Analysing Weather Dataset

November 21, 2018

```
In [99]: import pandas as pd
         import numpy as np
         import seaborn as sb
         % matplotlib inline

         global_df = pd.read_csv('global_data.csv')
         city_df = pd.read_csv('city_data.csv')
```

1 Outline

1. Tools : Google Sheet ,Jupyter Notebook , Python

2 SQL Query

2.0.1 Query the Global Data

2. to get the global data
3. Select *
4. From global_data

2.0.2 Query the City Data

6. To get the city which is Kuala Lumpur in the country Malaysia weather dataset
7. Select *
8. From city_data
9. Where Country = 'Malaysia'

2.0.3 To find the city near me with my country

11. Select city
12. From city_list
13. Where Country = 'Malaysia'

3 Process

1. Fix the datatpes

2. Take the average of 10 years between 1825 and 2013 for local data and 1804 and 2015 for global data.

4 Key Considerations

1. Using moving average to keep the line as smooth as possible.
2. Making sure that the two line chart has different colours so that the chart can be visualize clearly.
3. Making sure the y- axis and x - axis shown the labelled that is appropriate for the data.

```
In [100]: global_df.head()
```

Out[100]:

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

```
In [101]: city_df.head()
```

Out[101]:

	year	avg_temp
0	1839	25.74
1	1840	25.96
2	1841	26.10
3	1842	26.18
4	1843	26.25

```
In [102]: # Check the datatypes of global_df
          global_df.dtypes
```

```
Out[102]: year          int64
          avg_temp    float64
          dtype: object
```

```
In [103]: # Drop city dataframe missing values
          city_df.dropna()
```

Out[103]:

	year	avg_temp
0	1839	25.74
1	1840	25.96
2	1841	26.10
3	1842	26.18
4	1843	26.25
5	1844	25.77
6	1845	25.64
7	1846	26.44
8	1847	25.89
9	1850	26.06

10	1851	26.13
11	1852	26.02
12	1853	26.26
13	1854	25.98
14	1855	26.12
15	1856	26.21
16	1858	26.21
17	1859	26.27
18	1860	25.97
19	1861	25.93
20	1863	26.11
21	1864	25.95
22	1865	26.22
23	1866	26.22
24	1867	26.12
25	1868	26.14
26	1869	25.95
27	1870	25.59
28	1871	25.68
29	1872	26.23
..	...	...
141	1984	26.59
142	1985	26.83
143	1986	26.93
144	1987	27.27
145	1988	27.15
146	1989	26.85
147	1990	27.21
148	1991	27.04
149	1992	27.05
150	1993	26.99
151	1994	27.00
152	1995	27.05
153	1996	27.04
154	1997	27.29
155	1998	27.89
156	1999	26.95
157	2000	27.14
158	2001	27.24
159	2002	27.57
160	2003	27.36
161	2004	27.35
162	2005	27.59
163	2006	27.29
164	2007	27.23
165	2008	27.12
166	2009	27.47
167	2010	27.69

```
168  2011     27.27
169  2012     27.36
170  2013     27.80

[171 rows x 2 columns]
```

In [104]:
```
# convert city dataframe to integer
city_df.astype(int)
```

Out[104]:

	year	avg_temp
0	1839	25
1	1840	25
2	1841	26
3	1842	26
4	1843	26
5	1844	25
6	1845	25
7	1846	26
8	1847	25
9	1850	26
10	1851	26
11	1852	26
12	1853	26
13	1854	25
14	1855	26
15	1856	26
16	1858	26
17	1859	26
18	1860	25
19	1861	25
20	1863	26
21	1864	25
22	1865	26
23	1866	26
24	1867	26
25	1868	26
26	1869	25
27	1870	25
28	1871	25
29	1872	26
..	...	...
141	1984	26
142	1985	26
143	1986	26
144	1987	27
145	1988	27
146	1989	26
147	1990	27

```
148  1991        27
149  1992        27
150  1993        26
151  1994        27
152  1995        27
153  1996        27
154  1997        27
155  1998        27
156  1999        26
157  2000        27
158  2001        27
159  2002        27
160  2003        27
161  2004        27
162  2005        27
163  2006        27
164  2007        27
165  2008        27
166  2009        27
167  2010        27
168  2011        27
169  2012        27
170  2013        27

[171 rows x 2 columns]
```

In [105]:
```
# Convert global data frame dataset to integer
global_df.astype(int)
```

Out[105]:

	year	avg_temp
0	1750	8
1	1751	7
2	1752	5
3	1753	8
4	1754	8
5	1755	8
6	1756	8
7	1757	9
8	1758	6
9	1759	7
10	1760	7
11	1761	8
12	1762	8
13	1763	7
14	1764	8
15	1765	8
16	1766	8
17	1767	8

```
18   1768         6
19   1769         7
20   1770         7
21   1771         7
22   1772         8
23   1773         8
24   1774         8
25   1775         9
26   1776         8
27   1777         8
28   1778         8
29   1779         8
..    ...       ...
236  1986         8
237  1987         8
238  1988         9
239  1989         8
240  1990         9
241  1991         9
242  1992         8
243  1993         8
244  1994         9
245  1995         9
246  1996         9
247  1997         9
248  1998         9
249  1999         9
250  2000         9
251  2001         9
252  2002         9
253  2003         9
254  2004         9
255  2005         9
256  2006         9
257  2007         9
258  2008         9
259  2009         9
260  2010         9
261  2011         9
262  2012         9
263  2013         9
264  2014         9
265  2015         9

[266 rows x 2 columns]
```

In [108]:
```
# Calculate the moving average of global average temperature across 10 years
global_df['moving_average'] = global_df['avg_temp'].rolling(window = 10).mean()
```

In [109]:
```
# Calculate the moving average of city average temperature acorss 10 years
city_df['moving_average'] = city_df['avg_temp'].rolling(window = 10).mean()
```

In [110]:
```
# Drop missing values
global_df.dropna()
```

Out[110]:

	year	avg_temp	moving_average
9	1759	7.99	8.030
10	1760	7.19	7.877
11	1761	8.77	7.956
12	1762	8.61	8.239
13	1763	7.50	8.150
14	1764	8.40	8.143
15	1765	8.25	8.132
16	1766	8.41	8.088
17	1767	8.22	8.008
18	1768	6.78	8.012
19	1769	7.69	7.982
20	1770	7.69	8.032
21	1771	7.85	7.940
22	1772	8.19	7.898
23	1773	8.22	7.970
24	1774	8.77	8.007
25	1775	9.18	8.100
26	1776	8.30	8.089
27	1777	8.26	8.093
28	1778	8.54	8.269
29	1779	8.98	8.398
30	1780	9.43	8.572
31	1781	8.10	8.597
32	1782	7.90	8.568
33	1783	7.68	8.514
34	1784	7.86	8.423
35	1785	7.36	8.241
36	1786	8.26	8.237
37	1787	8.03	8.214
38	1788	8.45	8.205
..	...	...	...
236	1986	8.83	8.827
237	1987	8.99	8.841
238	1988	9.20	8.892
239	1989	8.92	8.911
240	1990	9.23	8.936
241	1991	9.18	8.937
242	1992	8.84	8.957
243	1993	8.87	8.941
244	1994	9.04	8.976
245	1995	9.35	9.045

```
246  1996      9.04           9.066
247  1997      9.20           9.087
248  1998      9.52           9.119
249  1999      9.29           9.156
250  2000      9.20           9.153
251  2001      9.41           9.176
252  2002      9.57           9.249
253  2003      9.53           9.315
254  2004      9.32           9.343
255  2005      9.70           9.378
256  2006      9.53           9.427
257  2007      9.73           9.480
258  2008      9.43           9.471
259  2009      9.51           9.493
260  2010      9.70           9.543
261  2011      9.52           9.554
262  2012      9.51           9.548
263  2013      9.61           9.556
264  2014      9.57           9.581
265  2015      9.83           9.594

[257 rows x 3 columns]
```

In [111]:
```
# Drop the missing values
city_df.dropna()
```

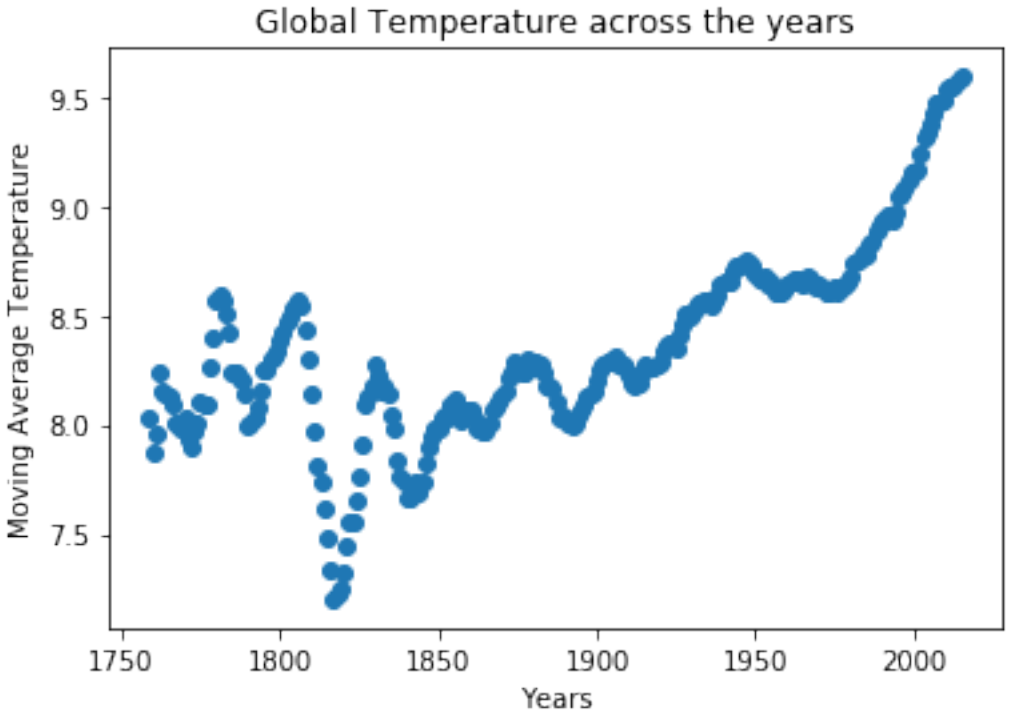
Out[111]:

	year	avg_temp	moving_average
9	1850	26.06	26.003
10	1851	26.13	26.042
11	1852	26.02	26.048
12	1853	26.26	26.064
13	1854	25.98	26.044
14	1855	26.12	26.031
15	1856	26.21	26.075
16	1858	26.21	26.132
17	1859	26.27	26.115
18	1860	25.97	26.123
19	1861	25.93	26.110
20	1863	26.11	26.108
21	1864	25.95	26.101
22	1865	26.22	26.097
23	1866	26.22	26.121
24	1867	26.12	26.121
25	1868	26.14	26.114
26	1869	25.95	26.088
27	1870	25.59	26.020
28	1871	25.68	25.991
29	1872	26.23	26.021

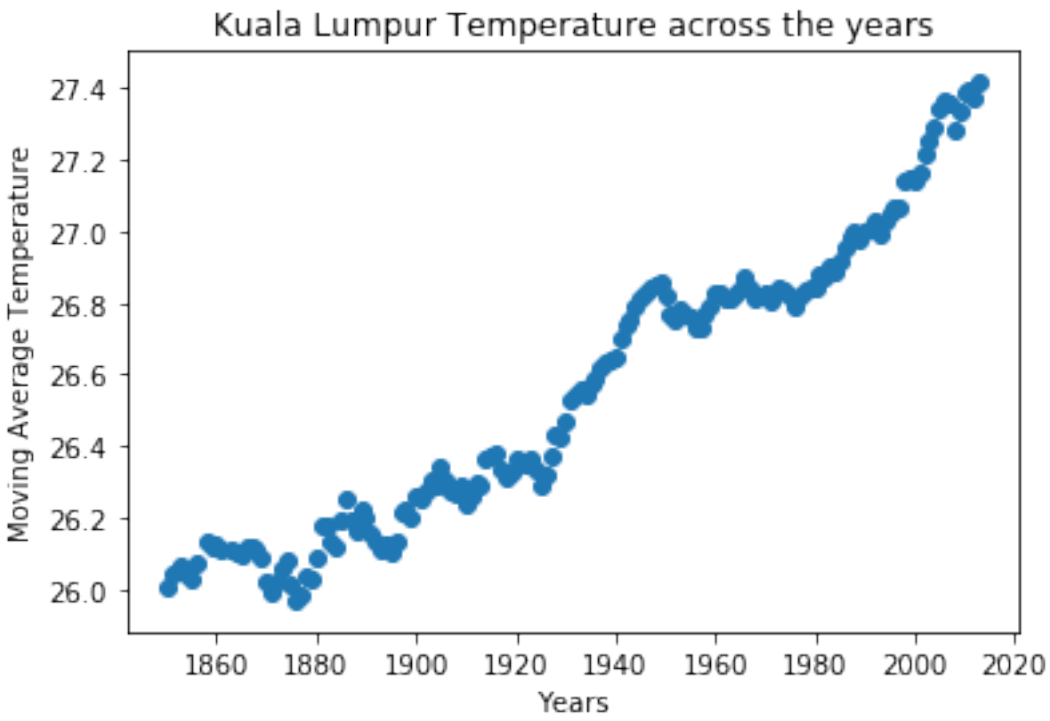
```
30   1873     26.47          26.057
31   1874     26.16          26.078
32   1875     25.59          26.015
33   1876     25.75          25.968
34   1877     26.31          25.987
35   1878     26.61          26.034
36   1879     25.88          26.027
37   1880     26.21          26.089
38   1881     26.60          26.181
..    ...       ...             ...
141  1984     26.59          26.891
142  1985     26.83          26.918
143  1986     26.93          26.951
144  1987     27.27          26.983
145  1988     27.15          26.999
146  1989     26.85          26.976
147  1990     27.21          26.999
148  1991     27.04          27.008
149  1992     27.05          27.027
150  1993     26.99          26.991
151  1994     27.00          27.032
152  1995     27.05          27.054
153  1996     27.04          27.065
154  1997     27.29          27.067
155  1998     27.89          27.141
156  1999     26.95          27.151
157  2000     27.14          27.144
158  2001     27.24          27.164
159  2002     27.57          27.216
160  2003     27.36          27.253
161  2004     27.35          27.288
162  2005     27.59          27.342
163  2006     27.29          27.367
164  2007     27.23          27.361
165  2008     27.12          27.284
166  2009     27.47          27.336
167  2010     27.69          27.391
168  2011     27.27          27.394
169  2012     27.36          27.373
170  2013     27.80          27.417

[162 rows x 3 columns]
```

In [125]:
```
# Plotting the Global Temperature across the years
x = global_df['year']
y = global_df['moving_average']

plt.scatter(x , y)
```

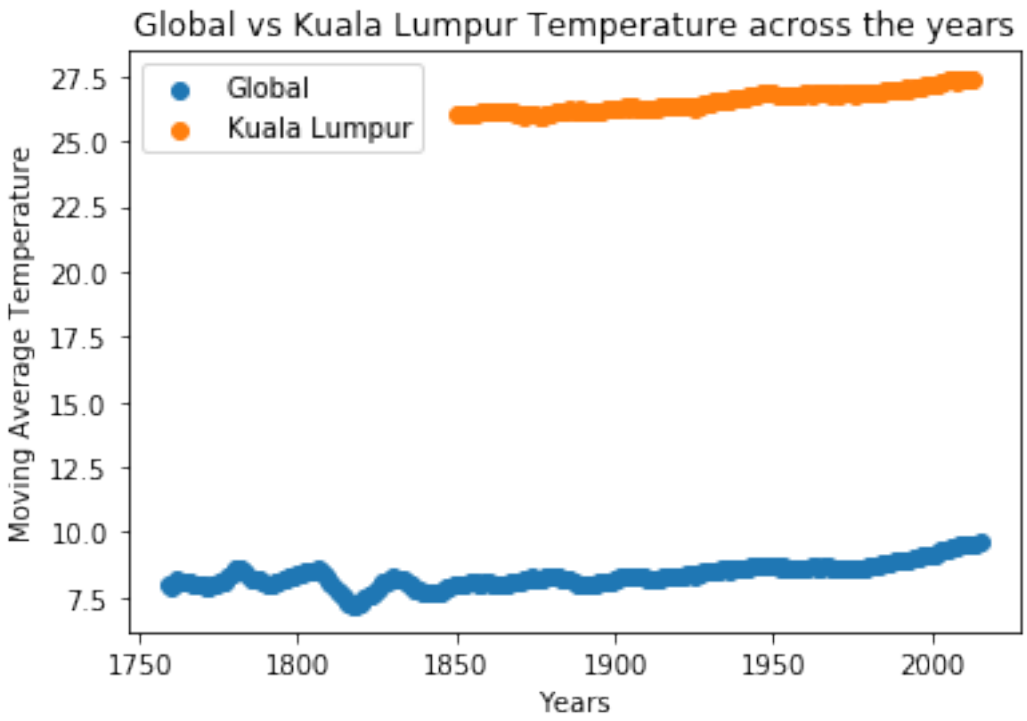
```
plt.xlabel('Years')
plt.ylabel('Moving Average Temperature')
plt.title('Global Temperature across the years');
```



In [123]:
```
#Plotting Kuala Lumpur Temperature across the year
x = city_df['year']
y = city_df['moving_average']

plt.scatter(x , y)
plt.xlabel('Years')
plt.ylabel('Moving Average Temperature')
plt.title('Kuala Lumpur Temperature across the years');
```



In [124]:
```
# Plotting Global Temperature vs Kuala Lumpur Temperature
x = global_df['year']
y = global_df['moving_average']

plt.scatter(x , y , label = 'Global')

x = city_df['year']
y = city_df['moving_average']

plt.scatter(x , y , label = 'Kuala Lumpur')
plt.xlabel('Years')
plt.ylabel('Moving Average Temperature')
plt.legend()
plt.title('Global vs Kuala Lumpur Temperature across the years');
```



Similarities and Differences : 1. Both temperature for Local and Global has been relatively stable throughout the years. 2. By analysing the local trend, we can observe that the local temperature has dropped significantly in 1870 to 25.59 Celsius and gradually increases to 27.8 Celsius in 2013. 3. While for global temperature, the global temperature has dropped to its lowest in 1840 which is 7.74 Celsius and rose steadily to 9.63 in 2015 which is the highest temperature for the average global temperature. 4. Both experience rapid rise in temperature in the year 1870 to 2015. Both data shows that global and local temperature recorded their hottest years in the latest data collected which is 2013 for local and 2015 for global. 5. For Kuala Lumpur, the weather is relatively hotter throughout the years compare to the global temperature.

In [ ]: