Name: Ahmed Eldably
Date: May 3rd, 2020
Udacity Data Analysis Nanodegree
Project 1: Explore Weather Trends

Overview:

This project aims to analyze the local temperature in Berlin, Germany, and the global temperature data. The dataset was given by Udacity where I had to extract, manipulate, and visualize data.

Goals

1. Extraction of the yearly average temperature of Berlin data compared to the global one and export it to a CSV file.
2. Calculating the moving average for both average temperatures.
3. Creating a line chart that compares Berlin's temperatures with global temperatures.
4. Making observations based on the chart.

Project steps:

Step 1: Extracting data using SQL:

1. There are 3 datasets to extract from: city_list, city_data & global_data. So firstly, let's see what does the data look like.

   SELECT * FROM city_list;

Output   342 results   ⬇ Download CSV

city	country
Abidjan	Côte D'Ivoire
Abu Dhabi	United Arab Emirates
Abuja	Nigeria
Accra	Ghana
Adana	Turkey
Adelaide	Australia
Agra	India
Ahmadabad	India

```sql
SELECT * FROM city_data;
```

Output   70792 results   ⬇ Download CSV

year	city	country	avg_temp
1849	Abidjan	Côte D'Ivoire	25.58
1850	Abidjan	Côte D'Ivoire	25.52
1851	Abidjan	Côte D'Ivoire	25.67
1852	Abidjan	Côte D'Ivoire	
1853	Abidjan	Côte D'Ivoire	
1854	Abidjan	Côte D'Ivoire	
1855	Abidjan	Côte D'Ivoire	
1856	Abidjan	Côte D'Ivoire	26.28

```sql
SELECT * FROM global_data;
```

Output	266 results		↓ Download CSV
year		**avg_temp**	
1750		8.72	
1751		7.98	
1752		5.78	
1753		8.39	
1754		8.47	
1755		8.36	
1756		8.85	
1757		9.02	

2. Before extracting, the average temperature column name should be changed since both datasets have the same column name.

   ```sql
   ALTER TABLE global_data RENAME COLUMN avg_temp TO global_avg_temp;
   ALTER TABLE city_data RENAME COLUMN avg_temp TO city_avg_temp;
   ```

3. After changing column names, let's extract the yearly average temperature in Berlin and the globe year by year.

   ```sql
   SELECT global_data.year, city_data.city_avg_temp,
   global_data.global_avg_temp FROM global_data JOIN city_data ON
   global_data.year = city_data.year WHERE city like 'Berlin';
   ```

Now we have the option of downloading the file as CSV format. The file name is "results.csv"

Step 2: Creating a line chart using Python:

1. Using Pandas library, we will be able to manipulate the data.

```
# importing temperature data to a dataframe
df = pd.read_csv("results.csv")
# A sample of the data
df.head()
```

	year	city_avg_temp	global_avg_temp
0	1750	9.83	8.72
1	1751	9.75	7.98
2	1752	4.84	5.78
3	1753	8.72	8.39
4	1754	8.49	8.47

2. To have a more clear visualization, we have to calculate the moving average instead of the average. Moving averages are used to smooth out data to make it easier to observe long term trends and not get lost in daily fluctuations.

   global_mov_avg = df["global_avg_temp"].rolling(window = 150).mean()
   berlin_mov_avg = df["city_avg_temp"].rolling(window = 150).mean()

3. Visualizing data using Matplotlib, by creating a line chart that compares Berlin's temperatures with global temperatures.

```python
# The global moving average temperature of Berlin and the globe
plt.plot( df["year"], global_mov_avg, label = "Global")
plt.plot( df["year"], berlin_mov_avg, label = "Berlin")
plt.xlabel("Years")
plt.ylabel("Temperature (°C)")
plt.title("The Average Temperature of Berlin & the Globe")
plt.legend()
plt.show()
```
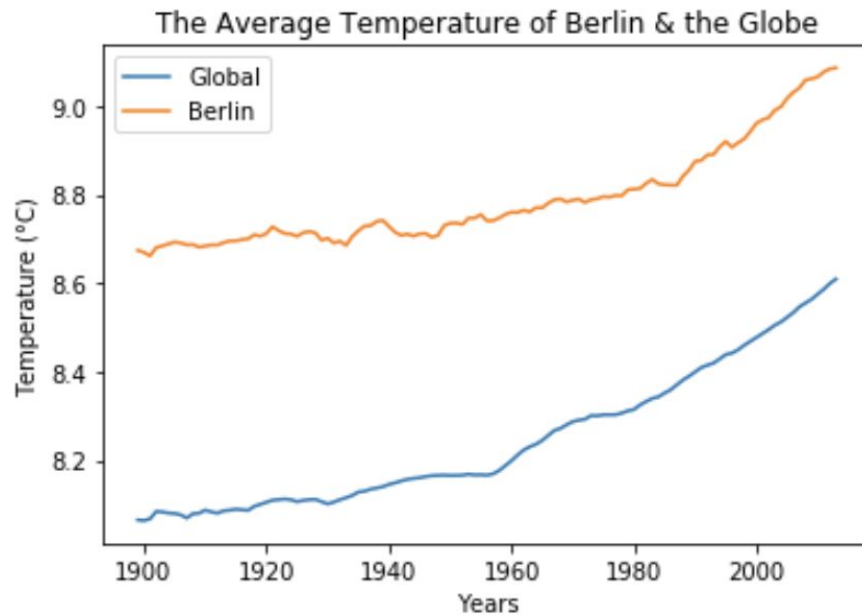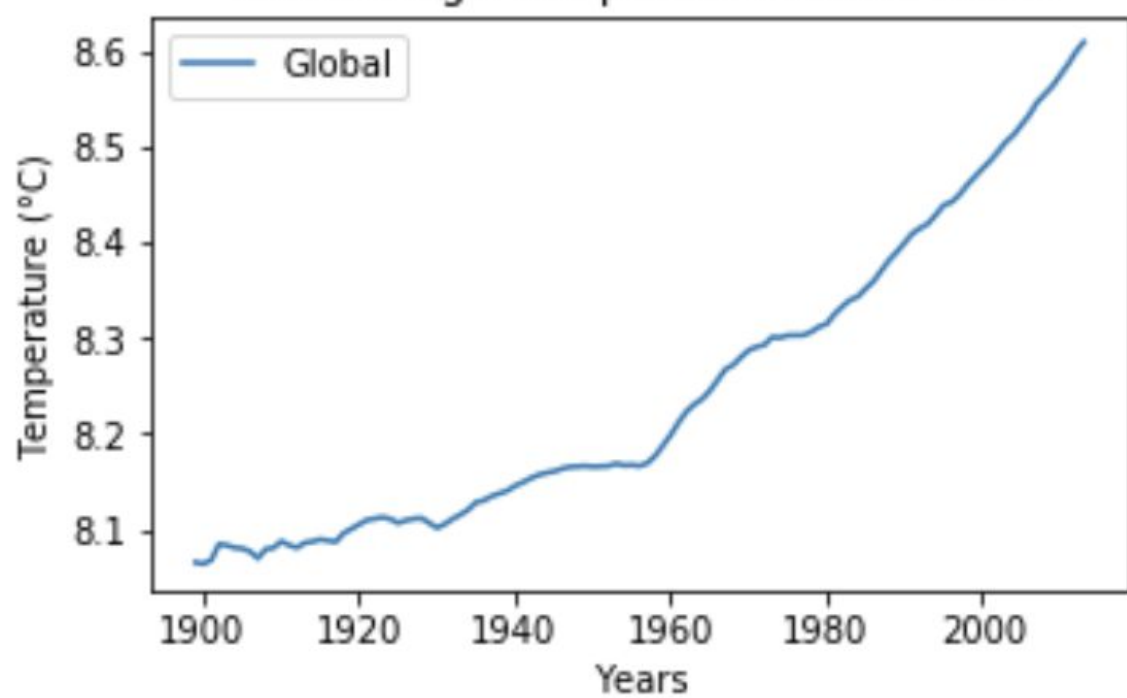

Plotting every chart separately.

The Average Temperature of Berlin
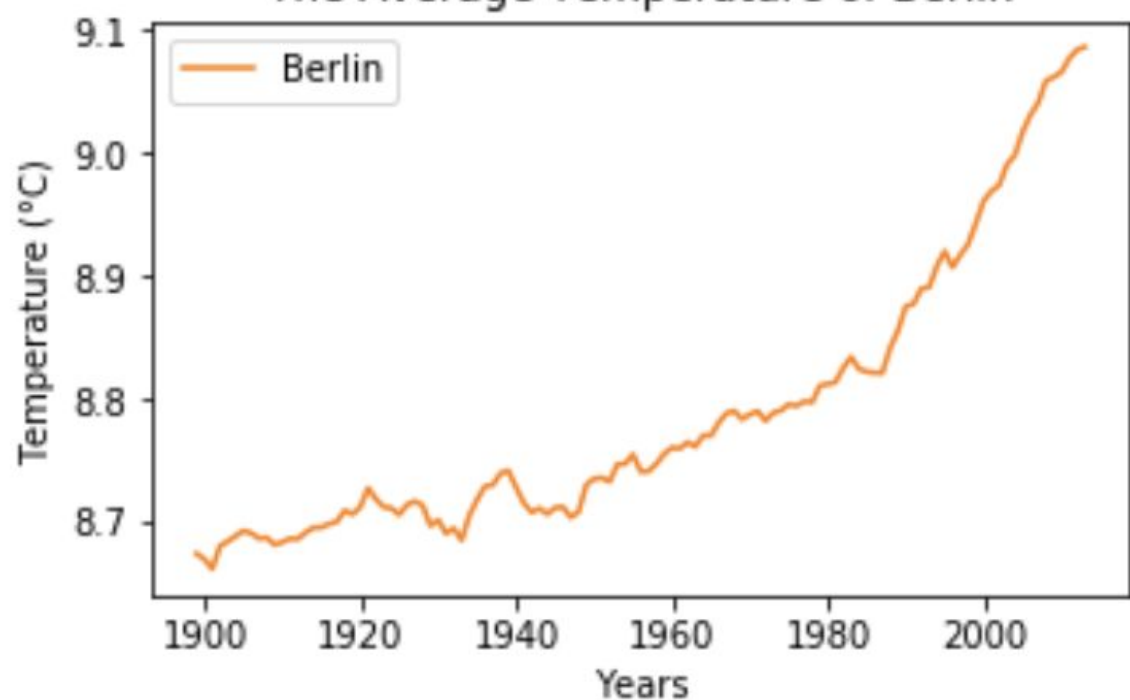
The Average Temperature of the Globe

Observations based on the line chart:

1. The temperature in Berlin is a little bit hotter than the global average temperature and the difference seems to be consistent over time.
2. The temperature is getting increased in both cases during the whole period.
3. There is a significant rise in the temperature in Berlin starting from the 1980s upwards.
4. The yearly global avg temperatures seem to be increasing rapidly starting from 1960. The world is getting much hotter year after year. There is an uptrend in the temperature rise.
5. The mean average temperature in Berlin between 1750 and 2013 is 8.917727 °C.

References:

1. https://www.investopedia.com/terms/m/movingaverage.asp
2. https://towardsdatascience.com/implementing-moving-averages-in-python-1ad28e636f9d