МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ

РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение

высшего образования

«Московский государственный технический университет имени Н.Э. Баумана

(национальный исследовательский университет)»

**ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА**

**по курсу**

«Data Science Pro»

Прогнозирование продаж неликвидных товаров аптечного ассортимента в сети аптек

Слушатель Степанченко Виктория Васильевна

Москва, 2024

**СОДЕРЖАНИЕ**

1. Введение
2. Аналитическая часть
   1. Постановка задачи.
   2. Предварительная работа с данными
   3. Разведочный анализ данных и предобработка данных
3. Описание используемых методов
4. Практическая часть

4.1.Разработка и обучение модели

4.2.Тестирование модели

4.3.Разработка приложения

1. Создание удаленного репозитория   
   и загрузка результатов работы на него.

**Введение**

Перераспределение товарных запасов аптечной сети является одним из основных бизнес-процессов большинства аптечных сетей. Это эффективный инструмент для достижения оптимизации товарных запасов сети в целом и конкретного подразделения в частности. Оборачиваемость запасов имеет особое значение в производственных и торговых компаниях с интенсивным управлением запасами. Низкий уровень оборачиваемости запасов может указывать на затоваривание, моральное устаревание или недостатки в линейке продуктов или маркетинговых усилиях. Однако в некоторых случаях, низкое значение оборачиваемости может быть экономически обоснованным, например, в ожидании быстрого роста цен или ожидаемого дефицита на рынке. Вместе с тем, при низкой и/или медленной оборачиваемости запасов определенного подразделения растут и их затраты на складирование. Напротив, высокая скорость оборота может указывать на недостаточный уровень запасов, что может привести к убыткам в бизнесе из-за их нехватки. Кроме того, товар, запасы которого продаются (оборачиваются) один раз в год, имеет более высокую стоимость хранения, чем товар, запасы которого за это время оборачиваются дважды, трижды или более. Оборачиваемость запасов также указывает на активность хозяйственной деятельности (торговых операций) компании.

Чем выше скорость оборачиваемости запасов, тем ниже затраты капитала, и наоборот. Капитал, находящийся на складе (в аптеке), недоступен для других целей (инвестирование, обслуживание долга) и может быть высвобожден только за счет увеличения оборачиваемости запасов. Таким образом, скорость оборачиваемости запасов также влияет на ликвидность компании. Высокий уровень оборачиваемости запасов также повышает рентабельность и, следовательно, прибыль, пока выручка от продаж товаров остается постоянным. Чем выше определенные виды затрат, тем выше должна быть скорость оборачиваемости запасов для компенсации роста указанных затрат. То есть, увеличение оборачиваемости запасов позволяет снизить складские расходы, что позволяет компании экономить средства на аренду, коммунальные услуги, страховку, на защиту от кражи и другие затраты на поддержание товарно-материальных запасов.

Основные причины образования неликвидного товара в подразделениях аптечной сети:

1. Товары аптечного ассортимента, заказанные под конкретного покупателя, но не востребованные им.
2. Сезонность некоторых групп товаров аптечного ассортимента
3. Резкое снижение покупательского спроса на конкретную позицию, в связи, например, со сменой или увольнением профильного врача в близлежащем лечебно-профилактическом учреждении, постоянным назначением которого, являлась эта позиция.
4. Снижение спроса после пиковой востребованности товаров аптечного ассортимента. Примером является снижение востребованности после окончания COVID-19 на системы диагностики вируса, а так же на лекарственные препараты, входящие в рекомендации к лечению. Другим примером являются пики востребованности на дефектурные товары. Дефектура (отсутствие или снижение предложений на рынке) появляется по разным причинам и может быть полной (прекращение производства, импортирования) или временной (при получении нового регистрационного удостоверения лекарственный препарат пропадает с рынка на срок, требуемый для выпуска партий с новым удостоверением).
5. Появлением дженериков ( лекарственное средство, содержащее химическое вещество — активный фармацевтический ингредиент, идентичный запатентованному компанией — первоначальным разработчиком лекарства. Дженерики производятся и продаются после истечения срока действия патента и, соответственно, монополии компании-патентодержателя другими производителями под международным непатентованным названием либо под коммерческим названием, отличающимся от фирменного названия разработчика препарата). При выводе на рынок дженериков производители сохраняют низкое ценовое предложение, а так же используют маркетинговые активности (рекламу, фарм. кружки, информирование врачей и профильных специалистов), что может вести к резкому падению спроса на оригинальный лекарственный препарат.

**Постановка задачи**

В настоящей работе изучаются методы прогнозирования продаж конкретного товара с идентификаций по уникальному коду для выявления целевого подразделения для перераспределения неликвидов. Польза от успешной разработки методов заключается в единообразии подхода, скорости и уменьшения временных затрат на принятие решения по движению неликвидов внутри аптечной сети.

В работу поступили данные из товароучетной системы аптечной сети. Данные собирались в автоматическом режиме на всех подразделениях сети в режиме реального времени (регистрация продаж с использованием контрольно – кассовых машин).

Исходные данные: файлы в формате MS Excel разделенные на 2 временных промежутка - Train data – с 01.04.2022 по 31.03.2024 года, Test data – с 01.04.2024 по 31.07.2024 года. Данные представляют собой выборку продаж пяти случайных подразделений аптечной сети. Train data содержит 359436 записей, число параметров – 5. Test data содержит 54293 записи, число параметров – 5. Файлы идентичны по структуре и не требуют дополнительной обработки (Рис.1).

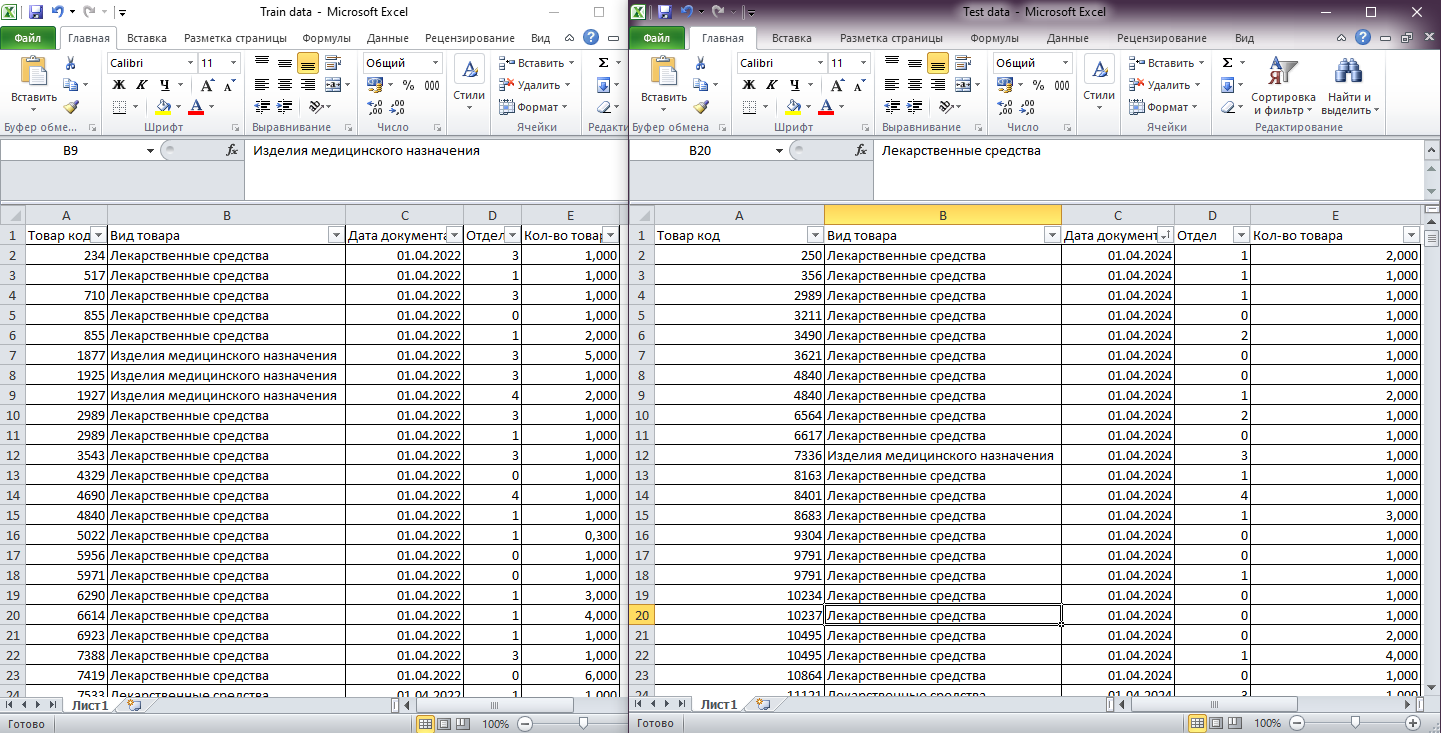


Рисунок 1 – Вид сырых данных и снимки экрана c содержимым   
обоих из файлов xlsx.

Данные были выгруженных мной из товароучетной системы, поэтому распределение дат и название были определены заранее с целью использования в работе.

Датасет представляет из себя временной ряд, собранный по пяти подразделениям, за двухлетний период по 12769 уникальным кодам товара, с разделением по виду товара.

Временной ряд (динамический ряд, ряд динамики) — собранный в разные моменты времени статистический материал о значении каких-либо параметров (в простейшем случае одного) исследуемого процесса. Каждая единица статистического материала называется измерением или отсчётом, также допустимо называть его уровнем на указанный с ним момент времени. Во временном ряде для каждого отсчёта должно быть указано время измерения или номер измерения по порядку. Временной ряд существенно отличается от простой выборки данных, так как при анализе учитывается взаимосвязь измерений со временем, а не только статистическое разнообразие и статистические характеристики выборки

Задача состоит в том, чтобы при поступлении запроса на определенный код неликвидного товара были сформированы прогнозы по реализации на каждом подразделении и выбрано подразделение с лучшим шансом его реализации за 4 месяца.

# **Предварительная работа с файлами**

Для удобства названия колонок были переименованы с русского языка на английский. Наименования подразделений унифицированы до порядковых номеров от 0 до 4.

**Разведочный анализ данных и предобработка данных**

Разведочный анализ данных ([англ.](https://ru.wikipedia.org/wiki/%D0%90%D0%BD%D0%B3%D0%BB%D0%B8%D0%B9%D1%81%D0%BA%D0%B8%D0%B9_%D1%8F%D0%B7%D1%8B%D0%BA) exploratory data analysis, EDA) — анализ основных свойств данных, нахождение в них общих закономерностей, распределений и аномалий, построение начальных моделей, зачастую с использованием инструментов [визуализации](https://ru.wikipedia.org/wiki/%D0%92%D0%B8%D0%B7%D1%83%D0%B0%D0%BB%D0%B8%D0%B7%D0%B0%D1%86%D0%B8%D1%8F_%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D1%85).

Понятие введено математиком [Джоном Тьюки](https://ru.wikipedia.org/wiki/%D0%A2%D1%8C%D1%8E%D0%BA%D0%B8,_%D0%94%D0%B6%D0%BE%D0%BD), который сформулировал цели такого анализа следующим образом:

* максимальное «проникновение» в данные,
* выявление основных структур,
* выбор наиболее важных переменных,
* обнаружение отклонений и аномалий,
* проверка основных гипотез,
* разработка начальных моделей.

Были использованы библиотеки:

1. pandas: Pandas — это библиотека для работы с данными, которая предоставляет удобные структуры данных, такие как DataFrame, и множество функций для анализа и манипуляции данными. Она идеально подходит для работы с временными рядами, так как позволяет легко хранить и анализировать временные данные.

2. numpy: NumPy — это библиотека для работы с массивами и матрицами чисел. Она полезна при выполнении вычислений над данными временных рядов.

3. matplotlib и seaborn: Эти библиотеки позволяют создавать графики и визуализации данных, что особенно важно при анализе временных рядов.

4. statsmodels: Statsmodels — это библиотека для статистического анализа данных. Она содержит множество методов для анализа временных рядов, включая модели ARIMA и SARIMA.

5. scikit-learn: Scikit-learn предоставляет множество инструментов для машинного обучения, включая модели регрессии и классификации, которые можно использовать для анализа временных рядов.

Датасет велик, однороден, собран качественно, ошибок при первоначальном разведочном анализе выявлено не было.

Были применены методы библиотеки pandas для проведения разведочного анализа (exploratory data analysis – EDA) датасета (sample, head, tail, shape, info, columns): изучили структуру и характеристики данных, типы данных, наличие пропусков, дубликатов, ошибок заполнения столбцов датасета.

Тип данных - числа с плавающей точкой, двойной точности **(float64)**. Пропуски отсутствуют. Описательная статистика и размерность данных представлена на Рис. 2.

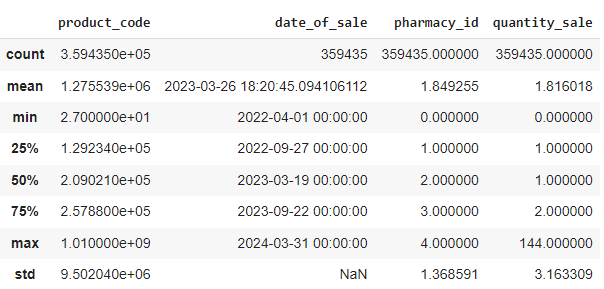


Рисунок 2– Описательная статистика и размерность  
 исходных данных после грубой очистки.

По виду товара продажи в датасете распределены следующим образом (Рис. 3):

Основную долю реализации стоставляют лекарственные средства, на втором месте изделия медицинского назначения, на третьем месте полиграфическая продукция. Полиграфической продукцией можно пренебречь – это упакет – майка, и эти данные не несут ценности. В продажах изделий мед назначения большую долю занимают бахилы, маски и контейнеры – товары первого спроса, продающиеся часто, они не бывают неликвидами, особенно в нашей аптечной сети. Основным анализируемым видом товара являются лекарственные средсва. В связи с этим перед анализом данных признак «Вид товара» удален.

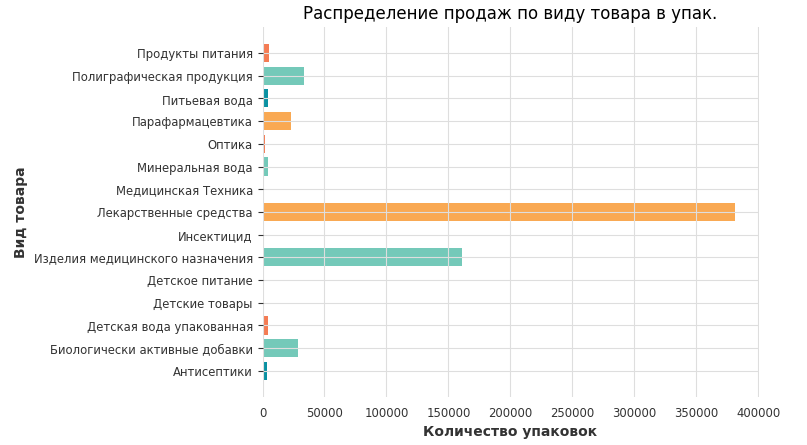


Рисунок 3 – Распределение продаж по виду товара.

Исходя из условий задачи, что необходимо производить анализ по каждому подразделению отдельно оба датасета были разделены на 5 датасетов по признаку «Отдел».

Аптеки, представленные в датасете, имеют разный товарооборот (Рис. 4)

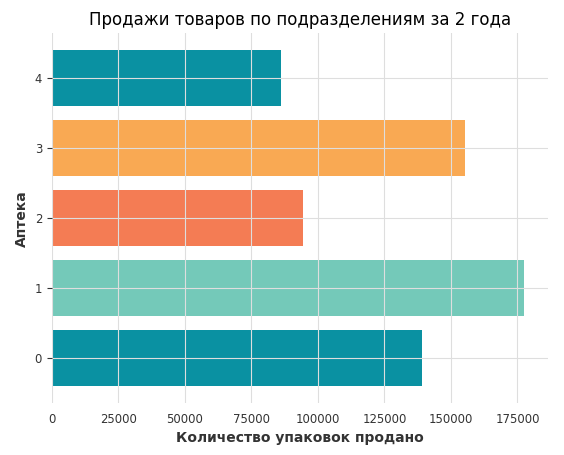


Рисунок 4– Распределение продаж по виду товара.

Затем была введена переменная «Код товара» и произведен окончательный отбор датасетов с продажами именно по этому коду по каждому подразделению отдельно. Продажи в датасетах были представлены по дням, для снижения частотности были преобразованы до суммы продаж в каждом интервале - месяце.

Был произведен поиск выбросов – пикового спроса, при помощи фильтра Хэмплея. Главная цель Хэмплея – найти и заменить выбросы в заданном временном ряду. Для этого в своей основе он использует скользящее среднее с заданным окном. Для каждой итерации или окна фильтр вычисляет медиану и стандартное отклонение. Оно выражается в среднем абсолютном значении и обозначается как MAD.



Чтобы MAD стал последовательной оценкой стандартного отклонения надо умножить его на постоянный коэффициент k. Коэффициент зависит от распределения. Мы считаем, что данные подчиняются распределению Гаусса, поэтому берём коэффициент равным 1,4826.



Если значение медианы окна скользящего среднего больше чем *х* стандартных отклонений, то это – выброс.

Фильтр Хэмплея имеет 2 настраиваемых параметра:

1. размер раздвижного окна
2. количество стандартных отклонений, которые идентифицируют выброс. В датасете применили 3 стандартных отклонения.

Для визуализации был использован Boxplot (Рис. 5) или Ящик с усами. Ящик с усами, диаграмма размаха, усиковая диаграмма, коробчатая диаграмма, блочная диаграмма с ограничителями выбросов (англ. box-and-whiskers diagram, box-and-whiskers plot, box plot) — график, использующийся в описательной статистике, компактно изображающий одномерное распределение вероятностей.

Такой вид диаграммы в удобной форме показывает медиану (или, если нужно, среднее), нижний и верхний квартили, минимальное и максимальное значение выборки и выбросы. Несколько таких ящиков можно нарисовать бок о бок, чтобы визуально сравнивать одно распределение с другим; их можно располагать как горизонтально, так и вертикально. Расстояния между различными частями ящика позволяют определить степень разброса (дисперсии) и асимметрии данных и выявить выбросы. 

Границами ящика служат первый и третий квартили (25-й и 75-й процентили соответственно), линия в середине ящика — медиана (50-й процентиль).[5] Концы усов — края статистически значимой выборки (без выбросов), и они могут определяться несколькими способами. Наиболее распространённые значения, определяющие длину «усов»:

Рисунок 5 –Фильтр Хэмплея и Boxplot

1. Минимальное и максимальное наблюдаемые значения данных по выборке
2. Разность первого квартиля и полутора межквартильных расстояний; сумма третьего квартиля и полутора межквартильных расстояний

Для двух случайных датасетов из 5 сформированных строим графики (Рис. 6):

1. график временного ряда — это линейный график, который показывает данные, такие как измерения, продажи или частоты за определенный период времени. Они могут использоваться для отображения закономерности или тенденции в данных и полезны для составления прогнозов на будущее, таких как прогнозирование погоды или финансовый рост.
2. график скользящего среднего. Скользя́щее сре́днее (англ. moving average, MA) — общее название для семейства функций, значения которых в каждой точке определения равны некоторому среднему значению исходной функции за предыдущий период. Скользящие средние обычно используются с данными временных рядов для сглаживания краткосрочных колебаний и выделения основных тенденций или циклов
3. график автокорреляции. Автокорреляционная функция (АКФ) — зависимость взаимосвязи между функцией (сигналом) и её сдвинутой по аргументу функции копией от величины сдвига. Автокорреляционная функция играет важную роль в математическом моделировании и анализе временных рядов, показывая характерные времена для исследуемых процессов.

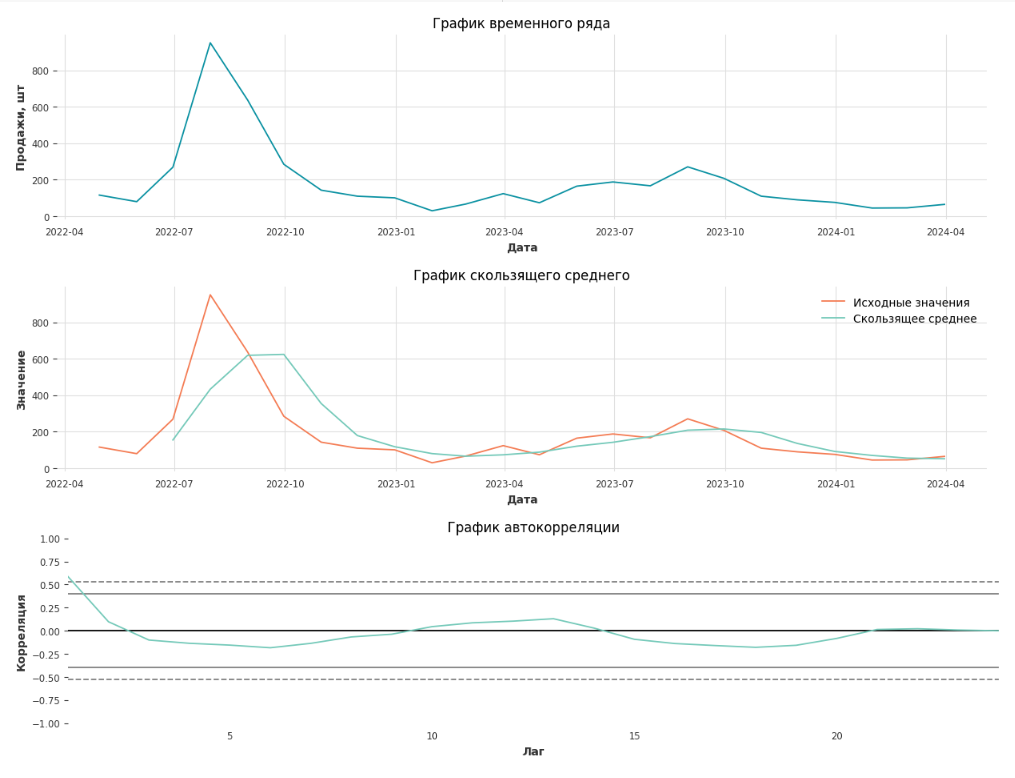


Рисунок 6 –Графики временного ряда

Проверка на стационарность ряда.

Стационарность - это свойство временного ряда, которое означает, что его средние и стандартные отклонения не меняются со временем. Если временной ряд является стационарным, то его можно легко анализировать и прогнозировать.

Нестационарный временной ряд может иметь тренд (постоянный рост или падение), цикличность (повторение циклов) или сезонность (повторение определенных событий в разное время года).

Один из способов проверки стационарности временных рядов - это использование графических методов, таких как график временных рядов, график автокорреляции и график частной автокорреляции. График временного ряда позволяет визуализировать изменения значений ряда с течением времени, график автокорреляции показывает корреляцию между значениями ряда в разные периоды времени, а график частной автокорреляции учитывает корреляцию только между двумя значениями, пропуская все промежуточные значения.

Было проведено тестирование на стационарность ряда при помощи расширенного теста Тест Дики — Фуллера (DF-тест, Dickey — Fuller test). Тест Дики — Фуллера это методика, которая используется в прикладной статистике и эконометрике для анализа временных рядов для проверки на стационарность. Временной ряд называется «стационарным», если он не имеет тренда, демонстрирует постоянную дисперсию во времени и имеет постоянную структуру автокорреляции во времени.

Один из способов проверить, является ли временной ряд стационарным, — это выполнить расширенный тест Дики-Фуллера, в котором используются следующие нулевая и альтернативная гипотезы:

1. временной ряд является нестационарным. Другими словами, он имеет некоторую структуру, зависящую от времени, и не имеет постоянной дисперсии во времени.
2. временной ряд является стационарным.

Если p-значение из теста меньше некоторого уровня значимости (например, α = 0,05), то мы можем отвергнуть нулевую гипотезу и сделать вывод, что временной ряд является стационарным.

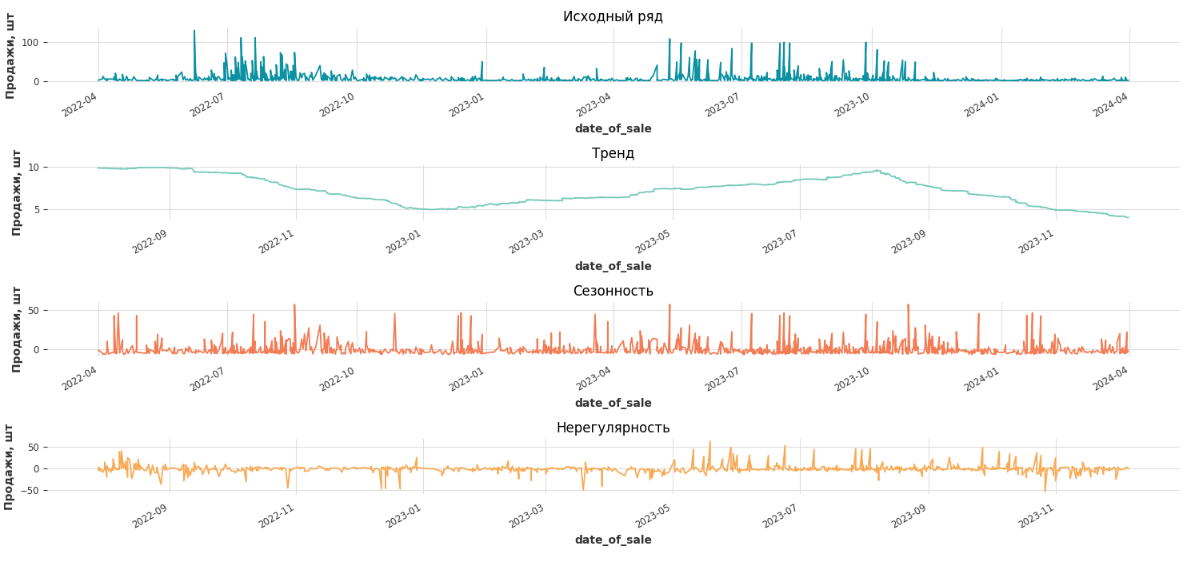
Тест был проведен на данных с разной частотностью – по дням и по месяцам, для каждой пары набора данных.

Значения в тесте по сводным данным по всем подразделениям меньше уровня значимости (в основном это 0.05), то позволительно отклонить нулевую гипотезу о нестационарности данных и считать их стационарными.

В тестах по каждой аптеке по ряду с данными по месяцам ряды не стационарны по всем подразделениям, кроме Аптеке №1 и №4.

Ввиду вышесказанного и того условия, что код товара будет меняться, и всегда есть шанс получить датасет с нестационарными рядами, следует преобразовать ряд для достижения стационарности методом дифференцирования.

Были построены графики компонент временных рядов. Для набора данных по дням и месяцам по всей аптечной сети одновременно (Рис . 7)



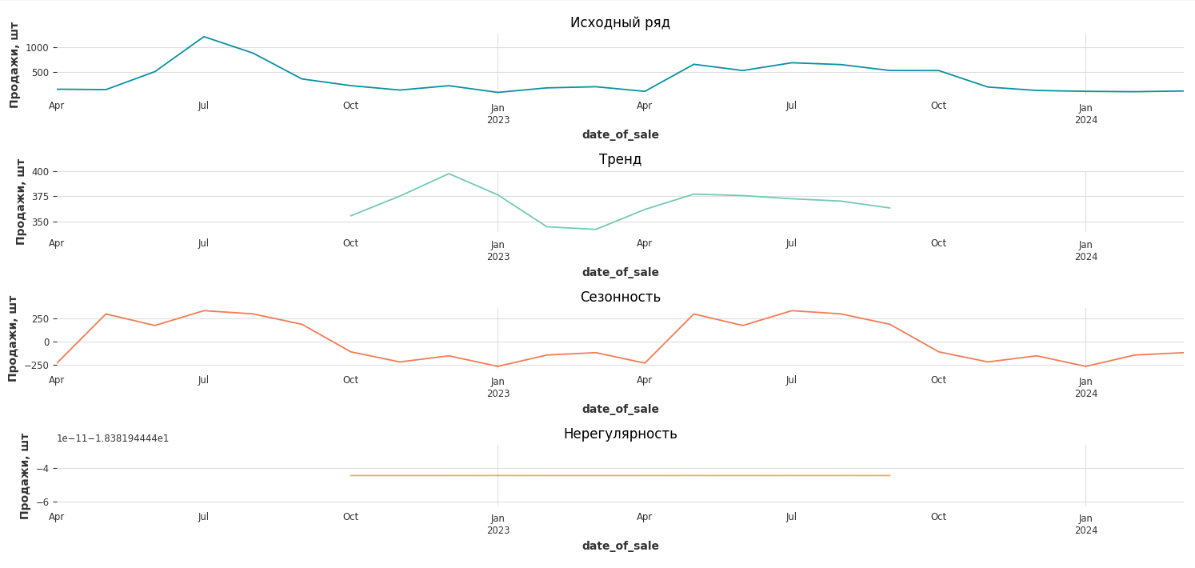


Рисунок 7 – Компоненты временных рядов

Компоненты временных рядов:

\* Тренд (Trend): показывает общее направление движения данных временного ряда на протяжении длительного времени. Это может быть восходящий тренд (увеличение значений), нисходящий тренд (уменьшение значений) или отсутствие тренда (значения остаются примерно на одном уровне).

\* Сезонность (Seasonality): повторяющиеся колебания данных временного ряда, которые происходят с регулярными интервалами времени, например, ежемесячно, ежеквартально или ежегодно.

\* Цикличность (Cyclical): колебания временного ряда, которые происходят на более длительных временных интервалах и не обязательно регулярны. Циклы могут быть связаны с экономическими, политическими или другими внешними факторами.

\* Нерегулярность/остатки (Irregularity/residuals): представляют собой случайные и непредсказуемые изменения в данных временного ряда, которые не связаны с трендом, сезонностью или цикличностью.

**Описание используемых методов**

ARIMA (англ. autoregressive integrated moving average, иногда модель Бокса — Дженкинса, методология Бокса — Дженкинса) — интегрированная модель авторегрессии — скользящего среднего — модель и методология анализа временных рядов. Является расширением моделей ARMA для нестационарных временных рядов, которые можно сделать стационарными взятием разностей некоторого порядка от исходного временного ряда (так называемые интегрированные или разностно-стационарные временные ряды). Модель ARIMA(p,d,q) означает, что разности временного ряда порядка d подчиняются модели ARMA(p,q)

ARIMA — это статистическая модель, которая используется для анализа и прогнозирования данных динамических рядов.

Авторегрессионная (AR) часть относится к использованию зависимости между текущим наблюдением и определенным количеством предыдущих наблюдений. Например, чтобы предсказать погоду на завтра, мы смотрим погоду в предыдущие дни.

Интегрированная (I) часть обозначает, что данные временного ряда преобразуются с целью сделать ряд стационарным. Под стационарностью подразумевается такое свойство временного ряда, при котором его статистические характеристики не меняются во времени. Например, среднее и дисперсия остаются постоянными со временем. Обычно это достигается путем применения операции разности к ряду один или несколько раз.

Скользящее среднее (MA) относится к моделированию ошибки между реальным наблюдением и прогнозируемым значением, используя комбинацию значений ошибок предыдущих шагов.

Модель ARIMA обычно обозначается как ARIMA (p, d, q), где p, d и q — параметры модели:

p — порядок авторегрессии, который описывает количество предыдущих наблюдений, принимаемых во внимание в модели.

d — порядок интегрирования, который указывает, сколько раз нужно взять разность временного ряда для достижения стационарности.

q — порядок скользящего среднего, который связан с количеством предыдущих ошибок, используемых в модели.

Выбор подходящих значений (p, d, q) в модели ARIMA — это искусство и наука, требующая обращения с автокорреляционными и частными автокорреляционными графиками, а также применения информационных критериев.

Авторегрессионные модели (AR-модели) используют прошлые значения ряда для прогнозирования его будущих значений. Эта модель предполагает, что текущее значение ряда зависит от его предыдущих значений. Наиболее популярной AR-моделью является модель первого порядка (AR(1)), которая предполагает, что текущее значение ряда зависит только от его предыдущего значения.

Скользящее среднее (MA-модели). Эта модель использует прошлые значения ошибок - разницу между фактическими значениями ряда и его прогнозируемыми значениями, для прогнозирования будущих значений. Наиболее популярной MA-моделью является модель первого порядка (MA(1)), которая предполагает, что текущая ошибка зависит только от ее предыдущего значения.

**Разработка и обучение моделей**

Для прогнозирования и отбора результатов, наиболее полно отражающих реальные продажи за прогнозируемый период ( используется датасет Test data с аналогичной разбивкой общего датасета на датасеты по подразделениям, в показателями в суммовом эквиваленте за период 4 мес по каждому подразделению по тестируемому коду товара) выполнена разработка следующих моделей:

* Авторегрессивная модель (Auto-Regressive Model, AR)
* Модель SARIMAX
* Метод скользящего среднего (Moving Average)
* Модель ARIMA

Метод XGBoost разработан в качестве примера возможности анализа временных рядов. В рамках задачи, ввиду того что код товара, переданный модели может быть любой, то нет возможности подбора оптимальных параметров для данной модели. На выбранном примере метода показал неплохой результат прогнозирования, но не был выбран в качестве основной модели.

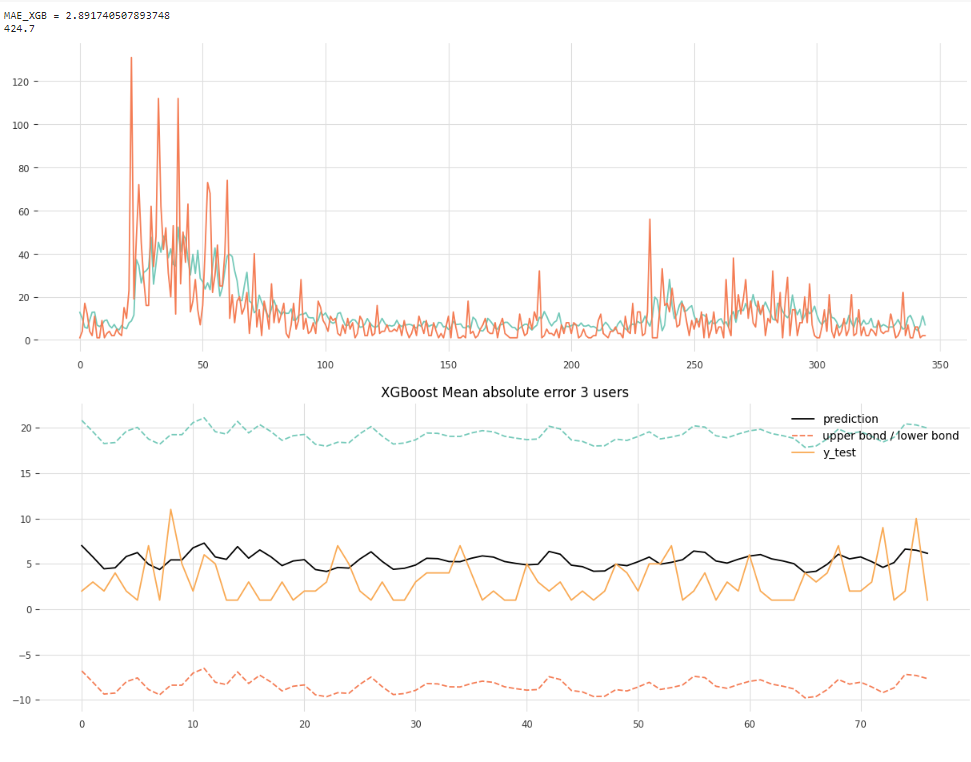


Рисунок 7 – Результат применения XGBoost

Вывод по применению моделей:

Сравнение методов производилось не по ошибкам, а по соответствию рассчитанных методом значений со значениями реальных продаж за этот период.

Наиболее близким к тестовому результату оказался метод, близкий к тому, что обычно используется для прогноза закупок и продаж по средним величинам продаж за предыдущий сопоставимый период - Метод скользящего среднего (Moving Average)

Сопоставимые показатели по 3/5 подразделений показала модель ARIMA. Модели Авторегрессии и SARIMAX – показали сопоставимые результаты в 1/ 5 случаев (Рис. 8).

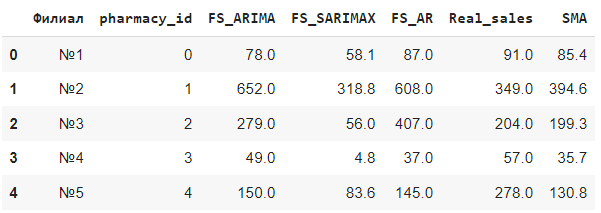


Рисунок 8 – Сравнение прогнозов продаж за 4 мес. моделями

# **Разработка приложения**

Приложение для получения информации о наилучшем подразделении для перераспределения и дальнейшей реализации неходового товара будет разработано с помощью инструмента Flask**.** Приложение будет иметь минималистичный стиль, будет иметь 2 окна для ввода информации: «Код товара» и остаточный срок годности в месяцах. После ввода данных необходимо нажать кнопку «Ввод» При вводе информации в первое окно будут производиться расчеты прогноза реализации за квартал. Если во второе окно вводится остаточный срок годности 6 месяцев и менее рассчитанные показатели уменьшаются вдвое. В поле ниже окон для запроса будет выводиться результат в виде номера подразделения, куда необходимо переместить неликвидный товар или в случае, если согласно расчетам такого подразделения не будет выявлено, выведется «Товар не востребован».

# **Создание удаленного репозитория и загрузка на него результатов исследовательской работы:**

https://github.com/vetalaz/VKR

**Используемая литература**

1. Commercial Loan Analysis: principles and techniques for credit analysts and lenders By Kenneth R. Pirok
2. Financial Analysis Reports. Bruin Financial Management. Дата обращения: 28 июля 2019.
3. Willy Schneider/Alexander Hennig, Lexikon Kennzahlen für Marketing und Vertrieb, 2008, S. 209
4. What does “generic” mean? (англ.). WTO (сентябрь 2006). Дата обращения: 9 сентября 2014.
5. П. Брюс, Э. Брюс. 1. Разведочный анализ данных // Практическая статистика для специалистов Data Science. — СПб.: БХВ-Петербург, 2018. — С. 19—58. — 304 с.
6. Frigge, M.[англ.]; Hoaglin, D. C.[англ.]; Iglewicz, B.[англ.]. Some Implementations of the Boxplot (англ.) // The American Statistician[англ.] : journal. — 1989. — Vol. 43, no. 1. — P. 50—54.
7. Альтхоф К. #Сам себе программист — М.: Бомбора, 2022
8. Грас, Джоэл. Data Science. Наука о данных с нуля - СПб.:  
    БХВ-Петербурr, 2021
9. Стивен Б. Акелис. Скользящие средние значения (Moving Averages) // Технический анализ от А до Я. Полный набор инструментов торговли… от «Абсолютного индекса ширины» до «Японских свечей» = Technical Analysis from A to Z: Covers Every Trading Tool... from the Absolute Breadth Index to the Zig Zag / Пер. с англ. М. Волкова, А. Лебедева. — М.: Диаграмма, 1999. — С. 26—29. — 376 с. — ISBN 978-5-902537-13-7, 5-900082-05-09, ГРНТИ 06.73, ББК 65.526. Архивировано 14 апреля 2012 года.
10. Стивен Б. Акелис. Скользящие средние значения (Moving Averages) // Технический анализ от А до Я. Полный набор инструментов торговли… от «Абсолютного индекса ширины» до «Японских свечей» = Technical Analysis from A to Z: Covers Every Trading Tool... from the Absolute Breadth Index to the Zig Zag / Пер. с англ. М. Волкова, А. Лебедева. — М.: Диаграмма, 1999. — С. 198—205. — 376 с. — ISBN 978-5-902537-13-7, 5-900082-05-09, ГРНТИ 06.73, ББК 65.526
11. https://datascientest.com/en/sarimax-model-what-is-it-how-can-it-be-applied-to-time-series
12. https://pandas.pydata.org/
13. https://habr.com/ru/companies/mvideo/articles/769190/
14. https://www.statsmodels.org/dev/generated/statsmodels.tsa.arima.model.ARIMA.html
15. https://habr.com/ru/companies/otus/articles/732080/
16. https://habr.com/ru/articles/752662/