Уровень 1

Задание 1

1. RapidMiner - бесплатный инструмент для анализа данных и машинного обучения, включая функционал text mining.

2. Orange - бесплатное программное обеспечение для визуализации и анализа данных, включая возможности text mining.

3. KNIME - бесплатный инструмент для создания и выполнения аналитических рабочих процессов, включая модули для text mining.

4.GATE (General Architecture for Text Engineering) - открытая платформа для разработки и интеграции инструментов text mining.

5. NLTK (Natural Language Toolkit) - библиотека Python для обработки естественного языка, включая возможности text mining.

6. SpaCy - еще одна библиотека Python для обработки естественного языка, предоставляющая инструменты для text mining.

7. TextBlob - простая библиотека Python для анализа текста, включая функции text mining.

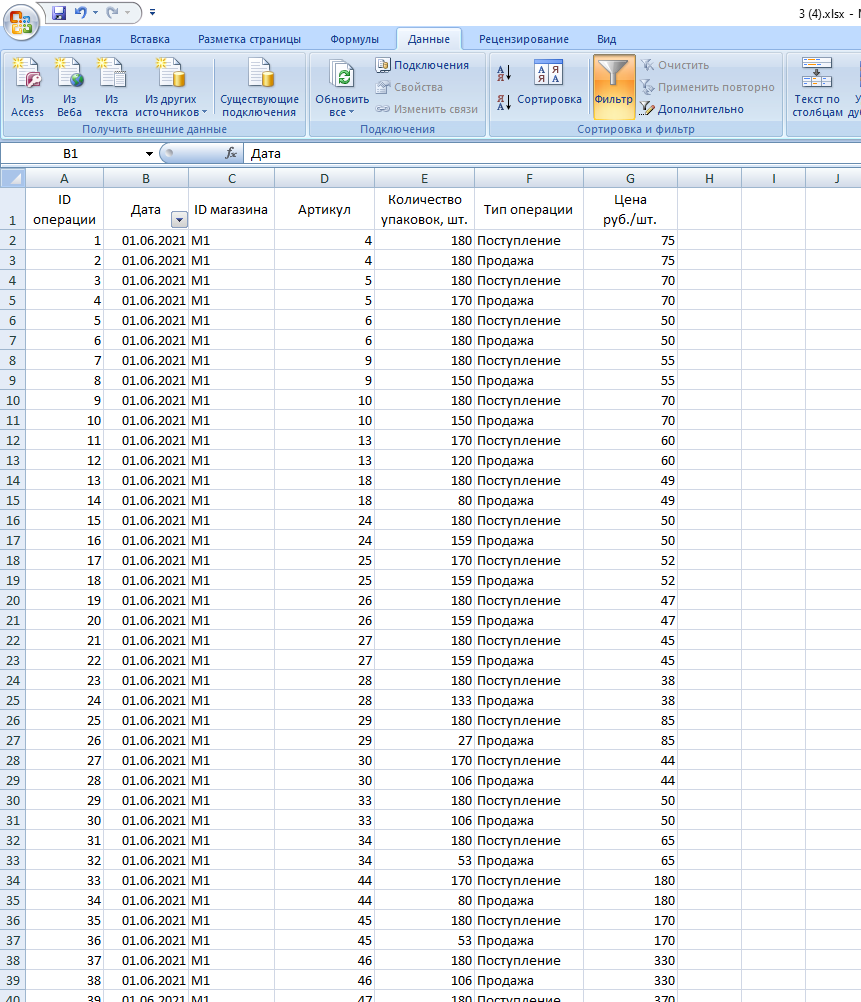
8. Stanford NLP - набор инструментов для обработки естественного языка, включая модули для text mining.

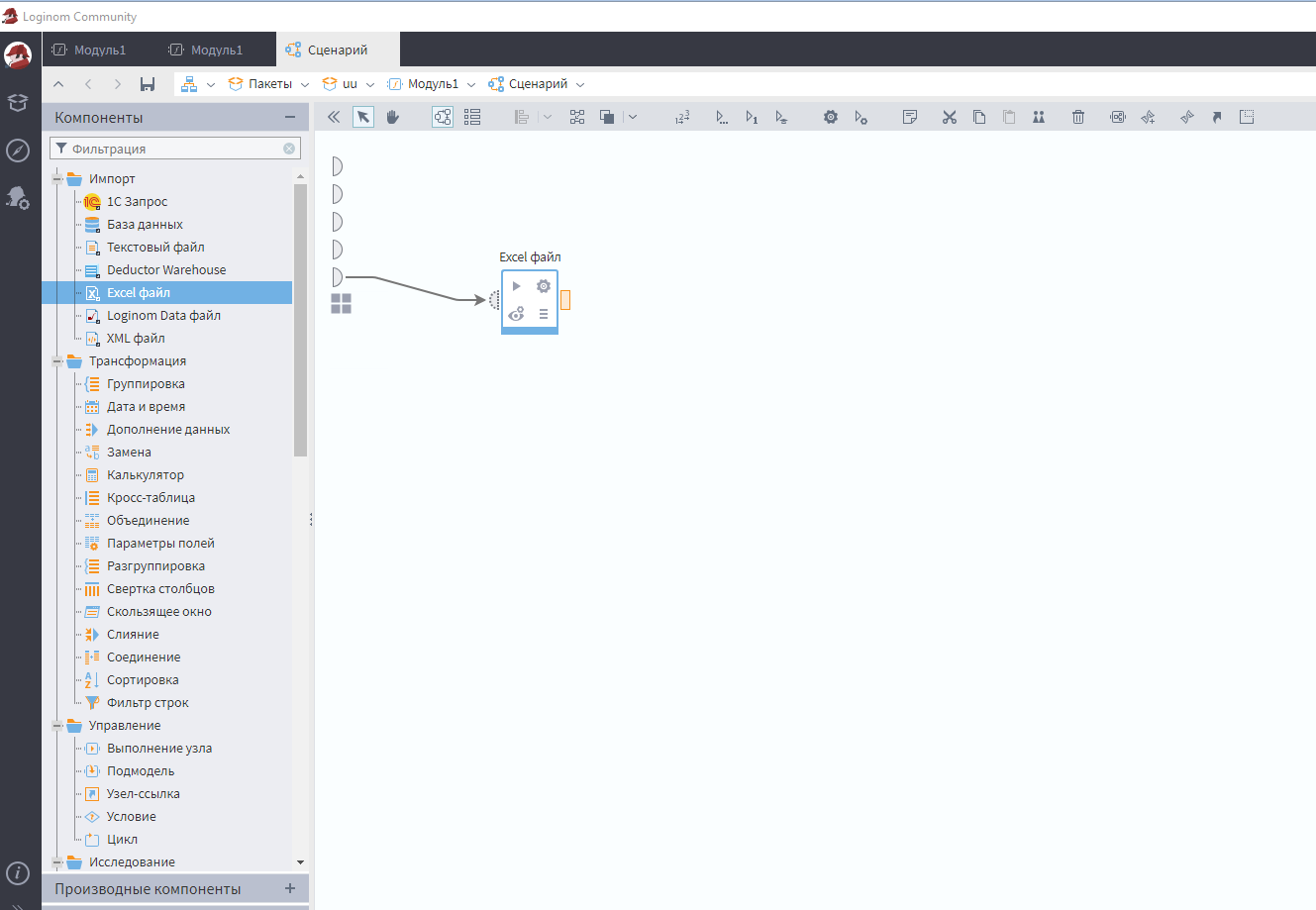
9. IBM Watson Natural Language Understanding - платформа IBM для анализа текста с использованием искусственного интеллекта.

10. Linguamatics - коммерческое решение для анализа текста и извлечения информации из него, включая text mining.

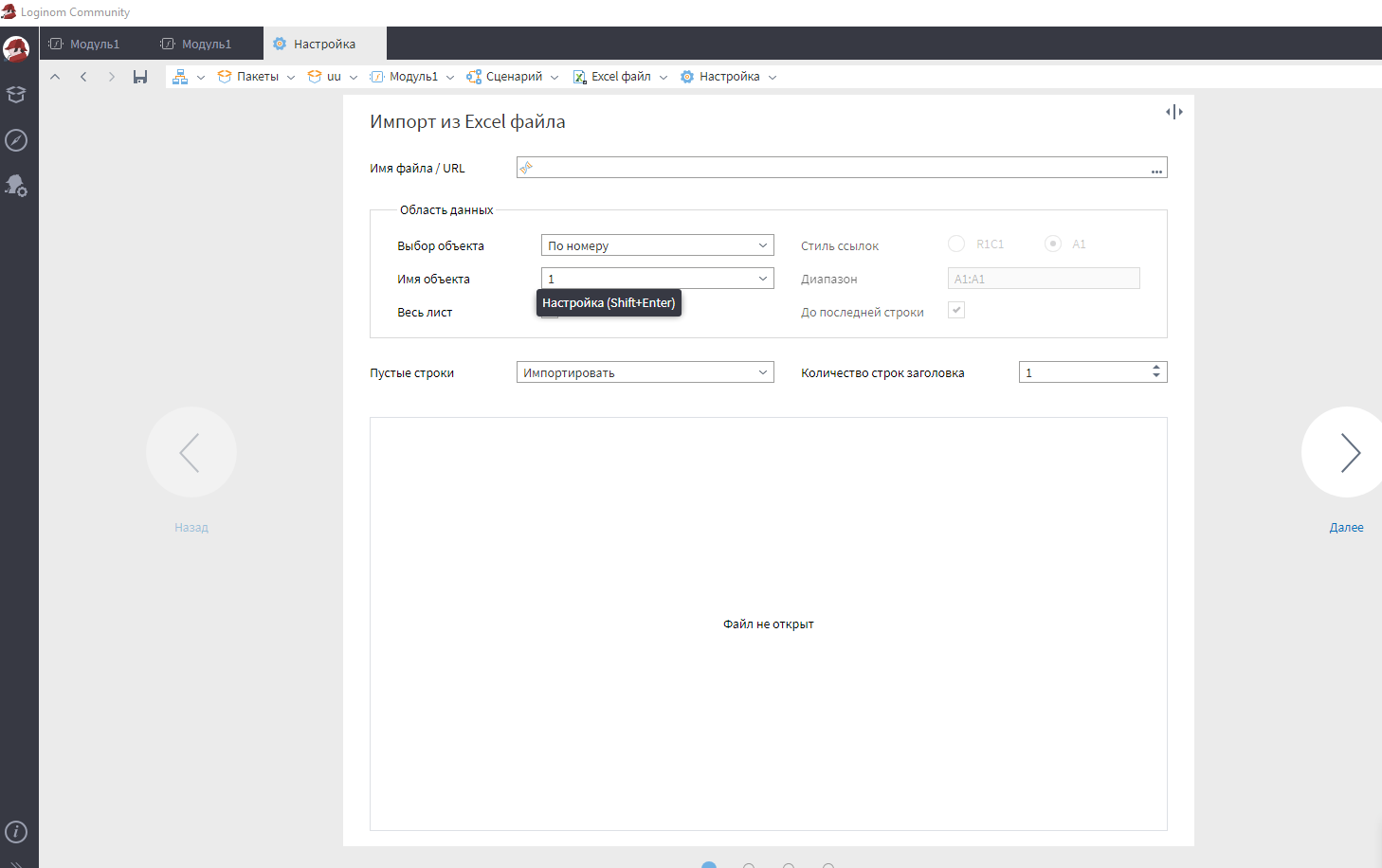
**Loginom Community**

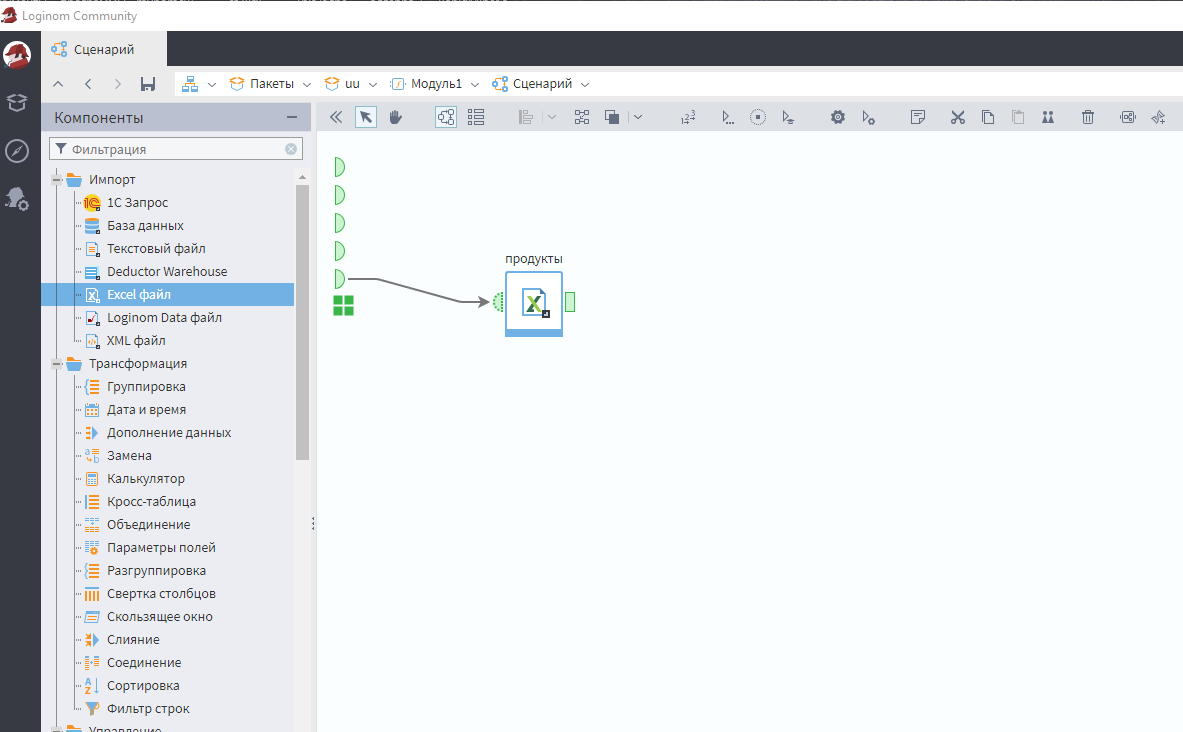
Для анализа возьмем файл в формате Exсel (база данных магазинов)



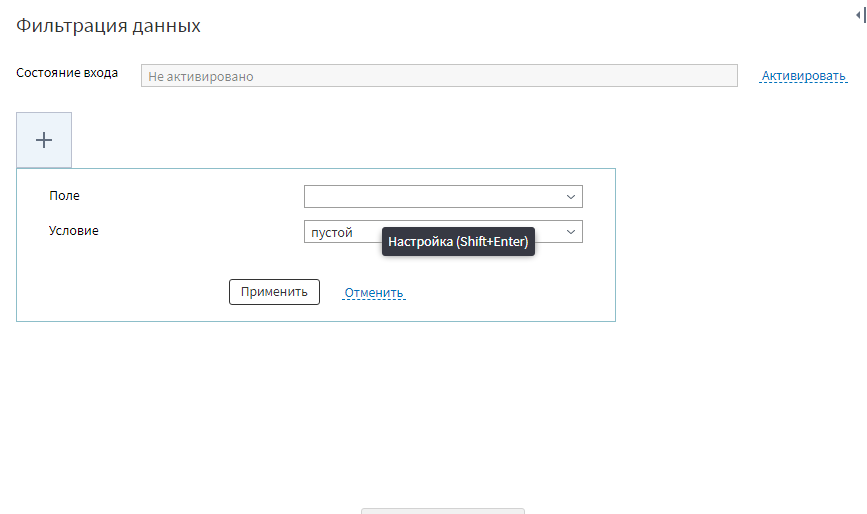
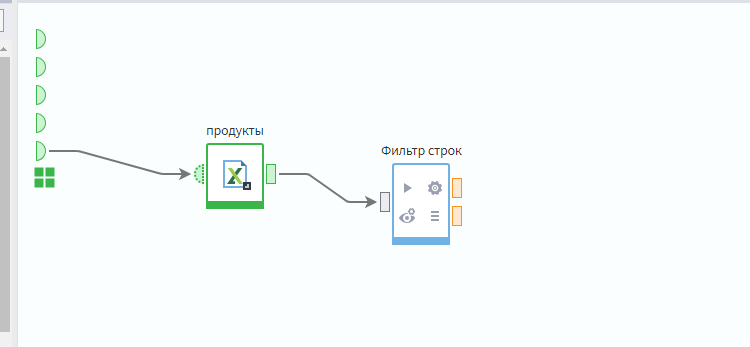


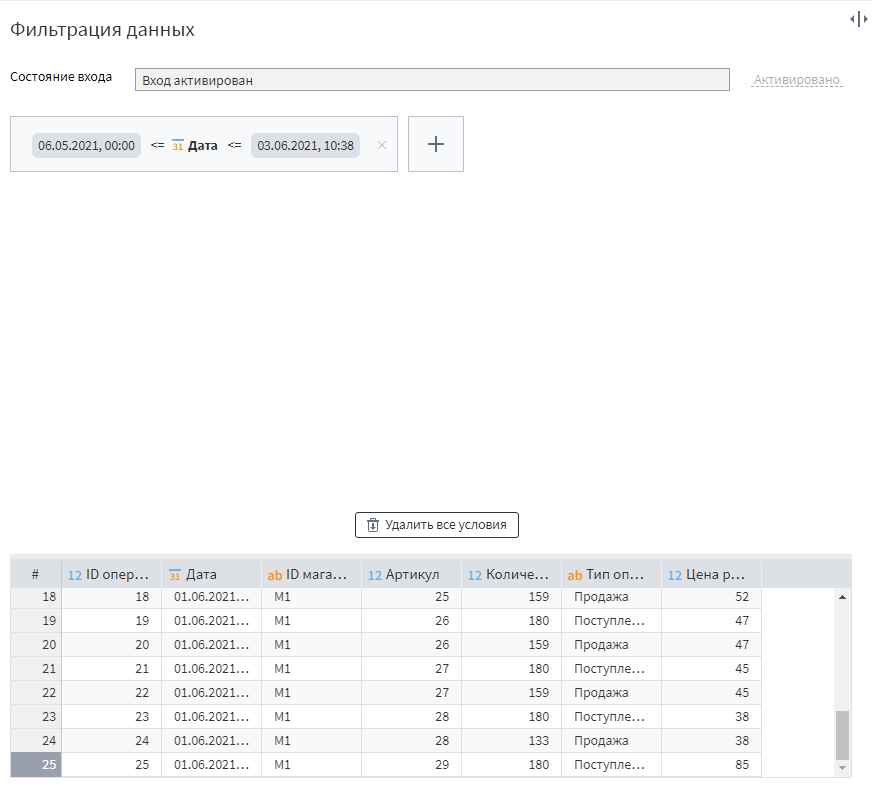
Создаем файл в рабочем поле, нажимаем значок найстроек



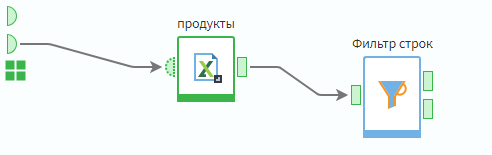


Дальше необходимо отфильтровать данные

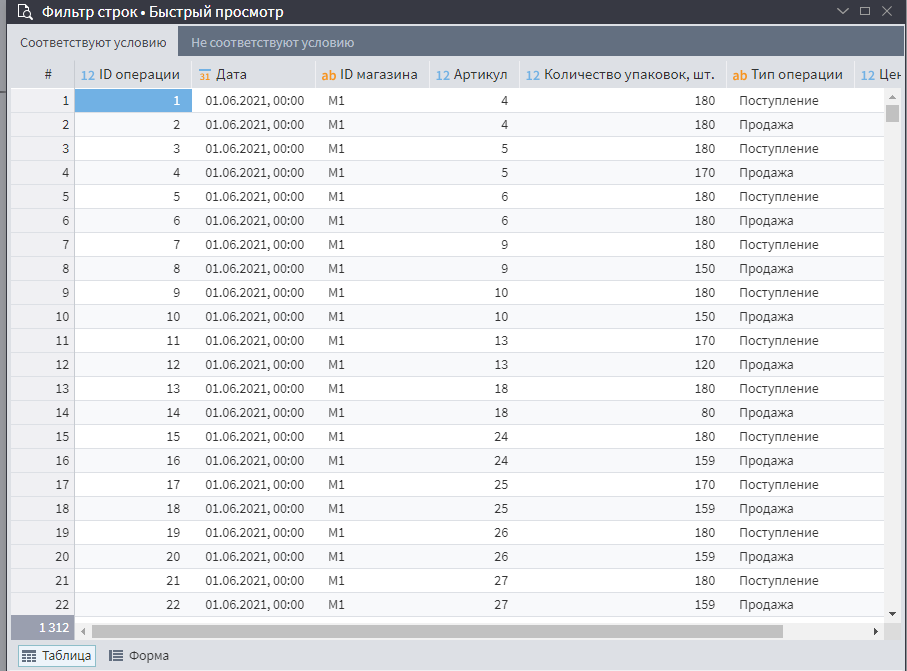


Фильтруем по дате

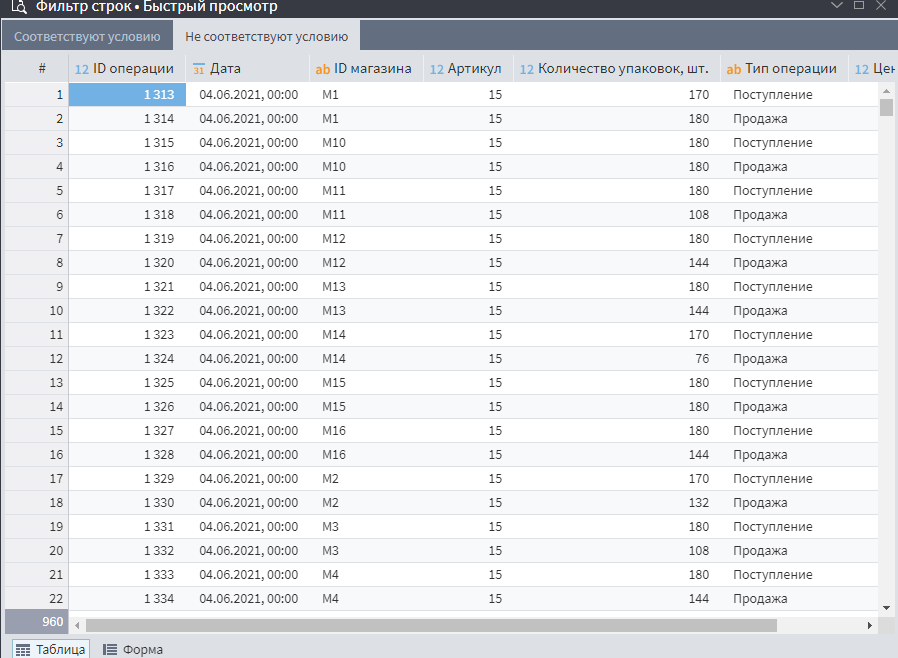
Применяем фильтр



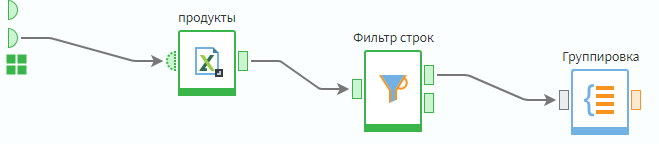
Смотрим, что получилось



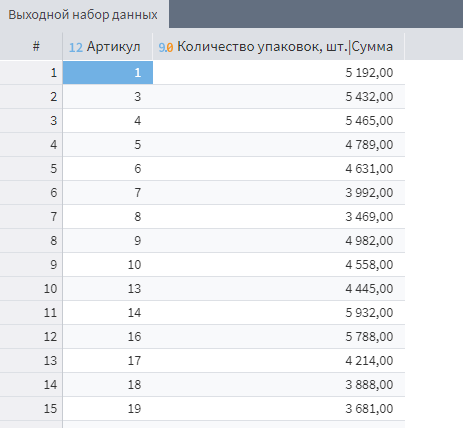
Данные отфильтровались по указанному временному интервалу



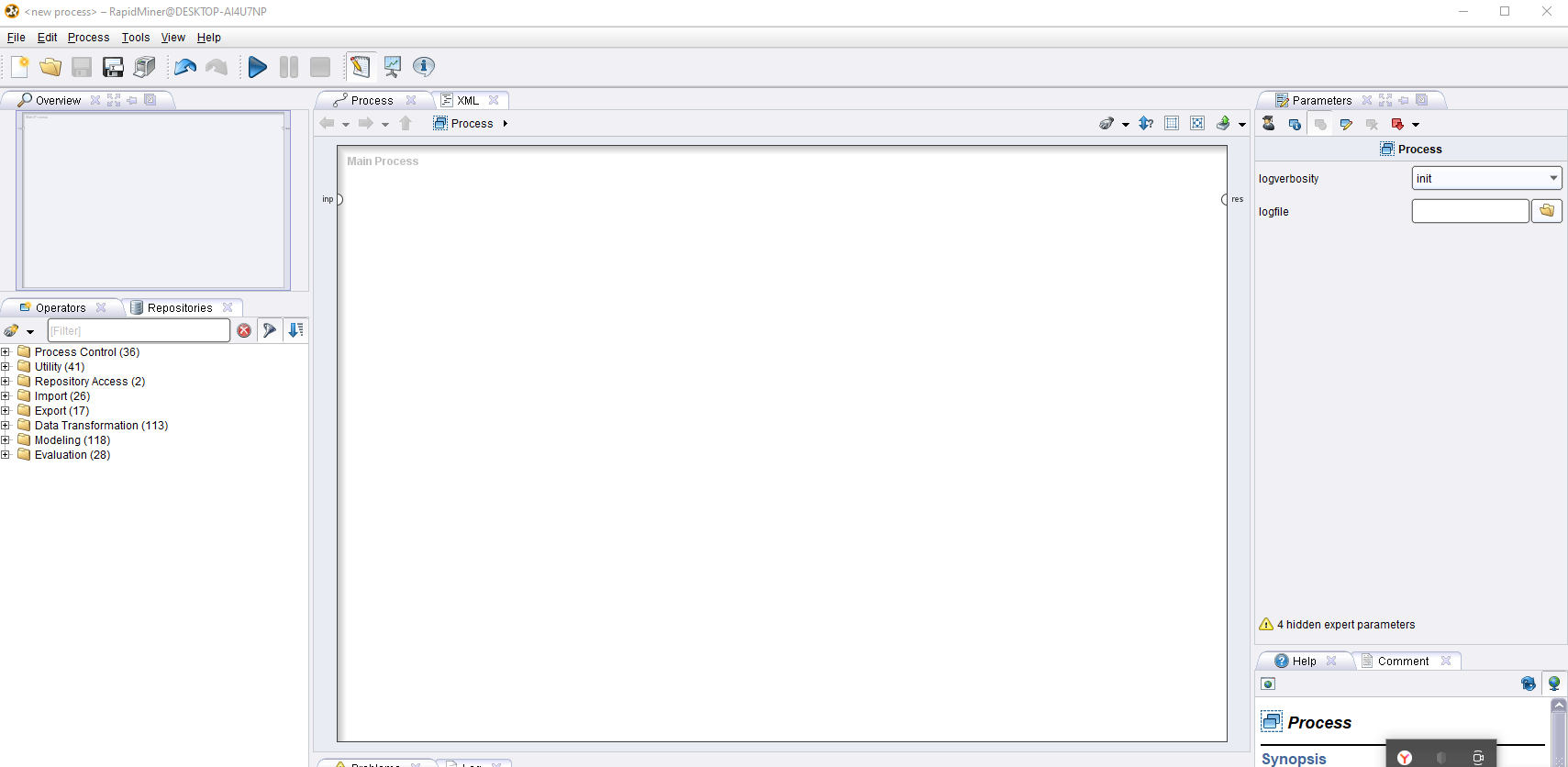
Группируем данные



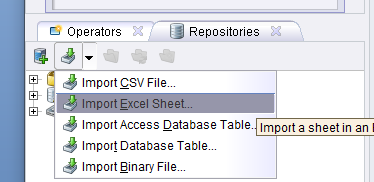
Посмотрим, на какую сумму в магазинах содержится каждый товар

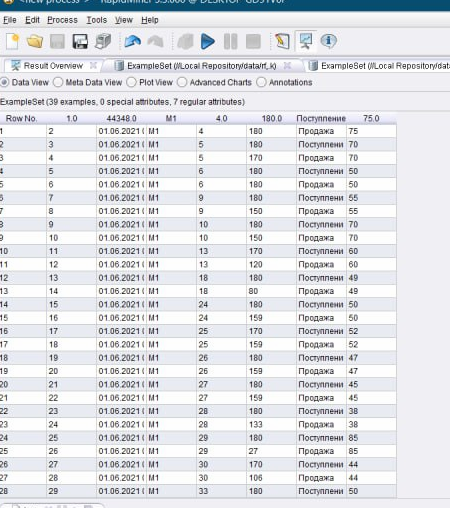


**RapidMiner 5**

****

Импортируем таблицу Excel

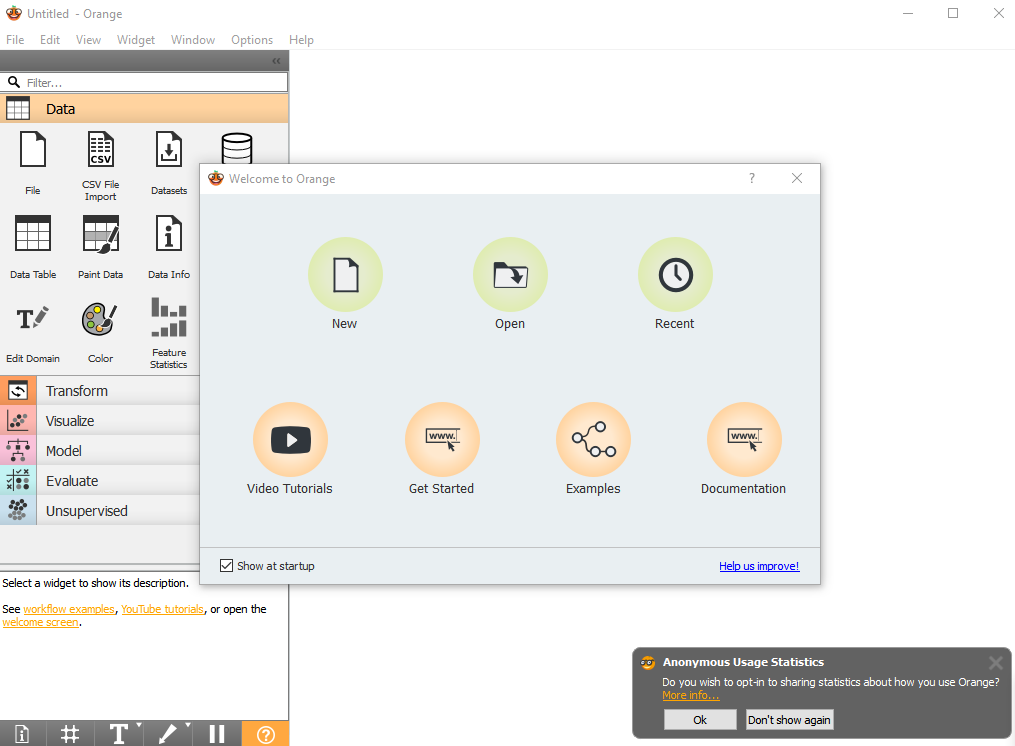
****

****

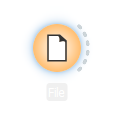
****

****

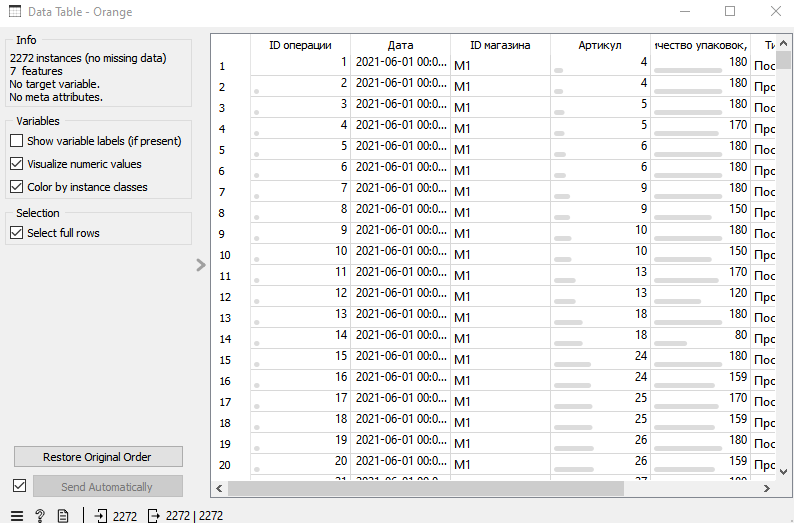
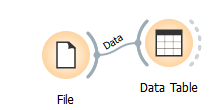
**Orange**

****

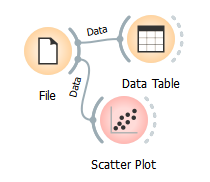
Загружаем новый файл

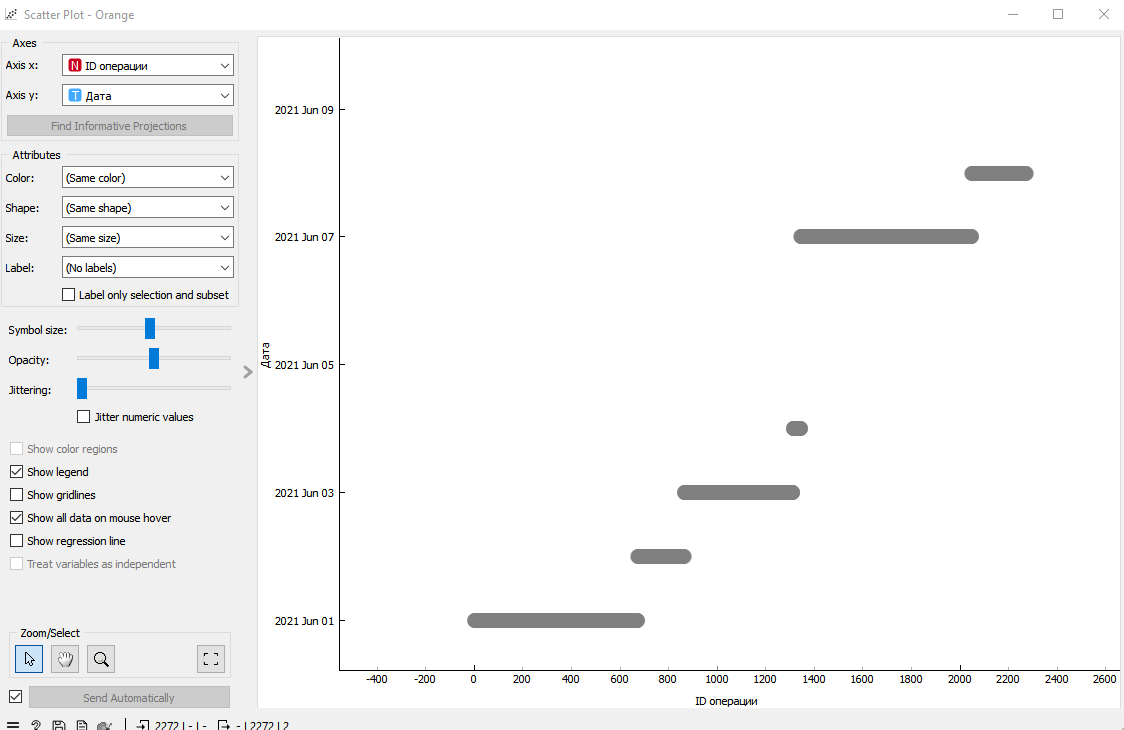
****

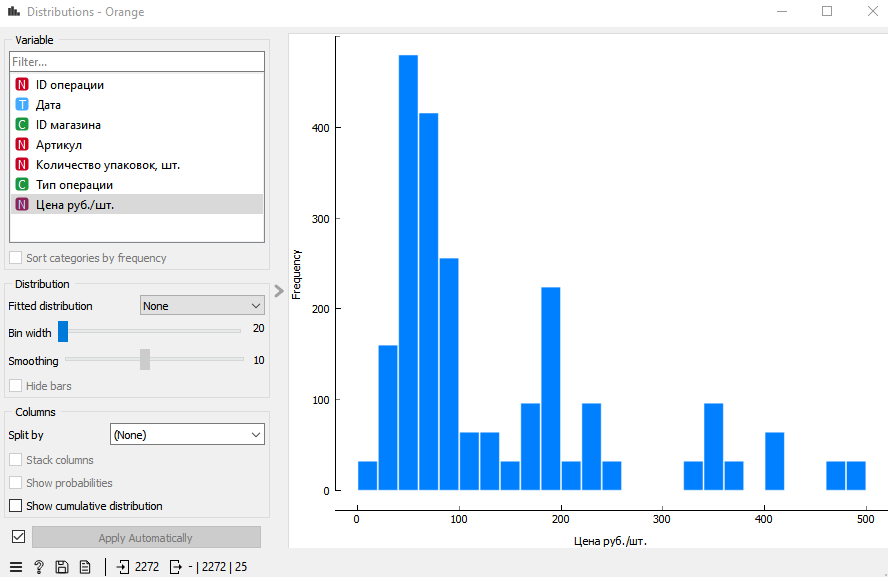
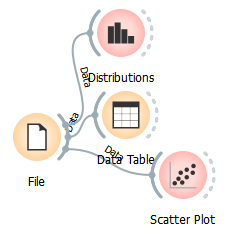
Просмотрим данные, которые мы загрузили

****

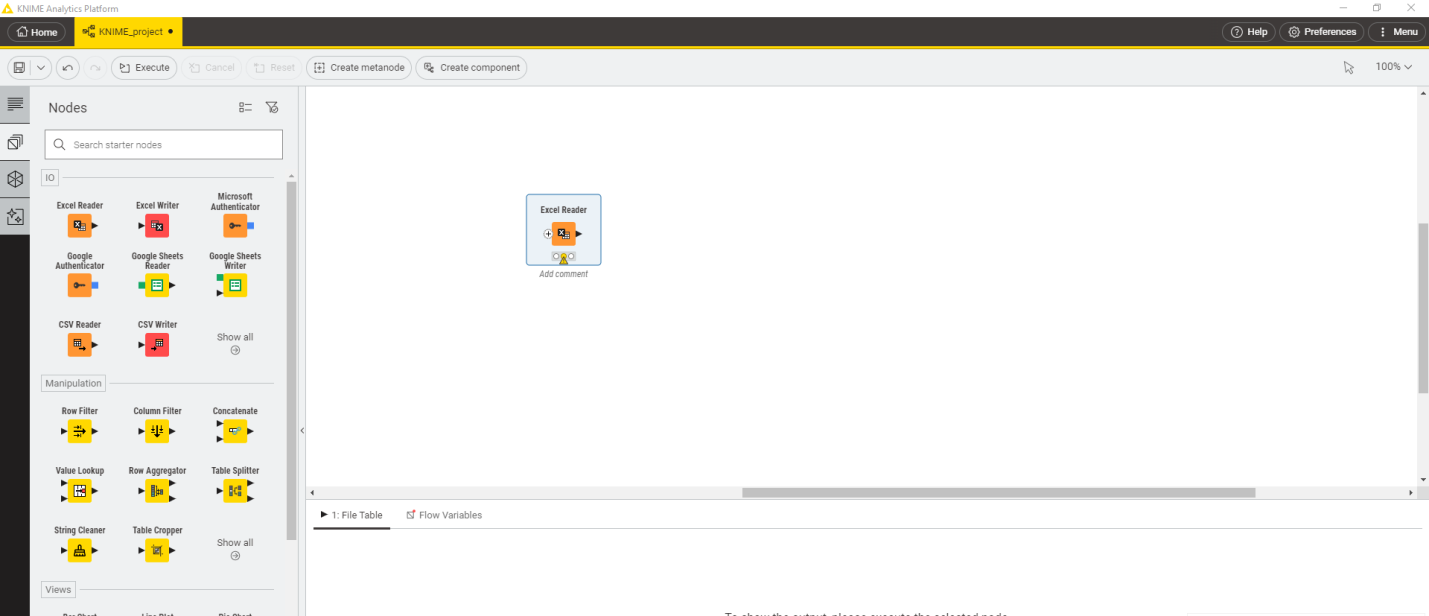
Визуализируем данные

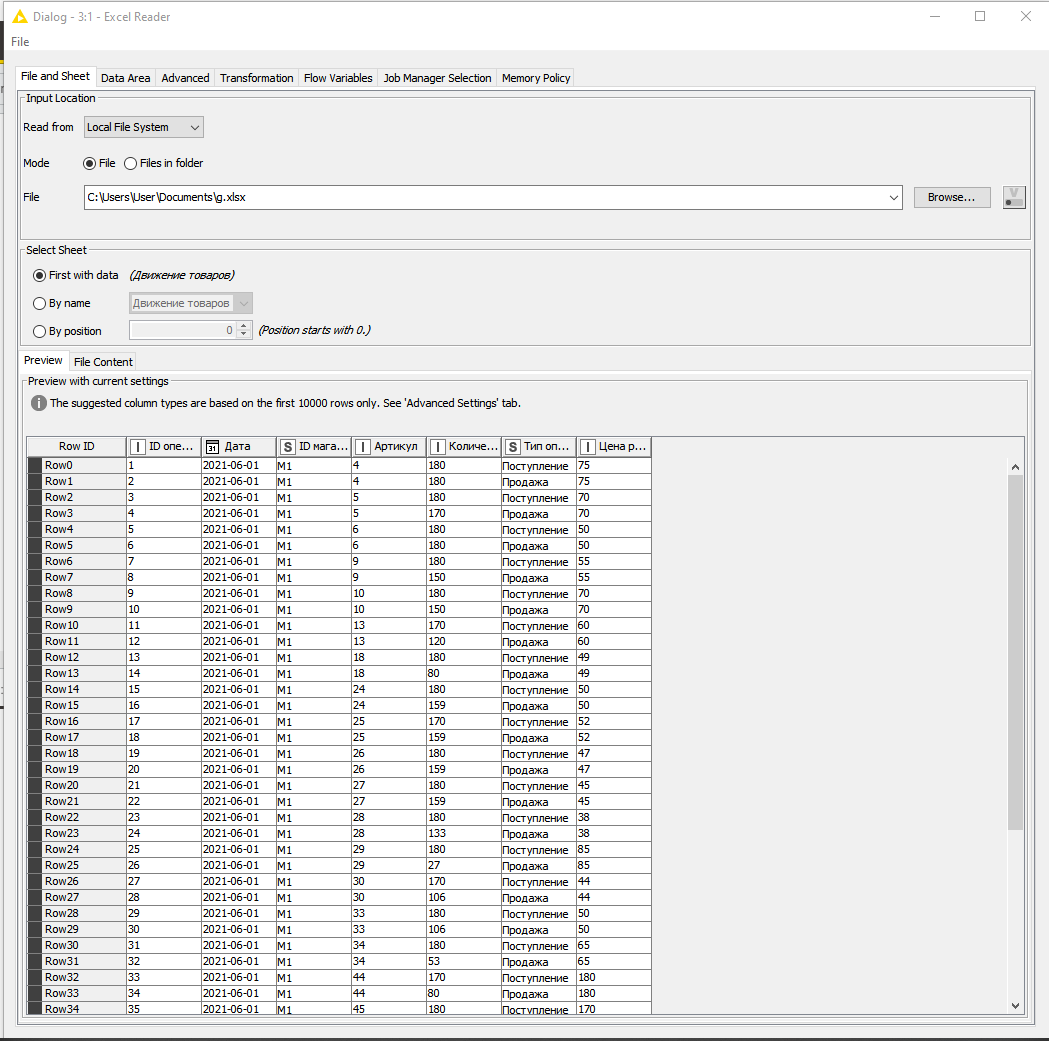
****

****

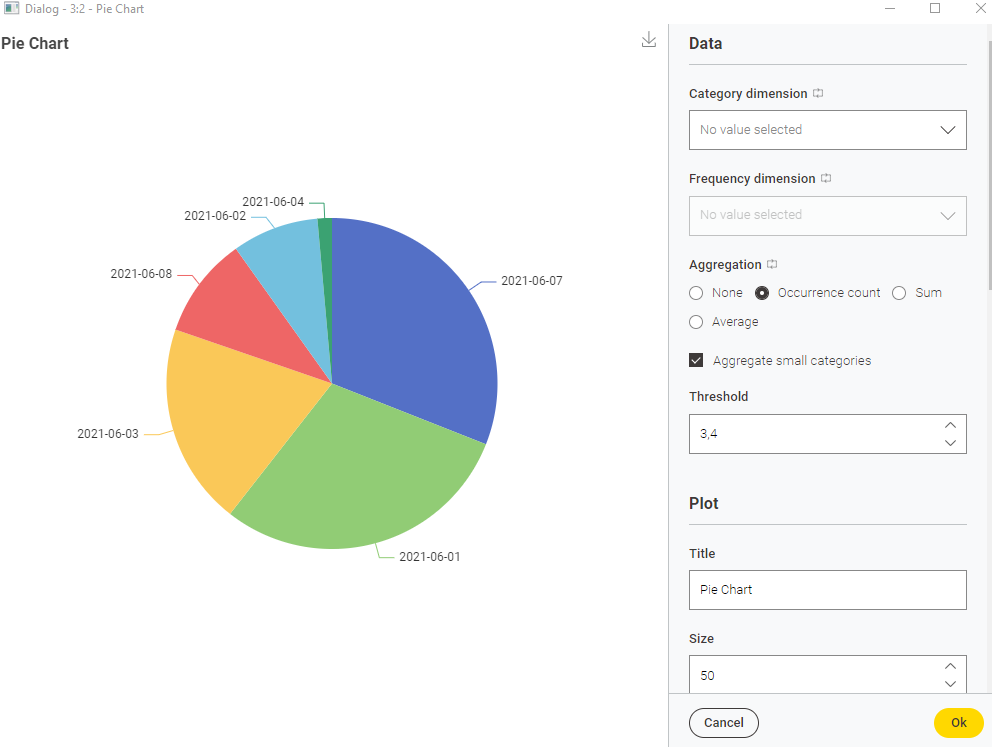
****

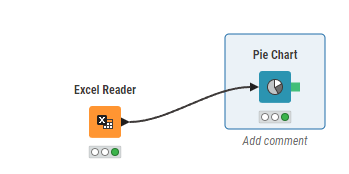
Knime

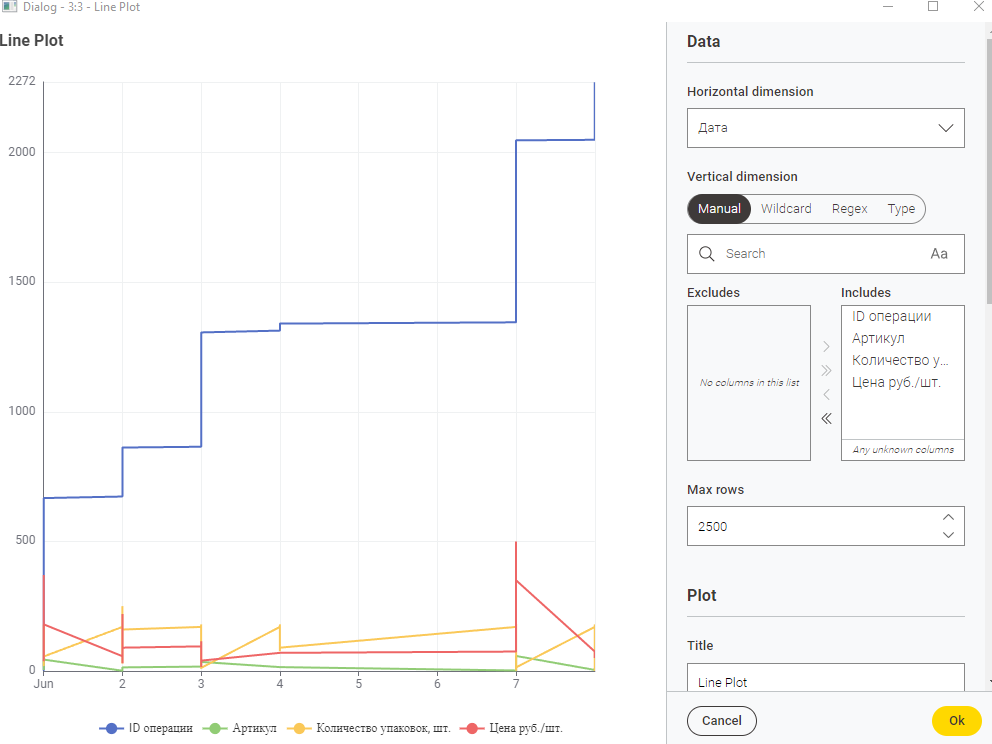
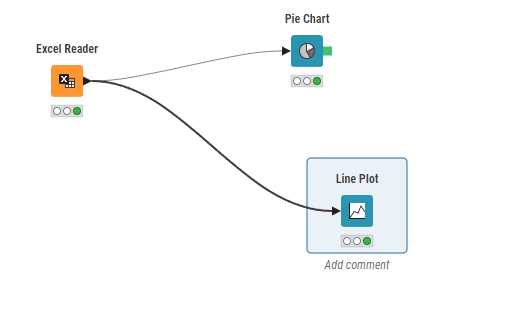
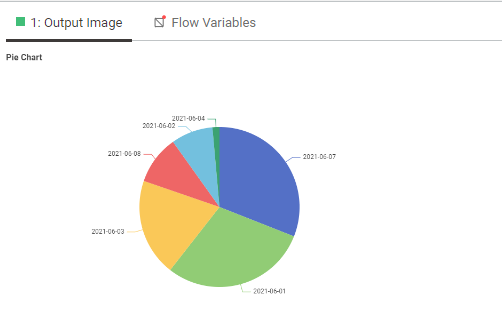


****

Визуализация данных

****

****

****

1. Loginom Community:

- Описание: Loginom Community — это инструмент для визуализации данных и создания моделей машинного обучения без необходимости программирования.

- Плюсы:

- Простой в использовании интерфейс.

- Предоставляет широкие возможности для анализа данных и построения моделей.

- Может быть бесплатным для некоторых функций.

- Минусы:

- Ограниченные возможности бесплатной версии.

- Не так мощный как некоторые другие инструменты.

2. RapidMiner:

- Описание: RapidMiner — мощный инструмент для анализа данных, предоставляющий широкий спектр функций для машинного обучения и аналитики.

- Плюсы:

- Обширные возможности для работы с данными и построения моделей.

- Поддерживает различные методы машинного обучения.

- Имеет большое сообщество пользователей.

- Минусы:

- Не всегда прост в использовании для начинающих.

- Платный доступ к расширенным функциям.

3. Orange:

- Описание: Orange — инструмент для визуального программирования, предназначенный для анализа данных и машинного обучения.

- Плюсы:

- Интуитивно понятный визуальный интерфейс.

- Широкий выбор методов анализа данных и машинного обучения.

- Бесплатное и открытое программное обеспечение.

- Минусы:

- Могут возникнуть ограничения при работе с большими объемами данных.

- Некоторые продвинутые функции доступны только в платной версии.

4. KNIME:

- Описание: KNIME — платформа с открытым исходным кодом для интеграции данных, анализа и моделирования.

- Плюсы:

- Мощные возможности интеграции данных из различных источников.

- Поддерживает широкий спектр методов машинного обучения и анализа данных.

- Большое сообщество пользователей и множество плагинов.

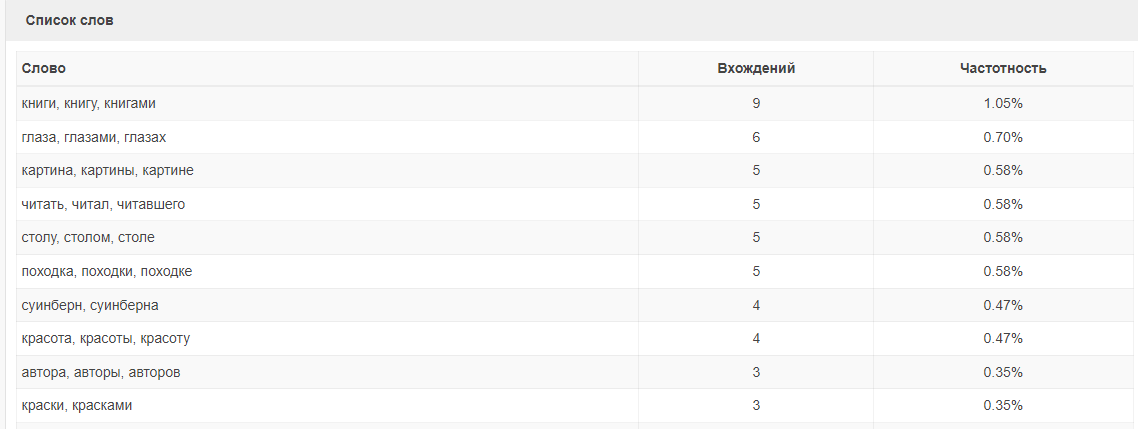
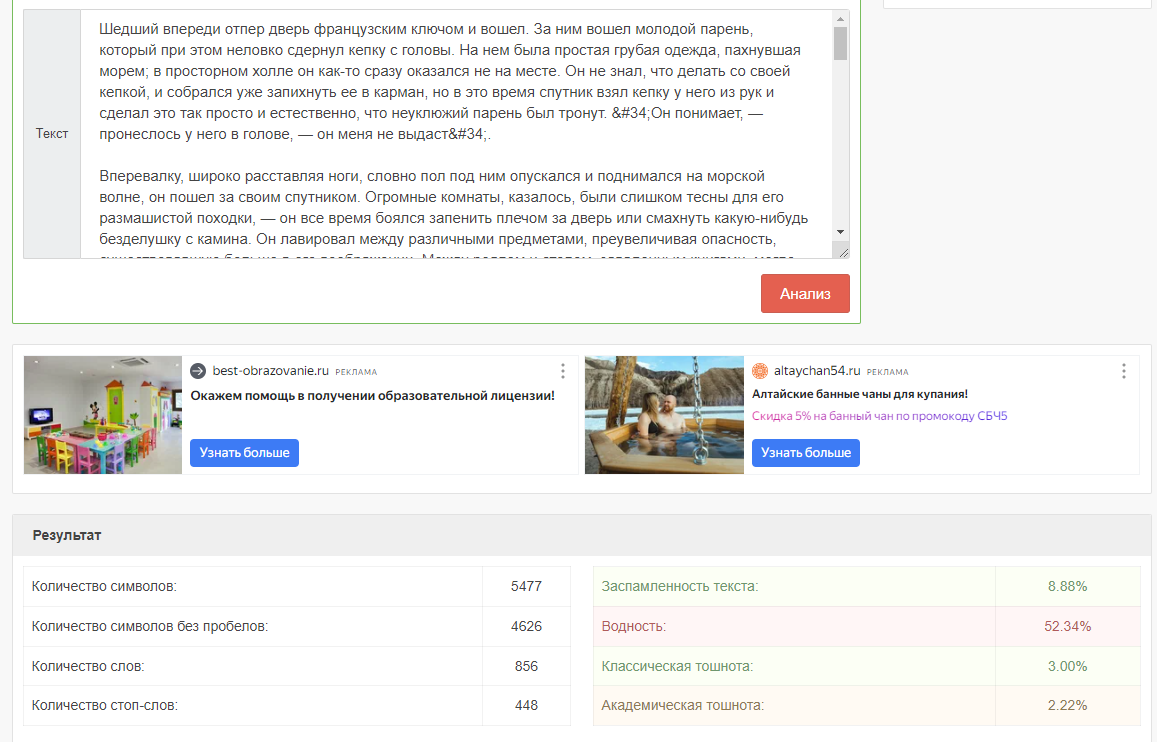
- Минусы:

- Может потребоваться время для освоения из-за сложности интерфейса.

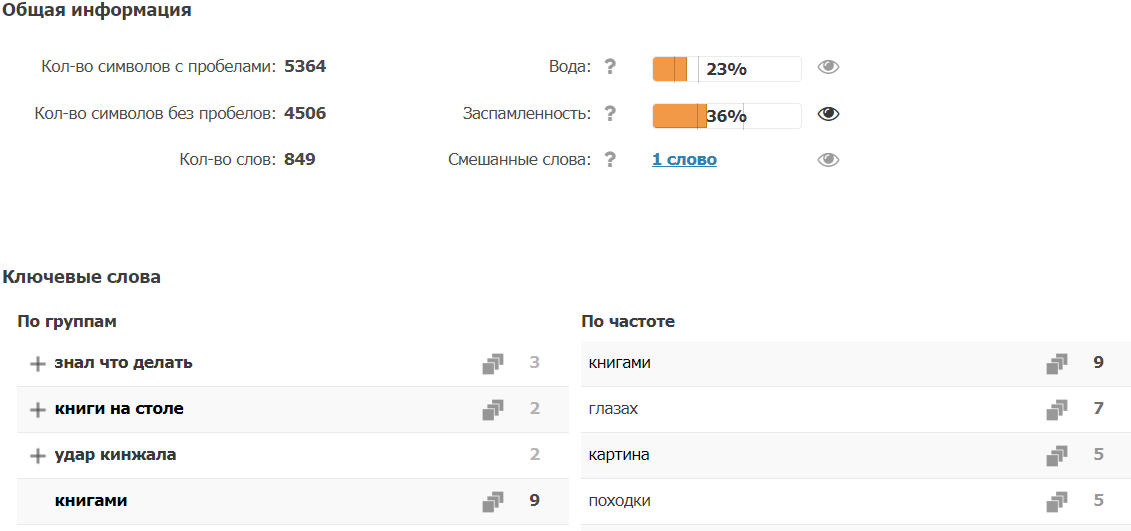
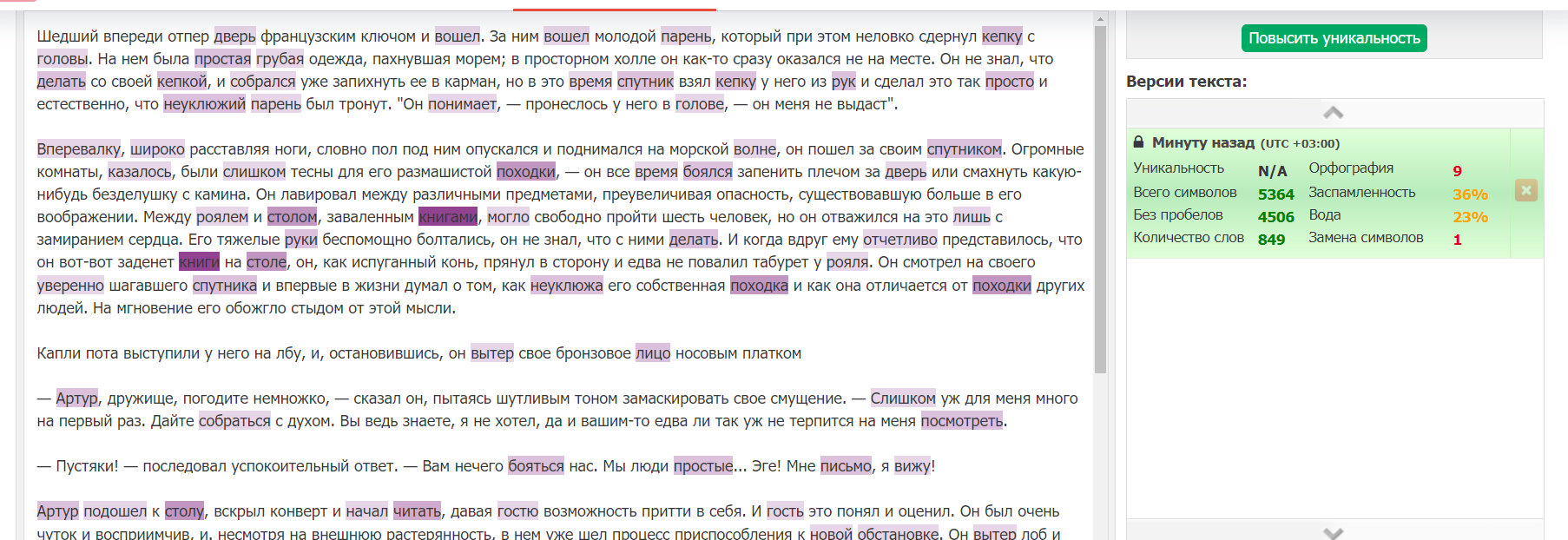
- Некоторые продвинутые функции могут быть доступны только в платной версии.

**Анализаторы текста**

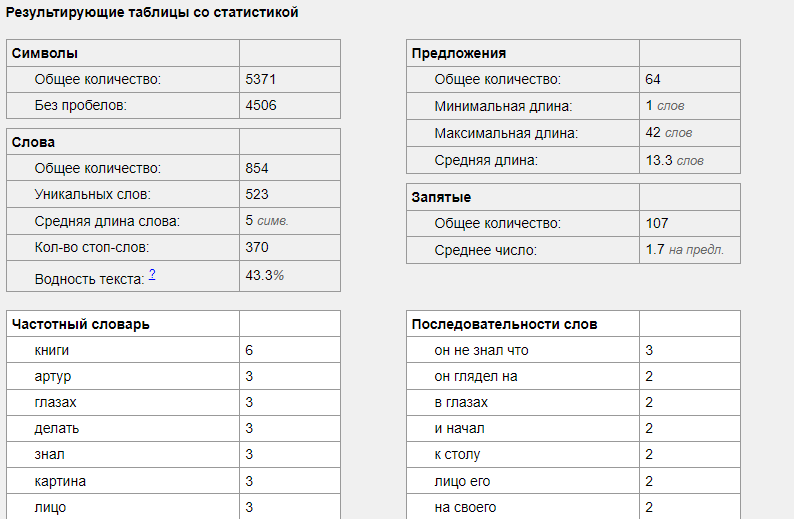
**SeoLik**

****

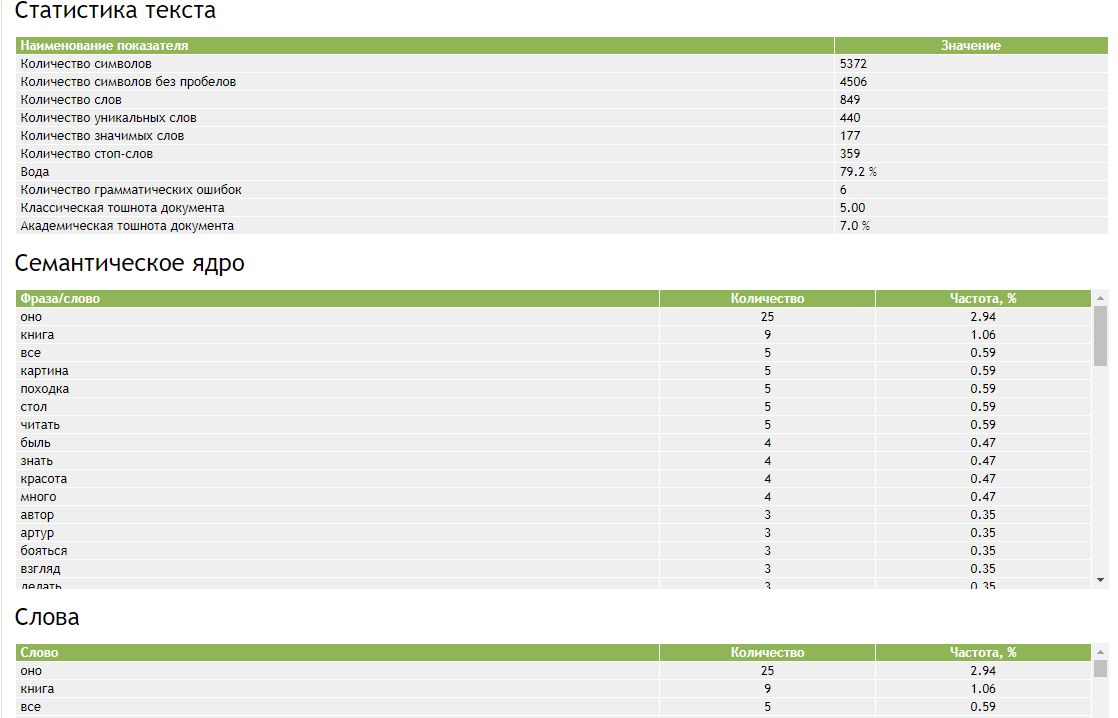
**Text.ru**

****

**textanalyzer.pro**

****

**advego.com**

****

Уровень 2

Задание 2

1.

|  |  |  |  |
| --- | --- | --- | --- |
| Слово | Пример, рус | Пример, англ | Ваш комментарий о возможностях морфологического анализатора |
| Существительное |  |  | В русском дана подробная характеристика. В английском даны различные интерпритации. |
| Глагол |  |  | В русском дана подробная характеристика. В английском даны различные интерпритации. |
| Прилагательное |  |  | В русском дана подробная характеристика. В английском даны различные интерпритации. |
| Наречие |  |  | В русском дана подробная характеристика. В английском даны различные интерпритации. |

2..

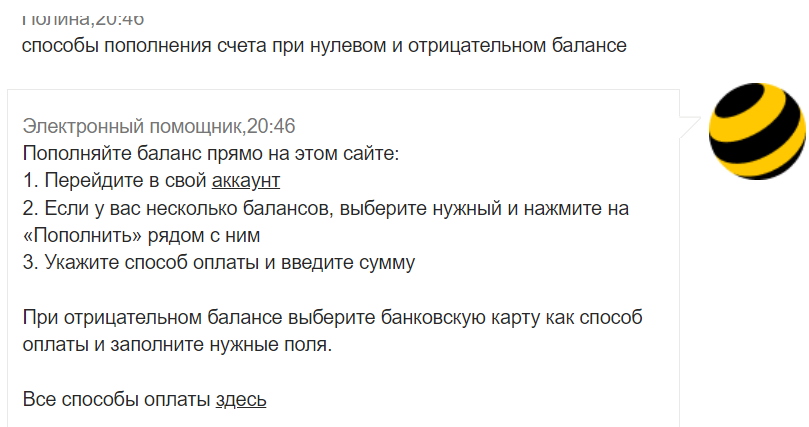
|  |  |
| --- | --- |
| Предложение, рус. | Лингвистическая информация,  представлена в синтаксическом анализе |
| 1 Печален я: со мною друга Нет, с кем долгую запил бы я разлуку, кому бы мог пожать от сердца руку и пожелать весёлых много лет | Это сложное предложение с разными видами связи: бессоюзной и подчинительной, состоит из двух частей (блоков), связанных бессоюзно; вторая часть раскрывает причину того, о чём говорится в первой; I часть по структуре представляет собой простое предложение; II часть - это сложноподчинённое предложение с двумя придаточными определительными, с однородным соподчинением. |
| 2Солнце закатилось, и ночь последовала за днем без промежутка, как это обыкновенно бывает на юге… | Повествовательное, невосклицательное, союзное, сложное (с разными видами связи: 1 и 2 — сложносочиненное, 2 и 3 сложноподчиненное): 1-ое: двусоставное, нераспространенное, полное, не осложнено [главное I или блок I]; 2-ое: двусоставное, распространенное, полное, не осложнено [главное II или блок II]; 3-ье: двусоставное, распространенное, полное, не осложнено (придаточное ко 2-му). |
| 3 Мне **надоело** жить в городе, и весной **я уехал** в деревню к знакомому рыбаку Михею | Предложение повествовательное, невосклицательное, сложное, состоит из 2 частей, связанных сочинительной связью, ССП с соединительным одиночным союзом И, выражающим причинно-следственные отношения между частями: 1 часть - односоставная, безличная, распространённая, полная, не осложнена; 2 часть - двусоставная, распространённая, полная, не осложнена. |
| Общий комментарий: сайт "Link Grammar for Russian" представляет собой ресурс, посвященный анализу и описанию грамматической структуры русского языка с использованием метода link grammar. Link grammar - это формальная грамматическая теория, которая анализирует связи между словами в предложениях. | |

3.

|  |  |
| --- | --- |
| Предложение, англ. | Лингвистическая информация,  представлена в синтаксическом анализе |
| 1 It's raining. | - It (subject) is raining (predicate). |
| 2 My father knows that he works as a teacher. | My father (subject) knows (predicate) that he works as a teacher.  - that (subordinating conjunction) introduces the subordinate clause "he works as a teacher."  - he (subject) works (predicate) as a teacher. |
| 3 The sun was shining, and our children decided to go to the beach. | - The sun (subject) was shining (predicate).  - our children (subject) decided (predicate) to go to the beach.  - and (conjunction) connects the two independent clauses. |
| Общий комментарий: Stanford Parser - это инструмент для анализа синтаксической структуры предложений в естественном языке. Он использует технологии глубокого обучения и машинного обучения для автоматического разбора предложений и определения связей между словам Stanford Parser работает путем обработки входного текста и создания структурированного дерева зависимостей между словами в предложении. | |

**Автоматический синтез диалогов**

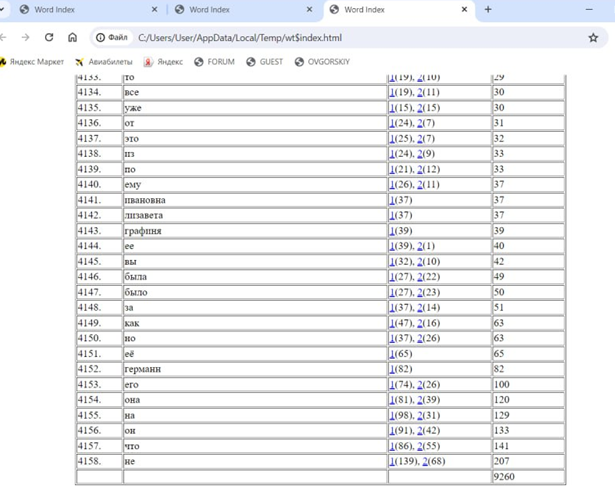
**Задание 2**



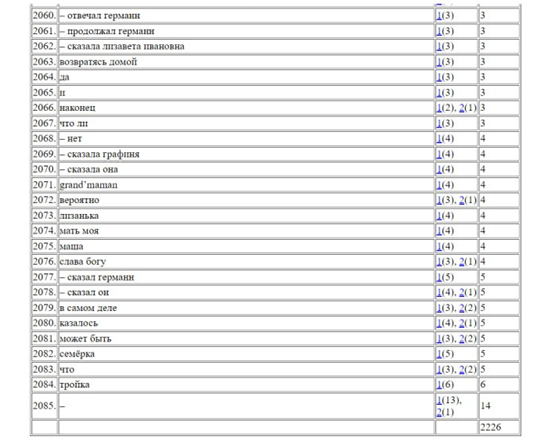
**Проектное задание 8 Прикладные программные средства**

**Возьмем произведения А.С.Пушкина «Пиковая дама» и «Метель».**

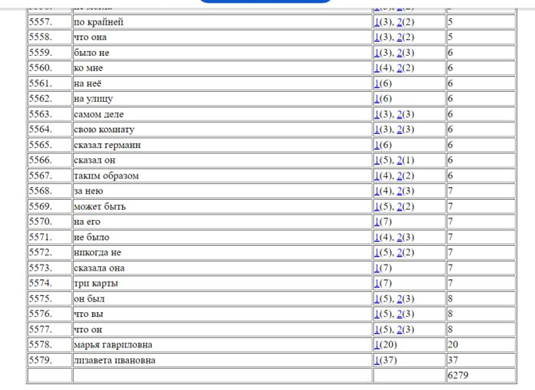
**Проведем анализ частотности слов (найдем самые частотные слова)**

****

**Самые частотные выражения**

****

**N-Grams**

****

**Вывод, сделанный на основе полного анализа.**

1. Устойчивые словоформы:

- В произведениях Пушкина часто встречаются слова, связанные с природой, эмоциями, чувствами, а также слова, отражающие общественные и философские аспекты.

- Особенно в "Метели" присутствуют метафоры и образы, связанные с природными явлениями, что характерно для романтического стиля писателя.

- В "Пиковой даме" можно наблюдать более строгую логику и точность в использовании слов, что связано с жанром повести.

2. Идиостиль писателя:

- Идиостиль Пушкина отличается изяществом, элегантностью и глубоким пониманием человеческой природы.

- Он умело использует язык для передачи эмоций, мыслей и образов, создавая неповторимую атмосферу в своих произведениях.

3. Интересы и приоритеты автора:

- В период написания "Метели" и "Пиковой дамы" Пушкин был заинтересован в изучении человеческих отношений, влиянии общественных норм на индивидуальность, а также в философских и моральных аспектах жизни.

- В "Метели" автор обращается к теме любви, судьбы, свободы, в то время как "Пиковая дама" затрагивает тему игры, страсти и жертвоприношений.

В целом, идиостиль Пушкина отличается глубоким пониманием человеческой души, изяществом языка и философским подходом к изображению жизни.

**Уровень 3.**

**Задание 2.**

**1.**

|  |  |  |  |
| --- | --- | --- | --- |
| Слово | Машинная основа | Слово | Машинная основа |
| семья | Сем | окружать | Окруж |
| весна | Весн | сидеть | Сид |
| дело | Дел | друг | Друг |
| пианино | пианин | ты | Ты |
| стремительный | Стремител | желтый | Желт |
| развивать | развив | петь | Пет |

**2.** основе город# соответствуют типовые парадигмы для описания слов городской и городовой;

- основе прыг# соответствуют типовые парадигмы для описания слов прыгать и прыгающий.

**3.** Исходные формы выделенных словоформ:

1. косой (прилагательное) - коса (существительное)

2. косой (прилагательное) - коса (существительное)

3. косить (глагол) - косить (глагол)

4. косить (глагол) - коса (существительное)

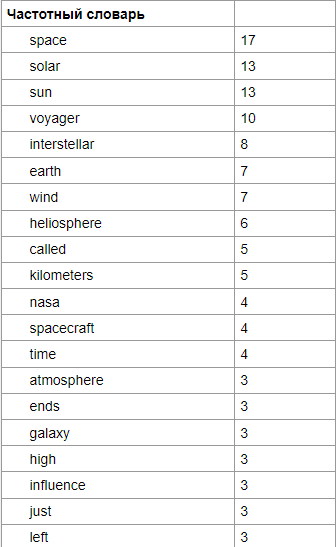
Данный текст иллюстрирует следующие виды омонимии:

- Гомонимия слов разных частей речи: "косой" как прилагательное и существительное;

- Гомонимия слов одной части речи: "косить" как глагол и "коса" как существительное.

4.

Берем текст про космос



You may see where I’m going with this. That line you drew in your head is arbitrary and, I dare say, even wrong.

First, well past this distance, there are icy bodies called [trans-Neptunian objects](https://www.youtube.com/watch?v=ZJscxTyI__s) (TNOs) that are nonetheless still bound to the sun by gravity. Some TNOs orbit the sun in a flattish torus called the Kuiper belt, and others orbit much, *much* farther out from a very roughly spherical halo called the Oort cloud that potentially stretches for a *trillion* kilometers around our star. On that scale, even the outer planets orbiting the sun seem huddled close-in.

And second, well, setting such outer limits depends on how you define what the solar system is and what’s outside it.

I was reminded of this because of a space news story that came out just last week, and it’s good news (a rare gem): engineers have been able to get Voyager 1 talking to Earth again. The deep-space probe was launched in 1977 and [is now a staggering 24 billion kilometers from Earth](https://voyager.jpl.nasa.gov/mission/status/), which is more than 160 times farther away from our planet than the sun is. Last November the spacecraft suffered a hardware glitch that scrambled its communications, and engineers had to get clever by rerouting software around the bad component. After they uploaded the fix, Voyager 1 appears to be working better, and they expect it will be back to full operational duty in the next few months.

This reminded me of something that happened in September 2013, when Voyager 1 was “just” 19 billion kilometers from Earth: [NASA announced that the spacecraft had entered interstellar space in August 2012](https://www.nasa.gov/news-release/nasa-spacecraft-embarks-on-historic-journey-into-interstellar-space/). At the time, a lot of people talked about how Voyager had finally “left the solar system.”

And here’s where we run into that second issue of where the solar system “ends.” By any real definition, even the fuzzy ones, Voyager 1 was still well within the solar system—certainly, it was (and still is, and will be for some time) closer to the sun than most of the TNOs in the black depths of space—yet NASA was correct: Voyager 1 is *also* in interstellar space.

How can this be?

This confusion arises because of two different ways of thinking of what defines the solar system. In this case, we’re comparing the sun’s gravitational influence, exerted upon the objects orbiting it, and its magnetic influence, delivered to deep space by its solar wind.

[The solar wind](https://www.scientificamerican.com/article/the-eclipse-is-coming-and-solar-science-will-never-be-the-same/) is a stream of subatomic particles the sun continuously blows into space. It flows away from the sun at high speed, nearly two million kilometers per hour, and consists of electrons, protons, neutrons and some heavier atomic nuclei as well. It’s not clear what accelerates the wind to such high speeds. Scientists know the sun’s magnetism is the driving force, but the exact mechanism still isn’t understood.

If space were truly empty, the solar wind would expand forever, flowing out into the galaxy and, because it moves at such high speed, eventually exiting the Milky Way entirely. But space—despite the name—is not empty. The vast volume between the stars does in fact have matter in it. It’s not much, to be sure: roughly one subatomic particle per cubic centimeter on average (although that can change hugely depending on where exactly in space you are). The air you’re breathing right now is some 10 quintillion times denser, so this interstellar matter is thin gruel indeed, but it’s enough.

As the solar wind plows into this ethereally thin cosmic vapor, it loses momentum and slow down, eventually coming to a halt. This region where it stalls out, poetically called the *heliopause,* marks the exterior boundary of the *heliosphere,* the volume of space dominated by the sun’s solar wind. Within the heliopause region, the sun’s magnetic influence wanes and that of the interstellar medium—the material between the stars—strengthens.

This shift is just what Voyager 1 detected in 2012. Several measurements showed that the interstellar medium dominated the region of space the spacecraft was passing through and that it had left the heliosphere behind.

So while Voyager 1 was still well inside the solar system, the space around it was influenced more by the galaxy itself than the sun.

As usual, when dealing with scientific matters, you need to be careful to define your terms.

And in the interest of open scientific honesty, I’ll admit [I’ve made this mistake myself](https://slate.com/technology/2013/03/voyager-1-the-spacecraft-has-apparently-left-the-solar-system.html). I wrote in early 2013 that Voyager 1 had left the solar system when, in fact, NASA said at the time that it had not. (This happened so often in media over the years that the webcomic xkcd, in its usual cheeky style, [had something to say about this topic as well](https://xkcd.com/1189/). NASA later confirmed after reviewing its data that the spacecraft had actually entered interstellar space in 2012.) But I also pointed out at the time how hard-and-fast definitions of even where the sun’s heliosphere ends are complicated and difficult to pin down. These regions are squishy and in flux, lacking any easily measured delineation.

If any of this sounds familiar, that’s because it’s reminiscent of pondering where Earth’s atmosphere ends and outer space begins—a quandary encapsulated by the debate over something called the Kármán line. [I covered this in a recent article](https://www.scientificamerican.com/article/where-does-outer-space-begin/), and there are some similarities; in both cases, we’re dealing with a sort-of atmosphere—the heliosphere carved out by the solar wind and Earth’s enveloping shroud of air—and where it impinges on the environment of deeper space. The difference here is that Earth’s atmosphere fades away gradually with altitude, blending seamlessly with the near vacuum of space, whereas the heliosphere does have a boundary. That neutral zone (to borrow a *Star Trek*–ism) is wide, certainly—tens of billions of kilometers through—but it’s small compared to the immense size of the heliosphere itself.

With Voyager 1 having long passed the interstellar version of the Kármán line, it’s well on its way into the galaxy. It will hopefully continue to take measurements of the interstellar medium and begin transmitting them back to Earth soon once again. Even after 46 years, it’s still breaking boundaries.

**5.** - Сам - выражает идею индивидуальности, относящейся к субъекту действия или являющейся основой для других объектов.

- Кто-то - указывает на неопределенное лицо или существо, отличное от говорящего и адресата.

- Иметь - обозначает наличие у субъекта какого-либо объекта, качества или свойства.

- Заставлять - означает принуждать, вынуждать кого-либо совершать какое-либо действие.

- Переставать - указывает на прекращение действия, переход из состояния активности в состояние пассивности.

- Начинать - обозначает начало действия, процесса или состояния.

- Не - отрицательная частица, используемая для отрицания действия или свойства.

- Приобретать - получать что-то новое, становиться обладателем.

- Утрачивать - терять что-то, лишаться владения.

- Сохранять - поддерживать в наличии, не терять.

- Лишать - лишать кого-то чего-то, отбирать у кого-то что-то.

**6.**

А              Говорите!

Б    Меня волнует погода.

А              Как давно вас волнует погода?

Б     С тех пор, как умерла моя мать.

А              Расскажите больше о Вашей семье.

Б      Они все в области компьютеров.

А              Машины пугают вас.

Б      Нет

А-машина, Б-человек. У человека больше эмоциональной составляющей, у машины все очень прямолинейно и холодно.