Лабораторная работа 1: Обработка данных с использованием NumPy и Pandas

Цель
Цель данной лабораторной работы — научиться использовать библиотеки NumPy и Pandas для обработки и анализа данных. Студенты будут работать с массивами NumPy и объектами DataFrame Pandas, осваивая методы фильтрации, агрегирования и преобразования данных.
Описание
В рамках лабораторной работы студенты будут выполнять задачи по обработке данных, используя библиотеки NumPy и Pandas. Задания включают работу с массивами NumPy для выполнения числовых операций и работу с DataFrame Pandas для манипуляций с табличными данными. Лабораторная работа включает в себя как простые операции, так и более сложные задачи, такие как объединение и агрегация данных.

Общие задачи
1. **Работа с массивами NumPy**: Выполните операции над массивами, включая создание, индексирование, фильтрацию и агрегирование данных.
2. **Работа с DataFrame Pandas**: Выполните манипуляции с DataFrame, включая создание, фильтрацию, группировку и агрегирование данных.
3. **Анализ данных**: Используйте методы Pandas для анализа данных, включая вычисление статистических характеристик и преобразование данных.
4. **Отчёт**: Подготовьте отчёт о выполненной работе, включая описание методов, результаты и визуализации.

Индивидуальные задания
Каждому студенту предоставляется уникальная задача в соответствии с его номером в списке группы (смотреть в SSO).

1. **Анализ данных о погоде:** Создать массивы NumPy с данными о температуре и влажности. С помощью Pandas создать DataFrame и рассчитать средние значения температуры и влажности по дням недели.
2. **Обработка данных о продажах:** Загрузить данные о продажах из CSV-файла в DataFrame. Преобразовать цены в массив NumPy и рассчитать общую выручку по каждому продукту.
3. **Работа с временными рядами:** Создать массив NumPy с временными метками и соответствующими значениями. Использовать Pandas для создания временного ряда и выполнения операции ресемплинга (например, переход от минутных данных к часовым).
4. **Анализ данных о студентах:** Создать DataFrame с данными о студентах (имена, оценки, возраст). Преобразовать оценки в массив NumPy и рассчитать средний балл и стандартное отклонение по курсу.
5. **Анализ данных о населении:** Создать массив NumPy с данными о населении и площади городов. С помощью Pandas создать DataFrame и рассчитать плотность населения для каждого города.
6. **Работа с пропущенными значениями:** Загрузить DataFrame с пропущенными значениями, создать массив NumPy на основе одного из столбцов и заполнить пропущенные значения средним значением.

7. **Фильтрация и сортировка данных:** Создать DataFrame с данными о товарах и их ценах. Использовать NumPy для фильтрации цен и Pandas для сортировки товаров по возрастанию или убыванию цены.
8. **Анализ данных о посещаемости:** Создать DataFrame с данными о посещаемости студентов на лекциях. Использовать NumPy для вычисления процентного соотношения посещаемости и Pandas для группировки студентов по успеваемости.
9. **Работа с категориальными данными:** Загрузить данные о клиентах (имя, возраст, категория клиента) в DataFrame. Преобразовать категории в числовой массив NumPy и провести анализ по возрастным группам.
10. **Преобразование данных:** Загрузить набор данных о погоде в DataFrame. Преобразовать температуры из градусов Цельсия в Фаренгейты с использованием массивов NumPy и добавить новый столбец в DataFrame.
11. **Анализ данных о зарплатах:** Создать DataFrame с данными о зарплатах сотрудников и их возрасте. Использовать NumPy для расчета корреляции между возрастом и зарплатой и визуализировать результаты с помощью Pandas.
12. **Агрегация данных:** Создать DataFrame с данными о продажах в разных магазинах. Использовать NumPy для расчета суммарных продаж по каждому магазину и Pandas для группировки данных по регионам.
13. **Работа с многомерными данными:** Создать 3D-массив NumPy с данными о продажах в нескольких регионах и создать DataFrame, представляющий суммарные продажи по каждому продукту и региону.
14. **Анализ данных о трафике:** Загрузить данные о трафике на дорогах в DataFrame. Преобразовать данные в массив NumPy и рассчитать среднюю скорость движения по каждому участку дороги.
15. **Объединение данных:** Загрузить два DataFrame с данными о продажах и возвратах. Использовать NumPy для расчета чистой выручки и объединить результаты в одном DataFrame.

Вопросы для подготовки

1. Что такое библиотека NumPy и какие основные операции она поддерживает?
2. Как создавать и манипулировать массивами NumPy?
3. Какие методы используются для фильтрации и агрегирования данных в NumPy?
4. Что такое DataFrame в Pandas и как его создать?
5. Какие методы фильтрации и очистки данных доступны в Pandas?
6. Как выполнять группировку и агрегирование данных в Pandas?
7. Как преобразовывать данные в Pandas и какие операции для этого доступны?
8. Какие методы объединения DataFrame существуют в Pandas и как их использовать?
9. Как визуализировать данные в Pandas и какие графики можно построить?
10. В чём отличие между использованием NumPy и Pandas для обработки данных и когда следует использовать каждую из библиотек?

Критерии Оценки:

- Написание кода для решения индивидуальной задачи: 2 балл
- Объяснение и понимание написанного кода во время защиты: 1 балла
- Ответ на один из теоретических вопросов, выбранный преподавателем: 1 балл

***** Обязательно предоставьте отчет в формате Word, соответствующий всем стандартам.***