**АВТОНОМНАЯ НЕКОММЕРЧЕСКАЯ ОРГАНИЗАЦИЯ ПРОФЕССИОНАЛЬНАЯ ОБРАЗОВАТЕЛЬНАЯ ОРГАНИЗАЦИЯ МОСКОВСКИЙ МЕЖДУНАРОДНЫЙ КОЛЛЕДЖ ЦИФРОВЫХ ТЕХНОЛОГИЙ**

**«АКАДЕМИЯ ШАГ»**

**ИНДИВИДУАЛЬНЫЙ ПРОЕКТ**

Уровень профессионального образования:

Среднее профессиональное образование

Программа подготовки специалистов среднего звена

по специальности:

09.02.07 Информационные системы и программирование

Квалификация: Программист

Учебный предмет: Математика

Тема: Математический анализ в IT-проекте

Преподаватель:

Кульпина Татьяна Александровна

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

подпись инициалы фамилия

Обучающийся:

Зотеев Владислав Александрович

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

подпись дата инициалы фамилия

**Москва 2023**

Оглавление

[1. Введение 3](#_Toc134128057)

[1.1 Определение темы и ее актуальность 3](#_Toc134128058)

[1.2 Цели и задачи проекта 4](#_Toc134128059)

[2. Основные понятия и определения 5](#_Toc134128060)

[2.1 Математический методы анализа 5](#_Toc134128061)

[2.2 Инструменты математического анализа в IT 7](#_Toc134128062)

[2.3 Применение математического анализа в IT 8](#_Toc134128063)

[3. Анализ и формулировка проблемы 11](#_Toc134128064)

[3.1 Описание конкретного IT-проекта 11](#_Toc134128065)

[3.2 Выявление проблем и потребностей, связанных с математическим анализом 11](#_Toc134128066)

[3.3 Оценка исходных данных 12](#_Toc134128067)

[4. Подготовка данных 14](#_Toc134128068)

[5. Какие показатели оцениваем 14](#_Toc134128069)

[6. Анализ данных 15](#_Toc134128070)

[7. Построение модульной системы 16](#_Toc134128071)

[8. Вывод 16](#_Toc134128072)

[9. Список источников, использованных для подготовки проекта 17](#_Toc134128073)

[10. Приложение 19](#_Toc134128074)

# Введение

## Определение темы и ее актуальность

Тема: "Математический анализ в IT-проекте". Определение темы: Исследование применения методов и инструментов математического анализа для решения задач в рамках разработки и оптимизации IT-проектов, а также оценка их эффективности и влияния на результаты проекта.

Актуальность темы: Математический анализ является одним из ключевых инструментов в области информационных технологий. Он позволяет анализировать и оптимизировать сложные системы, предсказывать их поведение, а также принимать обоснованные решения на основе аналитических моделей и алгоритмов.

С развитием технологий, в частности машинного обучения, искусственного интеллекта и больших данных, математический анализ становится еще более актуальным для IT-проектов. Он помогает разработчикам и аналитикам эффективно обрабатывать огромные объемы информации, выявлять закономерности и тенденции, а также создавать новые алгоритмы и технологии, которые могут значительно повысить качество и эффективность IT-проектов.

Актуальность темы также связана с тем, что современные IT-проекты становятся все более масштабными и сложными, а значит, требуют более глубокого и системного подхода к анализу и оптимизации их работы. Математический анализ предоставляет необходимые инструменты для этого и позволяет существенно сократить время разработки и улучшить качество продуктов и услуг.

Таким образом, изучение и применение методов математического анализа в IT-проектах является актуальной и важной задачей для специалистов в области информационных технологий. Особенно это касается разработчиков, аналитиков и архитекторов систем, которые должны быть знакомы с различными аспектами математического анализа и его применения для решения сложных проблем и создания инновационных продуктов.

## Цели и задачи проекта

Цель проекта:

Использовать математический анализ для оптимизации процессов и повышения эффективности, автоматизировании анализа данных продаж магазина.

В проекте по разработке системы аналитики данных для магазинов задачи могут быть следующими:

1. Разработка системы аналитики данных продаж. Эта задача включает в себя сбор и анализ данных.
2. Разработка модульной системы генерации отчета. Эта задача включает в себя создание инструментов, которые позволяют пользователю легко генерировать отчеты на основе данных, собранных системой аналитики продаж.

В целом, эти задачи помогут создать систему аналитики данных, которая позволит управлять магазином более эффективно, оптимизировать продажи и увеличить прибыль.

# Основные понятия и определения

## Математический методы анализа

В IT-проектах широко распространено использование методов математического анализа, включает использование различных математических понятий и методов для достижения поставленной цели и решения задач. В рамках проекта по разработке системы аналитики данных, могут быть использованы следующие методы математического анализа:

1. Статистический анализ: применение методов статистики для описания, анализа и интерпретации данных, сбора информации о продажах, клиентах и ассортименте товаров. Статистический анализ может включать описательную статистику, корреляционный анализ, регрессионный анализ и дисперсионный анализ.
2. Машинное обучение: использование алгоритмов и моделей для обработки и анализа больших объемов данных, выявление закономерностей и построение прогнозных моделей. В рамках проекта машинное обучение может быть использовано для разработки модели предсказания продаж на основе статистических данных.
3. Оптимизация: применение математических методов для определения наилучших решений или стратегий, основанных на заданных критериях и ограничениях. В контексте проекта оптимизация может быть использована для улучшения управления запасами, определения оптимального ассортимента товаров и разработки стратегии продаж.
4. Графический анализ: использование графических методов и инструментов для визуализации данных и анализа результатов. В рамках проекта графический анализ может быть использован для разработки модульной системы генерации отчетов, позволяющей наглядно представить информацию о продажах, клиентах и товарах.
5. Вероятностный анализ: применение теории вероятности для оценки рисков, неопределенности и вероятности наступления различных событий. В контексте проекта вероятностный анализ может быть использован для определения вероятности успеха различных маркетинговых кампаний, оценки рисков, связанных с изменениями на рынке, и прогнозирования продаж с учетом возможных сценариев развития событий.
6. Линейная алгебра: применение понятий и методов линейной алгебры, таких как векторы, матрицы и системы линейных уравнений, для обработки и анализа данных. В контексте проекта линейная алгебра может быть использована для оптимизации распределения ресурсов, таких как рекламный бюджет, или решения задач, связанных с управлением запасами.
7. Численные методы: использование численных алгоритмов для приближенного решения математических задач, которые сложно или невозможно решить аналитически. В рамках проекта численные методы могут быть использованы для решения задач оптимизации, прогнозирования продаж и анализа данных, особенно при работе с большими объемами информации.
8. Анализ временных рядов: применение методов анализа данных, представленных в виде последовательности значений, измеренных через равные временные интервалы. В контексте проекта анализ временных рядов может быть использован для изучения динамики продаж, определения сезонности и выявления трендов, что поможет в прогнозировании будущих продаж.

Используя вышеуказанные понятия и методы математического анализа, IT-проекта сможет эффективно решать поставленные задачи.

## Инструменты математического анализа в IT

В рамках IT-проекта для разработки системы аналитики данных магазинов, могут быть использованы различные инструменты математического анализа. Ниже приведены некоторые из них, которые могут быть применены для решения поставленных задач:

1. Для реализации математического анализа данных могут быть использованы различные языки программирования такие как: R, Python (с библиотеками, такими как pandas, NumPy и SciPy), MATLAB и другие, которые позволяют проводить статистический анализ данных, строить графики и моделировать.
2. Алгоритмы машинного обучения: для разработки модели предсказания продаж на основе статистических данных можно использовать алгоритмы машинного обучения, такие как линейная регрессия, деревья решений, случайный лес, градиентный бустинг, нейронные сети и другие. Библиотеки, такие как scikit-learn (Python) и caret (R), предоставляют множество алгоритмов и функций для этой цели.
3. Визуализация данных: для создания модульной системы генерации отчетов, инструменты визуализации данных, такие как Matplotlib, Seaborn (Python), ggplot2 (R) или Tableau, могут быть использованы для создания наглядных графиков и диаграмм, представляющих информацию о продажах, клиентах и товарах.
4. Обработка больших данных: для анализа и обработки больших объемов данных, связанных с продажами, клиентами и товарами, можно использовать технологии и инструменты для обработки больших данных, такие как Apache Hadoop, Apache Spark, Dask (Python) или базы данных с поддержкой распределенных вычислений, такие как Google BigQuery или Amazon Redshift.
5. Анализ социальных сетей: для изучения влияния социальных медиа на продажи, определения наиболее влиятельных пользователей и маркетинговых каналов можно использовать методы анализа социальных сетей, такие как выявление сообществ, анализ центральности и другие.
6. Инструменты для сбора и хранения данных: для систематического сбора и хранения данных о продажах, клиентах и товарах, можно использовать инструменты и технологии, такие как веб-скрейпинг, API, базы данных SQL и NoSQL, облачные хранилища данных и другие.

Применение этих методов и инструментов математического анализа в IT позволит разработать систему аналитики данных для магазинов, с эффективной модульной системой генерации отчетов и предсказанием продаж на основе статистических данных. В результате магазин сможет повысить свою эффективность, увеличить прибыльность и улучшить удовлетворенность клиентов.

## Применение математического анализа в IT

* Нейронные сети: использование алгоритмов нейронных сетей для анализа временных рядов и прогнозирования продаж на основе исторических данных. Нейронные сети могут помочь выявить скрытые закономерности в данных и создать более точные прогнозы.
* Дифференциальные уравнения: использование дифференциальных уравнений для моделирования динамических процессов в IT-системах, таких как изменение трафика или нагрузки на сервера. Дифференциальные уравнения могут помочь определить оптимальные стратегии управления ресурсами и прогнозирования будущих изменений в системе.
* Теория графов: использование теории графов для анализа социальных сетей и связей между клиентами. Теория графов может помочь определить наиболее эффективные маркетинговые каналы и разработать алгоритмы оптимизации процессов в IT-системах.
* Теория информации: использование теории информации для определения оптимальных методов хранения, передачи и обработки данных. Теория информации может помочь оптимизировать процессы в IT-системах и улучшить эффективность сбора и анализа данных.
* Статистический анализ: использование методов статистики для анализа данных о продажах, клиентах и ассортименте товаров. Статистический анализ может помочь определить закономерности в данных и создать более точные прогнозы продаж.
* Машинное обучение: использование алгоритмов машинного обучения для анализа больших объемов данных и создания прогнозных моделей. Машинное обучение может помочь создать более точные прогнозы продаж и определить наиболее эффективные маркетинговые каналы.
* Оптимизация: использование математических методов оптимизации для определения наилучших решений или стратегий. Оптимизация может помочь улучшить управление запасами, определить оптимальный ассортимент товаров и разработать стратегию продаж.
* Графический анализ: использование графических методов и инструментов для визуализации данных и анализа результатов. Графический анализ может помочь наглядно представить информацию о продажах, клиентах и товарах и помочь принимать решения на основе данных.
* Вероятностный анализ: использование теории вероятности для оценки рисков и прогнозирования будущих продаж. Вероятностный анализ может помочь определить вероятность успеха маркетинговых кампаний и оценить риски, связанные с изменениями на рынке.
* Линейная алгебра: использование понятий и методов линейной алгебры для обработки и анализа данных. Линейная алгебра может помочь оптимизировать распределение ресурсов и решать задачи, связанные с управлением запасами.
* Численные методы: использование численных алгоритмов для решения математических задач, которые сложно или невозможно решить аналитически. Численные методы могут помочь решать задачи оптимизации, прогнозирования продаж и анализа данных, особенно при работе с большими объемами информации.
* Анализ временных рядов: использование методов анализа данных в виде последовательности значений, измеренных через равные временные интервалы, для прогнозирования будущих продаж. Анализ временных рядов может помочь выявить сезонность и тренды в данных и создать более точные прогнозы.

# Анализ и формулировка проблемы

## Описание конкретного IT-проекта

Задачей нашего проекта будет разработкой системы помощи аналитикам данных, основываясь на данных интернет-магазина спортивной обуви StockX. В данном контексте система аналитики данных должна обеспечивать автоматизированный процесс сбора, обработки и анализа информации о продажах, клиентах и ассортименте товаров.

Для решения поставленной задачи:

1. Оценить исходных данные
2. Выявить показатели, которые возможно оценить математическими методами анализа
3. Построить алгоритмы для оценки показателей
4. Построить систему генерации отчета по полученным данным

Таким образом, основная цель данного IT-проекта заключается в создании комплексной системы аналитики данных, которая будет способна обеспечивать эффективное изучение и анализ информации о продажах и ассортименте товаров. Это позволит магазинам улучшить свою стратегию продаж, оптимизировать управление запасами, повысить уровень удовлетворенности клиентов и, в итоге, увеличить прибыльность бизнеса.

## Выявление проблем и потребностей, связанных с математическим анализом

Проблемы и потребности, связанные с математическим анализом в рамках проекта на тему "Математический анализ в IT":

* Интеграция системы аналитики данных с существующими системами магазина: внедрение системы аналитики данных может потребовать интеграции с существующими системами управления магазином, такими как системы управления товарными запасами, CRM-системы и другие. Необходимо обеспечить гладкую интеграцию и совместимость с текущими решениями.
* Обеспечение безопасности и конфиденциальности данных: при работе с чувствительными данными магазинов, такими как данные о продажах, заказах и клиентах, необходимо гарантировать их безопасность и конфиденциальность. В рамках проекта требуется разработать механизмы защиты данных и предоставления доступа только авторизованным пользователям.
* Обеспечение масштабируемости системы аналитики данных: с учетом того, что проект может быть применен не только к интернет-магазину спортивной обуви StockX, но и к другим розничным предприятиям, важно предусмотреть возможность масштабирования системы аналитики данных для работы с большим количеством магазинов и отраслями.

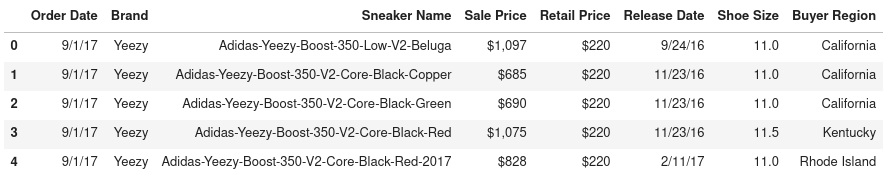
## Оценка исходных данных

Для задач нашего проекта мы получили данные продаж интернет-магазина StockX.

Таблица продаж магазина StockX за период 01.09.2017 - 13.02.2019

В таблице ниже записаны данные, с которыми мы работаем:

1. столбцом “Order Date” - данные о дате заказа.
2. столбцом “Brand” - данные о названии бренда модели.
3. столбцом “Sneaker Name” - данные о названии модели.
4. столбцом “Sale Price” - данные о цене продажи.
5. столбцом “Retail Price” - данные о цене закупки.
6. столбцом “Release Date” - данные о дате продажи.
7. столбцом “Shoe Size” - данные о размере модели.
8. столбцом “Buyer Region” - данные о регионе покупателя.



Описание колонок из таблицы выше, с которыми мы работаем:

1. столбец “#” - Нумерация строк.
2. столбец “Columns” - Строки с названием.
3. столбец “Non-Null” - Означает, что строка не пуста и содержит информацию.
4. столбец “Count” - указывает на количество заполненных значений в соответствующей строке.
5. столбец “Dtype” - Означает, как они будут представлены в памяти компьютера.

RangeIndex: 99956 entries, 0 to 99955  
Data columns (total 8 columns):  
 # Column Non-Null Count Dtype   
--- ------ -------------- -----   
 0 Order Date 99956 non-null object   
 1 Brand 99956 non-null object   
 2 Sneaker Name 99956 non-null object   
 3 Sale Price 99956 non-null object   
 4 Retail Price 99956 non-null object   
 5 Release Date 99956 non-null object   
 6 Shoe Size 99956 non-null float64  
 7 Buyer Region 99956 non-null object   
dtypes: float64(1), object(7)  
memory usage: 6.1+ MB

# Подготовка данных

В результате оценки данных выяснили:

1. Колонки Order Data представлены форматом object, надо перевести в datetime.
2. Колонки Order Data представлены форматом object, надо перевести в datetime.
3. Колонки Sale Price представлены форматом object, надо убрать значок “$” и перевести в int.
4. Колонки Retail Price представлены форматом object, надо убрать значок “$” и перевести в int.
5. В данных нет пропусков.

# Какие показатели оцениваем

* + - 1. Зависимость продаж моделей кроссовок каждого бренда от размера обуви может помочь выявить самые популярные размеры для каждой модели обуви. Полученная информация может помочь при принятии решения при закупке следующей партии и снизить количество единиц товара на складе, увеличить оборот.
      2. Количество продаж моделей кроссовок каждого бренда от месяца позволяет опередить какие модели формируют сезонный спрос.
      3. Зависимость продаж моделей кроссовок каждого бренда от цены за пару позволяет опередить влияние цены на продажи единиц товара.
      4. Зависимость продаж моделей кроссовок каждого бренда от даты продажи позволяет опередить влияние пики продаж.
      5. Прогноз продаж по неделям позволяет предсказывать продажи на основе исторических данных, прогнозировать будущие продажи с большей точностью.

# Анализ данных

1. Для анализа зависимости продаж моделей кроссовок различных брендов от размера обуви применяем метод визуализации данных, построив график в виде тепловой карты.

2. Для анализа зависимости продаж моделей кроссовок каждого бренда от цены за пару используем метод визуализации данных, строим график баров (столбиковый).

3, 4. Для анализа зависимости продаж моделей кроссовок каждого бренда от месяца используем метод визуализации данных, строим график типа scatter plot и график временной ряд график plot.

5. Для анализа прогноза продаж используем метод анализа временных рядов, для анализа выбрана модель SARIMA.

Модель SARIMA (Seasonal AutoRegressive Integrated Moving Average) является одним из распространенных методов анализа временных рядов и позволяет прогнозировать значения временного ряда на основе его прошлых значений. SARIMA учитывает сезонность, тренды, циклы и шум в данных, что делает ее очень полезной для прогнозирования продаж в интернет-магазине StockX.

В ходе анализа временных рядов был использован «тест Дики – Фуллера» - для анализа стационарности временного ряда. Тест Дики-Фуллера — это статистический тест, используемый для проверки наличия единичных корней во временных рядах. Он позволяет определить, является ли временной ряд стационарным или нет.

# Построение модульной системы

Архитектура проекта построена с учетом возможности отключения или подключения дополнительного функционала.

Гибкость системы была достигнута путем разделения основного кода на функциональные модули.

Разделение на модули было реализовано за счет разделения кода проекта на основную часть и отдельные функции, каждая из функций выполняет независимую часть работы и может быть исключена из общего алгоритма.

# Вывод

Мы разработали систему помощи аналитикам данных. Благодаря нашей системе аналитик получает возможность быстрой генерации отчетов по анализируемым показателям с минимальным вмешательством человека за короткий промежуток времени.

Формирование представленного в приложении отчета на ПК среднего уровня занимает меньше минуты.

Разработанная система анализа данных продаж на примере интернет-магазина спортивной обуви StockX позволит:

* автоматизировать процесс сбора, хранения и анализа данных о продажах, что позволит увеличить точность прогнозирования спроса на товары и оптимизировать запасы.
* автоматизировать процесс создания отчетов на основе данных аналитики продаж и других источников данных. Это сократит время на создание отчетов и повысит качество принимаемых решений.
* предсказывать спрос на товары с большой точностью и оптимизировать стратегию ценообразования. Это повысит прибыльность магазина и улучшит качество обслуживания клиентов.

Таким образом, применение математического методов анализа является эффективным инструментом для повышения эффективности работы интернет-магазина.

При разработке аналогичных систем могут возникнуть следующие проблемы:

* Преодоление проблемы недостатка данных или некачественных данных: в некоторых случаях магазины могут столкнуться с проблемой недостатка данных или некачественных данных для обучения модели предсказания продаж. В таких случаях необходимо использовать методы улучшения качества данных или применять подходы, способные работать с неполными данными.

# Список источников, использованных для подготовки проекта

1. "IEEE Transactions on Knowledge and Data Engineering" - журнал, посвященный теории и практике обработки и анализа данных.
2. Айвазян С. А., Даниелян Э. А., Саакян А. А. Прикладная статистика: классификация и снижение размерности. — М.: Финансы и статистика, 1989. — 607 с.
3. Дьяконов А. Г. Курс лекций по математической статистике. — М.: МАКС Пресс, 2005. — 320 с.
4. Кун Т. Python для сложных задач: наука о данных и машинное обучение. — СПб.: Питер, 2018. — 706 с.
5. Харт П. Е. и др. Статистические методы машинного обучения / Пер. с англ. под ред. Д. В. Черных, О. В. Кузнецова, А. А. Черных. — М.: Лори, 2017. — 547 с.
6. Книга "Математический анализ: Учебное пособие" авторов Н.П. Белугина и М.А. Чеснокова.
7. Официальная документация библиотек и инструментов, используемых в проекте, таких как Python, Pandas, Matplotlib
8. Книга “Python и анализ данных автора” Уэс Маккинни
9. Книга "Data Science for Business" авторов Foster Provost и Tom Fawcett.
10. Статьи и руководства на сайте Kaggle
11. Data Science Наука о данных с нуля, Джоэл Грас
12. Наука о данных. Учебный курс С. Скиена.
13. Документация библиотек Python, таких как NumPy, Pandas, Scikit-learn, для более глубокого понимания принципов работы и использования этих инструментов.

# Приложение