

# Data Analysis



## Exploratory Data Analysis (EDA) Knime



Edgar Rios Linares

# Problem

## ▼ Problem description

### Context

The original dataset contains 1000 records with 20 categorical attributes prepared by Prof. Hofmann.

In this dataset, each record represents a person taking out a loan from a bank.

Each person is classified as **good** or **bad** credit risk based on the set of attributes.

Link to the original dataset at

[UCI Machine Learning](#)

### Content

The selected attributes:

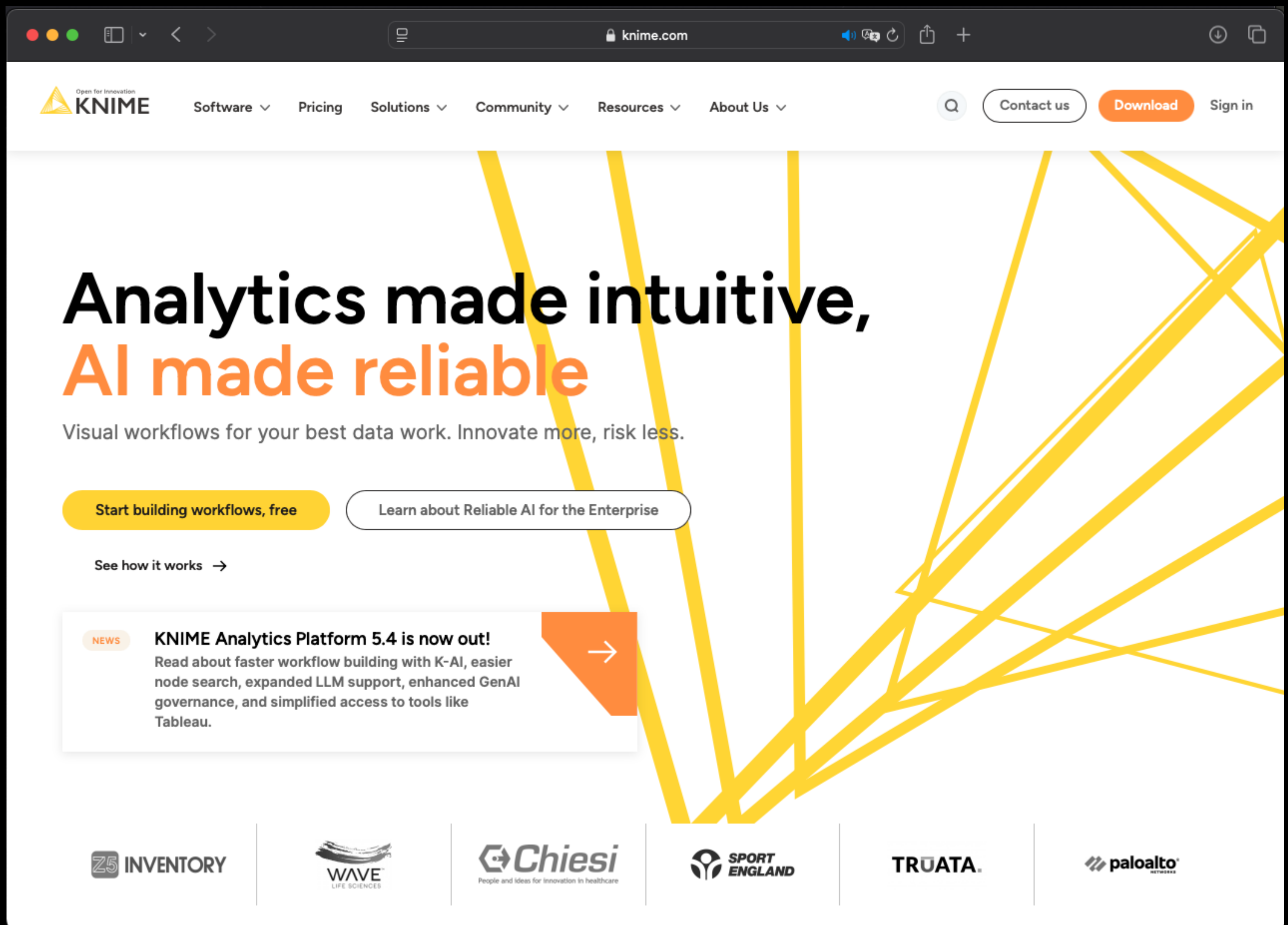
- **Age** (numeric)
- **Sex** (text: male, female)
- **Job** (numeric: 0 - unskilled and non-resident, 1 - unskilled and resident, 2 - skilled, 3 - highly skilled)
- **Housing** (text: own, rent, or free)
- **Saving accounts** (text - little, moderate, quite rich, rich)
- **Checking account** (numeric, in DM - Deutsch Mark)
- **Credit amount** (numeric, in DM)
- **Duration** (numeric, in month)
- **Purpose**(text: car, furniture/equipment, radio/TV, domestic appliances, repairs, education, business, vacation/others)
- **Risk** (Value target - Good or Bad Risk)

### Objective

Train a model to predict from new data whether a person applying for a loan represents a good or bad risk

# Download

Go website Knime, download & install app  
<https://www.knime.com>



The screenshot shows the Knime website homepage. The browser address bar displays "knime.com". The navigation bar includes the Knime logo, "Software", "Pricing", "Solutions", "Community", "Resources", and "About Us". On the right, there are links for "Contact us", "Download", and "Sign in". The main content area features the headline "Analytics made intuitive, AI made reliable" with "AI made reliable" in orange. Below this is the tagline "Visual workflows for your best data work. Innovate more, risk less." Two buttons are present: "Start building workflows, free" and "Learn about Reliable AI for the Enterprise". A link "See how it works" with a right arrow is also shown. A news box on the left titled "KNIME Analytics Platform 5.4 is now out!" contains details about new features like K-AI, LLM support, and GenAI governance. The footer displays logos for partner companies: Z5 INVENTORY, WAVE LIFE SCIENCES, Chiesi, SPORT ENGLAND, TRUATA, and paloalto NETWORKS.

Open for Innovation  
**KNIME**

Software ▾ Pricing Solutions ▾ Community ▾ Resources ▾ About Us ▾

Q Contact us Download Sign in

# Analytics made intuitive, AI made reliable

Visual workflows for your best data work. Innovate more, risk less.

Start building workflows, free

Learn about Reliable AI for the Enterprise

See how it works →

**NEWS** **KNIME Analytics Platform 5.4 is now out!**  
Read about faster workflow building with K-AI, easier node search, expanded LLM support, enhanced GenAI governance, and simplified access to tools like Tableau.

**Z5 INVENTORY**

**WAVE**  
LIFE SCIENCES

**Chiesi**  
People and Ideas for Innovation in healthcare

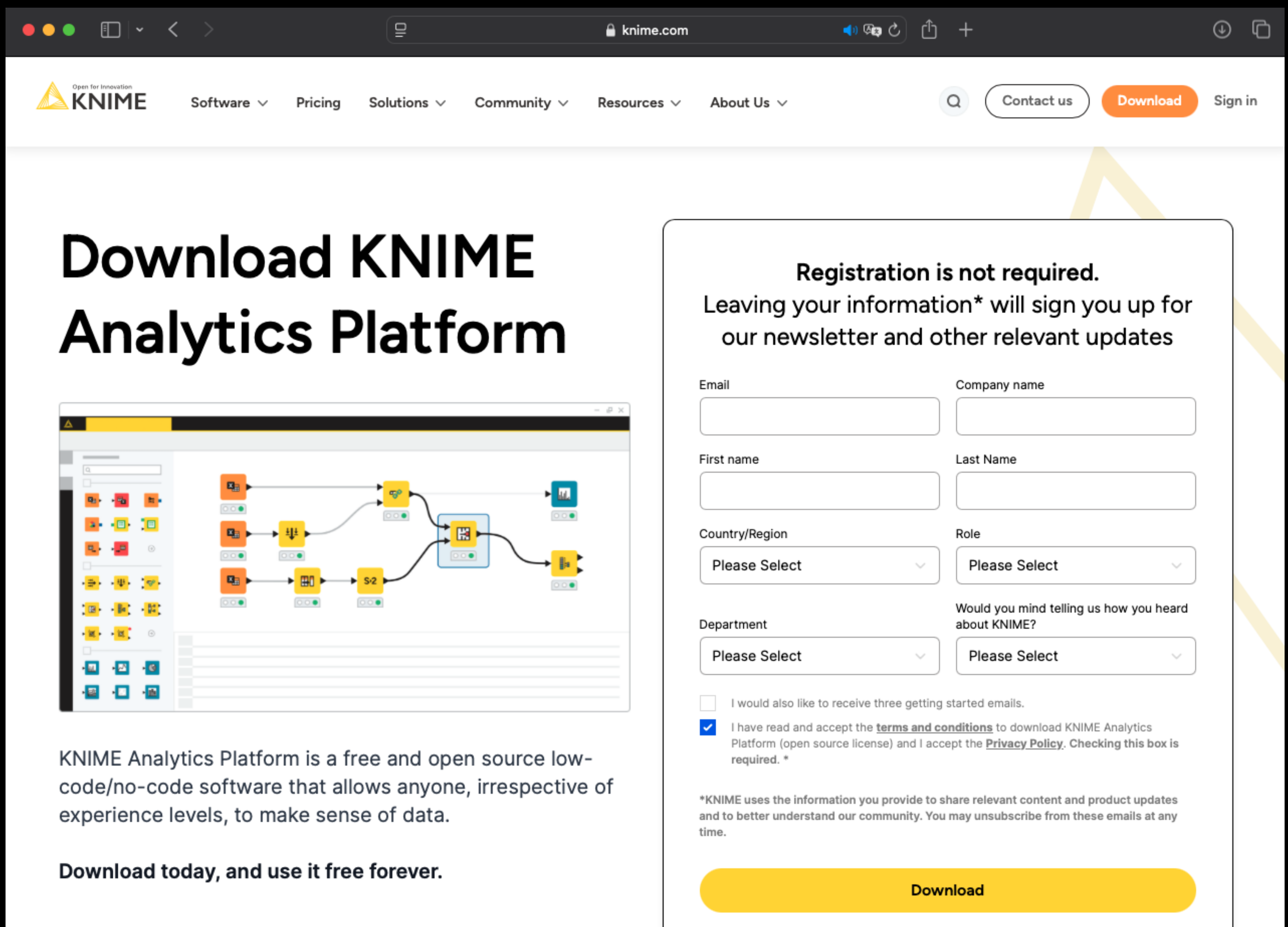
**SPORT ENGLAND**

**TRUATA**

**paloalto**  
NETWORKS


# Download

# Don't fill form, only check advise & click download



# Download

Select your SO, install & launch app



Open for Innovation

Software

Pricing

Solutions

Community

Resources

About Us

Q

Contact us

Download

Sign in

Windows

Microsoft Defender SmartScreen may block download in its attempt to prevent malicious software installations.  
To solve the problem [click here](#).

KNIME Analytics Platform for Windows (installer)	<a href="#">Download</a> (591 MB)
KNIME Analytics Platform for Windows (self-extracting archive)	<a href="#">Download</a> (593MB)
KNIME Analytics Platform for Windows (zip archive)	<a href="#">Download</a> (709 MB)

Linux

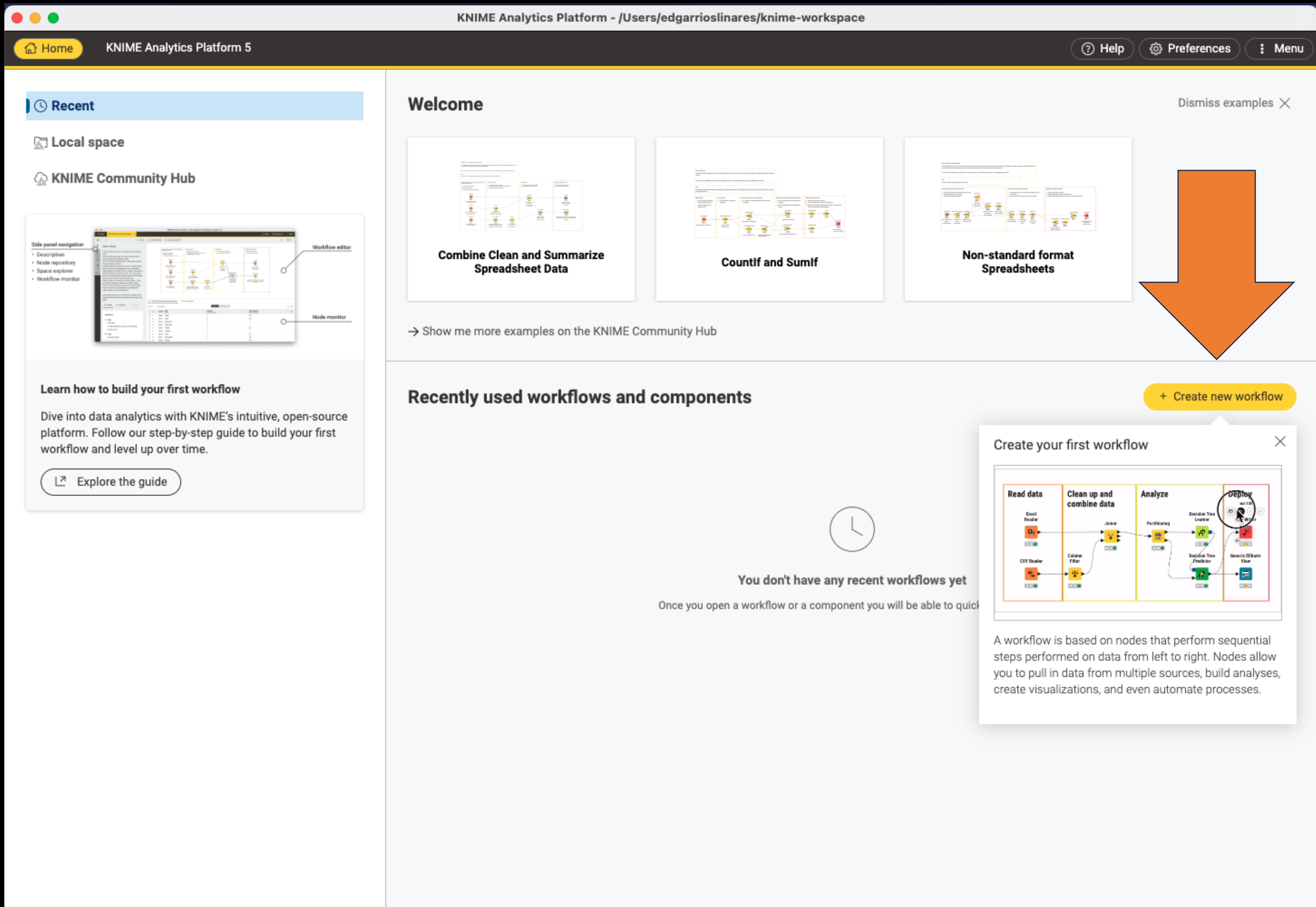
KNIME Analytics Platform for Linux	<a href="#">Download</a> (718 MB)
------------------------------------	-----------------------------------

Mac

KNIME Analytics Platform for macOS x86_64 (Intel)	<a href="#">Download</a> (671 MB)
---	-----------------------------------



# New Project



# Load File

HomeKNIME\_EDA+HelpPreferencesMenu

ExecuteCancelResetCreate metanodeCreate component

100%

InfoNodesExplorerK-AIMonitor

Nodes

Search all nodes

IO

Excel ReaderExcel WriterMicrosoft Authenticator

Google Sheets ReaderGoogle Sheets Writer

CSV ReaderCSV WriterShow all

Manipulation

Row FilterColumn FilterConcatenate

Value LookupRow AggregatorJoiner

String CleanerGroupByShow all

Views

Bar ChartLine PlotPie Chart

Stacked Area ChartScatter PlotStatistics View

HeatmapHistogramShow all

CSV Reader

Add comment

1: File Table

Flow Variables

Rows: 1000 | Columns: 11

TableStatistics

SearchFilter

	#	RowID	Column0 Number (inte...	age Number (inte...	sex String	job Number (inte...	housing String	saving_ac... String	checking ... String	credit am... Number (inte...	duratio Number (
<input type="checkbox"/>	1	Row0	0	67	male	2	own	NA	little	1169	6
<input type="checkbox"/>	2	Row1	1	22	female	2	own	little	moderate	5951	48
<input type="checkbox"/>	3	Row2	2	49	male	1	own	little	NA	2096	12
<input type="checkbox"/>	4	Row3	3	45	male	2	free	little	little	7882	42
<input type="checkbox"/>	5	Row4	4	53	male	2	free	little	little	4870	24
<input type="checkbox"/>	6	Row5	5	35	male	1	free	NA	NA	9055	36
<input type="checkbox"/>	7	Row6	6	53	male	2	own	quite rich	NA	2835	24
<input type="checkbox"/>	8	Row7	7	35	male	3	rent	little	moderate	6948	36
<input type="checkbox"/>	9	Row8	8	61	male	1	own	rich	NA	3059	12

# Load File

Home

KN

Nodes

Info

Nodes

Explorer

K-AI

Monitor

Nodes

Search

IO

Excel Reader

Google Authenticator

CSV Reader

Manipulation

Row Filter

Value Lookup

String Cleaner

Views

Bar Chart

Stacked Area Chart

Heatmap

Dialog - 3:1 - CSV Reader

SettingsTransformationAdvanced SettingsLimit RowsEncodingFlow VariablesJob Manager SelectionMemory Policy

Input location

Read fromLocal File System

ModeFileFiles in folder

File/Users/edgarrioslinares/Documents/academicos/ideas\_proyectos\_IA/365\_AI\_Journey/Day\_029/german\_credit\_data.csvBrowse...

Reader options

Format

Autodetect format

Column delimiterRow delimiterLine breakCustom\n

Quote charQuote escape char

#Comment char

Has column headerHas RowID

Support short data rowsPrepend file index to RowID

Preview

The suggested column types are based on the first 10000 rows only. See 'Advanced Settings' tab.

Row ID	I	Colum...	I	age	S	sex	I	job	S	housing	S	saving...	S	checki...	I	credit ...	I	duration	S	purpose	S	risk
Row0	0			67		male		2		own		NA		little		1169		6		radio/TV		good
Row1	1			22		female		2		own		little		moderate		5951		48		radio/TV		bad
Row2	2			49		male		1		own		little		NA		2096		12		education		good
Row3	3			45		male		2		free		little		little		7882		42		furniture/...		good
Row4	4			53		male		2		free		little		little		4870		24		car		bad
Row5	5			35		male		1		free		NA		NA		9055		36		education		good
Row6	6			53		male		2		own		quite rich		NA		2835		24		furniture/...		good
Row7	7			35		male		3		rent		little		moderate		6948		36		car		good
Row8	8			61		male		1		own		rich		NA		3059		12		radio/TV		good
Row9	9			28		male		3		own		little		moderate		5234		30		car		bad
Row10	10			25		female		2		rent		little		moderate		1295		12		car		bad
Row11	11			24		female		2		rent		little		little		4308		48		business		bad
Row12	12			22		female		2		own		little		moderate		1567		12		radio/TV		good
Row13	13			60		male		1		own		little		little		1199		24		car		bad
Row14	14			28		female		2		rent		little		little		1403		15		car		good
Row15	15			32		female		1		own		moderate		little		1282		24		radio/TV		bad
Row16	16			53		male		2		own		NA		NA		2424		24		radio/TV		good
Row17	17			25		male		2		own		NA		little		8072		30		business		good

OKApplyCancel?



# View Statistics

HomeKNIME\_EDA+HelpPreferencesMenu

ExecuteCancelResetCreate metanodeCreate component100%

Nodes

Value Lookup

Row Aggregator

Joiner

String Cleaner

GroupBy

Statistics View

Views

Bar Chart

Line Plot

Stacked Area Chart

Scatter Plot

Heatmap

Histogram

Show all

Analytics

Decision Tree Predictor

Decision Tree Learner

Regression Predictor

Linear Regression Learner

Scorer

k-Means

Statistics

Logistic Regression...

Show all

CSV Reader

Statistics View

Add comment

ViewFlow Variables

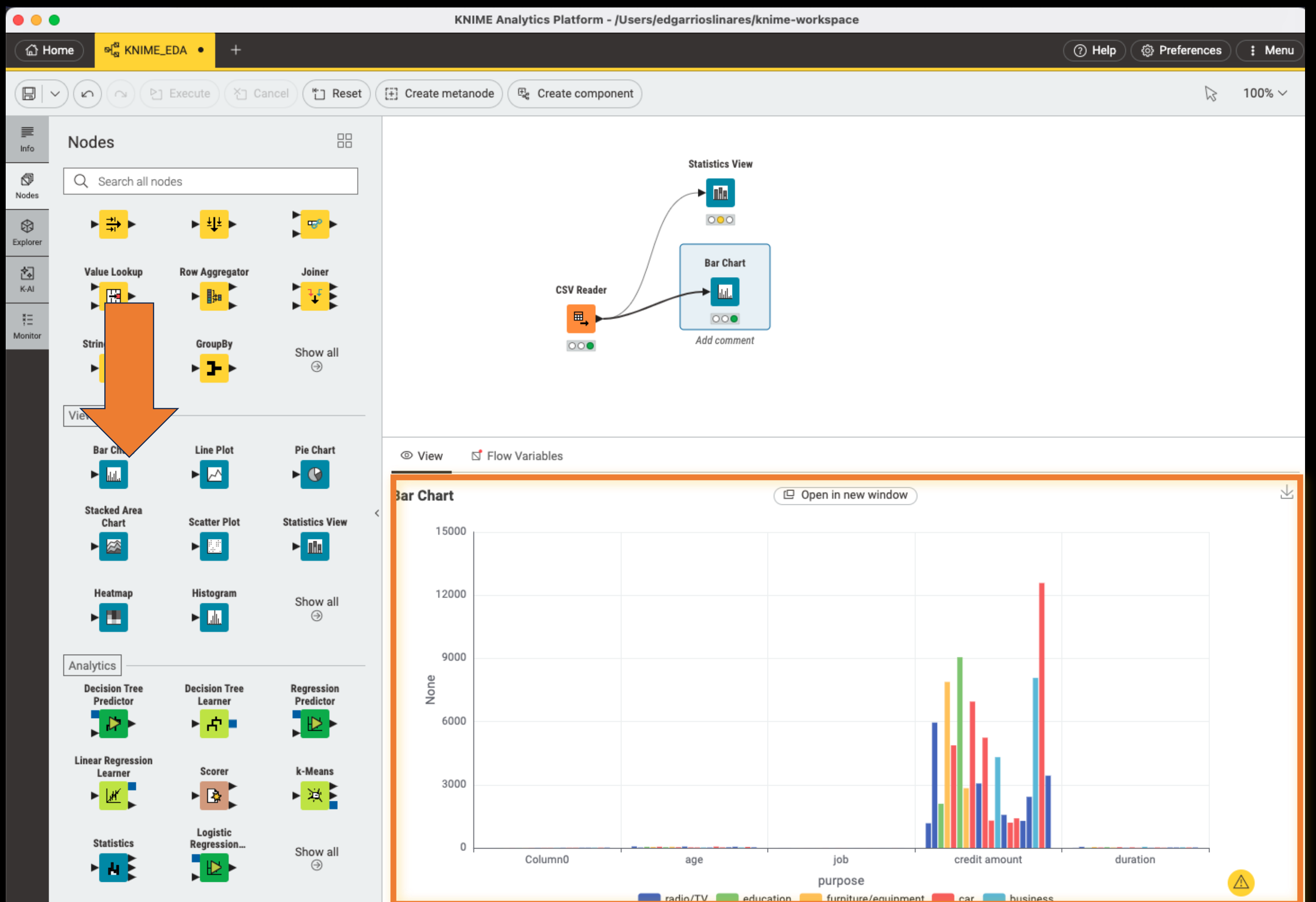
Statistics

Open in new window

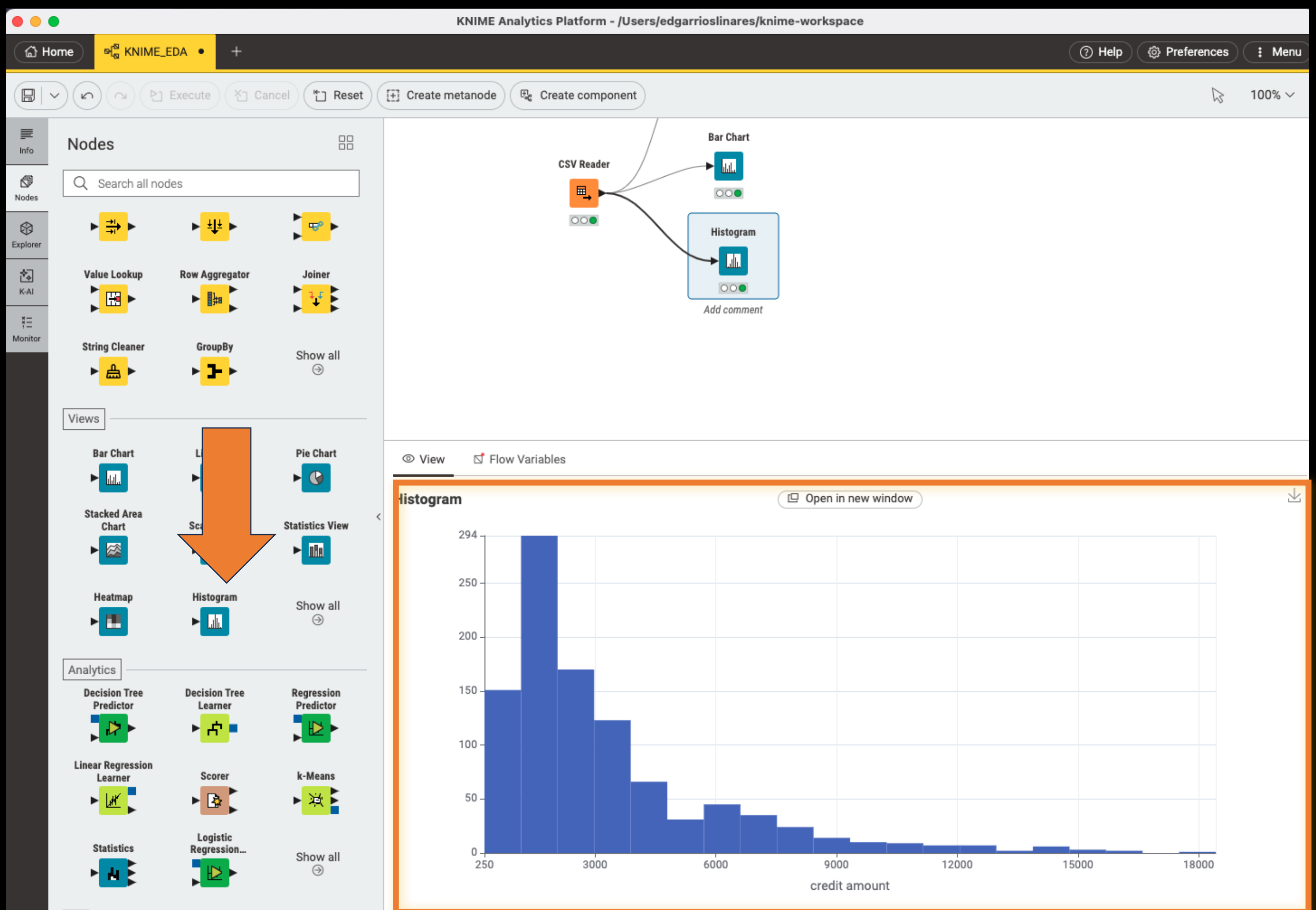
Rows: 11 | Columns: 14

Name	Type	# Missing val...	# Unique valu...	Minimum	Maximum	25% Quantile	50% Quantile ...	75% Quantile	Mean
Column0	Number (intege	0	1000	0	999	249.25	499.5	749.75	499.5
age	Number (intege	0	53	19	75	27	33	42	35.546
sex	String	0	2	?	?	?	?	?	?
job	Number (intege	0	4	0	3	2	2	2	1.904
housing	String	0	3	?	?	?	?	?	?
saving_account	String	0	5	?	?	?	?	?	?
checking accou	String	0	4	?	?	?	?	?	?
credit amount	Number (intege	0	921	250	18,424	1,364.5	2,319.5	3,972.75	3,271.258
duration	Number (intege	0	33	4	72	12	18	24	20.903

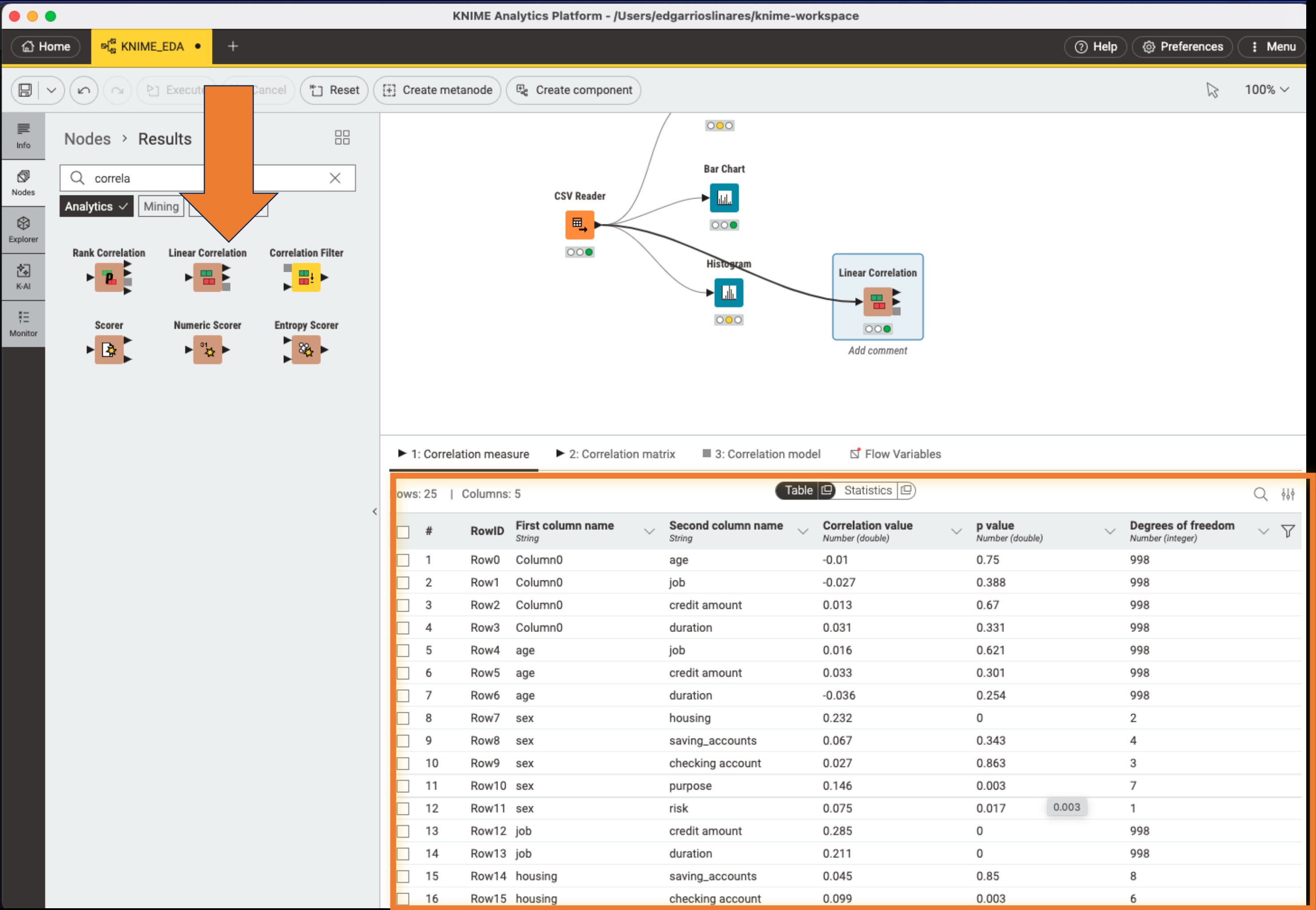
# View Bar Chart



# View Histogram

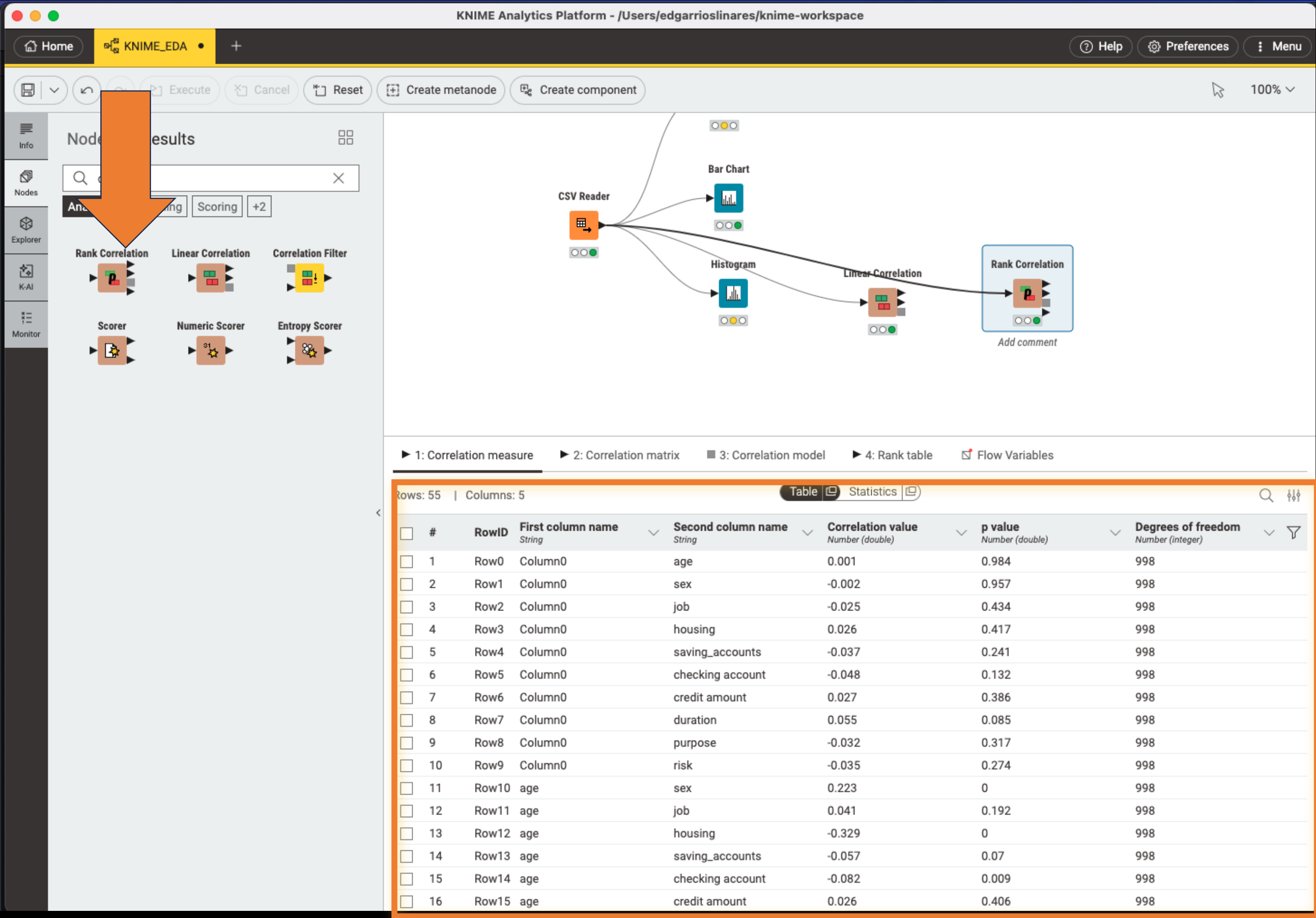


# View Linear Correlation

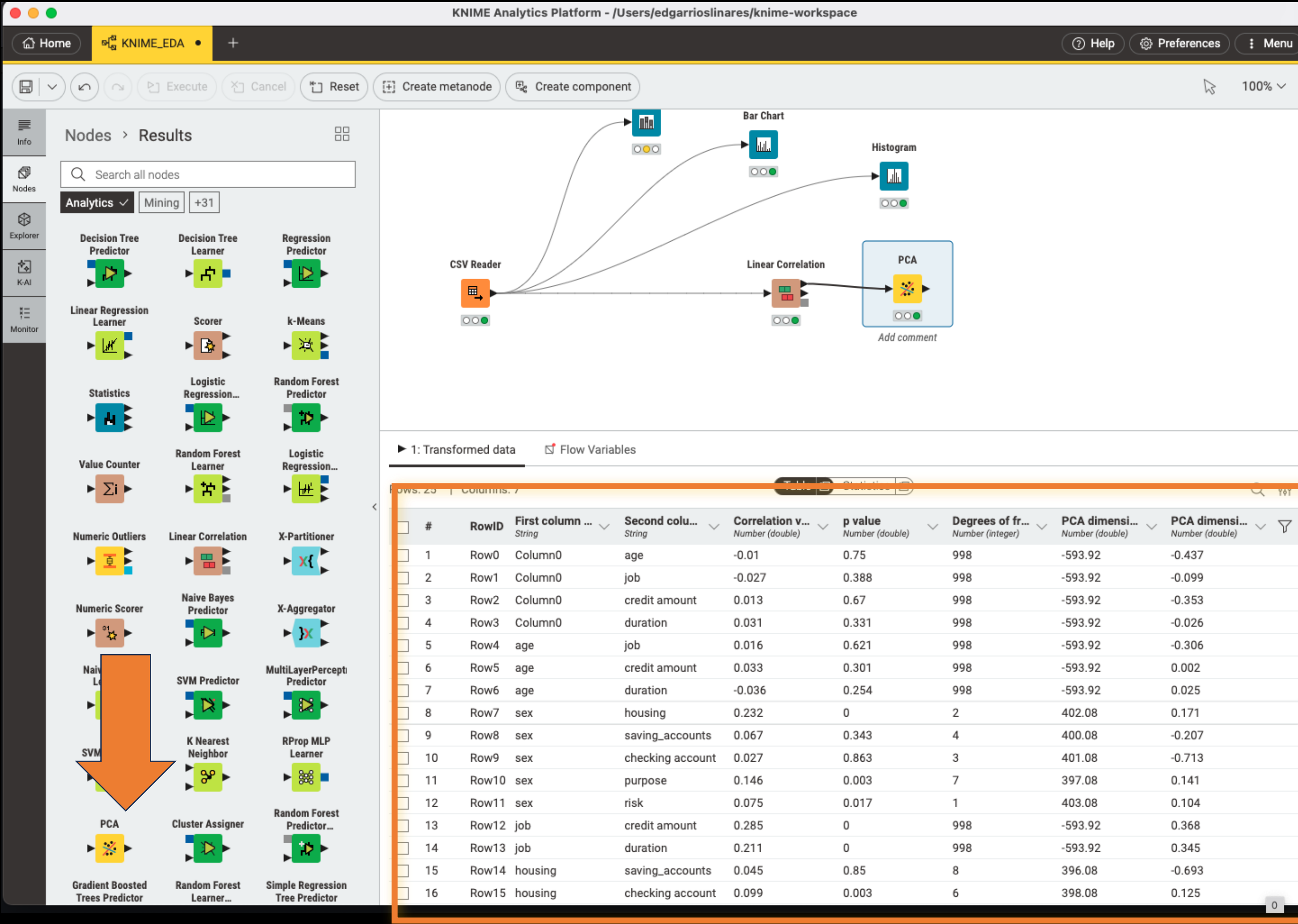




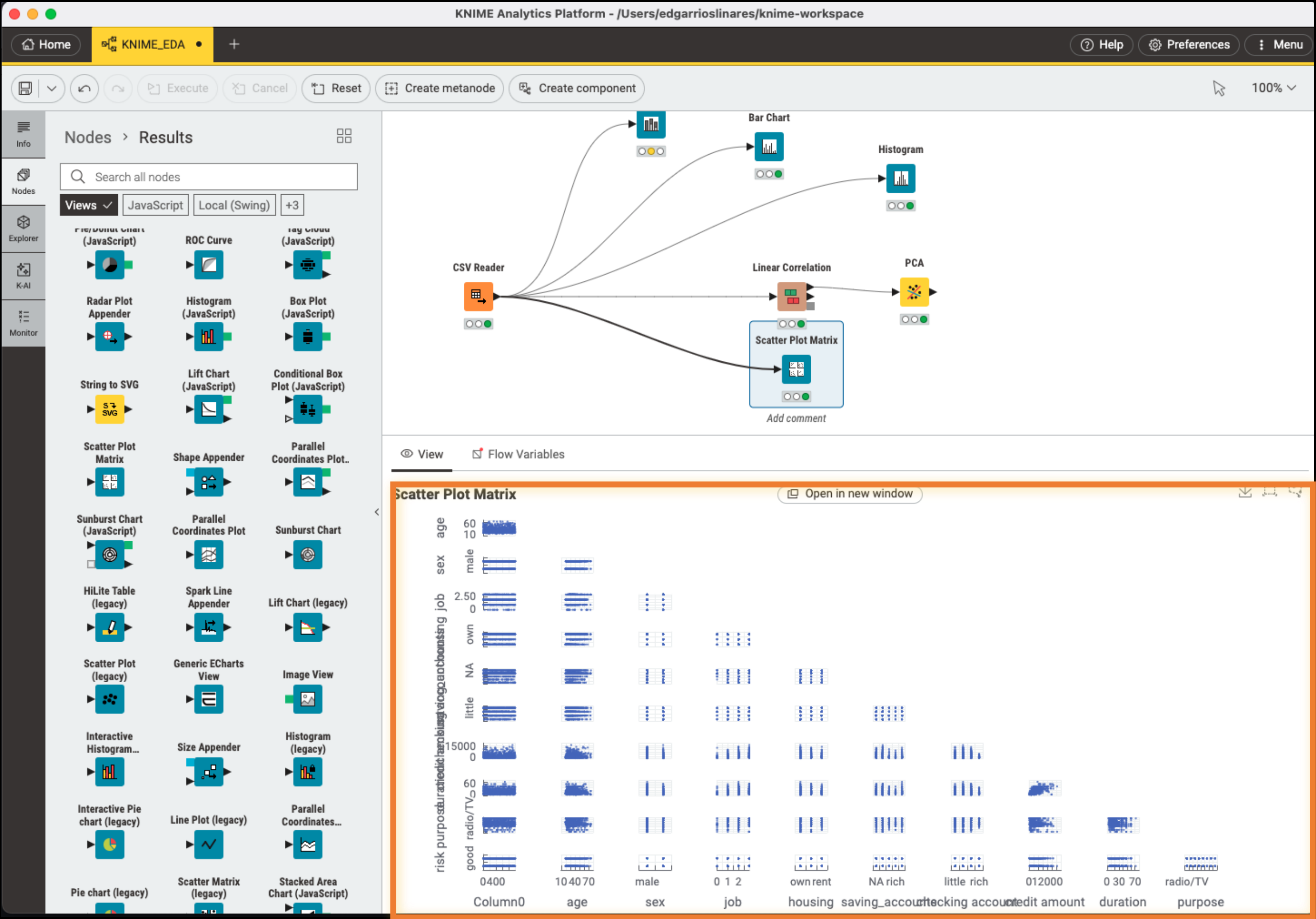
# View Rank Correlation



# Make PCA



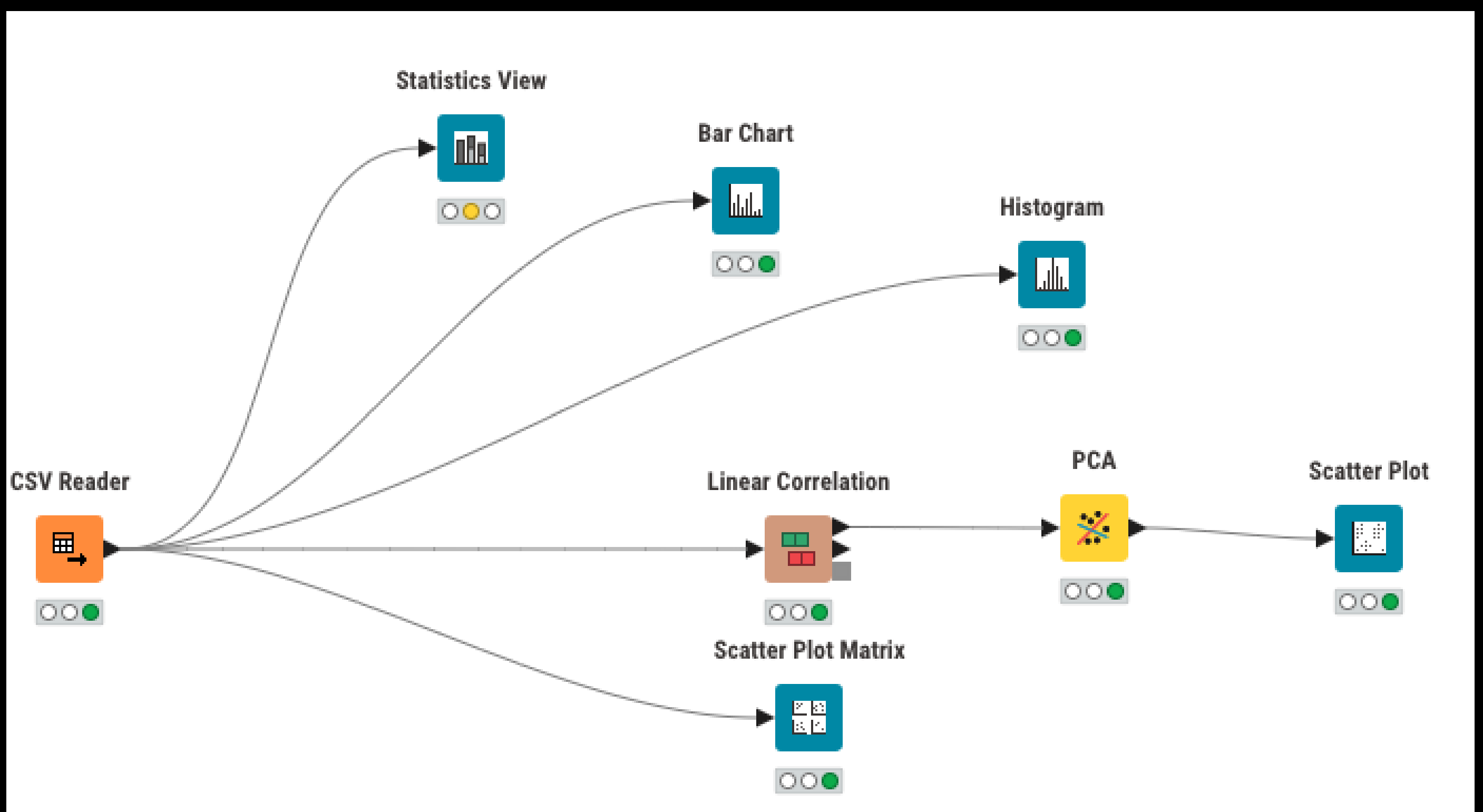
# View Scatter Plot Matrix





# The work is done, great job!"

## You have a basic EDA quickly





Educator in AI

**Artificial  
Intelligence**

**Data Engineering**



**Machine Learning**

**Data Science**

📌 **Linkedin** —> <https://www.linkedin.com/in/erlinares/>

👋 **Follow us on X**: <https://x.com/erlinares><sup>[SEP]</sup>

💻 **GitHub**: [https://github.com/erlinares/365\\_AI\\_Journey/](https://github.com/erlinares/365_AI_Journey/)

💬 **Discord**: <https://discord.gg/5fFM2zh8>



**Edgar Rios Linares**