

Geekbrains

# **Анализ продаж молочной продукции и оптимизация управления запасами на молочных фермах**

Программа: Цифровые процессы  
Аналитик больших данных  
Батыль Михаил Сергеевич

Новосибирск  
2024 г.

### ***Цель проекта:***

Изучить особенности анализа данных по продажам молочной продукции и оптимизация складских запасов. Способы формирования интерактивных дашбордов в целях автоматизации процессов на производстве.

### ***Задачи:***

Изучить литературу, материалы и статьи касающиеся темы анализа продаж, общая эффективность производства, маркетинговые исследования, бережливое производство, анализ производства.

Рассмотреть несколько вариантов имеющихся платформ для создания дашбордов, обработки данных и выбрать программное обеспечение, соответствующее задачи автоматизации.

Произвести анализ эффективности молочных ферм с учетом местоположения, площади земель и поголовья коров.

Сформировать модель продаж и распределения различных молочных продуктов по различным брендам и регионам.

Оптимизировать управление запасами путем отслеживания количества запасов, минимальных пороговых значений и количества повторных заказов.

Разработать дашборд, отражающий результаты анализа данных по продажам и запасам, позволяющий вручную оптимизировать рекомендуемые объемы и ассортимент заказа производителя.

### ***Инструменты:***

GitHub, VScode, Power BI (или аналог), Power Query, Python, Excel, draw io.

## Оглавление

Введение .....	4
Глава 1. Основы автоматизации процесса, связанных с аналитикой данных .....	6
1.1. Что такое автоматизация бизнес-процесса.....	6
1.2. Какой бизнес нужно автоматизировать .....	9
1.3. Какие требования существуют к автоматизации процесса .....	10
1.4. Способы сбора и анализа данных. ....	12
1.5. Что такое дашборд. Как он может повысить эффективность аналитических процессов.....	13
Глава 2. Подготовка к разработке дашборда .....	16
2.1. Сбор требований к разработке дашборда.....	16
2.2. Сбор и обработка данных для анализа бизнес-процесса продажи молочных товаров. ....	19
2.3. Сформировать список отслеживаемых метрик, необходимых для автоматизации, разработки дашборда.....	22
2.4. Составить предварительную визуализацию дашборда.....	22
2.5. Создать модуль обработки и анализа данных, а также протестировать его на данных компаний по производству молочной продукции .....	24
2.6. Произвести анализ предпочтений и покупательского поведения клиентов в зависимости от местоположения и каналов продаж .....	52
2.7. Изучить влияние условий хранения и сроков годности на качество и доступность молочной продукции .....	65
2.8. Провести ABC анализ продаж.....	70
Глава 3. Описание процесса создания системы.....	74
3.1. Разработка дашборда.....	74
3.2. Подведение итогов. Разработка предложений по улучшению процесса продаж и оптимизации производства. ....	78
Список литературы:.....	80
Приложения:.....	81

## **Введение**

На рынке прогнозируется уменьшение количества поставщиков молочной продукции, падение их производительности в связи со сложной экономической ситуацией, сезонным снижением цен в летний период и неутешительными для поставщиков прогнозами падения цен на сырье осенью. Небольшие поставщики, вероятно, не смогут обеспечивать поставку сырья на прежних условиях. Как сохранить прежних поставщиков, а в случае с новыми, наладить сотрудничество и предложить оптимальные условия работы и оплаты за сырье?

Чтобы гибко реагировать на изменение покупательского спроса, предприятиям важно отслеживать потребности рынка и периодически менять ассортимент. В сложных экономических условиях это становится как никогда актуальным, поскольку повышается риск того, что выпущенная продукция не будет продаваться, что приведет к затовариванию складов и потере денег — ведь молочная продукция в основном имеет короткий срок годности и, следовательно, необходимо своевременно ее реализовывать.

Проанализировать покупательский спрос и вовремя пересмотреть стратегию продаж поможет аналитика с классификацией по продажам продукции как по наименованиям (видам), так и по регионам поставки и сегментам. Руководителям отделов продаж (сбыта) и менеджерам по продажам в настоящее время важно анализировать потребности покупателей именно своего региона, поскольку на полках магазина, чаще всего и представлена продукция местных производителей.

Таким образом, используя аналитические отчеты, руководители предприятия могут получать необходимую для анализа информацию и своевременно принимать решение о выпуске наиболее востребованной и приносящей прибыль продукции.

На фермах и молокоперерабатывающих предприятиях выпуск продукции осуществляется под заказы клиентов. Как правило, это постоянные клиенты и спрос у них более-менее стабильный, поэтому производству понятно: что производить и в каком количестве. Но, даже у постоянных клиентов могут изменяться планы по количеству и ассортименту закупаемой продукции и эти колебания необходимо отслеживать, чтобы своевременно скорректировать планы производства. Для решения этой задачи начальник производства и сотрудники планового отдела могут использовать инструменты планирования производства и расчета потребностей в материалах с учетом отраслевой специфики. Программа поможет рассчитать, когда и сколько сырья и материалов (в зависимости от их качества) потребуется на выпуск продукции в соответствии с планами производства, а также запланировать выпуск полуфабрикатов.

Немаловажная задача — это обеспечение полной загрузки производственных мощностей. Ведь потери от простоев оборудования для заводов могут быть критичными.

В период, когда предприятия стремятся оптимизировать расходы, становится буквально обязательным не только своевременно отгружать продукцию, но и выстраивать маршруты автомобилей, развозящих продукцию клиентам так, чтобы максимально сократить затраты на доставку.

**Цифровизация всех бизнес-процессов — неизбежное будущее ферм молокоперерабатывающих предприятий, ферм, а своевременная, корректная аналитика поступающих данных, снижает риски в части принятия неверных действий, шагов оптимизации бизнеса.**

**Цифровая трансформация** – глубокая реорганизация бизнес-процессов с широким применением цифровых инструментов для их исполнения, которая приводит к существенному улучшению их характеристик (сокращение времени выполнения, исчезновению целых групп подпроцессов, сокращению

ресурсов, затрачиваемых на выполнение процессов) и/или появлению принципиально новых их качеств и свойств.

Результаты автоматизации и цифровизации в виде внутренних автоматизированных и прошедших реинжиниринг процессов, внедренных автоматизированных систем и надежных данных, во многих случаях служат основой для **цифровой трансформации**.

## **Глава 1. Основы автоматизации процесса, связанных с аналитикой данных**

### **1.1. Что такое автоматизация бизнес-процесса**

Помните старый мультфильм про Нехочуху? Ленивый мальчишка попал в страну великого Нехочухи, где большинство действий выполняли роботы. Надо сказать, мультипликаторы изобразили безрадостную картину автоматизации: человек внутри роботизированных процессов распадается как личность и теряет навыки социальной адаптации. Это идёт явно вразрез с любимым нами «Электроником», который как раз проповедует ценность соединения автоматизации и человеческого начала. А вот современный бизнес об автоматизации думает, как о вынужденной, как о неоправданной трате, как о единственной волшебной таблетке от всех проблем, и в целом относится к ней с подозрением. А это ошибочный подход. Управленческая концепция BPMN рассматривает бизнес-процессы как важные ресурсы предприятия, и предполагает управление ими как одну из ключевых организационных систем. Ну а автоматизация — это как раз процесс внедрения и эксплуатации программных средств для оптимизации и ускорения протекания бизнес-процессов. Автоматизация упорядочивает, налаживает оперативную работу и управление задачами. Все сотрудники (причём совершенно не важно,

удалённые или офисные) работают в едином информационном пространстве АСУ (PM, ERP, CRM...), координируют свои действия, получают и делегируют задачи, видят свои задания внутри проектов и бизнес-процессов, распределяют свою нагрузку во времени. Руководителю больше не нужно вызывать каждого к себе и разжёвывать задачи — достаточно поставить их в системе и проконтролировать выполнение. Вся оперативная работа ведётся в единой системе, и любой сотрудник может получить доступ к данным в рамках своих компетенций, а также запросить доступ или выгрузку данных, если в этом есть необходимость. Вне автоматизации всё это представляет собой сплошную стихийную коммуникацию, в которой несложно забыть, ошибиться или просто «незаметно» не выполнить задачу.

Человеческий фактор — это, пожалуй, самое первое, с чем борется автоматизация в компании. Вот примерный минимум корпоративных проблем, с которыми сталкивается малый и средний бизнес: упущенные возможности контакта с клиентами, забытые и нигде не зафиксированные данные, растянутые во времени элементарные задачи, офисная и удалённая прокрастинация, работа «налево» по профилю компании, «свои» приватные клиенты менеджеров, хищение коммерческой информации и проч. Когда в руках, а точнее на серверах бизнеса появляется программное обеспечение для автоматизации бизнес-процессов всех уровней (управляющие, операционные, поддерживающие), большинство этих проблем легко и относительно безболезненно решаются.

**Автоматизация — это не лекарство от всех болезней бизнеса и не волшебная кнопка «Сделать все правильно».**

На основе данных по аудиту ферм специалистами сервиса Dairy Production Analytics партнёрами в разных странах, видим следующую проблематику в данном вопросе:

- Полное или частичное отсутствие в системах автоматизации достоверных данных для контроля производственных процессов

- Отсутствие квалифицированного персонала, который вносит неавтоматизированные данные в ручном режиме в системы автоматизации
- Отсутствие регламентов по внесению этих данных
- Отсутствие единого формата внесения данных, при котором одним оператором могут вноситься одни значения, другим – совершенно другие, но подразумевающие в итоге одинаковую информацию (как пример, один оператор в систему управления стадом вносит заболевание животного «Мастит», другой «Маст», «Маст.», «М» и т.п., по факту подразумевая мастит, но для системы автоматизации это, тем не менее, разные значения показателя)
- Отсутствие анализа этих данных на уровне систем автоматизации (основной аргумент при этом – там недостоверные данные, нам нечего там анализировать!)
- И самое главное – отсутствие понимания, для чего необходимы все эти данные.

В свою очередь, отсутствие достоверных данных порождает отсутствие контроля за ключевыми показателями на ферме. Приведем набор таких показателей:

- Количество поголовья (тёлки, нетели, бычки, коровы первой, второй, третьей и более лактаций, дойное поголовье, фуражное поголовье)
- Объёмы и показатели производства молока (надой на дойную корову, на фуражную корову, в целом по стаду за день, месяц, год, относительные показатели по надою и так далее)
- Показатели по выбытию животных, заболеваниям и их причинах
- Лактационные кривые по разным лактациям и по разным группам дойного поголовья
- Данные по оплодотворяемости коров и тёлочек
- Данные по отёлам и изменению поголовья стада после отёлов
- Данные по кормлению, стоимости и структуре кормов
- И так далее...



Но есть и конкретные критические показатели, контроль за которыми особенно важен:

- Ежегодный процент выбраковки (количество выбывших коров, делённое на среднегодовое поголовье фуражных коров)
- Средний день доения
- Средний сервис-период
- Средний межотёльный интервал
- Процент стельности по стаду (количество стельных коров, делённое на количество голов в стаде)
- Средний расход доз семени на одну голову
- Средний выход тёлочек
- PR-коэффициент
- Процент выбывших животных из новотельной группы
- Возраст выбывших животных до 60-го дня доения
- Процент животных в группе «Мастит»
- Количество повторных заболеваний маститом
- Процент животных в группе «Карантин»
- Процент хромоты животных

И это ещё не полный список ключевых показателей. Очевидно, что без должного уровня автоматизации и качества данных, мы не можем контролировать эти показатели, особенно в рамках ферм с большим (более 500 голов) поголовьем.

Все вышеперечисленные метрики, а также управление складскими запасами, напрямую зависят от службы сбыта готовой продукции. Проводя полный и своевременный анализ продаж (воронки продаж), гарантируем корректное управление производственной системой.

## **1.2.Какой бизнес нужно автоматизировать**

Любой бизнес необходимо автоматизировать. У компаний существуют разные метрики, бизнес-процессы, но есть группа целей автоматизации, которая универсальна для большинства организаций.

- **Экономия** — одна из ключевых целей автоматизации. Вложившись в программное обеспечение, компании используют его либо для увеличения производительности труда, либо для сокращения затрат. Таким образом, компания как минимум экономит время и деньги на изменении структуры сделок и продаж, увеличивая при этом оборот и прибыль. Чтобы автоматизация оказалась откровенно убыточной, нужно просто забросить купленную программу и забыть о потраченных деньгах.
- **Упорядочить и ускорить бизнес-процессы** — ключевая задача автоматизации. Сокращение времени на каждую операцию, определение чётких этапов, ответственных и сроков делает процесс эффективным, прозрачным и отлаженным. Чёткое распределение задач между сотрудниками минимизирует человеческий фактор и сокращает связанные с ним риски.
- **Увеличение прозрачности бизнеса в целом** — важная цель. Автоматизация помогает сотрудникам работать с клиентами и друг с другом в единой информационной системе и не конфликтовать «на перedelе» клиентов и следовать индивидуальным KPI; клиентам — получать своевременное и качественное обслуживание; руководителю — видеть реальные результаты работы сотрудников. Такая прозрачность не только формирует позиционирование компании, но и ведёт к росту выручки благодаря исключению конфликтов и лояльности клиентов, а также стандартизации процессов (стандартный, фиксированный бизнес-процесс — признак надёжной компании).

### 1.3. Какие требования существуют к автоматизации процесса

Прежде всего нужно выбрать правильные процессы для автоматизации:

- Большие процессы — ветвящиеся, сложные, затрагивающие несколько сотрудников;
- Повторяющиеся с каким-либо периодом (запуск акций, отгрузки, стандартные продажи и т.д.);
- Влияющие на все остальные процессы и активности в организации (обучение новых сотрудников, выпуск релиза продукта);
- Требовательные к срокам исполнения (запуск акций и промо, поддержка клиентов, некоторые внутренние процессы управления инфраструктурой и т.д.);
- Рутинные процессы — которые отнимают много времени, часто повторяются и при этом не требуют особо интеллектуальных действий (закупки, отгрузки, формирование пакетов документов и проч.)

Обратимся к теории бережливого производства. Необходимо смотреть на приоритет автоматизации через основные потери. Если изобразить потери на диаграмме, то сможем увидеть, что в любом процессе ценность занимает самую малую часть, все остальное это разного вида потери, которые бывают первого и второго рода (график 1).

Потери 1 рода — это действия, не создающие ценность, но без которых невозможно обойтись. Например транспортировка, оформление документов. Их невозможно удалить из процесса, но необходимо стремиться сокращать.

А вот потери 2 рода — это действия, не создающие ценности вообще, и их можно и нужно исключать из процесса полностью. Например ожидание, запасы, брак и т.д.

Потери можно найти в любом процессе, будь то производство, оказание услуг различного характера или здравоохранение. Для того что бы устранить данные потери, необходимо уметь их распознавать и знать способы борьбы с ними.

В классической теории бережливого производства были выделены 7 видов потерь:

- Перепроизводство
- Ожидание
- Запасы
- Излишняя транспортировка
- Излишнее перемещение людей
- Брак
- Излишняя обработка

В дальнейшем, совершенствуя теорию, исследователи дополнили первоначальный список еще одним видом потерь:

- Неиспользованный человеческий потенциал

#### **1.4.Способы сбора и анализа данных.**

Есть несколько методов сбора, необходимых для анализа данных. Они перечислены в порядке увеличения стоимости. Выгрузка из учетных систем. Обычно в учетных системах имеются механизмы экспорта данных или существует API для доступа к ним. Поэтому извлечение нужной информации из систем учета чаще всего относительно несложная операция. Получение сведений из косвенных данных. О многих показателях можно судить по косвенным признакам. Например, можно оценить реальное финансовое положение жителей определенного региона. В большинстве случаев имеется несколько товаров, предназначенных для выполнения одной и той же функции, но отличающихся по цене: товары для бедных, среднеобеспеченных и состоятельных. При наличии данных о продажах по регионам, можно проанализировать пропорции, в которых продаются товары для каждой из категорий клиентов: чем больше доля дорогой продукции, тем более состоятельны в среднем жители данного региона. Использование открытых источников. Большое количество данных присутствует в открытых источниках, таких как статистические сборники,

отчеты корпораций, опубликованные результаты маркетинговых исследований и прочее. Покупка сведений у соцсетей, мобильных операторов и дата-брокеров. На рынке работает много компаний, которые занимаются сбором и продажей данных. Они предоставляют посредством API систематизированную информацию различного плана: кредитоспособность, клиентские предпочтения, цены на продукцию, геолокация и т.д. Проведение собственных маркетинговых исследований и аналогичных мероприятий по сбору данных. Это может быть достаточно дорогостоящим мероприятием, но, в любом случае, такой вариант сбора данных возможен. Ввод данных «вручную», когда данные вводятся по различного рода экспертным оценкам сотрудниками организации. Данный метод достаточно трудоемкий и требует постоянного выделения ресурсов для обеспечения актуальности сведений. Очевидно, что если эксперт указал некоторый фактор как важный, то исключать его неразумно. Мы рискуем провести анализ, ориентируясь на второстепенные малозначащие факторы, и, следовательно, получить модель, которая будет давать плохие и нестабильные результаты. Такая модель не представляет практической ценности.

Собранные данные нужно преобразовать к единому формату. Идеальный случай — загрузка в базу или витрину данных. Но можно использовать и более легковесные форматы, например, Excel или текстовый файл с разделителями. Данные обязательно необходимо стандартизировать, т.е. одна и та же информация везде должна описываться одинаково. Обычно проблемы с единообразным представлением возникают при сборе информации из разнородных источников.

### **1.5. Что такое дашборд. Как он может повысить эффективность аналитических процессов**

Термин business intelligence впервые появился в бизнес-литературе 1865 года. Его упомянул Ричард Миллар Девенс в своей книге «Энциклопедия бизнес-анекдотов». Он писал, что банкир Сэр Генри Фюрненс преуспел благодаря

своему Business Intelligence — пониманию политики, нестабильности рынков и рыночных взаимоотношений.

Такое определение термина business intelligence ближе к бизнес-анализу, чем к техническому описанию и так как в 1865 году технологии были неразвиты, можно сделать вывод, что истоки термина уходят к бизнес-аналитике.

В 1958 году сотрудник IBM по имени Ганс Питер Луи написал научную статью A Business Intelligence System, полностью посвящённую business intelligence. Её можно считать точкой зарождения BI в его современном понимании, так как в статье речь шла о работе с данными, а сам термин BI определялся как «автоматизированная система, разработанная с целью распределения информации различных направлений деятельности любой индустриальной, научной или государственной организацией».

BI или Business Intelligence, если объяснять простыми словами, то BI — это Excel + Power Point в одном инструменте, в котором технических возможностей больше, чем в самих Excel и Power Point, например в BI можно настроить автопреобразование и автообновление данных, а также делать не статичные презентации как в Power Point, а интерактивные дашборды, в которых можно менять контекст данных в онлайн режиме.

Если определять, что такое BI более научно, то BI - это автоматизация аналитических процессов для презентации данных в красивой форме. В данном определении акцент делается на 2 слова: автоматизация и презентация. Автоматизация помогает высвободить время аналитиков от рутинных задач, на которые аналитики тратят до 80% своего времени, а презентационный слой помогает делать интерактивные дашборды, которые в онлайн режиме показывают данные.

Если говорить о миссии BI, какую ценность BI несет для компаний и предприятий, миссия заключается в том, чтобы обеспечить организацию оперативными данными для помощи в принятии управленческих решение. BI делает решения более очевидными, понятными, легкими для объяснения, а также

ускоряет их принятие, что крайне важно в условиях быстроменяющейся реальности. Например, без внедренной ВІ системы вы чаще всего будете решать проблемы по мере их обнаружения, а при помощи ВІ вы можете либо реагировать до их появления, либо быстрее принимать необходимые действия. Допустим, делаем аналитику прогнозирования амортизации оборудования - при помощи ВІ возможно настроить отправку уведомления сотруднику, когда необходимо начать заниматься обслуживанием оборудование, чтобы оно не вышло из строя.

Теперь давайте разберемся с тем, какие задачи ВІ решает, а для каких задач есть более профильные инструменты.

Профильные задачи для ВІ инструментов:

1. Автоматизация аналитических процессов. Очень часто внутри компаний одна и та же аналитика нужна к одним и тем же встречам. Раньше отчёты для них каждый раз готовил аналитик. Сейчас можно один раз настроить автоматическую аналитику при помощи ВІ-систем. Кейс: аналитик две недели готовит отчёт к встрече. После внедрения ВІ его время освобождается. В течение двух недель аналитик забирает данные из одной и той же системы, производит одни и те же преобразования (формат столбцов, расчеты, сводные таблицы) и производит много рутинной работы, ВІ позволяет 1 раз настроить этот процесс и высвободить время для более креативных задач.
2. Создание автоматических презентаций. Часто на встречах могут показывать презентации в PowerPoint с одинаковым набором графиков, таблиц и показателей. Меняются только выводы, основанные на данных. Создание таких презентаций тоже можно автоматизировать в ВІ. Кейс: на постоянной встрече показывают одну и ту же
3. Alarm-отчёты. Многим руководителям не хочется погружаться в аналитику детально, либо у них нет на это времени. Часто они просят сообщать о чём-либо, когда возникают проблемы. Отчёты для них можно создать, когда известны границы для какого-то значения и подсвечивать, когда показатели выходят за

пределы нормы. Кейс: руководителю приходит уведомление о том, что что-то идёт не так. Допустим у нас есть необходимость отслеживать план факт продаж ритейл магазина раз в час и если факт ниже плана на 20%, то система отправляет профильному сотруднику уведомление о том, что что-то идет не так и необходимо идти разбираться. Таким образом компания реагирует на проблему сразу, а не через какое-то время и если проблему можно решить быстро, то количество упущенной прибыли будет минимизировано.

## **Глава 2. Подготовка к разработке дашборда**

### **2.1. Сбор требований к разработке дашборда.**

Можно выделить два основных типа дашбордов:

**Операционный** дашборд нацелен на предоставление пользователям критически важной информации для оперативного принятия решений. Актуальность информации в таких дашбордах должна стремиться к online. Это наборы ключевых метрик и информации об отклонениях, статусах и т.п. Обычно пользователями таких дашбордов являются менеджеры нижнего звена ответственные за текущую деятельность в организации (управление грузоперевозками, складами, закупками, продажами). Такой дашборд может содержать большое количество элементов, охватывающих все возможные метрики, которые пользователь хочет наблюдать.

**Аналитический** дашборд не несет таких требований ко времени как операционный и призван дать пользователю информацию о трендах и прогнозах, исследование взаимосвязи различных факторов. Обычно их используют аналитики и руководители организации для принятия стратегических решений, основанных на данных. В отличии от операционного дашборда, аналитический



не стоит перегружать большим количеством элементов, а сконцентрироваться на основных показателях нужных для принятия решений и тем более лаконичным должен быть дашборд чем выше уровень руководителя, который работает с ним.

После того как определены пользователи дашборда, необходимо определить на какие вопросы пользователя должен ответить дашборд. Для этого требуется изучить логику принятия решений, по которой пользователь ищет ответ на свой вопрос. Должны быть определены метрики, на которые опирается пользователь в решении задачи и их взаимосвязь. Исходя из этого можно начинать сбор данных и формирование витрины, содержащей всю необходимую информацию для ответа на поставленные вопросы. Также следует получить требования по доступу к данным в зависимости от пользователя. Например, это могут быть ограничения по региону или департаменту/отделу пользователь, которого просматривает дашборд.

Ни один проект, связанный с данными, не обходится без стадии Data Profiling (Профайлинг). Что это такое?

Data Profiling (Профайлинг) — это процесс исследования данных, понимания структуры, взаимосвязей, контекста и возможного применения этих данных в дальнейшем анализе.

Необходимо понять значение данных с точки зрения бизнеса и текущих процессов, а также описать как эти данные могут быть использованы в анализе.

Профайлинг данных помогает:

В проектах связанных с исследованием систем-источников данных. В процессе data profiling (профайлинг) могут быть подчеркнуты риски низкого качества данных, что повлияет на дальнейшее использование данных в BI инструментах и Data-продуктах.

При миграции данных из одной системы в другую. В процессе переезда важно понимать структуру, взаимосвязи и бизнес-контекст данных, также важно

проверить качество данных и при необходимости повысить его с помощью интеграционных и ETL/ELT процессов.

В процессе построение BI или Data продукта, которым пользуются конечные бизнес-пользователи. Data Profiling (профайлинг) позволит подчеркнуть все риски качества данных, а также увеличит шансы применения данных в нужном ключе.

**Важное правило:** Data Profiling (профайлинг) необходимо планировать как можно раньше (на первых шагах) работы над проектом и продуктом. Это упростит понимание данных, а также даст возможность управлять поставленными требованиями с большей гибкостью.

Data Profiling должен идти сразу после анализа и формулирования бизнес-требований. Вы должны понять, какие данные у вас есть и какие требования с помощью этих данных вы можете покрыть

Важно оценить качество данных до этапа разработки. Может случиться так, что качество одного источника данных очень низкое и придется искать другой источник. Это лучше делать в начале, чем потом мигрировать все решение на другой источник

Необходимо проверять данные систематически. Нельзя слепо верить документации или описанной модели данных, они могут быть устаревшими или не совпадать с реальной структурой данных

Правильные вопросы при профайлинге данных:

Какие таблицы у нас есть и какие атрибуты они в себя включают?

Какие атрибуты включены в первичный ключ? Какой тип ключа используется (простой, составной, суррогатный)?

Какой тип историчности и обновления данных используется в таблице?

Как часто обновляются данные в таблице?

Сколько пустых (blank или null) значений содержит каждый атрибут таблицы? Какой процент от общего количество строк содержит пустые значения?

Сколько уникальных значений содержит каждый атрибут таблицы?

Какие самые популярные значения атрибутов? Если числовые значения, то какое минимальное, максимальное и среднее значение?

Какие корреляции есть между значениями атрибутов? Есть ли в этом зависимость от описываемого бизнес-процесса?

Как итогом после успешного профайлинга данных:

- Понимание данных, атрибутов, взаимосвязей
- Понимание бизнес-области применения данных
- Описанное качество данных и возможное улучшение этого качества при помощи дальнейшей обработки данных
- Сокращение процесса разработки путем проведения профайлинга и выявления всех зависимостей перед самой разработкой
- Большее покрытие бизнес-требований

После того как все требования собраны, необходимо подготовить макет дашборда, в виде схемы расположения всех его элементов и используемый тип визуализаций и согласовать его с Заказчиком. Макет может быть подготовлен как в BI инструменте, так и с использованием графических редакторов, таблиц Excel или любым другим удобным способом.

## **2.2. Сбор и обработка данных для анализа бизнес-процесса продажи молочных товаров.**

Рост потребления молока и молочных продуктов является в настоящее время общемировой тенденцией. Для обеспечения устойчивого роста

производства молока необходимо повысить эффективность деятельности молочных ферм. Это требует, прежде всего, формализации и глубокого анализа протекающих на фермах бизнес-процессов. Производство и служба сбыта напрямую зависят друг от друга. Чем ближе находишься к пересечению двух процессов сбыта и производства, тем больше экономический эффект у компании, гармоничнее работают бизнес-процессы.

Рассмотрим набор данных (dataset) о продаже молочных товаров в конкретных штатах и союзных территориях Индии, комплексный набор данных о молочных фермах, продуктах, продажах и отслеживании запасов на складе. На схеме рис.1. представлена структура (точка) пересечения отдела продаж и производства. Жизненный цикл формирования коммерческого предложения на заявку потребителя.

Набор данных о продажах молочных продуктов сформирован подробный и информативный, касается молочных ферм, молочных продуктов, молокоперерабатывающих предприятий, продаж и управления запасами. Этот набор данных включает в себя широкий спектр информации, включая местоположение фермы, земельную площадь, поголовье коров, размер фермы, даты производства, сведения о продукте, информацию о бренде, количества, цены, срок годности, условия хранения, сроки годности, информацию о продажах, местонахождение клиентов и т. д. каналы продаж, объемы запасов, пороговые значения запасов и объемы повторного заказа.

Функции (Реквизиты):

- Местоположение: Географическое положение молочной фермы.
- Общая площадь земли (акры): общая площадь земли, занимаемая молочной фермой.
- Количество коров: количество коров на молочной ферме.
- Размер фермы: Размер молочной фермы (в кв. км).
- Дата: дата записи данных.

- Идентификатор продукта: уникальный идентификатор каждого молочного продукта.
- Название продукта: Название молочного продукта.
- Бренд: бренд, связанный с молочным продуктом.
- Количество (л/кг): Количество доступного молочного продукта.
- Цена за единицу: Цена за единицу молочного продукта.
- Общая стоимость: общая стоимость доступного количества молочного продукта.
- Срок годности (дни): Срок годности молочного продукта в днях.
- Условия хранения: Рекомендуемые условия хранения молочного продукта.
- Дата производства: дата производства молочного продукта.
- Дата истечения срока годности: дата истечения срока годности молочного продукта.
- Проданное количество (литры/кг): количество проданного молочного продукта.
- Цена за единицу (проданного): Цена за единицу, по которой был продан молочный продукт.
- Прибл. Общий доход (INR): приблизительный общий доход, полученный от продажи молочного продукта.
- Местоположение покупателя: Местоположение покупателя, купившего молочный продукт.
- Канал продаж: Канал, через который продавался молочный продукт (Розничная, Оптовая, Интернет).
- Количество на складе (литры/кг): количество молочного продукта, оставшегося на складе.
- Минимальный порог запаса (литры/кг): Минимальный порог запаса молочного продукта.

- Количество для повторного заказа (литры/кг): Рекомендуемое количество молочного продукта для повторного заказа.

### **2.3. Сформировать список отслеживаемых метрик, необходимых для автоматизации, разработки дашборда.**

Потенциальный вариант ключевых показателей для анализа:

- Анализ эффективности молочных ферм с учетом местоположения, площади земель и поголовья коров.
- Понимание моделей продаж и распределения различных молочных продуктов по различным брендам и регионам.
- Изучение влияния условий хранения и сроков годности на качество и доступность молочной продукции.
- Анализ предпочтений и покупательского поведения клиентов в зависимости от местоположения и каналов продаж.
- Оптимизация управления запасами путем отслеживания количества запасов, минимальных пороговых значений и количества повторных заказов.
- Определим потери в ценовой политики компании, в продажах продукции (ценообразование), рассчитаем упущенную выгоду по каждой ферме, относительно остатков на складе.
- Проведем ABC анализ продаж, по основным продуктовым группам.

### **2.4. Составить предварительную визуализацию дашборда**

Дашборд будет сформирован по следующим направлениям, в первую очередь анализ продаж, общий выпуск продукции по каждой ферме в рамках конкретного периода. Остатки на складе, потери типа перепроизводства.

Разделение по территориальному расположению, по брендам и продуктовой группе. Разбивка в процентном соотношении, канал продажи, относительно каждой локации, с возможностью глубокой аналитики по каждой группе и конечному потребителю. Территориальная карта расположения фермы и конечных потребителей. Сформирована динамика для контроля получения общего дохода (INR) в конкретном промежутке времени. В рамках данного дашборда, возможно будет проводить мониторинг и анализ объема производящего продукта и получаемого дохода. Возможно определить динамику и увидеть потенциальные узкие места (потери).

Во вторую очередь, дашборд анализ продаж (план-факт), проданное количество, количество остатков на складе, упущенная выгода, уставки типа минимальный порог запаса молочного продукта, рекомендуемое количество молочного продукта для повторного заказа. Воронка проданного количества, остатки на складе, потенциал для повторного заказа и минимальный порог для повторного заказа, вывели потери, рассчитали упущенную выгоду. Также возможно будет посмотреть динамику стоимости каждого продукта, прайс и итоговая стоимость продаж, определить динамику и сформировать модель стабилизации стоимости, соблюдение прайса и потенциал для корректировки стоимости относительно фактической стоимости продажи, повышение стоимости продукта. Сформировать диаграмму распределения продуктовой группы в процентном соотношении.

В третью очередь, дашборд анализ производства (складские запасы).

Площадь фермы, количество коров, доступное количество молока, динамика остатков, дата производства и срок годности, условия хранения, разбивка по периодам (в рамках одного месяца), динамика производства.

На основе данных показателей и возможности мониторинга динамики ключевых показателей, возможно будет оперативно реагировать на изменения объемов производства, покупательского спроса, типах продуктовых группах и

вносить корректировки в стоимость продукции, для увеличения процента прибыли.

## **2.5. Создать модуль обработки и анализа данных, а также протестировать его на данных компаний по производству молочной продукции**

Для обработки данных и формирования анализа будем использовать python библиотеку pandas, matplotlib, seaborn в редактор исходного кода Visual Studio Code. Основные показатели отобразим в Power BI, для возможности производить мониторинг и оперативно реагировать на показатели.

Загружаем файл с данными из локального диска формата .csv, визуально изучаем его, определяем дальнейшие шаги, нашего анализа. Производим корректировку названия столбцов, для удобства использовать их в коде программы. Сформировал следующее обозначение (приложение 2 во вложении).

По ходу анализа, определил количество ферм, согласно их расположению и площади. Распределение по бренду и количеству выпущенной продукции. Локации ферм - Telangana, Uttar Pradesh, Tamil Nadu, Maharashtra, Karnataka, Bihar, West Bengal, Madhya Pradesh, Chandigarh, Delhi, Gujarat, Kerala, Jharkhand, Rajasthan, Haryana. Сформировал продуктовую группу, тип продукции: 'Ice Cream' - мороженое, 'Milk' - молоко, 'Yogurt' - йогурт, 'Cheese' - сыр, 'Buttermilk' - пахта, 'Curd' - творог, 'Paneer' - панир, 'Lassi' - лассии, 'Ghee' - топленое масло, 'Butter' - масло. Каналы продаж - 'Wholesale' -оптовая, 'Online' - онлайн, 'Retail' - розница.

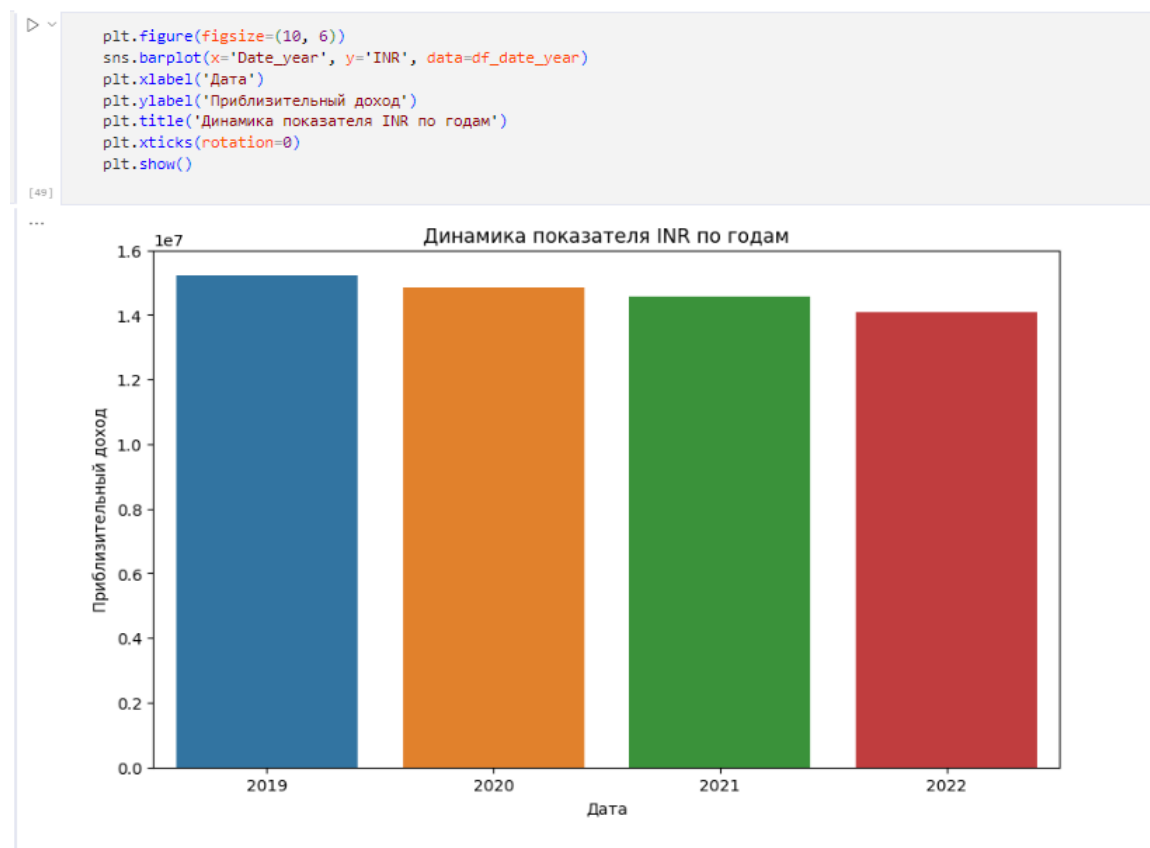
Очистил датасет, проверил на дубликаты и удалил повторяющие данные. Первое с чего начал процесс анализа, с определения общей стоимости произведенного и доступного количества молочной продукции за весь период с 2019-2022 г. Вывел общую стоимость доступного количества молочного продукта и полученный общий доход, от продажи молочного продукта, за данный период времени. Количество доступного молочного продукта,



количество проданного молочного продукта и количество молочного продукта, оставшегося на складе, зафиксировал данные показатели.

Результат - большие потери, относительно информации, которая предоставлена в этом датасете.

Формирую и вывожу динамику показателей общего дохода по периоду (год):



Вывод: отрицательная динамика дохода в сравнение с предыдущим годом, БЕРЕМ ВО ВНИМАНИЕ.

Вывод: просматривается отрицательная динамика общего дохода по году в сравнение с предыдущим 2019 годом.

Интересные показатели, стабильная отрицательная динамика в течение долгого промежутка времени. Три года без конкретных действий, без оптимизации процессов, попытки изменить положение и скорректировать производство плюс модель продаж.

Формирую и вывожу диаграмму количество доступного молочного продукта, динамика по годам.

	Date_year	Quan
1	2020	558973.06
0	2019	552693.55
2	2021	534675.60
3	2022	518980.53

```
plt.figure(figsize=(10, 6))
sns.barplot(x='Date_year', y='Quan', data=df_date_year_Quan)
plt.xlabel('Дата')
plt.ylabel('Количество доступного молочного продукта')
plt.title('Динамика показателя Quan по годам')
plt.xticks(rotation=0)
plt.show()
```

[82] ✓ 0.4s



Вывод: диаграмма показывает, что год 2020, был самым проблемным, общий доход пошел на спад, объем производства в максимальных показателях, самые большие потери. Так как по объему выпускаемой продукции, производство работало аналогично показателям 2019 года плюс корректирующий показатель роста, а по сбыту и продажам был спад. Предположение, 2019 год — это начало эпидемии по всему миру, коронавирус, резкие ограничение и снижение спроса в каждом направлении.

Продолжаю анализ, в первую очередь анализирую ценовую политику, какой плановый показатель стоимости (прайс) и факт по стоимости продажи продукта, отличаются эти показатели. Формирую новый столбец, количество проданного молочного продукта, умноженного на стоимость за единицу, по

которой был продан молочный продукт (фактическая стоимость). Итоговый результат, общий доход с учетом фактической стоимости продукта, равен показателям (INR) приблизительного общего дохода, который рассчитывается следующим образом, объем произведенного молочного продукта, умноженный на стоимость по прайсу, ожидаемый эффект, полученный от продажи молочного продукта.

Далее, необходимо рассмотреть в динамике, распределение фактической стоимости, относительно стоимости по прайсу. В каких случаях, стоимость продажи была меньше или больше прайса. Данный процесс, позволит определить узкие места и потенциал для роста, назовем данный показатель «Упущенная выгода».

Производим разбивку общей прибыли по годам и локациям фермы, группируем по расположению фермы и проданному количеству молочного продукта, умноженного на медианное значение стоимости, по которой был продан данный тип продукта.

Анализирую полученные данные и определяю упущенную выгоду, фактическая стоимость минус ожидаемая стоимость продукта (фактическая стоимость продукта минус стоимость продукта в прайсе).

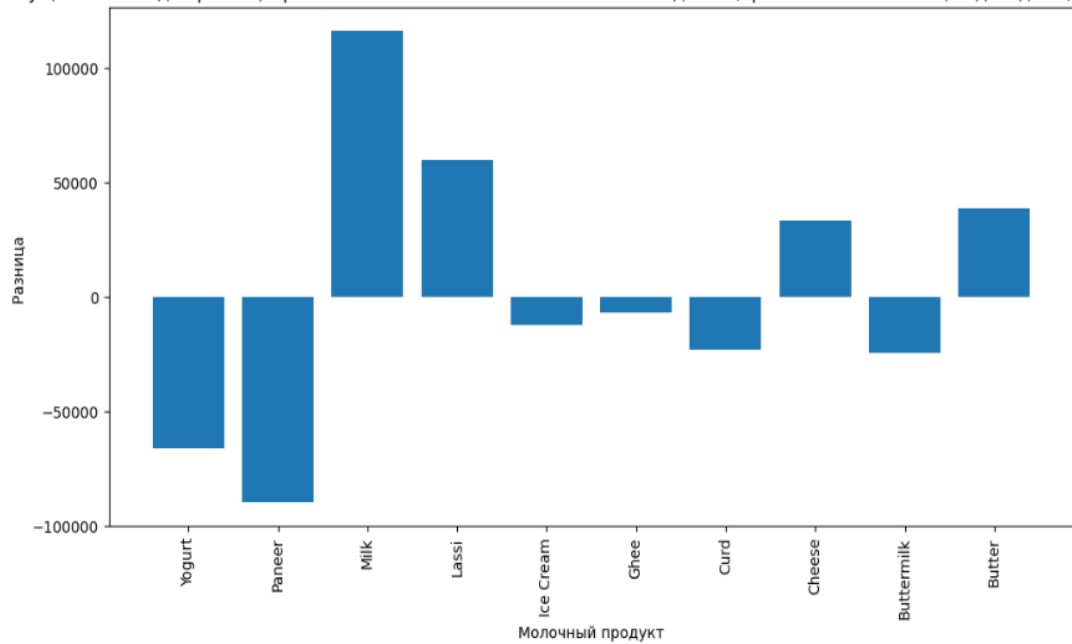
Для анализа, возьмём показатель "INR" (приблизительный) общий доход за показатель «ожидания», построим гистограмму относительно ожидания и факт (проданное кол-во продукта \* на фактическую стоимость, по которой продали продукт), сравним два показателя.

```

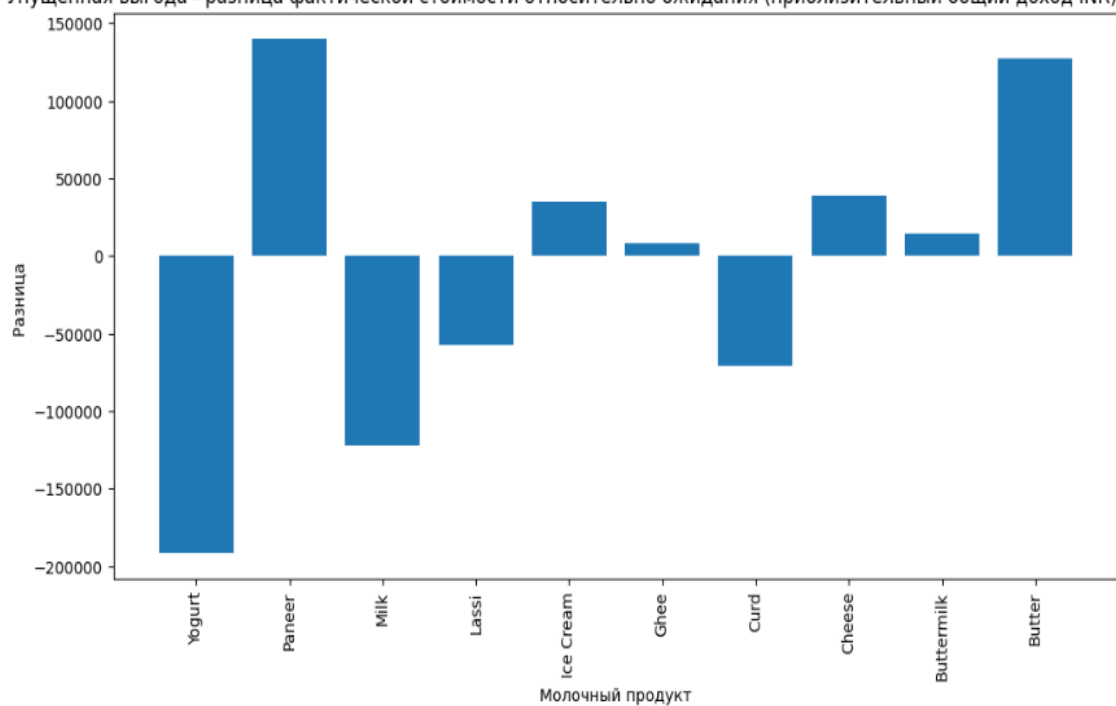
for date_Price in df_Price_INR:
    plt.figure(figsize=(12, 6))
    df_Price_INR_lost = df_lost_profit_Price[df_lost_profit_Price['Date_year'] == date_Price]
    x = df_Price_INR_lost['Prod_Name']
    plt.bar(x, df_Price_INR_lost['lost_profit_INR'])
    plt.xlabel('Молочный продукт')
    plt.ylabel('Разница ')
    plt.title(f'Упущенная выгода - разница фактической стоимости относительно ожидания (приблизительный общий доход INR). - {date_Price} год')
    plt.xticks(rotation=90)
    plt.show()

```

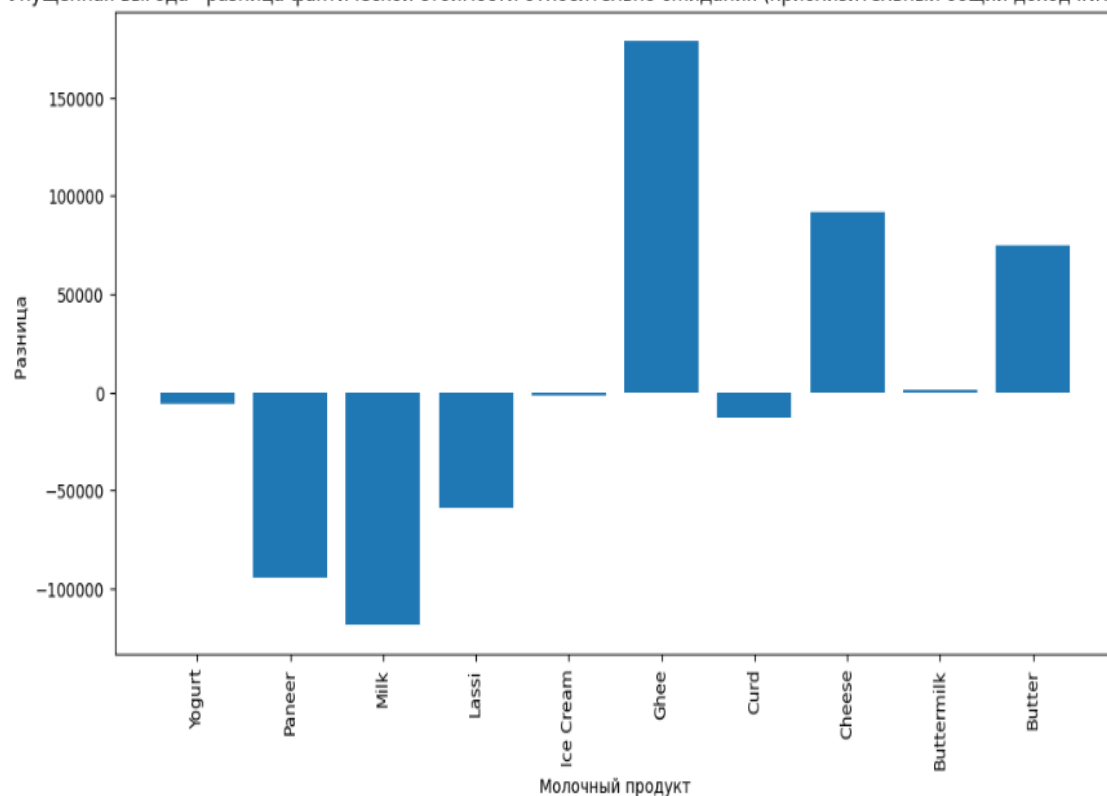
Упущенная выгода - разница фактической стоимости относительно ожидания (приблизительный общий доход INR). - 2022 год



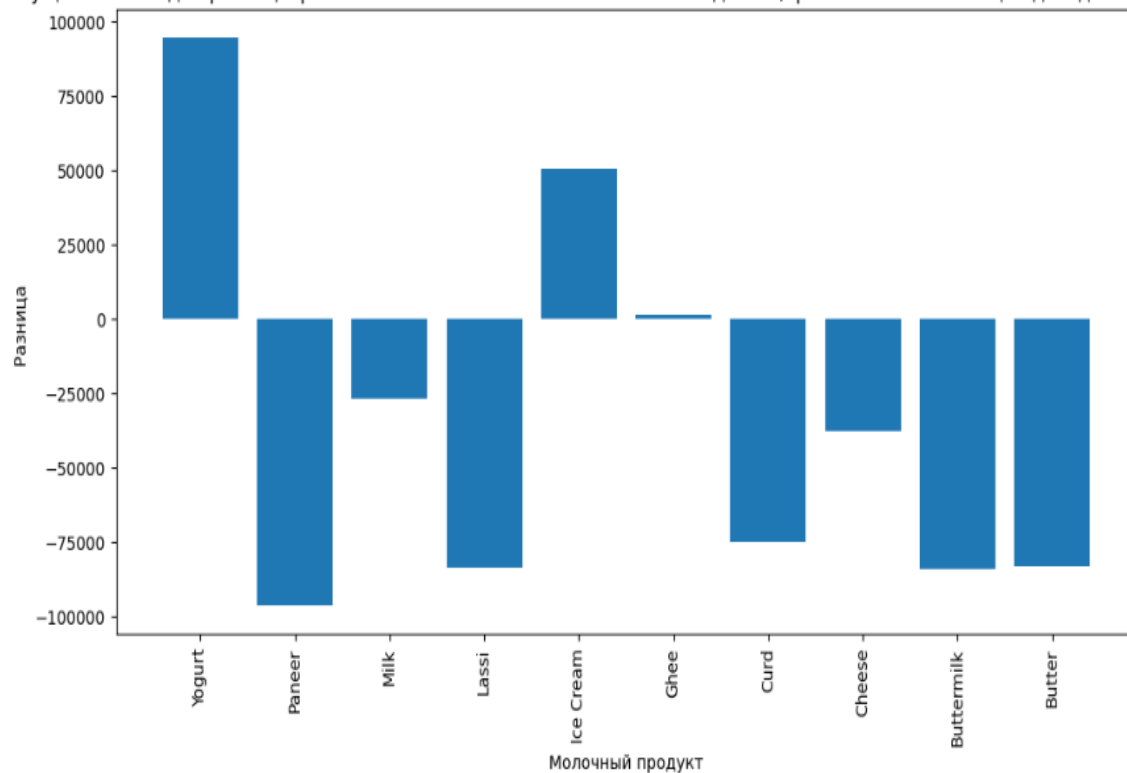
Упущенная выгода - разница фактической стоимости относительно ожидания (приблизительный общий доход INR). - 2021 год



Упущенная выгода - разница фактической стоимости относительно ожидания (приблизительный общий доход INR). - 2020 год



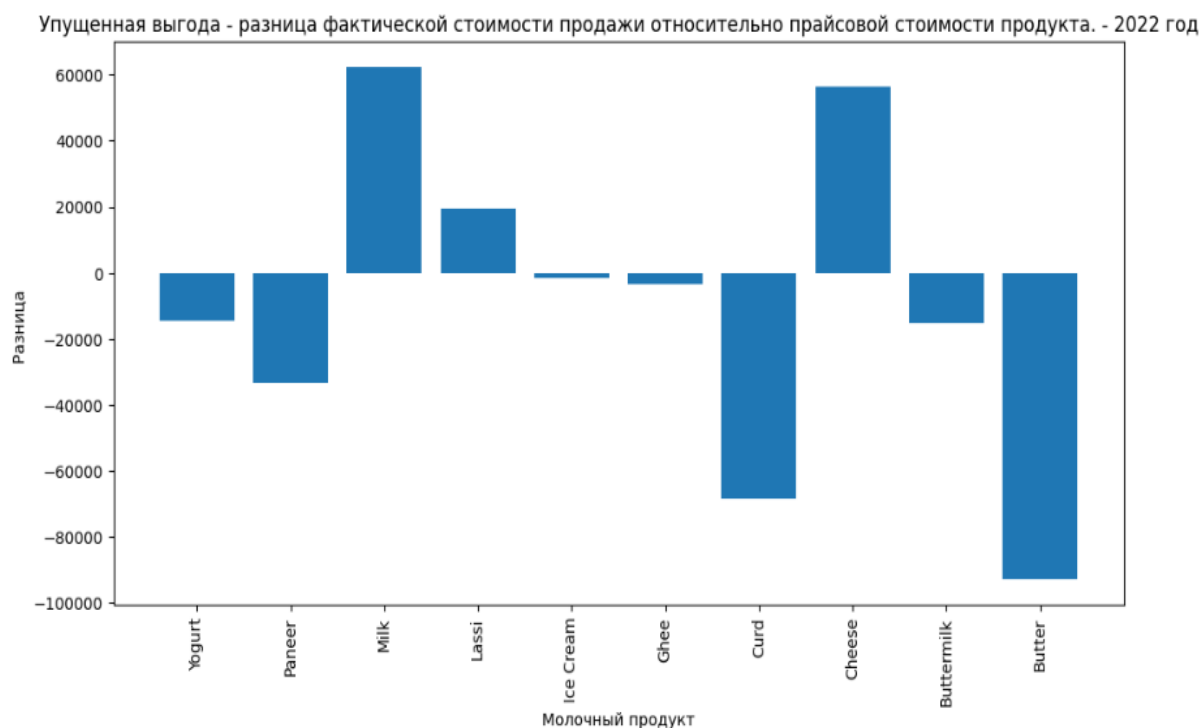
Упущенная выгода - разница фактической стоимости относительно ожидания (приблизительный общий доход INR). - 2019 год



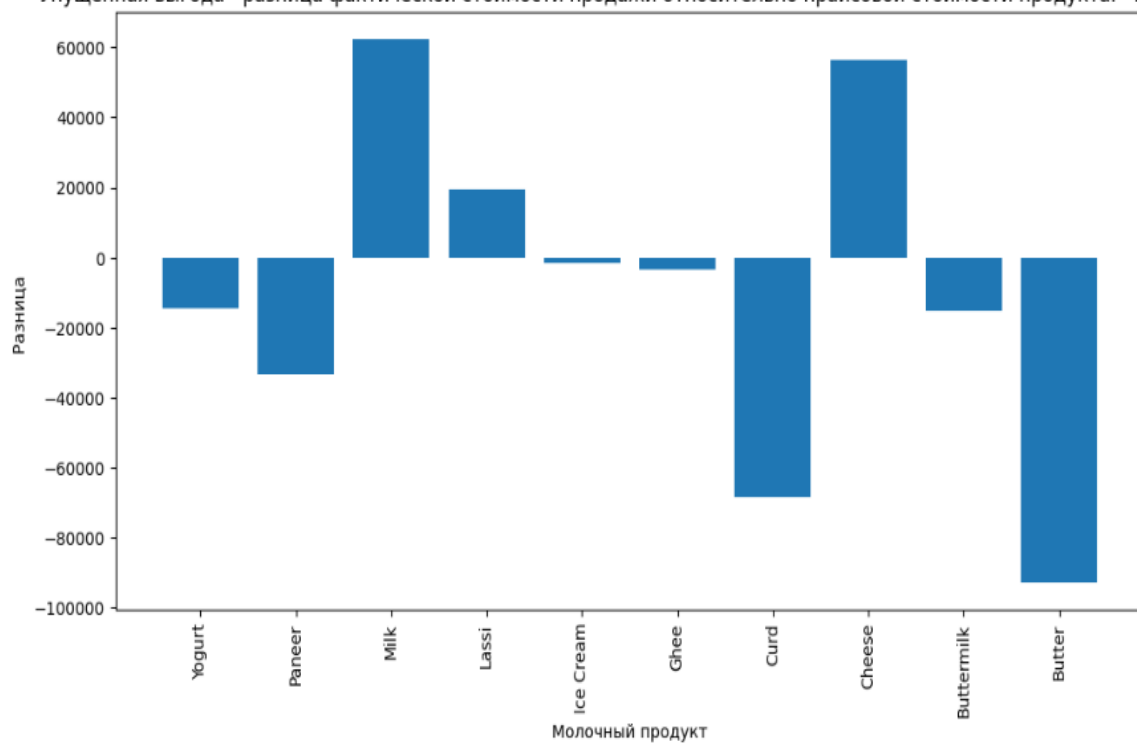
Вывод: сгруппировал данные по количеству проданного продукта, по какой стоимости был продан продукт, сравнил с установленным прайсом на данный тип продукта (ожидание). Произвел расчёт упущенной выгоды и наоборот,

позиции, где продавали дороже чем прайс (перспектива развития). Если продажи были выше прайса, отклонения были стабильные, имеет смысл поднять стоимость в прайсе, взять за базовый уровень. Тем самым, возможно будет скорректировать КПЭ у отдела продаж, мотивировать для улучшения данного показателя, увеличение прибыли, но в любом случае необходимо провести ABC-XYZ анализ, только после этого производить корректировку стоимости в прайсе.

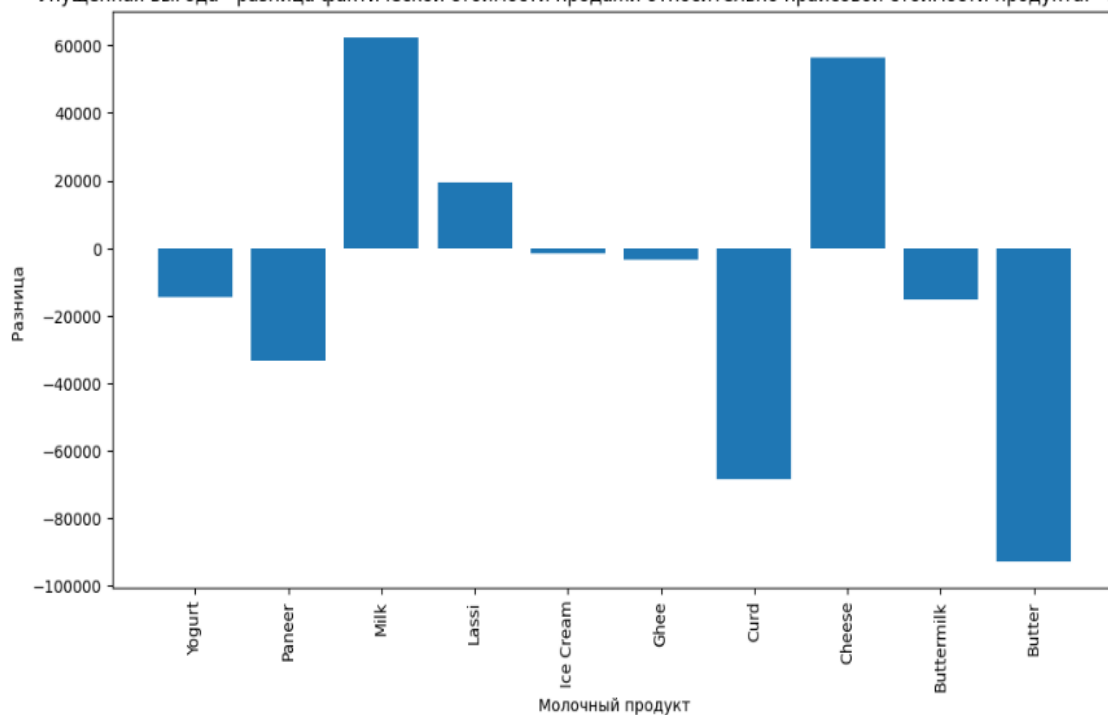
Рассмотрим упущенную выгоду относительно прайсовой стоимости и фактической стоимости продажи продукта:

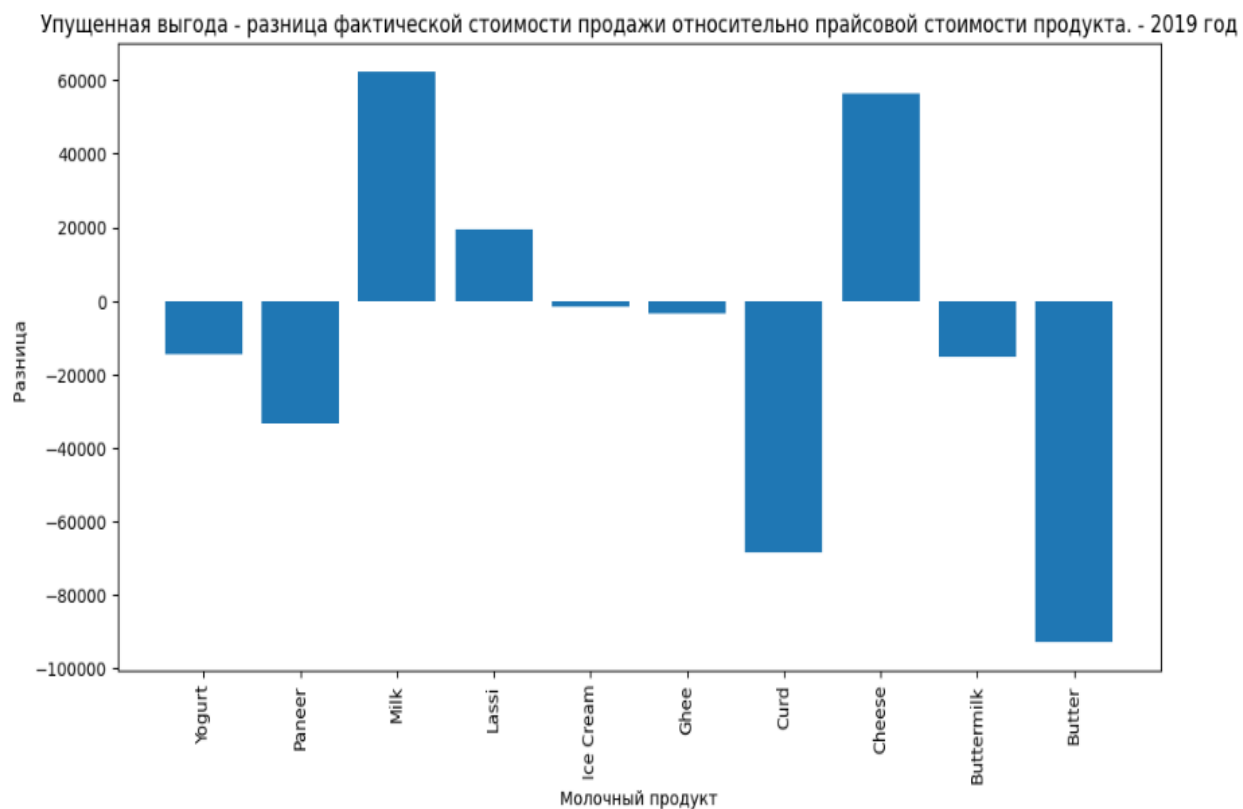


Упущенная выгода - разница фактической стоимости продажи относительно прайсовой стоимости продукта. - 2021 год



Упущенная выгода - разница фактической стоимости продажи относительно прайсовой стоимости продукта. - 2020 год





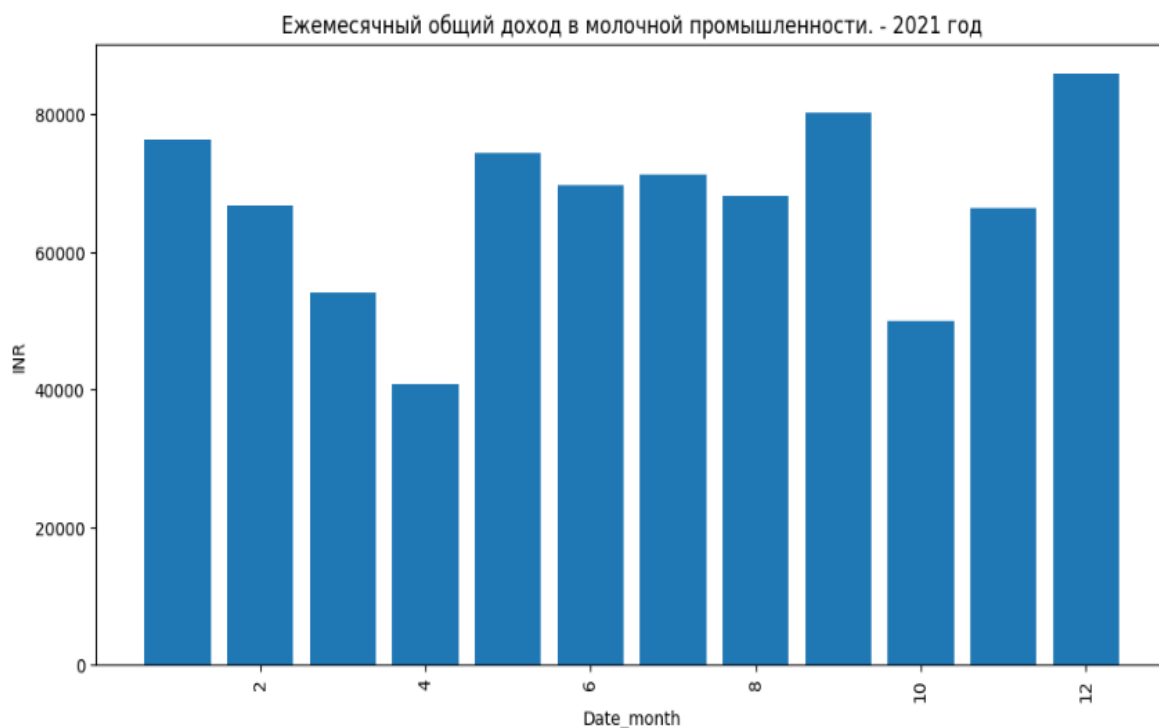
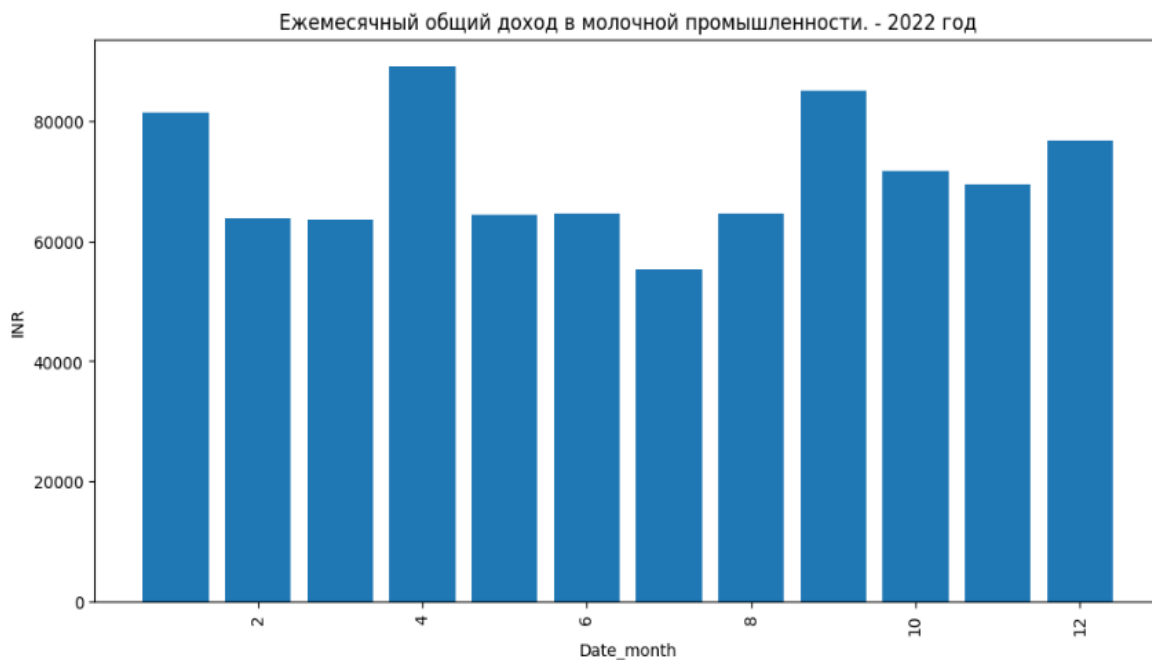
Вывод: необходимо пересмотреть модель продаж, проводить анализ, выявлять причину, почему в некоторой группе продуктов, требуется постоянно давать скидки, а по остальной части продавать дороже. По итогу, картина компенсируется, потери выравниваются и итоговые показатели выравниваются. Перспективы на развитие имеются как со стороны производства, так и со стороны отдела продаж, изменение стоимости прайса (доходная часть/маржинальный доход). В первую очередь, я рекомендую сделать акцент на корректировке модели продаж.

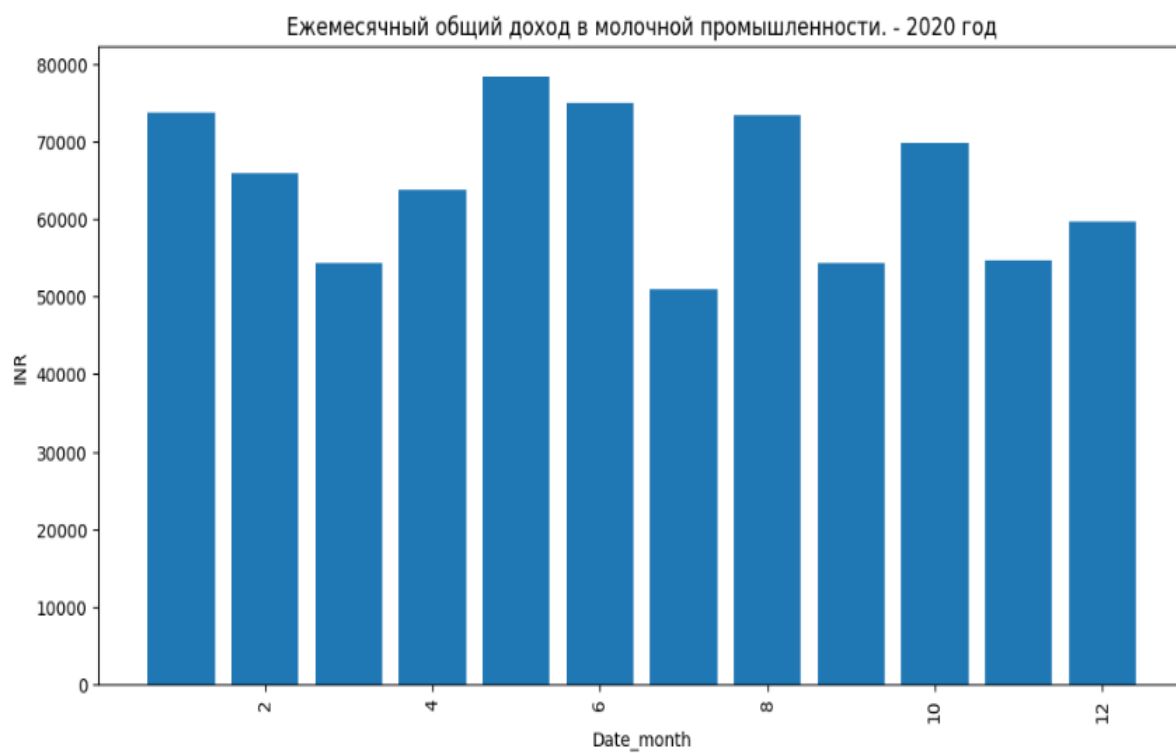
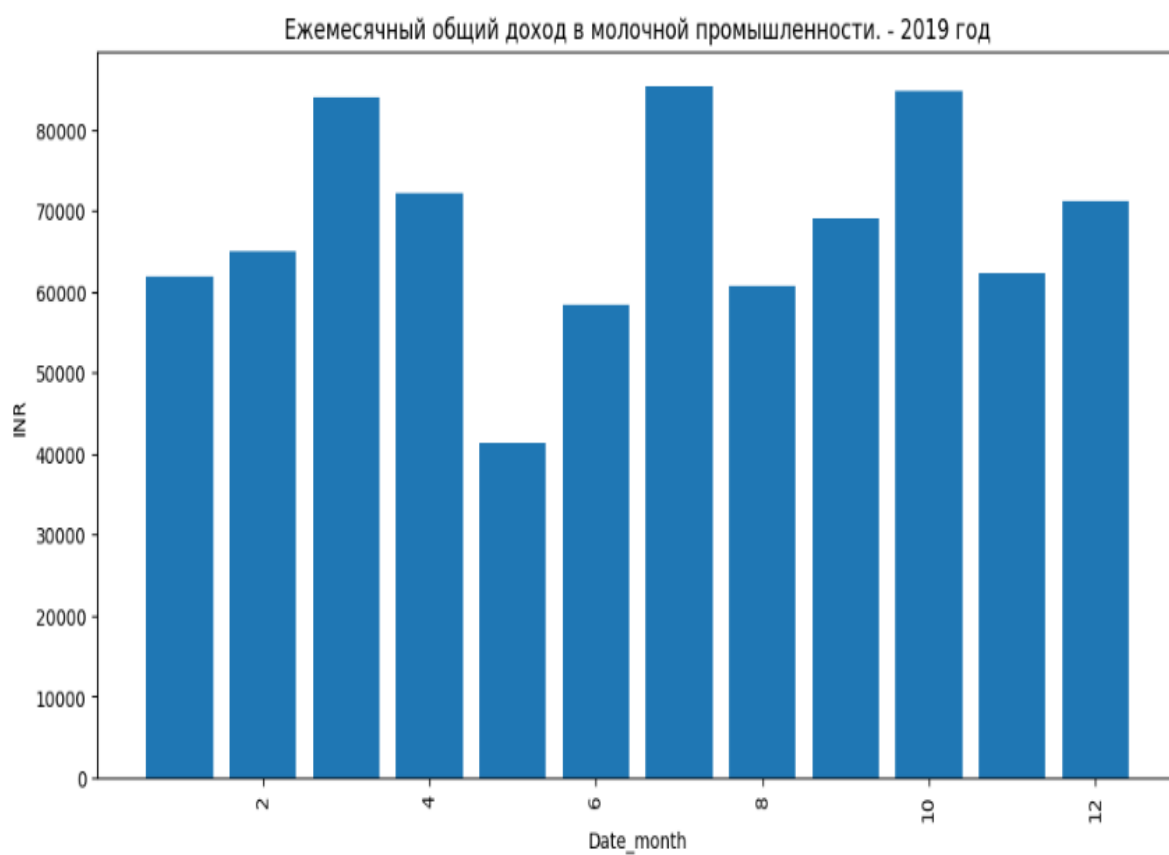
Произвел группировку по месяцу, для определения ежемесячного общего дохода.



```
df_monthly_date = df_monthly.groupby('Date_month').agg({
    'INR': 'sum'
}).reset_index()
```

```
for date_monthly in df_date_monthly:
    plt.figure(figsize=(12, 6))
    df_dateMonthly = df_monthly[df_monthly['Date_year'] == date_monthly]
    x = df_dateMonthly['Date_month']
    plt.bar(x, df_dateMonthly['INR'])
    plt.xlabel('Date_month')
    plt.ylabel('INR')
    plt.title(f'Ежемесячный общий доход в молочной промышленности. - {date_monthly} год')
    plt.xticks(rotation=90)
    plt.show()
```





Вывод: общий доход в 2022 году был более равномерным по сравнению с другими. Январь, апрель, сентябрь и декабрь — месяцы с многообещающими продажами. Июль показывает большинство изменений.

Произвожу группировку, приблизительный общий доход по бренду и разбивка по месяцам.

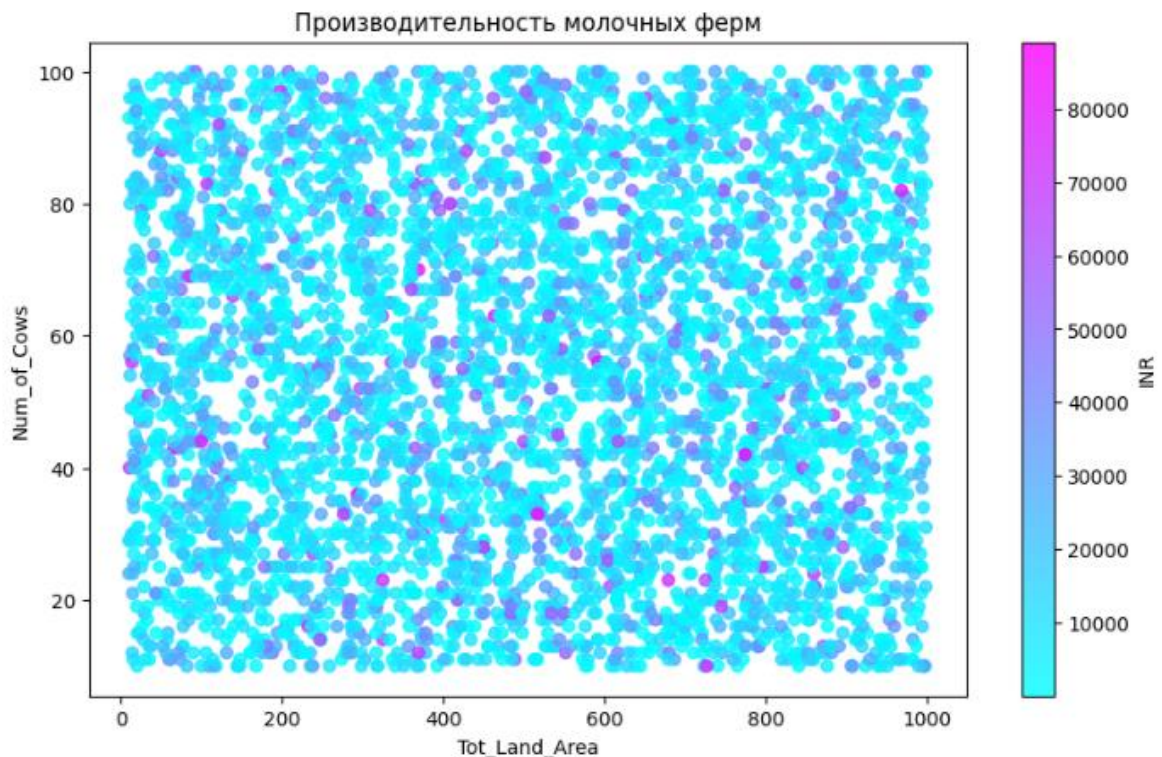
```
sns.lineplot(data=monthly_brand_total, x='Date_month', y='INR', hue='Brand')
plt.title('Общая стоимость по месяцам относительно бренда')
plt.xticks(range(1, 13))
plt.legend(loc='center left', bbox_to_anchor=(1, 0.5))
plt.show()
```



На мой взгляд более-менее стабильная динамика по общей доходности у бренда Britannia Industries.

Произвожу группировку, локация, общая площадь фермы, количество коров, формирую диаграмму производительность фермы с учетом количества коров.

```
plt.figure(figsize=(10, 6))
plt.scatter(df_farm_performance['Tot_Land_Area'], df_farm_performance['Num_of_Cows'],
           c=df_farm_performance['INR'], cmap='cool', alpha=0.8)
plt.xlabel('Tot_Land_Area')
plt.ylabel('Num_of_Cows')
plt.title('Производительность молочных ферм')
plt.colorbar(label='INR')
plt.show()
```



Вывод: на первый взгляд более производительней являются фермы с площадью от 200 до 650 единиц, количество коров от 40 до 80, а также с более высокой доходностью. Но явного преимущества и зависимости от площади и количества коров не наблюдается.

Вывожу аналитические показатели из датасета по столбцам count, mean, std, min, 25%, 50%, 75%, max.

```
df.describe(include=['object'])
```

✓ 0.0s

	Loc	Farm_Size	Prod_Name	Brand	Storage_Con	Cust_Loc	Sales_Channel
count	4325	4325	4325	4325	4325	4325	4325
unique	15	3	10	11	5	15	3
top	Delhi	Large	Curd	Amul	Refrigerated	Delhi	Retail
freq	525	1462	479	1053	2459	499	1478

```
df.describe(include=['float'])
```

✓ 0.0s

	Tot_Land_Area	Quan	Price_per_Unit	Tot_Value	Price_per_Unit(sold)	INR	Min_Stock_Threshold	Reo_Quan
count	4325.000000	4325.000000	4325.000000	4325.000000	4325.000000	4325.000000	4325.000000	4325.000000
mean	503.483073	500.652657	54.785938	27357.845411	54.77914	13580.265401	55.826143	109.107820
std	285.935061	288.975915	26.002815	21621.051594	26.19279	14617.009122	26.301450	51.501035
min	10.170000	1.170000	10.030000	42.516500	5.21000	12.540000	10.020000	20.020000
25%	252.950000	254.170000	32.460000	9946.814500	32.64000	2916.650000	32.910000	64.280000
50%	509.170000	497.550000	54.400000	21869.652900	54.14000	8394.540000	56.460000	108.340000
75%	751.250000	749.780000	77.460000	40954.441000	77.46000	19504.550000	79.010000	153.390000
max	999.530000	999.930000	99.990000	99036.369600	104.51000	89108.900000	99.990000	199.950000

```
df.describe(include=['int64'])
```

✓ 0.0s

	Num_of_Cows	Prod_ID	Shelf_Life	Quan_Sold	Quan_in_Stock
count	4325.000000	4325.000000	4325.000000	4325.000000	4325.000000
mean	54.963699	5.509595	29.127630	248.095029	252.068671
std	26.111487	2.842979	30.272114	217.024182	223.620870
min	10.000000	1.000000	1.000000	1.000000	0.000000
25%	32.000000	3.000000	10.000000	69.000000	66.000000
50%	55.000000	6.000000	22.000000	189.000000	191.000000
75%	77.000000	8.000000	30.000000	374.000000	387.000000
max	100.000000	10.000000	150.000000	960.000000	976.000000

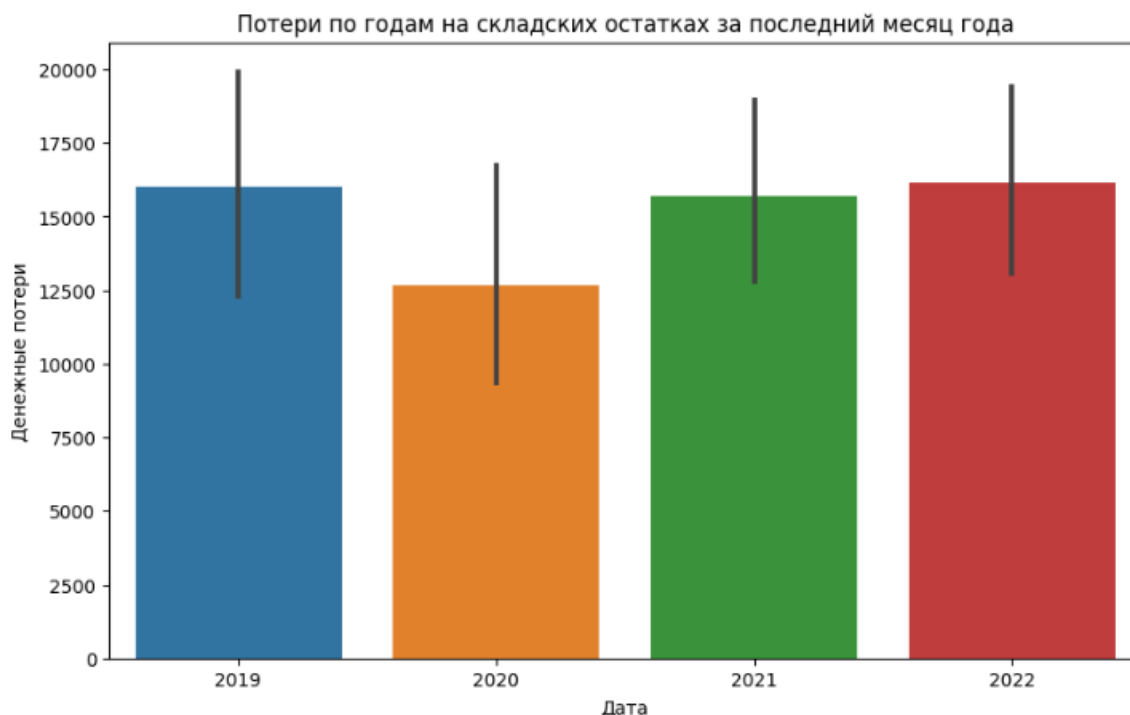
В первую очередь обращаю внимание на последние два столбца, 'Quan\_Sold' и 'Quan\_in\_Stock', показатели по остатку на складе завышены, данный формат реквизитов датасета, не позволяют сделать подробный и точный анализ остатков на складе, необходимо рассматривать в более короткие промежутки времени (периоды) и с фиксацией необходимой информации в датасете. Допустим в разрез одного месяца, сколько было произведено, какой остаток на складе, сколько было реализовано в следующем месяце из остатков предыдущего месяца, сколько продукции было отправлено на утилизацию, какой объем переместили в следующий период. Если делать анализ по складским остаткам в период одного года, как вариант, рассмотреть только последний месяц текущего года с выводом медианного значения.

Сформируем датасет и сделаем советующую группировку и построим диаграмму.

```

plt.figure(figsize=(10, 6))
sns.barplot(x='Date_year', y='Quan_in_Stock*Price_per_Unit', data=df_date_stock_tot)
plt.xlabel('Дата')
plt.ylabel('Денежные потери')
plt.title('Потери по годам на складских остатках за последний месяц года')
plt.xticks(rotation=0)
plt.show()

```



Вывод: из barplot видно, что большие потери из-за складских остатков, были в 2022. В общем, каждый год имеет большие потери по факту окончания года. Необходимо сделать обследование производства, организовать операционную трансформацию на производстве. Запустить проект по стратегическому планированию и разработать прогнозную модель продаж.

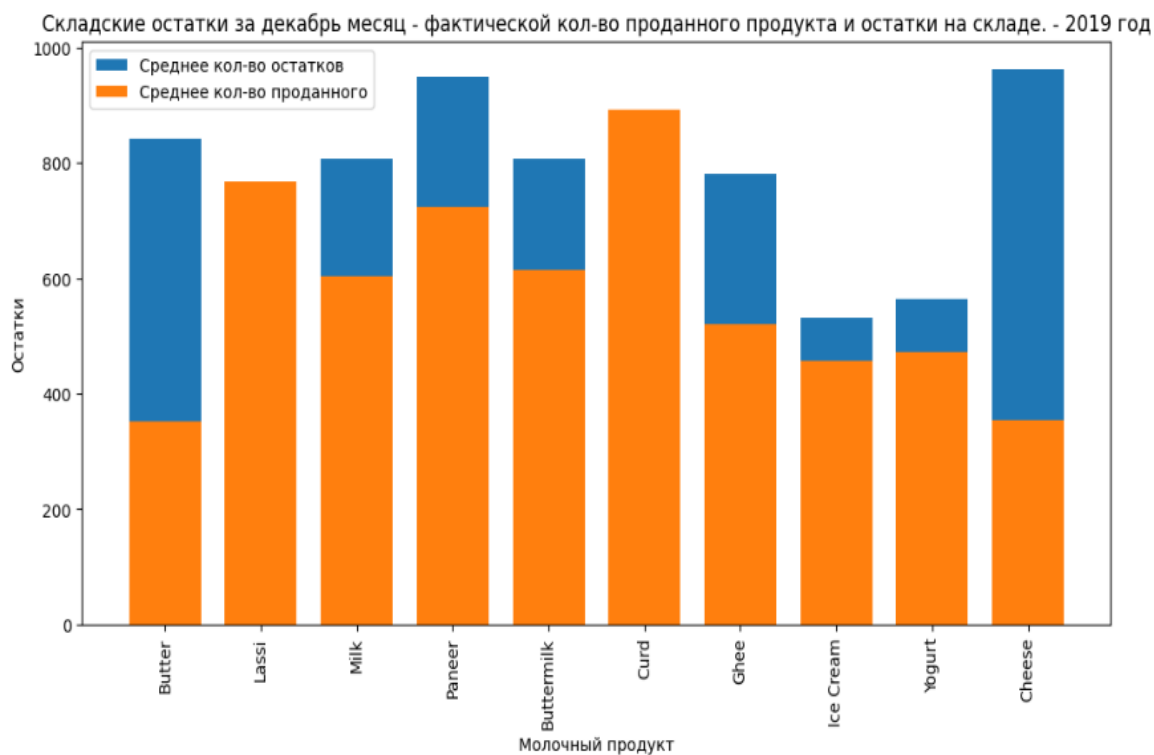
В модели мониторинга ключевых показателей, построенную на площадке power BI, возможно будет рассмотреть данные показатели в динамике, в разрезе дня, месяца, года, производить глубокий мониторинг.

Продолжаем... группирую, визуализирую складские остатки за декабрь месяц – фактическое количество проданного продукта и остатки на складе в разрезе по годам:

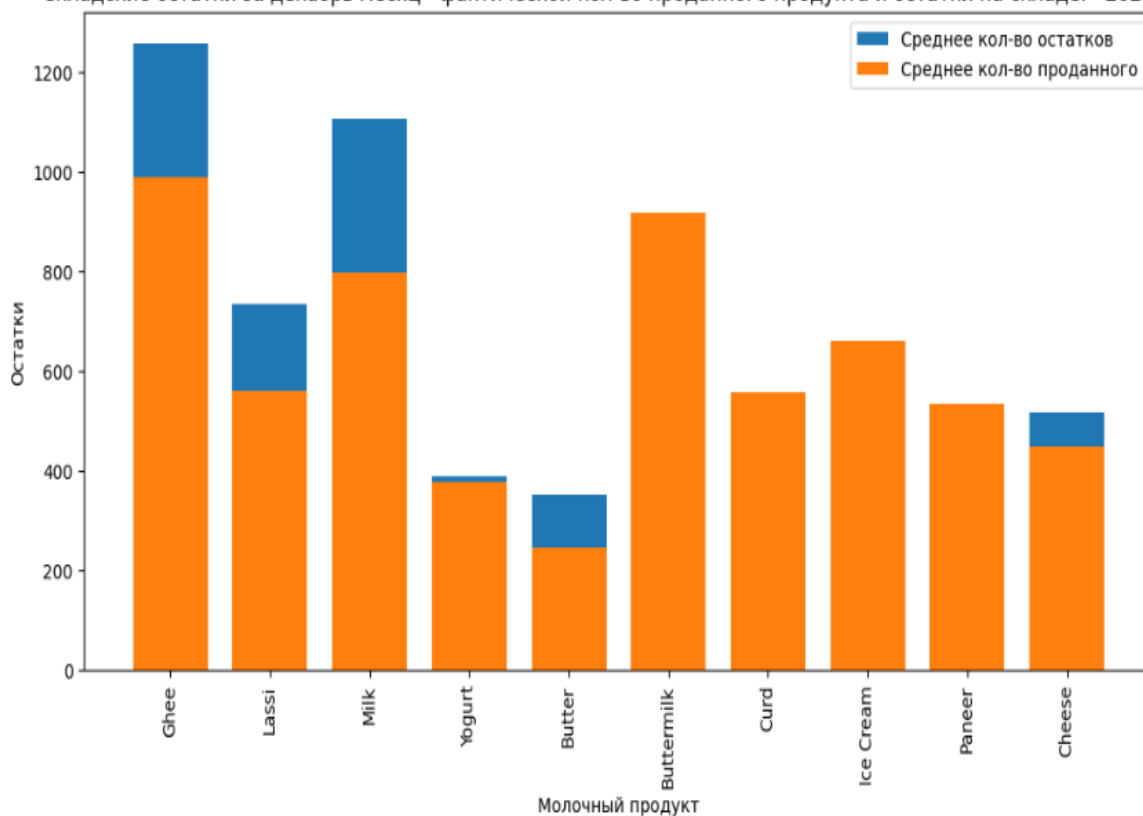
```

for date_Sold_Stock in df_date_Sold_Stock:
    plt.figure(figsize=(12, 6))
    width = 0.75
    df_dateSoldStock = df_date_stock[df_date_stock['Date_year'] == date_Sold_Stock]
    x = df_dateSoldStock['Prod_Name']
    plt.bar(x, df_dateSoldStock['Quan_in_Stock'], width, label='Среднее кол-во остатков')
    plt.bar(x, df_dateSoldStock['Quan_Sold'], width, label='Среднее кол-во проданного')
    plt.xlabel('Молочный продукт')
    plt.ylabel('Остатки')
    plt.title(f'Складские остатки за декабрь месяц - фактической кол-во проданного продукта и остатки на складе. - {date_Sold_Stock} год')
    plt.xticks(rotation=90)
    plt.legend()
    plt.show()

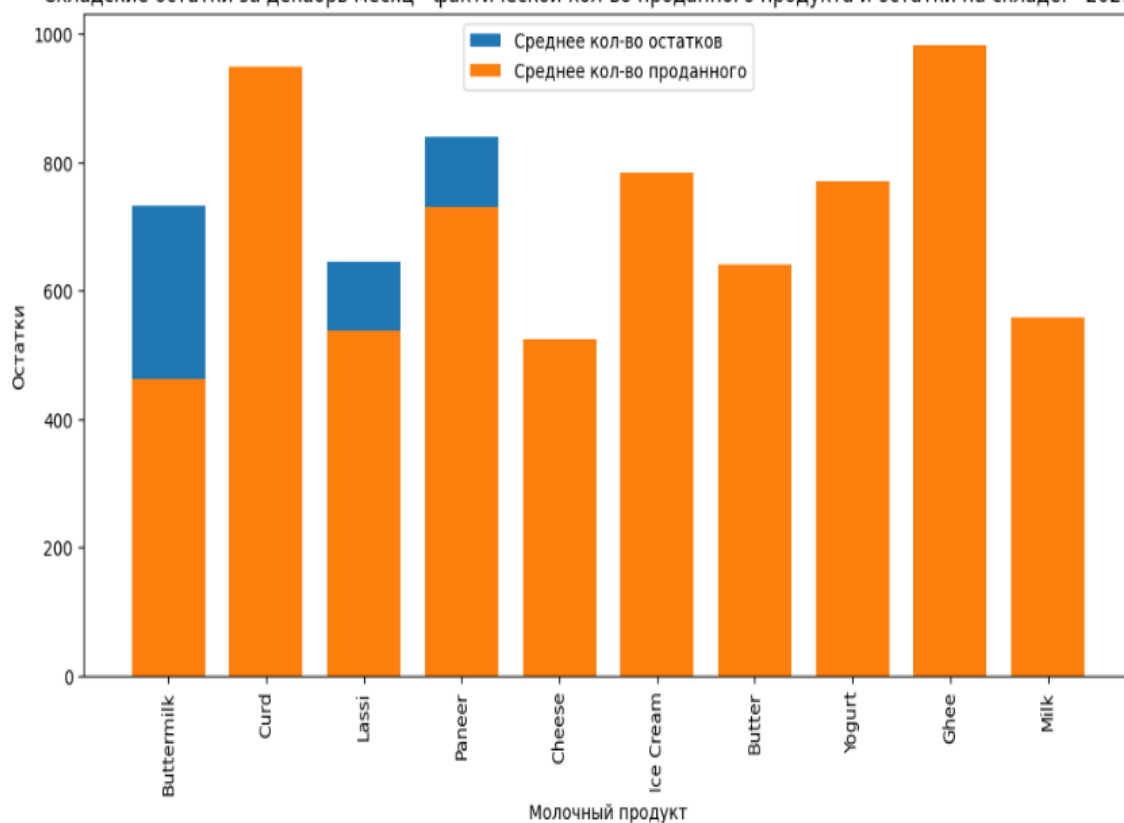
```



Складские остатки за декабрь месяц - фактической кол-во проданного продукта и остатки на складе. - 2020 год

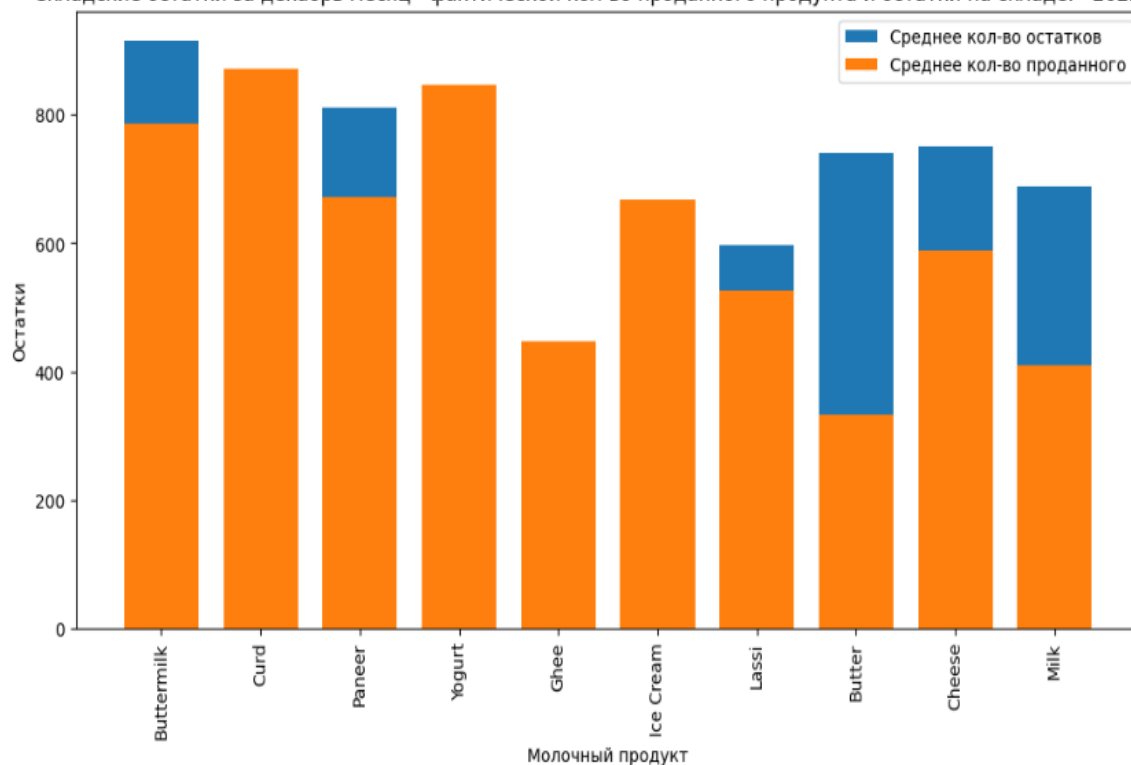


Складские остатки за декабрь месяц - фактической кол-во проданного продукта и остатки на складе. - 2021 год





Складские остатки за декабрь месяц - фактической кол-во проданного продукта и остатки на складе. - 2022 год



Вывод: из данных диаграмм видно, что выше сказанное необходимо принять во внимание и незамедлительно начать процесс анализа и устранения потерь, большой показатель фактических остатков на складе, разбивка по продуктовой группе и годам. Для данного типа продукции, с минимальным сроком годности и небольшим временем производства, складские остатки должны быть минимальные, близкие к нулю, производство - отгрузка – продажа.

Рассмотрим распределение бренда и типов молочных продуктов, а также их зависимость.

Проверяем частоту встречаемости локаций (ферм):

```
df['Loc'].value_counts(normalize=True, dropna=False)
```

```
Loc
Delhi          0.121387
Chandigarh     0.120000
Uttar Pradesh  0.063815
Gujarat        0.061734
Karnataka      0.060347
Madhya Pradesh 0.059884
Rajasthan      0.059191
Maharashtra    0.058960
Haryana        0.058497
Kerala         0.057572
Telangana      0.057341
Jharkhand      0.057341
Bihar          0.056647
West Bengal    0.055723
Tamil Nadu     0.051561
Name: proportion, dtype: float64
```

Проверяем частоту встречаемости бренда:

```
df['Brand'].value_counts(normalize=True, dropna=False)
```

```
Brand
Amul          0.243468
Mother Dairy  0.233526
Raj           0.158382
Sudha         0.149827
Dodla Dairy   0.051329
Palle2patnam  0.048786
Dynamix Dairies 0.024509
Warana        0.024046
Parag Milk Foods 0.023584
Passion Cheese 0.022197
Britannia Industries 0.020347
Name: proportion, dtype: float64
```

Проверяем частоту встречаемости типа продукта:

```
df['Prod_Name'].value_counts(normalize=True, dropna=False)
```

Prod_Name	
Curd	0.110751
Lassi	0.103353
Paneer	0.101965
Yogurt	0.101040
Buttermilk	0.100578
Butter	0.099653
Milk	0.099191
Ice Cream	0.097803
Ghee	0.092948
Cheese	0.092717

Name: proportion, dtype: float64

Вывод: самые востребованные фермы Delhi и Chandigarh, так как за период 2019, 2020, 2021, 2022, чаще всего молочная продукция поступала именно с этих локации.

Строим круг с разбивкой рыночной доли молочных брендов:

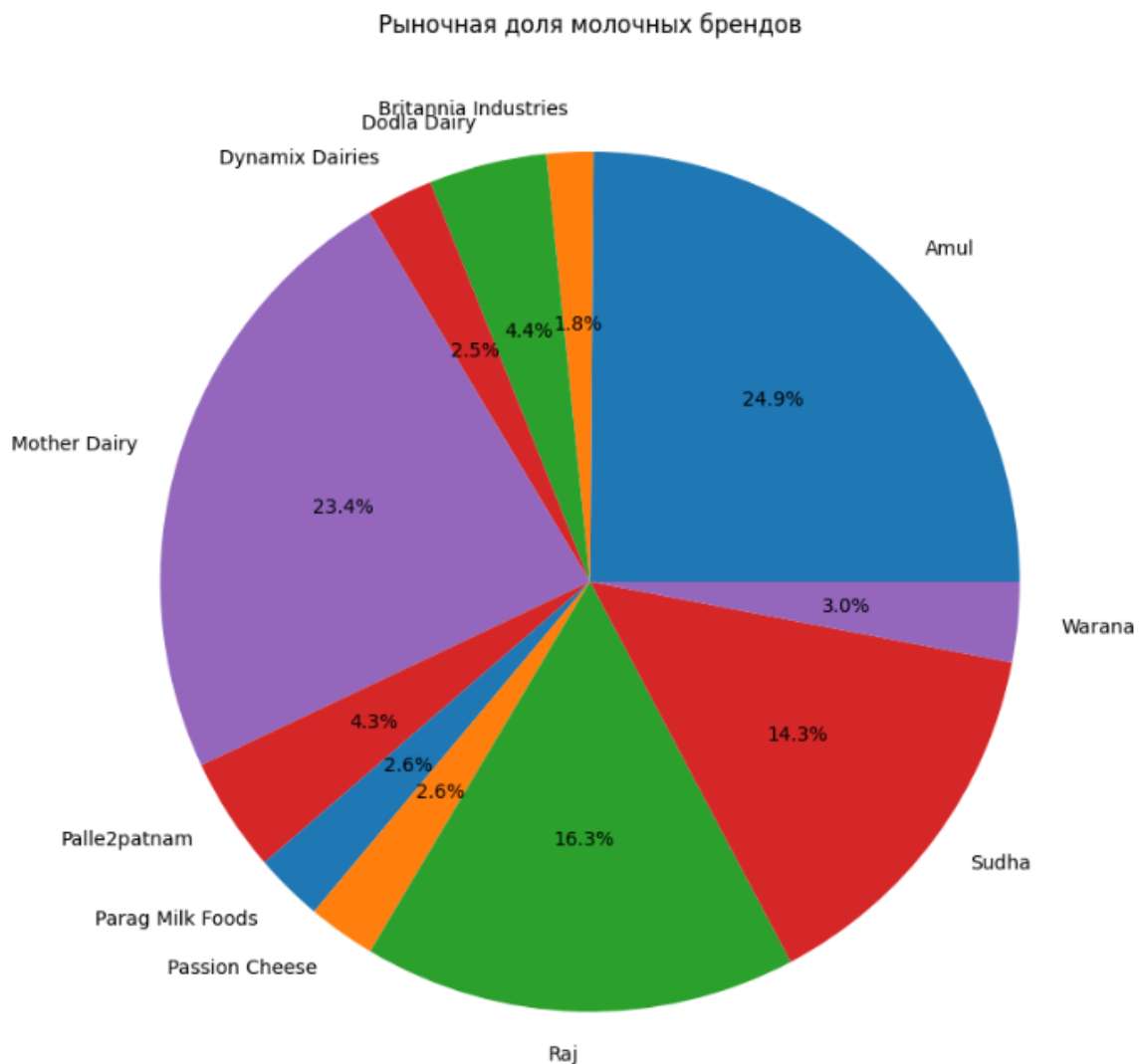
```

df_brand_market_share = df.groupby('Brand') \
    .agg({'INR': 'sum'}) \
    .reset_index()

plt.figure(figsize=(10, 10))
colors = ['tab:blue', 'tab:orange', 'tab:green', 'tab:red', 'tab:purple', 'tab:red']

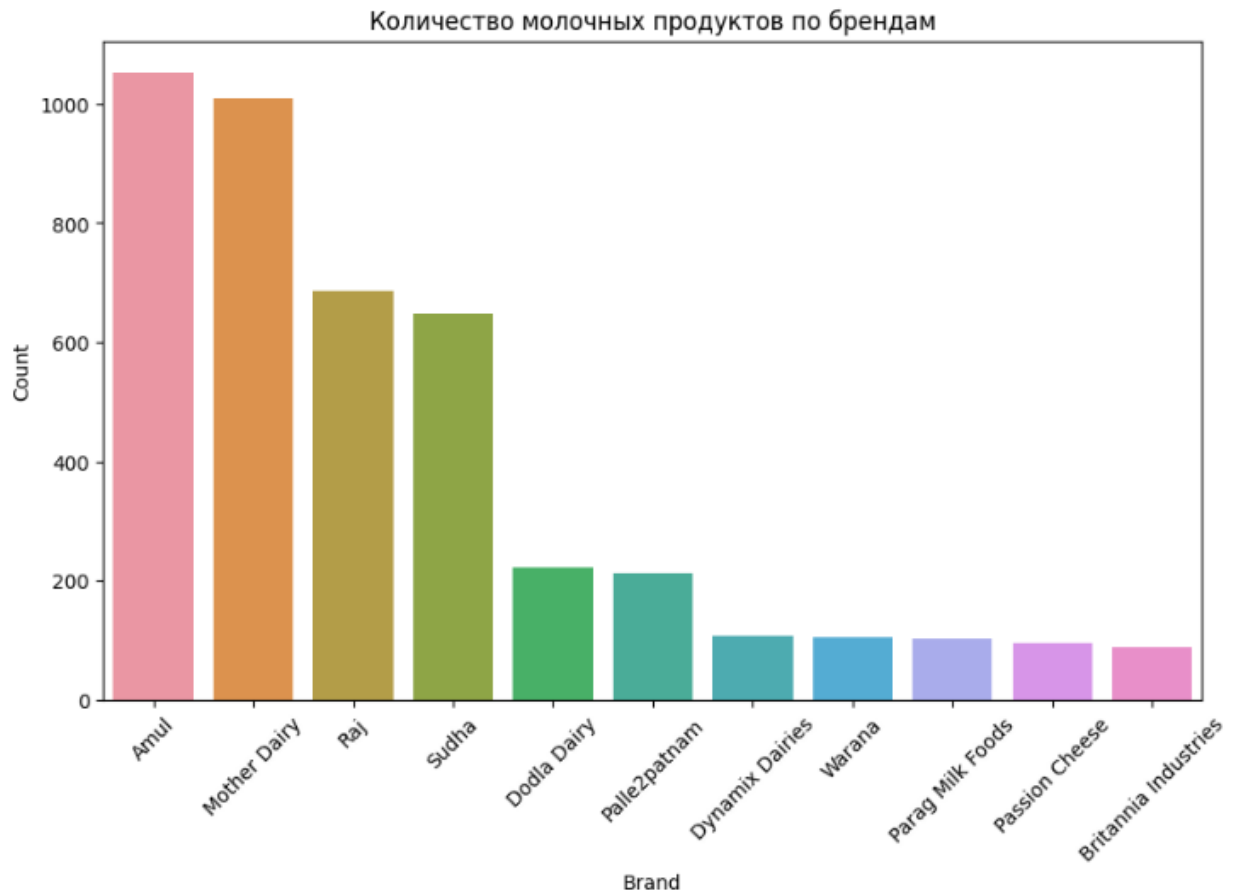
plt.pie(df_brand_market_share['INR'],
        labels=df_brand_market_share['Brand'],
        autopct='%1.1f%%',
        colors=colors)
plt.title('Рыночная доля молочных брендов')
plt.show()

```



Вывод: самая высокая доля рынка = Амуль (24,9%), самая низкая доля рынка = Britannia Industries (1,8%).

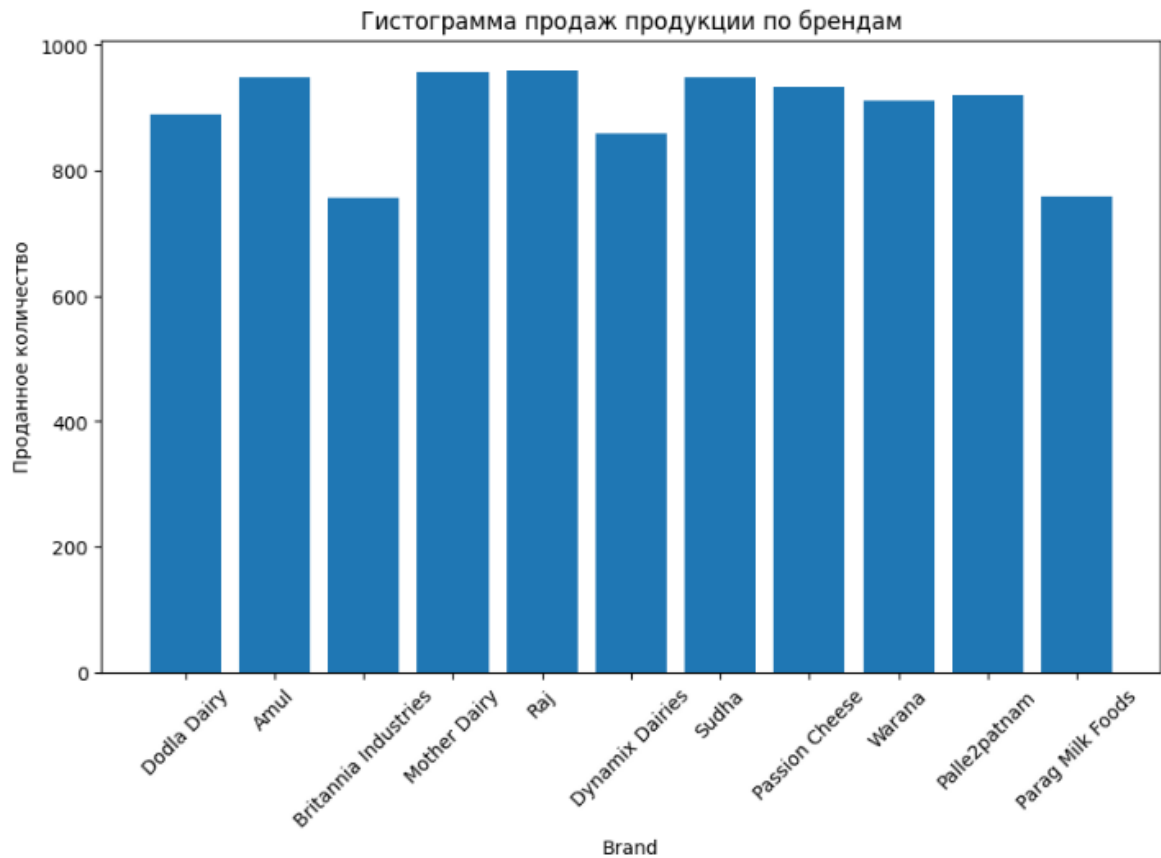
```
df_brand_counts = df['Brand'].value_counts()
plt.figure(figsize=(10, 6))
sns.barplot(x=df_brand_counts.index, y=df_brand_counts.values)
plt.xlabel('Brand')
plt.ylabel('Count')
plt.title('Количество молочных продуктов по брендам')
plt.xticks(rotation=45)
plt.show()
```



Вывод: график позволяет сравнить бренды, у которых объем продаж выше или ниже. Самые высокие продажи у бренда Amul. Самые низкие продажи Britannia Industries.

Строю гистограмму продаж продукции по брендам:

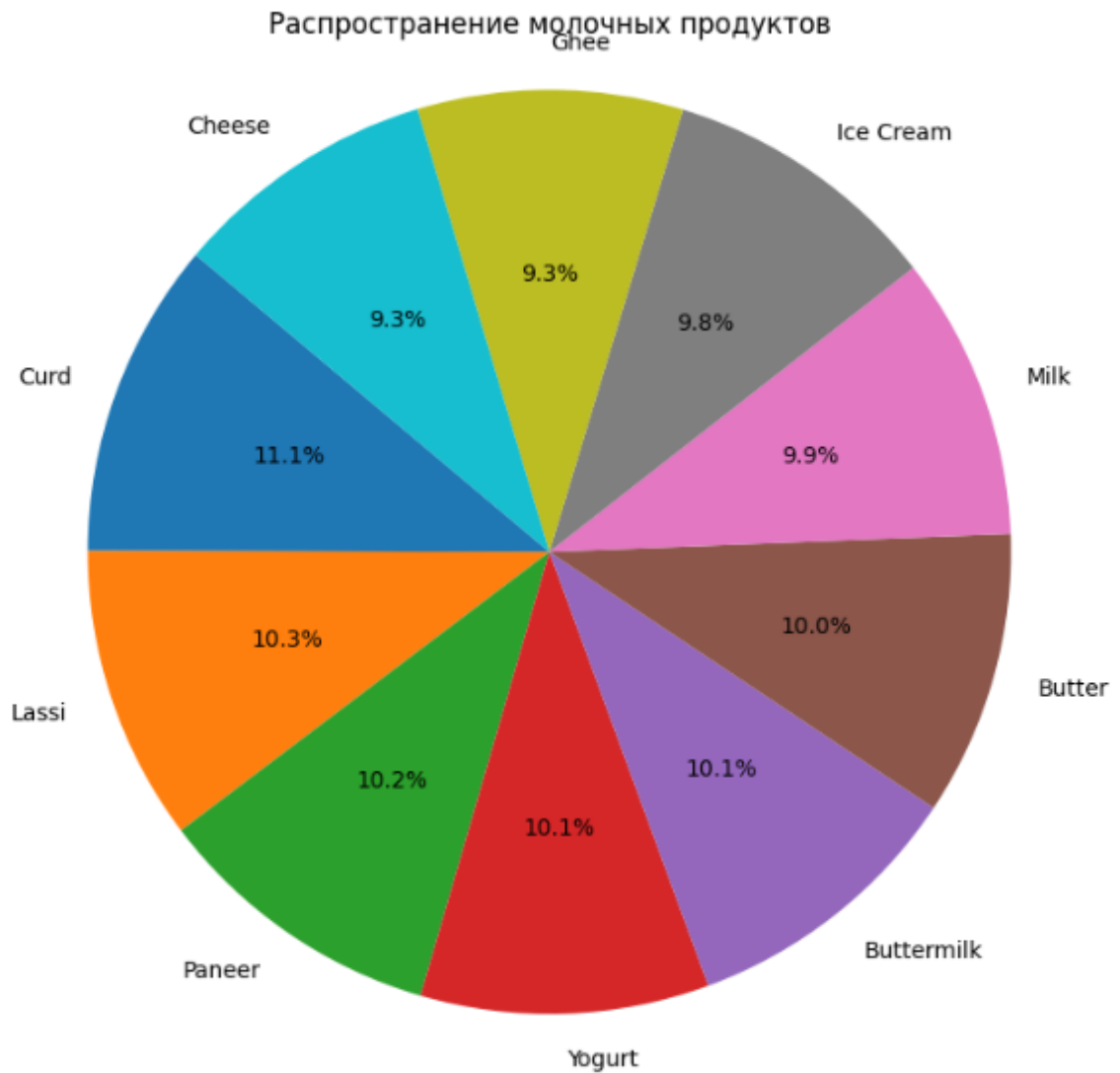
```
plt.figure(figsize=(10, 6))
plt.bar(df['Brand'], df['Quan_Sold'])
plt.xlabel('Brand')
plt.ylabel('Проданное количество')
plt.title('Гистограмма продаж продукции по брендам')
plt.xticks(rotation=45)
plt.show()
```



Вывод: топ самых популярных и продаваемых брендов Amul, Mother Dairy, Raj, Sudha. Топ типа продукции: Curd, Lassi, Paneer, Yogurt, Buttermilk.

Строим распределение молочных продуктов:

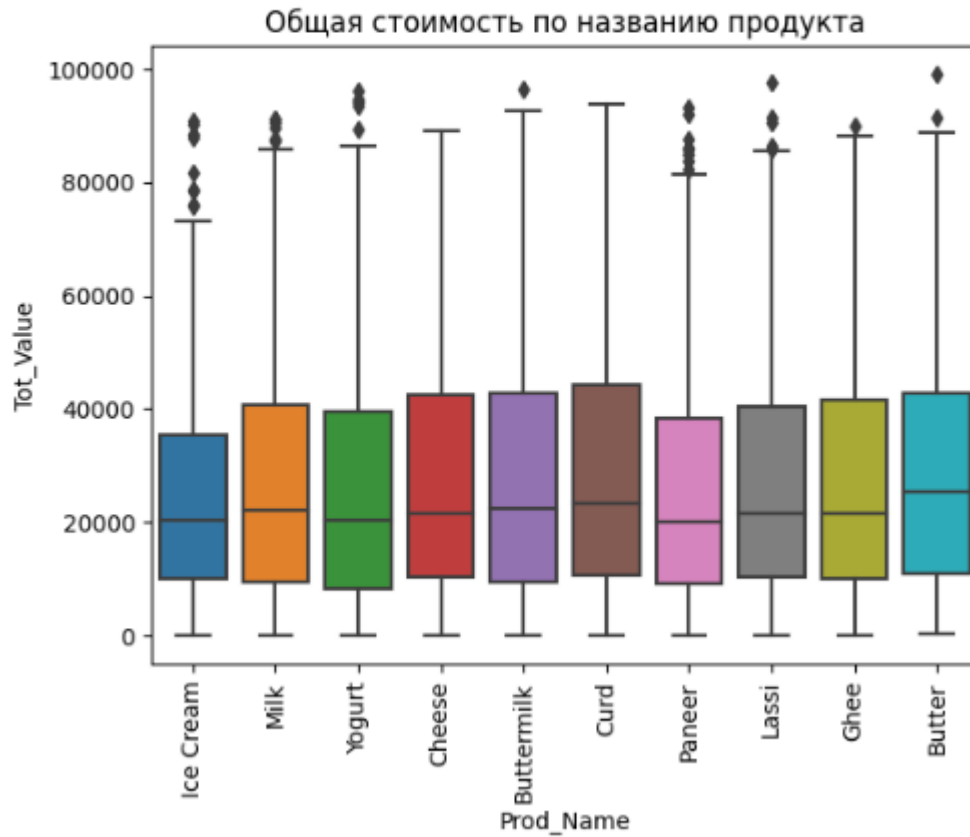
```
plt.figure(figsize=(8, 8))
plt.pie(df_product_counts, labels=df_product_names, autopct='%1.1f%%', startangle=140)
plt.title('Распространение молочных продуктов')
plt.axis('equal')
plt.show()
```



Вывод: среди всех молочных продуктов наибольшее распространение имеет творог.

Общая стоимость продукции:

```
sns.boxplot(data=df, x='Prod_Name', y='Tot_Value')  
plt.title('Общая стоимость по названию продукта')  
plt.xticks(rotation=90)  
plt.show()
```

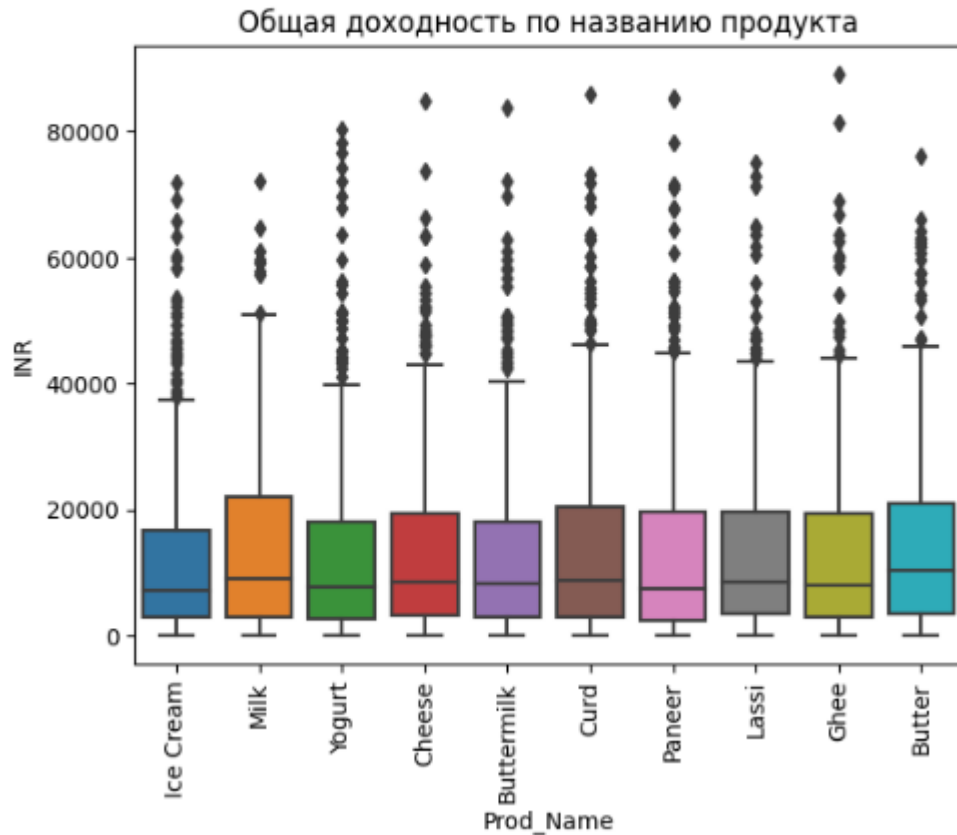


Вывод: наблюдаем большое количество выбросов, информацию необходимо дополнительно фильтровать, по общей стоимости продукции лидирует Curd, Buttermilk, Cheese.



Строю boxplot общая доходность по названию продукта:

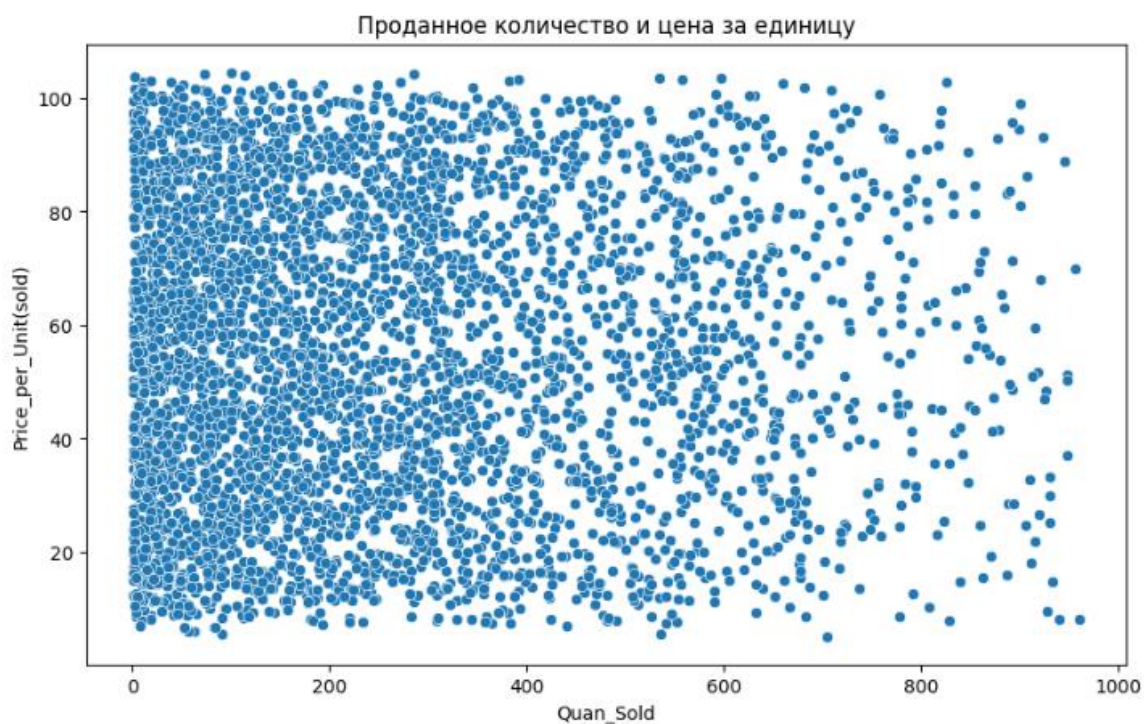
```
sns.boxplot(data=df, x='Prod_Name', y='INR')
plt.title('Общая доходность по названию продукта')
plt.xticks(rotation=90)
plt.show()
```



Вывод: наблюдаем большое количество выбросов, информацию необходимо дополнительно фильтровать, по общей доходности лидируют Milk, Curd, Butter.

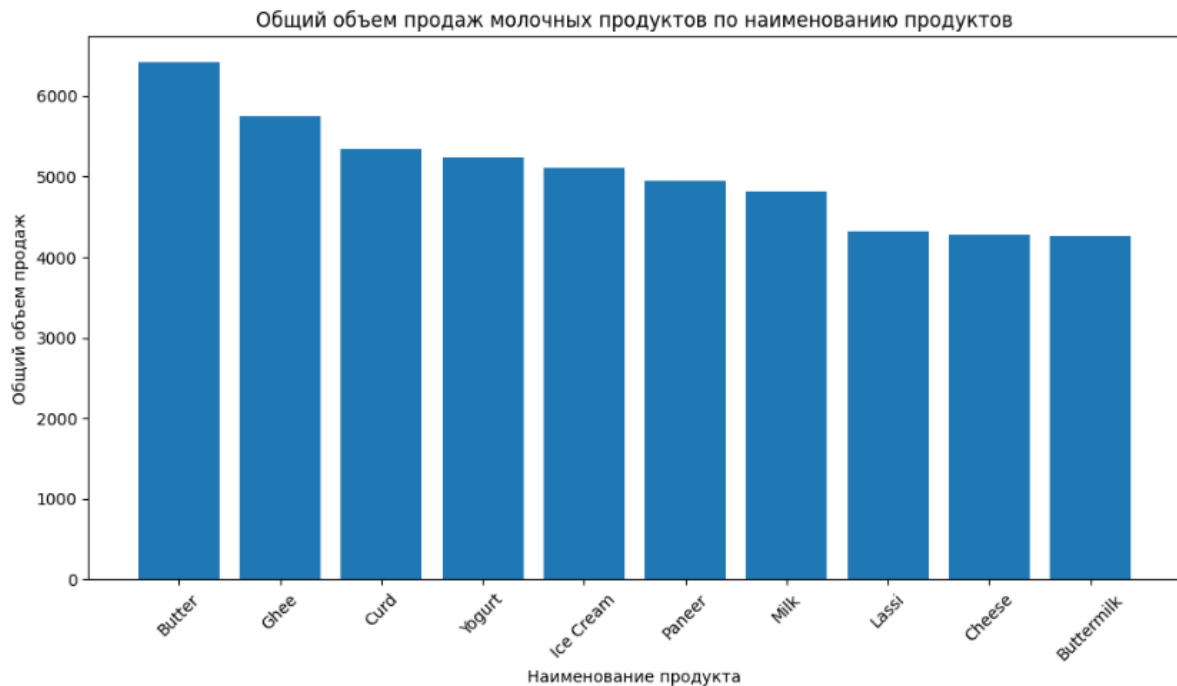
Произвожу анализ проданного количества и цену за единицу, группирую по количеству проданного продукта и фактической стоимости продукта.

```
plt.figure(figsize=(10, 6))
sns.scatterplot(x='Quan_Sold', y='Price_per_Unit(sold)', data=df)
plt.xlabel('Quan_Sold')
plt.ylabel('Price_per_Unit(sold)')
plt.title('Проданное количество и цена за единицу')
plt.show()
```



Вывод: Основная часть продукции продается в количестве от 1 до 200 единиц.

```
plt.figure(figsize=(12, 6))
df_product_sales_sorted = df_product_sales.sort_values('Quan_Sold', ascending=False)
plt.bar(df_product_sales_sorted['Prod_Name'], df_product_sales_sorted['Quan_Sold'])
plt.xlabel('Наименование продукта')
plt.ylabel('Общий объем продаж')
plt.title('Общий объем продаж молочных продуктов по наименованию продуктов')
plt.xticks(rotation=45)
plt.show()
```

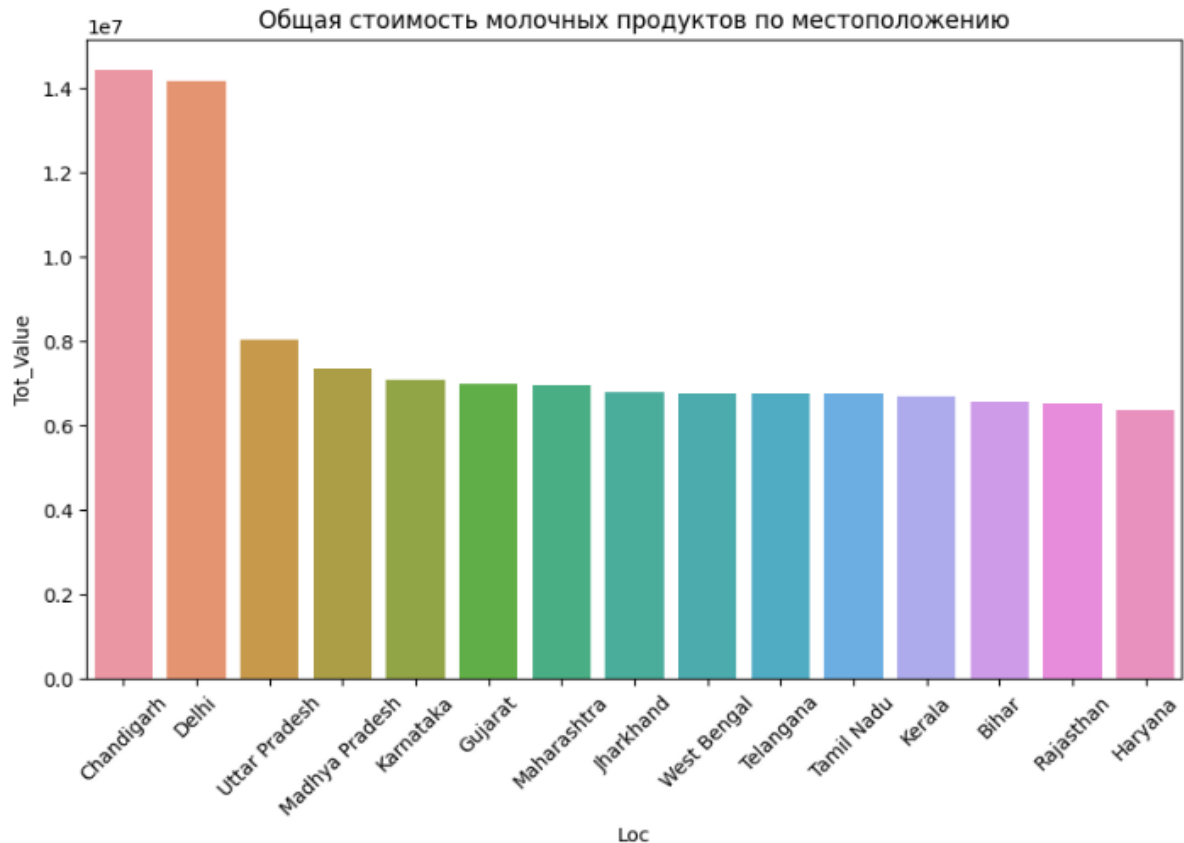


Вывод: по общему объему продаж лидирует Butter, Ghee, Curd, Yogurt.

Группирую данные по местоположению и рассчитаю сумму общей стоимости для каждого местоположения.

```
plt.figure(figsize=(10, 6))
sns.barplot(x='Loc', y='Tot_Value', data=df_location_tot_value)
plt.xlabel('Loc')
plt.ylabel('Tot_Value')
plt.title('Общая стоимость молочных продуктов по местоположению')
plt.xticks(rotation=45)
plt.show()
```

✓ 0.3s



Вывод: в Chandigarh, Delhi самая высокая общая стоимость молочных продуктов.

На этом этапе заканчиваем данный блок аналитики, переходим к другой части, каналы продаж и складские остатки.

## 2.6. Произвести анализ предпочтений и покупательского поведения клиентов в зависимости от местоположения и каналов продаж

В данном блоке, произведем анализ каналов продаж, определим потери в виде складских остатков и рассчитаем упущенную выгоду.

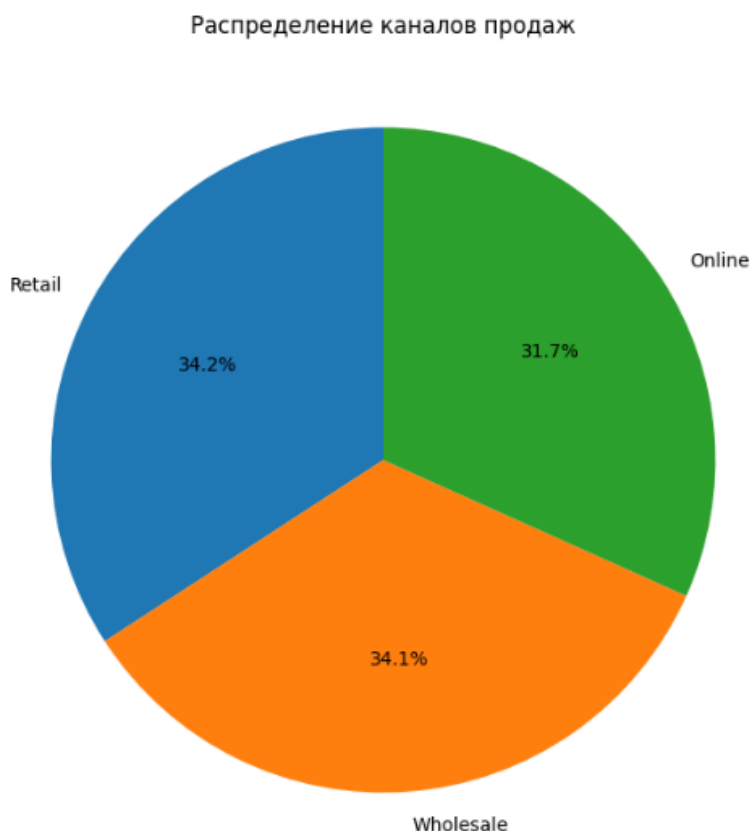
Формирую датасет, группировку, берем для анализа каналы продаж, строю распределение каналов продаж по типу.

```

df_sales_channel_counts = df['Sales_Channel'].value_counts()
plt.figure(figsize=(8, 8))
plt.pie(df_sales_channel_counts.values, labels=df_sales_channel_counts.index, autopct='%1.1f%%', startangle=90)
plt.title('Распределение каналов продаж')
plt.show()

```

✓ 0.1s

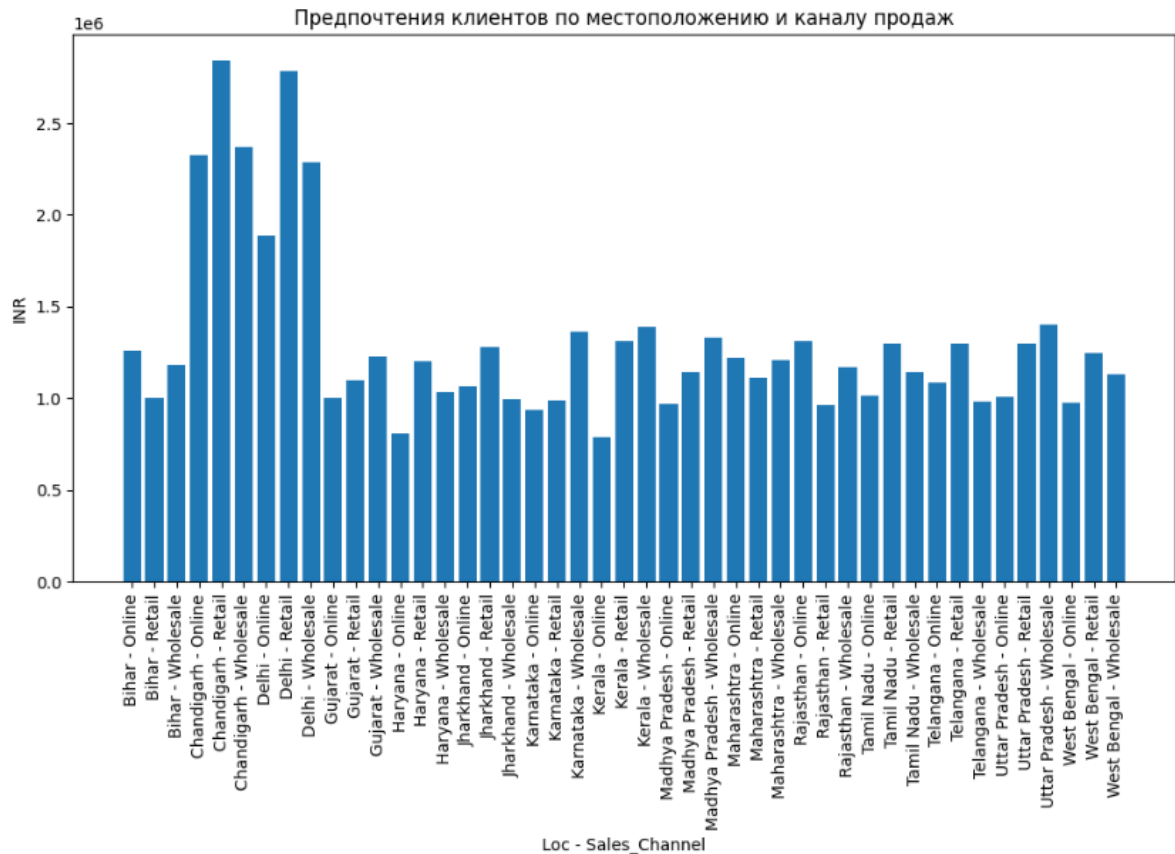


Вывод: большая часть товаров реализуется через розницу и оптом.

Рассмотрим предпочтение клиентов по местоположению и каналу продаж. Сгруппирую данные по «Местоположению» и «Каналу продаж» и рассчитаю сумму «Прибл. Общий доход (INR)» и «Проданное количество (литры/кг)».

```
plt.figure(figsize=(12, 6))
x = df_customer_preferences['Loc'] + ' - ' + df_customer_preferences['Sales_Channel']
plt.bar(x, df_customer_preferences['INR'])
plt.xlabel('Loc - Sales_Channel')
plt.ylabel('INR')
plt.title('Предпочтения клиентов по местоположению и каналу продаж')
plt.xticks(rotation=90)
plt.show()
```

✓ 0.5s



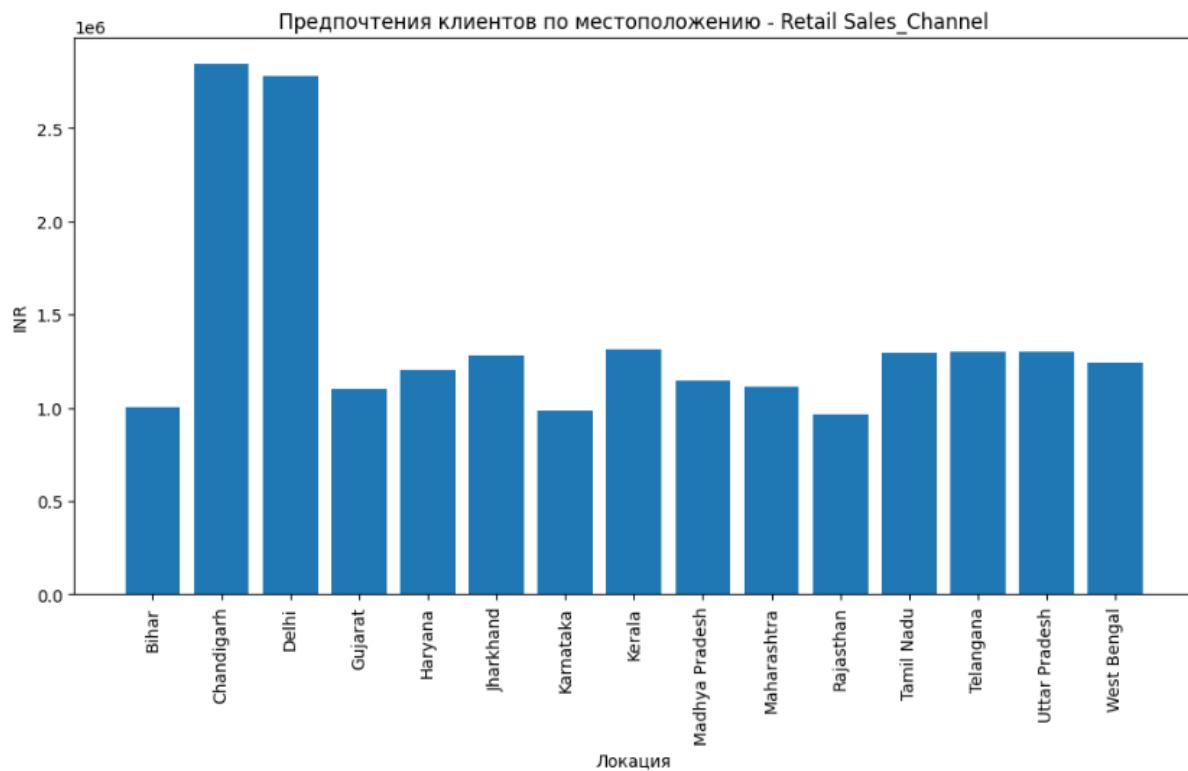
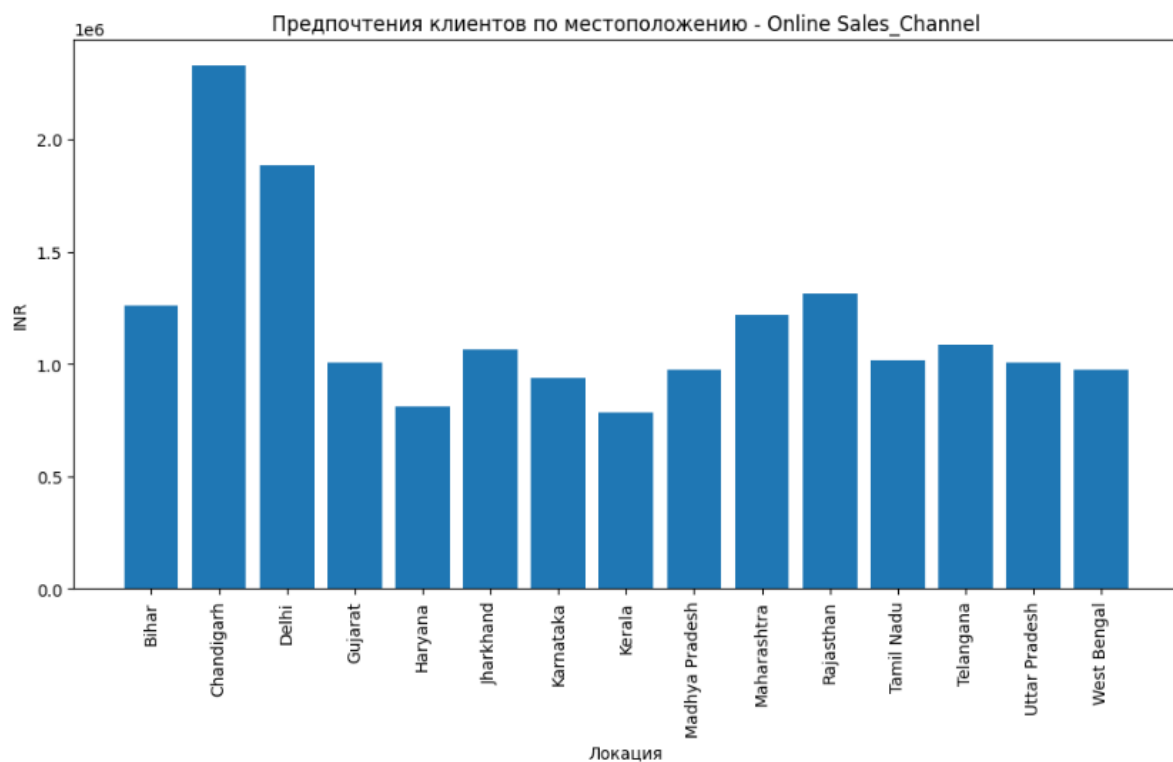
Вывод: основные предпочтения Chandigarh-Online, Chandigarh-Retail, Chandigarh-Wholesale, Delhi-Retail, Delhi-Wholesale.

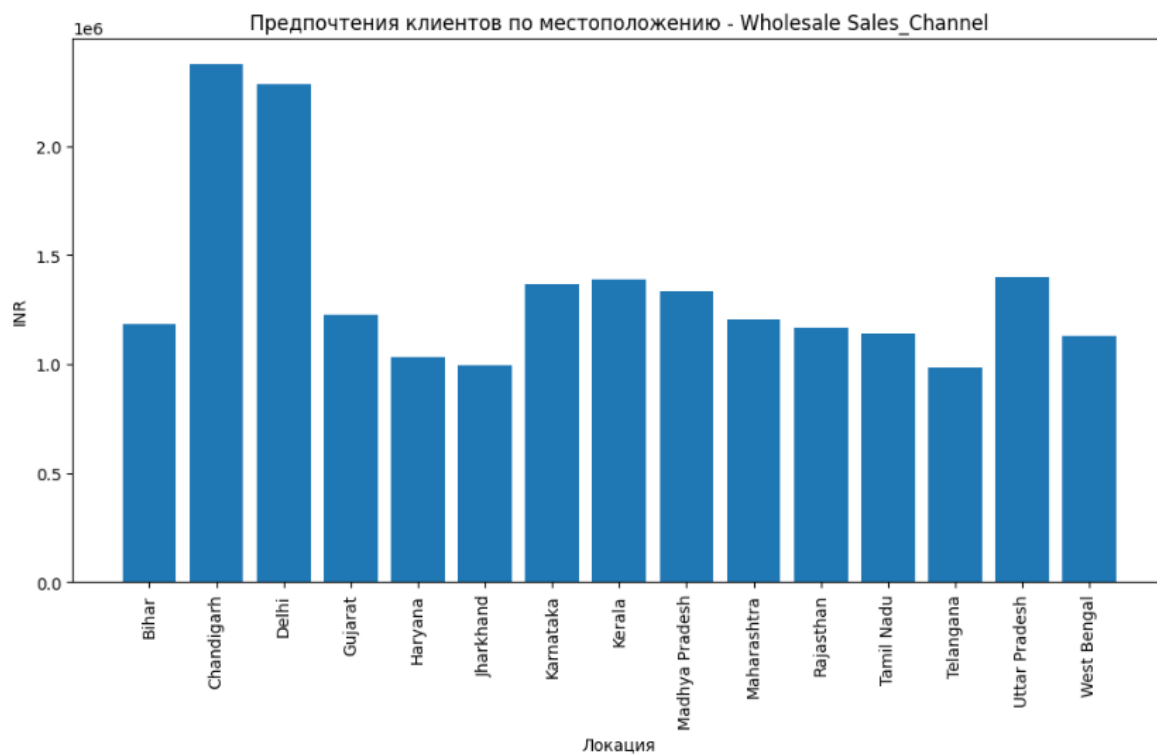
Строю гистограмму предпочтения клиентов по местоположению и типу канала продаж.

```

for sales_channel in df_sales_channels:
    plt.figure(figsize=(12, 6))
    df_sales_channel_data = df_customer_preferences[df_customer_preferences['Sales_Channel'] == sales_channel]
    x = df_sales_channel_data['Loc']
    plt.bar(x, df_sales_channel_data['INR'])
    plt.xlabel('Локация')
    plt.ylabel('INR')
    plt.title(f'Предпочтения клиентов по местоположению - {sales_channel} Sales_Channel')
    plt.xticks(rotation=90)
    plt.show()

```



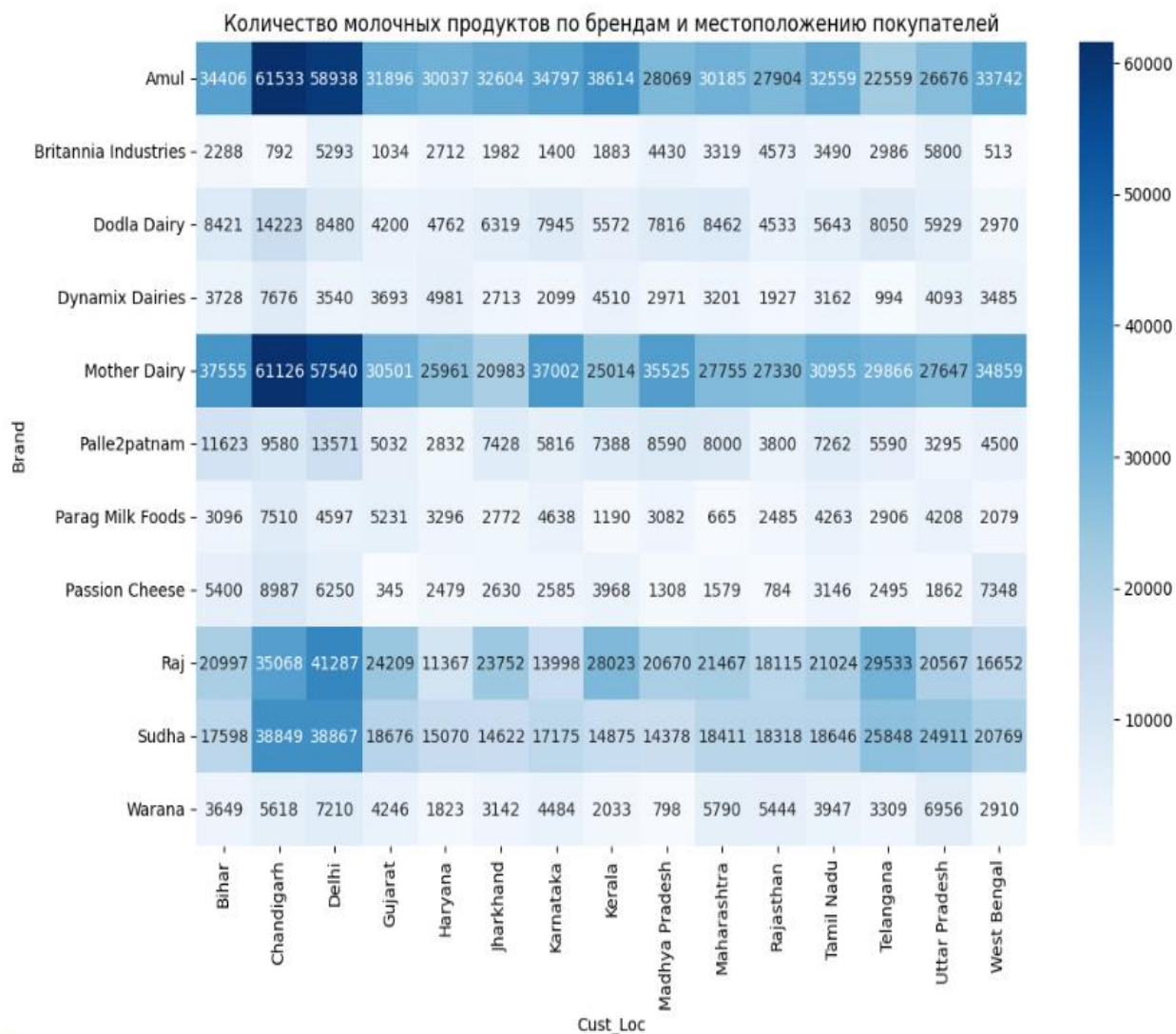


Вывод: подтверждаю выше сказанное, самые востребованные локации по типу продаж, это Chandigarh, Delhi.

Построим тепловую карту, количество молочных продуктов по брендам и местоположению покупателей.

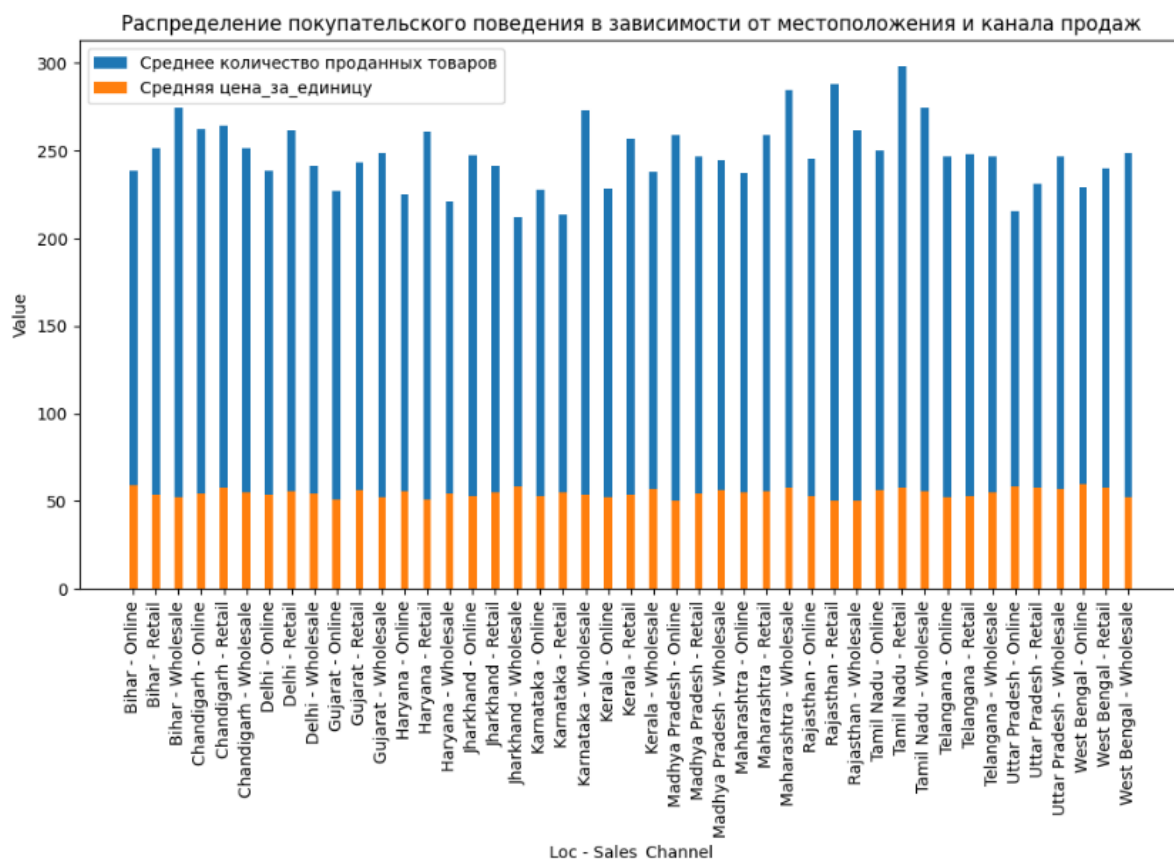


```
df_brand_customer_matrix = df.pivot_table(values='Quan', index='Brand', columns='Cust_Loc', aggfunc='sum')
plt.figure(figsize=(12, 8))
sns.heatmap(df_brand_customer_matrix, cmap='Blues', annot=True, fmt='.0f')
plt.xlabel('Cust_Loc')
plt.ylabel('Brand')
plt.title('Количество молочных продуктов по брендам и местоположению покупателей')
plt.show()
```



Формирую датасет локация, канал продаж, произведем группировку количество проданного молочного продукта, цена за единицу, по которой был продан молочный продукт. Произведем анализ покупательского поведения в зависимости от местоположения и канала продаж.

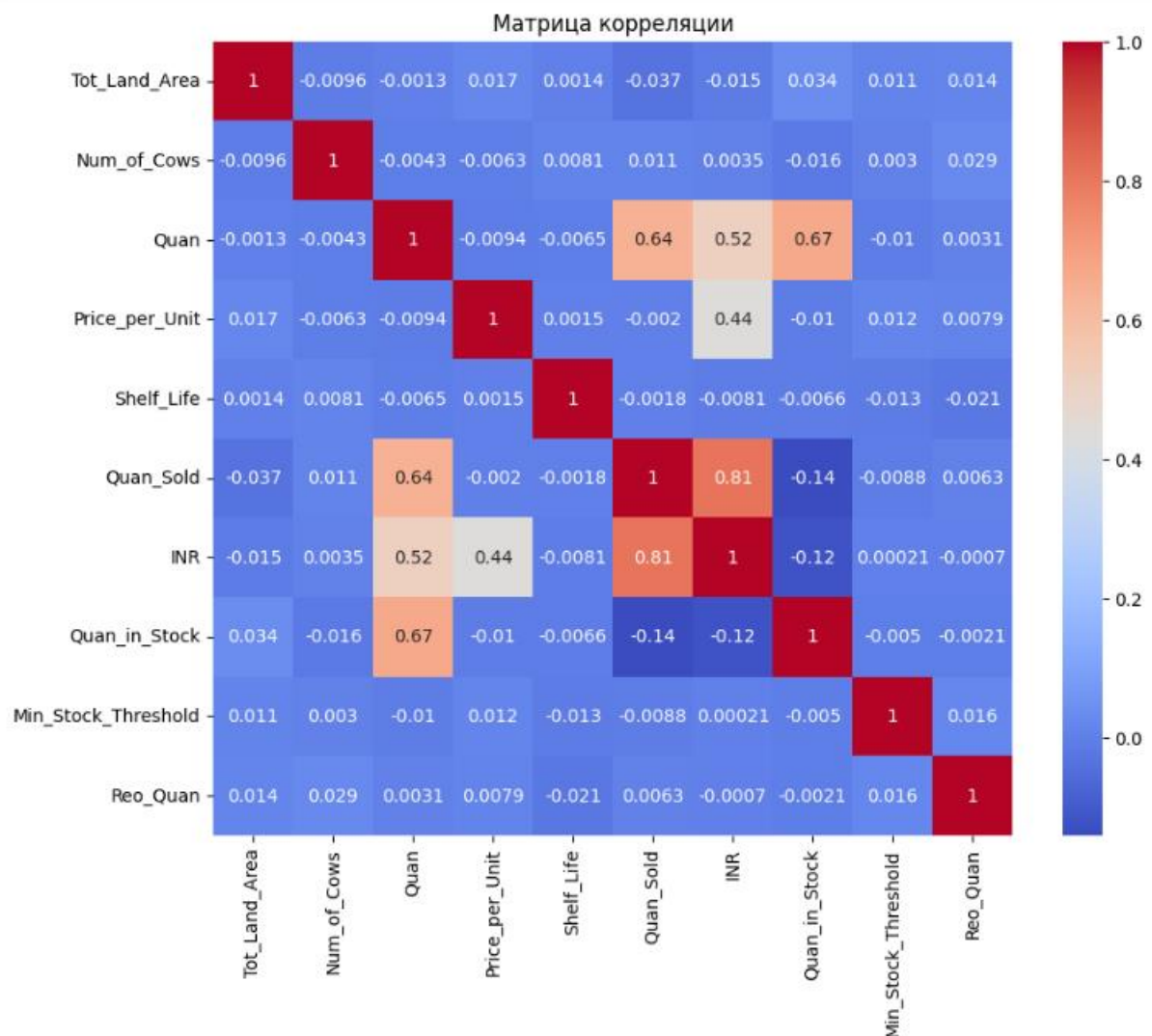
```
plt.figure(figsize=(12, 6))
width = 0.35
x = df_buying_behavior['Loc'] + ' - ' + df_buying_behavior['Sales_Channel']
plt.bar(x, df_buying_behavior['Quan_Sold'], width, label='Среднее количество проданных товаров')
plt.bar(x, df_buying_behavior['Price_per_Unit(sold)'], width, label='Средняя цена_за_единицу')
plt.xlabel('Loc - Sales_Channel')
plt.ylabel('Value')
plt.title('Распределение покупательского поведения в зависимости от местоположения и канала продаж')
plt.xticks(rotation=90)
plt.legend()
plt.show()
```



Вывод: канал розничных продаж в Tamil Nadu - retail имеет самый высокий средний объем продаж и одну из самых высоких среднюю цену за единицу, аналогична ситуация в Maharashtra-Wholesale, самые прибыльные локации. Самая высокая средняя цена за единицу: Bihar, Chandigarh, Gujarat, Haryana, Jharkhand, Kerala, Tamil Nadu, West Bengal - но необходимо увеличивать количество проданных товаров, потенциально узкое место, требуется модель продаж (корректировка), стратегическое планирование.

Построим корреляционную матрицу реквизитов (функций) датасета.

```
df_corr_matrix = df[['Tot_Land_Area', 'Num_of_Cows', 'Quan', 'Price_per_Unit',
                    'Shelf_Life', 'Quan_Sold', 'INR',
                    'Quan_in_Stock', 'Min_Stock_Threshold',
                    'Reo_Quan']].corr()
plt.figure(figsize=(10, 8))
sns.heatmap(df_corr_matrix, annot=True, cmap='coolwarm')
plt.title('Матрица корреляции')
plt.show()
```



Вывод: наиболее сильная корреляция наблюдается между приблизительным общим доходом (INR) и проданным количеством (литры/кг). А также между произведенным продуктом, проданным и остатками на складе.

Определим тип продукции и какой объем складского остатка.

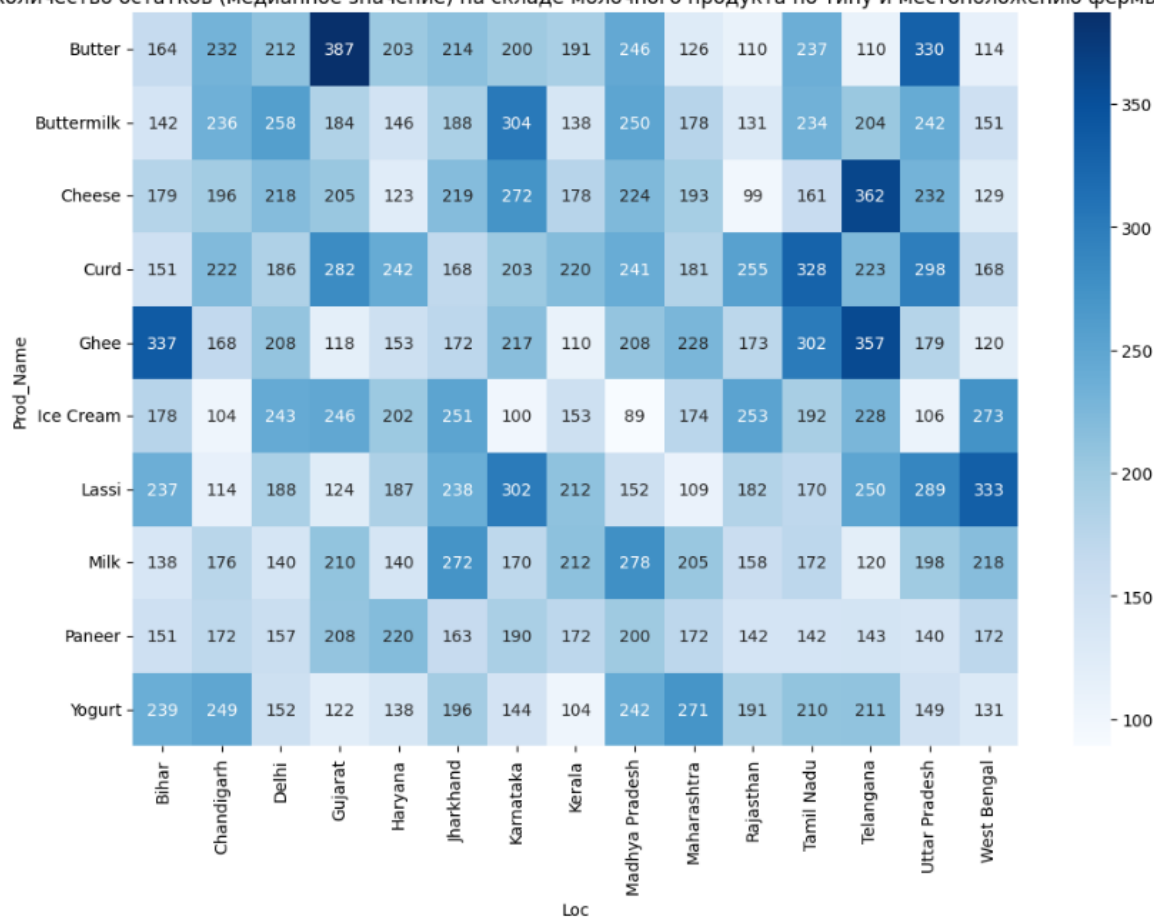
Группирую датасет, год, месяц, локация фермы, продуктовая группа. Агрегируем сумму количества доступного молочного продукта, сумму

количества проданного молочного продукта, медиана цены за единицу (по которой был продан молочный продукт), медиана количества молочного продукта, оставшегося на складе, медиана цены прайса за единицу молочного продукта, сумма общего дохода, полученный от продажи молочного продукта.

Строю тепловую карту.

```
df_stocks_matrix = df_stocks.pivot_table(values='Quan_in_Stock', index='Prod_Name', columns='Loc', aggfunc='median')
plt.figure(figsize=(12, 8))
sns.heatmap(df_stocks_matrix, cmap='Blues', annot=True, fmt='.0f')
plt.xlabel('Loc')
plt.ylabel('Prod_Name')
plt.title('Количество остатков (медианное значение) на складе молочного продукта по типу и местоположению фермы')
plt.show()
```

Количество остатков (медианное значение) на складе молочного продукта по типу и местоположению фермы



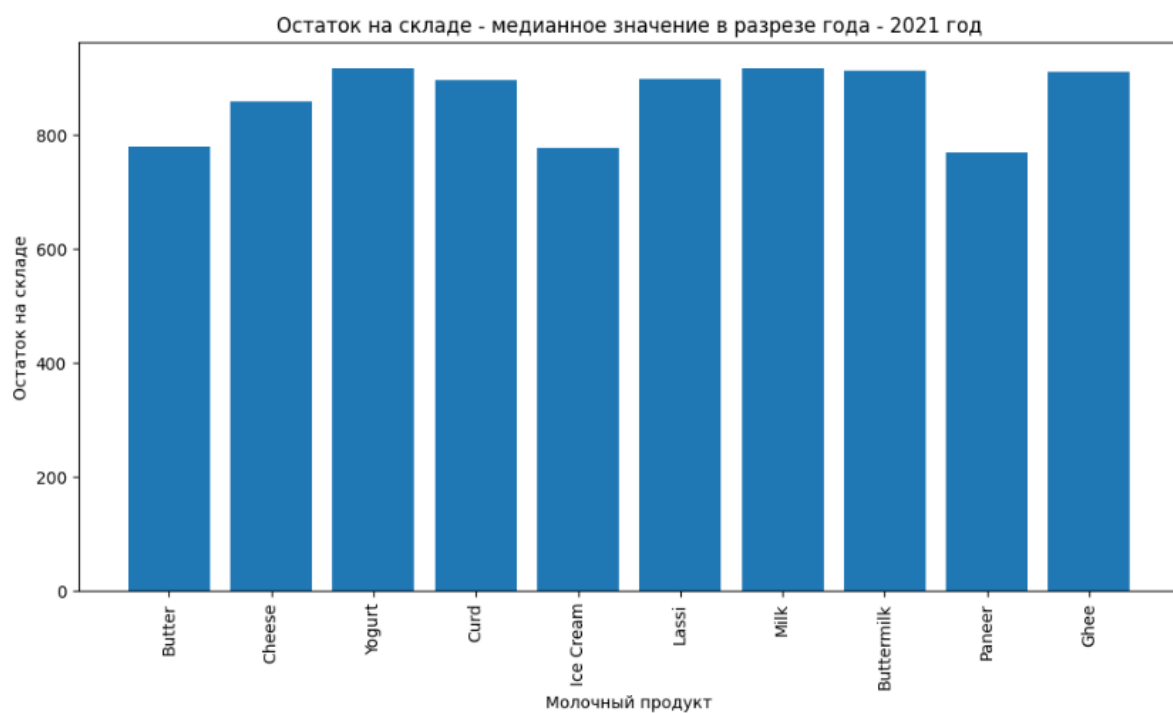
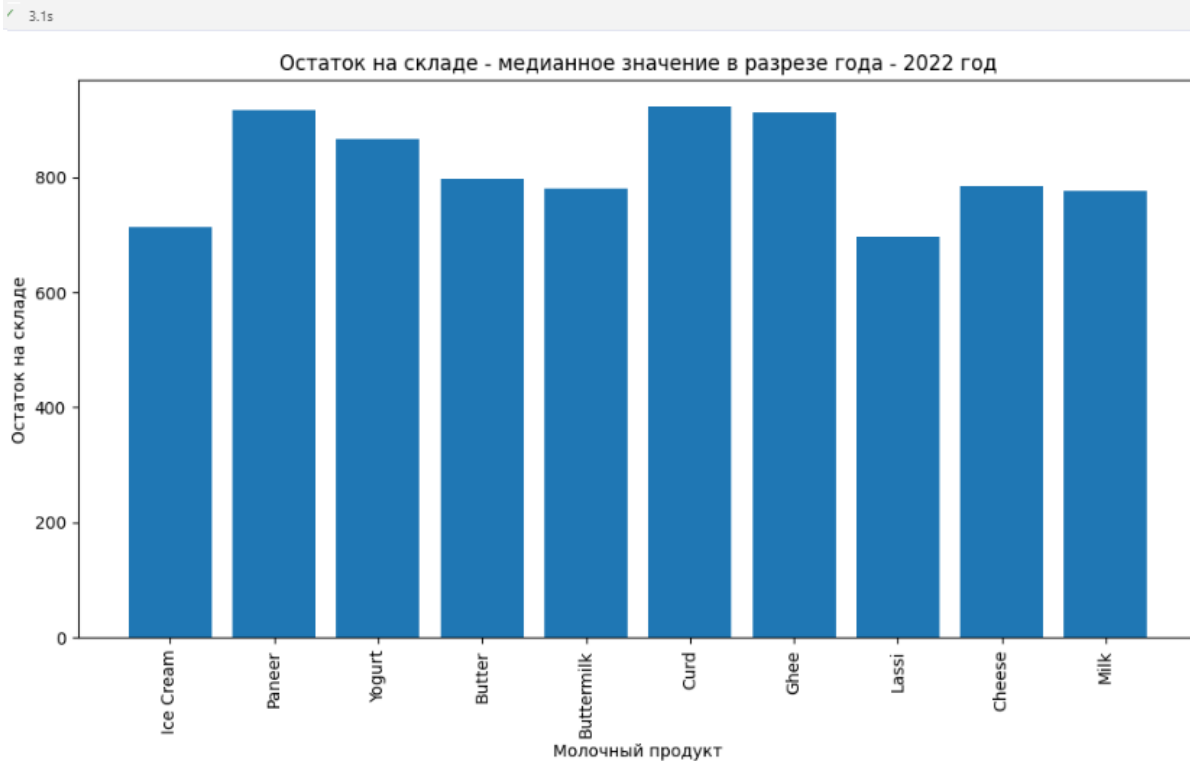
Произвел

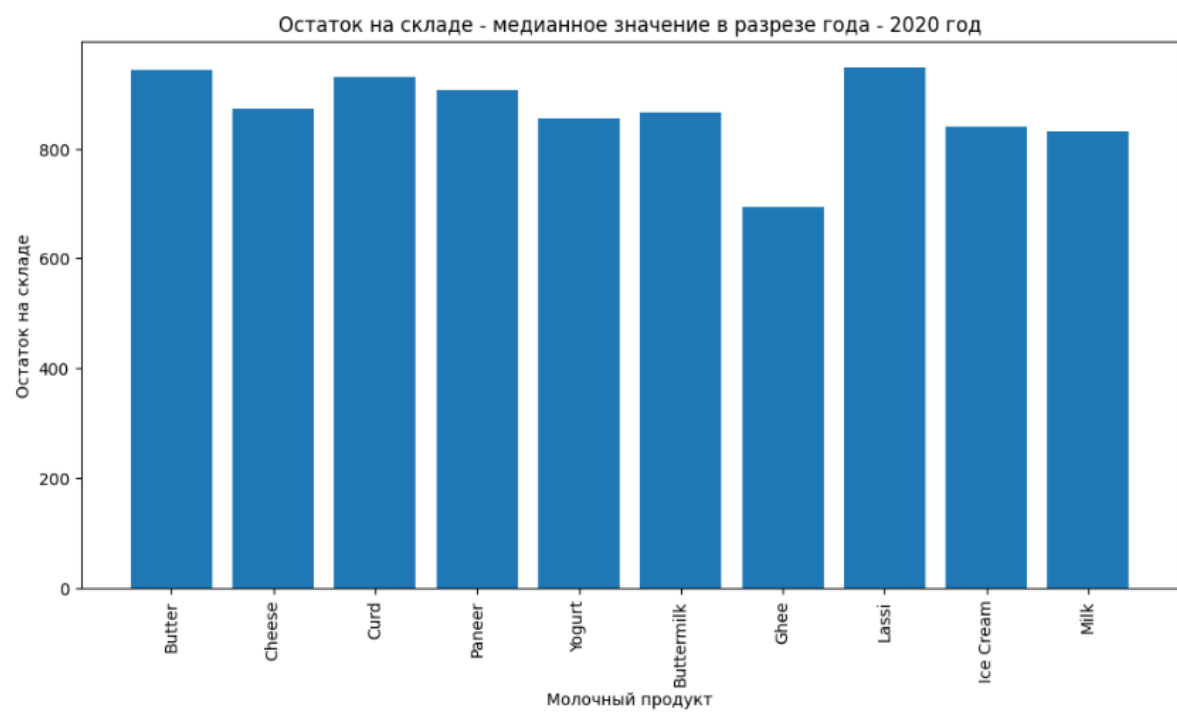
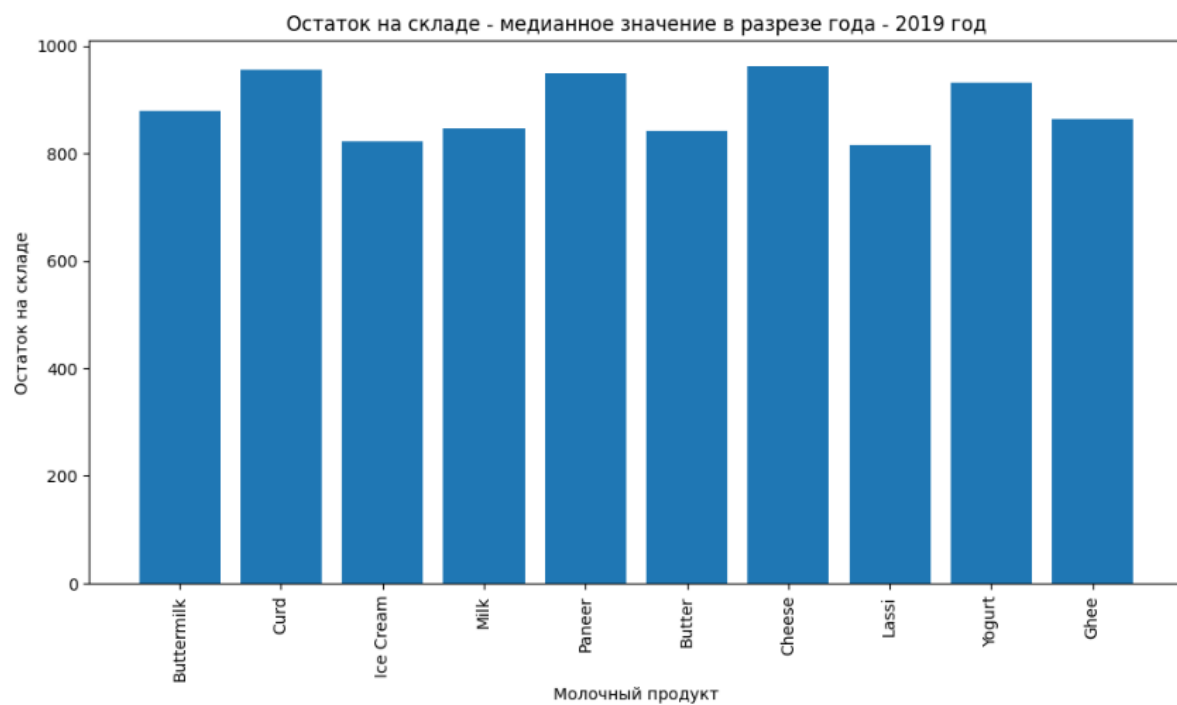
разбивку

по

годам:

```
for date_stocks_year in df_stocks_year:
    plt.figure(figsize=(12, 6))
    df_stocks_year_quan = Prod_Name_stocks[Prod_Name_stocks['Date_year'] == date_stocks_year]
    x = df_stocks_year_quan['Prod_Name']
    plt.bar(x, df_stocks_year_quan['Quan_in_Stock'])
    plt.xlabel('Молочный продукт')
    plt.ylabel('Остаток на складе')
    plt.title(f'Остаток на складе - медианное значение в разрезе года - {date_stocks_year} год')
    plt.xticks(rotation=90)
    plt.show()
```

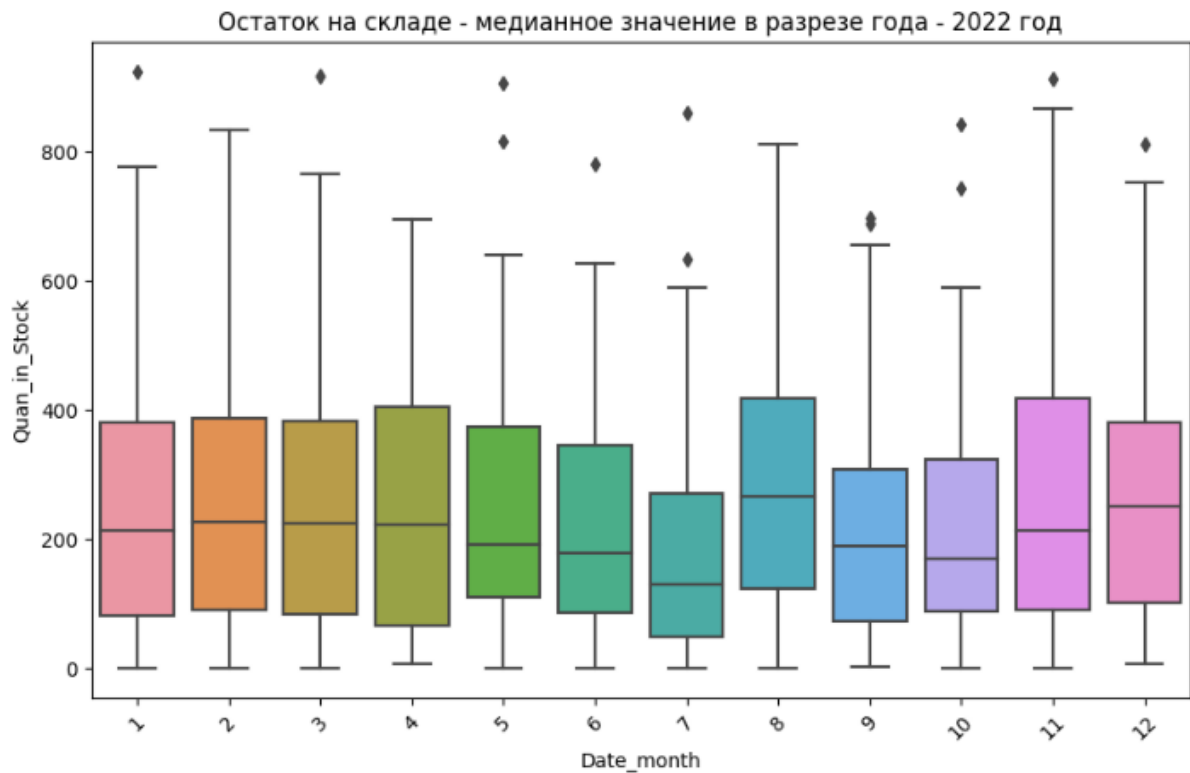


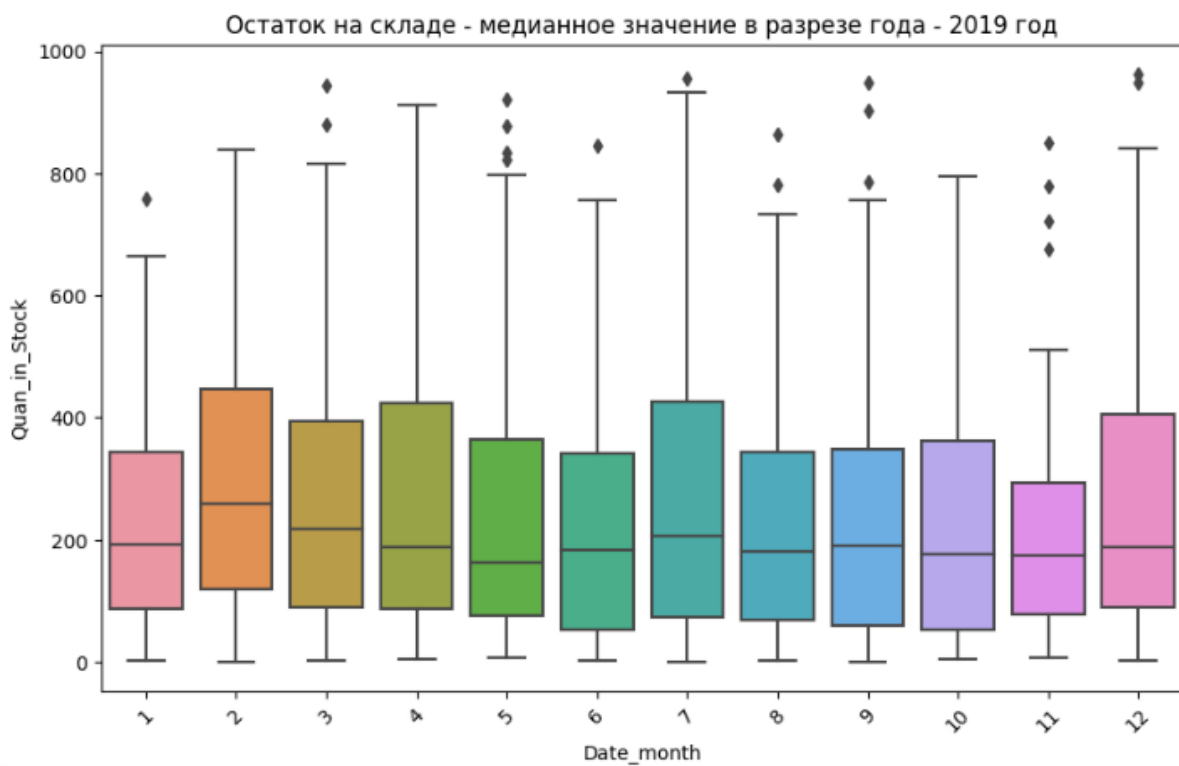
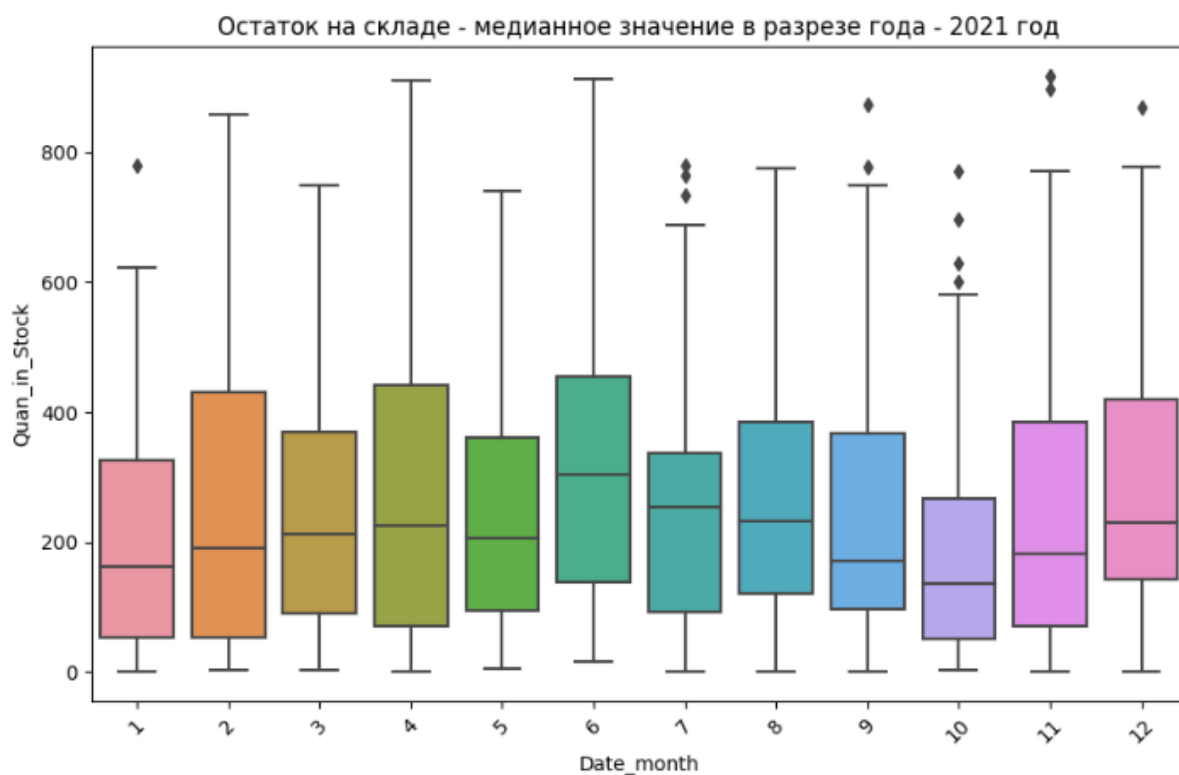


```

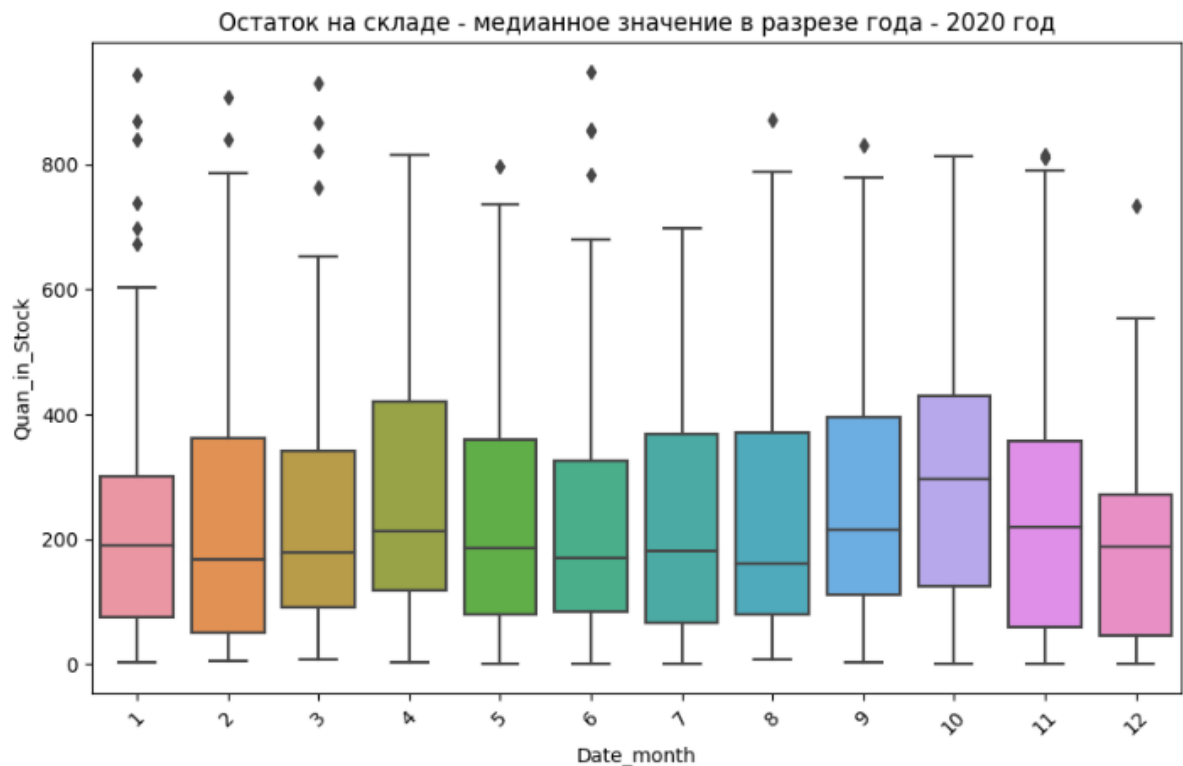
for date_stocks_year in df_stocks_year:
    plt.figure(figsize=(10, 6))
    df_stocks_year_quan = Prod_Name_stocks[Prod_Name_stocks['Date_year'] == date_stocks_year]
    x = df_stocks_year_quan['Date_month']
    sns.boxplot(x='Date_month', y='Quan_in_Stock', data=df_stocks_year_quan)
    plt.xlabel('Date_month')
    plt.ylabel('Quan_in_Stock')
    plt.title(f'Остаток на складе - медианное значение в разрезе года - {date_stocks_year} год')
    plt.xticks(rotation=45)
    plt.show()

```









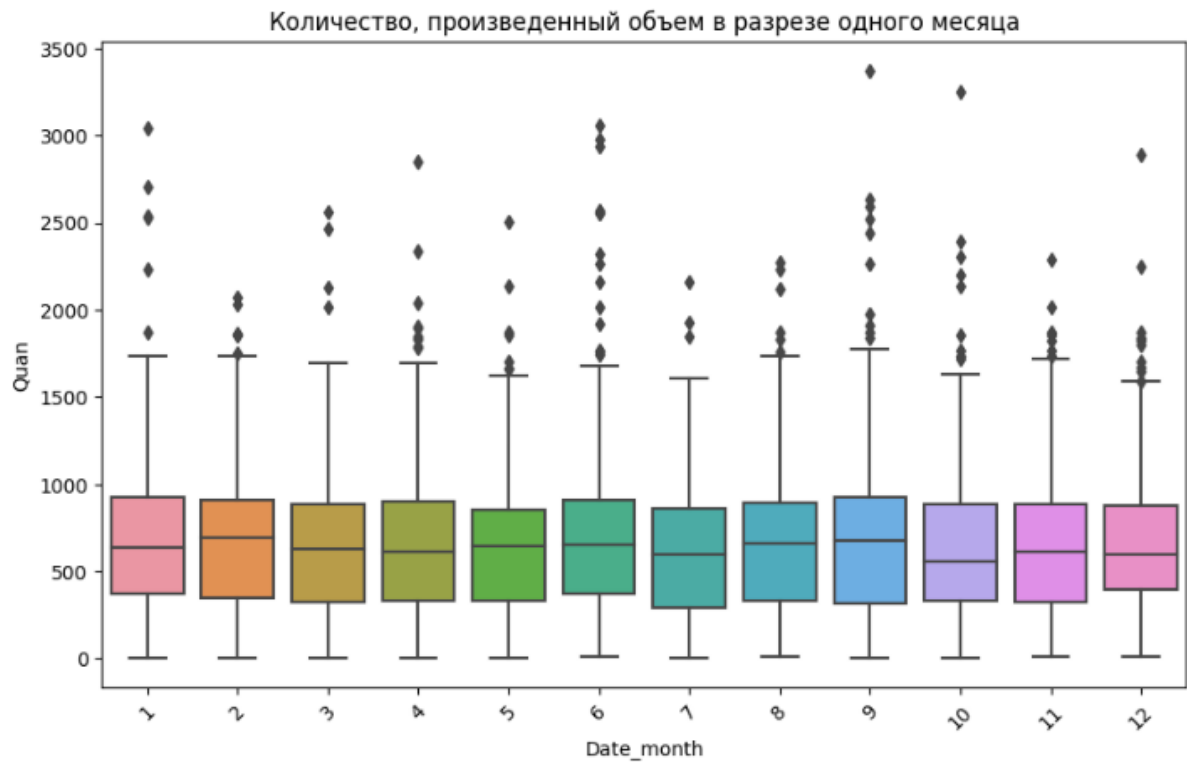
Вывод: каждый объем остатка на складе, с низким сроком хранения продукта, является потерей для компании/фермы, на тепловой карте видно, какая продуктовая группа и на какой локации остается в остатках производства. Требуется сфокусироваться на данных объектах и провести аудит процесса производства и скорректировать отдел продаж (сбыта), требуется корректировка оперативного планирования, требуется оптимизация модели продаж. Значение сильно превышает минимальный порог остатков и уровень для потенциала.

## 2.7. Изучить влияние условий хранения и сроков годности на качество и доступность молочной продукции

Беру для анализа, параметр условие хранения продукта. Формирую boxplot месяц и количество доступного молочного продукта.

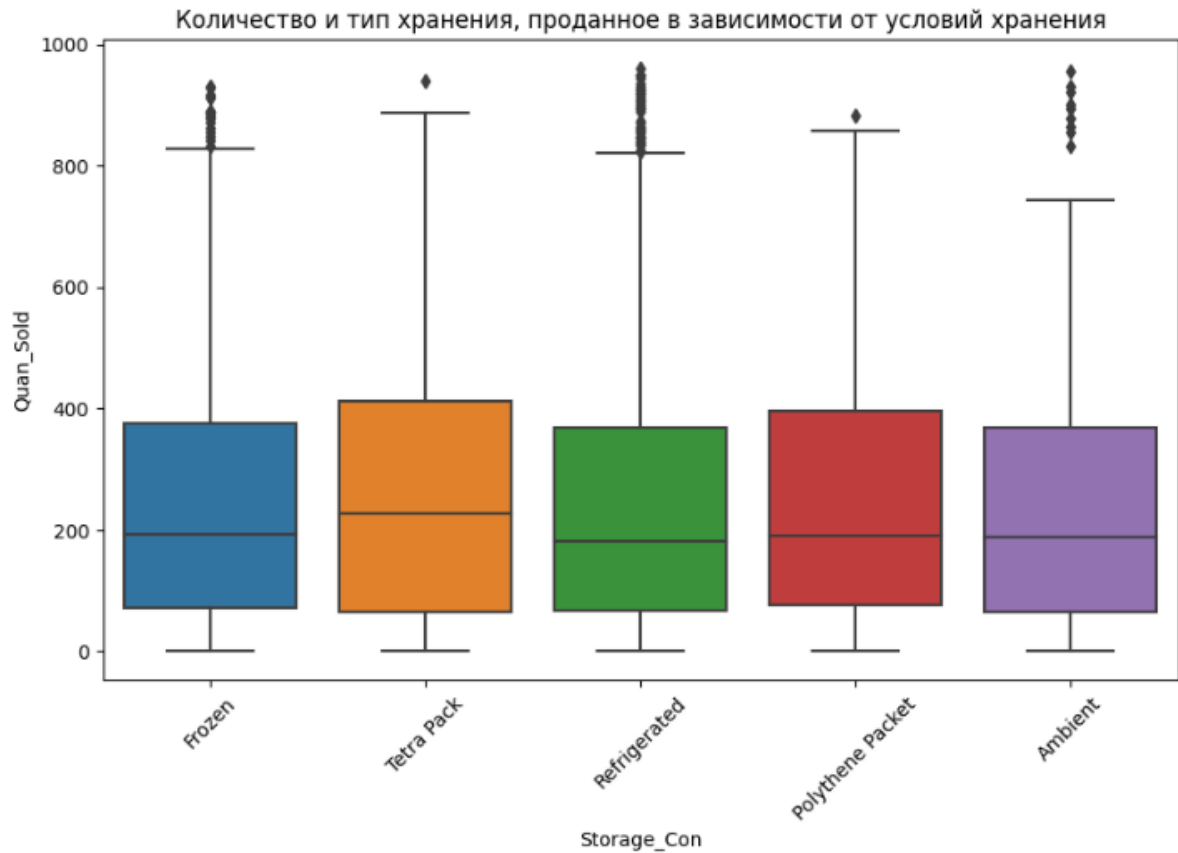
```
plt.figure(figsize=(10, 6))
sns.boxplot(x='Date_month', y='Quan', data=Prod_Name_stocks)
plt.xlabel('Date_month')
plt.ylabel('Quan')
plt.title('Количество, произведенный объем в разрезе одного месяца')
plt.xticks(rotation=45)
plt.show()
```

0.6s



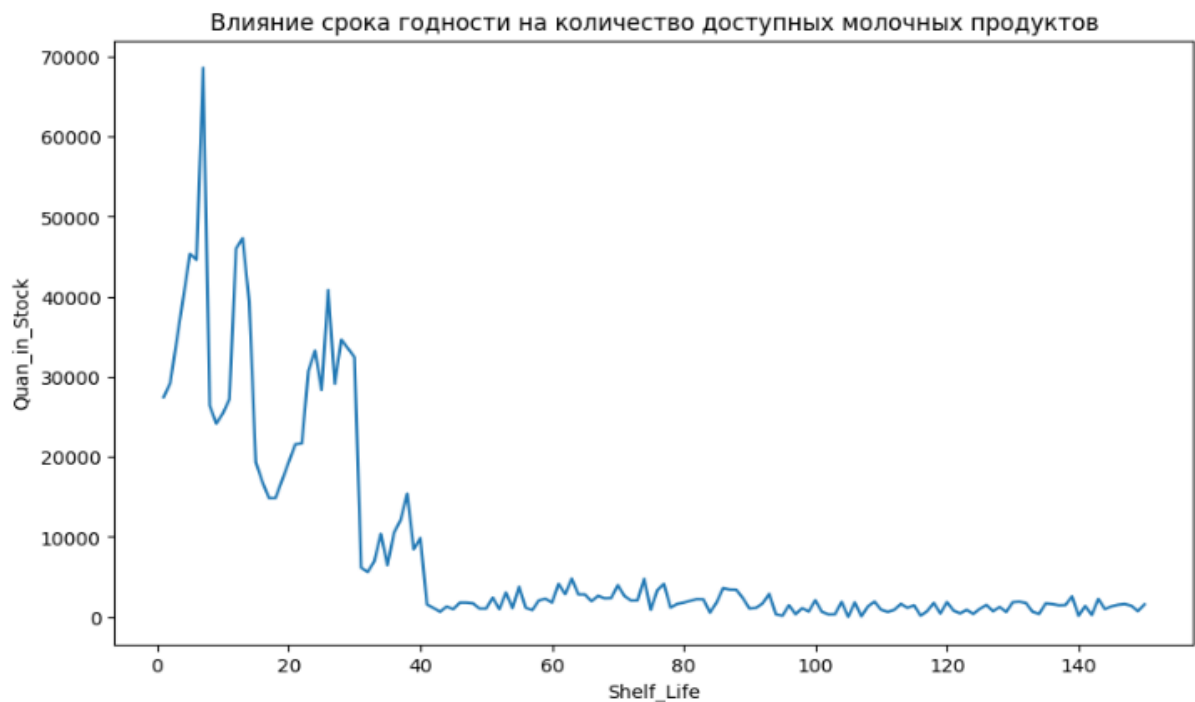
## Проданное количество в зависимости от условий хранения.

```
plt.figure(figsize=(10, 6))
sns.boxplot(x='Storage_Con', y='Quan_Sold', data=df)
plt.xlabel('Storage_Con')
plt.ylabel('Quan_Sold')
plt.title('Количество и тип хранения, проданное в зависимости от условий хранения')
plt.xticks(rotation=45)
plt.show()
```

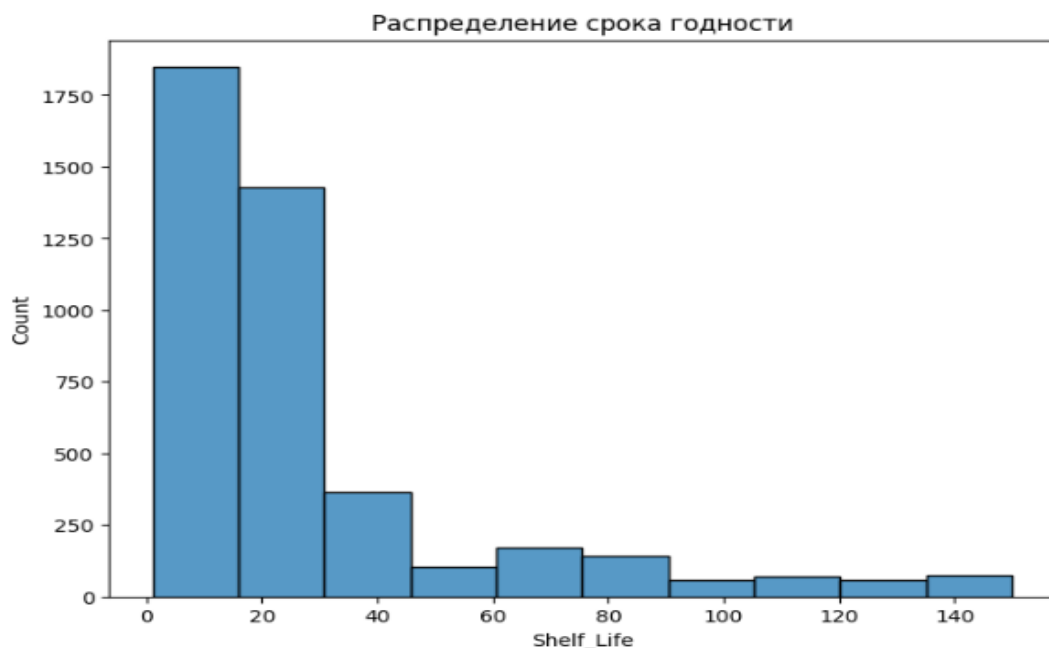


Строю распределение срока годности и влияние срока годности на количество доступных молочных продуктов. Сделаю анализ, влияние условий хранения на количество доступных молочных продуктов.

```
plt.figure(figsize=(10, 6))
plt.plot(df_shelf_life_impact['Shelf_Life'], df_shelf_life_impact['Quan_in_Stock'])
plt.xlabel('Shelf_Life')
plt.ylabel('Quan_in_Stock')
plt.title('Влияние срока годности на количество доступных молочных продуктов')
plt.show()
```



```
plt.figure(figsize=(8, 6))
sns.histplot(df['Shelf_Life'], bins=10)
plt.xlabel('Shelf_Life')
plt.ylabel('Count')
plt.title('Распределение срока годности')
plt.show()
```



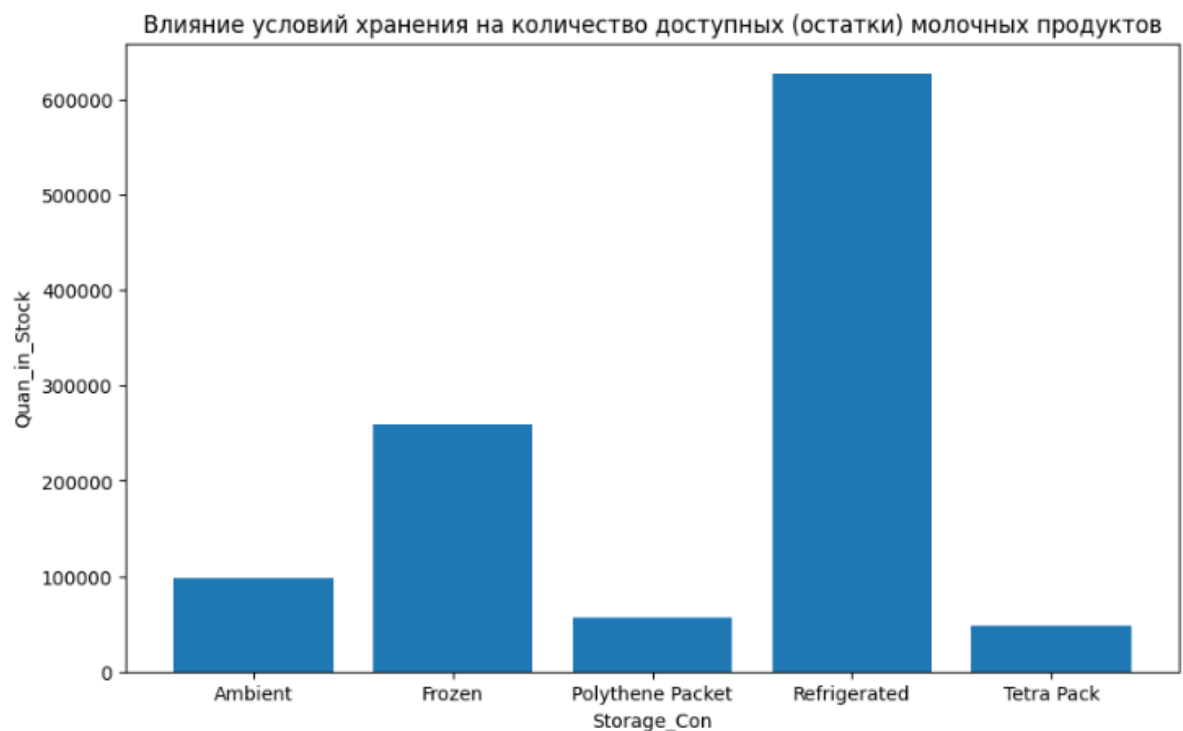
Вывод: получил представление об общем диапазоне срока годности и частоте различных значений срока годности. В основном, срок годности от 5 до 20 дней.

```

df_shelf_life_impact = df.groupby(['Shelf_Life']).agg({
    'Quan_in_Stock': 'sum'
}).reset_index()
plt.figure(figsize=(10, 6))
plt.bar(df_storage_impact['Storage_Con'], df_storage_impact['Quan_in_Stock'])
plt.xlabel('Storage_Con')
plt.ylabel('Quan_in_Stock')
plt.title('Влияние условий хранения на количество доступных (остатки) молочных продуктов')
plt.show()

```

✓ 0.1s

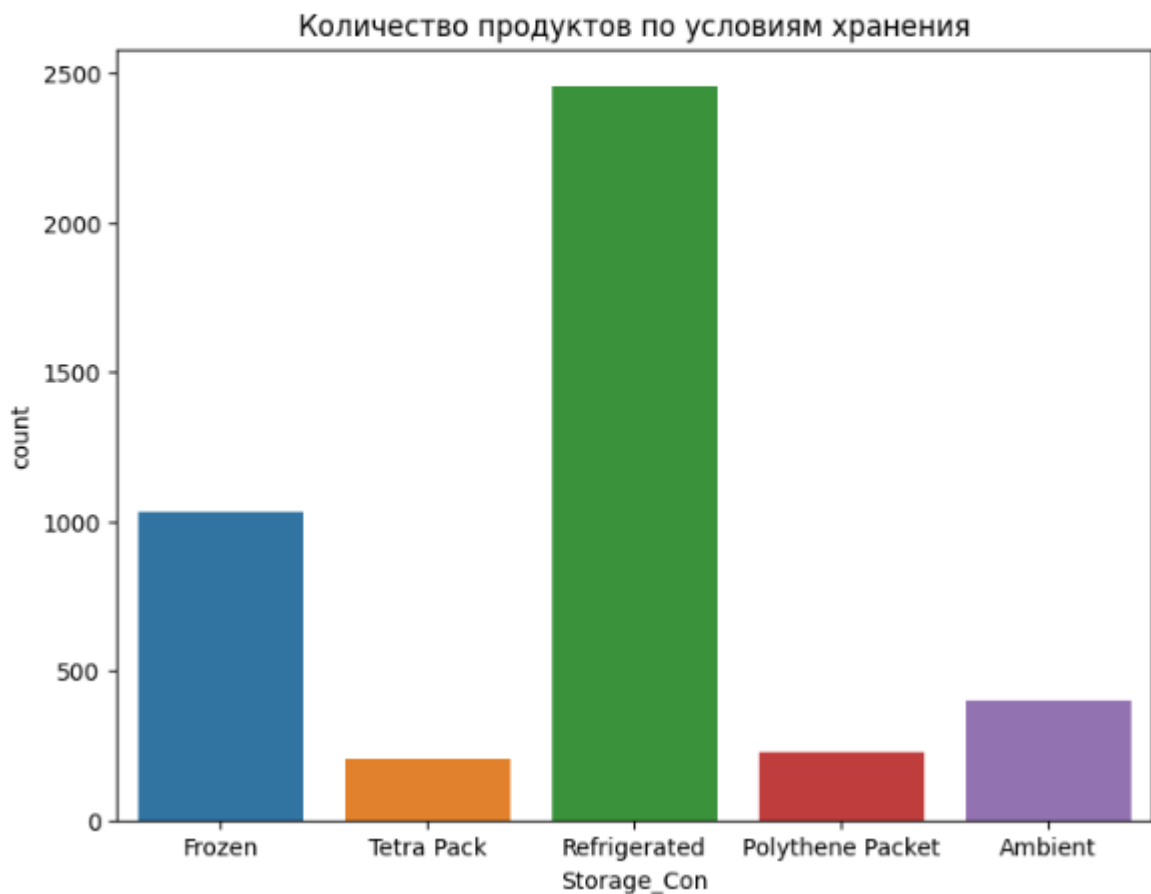


Вывод: самые перспективные варианты хранения, это заморозка и холодильник.

Произвел группировку количества продуктов по условиям хранения.

```
plt.figure(figsize=(8, 6))
sns.countplot(x='Storage_Con', data=df)
plt.title('Количество продуктов по условиям хранения')
plt.show()
```

✓ 0.1s



Вывод: основной тип хранения, который позволяет максимально долго хранить продукт, это холодильник и морозилка (в замороженном виде).

## 2.8. Провести ABC анализ продаж

Возьмем самый прибыльный год и проведем многомерный ABC - анализ продаж, по продуктовой группе.

ABC-анализ — это способ классификации ресурсов по степени их влияния на прибыль компании. Чаще всего метод ABC применяют к товарному ассортименту, чтобы понять рентабельность групп товаров или отдельных

наименований и скорректировать ассортимент. Поэтому дальше мы будем рассказывать о методе именно в контексте анализа продуктовой группы.

Формирую датасет `df_abc`, в котором фиксирую следующие столбцы: год, бренд, общая доходность, количество проданного молочного продукта, название молочного продукта, стоимость прайс и фактическая стоимость продажи.

Так как самый прибыльный год 2019, берем его для анализа. Группирую продуктовую группу по данному периоду, произвожу агрегацию общей доходности, количества проданного молочного продукта, бренд, медиану значения стоимости по прайсу и медиану фактической стоимости продажи, сортирую по количеству проданного молочного продукта. В результате возможно посмотреть разницу фактической стоимости от стоимости по прайсу. Посмотрим разницу фактической стоимости от прайса, для начала работы с моделью продаж, необходимо посмотреть значения отклонений. Смотрим медианные значения. Вывел процентное отклонение факта от прайса в столбце «`Raznic%`». Видно, что продажи осуществлялись как с завышенной, так и с заниженной стоимостью. В рамках 2019 года наблюдаем следующее, разница между стоимостью прайс и фактической стоимостью продаж, незначительная, что скорей всего и вывело этот год в лидеры. После 2019 необходимо было производить ABC-анализ, формировать стратегическую модель планирования, определить, где возможно поднять стоимость, поработать с отрицательными значениями, сформировать корректный прайс лист, скорректировать объем производства. В любом случае необходимо делать ABC-анализ, инструмент позволяет вносить своевременные корректировки.

Добавим в датасет столбец выгода/маржа `profit`, так как из источника у нас такой информации нет, сделаем значения рандомно в диапазоне до 30, что будет равна % маржинальности.

Выбираю характеристику для анализа. Считаю относительный вклад каждой позиции в общий «котел». Сортирую по убыванию величины вклада.

Нахожу накопительный итог вклада. Разбиваю на группы А (накопительный итог до 80%), Б (накопительный итог до 95%) и С (накопительный итог до 100%). После делаю тоже самое, только по всем 3 характеристикам - количество, общий доход, % доходности.

Результат, сформирован список по продуктовым группам и количеству в каждой группе в период 2019 года, подробный код анализа добавил в приложения:

groups			
	final_group	count	perc
0	AAA	185	0.166817
1	AAB	95	0.085663
2	AAC	96	0.086564
3	ABA	26	0.023445
4	ABB	13	0.011722
5	ABC	9	0.008115
6	BAA	50	0.045086
7	BAB	17	0.015329
8	BAC	27	0.024346
9	BBA	94	0.084761
10	BBB	25	0.022543
11	BBC	33	0.029757
12	BCA	22	0.019838
13	BCB	6	0.005410
14	BCC	7	0.006312
15	CAA	9	0.008115
16	CAB	3	0.002705
17	CAC	7	0.006312
18	CBA	39	0.035167
19	CBB	15	0.013526
20	CBC	8	0.007214
21	CCA	165	0.148783
22	CCB	72	0.064923
23	CCC	86	0.077547

Бизнес выводы ABC-анализа продаж:

Группа А – самые важные ресурсы, приносят максимальную прибыль и продажи. Компания будет нести большие потери при резком снижении эффективности данной группы ресурсов, а следовательно, ресурсы группы А должны жестко контролироваться, четко прогнозироваться, часто мониториться, быть максимально конкурентоспособными и не терять свои сильные стороны.



На данную группу ресурсов должны быть выделены максимальные инвестиции, лучшие ресурсы. Успехи группы А должны быть проанализированы и максимально транслироваться на другие категории.

Группа В – группа ресурсов, которые обеспечивают хорошие стабильные продажи/ прибыль компании. Данные ресурсы также важны для компании, но могут модерироваться более спокойными и умеренными темпами. Данные ресурсы относительно стабильны в краткосрочной перспективе. Инвестиции в данный вид ресурсов компании не значительны и необходимы только для поддержания существующего уровня.

Группа С – наименее важная группа в компании. Обычно ресурсы группы С тянут компанию вниз или не приносят дохода. При анализе данной группы необходимо быть очень внимательным и в первую очередь понять причину низкого вклада.

Дополнительно:

Группа ААА - посмотреть XYZ-анализ, чтобы принять решение о поднятии цены и увеличение объема производства. Если группа Х - можно поднимать цену на короткий период, чтобы увеличить маржинальность. Тестем в течение небольшого срока, чтобы оценить динамику продаж. Если группа Y/Z, лучше не увеличивать цену, потому что сложно будет оценить влияние. Лучше понять, почему спрос нестабильный. И либо поднять цену, но аккуратно спланировать эксперимент на более длинный срок, либо не повышать цену.

Группа ССА - оценить объективность причин, часть вывести из ассортимента, попробовать перевести в группу повыше по количеству (а может и по выручке за счет этого). Объективность причин - проверить качество, не просроченный ли он и др.

Подтянуть в группу В по количеству можно за счет акции 1+1 или 2+1 даже, как вариант. Можно увеличить средний чек и количество продаж за счет такой акции, стимулируя покупать больше и др. варианты.

Группа ААВ - нужно просто найти более выгодную цену поставки, чтобы стать ААА. Либо попробовать немного подтянуть цену, если это группа Х по XYZ-анализу.

Группа АВА - судя по всему, нужно увеличить количество продаж, не упав в прибыли, плюс реализовывать объем производства, сокращать складские остатки. Можно попробовать постепенно, сделать акцию формата 1+1 или другой вариант акции. Ищем другого поставщика или сбиваем цену у текущего, чтобы увеличить прибыль.

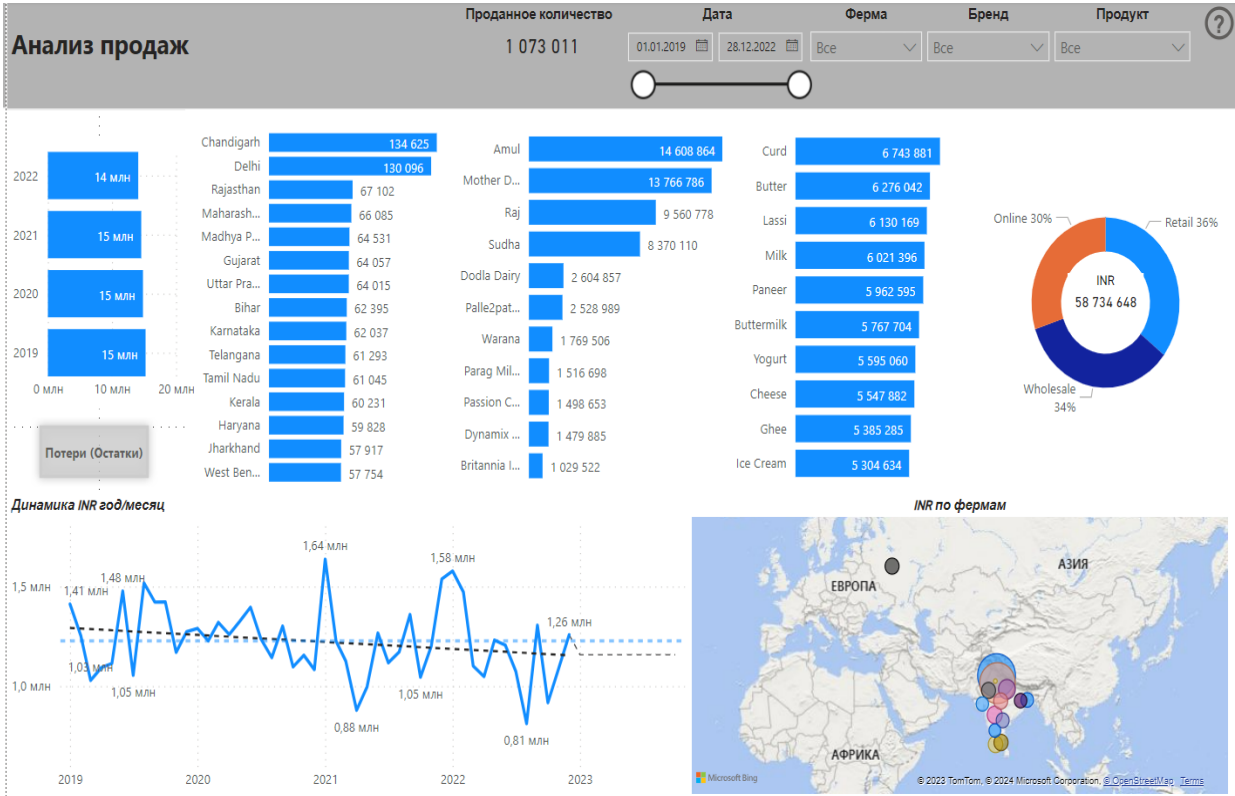
### **Глава 3. Описание процесса создания системы**

#### **3.1. Разработка дашборда**

В завершение анализа данных, разработал дашборд с основными показателями, для мониторинга, контроля показателей, оперативного реагирования, в случае выхода из установленных показателей и возможности глубокой аналитики в разрезе день, месяц, год.

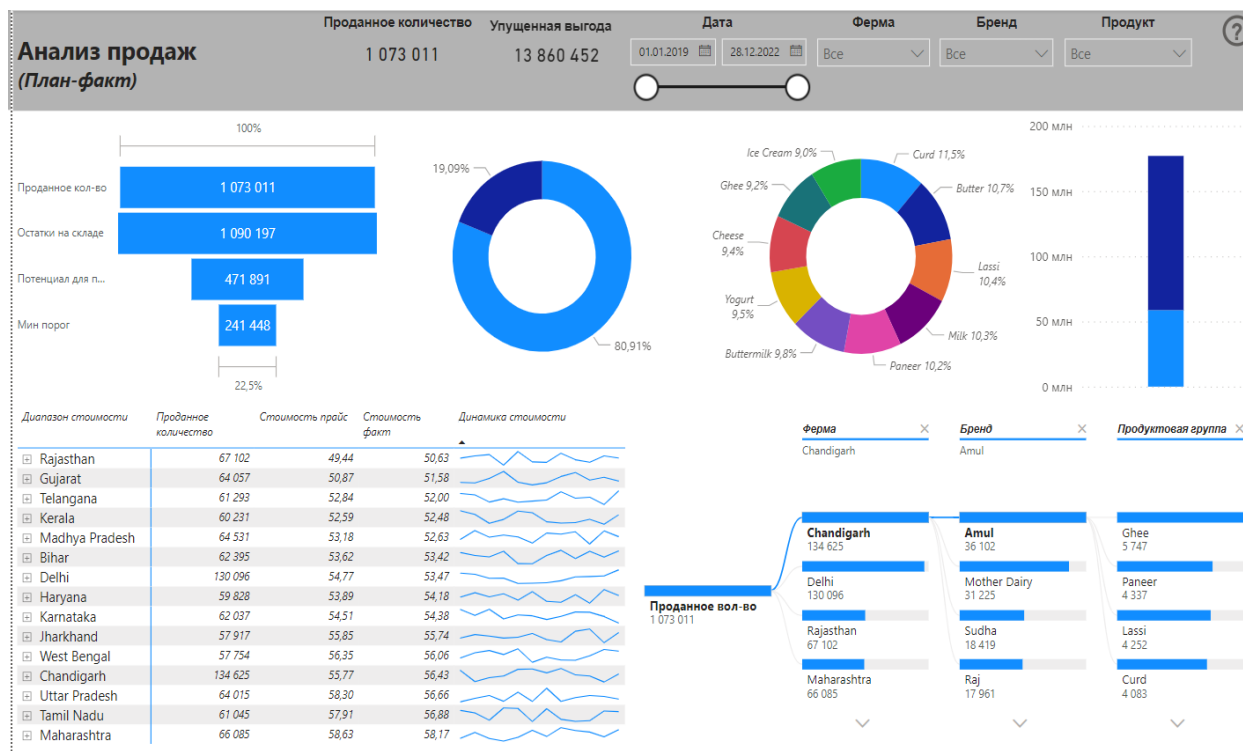
Первый блок, направлен на основные показатели продаж, объем произведенного продукта и продажи конечному потребителю, полученный доход(прибыль). С возможностью распределения по локациям, бренду и продуктовой группе. Основные типы канала продаж, с разбивкой на проценты. Сформирована динамика продаж, с прогнозом на следующий год. Фермы и конечных заказчиков отобразил на карте.

# Дашборд анализ продаж



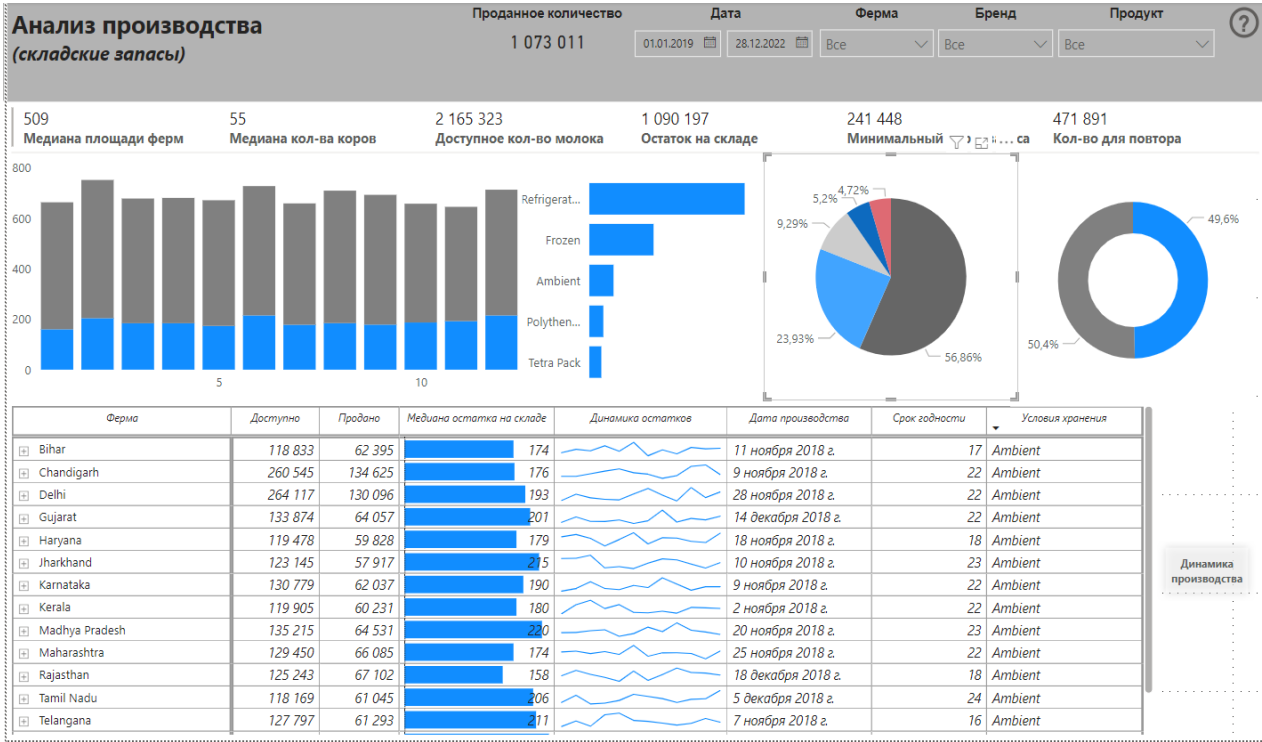
Второй блок, направлен на показатели сравнения плана с фактом и расчёт упущенной выгоды, распределение долей продуктовой группы, стоимость продукта, стоимость продажи продукта (прайс – факт). Построена воронка проданного объема молочного продукта, остаток на складе, потенциал для повторного заказа и минимальный порог для повторного заказа. На основаи этих проказателей, произвел расчет упущенной выгоды, минимальный порог к повторному заказу и потенциальный объем к повторному заказу. Данные показатели, необходимо корректировать и оптимизировать производственную программу. Цель, минимизировать разницу между потенциалом для повтоного заказа и минимальным порогом к повторному заказу.

## Дашборд анализ продаж (План-факт)



Третий блок, направлен на анализ производства (складские запасы), объем произведенного молочного продукта, площадь фермы, количество коров, доступное количество молока, остатки на складе. Разбивка по периодам один месяц, так как данный показатель необходимо рассматривать в разрезе одного месяца, выпуск- продажа – остатки. Контроль срока годности, относительно типа хранения. Контроль динамики появления остатков на складе и динамика производства по каждой продуктовой группе.

# Дашборд анализ производства (складские запасы)



### **3.2. Подведение итогов. Разработка предложений по улучшению процесса продаж и оптимизации производства.**

Вывод: после проделанного анализа, сформировал следующие предложения по улучшению процесса продаж и оптимизации производства.

1. Расширить информацию в датасете, добавить дополнительные показатели. Например, отслеживать объем производства и количество остатков на конец периода (месяц), фиксировать какой объем перешёл на следующий период, сколько ушло в отходы (потери), рассчитывать стоимость потерь, типа перепроизводство, дополнительные затраты.
2. Осуществлять постоянный, оперативный мониторинг показателей, один раз в квартал производить ABC-анализ по продуктовой группе и XYZ -анализ, по итогу производить корректировку в планирование производства, модель продаж.
3. Разработать модель продаж, согласно стратегическому планированию (прогноз не менее одного года).
4. Сократить объем Reorder Quantity (liters/kg) рекомендуемого количества молочного продукта для повторного заказа, до показателей Minimum Stock Threshold (liters/kg) минимального порога запасов молочного продукта. Минимизировать показатель «упущенная выгода», сформировать программу по сокращению складских остатков.
5. Для сокращения складских остатков, необходимо разработать дорожную карту по выходу на внешний рынок, расширить территорию продаж продуктовой группы, в первую очередь использовать продукцию из группы В и С (из ABC-анализа).
6. Осуществлять контроль производственной программы, в рамках одного периода, (например, в течение одного месяца), разработать стратегическое планирование производства. Для данного типа продукции, с минимальным сроком годности и небольшим временем производства, складские остатки должны быть минимальные, близкие к нулю, производство - отгрузка –

продажа. Формировать плановые показатели, производить сравнение с фактическими, отслеживать динамику, в течение определенного периода времени (один месяц).

7. Изменить подход к продажам, пересмотреть КПЭ сотрудников отдела продаж.
8. Отслеживать динамику, делать анализ продаж и потребности конечных заказчиков, а также формирование складских остатков.
9. Пересмотреть концепцию стоимости продуктовой группы, где возможно повысить стоимость, начать процесс корректировки, взять за базовый уровень (доходная часть/маржинальный доход). В случае, где стоимость была ниже прайса, провести аудит процесса, определить узкие места, проблемы, это может быть переизбыток продукта на рынке, высокая конкуренция, неэффективная работа отдела продаж, неграмотное ценообразование. Выявить причину, почему в некоторой группе продуктов, требуется постоянно давать скидки, а по остальной части продавать дороже. В итоге, общая картина по доходу компенсируется, потери на длительном промежутке (допустим в рамках одного года) времени не заметны, а итоговые (факт) показатели выравниваются относительно прогноза (плановые). Произвести оценку стоимости конкурентов. Перспектива на развитие имеется как со стороны производства, так и со стороны отдела продаж. В дальнейшем делать контроль соблюдения ценовой политики, соблюдение прайса, в случае отклонения фиксировать причину.
10. В общем, каждый год присутствуют потери по факту окончания календарного (финансового) года. Необходимо сделать обследование производства, организовать операционную трансформацию на производстве. А также, повторно, запустить проект по стратегическому планированию. Проводить SWOT- анализ, поможет дать объективную оценку деятельности компании.
11. В ходе проведения аналитики, наблюдаем большое количество выбросов в датасете, необходимо дополнительно фильтровать данные.

## Список литературы:

1. <https://v8.1c.ru/metod/article/molokozavod-kak-obespechit-effektivnoe-proizvodstvo-i-sokhranit-sbyt-v-period-krizisa.htm?>
2. <https://habr.com/ru/companies/regionsoft/articles/520242/>
3. <https://dzen.ru/a/YK5JhhfRcSg4337H?experiment=919795&experiment=919795>
4. [https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.sort\\_values.html](https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.sort_values.html)
5. <https://habr.com/ru/companies/ruvds/articles/494720/>
6. <https://datastart.ru/blog/read/10-priemov-python-pandas-kotorye-sdelayut-vashu-rabotu-bolee-effektivnoy>
7. <https://www.unisender.com/ru/glossary/chto-takoe-abc-analiz-i-kak-ego-provesti/>
8. <https://learn.microsoft.com/ru-ru/dax/calculate-function-dax>
9. <https://www.youtube.com/@itresume7464>
10. <https://ru.stackoverflow.com/questions/959525/%D0%9A%D0%B0%D0%BA-%D0%BF%D0%BE%D0%BB%D1%83%D1%87%D0%B8%D1%82%D1%8C-%D0%B3%D0%BE%D0%B4-%D0%B8%D0%B7-%D0%BF%D0%BE%D0%BB%D0%BD%D0%BE%D0%B9-%D0%B4%D0%B0%D1%82%D1%8B-%D0%B2-%D1%81%D1%82%D0%BE%D0%BB%D0%B1%D1%86%D0%B5-dataframe>
11. <https://www.codecamp.ru/blog/pandas-count-unique-values/>
12. <https://Github.com>
13. <https://www.kaggle.com/>
14. Книга Масааки Имаи «Стратегический кайдзен: как изменить ДНК компании и стать лидером отрасли» 28 июня 2022 г.
- 15.



## Приложения:

### Приложения №1. Блок схема процесса

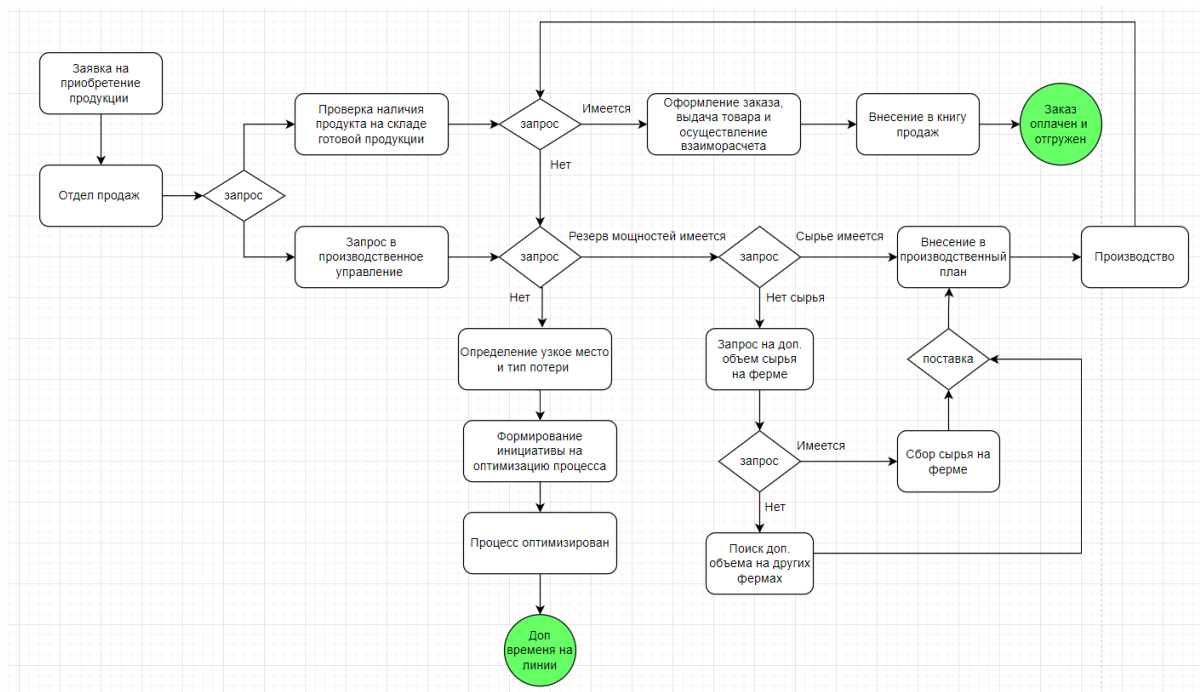


Рис. 1. Жизненный цикл формирования коммерческого предложения на заявку

### Приложение №2. Скорректированный функционал (реквизиты столбцов)

'Location': 'Loc' - Местоположение: Географическое положение молочной фермы.

'Total Land Area (acres)': 'Tot\_Land\_Area' - Общая площадь земли (акры): общая площадь земли, занимаемая молочной фермой.

'Number of Cows': 'Num\_of\_Cows' - Количество коров: количество коров на молочной ферме.

'Farm Size': 'Farm\_Size' - Размер фермы: Размер молочной фермы (в кв.км).

'Date': 'Date' - Дата: дата записи данных.

'Product ID': 'Prod\_ID' - Идентификатор продукта: уникальный идентификатор каждого молочного продукта.

'Product Name': 'Prod\_Name' - Название продукта: Название молочного продукта.

'Brand': 'Brand' - Бренд: бренд, связанный с молочным продуктом.

'Quantity (liters/kg)': 'Quan' - Количество (л/кг): Количество доступного молочного продукта.

'Price per Unit': 'Price\_per\_Unit' - Цена за единицу: Цена за единицу молочного продукта.

'Total Value': 'Tot\_Value' - Общая стоимость: общая стоимость доступного количества молочного продукта.

'Shelf Life (days)': 'Shelf\_Life' - Срок годности (дни): Срок годности молочного продукта в днях.

'Storage Condition': 'Storage\_Con' - Условия хранения: Рекомендуемые условия хранения молочного продукта.

'Production Date': 'Prod\_Date' - Дата производства: дата производства молочного продукта.

'Expiration Date': 'Exp\_Date' - Дата истечения срока годности: дата истечения срока годности молочного продукта.

'Quantity Sold (liters/kg)': 'Quan\_Sold' - Проданное количество (литры/кг): количество проданного молочного продукта.

'Price per Unit (sold)': 'Price\_per\_Unit(sold)' - Цена за единицу (проданного): Цена за единицу, по которой был продан молочный продукт.

'Approx. Total Revenue (INR)': 'INR' - Прибл. Общий доход (INR): приблизительный общий доход, полученный от продажи молочного продукта.

'Customer Location': 'Cust\_Loc' - Местоположение покупателя: Местоположение покупателя, купившего молочный продукт.

'Sales Channel': 'Sales\_Channel' - Канал продаж: Канал, через который продавался молочный продукт (Розничная, Оптовая, Интернет).

'Quantity in Stock (liters/kg)': 'Quan\_in\_Stock' - Количество на складе (литры/кг): количество молочного продукта, оставшегося на складе.

'Minimum Stock Threshold (liters/kg)': 'Min\_Stock\_Threshold' - Минимальный порог запаса (литры/кг): Минимальный порог запаса молочного продукта.

'Reorder Quantity (liters/kg)': 'Reo\_Quan'- Количество для повторного заказа (литры/кг): Рекомендуемое количество молочного продукта для повторного заказа.

Приложение №3. Код ABC – анализа (полный код добавлен файлом в пакет документации) фрагмент:

```
df_abc_multi = df_abc[['Prod_Name', 'INR', 'Quan_Sold', 'profit']]
✓ 0.0s
```

```
df_abc_multi.info()
✓ 0.0s
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 1109 entries, 3 to 4324
Data columns (total 4 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   Prod_Name    1109 non-null   object
1   INR          1109 non-null   float64
2   Quan_Sold    1109 non-null   int64
3   profit       1109 non-null   float64
dtypes: float64(2), int64(1), object(1)
memory usage: 43.3+ KB
```

```
index = 'Prod_Name'
cols = list(df_abc_multi.columns)
cols.remove(index)
for col in cols:
    grouped_df[f'rel_{col}'] = grouped_df[col] / sum(grouped_df[col])
    grouped_df = grouped_df.sort_values(f'rel_{col}', ascending=False)
    grouped_df[f'cumsum_{col}'] = grouped_df[f'rel_{col}'].cumsum()
    grouped_df[f'abc_{col}'] = np.where(grouped_df[f'cumsum_{col}'] < 0.8, 'A', np.where(grouped_df[f'cumsum_{col}'] < 0.95, 'B', 'C'))
grouped_df[cols + [f'abc_{col}' for col in cols]].head(5)
✓ 0.0s
```

	INR	Quan_Sold	profit	abc_INR	abc_Quan_Sold	abc_profit
3906	21757.02	226	0.29	A	B	A
2362	3630.92	86	0.29	C	C	A
4193	1580.61	47	0.29	C	C	A
683	37004.00	400	0.29	A	A	A
171	8587.00	277	0.29	B	A	A

Проводим анализ датасета

```
grouped_df['final_group'] = grouped_df['abc_INR'] + grouped_df['abc_Quan_Sold'] + grouped_df['abc_profit']
groups = grouped_df.groupby('final_group')['final_group'].agg({'count'}).reset_index()
groups['perc'] = groups['count'] / sum(groups['count'])
✓ 0.0s
```

```
groups
✓ 0.0s
```

groups			
	final_group	count	perc
0	AAA	185	0.166817
1	AAB	95	0.085663
2	AAC	96	0.086564
3	ABA	26	0.023445
4	ABB	13	0.011722
5	ABC	9	0.008115
6	BAA	50	0.045086
7	BAB	17	0.015329
8	BAC	27	0.024346
9	BBA	94	0.084761
10	BBB	25	0.022543
11	BBC	33	0.029757
12	BCA	22	0.019838
13	BCB	6	0.005410
14	BCC	7	0.006312
15	CAA	9	0.008115
16	CAB	3	0.002705
17	CAC	7	0.006312
18	CBA	39	0.035167
19	CBB	15	0.013526
20	CBC	8	0.007214
21	CCA	165	0.148783
22	CCB	72	0.064923
23	CCC	86	0.077547

