

project1__NB

December 2, 2019

DAND: Explore Weather Trends

0.0.1 Steps Taken to Complete the Project :

- **Data Acquisition** to acquire the average global temperature and Riyadh's average temperature using two queries
- **Data cleaning** using python to impute median to NaN values
- **Moving averages creation** by using the rolling method in python
- **Line chart creation** to visualize the created moving averages
- **Insights extraction** from the line chart

0.0.2 Queries Used to Extract Data

- To extract global temps

```
SELECT * FROM global_data
```

- To extract nearest city temps

```
SELECT city,* FROM city_data WHERE city ='Riyadh'
```

```
[31]: import pandas as pd
import matplotlib.pyplot as plt
```

```
[3]: !ls
```

```
global_data.csv  nearest_CityTemp.csv  project1_NB.ipynb
```

```
[6]: df_global= pd.read_csv('global_data.csv')
```

```
[7]: df_riyadh = pd.read_csv('nearest_CityTemp.csv')
```

```
[9]: df_riyadh
```

```
[9]:   year    city    country  avg_temp
0   1843  Riyadh  Saudi Arabia    24.74
```

1	1844	Riyadh	Saudi Arabia	15.45
2	1845	Riyadh	Saudi Arabia	20.82
3	1846	Riyadh	Saudi Arabia	NaN
4	1847	Riyadh	Saudi Arabia	NaN
..
166	2009	Riyadh	Saudi Arabia	26.71
167	2010	Riyadh	Saudi Arabia	27.37
168	2011	Riyadh	Saudi Arabia	26.40
169	2012	Riyadh	Saudi Arabia	26.83
170	2013	Riyadh	Saudi Arabia	27.78

[171 rows x 4 columns]

```
[11]: #Merging the two dataframes
df=pd.merge(df_global,df_riyadh,on='year')
```

```
[12]: df
```

```
[12]:
```

	year	avg_temp_x	city	country	avg_temp_y
0	1843	8.17	Riyadh	Saudi Arabia	24.74
1	1844	7.65	Riyadh	Saudi Arabia	15.45
2	1845	7.85	Riyadh	Saudi Arabia	20.82
3	1846	8.55	Riyadh	Saudi Arabia	NaN
4	1847	8.09	Riyadh	Saudi Arabia	NaN
..
166	2009	9.51	Riyadh	Saudi Arabia	26.71
167	2010	9.70	Riyadh	Saudi Arabia	27.37
168	2011	9.52	Riyadh	Saudi Arabia	26.40
169	2012	9.51	Riyadh	Saudi Arabia	26.83
170	2013	9.61	Riyadh	Saudi Arabia	27.78

[171 rows x 5 columns]

```
[25]: #Checking for whether there are NaN values or not
df[df['avg_temp_y'].isnull() == True]
```

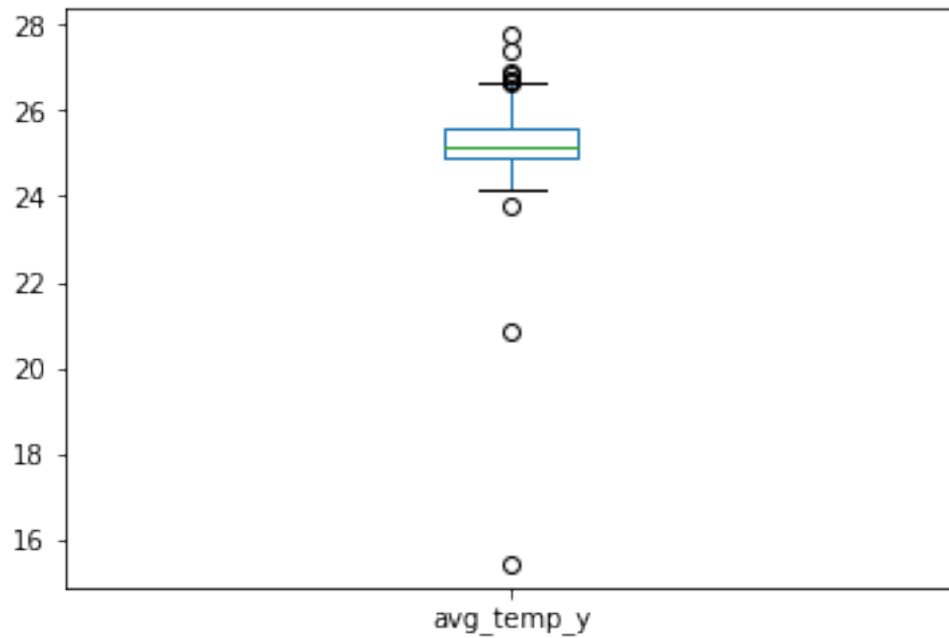
```
[25]:
```

	year	avg_temp_x	city	country	avg_temp_y
3	1846	8.55	Riyadh	Saudi Arabia	NaN
4	1847	8.09	Riyadh	Saudi Arabia	NaN

Note: This is an issue, we need to impute either the median or the mean

```
[35]: df['avg_temp_y'].plot.box()
```

```
[35]: <matplotlib.axes._subplots.AxesSubplot at 0x7fdf84dd6550>
```

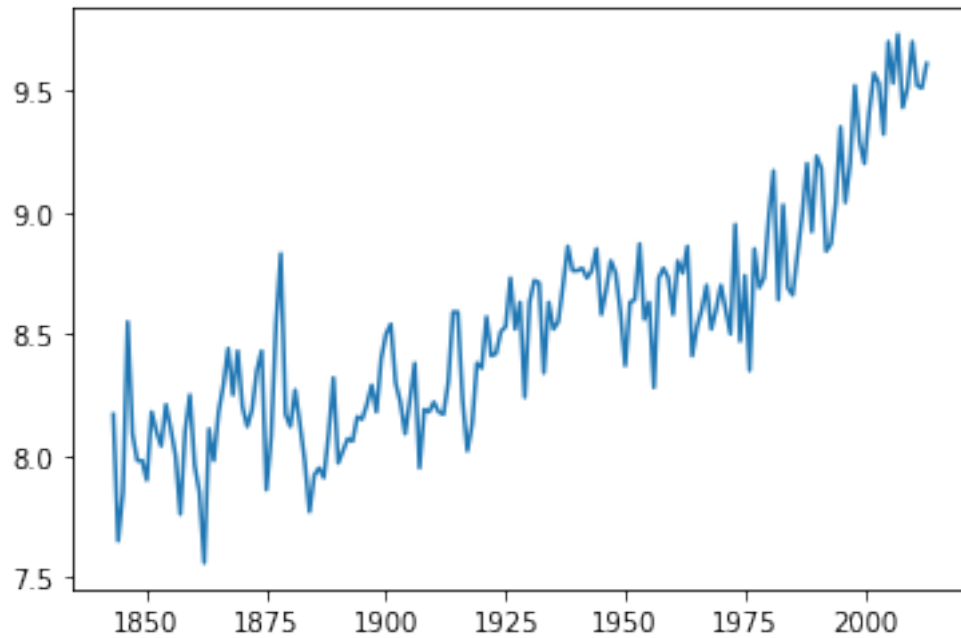


We need to impute the median since we have a lot of outliers

```
[50]: df['avg_temp_y'].fillna((df['avg_temp_y'].median()), inplace=True)
```

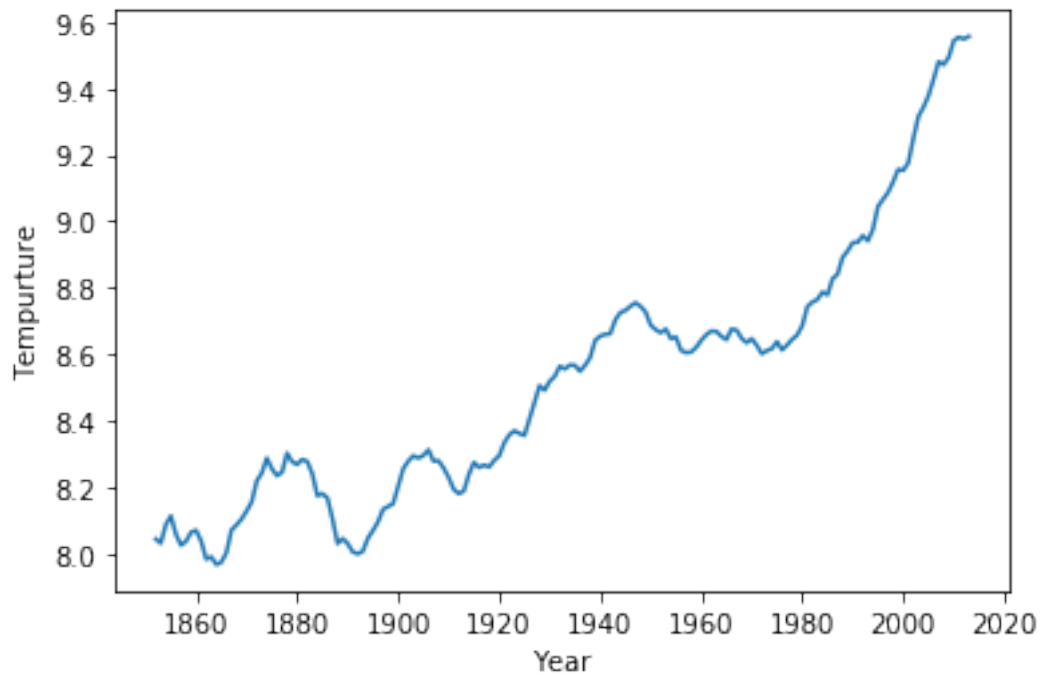
0.0.3 Moving Averages Creation

```
[55]: plt.plot(df['year'],df['avg_temp_x']);
```

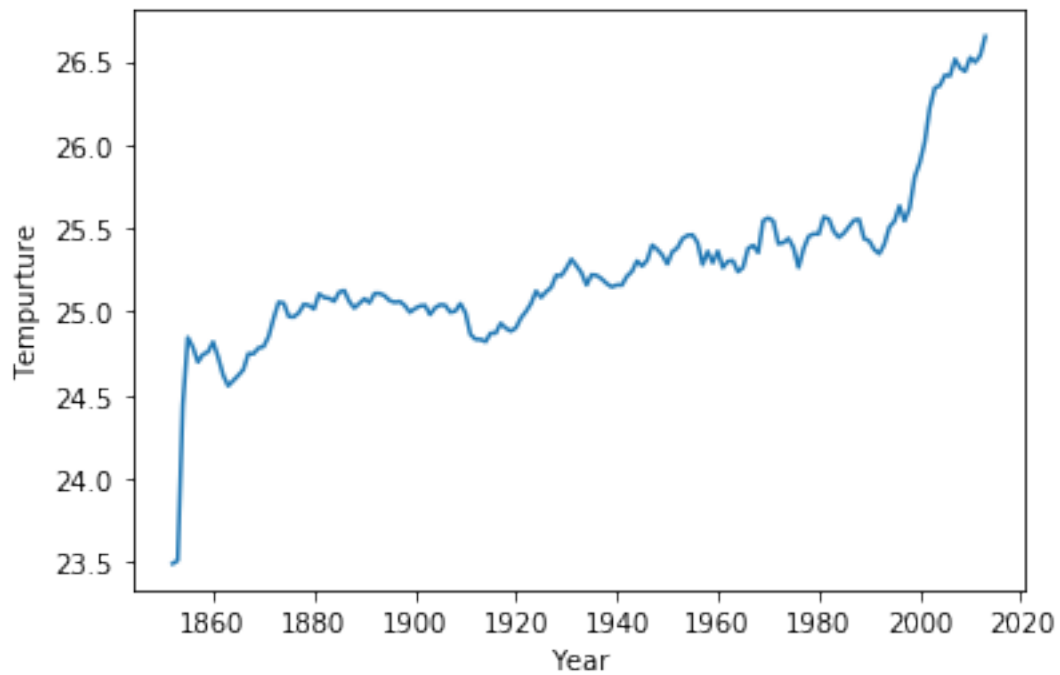


```
[67]: #Creating moving averages for a window of 10
short_rolling = df['avg_temp_x'].rolling(window=10).mean()
short_rolling2 = df['avg_temp_y'].rolling(window=10).mean()

[90]: plt.plot(df['year'],short_rolling, label = "Global Temp");
plt.xlabel('Year')
plt.ylabel('Tempurture');
```



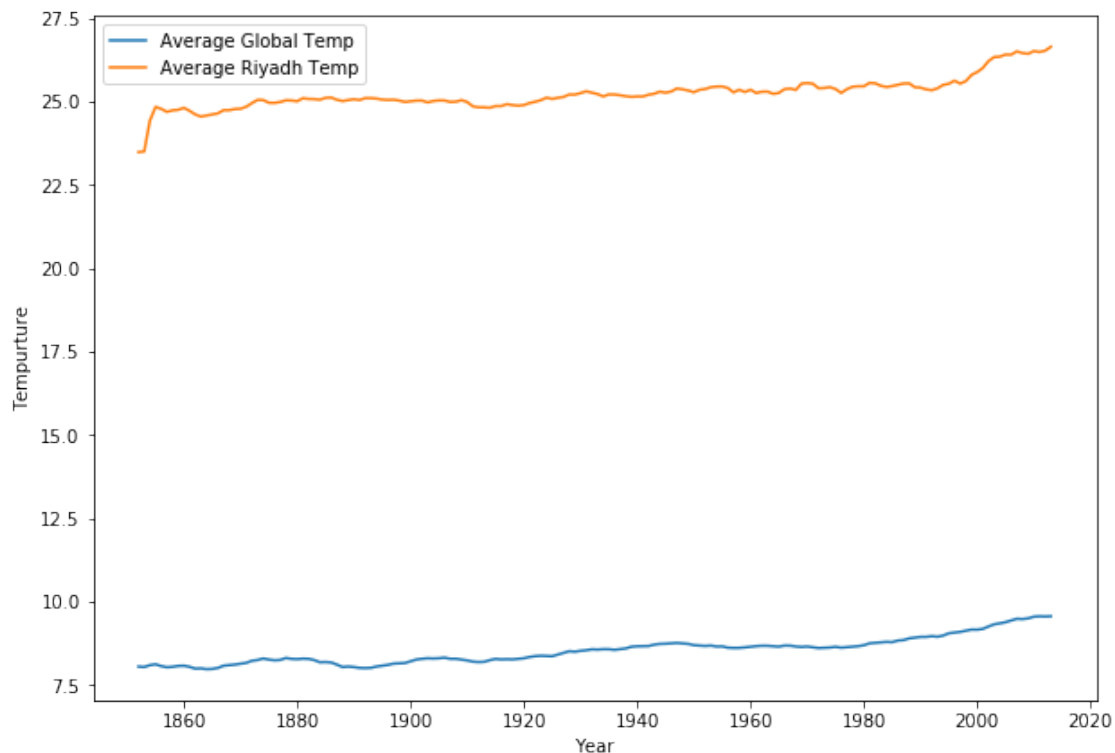
```
[89]: plt.plot(df['year'],short_rolling2, label = "Riyadh Temp");
plt.xlabel('Year')
plt.ylabel('Tempurture');
```



```
[91]: plt.figure(figsize=(10,7))

plt.plot(df['year'], short_rolling, label = "Average Global Temp")
plt.plot(df['year'], short_rolling2, label = "Average Riyadh Temp")
plt.xlabel('Year')
plt.ylabel('Tempurture')
plt.legend()
```

[91]: <matplotlib.legend.Legend at 0x7fdf4a3969b0>



Observations

- Riyadh's average temperature is way above the global average temperature
- Both the global average and Riyadh average temperature seem to be in an increasing pattern each year
- Both the global average and Riyadh's average temperature seems to increase sharply in almost the last 20 years
- In general, the global average temperature seems to have higher fluctuations than Riyadh's average temperature