UDACITY DATA ANALYST NANODEGREE

PROJECT 1

EXPLORING WEATHER TRENDS

UDACITY HONOR CODE

I confirm that this submission is my own work. I have not used code from any other Udacity student's or graduate's submission of the same project. I have correctly attributed all code I have obtained from other sources, such as websites, books, forums, blogs, GitHub repos, etc. I understand that Udacity will check my submission for plagiarism, and that failure to adhere to the Udacity Honor Code may result in the cancellation of my enrollment.

AIM

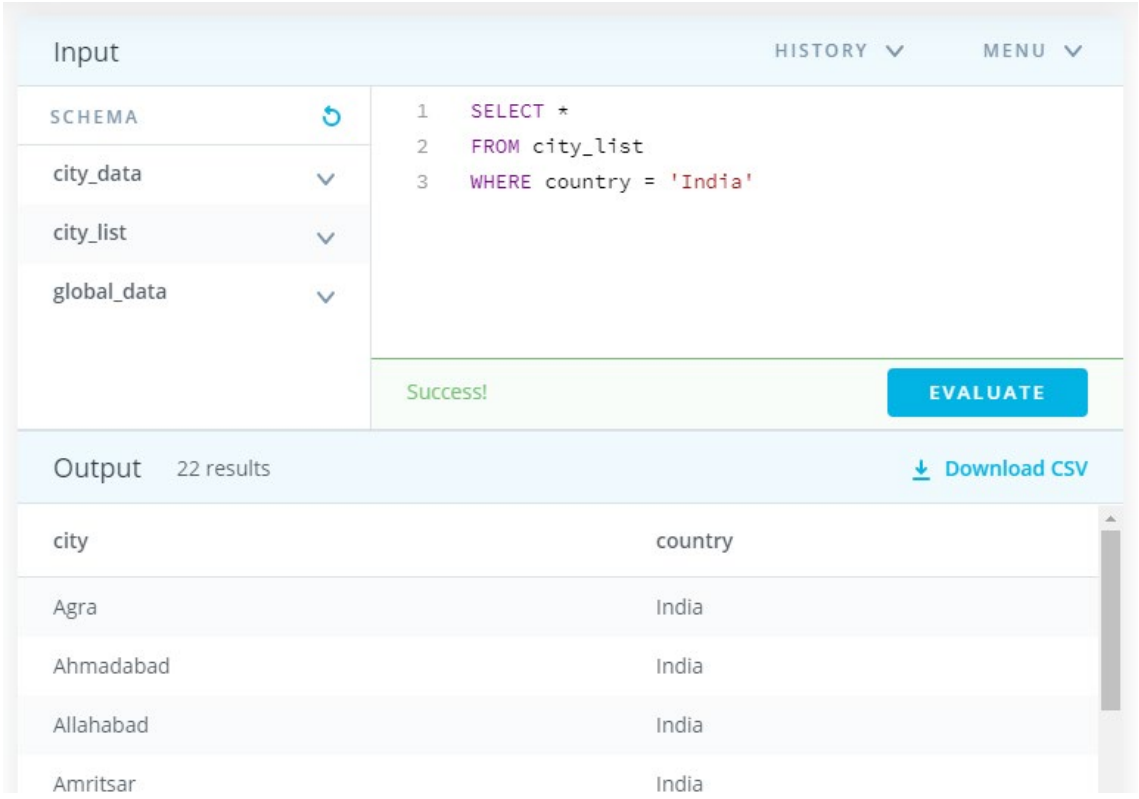
In this project, I will analyze local and global temperature data and compare the temperature trends of where I live to overall global temperature trends.

STEPS TAKEN

1. EXTRACTING THE DATA

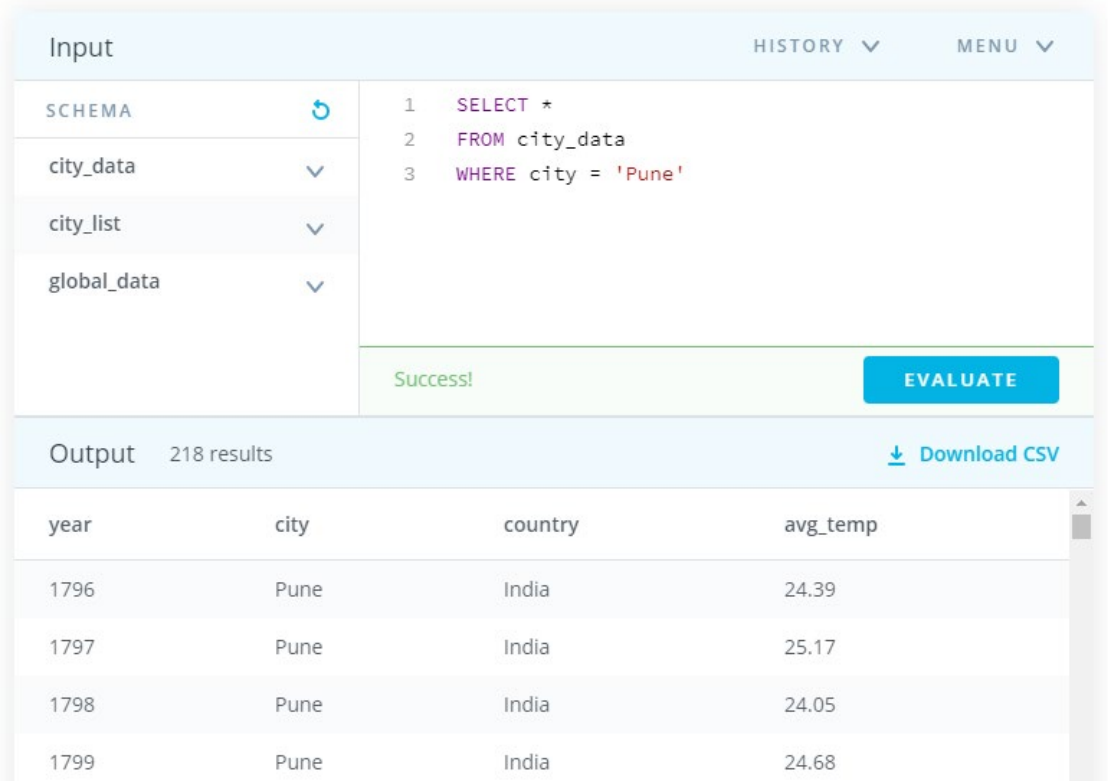
I have used SQL queries to extract data from Udacity database. I have used commands **SELECT**, **FROM**, **WHERE** and **ORDER BY**.

- I live in India hence I filtered a list of cities based on the country field. A manual search was needed to find the closest city in the list using the table 'city_list'.

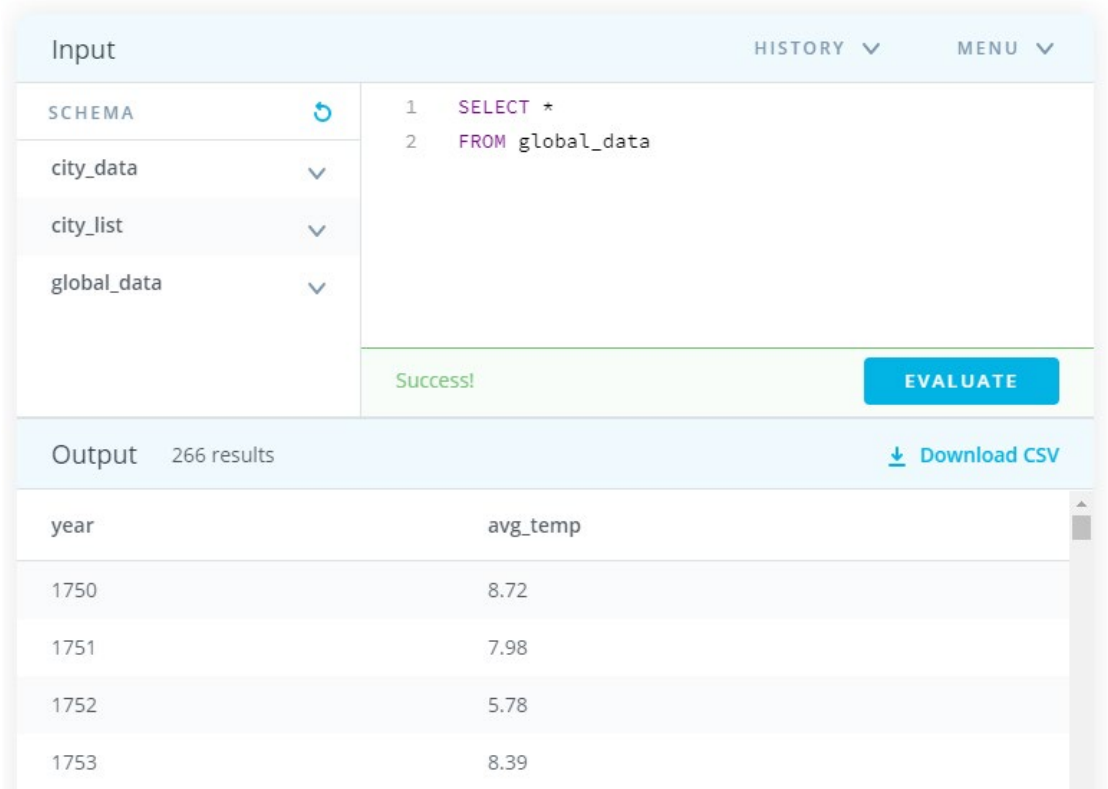


Maharshi Saurabh Mishra

- Then, I filtered the data based on the found city i.e. ‘Pune’ in the table ‘city_data’. And downloaded the CSV file containing this data.

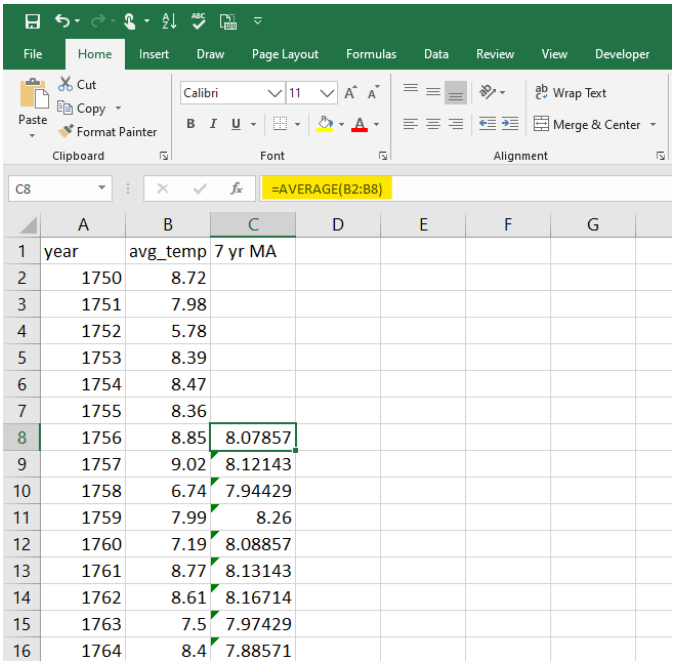


- Finally, I entered a query for extracting the global temperature data from the table ‘global_data’ and downloaded the CSV containing this data.



2. COPYING THE DATA TO MICROSOFT EXCEL

- I copied the data from the CSV to the Microsoft Excel program and found out the average for first 7 years and dragged the fill handle to the bottom. I did it for both global and cities data.



File Home Insert Draw Page Layout Formulas Data Review View Developer

Cut Copy Format Painter Paste Clipboard | Calibri 11 Font | Wrap Text Merge & Center Alignment

C8 =AVERAGE(B2:B8)

	A	B	C	D	E	F	G
1	year	avg_temp	7 yr MA				
2	1750	8.72					
3	1751	7.98					
4	1752	5.78					
5	1753	8.39					
6	1754	8.47					
7	1755	8.36					
8	1756	8.85	8.07857				
9	1757	9.02	8.12143				
10	1758	6.74	7.94429				
11	1759	7.99	8.26				
12	1760	7.19	8.08857				
13	1761	8.77	8.13143				
14	1762	8.61	8.16714				
15	1763	7.5	7.97429				
16	1764	8.4	7.88571				

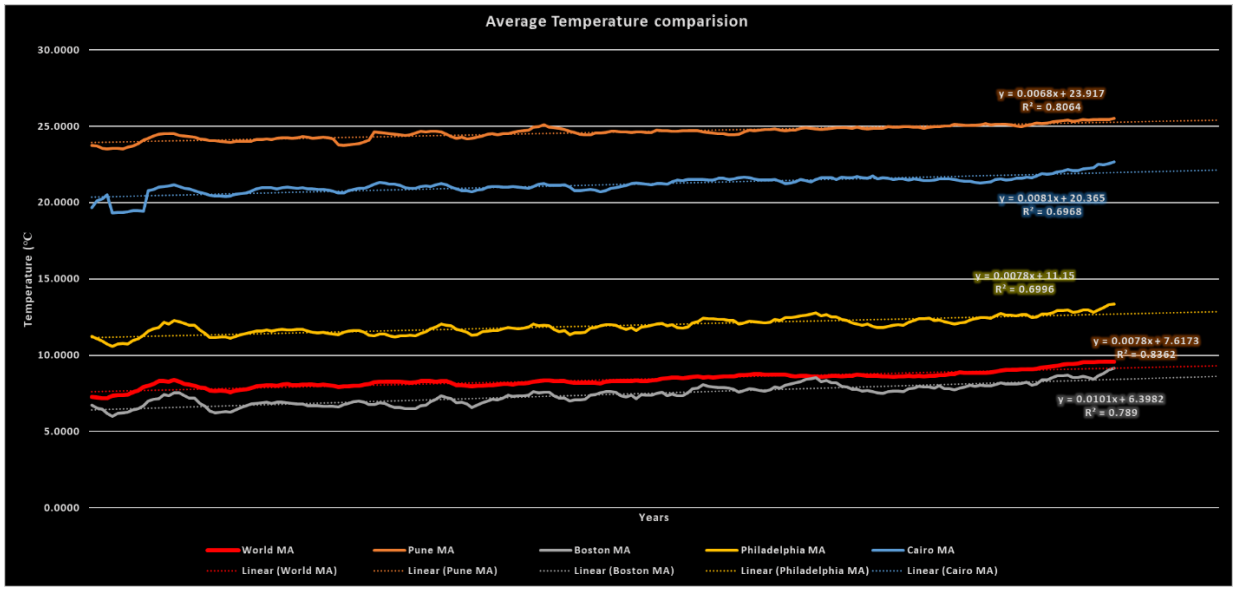
- I have added multiple cities to get a better understanding from the data trends. Along with the global and Pune data, I have included Boston, Philadelphia, and Cairo and found out their moving average. Then, I aggregated the moving averages data in one sheet as shown below from 1814 for the sake of uniformity since Cairo's moving average is available from 1814.

	A	B	C	D	E	F	G
1	Year	World MA	Pune MA	Boston MA	Philadelphia MA	Cairo MA	
2	1814	7.2671	23.7600	6.7143	11.2357	19.6829	
3	1815	7.2114	23.7000	6.5457	11.0614	20.1386	
4	1816	7.1914	23.5575	6.4400	10.9414	20.2300	
5	1817	7.2000	23.5200	6.2014	10.7400	20.5086	
6	1818	7.3386	23.5667	5.9843	10.5757	19.3086	
7	1819	7.3843	23.5629	6.2000	10.7471	19.3629	
8	1820	7.3671	23.5114	6.2271	10.7614	19.3729	
9	1821	7.4386	23.6186	6.2500	10.7300	19.4014	
10	1822	7.5743	23.7143	6.4271	10.9557	19.4614	
11	1823	7.6857	23.8786	6.4886	11.0814	19.4900	
12	1824	7.9100	24.0886	6.6986	11.3129	19.4271	
13	1825	7.9900	24.1943	6.9971	11.5886	20.7700	
14	1826	8.1314	24.3514	7.1186	11.7286	20.8600	
15	1827	8.3014	24.4900	7.1414	11.7957	21.0100	
16	1828	8.3129	24.5129	7.4171	12.1457	21.0614	
17	1829	8.2771	24.5100	7.3271	12.0443	21.0886	
18	1830	8.3914	24.5214	7.5357	12.2743	21.1657	
19	1831	8.2614	24.4129	7.5514	12.2029	21.0343	
20	1832	8.1271	24.3771	7.3614	12.0700	20.9200	
21	1833	8.0771	24.3271	7.2071	11.9629	20.9014	
22	1834	7.9829	24.2871	7.2000	11.9571	20.7671	
23	1835	7.8714	24.1671	6.8457	11.6471	20.6500	
24	1836	7.8371	24.1286	6.6371	11.4429	20.6043	
25	1837	7.6743	24.0657	6.3329	11.1643	20.4643	
26	1838	7.6557	24.0529	6.2200	11.1700	20.4200	
27	1839	7.6814	24.0171	6.2600	11.1786	20.4529	
28	1840	7.6514	23.9929	6.2914	11.1871	20.4171	
29	1841	7.5857	23.9229	6.2800	11.1200	20.4486	
30	1842	7.6757	24.0114	6.4171	11.2214	20.5557	
31	1843	7.7429	24.0057	6.5629	11.3429	20.5900	

ALL-IN-ONE | WORLD | Pune | Boston | Philadelphia | Cairo

3. PREPARING A LINE GRAPH AND MODEL OF THE DATA TRENDS

- A line chart is made from the information in order to visualize the trends in the temperature across the years. A larger picture of the chart can be found at https://i.imgur.com/FkV9qqb.png.



- The chart is a simple line chart for multiple sets of data (moving averages of different various temperature data).
- The legends indicate the different color coding for different places. The dotted line for each line graph indicates the nearest fit according to the charted data.
- The R^2 value indicates how much of the line graph is explained by a given model.
- The dotted line plots the linear models for the corresponding dataset.
- The dotted line also forecasts the data trends to show future possibilities.
- Along with the R^2 value, the equation for the model is also given which the graph is compared to.

4. CALCULATING THE CORRELATION COEEFICIENT

- Correlation coefficient is calculated for two sets of data to understand the strength of their relationship. It is a number between +1 and -1. A value between 0.7 and 1.0 indicate strong positive relationship while a value between -1.0 and -0.7 indicate strong negative relationship.
- Correlation coefficient can be calculated in Excel by enabling the Analysis ToolPak Add-in. The dataset is then selected and correlation coefficient is calculated by selecting the particular option in the Data tab under Data Analysis. (MA stands for moving average)

	World MA	*Pune MA*	*Boston MA*	*Philadelphia MA*	*Cairo MA*
World MA	1				
Pune MA	0.931172	1			
Boston MA	0.913434	0.87512	1		
Philadelphia MA	0.909935	0.84646223	0.97515551	1	
Cairo MA	0.912939	0.86929633	0.84431229	0.849690092	1

OBSERVATIONS

- ✓ Looking at the data visualization for Pune, it can be concluded that this city is located near the equator since it has an average temperature a lot higher than world average temperature.
- ✓ The rising data model representing the line graph for each dataset indicate we are experiencing a rise in average temperature over the years.
- ✓ A rise in temperature is seen during the period of Industrial Revolution (till 1840).
- ✓ The dip in the temperature at the start refers to the event "Year Without a Summer" in 1816. The year 1816 is known as the Year Without a Summer (also the Poverty Year and Eighteen Hundred and Froze to Death) because of severe climate abnormalities that caused average global temperatures to decrease by 0.4–0.7 °C (0.72–1.26 °F). This resulted in major food shortages across the Northern Hemisphere. [1]
- ✓ A higher correlation coefficient is seen between Philadelphia and Boston than others as they are situated very close to each other and the local factors affect ting the region affected both the places. Fun fact: they are just 300 miles away by road.

REFERENCES

1. Year Without a Summer (https://en.wikipedia.org/wiki/Year_Without_a_Summer)