

UDACITY: DATA ANALYST NANODEGREE

PROJECT 1: EXPLORE WEATHER PATTERNS

Report By: Havisha Kalwad

OVERVIEW

Global Warming has been a topic of concern in the last few years. The signs of global warming are seen increasingly every year. The most convincing proof of the occurrence of global warming is through data. Analyzing the temperature trends over the last few years would give a very precise and realistic overview of the situation. This report gives a clear insight of the changes in the average global temperature and average temperature in San Jose over the last few years.

METHOD

The data was retrieved from a database using Structured Query Language in the form of a .csv file and was analyzed on Jupyter notebook using Python programming language and matplotlib, pandas libraries. The data was analyzed using a 25 year moving average to make the trends clear.

RETRIEVING DATA FROM DATABASE

Data was retrieved from a database using SQL. The data obtained from the database was stored in the form of a .CSV file.

- To check the list of cities and countries available on the database:

```
SELECT * FROM city_list;
```

Observed that San Jose city was present in the table under the country name United States.

- To rename the column names:

```
ALTER TABLE city_data RENAME COLUMN avg_temp to local_avg_temp;
```

```
ALTER TABLE global_data RENAME COLUMN avg_temp to  
global_avg_temp;
```

- To download the required data of year, local_avg_temp, global_avg_temp into a table to further analyze the data.

```
SELECT global_data.year, global_data.global_avg_temp,
city_data.local_avg_temp FROM global_data, city_data WHERE
(global_data.year = city_data.year) AND (city_data.city = 'San Jose' AND
city_data.country = 'United States')
```

The data obtained from the database with the above query was downloaded into a .csv file. Renamed the .csv file to Global_and_local_temp_data.csv

ANALYZING DATA

- The data obtained was analyzed on Jupyter notebook. As shown in Fig. 1.

```
1. Importing necessary Libraries

In [3]: import pandas as pd
import matplotlib.pyplot as plt
import numpy as np

2. Importing the data

In [6]: df = pd.read_csv('Global_and_local_temp_data.csv')

In [7]: df

Out[7]:
```

	year	global_avg_temp	local_avg_temp
0	1849	7.98	14.12
1	1850	7.90	13.80
2	1851	8.18	14.39
3	1852	8.10	13.81
4	1853	8.04	14.40
...
160	2009	9.51	15.02
161	2010	9.70	14.67
162	2011	9.52	14.50
163	2012	9.51	15.05
164	2013	9.61	16.23

165 rows x 3 columns

Fig. 1. Importing data into Jupyter Notebook

- The Moving Averages was used to smoothen the data lines by filtering out the noise caused by random short term fluctuations and also to help observe the long term data trends as shown in Fig. 2.

3. Finding Moving Average

Moving Averages help to smoothen the lines by filtering out the noise caused by random short-term fluctuations and helps to observe long term trends in the data.

```
In [13]: global_moving_avg = df['global_avg_temp'].rolling(25).mean()
```

```
In [14]: local_moving_avg = df['local_avg_temp'].rolling(25).mean()
```

Fig. 2. Calculating moving average

- The data from global average temperature and San Jose average temperature trends was used to further analyze and find differences and similarities I the trends.

RESULTS AND OBSERVATIONS

- The Global average temperature has increased from 7.5 degrees Celcius to 9.7 degrees Celsius. The San Jose average temperature has increased from 13.8 degrees Celsius to 15.2 degrees Celsius. This clearly indicates the rise in temperatures over the years.
- Fig. 3 clearly indicates the general rise in the average temperature trends over the years, both locally and globally.

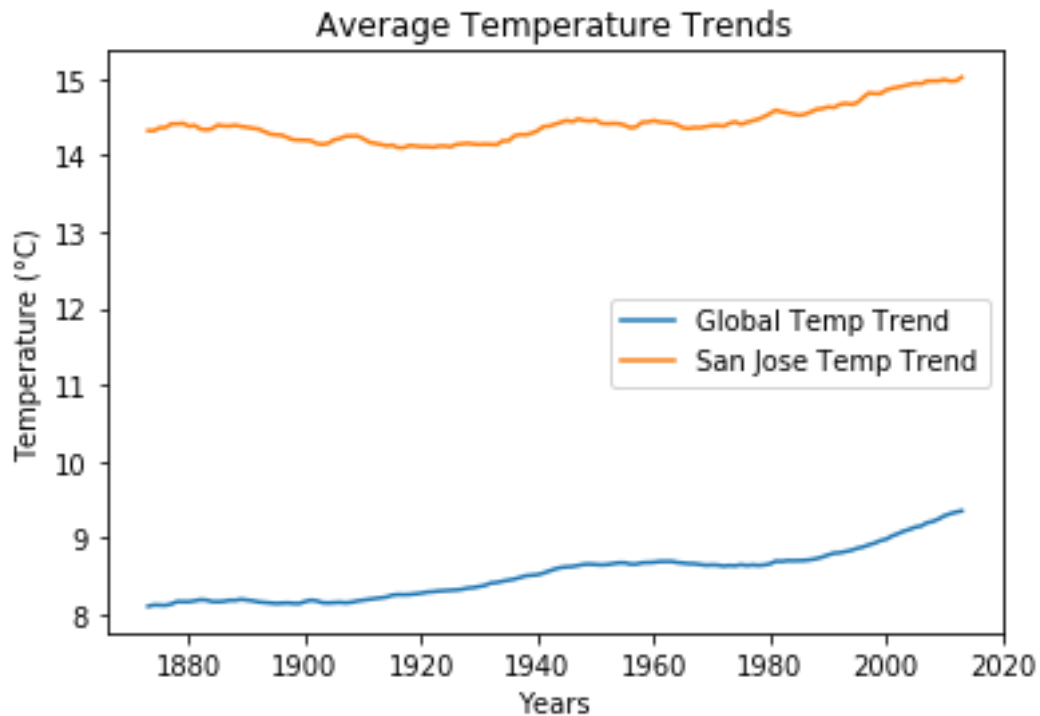


Fig. 3. Average Temperature Trends

- The metrics obtained for Global temperature trends and San Jose temperature trends as shown in Fig. 4. Global temperature has a mean of 8.5 degrees Celsius and the mean San Jose temperature is 14.4 degrees Celsius. This indicates that San Jose is relatively warmer than the global average all throughout the year. Standard deviation in the global temperatures is 0.46 while for San Jose it is 0.5. This indicates that the differences in the temperature for local and global temperatures are almost relatable.

5. Average Temperature

```
In [16]: df['global_avg_temp'].describe()
```

```
Out[16]: count      165.000000  
mean         8.554545  
std          0.460165  
min          7.560000  
25%          8.190000  
50%          8.530000  
75%          8.770000  
max          9.730000  
Name: global_avg_temp, dtype: float64
```

```
In [17]: df['local_avg_temp'].describe()
```

```
Out[17]: count      165.000000  
mean        14.450788  
std         0.504885  
min         13.220000  
25%         14.120000  
50%         14.390000  
75%         14.760000  
max         16.230000  
Name: local_avg_temp, dtype: float64
```

Fig. 4. Temperature trends

DIFFERENCES BETWEEN GLOBAL AND LOCAL TEMPERATURES

- San Jose has a much warmer temperature all throughout the year as compared to the global averages.
- San Jose has a 3.22 degree Celsius (max – min = 3.22) average temperature range while global average temperature range (max – min = 2.17) was about 2.17 degree Celsius.

SIMILARITIES BETWEEN GLOBAL AND LOCAL TEMPERATURES

- Both San Jose and the global temperatures have almost the same curves indicating similar trends over the past years.
- Both San Jose and Global average temperatures have has a steady increase in the temperatures indicating global warming.