# Udacity_Project 1_Explore Weather Trends

July 18, 2020

#

Udacity Data Analysis Nanodegree

##

Project: Weather Trends Dataset

###

Noaman Mangera, June 2020

### 0.1 Table of Contents

Data Wrangling

Exploratory Data Analysis

Conclusions

[47]:
```
#import necessary modules
import numpy as np
import pandas as pd
import seaborn as sns
```

### 0.2 Data Wrangling

[48]:
```
#read in dataset with London temperatures
df_london = pd.read_csv(r'C:\Users\noama\LondonTemperature.csv')
```

[49]:
```
#inspect first 5 rows of dataset with temperature data on London
df_london.head()
```

[49]:

| | year | city | country | avg_temp |
|---|---|---|---|---|
| 0 | 1743 | London | United Kingdom | 7.54 |
| 1 | 1744 | London | United Kingdom | 10.34 |
| 2 | 1745 | London | United Kingdom | 4.13 |
| 3 | 1746 | London | United Kingdom | NaN |
| 4 | 1747 | London | United Kingdom | NaN |

[50]:
```
#read in dataset with global temepratures
df_global = pd.read_csv(r"C:\Users\noama\GlobalTemperature.csv")
```

[51]:
```
#inspect first 5 rows of dataset with global temperature data
df_global.head()
```

[51]:

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

[52]:
```
#combine London data with global data
df = pd.merge(df_london, df_global, how='outer', on='year',␣
 ↪suffixes=('_london', '_global'))
```

[53]:
```
#inspects first 5 rows of combined temperature dataset
df.head()
```

[53]:

| | year | city | country | avg_temp_london | avg_temp_global |
|---|---|---|---|---|---|
| 0 | 1743 | London | United Kingdom | 7.54 | NaN |
| 1 | 1744 | London | United Kingdom | 10.34 | NaN |
| 2 | 1745 | London | United Kingdom | 4.13 | NaN |
| 3 | 1746 | London | United Kingdom | NaN | NaN |
| 4 | 1747 | London | United Kingdom | NaN | NaN |

[54]:
```
#remove city and country columns from merged datset
df = df.drop(columns=['city', 'country'])
```

## Exploratory Data Analysis

[55]:
```
#inspect number of rows and columns
df.shape
```

[55]: (273, 3)

[56]:
```
#calculate 10 year rolling for london temperatures
df['roll10_London'] = df.avg_temp_london.rolling(10).mean()
```

[57]:
```
#calculate 10 year rolling for global temperatures
df['roll10_Global'] = df.avg_temp_global.rolling(10).mean()
```

[58]:
```
#calculate summary statistics for numerical variables
df.describe()
```

[58]:
```
              year  avg_temp_london  avg_temp_global  roll10_London  \
count   273.000000       267.000000       266.000000     255.000000
mean   1879.000000         9.435880         8.369474       9.450353
std      78.952517         0.751648         0.584747       0.370030
min    1743.000000         4.130000         5.780000       8.802000
25%    1811.000000         9.040000         8.082500       9.203000
50%    1879.000000         9.420000         8.375000       9.415000
75%    1947.000000         9.880000         8.707500       9.620500
max    2015.000000        11.190000         9.830000      10.701000

       roll10_Global
count     257.000000
mean        8.353961
std         0.452483
min         7.203000
25%         8.059000
50%         8.275000
75%         8.642000
max         9.594000
```

[59]:
```
#reshape merged dataset into tidy form
df_melt_roll10 = df.melt(id_vars=['year', 'avg_temp_london',␣
 ↪'avg_temp_global'], var_name='Location',  value_name='Temperature', )
```

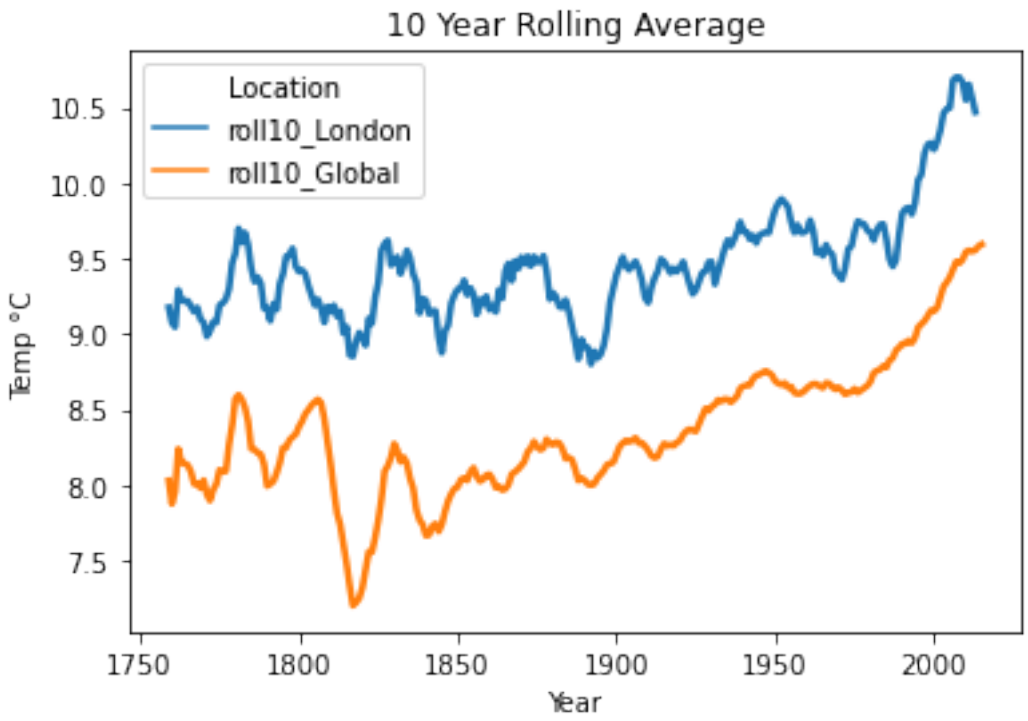
[60]:
```
#inspect first five rows of tidy dataset
df_melt_roll10.head()
```

[60]:

| | year | avg_temp_london | avg_temp_global | Location | Temperature |
|---|---|---|---|---|---|
| 0 | 1743 | 7.54 | NaN | roll10_London | NaN |
| 1 | 1744 | 10.34 | NaN | roll10_London | NaN |
| 2 | 1745 | 4.13 | NaN | roll10_London | NaN |
| 3 | 1746 | NaN | NaN | roll10_London | NaN |
| 4 | 1747 | NaN | NaN | roll10_London | NaN |

[61]:
```
#create line plot of rolling average temperatures
g = sns.lineplot(df_melt_roll10.year, df_melt_roll10.Temperature,␣
 ↪hue=df_melt_roll10.Location, palette="tab10", linewidth=2.5);

#set title and axis labels
g.set(title = '10 Year Rolling Average', xlabel = 'Year', ylabel = 'Temp °C' );
```




[62]:
```
#export visualisation
g.get_figure().savefig('10 Year Rolling Average.png')
```

## Conclusion

**Question: Is London hotter or cooler compared to the global average?** > Answer: London is warmer than the global average, by around 1 degree

**Question: Has the difference been consistent over time?** > Answer: Yes, the difference appears to be consistent over time.

**Question: How does London's temperature compare to the changes in the global average over time?** > Answer: The trends align over time.

**Question: What does the overall trend look like?** > Answer: Both are trending upwards.

**Question: Is the world getting hotter or cooler?** > Answer: The world has been getting warmer.

**Question: Has the trend been consistent over the last few hundred years?** > Answer: This trend has been consistent over the last 150 years.

[ ]: