EXPLORE WEATHER TRENDS

Introduction:

My name is Abas Omar.

In this project we are tasked with exploring global weather trends and comparing it with the major city we live nearest to which in my case is Oslo.

I will use SQL to extract the data in CSV from global_data and city_data.
Use Jupyter notebook for Python and visualization.
Use Pandas to create DataFrame from the dataset, and find moving averages.
Using google docs I will then summarize what I found, including plots and code.

Extracting Data with SQL

We extract oslo_data from city_data, then global_data and finally join tables.
I found this new table useful at first as it allowed me to plot and compare with pandas.plot, but as you lose some years in the inner-join operation, I decided I would separate tables for the final results. Commenting out the parts from the query I didn't want.

```
WITH oslo_data AS(
        SELECT c.year, c.avg_temp oslo_avg
        FROM city_data c
        WHERE  c.city LIKE '%Oslo%'  AND c.avg_temp IS NOT NULL)

SELECT g.year, avg_temp global_avg, oslo_avg
FROM global_data g
JOIN oslo_data o
ON o.year = g.year
```

I would also try to extract the cumulative sum of the average temperature, but it was too volatile without moving averages.

```
WITH oslo_data AS (
        SELECT c.year, c.avg_temp oslo_avg,
                CASE WHEN (avg_temp - LAG(avg_temp) OVER () ) > 0 THEN 1 ELSE -1
                END AS temp_diff

        FROM city_data c
        WHERE  c.city LIKE '%Oslo%' AND c.avg_temp IS NOT NULL),
        g_data AS (
        SELECT g.year, g.avg_temp,
                CASE WHEN (avg_temp - LAG(avg_temp) OVER () ) > 0 THEN 1 ELSE -1 END
                AS temp_diff

        FROM global_data g
        WHERE g.avg_temp IS NOT NULL)

SELECT year, avg_temp, SUM(temp_diff) OVER (ORDER BY year) AS running_temp

FROM g_data g
```

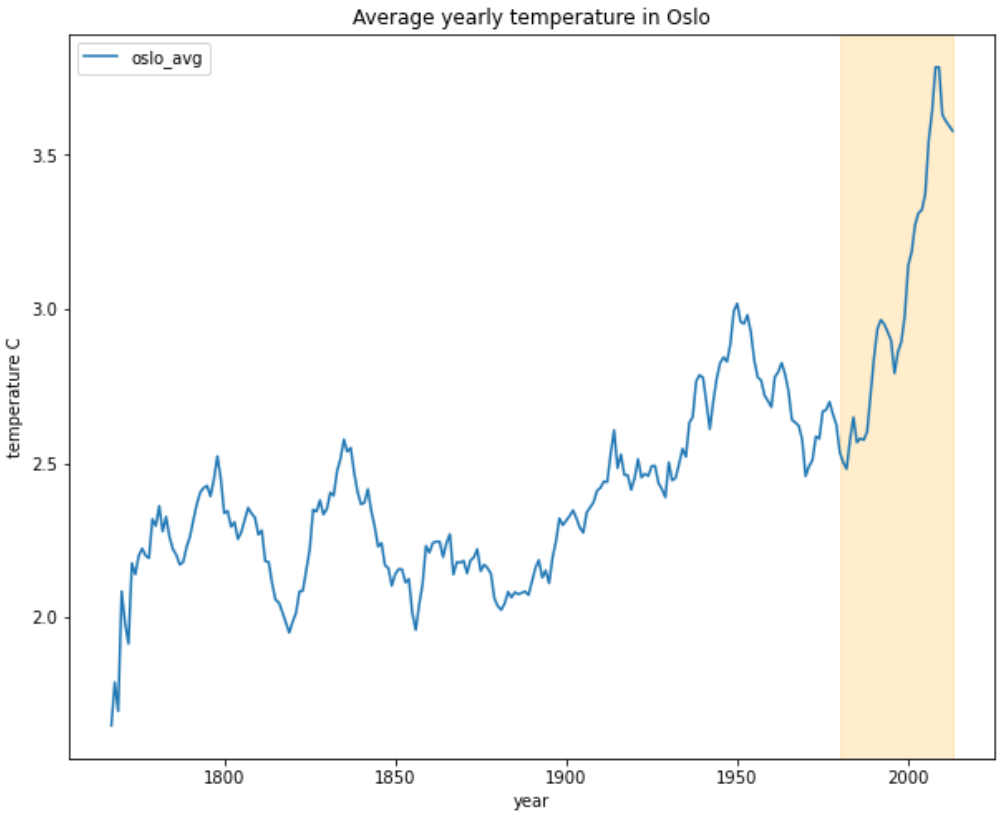
Finding moving average in Jupyter (with Python, Pandas, Numpy)

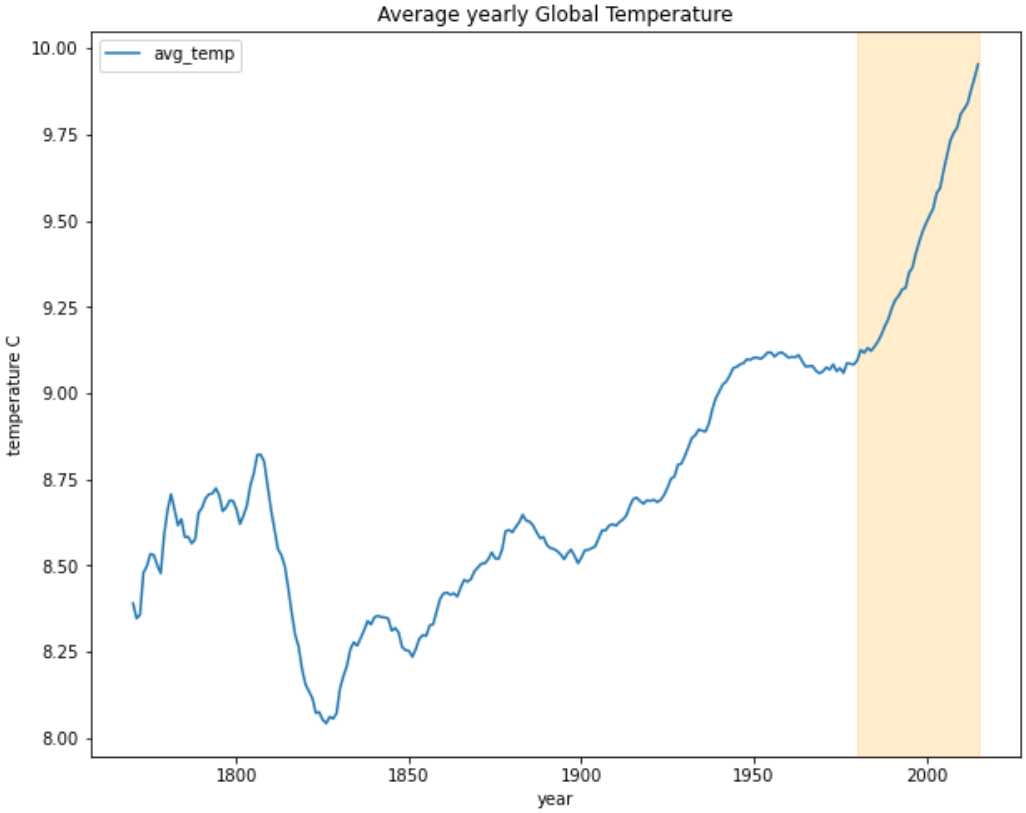
The code is included in the sources page below.
We import the pandas, numpy and matplotlib libraries.
Read the data into DataFrame using pd.read_csv function and start with finding the moving average.

We start with selecting a time delta, I tested different values and settled on 20 years. You lose the first 19 data points but given the holes in the data and how similar it looks at the start using different dt, it doesn't have a significant effect on the trends. Had it been the other way around (losing the last 19 data points) it would have had a major effect.

We use the movingAvg function in the Jupyter notebook to calculate the moving average, and drop the rows with Null values.

We plot the values we find in the figures below.


Average yearly temperature in Oslo
oslo_avg
temperature C
year
3.5
3.0
2.5
2.0
1800
1850
1900
1950
2000


Average yearly Global Temperature
avg_temp
temperature C
year
10.00
9.75
9.50
9.25
9.00
8.75
8.50
8.25
8.00
1800
1850
1900
1950
2000

Interpreting the figures

The temperature in my city varies a lot, over the seasons as well as year to year. Being in the northern part of the north we experience cold winters, warm winters, cool summers and hot summers.

This is given, as we are comparing global temperature to local one. A drought taking place in one region doesn't mean that the rest of the world is experiencing a drop in rainfall. Same for floods in one region of the world not being indicative of global increase in rainfall.

That being said, we have a clear trend from the earliest data points that while the temperature drops for years, eventually it recovers and increases.
The same can't be said for the increases that we see from the 1850s to this day.
The trend, which is more obvious in the global temperature, is upwards.

In Oslo, while it's more difficult, you can draw an upwards line from the 1850s to 1950s that shows an increase of approximately 1 degree celsius.

In the figures we highlight the period between 1980 and the most recent data-points. And we see sharp increase from the 1980s in the global temperature, and 1970s in the Oslo temperature which falls again only to pick up in the 80s again matching the global trend.

A nearly continuous increase in temperature that takes the global moving average in 1980 from 9.23 to 10.17 in 2015. And 2.58 in 1980 to 3.81 in 2013 in Oslo.

In Oslo between 1970s and mid-2010s the average increase of temperature was 0.0670. And the average decrease was -0.0476. The temperature increased 27 times and decreased 17 times.

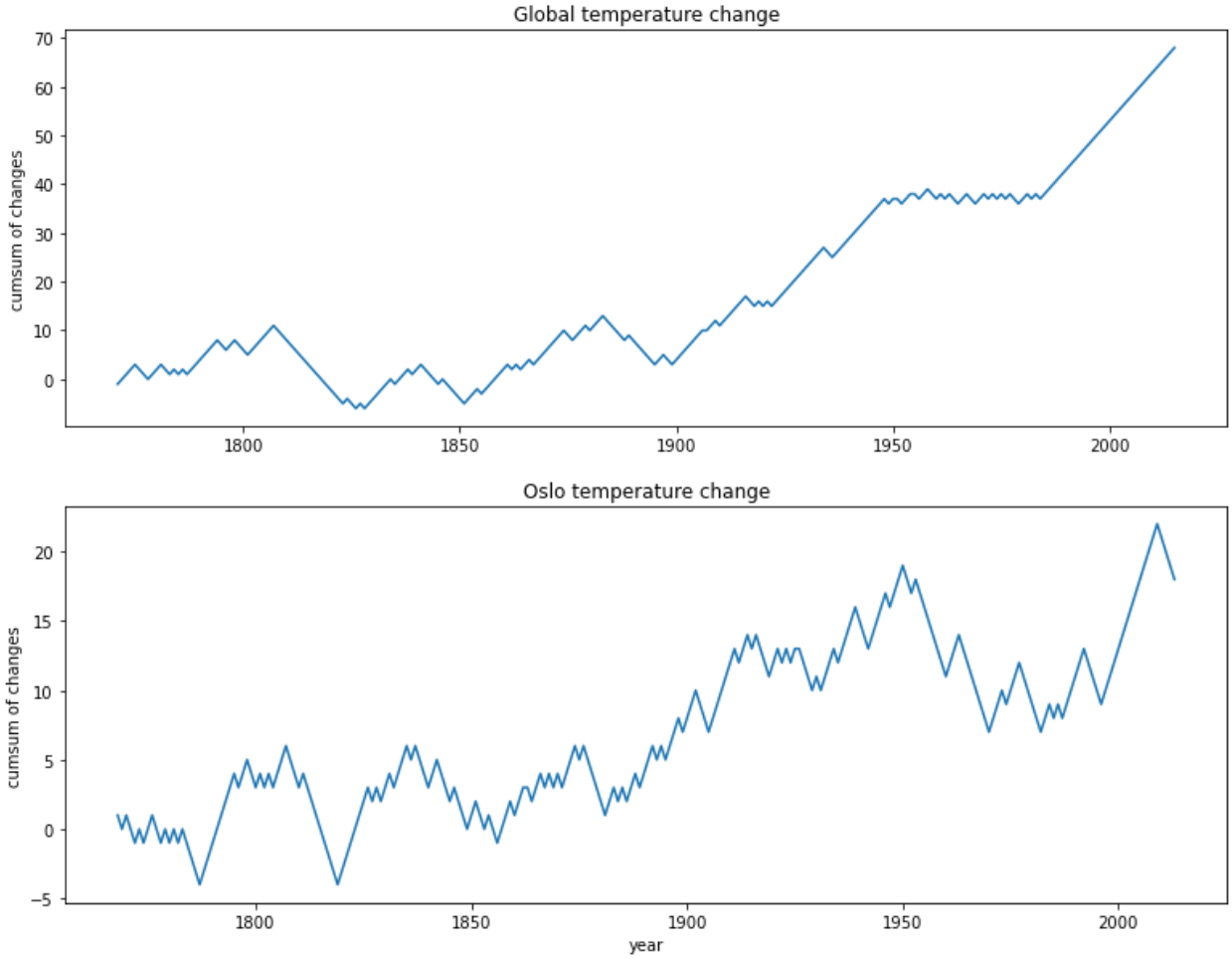
In the world for the same period the numbers are, 0.0245 increase and -0.0087 decrease. Increasing 39 times and decreasing only 7 times.

In comparison, if we look at the first 100 years of the moving average data points, we see the complete opposite. The differences are much more narrow and the changes almost even.

Looking at the numbers for the first 100 years in global temperature we get:
Average change of [0.0275, -0.0287]. With 53 increases and 47 decreases.

For Oslo the average temperature changes are: [0.0557, -0.0490] and number of changes 53 increases, 47 decreases.

With this I wanted to visualize this and make the differences clear.
In the figures below I take the cumulative sum of the moving averages where decreases are -1 and increases are 1. And we see as expected something that is similar to the figures above. Oslo being volatile, we still have the clear visual of the global trend where the increases outpace the decreases significantly since the start of 1900s.



```
#EXPLORING GLOBAL WEATHER TRENDS

#LOAD FILES
#CREATE DATAFRAMES
#ANALYSE
#PLOT RESULTS

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

path = r'C:\Users\omara\Documents\DatAn\Data Analysis\DAND\proj1\data\\'
file_oslo = 'oslo_nonnull.csv'
file_world = 'global_temp.csv'
def readcsv(filename):
    return pd.read_csv(path+filename)

df_oslo = readcsv(file_oslo)
df_world = readcsv(file_world)

dt = 20
def movingAvg(a):
    MA = np.zeros(len(a))
    for i in range(dt,len(a+1)):
        MA[i]= np.sum(a[i-dt:i+1])/dt
    return MA

MA = pd.DataFrame({
    'year': df_oslo['year'],
    'oslo_avg': movingAvg(df_oslo['oslo_avg'].values)
}).iloc[dt:]

o_plot = MA.plot(x='year', figsize=(10,8), title='Average yearly temperature in Oslo')
o_plot.set(ylabel='temperature C')
o_plot.axvspan(1980, 2013, color='orange', alpha=0.2)
```

```
MA_g = pd.DataFrame({
    'year': df_world['year'],
    'avg_temp': movingAvg(df_world['avg_temp'].values)
}).iloc[dt:]

g_plot = MA_g.plot(x='year', figsize=(10,8), legend='global temperature', title='Average
yearly Global Temperature')
g_plot.set(ylabel='temperature C')
g_plot.axvspan(1980, 2015, color='orange', alpha=0.2)

def mapper(y):
    x = np.copy(y)
    x[x > 0] = 1
    x[x < 0] = -1
    return x

MA_diff = (MA['oslo_avg'].values[1:] - MA['oslo_avg'].values[:-1])
MAg_diff = (MA_g['avg_temp'].values[1:] - MA_g['avg_temp'].values[:-1])

MA_change = mapper(MA_diff)
MAg_change = mapper(MAg_diff)

##Changes in temp since the 1970s for Oslo
MA70s = MA_diff[:100]#[-44:]
print('temp increase:', np.mean(MA70s[MA70s > 0]))
print('temp decrease:', np.mean(MA70s[MA70s < 0]))

print('number of increases:', len(MA70s[MA70s > 0]))
print('number of decreases:', len(MA70s[MA70s < 0]))

##Changes in temp since the 1970s for World
MAg70s = MAg_diff[:100]#[-46:]
print('temp increase:', np.mean(MAg70s[MAg70s > 0]))
print('temp decrease:', np.mean(MAg70s[MAg70s < 0]))

print('number of increases:', len(MAg70s[MAg70s > 0]))
print('number of decreases:', len(MAg70s[MAg70s < 0]))
```

```
#CUMSUM of yearly (moving avg) changes
MA_cumsum = np.cumsum(MA_change)
MAg_cumsum = np.cumsum(MAg_change)

fig, (ax1, ax2) = plt.subplots(2, 1, sharex=False, figsize=(13,10))
ax1.plot(MA_g['year'].values[1:], MAg_cumsum)
ax1.set(title='Global temperature change',  ylabel='cumsum of changes')
ax2.plot(MA['year'].values[1:], MA_cumsum)
ax2.set(title='Oslo temperature change', xlabel='year',  ylabel='cumsum of changes')
```