**МИНОБРНАУКИ РОССИИ**

**федеральное государственное бюджетное образовательное**

**учреждение высшего образования**

**«САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ**

**ЭКОНОМИЧЕСКИЙ УНИВЕРСИТЕТ»**

**(СПбГЭУ)**

Факультет информатики и прикладной математики  
Кафедра прикладной математики и экономико-математических методов

**ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА**

на тему: «Инструментальное обеспечение поддержки принятия решений на примере АО “Вертекс”»

Направление: 01.03.02 – Прикладная математика и информатика

Направленность: Прикладная математика и информатика в экономике и управлении

Обучающаяся очной формы обучения группы ПМ-2101

Воронина Мария Игоревна \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

*(ФИО) (подпись)*

**Руководитель** **ВКР** к.ф.-м.н., доцент, Лебедева Людмила Николаевна

*(ученая степень, ученое звание, ФИО)*

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

*(подпись)*

Нормоконтроль пройден «\_\_\_»\_\_\_\_\_\_\_\_\_\_ 2025 г.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

*(подпись лица, проводившего нормоконтроль)*

«Допущен(а) к защите» «\_\_\_»\_\_\_\_\_\_ \_\_\_\_\_2025 г.

Заведующий кафедрой

д.т.н., профессор Фридман Григорий Морицович

*(подпись)*

Санкт-Петербург

2025г.

**МИНОБРНАУКИ РОССИИ**

**федеральное государственное бюджетное образовательное**

**учреждение высшего образования**

**«САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ**

**ЭКОНОМИЧЕСКИЙ УНИВЕРСИТЕТ»**

**(СПбГЭУ)**

Факультет информатики и прикладной математики  
Кафедра прикладной математики и экономико-математических методов

|  |
| --- |
| УТВЕРЖДАЮ: Заведующий кафедрой ПМ и ЭММ д.т.н., профессор Фридман Г. М.  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  (подпись)  «\_\_\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_2025 г. |

**ЗАДАНИЕ**

**на выполнение ВКР**

обучающейся Ворониной Марии Игоревне группы ПМ-2101

1. Тема ВКР: «Инструментальное обеспечение поддержки принятия решений на примере АО «Вертекс»»

2. Цель и задачи ВКР: разработка приложения с пользовательским графическим интерфейсом для обеспечения поддержки принятия решений. К задачам относятся: изучение и анализ фармацевтического рынка России, моделирование продаж оригинальных лекарственных препаратов и генериков на рынке России, изучение и анализ способов веб-скрапинга данных, исследование структуры DOM-дерева онлайн-ресурсов, разработка алгоритма сбора данных о фармацевтической продукции с онлайн-ресурсов по заданному словарю, разработка веб-приложения для работы с собранными данными и отчетностью компании.

3. Срок сдачи оформленной ВКР на кафедру (с сопроводительными документами):

«\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 2025 г.

4. Перечень вопросов, подлежащих разработке и изложению в ВКР:

* общее описание фармацевтического рынка России: динамика, импортзамещение, онлайн-продажи в аптечном сегменте;
* моделирование продаж лекарственных препаратов на фармацевтическом рынке России;
* описание способов веб-скрапинга данных и проведение их сравнения;
* проведение анализа структуры DOM-дерева онлайн-источников данных;
* реализация алгоритма сбора данных о фармацевтической продукции из открытых источников;
* реализация веб-приложения для работы с данными о фармацевтической продукции и отчетность компании.

5. Консультанты по смежным вопросам ВКР (с указанием относящихся к ним разделов работы): –

«\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 2025 г.

Руководитель ВКР

к.ф.-м.н., доцент \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_/ (Лебедева Л.Н.)

*(подпись)*

Обучающаяся: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_/ (Воронина М.И.)

*(подпись)*

**РЕФЕРАТ**

с. 97, рис. 51, табл. 1, источн. 20, прил. 2

ФАРМАЦЕВТИЧЕСКИЙ РЫНОК РОССИИ, СИСТЕМНАЯ ДИНАМИКА, ОНЛАЙН-ПРОДАЖИ В АПТЕЧНОМ СЕГМЕНТЕ, DOM-ДЕРЕВО, ВЕБ-СКРАПИНГ ДАННЫХ, ВЕБ-ПРИЛОЖЕНИЕ, ВНЕШНИЙ ФИНАНСОВЫЙ АНАЛИЗ

Объект исследования – фармацевтический рынок России.

Предмет исследования – веб-скрапинг данных о фармацевтической продукции из открытых источников.

Цель работы – разработка приложения с пользовательским графическим интерфейсом для обеспечения поддержки принятия решений.

Методы и методология исследования: модели системной динамики, инструменты для работы с HTTP-запросами, управления браузерами, тестирования и отладки сайтов, инструмент для разработки веб-приложений, внешний финансовый анализ компании.

В результате: исследована важность онлайн-продаж в аптечном сегменте фармацевтического рынка России, разработана потоковая модель продаж лекарственных препаратов на рынке России в условиях импортзамещения, изучены и проанализированы способы веб-скрапинга, исследованы структуры DOM-дерева онлайн-ресурсов, разработан алгоритм сбора данных о фармацевтической продукции и разработано веб-приложение для обеспечения поддержки принятия решений.

Степень внедрения – результаты приняты к внедрению.

**СОДЕРЖАНИЕ**

[ВВЕДЕНИЕ 7](#_Toc199707604)

[1. ФАРМАЦЕВТИЧЕСКИЙ РЫНОК РОССИЙСКОЙ ФЕДЕРАЦИИ 9](#_Toc199707605)

[1.1 Общее состояние фармацевтического рынка России 9](#_Toc199707606)

[1.2 Влияние импортзамещения 11](#_Toc199707607)

[1.3 Аптечные продажи и онлайн-продажи 15](#_Toc199707608)

[1.4 АО «Вертекс» на фармацевтическом рынке России 16](#_Toc199707609)

[1.5 Выводы по разделу 1 17](#_Toc199707610)

[2. МОДЕЛИРОВАНИЕ ПРОДАЖ ОРИГИНАЛЬНЫХ ЛЕКАРСТВЕННЫХ ПРЕПАРАТОВ И ГЕНЕРИКОВ НА РЫНКЕ РОССИИ …………………………………………………………………………...19](#_Toc199707611)

[2.1 Разработка модели продаж оригинальных лекарственных препаратов и генериков на рынке 19](#_Toc199707612)

[2.2 Построение прогноза и исследование некоторых стратегий продвижения российских генериков 29](#_Toc199707613)

[2.3 Вывод по разделу 2 34](#_Toc199707614)

[3. РАЗРАБОТКА АЛГОРИТМА СБОРА ДАННЫХ О ФАРМАЦЕВТИЧЕСКОЙ ПРОДУКЦИИ 36](#_Toc199707615)

[3.1 Использование модуля Requests и сессий с подстановкой фиктивных заголовков 36](#_Toc199707616)

[3.2 Использование библиотеки Selenium и разработанного для нее пакета selenium\_stealth 39](#_Toc199707617)

[3.3 Изучение и анализ структуры источников данных 41](#_Toc199707618)

[3.4 Реализация алгоритма 47](#_Toc199707619)

[3.5 Выводы по разделу 3 57](#_Toc199707620)

[4. РЕАЛИЗАЦИЯ ВЕБ-ПРИЛОЖЕНИЯ ДЛЯ ОБЕСПЕЧЕНИЯ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ 58](#_Toc199707621)

[4.1 Архитектура приложения 58](#_Toc199707622)

[4.2 Глобальные переменные 60](#_Toc199707623)

[4.3 Первый отдел – обновление данных о фармацевтической продукции …… 61](#_Toc199707624)

[4.4 Второй отдел – просмотр данных о фармацевтической продукции и финансовой отчетности компании 65](#_Toc199707625)

[4.5 Третий отдел – общий анализ рынка, финансовой отчетности и отдельных лекарственных препаратов 72](#_Toc199707626)

[4.6 Выводы по разделу 4 9](#_Toc199707627)2

[ЗАКЛЮЧЕНИЕ 9](#_Toc199707628)3

[СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ 9](#_Toc199707629)5

[ПРИЛОЖЕНИЕ А Алгоритм сбора данных о фармацевтической продукции 98](#_Toc199707630)

[ПРИЛОЖЕНИЕ Б Веб-приложение для обеспечения поддержки принятия решений 115](#_Toc199707634)

# **ВВЕДЕНИЕ**

Задача сбора данных для их дальнейшего анализа, изучения, хранения, обмена и т.д. решается практически в любой компании. Причем, данные могут быть собраны из совершенно разных областей и для совершенно разных целей. Например, для анализа деятельности конкурентов и сравнения результатов их работы с результатами работы компании для того, чтобы в дальнейшем можно было определить стратегию поведения на рынке или чтобы оценить положение компании на рынке.

Существует много методов для решения данной задачи. И один из них – это веб-скрапинг. Веб-скрапинг – это автоматизированный сбор информации с онлайн-ресурсов. Для многих веб-скрапинг более известен, как парсинг данных. Конечно, принято считать, что веб-скрапинг и парсинг – это одно и то же, однако, парсинг все же является этапом веб-скрапинга, на котором из собранных данных извлекается нужная информация [2].

В компании АО «Вертекс», в отделе разработки и развития внутренних систем и сервисов, решается данная задача. Работники отдела собирают информацию с онлайн-витрин аптек о фармацевтической продукции своей компании такой, как косметические средства, лекарственные средства и т.д, для того, чтобы в дальнейшем передать собранные данные в отдел аналитиков.

Передо мной была поставлена задача по разработке веб-приложения для работы с данными о фармацевтической продукции компании и других фармацевтических производителей из открытых источников, а так же с данными бухгалтерской отчетности компании. Такое приложение и разработанный для него алгоритм сбора соответствующих данных должны облегчить работу как отделу разработки и развития, так и отделу аналитиков. Алгоритм будет полезен первому отделу тем, что автоматизирует ручной труд по сбору данных с онлайн-ресурсов, формирует уже отредактированные данные и загружает их в базу данных. А веб-приложение обеспечит отделу аналитиков поддержку принятия решений: оно позволяет обновлять данные в базе данных на текущую дату, используя разработанный алгоритм, просматривать и скачивать данные из базы, используя фильтры, просматривать финансовую отчетность компании. Помимо этого, веб-приложение позволяет увидеть общий анализ фармацевтического рынка в целом, анализ ценообразования конкретного лекарственного препарата и внешний финансовый анализ деятельности компании. С помощью данного приложения аналитики смогут проанализировать ситуацию на рынке, оценить конкурентоспособность своей организации и ее место на рынке, а также принимать решения, исходя из анализа данных.

Однако перед разработкой веб-приложения необходимо изучить и проанализировать фармацевтический рынок России, так как понимание его процессов, динамики и состояния в целом важно для оценки собранных данных и определения их ценности в будущем разделе аналитики веб-приложения. Кроме того, исследование данного объекта поможет определить, какой именно анализ необходимо отображать в веб-приложении, чтобы оно успешно выполняло свою цель – обеспечение поддержки принятия решений.

# **ФАРМАЦЕВТИЧЕСКИЙ РЫНОК РОССИЙСКОЙ ФЕДЕРАЦИИ**

В данном разделе был изучен и проанализирован объект исследования выпускной квалификационной работы – фармацевтический рынок России. А именно: его общее состояние в течение последних лет, влияние импортзамещения, быстрый рост числа генериков и возросшая значимость аптечных продаж, в том числе онлайн-продаж. Кроме того, в данном разделе описана компания по фармацевтическому производству АО «Вертекс» и ее место на фармацевтическом рынке России.

## **Общее состояние фармацевтического рынка России**

Фармацевтическая отрасль является одной из ключевых в структуре любой национальной экономики. Изучение, развитие и поддержка государством данного направления помогут значительно повысить уровень и продолжительность жизни жителей страны за счет обеспечения их и медицинских учреждений необходимыми и качественными лекарственными препаратами.

В России последние годы динамика фармацевтического рынка положительная, что можно увидеть на рисунке 1.1, причем, если в 2020-2021 годах объем рос за счет государственных закупок на фоне пандемии короновируса, то, начиная с 2022, большую роль играют аптечные продажи, о чем будет рассказано чуть позже. Однако из-за введенных в 2022 году мировыми государствами санкций против России и появившегося вследствие этого повышения курса доллара динамика фармацевтического рынка в долларах оказалась отрицательная, а доля российской фармацевтики на мировом рынке снизилась. В конце 2024 года и в начале 2025 года ситуация начала немного стабилизироваться: многие иностранные компании стали возвращаться на российский рынок, некоторые санкции были отменены, налажены отношения с США [5]. Ввиду этого было спрогнозировано значительное повышение емкости фармацевтического рынка России, что можно увидеть на рисунке 1.2.

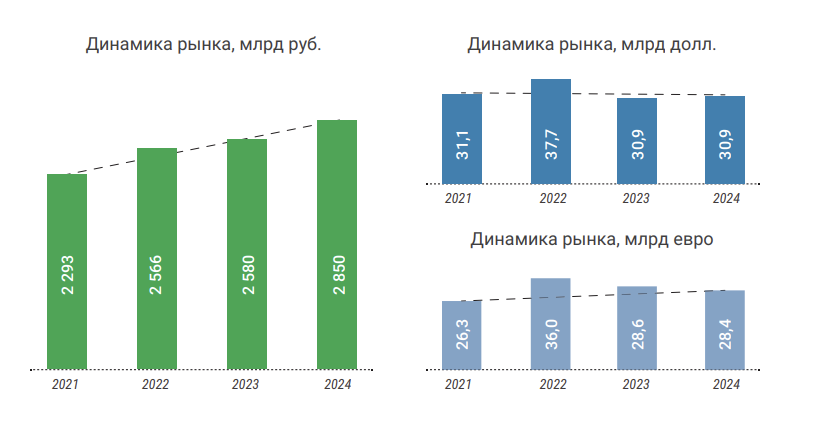


Рисунок 1.1 – Динамика фармацевтического рынка

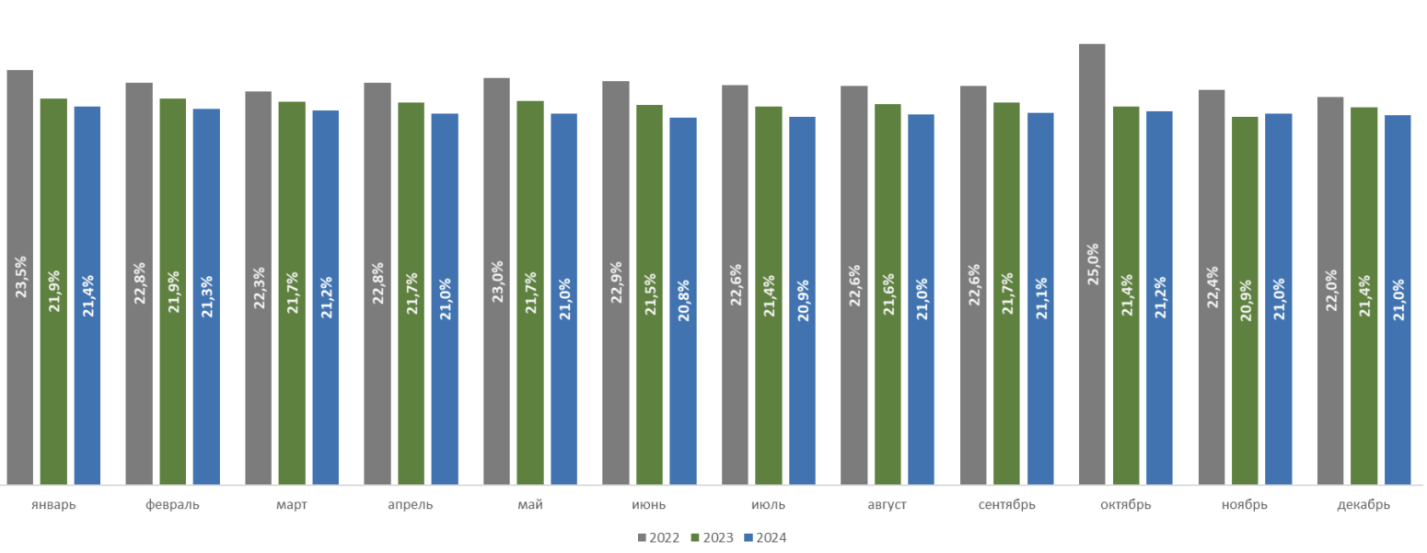


Рисунок 1.2 – Прогноз емкости фармацевтического рынка России, млрд. руб.

Однако вместе с динамикой фармацевтического рынка выросла наценка на лекарственные препараты и уровень инфляции (рисунок 1.3 и 1.4). На повышение повлияло много причин, но основные три были следующие: ставка Центрального Банка, которая за 2024 год повышалась три раза и сделала кредиты для бизнеса практически недоступными, слишком дорогими; снижение государственного финансирования и санкции, которые значительно усложнили закупки импортных препаратов, заменить которые пока невозможно [5]. Таким образом, с помощью наценки фармацевтические компании пытаются компенсировать рост стоимости производства.

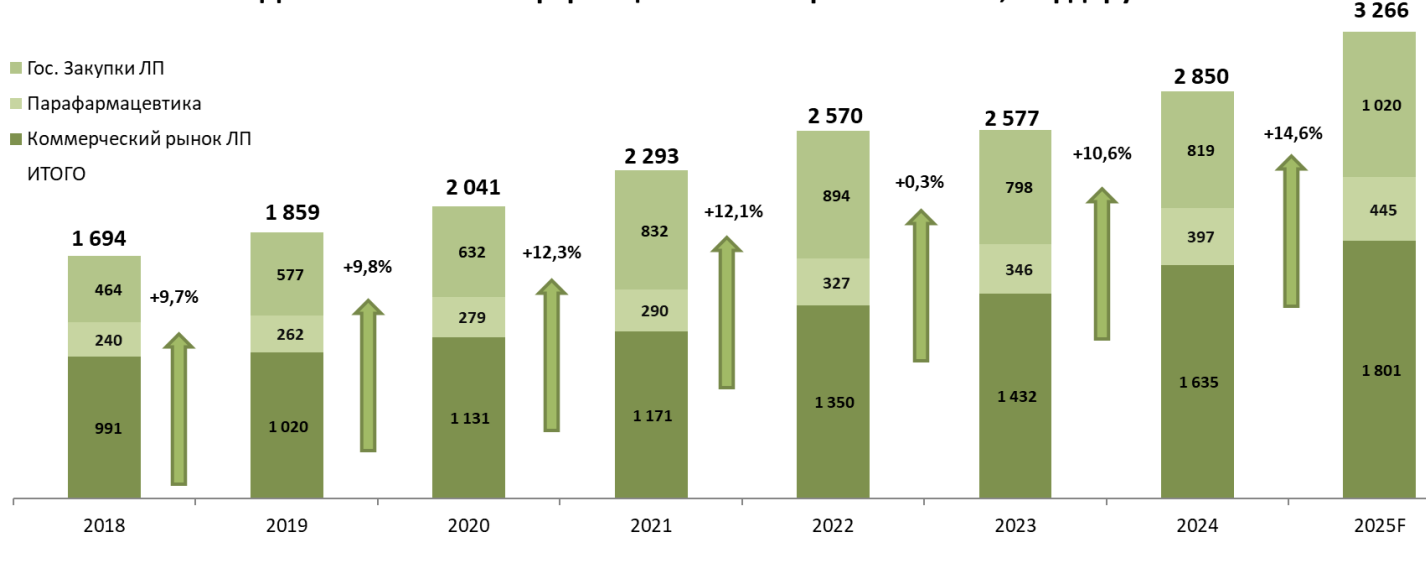


Рисунок 1.3 – Динамика аптечной наценки на лекарства, руб

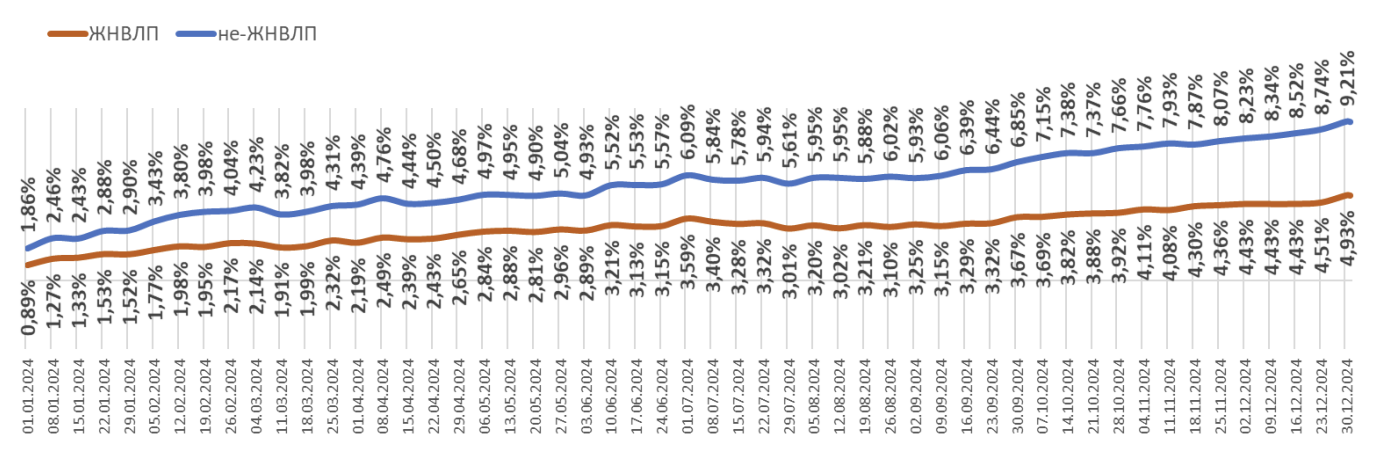


Рисунок 1.4 – Динамика уровня инфляции

Получается, что российский фармацевтический рынок динамично растет и развивается с каждым годом. Однако довольно высокий уровень инфляции и наценки, особенно на жизненно необходимые и важные лекарственные препараты, сильно осложняет доступность лекарственных средств для малообеспеченных людей, нуждающихся в них.

## **Влияние импортзамещения**

Как уже было сказано выше, в 2022 году против России мировыми государствами были введены санкции. Многие крупные иностранные компании покинули российский рынок. На фармацевтическом рынке развивалась следующая ситуация: иностранные компании с него не уходили, но очень многие полностью отказались от проведения совместных клинических испытаний и исследований, запуске новых лекарственных препаратов, а также некоторые компании отказались от поставки оборудования и сырья для производства лекарственных средств. Кроме того, некоторые американские компании отказались от поставки в Россию необходимых вакцин, таких как, вакцины от ветряной оспы, паротита, краснухи и кори.

Как можно понять из всего вышесказанного, у России появилась необходимость в поддержке и развитии отечественной фарминдустрии. За это время были приняты акты Правительства и Министерства о приоритизации российской продукции на отечественном рынке, с 2019 года по 2024 год действовал национальный проект «Здравоохранение», который определил векторы государственной политики в сфере защиты здоровья, многим фармацевтическим производителям и научным институтам были выданы патенты на производство и оказана государственная поддержка в клинических исследованиях. Кроме того, с 1 января 2025 года начал действовать закон о едином рынке для лекарственных препаратов, суть которого заключается в том, что Министерство Здравоохранения переходит на реестровую систему учета регистрации лекарственных препаратов. Помимо этого, в июне 2023 года была принята стратегия развития фармацевтической отрасли «Фарма-2030», основной целью которой является наращивание потенциала независимости российских фармацевтических компаний. Данная стратегия работает, в том числе, за счет принятого 1 января 2025 года закона о механизме «второй лишний»: иностранные препараты, производство которых не локализовано в России, будут отвергаться на торгах, если поступить хотя бы одна заявка от российского производителя [8]. Не стоит забывать и о государственных закупках, доля которых до сих пор остается значимой в общем объеме фармацевтического рынка, что можно увидеть на рисунке 1.5.

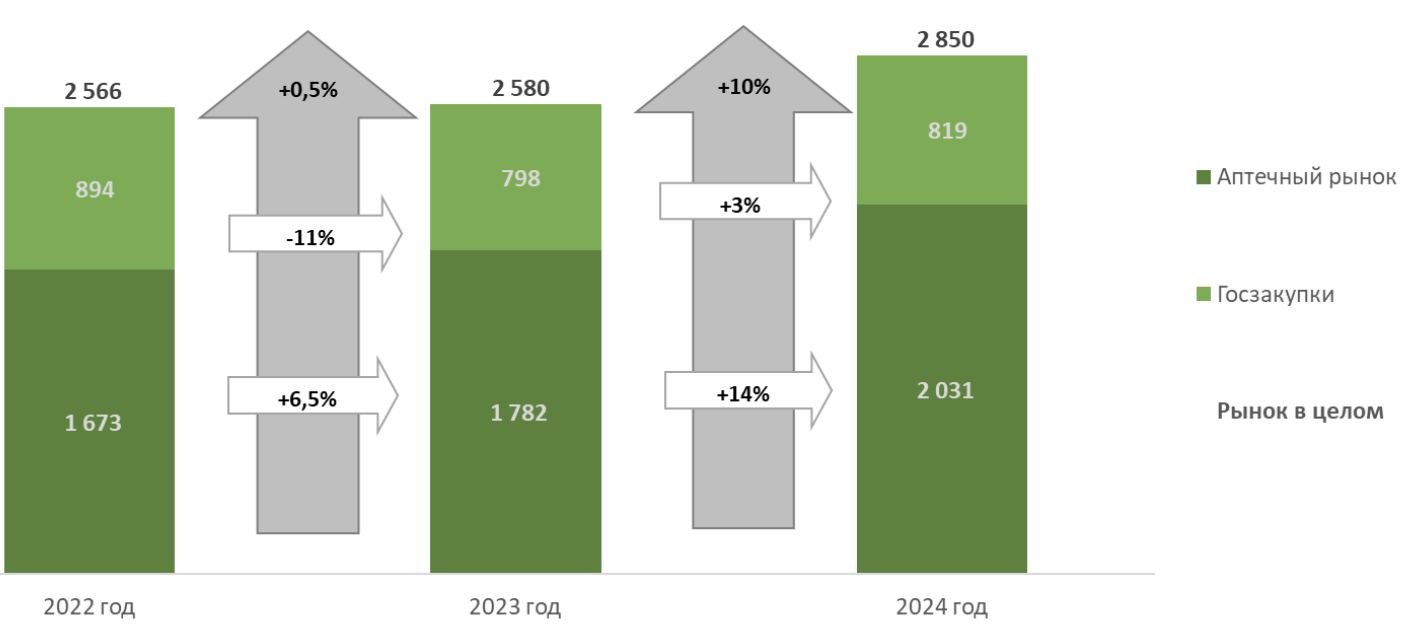


Рисунок 1.5 – Емкость фармацевтического рынка России, млрд. руб

Результаты всех принятых мер поддержки фармацевтической отрасли можно считать успешными. Во-первых, доля локализованных препаратов, то есть отечественного производства, выросла, что можно увидеть на рисунке 1.6. Конечно, российский рынок все еще остается зависимым от импортных препаратов, но все больше и больше российских фармацевтических производителей создают качественные продукты, которые постепенно обретают популярность у населения [5].

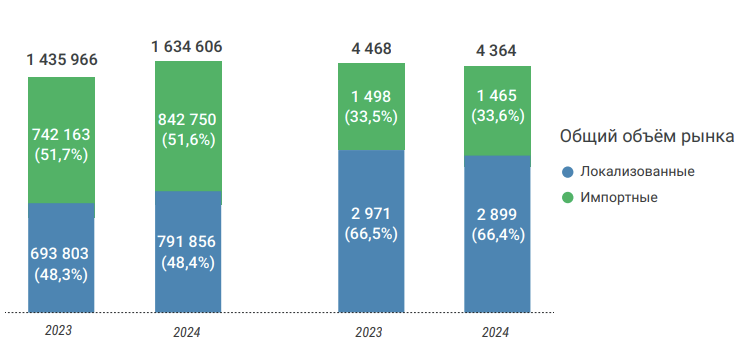


Рисунок 1.6 – Соотношение продаж импортных и локализованных препаратов, млн. руб

Во-вторых, с каждым годом растет доля генериков, что можно увидеть на рисунке 1.7. Фармацевтическое производство России развивается и замещает импортные оригинальные препараты, которые в связи с ситуацией не доступны на фармацевтическом рынке. А так как они дешевле оригинальных, но содержат то же самое действующее вещество, то генерики пользуются большой популярностью на фармацевтическом рынке России [6].

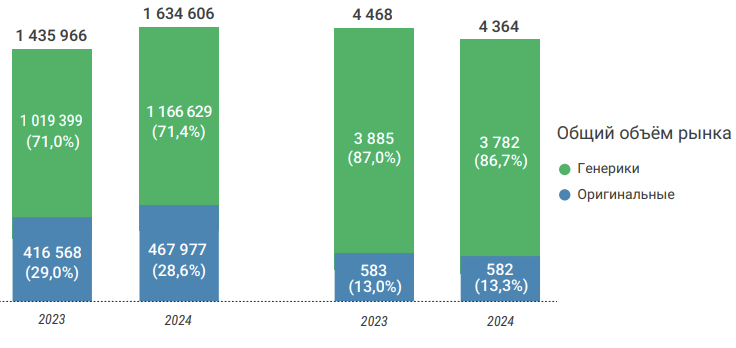


Рисунок 1.7 – Соотношение продаж оригинальных и генериковых препаратов, млн. руб

И, в-третьих, все больше и больше российских компаний входит в ТОП-20 фармацевтических производителей на рынке России. Благодаря поддержке государства, льготам, патентам и отсутствию на фармацевтическом рынке иностранных компаний, их отказу от клинических исследований и изучений российские фармацевтические производители развивают свое производство. Помимо этого, развивается не только разработка и производство российского оборудования, необходимого для производства лекарственных препаратов, но и логистические сети, обеспечивающие поставки сырья [5].

## **Аптечные продажи и онлайн-продажи**

Когда было рассказано про общее состояние фармацевтического рынка, то было упомянуто, что большую роль в положительной динамике рынка стали играть аптечные продажи. И это правда: в 2020-2022 годах ведущая роль была у государственных закупок на фоне пандемии короновируса. Государство обеспечивало медицинские учреждения всеми необходимыми лекарственными препаратами, медицинским оборудованием, инструментами и расходными материалами, которые закупала прямо у производителя. А вот начиная с 2023 года популярность аптечных продаж сильно возросла, что можно увидеть на рисунке 1.8. Причем растет не только выручка аптечного рынка России, но и количество аптечных точек. Это связно, во-первых, с усилением поддержки государством бизнеса, а, во-вторых, с выросшей доступностью лекарственных препаратов, вызванной увеличением доходов людей [6].



Рисунок 1.8 – Динамика аптечного рынка России

Однако последние годы положительная динамика не только у прямых аптечных продаж, но и у онлайн-продаж, что можно увидеть на рисунке 1.9. Они стали играть важную роль на фармацевтическом рынке России благодаря своим достоинствам. Покупатель всегда может на определенном сайте узнать, есть ли в наличии необходимое ему лекарственное средство и в каком аптечном пункте его можно забрать или заказать его с доставкой на дом. Так как онлайн-аптек бесчисленное множество, то с большой вероятностью человек сможет найти нужный лекарственный препарат. Но стоит сказать, что в данной ситуации кроется большой минус: некоторые веб-ресурсы могут быть непроверенными и нелицензированными продавцами, которые продают либо просроченные препараты, либо подделки. Поэтому в современной ситуации стоит осторожнее относиться к онлайн-ресурсам [5].

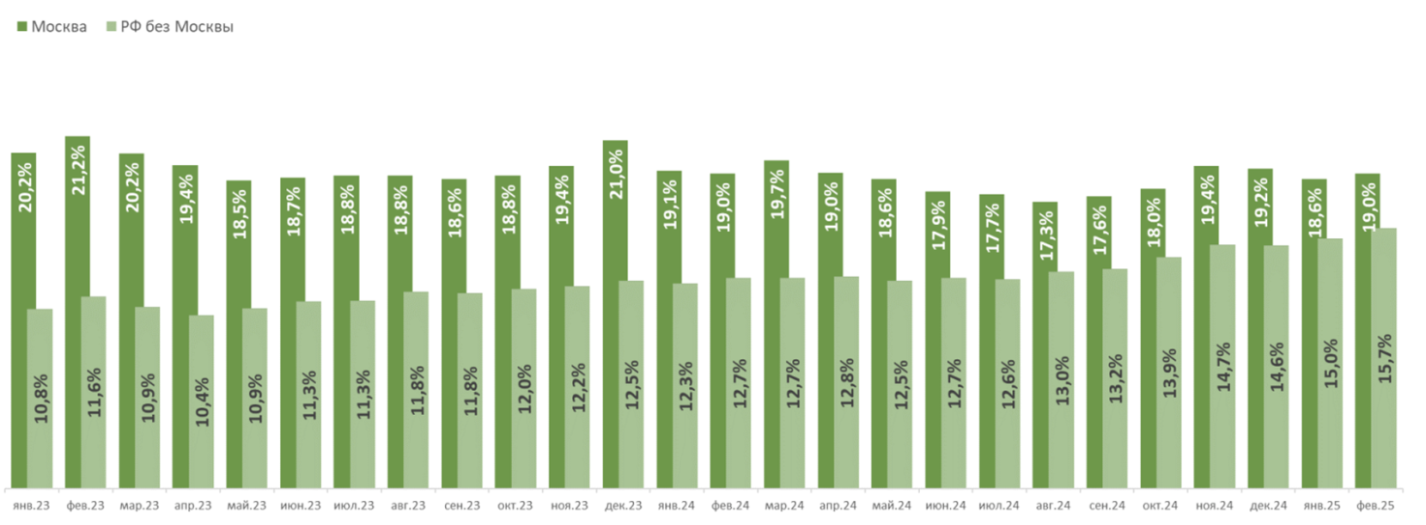


Рисунок 1.9 – Динамика доли онлайн-продаж на аптечном рынке, руб

Таким образом, аптечные продажи и, в частности, онлайн-продажи начали играть важную роль на фармацевтическом рынке России, что делает их отслеживание и изучение необходимой задачей.

## **АО «Вертекс» на фармацевтическом рынке России**

АО «Вертекс» - это один из крупнейших фармацевтических производителей России. Компания владеет большим производственным комплексом на участке «Новоорловская», уникальной экономической зоне города Санкт-Петербурга [7].

АО «Вертекс» был основан в 1999 году и начинал свое производство с линейки косметических препаратов, таких как, например, линейка средств по уходу за волосами «Alerana». За неполных 26 лет компания сумела достичь успеха на фармацевтическом рынке: она получила лицензию на производство лекарственных препаратов, добилась двух премий Правительства Санкт-Петербурга, стала лауреатом премии Правительства РФ в области качества [7]. И на протяжении последних лет она входит в ТОП-20 производителей России, что можно увидеть на рисунке 1.10.



Рисунок 1.10 – ТОП-20 российских производителей

На сегодняшний момент АО «Вертекс» является лидером по объему производства рецептурных препаратов и входит в тройку российских производителей лекарств, оказавших наибольшее влияние на фармацевтический рынок России.

## **Выводы по разделу 1**

Если подытожить все вышесказанное, то можно с уверенностью сказать, что фармацевтический рынок России движется в нужном, успешном направлении. С каждым годом его емкость растет, активно развивается отечественное фармацевтическое производство, заменяя импортные лекарственные препараты, создается необходимое оборудование для проведения клинических исследований, испытаний, а также для производства лекарственных средств. Помимо государственных закупок большую роль начали играть аптечные продажи, в частности онлайн-продажи, что делает задачу их отслеживания, исследования и анализа важной для любой фармацевтической компании. Особенно для такой крупной компании, как АО «Вертекс». Информация об онлайн-продажах ее лекарственных препаратов, а также лекарственных препаратов с таким же действующим веществом, произведенных другими фирмами, позволит АО «Вертекс» проанализировать свою деятельность и оценить результаты своей работы. Например, оценить свою конкурентоспособность на фармацевтическом рынке России или определить наценку на препараты, сравнивая цены на свою и постороннюю продукцию.

# **МОДЕЛИРОВАНИЕ ПРОДАЖ ОРИГИНАЛЬНЫХ ЛЕКАРСТВЕННЫХ ПРЕПАРАТОВ И ГЕНЕРИКОВ НА РЫНКЕ РОССИИ**

Как уже было сказано выше, значимость онлайн-продаж в аптечном сегменте с каждым годом повышается, поэтому их отслеживание, сбор данных и дальнейшее исследование становятся важными задачами для любого фармацевтического производителя. И они не менее важны для работы веб-приложения, так как на основе их анализа пользователи-аналитики смогут принимать решения по поводу будущих стратегий поведения на рынке, оценить конкурентоспособность и представить план дальнейшего развития продукции компании.

Но прежде чем приступить к разработке алгоритма по сбору соответствующих данных и веб-приложения на его основе, необходимо понять в целом, на каких принципах основан процесс продаж лекарственных препаратов на российском рынке, от чего он зависит и какой прогноз ждет рынок в будущем.

Именно для лучшего понимания процесса продаж фармацевтической продукции в данном разделе была разработана соответствующая упрощенная модель, на основе которой будет проведен прогноз на 5 лет вперед, чтобы осознать перспективы российского фармацевтического рынка.

## **Разработка модели продаж оригинальных лекарственных препаратов и генериков на рынке**

Для разработки модели были изучены следующие источники: статья про национальную фармацевтическую политику в Иране [19] и статья про моделирование распространения сердечнососудистых заболеваний в Австралии и стратегий по их предотвращению [20]. Первая статья помогла разобраться в причинно-следственных связях между фармацевтическим рынком и государством, которое на него влияет, и осознать, что от продвижения властями локальных препаратов прямо пропорционально зависит объем их продаж. Вторая статья оказалась полезна тем, что объяснила степень влияния уровня доверия пациентов и уровня одобрения врачами лекарственных препаратов на их объем продаж.

Таким образом, на основе описанных выше научных статей удалось разработать упрощенную модель продаж оригинальных лекарственных препаратов и генериков на фармацевтическом рынке России. Ее полезность заключается в том, что, во-первых, перед разработкой веб-приложения будет более детально изучен процесс продаж препаратов, что даст необходимую информацию для понимания того, насколько полезны собранные данные и какой анализ можно провести на их основе, а, во-вторых, будет построен прогноз по продажам на ближайшие пять лет, чтобы можно было увидеть тенденцию развития рынка России.

Модель разработана с помощью инструмента для имитационного моделирования AnyLogic, модельное время в ней будет исчисляться в годах, так как удалось найти данные для ее работы только на конец каждого года. В основе работы модели будут 4 накопителя. Первый из них – AllDrugs – показывает, сколько миллионов упаковок всех лекарственных препаратов поступило в продажу на текущий год. Следующие 3 накопителя отвечают за определенную категорию лекарственных препаратов: российские генерики, импортные генерики и оригинальные бренды. Такие категории были выбраны в связи с необходимостью исследования дальнейших перспектив российских генериков на рынке в ситуации импортзамещения.

Как известно, генерики – это лекарственные препараты, которые являются аналогами оригинальных брендов. И в текущей политической ситуации необходимо рассмотреть, во-первых, насколько конкурентоспособны российские импортзамещенные аналоги по сравнению с оригинальными импортными брендами, а, во-вторых, оценить их возможные перспективы развития исходя из ситуации на рынке.

В целом модель представлена на рисунке 2.1. Далее будет более подробно рассказано про каждый их объектов.

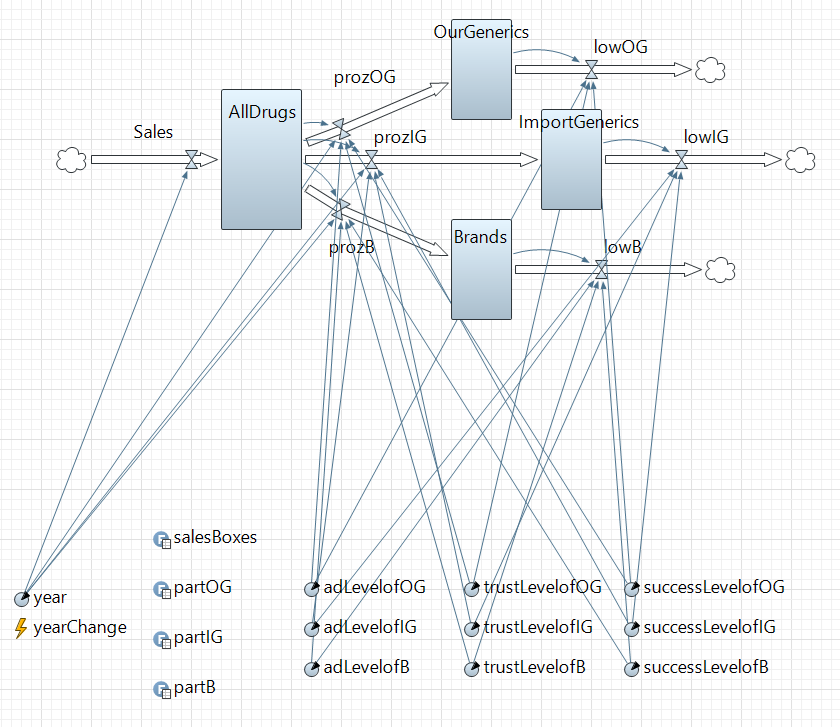


Рисунок 2.1 – Модель продаж оригинальных лекарственных препаратов и генериков на российском рынке

Первый из 4 возможных накопителей был описан выше. Второй накопитель – OurGenerics – отвечает за миллионы проданных упаковок российских генериков, которые являются локализованными версиями ушедших лекарственных препаратов. Третий накопитель, как можно понять по его названию (ImportGenerics), отвечает за импортные генерики, которые пришли на смену ушедших оригинальных брендов. После 2022 года, когда на Россию были наложены санкции, многие оригинальные бренды сократили количество своих поставок, а то и вовсе ушли с рынка. Но некоторые из них пока не представлялось возможным заменить российскими аналогами, поэтому на рынке в большом количестве начали появляться генерики брендов, которые пришли из содружественных стран: Китая, Индии, Пакистана и т.д. Соответственно, их тоже стоит учитывать, так как в текущей ситуации на фармацевтическом рынке России они играют значимую роль. И последний накопитель – это накопитель, отвечающий за миллионы упаковок оригинальных брендовых препаратов, которые остались на рынке и заменить которые пока не возможно.

Перейдем к переменным, от которых будут зависеть потоки изменения описанных выше накопителей. Первая переменная year отвечает за текущий год в модельном времени. Она необходима для того, чтобы в дальнейшем от ее значения срабатывали табличные функции модели. Вторая переменная представляет собой особый объект AnyLogic – Event. Это событие, которое происходит с указанной частотой и зависит от модельного времени. В данной модели оно весьма простое – всего лишь увеличивает текущий год на единицу на каждой итерации модели.

Теперь будут рассмотрены очень важные переменные, объекты AnyLogic – табличные функции. Данные объекты применяются для описания сложных нелинейных зависимостей, которые практически невозможно реализовать, используя стандартные функции. Другое их применение заключается в следующем: предположим, что есть данные эксперимента, собранные с некоторой периодичностью, и их необходимо привести к непрерывному виду, чтобы в дальнейшем экстраполировать их для нового, неизвестного периода. И именно с данной целью будут использоваться табличные функции в разработанной модели, чтобы с их помощью построить прогноз на 5 лет вперед.

Первая табличная функция – salesBoxes – отражает, сколько миллионов упаковок всех лекарственных препаратов в среднем поступило на рынок за указанный год. Ее реализацию можно увидеть на рисунке 2.2. Самое главное – это то, что интерполяция функции будет реализована с помощью сплайна, так как этот способ наиболее точный для прогноза значений среди всех предлагаемых. А данные для нее были собраны из отчетов по фармацевтическому рынку со специального ресурса, который специализируется на подобном, так что они могут считаться исторически верными [5].



Рисунок 2.2 – Табличная функция для отражения общего количества миллионов упаковок, поступающих на продажу

Следующие табличные функции будут отображать информацию о том, сколько процентов от поступивших в продажу миллионов упаковок составляют три соответствующие категории. Они точно таким же образом интерполированы сплайнами и точно также будут на этой основе экстраполировать данные для прогноза. Данные из них взяты из тех же самых профессиональных аналитических отчетов по фармацевтическому рынку России с соответствующих ресурсов, поэтому могут считаться исторически верными и значимыми [5]. И х можно увидеть на рисунках 2.3, 2.4 и 2.5 соответственно.

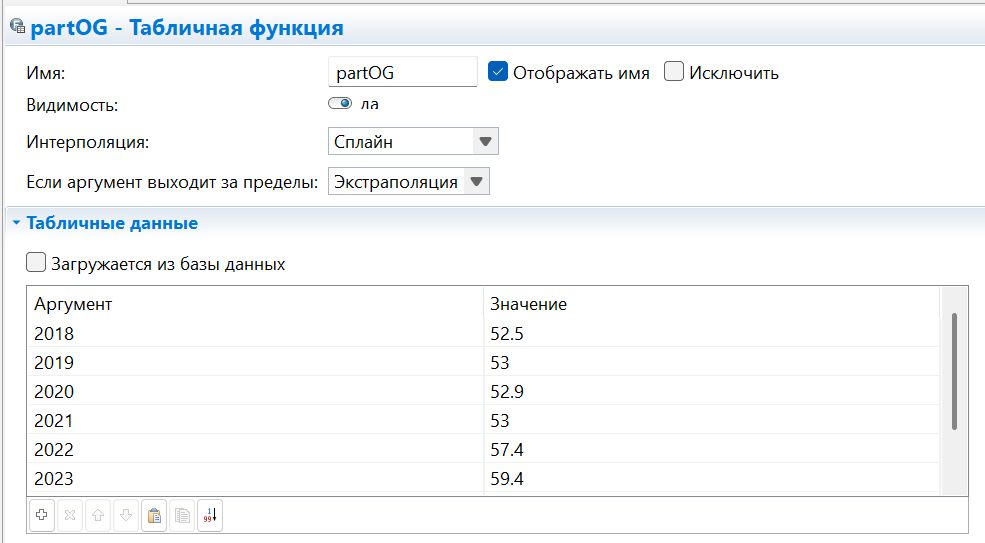


Рисунок 2.3 – Табличная функция для отражения доли русских генериков от всех поступивших на продажу препаратов

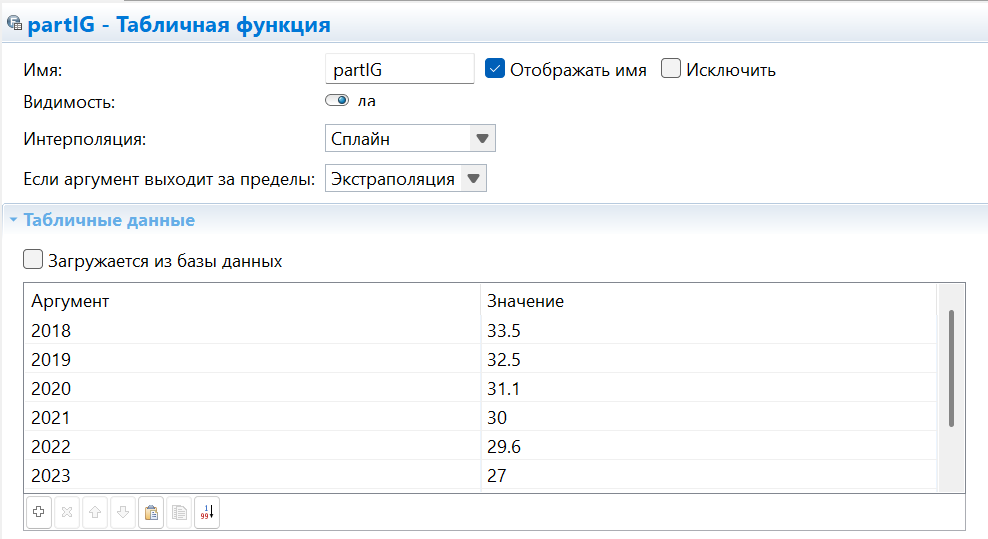


Рисунок 2.4 – Табличная функция для отражения доли импортных генериков от всех поступивших на продажу препаратов

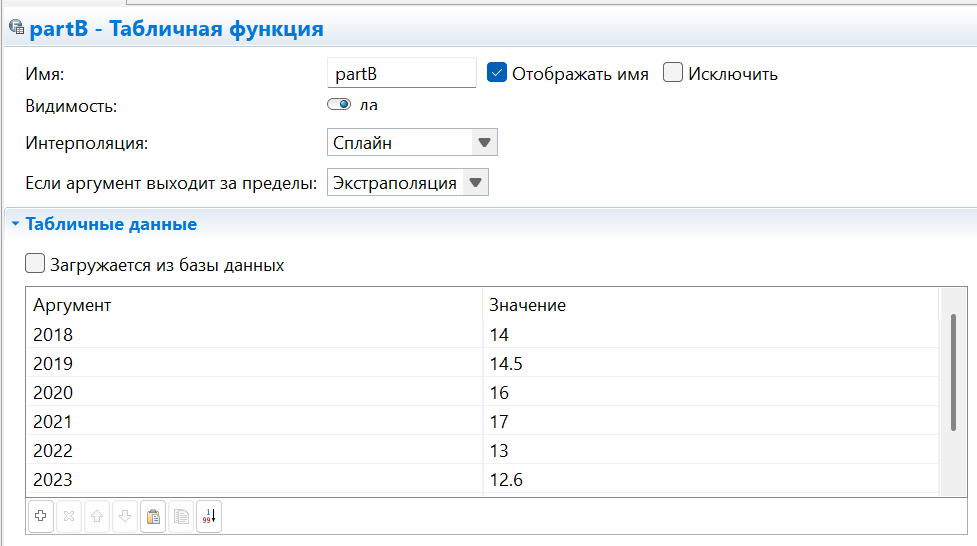


Рисунок 2.5 – Табличная функция для отражения доли импортных генериков от всех поступивших на продажу препаратов

И остались последние шесть параметров, по три для каждой из трех возможных категорий лекарственных препаратов. Про каждый из них было уже сказано выше, когда было представлено описание источников, на основе которых разрабатывалась данная модель. Первый, который для каждой из категорий называется adLevel, представляет собой уровень продвижения данной категории лекарственных препаратов. Сюда можно включить: рекламу продукции, сотрудничество с медицинскими учреждениями и поставки в них лекарственных препаратов и т.д. Интерпретировать данную переменную можно следующим образом: чем ближе к нулю, тем меньше уровень продвижения, и наоборот. Как ни странно, после изучения множества аналитических источников выяснилось, что уровень продвижения российских генериков по сравнению с оригинальными брендами весьма мал и составляет примерно 0.33 [6]. Казалось бы, политическая ситуация требует того, чтобы созданные отечественные аналоги как можно больше продвигались на рынке, но этого не происходит. Уровень продвижения брендов остается самым высоким и составляет примерно 0.76. Другая же ситуация с импортными генериками: уровень их распространения очень мал, так как они только-только начали появляться на российском фармацевтическом рынке и не могут выдержать конкуренции ни с российскими генериками, ни тем более с брендами. Данный уровень можно оценить с помощью аналитических отчетов примерно как 0.15.

Второй параметр для всех называется trustLevel и представляет собой уровень доверия населения к данной категории препаратов. Примерное его значение можно определить с помощью проводимых официальных опросов среди населения [6]. И можно заметить, что по сравнению с 2010-ыми годами уровень доверия к российским генерикам значительно вырос и составил примерно 78%. В основном это происходит из-за того, что данный препарат работает практически так же, как и оригинальный бренд, но стоит гораздо дешевле и доступен практически всем слоям населения. Однако, несмотря на ситуацию, уровень доверия к оригинальным брендам остается значительно выше (примерно 91%), на что еще большей частью влияет мнение лечащих врачей, о чем будет сказано позже. А вот у импортных генериков можно сказать, что наблюдается стабильность: уровень доверия к ним очень низок и составляет 30%. Связано это с тем, что они только недавно появились на рынке и практически не проверены в деле.

Последний важный параметр, который влияет на потоки изменения продаж препаратов, - это successLevel для каждой из трех категорий. Он означает уровень удовлетворенности лекарственными препаратами, но уже не у населения, а непосредственно у врачей, у профессионалов, которые могут оценить, насколько он успешно справляется со своей задачей. И, как показали опросы и исследования, данный уровень у врачей для российских генериков очень низкий, несмотря на воодушевленность населения [6]. Многие из них отмечают, что данные генерики не выполняют своей задачи, как оригинальные бренды, или вообще не действуют на пациентов. Поэтому в результате всех изучений уровень удовлетворенности российскими препаратами составил всего 4,4%. Про импортные генерики вообще говорить не стоит: только 2,8% врачей допускают их использование. А вот оригинальными брендами довольны 69% процентов профессионалов, так как они доказали свое качество и полезность.

После того, как были описаны все объекты и параметры, необходимые для реализации модели, можно описать процесс изменения количества проданных упаковок лекарственных препаратов. И самый первый поток под названием Sales отображает, сколько миллионов упаковок примерно поступило на продажу за текущий год. Как можно понять, он зависит от табличной функции, связанной с продажами всех лекарственных препаратов. На самом деле он очень прост: всего поступает столько препаратов, сколько указано в табличной функции за текущий год, что можно увидеть на рисунке 2.6.

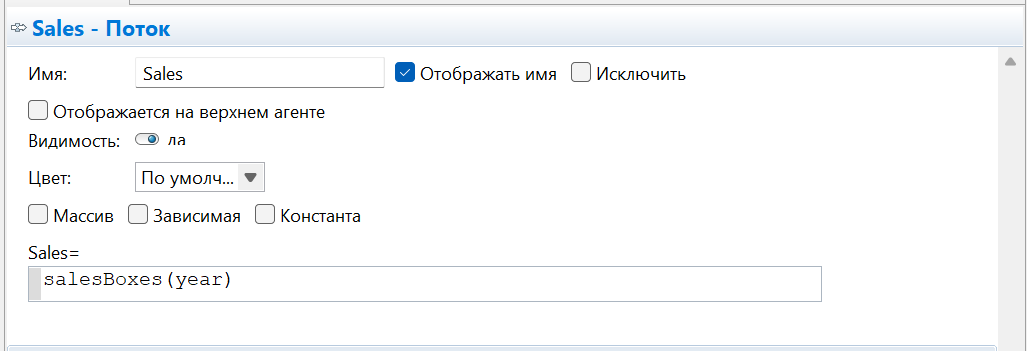


Рисунок 2.6 – Поток Sales

Следующие три потока будут описаны на примере одного из них – prozOG. Они несут один и тот же смысл, просто зависеть будут от соответствующих категории переменных. Данный поток показывает, сколько миллионов упаковок из всех поступивших в продажу является российскими генериками. Рассчитывается он по следующему принципу: есть табличная функция, которая отвечает за процент соответствующей категории от всех поступивших в продажу лекарственных препаратов. Соответственно, в накопитель, отвечающий за российские генерики, будет поступать данный процент препаратов от всех проданных в зависимости от года. Но сверху к данному проценту будут добавлены еще некоторые проценты. Первый из этих процентов представляет собой уровень доверия населения, умноженный на уровень продвижения препаратов. Смысл в этом следующий: данный процент препаратов может очень хорошо продвигаться и быть популярным у населения, поэтому его очень быстро раскупают, из-за чего нужны поставки сверх той нормы, которая установлена. И как раз этим процентом и добавляются дополнительные препараты, которые хорошо продаются на рынке. Тот же самый смысл и у второго надбавочного процента: некоторые препараты очень популярны у врачей, поэтому процент их удовлетворенности тоже стоит учитывать, как надбавку. Соответствующую реализацию можно увидеть на рисунке 2.7. Для оставшихся двух категорий потоки будут реализованы подобным образом, только с указанием подходящих для категорий параметров и объектов.

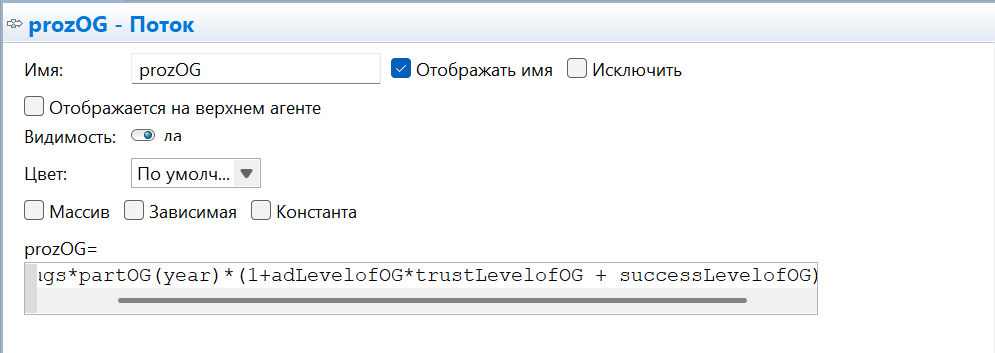


Рисунок 2.7 – Поток prozOG

И, наконец, остались потоки, которые отвечают за уменьшение количества продаж в миллионах упаковках. Они тоже будут показаны на примере одной категории – российских генериков. В данных потоках процент уменьшения общего объема проданных упаковок соответствующей категории рассчитывается исходя их двух ситуаций. Первая ситуация: процент препаратов, которые не были продвинуты в достаточной степени для того, чтобы стать популярными у населения, поэтому их количество в общем объеме снижается. Вторая ситуация: препараты, которые были продвинуты достаточно, но не получили одобрения ни у населения, ни у врачей, поэтому число их продаж в миллионах упаковок тоже снизилось. Пример данного потока можно увидеть на рисунке 2.8. Для оставшихся двух категорий все происходит подобным образом, только поток зависит от соответствующих параметров и объектов.

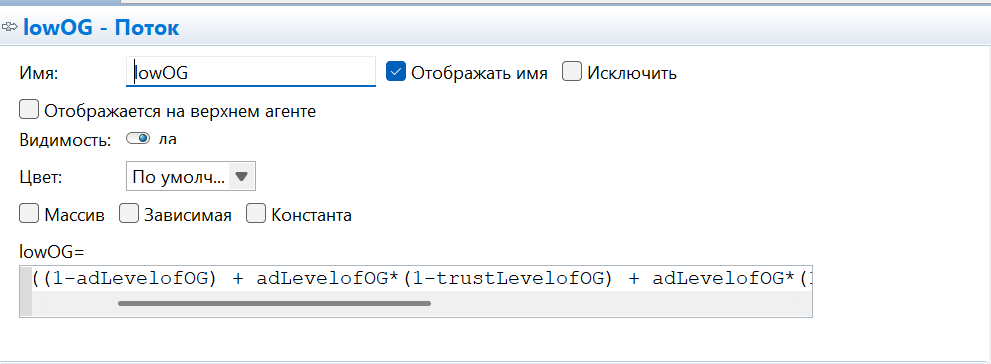


Рисунок 2.8 – Поток lowOG

Таким образом, была разработана модель, которая отражает продажи оригинальных брендов и генериков на российском фармацевтическом рынке. Следующий шаг – это построение на ее основе прогноза на ближайшие пять лет и оценка различных стратегий.

## **Построение прогноза и исследование некоторых стратегий продвижения российских генериков**

В данной модели имеется все для построения прогноза на ближайшие пять лет: постоянные параметры и табличные функции, которые смогут экстраполировать соответствующие данные для будущих лет. Пять лет было выбрано потому, что это самый оптимальный срок для прогноза: на меньшее количество лет прогнозировать не имеет смысла, так как не будет видна тенденция, а на большее – ситуация может поменяться очень сильно и прогноз будет неверным.

Для начала будет запущен эксперимент с текущими данными, определенными и взятыми из аналитических отчетов. Для этого, будет создан временной график, на котором будут отражаться продажи в миллионов упаковок по каждой из категорий. Устанавливаем во времени эксперимента, что он идет с 2018 по 2029 года. Данные имеются с 2018 по 2024 года, а на последующие пять лет будет получен прогноз, который можно увидеть на рисунке 2.9.

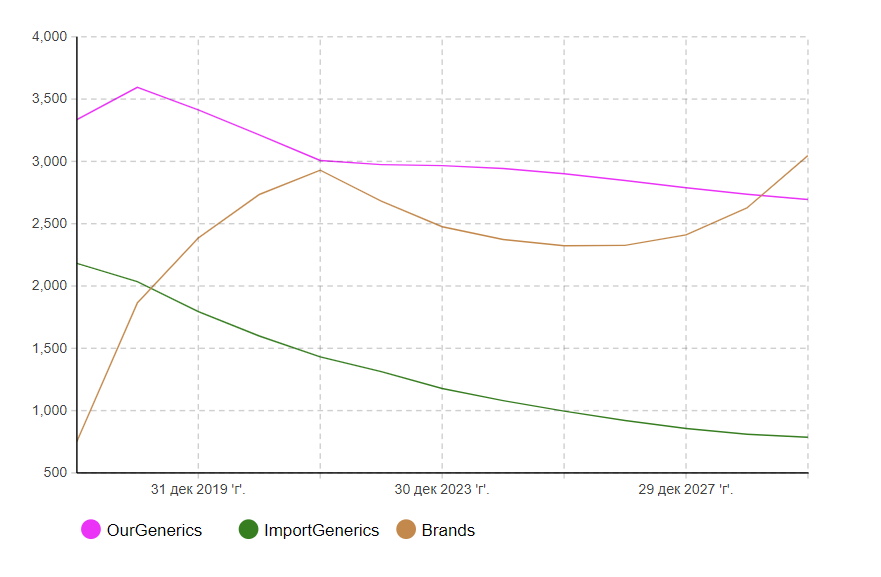


Рисунок 2.9 – Прогноз по продажам на пять лет

Исходя из полученного прогноза, можно сказать следующее: продажи оригинальных брендов будут с каждым годом возрастать, что может являться правдой, так как с 2024 года налаживается политическая ситуация между США и Россией. Она может повлечь за собой налаживание поставок импортных брендов и возобновление совместного создания лекарственных препаратов и их клинических испытаний. Соответственно, будет расти их уровень продвижения и доверия среди населения, а значит будут расти и продажи оригинальных брендов. С импортными генериками тоже все ясно: из-за возобновления отношений между странами, а также из-за развития отечественного фармацевтического производства их распространение и продажи будут с каждым годом снижаться. Но, к сожалению, будут снижаться продажи и российских генериков. Не так сильно, как импортных, но все же будут, особенно по сравнению с оригинальными брендами, которые снова займут лидирующие позиции на рынке России.

Преобладание оригинальных брендов и снижение продаж российских генериков весьма печально для отечественной фармацевтики. Поэтому можно рассмотреть три стратегии продвижения и развития российских генриков и построить прогноз для них.

Первая стратегия – это повышение уровня продвижения соответствующих лекарственных препаратов. Как уже было сказано выше, для отечественных лекарств практически не осуществляется никакого продвижения, особенно по сравнению с оригинальными брендами, поэтому стоит узнать, как измениться ситуация, если повысить данный уровень. Если указать уровень продвижения равный 0.48, то можно заметить на рисунке 2.10, что объемы продаж хоть и падают, но не так заметно, как изначально, они остаются практически на уровне продаж оригинальных препаратов.

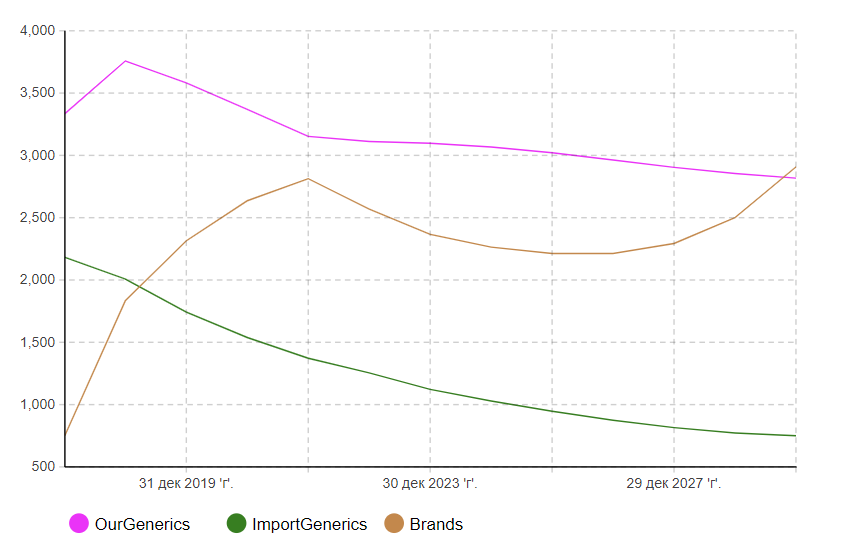


Рисунок 2.10 – Первый вариант прогноза первой стратегии

А если попытаться поднять уровень продвижения до 0.73, практически такой же, как и у оригинальных брендов, то на рисунке 2.11 можно увидеть, что продажи российских генериков превышают продажи оригинальных брендов. Это означает, что при должном уровне продвижения российские генерики являются вполне себе конкурентоспособными на фармацевтическом рынке.

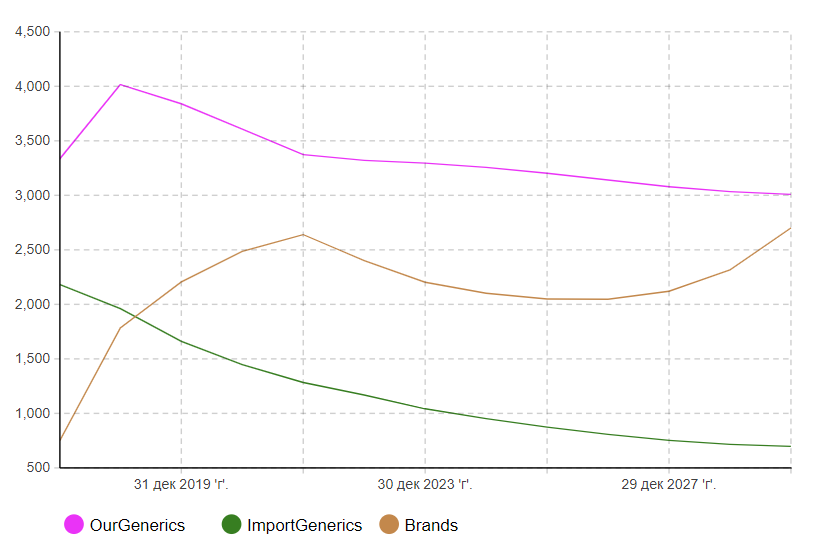


Рисунок 2.11 – Второй вариант прогноза первой стратегии

Вторая стратегия – это повышения уровня одобрения врачей за счет улучшения качества генериков и проведения более серьезных клинических испытаний. Можно повысить уровень удовлетворенности врачей до 12.4%, а затем до 24.4%. Результаты прогноза по этим двум случаям можно увидеть на рисунках 2.12 и 2.13 соответственно. И снова будет видна та же картина, что и в предыдущей стратегии: сначала объемы продаж практически выравниваются, а затем объем продаж российских генериков становится выше, чем у брендов.

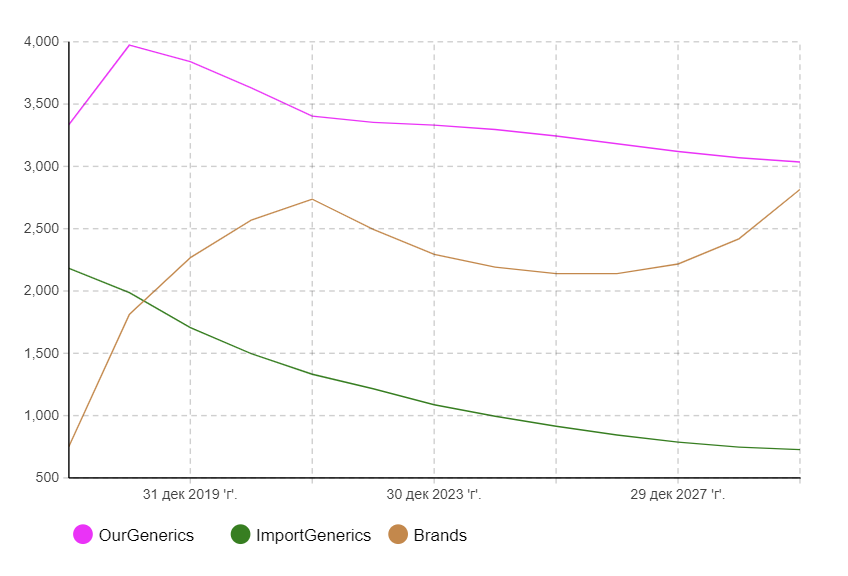


Рисунок 2.12 – Первый вариант прогноза второй стратегии

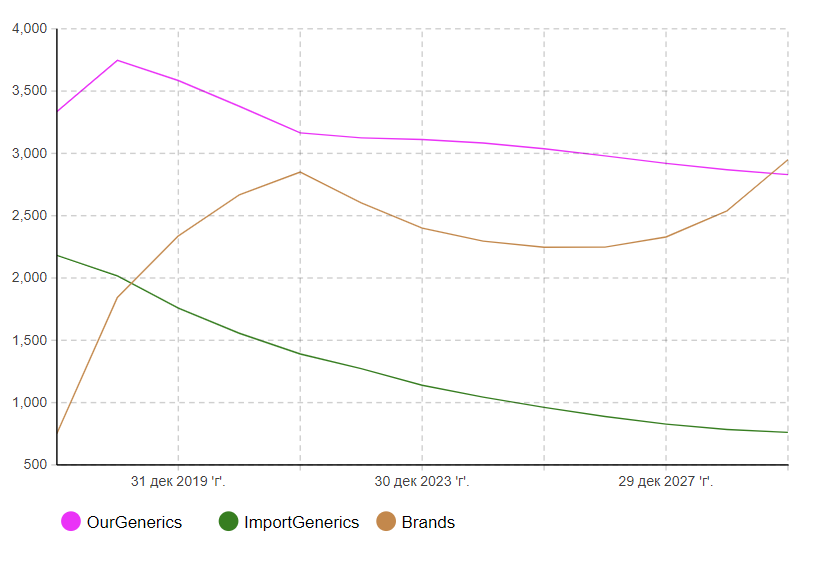


Рисунок 2.13 – Второй вариант прогноза второй стратегии

И третья стратегия заключается в повышении обоих описанных показателей. Конечно, обеспечить подобное развитие за ближайшие годы практически невозможно, но можно промоделировать ситуацию, чтобы выяснить, как изменить положение российских генериков на рынке. Возьмем уровень продвижения 0.63, а уровень удовлетворения врачей – 20.4%. Прогноз можно увидеть на рисунке 2.14. Разница между объемами продаж отечественных генериков и брендов стала еще больше, что означает несомненную конкурентоспособность генериков, но только при соответствующих условиях, развить которые еще только предстоит.

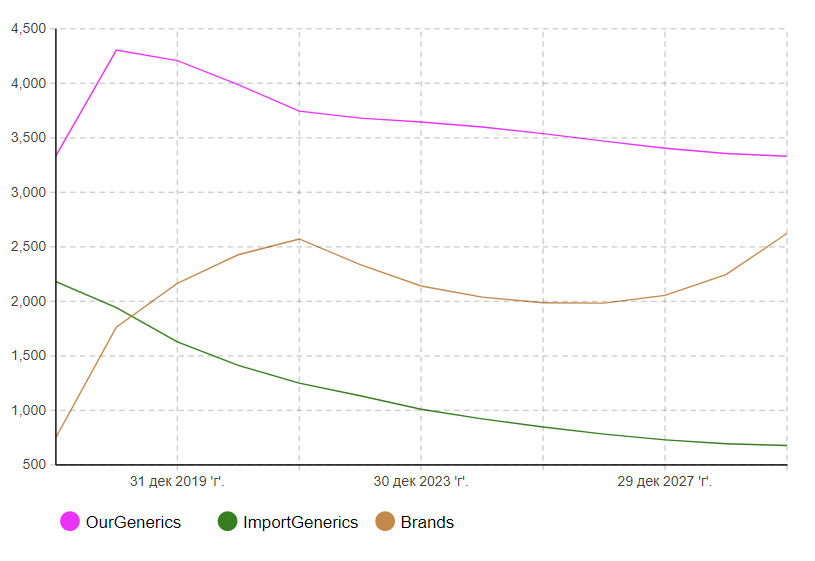


Рисунок 2.14 – Прогноз третьей стратегии

Таким образом, данные стратегии отражают возможное поведение российских генериков на рынке при улучшении условий их продаж.

## **Вывод по разделу 2**

В данном разделе были изучены научные статьи про фармацевтику различных стран, на основе которых была разработана упрощенная потоковая модель продаж оригинальных лекарственных препаратов и генериков на рынке. Кроме того, исходя из текущей ситуации, был получен прогноз на пять лет, который показал недостатки в продвижении отечественных генериков, которые призваны импортзамещать ушедшие бренды. Поэтому были рассмотрены возможные стратегии улучшения положения российских препаратов на рынке, которые показали, что при должном улучшении ситуации генерики способны составлять конкуренцию брендам.

# **РАЗРАБОТКА АЛГОРИТМА СБОРА ДАННЫХ О ФАРМАЦЕВТИЧЕСКОЙ ПРОДУКЦИИ**

В данном разделе были изучены и проанализированы различные способы веб-скрапинга данных, а конкретнее: способы работы с HTTP-запросами на языке программирования Python, а также был выбран наиболее эффективный способ, который будет использован в работе итогового приложения. Кроме того, были исследованы структуры DOM-дерева онлайн-ресурсов с целью выяснения возможности автоматизированного сбора данных и подбора источников данных для работы алгоритма. И на основе данных исследований был разработан алгоритм сбора данных о фармацевтической продукции компании.

## **Использование модуля Requests и сессий с подстановкой фиктивных заголовков**

Самый известный модуль языка программирования Python для работы с HTTP-запросами – это Requests. С помощью него можно устанавливать соединение с онлайн-ресурсами, посылать различные запросы и получать ответы. В частности, возможно послать запрос на получение структуры веб-страницы [15].

Работать с данным модулем весьма просто и понятно: необходимо послать запрос GET указанному онлайн-ресурсу с помощью его url-адреса, таким образом, установив соединение с ним, а затем указать, что именно надо получить в ответ от веб-страницы. Для дальнейшей работы приложения необходимо от сайта онлайн-аптеки получить его DOM-дерево (см. рисунок А.1).

Однако данный модуль весьма неэффективен для работы с онлайн-ресурсами. Так как он использует протокол передачи данных HTTP для запросов, а все веб-сайты уже давно перешли на протокол HTTPS, то лишь с очень малым количеством сайтов удается установить успешное соединение, что весьма затрудняет дальнейшую работу.

Проверить, успешно ли соединение, можно с помощью метода status\_code. Если ответ получился 200, то это означает, что соединение было установлено успешно, и запрос выполнился. Однако на данный момент подавляющее большинство сайтов хорошо защищено и умеет распознавать подобного рода запросы. Поэтому вместо кода 200 можно получить 403 (данный код означает, что дальнейшие запросы запрещены, так как у пользователя нет прав и доступа), 401 (пользователь, посылающий запрос, не авторизирован, следовательно, не может получить информацию), 511 (требуется сетевая аутентификация) и многие другие [2].

Решить данную проблему с передачей запросов можно с помощью библиотеки Requests.session [15]. Она позволяет создавать сессии с различными заголовками для установки соединения с веб-страницами.

Сессия – это временной промежуток, период которого охватывает время между тем, как пользователь впервые перешел по url-ссылке и до того момента, когда он закончил работать с онлайн-ресурсом. Она способствует качественной передачи данных между последовательными HTTP-запросами. У каждой сессии есть заголовки, которые позволяют клиенту и серверу отправлять дополнительную информацию с запросом или ответом. Бывают основные заголовки, которые применяются к запросам и ответам, а бывают отдельно заголовки запроса и ответа.

Проблема с доступом к содержимому веб-страницы решается с помощью правильно установленных заголовков сессии. При простом использовании модуля Requests никаких заголовков не передается в запросе, поэтому современная защита онлайн-ресурсов однозначно понимает, что с ней взаимодействует «неправильный» пользователь, и блокирует его доступ [2].

Необходимые заголовки сессии возможно узнать, посмотрев с помощью специальных программных средств, какая информация передается онлайн-ресурсу, когда посылается запрос GET. В качестве подобных программных средств могут выступать: инструменты для нагрузочного тестирования, инструменты для создания и отладки сайтов, встроенные практически в любой имеющийся браузер, инструменты для захвата и анализа сетевого трафика.

При разработке приложения для работы с данными используется инструмент для нагрузочного тестирования – Jmeter. Данное программное средство предоставляет простой и понятный интерфейс не только для работы с HTTP-запросами и информацией, получаемой после их выполнения, но и для анализа работы всего сервера при высокой нагрузке на нем [10].

Последовательность действий для того, чтобы узнать необходимые заголовки, заключается в следующем: в начале, необходимо выписать SSL-сертификат для того, чтобы без препятствий взаимодействовать с онлайн-ресурсами, работающими по протоколу безопасности HTTPS (см. рисунок А.2), затем создать объекты, с помощью которых будет устанавливаться соединение, отправляться HTTP-запросы и куда будут записаны все результаты работы (см. рисунок А.3). В конце, нужно включить работу прокси-серверов и указать порт данного инструмента (см. рисунок А.4). Информация о том, что передают в заголовках запроса, будет представлена в удобном для пользователя виде (см. рисунок А.5).

После того, как с помощью данного способа были проверены необходимые сайты онлайн-аптек, выяснилось, что почти все они передают одну и ту же информацию в заголовках запроса (за исключением cookies, которые для каждого онлайн-ресурса генерируются им лично), поэтому создание функции для получения DOM-дерева с помощью сессии и правильных заголовков упростилось.

Принцип работы указанной выше функции прост: создается объект сессии, для которого устанавливаются выясненные раньше заголовки, а затем онлайн-ресурсу передается запрос GET для получения его структуры (см. рисунок А.6).

Однако, как было выяснено в ходе исследований веб-страниц онлайн-ресурсов, данный способ тоже не всегда дает положительный результат. Самая частая ошибка, которая возникает при подобном способе получения структуры сайта, – это ошибка сертификации. Защита некоторых сайтов способна распознать сертификаты прокси-сервера и не дать установить соединение. Вторая часто встречающаяся проблема – это внутренняя защита сайта в виде особой кодировки, ключ от которой пользователь не имеет возможности получить. Более подробно данные сложности будут рассмотрены в следующем разделе, где будет проведен анализ онлайн-ресурсов и где будут подобраны сайты, удобны для получения информации о продукции компании.

В связи с подобными трудностями было принято решение рассмотреть еще одну библиотеку для работы с веб-ресурсами на языке программирования Python – Selenium [16].

## **Использование библиотеки Selenium и разработанного для нее пакета selenium\_stealth**

Selenium — это популярный инструмент для автоматизации действий в веб-браузере, который помогает выполнять различные задачи, такие как тестирование веб-приложений, веб-скрапинг и автоматизация повторяющихся действий. С его помощью можно взаимодействовать с интерфейсом веб-сайта. Например, имитируя действия пользователя, нажимать на кнопки, раскрывать списки, вводить данные и т.д. Однако чаще всего данную библиотеку не рекомендуют к использованию для скрапинга из-за больших временных затрат [2].

Но, так как данный инструмент давно существует, то современная защита онлайн-ресурсов научилась распознавать его использование. Selenium в основе своей работы использует веб-драйвер браузера, выбранного пользователем, у которого имеется свойство window.navigator.webdriver. Когда веб-драйвер используется с помощью Selenium, то значение этого свойства становится True, что передается в JavaScript сайту, и поэтому защита может распознать его использование.

Для того, чтобы современная защита онлайн-ресурсов при подключении не распознавала данную библиотеку, было принято решение совместно с ней использовать созданный недавно пакет Selenium\_stealth — специальный пакет Python для библиотеки Selenium, который позволяет маскировать использование WebDriver [17].

Таким образом, принцип работы функции для получения DOM-дерева веб-страницы аптеки на основе Selenium следующий: создаем объект веб-драйвера и передаем ему необходимые опции для оптимизации работы (см. рисунок А.7). Как уже было сказано выше данную библиотеку специалисты не рекомендуют использовать для скрапинга данных из-за больших временных затрат. Но ее работу можно ускорить, отключив или настроив некоторые параметры и свойства веб-драйвера. Так, были отключены следующие свойства: голова веб-драйвера, которая позволяет запускать браузер и отражать все действия пользователя в нем, расширения и «песочница» браузера, а так же свойства, которые сигнализируют о том, что данный веб-драйвер используется с помощью Selenium. Данные действия позволят значительно ускорить процесс получения структуры веб-страницы онлайн-ресурса [2].

В конце необходимо скрыть веб-драйвер с помощью пакета Selenium\_stealth, после чего можно передать запрос и получить необходимую информацию, а именно структуру веб-страницы (см. рисунок А.8).

Способ получения структуры сайта с помощью сессии работает быстрее, но не всегда дает возможность подключения к онлайн-ресурсам из-за проблем, описанных выше, а способ с помощью Selenium такую возможность предоставляет, но работает значительно медленнее предыдущего. Окончательное решение о том, какой способ целесообразнее использовать в приложении, будет принято в следующем подразделе после изучения и анализа структуры онлайн-ресурсов.

## **Изучение и анализ структуры источников данных**

Изучение и анализ онлайн-ресурсов проводились с целью выяснения возможности автоматизированного сбора данных. Такая возможность была выявлена, но лишь на определенной категории сайтов, о которой будет рассказано в конце данного раздела, после описания всех ситуаций, возникших в ходе исследования.

Первый этап исследования заключался в том, что онлайн-источники были проверены на возможность веб-скрапинга описанными выше способами. Часть результатов работы представлена в таблице А.1. В ходе изучения и анализа структуры источников данных было рассмотрено более 50 сайтов онлайн-аптек, но так как из раза в раз сложности работы и ошибки первого этапа повторялись, то было принято решение в таблицу вынести некоторые примеры данных ошибок, а также url-адреса тех источников, которые не вызвали никаких проблем.

Самая распространенная проблема, возникающая при попытке получить DOM-дерево веб-страницы аптеки, *–* это капча, или проверка пользователя на то, является ли он человеком, а не роботом (ботом) [2]. Как уже было сказано, благодаря развитым современным технологиям защита большинства веб-ресурсов довольно серьезная, и ее невозможно обойти никакими программными средствами. Соответственно, данную проверку также невозможно обойти, что делает онлайн-ресурс неподходящим для скрапинга данных.

Вторая по частоте встречаемости – это невозможность некоторых веб-страниц выдержать несколько последовательных запросов. Несомненно, можно сделать так, чтобы между последовательными запросами проходил некоторый промежуток времени, за который онлайн-ресурс сумеет восстановиться. Однако, во-первых, для каждого сайта аптеки этот промежуток будет разным, что сводит на ноль автоматизацию работы алгоритма сбора карточек товаров, во-вторых, эта задержка будет увеличивать и без того достаточно большое время работы приложения, что является невыгодным. Следовательно, данные ресурсы рассматривать не стоит.

Не менее частая проблема – это скрипты. Данная проблема была обнаружена с помощью Chrome Developer Tools – набора инструментов, встроенных в браузер Google Chrome, для создания и отладки сайтов [12]. С их помощью работа по изучению и анализу структуры сайтов стала заметно проще. Как известно, большинство современных веб-ресурсов написано не только с помощью языков HTML и CSS, но и с помощью языка программирования JavaScript. Он предназначен для управления элементами на сайте, и с помощью него можно писать сложные функции и действия, которые позволят «украсить» веб-сайт [2]. Таким образом, данная проблема заключается в следующем: большинство элементов веб-страницы, включая строку поиска и карточки товаров, «отрисовано» с его помощью, поэтому они не присутствуют в DOM-дереве, что делает их поиск невозможным.

И последняя из самых частых проблем, которую тоже можно обнаружить с помощью Chrome Developer Tools, – это ошибки создателей веб-страниц в описании товаров, а также неподходящее описание карточек товаров. Ошибки могут быть любые: лекарственный препарат, произведенный одной фармацевтической компанией, приписывают другой, ошибки в написании названия лекарственного препарата, ссылка на другое лекарственное средство, а не на то, что указано в заголовке и т.д. Но все они делают онлайн-ресурсы неподходящими для работы алгоритма, так как значительно затрудняют поиск продукции. Под неподходящим описанием карточек товаров имеется в виду следующее: в ходе работы алгоритма на веб-странице с результатами поиска находятся ссылки, которые содержат название соответствующего лекарственного препарата, но у некоторых источников данных в ссылках нет ничего, чтобы указывало на данный препарат, что делает его поиск невозможным. Однако стоит сказать, что в ходе исследований были встречены очень редкие ошибки: ошибка сертификации и ошибка кодировки. Первая заключалась в том, что защита некоторых веб-ресурсов была способна распознать SSL-сертификат, выписанный специально для прокси-сервера JMeter, поэтому доступ к данному ресурсу был запрещен (см. рисунок 3.1). Вторая заключалась в том, что у защиты сайта аптеки была своя собственная кодировка, ключи от которой генерировались при каждом заходе пользователя на браузер, поэтому при попытке скрапинга данных информация передавалась в запросе, но без ключа кодировки, что делало ее невозможным для прочтения (см. рисунок 3.2).

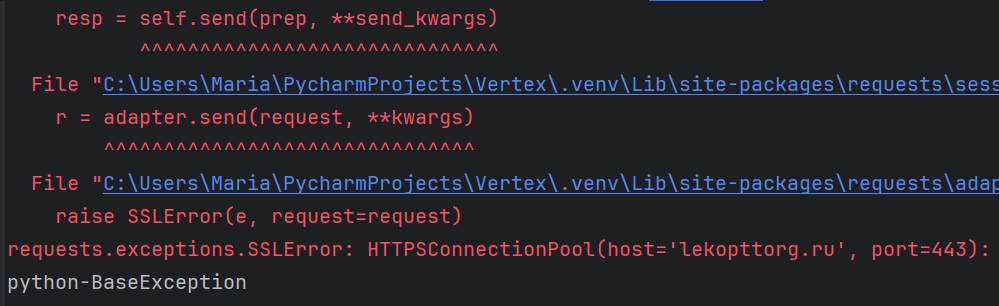


Рисунок 3.1 – Ошибка сертификации

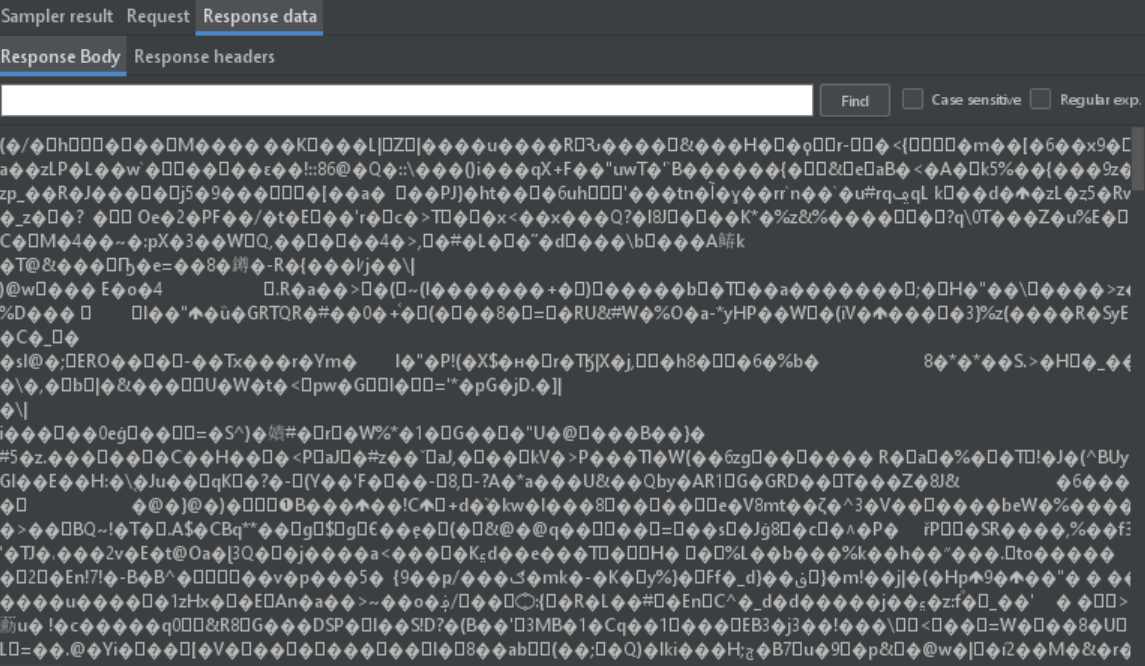


Рисунок 3.2 – Ошибка кодировки

Таким образом, после изучения и анализа более 50 веб-ресурсов были выбраны онлайн-ресурсы, при работе с которыми не возникают приведенные выше ошибки и проблемы. Последний шаг по исследованию источников данных – это изучение структуры с помощью Chrome Developer Tools, чтобы выявить закономерности для создания алгоритма.

И первая закономерность, которая стала видна при изучении, – это структура строки поиска. Для того, чтобы с ее помощью построить url-адрес поискового запроса, важно, чтобы она был следующего вида, показанного на рисунке 3.3. Во-первых, элемент с тегом input, обозначающий поисковую строку, должен содержать в значениях своих атрибутов или в наименованиях атрибутов слова «найти», «поиск», «искать», «search» и другие вариации слов, как на русском языке, так и на английском. Это поможет найти его в DOM-дереве без необходимости в обходе всего дерева. Во-вторых, данный элемент должен иметь атрибут name, который участвует в построении адреса поискового запроса. В-третьих, элемент с тегом input должен лежать в элементе с тэгом form, у которого среди атрибутов есть action и метод GET, которые также участвуют в построении адреса поискового запроса. Как можно понять – это идеальный случай и таких структур очень мало. Могут быть отклонения такие, как отсутствие некоторых необходимых атрибутов, метода GET или элемента form. Главное, чтобы в элементе поисковой строки содержались все возможные варианты слов «поиск» и «найти».

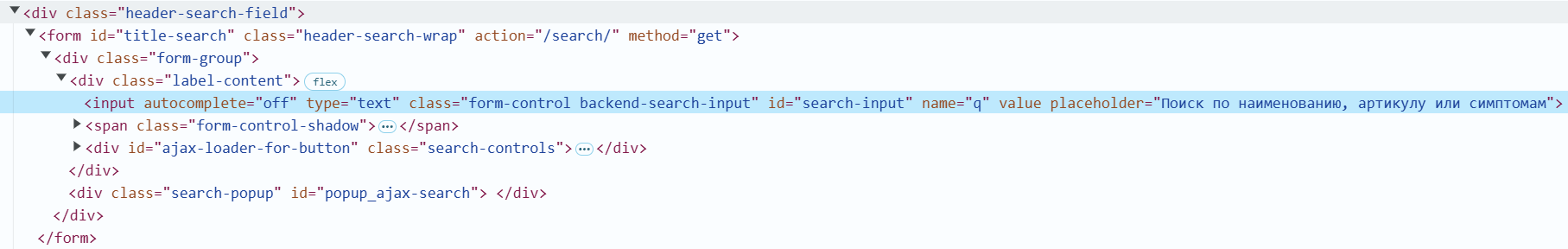


Рисунок 3.3 – Идеальная структура поисковой строки

Вторая закономерность – это структура карточек товаров на странице с результатами поискового запроса, ее пример можно увидеть на рисунке 3.4. Во-первых, в ссылке, т.е. в элементе с тэгом a, которая ведет на карточку с товаром, должно быть название данного препарата или действующего вещества, чтобы при поиске она была найдена. Как уже говорилось чуть выше, на этой закономерности и возникает ошибки, связанные с неправильными названиями или ссылками. Во-вторых, в карточке должен быть указан производитель товара, причем либо в отдельном тэге, либо прямо в ссылке.



Рисунок 3.4 – Необходимая структура карточки товара на странице с результатами поискового запроса

Третья закономерность – это кнопки для переключения между страницами с результатами поискового запроса. Их может и не быть, но вот когда они имеются, то должны выглядеть примерно таким же образом, как представлены на рисунке 3.5. У их элементов среди атрибутов должны быть слова «page», «pagination» и другие подобные. А в их ссылках должны быть слова «след», «следующая», «вперед» или «next», чтобы среди них можно было найти следующую страницу результатов. Иногда могут быть слова «navig» или «navigation» вместо «page» и других.



Рисунок 3.5 – Пример кнопок переключения между страницами с результатами запроса

И последняя, четвертая закономерность, - это наличие в карточке товара элемента, обозначающего цену лекарственного препарата. Он должен содержать слово «price» в обязательно порядке, а вот остальные слова, которые бы выделяли данный элемент на фоне остальных, для каждого онлайн-ресурса будут разные. Пример подобного элемента можно увидеть на рисунке 3.6.

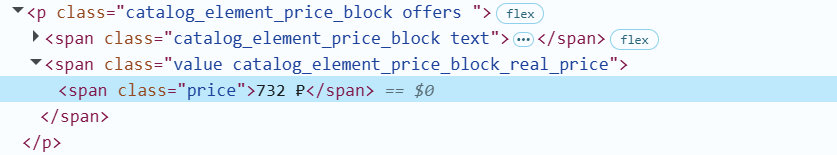


Рисунок 3.6 – Пример элемента в карточке товара, обозначающего цену препарата

На данном этапе все онлайн-источники, подобранные на прошлом этапе, были проверены на выявленные закономерности. И среди них были отобраны те, которые подходят под все указанные выше требования.

Выбранные источники данных, с которыми не возникает проблем при попытке веб-скрапинга и чья структура соответствует требованиям закономерностей, представлены в таблице 3.1. Также в таблице 3.1 приведено сравнение времени работы алгоритма поиска карточек товаров косметической продукции компании (т.е. 8 наименований брендов) с использованием функции подключения с помощью Request.session и с использованием функции подключения с помощью Selenium.

Таблица 3.1 - Перечень подходящих сайтов и сравнение времени работы алгоритма с двумя способами подключения

| **URL-адрес сайта** | **Requests.session** | **Selenium** |
| --- | --- | --- |
| https://wer.ru | 1 мин 47 с | 2 мин 32 с |
| https://vn1.ru | 2 мин 44 с | 4 мин 02 с |
| https://aptekaantey.ru | 1 мин 24 с | 2 мин 53 с |
| https://www.gorapteka.ru | 1 мин 52 с | 3 мин 06 с |
| http://spb.vapteke.ru | 4 мин 36 с | 6 мин 13 с |
| https://megapteka.ru | 1 мин 41 с | 2 мин 58 с |

Исходя из результатов сравнения, было принято решение в приложении использовать поиск структуры онлайн-ресурса с помощью создания сессии и подбора для нее заголовков, потому что, во-первых, все подходящие источники данных работают без указанных выше ошибок, а, во-вторых, данный способ работает быстрее, что делает работу приложения эффективнее.

## **Реализация алгоритма**

В целом процесс работы алгоритма представлен на рисунке 3.7. Его можно описать следующим образом: на главной веб-странице онлайн-ресурса алгоритм находит поисковую строку, с помощью структуры которой и словаря фармацевтической продукции строится url-адрес поискового запроса, затем алгоритм переходит на страницу с результатами поискового запроса. На данной странице он ищет необходимые карточки товаров, учитывая переключение между страницами с результатами поиска. А затем он рассматривает каждую найденную карточку товара и парсит из нее информацию о производителе лекарственного препарата, цене и полном наименовании. Алгоритм сохраняет всю найденную информацию в базу данных в соответствующую итоговую таблицу. Словарь фармацевтической продукции получен из конфигурационных файлов, которые были сделаны для упрощения работы, а именно: если пользователь захочет добавить новый продукт компании, то он сможет это сделать, просто записав информацию в соответствующий файл. Также в базе данных есть таблица с фармацевтическими производителями, информация о которых была найдена в процессе веб-скрапинга. Она, во-первых, используется для того, чтобы в итоговую таблицу по соответствующим поисковым словам подставлять необходимого производителя, а, во-вторых, в процессе могут быть найдены новые производители, и информация о них сохраняется в данную таблицу для дальнейшей обработки. Причем алгоритм в процессе работы предупреждает, если был найден новый производитель.



Рисунок 3.7 – Блок-схема работы алгоритма сбора данных

Теперь будет рассмотрен каждый этап работы алгоритма и его реализация.

Первое, с чего стоит начать описание реализации алгоритма, – это глобальные переменные. Одни из них – это словарь фармацевтической продукции, словарь с категориями лекарственными препаратами и словарь с онлайн-ресурсами. Каждый из них получен из соответствующего конфигурационного файла, пример одного из них можно увидеть на рисунке А.9. Как происходит процесс присваивания значений данным словарям, изображено на рисунке А.10. И еще одна не менее важная переменная – это таблица с фармацевтическими производителями, полученная из базы данных. Для создания подключения к базе данных и дальнейшей работе с ней использовалась библиотека SQLAlchemy [18]. Данная библиотека позволяет работать с базами данных в среде программирования Python, смотреть таблицы, писать запросы, получать их результаты и т.д. Значения из данной таблицы получены следующим образом: сначала создано соединение с базой данных, а затем с помощью модуля Pandas и специальной функции из него получена таблица с фармацевтическими производителями в формате Dataframe. Пример данных из таблицы с фармацевтическими производителями можно увидеть на рисунке 3.8. А весь процесс создания соединения и получения таблицы изображен на рисунке А.11.

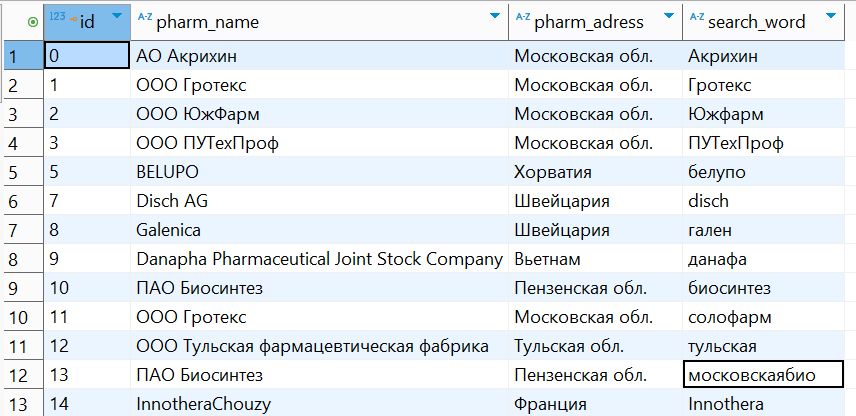


Рисунок 3.8 – Пример данных из таблицы с производителями из базы данных

Весь процесс работы алгоритма реализован с помощью двух функций: главной функции алгоритма и вспомогательной функции алгоритма. В дальнейшем они именно так и будут именоваться. Первая функция отвечает за работу всего алгоритма, а вспомогательная функция ищет карточки товаров на текущем онлайн-ресурсе.

Первый этап работы главной функции алгоритма – это присвоение значений локальным переменным и внедрение глобальных переменных в функцию. Среди локальных переменных присутствуют: результирующая таблица с собранными данными, информация из которой будет перенесена в итоговую таблицу в базе данных, списки с поисковыми словами для нахождения необходимых элементов в DOM-дереве онлайн-ресурса (это закономерности, описанные в подразделе 3.3), а так же текущая дата. Посмотреть локальные переменные главной функции алгоритма можно на рисунке А.12.

Второй этап – это запуск цикла по всем онлайн-источникам из списка, который был дан в качестве аргумента главной функции алгоритма, нахождение элемента поисковой строки в DOM-дереве текущего источника и построение части url-адреса поискового запроса.

Процесс нахождения элемента поисковой строки происходит следующим образом: в словаре онлайн-ресурсов находится url-адрес соответствующего ресурса, а затем с помощью описанного выше способа веб-скрапинга изымается DOM-дерево источника. Работа с деревом онлайн-источника происходит с помощью библиотеки BeautifulSoup [11]. Это библиотека для работы с XTML-документами и DOM-деревьями, с ее помощью можно находить элементы, строить css-запросы, работать с родителями, детьми, соседями элементов и т.д [2]. Данная библиотека будет незаменима на протяжении работы всего алгоритма. Так вот, после изъятия DOM-дерево в нем происходит поиск элемента, обозначающего поисковую строку. Как говорилось в подразделе выше, для этого с помощью селектора находится элемент с тэгом input, который обладает «идеальной» структурой, пример которой можно увидеть на рисунке 3.3. Если подобный не найден, то рассматриваются другие возможные случаи. Найденный элемент и связанные с ним элементы (например, элемент из «идеальной» структуры с тэгом form) запоминаются. Процесс поиска можно увидеть на рисунке А.13.

Последнее действие второго этапа – это построение части url-адреса поискового запроса. Эта часть будет одинакова для определенного онлайн-ресурса, меняться в полном адресе будет лишь поисковое слово для лекарственного препарата. Часть будет строиться на основе найденного элемента поисковой строки. В ходе исследований было выяснено, что поисковый запрос строится следующим образом: url-адрес источника + action + name + слово поиска, где action и name – это значения соответствующих атрибутов (все описано в подразделе 3.3). Поэтому сначала на основе найденного элемента поисковой строки будет собрана часть адреса поискового запроса, а именно: из элементов input и form (при наличии) достанут значения соответствующих атрибутов. Весь описанный процесс можно увидеть на рисунке А.14.

Третий этап работы главной функции алгоритма – это запуск цикла по лекарственным препаратам, которые указаны в аргументе функции, построение полного адреса поискового запроса и запуск второстепенной функции алгоритма, которая ищет карточки товаров. Построение полного адреса происходит следующим образом: к уже имеющейся части добавляется поисковое слово для данного препарата, которое берется из словаря фармацевтической продукции. Причем, если поисковое слово состоит из нескольких слов, то они записываются в адрес через «+», чтобы было выяснено в ходе исследований. А затем запускается функция поиска карточек товаров на странице с результатами поискового запроса. Правда процесс слегка усложняется тем, что есть онлайн-источники, где при данном поисковом слове находятся все возможные препараты, а есть те, где на поисковое слово находятся только соответствующие ему препараты. Поэтому на некоторых ищут по действующему веществу, а где-то по поисковым словам. Весь данный процесс можно увидеть на рисунках А.15 и А.16. Стоит упомянуть, что перед построением полного адреса еще происходит поиск категории лекарственного препарата из соответствующего словаря, чтобы добавить ее в итоговую таблицу.

Второстепенная функция начинает свою работу с того, что с помощью выбранного способа веб-скрапинга изымает DOM-дерево страницы с результатами поискового запроса по указанному url-адресу, что можно увидеть на рисунке А.17.

Первый этап работы второстепенной функции алгоритма – это построение списка поисковых слов для нахождения правильных карточек товаров. Это важно, так как названия препаратов в ссылках на них на онлайн-ресурсах могут быть названы по-разному. Например, могут быть названы по-русски или по-английски. Если название состоит из двух слов, то они могут быть написаны через дефис или через подчеркивание. Поэтому важно учитывать все возможные варианты написания, чтобы найти все подходящие карточки товаров и никакие не пропустить. Построение списка происходит по-разному: если рассматривается онлайн-источник, где необходимые карточки можно найти по действующему веществу, то тогда в список добавляются только все возможные варианты написания данного действующего вещества, а вот если на онлайн-источнике необходимо искать именно по названиям, то в список добавляются все названия препаратов и их все возможные вариации. Реализация данного этапа изображена на рисунке А.18.

Дальше начинается второй этап работы второстепенной функции – это поиск карточек товаров и переключение между страницами с результатами поискового запроса. Для этого запускается бесконечный цикл, который заканчивается в том случае, когда у нас не будет больше кнопок переключения для того, чтобы перейти на следующую страницу с результатами. На каждой итерации цикла происходит следующее: в DOM-дереве страницы с результатами поискового запроса находятся все элементы- ссылки, т.е. элементы с тэгом а, среди которых ищутся карточки товаров. Они находятся по списку поисковых слов, составленному на первом этапе работы алгоритма. А затем отсеиваются те ссылки, которые являются отзывами на товары, фильтрами или еще какими-нибудь посторонними элементами. Последнее действие – это отсеивание одинаковых ссылок, так как иногда ссылка на изображение и на название является одной и той же.

Затем, после нахождения карточек товаров на текущей странице с результатами запроса, проверяется первая возможная ситуация с кнопками переключения между страницами. А именно: находятся все элементы-ссылки, которые содержат слова из соответствующих списков (это третья закономерность из подраздела 3.3), и если подобных элементов нет, то проверяется наличие элементов-списков, которые содержат подобные слова. Если элементы списки находятся, то они запоминаются, а если нет, то список элементов остается пустым, но и в том, и в другом случае вспомогательная функция переходит ко второй возможной ситуации с кнопками пересечения. Процесс нахождения карточек товаров и первую ситуацию можно увидеть на рисунке А.19.

Вторая возможная ситуация заключается в следующем: у нас не нашлось никаких элементов. Это может быть в следующих случаях: во-первых, на странице вообще нет кнопок переключения, так как результатов слишком мало, во-вторых, на странице нет кнопок переключения в виде элементов-ссылок или элементов-списков, которые содержат заданные слова, но есть другие элементы и, в-третьих, алгоритм прошел по всем возможным кнопкам и посмотрел все возможные страницы с результатами поиска. Поэтому в данной ситуации проверяется три подобных случая. Сначала проверяется, были ли найдены вообще кнопки переключения, и если нет, то ищутся элементы с тэгом div, которые содержат разнообразные вариации слова «навигация», а если и таких элементов нет, то цикл заканчивается, так как кнопок переключения по данному запросу не существует. Если же уже были найдены какие-то кнопки переключения раньше, то цикл, соответственно, тоже заканчивает работу. А вот если были найдены элементы с тэгом div, то из них изымается ссылка на следующую страницу с результатами поиска, изымается ее DOM-дерево и цикл переходит на следующую итерацию. Реализацию второй ситуации среди кнопок переключения между страницами с результатами поиска можно увидеть на рисунке А.20.

Третья ситуация с кнопками переключения – это нахождение ровно одного элемента-ссылки, подходящего под слова из списка. И тут возможны тоже разные варианты: иногда, когда достигнут конец в кнопках переключения, то есть просмотрены все возможные страницы с результатами, кнопка переключения на следующую страницу становиться недействительной, так как у нее отсутствует ссылка или присутствует атрибут disabled, поэтому в данном случае необходимо закончить цикл. А вот если найденная кнопка активна, то стоит повторить все то же самое, что и во второй ситуации: достать ссылку на следующую страницу, перейти и изъять DOM-дерево. Третья ситуация изображена на рисунке А.21.

И последняя, четвертая, ситуация – это нахождение нескольких подходящих элементов ссылок. В данной ситуации алгоритм действует следующим образом: среди найденных элементов отсеивает те, которые недействительны. Если после этого не осталось элементов, то цикл следует заканчивать, а если остались, то берется самый первый и совершаются те же действия, что и в ситуации выше, с переходом на следующую итерацию цикла. Ситуацию можно увидеть на рисунке А.22.

После окончания работы цикла вспомогательная функция алгоритма возвращает список всех найденных карточек товаров, ссылок на них.

Четвертый этап главной функции алгоритма – запуск цикла по найденным карточкам товаров, парсинг необходимых данных из каждой и импорт собранных данных в итоговую таблицу в базе данных.

Сначала происходит парсинг полного наименования лекарственного препарата. Это реализуется очень просто – чаще всего наименования содержатся в элементах с тэгом h1, который стоит на первом месте в DOM-дереве. Поэтому для каждой карточки строится ее url-адрес, а затем с помощью выбранного способа веб-скрапинга по данному адресу изымается дерево веб-страницы карточки товара. И в данном дереве с помощью селектора находятся описанные выше элементы, что можно увидеть на рисунке А.23.

Затем алгоритмом находится цена товара. Для этого в DOM-дереве веб-страницы карточки товара находятся элементы, которые могут содержать информацию о цене препарата. Это элементы с тэгом span или div, которые могут содержать различные комбинации со словом «price» (четвертая закономерность из раздела выше). Данные комбинации были выяснены в ходе исследования структуры онлайн-ресурсов, поэтому при реализации поиска элемента цены они все перебираются, что можно увидеть на рисунке А.24. Если не найден элемент, или он не содержит никакой информации, то вместо цены у препарата будет null.

После поиска элемента цены, необходимо извлечь из него информацию. Реализуется это с помощью регулярных выражений и модуля для работы с ними Re [14]. Рассматриваются все возможные ситуации, которые также были выяснены в ходе исследования онлайн-ресурсов. Где-то цена может быть написана через пробел, где-то через неразрывный пробел. Какие-то онлайн-источники пишут копейки через точку, а какие-то через запятую. Необходимо учесть все ситуации при поиске информации о цене препарата, что изображено на рисунке А.25.

И, наконец, последнее, что необходимо сделать на данном этапе – это спарсить информацию о фармацевтическом производителе в карточке товара. Как уже было сказано выше, это парсинг данных, который будет уникален для каждого онлайн-ресурса. Где-то информация о производителе написана в элементах с тэгом h2, где-то в элементах с тэгом div, а на некоторых источниках она вообще указана в полном наименовании лекарственного препарата. Поэтому поиск будет уникален для каждого онлайн-источника. Реализацию данных вариантов можно увидеть на рисунках А.26, А.27 и А.28.

А затем с найденной информацией о производителе происходит следующее: в тексте ищутся поисковые слова из таблицы с фармацевтическими производителями, которая является глобальной переменной. Если поисковое слово находится, то из таблицы подставляется то наименование, которое этому поисковому слову соответствует. Сделано это затем, что на каждом онлайн-источнике производители называются по-разному, а для дальнейшей работы с данными необходимо, чтобы одни и те же производители были записаны одинаково. Поэтому в базе была создана таблица, где напротив правильного наименования каждого производителя стоит поисковое слово. А если поисковое слово не найдено, то в таблицу записывается информация о новом найденном производителе, причем алгоритм предупреждает о том, что была записана информация. Весь описанный процесс можно увидеть на рисунке А.29. Так же на данном рисунке изображен процесс записи данных в промежуточную таблицу, а именно в порядке: дата обращения, источник, категория, действующее вещество, производитель, наименование и цена препарата.

Завершающее действие главной функции и всего алгоритма – это запись собранных данных из промежуточной таблицы в итоговую таблицу в базу данных. Для этого точно также создаем соединение и с помощью метода INSERT вставляем необходимые данные в итоговую таблицу. Весь этот процесс изображен на рисунке А.30.

В поиске элементов DOM-дерева, которые среди атрибутов или содержания атрибутов содержат необходимые слова, поиске кнопок переключения между страницами с результатами поискового запроса, в отсеивании ненужных элементов среди выбранных алгоритму помогают вспомогательные функции, которые можно увидеть на рисунках А.31, А.32 и А.33.

В итоге работы алгоритма получается следующая таблица, которую можно увидеть на рисунке 3.9.

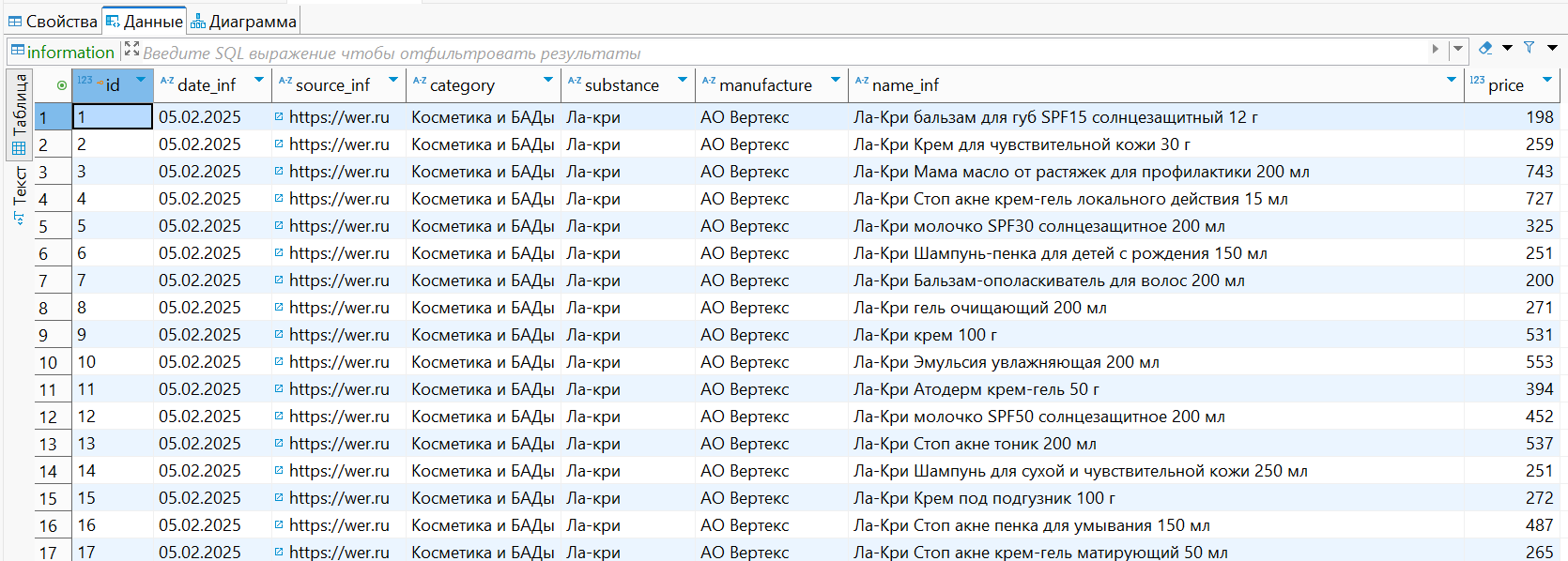


Рисунок 3.9 – Пример собранных данных

Собранные данные получаются отформатированными и подходящими для дальнейшего исследования и анализа.

## **Выводы по разделу 3**

Как уже было сказано в начале данного раздела, в ходе изучения и анализа источников данных была выявлена категория онлайн-ресурсов, с которой практически не возникало сложностей при попытке получения DOM-дерева и при дальнейшей работе алгоритма сбора карточек товаров. Онлайн-источники, принадлежащие данной категории, представлены и в таблице 3.1.

Кроме того, были изучены и проанализированы различные способы веб-скрапинга данных, среди которых был выбран наиболее эффективный для работы приложения. Помимо этого, в результате исследований DOM-деревьев онлайн-ресурсов была выявлена возможность автоматизированного сбора данных и был намечен план с этапами разработки алгоритма.

Данный план был полностью реализован в алгоритме, который возвращает отформатированные данные, удобные для дальнейшего использования.

# **РЕАЛИЗАЦИЯ ВЕБ-ПРИЛОЖЕНИЯ ДЛЯ ОБЕСПЕЧЕНИЯ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ**

В данном разделе были изучены собранные данные о фармацевтической продукции в онлайн-сегменте фармацевтического рынка России и финансовая отчетность компании АО «Вертекс». Кроме того, был проведен анализ имеющихся данных, включая общий анализ фармацевтического рынка, анализ ценообразования отдельных лекарственных препаратов и внешний финансовый анализ деятельности компании. Для удобства работы как с реализованным алгоритмом, так и с проведенными различными анализами было разработано веб-приложение с пользовательским графическим интерфейсом.

Для создания веб-приложения используется фреймворк Dash, библиотека для которого полностью реализована для языка программирования Python. Он позволяет создавать интерактивные графические интерфейсы, где пользователи могут взаимодействовать с элементами приложения, графиками, проводить аналитику данных и т.д. Самое главное, что данный фреймворк не требует глубоких знаний HTML и CSS, достаточно просто иметь базовые навыки в данной области. Его работа связана с библиотеками Plotly, Flask и React, поэтому пользователю предоставлено множество возможностей для создания разнообразных графиков и красочных графических интерфейсов.

## **Архитектура приложения**

Полную архитектуру веб-приложения можно увидеть на рисунке 4.1. У него есть основное тело, состоящее из трех отделов.

Первый отдел отвечает за обновление информации об онлайн-продаже фармацевтической продукции как АО «Вертекс», так и других производителей на текущую дату. Соответственно, данный раздел связан с разработанным и реализованным алгоритмом сбора данных, для работы которого необходимы реализация способа веб-скрапинга, вспомогательные функции поиска и отсеивания элементов, база данных и конфигурационные файлы, описанные в разделе выше.

Второй отдел отвечает за возможность просмотра данных из базы данных о фармацевтической продукции на онлайн-рынке и бухгалтерской отчетности компании. Кроме того, пользователь имеет возможность скачать те данные, которые ему нужны для работы. Для работы отдела необходимы база данных и файлы формата Excel, которые содержат всю финансовую информацию об отчетности. Для удобства все данные вынесены в отдельный файл формата Python.

И последний, третий, отдел отвечает за аналитику. Он в свою очередь тоже разделен на три подотдела. В первом можно увидеть общий анализ фармацевтического рынка по самым главным показателям, которые интересуют компанию АО «Вертекс» и необходимы им для принятия решений. Второй дает пользователю возможность получить анализ ценообразования для конкретного лекарственного препарата, исходя из его различных дозировок. И третий – это внешний финансовый анализ по тем отчетностям, которые удалось получить со сторонних ресурсов. Для работы данного отдела необходим конфигурационный файл, который содержит информацию о дозировках для всех возможных лекарственных препаратов, а также графики, данные для которых рассчитаны и построены на основе исходных. Рассчитанные данные и графики вынесены в отдельный файл формата Python.

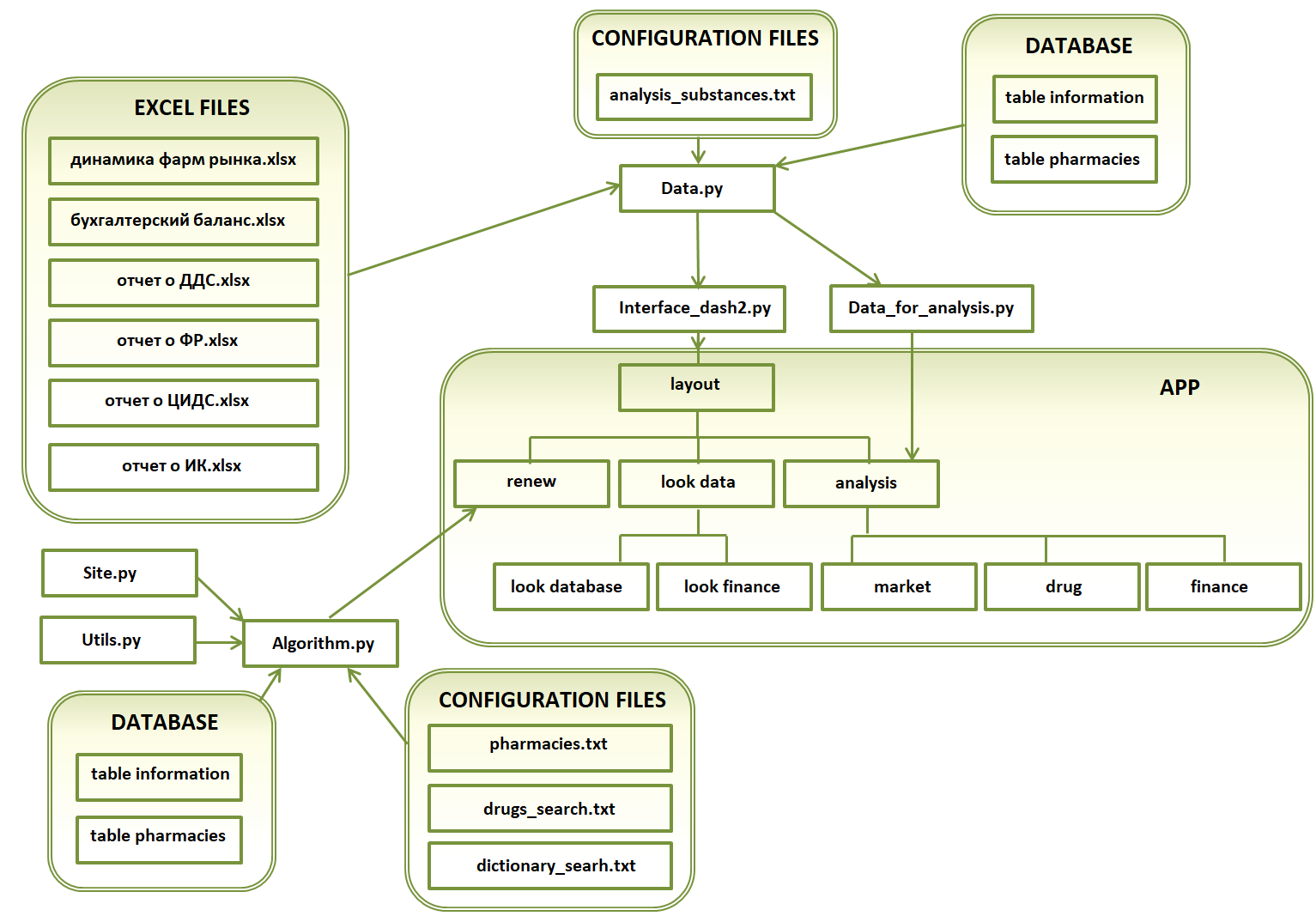


Рисунок 4.1 – Архитектура веб-приложения

В следующих разделах будет более подробно рассказано про реализацию каждого из описанных отделов веб-приложения.

## **Глобальные переменные**

Самые первые переменные, о которых стоит сказать, – это словарь дозировок для лекарственных препаратов и информация из базы данных. Для словаря данные содержатся в конфигурационном файле и представляют собой следующее: ключ – это действующее вещество, а значение – это список всех возможных дозировок данного препарата. Пример информации из конфигурационного файла можно увидеть на рисунке Б.1. Подключение же к базе данных и чтение информации из нее реализовано точно так же, как и описано в разделе выше, когда было рассказано про работу алгоритма сбора данных [18]. Сначала создается подключение к базе данных, а затем с помощью специальных функций модуля pandas и данного подключения из базы считывается необходимая информация: таблица с собранными данными о фармацевтической продукции, списки всех возможных категорий, действующих веществ и фармацевтических производителей лекарственных препаратов. Весь процесс чтения информации для описанных глобальных переменных изображен на рисунке Б.2.

Все оставшиеся глобальные переменные – это таблицы различных бухгалтерских отчетов и динамики фармацевтического рынка. Они с помощью встроенных функций модуля pandas считываются в переменные, что изображено на рисунках Б.3 и Б.4. Было принято решение каждую таблицу из каждого отчета считать в отдельную переменную, чтобы упростить дальнейшую работу с ними. Пример одной из таких таблиц можно увидеть на рисунке 4.2.

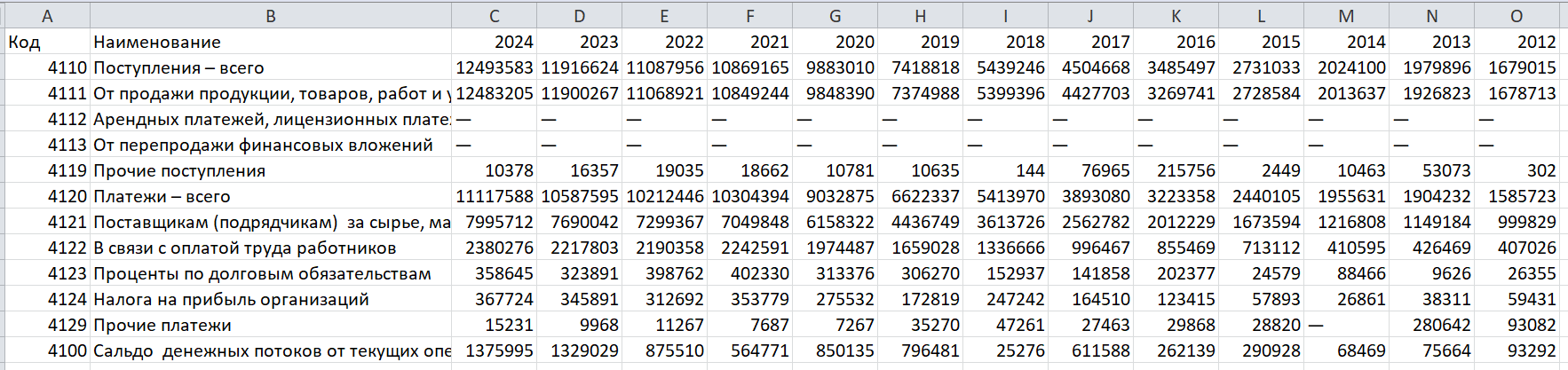


Рисунок 4.2 – Пример одной из таблиц из финансовой отчетности

Таким образом, были получены глобальные переменные, содержащие данные для дальнейшей работы, а именно для второго отдела, где их можно будет посмотреть и скачать, и для третьего отдела, где будет проведен их анализ для поддержки принятия решений.

## **Первый отдел – обновление данных о фармацевтической продукции**

Работа всего веб-приложения начинается с того, что создается объект Dash, который и является главным его телом [13]. Помимо создания объекта были скачаны специальные стили для него и присвоены объекту с помощью специального атрибута. Конечно, можно не скачивать данные стили, так как библиотека dash предлагает встроенные, но у них гораздо меньше возможностей для реализации красочного интерфейса, который предлагал бы различные возможности взаимодействия с элементами.

Приложения, которые написаны на Dash, делятся на две части: layout – это первая часть, которая отвечает за тело веб-приложения, т.е. за его внешний вид, и callbacks – это функции, которые отвечают за взаимодействия с написанным интерфейсом [13].

В первой части содержатся все элементы приложения: HTML-элементы, которые можно увидеть в любом DOM-дереве онлайн-ресурсов и на которых строится весь внешний вид ресурса [16], и Core-элементы, которые отвечают за взаимодействия с пользователем и отражения данных [13].

А во второй части реализованы функции изменения интерфейса приложения при каком-либо взаимодействии с ним. Например, при нажатии на определенную кнопку пользователем в соответствующем элементе Div должно поменяться содержимое. И в данной части будет описана функция, которая соответствующим образом поменяет содержимое, если пользователь взаимодействует с кнопкой.

Поэтому после того, как был создан объект веб-приложения, необходимо реализовать его первую часть – внешний вид. Все элементы реализуемого интерфейса будут находиться в основном элементе с тэгом Div, как в своеобразном контейнере. И самый первый из них – это изображение логотипа компании АО «Вертекс», который будет подтверждать, что данное приложение было разработано специально для нее. Следующий и последний элемент, который находится в основном, – это элемент Tabs. Он является Core-элементом и отвечает за создание вкладок в веб-приложение, между которыми можно переключаться и смотреть различную информацию в зависимости от того, какая из вкладок сейчас просматривается. Как уже было сказано ранее, в данном реализуемом приложении присутствуют три отдела, поэтому в описанном выше элементе будет три вкладки.

Первая вкладка – это отдел, который отвечает за обновление данных. Его визуальное оформление будет весьма простым: пользователь будет видеть надпись, которая предлагает обновить данные, и кнопку, нажав на которую, можно запустить процесс обновления, т.е. запустить разработанный алгоритм. Создание объекта веб-приложения и начало работы с его внешним видом, а конкретнее с первым отделом, можно увидеть на рисунке Б.5. Каждому элементу присвоен собственный стиль, который создается на основе CSS. Для некоторых элементов будут описаны неординарные границы и отступы, для других – цвета, шрифты и так далее, а для третьих вообще созданы гибкие контейнеры, предоставляющие возможность самого разнообразного размещения элементов. В веб-приложении большую роль будут играть цвета, которые подобраны максимально похожими на цвета логотипа компании АО «Вертекс». Было принято решение, чтобы все элементы, отвечающие за вкладки, имели такой же оттенок синего цвета, что и на логотипе. А все контейнеры с информацией – тот же оттенок зеленого.

Самое главное в работе первого отдела – это реализованный для него callback, отвечающий за запуск алгоритма. Самое важное в создании подобных функций – это определение, на взаимодействие с каким элементом она будет реагировать [13]. Данный элемент будет фигурировать в специальном объекте callback, который называется Input. Он должен быть у каждой функции и должен содержать среди аргументов наименование атрибута id того элемента, с которым необходимо взаимодействовать, а также наименование того атрибута, на взаимодействие с которым функция будет реагировать. Что это означает: во-первых, атрибут id можно присвоить любому элементу библиотеки Dash вне зависимости от его принадлежности к одному из двух видов, описанных выше. Любой элемент может спокойно обойтись без этого атрибута, и приложение все равно будет работать, если у него нет callbacks. Однако важно знать, что если у приложения есть вторая часть, то тогда те элементы, с которыми пользователь может взаимодействовать и на которые должны реагировать callbacks, обязаны иметь атрибуты id. Причем для каждого из этих элементов значение атрибута обязано быть уникальным, так как двух элементов с одинаковым значением id быть не может, приложение просто не будет работать. А, во-вторых, у различных элементов есть различные атрибуты, которые отвечают за разнообразное взаимодействие с ними. Например, у элемента-кнопки есть только атрибут, который отвечает за количество кликов по ней, а у элемента-выбора даты есть атрибуты, которые отвечают за кликанье по этому элементу, за изменение начальной и конечной даты и т.д. Поэтому в Input, помимо id элемента, с которым будут взаимодействовать, необходимо указать тот атрибут, на взаимодействие с которым будет реагировать callback.

Поимо Input у объекта callback может быть еще элемент Output. Он не обязателен, его может не быть, а может быть либо один, либо несколько. Данный объект отвечает за элемент, который будет изменяться при взаимодействии с элементом, указанным в Input. Точно так же в нем необходимо указать уникальное значение id и атрибут, который будет изменяться. Как уже было сказано, данного объекта может и не быть, так как, например, при определенном взаимодействии могут меняться не элементы веб-приложения, а глобальные переменные. Или же при взаимодействии может меняться один или несколько элементов веб-приложения. Самое главное при создании callback с объектом Output – это необходимость того, чтобы функция возвращала данные в том формате, который должен быть у указанного атрибута. Например, в объекте Output указано, что при взаимодействии должен изменяться атрибут children у некоторого элемента. Значит, функция обязана возвращать данные в формате списка, содержащего различные элементы. Если это будет не так, то работа приложения нарушится.

Callback, отвечающего за первый отдел веб-приложения, будет реагировать на кликанье по кнопке обновления данных. Поэтому в объекте Input будут указаны id данной кнопки и атрибут n\_clicks, отвечающий за количество кликов по ней. А в объекте Output будут указаны id элемента-вкладки первого отдела и атрибут children, отвечающий за его содержимое. Функция будет совершать следующее: при нажатии на кнопку проверяется наличие данных на текущую дату в подключенной заранее базе. Если данные уже существуют, то функция возвращает исходное содержимое вкладки с добавленной строкой, информирующей о том, что такие данные уже занесены в базу. Если их нет, то запускается алгоритм, описанный в разделе выше, и после его работы обновляется содержимое вкладки, и появляется строка, информирующая о том, что алгоритм отработал и собранные данные занесены в базу. Реализацию данного callback можно увидеть на рисунке Б.6.

В итоге получается следующий вид данного отдела веб-приложения, который изображен на рисунке 4.3.

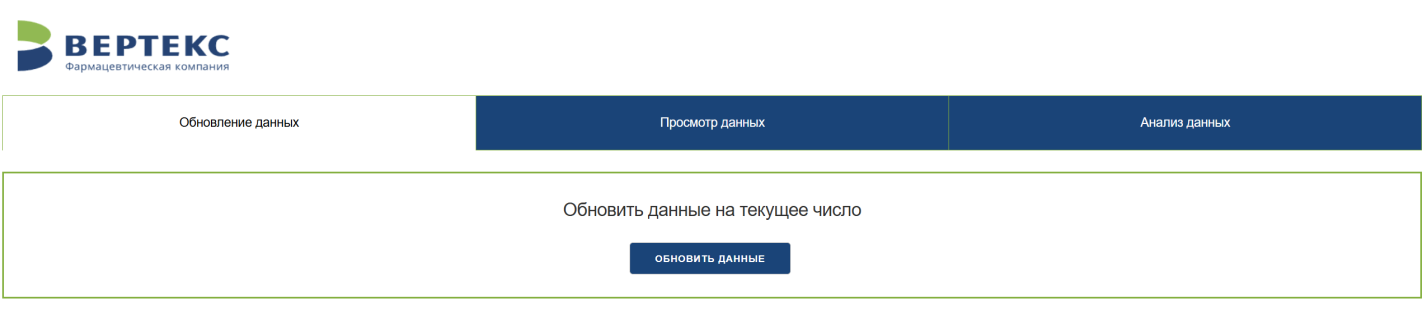


Рисунок 4.3 – Первый отдел веб-приложения

Таким образом, получился простой и понятный интерфейс для работы с разработанным алгоритмом сбора данных. Он заметно упрощает работу, ведь теперь нет нужды запускать алгоритм через интерпретатор и отслеживать его процесс, достаточно просто нажать на кнопку и подождать некоторое время, пока callback сделает все работу.

## **Второй отдел – просмотр данных о фармацевтической продукции и финансовой отчетности компании**

Данный раздел был включен для того, чтобы пользователь не взаимодействовал напрямую с базой данных, особенно если у него нет навыков написания SQL-запросов.

Внутри второй отдел точно таким же образом делится на два подотдела. Первый отвечает за просмотр данных из базы данных, которая была собрана и накоплена в течение всего времени с момента создания алгоритма. Причем пользователь имеет возможность воспользоваться фильтрами: настроить даты, за которые необходимо собрать информацию, выбрать категории, действующие вещества или фармацевтических производителей необходимых лекарственных препаратов. Помимо просмотра данных, их можно скачать в формате Excel для дальнейшей работы с ними. Как уже было сказано в разделе выше, в процесс работы алгоритма собранные данные форматируются и предоставляются в удобном для анализа, аналитики виде.

Второй подотдел отвечает за просмотр данных бухгалтерской отчетности. Чтобы человеку не надо было заходить в различные файлы, искать необходимые отчетности и таблицы, понимать, какая из них за что отвечает, был создан удобный пользовательский графический интерфейс. Во-первых, он отражает названия всех пяти возможных отчетов, а, во-вторых, в каждом таком отчете подписана таблица с соответствующими данными за указанный период. Подобная реализация заметно упрощает работу с финансами компании.

Реализация второго отдела не отличается от реализации первого отдела. Это точно такой же элемент-вкладка, стиль которого совпадает со стилем элемента-вкладки первого отдела. А вот внутри данной вкладки лежат два элемента: надпись, информирующая о том, что пользователь может выбрать, какую информацию хотел бы увидеть, и еще две вкладки, отвечающие за подразделы.

Сначала стоит разобрать подотдел, отвечающий за работу с базой данных. Он точно так же реализован с помощью элемента-вкладки, внутри которого находятся фильтры по датам, категориям, действующим веществам и фармацевтическим производителям. Данные фильтры реализованы с помощью Core-элементов, а именно выпадающих списков с возможностью множественного выбора и ранжировки дат [13]. Информация о возможных вариантах для этих списков взята из глобальных переменных, о которых упоминалось ранее. Помимо фильтров в данном подотделе есть еще две кнопки: первая отвечает за вывод таблицы с выбранной информацией из базы данных, а вторая – за скачивание выбранной информации в формате Excel. Рядом с кнопкой скачивания располагается элемент для этого. Он никак визуально не отображается, а необходим для взаимодействия с загрузками браузера. Так как каждый из этих элементов и даже элемент-вкладка являются интерактивными, то у всех них есть уникальные значения атрибута id. Реализацию данного отдела и подотдела можно увидеть на рисунке Б.7.

Для данного подотдела написаны свои callbacks. Первые пять, которые изображены на рисунке Б.9, отвечают за сохранение результатов фильтров, чтобы в дальнейшем из базы данных согласно ним выбрать информацию. Как можно увидеть, в данных функциях есть только объекты Input, так как в зависимости от выбора фильтров меняться будут не элементы веб-приложения, а локальные переменные. Каждая функция связана с определенным Core-элементом, о чем свидетельствует значение атрибута id, и откликается на изменение его значения. Результат выбора по каждому из фильтров сохраняется в определенную переменную.

Шестой callback отвечает за взаимодействие с кнопкой просмотра данных, поэтому значение ее атрибута id и переменная, отвечающая за клики, указаны в объекте Input. В объекте же Output указано, что при нажатии на данную кнопку будет изменяться содержимое данного подотдела. Реализация процесса изменения показана на рисунках Б.10 и Б.11. При нажатии на кнопку происходит следующее: таблица, содержащая собранную информацию о фармацевтической продукции из базы данных, фильтруется. Сначала по фильтрам, отвечающим за категории, действующие вещества и фармацевтических производителей. Если пользователь не выбрал какой-то фильтр, то процесс отбора данных согласно нему пропускается. А вот если пользователем был сделан выбор, то с помощью специальных методов модуля pandas таблица сортируется, и из нее отбираются только те данные, которые подходят под соответствующие значения. После происходит фильтрация по датам, и в результате получается готовая таблица с данными, которые необходимы для пользователя. Когда процесс отбора данных был выполнен, то выбор пользователя очищается. Данный callback возвращает исходное содержимое первого подотдела, но с добавлением таблицы с выбранными согласно фильтрам данными. Таблица генерируется с помощью вспомогательной функции, которую можно увидеть на рисунке Б.12. В качестве аргумента данной функции передается Dataframe из модуля pandas, а также указывается максимальное количество строк, которое стоит вывести в веб-приложении. Это сделано для того, чтобы не выводить все то, огромное количество данных, которое есть в базе. Пользователь сможет просмотреть данные и примерно понять, как они устроены. Таблица строится с помощью HTML-элементов, и точно также имеет заголовки, столбцы, строки, которые считываются из указанного в аргументах Dataframe.

Седьмой и последний callback позволяет скачать выбранные пользователем данные в формате Excel. В объекте Input указана кнопка скачивания данных, при нажатии на которую будет срабатывать данная функция. А вот в Output указано, что изменяться будет тот самый элемент скачивания, описанный чуть выше. Его атрибуту value будет присваиваться соответствующее значение, что повлечет за собой взаимодействие с браузером и загрузку документа, который был передан, как значение атрибута. Процесс работы данной функции прост: сначала производится та же самая фильтрация данных из базы, как и в прошлом callback, а затем полученная таблица с помощью специальных методов модуля pandas переводится в формат Excel и передается в качестве значения в атрибут элемента скачивания. Реализацию описанного процесса можно увидеть на рисунке Б.13.

Теперь стоит описать второй подраздел, отвечающий за просмотр финансовой отчетности. Для его реализации не потребуются callbacks, так как пользователь будет взаимодействовать только с элементами-вкладками. Каждая из них – это отдельный отчет, который был установлен законодательством как утвержденный набор форм. В бухгалтерскую отчетность обязательно входят: бухгалтерский баланс, отчет о финансовых результатах, отчет о движении денежных средств, отчет о целевом использовании денежных средств и отчет об изменении капитала. Прежде чем приступить к описанию реализации работы второго подотдела, необходимо упомянуть о том, что каждый из отчетов представляет и какую ценность несет.

Бухгалтерский баланс – это основная форма финансовой отчетности компании. Он предоставляет данных об активах и пассивах компании. В активах раскрывается информация об имуществе организации, а в пассивах содержатся данные о способах формирования описанного имущества. Таким образом, данные отчет помогает понять имущественное положение определенной компании на рынке [3].

Отчет о финансовых результатах предоставляет данные о доходах и расходах определенной компании. Благодаря данной форме становится возможно проследить формирование конечного финансового результата деятельности организации. А превышение доходов над расходами показывает тот факт, что за отчетный период была получена чистая прибыль [3].

Третья форма, или отчет о движении денежных средств, позволяет узнать информацию о движении денежных потоков. Причем, процесс изменения денежных средств происходит в результате инвестиционной, финансовой и операционной деятельности организации. С помощью данного отчета можно произвести анализ притоков и оттоков, чтобы оценить платежеспособность компании [3].

Вторая из основных форм финансовой отчетности – это отчет о целевом использовании денежных средств. Данную форму обычно составляют некоммерческие компании или отделы организаций, которые не связаны с предпринимательством, для органа статистики, налоговой службы и т.д. С помощью нее можно узнать информацию о движении денежных средств, которые были направлены на членский взносы, благотворительность или пожертвования [3].

И последний из отчетов, который будет представлен в реализуемом веб-приложении, – это отчет об изменении капитала. С помощью данной формы можно узнать все о резервном, уставном и добавочном капиталах компании. Причем в ней еще отражена информация о нераспределенной прибыли, что в сумме с указанными выше капиталами дает понять, что происходит с собственным капиталом компании [3].

Финансовая отчетность считается публичным документом и должна находиться в открытом доступе, чтобы потенциальные покупатели, инвесторы и другие лица могли ее посмотреть в любой момент времени. Данная документация публикуется на специальном онлайн-ресурсе Федеральной Налоговой Службы, который называется БФО [9]. Все бухгалтерские данные, которые присутствуют в данном веб-приложении, взяты оттуда с согласия представителей компании АО «Вертекс».

А реализация просмотра данных о бухгалтерской отчетности компании очень проста. Для каждого отчета создан свой собственный элемент-вкладка, стиль которого совпадает со стилями всех остальных вкладок, описанных выше. И в содержимом каждой из вкладок указаны соответствующие определенному отчету таблицы и их заголовки. Таблицы созданы с помощью вспомогательной функции, указанной на рисунке Б.14. Она отличается от предыдущей тем, что, во-первых, выводятся все строки таблицы, а, во-вторых, для таблицы можно установить при желании определенный стиль. На рисунке Б.8 можно увидеть пример реализации одного из отчетов во втором подразделе. Все остальные реализованы подобным образом, просто с соответствующими им таблицами.

В итоге получается следующий вид второго отдела веб-приложения, который можно увидеть на рисунках 4.4, 4.5 и 4.6.

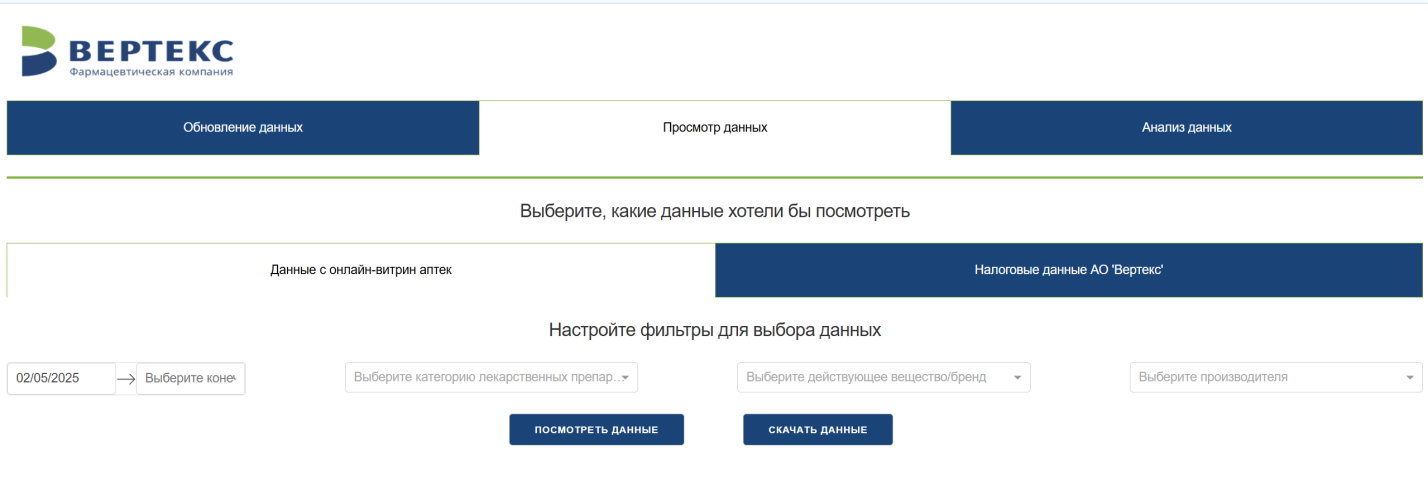


Рисунок 4.4 – Первый подраздел второго раздела веб-приложения

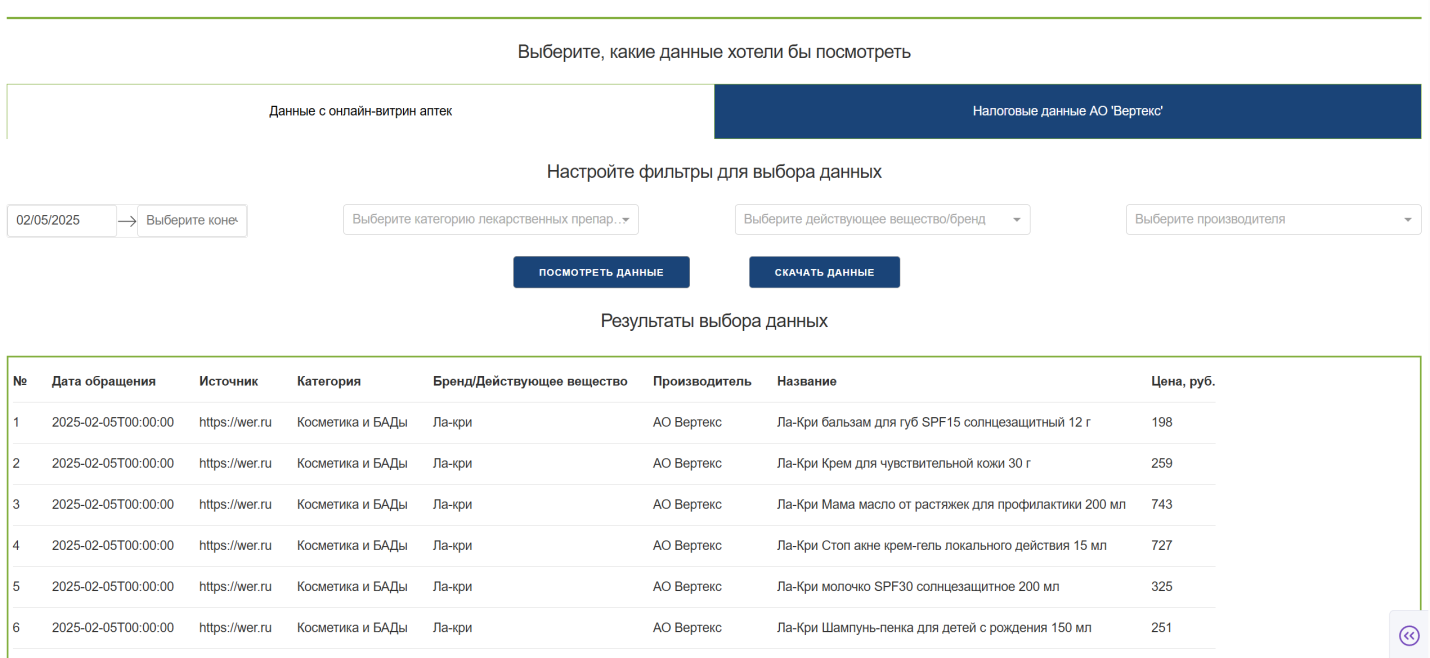


Рисунок 4.5 – Просмотр выбранных данных во втором подразделе второго раздела веб-приложения

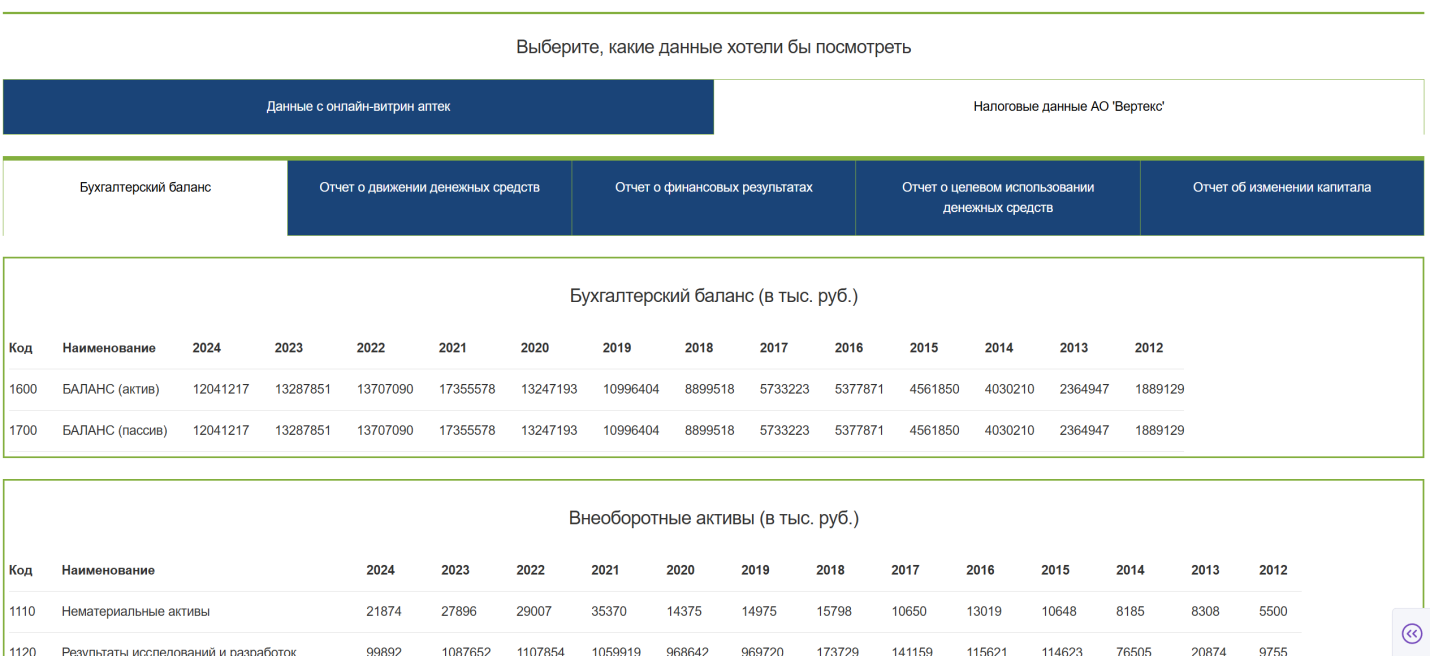


Рисунок 4.6 – Второй подраздел второго раздела веб-приложения

Таким образом, данный отдел веб-приложения заметно облегчает работу пользователю как с базой данных, так и с финансовой отчетность. Он позволяет человеку, не знающему SQL-запросы, легко и просто фильтровать данные о фармацевтической продукции в онлайн-сегменте аптечного рынка, взятые из базы данных, а также скачивать полученную информацию для дальнейшей работы. Кроме того, в нем реализован понятный и удобный интерфейс для просмотра бухгалтерской отчетности компании АО «Вертекс».

## **Третий отдел – общий анализ рынка, финансовой отчетности и отдельных лекарственных препаратов**

Данный отдел непосредственно обеспечивает реализацию главной цели данного веб-приложения, а именно: обеспечение поддержки принятия решений. Он предоставляет анализ фармацевтического рынка в целом, анализ ценообразования конкретных лекарственных препаратов и внешний финансовый анализ по значимым для компании параметрам.

Как можно понять из сказанного выше, данный отдел делиться на три раздела соответственно. Каждый раздел реализован точно так же, как и все до него, что можно увидеть на рисунках Б.15, Б.26 и Б.29. Причем стили у всех элементов-вкладок совпадают, о чем говорилось уже несколько раз. Сделано это для того, чтобы веб-приложения выглядело одновременно стильно и минималистично.

Первый раздел отвечает за общий анализ фармацевтического рынка. С представителями компании было обговорено, какие именно показатели интересуют отдел аналитиков. В результате обсуждения были выбраны следующие параметры: диаграмма долей от общего объема рынка ТОП-20 фармацевтических компаний, диаграмма долей от общего объема рынка ТОП-20 аптечных точек, динамика розничных продаж и государственных закупок по ФО РФ и графики, отражающие изменения средней цены в онлайн-сегменте рынка на безрецептурные и рецептурные препараты по 4 сегментам.

Все данные по ТОП-20 фармацевтическим компаниям и аптечным точкам, их долях, а также динамика государственных закупок и розничных продаж были взяты со специальных источников, которые профессионально занимаются сбором данных, разбором и анализом фармацевтического рынка России [9]. Бухгалтерская отчетность описана в подразделе выше.

Как уже было сказано выше, фреймворк Dash работает в том числе на основе библиотеки Plotly, с помощью которой можно строить самые разнообразные графики для изучения и анализа данных [13]. Так вот, круговые диаграммы, которые отражают самые крупные доли от общего объема рынка ТОП-20 производителей и аптечных сетей на текущий год будут созданы с помощью данной библиотеки. Их реализация изображена на рисунке Б.17, где можно увидеть, что создаются специальные объекты, которым в качестве аргументов передаются соответствующие глобальные переменные, содержащие нужную информацию. А затем данные объекты круговых диаграмм передаются в специальные Core-элементы, отвечающие за отображение графиков в веб-приложении. Весь данный процесс можно увидеть на рисунке Б.15.

Следующий этап – это отображение данных о динамике розничных продаж и государственных закупок по Федеральным Округам России. Все необходимые данные содержаться в соответствующих глобальных переменных, поэтому с помощью вспомогательной функции на рисунке Б.14 отображаем их в таблице. Как показано на рисунке Б.15, данные таблицы с соответствующими подписями помещены в отдельный div-элемент со стилем, подходящим под веб-приложение.

Последнее и самое трудоемкое, что необходимо сделать для данного раздела, – это создание графиков изменения средних цен на безрецептурные и рецептурные препараты по 4 сегментам. Именно для них были собраны данные в базу с помощью разработанного алгоритма. Данные графики позволят оценить конкурентоспособность АО «Вертекс» в онлайн-сегменте фармацевтического рынка России. Как было выяснено из соответствующих источников по аналитике рынка, традиционно лекарственные препараты разделяются на 4 сегмента: стоимость выше 1000 руб., в пределе 500-1000 руб., в пределе 100-500 руб. и меньше 100 руб. Поэтому необходимо оценить изменение цен именно в данных диапазонах.

Для начала из глобальной переменной, связанной с базой данных и представляющей собой таблицу с собранными данными, необходимо выделить записи о 4 сегментах безрецептурных, а затем и рецептурных лекарств. После чего с помощью специальных функций группировки модуля pandas производится поиск средних цен для каждой даты в выбранных сегментах и категориях, что изображено на рисунке Б.18. Следующий шаг – это нахождение в каждой категории и сегменте 6 самых преуспевающих компаний, чей продукции больше всего. Для этого с помощью все тех же встроенных функций модуля Pandas считаем, сколько раз встречается среди продукции каждая компания, находим 6 самых частых и узнаем их наименования. Весь данный процесс можно увидеть на рисунке Б.19. А затем точно так же с помощью группировки находим для каждой даты и компании в эту дату среднюю цену на препараты соответствующей категории и сегмента (см. рисунок Б.20). Предпоследние шаги для создания описанных выше графиков – это извлечение данных о средних ценах, создание таблиц, где они будут храниться для дальнейшего использования, и их заполнение извлеченными данными о средних ценах для каждой из 6 компаний в определенной категории и сегменте. Весь этот процесс реализован с помощью все того же модуля по работе с таблицами. Его этапы можно увидеть на рисунках Б.21, Б.22 и Б.23.

После описанной выше подготовки данных можно создавать графики. Для этого будет использована уже упомянутая ранее библиотека Plotly, в частности объект Figure. В данной библиотеке есть объекты, которые заранее отвечают за определенные графики: как, например, с круговой диаграммой, описанной выше. А есть такой объект, как Figure, на который можно добавлять «следы», представляющие собой различные графики, такие как линии, гистограммы, распределения и так далее. Так вот отображение изменения средних цен будет следующим: основной будет гистограмма, отражающая среднюю цену на всю категорию в данном сегменте по всему рынку, а поверх нее графиками будут идти средние цены на данную категорию в указанном сегменте для определенной компании. Подобные объекты, реализацию которых можно увидеть на рисунках Б.24 и Б.25, будут сделаны для каждой категории и для каждого сегмента в ней. И последний этап реализации первого раздела третьего отдела веб-приложения – это размещение созданных графиков в теле веб-приложения, что показано на рисунке Б.16.

Второй раздел третьего отдела веб-приложения отвечает за анализ ценообразования определенного лекарственного препарата в онлайн-сегменте рынка. Он реализован точно также с помощью элемента-вкладки с определенным стилем. Кроме того, как и в первом разделе второго отдела в нем присутствуют фильтры для выбора данных. Для того, чтобы получить анализ, необходимо настроить даты, за которые желательно узнать информацию, и выбрать нужные действующие вещества из выпадающего списка. Реализация всего описанного выше изображена на рисунке Б.26. Вся суть работы данного раздела будет заключаться в написанном для него Callback. В объекте Input указано, что данная функция будет реагировать на нажатие соответствующей кнопки, а в объекте Output – что изменяться будет содержимое данного раздела, т.е. его элемента-вкладки. А при взаимодействии с кнопкой происходит следующее: все выборы в фильтрах запоминаются и очищаются, а затем происходит выбор данных из таблицы, связанной с базой данных, согласно фильтрам. После чего для каждого выбранного действующего вещества будут построены свои графики. Это происходит следующим образом: на каждой итерации цикла по списку выбранных действующих веществ в соответствующей глобальной переменной выбираются по ключу все дозировки для текущего вещества. И далее начинается цикл по дозировкам: среди отфильтрованных ранее данных выбираются те, которые подходят под текущую дозировку, группируются по датам и рассчитываются средние цены в каждую из дат. А затем по собранным данным строится график с помощью уже известного объекта Figure. В результате работы получаются наглядные графики изменения средних цен на все возможные дозировки по всем выбранным лекарственным препаратам. Реализацию функции callback можно увидеть на рисунках Б.27 и Б.28.

И остался самый последний раздел третьего отдела веб-приложения, который отвечает за внешний финансовый анализ по тем параметрам, которые наиболее важны для компании АО «Вертекс». Для финансового анализа будет использована имеющаяся бухгалтерская отчетность, которую можно посмотреть во втором отделе веб-приложения. Третий раздел оформлен точно так же, как и все остальные элементы-вкладки, что можно увидеть на рисунке Б.29.

Внешний финансовый анализ обычно проводится заинтересованными сторонами, такими как инвесторы, кредиторы и так далее. Он основан на данных из открытых источников, в частности на бухгалтерской отчетности, и призван обеспечить сторонних наблюдателей информацией о текущем финансовом состоянии компании, ее положении на рынке. Данный анализ позволяет заинтересованным сторонам оценить, насколько выгодно взаимодействовать с необходимой им компанией.

Однако в некоторых ситуациях внешний анализ может быть полезен и самой компании, чтобы «посмотреть на себя со стороны», особенно когда она пытается привлечь инвесторов. Конечно, вне всяких сомнений то, что внутренний анализ дает более полную и четкую оценку деятельности организации, но для его проведения необходимо обладать соответствующей квалификацией и доступом к внутренним отчетам компании. Так как при выполнении данной выпускной квалификационной работы необходимого доступа не было, то совместно с представителем компании было принято решение о проведении внешнего финансового анализа, который позже работники компании дополнят еще и внутренним.

Первые параметры, которые будут представлены в данном разделе, – это выручка и ее горизонтальный анализ – расчет темпов роста. Выручкой называется полная сумма или основной доход, который организация получает либо от продажи своих товаров, либо от реализации своих услуг. С ее помощью можно оценить, насколько продукция компании популярна и актуальна на рынке страны или мировом рынке. А рассчитанные темпы роста выручки за каждый год помогут выявить тренды и определить по ним, является ли текущая стратегия поведения организации успешной [1].

В формуле (3.1) показано, как рассчитать темпы роста выручки для каждого года.

|  |  |
| --- | --- |
| , | (3.1) |

где – это темп роста выручки текущего года;

– выручка за текущий отчетный год;

– выручка за прошлый отчетный год.

Вторые параметры, которые будут рассмотрены в данном анализе в веб-приложении, – это активы компании АО «Вертекс». В данной секции будут рассмотрены три вида активов: общие активы, внеоборотные активы и чистые активы. Два первых параметра деятельности компании, как уже было сказано выше, можно найти в бухгалтерском балансе. Общие активы представляют собой следующее: каждая организация владеет собственностью, а именно деньгами (капиталом), землей и недвижимостью, и вот все это образует собой общие активы компании. Они показывают общее финансовое состояние компании на текущий отчетный год. А внеоборотные активы входят в общие, но показывают только те средства, которые за счет деятельности компании приносят доход за один или более год. Совершенно другая ситуация с чистыми активами. Это единственный параметр в отчетности, который регламентирован государством для подсчета. Данный показатель представляет собой разность между общими активами компании и суммой ее обязательств, т.е. с его помощью эксперты оценивают способность компании покрывать свои расходы. Таким образом, необходимо отразить данные экономические показатели деятельности организации на одном графике, так как при этом сразу предоставляется возможность оценить динамику финансовой деятельности компании. Сторонний наблюдатель сможет оценить положение компании на рынке, актуальность и популярность ее продукции или услуг, а также способность выплачивать все свои обязательства [1].

И третьи, тоже очень важные параметры, объем которых необходимо отразить на графиках во внешнем финансовом анализе, – это выручка, чистая прибыль и EBIT. Про выручку уже было сказано чуть выше. Чистая прибыль представляет собой прибыль после вычета всех расходов. Причем расход включают в себя как расходы на операционную деятельность, так и налог на прибыль, амортизацию и так далее. То есть, другими словами, она отражает информацию о том, сколько действительно заработала организация за отчетный период. EBIT точно так же является прибылью после вычета расходов, но вот только после вычета операционных расходов и никаких других. Операционные расходы показывают количество денежных средств, которое компания тратит либо на производство продукции, либо на оказание услуг, т.е. на свою основную деятельность. Получается, что данный показатель хранит информацию о заработках организации, которые не зависят от влияния налоговых служб. Сравнение этих трех показателей и отображение на одном графике очень важны для анализа экономической деятельности компании, так как позволять оценить верность ее стратегий, стабильность и возможность взаимодействия с налоговыми органами [1].

Выручка и чистая прибыль представлены в отчете о финансовых результатах и в отчете об изменении капитала соответственно, а вот EBIT необходимо рассчитать по формуле (3.2).

|  |  |
| --- | --- |
| , | (3.2) |

где – чистая прибыль за текущий отчетный год;

– налог на прибыль за текущий отчетный год;

– проценты по кредитам за текущий отчетный год.

Таким образом, были описаны все экономические показатели, которые необходимо либо найти в соответствующей отчетности, либо рассчитать для того, чтобы отобразить их динамику на графиках. Теперь необходимо перейти к последнему этапу работы над веб-приложением – к расчету различных коэффициентов, чтобы оценить ту или иную способность компании АО «Вертекс».

Самое первое, что нужно оценить во внешнем финансовом анализе, – это ликвидность компании. Оценка ликвидности позволяет заинтересованным сторонам, кредиторам и инвесторам, понять, как быстро и насколько эффективно интересующая их организация способна выплатить все долги по краткосрочным обстоятельствам. Вдруг возникнет ситуация, что инвестор хочет вложить деньги в предприятие, а анализ показывает, что будет очень сложно минимизировать риски потерь. Для данной оценки есть три основных показателя, которые возможно рассчитать по имеющейся финансовой отчетности, а именно: коэффициенты текущей ликвидности, быстрой ликвидности и абсолютной ликвидности [1].

Первый параметр связан с оборотными активами, которые отличаются от упомянутых ранее внеоборотных тем, что отражают информацию о тех показателях, которые приносят доход организации несколько раз за год. Получается, что с помощью данного коэффициента возможно оценить способность компании покрыть краткосрочные обязательства с помощью описанных активов.

По формуле (3.3) можно рассчитать коэффициент текущей ликвидности.

|  |  |
| --- | --- |
|  | (3.3) |

где – коэффициент текущей ликвидности;

– оборотные активы за текущий отчетный год;

– краткосрочные обязательства за текущий отчетный год.

Второй коэффициент при своем расчете из оборотных активов исключает запасы, поэтому является более строгим в оценке, чем первый. Запасы могут спасти организацию в некоторых кризисных ситуациях, поэтому в более строгой оценке их не надо учитывать при расчете возможности компании покрыть краткосрочные обязательства.

По формуле (3.4) считается коэффициент быстрой ликвидности.

|  |  |
| --- | --- |
|  | (3.4) |

где – коэффициент быстрой ликвидности;

– оборотные активы за текущий отчетный год;

– запасы за текущий отчетный год;

– краткосрочные обязательства за текущий отчетный год.

Последний коэффициент позволяет оценить быстроту компании, с которой она сможет расплатиться по всем краткосрочным обязательством, используя из активов только деньги.

По формуле (3.5) можно рассчитать коэффициент абсолютной ликвидности.

|  |  |
| --- | --- |
|  | (3.5) |

где – коэффициент абсолютной ликвидности;

– денежные средства за текущий отчетный год;

– краткосрочные обязательства за текущий отчетный год.

Вторая не менее важная оценка – это оценка рентабельности компании, которая позволит понять эффективность использования вложенных ресурсов и капитала организации для получения прибыли. На данную способность в первую очередь смотрят при изучении внешнего финансового анализа компании, так как она дает информацию о конкурентоспособности компании на рынке [1]. Данный показатель точно так же, как и предыдущий основывается на трех оценках.

Самая первая из них называется ROA, или рентабельность активов. Она оценивает способность организации генерировать прибыль с помощью собственных активов.

Рассчитать данный показатель можно по формуле (3.6).

|  |  |
| --- | --- |
|  | (3.6) |

где – чистая прибыль за текущий отчетный год;

– средние активы за текущий отчетный год;

– максимальные активы за текущий отчетный год;

– минимальные активы за текущий отчетный год.

Вторая оценка – это ROE. Она называется рентабельность собственного капитала и отражает способность компании получать прибыль за счет денежных средств акционеров данной организации. Последнее в свою очередь является собственным капиталом компании. Он рассчитывается, как разность между всем капиталом и всеми обязательствами компании, т.е. суммой краткосрочных и долгосрочных обязательств.

С помощью формулы (3.7) можно узнать описанный показатель компании.

|  |  |
| --- | --- |
|  | (3.7) |

где – чистая прибыль за текущий отчетный год;

– собственный капитал за текущий отчетный год;

– капитал за текущий отчетный год;

– краткосрочные обязательства за текущий отчетный год;

– долгосрочные обязательства за текущий отчетный год.

И последняя в данной секции оценка – это проверка рентабельности продаж (ROS). На самом деле смысл у нее очень простой: она показывает, какой процент от всей выручки составляет чистая прибыль, т.е. предоставляет информацию о том, как компания управляет своими расходами.

С помощью формулы (3.8) можно рассчитать описанный показатель компании.

|  |  |
| --- | --- |
|  | (3.8) |

где – чистая прибыль за текущий отчетный год;

– выручка за текущий отчетный год.

Третья оценка, которая будет играть важную роль в реализуемом веб-приложении, является оценкой финансовой устойчивости организации. Ее важность очень велика, поскольку с ее помощью сторонние наблюдатели смогут понять, способна ли компания пережить кризисную ситуацию и спасти свою платежеспособность [1]. Будет рассмотрено три основных коэффициента для определения данной эффективности.

Коэффициент автономии по сути своей является показателем доли собственного капитала от общих активов компании и позволяет оценить нужду организации в заемных средствах.

Посчитать его можно по формуле (3.9).

|  |  |
| --- | --- |
|  | (3.9) |

где – коэффициент автономии;

– собственный капитал за текущий отчетный год;

– общие активы за текущий отчетный год;

– капитал за текущий отчетный год;

– краткосрочные обязательства за текущий отчетный год;

– долгосрочные обязательства за текущий отчетный год.

Следующий коэффициент позволяет оценить уровень долговой нагрузки организации за текущий отчетный период. Называется он коэффициент финансовой зависимости.

С помощью формулы (3.10) можно посчитать описанный показатель.

|  |  |
| --- | --- |
|  | (3.10) |

где – коэффициент финансовой зависимости;

– заемные средства за текущий отчетный год;

– общие активы за текущий отчетный год;

– краткосрочные обязательства за текущий отчетный год;

– долгосрочные обязательства за текущий отчетный год.

Коэффициент покрытия процентов полезен тем, что позволяет узнать, насколько способна компания покрывать проценты по кредитам с заемных средств.

Необходимо воспользоваться формулой (3.11) для его расчета.

|  |  |
| --- | --- |
|  | (3.11) |

где – коэффициент покрытия процентов;

– операционная прибыль за текущий отчетный год;

– проценты по кредитам за текущий отчетный год.

Четвертая, используемая во внешнем финансовом анализе оценка – это оценка платежеспособности организации, или, другими словами, насколько эффективно, своевременно и в каком объеме компания закрывает свои экономические обязательства. С помощью нее инвесторы и кредиторы могут оценить состояние компании, если вдруг произойдет кризис [1]. В данной секции всего два основных показателя-коэффициента.

Коэффициент покрытия обязательств важен тем, что показывает способность организации погасить все обязательство с помощью ее активов. А вот второй коэффициент предоставляет данные о том, как организация покрывает свои проценты по долговым обязательствам лишь за счет операционной прибыли, т.е. прибыли, полученной за основную деятельность предприятия.

С помощью формулы (3.12) можно посчитать первый описанный показатель.

|  |  |
| --- | --- |
|  | (3.12) |

где – коэффициент покрытия обязательств;

– заемные средства за текущий отчетный год;

– общие активы за текущий отчетный год;

– краткосрочные обязательства за текущий отчетный год;

– долгосрочные обязательства за текущий отчетный год.

Второй коэффициент считается для внешнего финансового анализа по формуле (3.13).

|  |  |
| --- | --- |
|  | (3.13) |

где – коэффициент обслуживания долга;

– операционная прибыль за текущий отчетный год;

– проценты по кредитам за текущий отчетный год.

Предпоследний анализ экономической деятельности компании является оценкой деловой активности организации. Она отражает способность компании поддерживать стабильность прибыли и объема продаж только лишь за счет ее активов и ничего больше [1]. Снова в ней играют важную роль три коэффициента, про которые будет рассказано ниже.

Один из них, оборачиваемость активов, помогает оценить полезность использования активов компании для получения выручки. Второй, который называется оборачиваемость запасов, отвечает за показатель быстроты компании в распродаже всех своих запасов. А последний коэффициент – оборачиваемость дебиторской задолженности – показывает инвесторам и кредиторам способность компании в кратчайшие сроки собрать денежные средства, полученные у дебиторов.

Рассчитать первый из описанных показателей можно по формуле (3.14).

|  |  |
| --- | --- |
|  | (3.14) |

где – оборачиваемость активов за текущий отчетный год;

– выручка за текущий отчетный год;

– средние активы за текущий отчетный год;

– максимальные активы за текущий отчетный год;

– минимальные активы за текущий отчетный год.

Для второго используется формула (3.15).

|  |  |
| --- | --- |
|  | (3.15) |

где – оборачиваемость запасов за текущий отчетный год;

– себестоимость продаж за текущий отчетный год;

– средние запасы за текущий отчетный год;

– максимальные запасы за текущий отчетный год;

– минимальные запасы за текущий отчетный год.

И, наконец, по формуле (3.16) можно рассчитать последний из трех показателей.

|  |  |
| --- | --- |
|  | (3.16) |

где – оборачиваемость дебиторской задолженности за текущий отчетный год;

– выручка продаж за текущий отчетный год;

– средняя дебиторская задолженность за текущий отчетный год;

– максимальная дебиторская задолженность за текущий отчетный год;

– минимальная дебиторская задолженность за текущий отчетный год.

Последняя из оценок, которая будет указана во внешнем финансовом анализе в данном веб-приложении, - это оценка рисков деятельности компании. Показатели, включенные в данную оценку, позволят увидеть, сможет ли изучаемая организация в условиях экономической нестабильности держаться на плаву [1]. Для этого необходимо рассчитать всего два показателя: волатильность прибыли за некоторые периоды и коэффициент задолженности. Первый из них позволяет оценить рыночные риски, в то время как второй – кредитные риски.

Для того, чтобы произвести расчет первого показателя, необходимо воспользоваться формулой (3.17).

|  |  |
| --- | --- |
|  | (3.17) |

где – волатильность прибыли за определенный отчетный период;

– максимальная прибыль за определенный отчетный период;

– минимальная прибыль за определенный отчетный период;

– средняя прибыль за определенный отчетный период.

Для того, чтобы произвести расчет второго показателя, необходимо воспользоваться формулой (3.18).

|  |  |
| --- | --- |
|  | (3.18) |

где – коэффициент задолженности за текущий отчетный год;

– заемные средства за текущий отчетный год;

– собственный капитал за текущий отчетный год;

– капитал за текущий отчетный год;

– краткосрочные обязательства за текущий отчетный год;

– долгосрочные обязательства за текущий отчетный год.

Таким образом, были описаны все экономические показатели и формулы для их расчета, которые необходимы и важны для проведения внешнего финансового анализа. Реализацию расчетов и создания графиков для них можно увидеть на рисунках Б.30, Б.31, Б.32, Б.33, Б.34 и Б.35. А на рисунке Б.29 показано, как они были перенесены в элементы тела веб-приложения.

Теперь стоит пройтись по результатам финансового анализа деятельности АО «Вертекс». И начнем с оценки выручки и ее темпов роста, что можно увидеть на рисунке 4.7. По сравнению с 2010-ыми годами темпы роста выручки значительно возросли за последнее время, что является хорошим показателем для компании, так как означает, что ее продукция популярна на рынке. Конечно, были спады в темпах, например ковидный 2020 год, что хорошо видно и на графике, и в таблице с темпами роста, но в основном компания с каждым годом растет и развивается в нужном направлении.

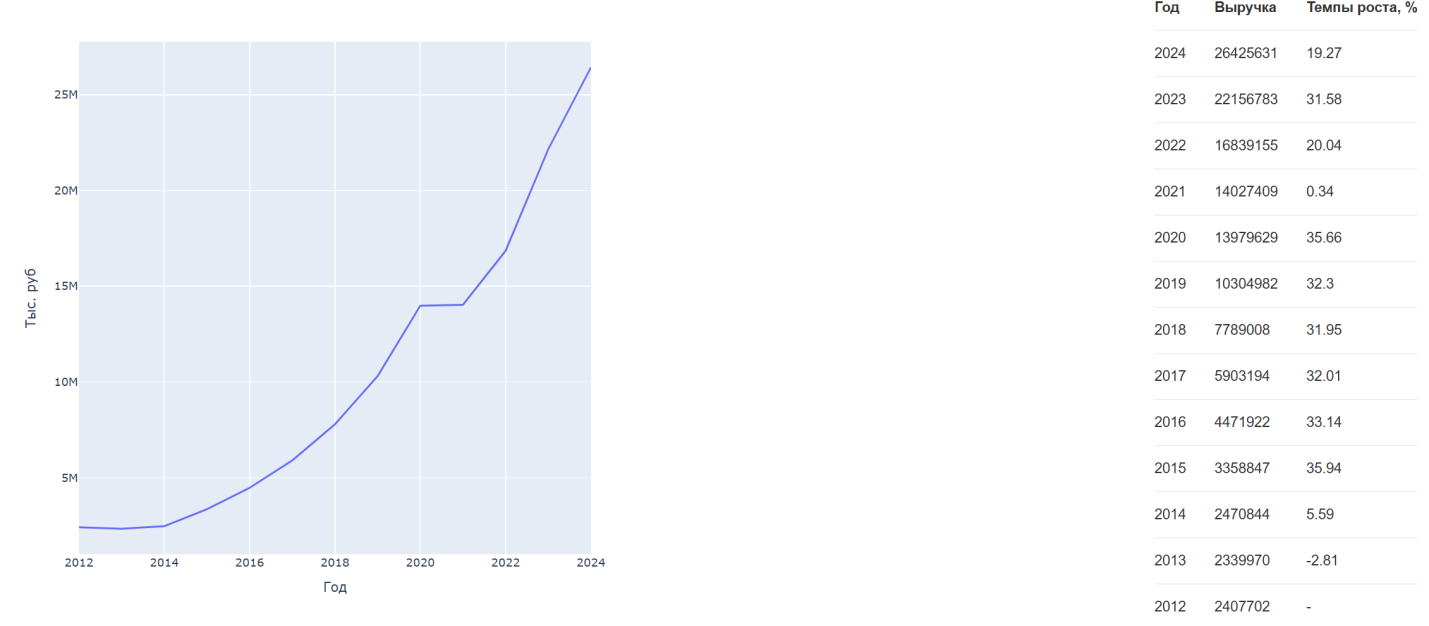


Рисунок 4.7 – Динамика выручки АО 'Вертекс'

По динамике активов сразу можно заметить два спада: 2020 и 2022 годы. В первый год распространялась пандемия, а во второй – на Россию наложили санкции, поэтому некоторые зарубежные точки сбыта продукции пропали у компании. Именно из-за этих ситуаций общее состояние компании уменьшилось в эти годы, но в целом его рост очень заметен, особенно с первыми годами отчетности. А динамика внеоборотных и чистых активов показывает, что компания способна покрывать все свои расходы и выплачивать проценты вовремя за счет заработанных на операционной деятельности средств. Соответственно, это означает, что АО «Вертекс» является надежным для инвесторов и кредиторов, потому что их средства будут в безопасности. Динамика активов компании изображена на рисунке 4.8.



Рисунок 4.8 – Динамика активов АО 'Вертекс'

На основе главных показателей рентабельности можно с уверенностью сказать, что за последние годы компания научилась уверенно использовать и применять свои активы для того, чтобы заработать прибыль. С каждым годом ее чистая прибыль растет, что показывает тот факт, что компания выбрала правильную стратегию поведения на рынке и ее продукция становится популярной. А увеличивающийся EBIT говорит о том, что компания способна покрыть свои расходы от производства лекарственных препаратов и при этом остаться в плюсе, что можно увидеть на рисунке 4.9.

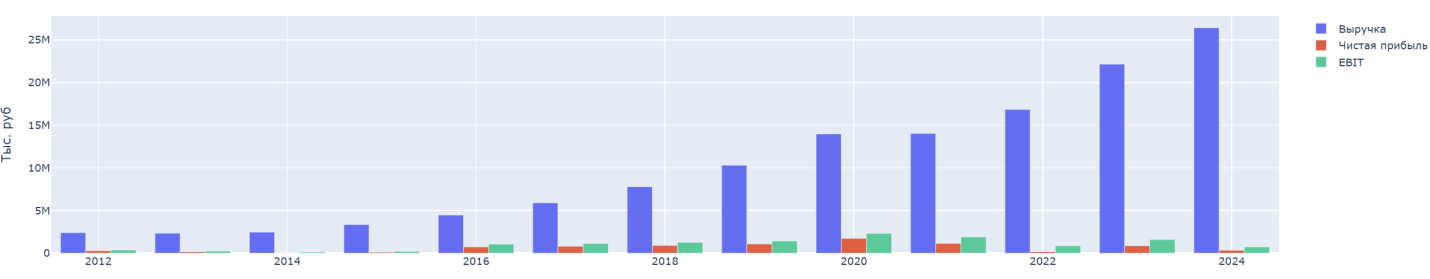


Рисунок 4.9 – Основные показатели рентабельности АО 'Вертекс'

Теперь перейдем к рассчитанным оценкам деятельности и коэффициентам в них. Первая – оценка ликвидности, что изображена на рисунке 4.10. Для нее были рассчитаны соответствующие коэффициенты, описанные выше. По ним можно судить о том, что в 2010-ые у компании были некоторые сложности с выплатами краткосрочных обязательств. Да, компания не уходила в минус, но тогда краткосрочные обстоятельства представляли для них некоторые сложности. А вот начиная с 2019 года, когда у компании резко возросли доходы и популярность на рынке, она стала способная быстро и эффективно закрывать свои краткосрочные обязательства, что, несомненно, является хорошим показателем для сторонних людей, оценивающих данную компанию.



Рисунок 4.10 – Оценка ликвидности АО 'Вертекс'

Следующая оценка – это оценка рентабельности компании. Были рассчитаны соответствующие показатели, по которым видно, что компания разумно использует активы и продажи для получения прибыли. Причем с каждым годом ситуация становится все лучше и лучше. Но, к сожалению, исходя из расчетов компания неэффективно использует собственный капитал, так как расходы на него превышают доходы. В целом ситуация ровно такая же, как была описана выше при оценке основных показателей, но данные коэффициенты позволили более подробно узнать про ситуацию и понять, где могут быть проблемы. Оценку рентабельности можно увидеть на рисунке 4.11.

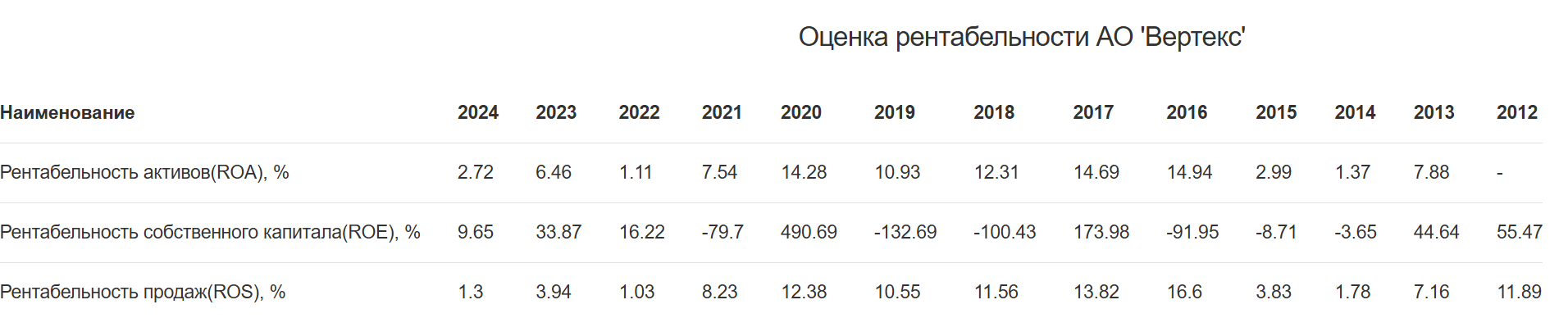


Рисунок 4.11 – Оценка рентабельности АО 'Вертекс'

Оценка финансовой устойчивости, что изображена на рисунке 4.12, показывает, что с долговой нагрузкой у компании все отлично. В 2010-х годах она была зависима от заемных средств, но данная зависимость с каждым годом уменьшалась, а с 2019 вообще исчезла. Данная оценка позволяет узнать, что компания всегда покрывает свои проценты по кредитам вовремя и в полном объеме, что доказывает полученную ранее оценку рентабельности. Для кредиторов данная компания считается надежной.

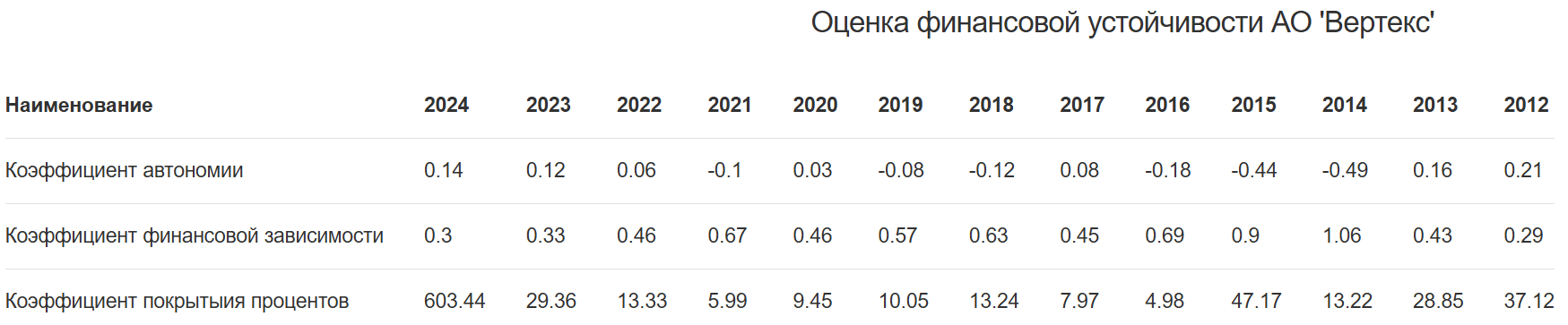


Рисунок 4.12 – Оценка финансовой устойчивости АО 'Вертекс'

Оценка платежеспособности компании изображена на рисунке 4.13. Она позволяет инвесторам и кредиторам убедиться в том, что компания способна покрыть свои расходы за счет операционной прибыли и активов. То есть она показывает, что компания способна покрыть расходы на производство препаратов. Кроме того, видно, что с каждым годом эта способность только возрастает, что означает устойчивость компании в кризисных ситуациях.

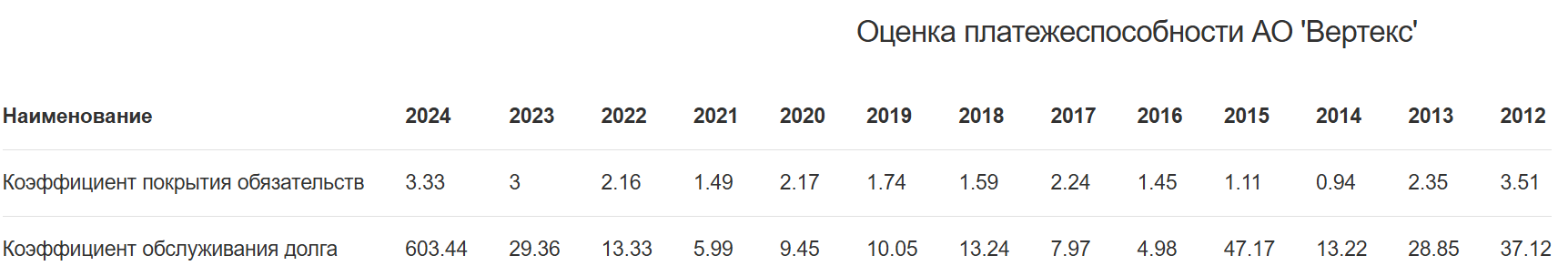


Рисунок 4.13 – Оценка платежеспособности АО 'Вертекс'

Пятый анализ называется оценка деловой активности компании. С помощью нее можно узнать может ли компания поддерживать стабильность объемов продаж и получения прибыли только за счет своих активов. Так вот АО «Вертекс» показывает абсолютную стабильность за весь период отчетности, что означает, что компания эффективно использует свои активы для операционной деятельности и в случае кризисной ситуации сможет выстоять на рынке. Это является хорошим показателем для инвесторов и кредиторов. Данный анализ можно увидеть на рисунке 4.14.

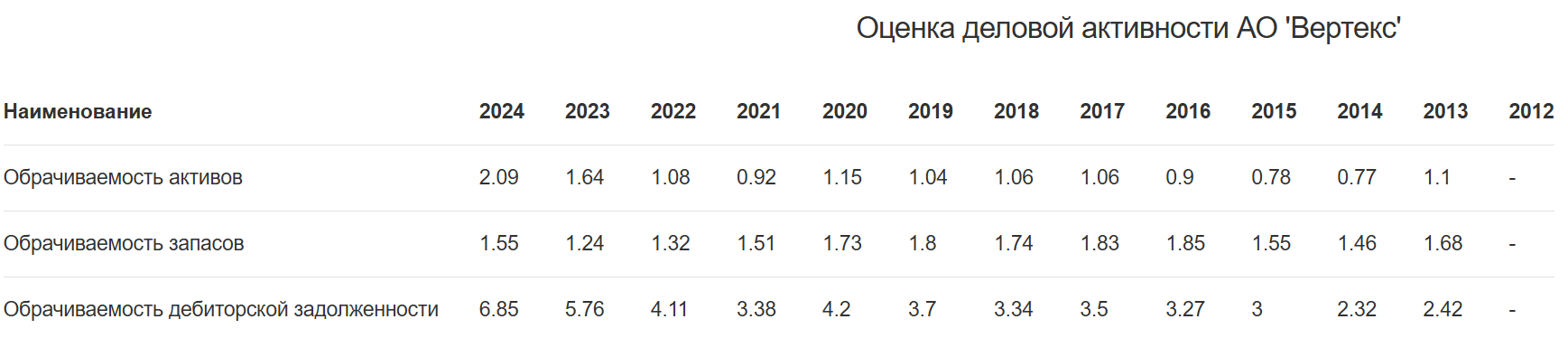


Рисунок 4.14 – Оценка деловой активности АО 'Вертекс'

И последняя оценка будет очень полезна для инвесторов и кредиторов, так как позволит оценить все риски, связанные с данной компанией. Она изображена на рисунке 4.15 и показывает, что раньше у нее был шанс на рыночные риски, но за текущий период он заметно снизился, и состояние компании стабилизировалось. А вот с кредитными рисками все заметно хуже: оценка показывает, что они имеются и даже в текущем периоде.

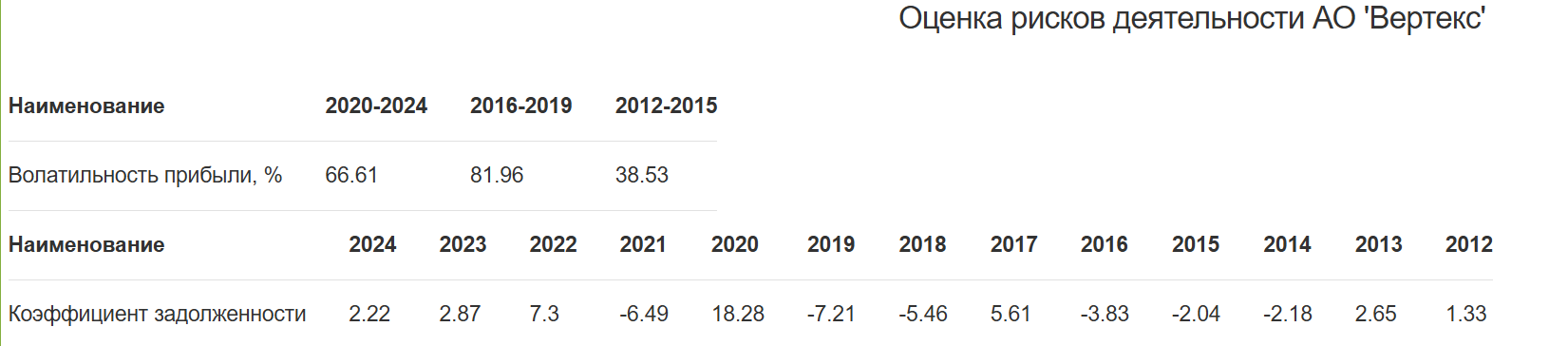


Рисунок 4.15 – Оценка рисков деятельности АО 'Вертекс'

Получился третий отдел приложения следующего вида, что можно увидеть на рисунках 4.16, 4.17 и 4.18.

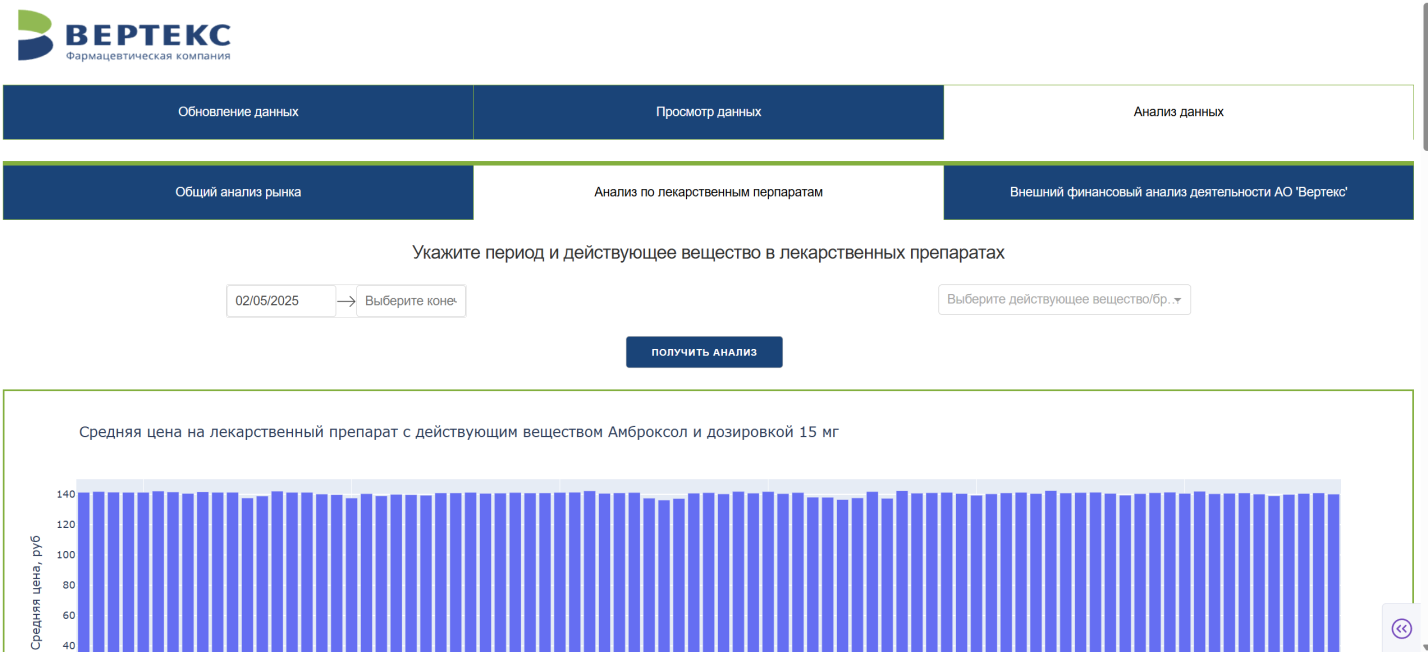


Рисунок 4.16 – Первый раздел третьего отдела веб-приложения



Рисунок 4.17 – Второй раздел третьего отдела веб-приложения

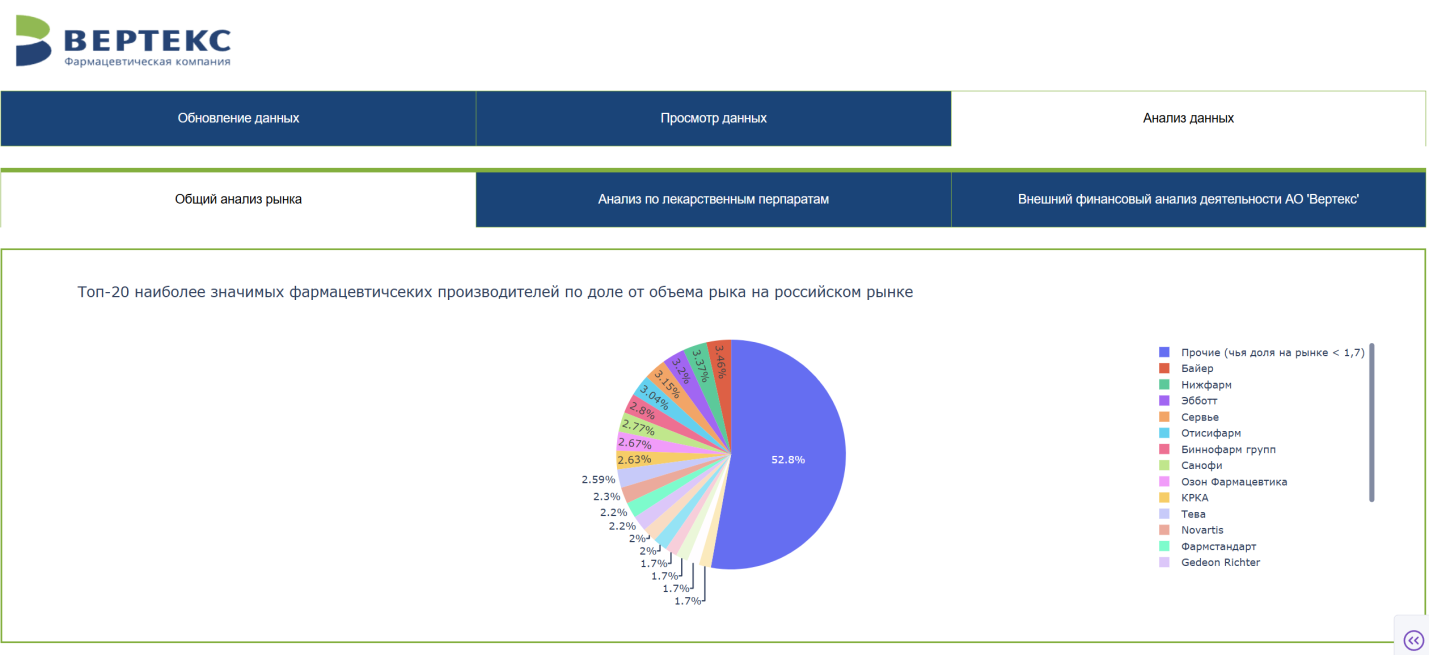


Рисунок 4.18 – Третий раздел третьего отдела веб-приложения

Получился третий отдел с удобным, а самое главное простым и понятным пользовательским графическим интерфейсом. Теперь для заинтересованных сторон работы весьма упростилась, ведь все графики уже заранее построены, а показатели рассчитаны.

## **Выводы по разделу 4**

Подведя итоги по всему вышеописанному, можно с уверенность сказать, что цель диплома выполнена. Получилось простое и понятное веб-приложение с широким набором возможностей, которые призваны обеспечить поддержку принятия решений для компании АО «Вертекс». Причем разработанное приложение облегчает работу двум отделам, о чем было сказано выше.

# **ЗАКЛЮЧЕНИЕ**

В процессе работы над выпускной квалификационной работой было проведено исследование и анализ фармацевтического рынка России за последние годы. В том числе, было изучено влияние импортзамещения на рынок в целом и развитие розничных продаж в аптечном сегменте. Выяснилось, что, несмотря на текущую ситуацию, фармацевтический рынок России развивается и движется в нужном направлении. Это проявляется в государственной поддержке, когда создаются национальные проекты здравоохранения и выдаются льготы, патенты и денежные средства на клинические исследования, испытания и фармацевтическое производство. Кроме того, развивается отечественное производство и все больше и больше российских фармацевтических компаний выходит в ТОП-20. Также было выяснено, что последние годы растут розничные продажи, особенно в онлайн-сегменте фармацевтического рынка. Онлайн-продажи становятся популярными, поэтому у каждой фармацевтической компании появилась необходимость собирать по ним данные и анализировать их.

Кроме того, была разработана потоковая модель продаж оригинальных лекарственных препаратов и генериков на фармацевтическом рынке России. С помощью нее был построен прогноз на пять ближайших лет, оценена ситуация и представлены возможные стратегии поведения российских геенриков, которые призваны улучшить их положения на рынке.

Для работы с онлайн-ресурсами были изучены разнообразные инструменты и библиотеки языка программирования Python. Кроме того, были исследованы и проанализированы разные способы веб-скрапинга данных. В ходе изучения структуры DOM-дерева различных онлайн-источников были выявлены закономерности и построен план разработки алгоритма. Следуя данному плану, удалось автоматизировать сбор данных о фармацевтической продукции из открытых источников и создать алгоритм.

И, наконец, было разработано веб-приложение для обеспечения поддержки принятия решений. Оно предлагает возможности по обновлению информации в базе данных, а также просмотру ее содержимого и финансовой отчетности компании. А для изучения ситуации на рынке и оценки положения компании АО «Вертекс» на нем был проведен был создан специальный раздел в указанном веб-приложении. В нем предоставлен общий анализ рынка, внешний финансовый анализ экономической деятельности компании и анализ ценообразования для конкретных лекарственных препаратов.

Поставленная цель и задачи выпускной квалификационной работы выполнены в полном объеме.

# **СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ**

1. Банк В.Р. Финансовый анализ: учебное пособие / Банк В.Р., Банк С.В., Тараскина А.В. – М.: Проспект, 2007. – 344 с.
2. Амбудж Агравал. Автоматизация предприятия с помощью Python / Амбудж Агравал. – 2022. – 202 с. – URL: https://coollib.cc/b/615564-ambudzh-agraval-avtomatizatsiya-predpriyatiya-s-pomoschyu-python/readp (дата обращения: 12.02.2025).
3. Дьяконова О.С. Бухгалтерская отчетность, как источник экономической информации / Дьяконова О.С., Моргун Н.А. // Вестник университета. – 2016. – №11. – URL: https://cyberleninka.ru/article/n/buhgalterskaya-otchetnost-kak-istochnik-ekonomicheskoy-informatsii (Дата обращения: 04.05.25)
4. Райан Митчелл. Современный скрапинг веб-сайтов с помощью Python / Райан Митчелл. – 2021. – 336 с. – URL: https://www.litres.ru/book/r-mitchell/sovremennyy-skraping-veb-saytov-s-pomoschu-python-pdf-epub-64822502/ (дата обращения: 28.02.2025).
5. Аналитические отчеты [Электронный ресурс] // dsm.ru: сайт / DSM Group. – URL: https://dsm.ru/news-reports/ (дата обращения: 14.03.25).
6. Аналитические отчеты [Электронный ресурс] // alpharm.ru: сайт / ALPHA RESEARCH&MARKTING – URL: https://alpharm.ru/ (дата обращения: 14.03.25).
7. Главная [Электронный ресурс] // vertex.spb.ru: сайт / Вертекс – URL: https://vertex.spb.ru/ (дата обращения: 06.02.25).
8. Программы [Электронный ресурс] // minzdrav.gov.ru: сайт / Министерство здравоохранения Российской Федерации – URL: https://minzdrav.gov.ru/ministry/programms (дата обращения: 15.03.25).
9. Ресурс БФО [Электронный ресурс] // bo.nalog.gov.ru: сайт / Государственный информационный ресурс бухгалтерской (финансовой) отчетности – URL: https://bo.nalog.gov.ru/#balance (дата обращения: 02.05.25).
10. Apache JMeter. Выпуск 5.6.3 : программа / Apache Software Foundation, 2024. – Лицензия: Apache License 2.0 – URL: https://jmeter.apache.org/usermanual/index.html (дата обращения: 11.02.25).
11. Beautiful Soup. Версия 4.13.0 : библиотека Python / Crummy, 2024. – Лицензия: MIT. – URL: https://www.crummy.com/software/BeautifulSoup/bs4/doc/ (дата обращения: 25.03.25).
12. Chrome DevTools. Выпуск 137.0.7151.56 : программа / Google, 2024. – Лицензия: BSD – URL: https://developer.chrome.com/docs/devtools?hl=ru (дата обращения: 19.02.25).
13. Dash. Версия 3.0.4 : библиотека Python / Plotly, 2024. – Лицензия: Python Software Foundation License. – URL: https://dash.plotly.com (дата обращения: 28.04.25).
14. Re. Версия 3.13.4 : библиотека Python / Python Software Foundation, 2024. – Лицензия: Python Software Foundation License. – URL: https://docs.python.org/3/library/re.html (дата обращения: 07.04.25).
15. Requests. Версия 2.32.3 : библиотека Python / Python Software Foundation, 2024. – Лицензия: Python Software Foundation License. – URL: https://pypi.org/project/requests/ (дата обращения: 10.02.25).
16. Selenium. Версия 3.141.0 : библиотека Python / Selenium, 2020. – Лицензия: Apache License 2.0. – URL: https://selenium-python.readthedocs.io/ (дата обращения: 19.02.25).
17. Selenium-stealth. Версия 1.0.6 : библиотека Python / Selenium, 2020. – Лицензия: MIT. – URL: https://pypi.org/project/selenium-stealth/ (дата обращения: 21.02.25).
18. SQLAlchemy. Версия 2.0.41 : библиотека Python / Mike Bayer, 2025. – Лицензия: MIT. – URL: https://www.sqlalchemy.org/ (дата обращения: 20.04.25).
19. Akbar Abdollahiasl. A system dynamics model for national drug policy / Akbar Abdollahiasl, Abbas Kebriaeezadeh, Rassoul Dinarvand, Mohammad Abdollahi, Abdol Majid Cheraghali, Mona Jaberidoost, Shekoufeh Nikfar // DARU Journal of Pharmaceutical Sciences. – 2014. – С. 22-34. – URL: https://link.springer.com/journal/40199 (дата обращения: 15.03.25).
20. Peng C.Q. Gazing through time and beyond the health sector: Insights from a system dynamics model of cardiovascular disease in Australia / Peng C.Q., Lawson K.D., Heffeman M., McDonnell G., Liew D., Lybrand S. // PLoS ONE. – 2021. – №16 – URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0257760 (дата обращения: 16.03.25).