

Санкт-Петербургский государственный университет

Кафедра информационно-аналитических систем

Группа 23.Б10-мм

Прогнозирование рынка подсолнечного масла

Мусаев Немат Адалат оглы

Отчёт по учебной практике
в форме «Решение»

Научный руководитель:
старший преподаватель кафедры информационно-аналитических систем Азимов Р. Ш.

Санкт-Петербург
2025

Оглавление

Введение	3
1. Постановка задачи	4
2. Обзор существующих подходов	5
3. Решение	8
3.1. Анализ данных и EDA	8
3.2. Прогноз цен на масло	8
3.3. Прогноз выручки клиентов (временные ряды)	10
3.4. Анализ влияния бизнес-решений	12
3.5. Визуализация результатов	13
4. Заключение	14
Список литературы	15

Введение

Рынок подсолнечного масла относится к числу экономически значимых и нестабильных рынков. На динамику цен и объёмов реализации влияют сезонность сельскохозяйственного производства, урожайность, региональные особенности, валютный курс, экспортные пошлины. В результате компании и аналитики сталкиваются с задачей прогнозирования показателей в условиях неопределённости и неполной информации ввиду также вышеупомянутых факторов.

В данном проекте рассматривается задача анализа и прогнозирования цен на подсолнечное масло, а также выручки клиентов на основе исторических данных. Работа ориентирована на практическое применение методов анализа данных и временных рядов и может рассматриваться как основа для выпускной квалификационной работы.

Актуальность проекта связана с тем, что подобные задачи регулярно возникают в бизнесе и экономике, а их решение требует сочетания статистических методов, машинного обучения и корректной интерпретации результатов. При этом задача не является тривиальной и позволяет продемонстрировать владение современными инструментами анализа данных.

1. Постановка задачи

В рамках исследования для рынка подсолнечного масла были поставлены следующие задачи:

- Провести разведочный анализ данных (EDA) для исследования динамики цен и выручки.
- Выполнить feature engineering для создания признаков, пригодных для построения моделей машинного обучения.
- Реализовать модель прогнозирования цен с применением Lasso-регрессии.
- Построить и сравнить модели прогнозирования временных рядов для выручки клиентов:
 - Simple Exponential Smoothing;
 - Модель Холта–Винтерса (Holt–Winters);
 - SARIMA;
 - Ансамблевый подход (Ensemble).
- Провести А/В тестирование разработанных моделей временных рядов для определения наиболее эффективного подхода к прогнозированию.
- Проанализировать влияние бизнес-решений (в частности, увеличения посевых площадей) на выручку и конверсии в покупку с использованием статистических критериев.
- Визуализировать исходную динамику данных, результаты прогнозирования и итоги статистических проверок.

2. Обзор существующих подходов

Задачи анализа и прогнозирования экономических показателей на товарных рынках являются одним из наиболее распространённых направлений прикладной аналитики данных. Особенно это относится к аграрным рынкам, где временные ряды, как правило, обладают выраженной сезонностью, трендами и высокой чувствительностью к внешним факторам. Рынок подсолнечного масла не является исключением и представляет собой наглядный пример сложной, но хорошо формализуемой предметной области.

В научных исследованиях и прикладных проектах для решения подобных задач широко используются методы анализа временных рядов. Классические модели, такие как:

Модели для анализа и прогноза временных рядов

<

- Simple Exponential Smoothing и Holt–Winters
- Holt–Winters [1]
- Sarima [4]
- Ensambles

>

Simple Exponential Smoothing и Holt–Winters применяются для краткосрочного прогнозирования спроса и выручки. Их основное преимущество заключается в относительной простоте и интерпретируемости (хотя и используются немало формул для прогнозирования, а также рекурсии), а также в способности учитывать тренд и сезонность, что делает их удобными для первичного анализа и базовых прогнозов.+

Модели класса ARIMA и SARIMA считаются стандартом в эконометрике и анализе временных рядов. Они позволяют учитывать автокорреляцию, сезонные компоненты. Несмотря на необходимость акку-

ратного подбора параметров и проверки стационарности, данные модели активно применяются на практике для прогнозирования цен, объёмов продаж и финансовых показателей. Во многих прикладных работах именно SARIMA показывает наилучшие результаты для месячных данных с устойчивой сезонностью, что подтверждает целесообразность её использования в рамках данного проекта.

Помимо классических временных рядов, в последние годы всё чаще используются регрессионные модели и методы машинного обучения. Для прогнозирования цен на сырьевые и продовольственные товары применяются линейные и регуляризованные модели, такие как:

Линейная регрессия

<

- Lasso [3] регуляризация
- Ridge регуляризация

> Ridge и Lasso. Первая подбирает более удачные веса (метрики качества более угодны) благодаря ограничению по норме вектора весов, а вторая подбирает веса, одни из которых могут быть нулями, что показывает нагляднее степень влияния признаков на целевую переменную, благодаря ограничению по сумме весов. Их ключевое преимущество состоит в возможности учитывать большое количество факторов и автоматически отбирать наиболее значимые признаки. Feature engineering в подобных задачах играет важную роль и включает создание лаговых признаков, календарных переменных и показателей, отражающих внешнюю экономическую среду.

Отдельное направление исследований связано с оценкой влияния управлеченческих и бизнес-решений. В экономической аналитике для этого широко применяются методы статистического вывода. Параметрические и непараметрические тесты, такие как z-тест и U-тест Манна–Уитни [2], позволяют сравнивать распределения показателей до и после

внедрения изменений и делать выводы о наличии или отсутствии статистически значимого эффекта. Данные подходы активно используются в А/В тестировании и анализе эффективности бизнес-стратегий.

С точки зрения сложности, задачи прогнозирования рынка подсолнечного масла нельзя считать тривиальными. Данные характеризуются шумом, неполной наблюдаемостью и влиянием факторов, которые не всегда можно напрямую включить в модель. Это требует сравнения различных подходов, аккуратной настройки моделей. Одновременно с этим наличие исторических данных и хорошо изученных методов делает данные задачи выполнимыми и практически значимыми.

Таким образом, проведённый обзор показывает, что:

Выводы

- Подобные задачи являются выполнимыми и могут быть успешно решены при наличии данных и корректного выбора моделей.
- Задачи прогнозирования и оценки результатов решений достаточно сложны за счёт сезонности, шума в данных и влияния внешних факторов, что делает их интересными и уместными для выполнения выпускной квалификационной работы.

3. Решение

В данном разделе описывается практическая часть проекта, включающая анализ данных, построение моделей прогнозирования и интерпретацию полученных результатов. Основной акцент сделан на последовательном применении методов анализа данных — от первичного исследования до оценки бизнес-эффектов.

3.1. Анализ данных и EDA

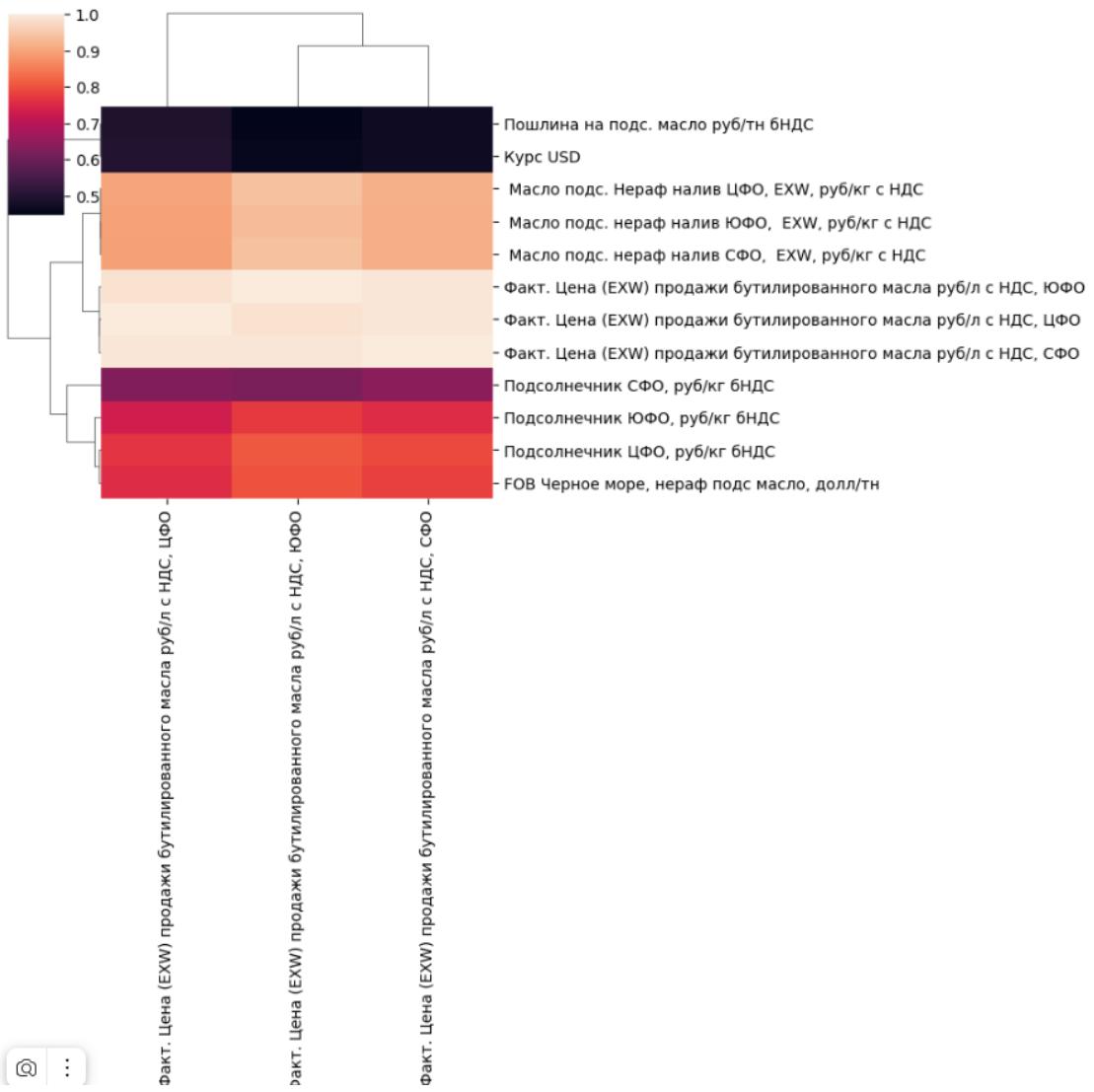
На первом этапе был выполнен разведочный анализ данных (EDA). Были изучены временные ряды цен и выручки, проверено наличие пропусков, выбросов и аномальных значений, характерных для рынка подсолнечного масла.

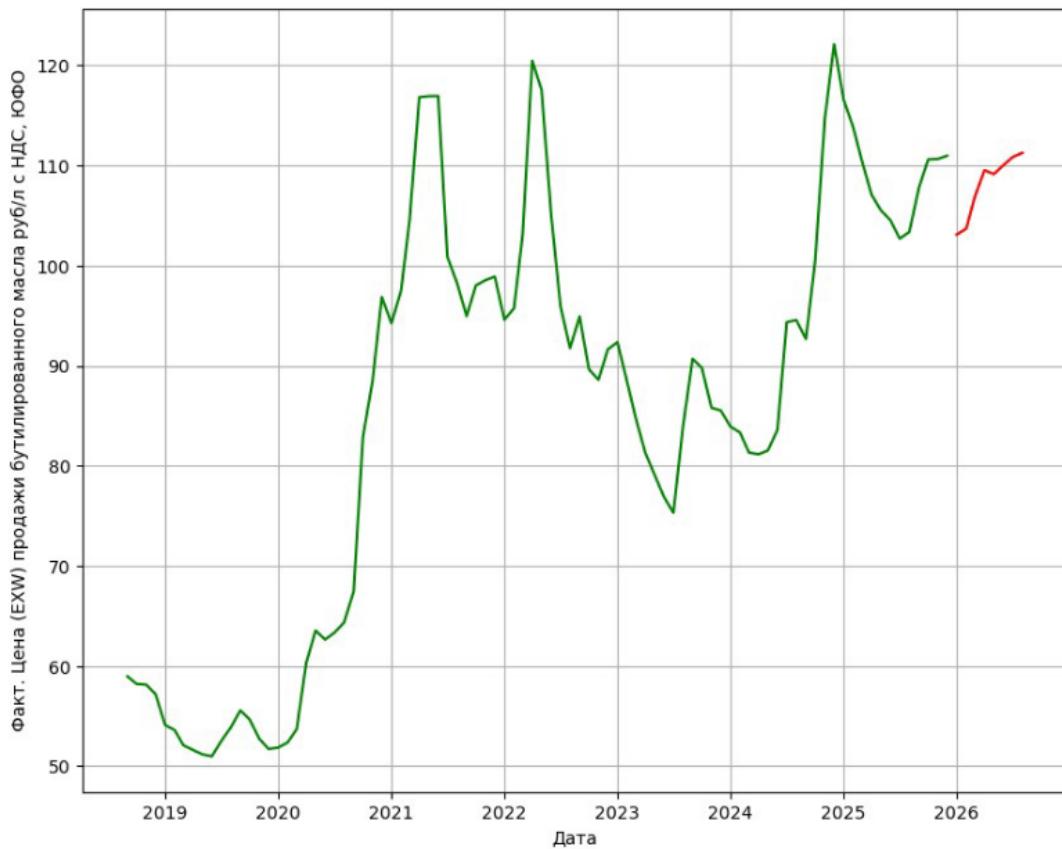
В рамках EDA также анализировались распределения показателей и их изменения во времени. Это позволило сделать предварительные выводы о структуре данных и выбрать подходящие методы дальнейшего моделирования.

3.2. Прогноз цен на масло

Для прогнозирования цен была использована Lasso-регрессия. Перед обучением модели был проведён feature engineering: добавлены лаговые признаки, календарные переменные и показатели, отражающие динамику рынка.

Выбор Lasso был обусловлен необходимостью отбора наиболее значимых факторов и снижения риска переобучения. Модель показала стабильные результаты на тестовом периоде, а анализ коэффициентов позволил интерпретировать влияние отдельных признаков на формирование цены.

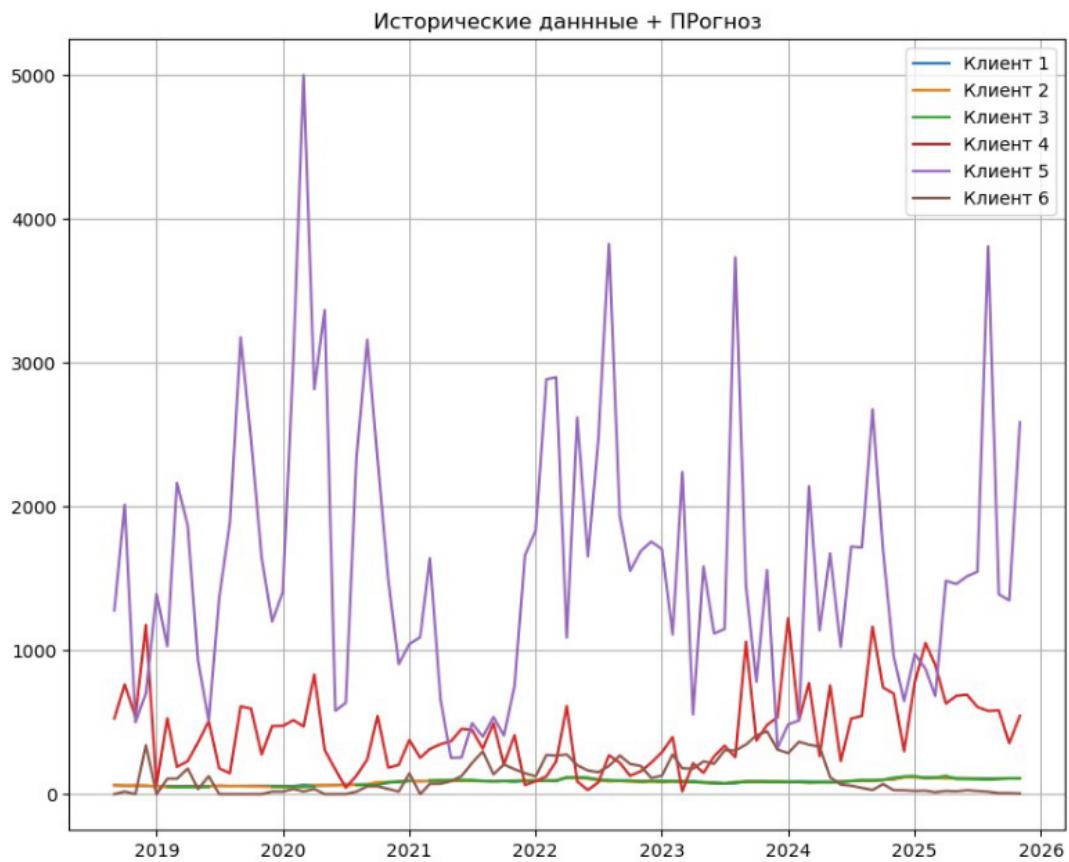




Пример для ”Факт. Цена (EXW) продажи бутилированного масла руб/л с НДС, ЦФО”

3.3. Прогноз выручки клиентов (временные ряды)

Для прогнозирования выручки клиентов были реализованы и сравниены несколько моделей временных рядов: Simple Exponential Smoothing, Holt–Winters, SARIMA и ансамбль моделей. Сравнение проводилось в формате А/В тестирования с использованием стандартных метрик качества.



Исторические данные для клиентов



ПРогноз для клиента №4 на модели Holt-Winters

Результаты показали, что модель SARIMA обеспечивает наилучшее качество прогноза для рассматриваемых данных, что связано с её способностью учитывать как сезонность, так и автокорреляцию временного ряда. Поэтому данная модель бы-

ла выбрана как наиболее выгодная для практического приме-

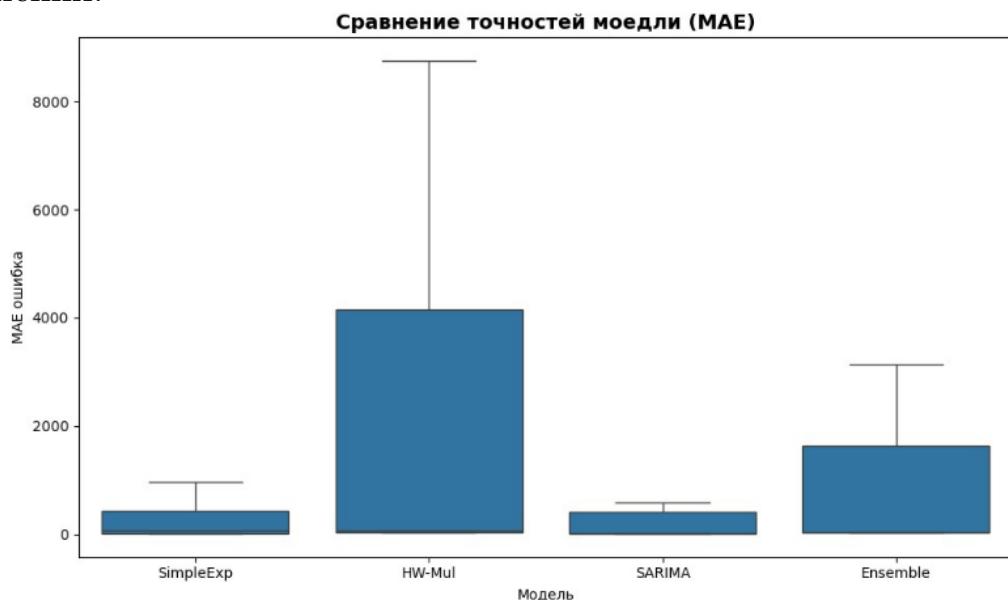
СРЕДНИЕ МЕТРИКИ ПО СТРАТЕГИЯМ:

Strategy	MAE	MAPE	R2	P-value	Significant
Ensemble	930.3569	0.9409	-8.9055	0.0001	0
HW-Mul	2461.2087	2.0405	-63.0759	0.0000	0
SARIMA	212.1905	0.3974	-0.7516	0.1353	2
SimpleExp	293.3763	1.1364	-1.5657	0.0002	0

ЛУЧШИЕ МОДЕЛИ ДЛЯ КАЖДОГО SKU (по MAE):

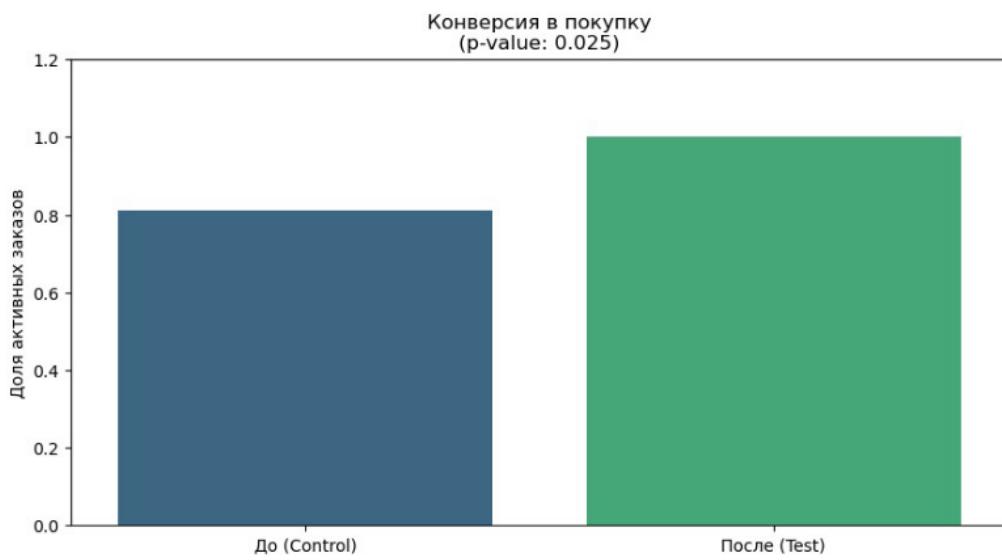
- Итог Сумма по полю Вес реализации, т: SARIMA (MAE=573.91)
Итог Сумма по полю чистая цена руб/л с НДС: SARIMA (MAE=17.32)
Клиент 1: SARIMA (MAE=16.22)
Клиент 2: SARIMA (MAE=17.36)
Клиент 3: SARIMA (MAE=17.12)
Клиент 4: SARIMA (MAE=257.84)
Клиент 5: SARIMA (MAE=585.56)
Клиент 6: SimpleExp (MAE=130.35)

нения.



3.4. Анализ влияния бизнес-решений

Дополнительно в рамках проекта был проведён анализ влияния бизнес-решений на выручку. В частности, рассматривался эффект увеличения посевных площадей.



Для оценки статистической значимости изменений использовались U-тест Манна–Уитни и z-тест. Применение данных методов позволило сделать вывод о наличии положительного и статистически значимого эффекта на выручку.

3.5. Визуализация результатов

В ходе работы были построены графики динамики цен и выручки, прогнозные значения моделей, а также визуализации результатов статистических тестов. Визуализация использовалась не только для представления итоговых результатов, но и как инструмент анализа, позволяющий лучше понять поведение данных и качество построенных моделей.

4. Заключение

В ходе выполнения проекта были реализованы методы анализа данных, машинного обучения и анализа временных рядов для решения задачи прогнозирования показателей рынка подсолнечного масла. Проведён EDA, построены модели прогнозирования цен и выручки, выполнено сравнение моделей и анализ эффекта бизнес-решений.

Полученные результаты показывают, что подобные задачи являются решаемыми и практически значимыми, но при этом требуют аккуратного подхода к данным и выбору методов. За счёт сочетания прогнозирования, статистического анализа и интерпретации результатов проект является достаточно сложным и интересным для использования в качестве ВКР

Список литературы

- [1] Hyndman Rob J. Holt-Winters' seasonal method. — URL: <https://otexts.com/fpp2/holt-winters.html>.
- [2] Hyndman Rob J. U-test критерий АБ. — URL: https://ru.wikipedia.org/wiki/U-критерий_Манна_-_Уитни.
- [3] Kirenz Jabn. Lasso regression. — URL: <https://kirenz.github.io/regression/docs/lasso.html>.
- [4] Sarima' seasonal method. — 08-12-2025. — URL: <https://www.geeksforgeeks.org/machine-learning/sarima-seasonal-autoregressive-integrated-moving-average/>.

[1] [4] [3] [2]