**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:**

1. Output a table that has the overall average of each country
2. Classify the countries Temperature into “Low/Mid/High”
3. 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
4. Output a table that shows the top 5 countries that have the largest difference from the global Temp
5. Draw a histogram for the yearly global temperatures
6. 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**

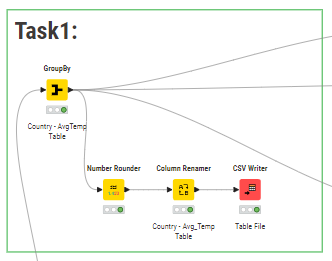
**Task1:** Output a table that has the overall average of each country

**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:**

**A screenshot of a computer

AI-generated content may be incorrect.**

**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

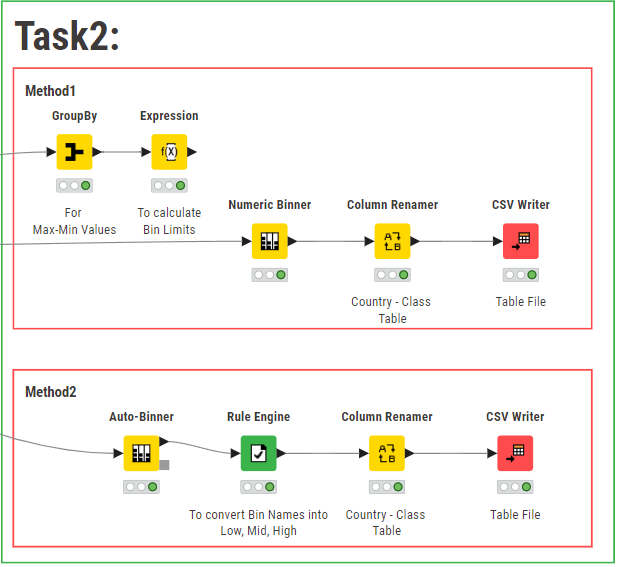
**Method 2:** Using Auto Binner

**Challenges & Decisions:**

* The **Auto Binner** doesthe 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)

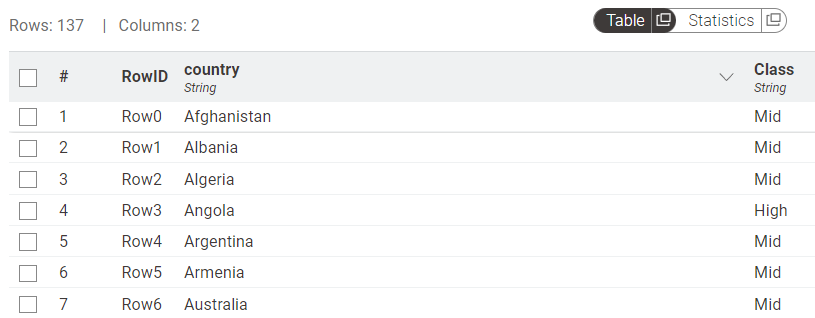
**Nodes Used:** CSV Reader, GroupBy, Auto Binner, Rule Engine, Column Renamer, CSV Writer

**Workflow:**

****

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

**Sample Output:**

****

**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

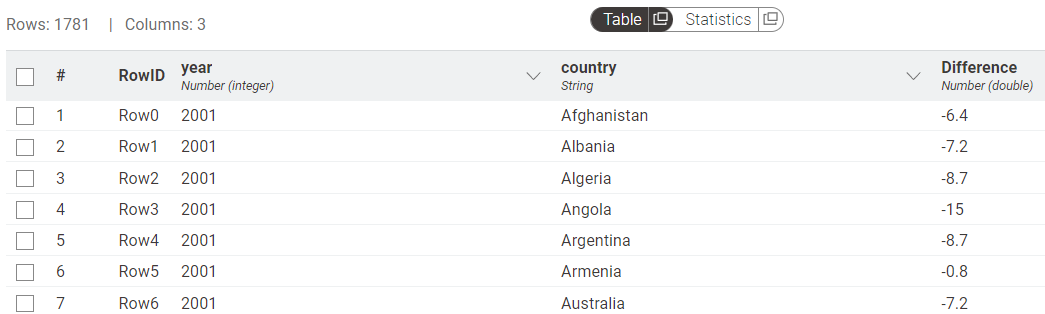
**Workflow:**

**A computer screen shot of a diagram

AI-generated content may be incorrect.**

**Output File:** Task3 Table.csv

**Sample Output:**

****

**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:**

**A diagram of a group

AI-generated content may be incorrect.**

**Output File:** Task4 Table.csv

**Sample Output:**

**A screenshot of a computer

AI-generated content may be incorrect.**

**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:**

**A screen shot of a computer

AI-generated content may be incorrect.**

**Output File:** Task5 Image.svg

**Sample** **Output:**

**A graph with blue bars

AI-generated content may be incorrect.**

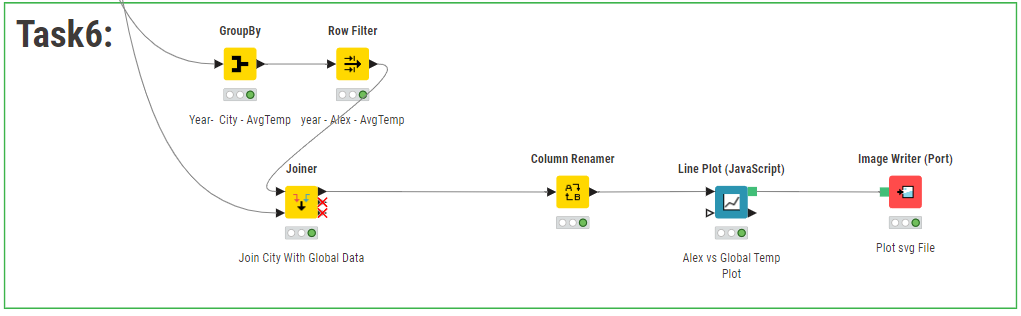
**Task6:** Draw a chart to compare between any city and global average temperature over the past years

**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

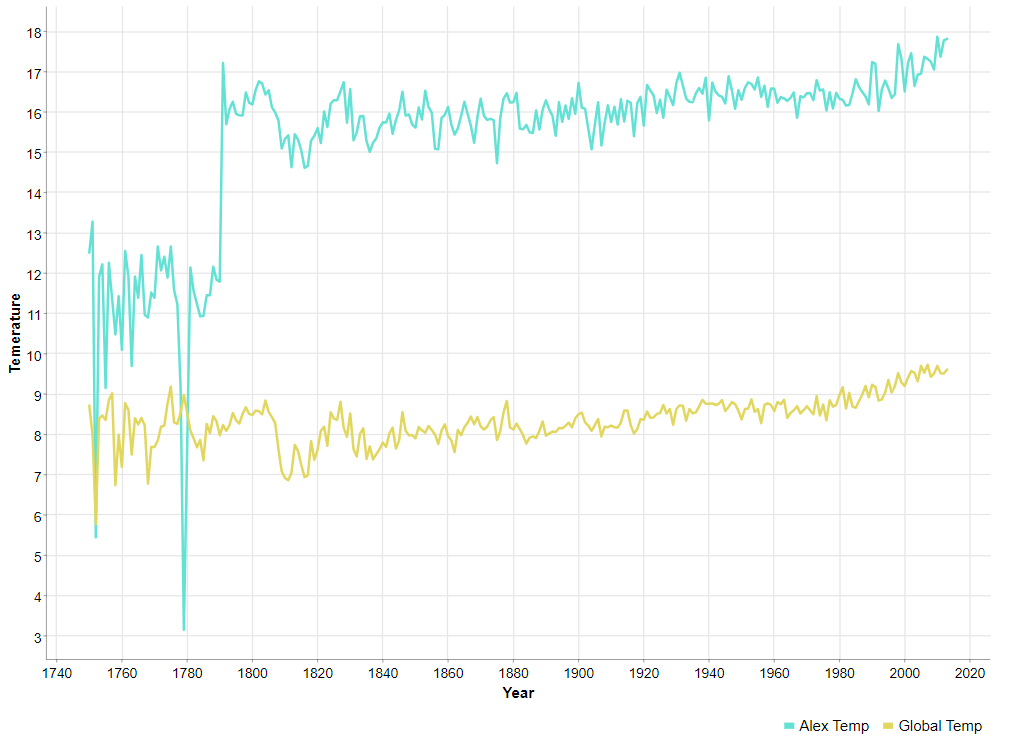
**Nodes Used:** CSV Reader, GroupBy, Row Filter, Joiner, Column Renamer, Line Plot (JavaScript), Image Writer (Port)

**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