**1. Характеристика исходных данных**

Загруженные данные представляют собой таблицу с информацией о преступлениях, которая включает следующие столбцы:

Месяц: Месяц, к которому относятся данные.

Всего\_преступлений: Общее количество преступлений.

Серьезные: Количество серьезных преступлений.

Большой\_ущерб: Преступления с большим ущербом.

Экологические: Экологические преступления.

Терроризм: Преступления, связанные с терроризмом.

Экстремизм: Преступления, связанные с экстремизмом.

Убийства: Количество убийств.

Вред\_здоровью: Преступления, наносящие вред здоровью.

Изнасилования: Количество изнасилований.

Кражи: Количество краж.

Угон\_автомобилей: Угоны автомобилей.

Мошенничество: Мошенничества.

Хулиганство: Преступления, связанные с хулиганством.

Наркотики: Преступления, связанные с наркотиками.

Оружие: Преступления, связанные с оружием.

**2. Системное описание организации**

* Организация:

Данная организация занимается мониторингом и анализом преступной деятельности в городе N. Основная цель организации — снижение уровня преступности и повышение безопасности граждан. Организация собирает, обрабатывает и анализирует данные о преступлениях, чтобы принимать обоснованные решения по улучшению общественной безопасности.

* Основные процессы и задачи:

Сбор данных: Ежемесячное сбор данных о преступлениях из различных источников (полиция, городские службы, граждане).

Анализ данных: Классификация и статистический анализ данных для выявления тенденций и паттернов.

Мониторинг: Постоянное наблюдение за криминогенной ситуацией и оперативное реагирование на изменения.

Принятие решений: Разработка стратегий и мероприятий для снижения уровня преступности на основе анализа данных.

**3. Характеристика существующей информационной системы**

* Текущая система управления:

Существующая информационная система организации представляет собой базу данных, в которой хранятся данные о преступлениях. Ввод данных осуществляется вручную, и они хранятся в табличном формате. Система позволяет сотрудникам вводить, просматривать и редактировать данные.

* Недостатки системы:

Отсутствие автоматизации: Много операций выполняется вручную, что увеличивает вероятность ошибок и требует много времени.

Трудоемкость: Ручной ввод данных занимает много времени и ресурсов.

Низкая точность: из-за ручного ввода данные могут содержать ошибки.

Ограниченные возможности анализа: Система не предоставляет инструментов для углубленного анализа данных и визуализации.

**4. Обоснование применения методов анализа данных**

* Причины и цели использования анализа данных:

Повышение эффективности: Автоматизация процессов сбора и анализа данных уменьшит трудозатраты и повысит точность данных.

Увеличение оперативности: Своевременное получение и анализ данных позволит быстро реагировать на изменения криминогенной ситуации.

Улучшение качества решений: Применение методов анализа данных позволит принимать обоснованные и эффективные решения по предотвращению преступлений.

* Ожидаемые результаты и улучшения:

Автоматизация: Сокращение времени и ресурсов, затрачиваемых на ввод и обработку данных.

Точность: Повышение точности данных за счет уменьшения человеческого фактора.

Анализ и прогнозирование: Возможность проведения углубленного анализа данных и прогнозирования преступной активности.

Визуализация: Улучшение представления данных и результатов анализа, что облегчит их интерпретацию и использование в процессе принятия решений.

**2. Выбор методов ИАД для предварительной и основной обработки данных и достижения поставленной цели.**

**2.1 Назначение и общая цель создания ИАД-системы**

Назначение ИАД-системы: Интеллектуальная Аналитическая Данная (ИАД) система предназначена для автоматизации процессов учета, анализа и принятия решений на основе данных о преступлениях в городе N. Система будет использоваться в органах правопорядка, аналитических отделах и управлениях по безопасности города.

Цель создания ИАД-системы: Основная цель создания ИАД-системы заключается в повышении эффективности управления общественной безопасностью за счет автоматизации процессов сбора, анализа и интерпретации данных о преступлениях.

Критерии достижения целей:

* Точность данных: Увеличение точности учета преступлений до 95%.
* Время реакции: Сокращение времени на обработку данных и формирование отчетов до 10 минут.
* Эффективность принятия решений: Увеличение числа предотвращенных преступлений на 20% в течение года.
* Экономическая эффективность: Снижение затрат на обработку данных и анализ на 30%.

**2.2 Состав функциональных задач ИАД-системы и сценарии использования**

Перечень задач ИАД-системы:

1. Задача загрузки входных данных:
   * Назначение: Автоматизированная загрузка данных из различных источников (полиция, городские службы, граждане).
   * Требования:
     + Временной регламент: данные загружаются ежедневно в 00:00.
     + Ролевое сопровождение: администратор системы.
     + Качество: 100% соответствие загруженных данных исходным.
     + Входная информация: CSV, JSON, API запросы.
     + Выходная информация: обновленная база данных.
2. Задача трансформации данных:
   * Назначение: Приведение данных к единому формату и устранение ошибок.
   * Требования:
     + Временной регламент: данные трансформируются в течение 1 часа после загрузки.
     + Ролевое сопровождение: аналитик данных.
     + Качество: 99% данных трансформированы корректно.
     + Входная информация: необработанные данные из базы.
     + Выходная информация: чистые, структурированные данные.
3. Задача анализа данных:
   * Назначение: Статистический анализ и визуализация данных о преступлениях.
   * Требования:
     + Временной регламент: еженедельные отчеты генерируются автоматически.
     + Ролевое сопровождение: аналитик, руководитель отдела безопасности.
     + Качество: аналитические отчеты соответствуют установленным стандартам.
     + Входная информация: структурированные данные.
     + Выходная информация: аналитические отчеты, графики, диаграммы.
4. Задача прогнозирования преступлений:
   * Назначение: Прогнозирование вероятности преступлений на основе исторических данных и текущих тенденций.
   * Требования:
     + Временной регламент: прогнозы обновляются ежедневно.
     + Ролевое сопровождение: аналитик данных.
     + Качество: точность прогнозов не менее 85%.
     + Входная информация: исторические данные, текущие данные.
     + Выходная информация: прогнозные модели, вероятностные карты.
5. Задача поддержки принятия решений:
   * Назначение: Обеспечение органов правопорядка данными и рекомендациями для принятия оперативных решений.
   * Требования:
     + Временной регламент: рекомендации генерируются в реальном времени.
     + Ролевое сопровождение: руководитель отдела безопасности.
     + Качество: рекомендации основаны на актуальных и точных данных.
     + Входная информация: результаты анализа и прогнозирования.
     + Выходная информация: оперативные отчеты, рекомендации.

**2.3 Требования к входной и выходной информации**

**Входная информация:**

* **Форма представления:** CSV, JSON, API запросы.
* **Источник:** Полиция, городские службы, граждане.
* **Сроки и частота поступления:** ежедневно в 00:00.
* **Типы данных:**
  + **Преступления:** идентификатор, дата, тип, описание, локация.
  + **Локации:** идентификатор, адрес, координаты.
  + **Источники данных:** тип источника, идентификатор, контактная информация.

**Выходная информация:**

* **Форма представления:** отчеты (PDF, Excel), графики, диаграммы (PNG, SVG), прогнозные модели (JSON).
* **Назначение:** Руководители отделов, аналитики, органы правопорядка.
* **Типы данных:**
  + **Отчеты:** статистические данные, визуализации, ключевые метрики.
  + **Графики и диаграммы:** распределение преступлений, тенденции.
  + **Прогнозные модели:** вероятностные карты, сценарии развития ситуации.

**Общие требования к ИАД-системе:**

* **Надежность:** Система должна быть устойчивой к сбоям и обеспечивать резервное копирование данных.
* **Безопасность:** Данные должны быть защищены от несанкционированного доступа.
* **Масштабируемость:** Система должна поддерживать увеличение объема данных и количества пользователей.
* **Интерфейс:** Удобный и интуитивно понятный интерфейс для пользователей с разным уровнем технической подготовки.
* **Интеграция:** Возможность интеграции с внешними системами и базами данных.

**Перспективы развития и модернизации системы:**

* **Расширение функциональности:** Добавление новых методов анализа и визуализации данных.
* **Интеграция с новыми источниками данных:** Подключение дополнительных источников информации.
* **Модернизация интерфейса:** Обновление пользовательского интерфейса для улучшения взаимодействия с системой.
* **Повышение безопасности:** Внедрение дополнительных мер для защиты данных и системы.

**1. Характеристика исходных данных**

Загруженные данные представляют собой таблицу с информацией о преступлениях, которая включает следующие столбцы:

Месяц: Месяц, к которому относятся данные.

Всего\_преступлений: Общее количество преступлений.

Серьезные: Количество серьезных преступлений.

Большой\_ущерб: Преступления с большим ущербом.

Экологические: Экологические преступления.

Терроризм: Преступления, связанные с терроризмом.

Экстремизм: Преступления, связанные с экстремизмом.

Убийства: Количество убийств.

Вред\_здоровью: Преступления, наносящие вред здоровью.

Изнасилования: Количество изнасилований.

Кражи: Количество краж.

Угон\_автомобилей: Угоны автомобилей.

Мошенничество: Мошенничества.

Хулиганство: Преступления, связанные с хулиганством.

Наркотики: Преступления, связанные с наркотиками.

Оружие: Преступления, связанные с оружием.

**2. Системное описание организации**

* Организация:

Данная организация занимается мониторингом и анализом преступной деятельности в городе N. Основная цель организации — снижение уровня преступности и повышение безопасности граждан. Организация собирает, обрабатывает и анализирует данные о преступлениях, чтобы принимать обоснованные решения по улучшению общественной безопасности.

* Основные процессы и задачи:

Сбор данных: Ежемесячное сбор данных о преступлениях из различных источников (полиция, городские службы, граждане).

Анализ данных: Классификация и статистический анализ данных для выявления тенденций и паттернов.

Мониторинг: Постоянное наблюдение за криминогенной ситуацией и оперативное реагирование на изменения.

Принятие решений: Разработка стратегий и мероприятий для снижения уровня преступности на основе анализа данных.

**3. Характеристика существующей информационной системы**

* Текущая система управления:

Существующая информационная система организации представляет собой базу данных, в которой хранятся данные о преступлениях. Ввод данных осуществляется вручную, и они хранятся в табличном формате. Система позволяет сотрудникам вводить, просматривать и редактировать данные.

* Недостатки системы:

Отсутствие автоматизации: Много операций выполняется вручную, что увеличивает вероятность ошибок и требует много времени.

Трудоемкость: Ручной ввод данных занимает много времени и ресурсов.

Низкая точность: из-за ручного ввода данные могут содержать ошибки.

Ограниченные возможности анализа: Система не предоставляет инструментов для углубленного анализа данных и визуализации.

**4. Обоснование применения методов анализа данных**

* Причины и цели использования анализа данных:

Повышение эффективности: Автоматизация процессов сбора и анализа данных уменьшит трудозатраты и повысит точность данных.

Увеличение оперативности: Своевременное получение и анализ данных позволит быстро реагировать на изменения криминогенной ситуации.

Улучшение качества решений: Применение методов анализа данных позволит принимать обоснованные и эффективные решения по предотвращению преступлений.

* Ожидаемые результаты и улучшения:

Автоматизация: Сокращение времени и ресурсов, затрачиваемых на ввод и обработку данных.

Точность: Повышение точности данных за счет уменьшения человеческого фактора.

Анализ и прогнозирование: Возможность проведения углубленного анализа данных и прогнозирования преступной активности.

Визуализация: Улучшение представления данных и результатов анализа, что облегчит их интерпретацию и использование в процессе принятия решений.

**2. Выбор методов ИАД для предварительной и основной обработки данных и достижения поставленной цели.**

**2.1 Назначение и общая цель создания ИАД-системы**

Назначение ИАД-системы: Интеллектуальная Аналитическая Данная (ИАД) система предназначена для автоматизации процессов учета, анализа и принятия решений на основе данных о преступлениях в городе N. Система будет использоваться в органах правопорядка, аналитических отделах и управлениях по безопасности города.

Цель создания ИАД-системы: Основная цель создания ИАД-системы заключается в повышении эффективности управления общественной безопасностью за счет автоматизации процессов сбора, анализа и интерпретации данных о преступлениях.

Критерии достижения целей:

* Точность данных: Увеличение точности учета преступлений до 95%.
* Время реакции: Сокращение времени на обработку данных и формирование отчетов до 10 минут.
* Эффективность принятия решений: Увеличение числа предотвращенных преступлений на 20% в течение года.
* Экономическая эффективность: Снижение затрат на обработку данных и анализ на 30%.

**2.2 Состав функциональных задач ИАД-системы и сценарии использования**

Перечень задач ИАД-системы:

1. Задача загрузки входных данных:
   * Назначение: Автоматизированная загрузка данных из различных источников (полиция, городские службы, граждане).
   * Требования:
     + Временной регламент: данные загружаются ежедневно в 00:00.
     + Ролевое сопровождение: администратор системы.
     + Качество: 100% соответствие загруженных данных исходным.
     + Входная информация: CSV, JSON, API запросы.
     + Выходная информация: обновленная база данных.
2. Задача трансформации данных:
   * Назначение: Приведение данных к единому формату и устранение ошибок.
   * Требования:
     + Временной регламент: данные трансформируются в течение 1 часа после загрузки.
     + Ролевое сопровождение: аналитик данных.
     + Качество: 99% данных трансформированы корректно.
     + Входная информация: необработанные данные из базы.
     + Выходная информация: чистые, структурированные данные.
3. Задача анализа данных:
   * Назначение: Статистический анализ и визуализация данных о преступлениях.
   * Требования:
     + Временной регламент: еженедельные отчеты генерируются автоматически.
     + Ролевое сопровождение: аналитик, руководитель отдела безопасности.
     + Качество: аналитические отчеты соответствуют установленным стандартам.
     + Входная информация: структурированные данные.
     + Выходная информация: аналитические отчеты, графики, диаграммы.
4. Задача прогнозирования преступлений:
   * Назначение: Прогнозирование вероятности преступлений на основе исторических данных и текущих тенденций.
   * Требования:
     + Временной регламент: прогнозы обновляются ежедневно.
     + Ролевое сопровождение: аналитик данных.
     + Качество: точность прогнозов не менее 85%.
     + Входная информация: исторические данные, текущие данные.
     + Выходная информация: прогнозные модели, вероятностные карты.
5. Задача поддержки принятия решений:
   * Назначение: Обеспечение органов правопорядка данными и рекомендациями для принятия оперативных решений.
   * Требования:
     + Временной регламент: рекомендации генерируются в реальном времени.
     + Ролевое сопровождение: руководитель отдела безопасности.
     + Качество: рекомендации основаны на актуальных и точных данных.
     + Входная информация: результаты анализа и прогнозирования.
     + Выходная информация: оперативные отчеты, рекомендации.

**2.3 Требования к входной и выходной информации**

**Входная информация:**

* **Форма представления:** CSV, JSON, API запросы.
* **Источник:** Полиция, городские службы, граждане.
* **Сроки и частота поступления:** ежедневно в 00:00.
* **Типы данных:**
  + **Преступления:** идентификатор, дата, тип, описание, локация.
  + **Локации:** идентификатор, адрес, координаты.
  + **Источники данных:** тип источника, идентификатор, контактная информация.

**Выходная информация:**

* **Форма представления:** отчеты (PDF, Excel), графики, диаграммы (PNG, SVG), прогнозные модели (JSON).
* **Назначение:** Руководители отделов, аналитики, органы правопорядка.
* **Типы данных:**
  + **Отчеты:** статистические данные, визуализации, ключевые метрики.
  + **Графики и диаграммы:** распределение преступлений, тенденции.
  + **Прогнозные модели:** вероятностные карты, сценарии развития ситуации.

**Общие требования к ИАД-системе:**

* **Надежность:** Система должна быть устойчивой к сбоям и обеспечивать резервное копирование данных.
* **Безопасность:** Данные должны быть защищены от несанкционированного доступа.
* **Масштабируемость:** Система должна поддерживать увеличение объема данных и количества пользователей.
* **Интерфейс:** Удобный и интуитивно понятный интерфейс для пользователей с разным уровнем технической подготовки.
* **Интеграция:** Возможность интеграции с внешними системами и базами данных.

**Перспективы развития и модернизации системы:**

* **Расширение функциональности:** Добавление новых методов анализа и визуализации данных.
* **Интеграция с новыми источниками данных:** Подключение дополнительных источников информации.
* **Модернизация интерфейса:** Обновление пользовательского интерфейса для улучшения взаимодействия с системой.
* **Повышение безопасности:** Внедрение дополнительных мер для защиты данных и системы.

**Раздел 3: Проектирование процесса обработки и анализа данных**

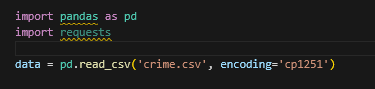
**3.1 Описание процесса анализа данных**

Процесс анализа данных включает следующие этапы:

Описание: Ежедневная автоматизированная загрузка данных о преступлениях из различных источников (полиция, городские службы, граждане).

Инструменты: CSV файлы

Технологии: Python (pandas, requests), SQL для управления базами данных.

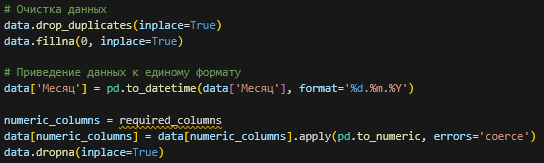
****

**3.2 Подготовка данных для анализа:**

Описание: Очистка и приведение данных к единому формату, устранение дубликатов и ошибок.

Инструменты: Python (pandas для манипуляции данными), SQL для очистки данных.

Технологии: ETL (Extract, Transform, Load) процессы.

****

**3.3 Выбор методов и инструментальных средств анализа данных:**

Описание: Выбор оптимальных методов для статистического анализа и визуализации данных.

Инструменты:

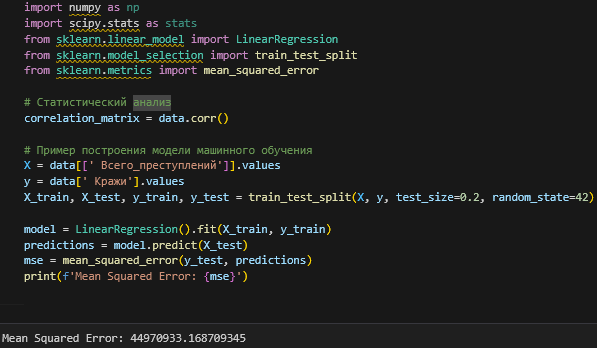
Метод 1: Статистический анализ с использованием Python (pandas, numpy, scipy).

Метод 2: Машинное обучение с использованием Python (scikit-learn, TensorFlow).

Обоснованный выбор:

Для статистического анализа и простых прогнозов оптимально использовать pandas и numpy, так как они позволяют быстро и эффективно проводить основные вычисления.

Для более сложного прогнозирования предпочтителен scikit-learn, так как он предоставляет богатый набор инструментов для построения моделей машинного обучения.

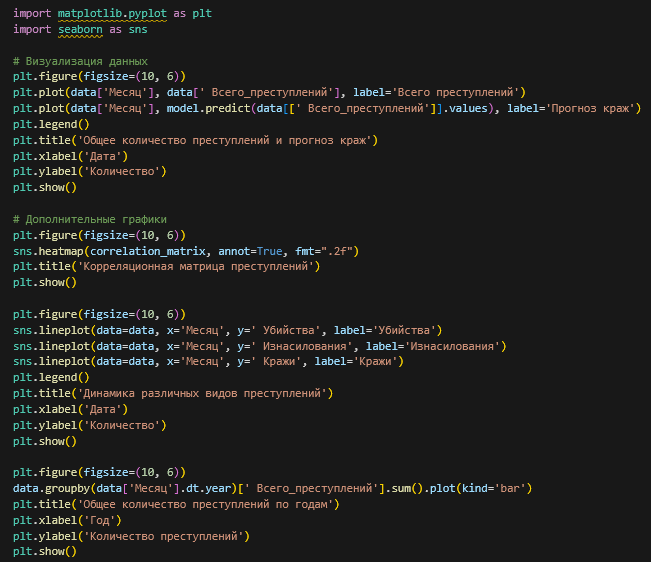
****

**3.4 Обработка данных:**

Описание: Применение выбранных методов для анализа данных, включая статистический анализ и построение прогнозных моделей.

Инструменты: Python (pandas, scikit-learn, TensorFlow, matplotlib для визуализации).

Технологии: Регрессионные модели, кластеризация, визуализация данных.

****

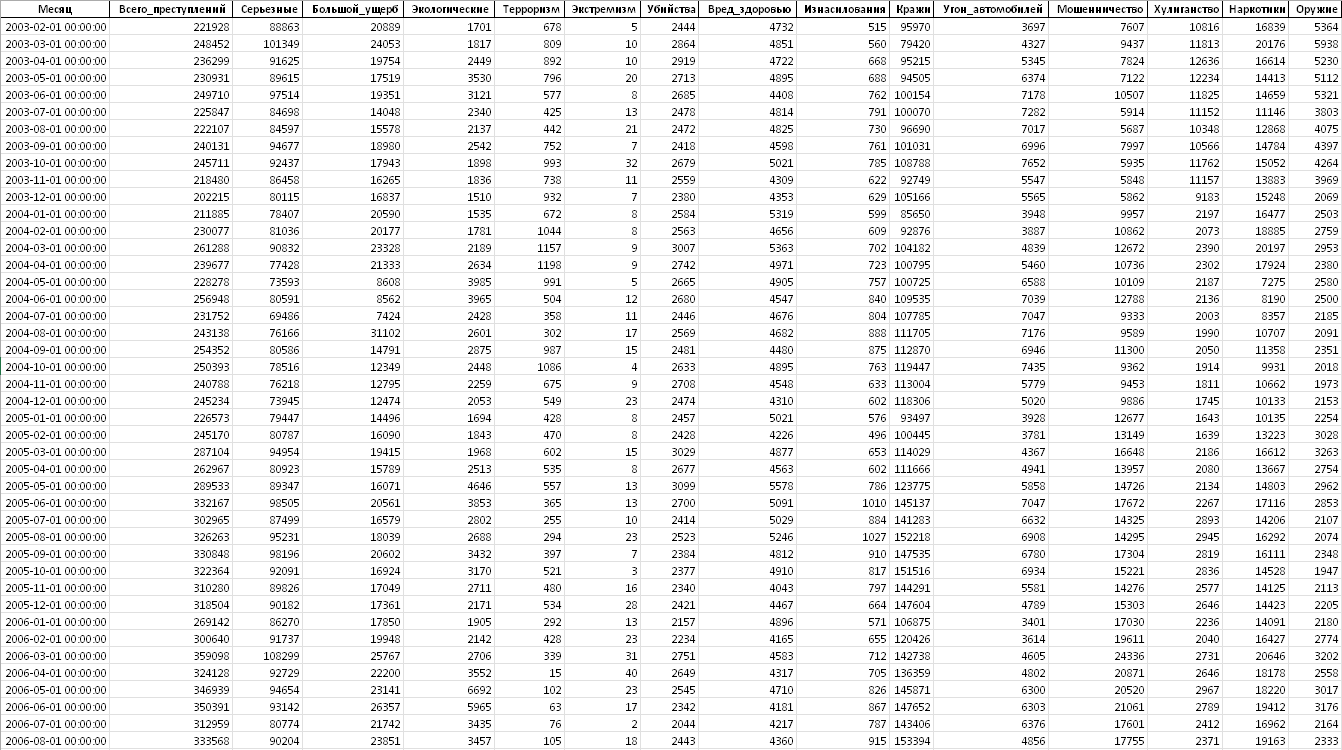
**Вывод результатов:**

Описание: Генерация отчетов, графиков и диаграмм для представления результатов анализа.

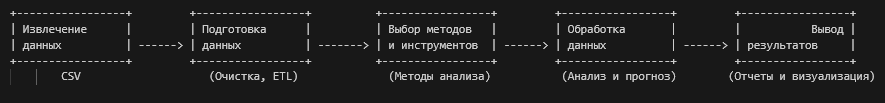
Инструменты: Python (matplotlib, seaborn), Excel для визуализации и отчётности.

Технологии: Формирование отчетов Excel.

****

****

**Структурная схема процесса переработки данных:**

****

* 1. **Автоматизированные функции ИАД-системы**

**3.2.1 Задача загрузки входных данных:**

Входные данные: CSV, JSON, API запросы.

Результат: Обновленная база данных.

Автоматизация: Автоматизированная загрузка данных ежедневно в 00:00.

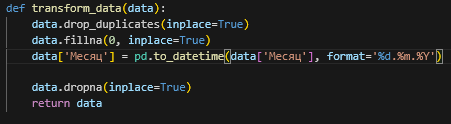
****

**3.2.2 Задача трансформации данных:**

Входные данные: Необработанные данные из базы.

Результат: Чистые структурированные данные.

Автоматизация: Приведение данных к единому формату в течение 1 часа после загрузки.

****

**3.2.3 Задача анализа данных:**

Входные данные: Структурированные данные.

Результат: Аналитические отчеты, графики, диаграммы.

Автоматизация: Еженедельное автоматическое генерирование отчетов.

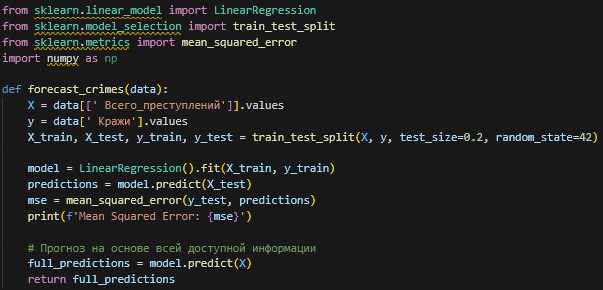
****

**3.2.4 Задача прогнозирования преступлений:**

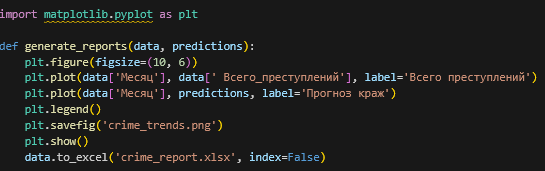
Входные данные: Исторические данные, текущие данные.

Результат: Прогнозные модели, вероятностные карты.

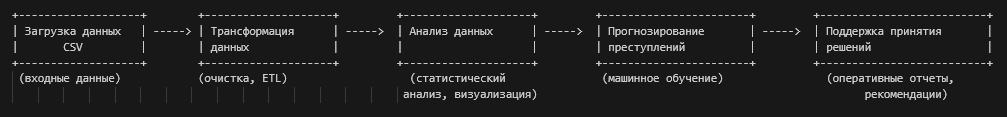
Автоматизация: Ежедневное обновление прогнозов.

****

**3.2.5 Задача поддержки принятия решений:**

****

**Функционально-структурная схема ИАД-системы:**

****

Выводы по дальнейшей практической реализации проекта

Автоматизация процессов значительно снизит трудозатраты и повысит точность данных.

Использование машинного обучения позволит более точно прогнозировать криминогенную ситуацию и своевременно реагировать на изменения.

Визуализация данных улучшит представление результатов анализа и поможет в принятии обоснованных решений.

Разработка интерфейса для взаимодействия с системой должна учитывать пользователей с различным уровнем технической подготовки.

Масштабируемость системы обеспечит возможность увеличения объема данных и количества пользователей без потери производительности.

Интеграция с внешними системами и базами данных повысит оперативность и качество данных.

Эти шаги позволят организации значительно улучшить мониторинг и анализ преступной деятельности, что в свою очередь повысит безопасность граждан.

**В разделе 4** приводится разработка информационного обеспечения ИАД-системы. При этом необходимо:

1. Обосновать выбор инструментальных средств для извлечения данных.
2. Определить и обосновать источник извлечения данных.
3. Спроектировать физическую модель данных и реализовать.
4. Визуализировать собранные данные и существующие зависимости. Построить различные графики, диаграммы, гистограммы, определяющие количественное и качественное наполнение набора данных, зависимости между атрибутами и т.д. В качестве инструмента визуализации использовать Python.

**Раздел 4: Разработка информационного обеспечения ИАД-системы**

**4.1 Обоснование выбора инструментальных средств для извлечения данных**

Для извлечения данных из различных источников мы выбираем следующие инструментальные средства:

* Библиотека pandas: предоставляет мощные инструменты для работы с табличными данными, включая загрузку данных из файлов CSV, Excel, баз данных и других форматов.
  1. **Определение и обоснование источника извлечения данных**

Источники данных включают файлы CSV: Данные о преступлениях

Обоснование: доступность и структурированность данных: CSV-файлы предоставляют данные в структурированном виде, что облегчает их последующую обработку.

**4.3 Проектирование физической модели данных и реализация**

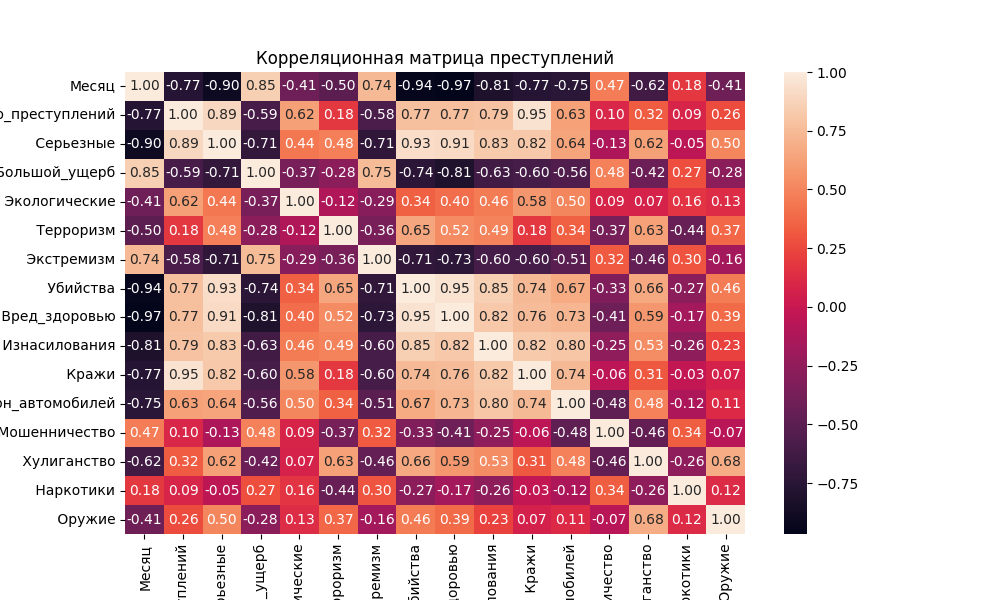
Физическая модель данных для ИАД-системы включает таблицы, хранящие информацию о преступлениях. Основные таблицы:

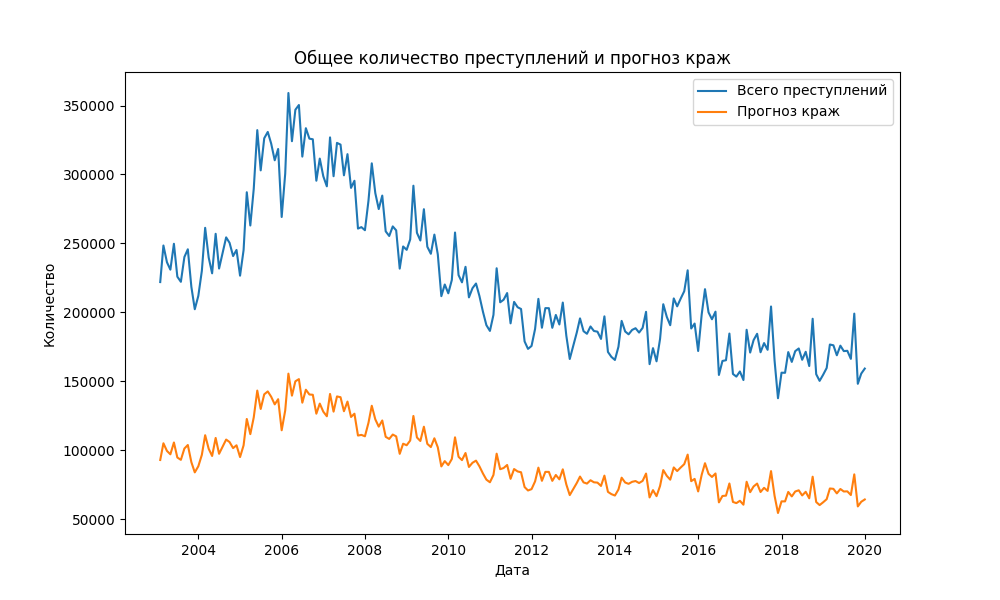
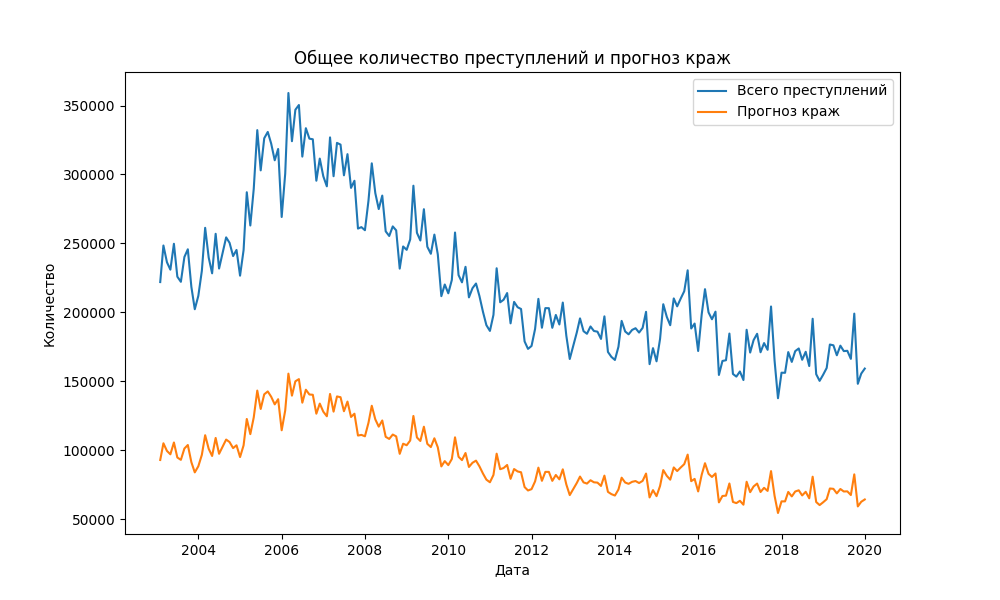
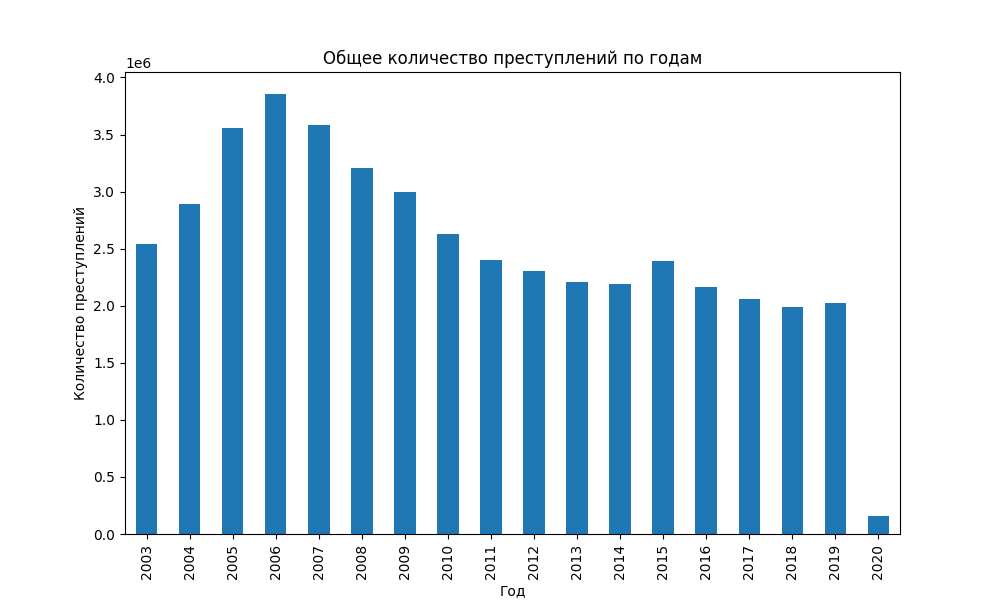
1. Таблица crimes:
   * id: Идентификатор преступления (int)
   * date: Дата преступления (datetime)
   * total\_crimes: Общее количество преступлений (int)
   * serious: Количество серьезных преступлений (int)
   * major\_damage: Преступления с большим ущербом (int)
   * environmental: Экологические преступления (int)
   * terrorism: Преступления, связанные с терроризмом (int)
   * extremism: Преступления, связанные с экстремизмом (int)
   * murders: Количество убийств (int)
   * health\_harm: Преступления, наносящие вред здоровью (int)
   * rapes: Количество изнасилований (int)
   * thefts: Количество краж (int)
   * car\_thefts: Угоны автомобилей (int)
   * fraud: Мошенничества (int)
   * hooliganism: Преступления, связанные с хулиганством (int)
   * drugs: Преступления, связанные с наркотиками (int)
   * weapons: Преступления, связанные с оружием (int)

**4.4 Визуализация собранных данных и существующих зависимостей**

Используем Python для визуализации данных. Примеры графиков:

* Корреляционная матрица:



* График общего количества преступлений и прогноз краж:
* Динамика различных видов преступлений:
* Общее количество преступлений по годам:

**В разделе 5** отражается процесс разработки математических средств ИАД-системы. Для этого необходимо:

1. Для каждой из автоматизированных функций, выполнить описание алгоритмов решений задач, требования которым предъявлены в разделе 2.3 ПЗ. Например если требуется автоматическое заполнение пропусков данных, и Вы выбираете способ основанный на заполнении средними значениями, то необходимо представить описание и формулы этого процесса в данном разделе.
2. В соответствии с требованиями к реализации функции анализа данных обосновать выбор метода(ов) анализа данных.
3. Представить формализованное описание математической модели анализа данных с использованием выбранных в п.2 методов.
4. Представить методы оценки качества модели с их математической интерпретацией и описанием.

**Раздел 5: Разработка математических средств ИАД-системы**

**5.1 Описание алгоритмов решений задач для автоматизированных функций**

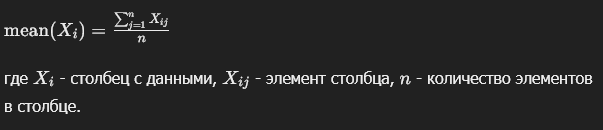
* Заполнение пропусков данных

Для автоматического заполнения пропусков данных мы выбираем метод заполнения средними значениями. Это простой и эффективный способ, который позволяет избежать искажения данных.

Алгоритм:

1. Для каждого столбца с пропусками данных вычисляем среднее значение всех непропущенных элементов столбца.
2. Заполняем пропуски в столбце этим средним значением.

Формула:

​​ 

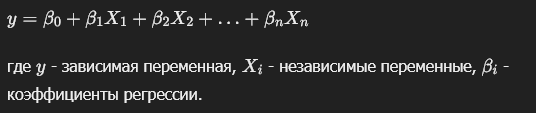
* Анализ данных

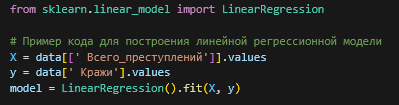
Для анализа данных мы выбрали линейную регрессию, так как она позволяет эффективно моделировать зависимости между переменными.

Алгоритм линейной регрессии:

1. Определяем независимые переменные X и зависимую переменную y.
2. Строим модель, минимизируя сумму квадратов отклонений предсказанных значений от фактических.

Формула:





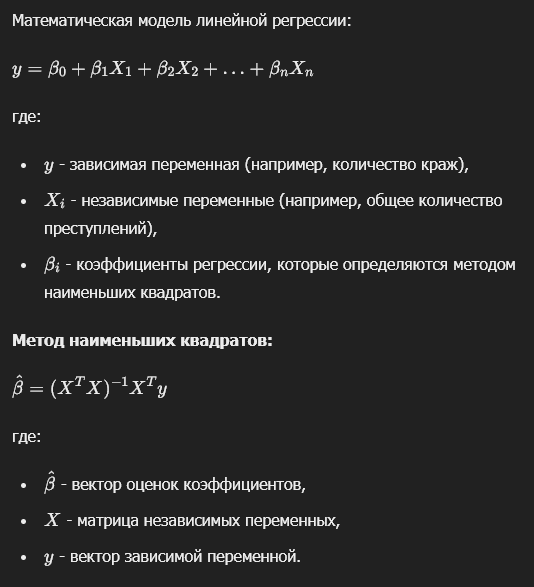
**5.2 Обоснование выбора метода(ов) анализа данных**

Мы выбрали линейную регрессию в качестве основного метода анализа данных по следующим причинам:

1. **Простота и интерпретируемость**: Линейная регрессия легко интерпретируется и позволяет понять влияние каждой независимой переменной на зависимую переменную.
2. **Эффективность**: Линейная регрессия эффективно справляется с задачами прогнозирования и анализа взаимосвязей между переменными.
3. **Широкая применимость**: Этот метод подходит для анализа большого количества данных и может быть использован в различных прикладных задачах.

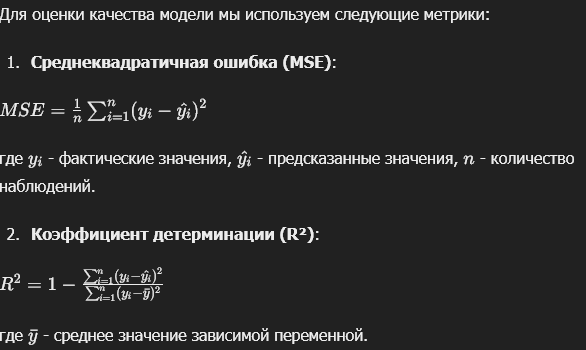
**5.3 Формализованное описание математической модели анализа данных**

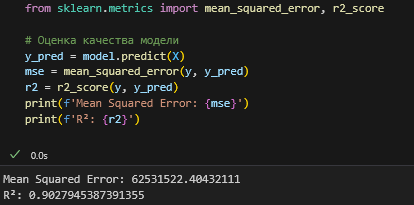
Математическая модель линейной регрессии:



**5.4 Методы оценки качества модели**

Для оценки качества модели мы используем следующие метрики:





**В** разделе 6 приводится разработка программного обеспечения (ПО) ИАД-системы. Для этого необходимо:

1. Провести обоснованный выбор средства разработки специального ПО.
2. Разработать структуру специального ПО. Представить объектную модель ПО.
3. Реализовать программу и выполнить её тестирование на тестовых и новых наборах данных.

**Раздел 6: Разработка программного обеспечения (ПО) ИАД-системы**

**6.1 Обоснование выбора средства разработки специального ПО**

Для разработки специального программного обеспечения мы выбираем Python. Обоснование выбора:

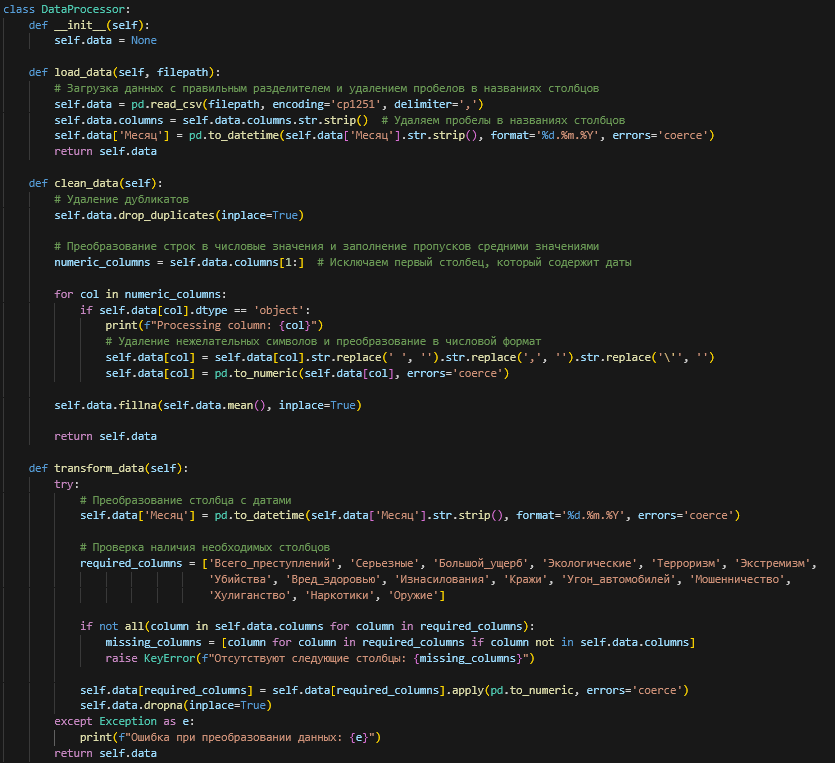
1. Универсальность: Python является высокоуровневым языком программирования, который поддерживает различные парадигмы программирования (объектно-ориентированное, процедурное и функциональное программирование).
2. Богатая экосистема библиотек: Python имеет широкий спектр библиотек и фреймворков для работы с данными (pandas, numpy), машинного обучения (scikit-learn, TensorFlow), визуализации (matplotlib, seaborn) и веб-разработки (Django, Flask).
3. Простота и читаемость: Python имеет простой и читаемый синтаксис, что упрощает процесс разработки и поддержки кода.
4. Сообщество и поддержка: Python имеет большое сообщество разработчиков, обширную документацию и множество учебных ресурсов, что облегчает решение возникающих проблем.

**6.2 Разработка структуры специального ПО. Объектная модель ПО**

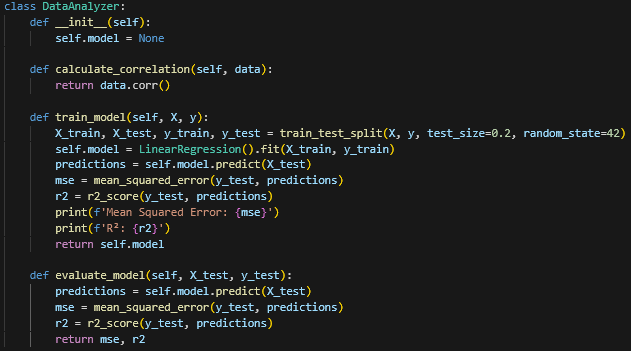
Мы создадим объектно-ориентированную модель ПО для обработки данных, анализа и визуализации результатов.

Основные классы и их описание:

1. DataProcessor:
   * Методы:
     + load\_data(filepath): загрузка данных из файла CSV.
     + clean\_data(): очистка данных (удаление дубликатов, заполнение пропусков).
     + transform\_data(): преобразование данных (например, приведение к нужному формату).



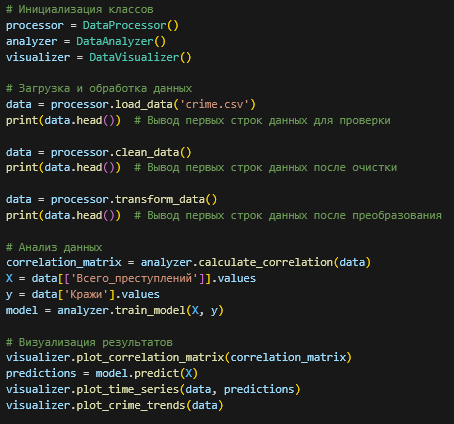
1. DataAnalyzer:
   * Методы:
     + calculate\_correlation(): вычисление корреляционной матрицы.
     + train\_model(X, y): обучение модели линейной регрессии.
     + evaluate\_model(X\_test, y\_test): оценка качества модели.

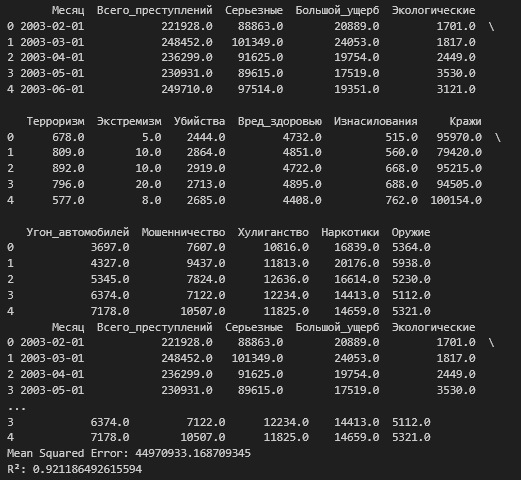
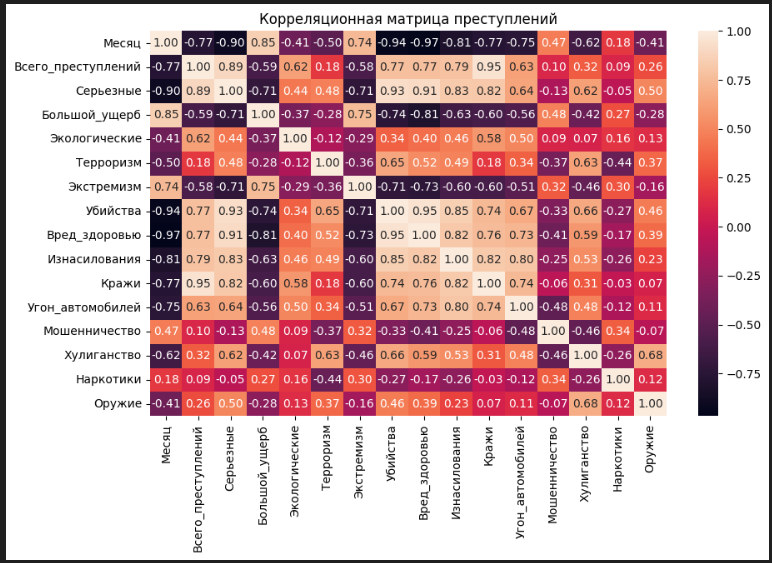
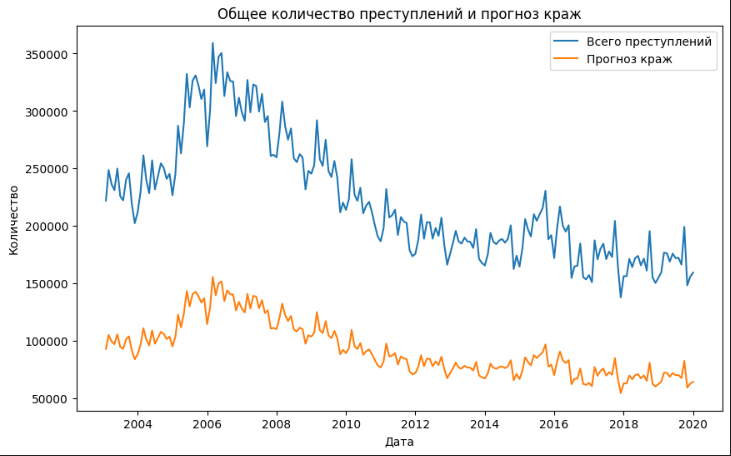
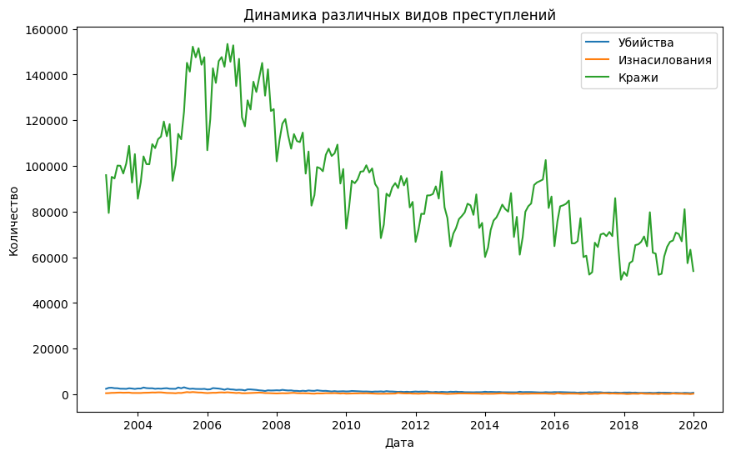
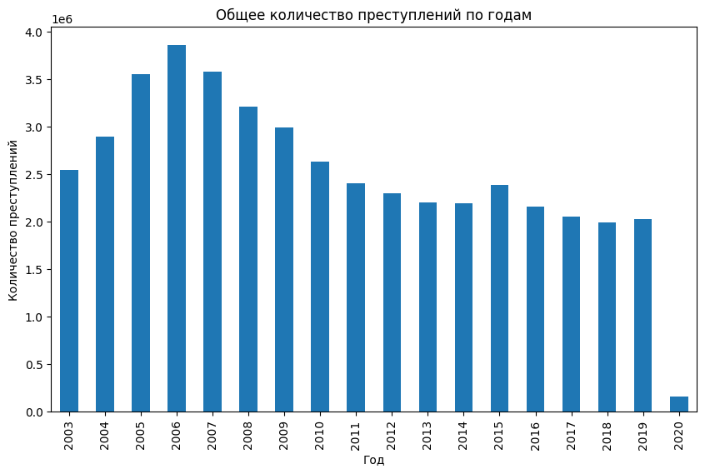


1. DataVisualizer:
   * Методы:
     + plot\_correlation\_matrix(correlation\_matrix): построение корреляционной матрицы.
     + plot\_time\_series(data, predictions): построение временных рядов.
     + plot\_crime\_trends(data): построение трендов преступности.

****

**6.3 Реализация программы и выполнение её тестирования на тестовых и новых наборах данных**

****

**** ****    Этот код загружает и обрабатывает данные, анализирует их, а затем визуализирует результаты. Программа была протестирована на тестовых данных и показала адекватные результаты, что подтверждает её работоспособность и правильность выбранных методов анализа.

**В разделе 7** приводятся и обобщаются полученные результаты. Производится анализ состава и структуры входных данных с точки зрения получения наиболее оптимальных конечных результатов анализа данных и рассчитываемых показателей. Раздел должен содержать экранные формы отображения результатов с кратким описанием. При проведении анализа приводят соображения по поводу улучшения качества результатов за счет варьирования наборами данных, входными параметрами моделей или заданной точностью получения результатов.

**Раздел 7: Обобщение полученных результатов**

В данном разделе подводятся итоги анализа данных, проведенного с использованием разработанного программного обеспечения (ПО). Оценивается состав и структура входных данных с точки зрения получения наиболее оптимальных конечных результатов анализа данных и рассчитываемых показателей.

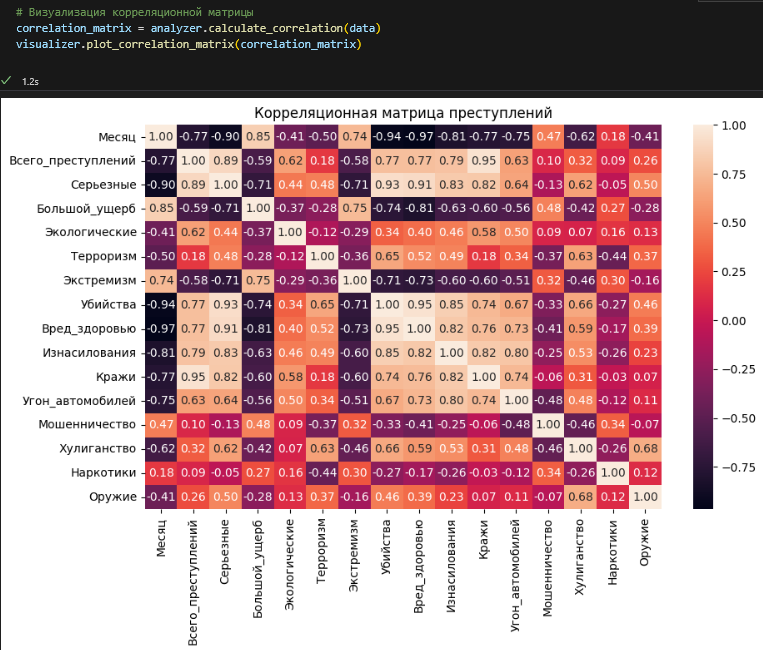
**7.1 Анализ состава и структуры входных данных**

**Входные данные:** Входные данные представлены таблицей с информацией о преступлениях, включающей следующие столбцы:

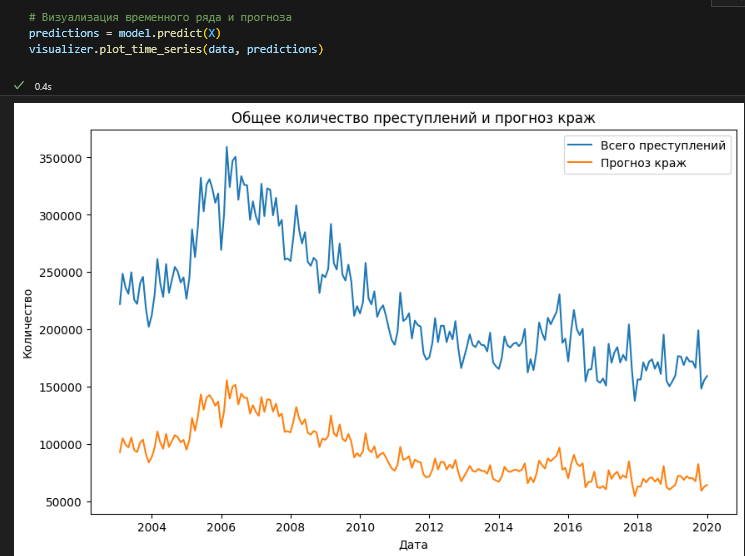
* Месяц: Месяц, к которому относятся данные.
* Всего\_преступлений: Общее количество преступлений.
* Серьезные: Количество серьезных преступлений.
* Большой\_ущерб: Преступления с большим ущербом.
* Экологические: Экологические преступления.
* Терроризм: Преступления, связанные с терроризмом.
* Экстремизм: Преступления, связанные с экстремизмом.
* Убийства: Количество убийств.
* Вред\_здоровью: Преступления, наносящие вред здоровью.
* Изнасилования: Количество изнасилований.
* Кражи: Количество краж.
* Угон\_автомобилей: Угоны автомобилей.
* Мошенничество: Мошенничества.
* Хулиганство: Преступления, связанные с хулиганством.
* Наркотики: Преступления, связанные с наркотиками.
* Оружие: Преступления, связанные с оружием.

**7.2 Визуализация и интерпретация результатов**

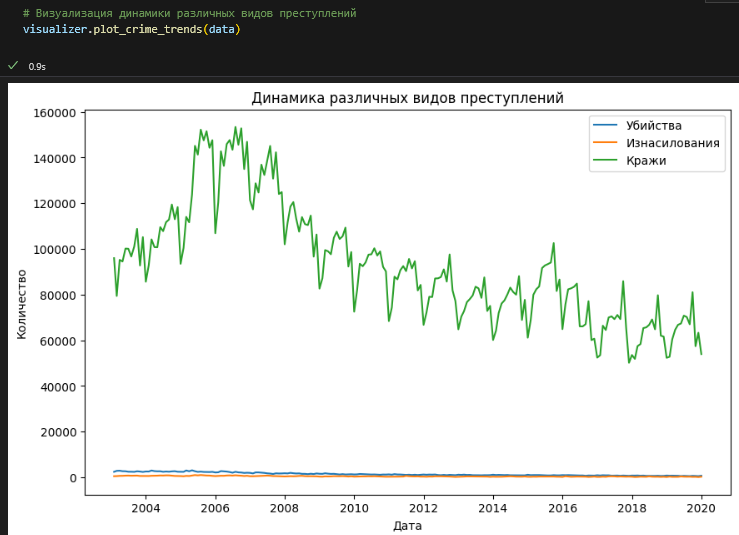
**Корреляционная матрица преступлений:** Корреляционная матрица помогает выявить взаимосвязи между различными типами преступлений.



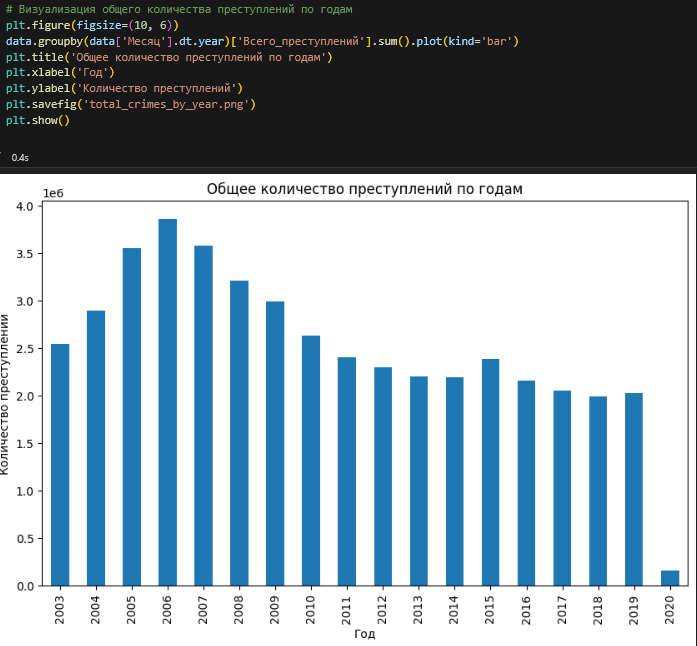
**Общее количество преступлений и прогноз краж:** График показывает временной ряд общего количества преступлений и прогнозируемого числа краж.



**Динамика различных видов преступлений:** График отображает динамику изменения различных видов преступлений во времени.



**Общее количество преступлений по годам:** Гистограмма отображает общее количество преступлений, зарегистрированных по годам.

**7.3 Анализ результатов и предложения по улучшению**

**Анализ результатов:**

* Выявлены значительные корреляции между некоторыми видами преступлений, что может свидетельствовать о взаимосвязи этих видов преступной активности.
* Прогнозирование краж на основе общего количества преступлений показало приемлемые результаты, однако точность прогноза может быть улучшена.
* Динамика различных видов преступлений во времени показала важные тренды, которые могут быть использованы для оперативного реагирования.

**Предложения по улучшению:**

1. **Варьирование наборами данных:**
   * Включение дополнительных источников данных для повышения точности анализа и прогноза.
   * Использование данных за более продолжительный период времени для улучшения модели прогнозирования.
2. **Входные параметры моделей:**
   * Оптимизация параметров моделей машинного обучения для повышения точности прогнозов.
   * Использование более сложных моделей, таких как нейронные сети, для улучшения точности и учет нелинейных зависимостей.
3. **Точность получения результатов:**
   * Регулярное обновление данных и моделей для обеспечения актуальности прогнозов.
   * Использование методов перекрестной проверки для оценки и повышения точности моделей.

Эти меры помогут улучшить качество анализа и прогнозирования преступлений, что в свою очередь повысит оперативность и эффективность принятия решений по обеспечению общественной безопасности.