

Introduction

Analyzing Boston city and global temperature data and compare the temperature trends to overall global temperature trends.

For this project I have used python for coding with libraries like pandas , numpy and matplotlib.

```
In [8]: import sys
if "your\path\to\xelatex" not in sys.path:
    print('adding path') # I just add this to know if the path was present or not.
    sys.path.append("your\path\to\xelatex")

File "<ipython-input-8-832418ccb5c3>", line 2
if "C:\your\path\to\xelatex" not in sys.path:

SyntaxError: (unicode error) 'unicodeescape' codec can't decode bytes in position 15-17: truncated \xxx escape
```

Analysis

```
In [1]: #check directory
import os
os.getcwd()

Out[1]: 'C:\\Users\\Shivam'
```

```
In [2]: #change working directory
os.chdir("C:\\Users\\Shivam\\Downloads")
```

```
In [3]: import pandas as pd
import numpy as np
```

```
In [4]: #load data
df_city = pd.read_csv("results (1).csv")
df_global = pd.read_csv("results (2).csv")
```

```
In [5]: df_city.head(10)

Out[5]:
```

	year	city	country	avg_temp
0	1743	Boston	United States	1.19
1	1744	Boston	United States	9.63
2	1745	Boston	United States	-1.37
3	1746	Boston	United States	NaN
4	1747	Boston	United States	NaN
5	1748	Boston	United States	NaN
6	1749	Boston	United States	NaN
7	1750	Boston	United States	7.88
8	1751	Boston	United States	8.60
9	1752	Boston	United States	0.36

```
In [6]: #impute the missing values with mean
df_city["avg_temp"] = df_city["avg_temp"].fillna(df_city["avg_temp"].mean())
```

```
In [7]: df_city["avg_temp"] = df_city["avg_temp"].round(2)
```

```
In [8]: df_city.head(10)

Out[8]:
```

	year	city	country	avg_temp
0	1743	Boston	United States	1.19
1	1744	Boston	United States	9.63
2	1745	Boston	United States	-1.37
3	1746	Boston	United States	7.26
4	1747	Boston	United States	7.26
5	1748	Boston	United States	7.26
6	1749	Boston	United States	7.26
7	1750	Boston	United States	7.88
8	1751	Boston	United States	8.60
9	1752	Boston	United States	0.36

```
In [9]: #check if there is any other missing values
df_city.isna().sum()

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

```
In [10]: df_global.tail()

Out[10]:
```

	year	avg_temp
261	2011	9.52
262	2012	9.51
263	2013	9.61
264	2014	9.57
265	2015	9.83

```
In [11]: #check missing values for global data
df_global.isnull().any()

Out[11]: year      False
avg_temp  False
dtype: bool
```

Calculate moving avg for city data

For calculating the moving average, I have taken average of previous 7 days average temperature data.

```
In [12]: df_city.shape

Out[12]: (271, 4)
```

```
In [13]: df_city["moving_avg"] = float(0)
```

```
In [14]: for i in range(6,len(df_city)):
    avg_t = 0
    avg_t = np.mean(df_city["avg_temp"][i-6:i])
    df_city["moving_avg"][i] = avg_t

C:\Users\Shivam\Anaconda3\envs\py35\lib\site-packages\ipykernel_launcher.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
after removing the cwd from sys.path.
```

```
In [15]: df_city.tail(20)

Out[15]:
```

	year	city	country	avg_temp	moving_avg
251	1994	Boston	United States	7.96	8.155000
252	1995	Boston	United States	8.13	8.165000
253	1996	Boston	United States	7.93	8.268333
254	1997	Boston	United States	7.87	8.056667
255	1998	Boston	United States	9.63	7.855000
256	1999	Boston	United States	9.30	8.226667
257	2000	Boston	United States	8.00	8.470000
258	2001	Boston	United States	8.91	8.476667
259	2002	Boston	United States	9.02	8.606667
260	2003	Boston	United States	7.78	8.788333
261	2004	Boston	United States	8.09	8.773333
262	2005	Boston	United States	8.56	8.516667
263	2006	Boston	United States	9.48	8.393333
264	2007	Boston	United States	8.36	8.640000
265	2008	Boston	United States	8.53	8.548333
266	2009	Boston	United States	8.07	8.466667
267	2010	Boston	United States	9.58	8.515000
268	2011	Boston	United States	9.12	8.763333
269	2012	Boston	United States	10.06	8.856667
270	2013	Boston	United States	10.38	8.953333

```
In [16]: df_global.shape

Out[16]: (266, 2)
```

```
In [17]: df_global["moving_avg"] = float(0)
```

```
In [18]: for i in range(6,len(df_city)):
    avg_t = 0
    avg_t = np.mean(df_global["avg_temp"][i-6:i])
    df_global["moving_avg"][i] = avg_t

C:\Users\Shivam\Anaconda3\envs\py35\lib\site-packages\ipykernel_launcher.py:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
after removing the cwd from sys.path.
```

```
In [19]: df_global.head(10)

Out[19]:
```

	year	avg_temp	moving_avg
0	1750	8.72	0.000000
1	1751	7.98	0.000000
2	1752	5.78	0.000000
3	1753	8.39	0.000000
4	1754	8.47	0.000000
5	1755	8.36	0.000000
6	1756	8.85	7.950000
7	1757	9.02	7.971667
8	1758	6.74	8.145000
9	1759	7.99	8.305000

```
In [20]: df = df_city.iloc[13:, [4]]
df

Out[20]:
```

	moving_avg
13	6.036667
14	6.016667
15	5.691667
16	6.646667
17	6.555000
...	...
266	8.466667
267	8.515000
268	8.763333
269	8.856667
270	8.953333

258 rows × 1 columns

```
In [21]: df_g = df_global.iloc[6:264, [2]]
```

```
In [22]: df_g

Out[22]:
```

	moving_avg
6	7.950000
7	7.971667
8	8.145000
9	8.305000
10	8.238333
...	...
259	9.540000
260	9.536667
261	9.600000
262	9.570000
263	9.566667

258 rows × 1 columns

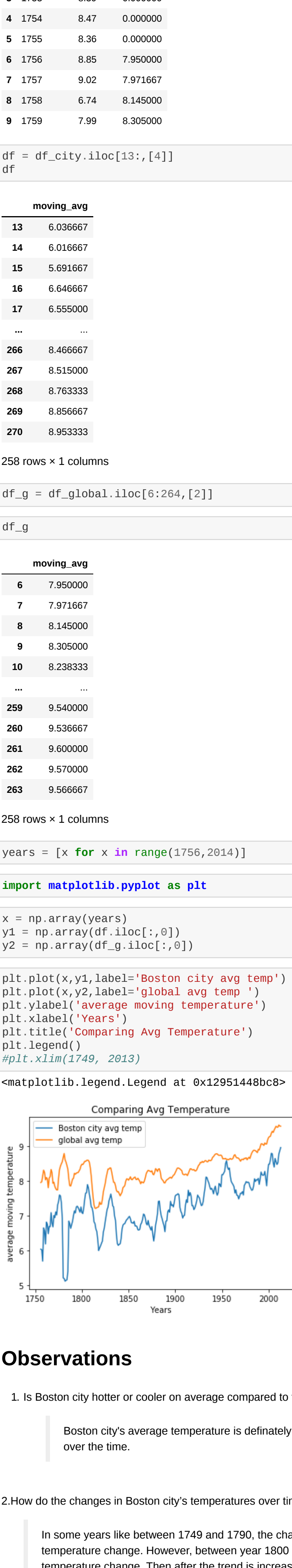
```
In [23]: years = [x for x in range(1756, 2014)]
```

```
In [24]: import matplotlib.pyplot as plt
```

```
In [25]: x = np.array(years)
y1 = np.array(df.iloc[:, 0])
y2 = np.array(df_g.iloc[:, 0])
```

```
In [26]: plt.plot(x, y1, label='Boston city avg temp')
plt.plot(x, y2, label='global avg temp')
plt.ylabel('average moving temperature')
plt.xlabel('years')
plt.title('Comparing Avg Temperature')
plt.legend()
#plt.xlim(1749, 2013)

Out[26]: <matplotlib.legend.Legend at 0x12951448bc8>
```



Observations

1. Is Boston city hotter or cooler than the global average? Has the difference been consistent over time?

Boston city's average temperature is definitely cooler than the average global temperature consistently over the time.

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

In some years like between 1749 and 1790, the change in Boston city's temperature is opposite to the global temperature change. However, between 1790 and 1850, the changes are proportional to the global temperature change. Then after the trend is increasing in average temperature for both Boston city and global.

3.What does the overall trend look like? Is the world getting hotter or cooler? Has the trend been consistent over the last few hundred years?

Yes the world is getting hotter and the trend is consistent over the few hundred years.

1. Some major differences and similarities between Boston City's and global average temperature?

In around 1780's and 1790's, there was a drastic difference between both of the cities. This means the Boston city got cooler and other part of the world too hot which made the global avg temp. to rise.

In 1950, we see a unusual scenario where Boston City's and global average temperature was same. This mean Boston got hot and other part got cold. This is unusual because over the past years this has never happened. Boston's avg temp was cooler than the global average temp in this dataset.

Reference