Лабораторная работа № 1

Создание первого сценария в Loginom

На рис. 1 показаны основные блоки интерфейса Loginom:

- 1. **Главное меню** позволяет пользователю начать/завершить работу, получить доступ к пакетам и настройкам платформы и др. Состав элементов *Главного меню* может изменяться в зависимости от редакции платформы (Server/Desktop), прав пользователя, а также при работе с визуализаторами.
 - 2. Адресная строка строка, содержащая путь к открытому объекту.
- 3. **Рабочее пространство** область, в которой осуществляются основные действия по построению сценария, настройке подключений, отчетов, администрированию платформы и т.п. Состав элементов, доступных команд и визуальное отображение этой области зависят от того, какая страница платформы находится в активном состоянии. На рис. 1 показано рабочее пространство страницы «Сценарий».
- 4. **Панель** «**Процессы**» панель, содержащая подробную информацию о процессах обработки данных узлами Сценария.

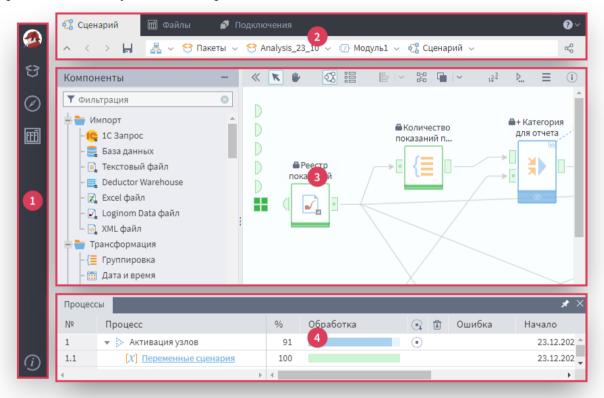


Рисунок 1. Основные блоки интерфейса

Практическое задание

1 Загрузка данных в Loginom

Загрузите данные из файла специального формата Loginom Data файл (*.lgd) из панели **Компоненты** выберите Loginom Data файл и мышью перетащите в **Рабочее**

пространство, получаем узел вида: . Далее этот узел необходимо настроить.

файл

файл

Наведите мышкой на узел и увидите несколько иконок, которые обозначают Loginom Data

действия с файлом . Выберите **Настройка** и в окне **Импорт из Loginom Data** файла в строке **Имя файла** выберите файл *Товары_для_творчества_филиал_T3_1.lgd* (рис. 2).

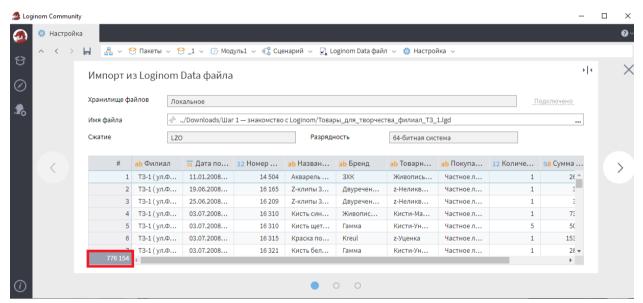


Рисунок 2. Импорт из Loginom Data файла

Loginom распознал столбцы и строки данных, которых в этом наборе данных 776 тысяч. Нажмите 2 раза **Далее** и в мастере настройки **Описание узла** задайте узлу метку *Данные по филиалу Т3-1*. Такие метки делаются для того чтобы в дальнейшем было легче разобраться в построенном сценарии.

Далее запустите узел нажав кнопку Выполнить. Все данные считались - узел стал

зеленым и его название изменилось . Иногда бывает очень полезно просмотреть

Данные по филиалу Т3-1 какие данные попали для исследования, для этого сделаем **Визуализацию данных**.

Наведите курсор мыши на узел и выберите Настройка визуализатора

филиалу ТЗ-1

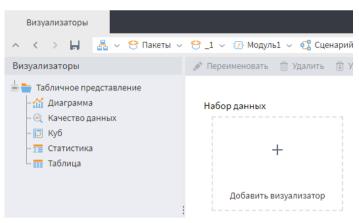


Рисунок 3. Настройка визуализатора

Визуализаторы можно использовать разные (Диаграмма, Куб, Таблицы и др.) (рис. 3). Выберем визуализатор Статистика, для этого с помощью мыши перетаскиваем на поле Набор Данных (рис. 3). Зайдите внутрь визуализатора, щёлкнув на



Посмотрите, как выглядят данные после загрузки (рис. 4). Если часть данных не отобразилась щелкните на кнопке Уникальные значения и Порядковые статистики в адресной строке. В строке метки Дата покупки можно увидеть, что данные у нас есть с января 2008 г. по конец 2009 г. В строке метки Номер чека имеются пропуски (820 пропусков). Пропуски в номерах чеков скорее всего были или из-за отмены чеков или каких-то возвратов или закрытия смены, на данном этапе обучения это не важно. Но если бы данные исследовались в действительности, необходимо было бы предварительно обработать пропуски, перед тем как анализировать (это будет рассмотрено в дальнейшем).

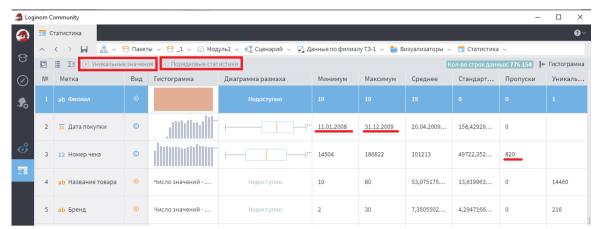


Рисунок 4. Визуализатор Статистика.

Вернуться в сценарий можно либо нажав дважды кнопку или в адресной строке выбрав

Сценарий

2 Простой расчет

В этом разделе организуем фильтрацию данных по последнем году (2009 как выяснили ранее), сгруппируем продажи по брендам, в калькуляторе выполним расчет суммы по всем продажам и процента по каждому бренду. Это позволит узнать какой бренд является самым лучшим, на какие лучше обратить внимание, а также если процент продаж по бренду невысок, то скорее всего бренд можно либо выводить из продажи либо проводить какие-то акции.

2.1 Фильтрация

Для организации фильтрации по 2009 году нужно выбрать из панели **Компоненты** раздел **Трансформация** и далее перетащить

рабочее пространство сценария и соединяем выходной и входной порт и двух узлов:



Более подробно ознакомиться с обработчиком **Фильтр строк** можно по ссылке: https://help.loginom.ru/userguide/processors/transformation/row-filter/. Перейдем в настройку узла **Фильтр строк**, для этого наводим курсор мыши на этот узел и выбираем **Настройка**. Для удобства лучше активировать **Состояние входа**, это позволит увидеть какие есть поля на входе. Далее нажмите + и в строке **Поле** откроется список доступных полей (рис. 5).

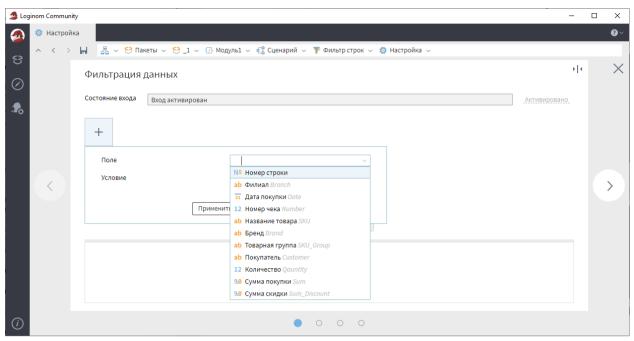
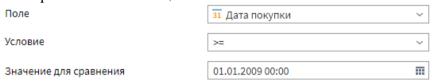


Рисунок 5. Фильтрация данных.

Из списка полей выберите поле **Дата покупки**, введите **Условие** >= и **Значение** для **сравнения** 01.01.2009, поскольку нас интересует 2009 год, а диапазон данных в этом наборе данных с января 2008 г. по конец 2009 г.:



Если вы что-то написали неправильно, ошиблись Loginom подсвечивает поле. С помощью + можно добавить еще условия, ну а пока условия для фильтрации будут выглядеть так:



700 тысяч строк практически мгновенно были отсортированы. Далее нажимаем кнопку Далее, потом кнопку Выполнить. В рабочем пространства сценария дважды щелкнуть мышкой на узле фильтрации и переименовать Фильтр строк в Выбор 2009 года. У узла Фильтр строк 2 выхода: верхний порт это те данные, который соответствует условию, нижний порт - данные не соответствуют условию. Будем работать с теми данными которые соответствуют условию, которое задали. Эти данные можно посмотреть щёлкнув мышью по верхнему порту узда Выбор 2009 года (рис. 6).

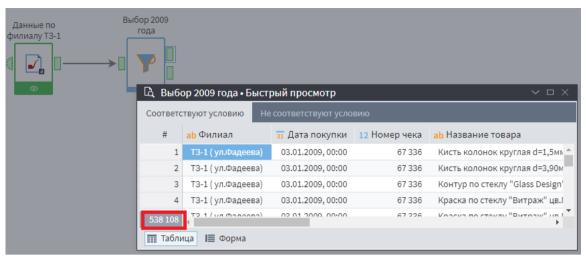


Рис. 6. Результаты фильтрации данных.

Как видно из свыше 776 тысяч строк осталось около 538 тысяч и все они соответствуют требованиям.

2.2 Группировка товаров и брендов

Обработчик **Группировка** перетащите на рабочее пространство и подайте на вход данные, соответствующие условию фильтрации, то есть соедините верхний порт узла **Выбор 2009 года** с входным портом узла **Группировка**:



Более подробно ознакомиться с обработчиком Группировка можно по ссылке: https://help.loginom.ru/userguide/processors/transformation/grouping.html. В узле Группировка выберете Настройка . Что бы узнать сколько принес каждый бренд с помощью мыши перетащите Бренд из столбца Доступные поля в правый столбец Выбранные поля в Группу, а Сумму покупки перетащите в правый столбец Выбранные поля в Показатели (используем сумму покупки как показатель по умолчанию) (рис. 7).

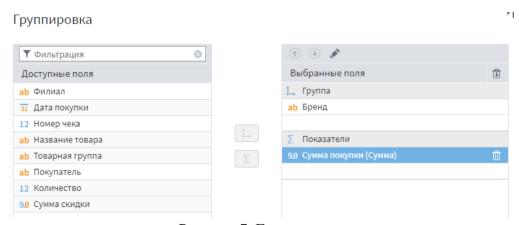


Рисунок 7. Группировка

Нажмите Далее, потом Выполнить. Два раза щелкните мышкой по выходному полю узла Группировка, для быстрого просмотра результатов (рис. 8). Получаем практически мгновенно расчет по каждому бренду. Видно сколько в деньгах принес каждый бренд, видно также, что брендов не так много (203).

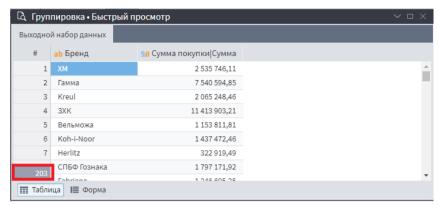


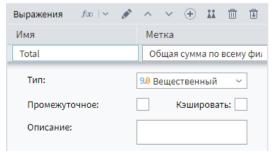
Рисунок 8. Результаты группировки

2.3 Расчет суммы по всем продажам и процента по каждому бренду

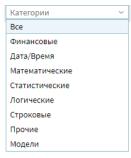
Посчитайте процент по каждому брендом с помощью узла калькулятор. На панели **Компоненты**, в разделе **Трансформация** выберите Калькулятор на рабочее пространство и соедините выходной порт узла **Группировка** с входным портом узла **Калькулятор**.



Более подробно ознакомиться с обработчиком **Калькулятор** можно по ссылке: https://help.loginom.ru/userguide/processors/transformation/calc/. В узле **Калькулятор** выберете **Настройка**. Сначала нужно посчитать общую сумму, для этого в разделе **Выражение** в поле **Имя** задать имя *Total*, а в поле Метка ввести *Общая сумма по всему филиалу*. Имя должно быть написано латиницей, метка задается для удобства на русском языке.



Что бы посчитать Общую сумму по всему филиалу воспользуемся одной из функций.



В разделе Список функций узла Калькулятор в списке Категории выберите тип функции Статистические и далее функцию Stat("ИмяПоля", "Тип"). Внизу справа можно посмотреть описание, что может данная функция. В частности функция Stat возвращает значение одного из статистических показателей. Тип может принимать одно из следующих значений:

- Min минимальное значение;
- Мах максимальное значение;
- Avg среднее значение;
- StdDev стандартное отклонение;
- Sum сумма значений;
- SumSq сумма квадратов значений;
- Count количество значений;
- UniqueCount количество уникальных значений;
- NullCount количество пропущенных значений;
- Mode мода (значение, которое встречается наиболее часто).

Сумму сделайте по полю Сумма покупки. Для этого сначала установите курсор в кавычках **ИмяПоля** и щелкните **Сумма покупки** | **Сумма** в разделе **Поля/Переменные**. Далее установите курсор в кавычках **Тип** и вручную впечатайте Sum.

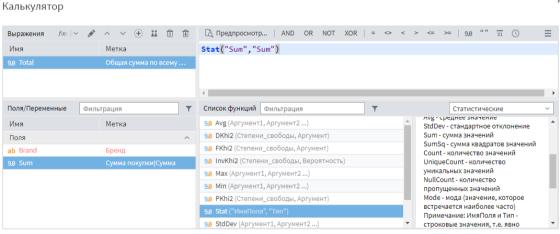


Рисунок 9. Расчет суммы по всем продажам

Добавьте еще одно выражение с помощью кнопки $\stackrel{\textcircled{\ }}{}$. Назовите выражение *Brend*, установите метку % *по бренду*. В рабочую область калькулятора введите формулу (**Sum/Total)*100** (рис. 10).

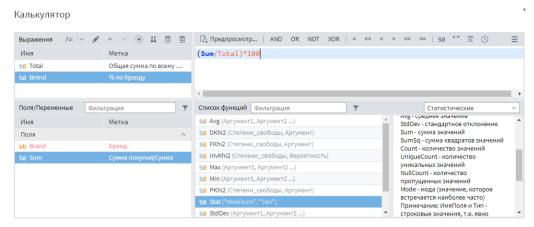
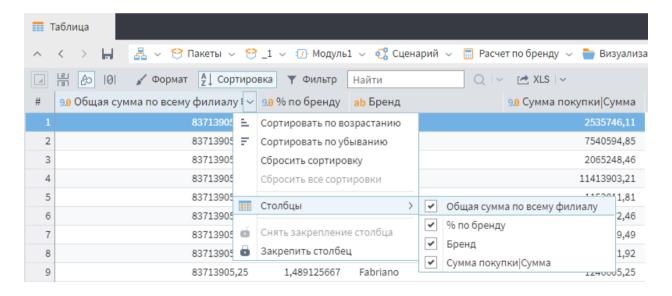


Рисунок 10. Расчет процента по каждому бренду

Нажмите Далее, измените название узла в разделе Описание узла на *Расчет по бренду*. Нажмите Выполнить. Далее посмотрит результаты расчета с помощью визуализатора. Наведите курсор мыши на узел Расчет по бренду и выберите Настройка визуализатора . На поле Входной набор данных перетащите визуализатор Таблица, откройте его.

В данной ситуации столбец Общая сумма по всему филиалу содержит одинаковые значения и не нужен, поэтому необходимо убрать этот столбец - снять галочку:



Для удобства работы отсортируйте по убыванию по полю % по бренду. Больше всего выручки принес бренд 3ХК - больше, чем 13 %, ну и соответственно дальше есть бренды Гамма., Faber-Castell вносящие 9, 4.4 и т.д. %.

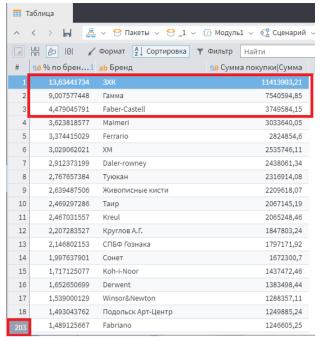


Рисунок 11. Результаты анализа (файл Простой расчет)

В самом верху таблицы находятся те бренды, на которые надо обращать особое внимание, которые приносят больше всего прибыли. На те бренды, которые находятся внизу таблицы тоже стоит обратить внимание и разобраться стоит ли их закупать и держать, чтобы предлагать клиентам.

Вернитесь в Сценарии и сохраните файл с помощью кнопки под именем Простой расчет.

3 Модификация сценария

Модифицируем разработанный сценарий. Добавим новые данные с помощью узла **Loginom Data файл,** далее выберите **Hactpoйка** и загрузите файл *Товары_для_творчества_все_кроме_Т3_1.lgd*. Этот набор данных содержит около 1.5 млн. записей (для сравнения попробуйте обработать такой набор данных, например в MS Excel, скорее всего это просто не удастся). Нажимаем **Далее** и в **Описании узда** зададим название *Остальные филиалы*, нажимаем **Выполнить**. Все практически полтора миллиона строк считались.

Объединение эти данные с помощью компонента Объединение из раздела Трансформация панели Компоненты. Более подробно ознакомиться с обработчиком Объединение можно по ссылке: https://help.loginom.ru/userguide/processors/transformation/union.html. В компоненте Объединение есть два входных порта (в принципе порты можно еще добавлять). Соедините выходной порт узла Остальные филиалы со вторым входным портом узла Объединение. Соедините выходной порт узла Данные по филиалу ТЗ-1 с первым входным портом узла Объединение одним из приведённых способов:

• разорвите связь (щёлкните правой кнопкой мышки на связи и выберите Удалить связь) между узлом Данные по филиалу ТЗ-1 и Выбор 2009 года и соедините заново указанные узлы;

• перетащите соединение (синий квадратик) с входного порта **Выбор 2009 года** на первый входной порт узла **Объединение**.

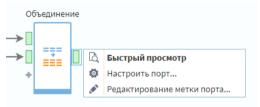
Проведем настройку узла **Объединение**. Сопоставим все данные по двум наборам данных, название узла оставим тем же (рис. 12). Нажмите **Далее**, потом **Выполнить**.

Объединение

| № Главная таблица | ица |
|---|-----|
| | |
| 1 ab Филиал ✓ ab Филиал | |
| 2 ₃ Дата покупки ✓ ₃ Дата покупки | |
| 3 12 Номер чека ✓ 12 Номер чека | |
| 4 ab Название товара ✓ ab Название товара | |
| 5 <mark>ab</mark> Бренд ✓ ab Бренд | |
| 6 ab Товарная группа 🗸 ab Товарная группа | |
| 7 ab Покупатель 🗸 ab Покупатель | |
| 8 12 Количество ✓ 12 Количество | |
| 9 9.0 Сумма покупки 🗸 9.0 Сумма покупки | |
| 10 9,0 Сумма скидки 🗸 9,0 Сумма скидки | |

Рисунок 12. Настройка узла Объединение

Нажмите правой кнопкой мыши на выходном порту узла **Объединения** и из контекстного меню выберите **Быстрый просмотр** (это еще один способ). В полученном <u>объединенном</u> наборе данных более 2 млн записей, убедитесь, что имеется информация по разным филиалам.



Соедините выходной порт узла **Объединение** с выходным портом узда **Выбор 2009** года. Итоговая схема показана на рис. 13.



Рисунок 13. Итоговый сценарий.

Запустите весь сценарий нажав кнопку **Выполнить все** или запустите последний узел для того чтобы были произведены расчеты для 2 млн записей.

С помощью визуализатора **Таблица** сравним результаты сценария файла *Простой* расчет (рис. 11) с полученными данными после объединения (рис. 14).

| €8 € | ценарий 🏢 | Таблица | |
|------|----------------|------------------------|--------------------------|
| ^ | < → 📙 🔠 | ∨ 😌 Пакеты ∨ 😌 _2 ∨ ः(| √ Модуль1 ∨ € √ Сценарий |
| | 를 🕭 101 🖌 | Формат 🛕 Сортировка 🔻 | Фильтр Найти |
| # | 9.0 % по брен1 | ab Бренд | 9.0 Сумма покупки Сумма |
| 1 | 12,87196421 | ЗХК | 26416190,33 |
| 2 | 8,764612206 | Гамма | 17986972,34 |
| 3 | 4,336149354 | Faber-Castell | 8898762,05 |
| 4 | 3,914674448 | XM | 8033799,94 |
| 5 | 3,455775141 | Maimeri | 7092034,47 |
| 6 | 3,260091719 | Ferrario | 6690447,7 |
| 7 | 2,702516093 | Туюкан | 5546176,04 |
| 8 | 2,698375786 | Живописные кисти | 5537679,19 |
| 9 | 2,50786091 | Daler-rowney | 5146699,45 |
| 10 | 2,481121103 | Kreul | 5091823,3 |
| 11 | 2,380238327 | Таир | 4884788,96 |
| 12 | 2,34280247 | СПБФ Гознака | 4807962,09 |
| 13 | 2,207943771 | Сонет | 4531201,45 |
| 14 | 1,710195114 | БМ "Передвижник" | 3509708,3 |
| 15 | 1,684802604 | Круглов А.Г. | 3457597,11 |
| 16 | 1,486569342 | Koh-i-Noor | 3050777,49 |
| 17 | 1,483316027 | Подольск Арт-Центр | 3044100,95 |
| 18 | 1,412010196 | Fabriano | 2897765,21 |
| 217 | 1,366601774 | изд. АСТ | 2804576,83 |

Рисунок 14. Результаты анализа объединенного файла.

Как видно лидеры остались прежними 3ХК, Гамма., Faber-Castell, однако сумма и проценты изменились, а также увеличилось количество брендов (217) за счет добавленного набора данных.

Сохраните объединенный сценарий в новом файле Простой расчет2. Для этого в

Главном меню выберите Пакеты и дале Сохранить как...

Задания для самостоятельной работы

- 1. Создайте новый пакет и сохраните его под именем test1.lgp.
- 1.1. Создайте и сохраните в любом текстовом редакторе файл следующего вида:

a,1,4.5,b,c,26/04/2007,d a1,0,5,b1,c1,,d1

- 1.2. Импортируйте его в Loginom, корректно настроив параметры импорта. Метку узла переименуйте в *Пример импорта файла*. В комментарии к узлу впишите: *Текстовый файл с разделителями-запятыми*.
- 1.3. В настройках узла задайте следующие метки к столбцам: *Поле*1, *Поле*2, *Поле*3 и т.д. *Поле*6 задайте тип *Дата/время*.
- 1.4. Экспортируйте набор данных в текстовый файл с настройками, предлагаемыми по умолчанию. Импортируйте только что экспортированный файл в Loginom. В

импортированном наборе данных измените тип столбца *Поле*2 на *погический*. Сохраните пакет.

- 1.5. Добавьте к первому узлу с импортированными данными с помощью визуализатор Статистика. Имеются пропуски данных?
- 1.7. Создайте еще один визуализатор **Таблица**. Сделайте первые три столбца невидимыми. Сохраните конфигурацию визуализатора под названием «К2».
- 1.8. Вернитесь к конфигурации К1. В визуализаторе **Таблица** установите фильтр «*Поле*6 = не пустой». Сохраните пакет.

2. Создайте новый проект test2.lgp.

- 2.1. Импортируйте в него текстовый файл CreditSample.txt.
- 2.2. Отсортируйте (компонент **Сортировка**) этот набор данных по следующим полям в порядке возрастания: *Срок ссуды, Размер ссуды, Количество иждивенцев*.
- 2.3. Сделайте следующую замену (компонент Замена) в поле Семейное положение: значение Да измените на Женат/замужем, Нет на Холост/Не замужем.
- 2.4. Сделайте следующую замену (после предыдущего узла **Замена**) в поле Количество иждивенцев: значение 0 на Нет, 1 без изменений, 2 и 3 2 и более.
- 2.5. Отфильтруйте набор данных, полученный в п. 2.4 по полю Иждивенцы так, чтобы в выходной набор попали только строки, у которых значение в поле Иждивенцы не равно Нет. Сколько записей прошло через фильтр?
- 2.6. Продолжите фильтровать набор данных, полученный в п. 2.5. Наложите следующий фильтр, в который попадают все записи, удовлетворяющие условиям а либо условиям b:
 - а. Размер ссуды от 2000 до 5000, Цель ссуды Покупка товара.
 - b. Цель ссуды Иное. 9
- 2.7. Сколько записей прошло через фильтр?
- 2.8. Отсортируйте последний набор данных по полю Код.

3. Создайте новый пакет test3.lgp.

- 3.1. Импортируйте в него текстовый файл CreditSample.txt.
- 3.2. С помощью обработчика **Калькулятор** создайте новое поле **Дата обработки**, значения в котором равны текущей дате.
- 3.3. Создайте новое поле Размер ссуды у.е., который рассчитывается делением на 70 поля Размер ссуды, руб. Все значения в новом поле должны быть округлены до второго знака.
- 3.4. Создайте новое поле Флаг, значение в котором истинно, если выполняется условие:

Среднемесячный доход > 2000 и Наличие недвижимости = Да.

3.5. Создайте еще один столбец, значение в котором равно 1, если выполняется условие:

Флаг = TRUE и Давать кредит = FALSE.

- 3.6. Создайте новое поле RATE, в котором хранится значение в поле Срок ссуды, возведенное в степень 0,6.
- 3.7. Создайте новое поле Сегмент, которое делит всех заемщиков на сегменты по следующим правилам (используйте функцию IF/IFF):
- 1) ЕСЛИ Возраст >= 50 и Среднемесячный доход < 6000 ТО Сегмент = Сегмент 1
- 2) ЕСЛИ Возраст < 30 ТО Сегмент = Сегмент 2
- 3) Сегмент = Сегмент 3 во всех остальных случаях, не удовлетворяющим п. 1) и 2).

4. Создайте новый пакет test4.lgp.

- 4.1. Импортируйте в него файл *Товары.txt*.
- 4.2. Исключите из набора записи, в которых для товара отсутствует артикул.
- 4.3. С помощью функций компонента **Калькулятор** рассчитайте, сколько месяцев прошло от даты последней продажи каждого товара до 01.04.2018.
- 4.4. Отсортируйте набор данных по количеству месяцев по убыванию.
- 4.5. Добавьте в набор поле логического типа **Вывод из продажи** и установите значение **true** для товаров, у которых от даты продажи до 01.04.2018 прошло более 10 месянев.
- 4.6. Ответьте на вопросы:
 - У какого количества товаров отсутствует артикул?
 - Сколько месяцев прошло с продажи отдельных товаров?
 - Какие товары не продавались дольше всего? Как долго?
 - Какое количество товаров пора выводить из продажи?

Контрольные вопросы

- 1. Детально опишите основные области интерфейса Loginom.
- 2. Опишите способы импорта и экспорта данных, включая подключения из баз данных.
- 3. Какое расширение имеют Loginom Data файлы.
- 4. Как задать метку узлу сценария и для чего они нужны.
- 5. Как визуализировать данные в узле, какие типы визуализаторов доступны в Loginom?
- 6. Какую первичную информацию о наборе данных можно получить после использования визуализатора Статистика?
- 7. Назовите назначение фильтрации данных в Loginom. Опишите обработчик Фильтр строк и его основные особенности.
- 8. В каких случаях в обработчике Фильтр строк выводится такое сообщение: Достигнуто максимальное количество строк сканирования: 1000?
- 9. Опишите назначение обработчика Группировка данных в Loginom и его основные особенности.
- 10. Что такое функции агрегации?
- 11. Какие области для ввода данных можно выделить в мастере настройки Калькуляторе Loginom. Охарактеризуйте их.
- 12. Какие способы расчета выражений существует в калькуляторе.
- 13. Можно ли установить для разных полей в калькуляторе разные способы расчета.

- 14. Опишите каким может быть аргумент Тип в функции Stat?
- 15. Назовите назначение обработчика Объединение и его основные особенности.
- 16. Для каких полей возможно сопоставление в обработчике Объединение.
- 17. С помощью каких элементов управления осуществляется ручная настройка в мастере настройки обработчика Объединение.
- 18. Для чего применяются префиксы в обработчике Объединение?
- 19. Назовите 2 способа быстрого просмотра результатов расчёта в узлах.
- 20. Назовите способы запуска сценария.
- 21. Как в визуализаторе Таблица скрыть столбец и отсортировать данные.