минобрнауки россии

федеральное государственное бюджетное

образовательное учреждение высшего образования

«ЧЕРЕПОВЕЦКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

|  |
| --- |
| Иформационных технологий |
| наименование института (факультета) |
| Математическое и программное обеспечение ЭВМ |
| наименование кафедры |

ОТЧЁТ

по учебной: технологической практике 1

Листов 27

17 вариант

Студента Пикаловой Анастасии Сергеевны

Группы 1ПИб-01-21оп\_\_ \_\_

Место прохождения практики

ФГБОУ ВО «Череповецкий государственный университет» кафедра математического и программного обеспечения ЭВМ, компьютерный класс

Студент \_\_\_\_\_\_\_\_\_\_\_\_

(подпись)

Руководитель практики

от кафедры МПО ЭВМ доцент к.т.н. Юдина О.В.\_\_\_\_\_\_\_\_

(должность) (подпись, Ф.И.О.)

2022 год

Оглавление

Оглавление

[**Часть 1** 4](#_Toc121916319)

[**Раздел 1. Изучение возможностей инструментов для анализа данных** 4](#_Toc121916320)

[**Раздел 2. Описание выполнения практического задания** 14](#_Toc121916321)

[**Работа с данными «Adult»** 14](#_Toc121916322)

[**Работа с данными «Hausing»** 16](#_Toc121916323)

[**Результаты работы** 19](#_Toc121916324)

[**Часть 2** 20](#_Toc121916325)

[**Решение оптимизационных задач в EXCEL, с использованием** 20](#_Toc121916326)

[**пакета анализа** 20](#_Toc121916327)

[**Задание №1** 20](#_Toc121916328)

[**Задание №2** 23](#_Toc121916329)

[**Вывод:** 27](#_Toc121916330)

[**Литература** 29](#_Toc121916331)

**Введение**

Навыки работы с данными необходимы при работе в таких областях как Data Mining, Big Data, машинное обучение. Для решения этой задачи в настоящее время создано много пакетов прикладных программ, позволяющих получать оценки, делать выводы, проводить необходимые расчеты.

Целью технологической практики на втором курсе является обзорное знакомство с некоторыми из этих пакетов. В дальнейшем часть из них будут использоваться как инструментарий.

По итогам работы необходимо составить отчет, представляющий собой описание выполнения заданий и полученные решения. Титульный лист отчета представлен.

# **Часть 1**

## **Раздел 1. Изучение возможностей инструментов для анализа данных**

**KNIME**

KNIME Analytics Platform – это программная платформа анализа, интеграции данных и подготовки отчётности с открытым исходным кодом.

Программный продукт KNIME Analytics Platform от компании KNIME предназначена для исследования данных с использованием обширного инструментария аналитических инструментов: прогнозной аналитики, администрирования баз данных, импорта/экспорта данных, совместной работы с информацией и другого. Программный продукт может использоваться отдельными исследователями и организациями любых размеров.

Аналитическая платформа KNIME предоставляет пользователям возможности визуально создавать потоки данных (конвейеры), выборочно выполнять отдельные или все шаги анализа, а затем проверять результаты, модели, используя интерактивные виджеты и представления.

Аналитическая система KNIME Analytics Platform может быть использована как в виде отдельно-стоящего приложения на базе дистрибутива с открытым исходным кодом, так и в форме аренды облачной версии системы.

В Knime процесс программирования логики осуществляется через создание Workflow. Workflow состоит из узлов, которые выполняют ту или иную функцию (например чтение данных из БД, трансформация, визуализация). Узлы, соответственно, соединяются между собой стрелочками, которые показывают направление движение данных.

Результатом работы программы служит открывающееся окно с подсчитанными значениями, т.е. будет выводить визуализацию множеств решений системы.

**Преимущества:**

Программа подходит для людей, которые хотят анализировать данные и не владеет навыками программирования, и кто хочет изучить набор реализованных алгоритмов и узнать что-то новое. Простой, гибкий, функциональный инструмент для анализа данных, в том числе для тех, кто не владеет навыками программирования, который позволяет быстро создавать простые и средней сложности потоки работ. Бесплатная десктопная версия (компьютерная версия). Доступен на macOS, Windows, Linux. Интегрирование с широким спектром языков программирования, таких как Java, R, Python и многими другими для Data scientists. Наглядно видно все что происходит с данными, по каким веткам они перемещаются и как преобразуются, для начинающих аналитиков. Доступна большая библиотека примеров. Исполнение workflow на сервере и предоставление доступа к результатам работы через REST API. Данная функциональность доступна при покупке KNIME-Server. Полный дистрибутив KNIME-а со всеми плагинами весит почти 2 гигабайта. В данный дистрибутив входит большое количество сторонних библиотек (например JFreeChart), которые становятся доступны в виде узлов. Реализована возможность сделать операцию Pivot прямо на базе данных или на данных загруженных в локальный кеш. Работа с Hadoop и другими BigData источниками.

**Минусы:**

Возможность использования KNIME Analytics Platform на Android-устройствах отсутствует. Использование системы KNIME Analytics Platform на мобильном устройстве не предусмотрено. Недостаточно гибкое поведение узлов. Программирование сложных flow может вылиться в длинную и сложную диаграмму с циклами, условиями что приведет к сложности чтения и отладки.

**SPSS**

Statistical Package for the Social Sciences – это аналитическое программное обеспечение от компании IBM, позволяющее производить продвинутый статистический анализ деловых данных, охватывая решение всех задач от планирования и сбора данных до непосредственного анализа, построения бизнес-отчётности и извлекания полезных сведении из исследуемых данных.

Продвинутые статистические процедуры и визуализация в системе СПСС обеспечивают надежную, удобную и интегрированную платформу для понимания предметных данных и решения сложных бизнес- и исследовательских задач, обеспечивая в целом увеличение доходов, обыгрывание конкурентов, проведение исследований и принятия качественных решений на базе фактических данных. Программное обеспечение поддерживает следующие методы статистических исследований: регрессионный анализ, деревья решений, прогнозирование, нейронные сети, категоризация, комбинационный анализ, сложные выборки и другие.

**В числе важных особенностей информационно-аналитической системы SPSS Statistics можно отметить:**

* Система обеспечивает индивидуальную настройку функций и пользовательских интерфейсов для различных уровней квалификации и функциональных обязанностей;
* SPSS Statistics охватывает все основные части комплексного аналитического процесса – от подготовки данных и управления ими до анализа и отчётности;
* Предоставляет готовые к использованию шаблоны графиков и отчётов, позволяя легко переводить полученные результаты в презентабельный и понятный вид для других заинтересованных лиц.

**Преимущества:**

* Обеспечивает быстрое понимание больших и сложных наборов данных с помощью современных статистических процедур, которые помогают обеспечить высокую точность и качество принятия решений;
* Позволяет использовать программные расширения на языках программирования Python и R для интеграции с программным обеспечением с открытым исходным кодом;
* Облегчает выбор и управление программным обеспечением, предоставляя гибкие варианты развертывания;
* Есть бесплатная пробная версия, а также платная версия с расширениями. Имеется лицензия для обучающихся, где надо предоставить место обучения и телефон;
* Программа представляется в виде excel. Пользователю нужно знать базовую работу в excel. Открывается окно вывода, где также высвечиваются ошибки в базе данных.

**SAS**

Основные приложения SAS — настраиваемые системы Business Intelligence для финансового менеджмента, управления рисками, маркетинга, управления цепочками поставок. Все решения базируются на общей технологической платформе, которая обеспечивает базовые функциональные возможности, необходимые всем приложениям:

* Процесс извлечения данных из разнородных источников с последующей их обработкой и очисткой;
* Хранение данных в специализированном аналитическом хранилище данных;
* Формирование и доставка пользователям аналитических отчётов различного уровня сложности;
* Углублённая аналитика — среда для проведения углублённого анализа данных, описательного и прогнозного моделирования, прогнозирования временных рядов, оптимизации и тому подобных задач.

На основе общей платформы строятся решения SAS для решения задач функциональных направлений бизнеса:

* Управление эффективностью включает процессно-ориентированное управление, составление консолидированной отчётности, бюджетирование и финансовое планирование;
* Анализ клиентской базы (клиентская аналитика) включают управление маркетинговыми кампаниями, автоматизацию кросс-продаж, поведенческую сегментацию клиентов;
* Управление маркетинговыми ресурсами;
* Управление рисками (кредитными, операционными, рыночными и др.);
* Борьба с мошенничеством и отмыванием денег;
* Кредитный скоринг;
* Управление цепочками поставок;
* Прогнозирование спроса.

Функциональные решения настраиваются в соответствии с отраслевой спецификой для банков, страховых компаний, провайдеров телекоммуникационных услуг, транспортных, энергетических, промышленных и других компаний.

**Преимущества и недостатки SAS:**

* SAS требует дорогое программное обеспечение;
* Простой синтаксис, быстрое обучение “с нуля”;
* Отладка кода проходит значительно проще, чем на R;
* Интеграция с БД (Oracle/Teradata);
* Удобный формат выходных данных (особенно таблиц);
* Мощная поддержка со стороны компании SAS;
* Многолетний успешный опыт эксплуатации компаниями разной величины, с разными задачами и разным объёмом входных данных. В частности, России SAS используют ОАО “РЖД”, МТС, ЦБ РФ, а также ведущие банки, среди которых Сбербанк, Альфабанк, Тинькофф и многие другие.
* Профессиональное использование языка предполагает покупку программного продукта;
* Исходники многих исполняемых алгоритмов SAS не являются публичными, следовательно изучение работы языка сильно ограничено;
* SAS значительно уступает в производительности R;
* С точки зрения объёма кода SAS также зачастую сильно проигрывает (иногда в несколько раз);
* Компания активно продвигает свой продукт среди студентов, причём абсолютно бесплатно.

Также на официальном сайте у вас есть возможность записаться на экспресс-курсы по изучению языка (причём на русском языке). Рекомендуемая длительность обучения в каждом разделе составляет 3 дня, требования и программу можно посмотреть прямо рядом с торжественной кнопкой записи.

Научиться программировать на SAS Base можно на бесплатном интерфейсе под названием SAS University Edition или сокращенно SAS UE. Скачать его можно на сайте SAS абсолютно бесплатно. На ПК SAS UE запускается под виртуальной машиной и требует установки программного обеспечения виртуализации. Для Windows подойдет Oracle VM VirtualBox и Vmware Player, но чтобы работать в этой программе стоит покупать лицензию.

**STATISTICA 6.0**

Программный пакет для статистического анализа, разработанный компанией StatSoft, реализующий функции анализа данных, управления данными, добычи данных, визуализации данных с привлечением статистических методов.

Пакет обладает широкими графическими возможностями, позволяет выводить информацию в виде различных типов графиков (включая научные, деловые, трёхмерные и двухмерные графики в различных системах координат, специализированные статистические графики — гистограммы, матричные, категорированные графики и др.), все компоненты графиков настраиваются.

Линейка STATISTICA 6 основана на самых современных технологиях, полностью соответствует последним достижениям в области IT, позволяет решать любые задачи в области анализа и обработки данных, идеально подходит для применения в любой области: маркетинге, финансах, страховании, экономике, бизнесе, промышленности, медицине и др.

STATISTICA 6 полностью переведена на русский язык, включая интерфейс, документацию и Электронное справочное руководство.

Размер параметров неограничен. Размер текста в названиях переменных (8 символов) и наблюдений. Ранее существовали ограничения на размер файла данных (порядка 5000х20000). Сейчас это ограничение для большинства видов анализа снято. Весь интерфейс пользователя полностью настраивается. Например, можно создать собственные панели инструментов, диалоги, меню и совершенно новые упрощенные (или наоборот продвинутые) способы выполнения специальных аналитических процедур.

Язык Visual Basic, интегрированный в STATISTICA 6.0, сочетает все функциональные возможности языков STATISTICA Basic и SCL (реализованных в версии 5) с большим количеством новшеств. Имеются система управления графическими стилями, а также и мощные инструменты для трехмерного разделения данных гиперплоскостями.

**Преимущества решения:**

* Наличие качественных инструментов для построения графиков и схем;
* Сформированные отчеты можно анализировать в Word;
* Высокая точность расчетов, ошибки исключены;
* Есть руководство по решению статистических операций;
* Разнообразные методы редактуры графиков, схем, диаграмм;
* Возможность запуска системы через Excel;
* Доступность средств детализации, качественной графики.

**Оно включает выполнение таких процедур, как:**

* Автокорреляции;
* Множественная регрессия;
* Аппроксимация;
* Графический анализ таблиц;
* Вычисление экстремумов;
* Подгонка распределений;
* Байесовский анализ и т.д.

Таковы графические возможности пакета Statistica. Это лишь некоторые инструменты, входящие в модули программного обеспечения. В состав группы, которая позволяет использовать разведочные технологии, входят: деревья классификации, кластерный, дискриминантный и факторный анализ, деревья решений и другие позиции.

Также есть портативная версия. Доступ к portable программе осуществляется через электронную почту или облако. Все преимущества, инструменты и функции комплекса будут доступны.

Можно скачать бесплатно, анализируемые данные представляются в STATISTICA в виде электронной таблицы, подобно тому, как это делается, например, в программе MS Excel.

**Minitab**

Флагманский продукт компании, Minitab Statistical Software, представляет собой мощный статистический пакет для решения самых разнообразных задач анализа данных. Minitab содержит все необходимые инструменты для первичной обработки данных, анализа результатов и подготовки презентационных материалов и отчетов. Пользователи легко могут корректировать входные данные, обрабатывая пропуски или опечатки, удалять лишние пробелы и контролировать размер данных, импортируемых из MS Excel, баз данных или других источников. При изменении данных все графики автоматически обновляются, так что пользователю не нужно беспокоиться об их актуальности. Мощные возможности по форматированию данных внутри рабочих документов позволяют легко выделять подмножества часто встречающихся значений, выбросов, бракованных изделий и многого другого. Все графические материалы, созданные в Minitab, легко можно экспортировать в MS Word или MS PowerPoint для создания презентаций и отчетов. Новая схема лицензирования предоставляет пользователям Minitab возможность работы как в классическом приложении, так и в новой полнофункциональной облачной версии! Благодаря единой учетной записи, пользователи могут работать в Minitab на любом устройстве с доступом в интернет: дома, в дороге или на работе. Облачная версия Minitab не только позволяет выполнять все виды анализа и визуализировать данные, но легко делиться ими с коллегами. В Minitab пользователи смогут найти все необходимые инструменты для анализа данных, а с помощью встроенных подсказок получить необходимые результаты и ответить на самые разнообразные вопросы.

Основные возможности Minitab:

* Базовая статистика;
* Регрессионный и дисперсионный анализ (ANOVA);
* Предсказательная аналитика;
* Анализ измерительных систем;
* Планирование экспериментов;
* Контрольные карты;
* Анализ надежности и выживаемости;
* Встроенный помощник.

После завершения анализа через помощника пользователь получает серию отчетов, которая содержит детальную интерпретацию результатов. Итоговый отчет содержит основные выводы по проведенному анализу вместе с комментариями, которые позволяют сформулировать ответ в контексте рассматриваемой проблемы. Диагностический отчет содержит более детальную информацию о результатах, включая, например, информацию о выбросах в данных, на которые необходимо обратить внимание. Наконец, отчетная карточка содержит информацию о надежности результатов анализа - достаточен ли размер выборки, распределены ли данные по нормальному закону и т.д.

**R + R-Studio**

R-STUDIO это группа надежных, эффективных и рентабельных утилит для восстановления диска от компании R-Tools Technology, Inc. Первоначально предназначенная для специалистов в области восстановления информации, в дальнейшем R-Studio стала многофункциональным, удобным и универсальным инструментом для восстановления данных. Сочетая новейшие уникальные технологии восстановления файлов и диска с удобным пользовательским интерфейсом, R-Studio позволяет эффективно решать задачи как компаниям и частным лицам, занимающимся восстановлением данных, так и начинающим пользователям.

R-Studio запускается на Mac, Windows и Linux ОС и позволяет восстанавливать данные с локальных дисков, съемных носителей, значительно поврежденных дисков, незагружающихся дисков, а также с других компьютерах в локальной сети или в Интернете. Благодаря своей гибкости и универсальности R-Studio является идеальным решением как при выполнении простых, так и комплексных задач по восстановлению данных.

R-Studio позволяет быстро и эффективно восстанавливать утраченные данные с поврежденного, форматированного, переформатированного или удаленного диска (раздела) любого локального или удаленного компьютера вне зависимости от используемой системы (платформы).

Хотелось бы отметить, что это приятная отличительная черта пакета - очень простая и очевидная навигация, позволяющая выбрать наиболее удобный способ выполнения каждой операции. Если по каким-либо причинам автоматического распознавания не произошло, программа позволяет создать виртуальный массив и работать с ним как с реальным. Главное - знать порядок расположения дисков в нем.

Можно увидеть все накопители, обнаруженные в системе. В случае с жесткими дисками наглядно представлены все существующие логические разделы и неразмеченные области, но работой с винчестерами программа не ограничивается, поддерживая также CD- и DVD-диски, дискеты, различные флеш-карты и USB-накопители. Все это удобно представлено в левой части окна в виде древовидной структуры, напоминающей привычный "Проводник". Правая часть окна отведена под справочную информацию о выбранном носителе и может похвастаться полным и удобным представлением действительно необходимых параметров и характеристик - от файловой системы до физической геометрии диска.

Освоение азов работы с программой является достаточно несложным для тех, кто имеет общее представление о принципах хранения информации в компьютере.

Для повышения шансов восстановления определенных файлов пользователь может добавлять собственные сигнатуры файлов любой степени сложности. Добавить собственную сигнатуру в файлы известных типов можно либо посредством пользовательского интерфейса, либо путем непосредственного редактирования файла описания сигнатуры.

При помощи R-Studio можно создавать файлы образа диска, являющиеся точной, "байт в байт" копией любого объекта панели Диски (Drives) - такого, как жесткий диск, логический диск или сложный RAID. Образы весьма полезны тогда, когда существует вероятность полной утраты данных вследствие неисправности аппаратной части компьютера. Файлы образов дисков могут обрабатываться программой как обычный диск. Также восстанавливать данные по сети можно, если компьютер не загружается по причине повреждения загрузочного раздела или системных файлов.

Программа от официального производителя платная. Имеются несколько видов лицензий. Также эту программу можно скачать и бесплатно или скачать демо версию.

**Python**

Python — это интерпретируемый, высокоуровневый объектно-ориентированный язык общего назначения, используемый для разработки API, искусственного интеллекта, веб-разработки, интернета вещей и так далее.

Отчасти Python стал так популярен благодаря специалистам в области data science. Это один из самых простых языков для изучения. Он предлагает множество библиотек, которые применяются на всех этапах анализа данных. Поэтому язык однозначно подходит для этих целей.

После десятилетий разработки с открытым исходным кодом Python предоставляет широкие функциональные возможности за счет мощных библиотек для работы со статистикой и числами. В основном они бесплатные:

* NumPy и Pandas упрощают анализ данных и управление ими;
* Matplotlib позволяет создавать привлекательные визуализации данных;
* Scikit-learn предназначена для простого и эффективного прогнозного анализа данных;
* TensorFlow и PyTorch обеспечивают возможности машинного обучения и глубокого обучения.

Как правило, проект обработки и анализа данных предназначен для получения аналитических сведений в конкретном сценарии или для проверки гипотезы.

NumPy — это библиотека Python, которая предоставляет функциональные возможности, сравнимые с математическими инструментами, такими как MATLAB и R. Несмотря на то, что в NumPy значительно упрощено взаимодействие с пользователем, она предлагает большой набор математических функций.

Pandas — это очень популярная библиотека Python для анализа и обработки данных. Pandas похожа на Excel для Python, поскольку включает простые в использовании функции для таблиц данных.

Python приобрел популярность не так давно. В 2020 Python стал четвертым в списке самых используемых языков программирования после JavaScript, HTML/CSS и SQL — его используют 44,1% разработчиков.

Python работает практически на всех известных платформах – от карманных компьютеров и смартфонов до серверов сети. Его используют Google, Intel, Cisco и Hewlett-Packard, на нем работают популярные площадки YouTube, «ВКонтакте», DropBox.

**Orange**

Программный продукт Orange (рус. Оранж), разрабатываемая Лабораторией биоинформатики Люблянского университета, предназначена для интеллектуального анализа данных (ИАД), статистических исследований и визуализации данных. Компоненты аналитической платформы называются виджетами, и они варьируются от минималистичной визуализации данных, выбора подмножеств и предварительной обработки до эмпирической оценки алгоритмов обучения и прогностического моделирования. Система станет эффективным инструментом в руках аналитика данных, исследователя и учёного.

В программном обеспечении Orange Data Mining применяется визуальное программирование, которое реализуется удобным графическим интерфейсом. В рамках визуального программирования аналитические процедуры создаются путём связывания предопределенных или разработанных пользователем блоков (виджетов), в то время как продвинутые пользователи могут использовать Orange в качестве программной библиотеки Python для манипулирования данными и создания новых блоков (виджетов).

Система Orange является инструментом для визуализации и анализа данных с открытым исходным кодом. Интеллектуальный анализ данных проводится путем визуального программирования и с помощью Python сценариев. Рабочее пространство состоит из виджетов и связей между ними.

Каждый виджет имеет свой тип. Тип виджета можно определить по его иконке. Виджеты сгруппированы по разделам: Data, Visualization, Predictions и пр. Группа виджета определяет цвет иконки. Каждый виджет имеет множество (возможно, пустое) входных и множество выходных сигналов. Сигнал определяет данные, которые поступают на вход виджету или являются его результатом. При получении входного сигнала виджет выполняет определенные действия и оповещает связанные с ним виджеты путем отправки им соответствующих сигналов.

**Возможности программы:**

* Чтение, запись и верификация SEEPROM. В отличие от известных аналогов, все алгоритмы чтения и записи написаны на специальном макроязыке HPLTM, что позволяет пользователю не только изучать и корректировать протоколы обмена, но и самостоятельно добавлять любые SEEPROM (если аппаратная часть программатора позволяет это сделать);
* Поддержка аппаратной части программаторов любого типа с помощью специализированных драйверов;
* Чтение и запись в файл буфера обмена в bin и hex формате и его редактирование в шестнадцатеричном и символьном виде;
* Режим записи только в необходимые ячейки EEPROM;
* Интерфейс "Turbo Vision" с поддержкой мыши и "горячих" кнопок.

## **Раздел 2. Описание выполнения практического задания**

### **Работа с данными «Adult»**

1) Создан виджет «Datasets», в котором выбраны данные «Adult» с краткой аннотацией (рис.1- 2):

«Взрослый (1996), из репозитория UCI ML

Извлечение было сделано Барри Беккером из базы данных переписи 1994 года. Набор достаточно чистых записей был извлечен с использованием следующих условий: ((AAGE>16) && (AGI>100) && (AFNLWGT>1)&& (HRSWK>0)). Задача прогнозирования состоит в том, чтобы определить, зарабатывает ли человек более 50 тысяч в год.

Рекомендации

Рон Кохави (1996) Повышение точности классификаторов наивного Байеса: гибрид дерева решений. Материалы второй международной конференции по обнаружению знаний и интеллектуальному анализу данных.»

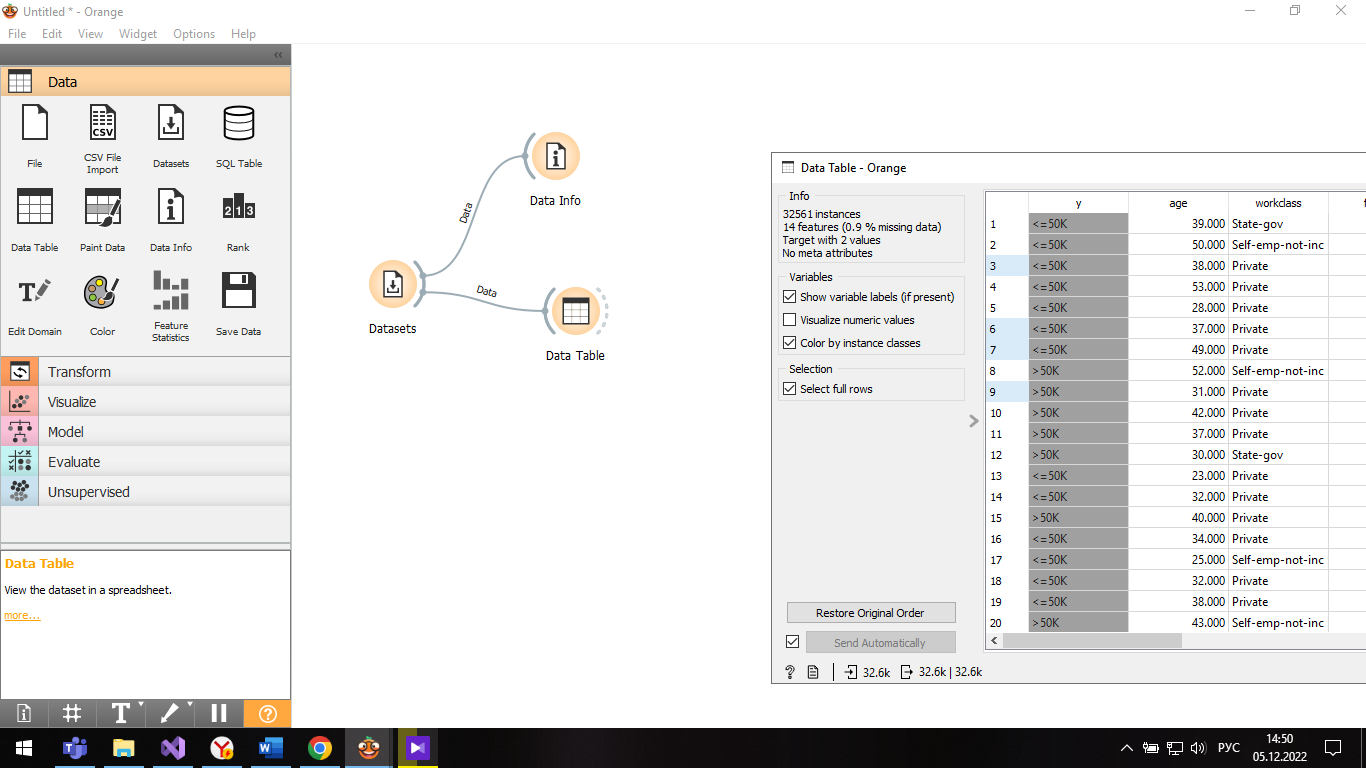


Рис. 1. Создание виджета «Datasets»

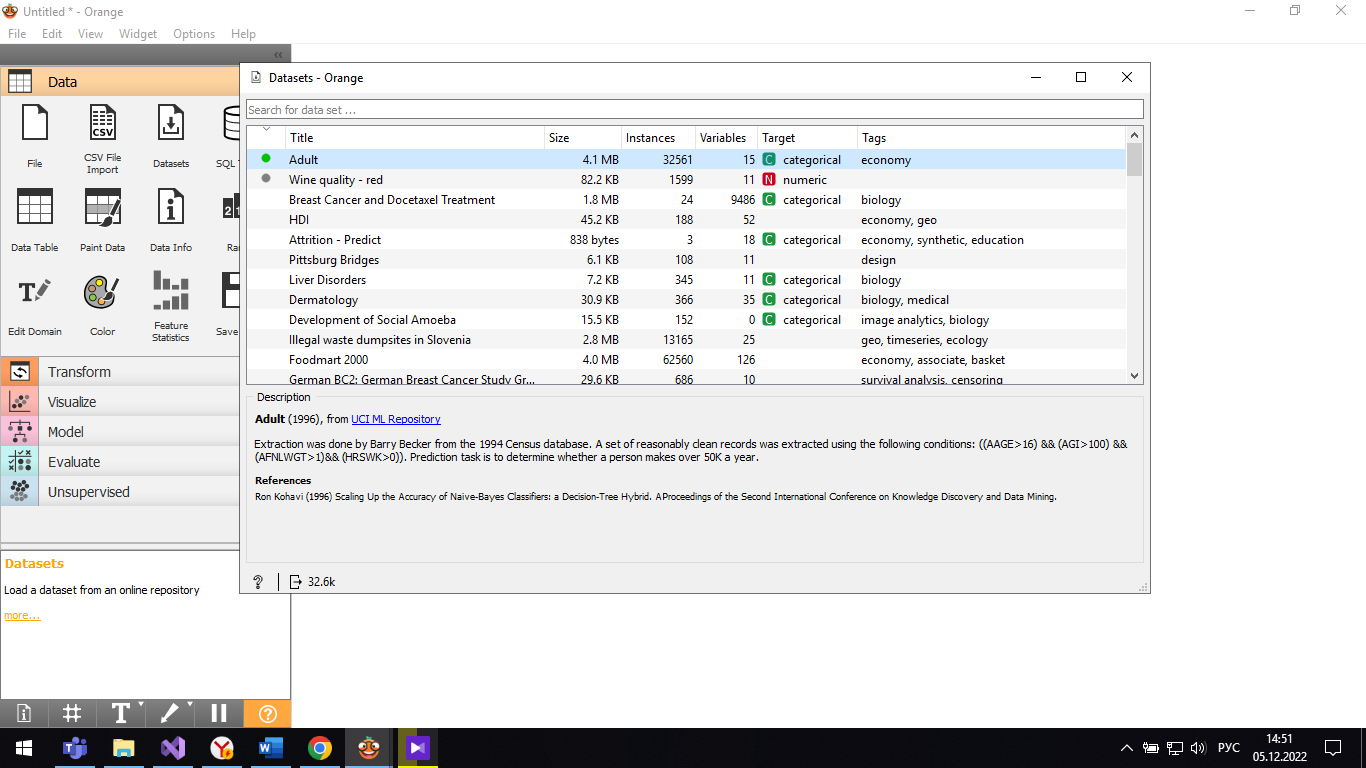


Рис. 2. «Adult» с краткой аннотацией

2) Двойным кликом на виджет «Data Table» удалось визуализировать таблицу (рис.3):



Рис. 3. Представление табличных данных из «Adult»

3) Добавлен виджет «Scatter Plot», чтобы можно было просмотреть визуализация данных. В ходе изменений параметров виджета «Scatter Plot» получены разные отображения данных «Adult» (рис.4-7):

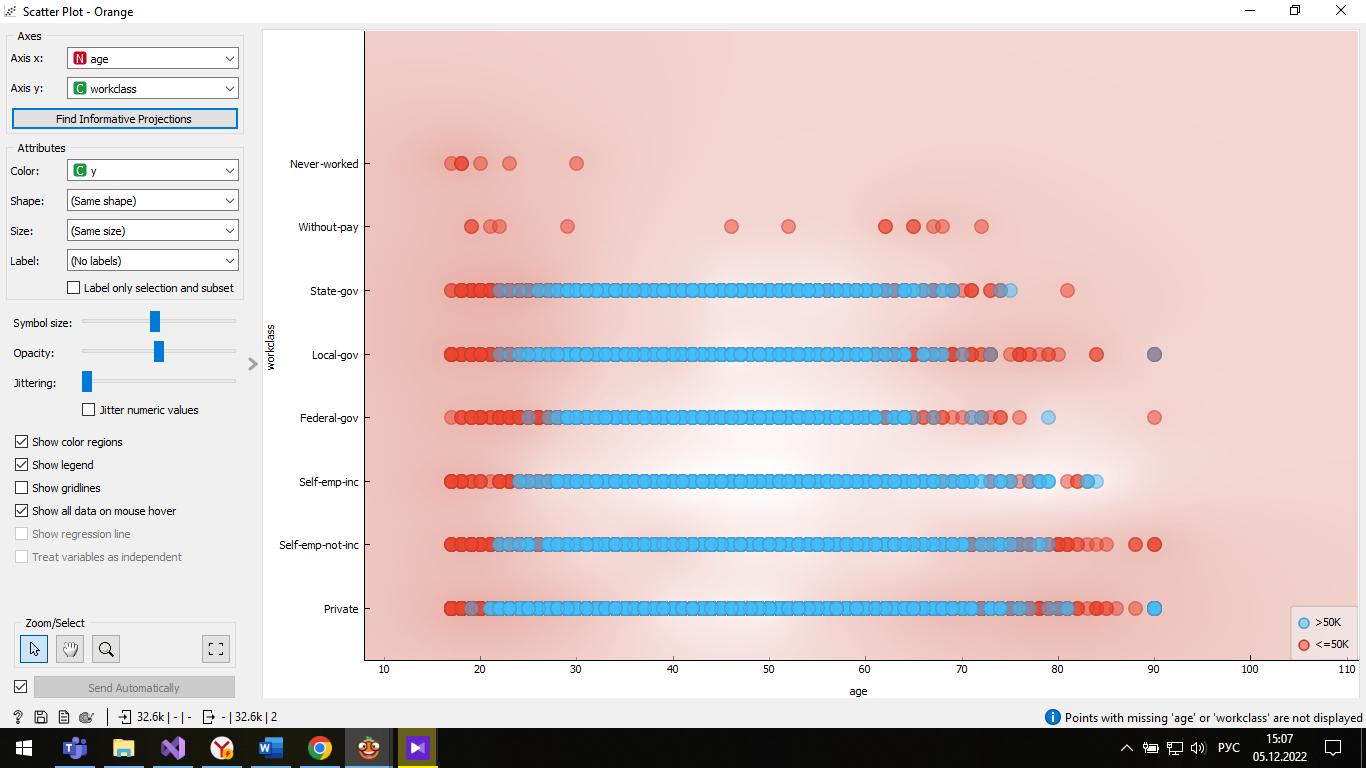


Рис. 4. Представление табличных данных из «Adult»

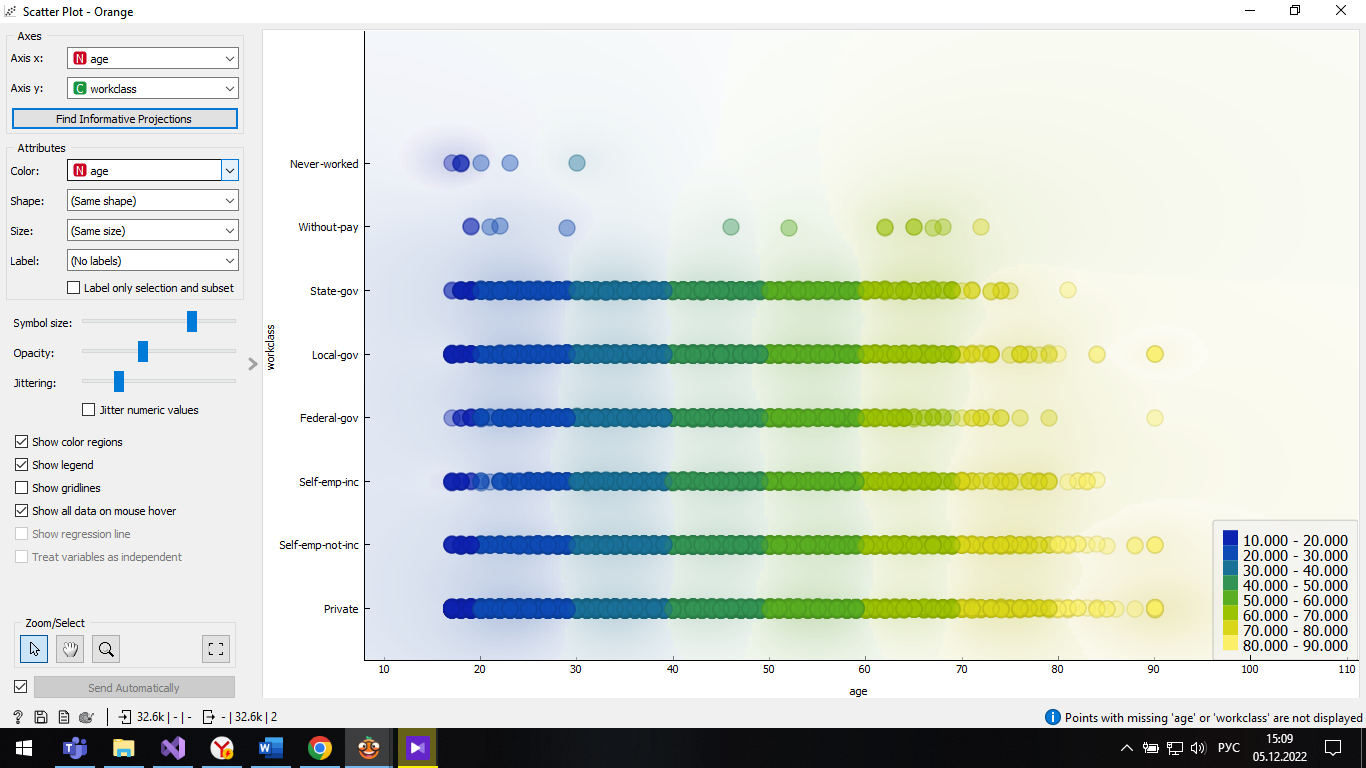


Рис. 5. Представление табличных данных из «Adult»

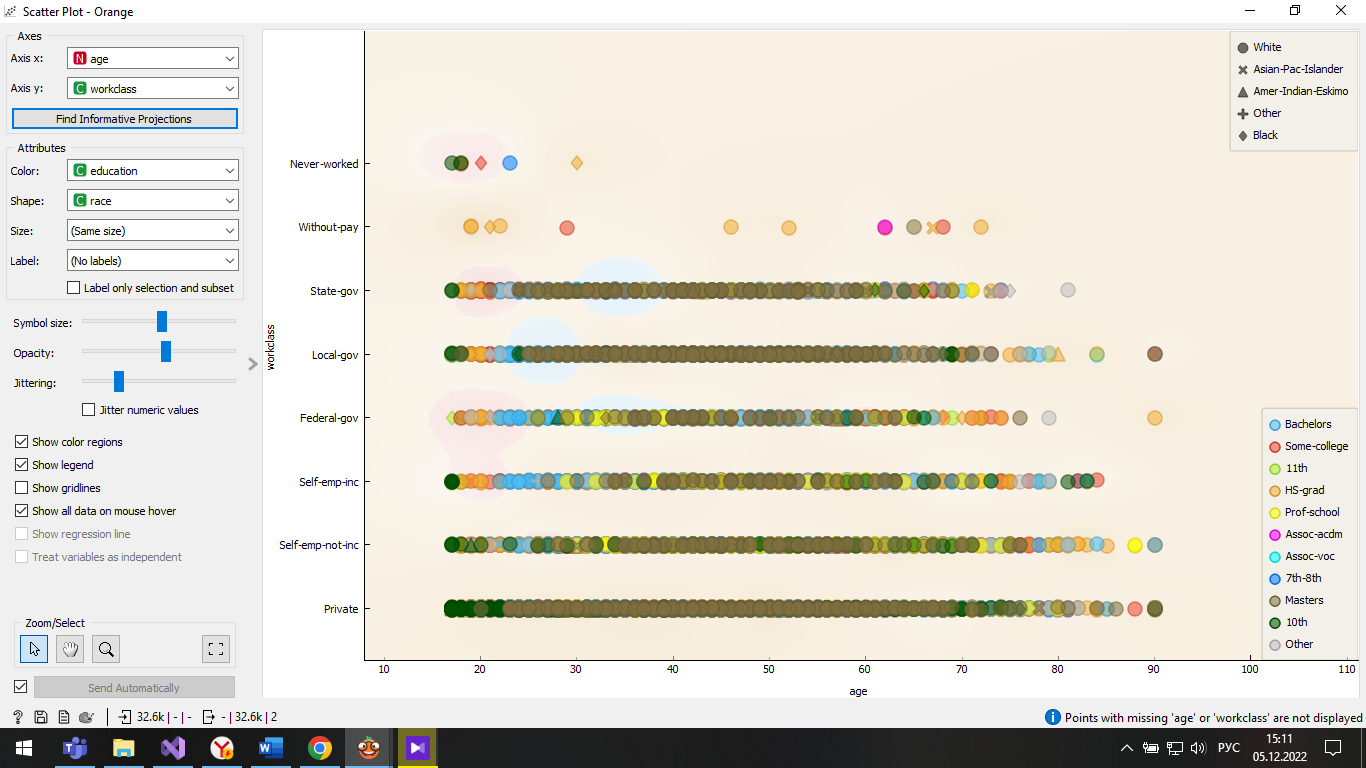


Рис. 6. Представление табличных данных из «Adult»

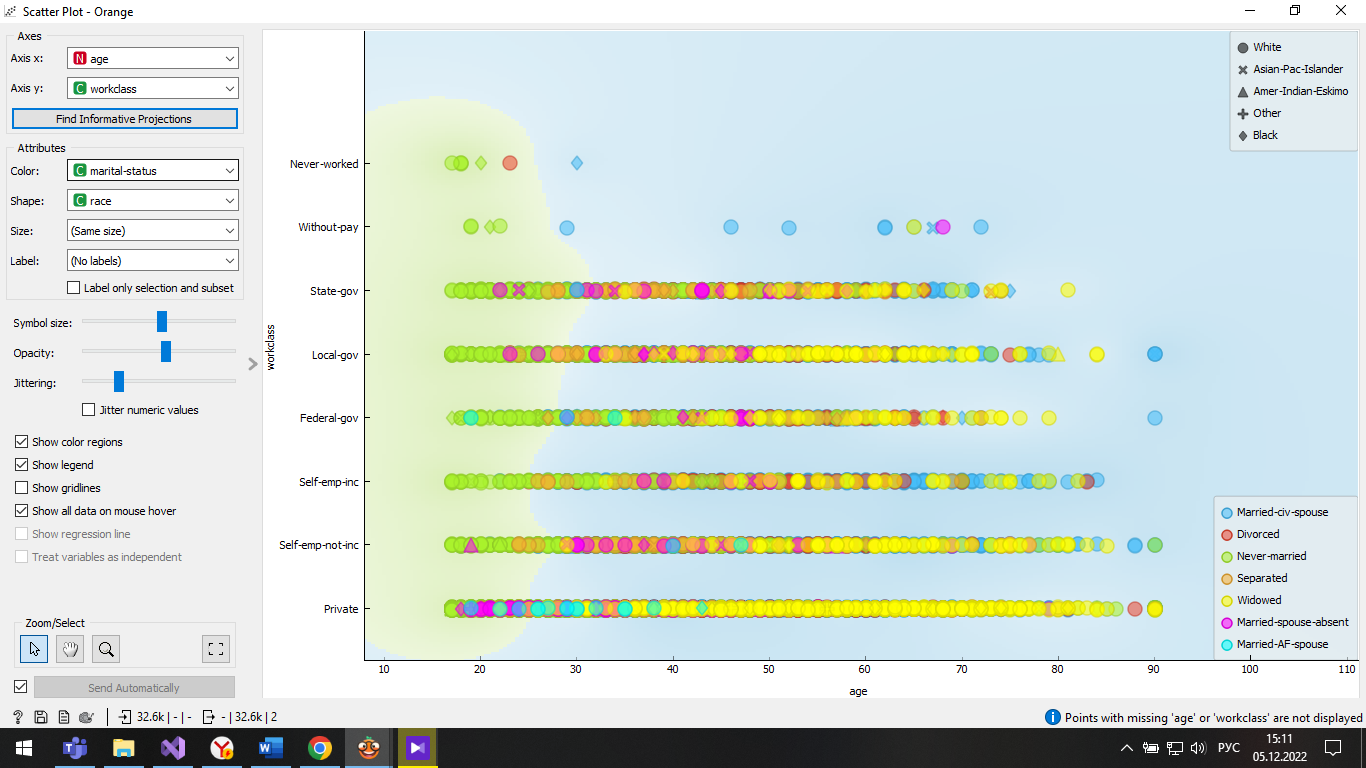


Рис. 7. Представление табличных данных из «Adult»

### **Работа с данными «Hausing»**

1) Создан виджет «Datasets», в котором выбраны данные «Hausing» с краткой аннотацией (рис.8):

«Жилье (1979), из StatLib

Этот набор данных содержит информацию, собранную Службой переписи населения США, о жилье в районе Бостона. Он был получен из архива StatLib и широко использовался в литературе для тестирования алгоритмов. Эти данные изначально были частью репозитория машинного обучения UCI и были удалены. Цель состоит в том, чтобы спрогнозировать значение цен на дом, используя заданные характеристики.

Рекомендации

Харрисон Д. и Рубинфельд Д.Л. (1978) Гедонистические цены и спрос на чистый воздух. J. Окружающая среда. Экономика и управление 5, 81-102.»

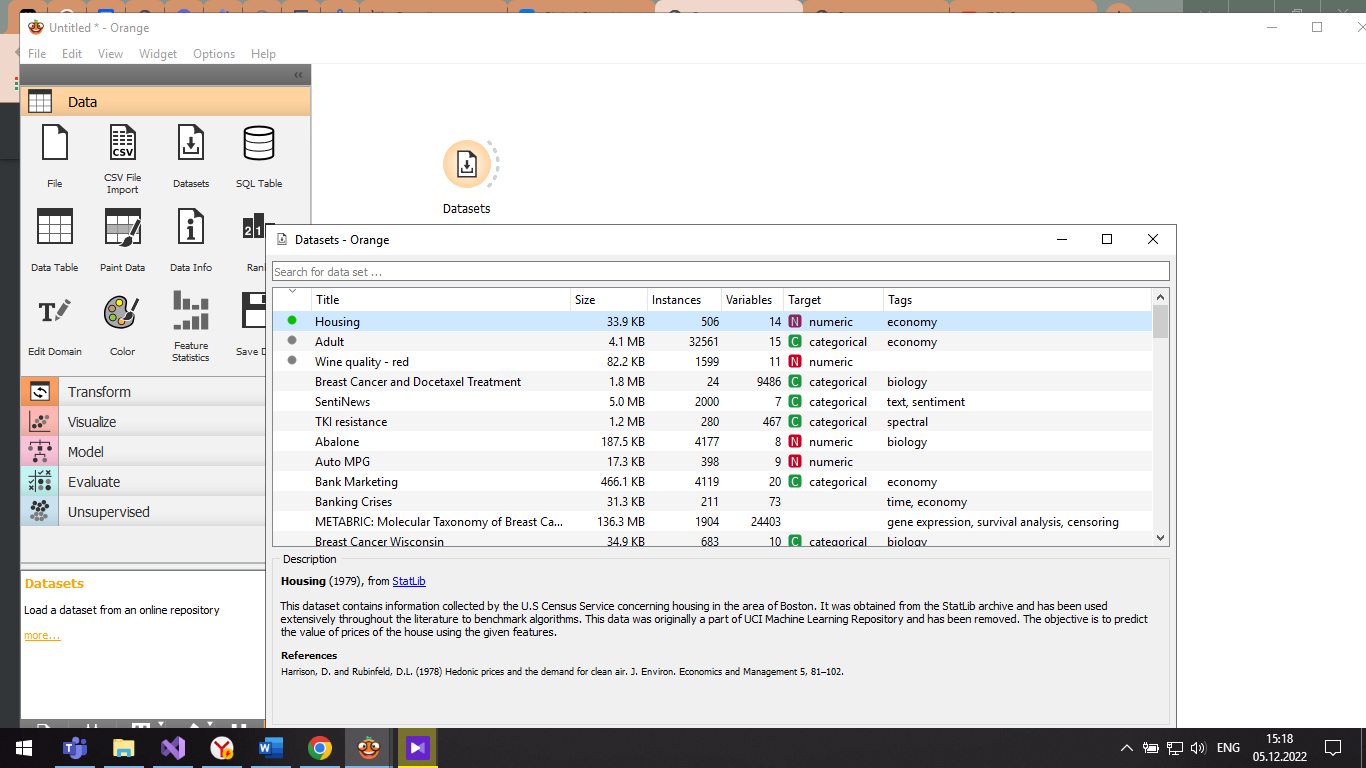


Рис. 8. «Hausing» с краткой аннотацией

2) Двойным кликом на виджет «Distributions» удалось посмотреть данные графическими представлениями, в частности, об их распределении (рис.9):

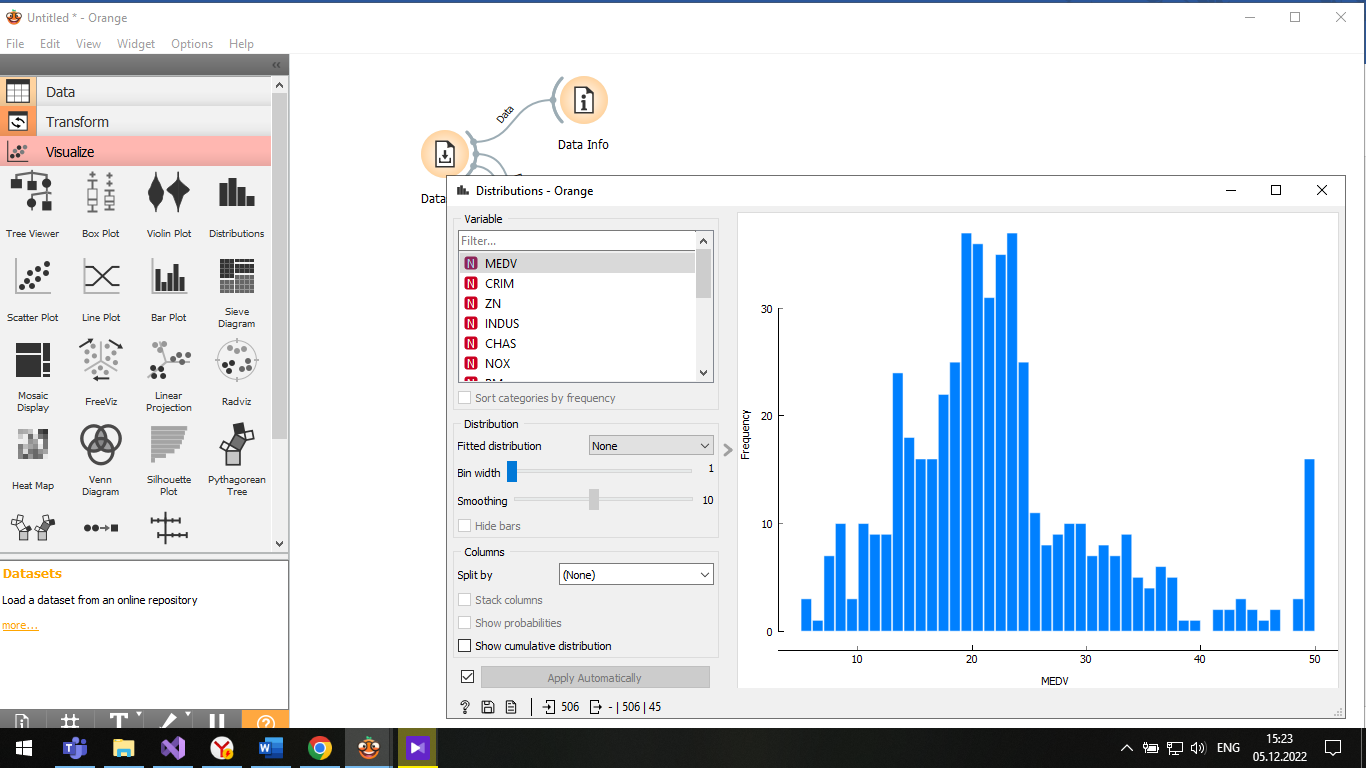


Рис. 9. Графическое представление данных «Hausing»

3) Для того, чтобы построить регрессию, использованы виджеты «Linear Regression» и «Test and Score» и соединить их с «Data Table» (рис.10). В полученной таблице видны значения:

* 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.

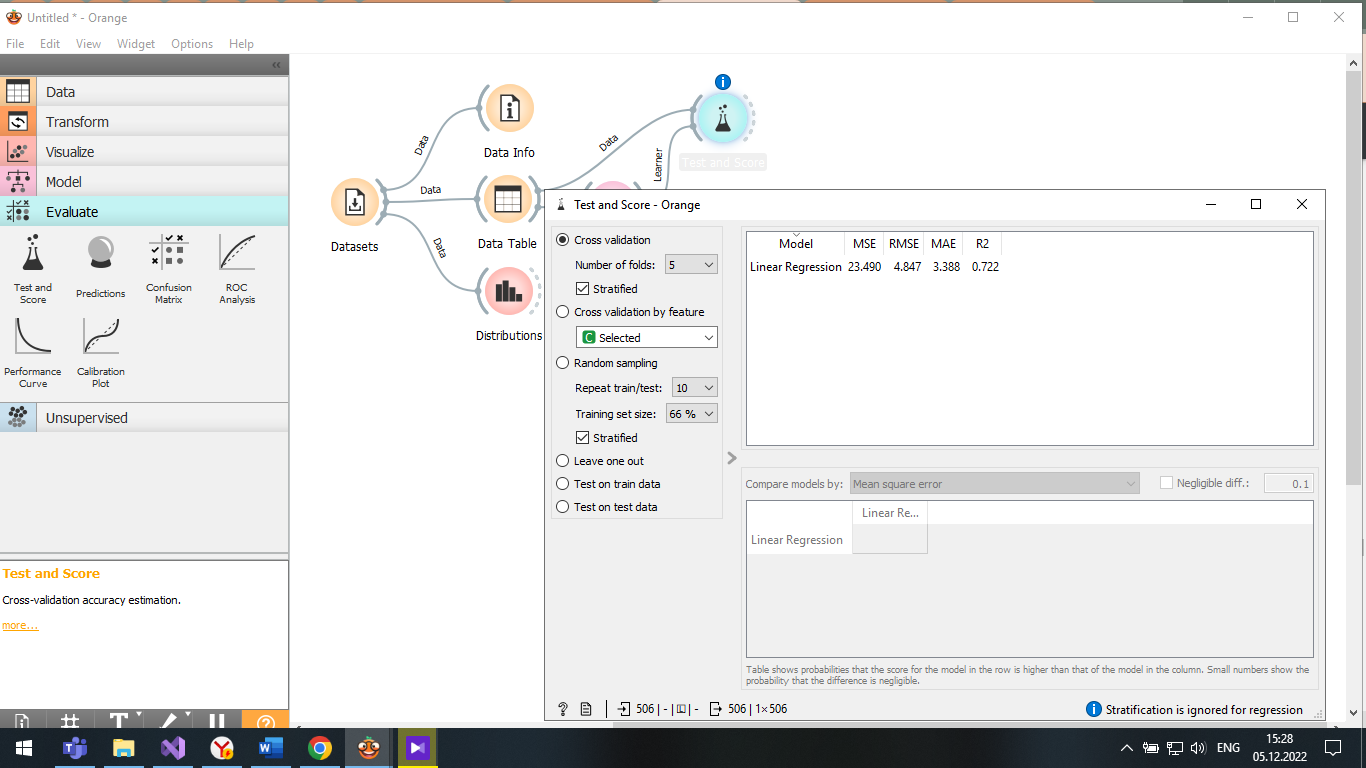


Рис. 10. Табличка с оценкой регрессии

4) Для улучшения расчётов добавлен препроцессинг и построены деревья решений(рис.11-12):

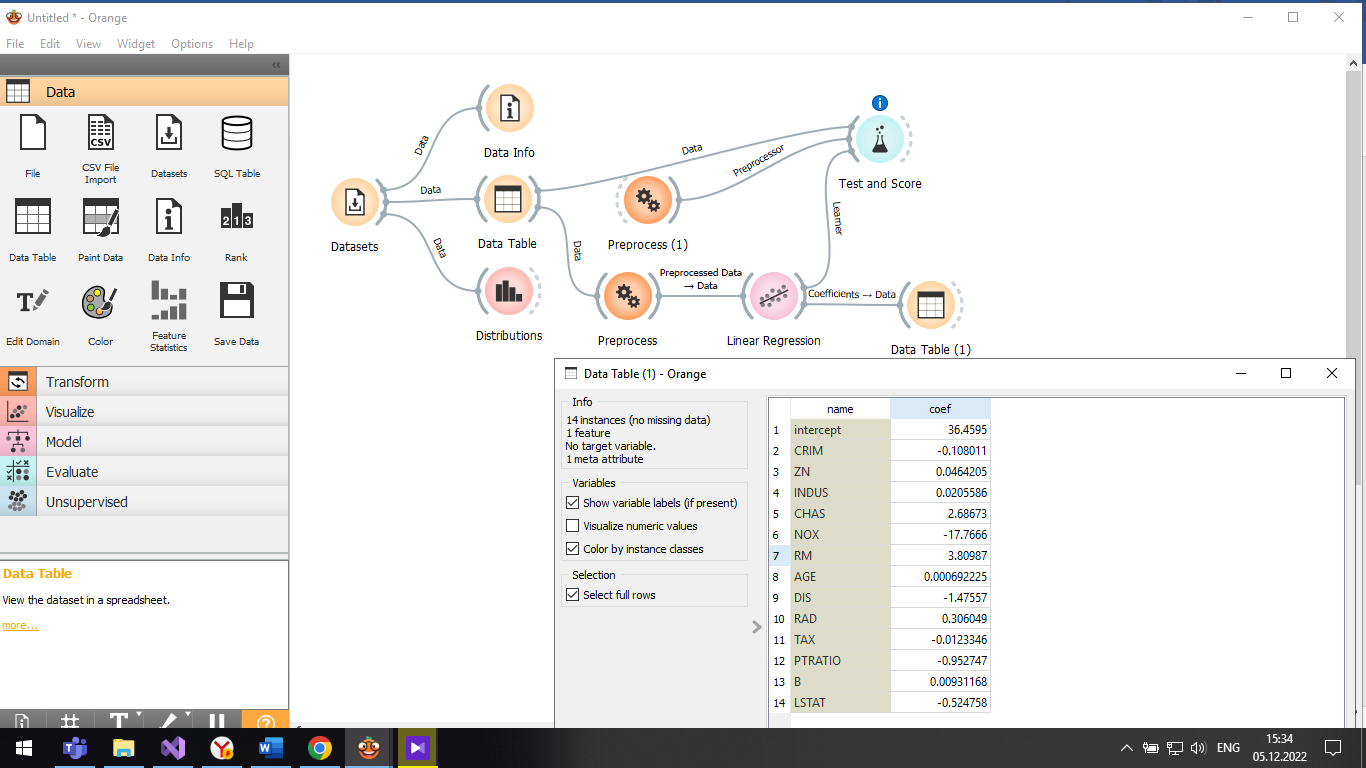


Рис. 11. Добавление препроцессинга

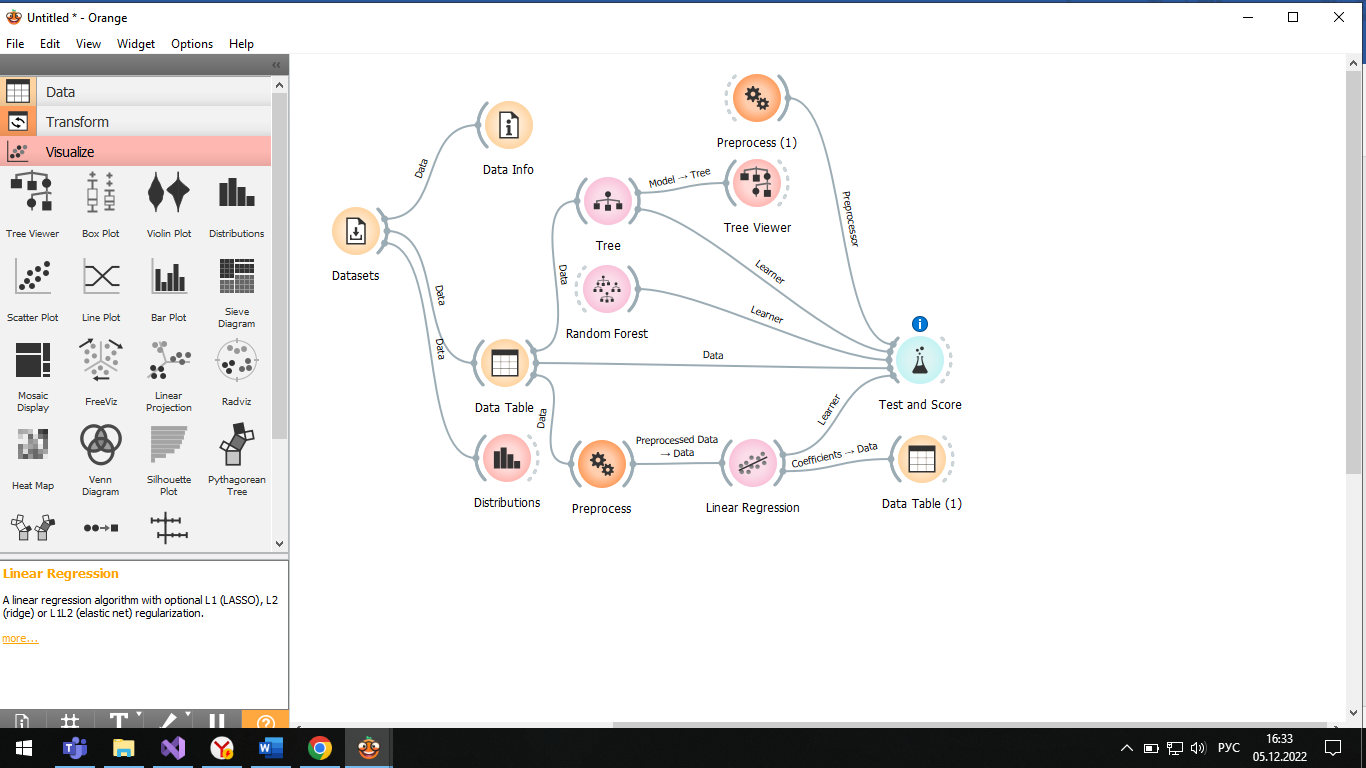


Рис. 12. Добавление деревья решений

4) В ходе работы добавились коэффициенты регрессии для дерева и линии регрессии и продемонстрирована таблица зависимостей, которые влияют на результат (рис.13):

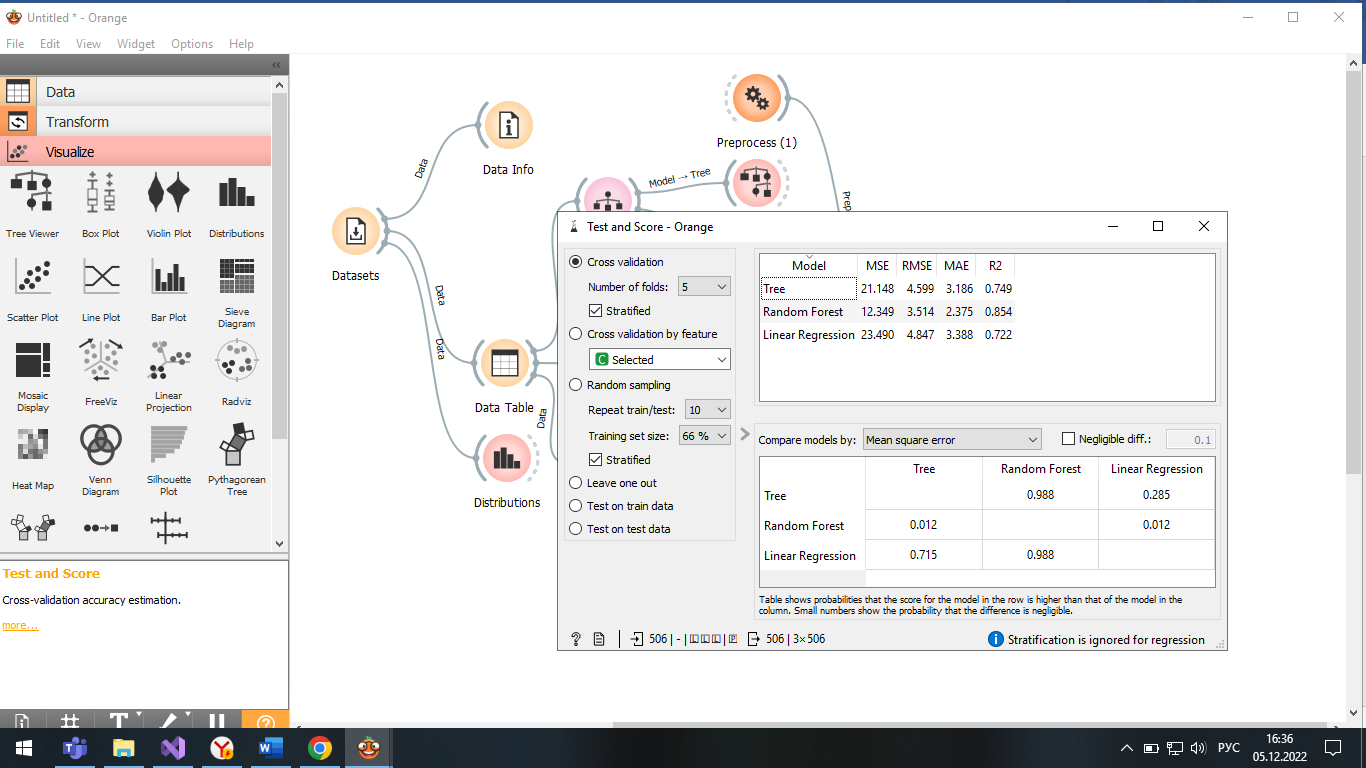


Рис. 13. Результат «Test and Score»

5) Двойным кликом на виджет «Tree Viewer» удалось посмотреть на визуализацию полученного дерева (рис.14):

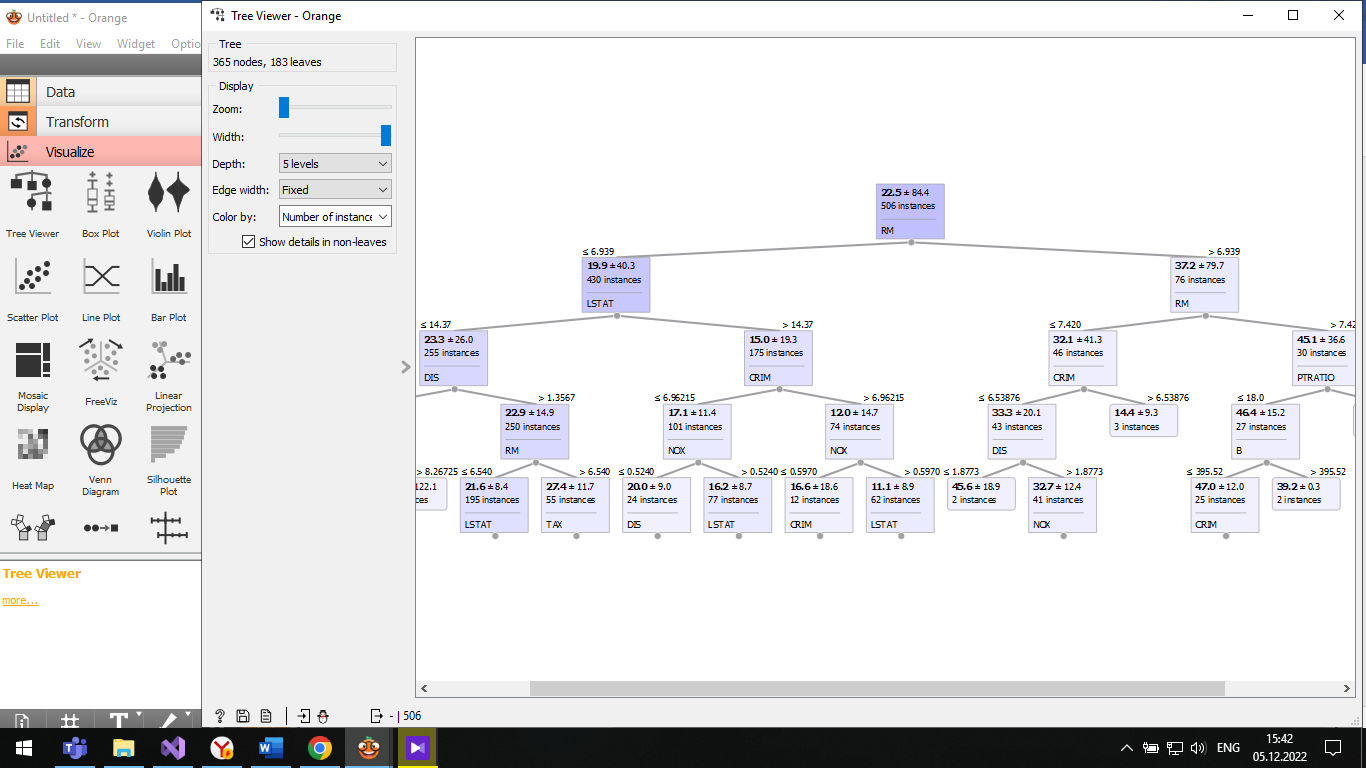


Рис. 14. Результат «Tree Viewer»

## **Результаты работы**

Программа подходит тем, кто работает с большими и малыми данными. Результаты отображаются в удобном формате как в табличным, так и визуальным способом. Данные легко (автоматически) обрабатываются, что позволяет сделать анализ данных точнее. В программе также можно делать регрессию, распределение, сортировку и многое другое.

# **Часть 2**

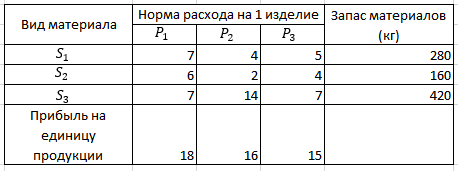
## **Решение оптимизационных задач в EXCEL, с использованием**

## **пакета анализа**

### **Задание №1**

Для изготовления трёх видов изделий Р1, Р2, и Р3 используют три вида материалов: S1, S2, S3. Запасы материалов, технологические нормы расхода материалов на каждое изделие и цена единицы изделия приведены в таблице 1. Составить план выпуска изделий, обеспечивающих их максимальный выпуск по стоимости.

Табл.1



Составим математическую модель для решения поставленной задачи.

Обозначим переменные:

x1 – объем произведенной продукции вида Р1;

x2 – объем произведенной продукции вида Р2;

x3 – объем произведенной продукции вида Р3;

Поскольку производство продукции ограничено имеющимися в распоряжении предприятия ресурсами и спросом на данную продукцию, а также учитывая, что объем изготовляемой продукции не может быть отрицательным, должны выполняться следующие неравенства:

Прибыль от реализации продукции составит:Р=

Создадим на рабочем листе таблицу для ввода исходных данных (табл.2):

Табл.2



1) За переменные возьмем числа с оптимальным решением, значение этих ячеек будет получено в результате решения задачи;

2) Впишем целевую функцию ;

3) Напишем левые части ограничения:

;

;

, где за возьмем значения переменных;

Напишем правые части ограничения:

.

Откроем «**Поиск решения»** и запишем значения (рис. 15).

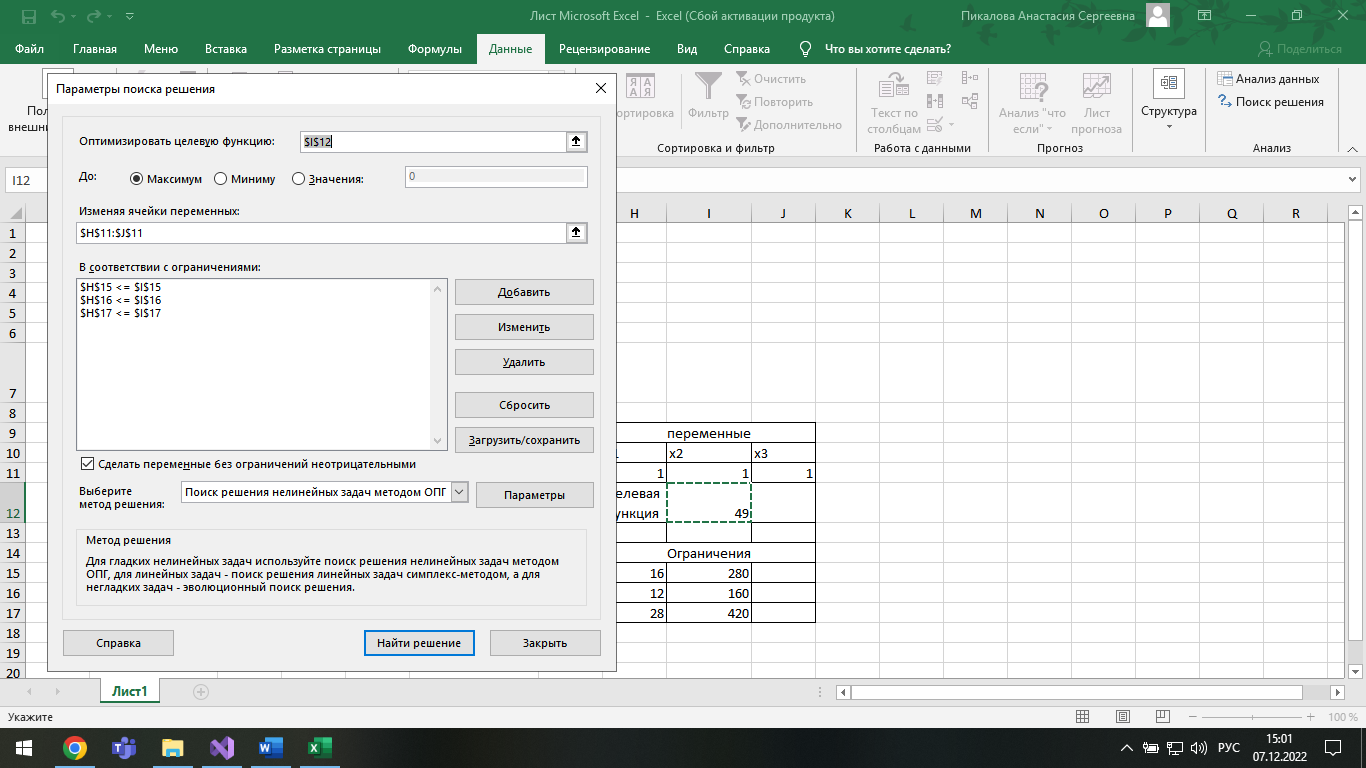


Рис.15. Параметры поиска решения

После выбора кнопки «Найти» решение отобразится окно «Результаты

поиска решения». Для сохранения полученного решения необходимо установить переключатель. Сохранить найденное решение и нажать кнопку ОК. После чего на рабочем листе отобразится решение задачи (рис. 16-17).

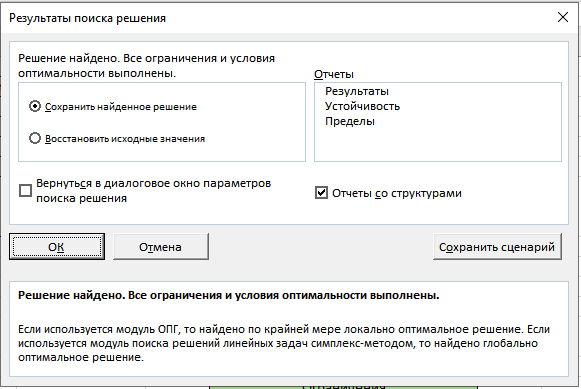


Рис.16. Результаты поиска решения

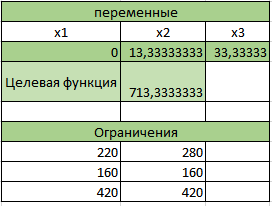


Рис.17. Решение задачи

Таким образом, максимальная прибыль при реализации продукции будет получена в размере 713,33 д. е. при следующем плане производства:

13,33 – объем продукции типа ;

33,33 – объем продукции типа .

Кроме вставки оптимальных значений в изменяемые ячейки, Поиск решения позволяет представлять результаты в виде трех отчетов: Результаты, Устойчивость и Пределы.

Отчет по результатам содержит три таблицы: в первой приведены сведения о целевой функции до начала вычисления и окончательное значение, во второй – значения искомых переменных: исходные и полученные в результате решения задачи, в третьей – результаты оптимального решения для ограничений (рис. 18).

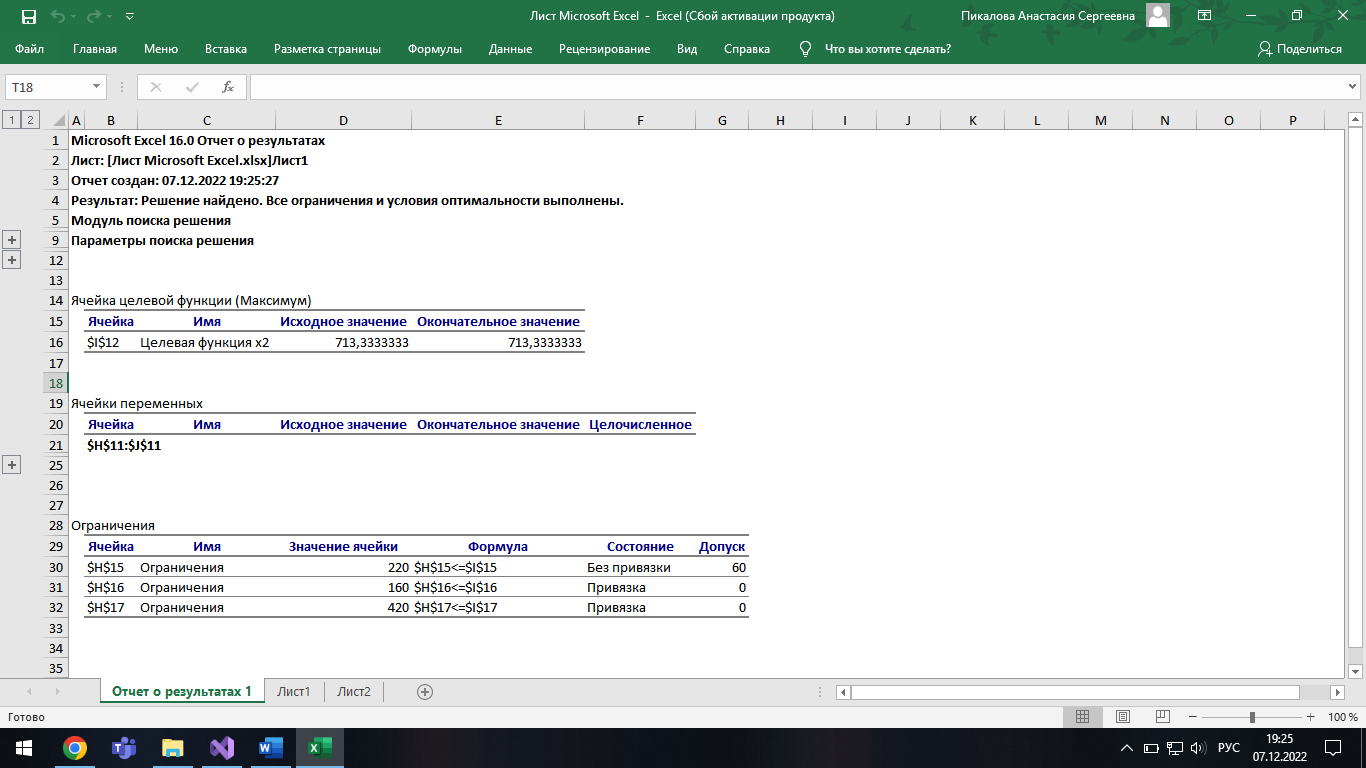


Рис.18. Отчет по результатам

### **Задание №2**

Требуется найти минимальную по стоимости смесь сырья для изготовления пищевых концентратов, которые должны содержать питательные вещества (П). Эти вещества содержатся в сырье (М) в различных сочетаниях. Содержание питательных веществ в сырье и готовом продукте, а также цена на каждый вид сырья показаны в таблице 3.

Табл.3



Составим математическую модель для решения поставленной задачи.

Обозначим переменные:

x1 – объем сырья вида М1;

x2 – объем сырья вида М2;

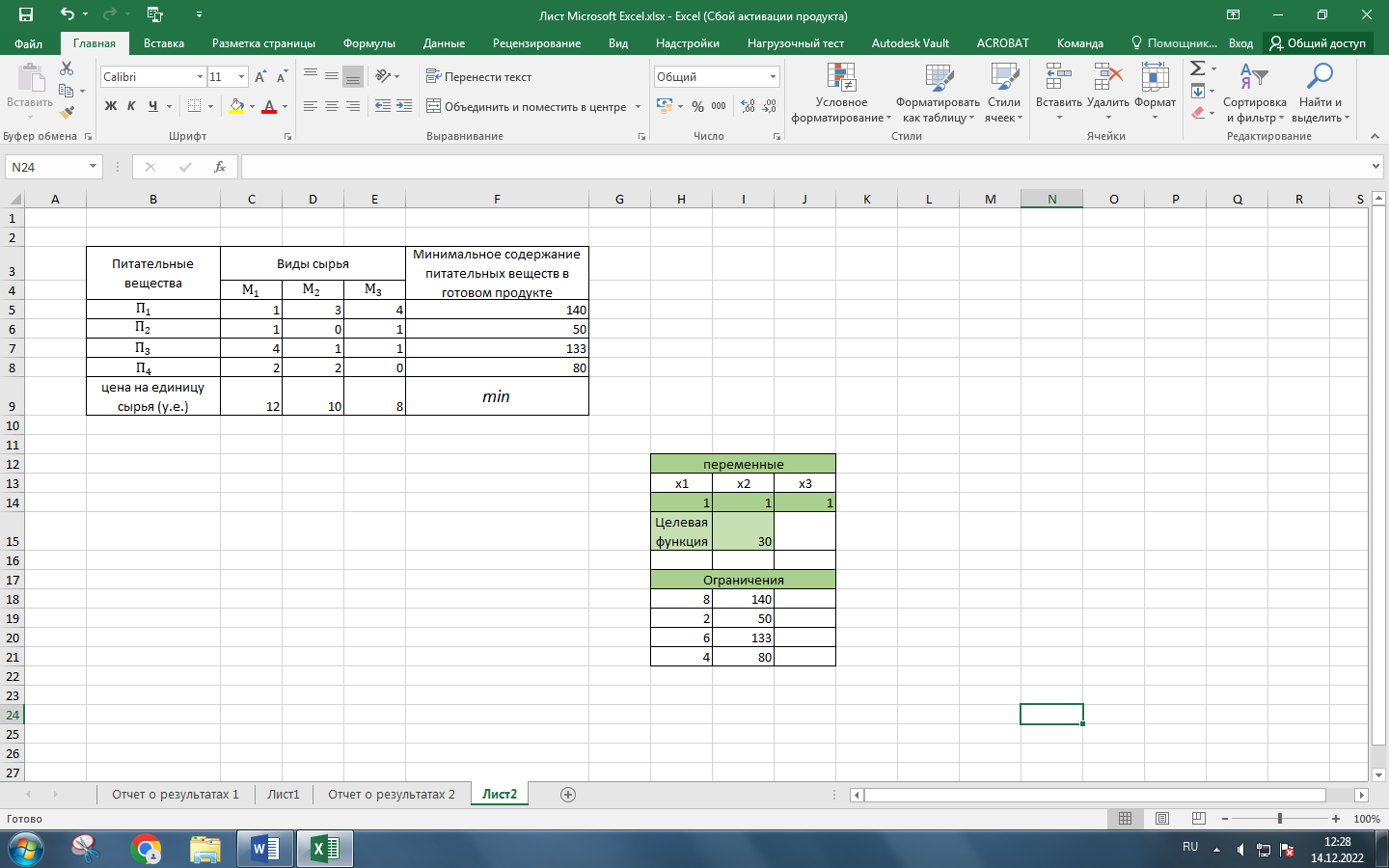
x3 – объем сырья вида М3;

Поскольку производство сырья ограничено минимальным содержанием питательного вещества и ценой на сырье, а также учитывая, что объем изготовляемой сырья не может быть отрицательным, должны выполняться следующие неравенства:

Прибыль от реализации продукции составит:Р=

Создадим на рабочем листе таблицу для ввода исходных данных (табл.4):

Табл.4



1) За переменные возьмем числа с оптимальным решением, значение этих ячеек будет получено в результате решения задачи;

2) Впишем целевую функцию ;

3) Напишем левые части ограничения:

;

;

;

, где за возьмем значения переменных;

Напишем правые части ограничения:

.

Откроем «**Поиск решения»** и запишем значения (рис. 19).

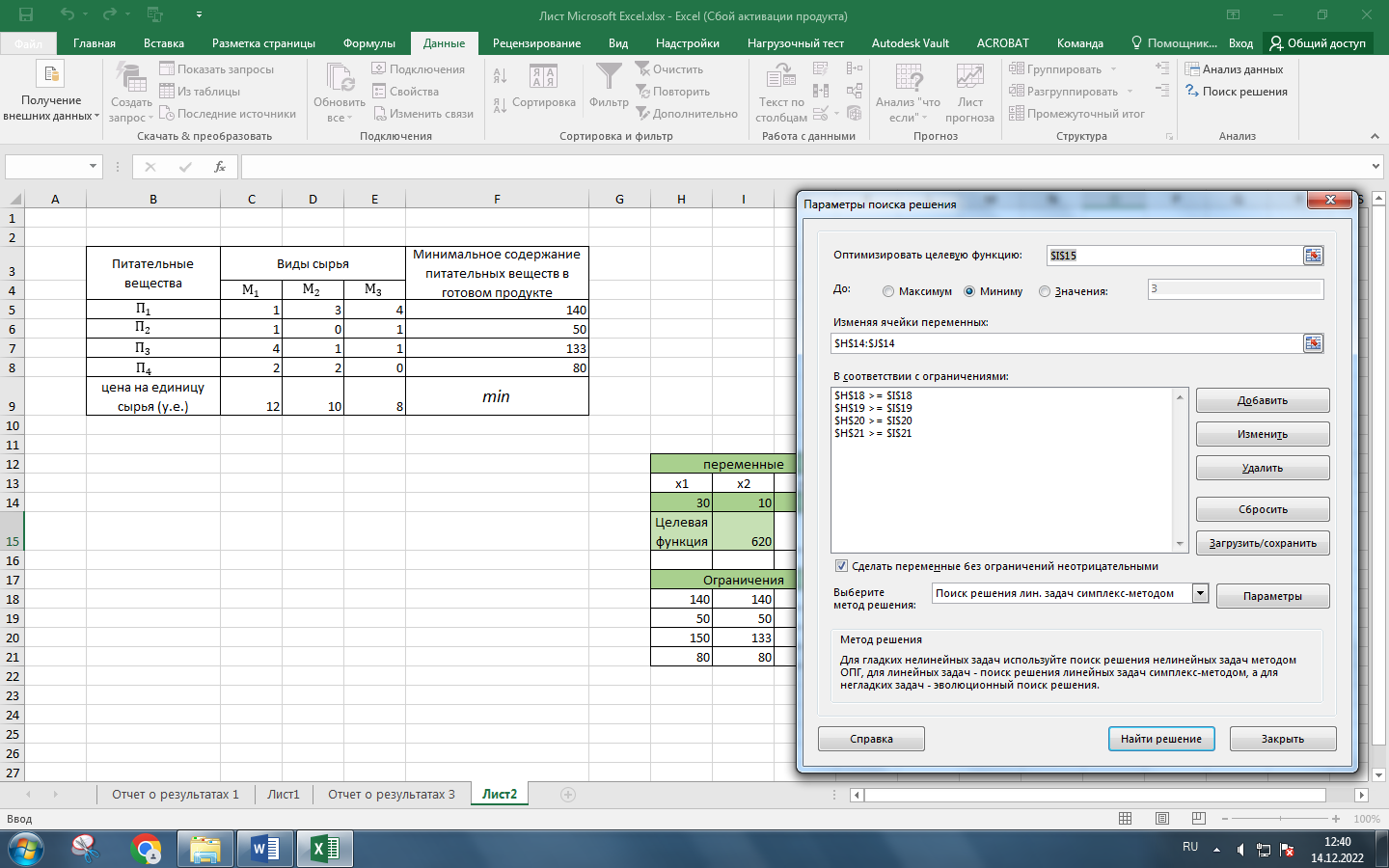


Рис.19. Параметры поиска решения

После выбора кнопки «Найти» решение отобразится окно «Результаты

поиска решения». Для сохранения полученного решения необходимо установить переключатель. Сохранить найденное решение и нажать кнопку ОК. После чего на рабочем листе отобразится решение задачи (рис. 16-17).

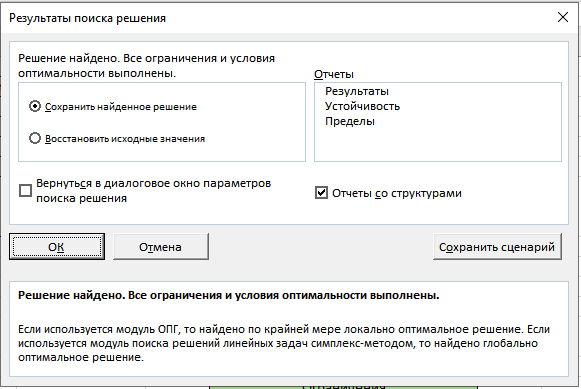


Рис.16. Результаты поиска решения

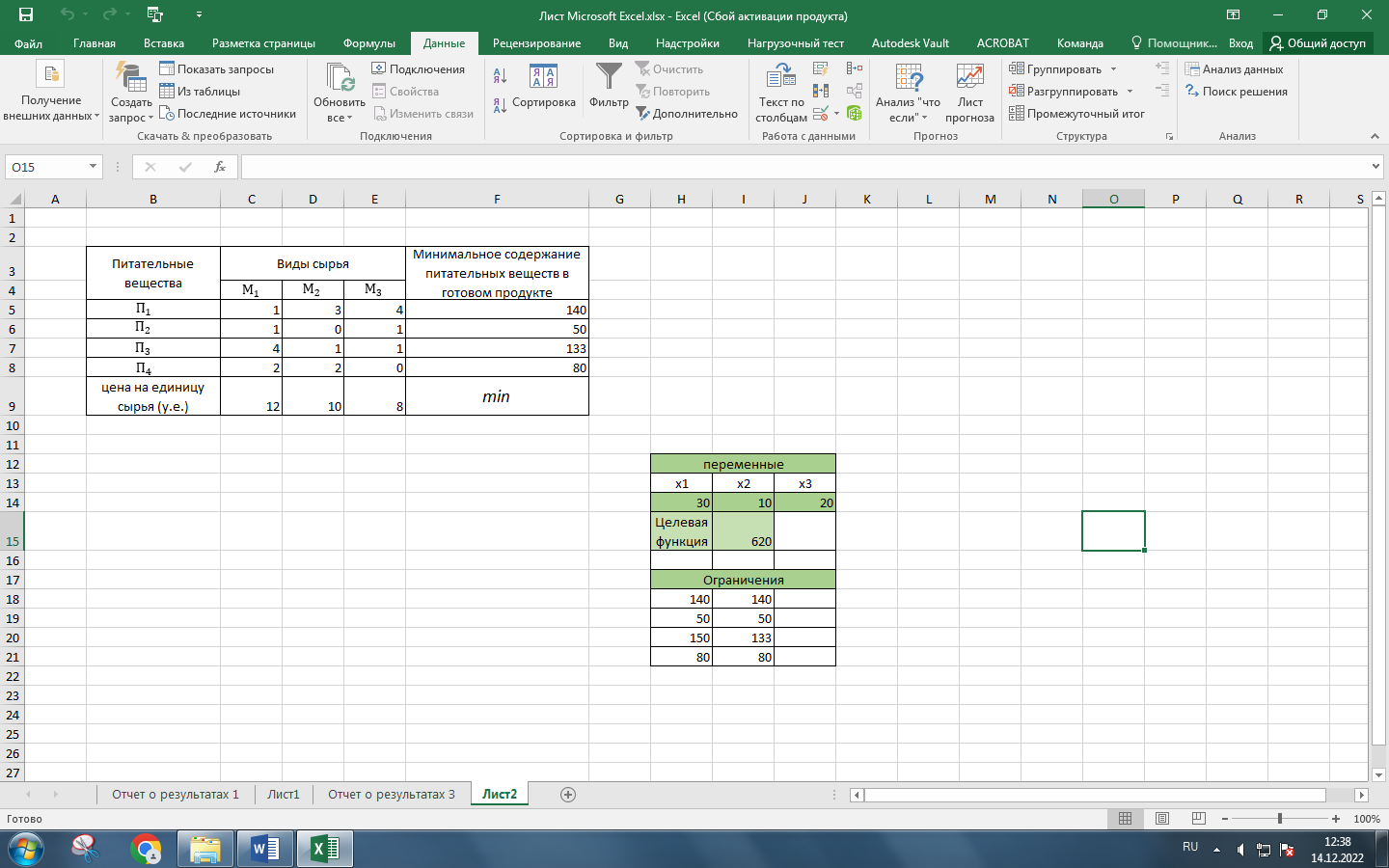


Рис.17. Решение задачи

Таким образом, минимальная прибыль при реализации продукции будет получена в размере 620 д. е. при следующем плане производства:

30 – объем продукции типа ;

10 – объем продукции типа ;

20 – объем продукции типа .

Кроме вставки оптимальных значений в изменяемые ячейки, Поиск решения позволяет представлять результаты в виде трех отчетов: Результаты, Устойчивость и Пределы.

Отчет по результатам содержит три таблицы: в первой приведены сведения о целевой функции до начала вычисления и окончательное значение, во второй – значения искомых переменных: исходные и полученные в результате решения задачи, в третьей – результаты оптимального решения для ограничений (рис. 18).

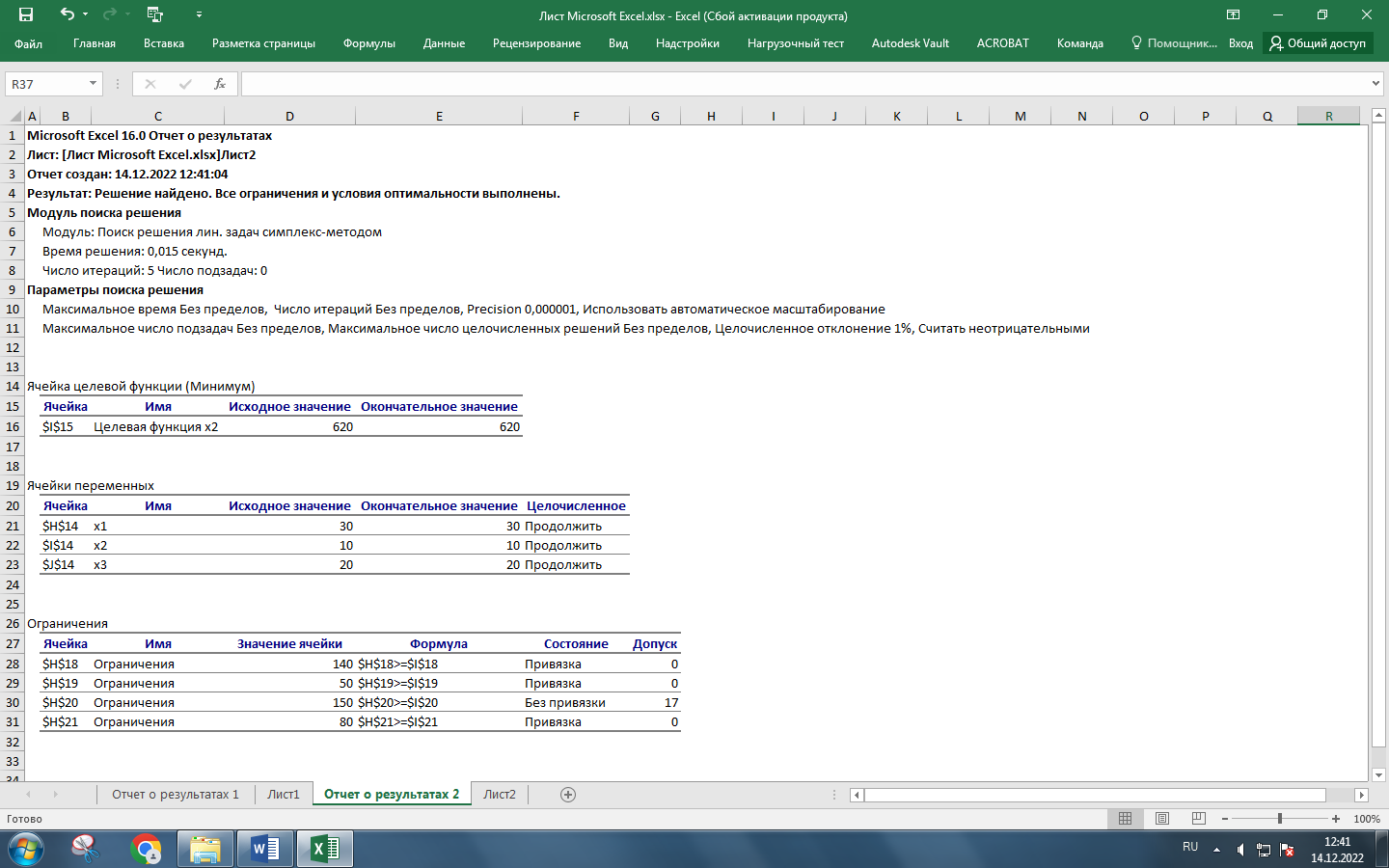


Рис.18. Отчет по результатам

# **Вывод:**

В ходе работы были рассмотрены инструменты для работы с данными (KNIME Analytics Platform, Statistical Package for the Social Sciences, SAS, STATISTICA 6.0, Minitab, Orange3, Python, R + R-Studio). Многие из программ имеют бесплатную версию или имеют возможность приобретения студенческой лицензии, есть программы, в которых не обязательны знания программирования, но большинство требуют базовых знаний. Сделана работа в программе Orange, и сделаны выводы по работе и сделана работа в программе Excel, где было изучено решение оптимизационных задач.

В процессе выполнения заданий по технологической практике частично были освоены компетенции:

* Способна использовать современные информационные технологии и программные средства, в том числе отечественного производства, при решении задач профессиональной деятельности (ОПК-2);
* Способна решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности (ОПК-3).

## **Литература**

1. Официальный сайт r-studio: <https://www.r-studio.com/ru/>;
2. Что такое Python: чем он хорош, где пригодится и как его выучить: <https://netology.ru/blog/python>
3. Официальный сайт Orange: <https://orangedatamining.com/>
4. Страница Википедии о Minitab : <https://ru.wikipedia.org/wiki/Minitab>
5. Статистический анализ данных. STATISTICA 6.0: <http://statlab.kubsu.ru>
6. SAS Visual Analytics : <https://www.tadviser.ru/index.php/%D0%9F%D1%80%D0%BE%D0%B4%D1%83%D0%BA%D1%82:SAS_Visual_Analytics>
7. Программа для статистической обработки данных SPSS Statistics: <http://ecoshop.biz/?action=748M>
8. Всё о KNIME: [https://analytikaplus.ru/vsyo-o-tableau-opisanie-preimushhestva-produkty-resheniya-prakticheskie-kejsy/#:~:text=KNIME%20Analytics%20Platform%20%E2%80%94%20%D1%8D%D1%82%D0%BE%20%D0%B1%D0%B5%D1%81%D0%BF%D0%BB%D0%B0%D1%82%D0%BD%D0%B0%D1%8F,%D0%BF%D1%80%D0%BE%D1%86%D0%B5%D1%81%D1%81%20%D0%B8%20%D1%84%D1%83%D0%BD%D0%BA%D1%86%D0%B8%D0%B8%20%D0%BE%D1%80%D0%B3%D0%B0%D0%BD%D0%B8%D0%B7%D0%B0%D1%86%D0%B8%D0%B8%20%D1%83%D0%B7%D0%BB%D0%BE%D0%B2](https://analytikaplus.ru/vsyo-o-tableau-opisanie-preimushhestva-produkty-resheniya-prakticheskie-kejsy/#:~:text=KNIME%20Analytics%20Platform%20%E2%80%94%20%D1%8D%D1%82%D0%BE%20%D0%B1%D0%B5%D1%81%D0%BF%D0%BB%D0%B0%D1%82%D0%BD%D0%B0%D1%8F,%D0%BF%D1%80%D0%BE%D1%86%D0%B5%D1%81%D1%81%20%D0%B8%20%D1%84%D1%83%D0%BD%D0%BA%D1%86%D0%B8%D0%B8%20%D0%BE%D1%80%D0%B3%D0%B).