Platform: Udacity
Program: Data Analyst NanoDegree
Student: Laman Mammadli
Project: Exploring Weather Trends

Exploring Weather Trends

City Selection for Analysis

I am from Azerbaijan and now living in Poland. For these reason, I chose the capital cities of these countries. Berlin is the first city that I travelled, and Istanbul is a mysterious city for me. I intended to include these four cities (Baku, Berlin, Istanbul and Warsaw) to my analysis.

Accessing Data with SQL

Step 1. I use the SQL Workspace to extract data from the temperatures' database, then download the results to a CSV. Code is below:

SELECT cd.city, cd.year, cd.country, cd.avg_temp as avg_city_temp, gd.avg_temp as avg_global_temp
FROM city_data as cd
INNER JOIN global_data as gd
ON cd.year = gd.year
WHERE cd.country = 'Poland' OR cd.country = 'Germany' AND cd.city = 'Berlin' OR cd.country = 'Azerbaijan' AND cd.city = 'Baku' OR cd.country = 'Turkey' AND cd.city = 'Istanbul'
ORDER BY cd.city;

```
Input                                              HISTORY  ∨         MENU  ∨
  1   SELECT cd.city, cd.year, cd.country, cd.avg_temp as avg_city_temp, gd.avg_temp
      as avg_global_temp
  2   FROM city_data as cd
  3   INNER JOIN global_data as gd
  4   ON cd.year = gd.year
  5   WHERE cd.country = 'Poland' OR cd.country = 'Germany' AND cd.city = 'Berlin' OR
      cd.country = 'Azerbaijan' AND cd.city = 'Baku' OR cd.country = 'Turkey' AND
      cd.city = 'Istanbul'
```

Success!   EVALUATE

Output   998 results   ⬇ Download CSV

city	year	country	avg_city_temp	avg_global_temp
Baku	1808	Azerbaijan	5.50	7.63
Baku	1809	Azerbaijan	9.42	7.08
Baku	1810	Azerbaijan	9.47	6.92

```
In [2]:    import pandas as pd

           temp_df = pd.read_csv('temp_result_sql.csv')
           temp_df.head()
```

Out[2]:

	city	year	country	avg_city_temp	avg_global_temp
0	Baku	1808	Azerbaijan	5.50	7.63
1	Baku	1809	Azerbaijan	9.42	7.08
2	Baku	1810	Azerbaijan	9.47	6.92
3	Baku	1811	Azerbaijan	9.58	6.86
4	Baku	1812	Azerbaijan	9.71	7.05

```
In [3]:    temp_df.tail()
```

Out[3]:

	city	year	country	avg_city_temp	avg_global_temp
993	Warsaw	2009	Poland	8.50	9.51
994	Warsaw	2010	Poland	7.67	9.70
995	Warsaw	2011	Poland	8.84	9.52
996	Warsaw	2012	Poland	8.48	9.51
997	Warsaw	2013	Poland	9.32	9.61

```
In [4]:    temp_df.shape
```

Out[4]:   (998, 5)

```
In [5]:    temp_df.dtypes
```

```
Out[5]:    city                object
           year                 int64
           country             object
           avg_city_temp      float64
           avg_global_temp    float64
           dtype: object
```

```
In [7]:    temp_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 998 entries, 0 to 997
Data columns (total 5 columns):
 #   Column           Non-Null Count   Dtype
---  ------           --------------   -----
 0   city             998 non-null     object
 1   year             998 non-null     int64
 2   country          998 non-null     object
 3   avg_city_temp    998 non-null     float64
 4   avg_global_temp  998 non-null     float64
dtypes: float64(2), int64(1), object(2)
memory usage: 39.1+ KB

 3        False
 4        False
           ...
 993      False
 994      False
 995      False
 996      False
 997      False
Length: 998, dtype: bool
```

```
In [11]:   temp_df.duplicated().sum()
```

```
Out[11]:   0
```

Step 3. As seen from the result the data is clean. So, I downloaded 'CSV' file and opened it on Google Sheet.

Step 4. I calculated Moving Averages and created Line charts for four cities. You can see below: I calculated moving average temperature by decade.
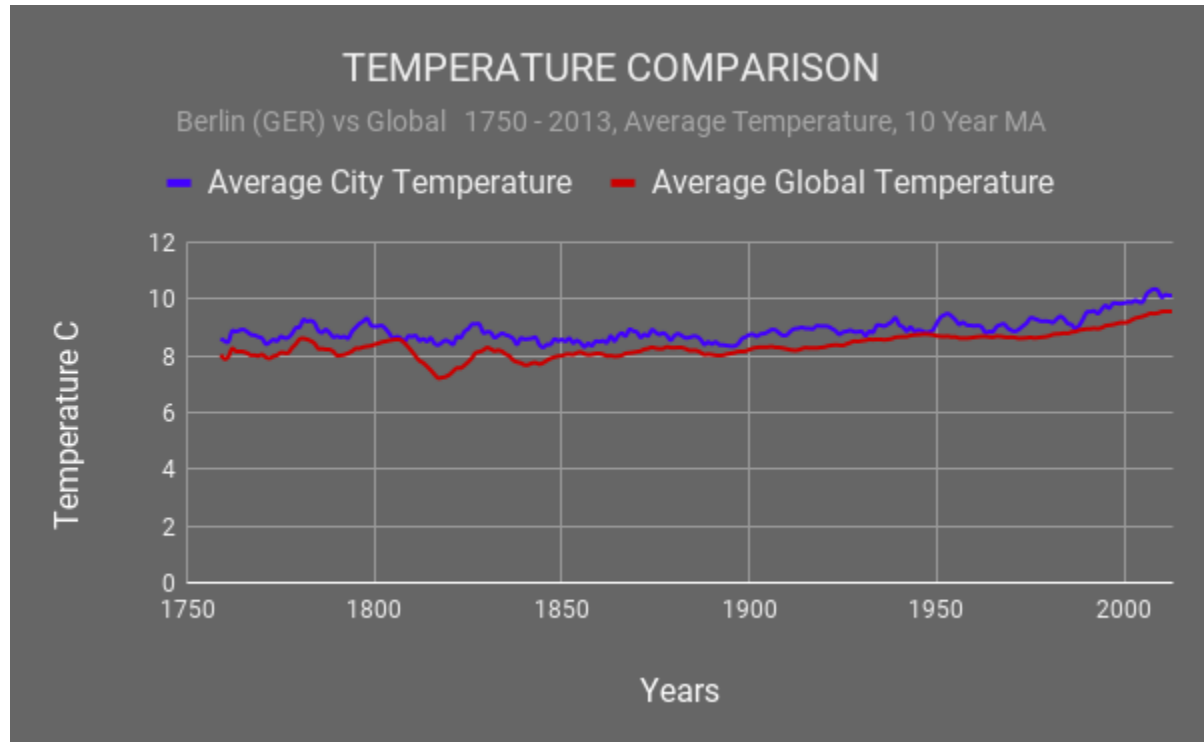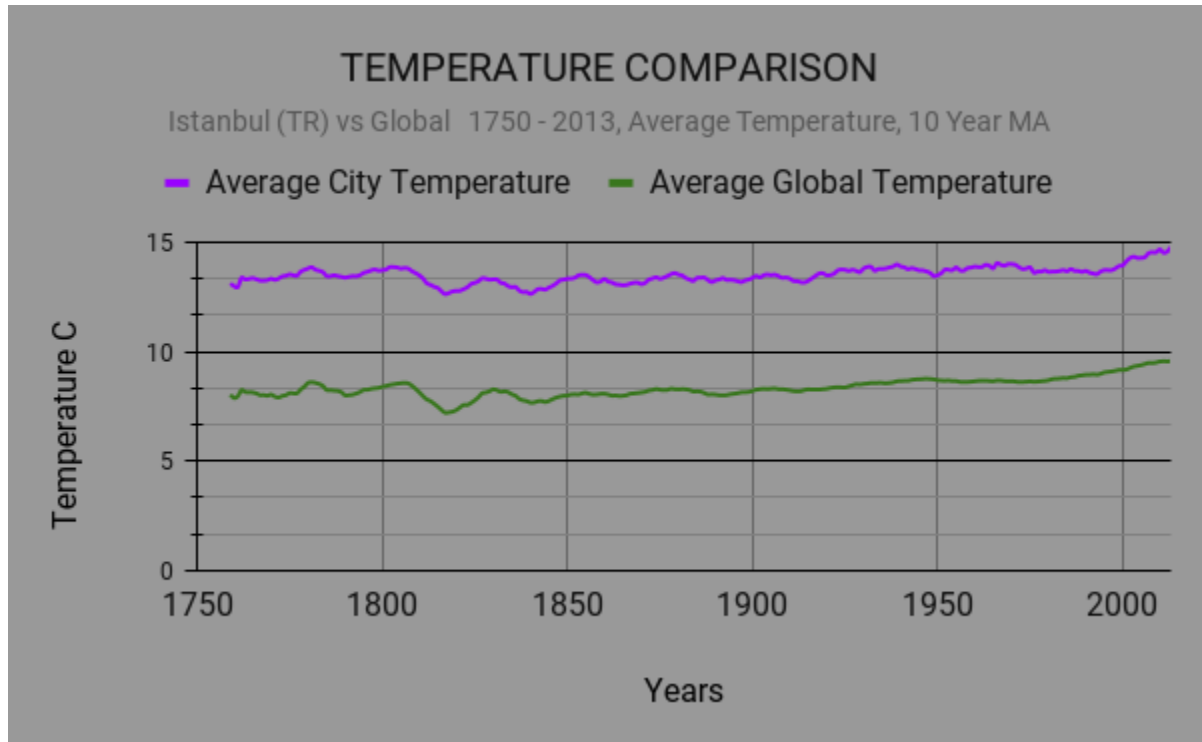
BAKU

Observation:

The line chart displays that average city and global temperature is going upward year by year. And the average city temperature is more than the global average temperature.

TEMPERTURE COMPARISON
Baku (AZE) vs Global   1808 - 2013, Average Temperature, 10 Year MA
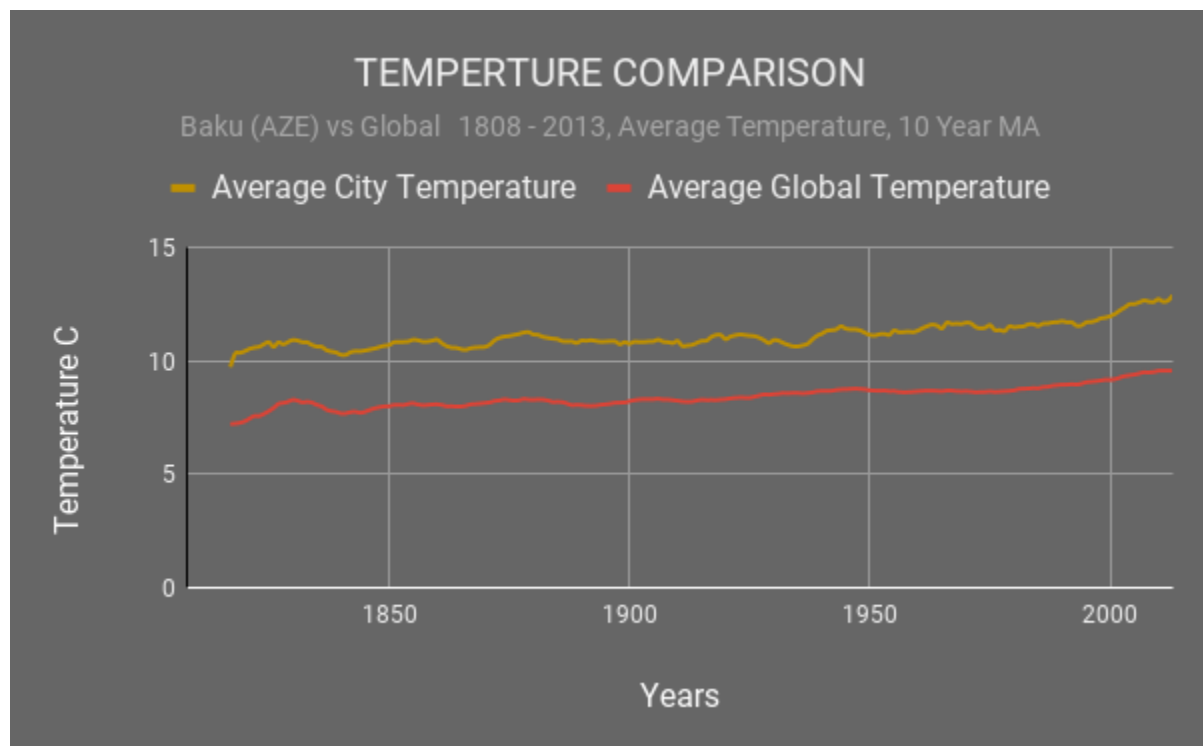
BERLIN

Observation:

The line chart displays that average city and global temperature is going upward year by year. And the average city temperature is more than the global average temperature. However, if we compare the result with previous one, we can see that the difference between the Berlin city average temperature and global average temperature is less than the difference between the Baku and global temperature.

ISTANBUL

Observation:

The line chart displays that average city and global temperature is going upward year by year. And the average city temperature is more than the global average temperature. The difference is bigger between Istanbul average temperature and average global temperature, like in Baku.

WARSAW

Observation:

The line chart displays that average city and global temperature is going upward year by year. And the average city temperature is less than the global average temperature, unlike the previous city results. The difference between the average city temperature and average global temperature is small, like in Berlin.
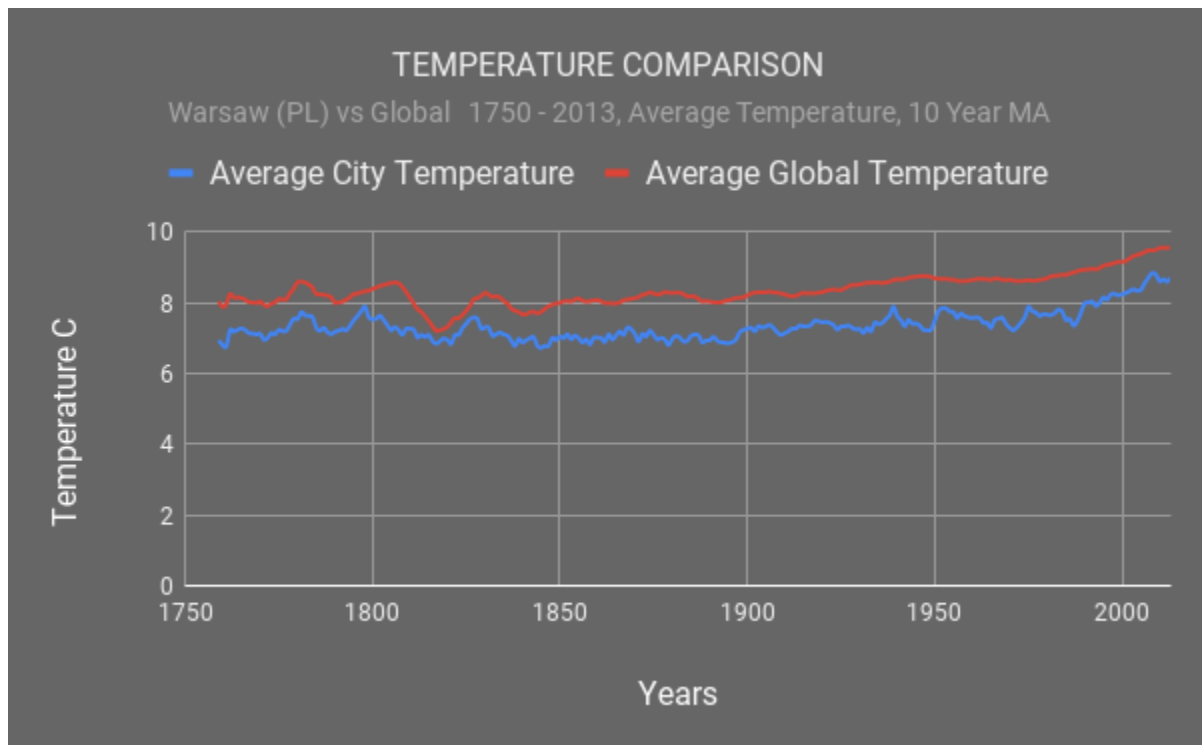
Step 5. I used 'matplotlib' library in Jupyter Notebook to see the correlation. As seen from the scatter plot below, there is a positive correlation, as average global temperature is increasing year by year, since 1900s.

```
In [14]: %matplotlib inline
         temp_df.plot(x = 'year', y = 'avg_global_temp', kind = 'scatter')
```

Out[14]: \<AxesSubplot:xlabel='year', ylabel='avg_global_temp'\>