Southampton Solent University

School of Computing

**Module Title :** DATA SCIENCE

**Unit Code :** COM603

**Assessment Value :** 100%

**Level :** level 6

**Student Name :** P.Manojkumar

**Student Number :** 15166902

**Course title :** BSc (Hons) in Software Engineering

**Seminar Tutor :** G.ingayaran

Table of Contents

[**Task 1** 2](#_Toc43059046)

[Global warming 2](#_Toc43059047)

[**Task 2** 6](#_Toc43059048)

[Advantages of NoSQL database 6](#_Toc43059049)

[**Task 3** 9](#_Toc43059050)

[**Task 4** 10](#_Toc43059051)

[Access data from database 10](#_Toc43059052)

[Delete data from the database 11](#_Toc43059053)

[Update data 12](#_Toc43059054)

[**Task 5** 13](#_Toc43059055)

[**Task 6** 16](#_Toc43059056)

[**Task 7** 18](#_Toc43059057)

[**Task 8** 21](#_Toc43059058)

[**Task 9** 23](#_Toc43059059)

[Apriori algorithm 23](#_Toc43059060)

[References 25](#_Toc43059061)

# **Task 1**

## Raise in global average temperature

A different and non-natural temperature changes in the atmosphere the uncommonly quick improvement in Earth's over the previous century principally because of the greenhouse gases speeded by human beings by burning carbon contained fuels.

Atmosphere impacts where we live, our development, and our prosperity. Every types of plant and creature has adjusted to live inside a particular climatic change. A worldwide temperature on the earth's surface increment of about 0.7 0C in the twentieth century

The main part of this report is to make a more knowledge of the scientific premise of environmental change and its impacts on forest industry and the planet.

**Reason why scientists think global warming is not natural**

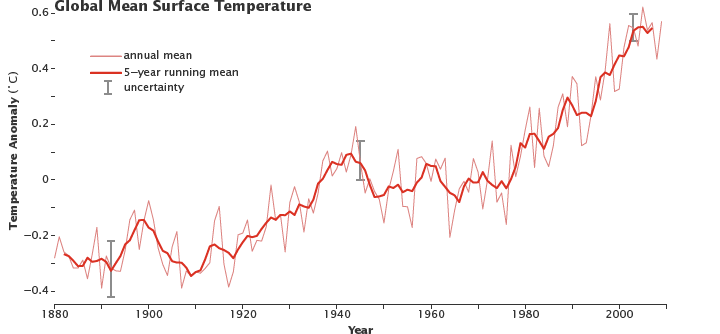
In this planet even before the industrial and technological improvement and revolution, planets average temperature gradually increased due to the characteristic reason's inconsequential to people’s action. This kind of characteristic reasons happens still now, yet these people’s impact is excessively rare compared, so explaining the global warming is too hard.

Figure mean surface temperature

This is the chart that shows how the planet mean temperature gradually increasing from 1880 to 20th contrary.

**The increased greenhouse gases and global effect**

The main reason the researchers and the scientists think the presently peoples are the one who burning fuels releasing carbon gases in the atmosphere the carbon gases in the air at a very fast rate, for mostly by using petroleum fuels, yet additionally also from chopping down carbon reducing trees from the technological revolution started mankind started to destroy the forests form the 17th century humans released CO2 gases to the atmosphere by burning down trees and burning coals form industries due to that reason the carbon gases increased from 38 percentage to 149 percentage by the start of 20th century.

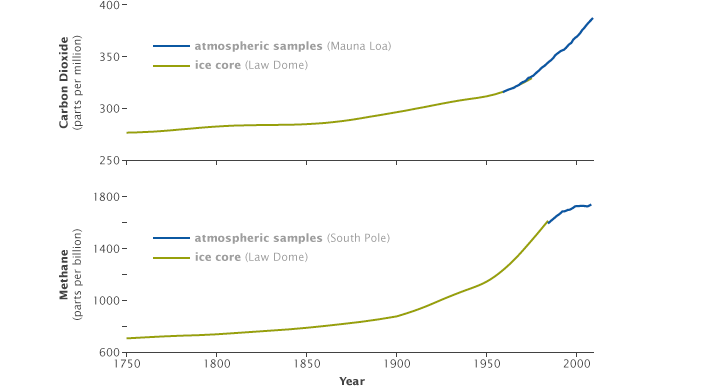
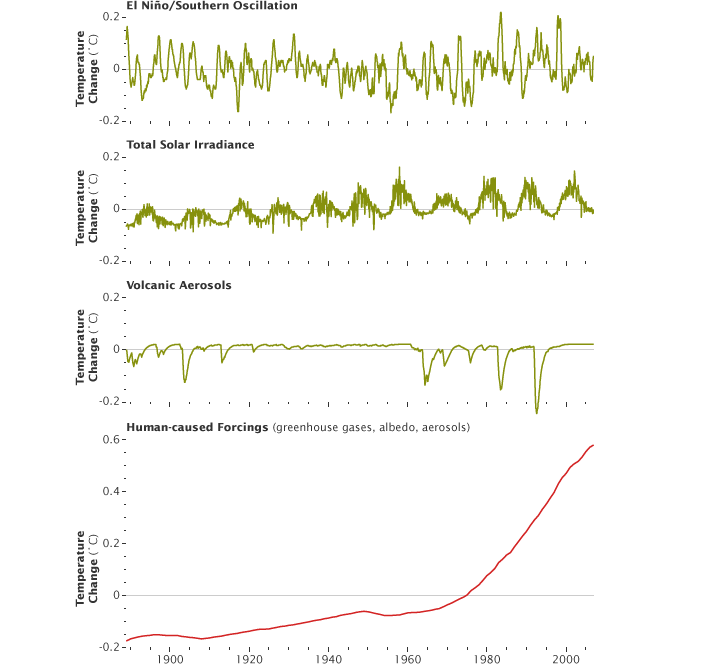


Figure CO2 and NH4 range

* Top chart shows increased Co2 gases in atmosphere and bottom chart shows increased (methane) CH4 along with the Industrial Revolution from 1750.
* Green line shows measurement of artic ice cores increases along with direct atmospheric measurements show the increase of both gases over time.



Proof of global warming is not natural

* The warming phase of the sea temperature is known as [El Niño](https://en.wikipedia.org/wiki/El_Ni%C3%B1o) happened by underwater volcanos
* As well as volcanoes are dynamic event happens all the globe, and keep on producing (Co2) usually as in a constant range, the measure of CO2 the volconoes discharge are incredibly little contrasted with human emanations.

**Solutions by data science to help in global warming**

From breaking down big datasets we realize that our planet lost what could be compared to 40 football fields for each minute a year ago by manmade faults. So we can reduce this types of faults in future by minimizing them.

There are so many problems that big data have helped, for instance by identifying the locations where most of the dangerous gases released to the atmosphere and finding out the cause of the problem or reason.

Data created by IOT systems can assist us with conserving energy and water, yet this information likewise helps researchers with developing information on environmental change that better convey to the open the genuine situation.

Information science additionally helps to understand of how various businesses contributes to environmental changes. And find out where cutting down forests are happening..

# **Task 2**

## Advantages of NoSQL database

**Rarely used of ETL(extract, transform, and load)**

Databases like NoSQL databases enables you to store coma separated files or simple structures, while record NoSQL gives capacity to deal with a scope of level or nested structures.

* XML format
* JSON format

Having able to deal with these formats locally in a Databases like NoSQL reduces the numbers of lines in a code that need to write to change the needed format from the format. This is known as extract, transform, and load (ETL).

**Support for unstructured data**

Most by far of data in big business system are not structured. Numerous Databases like NoSQL can deal with organizing of nonstructural content and change them to a structured data.

Having the option to manage unstructured data highly builds information and can assist associations with taking better choices.

**Capacity to deal with change after some time**

The normal definition of Databases like NoSQL, these databases entirely equipped for taking tasks like managing frequent changes. This is the natural of NoSQL database.

Some NoSQL databases provides all inclusive index to the qualities, structure, and information in raw data. For example Microsoft Document Database gives this feature

By chance document structure changes, these associations can utilize the data quickly, as opposed to weight for a while to rewrite the data.

**Hardware commodity**

Databases like NoSQL capable of manage dividing of a database over a few data servers. Along these lines, if your information stockpiling necessities develop excessively, you can keep on including modest servers and implement your data to those servers and making them fill in as a solitary information administration.

Differentiation this to the social database world where you have to purchase new, progressively ground-breaking and along these lines increasingly costly equipment to scale up. If you somehow managed to increase the amount of information you want to store to database, you would effortlessly fourfold the expense of the equipment you need.

Giving solidness and high accessibility of a database. By utilizing reasonable equipment and capacity of a NoSQL's significant resources. Having the option to do as such while giving liberal versatility to numerous utilizations likewise doesn't do any harm!

**Capable of handling many data structure**

Lots of applications wants basic data storing needs, on the other hand various applications needs complex and different data structures to be stored. So these kind of database comes in action they offer help to store these kind of data structures

Related data gathered by section by section and in to bigable clones.

Also can view hierarchal structure of the data stored in the database.

A trap of interrelated data can be portrayed deftly and related in triple and chart stores.

**Seller decision**

The NoSQL business is flooded with databases, however most of the famous companies are using this databases for more than many years. For instance, java Oracle, Microsoft windows, and IBM as of late came into this market. Therefore, numerous sellers are focusing on specific crowds and use own blend of development.

Open source variations can be accessible for most of these databases, that can empowers organizations to investigate as well as begin utilizing these databases at negligible hazard. This kind of organizations would then be able to take their new techniques to a creation stage by utilizing undertaking contributions.

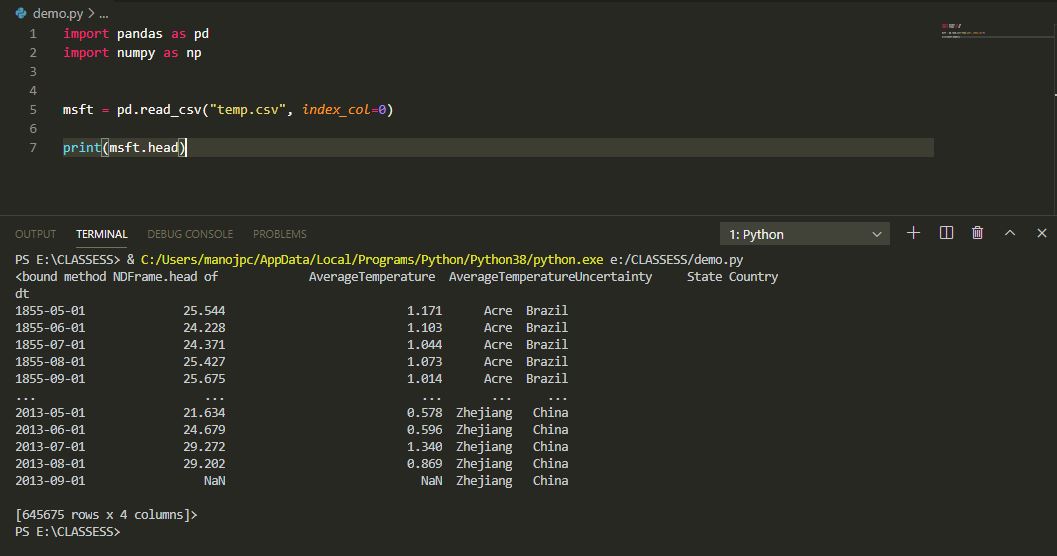
**No inheritance code**

In recent years databases like NoSQL don’t have inheritance code, that can implies they don't have to offer help for old equipment stages or keep abnormal and rarely utilized usefulness refreshed.

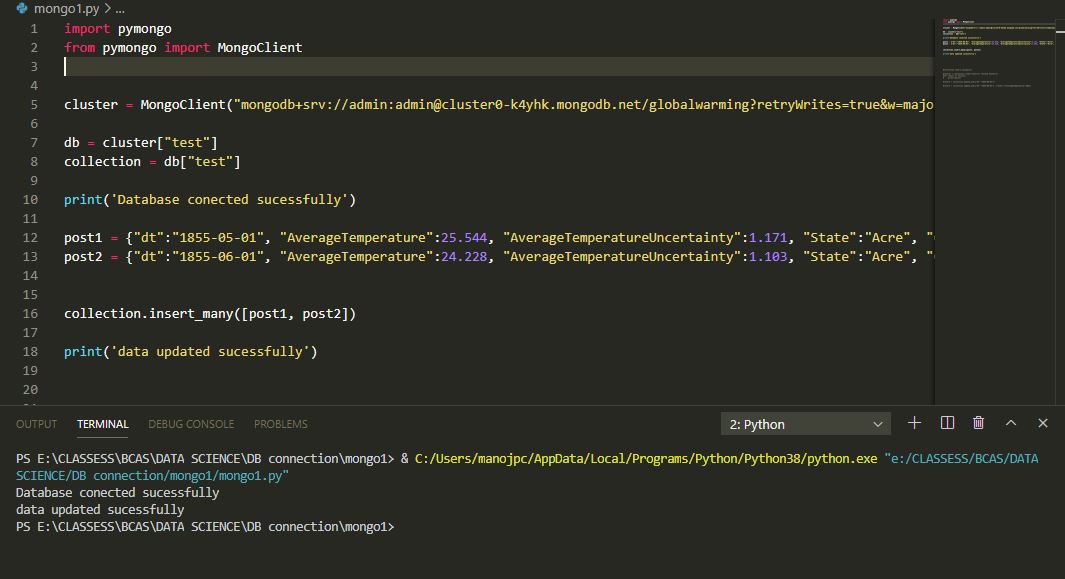
NoSQL databases appreciate a fast pace regarding improvement and development. New highlights are discharged constantly, and new and existing highlights are refreshed much of the time.

# **Task 3**

## Insert data set as CSV



## Mongo dB connection



Mongo DB link :

# **Task 4**

## Access data from database

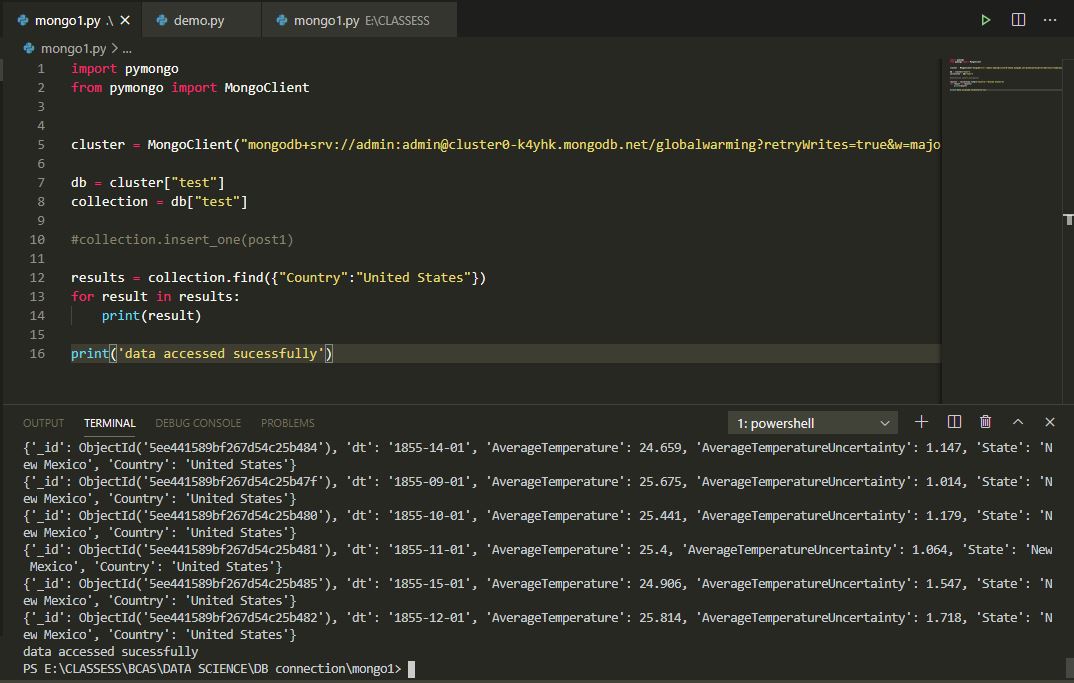
results = collection.find({"Country":"United States"})

for result in results:

    print(result)

This code gets all the data that has column name “country”

Result

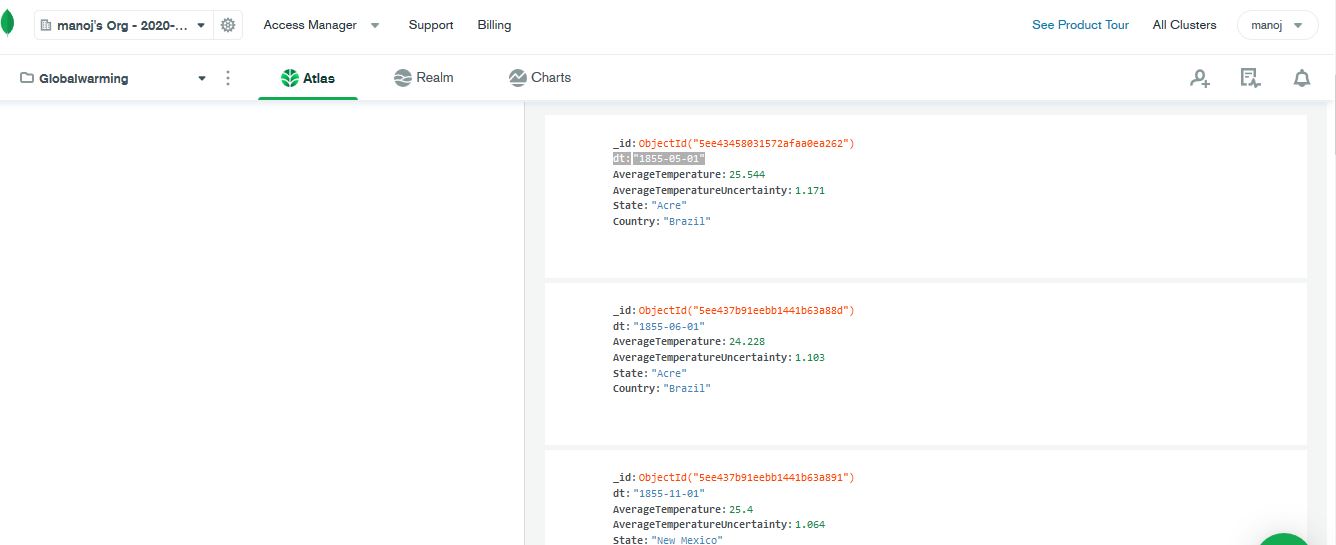


## Delete data from the database

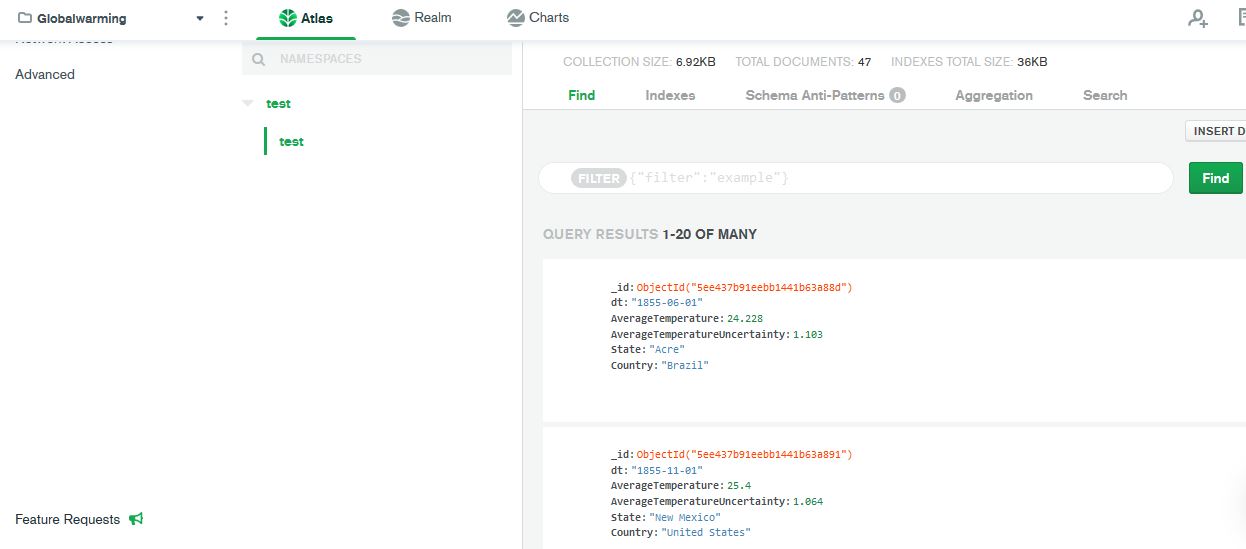
Result = collection.delete\_one({"dt":"1855-05-01"})

This code delete the data with date 1855-05-01

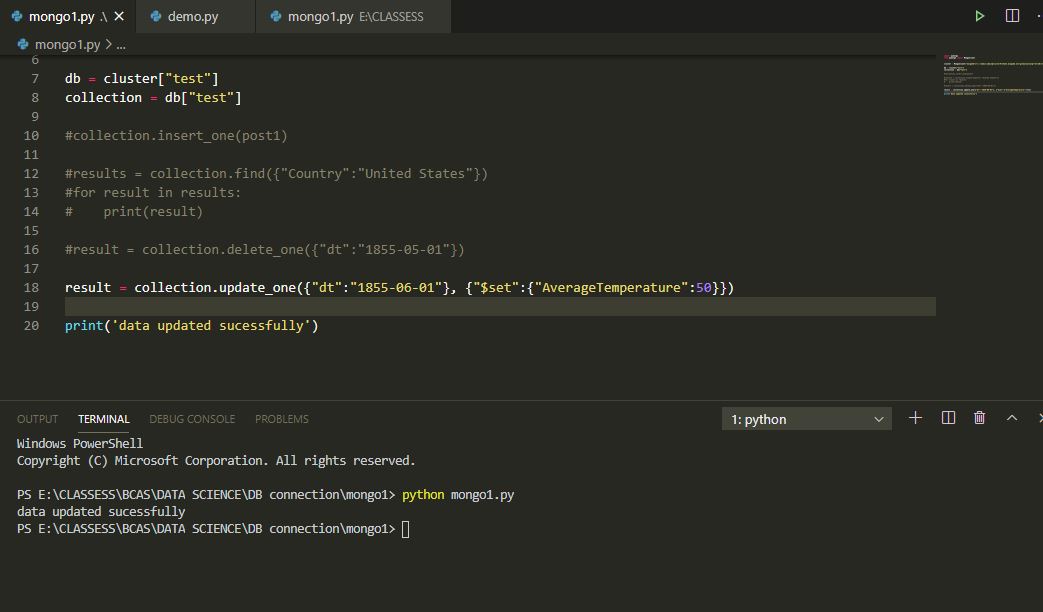
Result



After

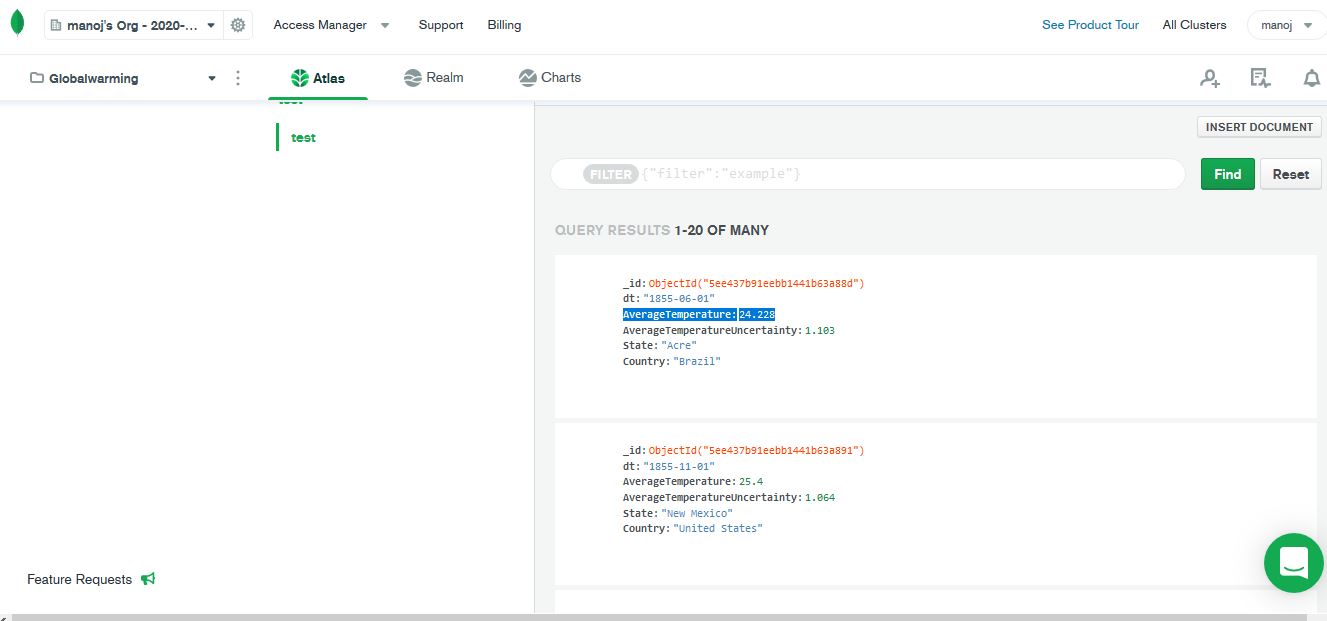


## Update data

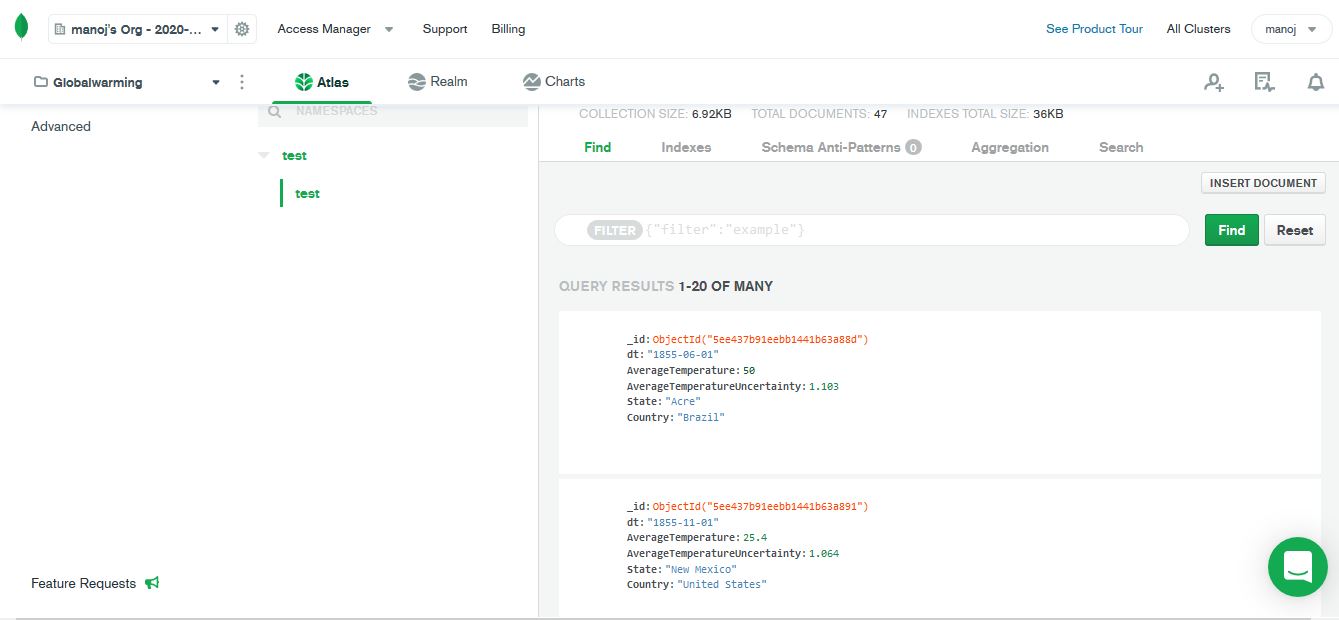


Result

Before update



After update



You can see the reverent data is being updated with new value.

# **Task 5**

Data mining is the process of looking at huge amount of information to create new information. Naturally, you may feel that information "mining" alludes to the extraction of new information, yet this isn't the situation; rather, information mining is tied in with extrapolating designs and new information from the information you've just gathered.

Depending on strategies and advances from the crossing point of database the executives, insights, and AI, masters in information mining have committed their vocations to better seeing how to process and make inferences from tremendous measures of data.

**Data Mining Techniques**

Information mining is exceptionally powerful, inasmuch as it draws upon at least one of these procedures:

1. **Tracking patterns.**

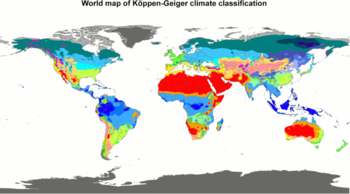
One of the most fundamental techniques in data mining is figuring out how to perceive designs in your informational indexes. This is typically an acknowledgment of some distortion in your information occurring at normal spans, or a back and forth movement of a specific variable after some time. For example, predicting rainfall in winter seasons.

**2. Classification.** Classification is a more complex data mining technique that forces you to collect various attributes together into discernable categories, which you can then use to draw further conclusions, or serve some function.

Geographers perceive various variables that influence a district's atmosphere:

* latitude
* elevation
* proximity to huge water bodies, mountains, or other surface highlights
* ocean flow designs
* long-term air flow

Together, these variables control the scope of temperatures and the measure of downpour as well as snow every locale gets as the year progressed. These elements control atmosphere, and thus, atmosphere controls environment—the kinds of local plants and creatures that live in an area.



1. **Association.**

Association is same as tracking patterns, however is increasingly explicit to conditionally connected factors. For this situation, you'll search for explicit occasions or traits that are exceptionally corresponded with another occasion or property; for instance, when a temperature increases in a particular area there might be a chance for tornado.

**4. Outlier detection.**

In many cases, simply recognizing the overarching pattern can’t give you a clear understanding of your data set. You also need to be able to identify anomalies, or outliers in your data.

**5. Clustering.**

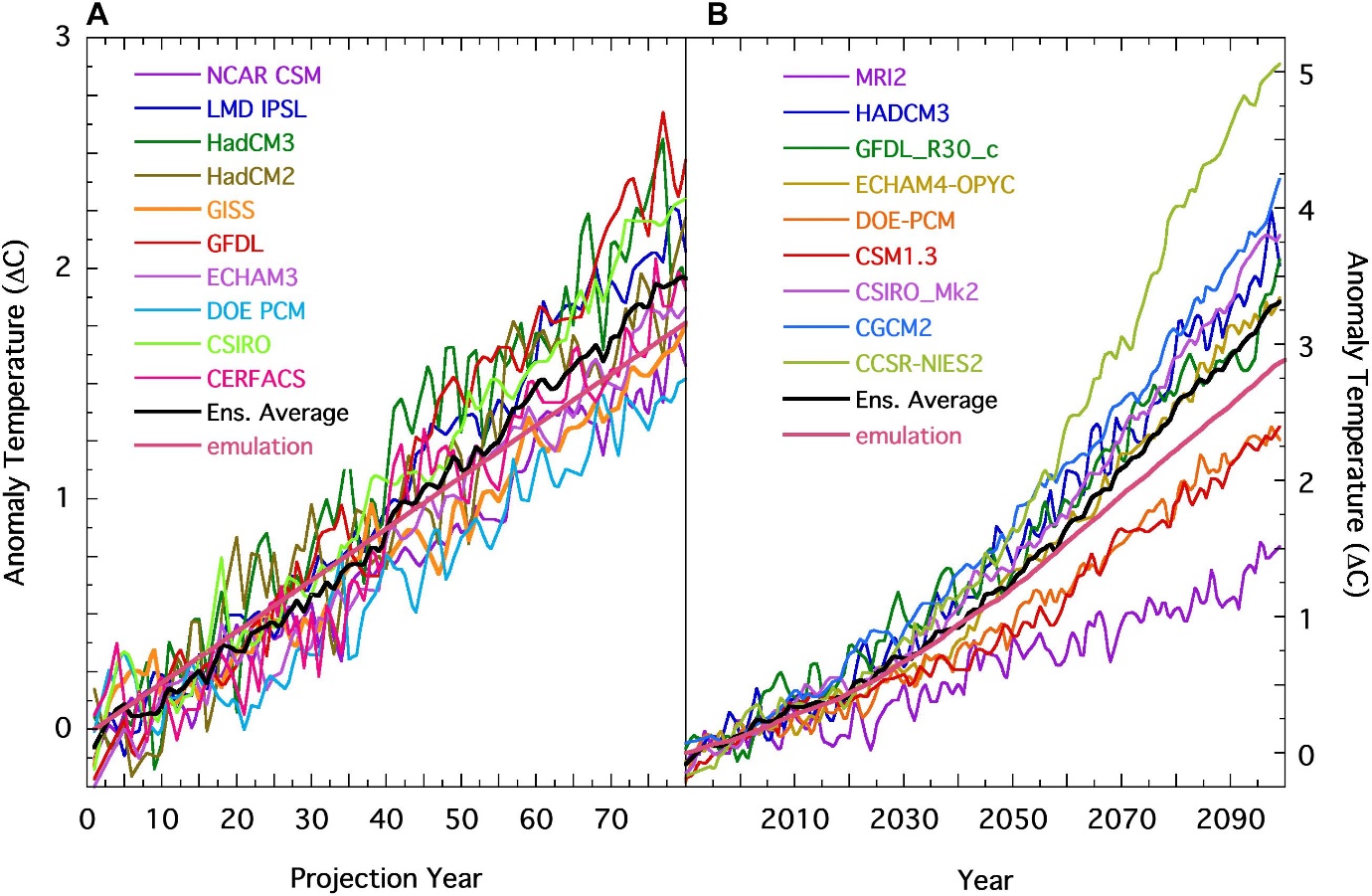
Clustering is fundamentally the same as order, yet includes gathering pieces of information dependent on their similarities. For instance, you can cluster global temperature rise and ice melting in poles of the planet and have a prediction what happens in future.

**6. Regression.**

Regression, used primarily as a form of planning and modeling, is used to identify the likelihood of a certain variable, given the presence of other variables.

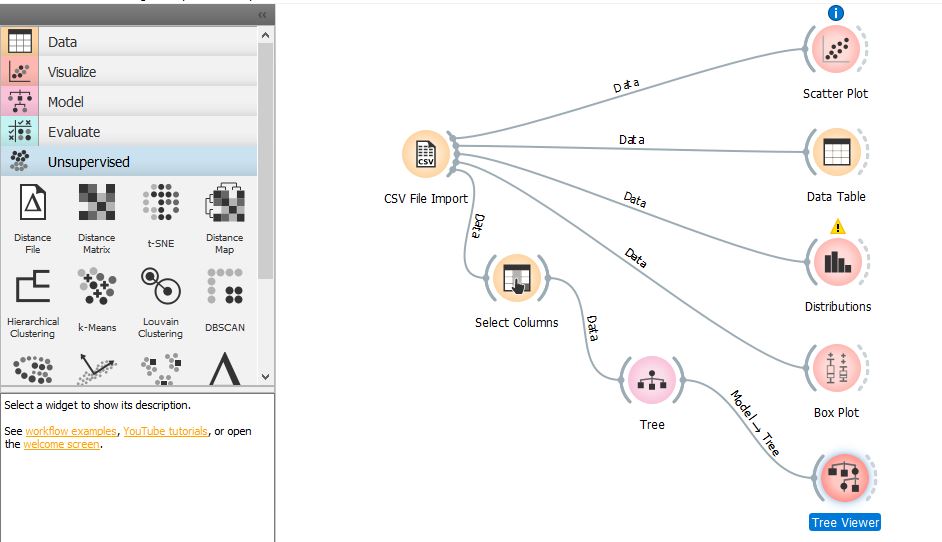
**7. Prediction.**

Prediction is one of the most significant information mining procedures, since it's utilized to extend the sorts of information you'll find later on. As a rule, simply perceiving and understanding recorded patterns is sufficient to graph a to some degree precise forecast of what will occur later on. For instance, you can predict the future environmental temperature by analyzing historical data of resent years.

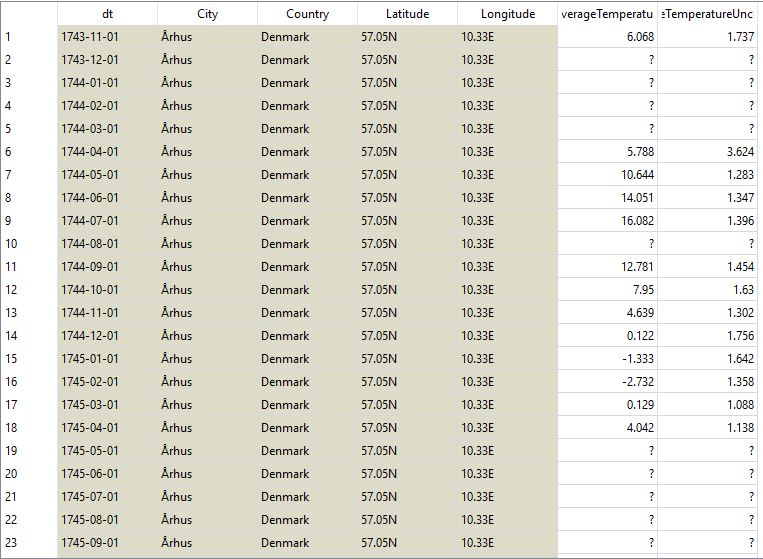


# **Task 6**

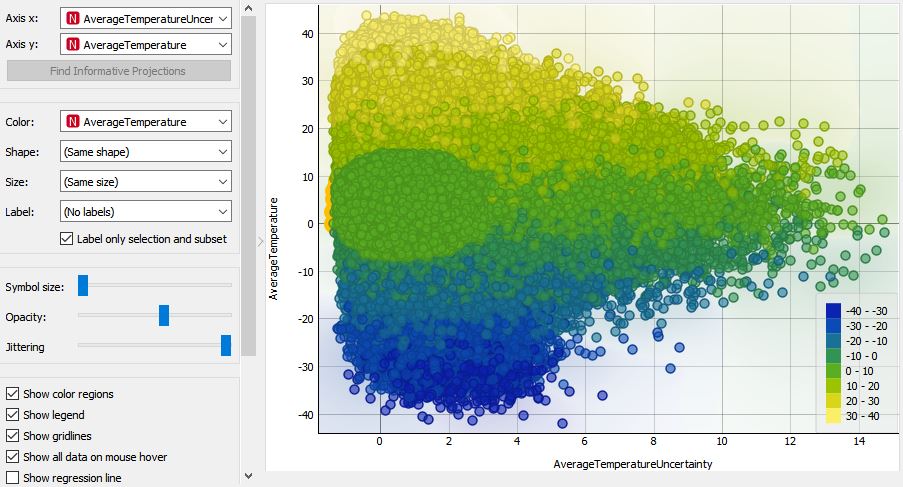
Importing csv file to orange



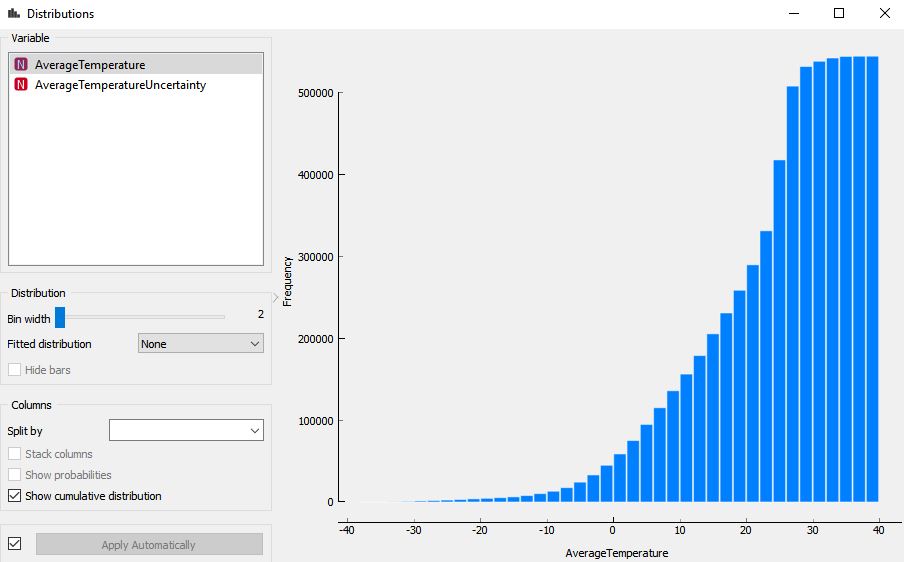
Data table



Scatter plot



Distributions



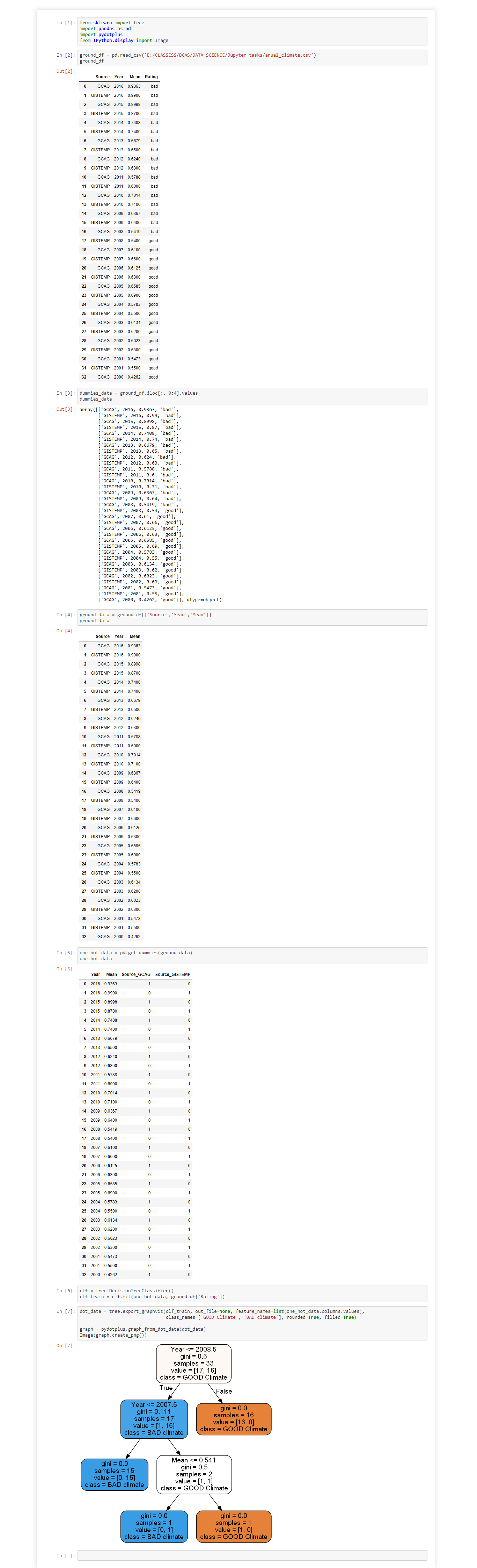
# **Task 7**

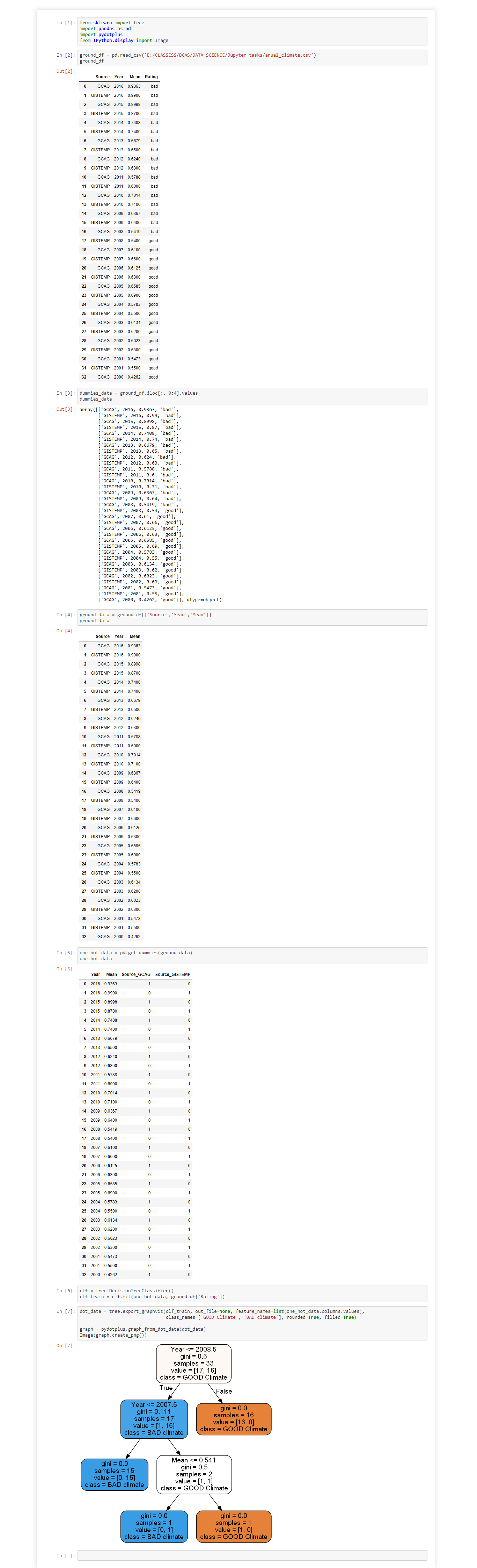
this is the method to cluster data for global mean temperature

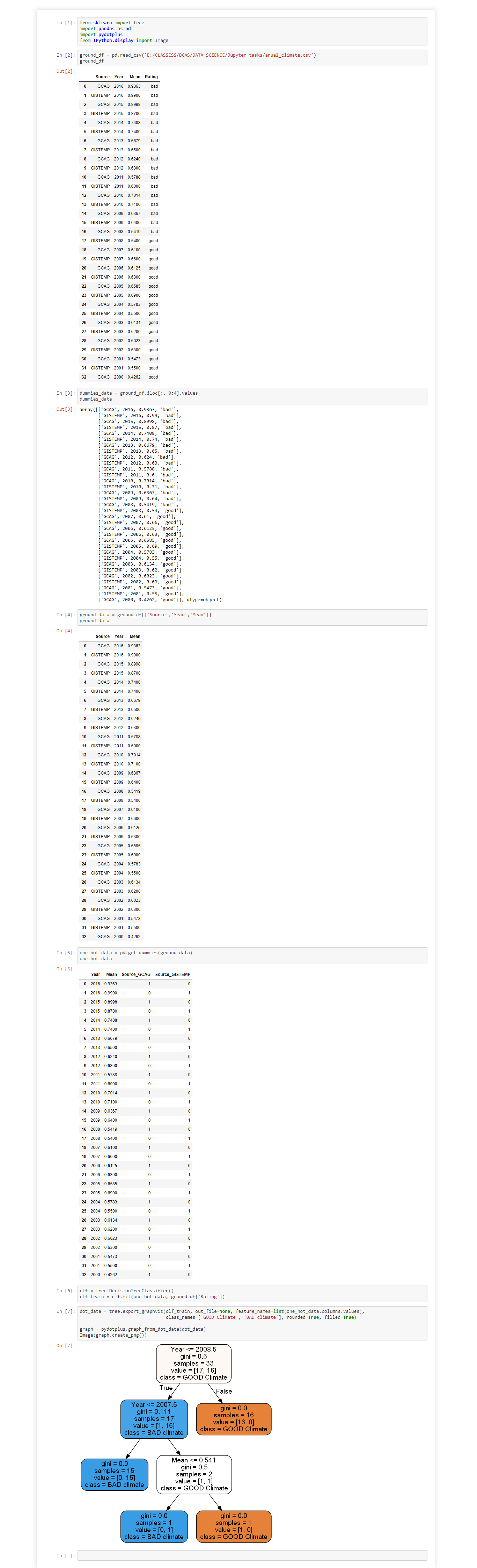


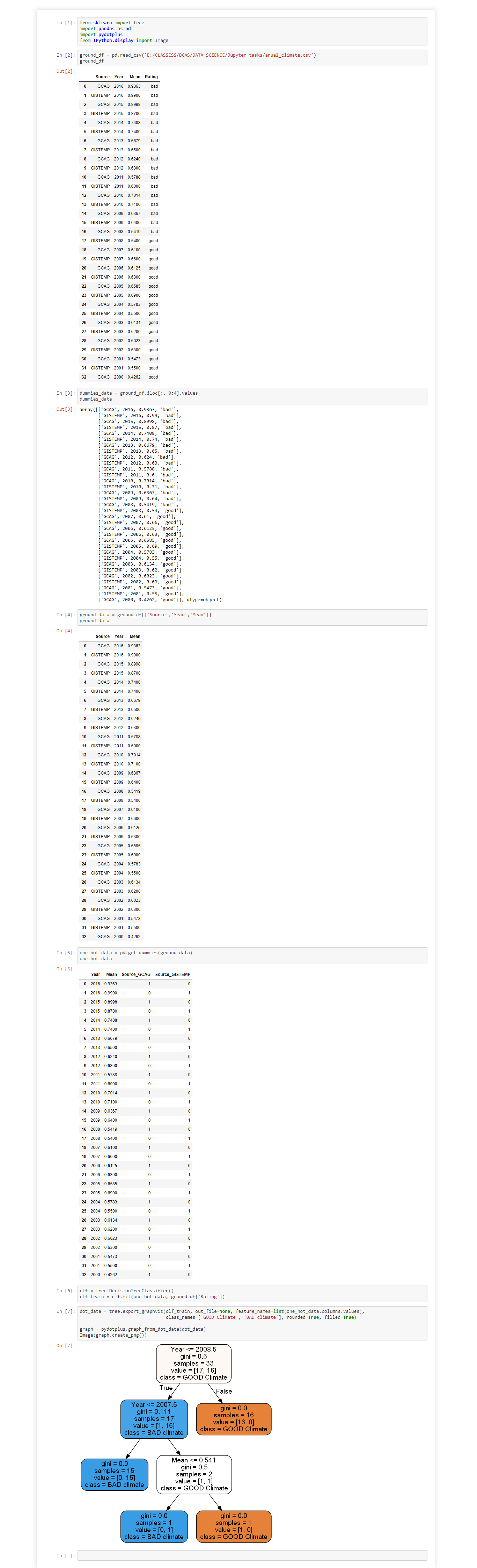


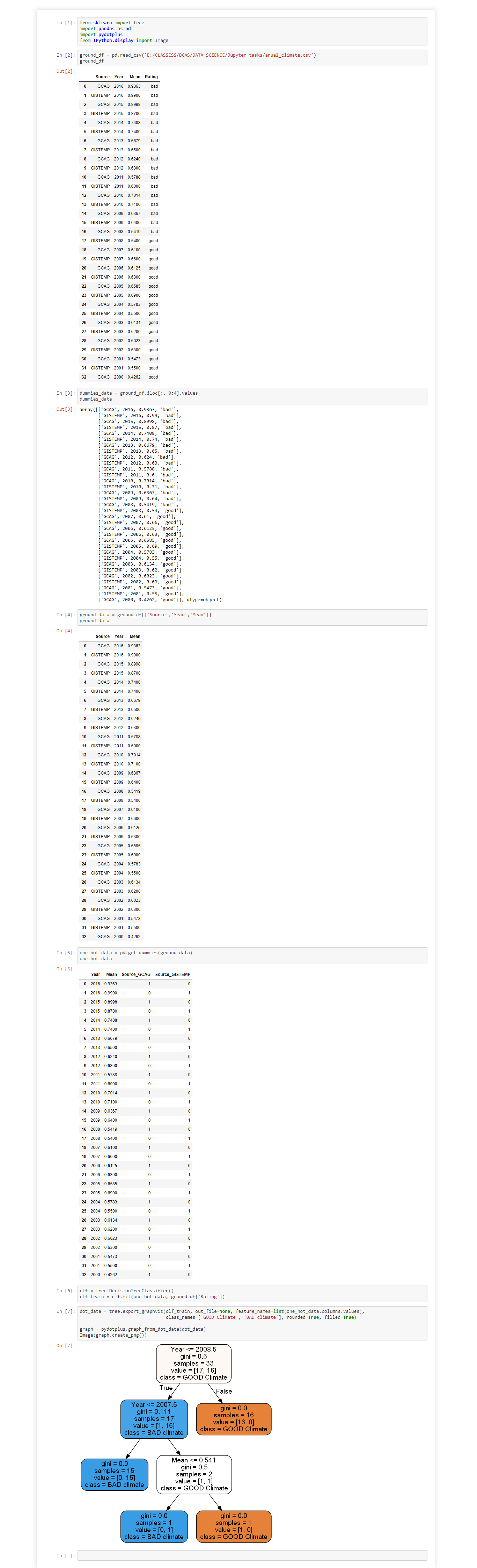
# **Task 8**











# **Task 9**

## **Apriori algorithm**

**Apriori algorithm is a algorithm for frequent item set mining and association rule learning over exchange databases. Its followed by finding the incessant individual things in the database and extending them to bigger and bigger thing sets as long as those thing sets show up adequately frequently in the database. The successive thing sets controlled by Apriori can be utilized to decide association rules which feature general patterns in the database.**

**Limitations**

* Apriori calculation can be extremely moderate.
* For example, if the exchange DB has 104 successive 1-itemsets, they will create 107 competitor 2-itemsets significantly in the wake of utilizing the descending conclusion.
* To register those with sup more than min sup, the database should be checked at each level. It needs (n +1 ) filters, where n is the length of the longest patern

**Methods to Improve Apriori’s Efficiency**

* Hash-based itemset counting: A k-itemset whose relating hashing pail tally is beneath the limit can't be visit
* Transaction reduction: An exchange that doesn't contain any continuous k-itemset is futile in ensuing outputs
* Partitioning: Any itemset that is possibly visit in DB must be visit in any event one of the segments of DB.
* Sampling: mining on a subset of given information, lower bolster edge + a technique to decide the culmination
* Dynamic itemset counting: include new competitor itemsets just when the entirety of their subsets are evaluated to be visit

**Apriori Advantages/Disadvantages**

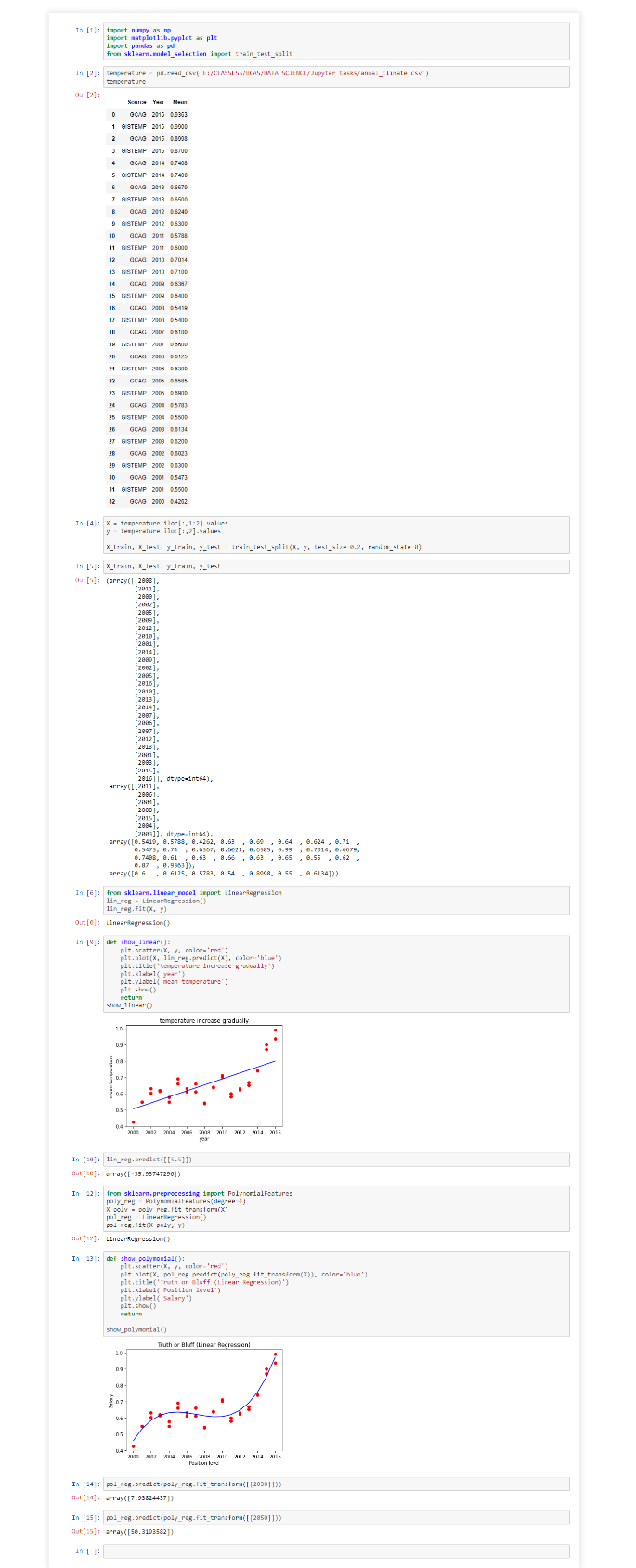
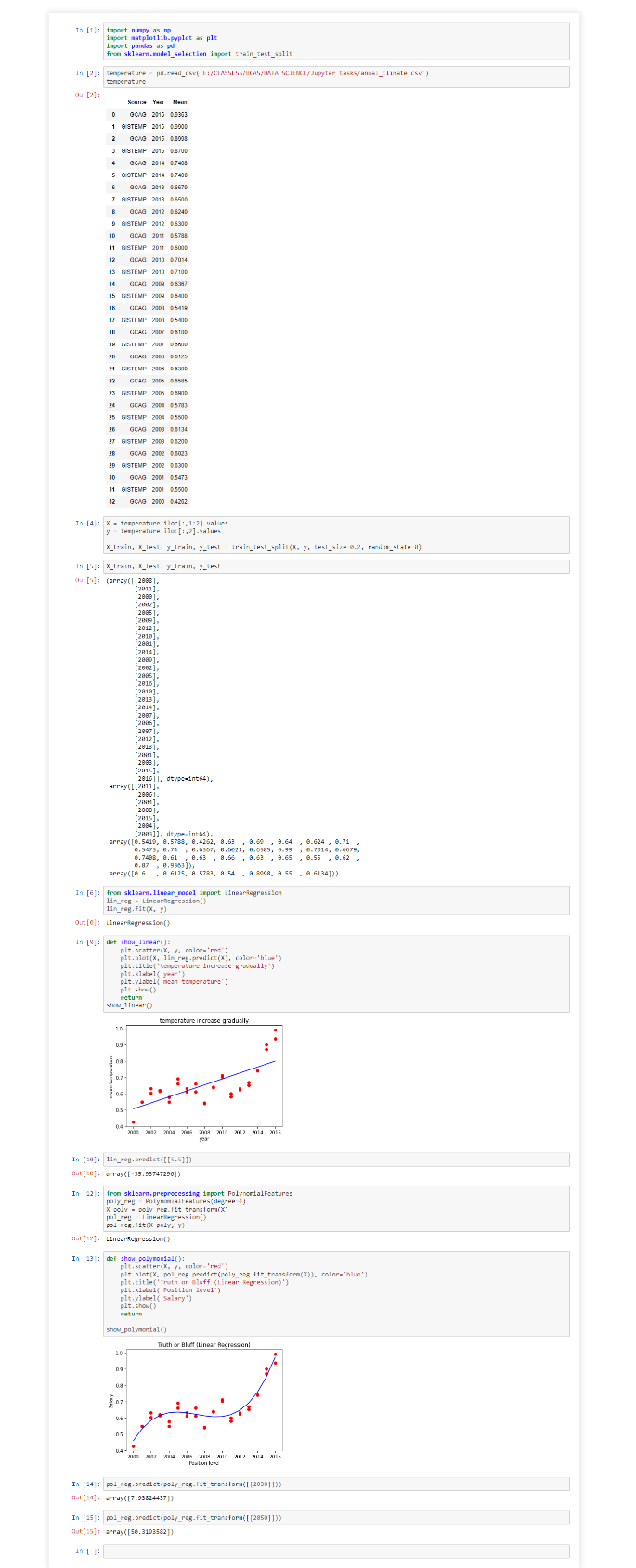
Advantages

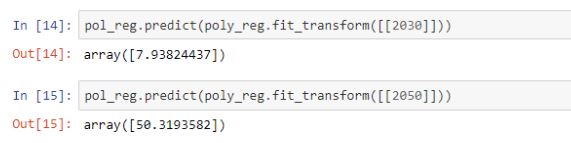
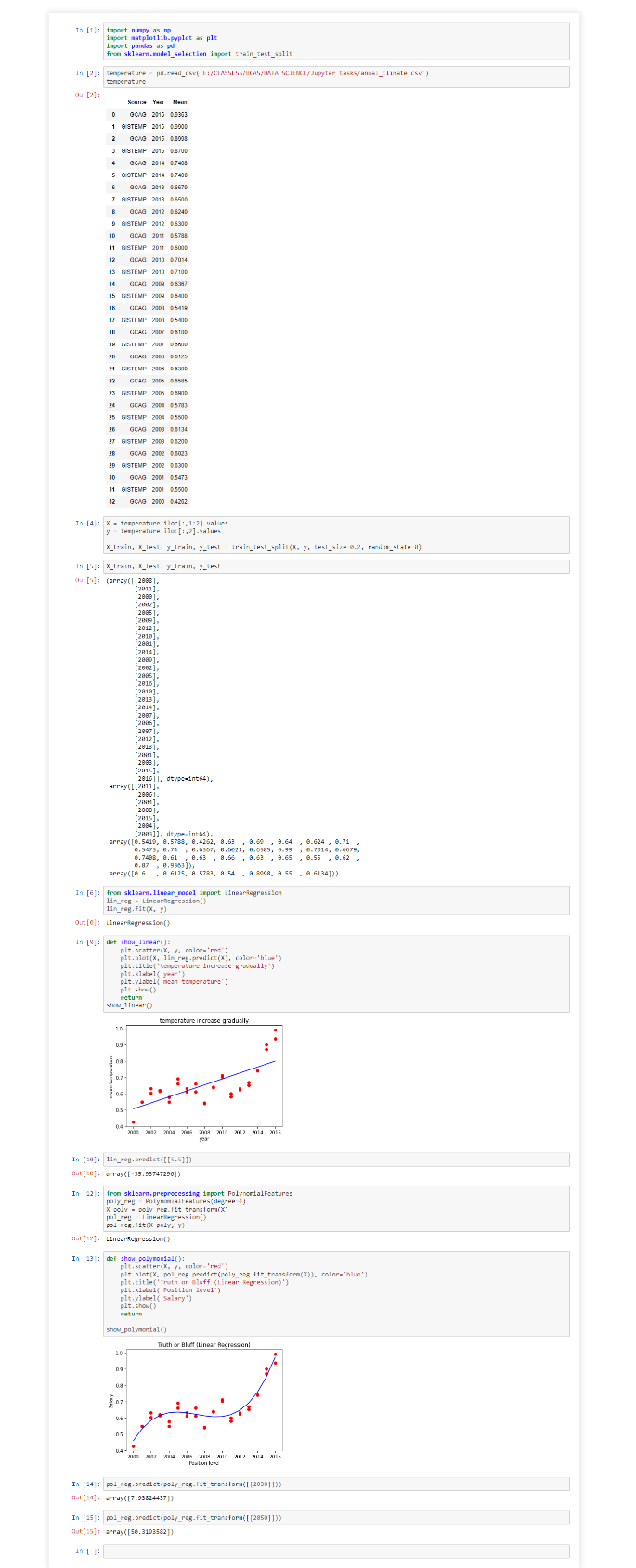
* Uses enormous itemset property
* Easily parallelized
* Easy to execute

Disadvantages

* Assumes exchange database is memory occupant.
* Requires numerous database checks

# **Task 10**





In the last line I predicted the mean temperature in 2050 will be 50.31035

# References

**There are no sources in the current document.**