

Знакомство с системой Orange

Определим ключевые термины, которые помогут в дальнейшем понимании:

Виджет - основная точка обработки любых действий с данными. Виджет выполняет действия в зависимости от того, что вы выберете в селекторе виджетов в левой части экрана.

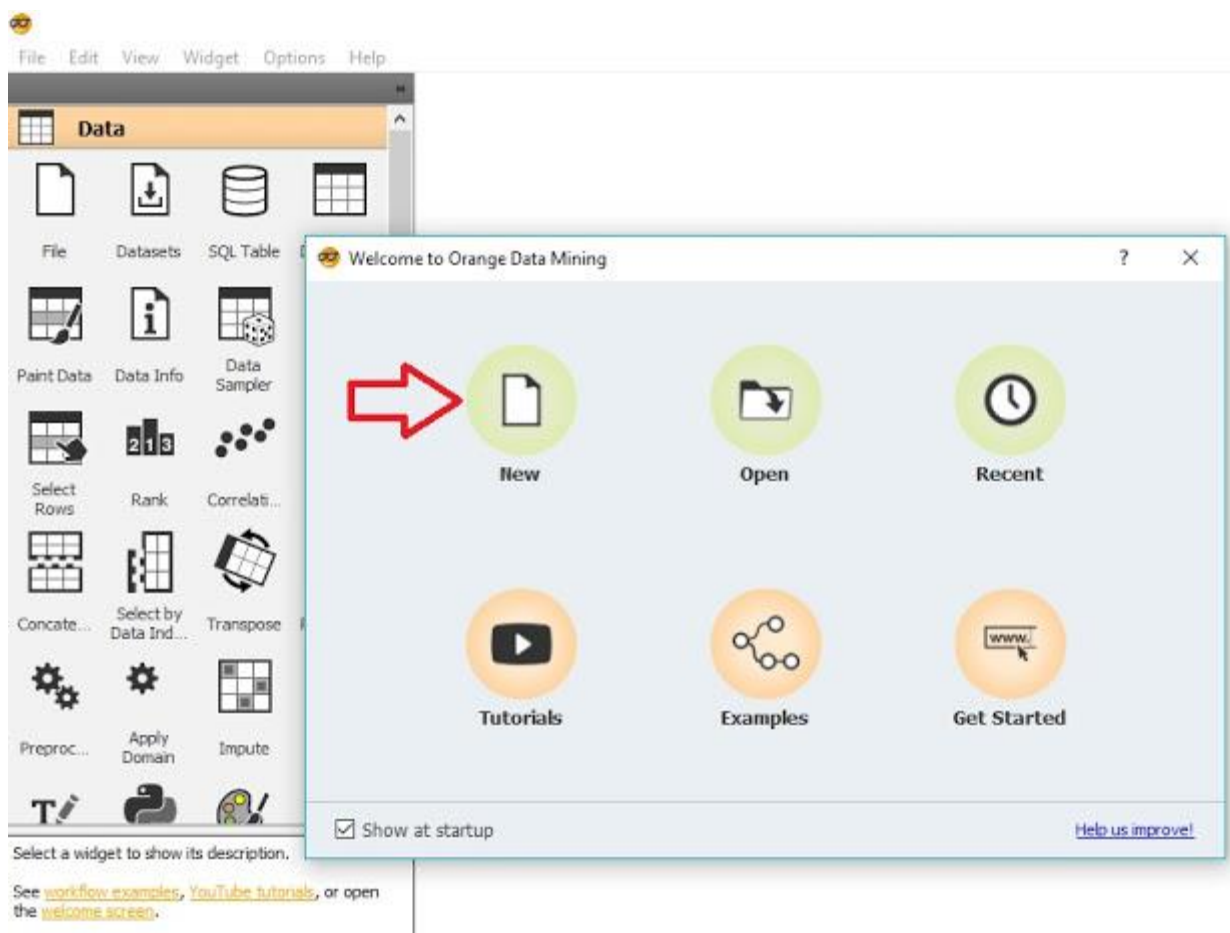
Рабочий процесс — это последовательность шагов или действий, которые вы выполняете на платформе для решения задачи. Рабочий процесс предполагает передачу данных от виджета к виджету.

Перейдите к разделу «Примеры рабочих процессов» на начальном экране, чтобы изучить варианты дополнительных рабочих процессов и используемых моделей.

Создание первого рабочего процесса

Нажмите «New» и создайте первый рабочий процесс.

Это первый шаг на пути к решению любой задачи. Обдумайте, какие шаги необходимо предпринять для достижения конечной цели — алгоритм построения процесса.



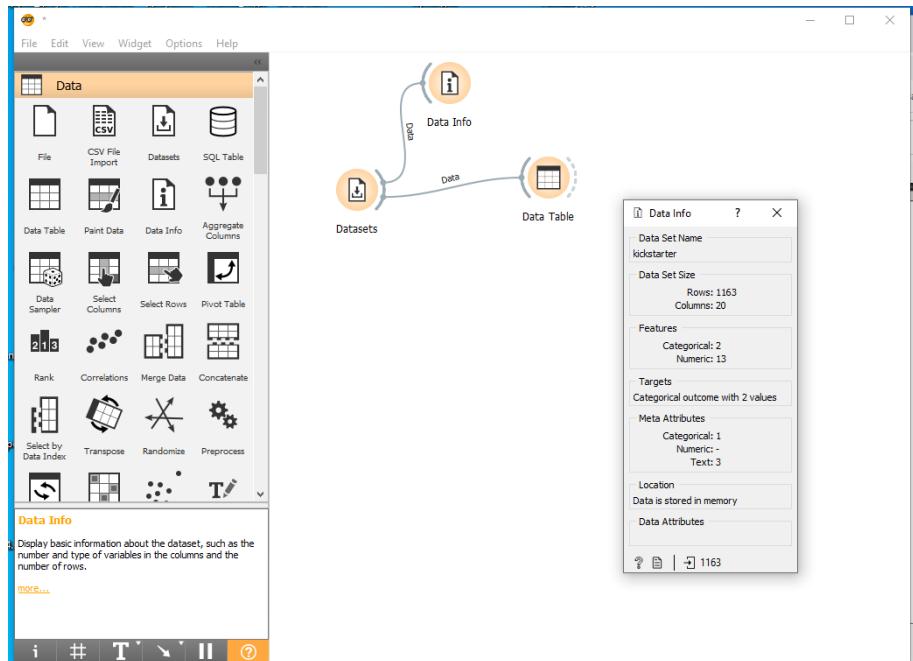
Импорт данных в Orange

Шаг 1: Нажмите на вкладку «Data» в меню выбора виджетов и перетащите виджет «File» в пустой рабочий процесс.

Шаг 2: Загрузка данных. Данные можно загрузить из файла (виджет «File») или воспользоваться одним из встроенных датасетов (выбрать виджет «Datasets»).

Воспользуемся готовым датасетом.. Выберем «Datasets» и загрузить данные оттуда, например, воспользуемся любым набором

Шаг 3: Как только вы сможете увидеть структуру набора данных с помощью виджета, вернитесь, закрыв это меню.

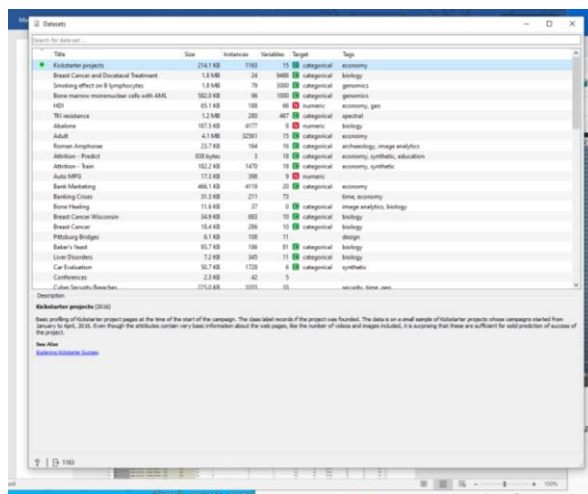


Шаги 3 и 4

Пример. Работа с виджетами. Мы загрузили виджеты датасет, таблица и информационный

*Для работы можно выбрать уже имеющийся в Orange датасет. Для этого нужно кликнуть по виджету. В данном случае выбран **Kickstarter projects***

Когда вы выбираете набор данных, в нижнем поле можете прочитать описание ситуации или получить на него ссылку.



Меню виджета Datasets

Шаг 4: Поскольку нам нужна таблица данных, чтобы лучше визуализировать наши результаты, мы используем виджет «Data Table». Перетаскиваем его в рабочее поле и соединяем с виджетом с данными. На появившейся соединительной линии кликаем и в появившейся таблице устанавливаем связь, если ее не было, нажимаем «Ок» Для применения виджетов с инструментами и методами данные должны быть именно в таблице.

Data Table

Info

1163 instances (no missing data)

15 features

Target with 2 values

4 meta attributes

Variables

☐ Show variable labels (if present)

☐ Visualize numeric values

☐ Color by instance classes

Selection

☐ Select full rows

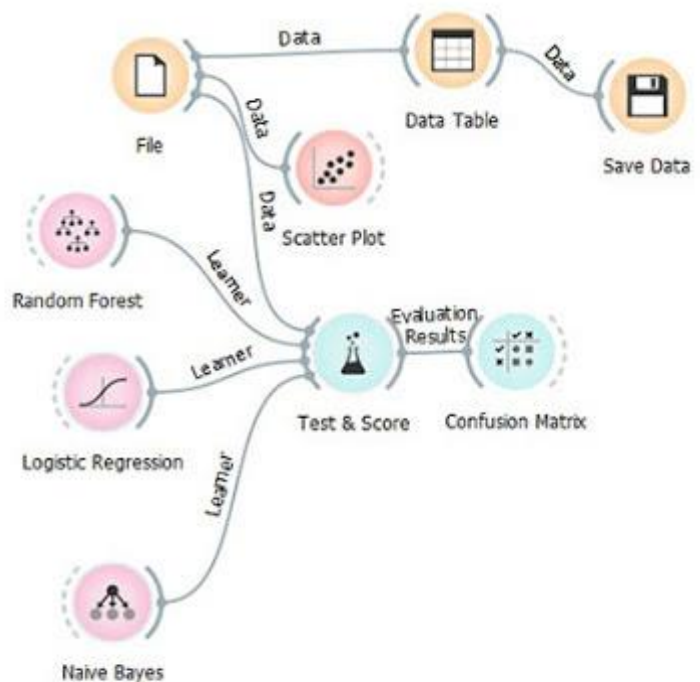
Restore Original Order

☒ Send Automatically

Funded	URL	Title	Year	Month	Type	Has FB	Backed Projects	Previous Projects	Creator Desc Len	Title Len	Goal	Duration	Pledge Levels	Min Pledge Tiers
no	https://www.k...	Pixelart: Cha...	2016	Apr	Art	1	11	2	125	57	2829.59	53	7	1.14
no	https://www.k...	Smart shop loc...	2016	Apr	Art	1	0	0	111	27	28295.87	51	3	1.14
no	https://www.k...	Minimal Hour...	2016	Apr	Art	0	4	0	284	52	766.25	30	8	1.51
no	https://www.k...	Neofu Altering...	2016	Mar	Art	0	0	0	179	41	1439.10	24	5	7.00
no	https://www.k...	Nintendo NES &...	2016	Mar	Art	0	0	0	51	41	1000.00	30	2	5.00
no	https://www.k...	Day and Night...	2016	Mar	Art	0	0	0	162	30	800.00	28	1	10.00
no	https://www.k...	Fund an Art Sh...	2016	Feb	Art	1	2	0	110	55	2000.00	50	9	5.00
no	https://www.k...	Trump that T...	2016	Mar	Art	1	1	0	12	58	1000.00	30	3	1.00
yes	https://www.k...	Once Upon a Pl...	2016	Apr	Art	1	0	0	285	19	1079.32	30	16	7.00
yes	https://www.k...	Under the Hoo...	2016	Apr	Art	1	12	8	55	49	10000.00	30	19	5.00
yes	https://www.k...	KONORO	2016	Mar	Art	0	0	0	637	6	792.28	45	5	6.00
yes	https://www.k...	Draw Cool Sh'e...	2016	Apr	Art	1	0	0	317	52	996.12	21	4	0.77
yes	https://www.k...	Elleforthecomp...	2016	Mar	Art	0	0	0	204	45	2000.00	30	14	10.00
yes	https://www.k...	PT Apparel	2016	Mar	Art	1	0	0	337	10	600.00	45	6	5.00
yes	https://www.k...	Epocha - Hand...	2016	Apr	Art	1	143	18	221	55	1000.00	25	3	20.00
yes	https://www.k...	The Little ABC...	2016	Apr	Art	1	13	9	176	29	863.46	20	8	4.00
yes	https://www.k...	Bull & Fur	2016	Mar	Art	0	0	0	161	14	14000.00	30	7	10.00
yes	https://www.k...	Pens & Pedals	2016	Apr	Art	1	3	0	514	17	3000.00	21	5	10.00
yes	https://www.k...	BCU Illustration...	2016	Feb	Art	1	0	0	26	21	1439.10	58	7	7.00
yes	https://www.k...	78 Tarot Carnival	2016	Mar	Art	1	104	4	498	17	20000.00	29	14	1.00
yes	https://www.k...	Deep Sea Garde...	2016	Mar	Art	1	0	2	272	30	500.00	30	8	1.00
yes	https://www.k...	La Petite Mech...	2016	Feb	Art	1	0	0	518	17	6791.01	60	9	6.00
yes	https://www.k...	Prohibition Fir...	2016	Mar	Art	1	0	0	188	64	143.91	30	9	4.00
yes	https://www.k...	HOLLY SIMPLE...	2016	Mar	Art	1	2	0	511	57	5244.00	36	12	12.00
yes	https://www.k...	Ulysses et Plucl...	2016	Mar	Art	1	4	0	380	39	4037.37	45	14	4.00
yes	https://www.k...	Special Edition...	2016	Mar	Art	1	0	2	498	58	100.00	30	6	1.00
yes	https://www.k...	Magnificent M...	2016	Mar	Art	1	2	0	329	47	4000.00	29	7	2.00
yes	https://www.k...	The Future is N...	2016	Mar	Art	0	10	0	210	30	20373.03	34	9	1.10
yes	https://www.k...	Day of the Dea...	2016	Mar	Art	1	21	24	120	47	250.00	14	9	2.00
yes	https://www.k...	Yin Yang Sugar...	2016	Mar	Art	0	42	9	49	41	250.00	15	7	2.00
yes	https://www.k...	Best Season Ever	2016	Mar	Art	0	14	0	255	16	14000.00	30	42	1.00
yes	https://www.k...	New Frontiers...	2016	Mar	Art	1	18	2	577	45	2298.75	25	16	4.00
yes	https://www.k...	Drakon Infinitu...	2016	Mar	Art	1	6	0	534	30	600.00	30	8	1.00
yes	https://www.k...	Pookie Cat Tar...	2016	Mar	Art	0	1	2	457	60	1131.84	30	13	1.10
yes	https://www.k...	Little Batman V...	2016	Mar	Art	1	29	6	96	33	400.00	35	12	2.00
yes	https://www.k...	Liam Irons Artw...	2016	Mar	Art	1	0	0	499	18	503.68	30	3	21.00
yes	https://www.k...	The Art of Beer...	2016	Feb	Art	0	13	0	505	55	1128.00	32	8	1.00
yes	https://www.k...	Madam Memo...	2016	Mar	Art	1	13	0	498	62	950.00	30	12	1.00
no	https://www.k...	A graphic novel...	2016	Apr	Art	1	0	0	215	55	2500.00	30	1	15.00
no	https://www.k...	A database of fr...	2016	Mar	Art	0	0	0	628	47	3395.50	60	4	5.00
no	https://www.k...	Iga Shadow De...	2016	Apr	Art	0	0	0	503	38	1200.00	15	2	5.00
no	https://www.k...	Raven Revolution...	2016	Apr	Art	0	0	2	514	57	3397.74	21	4	4.00
no	https://www.k...	A Half-Century...	2016	Apr	Art	1	0	0	513	49	2000.00	30	6	5.00
no	https://www.k...	Childrens book...	2016	Apr	Art	0	0	0	523	52	14290.95	30	1	26.00
no	https://www.k...	Amateur Artist...	2016	Mar	Art	1	6	0	503	30	4000.00	31	16	4.00

Вид таблицы

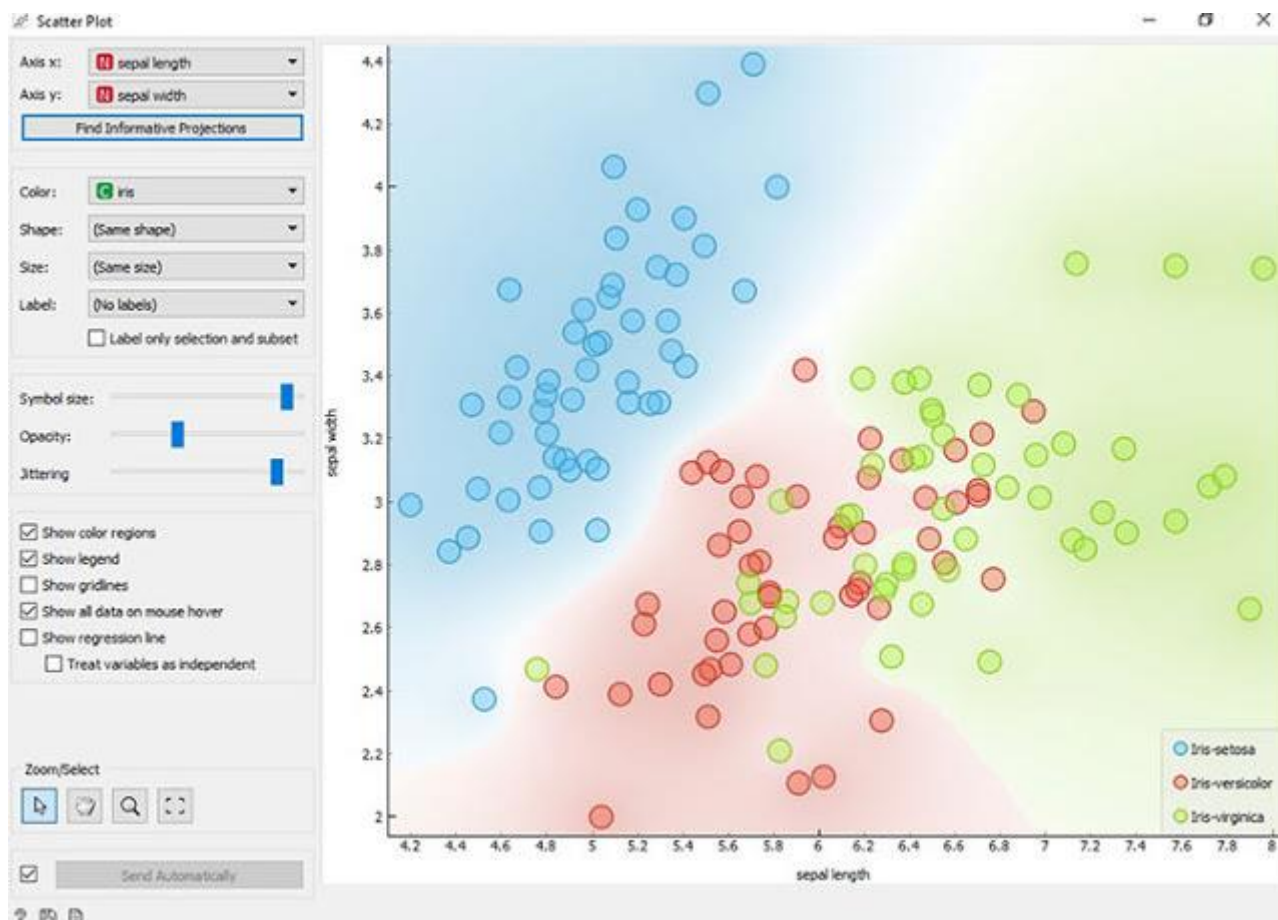
Шаг 5. Теперь дважды щёлкните виджет, чтобы визуализировать таблицу.



Визуализация данных при помощи Orange

Виджет Scatter Plot один из самых популярных в среде Orange. Нажмите на полукруг перед виджетом «File», перетащите его в пустое место в рабочем процессе и выберите виджет «Scatter Plot».

Как только создадите виджет Scatter Plot, дважды щёлкните по нему и изучите данные. Вы можете выбрать оси X и Y, цвета, формы, размеры и другие настройки.



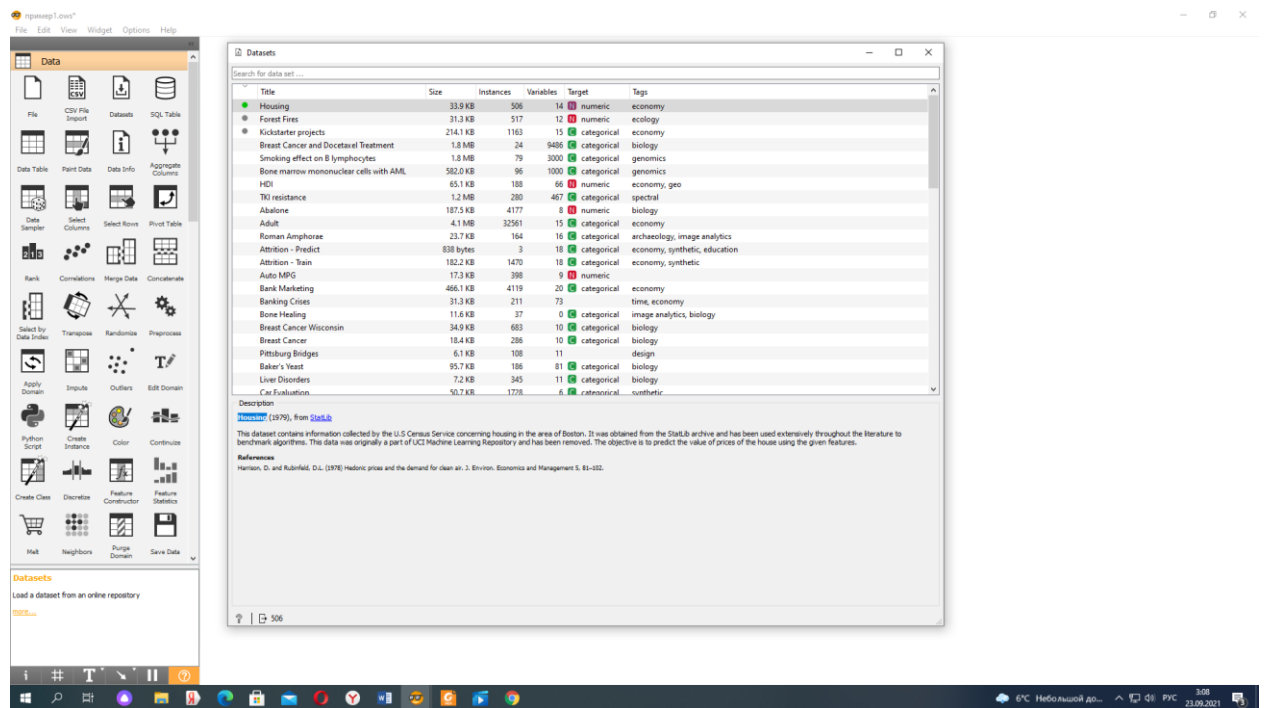
Экспериментируйте, добавляя или меняя виджеты в вашем рабочем процессе.

Использование инструментов работы с данными.

Попробуем построить регрессию. Создайте новый файл.

Выберите «Datasets» щелкните по нему, увидите список. Выберите набор данных *Hausing*

Он переместится вверх, появится зеленая метка, а внизу его краткое описание



Краткое описание набора:

Данные о ценах на жилье в Бостоне, опубликованные Харрисоном Д. и Рубинфельдом Д. Л. 'Hedonic'

Переменные по порядку:

Уровень преступности на душу населения по городам

ZN доля земли под жилую застройку зонирована на участки площадью более 25 000 кв. Футов.

INDUS доля акров, не относящихся к розничной торговле, на город

CHAS фиктивная переменная реки Чарльз (= 1, если участок ограничивает реку; 0 в противном случае)

Концентрация оксидов азота NOX (частей на 10 миллионов)

RM среднее количество комнат в доме

ВОЗРАСТ Доля единиц, занимаемых владельцами, построенных до 1940 г.

DIS взвешивала расстояния до пяти бостонских центров занятости

Индекс доступности радиальных автомобильных дорог РАД

НАЛОГ Полная ставка налога на имущество за 10 000 долларов США.

PTRATIO соотношение учеников и учителей по городам

В 1000 $(B_k - 0,63)^2$, где B_k - доля черных по городам.

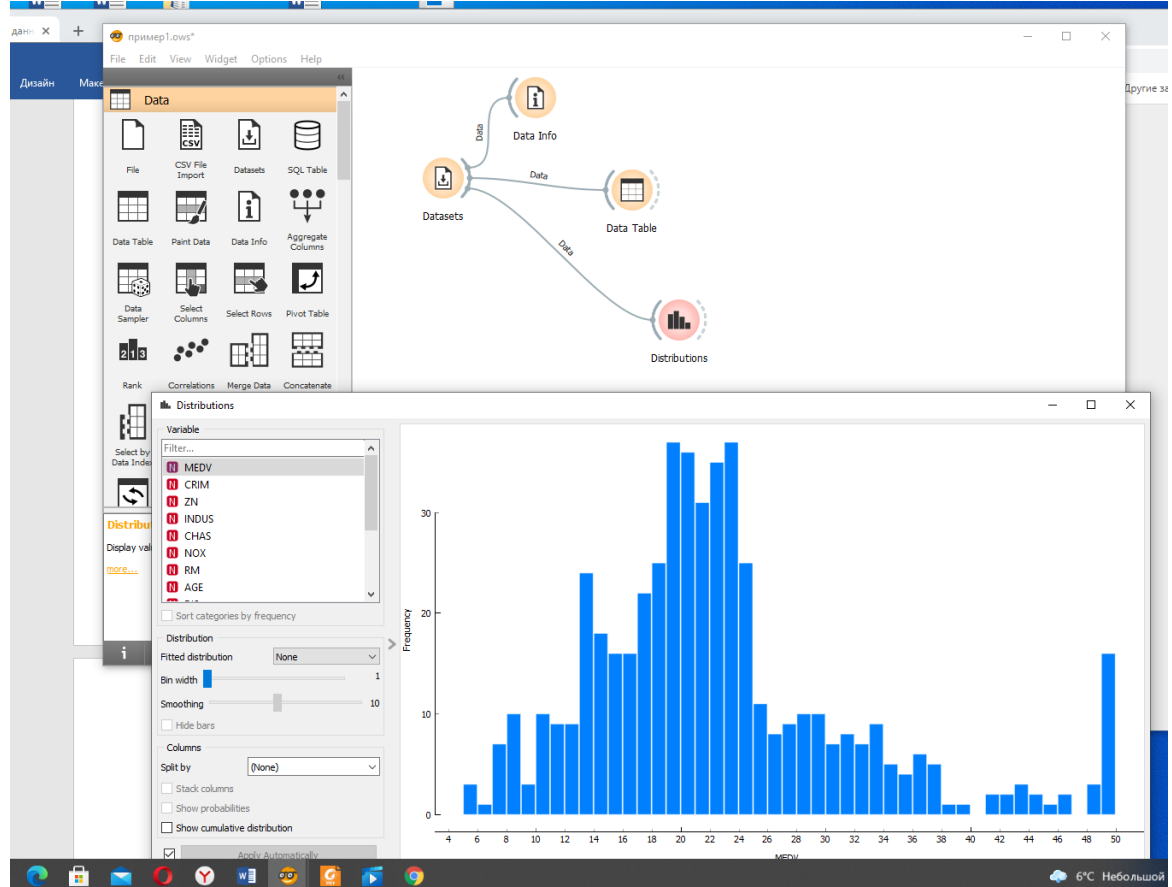
LSTAT% более низкий статус населения

MEDV Средняя стоимость частных домов в 1000 долларов

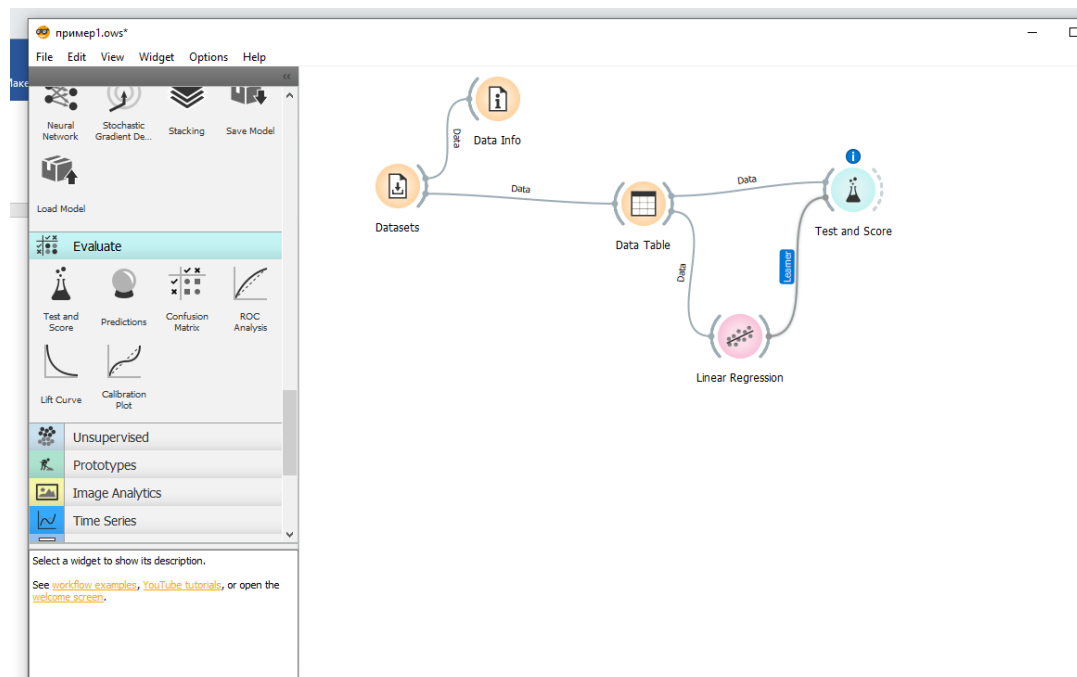
Посмотреть данные можно с помощью «Data Table», он же отвечает за подготовку данных.

Обратите внимание, там помечена целевая переменная. Перед построением каких-либо моделей полезно посмотреть и изучить сами данные. Для этого можно воспользоваться графическими представлениями, в частности, об их распределении. Например, Виджет Distributions. С его

помощью можно посмотреть как распределены числовые переменные

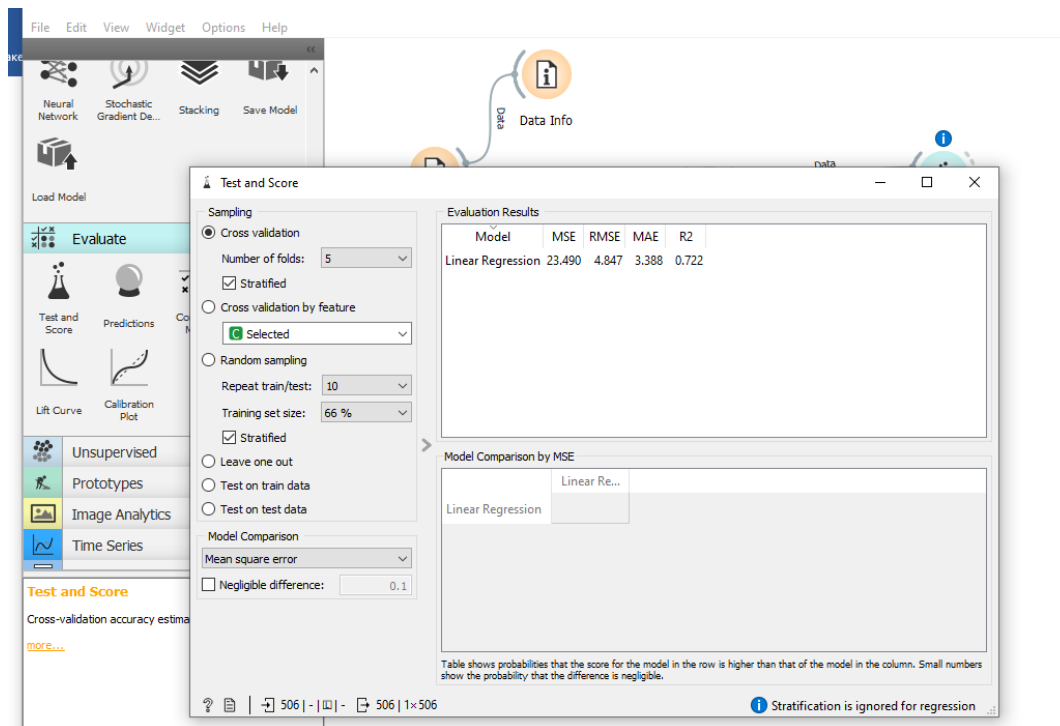


Для того, чтобы построить регрессию, мы используем виджеты:



И устанавливаем связи между ними.

Необходимо посмотреть оценки регрессии. Для этого нужно кликнуть по виджету Test and Score, увидим табличку с оценками:



Смотрим на значения оценок

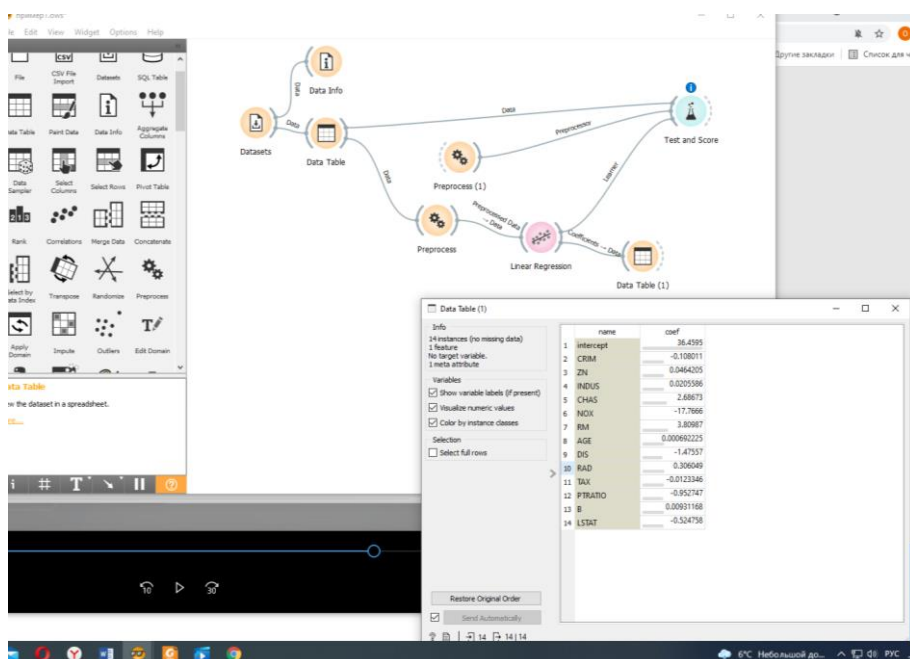
MSE средний квадрат ошибки

RMSE – среднеквадратическое отклонение, это просто квадратный корень из MSE

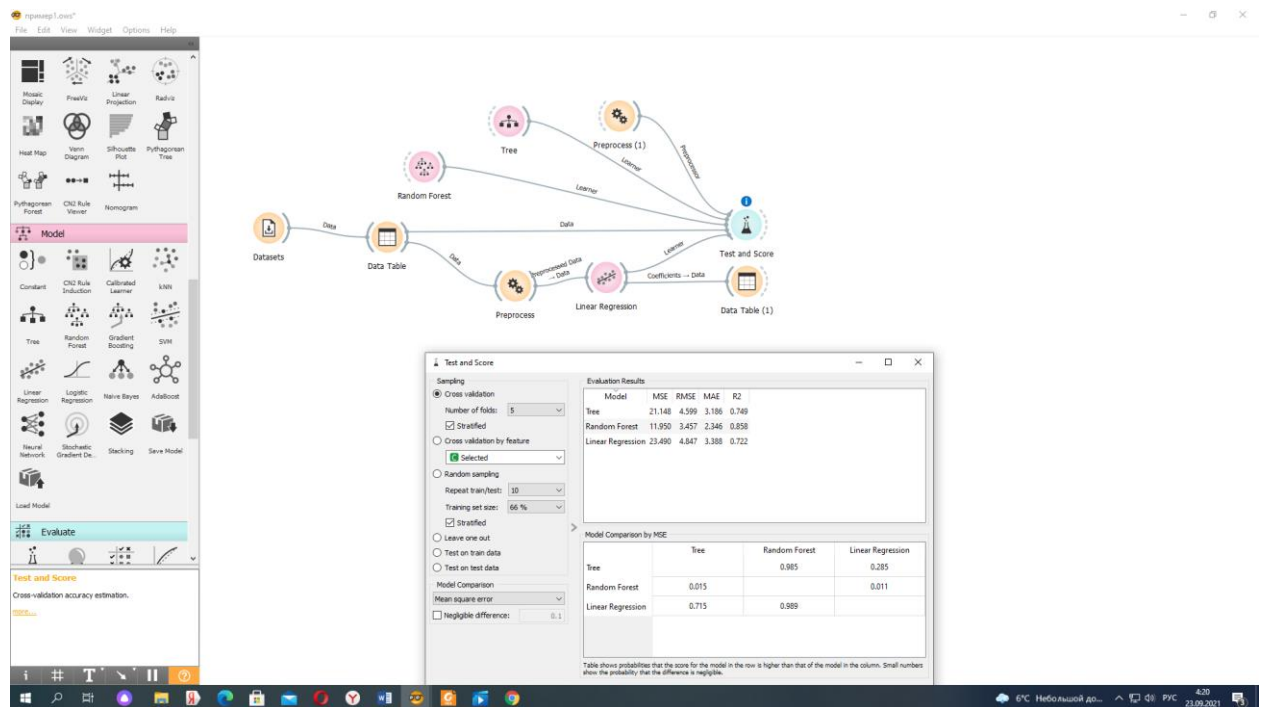
MAE ошибка рассчитывается как среднее абсолютных разностей между целевыми значениями и прогнозами. MAE - это линейная оценка, которая означает, что **все индивидуальные различия взвешены одинаково** в среднем

Коэффициент детерминации, или R^2

Добавим препроцессинг для улучшения расчётов и попробуем построить деревья решений



В таблице мы можем проанализировать коэффициенты регрессии и проанализировать, как они влияют на результат



Посмотрим на визуализацию дерева:

