

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РФ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«ПЕНЗЕНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
ПОЛИТЕХНИЧЕСКИЙ ИНСТИТУТ

Факультет
вычислительной техники

Кафедра
МОиПЭВМ

Направление подготовки 09.03.02 «Информационные системы и технологии»

БАКАЛАВРСКАЯ РАБОТА

на тему

**Статистический анализ данных
инструментальными средствами
платформы Logiном**

Студент _____ Садомов Данила Константинович
(подпись, дата) (ФИО полностью)

Руководитель _____ Макарычев П.П.
(подпись, дата) (фамилия, инициалы)

Нормоконтролёр _____ Такташкин Д.В.
(подпись, дата) (фамилия, инициалы)

Работа допущена к защите (протокол заседания кафедры от _____ № _____)

Заведующий кафедрой _____ Козлов А.Ю.
(подпись) (фамилия, инициалы)

Работа защищена с отметкой _____ (протокол заседания ГЭК от _____ № _____)

Секретарь ГЭК _____ Попова Н.А.
(подпись) (фамилия, инициалы)

Пенза, 2023

Федеральное государственное бюджетное образовательное учреждение
высшего образования

ПЕНЗЕНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ПОЛИТЕХНИЧЕСКИЙ ИНСТИТУТ

Факультет
Вычислительной техники

Кафедра
МО и ПЭВМ

«Утверждаю»
Заведующий кафедрой
_____ А.Ю. Козлов
«__» _____ 2023 г

**ЗАДАНИЕ
НА ВЫПУСКНУЮ КВАЛИФИКАЦИОННУЮ РАБОТУ
БАКАЛАВРА**

1. Студент Садомов Данила Константинович гр 19ВИ1 факультета
<Фамилия, имя, отчество полностью>
вычислительной техники

2. Тема работы Статистический анализ данных инструментальными средствами
платформы Logipom

Тема утверждена приказом ПГУ № _____ от " ____ " _____ 20 ____

2. Руководитель работы д.т.н. профессор каф. «МО и ПЭВМ» Макарычев П.П.

3. Задание на работу (назначение разработки, исходные данные и т.п.)
Произвести статистический анализ данных инструментальными средствами
платформы Logipom. Рассмотреть существующие способы анализа и выбрать наиболее
подходящий. Изучить платформу Logipom и обосновать выбор данного приложения. Logipom
должен принимать на вход базу данных, обрабатывать ее и выдавать проведенный анализ,
выраженный через графические представление.

4. Перечень подлежащих решению вопросов:

1) анализ требований проведения анализа;

2) постановка задачи выполнения анализа;

3) решение поставленной задачи проведения анализа;

4) интерпретация полученного результата;

5) оформление пояснительной записки.

5. Календарный график выполнения работы

п/п	Наименование этапов работы	Объем работы	Срок выполнения	Подпись руководителя
	Анализ требований проведения анализа	10%	12.04.2023 – 16.04.2023	
	Постановка задачи выполнения анализа	20%	17.04.2023 – 26.04.2023	
	Решение поставленной задачи анализа	30%	27.04.2023 – 11.05.2023	
	Интерпретация полученного результата	20%	12.05.2023 – 21.05.2023	
	Оформление пояснительной записки	20%	22.05.2023 – 01.06.2023	

Дата выдачи задания «22» декабря 2022г.

Руководитель бакалаврской работы _____
(подпись, дата)

Задание к исполнению принял студент _____
(подпись, дата)

Работу к защите допустить
Декан ФВТ _____ Л.Р. Фионова
(подпись, дата)

Реферат

Пояснительная записка содержит 58 листов, 27 рисунков, 10 таблицы, 37 использованных источников и 2 приложения.

СТАТИСТИЧЕСКИЙ АНАЛИЗ, LOGINOM, МЕТОДЫ АНАЛИЗА, КОРРЕЛЯЦИЯ, ФАКТОРЫ, БАЗЫ ДАННЫХ.

Объект исследования: статистический анализ данных.

Цель работы: выполнить статистический анализ данных с помощью инструментальных средств Loginom.

Средство выполнение задач: Loginom.

Результат работы: произведен статистический анализ данных инструментальными средствами платформы Loginom.

Область применения: анализ данных.

					ПГУ 09.03.02 – 8БР191.19 ПЗ			
Изм.	Лист	№ докум.	Подпись	Дата	Статистический анализ данных инструментальными средствами платформы Loginom. Пояснительная записка.	Лит.	Лист	Листов
Разраб.		Садомов. Д.К.						
Провер.		Макарычев П.П.					4	58
Реценз.						группа 19ВИ1		
Н. Контр.		Такташкин Д.В.						
Утверд.		Козлов А.Ю.						

Содержание

Введение.....	6
1 Постановка задачи и анализ требований	8
1.1 Постановка задачи и способ решения	8
1.2 Преимущества выбранной платформы над аналогами.....	8
1.3 Анализ требований и предметной области	9
1.4 Оценка бюджета разработки.....	11
2 Статистический анализ данных.....	13
2.1 Методы статистического анализа	14
2.1.1 Метод корреляции	14
2.1.2 Факторный метод.....	15
2.1.3 Дескриптивная статистика.....	16
2.1.4 Регрессионный анализ.....	17
2.1.5 Анализ дисперсии.....	18
2.1.6 Тестирование гипотез.....	19
2.1.7 Анализ временных рядов	20
2.1.8 Анализ главных компонент	21
2.2 Обоснование выбора методов статистического анализа данных	22
3 Платформа решения задач анализа Loginom	25
3.1 Интерфейс программы Loginom	25
3.2 Компоненты платформы Loginom	27
4 Выполнение статистического анализа данных	35
4.1 Импортирование данных	36
4.2 Корреляционный анализ данных	38
4.3 Факторный анализ данных.....	40
4.4 Выводы проведенного анализа.....	43
Заключение	45
Список использованных источников	46
Приложение А. Глоссарий	51
Приложение Б. Таблицы данных по Пензенской области.....	53

Введение

Международные, государственные и частные организации все чаще открывают свои данные для повторного использования. Большая часть открытых данных связана со статистикой: демографическими (например, данные переписи), экономическими и социальными показателями (например, количество новых предприятий, уровень безработицы). Открытые многомерные статистические данные сегодня составляют важную основу для ускорения социально-экономического развития путем создания новых общественно значимых государственных, муниципальных некоммерческих и коммерческих продуктов.

В наши дни актуальность статистического анализа неоспорима, поскольку есть доступ к огромному количеству данных, которые нужно обрабатывать и анализировать в поисках новых знаний и решений. Каждый день появляются новые методы и инструменты статистического анализа, увеличивая его эффективность и достоверность результатов.

Loginot – это платформа для анализа данных, которая позволяет собирать, обрабатывать, анализировать и визуализировать данные из разных источников.

Основная функция Loginot – это автоматизированный анализ данных, который осуществляется при помощи технологий и методологий бизнес-аналитики. С помощью Loginot можно создавать отчеты, dashboard (приложение А), аналитические модели и прогнозы, проводить корреляционный и регрессионный анализы.

Loginot также позволяет проводить анализ данных в режиме реального времени, что особенно важно для бизнеса, где каждая минута может стоить дорого.

Одной из основных преимуществ Loginot является его удобный и интуитивно понятный интерфейс, который позволяет быстро и легко создавать аналитические отчеты и настраивать параметры анализа.

Кроме того, Loginom имеет широкие возможности по интеграции с различными источниками данных, включая веб-страницы, базы данных, файлы, API и другие системы.

1 Постановка задачи и анализ требований

1.1 Постановка задачи и способ решения

Данная выпускная квалификационная работа направлена на решение задач статистического анализа данных инструментальными средствами Logiном.

Первой задачей аналитика является сбор данных, статистический анализ, которых требуется провести. После формирования базы данных с помощью таблиц или иных способов представления данных.

При помощи платформы Logiном аналитик должен импортировать базу данных и провести статистический анализ используя инструментальные средства, такие как, корреляционный и факторный анализы.

Итогом проделанного анализа должно быть графическое представление данных, на основании которого будут делаться выводы.

Для решений поставленных задач была выбрана платформа Logiном отечественного производства, которая предоставляет инструментальные средства корреляционного и факторного анализов данных, а также дает возможность импортировать данные из вне платформы, что отлично подходит для выполнения поставленной задачи.

1.2 Преимущества выбранной платформы над аналогами

Далее будут перечислены основные преимущества данной платформы на фоне аналогов:

- масштабируемость и высокая производительность, что позволяет управлять большими объемами данных и высокой нагрузкой на систему;
- удобный интерфейс и простота использования, что позволяет максимально быстро и эффективно управлять пользователями и их правами;
- разнообразие опций автоматизации и настройки системы, что упрощает процессы работы с множеством пользователей;

- гибкая настройка под нужды конкретного бизнеса и возможность интеграции с различными информационными системами;
- бесплатная, отечественного производства.

Эти преимущества делают платформу Loginom одной из лучших на рынке и удобной в использовании для управления пользователями в организациях любого типа.

1.3 Анализ требований и предметной области

Решение задачи аналитика, должно позволять проводить корреляционный и факторный анализы статистических данных с помощью платформы Loginom. Для разрабатываемого сценария можно выделить следующие требования:

- в сценарии должна быть возможность импортировать базу данных из таблицы;
- сценарий должен передавать данные для корреляционного или факторного анализа в компоненты исследования платформы Loginom.
- Существует один вид пользователя, который будет использовать данный сценарий статистического анализа данных: аналитик.

Для отображения функций данного действующего лица была построена диаграмма вариантов использования, которая приведена на рисунке 1.

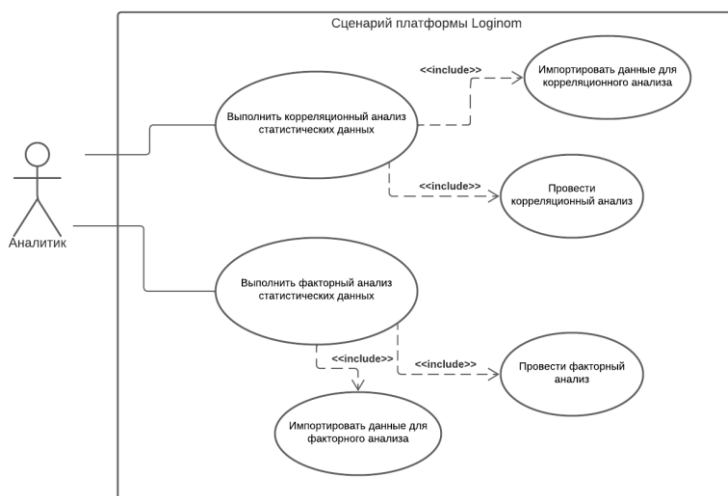


Рисунок 1 – Диаграмма вариантов использования сценария платформы

Краткое описание действующего лица, изображенного на диаграмме вариантов использования, представлено в таблице 1.

Таблица 1 – Описание действующего лица

Пользователь	Описание
Аналитик	Пользователь может выполнить корреляционный или факторный анализ данных

Для аналитика можно выделить следующие варианты использования:

- выполнить корреляционный анализ данных;
- выполнить факторный анализ данных.

Краткое описание вариантов использования описано в таблице 2.

Таблица 2 – Краткое описание вариантов использования

Исполнитель	Цель	Краткое описание
Аналитик	Выполнить корреляционный анализ статистических данных	Пользователь проводит анализ данных с помощью выбранного корреляционного метода
Аналитик	Выполнить факторный анализ данных	Пользователь проводит анализ данных с помощью выбранного факторного метода
Аналитик	Провести корреляционный анализ данных	Пользователь использует инструментальное средство корреляционного анализа
Аналитик	Провести факторный анализ данных	Пользователь использует инструментальное средство факторного анализа
Аналитик	Импортировать данные для корреляционного анализа	Пользователь передает базу данных из файла в сценарий для корреляционного анализа
Аналитик	Импортировать данные для факторного анализа	Пользователь передает базу данных из файла в сценарий для факторного анализа

Таким образом были изучены основные функциональные требования к сценарию по аналитики статистических данных на платформе Loginom.

1.4 Оценка бюджета разработки

При помощи онлайн конструктора таблиц, был разработан план по выполнения статистического анализа данных (рисунок 2).

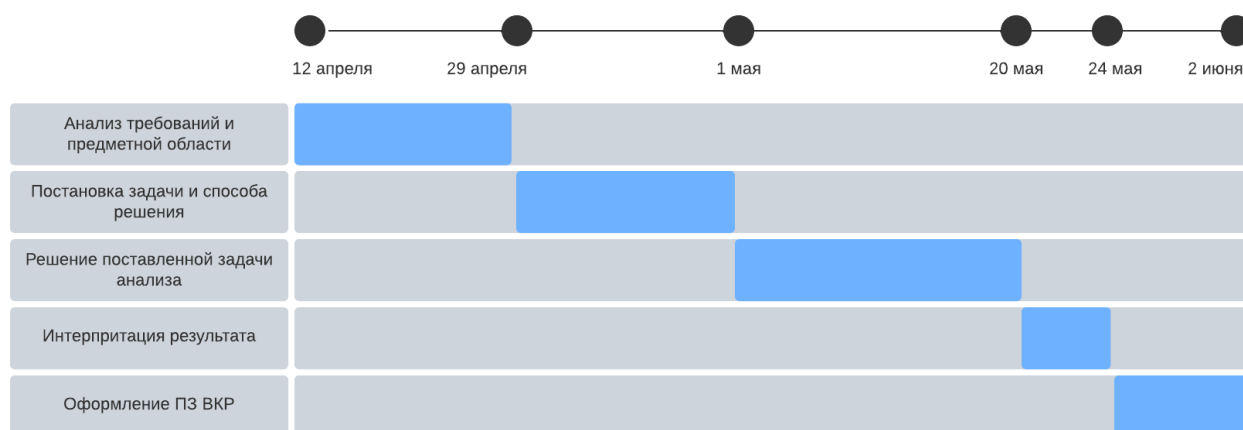


Рисунок 2 – Диаграмма Ганта

После определения задач проект, необходимо проанализировать ресурсы распределения на каждую задачу.

В ходе выполнения работы были оценены ресурсы проекта. Они представляют из себя необходимые для завершения процесса решения задач анализа компоненты, такие как оборудование, трудовые затраты, оплата аналитикам. Для полноценной оценки стоимости проекта было определено количество часов, потраченное на выполнение задач. Так же были проанализированы средние заработные платы аналитиков по России. Далее представлена таблица 3, которая описывает затраченные ресурсы на выполнение поставленной задачи статистического анализа данных инструментальными средствами платформы Loginom.

Исполнителем является аналитик, этапы: анализ требований, постановка задачи и способа решения, решение поставленной задачи анализа и интерпретация результатов анализа.

Таблица 3 – Результаты анализа затрат

Исполнитель	Этап	Время выполнения, часов	Денежные выплаты, рублей в час
Аналитик	Анализ требований и предметной области	25	430
	Постановка задачи и способа решения	15	
	Решение поставленной задачи анализа	45	
	Интерпретация результатов анализа	10	
Loginom	-	-	-
Итого			40850

В результате проведения функционально-стоимостного анализа проекта по выполнению статистического анализа данных была определена суммарная стоимость, равная 40850 рублей, которая учитывает расходы на все этапы выполнения проекта.

2 **Статистический анализ данных**

Статистический анализ – это процесс сбора, обработки и интерпретации данных с целью выявления закономерностей и делающихся выводов на основе численных методов и моделей.

Статистический анализ нужен для принятия важных решений, основанных на объективных данных. Он может помочь узнать тенденции, сравнить результаты, оценить варианты, провести анализ рисков и т. д. В целом, статистический анализ помогает более точно оценить текущую ситуацию и прогнозировать будущие изменения [1].

Анализ данных имеет множество значений в современном мире. Он используется в различных областях, включая экономику, медицину, социологию, психологию и многие другие.

Один из основных примеров использования статистического анализа данных – это принятие бизнес-решений. Многие компании используют данные, чтобы определить успешность продуктов, определить оптимальные цены и расписание продаж. Статистический анализ данных также может помочь компаниям измерить производительность и определить области, которые нуждаются в улучшении.

Статистический анализ данных также используется для оценки качества услуг и продуктов в медицинской и фармацевтической промышленности. Врачи и исследователи используют статистический анализ, чтобы определить, какие методы лечения наиболее эффективны, а также оценить безопасность новых лекарств и процедур.

В социологии и политической науке статистический анализ данных используется для измерения настроений общества и определения тенденций в политических процессах. Статистические методы также могут помочь рассчитать размер и динамику различных групп или населенных пунктов, что

может помочь правительствам и органам управления в планировании развития регионов.

Тем самым можно прийти к выводу о том, что статистический анализ данных играет важную роль в современном мире, помогая принимать обоснованные решения в различных областях.

2.1 Методы статистического анализа

Методы статистического анализа – это совокупность математических инструментов и технологий, которые используются для обработки, интерпретации и создание выводов из статистических данных [2].

Методы статистического анализа данных – это различные подходы и техники, используемые для извлечения значимой информации из наборов данных. Они включают в себя множество методов, таких как описательная статистика, регрессионный анализ, корреляционный анализ, анализ переменных, факторный анализ, анализ временных рядов, анализ кластеров, Анализ главных компонент и т. д.

Они могут быть использованы для выявления трендов, анализа паттернов, поиска взаимосвязей между переменными, прогнозирования будущих значений или для деления наблюдений на группы, которые имеют схожие характеристики.

2.1.1 Метод корреляции

Метод корреляции – это метод, используемый для оценки степени связи различных переменных, которые могут быть связаны между собой [3]. Метод корреляции (также известен как коэффициент корреляции) используется для определения степени связи (корреляции) между двумя переменными. Корреляция может быть положительной (когда значения одной переменной увеличиваются со значением другой переменной) или отрицательной (когда значения одной переменной уменьшаются со значением другой переменной).

Шаги метода корреляции:

- 1) выбор двух переменных, которые будут сравниваться;
- 2) собрать данные для каждого из параметров;
- 3) подсчитать среднее значение для каждого параметра;
- 4) вычислить соответствующие отклонения для каждого параметра от их среднего значения;
- 5) умножить отклонения для каждого параметра, чтобы получить соответствующие произведения;
- 6) сложить все произведения из шага 5 вместе;
- 7) разделите результат из шага 6 на произведение стандартных отклонений для каждого параметра;
- 8) полученное число является коэффициентом корреляции.

Чем ближе значение коэффициента корреляции к минус один или плюс один, тем сильнее связь между двумя переменными. Если коэффициент корреляции ближе к нулю, то связь между переменными слабее. Если коэффициент корреляции равен нулю, то связь между двумя переменными отсутствует. Коэффициент корреляции может быть в диапазоне от минус один до плюс один.

2.1.2 Факторный метод

Факторный анализ статистических данных – это метод анализа множества переменных (факторов), который позволяет выявить главные факторы, оказывающие значимое влияние на изучаемые явления, и установить взаимосвязь между ними [4].

Этот метод используется для сжатия и упрощения данных, чтобы можно было увидеть более ясную картину, скрытую за множеством различных переменных. Факторный анализ позволяет определить число основных факторов, которые объясняют наибольшую часть вариации данных и являются ответом на главные вопросы исследования.

Факторный анализ широко используется в многих областях, включая науку, экономику, социологию, маркетинг и другие дисциплины. Он помогает выявить основные факторы, определяющие изучаемые явления, разработать модели, улучшить прогнозирование и принять лучшие решения.

2.1.3 Дескриптивная статистика

Дескриптивная статистика – это методика, которая используется для описания и анализа исследовательских данных, используя различные меры центральной тенденции и меры разброса [5]. Этот метод предназначен для описания основных характеристик данных, таких как:

- среднее значение;
- медиана;
- мода;
- дисперсия;
- стандартное отклонение;
- квартили;
- межквартильный размах;
- диапазон.

Среднее значение – это среднее значение данных, которое получается путем деления суммы всех значений на количество значений.

Медиана – это значение, которое находится в центре набора данных. Для нахождения медианы данные сортируются по порядку и выбирается среднее значение.

Мода – это самое частое значение, которое встречается в наборе данных.

Дисперсия – это мера, которая отображает, насколько данные разнообразны. Чем больше дисперсия, тем больше разброс данных.

Стандартное отклонение – это мера распределения значений вокруг среднего значения. Чем больше стандартное отклонение, тем больше разброс данных.

Квартили – это значения, которые делят данные на четыре равные части. Первый квартиль – это значение, которое меньше 25 процентов значений. Второй квартиль – это медиана, которая находится на 50 процентов значений. Третий квартиль – это значение, которое меньше семидесяти пяти процентов значений.

Межквартильный размах – это разница между третьим и первым квартилями.

Диапазон – это разница между максимальным и минимальным значениями.

Эти показатели являются основными характеристиками распределения данных и могут использоваться для определения формы распределения, наличия выбросов и других статистических свойств данных.

2.1.4 Регрессионный анализ

Регрессионный анализ – это метод, используемый для моделирования отношений между зависимой переменной и независимыми переменными. Целью регрессионного анализа является построение модели, которая предсказывает значения зависимой переменной на основе значений независимых переменных [6].

Первым шагом в регрессионном анализе является определение зависимой и независимых переменных. Зависимая переменная – это та, которую надо предсказывать, а независимые переменные – это те, которые могут влиять на зависимую переменную.

Затем проводится анализ данных и создается модель регрессии. Существует два вида моделей регрессии: простая линейная регрессия и множественная линейная регрессия.

Простая линейная регрессия используется в том случае, когда имеем только одну независимую переменную, которая влияет на зависимую переменную. Множественная линейная регрессия используется тогда, когда есть

несколько независимых переменных, которые влияют на зависимую переменную.

После того, как модель регрессии была создана, ее необходимо проверить на точность предсказания. Для этого используются различные статистические методы, такие как коэффициент детерминации и F-тест.

Когда модель регрессии была проверена на точность, можем использовать ее для предсказания значений зависимой переменной на основе значений независимых переменных. Это может помочь нам понять, как изменения в независимых переменных могут повлиять на зависимую переменную и какие действия можем предпринять, чтобы достичь желаемого результата.

2.1.5 Анализ дисперсии

Анализ дисперсии – это статистический метод, используемый для определения того, насколько значимыми являются различия между средними значениями двух и более групп данных [7].

Процесс анализа дисперсии включает в себя сравнение разброса данных внутри каждой группы с разбросом между группами. Если внутригрупповой разброс мал по сравнению с межгрупповым разбросом, можно сделать вывод о том, что группы статистически значимо различаются.

Последовательность шагов при анализе дисперсии:

- формулируется гипотеза о равенстве средних значений в группах;
- собираются данные, которые находятся в соответствии с гипотезой;
- вычисляется общая дисперсия для всех данных и разброс, обусловленный фактором (в данном случае группами);
- вычисляется F-статистика – отношение межгруппового разброса к внутригрупповому разбросу;
- сравниваются значения F-статистики со значением критической F-статистики, определенной на уровне значимости;

– если значение F больше критической F , то гипотеза о равенстве средних значений отклоняется.

Анализ дисперсии используется в различных областях, таких как экспериментальная психология, биология, медицина, инженерия, экономика, программирование и т. д.

2.1.6 Тестирование гипотез

Тестирование гипотез – это метод, используемый для проверки статистических гипотез и оценки их значимости. Ниже приведены шаги, которые обычно включаются в процесс тестирования гипотез:

- формулировка гипотезы: в этом шаге определяем теорию, которую хотим проверить, гипотеза должна быть ясной, конкретной и проверяемой;
- определение критериев успеха: в этом шаге определяем, что будет служить доказательством того, что гипотеза верна или ложна;
- сбор данных: в зависимости от гипотезы, можем собирать данные, чтобы проверить ее, это может включать опросы, эксперименты, анализ данных и т. д.;
- анализ данных: после сбора данных анализируем их, чтобы посмотреть, соответствует ли гипотеза нашим критериям успеха или нет;
- выводы: в завершении тестирования гипотез делаем выводы на основе наших результатов;
- документирование результатов: результирующие данные и выводы должны быть очевидны, тщательно идентифицированы и проверены.

Если гипотеза была подтверждена, можем сделать общие выводы и дать рекомендации на основе наших результатов. Если гипотеза была опровергнута, можем сделать вывод о том, что наша теория не верна и пересмотреть ее.

Чтобы обеспечить надежный результат, данный процесс также должен быть документирован и передан для дальнейшего изучения.

Этот процесс требует точности, аккуратности и объективности. Важно знать, как правильно тестировать гипотезы, чтобы получить наиболее полезную и достоверную информацию.

Таким образом тестирование гипотез отлично подходит для проверки статистических гипотез и оценки их значимости [8].

2.1.7 Анализ временных рядов

Анализ временных рядов – это метод, который используется для предсказания будущих значений переменных на основе их предыдущих значений [9].

Анализ временных рядов является важным инструментом в исследовании и прогнозировании различных процессов и явлений, которые изменяются во времени. Шаги, необходимые для проведения анализа временных рядов:

- сбор данных: собираются исторические данные о событиях, объектах или явлениях, которые происходили в определенный период времени;
- визуализация данных: данные представляются в виде графиков, чтобы проанализировать особенности временного ряда, такие как тренды, сезонность и цикличность;
- предварительный анализ: проводятся статистические тесты и анализ данных для выявления аномалий, выбросов и пропущенных значений.
- оценка тренда: тренд определяет общее направление изменения временного ряда, для определения тренда можно использовать метод скользящей средней или экспоненциального сглаживания;
- анализ сезонности: сезонность отображает регулярные повторяющиеся паттерны в временных рядах;
- анализ цикличности: цикличность определяется периодическими колебаниями в данных;

- прогнозирование: на основе анализа тренда, сезонности и цикличности можно предсказать будущую тенденцию изменения временного ряда;
- оценка точности прогноза: оценивается точность прогнозирования с помощью сравнения фактических данных с прогнозируемыми значениями.

Анализируя сезонность можно определить, в каких периодах времени наибольшее количество событий или явлений происходит.

Анализ цикличности может помочь в выявлении трендов, связанных с экономическими или другими циклическими явлениями.

Для прогнозирования могут использоваться различные методы, такие как метод ARIMA (авторегрессионная интегрированная скользящая средняя), метод экспоненциального сглаживания и другие.

Метрики, используемые для оценки точности, могут включать среднеквадратичную ошибку, коэффициент детерминации и другие.

Суммируя, анализ временных рядов позволяет выявлять тренды, сезонность и цикличность в данных, что помогает в более эффективном управлении бизнес-процессами, прогнозировании экономических или социальных явлений и в других областях.

2.1.8 Анализ главных компонент

Анализ главных компонент – это метод, который используется для редуцирования размерности данных и выделения главных компонент из множества переменных [10].

Главные компоненты выбираются таким образом, чтобы они максимально объясняли исходную дисперсию данных и были ортогональны друг другу.

Анализ главных компонент широко используется в различных областях, таких как статистика, машинное обучение и data science. Он позволяет снизить размерность данных, упростить их интерпретацию и улучшить производительность модели.

Процедура анализа главных компонент состоит из нескольких шагов:

- стандартизация данных: перед началом процедуры РСА необходимо стандартизировать данные путем вычитания среднего значения и деления на стандартное отклонение;
- расчет ковариационной матрицы: затем рассчитывается ковариационная матрица, которая показывает, как различные признаки взаимодействуют друг с другом;
- расчет главных компонент: затем производится расчет главных компонент, которые являются линейными комбинациями исходных признаков;
- выбор числа главных компонент: необходимо выбрать оптимальное число главных компонент, которые нужно оставить;
- преобразование данных: после выбора количества главных компонент происходит преобразование данных путем проекции на новые оси.

Главные компоненты сортируются в порядке убывания доли объясненной дисперсии.

Обычно выбирают количество главных компонент, которые объясняют 80 процентов дисперсии.

Преимущества использования метода анализа главных компонент включают более простую и интерпретируемую модель, улучшенную производительность модели и сокращение размерности данных. Однако, анализ главных компонент может потерять важную информацию из данных, если не выбрать правильное количество главных компонент или если исходные признаки имеют сложные взаимосвязи.

2.2 Обоснование выбора методов статистического анализа данных

В данной разработке статистический анализ данных будет производиться с помощью таких инструментальных средств как корреляционный и факторный анализы.

Корреляционный анализ данных выбирается в тех случаях, когда требуется оценить степень взаимосвязи между двумя или более переменными. Например, при исследовании зависимости между ростом и весом человека, между доходом и уровнем образования, между интенсивностью трафика и загруженностью серверов и т. д.

Корреляционный анализ позволяет оценить силу и направление взаимосвязи между переменными, а также выявить выбросы и аномалии в данных. Это позволяет принимать более обоснованные решения на основе данных и избегать ошибок при интерпретации результатов.

Кроме того, корреляционный анализ может помочь выявить скрытые зависимости, которые могут не быть очевидными при первом взгляде на данные. Например, при исследовании зависимости между потреблением кофе и заболеваниями сердца может быть обнаружено, что существует сильная скрытая зависимость между потреблением кофе и курением, что приводит к недостоверности результатов.

Таким образом, корреляционный анализ данных является мощным инструментом для выявления зависимостей и определения степени взаимосвязи между переменными, что позволяет принимать более обоснованные решения в различных областях деятельности, включая науку, медицину, экономику, социологию, технику и т. д.

Когда имеем дело с большим количеством данных, нам необходимо иметь эффективный метод для их анализа. Один из наиболее эффективных методов – это факторный анализ данных. Преимущества этого метода перед остальными будут описаны далее.

Во-первых, факторный анализ данных позволяет нам выделить скрытые факторы, которые оказывают влияние на данные. Эти факторы могут оказаться критическими и могут помочь нам понять, какие переменные наиболее важны для нашего анализа.

Во-вторых, факторный анализ данных упрощает процесс анализа данных. Это связано с тем, что метод позволяет свести множество переменных к более небольшому количеству факторов, которые объясняют основную часть вариации в данных. Таким образом, можем сократить нагрузку на нашу систему и уменьшить вероятность ошибок.

В-третьих, факторный анализ данных не является предпочтительным методом только из-за своей эффективности, но также из-за его гибкости. Этот метод может быть использован в различных ситуациях и применен к различным видам данных.

Тем самым, факторный анализ данных является приоритетным методом анализа данных, потому что позволяет выделить скрытые факторы, снизить нагрузку на систему и обладает гибкостью.

Итого, корреляционный и факторный анализы статистических данных – два наиболее подходящих метода для статистического анализа данных.

3 Платформа решения задач анализа Loginom

Loginom – это платформа, разработанная для решения задач анализа данных и построения бизнес-отчетов. Она позволяет проводить широкий спектр статистических анализов, таких как дисперсионный анализ, корреляционный анализ, регрессионный анализ, анализ временных рядов и др.

Данная платформа является мощным инструментом для обработки больших объемов данных и создания гибких и интуитивно понятных отчетов. Благодаря своей функциональности, платформа Loginom может быть использована в различных сферах, например, в маркетинге, финансах, телекоммуникациях и промышленности. В Loginom используются инновационные технологии, которые обеспечивают быстрый и точный анализ данных.

3.1 Интерфейс программы Loginom

После открытия приложения, в случае настольной редакции Loginom, или авторизации, в случае серверной редакции Loginom, появляется страница «Начало» (рисунок 3).

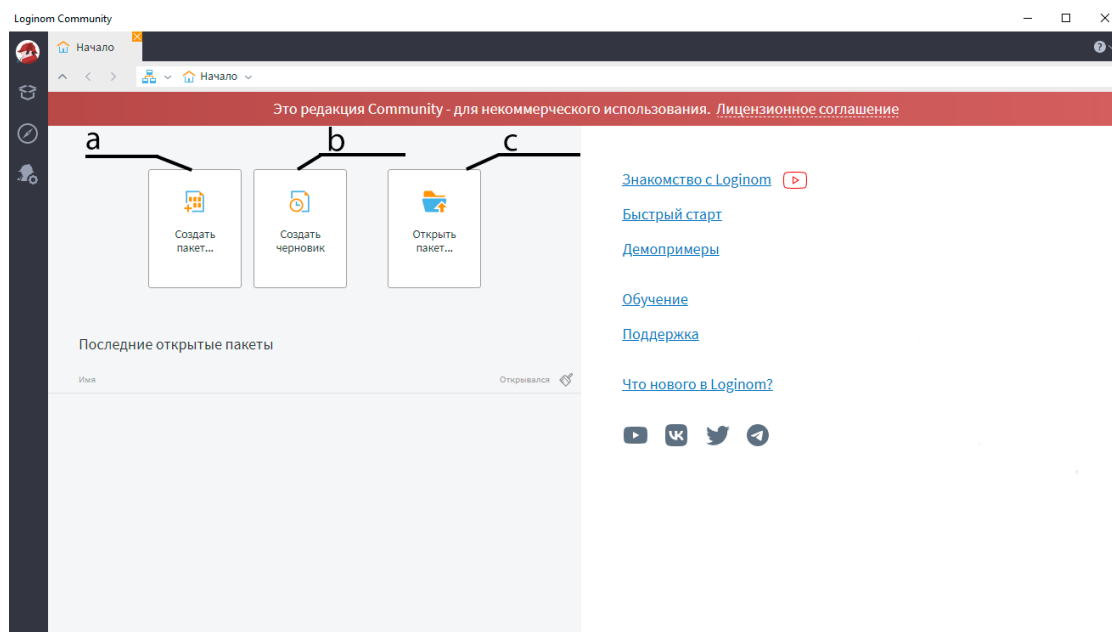


Рисунок 3 – Страница «Начало»

Данная страница содержит следующие команды для манипуляции с пакетами:

- а) «Создать новый пакет» – позволяет создать и сохранить новый «Пакет»;
- б) «Создать черновик» – создает временный «Пакет» и позволяет работать с ним, не сохраняя его;
- с) «Открыть пакет» – позволяет открыть существующий «Пакет»;

Ниже расположена графа «Последние открытые пакеты», включающая в себя историю открытия пакетов.

После создания или открытия пакета появится главное рабочее окно программы, состоящее из трех основных блоков (рисунок 4).

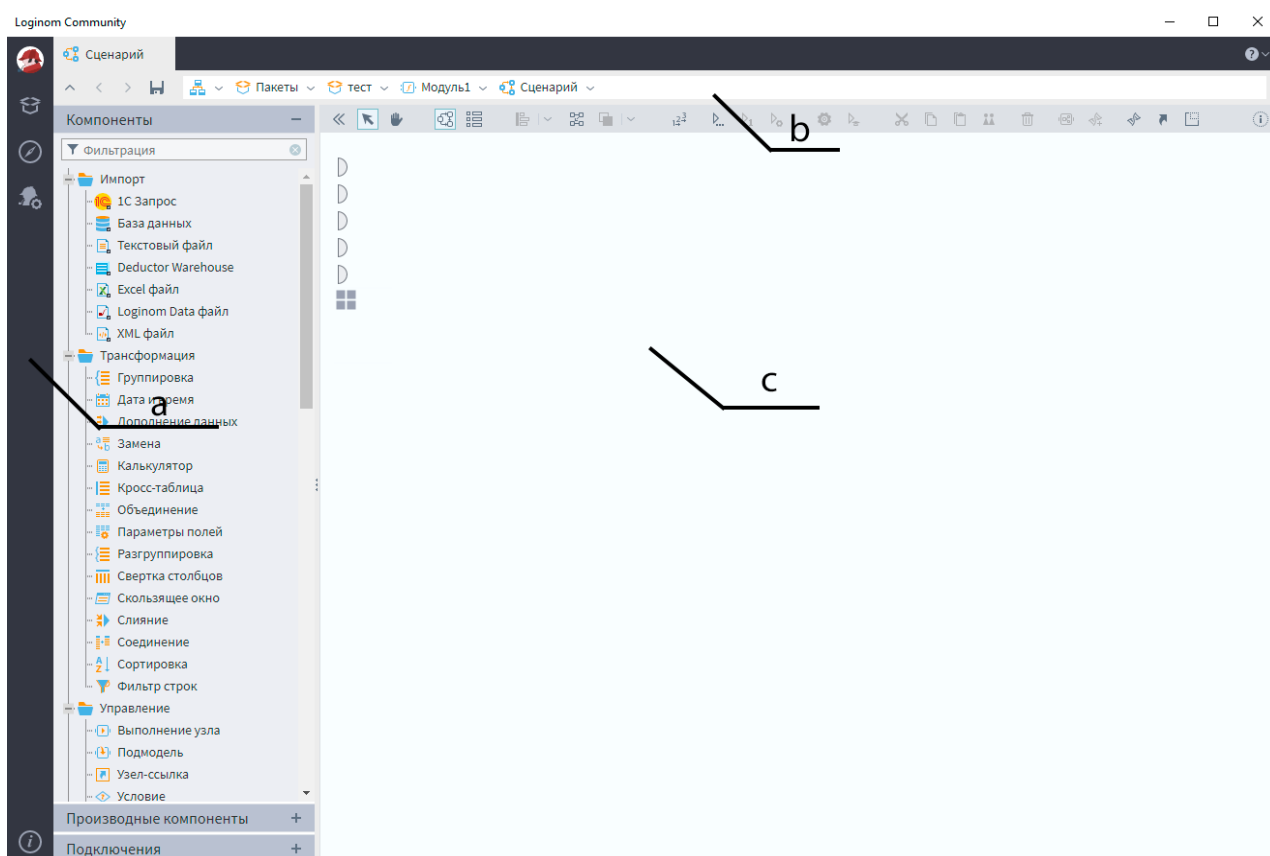


Рисунок 4 – Рабочее окно Loginom

Основные блоки рабочего окна:

- а) главное меню – панель с кнопками для манипуляции с различными настройками;
- б) адресная строка – строка, содержащая путь к открытому объекту;
- с) рабочее пространство – панель компонентов, панель инструментов и область построения Сценариев.

3.2 Компоненты платформы Loginom

Платформа Loginom имеет большое количество разнообразных инструментальных средств, помогающих провести информационный анализ. Рассмотрим некоторые из них.

Компоненты «Импорт» отвечают за импортирование данных для анализа (рисунок 5).

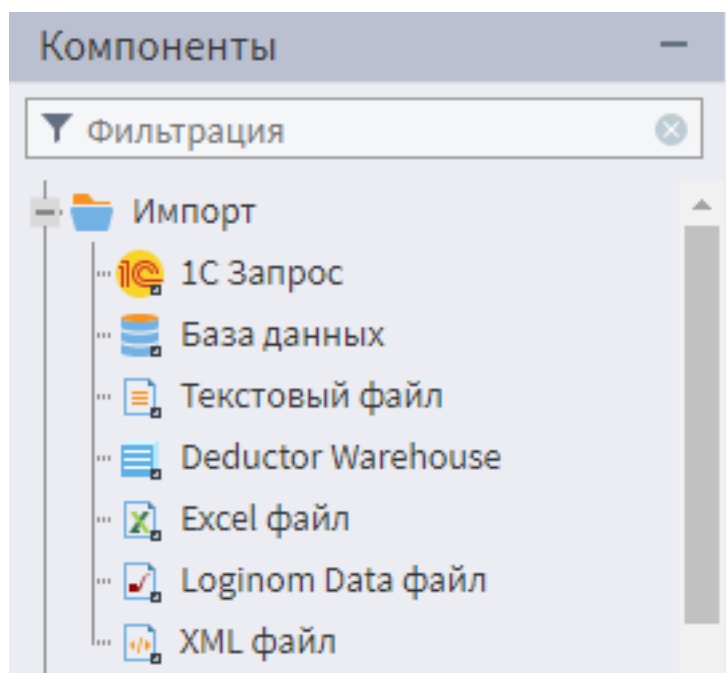


Рисунок 5 – Компоненты «Импорт»

Платформа Loginom предоставляет список из нескольких возможных компонентов, с помощью которых производится импорт данных. Например, данные могут быть предоставлены из текстового файла, XML файла или Excel файл (смотри рисунок 4) [11].

Далее идут компоненты «Трансформация» (рисунок 6).

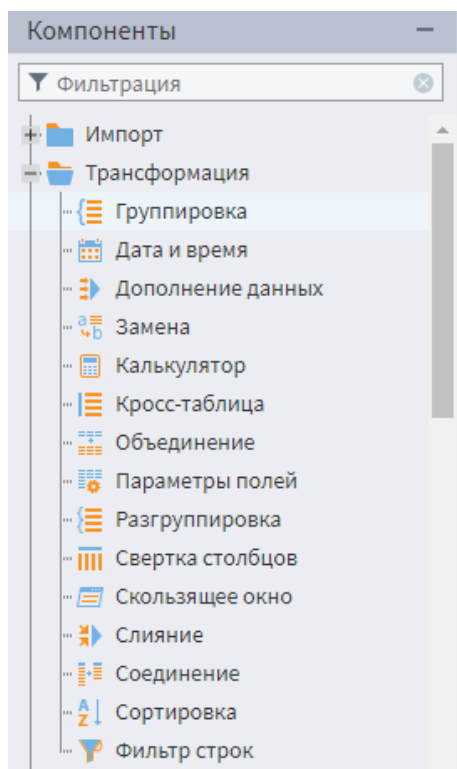


Рисунок 6 – Компоненты «Трансформация»

Группа компонентов для первоначальной подготовки и простой обработки исходных наборов данных. С помощью этих компонентов пользователь может сортировать, группировать, сливать и т. д. импортированные данные [12].

Блок компонентов «Управление» нужен, например, для добавления условия или цикла выполнения узла сценария (рисунок 7).

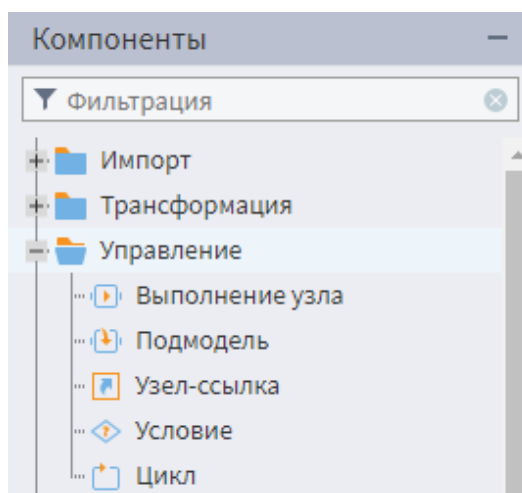


Рисунок 7 – Компоненты «Управление»

Компоненты предназначены для оптимизации сценариев с помощью создания подмоделей и повторного использования узлов, а также формирования логики выполнения сценариев при помощи условий и циклов [13].

Компоненты для корреляционного и факторного анализов, а также автокорреляция и поиск дубликатов и противоречий находятся в разделе «Исследование» (рисунок 8).

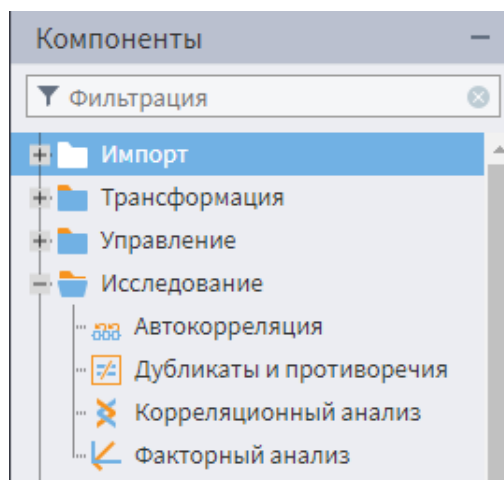


Рисунок 8 – Компоненты «Исследование»

Методы для статистического анализа, используемые для выполнения задач данной работы будут взяты из этого раздела [14].

Компоненты «Предобработка» (рисунок 9).

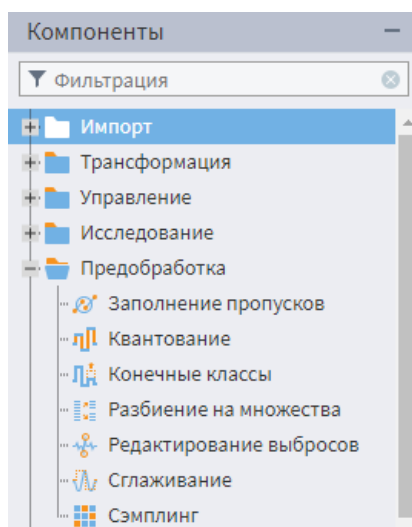


Рисунок 9 – Компоненты «Предобработка»

Данные компоненты нужны для предварительной обработки данных для их дальнейшего использования в алгоритмах Data Mining. Применяются такие методы, как заполнение пропусков, «сэмплинг», редактирование выбросов и другие [15].

Блок компонентов «Data Mining» (рисунок 10).

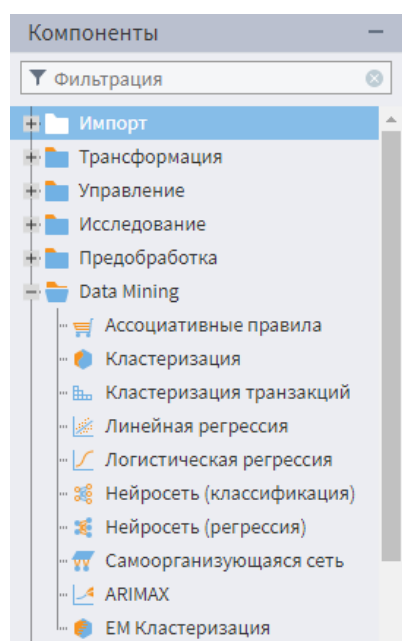


Рисунок 10 – Компоненты «Data Mining»

Компоненты, входящие в эту группу, являются инструментами для реализации различных методов Data Mining: кластеризация, ассоциативные правила и другие [16].

Блок компонентов «Программирование» (рисунок 11).

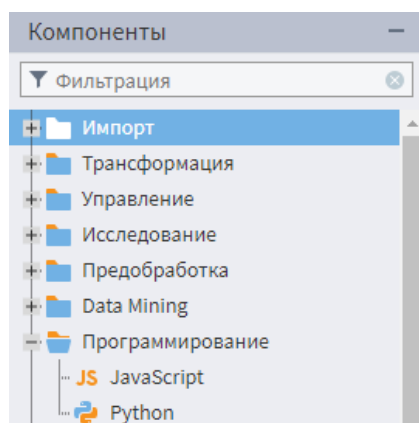


Рисунок 11 – Компоненты «Программирование»

Компоненты этой группы используются для обработки данных с помощью языков программирования JavaScript и Python [17].

В Loginom есть компоненты, предназначенные для работы с переменными. Они выделены в отдельную группу на панели Компоненты. Блок компонентов «Переменные» (рисунок 12).

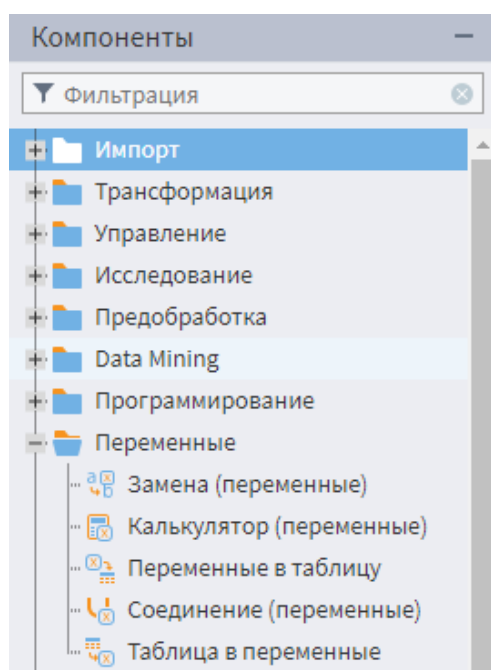


Рисунок 12 – Компоненты «Переменные»

Значения переменных могут быть использованы в вычислениях, например, в узлах условие, калькулятор и других. Так же значения переменных могут использоваться в качестве управляющих переменных для задания параметров настройки узлов [18].

Блок компонентов «Дерево данных». Дерево представляет собой структуру данных, состоящую из узлов, которые связаны между собой. Каждый узел может быть родительским или дочерним для другого узла. Корневой узел является первым и основным элементом дерева. Узлы могут быть контейнерами, то есть содержать другие узлы, либо массивами, когда в рамках одного родительского узла может содержаться несколько экземпляров одного и того же узла. Древовидные данные используются в

программировании и могут передаваться через порты или пакеты (рисунок 13) [19].

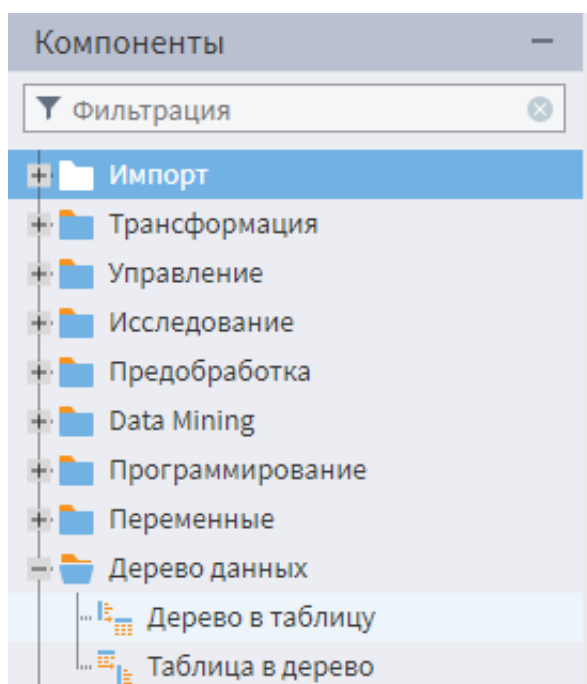


Рисунок 13 – Компоненты «Дерево данных»

Компоненты блока «Интеграция» позволяют осуществлять обмен данными между платформой Logiот и сторонними внешними системами (рисунок 14).

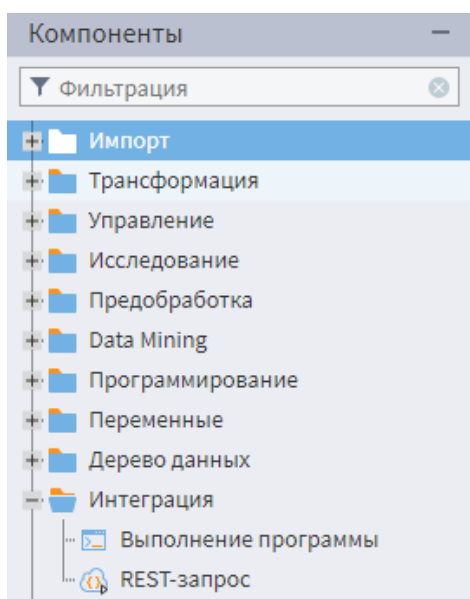


Рисунок 14 – Компоненты «Интеграция»

Механизмы интеграции предназначены для обмена данными между платформой Loginom и сторонними внешними системами. Loginom имеет возможность работать как с веб-сервисами, так и базами данных и файловыми хранилищами данных (электронные таблицы, текстовые файлы и т. д.) [20].

Компоненты блока «Экспорт» позволяют выгружать Результаты обработки и анализа данных во внешние источники (рисунок 15).

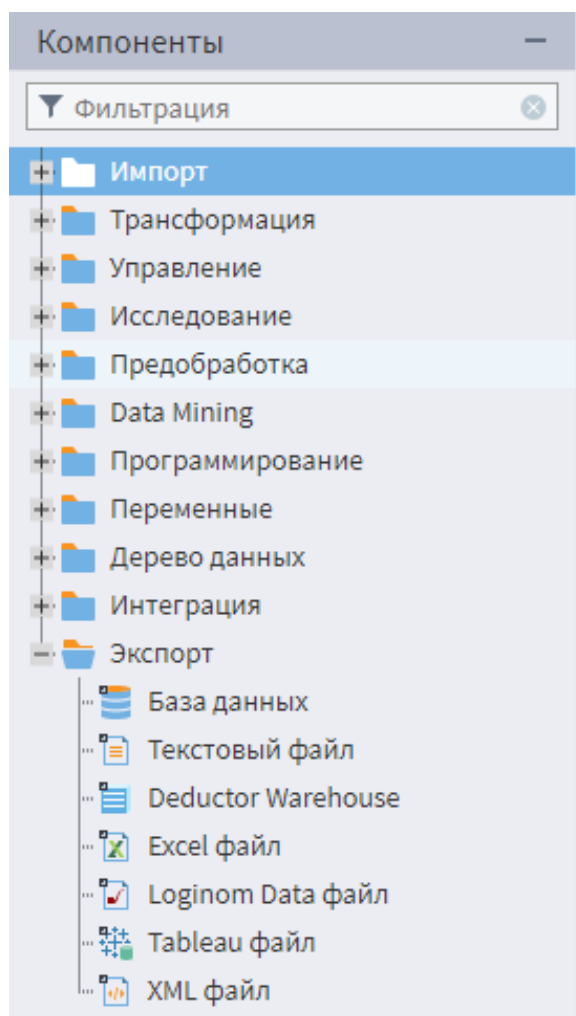


Рисунок 15 – Компоненты «Экспорт»

В Loginom можно экспортировать данные в текстовый файл, XML-файл, электронные таблицы, внутренний формат программы, а также в базы данных [21].

Из всего набора вышеперечисленных компонентов для выполнения поставленной задачи будут использоваться «Импорт» и «Исследования» для загрузки данных, корреляционного и факторного анализов.

4 Выполнение статистического анализа данных

Для выполнения статистического анализа данных была сформирована таблица, состоящая из официальных данных об Пензенской области, опубликованных в открытых источниках (таблица 4) [22 – 27].

Таблица 4 – Данные для проведения статистического анализа

Год	Население, человек	Денежный доход в расчете на душу населения, рублей	Потребительские расходы в расчете на душу населения, рублей	Потребление электроэнергии, млн кВт час	Величина прожиточного минимума на душу населения, рублей	Экспорт товаров, млн долларов США	Импорт товаров, млн долларов в США	ВРП, млн рублей
2013	1368657	16803,6	14169,1	4695,5	5440	217,4	280,5	270436,8
2014	1360587	18721,4	15832,9	4768,5	6571	189,9	223,7	295238,7
2015	1355618	21452,7	17024,7	4732,7	8096	278,8	171,1	343328,6
2016	1348703	20627,6	16776,3	4699	8428	204,8	157,2	358258,2
2017	1341526	21364,3	17526	4788,4	8690	247,3	256,3	376076,2
2018	1331655	21821,1	18254,1	4858,7	8841	298,6	272,5	411028,7
2019	1318103	22989,7	19504	4808,9	9337	308,4	252,4	448521,1
2020	1305563	24135	19482,3	4642,7	9512	403,9	301,5	484758,4
2021	1290898	26415,8	22508,3	4829,9	9610	358	359,4	537290
2022	1261103	27539,3	23306,5	4909,3	11692	303	305,3	560393,4

Таблица состоит из данных об количестве населения Пензенской области, денежных доходах в расчете на душу населения, потребительских расходах в расчете на душу населения, потреблении электроэнергии выраженную в миллионах киловатт-час, величине прожиточного минимума на душу населения выраженную в рублях, экспорте товаров в страны СНГ выраженную в миллионах долларов США, импорте товаров в страны СНГ выраженную в миллионах долларов США и валовом региональном продукте выраженном в миллионах рублей по годам за последние десять лет (приложение Б). С полученными данными будет проводиться статистический анализ в качестве корреляционного и факторного анализов.

4.1 Импорт данных

Прежде чем выполнять корреляционный анализ, данные, в качестве таблицы, надо импортировать в сценарий платформы Loginom с помощью компонента «Excel файл» (рисунок 16).

Импорт из Excel файла

Хранилище файлов: Локальное Подключено

Имя файла: ../Данные по Пензенской обл..xlsx

Область данных

Выбор объекта: По номеру Стиль ссылок: ☐ R1C1 ☒ A1

Имя объекта: 1 Диапазон: A1:A1

Весь лист: ☒ До последней строки: ☒

Пустые строки: Импортировать Количество строк заголовка: 1

#	A	B	C	D	E	F	G	H	I
1	Год	Населени	Денежный	Потребит	Потребле	Величина	Экспорт т	Импорт то	Валовый
2	2 013	1 368 657	16 803,60	14 169,10	4 695,50	5 440	217,40	280,50	270 436,80
3	2 014	1 360 587	18 721,40	15 832,90	4 768,50	6 571	189,90	223,70	295 238,70
4	2 015	1 355 618	21 452,70	17 024,70	4 732,70	8 096	278,80	171,10	343 328,60
5	2 016	1 348 703	20 627,60	16 776,30	4 699	8 428	204,80	157,20	358 258,20
6	2 017	1 341 526	21 364,30	17 526	4 788,40	8 690	247,30	256,30	376 076,20
7	2 018	1 331 655	21 821,10	18 254,10	4 858,70	8 841	298,60	272,50	411 028,70
8	2 019	1 318 103	22 989,70	19 504	4 808,90	9 337	308,40	252,40	448 521,10
9	2 020	1 305 563	24 135	19 482,30	4 642,70	9 512	403,90	301,50	484 758,40
10	2 021	1 290 898	26 415,80	22 508,30	4 829,90	9 610	358	359,40	537 290
11	2 022	1 261 103	27 539,30	23 306,50	4 909,30	11 692	303	305,30	560 393,40

Рисунок 16 – Импорт таблицы

Выбираем объекты по номеру, значение «Имя объекта» записываем равное единице и устанавливаем галочку напротив «Весь лист».

Далее настраиваем поля, выставляем тип и вид данных или нажимаем кнопку «Определить типы данных», тогда они определяться автоматически (рисунок 17).

Настройка соответствия между столбцами

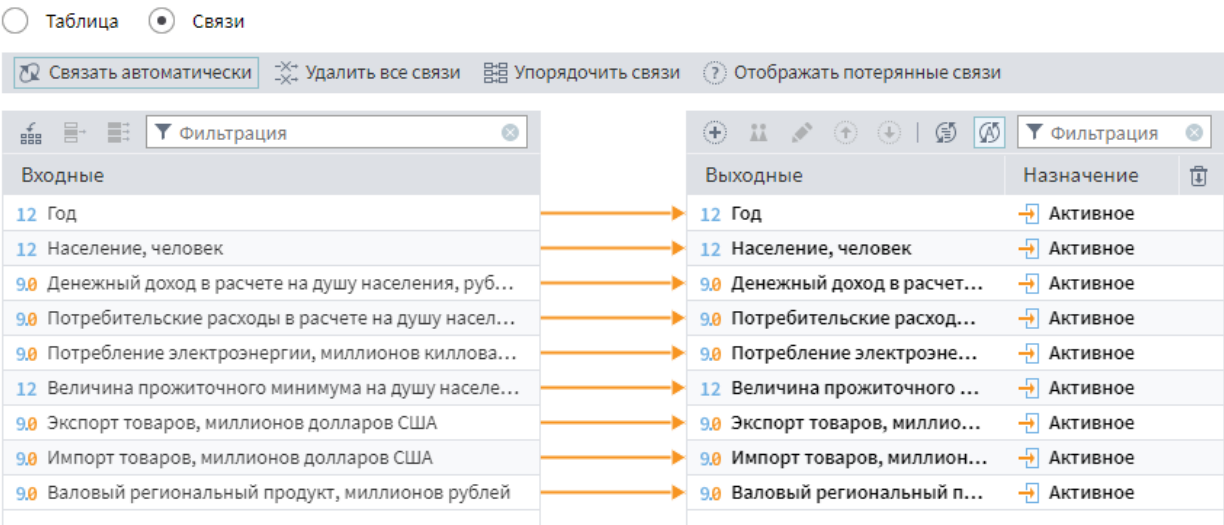


Рисунок 19 – Настройка связей

Определяем соответствия и завершаем настройку. Далее переходим к проведению корреляционного анализа.

4.2 Корреляционный анализ данных

Для проведения корреляционного анализа добавим одноименный компонент и приступим к его настройке (рисунок 20).

Корреляционный анализ

- ☒ Коэффициент корреляции Пирсона
- ☐ Экстремум взаимнокорреляционной функции
- ☐ Коэффициент Тау-В Кендалла
- ☐ Коэффициент корреляции Спирмена

Входные колонки	Набор 1	Набор 2
12 Население, человек	☐	☒
9.0 Денежный доход в расчете на душу населения, рублей	☐	☒
9.0 Потребительские расходы в расчете на душу населения, рублей	☐	☒
9.0 Потребление электроэнергии, миллионов киловатт-часов	☐	☒
12 Величина прожиточного минимума на душу населения, рублей	☐	☒
9.0 Экспорт товаров, миллионов долларов США	☐	☒
9.0 Импорт товаров, миллионов долларов США	☐	☒
9.0 Валовый региональный продукт, миллионов рублей	☒	☐
12 Год	☐	☐

Рисунок 20 – Настройка корреляционного анализа

Выбираем «Коэффициент корреляции Пирсона», значение первого набора данных выбираем «Валовый региональный продукт», вторым набором выбираем все остальные, кроме «Год». Таким образом проведем корреляционный анализ между данными о валовом внутреннем продукте и множестве данных по Пензенской области. Чтобы увидеть результат, открываем настройки визуализации и выбираем табличный вид. Результат анализа (рисунок 21).

#	ab Поле1.Метка	ab Поле2.Метка	9.0 Пирсона
1	Валовый региональный прод...	Население, человек	-0,9786123804
2	Валовый региональный прод...	Денежный доход в расчете на душу населе...	0,9803380567
3	Валовый региональный прод...	Потребительские расходы в расчете на ду...	0,9824351479
4	Валовый региональный прод...	Потребление электроэнергии, миллионов ...	0,5066644538
5	Валовый региональный прод...	Величина прожиточного минимума на ду...	0,927482232
6	Валовый региональный прод...	Экспорт товаров, миллионов долларов США	0,7951730313
7	Валовый региональный прод...	Импорт товаров, миллионов долларов США	0,6426138377

Рисунок 21 – Результат корреляционного анализа

В первом столбце оказывается первый набор данных, во втором – второй, соответственно, в третьем – коэффициент корреляции.

Все результаты были перепроверены и рассчитаны по формуле нахождения коэффициента корреляции Пирсона [29]:

$$r_{xy} = \frac{n \times \sum (x_i \times y_i) - (\sum x_i) \times (\sum y_i)}{\sqrt{[n \times \sum x_i^2 - (\sum x_i)^2] \times [n \times \sum y_i^2 - (\sum y_i)^2]}}$$

где r_{xy} – коэффициент корреляции;

n – количество пар наборов данных;

x_i – значения первого набора данных;

y_i – значения второго набора данных.

Интерпретируем полученный результат. Для наглядности результат представлен в виде диаграммы (рисунок 22).

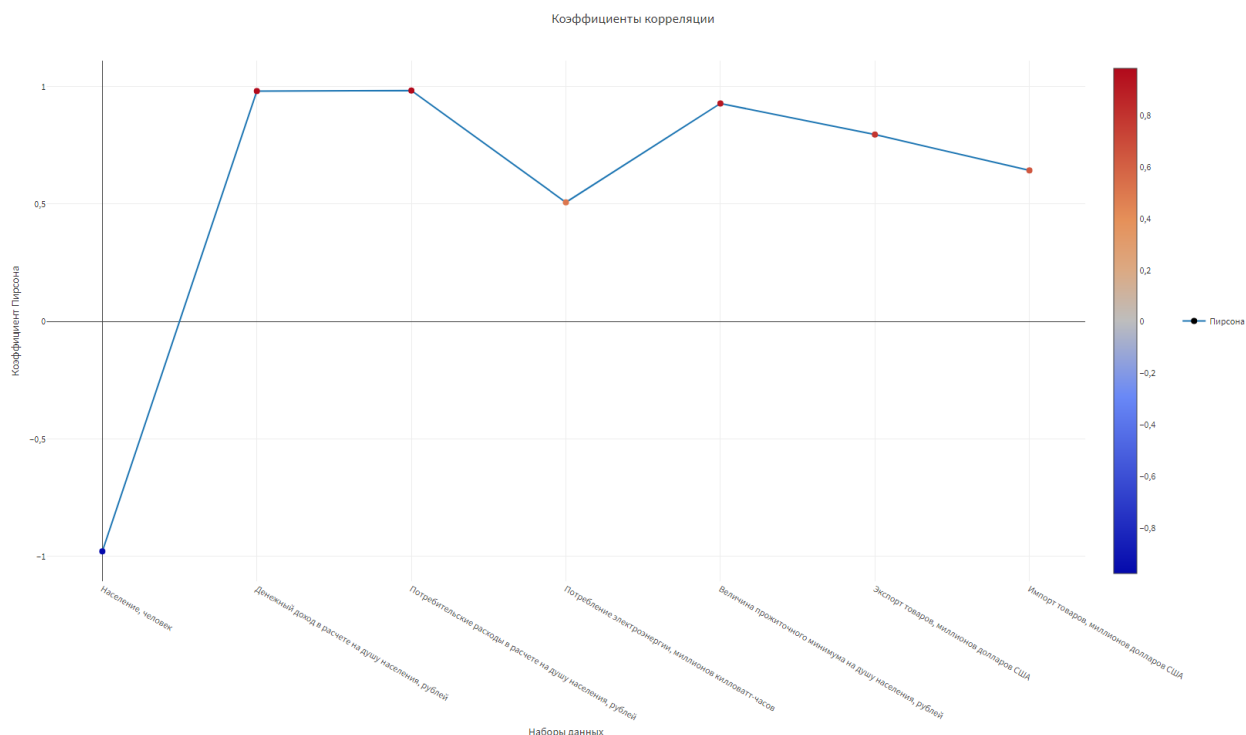


Рисунок 22 – Диаграмма коэффициентов корреляции

По результатам проведенного анализа данных, полученных с официального сайта статистики по пензенской области, можно сделать вывод, что значение внутреннего валового продукта сильно коррелируется с такими показателями, как денежный доход на душу населения, потребительский расход на душу населения и величина прожиточного минимума, значения коэффициента большего ноль целый пяти десятых принято считать значительным, по итогам результата все полученные коэффициенты именно такие. Важно отметить то, что значение коэффициента между ВРП и количеством населения отрицательно, но велико по модулю, это означает сильную обратно пропорциональную зависимость [30].

4.3 Факторный анализ данных

Для проведения факторного добавим одноименный компонент и приступим к его настройке. Настроим выходные данные (рисунок 23).

Настройка входных столбцов

Фильтрация			
Метка	Имя	Вид данных	Назначение
12 Год	God	Непрерывн...	Не задано
12 Население, человек	Naselenie__chelovek	Непрерывн...	Используемое
9.0 Денежный доход в расчете на душу ...	Denezhniy_dokhod_v_raschete_na_dus...	Непрерывн...	Используемое
9.0 Потребительские расходы в расчет...	Potrebitelskie_raskhody_v_raschete_na...	Непрерывн...	Используемое
9.0 Потребление электроэнергии, милл...	Potreblenie_elektroenergii__millionov_...	Непрерывн...	Используемое
12 Величина прожиточного минимума ...	Velichina_prozhitochnogo_minimuma_n...	Непрерывн...	Используемое
9.0 Экспорт товаров, миллионов долла...	Eksport_tovarov__millionov_dollarov_S...	Непрерывн...	Используемое
9.0 Импорт товаров, миллионов долла...	Import_tovarov__millionov_dollarov_SS...	Непрерывн...	Используемое
9.0 Валовый региональный продукт, м...	Valoviy_regionalniy_produkt__millionov...	Непрерывн...	Не задано

Рисунок 23 – Настройка входных данных

Назначение у всех используется, кроме «Год» и ВРП, так как все остальные являются его факторами.

Далее настроим критерии значимости и метод получения окончательного решения (рисунок 24).

Факторный анализ

Критерий значимости факторов

☐ По собственному значению

Порог собственного значения

0,5

☐ По дисперсии

Порог дисперсии, %

90

☒ Задать число факторов

Число факторов

3

Метод получения окончательного решения

Метод вращения

☒ Без вращения
 ☐ Варимакс
 ☐ Квартимакс

☐ Ограничить число выходных факторов

Число факторов

10

Рисунок 24 – Настройка факторного анализа

Критерий значимости факторов ставим «Задать число факторов», число факторов три. Метод получения окончательного решения – без вращения. Метод Варимакс – ортогональный метод вращения, минимизирующий число переменных с высокими нагрузками на каждый фактор. Этот метод упрощает

интерпретацию факторов. Метод Квартимакс – метод вращения, который минимизирует число факторов, необходимых для объяснения каждой переменной. Этот метод упрощает интерпретацию наблюдаемых переменных [28]. Модель факторного анализа:

$$y_i = a_0 + a_1 \times y_{i-1} + a_2 \times y_{i-2} + a_3 \times y_{i-3},$$

где y_i – фактор;

a – коэффициент.

В качестве результата выполненного факторного анализа имеем две таблицы. Первая таблица факторов (рисунок 25).

#	9.0 Фактор1	9.0 Фактор2	9.0 Фактор3	12 Населен...	9.0 Денеж...	9.0 Потреб...	9.0 Потребл...	12 Велич...1	9.0 Экс...	9.0 Импорт...	12 Год
1	-1,430580558	-0,732956606	1,652362376	1368657	16803,6	14169,1	4695,5	5440	217,4	280,5	2013
2	-1,10735488	0,6960263691	0,6058389679	1360587	18721,4	15832,9	4768,5	6571	189,9	223,7	2014
3	-0,6258315953	0,01626876276	-1,358434553	1355618	21452,7	17024,7	4732,7	8096	278,8	171,1	2015
4	-0,8705237372	0,475576977	-1,694349665	1348703	20627,6	16776,3	4699	8428	204,8	157,2	2016
5	-0,2448596424	0,4381422947	0,2104882446	1341526	21364,3	17526	4788,4	8690	247,3	256,3	2017
6	0,1697662058	0,5873153338	0,6813550481	1331655	21821,1	18254,1	4858,7	8841	298,6	272,5	2018
7	0,366411515	0,2215797728	-0,265433897	1318103	22989,7	19504	4808,9	9337	308,4	252,4	2019
8	0,6119998025	-2,49669909	-0,7396647113	1305563	24135	19482,3	4642,7	9512	403,9	301,5	2020
9	1,367175967	-0,5016710902	1,02326017	1290898	26415,8	22508,3	4829,9	9610	358	359,4	2021
10	1,763796923	1,296417331	-0,1154224873	1261103	27539,3	23306,5	4909,3	11692	303	305,3	2022

Рисунок 25 – Таблица факторов

В ней прописаны значения трех факторов за каждый год. Значения факторных нагрузок представлены во второй таблице (рисунок 26).

#	ab Метка	9.0 Фактор1	9.0 Фактор2	9.0 Фактор3
1	Население, человек	-0,9797942139	-0,009351701021	0,02660586507
2	Денежный доход в расчете на душу населения, рублей	0,9761295215	0,01087093981	-0,1849344056
3	Потребительские расходы в расчете на душу населения, рублей	0,9871965358	0,07001537011	-0,06161635114
4	Потребление электроэнергии, миллионов киловатт-часов	0,600096269	0,7135485622	0,3082272432
5	Величина прожиточного минимума на душу населения, рублей	0,917852351	0,1466571734	-0,3375639074
6	Экспорт товаров, миллионов долларов США	0,7729475824	-0,5693044015	-0,06987217981
7	Импорт товаров, миллионов долларов США	0,6810848279	-0,3107661637	0,6552649178

Рисунок 26 – Таблица факторных нагрузок

Факторный анализа позволили снизить количество влияющих факторов с восьми до трех, что является хорошим результатом [30]. Интерпретируем результаты. Для наглядности была визуализирована диаграмма (рисунок 27).

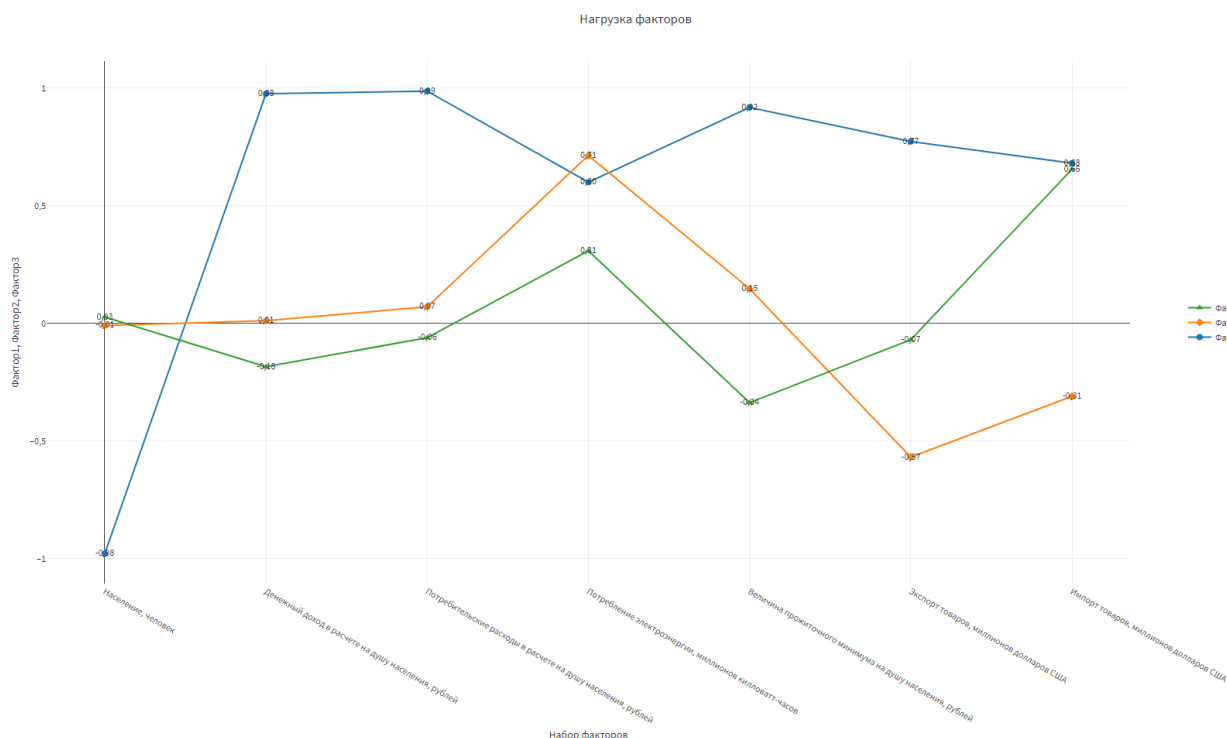


Рисунок 27 – Диаграмма нагрузок факторов

Трехфакторный анализ позволил выявить наиболее важные факторы, ими являются значения большие ноль целых пяти десятых [31-37], таких несколько, все они коррелируются между собой, а значит могут быть взяты в качестве одного обобщенного, ранее не выявленного фактора. Например, большую корреляция по первому фактору имеют «Денежный доход в расчете на душу населения» и «Потребительский расход в расчете на душу населения», два этих параметра сильно коррелируются между собой и могут быть объединены в один, ранее не выявленный.

4.4 Выводы проведенного анализа

Выполнив статистический анализ данных по Пензенской области инструментальными средствами платформы LogiTom, можно сделать некоторые выводы. Например, при отрицательном демографическом росте, то есть при депопуляции, показатели экономической и потребительской сфер растут с каждым годом, с каждым годом все увеличивается ВРП, величина прожиточного минимума, доходы и расходы населения. Некоторые показатели явно связаны

между собой, то есть имеют выявленную корреляцию. Выполнив факторный анализ между наборами данных по нескольким показателям, пришли к выводу, что многие из них связаны между собой и влияют на изучаемый ВРГ как три общих факторов.

Заключение

В результате выполнения выпускной квалификационной работы, был проведен статистический анализ данных инструментальными средствами Loginom.

Была поставленная задача и проведен анализ требований работы. Изучены статистические методы анализа данных и выбраны более подходящие для выполнения поставленной задачи. Составлена база данных основанная на реальной статистике Пензенской области за последние десять лет. Изучена платформа анализа данных Loginom, а также ее компоненты, с помощью которых был произведен статистический анализ.

Аналитическая платформа Loginom отлично справилась с поставленной задачей выполнения статистического анализа данных. Благодаря инструментальным средствам предоставленным платформой, получилось осуществить корреляционный и факторный анализы, были получены, визуализированы, а также интерпретированы результаты.

Итогами проведенной работы стали выводы, сделанные на основе проделанного анализа статистических данных Пензенской области.

Список использованных источников

1. Статья «Статистический анализ» [Электронный ресурс]. URL: <https://datascience.eu/ru/> (дата обращения 18.04.2023).
2. Статья «7 методов статистического анализа, которые может применять каждый» [Электронный ресурс]. URL: <https://4brain.ru/blog/> (дата обращения 18.04.2023).
3. Статья «Коэффициент корреляции» [Электронный ресурс] URL: <http://statistica.ru/theory/koeffitsient-korrelyatsii/> (дата обращения 18.04.2023).
4. Статья «Факторный анализ» [Электронный ресурс]. URL: <https://www.ibm.com/docs/ru/spss-statistics/25.0.0?topic=features-factor-analysis> (дата обращения 18.04.2023).
5. Статья «Методы описательной статистики» [Электронный ресурс]. URL: <https://sibac.info/blog/metody-opisatelnoy-statistiki> (дата обращения 18.04.2023).
6. Статья «Основы регрессионного анализа» [Электронный ресурс]. URL: <https://pro.arcgis.com/ru/pro-app/2.9/tool-reference/spatial-statistics/regression-analysis-basics.htm> (дата обращения 18.04.2023).
7. Статья «Дисперсионный анализ» [Электронный ресурс]. URL: <http://www.iki.rssi.ru/magbase/REFMAN/STATTEXT/modules/stanman.html> (дата обращения 18.04.2023).
8. Статья «Что такое тестирование гипотез и как его проводить?» [Электронный ресурс]. URL: <https://lpgenerator.ru/blog/2018/12/24/chto-takoe-testirovanie-gipotez-i-kak-ego-provodit/> (дата обращения 18.04.2023).
9. Статья «Анализ временных рядов» [Электронный ресурс]. URL: <http://www.iki.rssi.ru/magbase/REFMAN/STATTEXT/modules/sttimser.html> (дата обращения 18.04.2023).

10. Статья «Как работает метод главных компонент на простом примере» [Электронный ресурс]. URL: <https://habr.com/ru/articles/304214/> (дата обращения 18.04.2023).

11. Документация Loginom Help // Импорт [Электронный ресурс]. URL: <https://help.loginom.ru/userguide/integration/import/index.html> (дата обращения 20.04.2023).

12. Документация Loginom Help // Трансформации [Электронный ресурс]. URL: <https://help.loginom.ru/userguide/processors/transformation/index.html> (дата обращения 20.04.2023).

13. Документация Loginom Help // Управления [Электронный ресурс]. URL: <https://help.loginom.ru/userguide/processors/control/index.html> (дата обращения 20.04.2023).

14. Документация Loginom Help // Исследования [Электронный ресурс]. URL: <https://help.loginom.ru/userguide/processors/scrutiny/index.html> (дата обращения 20.04.2023).

15. Документация Loginom Help // Предобработка [Электронный ресурс]. URL: <https://help.loginom.ru/userguide/processors/preprocessing/index.html> (дата обращения 20.04.2023).

16. Документация Loginom Help // Data Mining [Электронный ресурс]. URL: <https://help.loginom.ru/userguide/processors/datamining/index.html> (дата обращения 20.04.2023).

17. Документация Loginom Help // Программирование [Электронный ресурс]. URL: <https://help.loginom.ru/userguide/processors/programming/index.html> (дата обращения 20.04.2023).

18. Документация Loginom Help // Переменные [Электронный ресурс]. URL: <https://help.loginom.ru/userguide/workflow/variables/index.html> (дата обращения 20.04.2023).

19. Документация Loginom Help // Дерево данных [Электронный ресурс]. URL: <https://help.loginom.ru/userguide/processors/data-trees/index.html> (дата обращения 20.04.2023).

20. Документация Loginom Help // Интерпретации [Электронный ресурс]. URL: <https://help.loginom.ru/userguide/integration/index.html> (дата обращения 20.04.2023).

21. Документация Loginom Help // Экспорт [Электронный ресурс]. URL: <https://help.loginom.ru/userguide/integration/export/index.html> (дата обращения 20.04.2023).

22. Территориальный орган Федеральной службы государственной статистики по Пензенской области // Валовый региональный продукт [Электронный ресурс]. URL: https://58.rosstat.gov.ru/Gross_Regional_Product_1 (дата обращения 21.04.2023).

23. Территориальный орган Федеральной службы государственной статистики по Пензенской области // Промышленное производство [Электронный ресурс]. URL: https://58.rosstat.gov.ru/industrial_production (дата обращения 21.04.2023).

24. Территориальный орган Федеральной службы государственной статистики по Пензенской области // Неравенство и бедность [Электронный ресурс]. URL: https://58.rosstat.gov.ru/inequality_poverty (дата обращения 21.04.2023).

25. Территориальный орган Федеральной службы государственной статистики по Пензенской области // Экспорт и импорт [Электронный ресурс]. URL: https://58.rosstat.gov.ru/international_trade (дата обращения 21.04.2023).

26. Территориальный орган Федеральной службы государственной статистики по Пензенской области // Демография [Электронный ресурс]. URL: <https://58.rosstat.gov.ru/demography> (дата обращения 21.04.2023).

27. Территориальный орган Федеральной службы государственной статистики по Пензенской области // Денежные доходы и расходы населения

[Электронный ресурс]. URL: https://58.rosstat.gov.ru/living_standard (дата обращения 21.04.2023).

28. Статья «Вращение факторов для факторного анализа» [Электронный ресурс]. URL: https://www.ibm.com/docs/ru/spss-statistics/25.0.0?topic=SSLVMB_25.0.0/spss/base/idh_fact_rot.htm#idh_fact_rot (дата обращения 21.04.2023).

29. Син Такхаси, Занимательная статистика. Факторный анализ. – М.: ДМК Пресс, 2015. – 248 с.

30. Жукова, Г.С. Математический анализ: учебник. Том 1 / М.Ф. Рушайло; под ред. Г.С. Жуковой. – Москва: ИНФРА-М, 2023. – 388 с.

31. Жукова, Г.С. Математический анализ: учебник. Том 2 / М.Ф. Рушайло; под ред. Г.С. Жуковой. – Москва: ИНФРА-М, 2020. – 518 с. – 015. – 248 с.

32. Гитис, Л.Х. Факторный анализ; Лекция 1. Постановка задачи и области применения. – М.: Издательство Московского государственного горного университета, 2000. – 24 с.

33. Документация IBM // Факторный анализ [Электронный ресурс]. URL: <https://www.ibm.com/docs/ru/spss-statistics/25.0.0?topic=features-factor-analysis> (дата обращения 21.06.2023).

34. Документация IBM // Описательные статистики факторного анализа [Электронный ресурс]. URL: <https://www.ibm.com/docs/ru/spss-statistics/25.0.0?topic=analysis-factor-descriptives> (дата обращения 21.06.2023).

35. Документация IBM // Значение факторов в процедуре факторного анализа [Электронный ресурс]. URL: <https://www.ibm.com/docs/ru/spss-statistics/25.0.0?topic=analysis-factor-scores> (дата обращения 21.06.2023).

36. Документация IBM // Выделение факторов в процедуре Факторный анализ [Электронный ресурс]. URL: <https://www.ibm.com/docs/ru/spss-statistics/25.0.0?topic=analysis-factor-extraction> (дата обращения 21.06.2023).

37. Документация IBM // Параметры процедуры Факторный анализ [Электронный ресурс]. URL: <https://www.ibm.com/docs/ru/spss-statistics/25.0.0?topic=analysis-factor-options> (дата обращения 21.06.2023).

Приложение А
ГЛОССАРИЙ
(справочное)

База данных (БД) – специально организованная и структурированная коллекция данных, где каждая порция информации сохранена в отдельном месте и может быть легко найдена и извлечена при необходимости.

Диаграмма вариантов использования (ДВИ) – это графическое представление взаимодействия между системой и ее окружением, которое помогает описать, как пользователи будут взаимодействовать с системой и как система будет им на это отвечать.

Диаграмма Ганта (ДГ) – это инструмент визуализации, который используют для планирования проектов. Он состоит из горизонтальных полос, которые представляют задачи, расположенных вдоль временной шкалы. График помогает проследить, какие задачи выполняются и когда, а также определить пути критического планирования и выявить возможные задержки в процессе работы.

Dashboard – это инструмент, который представляет информацию в виде графиков, таблиц и других элементов управления на одном экране для того, чтобы пользователь мог быстро и легко получить доступ к нужной информации и принимать решения.

Приложение Б
ТАБЛИЦЫ ДАННЫХ ПО ПЕНЗЕНСКОЙ ОБЛАСТИ
(рекомендованное)

Таблица Б.1 – Численность постоянного населения Пензенской области

Год	Все население, человек	В том числе	
		городское население, человек	сельское население, человек
2010	1373236	915930	457306
2011	1384006	929976	454030
2012	1376538	930155	446383
2013	1368657	927210	441447
2014	1360587	924882	43705
2015	1355618	923241	432377
2016	1348703	920885	427818
2017	1341526	916593	424633
2018	1331655	912181	419474
2019	1318103	905388	412715
2020	1305563	89910	405653
2021	1290898	892116	398782
2022	1261103	866992	394111

Таблица Б.2 – Среднедушевные денежные доходы и потребительские расходы

Год	Денежные доходы в расчете на душу населения, рублей	в % к аналогичному периоду предыдущего года	Потребительские расходы в расчете на душу населения, рублей	в % к аналогичному периоду предыдущего года
2013	16803,6	х	14169,1	х

Продолжение таблицы Б.2

Год	Денежные доходы в расчете на душу населения, рублей	в % к аналогичному периоду предыдущего года	Потребительские расходы в расчете на душу населения, рублей	в % к аналогичному периоду предыдущего года
2014	18721,4	111,4	15832,9	111,7
2015	21452,7	114,6	17024,7	107,5
2016	20627,6	96,2	16776,3	98,5
2017	21364,3	103,6	17526,0	104,5
2018	21821,1	102,1	18254,1	104,2
2019	22989,7	105,4	19504,0	106,8
2020	24135,0	105,0	19482,3	99,9
2021	26415,8	109,5	22508,3	115,5
2022	27519,3	112,3	23869,6	118,8

Таблица Б.3 – Потребление электроэнергии

Год	Потребление электроэнергии, миллионов киловатт-часов
2009	4259,0
2010	4357,9
2011	4460,7
2012	4610,1
2013	4695,5

Продолжение таблицы Б.3

2014	4768,5
2015	4732,7
2016	4699,0
2017	4788,4
2018	4858,7
2019	4808,9
2020	4642,7
2021	4829,9
2022	4909,3

Таблица Б.4 – Величина прожиточного минимума на душу населения

Год	Величина прожиточного минимума на душу населения, рублей	в том числе по социально-демографическим группам населения		
		трудоспособн ое население, рублей	пенсионеры, рублей	дети, рублей
2013	5 440	-	-	-
2014	6571	7083	5442	6517
2015	8096	8723	6677	8275
2016	8428	9069	6936	8773
2017	8690	9345	7172	9027
2018	8841	9527	7328	9238
2019	9337	10117	7758	9628
2020	9512	10288	7889	9877
2021	9610	10396	7953	10002

Продолжение таблицы Б.4

Год	Величина прожиточного минимума на душу населения, рублей	в том числе по социально-демографическим группам населения		
		трудоспособн ое население, рублей	пенсионеры, рублей	дети, рублей
2022	11692	12745	10319	10319

Таблица Б.5 – Экспорт и импорт товаров

Год	Экспорт	Дальнее зарубежье	Страны СНГ	Импорт	Дальнее зарубежье	Страны СНГ
2009	137,1	46,5	90,6	101,4	88,2	13,2
2010	115,6	54,9	60,7	145,3	134,8	10,5
2011	149,9	83,8	66,1	183,0	168,0	15,0
2012	307,0	112,0	195,0	232,7	186,0	46,7
2013	217,4	71,0	146,4	280,5	226,1	54,4
2014	189,9	72,7	117,2	223,7	178,4	45,3
2015	278,8	186,1	92,7	171,1	142,6	28,5
2016	204,8	84,0	120,8	157,2	130,8	26,4
2017	247,3	114,8	132,5	256,3	225,1	31,2
2018	298,6	151,3	147,3	272,5	233,6	38,9
2019	308,4	146,8	161,6	252,4	204,5	47,8
2020	403,9	236,9	167,0	301,5	256,2	45,3
2021	358,0	172,3	185,7	359,4	294,2	65,2
2022	303,1	-	-	305,3	-	-

Таблица Б.6 – Валовой региональный продукт

Год	Валовой региональный продукт (в текущих основных ценах), млн. руб.	Индекс физического объема валового регионального продукта (в постоянных ценах; в процентах к предыдущему году)	Валовой региональный продукт на душу населения, руб.	Региональная структура валового регионального продукта, %
2009	147185,1	95,3	105487,4	0,5
2010	172166,7	101,8	124020,7	0,5
2011	213401,2	107,9	154608,1	0,5
2012	239962,5	106,7	174823,6	0,5
2013	270436,8	104,6	198177,1	0,5
2014	295238,7	103,5	217390,7	0,5
2015	343328,6	103,9	253911,0	0,5
2016	358258,2	97,0	266340,2	0,5
2017	376076,2	102,1	281369,7	0,5
2018	411028,7	103,2	310238,7	0,5
2019	448521,1	103,1	341904,1	0,5
2020	484758,4	103,1	373420,3	0,5
2021	537290,0	101,8	418946,1	0,4
2022	560393,4	-	-	-