Data Analysis

Exploratory Data Analysis (EDA) Orange Data Mining

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)

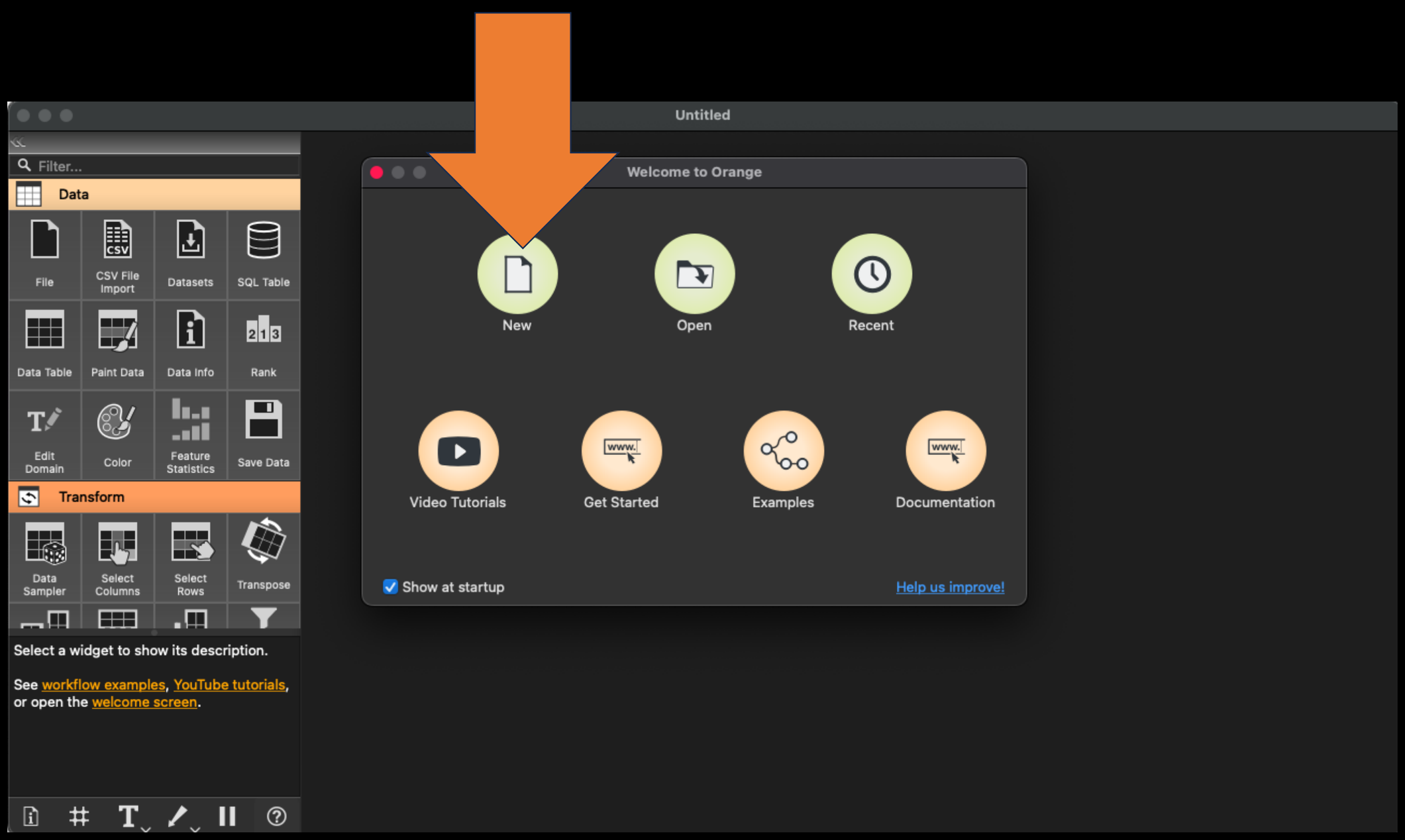
Objective

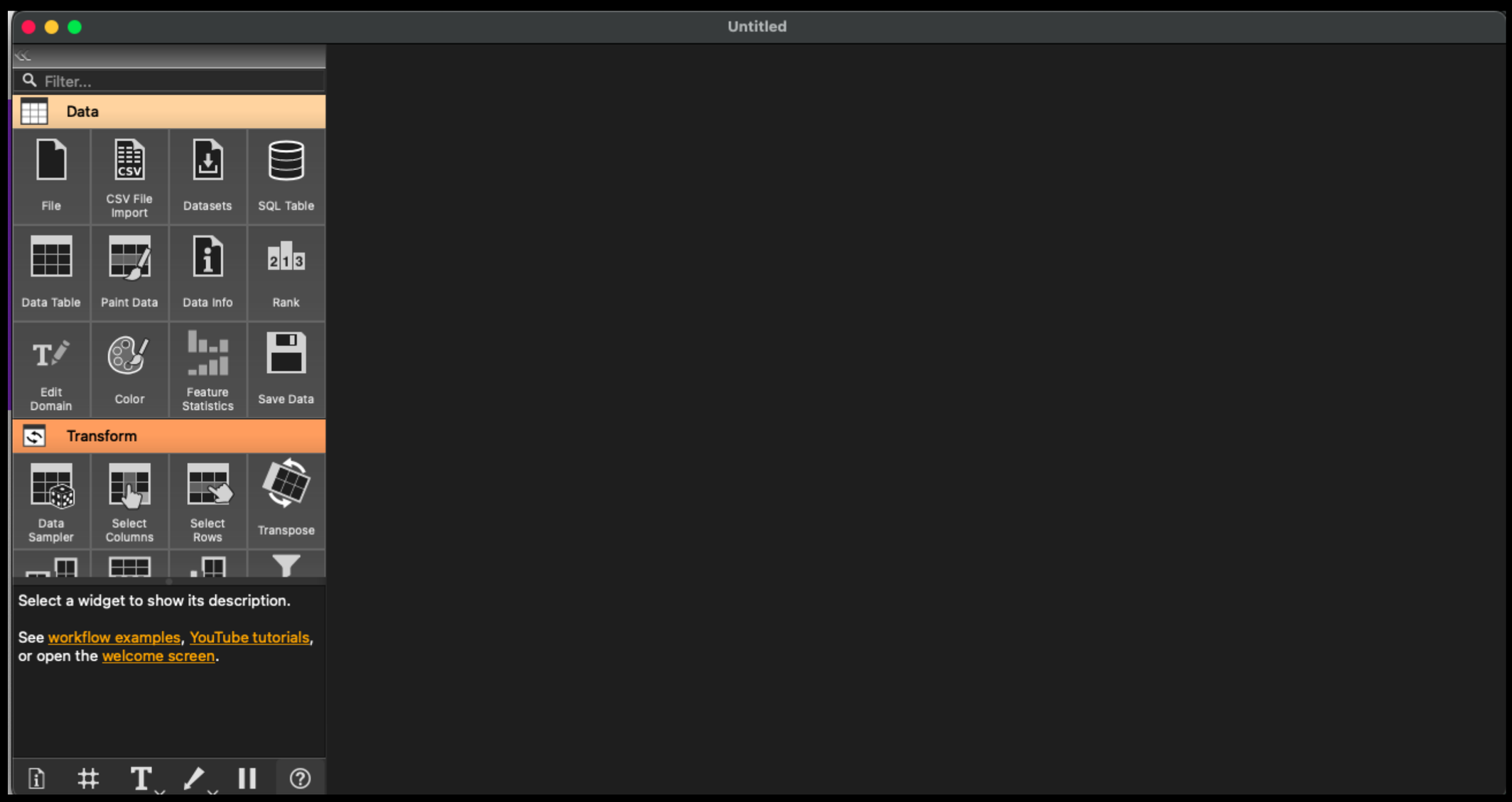
Train a model to predict from new data whether a person applying for a loan represents a good or bad risk

Download

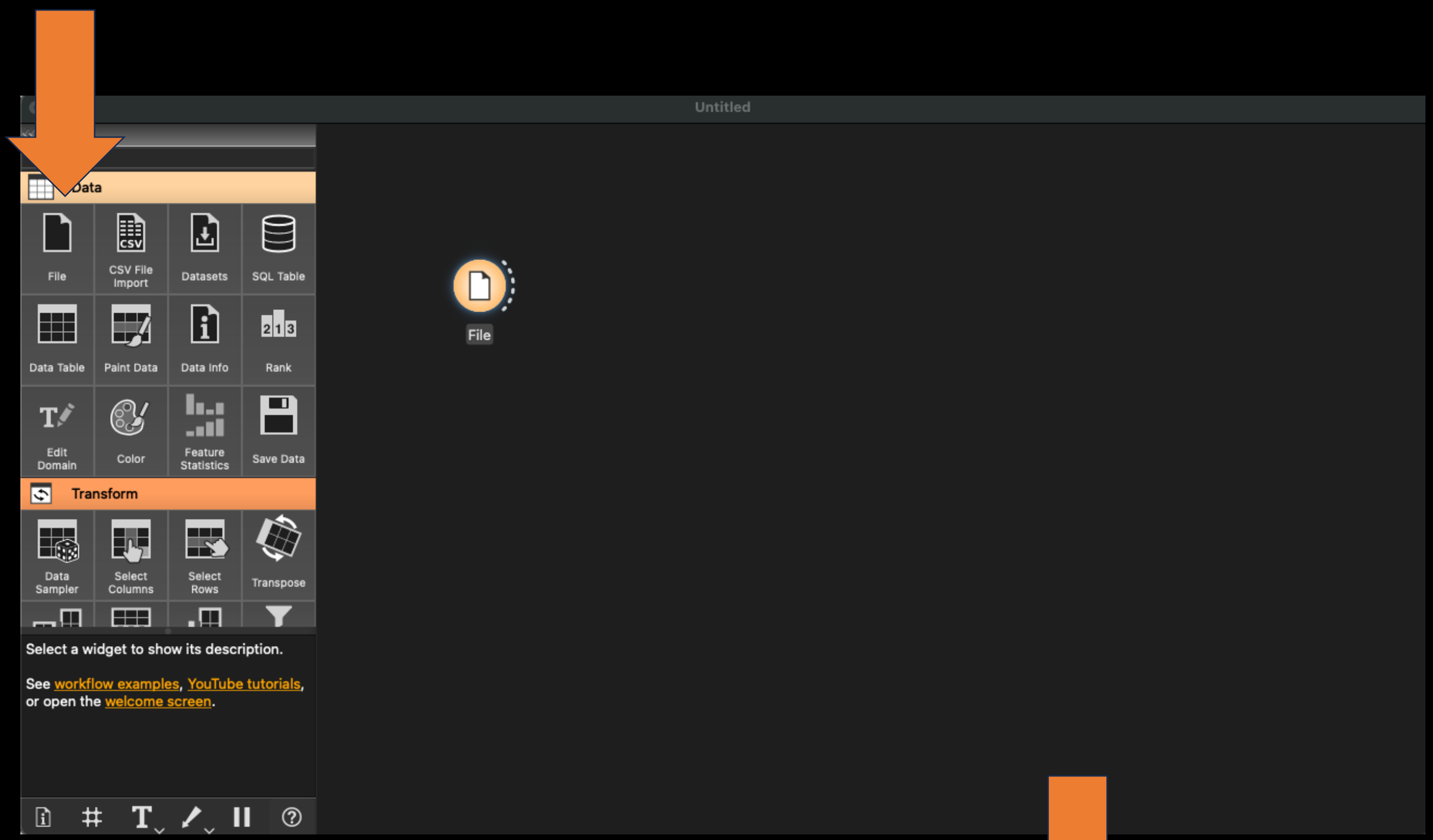
Go website Orange, download & install app
https://orangedatamining.com

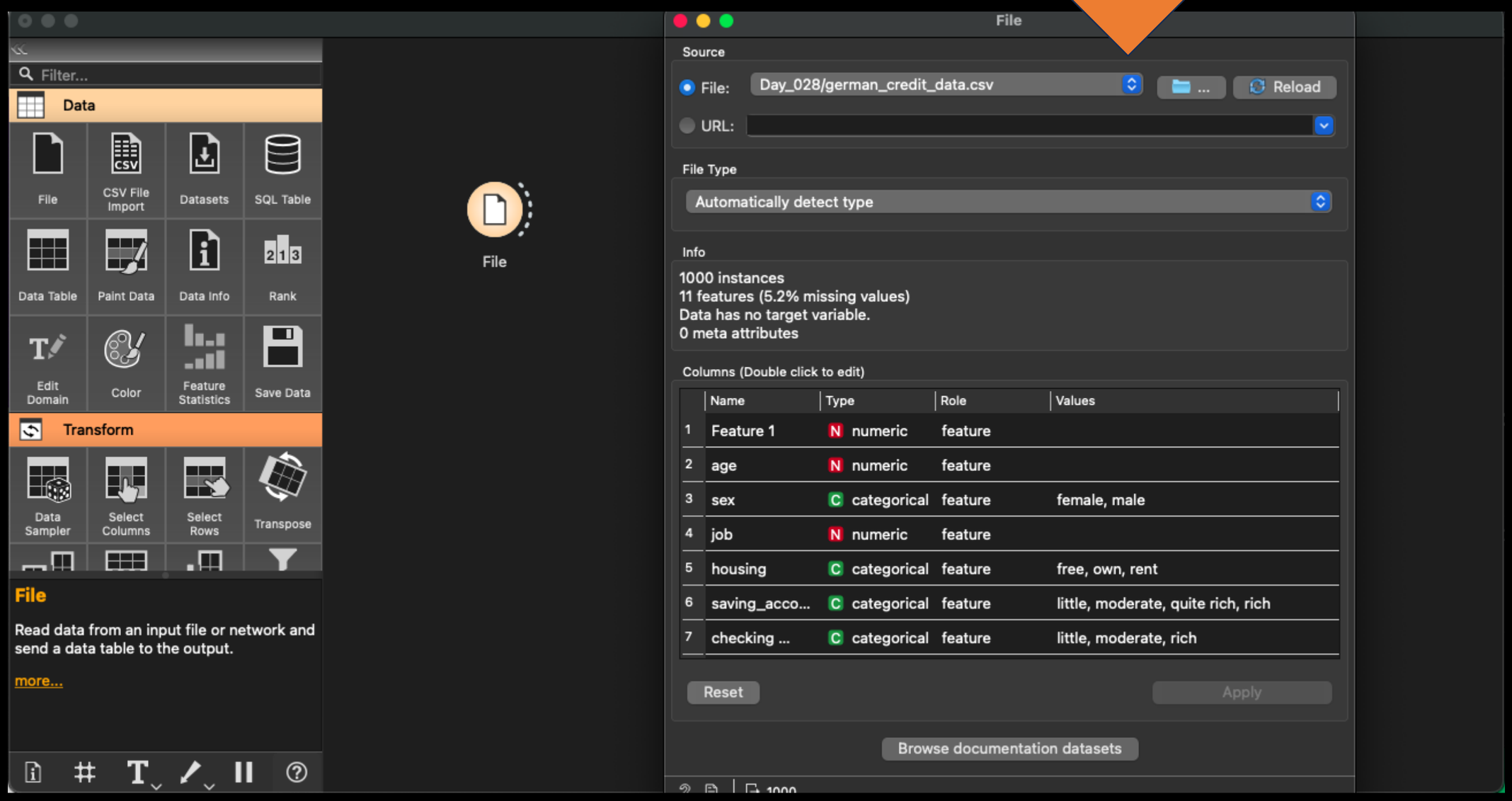
New Project


Untitled
Filter...
Data
File
CSV File Import
Datasets
SQL Table
Data Table
Paint Data
Data Info
Rank
Edit Domain
Color
Feature Statistics
Save Data
Transform
Data Sampler
Select Columns
Select Rows
Transpose
Select a widget to show its description.
See workflow examples, YouTube tutorials, or open the welcome screen.
Welcome to Orange
New
Open
Recent
Video Tutorials
Get Started
Examples
Documentation
Show at startup
Help us improve!

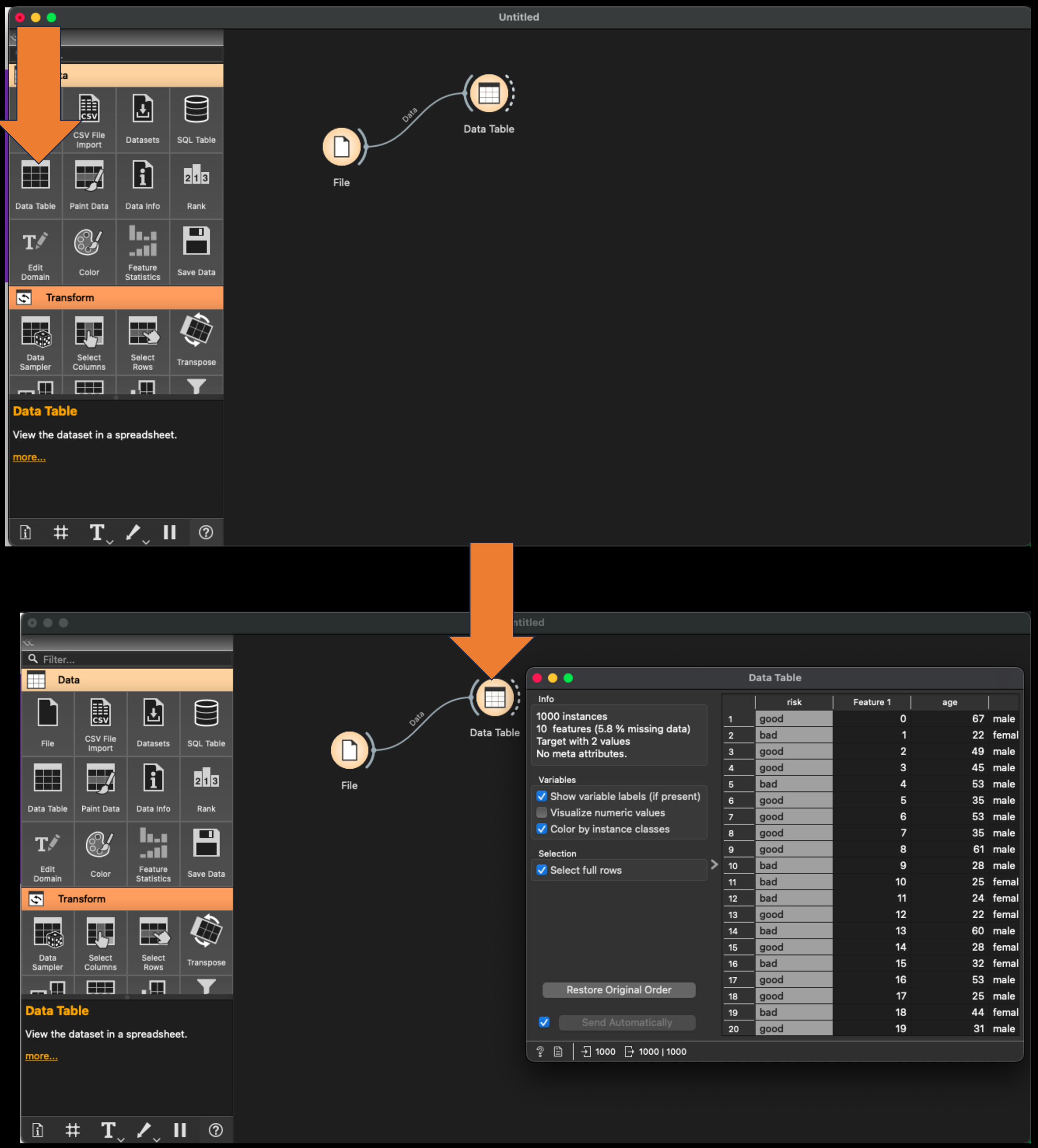

Untitled
Filter...
Data
File
CSV File Import
Datasets
SQL Table
Data Table
Paint Data
Data Info
Rank
Edit Domain
Color
Feature Statistics
Save Data
Transform
Data Sampler
Select Columns
Select Rows
Transpose
Select a widget to show its description.
See workflow examples, YouTube tutorials, or open the welcome screen.

Load File

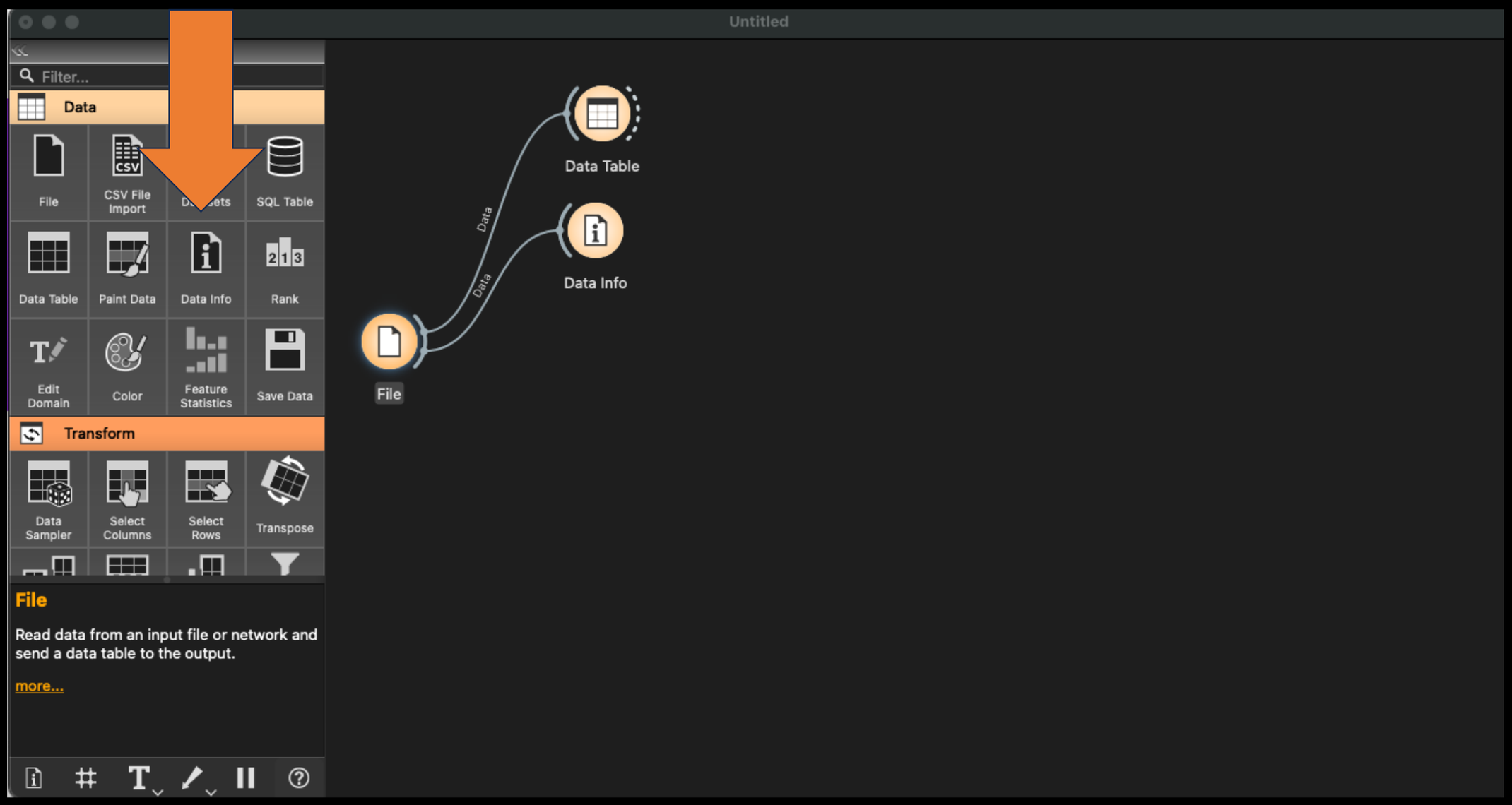

Untitled
Data
File
CSV File Import
Datasets
SQL Table
Data Table
Paint Data
Data Info
Rank
Edit Domain
Color
Feature Statistics
Save Data
Transform
Data Sampler
Select Columns
Select Rows
Transpose
Select a widget to show its description.
See workflow examples, YouTube tutorials, or open the welcome screen.
File

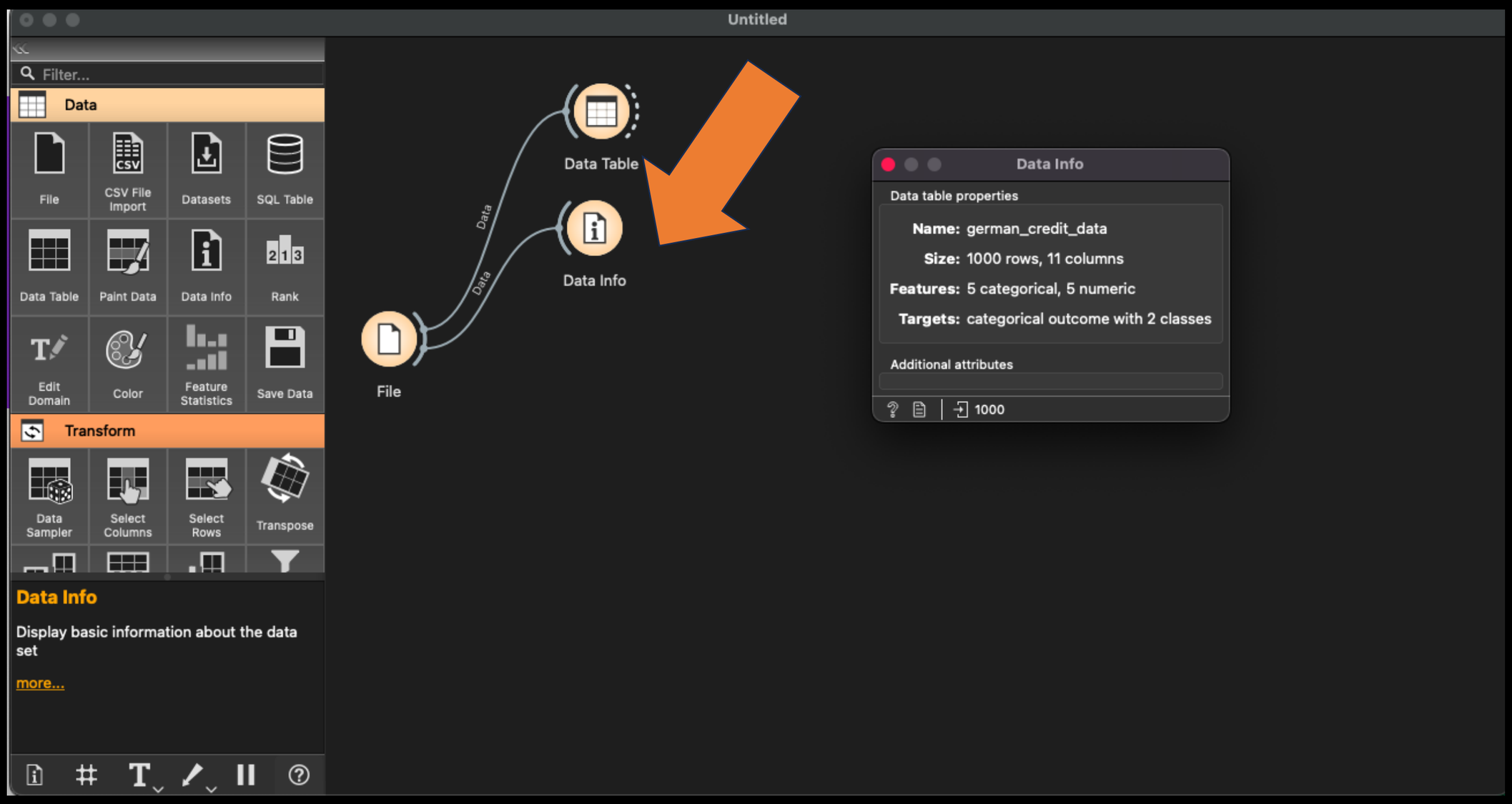

Filter...
Data
File
CSV File Import
Datasets
SQL Table
Data Table
Paint Data
Data Info
Rank
Edit Domain
Color
Feature Statistics
Save Data
Transform
Data Sampler
Select Columns
Select Rows
Transpose
File
Read data from an input file or network and send a data table to the output.
more...
File
Source
File: Day_028/german_credit_data.csv
Reload
URL:
File Type
Automatically detect type
Info
1000 instances
11 features (5.2% missing values)
Data has no target variable.
0 meta attributes
Columns (Double click to edit)
Name Type Role Values
1 Feature 1 numeric feature
2 age numeric feature
3 sex categorical feature female, male
4 job numeric feature
5 housing categorical feature free, own, rent
6 saving_acco... categorical feature little, moderate, quite rich, rich
7 checking ... categorical feature little, moderate, rich
Reset
Apply
Browse documentation datasets

Data Table

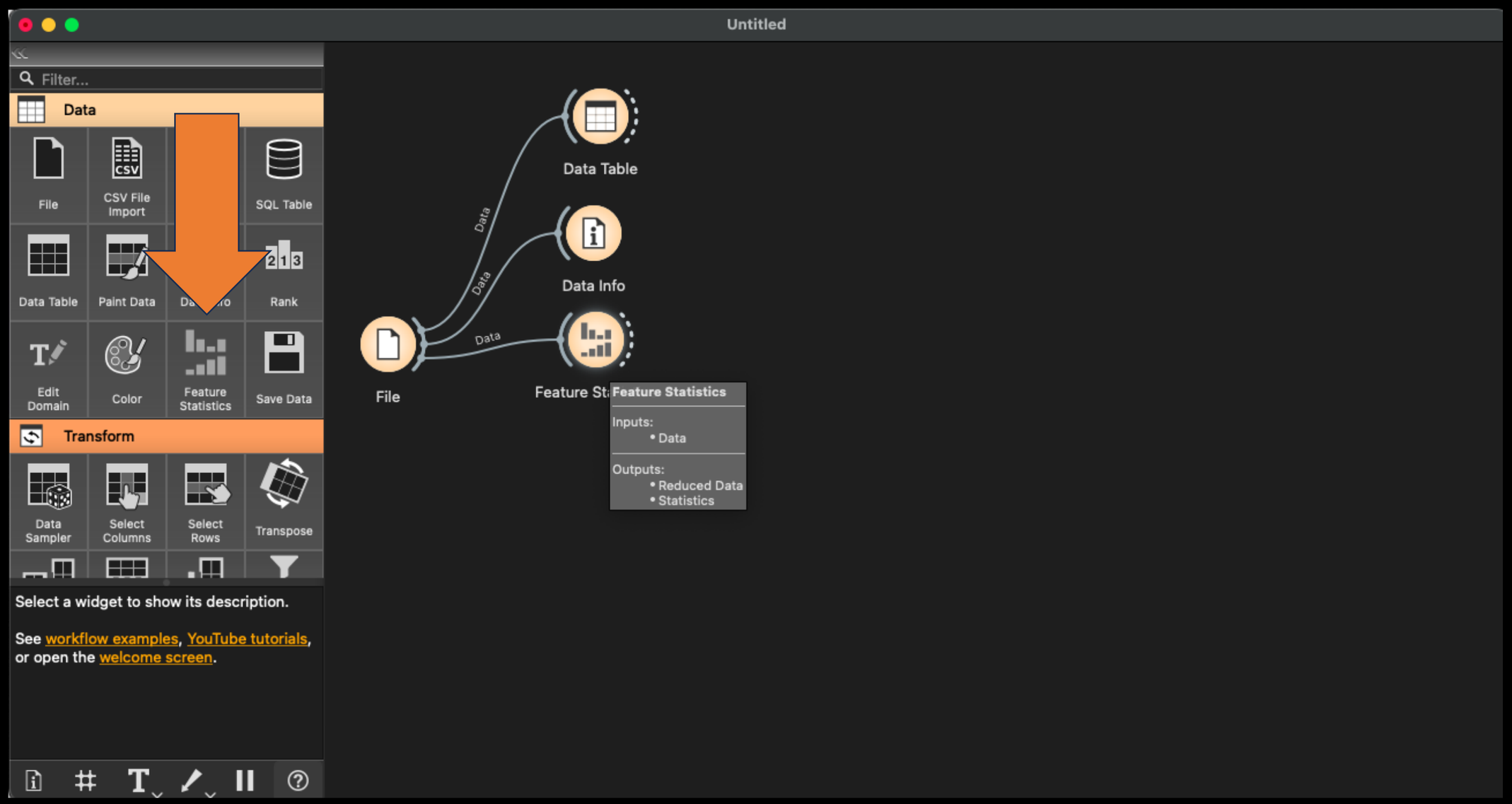

Untitled
CSV File Import
Datasets
SQL Table
Data Table
Paint Data
Data Info
Rank
Edit Domain
Color
Feature Statistics
Save Data
Transform
Data Sampler
Select Columns
Select Rows
Transpose
Data Table
View the dataset in a spreadsheet.
more...
Data
File
Data
Data Table
Filter...
Data Table
Info
1000 instances
10 features (5.8 % missing data)
Target with 2 values
No meta attributes.
Variables
Show variable labels (if present)
Visualize numeric values
Color by instance classes
Selection
Select full rows
Restore Original Order
Send Automatically
1000
1000 | 1000
risk
Feature 1
age
1 good 0 67 male
2 bad 1 22 femal
3 good 2 49 male
4 good 3 45 male
5 bad 4 53 male
6 good 5 35 male
7 good 6 53 male
8 good 7 35 male
9 good 8 61 male
10 bad 9 28 male
11 bad 10 25 femal
12 bad 11 24 femal
13 good 12 22 femal
14 bad 13 60 male
15 good 14 28 femal
16 bad 15 32 femal
17 good 16 53 male
18 good 17 25 male
19 bad 18 44 femal
20 good 19 31 male

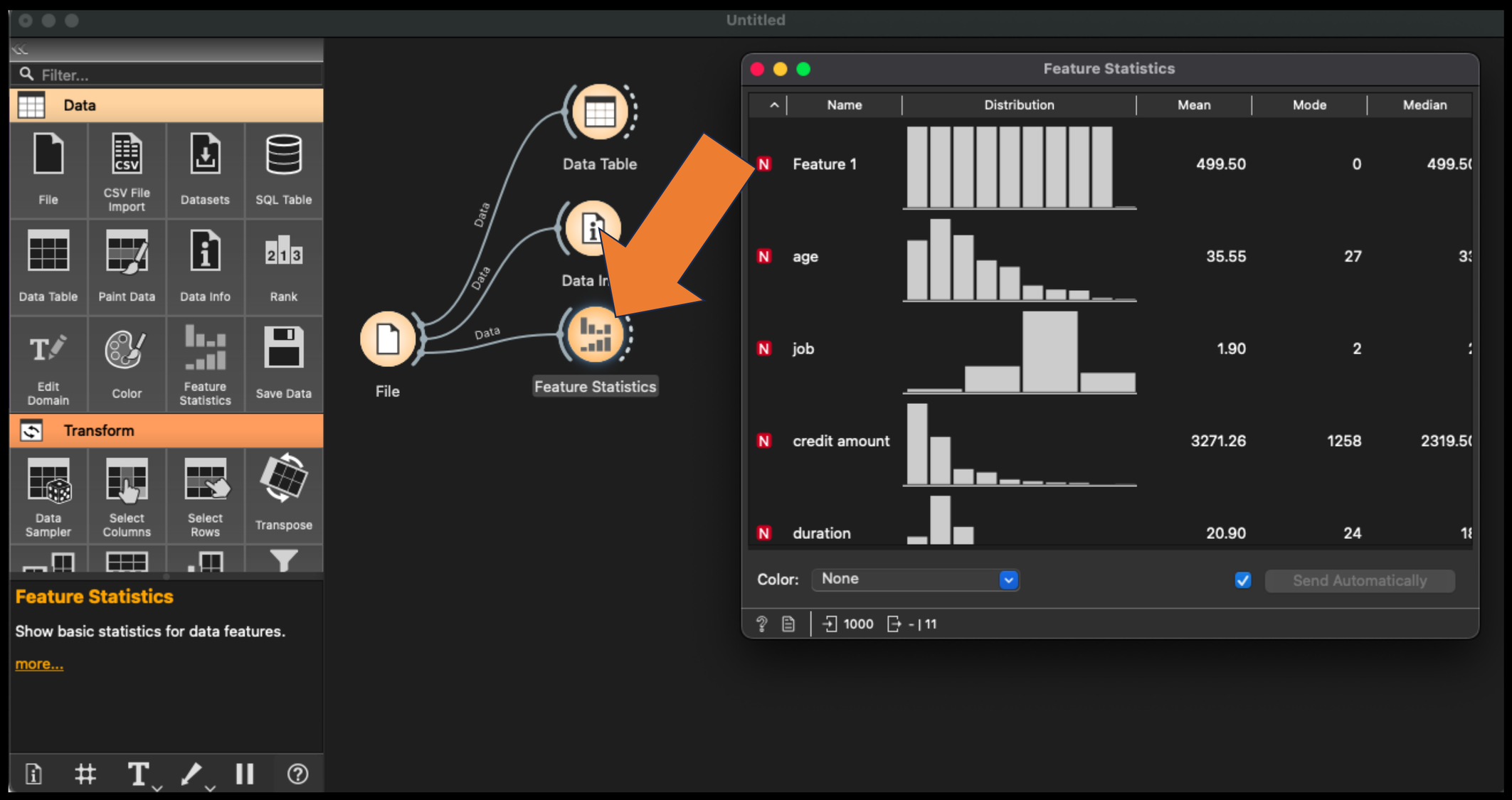
Data Info


Untitled
Filter...
Data
File
CSV File Import
Datasets
SQL Table
Data Table
Paint Data
Data Info
Rank
Edit Domain
Color
Feature Statistics
Save Data
Transform
Data Sampler
Select Columns
Select Rows
Transpose
File
Read data from an input file or network and send a data table to the output.
more...
Data
Data Table
Data Info
File

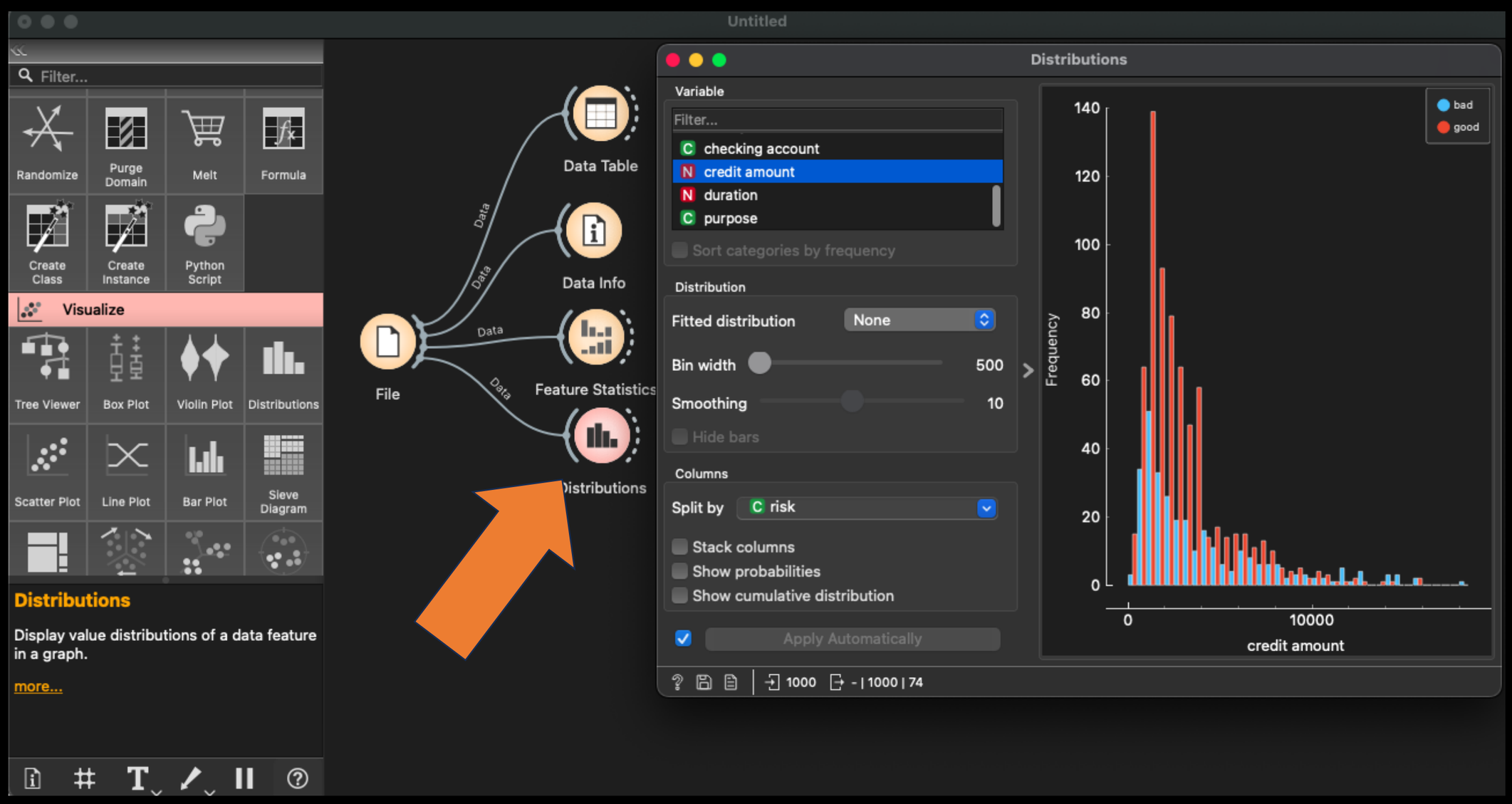

Untitled
Filter...
Data
File
CSV File Import
Datasets
SQL Table
Data Table
Paint Data
Data Info
Rank
Edit Domain
Color
Feature Statistics
Save Data
Transform
Data Sampler
Select Columns
Select Rows
Transpose
Data Info
Display basic information about the data set
more...
Data
Data Table
Data Info
File
Data Info
Data table properties
Name: german_credit_data
Size: 1000 rows, 11 columns
Features: 5 categorical, 5 numeric
Targets: categorical outcome with 2 classes
Additional attributes
1000

Feature Statistics

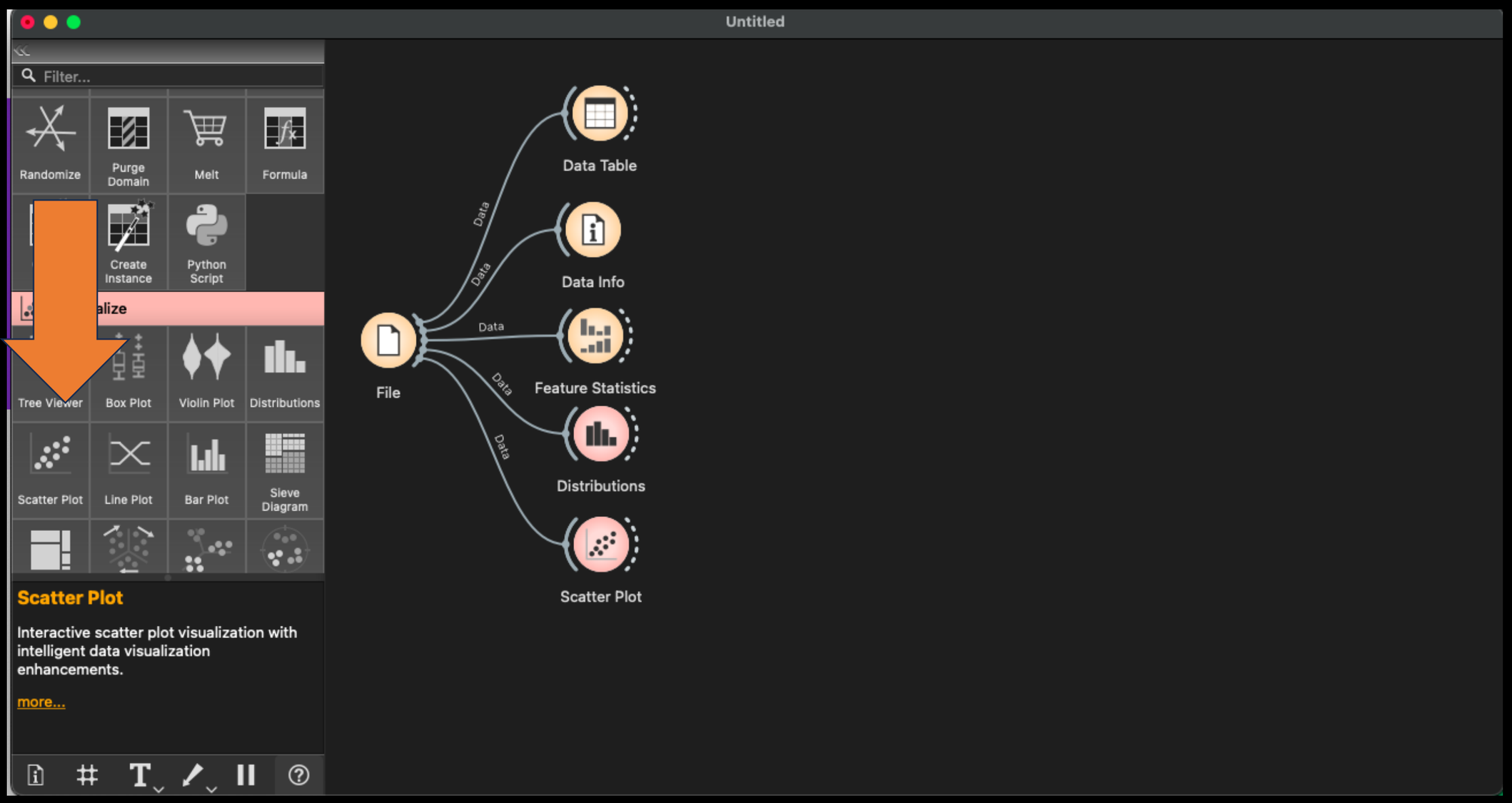


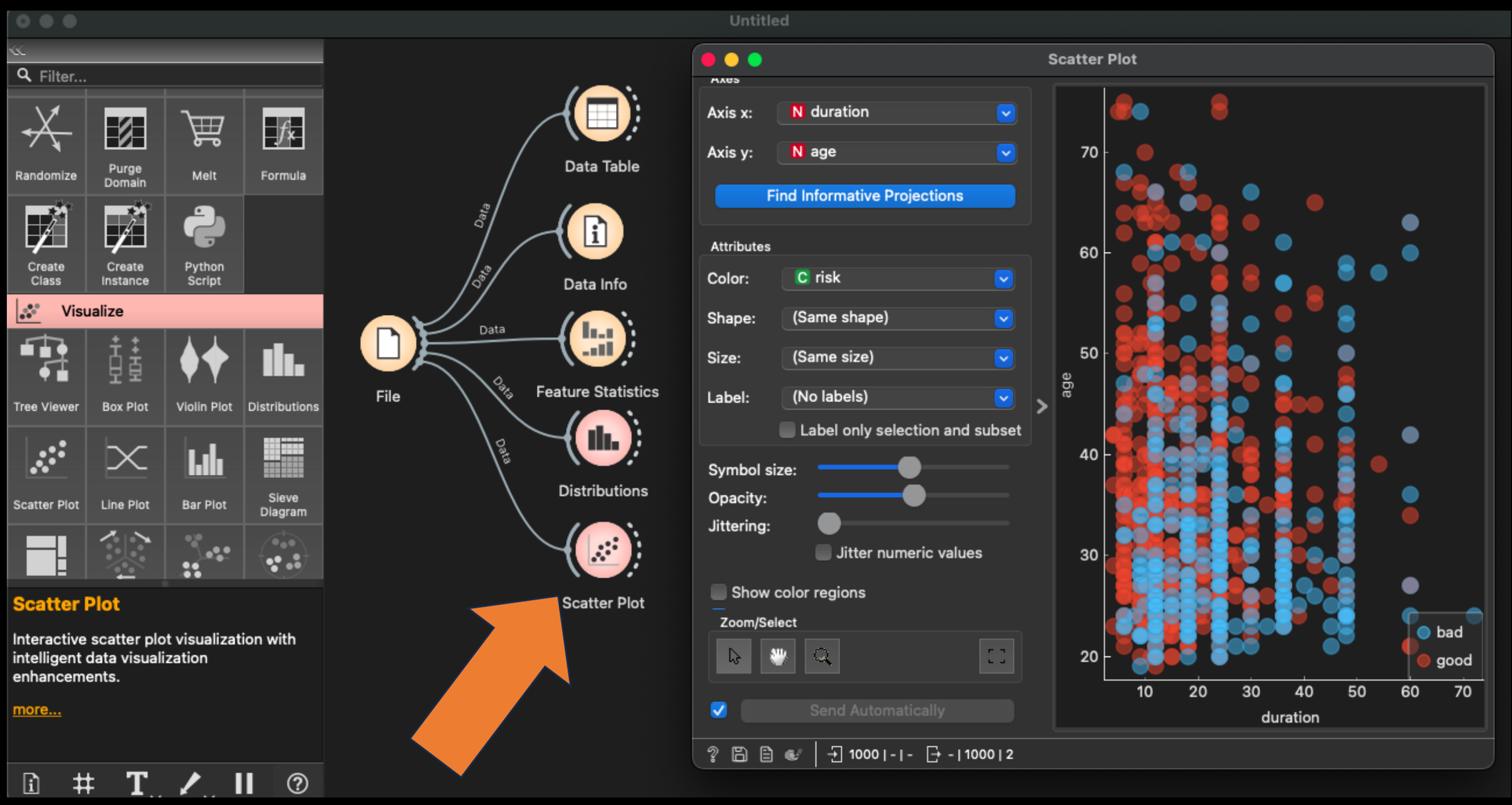


Distributions

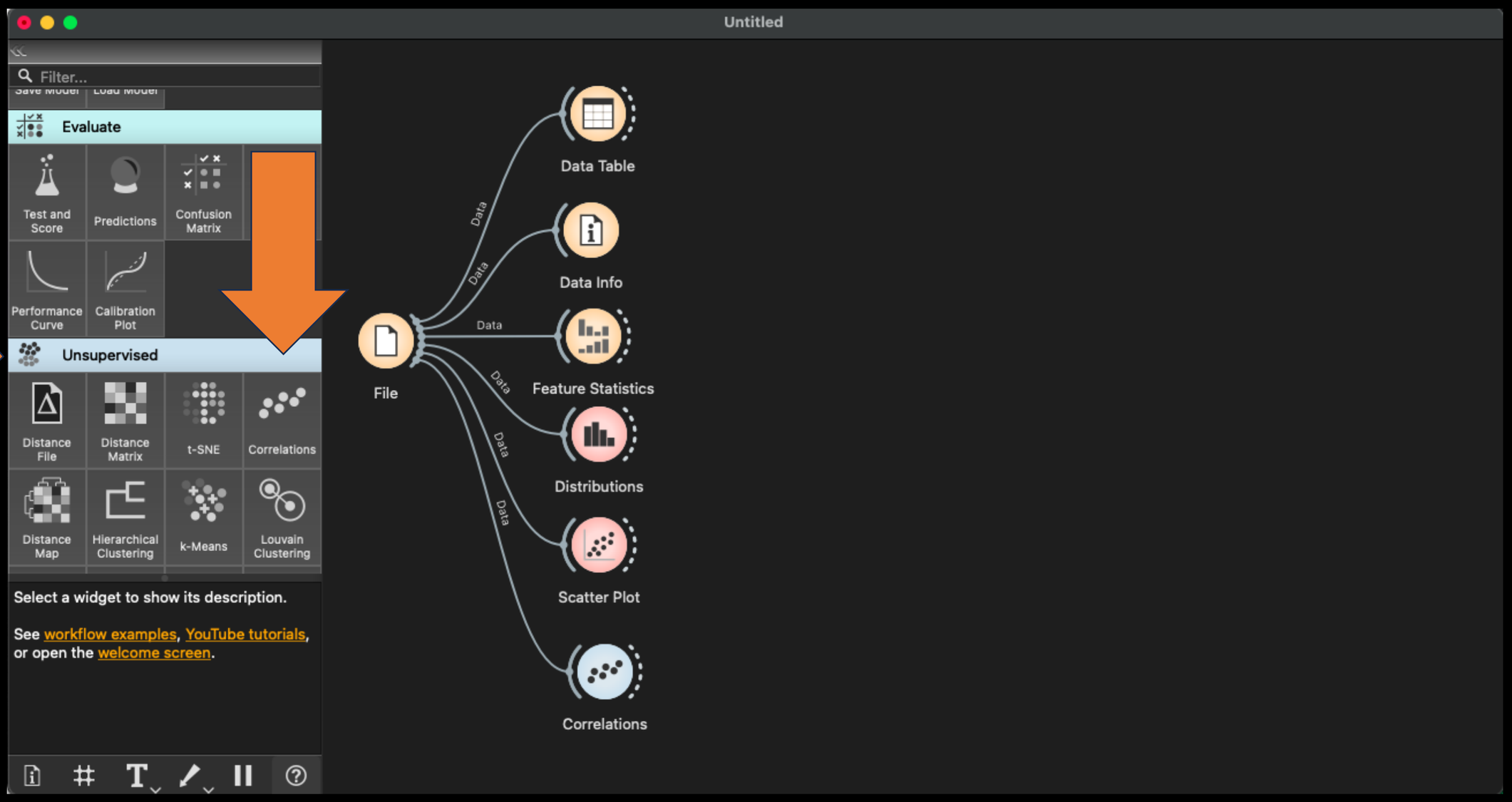


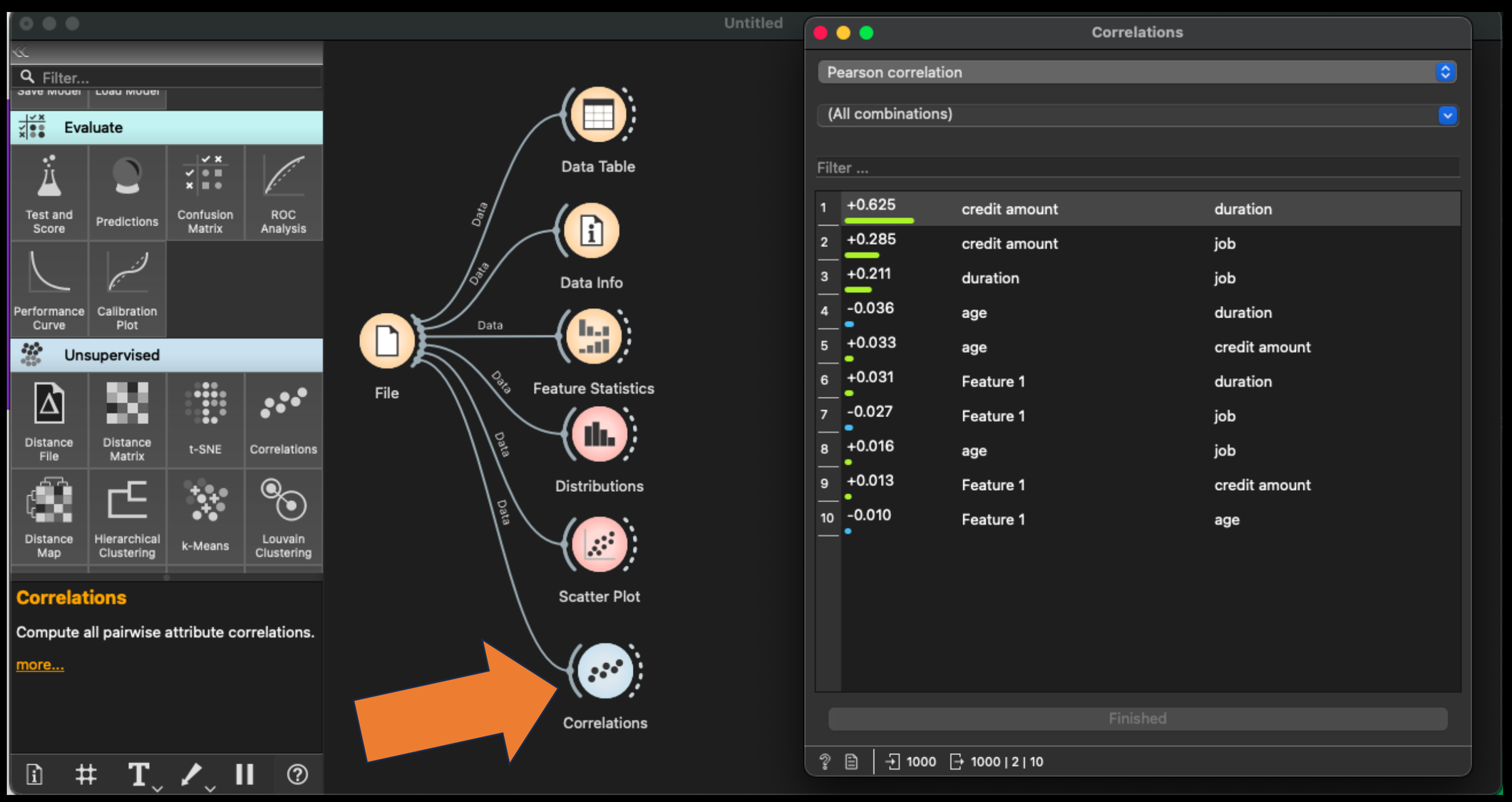
Scatter Plot





Correlations





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

💻 GitHub: https://github.com/erlinares/365_AI_Journey/

💬 Discord: https://discord.gg/5fFM2zh8

Edgar Rios Linares