EXPLORING WEATHER TRENDS

Analyzing data

Finding the nearest city:

```
select * from city_list;
```

I viewed the list of cities that are available in the city_list table and I decided to choose Surat as it is the closest city to my residence.

Global Data:

```
select * from global_data;
```

The data in the global_data table is from 1750 to 2015. The result is the file **global_data.csv**

Surat city data:

```
select year, avg_temp from city_data
where city like 'Surat';
```

Surats' data in the city_data table is from the year 1796 to 2013 and misses some of the year's data in between. The result is in the file **city_data.csv**

Finding which years data is missing in global temperature and city temperature data:

After running the below 2 commands, I came to know that only the city_data table misses the years and the global_data table has all the year's data ranging from 1750 to 2015

1. The temperature of a particular year that is available in global_data but not in city_data

```
select gd.year
from global_data gd
where year not in
(select cd.year
  from city_data cd
  where city like 'Surat');
```

The query had 48 results which means that city_data has 48 missing data in comparison to global_data

2. The temperature of a particular year that is available in city_data but not in global_data

```
select cd.year
from city_data cd
where city like 'Surat'
and year not in
(select gd.year
  from global_data gd);
```

The query had 0 results which means that the global_data table has temperature for all the years which are available in the city_data table for the city Surat.

Gathering required data only

Now as city_data is having missing values I filled the missing data with mean as the data does not have a high amount of change rates.

The below query does the following

1. Table t1: selects data from global_data table
2. Table t2: selects Surat city data from city_data table
3. Coalesce missing data of city with mean which was calculated earlier
4. Left join based on condition, if the year does not exists in t2 table then assume it as t1 year, as the city temp will not be available, the mean will be filled in it

```
with t1 as (
  select gd.year gdyear,
  gd.avg_temp gdtemp
  from global_data gd),
t2 as (
      select cd.year cdyear,
  cd.avg_temp cdtemp
  from city_data cd
  where city like 'Surat'
)
select gdyear, gdtemp,
coalesce(cdtemp, 26.32)
from t1
left join t2
on t1.gdyear =
case
      WHEN t2.cdyear IS NULL THEN t1.gdyear
      else t2.cdyear
END;
```

The query result is attached in file **final_data.csv**

Gathering additional data for further analysis

1. Getting the temperature average of all the cities of India. If it is in Gujarat state then the state column will have value ‘Gujarat’ or it will have value ‘Rest of India’. (Surat city is in Gujarat). The data is in the file **indian_cities_temp.csv**

```
select city,
round(avg(avg_temp), 2) as year_avg,
case
when city in ('Ahmadabad', 'Rajkot', 'Surat', 'Vadodara') then 'Gujarat'
 else 'Rest of India'
 end as state
from city_data
where city in
(select city from city_list where country like 'India')
group by city
order by year_avg desc;
```

2. Getting data of only Gujarat state for all the years.

```
with guj as (
      select year,
  max(case when city like 'Surat' then avg_temp end) as surat,
  max(case when city like 'Rajkot' then avg_temp end) as rajkot,
  max(case when city like 'Ahmadabad' then avg_temp end) as ahmadabad,
  max(case when city like 'Vadodara' then avg_temp end) as vadodara
  from city_data
  where city in ('Surat', 'Rajkot', 'Ahmadabad', 'Vadodara')
  group by city_data.year
  order by year
)
select gd.year as Year,
gd.avg_temp as Global,
coalesce(guj.surat, 26.31) as Surat,
coalesce(guj.rajkot, 26.07) as Rajkot,
coalesce(guj.ahmadabad, 26.53) as Ahmadabad,
coalesce(guj.vadodara, 26.31) as Vadodara
from global_data gd
left join guj on
gd.year = guj.year
order by gd.year;
```

The above query will do the following.

a. Table guj: contains data of cities of Gujarat and it is aggregated by year.
b. The null values will be filled by mean for each city

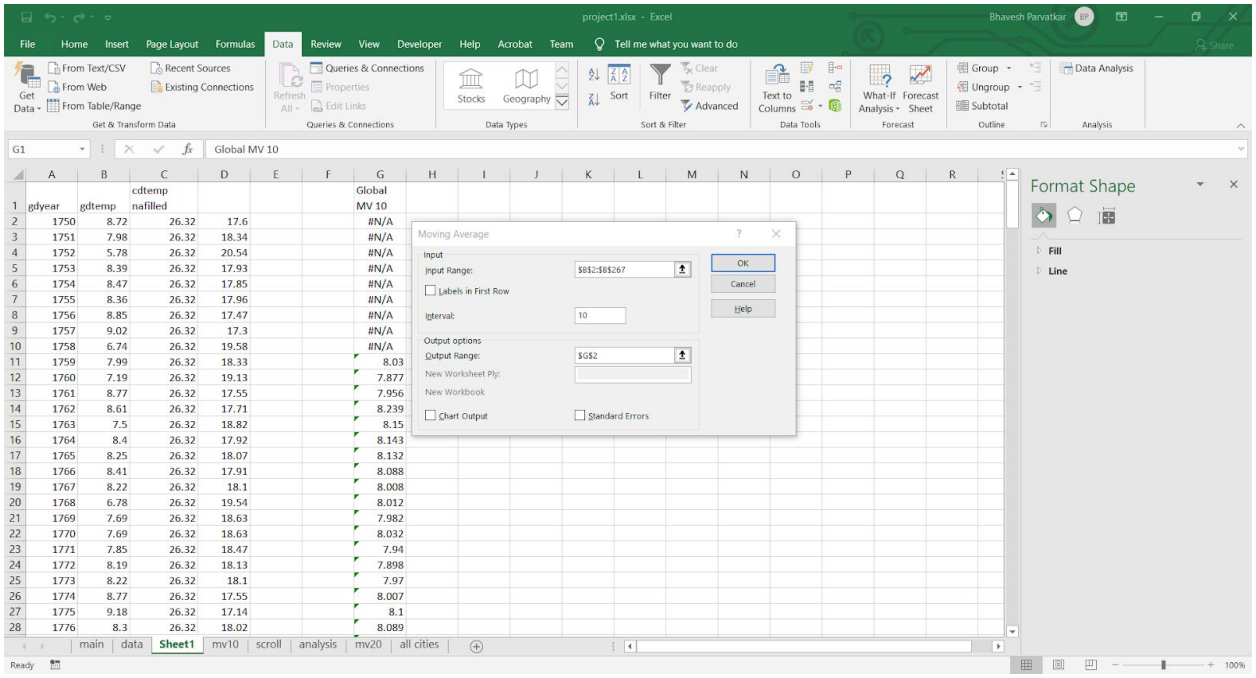
Example of result:

Year	Global	Surat	Rajkot	Ahmadabad	Vadodara
1800	8.48	26.37	26.1	26.56	26.37
1801	8.59	25.6	25.63	25.73	25.6
1802	8.58	26.84	26.55	27.01	26.84
1803	8.5	26.6	26.31	26.77	26.6
1804	8.84	26.95	26.63	27.09	26.95

Tools

Excel
I used Excel as it was based on my preference.

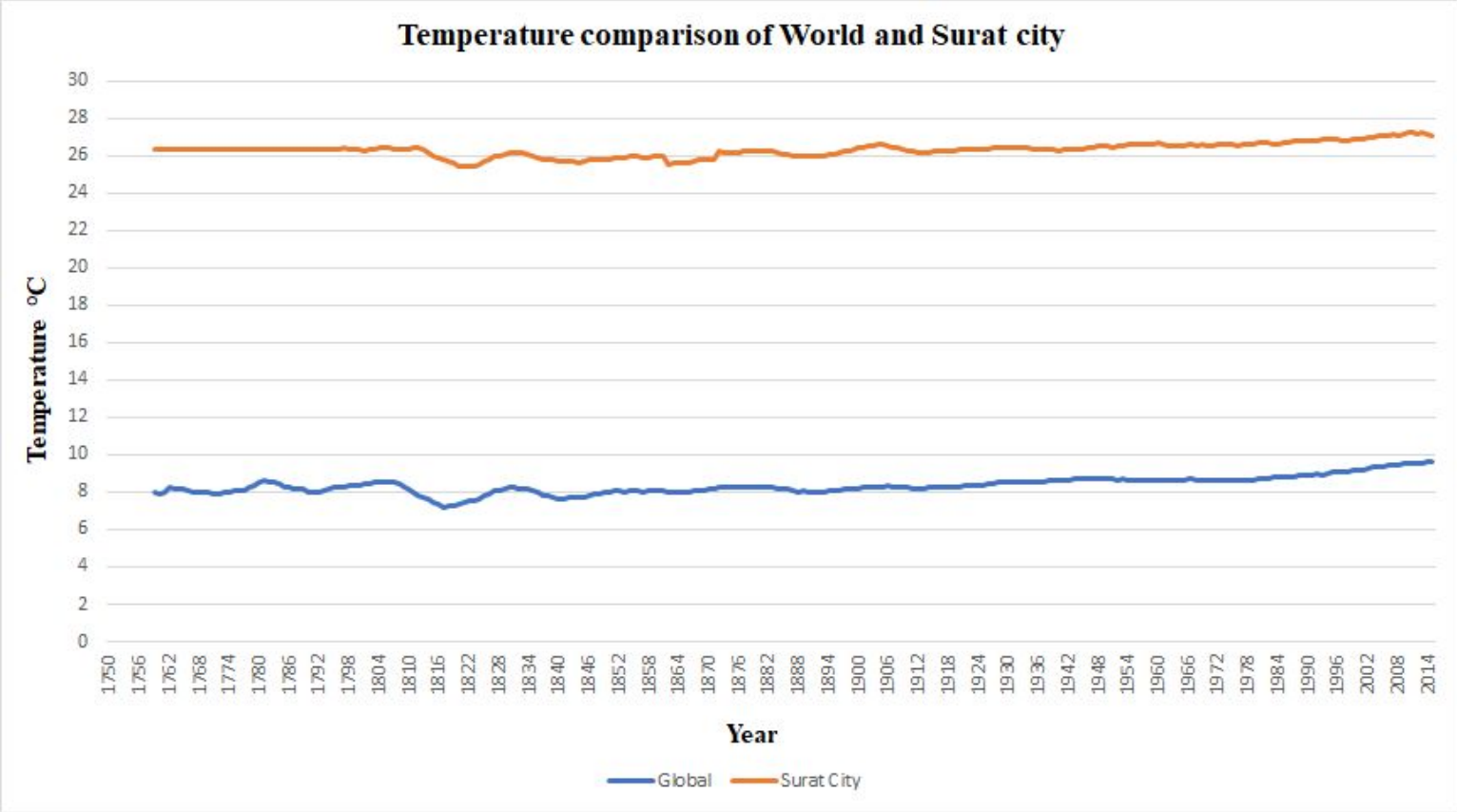
Calculating Moving Average
I used the Moving Average tool available in the Data Analysis section in Excel. I tried options ranging from Moving Average 2 to 20 with an interval of 2 which I tested and settled for Moving Average of 10 as it showed the necessary details of the temperature movement along with highlighting the trend of the temperature.



Using scroll bar

Using a scroll bar was a personal preference as there is too much data, it's better to keep a slider to use it as a reference while observing temperature for a particular year. I used a scroll bar that is available in the Developers Tab in excel.

Observations



1. **Is your city hotter or cooler on average compared to the global average? Has the difference been consistent over time?**

Answer: Surat is hotter on average compared to the global average. The difference has been almost consistent over time as seen on the graph.

2. **How do the changes in your city's temperatures over time compare to the changes in the global average?**

Answer: The change in the global average temperature has shown effect in the temperature in Surat too. For instance, when the temperature fell globally during the 1810s by 1℃, the

temperature in Surat fell relatively to around 1℃. Similarly, the temperature in Surat rose back by 1℃ by 1830 too when the global temperature rose back by 1℃.

3. **What does the overall trend look like? Is the world getting hotter or cooler? Has the trend been consistent over the last few hundred years?**

Answer: The overall temperature is increasing at a steady rate. If we look at the temperature from the year 1900, it has been increasing at a consistent rate. The world is getting hotter and the trend has been consistent over the last few hundred years.

4. **How much has the average temperature risen according to the data?**

Answer: If we compare the data of the past 2 decades, the global average temperature has risen by around 1.2℃ whereas for Surat the average temperature has risen by 0.9℃.

Year	Global Temp	Surat Temp
1759	8.03	26.32
1800	8.38	26.33
2013	9.55	27.21

5. **Which century was the hottest? 19th or 20th?**

Answer:

Century	Global	Surat City
19	8.0093	25.95645161
20	8.6381	26.5262

> The average for each century was taken for both global and Surats' temperature and we can conclude that the 20th century was the hottest century compared to the 19th century.

6. **What was the hottest and the coldest year for Global Temperature and Surat City Temperature? What can you conclude from the data?**

Answer: The highest and the lowest temperature based on the Moving Average temperature is given in the table below. We can see that the Global temperature has a range of 7.203 ℃ to 9.594 ℃ while Surat City has a temperature range of 25.401 ℃ to 27.24 ℃.

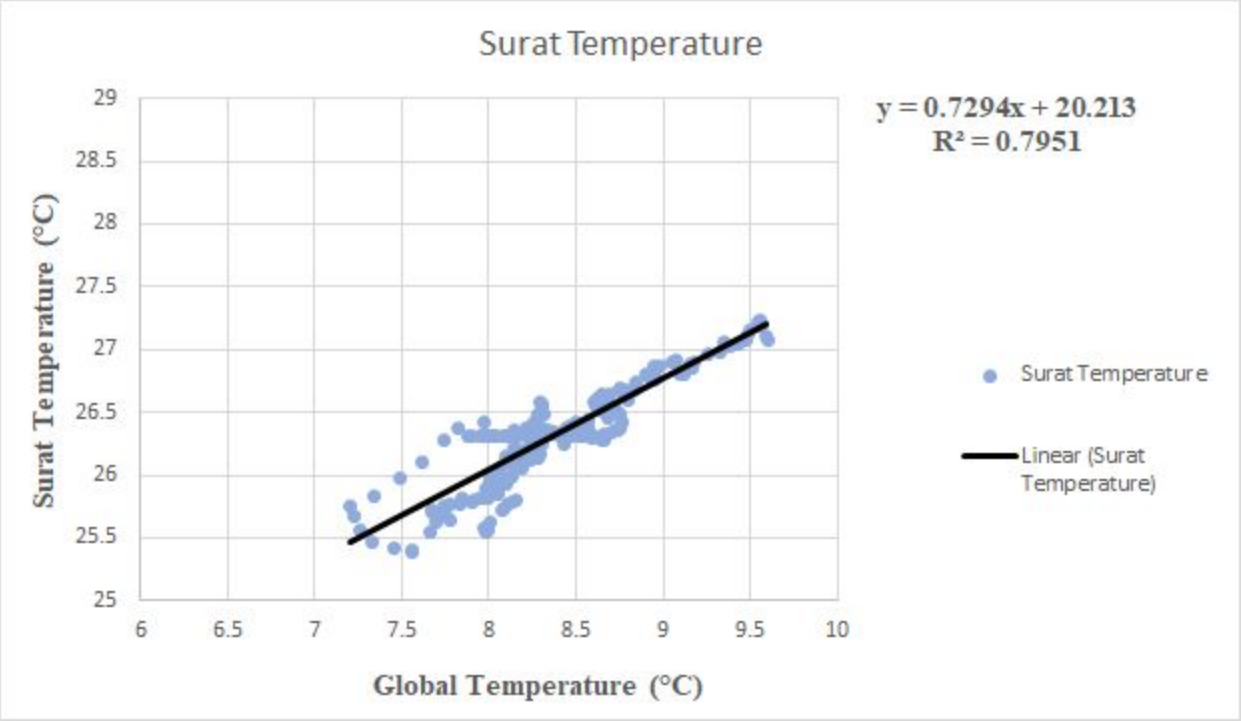
Temp Type	Year	Global	Year2	Surat City
Min	1817	7.203	1822	25.401
Max	2015	9.594	2011	27.24

We can see that Surat City touched its lowest temperature after the Global temperature hit its lowest. However, in the case of hottest year Global temperature hit its highest after Surat City has hit its highest. We can see that the temperature for Surat is more affected when the global temperature is high as Surat is a hotter region.

7. **What is the correlation coefficient?**

Answer: The correlation coefficient is 0.891694777. I have used Pearson's correlation formula which is available in excel. The correlation is very strong positively between Global and Surat temperature which means that the temperature of Surat is affected based on Global temperature too.

Here is a scatter plot of Global vs Surat temperature.



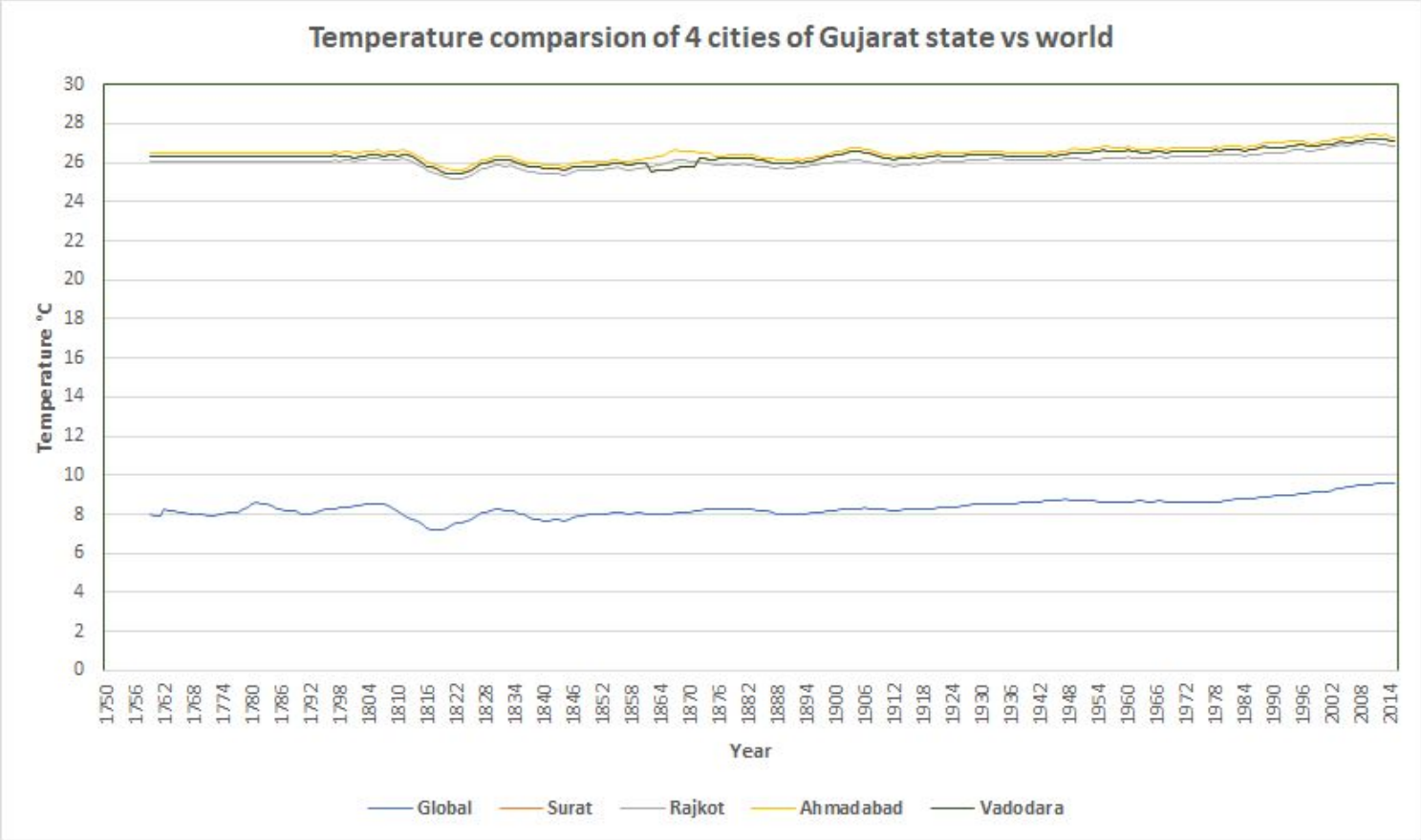
8. **Surat is in the Gujarat state of India, what is the temperature comparison with other cities of Gujarat?**

Answer: The dataset had 4 cities that are in the Gujarat state. The data is available in the file **gujarat_temp.csv**. The cities are Ahmadabad, Rajkot, Surat, and Vadodara. After comparing the temperature trends of all 4 states we can state that:

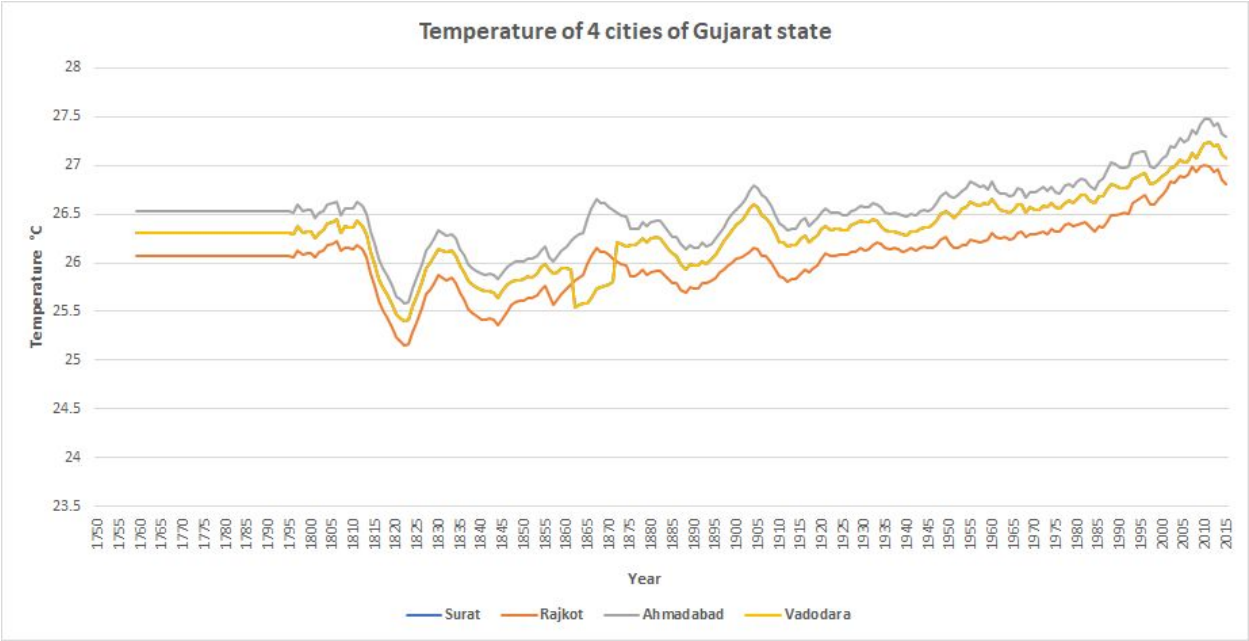
a. Ahmadabad always had the highest temperature
b. Rajkot mostly had the least temperature
c. Surat and Vadodara had the same temperature

In short, Rajkot > Surat = Vadodara > Ahmadabad.
We can conclude that:

a. The temperature of Surat will mostly lie between the temperature of Ahmadabad and Rajkot.
b. Surat and Vadodara will always have similar temperatures.



Viewing only the 4 cities data (The trend line of Surat can't be seen because Surat and Vadodara have the same temperature for all the years). (On the next page)



9. **What has been the average temperature for the cities of India for the last 2 decades?**
Answer: Around 22 cities of India were available in the city_data dataset. The data is available in the file **indian_cities_temp.csv**. The lowest temperature average is of Ludhiana city which is 23.56 and the highest in the Hyderabad city with an average temperature of 26.73. All the 4 cities of Gujarat are in the top 5 list. Mostly the Northen cities are colder than the rest of the country.

The top 5 result is:

Country	City	Average	State
India	Hyderabad	26.73	Rest of India
India	Ahmadabad	26.53	Gujarat
India	Surat	26.32	Gujarat
India	Vadodara	26.32	Gujarat
India	Rajkot	26.08	Gujarat

The map is shown on the next page.

Indian states temperature average since last 2 decades