Exploring Weather Trends

Author: Felipe Millacura

Your submission should be a PDF that includes:

```
An outline of steps taken to prepare the data to be visualized in the chart, such as:
    What tools did you use for each step? (Python, SQL, Excel, etc)
    How did you calculate the moving average?
    What were your key considerations when deciding how to visualize the trends?
Line chart with local and global temperature trends
At least four observations about the similarities and/or differences in the trends
```

Extracting the SQL Data

In order to extract global data

```
SELECT *
FROM global_data
```

In order to extract city data from Edinburgh, London, Santiago and California

```
SELECT year, avg_temp, city, country
FROM city_data
WHERE country = 'United Kingdom'
OR country = 'Chile'
OR country = 'Japan'
OR country = 'Australia'
OR country = 'South Africa'
```

Analysing in Python

In [150]:

```
# Imports libraries

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

In [40]:

```
# Imports data

global_data = pd.read_csv('./global_data.csv')

cities_data = pd.read_csv('./five_countries.csv')
```

In [87]:

```
#Calculates rolling windows

global_data['MA7'] = global_data['avg_temp'].rolling(7, min_periods=7).mean()

global_data['MA14'] = global_data['avg_temp'].rolling(14, min_periods=14).mean()

global_data.set_index('year')
```

Out[87]:

year	avg_temp	MA7	MA14
1750	8.72	NaN	NaN
1751	7.98	NaN	NaN
1752	5.78	NaN	NaN
1753	8.39	NaN	NaN
1754	8.47	NaN	NaN
...	...	...	...
2011	9.52	9.588571	9.497143
2012	9.51	9.561429	9.496429
2013	9.61	9.572857	9.519286
2014	9.57	9.550000	9.545714
2015	9.83	9.607143	9.575714

266 rows × 3 columns

In [104]:

```
# Pivots table

cities_data_pv = cities_data.pivot_table(index='year', columns='city', values='avg_temp', aggfunc='mean')

year = cities_data_pv.reset_index()[['year']]
```

In [122]:

```
# Replaces data with 7 day moving average

cities_MA7 = cities_data_pv.reset_index().drop('year', 1).rolling(7, min_periods=7).mean()

# Replaces data with 14 day moving average

cities_MA14 = cities_data_pv.reset_index().drop('year', 1).rolling(14, min_periods=14).mean()
```

In [131]:

```
# Adds back global, year and drops NaN values into MA7 dataset

cities_MA7['year'] = year

cities_MA7['Global'] = global_data['MA7']

cities_MA7 = cities_MA7.dropna()

cities_MA7
```

Out[131]:

city	Adelaide	Belfast	Birmingham	Brisbane	Canberra	Cardiff	Durban	Edinburgh	Hiroshima	Johannesburg	...	Melbourne	Na
135	15.940000	8.290000	8.634286	19.334286	11.444286	8.634286	19.914286	7.200000	14.800000	14.677143	...	13.217143	14.56
136	15.991429	8.238571	8.557143	19.204286	11.358571	8.557143	19.950000	7.161429	14.788571	14.711429	...	13.245714	14.52
137	15.975714	8.320000	8.618571	19.192857	11.337143	8.618571	19.872857	7.267143	14.610000	14.632857	...	13.212857	14.33
138	15.898571	8.190000	8.504286	19.088571	11.271429	8.504286	19.842857	7.151429	14.538571	14.620000	...	13.165714	14.25
139	15.912857	8.285714	8.670000	19.155714	11.297143	8.670000	19.900000	7.272857	14.457143	14.692857	...	13.167143	14.16
...	...	...	...	...	...	...	...	...	...	...	...	...	
261	16.814286	9.865714	10.258571	20.247143	12.387143	10.258571	21.035714	8.848571	16.061429	15.844286	...	14.182857	15.67
262	16.922857	9.838571	10.204286	20.285714	12.467143	10.204286	20.984286	8.831429	16.038571	15.781429	...	14.264286	15.66
263	16.945714	9.622857	9.997143	20.340000	12.487143	9.997143	21.045714	8.614286	16.102857	15.748571	...	14.298571	15.74
264	16.982857	9.617143	10.048571	20.262857	12.460000	10.048571	20.990000	8.618571	15.981429	15.748571	...	14.362857	15.61
265	16.970000	9.527143	9.950000	20.172857	12.361429	9.950000	21.004286	8.508571	15.950000	15.691429	...	14.347143	15.59

131 rows × 25 columns

In [132]:

```
# Adds back global, year and drops NaN values into MA14 dataset

cities_MA14['year'] = year

cities_MA14['Global'] = global_data['MA14']

cities_MA14 = cities_MA14.dropna()

cities_MA14
```

Out[132]:

city	Adelaide	Belfast	Birmingham	Brisbane	Canberra	Cardiff	Durban	Edinburgh	Hiroshima	Johannesburg	...	Melbourne	Na
142	15.954286	8.272143	8.547143	19.265714	11.355714	8.547143	19.962143	7.201429	14.619286	14.707857	...	13.230000	14.33
143	15.996429	8.265000	8.500714	19.234286	11.342857	8.500714	20.002857	7.197857	14.685000	14.723571	...	13.257143	14.41
144	15.970714	8.257857	8.455000	19.182857	11.337143	8.455000	19.991429	7.193571	14.673571	14.709286	...	13.257857	14.40
145	15.910000	8.169286	8.378571	19.142857	11.286429	8.378571	19.995000	7.106429	14.658571	14.704286	...	13.228571	14.38
146	15.902857	8.315714	8.555714	19.158571	11.305714	8.555714	19.988571	7.265714	14.604286	14.712143	...	13.243571	14.33
...	...	...	...	...	...	...	...	...	...	...	...	...	
261	16.657143	9.664286	10.070000	20.181429	12.175714	10.070000	20.992143	8.638571	15.933571	15.735714	...	13.979286	15.59
262	16.746429	9.669286	10.050714	20.219286	12.281429	10.050714	20.990714	8.649286	15.982143	15.712143	...	14.087143	15.64
263	16.789286	9.640714	10.045714	20.238571	12.358571	10.045714	21.065714	8.629286	16.047857	15.775000	...	14.171429	15.71
264	16.819286	9.632143	10.078571	20.218571	12.360714	10.078571	21.076429	8.635000	16.044286	15.782857	...	14.190000	15.72
265	16.859286	9.609286	10.046429	20.180000	12.312857	10.046429	21.077857	8.600714	15.959286	15.767143	...	14.237143	15.63

124 rows × 25 columns

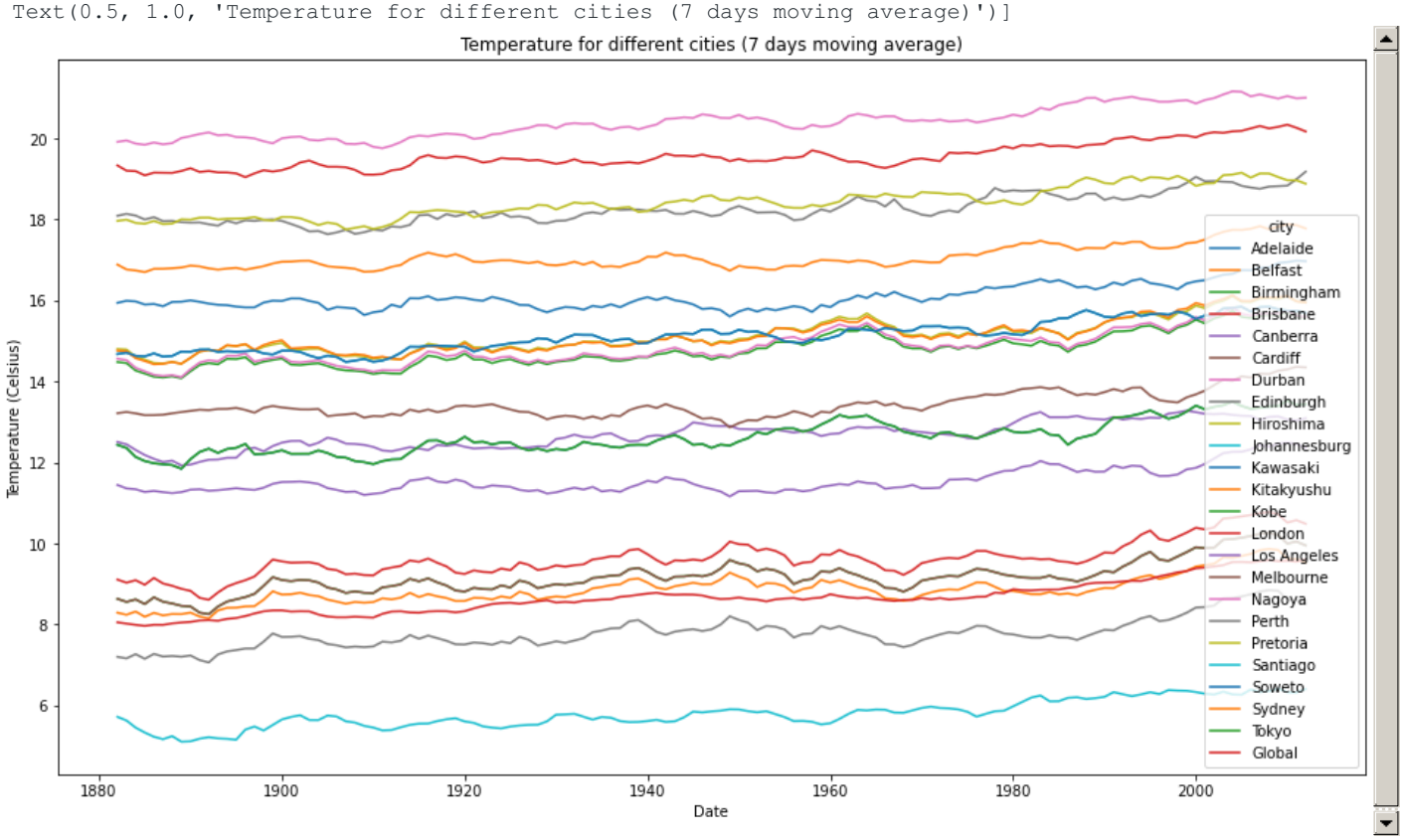
In [148]:

```
# generates plot for MA7

cities_MA7.set_index('year').plot(figsize=(16, 9)
      ).set(xlabel="Date",
            ylabel="Temperature (Celsius)",
            title="Temperature for different cities (7 days moving average)")
```

Out[148]:

```
[Text(0.5, 0, 'Date'),
 Text(0, 0.5, 'Temperature (Celsius)'),
 Text(0.5, 1.0, 'Temperature for different cities (7 days moving average)')]
```



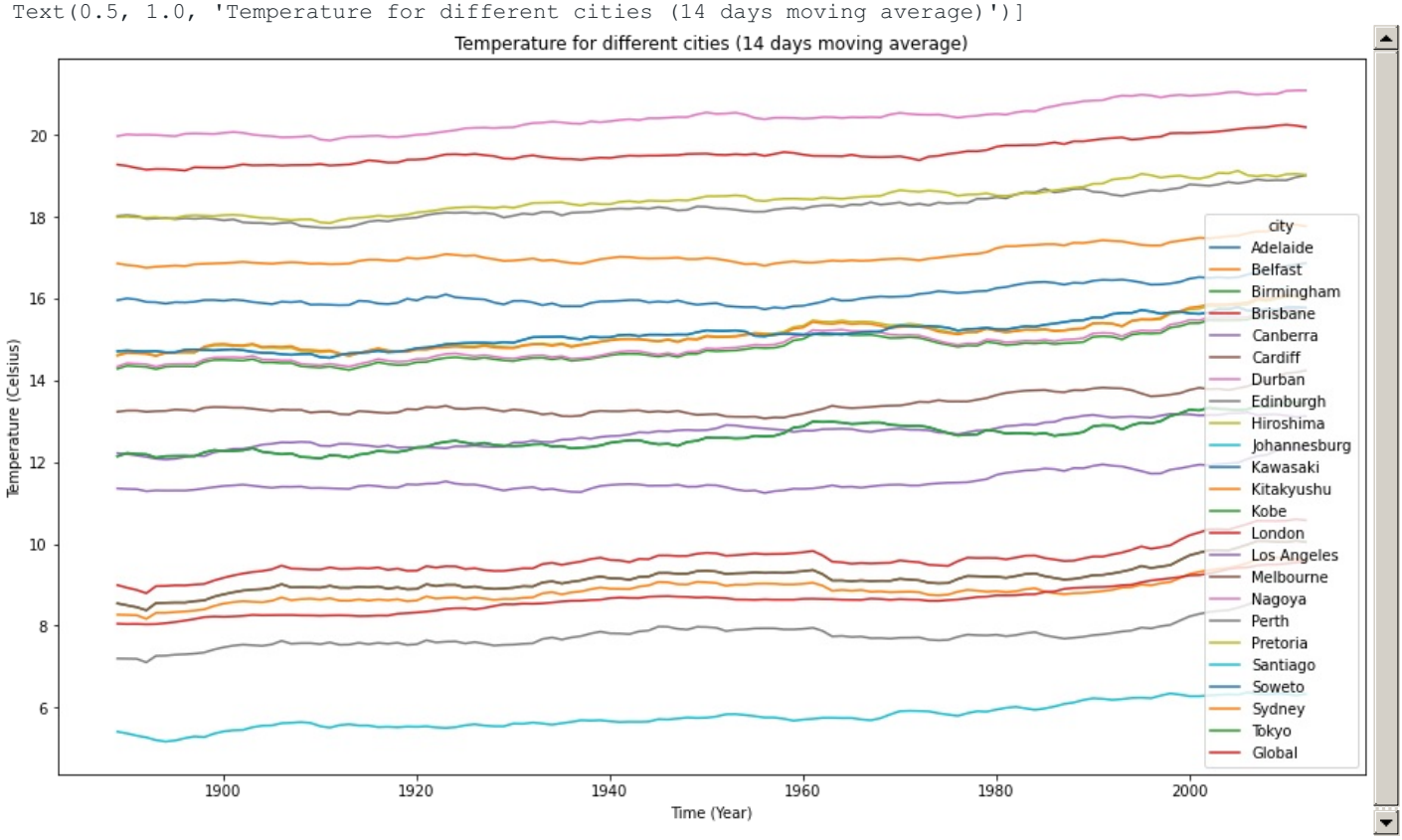
In [142]:

```
# generates plot for MA14

cities_MA14.set_index('year').plot(figsize=(16, 9)
      ).set(xlabel="Time (Year)",
            ylabel="Temperature (Celsius)",
            title="Temperature for different cities (14 days moving average)")
```

Out[142]:

```
[Text(0.5, 0, 'Time (Year)'),
 Text(0, 0.5, 'Temperature (Celsius)'),
 Text(0.5, 1.0, 'Temperature for different cities (14 days moving average)')]
```



In [ ]:

```
cities_MA7.max()
```

In [155]:

```
cities_MA7.min()
```

Out[155]:

```
city
Adelaide          15.602857
Belfast            8.148571
Birmingham         8.252857
Brisbane          19.042857
Canberra          11.160000
Cardiff            8.252857
Durban            19.758571
Edinburgh          7.061429
Hiroshima         14.438571
Johannesburg      14.481429
Kawasaki          11.844286
Kitakyushu        14.425714
Kobe              14.080000
London             8.610000
Los Angeles       11.927143
Melbourne         12.865714
Nagoya            14.105714
Perth             17.635714
Pretoria          17.752857
Santiago           5.105714
Soweto            14.481429
Sydney            16.698571
Tokyo             11.844286
year            1882.000000
Global             7.964286
dtype: float64
```

Observations about the similarities and differences between the world averages and your city's averages, as well as overall trends. Here are some questions to get you started.

Is your city hotter or cooler on average compared to the global average? Has the difference been consistent over time?

Santiago is the city showing the lowest temperature of all, fluctuating below the 10 degrees Celsius

How do the changes in your city's temperatures over time compare to the changes in the global average?

Global temperature fluctuates aroind 8 and 9 degrees, but Santiago is still below that range showing values of up to 2 degrees lower

What does the overall trend look like? Is the world getting hotter or cooler?

From this analysis we can clearly observe that global temperature is going up, and that on each country an uptrend can be seen

What is the hottest city?

From all five countries analysed the hottest city is present in South Africa, and corresponds to Durban

What are the second best?

The second coolest city is Edinburgh in United Kingdom and the second hotest is Brisbane, Australia

In general there is no major differences between MA7 and MA14 analysis