

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«Тольяттинский государственный университет»

Институт математики, физики и информационных технологий
(наименование института полностью)

Кафедра «Прикладная математика и информатика»
(наименование)

09.04.03 Прикладная информатика
(код и наименование направления подготовки)

Управление корпоративными информационными процессами
(направленность (профиль))

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА (МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ)

на тему «Методы и модели интеллектуального анализа данных в задачах управления предприятиями сферы торговли»

Обучающийся

С.Н. Косолапов

(И.О. Фамилия)

(личная подпись)

Научный
руководитель

кандидат экономических наук,
доцент Т.А. Раченко

(ученая степень, звание, И.О. Фамилия)

Тольятти 2023

Оглавление

Введение.....	3
Глава 1 Анализ проблематики исследования.....	7
1.1 Обзор научных работ по тематике исследования.....	7
1.2 Общая характеристика технологий Big Data	8
1.3 Анализ методов обработки больших данных.....	28
1.4 Анализ практик использования систем больших данных в сфере торговли	35
Глава 2 Анализ решения исследовательской задачи.....	39
2.1 Постановка задач обработки больших данных	39
2.2 Описание и обоснование выбора средств автоматизации обработки больших данных	47
Глава 3 Статистический анализ данных предприятий торговли с использованием методов обработки больших данных	56
3.1 Постановка задач обработки больших данных	55
3.2 Описание средств автоматизации обработки больших данных.....	59
Глава 4 Апробация результатов исследования	62
4.1 Проведение статистической обработки массива данных	62
4.2 Апробация результатов по статистике продаж.....	87
Заключение	109
Список используемых источников.....	112

Введение

В настоящее время с развитием современных технологий обработки больших данных и созданием программного обеспечения, позволяющего осуществлять их обработку, созданы возможности использования указанных систем в планировании деятельности компаний сфере торговли.

Основными проблемами в деятельности компаний, работающих в сфере торговли, являются: необходимость оптимизации номенклатуры реализуемого товара, невозможность оптимального использования складских площадей и времени нахождения продукции ограниченного срока годности, предназначенной для транспортировки, на складах. Также проблемным фактором является обеспечение эффективности при взаимодействии с партнерами и клиентами в части вероятности срыва обязательств по срокам поставки производимой продукции, транспортировки грузов и проведения платежей за поставленную продукцию. Технологии анализа больших данных обеспечивают возможности формирования отчетности, на основании которой выдаются рекомендации по повышению эффективности работы организаций.

Цель данной работы: Применение технологий анализа и использования больших данных в сфере торговли.

Задачи работы:

- анализ источников по теме исследования;
- анализ теоретических аспектов использования систем Big Data, алгоритмов обработки информации;
- анализ областей использования больших данных в сфере торговли;
- анализ бизнес-процессов сферы торговли, предполагающих использование систем больших данных;
- выбор программных решений для автоматизации работы компаний сфере торговли с использованием технологий Big Data;

- анализ аппаратного обеспечения, применяемого для обработки больших данных в сфере торговли;
- формирование статистических данных (выборки) для апробации решения;
- анализ полученных результатов;
- оценка эффективности применения инструментария по обработке больших данных в сфере торговли.

Объект исследования: интеллектуальный анализ данных в целях принятия управленческих решений предприятиями сферы торговли.

Предмет исследования: методы и модели интеллектуального анализа данных в задачах управления предприятиями сферы торговли.

Научная новизна: возможность использования полученных результатов в условиях предприятий сферы торговли в статистическом анализе данных реализации товаров.

Гипотеза исследования: применение технологий больших данных в сфере торговли позволит оптимизировать ассортиментную политику, снизить различного вида издержки и повысить эффективность управления торговыми предприятиями.

Практическая значимость: возможность использования результатов исследования в сфере торговли при использовании систем обработки больших данных с целью планирования реализации товаров и повышения эффективности управления торговых предприятий.

В работе используются методы: системно-информационный анализ; статистическая обработка больших данных, синтез, включенное наблюдение, математическое моделирование.

Теоретическая методология исследования: историко-логический анализ с упором на труды: Григорьева А. А., Исаева Е. А., Тарасова П. А. «Передача, хранение и обработка больших объемов научных данных», Даниловой Н.В., Белявского Г. И. «Анализ данных. Машинное обучение», а

также зарубежные исследования Korjakovskij I. «Safer reinforcement learning for robotics». Сравнительный анализ существующих систем, основываясь на трудах Ерохина В. В., Раяновой А. Р., Кафтарева А. Б. «Статистические методы анализа в бизнес-информатике», Железнова М. М. «Методы и технологии обработки больших данных».

Практическая методология исследования - применение информационных средств и программного обеспечения для возможности статистической обработки данных в Pandas - Плас Д. В. «Python для сложных задач: наука о данных и машинное обучение».

Предполагаемые научные результаты: результаты обработки больших данных позволят оптимизировать деятельность исследуемой компании в части ассортимента реализуемой продукции, снижения издержек.

Апробация и внедрение результатов работы велись в течение всего исследования. За время проведения исследования были опубликованы следующие научные статьи:

- «Обработка и интеллектуальный анализ данных». Представлена на VIII Международной научно-практической конференции (школы-семинара) молодых ученых «Прикладная математика и информатика: современные исследования в области естественных и технических наук» (20-22 апреля 2022 г., место проведения – Тольяттинский государственный университет);

- «Применение экономико-математических методов в анализе дебиторской задолженности». Представлена на IX Международной научно-практической конференции (школы-семинара) молодых ученых «Прикладная математика и информатика: современные исследования в области естественных и технических наук» (18-20 апреля 2023 г., место проведения - Тольяттинский государственный университет);

- «Совершенствование методики анализа коэффициента «критической» ликвидности». Представлена на IX Международной научно-практической конференции (школы-семинара) молодых ученых «Прикладная математика и информатика: современные исследования в области естественных и

технических наук» (18-20 апреля 2023 г., место проведения - Тольяттинский государственный университет);

- «Поддержка принятия управленческих решений методом интеллектуального анализа данных на предприятиях сферы торговли». Представлена на XIII Международной научно-практической конференции: «Современное состояние и перспективы развития науки и образования: проблемы и решения» (7 марта 2023г., место проведения: Россия, город-курорт Анапа);

- «О поддержке принятия управленческих решений на предприятиях сферы торговли методом интеллектуального анализа данных». Представлена на XIII Международной научно-практической конференции «Современное состояние и перспективы развития науки и образования: проблемы и решения» (7 марта 2023г., место проведения: Россия, город-курорт Анапа).

На защиту выносятся:

– методы и модели интеллектуального анализа данных для выстраивания эффективной системы управления предприятий сферы торговли;

– результаты апробации предлагаемого проектного решения.

Работа изложена на 116 страницах и включает 62 рисунка, 6 таблиц, 92 источника.

Глава 1 Анализ проблематики исследования

1.1 Обзор научных работ по тематике исследования

В рамках данной работы проведён анализ источников в области использования технологий интеллектуального анализа данных. В работах [1, 4, 8] проведён анализ основных направлений использования интеллектуальных систем для решения задач статистического анализа. Показано, что системы обработки больших данных позволяют выявлять основные закономерности исследуемых параметров, выявлять аномальные наблюдения, проводить анализ поведения систем при изменении какого-либо из параметров.

В работах [2, 24, 26, 30] проведено описание математических моделей алгоритмов анализа больших данных. Проведён обзор основных принципов группировки данных, оптимизации расчетов. Также рассмотрен порядок выявления характера распределения, выявления зависимостей и трендов.

В работах [3, 14, 15, 19] описан инструментарий, посредством которого осуществляется обработка массивов больших данных. Рассмотрены библиотеки Python, MATLAB. Описан порядок работы с библиотеками, настройка систем расчетов, принципы графического вывода информации.

В работах [5, 11, 28, 37, 50] проведён анализ использования систем интеллектуального анализа больших данных в сфере торговли. Описан порядок обработки информации, включающей данные по продажам продукции в разрезе товарных позиций, формата торговых точек. Описаны процессы анализа эффективности маркетинговых политик, расчеты параметров скидок, при которых наблюдается рост прибыльности компании при минимизации издержек, связанных с расходами на проведение бонусных акций. Также описан порядок проведения анализа эффективности использования складских площадей в части объемов товарных запасов и

сроков хранения товара, оптимальной штатной численности сотрудников, распределения рабочего времени.

В работах [18, 21, 22, 23, 24] проведён анализ использования систем интеллектуального анализа больших данных в области транспортного обслуживания и логистики. Рассмотрены основные принципы планирования бюджетов на транспортные расходы компаний на основе анализа зависимости показателей расхода ГСМ, пробега, сервисного обслуживания транспорта.

Применение технологий и анализа больших данных даёт определенные преимущества при принятии управленческих решений в рассматриваемой области, позволяет оптимизировать бизнес-процессы, эффективно использовать имеющиеся ресурсы.

1.2 Анализ существующих методов и моделей интеллектуального анализа данных в вопросах управления предприятиями сферы торговли

Перед тем как осуществить анализ существующих методов и моделей интеллектуального анализа данных в вопросах управления предприятиями сферы торговли целесообразно рассмотреть текущее положение дел в данной сфере.

Оборот розничной торговли в 2022 году составил 42 трлн 512,5 млрд рублей, или 93,3% (в сопоставимых ценах) к 2021 году, в декабре 2022 года - 4 трлн 201,1 млрд рублей, или 89,5% к уровню соответствующего периода предыдущего года.[28]

Согласно данным статистики, «в 2022 году оборот розничной торговли на 95,4% формировался торгующими организациями и индивидуальными предпринимателями, осуществляющими деятельность вне рынка, доля розничных рынков и ярмарок составила 4,6% (в 2021 году - 95,4% и 4,6%

соответственно), в декабре 2022 года - 95,3% и 4,7% (в декабре 2021 года - 95,5% и 4,5% соответственно)». [22]

По данным Росстата, «в 2022 году в структуре оборота розничной торговли доля пищевых продуктов, включая напитки, и табачных изделий составила 49,5%, непродовольственных товаров - 50,5% (в 2021 году - 47% и 53% соответственно), в декабре 2022 года - 50,7% и 49,3% (в декабре 2021 года - 47,5% и 52,5% соответственно). В декабре 2022 года объемы продаж через интернет по крупным и средним торговым организациям увеличились по сравнению с ноябрем 2022 года на 15% (в сопоставимых ценах), с декабрем 2021 года - уменьшились на 3,6%». [28]

«2022 год для российской экономики ознаменовался наличием целого ряда вызовов, связанных с обострением мировой политической ситуации: уход иностранных компаний с российского рынка, общая переориентация с Запада на Восток в вопросе установления и развития международных экономических связей, поиск новых логистических цепочек поставки товаров, беспрецедентные санкционные ограничения. Всё это не могло не сказаться на состоянии рынка.

Отрасль электронной коммерции (e-commerce), тем не менее, продемонстрировала рост по итогам текущего года, несмотря на уход крупных поставщиков, связанное с этим снижение ассортимента и снижение покупательской способности»[22].

Отметим, что «развитие рынка e-commerce и e-grocery (продовольственная онлайн-торговля) в своё время резко ускорила пандемия COVID-19. По итогам прошлого года рынок онлайн-торговли вырос на 45,4%, и сейчас рост продолжается: за последний год рост количества онлайн-покупок составил более 60%. Несмотря на то, что некоторые ниши существенно просели на фоне всех кризисных событий, в целом лишь в первом полугодии 2022 года интернет-продажи непродовольственных товаров в России увеличились на 51,5%, составив почти 2 трлн рублей, а продовольственная онлайн-торговля выросла на 100%. Эксперты

утверждают, что в течение нескольких лет объем рынка e-grocery в России увеличится еще на треть и превысит 15,3 трлн рублей, что сделает продукты важнейшим драйвером онлайн-продаж. В отчете, который опубликовало Forbes, речь идет о росте продажи продуктов в интернете более чем в пять раз к 2026 году»[27].

«Лишь в первом полугодии 2022 отечественные ретейлеры заработали в онлайн 301 млрд рублей, что вдвое превышает показатели аналогичного периода 2021 года. Предсказывается рост показателя до 940 млрд рублей в 2023 году»[23].

По мнению ряда экспертов, «лидерство по объемам онлайн-продаж в России сохраняют за собой такие известные маркетплейсы как Wildberries и Ozon, рядом с ними стоит Яндекс Маркет. Оборот Ozon лишь за 9 месяцев 2022 года прибавил 98% год к году и достиг 536,2 млрд рублей. Оборот Wildberries вырос почти на 100%, превысив 1 трлн рублей (GMV, оборот от продаж товаров и услуг с учетом возвратов и скидок), рост продаж в штуках составил более 70%. Только в 3 квартале 2022 года оборот компании увеличился на 95% до 420,2 млрд рублей, что стало для неё новым рекордом. За отчетный период оборот компании (GMV с учётом сервисов и за вычетом возвратов) вырос на 74% по сравнению с аналогичным периодом прошлого года и составил 188,1 млрд рублей. При этом оборот маркетплейса, т. е. продажи продавцов, показали рост более чем в 2 раза (на 104%) до 147,1 млрд рублей[22].

Отдельно стоит отметить, что по итогам ноября 2022 года Wildberries и Ozon впервые опередили китайскую площадку AliExpress по продажам в России. Согласно данным МТС, обнародованным в декабре 2022 года, более 50% продаж пришлось на Wildberries, Ozon же занял второе место с долей в 29%[27].

Также стоит отметить, что в декабре 2022 года состоялась конференция для онлайн-предпринимателей «Tinkoff eCommerce», в рамках которой эксперты обсудили не только основные события уходящего, 2022 года, но и

прогнозы на 2023. Говоря о ключевых тенденциях, прежде всего, выделяется развитие e-commerce в сторону «вертикализации» рынка и соответствующей ей мультиплатформенности брендов[22].

На данный момент, главную роль в отечественной электронной торговле играют такие крупные игроки как универсальные маркетплейсы. При этом, согласно исследованию, 75% продавцов изначально заходят только на один маркетплейс, подключая две и более площадки уже спустя год-два. В дальнейшем ожидается, во-первых, увеличение скорости данного процесса размещения бренда на разных площадках (собственно, переход к мультиплатформенности бизнеса, то есть присутствие бренда сразу везде), а во-вторых, рост количества нишевых площадок для размещения[27].

В ближайшие годы могут появиться более 100 нишевых маркетплейсов как в B2C, так и в B2B-сегменте, — утверждают эксперты. Отдельно здесь стоит отметить, что весьма перспективным направлением является создание и внедрение современных онлайн-систем закупочных процедур, особенно в B2B сфере. По данным компании RAEX Аналитика, в 2022 году объемы малых закупок выросли на 15,6%, а доля малых закупок в общем объеме закупок удвоилась»[43].

«Мультиплатформенность как тренд тесно связана и с повышением внимания к возможностям продвижения на сайтах и в приложениях социальных сетей. За последние годы существенно выросла популярность Facebook Marketplace и Instagram Shopping (принадлежат к Мета, запрещённой в РФ), так что, возможно, стоит ожидать создание и развитие подобных сервисов на базе отечественных социальных сетей»[54].

«Ещё одной интересной тенденцией e-commerce является изменение логистики. Согласно данным исследования Data Insight, опубликованном в ноябре 2022 года, последние три года интернет-магазины осуществляли сами около 74% доставок, однако в 2022 году показатель вырос ещё на 12%. Отказ от услуг логистических операторов обещает привести, с одной стороны, к

консолидации российского рынка логистики, с другой — к большому развитию интернет-магазинами собственных систем доставки»[46].

«Автоматизация бизнес-процессов, в свою очередь, связана с увеличением роли искусственного интеллекта (ИИ) в организации предпринимательской деятельности. Стоит ожидать развитие таких форм ИИ как чат-боты, голосовые помощники, рекомендации по продуктам, персонализация рекламы и электронной коммерции в целом, а также введение искусственных алгоритмов, способных, например, вести документооборот или отслеживать наличие товаров на складах. В тренде находятся и технологии дополненной реальности (AR), позволяющие создавать такие формы привлечения клиентов как виртуальные примерочные, специализированные сервисы для подбора косметики, студии дизайна интерьеров и так далее»[83].

«Прогнозы экспертов касательно судьбы электронной коммерции в 2023 году можно обозначить как позитивные. Ожидается дальнейший рост рынка, окончательно оправившегося от вызовов уходящего года»[77].

«Деятельность любого торгового предприятия независимо от его сегмента и сектора сложна и многообразна. Деятельность торгового предприятия регулируется и управляется людьми».

«Управление – есть сознательное воздействие человека на объекты и процессы в целях придания деятельности предприятия коммерческой направленности и получения определенных результатов»[52].

«Управление коммерческой деятельностью предприятий сферы торговли ставит своей непосредственной задачей внести определенную упорядоченность в коммерческие и торговые процессы, организовать совместные действия работников, участвующих в этих процессах, достичь согласованности и координации действий»[18].

«При этом управление направлено на оптимизацию работы сотрудников в целях повышения эффективности коммерческих процессов и достижения конечных целей предприятия»[24].

«В управлении деятельностью торгового предприятия присутствуют две стороны: управляющие и управляемые»[18]. (Рисунок 1).



Рисунок 1 - Стороны управления деятельностью торгового предприятия[18]

«Тех, кто управляет, называют субъектами управления (администраторы, руководители, управляющие)»[61].

«Объекты управления - это те, кем управляют (работники, коллективы), и то, чем управляют (экономика, торговая деятельность, торговый процесс). Взаимодействие субъектов и объектов посредством управляющих воздействий и обратной связи позволяет целенаправленно управлять многосторонней деятельностью торгового предприятия»[27].

«Управляющее воздействие осуществляется посредством использования законов, указов, планов, программ, постановлений, нормативов, рекомендаций, инструкций, материальных и финансовых стимулов»[46].

«Обратные связи — это результаты непосредственных наблюдений и контроля со стороны субъекта управления, статистическая и текущая отчетность, бухгалтерская документация»[34].

«В современных условиях рынка возникла потребность в расширении управленческих задач, разработке новых приемов и методов управления коммерческой деятельностью торговых организаций, пригодных для

хозяйствующих субъектов разных форм собственности и организационно-правовых форм»[18].

«Процесс управления торговым предприятием должен быть основан на принципах рыночной экономики и методологии современного управления. Система управления, ориентированная на рынок, означает не только организацию структуры и взаимосвязанную совокупность задействованных процессов предприятия, но и их сочетание со всеми внешними факторами»[46].

«Управление коммерческой деятельностью предприятий сферы торговли ставит своей непосредственной задачей обеспечить определенную упорядоченность в коммерческие и торговые процессы, организовать совместные действия работников, участвующих в этих процессах, достичь согласованности и координации действий»[72].

«При этом управление направлено на оптимизацию работы сотрудников для повышения эффективности коммерческих процессов и достижения конечных целей предприятия. Коммерческая деятельность ориентирована на реализацию товаров и услуг в соответствии с требованиями рынка»[58].

«В практической деятельности торговых предприятий важное значение приобретает оценка эффективности каждой торговой сделки и коммерческой деятельности предприятия в целом, что позволяет принимать обоснованные управленческие решения. В связи с этим вопросы повышения эффективности управления коммерческой деятельностью торговых предприятий имеют важное прикладное значение»[16].

Ключевым способом повышения эффективности в задачах управления предприятиями сферы торговли является применение методов интеллектуального анализа данных [74].

«Термин «интеллектуальный анализ данных»(Data Mining) фигурировал в академических журналах еще в 1970 году, но по-настоящему популярным он стал только в 1990-х после появления глобальной сети

Интернет[62]. Тогда торговым компаниям потребовалось анализировать большие объемы разнородных данных, чтобы отыскать нетривиальные паттерны и научиться предсказывать поведение клиентов. Обычные модели статистики оказались неспособны справиться с этой задачей» [30].

«Первые системы Data Mining предназначались для обработки данных о продажах в супермаркетах по нескольким параметрам, включая их объем по регионам и тип продукта»[19].

Специалист по обработке данных Григорий Пятецкий-Шапиро определяет «Data Mining» как «процесс обнаружения в сырых данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности» [26].

Интеллектуальный анализ данных – это во многом прикладная теория, число приложений которой к реальным бизнес-задачам растёт с каждым годом[84]. Востребованность специалистов по интеллектуальному анализу данных постоянно возрастает, как и доля финансирования разработок в этой области. Появляется всё больше программных решений для анализа данных, в том числе с открытым исходным кодом [41].

В настоящее время наиболее широкое использование технология «Data Mining» получила в сфере торговли[56].

«Крупным торговым сетям Data Mining позволяет анализировать покупательские корзины, чтобы улучшать рекламу, создавать запасы товаров на складах и планировать, как их разложить на витринах, открывать новые магазины и выявлять потребности разных категорий клиентов»[54]. «Например, российская сеть «Лента» проанализировала данные карт лояльности более 90% своих покупателей и поделила аудиторию на определенные сегменты по покупательскому поведению. В частности, ритейлер выделил сегмент покупающих только базовые продукты и мужчин, которые чаще приобретали только напитки и снеки. Это позволило оптимизировать ассортимент и управлять выкладкой и ценами»[61].

«Крупнейшая в мире платформа электронной коммерции «Amazon» в октябре 2021 года представила инструмент, который предоставит продавцам доступ к информации о том, что в настоящее время ищут покупатели, и тем самым поможет упростить выбор продуктов для продажи»[43].

«Также широко известен пример, когда специалисты одного из супермаркетов крупнейшей международной сети Wal-Mart выявили, что по пятницам вечером пиво часто продается вместе с детскими подгузниками. Менеджеры Wal-Mart умело этим воспользовались: поставив на полки рядом с подгузниками одни из самых дорогих марок пива, удалось добиться «геометрического роста» его продаж»[54].

«Пример определения таких неочевидных связей - это и есть основная задача этой технологии, она позволяет лучше понимать потребности покупателя и его поведение, что, несомненно, нужно использовать для достижения финансовых целей» [28].

Здесь также следует рассмотреть вопросы управления организациями сферы торговли. Известно, что удержание существующего клиента обходится гораздо дешевле для торговой компании, чем привлечение нового[37].

«Представляют также большой интерес те клиенты, которые вероятно откажутся от услуг компании. Возврат ушедшего клиента оцениваются экспертами в сотни раз превышающими стоимость по удержанию. Поэтому предотвратить уход хорошего клиента - одна из важнейших задач любого успешного бизнеса»[68].

«Важным аспектом в работе отделов маркетинга и продаж предприятий сферы торговли является составление целостного представления о клиентах, информация об их особенностях, характеристиках, структуре клиентской базы»[11].

«Сегментацию клиентской базы обычно проводят, используя правило Парето «80/20»: примерно 80% доходов приносят 20% клиентов, поэтому максимальное внимание стоит уделить этим 20%. К сожалению, многие

торговые компании не выполняет даже эту базовую сегментацию, чтобы выделить наиболее прибыльных клиентов»[19].

«Методы выявления потребностей покупателей можно условно разделить на два вида:

- количественные - методы на основе анализа данных о совершаемых покупках (например, чеков);

- качественные - на основе анализа результатов специально проводимых для этого маркетинговых исследований»[26].

«В обоих случаях большинство российских компаний, даже с самыми сильными аналитическими отделами, для анализа обычно пользуются лишь стандартными статистическими методами. Использовать только статистические методы нерационально, т.к. теряется множество нужной и полезной информации о покупателях, которую обычные статистические методы обнаружить не могут из-за того, что они для этого не предназначены»[57]. «Статистические методы преимущественно ориентированы лишь на обобщение информации, а не на ее глубокий анализ, который необходим для внимательного и всестороннего изучения клиентов предприятий сферы торговли. Наличие в таких компаниях человека или лиц, выполняющих роль бизнес-аналитиков, обязательно подтолкнет руководство использовать средства Data Mining- это лишь вопрос времени»[37].

«Рассмотрим некоторые ключевые методы интеллектуального анализа данных в сфере торговли:

- ассоциация;
- классификация;
- кластеризация;
- прогнозирование;
- дерево решений;
- метод комбинаций»[44].

«Ассоциация (или отношение) является наиболее известным методом интеллектуального анализа данных. Данный метод заключается

в сопоставление двух или более элементов чаще всего одного и того же типа. Примером ассоциативной закономерности (ассоциативного правила) служит правило, указывающее, что из события X следует событие Y »[88].

Ассоциативные правила помогают выявлять группы товаров, как правило, приобретаемые совместно. Знание этих правил позволяет соответствующим образом размещать товары на прилавках, стимулируя интенсивность их продаж [47].

«Метод классификации используют при описании нескольких атрибутов для идентификации определенного класса, а также в качестве входных данных для других методов. Для определения классификации применяют деревья принятия решений. Каждый класс обладает определенными свойствами, которые характеризуют его объекты»[82].

«Наглядный пример классификации - задача выяснения придет ли покупатель вновь в нашу торговую точку. Человек, проводящий анализ, при этом оперирует двумя известными ему классами - «придет» и «не придет». Итак, задача классификации - определить к какому классу относятся те или иные данные. При этом множество классов, к одному из которых впоследствии можно отнести исследуемый объект, заранее известно. Отнести покупателя к той или иной группе помогает анализ его характеристик - возраста, истории покупок, уровня дохода, семейного положения»[74].

«Кластеризация позволяет использовать общие атрибуты различных классификаций в целях выявления кластеров. Исследуя один или более атрибутов можно сгруппировать отдельные элементы данных, вместе получая структурированное заключение. На простом уровне при кластеризации используется один или несколько атрибутов в качестве основы для определения кластера сходных результатов. Кластеризация полезна при определении различной информации, потому что она коррелируется с другими примерами, так что можно увидеть, где подобию и диапазоны согласуются между собой» [47].

«Прогнозирующие методы используют значения одних переменных для предсказания неизвестных (целевых) значений переменных. В сочетании с другими методами интеллектуального анализа данных прогнозирование предполагает анализ тенденций классификацию сопоставление с моделью отношения»[73].

«Метод под названием прогнозирование хорошо знаком управляющим предприятий сферы торговли. Так анализируя данные прошлых периодов можно построить прогноз на будущее - причем, чем подробнее исторические данные и чем больше анализируемый отрезок времени, тем точнее получатся результаты. Этот метод применяется для прогнозирования структуры сбыта, характеризующегося сезонными колебаниями или потребности в кадрах. Последовательные модели используются для анализа долгосрочных данных — полезный метод выявления тенденций или регулярных повторений подобных событий»[6].

«Дерево решений, связанное с большинством других методов, используют в рамках критериев отбора, а также для поддержки выбора определенных данных в рамках общей структуры. Дерево решений начинают с простого вопроса, который имеет два ответа (но возможно и больше). Каждый ответ приводит к следующему вопросу помогая классифицировать и идентифицировать данные или делать прогнозы. Деревья решений чаще всего используются с системами классификации информации системами прогнозирования, где различные прогнозы могут основываться на прошлом историческом опыте, который помогает построить структуру дерева решений и получить результат»[22].

Так в задачах управления предприятиями сферы торговли потенциальным инвесторам очень важно получить информацию о целесообразности финансовых влияний в открытие новых торговых объектов, а сотрудникам торгово-закупочных отделов важно получать информацию о целесообразности введения в ассортимент торгового объекта

новых позиций ассортимента продукции. Эти и многие другие задачи позволяет решить метод дерево решений [66].

«Метод комбинации используется довольно редко. Он подобен методам классификации и кластеризации»[9].

«При всех основных методах часто имеет смысл записывать и впоследствии изучать полученную информацию. Для некоторых методов это совершенно очевидно. Например, при построении последовательных моделей и обучении в целях прогнозирования анализируются исторические данные из разных источников и экземпляров информации. В других случаях этот процесс может быть более ярко выраженным. Деревья решений редко строятся один раз и никогда не забываются. При выявлении новой информации событий и точек данных может понадобиться построение дополнительных ветвей или даже совершенно новых деревьев. Некоторые из этих процессов можно автоматизировать. Например, построение прогностической модели для выявления мошенничества с кредитными картами сводится к определению вероятностей, которые можно использовать для текущей транзакции с последующим обновлением этой модели при добавлении новых (подтвержденных) транзакций. Затем эта информация регистрируется так, что в следующий раз решение можно будет принять быстрее»[58].

«Интеллектуальный анализ данных опирается на построение подходящей модели и структуры, которые можно использовать для обработки выявления и создания необходимой информации»[9].

«Аналитические переменные для данных полученных из множества различных источников можно составить в единую определенную структуру (например, создать класс покупателей определенных уровней и возрастов или класс ошибок определенного типа). В зависимости от источника данных важно выбрать правильный способ построения и преобразования этой информации, каким бы ни был метод окончательного анализа данных»[54].

Рассмотрим основные алгоритмы интеллектуального анализа данных в задачах управления предприятиями сферы торговли:

- «- алгоритм дерева принятия решений;
- алгоритм кластеризации;
- алгоритм кластеризации последовательностей;
- алгоритм линейной регрессии;
- алгоритм временных рядов;
- алгоритм взаимосвязей;
- алгоритм нейронной сети;
- алгоритм Байеса»[22].

«Алгоритм дерева принятия решений представляет собой алгоритм линейной регрессии в прогнозировании непрерывных величин и алгоритм классификации для использования в дискретных случаях. Для дискретных величин алгоритм осуществляет прогнозирование на основе связи между входными столбцами в наборе данных»[34].

«Алгоритм дерева принятия решений строит модель интеллектуального анализа данных путем создания ряда разбиений в дереве. Эти разбиения представлены как узлы. Алгоритм добавляет узел к модели каждый раз, когда выясняется, что входной столбец имеет корреляцию с прогнозируемым столбцом. Способ, которым алгоритм определяет разбиение, отличается в зависимости от того, прогнозирует ли он непрерывный столбец или дискретный столбец. Алгоритм дерева принятия решений использует оценку компонентов для управления выбором наиболее полезных атрибутов. Выбор компонентов играет важную роль в предотвращении использования малозначительных атрибутов»[44].

«Одной из часто встречающихся проблем в моделях интеллектуального анализа данных является чрезмерная чувствительность к небольшим различиям в обучающих данных. В этих случаях модель называется чрезмерно оснащенной или чрезмерно обученной. Во избежание чрезмерного

оснащения модели в алгоритме дерева принятия решений используются методики контроля роста дерева»[56].

«Алгоритм кластеризации представляет собой алгоритм разбиения множества. Он использует итерационные методы для группировки объектов в кластеры, содержащие сходные характеристики. Такая группировка полезна для просмотра данных, выявления в них аномалий и создания прогнозов. Модели кластеризации определяют связи в наборе данных, который невозможно логически получить с помощью случайного наблюдения. Алгоритм кластеризации обучает модель на основе связей, существующих в данных и на основе кластеров, идентифицированных алгоритмом»[28].

«Алгоритм кластеризации сначала определяет связи в наборе данных и формирует ряд кластеров на основе этих связей. Точечная диаграмма является хорошим способом визуально представить, как алгоритм группирует данные. После первого определения кластеров алгоритм вычисляет, как кластеры представляют группирование точек, а затем пытается повторно создать кластеры, которые лучше представляют исходные данные»[19].

«Алгоритм последовательно выполняет этот процесс до тех пор, пока не достигнет результата, улучшить который будет невозможно. Можно настраивать работу данного алгоритма, выбирая конкретный метод объединения в кластеры, ограничивая максимальное количество кластеров или изменяя размер несущего множества, необходимый для создания кластера. После обучения модели результаты хранятся в виде набора закономерностей, которые можно исследовать или делать на их основе прогнозы. Предусмотрена возможность создавать запросы, возвращающие прогнозы того, соответствуют ли новые данные обнаруженным кластерам, или предоставляющие описательные статистические данные о кластерах»[48].

«Алгоритм кластеризации последовательностей представляет собой алгоритм анализа последовательностей. Его можно использовать для изучения данных о событиях, которые можно связать в логическую последовательность. Алгоритм находит самые распространенные последовательности, выполняя группирование или кластеризацию идентичных последовательностей»[41].

«К алгоритмам кластеризации последовательностей можно отнести данные, которые создаются когда пользователи просматривают веб-сайт; данные, которые описывают порядок, в котором клиент добавляет в корзину товары, выбранные в Интернет-магазине»[62].

«Алгоритм кластеризации последовательностей – это гибридный алгоритм, использующий методы Марковских цепей для определения кластеров и их последовательностей. Одной из особенностей алгоритма кластеризации последовательностей является то, что он использует ряд событий или переходов между состояниями в наборе данных, например, ряд приобретений продуктов или последовательность переходов между сайтами для конкретного пользователя Интернета»[64].

«Алгоритм изучает вероятность переходов и измеряет различия между всеми возможными последовательностями в наборе данных, чтобы определить, какие последовательности лучше всего использовать в качестве исходных данных для кластеризации»[58].

«Алгоритм линейной регрессии является разновидностью алгоритма дерева принятия решений, помогающего рассчитать линейную связь между зависимой и независимой переменной, а затем использовать эту связь при прогнозировании. Связь принимает вид линейной формулы, представляющей заданный набор данных. Если построить график, то каждой точке будет соответствовать ошибка, связанная с ее расстоянием от линии регрессии. Коэффициенты a и b в уравнении регрессии регулируют угол и положение линии регрессии. Можно получать регрессивное уравнение с помощью

подбора коэффициентов a и b до тех пор, пока сумма ошибок, связанных с этими точками, не станет минимальной»[31].

«Существуют другие типы регрессии, в которых используется несколько переменных, а также нелинейные методы регрессии. Однако линейная регрессия является полезным и широко известным методом моделирования ответа на изменение в каком-либо базовом факторе. Линейную регрессию можно использовать для определения связи между двумя непрерывными столбцами. В задачах управления предприятиями сферы торговли алгоритм линейной регрессии используют для вычисления линии тренда в данных продаж»[73].

«Алгоритм временных рядов обеспечивает прогноз непрерывных значений, таких как продажи продуктов, во времени. В отличие от других алгоритмов Майкрософт, таких как деревья решений, модель временных рядов не требует дополнительных столбцов новых сведений, чтобы прогнозировать тенденцию. С помощью модели временных рядов можно прогнозировать тенденции на основе только исходного набора данных, использованного для создания модели»[56].

«При прогнозировании можно вводить в модель новые данные и автоматически задействовать их при анализе тенденций. Сочетание исходных данных и прогнозируемых данных называется рядом. Важной характеристикой алгоритма временных рядов является его способность выполнять перекрестный прогноз. При обучении алгоритма двумя отдельными, но связанными друг с другом рядами можно использовать итоговую модель для прогнозирования исхода одного ряда на основе поведения другого ряда. Например, наблюдаемые продажи одного продукта могут оказать влияние на прогнозируемые продажи другого продукта. Перекрестные прогнозы также полезны при создании общей модели, которую можно применить к нескольким рядам. Например, прогнозы по продажам для конкретного региона нестабильны, так как в ряду недостаточно данных хорошего качества. Общую модель можно обучить на

среднем значении нескольких регионов, а затем применить модель к отдельным рядам, чтобы подготовить более стабильные прогнозы для каждого региона»[48].

«Алгоритм взаимосвязей полезен для выработки рекомендаций покупателям по дальнейшим приобретениям на основе тех товаров, которые они уже купили, или к которым проявили интерес. Алгоритм взаимосвязей полезно использовать для анализа потребительской корзины. Правила, определяемые алгоритмом, могут использоваться для прогнозирования вероятных будущих покупок покупателей на основе элементов, уже имеющихся в корзине покупателя. Алгоритм взаимосвязей прослеживает набор данных для поиска элементов, которые находятся в наборе одновременно. Затем алгоритм группирует в наборы любые связанные элементы, найденные в количестве вариантов, не меньшем минимального параметра»[15].

«Алгоритм нейронной сети сочетает каждое возможное состояние входного атрибута с каждым возможным состоянием прогнозируемого атрибута и использует обучающие данные для вычисления вероятностей. Далее эти вероятности можно использовать для классификации или регрессии, а также для прогнозирования исхода прогнозируемого атрибута на основе входных атрибутов»[86].

«В модели интеллектуального анализа данных, которая создается при помощи алгоритма нейронной сети, может содержаться несколько сетей, что определяется количеством столбцов данных, используемых для анализа. Алгоритм нейронной сети полезен при анализе сложных входных данных, полученных, например, в результате торгового процесса, для которых предоставляется значительный объем обучающих данных, но при этом затруднено образование производных правил при помощи других алгоритмов»[51].

«Существуют следующие сценарии использования алгоритма нейронной сети:

- анализ маркетинговой составляющей торговой компании и рекламы. Например, измерение эффективности прямой почтовой рассылки или рекламной кампании, проводимой по радио;

- прогнозирование изменений цен на акции, колебаний валютных курсов или других изменчивых финансовых данных из данных с предысторией;

- анализ торговых процессов;

- интеллектуальный анализ текста;

- любая прогнозирующая модель, которая анализирует сложные связи между большим количеством входных атрибутов и сравнительно малым количеством выходных атрибутов»[7].

«Алгоритм Байеса является алгоритмом классификации в прогнозирующем моделировании. Данный алгоритм использует теорему Байеса, но не учитывает зависимости, которые могут существовать между событиями. Этот алгоритм можно использовать для первоначального исследования данных, а затем применять результаты для создания дополнительных моделей интеллектуального анализа с другими алгоритмами, требующими большего количества вычислений и являющимися более точными» [11].

«Упрощенный алгоритм Байеса рассчитывает вероятность состояния каждого входного столбца при каждой возможной реализации прогнозируемого столбца, т.е. в предположении, что некоторое вероятное событие уже произошло. После обучения модели результаты хранятся в виде набора закономерностей, которые можно исследовать или делать на их основе прогнозы. Можно создавать запросы, возвращающие прогнозы о связи новых данных с прогнозируемым атрибутом, или получать статистику, описывающую взаимосвязи, обнаруженные моделью»[6].

Модель интеллектуального анализа данных создается путем применения алгоритма к данным. Но это больше, чем алгоритм или контейнер метаданных: это набор данных, статистик и шаблонов, которые

можно применять к новым данным для формирования прогнозов и вывода взаимосвязей. Структура интеллектуального анализа данных и модель интеллектуального анализа данных являются отдельными объектами. В структуре интеллектуального анализа данных хранятся сведения, определяющие источник данных. Модель интеллектуального анализа данных содержит сведения, полученные по итогам статистической обработки данных, например закономерности, обнаруженные в результате анализа»[18].

«Модель интеллектуального анализа данных будет пуста до тех пор, пока не будут обработаны и проанализированы данные, переданные структурой интеллектуального анализа данных. После обработки модель интеллектуального анализа данных содержит метаданные, результаты и привязки к структуре интеллектуального анализа данных»[4].

«Например, одни и те же данные можно использовать для создания нескольких моделей, использующих алгоритм кластеризации, алгоритм дерева принятия решений и упрощенный алгоритм Байеса. В каждом из типов моделей создаются различные наборы шаблонов, наборов элементов, правил и формул, которые могут применяться при прогнозировании. Как правило, каждый из алгоритмов анализирует данные по-своему, поэтому содержимое получаемой модели также организуется в различные структуры. В одном из типов моделей данные и шаблоны могут группироваться в кластеры; в модели другого типа данные могут быть упорядочены с помощью деревьев, ветвей и правил, разделяющих и определяющих данные»[37].

«Модель также зависит от данных, на которых проводилось ее обучение: даже те модели, обучение которых производилось на основе одной и той же структуры интеллектуального анализа данных, могут выдавать различные результаты, если во время анализа фильтрация данных выполнялась по-разному или использовались разные начальные значения. Однако фактические данные не хранятся в сводной статистике, доступной только для модели, хранятся с фактическими данными, находящимися в

структуре интеллектуального анализа данных. Если при обучении модели были созданы фильтры данных, то определения фильтров также сохраняются в объекте модели»[25].

«Чтобы создать модель интеллектуального анализа данных, необходимо выполнить следующие действия:

- создать базовую структуру интеллектуального анализа данных и включите в нее столбцы данных, которые могут потребоваться;
- выберите алгоритм, который наилучшим образом подходит для аналитической задачи;
- выбрать столбцы из структуры для использования в модели и указать, как их следует использовать, в каком столбце содержится результат, который нужно спрогнозировать, какие столбцы предназначены только для ввода и др.;
- задать дополнительные параметры для тонкой настройки обработки, проводимой алгоритмом;
- заполните модель данными, выполнив обработку структуры и модели»[31].

Таким образом, «интеллектуальный анализ данных представляет собой процесс обнаружения пригодных к использованию сведений в крупных наборах данных. В интеллектуальном анализе данных применяется математический анализ для выявления закономерностей и тенденций, существующих в данных. Обычно такие закономерности трудно обнаружить, поскольку связи слишком сложны. Модели интеллектуального анализа данных также можно создавать программным образом с помощью объектов АМО или XML для аналитики, а также клиента интеллектуального анализа данных для Excel и других клиентов. В настоящее время интеллектуальный анализ данных является важнейшим элементом для формирования высокого уровня конкурентоспособности для предприятий сферы торговли»[8].

1.3 Анализ методов обработки больших данных

«В деятельности современных компаний возникает множество задач, связанных с необходимостью обработки больших массивов информации в целях получения необходимой аналитической информации в различных направлениях деятельности»[17].

«Одной из основных проблем в аналитической работе является большое количество разрозненных механизмов, обеспечивающих возможности выявления закономерностей, что связано с необходимостью анализа множества факторов, фигурирующих в рамках осуществления основной деятельности»[6].

«Обработка данных массивов информации ввиду их большого объема предполагает необходимость использования аппарата Big Data, систем искусственного интеллекта, позволяющих выявлять закономерности в параметрах спроса на определенные виды продукции и зависимости объемов продаж от применения различных маркетинговых политик»[23].

«Технологии больших данных обеспечивают возможности обработки полной информации об исследуемых объектах. Например, в случае компаний сфере торговли, объектом анализа могут являться объемы перевозимых грузов, статистика видов заказов, их стоимости, истории проводимых расчетов, наличие прецедентов, связанных с нарушениями контрактных обязательств. Обладая необходимыми технологиями, полученные сведения можно использовать в различных целях - для реализации политики взаимодействия с клиентами, при выборе поставщиков по информации, имеющейся по нескольким контрагентам, при разработке рекламных компаний»[44].

Таким образом, «использование систем Big data обеспечивает возможности: сегментации контрагентов, обеспечения эффективности взаимодействия с ними, визуализации данных по статистике расчетов,

разработки и тестирования партнерских программ, составление прогнозов по ожидаемым объемам производства»[36].

На рисунке 2 приведена диаграмма методов обработки больших данных.

Методы обработки больших данных включают [21]:

Краудсорсинг - привлечение к решению какой-либо проблемы большой группы людей. Технология предполагает проведение просмотра больших массивов информации выбранной группой людей, который в ручном режиме осуществляют выбор данных по определенному признаку.

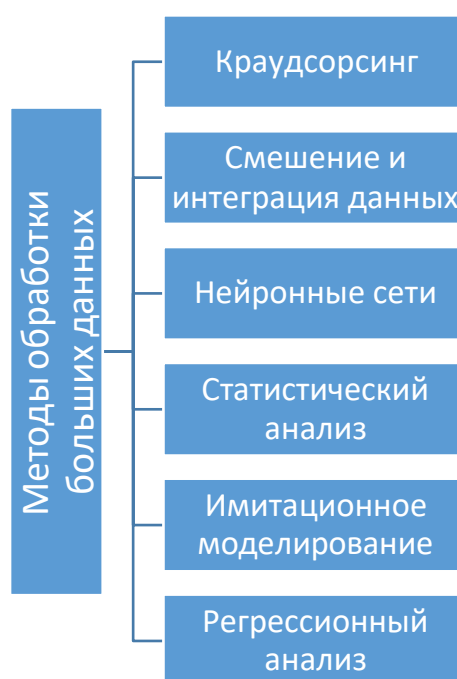


Рисунок 2 – Методы обработки больших данных [21]

Далее проведено описание алгоритма обработки больших данных с использованием указанной методологии. Данная технология предполагает возможности приведения больших массивов информации к виду, пригодному для обработки, что предполагает выявление ошибок в записях и их исправление, удаление данных, идентификация которых невозможна. Обработанные массивы обеспечивают возможности проведения дальнейшей

обработки с использованием стандартных систем автоматизации. Недостатком данного алгоритма является его высокая стоимость, так как для его проведения требуется привлечение сотрудников, а также низкую скорость обработки данных и вероятность ошибок, связанных с влиянием человеческого фактора [14].

Данная операция возможна, когда операция приведения данных к виду, доступному для обработки в указанной предметной области, является единичной. Когда поток обрабатываемых данных является непрерывным и требует постоянного выполнения указанных операций, эффективно использовать системы машинного обучения, в рамках которых искусственный интеллект позволяет проводить обработку записей в потоке информации [42].

Использование систем краудсорсинга в сфере торговли обеспечивает возможности разработки правил отнесения данных в массиве к различным категориям в части, например, распознавания вида пищевой продукции по фото, отнесению качественных признаков к заданному сорту. Примером использования данного метода является работа компании «Русагро», в которой используется оборудование по сортировке продукции, использующее систему распознавания, настройка которой проведена через метод краудсорсинга [51].

Системы смещения и интеграции данных.

Обработка систем больших данных зачастую предполагает необходимость получения потоков структурированной информации в различных форматах из различных источников.

В данном случае для проведения анализа необходимо приведение разнородной информации к единому формату. Операция смещения и интеграции предполагает выполнение операций, включающих [46]:

– настройку единой структуры информации, которую предполагается обрабатывать в рамках решения задач, связанных с использованием больших данных;

– слияние записей через поиск дополнительных сведений, соответствующих объекту, в таблицах, полученных из альтернативных источников;

– удаление данных, не являющихся объектом для анализа и не имеющих отношения к предметной области, либо данных, идентификация которых невозможно, либо принадлежность которой однозначно установить невозможно.

Интеграция данных проводится в информационных системах, где различные подразделения предприятий осуществляют обработку информации по смежным вопросам (например, интеграция может проводиться при работе с данными, полученными от отделов маркетинга, продаж, расчетов с поставщиками, кассовой информацией, данными, полученными из систем банковского документооборота др.).

Также интеграция данных бывает актуальна при реорганизации в государственных структурах или изменениях в системах работы с информацией (например, при необходимости анализа прав граждан на получение субсидий от органов соцзащиты необходимо получение данных из ПФР с использованием идентификатора, которым является СНИЛС. Для интеграции данный реквизит вводится в карточки льготополучателей в картотеках социальной защиты, после чего становится возможным проводить интеграцию данных из этих государственных структур) [10].

Технология проведения слияния информации включает этапы [33]:

- приведение к единому формату;
- установка связей между идентифицирующими атрибутами;
- выбор параметров, по которым проводится слияние;
- удаление посторонних записей;
- получение итогового файла.

Файлы, приведенные к единому формату, дают возможность анализа по выбранным показателям, предварительно проведя проверку корректности загрузки) [52].

На рисунке 3 приведена схема алгоритма смешения и интеграции данных.

Технологии смешения и интеграции данных при их использовании в технологии сфере торговли используются при интеграции данных о выпуске продукции, поступающих от производственного отдела и отдела сбыта, что позволяет отслеживать технологическую цепочку от производства до реализации. Примером внедрения данной технологии является работа компании «Мираторг» в технологической цепочке учета производства и сбыта продукции.

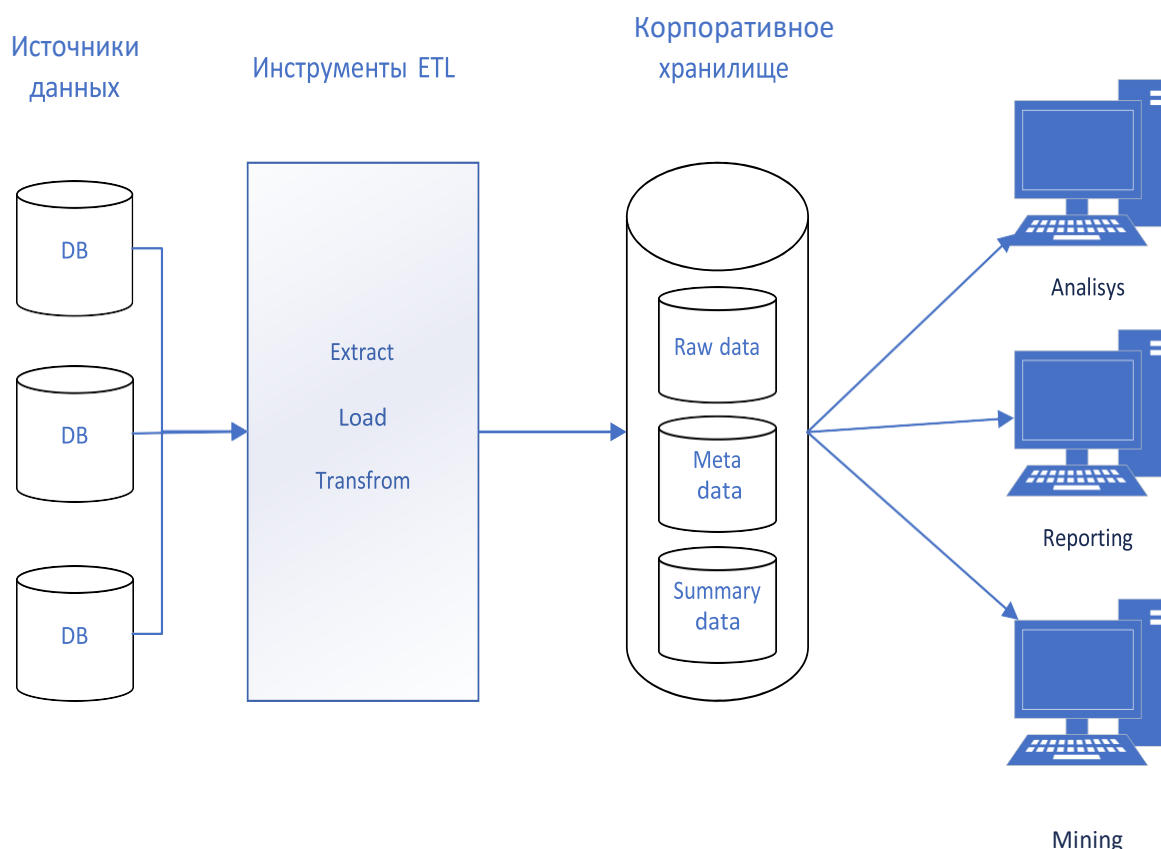


Рисунок 3 – Схема алгоритма смешения и интеграции данных
(составлено автором)

Далее проведено описание алгоритмов обработки больших данных методом машинного обучения и с использованием нейронных сетей [30].

Системы машинного обучения предполагают построение правил, позволяющих компьютерным системам проводить классификацию и систематизацию информации на основе результатов, полученных при обучении системы на основе анализа потоков данных. Порядок обработки данных через систему машинного обучения предполагает получение входного потока данных, по результатам обработки которого система выдает результат отнесения входных данных к заданной категории. Решение указанных задач возможно при предварительном проведении обучения системы (задания правил, в соответствии с которыми предполагается классификация объекта). В случае использования данной технологии в системах сфере торговли проводится анализ качественных характеристик реализуемого товара при отнесении их к определённому виду, сорту, признакам расчета стоимости.

Стадии настройки правил включают [40]:

- настройку нейронов, позволяющих проводить распознавание входных данных и определять связи между заданными признаками;
- анализ тестовой выборки;
- проверку функциональности системы и, при необходимости, исправление правил.

После обучения и тестирования система может использоваться при работе с массивами больших данных. На рисунке 4 приведена структура типовой нейросети.

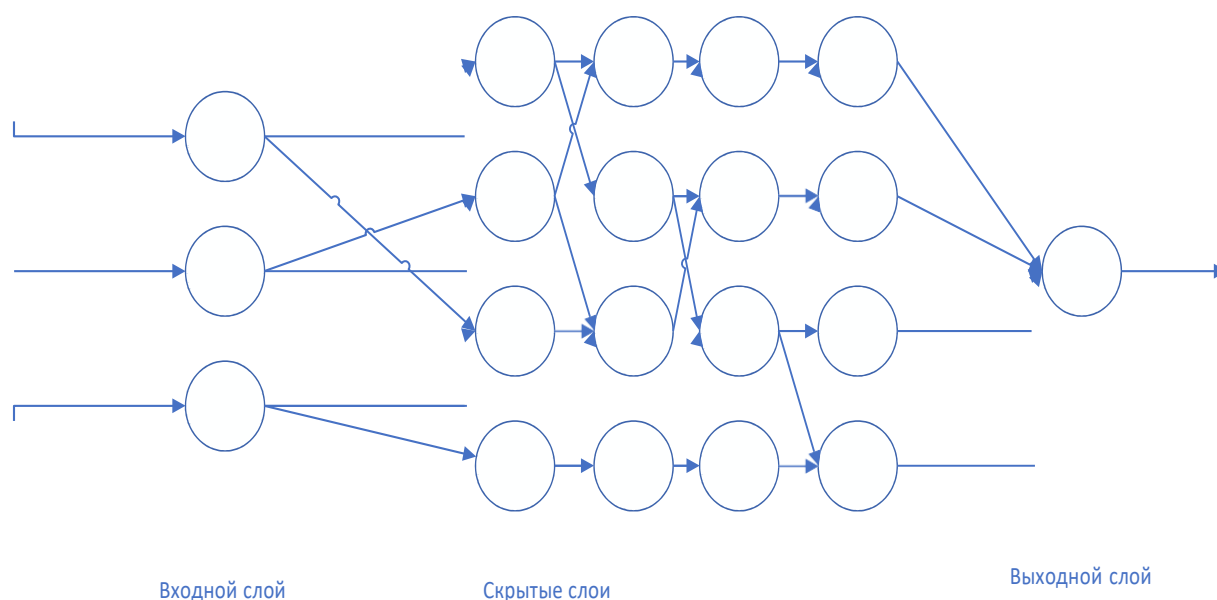


Рисунок 4 – Структура типовой нейросети (составлено автором)

Как показано на рисунке 4, порядок обработки информации включает: *подачу данных на вход*, обработку информации посредством работы с правилами, *выдачу обработанных данных*.

Использование нейросетей эффективно при необходимости проведения сортировки массивов, поддержки принятия решений в условиях неопределенности.

1.4 Анализ практик использования систем больших данных в сфере торговли

Пример использования нейросетей в сфере торговли – выявление брака в партии товара, получаемой от поставщиков. Примером внедрения данной технологии является работа компании «Юг Руси» в части приема и сортировки сырья для производства продукции определенных сортов.

Типовыми задачами для нейросетей являются: обработка изображений, выявление признаков активности злоумышленников, вредоносных систем в

системах информационной безопасности, признаков мошеннических действий в банковских информационных системах [17].

Технологии предиктивной аналитики при обработке больших данных.

Данные системы используются для решения задач прогнозирования поведения системы на основе массивов данных, полученных в динамике. Задачи данного типа предполагают выявление корреляции между различными параметрами потока данных, на основе которого строятся характеристики зависимостей, позволяющие вычислять вероятности поведения системы в перспективе. Использование данных систем позволяет рассчитывать прогнозные объемы по клиентской базе, спросу на продукцию определённого вида, анализировать поведение системы при различных параметрах кредитования, активности конкурентов. Полученные результаты позволяют своевременно принимать необходимые решения, корректировать политику компании, что обеспечивает защиту от негативных последствий в будущем. Эффективность данного метода обработки больших данных связана с отсутствием факторов форс-мажора, отсутствием явной зависимости моделируемого объекта от человеческого фактора, в отсутствие которых построенные прогнозы являются достаточно достоверными [9].

Примерами областей использования систем предиктивной аналитики являются [11]:

- в случае сфере торговли: планирование производства;
- результаты маркетинговых акций;
- получение данных по параметрам транспортного обслуживания компаний;
- выявление признаков вероятности неисполнения обязательств партнерами;
- динамика прибыльности и эффективности деятельности организаций.

Примером внедрения системы в предприятиях торговли является деятельность компании «Мираторг» в части анализа сбыта продукции и планирования ее реализации в различных клиентских сегментах.

Область использования систем больших данных также включает работу систем умного дома и умного производства, систем охраны, диспетчеризации. В рамках проведения обучения система проводит сбор информации о штатных и нештатных режимах работы технических систем, позволяют выявлять признаки неисправностей и определять порядок их устранения[22].

Далее проведено описание методов имитационного моделирования как метода обработки больших данных.

Специфика задач указанного типа связана с выявлением зависимостей параметров системы при изменении внешних факторов. Использование модели для данных целей позволяет без значительных финансовых затрат провести анализ поведения системы, получить данные по исследуемым характеристикам.

В случае сфере торговли имитационные методы могут использоваться при расчете параметров ожидаемого эффекта от производства продукции определенного вида в заданных объемах (моделируемой системой является состояние рынка сбыта). Существует опыт внедрения имитационных методов в деятельности ООО «КДВ Групп» в части расчета плановых показателей реализации товара.

Примеры имитационных моделей [6]:

- анализ параметров получения прибыльности банков при изменении ставок по кредитам и депозитам;
- анализ динамики задолженности по оплате за услуги ЖКХ при изменении тарифов;
- анализ демографической динамики при изменении параметров социальных программ;

– анализ состояния рынка труда в зависимости от количества мест в учебных заведениях.

На основании данных, полученных в имитационных моделях, проводится принятие управленческих решений, позволяющих решать поставленные задачи в области экономики, управления, социальной сферы.

В отличие от модели, рассмотренной выше, при работе с имитационными моделями, работа осуществляется не с данными, полученными из реальной системы, а со сгенерированными, что влияет на точность получаемых результатов. При этом получение реальных данных не всегда возможно в силу объективных обстоятельств.

Далее проведём обзор методов статистического анализа при обработке массивов больших данных.

Статистический аппарат при проведении анализа больших данных включает обработку, предполагающую операции по [13]:

- динамике показателей в абсолютном и относительном выражениях;
- получению статистических характеристик выборки;
- анализ средних значений различных видов;
- анализ распределения по выборке и др.

Использование статистического анализа в сфере торговли эффективно при расчёте показателей объёмов продаж товаров и выручки в динамике, составлении аналитической отчетности в качестве средства поддержки принятия управленческих решений. Использование алгоритмов статистического анализа характерно для большинства предприятий сферы торговли в службах экономического профиля, работы с поставщиками и клиентами. В статистическом аппарате, проводится изучение массовых социально-экономических явлений, каждое из которых может выражаться через различные количественные показатели по выбранным признакам, которые могут включать уровень цен, оплаты труда, налогов, выручки и др.

Выводы по разделу 1

В рамках проведённого анализа основных направлений использования систем интеллектуального анализа больших данных в сфере торговли было показано, что инструменты обработки больших массивов информации позволяют выявлять наличие зависимостей между изучаемыми параметрами, проводить расчёты статистических показателей, оценивать эффективность ассортиментной и бонусной политики. Таким образом, в рамках первого главы осуществлен обзор научных работ по теме исследования, рассмотрена общая характеристика технологий интеллектуального анализа данных, проанализированы методы обработки больших данных и практик использования систем больших данных в сфере торговли.

Глава 2 Анализ решения исследовательской задачи

2.1 Постановка задач обработки больших данных

Данные - это поддающееся многократной интерпретации представление информации в формализованном виде, пригодном для передачи, связи или обработки [11].

Таблица 1 - Характеристика 7V данных

Характеристика 7V данных	
Характеристика	Описание
«Объем (Volume)	К 2025 году мировой объем цифровых данных, по прогнозу компании IDC, вырастет до 175 зеттабайт (40 зеттабайт в 2021 году). Большая часть этих данных будет постоянно изменяться в реальном времени. Препных инструментов уже недостаточно. Для обработки требуются другие подходы и действия.
Скорость (Velocity)	Под скоростью понимается как сама скорость прироста данных, так и необходимость высокоскоростной обработки и получения результатов.
Разнообразие (Variety)	Обработка широкого спектра источников, типов и качества информации: изображения, видео- и аудиозаписи, текстовые файлы и базы данных. Эту информацию сначала обрабатывают и только потом анализируют»[2].
«Визуализация (Visualization)	Составление многомерных диаграмм и графиков для наглядного использования
Изменчивость (Volatility)	Значение данных может меняться. Например, смысловая нагрузка переводного текста в зависимости от обстоятельств
Ценность (Value)	Принятие верного решения на основе данных и получение из этого прибыли или ценное знание о клиентах.
Достоверность (Veracity)	Только на основе достоверных данных получается верный результат, который используется для принятия решений»[7].

Данные, которыми можно оперировать в Data Mining, можно разделить на семь категорий (Таблица 2).

Выделяют семь характеристик данных, которые показывают не только объем, но и другие категории, важные для обработки и анализа.

Перечень называют 7V по первым буквам характеристик [7].
Характеристики и описание к ним приведены в таблице 1.

Таблица 2 – Категории данных Data Mining

Категории данных Data Mining	
Наименование категории	Описание
«Структурированные	Зависят от модели данных и хранятся в фиксированных полях внутри записей
Неструктурированные	Трудно подогнать под конкретную модель, зависят от контекста или носят переменный характер
Графовые	Такие структуры данных предназначены для отображения отношений между объектами, их взаимосвязей и влияния
На естественном языке	Структурированные текстовые данные. Для отработки требуется знание лингвистики и специальных технологий
Аудио, видео, изображения	Данные этих форматов наиболее привычны пользователям. Их легко понимать, но для компьютера их обработка все еще – сложная задача
Машинные	Данные, генерируемые различными датчиками
Потоковые	Могут принимать почти любую из перечисленных в таблице форм, но поступают в систему при возникновении событий, а не загружаются в хранилище отдельными массивами»[8].

В рамках данной работы проведено решение задач по обработке массивов больших данных в динамике 2020-2022 гг., включающих:

- статистику реализации товара торговыми компаниями в разрезе видов товарных позиций, форматов магазинов, способов расчетов;
- получение выручки от реализации продукции;
- затратами на хранение продукции;
- штатной численности компаний;
- показатели продажи импортной продукции;
- получением прибыльности от деятельности, связанной с реализацией товаров;

– налоговыми перечислениями, связанными с ведением продаж товаров;

– получением субсидий на ведение продаж товаров;

– данные по задолженности по кредитам.

Полученная информация позволяет выявлять структуру реализации товара, прибыльность от реализации товаров отдельных видов, эффективность проводимой маркетинговой политики.

Также статистический анализ больших данных статистики производства и реализации пищевой продукции обеспечивает возможности анализа фонда заработной платы в отрасли, проведения расчета налоговой базы, занятости специалистов, динамики доходов и расходов, выявлять убыточные компании в динамике.

В сфере торговли часто приходится сталкиваться с количественными данными, которые можно собрать и упорядочить, т.е. структурировать. Но большая часть информации представлена документами, письмами, видео и др. - т.е. в неструктурированном виде. В работе нам понадобится выбирать подходящие инструменты и методы хранения, обработки и анализа таких данных.

«Для анализа выбранной совокупности по варьирующим(количественно изменяющимся) показателям в статистике используются средние величины» [49].

«Средние величины представляют собой обобщающую количественную характеристику совокупности однородных явлений в разрезе одного изменяющегося признака. В экономической практике используется широкий круг показателей, вычисленных в виде средних величин»[46].

«Примером обобщающих показателей доходов работников может являться значение среднего дохода одного работника, что определяется через отношение фонда заработной платы и выплат социального характера за анализируемый промежуток времени (в годовом, квартальном, месячном

выражении) к количеству сотрудников. Для сотрудников с сопоставимым уровнем доходов можно проводить расчет структуры затрат на определенные нужды. Аналогично можно проводить анализ рабочего времени, производительности труда и т.д.» [17].

«Одним из важнейших свойств средних величин является то, что с их использованием возможна интерпретация выбранного показателя через единый показатель, несмотря на количественные различия его у отдельных единиц в выборке, и позволяет выражать то общее, что присуще всем единицам анализируемой совокупности. Таким образом, через характеристику единицы совокупности средняя величина характеризует всю совокупность в целом»[53].

С использованием средних величин посредством математического аппарата закона больших чисел возможно выявление основных тенденций развития, установка закономерностей, анализ точности построенных моделей.

«С использованием средних величин возможно сравнение показателей, относящихся к совокупностям с различным количеством единиц. Важнейшим условием научного использования средних величин в статистическом анализе общественных явлений является однородность совокупности, для которой проводится вычисление средних»[74].

«В рамках проведения статистического анализа выделяются следующие виды средних величин:

1. Используемые весовые коэффициенты:

- а) невзвешенные средние величины;
- б) взвешенные средние величины.

2. По формулам вычисления:

- а) средние арифметические величины;
- б) средние гармонические величины;
- в) средние геометрические величины;
- г) средние квадратические, кубические и др. величины.

3. По технологии обработки выборки:

- а) групповые средние величины;
- б) общие средние величины»[14].

«Различие средних величин связано со спецификой учета характеристик, оказывающих влияние на влияющие на осредняемые величины:

Когда средние величины рассчитываются для признака, без учета влияния на него каких-либо других характеристик, то такие средние величины называются средними невзвешенными или простыми средними.

При наличии данных о наличии влияния на осредняемую величину некоторого количества факторов, которые необходимо учитывать при расчете для корректного расчета средней величины, то проводится расчёт средней взвешенной»[2].

«По форме расчета выделяется несколько типов средних величин, образованных из единой степенной средней величины, вычисляемой по формуле:

$$\bar{x} = \sqrt[k]{\frac{\sum_{i=1}^n x_i^k}{n}}, \quad (1)$$

Где \bar{x} – вычисляемое среднее значение;

k – значение показателя степени средней;

x – текущее значение (вариант) осредняемого признака;

i – i-тый элемент совокупности;

n – количество проведённых наблюдений (количество единиц в выборочной совокупности)» [37].

Для различных показателей степени k определяется тип вычисленных типов средних величин, описание которых приведено в таблице 3.

Алгоритм расчета среднего значения включает показатели [20]:

- Анализ исходного соотношения для анализируемой характеристики.
- Оценка недостающих данных для расчета исходного соотношения.

– Проведение расчета средней величины.

«Вместе с расчетом средней арифметической, в статистике применяется средняя гармоническая величина, обратная средней арифметической из обратных значений признака. Как и средняя арифметическая, она может быть простой и взвешенной. Используется она в тех случаях, когда необходимые весовые коэффициенты (f_i) в исходных данных не заданы непосредственно, а входят в качестве сомножителей водни из имеющихся показателей»[18].

Таблица 3 - Типы вычисляемых типов средних величин в зависимости от степени

Степень вычисляемой средней величины (k)	Наименование вычисленной средней
-1	Среднее гармоническое
0	Среднее геометрической
1	Среднее арифметическое
2	Среднее квадратичное
3	Среднее кубическое

«Средняя гармоническая простая рассчитывается по формуле

$$x = \frac{n}{\sum \frac{1}{x_i}}, \quad (2)$$

Таким образом, она является обратной величиной средней арифметической простой из обратных значений признака. Областью применения данного показателя является расчёт временных затрат на выполнение технологических операций»[59].

«Использование средней геометрической осуществляется в случаях, когда отдельные значения признаков являются, как правило, относительными

величинами динамики, которые строятся в формате цепных величин, и могут представляться как отношение к предыдущему уровню каждого уровня в ряду динамики, т.е. описывает средние коэффициенты роста»[8].

«Вычисление средней геометрической производится через извлечение корня степени из произведений отдельных значений — вариантов признака x :

$$\bar{x} = \sqrt[n]{x_1 * x_2 * \dots * x_n} = \quad (3)$$

где n — количество вариантов; $\sqrt[n]{\prod x_i}$

\prod — оператор произведения» [63].

«В некоторых случаях в решении экономических задач возникает необходимость в расчете средних размеров признаков, выраженных через квадратные или кубические единицы измерения (когда при описании предметной области учитываются значения второй или третьей степени вычисляемого показателя). В таких случаях используются средние квадратические (например, при расчете средних величин сторон и квадратных участков, средних значений диаметров труб, стволов и др.) и средние кубические (например, при расчете средней длины стороны и кубов)»[30].

«Средняя квадратическая простая является квадратным корнем из частного от деления суммы квадратов отдельных значений признака на их число:

$$\bar{x} = \sqrt{\frac{x_1^2 + x_2^2 + \dots + x_n^2}{n}} = \sqrt{\frac{\sum x_i^2}{n}}, \quad (4)$$

где x_1, x_2, \dots, x_n - значения признака,

n - их количество» [63].

«Формула расчёта средней квадратической взвешенной:

$$\bar{x} = \sqrt{\frac{\sum x_i^2 f_i}{\sum f_i}}, \quad (5)$$

где f -весовые коэффициенты»[22].

«При расчёте средней кубической вычисляется кубический корень из частного при делении суммы кубов отдельных значений признака на их количество:

$$x = \sqrt[3]{\frac{\sum x_i^3}{n}}, \quad (6)$$

где x_1, x_2, \dots, x_n - значения признака, n - их количество»[63].

«Формула расчета средней кубической взвешенной:

$$\bar{x} = \sqrt[3]{\frac{\sum x_i^3 f_i}{\sum f_i}}, \quad (7)$$

где f -весовые коэффициенты»[27].

Использование средней квадратичной и средней кубической имеет небольшую область применения и включает расчёт показателей вариации или расчет характеристик при обработке экспериментальных данных [25].

«Вычисление средних показателей возможно не для всей выборки, а для некоторого ее сегмента. В качестве примера такого вида средних можно рассматривать быть среднюю прогрессивную как элемент из частных средних, вычисляемый не для всех, а только для "лучших" (например, для показателей выше или ниже средних индивидуальных)»[27].

«Для характеристики структуры вариационных рядов используются структурные средние. В экономической статистике наиболее часто используемыми показателями являются мода и медиана»[63].

Мода – показатель случайной величины, который встречается в изучаемой выборке с максимальной вероятностью, либо его частота в выборке является наибольшей [27].

«В интервальных вариационных рядах моду определяют приближенно по формуле:

$$M_o = x_{M_o} + i_{M_o} \cdot \left(\frac{f_{M_o} - f_{M_o-1}}{f_{M_o} - f_{M_o-1} + f_{M_o} - f_{M_o+1}} \right).$$

$$\overline{f_{M_0} - f_{M_0-1}} + \overline{f_{M_0} - f_{M_0+1}}, \quad (8)$$

где x_{Mo} - минимальный показатель выборки, содержащей моду;

i_{Mo} - значение величины модального интервала;

f_{Mo} - показатель частоты модального интервала;

f_{Mo-1} - показатель частоты интервала, предшествующего модальному;

f_{Mo+1} - показатель частоты интервала, следующего за модальным» [63].

«Мода используется при решении некоторых прикладных задач. Так, например, при изучении товарооборота рынка используется модальная цена, при изучении уровня спроса на обувь, одежду используются модальные размеры обуви и одежды и др.»[43].

«Медиана — это численное значение признака у той единицы совокупности, находящейся в середине ранжированного ряда (построенного в порядке возрастания, либо убывания значения изучаемого признака). Медиана представляет собой серединную варианту, т.к. она делит совокупность на две одинаковые части»[71]. Например, в экономике используется параметр медианной заработной платы, которая характеризует реальный размер оплаты труда в организации (позволяя не учитывать верхний сегмент с высокими показателями заработной платы) [15].

Повышение точности уровня статистических измерений обеспечивается через рост объема анализируемой выборки. Таким образом, статистическая обработка больших данных позволяет получать результаты с высоким уровнем достоверности [26].

2.2 Описание и обоснование выбора средств автоматизации обработки больших данных

«OLAP (Online Analytical Processing) - это система аналитической обработки данных. Она предназначена для подготовки отчетности, построения прогностических сценариев и выполнения статистических

расчетов на базе больших информационных массивов, обладающих сложной структурой»[34].

«Компонентами OLAP являются:

- база данных (БД) - источник, из которого осуществляется загрузка данных для проведения обработки. Формат базы данных определяется разновидностью OLAP системы и порядком выполнения действий OLAP сервера. Чаще всего пользуются реляционными и многомерными БД и хранилищами данных;

- OLAP сервер - ядро системы, с помощью которого проводится обработка многомерных структур данных, и обеспечивается связь между БД и пользователями систем;

- приложения для работы пользователей, в которых формируются запросы и визуализируются полученные ответы.

- Специфика обработки данных OLAP системами состоит в построении многомерных, то есть имеющих большое количество связей между отдельными элементами, массивов информации»[55].

«Для формирования таких массивов OLAP система собирает данные из различных источников (например, из хранилищ данных, из информационных систем управления предприятием (ERP) или из системы взаимодействия с клиентами (CRM)). После этого информация обрабатывается на OLAP сервере и передается в пользовательские приложения»[18].

«Хранение и обработка данных с применением OLAP систем могут осуществляться:

- непосредственно на рабочих местах пользователей;
- в форме реляционных баз данных - при совместной работе OLAP систем с ERP или CRM-системами;
- в форме многомерных баз данных на обособленных серверах»[82].

«OLAP системы используются главным образом в сфере принятия стратегических управленческих решений и используются для

бюджетирования, построения прогностических моделей, подготовки финансовой отчетности, хранения результатов»[26].

В качестве платформы для анализа данных выбран язык программирования Python, возможности которого позволяют создавать оконные, мобильные приложения, Web-приложения. Также инструментарий данной платформы позволяет решать задачи, связанные с анализом данных. Python имеет множество полезных библиотек, такие как Matplotlib, Pandas, NumPy [15].

В качестве платформы-интерпретатора выбрана платформа Google Colab. Данная платформа обеспечивает возможности коллективной работы с информационными ресурсами различных форматов, включая возможности разработки приложений на языке Python.

Одним из наиболее распространённых направлений использования указанной среды является работа с большими данными, что включает обработку массивов информации, их группировку, статистическую обработку, возможности по прогнозированию, изучению параметров выборок и распределений.

«Для использования среды Python в качестве инструмента анализа данных необходимо использование специализированных библиотек, в возможности которых входят инструментов статистики, графические сервисы и другие технологии, позволяющие проводить обработку информации»[31].

В качестве инструментов анализа данных в магистерской диссертации выбраны среды Pandas, Matplotlib, Numpy.

Pandas

«Данная библиотека позволяет работать с технологиями анализа данных, обеспечивает высокий уровень производительности при проведении обработки больших массивов информации. Система имеет открытый код, что позволяет разрабатывать собственные процедуры обработки данных»[22].

«Библиотека Pandas используется в различных академических и коммерческих областях, включая искусственный интеллект (Artificial Intelligence), финансы, нейробиологию, экономику, статистику, рекламу, веб-аналитику и многое другое. Рассмотрим основные возможности библиотеки:

DataFrame- быстрый и эффективный инструмент для манипулирования данными со встроенной индексацией. Методы, требующие высокой производительности, написаны на C или Cython.

Позволяет читать и записывать данные разных форматах: CSV (comma-separated values), таблицы Excel, базы данных SQL, иерархический формат HDF. Всего насчитывается 19 поддерживаемых форматов.

Удобный инструмент для работы с отсутствующими данными»[16].

«Также стоит отметить следующее немаловажные факторы как:

- простое управление беспорядочными данными в упорядоченной форме;
- гибкое изменение форм: добавление, удаление, присоединение новых или старых данных;
- интеллектуальное индексирование, манипулирование и управление столбцами и строками;
- мощный инструмент для агрегирования и преобразования данных, в том числе и большого размера (BigData);
- быстрое слияние и объединение наборов данных, например, два и более объектов DataFrame;
- поддержка иерархического индексирования, то есть возможность объединения столбцов под общей категорией (MultiIndex);
- поддержка работы с датами и временем»[28].

Matplotlib

Matplotlib – является основной библиотекой, позволяющей осуществлять визуализацию данных, обработка которых проведена в среде Python.

Поддерживаются линейный, полиномиальный форматы, гистограммы, 3-D модели и другие форматы визуализации данных.

Requests

«Данная модель обеспечивает возможности создания Web-приложений, обмениваться данными с Web-серверами»[34].

«Функционал библиотеки включает:

- возможности работы с HTTP-подключениями;
- работу с Cookie, что включает передачу и получение данных в формате идентификатор – наименование;
- работа с кодированием данных;
- работа с пользовательскими учетными записями;
- поддержка файлового обмена,
- другие сервисы, используемые в рамках работы с Web-ресурсами»[27].

Numpy

Numeric Python (NumPy) - это система, используемая для расчетов характеристик многомерных данных, используемых в приложениях, где необходима статистическая обработка информации. Входные данные должны быть проиндексированы, размерность их может задаваться пользователем. Запись массива в память осуществляется в соответствии с заданными настройками и проводится либо в определенном порядке, либо беспорядочно [29].

Отличие массива от набора данных (списка или кортежа). Величины, входящие в массив, являются однотипными и их число определяется в процессе инициализации. Элементы массива не являются объектами, это переменные в обычном понимании данного термина. Использование массивов позволяет оптимизировать системные ресурсы при работе с потоками данных. Данная библиотека позволяет проводить математические операции с данными различных типов и объема. Реализованы возможности

обработки массивов, обработки данных с помощью аппарата линейной алгебры, преобразований различного вида [84].

«Анализ данных – это комплекс математических дисциплин, в котором проводятся построения и исследования универсальных математических методов и вычислительных алгоритмов извлечения знаний из экспериментальных (в широком смысле) данных»[20].

В практической части работы необходимо осуществить интеллектуальный анализ данных для решения задачи по оптимизации ассортиментной политики и снижению издержек предприятий сферы торговли.

Вопросы оптимизации ассортиментной политики для торгового предприятия имеют особое значение, поскольку непродуманные решения относительно ассортимента способствуют избыточным скоплением товара, снижением рентабельности. Таким образом, оптимизация ассортиментной политики приводит к снижению издержек у предприятий сферы торговли.

Для решения данной задачи осуществим интеллектуальный анализ данных. В достижении этой цели произведем обработку массивов слабоструктурированных данных, полученных по данным статистики продаж за период 2020-2022 годов в целях анализа статистических показателей в области продаж, выявления закономерностей и зависимостей.

Этапы проведения интеллектуального анализа данных в целях решения поставленной задачи:

- постановка задачи;
- подготовка данных;
- изучение данных;
- построение моделей;
- исследование и проверка моделей.

В рамках проведения оптимизации ассортиментной политики целесообразно провести анализ динамики рынка, структуры спроса, финансовых показателей по товарным группам, отдельным товарам.

В практической части работы будет выполнен интеллектуальный анализ данных по продажам товаров различной номенклатуры. Анализ будет производиться по трем массивам слабоструктурированных данных, затем будут построены соответствующие модели.

В качестве инструментов интеллектуального анализа выбраны:

- группировка;
- вычисление относительных величин;
- графическое представление данных;
- корреляция;
- кластеризация.

Выводы по второму разделу

Таким образом, интеллектуальный анализ данных представляет собой процесс обнаружения пригодных к использованию сведений в крупных наборах данных. В интеллектуальном анализе данных применяется математический анализ для выявления закономерностей и тенденций, существующих в данных. В рамках проведённого анализа основных направлений использования систем интеллектуального анализа больших данных в сфере торговли было показано, что инструменты обработки больших массивов информации позволяют выявлять наличие зависимостей между изучаемыми параметрами, проводить расчёты статистических показателей, оценивать эффективность ассортиментной и бонусной политики. Следовательно, технологии интеллектуального анализа позволяют получать конкурентные преимущества для предприятий сферы торговли вследствие наличия возможностей проведения оптимизации технологии работы с поставщиками, клиентами, торговыми площадями и персоналом.

Проведена постановка задач, связанных с расчетом показателей по продажам товаров в зависимости от различных показателей. Проведён выбор инструментария для проведения анализа. Также определены критерии и показатели, по которым предполагается проведение расчётов.

Следовательно, использование системы статистической обработки информации обеспечивает возможность анализа деятельности предприятий, оптимизации структуры производства, финансирования основной деятельности, эффективного использования кредитных ресурсов и государственных субсидий.

В качестве платформы для анализа данных выбран язык программирования Python, возможности которого позволяют создавать оконные, мобильные приложения, Web-приложения. Также инструментарий данной платформы позволяет решать задачи, связанные с анализом данных. Python имеет множество полезных библиотек, такие как Matplotlib, Pandas, NumPy [15].

В качестве платформы-интерпретатора выбрана платформа Google Colab, так как данная платформа обеспечивает возможности коллективной работы с информационными ресурсами различных форматов, включая возможности разработки приложений на языке Python. В качестве инструментов анализа данных выбраны среды Pandas, Matplotlib, Numpy.

Одним из наиболее распространённых направлений использования указанной среды является работа с большими данными, что включает обработку массивов информации, их группировку, статистическую обработку, возможности по прогнозированию, изучению параметров выборок и распределений.

Глава 3 Статистический анализ данных предприятий торговли с использованием методов обработки больших данных

3.1 Постановка задач обработки больших данных

На этапе постановки задачи определяем цель - проведение интеллектуального анализа данных для решения задачи по оптимизации ассортиментной политики и снижению издержек предприятий сферы торговли. В качестве исходных данных для анализа взят массив больших данных в динамике 2020-2022 гг., включающих:

- статистику реализации товара торговыми компаниями в разрезе видов товарных позиций, форматов магазинов, способов расчетов;
- получение выручки от реализации продукции;
- затратами на хранение продукции;
- штатной численности компаний;
- показатели продажи импортной продукции;
- получением прибыльности от деятельности, связанной с реализацией товаров;
- налоговыми перечислениями, связанными с ведением продаж товаров;
- получением субсидий на ведение продаж товаров;
- данные по задолженности по кредитам.

Такая структура данных для интеллектуального анализа не является случайной. Для корректного решения поставленной задачи целесообразно провести анализ динамики рынка, структуры спроса, различных финансовых показателей.

Полученная информация позволяет выявлять структуру реализации товара, прибыльность от реализации товаров отдельных видов.

Также интеллектуальный анализ больших данных на основе статистики производства и реализации товаров обеспечивает возможности анализа фонда заработной платы в отрасли, динамики доходов и расходов, выявления убыточных торговых компаний в динамике.

«Программная реализация систем больших данных предполагает необходимость использования специализированных программных средств, которые обеспечивают возможности по итогам обработки данных построение аналитической отчетности, построения моделей состояния системы предприятий при изменении определенных параметров. Таким образом, при внедрении систем Big Data в компаниях сферы торговли необходимо провести выбор аппаратных, программных решений, СУБД, а также систему настройки шаблонов обработки»[13].

На рисунке 5 приведена BPMN-диаграмма бизнес-процесса использования технологий больших данных. В работе системы предполагается наличие сценариев аналитика и ИТ-специалиста. Сценарий аналитика включает работу аналитика, к которой относится получение и обработку массива данных, формирование отчетности. Сценарий ИТ-специалиста включает настройку скриптов обработки и шаблонов вывода данных.

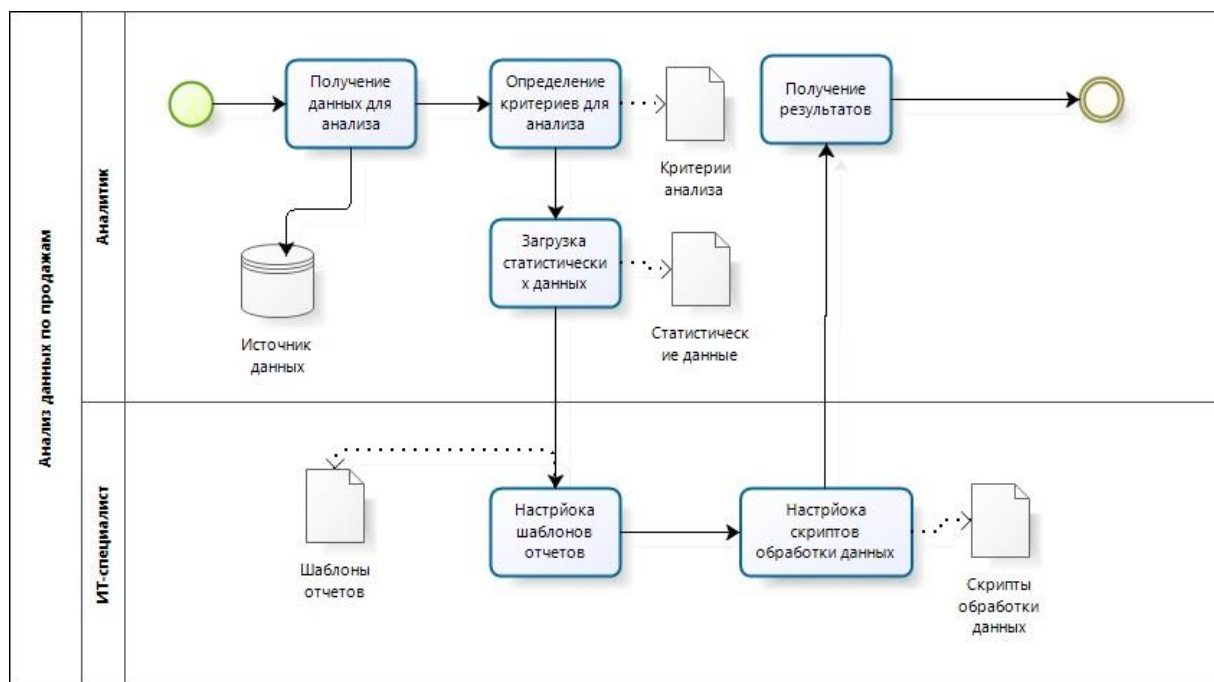


Рисунок 5 – BPMN-диаграмма бизнес-процесса использования технологий больших данных

На рисунке 6 приведена диаграмма вариантов использования технологий больших данных.

На рисунке 7 приведена диаграмма компонентов. На

рисунке 8 представлена диаграмма деятельности.

Результаты проделанной работы позволили сделать выводы о том, что системы больших данных при использовании в компаниях сфере торговли обеспечивают возможности решения задач различного направления. Результатом обработки больших данных может быть выработка оптимальной ассортиментной политики, определение путей для снижения издержек, определение оптимальных поставщиков и логистических цепочек.

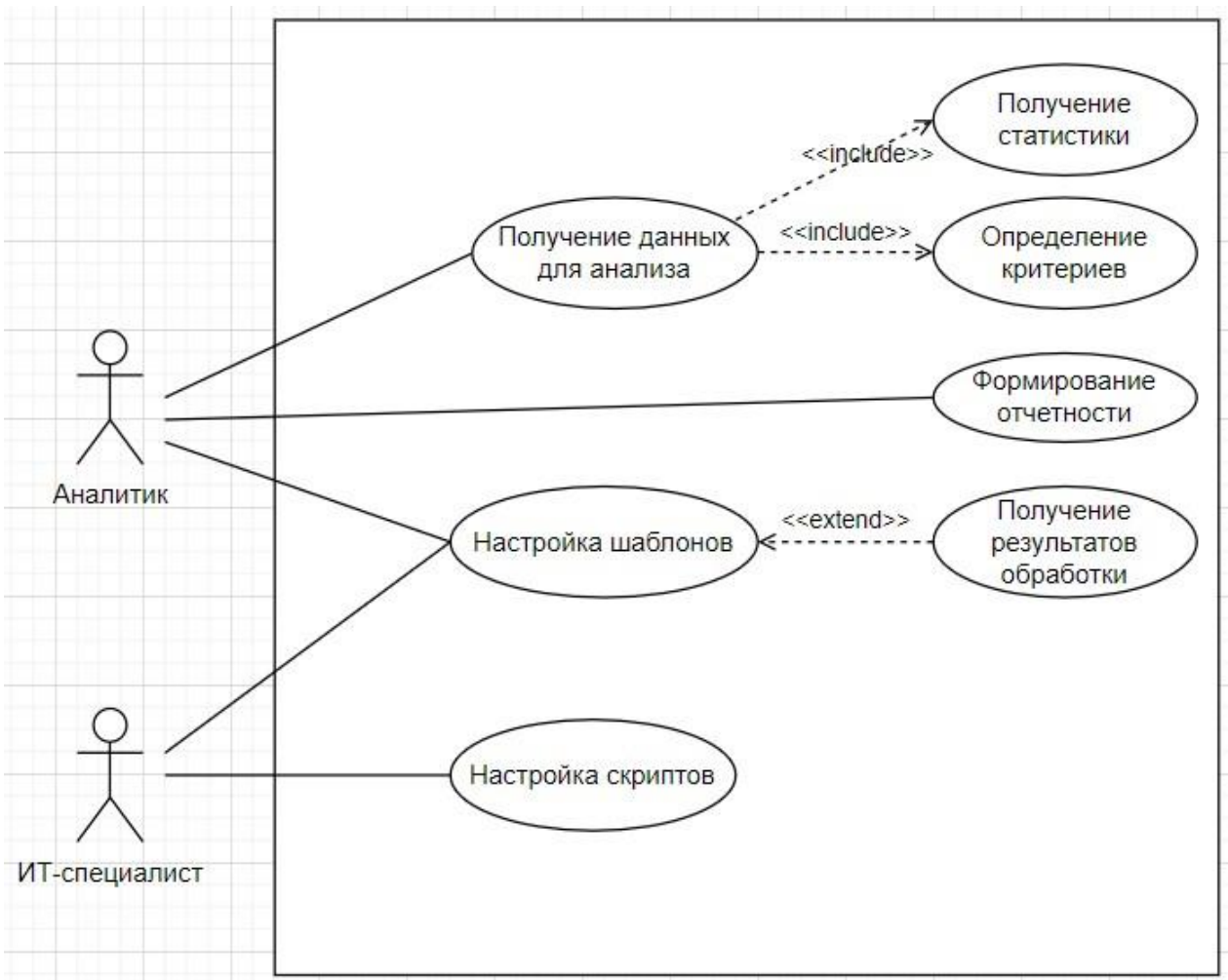


Рисунок 6 – Диаграмма вариантов использования технологий больших данных

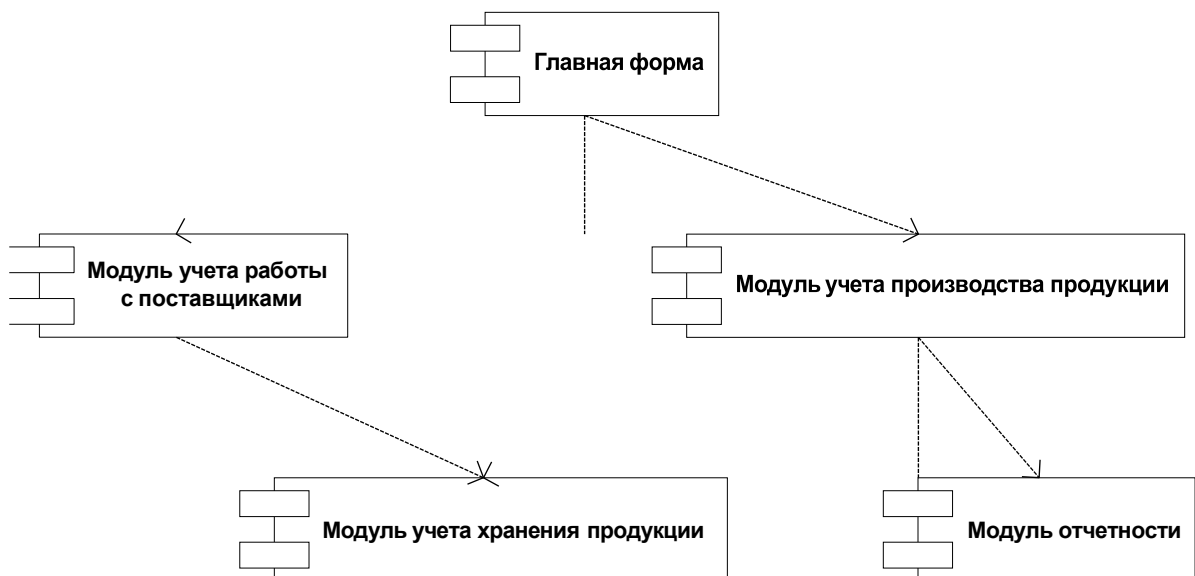


Рисунок 7 – Диаграмма компонентов

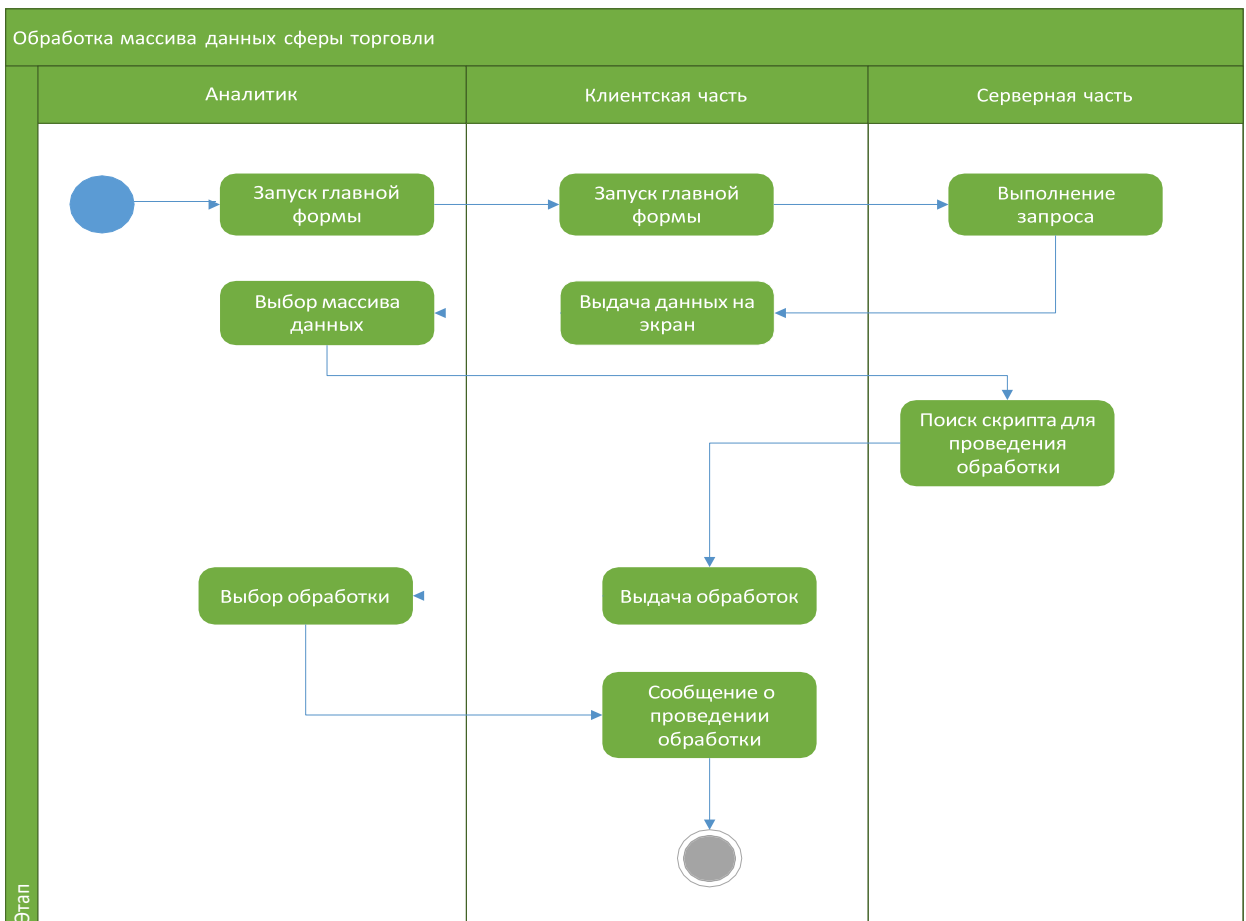


Рисунок 8 – Диаграмма деятельности

3.2 Описание средств автоматизации обработки больших данных

«OLAP (Online Analytical Processing) - это система аналитической обработки данных. Она предназначена для подготовки отчетности, построения прогностических сценариев и выполнения статистических расчетов на базе больших информационных массивов, обладающих сложной структурой»[34].

«Компонентами OLAP являются:

- база данных (БД) - источник, из которого осуществляется загрузка данных для проведения обработки;

- OLAP сервер - ядро системы, с помощью которого проводится обработка многомерных структур данных, и обеспечивается связь между БД и пользователями систем;
- приложения для работы пользователей, в которых формируются запросы и визуализируются полученные ответы.
- Специфика обработки данных OLAP системами состоит в построении многомерных, то есть имеющих большое количество связей между отдельными элементами, массивов информации»[55].

Архитектурно – технологическая схема OLAP представлена на рисунке 9.

В качестве платформы для анализа данных выбран язык программирования Python, возможности которого позволяют создавать оконные, мобильные приложения, Web-приложения. Также инструментарий данной платформы позволяет решать задачи, связанные с анализом данных. Python имеет множество полезных библиотек, такие как Matplotlib, Pandas, NumPy, Sklearn.cluster.

В качестве платформы-интерпретатора выбрана платформа Google Colab. Данная платформа обеспечивает возможности коллективной работы с информационными ресурсами различных форматов, включая возможности разработки приложений на языке Python.

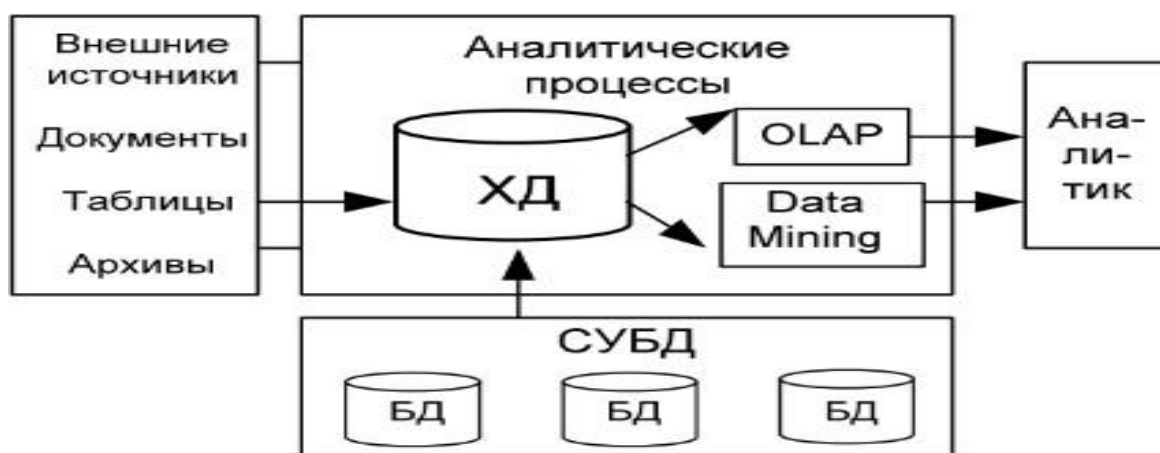


Рисунок 9 - Архитектурно – технологическая схема OLAP[Составлено автором]

Одним из наиболее распространённых направлений использования указанной среды является работа с большими данными, что включает обработку массивов информации, их группировку, статистическую обработку, возможности по прогнозированию, изучению параметров выборок и распределений.

«Для использования среды Python в качестве инструмента анализа данных необходимо использование специализированных библиотек, в возможности которых входят инструменты статистики, графические сервисы и другие технологии, позволяющие проводить обработку информации»[31].

В качестве инструментов анализа данных выбраны среды Pandas, Matplotlib, Numpy, Sklearn.cluster.

Выводы по разделу

Проведена постановка задач, связанных с расчетом показателей по продажам товаров в зависимости от различных показателей. Проведён выбор инструментария для проведения анализа. Также определены критерии и показатели, по которым предполагается проведение расчётов.

Глава 4 Апробация результатов исследования

4.1 Проведение статистической обработки массива данных

На этапе подготовки данных определены источники данных для осуществления интеллектуального анализа - массивы больших данных в динамике 2020-2022 гг.

В соответствии с поставленными задачами проведен интеллектуальный анализ массива данных, структура которого включает:

- наименование компании, работающей в сфере торговли;
- данные о видах реализуемых товаров;
- данные о реализации видов продукции по временным периодам;
- данные о налоговых выплатах;
- данные о реализации импортной продукции;
- данные о качестве реализуемого товара;
- данные о затратах на хранение товара;
- данные о количестве работников.

В таблице 4 приведён фрагмент источника данных, представленных для анализа.

Таблица 4 - Фрагмент источника данных, представленных для анализа

Компания	Продукт	Выручка	Затраты	Налоги	Количество работников	Процент брака	Год	Процент импорта
ООО Продукты	Молоко	9301184	1225232	1760382	147	10	2021	12
ООО Продукты	Масло	5336586	1912387	2905812	188	3	2020	6
ИП Иванов	Чай	2578295	1790496	2924583	61	5	2022	3
ООО Магазин	Творог 5%	9239042	2876587	600775	212	2	2022	6
ИП Васильев	Макаронные изделия	1087870	462651	659082	208	3	2022	5

Продолжение таблицы 4

Компания	Продукт	Выручка	Затраты	Налоги	Количество работников	Процент брака	Год	Процент импорта
ИП Семенов	Чай	9367546	238297	2317842	223	10	2022	9
ООО Ромашка	Макаронные изделия	5676465	1697983	1695799	234	4	2021	3
ООО У дома	Белый хлеб	3987197	2519305	2556902	232	2	2020	1
ООО Продукты	Плавленый сыр	2589544	1689138	1598052	210	9	2021	10
ИП Иванов	Белый хлеб	3009772	2947162	1863158	3	7	2021	20
ИП Петров	Крупы	3750347	1184581	496880	134	5	2022	10
ИП Козлов	Кофе	9799319	860076	555102	80	7	2020	6
ООО Люкс	Белый хлеб	8150273	743536	999158	200	1	2020	20
ООО Ярче	Творог 2%	2418934	1092164	1186410	295	5	2021	4
ООО Ландыш	Макаронные изделия	5666680	2428407	963746	31	8	2022	10
ООО Магазин	Сок яблочный	9008428	1960802	1433985	88	9	2020	18
ООО Магазин	Плавленый сыр	8480525	2867499	28108	235	8	2020	9
ООО У дома	Сок персиковый	4384557	258051	1489288	103	8	2022	1
ИП Иванов	Творог 2%	6167590	910818	2400293	202	8	2022	14
ИП Козлов	Макаронные изделия	7439872	1786718	2440809	205	1	2020	1
ООО Ландыш	Белый хлеб	1010421	212427	2836459	39	10	2021	6
ИП Иванов	Плавленый сыр	1715985	1481183	2145368	48	6	2021	2
ООО Магазин	Черный хлеб	3212318	2429985	1518413	151	9	2021	3
ООО Продуктовый	Чай	6966487	329482	1756894	55	5	2021	19
ООО Люкс	Масло	7496274	724786	2209979	131	6	2022	12
ИП Семенов	Масло	2626351	2524439	1450863	56	7	2021	10
ООО Ромашка	Белый хлеб	4422753	103536	2478473	143	8	2020	11
ИП Петров	Плавленый сыр	3270107	1657380	722463	27	4	2022	11
ООО Магазин	Кофе	9573184	1383305	676259	244	4	2020	14
ООО У дома	Крупы	3580395	60340	2243536	85	7	2020	7
ООО Ромашка	Творог 2%	7384536	2577771	1363862	127	7	2020	4
ООО Ландыш	Кофе	6137417	1171657	2653278	105	8	2021	2
ИП Иванов	Масло	8336010	2170783	2696853	292	2	2022	10

Продолжение таблицы 4

Компания	Продукт	Выручка	Затраты	Налоги	Количество работников	Процент брака	Год	Процент импорта
ИП Петров	Сок яблочный	7796584	906015	2691370	291	7	2021	3
ИП Иванов	Чай	5623017	1812982	2890711	258	6	2020	3
ООО Продукты	Макаронные изделия	8267940	890545	686325	40	7	2020	10
ООО Магазин	Плавленый сыр	2828205	302984	1392855	204	6	2022	5
ИП Петров	Белый хлеб	8810629	2299245	1139720	186	2	2021	10
ООО Продуктовый	Чай	9267385	965438	6435	100	10	2021	14
ИП Петров	Творог 2%	2349031	435079	2345615	29	10	2020	19
ООО Продукты	Творог 2%	4160153	1562885	505278	210	9	2020	18
ООО Магазин	Сок яблочный	8149858	239429	935603	139	2	2021	12
ООО У дома	Черный хлеб	200415	2411829	2720237	112	7	2021	7
ООО Ландыш	Черный хлеб	5898218	816741	2115668	35	5	2020	18
ООО Магазин	Крупы	4092486	1302972	1478500	193	7	2021	18
ИП Семенов	Кофе	5688591	1017854	2022799	250	2	2020	9
ИП Петров	Творог 5%	3752999	2612745	410922	195	1	2022	15
ООО Продуктовый	Масло	1392775	959427	2347690	224	6	2020	16
ИП Петров	Сок персиковый	3243759	2191603	1171186	53	10	2021	15
ООО Ромашка	Творог 2%	6733745	1899930	1757471	229	2	2020	7
ООО Ландыш	Черный хлеб	8668669	1813167	1402211	81	9	2020	5
ИП Васильев	Черный хлеб	6632968	371920	2416094	277	2	2020	2
ООО Магазин	Молоко	4128831	2485582	1294763	285	7	2021	10
ИП Васильев	Белый хлеб	9031336	441404	543381	288	1	2020	10
ООО У дома	Плавленый сыр	3064937	1717145	2561814	285	5	2020	5
ООО Продукты	Творог 5%	7397904	49272	402868	169	1	2020	18
ООО Магазин	Масло	6693557	1692798	454565	10	7	2020	16
ООО Ландыш	Плавленый сыр	8470000	2131736	1147653	114	10	2022	8
ООО Продукты	Белый хлеб	4057643	2289029	2487836	170	2	2022	16
ИП Козлов	Крупы	6345739	641940	2786101	158	7	2021	18

Продолжение таблицы 4

Компания	Продукт	Выручка	Затраты	Налоги	Количество работников	Процент брака	Год	Процент импорта
ООО Продуктовый	Сок персиковый	126279	1618797	1576245	20	8	2020	3
ИП Семенов	Черный хлеб	9120544	1709911	1387451	122	3	2022	2
ИП Петров	Макаронные изделия	2831347	2880683	103064	12	6	2022	11
ИП Петров	Кофе	2421572	2285686	1845529	8	2	2021	12
ИП Семенов	Черный хлеб	6942285	2735996	1618437	207	7	2021	10
ИП Иванов	Творог 2%	1490704	215154	1077543	189	10	2020	13
ООО Люкс	Сок персиковый	4871382	462937	844120	130	1	2022	5
ООО Люкс	Кофе	9579138	2432811	1821673	149	4	2022	9
ИП Козлов	Масло	782467	2920843	1627930	169	1	2021	5
ИП Иванов	Молоко	2380322	762363	2241027	199	7	2020	3
ИП Козлов	Черный хлеб	5557674	221753	1022357	285	10	2021	18
ООО Продукты	Творог 5%	3537861	2333358	2404176	17	10	2020	3
ООО Продуктовый	Черный хлеб	8157772	1844461	1813844	57	7	2022	11
ООО Продуктовый	Черный хлеб	2645328	1309732	1831949	112	6	2020	1
ООО Продукты	Творог 2%	3543025	126920	2386140	287	6	2021	9
ИП Козлов	Масло	9521485	245060	815403	274	7	2021	20
ООО Люкс	Белый хлеб	9662871	1765893	2901109	177	10	2022	17
ООО У дома	Чай	7838071	2849537	1254897	14	1	2021	5
ООО Люкс	Творог 2%	8086893	1735130	549649	99	4	2022	5
ИП Козлов	Кофе	8797881	1070362	563562	128	2	2021	15
ООО Продукты	Кофе	3550990	1261848	548455	132	4	2022	19
ООО Люкс	Молоко	8981862	856414	1729748	45	10	2020	12
ИП Иванов	Плавленый сыр	6966221	798838	2237889	72	5	2021	15
ООО Продукты	Крупы	3746800	1423383	2393423	127	9	2022	14
ООО Продукты	Крупы	4034090	235096	2640143	97	3	2022	17
ООО Люкс	Кофе	3470881	983828	659131	162	8	2020	15
ИП Васильев	Черный хлеб	850634	1317261	750771	3	6	2021	2
ООО У дома	Творог 5%	7816296	341355	435243	211	5	2020	11

Продолжение таблицы 4

Компания	Продукт	Выручка	Затраты	Налоги	Количество работников	Процент брака	Год	Процент импорта
ООО Продуктовый	Макаронные изделия	9974455	1937270	1925490	117	2	2021	14
ООО Люкс	Белый хлеб	5646312	682526	895815	237	1	2022	18
ООО Продукты	Сок персиковый	1174431	195089	14485	200	1	2021	7
ИП Семенов	Плавленный сыр	947720	935367	15056	293	9	2022	2
ООО Продуктовый	Черный хлеб	4952732	477045	1159896	194	2	2022	8
ИП Васильев	Макаронные изделия	8968861	1220197	538104	182	4	2020	18
ИП Иванов	Сок персиковый	4144179	1284645	2159799	6	1	2022	1
ООО Люкс	Белый хлеб	8820287	2979599	1665155	98	9	2021	15
ООО У дома	Сок персиковый	6901474	520851	1976687	226	5	2021	14
ИП Козлов	Сок яблочный	7894675	11437	2580771	226	6	2022	17
ООО Люкс	Масло	2358207	1603693	1887247	153	2	2020	9
ООО У дома	Творог 2%	1607584	89486	1811758	176	4	2022	14
ИП Семенов	Творог 2%	7551537	445844	185029	44	5	2020	11
ООО Люкс	Сок яблочный	59664	994907	17568	194	10	2022	6
ООО Люкс	Макаронные изделия	201612	1837329	155638	106	9	2022	19
ИП Васильев	Кофе	7155346	661411	1001848	285	7	2020	3
ООО Ромашка	Молоко	7850826	186525	49265	82	1	2020	11
ООО Люкс	Сок яблочный	8748388	2119323	1952641	273	2	2022	17
ИП Козлов	Творог 5%	4837527	1074468	505485	47	4	2021	13
ООО Магазин	Сок персиковый	6210813	2028372	2731384	120	10	2021	10
ООО У дома	Творог 5%	8502495	2555124	1657150	7	1	2021	2
ООО Продуктовый	Кофе	1334590	2688282	1400700	78	10	2021	18
ИП Козлов	Кофе	5075864	702479	2659584	209	1	2022	15
ООО Продукты	Масло	2617840	1767532	597787	277	5	2020	14
ИП Иванов	Белый хлеб	2632355	2950791	899563	158	7	2020	2
ИП Петров	Чай	7253600	1791236	2281783	177	6	2021	6

Продолжение таблицы 4

Компания	Продукт	Выручка	Затраты	Налоги	Количество работников	Процент брака	Год	Процент импорта
ИП Петров	Сок персиковый	7777465	1508086	434293	130	9	2022	13
ИП Козлов	Творог 5%	199504	2200479	861509	184	1	2020	11
ИП Семенов	Творог 2%	7882174	1004117	2528452	108	6	2021	3
ООО Продуктовый	Белый хлеб	5098379	1920159	2817059	83	4	2020	3
ООО Люкс	Молоко	618355	1031729	2693034	213	2	2022	17
ООО Магазин	Творог 5%	2170491	1369566	1335267	267	7	2020	1
ИП Козлов	Творог 2%	9875860	1119186	377755	167	3	2020	11
ИП Семенов	Плавленный сыр	8696433	1299580	2654915	221	5	2021	17
ООО Ландыш	Сок яблочный	2491202	1859854	629453	227	2	2021	2
ООО Ландыш	Чай	8671895	1054979	412729	178	8	2020	11
ООО Ромашка	Плавленный сыр	5852232	1930698	1392007	14	1	2020	20
ООО Магазин	Творог 5%	8504236	1926582	1388520	225	9	2022	12
ООО Магазин	Творог 2%	2787649	1541318	2152786	260	2	2021	2
ООО У дома	Масло	2118277	202937	2711801	290	3	2022	13
ИП Петров	Творог 5%	981867	2455854	283393	139	2	2020	8
ООО Продукты	Плавленный сыр	8969516	2931845	519163	68	4	2022	9
ООО Магазин	Плавленный сыр	8071215	2870742	1429329	235	7	2021	20
ООО Ромашка	Творог 2%	6534746	459407	847071	211	5	2021	5
ИП Васильев	Чай	5322811	2527521	193815	211	7	2020	16

При решении задачи оптимизации ассортиментной политики предприятий сферы торговли необходимо опираться на ряд показателей. Одним из таких показателей является маржинальный доход, который представляет собой разницу между выручкой от реализации и расходами в процессе производства или закупки продукции.

На рисунке 10 приведена процедура расчета маржинального дохода от реализации товара в компаниях сфере торговли (фрагмент). Метод статистической обработки: группировка по предприятиям, продукции и годам.

На рисунке 11 приведен отчет по суммарной выручке компаний с наибольшими оборотами.

На рисунке 12 приведена доля компаний с наибольшим оборотом в форме круговой диаграммы.

```
[ ] """ Добавление столбца """
df['Маржинальный доход'] = df['Выручка'] - df['Затраты'] - df['Налоги']
df.head(15)
```

	Компания	Продукт	Выручка	Затраты	Налоги	Количество работников	Процент брака	Год	Процент экспорта	Маржинальный доход
0	ООО Продукты	Молоко	9301184	1225232	1760382	147	10	2021	12	6315570
1	ООО Продукты	Масло	5336586	1912387	2905812	188	3	2020	6	518387
2	ИП Иванов	Чай	2578295	1790496	2924583	61	5	2022	3	-2136784
3	ООО Магазин	Творог 5%	9239042	2876587	600775	212	2	2022	6	5761680
4	ИП Васильев	Макаронные изделия	1087870	462651	659082	208	3	2022	5	-33863
5	ИП Семенов	Чай	9367546	238297	2317842	223	10	2022	9	6811407
6	ООО Ромашка	Макаронные изделия	5676465	1697983	1695799	234	4	2021	3	2282683
7	ООО У дома	Белый хлеб	3987197	2519305	2556902	232	2	2020	1	-1089010
8	ООО Продукты	Плавленый сыр	2589544	1689138	1598052	210	9	2021	10	-697646
9	ИП Иванов	Белый хлеб	3009772	2947162	1863158	3	7	2021	20	-1800548
10	ИП Петров	Крупы	3750347	1184581	496880	134	5	2022	10	2068886
11	ИП Козлов	Кофе	9799319	860076	555102	80	7	2020	6	8384141
12	ООО Люкс	Белый хлеб	8150273	743536	999158	200	1	2020	20	6407579
13	ООО Ярче	Творог 2%	2418934	1092164	1186410	295	5	2021	4	140360
14	ООО Ландыш	Макаронные изделия	5666680	2428407	963746	31	8	2022	10	2274527

Рисунок 10 - Процедура расчета маржинального дохода от реализации товара в компаниях сфере торговли

```
df_2 = df.groupby('Компания').aggregate({'Выручка' : 'sum'})
df_2
```

Компания	Выручка
ООО Люкс	3862736296
ООО Магазин	4317543358
ООО Продуктовый	4039662915
ООО Ромашка	3895790924
ООО У дома	3908141744
ООО Ярче	3889118800
ИП Васильев	3850791148
ИП Иванов	3730938254
ИП Козлов	3871563341
ИП Петров	3763777644
ИП Семенов	3812975384
ООО Ландыш	3824733399

Рисунок 11 - Отчет по суммарной выручке компаний с наибольшими оборотами

На рисунке 13 приведена диаграмма по объемам выручки от реализации товаров различной номенклатуры.

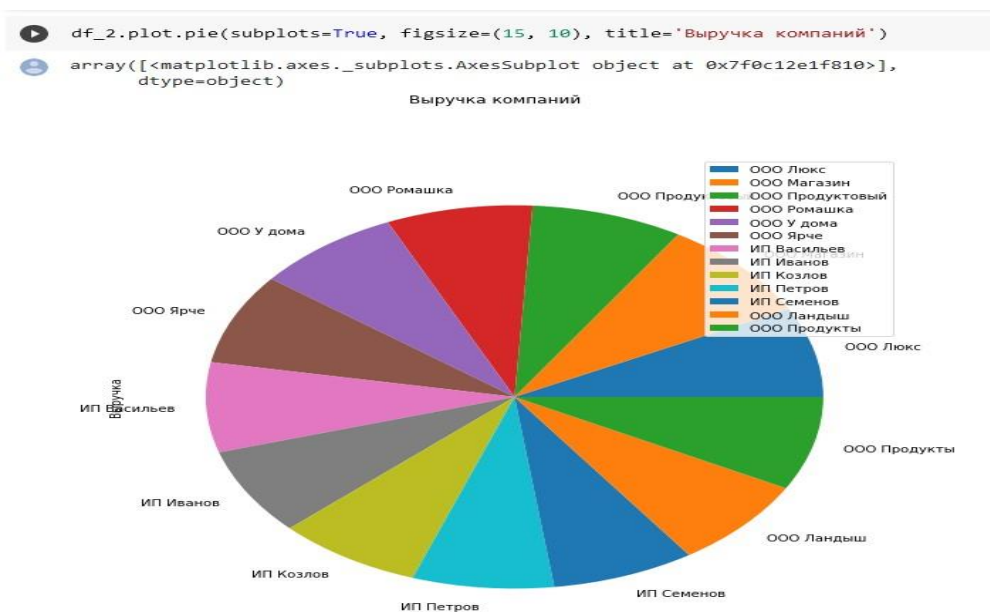


Рисунок 12 – Диаграмма оборотов компаний с наибольшей величиной выручки

```
[ ] fig = plt.figure(figsize=(6, 4))
ax = fig.add_subplot()

x = df_5['index']
y = df_5['Продукт']
ax.barh(x, y)
ax.grid()
```

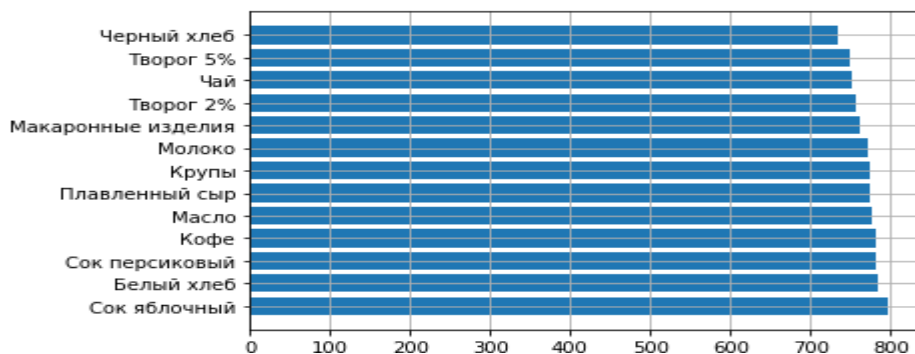


Рисунок 13 – Диаграмма по объемам выручки от реализации товаров различной номенклатуры

На рисунке 14 приведён расчет объемов реализации импортной продукции. Использован метод статистической группировки. Показано, что в рассматриваемый период предприятиями осуществлялся значительный объем продажи импортных товаров. Здесь стоит отметить, что продажа импортных товаров в современной политической ситуации является деятельностью рискованной. Тем не менее, реализация импортной продукции продолжает являться серьезной статьей дохода предприятий сферы торговли.

```
df_8 = df.groupby('Компания').aggregate({'Экспорт (в рублях)': 'sum'})
df_8 = df_8.reset_index()
df_8
```

	Компания	Экспорт (в рублях)
0	ООО Люкс	4.052500e+08
1	ООО Магазин	4.566508e+08
2	ООО Продуктовый	4.176019e+08
3	ООО Ромашка	4.001810e+08
4	ООО У дома	4.020618e+08
5	ООО Ярче	3.970336e+08
6	ИП Васильев	4.194660e+08
7	ИП Иванов	3.892349e+08
8	ИП Козлов	4.170755e+08
9	ИП Петров	3.999840e+08
10	ИП Семенов	3.974181e+08
11	ООО Ландыш	3.890446e+08
12	ООО Продукты	4.327868e+08

Рисунок 14 – Расчет объемов продаж импортных товаров

На рисунке 15 приведён график сводных показателей продаж импортных товаров компаниями, имеющими наибольший объем.


```
[ ] plt.xticks(rotation=45, ha='right')
plt.title("Экспорт(в рублях) по компаниям")
plt.plot(df_8['Компания'], df_8['Экспорт (в рублях)'])
```

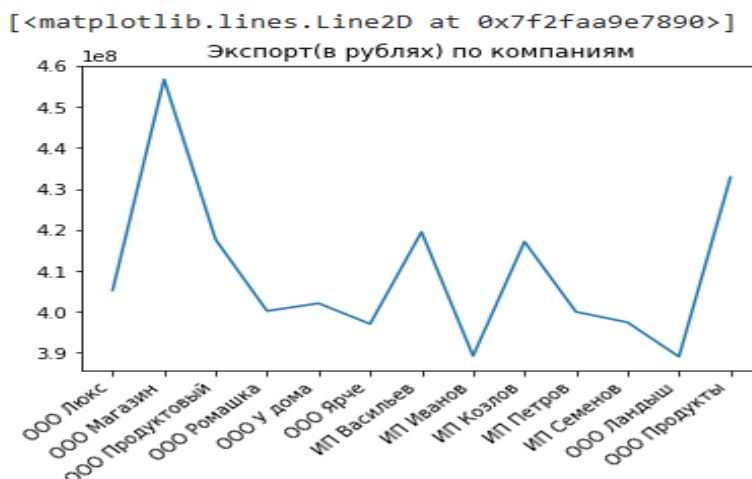


Рисунок 15 – График сводных показателей продаж импортных товаров компаниями, имеющими наибольший объем

В задачах управления предприятиями сферы торговли важное значение приобретает получение данных, которые могут позволить рассчитать себестоимость единицы выпускаемой продукции, рентабельность продаж. Такие данные содержит отчет по затратам при производстве продукции.

На рисунке 16 приведён годовой отчет по затратам при производстве пищевой продукции.

```
[ ] df_9 = df.groupby('Год').aggregate({'Затраты' : 'sum'})
df_9 = df_9.reset_index()
df_9
```

	Год	Затраты
0	2020	4886628841
1	2021	5069703857
2	2022	4989577060

Рисунок 16 – Отчет по издержкам торговых компаний в разрезе отчетных периодов

Как показано на рисунке 16, в рассматриваемый период максимальный объем издержек торговых компаний пришелся на 2021 год, далее отмечено снижение, связанное с сокращением валового объема реализации товара на текущий год.

На рисунке 17 приведена диаграмма затрат на производство продукции по расчетным периодам. Используются статистические методы группировки и вычисления относительных величин.

На рисунке 18 приведен отчет по величине выручки от продаж. Как показано на рисунке - наибольший удельный вес в выручке за проданные товары занимает позиция «периковый сок».

На рисунке 19 приведено графическое представление по выручке от сбыта товаров.

На рисунке 20 приведён отчет по издержкам торговых предприятий. На рисунке 21 – его графическое представление. Как показано на рисунке 20 - наибольший уровень затрат приходится на продажи продукции по позиции «Белый хлеб». Используются статистические методы группировки, расчета относительных показателей.

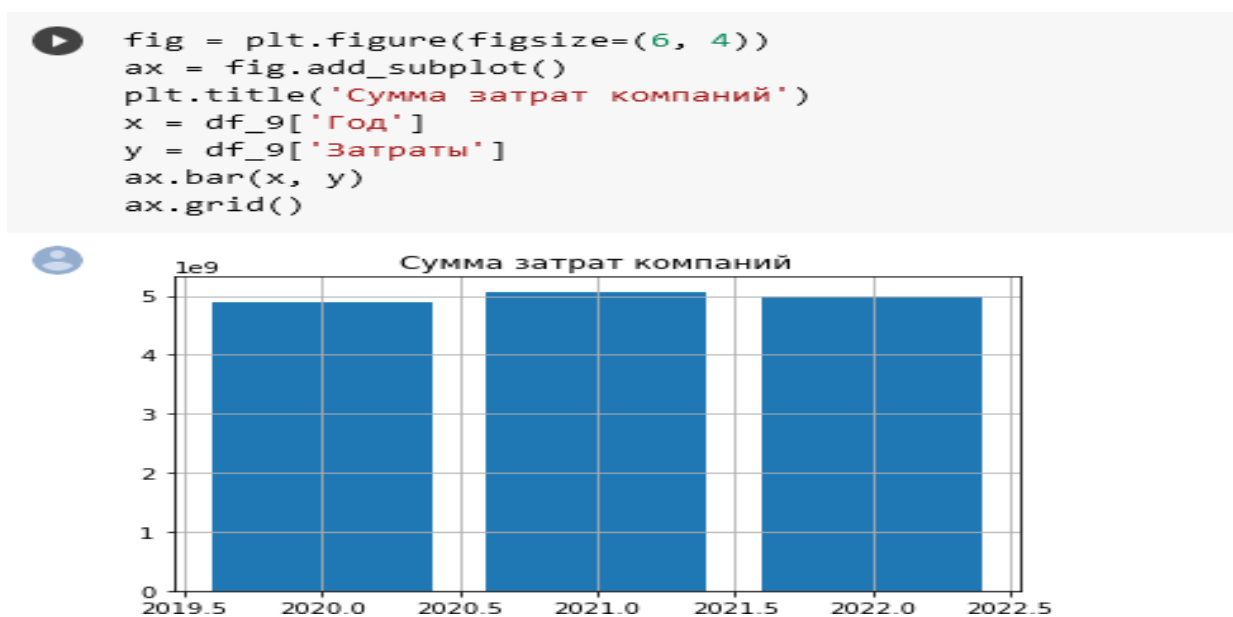


Рисунок 17 – Отчет по затратам на производство пищевой продукции в разрезе отчетных периодов (графическое представление)

```
[ ] df_11 = df.groupby('Продукт').aggregate({'Выручка' : 'sum'})
df_11 = df_11.reset_index()
df_11
```

	Продукт	Выручка
0	Белый хлеб	3992706095
1	Кофе	3999264483
2	Крупы	3754562562
3	Макаронные изделия	3847264139
4	Масло	4014606655
5	Молоко	3953285474
6	Плавленный сыр	3955469433
7	Сок персиковый	4040468510
8	Сок яблочный	3982530326
9	Творог 2%	3836278038
10	Творог 5%	3789580481
11	Чай	3919932268
12	Черный хлеб	3694921339

Рисунок 18 - Отчет по величине выручки от сбыта товаров

```
fig = plt.figure(figsize=(6, 4))
ax = fig.add_subplot()
plt.title('Выручка по продуктам')
x = df_11['Продукт']
y = df_11['Выручка']
ax.barh(x, y)
ax.grid()
```



Рисунок 19 - Отчет по величине выручки от сбыта товаров
(графическое представление)

```
df_12 = df.groupby('Продукт').aggregate({'Затраты' : 'sum'})
df_12 = df_12.reset_index()
df_12
```

	Продукт	Затраты
0	Белый хлеб	1206373307
1	Кофе	1189194424
2	Крупы	1140393664
3	Макаронные изделия	1174202171
4	Масло	1164662748
5	Молоко	1128935112
6	Плавленный сыр	1107797380
7	Сок персиковый	1180277386
8	Сок яблочный	1199458785
9	Творог 2%	1105405067
10	Творог 5%	1127569258
11	Чай	1108376543
12	Черный хлеб	1113263913

Рисунок 20 – Отчет по затратам на производство продукции

```
[ ] fig = plt.figure(figsize=(6, 4))
ax = fig.add_subplot()
plt.title('Затраты по продуктам')
x = df_12['Продукт']
y = df_12['Затраты']
ax.barh(x, y)
ax.grid()
```



Рисунок 21 – Отчет по затратам на производство продукции

(графическое представление)

В задачах управления очень важно принимать грамотные и взвешенные решения. Для этого осуществим «корреляционный анализ, задача которого сводится к измерению тесноты связи между факторами, выявлению неизвестных причин связей и оценке факторов, вызывающих максимальное влияние на результат»[9].

Корреляционный анализ по показателям «маржинальный доход» и «количество работников» и результат расчёта показан на рисунке 22.

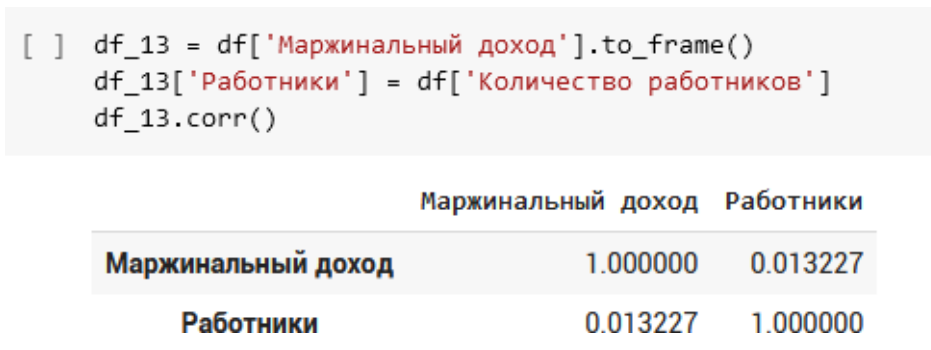


Рисунок 22 – Корреляционный анализ по показателям «маржинальный доход» и «количество работников»

Как показано на рисунке 22, корреляционная зависимость между указанными параметрами отсутствует.

Далее проведён корреляционный анализ по показателям «выручка» и «процент брака». Результат расчёта показан на рисунке 23.

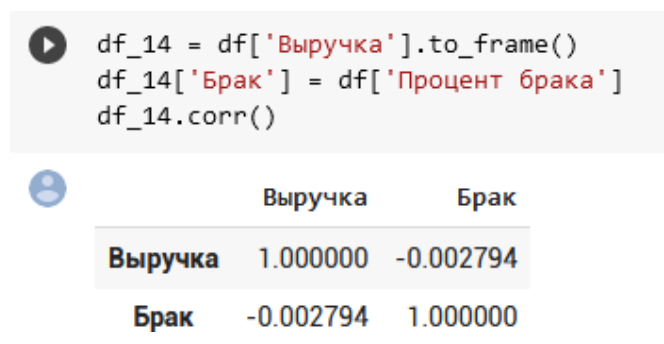


Рисунок 23 – Корреляционный анализ по показателям «выручка» и «процент брака»

Далее проведем исследование файла данных с показателями финансовой деятельности пищевых предприятий. В таблице 5 приведен фрагмент файла данных.

Таблица 5 - Фрагмент файла данных по финансовому анализу деятельности пищевых предприятий

Компания	Дебиторская задолженность	Кредиторская задолженность	Субсидии	Год
ИП Васильев	1032375	495160	598912	2021
ООО У дома	883508	4360708	909087	2020
ООО Продукты	2232248	1086130	735505	2021
ИП Иванов	3066631	2311238	146467	2020
ООО У дома	284078	3380706	812077	2022
ООО Ярче	4926962	1019984	887426	2022
ООО Магазин	201037	673142	978487	2022
ООО Магазин	4917631	4295111	158214	2021
ООО Ярче	139584	250108	752736	2020
ООО У дома	4495241	1267255	403928	2020
ООО У дома	4353107	4657403	809519	2020
ООО Ландыш	367094	4060111	445838	2022
ИП Козлов	3740801	4628921	268986	2022
ИП Козлов	2880547	4945436	361792	2022
ООО Магазин	160540	3640677	428815	2021
ИП Иванов	1027677	870123	915119	2021
ООО У дома	744405	1205127	861774	2021
ООО У дома	4095705	2074415	578771	2021

Продолжение таблицы 5

Компания	Дебиторская задолженность	Кредиторская задолженность	Субсидии	Год
ИП Петров	499237	277182	833758	2022
ИП Семенов	4247332	4943200	616908	2020
ООО Ромашка	2976446	4218428	121207	2022
ИП Иванов	3386768	474075	253632	2020
ИП Семенов	1794570	4227177	377678	2020
ИП Семенов	2675162	1203732	267783	2021
ООО Люкс	1527842	2576643	787044	2021
ИП Петров	3625567	2556171	217306	2020
ИП Семенов	1292169	4143877	885541	2022
ООО У дома	4466255	2687638	606375	2020
ООО Продуктовый	4343670	3342803	190717	2022
ИП Васильев	2255637	4409390	881615	2021
ООО Продуктовый	611654	4646169	289212	2021
ООО Ярче	2107071	3122300	593175	2022
ООО Ярче	3863029	694793	812865	2022
ООО Ландыш	4314534	1776565	137207	2022
ООО Ярче	1547195	2095694	611139	2022
ИП Васильев	3514434	4209776	472703	2021
ООО Продукты	1186404	3107293	318173	2020

Продолжение таблицы 5

Компания	Дебиторская задолженность	Кредиторская задолженность	Субсидии	Год
ООО Продуктовый	2442653	1553904	218175	2020
ООО Магазин	1531860	1256605	581917	2021
ИП Семенов	1450825	3410522	684347	2020
ООО Ярче	2695954	418830	493239	2022
ООО Продуктовый	618893	4071844	295980	2021
ООО Продукты	1005644	4543239	360850	2020
ООО Продуктовый	2972007	2474447	511538	2022
ИП Семенов	1816909	836439	937666	2020
ООО У дома	103458	3411017	776505	2020
ООО Ландыш	4652354	1314479	669772	2020
ООО Продуктовый	171192	3752221	599836	2021
ООО Ромашка	4917530	2649349	605796	2020
ИП Васильев	3246053	1621245	275774	2022
ИП Иванов	1824723	797916	822188	2020
ООО У дома	4791952	129808	553238	2021
ООО У дома	706520	3714655	414500	2020
ИП Козлов	807804	3847867	651818	2020
ООО Магазин	4886726	4858865	342919	2022
ООО Люкс	1062029	2926765	127546	2021
ООО У дома	829663	1027353	661564	2022
ООО Магазин	3420885	1859438	478386	2021
ООО Ландыш	1553094	2675106	402047	2022

Продолжение таблицы 5

Компания	Дебиторская задолженность	Кредиторская задолженность	Субсидии	Год
ООО Люкс	4969067	2892123	754312	2022
ООО Ярче	1933201	581721	405167	2022
ИП Семенов	2472542	2597700	278775	2020
ИП Козлов	4207374	501356	879666	2020
ООО Ландыш	1935564	2236649	991617	2021

Файл данных содержит информацию по дебиторской и кредиторской задолженности, отрасли продаж, объему бонусов, данные о расчетном периоде в форме компенсации налогов.

Причина, по которой автор считает целесообразным рассмотреть данный массив данных, заключается в том, что эффективное управление торговым предприятием – это также вопрос обеспечения приемлемого уровня риска. «Увеличение размера кредиторской задолженности как бесплатного источника финансирования и ее доли в заемном капитале предприятия, как правило, способствует росту прибыли от продаж, но и увеличивает риски неуплаты организации по своим обязательствам»[24].

На рисунке 24 приведён отчет о финансовой деятельности предприятий.

Как показано на рисунке 24, финансовая деятельность компаний осуществляется как с положительным, так и с отрицательным финансовым результатом. На рисунке 25 приведён отчет по суммарной дебиторской задолженности по предприятиям сфере торговли.

```
[ ] df2['Сальдо'] = df2['Дебиторская задолженность'] + df2['Субсидии'] - df2['Кредиторская задолженность']
df2
```

	Компания	Дебиторская задолженность	Кредиторская задолженность	Субсидии	Год	Отрасль	Сальдо
0	ИП Васильев	1032375	495160	598912	2021	Производство круп	1136127
1	ООО У дома	883508	4360708	909087	2020	Производство круп	-2568113
2	ООО Продукты	2232248	1086130	735505	2021	Производство консервов	1881623
3	ИП Иванов	3066631	2311238	146467	2020	Молочное производство	901860
4	ООО У дома	284078	3380706	812077	2022	Производство напитков	-2284551
...
9995	ООО Ромашка	3985140	352252	580546	2021	Производство круп	4213434
9996	ИП Козлов	2880031	473519	308859	2021	Молочное производство	2715371
9997	ИП Васильев	1187420	3016803	304840	2022	Производство кондитерских изделий	-1524543
9998	ИП Петров	495681	3329939	760874	2020	Производство круп	-2073384
9999	ООО Ромашка	1948411	4931220	100653	2021	Производство напитков	-2882156

10000 rows x 7 columns

Рисунок 24 - Отчет о финансовой деятельности предприятий (фрагмент)

```
df2_1 = df2.groupby('Год').aggregate({'Дебиторская задолженность' : 'sum'})
df2_1 = df2_1.reset_index()
df2_1
```

	Год	Дебиторская задолженность
0	2020	8524943762
1	2021	8358373512
2	2022	8573884815

Рисунок 25 - Отчет по суммарной дебиторской задолженности по предприятиям сфере торговли

Как показано на рисунке 25, в настоящее время отмечается рост уровня дебиторской задолженности предприятий.

На рисунке 26 приведена графическая интерпретация данного отчета. Используются методы статистической группировки и расчета относительных показателей.

```
[ ] fig = plt.figure(figsize=(6, 4))
ax = fig.add_subplot()
plt.title('Дебиторская задолженность всех компаний')
x = df2_1['Год']
y = df2_1['Дебиторская задолженность']
ax.bar(x, y)
ax.grid()
```



Рисунок 26 - Отчет по суммарной дебиторской задолженности по предприятиям сфере торговли (графическая интерпретация)

На рисунке 27 приведён отчет по суммарной кредиторской задолженности по предприятиям сфере торговли.

```
[ ] df2_2 = df2.groupby('Год').aggregate({'Кредиторская задолженность' : 'sum'})
df2_2 = df2_2.reset_index()
df2_2
```

	Год	Кредиторская задолженность
0	2020	8505286622
1	2021	8401851752
2	2022	8559045545

Рисунок 27 - Отчет по суммарной кредиторской задолженности по предприятиям сфере торговли

Как показано на рисунке 27, в настоящее время отмечается рост уровня дебиторской задолженности предприятий. На рисунке 28 приведена графическая интерпретация данного отчета. Используются методы статистической группировки и расчета относительных показателей.

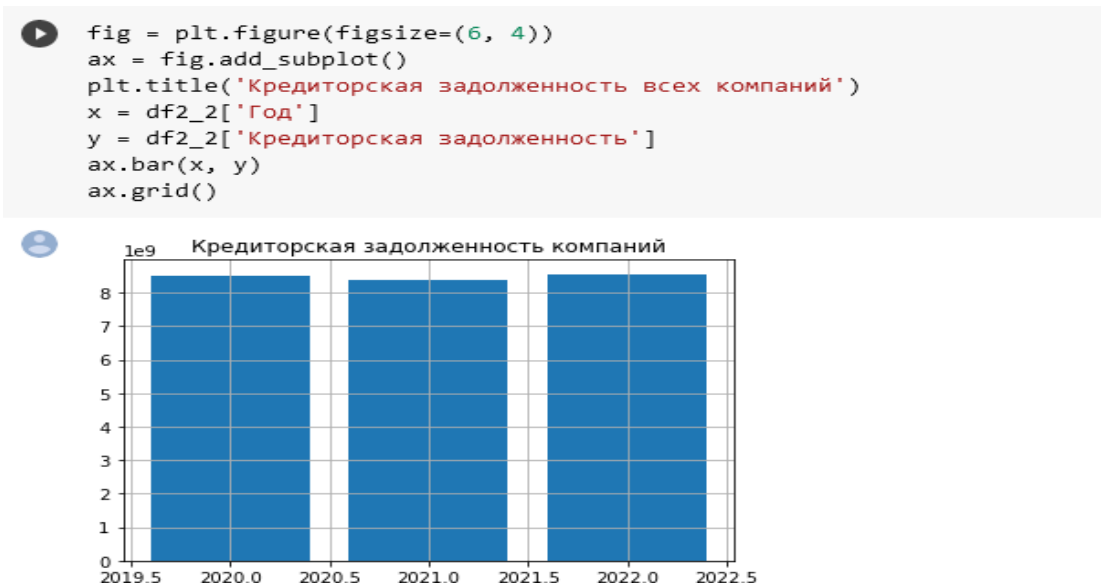


Рисунок 28 - Отчет по суммарной кредиторской задолженности по предприятиям сфере торговли (графическая интерпретация)

Далее на рисунке 29 проведен отчет по суммарной дебиторской задолженности по компаниям (использован метод группировки).

```

df2_3 = df2.groupby('Компания').aggregate({'Дебиторская задолженность': 'sum'})
df2_3

```

Компания	Дебиторская задолженность
ООО Люкс	1986533706
ООО Магазин	1927563725
ООО Продуктовый	1970254245
ООО Ромашка	2006289998
ООО У дома	1925487494
ООО Ярче	1987470905
ИП Васильев	1931900941
ИП Иванов	1973444731
ИП Козлов	1976104551
ИП Петров	1881505034
ИП Семенов	1991044306
ООО Ландыш	1975088643
ООО Продукты	1924513810

Рисунок 29 - Отчет по суммарной кредиторской задолженности по предприятиям сфере торговли

На рисунке 30 приведено графическое представление указанного отчета.

На рисунке 31 приведён сводный отчет по дебиторской задолженности по отраслям. Использован метод статистической группировки.

Как показано на рисунке 32, наибольший уровень дебиторской задолженности наблюдается в отрасли продаж кондитерских изделий, наименьший уровень - в отрасли продаж круп.

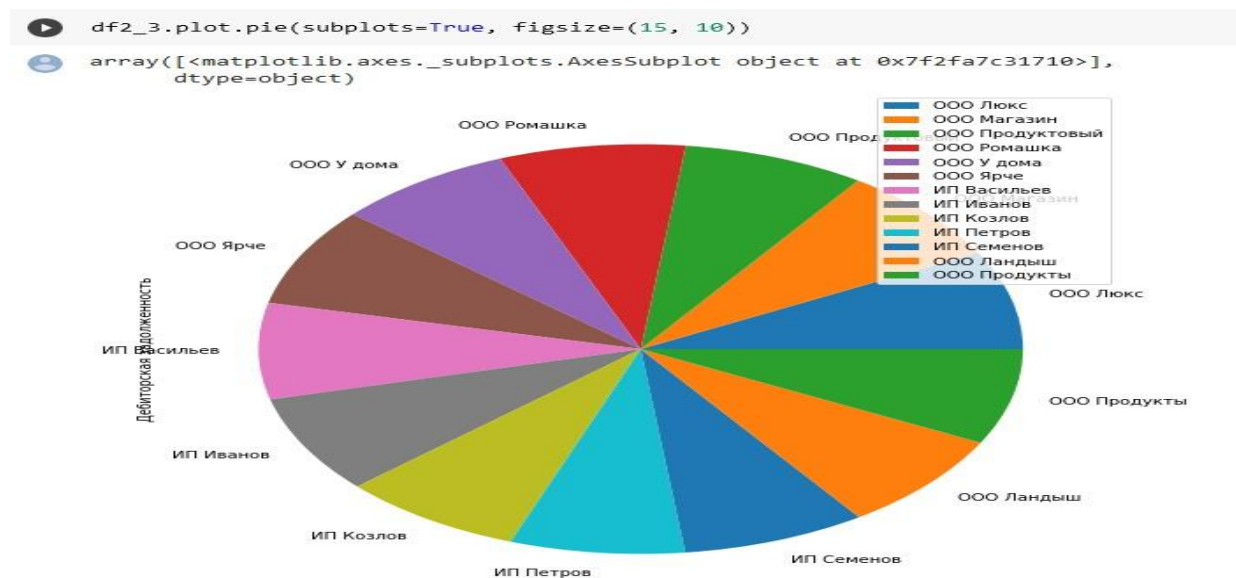


Рисунок 30 - Отчет по суммарной кредиторской задолженности по предприятиям сфере торговли (графическое представление)

```
[ ] df2_4 = df2.groupby('Отрасль').aggregate({'Дебиторская задолженность' : 'sum'})
df2_4 = df2_4.reset_index()
df2_4
```

	Отрасль	Дебиторская задолженность
0	Молочное производство	5125589472
1	Производство кондитерских изделий	5303392359
2	Производство консервов	4969412908
3	Производство круп	4966450942
4	Производство напитков	5092356408

Рисунок 31 - Сводный отчет по дебиторской задолженности по отраслям

```
fig = plt.figure(figsize=(6, 4))
ax = fig.add_subplot()
x = df2_4['Отрасль']
y = df2_4['Дебиторская задолженность']
ax.barh(x, y)
ax.grid()
```

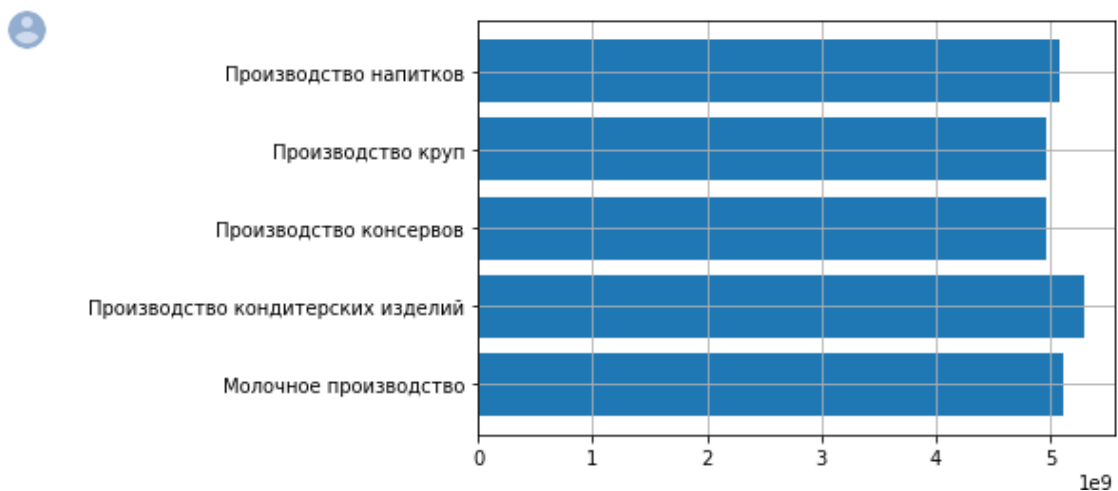


Рисунок 32 - Сводный отчет по дебиторской задолженности по отраслям (графическое представление)

На рисунке 33 приведён сводный отчет по кредиторской задолженности по отраслям. Использован метод статистической группировки. Как показано на рисунке 34, наибольший уровень

кредиторской задолженности наблюдается в отрасли продаж кондитерских изделий, наименьший уровень - в отрасли продаж круп.

```
[ ] df2_5 = df2.groupby('Отрасль').aggregate({'Кредиторская задолженность' : 'sum'})
df2_5 = df2_5.reset_index()
df2_5
```

	Отрасль	Кредиторская задолженность
0	Молочное производство	5103487973
1	Производство кондитерских изделий	5193633386
2	Производство консервов	5046672224
3	Производство круп	4972127418
4	Производство напитков	5150262918

Рисунок 33 - Сводный отчет по кредиторской задолженности по отраслям

```
[ ] fig = plt.figure(figsize=(6, 4))
ax = fig.add_subplot()
x = df2_5['Отрасль']
y = df2_5['Кредиторская задолженность']
ax.barh(x, y)
ax.grid()
```

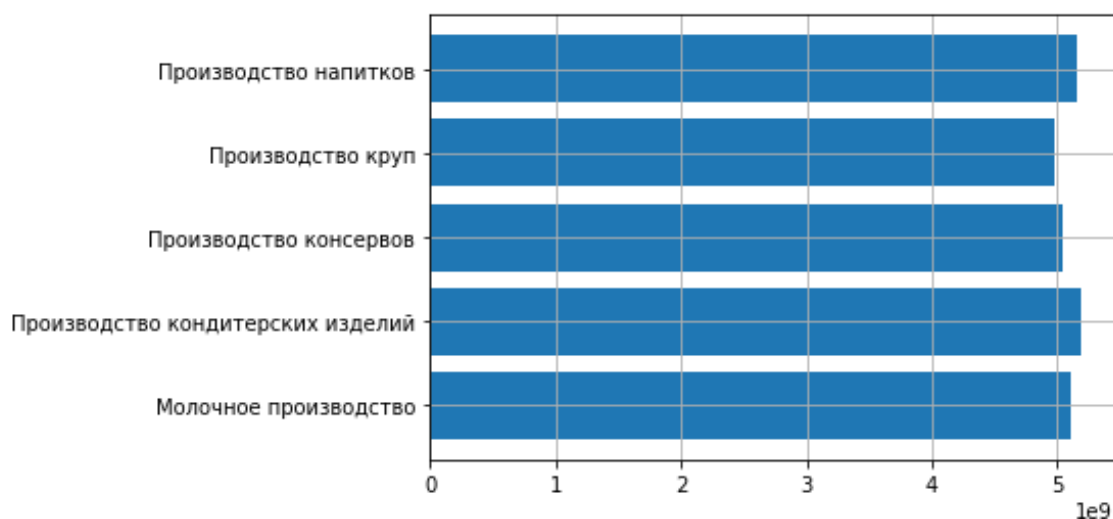


Рисунок 34 - Сводный отчет по дебиторской задолженности по отраслям (графическое представление)

На рисунке 35 приведён сводный отчет по финансовым результатам деятельности предприятий сфере торговли. Используются методы статистической группировки.

```
[ ] df2_7 = df2.groupby('Отрасль').aggregate({'Сальдо' : 'sum'})
df2_7 = df2_7.reset_index()
df2_7
```

	Отрасль	Сальдо
0	Молочное производство	1104488065
1	Производство кондитерских изделий	1251802051
2	Производство консервов	1002220468
3	Производство круп	1060101578
4	Производство напитков	1058007986

Рисунок 35 - Сводный отчет по финансовым результатам деятельности предприятий сфере торговли

На рисунке 36 приведено графическое представление данного отчета.

```
[ ] fig = plt.figure(figsize=(6, 4))
ax = fig.add_subplot()
x = df2_7['Отрасль']
y = df2_7['Сальдо']
ax.barh(x, y)
ax.grid()
```

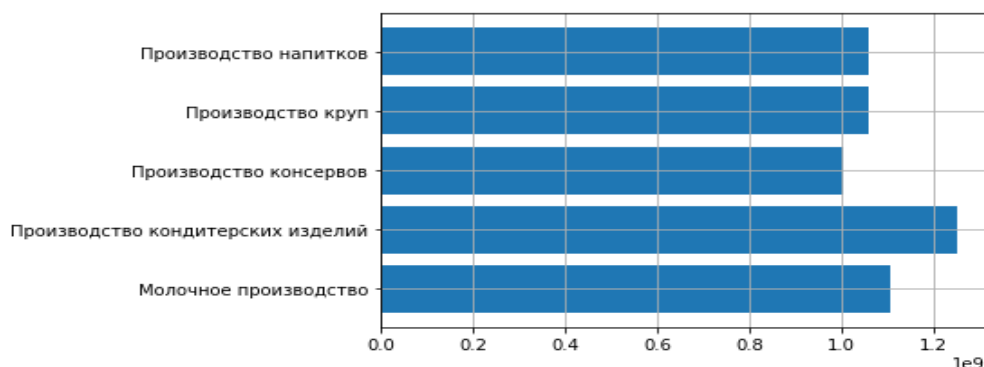


Рисунок 36 - Сводный отчет по финансовым результатам деятельности предприятий сфере торговли (график)

Как показано на рисунке 35, наибольшую эффективность показала отрасль производства кондитерских изделий.

Важно отметить, что Российская Федерация оказывает финансовую помощь предприятиям сферы торговли в виде субсидий, поскольку Правительство РФ понимает стратегическое значение данной отрасли в структуре экономики нашей страны.

В связи с этим на рисунке 37 приведен расчёт корреляции между финансовым результатом и объемом полученных субсидий.

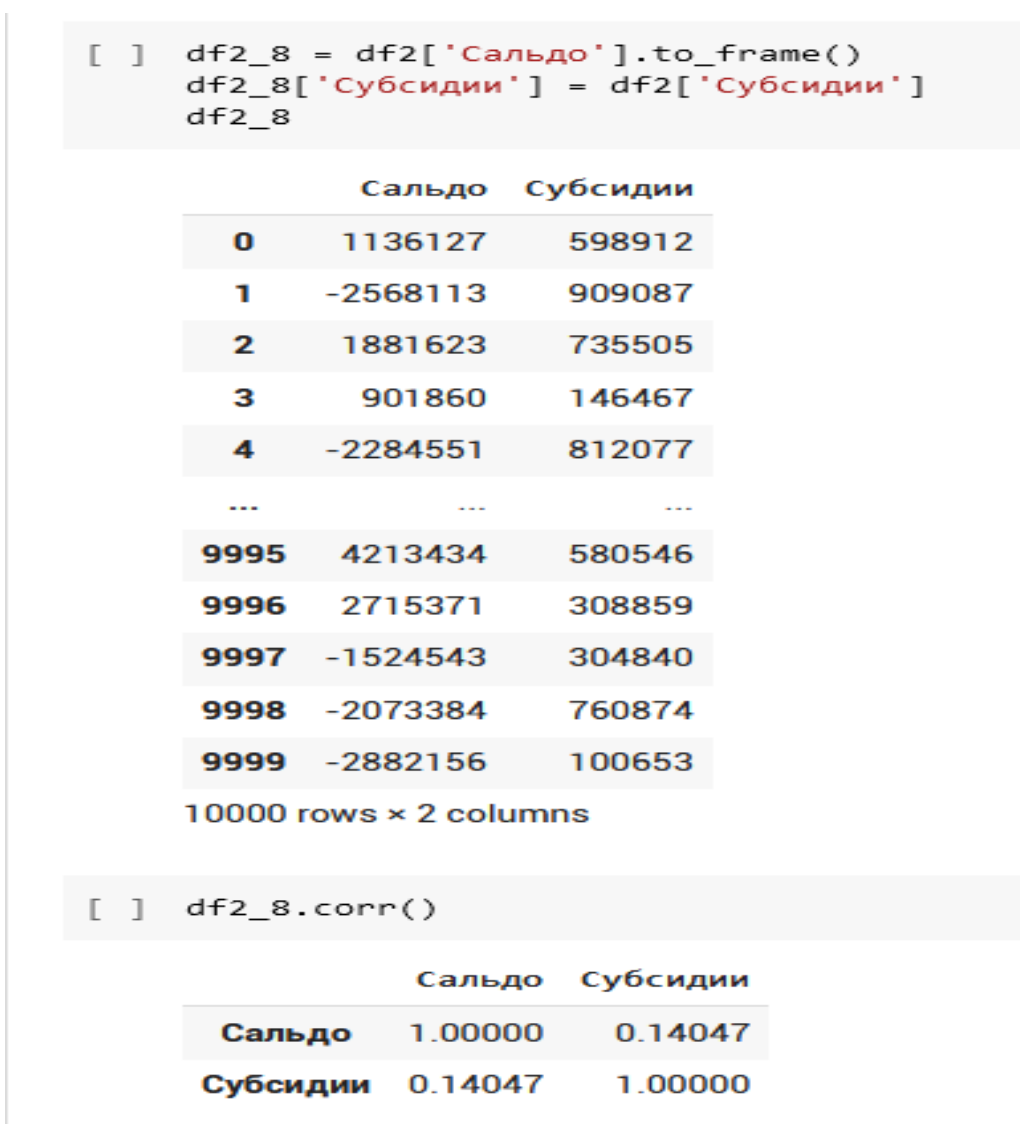


Рисунок 37 - Расчёт корреляции между финансовым результатом и объемом полученных субсидий

Как показано на рисунке 37, между указанными показателями корреляция отсутствует.

На рисунке 38 приведен расчёт корреляции между финансовым результатом и объемом кредиторской задолженности.

```
df2_9 = df2['Субсидии'].to_frame()  
df2_9['Кредиторская задолженность'] = df2['Кредиторская задолженность']  
df2_9
```

	Субсидии	Кредиторская задолженность
0	598912	495160
1	909087	4360708
2	735505	1086130
3	146467	2311238
4	812077	3380706
...
9995	580546	352252
9996	308859	473519
9997	304840	3016803
9998	760874	3329939
9999	100653	4931220

10000 rows x 2 columns

```
[ ] df2_9.corr()
```

	Субсидии	Кредиторская задолженность
Субсидии	1.000000	0.003986
Кредиторская задолженность	0.003986	1.000000

Рисунок 38 - Расчёт корреляции между финансовым результатом и объемом кредиторской задолженности

Как показано на рисунке 38, между указанными показателями корреляция отсутствует. Таким образом, поставленные задачи статистической обработки информации реализованы.

4.2 Апробация и анализ результатов по статистике продаж

Далее проведём анализ статистики продаж товаров в разрезе видов товарных позиций, форматов торговых точек. В таблице 6 приведен фрагмент исходных данных анализируемого массива.

Таблица 6 – Исходные данные

Вид продукции	Формат магазина	Дата	Выручка	Формат оплаты
Молочная	Маркетплейс	19.03.2023	1692483	наличные
Бытовая химия	Супермаркет	27.03.2023	2821627	наличные
Бытовая химия	Гипермаркет	29.03.2023	2736173	наличные
Электроника	Гипермаркет	10.03.2023	6819064	наличные
Кондитерская продукция	Гипермаркет	01.03.2023	7902809	банковская карта
Бытовая химия	Маркетплейс	28.03.2023	2752005	наличные
Хлебная	Магазин у дома	09.03.2023	549832	банковская карта
Кондитерская продукция	Магазин у дома	30.03.2023	3069652	банковская карта
Сантехника	Магазин у дома	10.03.2023	3532898	наличные
Хлебная	Маркетплейс	26.03.2023	6964775	банковская карта
Хлебная	Маркетплейс	01.03.2023	9630604	наличные
Молочная	Маркетплейс	07.03.2023	8485364	банковская карта
Молочная	Маркетплейс	28.03.2023	2837568	банковская карта
Молочная	Магазин у дома	10.03.2023	8504384	наличные
Кондитерская продукция	Маркетплейс	22.03.2023	7297989	банковская карта
Сантехника	Маркетплейс	01.03.2023	3545797	банковская карта
Кондитерская продукция	Гипермаркет	30.03.2023	4440008	наличные
Электроника	Магазин у дома	04.03.2023	2928329	наличные
Хлебная	Маркетплейс	06.03.2023	4339618	банковская карта
Бытовая химия	Магазин у дома	12.03.2023	8223836	наличные
Электроника	Гипермаркет	01.03.2023	2757230	наличные
Бытовая химия	Магазин у дома	24.03.2023	7985304	наличные
Молочная	Супермаркет	14.03.2023	6099754	банковская карта

Продолжение таблицы 6

Вид продукции	Формат магазина	Дата	Выручка	Формат оплаты
Сантехника	Маркетплейс	20.03.2023	4216666	банковская карта
Сантехника	Магазин у дома	11.03.2023	5943479	наличные
Хлебная	Супермаркет	17.03.2023	9798672	банковская карта
Кондитерская продукция	Гипермаркет	07.03.2023	4780535	банковская карта
Кондитерская продукция	Маркетплейс	12.03.2023	8978095	банковская карта
Сантехника	Гипермаркет	25.03.2023	6368425	наличные
Сантехника	Супермаркет	28.03.2023	7911266	наличные
Бытовая химия	Магазин у дома	10.03.2023	9238246	банковская карта
Сантехника	Супермаркет	21.03.2023	4892675	банковская карта
Бытовая химия	Гипермаркет	27.03.2023	7261840	наличные
Хлебная	Супермаркет	09.03.2023	9855224	наличные
Кондитерская продукция	Супермаркет	02.03.2023	2766221	наличные
Молочная	Супермаркет	07.03.2023	9450074	банковская карта
Молочная	Маркетплейс	20.03.2023	4230971	банковская карта
Электроника	Гипермаркет	13.03.2023	9768092	наличные
Хлебная	Маркетплейс	24.03.2023	7517197	банковская карта
Кондитерская продукция	Гипермаркет	05.03.2023	5839101	банковская карта
Сантехника	Гипермаркет	04.03.2023	2451169	наличные
Хлебная	Магазин у дома	21.03.2023	228562	наличные
Хлебная	Маркетплейс	07.03.2023	6841238	банковская карта
Сантехника	Магазин у дома	11.03.2023	9816469	наличные
Сантехника	Супермаркет	22.03.2023	6633081	банковская карта
Хлебная	Магазин у дома	14.03.2023	1634136	наличные
Сантехника	Гипермаркет	08.03.2023	6215671	наличные
Хлебная	Маркетплейс	19.03.2023	458576	банковская карта
Молочная	Маркетплейс	29.03.2023	8267768	наличные
Кондитерская продукция	Гипермаркет	05.03.2023	5693925	банковская карта
Молочная	Магазин у дома	19.03.2023	6322301	банковская карта
Сантехника	Супермаркет	25.03.2023	6929806	наличные
Электроника	Супермаркет	18.03.2023	5218887	банковская карта
Бытовая химия	Гипермаркет	16.03.2023	4165154	банковская карта
Электроника	Гипермаркет	28.03.2023	2904323	банковская карта
Электроника	Магазин у дома	01.03.2023	4865952	банковская карта
Молочная	Гипермаркет	23.03.2023	3179259	наличные
Электроника	Супермаркет	30.03.2023	210985	наличные
Бытовая химия	Супермаркет	22.03.2023	8271115	банковская карта
Хлебная	Маркетплейс	19.03.2023	6100079	наличные
Бытовая химия	Гипермаркет	22.03.2023	5413281	банковская карта

Продолжение таблицы 6

Вид продукции	Формат магазина	Дата	Выручка	Формат оплаты
Молочная	Супермаркет	17.03.2023	7398614	наличные
Кондитерская продукция	Маркетплейс	18.03.2023	4657806	банковская карта
Кондитерская продукция	Супермаркет	06.03.2023	1384167	наличные
Хлебная	Гипермаркет	17.03.2023	1085822	наличные
Сантехника	Гипермаркет	05.03.2023	389496	наличные
Молочная	Супермаркет	27.03.2023	8876334	наличные
Молочная	Супермаркет	20.03.2023	4032188	банковская карта
Молочная	Гипермаркет	20.03.2023	2218804	наличные
Электроника	Магазин у дома	06.03.2023	3554119	банковская карта
Бытовая химия	Магазин у дома	08.03.2023	1462918	банковская карта
Сантехника	Супермаркет	16.03.2023	317880	банковская карта
Электроника	Гипермаркет	01.03.2023	4573605	наличные
Бытовая химия	Маркетплейс	08.03.2023	4147935	наличные
Хлебная	Супермаркет	03.03.2023	1532045	банковская карта
Молочная	Супермаркет	17.03.2023	1987748	наличные
Кондитерская продукция	Маркетплейс	09.03.2023	9229624	наличные
Молочная	Супермаркет	25.03.2023	2922725	банковская карта
Хлебная	Магазин у дома	26.03.2023	8779205	банковская карта
Бытовая химия	Супермаркет	29.03.2023	9632846	банковская карта
Хлебная	Супермаркет	01.03.2023	1873679	банковская карта
Молочная	Гипермаркет	18.03.2023	7123334	наличные
Сантехника	Магазин у дома	06.03.2023	4966650	наличные
Кондитерская продукция	Гипермаркет	04.03.2023	6899548	наличные
Бытовая химия	Маркетплейс	07.03.2023	7123931	наличные
Бытовая химия	Магазин у дома	15.03.2023	7499649	наличные
Электроника	Гипермаркет	08.03.2023	464462	наличные
Бытовая химия	Магазин у дома	31.03.2023	1649604	банковская карта
Хлебная	Гипермаркет	30.03.2023	6297527	наличные
Электроника	Гипермаркет	16.03.2023	8552822	наличные
Хлебная	Гипермаркет	11.03.2023	9378239	банковская карта
Электроника	Гипермаркет	04.03.2023	155055	банковская карта
Молочная	Супермаркет	14.03.2023	5515046	банковская карта
Сантехника	Маркетплейс	19.03.2023	675321	наличные
Молочная	Супермаркет	23.03.2023	528380	наличные
Кондитерская продукция	Магазин у дома	15.03.2023	5839938	банковская карта
Кондитерская продукция	Гипермаркет	13.03.2023	841418	наличные
Электроника	Гипермаркет	12.03.2023	7032460	наличные

Продолжение таблицы 6

Вид продукции	Формат магазина	Дата	Выручка	Формат оплаты
Бытовая химия	Гипермаркет	17.03.2023	19283	наличные
Сантехника	Супермаркет	13.03.2023	9224263	наличные
Молочная	Маркетплейс	14.03.2023	5996679	наличные
Электроника	Маркетплейс	12.03.2023	9334257	банковская карта
Электроника	Магазин у дома	02.03.2023	3366554	банковская карта
Кондитерская продукция	Гипермаркет	29.03.2023	7366453	банковская карта
Сантехника	Супермаркет	28.03.2023	1419535	банковская карта
Кондитерская продукция	Магазин у дома	27.03.2023	7055461	наличные
Молочная	Гипермаркет	11.03.2023	9583923	наличные
Кондитерская продукция	Супермаркет	31.03.2023	4880652	банковская карта
Молочная	Супермаркет	22.03.2023	4271407	банковская карта
Кондитерская продукция	Магазин у дома	17.03.2023	2825968	наличные
Молочная	Гипермаркет	28.03.2023	9271689	банковская карта
Хлебная	Маркетплейс	17.03.2023	5902764	наличные
Электроника	Гипермаркет	15.03.2023	6081884	наличные
Хлебная	Гипермаркет	24.03.2023	9685171	наличные
Хлебная	Супермаркет	28.03.2023	7418177	наличные
Электроника	Маркетплейс	06.03.2023	9683553	наличные
Сантехника	Маркетплейс	30.03.2023	9833058	наличные
Электроника	Гипермаркет	16.03.2023	6920542	банковская карта
Бытовая химия	Гипермаркет	23.03.2023	7424890	банковская карта
Бытовая химия	Маркетплейс	31.03.2023	5906085	наличные
Сантехника	Магазин у дома	25.03.2023	1224885	наличные
Бытовая химия	Супермаркет	20.03.2023	8730071	банковская карта
Электроника	Магазин у дома	05.03.2023	8602008	банковская карта
Молочная	Супермаркет	19.03.2023	2190508	наличные
Хлебная	Супермаркет	23.03.2023	9788856	банковская карта
Молочная	Магазин у дома	29.03.2023	4022931	наличные
Кондитерская продукция	Маркетплейс	26.03.2023	3695349	банковская карта
Сантехника	Магазин у дома	08.03.2023	9469140	наличные
Электроника	Гипермаркет	04.03.2023	1998685	наличные
Молочная	Маркетплейс	05.03.2023	4416359	банковская карта
Кондитерская продукция	Супермаркет	24.03.2023	1255933	наличные
Молочная	Гипермаркет	27.03.2023	6366575	наличные
Кондитерская продукция	Магазин у дома	03.03.2023	4899914	наличные
Бытовая химия	Супермаркет	21.03.2023	9414340	наличные

Продолжение таблицы 6

Вид продукции	Формат магазина	Дата	Выручка	Формат оплаты
Кондитерская продукция	Супермаркет	18.03.2023	6200770	банковская карта
Электроника	Гипермаркет	04.03.2023	8676391	наличные
Электроника	Маркетплейс	13.03.2023	8884095	наличные
Бытовая химия	Супермаркет	28.03.2023	643192	банковская карта
Хлебная	Супермаркет	29.03.2023	2907660	банковская карта
Электроника	Магазин у дома	01.03.2023	3236226	банковская карта
Электроника	Магазин у дома	09.03.2023	7372560	банковская карта
Молочная	Гипермаркет	04.03.2023	3107265	наличные
Сантехника	Маркетплейс	04.03.2023	1195919	банковская карта
Хлебная	Гипермаркет	29.03.2023	5549400	наличные
Хлебная	Магазин у дома	10.03.2023	7896291	наличные
Сантехника	Гипермаркет	26.03.2023	6931618	наличные
Кондитерская продукция	Супермаркет	01.03.2023	5462221	наличные
Бытовая химия	Магазин у дома	11.03.2023	9600487	банковская карта
Электроника	Гипермаркет	18.03.2023	4645068	наличные
Молочная	Супермаркет	10.03.2023	1161973	банковская карта
Бытовая химия	Магазин у дома	12.03.2023	4139761	банковская карта
Хлебная	Маркетплейс	05.03.2023	2133590	банковская карта
Молочная	Супермаркет	28.03.2023	539855	банковская карта
Кондитерская продукция	Магазин у дома	23.03.2023	3221705	наличные
Бытовая химия	Маркетплейс	04.03.2023	7693553	наличные
Хлебная	Магазин у дома	20.03.2023	294740	банковская карта
Электроника	Магазин у дома	30.03.2023	1237844	банковская карта
Сантехника	Супермаркет	25.03.2023	861669	банковская карта
Электроника	Супермаркет	07.03.2023	6856185	наличные
Бытовая химия	Маркетплейс	11.03.2023	2127895	наличные
Электроника	Маркетплейс	21.03.2023	3415217	наличные
Электроника	Маркетплейс	24.03.2023	5862728	наличные
Хлебная	Супермаркет	31.03.2023	6687302	наличные
Сантехника	Супермаркет	15.03.2023	8561865	наличные
Кондитерская продукция	Маркетплейс	22.03.2023	2667405	наличные
Молочная	Маркетплейс	28.03.2023	139073	банковская карта
Бытовая химия	Маркетплейс	17.03.2023	9173217	наличные
Электроника	Маркетплейс	25.03.2023	2302261	банковская карта
Сантехника	Магазин у дома	09.03.2023	1856984	наличные
Молочная	Супермаркет	21.03.2023	5627265	наличные
Хлебная	Гипермаркет	06.03.2023	5492212	банковская карта
Сантехника	Супермаркет	23.03.2023	6841575	наличные

Продолжение таблицы 6

Вид продукции	Формат магазина	Дата	Выручка	Формат оплаты
Электроника	Гипермаркет	07.03.2023	3619700	банковская карта
Молочная	Магазин у дома	22.03.2023	1852933	наличные
Кондитерская продукция	Супермаркет	17.03.2023	931774	банковская карта
Молочная	Гипермаркет	09.03.2023	4394860	наличные
Электроника	Гипермаркет	02.03.2023	7106279	наличные
Кондитерская продукция	Гипермаркет	31.03.2023	798576	банковская карта
Бытовая химия	Гипермаркет	27.03.2023	3149721	банковская карта
Сантехника	Магазин у дома	11.03.2023	4914637	наличные
Хлебная	Супермаркет	16.03.2023	2569157	банковская карта
Сантехника	Супермаркет	15.03.2023	7510006	наличные
Хлебная	Супермаркет	14.03.2023	9594849	банковская карта
Бытовая химия	Супермаркет	15.03.2023	6108718	банковская карта
Электроника	Магазин у дома	22.03.2023	9370981	наличные
Электроника	Магазин у дома	01.03.2023	7976456	банковская карта
Хлебная	Супермаркет	09.03.2023	8098299	банковская карта
Кондитерская продукция	Гипермаркет	17.03.2023	9935347	банковская карта
Сантехника	Маркетплейс	01.03.2023	6283674	наличные
Кондитерская продукция	Магазин у дома	14.03.2023	4636369	наличные
Электроника	Маркетплейс	19.03.2023	2573407	банковская карта
Кондитерская продукция	Супермаркет	27.03.2023	2427187	банковская карта
Хлебная	Маркетплейс	12.03.2023	6174986	наличные
Хлебная	Супермаркет	12.03.2023	9922973	банковская карта
Бытовая химия	Маркетплейс	19.03.2023	422251	наличные
Кондитерская продукция	Гипермаркет	22.03.2023	24657	банковская карта
Молочная	Магазин у дома	26.03.2023	8323696	банковская карта

Таким образом, мы осуществляем этап изучения данных, чтобы понять насколько адекватно подготовленный массив данных представляет исследуемую предметную область. Здесь целесообразно проводиться поиск минимальных и максимальных значений параметров, анализ распределений значений и других статистических характеристик, сравнение полученных результатов с представлениями предметной области.

Результат анализа данных исследуемого массива приведён на рисунке 39.

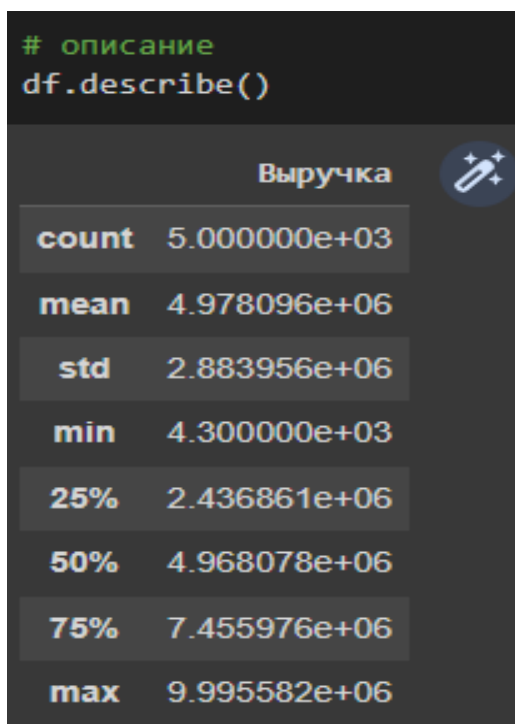


Рисунок 39 – Результат анализа данных исследуемого массива

По результатам анализа данных были выявлены минимальные, максимальные показатели, показатели отклонения статистической выборки. На рисунке 40 приведены параметры исследуемого источника данных.

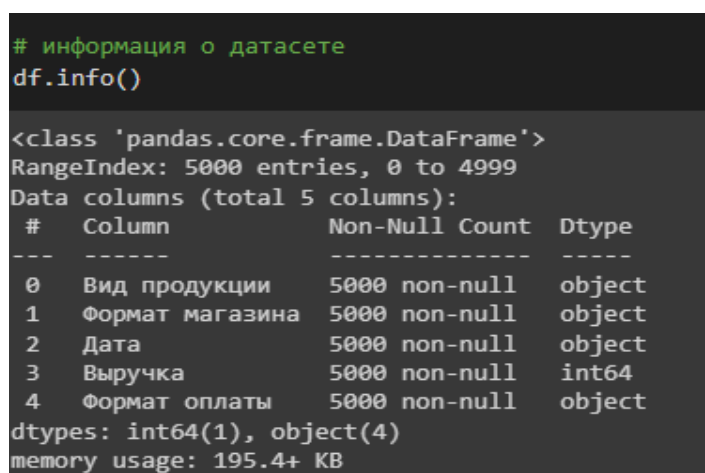


Рисунок 40 – Параметры исследуемого источника данных

На рисунке 41 приведены результаты расчета минимальной, максимальной, средней, суммарной выручки. Эти данные позволят оценить финансовые показатели предприятий сферы торговли и тем самым способствуют принятию управленческих решений.

```
[16] # преобразование в формат даты
df['Дата'] = pd.to_datetime(df['Дата']).dt.date
```

```
[6] # минимальная выручка
df['Выручка'].min()
4300
```

```
[7] #максимальная выручка
df['Выручка'].max()
9995582
```

```
[8] # средняя выручка
df['Выручка'].mean()
4978096.1344
```

Рисунок 41 – Результаты расчета минимальной, максимальной, средней, суммарной выручки

На рисунке 42 приведён график динамики суммарной выручки по датам.

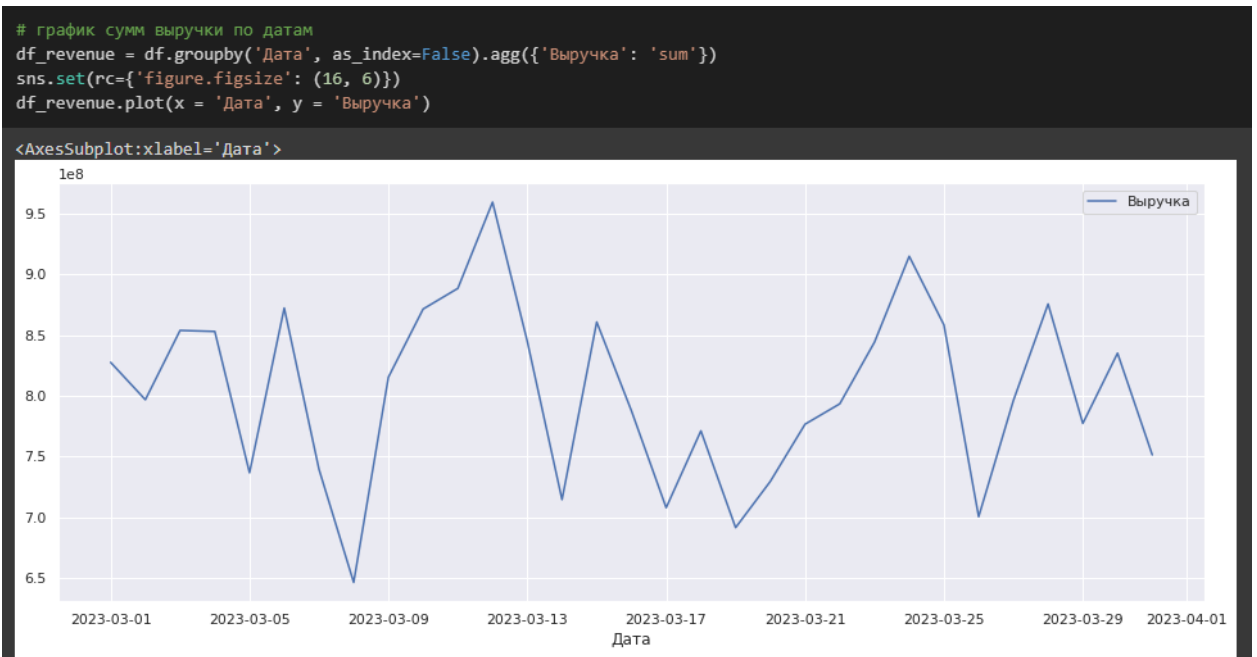


Рисунок 42 – График динамики суммарной выручки по датам

На рисунке 43 приведён результат суммирования выручки по датам.

```

[33] df_revenue

```

	Дата	Выручка
0	2023-03-01	827521181
1	2023-03-02	796734682
2	2023-03-03	853856745
3	2023-03-04	852995933
4	2023-03-05	736517640
5	2023-03-06	872298147
6	2023-03-07	739452886
7	2023-03-08	646158493
8	2023-03-09	815289072
9	2023-03-10	871479200
10	2023-03-11	888511129
11	2023-03-12	959584398
12	2023-03-13	844893842
13	2023-03-14	714396700
14	2023-03-15	860775655
15	2023-03-16	787571478
16	2023-03-17	707720063
17	2023-03-18	771043465
18	2023-03-19	691290481

Рисунок 43 – Результат суммирования выручки по датам

На рисунке 44 приведен график средней выручки по датам. На рисунке 45 – табличная интерпретация результатов.

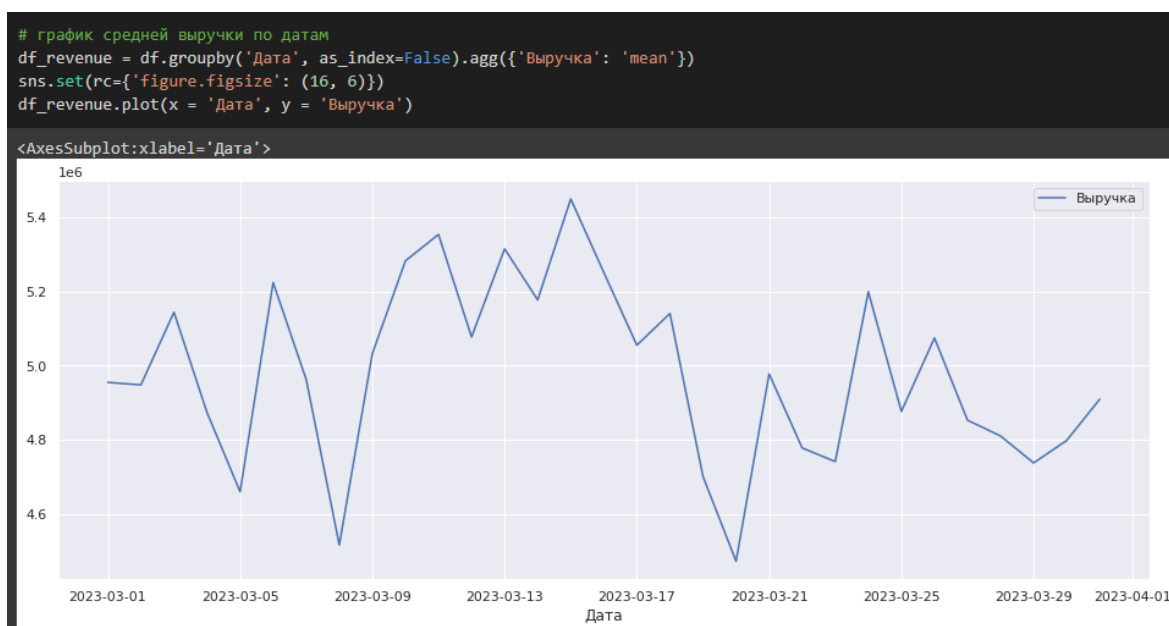


Рисунок 44 – График средней выручки по датам

```
[35] df_revenue
```

	Дата	Выручка
0	2023-03-01	4.955217e+06
1	2023-03-02	4.948663e+06
2	2023-03-03	5.143715e+06
3	2023-03-04	4.874262e+06
4	2023-03-05	4.661504e+06
5	2023-03-06	5.223342e+06
6	2023-03-07	4.962771e+06
7	2023-03-08	4.518591e+06
8	2023-03-09	5.032649e+06
9	2023-03-10	5.281692e+06
10	2023-03-11	5.352477e+06
11	2023-03-12	5.077166e+06
12	2023-03-13	5.313798e+06
13	2023-03-14	5.176788e+06
14	2023-03-15	5.447947e+06
15	2023-03-16	5.250477e+06
16	2023-03-17	5.055143e+06
17	2023-03-18	5.140290e+06
18	2023-03-19	4.702656e+06
19	2023-03-20	4.474962e+06

Рисунок 45 – Табличная интерпретация результатов

На рисунке 46 приведена диаграмма средней выручки по товарным позициям.

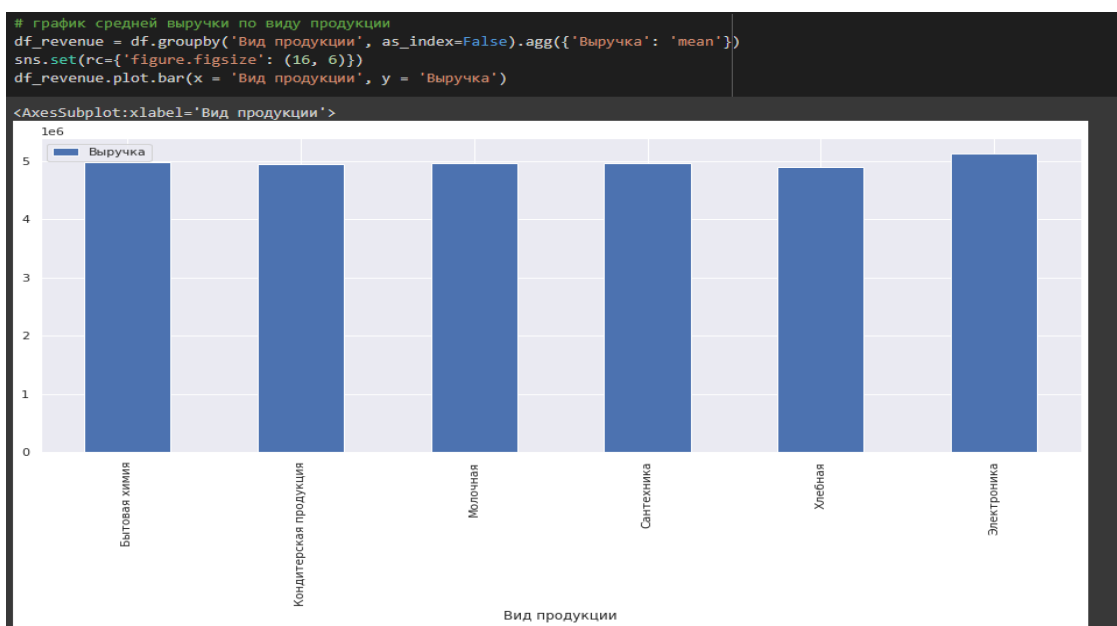


Рисунок 46 – Диаграмма средней выручки по товарным позициям

Как показано на рисунке 46, в рассматриваемой выборке в указанный период наибольший объем средней выручки соответствует продажам электроники, наименьший – продажам хлебобулочных изделий. На рисунке 47 приведена табличная интерпретация расчётов.

```
[37] df_revenue
```

	Вид продукции	Выручка
0	Бытовая химия	4.986353e+06
1	Кондитерская продукция	4.938295e+06
2	Молочная	4.961475e+06
3	Сантехника	4.955746e+06
4	Хлебная	4.887977e+06
5	Электроника	5.135116e+06

Рисунок 45 – Табличная интерпретация результатов

На рисунке 48 приведена диаграмма суммарной выручки по товарным ПОЗИЦИЯМ.

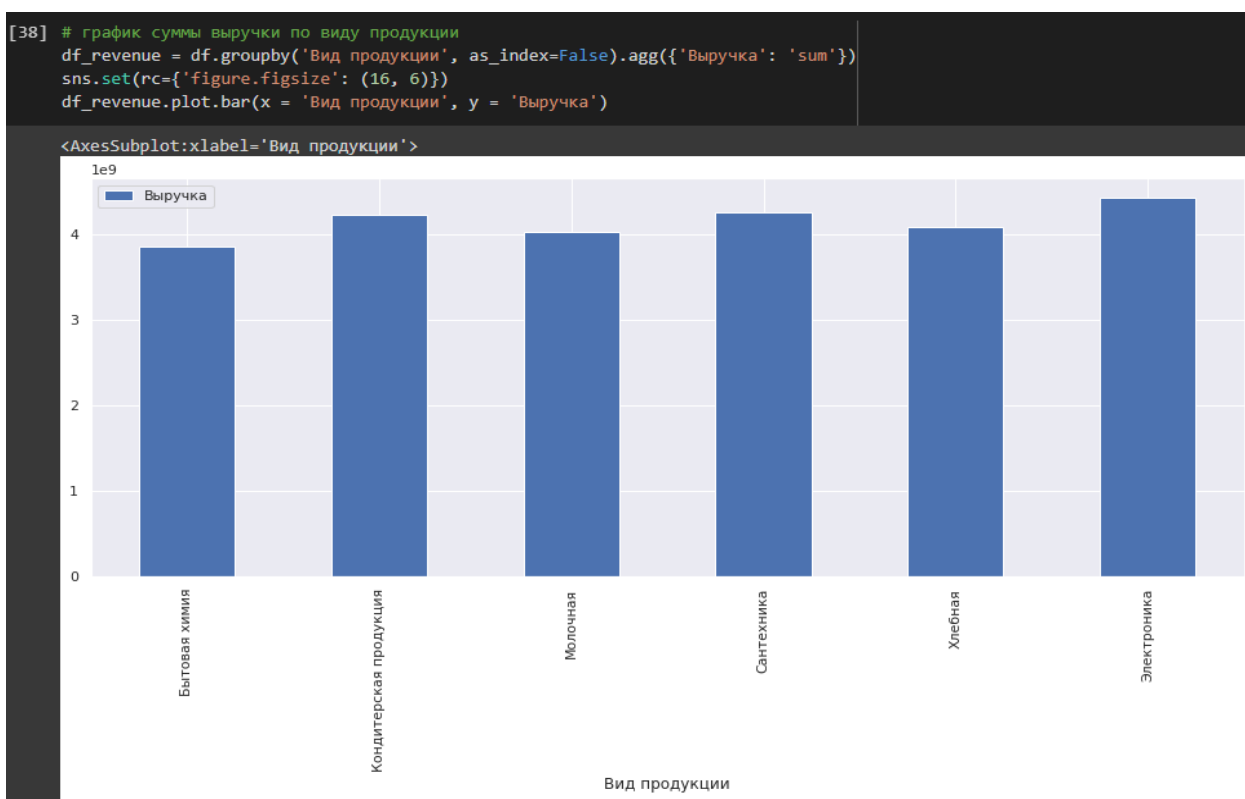


Рисунок 48 – Диаграмма суммарной выручки по товарным позициям

Как показано на рисунке 48, в рассматриваемой выборке в указанный период наибольший объем средней выручки соответствует продажам электроники, наименьший – продажам хлебобулочных изделий. На рисунке 49 приведена табличная интерпретация расчётов.

```
[39] df_revenue
```

	Вид продукции	Выручка
0	Бытовая химия	3859437018
1	Кондитерская продукция	4237056898
2	Молочная	4023756060
3	Сантехника	4252029774
4	Хлебная	4081460764
5	Электроника	4436740158

Рисунок 49 – Табличная интерпретация результатов

На рисунке 50 приведена диаграмма суммарной выручки по форматам магазинов.



Рисунок 50 - Диаграмма суммарной выручки по форматам магазинов

Как показано на рисунке 50, наибольший объем выручки приходится на магазины формата «Супермаркет», минимальный – на маркетплейсы. На рисунке 51 приведена таблица с результатами расчетов.

```
[41] df_revenue
```

	Формат магазина	Выручка
0	Гипермаркет	6501770887
1	Магазин у дома	5951623043
2	Маркетплейс	5938026639
3	Супермаркет	6499060103

Рисунок 51 – Таблица с результатами расчетов

На рисунке 52 приведена диаграмма средней выручки по форматам магазинов.

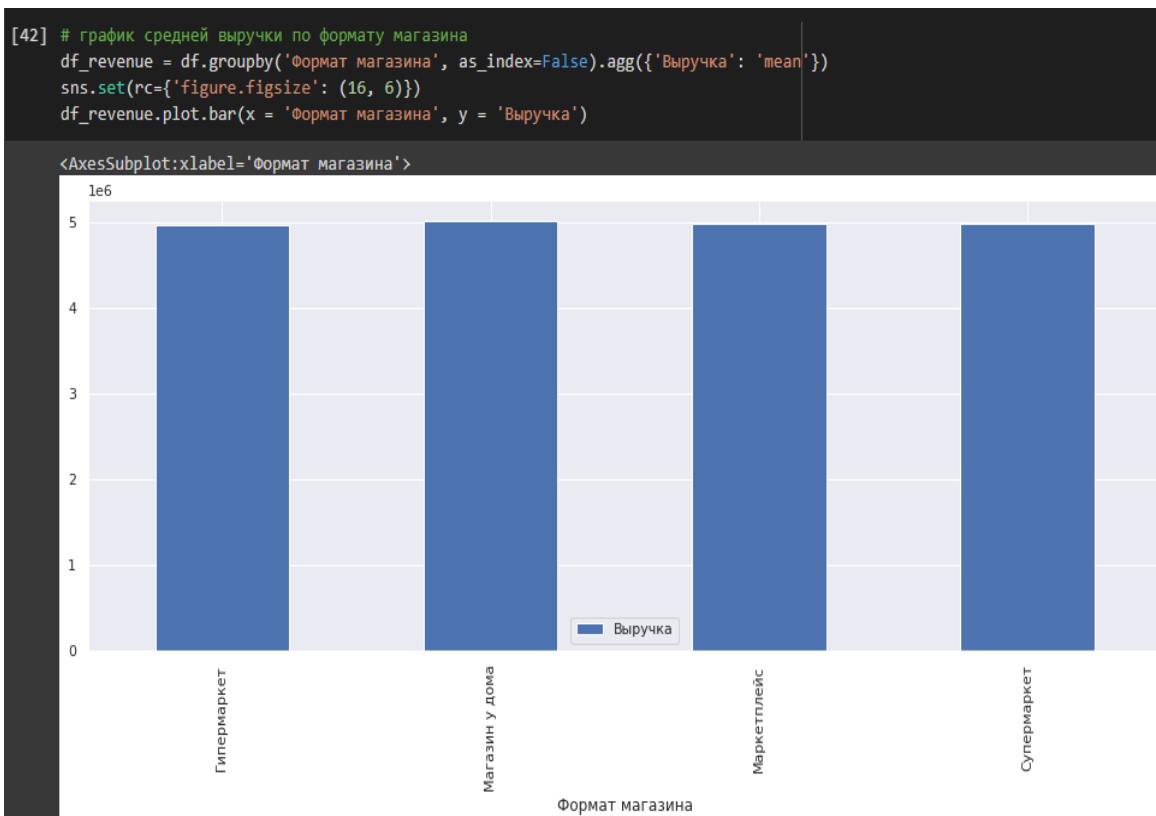


Рисунок 52 - Диаграмма средней выручки по форматам магазинов

По результатам расчётов показано, что наибольшая средняя выручка приходится на формат «Магазин у дома», наименьшая – на гипермаркеты.

```
[43] df_revenue
```

	Формат магазина	Выручка
0	Гипермаркет	4.959398e+06
1	Магазин у дома	5.005570e+06
2	Маркетплейс	4.973222e+06
3	Супермаркет	4.976309e+06

Рисунок 53 – Таблица с результатами расчета

Внедрение систем эквайринга – это значительная статья расхода. Поэтому управленческому персоналу предприятий сферы торговли важно получить точную информацию о необходимости внедрения данной системы. На рисунке 54 приведена диаграмма популярности форм расчётов.



Рисунок 54 - Диаграмма популярности форм расчётов

Как показано на рисунке 55, формы расчетов наличными и безналичными средствами составляют примерно одинаковые доли.

```
df_revenue
```

	Формат оплаты	Выручка
0	банковская карта	4.940730e+06
1	наличные	5.016095e+06

Рисунок 55 – Таблица с результатами расчетов

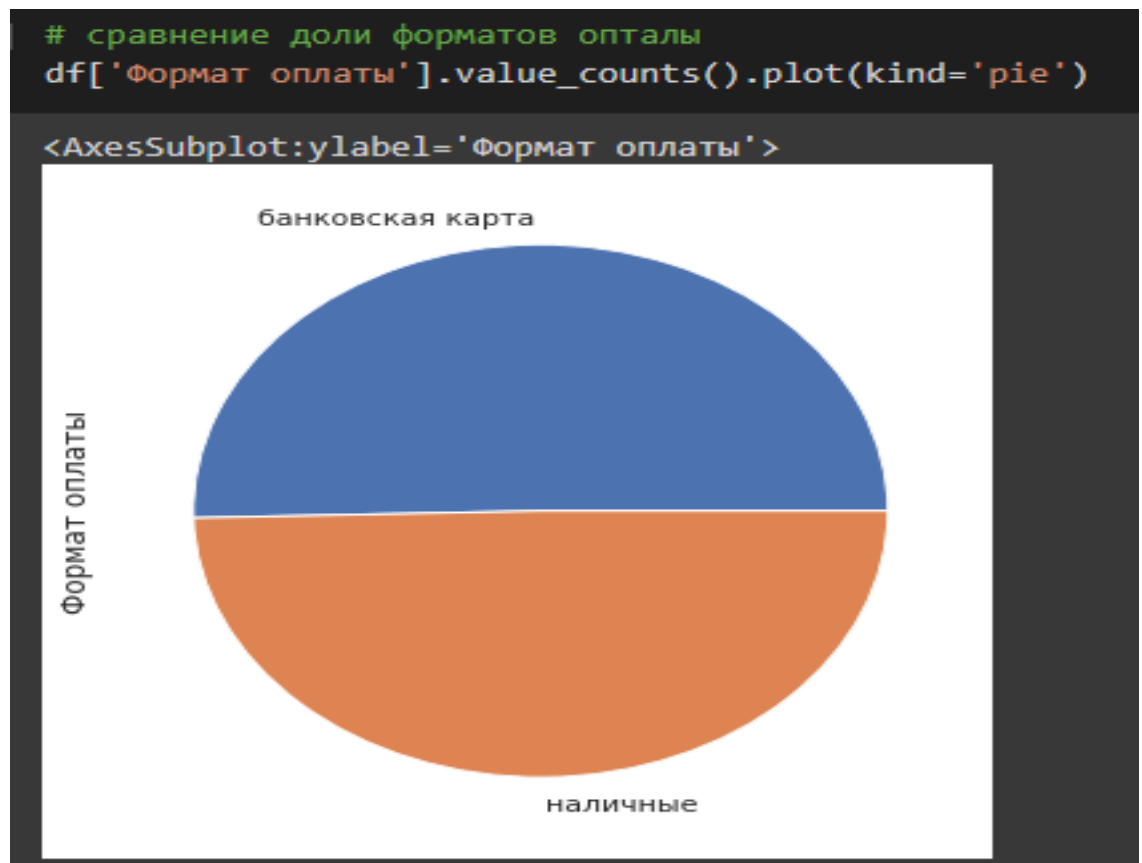


Рисунок 56 - Диаграмма популярности форм расчётов (круговая)

Далее проведен тест на тип распределения выборки (Рисунок 57).

```

import scipy

stat, p = scipy.stats.shapiro(df['Выручка']) # тест Шапиро-Уилк
print('Statistics=%.3f, p-value=%.3f' % (stat, p))

alpha = 0.05
if p > alpha:
    print('Принять гипотезу о нормальности')
else:
    print('Отклонить гипотезу о нормальности')

Statistics=0.954, p-value=0.000
Отклонить гипотезу о нормальности

```

Рисунок 57 – Результаты теста на признаки нормального распределения

Как показано на рисунке 57, гипотеза о нормальности распределения данных по выручке отклоняется.

Далее осуществим построение моделей на основе структуры данных. Модель будет включать указание на алгоритм интеллектуального анализа данных, его параметры и анализируемые данные. При определении модели использованы различные фильтры. Таким образом, не все имеющиеся в описании структуры данные будут использоваться каждой созданной для нее моделью. На рисунке 58 показано как в рамках решения задачи исследования для одной структуры создается несколько моделей, использующие различные наборы столбцов и фильтров.

Для построения моделей целесообразнее всего использовать алгоритм кластерного анализа (K-means) или K-средних с заранее известным числом кластеров. Для удобства выбрано число кластеров, равное 3. Для визуализации используем модуль «Sklearn.cluster» (Python).

На рисунке 59 представлена модель «Средняя выручка по товарным позициям».



Рисунок 58 – Модели, созданные в рамках решения задачи исследования

В модели, изображенной на рисунке 59, выделено три кластера. Модель полностью соответствует данным диаграммы на рисунках 46-47, где наибольшая средняя выручка принадлежит товарной позиции «Электроника», чуть меньше средняя выручка у «Бытовой техники». На третьем месте - категория «Молочные продукты».

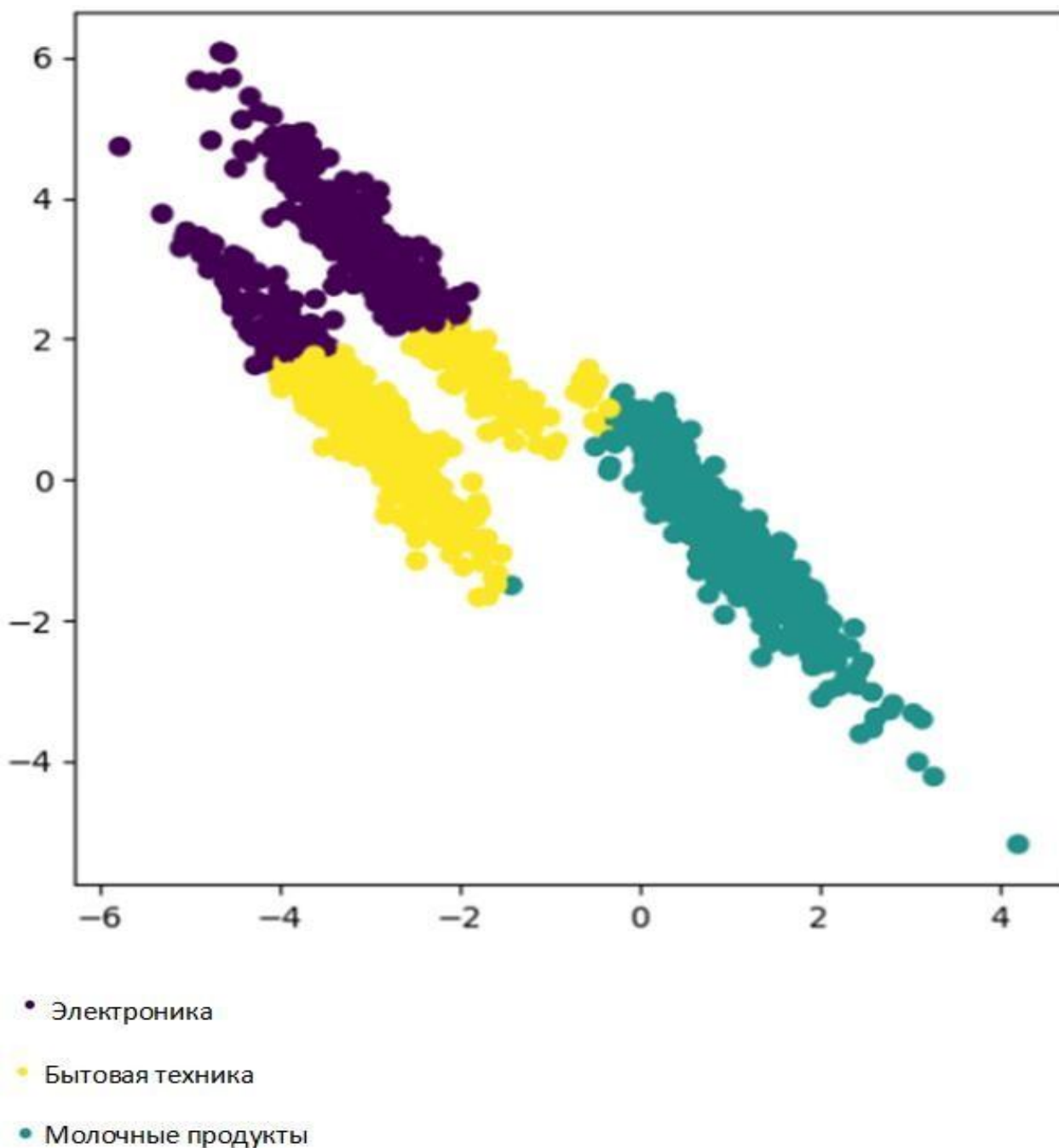


Рисунок 59 – Модель «Средняя выручка по товарным позициям»

На рисунке 60 представлена модель «Суммарная выручка по товарным позициям».

В модели, изображенной на рисунке 60, также выделено три кластера. Модель полностью соответствует данным диаграммы на рисунках 48-49, где наибольшая суммарная выручка принадлежит товарной позиции

«Электроника», чуть меньше суммарная выручка у категории «Сантехника». На третьем месте - «Кондитерские изделия».

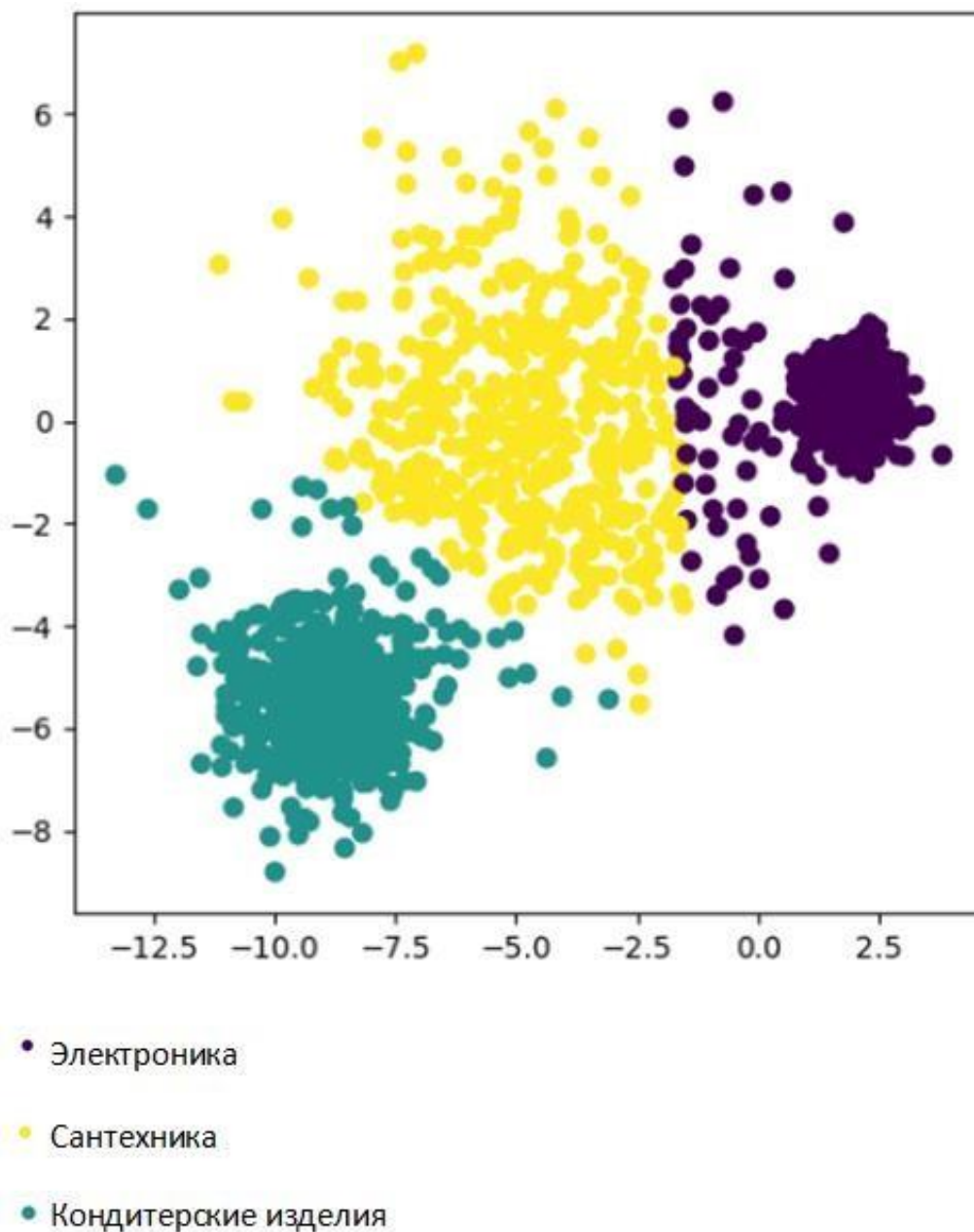


Рисунок 60 – Модель «Суммарная выручка по товарным позициям»

На рисунке 61 представлена модель «Средняя выручка по товарам магазинов».

В модели, изображенной на рисунке 61, выделено три кластера. Модель полностью соответствует данным диаграммы на рисунках 50-51, где

наибольшая средняя выручка по форматам магазинов принадлежит формату «Магазин у дома», на втором месте – «Супермаркет». На третьем месте - «Маркетплейс».

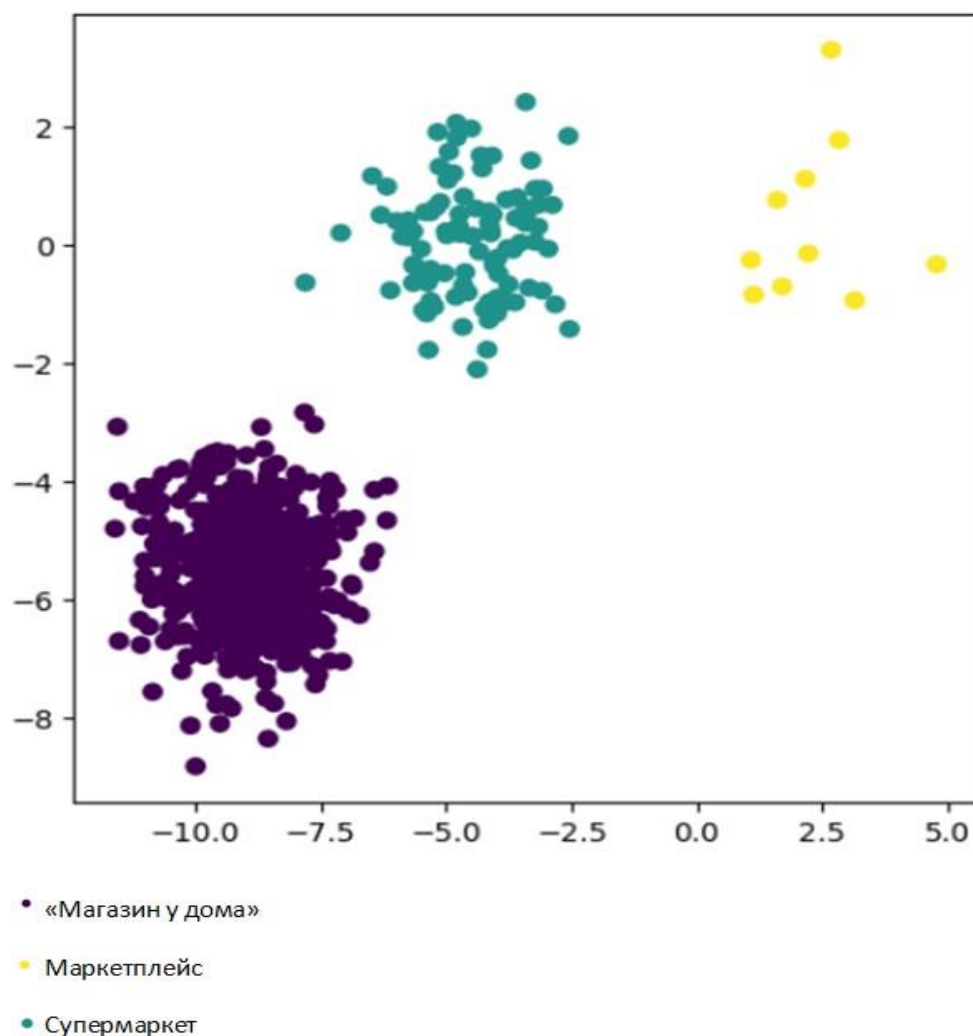


Рисунок 61 – Модель «Средняя выручка по форматам магазинов»

На рисунке 62 представлена модель «Популярность форм расчетов». Для построения данной модели был использован пакет «MiniBatchKMeans».

В модели, изображенной на рисунке 62, выделено два кластера. Модель полностью соответствует данным диаграммы на рисунках 52-53, где примерно одинаковые доли у формы расчетов банковская карта и наличные денежные средства.

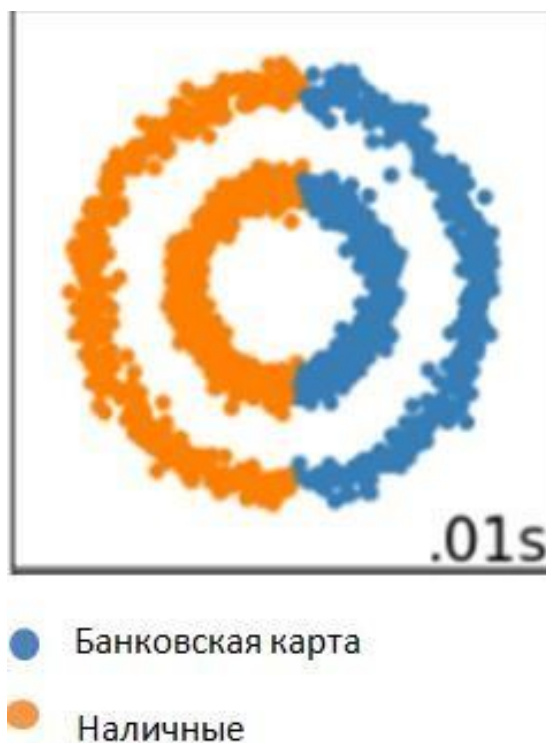


Рисунок 62 – Модель «Популярность форм расчетов»

Выводы по разделу:

В ходе апробации результатов исследования было проведено изучение статистических данных и проведен интеллектуальный анализ данных массивов по продажам товаров различной номенклатуры и динамике их финансового состояния за период 2020-2022 гг.

Интеллектуальный анализ данных был проведен в следующей последовательности:

- постановка задачи;
- подготовка данных;
- изучение данных;
- построение моделей;
- исследование и проверка моделей.

Чуть ранее на этапе постановки задачи определили цель исследования.

В ходе подготовки данных определили источники данных - массивы

информации для интеллектуального анализа и определили критерии, показатели, по которым производили исследование.

Затем осуществили изучение данных, чтобы понять насколько адекватно подготовленный массив данных представлял исследуемую предметную область. Произвели поиск минимальных и максимальных значений параметров, анализ распределений значений и других статистических характеристик. После чего построили и проверили модели.

В рамках осуществления задачи по оптимизации ассортиментной политики провели анализ динамики рынка, структуры спроса, финансовых показателей по товарным группам, отдельным товарам. С использованием сред Pandas, Matplotlib, NumPy, Sklearn.cluster была проведен интеллектуальный анализ, в результате которого получены данные, установлены зависимости между показателями, построены и проверены модели.

Также было проведено изучение финансовой отчетности компаний, работающих в отрасли пищевого производства, проведен анализ дебиторской и кредиторской задолженности. Причина, по которой автор считает целесообразным провести анализ дебиторской и кредиторской задолженности заключается в том, что эффективное управление торговым предприятием – это также вопрос обеспечения приемлемого уровня финансового риска.

По итогам проведённого анализа руководством осуществляется принятие управленческих решений, на основании которых проводится корректировка ассортиментной политики, сокращение издержек. Использование технологий больших данных позволяет вывести деятельность предприятий сфере торговли на качественно новый уровень.

Таким образом, технология анализа и использования больших данных в сфере торговли повышает эффективность управления компаниями, позволяет оптимизировать ассортиментную политику, снижать издержки. Следовательно, гипотеза исследования подтверждена и обоснована.

Заключение

В рамках данной работы проведен анализ теоретических и практических аспектов использования технологий обработки больших данных в сфере торговли. Показано, что в настоящее время технологии работы с массивами данных и накопление больших объемов информации предоставляет возможности проведения обработки, целью которой является выявление трендов, прогнозирование поведения изучаемых систем. Сферами применения технологий обработки больших данных являются: задачи маркетинга, государственного управления, планирование производства и реализации продукции, планирование транспортной инфраструктуры, строительства социальных объектов.

Использование больших данных является актуальным во множестве отраслей экономической деятельности, а также в сфере государственного управления, работе силовых структур, политической деятельности. Методика обработки массивов больших данных включает технологии, использование которых зависит от специфики обрабатываемых данных, периодичности работы с ними, представленных форматов. Использование технологий работы с большими данными в сфере торговли может включать множество направлений, связанных с анализом объемов производства, эффективностью расходования средств, проводимых маркетинговых акций. Также технологии больших данных в условиях предприятий торговли обеспечивают оптимизацию ассортиментной политики, что, в конечном итоге, повышает эффективность деятельности предприятий.

В рамках проведённого анализа основных направлений использования систем интеллектуального анализа больших данных в сфере торговли было показано, что инструменты обработки больших массивов информации позволяют выявлять наличие зависимостей между изучаемыми параметрами, проводить расчёты статистических показателей, оценивать эффективность ассортиментной и бонусной политики. Таким образом, технологии

интеллектуального анализа позволяют получать конкурентные преимущества для предприятий сферы торговли вследствие наличия возможностей проведения оптимизации технологии работы с поставщиками, клиентами, торговыми площадями и персоналом.

Изучены методы обработки больших данных. Показано, что выбор того или иного метода зависит от поставленных задач и специфики предметной области.

В практической части работы проведено изучение статистических данных по продажам товаров различной номенклатуры и динамике их финансового состояния за период 2020-2022 гг.

С использованием среды Pandas была проведена статистическая обработка массива, в результате которой получены сводные показатели по видам произведенной продукции, получаемого дохода, установлены зависимости между показателями.

Также было проведено изучение финансовой отчетности компаний, работающих в отрасли пищевого производства, проведен анализ дебиторской и кредиторской задолженности. Эффективности расходования государственных субсидий, сделаны выводы о наличии или отсутствии корреляции между различными показателями.

Таким образом, использование системы статистической обработки информации обеспечивает возможности анализа деятельности предприятий, оптимизации структуры производства, финансирования основной деятельности, эффективного использования кредитных ресурсов и государственных субсидий.

В практической части работы выполнен статический анализ данных по продажам товаров различной номенклатуры. В качестве параметров для анализа выбраны показатели объемов производства, финансовые результаты, учет качества продукции, штатного состава предприятий. В качестве инструментов статистического анализа выбраны: группировка, вычисление относительных величин, графическое представление данных, корреляция. По

результатам проведенных вычислений определены отрасли с наилучшими показателями деятельности, а также виды производимой продукции с наибольшим маржинальным доходом, проанализирована динамика финансовых показателей.

Таким образом, использование инструментов анализа больших данных в сфере торговли позволяет выявлять эффективность основных направлений производства по видам, выявлять взаимозависимости показателей. По итогам проведенного анализа руководством осуществляется принятие управленческих решений, на основании которых проводится корректировка ассортиментной политики, изменение подходов к финансированию производств, работе с персоналом.

Таким образом, использование технологий больших данных позволяет вывести деятельность предприятий сфере торговли на качественно новый уровень.

Список используемых источников

1. Applied methods of statistical analysis. Statistical computation and simulation : [AMSA'19] : proceedings of the International workshop, [Novosibirsk, Russia], 18-20 September 2019 / editors: Prof. Boris Lemeshko [et al.]. - Novosibirsk : Novosibirsk state technical university, 2019. - 573 с.
2. Evans R. Clustering for classification: using standard clustering methods to summarize datasets with minimal loss of classification accuracy [Text] / R. Evans. - Saarbrucken: VDM Verlag, 2018. - 108 с.
3. Jankowski N. Comparison of instances selection algorithms i. Algorithms survey / N. Jankowski, M. Grochowski // International conference on artificial intelligence and soft computing. - Springer, Berlin, Heidelberg, 2019. – 598 с.
4. Kordos M. Instance selection in logical rule extraction for regression problems / M. Kordos, S. Bialka, M. Blachnik // International Conference on Artificial Intelligence and Soft Computing. - Springer, Berlin, Heidelberg, 2013. – 175 с.
5. Korjakovskij I. Safer reinforcement learning for robotics : dissertation / by Ivan Koryakovskiy. - Moscow : Editus, 2018. - XII, 151 с.
6. Аврунев О. Е., Стасышин В. М. Бизнес-информатика. [Текст] учебное пособие: / О. Е. Аврунев, В. М. Стасышин. - Новосибирск: Изд-во НГТУ, 2018. – 121с.
7. Агамиров Л. В. Статистические методы анализа результатов научных исследований : учебное пособие / Л. В. Агамиров. - Москва: Изд-во МЭИ, 2018. - 71 с.
8. Алпайдин Э. Машинное обучение: новый искусственный интеллект / Этем Алпайдин ; [перевод с английского: Дмитрий Вибе, д.ф.-м.н.]. - Москва : Фонд развития промышленности, 2017. - XII, 191 с.
9. Бабиева Н. А., Раскин Л. И. Автоматизация ИТ-сервисов на

предприятиях. [Текст] : учебно-методическое пособие / Н. А. Бабиева, Л. И. Раскин. – М.: Инфра-М, 2018. – 208 с.

10. Баранчиков А. И. Управление ИТ-инфраструктурой организаций [Текст] : учебник / А. И. Баранчиков. - Рязань: РГУ, 2019. - 219 с.

11. Белоусов П. А., Марухина О. В., Скоморохов А. О. Машинное обучение и большие данные : учебное пособие / П. А. Белоусов, О. В. Марухина, А. О. Скоморохов. - Санкт-Петербург : ГУАП, 2021. - 119 с.

12. Беляева Т. М. Информационные технологии в юридической деятельности [Текст] : учебник /Беляева Т. М., Кудинов А. Т., Пальянова Н.В. - Москва: Проспект, 2018. - 349 с.

13. Бененсон М. З., Сорокин С. А. Проектирование баз данных большого объема : учебное пособие / М. З. Бененсон, С. А. Сорокин. - Москва: МИРЭА - Российский технологический университет, 2020. - 151 с.

14. Бессмертный И. А. Интеллектуальные системы: учебник и практикум / И. А. Бессмертный, А. Б. Нугуманова, А. В. Платонов. — Москва : Издательство Юрайт, 2022. — 243 с.

15. Бессмертный, И. А. Системы искусственного интеллекта : учебное пособие для вузов / И. А. Бессмертный. — 2-е изд., испр. и доп. — Москва : Издательство Юрайт, 2022. — 157 с.

16. Благирев А. Big data простым языком / Алексей Благирев. - Москва : АСТ : Времена, 2019. – 253с.

17. Болдырев А. В. Технологии хранения данных / А. В. Болдырев. - Ростов-на-Дону: ДГТУ, 2019. - 77 с.

18. Борисова С. В., Зейн А. Н., Дурова М. А. Технологии Big Data : учебное пособие / С. В. Борисова, А. Н. Зейн, М. А. Дурова. - Москва : Изд-во МЭИ, 2021. - 46 с.

19. Боченина М. В. Статистические индикаторы и тренды российского рынка / М. В. Боченина. - Санкт-Петербург : Изд-во Санкт-Петербургского государственного экономического университета, 2018. - 165 с.

20. Бродовская Е. В., Домбровская А. Ю. Большие данные в

исследовании политических процессов: учебное пособие / Е. В. Бродовская, А. Ю. Домбровская. - Москва : МПГУ, 2018. - 86

21. Бурков А. Машинное обучение без лишних слов / Андрей Бурков. - Санкт-Петербург: Питер, 2020. - 188 с.

22. Вартанов А. А., Лоскутова И. В., Миронов С. Н. Основы администрирования программных средств защиты информации: теория и практика: монография / Артур Александрович Вартанов, Инна Васильевна Лоскутова, Сергей Николаевич. - Москва: Радиотехника, 2018. - 163 с.

23. Вандер Плас Д. Python для сложных задач, наука о данных: и машинное обучение / Дж. Вандер Плас. - Санкт-Петербург: Питер, 2019. – 572с.

24. Васильев А. А. Теория вероятностей и математическая статистика: учебник и практикум / А. А. Васильев. — 2-е изд., испр. и доп. — Москва: Издательство Юрайт, 2022. — 232 с.

25. Васильков, А.В. Информационные системы и их безопасность [Текст] : Учебное пособие / А.В. Васильков, А.А. Васильков, И.А. Васильков. - М.: Форум, 2018. - 528 с.

26. Воронцов Ю. А., Груничев Ю. А., Козинец А. В. Инновационные технологии в IT. Нейросети: учебно-методическое пособие / Ю.А. Воронцов, Ю.А. Груничев, А.В. Козинец. - Москва : Медиа Паблишер, 2020. - 100 с.

27. Вьюгин В. В. Математические основы машинного обучения и прогнозирования / В. В. Вьюгин. - Изд. 2-е, испр. и доп. - Москва: Изд-во МЦНМО, 2018. – 236с.

28. Гмурман В. Е. Теория вероятностей и математическая статистика: учебник для вузов / В. Е. Гмурман. — 12-е изд. — Москва: Издательство Юрайт, 2022. — 479 с.

29. Григорьев Ю. А., Плужникова О. Ю. Концептуальное и логическое проектирование схемы базы данных в нотации Чена и с помощью CASE-средства AllFusion Erwin Data Modeler : учебно-методическое пособие : / Ю. А. Григорьев, О. Ю. Плужникова. - Москва: Спутник, 2019. - 33 с.

30. Григорьев А. А., Исаев Е. А., Тарасов П. А. Передача, хранение и обработка больших объемов научных данных : учебное пособие / А. А. Григорьев, Е. А. Исаев, П. А. Тарасов. - Москва : ИНФРА-М, 2021. – 205 с.
31. Данилова Н.В., Белявский Г. И. Анализ данных. Машинное обучение: учебное пособие / Н. В. Данилова, Г. И. Белявский. - Ростов-на-Дону: Изд-во Южного федерального университета, 2020. - 146 с.
32. Дейтел П. Д., Дейтел Х. М. Python : искусственный интеллект, большие данные и облачные вычисления / Пол Дейтел, Харви Дейтел. - Санкт-Петербург: Питер, 2020. - 861 с.
33. Дергачева И. В., Сарьян А. С. Цифровые технологии. Анализ больших данных: учебное пособие / И. В. Дергачева, А. С. Сарьян. - Ростов-на-Дону : ФГБОУ ВО РГУПС, 2019. - 114 с.
34. Донников Ю.Е. Проектирование баз данных: монография / Ю.Е. Донников. - Москва : Юстицинформ, 2021. – 219 с.
35. Дудин М. Н. Социально-экономическая статистика: учебник и практикум для вузов / М. Н. Дудин, Н. В. Лясников, М. Л. Лезина. — Москва : Издательство Юрайт, 2022. — 233 с.
36. Дудин М. Н. Теория статистики : учебник и практикум для вузов / М. Н. Дудин, Н. В. Лясников, М. Л. Лезина. — Москва : Издательство Юрайт, 2022. — 148 с.
37. Евтихов В. Г., Евтихова Н. В., Суворов С. В. Акселерация обработки данных : учебно-методическое пособие / В. Г. Евтихов, Н. В. Евтихов, С. В. Суворов. - Москва: Московский Политех, 2021. – 589с.
38. Ерохин В. В., Раянова А. Р., Кафтарева А. Б. Статистические методы анализа в бизнес-информатике : учебное пособие / В. В. Ерохин, А. Р. Раянова, А. Б. Кафтарева. - 2-е изд., перераб. и доп. - Москва : Спутник+, 2022. - 163 с.
39. Еськин Д. Л., Бакулин В. М. Основы защиты информации в компьютерных системах и сетях: учебное пособие / Д. Л. Еськин, В. М. Бакулин. - Волгоград : ВА МВД России, 2019. - 67 с.

40. Железнов М. М. Методы и технологии обработки больших данных: учебно-методическое пособие / М.М. Железнов. МГСУ. - Москва: Изд-во МИСИ - МГСУ, 2020. – 658с.
41. Замятин А. В. Интеллектуальный анализ данных: учебное пособие / А. В. Замятин. - Томск: Издательский Дом Томского государственного университета, 2020. - 193 с.
42. Зараменских Е. П. Менеджмент: бизнес-информатика : учебник и практикум / Е. П. Зараменских. — Москва : Издательство Юрайт, 2022. — 407 с.
43. Зимин В.В. Искусственный интеллект в информационной безопасности: учебное пособие / В. В. Зимин. - Кемерово: Кузбассвузиздат, 2018. – 499 с.
44. Зыков Р.В. Роман с Data Science: как монетизировать большие данные / Зыков Роман. - Санкт-Петербург: Питер : Прогресс книга, 2021. - 318 с.
45. Иванов Д. Б. Разработка системы управления данными: учебник / Иванов Д.Б. - Воронеж, 2018. – ВГУ, 2018. - 120 с.
46. Казакова, Н. А. Анализ финансовой отчетности. Консолидированный бизнес: учебник / Н. А. Казакова. — Москва: Издательство Юрайт, 2022. — 233 с.
47. Казанцев С.Я. Информационные технологии: учебное пособие / С. Я. Казанцев. - Москва: ЮНИТИ-Дана, 2020. - 351 с.
48. Калинина В. Н. Теория вероятностей и математическая статистика: учебник для вузов / В. Н. Калинина. — 2-е изд., перераб. и доп. — Москва: Издательство Юрайт, 2022. — 472 с.
49. Камалова Г. Г. Технологии анализа данных: монография / Камалова Гульфия Гафиятовна. - Саратов: Амирит, 2019. - 160 с.
50. Келлехер Д., Тирни Б. Наука о данных: базовый курс / Джон Келлехер, Брендан Тирни. - Москва : Альпина Пабlishер, 2020. – 220 с.
51. Килин П. М. Статистические методы сбора и обработки данных:

учебное пособие / П. М. Килин. - Тюмень : ТИУ, 2017. - 126 с.

52. Классен Р. К. Консервативные СУБД класса BigData с регулярным планом обработки запросов на кластерной платформе / Классен Роман Константинович. - Казань, 2019. - 26 с.

53. Котова Е. Е., Плахотников Д. П. Введение в аналитику больших данных: учебно-методическое пособие / Е. Е. Котова, Д. П. Плахотников. - Санкт-Петербург: Изд-во СПбГЭТУ, 2021. – 47с.

54. Кошкарров А. В. Основы аналитики больших данных: учебно-методическое пособие / А. В. Кошкарров. - Астрахань: Сорокин Роман Васильевич, 2018. - 75 с.

55. Кремер Н. Ш. Математическая статистика: учебник и практикум / Н. Ш. Кремер. — Москва : Издательство Юрайт, 2022. — 259 с.

56. Кувайскова Ю. Е., Клячкин В. Н. Статистические методы прогнозирования: учебное пособие / Ю. Е. Кувайскова, В. Н. Клячкин. - Ульяновск : УлГТУ, 2019. - 197 с.

57. Кудрявцев, В. Б. Интеллектуальные системы : учебник и практикум для вузов / В. Б. Кудрявцев, Э. Э. Гасанов, А. С. Подколзин. — 2-е изд., испр. и доп. — Москва : Издательство Юрайт, 2022. — 165 с.

58. Липина О.И. Большие данные: принятия решений, этика: курс лекций / ред.: О.И. Липина, А.А. Цветков. - Рязань : Полиспектр, 2021. - 88 с.

59. Логачёв М. С. Информационные системы и программирование. Администратор баз данных: учебник / М.С. Логачёв. - Москва: ИНФРА-М, 2020. - 438 с.

60. Лукьянова Е. А., Шимкевич Е. М., Ляпунова Т. В. Статистические методы анализа : учебное пособие / Е.А. Лукьянова, Е.М. Шимкевич, Т.В. Ляпунова. - Москва: Российский университет дружбы народов, 2020. – 116с.

61. Макаренко С. В. Методы аналитической обработки данных, полученных в результате исследований: учебное пособие / С.В. Макаренко. - Иркутск: Изд-во Иркутского национального исследовательского технического университета, 2021. - 91с.

62. Мамадалиева Л. Н., Хаконова И. М. Обучение математическим методам обработки экспериментальных данных в русле компетентностного подхода : учебное пособие / Л. Н. Мамадалиева, И. М. Хаконова. - Майкоп : ИП Кучеренко В. О., 2018. - 100 с.
63. Марц Н., Уоррен Д. Большие данные: принципы и практика построения масштабируемых систем обработки данных в реальном времени / Натан Марц, Джеймс Уоррен. - Москва: Вильямс, 2018. - 368 с.
64. Минаева Ю. В. Методы статистического и интеллектуального анализа данных : учебное пособие / Ю. В. Минаева. - Воронеж: Воронежский государственный технический университет, 2017. - 90 с.
65. Миронов А.М. Машинное обучение: учебное пособие / А.М. Миронов. - Москва: Макс-Пресс, 2018. - 21 с.
66. Михайлов Г. А. Статистическое моделирование. Методы Монте-Карло : учебное пособие для вузов / Г. А. Михайлов, А. В. Войтишек. — Москва : Издательство Юрайт, 2022. — 323 с.
67. Набережная Ж. Б., Набережная И. Б., Нимгирова А.С. Корреляция / Ж.Б. Набережная, И.Б. Набережная, А.С. Нимгирова. - Астрахань: Изд-во Астраханского государственного медицинского университета, 2017. - 40 с.
68. Назаров Д. М. Интеллектуальные системы: основы теории нечетких множеств: учебное пособие для вузов / Д. М. Назаров, Л. К. Конышева. — 3-е изд., испр. и доп. — Москва : Издательство Юрайт, 2022. — 186 с.
69. Никифоров С. Н. анализ бизнес-процессов: учебное пособие / С.Н. Никифоров. - Санкт-Петербург: СПбГАСУ, 2017. – 76с.
70. Основы технологий Big Data : учебное пособие / В. В. Ухлоva. - Воронеж: Издательский дом ВГУ, 2020. - 80 с.
71. Парамонов И. Ю., Смагин В. А., Косых Н. Е., Хомоненко А. Д. Методы и модели исследования сложных систем и обработки больших данных: монография / И. Ю. Парамонов, В. А. Смагин, Н. Е. Косых, А. Д. Хомоненко. - Санкт-Петербург: ЛАНЬ, 2020. – 234с.

72. Першина Е. Л., Чуканов С. Н. Машинное обучение [Электронный ресурс]: монография / Е.Л. Першина, С.Н. Чуканов. - Омск: ФГБОУ ВО "СибАДИ", 2017. – 258с.
73. Плас Д. В. Python для сложных задач: наука о данных и машинное обучение / Дж. Вандер Плас. - Санкт-Петербург: Питер, 2020. –572 с.
74. Политический анализ и прогнозирование: учебник для вузов / под общей редакцией В. А. Семенова. — 2-е изд. — Москва: Издательство Юрайт, 2022. — 433 с.
75. Полищук Ю. В., Боровский А. С. Базы данных и их безопасность : учебное пособие / Ю. В. Полищук, А. С. Боровский. - Москва: ИНФРА-М, 2021. - 207 с.
76. Рожков Н. Н. Статистические методы контроля и управления качеством продукции : учебное пособие для вузов / Н. Н. Рожков. — Москва: Издательство Юрайт, 2022. — 154 с.
77. Рудзей Г. Ф. Статистические методы анализа: учебное пособие / Г. Ф. Рудзей. - Новосибирск: Изд-во Сибирского государственного университета путей сообщения, 2017. – 119с.
78. Ручкин В. Н., Костров Б. В., Свирина А. Г. Системы искусственного интеллекта. Нейросети и нейрокомпьютеры: учебник для студентов высших учебных заведений / В. Н. Ручкин, Б. В. Костров, А. Г. Свирина. - Москва: Курс, 2018. – 282 с.
79. Сарычева Т. В., Иванова Д. Д., Фомичева Ю.П. Статистические подходы к анализу и прогнозированию развития цифровой экономики в Российской Федерации и ее регионах : монография / Сарычева Татьяна Владимировна, Иванова Дарья Дмитриевна, Фомичева Юлия Павловна. - Йошкар-Ола: Марийский государственный университет, 2020. - 158 с.

80. Селеменова Т. А., Крюкова М. С. Математические методы обработки экспериментальных данных: учебное пособие / Т. А. Селеменова, М. С. Крюкова. - Санкт-Петербург: Санкт-Петербургский университет ГПС МЧС России, 2019. – 166с.
81. Семёнычев В. К., Кожухова В. Н., Коробецкая А. А. Технологии и инструментарий анализа больших данных [Электронный ресурс] / В.К. Семёнычев, В.Н. Кожухова, А.А. Коробецкая. - Самара : ФГБОУ ВО "СГЭУ", 2019. – 589с.
82. Сенько А. Работа с BigData в облаках: обработка и хранение данных с примерами из Microsoft Azure / Александр Сенько. - Санкт-Петербург : Питер, 2019. – 445с.
83. Скиена С. С. Наука о данных: учебный курс / Стивен С. Скиена. - Санкт-Петербург: Диалектика, 2020. - 543 с.
84. Станкевич Л. А. Интеллектуальные системы и технологии : учебник и практикум / Л. А. Станкевич. — Москва : Издательство Юрайт, 2022. — 397 с.
85. Токарев В. В. Методы оптимизации: учебное пособие для вузов / В. В. Токарев. — Москва : Издательство Юрайт, 2022. — 440 с.
86. Уатт Д., Борхани Р., Катсаггелос А. Машинное обучение: основы, алгоритмы и практика применения / Джереми Уатт, Реза Борхани, Аггелос Катсаггелос ; перевод с английского Андрея Логунова. - Санкт-Петербург: БХВ-Петербург, 2022. - 612 с.
87. Шкурко В. Е. Статистические исследования в сферах предпринимательства и экономической безопасности: учебное пособие / В. Е. Шкурко. - Екатеринбург: Изд-во Уральского университета, 2020. – 172с.
88. Шолле Ф. Глубокое обучение на Python / Франсуа Шолле. - Санкт-Петербург: Питер, 2019. - 397 с.
89. Шорохова И. С., Мариев О. С., Кисляк Н. В. Статистический анализ экономических данных: учебное пособие / И. С. Шорохова, О. С.

Мариев, Н. В. Кисляк. - Екатеринбург: Изд-во Уральского университета, 2021.
– 338 с.

90. Ын А., Су К. Теоретический минимум по Big Data : всё, что нужно знать о больших данных / Анналин Ын, Кеннет Су. - Санкт- Петербург: Питер, 2020. - 205 с.

91. Яковлев, В. Б. Статистика: учебное пособие для вузов / В. Б. Яковлев. — 2-е изд., испр. и доп. — Москва : Издательство Юрайт, 2022. — 353 с.

92. Ямпольская, Д. О. Маркетинговый анализ: технология и методы проведения : учебник и практикум для вузов / Д. О. Ямпольская, А. И. Пилипенко. — Москва : Издательство Юрайт, 2022. — 268 с.