Udacity Project - Explore Weather Trends

Project by Abhi Vasanth
Email : abhivasanth90@gmail.com
Contact No. 1- 415 928 9707

Summary of the project

In this project, local and global temperature data are analysed and the temperature trends are compared between San Francisco (the city I live in) to overall global temperature trends.

Goal and process flow of the project

The goal of the project is to create a visualization and document the observations based on the temperature trends between San Francisco(SF) and the Overall global temperature trends.

Process flow - different stages

In order to accomplish this goal, we'll complete the following stages/steps in a sequence and illustrate the observations with the visualization.

Stage 1 : Extract data sets required for analysis (SQL)
Stage 2 : Data transformation/manipulation process (Google Sheets)
Stage 3 : Create a Data Visualization line chart. (Google Sheets)
Stage 4 : Perform Data analysis and observation based on Stage 3. (SQL and google sheets)

Stage 1 : Extract data sets required for analysis (SQL)

The following SQL queries were used to extract the data required for my analysis.

Explore Weather trends - sql queries with comments

-- Exploring city_list table to find my city San Francisco. Using the where clause. At first I used the LIKE operator to find cities starting with San and then took note of the exact name and format (case sensitivity scenario and then finally arrived to the below query)

```
SELECT city, country
FROM city_list
WHERE city = 'San Francisco';
```

-- Extracting data for San Francisco from the city_data table - The fields are year, city, country and avg_temp. Utilization of AND operator.

SELECT year, city, country, avg_temp
FROM city_data
WHERE country = 'United States'
AND city = 'San Francisco';

-- Extracting data for global temperature by year - extracting entire data set

SELECT year, avg_temp
FROM global_data;

Note : After evaluating each of these queries. I hit the download csv link to my local PC with a name for the file. (Especially query 2 and query 3)

The following are the extracted files:
San Francisco temperature data
Global temperature data

Stage 2 : Data transformation/manipulation process

In order to visualize our data in the form of a line chart, one of the requirement is to use the moving average instead of the yearly average. In order to perform this manipulation process, we use Google Sheets to calculate the moving average for each of our data set i.e San Francisco temperature data and Global temperature data.

Basic steps :
1. Decide what time period to be used for the moving average. For this project and experimentation reasons, I've used 20 years.
2. Add a derived column to the spreadsheet and name it Moving average. Go to the 21st cell of this derived column and add a formula =AVERAGE(B2:B21) and hit enter; where B2 to B21 are the cells from where we are going to reference the average temperature.
3. Once the value is updated, drag down the cell all the way to last cell to update the rest of the Moving average values
4. Repeat above steps for the global temperature data.
5. To perform data analysis on both the datasets side by side, I've combined both the data set in a single spreadsheet. (Please refer below link with steps performed in a google sheet)

Explore Weather Trends Global Vs SF

Stage 3 : Create a Data Visualization line chart. (Google Sheets)

In order to create a Data Visualization line chart to analyze the trends for the local temperature and global temperature, I've used Google sheets.

We used google sheets as a spreadsheet to manipulate the data set and calculate its moving average.

Steps to create a Line Chart (Smooth)

1. Select all from the data set to be analysed. (PFB - Screenshot)

Year (Global)	Global Average Temperature (Celsius)	20 Year Global Moving Average	Year (San Francisco)	City	SF Average Temperature (Celsius)	20 Year SF Moving Average
1994	9.04	8.8815	1994	San Francisco	14.21	14.7385
1995	9.35	8.912	1995	San Francisco	15.31	14.813
1996	9.04	8.9465	1996	San Francisco	15.56	14.855
1997	9.2	8.964	1997	San Francisco	15.75	14.911
1998	9.52	9.0055	1998	San Francisco	14.38	14.882
1999	9.29	9.0335	1999	San Francisco	14.41	14.853
2000	9.2	9.0445	2000	San Francisco	15.02	14.867
2001	9.41	9.0565	2001	San Francisco	15.25	14.8685
2002	9.57	9.103	2002	San Francisco	15	14.9185
2003	9.53	9.128	2003	San Francisco	15.43	14.9365
2004	9.32	9.1595	2004	San Francisco	15.37	14.9565
2005	9.7	9.2115	2005	San Francisco	15.17	15.0035
2006	9.53	9.2465	2006	San Francisco	15.02	15.0045
2007	9.73	9.2835	2007	San Francisco	14.94	15.004
2008	9.43	9.295	2008	San Francisco	15.05	15.0025
2009	9.51	9.3245	2009	San Francisco	15.02	15.031
2010	9.7	9.348	2010	San Francisco	14.67	15.028
2011	9.52	9.365	2011	San Francisco	14.5	15.028
2012	9.51	9.3985	2012	San Francisco	15.05	15.004
2013	9.61	9.4355	2013	San Francisco	16.23	15.067
2014	9.57	9.462				
2015	9.83	9.486				

2. Choose tab option Insert and select chart and then from the options choose line chart

   a) Use the options from Customize to edit/modify Line chart title, axis and legends
   b) Use the option from data to choose the appropriate style chart, series and axis from the data set.

Please find below screenshots for reference.
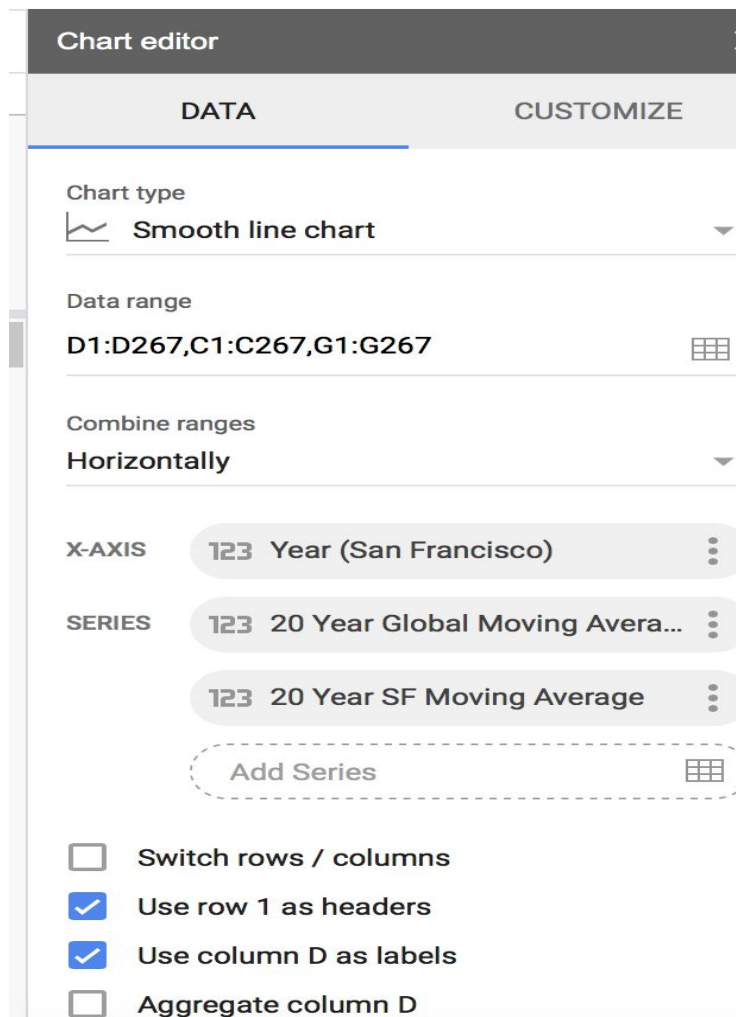
Key consideration when deciding how to visualize the trends?

- *Use the Line chart per the requirement. Ensure that the axis titles and legends are provided and displayed. Ensure that there is a unique color coding for the different temperature trends Global and SF temperature. Correct data set is used and series are provided with the moving average and not the yearly average. X axis and Y axis are given the appropriate names with units.*
- *Look for patterns in the line chart. Observe if the line is increasing by year. Compare the data against the two and see if we see a pattern in drop or rise. Observe factors that would influence the temperature.*
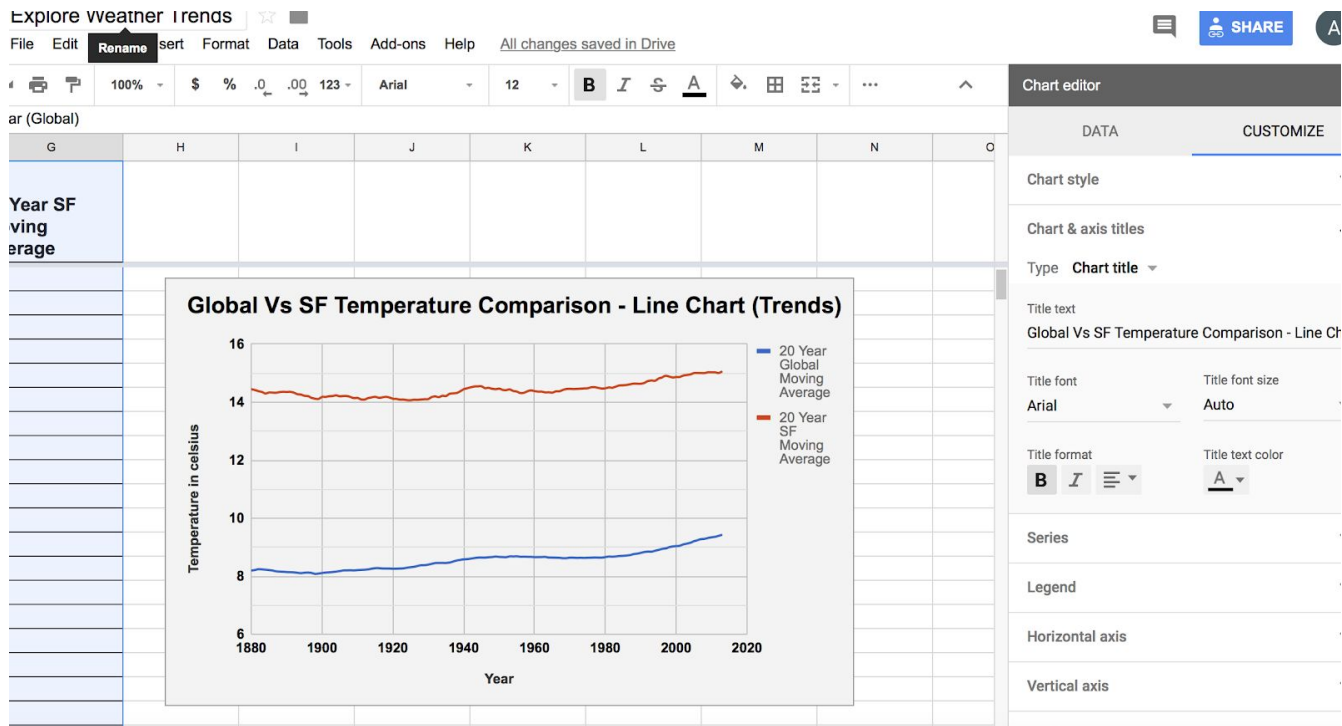
The Data Visualization Report : Global Vs SF Temperature trends comparison - Line Chart.

Fig a:
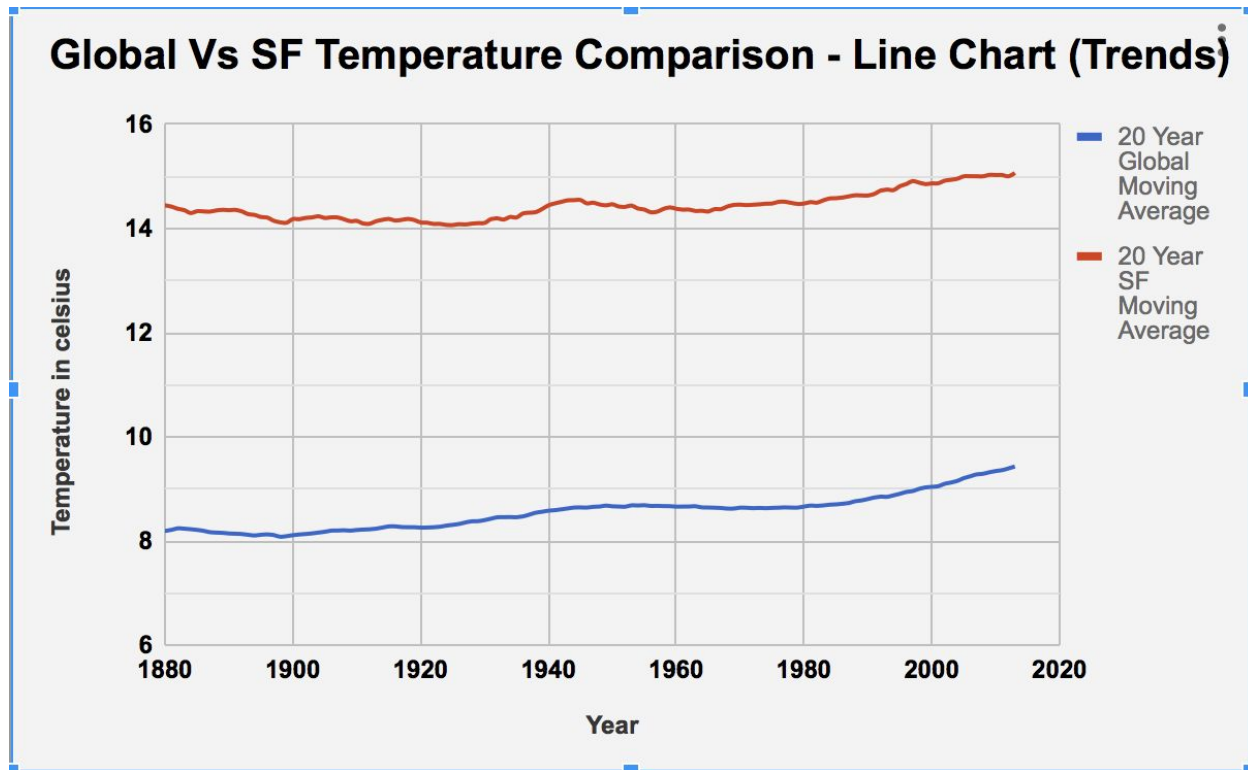
Perform Data analysis and observation based on Stage 3. (SQL and google sheets)

<u>Observations:</u>
1. San Francisco (SF) is hotter in terms of temperature average compared to the global average temperature. The difference between the two has been nearly consistent over time. Even though SF is a city which is located close to the ocean and it's usually windy and cold. The average temperature of SF is warmer compared to the average temperature globally. (Probably temperature from places like Russia and Mongolia have skewed the global temperature towards the colder side.)
2. There appears some correlation between the global average temperature to SF's average temperature. There's been an increase in temperature especially the last 30-40 years.
3. From the chart - fig a, with the recent years rise in technology, the rise in temperature for both San Francisco and the Global temperature is high likely due to global Warming. Greenhouse gases are causing climate change and also the reason for the rise in Temperature. Hence, possible correlation in rise in temperature at city and global level during the recent years. The trend for both local and global temperature appears to be on the rise for the upcoming years - based on the line chart trend fig a.
4. It appears there has not been a significant change in temperature for the last 200 years at a global level or at city level (SF). The temperature average has been been between 7 and 9.5 & 13 and 16 respectively. (I'd like to highlight that none of the temperature averages had a significant change like from 7 to 14 or 13 to 20 during the last 250 years)

   Fig a (see previous page) illustrates the trend where San Francisco shows higher temperature average compared to the Global average temperature.

   Additional observation using Dallas temperature data:
   1. Observed in all 3 cases :
      Global average temperature, SF average temperature and Dallas average temperature (moving average and yearly average). The highest recorded temperature has been the most latest year for their respective data sets. Please find below screenshot with the data set details. (See below screenshot with highlighted values - They represent the temperature in celsius)

      [Explore Weather Trends Global Vs SF Vs Dallas](#)

Year (Global)	Global Average Temperature (Celsius)	20 Year Global Moving Average	Year (San Francisco)	City	SF Average Temperature (Celsius)	20 Year SF Moving Average	year	city	avg_temp	20 Year Dallas Moving
1999	9.29	9.0335	1999	San Francisco	14.41	14.853	1999	Dallas	19.61	18.3985
2000	9.2	9.0445	2000	San Francisco	15.02	14.867	2000	Dallas	18.9	18.405
2001	9.41	9.0565	2001	San Francisco	15.25	14.8685	2001	Dallas	18.55	18.412
2002	9.57	9.103	2002	San Francisco	15	14.9185	2002	Dallas	18.19	18.413
2003	9.53	9.128	2003	San Francisco	15.43	14.9365	2003	Dallas	18.55	18.4805
2004	9.32	9.1595	2004	San Francisco	15.37	14.9565	2004	Dallas	18.58	18.486
2005	9.7	9.2115	2005	San Francisco	15.17	15.0035	2005	Dallas	19.13	18.5315
2006	9.53	9.2465	2006	San Francisco	15.02	15.0045	2006	Dallas	19.85	18.5825
2007	9.73	9.2835	2007	San Francisco	14.94	15.004	2007	Dallas	18.42	18.5895
2008	9.43	9.295	2008	San Francisco	15.05	15.0025	2008	Dallas	18.52	18.606
2009	9.51	9.3245	2009	San Francisco	15.02	15.031	2009	Dallas	18.42	18.653
2010	9.7	9.348	2010	San Francisco	14.67	15.028	2010	Dallas	18.69	18.639
2011	9.52	9.365	2011	San Francisco	14.5	15.028	2011	Dallas	19.69	18.7
2012	9.51	9.3985	2012	San Francisco	15.05	15.004	2012	Dallas	19.99	18.79
2013	9.61	9.4355	2013	San Francisco	16.23	15.067	2013	Dallas	20.45	18.9215
2014	9.57	9.462								
2015	9.83	9.486								

The above image highlights the highest observed temperatures in celsius. 2015 has 9.486 celsius for global average temperature, 15.067 in 2013 for SF, and 18.9215 in 2013 for Dallas.

The Data Visualization Report : Global Vs SF Vs Dallas Temperature trends comparison - Line Chart.

Fig b:

Global Vs SF Vs Dallas Temperature Comparison - Line Chart (Trends)