Explore Weather Trend by Trupti Patil

1. SQL CTE is used to extract clean data for city data New York, Pune and from Global data in a CSV file

```
/* t1 will find the average temp of New York */

with t1 as
(
  select
        round(avg(cd.avg_temp)) newyorkavg
  from
        city_data cd
  where
        cd.city = 'New York'
)
,

/* t2 will find the average temp of Pune */

t2 as
(
  select
        round(avg(cd.avg_temp)) puneavg
  from
        city_data cd
  where
        cd.city = 'Pune'
)
,

/* t3 will select and combine the average temp of New York, Pune and Global for
matching years from the database */

t3 as
(
  select
        cd.year,
        cd.avg_temp newyork_avg,
```

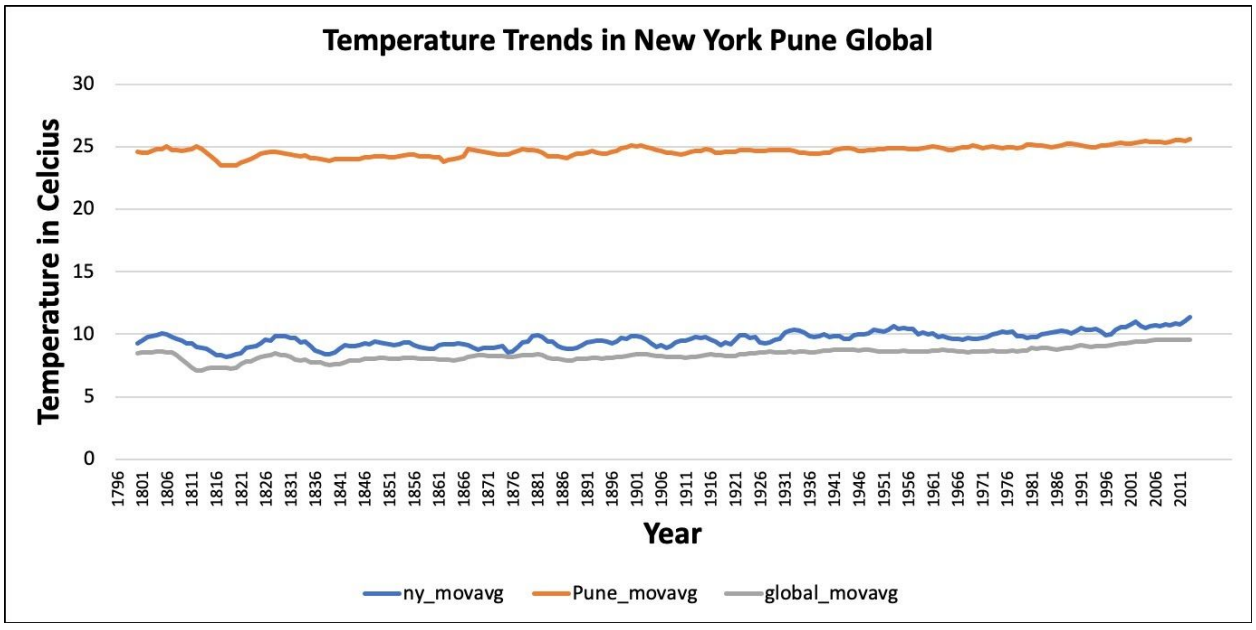
```
        pd.avg_temp pune_avg,
        gd.avg_temp global_avg
    from
        city_data cd
        join
        global_data gd
        on cd.year = gd.year
        join
        city_data pd
        on pd.year = gd.year
    where
        cd.city = 'New York'
        and pd.city = 'Pune'
)

/* Below join will replace null temp values with appropriate averages from t1 or t2 to give
clean avg_temp columns */

select
    t3.year,
    coalesce(t3.newyork_avg, t1.newyorkavg) newyorktemp,
    coalesce(t3.pune_avg, t2.puneavg) punetemp,
    t3.global_avg
from
    t3
    join
        t2
        on 1 = 1
    join
        t1
        on 1 = 1
```

2. The 5-year moving average for avg temperatures in New York, Pune and Global data from the CSV is calculated and used to plot line charts as shown below.

- Tool:
    - I used an Excel tool to calculate the moving average data of New York,Pune,Global temperatures.
- Method to Calculate:
    - I first calculated 10 yrs moving average. To calculate this create a new column ny_movavg. Find the average of the first 10 newyorktemp values using formula =Average(B2:B11) in the 11th cell of ny_movavg. B1 being the column heading is excluded. Then copy the formula in all the remaining cells in ny_movavg. I repeated the process to calculate pune_movavg using punetemp and global_movavg using globaltemp. I used these movavg columns to create a line chart.
    - Then I calculated 5 yrs moving average using formula =Average(B2:B6) in a similar process to draw another line chart.
- Key Considerations in visualization:
    - The 5 years moving average data was able to show more clearer fluctuations in the temperature trend. The 10 year moving average data was smoothing out the important curves over the years which could miss small but meaningful dips or increases in temperature. So I decided to use 5 yrs moving average data for final visualization below.



3. Observations

1. Average temperatures in cities Pune, New York and globally are rising above their mean.
2. Pune temperature is rising above 25 degrees, New York temperature is rising above 10 degrees.
3. Global temperatures are showing a slow upward trend from 8 degrees to 10 degrees as observed here from the 18th century till 21st century.
4. In the first decade of the 18th century there was a dip in temperature globally which is also confirmed by individual lines of New York and Pune.
5. Temperature trend has slowly grown since the 18th century with numerous intermittent up and down fluctuations.