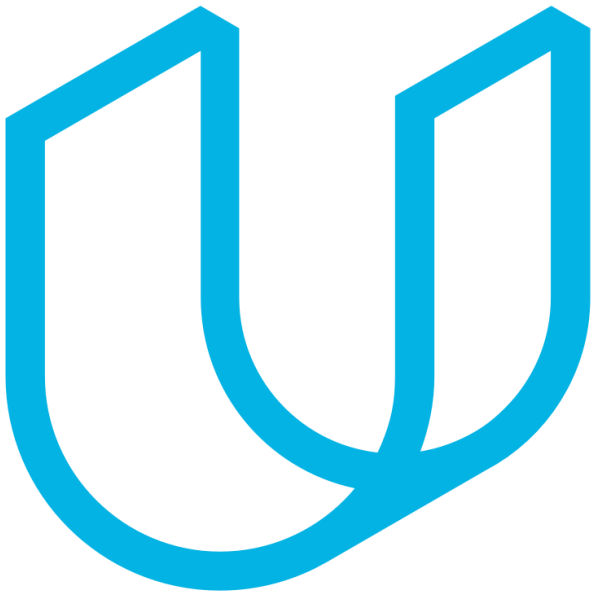
Project No 1:

Explore Weather Trends



UDACITY

Data Analyst Nanodegree Program

Submitted by:
Ali Salman

Project 1 – Explore Weather Trends

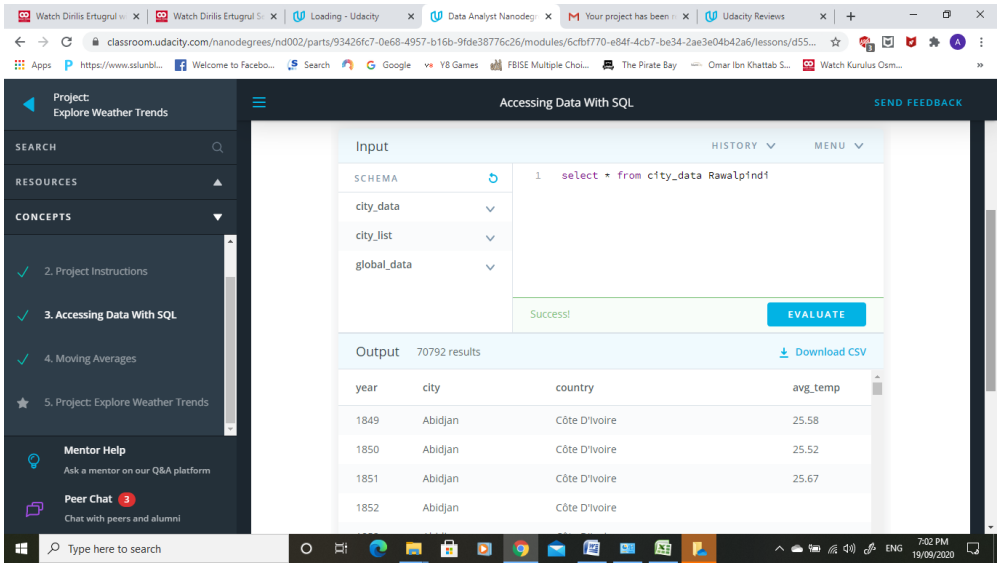
1. **Steps**

   a. **Exploring the Data**

      (1) I had to get to help in order to learn to use SQL workspace to extract data.

      (2) In order to open the list of cities I wrote ***Select * from City_list*** in my SQL workspace and pressed the button ***Evaluate***.

      (3) In order to open the data regarding average temperatures of cities I wrote ***Select * from City_data Rawalpindi*** in my SQL workspace and pressed the button ***Evaluate***. It opened the same complete file, instead of giving me the data of Rawalpindi only. However this time it was in alphabetic order with reference to city column. Screen shot of my SQL workspace is as under:-



      (4) In order to open the data regarding average temperatures of the globe I wrote ***Select * from global_data*** in my SQL workspace and pressed the button ***Evaluate***.

      (5) After opening each list I simply downloaded them and thanks that they were in Excel format, with which I am well conversant.

   b. Once the tables were available in Excel format, I applied Filter to separate the data of the city closest to my location.

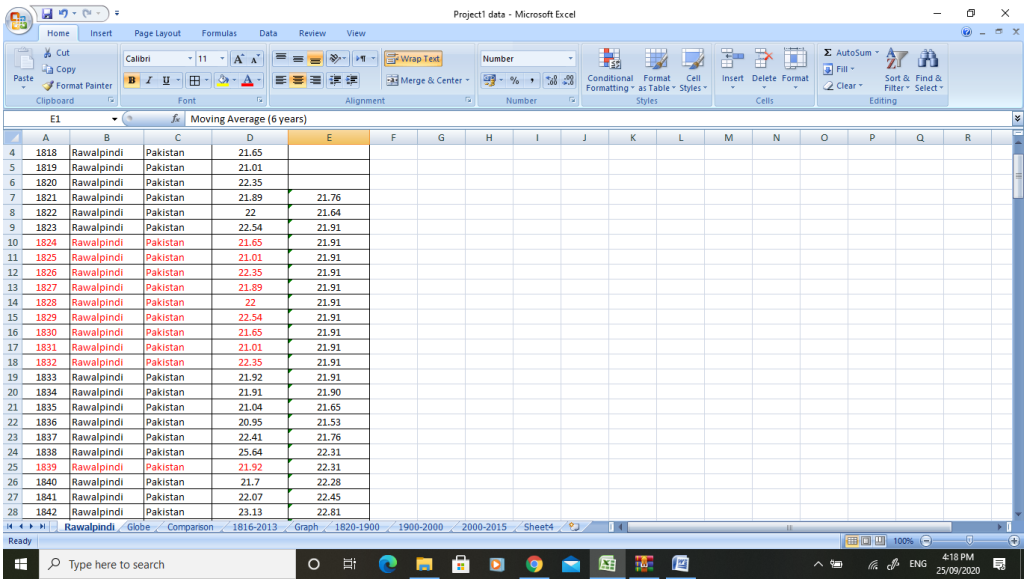
   c. The data of average temperatures of my city **"Rawalpindi" is of 198 years**, which is from 1816 till 2013.

d. There are certain years in which the data is missing and those years are 20, which means that the available data is of 178 years out of 198 years.

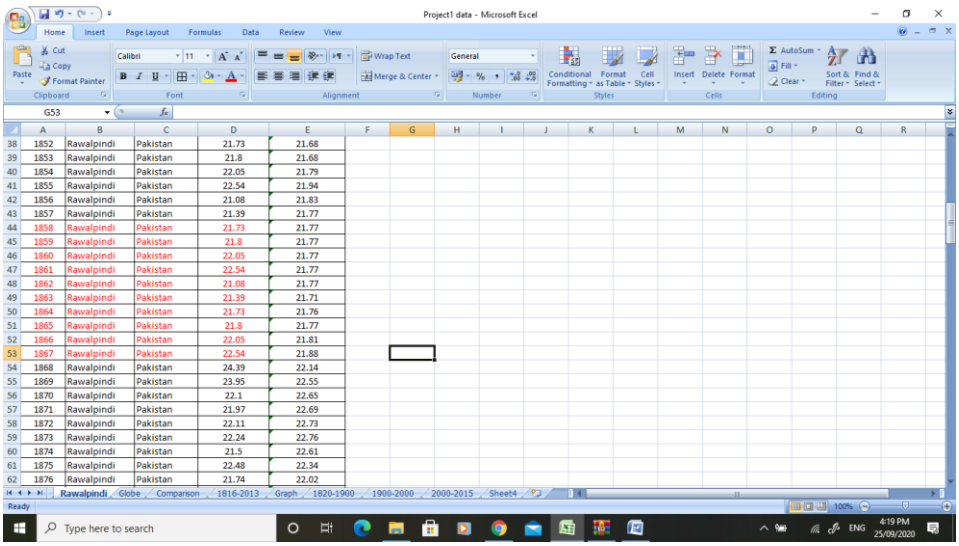
e. From year 1816 till 1823 **(8 years)** the data was available and after that from 1824 till 1832 **(9 years)** the data was missing. Details are as under:-

Serial	Years	Total Years	Remarks
(1)	1816 – 1823	8	Data was **available**
(2)	1824 – 1832	9	Data was **missing**
(3)	1833 – 1838	6	Data was **available**
(4)	1839	1	Data was **missing**
(5)	1840 - 1857	18	Data was **available**
(6)	1858 - 1867	10	Data was **missing**
(7)	1868 - 2013	146	Data was **available**
Total		**198**	Data of **20 years was missing** Data of **178 years was available**

f. Screen shots of predicted data from years 1824 – 1832, as well as 1839 is as under:-



g. Screen shots of predicted data from years 1858 – 1867 is as under:-



	A	B	C	D	E
38	1852	Rawalpindi	Pakistan	21.73	21.68
39	1853	Rawalpindi	Pakistan	21.8	21.68
40	1854	Rawalpindi	Pakistan	22.05	21.79
41	1855	Rawalpindi	Pakistan	22.54	21.94
42	1856	Rawalpindi	Pakistan	21.08	21.83
43	1857	Rawalpindi	Pakistan	21.39	21.77
44	1858	Rawalpindi	Pakistan	21.73	21.77
45	1859	Rawalpindi	Pakistan	21.8	21.77
46	1860	Rawalpindi	Pakistan	22.05	21.77
47	1861	Rawalpindi	Pakistan	22.54	21.77
48	1862	Rawalpindi	Pakistan	21.08	21.77
49	1863	Rawalpindi	Pakistan	21.39	21.71
50	1864	Rawalpindi	Pakistan	21.73	21.76
51	1865	Rawalpindi	Pakistan	21.8	21.77
52	1866	Rawalpindi	Pakistan	22.05	21.81
53	1867	Rawalpindi	Pakistan	22.54	21.88
54	1868	Rawalpindi	Pakistan	24.39	22.14
55	1869	Rawalpindi	Pakistan	23.95	22.55
56	1870	Rawalpindi	Pakistan	22.1	22.65
57	1871	Rawalpindi	Pakistan	21.97	22.69
58	1872	Rawalpindi	Pakistan	22.11	22.73
59	1873	Rawalpindi	Pakistan	22.24	22.76
60	1874	Rawalpindi	Pakistan	21.5	22.61
61	1875	Rawalpindi	Pakistan	22.48	22.34
62	1876	Rawalpindi	Pakistan	21.74	22.02

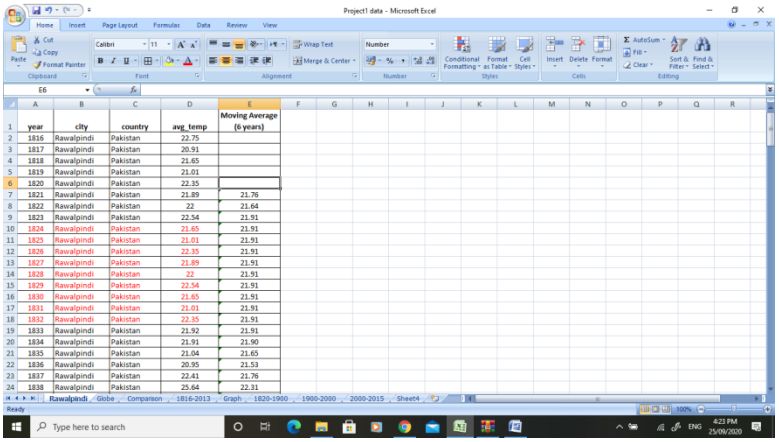
h. I had to predict the data basing on the average of first 8 years, hence the data of temperatures of each year was predicted on the basis of its previous 8 years.

i. The temperatures of globe are given from year 1750 till 2015, whereas the temperatures of my city “Rawalpindi” are given from 1816 till 2013. Therefore I have taken the global temperatures of the same time period i.e, 1816 – 2013.

j. Then I gathered both the data sets in one sheet of Excel.

k. **Moving Average**. Since the total values are 198, which is divisible by 6 hence the **moving average** has been calculated on the basis of each 6 years for both global data set as well as data set of the city.

l. In order to smooth out the noise by considering the data of all years for calculation of Moving Average, I have calculated the moving averages for all successive years (obviously not for first five years) so that the plotted graph is smoother.



	A	B	C	D	E
1	year	city	country	avg_temp	Moving Average (6 years)
2	1816	Rawalpindi	Pakistan	22.75	
3	1817	Rawalpindi	Pakistan	20.91	
4	1818	Rawalpindi	Pakistan	21.65	
5	1819	Rawalpindi	Pakistan	21.01	
6	1820	Rawalpindi	Pakistan	22.35	
7	1821	Rawalpindi	Pakistan	21.89	21.76
8	1822	Rawalpindi	Pakistan	22	21.64
9	1823	Rawalpindi	Pakistan	22.54	21.91
10	1824	Rawalpindi	Pakistan	21.65	21.91
11	1825	Rawalpindi	Pakistan	21.01	21.91
12	1826	Rawalpindi	Pakistan	22.35	21.91
13	1827	Rawalpindi	Pakistan	21.89	21.91
14	1828	Rawalpindi	Pakistan	22	21.91
15	1829	Rawalpindi	Pakistan	22.54	21.91
16	1830	Rawalpindi	Pakistan	21.65	21.91
17	1831	Rawalpindi	Pakistan	21.01	21.91
18	1832	Rawalpindi	Pakistan	22.35	21.91
19	1833	Rawalpindi	Pakistan	21.92	21.91
20	1834	Rawalpindi	Pakistan	21.91	21.90
21	1835	Rawalpindi	Pakistan	21.04	21.65
22	1836	Rawalpindi	Pakistan	20.95	21.53
23	1837	Rawalpindi	Pakistan	22.41	21.76
24	1838	Rawalpindi	Pakistan	25.64	22.31

2. **Line Chart**

a. I took help from Help Menu and after watching a video on youtube, I was able to make the line graph.

b. **Horizontal Axis (x-axis)**. Years have been taken on x-axis while keeping the major unit of 10 and starting the graph from 1820, as the first moving average is calculated in year 1821 and the graph goes till 2015.

c. **Vertical Axis (y-axis)**. Average temperatures in ^{o}C is taken on y-axis while keeping the major unit of 1.0 ^{o}C and starting the graph from 7.0 ^{o}C, as the lowest value of temperatures in 7.47 ^{o}C and the graph goes till 24.0 ^{o}C.

d. I added vertical as well as horizontal grid lines on my graph to make critical observations, which were removed later.

3. **Observations**

a. **City Temperatures**. In order to analyze the temperatures trends in **"Rawalpindi"**, I have divided the moving averages of temperatures into three brackets, i.e, **1820 – 1900**, **1901 – 2000** and **2000 – 2013**.

(1) **1820 – 1900**

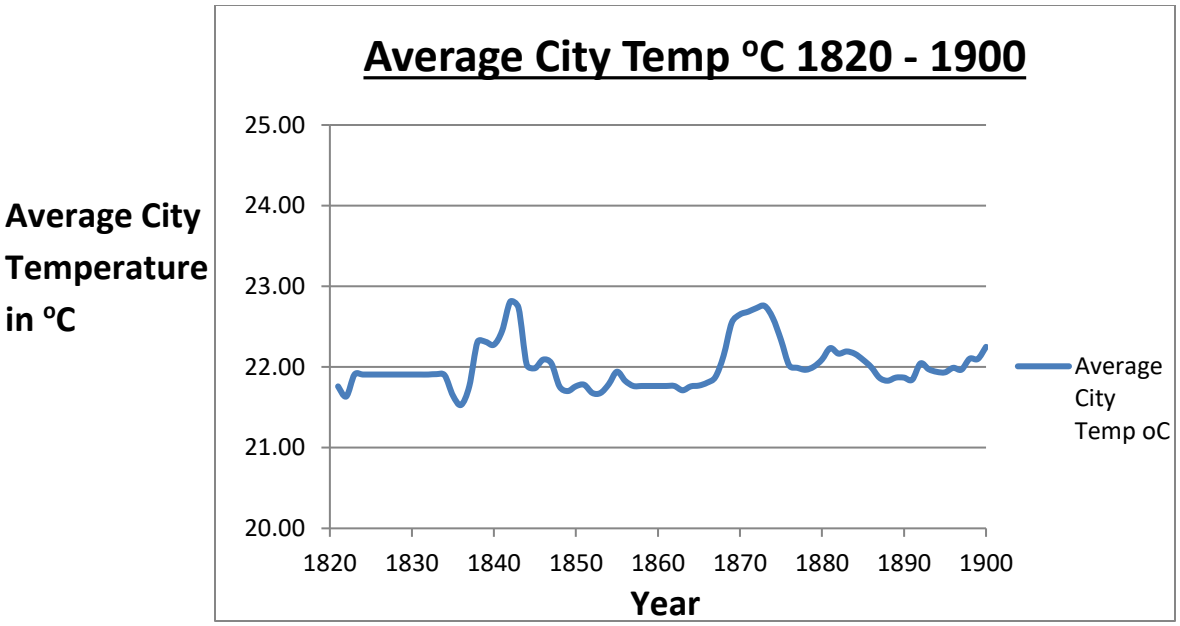
i. Previously I had made this graph, just by taking 14 values, as the moving averages of each sixth year was taken. Now I have taken **81 values by calculating the moving average on the basis of six years but for each successive year**.

ii. Average temperatures in this bracket is **22.01 ^{o}C**.

iii. There are total **81** reading in this bracket.

iv. Out of 81 reading **53 reading are below** the average line and **28 readings are above** the average line.

v. First reading of temperature in **1820 is 21.76 ^{o}C**, which rose to **22.25 ^{o}C in year 1900**, which shows a **clear rise in temperature over the century**, despite the fact that the **trend was not linear**, as the spikes are visible in the decades of 1840 – 1850 and later in 1870 - 1880.



(2) **1901 – 1999**

i. Previously I had made this graph, just by taking 16 values, as the moving averages of each sixth year was taken. Now I have taken **99 values by calculating the moving average on the basis of six years but for each successive year**.

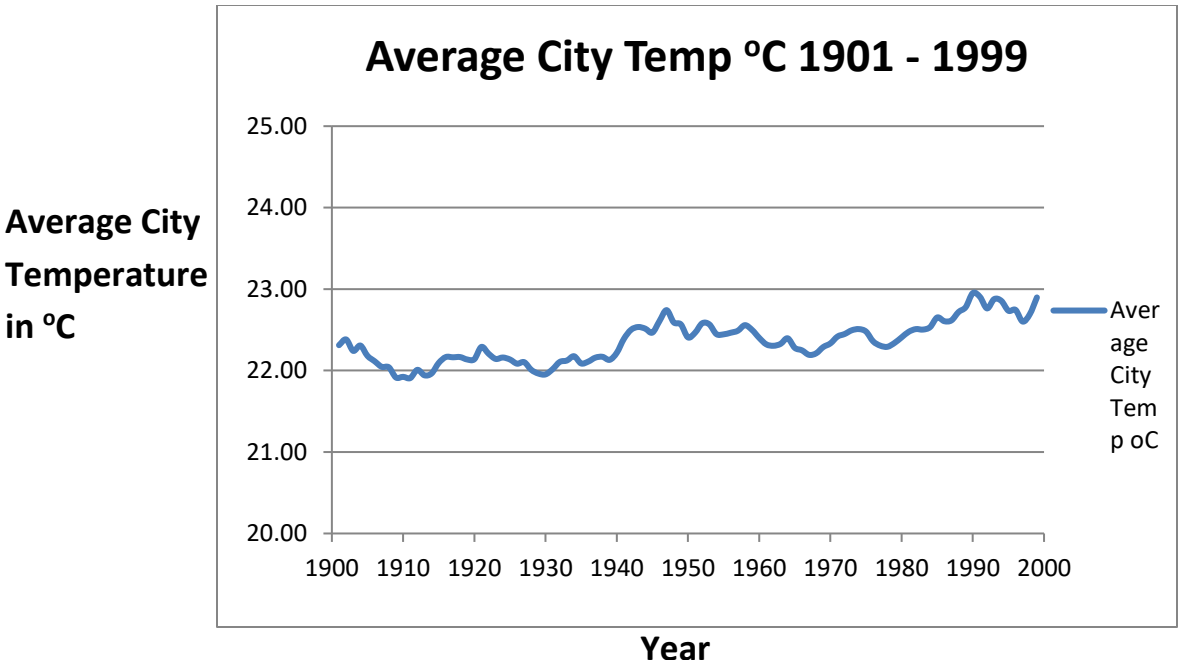
ii. Average temperatures in this bracket is **22.35 ºC**.

iii. There are total **99** reading in this bracket.

iv. Out of 99 reading **51 reading are below** the average line and **48 readings are above** the average line, which shows an equal trend.

v. First reading of temperature in **1901 is 22.31 ºC**, which rose to **22.90 ºC in year 1999**, which shows a **clear rise in temperature over the century**. The graph is **almost linear and shows a gradual upward trend**.

vi. Portion of 1985 – 1999 shows a clear rise in temperature, with no reduction.



(3) **<u>2000 – 2013</u>**

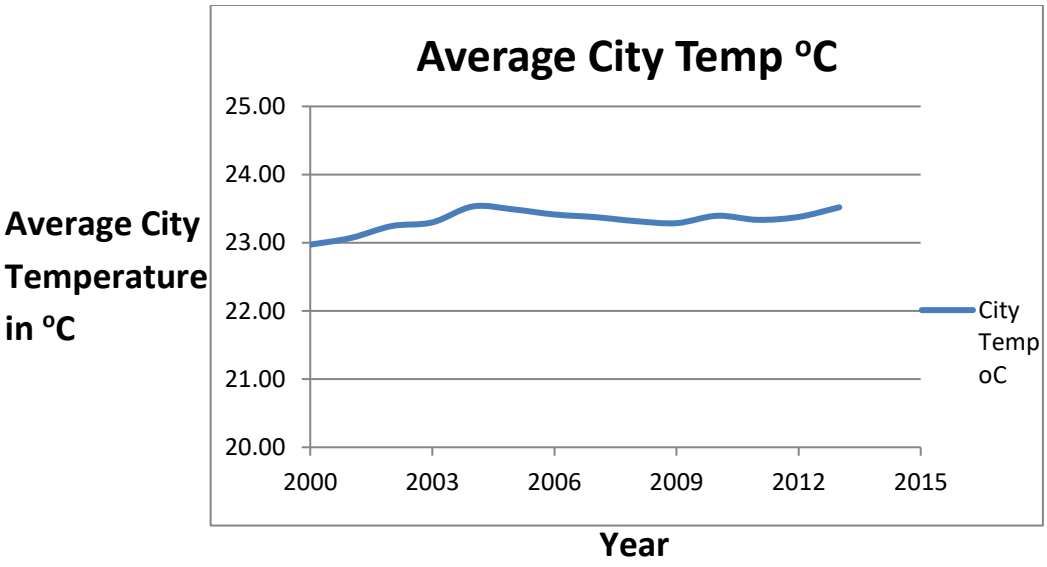
i. Average temperatures in this bracket is **23.33 ºC**.

ii. Previously I had made this graph, just by taking 3 values, as the moving averages of each sixth year was taken. Now I have taken **14 values by calculating the moving average on the basis of six years but for each successive year**.

iii. Out of 14 reading **6 reading are below** the average line and **8 readings are above** the average line, which shows an **upward trend**.

v. First reading of temperature in **2000 is 22.97 ºC**, which rose to **23.52 ºC in year 2013**, which shows a **clear rise in temperature over a period of 13 years**. The graph is **almost linear and shows a gradual upward trend**.

iv. First reading of temperature in 2001 is 23.15 ºC, which rose to 23.55 ºC in year 2013, which shows an upward trend.



b. **<u>Global Temperatures</u>**. In order to analyze the Global temperatures trends, I have divided the moving averages of temperatures into three brackets, i.e, **1820 – 1900**, **1901 – 1999** and **2000 – 2013**.

(1) **<u>1820 – 1900</u>**

i. Previously I had made this graph, just by taking 14 values, as the moving averages of each sixth year was taken. Now I have taken **81 values by calculating the moving average on the basis of six years but for each successive year**.

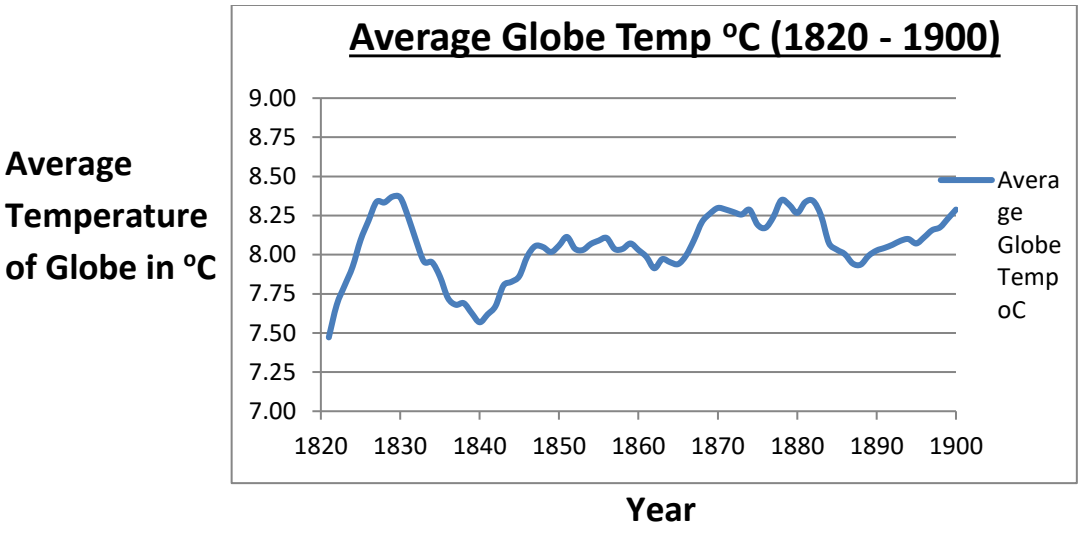
ii. Average temperatures in this bracket is **8.05 °C**.

iii. There are total **81** reading in this bracket.

iv. Out of 81 reading **38 reading are below** the average line and **42 readings are above** the average line and one reading is equal to the average.

v. First reading of temperature in **1820 is 7.47 °C**, which rose to **8.29 °C in year 1900**, which shows a **clear rise in temperature over the century**, despite the fact that the **trend was not linear**.

vi. In the last 50 years i.e, 1850- 1900 there are **51** reading and **18 (35 %) reading are below** the average line and **33 (65 %) readings are above** the average line, which shows that in the second half of the century the temperature rise was more consistent.



(2) **<u>1901 – 1999</u>**

i. Previously I had made this graph, just by taking 16 values, as the moving averages of each sixth year was taken. Now I have taken **99 values by calculating the moving average on the basis of six years but for each successive year**.

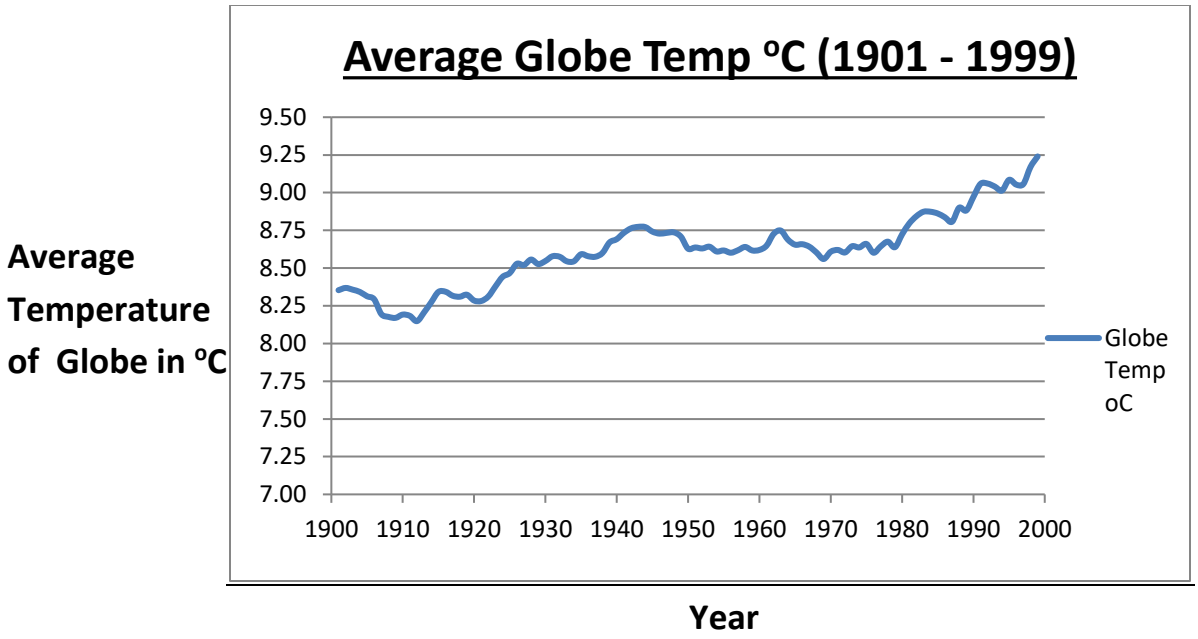
ii. Average temperatures in this bracket is **8.62 ºC**.

iii. There are total **99** reading in this bracket.

iv. Out of 99 reading **45 reading are below** the average line and **50 readings are above** the average line and four reading are equal to the average.

v. First reading of temperature in **1901 is 8.35 ºC**, which rose to **9.24 ºC in year 1999**, which shows a **rise of almost 1 ºC over the century**.

vi. In the last decade i.e, 1990 – 1999 a clear rise in average temperatures from 8.75 till 9.25 can be observed, which shows an upward trend in the temperatures.



(3) **2000 – 2013**

i. Previously I had made this graph, just by taking 3 values, as the moving averages of each sixth year was taken. Now I have taken **14 values by calculating the moving average on the basis of six years but for each successive year**.

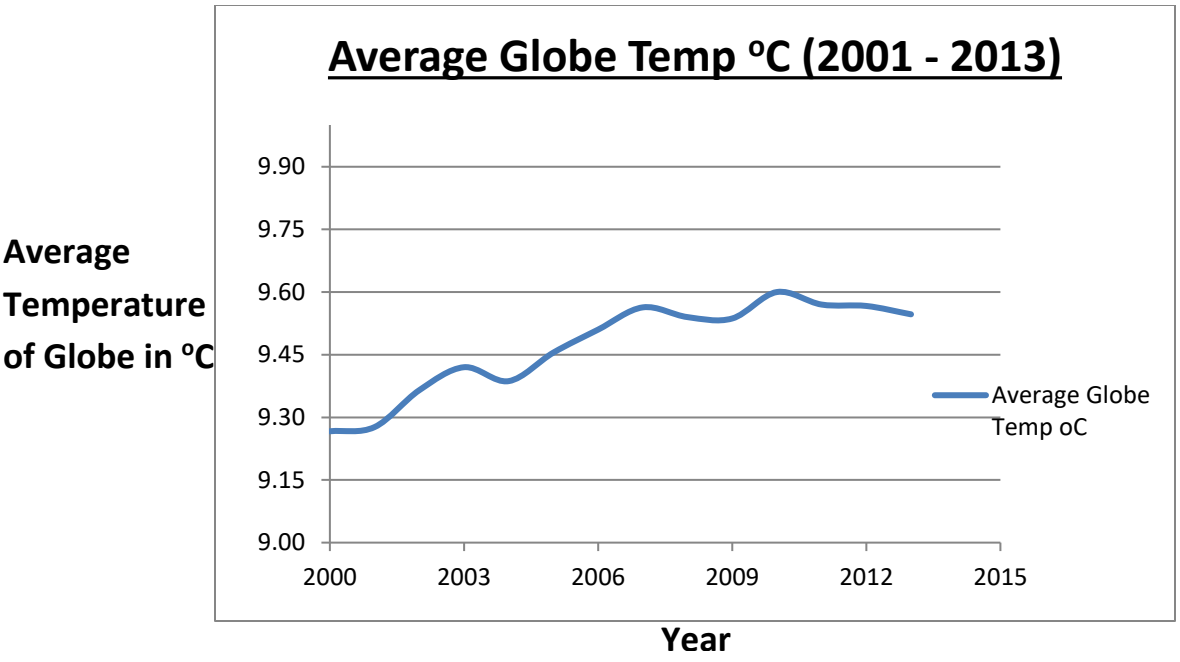
ii. Average temperatures in this bracket is **9.47 ºC**.

iii. There are total **14** reading in this bracket.

iv. Out of 14 reading **6 reading are below** the average line and **8 readings are above** the average line.

v. First reading of temperature in **2001 is 9.27 ºC**, which rose to **9.55 ºC in year 2013**.

vi. There a clear rise in temp from 2000 till 2010 which drops in last three years, this is evident from the graph.



c. **<u>Comparison between the Temperatures of Rawalpindi and Globe</u>**. Keeping the same brackets as guiding factors, comparison table is as under:-

Category	1820 - 1900		1901 – 1999		2001 - 2013	
	Rawalpindi	Globe	Rawalpindi	Globe	Rawalpindi	Globe
Lowest Average Temp	21.53	7.47	21.91	8.15	22.97	9.27
Highest Average Temp	22.81	8.37	22.95	9.24	23.53	9.60
Rise in Average Temp	1.28	0.90	1.04	1.09	0.56	0.33
Average Temp	22.01	8.05	22.35	8.62	23.33	9.47
Difference between average temp	13.96		13.73		13.86	

(1) The above table clearly shows that:-

(a) There is an **upward trend** in the average temperatures, whether we take the data of the globe or of Rawalpindi city.

(b) Average temperatures of Rawalpindi are above the average temperatures of the globe, which remained true throughout the period of research data. As is evident from the difference in the average temperatures i.e, 13.96, 13.73 and 13.86.

(c) Overall temperature rise in research period in the average temperature of globe was more than that of Rawalpindi city, which is evident from the table below:-

Ser	Cat	Rawalpindi	Globe
i.	lowest average temperature ^{O}C in	21.53	7.47
ii.	Highest average temperature ^{O}C	23.53	9.60
iii.	**Rise in Temperature ^{O}C**	**2.0**	**2.13**

(d) **<u>Average Temperature Rise</u>**

i. In the first century **1820 – 1899** the temperature rise of Rawalpindi city was more than that of the globe

ii. In the second century **1901 – 1999** the average temperature rise of globe is more than that of Rawalpindi city.

iii. In the research part of the third century **2000 – 2013** the temperature rise of Rawalpindi city is more than that of the globe.

(e) Above observation is supported by the fact that the trend of the difference between the average temperatures of Rawalpindi and Globe is decreasing, although the last bracket due to its less values does not conform this.

(f) The spikes shown in the graph of average temperatures of Rawalpindi city are not the same in that of the globe, which means that it might be due to some local activity in the city, having no effect or negligible effects on the average temperatures of the globe.

4. Line chart showing complete data and comparison the temperatures of Rawalpindi and that of the globe is as under:-

COMPARISON OF AVERAGE TEMPERATURES OF RAWALPINDI AND THE GLOBE