Data Analyst Nanodegree

Project Submission

Explore Weather Trends

Submitted by

Dishank Kalra

Explore-Weather-Trends | Github Repository

Dishank Kalra | LinkedIn

Extracting Data - To extract the dataset I used SQL Workspace provided in Udacity's web application. SQL Queries used for extracting data are as follows:

1. city_data - city_data.sql | (github.com)

2. global_data - global_data.sql | (github.com)

3. city_list - city_list.sql | (github.com)

4. list_Indian_cities - list_Indian_cities.sql | (github.com)

5. global_city_data (To find trend between Hyderabad and global Temperature) global_city_data.sql | (github.com)

Data Exploration - After Extracting data I used python to manipulate and analyze the data. In python, I used libraries like NumPy and Pandas for the exploration and analysis of data. For visualization, I used libraries like Seaborn and Matplotlib.
The city chosen for analysis was Hyderabad, India. There were two cities with name Hyderabad in India and Pakistan. The number of rows present in the data of Hyderabad, India, and global are 218 (Link to cell) and 266 Link to cell respectively.
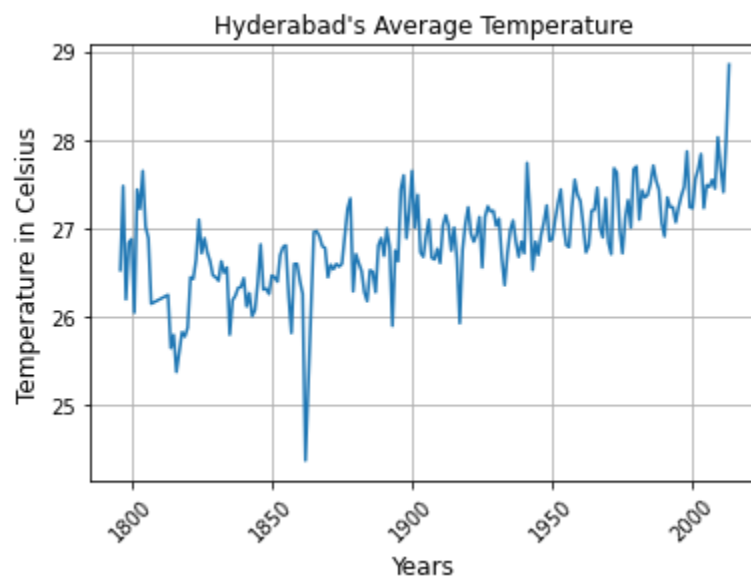
Line chart of Local Average Temperature Link to cell
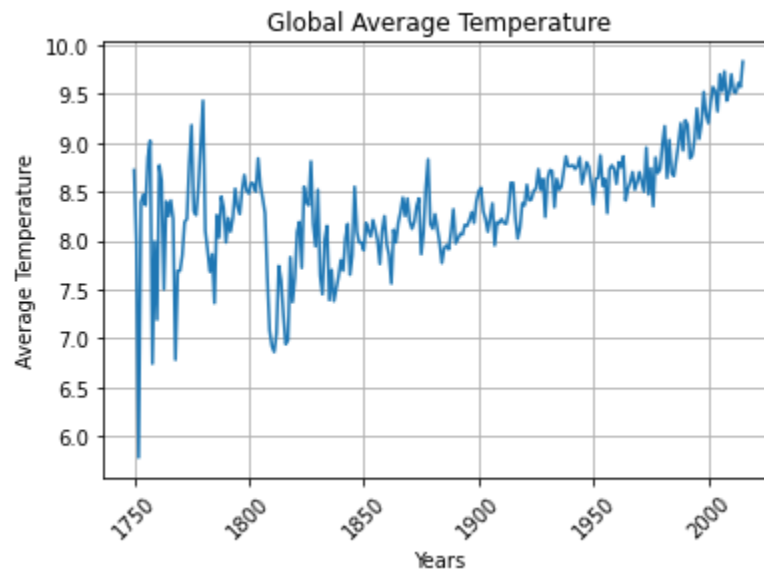
Line chart of Global Average Temperature [Link to cell](#)

Missing Values - In the data of Hyderabad there were 7 missing values present in the fourth column i.e. "avg_temp" [Link to cell](#)

Handling Missing Values - To handle missing values in the data I filled the missing values with the two methods instead of removing the rows with missing values. Two methods are as follows:
- Filled missing values with mean value of that column [Link to cell](#)
- Used interpolate method present in pandas to fill missing values [Link to cell](#)

Calculating Moving Averages -

- I have written a function, running_avg(window_size, rows, prev, new) to find running averages of average temperature [Link to cell](#)

  running_avg(window_size, rows, prev, new):
  window_size: Size of window to be considered for finding average
  rows: Number of rows in data
  prev: Column of avg_temp over which running averages have to be calculated
  new: Column with moving averages

```
1 def running_avg(window_size,rows,prev,new):
2   for i in range(rows - window_size + 1):
3     calc_running_avg = 0
4     for j in range(i, i + window_size):
5       calc_running_avg += prev.iloc[j]
6     new.iloc[i + window_size - 1] = calc_running_avg/window_size
7   return check_data
```

```
1 new = check_data['moving_avg']
2 prev = check_data['avg_temp']
3 rows = check_data.shape[0]
4 window_size = 25
```

- Calculating moving average using rolling method present in pandas [Link to cell](#)

Comparing Hyderabad's and Global Temperature -

- To compare the trend between local(Hyderabad) and global temperature we need to find years that are common in Hyderabad's and Global temperature. To find such data I wrote a SQL query, [global_city_data.sql | (github.com)](#)

```
1   /*
2   1. Comparing city(Hyderabad,India) and global data
3   2. To understand trends we have to compare temp. of same year
4   3. After getting desired dataset, will fill missing values
5   Note: Hyderabad is a city present in India and Pakistan both. I have chosen city present in India
6   */
7   -- Query:
8   SELECT
9     city_data.year,
10    city_data.avg_temp as city_temp,
11    global_data.avg_temp as global_temp
12  FROM city_data, global_data
13  WHERE city_data.year = global_data.year /* Selecting observations of same year */
14    AND city_data.city = 'Hyderabad'
15    AND city_data.country = 'India'
```

Snippet of SQL Query

- After extracting the desired data, there were 7 missing values in the column of city_temp, to fill the missing values I used interpolate method. Index of missing values are [12, 13, 14, 15, 16, 67, 68], Link to cell

	year	city_temp	global_temp
12	1808	NaN	7.63
13	1809	NaN	7.08
14	1810	NaN	6.92
15	1811	NaN	6.86
16	1812	NaN	7.05
67	1863	NaN	8.11
68	1864	NaN	7.98

Snippet of Missing values

Snippet of data after filling missing values using interpolation

- After filling in missing values, I calculated the moving average of city_temp and global_temp represented with column names, MA_city_temp and MA_global_temp respectively. Window size for calculating moving average was 8 Link to cell

```
year               1808.000000
city_temp            26.166667
global_temp           7.630000
Name: 12, dtype: float64

year               1809.000000
city_temp            26.183333
global_temp           7.080000
Name: 13, dtype: float64

year               1810.00
city_temp            26.20
global_temp           6.92
Name: 14, dtype: float64

year               1811.000000
city_temp            26.216667
global_temp           6.860000
Name: 15, dtype: float64

year               1812.000000
city_temp            26.233333
global_temp           7.050000
Name: 16, dtype: float64

year               1863.00
city_temp            25.24
global_temp           8.11
Name: 67, dtype: float64

year               1864.00
city_temp            26.10
global_temp           7.98
Name: 68, dtype: float64
```

	year	city_temp	global_temp	MA_city_temp	MA_global_temp
0	1796	26.53	8.27	NaN	NaN
1	1797	27.48	8.51	NaN	NaN
2	1798	26.20	8.67	NaN	NaN
3	1799	26.84	8.51	NaN	NaN
4	1800	26.88	8.48	NaN	NaN
5	1801	26.05	8.59	NaN	NaN
6	1802	27.44	8.58	NaN	NaN
7	1803	27.22	8.50	26.8300	8.51375
8	1804	27.65	8.84	26.9700	8.58500
9	1805	27.02	8.56	26.9125	8.59125

Snippet of Moving Average Temperature

Observation -

Moving Average Temperature comparison between Hyderabad and World -
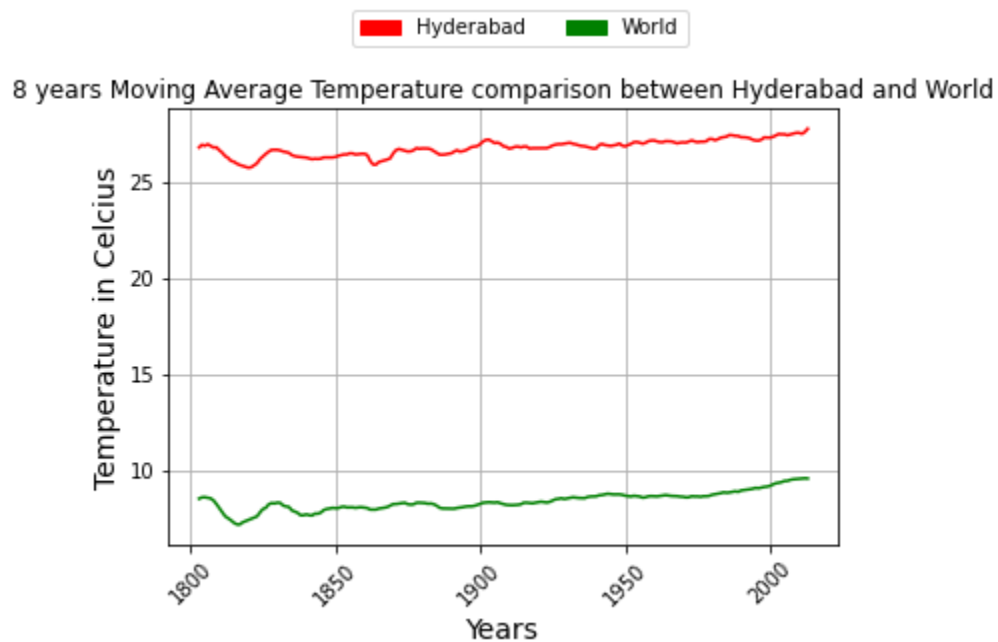
Link to cell

Comparison between Hyderabad and World's average temperature -

Link to cell

- The global average temperature varies between 8.27 to 9.61 degrees celsius between 1796 and 2013

- Hyderabad's average temperature varies between 26.53 to 28.85 degrees celsius between 1796 and 2013

- Hyderabad city is hotter than the Global Average temperature.

- There is an upward trend in temperature, hence the World is becoming hotter with every passing year.

	year	city_temp	global_temp	MA_city_temp	MA_global_temp
60	1856	26.24	8.00	26.51500	8.06500
61	1857	25.82	7.76	26.46000	8.03750
62	1858	26.60	8.10	26.47625	8.06250
63	1859	26.60	8.25	26.49500	8.07125
64	1860	26.41	7.96	26.49625	8.05375
65	1861	26.26	7.85	26.44125	8.03000
66	1862	24.38	7.56	26.14000	7.94875
67	1863	25.24	8.11	25.94375	7.94875
68	1864	26.10	7.98	25.92625	7.94625
69	1865	26.96	8.18	26.06875	7.99875

[Link to cell](#)

- There was a sudden decrease of **two degrees** in the average temperature of Hyderabad from **26.26** (in 1861) to **24.38** (in 1862) but there was no significant change in the Global Average temperature.

	year	city_temp	global_temp	MA_city_temp	MA_global_temp
year	1.000000	0.665212	0.765267	0.848356	0.859065
city_temp	0.665212	1.000000	0.771909	0.771301	0.745213
global_temp	0.765267	0.771909	1.000000	0.779534	0.881786
MA_city_temp	0.848356	0.771301	0.779534	1.000000	0.920007
MA_global_temp	0.859065	0.745213	0.881786	0.920007	1.000000

[Link to cell](#)

- The moving average of city Temperature is highly correlated with the Moving average of Global Temperature(correlation=0.92), so we can conclude that the temperature trend of both City and World is proportional.