Name: Syaidatul Salmah Nurbalqis Binti Saiful

Matric Number:17140336

WQD7009 INDIVIDUAL ASSIGNMENT

INTRODUCTION TO THE DATASET

The chosen dataset is from Kaggle with the name Japanese Cherry Blossom Data (link: https://www.kaggle.com/datasets/ryanglasnapp/japanese-cherry-blossom-data/data). This dataset shows the cherry blossom (sakura) first and full bloom dates from 1953 to 2023. Sakura blooming is strongly influenced by temperature as warmer conditions often leads to earlier blooming dates. The data was obtained from Japan Meteorological Agency's page on sakura blossoms and was translated from Japanese to English. The dataset has two files: one contains the first bloom dates, and the other one contains the full bloom dates of the cherry blossoms.

Prior to preprocessing, each dataset has 75 columns and 102 rows. Data preprocessing was done in Google Colab (link: https://colab.research.google.com/drive/1-G8NZcyjgqEcrZGaDBaPMn5sJ3FqiEpy?usp=sharing) with similar steps applied to both datasets. After processing, each dataset has 74 columns and 102 rows.

Preprocessing steps:

- Any missing data was replaced with NA.
- The text and date inputs were set as strings for better readability in Hbase.
- The column 'Currently Being Observed' was dropped as it is not relevant. If the dates are available in 2023, it shows that the site was observed.
- A new column 'site_ID' was added and was set as the row key for Hbase for easier data upload to Hbase.
- Prefix was added to the columns to match the Hbase column family structure.

Columns in the processed dataset:

Column name	Description
site_name	Name of the location where the sakura blossom observation took place.
1953 to 2023	70 columns in total, representing the cherry blossom blooming date for each year from 1953–2023.
average_first_bloom_date / average_first_bloom_date	The average blooming date over a 30-year period (1981–2010).
sakura_species	The cherry blossom species at the site.

QUERY AND ANALYSIS

Shell Commands

1. List all tables

```
hbase(main):013:0> list
TABLE
sakura_first_bloom
sakura_full_bloom
2 row(s) in 0.1460 seconds
```

The command 'list' lists all the tables in the Hbase database. Since we created two table, there should be two tables.

2. Hbase status

```
hbase(main):009:0> status
1 active master, 0 backup masters, 1 servers, 0 dead, 4.0000 average load
```

The command 'status' displays the status of the Hbase cluster. There is 1 Master Server that is managing the cluster, no Backup Master, 1 RegionServer in the cluster that is responsible for serving the data, 0 dead means all servers are running fine and the 4000 average load means that the single RegionServer is handling an average of 4000 regions.

3. Count rows

```
hbase(main):010:0> count 'sakura_first_bloom'
102 row(s) in 0.4690 seconds

=> 102
hbase(main):011:0> count 'sakura_full_bloom'
102 row(s) in 0.0770 seconds

=> 102
```

Since the dataset has 102 rows, we use the count command to verify the number of rows in the table and ensure that the CSV file has been uploaded correctly.

4. Describe table

```
hbase(main):014:0> describe 'sakura_first_bloom'
Table sakura_first_bloom is ENABLED
sakura_first_bloom
COLUMN FAMILIES DESCRIPTION
{NAME => 'metadata', DATA_BLOCK_ENCODING => 'NONE', BLOOMFILTER => 'ROW', REPLICATION_SCOPE
 => '0', VERSIONS => '1', COMPRESSION => 'NONE', MIN_VERSIONS => '0', TTL => 'FOREVER', KEE
P_DELETED_CELLS => 'FALSE', BLOCKSIZE => '65536', IN_MEMORY => 'false', BLOCKCACHE => 'true
'}
{NAME => 'year', DATA_BLOCK_ENCODING => 'NONE', BLOOMFILTER => 'ROW', REPLICATION_SCOPE =>
'0', VERSIONS => '1', COMPRESSION => 'NONE', MIN_VERSIONS => '0', TTL => 'FOREVER', KEEP_DE
LETED_CELLS => 'FALSE', BLOCKSIZE => '65536', IN_MEMORY => 'false', BLOCKCACHE => 'true'}
2 row(s) in 0.1260 seconds

hbase(main):015:0> describe 'sakura_full_bloom'
Table sakura_full_bloom is ENABLED
sakura_full_bloom
COLUMN FAMILIES DESCRIPTION
{NAME => 'metadata', DATA_BLOCK_ENCODING => 'NONE', BLOOMFILTER => 'ROW', REPLICATION_SCOPE => '0',
VERSIONS => '1', COMPRESSION => 'NONE', MIN_VERSIONS => '0', TTL => 'FOREVER', KEEP_DELETED_CELLS =>
 'FALSE', BLOCKSIZE => '65536', IN_MEMORY => 'false', BLOCKCACHE => 'true'}
{NAME => 'year', DATA_BLOCK_ENCODING => 'NONE', BLOOMFILTER => 'ROW', REPLICATION_SCOPE => '0', VERS
IONS => '1', COMPRESSION => 'NONE', MIN_VERSIONS => '0', TTL => 'FOREVER', KEEP_DELETED_CELLS => 'FA
LSE', BLOCKSIZE => '65536', IN_MEMORY => 'false', BLOCKCACHE => 'true'}
2 row(s) in 0.1300 seconds
```

The two tables, sakura_first_bloom and sakura_full_bloom, have similar structures with two column families: metadata and year. Both tables use the same settings: DATA_BLOCK_ENCODING is set to NONE, meaning no special encoding is applied; BLOOMFILTER is set to ROW, which optimizes lookups by row; only the latest version of data is kept with VERSIONS set to 1; TTL is FOREVER meaning the data never expires; and BLOCKCACHE is enabled to speed up data retrieval by caching it in memory.

5. Check the Hbase version

```
hbase(main):005:0> version
1.2.0-cdh5.10.0, rUnknown, Fri Jan 20 12:13:18 PST 2017
```

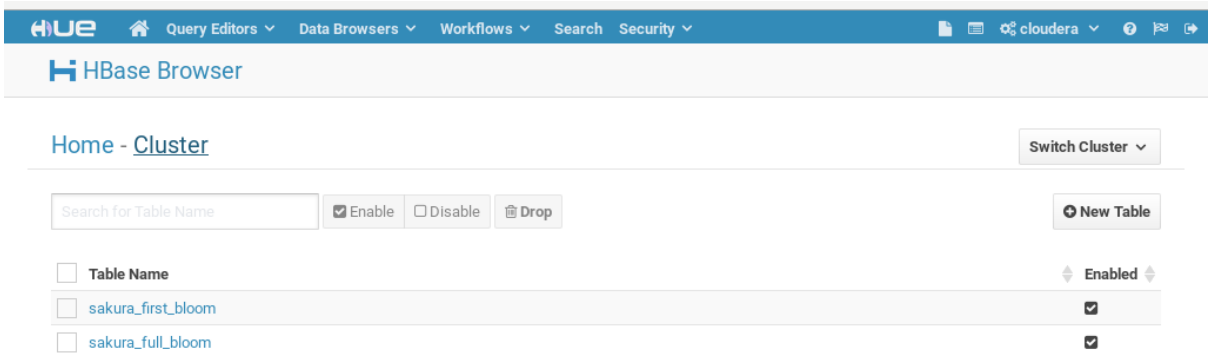
Checking the HBase version is important as some commands like ‘clone’ works for HBase version 2.x and later. Knowing the version will enable us to plan our queries command.

Data Definition Language (DDL) Queries

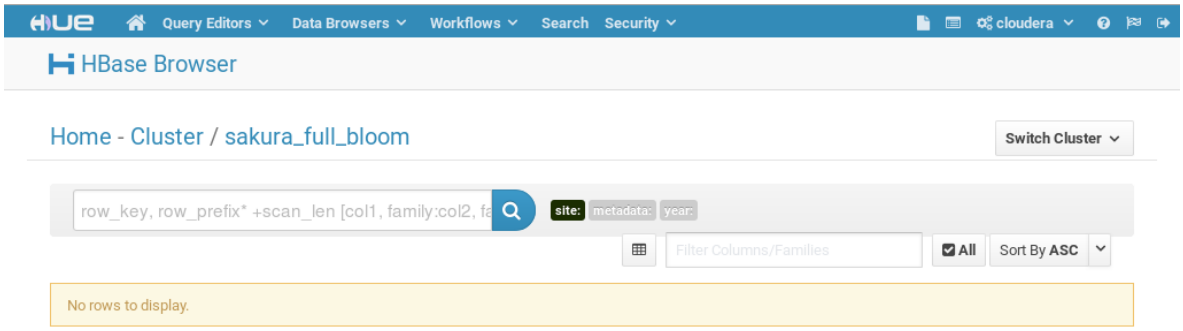
1. Create tables

```
hbase(main):001:0> create 'sakura_first_bloom','site','year','metadata'
0 row(s) in 2.4900 seconds

=> Hbase::Table - sakura_first_bloom
hbase(main):002:0> create 'sakura_full_bloom','site','year','metadata'
0 row(s) in 2.2520 seconds

=> Hbase::Table - sakura_full_bloom
```

Two tables, sakura_first_bloom and sakura_full_bloom, were created in HBase. Each table has three column families: site, year, and metadata. After the tables were created, they were verified in Hue, a web-based interface for interacting with Hadoop and HBase to ensure that they were set up correctly and are accessible for querying and analysis.



2. Drop column family

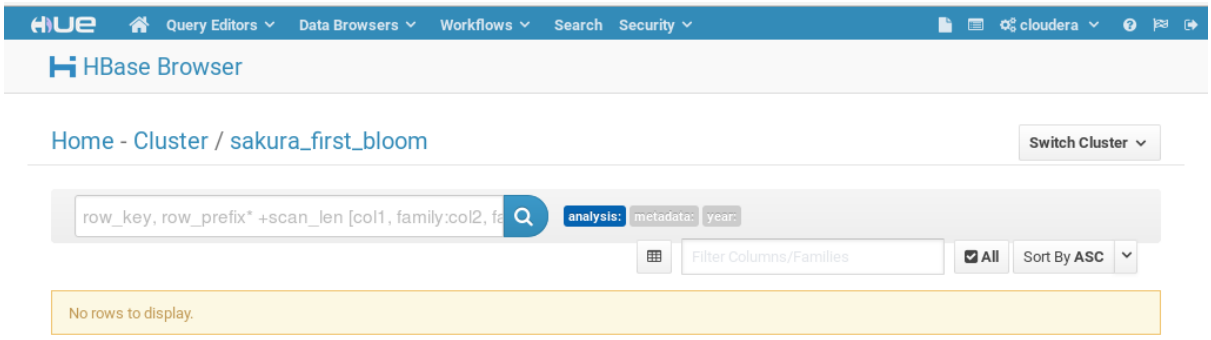


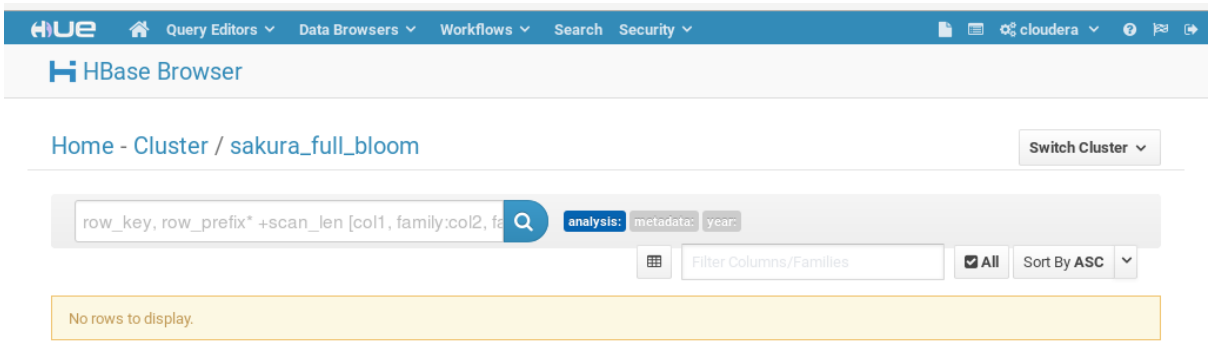

```
hbase(main):007:0> alter 'sakura_first_bloom', {METHOD => 'delete', NAME => 'site'}
Updating all regions with the new schema...
0/1 regions updated.
1/1 regions updated.
Done.
0 row(s) in 4.7120 seconds

hbase(main):008:0> alter 'sakura_full_bloom', {METHOD => 'delete', NAME => 'site'}
Updating all regions with the new schema...
0/1 regions updated.
1/1 regions updated.
Done.
0 row(s) in 4.3100 seconds
```

After uploading the CSV files to HBase, the column families were checked in Hue to verify the data. It was observed that there were no rows under the site column family for both the sakura_first_bloom and sakura_full_bloom tables. The site column family was dropped from both tables since it was not being used or populated with any data. This action was taken to clean up the schema and ensure that only relevant column families remain in the tables.

3. Alter table by adding new column family

```
=> Hbase::Table - sakura_full_bloom
hbase(main):003:0> alter 'sakura_first_bloom',{NAME => 'analysis'}
Updating all regions with the new schema...
0/1 regions updated.
1/1 regions updated.
Done.
0 row(s) in 4.4040 seconds

hbase(main):004:0> alter 'sakura_full_bloom',{NAME => 'analysis'}
Updating all regions with the new schema...
0/1 regions updated.
1/1 regions updated.
Done.
0 row(s) in 3.4030 seconds
```





A new column family was added to the table to store information derived from the analysis. The addition of the column family was successfully verified using Hue.

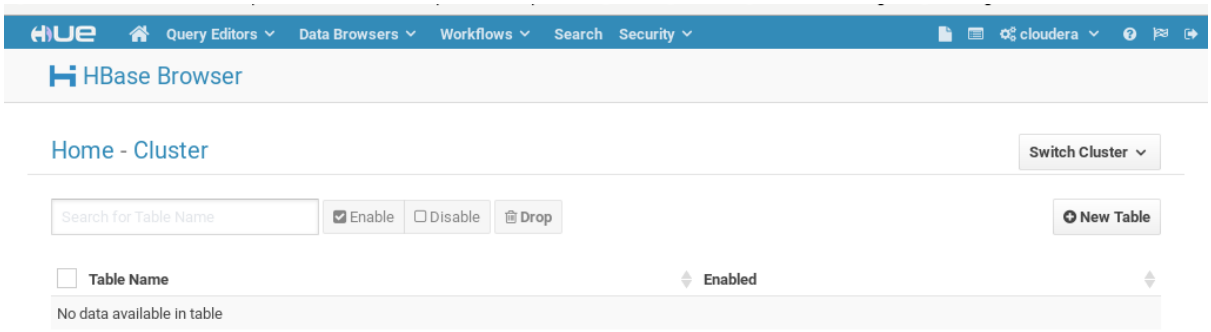
4. Disable table

```
hbase(main):009:0> disable 'sakura_first_bloom'
0 row(s) in 7.0160 seconds

hbase(main):010:0> disable 'sakura_full_bloom'
0 row(s) in 4.4110 seconds

hbase(main):011:0> drop 'sakura_first_bloom'
0 row(s) in 2.5300 seconds

hbase(main):012:0> drop 'sakura_full_bloom'
0 row(s) in 1.4690 seconds
```

Before deleting the table, we must disable the table. Disabling the table ensures that no operations can be performed on it that might interfere with the deletion process.

5. Delete table

```
hbase(main):011:0> drop 'sakura_first_bloom'
0 row(s) in 2.5300 seconds

hbase(main):012:0> drop 'sakura_full_bloom'
0 row(s) in 1.4690 seconds
```



After completing the analysis, the tables were deleted as they were created specifically for this project. To confirm the deletion, we checked in Hue which showed that no tables remained, verifying the successful removal.

Data Manipulation Language (DML) Queries

1. Retrieve data for Kyoto for specific years

```
hbase(main):003:0> get 'sakura_first_bloom','Kyoto','year:2004'
COLUMN                      CELL
 year:2004                  timestamp=1732677771433, value=2004-03-24
1 row(s) in 0.0750 seconds

hbase(main):004:0> get 'sakura_first_bloom','Kyoto','year:2023'
COLUMN                      CELL
 year:2023                  timestamp=1732677771433, value=2023-03-17
1 row(s) in 0.0260 seconds
```

Kyoto is one of the famous places in Japan for cherry blossoms viewing. The first bloom date in 2004 was 24th March, while in 2023, just 20 years later, it occurred a week earlier on 17th March. This shift highlights the impact of global warming, as warmer temperatures are accelerating the arrival of spring and influencing the timing of cherry blossom blooms.

```
hbase(main):005:0> get 'sakura_full_bloom','Kyoto','year:2004'
COLUMN                       CELL
 year:2004                   timestamp=1732677789420, value=2004-04-03
1 row(s) in 0.0870 seconds

hbase(main):006:0> get 'sakura_full_bloom','Kyoto','year:2023'
COLUMN                       CELL
 year:2023                   timestamp=1732677789420, value=2023-03-24
1 row(s) in 0.0310 seconds
```

Similarly, the full bloom date has also shifted earlier. In 2004, it occurred on 3rd April, while in 2023, it was on 24th March.

In 2004, it took 10 days for the cherry blossoms to go from first bloom to full bloom. However, in 2023, this period was shortened to just 7 days. This suggests that warmer temperatures in 2023 may have accelerated the blooming process causing the blossoms to reach full bloom more quickly after the first bloom.

2. Retrieve data for Osaka for specific years

```
hbase(main):001:0> get 'sakura_first_bloom','Osaka','year:2004'
COLUMN                CELL
 year:2004            timestamp=1732677771433, value=2004-03-23
1 row(s) in 0.2880 seconds

hbase(main):002:0> get 'sakura_first_bloom','Osaka','year:2023'
COLUMN                CELL
 year:2023            timestamp=1732677771433, value=2023-03-19
1 row(s) in 0.0160 seconds
```

Osaka is also a famous city to see cherry blossoms in Japan. The first bloom date in 2004 is 23rd March while in 2023 it was on 19th March, which was 4 days earlier than in 2004.

```
hbase(main):003:0> get 'sakura_full_bloom','Osaka','year:2004'
COLUMN                CELL
 year:2004            timestamp=1732677789420, value=2004-04-02
1 row(s) in 0.0360 seconds

hbase(main):004:0> get 'sakura_full_bloom','Osaka','year:2023'
COLUMN                CELL
 year:2023            timestamp=1732677789420, value=2023-03-27
1 row(s) in 0.0160 seconds
```

The full bloom date also shifted to an earlier date in 2023 occurring on 27th March while in 2004 it was on 2nd of April.

The time for the cherry blossoms to go from first bloom to full bloom in 2004 was 10 days while in 2023 this period was reduced to just 8 days. These shifts in blooming dates highlight the impact of climate change, as warmer temperatures are accelerating the blooming process.

3. Scan average bloom date for both first bloom and full bloom for Osaka and Kyoto

```
hbase(main):021:0> scan 'sakura_first_bloom', {
  FILTER => "RowFilter(=, 'regexstring:Kyoto|Osaka')",
  COLUMNS => ['metadata:average_first_bloom_date']
}
ROW                             COLUMN+CELL
 Kyoto                          column=metadata:average_first_bloom_date, timestamp=1732804702235, value=March 26
 Osaka                          column=metadata:average_first_bloom_date, timestamp=1732804702235, value=March 27
2 row(s) in 0.0300 seconds

hbase(main):025:0> scan 'sakura_full_bloom', {
  FILTER => "RowFilter(=, 'regexstring:Kyoto|Osaka')",
  COLUMNS => ['metadata:average_full_bloom_date']
}
ROW                             COLUMN+CELL
 Kyoto                          column=metadata:average_full_bloom_date, timestamp=1732804756780, value=April 4
 Osaka                          column=metadata:average_full_bloom_date, timestamp=1732804756780, value=April 4
2 row(s) in 0.0660 seconds
```

Kyoto and Osaka are both located in the Kansai region in central Japan which experiences a typical temperate climate. Over a 30-year period (1981–2010), the average first bloom date was March 26 in Kyoto and March 27 in Osaka while the average full bloom date for both cities falling on April 4. The time from first bloom to full bloom typically spans 9-10 days which is considered a normal blooming period for cherry blossoms.

The effects of climate change on blooming times are not clearly visible in the long-term average dates. The climates in Kyoto and Osaka also align with the usual spring season which spans from March to May as cherry blossoms are a natural indicator of the arrival of spring.

4. Delete row

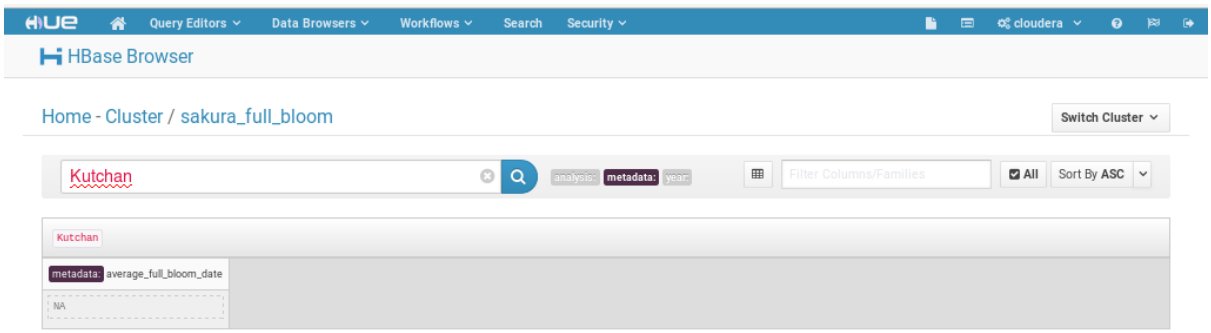
```
hbase(main):001:0> scan 'sakura_first_bloom', {COLUMNS => ['metadata:sakura_spec
ies']}
ROW                   COLUMN+CELL
 Abashiri             column=metadata:sakura_species, timestamp=1732804702235, v
                      alue=Sargent cherry (Prunus sargentii)
 Aikawa               column=metadata:sakura_species, timestamp=1732804702235, v
                      alue=NA
```

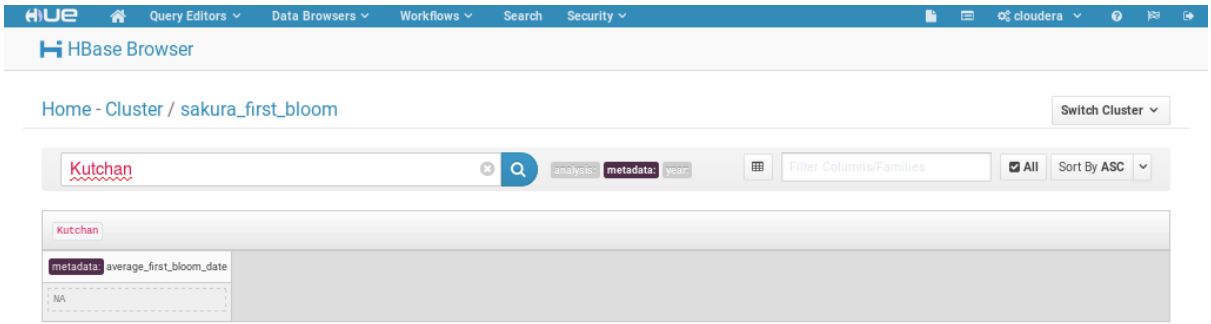
```
hbase(main):002:0> scan 'sakura_full_bloom', {COLUMNS => ['metadata:sakura_species']}
ROW                           COLUMN+CELL
 Abashiri                     column=metadata:sakura_species, timestamp=1732804756780, value=Sargent cherry (Prunus sargen
                              tii)
 Aikawa                       column=metadata:sakura_species, timestamp=1732804756780, value=NA
```

```
 Kushiro              column=metadata:sakura_species, timestamp=1732804702235, v
                      alue=Sargent cherry (Prunus sargentii)
 Kutchan              column=metadata:sakura_species, timestamp=1732804702235, v
                      alue=Until 1994 Sargent Cherry, from 1995 to 2006 they wer
                      e Yoshino Cherry.
 Kyoto                column=metadata:sakura_species, timestamp=1732804702235, v
                      alue=NA
```

The column `sakura_species` was scanned to identify the available species at the site. For the site "Kutchan," no specific current species name was provided. As a result, this entry was deleted to focus exclusively on the current cherry blossom species at each site. The deletion was successfully verified in Hue.

```
hbase(main):003:0> delete 'sakura_first_bloom', 'Kutchan', 'metadata:sakura_species'
0 row(s) in 0.1220 seconds

hbase(main):004:0> delete 'sakura_full_bloom', 'Kutchan', 'metadata:sakura_species'
0 row(s) in 0.0480 seconds
```





5. Scan data with a specific filter for species Taiwan cherry (Prunus campanulata)

```
hbase(main):011:0> scan 'sakura_first_bloom', {FILTER => "ValueFilter(=, 'binary:Taiwan cherry (Prunus campanulata)')"}
ROW                             COLUMN+CELL
 Iriomote Island                column=metadata:sakura_species, timestamp=1732677771433, value=Taiwan cherry (Prunus campanu
                                lata)
 Ishigaki Island                column=metadata:sakura_species, timestamp=1732677771433, value=Taiwan cherry (Prunus campanu
                                lata)
 Kumejima                       column=metadata:sakura_species, timestamp=1732677771433, value=Taiwan cherry (Prunus campanu
                                lata)
 Minami Daito Island            column=metadata:sakura_species, timestamp=1732677771433, value=Taiwan cherry (Prunus campanu
                                lata)
 Miyakojima                     column=metadata:sakura_species, timestamp=1732677771433, value=Taiwan cherry (Prunus campanu
                                lata)
 Nago                           column=metadata:sakura_species, timestamp=1732677771433, value=Taiwan cherry (Prunus campanu
                                lata)
 Naha                           column=metadata:sakura_species, timestamp=1732677771433, value=Taiwan cherry (Prunus campanu
                                lata)
 Naze                           column=metadata:sakura_species, timestamp=1732677771433, value=Taiwan cherry (Prunus campanu
                                lata)
 Yonaguni Island                column=metadata:sakura_species, timestamp=1732677771433, value=Taiwan cherry (Prunus campanu
                                lata)
9 row(s) in 0.2030 seconds
```

The cherry blossom species vary by region due to differences in climate and geography. There are 9 sites with Taiwan cherry (Prunus campanulata). All 9 sites are in Japan’s southernmost region with a subtropical climate with warm temperatures year-round.

6. Retrieve the first and full bloom dates in Miyakojima, a site with Taiwan cherry

```
hbase(main):017:0> get 'sakura_first_bloom', 'Miyakojima', 'year:1953'
COLUMN                          CELL
 year:1953                      timestamp=1732677771433, value=NA
1 row(s) in 0.0560 seconds

hbase(main):018:0> get 'sakura_first_bloom', 'Miyakojima', 'year:1978'
COLUMN                          CELL
 year:1978                      timestamp=1732677771433, value=1978-02-01
1 row(s) in 0.0620 seconds

hbase(main):019:0> get 'sakura_first_bloom', 'Miyakojima', 'year:2023'
COLUMN                          CELL
 year:2023                      timestamp=1732677771433, value=2023-01-15
1 row(s) in 0.0250 seconds
```

In 1953, the first bloom date is not available. After 25 years the first bloom date is on 1st February 1978 and 45 years later in 2023, it has an earlier blooming date which is 15th January.

```
hbase(main):020:0> get 'sakura_full_bloom', 'Miyakojima', 'year:1953'
COLUMN                          CELL
 year:1953                      timestamp=1732677789420, value=NA
1 row(s) in 0.0900 seconds

hbase(main):021:0> get 'sakura_full_bloom', 'Miyakojima', 'year:1978'
COLUMN                          CELL
 year:1978                      timestamp=1732677789420, value=NA
1 row(s) in 0.0350 seconds

hbase(main):022:0> get 'sakura_full_bloom', 'Miyakojima', 'year:2023'
COLUMN                          CELL
 year:2023                      timestamp=1732677789420, value=2023-02-03
1 row(s) in 0.0150 seconds
```

The full bloom date was not available for year 1953 and 1978 while in 2023 the full bloom date is 3rd of February.

In 2023, it took 19 days for the cherry blossoms to progress from first bloom to full bloom indicating stable weather conditions. This lack of fluctuations allowed the blossoming process to unfold more gradually rather than rapidly leading to a slower progression to full bloom.

7. Analyse first bloom changes over time for Miyakojima

```
hbase(main):016:0> scan 'sakura_first_bloom',{FILTER=>"PrefixFilter('Miyakojima')"}
ROW                           COLUMN+CELL
 Miyakojima                   column=metadata:average_first_bloom_date, timestamp=1732677771433, value=January 17
 Miyakojima                   column=metadata:sakura_species, timestamp=1732677771433, value=Taiwan cherry (Prunus campanu
                              lata)
 Miyakojima                   column=year:1953, timestamp=1732677771433, value=NA
 Miyakojima                   column=year:1954, timestamp=1732677771433, value=NA
 Miyakojima                   column=year:1955, timestamp=1732677771433, value=NA
 Miyakojima                   column=year:1956, timestamp=1732677771433, value=NA
 Miyakojima                   column=year:1957, timestamp=1732677771433, value=NA
 Miyakojima                   column=year:1958, timestamp=1732677771433, value=NA
 Miyakojima                   column=year:1959, timestamp=1732677771433, value=NA
 Miyakojima                   column=year:1960, timestamp=1732677771433, value=NA
 Miyakojima                   column=year:1961, timestamp=1732677771433, value=NA
 Miyakojima                   column=year:1962, timestamp=1732677771433, value=NA
 Miyakojima                   column=year:1963, timestamp=1732677771433, value=NA
 Miyakojima                   column=year:1964, timestamp=1732677771433, value=NA
 Miyakojima                   column=year:1965, timestamp=1732677771433, value=NA
 Miyakojima                   column=year:1966, timestamp=1732677771433, value=NA
 Miyakojima                   column=year:1967, timestamp=1732677771433, value=NA
 Miyakojima                   column=year:1968, timestamp=1732677771433, value=NA
 Miyakojima                   column=year:1969, timestamp=1732677771433, value=NA
 Miyakojima                   column=year:1970, timestamp=1732677771433, value=NA
 Miyakojima                   column=year:1971, timestamp=1732677771433, value=NA
 Miyakojima                   column=year:1972, timestamp=1732677771433, value=1972-01-29
 Miyakojima                   column=year:1973, timestamp=1732677771433, value=1973-02-05
 Miyakojima                   column=year:1974, timestamp=1732677771433, value=1974-01-15
 Miyakojima                   column=year:1975, timestamp=1732677771433, value=1975-01-30
 Miyakojima                   column=year:1976, timestamp=1732677771433, value=1976-01-20
 Miyakojima                   column=year:1977, timestamp=1732677771433, value=1977-01-17
 Miyakojima                   column=year:1978, timestamp=1732677771433, value=1978-02-01
 Miyakojima                   column=year:1979, timestamp=1732677771433, value=1979-01-26
 Miyakojima                   column=year:1980, timestamp=1732677771433, value=1980-02-03
 Miyakojima                   column=year:1981, timestamp=1732677771433, value=1981-01-28
 Miyakojima                   column=year:1982, timestamp=1732677771433, value=1982-01-20
 Miyakojima                   column=year:1983, timestamp=1732677771433, value=1983-01-17
```

```
cloudera@quickstart:~
File  Edit  View  Search  Terminal  Help
 Miyakojima                   column=year:2003, timestamp=1732677771433, value=2002-12-30
 Miyakojima                   column=year:2004, timestamp=1732677771433, value=2004-01-28
 Miyakojima                   column=year:2005, timestamp=1732677771433, value=2005-01-29
 Miyakojima                   column=year:2006, timestamp=1732677771433, value=2006-01-19
 Miyakojima                   column=year:2007, timestamp=1732677771433, value=2007-01-16
 Miyakojima                   column=year:2008, timestamp=1732677771433, value=2008-01-29
 Miyakojima                   column=year:2009, timestamp=1732677771433, value=2009-01-27
 Miyakojima                   column=year:2010, timestamp=1732677771433, value=2010-01-18
 Miyakojima                   column=year:2011, timestamp=1732677771433, value=2011-01-17
 Miyakojima                   column=year:2012, timestamp=1732677771433, value=2012-01-18
 Miyakojima                   column=year:2013, timestamp=1732677771433, value=2013-01-19
 Miyakojima                   column=year:2014, timestamp=1732677771433, value=2014-01-16
 Miyakojima                   column=year:2015, timestamp=1732677771433, value=2015-01-22
 Miyakojima                   column=year:2016, timestamp=1732677771433, value=2016-02-04
 Miyakojima                   column=year:2017, timestamp=1732677771433, value=2017-01-30
 Miyakojima                   column=year:2018, timestamp=1732677771433, value=2018-01-23
 Miyakojima                   column=year:2019, timestamp=1732677771433, value=2019-01-07
 Miyakojima                   column=year:2020, timestamp=1732677771433, value=2020-02-07
 Miyakojima                   column=year:2021, timestamp=1732677771433, value=2021-01-16
 Miyakojima                   column=year:2022, timestamp=1732677771433, value=2022-01-19
 Miyakojima                   column=year:2023, timestamp=1732677771433, value=2023-01-15
1 row(s) in 0.1210 seconds
```

In the 1970s, the first bloom dates start in late January. Since the 2000s, the first blooming in Miyakojima has typically begun in mid-January. Since Miyakojima is in Okinawa Prefecture, it experiences Japan's earliest cherry blossom season most likely due to its warm climate compared to other prefectures. The shift from late January to mid-January may subtly reflect the impact of global warming or changes in climate patterns which leads to warmer conditions earlier in the season.

8. Scan data with a specific filter for species Sargent cherry (Prunus sargentii)

```
hbase(main):012:0> scan 'sakura_first_bloom', {FILTER => "ValueFilter(=, 'binary:Sargent cherry (Prunus sargentii)')"}
ROW                           COLUMN+CELL
 Abashiri                     column=metadata:sakura_species, timestamp=1732677771433, value=Sargent cherry (Prunus sargen
                              tii)
 Asahikawa                    column=metadata:sakura_species, timestamp=1732677771433, value=Sargent cherry (Prunus sargen
                              tii)
 Hiroo                        column=metadata:sakura_species, timestamp=1732677771433, value=Sargent cherry (Prunus sargen
                              tii)
 Iwamizawa                    column=metadata:sakura_species, timestamp=1732677771433, value=Sargent cherry (Prunus sargen
                              tii)
 Kushiro                      column=metadata:sakura_species, timestamp=1732677771433, value=Sargent cherry (Prunus sargen
                              tii)
 Monbetsu                     column=metadata:sakura_species, timestamp=1732677771433, value=Sargent cherry (Prunus sargen
                              tii)
 Obihiro                      column=metadata:sakura_species, timestamp=1732677771433, value=Sargent cherry (Prunus sargen
                              tii)
 Rumoi                        column=metadata:sakura_species, timestamp=1732677771433, value=Sargent cherry (Prunus sargen
                              tii)
 Urakawa                      column=metadata:sakura_species, timestamp=1732677771433, value=Sargent cherry (Prunus sargen
                              tii)
 Wakkanai                     column=metadata:sakura_species, timestamp=1732677771433, value=Sargent cherry (Prunus sargen
                              tii)
10 row(s) in 0.0370 seconds
```

There are 10 sites with Sargent cherry (Prunus sargentii). All these sites are located on Hokkaido, Japan's northernmost island and prefecture. Hokkaido experience a cold climate with long harsh winters and relatively cool summers.

9. Retrieve the first and full bloom dates in Urakawa, a site with Sargent cherry (Prunus sargentii)

```
hbase(main):028:0> get 'sakura_first_bloom', 'Urakawa', 'year:1953'
COLUMN                          CELL
 year:1953                      timestamp=1732677771433, value=1953-05-09
1 row(s) in 0.0910 seconds

hbase(main):029:0> get 'sakura_first_bloom', 'Urakawa', 'year:1978'
COLUMN                          CELL
 year:1978                      timestamp=1732677771433, value=1978-05-16
1 row(s) in 0.0120 seconds

hbase(main):030:0> get 'sakura_first_bloom', 'Urakawa', 'year:2023'
COLUMN                          CELL
 year:2023                      timestamp=1732677771433, value=NA
1 row(s) in 0.0400 seconds
```

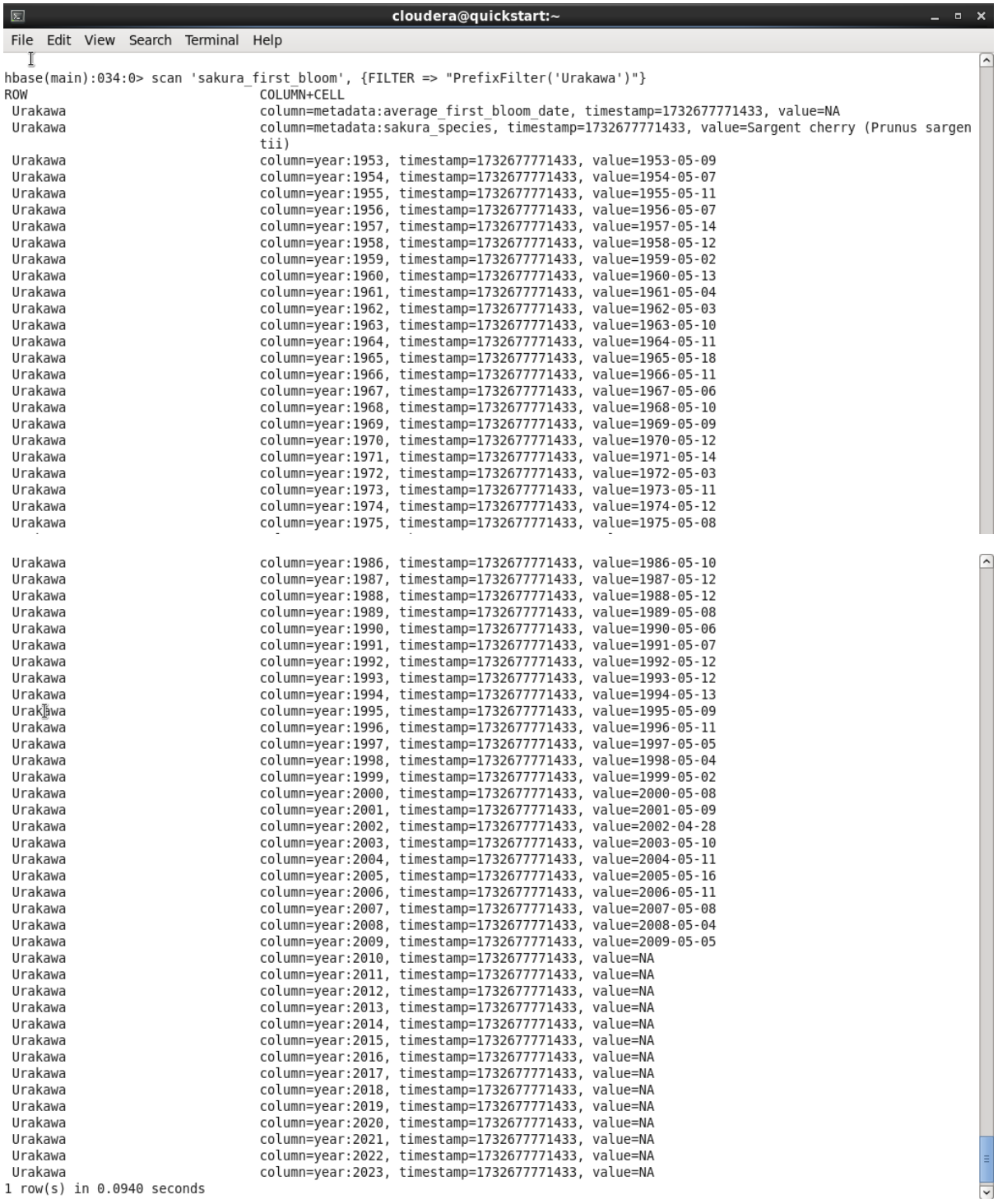
In 1953, the first bloom occurred on May 9th, but 25 years later, in 1978, it was delayed to May 16th.

```
hbase(main):031:0> get 'sakura_full_bloom', 'Urakawa', 'year:1953'
COLUMN                          CELL
 year:1953                      timestamp=1732677789420, value=1953-05-14
1 row(s) in 0.0380 seconds

hbase(main):032:0> get 'sakura_full_bloom', 'Urakawa', 'year:1978'
COLUMN                          CELL
 year:1978                      timestamp=1732677789420, value=1978-05-19
1 row(s) in 0.0150 seconds

hbase(main):033:0> get 'sakura_full_bloom', 'Urakawa', 'year:2023'
COLUMN                          CELL
 year:2023                      timestamp=1732677789420, value=NA
1 row(s) in 0.0100 seconds
```

In 1953, the full bloom occurred on May 14th, while 25 years later, in 1978, the first bloom was delayed until May 19th. The later bloom dates in 1978 compared to 1953 could indicate that the spring in 1978 was colder than in 1953 and may also be a byproduct of broader climate change effects.

The period from first bloom to full bloom in 1953 took 5 days, while in 1978 it took only 3 days. The shorter blooming period in 1978 suggests that temperatures may have risen more rapidly after the initial bloom potentially indicating the effects of climate change.

10. Analyse first bloom changes over time for Urakawa.

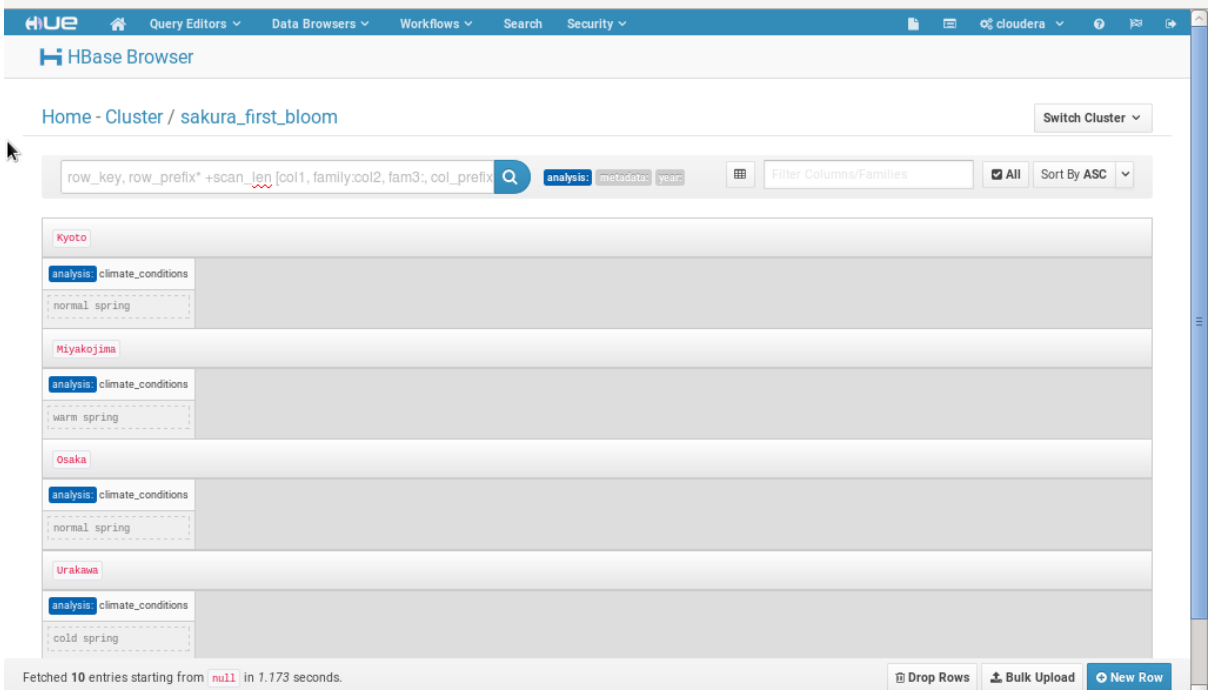


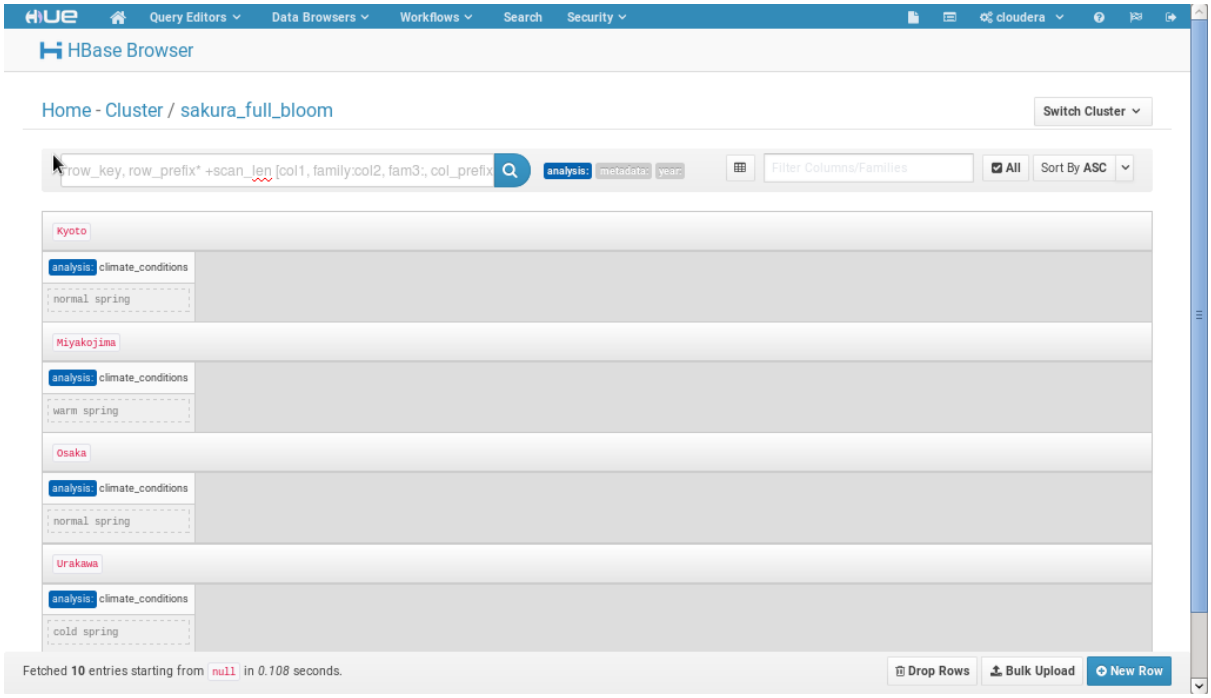
The first bloom dates in Urakawa in this 70-year period is typically in early to mid-May. Urakawa has a later cherry blossom season due to the cooler climate. In 2010, no first bloom was recorded. Since cherry blossom blooming is influenced by warm temperatures, this absence suggests that the temperature in Urakawa that year may not have been warm enough to trigger blooming. This anomaly could potentially indicate an effect of climate change as it might suggest colder spring conditions in the region.

11. Insert data into the new column

```
hbase(main):037:0> put 'sakura_first_bloom', 'Kyoto', 'analysis:climate_conditions', 'normal spring'
0 row(s) in 0.4440 seconds

hbase(main):038:0> put 'sakura_first_bloom', 'Osaka', 'analysis:climate_conditions', 'normal spring'
0 row(s) in 0.0530 seconds

hbase(main):039:0> put 'sakura_first_bloom', 'Urakawa', 'analysis:climate_conditions', 'cold spring'
0 row(s) in 0.0430 seconds

hbase(main):040:0> put 'sakura_first_bloom', 'Miyakojima', 'analysis:climate_conditions', 'warm spring'
0 row(s) in 0.0420 seconds

hbase(main):041:0> put 'sakura_full_bloom', 'Kyoto', 'analysis:climate_conditions', 'normal spring'
0 row(s) in 0.0860 seconds

hbase(main):042:0> put 'sakura_full_bloom', 'Osaka', 'analysis:climate_conditions', 'normal spring'
0 row(s) in 0.0290 seconds

hbase(main):043:0> put 'sakura_full_bloom', 'Urakawa', 'analysis:climate_conditions', 'cold spring'
0 row(s) in 0.0330 seconds

hbase(main):044:0> put 'sakura_full_bloom', 'Miyakojima', 'analysis:climate_conditions', 'warm spring'
0 row(s) in 0.0610 seconds
```

After analyzing the data for four sites—Kyoto, Osaka, Miyakojima, and Urakawa—a new column, climate_conditions, was created to record the spring conditions at each site. The data for this column was verified using Hue to ensure that it was successfully added.





SUMMARY

Global warming is a major driver of climate change, and its impact is evident in the shifting blooming dates of cherry blossoms. As the global temperatures rise, blooming dates have progressively become earlier while the blooming period has shortened. Japan's climate varies greatly across its regions because of its geographical diversity which affects how climate change influences blooming dates at different sites. Although the effects of climate change on blooming dates are subtle, they can disrupt the natural ecosystem by affecting plant cycles and the broader ecological balance.