Exploring Weather Trends

Importing neccessary libraries for the project :

In [2]:
```
import pandas as pd              # for dealing with data!
import matplotlib.pyplot as plt # for visualizing the data!
import numpy as np               # for calculating the moving average!
```

Extracting data from database

Firstly, I used SQL to download the data from the database. From the database I extracted data of global temperature and local temperature of 'Ahmadbad' and 'New Delhi' city using below mentioned queries:

select *from global_data*

select from city_data where city = 'Ahmadabad'

select * from city_data where city = 'New Delhi'

In [4]:
```
global_temp = pd.read_csv('global_data.csv') # importing 'global tempreature d
ata'
ahm_temp = pd.read_csv('Ahmadabad.csv')     # importing 'Ahmadabad tempreature
data' which is a data for one city over multiple years
delhi_temp = pd.read_csv('New Delhi.csv')   # importing 'New Delhi tempreature
data' which is a data for one city over multiple years
```

In [5]:
```
global_temp.head()  # Looking at some rows of the data
```

Out[5]:

	year	avg_temp
0	1750	8.72
1	1751	7.98
2	1752	5.78
3	1753	8.39
4	1754	8.47

In [6]:
```
ahm_temp.head() # Looking at some rows of Ahmadbad city temparature data
```

Out[6]:

	year	city	country	avg_temp
0	1796	Ahmadabad	India	26.35
1	1797	Ahmadabad	India	27.45
2	1798	Ahmadabad	India	25.82
3	1799	Ahmadabad	India	26.62
4	1800	Ahmadabad	India	26.56

In [7]:
```
delhi_temp.head() # Looking at some rows of New Delhi city temparature data
```

Out[7]:

	year	city	country	avg_temp
0	1796	New Delhi	India	25.03
1	1797	New Delhi	India	26.71
2	1798	New Delhi	India	24.29
3	1799	New Delhi	India	25.28
4	1800	New Delhi	India	25.21

In [9]:
```
delhi_temp.isna().sum() # To check any missing values in New Delhi city temperature dataset
```

Out[9]:
```
year         0
city         0
country      0
avg_temp    17
dtype: int64
```

In [8]:
```
ahm_temp.isna().sum() # To check any missing values in Ahmadabad city temperature dataset
```

Out[8]:
```
year         0
city         0
country      0
avg_temp    12
dtype: int64
```

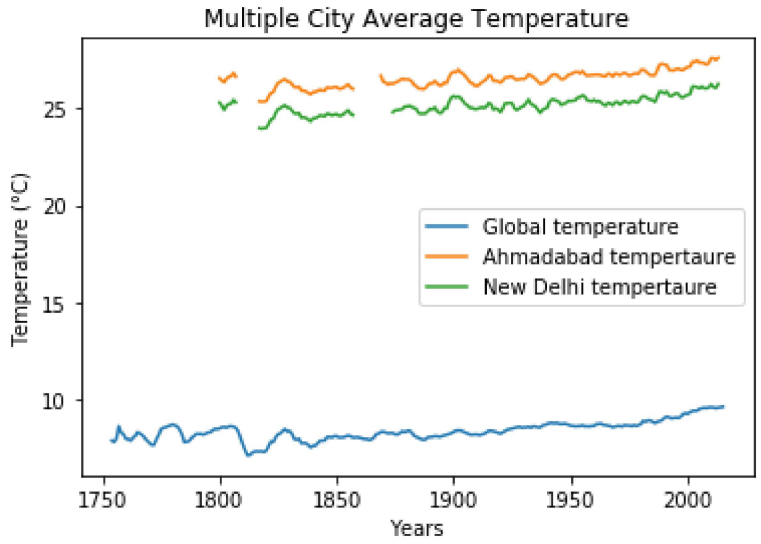
Moving Average for the Global data and Local data.

- Moving averages are used to smooth out data and to make it easier to observe long term trends and not get lost in short term fluctuations.

In [10]:
```
global_mv_avg = global_temp['avg_temp'].rolling(5).mean()
ahm_mv_avg = ahm_temp['avg_temp'].rolling(5).mean()
delhi_mv_avg = delhi_temp['avg_temp'].rolling(5).mean()
```

Visualizng Data using matplotlib Library :

In [12]:
```
#Local Data is as same as Ahmadabad
plt.plot(global_temp['year'],global_mv_avg,label='Global temperature')
plt.plot(ahm_temp['year'],ahm_mv_avg,label='Ahmadabad tempertaure')
plt.plot(delhi_temp['year'],delhi_mv_avg,label='New Delhi tempertaure')
plt.legend()
plt.xlabel("Years")
plt.ylabel("Temperature (°C)")
plt.title("Multiple City Average Temperature")
plt.show()
```



Key observations:

- In ahmadabad city and New Delhi temperature dataset, some missing values are present which can be clearly seen in the plot and same is shown in the notebook.
- Ahmadabad city and New Delhi temperature follows the same trend as the Global temperature.
- The Global temperature is less as compared to Ahmadabad and New Delhi temperature and this trend can be seen throuout.
- In the later part of the plot there are some fluctuations in ahmadabad city as well as new delhi temperature as compared to Global temperature because the Moving average is considered for 5 years window and if it is increased then plot will be smooth comparatively.
- From the plot above it can be seen that local city temperature and Global temperature which are positively corelated that means they should have positive corelation coefficient. Corelation coeffiecient indicates how strongly both variables are corelated to each other. If two variables are positively corelated then if the value of one variable increased then other varible value will also be increased.