

Разведочный анализ данных о продажах канцелярских товаров с помощью аналитической программы Loginom

Екимова Яна Сергеевна

Приамурский государственный университет имени Шолом-Алейхема

Студент

Аннотация

Целью исследования является проведение разведочного анализа данных о продажах канцелярских товаров с помощью аналитической платформы Loginom. В результате, которого получили фильтрацию по последнему году (2009); группировки продаж по брендам; сумму по всем продажам и процент по каждому бренду.

Ключевые слова: Loginom, узел, компонент, сценарий.

Exploratory analysis of data on sales of office supplies using the Loginom analytical program

Ekimova Yana Sergeevna

Sholom Aleichem Priamurskiy State University

Student

Abstract

The purpose of the study is to conduct an exploratory analysis of data on sales of office supplies using the Loginom analytical platform. As a result, we received filtering by the last year (2009); grouping of sales by brand; the sum of all sales and the percentage for each brand.

Key words: Loginom, node, component, script.

1. Введение

1.1. Актуальность

Актуальность данной темы заключается в том, что Loginom позволяет распределить информацию, так как нам нужно и сделать ее более наглядной и понятной. Loginom доступная и удобная программа с разными компонентами, которые имеют множество функций.

1.2. Обзор исследований

Н.А. Груданов проводил анализ лидеров рынка аналитических платформ, которые используются для машинного обучения [1]. А.А. Колесников и С. В. Пальмов описывали основные возможности Loginom [2]. Г.А. Поллак, И.А. Прохорова рассматривают теоретические основы алгоритмов DataMining и их применение для решения задач анализа [3]. Е.В. Морарь, П.А. Рыжковский показали возможности аналитической платформы

LOGINOM. Приведены примеры решения задач анализа данных для торговой компании [4]. В.Б. Яковлев рассмотрел основные методы анализа данных, используемые в аналитическом приложении Loginom [5].

1.3. Цель исследования

Цель исследования – выполнить разведочный анализ данных с помощью аналитической программы Loginom.

2. Материалы и методы

В данном исследовании используется программное обеспечение Loginom для анализа данных.

Данные для работы можно скачать по ссылке https://drive.google.com/drive/folders/1ua_fWrpHaxjFKb6XA92oBlqTIdGY3ooE?usp=share_link.

3. Результаты

Loginom – это платформа для реализации всех аналитических процессов: от интеграции и подготовки данных до моделирования, развертывания и визуализации.

1) Создадим черновик своего сценария.

Сценарий - главная составная часть Модуля и представляет собой последовательность шагов по обработке данных. Шаги задаются узлами из стандартных или производных компонентов.

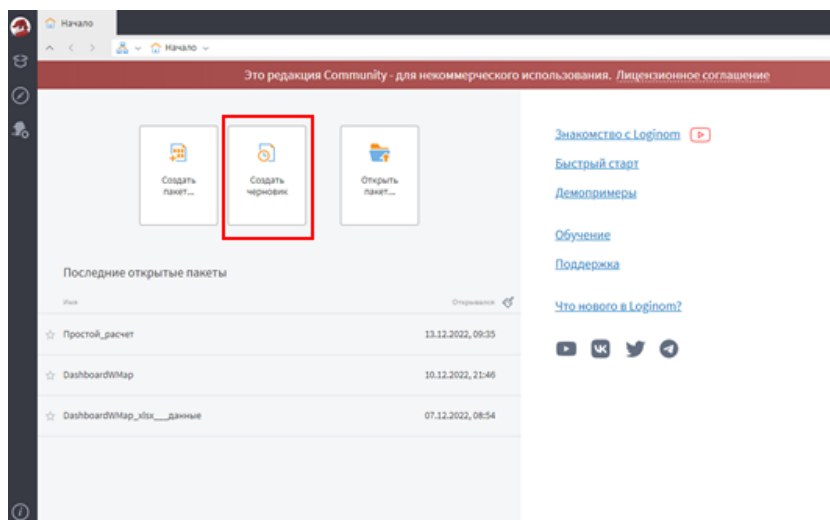


Рис.1. Создание сценария

2) Для того чтобы перейти к подключению, нужно в адресной строке нажать на пакеты и выбирать подключение.

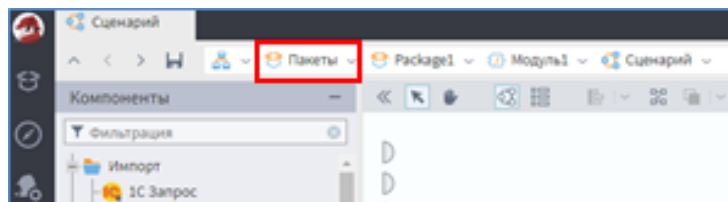


Рис.2-Подключение

Для того чтобы вернуться на сценарий необходимо нажать на кнопку вверх, или перейти в пакеты и кликнуть на слово сценарий.

3) Начнем с самого простого и загрузим данные из файла использовать нужно специальный формат файла LoginomData файл.

LoginomDataFile (LGD-файл) — это собственный формат файла данных, специально разработанный для использования с Loginom. Применение этого формата обеспечивает максимальную скорость при выполнении импорта и/или экспорта данных.

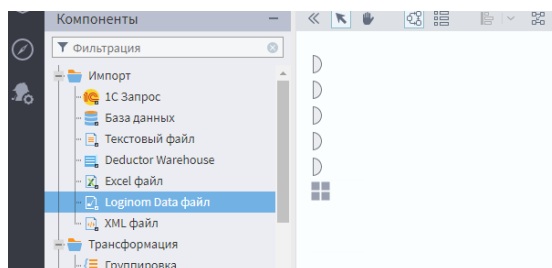


Рис.3-LoginomData файл

4) Из панели компонентов перетаскиваем мышью LoginomData файл на рабочее пространство.

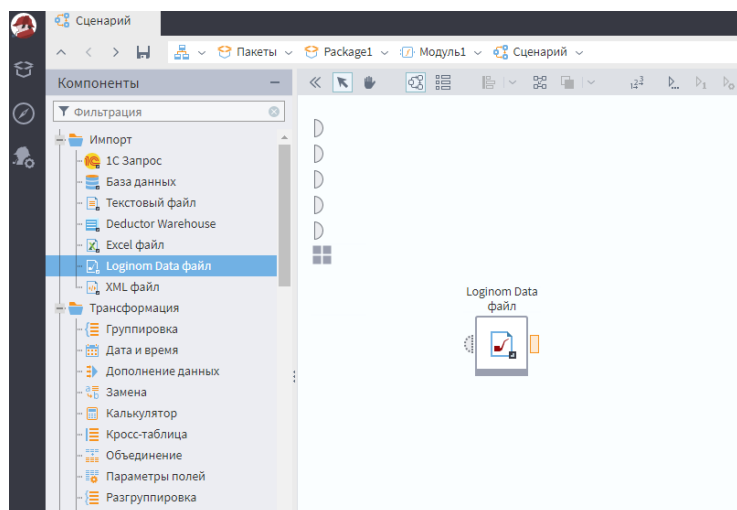


Рис.4- Рабочее поле

5) Для того, чтобы настроить этот узел, кликаем на значке и нажимаем на шестерёнку.

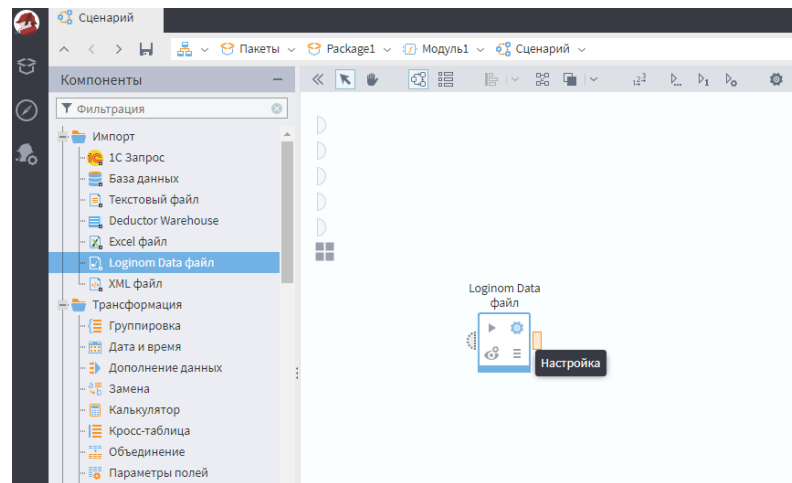


Рис.5- Настройка узла

б) Нажимаем один раз и переходим в настройки.

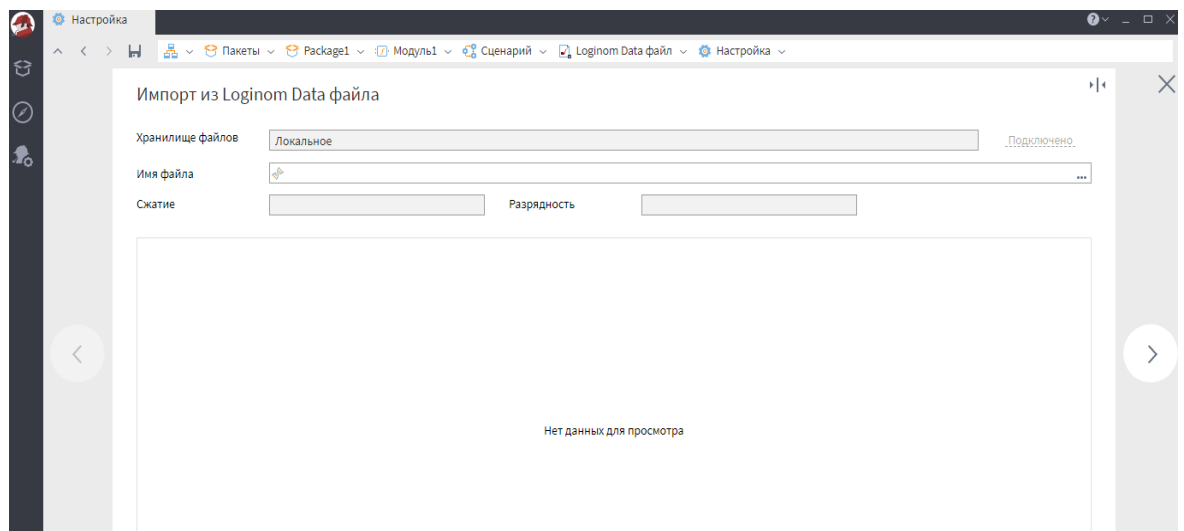


Рис.6-Настройка узла

Для того чтобы поработать с данными нужно их загрузить по ссылке, которая указана выше.

Загружаем файл “Товары_для_творчества_ТЗ_1”, нажимаем далее.

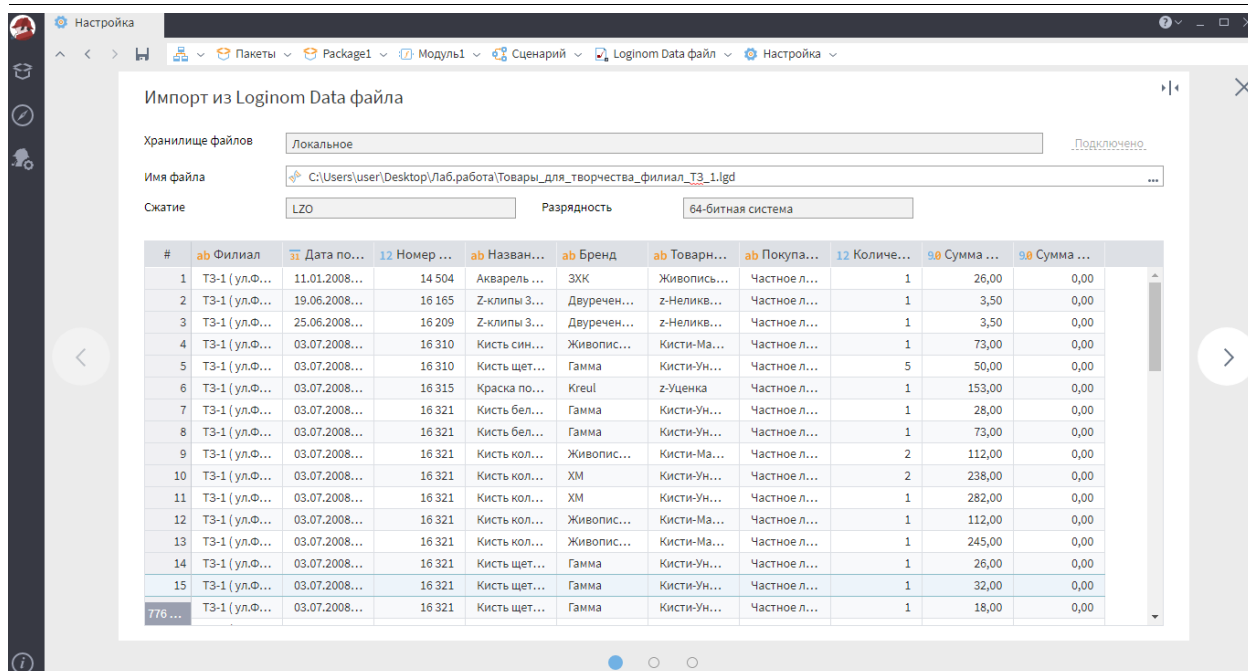


Рис.7-Загрузка файла

Чтобы было удобнее работать с файлом, назовем его “Данные по филиалу Т3-1” (рис.8.).

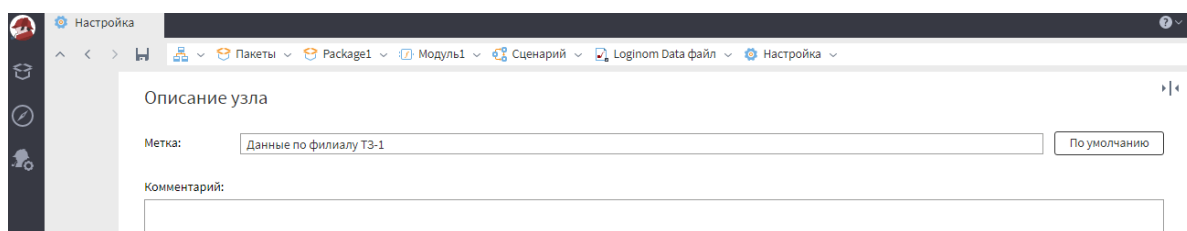


Рис.8. Метка узла

Далее запускаем узел, для этого нажимаем нижнюю кнопку выполнить. Обратите внимание, что узел стал зеленым, это означает, что данные уже в Loginom.

7) Посмотрим, какие данные попали для исследования, для этого нажимаем один раз на узел (рис.9) и переходим на настройку визуализаторов.

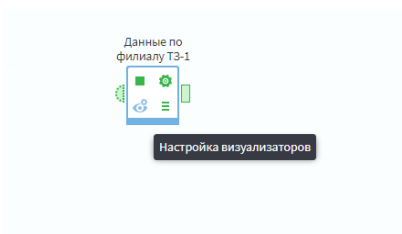


Рис.9- Настройка визуализатора

8) Нажимаем один раз и попадаем на визуализаторы. С помощью мыши перетаскиваем Статистику на рабочее поле. Далее нажимаем внутрь на «Войти».

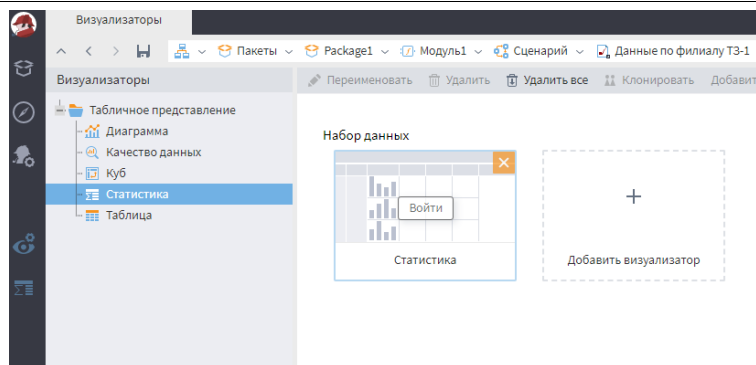


Рис.10- Настройка визуализатора

№	Метка	Вид	Гистограмма	Диаграмма размаха	Минимум	Максимум	Среднее	Стандарт...	Пропуски	Уникаль...
1	ab Филиал		Не вычислено	Недоступно						
2	31 Дата покупки			Не вычислено	11.01.2008	31.12.2009	20.04.2009...	158,42929...	0	
3	12 Номер чека			Не вычислено	14504	186822	101213	49722,352...	820	
4	ab Название товара		Не вычислено	Недоступно						
5	ab Бренд		Не вычислено	Недоступно						
6	ab Товарная группа		Не вычислено	Недоступно						
7	ab Покупатель		Не вычислено	Недоступно						
8	12 Количество			Не вычислено	1	309	1	2,4514127...	87	
9	9.0 Сумма покупки			Не вычислено	0	65380	149,71506...	354,05406...	0	
10	9.0 Сумма скидки			Не вычислено	-4,547473...	2343	2,1686994...	13,160016...	0	

Рис.11-Просмотр данных

После визуального анализа данных возвращаемся к сценарию.

Основные задачи анализа:

- 1) Провести фильтрацию по последнему году (2009);
- 2) Сгруппировать продажи по брендам;
- 3) В калькуляторе выполним расчет:
 - Сумма по всем продажам;
 - Процент по каждому бренду.

Это позволит узнать, какой бренд является самым лучшим, на какой лучше стоит обратить внимание или же какой бренд лучше убрать из продажи.

9) С помощью мыши перетаскиваем «Фильтр строк» на рабочее поле. Соединяем выходной и входной порты двух узлов.

Обработчик «Фильтр строк» позволяет выделить записи, которые удовлетворяют одному или нескольким условиям.

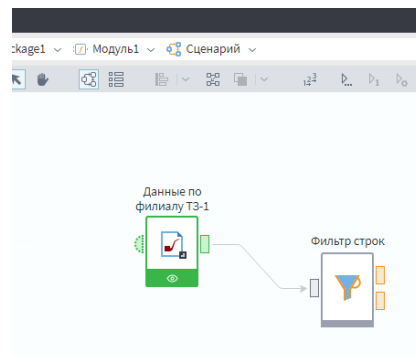


Рис.12-Соединение портов

10) Перейдем в настройку узла Фильтр строк. Для того, чтобы было удобнее дальше работать, лучше активировать состояние входа. Это позволит увидеть какие есть поля на входе. Нас сейчас интересует поле Дата покупки.

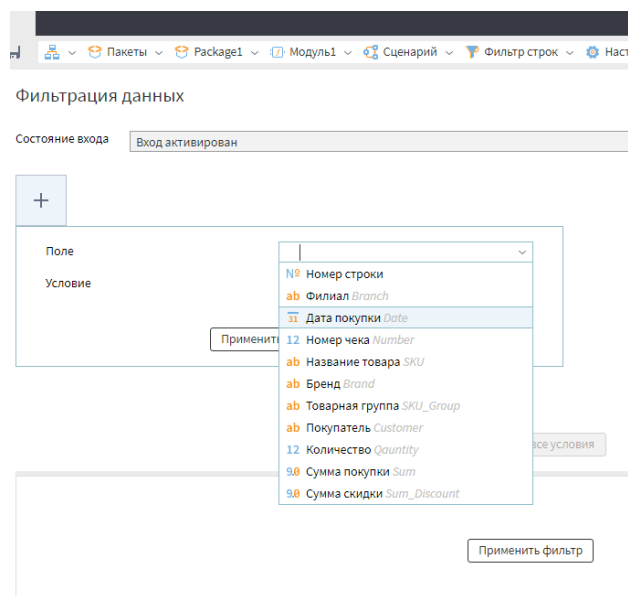


Рис.13-Настройка фильтр строк

11) Условие будет 2009 год, для этого поставим дату 01.01.2009

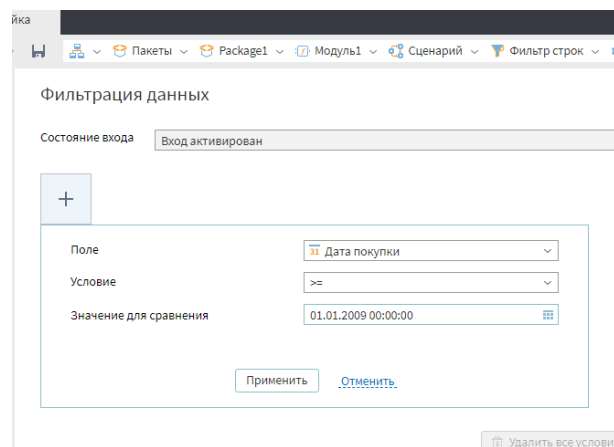


Рис.14- Создание условия

После того, как настроили условие нажимаем Применить. Снова нажимаем далее и нажимаем выполнить. Переименовать узел можно два раза щёлкнув на название узла. Переименуем его в Выбор 2009 года.

У фильтра строк два выхода, это:

- данные которые соответствуют выбранному условию;
- нижний порт, это данные которые не соответствуют выбранному условию.

Сейчас будем работать с теми данными которые будут соответствовать выбранному условию. Эти данные можно быстро посмотреть, нажав на верхний порт выхода и нажать на быстрый просмотр.

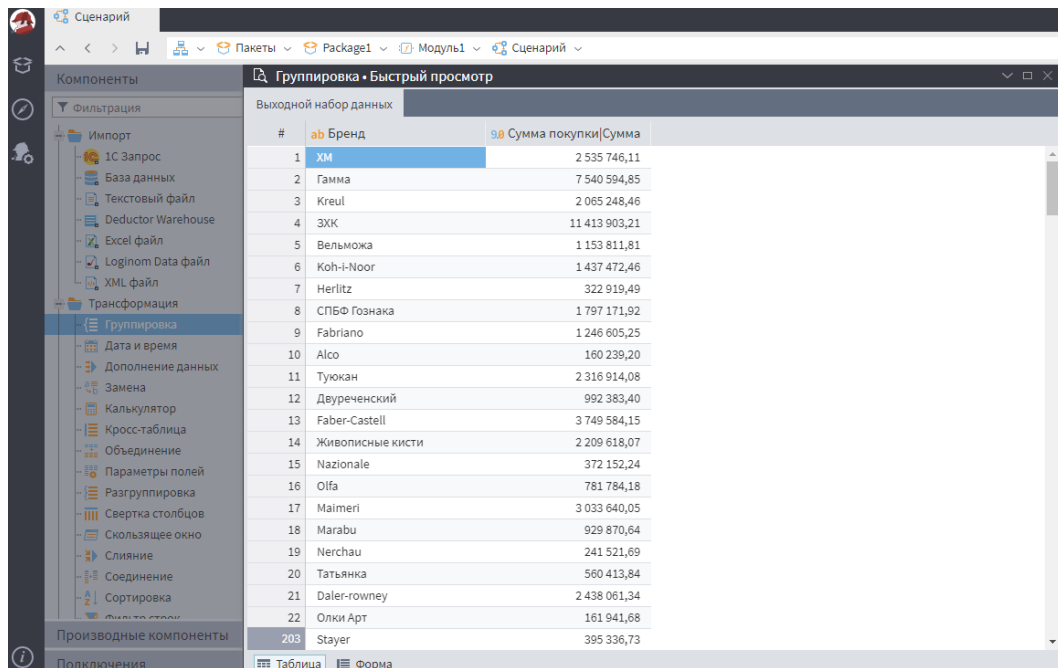
#	ab Филиал	Дата покупки	12 Номер чека	ab Название товара
1	T3-1 (ул.Фадеева)	03.01.2009, 00:00	67 336	Кисть колонок круглая d=1,5мм L=9мм серия 1105-1 №02
2	T3-1 (ул.Фадеева)	03.01.2009, 00:00	67 336	Кисть колонок круглая d=3,90мм L=16мм серия 431 №04
3	T3-1 (ул.Фадеева)	03.01.2009, 00:00	67 336	Контур по стеклу "Glass Design", флакон 80мл, черный (цв.№42774)
4	T3-1 (ул.Фадеева)	03.01.2009, 00:00	67 336	Краска по стеклу "Витраж" цв.№010 бесцветный флакон 25мл
5	T3-1 (ул.Фадеева)	03.01.2009, 00:00	67 336	Краска по стеклу "Витраж" цв.№221 малиновый флакон 25мл
6	T3-1 (ул.Фадеева)	03.01.2009, 00:00	67 336	Краска по стеклу "Витраж" цв.№419 синий флакон 25мл
7	T3-1 (ул.Фадеева)	03.01.2009, 00:00	67 336	Краска по стеклу "Витраж" цв.№504 изумрудный зеленый флако...
8	T3-1 (ул.Фадеева)	03.01.2009, 00:00	67 336	Краска по стеклу "Витраж" цв.№705 черный флакон 25мл
9	T3-1 (ул.Фадеева)	03.01.2009, 00:00	67 337	Гуашь "Мастер-класс", банка 100мл сер.2 белила цинковые (цв...
10	T3-1 (ул.Фадеева)	03.01.2009, 00:00	67 337	Гуашь "Мастер-класс", банка 100мл сер.3 стронциановый желты...
11	T3-1 (ул.Фадеева)	03.01.2009, 00:00	67 337	Папка на молнии А0, ткань, черный
12	T3-1 (ул.Фадеева)	03.01.2009, 00:00	67 339	Набор цветных карандашей, 12цв., метал.упаковка, "Sunpearl"
13	T3-1 (ул.Фадеева)	03.01.2009, 00:00	67 340	Набор бумаги поделочной 200x200мм, 100л, цветная
14	T3-1 (ул.Фадеева)	03.01.2009, 00:00	67 340	Планшет д/акв цел.100% 200г/кв.м 210x297мм 20л Розы
15	T3-1 (ул.Фадеева)	03.01.2009, 00:00	67 341	Набор салфеток д/декупажа 33x33см морские мотивы 6 мотивов...
16	T3-1 (ул.Фадеева)	03.01.2009, 00:00	67 342	Бумага пастельная 160г/кв.м, 700x1000мм "Tiziano" бледно-голу...
17	T3-1 (ул.Фадеева)	03.01.2009, 00:00	67 342	Гуашь "Мастер-класс", банка 100мл сер.2 белила цинковые (цв...
18	T3-1 (ул.Фадеева)	03.01.2009, 00:00	67 342	Гуашь "Сонет", банка 100мл сер.1 флуоресцентный, голубой, (цв...
19	T3-1 (ул.Фадеева)	03.01.2009, 00:00	67 343	Бумага пастельная 160г/кв.м, 700x1000мм "Tiziano" малиновый (...)
20	T3-1 (ул.Фадеева)	03.01.2009, 00:00	67 343	Кнопки канц. "Gagant", d:9.5мм, в карт.коробке, 100шт., золото
21	T3-1 (ул.Фадеева)	03.01.2009, 00:00	67 344	Холст на подрамнике лен100%, мелкозернистый 50x60см Б
22	T3-1 (ул.Фадеева)	03.01.2009, 00:00	67 344	Холст на подрамнике лен100%, мелкозернистый 50x80см

Рис.15- Просмотр данных

12) Теперь требуется сгруппировать товары по бренду, для этого используем компонент Группировка. Также переносим мышью данный компонент на рабочее поле и подаем данные соответствующие условию. Соединяем верхний порт узла Выбор 2009 года с входным портом узла Группировка.

Группировка выполняет действия над набором данных. Обработчик позволяет объединять записи избранных полей в группы, а для оставшихся полей вычислять статистические показатели (сумму, среднее, минимум и т. д.).

13) Снова настраиваем узел, но теперь уже группировки, нажав на сам узел и выбрав шестерёнку. В качестве группы выбираем поле бренд, перетащив поле бренд в правый столбик группы и используем сумму покупки как показатель. После этих действий нажимаем далее и получаем практически мгновенно расчет по каждому бренду.



#	ab Бренд	9.0 Сумма покупки Сумма
1	XM	2 535 746,11
2	Гамма	7 540 594,85
3	Kreul	2 065 248,46
4	ЗХК	11 413 903,21
5	Вельможа	1 153 811,81
6	Koh-i-Noor	1 437 472,46
7	Herlitz	322 919,49
8	СПБФ Гознака	1 797 171,92
9	Fabiano	1 246 605,25
10	Alco	160 239,20
11	Туюкан	2 316 914,08
12	Двуреченский	992 383,40
13	Faber-Castell	3 749 584,15
14	Живописные кисти	2 209 618,07
15	Nazionale	372 152,24
16	Olfa	781 784,18
17	Maimeri	3 033 640,05
18	Marabu	929 870,64
19	Nerchau	241 521,69
20	Татьянка	560 413,84
21	Daler-rowley	2 438 061,34
22	Олки Арт	161 941,68
203	Stayer	395 336,73

Рис.16-Расчет по каждому бренду

14) Теперь нужно посчитать процент по каждому бренду, это будем делать с помощью узла калькулятор. Также перетаскиваем данный компонент на рабочее поле. Соединяем узлы группировки и калькулятора с помощью выходного и входного порта. И снова начинаем настраивать.

15) Первое, что должны сделать, это посчитать общую сумму. Для этого нужно задать выражение Total, вместо выражения, напишем Общая сумма по всему филиалу.

Для того, чтобы посчитать общую сумму, воспользуемся одной из статистических функций. Нажимаем 2 раза на Stat, далее 2 раза щелкаем на Sum покупки и в типе вручную напишем Sum.

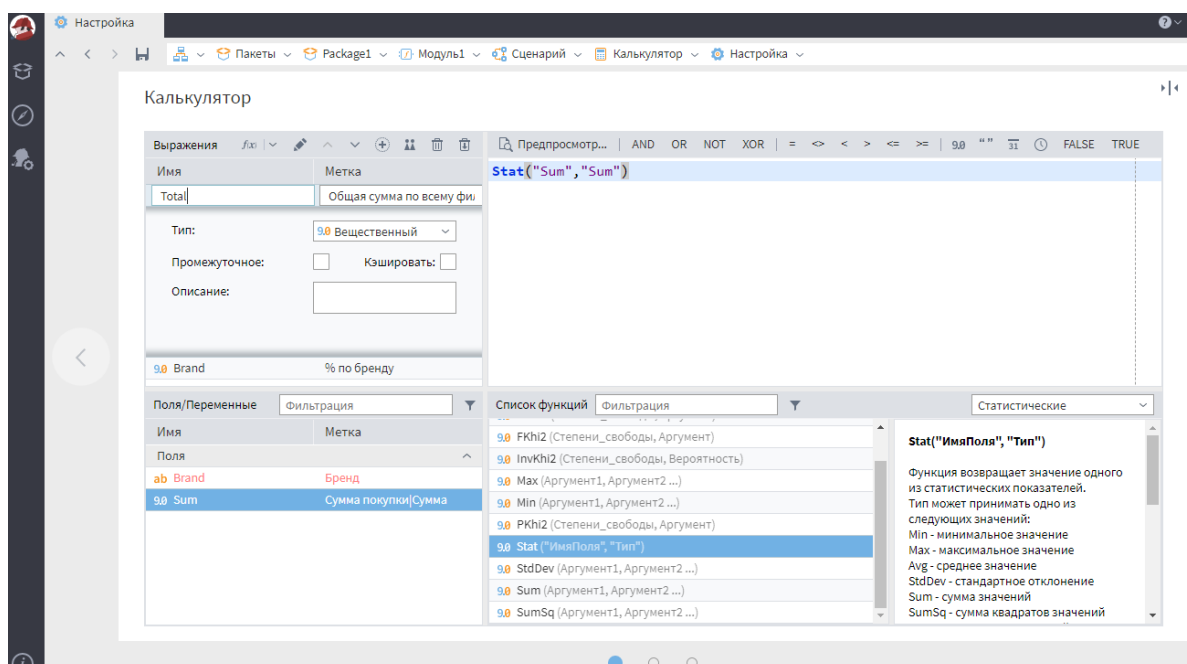


Рис.17-Настраиваем статистическую функцию

16) Далее добавим еще одно поле, нажав на плюсик

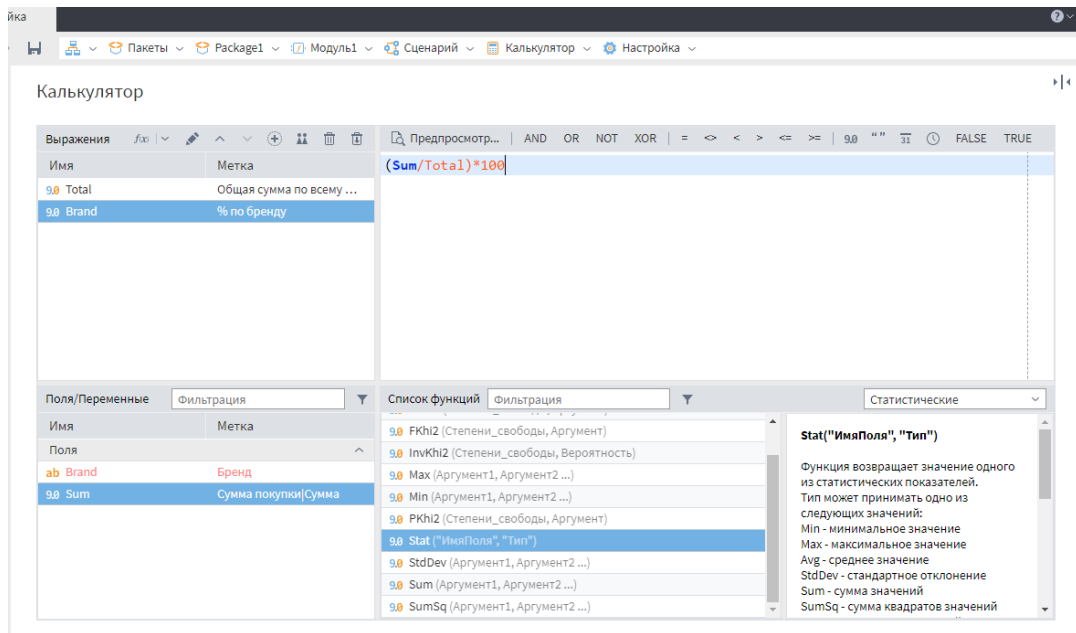


Рис.18- Добавление поля

Теперь нужно переименовать узел.

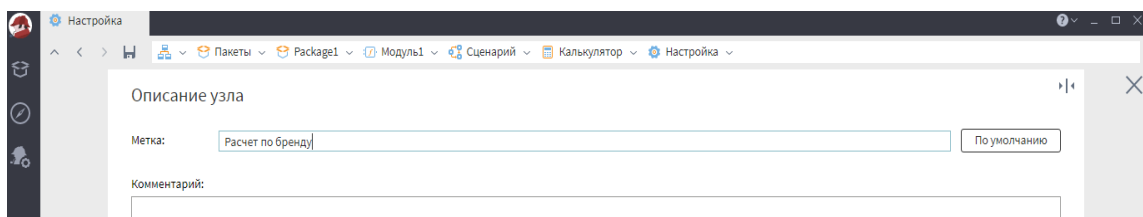


Рис.19-Переименовываем узел

После того, как переименовали узел, нажимаем выполнить.

17) Далее настраиваем визуализатор, щелкнув один раз по узлу Расчет по бренду.

18) Выбираем визуализатор таблица и заходим в него.нас не сильно интересует общая сумма по всему филиалу, поэтому этот столбец уберем, убрав галочку.

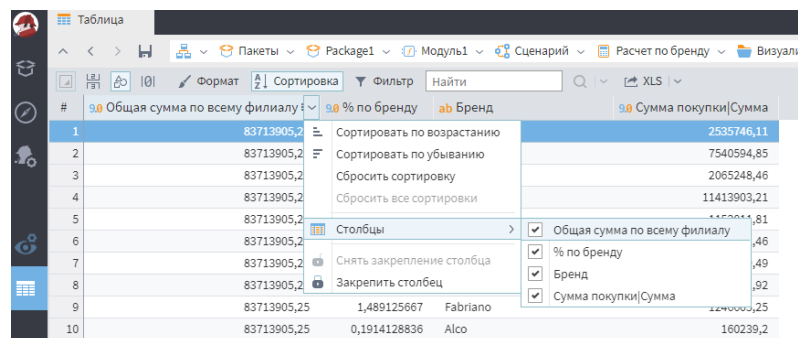
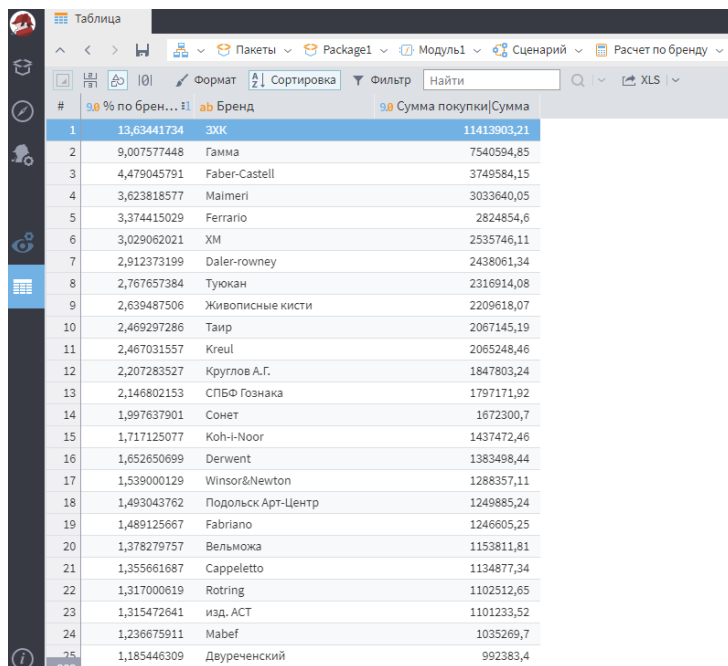


Рис.20-Настройка столбца

И чтобы было удобнее работать отсортируем по убыванию.



#	9,0 % по брен...	ab Бренд	9,0 Сумма покупки Сумма
1	13,63441734	ЗХК	11413903,21
2	9,007577448	Гамма	7540594,85
3	4,479045791	Faber-Castell	3749584,15
4	3,623818577	Maimeri	3033640,05
5	3,374415029	Ferrario	2824854,6
6	3,029062021	ХМ	2535746,11
7	2,912373199	Daler-rowney	2438061,34
8	2,767657384	Туюкан	2316914,08
9	2,639487506	Живописные кисти	2209618,07
10	2,469297286	Таир	2067145,19
11	2,467031557	Kreul	2065248,46
12	2,207283527	Круглов А.Г.	1847803,24
13	2,146802153	СПБФ Гознака	1797171,92
14	1,997637901	Сонет	1672300,7
15	1,717125077	Koh-i-Noor	1437472,46
16	1,652650699	Derwent	1383498,44
17	1,539000129	Winsor&Newton	1288357,11
18	1,493043762	Подольск Арт-Центр	1249885,24
19	1,489125667	Fabriano	1246605,25
20	1,378279757	Вельможа	1153811,81
21	1,355661687	Cappeletto	1134877,34
22	1,317000619	Rotring	1102512,65
23	1,315472641	изд. АСТ	1101233,52
24	1,236675911	Mabef	1035269,7
25	1,185446309	Деуреченский	992383,4

Рис.21-Отсортировка по убыванию

19) Теперь вернемся к сценарию и сохраним наш проект, нажав в верхнем углу на значок сохранить.

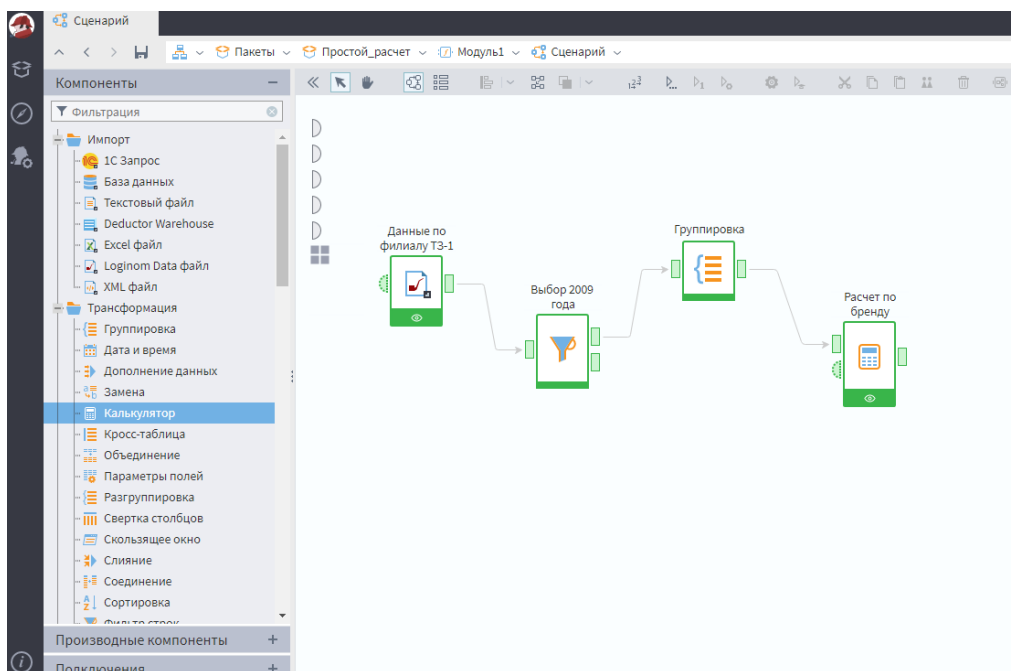


Рис.22-Сценарий

20) Давайте сейчас добавим новые данные. Для этого так же разместим на рабочем пространстве и настроим компонент LoginomData файл. И возьмем второй фотосетвсе_филиалы_кроме_ТЗ_1.

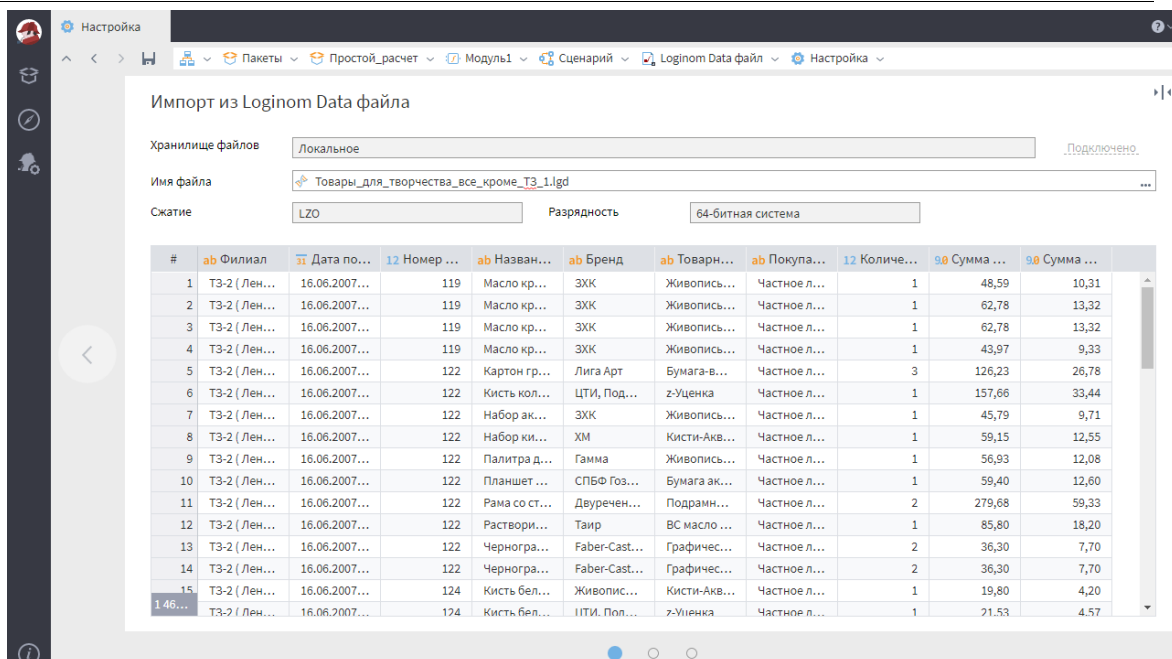


Рис.23-Настройка второго компонента LoginomData

Снова нажимаем далее и в описании узла задаем название остальные филиалы и нажимаем выполнить.

21) Теперь нужно объединить эти данные, для этого размещаем компонент объединение на рабочее поле. Далее подсоединяем порт Данные по филиалу Т31 к верхнему порту входа узла Объединение и узел Остальные филиалы к нижнему порту входа узла Объединение. И снова проведем тройку узла Объединение.

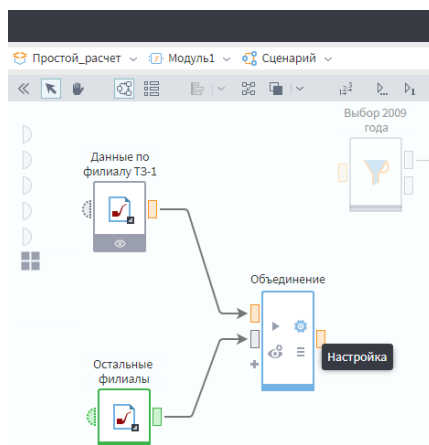


Рис.24-Подсоединение узлов

22) Далее сопоставим все данные по двум дата сетам.

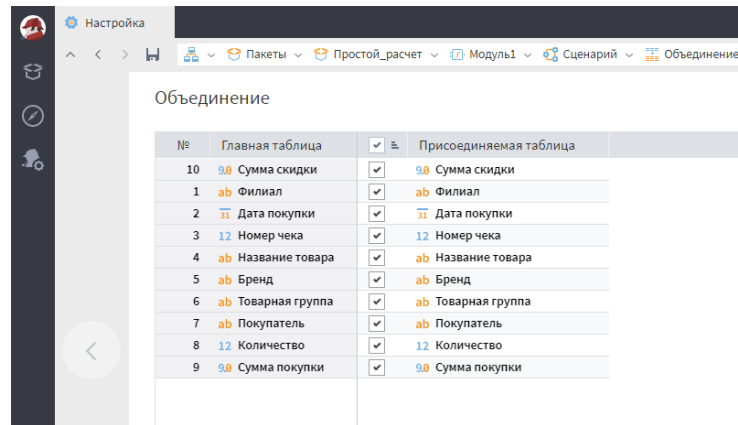


Рис.25-Сопоставка данных

Нажимаем далее-выполнить.

23) Теперь полученные данные подаем на вход созданного и сценария.

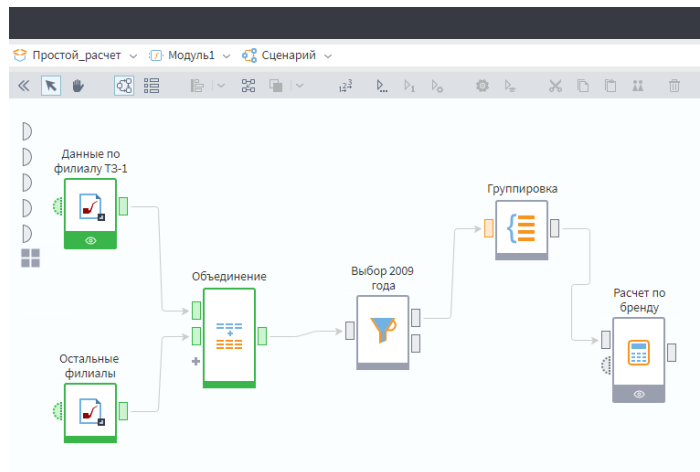


Рис.26-Подсоединение

24) Теперь можем запустить весь сценарий, нажав на кнопку выполнить все, либо нажать на наш последний узел один раз и далее нажать на кнопку выполнить

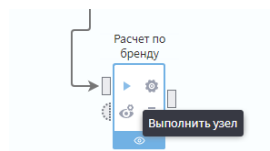


Рис.27-Узел

В итоге получается вот такой готовый сценарий.

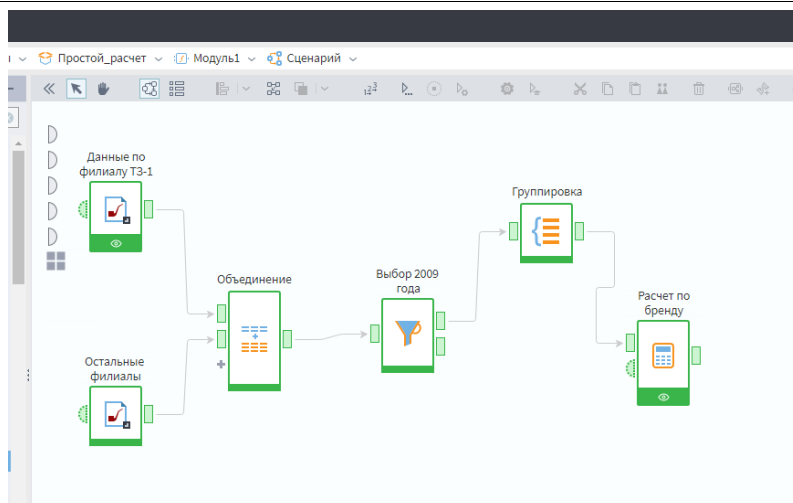


Рис.28- Сценарий

Если закрыли программу и при открытии стали неактивны компоненты (серые), то необходимо нажать на запуск и компоненты (узлы) станут активными.

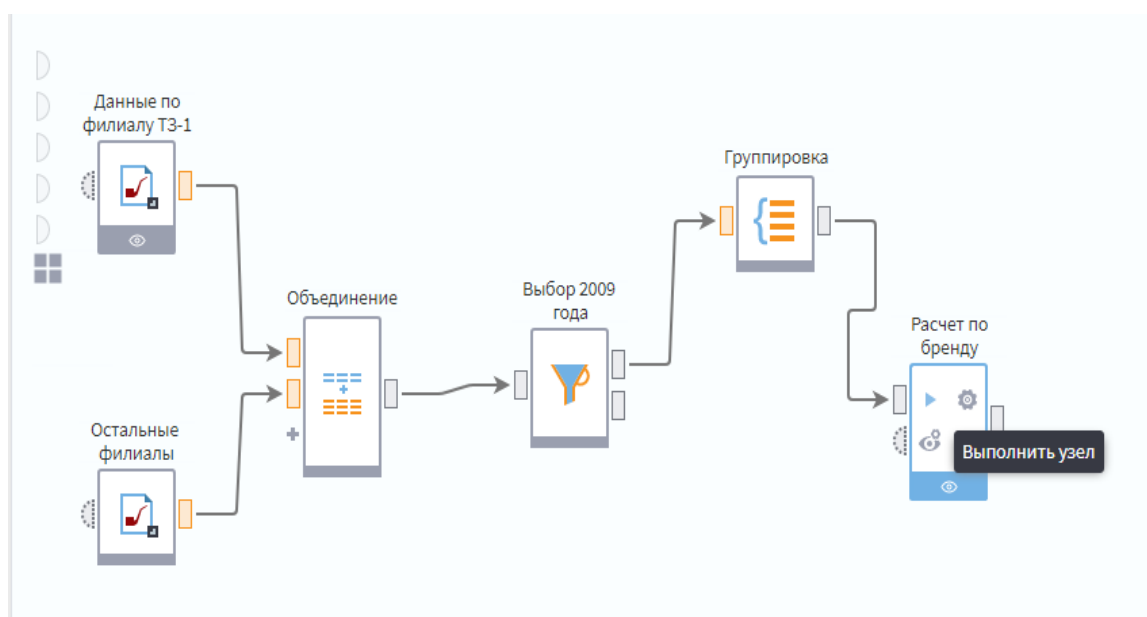


Рис.29- Запуск узлов

4. Выводы

В результате получили разведочный анализ данных о продажах канцелярских товаров с помощью аналитической платформы Loginom. В итоге была сделана: фильтрация по последнему году (2009); группировка продаж по брендам; сумма по всем продажам и процент по каждому бренду.

Библиографический список

1. Груданов Н. А. Обзор инструментальных средств для анализа данных // Physics and mathematics. 2020. №16. С. 35.
2. Колесников А. А., Пальмов С. В. Loginom: основные возможности // Форум

- молодых ученых. 2018. №. 10. С. 582-587.
3. Поллак Г.А., Прохорова И.А. Практический курс анализа данных на платформе Loginom. Саратов: Ай Пи Ар Медиа, 2022. 208 с.
 4. Морарь Е. В., Рыжковский П. А. О возможности применения аналитической платформы loginom для анализа данных торговой компании //информационный бюллетень Омского научно-образовательного центра ОмГТУ и ИМ СО РАН в области математики и информатики. 2020. С. 71-73.
 5. Яковлев В. Б. Анализ данных в аналитической платформе Loginom. Saarbrücken: Lap Lambert, 2020. 184 с.
 6. Евсюков В. В., Капустин А. В., Ильина Ю. А. Аналитическая платформа loginom-универсальный инструмент углубленной аналитики //Вестник Тульского филиала Финуниверситета. 2020. №. 1. С. 291-292.
 7. Королева Н. А. Платформа Loginom как революционный инструмент бизнес-аналитики //ИТ-технологии и корпоративные информационные системы в оптимизации бизнес-процессов. 2018. С. 19-22.
 8. Прокопенко Н.Ю., Тришин Д.В. Применение loginom для оптимизации процессов управления товарными запасами предприятий малого и среднего бизнеса //Межвузовский сборник статей лауреатов конкурсов. 2021. С. 216-222.