

# Global Temperature Analysis with KNIME

By: Mohamed Almatbaagy

---

## Topics

1. Overview
  2. Key Tasks
  3. Tools & Resources Used
  4. Workflow Summary
  5. Conclusion
- 

## 1. Overview

This project solves a real-world data analytics case using the KNIME platform, focusing on temperature trends across countries over the past 270 years.

**Given a dataset (in 2 CSV files) representing the history of temperatures of the world in around 270 years**

---

## 2. Key Tasks

### Using KNIME Analytics Software:

6. Output a table that has the overall average of each country
  7. Classify the countries Temperature into “Low/Mid/High”
  8. Output a table that has the difference between the average of the country in each year and the average global temp in the last 24 years
  9. Output a table that shows the top 5 countries that have the largest difference from the global Temp
  10. Draw a histogram for the yearly global temperatures
  11. Draw a chart to compare between any city and global average temperature over the past years
- 

## 3. Tools & Resources Used

1. KNIME Analytics Platform
  2. Google
  3. YouTube
  4. NodePit
  5. ChatGPT, DeepSeek
-

## 4. Workflow Summary

---

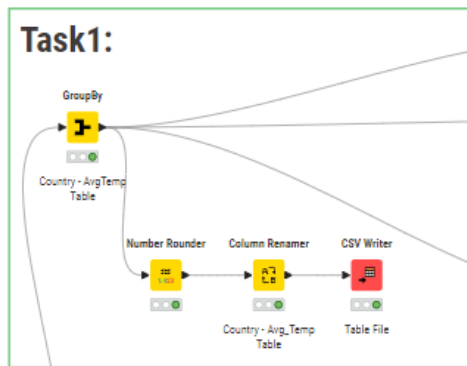
**Task1:** Output a table that has the overall average of each country

**Challenges & Decisions:**

- There were missing values in **avg\_temp** column from **city\_data table** so they were excluded during aggregation by **GroupBy** node
- Used **Number Rounder** to standardize temperature format for cleaner output

**Nodes Used:** CSV Reader, GroupBy, Number Rounder, Column Renamer, CSV Writer

**Workflow:**



**Output File:** Task1 Table.csv

**Sample Output:**

Rows: 137 | Columns: 2

#	RowID	country <small>String</small>	Avg_Temp <small>Number (double)</small>
<input type="checkbox"/> 1	Row0	Afghanistan	14.36
<input type="checkbox"/> 2	Row1	Albania	15.5
<input type="checkbox"/> 3	Row2	Algeria	16.43
<input type="checkbox"/> 4	Row3	Angola	23.69
<input type="checkbox"/> 5	Row4	Argentina	17.06
<input type="checkbox"/> 6	Row5	Armenia	8.37
<input type="checkbox"/> 7	Row6	Australia	15.86

---

**Task2:** Classify the countries Temperature into “Low/Mid/High”

**Challenges & Decisions:**

- Used **GroupBy & Expression** nodes to get Min, Max, Range and Bin limits
- Used **Number Rounder** to standardize temperature format for cleaner output

**Method 1:** Using Numeric Binner

**Challenges & Decisions:**

- The input values for bin limits in **Numeric Binner** need to be entered manually
- So I used **GroupBy & Expression** nodes to get Min, Max, Range and Bin limits

**Nodes Used:** CSV Reader, GroupBy, Expression, Numeric Binner, Column Renamer, CSV Writer

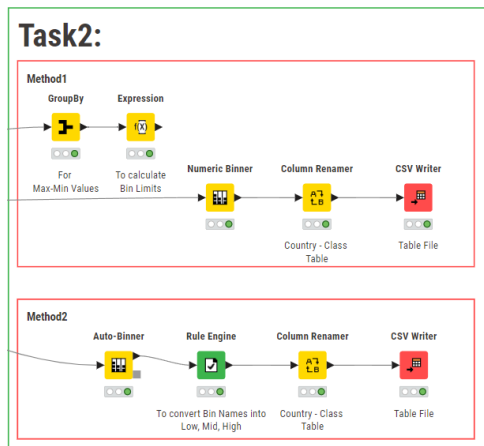
## Method 2: Using Auto Binner

### Challenges & Decisions:

- The **Auto Binner** does the calculations itself by specifying number of bins but can't write manual bin naming
- So I needed to use **Rule Engine** node to convert bin names into readable labels(Low, Mid, High)

**Nodes Used:** CSV Reader, GroupBy, Auto Binner, Rule Engine, Column Renamer, CSV Writer

### Workflow:



**Output File:** “Task2 Table1.csv” & “Task2 Table2.csv”

### Sample Output:

Rows: 137 | Columns: 2

#	RowID	country <i>String</i>	Class <i>String</i>
1	Row0	Afghanistan	Mid
2	Row1	Albania	Mid
3	Row2	Algeria	Mid
4	Row3	Angola	High
5	Row4	Argentina	Mid
6	Row5	Armenia	Mid
7	Row6	Australia	Mid

**Task3:** Output a table that has the difference between the average of the country in each year and the average global temp in the last 24 years

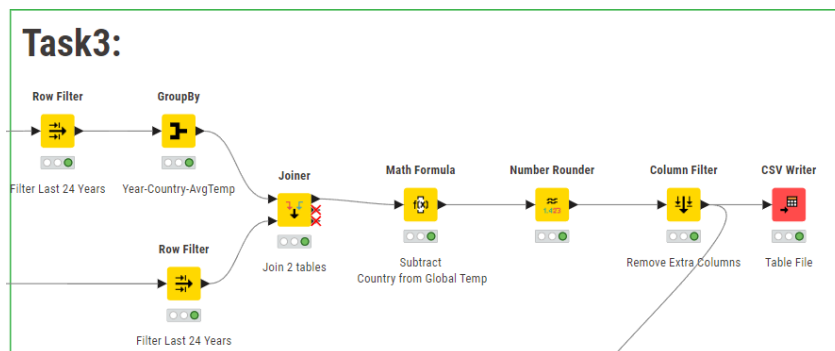
### Challenges & Decisions:

- Used **Row Filter** node to get the data of last 24 years from city\_data and global\_data tables

- There were some countries recorded their temp in the same year multiple times in City\_data table, so I used **GroupBy** node grouping by **year & country** columns aggregating **avg\_temp**
- Used **Joiner** node to get one table having year, country, country and global temperatures
- City\_table ends at 2013 year and global\_table ends at 2015 so I decided to eliminate the extra 2 years in global\_table while using **Joiner** configuration
- Used **Math Formula** node to get the difference between the two temperature columns

**Nodes Used:** CSV Reader, GroupBy, Row Filter, Joiner, Math Formula, Number Rounder, Column Filter, CSV Writer

### Workflow:



**Output File:** Task3 Table.csv

### Sample Output:

Rows: 1781 | Columns: 3

#	RowID	year <small>Number (integer)</small>	country <small>String</small>	Difference <small>Number (double)</small>
1	Row0	2001	Afghanistan	-6.4
2	Row1	2001	Albania	-7.2
3	Row2	2001	Algeria	-8.7
4	Row3	2001	Angola	-15
5	Row4	2001	Argentina	-8.7
6	Row5	2001	Armenia	-0.8
7	Row6	2001	Australia	-7.2

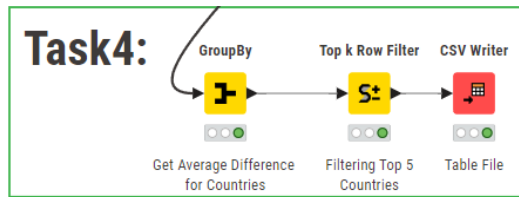
**Task4:** Output a table that shows the top 5 countries that have the largest difference from the global Temp

### Challenges & Decisions:

- Used **GroupBy** node to get the average temperature difference for each country in the last 24 years
- Used **Top K Filter** node to get the top 5 countries with largest differences

**Nodes Used:** Same as task3, GroupBy, Top K Row Filter, CSV Writer

### Workflow:



**Output File:** Task4 Table.csv

**Sample Output:**

Rows: 5 | Columns: 2

Table Statistics

<input type="checkbox"/>	#	RowID	country String	Mean(Difference) Number (double)
<input type="checkbox"/>	1	Row81	Mongolia	11.331
<input type="checkbox"/>	2	Row92	Norway	6.008
<input type="checkbox"/>	3	Row10	Russia	5.3
<input type="checkbox"/>	4	Row63	Kazakhstan	4.077
<input type="checkbox"/>	5	Row40	Finland	3.877

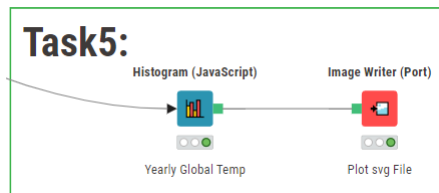
**Task5:** Draw a histogram for the yearly global temperatures

**Challenges & Decisions:**

- Used **Histogram (JavaScript)** node
- After several trials I used 10 bins for better resolution and readability of the temperature distribution over the last years

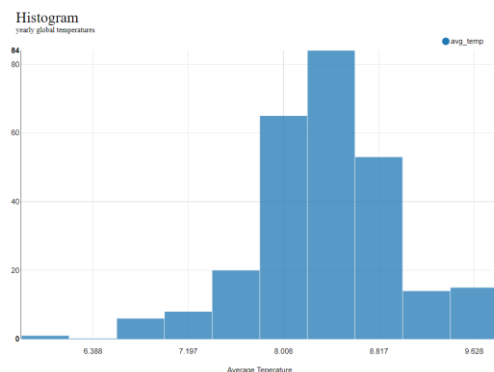
**Nodes Used:** CSV Reader, Histogram (JavaScript), Image Writer (Port)

**Workflow:**



**Output File:** Task5 Image.svg

**Sample Output:**



---

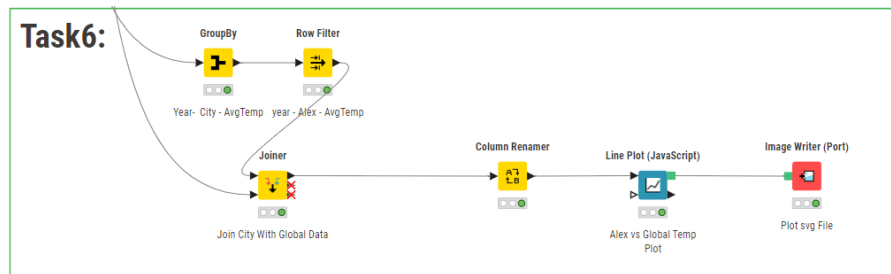
**Task6:** Draw a chart to compare between any city and global average temperature over the past years

**Challenges & Decisions:**

- Used Alexandria city to compare with global temperatures using **GroupBy & Row Filter** nodes
- Used **Row Filter** node to eliminate years with empty records for Alexandria
- Used **Line Plot (JavaScript)** node for side-by-side trend comparison

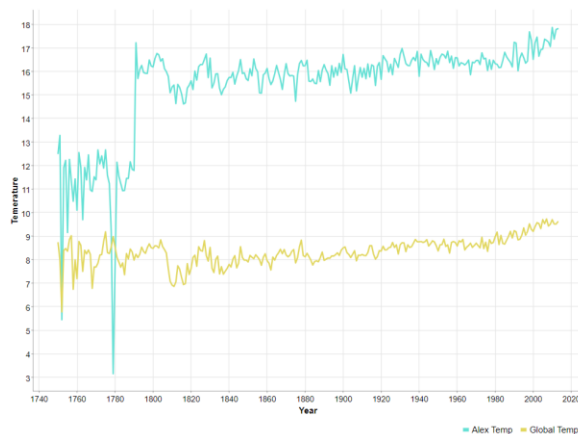
**Nodes Used:** CSV Reader, GroupBy, Row Filter, Joiner, Column Renamer, Line Plot (JavaScript), Image Writer (Port)

**Workflow:**



**Output File:** Task6 Image.svg

**Sample Output:**



---

## 5. Conclusion

- Efficiently used KNIME for full-cycle data analysis and visualization
- Applied advanced data manipulation techniques to extract key insights
- Created clear visualizations to highlight global and regional temperature trends
- Compared local vs. global data to identify meaningful patterns
- Demonstrated strong analytical thinking and fast adaptation to new tools