Amit Kumar Shankar   
Project 1

SQL query to extract data

**SELECT** **\***

**FROM** city\_list

**WHERE** country **in** **(**'United States'**)**

--nearest city is San Francisco

**SELECT** year**,** avg\_temp

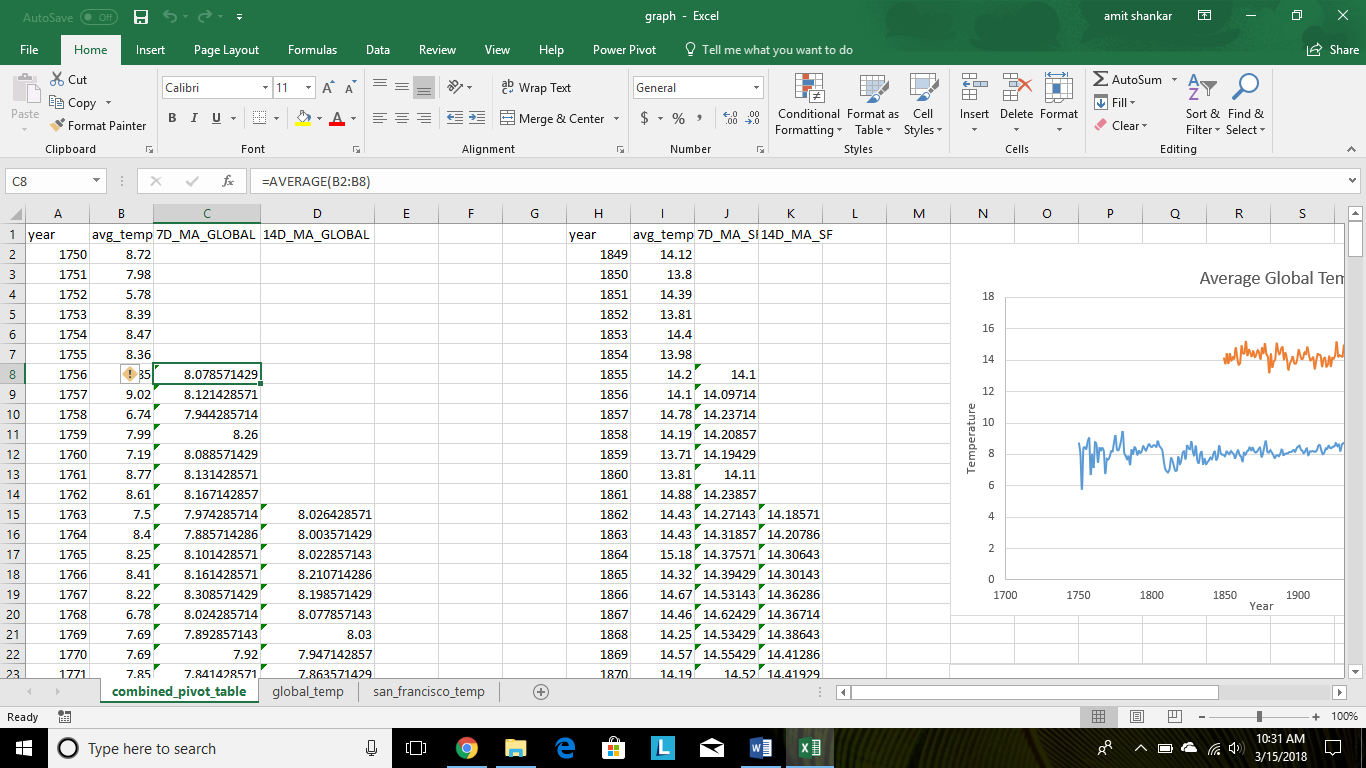
**FROM** city\_data

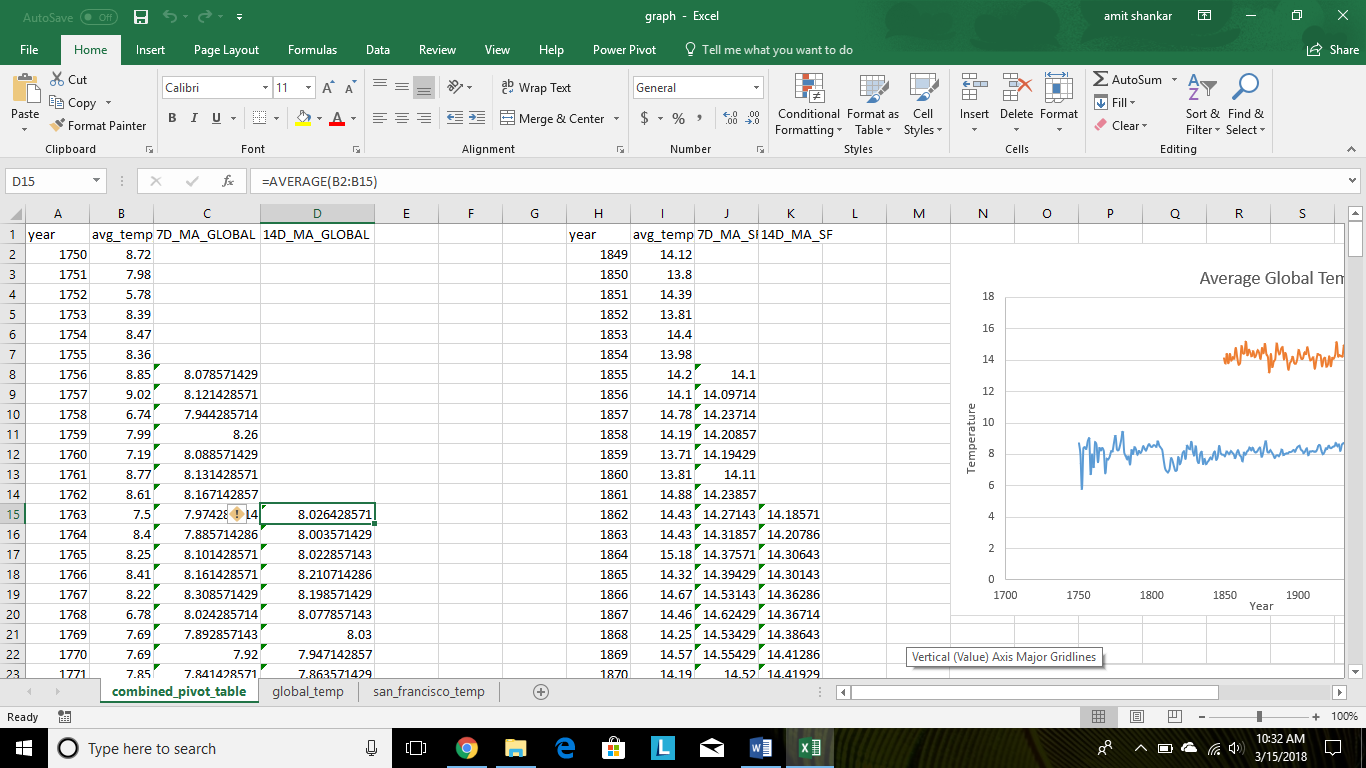
**WHERE** country **in** **(**'United States'**)** **AND** city **in** **(**'San Francisco'**)**

**SELECT** **\***

**FROM** global\_data

Excel Documentation  
Saved the two csv files on desktop and opened both the files in excel. Created a new sheet in a new excel file and copied and pasted both the csv files in on sheet and created the line graphs. The next two images show the excel formulas used to calculate the 7 day and the 14 day moving averages.





Below are the line graphs generated using the data.

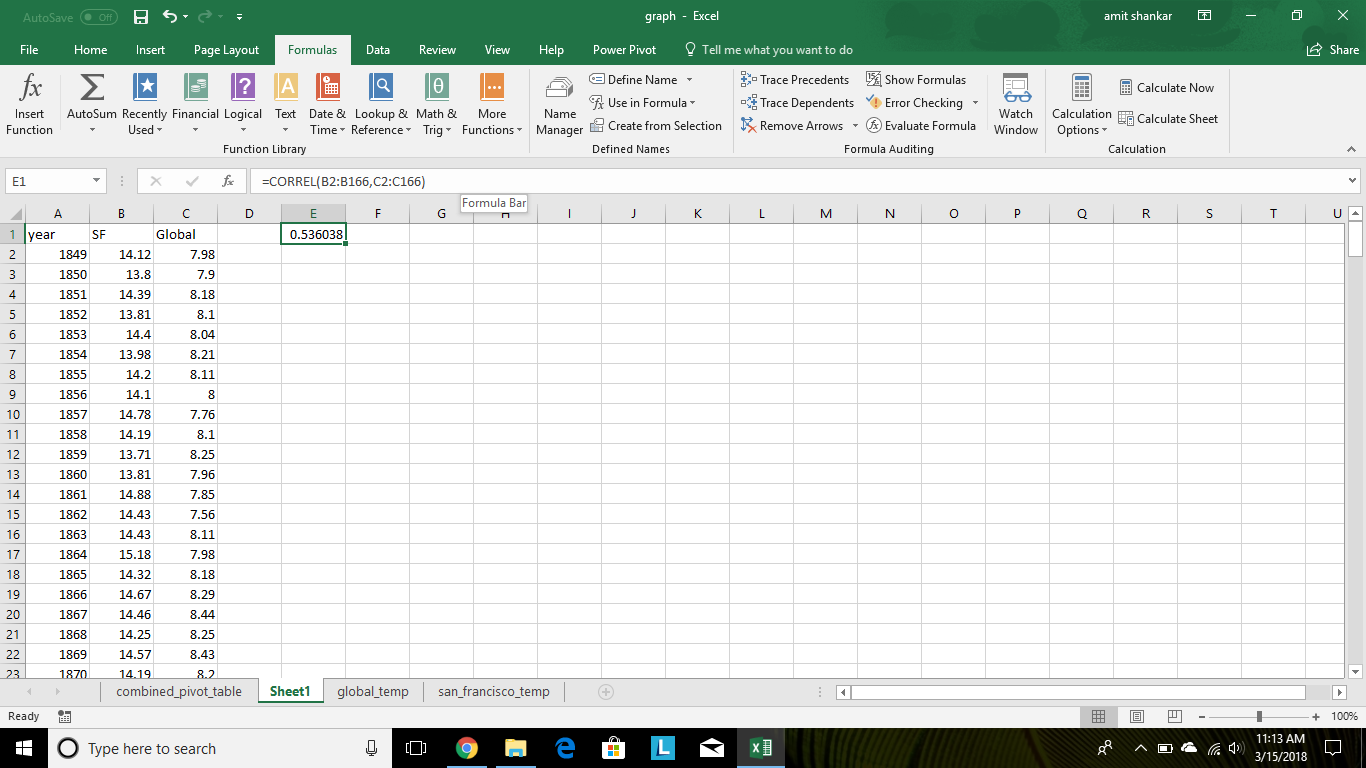
Graph 1 shows the actual temperature values against time. Because of the variations in the temperature and large number of data points, it makes it hard to see the overall trends. Therefore, the 7 day and 14 day moving average line graphs were created to identify the trends.

Observations

1. Both Global and San Francisco temperatures are increasing.
2. Temperature in San Francisco is higher (hotter) than the Global temperature.
3. The Global temperature is between 8-10 degrees while the temperature in San Francisco is between 14-16degrees. Overall, the temperature in San Francisco is approximately 4 degrees higher than the Global temperature.
4. The San Francisco and Global temperature patterns are very similar. Both increase or remain constant together. It could also be that the San Francisco temperature lags while the global temperature leads.
5. There are more data points available for the Global Temperature than the San Francisco temperature.
6. Overall, both Global and San Francisco temperature is increasing overtime and San Francisco is getting hotter.

Extra Points

The screen below shows the correlation coefficient between Global and San Francisco temperature to be 0.54. The shows that there is a positive correlation between the two temperatures. This also means that as one temperature goes up, the other goes up as well.



Looking at the Slope for the two graphs:

Its interesting to see that the slopes of both the 14 day MA graphs are postive and same (0.0045) but with a different y intercept. This means that the linear steepness of the graphs are same (moving at the same rate) but from different starting points ( the y interecept). The positive slope confirms an upward trend.