check the ideal setup for the database. Since version 3.2 of MongoDB was recently upgraded, the old support for the old setup was discontinued. A new arrangement had to be done. It was considered setting up, what they call, config servers for sharded clusters – deployed as a replica set or CSRS. Following their documentation, the setup for the Config Server started. An initial terminal instance was opened and the bin folder of MongoDB program was located and began the configuration of the Primary server instance using the default MongoDB port which is 27017. 2 back-ups for the primary was setup subsequently, using ports 27018 and 27019. That finalizes the first replica set, which is the config server that will act as the server for the research database setup.

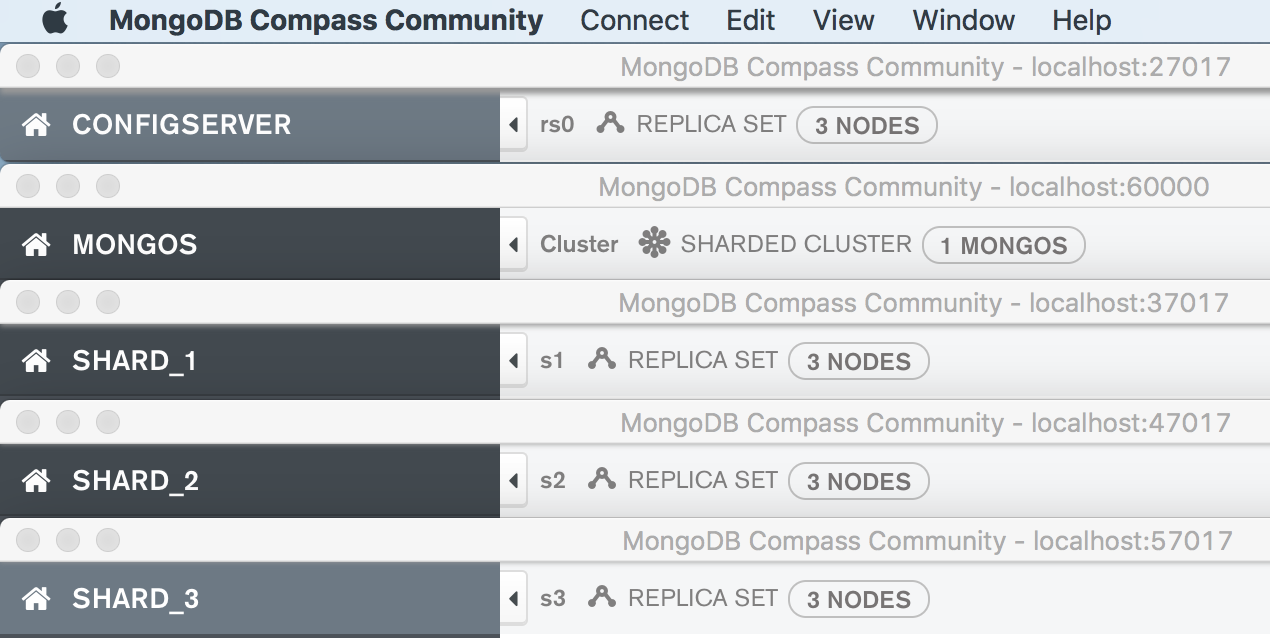
Thereafter, the next setup was for three more replica sets that was intended to act as shard servers for the database. These shard servers are similar to containers, they hold the dataset once loaded into MongoDB. The shard servers followed the logic of having a two-back-up structure for the config server. For uniformity, each shard server held its own ten-thousands’ place in ascending order. First shard server used 37017, followed by its back-ups having 37018 and 37019. Second shard server used 47017, followed by its back-ups having 47018 and 47019. The last shard server that was setup used 57017, followed by 57018 and 57019. After all the replica set has been configured, each primary server was implemented with higher priority than the back-ups. This has been set for instances where the primary unexpectedly crashes or go down.

The next phase before loading the dataset into MongoDB, is the mongos or the Sharded Cluster. This is the routing service for the MongoDB database technology for the shards to process the queries of users of the repository in the application layer. Once the config and shard servers were put into place, the mongos easily followed. The researchers just had to open another terminal window to select an unused port. For this research port 60000 was used. The config server hostname was added, which by default is localhost:27017, along with its back-ups. Successively, another tab from the terminal was opened, the port specified for the mongos was used to enter the Mongo Shell.

Right after mongos started running, the setup to link the shards were done. Each shard address was associated to the mongos having the same format of the default address and specifying each port, which is localhost:37017, localhost:47017 and localhost:57017. This is to ensure that when a dataset is loaded into MongoDB, all the replica set that is acting as shard servers will be able to have chunks of data distributed equally among them by the sharded cluster balancer.

Lastly, the researchers prepared the MongoDB database settings for the dataset import. This included a specific database and collection name using the mongos. A collection is where the dataset can be found once loaded in MongoDB. Sharding settings was also applied to the collection name and specific index key to target categorization for the Country attribute was generated. The tool mongoimport was used to import the dataset, now JSON file, to the Sharded Cluster. Chunks were manually split due to the size of the dataset being very small. Since each chunk of MongoDB can contain up to 64MB, the distribution of the data upon import remained on just one shard.

After all the aforementioned steps, the MongoDB database is already in the working phase. To make sure everything is working as expected, the researchers used another tool created also by the MongoDB team, this is the MongoDB Compass. It allowed graphical user interface (GUI) for the whole database, as seen on Figure 2. It lets users pull multiple window to view each server present on their database setup. On this study as stated before, one config server, three shard servers and one mongos instance was setup. The primary server ports are also displayed on the Compass GUI for address confirmation. Another beneficial feature of the MongoDB Compass is the next step that was taken which confirmed the successful importing of the dataset in the newly setup database. It is now visible in all the server shards and replicas. An additional test for the visibility of the documents loaded, queries were initiated from the mongos console window and checked responses from each shard which displayed the information expected as output.



**Fig 4: Screenshot of MongoDB Compass GUI for the newly setup database**

Furthermore, the researchers tested to insert a new document in each shard and was able to confirm when all of the test record appeared on each replica.

Once everything was in place, the aggregation steps needed to get the results for the study started. The MongoDB database stood also as the model to extract the information needed to fulfill the research objectives. The use of the MapReduce functions was the highlight of the querying of data. The MapReduce functions of MongoDB allowed the researchers to narrow down the dataset to get the average temperatures required to give information to answer the research questions in place, that led to the success of the study.

6.4 Results

The dataset contained 239,177 data in total count and 9,108 remained after the manual and automated pre-processing, and 2,640 was generated upon converting the data to JSON format. The researchers was able to identify the relevance of each data based on date, country, city, average temperature, longitude, and latitude. Then removed those data that are irrelevant to answer the research questions about getting the highest and lowest increase and decrease of temperature among the Countries and Cities included in the dataset.

The file used in this research covers 159 out of the 195 countries around the world. The vast number of countries used in this study exhibits better overview and understanding of the climate change around the world. As well as with the creation of an application, the researchers were able to verify the temperature changes for the past 10 years. These numbers show that the dataset can provide significant results and high accuracy on data analysis.

In this study the researchers grouped the data using MongoDB MapReduce Functions to retrieve the lowest and highest average temperature. It can be seen on Table 4, countries with the lowest average temperature, lowest recorded temperature, and its difference and percentage change in the Asian region. It is determined that Turkey and Iran had the significant temperature changes. In which, the values in the same table showed a temperature decrease from its average lowest temperature recorded for the past 10 years period.

Table 4. Temperature Comparison of Countries in Asia

**for the lowest recorded temperature**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Country** | **Temperature (ºC)** | | | **Percentage Change (%)** |
| **Average Lowest** | **Lowest Recorded** | **Difference** |
| Myanmar | 23.462 | 22.606 | 0.855 | -3.646 |
| Afghanistan | 1.753 | -0.193 | 1.946 | -111.009 |
| Singapore | 25.377 | 25.189 | 0.188 | -0.742 |
| Philippines | 24.648 | 24.165 | 0.483 | -1.958 |
| Japan | 0.965 | -0.153 | 1.118 | -115.850 |
| Thailand | 23.835 | 22.492 | 1.343 | -5.635 |
| Syria | 5.033 | 1.137 | 3.896 | -77.407 |
| Saudi Arabia | 13.241 | 10.736 | 2.505 | -18.918 |
| Bangladesh | 18.288 | 17.424 | 0.864 | -4.723 |
| Vietnam | 25.087 | 24.508 | 0.579 | -2.308 |
| South Korea | -3.684 | -5.274 | 1.590 | 43.144 |
| Turkey | -1.571 | -5.185 | 3.614 | 229.961 |
| Pakistan | 12.227 | 10.793 | 1.434 | -11.731 |
| Iran | -0.344 | -4.182 | 3.838 | 1,116.759 |
| China | -5.263 | -7.386 | 2.123 | 40.336 |
| Taiwan | 16.450 | 15.476 | 0.974 | -5.919 |
| Iraq | 8.977 | 4.236 | 4.741 | -52.811 |
| Indonesia | 25.508 | 24.871 | 0.637 | -2.496 |

In Table 5 shows the countries and the highest average temperature, highest recorded temperature, and its difference and percentage change in Asian region. It is determined that Vietnam and China had the significant temperature changes. In which, the values in the same table shows a temperature increase from its average highest temperature recorded for the past 10 years period. With the results from table 4 and 5, the researchers generated a multi-series range graph for the results visualization show in Figure. 5.

Table 5. Temperature Comparison of Countries in Asia

**for the highest recorded temperature**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Country** | **Temperature (ºC)** | | | **Percentage Change (%)** |
| **Average Lowest** | **Highest Recorded** | **Difference** |
| Myanmar | 29.715 | 30.505 | -0.790 | 2.659 |
| Afghanistan | 25.488 | 26.570 | -1.082 | 4.244 |
| Singapore | 27.219 | 27.766 | -0.547 | 2.009 |
| Philippines | 28.191 | 29.172 | -0.981 | 3.478 |
| Japan | 24.284 | 25.203 | -0.919 | 3.786 |
| Thailand | 29.507 | 30.279 | -0.772 | 2.615 |
| Syria | 28.510 | 29.734 | -1.224 | 4.294 |
| Saudi Arabia | 35.055 | 35.682 | -0.627 | 1.788 |
| Bangladesh | 29.018 | 30.236 | -1.218 | 4.197 |
| Vietnam | 28.555 | 30.216 | -1.661 | 5.817 |
| South Korea | 23.963 | 24.894 | -0.932 | 3.887 |
| Turkey | 25.085 | 26.497 | -1.412 | 5.630 |
| Pakistan | 33.755 | 34.605 | -0.850 | 2.518 |
| Iran | 24.541 | 25.795 | -1.254 | 5.109 |
| China | 27.899 | 29.883 | -1.984 | 7.111 |
| Taiwan | 27.838 | 28.189 | -0.351 | 1.262 |
| Iraq | 34.468 | 35.764 | -1.296 | 3.759 |
| Indonesia | 27.237 | 27.655 | -0.418 | 1.535 |

In Table 6 it showed the countries with the lowest average temperature, lowest recorded temperature, and its difference and percentage change in Europe region. It is determined that France and Germany had the significant temperature changes. In which, the values in table 6 shows a temperature decrease from its average lowest temperature recorded for the past 10 years period.

Table 6. Temperature Comparison of Countries in Europe

**for the lowest recorded temperature**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Country** | **Temperature (ºC)** | | | **Percentage Change (%)** |
| **Average Lowest** | **Lowest Recorded** | **Difference** |
| Ukraine | -6.218 | -11.586 | 5.368 | 86.336 |
| United Kingdom | 2.907 | 1.737 | 1.170 | -40.250 |
| Italy | 2.605 | 0.614 | 1.991 | -76.425 |
| France | 2.367 | -0.176 | 2.543 | -107.437 |
| Germany | -0.748 | -3.103 | 2.355 | 314.673 |
| Spain | 2.695 | 0.879 | 1.816 | -67.385 |
| Russia | -10.685 | -14.991 | 4.306 | 40.303 |

In Table 7 it showed the countries with the highest average temperature, highest recorded temperature, and its difference and percentage change in Europe region. It is determined that United Kingdom and France had the significant temperature changes. In which, the values in table 7 shows a temperature increase from its average highest temperature recorded for the past 10 years period.

Table 7. Temperature Comparison of Countries in Europe

**for the highest recorded temperature**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Country** | **Temperature (ºC)** | | | **Percentage Change (%)** |
| **Average Lowest** | **Highest Recorded** | **Difference** |
| Ukraine | 19.465 | 20.871 | -1.406 | 7.224 |
| United Kingdom | 16.388 | 19.211 | -2.823 | 17.226 |
| Italy | 21.057 | 22.553 | -1.496 | 7.105 |
| France | 18.088 | 21.661 | -3.573 | 19.751 |
| Germany | 18.241 | 20.649 | -2.409 | 13.204 |
| Spain | 21.206 | 23.144 | -1.938 | 9.137 |
| Russia | 17.846 | 19.976 | -2.130 | 11.935 |

With the results from Table 6 and 7, the researchers were able to generate a multi-series range graph for the results visualization show in Figure. 6. In Table 8 it showed the countries with the lowest average temperature, lowest recorded temperature, and its difference and percentage change in the Americas region. It is determined that United States and Chile had the significant temperature changes. In which, the values in table 8 shows a temperature decrease from its average lowest temperature recorded for the 10 years period.

Table 8. Temperature Comparison of Countries in the Americas

**for the lowest recorded temperature**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Country** | **Temperature (ºC)** | | | **Percentage Change (%)** |
| **Average Lowest** | **Lowest Recorded** | **Difference** |
| Colombia | 18.742 | 18.202 | 0.540 | -2.881 |
| Canada | -12.496 | -17.117 | 4.621 | 36.983 |
| United States | -4.113 | -8.590 | 4.477 | 108.845 |
| Brazil | 17.790 | 16.559 | 1.231 | -6.920 |
| Dominican Republic | 23.815 | 22.983 | 0.832 | -3.494 |
| Mexico | 12.007 | 10.729 | 1.278 | -10.647 |
| Chile | -1.079 | -3.816 | 2.737 | 253.530 |
| Peru | 13.273 | 12.099 | 1.174 | -8.846 |

In Table 9 showed the countries with the highest average temperature, highest recorded temperature, and its difference and percentage change in Europe region. It is determined that Canada and Chile had the significant temperature changes. In which, the values in table 9 shows a temperature increase from its average highest temperature recorded for the past 10 years period.

Table 9. Temperature Comparison of Countries in the Americas

**for the highest recorded temperature**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Country** | **Temperature (ºC)** | | | **Percentage Change (%)** |
| **Average Lowest** | **Highest Recorded** | **Difference** |
| Colombia | 22.158 | 23.096 | -0.938 | 4.232 |
| Canada | 19.776 | 22.524 | -2.748 | 13.898 |
| United States | 23.532 | 24.936 | -1.404 | 5.967 |
| Brazil | 26.089 | 27.193 | -1.104 | 4.232 |
| Dominican Republic | 27.328 | 27.869 | -0.541 | 1.979 |
| Mexico | 18.161 | 19.509 | -1.348 | 7.422 |
| Chile | 12.101 | 13.058 | -0.958 | 7.913 |
| Peru | 20.374 | 21.43 | -1.056 | 5.183 |

With the results from table 8 and 9, the researchers generated a multi-series range graph for the results visualization show in Figure 7. In Table 10 showed the countries with lowest average temperature, lowest recorded temperature, and its difference and percentage change in Africa region.

It is determined that Morocco and South Africa had the significant temperature changes. In which, the values in table 10 shows a temperature decrease from its average lowest temperature recorded for the 10 years period.

Table 10. Temperature Comparison of Countries in Africa

**for the lowest recorded temperature**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Country** | **Temperature (ºC)** | | | **Percentage Change (%)** |
| **Average Lowest** | **Lowest Recorded** | **Difference** |
| Angola | 19.936 | 19.039 | 0.897 | -4.498 |
| Egypt | 12.543 | 11.917 | 0.626 | -4.994 |
| Tanzania | 23.267 | 22.398 | 0.869 | -3.736 |
| Sudan | 22.816 | 20.555 | 2.261 | -9.911 |
| Somalia | 25.393 | 24.887 | 0.506 | -1.992 |
| Ivory Coast | 24.160 | 23.217 | 0.943 | -3.901 |
| Kenya | 14.189 | 13.304 | 0.885 | -6.239 |
| Congo | 21.496 | 20.822 | 0.674 | -3.135 |
| India | 18.308 | 17.385 | 0.923 | -5.040 |
| Senegal | 20.482 | 19.898 | 0.584 | -2.853 |
| South Africa | 11.462 | 9.997 | 1.465 | -12.782 |
| Zimbabwe | 14.870 | 13.381 | 1.489 | -10.016 |
| Nigeria | 20.618 | 19.514 | 1.104 | -5.354 |
| Ethiopia | 15.833 | 14.528 | 1.305 | -8.239 |

In Table 11 showed the countries with the highest average temperature, highest recorded temperature, and its difference and percentage change in Europe region. It is determined that South Africa and Zimbabwe had the significant temperature changes. In which, the values in table 11 shows a temperature increase from its average highest temperature recorded for the past 10 years period.

Table 11. Temperature Comparison of Countries in Africa

**for the highest recorded temperature**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Country** | **Temperature (ºC)** | | | **Percentage Change (%)** |
| **Average Lowest** | **Highest Recorded** | **Difference** |
| Angola | 25.650 | 26.327 | -0.677 | 2.639 |
| Egypt | 27.785 | 29.029 | -1.244 | 4.478 |
| Tanzania | 27.510 | 27.875 | -0.365 | 1.327 |
| Sudan | 33.735 | 34.435 | -0.700 | 2.076 |
| Somalia | 28.478 | 29.650 | -1.172 | 4.117 |
| Ivory Coast | 27.782 | 29.471 | -1.689 | 6.079 |
| Kenya | 17.195 | 17.97 | -0.775 | 4.510 |
| Congo | 24.809 | 25.622 | -0.813 | 3.278 |
| India | 32.999 | 33.878 | -0.880 | 2.665 |
| Senegal | 25.276 | 26.392 | -1.116 | 4.417 |
| South Africa | 19.863 | 21.193 | -1.331 | 6.699 |
| Zimbabwe | 23.443 | 25.549 | -2.106 | 8.983 |
| Nigeria | 30.468 | 32.047 | -1.579 | 5.181 |
| Ethiopia | 19.104 | 20.028 | -0.924 | 4.837 |

With the results from table 10 and 11, the researchers generated a multi-series range graph for the results visualization show in Figure 8.

7 VISUALIZATIONS

The following figures are visualizations of the results seen from the previous chapter of the paper which was generated by the MapReduce functions used on the MongoDB database of the research. Each figure is represented by tables as referred to from the results discussion.

# 

Fig 5: Significant Temperature Change in Asia

It can be seen on each Figure that only the major countries with cities of highest and lowest average temperature recorded has been visualized.

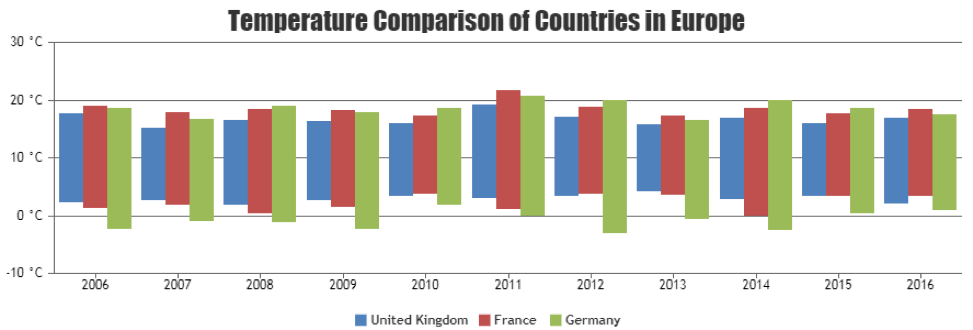


Fig 6: Significant Temperature Change in Europe

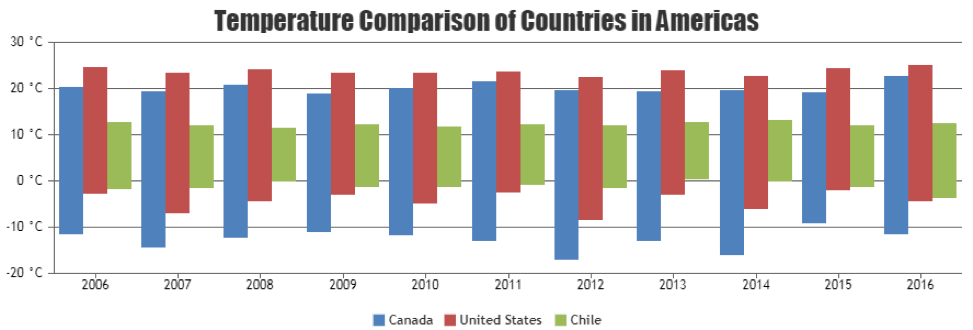


Fig 7: Significant Temperature Change in the Americas

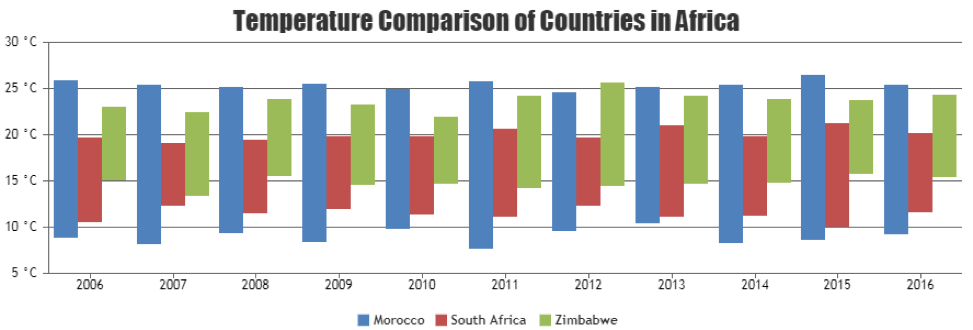


Fig 8: Significant Temperature Change in Africa

8 CONCLUSIONS AND FUTURE STUDY

This study demonstrated that pre-processing of datasets can reduce the extracted data such as rows and features that can help filter information needed for the research. The data analysis stage turns out to be valuable in retrieving and determining if there is information lacking as well as being able to supplement the values before moving forward with the study. Since these are conditional criteria that has been pre-determined, we can conclude that most of the results from the MongoDB MapReduce function flattened by the 3-layer pre-processing with the help of Java for Data Analysis is useful enough to determine significant temperature changes and understand climate change in each country. The results provided by this research study can now help determine better possible courses of action on other researchers, government and non-government agencies and their objectives in identifying critical climate changes in major cities of the world.

In this research, data analysis has been broken-down into three (3) stages that implements pragmatic approach and pre-processing for data extraction, modeling, enhancement, and determination of data validity and appropriateness for reduction strategy. Data analysis had been carried out using Climate Change: Earth Surface Temperature Data from 2006 to 2016, past 10 years.

There are possible enhancements that the researcher would like to implement in the future for further improvement of the capability when it comes with data analysis processes and come up linear regression methods. Furthermore, the implementation of different approaches using Java APIs and MongoDB could be beneficial in this kind of studies. We can expect more reliable and useful datasets, as well as the possible results for determining climate change.

8 REFERENCES

1. Usui, K. (2014). Marketing and Consumption in Modern Japan, 130.
2. Redshaw, C., Stahl-Timmins, W., Fleming, L., Davidson, I. & Depledge, M., (2013). Potential Changes in Disease Patterns and Pharmaceutical Use in Response to Climate Change, Journal of Toxicology and Environmental Health, Part B, 16:5, 285-320.

DOI: 10.1080/10937404.2013.802265

1. Plugge, E., Membrey, P., & Hawkins, T., (2010). Working with Data. In: The Definitive Guide to MongoDB. Apress, 47-81.

DOI: 10.1007/978-1-4302-3052-6\_4

1. Riebeek, H., & Simmon, R., (2007). Global Warming.

https://earthobservatory.nasa.gov/Features/GlobalWarming/global\_warming\_2007.pdf

1. Gunaratna, S., (2016, December). How climate change transformed the Earth in 2016.

https://www.cbsnews.com/news/climate-change-global-warming-transformed-the-earth-in-2016/

1. Abbes, H., & Gargouri, F., (2016, September). Big Data Integration: A MongoDB Database and Modular Ontologies based Approach.

DOI: 10.1016/j.procs.2016.08.099

1. Boicea, A., Radulescu, F., & Agapin, L. I., (2012, September). MongoDB vs Oracle--database comparison. In Emerging Intelligent Data and Web Technologies (EIDWT), 2012 Third International Conference on (pp. 330-335). IEEE.
2. Peral, J., Mate, A., & Marco, M., (2017, February). Application of Data Mining techniques to identify relevant Key Performance Indicators.

DOI: 10.1016/j.csi.2016.11.006

1. Farias, V., Sousa, F., Maia, J., Gomes, J., & Machado, J., (2017, August). Regression based performance modeling and provisioning for NoSQL cloud databases.

DOI: 10.1016/j.future.2017.08.061

1. Sherbakov, T., Malig, B., Guirguis, K., Gershunov, A., & Basu, R., (2017, September). Ambient temperature and added heat wave effects on hospitalizations in California from 1999 to 2009.

DOI: 10.1016/j.envres.2017.08.052

1. Lichtman, J., Leifheit-Limson, E., Jones, S., Wang, Y., & Goldstein, L., (2016, March). Average Temperature, Diurnal Temperature Variation, and Stroke Hospitalizations.

DOI: 10.1016/j.strokecerebrovasdis.2016.02.037