UDACITY

DATA ANALYST NANODEGREE PROGRAM

EXPLORING WEATHER TRENDS

PROJECT REPORT

Stephanie J. IRVINE

May 29, 2020

Contents

1 | Introduction

The aim of this project was to analyse both local and global temperature data and compare the temperature trends in my local city to overall global temperature trends. This report intends to effectively describe the steps taken to prepare data visualised in a line chart and indicates the tools used in each stage. Further, it demonstrates how the moving average was calculated and provides justification for the decisions made when visualising the trends. Additionally, this report includes the data visualisation and interprets this data to present observations and insights.

The overall project process can be summarised in the steps below.

Steps:

1. Extract the data from the database using SQL and export to CSV
2. Open the CSV using Excel and explore the data
3. Create a line chart in Excel to compare Edinburgh temperatures with global temperatures by plotting the moving average
4. Make observations about the trends, similarities and differences between the global averages and Edinburgh's averages

2 | Data Extraction

The first step was to extract the relevant data from the temperatures database using the SQL Workspace provided in the Udacity online classroom. The results were then exported to CSV. There were three tables in the database: city_list which contained a list of cities and countries in the database, city_data which held the average temperatures (ºC) for each city by year, and global_data which contained the average global temperatures (ºC) by year.

2.1 Finding the Closest Big City

The first step in extracting the relevant data was to explore the city_list table to discover which cities were included in the table. First the table was expanded to see which fields were included and it contained only city and country columns and then a SELECT * was performed on the city_list table to understand whether the data would be messy, to check if the case was upper or lower, and to check how the country I live in would be listed in the database as it could have been Scotland, United Kingdom, The United Kingdom, UK, Great Britain and so on. This initial SELECT * statement returned 342 results and after scrolling through the dataset it was clear this data was tidy and contained the 'United Kingdom'. An SQL query was executed as shown in Figure 1 using a SELECT statement on the city_list table and setting the WHERE clause to be equal to the United Kingdom to filter the selection and retrieve only cities listed in the UK. This query returned 5 results as demonstrated by Figure 1 with Edinburgh being the only Scottish city listed and my closest city held in the database.

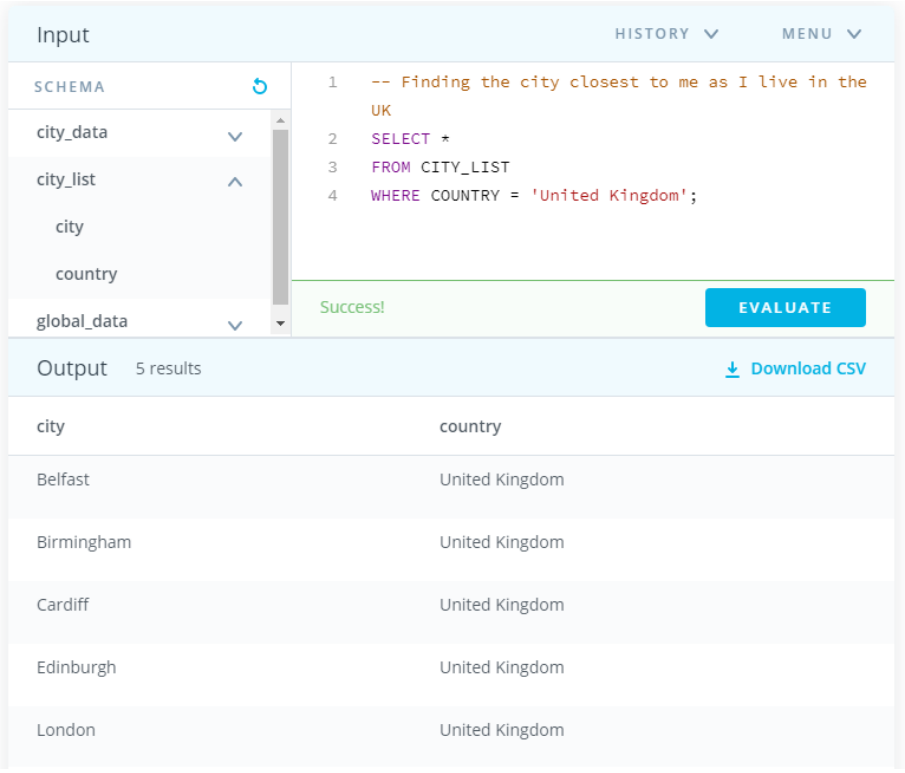


Figure 1: SQL query to find my closest city in the database

2.2 Extracting Edinburgh City Level Data

Once Edinburgh was selected as the city of interest from the database, a query to extract city level data from the city_data table was performed. Again the city_data table was expanded to initially gauge the fields included and then a query to extract all fields (year, city, country, avg_temp) for Edinburgh in the UK was executed by adding a WHERE clause to specify the city as Edinburgh and the country as the United Kingdom. As illustrated in Figure 2 this query returned 271 results therefore it was possible to scroll through the output quickly to gain insight into the time periods included in the dataset and ensure the result was what was expected from the query. The CSV was then downloaded so the data would be ready for exploration and analysis after the other necessary queries to extract all relevant data from the database had been executed.

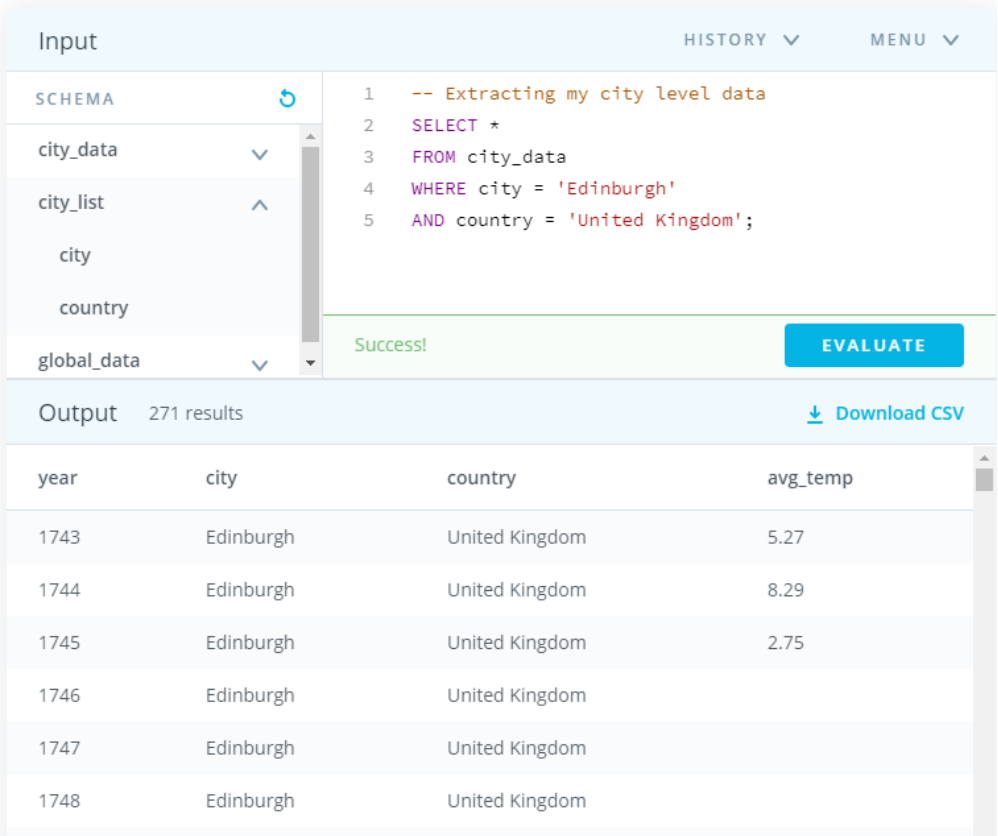


Figure 2: SQL query to extract Edinburgh data from the database

2.3 Extracting the Global Data

The final stage in data extraction was to extract the global data and export this to CSV. Before writing a query the first step was to expand the global_data table to see what columns comprised the table, these included just year and avg_temp. A SELECT * statement was performed as shown from Figure 3 and 266 results were outputted. The data was then exported to CSV. Following this final stage in data extraction all the relevant data had been gathered with these simple SQL queries and data exploration could commence.

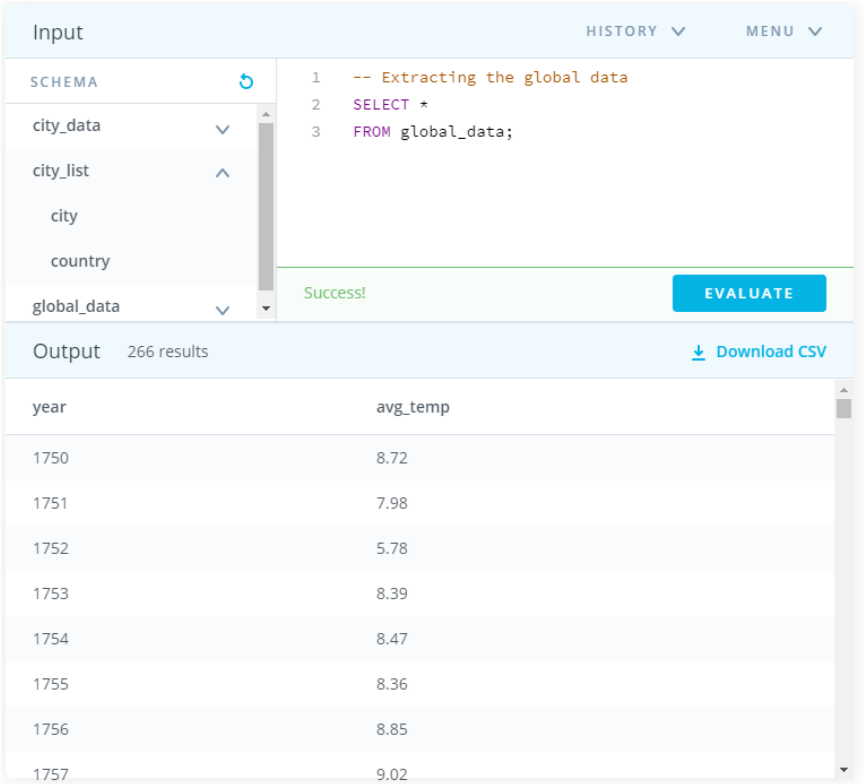


Figure 3: SQL query to extract global data from database

3 | Data Exploration

Once both the Edinburgh and global outputs were exported to CSV the data could be explored using Excel spreadsheets. After briefly examining the spreadsheets these were combined into one spreadsheet for improved efficiency and ease of comparison between Edinburgh data and city data. When investigating the data, the first point to note was that the Edinburgh dataset had data from the years 1743-2013 while the global dataset had data from the years 1750-2015. It was also important to note that there were 4 missing values from the

Edinburgh dataset from the years 1746-1749, therefore in order to clean the data only the years 1750-2013 were used for both datasets in the visualisation. To gain a better understanding of the data the mean was calculated for both the global and Edinburgh average temperatures and was found to be 8.359°C and 7.603°C respectively. Additionally, the median was found to be 8.365°C for the global average temperatures and 7.595°C for the Edinburgh average temperatures.

4 | Data Visualisation

Once the data had been cleaned the first step in the data visualisation process was to consider the period which should be selected for the moving average. It was important to plot the moving averages in this line chart to smooth out the data so that long term trends could be more easily observed [1]. A 5-year moving average was selected for this data ranging from 1750 to 2013 as this produced a clear and insightful visualisation which highlighted overall trends. To obtain the moving average from the Excel spreadsheet the AVERAGE () function was employed as described in the Udacity classroom. The line chart created is shown in Figure 4. The x-axis was formatted to span from the years 1750 to 2020 and the y-axis was formatted to show average temperatures between 6-10°C to highlight the data points more effectively.

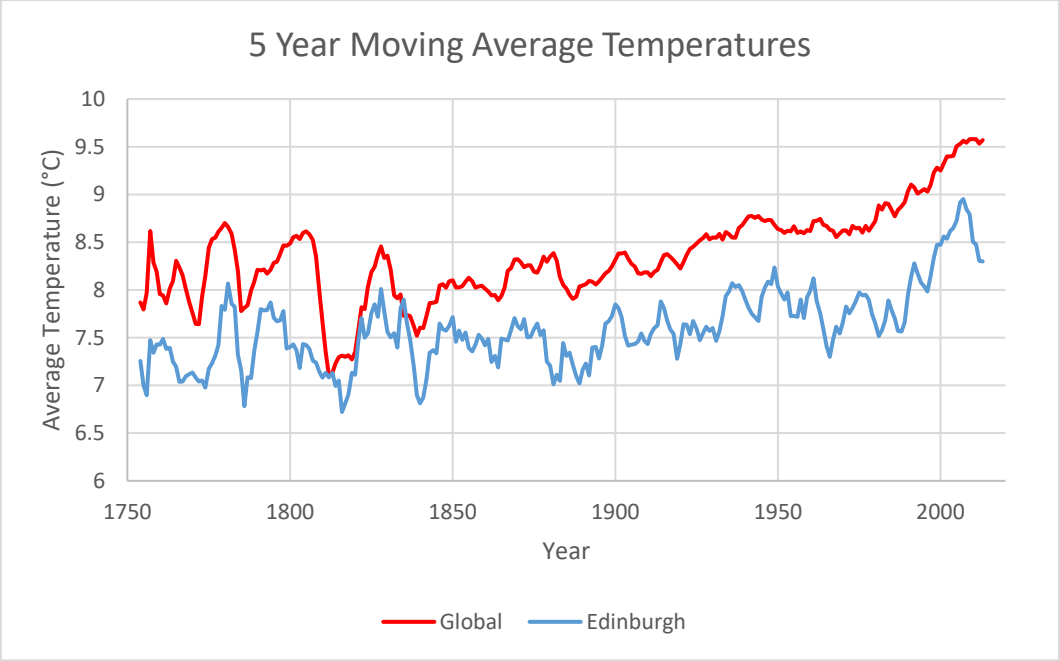


Figure 4: Line Chart of Global and Edinburgh Moving Average Temperatures

5 | Analysis and Discussion

From Figure 4 it is evident that Edinburgh is cooler on average compared to the global average temperature. This was expected from the initial data exploration stage where the median and mean were lower for Edinburgh's average temperatures compared to global average temperatures. Additionally, it can be inferred that the difference between Edinburgh's average temperature and the average global temperature has been relatively consistent over time with the exception of the time period 1810-1840 where the average temperature in Edinburgh is very close to the global average and even slightly surpasses the world average in 1835. However, on average Edinburgh is approximately 0.5-1.0°C cooler than the global temperature. Edinburgh's average temperature and the global average temperature both experience a decrease in temperature between 1807 to 1813, however the global temperature average decrease is more significant than the more gradual decrease exhibited by Edinburgh's average temperatures.

The trend in Edinburgh's average temperatures over time are similar to the global average temperature overall showing variance until around 1900 and then an overall increase from 1900 – 2000s. After 1900, the average temperature trend for both Edinburgh and global temperatures shows an overall increase however Edinburgh's temperature fluctuates more significantly than the global average temperature each year, with Edinburgh's average temperature decreasing slightly from 2009 to 2013. It is evident from this data that both the world and Edinburgh are becoming hotter overall which is expected due to the effects of global warming.

6 | Conclusions

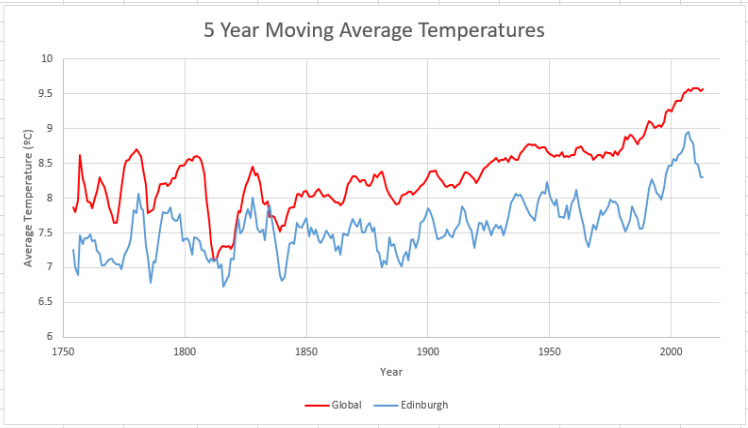
This report has discussed the steps taken to extract data from a database using SQL, covered data exploration and cleaning, described the process of data visualisation in Excel and has analysed the data and made observations about the similarities and differences between Edinburgh and global average temperatures.

7 | References

[1] Udacity. Moving Averages [Internet]. 2020 [cited 29 May 2020]. Available from: https://classroom.udacity.com

8 | Appendix – Spreadsheet with Global and Edinburgh Data

year	avg_temp_glob	avg_temp_edi	Global	Edinburgh		Mean Global	Mean Edi	Med Global	Med Edi
1750	8.72	8.41				8.359393939	7.60303	8.365	7.595
1751	7.98	8.16							
1752	5.78	4.78							
1753	8.39	7.51							
1754	8.47	7.42	7.868	7.256					
1755	8.36	7.14	7.796	7.002					
1756	8.85	7.64	7.97	6.898					
1757	9.02	7.65	8.618	7.472					
1758	6.74	6.87	8.288	7.344					
1759	7.99	7.85	8.192	7.43					
1760	7.19	7.13	7.958	7.428					
1761	8.77	7.93	7.942	7.486					
1762	8.61	7.13	7.86	7.382					
1763	7.5	6.92	8.012	7.392					
1764	8.4	7.13	8.094	7.248					
1765	8.25	6.87	8.306	7.196					
1766	8.41	7.14	8.234	7.038					
1767	8.22	7.14	8.156	7.04					
1768	6.78	7.2	8.012	7.096					
1769	7.69	7.24	7.87	7.118					
1770	7.69	6.96	7.758	7.136					
1771	7.85	6.86	7.646	7.08					
1772	8.19	6.95	7.64	7.042					
1773	8.22	7.25	7.928	7.052					
1774	8.77	6.86	8.144	6.976					
1775	9.18	7.95	8.442	7.174					
1776	8.3	7.15	8.532	7.232					
1777	8.26	7.34	8.546	7.31					
1778	8.54	7.82	8.61	7.424					
1779	8.98	8.9	8.652	7.832					
1780	9.43	7.75	8.702	7.792					
1781	8.1	8.53	8.662	8.068					
1782	7.9	6.28	8.59	7.856					
1783	7.68	7.66	8.418	7.824					
1784	7.86	6.36	8.194	7.316					



Joined_Data