

2.1. Введение в предобработку и очистку данных

Если анализируемые данные не соответствуют определенным критериям качества, то их предварительная обработка становится необходимым шагом для обеспечения удовлетворительного результата анализа. Необходимость в предварительной обработке данных возникает независимо от того, какие алгоритмы и технологии используются. Более того, эта задача может представлять самостоятельную ценность в областях, не имеющих непосредственное отношение к анализу данных. Очевидно, что исходные данные чаще всего нуждаются в очистке.

Задач, решаемых на этапе очистки данных множество: аномалии, пропуски, шумы и прочие. Ниже описаны отдельные механизмы решения задач очистки данных, реализованные в Loginform Community.

Заполнение пропусков

Часто бывает так, что в столбце некоторые данные отсутствуют в силу каких-либо причин (данные неизвестны, либо их забыли внести и т.п.). Раньше из-за этого пришлось бы убрать из обработки все строки, которые содержат пропущенные данные. Чтобы этого не происходило, в программе предусмотрен обработчик *Заполнение пропусков* для обработки пропусков, содержащий несколько методов для обработки как упорядоченных, так и не упорядоченных полей.

В любом случае, методы этого узла относятся к одномерному анализу, то есть работают (восстанавливают значение) исходя из информации по одному столбцу. Многомерные методы восстановления пропусков, как правило, более эффективны, особенно при решении задач предсказательной аналитики.

Доступны следующие методы обработки:

- *заменять средним* — выявленные пропуски заменяются средним значением столбца;
- *заменять медианой* — выявленные пропуски заменяются медианой, вычисленной по столбцу;
- *заменять наиболее вероятным* — выявленные пропуски заменяются наиболее вероятным значением по столбцу, замена производится на среднее значение из наиболее вероятного интервала, число интервалов варьируется в зависимости от объема выборки — чем она больше, тем больше интервалов;
- *заменять случайными значениями* — выявленные пропуски заменяются случайными значениями, которые формируются из диапазона от наименьшего до наибольшего значения столбца;
- *линейная интерполяция* — выявленные пропуски заменяются на вычисленные промежуточные значения линейной функции, полученной по известным значениям (условно проведя прямую между ними);

- *кубическая интерполяция* — выявленные пропуски заменяются на промежуточные значения кубического сплайна (сплайны третьей степени с непрерывной первой производной), полученного по известным значениям;
- *сплайн-интерполяция* — выявленные пропуски заменяются на промежуточные значения сплайн-функции, полученной по известным значениям;
- *оставить без изменения* — выявленные пропуски заполняться не будут;
- *удалять записи* — строки с выявленными пропусками исключаются из выходного набора данных;
- *заменять значением «Не задано»* — выявленные пропуски заменяются значением «Не задано».

Для каждого поля спектр доступных методов определяется тремя характеристиками данных одновременно:

- *упорядоченностью;*
- *типом;*
- *видом.*

В табл. 2.1 приведены все возможные варианты заполнения пропусков в зависимости от типа данных и того, считается набор данных упорядоченным, или нет.

Таблица 2.1. **Варианты заполнения пропусков**

Метод	Неупорядоченный набор		Упорядоченный набор	
	Непрерывное	Дискретное	Непрерывное	Дискретное
Заменять средним	+	+	+	+
Заменять медианой	+	+	+	+
Заменять наиболее вероятным	+	+	+	+
Заменять случайными значениями	+	+	+	+
Линейная интерполяция			+	+
Кубическая интерполяция			+	+
Сплайн-интерполяция			+	+
Оставить без изменения	+	+		
Удалять записи	+	+		
Заменять значением «Не задано»		+		+

Таким образом, методы восстановления пропусков зависят, в первую очередь, от того, как воспринимать столбец: упорядоченный или неупорядоченный набор данных.

Для неупорядоченных данных интересны два метода:

- *заменять наиболее вероятным* — замена производится на среднее значение из наиболее вероятного интервала, число интервалов варьируется в зависимости от объема выборки — чем она больше, тем больше интервалов;
- *заменять случайными значениями* — производится замена пропусков на случайное значение из распределения, параметры которого оцениваются из имеющихся значений в столбце.

Остальные методы — заменять средним, удалять, оставлять без изменений, заменять каким-либо значением, тривиальны и эффективны только для устранения единичных пропусков. Удаление большого числа пропусков — некорректный подход, так как теряется часть информации (поскольку остальные поля могут быть заполнены).

Метод интерполяции доступен только для упорядоченных данных, чаще всего это временные ряды. Восстановление пропусков в столбцах, значения в которых упорядочены, можно рассматривать как интерполяцию значений функции в точках, где она неизвестна.

Редактирование выбросов

Выбросы — это отклонения от нормального (ожидаемого) поведения чего-либо. Это может быть, например, резкое отклонение величины от ее ожидаемого значения. Под выбросами при этом подразумеваются значения данных, существенно отклоняющиеся от средних, а под экстремальными — значения, которые настолько сильно отклоняются от типичных значений, что перестают соответствовать логике исследуемых процессов и явлений.

Для автоматической корректировки выбросов и экстремальных значений в наборах данных предназначен обработчик *Редактирование выбросов*, в котором — доступны два метода выявления:

- *стандартное отклонение* — критерием является отклонение значения признака от среднего более, чем на заданное число стандартных отклонений. При этом данный параметр отдельно задается для выбросов и для экстремальных значений. Данный метод следует применять, если известно, что распределение данных близко к нормальному.

- *интерквартильный размах* — критерием является расстояние между 1-м и 3-м квартилями распределения значений признака. Если значение признака отклоняется от медианы более, чем на заданное число интерквартильного размаха, то оно считается аномальным. Данный параметр задается отдельно для выбросов и экстремальных значений. Этот метод можно применять и в случае, когда распределение данных отличается от нормального.

Как для выбросов, так и для экстремальных значений доступны следующие методы редактирования:

- *оставить без изменения*;
- *удалять записи* — исключить строки с аномальными значениями из набора данных;
- *заменять средним* — заменять аномалии средним значением столбца;
- *заменять медианой* — заменять аномалии медианой, вычисленной по столбцу;
- *заменять наиболее вероятным* — замена аномалий наиболее вероятным значением по столбцу, замена производится на среднее значение из наиболее вероятного интервала, число интервалов варьируется в зависимости от объема выборки — чем она больше, тем больше интервалов;
- *заменять заданным значением* — замена аномалий на значение, прописанное вручную;

- *ограничивать* — аномалии будут заменены значением границы, с которой начинается определение аномалии.

Для каждого поля спектр доступных методов определяется тремя характеристиками данных одновременно:

- *упорядоченностью*;
- *типом*;
- *видом*.

В табл. 2.2 приведены все возможные варианты редактирования выбросов в зависимости от типа данных и того, считается набор данных упорядоченным, или нет. Для экстремальных значений набор методов точно такой же.

Таблица 2.2. **Варианты редактирования выбросов**

Метод	Неупорядоченное поле		Упорядоченное поле	
	Непрерывное	Дискретное	Непрерывное	Дискретное
Оставить без изменения	+	+	+	+
Удалять записи	+	+		
Заменять средним	+	+	+	+
Заменять медианой	+	+	+	+
Заменять наиболее вероятным	+	+	+	+
Заменять заданным значением		+		+
Ограничивать	+	+	+	+

Нетрудно увидеть, что задача редактирования выбросов во многих случаях сводится к задаче заполнения пропусков — выброс «объявляется» пропущенным значением, и к нему применяются те или иные методы, доступные также и в узле *Заполнение пропусков*.

Представляет интерес метод *Ограничивать* — приведение экстремального значения или выброса к значению, превышение которого определяется как выброс (по умолчанию для выброса это 3-сигма, для экстремального значения 5-сигма).

Аналитику следует помнить, что важна последовательность применения узлов — можно сначала отредактировать выбросы, а затем заполнить пропуски, а можно — наоборот. В первом и втором случае результаты будут разными, и конкретная последовательность определяется исходными данными и решаемой задачей.

2.2. Методические указания

2.2.1. Неупорядоченный набор данных

В файле *Задача 2.1. Продажа автомобилей.xlsx*¹ имеются данные о числе продаж и перепродаж через четыре года автомобилей разных марок (тыс. шт.), а

¹ Здесь и далее исходные данные можно скачать с Яндекс.Диска по ссылке <https://yadi.sk/d/N8r-8cGBzYLUwg>

также различные характеристики этих автомобилей (цена, мощность двигателя, база шасси, емкость топливного бака и т.п.) (рис. 2.1).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Производитель	Модель	Продажи	Перепродажи	Тип	Цена	Объем двигателя	Мощность	Колесная база	Ширина	Длина	Масса	Емкость бака	Пробег
2	Acura	Integra	16,919	16,360	0	21,500	1,8	140	101,2	67,3	172,4	2,639	13,2	28
3	Acura	TL	39,384	19,875	0	28,400	3,2	225	108,1	70,3	192,9	3,517	17,2	25
158	Volkswagen	S80	18,969		0	36,000	2,9	201	109,9	72,1	189,8	3,600	21,1	24

Рис. 2.1

Требуется произвести комплексную оценку качества набора данных на основе количества обнаруженных пропусков и выбросов, а также определить методы их обработки.

Решение

Создадим новый пакет *Предобработка и очистка данных* (рис. 2.2–2.3).

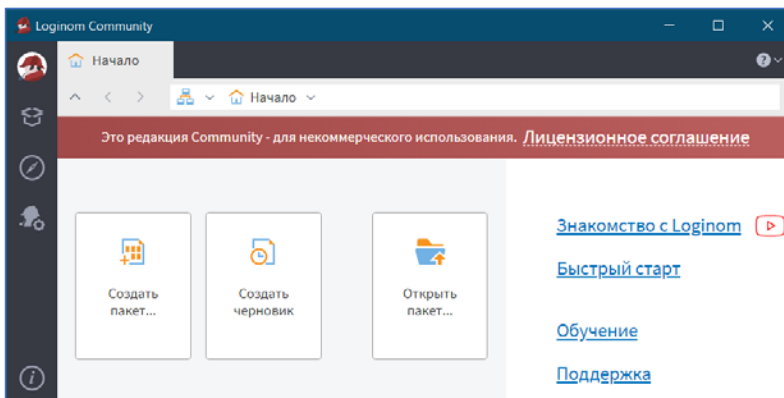


Рис. 2.2

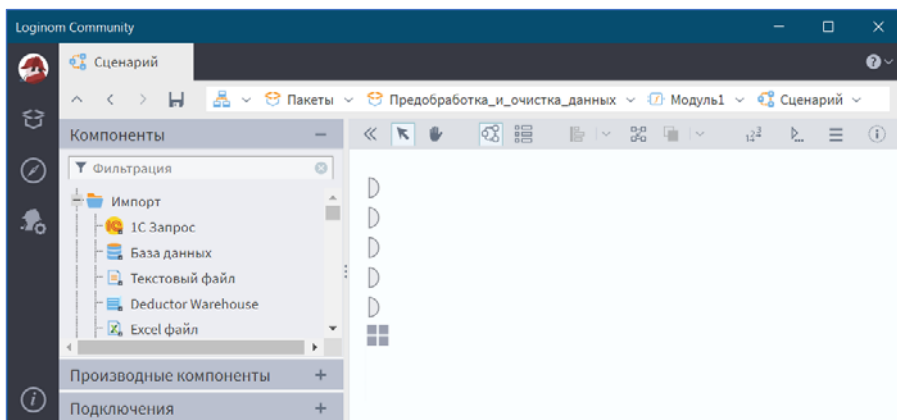


Рис. 2.3

Выполним импорт исходных данных. Для этого выберем в разделе *Импорт* компонент *Excel файл* и перенесем его в область построения (рис. 2.4). В результате создастся узел сценария, выполняющий действие импорта.

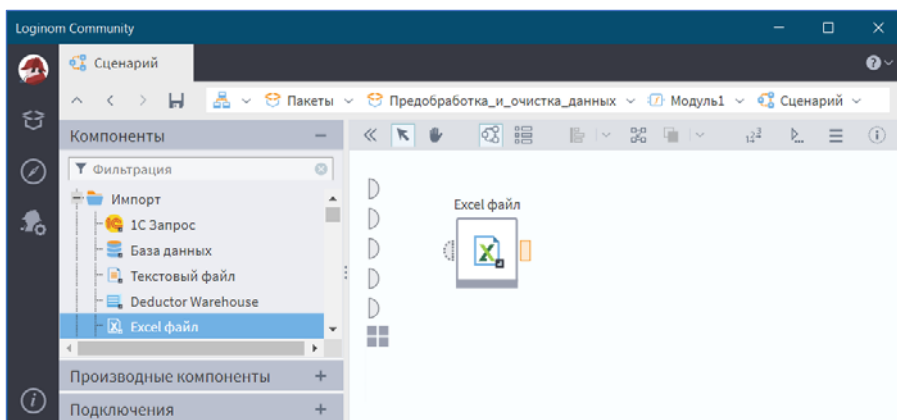


Рис. 2.4

При клике мышкой на узле отобразятся иконки возможных действий (рис. 2.5).

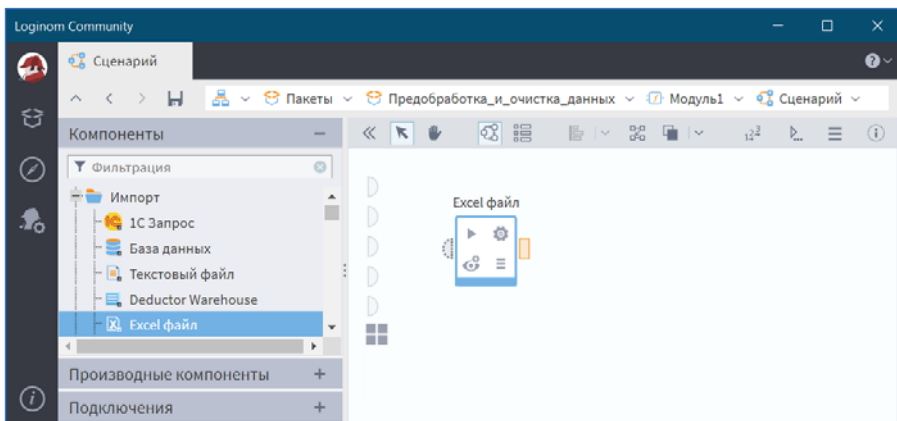


Рис. 2.5

Вызовем *Мастер настройки* (рис. 2.6).

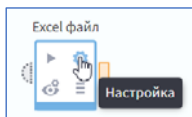


Рис. 2.6

Пройдем шаги мастера, указав в параметре *Имя файла* местоположение файла *Задача 2.1. Продажа автомобилей.xlsx* (рис. 2.7–2.9).

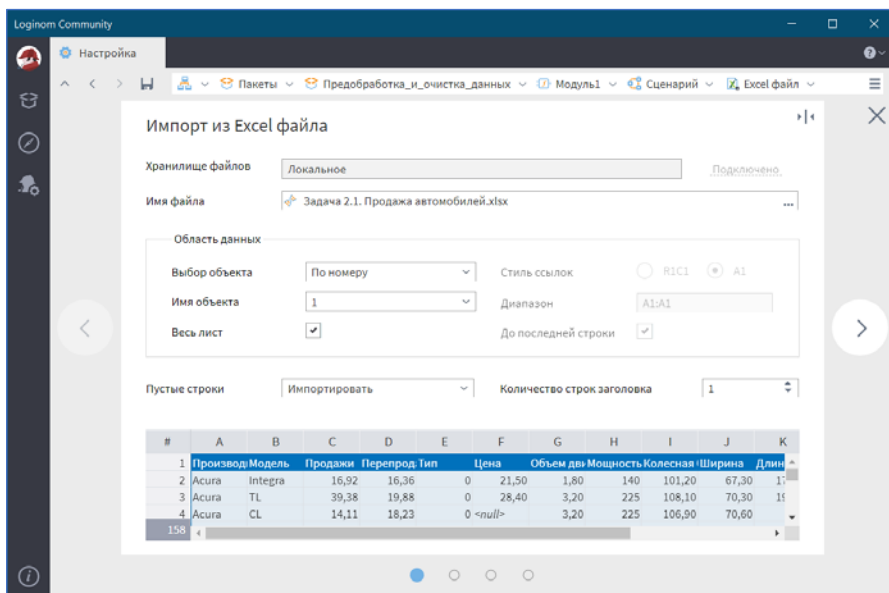


Рис. 2.7

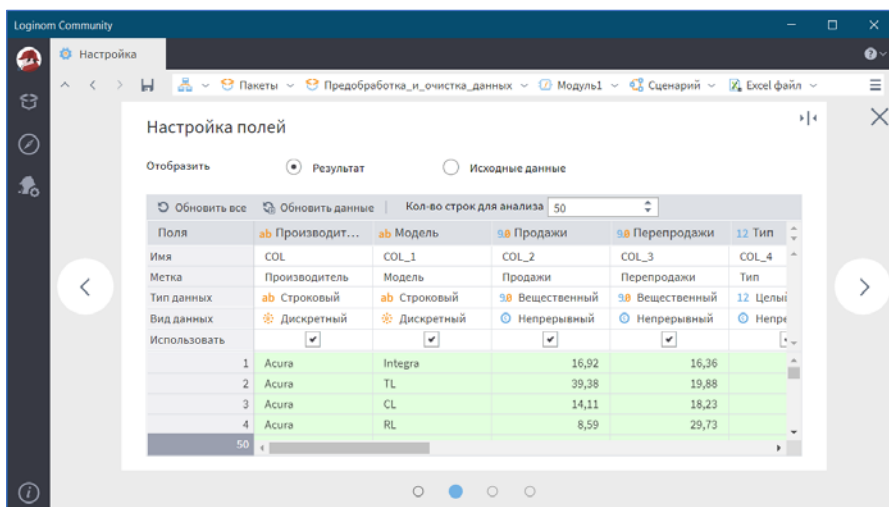


Рис. 2.8

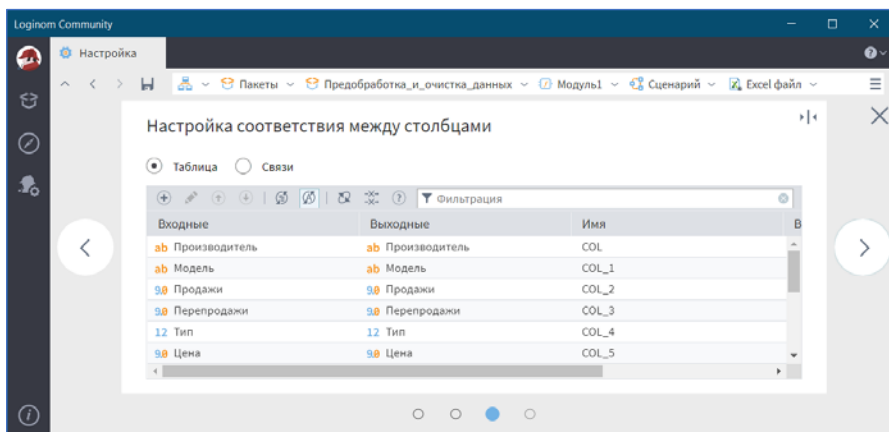


Рис. 2.9

После настройки узла выполним его, используя меню возможных действий (рис. 2.10–2.11).

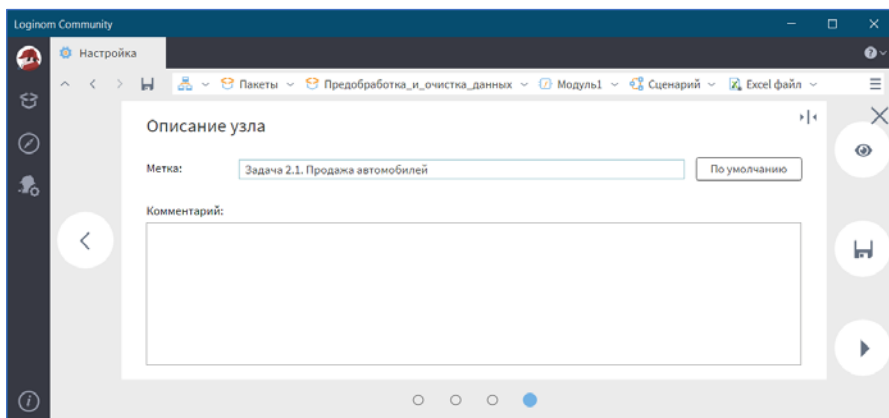


Рис. 2.10

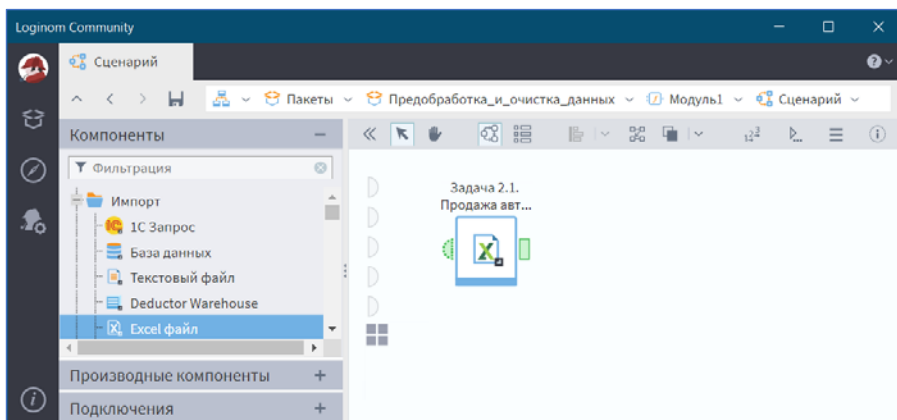


Рис. 2.11

Теперь в выходном порте узла присутствуют импортированные данные, которые можно увидеть, выбрав *Быстрый просмотр...* в контекстном меню порта *Набор данных* (рис. 2.12).

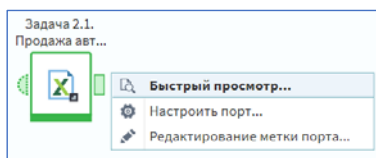


Рис. 2.12

Таблица с импортированными данными имеет вид (рис. 2.13).

#	ab	Производител...	ab	Модель	9,0	Продажи	9,0	Перепродажи	12	Тип	9,0	Цена	9,0
1		Acura		Integra		16,92		16,36		0		21,50	
2		Acura		TL		39,38		19,88		0		28,40	
3		Acura		CL		14,11		18,23		0			
4		Acura		RL		8,59		29,73		0		42,00	
5		Audi		A4		20,40		22,26		0		23,99	
6		Audi		A6		18,78		23,56		0		33,95	
7		Audi		A8		1,38		39,00		0		62,00	
8		BMW		323i		19,75				0		26,99	
9		BMW		328i		9,23		28,68		0		33,40	
10		BMW		528i		17,53		36,13		0		38,90	
11		Buick		Century		91,56		12,48		0		21,98	
12		Buick		Regal		39,35		13,74		0		25,30	
13		Buick		Park Avenue		27,85		20,19		0		31,97	

Рис. 2.13

Добавим визуализаторы к узлу сценария. Для этого кликнем мышкой по узлу и перейдем к настройкам визуализаторов (рис. 2.14–2.15).

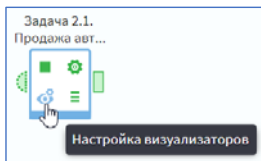


Рис. 2.14

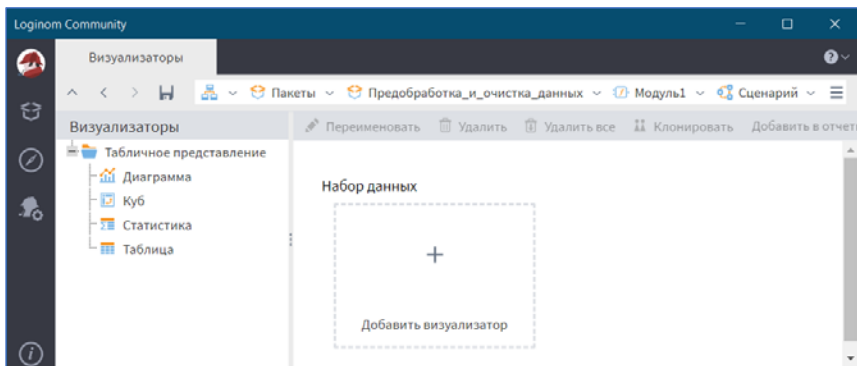


Рис. 2.15

Добавим визуализаторы *Таблица* и *Статистика*. Для этого выберем в дереве нужные визуализаторы и нажмем кнопку *Добавить визуализатор* у нужного выходного порта (рис. 2.16). Также это можно сделать, перетаскив мышкой необходимый визуализатор в область кнопки *Добавить визуализатор* у нужного выходного порта.

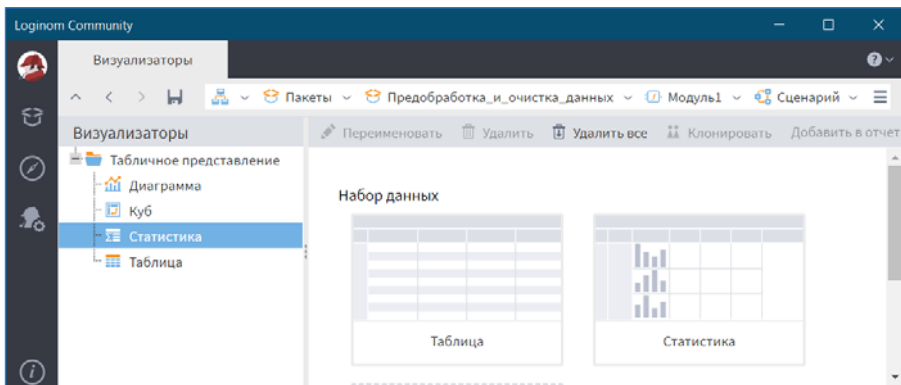


Рис. 2.16

Для каждого визуализатора может быть добавлен отчет. Для добавления отчета к визуализатору необходимо выбрать нужный визуализатор и нажать кнопку *Добавить в отчеты*, создав при необходимости новую группу или разместить отчет в существующую группу.

Откроем визуализатор *Таблица*. Для этого нажмем кнопку *Войти* (рис. 2.17).

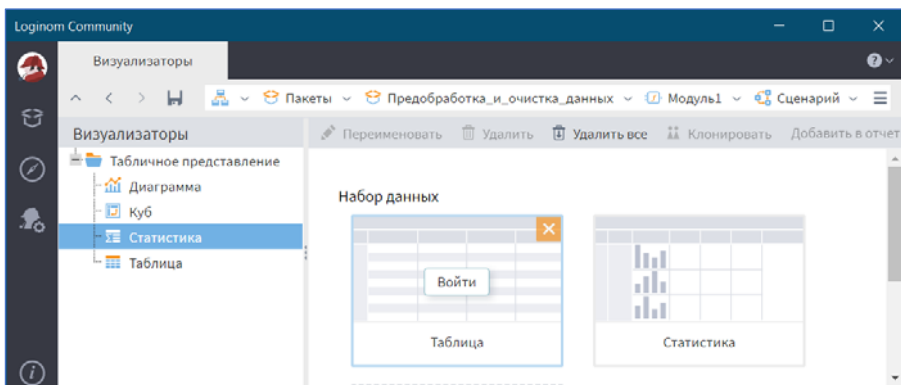


Рис. 2.17

В результате будет выведена таблица с исходными данными (рис. 2.18).

#	ab Произво...	ab Модель	98 Продажи	98 Перепродажи	12 Тип	98 Цена	98 Объем двигателя	12 Моу
1	Acura	Integra	16,919	16,36	0	21,5		1,8
2	Acura	TL	39,384	19,875	0	28,4		3,2
3	Acura	CL	14,114	18,225	0			3,2
4	Acura	RL	8,588	29,725	0	42		3,5
5	Audi	A4	20,397	22,255	0	23,99		1,8
6	Audi	A6	18,78	23,555	0	33,95		2,8
7	Audi	A8	1,38	39	0	62		4,2
8	BMW	323i	19,747		0	26,99		2,5
9	BMW	328i	9,231	28,675	0	33,4		2,8
10	BMW	528i	17,527	36,125	0	38,9		2,8

Рис. 2.18

Визуализатор *Статистика* предназначен для просмотра различных статистических показателей по каждому полю набора данных и представляет собой таблицу, в которой наименования полей набора данных расположены в строках, а наименования статистических показателей в столбцах. На пересечении, в

ячейках таблицы, расположены значения статистических показателей соответствующих полей (рис. 2.19).

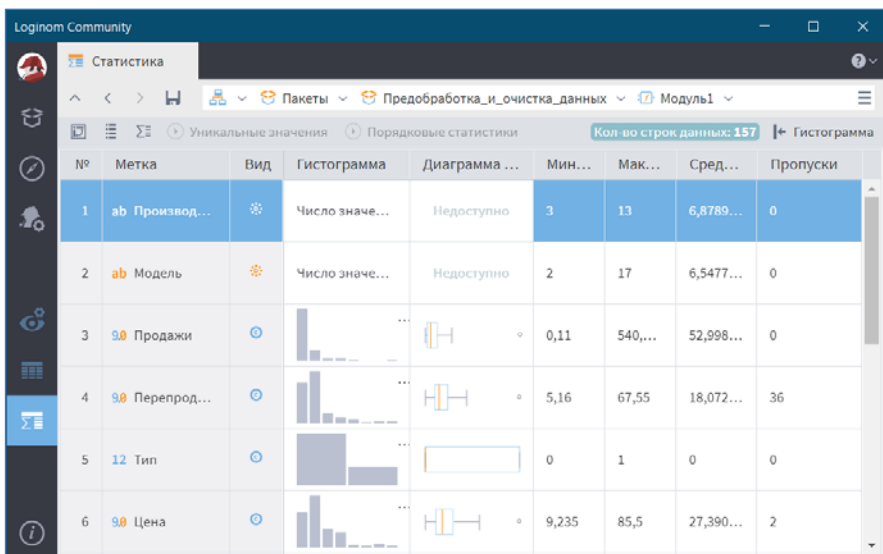


Рис. 2.19

По умолчанию в таблице отображаются 8 показателей:

- гистограмма;
- диаграмма размаха;
- минимум;
- максимум;
- среднее;
- стандартное отклонение;
- пропуски;
- уникальные.

В случае, если их недостаточно, весь перечень доступен с помощью операции *Настройка показателей* (рис. 2.20).

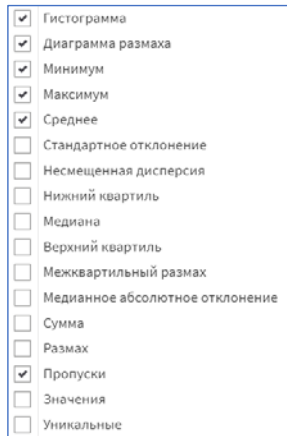


Рис. 2.20

Поскольку по десяти показателям выявлены пропуски, проведем их обработку на основе импортированных данных. Для этого переместим компонент *Заполнение пропусков* в рабочую область сценария. Последовательность обработки данных задается соединением выходного порта узла импорта с входным портом заполнения пропусков (рис. 2.21).

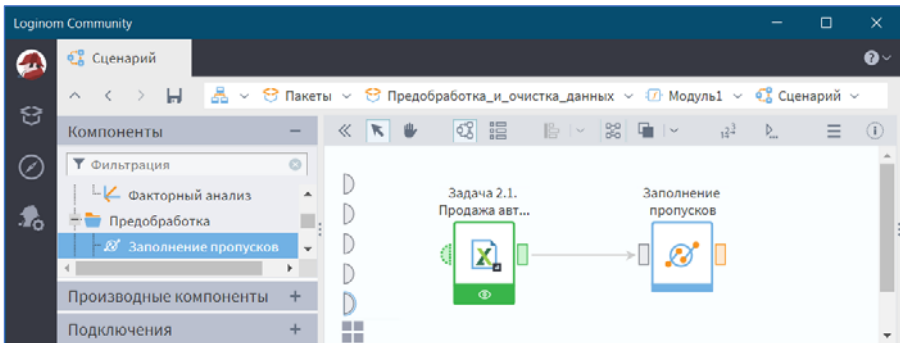


Рис. 2.21

Пройдем шаги *Мастера настройки*. На шаге 1 настроим параметры полей. Наибольшее количество пропусков имеется по столбцу *Перепродажи*. Ни один из возможных методов заполнения пропусков в данном случае применять нельзя, поскольку неизвестно будет ли в дальнейшем осуществляться перепродажа автомобилей. Следовательно, просто удалим данные записи. Для заполнения пропусков по остальным столбцам применим метод *Заменять наиболее вероятным* (рис. 2.22).

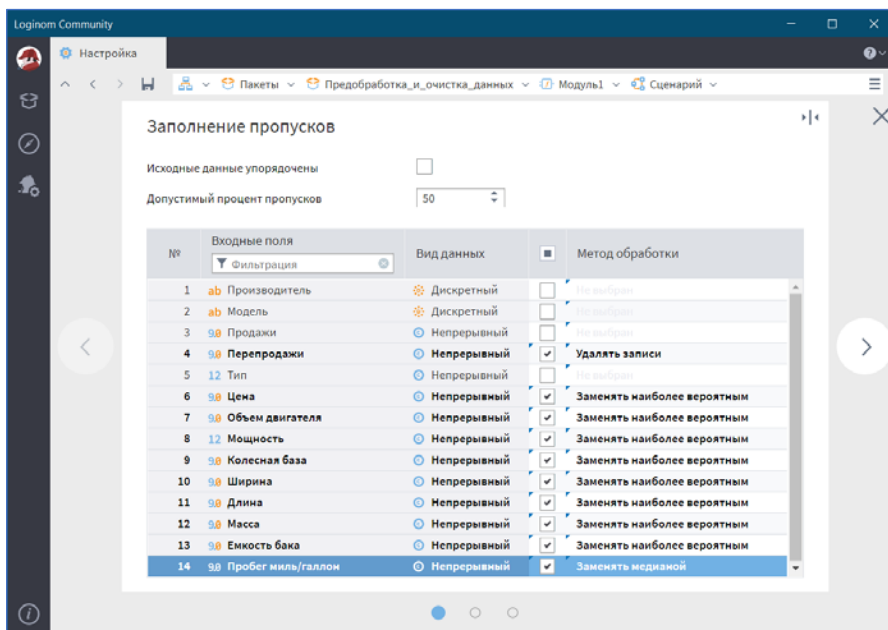


Рис. 2.22

На шаге 2 оставим стандартные параметры по умолчанию (рис. 2.23).

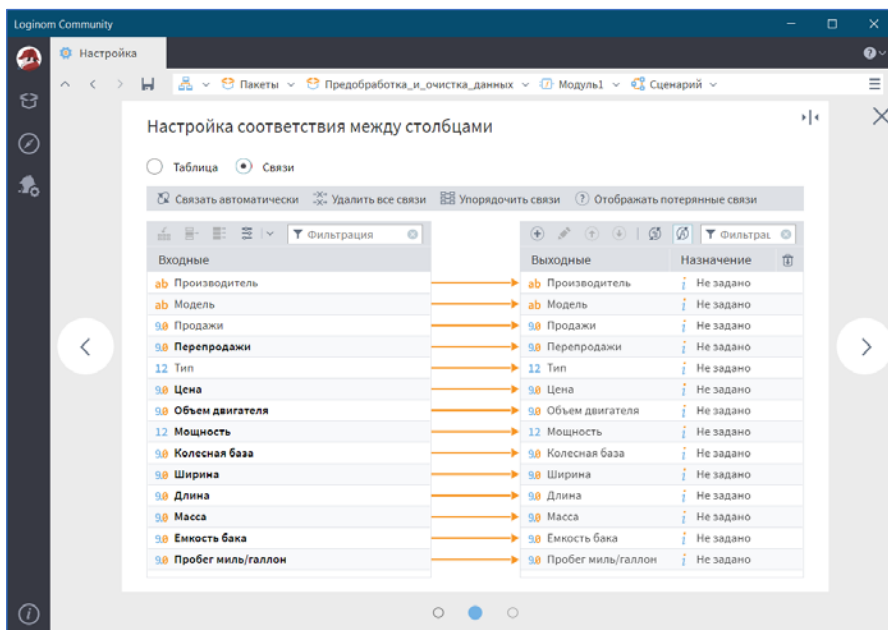


Рис. 2.23

Добавим визуализаторы *Таблица* и *Статистика* к узлу сценария. Для этого кликнем мышкой по узлу и перейдем к настройкам визуализаторов.

В визуализаторах будут представлены результаты обработки пропусков (рис. 2.24–2.25).

Loginom Community							
Таблица							
#	ab Производитель...	ab Модель	9.0 Продажи	9.0 Перепродажи	12 Тип	9.0 Цена	9.0
1	Acura	Integra	16,919	16,36	0	21,5	
2	Acura	TL	39,384	19,875	0	28,4	
3	Acura	CL	14,114	18,225	0	21,94583333	
4	Acura	RL	8,588	29,725	0	42	
5	Audi	A4	20,397	22,255	0	23,99	
6	Audi	A6	18,78	23,555	0	33,95	
7	Audi	A8	1,38	39	0	62	
8	BMW	328i	9,231	28,675	0	33,4	
9	BMW	528i	17,527	36,125	0	38,9	
10	Buick	Century	91,561	12,475	0	21,975	
121							
Страница 1 из 1							

Рис. 2.24

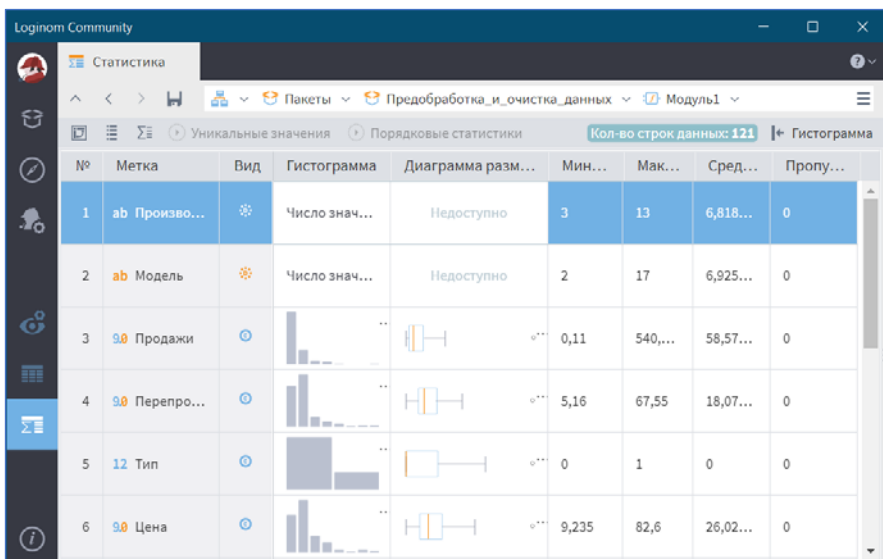


Рис. 2.25

Отредактируем выбросы и экстремальные значения. Для этого переместим компонент *Редактирование выбросов* в рабочую область сценария. Последовательность обработки данных задается соединением выходного порта узла заполнения пропусков с входным портом редактирования выбросов (рис. 2.26).

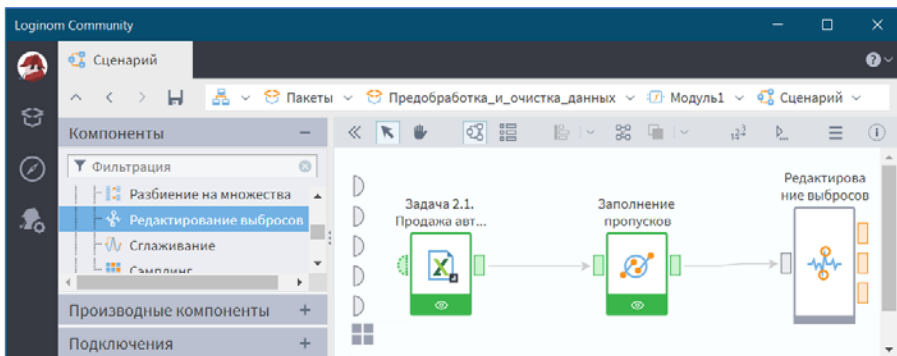


Рис. 2.26

Пройдем шаги *Мастера настройки*. На шаге 1 настроим параметры полей. Для всех вещественных переменных применим метод обработки выбросов *Ограничивать* (рис. 2.27).

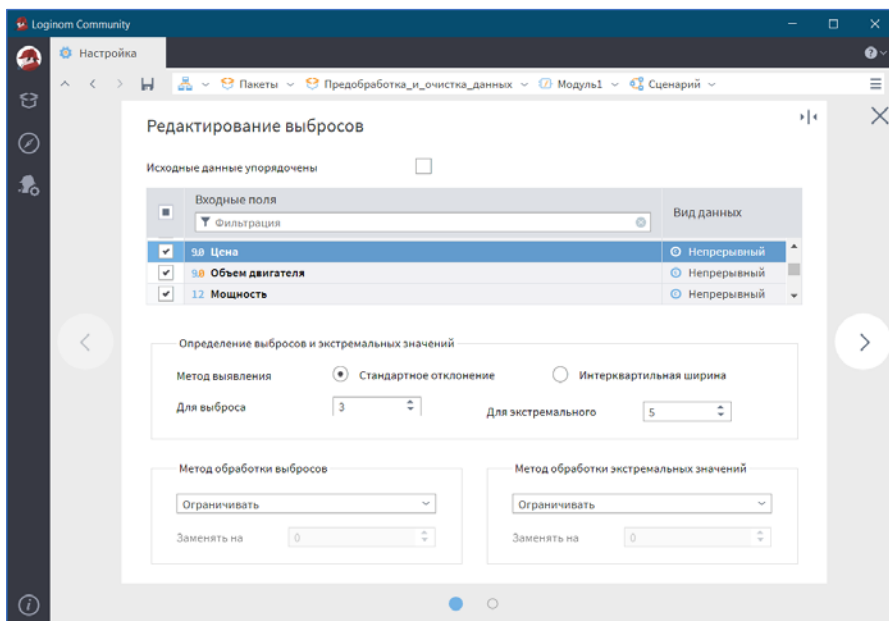


Рис. 2.27

Добавим визуализаторы *Выходной набор*, *Выбросы* и *Экстремальные значения* к узлу сценария. Для этого кликнем мышкой по узлу и перейдем к настройкам визуализаторов (рис. 2.28).

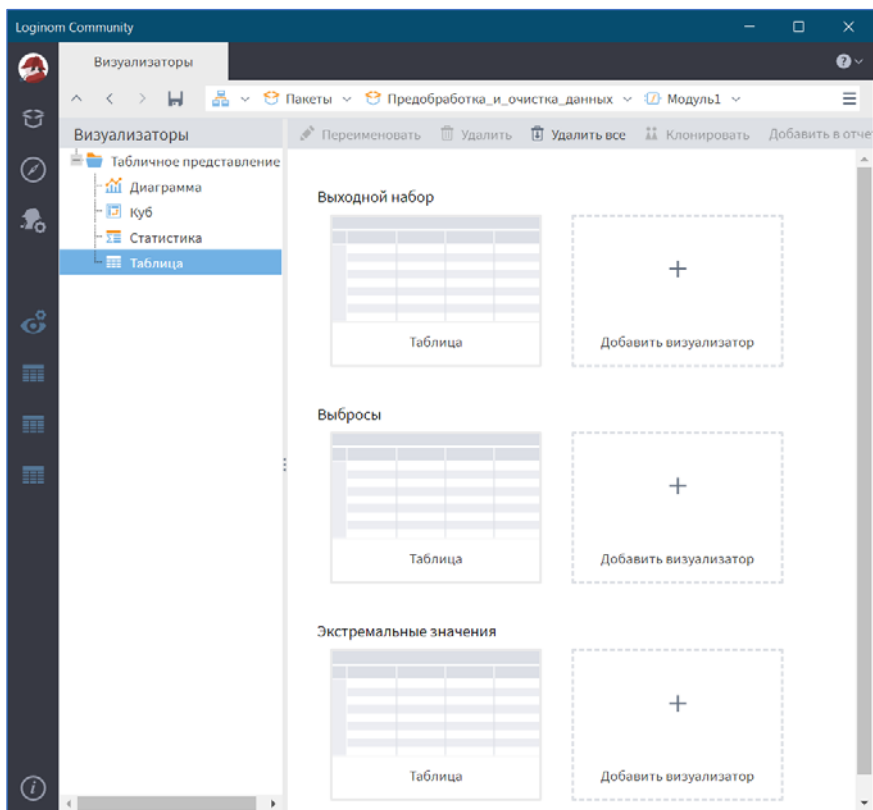


Рис. 2.28

В визуализаторах будут представлены результаты обработки пропусков (рис. 2.29–2.31).

Loginom Community

Таблица

Пакеты Предобработка_и_очистка_данных Модуль1

Формат | Сортировка | Фильтр | Найти | XLS | Детал

#	ab Производителе...	ab Модель	9.0 Продажи	9.0 Перепродажи	12 Тип	9.0 Цена	9.0 Объем дв
1	Acura	Integra	16,919	16,36	0	21,5	
2	Acura	TL	39,384	19,875	0	28,4	
3	Acura	CL	14,114	18,225	0	21,94583333	
4	Acura	RL	8,588	29,725	0	42	
5	Audi	A4	20,397	22,255	0	23,99	
6	Audi	A6	18,78	23,555	0	33,95	
7	Audi	A8	1,38	39	0	62	
8	BMW	328i	9,231	28,675	0	33,4	
9	BMW	528i	17,527	36,125	0	38,9	
10	Buick	Century	91,561	12,475	0	21,975	

121

К < | Страница 1 | > X

Страница 1 из 1

Рис. 2.29. Выходной набор

Loginom Community

Таблица

Пакеты Предобработка_и_очистка_данных Модуль1

Формат | Сортировка | Фильтр | Найти | XLS | Дет

#	ab Производителе...	ab Модель	9.0 Продажи	9.0 Перепродажи	12 Тип	9.0 Цена	9.0 Объем дви
1	Chevrolet	Metro	21,855	5,16	0	9,235	
2	Dodge	Viper	0,916	58,47	0	69,725	
3	Dodge	Ram Pickup	227,061	15,06	1	19,46	
4	Dodge	Ram Wagon	16,767	15,51	1	21,315	
5	Dodge	Ram Van	31,038	13,425	1	18,575	
6	Ford	F-Series	540,561	15,075	1	26,935	
7	Mercedes-Benz	S-Class	16,774	50,375	0	69,7	
8	Mercedes-Benz	SL-Class	3,311	58,6	0	82,6	
9	Porsche	Carrera Coupe	1,28	60,625	0	71,02	
10	Porsche	Carrera Cabriolet	1,866	67,55	0	74,97	
11	Toyota	Land Cruiser	9,835	34,08	1	51,728	

К < | Страница 1 | > X

Страница 1 из 1

Рис. 2.30. Выбросы

#	ab Производител...	ab Mode...	9.0 Продажи	9.0 Перепродажи	12 Тип	9.0 Цена	9.0 Объем двигателя
1	Ford	F-Series	540,561	15,075	1	26,935	4,6

Рис. 2.31. Экстремальные значения

Сохраним пакет (рис. 2.32).²

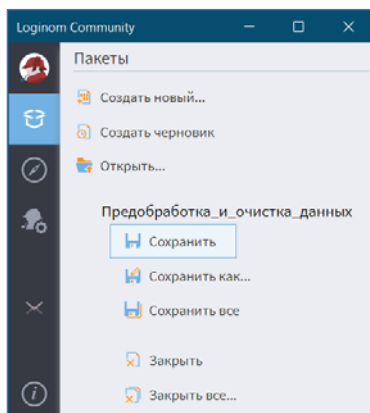


Рис. 2.32

2.2.2. Упорядоченный набор данных

В файле *Задача 2.2. Стоимость набора продуктов питания.xlsx* имеются данные о стоимости условного (минимального) набора продуктов питания в г. Москве за 5 лет (рис. 2.33).

	A	B	C
	№ периода	Дата	Стоимость условного (минимального) набора продуктов питания в г. Москве, руб.
1			
2	1	2016, янв.	4308,13
3	2	2016, фев.	4349,98
4	3	2016, мар.	
5	4	2016, апр.	4430,48
55	54	2020, июн.	5519,29

Рис. 2.33

² Данную процедуру следует применять по окончании выполнения каждого проекта.

Для прогнозирования данного показателя на перспективу необходимо заполнить пропуски, которые визуальны видны в таблице.

Решение

Выполним импорт исходных данных. Для этого создадим узел сценария, выполняющий действие импорта (рис. 2.34).

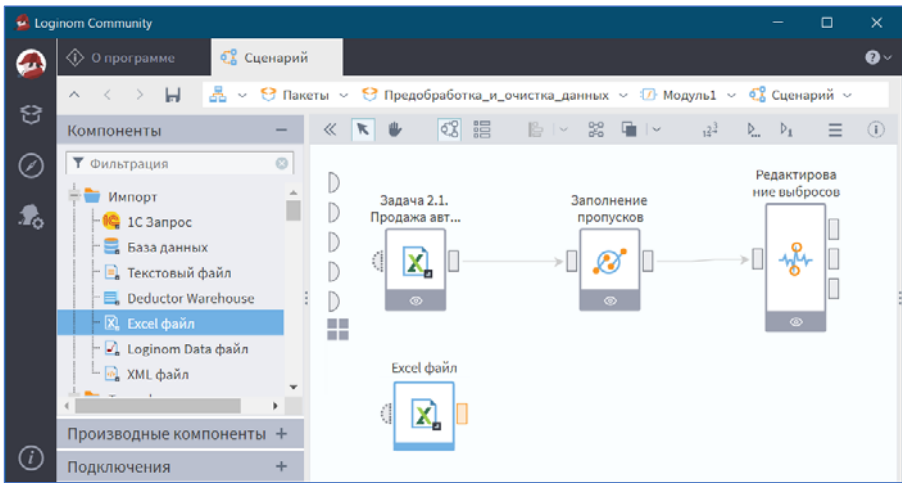


Рис. 2.34

Пройдем шаги *Мастера настройки*. На шаге *Настройка соответствия между столбцами* изменим назначение исходных столбцов данных на активное (рис. 2.35), указав в описании узла метку *Задача 2.2. Стоимость набора продуктов питания*.

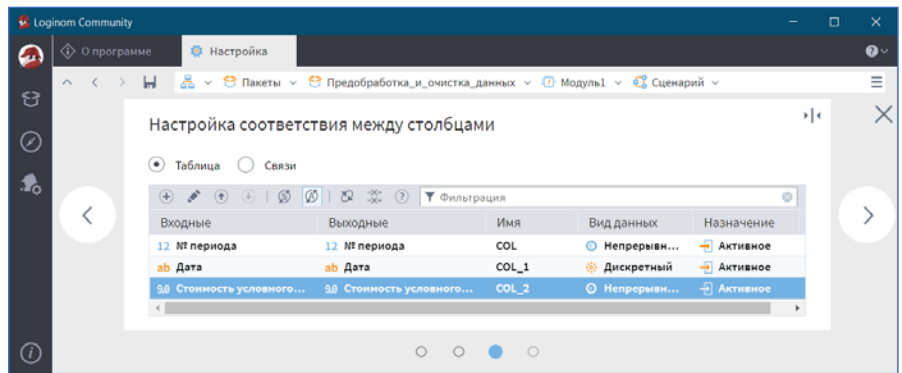


Рис. 2.35

Добавим визуализаторы *Таблица* и *Диаграмма* к узлу сценария (рис. 2.36).

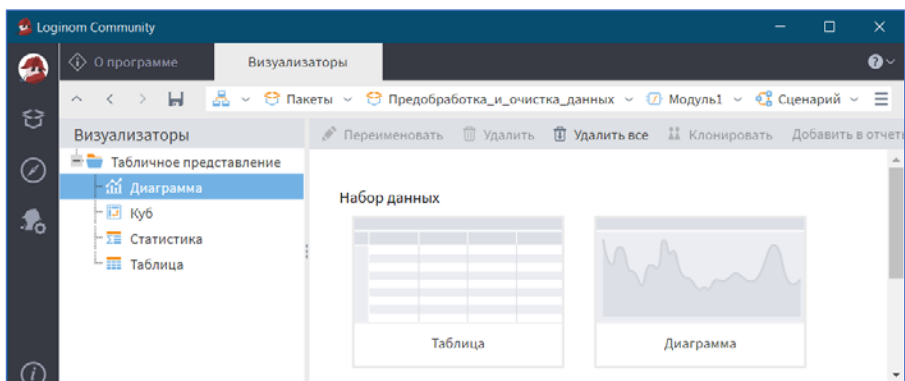


Рис. 2.36

Таблица с исходными данными и диаграмма имеют вид (рис. 2.37–2.38).

#	№ периода	ab Дата	9b Стоимость условного...
1	1	2016, янв.	4308,13
2	2	2016, фев.	4349,98
3	3	2016, мар.	
4	4	2016, апр.	4430,48
5	5	2016, май	4558,09
6	6	2016, июн.	4631,66
7	7	2016, июл.	4630,85
8	8	2016, авг.	4514,46
9	9	2016, сен.	4409,94
10	10	2016, окт.	
11	11	2016, ноя.	
12	12	2016, дек.	
55	13	2017, янв.	4505,39

Рис. 2.37

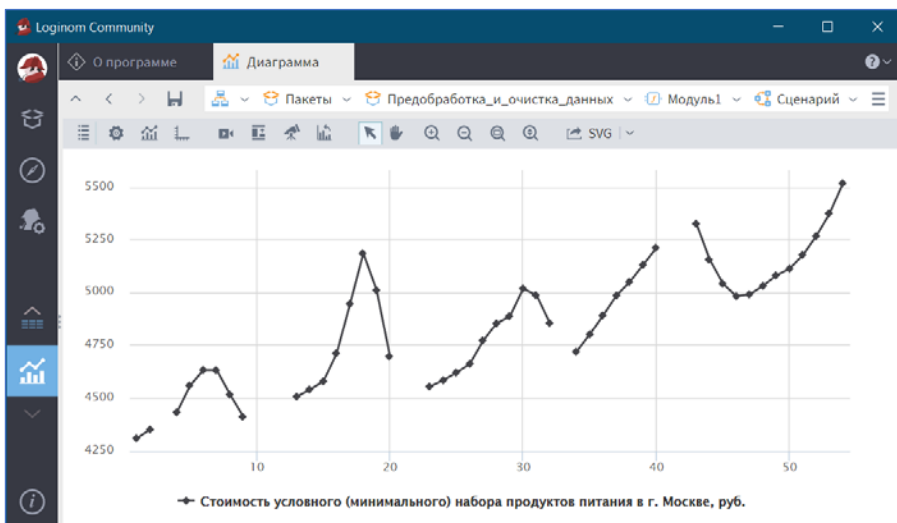


Рис. 2.38

Заполним пропущенные данные. Для этого переместим компонент *Заполнение пропусков* в рабочую область сценария. Последовательность обработки данных задается соединением выходного порта узла импорта с входным портом заполнения пропусков (рис. 2.39).

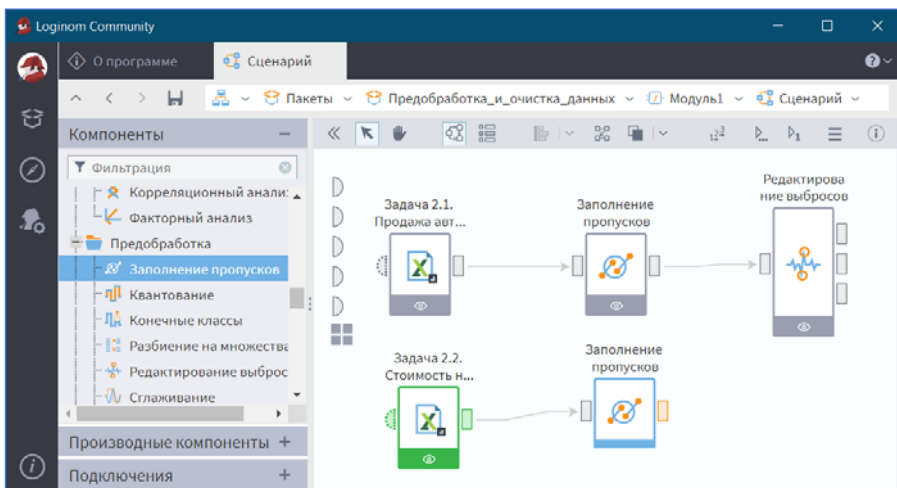


Рис. 2.39

Пройдем шаги *Мастера настройки*. На шаге 1 укажем *Исходные данные упорядочены* и настроим параметры полей. Для заполнения пропусков по

столбцу *Стоимость условного (минимального) набора продуктов питания* применим метод *Сплайн-интерполяция*, позволяющий соединить точки гладкой кривой (рис. 2.40).

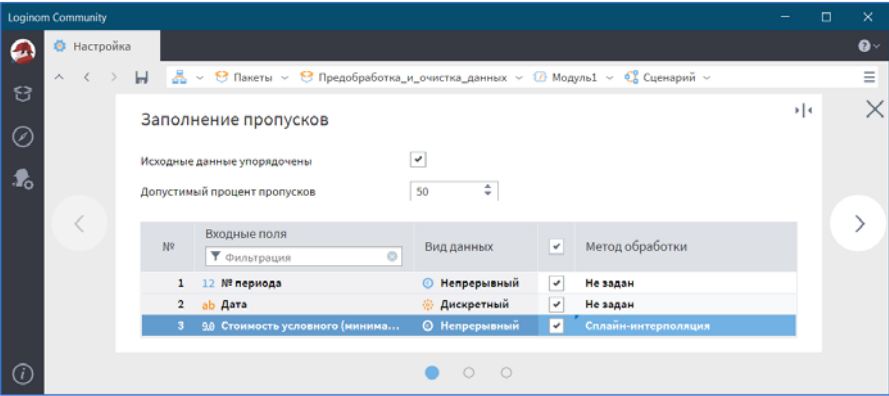


Рис. 2.40

На шаге 2 оставим стандартные параметры по умолчанию (рис. 2.41).

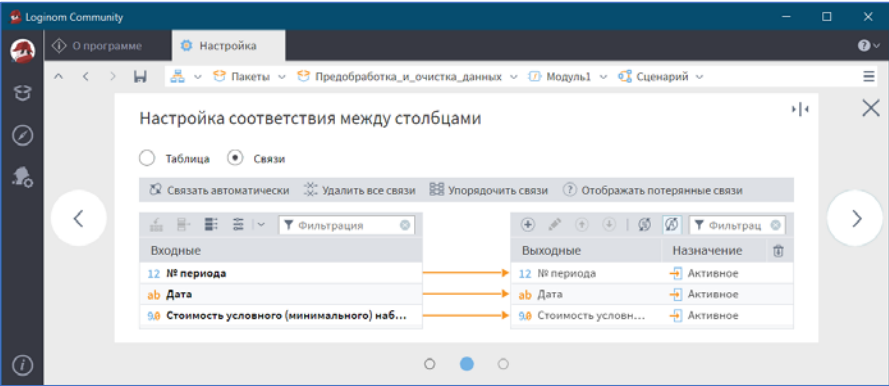


Рис. 2.41

Добавим визуализаторы *Таблица* и *Статистика* к узлу сценария. Для этого кликнем мышкой по узлу и перейдем к настройкам визуализаторов.

В визуализаторах будут представлены результаты обработки пропусков (рис. 2.42–2.43).

#	№ периода	Дата	Стоимость условного...
1	1	2016, янв.	4308,13
2	2	2016, фев.	4349,98
3	3	2016, мар.	4389,87
4	4	2016, апр.	4430,48
5	5	2016, май	4558,09
6	6	2016, июн.	4631,66
7	7	2016, июл.	4630,85
8	8	2016, авг.	4514,46
9	9	2016, сен.	4409,94
10	10	2016, окт.	4399,98
11	11	2016, ноя.	4424,57
12	12	2016, дек.	4465,70
55	13	2017, янв.	4505,39

Рис. 2.42

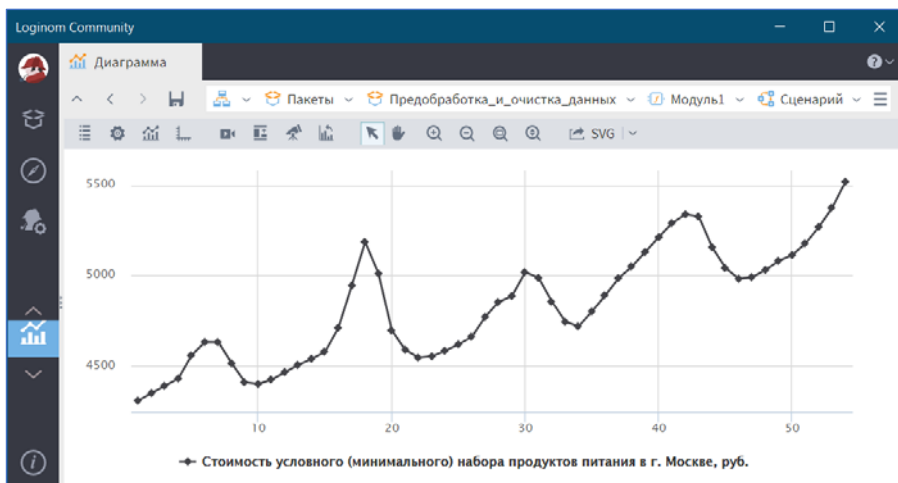


Рис. 2.43

2.3. Задания для самостоятельной работы

Задание 2.1

В файле *Задача 2.3.xlsx* имеется набор данных активных абонентов, регулярно пользующихся услугами сотовой связи в течение последних нескольких

месяцев: возраст, средняя продолжительность разговоров, продолжительность звонков днем, вечером и ночью за месяц, звонки в другие города и другие страны, доля звонков на стационарные телефоны, количество SMS за месяц (рис. 2.44).

	A	B	C	D	E	F	G	H	I	J
1	Код	Возраст, лет	Средняя продолжительность разговоров, мин.	Звонков днем за месяц	Звонков вечером за месяц	Звонков ночью за месяц	Звонки в другие города	Звонки в другие страны	Доля звонков на стационарные телефоны, %	Количество SMS за месяц
2	0	20	4	54	107	0	0	0	5	108
3	1	44	4	50	72	2	42	0	5	31
4	2	19		4	65	26	0	0	7	57
5	3	57	4	47	32	5	6	0	4	0
3565	3563	29	4	57	33	3	41	0	11	20
3566	3564	33	4	59	75	2	1	0	8	24

Рис. 2.44

Требуется произвести комплексную оценку качества набора данных на основе количества обнаруженных пропусков и выбросов, а также определить методы их обработки.

Задание 2.2

В файле *Задача 2.4. Цена на куриные яйца.xlsx* имеются данные о ежемесячной потребительской цене на куриные яйца в г. Санкт-Петербурге за 5 лет (рис. 2.45).

	A	B	C
1	№ периода	Дата	Средние потребительские цены на яйца куриные в г. Санкт-Петербурге, руб. за 10 шт.
2	1	2016, янв.	69,08
3	2	2016, фев.	67,97
4	3	2016, мар.	
5	4	2016, апр.	64,68
54	53	2020, май	66,36
55	54	2020, июн.	60,88

Рис. 2.45

Для выполнения декомпозиции данного показателя на составляющие (тренд, сезонность, остаток) необходимо заполнить пропуски, которые визуально видны в таблице.