KNIME Project - Netflix Data Analysis

Presented by Shih-Tien (Eric) Wang

What is KNIME ?

KNIME Analytics Platform is the open source software for creating data science.

KNIME Analytics Platform is the open-source software for creating data science.

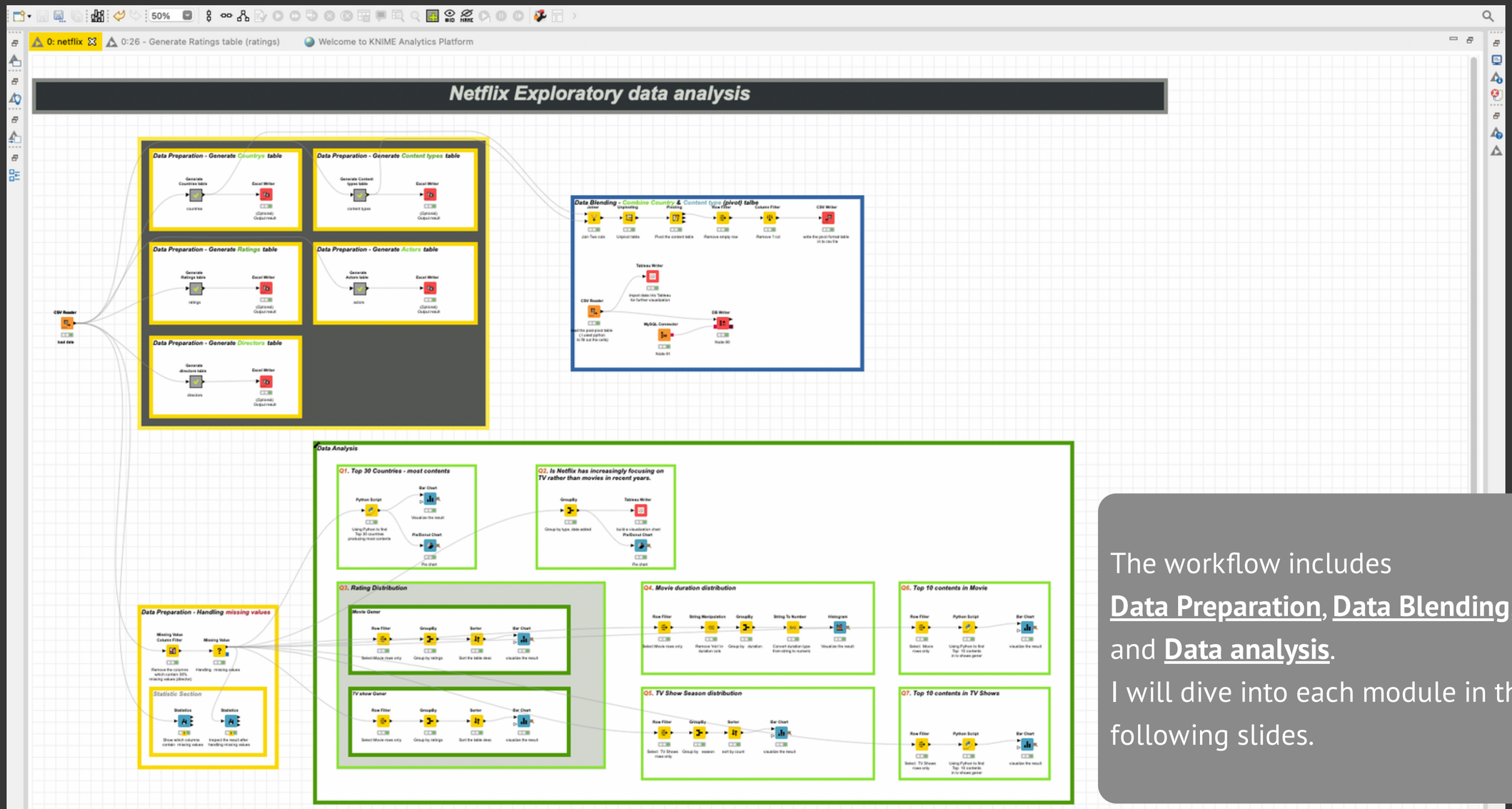
Personally, the most fantastic advantage is KNIME makes the Data Preprocessing process easier and more efficient. It's a powerful ETL platform.

Introduction of the project

The main object of the project is to analyze the data by KNIME.
I used KNIME to perform data preparation, data blending, and data analysis.

The dataset contains NETFLIX shows information (Title, Rating, Content-Type, etc...), and I used KNIME to find out some interesting insight.

Project Workflow

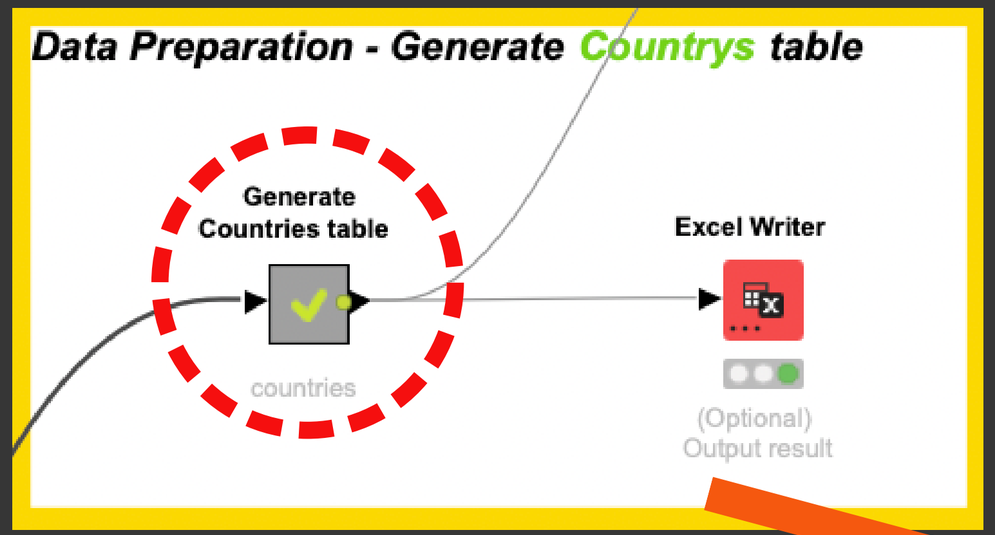


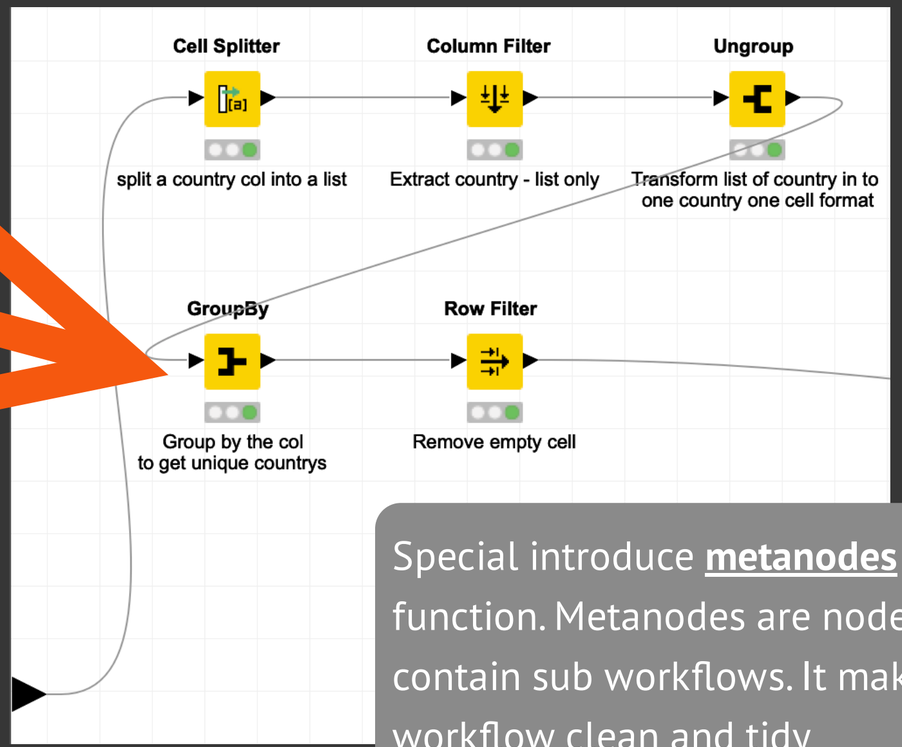
The workflow includes **Data Preparation**, **Data Blending**, and **Data analysis**.
I will dive into each module in the following slides.

Data Preparation - Generate metadata

In this phase, I load the dataset into KNIME and using **Row filter** and **Column filter** to generate metadata.

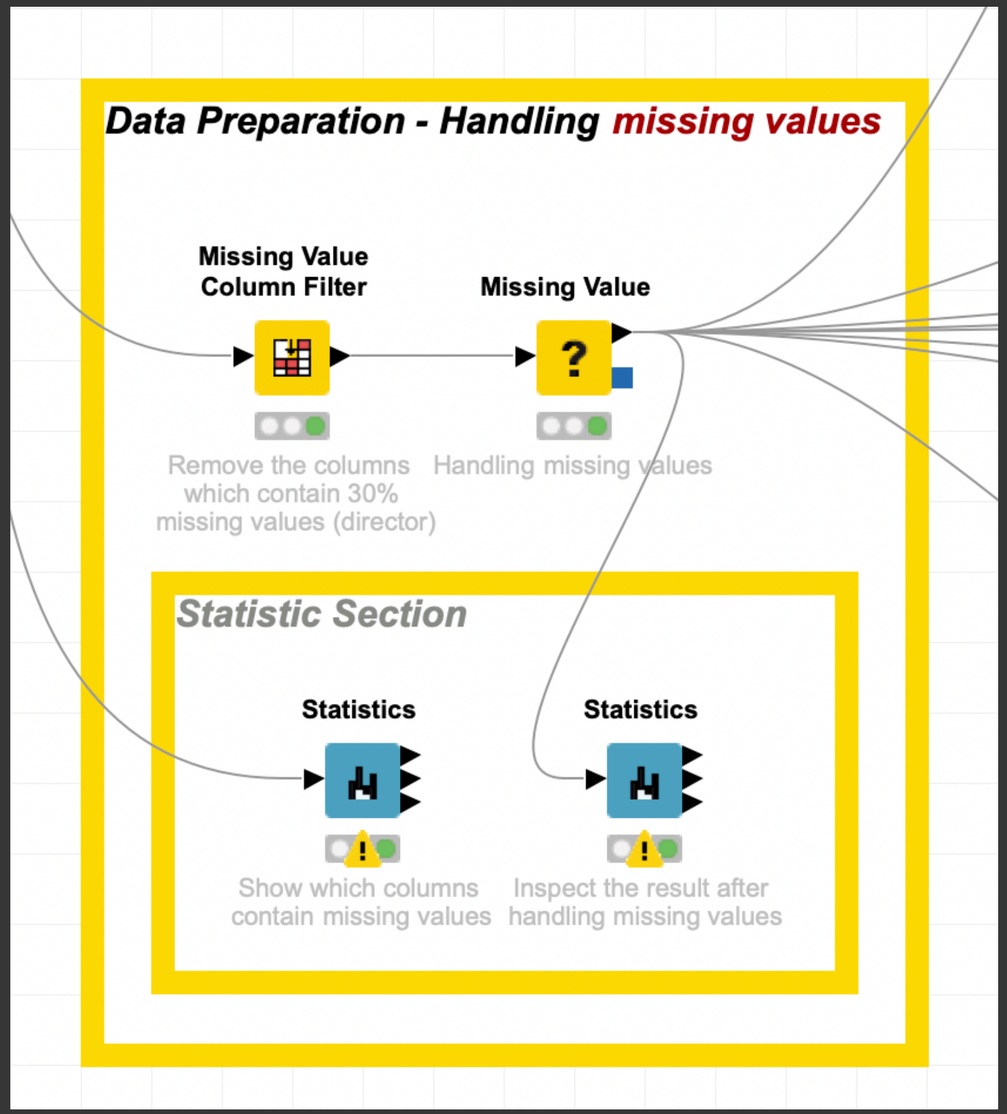
Data Preparation - Metanode





Special introduce **metanodes** function. Metanodes are nodes that contain sub workflows. It makes the workflow clean and tidy

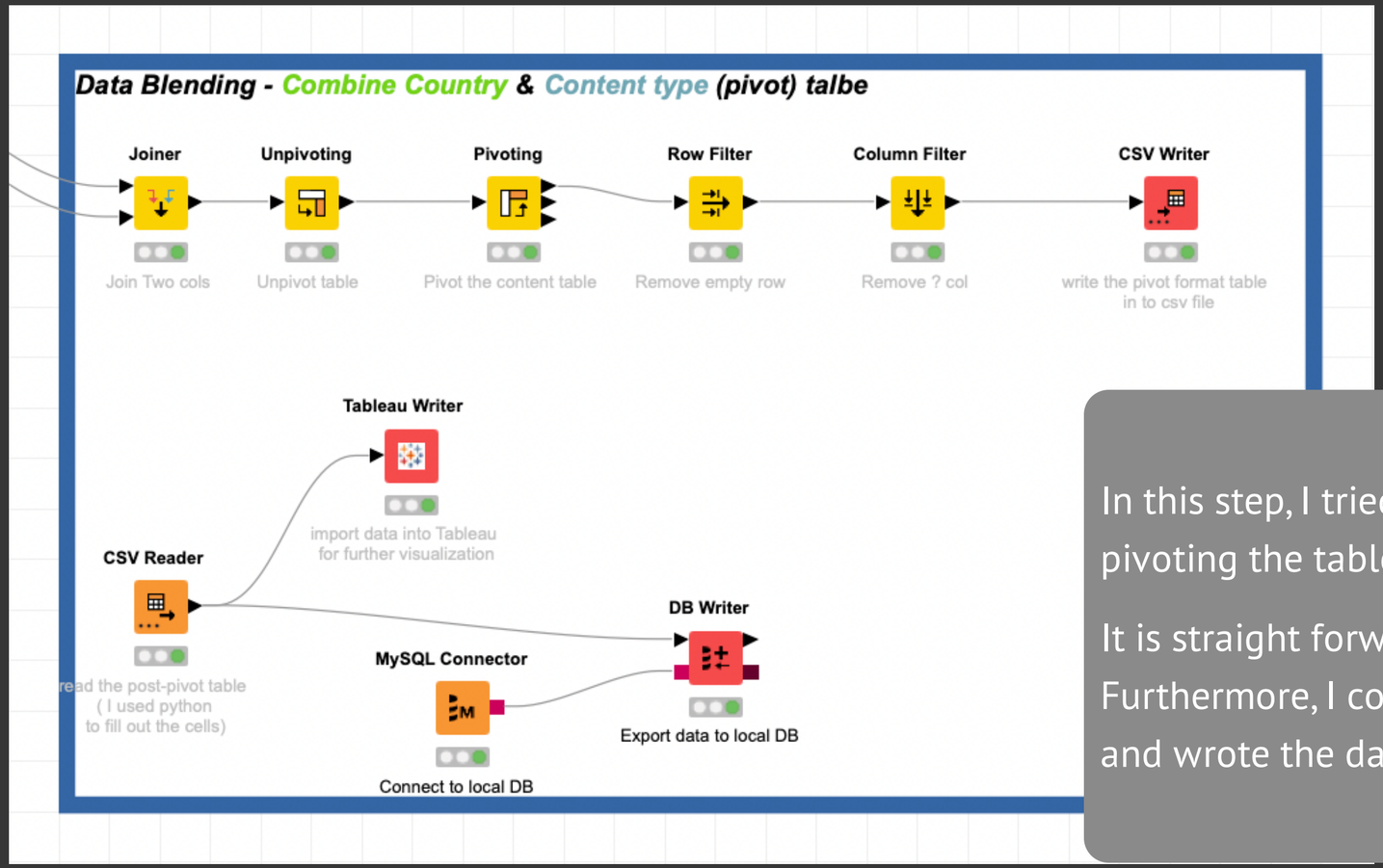
Data Preparation - Handling Missing Values



I used to struggle in handling missing values since the task is overwhelming and seems endless.

However, KNIME makes this process intuitive and easy to track. Moreover, KNIME provides some statistics functions for understanding the dataset.

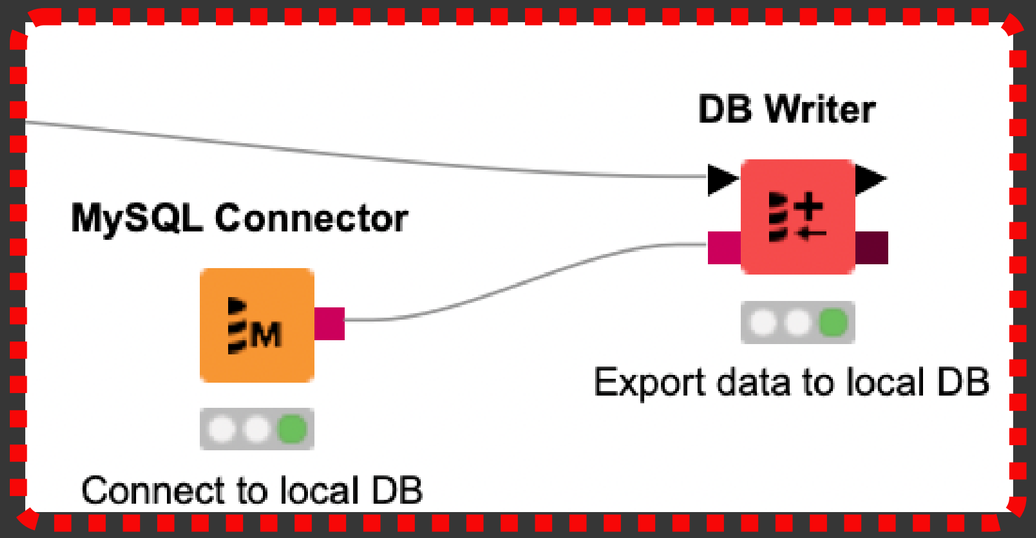
Data Blending



In this step, I tried to do advanced data manipulation - pivoting the table.

It is straight forward by the KNIME repository. Furthermore, I connected to the local database server and wrote the data from KNIME to the local DB.

Data Blending - DB connection



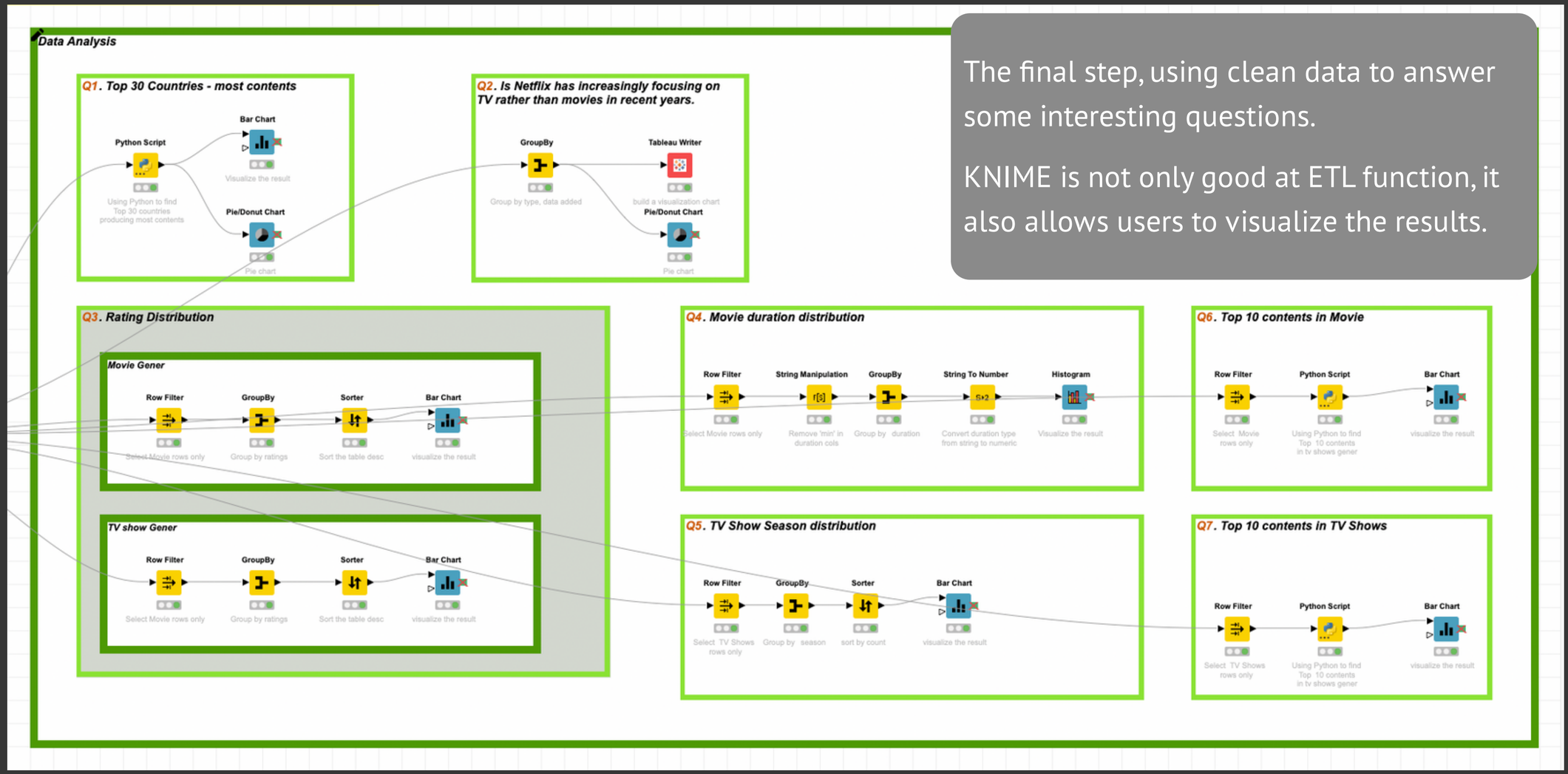

```
1 •  SELECT country, Action_Adventure FROM Knime.netflix_pivot
2    WHERE Action_Adventure > 50
3    ORDER BY Action_Adventure DESC
```

100% 34:1

Result Grid Filter Rows: Search Edit: Export/Import:

country	Action_Adventure
United States	333
India	130
United Kingdom	65
Hong Kong	64
China	53

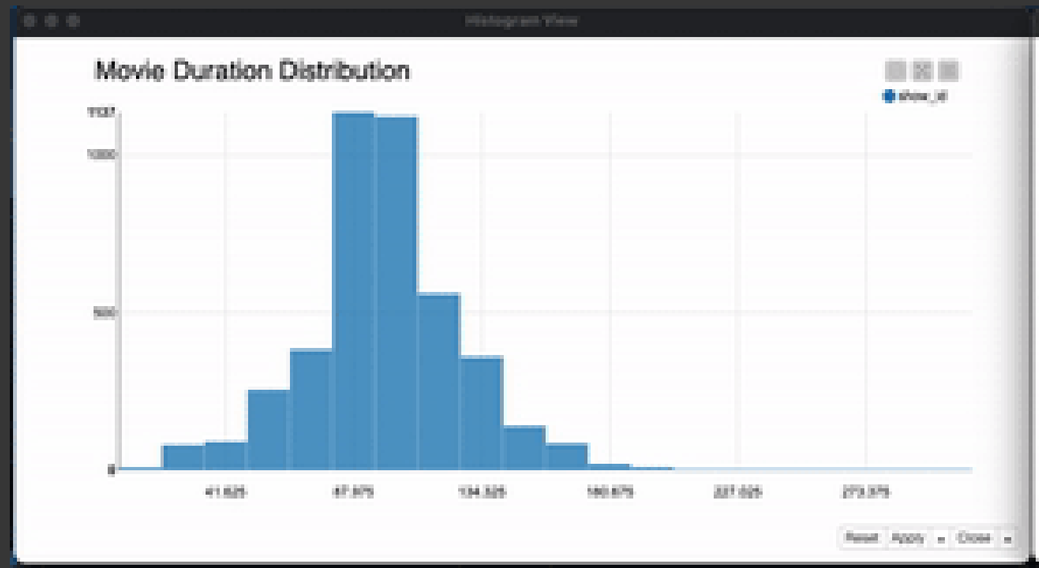
Data Analysis

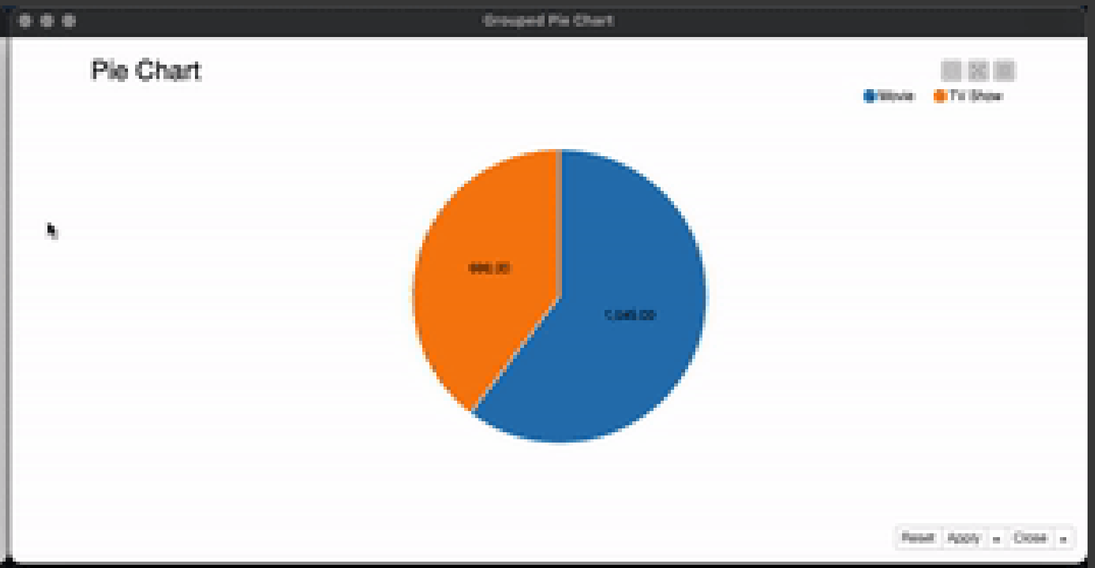


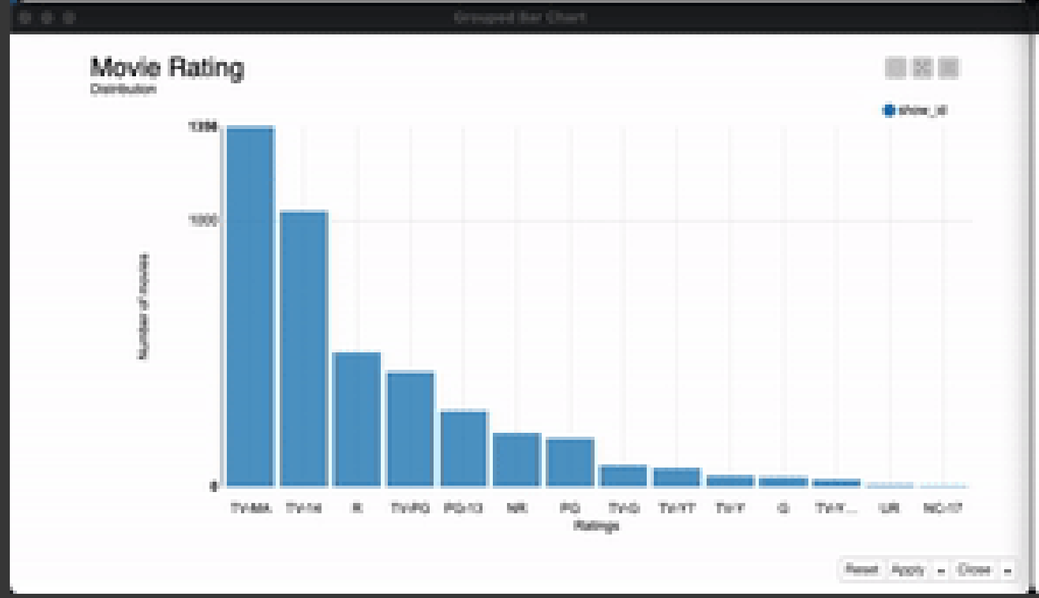
The final step, using clean data to answer some interesting questions.

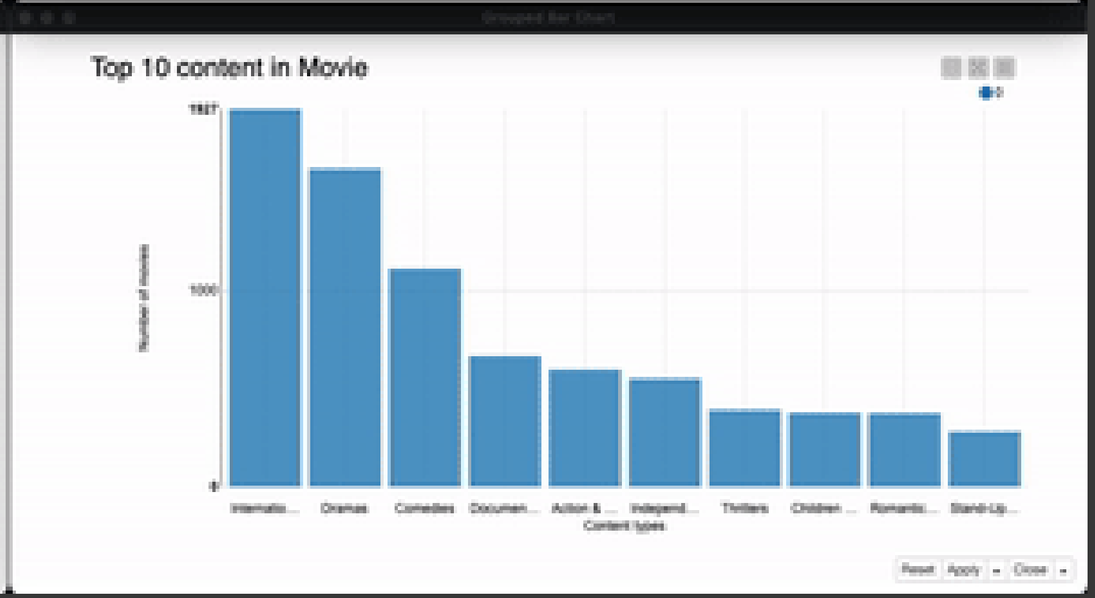
KNIME is not only good at ETL function, it also allows users to visualize the results.

Data Analysis - Vusualization









Conclusion and Reference

KNIME is excellent for data analysis.

The canvas allows me to see the whole workflow effortlessly. Therefore, I won't lose my object during data wrangling.

It's a great application and I will keep exploring its features.

Reference

* Knime website: https://www.knime.com/knime-analytics-platform
* Netflix dataset: https://www.kaggle.com/shivamb/netflix-shows