COM5940: NEW MEDIA BUSINESS MODEL & INNOVATION
REGRESSION AND CLASSIFICATION MODELS IN SUPERVISED LEARNING

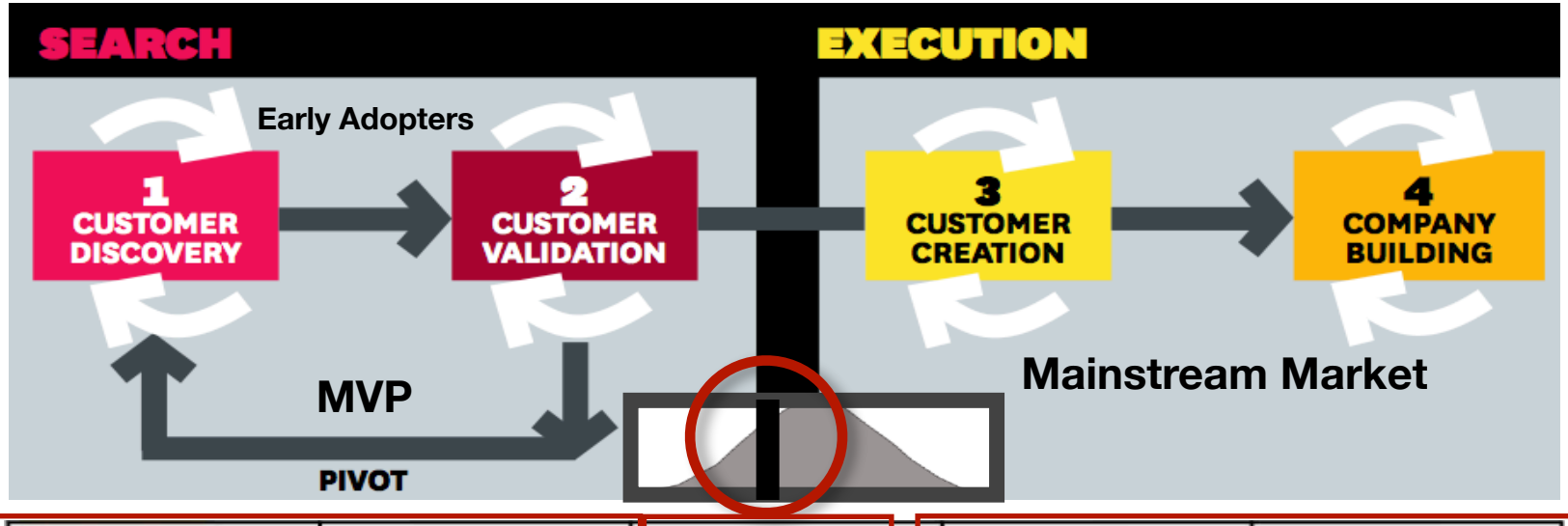
CUHK BUSINESS SCHOOL

Center for Entrepreneurship

Bernard Suen
Center for Entrepreneurship
Chinese University of Hong Kong

General Comments on 2nd Progress Report

Differences Between Business Model Canvas and Lean Canvas



PROBLEM | SOLUTION | UNIQUE VALUE PROPOSITION | UNFAIR ADVANTAGE | CUSTOMER SEGMENTS

KEY METRICS | CHANNELS

How

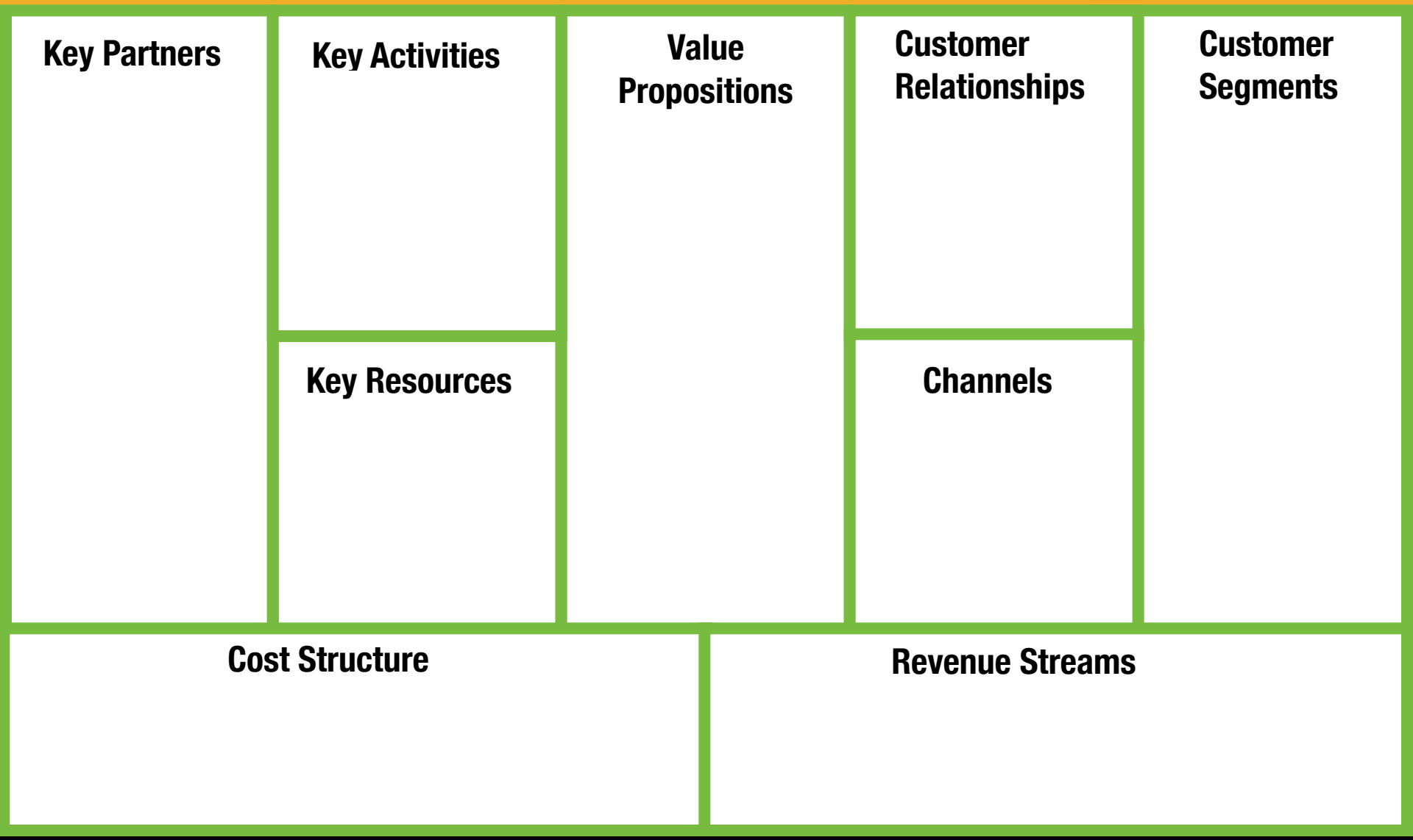
What/
Why

Why

Who

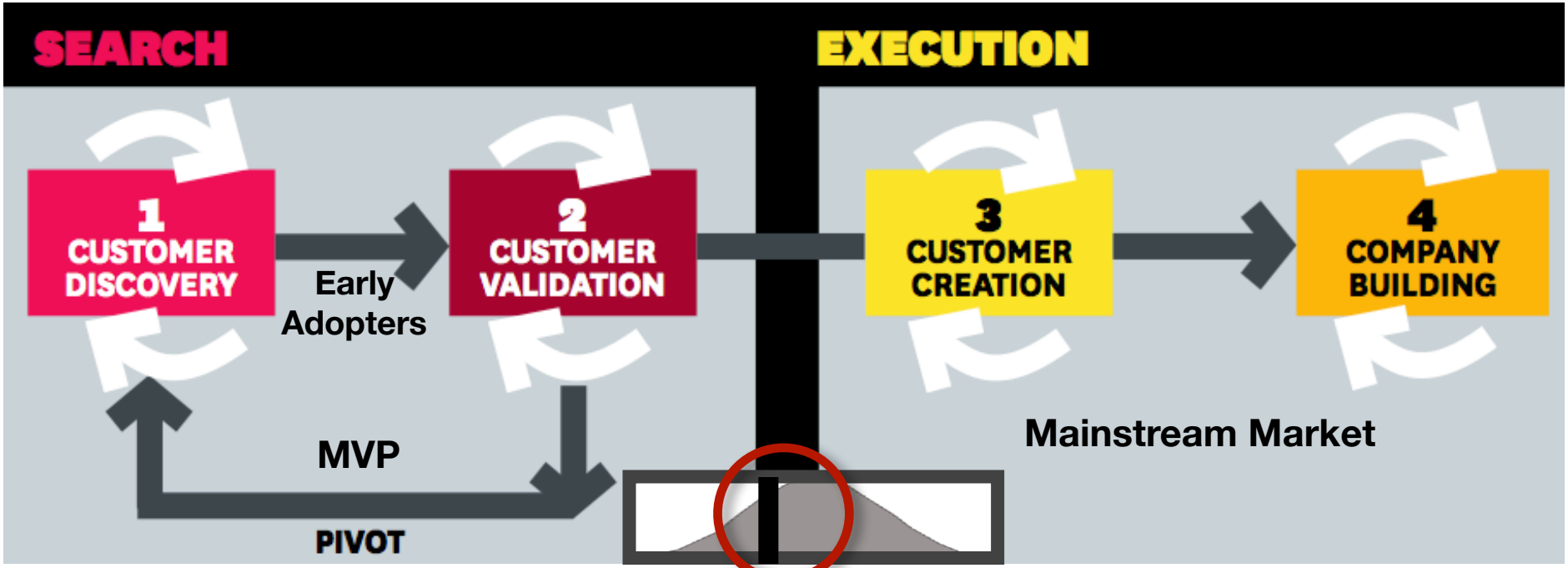
COST STRUCTURE | REVENUE STREAMS

How much


Key Partners
Key Activities
Key Resources
Value Propositions
Customer Relationships
Channels
Customer Segments
Cost Structure
Revenue Streams

Problem	Solution	Value Propositions	Unfair Advantage	Customer Segments
	Key Metrics		Channels	
Cost Structure			Revenue Streams	

“Why” the differences?



Problem	Solution	Value Propositions	Unfair Advantage	Customer Segments
	Key Metrics		Channels	
Cost Structure			Revenue Streams	

Key Partners	Key Activities	Value Propositions	Customer Relationships	Customer Segments
	Key Resources		Channels	
Cost Structure			Revenue Streams	

1. Some teams did not address the 2A3R concerns.
2. Problem definition not specific enough.
3. Data and content acquisition strategy has not addressed the key issues (e.g. domain knowledge & expertise)
4. Incentives for KOL, brand and other potential business partners to work with you are not adequately addressed.
5. Only one team mentioned data scraping to address the content shortage problem.
6. Not addressing where unfair advantage can be exploited (except one team).

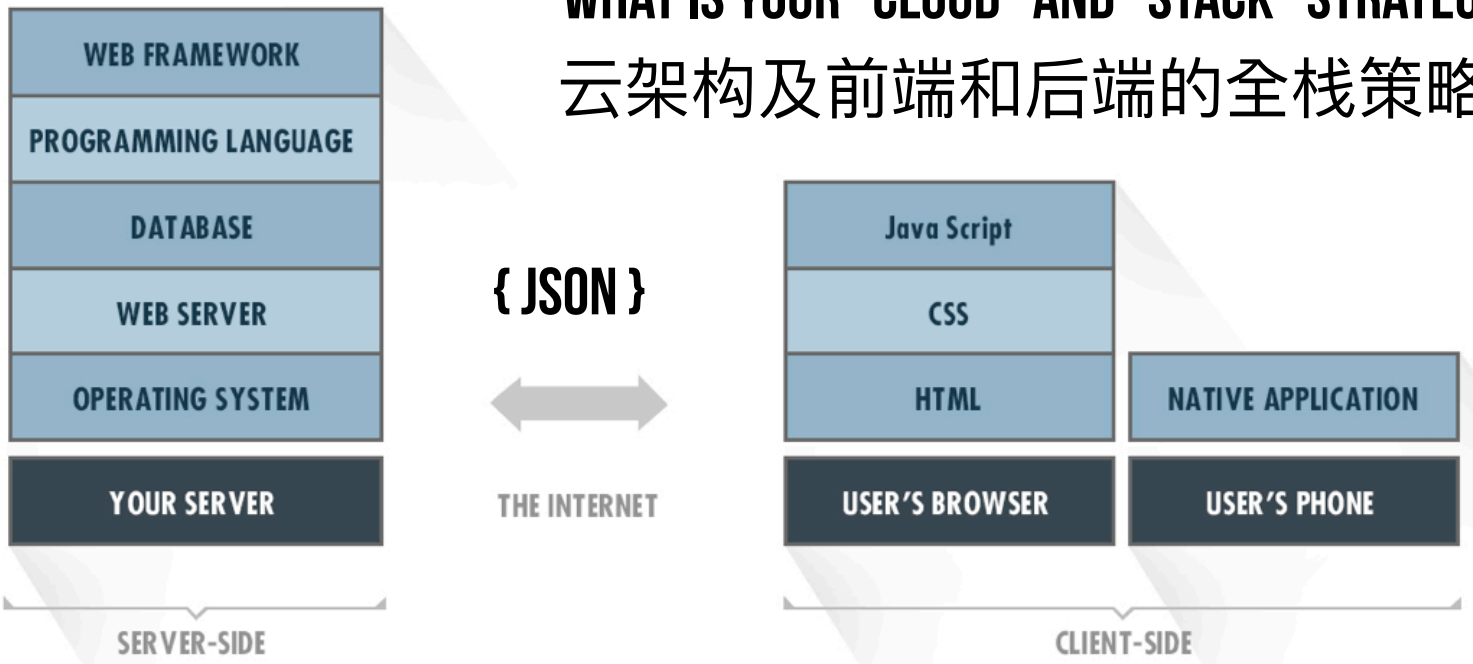
What has brought us here?

為甚么我们的学习会来到这一步!

API (e.g. REST)

WHAT IS YOUR "CLOUD" AND "STACK" STRATEGY?

云架构及前端和后端的全栈策略



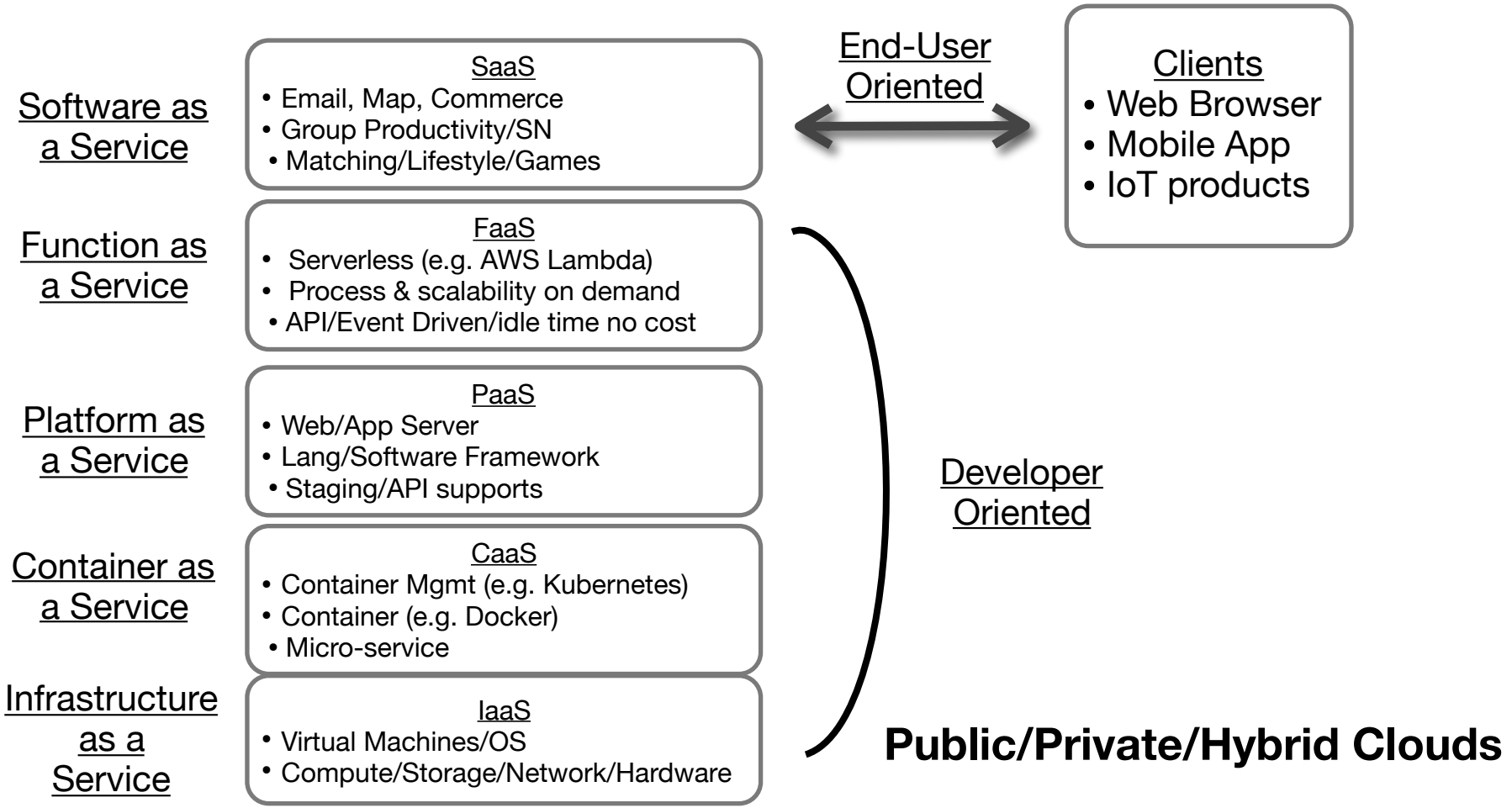
BACK-END TECHNOLOGY

后台

FRONT-END TECHNOLOGY

前台

云计算服务架构重组


Software as a Service
SaaS
• Email, Map, Commerce
• Group Productivity/SN
• Matching/Lifestyle/Games
End-User Oriented
Clients
• Web Browser
• Mobile App
• IoT products
Function as a Service
FaaS
• Serverless (e.g. AWS Lambda)
• Process & scalability on demand
• API/Event Driven/idle time no cost
Platform as a Service
PaaS
• Web/App Server
• Lang/Software Framework
• Staging/API supports
Developer Oriented
Container as a Service
CaaS
• Container Mgmt (e.g. Kubernetes)
• Container (e.g. Docker)
• Micro-service
Infrastructure as a Service
IaaS
• Virtual Machines/OS
• Compute/Storage/Network/Hardware
Public/Private/Hybrid Clouds

Unsupervised Learning

Supervised Learning

Reinforcement Learning

What has brought us here?

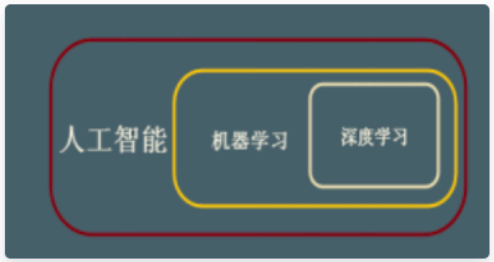
為甚么我们的学习会来到这一步!

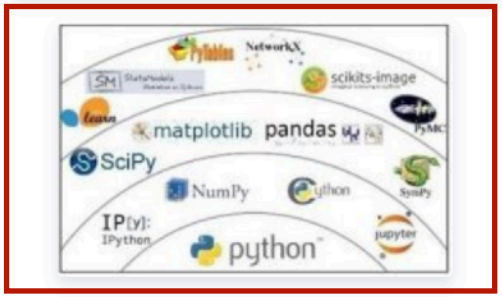
顶级程序员
319 篇文章 >

Python与人工智能的关系原来是这样的...

源 / AI时间　文 / 数据挖掘机

人工智能掀起了世界的新一波科技浪潮，如今，你要是不懂点AI、机器学习和python都不好意思说你是现代人，那么python究竟和人工智能什么关系，为什么人工智能把python也给带火了，今天就给大家简单介绍下python和人工智能的关系及应用，以及想要学人工智能的你，究竟需要学些什么python的知识，先来上两张图人工智能和python的图。

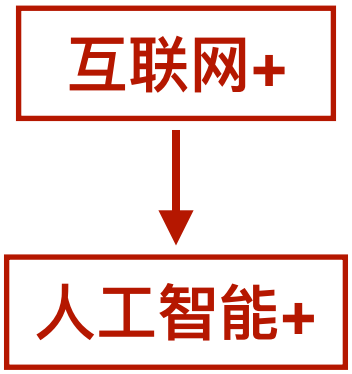




https://cloud.tencent.com/developer/article/1161025

从上图可以看出，人工智能包含常用机器学习和深度学习两个很重要的模块，而右图中python拥有matplotlib、Numpy、sklearn、keras等大量的库，像pandas、sklearn、matplotlib这些库都是做数据处理、数据分析、数据建模和绘图的库，基本上机器学习中对数据的爬取（scrapy）、对数据的处理和分析（pandas）、对数据的绘图（matplotlib）和对数据的建模（sklearn）在python中全都能找到对应的库来进行处理，所以，要想学习AI而不懂python，那就相当于想学英语而不认识单词，所以，python学起来吧。

PLATFORM (平台)



DEVELOPMENT
开发平台

DEPLOYMENT
发放营运平台

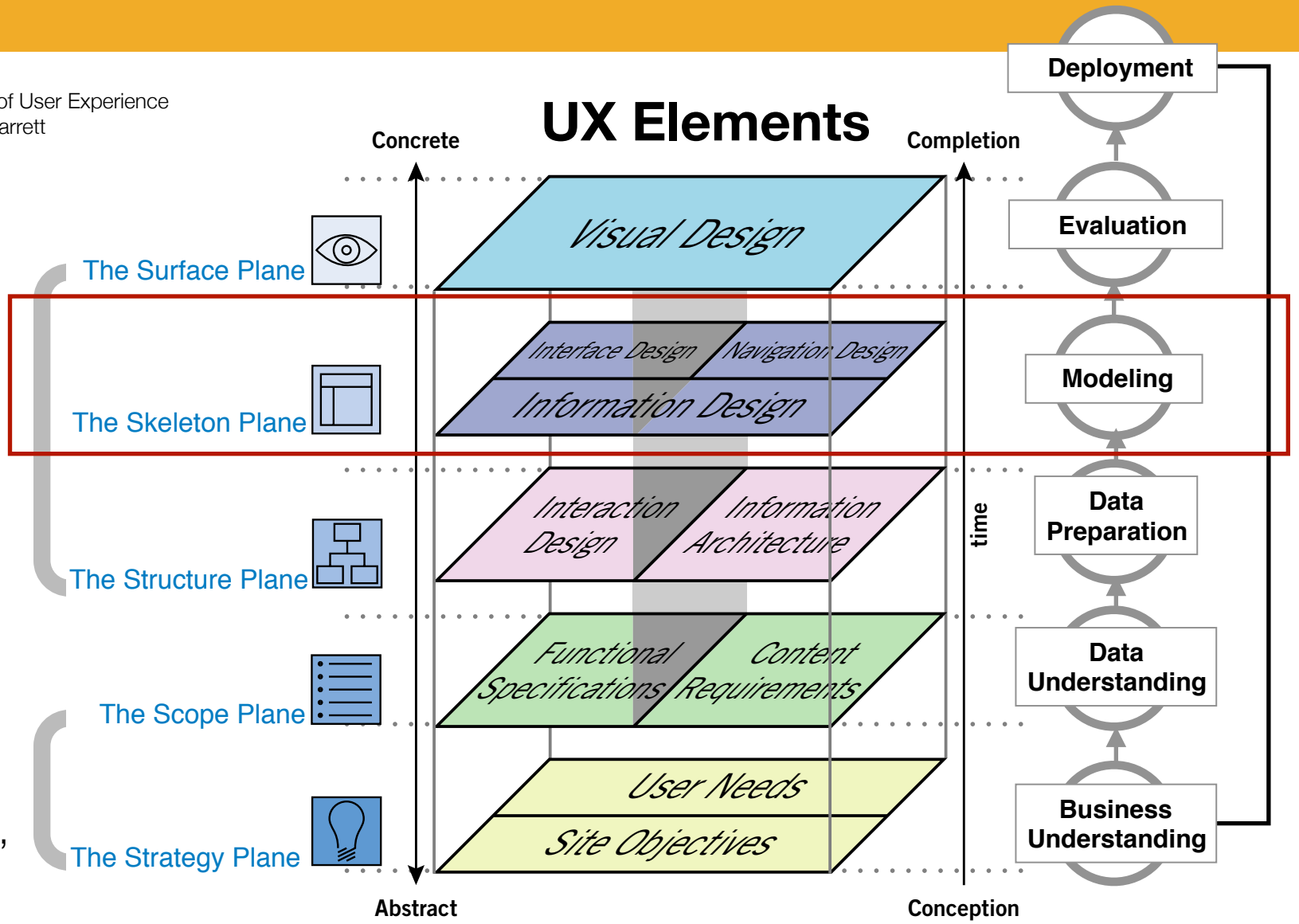
兼容和整合考虑

Source: By Hugo Bowen-Anderson August 15, 2018 Harvard Business Review

What data scientists do. We now know how data science works, at least in the tech industry. First, data scientists lay a solid data foundation in order to perform robust analytics. Then they use online experiments, among other methods, to achieve sustainable growth. Finally, they build machine learning pipelines and personalized data products to better understand their business and customers and to make better decisions. In other words, in tech, data science is about infrastructure, testing, machine learning for decision making, and data products.数据变成产品和运营的一个重要构成部分

How does Machine Learning fit into the Problem Solving Framework (CRISP-DM)?

Source: Elements of User Experience
by Jesse James Garrett



知乎

https://zhuanlan.zhihu.com/p/42418356

写文章

生产环境中进行机器学习模型部署（using Flask）

数据蛙datafrog
数据挖掘农民工

使用Flask REST API进行Python机器学习预测

2020-04-07 00:34:45　　分类：Python / Python Flask / 机器学习　　阅读(11)　　评论(0)

本文概述

http://www.srcmini.com/42615.html

AI科技大本营
2.5K 篇文章 >

为什么说爱奇艺是一家技术公司？

5

0

分享

AI无处不在

在近日举行的爱奇艺世界大会上，爱奇艺展示了AI技术和落地应用。爱奇艺首席技术官兼基础架构和智能内容分发事业群总裁刘文峰介绍称，在爱奇艺的整个运营流程中，AI发挥的作用无孔不入，贯穿视频内容的创作、生产、理解、分发、播放到变现以及客服整个流程中。

首先以智能生产环节为例。爱奇艺的内容主要来自三个部分，一部分是传统的采买，另外一部分是自制剧，以及爱奇艺号等原创用户产生的一些内容。对于自制部分，爱奇艺采用AI打造了一个爱创媒资系统，这个系统主要是针对导演、制作人的创作痛点。内容创作者在后期制作的过程中往往会被一些非原创性的工作所影响，以至于没有办法集中精力，而且拍摄素材繁多，导致1～2个小时的综艺后期制作的时间往往要花上8～10天才能最终看到成片。通过使用身份识别、表情识别、镜头识别等AI技术，类似工作可缩短至数分钟，剪辑师可以很方便地从繁杂的素材中挑选精彩片段，效率提高20%–30%。

比如，导演想要成百上千个小时的素材里找到吴亦凡生气的特写，AI可能只需要三步就能解决：第一是身份识别，找出吴亦凡；第二是表情识别，识别出生气的表情；第三是镜头识别，识别特写镜头。这样就能很快把导演想要的片段找出来。

https://cloud.tencent.com/developer/article/1430210

号外！号外！Python纳入高考内容了！人工智能时代就要来临了！ https://cloud.tencent.com/developer/article/1024699

2018-01-19 阅读 445

就在前几天，和一位高校的信息技术老师聊天，我得到了一个震惊的消息：明年，浙江省信息技术教材将不会在使用晦涩难懂的VB语言，而是改学更简单易懂的Python语言。也就是说：

Python语言将纳入高考内容之一。

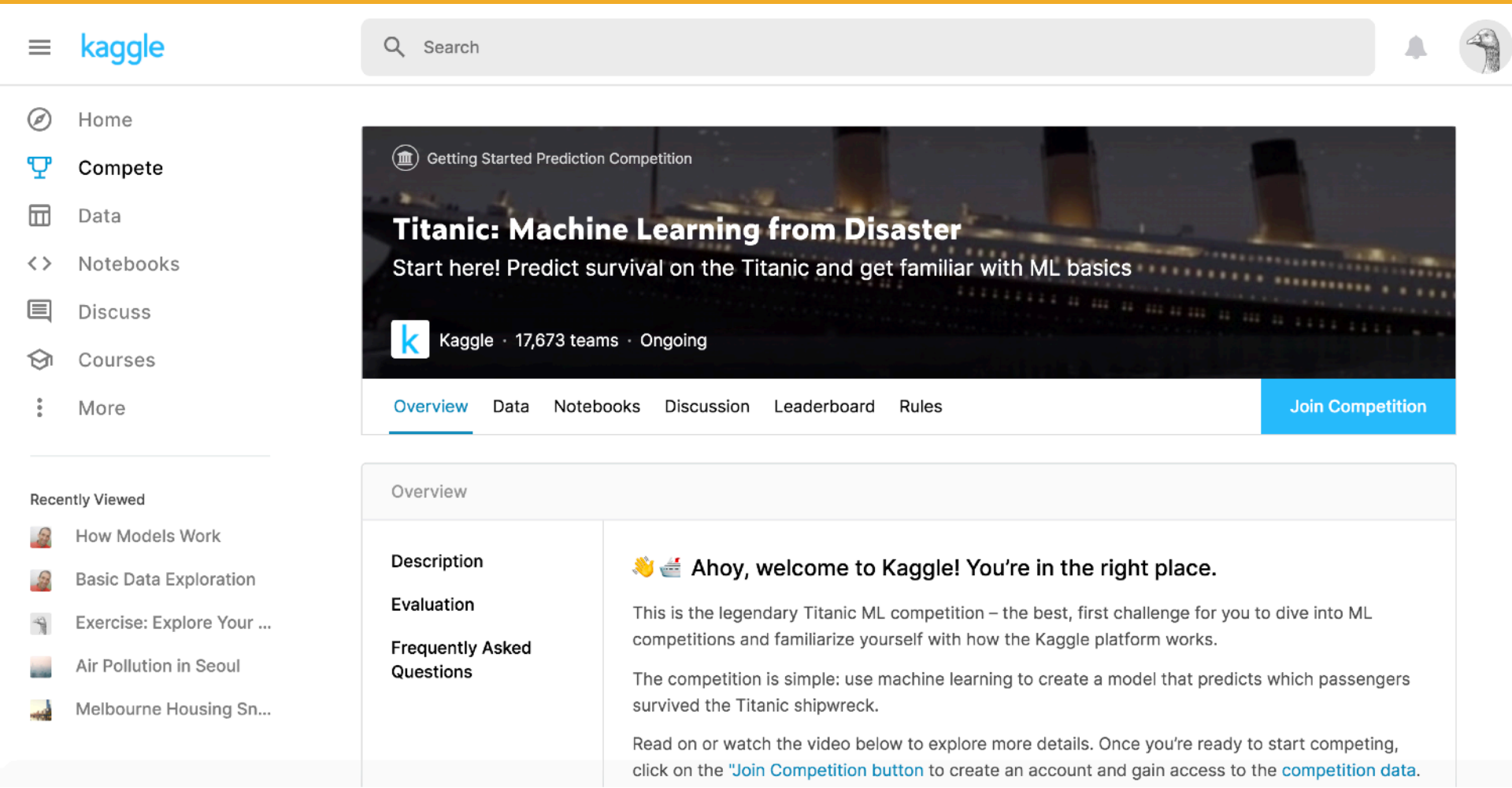
编程语言在升学中的比重逐渐加大，将要成为高考加分的一大利器。

Python将被纳入高考内容！

浙江省信息技术课程改革方案已经出台，Python确定进入浙江省信息技术高考，从2018年起浙江省信息技术教材编程语言将会从vb更换为Python。

其实不止浙江，教育大省北京和山东也确定要把Python编程基础纳入信息技术课程和高考的内容体系，Python语言课程化也将成为孩子学习的一种趋势。


kaggle
How to Get Started with Kaggle's Titanic Machine Learning Competition

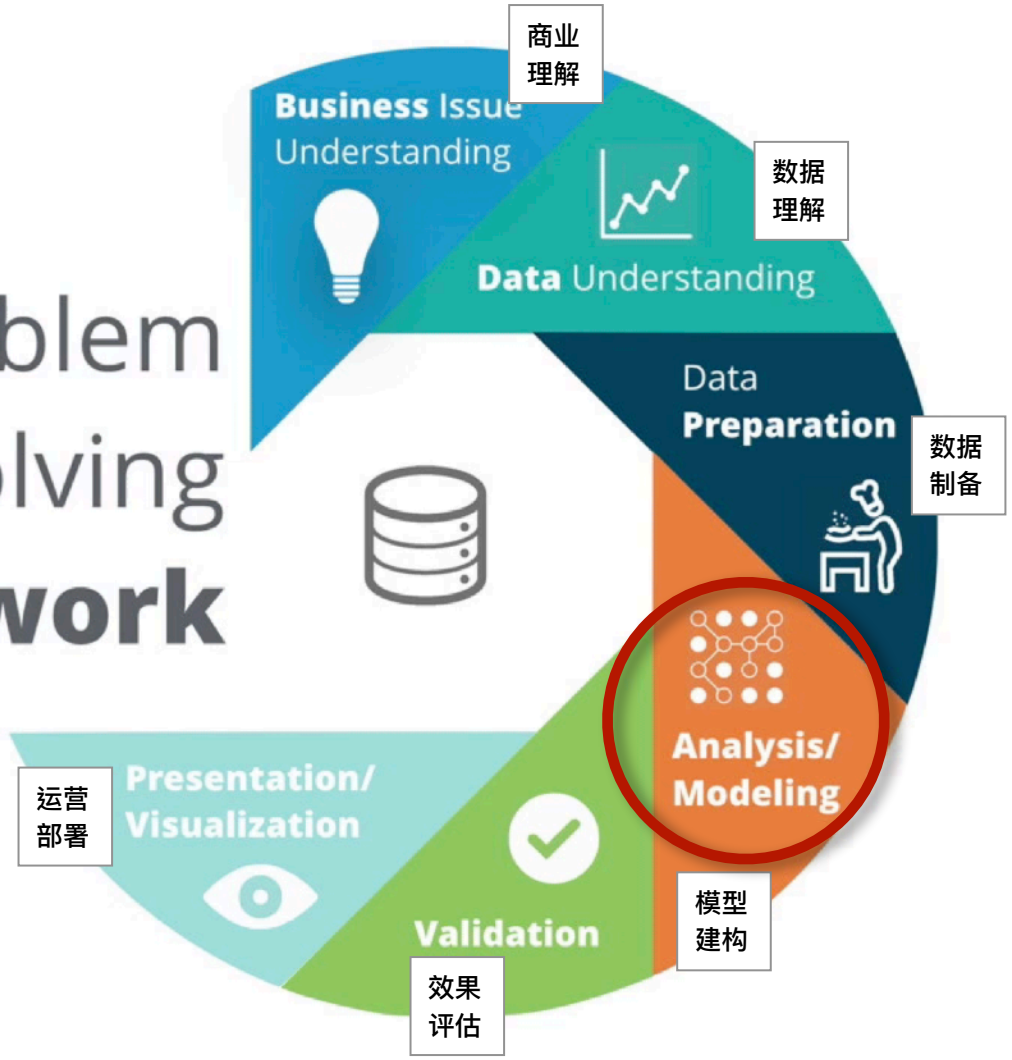


https://www.kaggle.com/c/titanic/

为什么Kaggle对找工作有帮助？如何入门？

猴子
中国科学院大学 电子与通信工程硕士

Problem Solving **Framework**



The Methodology Map

<table>
<tr><th colspan="4">聚焦未来</th><th>聚焦过去和现在</th></tr>
<tr><td colspan="4">Predict Outcome</td><td>Data Analysis</td></tr>
<tr><td colspan="3">Data Rich</td><td>Data Poor</td><td>Geospatial</td></tr>
<tr><td colspan="2">Numeric</td><td colspan="2">Classification</td><td>A/B Test</td><td>Segmentation</td></tr>
<tr><td>Continuous</td><td>Count</td><td>Binary</td><td>Non Binary</td><td></td><td>Aggregation</td></tr>
<tr><td>• Linear Regression,
• Multiple Regression</td><td>• Count Regression</td><td>• Logistic Regression
• Decision Tree</td><td>• Random Forest
• K-Nearest Neighbour</td><td></td><td>Descriptive</td></tr>
<tr><td colspan="4">e.g.
Scikit-learn, Tensorflow, Orange</td><td>e.g.
Google Optimize</td><td>e.g.
Google Analytics</td></tr>
</table>

Source: Udacity Model Selection Methodology Map

Predictive Analytics/Machine Learning/Deep Learning
预测分析/机器学习/深度学习

- Linear Regression
- Decision Tree
- Forest Model
- Boosted Model

- Count Regression

- Logistic Regression
- Decision Tree

- Forest Model
- Booted Model

Inferential Statistics & Hypothe-sis Testing
推论统计与假设检验

Explorative Data Analysis & Descriptive Statistics探索性数据分析与描述性统计

Source: Udacity Model Selection Methodology Map

Today's Agenda

1. Labelling of numerical and categorical data for supporting Supervised Learning.
2. Practical use of Unsupervised Learning
3. Separation of dataset into training and testing sets before building the model.
4. Use of regression or classification algorithms in training the model.
5. Model selection, training, testing, evaluation and deployment (as REST services)
6. Demonstration of regression and classification models in Orange3, Scikit-learn and TensorFlow.

Patterns in Unsupervised Learning

Classification & Regression in Supervised Learning

Demo in Orange, Scikit-learn, & Tensor Flow

Featuring Engineering & Dimension Reduction

https://www.youtube.com/watch?v=N9fDIAflCMY

With the possibility of having irrelevant and redundant features in confounding the outcome, how can we deal with them?

② Reduce the dimensions of the construct

a) Principal Component Analysis
b) Exploratory Factor Analysis

https://www.youtube.com/watch?v=TuSV2o3OcwI&list=PL7SDML5LXeEp6so26ZdKNZ6Ycvc55WnoG

Unsupervised Learning is used in **feature engineering** and **dimension reduction** by leveraging its power in **patter recognition without labelling**.

Features/Dimensions/Attributes/Independent Variables

Observations/Instances

	年龄	性别	聘有家佣	家庭人数	居住状况	教育程度	职业	每週买菜次数
1								
2								
3								
4								
5								
6								
7								

Features/Dimensions/Attributes/Independent Variables

Observations/Instances

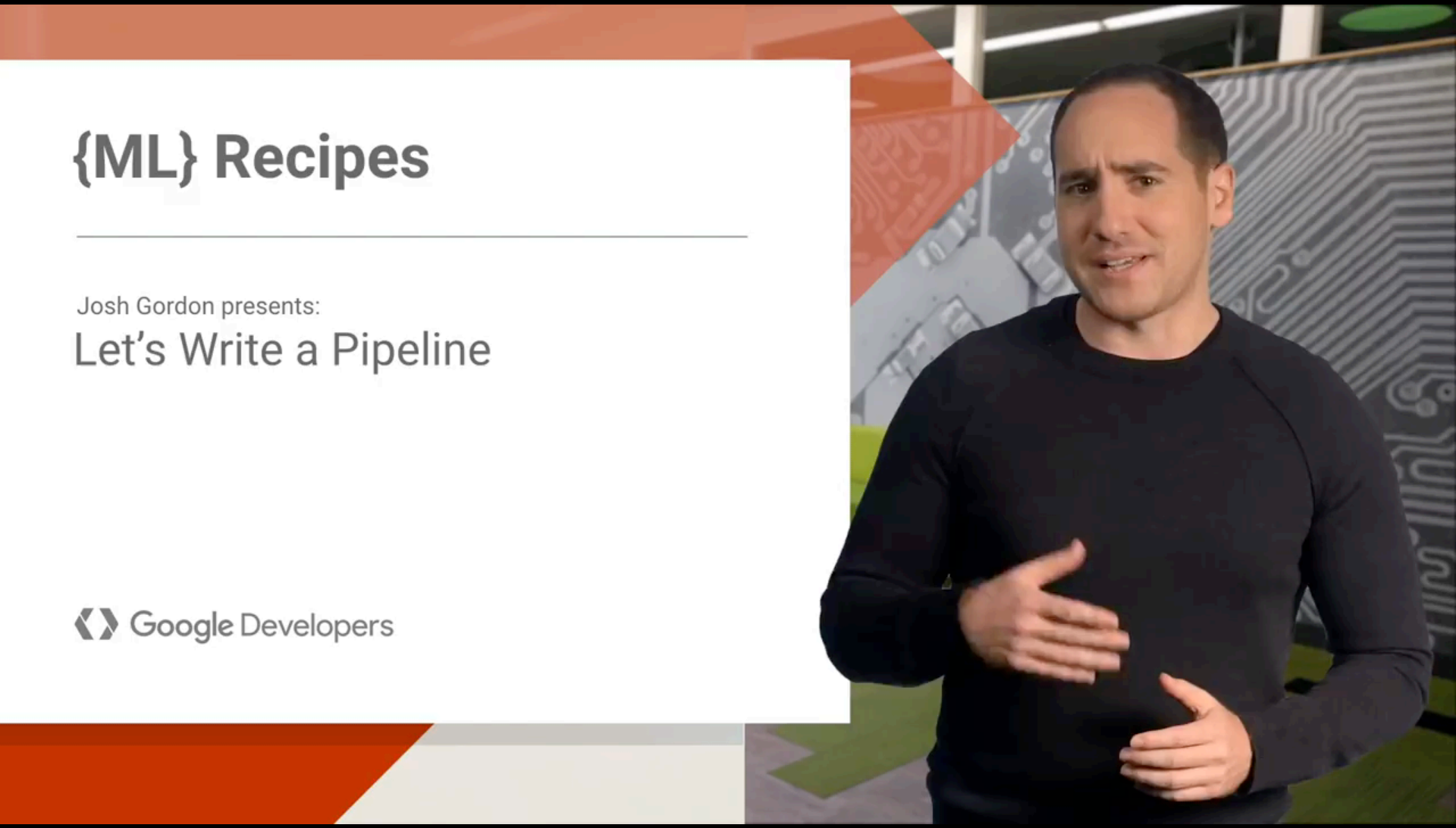
	年龄	性别	聘有家佣	家庭人数	居住状况	教育程度	职业	每週买菜次数	N+1	N+…
1										
2										
3										
4										
5										
6										
7										

Can anyone of the features be removed to focus on the key ones?

Patterns in Unsupervised Learning

Classification & Regression in Supervised Learning

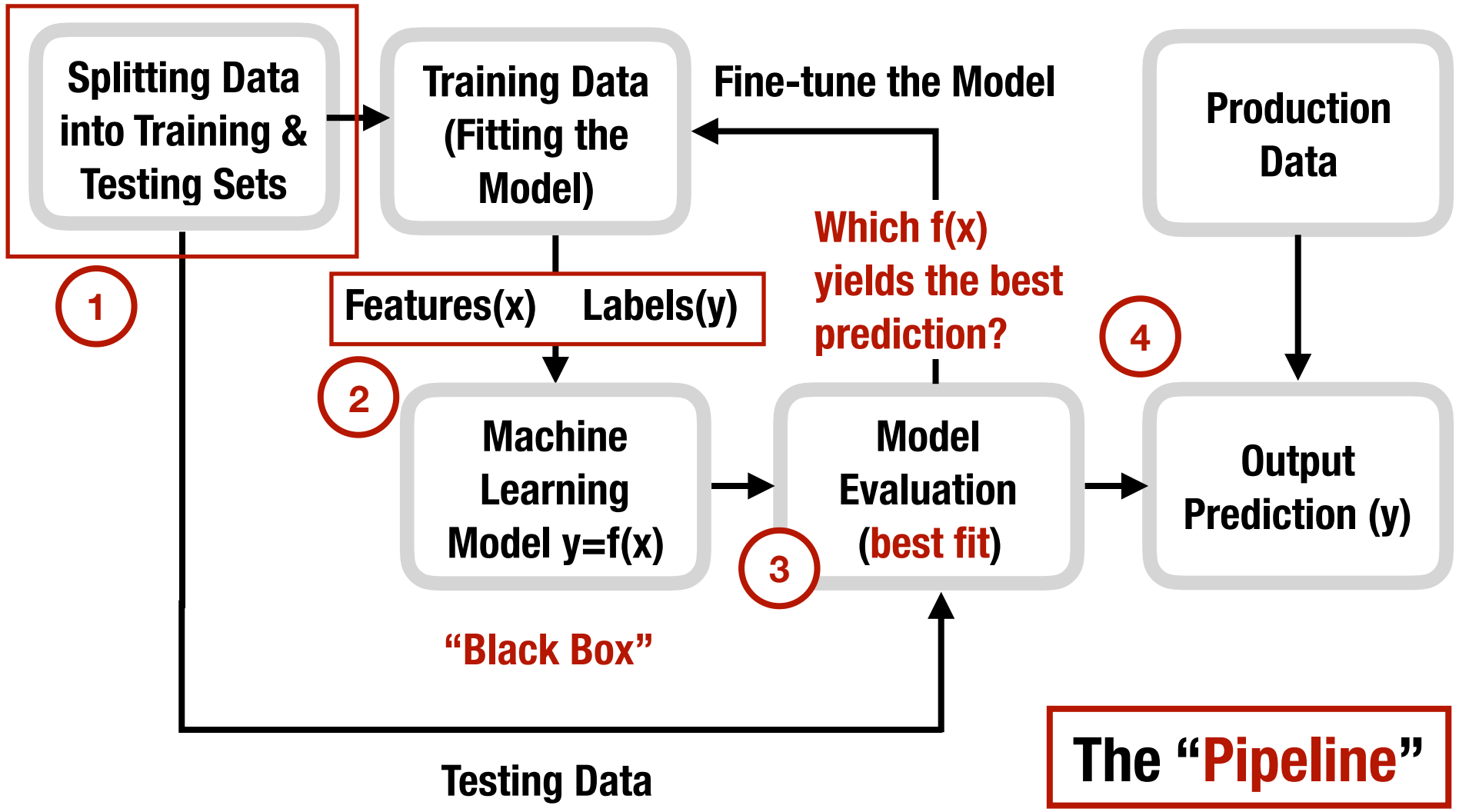
Demo in Orange, Scikit-learn, & Tensor Flow



https://www.youtube.com/watch?v=84gqSbLcBFE

"A **machine learning pipeline** is used to help **automate** **machine learning** **workflows**. They operate by enabling a sequence of data to be transformed and correlated together in a model that can be tested and evaluated to achieve an outcome, whether positive or negative."

Source: What is a Pipeline in Machine Learning? How to create one? by Shashanka M


Splitting Data into Training & Testing Sets
Training Data (Fitting the Model)
Fine-tune the Model
Production Data
1
Features(x) Labels(y)
Which f(x) yields the best prediction?
4
2
Machine Learning Model y=f(x)
Model Evaluation (best fit)
Output Prediction (y)
3
"Black Box"
Testing Data
The "Pipeline"

Why Splitting the Data into Training and Testing Sets in Prediction?

K-fold Cross Validation

https://www.youtube.com/watch?v=VcSNDMBTE-s

一个比喻

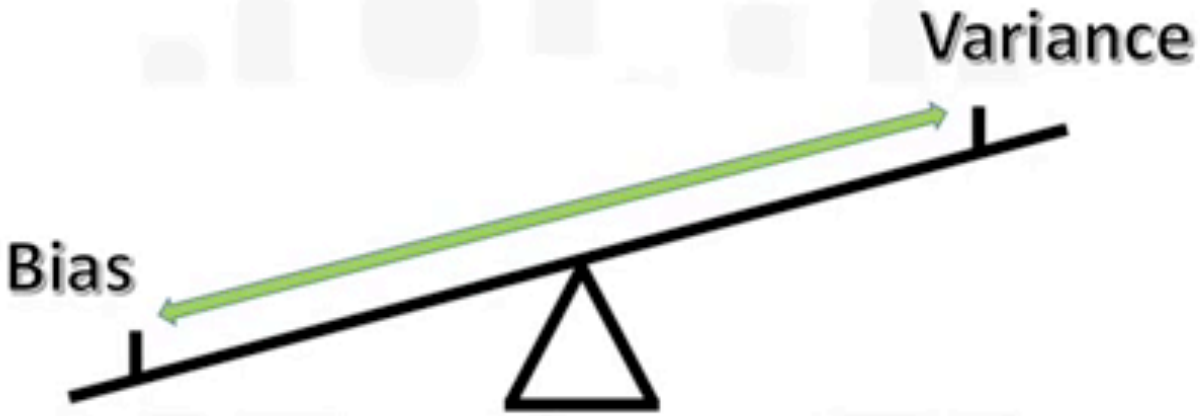
以考试成绩来预测学生表现

以旧试题来训练学生 ⟶ 重用旧试题来做评估

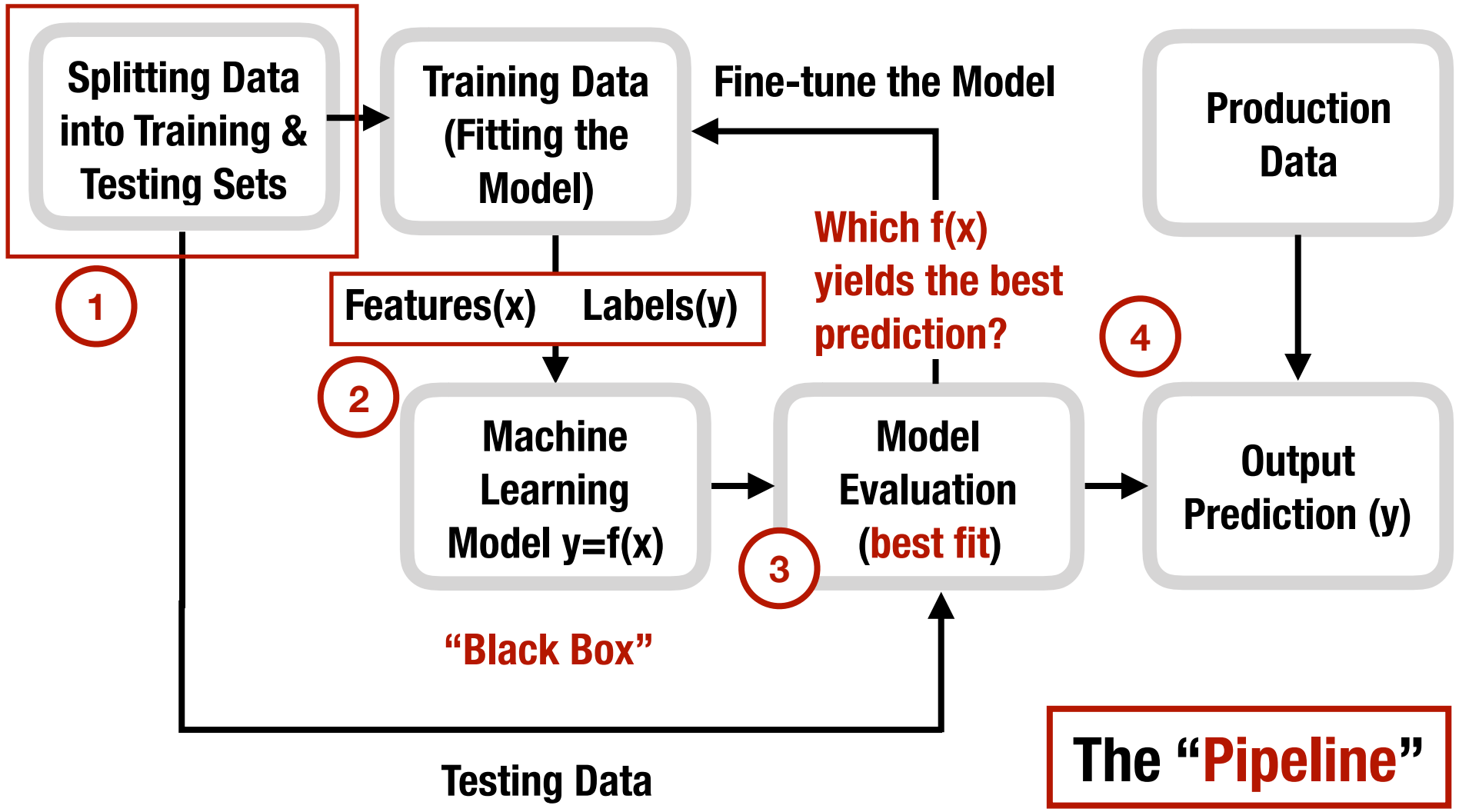
结果出来成绩优异能代表甚么?

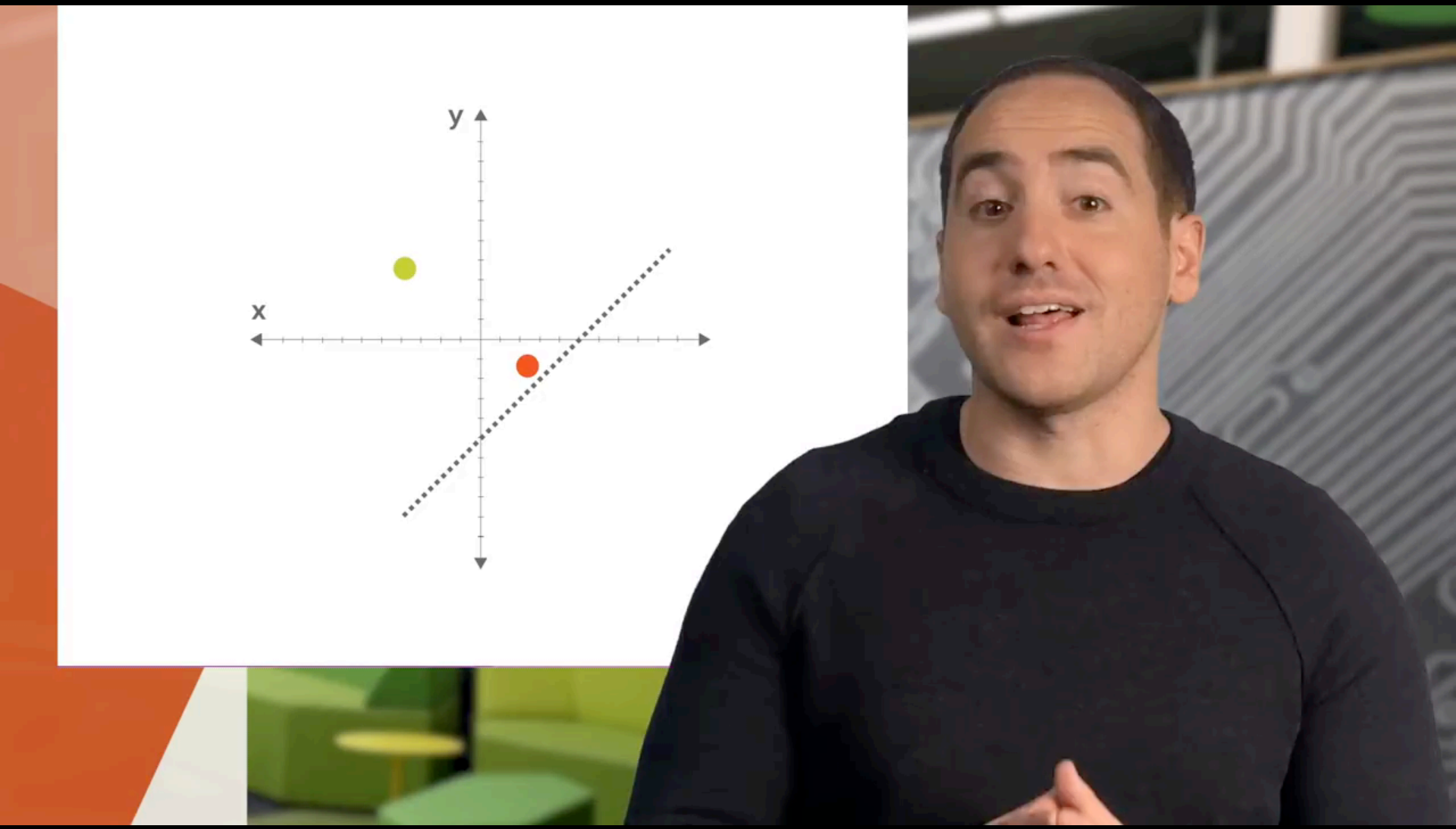
够竟我们要培训会考试的学生或是会思考的学生? 同样道理也适用於以Supervised Learning来训练ML模型.

Bias Variance Trade-Off



https://www.youtube.com/watch?v=hI8YJDpCxf0


Splitting Data into Training & Testing Sets
Training Data (Fitting the Model)
Fine-tune the Model
Production Data
1
Features(x) Labels(y)
Which f(x) yields the best prediction?
4
2
Machine Learning Model y=f(x)
Model Evaluation (best fit)
Output Prediction (y)
3
"Black Box"
Testing Data
The "Pipeline"

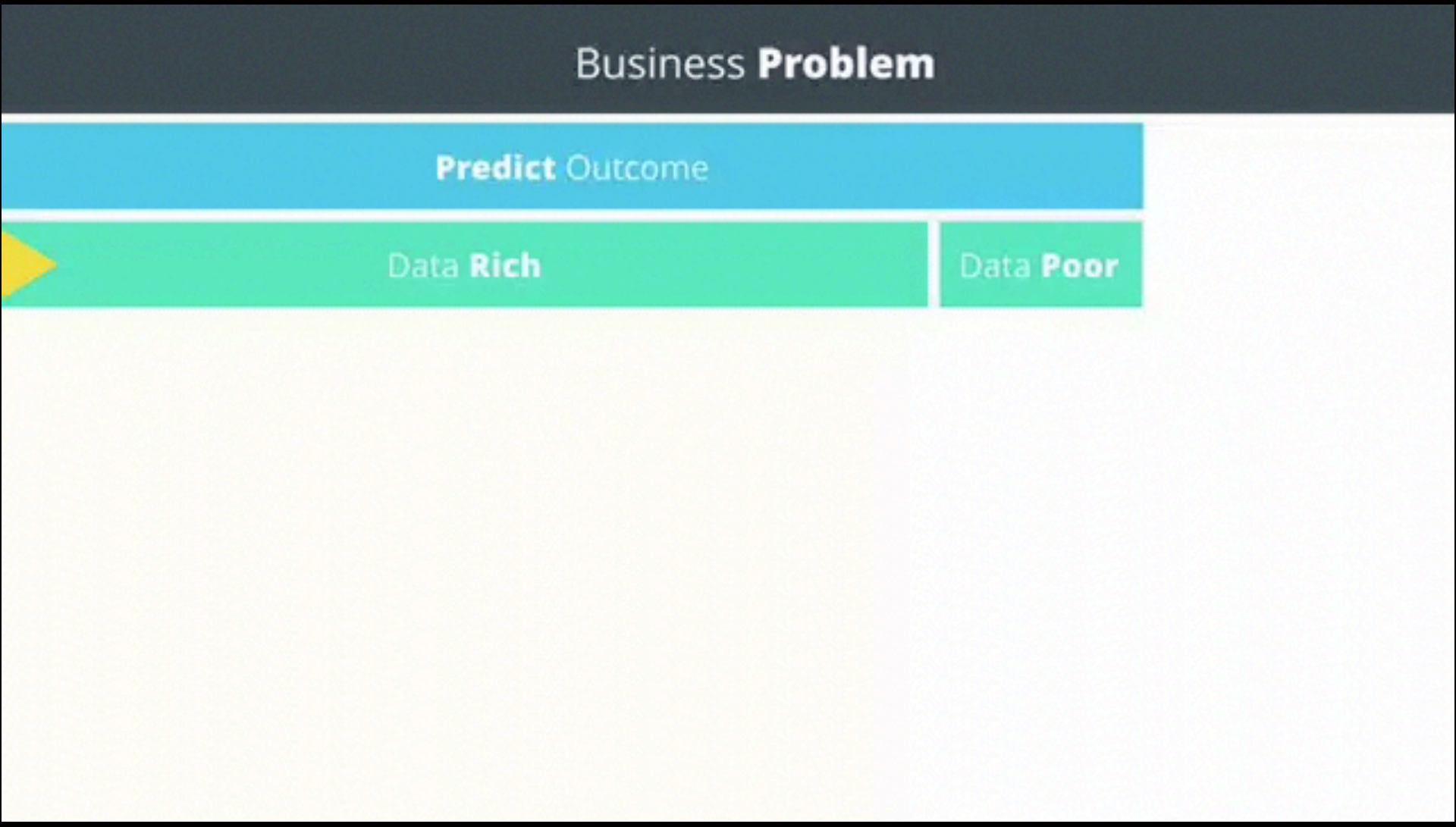


https://www.youtube.com/watch?v=84gqSbLcBFE

Which Model to Choose?

<table>
<tr><th colspan="5">聚焦未来</th><th>聚焦过去和现在</th></tr>
<tr><td colspan="5">Predict Outcome</td><td>Data Analysis</td></tr>
<tr><td colspan="4">Data Rich</td><td>Data Poor</td><td>Geospatial</td></tr>
<tr><td colspan="2">Numeric</td><td colspan="2">Classification</td><td>A/B Test</td><td>Segmentation</td></tr>
<tr><td>Continuous</td><td>Count</td><td>Binary</td><td>Non Binary</td><td></td><td>Aggregation</td></tr>
<tr><td>• Linear Regression,
• Multiple Regression</td><td>• Count Regression</td><td>• Logistic Regression
• Decision Tree</td><td>• Random Forest
• K-Nearest Neighbour</td><td></td><td>Descriptive</td></tr>
<tr><td colspan="4">e.g.
Scikit-learn, Tensorflow, Orange</td><td>e.g.
Google Optimize</td><td>e.g.
Google Analytics</td></tr>
</table>

Source: Udacity Model Selection Methodology Map



https://www.youtube.com/watch?v=LkpkL7KmGLA

https://www.youtube.com/watch?v=cAIPn6YRPc0

Let's learn a few basic ML algorithms (models)

K-Nearest Neighbour (Classification Model)

If K=11 and the new cell is between two (or more) categories, we simply pick the category that "gets the most votes".

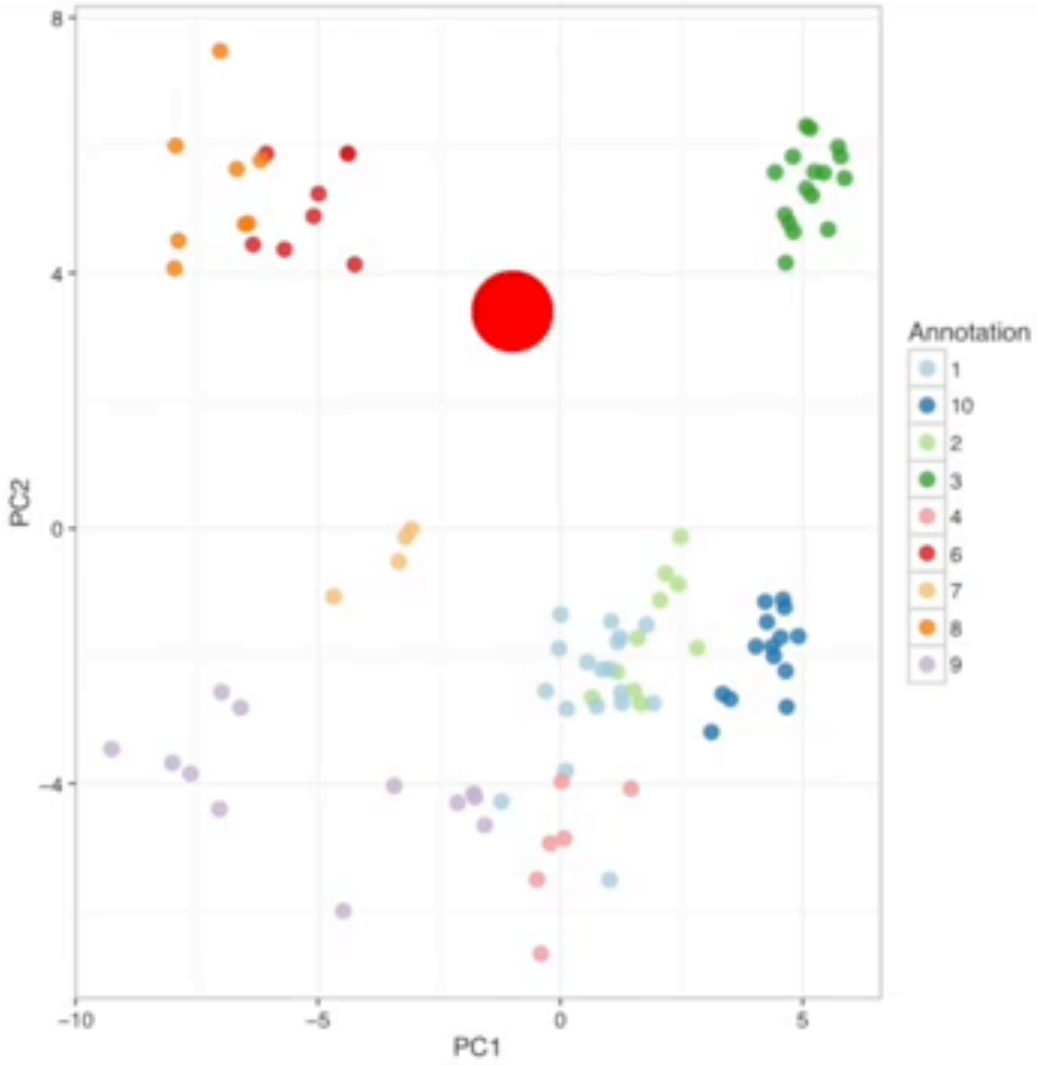
In this case....

7 nearest neighbors are **RED**.

3 nearest neighbors are **ORANGE**.

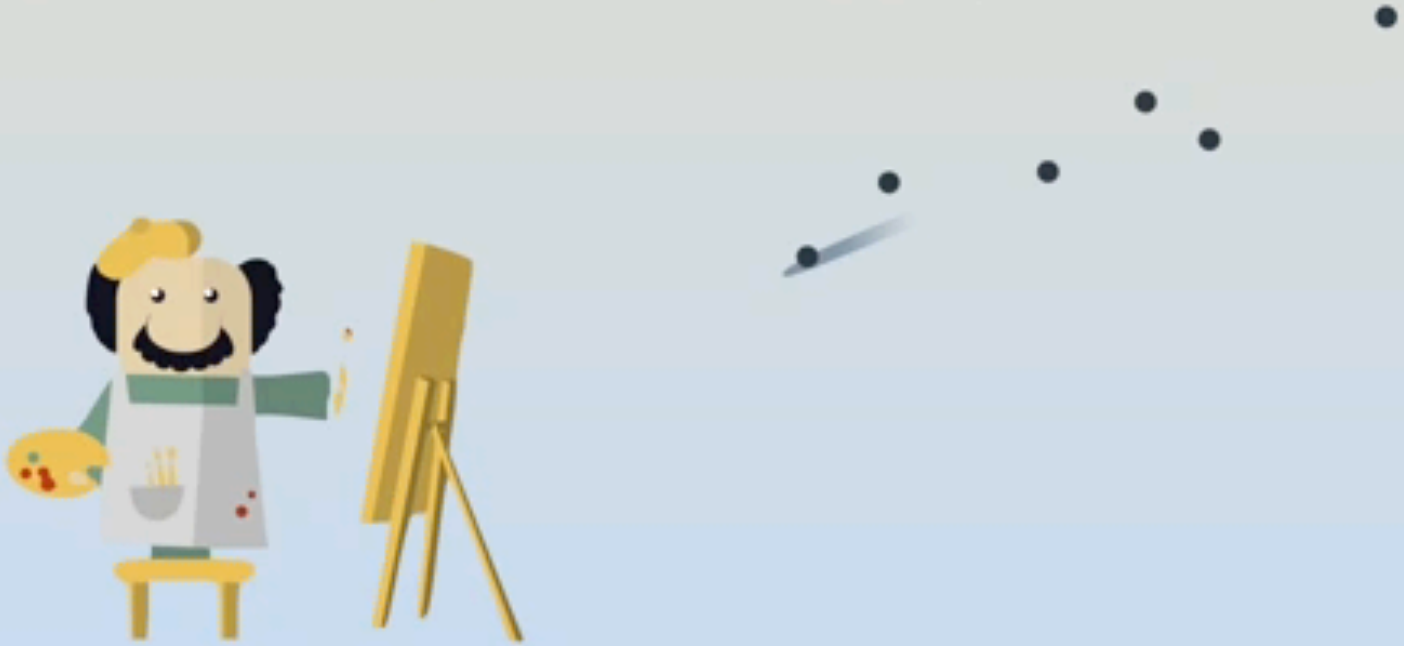
1 nearest neighbor is **GREEN**.

Since **RED** got the most votes, the final assignment is **RED**.



Linear Regression (Regression Model)

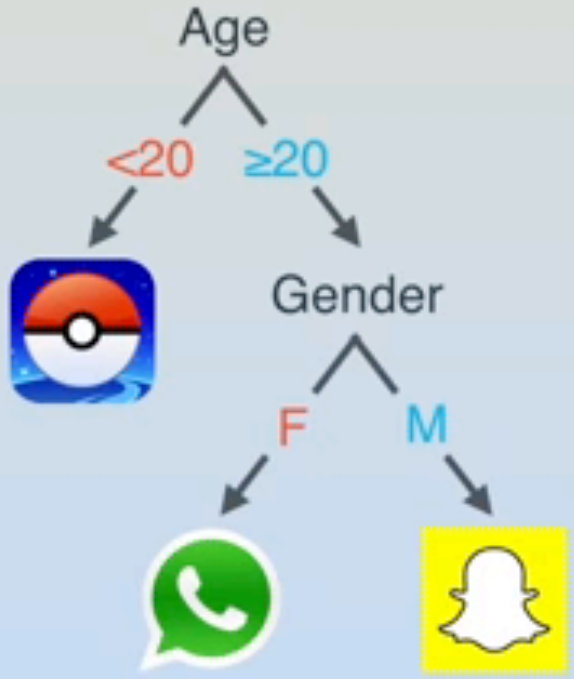
Linear and Polynomial Regression



Decision Tree (Classification Model)

Decision Tree

Gender	Age	App
F	15	
F	25	
M	32	
F	40	
M	12	
M	14	



Examples in Orange, Scikit-learn, & Tensor Flow

Patterns in Unsupervised Learning

Classification & Regression in Supervised Learning

Demo in Orange, Scikit-learn, & Tensor Flow

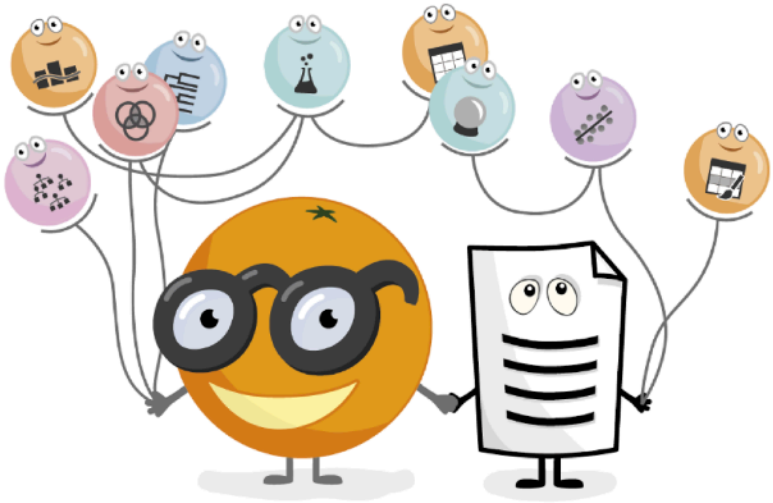
orange

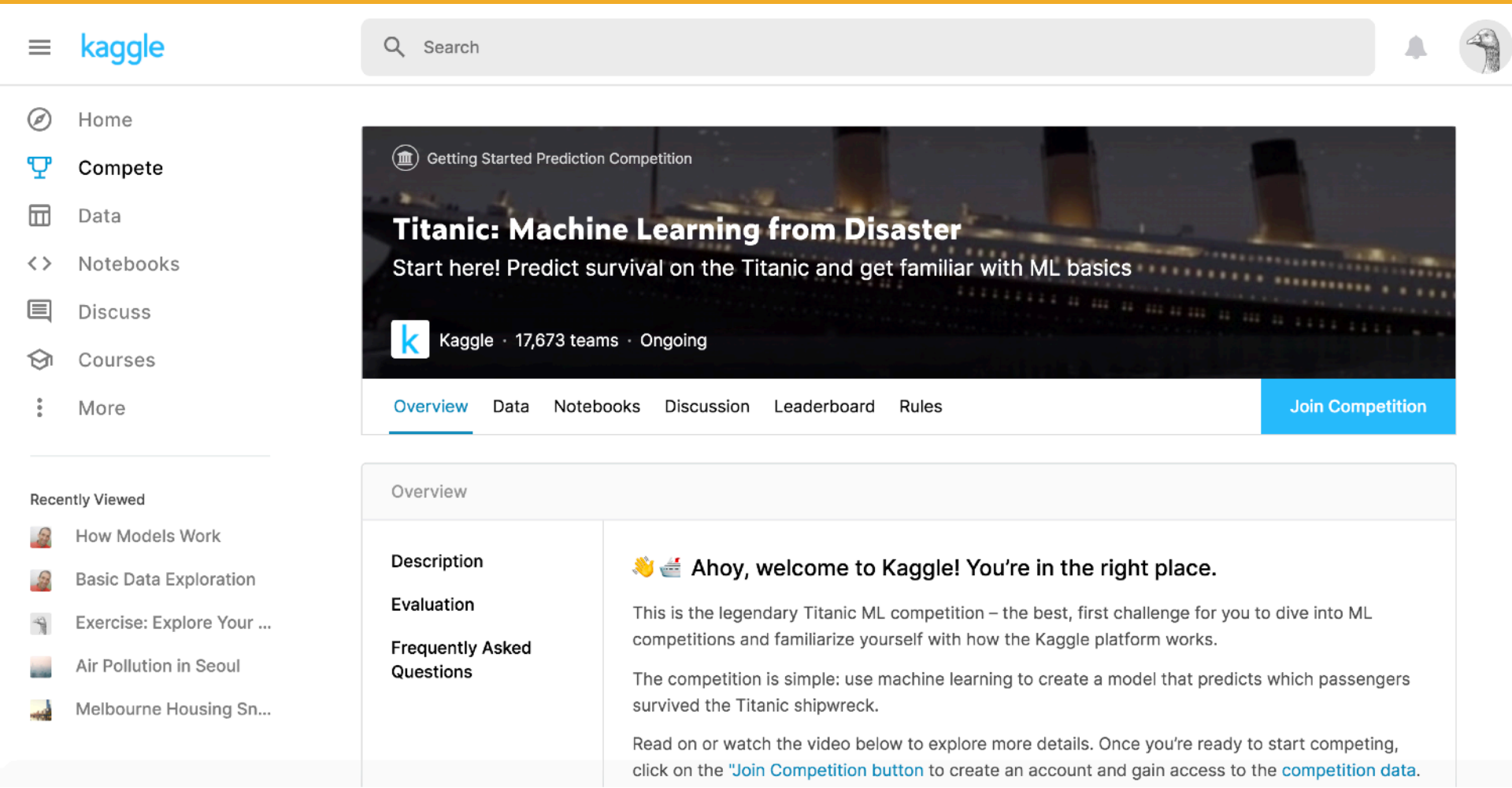
Features Screenshots Workflows Download Blog Docs Training Donate

Data Mining
Fruitful and Fun

Open source machine learning and data visualization for novice and expert. Interactive data analysis workflows with a large toolbox.

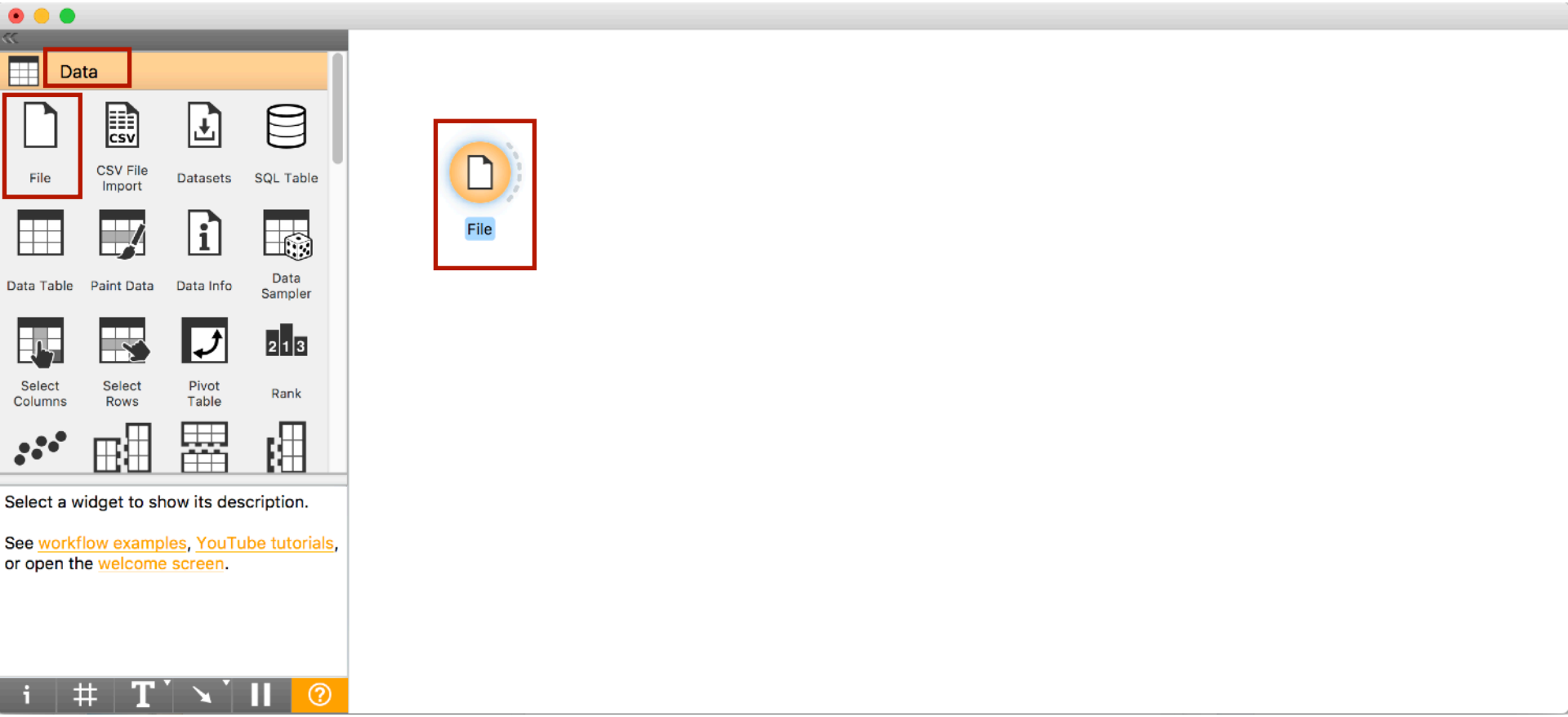
Download Orange

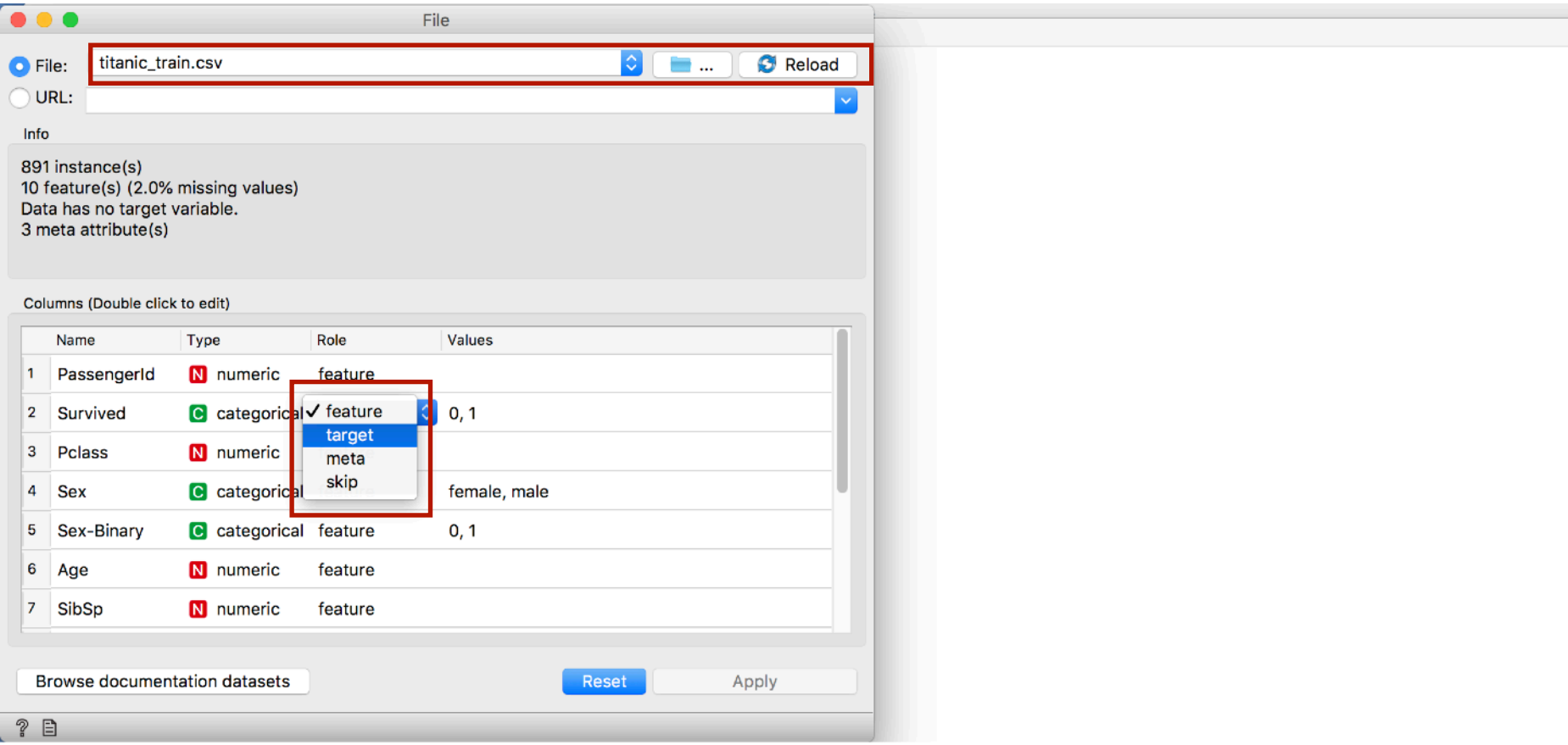


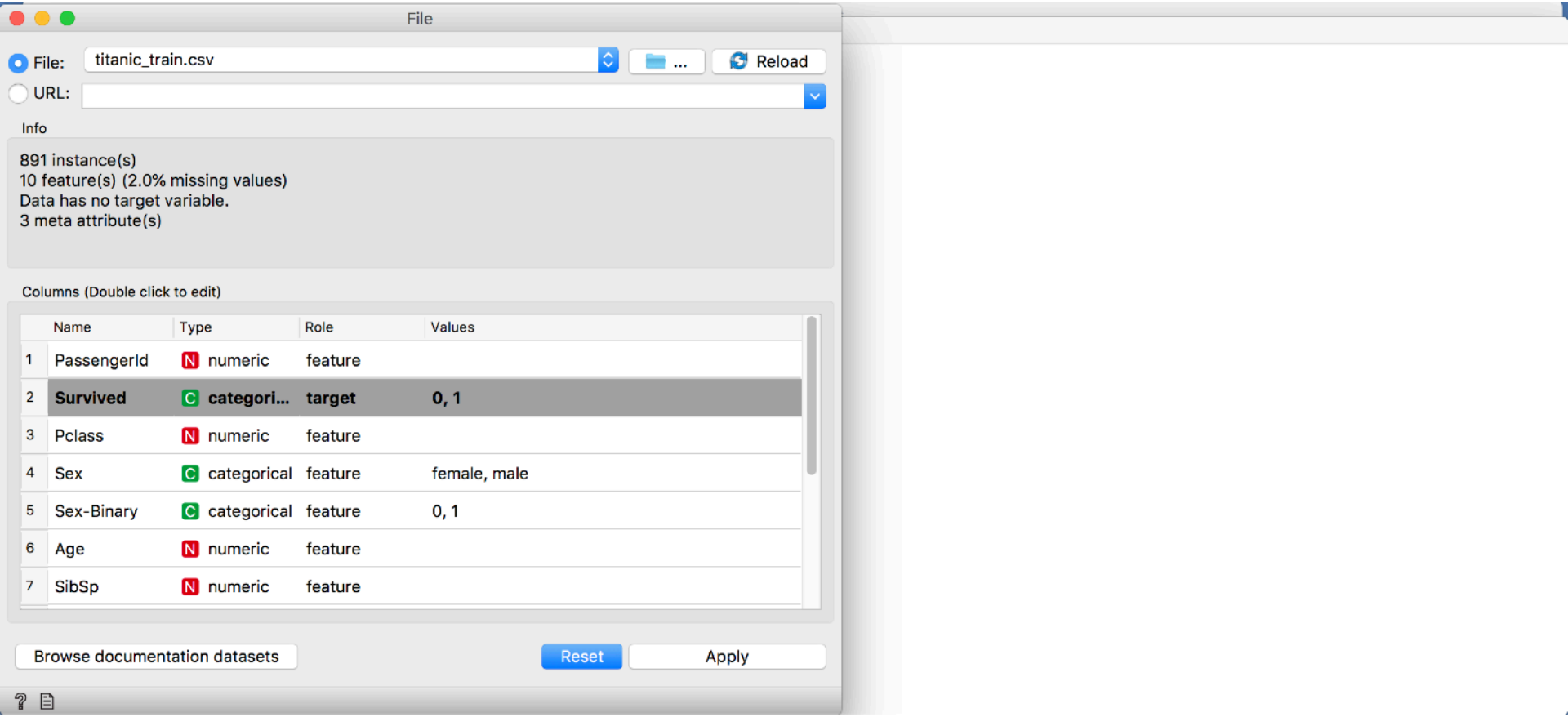


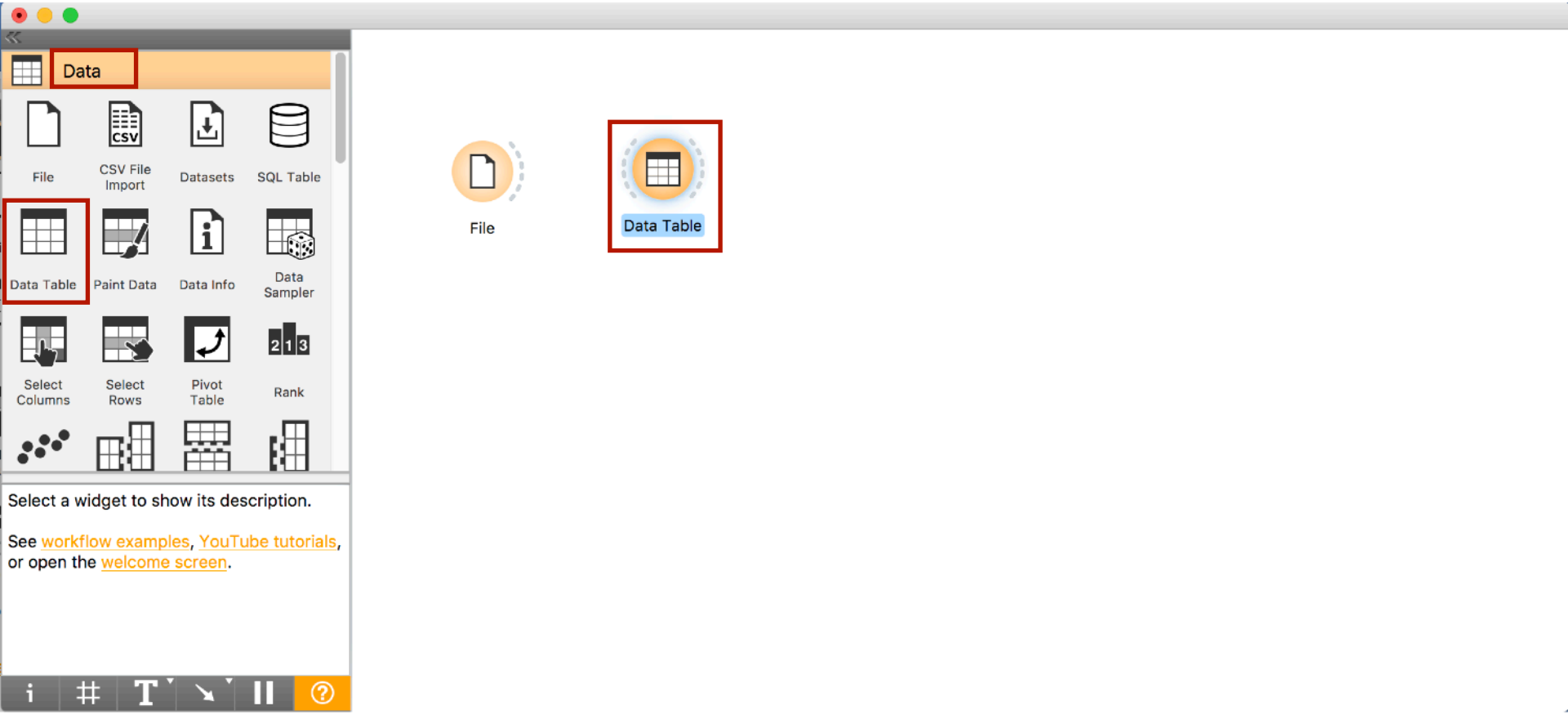
https://www.kaggle.com/c/titanic/

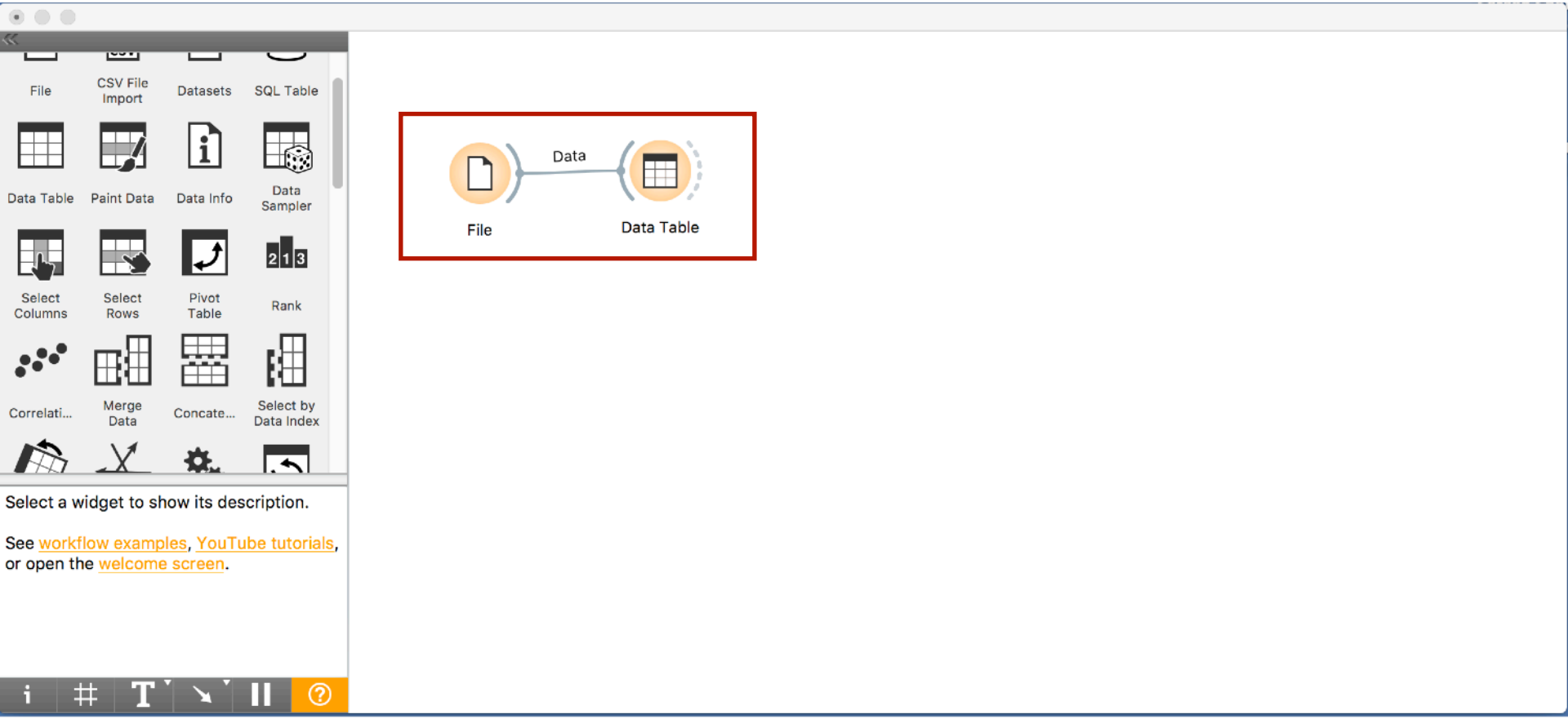
Let's train/build a Titanic survivor model pipeline in Orange to get you started in joining the Kaggle community.

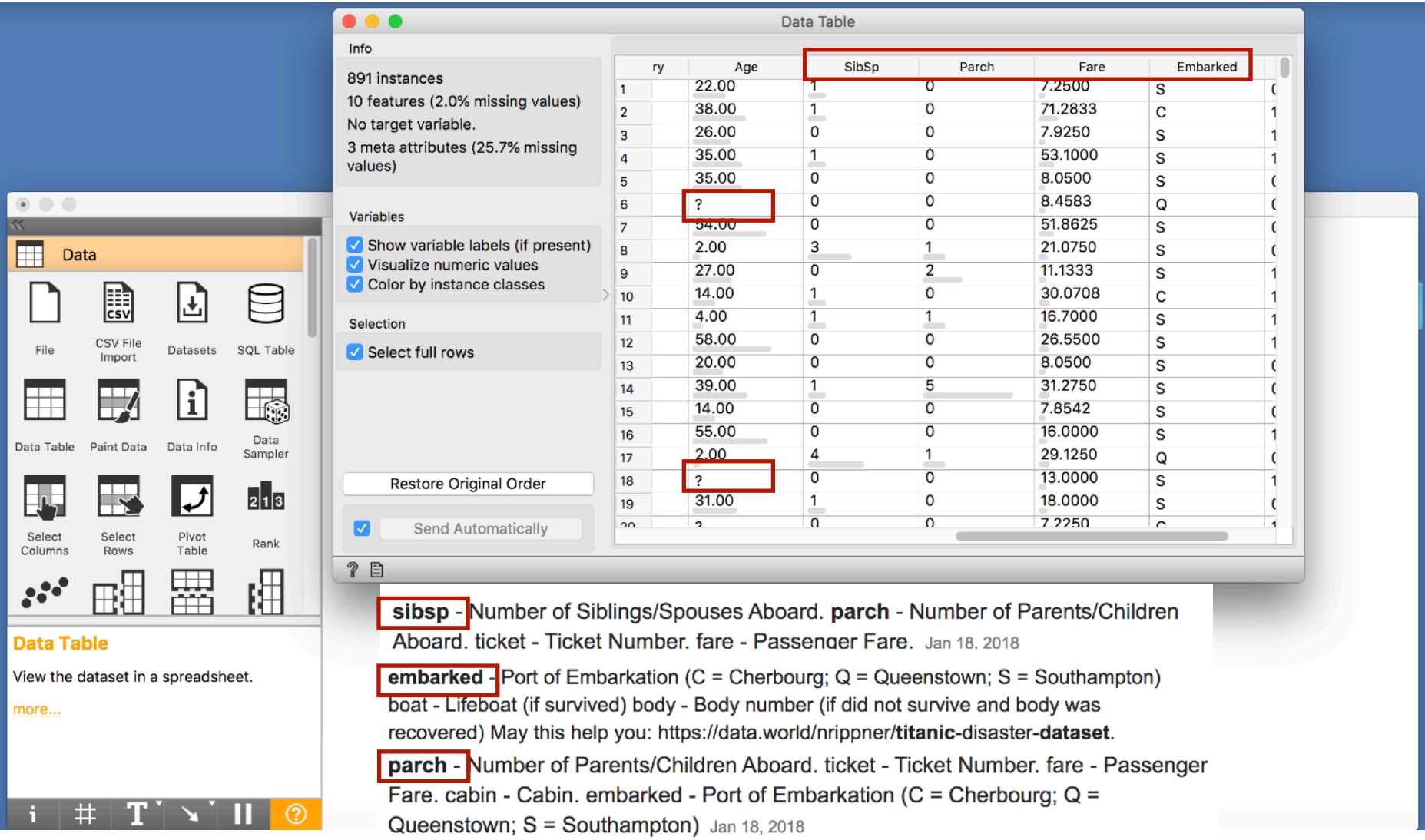

Data
File
CSV File Import
Datasets
SQL Table
Data Table
Paint Data
Data Info
Data Sampler
Select Columns
Select Rows
Pivot Table
Rank
Select a widget to show its description.
See workflow examples, YouTube tutorials, or open the welcome screen.
File

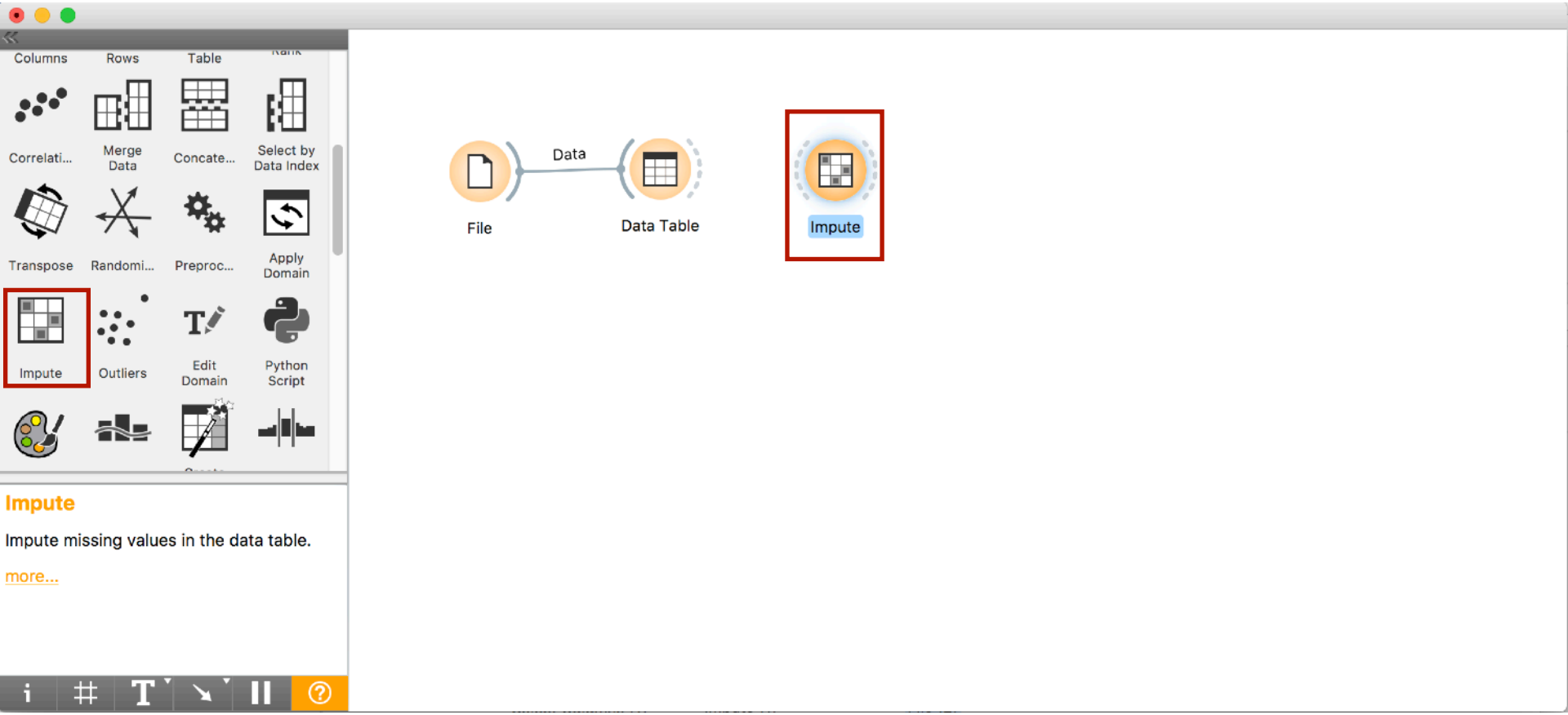

File
File:
titanic_train.csv
...
Reload
URL:
Info
891 instance(s)
10 feature(s) (2.0% missing values)
Data has no target variable.
3 meta attribute(s)
Columns (Double click to edit)
Name
Type
Role
Values
1 PassengerId numeric feature
2 Survived categorical feature 0, 1
3 Pclass numeric
4 Sex categorical female, male
5 Sex-Binary categorical feature 0, 1
6 Age numeric feature
7 SibSp numeric feature
feature
target
meta
skip
Browse documentation datasets
Reset
Apply

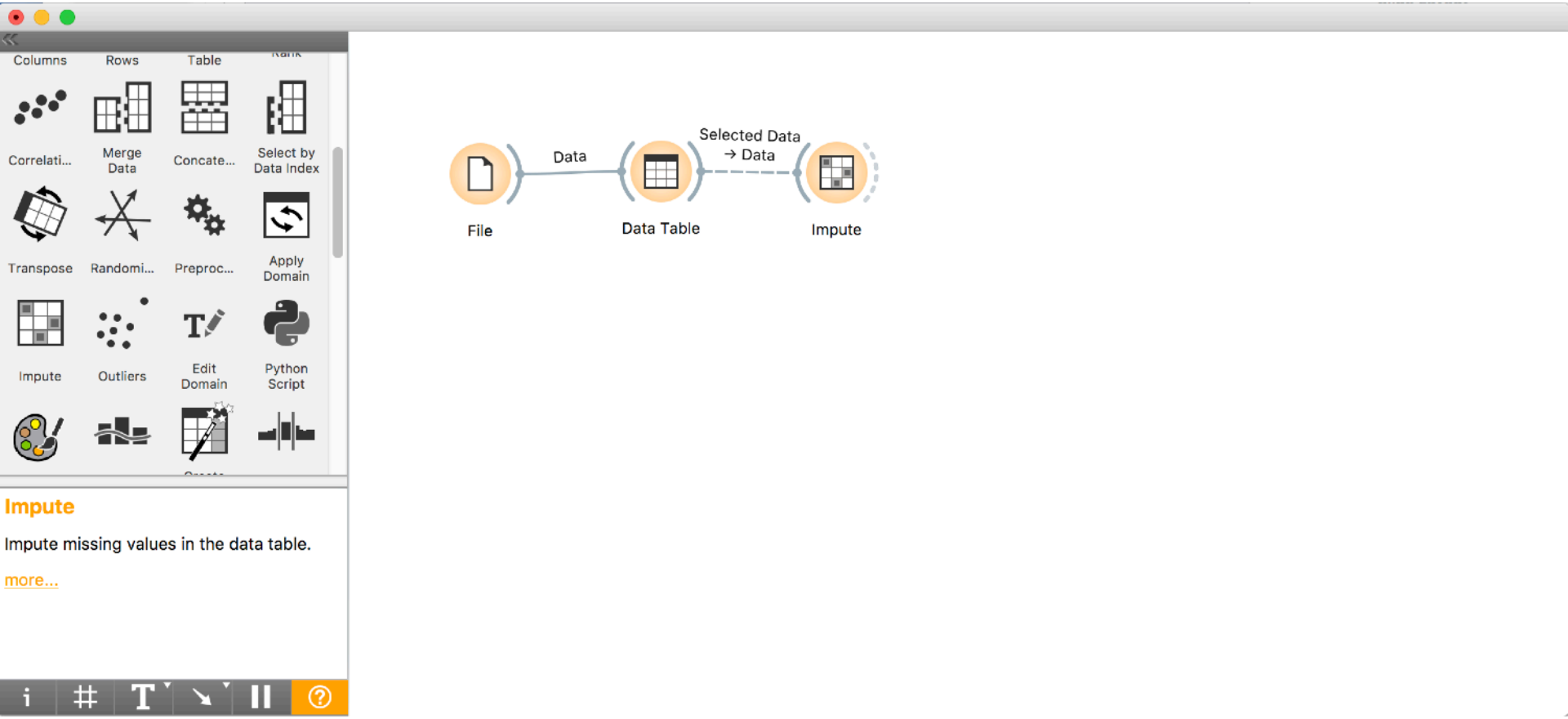

File
File:
titanic_train.csv
...
Reload
URL:
Info
891 instance(s)
10 feature(s) (2.0% missing values)
Data has no target variable.
3 meta attribute(s)
Columns (Double click to edit)
Name
Type
Role
Values
1 PassengerId N numeric feature
2 Survived C categori... target 0, 1
3 Pclass N numeric feature
4 Sex C categorical feature female, male
5 Sex-Binary C categorical feature 0, 1
6 Age N numeric feature
7 SibSp N numeric feature
Browse documentation datasets
Reset
Apply

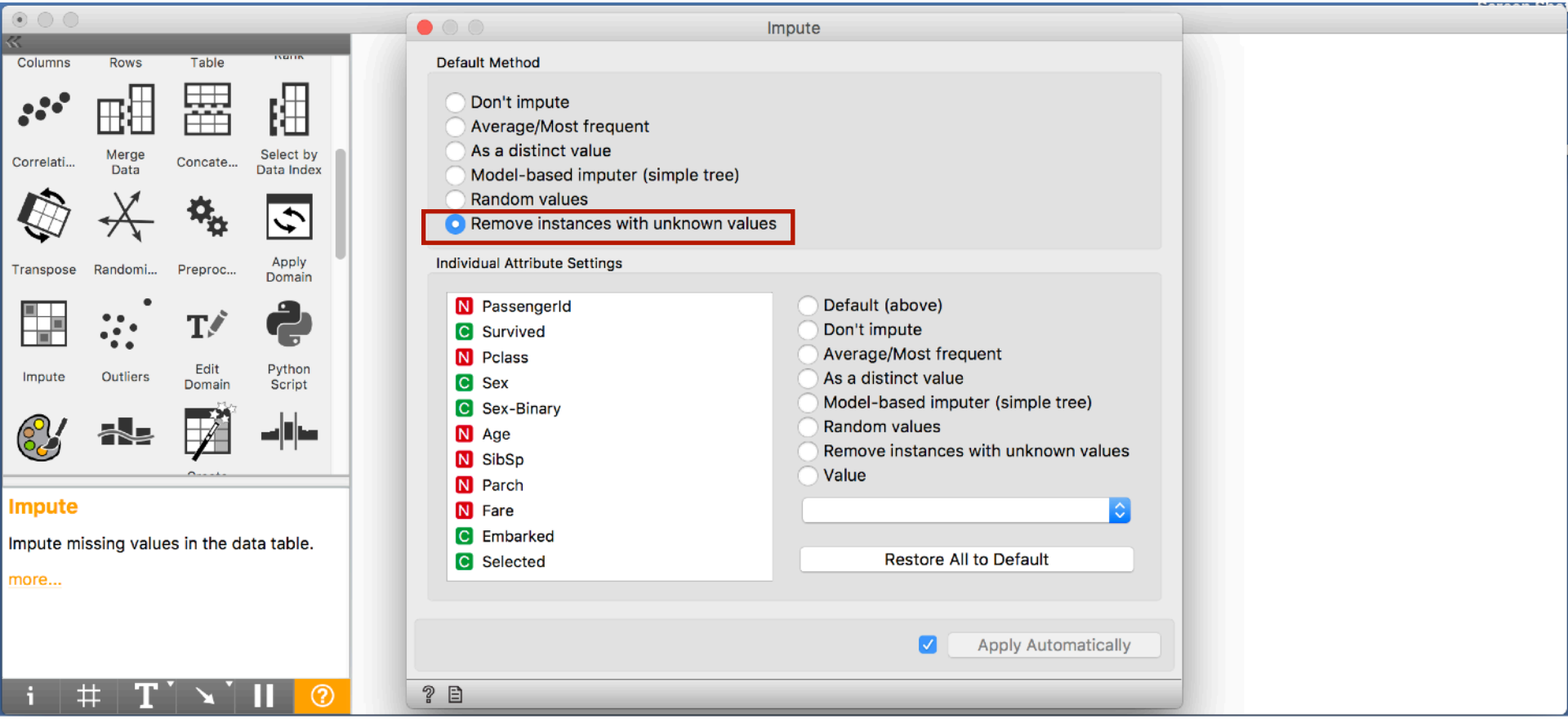

Data
File
CSV File Import
Datasets
SQL Table
Data Table
Paint Data
Data Info
Data Sampler
Select Columns
Select Rows
Pivot Table
Rank
Select a widget to show its description.
See workflow examples, YouTube tutorials, or open the welcome screen.
File
Data Table

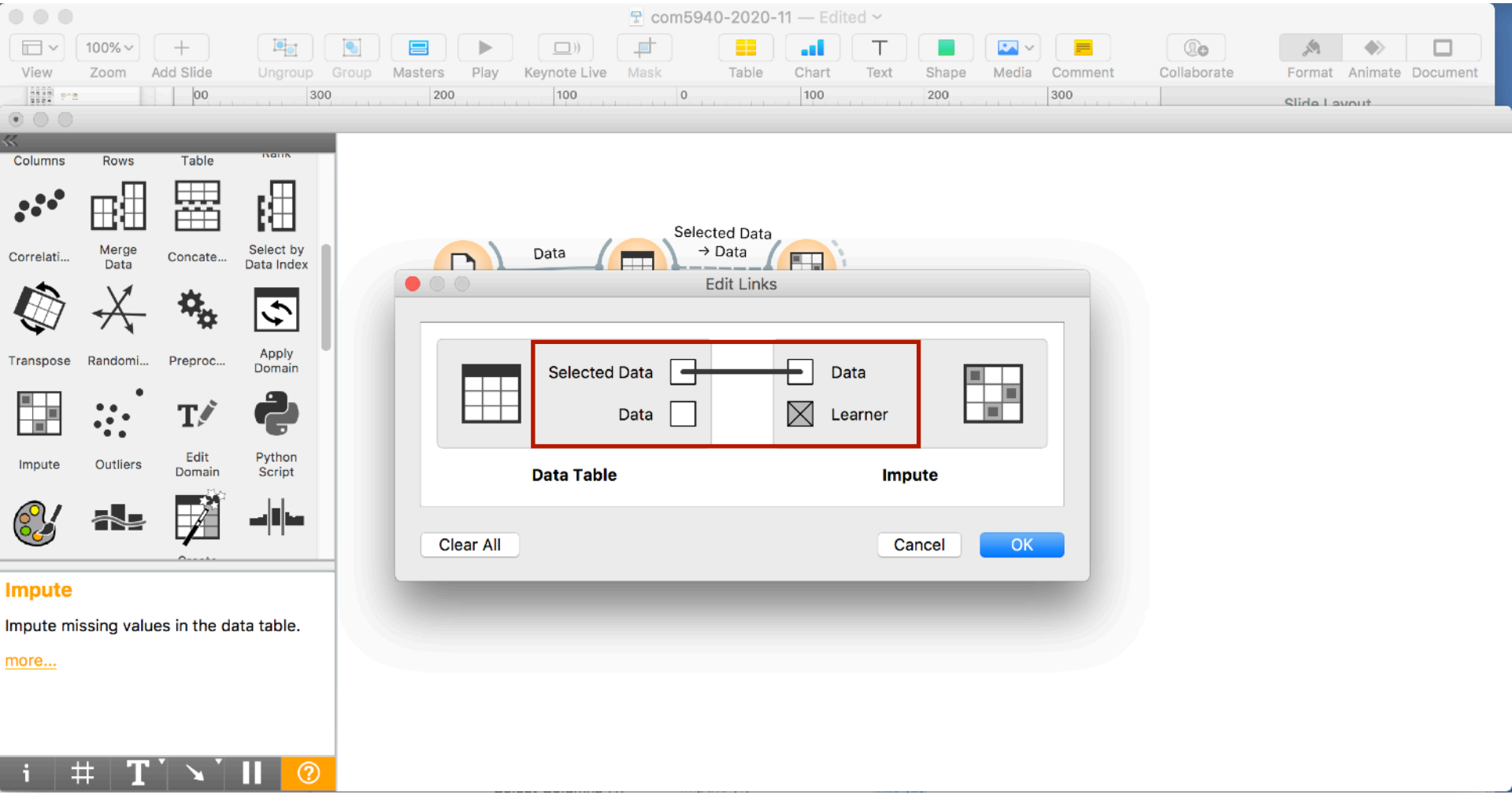

File
CSV File Import
Datasets
SQL Table
Data Table
Paint Data
Data Info
Data Sampler
Select Columns
Select Rows
Pivot Table
Rank
Correlati...
Merge Data
Concate...
Select by Data Index
Select a widget to show its description.
See workflow examples, YouTube tutorials, or open the welcome screen.
Data
File
Data Table

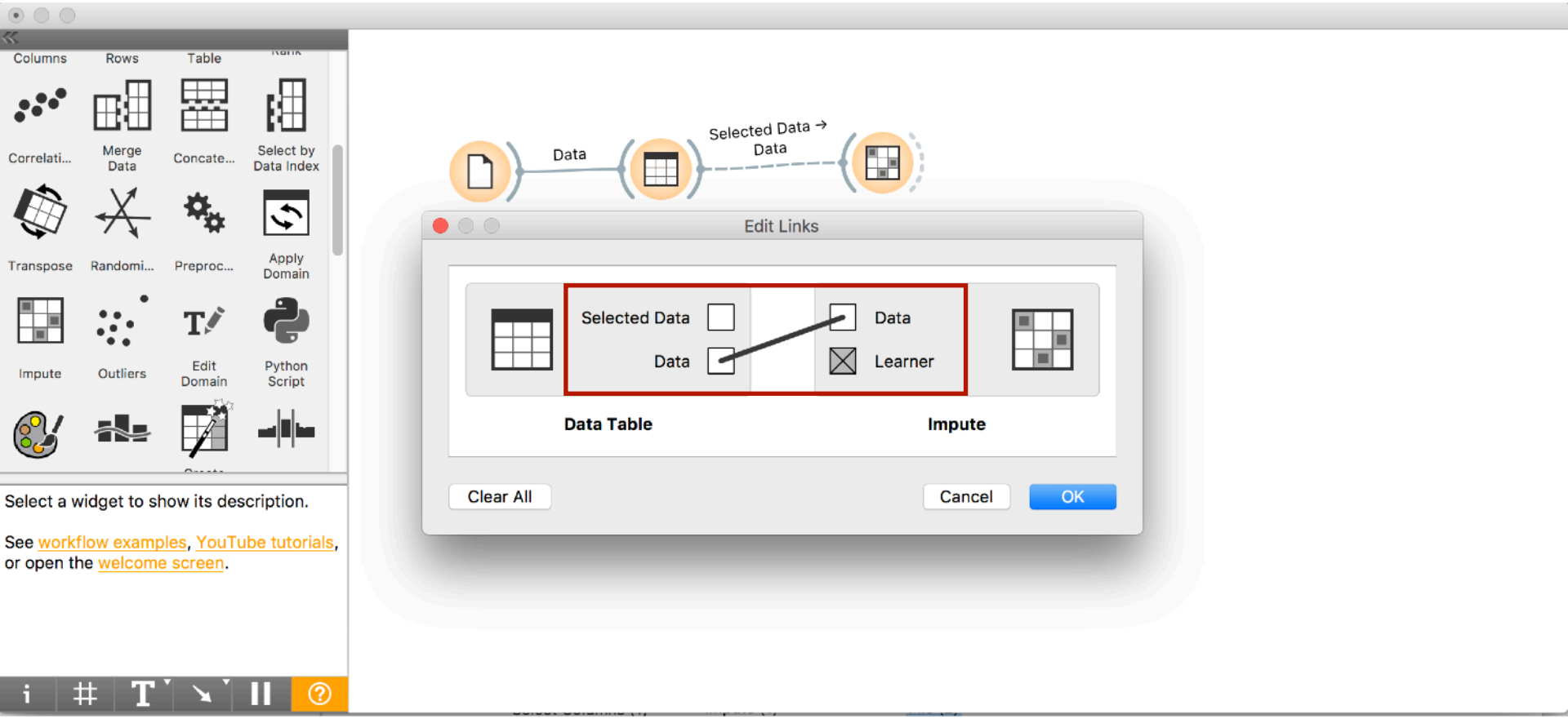

Data Table
Info
891 instances
10 features (2.0% missing values)
No target variable.
3 meta attributes (25.7% missing values)
Variables
Show variable labels (if present)
Visualize numeric values
Color by instance classes
Selection
Select full rows
Restore Original Order
Send Automatically
ry
Age
SibSp
Parch
Fare
Embarked
1 22.00 1 0 7.2500 S
2 38.00 1 0 71.2833 C
3 26.00 0 0 7.9250 S
4 35.00 1 0 53.1000 S
5 35.00 0 0 8.0500 S
6 ? 0 0 8.4583 Q
7 54.00 0 0 51.8625 S
8 2.00 3 1 21.0750 S
9 27.00 0 2 11.1333 S
10 14.00 1 0 30.0708 C
11 4.00 1 1 16.7000 S
12 58.00 0 0 26.5500 S
13 20.00 0 0 8.0500 S
14 39.00 1 5 31.2750 S
15 14.00 0 0 7.8542 S
16 55.00 0 0 16.0000 S
17 2.00 4 1 29.1250 Q
18 ? 0 0 13.0000 S
19 31.00 1 0 18.0000 S
Data
File
CSV File Import
Datasets
SQL Table
Data Table
Paint Data
Data Info
Data Sampler
Select Columns
Select Rows
Pivot Table
Rank
Data Table
View the dataset in a spreadsheet.
more...
sibsp - Number of Siblings/Spouses Aboard. parch - Number of Parents/Children Aboard. ticket - Ticket Number. fare - Passenger Fare. Jan 18, 2018
embarked - Port of Embarkation (C = Cherbourg; Q = Queenstown; S = Southampton) boat - Lifeboat (if survived) body - Body number (if did not survive and body was recovered) May this help you: https://data.world/nrippner/titanic-disaster-dataset.
parch - Number of Parents/Children Aboard. ticket - Ticket Number. fare - Passenger Fare. cabin - Cabin. embarked - Port of Embarkation (C = Cherbourg; Q = Queenstown; S = Southampton) Jan 18, 2018

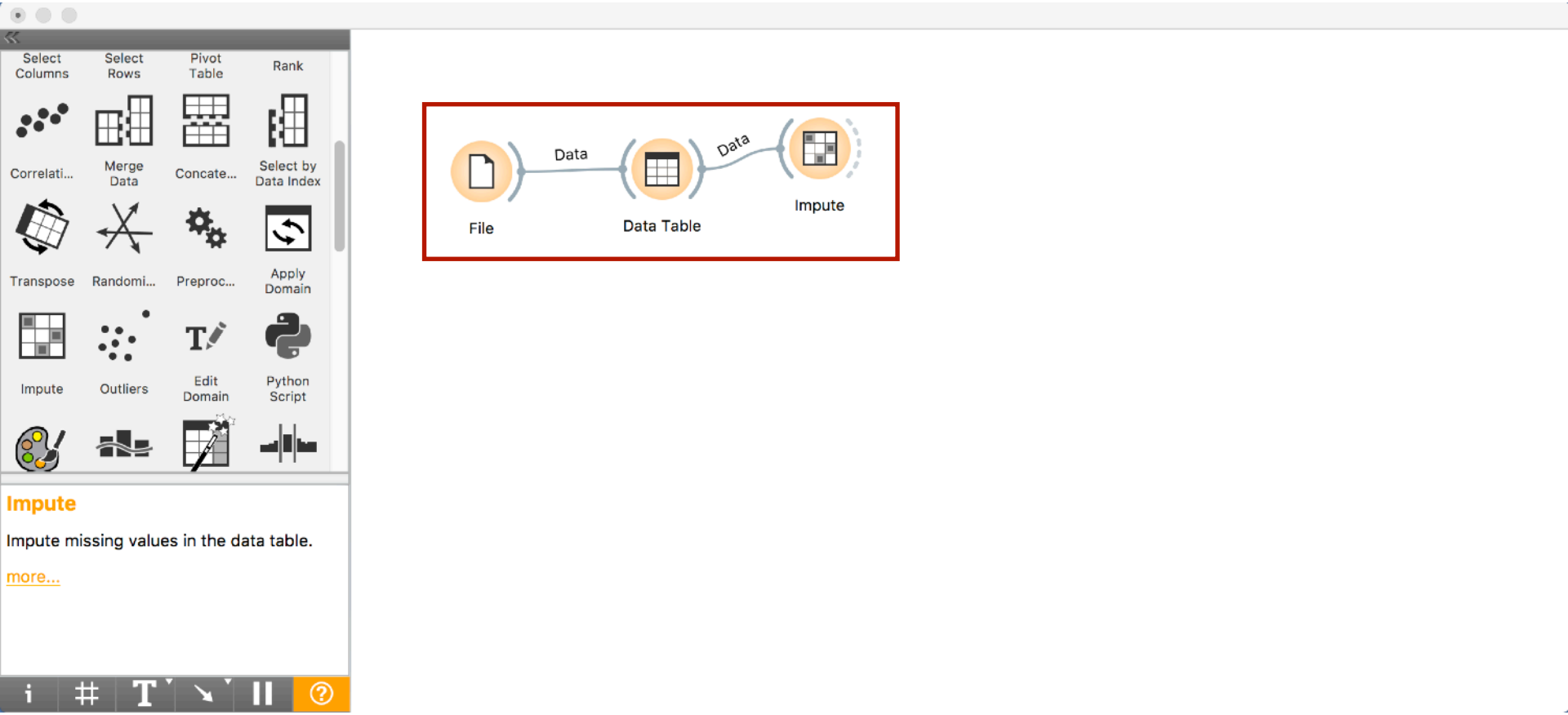

Columns
Rows
Table
Correlati...
Merge Data
Concate...
Select by Data Index
Transpose
Randomi...
Preproc...
Apply Domain
Impute
Outliers
Edit Domain
Python Script
Impute
Impute missing values in the data table.
more...
Data
File
Data Table
Impute

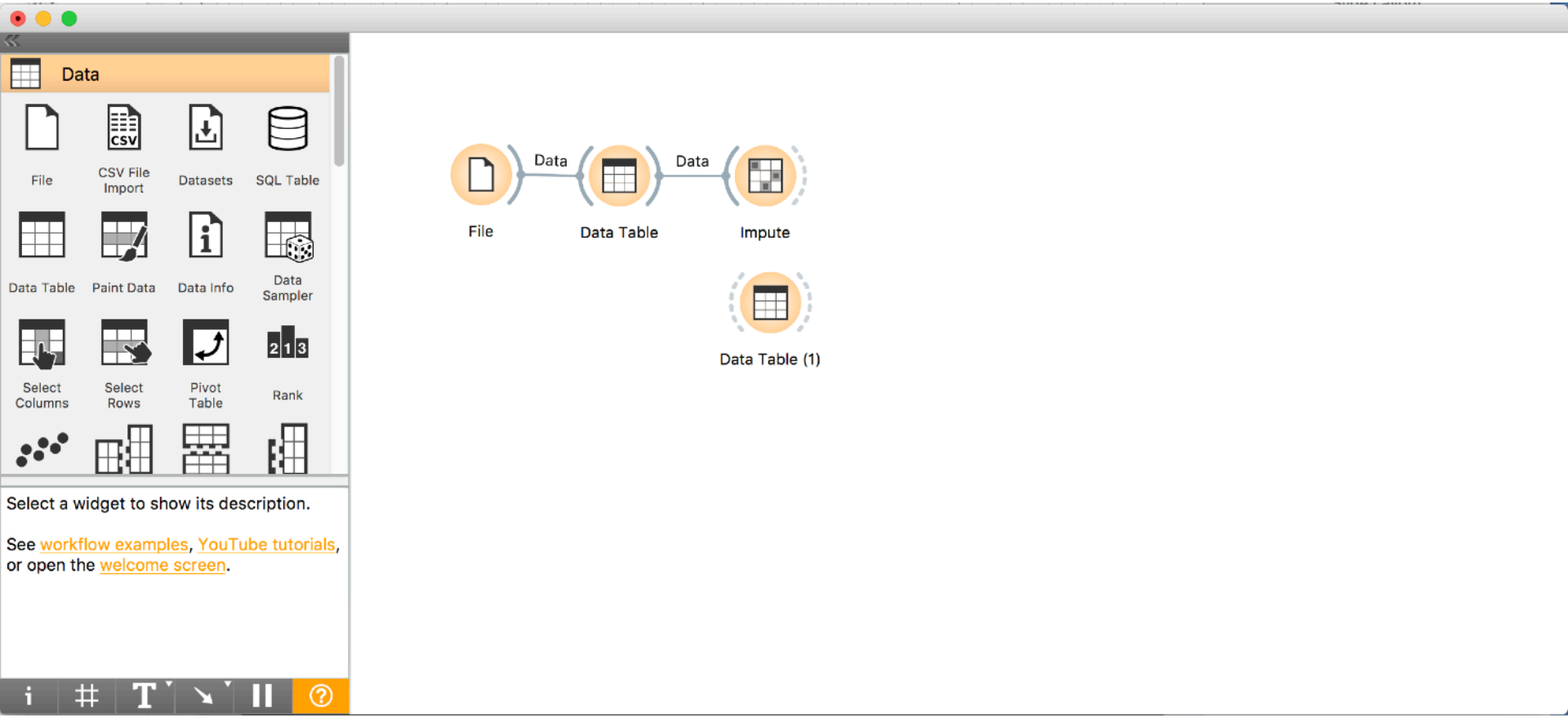

Columns
Rows
Table
Correlati...
Merge Data
Concate...
Select by Data Index
Transpose
Randomi...
Preproc...
Apply Domain
Impute
Outliers
Edit Domain
Python Script
Impute
Impute missing values in the data table.
more...
Data
Selected Data → Data
File
Data Table
Impute

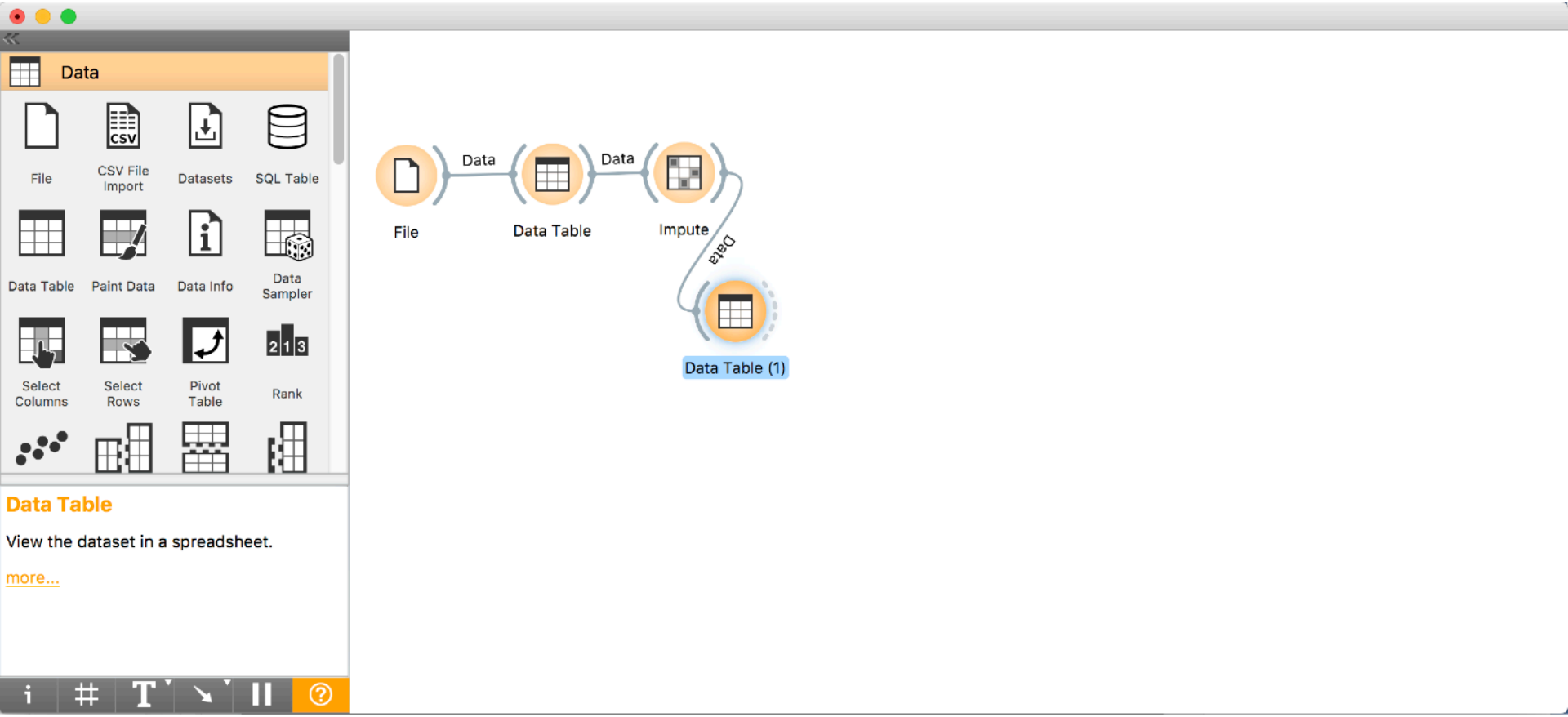

Columns
Rows
Table
Correlati...
Merge Data
Concate...
Select by Data Index
Transpose
Randomi...
Preproc...
Apply Domain
Impute
Outliers
Edit Domain
Python Script
Impute
Impute missing values in the data table.
more...
Impute
Default Method
Don't impute
Average/Most frequent
As a distinct value
Model-based imputer (simple tree)
Random values
Remove instances with unknown values
Individual Attribute Settings
PassengerId
Survived
Pclass
Sex
Sex-Binary
Age
SibSp
Parch
Fare
Embarked
Selected
Default (above)
Don't impute
Average/Most frequent
As a distinct value
Model-based imputer (simple tree)
Random values
Remove instances with unknown values
Value
Restore All to Default
Apply Automatically


com5940-2020-11 — Edited
View
Zoom
Add Slide
Ungroup
Group
Masters
Play
Keynote Live
Mask
Table
Chart
Text
Shape
Media
Comment
Collaborate
Format
Animate
Document
Columns
Rows
Table
Correlati...
Merge Data
Concate...
Select by Data Index
Transpose
Randomi...
Preproc...
Apply Domain
Impute
Outliers
Edit Domain
Python Script
Impute
Impute missing values in the data table.
more...
Data
Selected Data → Data
Edit Links
Selected Data
Data
Data
Learner
Data Table
Impute
Clear All
Cancel
OK

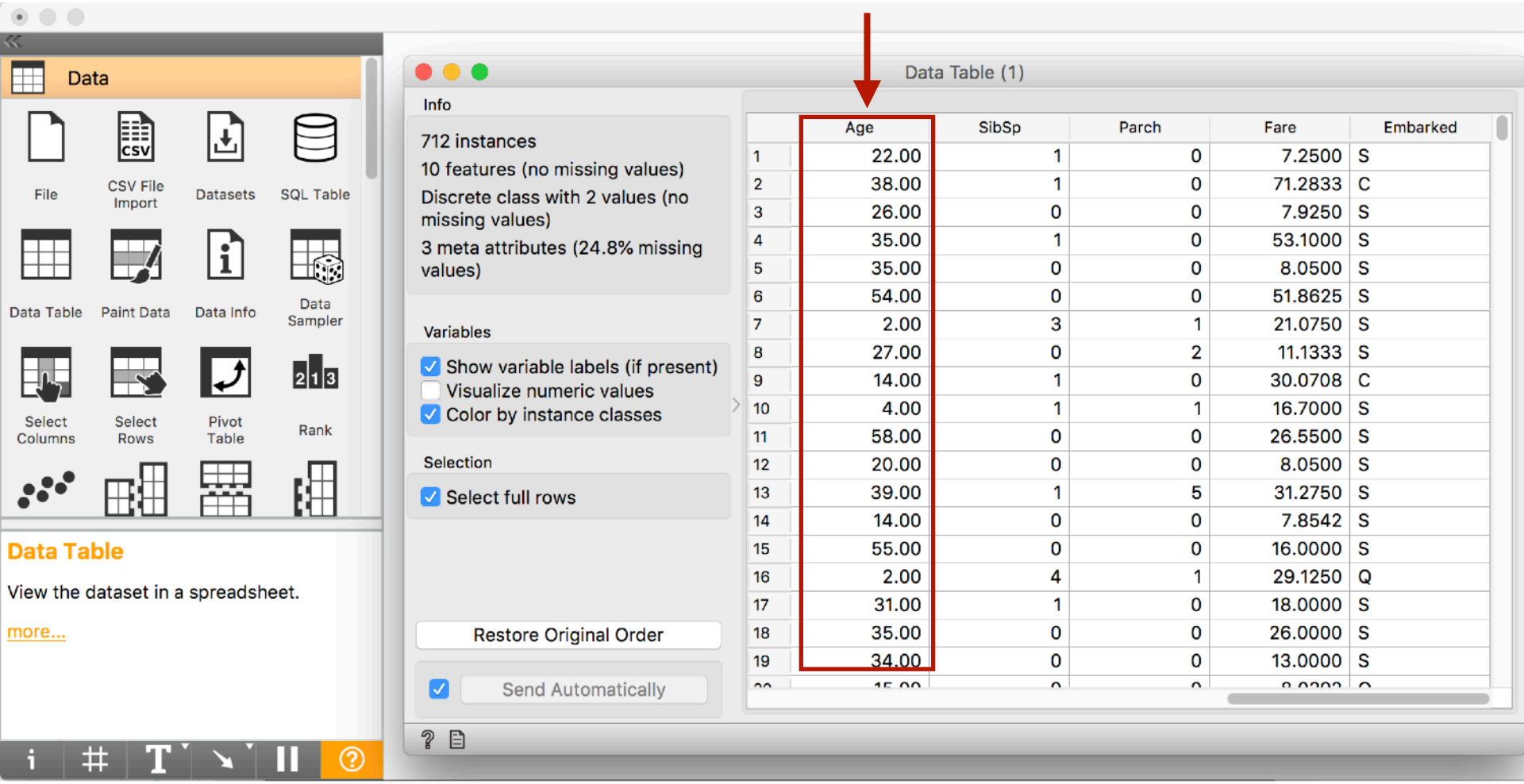

Columns
Rows
Table
Correlati...
Merge Data
Concate...
Select by Data Index
Transpose
Randomi...
Preproc...
Apply Domain
Impute
Outliers
Edit Domain
Python Script
Select a widget to show its description.
See workflow examples, YouTube tutorials, or open the welcome screen.
Data
Selected Data → Data
Edit Links
Selected Data
Data
Data
Learner
Data Table
Impute
Clear All
Cancel
OK


Select Columns
Select Rows
Pivot Table
Rank
Correlati...
Merge Data
Concate...
Select by Data Index
Transpose
Randomi...
Preproc...
Apply Domain
Impute
Outliers
Edit Domain
Python Script
Impute
Impute missing values in the data table.
more...
Data
Data
File
Data Table
Impute


Data
File
CSV File Import
Datasets
SQL Table
Data Table
Paint Data
Data Info
Data Sampler
Select Columns
Select Rows
Pivot Table
Rank
Select a widget to show its description.
See workflow examples, YouTube tutorials, or open the welcome screen.
Data
Data
File
Data Table
Impute
Data Table (1)

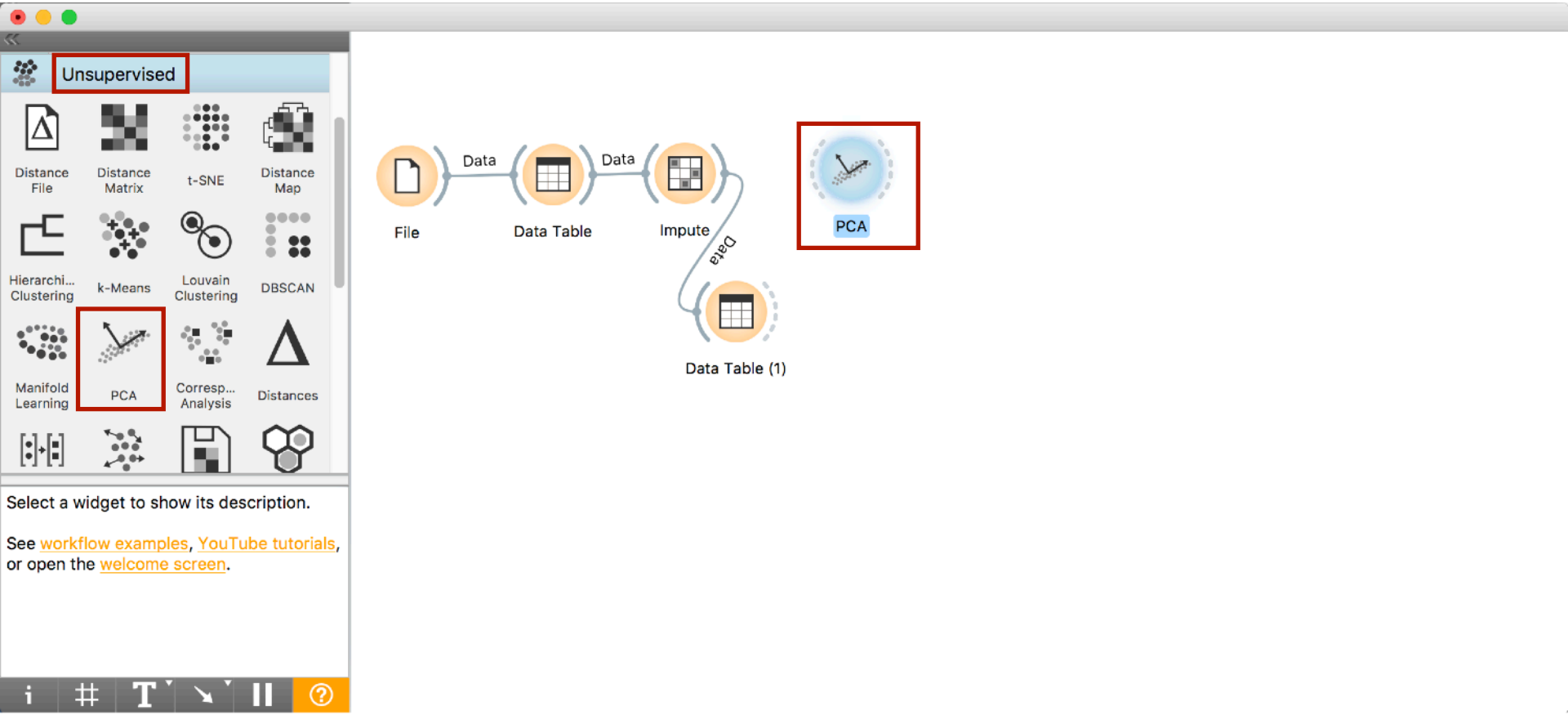

Data
File
CSV File Import
Datasets
SQL Table
Data Table
Paint Data
Data Info
Data Sampler
Select Columns
Select Rows
Pivot Table
Rank
Data Table
View the dataset in a spreadsheet.
more...
Data
Data
File
Data Table
Impute
Data
Data Table (1)

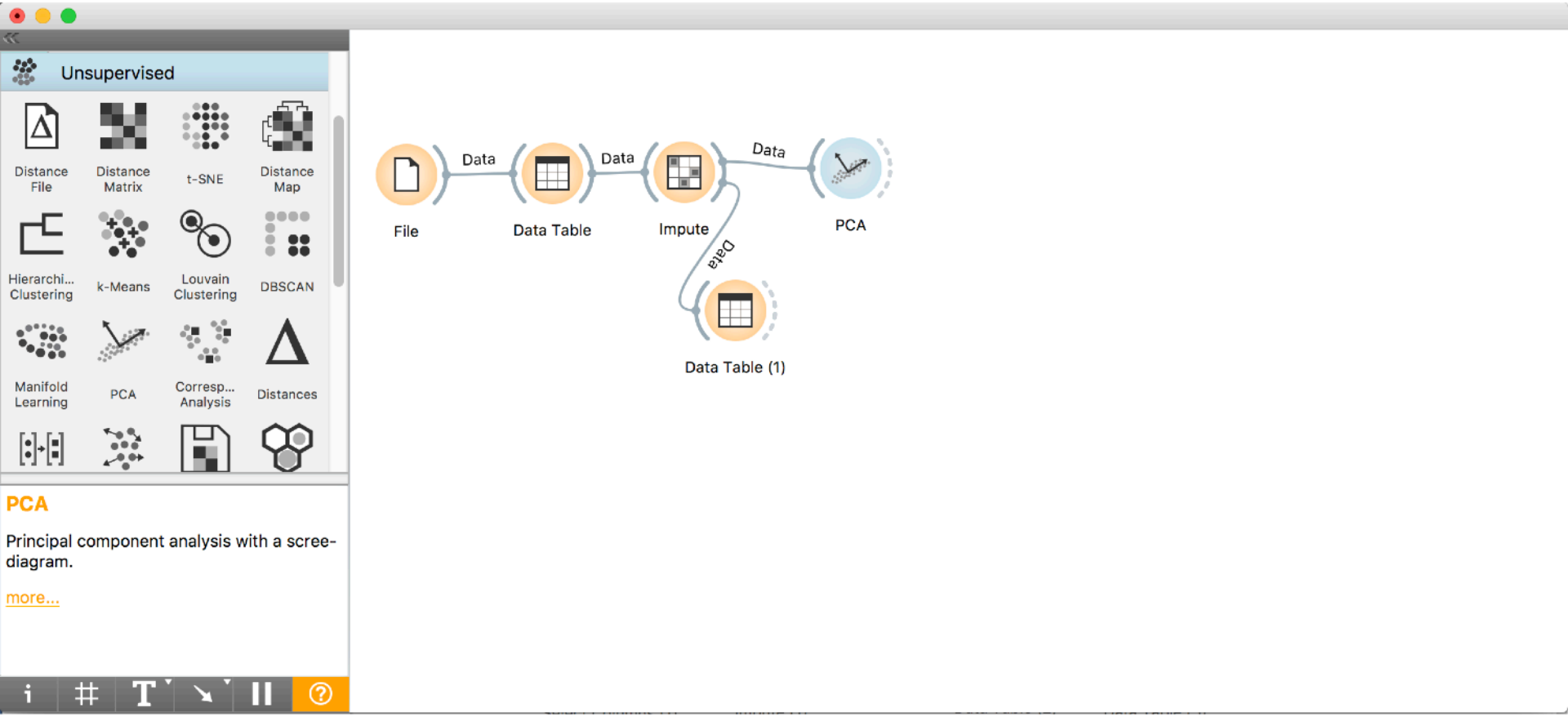
No more zeros

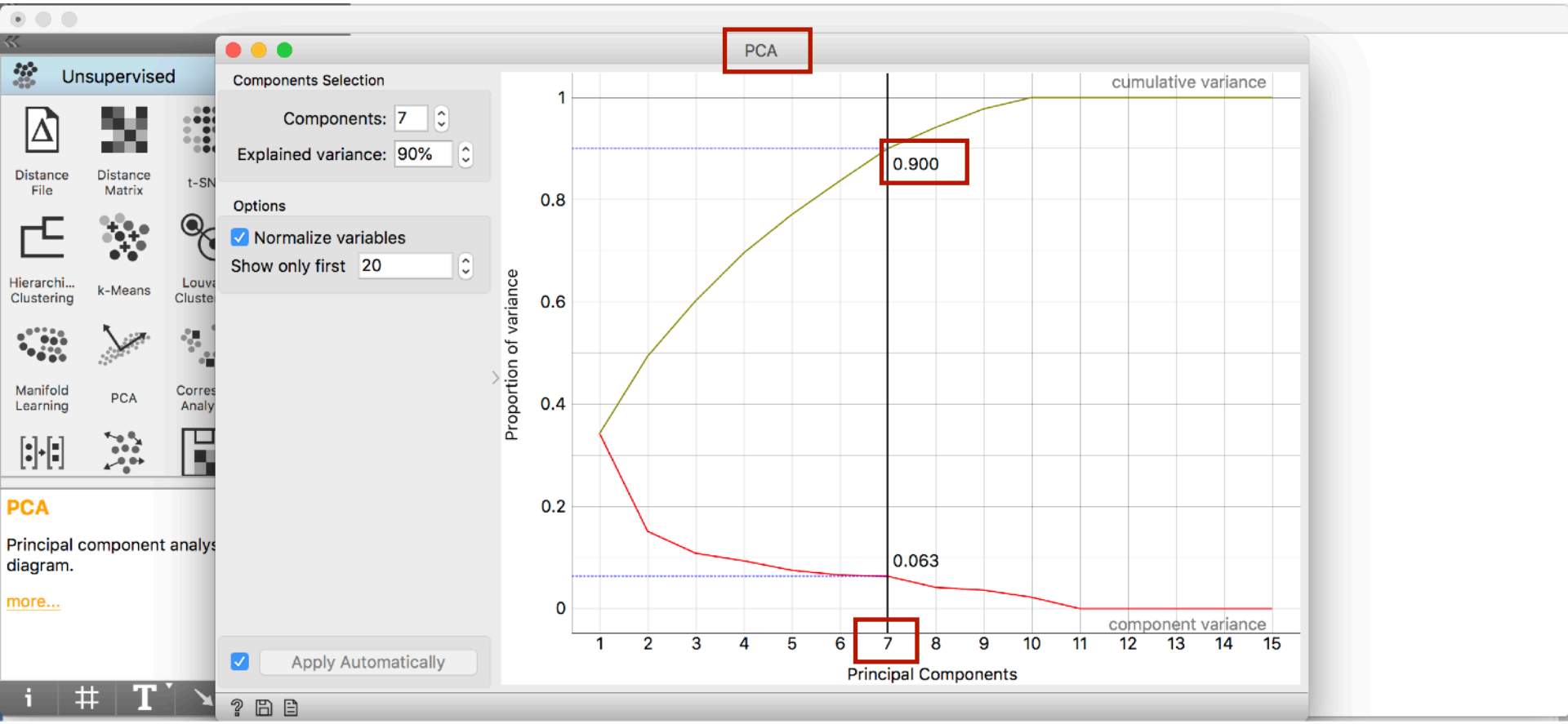


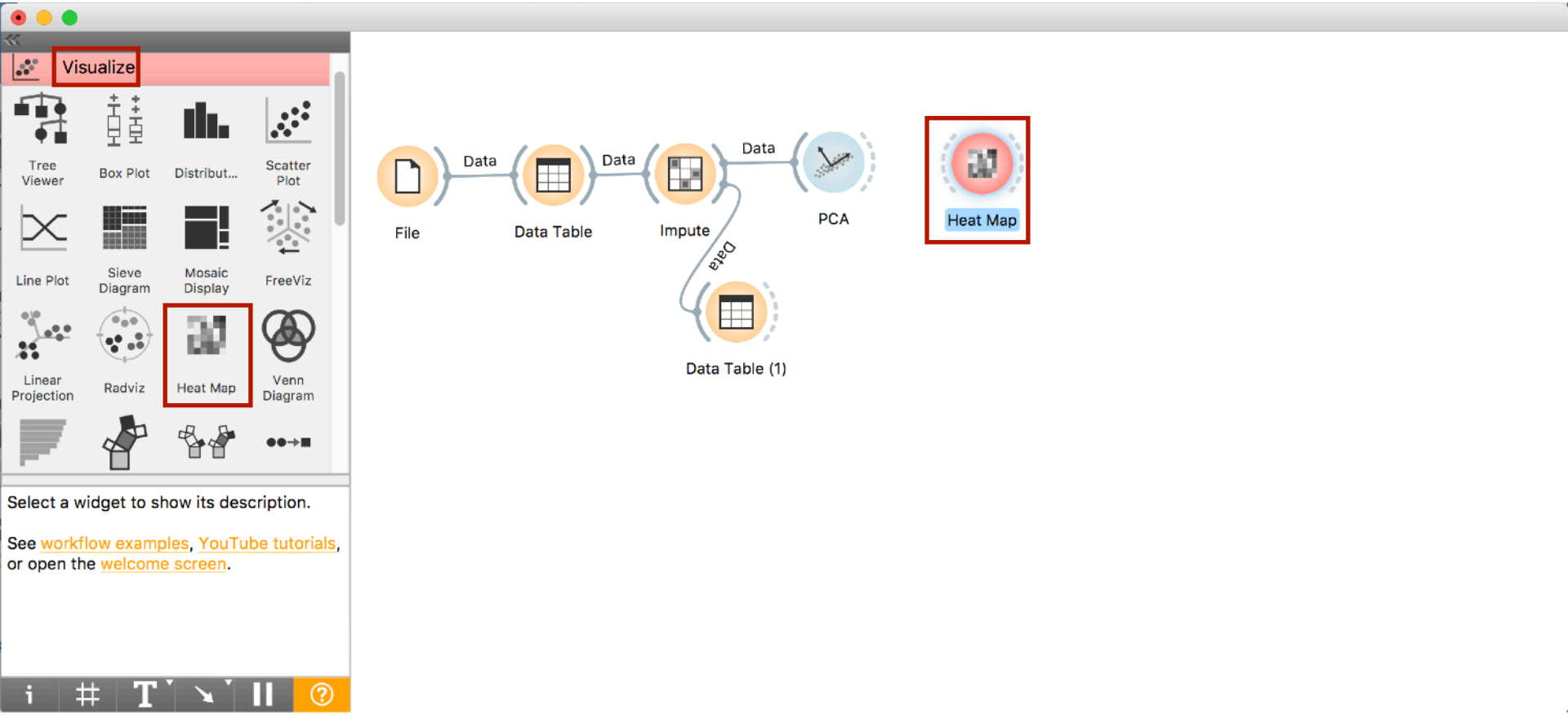
	Age	SibSp	Parch	Fare	Embarked
1	22.00	1	0	7.2500	S
2	38.00	1	0	71.2833	C
3	26.00	0	0	7.9250	S
4	35.00	1	0	53.1000	S
5	35.00	0	0	8.0500	S
6	54.00	0	0	51.8625	S
7	2.00	3	1	21.0750	S
8	27.00	0	2	11.1333	S
9	14.00	1	0	30.0708	C
10	4.00	1	1	16.7000	S
11	58.00	0	0	26.5500	S
12	20.00	0	0	8.0500	S
13	39.00	1	5	31.2750	S
14	14.00	0	0	7.8542	S
15	55.00	0	0	16.0000	S
16	2.00	4	1	29.1250	Q
17	31.00	1	0	18.0000	S
18	35.00	0	0	26.0000	S
19	34.00	0	0	13.0000	S

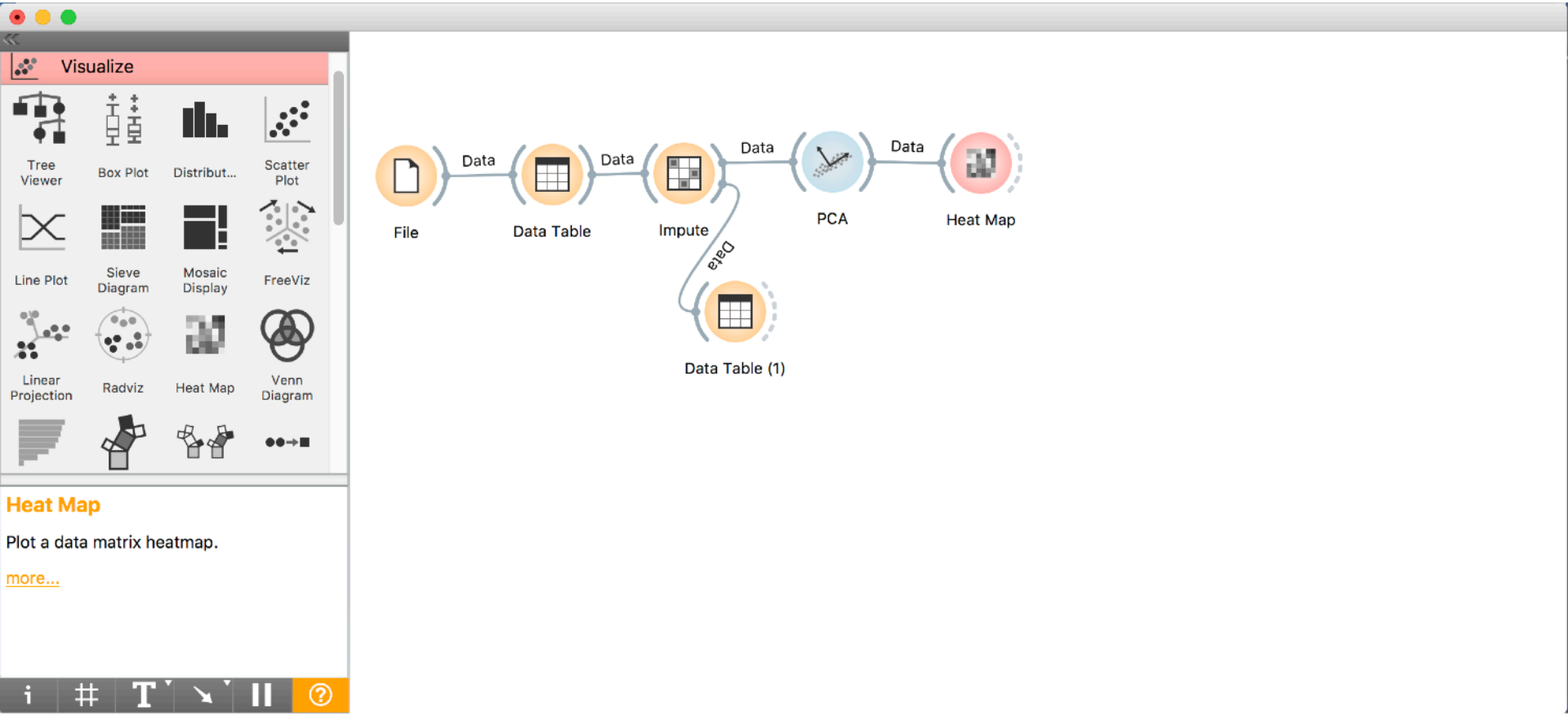
Dimension reduction and feature engineering in Orange using Unsupervised Learning for finding patterns.

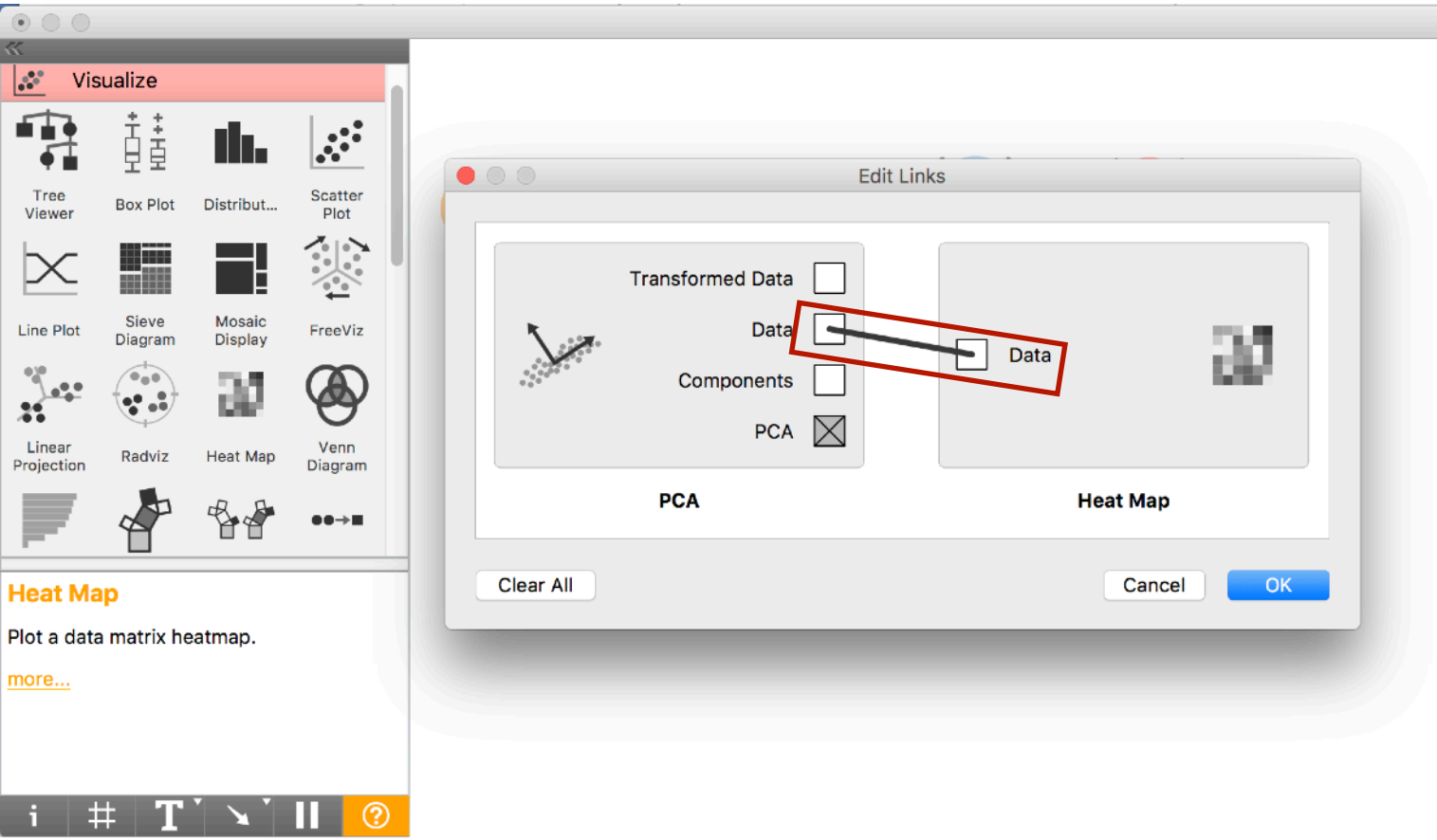

Unsupervised
Distance File
Distance Matrix
t-SNE
Distance Map
Hierarchi... Clustering
k-Means
Louvain Clustering
DBSCAN
Manifold Learning
PCA
Corresp... Analysis
Distances
Select a widget to show its description.
See workflow examples, YouTube tutorials, or open the welcome screen.
Data
Data
Data
File
Data Table
Impute
PCA
Data Table (1)

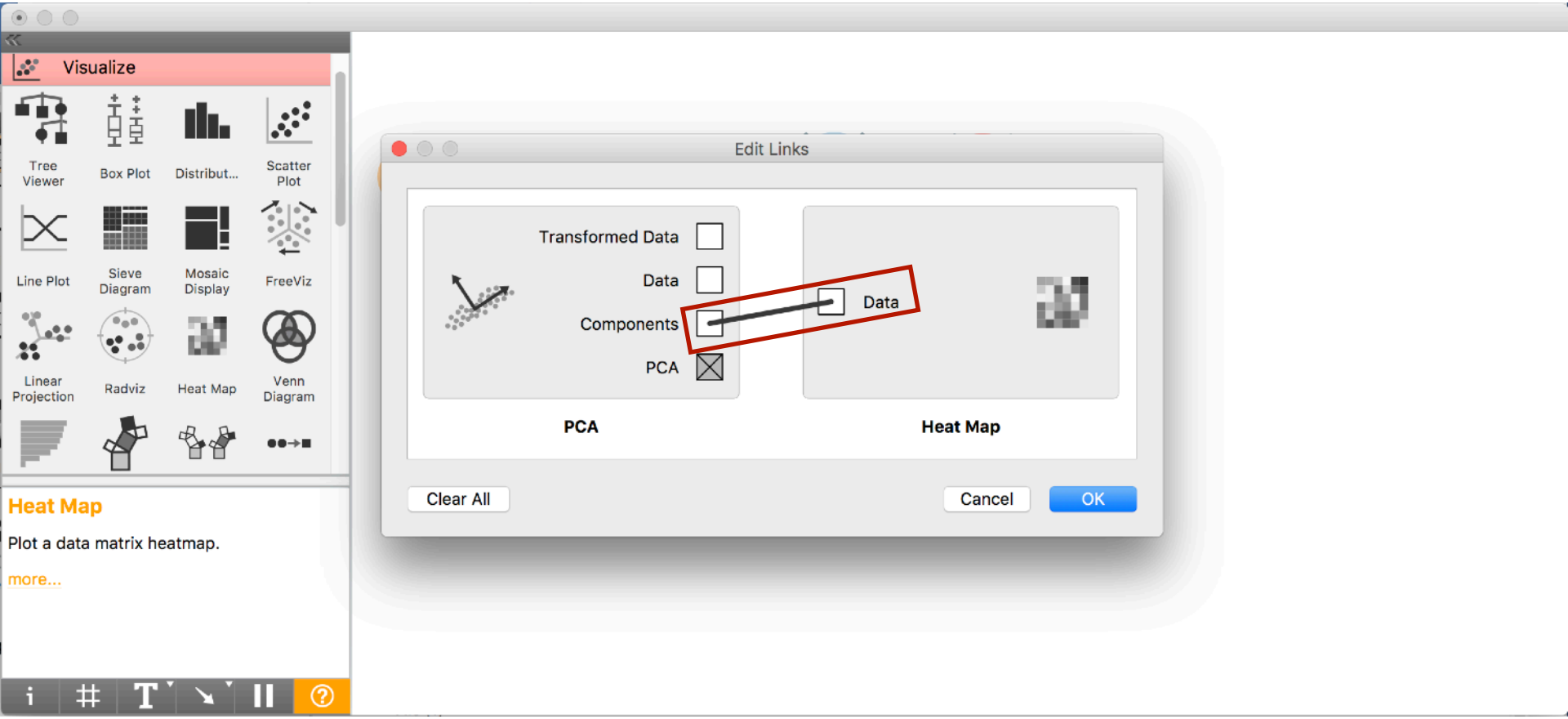

Unsupervised
Distance File
Distance Matrix
t-SNE
Distance Map
Hierarchi... Clustering
k-Means
Louvain Clustering
DBSCAN
Manifold Learning
PCA
Corresp... Analysis
Distances
PCA
Principal component analysis with a scree-diagram.
more...
Data
Data
Data
Data
File
Data Table
Impute
PCA
Data Table (1)

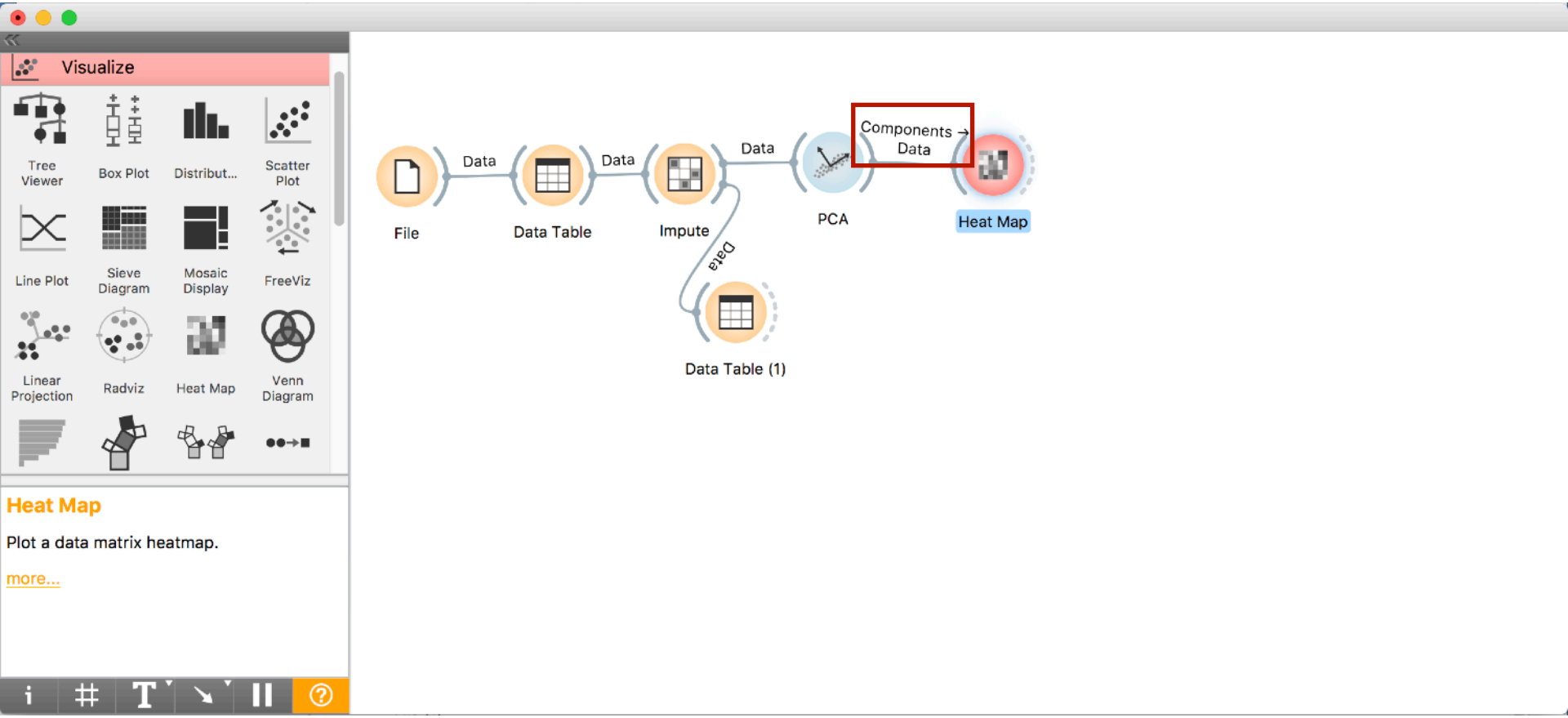

Unsupervised
Distance File
Distance Matrix
t-SN
Hierarchi... Clustering
k-Means
Louva Cluste
Manifold Learning
PCA
Corres Analy
PCA
Principal component analys
diagram.
more...
PCA
Components Selection
Components: 7
Explained variance: 90%
Options
Normalize variables
Show only first 20
Apply Automatically
cumulative variance
component variance
0.900
0.063
Proportion of variance
Principal Components
1
0.8
0.6
0.4
0.2
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

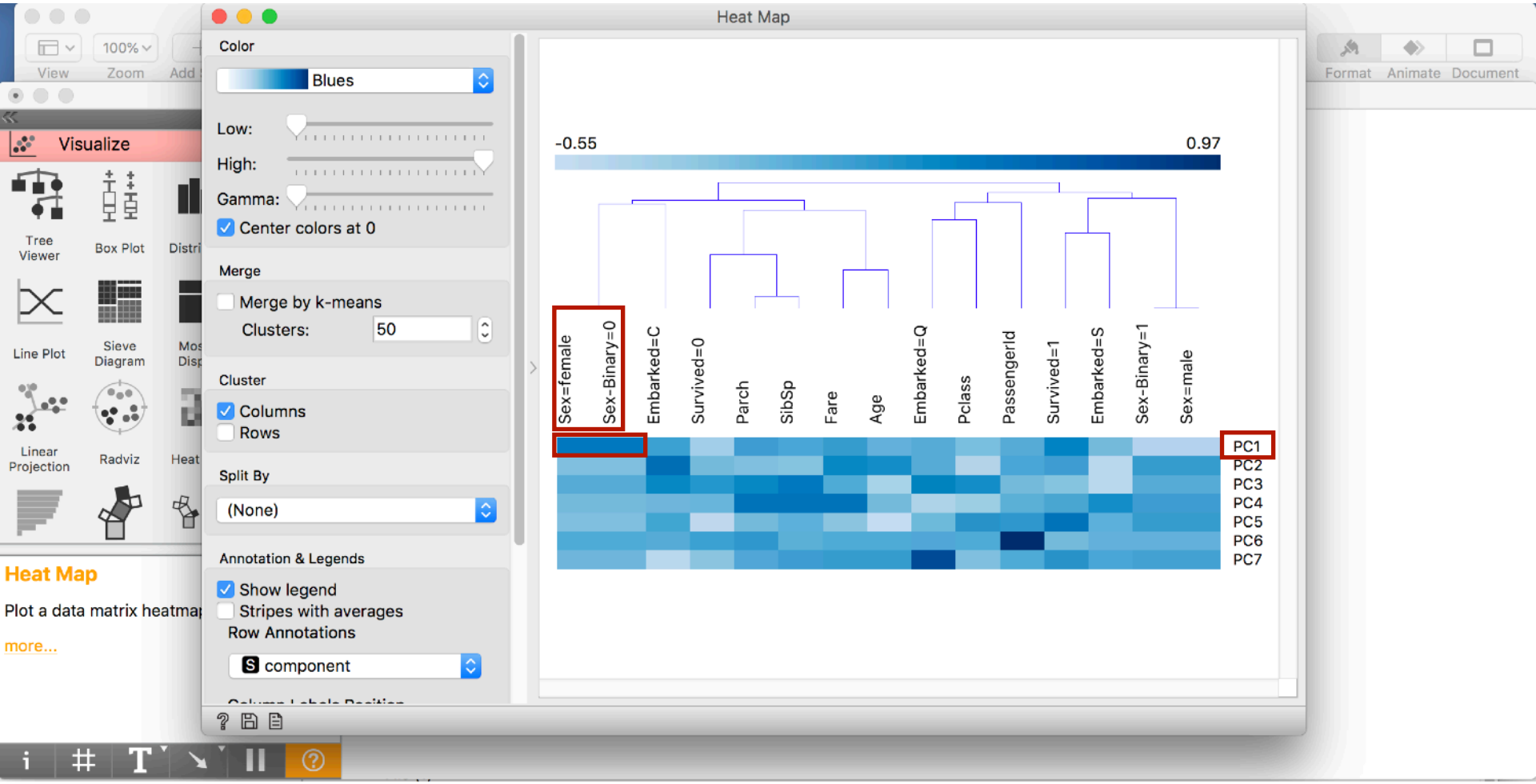

Visualize
Tree Viewer
Box Plot
Distribut...
Scatter Plot
Line Plot
Sieve Diagram
Mosaic Display
FreeViz
Linear Projection
Radviz
Heat Map
Venn Diagram
Select a widget to show its description.
See workflow examples, YouTube tutorials, or open the welcome screen.
Data
Data
Data
Data
File
Data Table
Impute
PCA
Heat Map
Data Table (1)

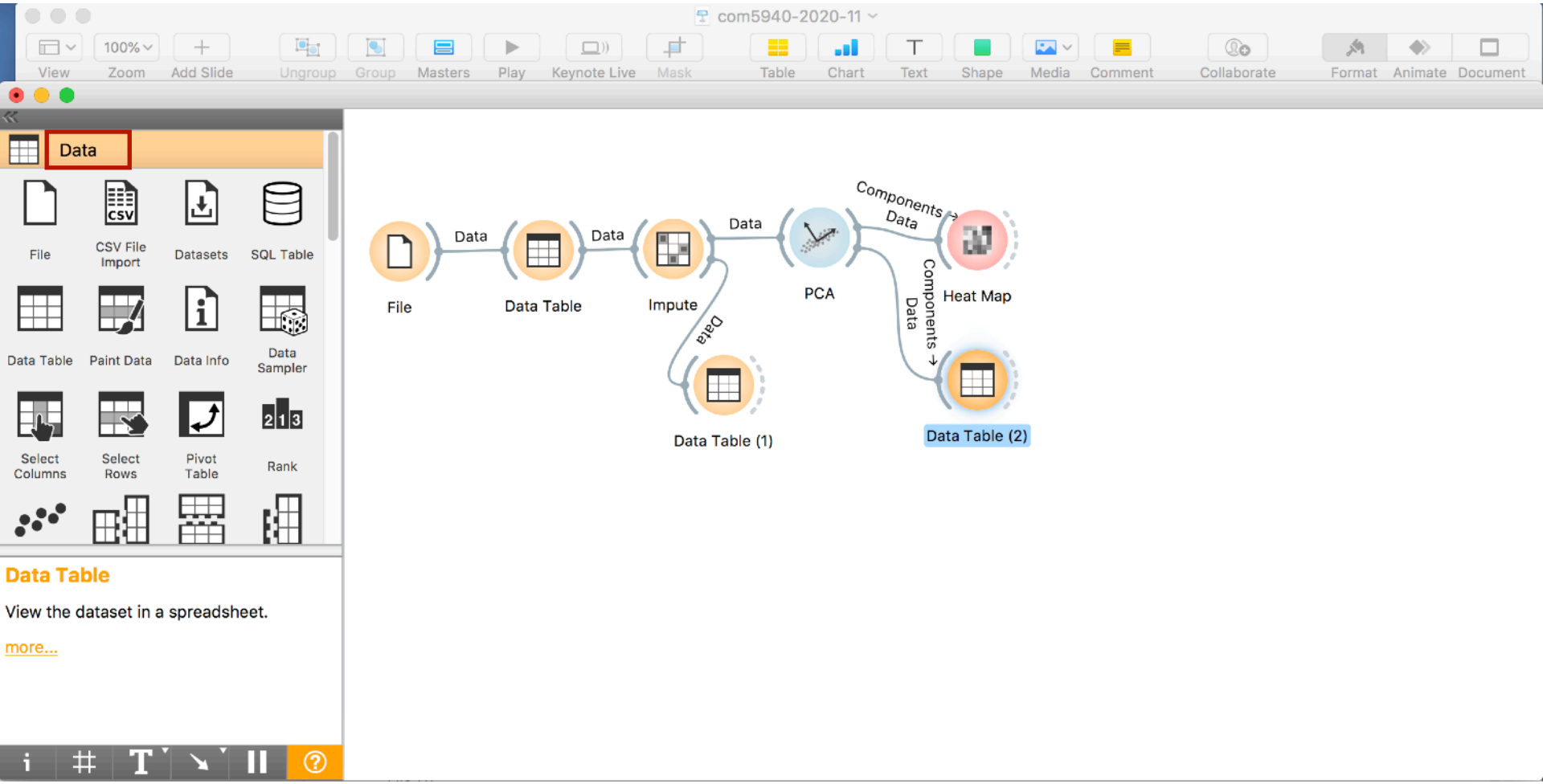

Visualize
Tree Viewer
Box Plot
Distribut...
Scatter Plot
Line Plot
Sieve Diagram
Mosaic Display
FreeViz
Linear Projection
Radviz
Heat Map
Venn Diagram
Heat Map
Plot a data matrix heatmap.
more...
Data
Data
Data
Data
File
Data Table
Impute
PCA
Heat Map
Data
Data Table (1)

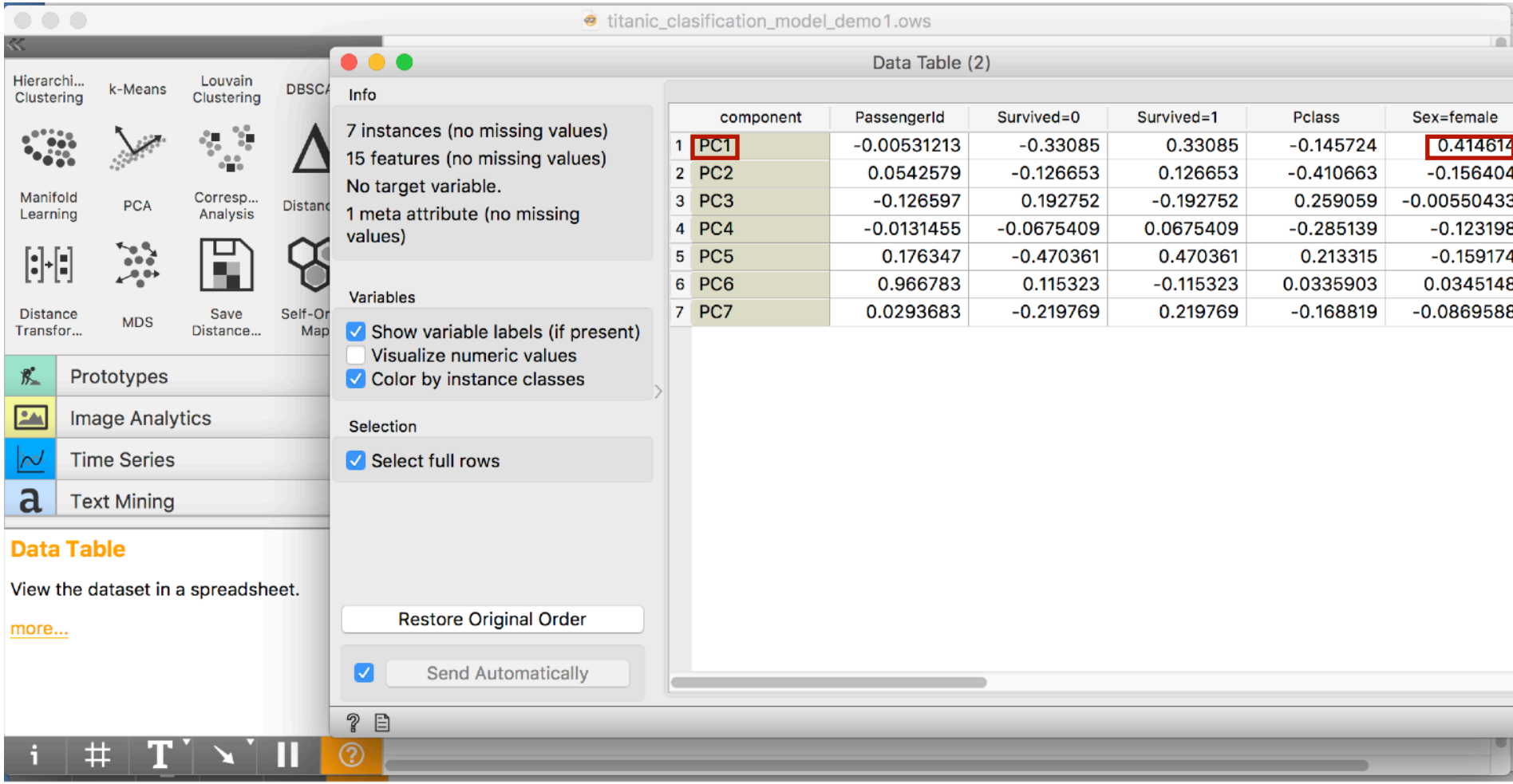

Visualize
Tree Viewer
Box Plot
Distribut...
Scatter Plot
Line Plot
Sieve Diagram
Mosaic Display
FreeViz
Linear Projection
Radviz
Heat Map
Venn Diagram
Heat Map
Plot a data matrix heatmap.
more...
Edit Links
Transformed Data
Data
Components
PCA
Data
PCA
Heat Map
Clear All
Cancel
OK

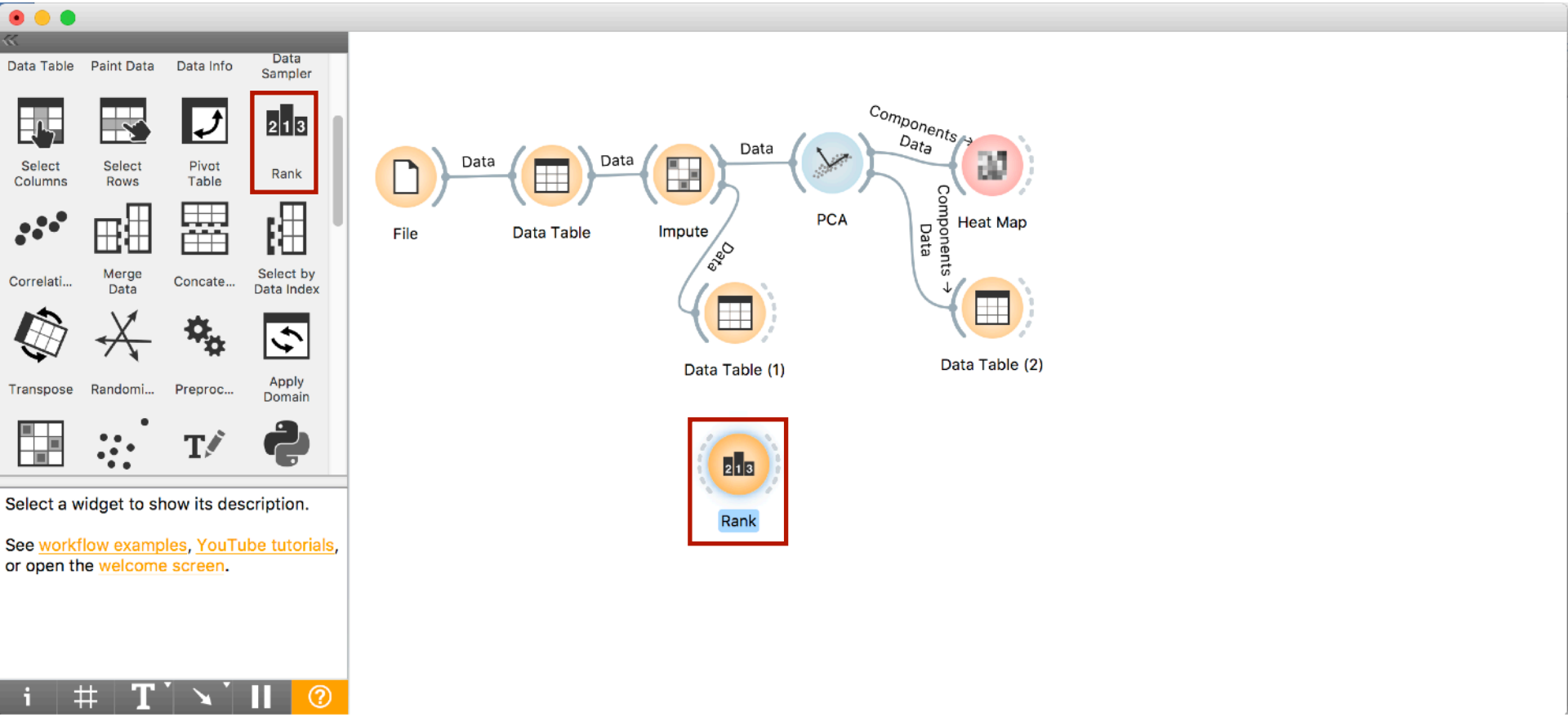

Visualize
Tree Viewer
Box Plot
Distribut...
Scatter Plot
Line Plot
Sieve Diagram
Mosaic Display
FreeViz
Linear Projection
Radviz
Heat Map
Venn Diagram
Heat Map
Plot a data matrix heatmap.
more...
Edit Links
Transformed Data
Data
Components
PCA
Data
PCA
Heat Map
Clear All
Cancel
OK

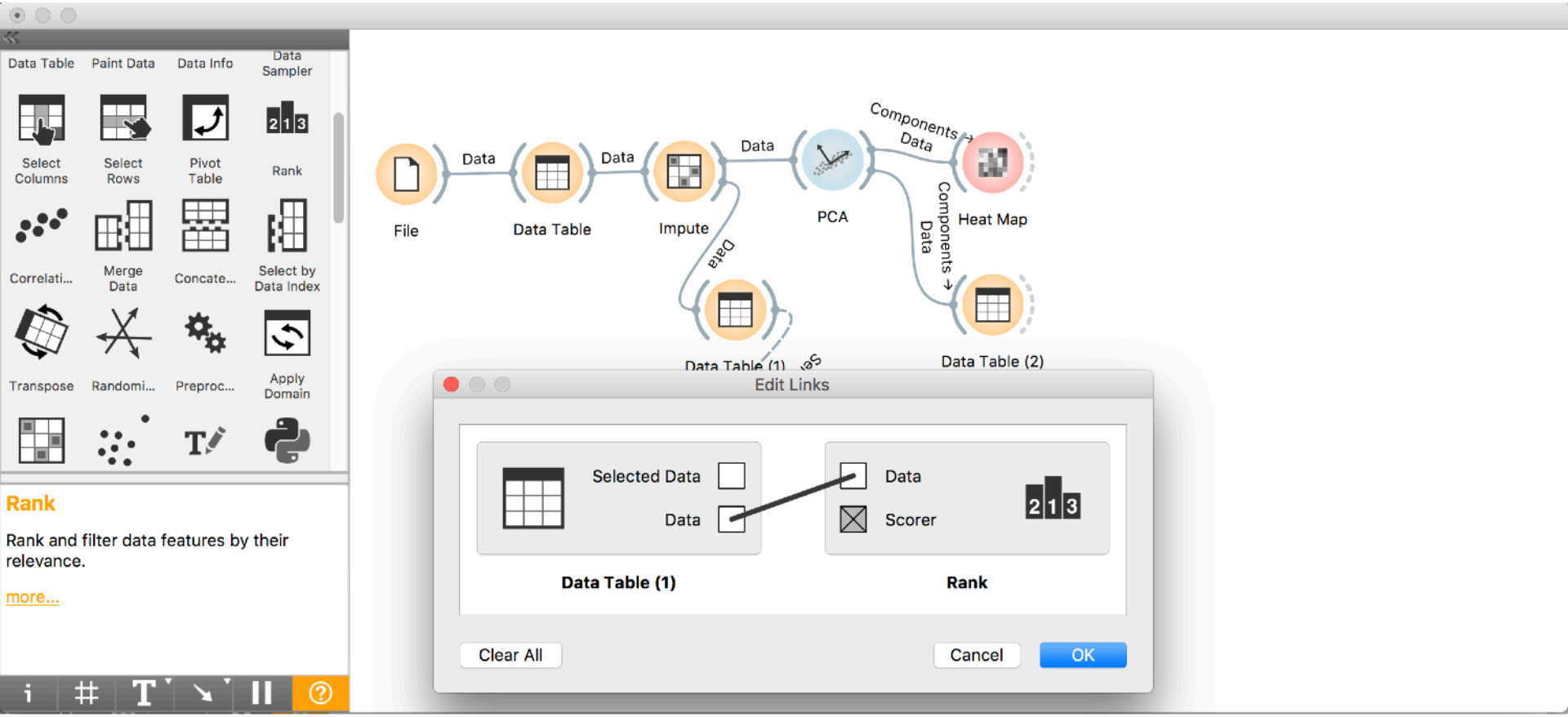

Visualize
Tree Viewer
Box Plot
Distribut...
Scatter Plot
Line Plot
Sieve Diagram
Mosaic Display
FreeViz
Linear Projection
Radviz
Heat Map
Venn Diagram
Heat Map
Plot a data matrix heatmap.
more...
Data
Data
Data
Components
Data
File
Data Table
Impute
PCA
Heat Map
Data
Data Table (1)

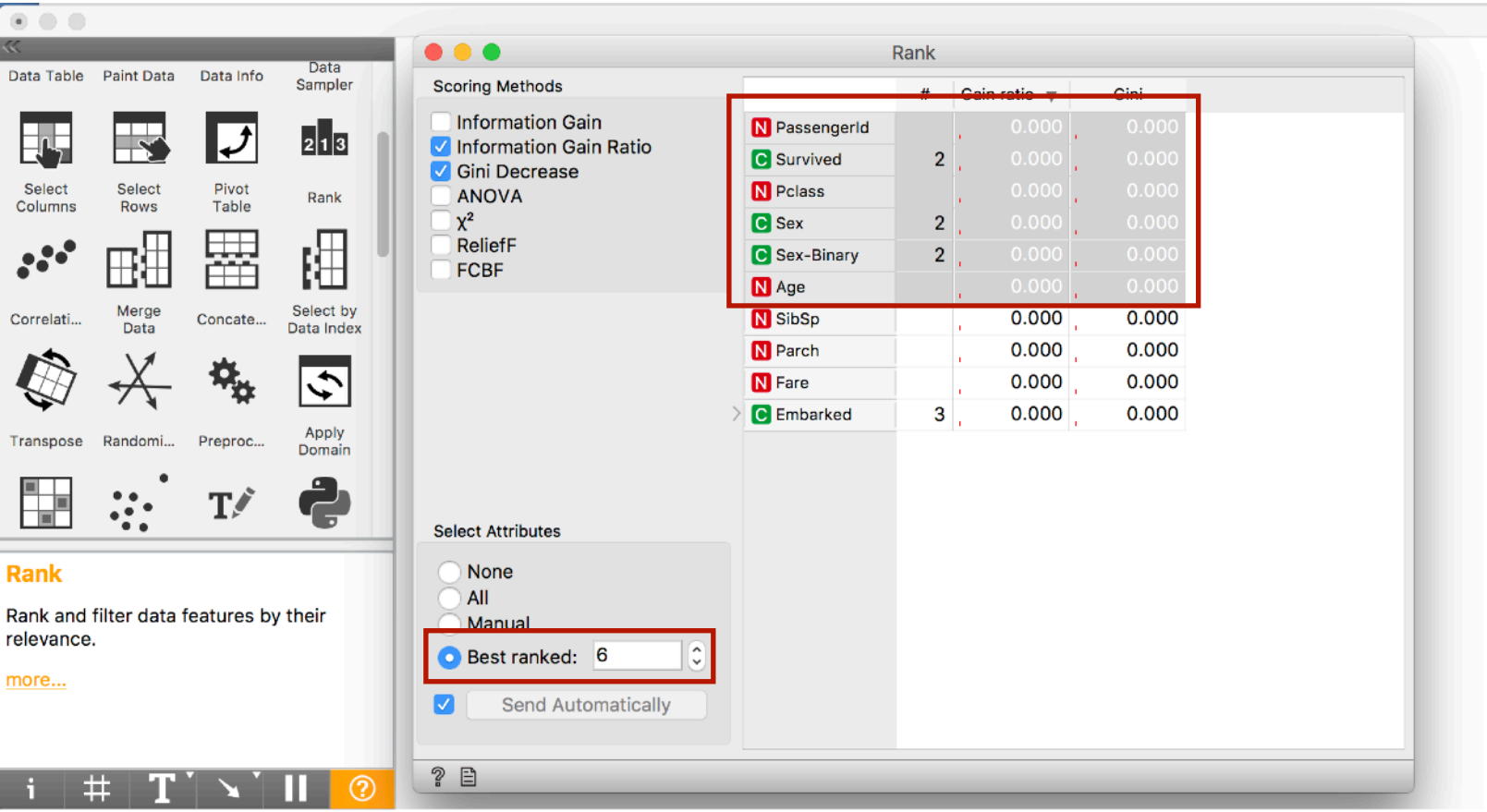

Heat Map
Color
Blues
Low:
High:
Gamma:
Center colors at 0
Merge
Merge by k-means
Clusters:
50
Cluster
Columns
Rows
Split By
(None)
Annotation & Legends
Show legend
Stripes with averages
Row Annotations
component
-0.55
0.97
Sex=female
Sex-Binary=0
Embarked=C
Survived=0
Parch
SibSp
Fare
Age
Embarked=Q
Pclass
PassengerId
Survived=1
Embarked=S
Sex-Binary=1
Sex=male
PC1
PC2
PC3
PC4
PC5
PC6
PC7
Visualize
Tree Viewer
Box Plot
Line Plot
Sieve Diagram
Linear Projection
Radviz
Heat Map
Plot a data matrix heatmap
more...
Format
Animate
Document

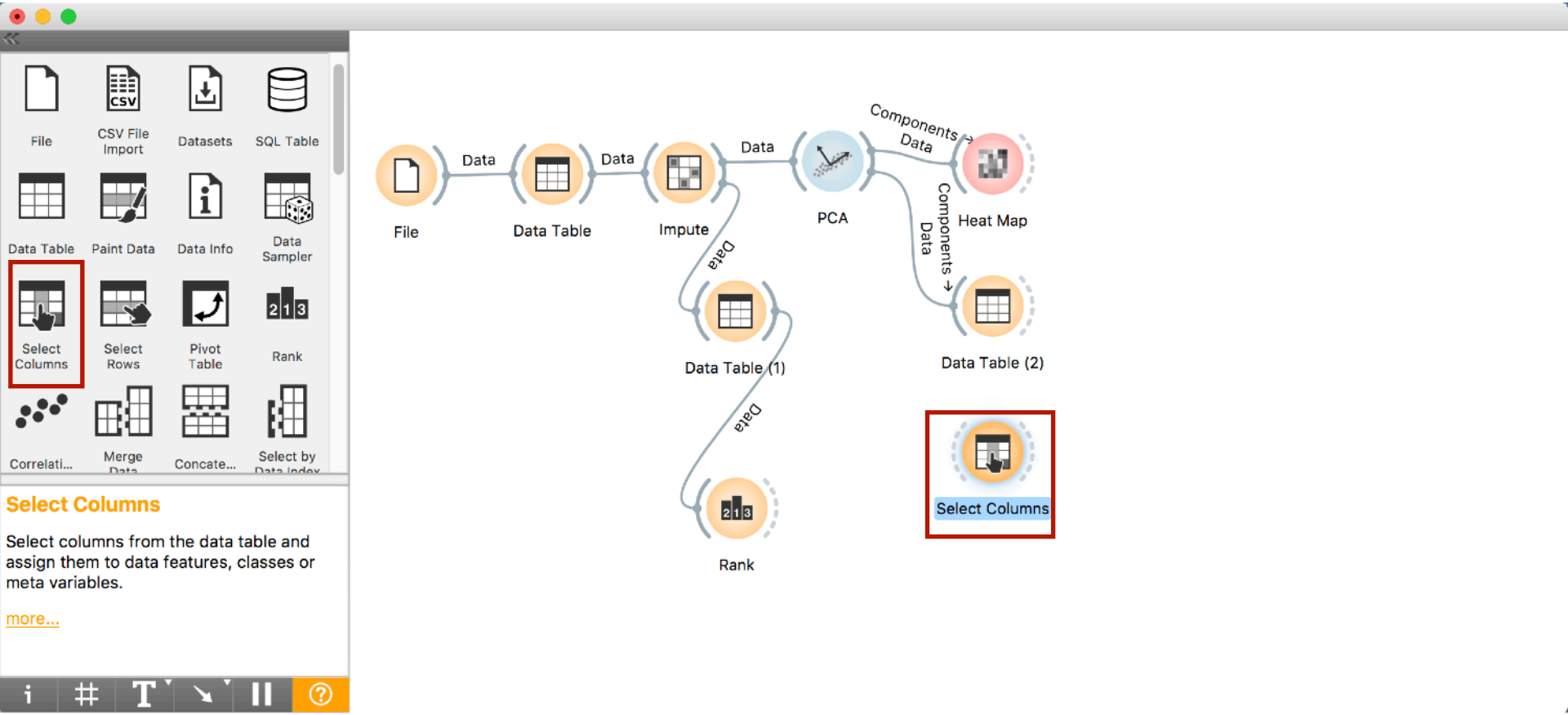

com5940-2020-11
View
Zoom
Add Slide
Ungroup
Group
Masters
Play
Keynote Live
Mask
Table
Chart
Text
Shape
Media
Comment
Collaborate
Format
Animate
Document
Data
File
CSV File Import
Datasets
SQL Table
Data Table
Paint Data
Data Info
Data Sampler
Select Columns
Select Rows
Pivot Table
Rank
Data Table
View the dataset in a spreadsheet.
more...
File
Data
Data Table
Data
Impute
Data
PCA
Components
Data
Heat Map
Components
Data
Data
Data Table (1)
Data Table (2)

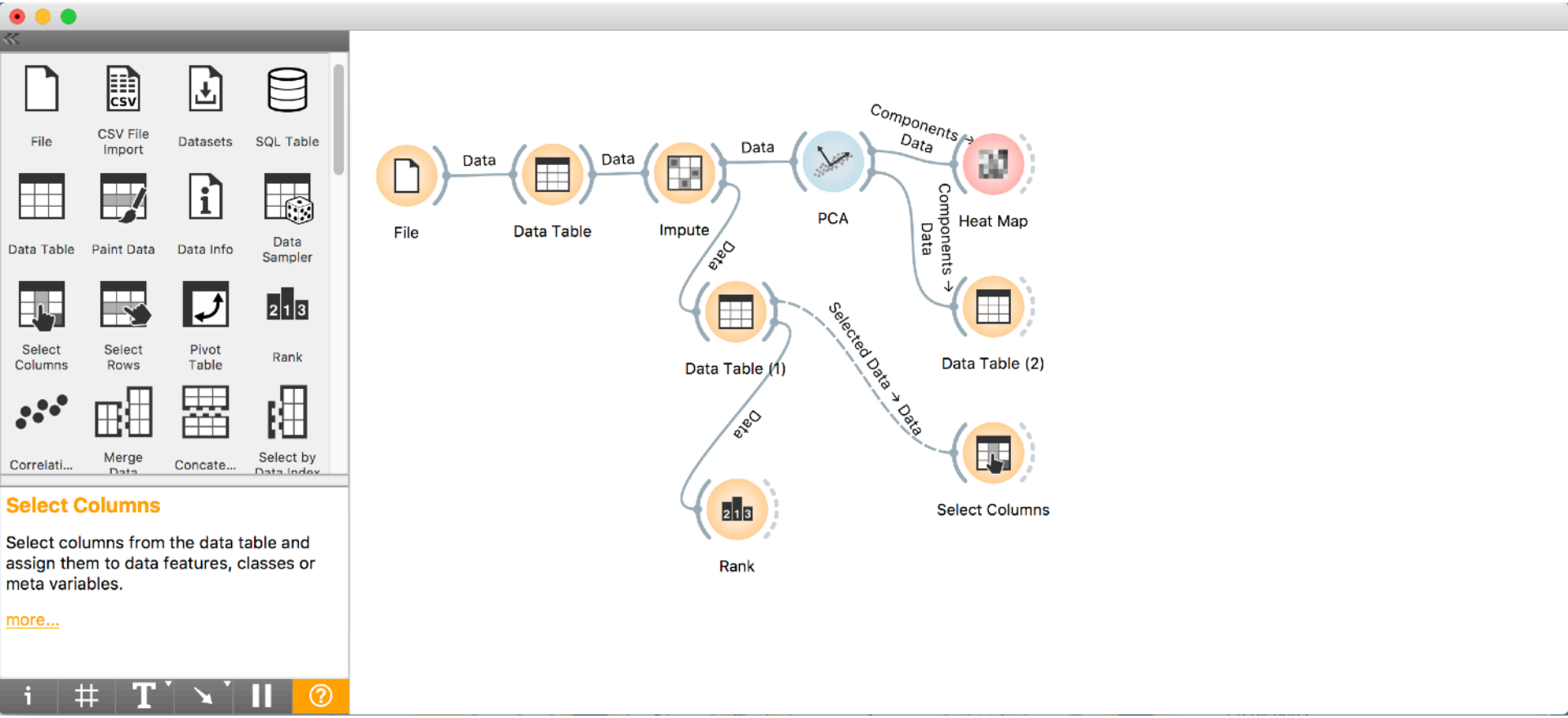

titanic_clasification_model_demo1.ows
Hierarchi... Clustering
k-Means
Louvain Clustering
DBSCA
Manifold Learning
PCA
Corresp... Analysis
Distanc
Distance Transfor...
MDS
Save Distance...
Self-Or Map
Prototypes
Image Analytics
Time Series
Text Mining
Data Table
View the dataset in a spreadsheet.
more...
Data Table (2)
Info
7 instances (no missing values)
15 features (no missing values)
No target variable.
1 meta attribute (no missing values)
Variables
Show variable labels (if present)
Visualize numeric values
Color by instance classes
Selection
Select full rows
Restore Original Order
Send Automatically
component PassengerId Survived=0 Survived=1 Pclass Sex=female
1 PC1 -0.00531213 -0.33085 0.33085 -0.145724 0.414614
2 PC2 0.0542579 -0.126653 0.126653 -0.410663 -0.156404
3 PC3 -0.126597 0.192752 -0.192752 0.259059 -0.00550433
4 PC4 -0.0131455 -0.0675409 0.0675409 -0.285139 -0.123198
5 PC5 0.176347 -0.470361 0.470361 0.213315 -0.159174
6 PC6 0.966783 0.115323 -0.115323 0.0335903 0.0345148
7 PC7 0.0293683 -0.219769 0.219769 -0.168819 -0.0869588

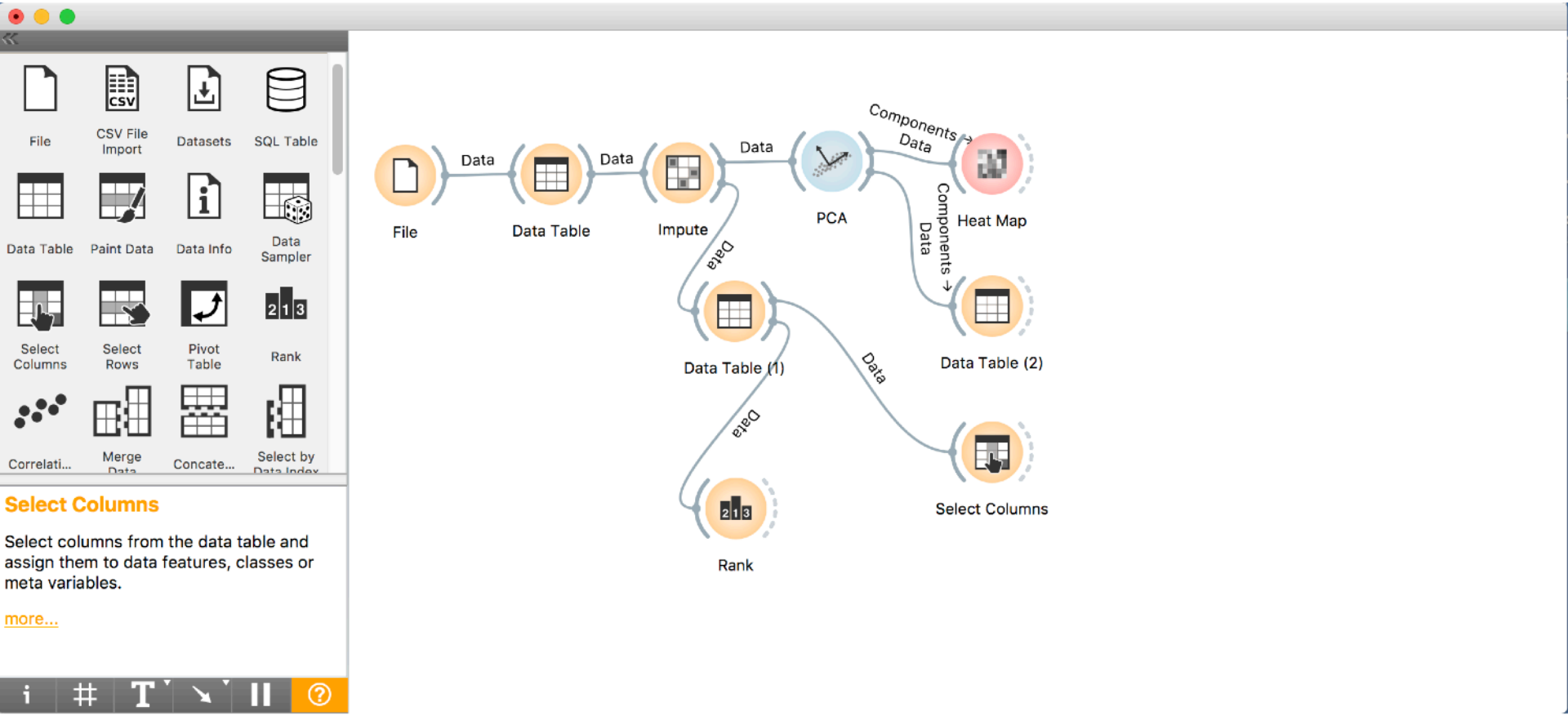

Data Table
Paint Data
Data Info
Data Sampler
Select Columns
Select Rows
Pivot Table
Rank
Correlati...
Merge Data
Concate...
Select by Data Index
Transpose
Randomi...
Preproc...
Apply Domain
Select a widget to show its description.
See workflow examples, YouTube tutorials, or open the welcome screen.
File
Data
Data Table
Data
Impute
Data
PCA
Components
Data
Heat Map
Components
Data
Data Table (2)
Data
Data Table (1)
Rank

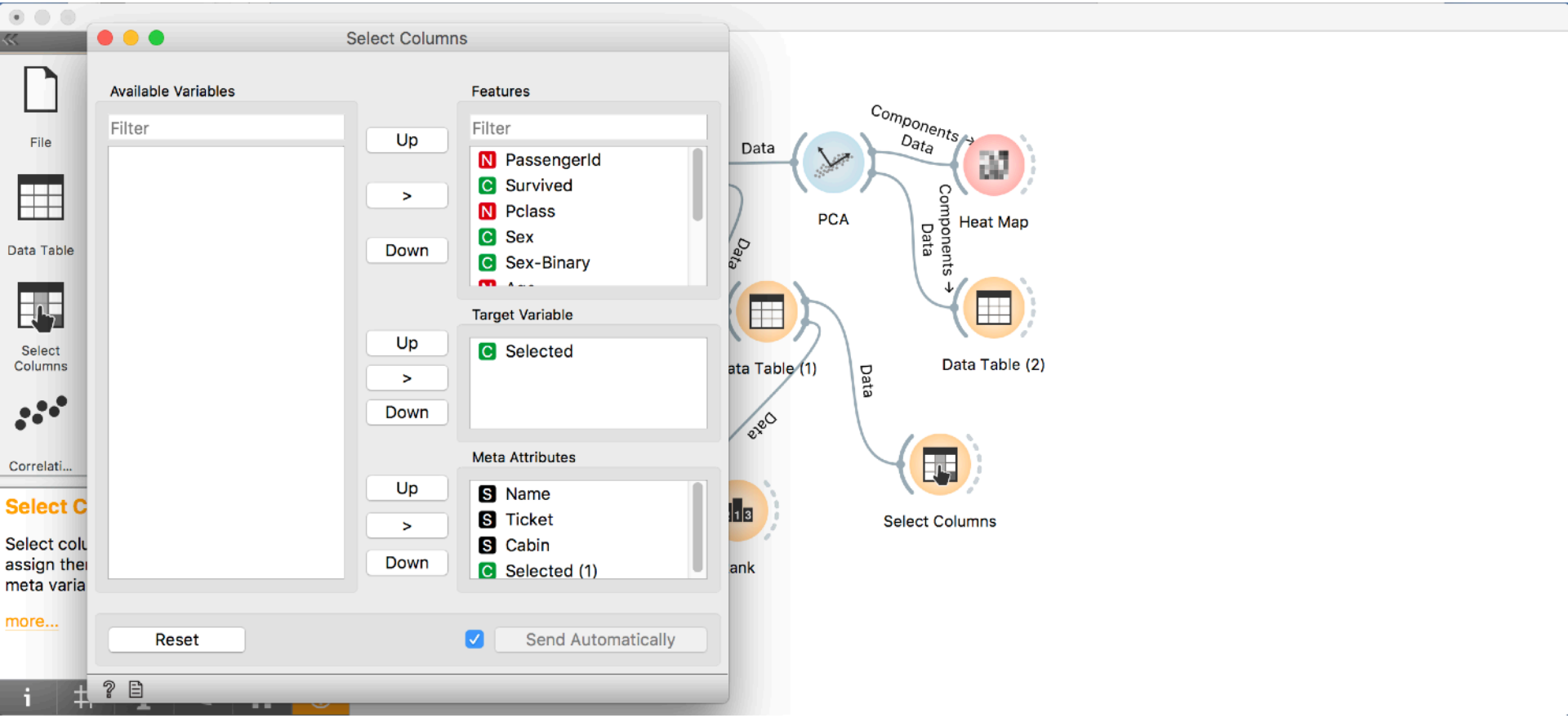

Data Table
Paint Data
Data Info
Data Sampler
Select Columns
Select Rows
Pivot Table
Rank
Correlati...
Merge Data
Concate...
Select by Data Index
Transpose
Randomi...
Preproc...
Apply Domain
Rank
Rank and filter data features by their relevance.
more...
File
Data
Data Table
Impute
PCA
Heat Map
Data Table (1)
Data Table (2)
Components
Edit Links
Selected Data
Scorer
Data Table (1)
Rank
Clear All
Cancel
OK

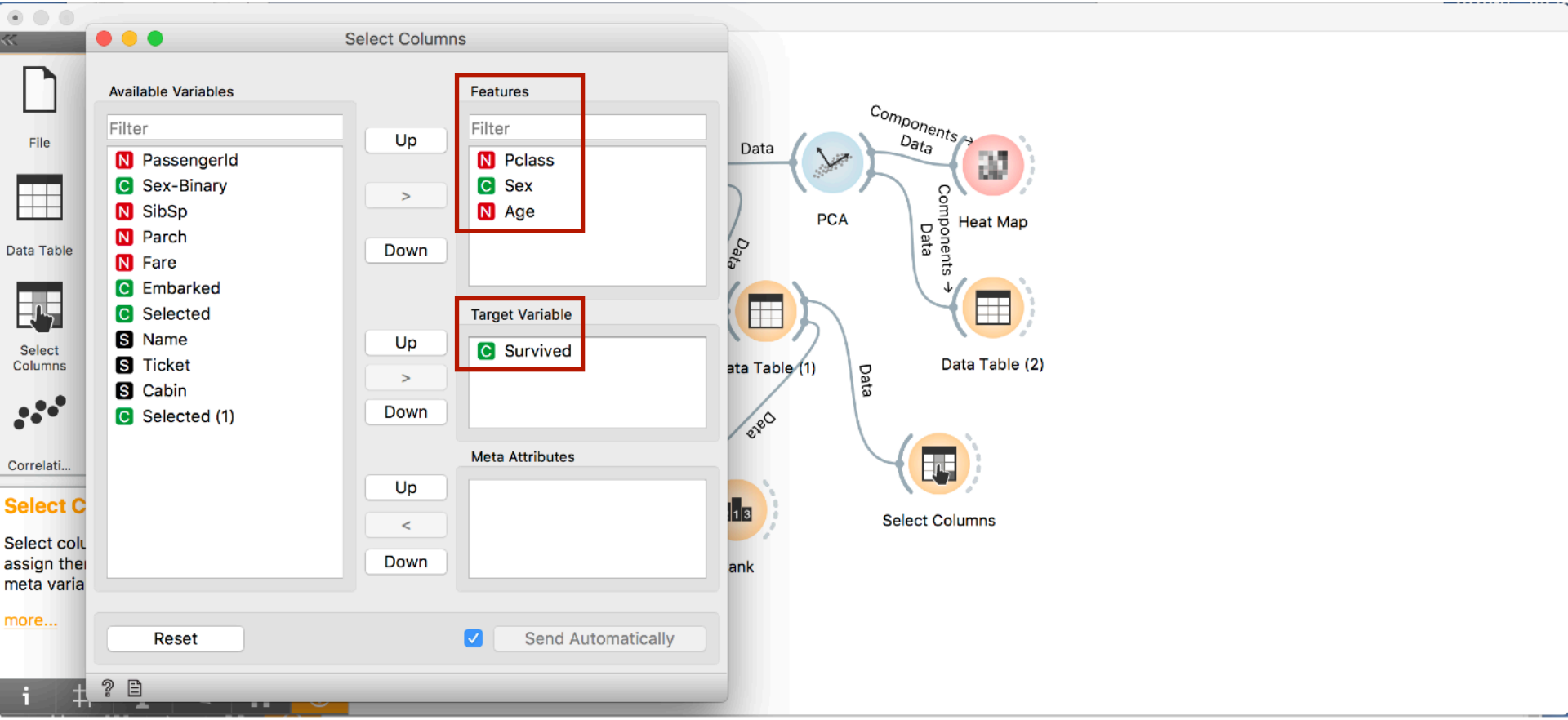

Data Table
Paint Data
Data Info
Data Sampler
Select Columns
Select Rows
Pivot Table
Rank
Correlati...
Merge Data
Concate...
Select by Data Index
Transpose
Randomi...
Preproc...
Apply Domain
Rank
Rank and filter data features by their relevance.
more...
Rank
Scoring Methods
Information Gain
Information Gain Ratio
Gini Decrease
ANOVA
χ²
ReliefF
FCBF
Gain ratio Gini
PassengerId 0.000 0.000
Survived 2 0.000 0.000
Pclass 0.000 0.000
Sex 2 0.000 0.000
Sex-Binary 2 0.000 0.000
Age 0.000 0.000
SibSp 0.000 0.000
Parch 0.000 0.000
Fare 0.000 0.000
Embarked 3 0.000 0.000
Select Attributes
None
All
Manual
Best ranked: 6
Send Automatically

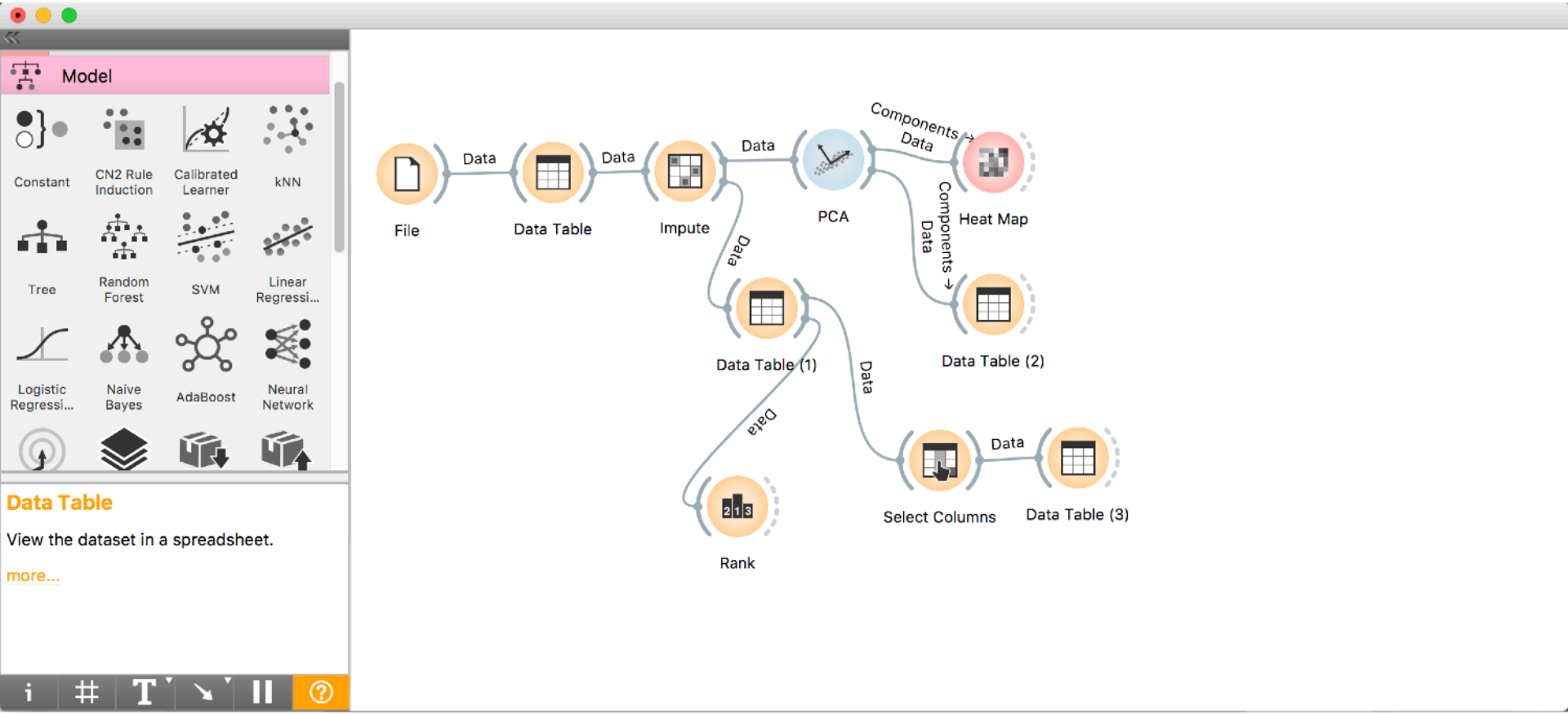

File
CSV File Import
Datasets
SQL Table
Data Table
Paint Data
Data Info
Data Sampler
Select Columns
Select Rows
Pivot Table
Rank
Correlati...
Merge Data
Concate...
Select by Data Index
Select Columns
Select columns from the data table and assign them to data features, classes or meta variables.
more...
Data
Data
Data
Components
Data
Components
Data
File
Data Table
Impute
PCA
Heat Map
Data
Data Table (1)
Data Table (2)
Data
Rank
Select Columns

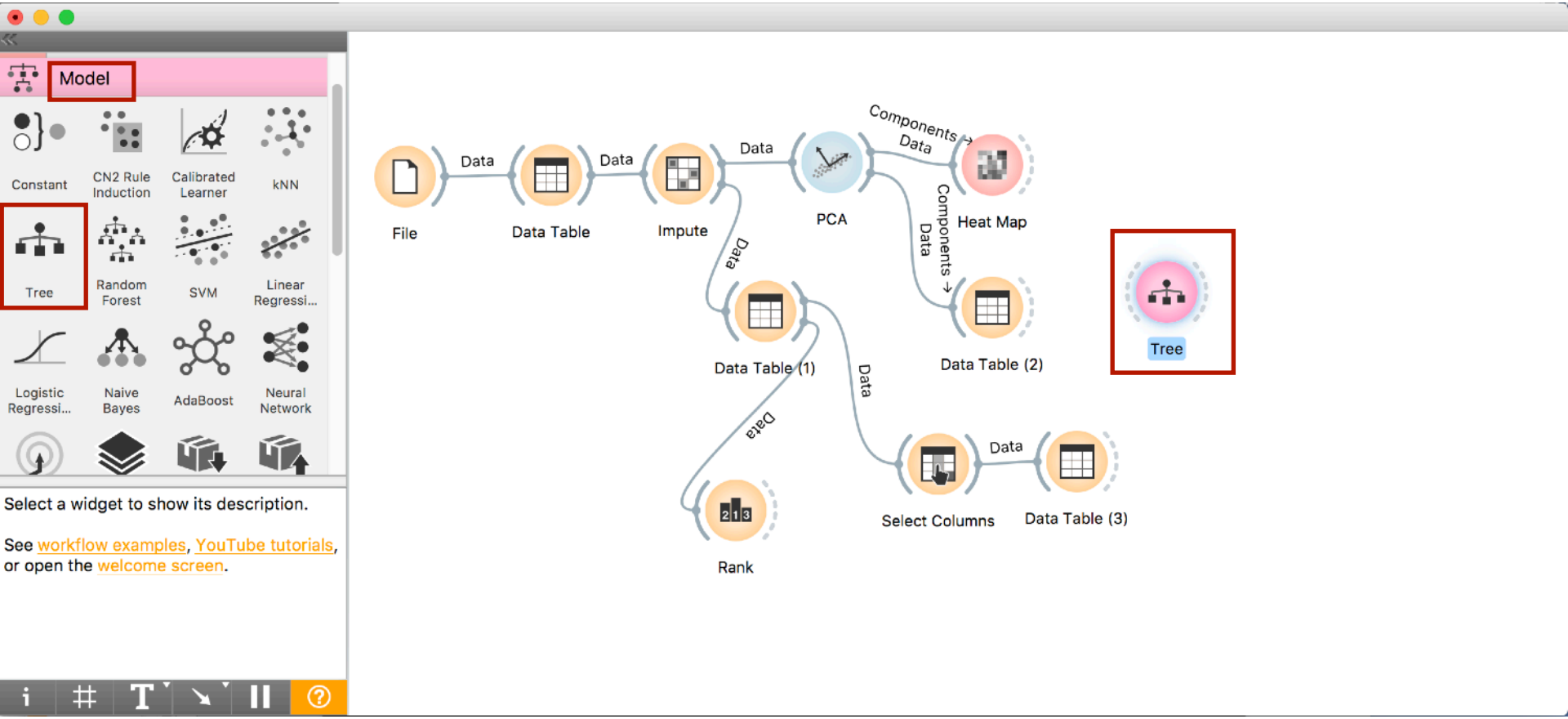

File
CSV File Import
Datasets
SQL Table
Data Table
Paint Data
Data Info
Data Sampler
Select Columns
Select Rows
Pivot Table
Rank
Correlati...
Merge Data
Concate...
Select by Data Index
Select Columns
Select columns from the data table and assign them to data features, classes or meta variables.
more...
Data
Data
Data
Components → Data
Components → Data
Data
Selected Data → Data
Data
File
Data Table
Impute
PCA
Heat Map
Data Table (1)
Data Table (2)
Rank
Select Columns

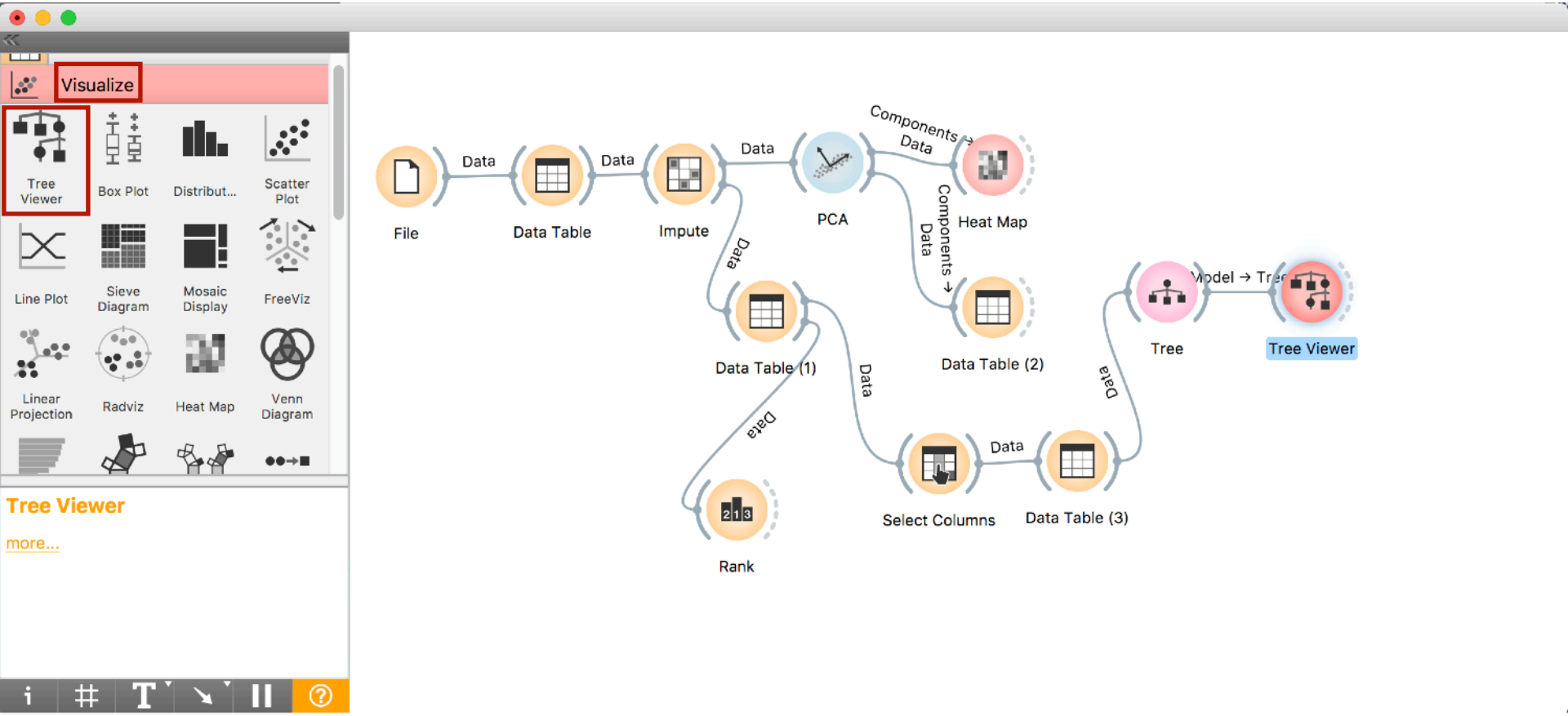

File
CSV File Import
Datasets
SQL Table
Data Table
Paint Data
Data Info
Data Sampler
Select Columns
Select Rows
Pivot Table
Rank
Correlati...
Merge Data
Concate...
Select by Data Index
Select Columns
Select columns from the data table and assign them to data features, classes or meta variables.
more...
File
Data
Data Table
Data
Impute
Data
PCA
Components
Data
Heat Map
Components
Data
Data Table (2)
Data
Data Table (1)
Data
Select Columns
Data
Rank

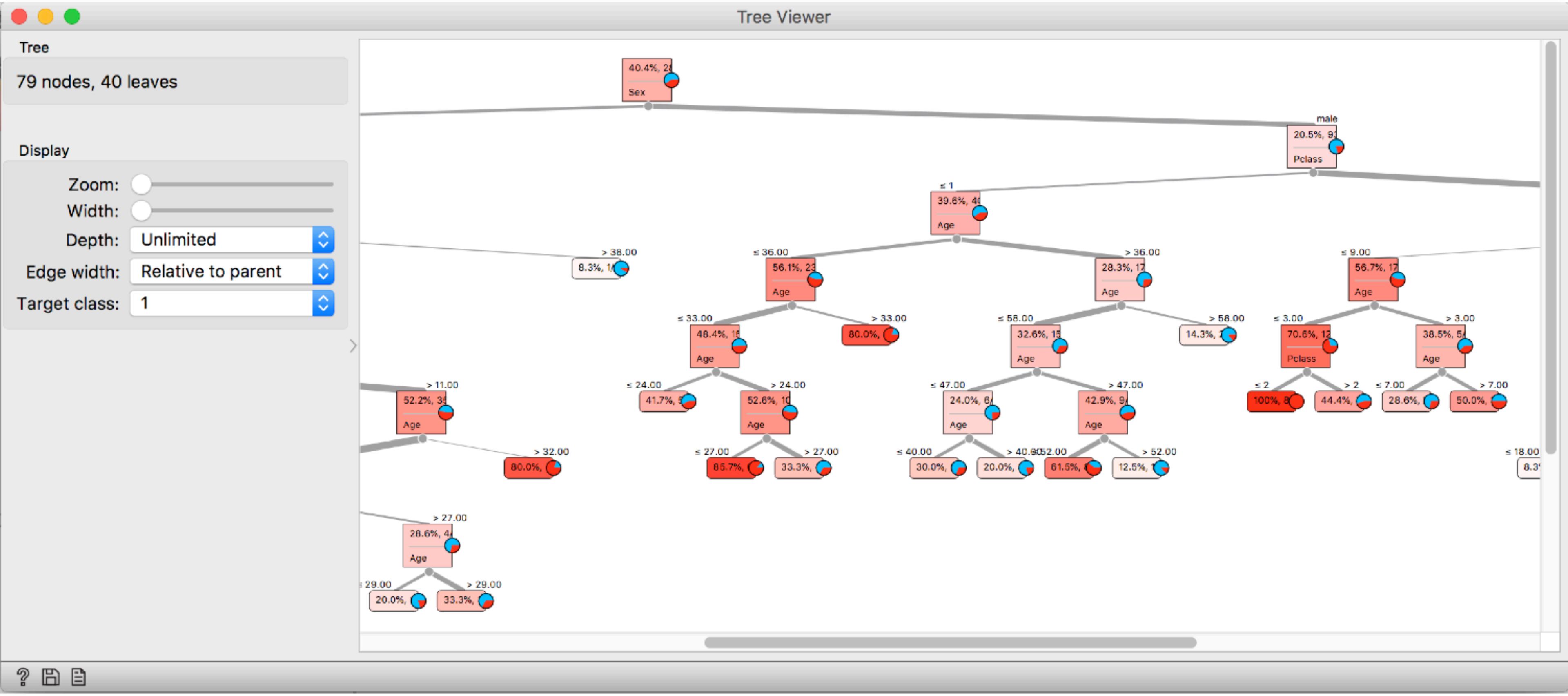

Select Columns
Available Variables
Filter
Up
>
Down
Features
Filter
PassengerId
Survived
Pclass
Sex
Sex-Binary
Target Variable
Selected
Meta Attributes
Name
Ticket
Cabin
Selected (1)
Reset
Send Automatically
File
Data Table
Select Columns
Correlati...
Select C
Select colu
assign the
meta varia
more...
Data
PCA
Components
Heat Map
Data Table (1)
Data Table (2)
Select Columns


Select Columns
Available Variables
Filter
PassengerId
Sex-Binary
SibSp
Parch
Fare
Embarked
Selected
Name
Ticket
Cabin
Selected (1)
Up
>
Down
Features
Filter
Pclass
Sex
Age
Target Variable
Survived
Meta Attributes
Reset
Send Automatically
File
Data Table
Select Columns
Correlati...
Select C
Select colu
assign the
meta varia
more...
Data
PCA
Components
Heat Map
Data Table (1)
Data Table (2)
Select Columns

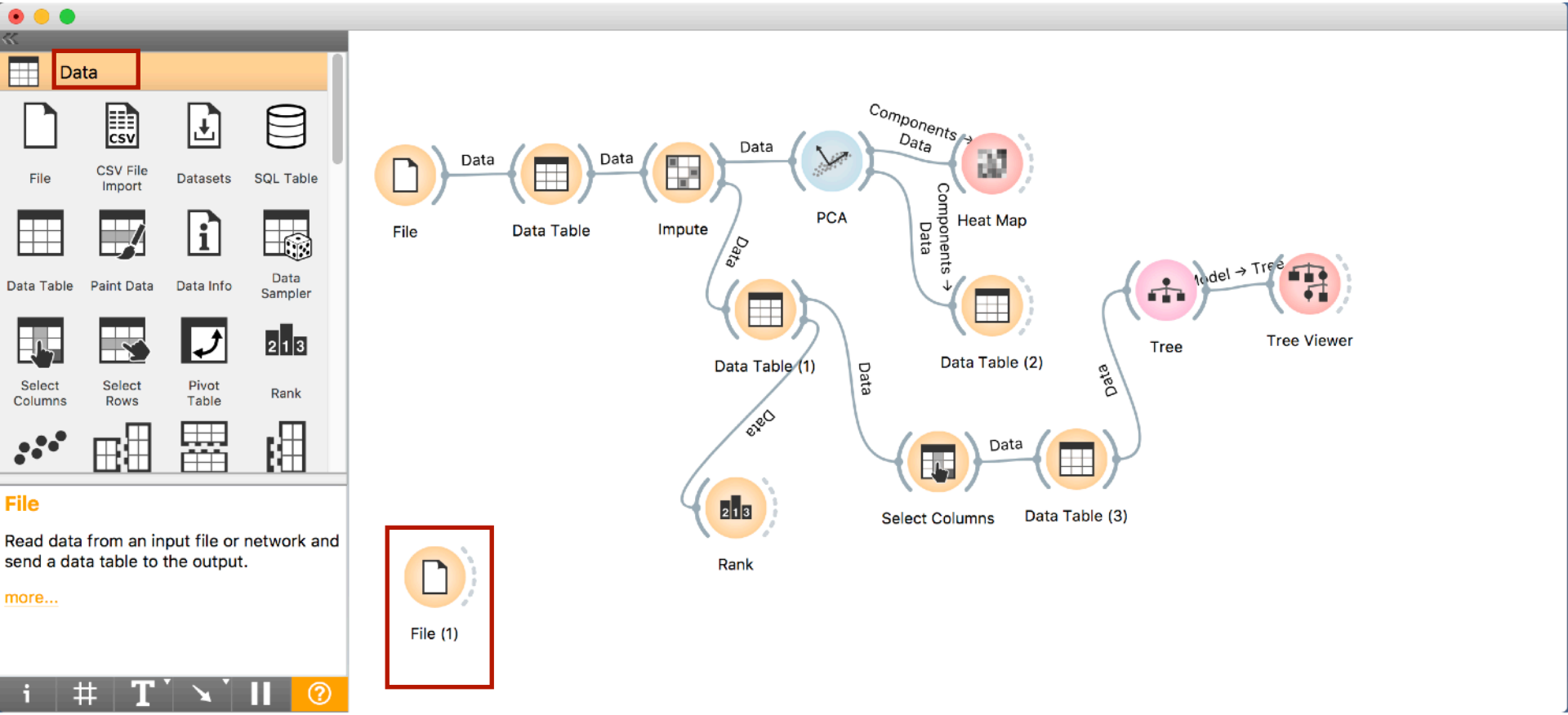

Model
Constant
CN2 Rule Induction
Calibrated Learner
kNN
Tree
Random Forest
SVM
Linear Regressi...
Logistic Regressi...
Naive Bayes
AdaBoost
Neural Network
Data Table
View the dataset in a spreadsheet.
more...
File
Data
Data Table
Data
Impute
Data
PCA
Components
Data
Heat Map
Components
Data
Data Table (2)
Data
Data Table (1)
Data
Rank
Data
Select Columns
Data
Data Table (3)

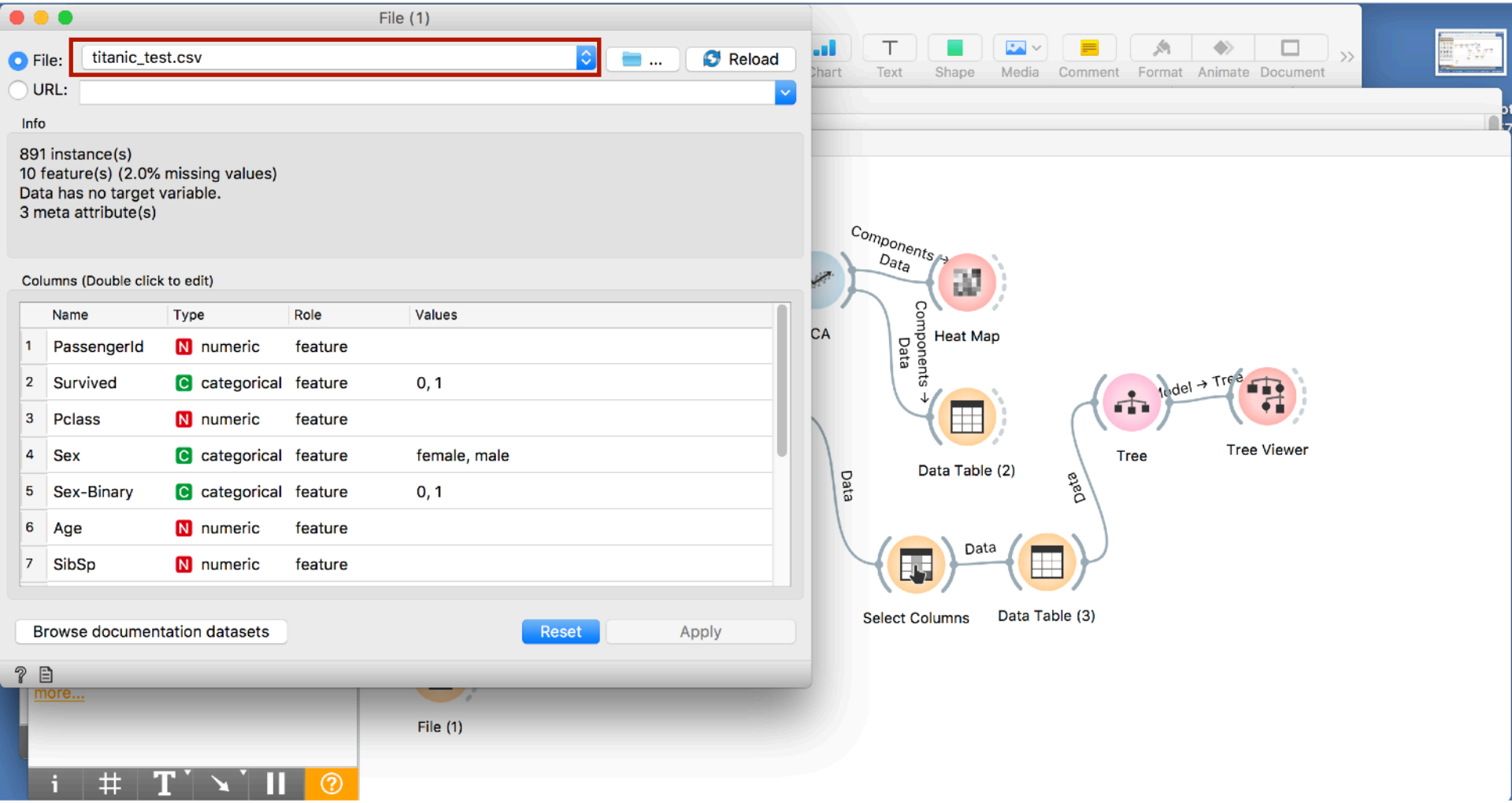

Model
Constant
CN2 Rule Induction
Calibrated Learner
kNN
Tree
Random Forest
SVM
Linear Regressi...
Logistic Regressi...
Naive Bayes
AdaBoost
Neural Network
Select a widget to show its description.
See workflow examples, YouTube tutorials, or open the welcome screen.
File
Data
Data Table
Data
Impute
Data
PCA
Components
Data
Heat Map
Components
Data
Data Table (2)
Data
Data Table (1)
Data
Rank
Data
Select Columns
Data
Data Table (3)
Tree

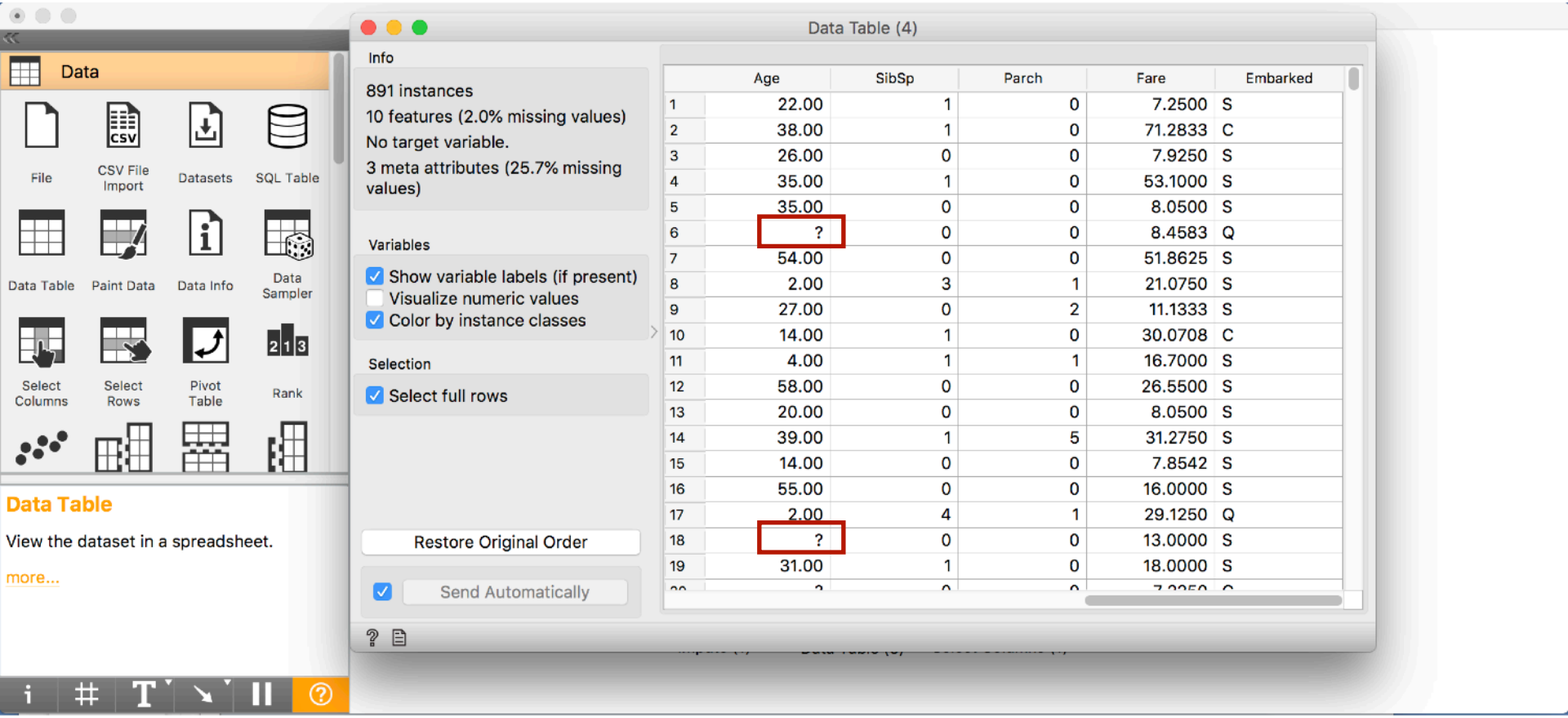

Visualize
Tree Viewer
Box Plot
Distribut...
Scatter Plot
Line Plot
Sieve Diagram
Mosaic Display
FreeViz
Linear Projection
Radviz
Heat Map
Venn Diagram
Tree Viewer
more...
File
Data
Data Table
Impute
PCA
Components
Heat Map
Data Table (1)
Data Table (2)
Rank
Select Columns
Data Table (3)
Tree
Tree Viewer

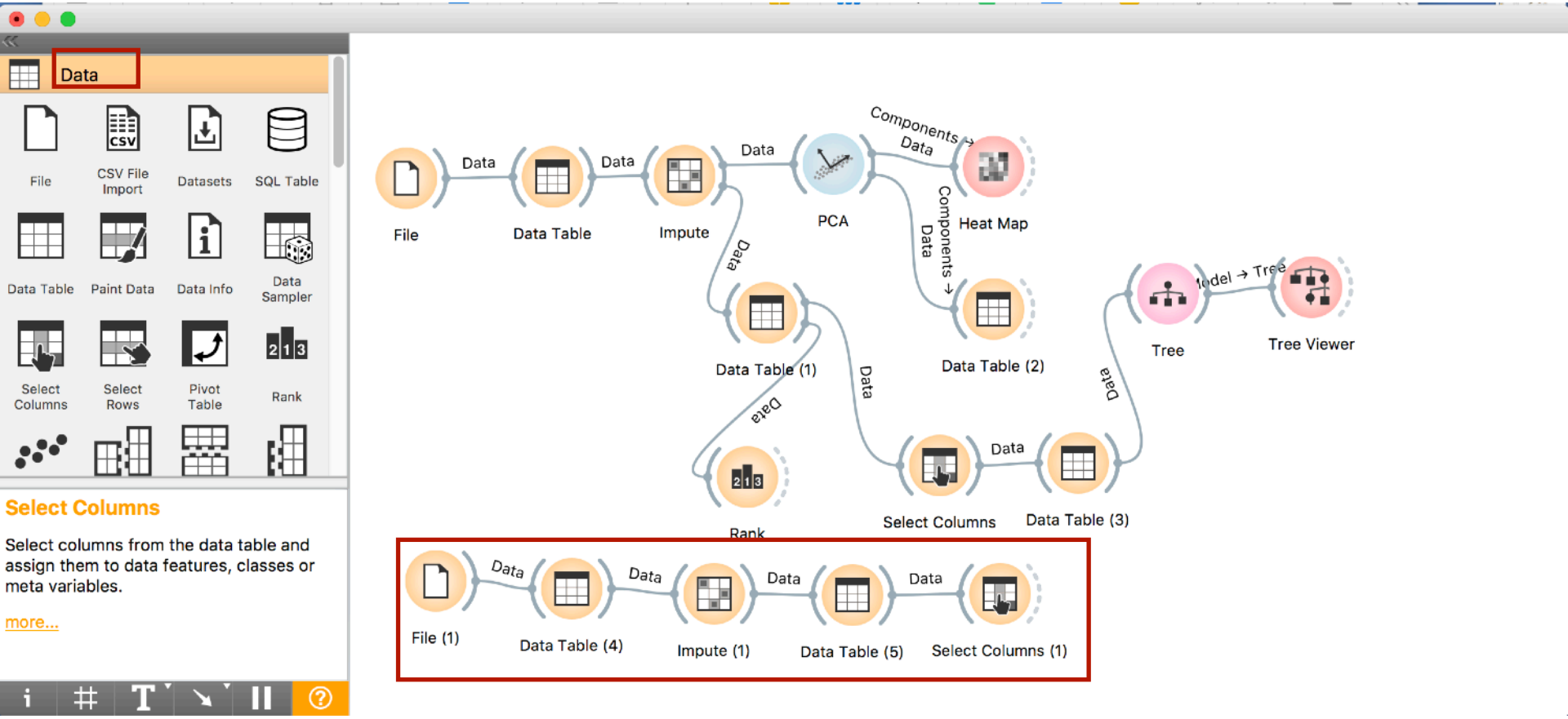

Tree Viewer
Tree
79 nodes, 40 leaves
Display
Zoom:
Width:
Depth:
Unlimited
Edge width:
Relative to parent
Target class:
1

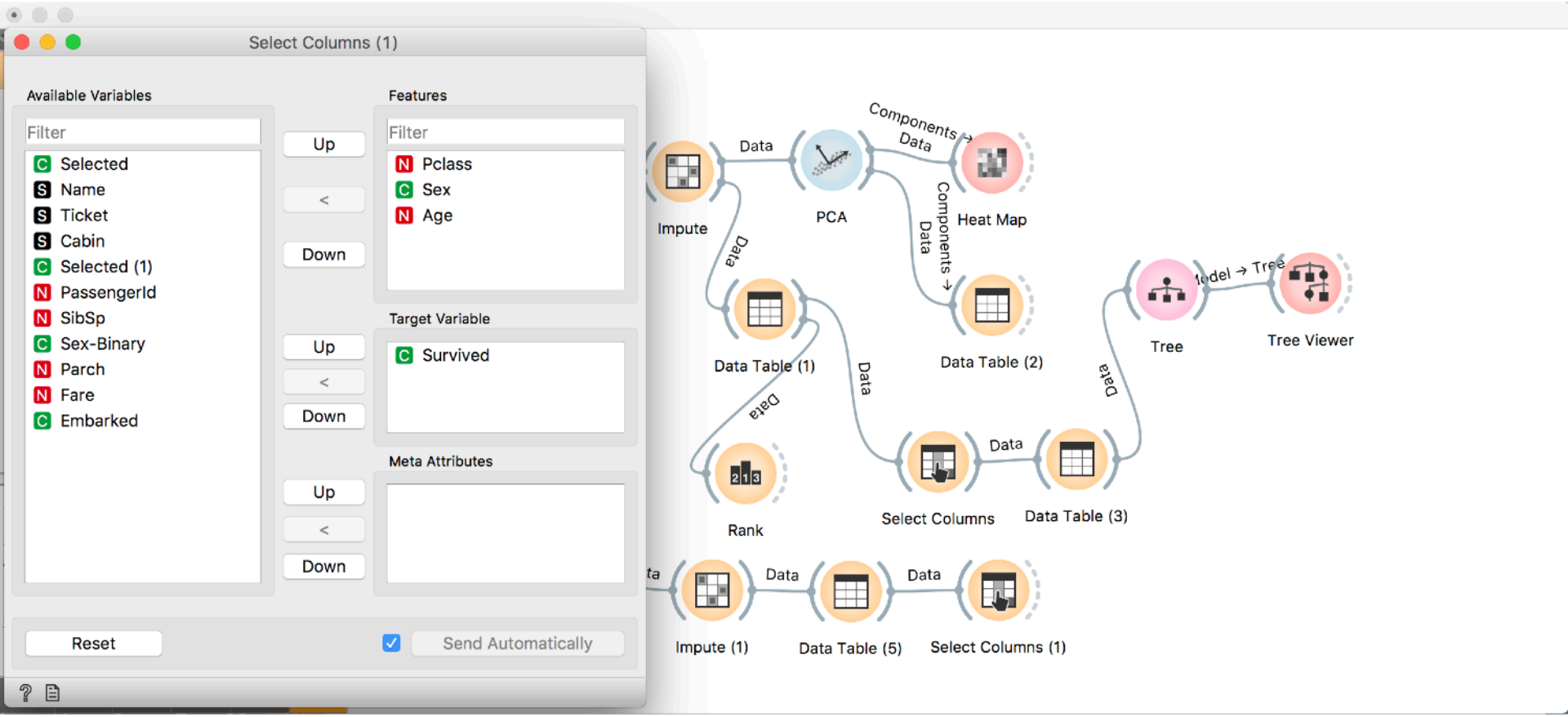
After training the model, it's time for testing it.

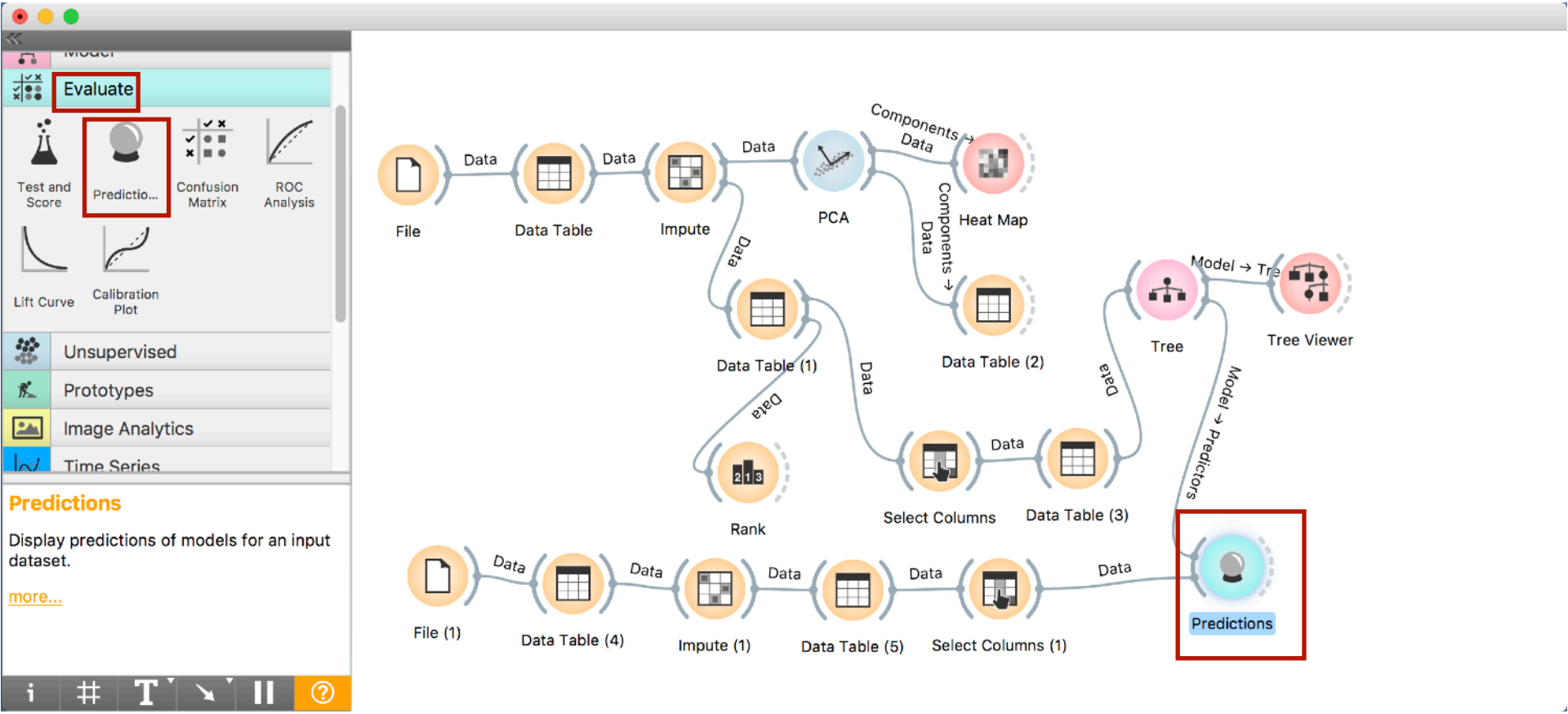

Data
File
CSV File Import
Datasets
SQL Table
Data Table
Paint Data
Data Info
Data Sampler
Select Columns
Select Rows
Pivot Table
Rank
File
Read data from an input file or network and send a data table to the output.
more...
File (1)
Data
Data
Data
Components → Data
Components → Data
Data
Data
Data
Data
Data
Model → Tree
File
Data Table
Impute
PCA
Heat Map
Data Table (1)
Data Table (2)
Tree
Tree Viewer
Rank
Select Columns
Data Table (3)

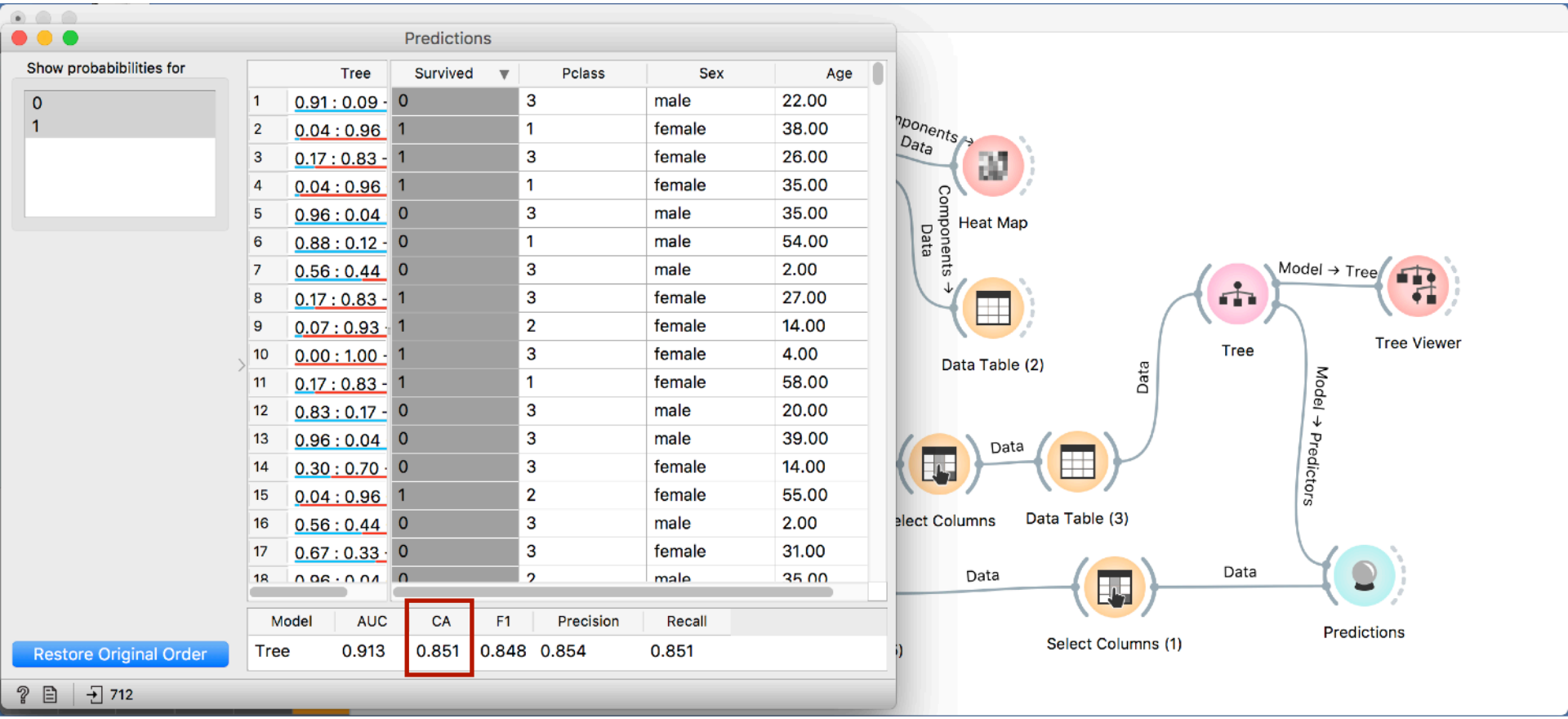

File (1)
File:
titanic_test.csv
...
Reload
URL:
Info
891 instance(s)
10 feature(s) (2.0% missing values)
Data has no target variable.
3 meta attribute(s)
Columns (Double click to edit)
Name
Type
Role
Values
1 PassengerId numeric feature
2 Survived categorical feature 0, 1
3 Pclass numeric feature
4 Sex categorical feature female, male
5 Sex-Binary categorical feature 0, 1
6 Age numeric feature
7 SibSp numeric feature
Browse documentation datasets
Reset
Apply
Components
Data
Heat Map
Data Table (2)
Tree
Tree Viewer
Select Columns
Data Table (3)
File (1)


Data
File
CSV File Import
Datasets
SQL Table
Data Table
Paint Data
Data Info
Data Sampler
Select Columns
Select Rows
Pivot Table
Rank
Data Table
View the dataset in a spreadsheet.
more...
Data Table (4)
Info
891 instances
10 features (2.0% missing values)
No target variable.
3 meta attributes (25.7% missing values)
Variables
Show variable labels (if present)
Visualize numeric values
Color by instance classes
Selection
Select full rows
Restore Original Order
Send Automatically
Age SibSp Parch Fare Embarked
1 22.00 1 0 7.2500 S
2 38.00 1 0 71.2833 C
3 26.00 0 0 7.9250 S
4 35.00 1 0 53.1000 S
5 35.00 0 0 8.0500 S
6 ? 0 0 8.4583 Q
7 54.00 0 0 51.8625 S
8 2.00 3 1 21.0750 S
9 27.00 0 2 11.1333 S
10 14.00 1 0 30.0708 C
11 4.00 1 1 16.7000 S
12 58.00 0 0 26.5500 S
13 20.00 0 0 8.0500 S
14 39.00 1 5 31.2750 S
15 14.00 0 0 7.8542 S
16 55.00 0 0 16.0000 S
17 2.00 4 1 29.1250 Q
18 ? 0 0 13.0000 S
19 31.00 1 0 18.0000 S

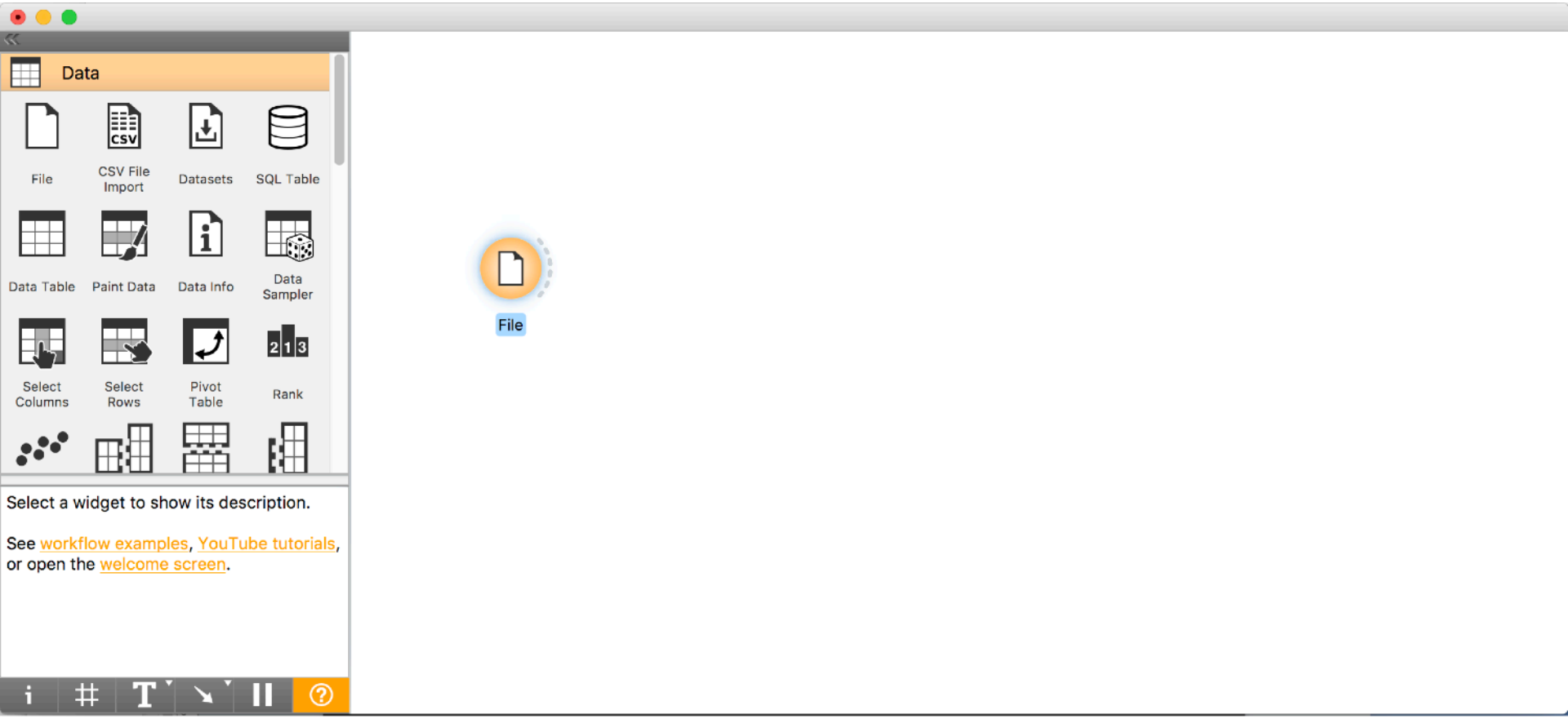

Data
File
CSV File Import
Datasets
SQL Table
Data Table
Paint Data
Data Info
Data Sampler
Select Columns
Select Rows
Pivot Table
Rank
Select Columns
Select columns from the data table and assign them to data features, classes or meta variables.
more...
File
Data
Data Table
Data
Impute
Data
PCA
Components → Data
Heat Map
Components → Data
Data Table (2)
Data
Data Table (1)
Data
Rank
Data
Select Columns
Data
Data Table (3)
Data
Tree
Model → Tree
Tree Viewer
File (1)
Data
Data Table (4)
Data
Impute (1)
Data
Data Table (5)
Data
Select Columns (1)

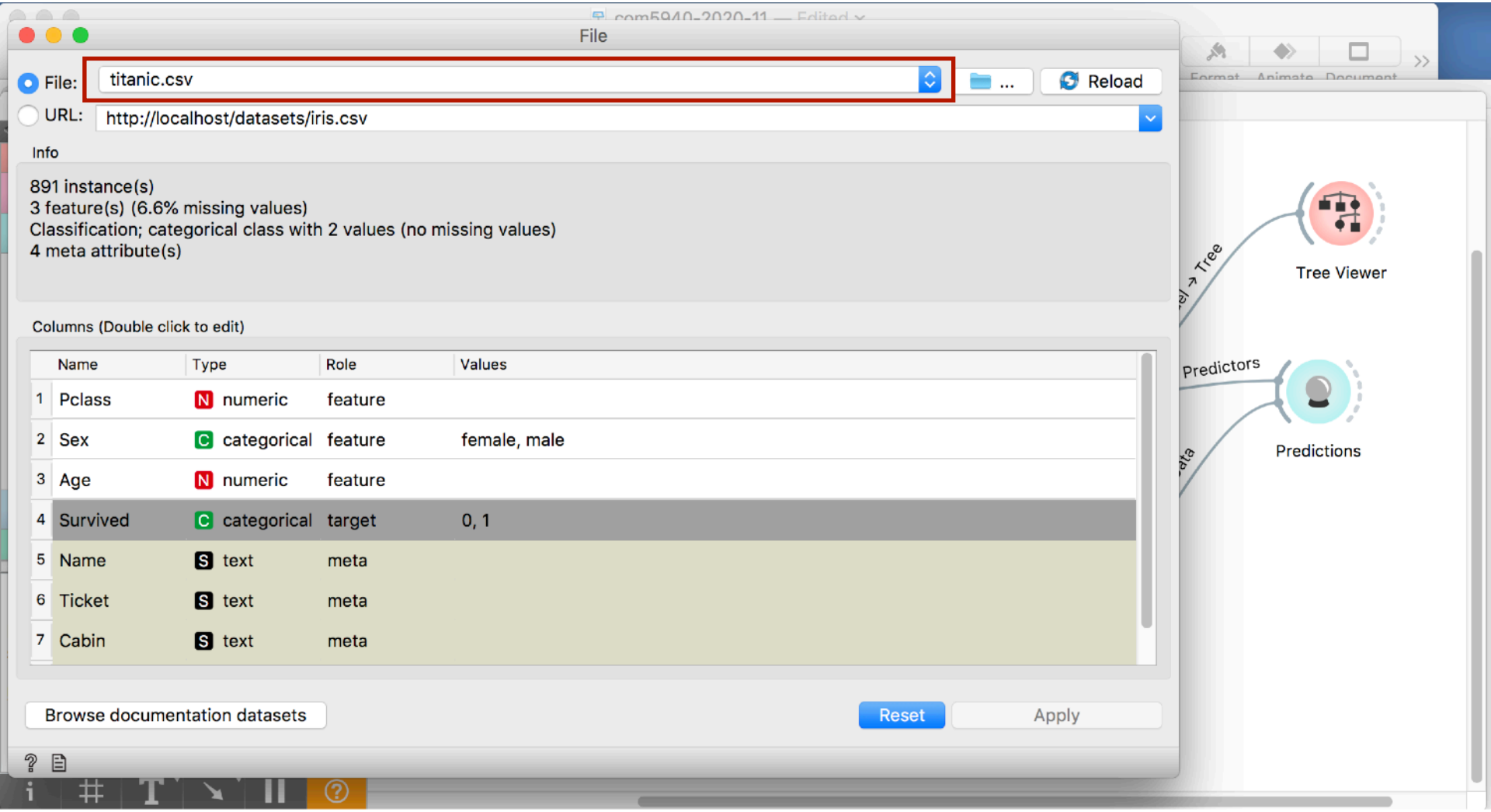

Select Columns (1)
Available Variables
Filter
Selected
Name
Ticket
Cabin
Selected (1)
PassengerId
SibSp
Sex-Binary
Parch
Fare
Embarked
Up
<
Down
Features
Filter
Pclass
Sex
Age
Target Variable
Survived
Meta Attributes
Reset
Send Automatically
Impute
Data
PCA
Components → Data
Heat Map
Data Table (1)
Data Table (2)
Rank
Select Columns
Data Table (3)
Tree
Tree Viewer
Impute (1)
Data Table (5)
Select Columns (1)

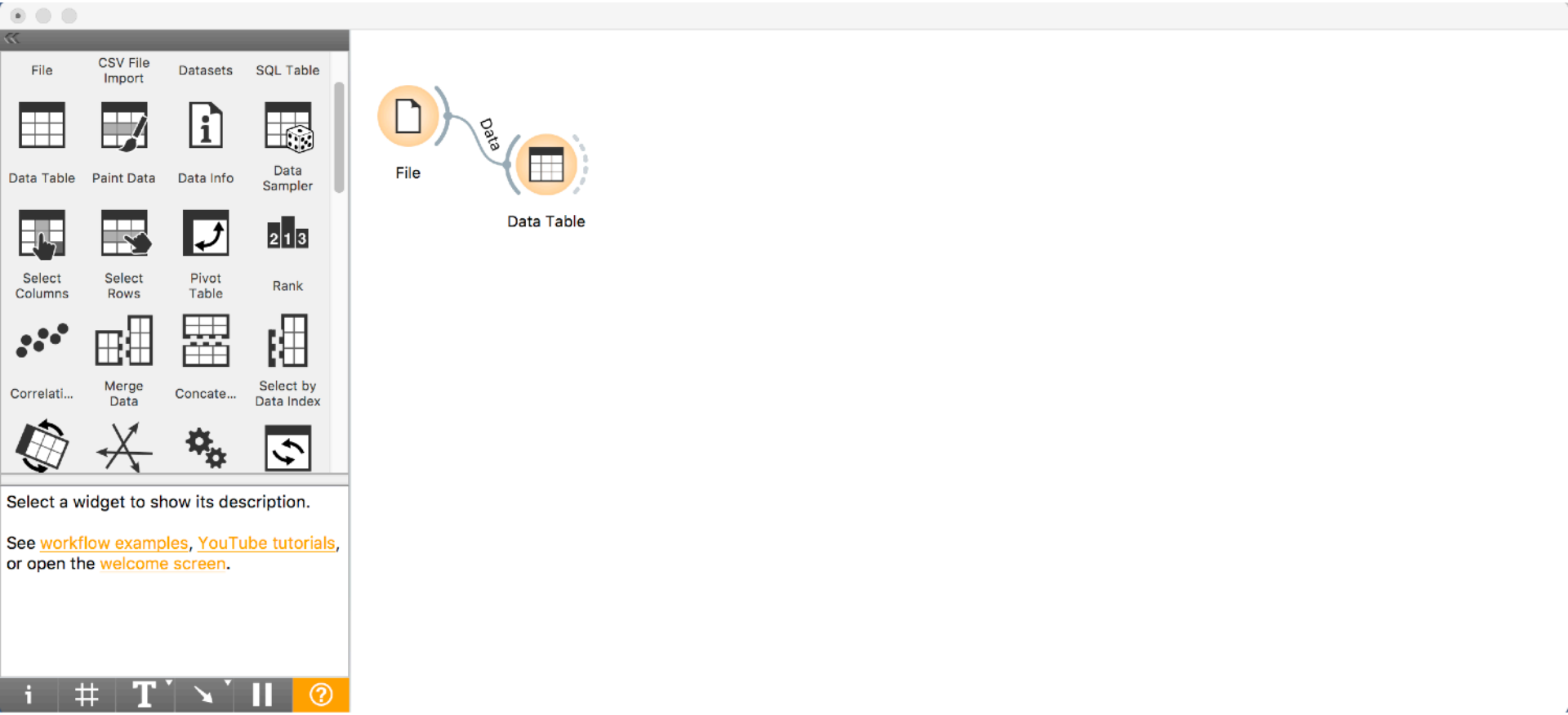

Evaluate
Test and Score
Predictio...
Confusion Matrix
ROC Analysis
Lift Curve
Calibration Plot
Unsupervised
Prototypes
Image Analytics
Time Series
Predictions
Display predictions of models for an input dataset.
more...
File
Data Table
Impute
PCA
Heat Map
Data Table (1)
Data Table (2)
Rank
Select Columns
Data Table (3)
Tree
Tree Viewer
File (1)
Data Table (4)
Impute (1)
Data Table (5)
Select Columns (1)
Predictions
Data
Components → Data
Model → Predictors

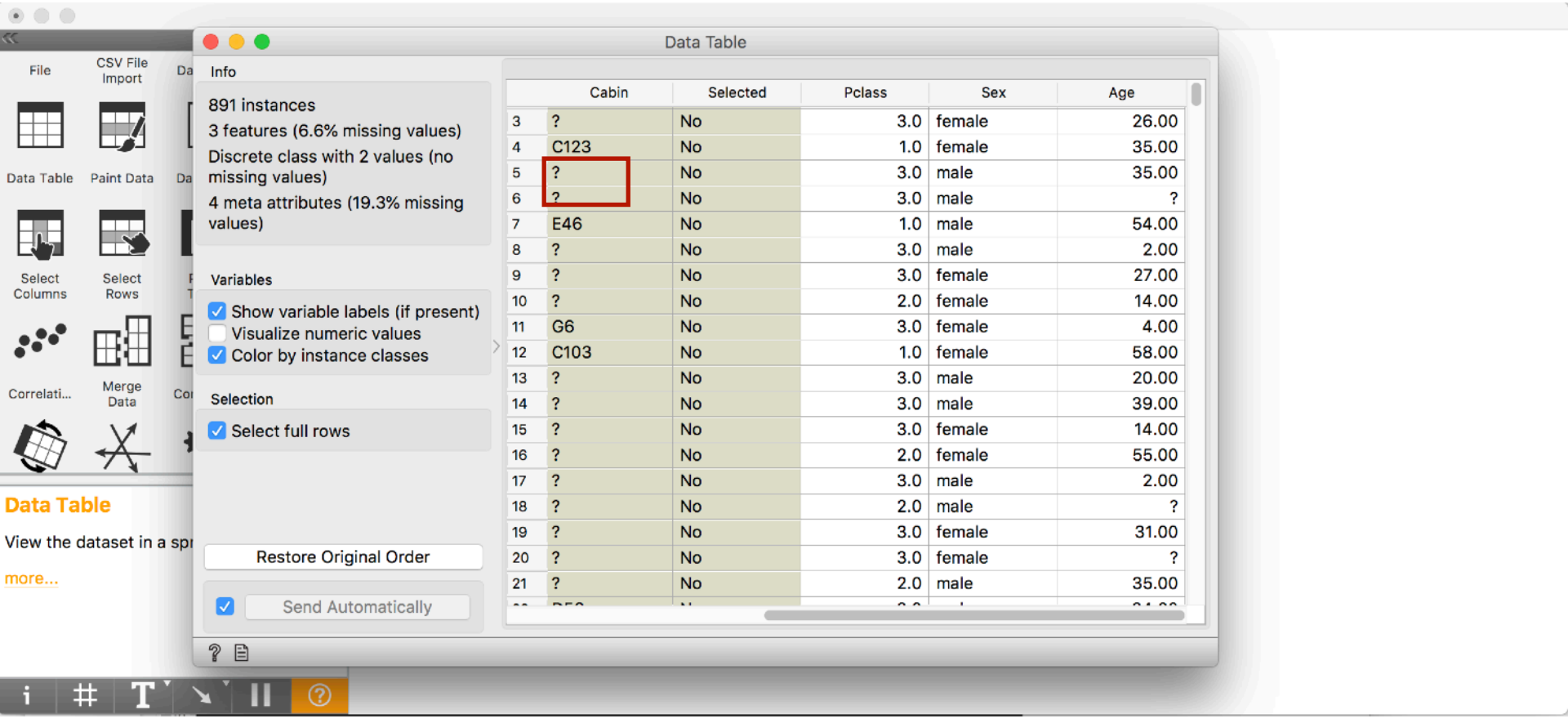

Predictions
Show probabilities for
0
1
Restore Original Order
Tree
Survived
Pclass
Sex
Age
1 0.91 : 0.09 0 3 male 22.00
2 0.04 : 0.96 1 1 female 38.00
3 0.17 : 0.83 1 3 female 26.00
4 0.04 : 0.96 1 1 female 35.00
5 0.96 : 0.04 0 3 male 35.00
6 0.88 : 0.12 0 1 male 54.00
7 0.56 : 0.44 0 3 male 2.00
8 0.17 : 0.83 1 3 female 27.00
9 0.07 : 0.93 1 2 female 14.00
10 0.00 : 1.00 1 3 female 4.00
11 0.17 : 0.83 1 1 female 58.00
12 0.83 : 0.17 0 3 male 20.00
13 0.96 : 0.04 0 3 male 39.00
14 0.30 : 0.70 0 3 female 14.00
15 0.04 : 0.96 1 2 female 55.00
16 0.56 : 0.44 0 3 male 2.00
17 0.67 : 0.33 0 3 female 31.00
Model AUC CA F1 Precision Recall
Tree 0.913 0.851 0.848 0.854 0.851
712
Heat Map
Components → Data
Components → Data
Data Table (2)
Data
Tree
Model → Tree
Tree Viewer
Model → Predictors
Data
elect Columns
Data Table (3)
Data
Data
Select Columns (1)
Predictions

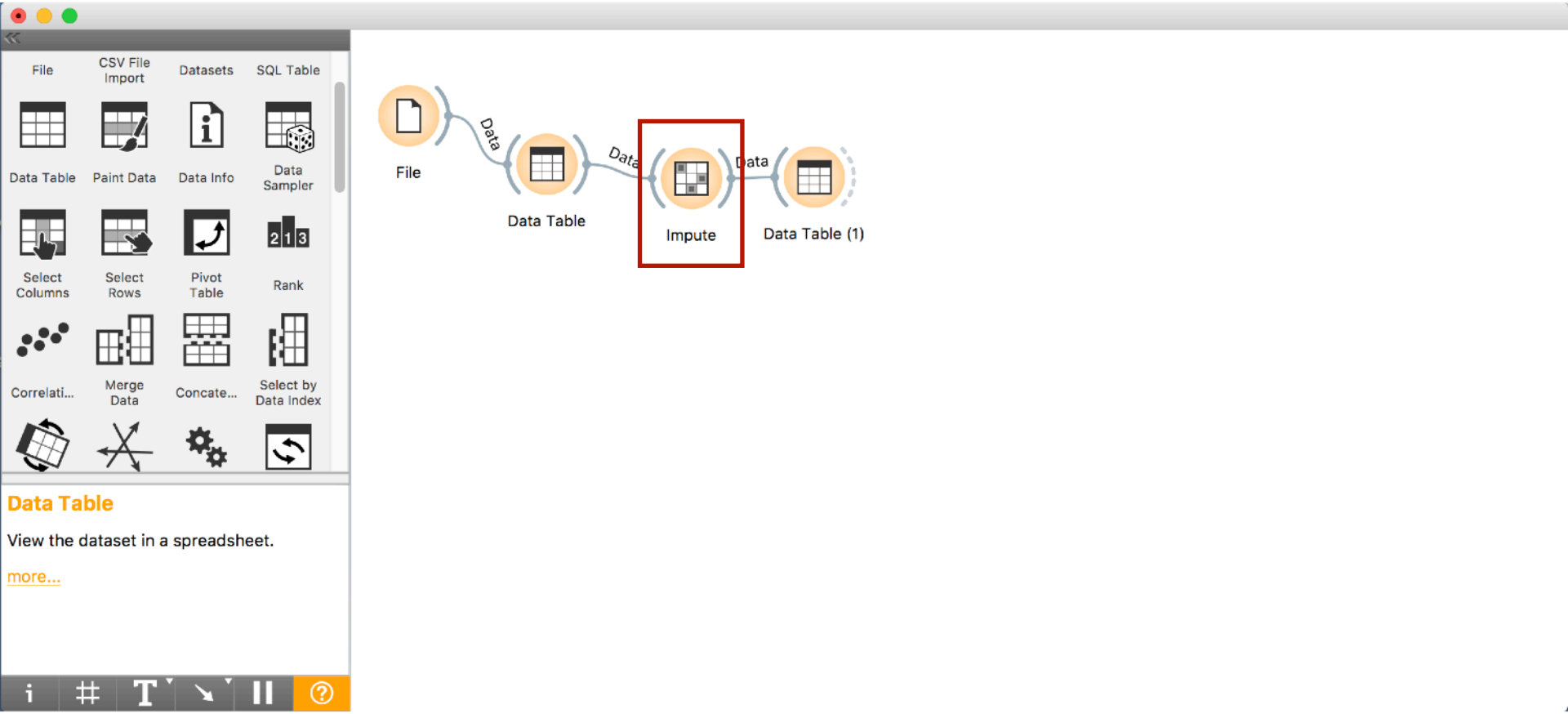
Another way of training and testing the model (when you only have one dataset not separated into training and testing sets).

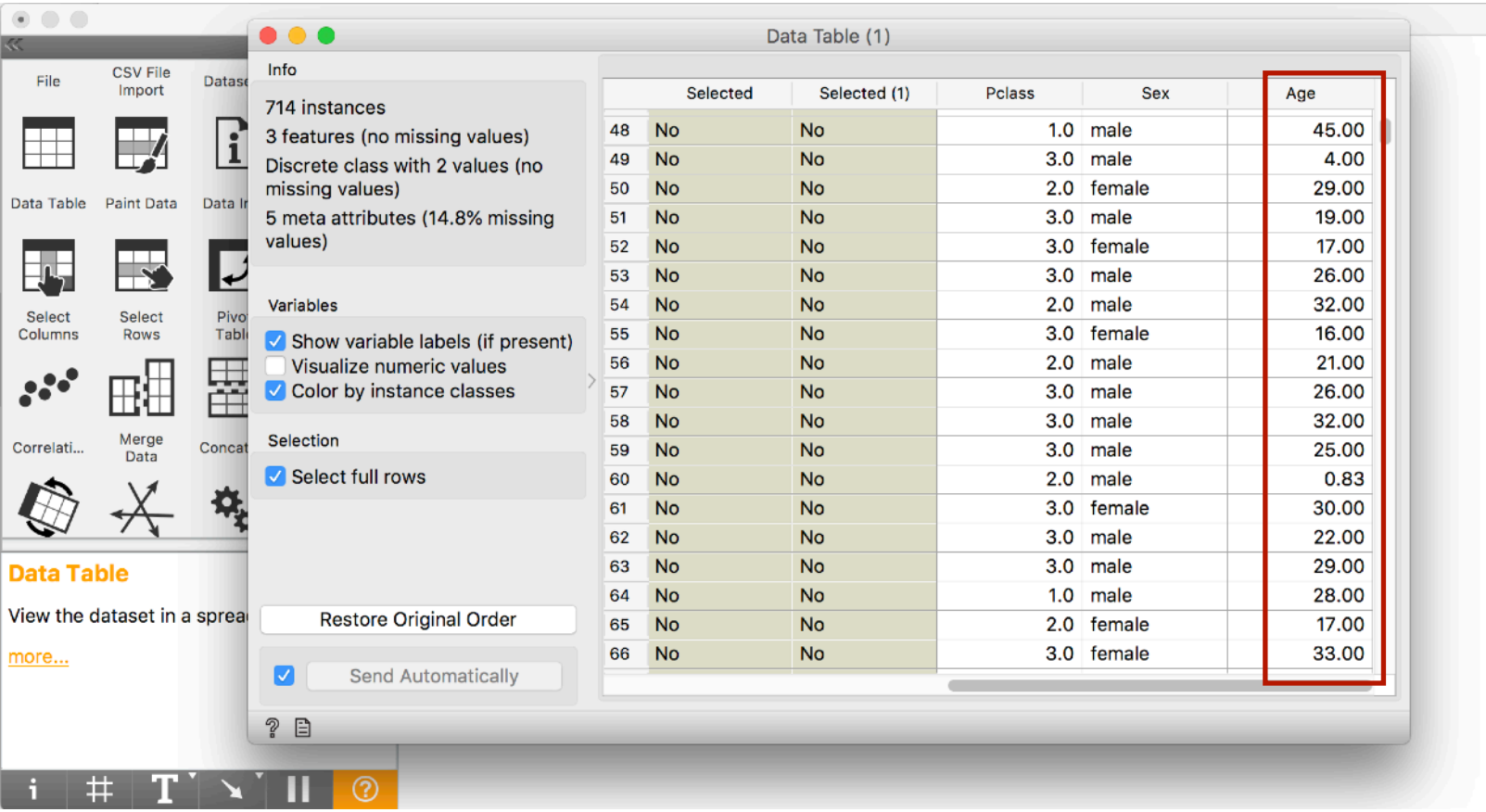

Data
File
CSV File Import
Datasets
SQL Table
Data Table
Paint Data
Data Info
Data Sampler
Select Columns
Select Rows
Pivot Table
Rank
Select a widget to show its description.
See workflow examples, YouTube tutorials, or open the welcome screen.
File

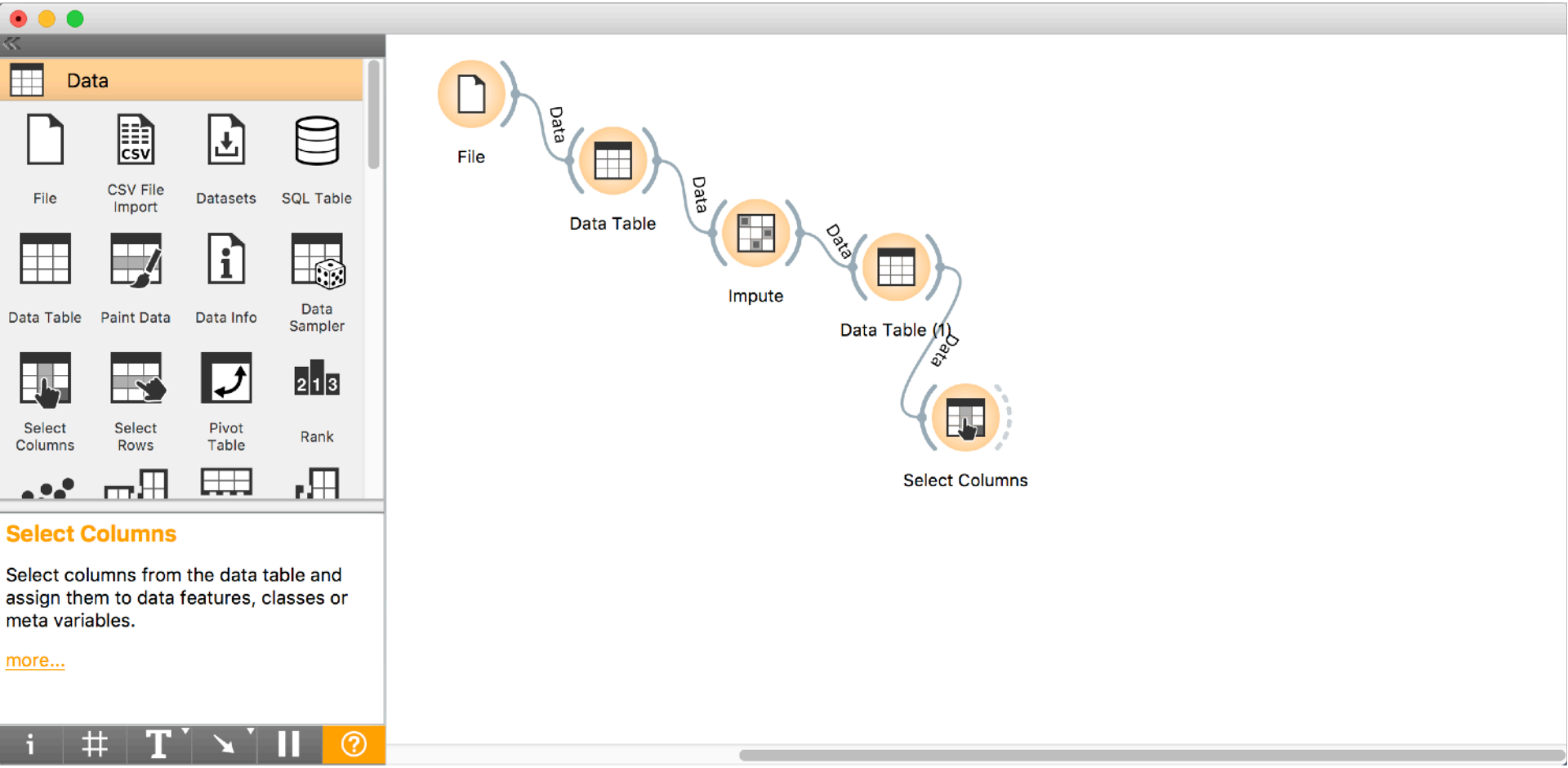

File
File:
titanic.csv
...
Reload
URL:
http://localhost/datasets/iris.csv
Info
891 instance(s)
3 feature(s) (6.6% missing values)
Classification; categorical class with 2 values (no missing values)
4 meta attribute(s)
Columns (Double click to edit)
Name
Type
Role
Values
1 Pclass numeric feature
2 Sex categorical feature female, male
3 Age numeric feature
4 Survived categorical target 0, 1
5 Name text meta
6 Ticket text meta
7 Cabin text meta
Browse documentation datasets
Reset
Apply
Tree Viewer
Predictors
Predictions

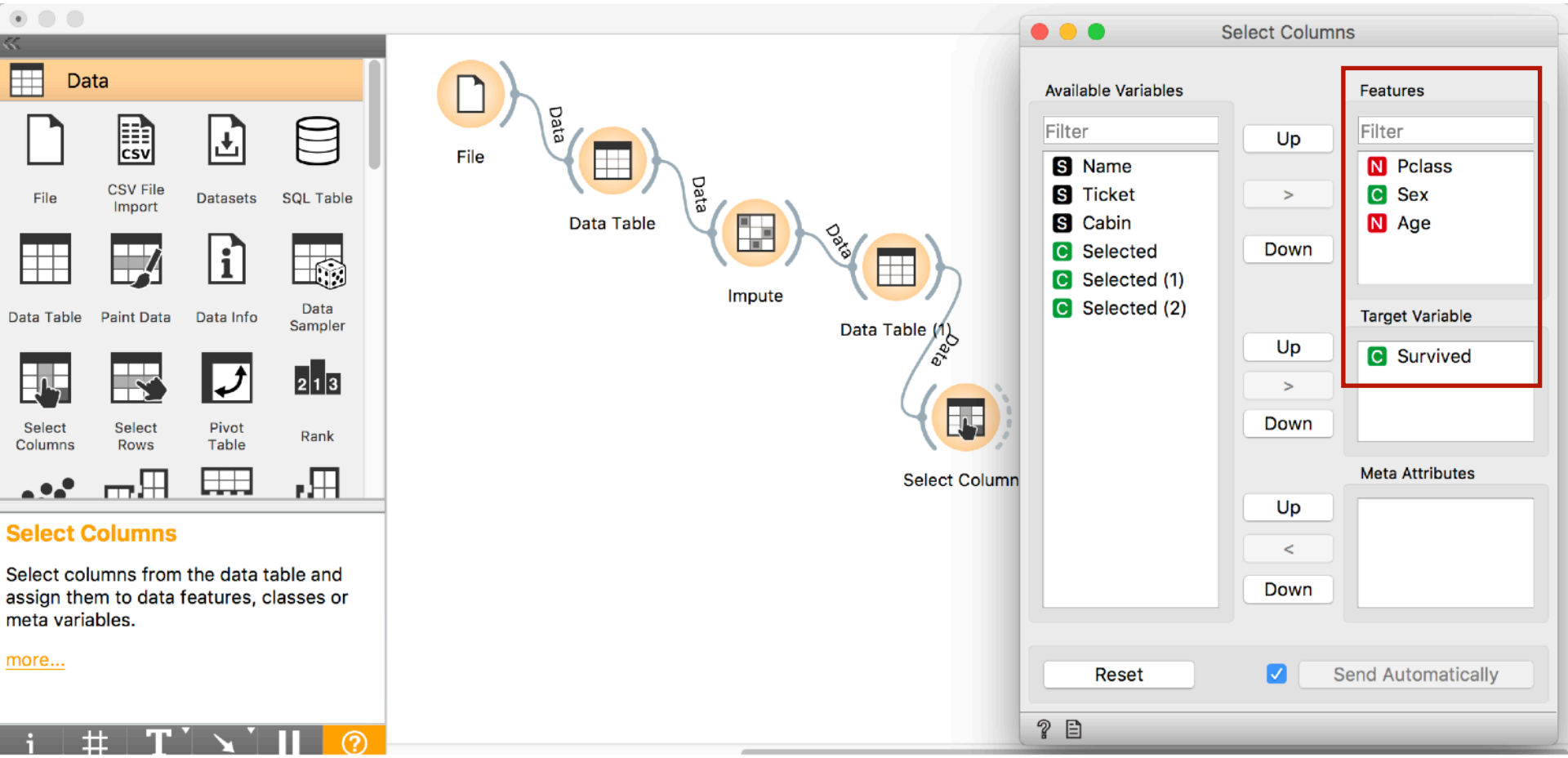

File
CSV File Import
Datasets
SQL Table
Data Table
Paint Data
Data Info
Data Sampler
Select Columns
Select Rows
Pivot Table
Rank
Correlati...
Merge Data
Concate...
Select by Data Index
Select a widget to show its description.
See workflow examples, YouTube tutorials, or open the welcome screen.
Data
File
Data Table

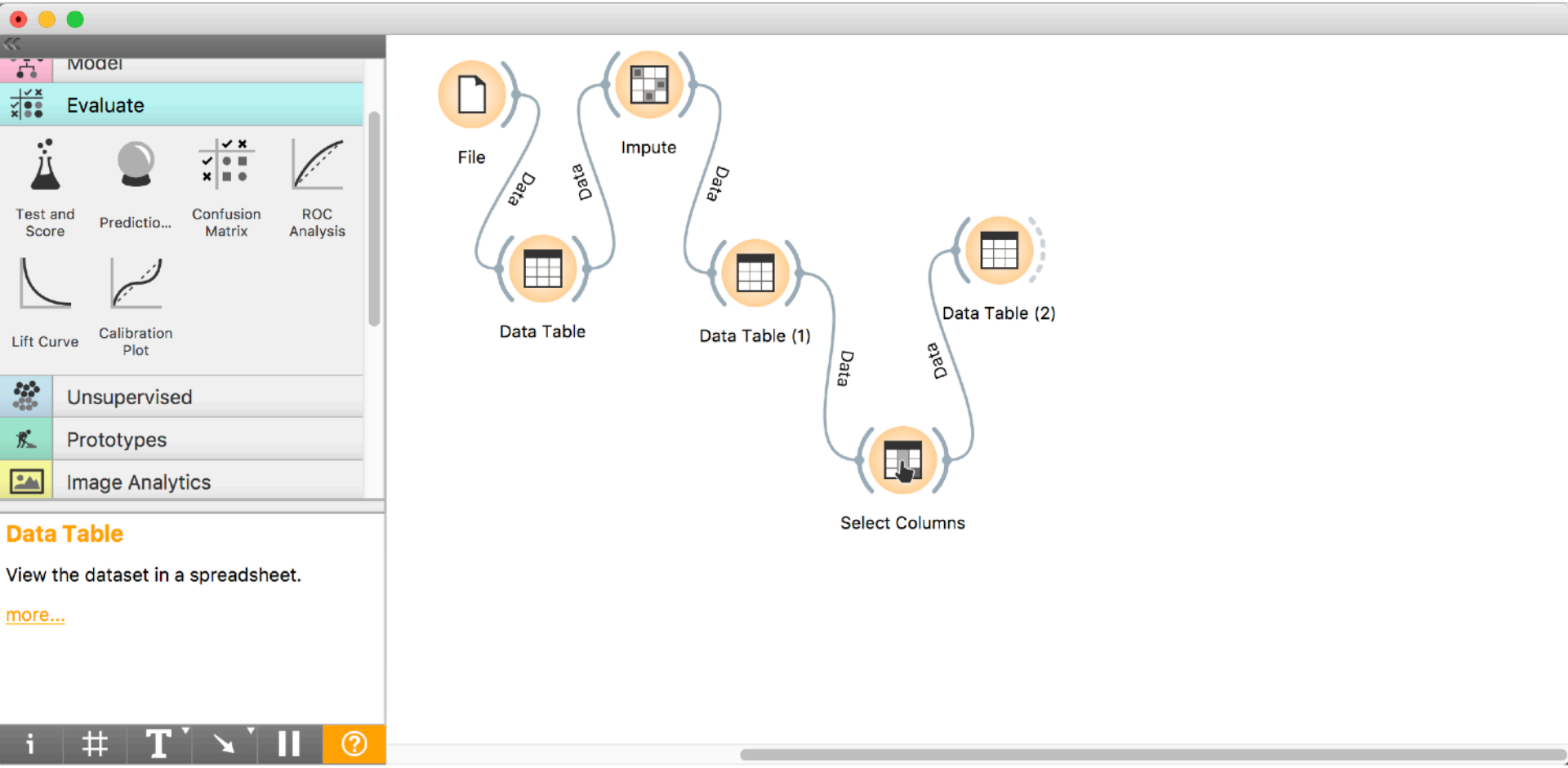

File
CSV File Import
Data Table
Paint Data
Select Columns
Select Rows
Correlati...
Merge Data
Data Table
View the dataset in a spr
more...
Data Table
Info
891 instances
3 features (6.6% missing values)
Discrete class with 2 values (no missing values)
4 meta attributes (19.3% missing values)
Variables
Show variable labels (if present)
Visualize numeric values
Color by instance classes
Selection
Select full rows
Restore Original Order
Send Automatically
Cabin
Selected
Pclass
Sex
Age
3 ? No 3.0 female 26.00
4 C123 No 1.0 female 35.00
5 ? No 3.0 male 35.00
6 ? No 3.0 male ?
7 E46 No 1.0 male 54.00
8 ? No 3.0 male 2.00
9 ? No 3.0 female 27.00
10 ? No 2.0 female 14.00
11 G6 No 3.0 female 4.00
12 C103 No 1.0 female 58.00
13 ? No 3.0 male 20.00
14 ? No 3.0 male 39.00
15 ? No 3.0 female 14.00
16 ? No 2.0 female 55.00
17 ? No 3.0 male 2.00
18 ? No 2.0 male ?
19 ? No 3.0 female 31.00
20 ? No 3.0 female ?
21 ? No 2.0 male 35.00

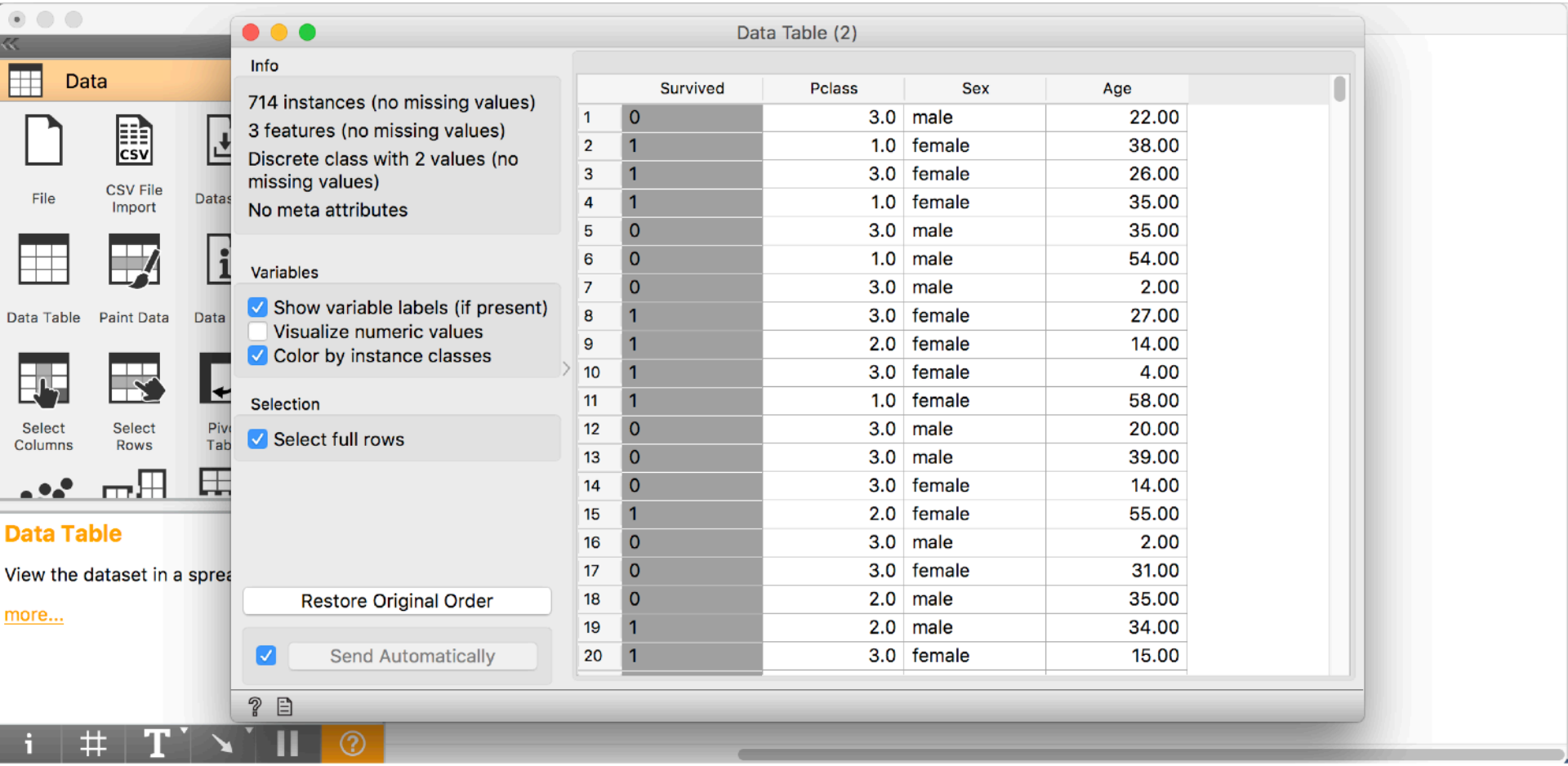

File
CSV File Import
Datasets
SQL Table
Data Table
Paint Data
Data Info
Data Sampler
Select Columns
Select Rows
Pivot Table
Rank
Correlati...
Merge Data
Concate...
Select by Data Index
Data Table
View the dataset in a spreadsheet.
more...
Data
Data
Data
File
Data Table
Impute
Data Table (1)


File
CSV File Import
Data Table
Paint Data
Select Columns
Select Rows
Merge Data
Data Table
View the dataset in a sprea
more...
Data Table (1)
Info
714 instances
3 features (no missing values)
Discrete class with 2 values (no missing values)
5 meta attributes (14.8% missing values)
Variables
Show variable labels (if present)
Visualize numeric values
Color by instance classes
Selection
Select full rows
Restore Original Order
Send Automatically
Selected
Selected (1)
Pclass
Sex
Age
48 No No 1.0 male 45.00
49 No No 3.0 male 4.00
50 No No 2.0 female 29.00
51 No No 3.0 male 19.00
52 No No 3.0 female 17.00
53 No No 3.0 male 26.00
54 No No 2.0 male 32.00
55 No No 3.0 female 16.00
56 No No 2.0 male 21.00
57 No No 3.0 male 26.00
58 No No 3.0 male 32.00
59 No No 3.0 male 25.00
60 No No 2.0 male 0.83
61 No No 3.0 female 30.00
62 No No 3.0 male 22.00
63 No No 3.0 male 29.00
64 No No 1.0 male 28.00
65 No No 2.0 female 17.00
66 No No 3.0 female 33.00

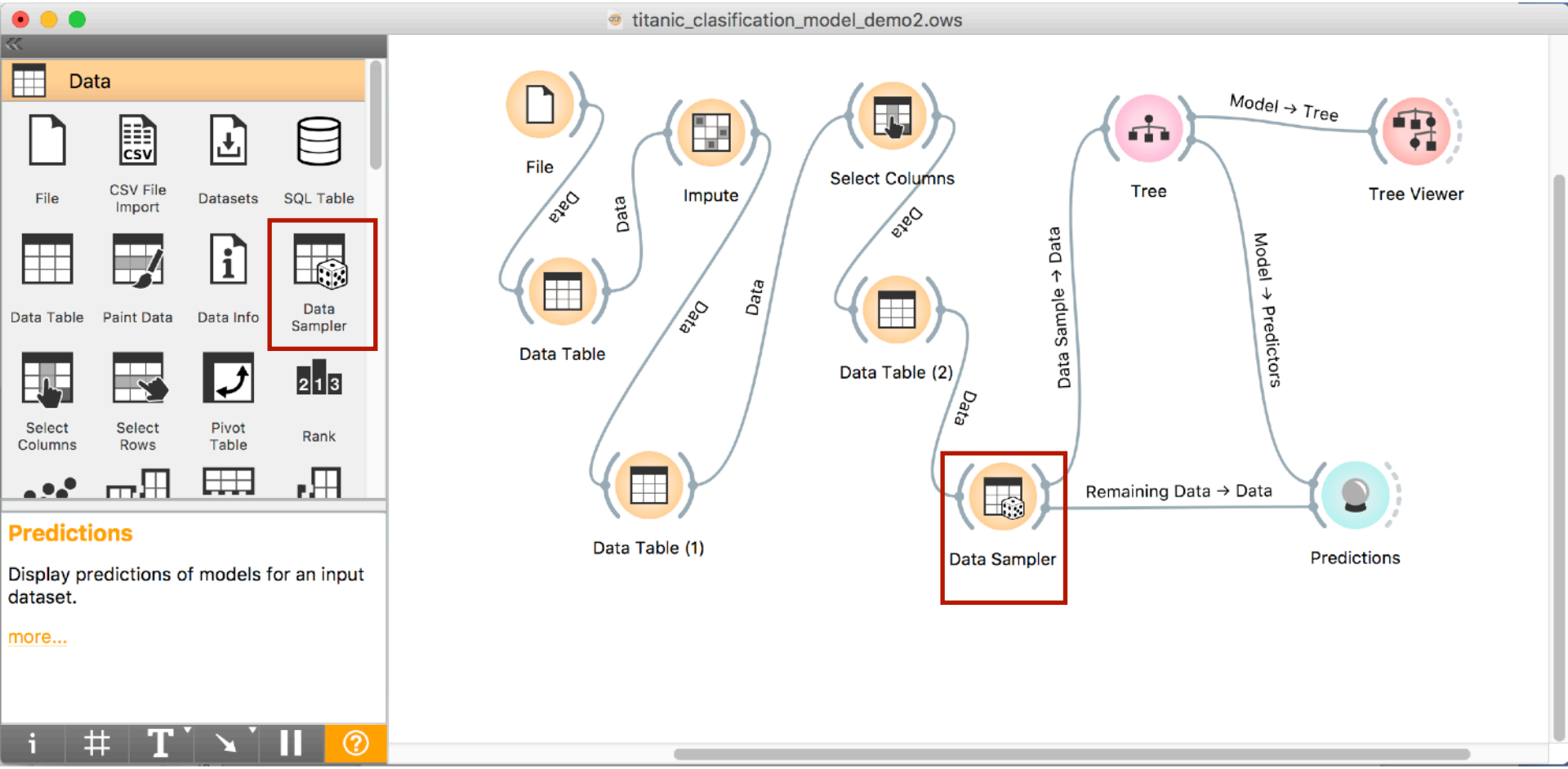

Data
File
CSV File Import
Datasets
SQL Table
Data Table
Paint Data
Data Info
Data Sampler
Select Columns
Select Rows
Pivot Table
Rank
Select Columns
Select columns from the data table and assign them to data features, classes or meta variables.
more...
File
Data
Data Table
Data
Impute
Data
Data Table (1)
Data
Select Columns

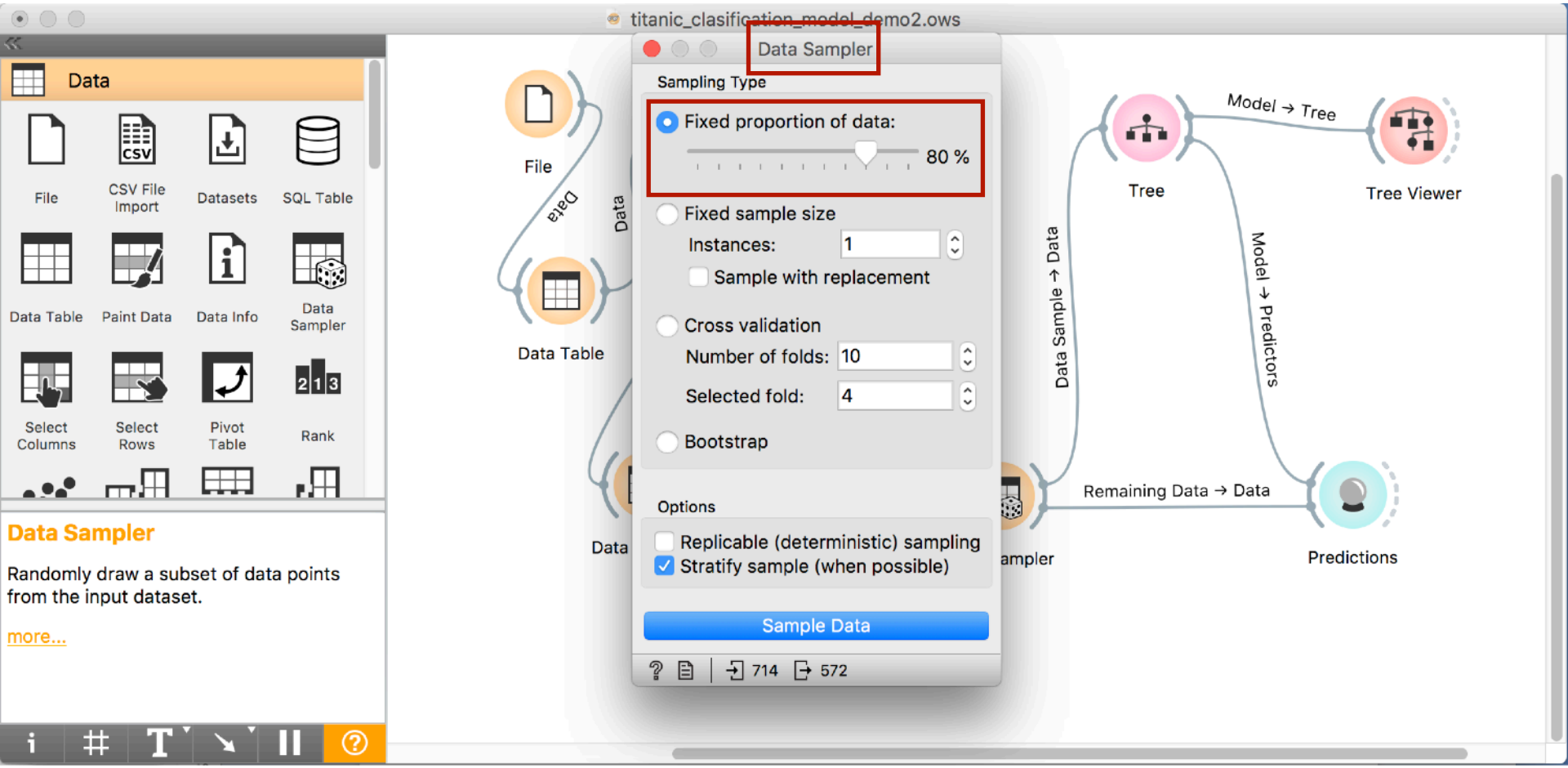

Data
File
CSV File Import
Datasets
SQL Table
Data Table
Paint Data
Data Info
Data Sampler
Select Columns
Select Rows
Pivot Table
Rank
Select Columns
Select columns from the data table and assign them to data features, classes or meta variables.
more...
File
Data
Data Table
Data
Impute
Data
Data Table (1)
Data
Select Column
Select Columns
Available Variables
Filter
Name
Ticket
Cabin
Selected
Selected (1)
Selected (2)
Up
>
Down
Up
>
Down
Up
<
Down
Features
Filter
Pclass
Sex
Age
Target Variable
Survived
Meta Attributes
Reset
Send Automatically

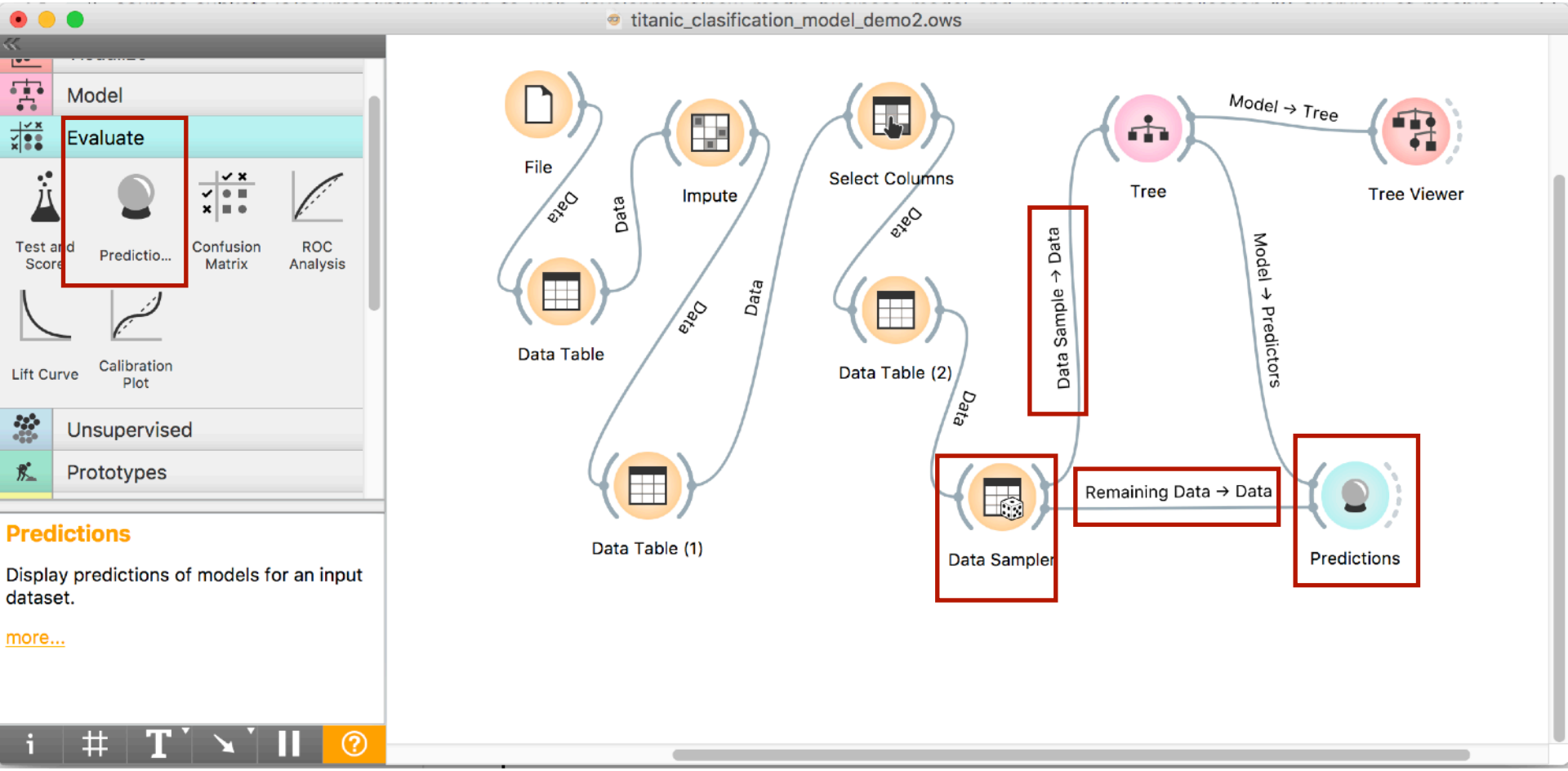

Model
Evaluate
Test and Score
Predictio...
Confusion Matrix
ROC Analysis
Lift Curve
Calibration Plot
Unsupervised
Prototypes
Image Analytics
Data Table
View the dataset in a spreadsheet.
more...
File
Impute
Data
Data
Data
Data
Data
Data Table
Data Table (1)
Data Table (2)
Select Columns

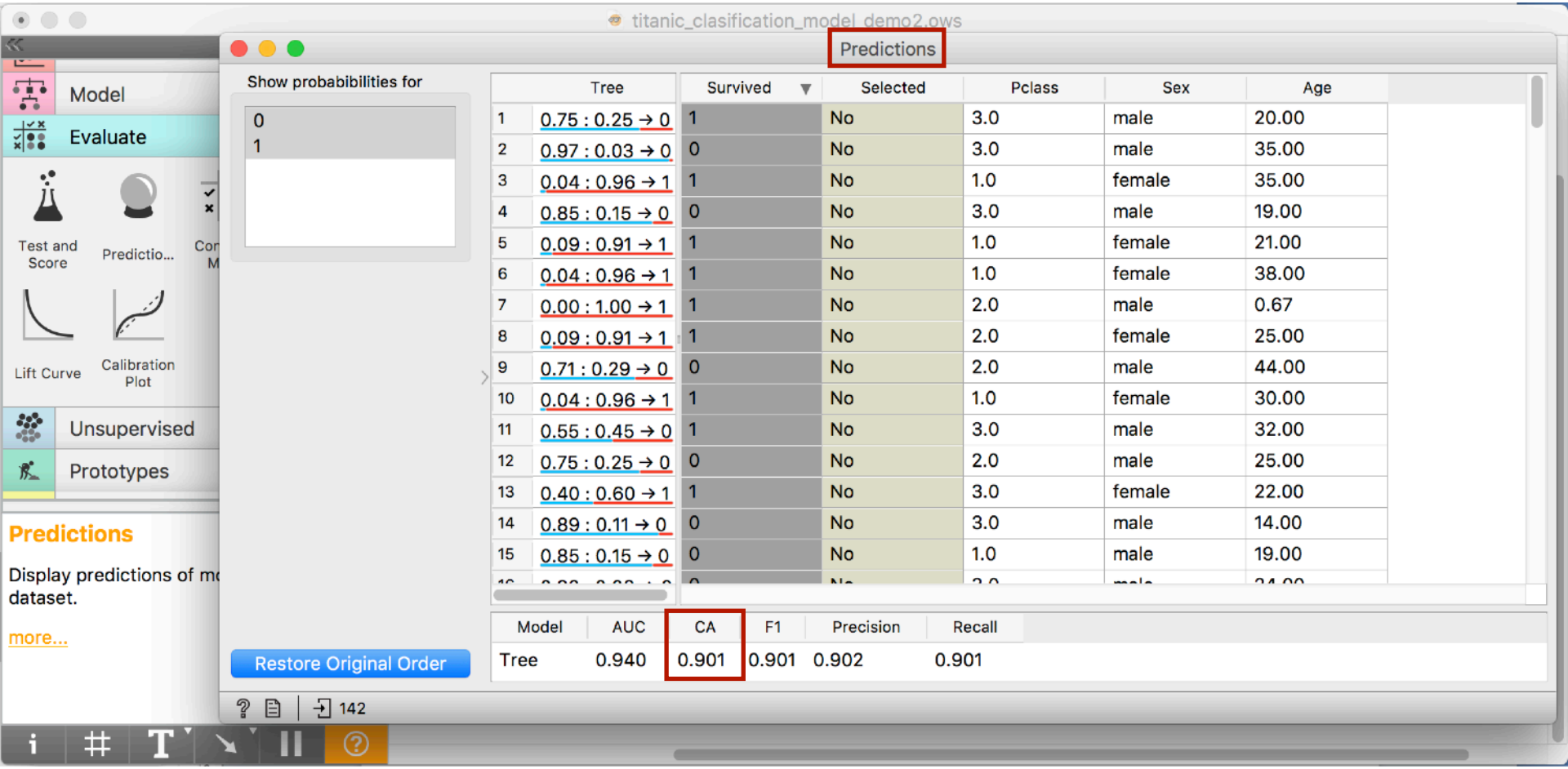


	Survived	Pclass	Sex	Age
1	0	3.0	male	22.00
2	1	1.0	female	38.00
3	1	3.0	female	26.00
4	1	1.0	female	35.00
5	0	3.0	male	35.00
6	0	1.0	male	54.00
7	0	3.0	male	2.00
8	1	3.0	female	27.00
9	1	2.0	female	14.00
10	1	3.0	female	4.00
11	1	1.0	female	58.00
12	0	3.0	male	20.00
13	0	3.0	male	39.00
14	0	3.0	female	14.00
15	1	2.0	female	55.00
16	0	3.0	male	2.00
17	0	3.0	female	31.00
18	0	2.0	male	35.00
19	1	2.0	male	34.00
20	1	3.0	female	15.00

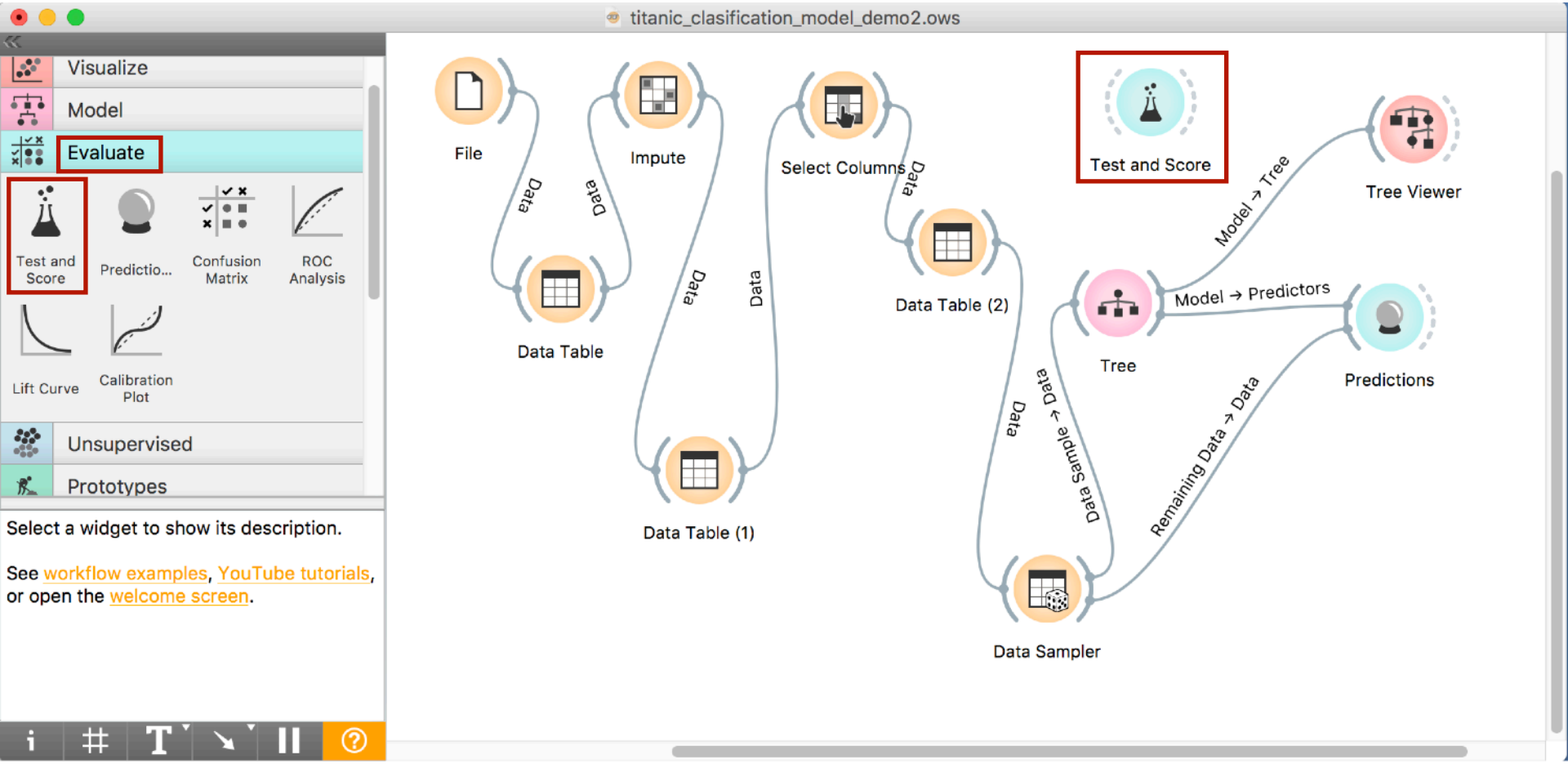

titanic_clasification_model_demo2.ows
Data
File
CSV File Import
Datasets
SQL Table
Data Table
Paint Data
Data Info
Data Sampler
Select Columns
Select Rows
Pivot Table
Rank
Predictions
Display predictions of models for an input dataset.
more...
File
Impute
Select Columns
Tree
Tree Viewer
Model → Tree
Data Table
Data Table (2)
Data Table (1)
Data Sampler
Predictions
Data
Data Sample → Data
Model → Predictors
Remaining Data → Data

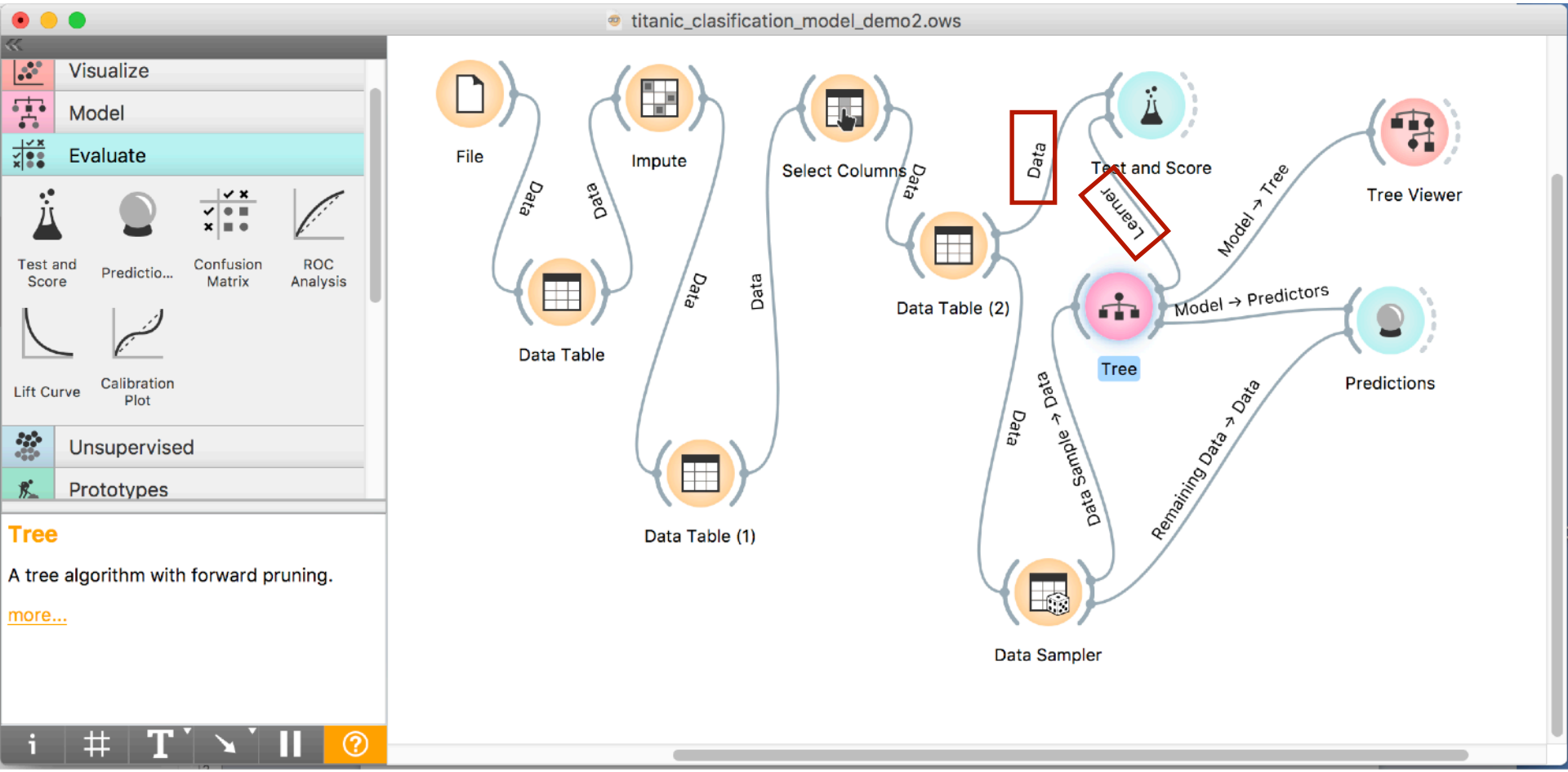

titanic_clasification_model_demo2.ows
Data
File
CSV File Import
Datasets
SQL Table
Data Table
Paint Data
Data Info
Data Sampler
Select Columns
Select Rows
Pivot Table
Rank
Data Sampler
Randomly draw a subset of data points from the input dataset.
more...
Data Sampler
Sampling Type
Fixed proportion of data:
80 %
Fixed sample size
Instances: 1
Sample with replacement
Cross validation
Number of folds: 10
Selected fold: 4
Bootstrap
Options
Replicable (deterministic) sampling
Stratify sample (when possible)
Sample Data
714
572
File
Data
Data Table
Data
Data
Tree
Model → Tree
Tree Viewer
Model → Predictors
Data Sample → Data
Remaining Data → Data
Predictions

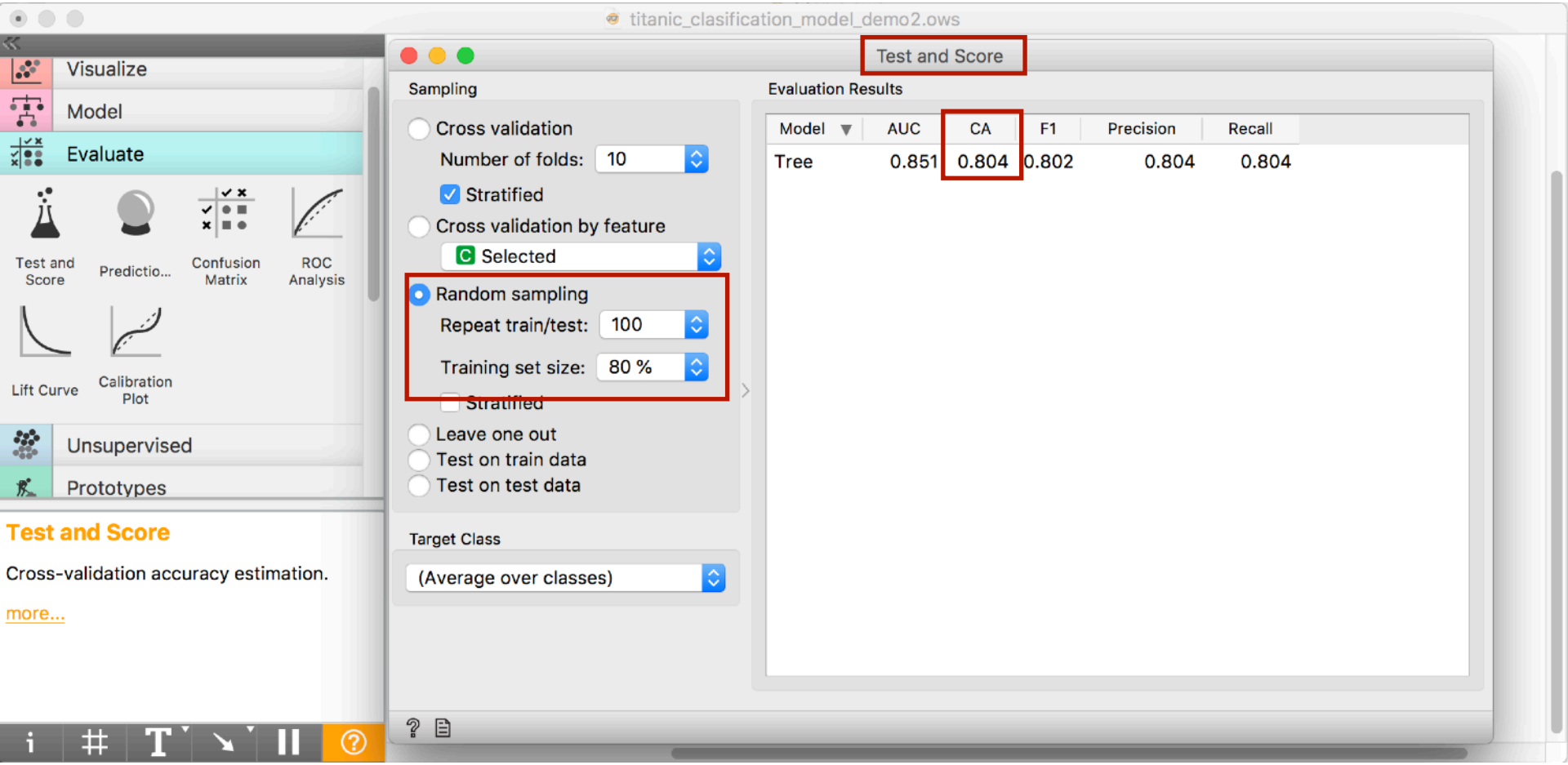

titanic_clasification_model_demo2.ows
Model
Evaluate
Test and Score
Predictio...
Confusion Matrix
ROC Analysis
Lift Curve
Calibration Plot
Unsupervised
Prototypes
Predictions
Display predictions of models for an input dataset.
more...
File
Impute
Select Columns
Tree
Tree Viewer
Model → Tree
Data Table
Data Table (1)
Data Table (2)
Data Sampler
Predictions
Data
Data Sample → Data
Model → Predictors
Remaining Data → Data


titanic_clasification_model_demo2.ows
Predictions
Model
Evaluate
Test and Score
Predictio...
Lift Curve
Calibration Plot
Unsupervised
Prototypes
Predictions
Display predictions of mo
dataset.
more...
Show probabibilities for
0
1
Restore Original Order
Tree
Survived
Selected
Pclass
Sex
Age
1 0.75 : 0.25 → 0 1 No 3.0 male 20.00
2 0.97 : 0.03 → 0 0 No 3.0 male 35.00
3 0.04 : 0.96 → 1 1 No 1.0 female 35.00
4 0.85 : 0.15 → 0 0 No 3.0 male 19.00
5 0.09 : 0.91 → 1 1 No 1.0 female 21.00
6 0.04 : 0.96 → 1 1 No 1.0 female 38.00
7 0.00 : 1.00 → 1 1 No 2.0 male 0.67
8 0.09 : 0.91 → 1 1 No 2.0 female 25.00
9 0.71 : 0.29 → 0 0 No 2.0 male 44.00
10 0.04 : 0.96 → 1 1 No 1.0 female 30.00
11 0.55 : 0.45 → 0 1 No 3.0 male 32.00
12 0.75 : 0.25 → 0 0 No 2.0 male 25.00
13 0.40 : 0.60 → 1 1 No 3.0 female 22.00
14 0.89 : 0.11 → 0 0 No 3.0 male 14.00
15 0.85 : 0.15 → 0 0 No 1.0 male 19.00
Model AUC CA F1 Precision Recall
Tree 0.940 0.901 0.901 0.902 0.901
142

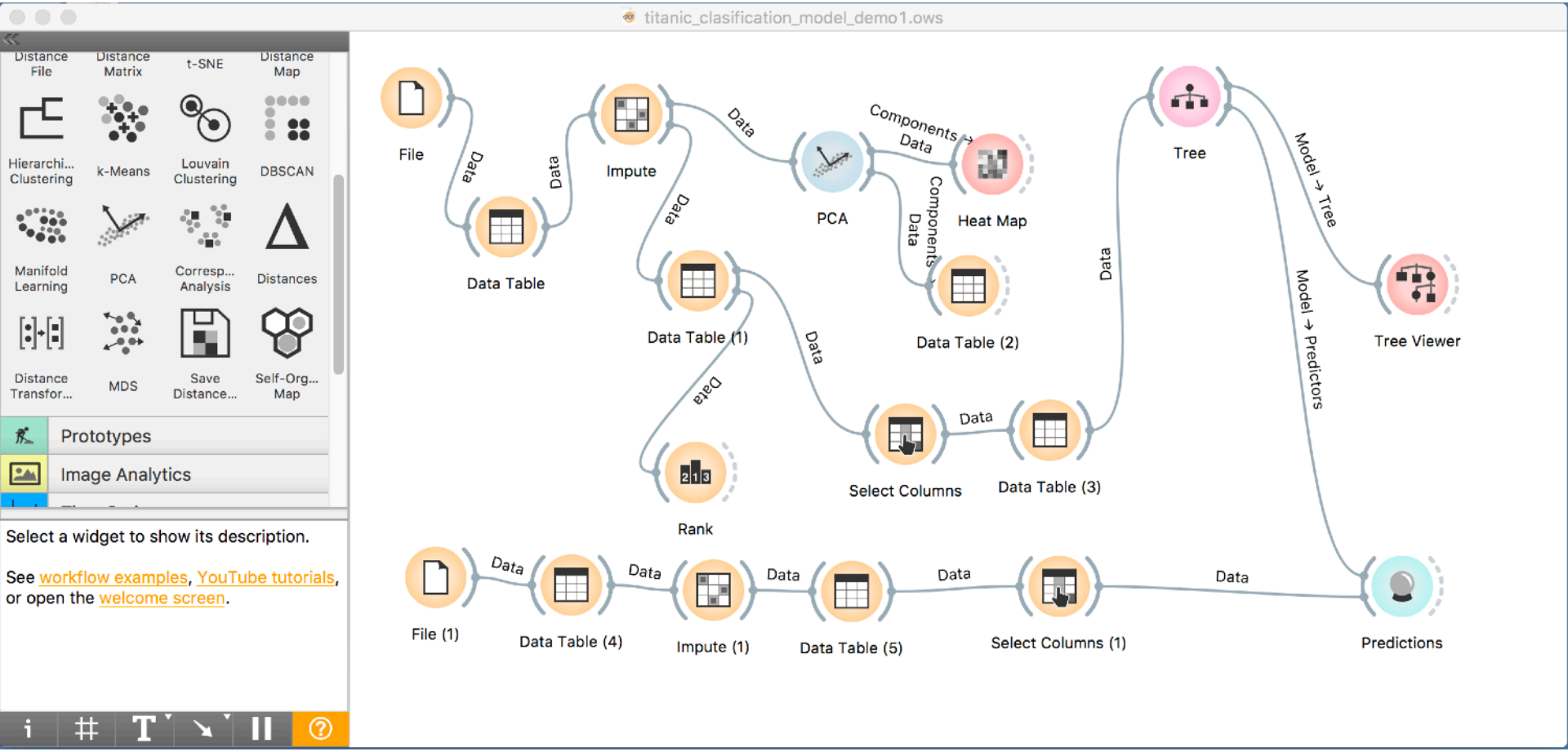
One more method

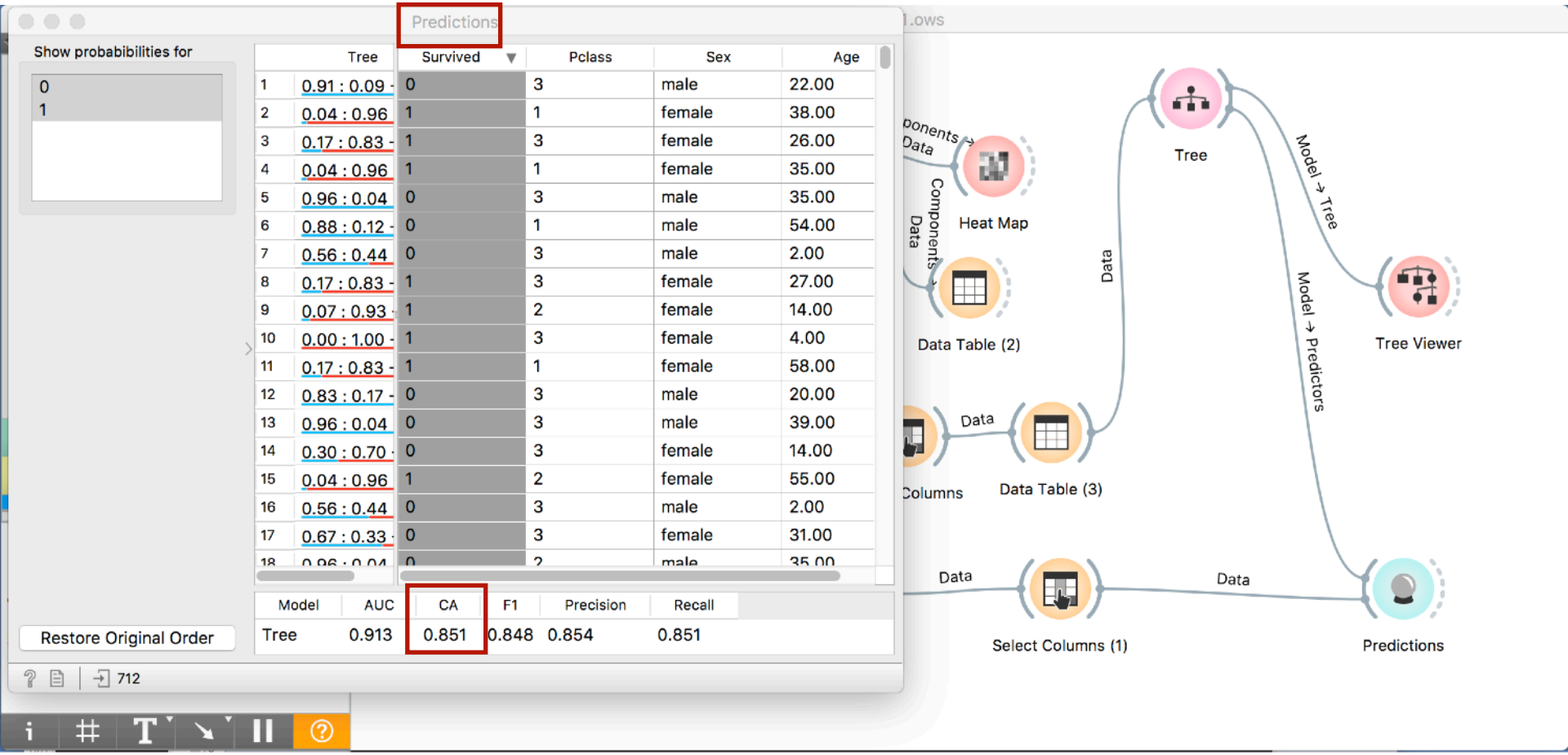

titanic_clasification_model_demo2.ows
Visualize
Model
Evaluate
Test and Score
Predictio...
Confusion Matrix
ROC Analysis
Lift Curve
Calibration Plot
Unsupervised
Prototypes
Select a widget to show its description.
See workflow examples, YouTube tutorials, or open the welcome screen.
File
Impute
Select Columns
Test and Score
Tree Viewer
Data Table
Data Table (1)
Data Table (2)
Tree
Predictions
Data Sampler
Data
Model → Tree
Model → Predictors
Data Sample → Data
Remaining Data → Data

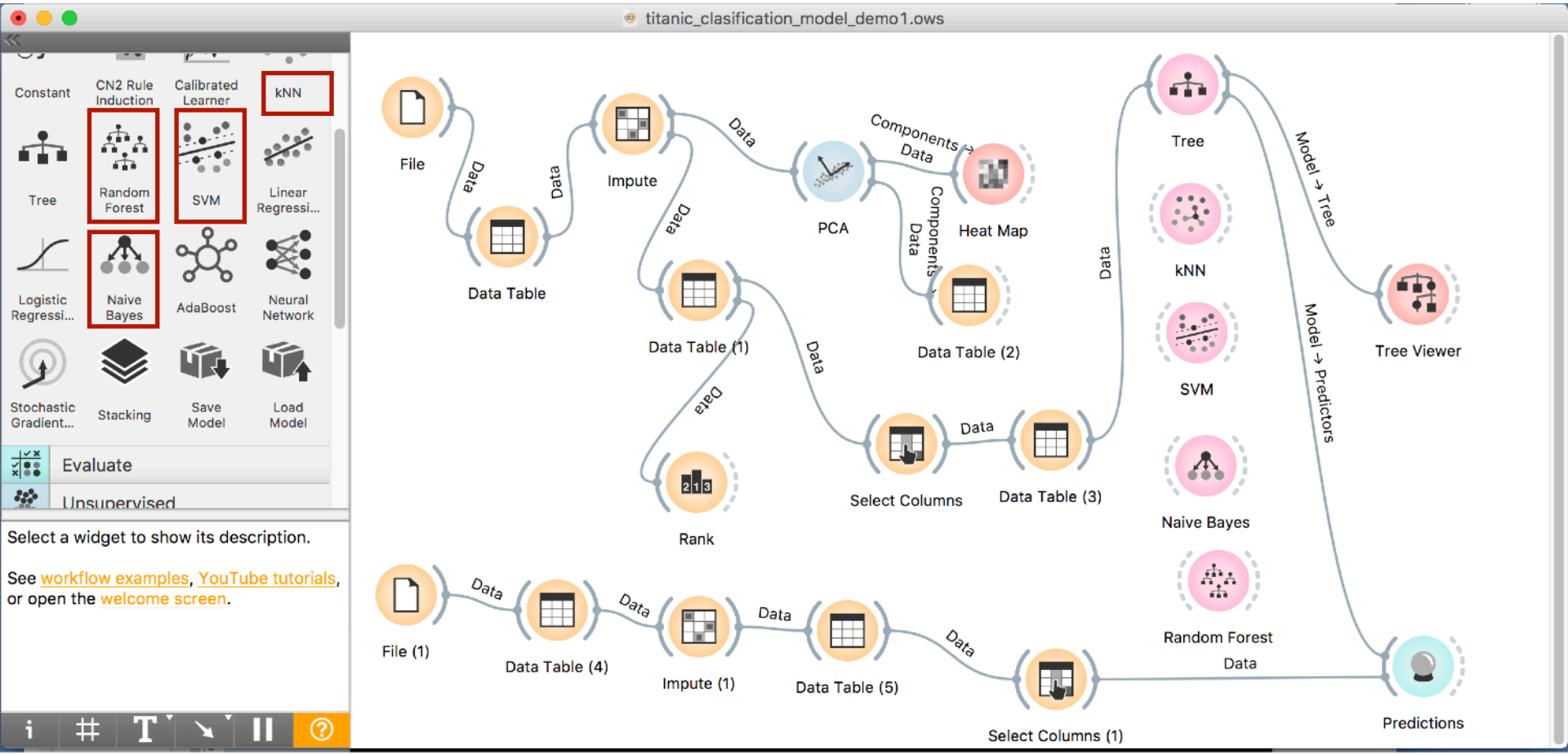

titanic_clasification_model_demo2.ows
Visualize
Model
Evaluate
Test and Score
Predictio...
Confusion Matrix
ROC Analysis
Lift Curve
Calibration Plot
Unsupervised
Prototypes
Tree
A tree algorithm with forward pruning.
more...
File
Impute
Select Columns
Data Table
Data Table (1)
Data Table (2)
Test and Score
Tree Viewer
Tree
Predictions
Data Sampler
Data
Learner
Model → Tree
Model → Predictors
Data Sample → Data
Remaining Data → Data

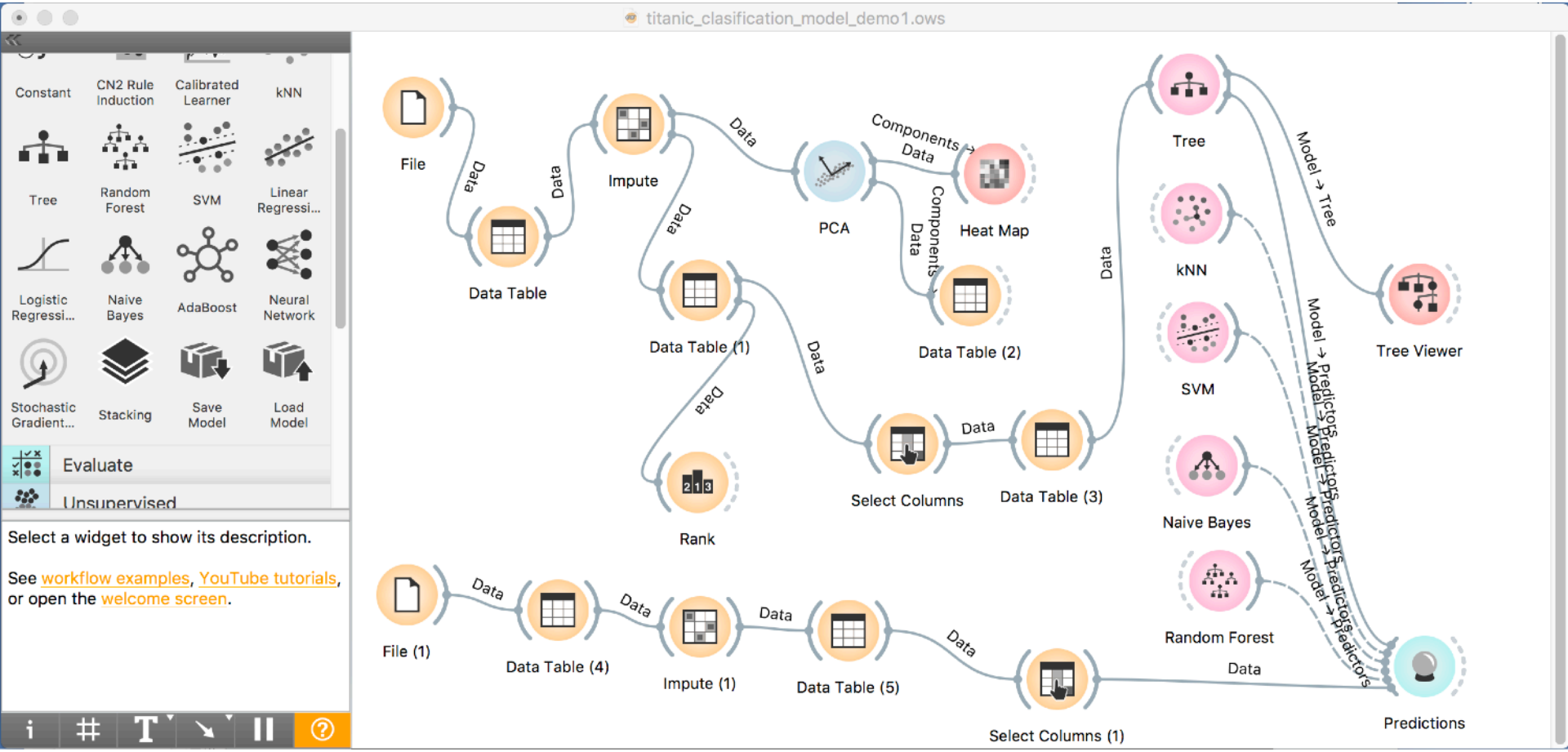

titanic_clasification_model_demo2.ows
Visualize
Model
Evaluate
Test and Score
Predictio...
Confusion Matrix
ROC Analysis
Lift Curve
Calibration Plot
Unsupervised
Prototypes
Test and Score
Cross-validation accuracy estimation.
more...
Test and Score
Sampling
Cross validation
Number of folds: 10
Stratified
Cross validation by feature
Selected
Random sampling
Repeat train/test: 100
Training set size: 80 %
Stratified
Leave one out
Test on train data
Test on test data
Target Class
(Average over classes)
Evaluation Results
Model
AUC
CA
F1
Precision
Recall
Tree
0.851
0.804
0.802
0.804
0.804

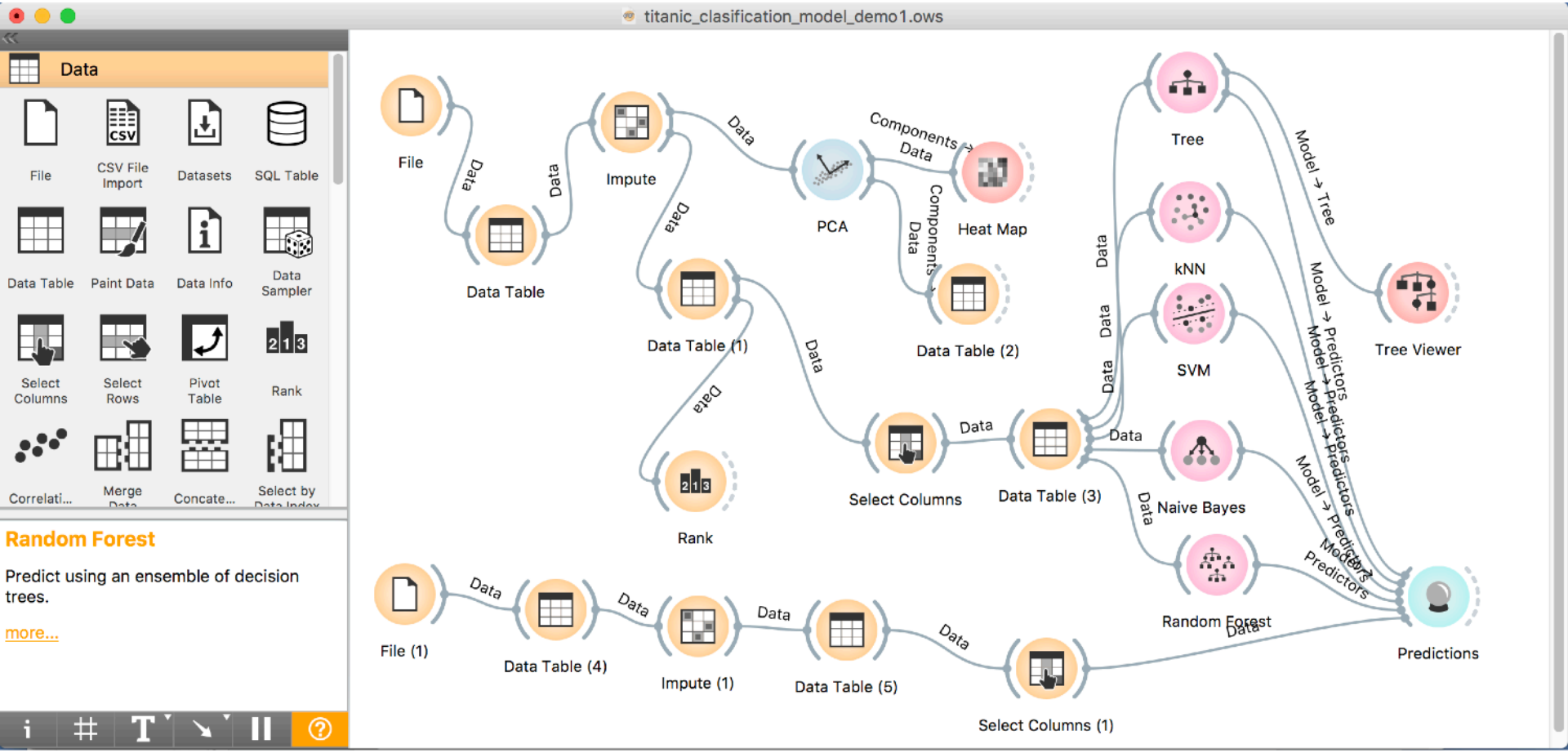
Try out other models for choosing the best fit.

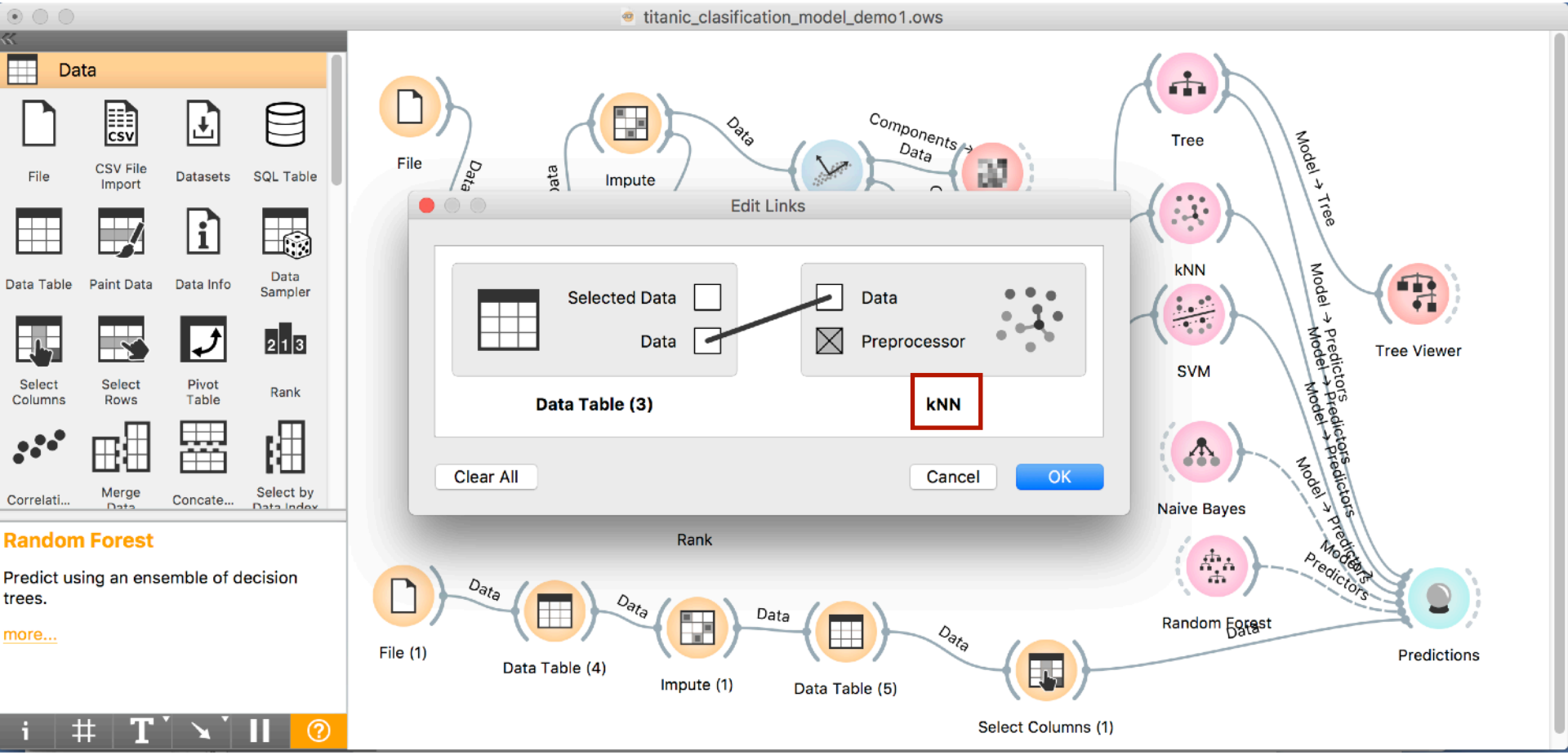

titanic_clasification_model_demo1.ows
Distance File
Distance Matrix
t-SNE
Distance Map
Hierarchi... Clustering
k-Means
Louvain Clustering
DBSCAN
Manifold Learning
PCA
Corresp... Analysis
Distances
Distance Transfor...
MDS
Save Distance...
Self-Org... Map
Prototypes
Image Analytics
Select a widget to show its description.
See workflow examples, YouTube tutorials, or open the welcome screen.
File
Data Table
Impute
Data Table (1)
Rank
PCA
Heat Map
Data Table (2)
Select Columns
Data Table (3)
Tree
Tree Viewer
Predictions
File (1)
Data Table (4)
Impute (1)
Data Table (5)
Select Columns (1)
Data
Components
Model → Tree
Model → Predictors

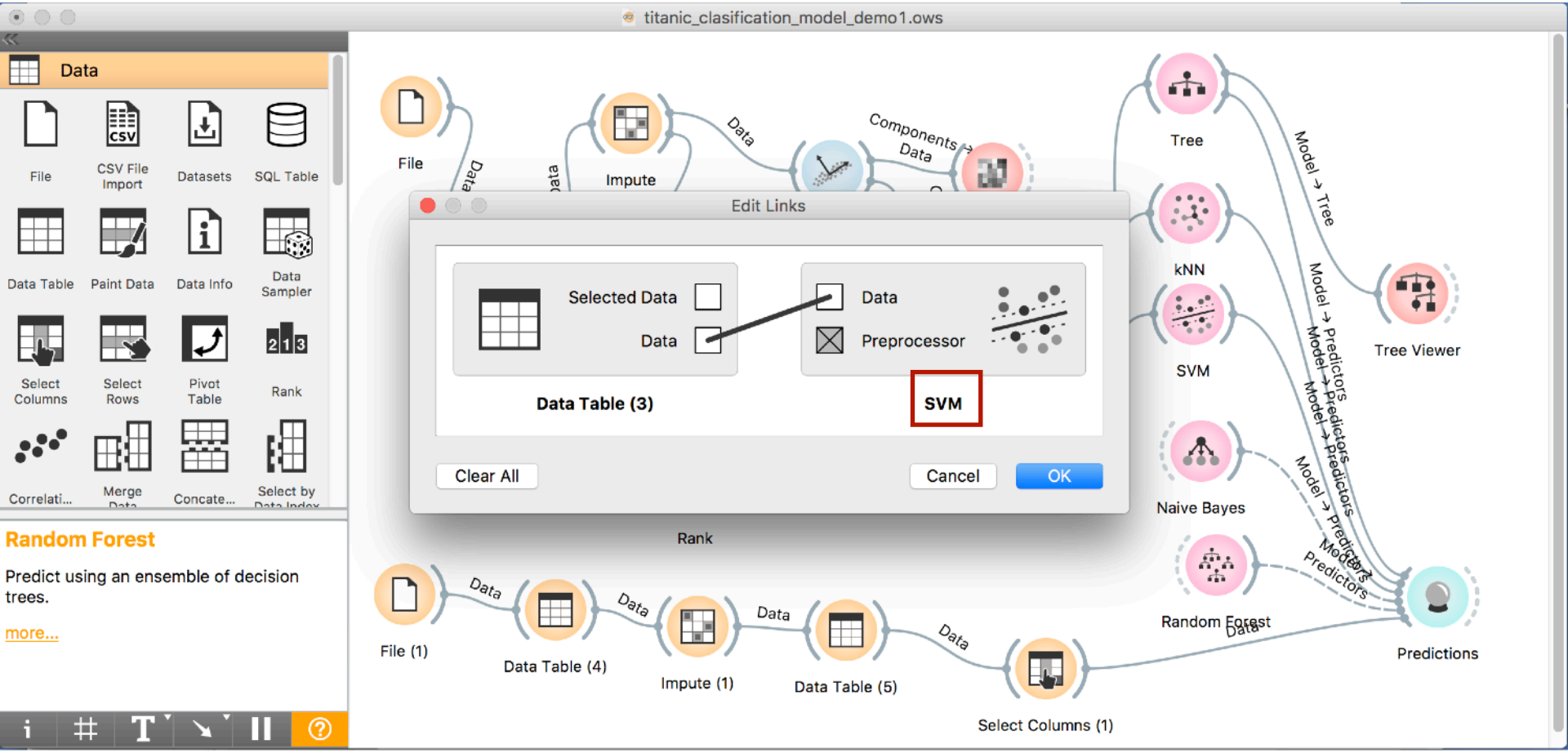

Predictions
Show probabilities for
0
1
Restore Original Order
712
Tree
Survived
Pclass
Sex
Age
1 0.91 : 0.09 0 3 male 22.00
2 0.04 : 0.96 1 1 female 38.00
3 0.17 : 0.83 1 3 female 26.00
4 0.04 : 0.96 1 1 female 35.00
5 0.96 : 0.04 0 3 male 35.00
6 0.88 : 0.12 0 1 male 54.00
7 0.56 : 0.44 0 3 male 2.00
8 0.17 : 0.83 1 3 female 27.00
9 0.07 : 0.93 1 2 female 14.00
10 0.00 : 1.00 1 3 female 4.00
11 0.17 : 0.83 1 1 female 58.00
12 0.83 : 0.17 0 3 male 20.00
13 0.96 : 0.04 0 3 male 39.00
14 0.30 : 0.70 0 3 female 14.00
15 0.04 : 0.96 1 2 female 55.00
16 0.56 : 0.44 0 3 male 2.00
17 0.67 : 0.33 0 3 female 31.00
Model AUC CA F1 Precision Recall
Tree 0.913 0.851 0.848 0.854 0.851
Heat Map
Data Table (2)
Tree
Tree Viewer
Data Table (3)
Select Columns (1)
Predictions
Data
Model → Tree
Model → Predictors

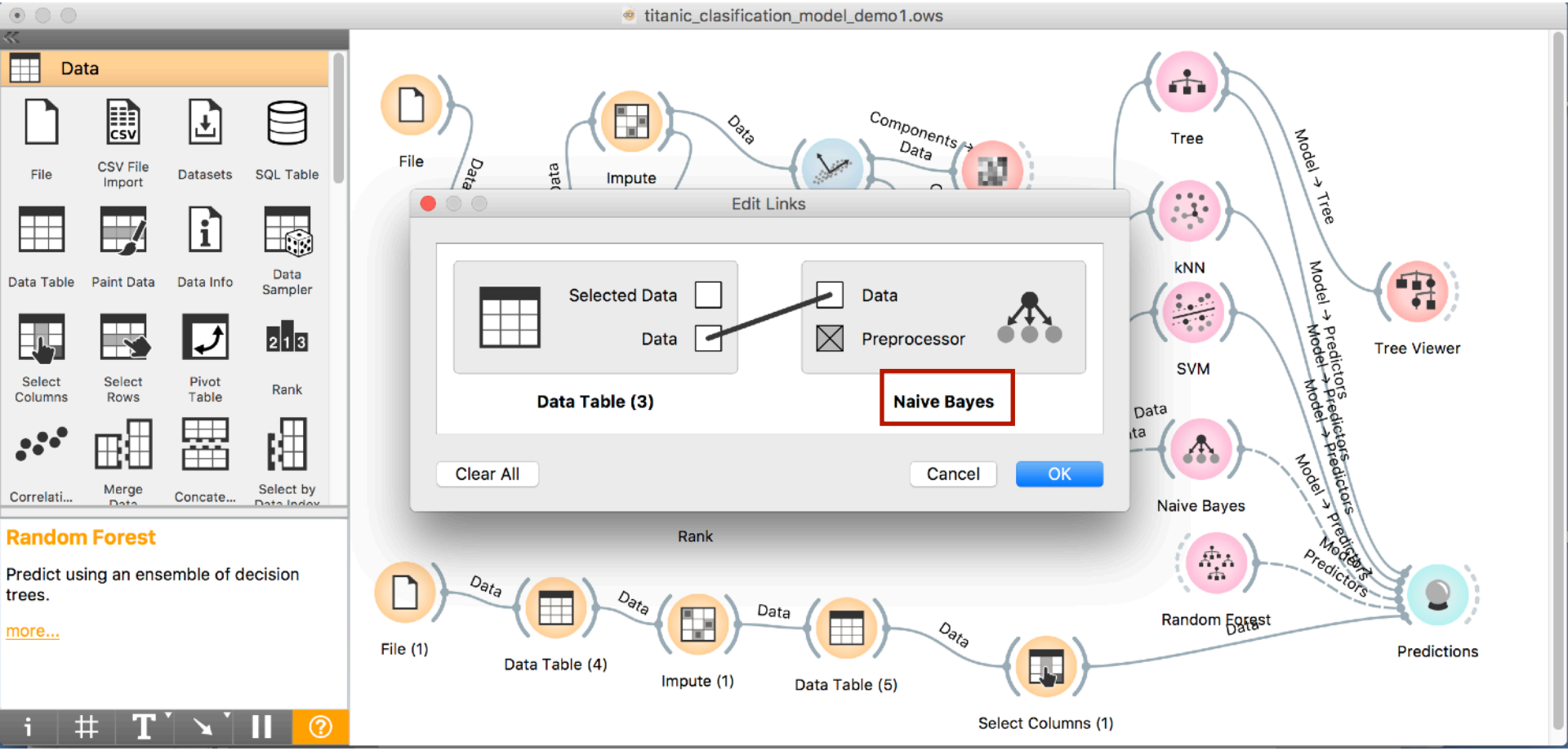

titanic_clasification_model_demo1.ows
Constant
CN2 Rule Induction
Calibrated Learner
kNN
Tree
Random Forest
SVM
Linear Regressi...
Logistic Regressi...
Naive Bayes
AdaBoost
Neural Network
Stochastic Gradient...
Stacking
Save Model
Load Model
Evaluate
Unsupervised
Select a widget to show its description.
See workflow examples, YouTube tutorials, or open the welcome screen.
File
Data Table
Impute
PCA
Heat Map
Data Table (1)
Data Table (2)
Rank
Select Columns
Data Table (3)
Tree
kNN
SVM
Naive Bayes
Random Forest
Tree Viewer
Predictions
File (1)
Data Table (4)
Impute (1)
Data Table (5)
Select Columns (1)
Data
Components
Model → Tree
Model → Predictors

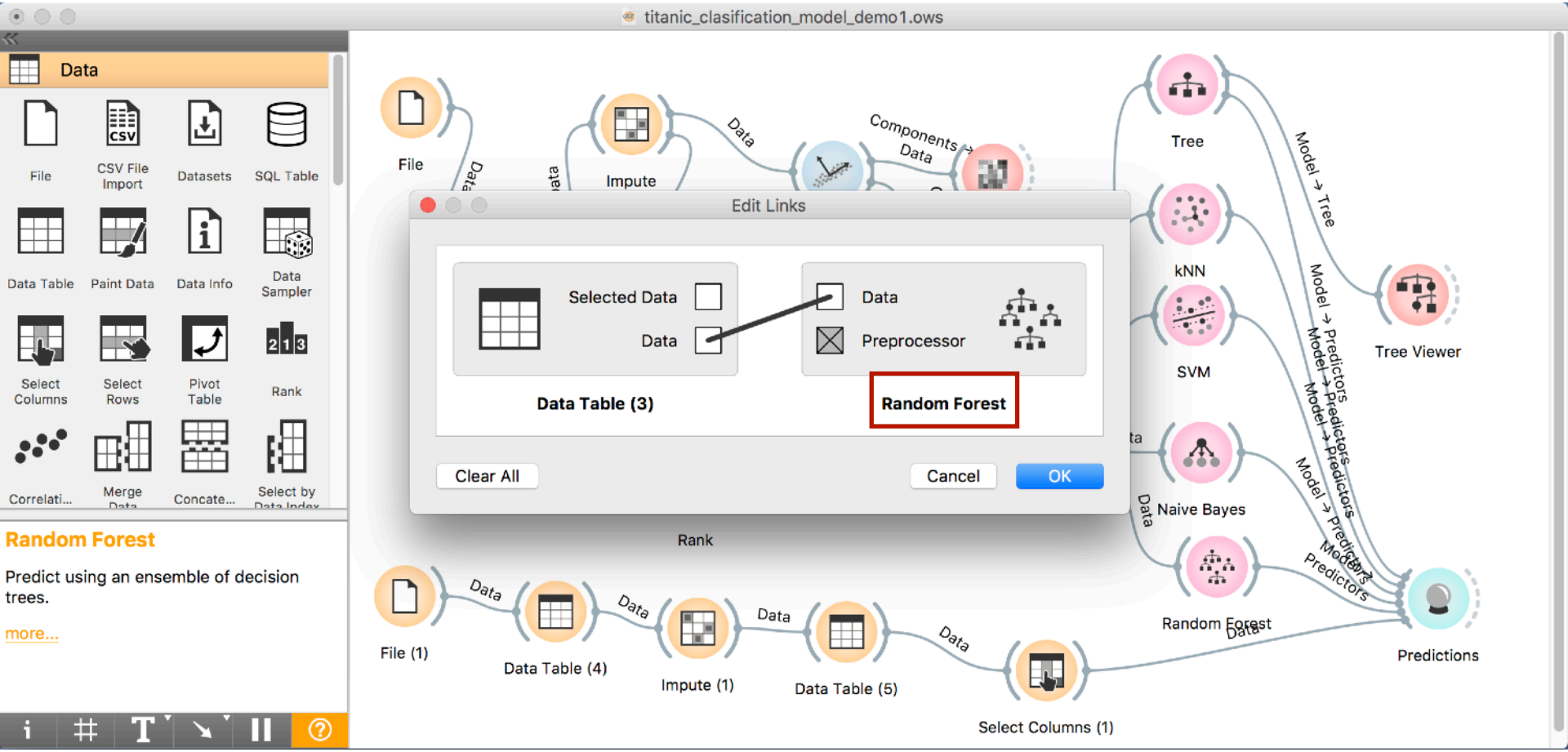

titanic_clasification_model_demo1.ows
Constant
CN2 Rule Induction
Calibrated Learner
kNN
Tree
Random Forest
SVM
Linear Regressi...
Logistic Regressi...
Naive Bayes
AdaBoost
Neural Network
Stochastic Gradient...
Stacking
Save Model
Load Model
Evaluate
Unsupervised
Select a widget to show its description.
See workflow examples, YouTube tutorials, or open the welcome screen.
File
Data Table
Impute
PCA
Heat Map
Data Table (1)
Data Table (2)
Rank
Select Columns
Data Table (3)
Tree
kNN
SVM
Naive Bayes
Random Forest
Tree Viewer
Predictions
File (1)
Data Table (4)
Impute (1)
Data Table (5)
Select Columns (1)
Data
Components
Model → Tree
Model → Predictors

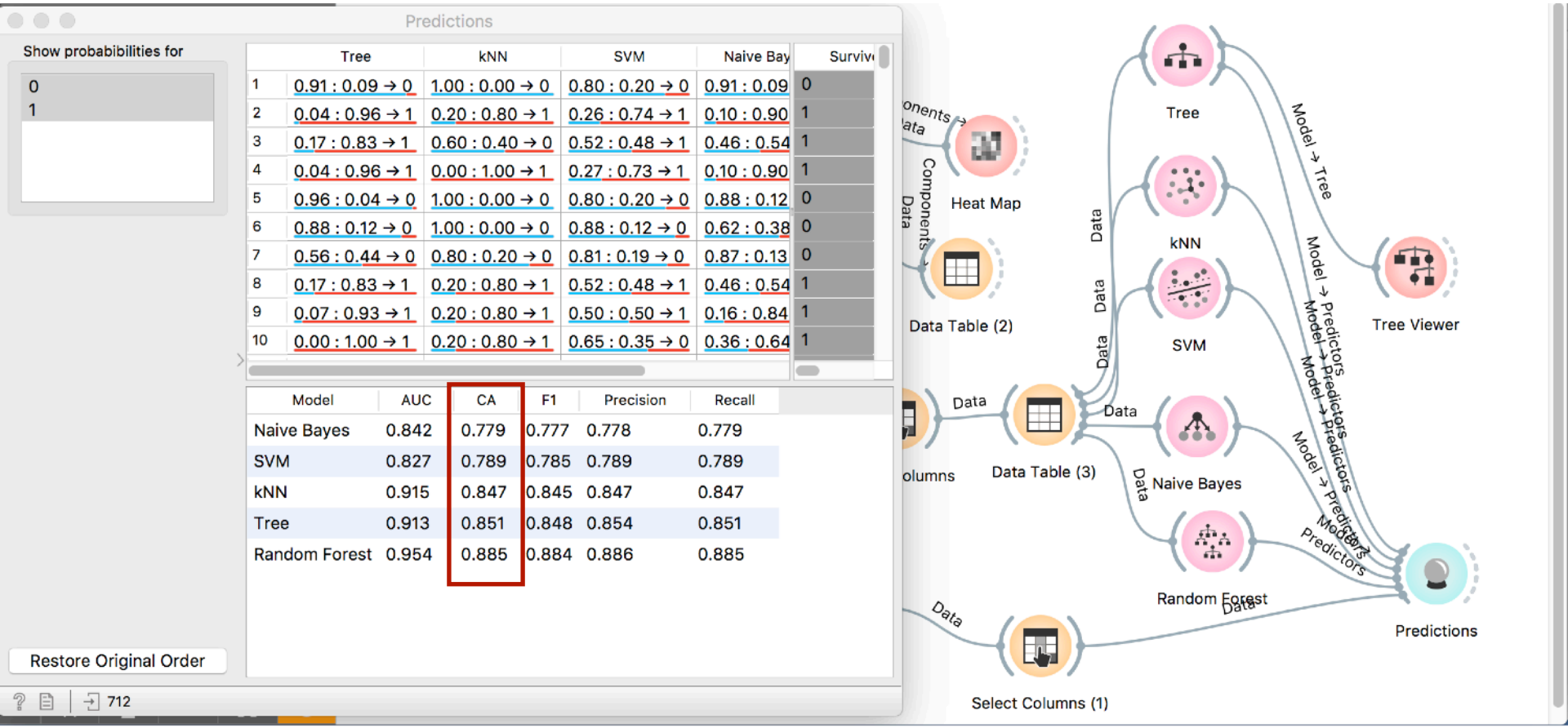

titanic_clasification_model_demo1.ows
Data
File
CSV File Import
Datasets
SQL Table
Data Table
Paint Data
Data Info
Data Sampler
Select Columns
Select Rows
Pivot Table
Rank
Correlati...
Merge Data
Concate...
Select by Data Index
Random Forest
Predict using an ensemble of decision trees.
more...
File
Data Table
Impute
PCA
Heat Map
Data Table (1)
Data Table (2)
Rank
Select Columns
Data Table (3)
Tree
kNN
SVM
Naive Bayes
Random Forest
Tree Viewer
Predictions
File (1)
Data Table (4)
Impute (1)
Data Table (5)
Select Columns (1)
Data
Components
Model → Tree
Model → Predictors

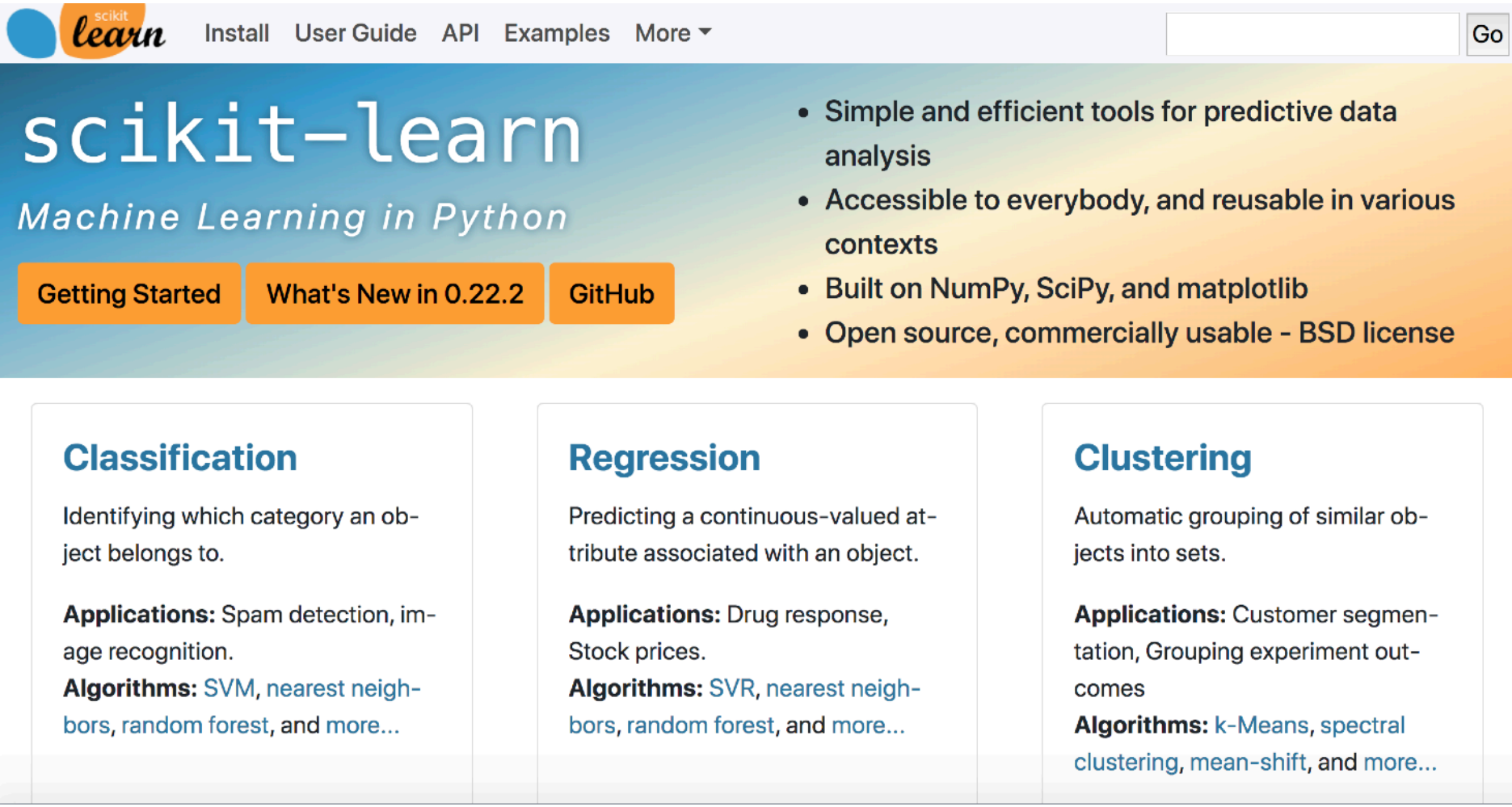

titanic_clasification_model_demo1.ows
Data
File
CSV File Import
Datasets
SQL Table
Data Table
Paint Data
Data Info
Data Sampler
Select Columns
Select Rows
Pivot Table
Rank
Correlati...
Merge Data
Concate...
Select by Data Index
Random Forest
Predict using an ensemble of decision trees.
more...
File
Impute
Data
Components
Edit Links
Selected Data
Data
Data
Preprocessor
Data Table (3)
kNN
Clear All
Cancel
OK
Rank
Tree
kNN
SVM
Naive Bayes
Random Forest
Tree Viewer
Model → Tree
Model → Predictors
Predictions
File (1)
Data Table (4)
Impute (1)
Data Table (5)
Select Columns (1)


titanic_clasification_model_demo1.ows
Data
File
CSV File Import
Datasets
SQL Table
Data Table
Paint Data
Data Info
Data Sampler
Select Columns
Select Rows
Pivot Table
Rank
Correlati...
Merge Data
Concate...
Select by Data Index
Random Forest
Predict using an ensemble of decision trees.
more...
File
Impute
Data
Components
Edit Links
Selected Data
Data
Data
Preprocessor
Data Table (3)
SVM
Clear All
Cancel
OK
Rank
Tree
kNN
SVM
Naive Bayes
Random Forest
Tree Viewer
Predictions
Model → Tree
Model → Predictors
File (1)
Data Table (4)
Impute (1)
Data Table (5)
Select Columns (1)


titanic_clasification_model_demo1.ows
Data
File
CSV File Import
Datasets
SQL Table
Data Table
Paint Data
Data Info
Data Sampler
Select Columns
Select Rows
Pivot Table
Rank
Correlati...
Merge Data
Concate...
Select by Data Index
Random Forest
Predict using an ensemble of decision trees.
more...
File
Impute
Data
Components
Edit Links
Selected Data
Data
Data
Preprocessor
Data Table (3)
Naive Bayes
Clear All
Cancel
OK
Rank
Tree
kNN
SVM
Naive Bayes
Random Forest
Tree Viewer
Predictions
Model → Tree
Model → Predictors
File (1)
Data Table (4)
Impute (1)
Data Table (5)
Select Columns (1)


titanic_clasification_model_demo1.ows
Data
File
CSV File Import
Datasets
SQL Table
Data Table
Paint Data
Data Info
Data Sampler
Select Columns
Select Rows
Pivot Table
Rank
Correlati...
Merge Data
Concate...
Select by Data Index
Random Forest
Predict using an ensemble of decision trees.
more...
Edit Links
Selected Data
Data
Data
Preprocessor
Data Table (3)
Random Forest
Clear All
Cancel
OK
Rank
File
Impute
Components
Tree
kNN
SVM
Naive Bayes
Random Forest
Tree Viewer
Predictions
Model → Tree
Model → Predictors
File (1)
Data Table (4)
Impute (1)
Data Table (5)
Select Columns (1)


Predictions
Show probabilities for
0
1
Restore Original Order
712
Tree
kNN
SVM
Naive Bay
Survive
1 0.91 : 0.09 → 0 1.00 : 0.00 → 0 0.80 : 0.20 → 0 0.91 : 0.09 0
2 0.04 : 0.96 → 1 0.20 : 0.80 → 1 0.26 : 0.74 → 1 0.10 : 0.90 1
3 0.17 : 0.83 → 1 0.60 : 0.40 → 0 0.52 : 0.48 → 1 0.46 : 0.54 1
4 0.04 : 0.96 → 1 0.00 : 1.00 → 1 0.27 : 0.73 → 1 0.10 : 0.90 1
5 0.96 : 0.04 → 0 1.00 : 0.00 → 0 0.80 : 0.20 → 0 0.88 : 0.12 0
6 0.88 : 0.12 → 0 1.00 : 0.00 → 0 0.88 : 0.12 → 0 0.62 : 0.38 0
7 0.56 : 0.44 → 0 0.80 : 0.20 → 0 0.81 : 0.19 → 0 0.87 : 0.13 0
8 0.17 : 0.83 → 1 0.20 : 0.80 → 1 0.52 : 0.48 → 1 0.46 : 0.54 1
9 0.07 : 0.93 → 1 0.20 : 0.80 → 1 0.50 : 0.50 → 1 0.16 : 0.84 1
10 0.00 : 1.00 → 1 0.20 : 0.80 → 1 0.65 : 0.35 → 0 0.36 : 0.64 1
Model AUC CA F1 Precision Recall
Naive Bayes 0.842 0.779 0.777 0.778 0.779
SVM 0.827 0.789 0.785 0.789 0.789
kNN 0.915 0.847 0.845 0.847 0.847
Tree 0.913 0.851 0.848 0.854 0.851
Random Forest 0.954 0.885 0.884 0.886 0.885
Heat Map
Data Table (2)
Data Table (3)
Select Columns (1)
Tree
kNN
SVM
Naive Bayes
Random Forest
Tree Viewer
Predictions
Data
Model → Tree
Model → Predictors


scikit learn
Install
User Guide
API
Examples
More
Go
scikit-learn
Machine Learning in Python
Getting Started
What's New in 0.22.2
GitHub
Simple and efficient tools for predictive data analysis
Accessible to everybody, and reusable in various contexts
Built on NumPy, SciPy, and matplotlib
Open source, commercially usable - BSD license
Classification
Identifying which category an object belongs to.
Applications: Spam detection, image recognition.
Algorithms: SVM, nearest neighbors, random forest, and more...
Regression
Predicting a continuous-valued attribute associated with an object.
Applications: Drug response, Stock prices.
Algorithms: SVR, nearest neighbors, random forest, and more...
Clustering
Automatic grouping of similar objects into sets.
Applications: Customer segmentation, Grouping experiment outcomes
Algorithms: k-Means, spectral clustering, mean-shift, and more...

Prev Up Next

scikit-learn 0.22.2
Other versions

Please cite us if you use the software.

7. Dataset loading utilities
7.1. General dataset API
7.2. Toy datasets
7.3. Real world datasets
7.4. Generated datasets
7.5. Loading other datasets

7.2. Toy datasets

scikit-learn comes with a few small standard datasets that do not require to download any file from some external website.

They can be loaded using the following functions:

`load_boston`([return_X_y])	Load and return the boston house-prices dataset (regression).
`load_iris`([return_X_y])	Load and return the iris dataset (classification).
`load_diabetes`([return_X_y])	Load and return the diabetes dataset (regression).
`load_digits`([n_class, return_X_y])	Load and return the digits dataset (classification).
`load_linnerud`([return_X_y])	Load and return the linnerud dataset (multivariate regression).
`load_wine`([return_X_y])	Load and return the wine dataset (classification).
`load_breast_cancer`([return_X_y])	Load and return the breast cancer wisconsin dataset (classification).

These datasets are useful to quickly illustrate the behavior of the various algorithms implemented in scikit-learn. They are however often too small to be representative of real world machine learning tasks.

7.2.1. Boston house prices dataset

Data Set Characteristics:

Toggle Menu

Let’s train the Titanic classifier in Python Codes with Scikit-learn.

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Code

Building and Visualizing a Decision Tree Classifier for the Titanic Dataset

In [68]:

```
# Import Pandas, Scikit-learn, Graphviz modules
from sklearn import tree
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.tree import export_graphviz
from sklearn.metrics import classification_report
from sklearn.metrics import accuracy_score
import pandas as pd
import pydotplus
from IPython.display import Image
```

In [69]:

```
# Import Titanic dataset
data = pd.read_csv('http://localhost/datasets/titanic.csv')
# Check the first 5 entries
data.head(5)
```

Out[69]:

	Pclass	Sex	Age	Survived	Name	Ticket	Cabin	Selected
0	continuous	discrete	continuous	discrete	string	string	string	discrete
1	NaN	NaN	NaN	class	meta	meta	meta	meta
2	3.0	male	22.0	0	Braund, Mr. Owen Harris	A/5 21171	NaN	No
3	1.0	female	38.0	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	PC 17599	C85	No
4	3.0	female	26.0	1	Heikkinen, Miss. Laina	STON/O2. 3101282	NaN	No

In [70]:

```
# Convert Male and Female into binary values of 0 and 1.
```

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Code

In [70]:

```
# Convert Male and Female into binary values of 0 and 1.
data.Sex[data.Sex == 'male'] = 1
data.Sex[data.Sex == 'female'] = 0
print(data)
```

```
        Pclass       Sex         Age  Survived  \
0    continuous  discrete  continuous  discrete
1           NaN       NaN         NaN     class
2           3.0         1        22.0         0
3           1.0         0        38.0         1
4           3.0         0        26.0         1
..          ...       ...         ...       ...
888         2.0         1        27.0         0
889         1.0         0        19.0         1
890         3.0         0         NaN         0
891         1.0         1        26.0         1
892         3.0         1        32.0         0

                                                Name            Ticket  \
0                                             string            string
1                                               meta              meta
2                                  Braund, Mr. Owen Harris         A/5 21171
3    Cumings, Mrs. John Bradley (Florence Briggs Th...          PC 17599
4                                Heikkinen, Miss. Laina  STON/O2. 3101282
..                                                  ...             ...
888                              Montvila, Rev. Juozas           211536
889                       Graham, Miss. Margaret Edith           112053
890            Johnston, Miss. Catherine Helen "Carrie"        W./C. 6607
891                              Behr, Mr. Karl Howell           111369
892                                Dooley, Mr. Patrick           370376

      Cabin  Selected
0    string  discrete
```

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Code

In [71]:
```
# Check the first 5 items.
data.head(5)
```

Out[71]:

	Pclass	Sex	Age	Survived	Name	Ticket	Cabin	Selected
0	continuous	discrete	continuous	discrete	string	string	string	discrete
1	NaN	NaN	NaN	class	meta	meta	meta	meta
2	3.0	1	22.0	0	Braund, Mr. Owen Harris	A/5 21171	NaN	No
3	1.0	0	38.0	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	PC 17599	C85	No
4	3.0	0	26.0	1	Heikkinen, Miss. Laina	STON/O2. 3101282	NaN	No

In [72]:
```
# Check frame structure
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 893 entries, 0 to 892
Data columns (total 8 columns):
Pclass      892 non-null object
Sex         892 non-null object
Age         715 non-null object
Survived    893 non-null object
Name        893 non-null object
Ticket      893 non-null object
Cabin       206 non-null object
Selected    893 non-null object
dtypes: object(8)
memory usage: 55.9+ KB
```

In [74]:
```
# Check number of rows and columns
data.shape
```

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Code

In [74]:
```
# Check number of rows and columns
data.shape
```

Out[74]: (893, 8)

In [55]:
```
# Slice the frame into selected features and target variable
data=data.loc[1:,'Pclass':'Survived']
data.head(5)
```

Out[55]:

	Pclass	Sex	Age	Survived
1	NaN	NaN	NaN	class
2	3.0	1	22.0	0
3	1.0	0	38.0	1
4	3.0	0	26.0	1
5	1.0	0	35.0	1

In [56]:
```
# Check null values
data.isnull().sum()
```

Out[56]:
```
Pclass        1
Sex           1
Age         178
Survived      0
dtype: int64
```

In [57]:
```
# Clean up by dropping the null entries
dataset = data.dropna()
```

In [58]:
```
# Ensure it is cleaned.
```

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Code

In [58]:
```
# Ensure it is cleaned.
dataset.isnull().sum()
```

Out[58]:
```
Pclass      0
Sex         0
Age         0
Survived    0
dtype: int64
```

In [59]:
```
# Check first 5 entries
dataset.head(5)
```

Out[59]:

	Pclass	Sex	Age	Survived
2	3.0	1	22.0	0
3	1.0	0	38.0	1
4	3.0	0	26.0	1
5	1.0	0	35.0	1
6	3.0	1	35.0	0

In [60]:
```
# Define x (features) and y (label/classifier)
x = dataset.iloc[:,:-1].values
y = dataset['Survived']
```

In [61]:
```
# Split the dataset into traning and testing sets (20% for testing)
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.20, random_state=27)
```

In [75]:
```
# Training/fitting the  model with the Decision Tree Classifier
tree_model = tree.DecisionTreeClassifier()
tree_model.fit(x_train,y_train)
```

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Code

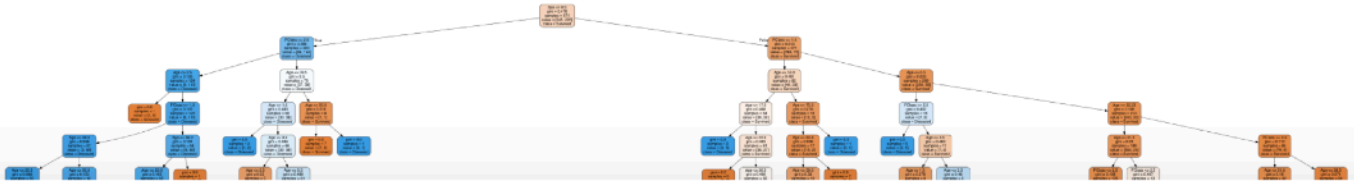
In [75]:
```
# Training/fitting the  model with the Decision Tree Classifier
tree_model = tree.DecisionTreeClassifier()
tree_model.fit(x_train,y_train)
```

Out[75]:
```
DecisionTreeClassifier(class_weight=None, criterion='gini', max_depth=None,
                       max_features=None, max_leaf_nodes=None,
                       min_impurity_decrease=0.0, min_impurity_split=None,
                       min_samples_leaf=1, min_samples_split=2,
                       min_weight_fraction_leaf=0.0, presort=False,
                       random_state=None, splitter='best')
```

In [76]:
```
# Define the feature and target variable names for displaying in the tree
titanic_feature_names = ['PClass',
 'Sex',
 'Age']
titanic_target_names = ['Survived', 'Diseased']
```

In [77]:
```
# Generate the tree
dot_data = tree.export_graphviz(tree_model, out_file=None,
                                feature_names=titanic_feature_names,
                                class_names=titanic_target_names,
                                rounded=True, filled=True)
# Draw graph
graph = pydotplus.graph_from_dot_data(dot_data)
# Show graph
Image(graph.create_png())
```

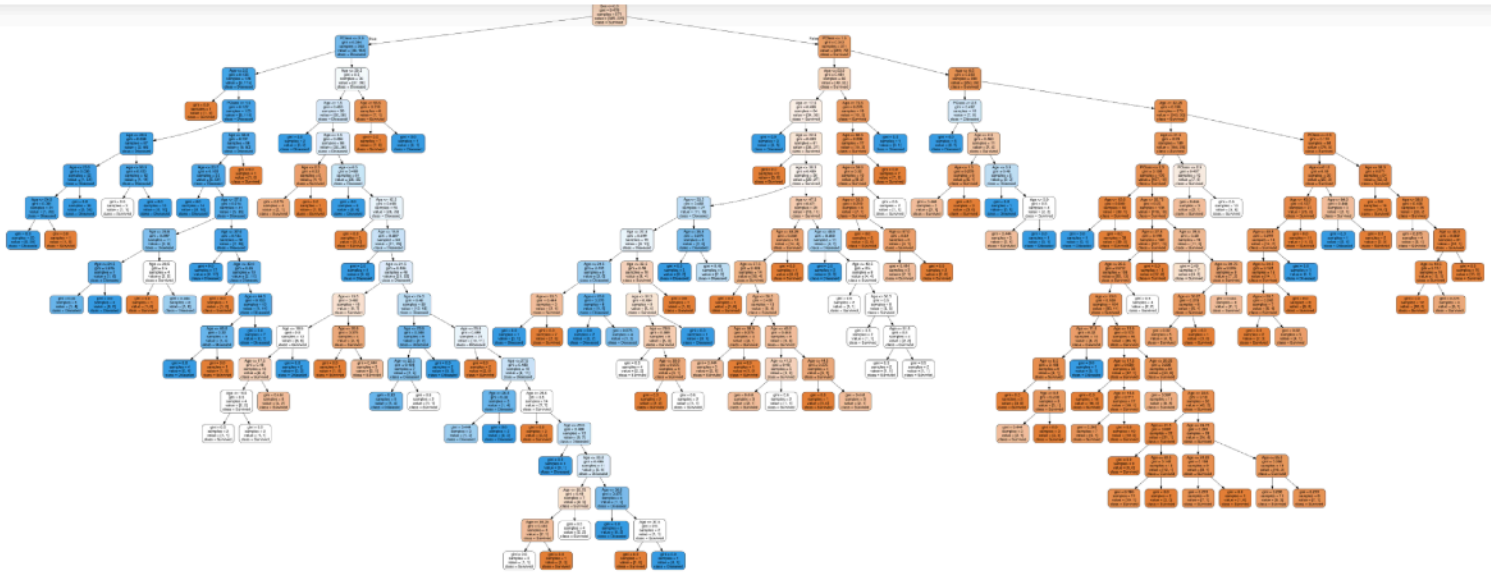
Out[77]:



File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Code

Out[77]:



In [66]:

```
# Evaluating the model by using the trained model against the testing dataset
clf_prediction = tree_model.predict(x_test)
```

In [67]:

```
# Display the accuarcy score
print(accuracy_score(clf_prediction, y_test))
```

```
0.7972027972027972
```

Adding the other models

In [79]:
```
# Import Pandas, Scikit-learn, Graphviz modules
from sklearn import tree
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.tree import export_graphviz
from sklearn.metrics import classification_report
from sklearn.metrics import accuracy_score
import pandas as pd
import pydotplus
from IPython.display import Image
from sklearn.neighbors import KNeighborsClassifier # K-Nearest Neibour Algorithm
from sklearn.naive_bayes import GaussianNB         # Gaussian Naive Bayes Algorithm
from sklearn.svm import SVC                        # Support Vector Machine Algorithm
from sklearn.ensemble import RandomForestClassifier # Random Algorithm
```

In [80]:
```
# Import Titanic dataset
data = pd.read_csv('http://localhost/datasets/titanic.csv')
# Check the first 5 entries
data.head(5)
```

Out[80]:

	Pclass	Sex	Age	Survived	Name	Ticket	Cabin	Selected
0	continuous	discrete	continuous	discrete	string	string	string	discrete
1	NaN	NaN	NaN	class	meta	meta	meta	meta
2	3.0	male	22.0	0	Braund, Mr. Owen Harris	A/5 21171	NaN	No
3	1.0	female	38.0	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	PC 17599	C85	No

In [80]:

```
# Import Titanic dataset
data = pd.read_csv('http://localhost/datasets/titanic.csv')
# Check the first 5 entries
data.head(5)
```

Out[80]:

	Pclass	Sex	Age	Survived	Name	Ticket	Cabin	Selected
0	continuous	discrete	continuous	discrete	string	string	string	discrete
1	NaN	NaN	NaN	class	meta	meta	meta	meta
2	3.0	male	22.0	0	Braund, Mr. Owen Harris	A/5 21171	NaN	No
3	1.0	female	38.0	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	PC 17599	C85	No
4	3.0	female	26.0	1	Heikkinen, Miss. Laina	STON/O2. 3101282	NaN	No

In [81]:

```
# Convert Male and Female into binary values of 0 and 1.
data.Sex[data.Sex == 'male'] = 1
data.Sex[data.Sex == 'female'] = 0
print(data)
```

```
        Pclass       Sex         Age  Survived  \
0   continuous  discrete  continuous  discrete
1          NaN       NaN         NaN     class
2          3.0         1        22.0         0
3          1.0         0        38.0         1
4          3.0         0        26.0         1
..         ...       ...         ...       ...
888        2.0         1        27.0         0
889        1.0         0        19.0         1
890        3.0         0         NaN         0
891        1.0         1        26.0         1
892        3.0         1        32.0         0
```

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Code

In [82]:
```
# Check the first 5 items.
data.head(5)
```

Out[82]:

	Pclass	Sex	Age	Survived	Name	Ticket	Cabin	Selected
0	continuous	discrete	continuous	discrete	string	string	string	discrete
1	NaN	NaN	NaN	class	meta	meta	meta	meta
2	3.0	1	22.0	0	Braund, Mr. Owen Harris	A/5 21171	NaN	No
3	1.0	0	38.0	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	PC 17599	C85	No
4	3.0	0	26.0	1	Heikkinen, Miss. Laina	STON/O2. 3101282	NaN	No

In [83]:
```
# Slice the frame into selected features and target variable
data=data.loc[1:,'Pclass':'Survived']
data.head(5)
```

Out[83]:

	Pclass	Sex	Age	Survived
1	NaN	NaN	NaN	class
2	3.0	1	22.0	0
3	1.0	0	38.0	1
4	3.0	0	26.0	1
5	1.0	0	35.0	1

In [84]:
```
# Check null values
data.isnull().sum()
```

Out[84]:
```
Pclass          1
Sex             1
```

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Code

In [84]:
```
# Check null values
data.isnull().sum()
```

Out[84]:
```
Pclass        1
Sex           1
Age         178
Survived      0
dtype: int64
```

In [85]:
```
# Ensure it is cleaned.
dataset.isnull().sum()
```

Out[85]:
```
Pclass      0
Sex         0
Age         0
Survived    0
dtype: int64
```

In [86]:
```
# Check first 5 entries
dataset.head(5)
```

Out[86]:

	Pclass	Sex	Age	Survived
2	3.0	1	22.0	0
3	1.0	0	38.0	1
4	3.0	0	26.0	1
5	1.0	0	35.0	1
6	3.0	1	35.0	0

In [87]:
```
# Define x (features) and y (label/classifier)
```

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Code

In [87]:
```
# Define x (features) and y (label/classifier)
x = dataset.iloc[:,:-1].values
y = dataset['Survived']
```

In [88]:
```
# Split the dataset into traning and testing sets (20% for testing)
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.20, random_state=27)
```

In [90]:
```
# Training/Fitting with Different Classifiers Setup
# Decision Tree Classifier
tree_model = tree.DecisionTreeClassifier()
# Support Vector Machine Classifier
SVC_model = SVC(gamma='scale')
# KNN model requires you to specify n_neighbors,
# the number of points the classifier will look at to determine what class a new point belongs to
KNN_model = KNeighborsClassifier(n_neighbors=5)
# KNN_model = KNeighborsClassifier(algorithm='scale', leaf_size=30, metric='minkowski',
#                      metric_params=None, n_jobs=None, n_neighbors=5, p=2)
# Gaussian Naiive Bayne model
GNB_model = GaussianNB()
# Random Forest
# Create the model with 100 trees
RF_model = RandomForestClassifier(n_estimators=100,
                                  bootstrap = True,
                                  max_features = 'sqrt')
```

In [91]:
```
# Start training/fitting
tree_model.fit(x_train,y_train)
SVC_model.fit(x_train, y_train)
KNN_model.fit(x_train, y_train)
GNB_model.fit(x_train, y_train)
RF_model.fit(x_train, y_train)
```

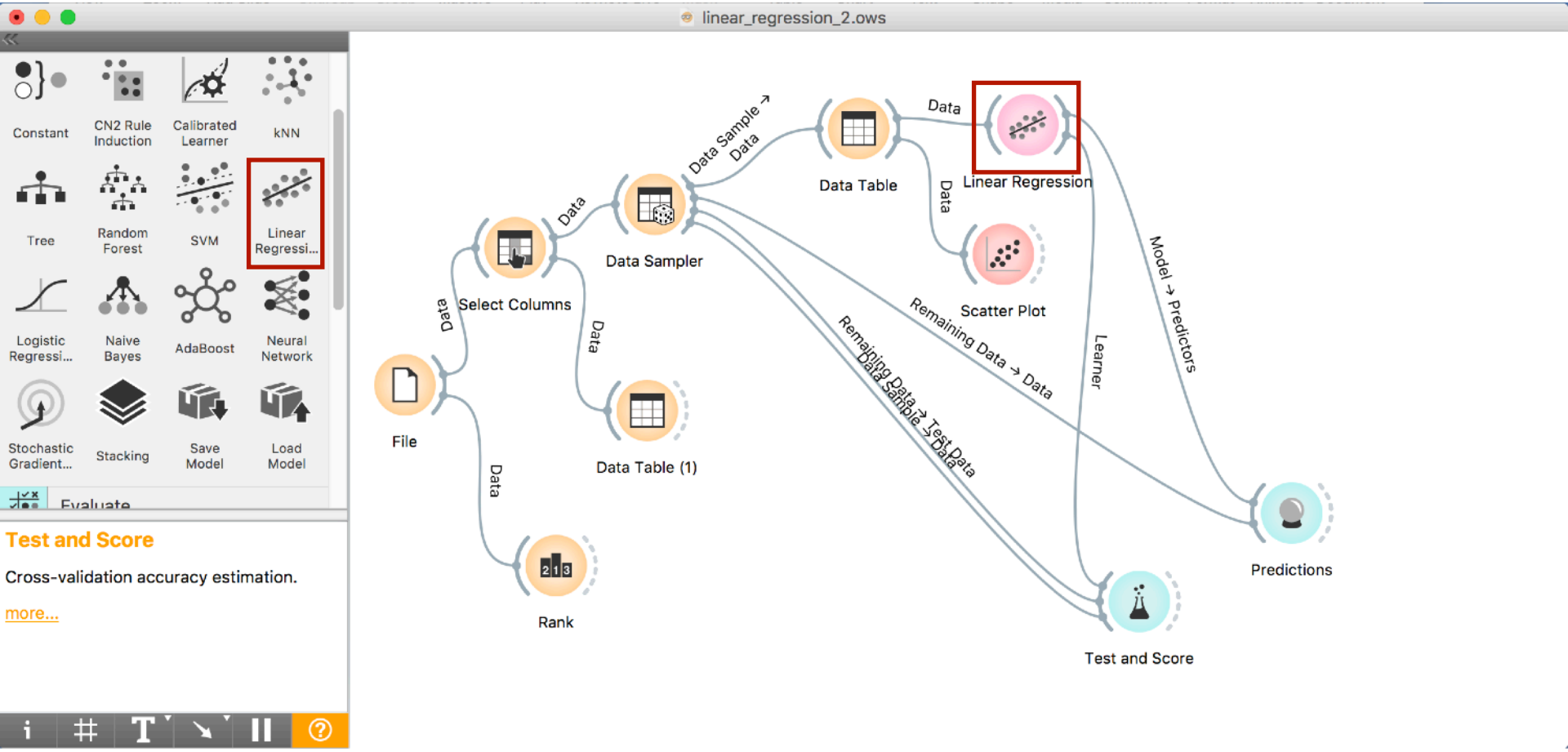
In [93]:
```
# Training completed. Use models to predict with the test data and store the outcome
SVC_prediction = SVC_model.predict(x_test)
KNN_prediction = KNN_model.predict(x_test)
GNB_prediction = GNB_model.predict(x_test)
TREE_prediction = tree_model.predict(x_test)
RF_prediction = RF_model.predict(x_test)
```

In [95]:
```
# Accuracy score is the simplest way to evaluate
print("SVC")
print(accuracy_score(SVC_prediction, y_test))
print("KNN")
print(accuracy_score(KNN_prediction, y_test))
print("GNB")
print(accuracy_score(GNB_prediction, y_test))
print("TREE")
print(accuracy_score(TREE_prediction, y_test))
print("RF")
print(accuracy_score(RF_prediction, y_test))
```

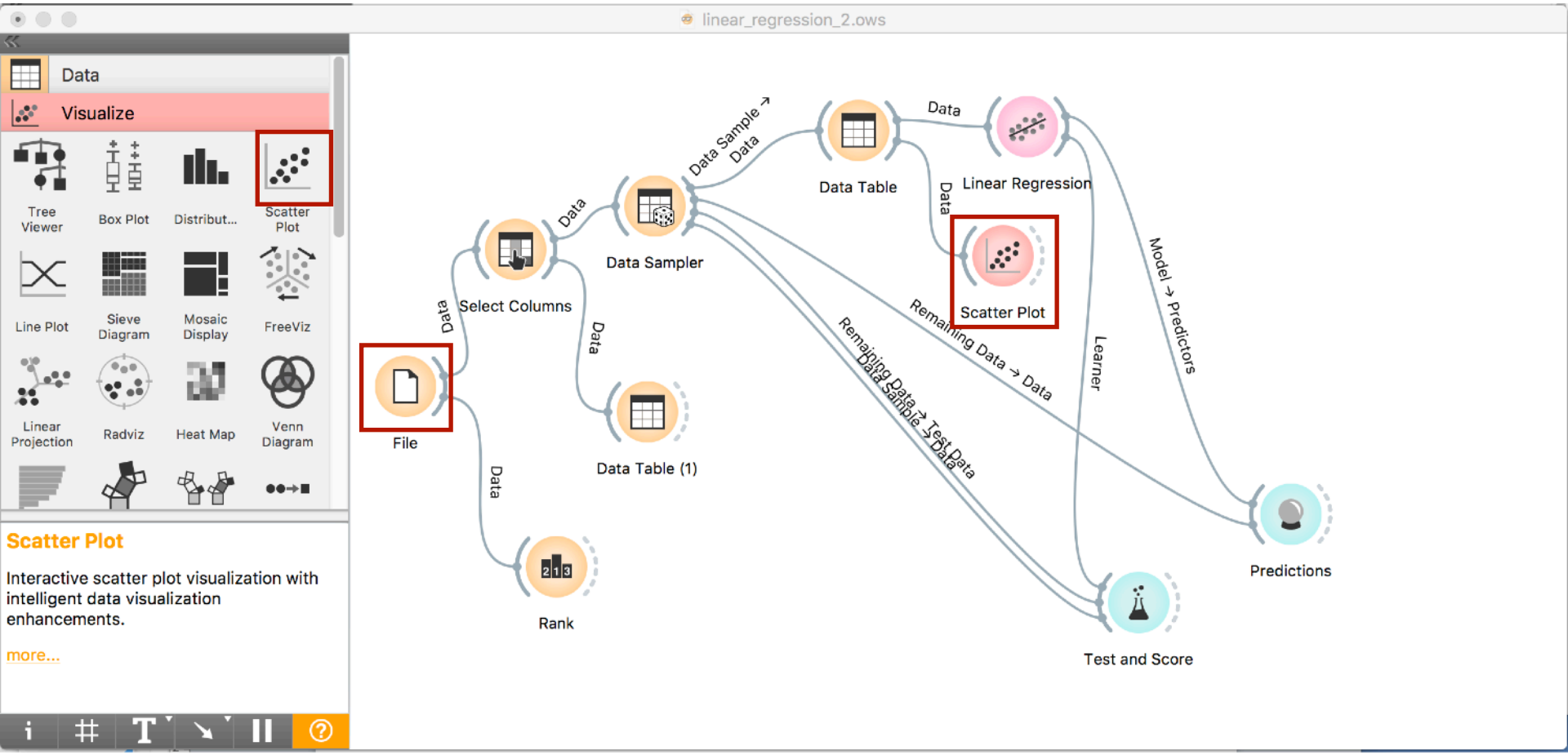
```
SVC
0.6083916083916084
KNN
0.8041958041958042
GNB
0.7272727272727273
TREE
0.7972027972027972
RF
0.8181818181818182
```

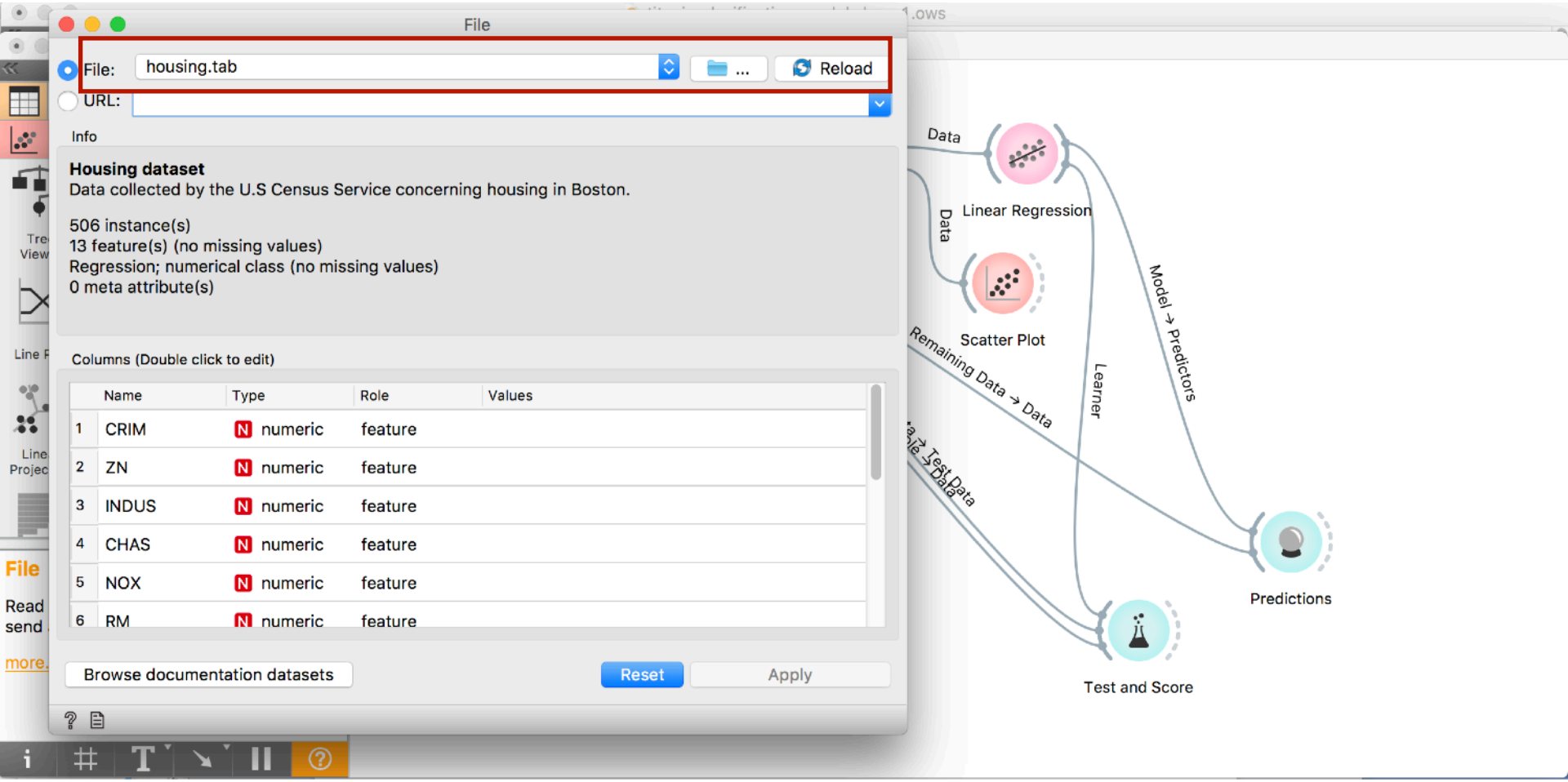
In [ ]:
```

```

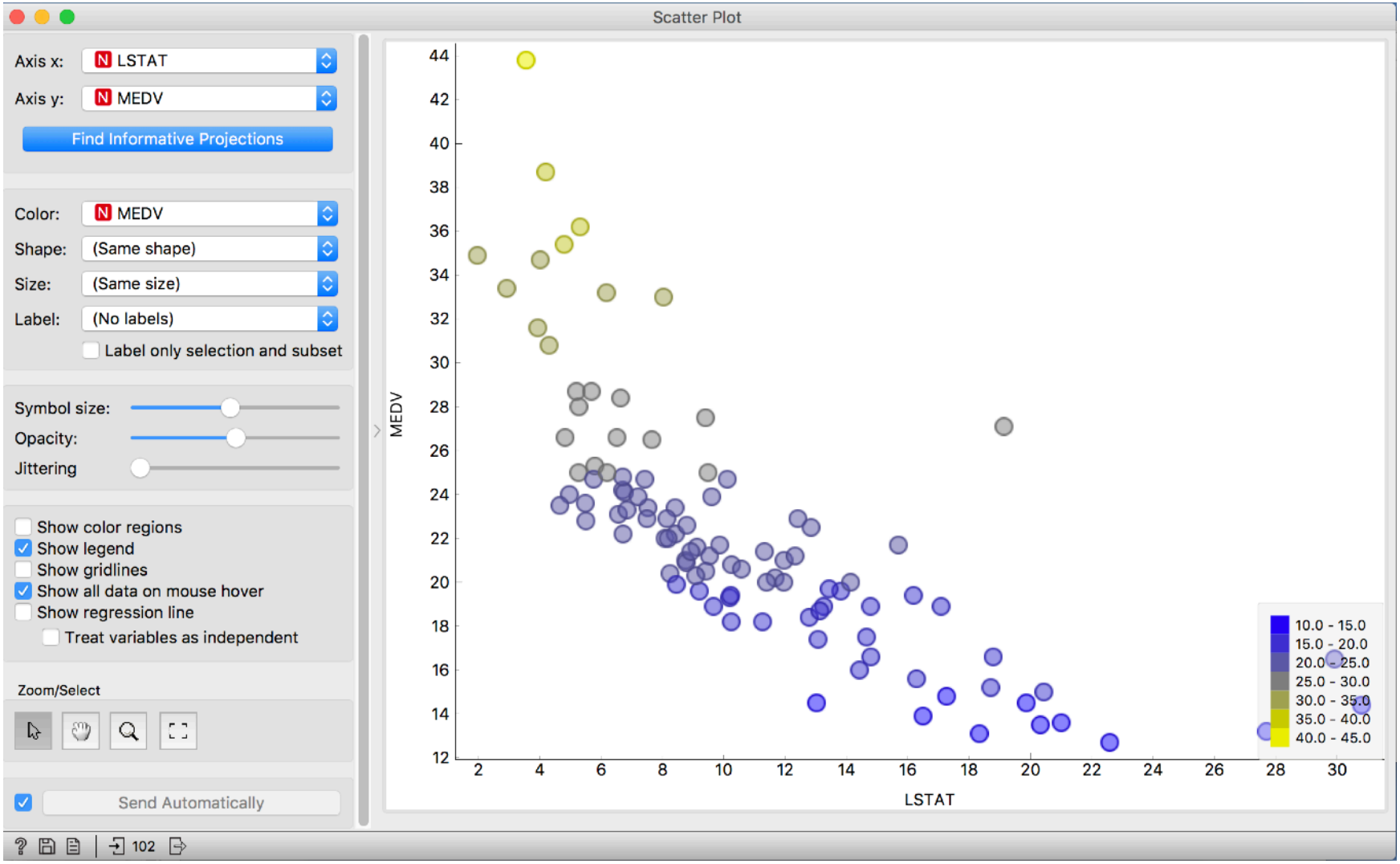
The same logic applies to Regression models.

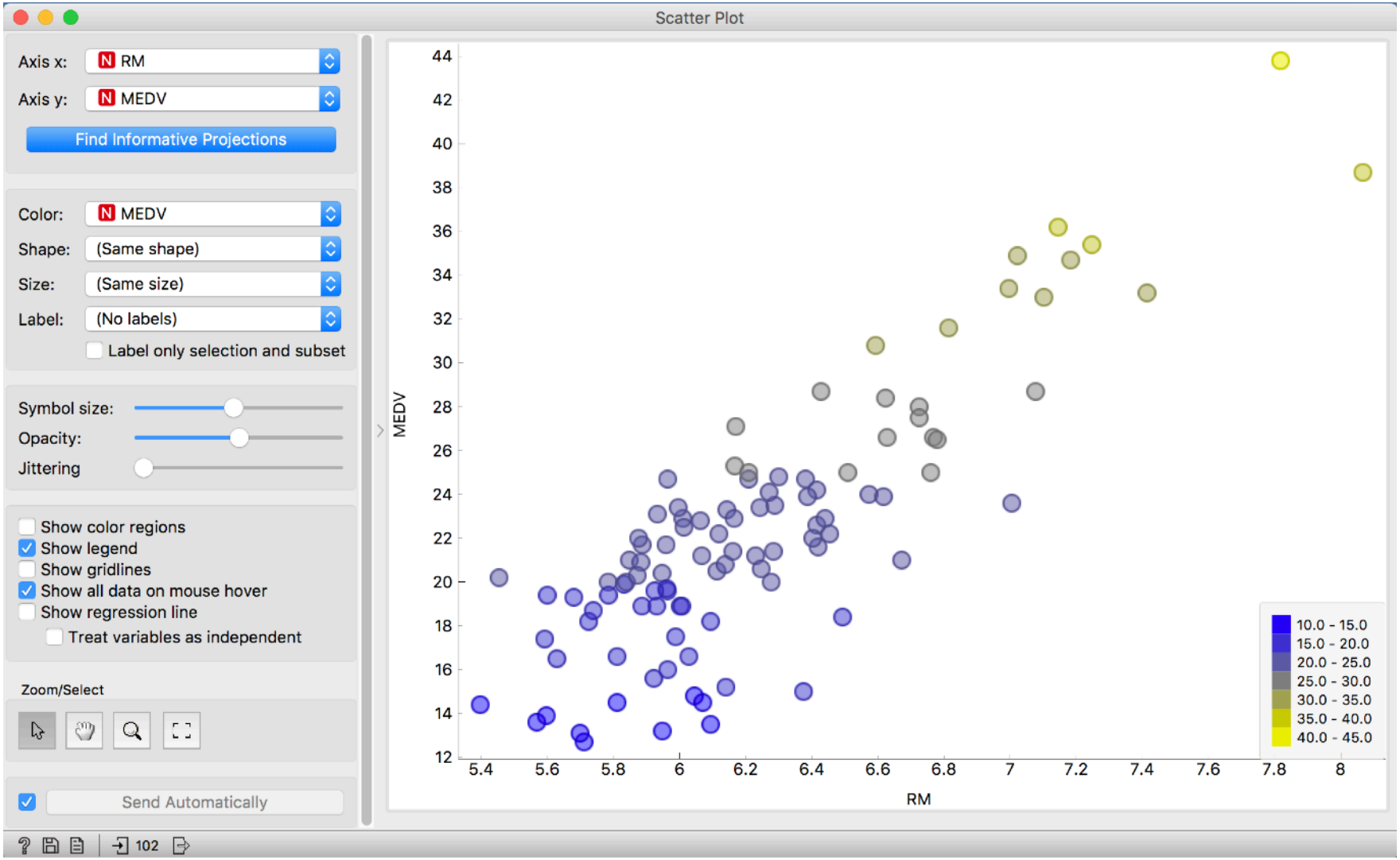

linear_regression_2.ows
Constant
CN2 Rule Induction
Calibrated Learner
kNN
Tree
Random Forest
SVM
Linear Regressi...
Logistic Regressi...
Naive Bayes
AdaBoost
Neural Network
Stochastic Gradient...
Stacking
Save Model
Load Model
Evaluate
Test and Score
Cross-validation accuracy estimation.
more...
File
Select Columns
Data Sampler
Data Table
Linear Regression
Scatter Plot
Data Table (1)
Rank
Test and Score
Predictions
Data
Data Sample
Remaining Data
Learner
Model
Predictors

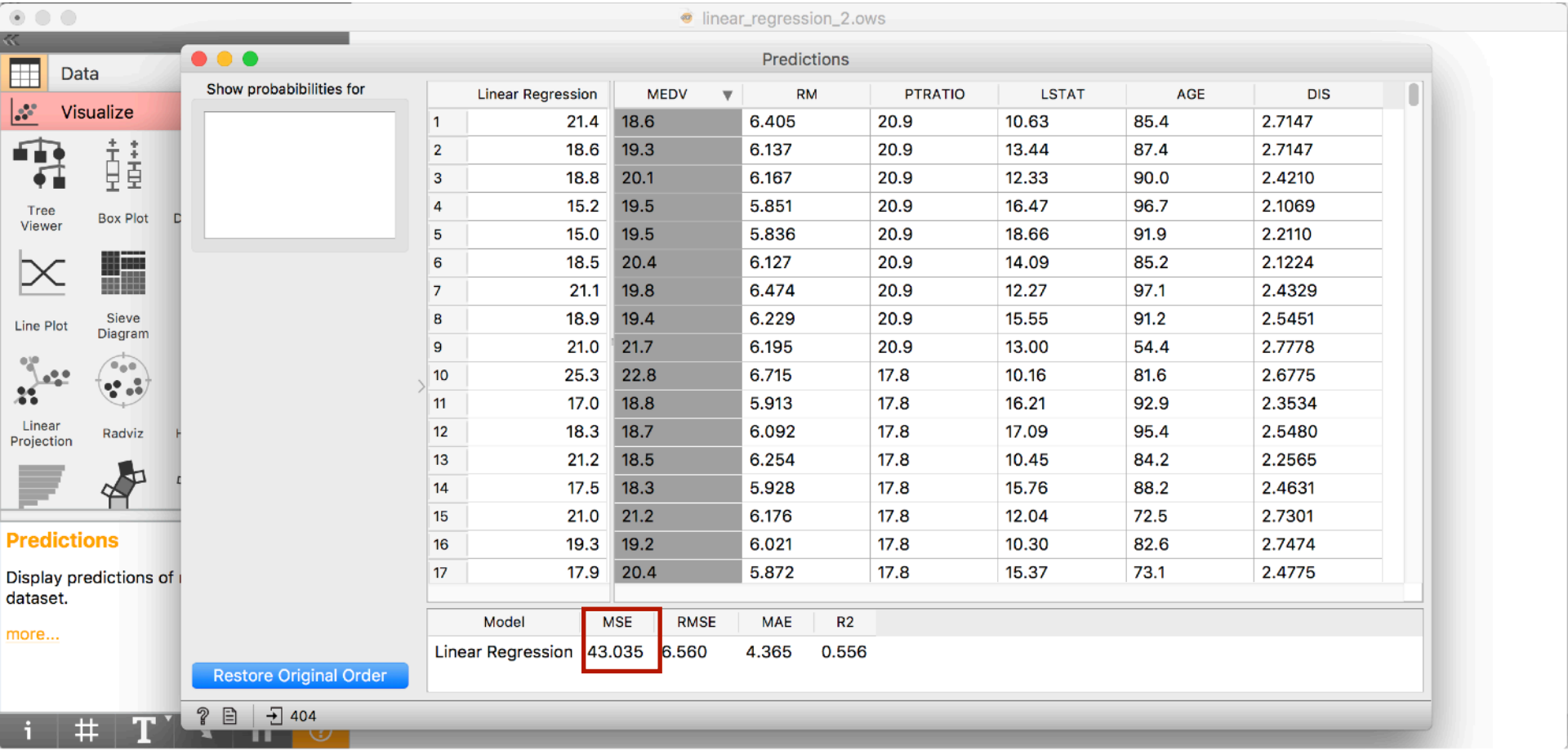
Same pipeline concept.


linear_regression_2.ows
Data
Visualize
Tree Viewer
Box Plot
Distribut...
Scatter Plot
Line Plot
Sieve Diagram
Mosaic Display
FreeViz
Linear Projection
Radviz
Heat Map
Venn Diagram
Scatter Plot
Interactive scatter plot visualization with intelligent data visualization enhancements.
more...
File
Select Columns
Data Sampler
Data Table
Linear Regression
Scatter Plot
Data Table (1)
Rank
Test and Score
Predictions
Data
Data Sample
Remaining Data
Learner
Model → Predictors
Remaining Data → Data
Remaining Data → Test Data
Data Sample → Data


File
File: housing.tab
...
Reload
URL:
Info
Housing dataset
Data collected by the U.S Census Service concerning housing in Boston.
506 instance(s)
13 feature(s) (no missing values)
Regression; numerical class (no missing values)
0 meta attribute(s)
Columns (Double click to edit)
Name
Type
Role
Values
1 CRIM numeric feature
2 ZN numeric feature
3 INDUS numeric feature
4 CHAS numeric feature
5 NOX numeric feature
6 RM numeric feature
Browse documentation datasets
Reset
Apply
Data
Linear Regression
Data
Scatter Plot
Remaining Data → Data
Test Data
Learner
Model → Predictors
Predictions
Test and Score

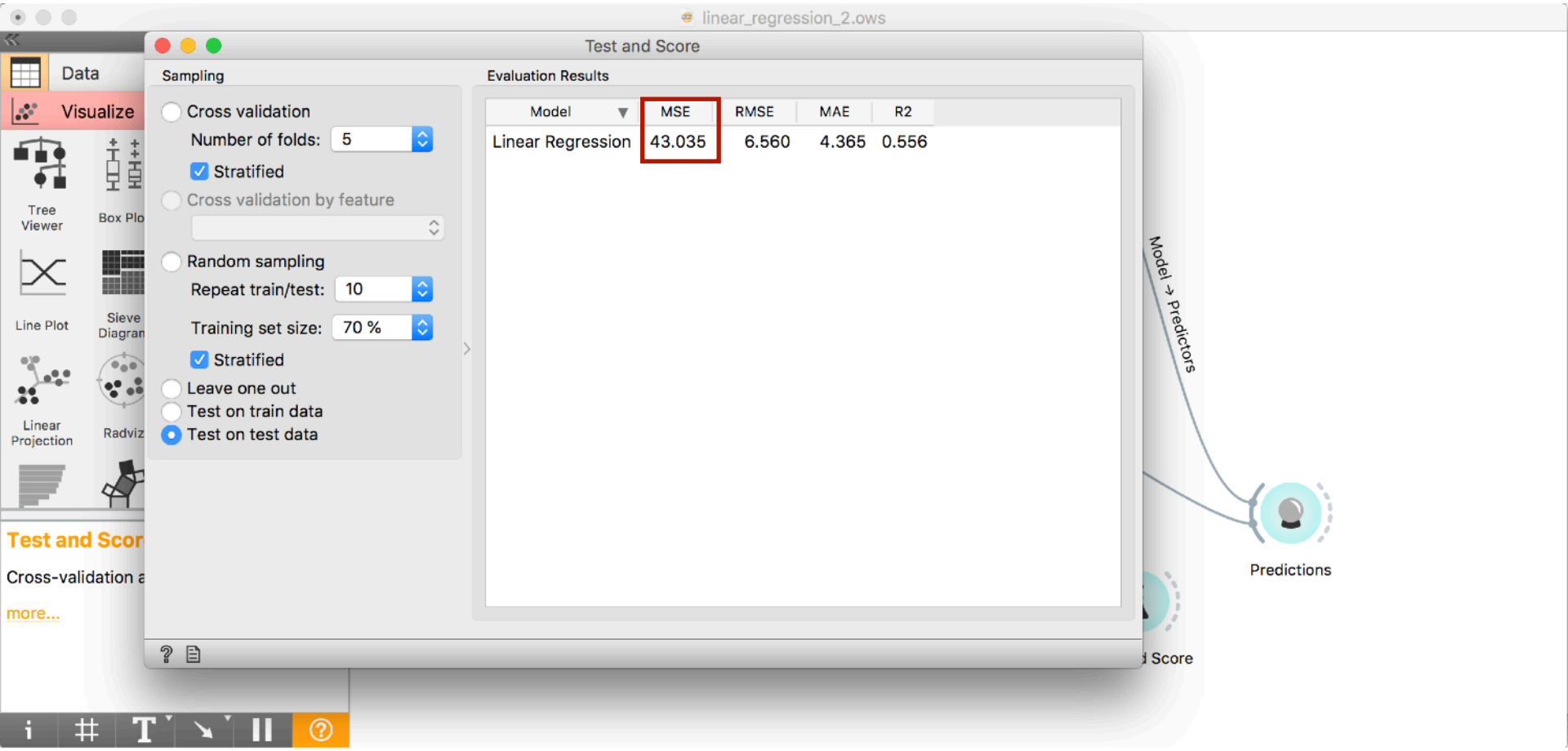

Scatter Plot
Axis x:
LSTAT
Axis y:
MEDV
Find Informative Projections
Color:
MEDV
Shape:
(Same shape)
Size:
(Same size)
Label:
(No labels)
Label only selection and subset
Symbol size:
Opacity:
Jittering
Show color regions
Show legend
Show gridlines
Show all data on mouse hover
Show regression line
Treat variables as independent
Zoom/Select
Send Automatically
102
MEDV
LSTAT
10.0 - 15.0
15.0 - 20.0
20.0 - 25.0
25.0 - 30.0
30.0 - 35.0
35.0 - 40.0
40.0 - 45.0


Scatter Plot
Axis x: RM
Axis y: MEDV
Find Informative Projections
Color: MEDV
Shape: (Same shape)
Size: (Same size)
Label: (No labels)
Label only selection and subset
Symbol size:
Opacity:
Jittering
Show color regions
Show legend
Show gridlines
Show all data on mouse hover
Show regression line
Treat variables as independent
Zoom/Select
Send Automatically
102
MEDV
RM
10.0 - 15.0
15.0 - 20.0
20.0 - 25.0
25.0 - 30.0
30.0 - 35.0
35.0 - 40.0
40.0 - 45.0



	Linear Regression	MEDV	RM	PTRATIO	LSTAT	AGE	DIS
1	21.4	18.6	6.405	20.9	10.63	85.4	2.7147
2	18.6	19.3	6.137	20.9	13.44	87.4	2.7147
3	18.8	20.1	6.167	20.9	12.33	90.0	2.4210
4	15.2	19.5	5.851	20.9	16.47	96.7	2.1069
5	15.0	19.5	5.836	20.9	18.66	91.9	2.2110
6	18.5	20.4	6.127	20.9	14.09	85.2	2.1224
7	21.1	19.8	6.474	20.9	12.27	97.1	2.4329
8	18.9	19.4	6.229	20.9	15.55	91.2	2.5451
9	21.0	21.7	6.195	20.9	13.00	54.4	2.7778
10	25.3	22.8	6.715	17.8	10.16	81.6	2.6775
11	17.0	18.8	5.913	17.8	16.21	92.9	2.3534
12	18.3	18.7	6.092	17.8	17.09	95.4	2.5480
13	21.2	18.5	6.254	17.8	10.45	84.2	2.2565
14	17.5	18.3	5.928	17.8	15.76	88.2	2.4631
15	21.0	21.2	6.176	17.8	12.04	72.5	2.7301
16	19.3	19.2	6.021	17.8	10.30	82.6	2.7474
17	17.9	20.4	5.872	17.8	15.37	73.1	2.4775

Model	MSE	RMSE	MAE	R2
Linear Regression	43.035	6.560	4.365	0.556


linear_regression_2.ows
Test and Score
Sampling
Cross validation
Number of folds: 5
Stratified
Cross validation by feature
Random sampling
Repeat train/test: 10
Training set size: 70 %
Stratified
Leave one out
Test on train data
Test on test data
Evaluation Results
Model
MSE
RMSE
MAE
R2
Linear Regression
43.035
6.560
4.365
0.556
Data
Visualize
Tree Viewer
Box Plo
Line Plot
Sieve Diagran
Linear Projection
Radviz
Test and Scor
Cross-validation a
more...
Model → Predictors
Predictions
d Score

The same coding logic applies to building Regression models in Sckit-learn

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Code

Building and Visualizing a Linear Regression Model

In [96]:
```
# Import Pandas, Matpolib, Numpy, and Scikit-learn modules
from matplotlib import pyplot as plt
import pandas as pd
import numpy as np
from sklearn import linear_model
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix
from sklearn.metrics import accuracy_score
```

In [97]:
```
# Import dataset
column_names = ['CRIM', 'ZN', 'INDUS', 'CHAS', 'NOX', 'RM', 'AGE', 'DIS', 'RAD', 'TAX', 'PTRATIO', 'B', 'LSTAT', 'N
data = pd.read_csv('http://localhost/datasets/housing.csv', header=None, delimiter=r"\s+", names=column_names)
# Check the first 5 entries
print(data.head(5))
```

```
      CRIM    ZN  INDUS  CHAS    NOX     RM   AGE     DIS  RAD    TAX  \
0  0.00632  18.0   2.31     0  0.538  6.575  65.2  4.0900    1  296.0
1  0.02731   0.0   7.07     0  0.469  6.421  78.9  4.9671    2  242.0
2  0.02729   0.0   7.07     0  0.469  7.185  61.1  4.9671    2  242.0
3  0.03237   0.0   2.18     0  0.458  6.998  45.8  6.0622    3  222.0
4  0.06905   0.0   2.18     0  0.458  7.147  54.2  6.0622    3  222.0

   PTRATIO       B  LSTAT  MEDV
0     15.3  396.90   4.98  24.0
1     17.8  396.90   9.14  21.6
2     17.8  392.83   4.03  34.7
```

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Code

In [98]:
```
# Display number of rows and columns
print(np.shape(data))
```

```
(506, 14)
```

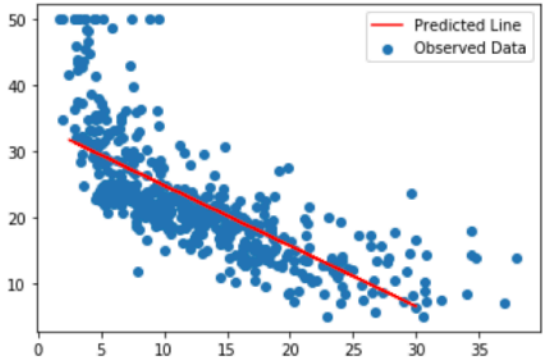
In [99]:
```
# Descriptive statistics of features
print(data.describe())
```

```
              CRIM          ZN       INDUS        CHAS         NOX          RM  \
count  506.000000  506.000000  506.000000  506.000000  506.000000  506.000000
mean     3.613524   11.363636   11.136779    0.069170    0.554695    6.284634
std      8.601545   23.322453    6.860353    0.253994    0.115878    0.702617
min      0.006320    0.000000    0.460000    0.000000    0.385000    3.561000
25%      0.082045    0.000000    5.190000    0.000000    0.449000    5.885500
50%      0.256510    0.000000    9.690000    0.000000    0.538000    6.208500
75%      3.677082   12.500000   18.100000    0.000000    0.624000    6.623500
max     88.976200  100.000000   27.740000    1.000000    0.871000    8.780000

              AGE         DIS         RAD         TAX     PTRATIO           B  \
count  506.000000  506.000000  506.000000  506.000000  506.000000  506.000000
mean    68.574901    3.795043    9.549407  408.237154   18.455534  356.674032
std     28.148861    2.105710    8.707259  168.537116    2.164946   91.294864
min      2.900000    1.129600    1.000000  187.000000   12.600000    0.320000
25%     45.025000    2.100175    4.000000  279.000000   17.400000  375.377500
50%     77.500000    3.207450    5.000000  330.000000   19.050000  391.440000
75%     94.075000    5.188425   24.000000  666.000000   20.200000  396.225000
max    100.000000   12.126500   24.000000  711.000000   22.000000  396.900000

            LSTAT        MEDV
count  506.000000  506.000000
mean    12.653063   22.532806
std      7.141062    9.197104
min      1.730000    5.000000
```

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Code

In [104]:
```
# Display scatter plot to fit the line
a=reg_model.predict(x_test[:None])
plt.scatter(x,y)
plt.plot(x_test,a,'r')
plt.legend(['Predicted Line','Observed Data'])
plt.show()
```

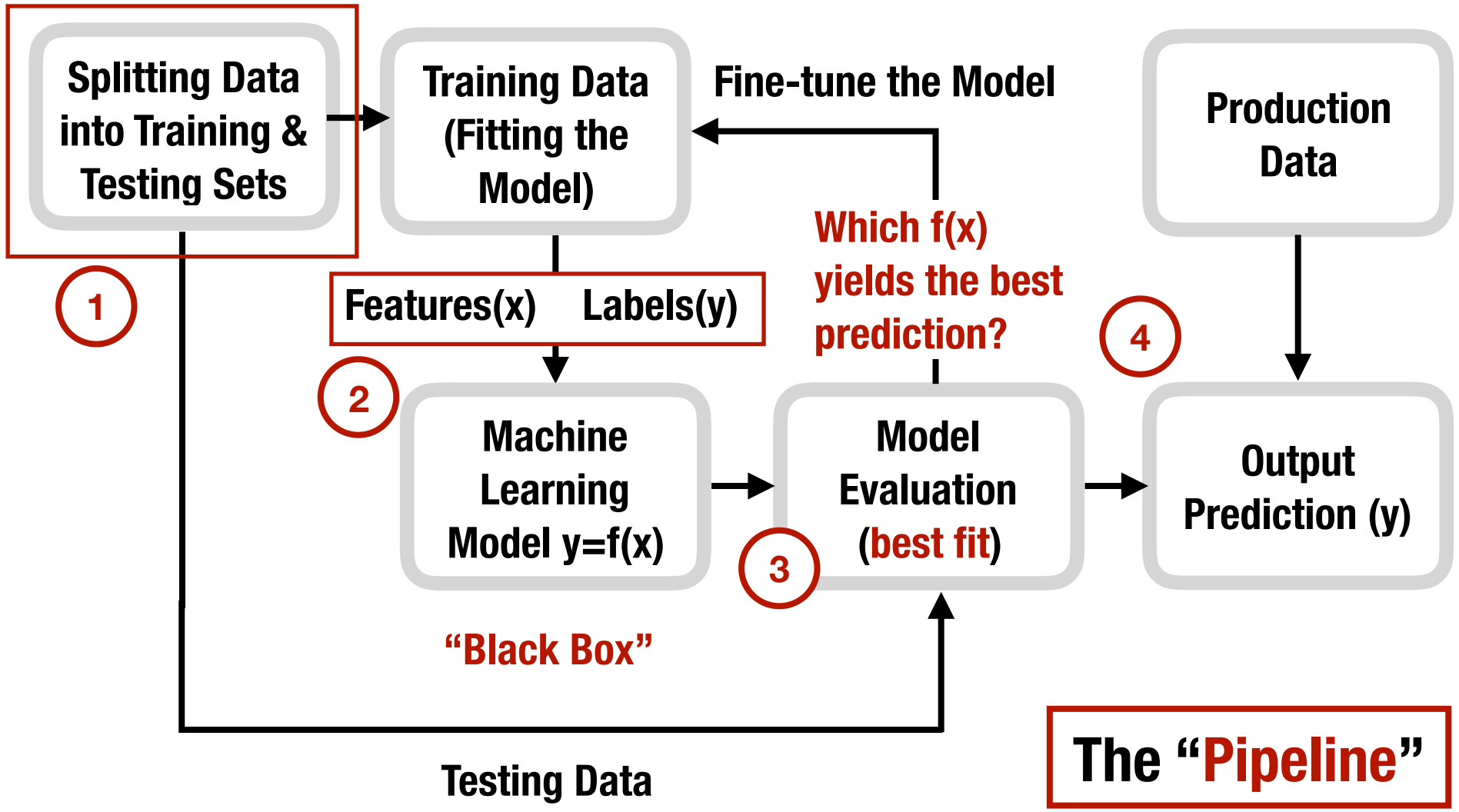


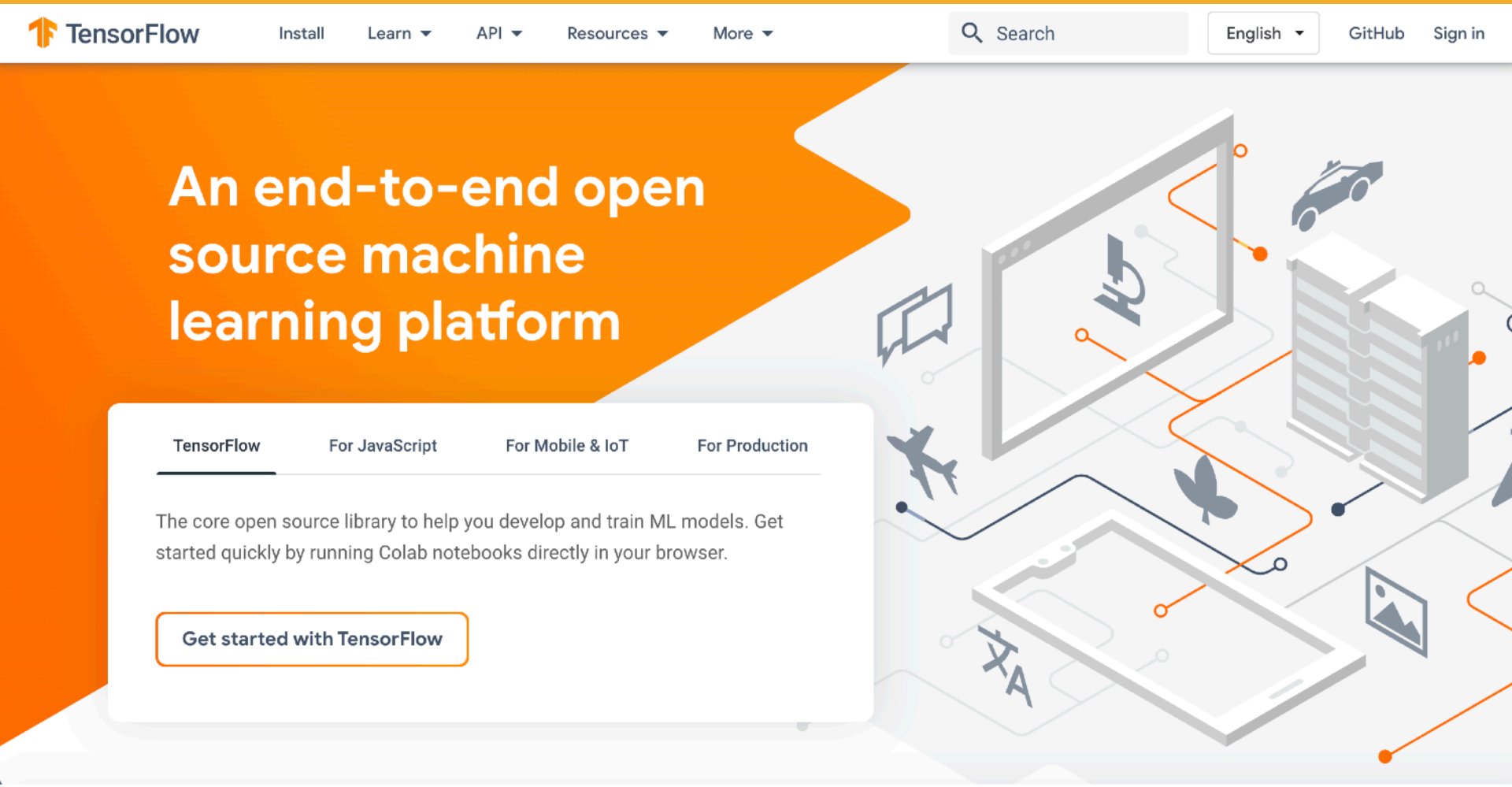
In [105]:
```
# Mean Square Error (MSE)
np.mean((a-y_test)**2)
```

Out[105]: 40.13019759337232

In [ ]:
```

```

Mastering the pipeline workflow is the key to get started.


Splitting Data into Training & Testing Sets
Training Data (Fitting the Model)
Fine-tune the Model
Production Data
1
Features(x) Labels(y)
Which f(x) yields the best prediction?
4
2
Machine Learning Model y=f(x)
Model Evaluation (best fit)
Output Prediction (y)
3
"Black Box"
Testing Data
The "Pipeline"



https://www.tensorflow.org/

TensorFlow Install Learn API Resources More Search English GitHub Sign in

TensorFlow Core

Overview Tutorials Guide TF 1

TensorFlow is an end-to-end open source platform for machine learning

TensorFlow makes it easy for beginners and experts to create machine learning models. See the sections below to get started.

See tutorials

See the guide

Tutorials show you how to use TensorFlow with complete, end-to-end examples.

Guides explain the concepts and components of TensorFlow.

https://www.tensorflow.org/overview/

beginner.ipynb

File Edit View Insert Runtime Tools Help

Share

ble of contents

opyright 2019 The TensorFlow Authors.

ensorFlow 2 quickstart for beginners

Section

+ Code + Text Copy to Drive

RAM Disk

Editing

Copyright 2019 The TensorFlow Authors.

1 cell hidden

TensorFlow 2 quickstart for beginners

View on TensorFlow.org Run in Google Colab View source on GitHub Download notebook

This short introduction uses Keras to:

1. Build a neural network that classifies images.
2. Train this neural network.
3. And, finally, evaluate the accuracy of the model.

This is a Google Colaboratory notebook file. Python programs are run directly in the browser—a great way to learn and use TensorFlow. To follow this tutorial, run the notebook in Google Colab by clicking the button at the top of this page.

1. In Colab, connect to a Python runtime: At the top-right of the menu bar, select *CONNECT*.
2. Run all the notebook code cells: Select *Runtime* > *Run all*.

TensorFlow 2 quickstart for beginners

Run in Google Colab | View source on GitHub | Download notebook

This short introduction uses Keras to:

1. Build a neural network that classifies images.
2. Train this neural network.
3. And, finally, evaluate the accuracy of the model.

This is a Google Colaboratory notebook file. Python programs are run directly in the browser—a great way to learn and use TensorFlow. To follow this tutorial, run the notebook in Google Colab by clicking the button at the top of this page.

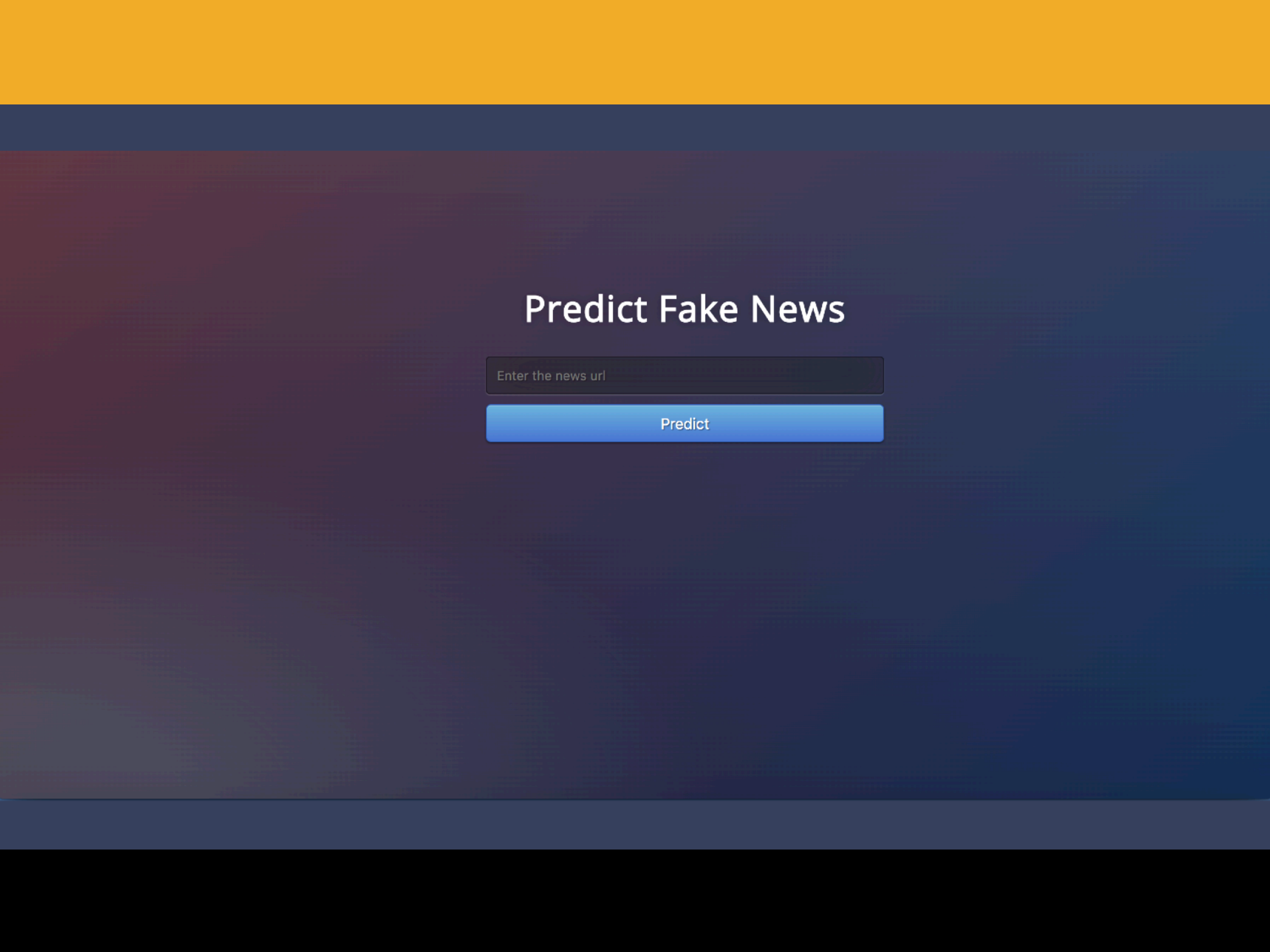
1. In Colab, connect to a Python runtime: At the top-right of the menu bar, select *CONNECT*.
2. Run all the notebook code cells: Select *Runtime* > *Run all*.

Download and install TensorFlow 2. Import TensorFlow into your program:

Note: Upgrade `pip` to install the TensorFlow 2 package. See the install guide for details.

https://www.youtube.com/watch?v=OXNC_sefxi4

Deploying ML+ Applications in Flask


Predict Fake News
Enter the news url
Predict

Training the Model

Building a Fake News Classifier & Deploying it Using Flask

Article Github link Demo Site

1. 1st Fake News
2. 1st Fake News
3. 1st Real News

In [108]:
```
# Importing the libraries
import pandas as pd
import numpy as np
from sklearn.metrics import classification_report, confusion_matrix
from sklearn.pipeline import Pipeline
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import MultinomialNB
from sklearn.feature_extraction.text import TfidfVectorizer
import pickle

#Importing the cleaned file containing the text and label
news = pd.read_csv('news.csv')
X = news['text']
y = news['label']
```

In [109]:
```
#Splitting the data into train
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)

#Creating a pipeline that first creates bag of words(after applying stopwords) & then applies Multinomial Naive Bay
pipeline = Pipeline([('tfidf', TfidfVectorizer(stop_words='english')),
                     ('nbmodel', MultinomialNB())])
```

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Code

In [109]:

```
#Splitting the data into train
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)

#Creating a pipeline that first creates bag of words(after applying stopwords) & then applies Multinomial Naive Bay
pipeline = Pipeline([('tfidf', TfidfVectorizer(stop_words='english')),
                     ('nbmodel', MultinomialNB())])

#Training our data
pipeline.fit(X_train, y_train)

#Predicting the label for the test data
pred = pipeline.predict(X_test)
```

In [110]:

```
plitting the data into train
train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)

reating a pipeline that first creates bag of words(after applying stopwords) & then applies Multinomial Naive Bayes
peline = Pipeline([('tfidf', TfidfVectorizer(stop_words='english')),
                   ('nbmodel', MultinomialNB())])

raining our data
peline.fit(X_train, y_train)

redicting the label for the test data
ed = pipeline.predict(X_test)
```

In [111]:

```
#Checking the performance of our model
print(classification_report(y_test, pred))
print(confusion_matrix(y_test, pred))

#Serialising the file
```

In [110]:

```
#Splitting the data into train
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)

#Creating a pipeline that first creates bag of words(after applying stopwords) & then applies Multinomial Naive Bay
pipeline = Pipeline([('tfidf', TfidfVectorizer(stop_words='english')),
                     ('nbmodel', MultinomialNB())])

#Training our data
pipeline.fit(X_train, y_train)

#Predicting the label for the test data
pred = pipeline.predict(X_test)
```

In [111]:

```
#Checking the performance of our model
print(classification_report(y_test, pred))
print(confusion_matrix(y_test, pred))

#Serialising the file
with open('model.pickle', 'wb') as handle:
    pickle.dump(pipeline, handle, protocol=pickle.HIGHEST_PROTOCOL)
```

```
              precision    recall  f1-score   support

        FAKE       0.98      0.70      0.82       655
        REAL       0.75      0.98      0.85       612

    accuracy                           0.84      1267
   macro avg       0.87      0.84      0.83      1267
weighted avg       0.87      0.84      0.83      1267

[[458 197]
 [ 10 602]]
```

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Code

In [1]:

```
#!pip install newspaper3k
# import nltk
# nltk.download('punkt')
```

In [*]:

```
#Importing the Libraries
import numpy as np
from flask import Flask, request,render_template
from flask_cors import CORS
import os
from sklearn.externals import joblib
import pickle
import newspaper
from newspaper import Article
import urllib

#Loading Flask and assigning the model variable
app = Flask(__name__)
CORS(app)

with open('model.pickle', 'rb') as handle:
    model = pickle.load(handle)

@app.route('/')
def main():
    return render_template('main.html')

#Receiving the input url from the user and using Web Scrapping to extract the news content
@app.route('/predict',methods=['GET','POST'])
def predict():
    url =request.get_data(as_text=True)[5:]
    url = urllib.parse.unquote(url)
    article = Article(str(url))
    article.download()
```

Running the Flask App

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Code

In [*]:

```
#Importing the Libraries
import numpy as np
from flask import Flask, request,render_template
from flask_cors import CORS
import os
from sklearn.externals import joblib
import pickle
import newspaper
from newspaper import Article
import urllib

#Loading Flask and assigning the model variable
app = Flask(__name__)
CORS(app)

with open('model.pickle', 'rb') as handle:
    model = pickle.load(handle)

@app.route('/')
def main():
    return render_template('main.html')

#Receiving the input url from the user and using Web Scrapping to extract the news content
@app.route('/predict',methods=['GET','POST'])
def predict():
    url =request.get_data(as_text=True)[5:]
    url = urllib.parse.unquote(url)
    article = Article(str(url))
    article.download()
    article.parse()
    article.nlp()
    news = article.summary
    #Passing the news article to the model and returing whether it is Fake or Real
```

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Code

```
    model = pickle.load(handle)

@app.route('/')
def main():
    return render_template('main.html')

#Receiving the input url from the user and using Web Scrapping to extract the news content
@app.route('/predict',methods=['GET','POST'])
def predict():
    url =request.get_data(as_text=True)[5:]
    url = urllib.parse.unquote(url)
    article = Article(str(url))
    article.download()
    article.parse()
    article.nlp()
    news = article.summary
    #Passing the news article to the model and returing whether it is Fake or Real
    pred = model.predict([news])
    return render_template('main.html', prediction_text='The news is "{}"'.format(pred[0]))

if __name__ == '__main__':
    from werkzeug.serving import run_simple
    run_simple('localhost', 9019, app)
# if __name__ == '__main__':
   # app.run(debug = True)
```

```
/Users/yssuen/opt/anaconda3/lib/python3.7/site-packages/sklearn/externals/joblib/__init__.py:15: DeprecationWarning:
sklearn.externals.joblib is deprecated in 0.21 and will be removed in 0.23. Please import this functionality directly
from joblib, which can be installed with: pip install joblib. If this warning is raised when loading pickled models,
you may need to re-serialize those models with scikit-learn 0.21+.
  warnings.warn(msg, category=DeprecationWarning)
 * Running on http://localhost:9019/ (Press CTRL+C to quit)
127.0.0.1 - - [20/Apr/2020 18:08:33] "GET / HTTP/1.1" 200 -
```

THANK YOU FOR YOUR TIME!

Popular Algorithms

Unsupervised Learning	Supervised Learning	Reinforcement Learning
Principal Component Analysis	K-Nearest Neighbour	Q-Learning
Hierarchical Clustering	Logical Regression	Markov Decision Processes
K-Means Clustering	Linear Regression	
Association Rules	Decision Tree	
Collaborative Filtering	Random Forest	
	Support Vector Machine	
	Naive Baynes	
	Artificial Neural Network	