**МИНИСТЕРСТВО ТРАНСПОРТА РОССИЙСКОЙ ФЕДЕРАЦИИ**

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ**

**РОССИЙСКИЙ УНИВЕРСИТЕТ ТРАНСПОРТА РУТ (МИИТ)**

Высшая инженерная школа

**ОТЧЕТ**

**по ознакомительной практике**

Выполнил: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Руководитель практики от

университета:

Руководитель практики от

организации: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Москва 2024

Оглавление

[Исследовательский этап 4](#_Toc173011357)

[1.1. Описание предметной области 4](#_Toc173011358)

[1.2. Описание задач 5](#_Toc173011359)

[Экспериментально-практическая часть 6](#_Toc173011360)

[1.1. Алгоритм, разработанный для реализации 6](#_Toc173011361)

[1.2. Реализация алгоритма 8](#_Toc173011362)

[1.3. Тестирование 12](#_Toc173011363)

[Заключение 13](#_Toc173011364)

[Список использованных источников. 15](#_Toc173011365)

[Приложение 1 16](#_Toc173011366)

**Введение**

В рамках данной практической работы, проходившей в компании "РЖД Бизнес Актив", я занимался анализом данных, связанный с деятельностью компании в области контейнерных грузоперевозок.

АО "РЖД Бизнес Актив" является мультимодальным контейнерным оператором, который занимается организацией перевозок на базе собственных и привлеченных активов. Диверсифицированная грузовая база и сбалансированный портфель активов обеспечивают "РЖД Бизнес Актив" устойчивое положение на транспортном рынке. В арсенале компании насчитывается 15 000 контейнеров и 11 терминалов, что позволяет эффективно оказывать услуги по импорту, экспорту и транзиту, а также осуществлять внутрироссийские перевозки.

Основная цель проекта заключалась в улучшении процесса сбора, обработки и визуализации данных для повышения эффективности управления и принятия решений в компании.

Работа включала в себя выполнение следующих задач:

* Сбор данных из CSV файлов, предоставленных компанией.
* Очистка и предобработка данных для их дальнейшего анализа.
* Вычисление ключевых метрик.
* Создание интерактивных отчетов для удобного представления результатов анализа.

В процессе работы я использовал современные инструменты и технологии, такие как Python и его библиотеки для анализа и визуализации данных.

# Исследовательский этап

## ****Описание предметной области****

Предметная область охватывает анализ данных, используемый для оптимизации логистических процессов и повышения эффективности деятельности компании "АО РЖД Бизнес Актив". Основное внимание уделялось применению методов анализа данных с целью улучшения управления ресурсами и качества предоставляемых услуг в сфере контейнерных грузоперевозок.

Основные элементы предметной области:

1. Сбор данных: Процесс включал извлечение информации из различных источников, таких как CSV файлы, предоставленные компанией. Эти данные содержали информацию о маршрутах перевозок, типах и объемах грузов, состоянии контейнеров, финансовых показателях и клиентской активности.
2. Предобработка данных: Этап включал очистку и преобразование данных для их дальнейшего анализа.
3. Анализ данных: Основной задачей анализа было вычисление ключевых метрик, таких как заполняемость контейнеров, доходы и расходы.
4. Визуализация данных: Создание интерактивных отчетов, которые помогли представлять результаты анализа в наглядной и доступной форме. Визуализации включали графики и различные диаграммы, что позволило лучше понимать текущую ситуацию и принимать обоснованные решения.

## **Описание задач**

Основные задачи в анализе данных:

1. Оптимизация использования активов: Эффективное использование контейнеров и терминалов требует тщательного анализа данных о заполняемости и состоянии контейнеров. Это позволяет минимизировать время простоя и увеличить доходность перевозок.
2. Финансовый анализ: Вычисление и анализ финансовых показателей, таких как доходы, расходы и прибыль, позволяют оценить рентабельность операций и принять меры для улучшения финансовых результатов.
3. Клиентский анализ: Изучение данных о клиентах и их активности помогает улучшить качество обслуживания, определить наиболее выгодных клиентов и разработать стратегии для увеличения клиентской базы.

# Экспериментально-практическая часть

## Алгоритм, разработанный для реализации

В рамках практики был разработан алгоритм для анализа данных, предназначенный для оптимизации логистических процессов и повышения эффективности работы компании "АО РЖД Бизнес Актив". Основные этапы алгоритма включают сбор данных, их предобработку, анализ, визуализацию и составление отчетов.

Ниже представлен детальный обзор каждого из этих этапов:

1. Сбор данных

Для сбора данных использовались три CSV файла, содержащие информацию за один месяц:

1. transports\_data.csv: Данные о перевозках, включая идентификаторы, даты начала и окончания перевозки, тип груза и маршрут.
2. containers\_data.csv: Информация о состоянии контейнеров, такие как идентификаторы контейнеров, текущая загрузка, вместимость и местоположение.
3. financials\_data.csv: Финансовые показатели, включая доходы, расходы и другие параметры.
4. Обработка данных
5. Очистка данных:

Проверка на дубликаты: Удаление повторяющихся записей для обеспечения уникальности данных.

Удаление пропусков: Удаление строк с пропущенными значениями для обеспечения полноты данных.

Проверка на аномалии: Выявление аномальных значений в данных.

1. Расчет метрик:

Доход и стоимость на отгрузку: показывает средние значение дохода и стоимости, полученных от одной отгрузки.

Общая валовая прибыль: показывает разницу между суммарным доходом и суммарными затратами на все отгрузки.

Общая сумма продаж: отражает суммарный доход, полученный от всех отгрузок.

Коэффициент заполнения 1 контейнера: показывает среднюю заполненность 1 контейнера в процентах.

Процент окупаемости контейнера: сколько процентов от стоимости контейнера, он принёс прибыли.

1. Визуализация данных и составление отчёта
2. Построение графиков и диаграмм: Создание визуализаций для представления ключевых метрик и трендов.
3. Создание интерактивных отчётов.

## Реализация алгоритма

Для решения задач были использованы следующие инструменты:

1. Python: Основной язык программирования.
2. Pandas: Для обработки данных.
3. Numpy: Для вычислений и анализа аномалий.
4. Plotly Express: Для создания графиков.
5. Dash: Для создания интерактивного дашборда.

Сбор данных:

Данные были загружены и объединены в одну таблицу с помощью библиотеки Pandas. После этого была выполнена проверка на наличие дубликатов, чтобы гарантировать уникальность каждой записи.

Анализ пропусков:

Были проанализированы строки с пропусками в ключевых столбцах: "доход", "стоимость", "тип\_груза", "маршрут" и "имя\_клиента". Это позволило выявить и устранить неполные данные, которые могли бы повлиять на точность анализа.

Поиск аномалий:

Проанализированы аномалии по следующим столбцам: "вес\_груза", "объём\_груза", "стоимость", "доход". Для этого были использованы два метода:

1. Метод стандартного отклонения.
2. Метод межквартильного размаха (IQR).

Вычисление метрик:

После очистки данных и удаления аномалий были вычислены следующие ключевые метрики:

1. Общая валовая прибыль
2. Общая сумма продаж
3. Процент окупаемости контейнера
4. Доход на отгрузку
5. Стоимость на отгрузку
6. Коэффициент заполнения одного контейнера

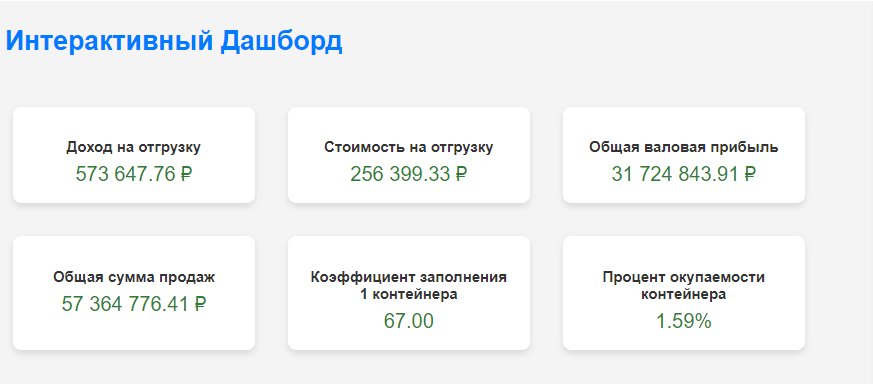
Визуализация данных:

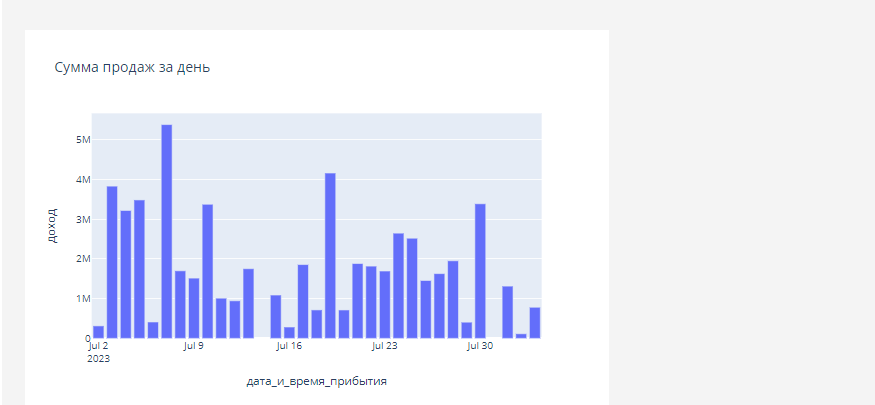
С использованием библиотеки Plotly были построены следующие графики:

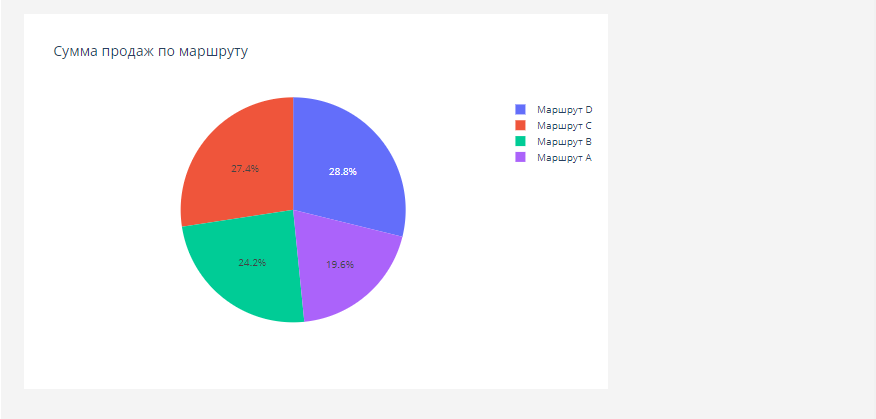
1. Круговая диаграмма, отображающая сумму продаж по маршруту.
2. Столбчатая диаграмма, показывающая сумму продаж по клиентам.
3. Столбчатая диаграмма, отображающая сумму продаж по дням.

Создание отчета:

С помощью библиотеки Dash был сформирован интерактивный отчет, который включил в себя вышеуказанные графики и рассчитанные метрики. На рис. 1 показан прототип отчета за месяц.







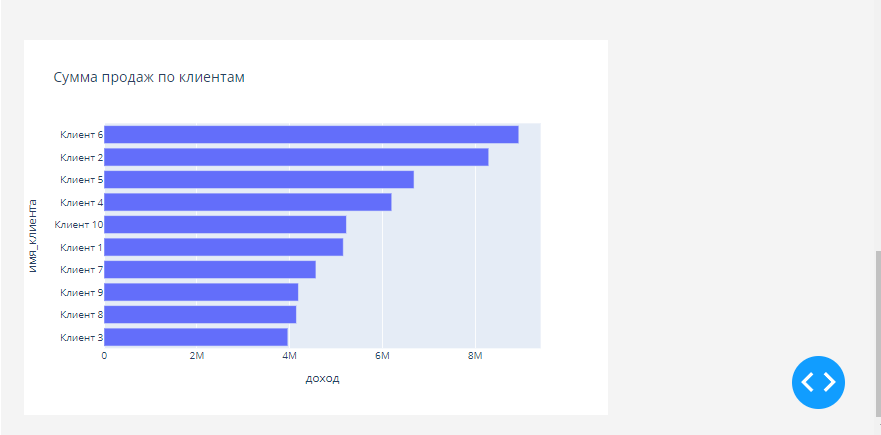


Рис. 1. Прототип интерактивного отчета, включающего метрики и визуализации.

Этот отчет позволяет получить представление о данных по перевозкам, а также выявить ключевые тенденции.

## Тестирование

Проект был протестирован для обеспечения точности и надежности всех этапов анализа данных и визуализации. Отчеты был сформированы для нескольких месяцев и показали свою точность и сопоставимость с графиками в Datalens, что подтверждает его надежность и информативность. Все результаты и визуализации соответствуют ожиданиям и предоставляют ценную информацию для дальнейшего анализа и принятия решений.

# Заключение

В ходе выполнения практической работы был разработан и реализован алгоритм для анализа данных компании "АО РЖД Бизнес Актив", специализирующейся на мультимодальных контейнерных перевозках. Основные этапы алгоритма включают сбор данных из CSV файлов, их предобработку, анализ, визуализацию и составление отчетов.

Разработанный алгоритм позволяет получать важную информацию о процессах компании, что способствует повышению эффективности управления и улучшению качества предоставляемых услуг.

Перспективы развития проекта:

1. Автоматизация сбора данных:

Внедрение системы автоматизированного сбора данных напрямую из базы данных компании. Для этого можно использовать инструменты, такие как Apache Airflow.

1. Использование SQL для обработки данных:

Перевод части предобработки данных на уровень базы данных с использованием SQL-запросов. Это позволит повысить эффективность обработки больших объемов данных и улучшить производительность системы.

1. Расширение функциональности дашбордов:

Добавление новых визуализаций и аналитических панелей для более глубокого анализа данных.

# Список использованных источников.

1. Pandas: <https://pandas.pydata.org/pandas-docs/stable/index.html>
2. Plotly: <https://plotly.com/python/>
3. Dash: <https://dash.plotly.com/>
4. Python Data Science Handbook by Jake VanderPlas: https://jakevdp.github.io/PythonDataScienceHandbook/

# Приложение 1