KGiSL INSTITUTE OF TECHNOLOGY

Coimbatore – 641035

Institution code: 7117

Big Data Analysis Using IBM Cloud Databases

MENTOR:

MRS.INDU POORNIMA.R

TEAM MEMBERS:

YOGANATHAN.R

SABAREESAN.R

VENKATESH.R

SIBIRAJ.M

Big Data Analysis Using IBM Cloud Databases

Problem Definition: The project involves delving into big data analysis using IBM Cloud Databases. The objective is to extract valuable insights from extensive datasets, ranging from climate trends to social patterns. The project includes designing the analysis process, setting up IBM Cloud Databases, performing data analysis, and visualizing the results for business intelligence.

DEVELOPMENT PART 2

Part 1: Import Libraries and Load Data :

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
data =
pd.read_csv('/content/drive/MyDrive/ColabNotebooks/GlobalLandTemperaturesByCity.csv')
```

Part 2: Convert 'dt' Column to DateTime :

```
# Convert the 'dt' column to a datetime format for time series analysis
data['dt'] = pd.to_datetime(data['dt'])
print("After converting 'dt' column to datetime:")
print(data['dt'].head())  # Display the first few rows of the 'dt' column to confirm the conversion
```

OUTPUT :

```
After converting 'dt' column to datetime:
0   1743-11-01
1   1743-12-01
2   1744-01-01
3   1744-02-01
4   1744-03-01
Name: dt, dtype: datetime64[ns]
```

Part 3: Set 'dt' Column as the Index for Time Series Analysis :

```
# Set the 'dt' column as the index for time series analysis
data.set_index('dt', inplace=True)
print("After setting 'dt' column as the index:")
print(data.head())
```

OUTPUT :

```
After setting 'dt' column as the index:
            AverageTemperature  AverageTemperatureUncertainty   City  Country  \
dt
1743-11-01               6.068                          1.737  Århus  Denmark
1743-12-01                 NaN                            NaN  Århus  Denmark
1744-01-01                 NaN                            NaN  Århus  Denmark
1744-02-01                 NaN                            NaN  Århus  Denmark
1744-03-01                 NaN                            NaN  Århus  Denmark

           Latitude Longitude
dt
1743-11-01   57.05N    10.33E
1743-12-01   57.05N    10.33E
1744-01-01   57.05N    10.33E
1744-02-01   57.05N    10.33E
1744-03-01   57.05N    10.33E
```

Part 4: Resample the Data to Monthly Averages :

```
# Resample the data to calculate monthly average temperatures
monthly_data = data['AverageTemperature'].resample('M').mean()
print("Monthly Average Temperatures:")
print(monthly_data.head())
```

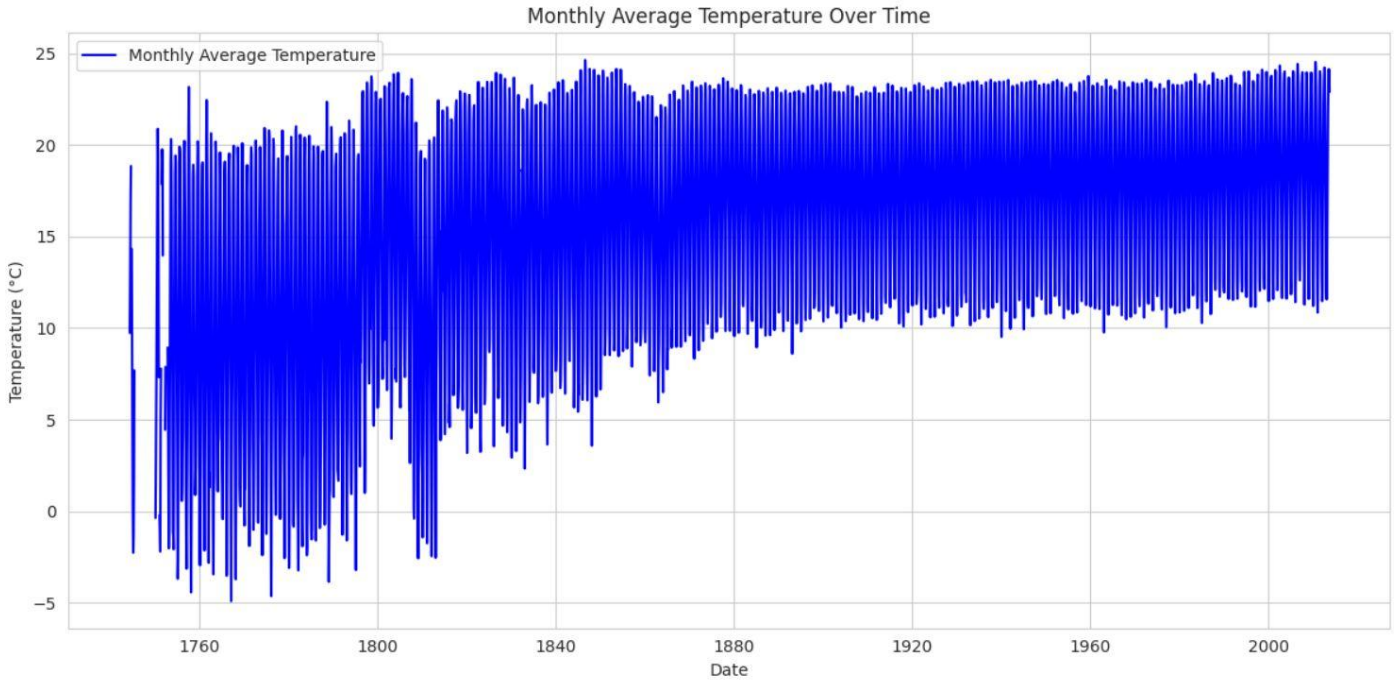
OUTPUT :

```
Monthly Average Temperatures:
dt
1743-11-30    4.882424
1743-12-31         NaN
1744-01-31         NaN
1744-02-29         NaN
1744-03-31         NaN
Freq: M, Name: AverageTemperature, dtype: float64
```

Part 5: Visualize the Time Series Data :

```
# Visualize the time series data
plt.figure(figsize=(12, 6))
sns.set_style('whitegrid')
plt.plot(monthly_data.index, monthly_data, label='Monthly Average Temperature', color='blue')
plt.title('Monthly Average Temperature Over Time')
plt.xlabel('Date')
plt.ylabel('Temperature (°C)')
plt.legend()
plt.tight_layout()
plt.show()
```

OUTPUT :


Monthly Average Temperature Over Time
Monthly Average Temperature
Temperature (°C)
25
20
15
10
5
0
−5
1760
1800
1840
1880
1920
1960
2000
Date