Explore Weather Trends

By Edgar Eguía Calcáneo

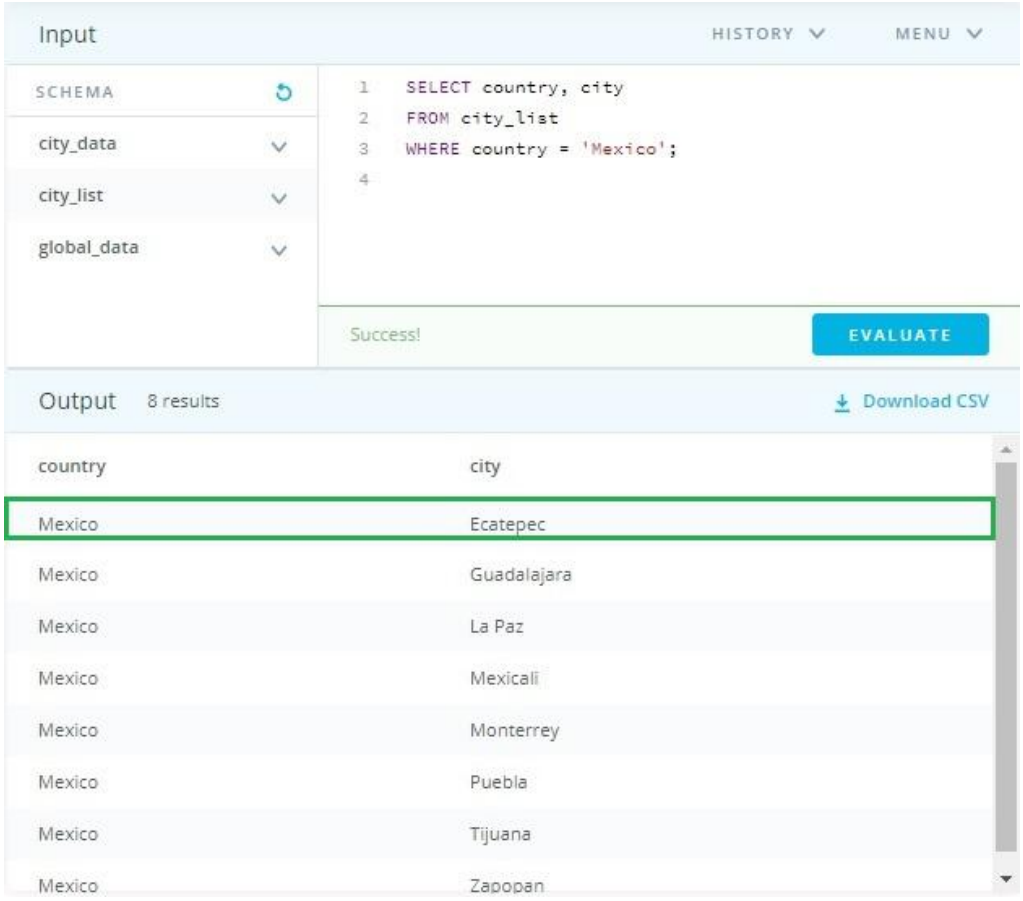
As stated in project requirements the objective of this document is to present the procedures, results and conclusions obtained through data analysis considering the information provided by Udacity database and data processing through tools such as SQL queries and Excel.

As first step information will be extracted from Udacity database, to get this done the next SQL queries are executed in Udacity SQL workspace.

This query is executed in order to find which city is the closest one to the city where I live.

```
SELECT country, city
FROM city_list
WHERE country = 'Mexico';
```

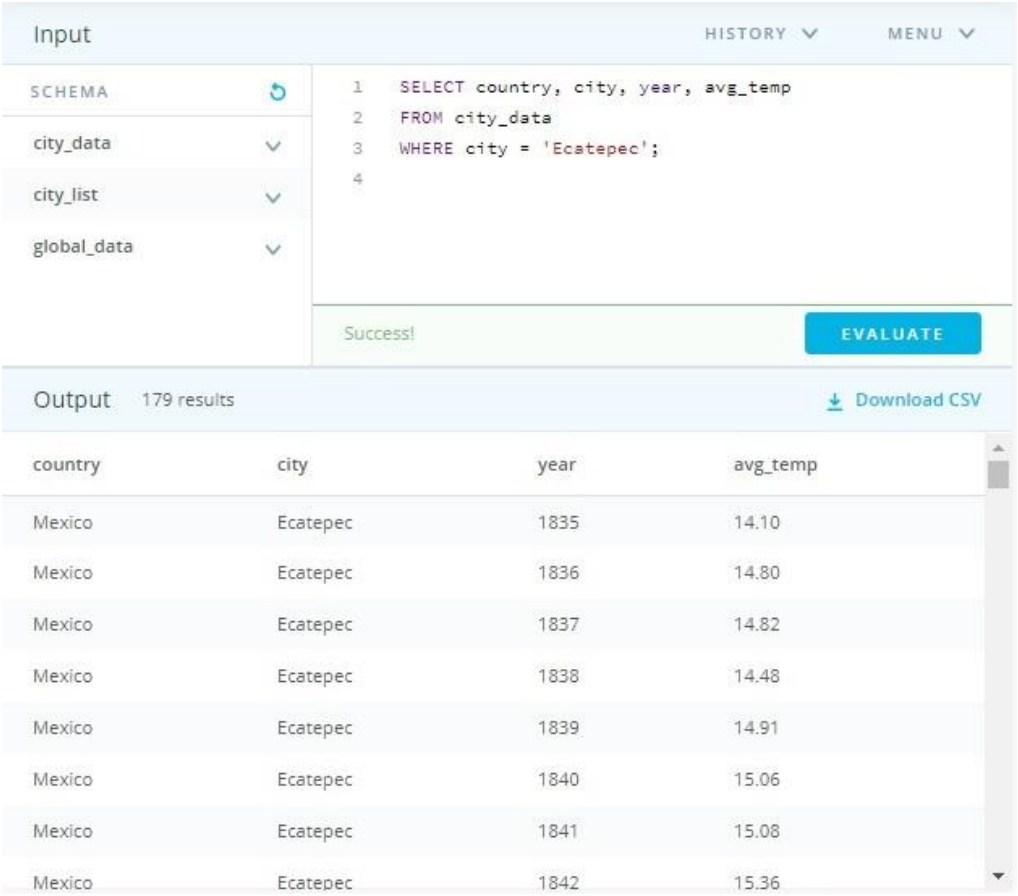
Udacity SQL workspace is shown in the following image.



This query is executed in order to find local (country: Mexico, city: Ecatepec) temperature from 1835 to 2013 (This period of time covers the whole set of data).

```
SELECT country, city, year, avg_temp
FROM city_data
WHERE city = 'Ecatepec';
```

Udacity SQL workspace is shown in the following image.

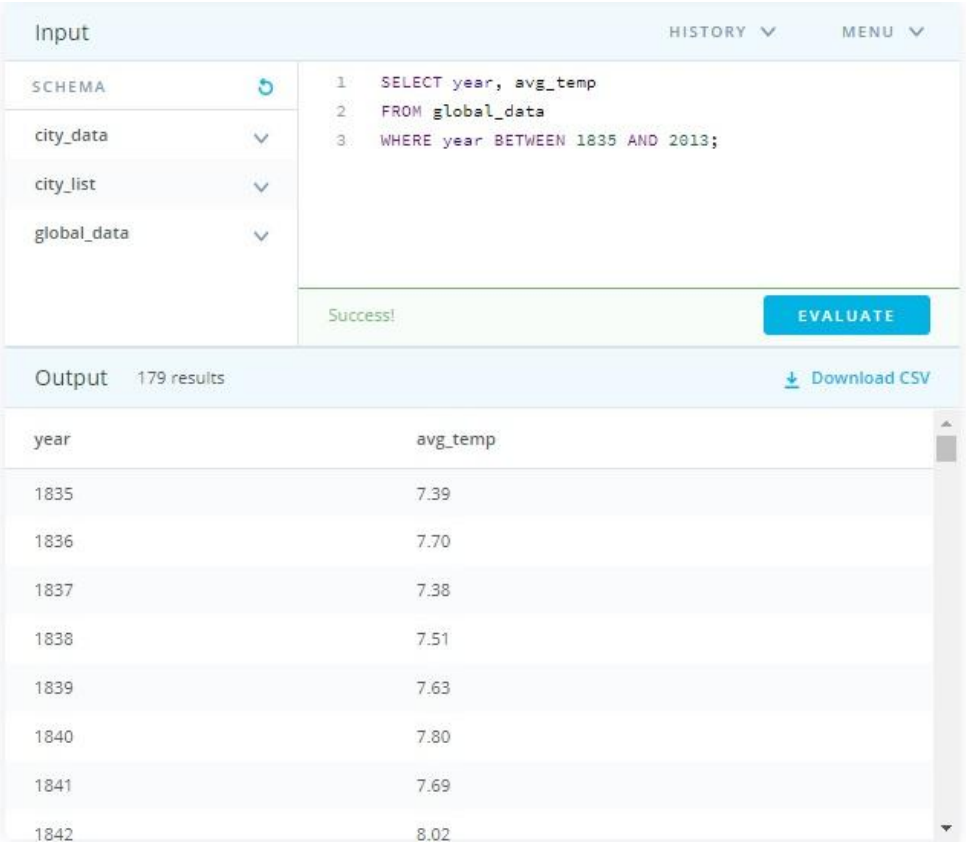


CSV file for local temperature is downloaded after SQL query is executed and information is verified through observation.

This query is executed in order to find global temperature from 1835 to 2013 (This period of time covers the subset of data which matches the set of data for local temperature).

```
SELECT year, avg_temp
FROM global_data
WHERE year BETWEEN 1835 AND 2013;
```

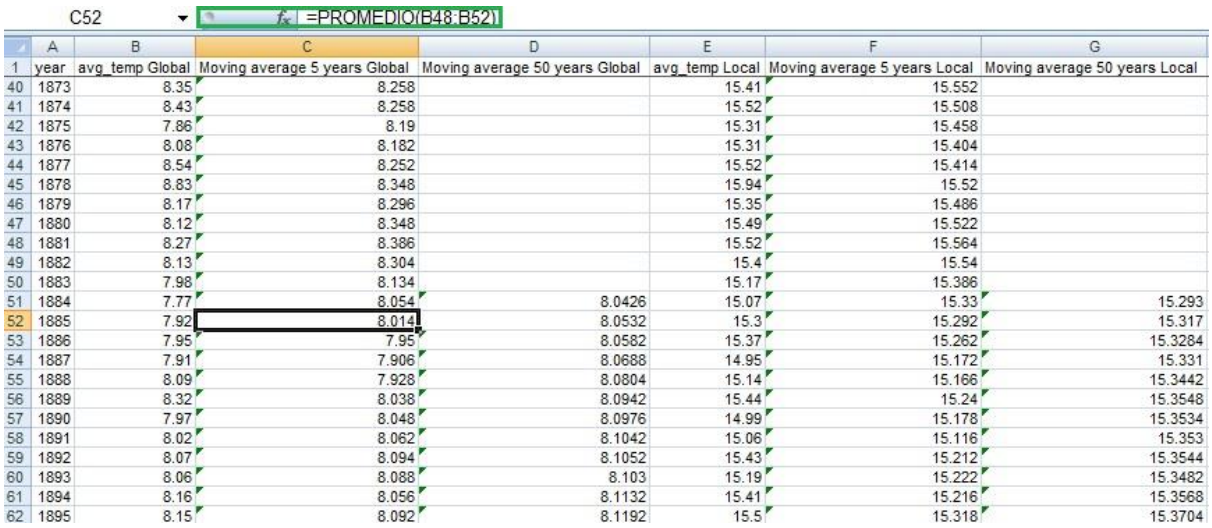
Udacity SQL workspace is shown in the following image.



CSV file for global temperature is downloaded after SQL query is executed and information is verified through observation.

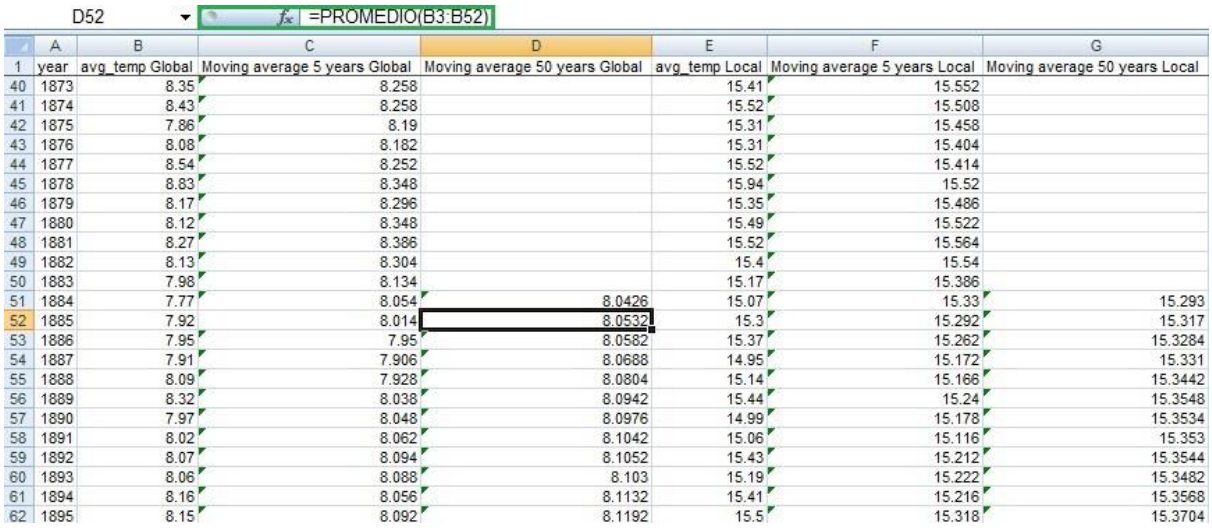
As second step CSV files information is put together in one Excel file in order to be able to apply average formulas to the information gathered. It's been considered to set 5 years moving average and 50 years moving average for both local and global temperatures so that conclusions can be made through chart behavior observations.

Excel sheet (part of it) is shown in the following image. Applied average formula (5 years moving average global) can be seen at the top of the image.


C52 =PROMEDIO(B48:B52)

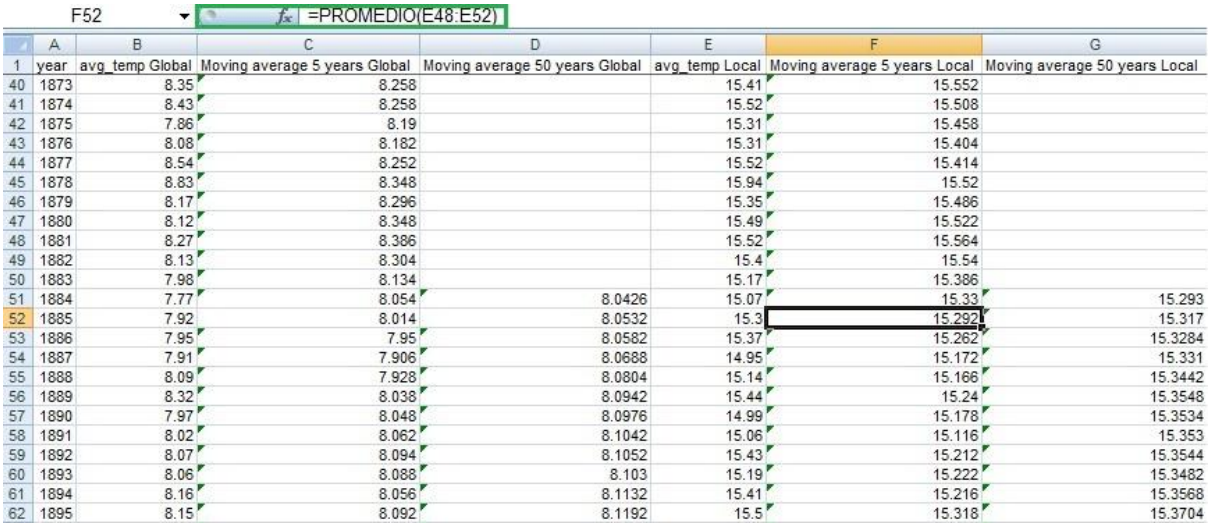
	A	B	C	D	E	F	G
1	year	avg_temp Global	Moving average 5 years Global	Moving average 50 years Global	avg_temp Local	Moving average 5 years Local	Moving average 50 years Local
40	1873	8.35	8.258		15.41	15.552	
41	1874	8.43	8.258		15.52	15.508	
42	1875	7.86	8.19		15.31	15.458	
43	1876	8.08	8.182		15.31	15.404	
44	1877	8.54	8.252		15.52	15.414	
45	1878	8.83	8.348		15.94	15.52	
46	1879	8.17	8.296		15.35	15.486	
47	1880	8.12	8.348		15.49	15.522	
48	1881	8.27	8.386		15.52	15.564	
49	1882	8.13	8.304		15.4	15.54	
50	1883	7.98	8.134		15.17	15.386	
51	1884	7.77	8.054	8.0426	15.07	15.33	15.293
52	1885	7.92	8.014	8.0532	15.3	15.292	15.317
53	1886	7.95	7.95	8.0582	15.37	15.262	15.3284
54	1887	7.91	7.906	8.0688	14.95	15.172	15.331
55	1888	8.09	7.928	8.0804	15.14	15.166	15.3442
56	1889	8.32	8.038	8.0942	15.44	15.24	15.3548
57	1890	7.97	8.048	8.0976	14.99	15.178	15.3534
58	1891	8.02	8.062	8.1042	15.06	15.116	15.353
59	1892	8.07	8.094	8.1052	15.43	15.212	15.3544
60	1893	8.06	8.088	8.103	15.19	15.222	15.3482
61	1894	8.16	8.056	8.1132	15.41	15.216	15.3568
62	1895	8.15	8.092	8.1192	15.5	15.318	15.3704

Excel sheet (part of it) is shown in the following image. Applied average formula (50 years moving average global) can be seen at the top of the image.


D52 =PROMEDIO(B3:B52)

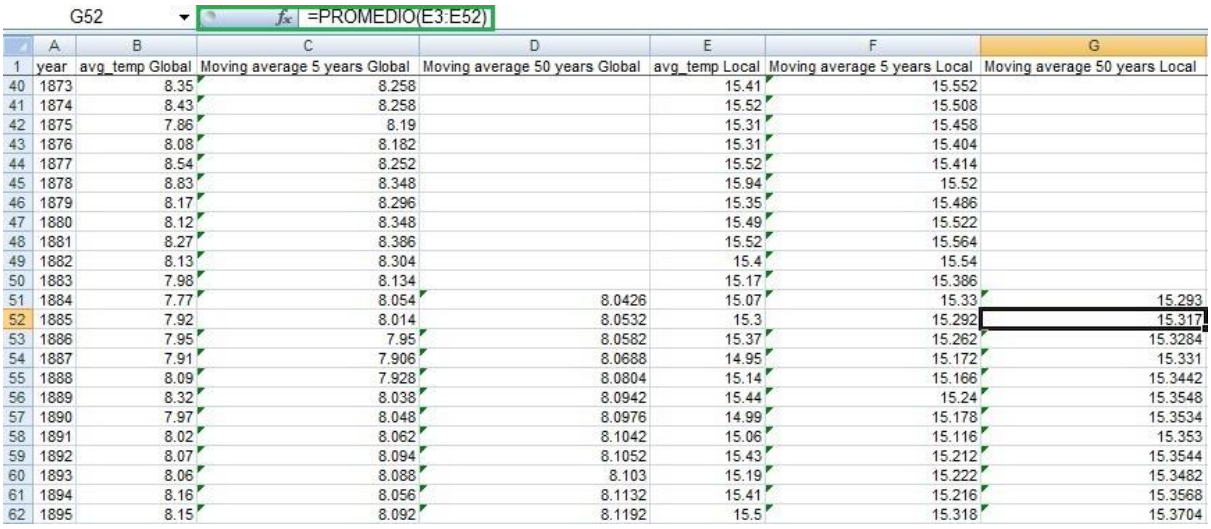
	A	B	C	D	E	F	G
1	year	avg_temp Global	Moving average 5 years Global	Moving average 50 years Global	avg_temp Local	Moving average 5 years Local	Moving average 50 years Local
40	1873	8.35	8.258		15.41	15.552	
41	1874	8.43	8.258		15.52	15.508	
42	1875	7.86	8.19		15.31	15.458	
43	1876	8.08	8.182		15.31	15.404	
44	1877	8.54	8.252		15.52	15.414	
45	1878	8.83	8.348		15.94	15.52	
46	1879	8.17	8.296		15.35	15.486	
47	1880	8.12	8.348		15.49	15.522	
48	1881	8.27	8.386		15.52	15.564	
49	1882	8.13	8.304		15.4	15.54	
50	1883	7.98	8.134		15.17	15.386	
51	1884	7.77	8.054	8.0426	15.07	15.33	15.293
52	1885	7.92	8.014	8.0532	15.3	15.292	15.317
53	1886	7.95	7.95	8.0582	15.37	15.262	15.3284
54	1887	7.91	7.906	8.0688	14.95	15.172	15.331
55	1888	8.09	7.928	8.0804	15.14	15.166	15.3442
56	1889	8.32	8.038	8.0942	15.44	15.24	15.3548
57	1890	7.97	8.048	8.0976	14.99	15.178	15.3534
58	1891	8.02	8.062	8.1042	15.06	15.116	15.353
59	1892	8.07	8.094	8.1052	15.43	15.212	15.3544
60	1893	8.06	8.088	8.103	15.19	15.222	15.3482
61	1894	8.16	8.056	8.1132	15.41	15.216	15.3568
62	1895	8.15	8.092	8.1192	15.5	15.318	15.3704

Excel sheet (part of it) is shown in the following image. Applied average formula (5 years moving average local) can be seen at the top of the image.


F52 =PROMEDIO(E48:E52)

	A	B	C	D	E	F	G
1	year	avg_temp Global	Moving average 5 years Global	Moving average 50 years Global	avg_temp Local	Moving average 5 years Local	Moving average 50 years Local
40	1873	8.35	8.258		15.41	15.552	
41	1874	8.43	8.258		15.52	15.508	
42	1875	7.86	8.19		15.31	15.458	
43	1876	8.08	8.182		15.31	15.404	
44	1877	8.54	8.252		15.52	15.414	
45	1878	8.83	8.348		15.94	15.52	
46	1879	8.17	8.296		15.35	15.486	
47	1880	8.12	8.348		15.49	15.522	
48	1881	8.27	8.386		15.52	15.564	
49	1882	8.13	8.304		15.4	15.54	
50	1883	7.98	8.134		15.17	15.386	
51	1884	7.77	8.054	8.0426	15.07	15.33	15.293
52	1885	7.92	8.014	8.0532	15.3	15.292	15.317
53	1886	7.95	7.95	8.0582	15.37	15.262	15.3284
54	1887	7.91	7.906	8.0688	14.95	15.172	15.331
55	1888	8.09	7.928	8.0804	15.14	15.166	15.3442
56	1889	8.32	8.038	8.0942	15.44	15.24	15.3548
57	1890	7.97	8.048	8.0976	14.99	15.178	15.3534
58	1891	8.02	8.062	8.1042	15.06	15.116	15.353
59	1892	8.07	8.094	8.1052	15.43	15.212	15.3544
60	1893	8.06	8.088	8.103	15.19	15.222	15.3482
61	1894	8.16	8.056	8.1132	15.41	15.216	15.3568
62	1895	8.15	8.092	8.1192	15.5	15.318	15.3704

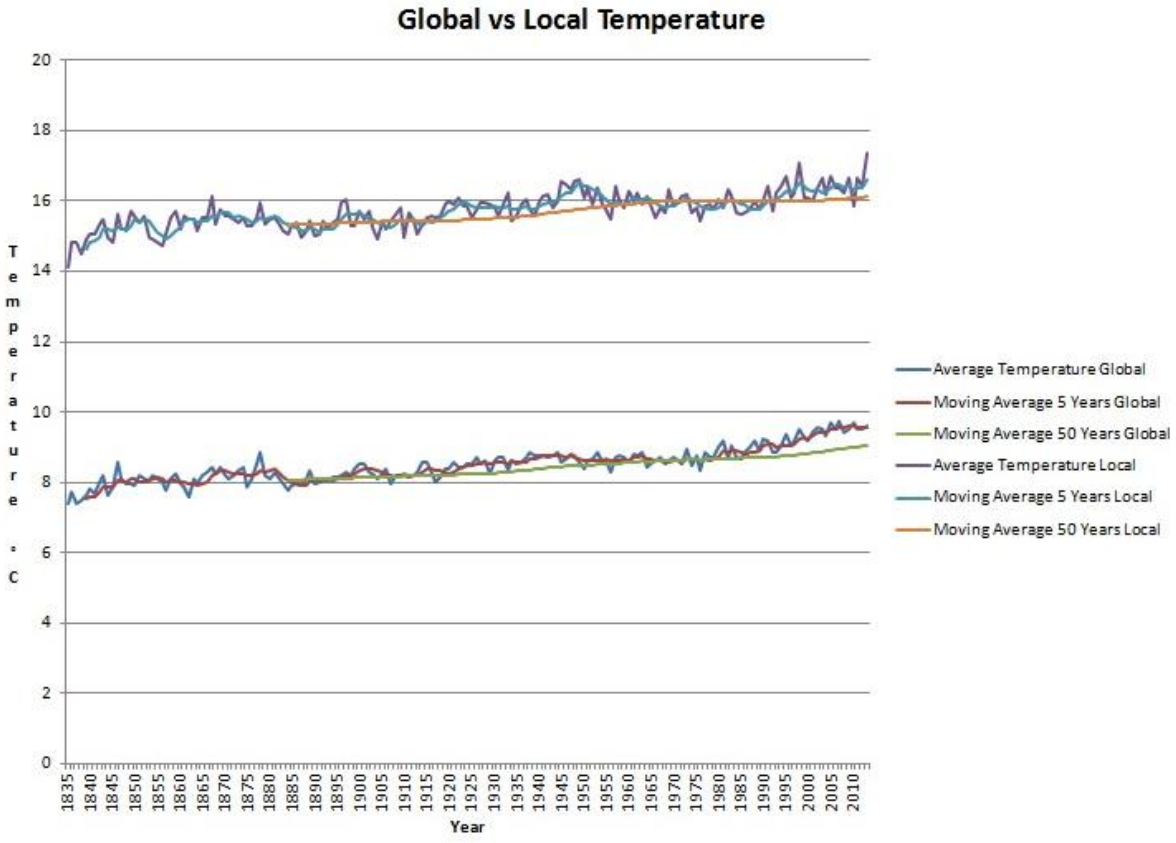
Excel sheet (part of it) is shown in the following image. Applied average formula (50 years moving average local) can be seen at the top of the image.


G52 =PROMEDIO(E3:E52)

	A	B	C	D	E	F	G
1	year	avg_temp Global	Moving average 5 years Global	Moving average 50 years Global	avg_temp Local	Moving average 5 years Local	Moving average 50 years Local
40	1873	8.35	8.258		15.41	15.552	
41	1874	8.43	8.258		15.52	15.508	
42	1875	7.86	8.19		15.31	15.458	
43	1876	8.08	8.182		15.31	15.404	
44	1877	8.54	8.252		15.52	15.414	
45	1878	8.83	8.348		15.94	15.52	
46	1879	8.17	8.296		15.35	15.486	
47	1880	8.12	8.348		15.49	15.522	
48	1881	8.27	8.386		15.52	15.564	
49	1882	8.13	8.304		15.4	15.54	
50	1883	7.98	8.134		15.17	15.386	
51	1884	7.77	8.054	8.0426	15.07	15.33	15.293
52	1885	7.92	8.014	8.0532	15.3	15.292	15.317
53	1886	7.95	7.95	8.0582	15.37	15.262	15.3284
54	1887	7.91	7.906	8.0688	14.95	15.172	15.331
55	1888	8.09	7.928	8.0804	15.14	15.166	15.3442
56	1889	8.32	8.038	8.0942	15.44	15.24	15.3548
57	1890	7.97	8.048	8.0976	14.99	15.178	15.3534
58	1891	8.02	8.062	8.1042	15.06	15.116	15.353
59	1892	8.07	8.094	8.1052	15.43	15.212	15.3544
60	1893	8.06	8.088	8.103	15.19	15.222	15.3482
61	1894	8.16	8.056	8.1132	15.41	15.216	15.3568
62	1895	8.15	8.092	8.1192	15.5	15.318	15.3704

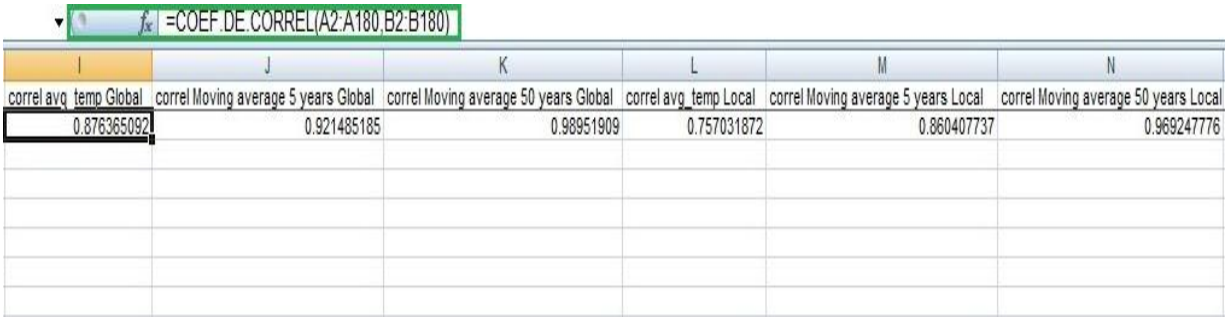
As third step a line chart (Temperature-Time) is plotted considering average global values, moving average 5 years global values, moving average 50 years global values, average local values, moving average 5 years local values, moving average 50 years local values (Excel sheet shown above).

Line chart presents the behavior (average) and trends (moving averages) of local and global temperature through time from 1835 to 2013.



As a fourth step correlation coefficients are computed in order to determine the degree to which temperature (averages, moving averages) and time variables are associated.

Excel sheet is shown in the following image. Applied correlation coefficient formula can be seen at the top of the image. Formula has considered as first argument year column values and as second argument avg_temp Global column values, Moving average 5 years Global column values, Moving average 50 years Global column values, avg_temp Local column values, Moving average 5 years Local column values, Moving average 50 years Local column values.



Finally observations and conclusions are made and listed as follows:

1. Ecatepec (local site) has always been hotter compared to the global average according to the chart (see Average Temperature Global, Average Temperature Local). Ecatepec minimum average temperatures are close to 14 °C in the 1830's and maximum average temperatures are close to 18 °C in the current decade, this means that Ecatepec has increased its temperature in 4 °C over the last 180 years (see Average Temperature Local).
2. Global minimum average temperatures are close to 8 °C in the 1830's and maximum average temperatures are close to 10 °C in the current decade, this means that the world has increased its temperature in 2 °C over the last 180 years (see Average Temperature Global).
3. Local or global average temperature plots don't allow the observer to determine trends easily, that's why moving average is being applied; 5 and 50 years periods are considered to see how the plots are smoothed (this way it will be possible to see if temperature is increasing or decreasing over time). It is clear that the longer the period is for the moving average the smoother the plot gets revealing that moving average is useful for forecast long-term trends (see Moving Average 50 Years Global, Moving Average 50 Years Local).
4. According to local and global moving average Ecatepec and the world have become hotter and the trends show that temperature will keep increasing over time. This conclusion has been reached in observation 1 and 2 but considering minimum and maximum values which are set at the beginning and at the end of time. Now, moving average is showing us clearly what is happening in the mean time and the plots show that over the last 180 years temperature has been increasing constantly (see Moving Average 50 Years Global, Moving Average 50 Years Local).
5. Correlation coefficients indicate strong positive relationship (cc > 0.75) between variables (temperature-time), meaning that for each year that goes by there is in general and increment in temperature. For this particular analysis correlation coefficients increase their values as moving average period is considered longer (comparison between 5 years period to 50 years period) leading us to the conclusion that Ecatepec and the world's temperature will keep increasing in the future (short and long term).
6. Answering the question posted "Can you estimate the average temperature in your city (Ecatepec) based on the average global temperature?" in project specification web page my answer is yes. According to line chart there is an almost constant difference between local and global values for each year (considering moving average 50 years to smooth plots which already includes average values in the calculation) leading us to think that this difference will be maintained in the years coming implying that local average temperature is predictable.