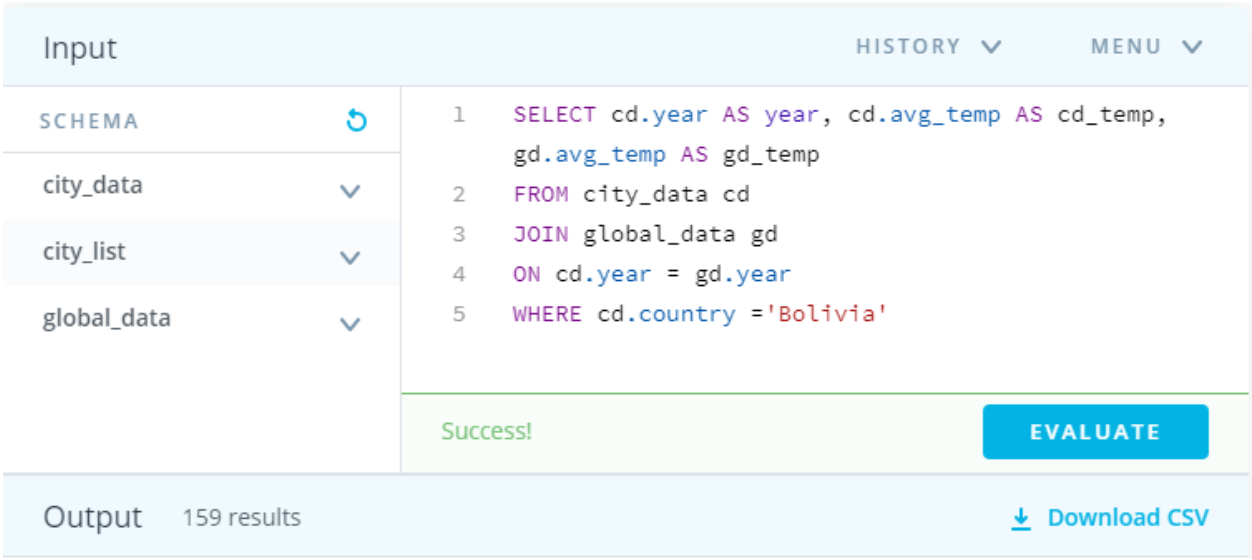
WEATHER TRENDS

The first step was to extract the data from the database Udacity provided. I extracted the global data temperature and La Paz city temperature using the following code:



Originally, global data had a size of 266 rows including data from 1750 until 2015; however, La Paz data had only 159 rows including data from 1855 until 2013, showing that I had less data from my city. Applying JOIN statement allowed me to extract data for years that existed in both tables.

After downloading the CSV file, I used Python to open it and calculate the moving averages. I tried some different sizes for moving averages, in the range of 5 to 25 and I finally decided to use a moving average of size 10.

I used pandas library to open CSV file and also calculate correlation between the two datasets, after that I used numpy arrays to hold the values and calculate moving average. Finally I used matplotlib library to plot the values. My code is:

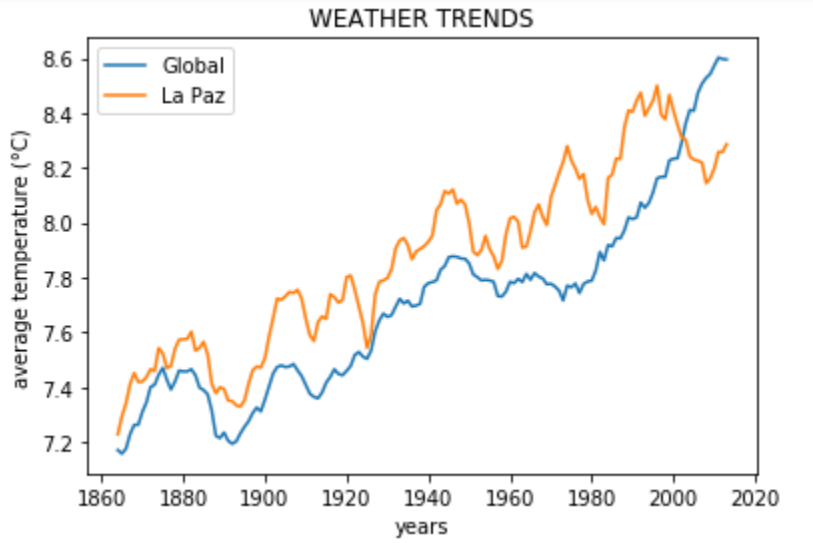
```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
```

```
#opening the file
weather_data= pd.read_csv('weather.csv', encoding='utf-8')
print(weather_data[['cd_temp','gd_temp']].corr())
weather= weather_data.values


#moving average size 10
size_avg=9
g_avg=np.zeros((len(weather)-size_avg))
c_avg=np.zeros((len(weather)-size_avg))
for i in range(len(weather)-size_avg):
    g_avg[i]=np.sum(weather[i:i+size_avg,2])/10
    c_avg[i]=np.sum(weather[i:i+size_avg,1])/10
years_avg=weather[size_avg:,0]

#plot the graph
%matplotlib inline
plt.plot(years_avg,g_avg,years_avg,c_avg)
plt.legend(('Global','La Paz'))
plt.title('WEATHER TRENDS')
plt.xlabel('years')
plt.ylabel('average temperature (°C)')
```

After running the code, I got the following figure:



And also the following correlation matrix:

	cd_temp	gd_temp
cd_temp	1.000000	0.720829
gd_temp	0.720829	1.000000

OBSERVATIONS:

- Temperature increases with time in both datasets, starting around 7.4 °C and increasing to above 8 °C, showing a positive correlation between them.
- The way data increases seems to be roughly lineal.
- The temperature in La Paz city also seems to have a roughly sinusoidal element besides lineal one as there seem to be cycles where the temperature increases more and then decreases again.
- La Paz city temperature is almost all the time higher than average global temperature.
- Since 1980 global temperature seems to be increasing faster than before and there is a steep slope.
- The correlation matrix show a high positive correlation between global data and La Paz city data as the result is 0.72.