UDACITY

Logout

‹ Return to Classroom

Explore Weather Trends

REVIEW | HISTORY

Meets Specifications

Great job!

Hi Udacious,

Thanks for fixing all the issues mentioned by the previous reviewer. This is a perfect submission now. **Congratulations!**

You demonstrated your ability to retrieve data from a SQL Database and derive interesting, accurate results from the output of your query. You were further able to manipulate this data using external software and create a meaningful visualization to demonstrate your observed results. This is a tremendously important skill and will prove useful throughout your career in data analytics.

Before you move on to your next lessons, take pride in the effort you've put into this project. I hope you found this exercise both challenging and rewarding. Keep up the exceptional work and effort here, and I look forward to seeing you rock those future submissions!

Kind regards,

Analysis

✓

- **The SQL query used to extract the data is included.**
- **The query runs without error and pulls the intended data.**

Great work here in extracting the data for your local city and comparing that to global temperatures. Your queries were spot on!

You could also make a simple join to get the whole data in a single query. It would be something like below.

```
SELECT
gd.year,
cd.city,
cd.avg_temp Your_city,
gd.avg_temp World
FROM city_data cd
     JOIN global_data gd ON gd.year = cd.year
WHERE cd.city = 'Your City'
ORDER BY gd.year
```

You could get different cities in a single script with a code like below.

```
SELECT
gd.year,
max(gd.avg_temp) as global_temp,
max(case when cd.city='London' then cd.avg_temp else null end) as london_temp,
max(case when cd.city='Cairo' then cd.avg_temp else null end) as cairo_temp,
max(case when cd.city='New York' then cd.avg_temp else null end) as new_york_temp
FROM city_data cd
     JOIN global_data gd ON gd.year = cd.year
WHERE cd.city in ('London','Cairo','New York')
GROUP BY gd.year
ORDER BY gd.year
```

If you're interested in bolstering your SQL mastery with more questions and puzzles, here are a couple of websites I often enjoy to looking for extra coding practice for SQL:

- https://www.hackerrank.com/domains/sql/select

You'll get a chance to practice increasingly difficult questions and learn how to interact with multiple tables at once.

✓ **Moving averages are calculated to be used in the line chart.**

Excellent work here in calculating the 7-year moving average for both Riyadh and Global temperatures.

Just to let you know that you can also calculate moving averages in SQL using window functions as shown below.

```
select
*,
avg(avg_temp) over (order by year asc ROWS 9 preceding) as moving_average_10
from city_data
where city='Your City'
order by year asc
```

If you used Jupyter notebooks to plot the data, you could get the moving averages as below.

```
df['moving_average_10'] = df['temperature'].rolling(10).mean()
```

To learn more about rolling function and see other examples, you can check:

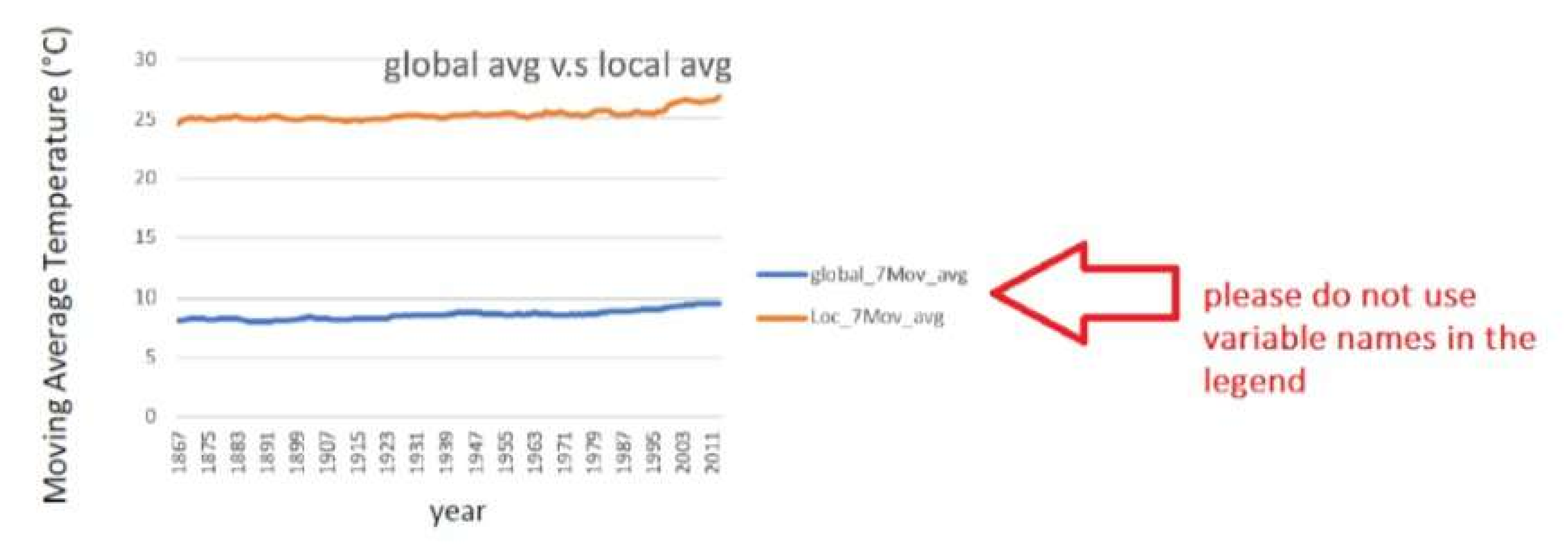
- https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.rolling.html

✓

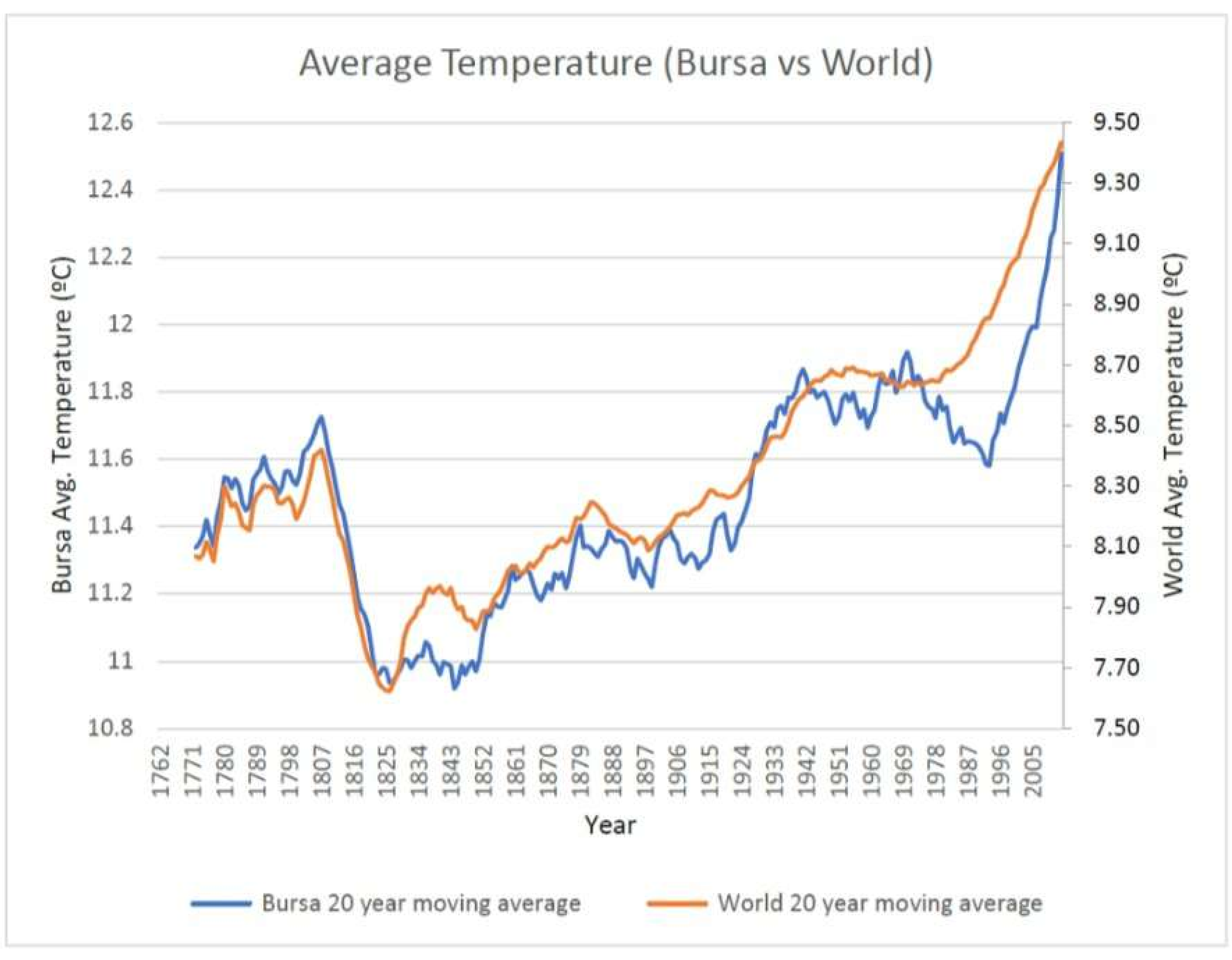
- **A line chart is included in the submission.**
- **The chart and its axes have titles, and there's a clear legend (if applicable).**

The line chart included in your submission looks nice! The chart contains a clearly represented title that explains the details of the presented line graph. I especially appreciate that you rotated the year values on the x-axis so that they don't overlap one another - or maybe this was automatic :). You even added the unit in which temperature is measured in the y-axis label. This attention to detail really goes a long way to help communicate your results to an audience.

One area of improvement I may suggest is not to use variable names in the legend as it would be difficult for a viewer to understand. Instead of "**Loc_7Mov_avg**", please use something like "**Riyadh 7-year moving average temperature**".



Also, this is very subjective but since the temperature difference is mostly stable, you could use a double y-axes plot. When I was doing the project, I did such a visualization which made it easier for me to see the changes (but as I said, this is subjective and there is no ground truth).



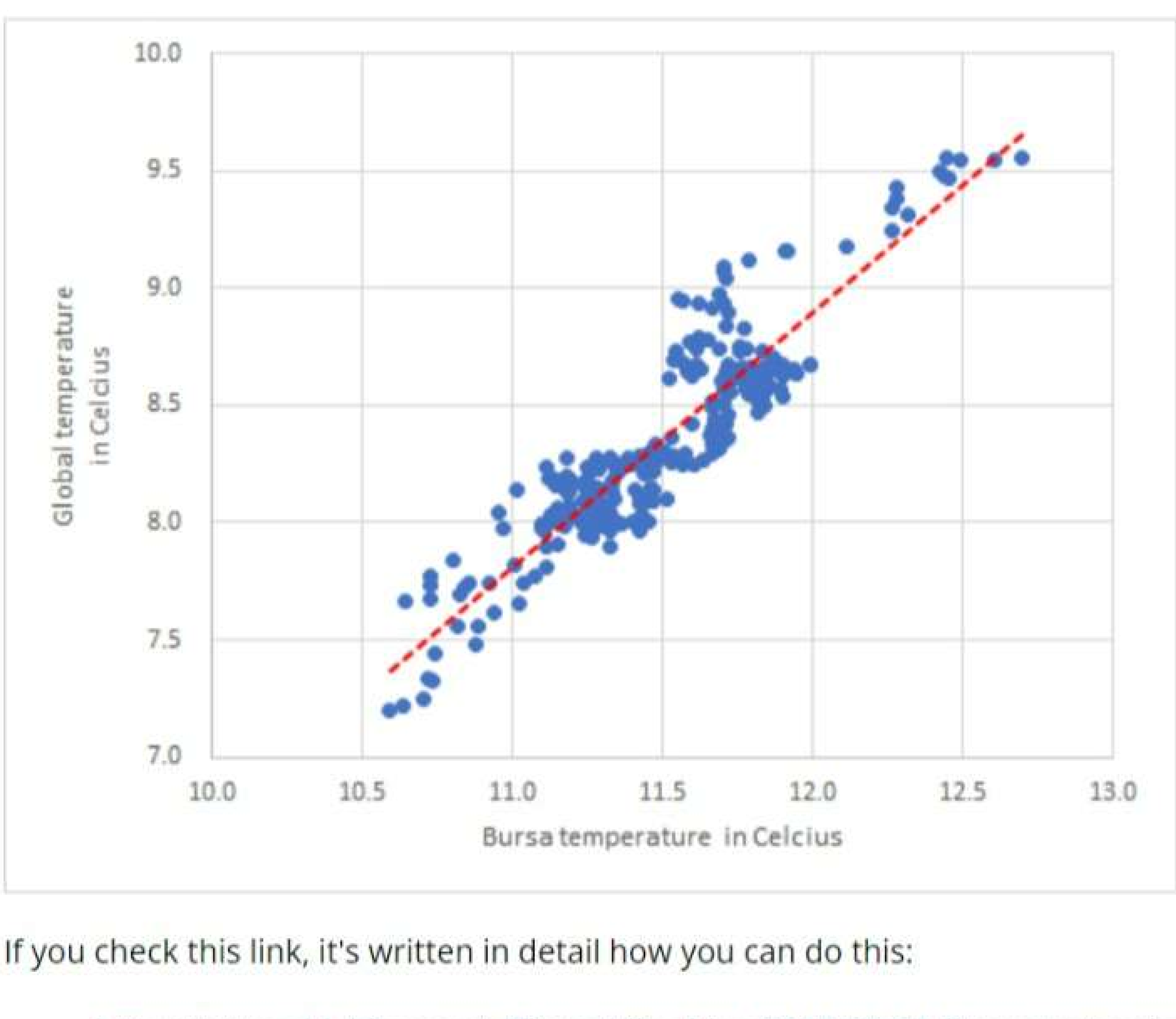
✓

- **The student includes four observations about their provided data visualization.**
- **The four observations are accurate.**

Great work in looking at the output of your graph and making observations from them.

You've done a nice job in your first project. One key thing we do in data analysis when comparing 2 numerical variables is check for their correlation.

In most cases, the global temperature is highly correlated with the local temperature. You can calculate this correlation using the "**correl**" function in excel. For my city the correlation was 0.91 which means that we can build a regression model to predict my local temperature using the global temperature with a high success. See the scatterplot below. The regression line is the red line we are trying to find.



If you check this link, it's written in detail how you can do this:

- https://www.ablebits.com/office-addins-blog/2018/08/01/linear-regression-analysis-excel/

My results were as below which means that we can predict local temperature for my city with this formula

> My local temperature = Global Temperature * 0.7769 + 5.005

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.917438817
R Square	0.841693982
Adjusted R Square	0.841060758
Standard Error	0.14897754
Observations	252

ANOVA

	df	SS	MS	F	Significance F
Regression	1	29.50111334	29.50111334	1329.219812	4.7513E-102
Residual	250	5.548576892	0.022194308		
Total	251	35.04969023			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	5.0051734	0.178163999	28.09306836	2.60486E-79	4.654279691	5.35606711	4.654279691	5.35606711
X Variable 1	0.776934263	0.021310119	36.45846695	4.7513E-102	0.734964018	0.818904507	0.734964018	0.818904507

As a data analyst, most of the time, we will try to find relations like these in the dataset.

This is a nice starter for regression and you can calculate the same for your city. Good luck!

DOWNLOAD PROJECT

RETURN TO PATH