



# Detecting Frauds in Investment Contracts Enrich Data Analytics with GenAI

Elisabeth Richter, Data Scientist @KNIME

2024/07/22

# Agenda

The Use Case: Fraud Detection in Investment Contracts

IQR Technique for Outlier Detection

LLM Interaction: Retrieval-Augmented Generation

Generative AI in KNIME

Wrap-up: To learn more about KNIME

Abstract geometric lines in the top right corner, consisting of several overlapping, thin white lines forming a complex, angular shape against the yellow background.

**Use Case:  
Fraud Detection in Investment Contracts**

# What is Fraud?

- *"Fraud is an intentionally deceptive action designed to provide the perpetrator with an unlawful gain or to deny a right to a victim."*

–Investopedia

- Fraud impacts people, industries, entities, services and the environment
- Common types of financial fraud include:
  - Tax fraud
  - Credit card fraud
  - Securities fraud
  - Insurance frauds
  - Bankruptcy fraud
- Luckily, severe financial frauds are rare, but the impact of tiny, frequent frauds can be significant



# Detecting wrong (fake?) contracts

Find suspicious contracts



**Investment Agreement** : C000004

**KNOW ALL MEN BY THESE PRESENTS:**

This Contract is entered by and between:

Bank A GmbH, whose principal place of address is at Debusstr. 18, Konstanz, Germany, 78460, (hereinafter referred to as "DEBTOR");

-and-

Mira Gleich, of legal age, (hereinafter referred to as "CREDITOR");

**WITNESSETH: That -**

WHERE AS, the DEBTOR provides the CREDITOR with the product "Private Investment"; NOW, THEREFORE, for and in consideration of the foregoing premises, the parties hereto agree as follows:

1. Product Description: The CREDITOR purchases the investment product "Private Investment". The DEBTOR makes investments with the amount paid by the CREDITOR.
2. Payment: The CREDITOR has to transfer a one time transfer in amount of 799.06 euros.
3. Payment Method: The DEBTOR may choose his method of payment to his convenience.
4. Agreement Modification: No modification of this Agreement shall be considered valid unless made in writing and agreed upon by both Parties.
5. Assignment of Rights: The CREDITOR can not transfer or assign this Agreement to a third party.
6. Severability: Should any provision found in this Agreement be held invalid, illegal, or unenforceable by any competent court, the same shall apply only to the provision and the rest of the remaining provisions hereto shall remain valid and enforceable.
7. Preferred means of communication: The CREDITOR's preferred means of communication is via email mira.gleich@provider.com

IN WITNESS WHEREOF, the Parties has executed this Agreement. 2016-04-05

**Name & Signature of Creditor**      **Name & Signature of Debtor**

\_\_\_\_\_

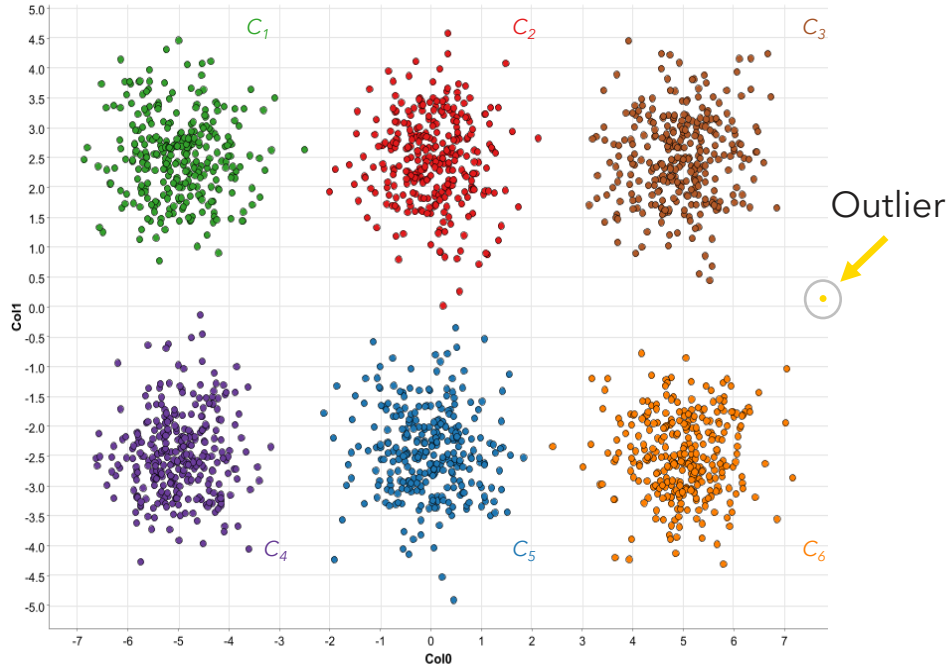
# Techniques for Outlier Detection

- Visualizations
  - **Statistics**
  - Clustering
  - Machine Learning
  - Deep Learning
- } Numeric data
- 
- String similarity metrics
- } Nominal data

# What is an Outlier?

- In statistics, an outlier is a data point that differs significantly from other observations

**Outlier detection techniques** are often used for fraud detection



# Fraud Detection Process

Banks monitor the risk of frauds regularly and notify their clients promptly if a suspicious transaction is detected



Bank A



Find Frauds



Send Warning  
Email in  
Customer's  
Language

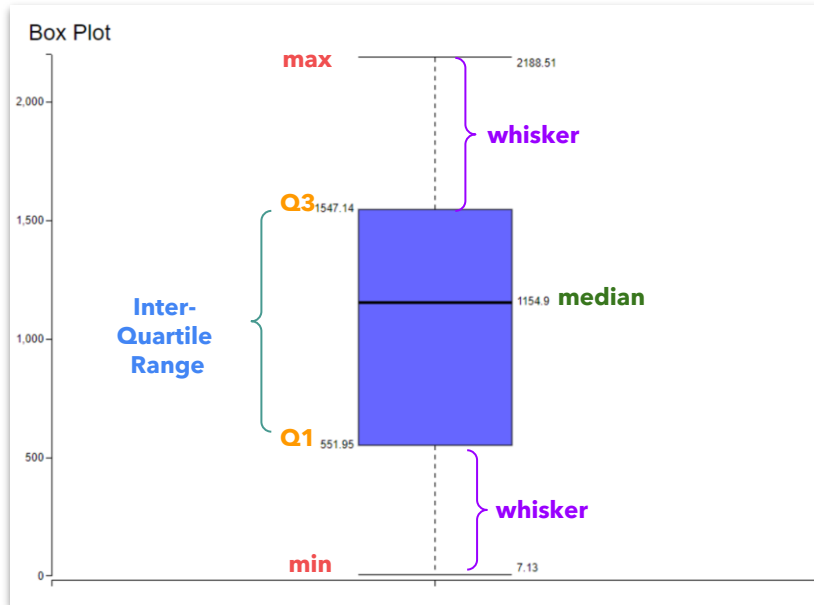
We will use  
**GenAI** to  
generate this  
email

Abstract geometric lines in the top right corner, consisting of several overlapping, thin white lines forming a complex, angular pattern against the yellow background.

# **IQR Technique for Outlier Detection**

# Finding Outliers through basic stats

## ■ IQR Technique



- $x$  is an outlier if:

$$\begin{cases} x > Q3 + k(IQR) \\ \text{or} \\ x < Q1 - k(IQR) \end{cases}$$

where  $k \geq 0$  and  $IQR = Q3 - Q1$

$k \uparrow$ : less outliers

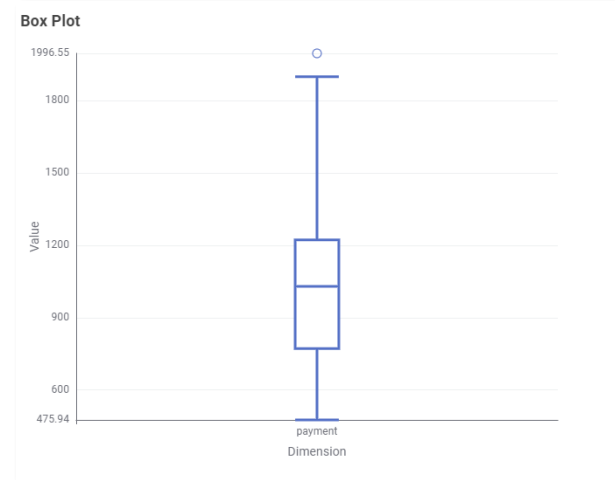
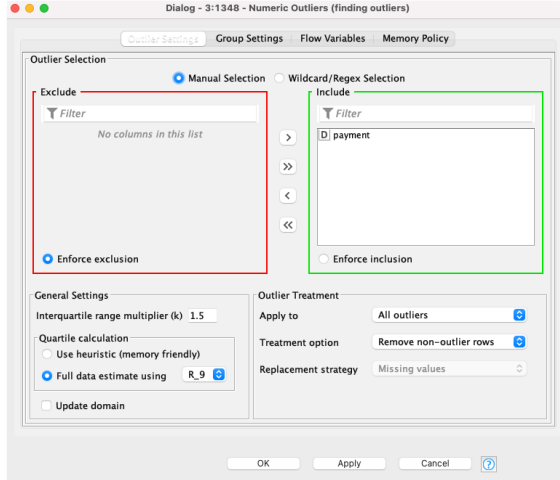
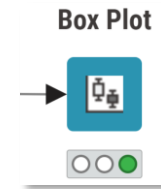
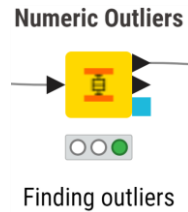
$k \downarrow$ : more outliers

Often,  $k = 1.5$

# Finding Outliers: Statistics

## IQR Technique

The Numeric Outliers node implements (and the Box Plot node visualizes) the IQR technique.

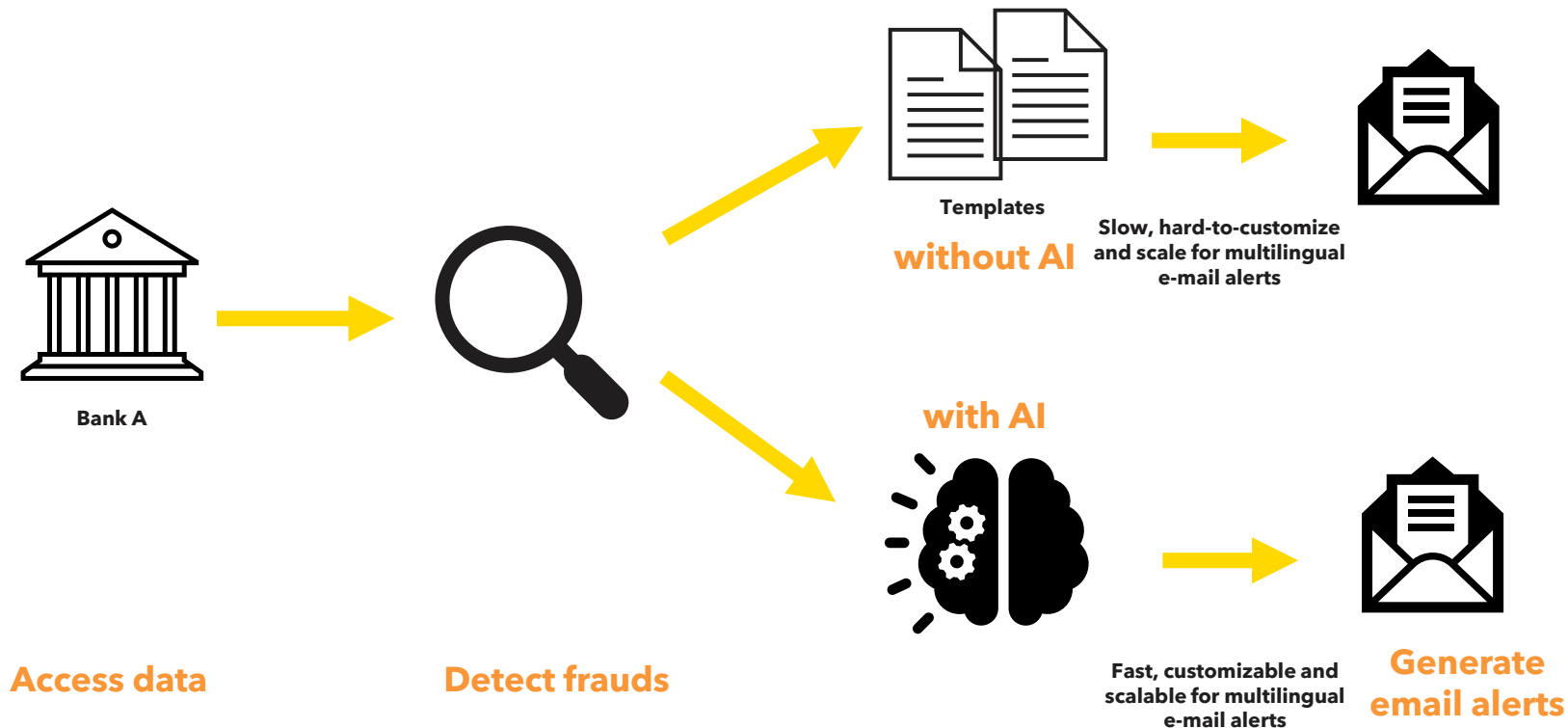


Abstract geometric lines in the top right corner, consisting of several overlapping, thin white lines forming a complex, angular pattern against the yellow background.

# LLM Interaction: Retrieval-Augmented Generation

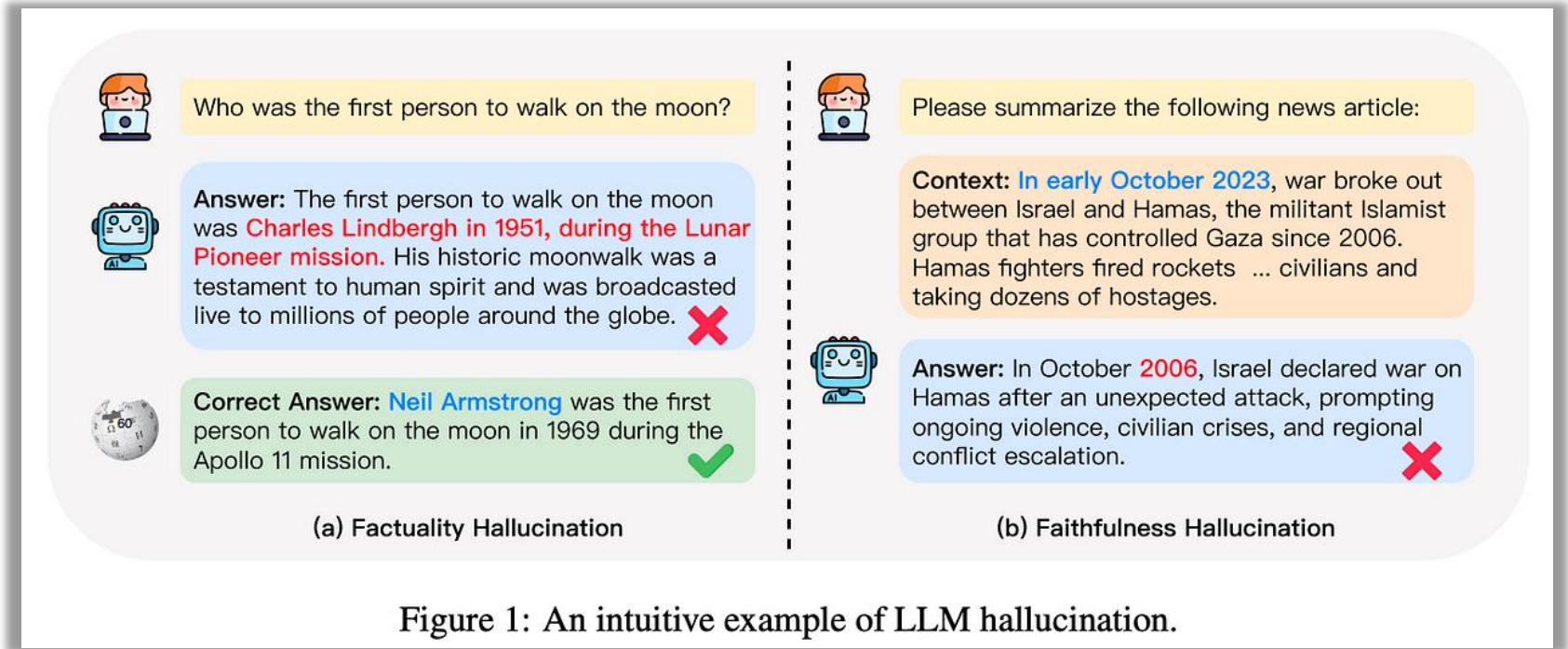
# Use GenAI for Email Alerts (no context)

Banks monitor the risk of frauds regularly and notify their clients promptly if a suspicious transaction is detected



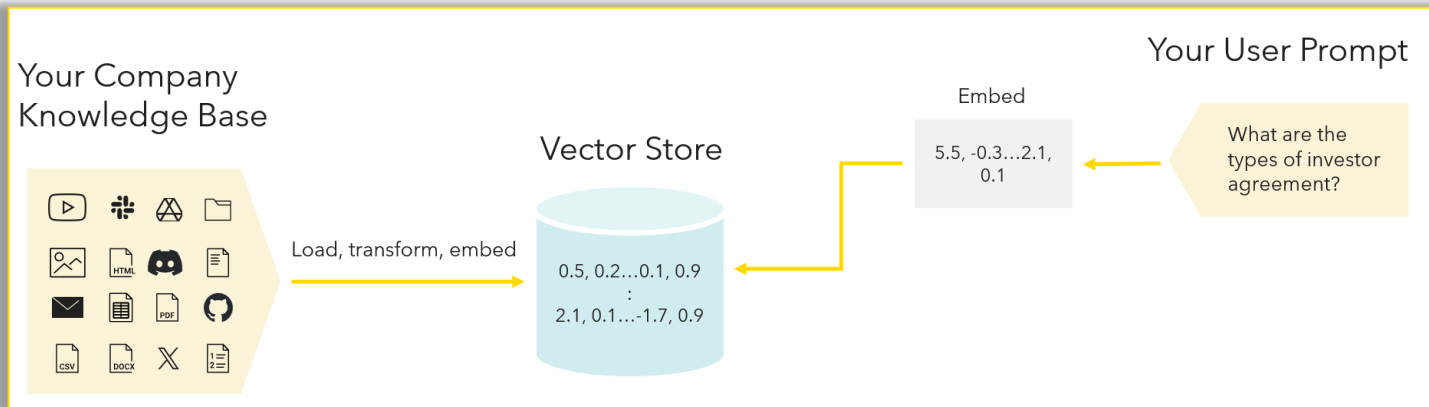
# LLM's Hallucinations & RAG

- LLMs are prone to hallucinate



# LLM's Hallucinations & RAG

- LLMs are prone to hallucinate
- Possible mitigation with **Retrieval-Augmented Generation (RAG)**
- Retrieval-Augmented Generation (RAG) is the process of optimizing the output of an LLM by referencing an authoritative knowledge base outside of its training data sources before generating a response.

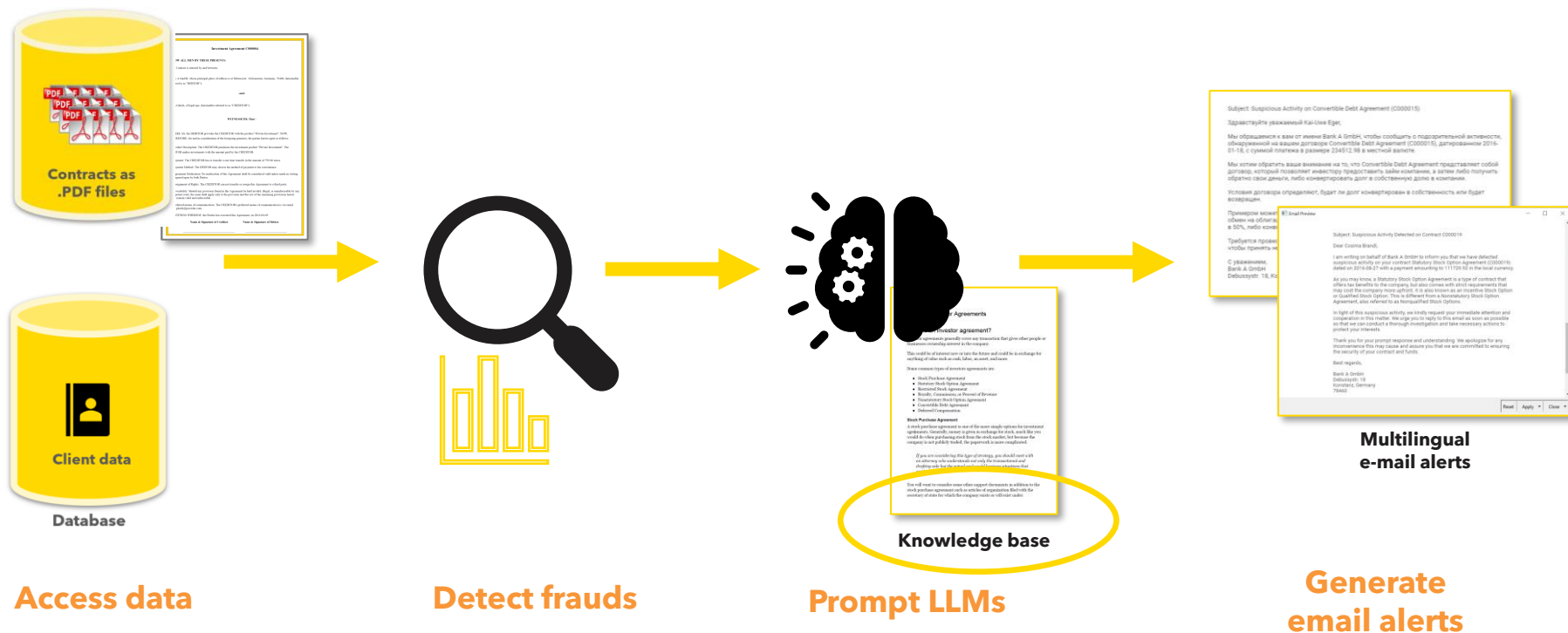


# LLM's Hallucinations & RAG

- LLMs are prone to hallucinate
- Possible mitigation with **Retrieval-Augmented Generation (RAG)**
- Retrieval-Augmented Generation (RAG) is the process of optimizing the output of an LLM by referencing an authoritative knowledge base outside of its training data sources before generating a response.
- In our case we could use the description of the various contract types.

# Use GenAI for Email Alerts with Context (RAG)

Add context from knowledge base with RAG (Retrieval Augmented Generation)



# What is an Embeddings Model?

The Oscar for best actress goes to ...



Embeddings Model



5	7	-2	3
---	---	----	---

The Oscar for best actress goes to ...



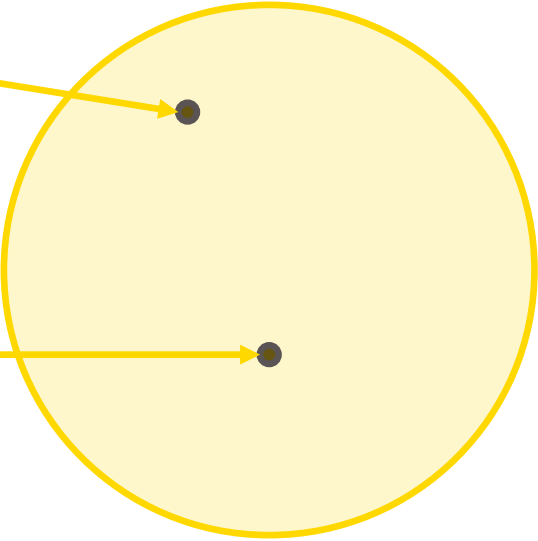
Embeddings Model



Argentina wins the Soccer world cup

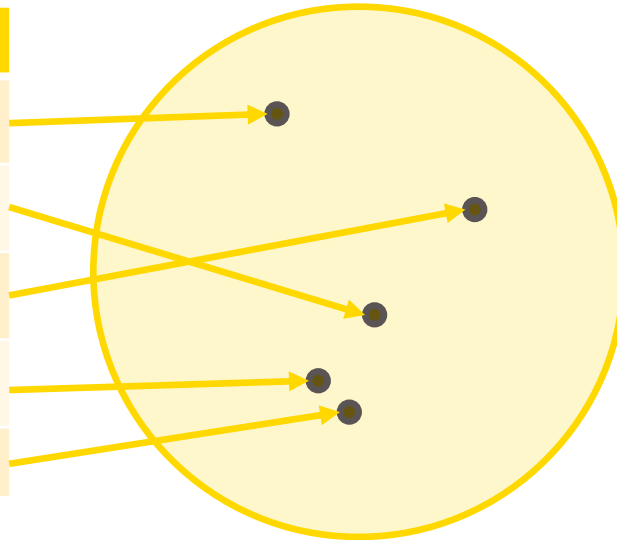


Embeddings Model



# What is a Vector Store?

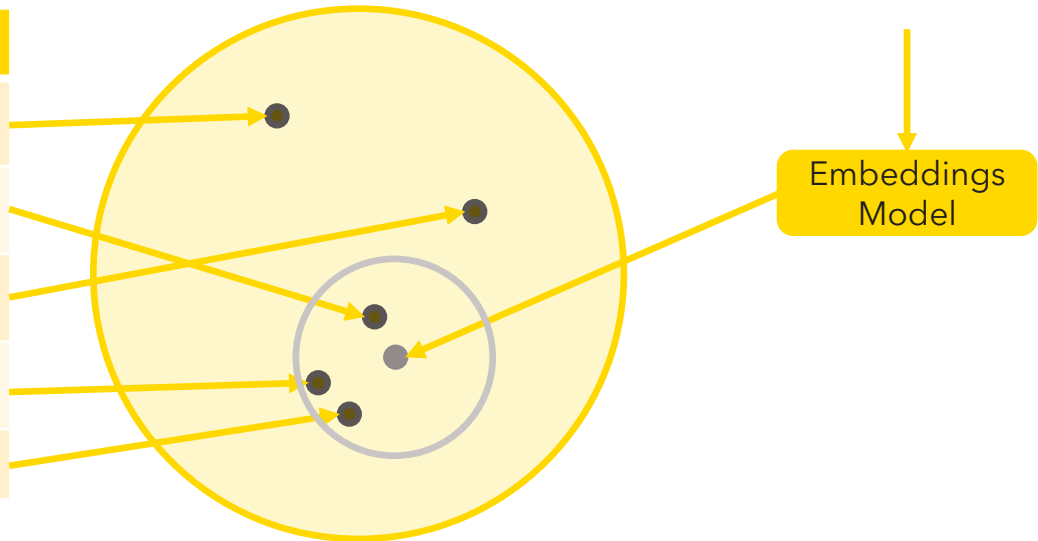
Document	Embedding
The 2023 Oscar for best actress goes to ...	5, 7, -2, 3
Argentina wins the soccer world cup	4, 6, -2, 4
The Kansas City Chiefs win the superbowl in 2023	4, 6, 5, -1
Bayern Munich wins the Bundesliga in 2023, again...	0, 4, 8, 9
VfB Stuttgart stays first class	0, 1, 9, 8



# How can we use it for Semantic Search?

Who won the soccer world cup?

Document	Embedding
The 2023 Oscar for best actress goes to ...	5, 7, -2, 3
Argentina wins the soccer world cup	4, 6, -2, 4
The Kansas City Chiefs win the superbowl in 2023	4, 6, 5, -1
Bayern Munich wins the Bundesliga in 2023, again...	0, 4, 8, 9
VfB Stuttgart stays first class	0, 1, 9, 8



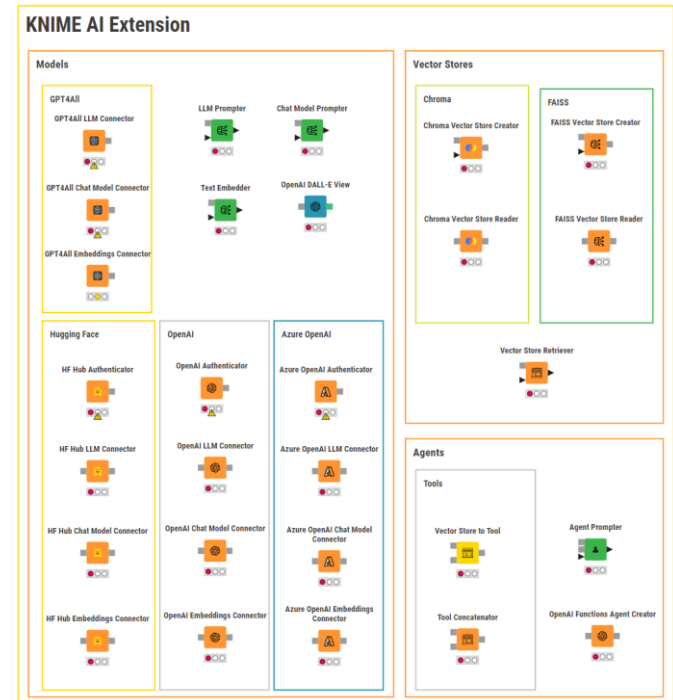
# Generative AI in KNIME



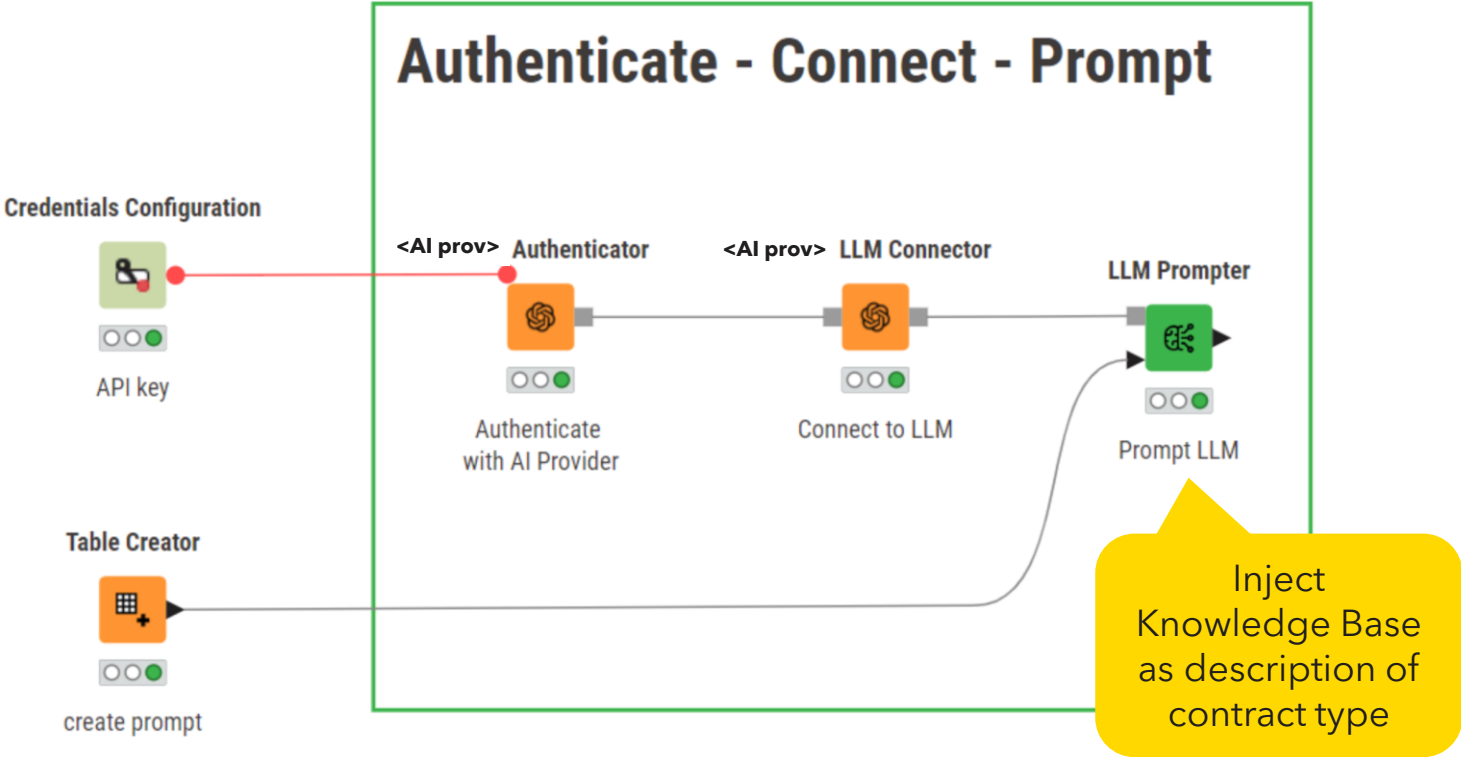
# KNIME AI Extension

LLMs can be leveraged with the [KNIME AI Extension](#) for connecting to open-source and closed-source models via API, or to open-source local models

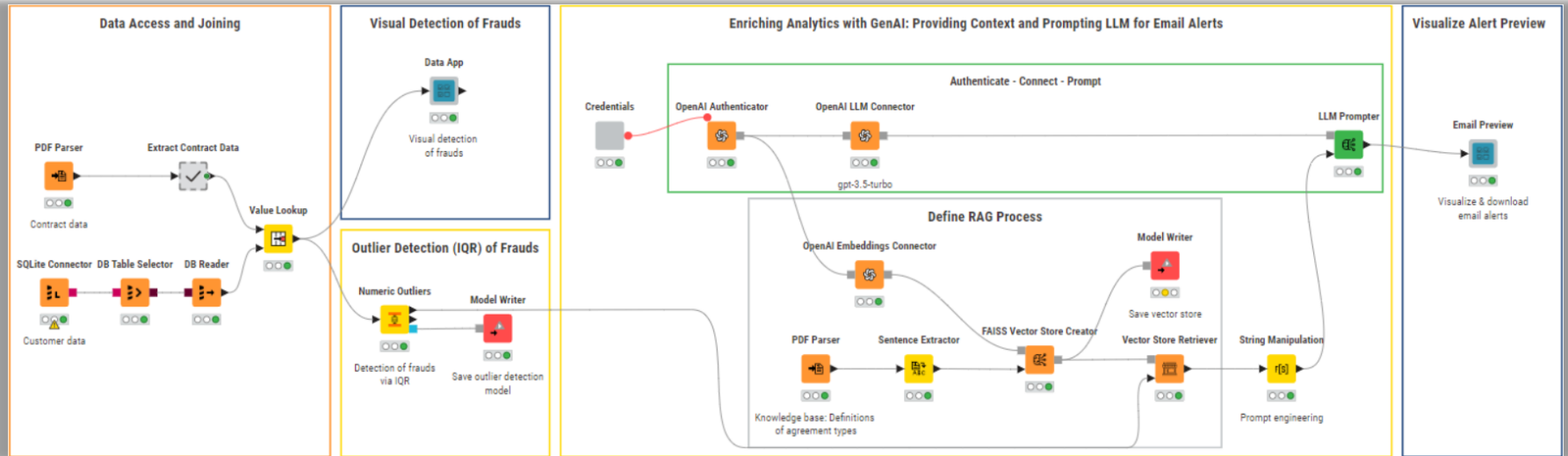
- Authenticators
- LLM Connectors
- Chat Model Connectors
- Embeddings Connectors
- LLM and Embeddings Prompters
- Local Connectors and Embedders
- Vector Stores
- Agents



# Simple Prompting: Authenticate - Connect - Prompt



# Let's have a look in KNIME Analytics Platform...



Download the workflow from the KNIME for Finance space:

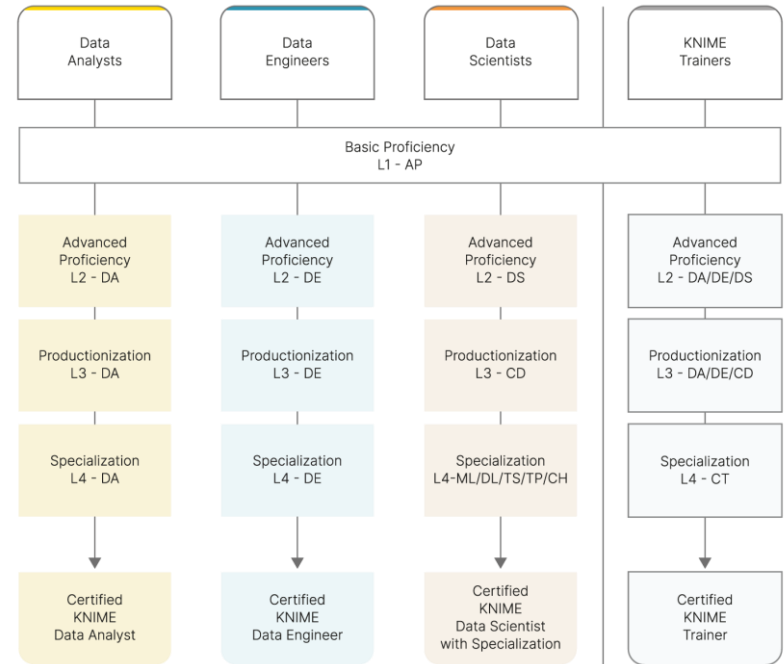
[Fraud Detection - Visualizations and IQR](#)

Abstract geometric lines in the top right corner, consisting of several overlapping, thin white lines forming various triangular and quadrilateral shapes against a yellow background.

**Wrap up:**  
**To learn more about KNIME**

# KNIME Learning Paths

- Courses from level **L1** to level **L4**
- Various professional profiles
- **Self-paced courses**
  - Videos and exercises at your own pace and for free
- **Instructor led courses**
  - Scheduled sessions and guided exercises in paid courses

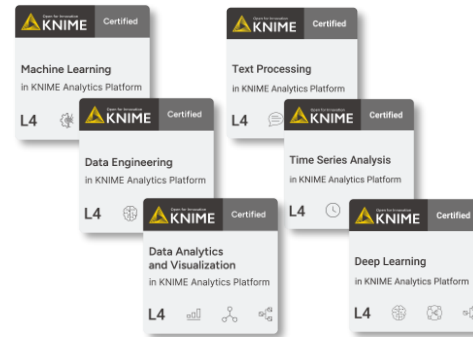
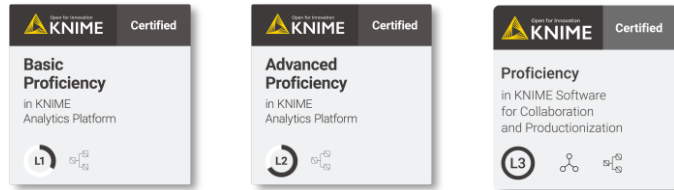


Find the **next course** at:

[knime.com/learning](https://knime.com/learning)

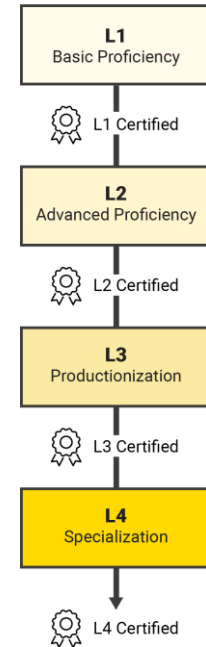
# KNIME Certification Exams

- From level **L1** to level **L4**
- Online examination
- Measure your expertise with KNIME software and data skills
- Get and share your badge!



## Teaching & Certifying

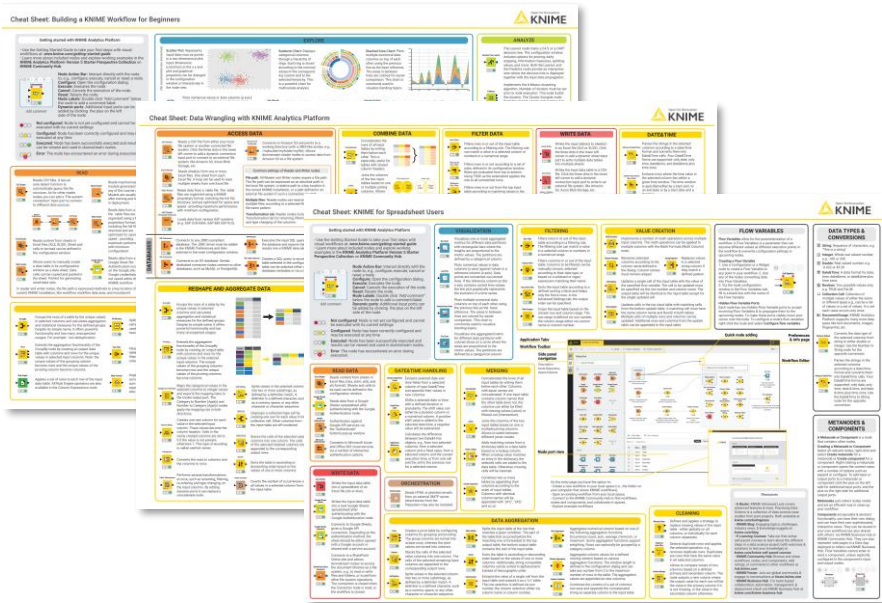
KNIME Expertise from level **L1** to level **L4**



Get **certified** at:

[knime.com/certification-program](https://knime.com/certification-program)

# KNIME Press: Cheat Sheets

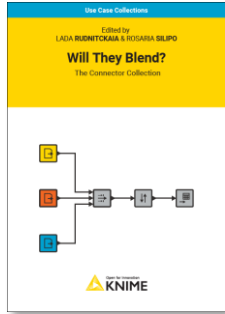


- Many cheat sheets available for:
  - Beginners
  - Spreadsheet Users
  - Machine Learning
  - Data Wrangling
  - Orchestration
  - and many more...

All **cheat sheets** are free and available for download at:

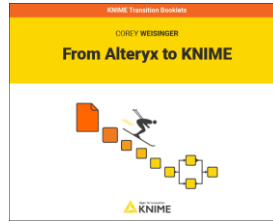
[knime.com/knimepress](https://www.knime.com/knimepress)

# KNIME Press: Books



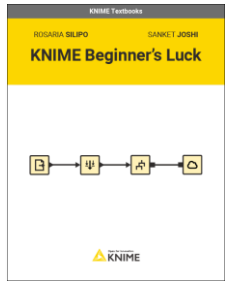
## Use Case Collections

Collections of classic and more innovative use cases around specific topics



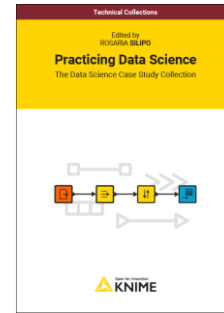
## Transition Booklets

Offer an easy onboarding into KNIME from other tools



## Textbooks

Well structured schoolbooks with plenty examples



## Technical Collections

Collections of specific technical topics to keep you up to date

All **KNIME Press books** are free and available for download at:

[knime.com/knimepress](https://knime.com/knimepress)

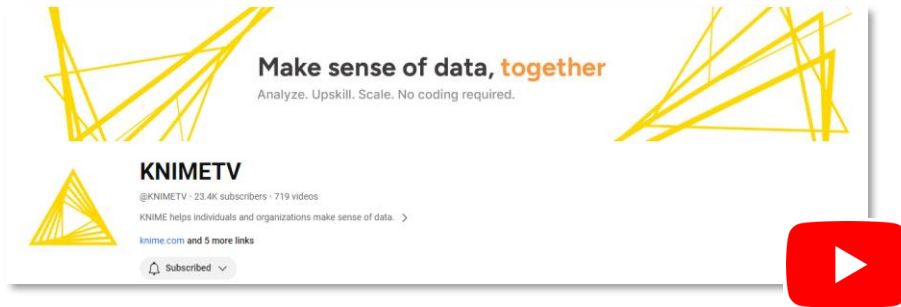
# KNIME TV and My Data Guest

- KNIME TV

- Webinars
- Panel discussions
- Educational videos
- Data Science topics
- Tutorials

- My Data Guest - Podcast

- Interviews with data professionals in the KNIME community

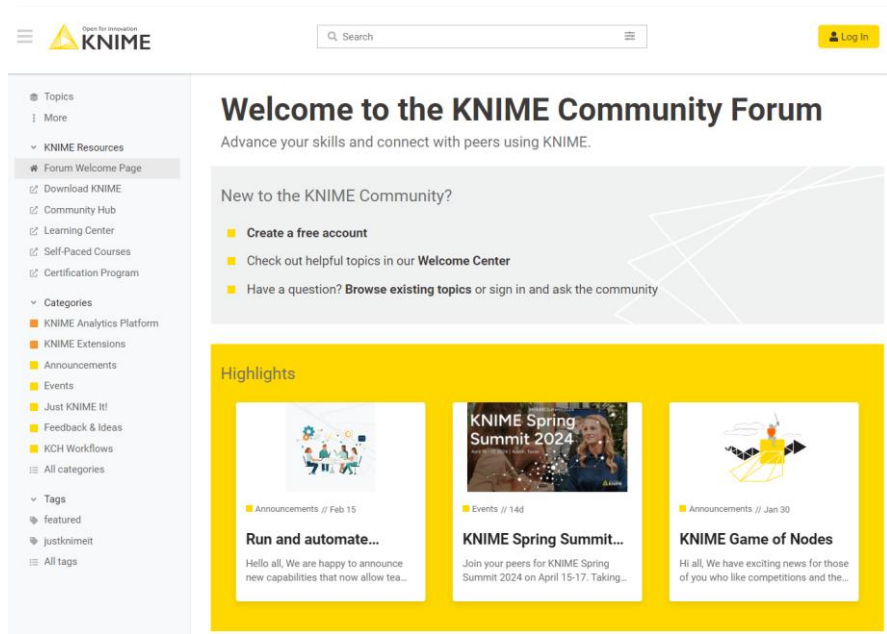


Find **KNIME multimedia content** at:

<https://www.youtube.com/@KNIMETV> and [Spotify](#)

# KNIME Forum

- Round-the-clock community support
- Specialized categories
  - KNIME Analytics Platform
  - KNIME Extensions
  - KNIME Business Hub
  - ...
- Multilingual support groups
  - French
  - Italian
  - Spanish
  - ...



Visit the KNIME Forum at:

<https://forum.knime.com/>

# Stay Connected with KNIME



## **Blog:**

[knime.com/blog](https://knime.com/blog)



## **KNIME Self-Paced Courses:**

[knime.com/knime-self-paced-courses](https://knime.com/knime-self-paced-courses)



## **Forum:**

[forum.knime.com](https://forum.knime.com)



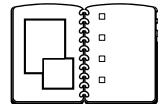
## **Email:**

[education@knime.com](mailto:education@knime.com)



## **KNIME Hub:**

[hub.knime.com](https://hub.knime.com)

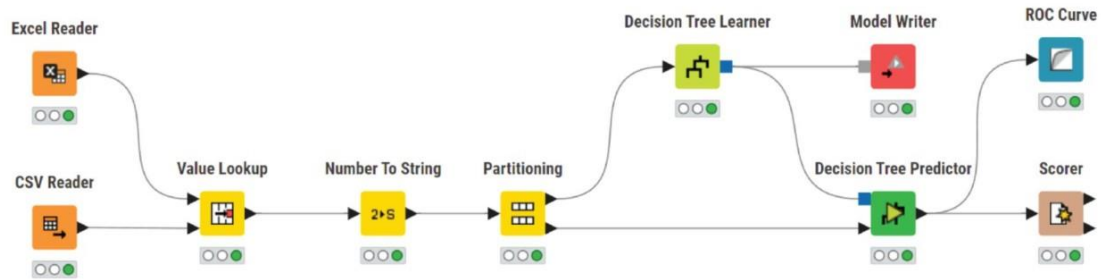


## **Medium Journal:**

[medium.com/low-code-for-advanced-data-science](https://medium.com/low-code-for-advanced-data-science)

Follow us on social media:





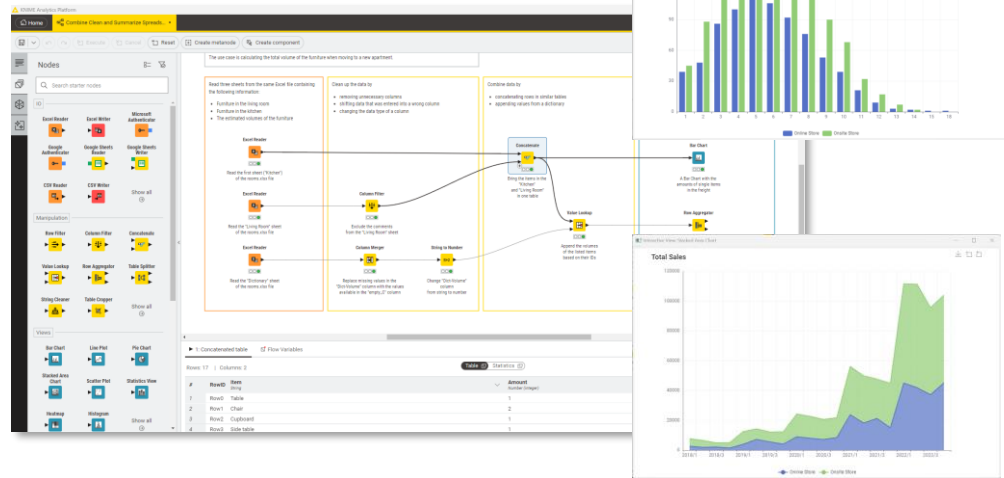
Thank you!

The top right corner of the slide features a series of overlapping, thin white lines that form a complex, abstract geometric pattern. These lines create a sense of depth and movement, resembling a stylized architectural or data visualization element.

# KNIME Analytics Platform

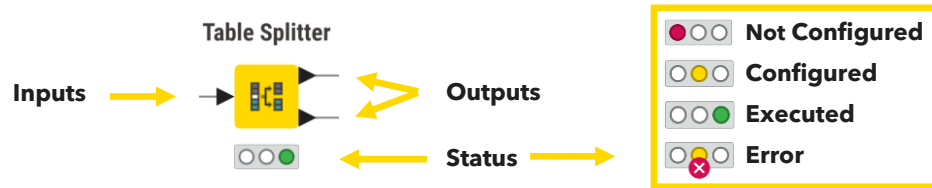
# What is KNIME Analytics Platform

- A tool to make sense of your data
  - Data analysis, Data Science, Data Engineering, AI
  - Transformation, visualization, reporting
- Open source & free
- Visual programming paradigm
  - No coding required

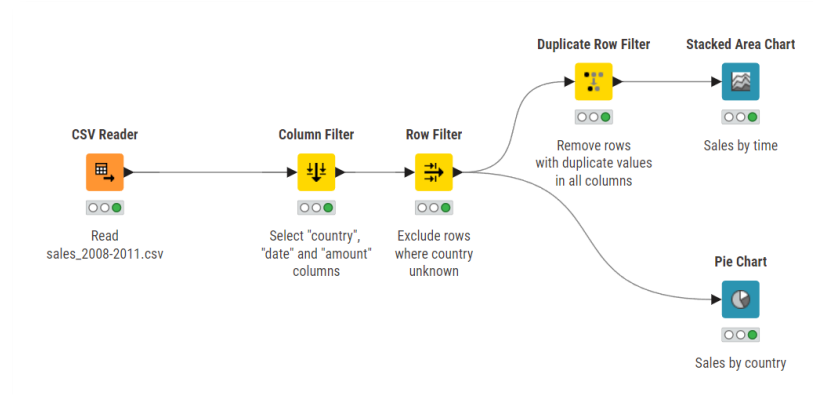


# Nodes and Workflows

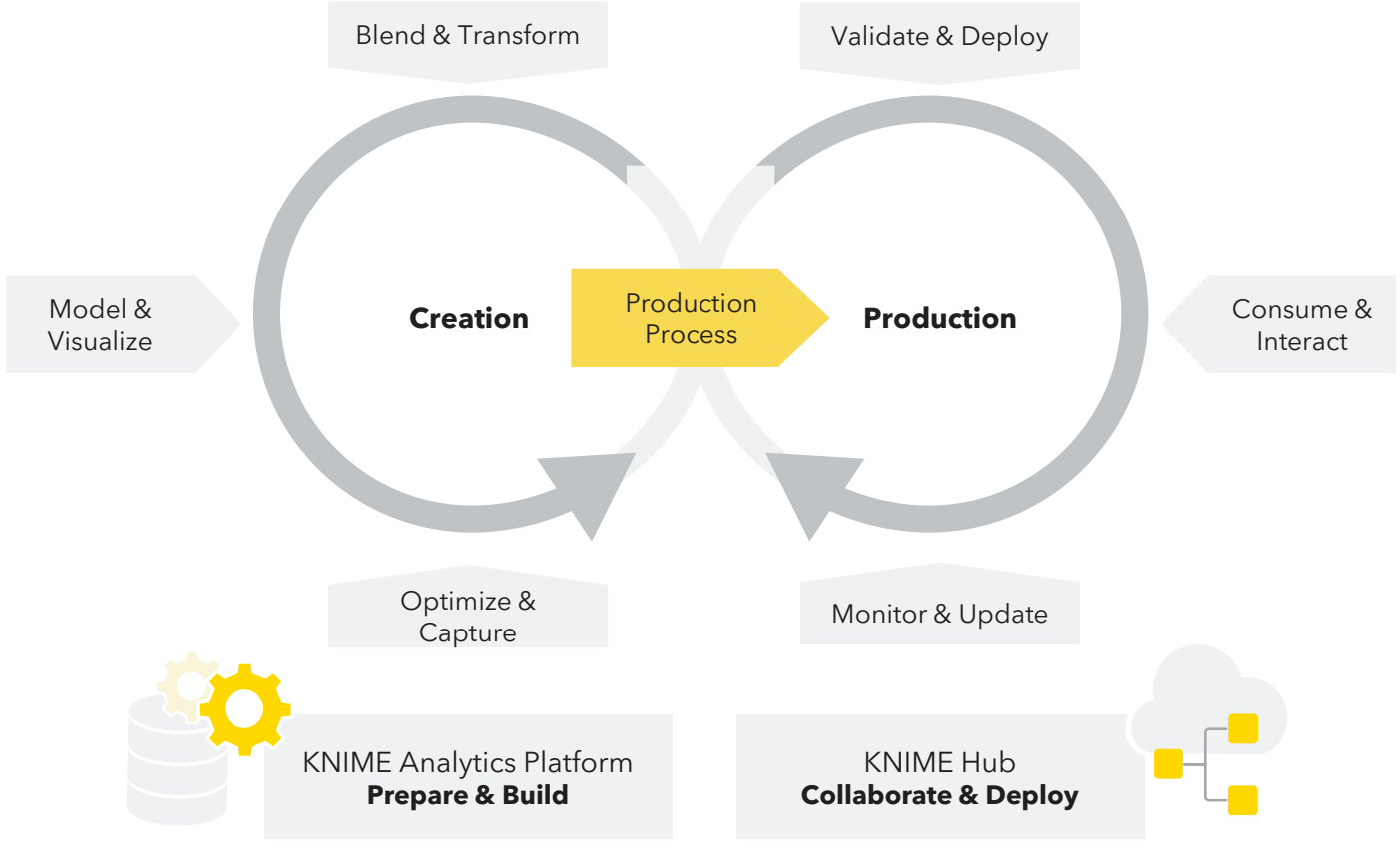
- **NODES** perform tasks on data



- Nodes are combined to create **WORKFLOWS**



# The Data Science Life Cycle



# Over 4000 Nodes included

