

Phase 4 project – BIG DATA ANALYSIS

PROBLEM STATEMENT:

- Continue building the big data analysis solution by applying advanced Analysis techniques and visualizing the results.
- Apply more complex analysis techniques, such as machine learning Algorithms, time series analysis, or sentiment analysis, depending on the Dataset and objectives.
- Create visualizations to showcase the analysis results. Use tools like Matplotlib, Plotly, or IBM Watson Studio for creating graphs and charts.

SOLUTION:

Certainly, building a big data analysis solution that incorporates advanced Techniques and visualizations is essential for deriving meaningful insights from Your data. Let's continue with the process:

Step 1:

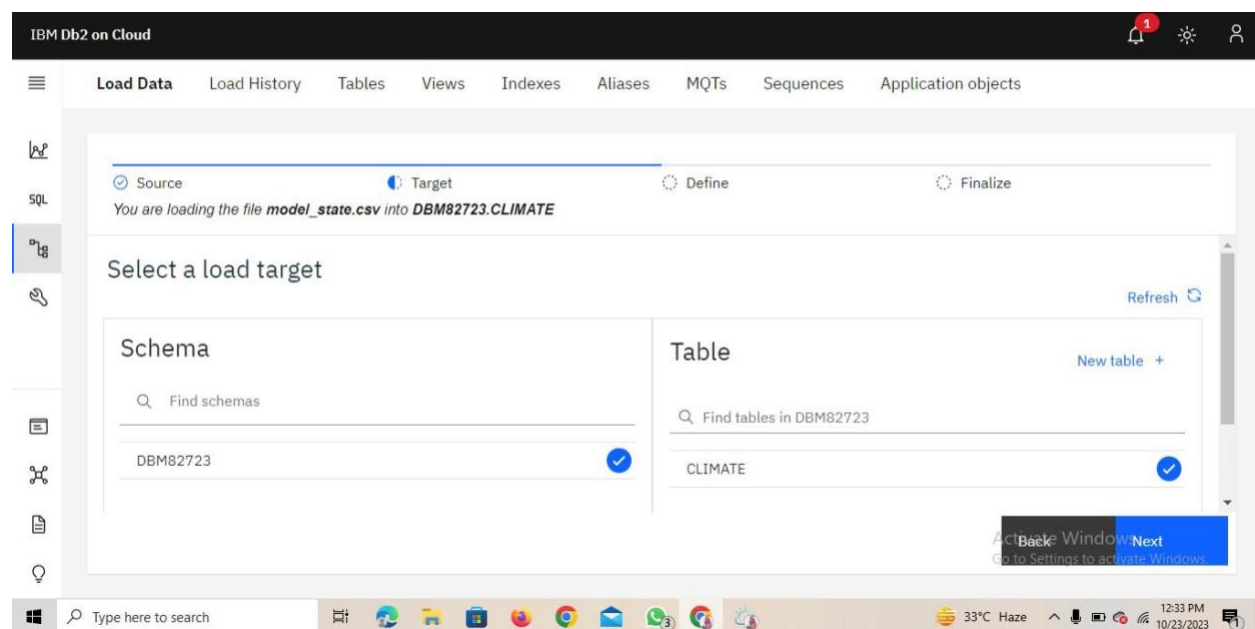
Download a CSV or xlsx file for upload in the DB2 database.

Example: open the wwv browser.

Search for the convenient topic to download database.(eg:kaggle,Data.world..)

Step 2:

Create a data table in IBM Cloud DB2 Database.



Step 3:

Upload the downloaded CSV. File in the database.

IBM Db2 on Cloud

Load Data | Load History | Tables | Views | Indexes | Aliases | MQTs | Sequences | Application objects

Source | Target | Define | Finalize

You are loading the file `model_state.csv` into `DBM82723.CLIMATE`

Code page (character encoding): 1208 (UTF-8) | Separator: , | Header in first row: ☒ | Time & date format: | Detect data types: ☐

	FIPS SMALLINT	FALL DECFLOAT	SPRING DECFLOAT	SUMMER DECFLOAT	WINTER DECFLOAT	MAX_WARMING_SEASON VARCHAR(6)	ANNUAL DECFLOAT
1	01	-0.19566843033509	-0.10586243386243	-0.32500881834215	0.458525573192233	Winter	-0.035047
2	04	1.203950617283951	1.384479717813051	1.274455026455033	1.388388007054677	Winter	1.319880
3	05	-0.04253968253968	0.266398589065250	0.058596119929444	0.532246913580247	Winter	0.214074
4	06	1.570920634920635	1.44924162257494E	1.478335097001771	1.412430335097001	Fall	1.480560
5	08	1.055308641975303	1.436910052910052	1.36784479717812E	1.838758377425037	Winter	1.438589
6	09	1.452003777777778	1.543707777777778	1.59067786506110E	2.622075208641073	Winter	1.901407

Back Window Next
Go to Settings to activate Windows

Step 4:

Finalize the uploading settings.

IBM Db2 on Cloud

Load Data | Load History | Tables | Views | Indexes | Aliases | MQTs | Sequences | Application objects

Source | Target | Define | Finalize

You are loading the file `model_state.csv` into `DBM82723.CLIMATE`

Review settings

Summary	
Code page:	1208 (Default)
Separator:	,
Time format:	HH:MM:SS (Default)
Date format:	YYYY-MM-DD (Default)

Option

Maximum number of warnings

1000

Back Activate Windows Begin Load
Go to Settings to activate Windows

Step 5:

Run the loaded data to check it is contain error or not.

The screenshot shows the 'Load Data' interface in IBM Db2 on Cloud. The 'Load details' section indicates the job is 'COMPLETE' with a status of 'My computer' and 'Target' 'model_state.csv' and 'DBM82723.CLIMATE'. A large blue donut chart shows the progress: 48 Rows read, 48 Rows loaded, and 0 Rows rejected. The text 'The data load job succeeded' is displayed. On the right, there are buttons for 'View Table' and 'Load More Data'. Below the chart, there are tabs for 'Status' and 'Settings'. To the right of the chart, there are tabs for 'Errors' (0) and 'Warnings' (0). A message 'No errors' is displayed with a note to 'Activate Windows'.

Step 6:

Create SQL queries to run the database table.

The screenshot shows the 'SQL' interface in IBM Db2 on Cloud. The 'Data objects' panel on the left shows a tree view with 'DBM82723' expanded, showing 'Tables', 'Views', 'MQTs', 'Aliases', and 'Nicknames'. The 'CLIMATE' table is selected. The 'Untitled - 1' editor shows the following SQL query:

```
1 SELECT STATE_NAME,max_warming_season
2 FROM CLIMATE
3 order by STATE_NAME;
```

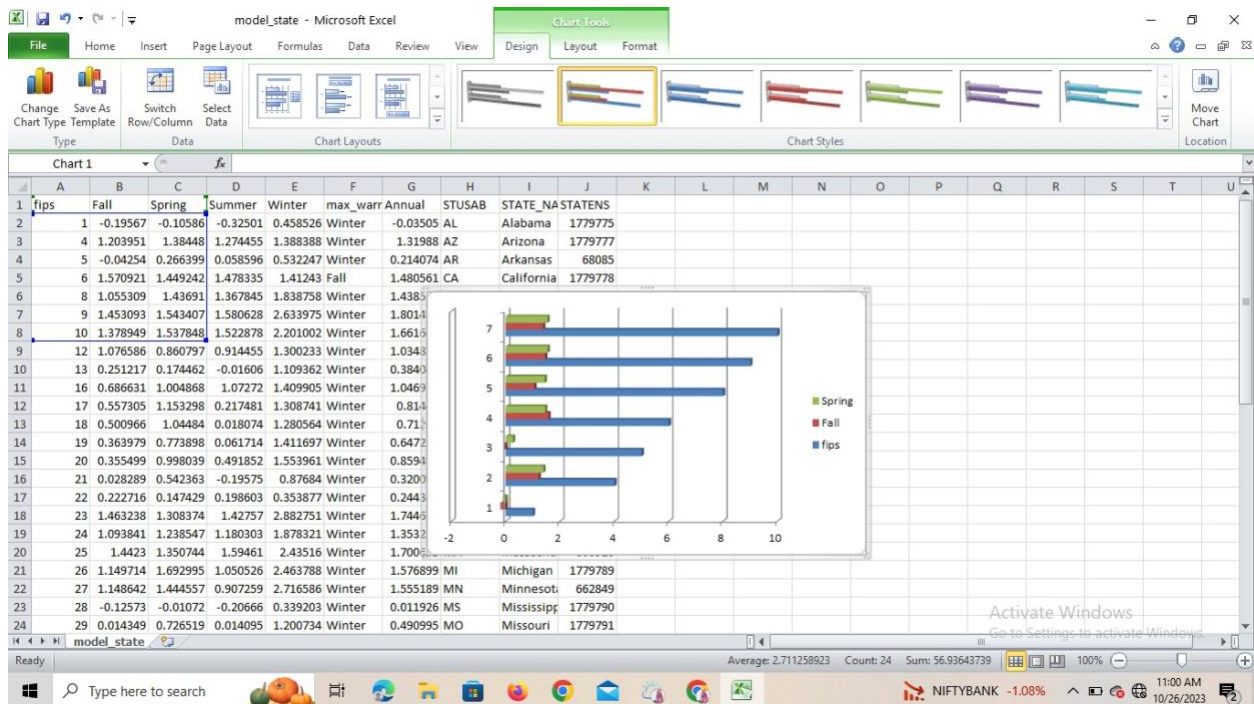
The 'Run all' button is visible. Below the editor, the 'History' tab shows a table of executed queries:

Script	Date	Status	Runtime
Untitled - 1	Oct 26, 2023 10:16:02 AM	✓ 1	0.006 s
SELECT STATE_NAME,max_warming_season FROM CLIMATE order b...		✓	0.006 s
Untitled - 1	Oct 26, 2023 10:15:39 AM	✗ 1	0.022 s

An 'Activate Windows' watermark is visible in the bottom right corner.

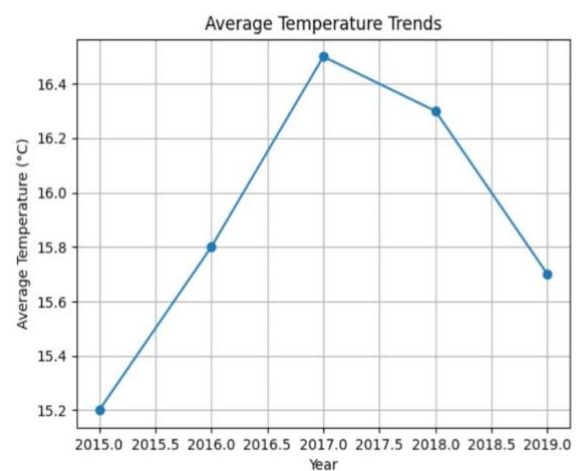
Step 7:

For development the analysis data we need to use the virtualization techniques in the datasets.



Step 8: Using python.

```
1 # Example Python code for creating a
  line chart using Matplotlib
2
3 import matplotlib.pyplot as plt
4
5 years = [2015, 2016, 2017, 2018, 2019]
6 avg_temperatures = [15.2, 15.8, 16.5,
  16.3, 15.7]
7 plt.plot(years, avg_temperatures,
  marker='o')
8 plt.title('Average Temperature Trends')
9 plt.xlabel('Year')
10 plt.ylabel('Average Temperature (°C)')
11 plt.grid(True)
12 plt.show()
```



Step 9:

Using Machine Learning techniques.

Select Appropriate Analysis Techniques:

Depending on the nature of your dataset and specific objectives, consider various

Advanced analysis techniques:

Machine Learning Algorithms: Use supervised or unsupervised machine learning

Algorithms like decision trees, random forests, support vector machines, or

Clustering algorithms for predictive modeling or pattern recognition.

Time Series Analysis: If your data involves time-based data points, use time Series analysis techniques to identify trends, seasonality, and forecast future Values.

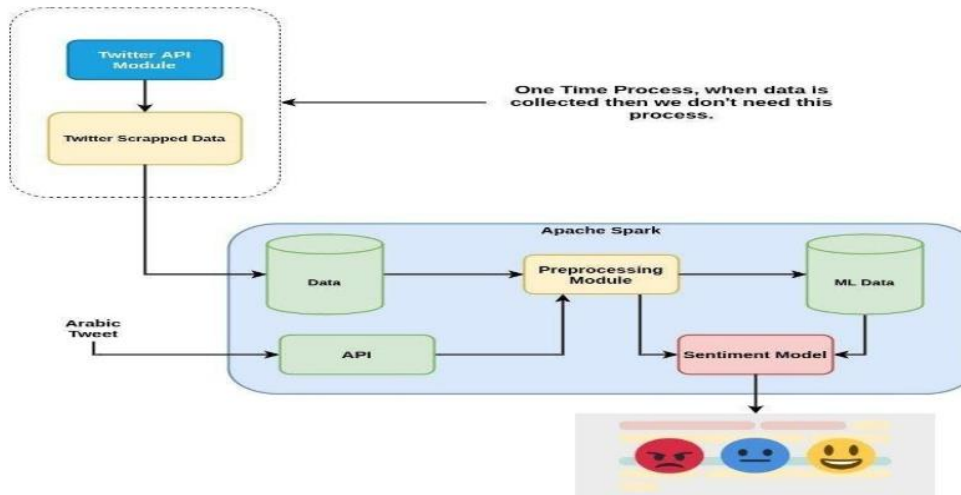
Sentiment Analysis: Apply natural language processing techniques to extract Sentiment from text data, useful for social media or customer reviews analysis.

Example:

```
# Example Python code for sentiment analysis using NLTK
import nltk

from nltk.sentiment import SentimentIntensityAnalyzer
nltk.download('vader_lexicon')

sia = SentimentIntensityAnalyzer()
text = "The weather is wonderful and the scenery is breathtaking."
sentiment_score = sia.polarity_scores(text)
print(sentiment_score)
```



Conclusion:

Thus the ,Continue building the big data analysis solution by applying advanced analysis techniques
And visualizing the results has been completed.