

Программа повышения квалификации
«Программирование на Python. Уровень 4. Анализ и визуализация данных:
Pandas, NumPy, Matplotlib»

состоит из 9 учебных тем:

Тема 1. Установка и настройка программного обеспечения. Базовый инструментарий.

Тема 2. Библиотека NumPy. Вычислительные задачи.

Тема 3. Эффективные вычисления с NumPy. Практические кейсы.

Тема 4. Библиотека Matplotlib. Визуализация данных.

Тема 5. Прикладная визуализация данных в Matplotlib. Практические кейсы.

Тема 6. Библиотека Pandas. Статистика и анализ данных.

Тема 7. Структурирование данных в Pandas. Практические кейсы.

Тема 8. Статистика и временные ряды.

Тема 9. Анализ статистики и агрегация данных. Практические кейсы.

Рабочая программа темы 1. Установка и настройка программного обеспечения. Базовый инструментарий.

Принципы организации и управления вычислениями. Используемые языки программирования и взаимосвязь между ними. Необходимые пакеты Python — NumPy, Matplotlib, Pandas, Jupyter. Дистрибутив Anaconda, его преимущества и недостатки. Установка пакетов в Linux. Установка пакетов в Windows. Возможные проблемы и методы их решения. Распределённая система контроля версий Git.

Рабочая программа темы 2. Библиотека NumPy. Вычислительные задачи.

Понятие массива и его основные характеристики. Структура библиотеки. Типы данных NumPy. Принципы векторизованных вычислений. Универсальные функции. Стандартные функции: арифметика, статистика, линейная алгебра, генерация случайных чисел.

Рабочая программа темы 3. Эффективные вычисления с NumPy. Практические кейсы.

Работа с массивами и матрицами: создание, изменение формы, булева индексация. Оптимизация циклов через векторизацию. Измерение производительности.

Рабочая программа темы 4. Библиотека Matplotlib. Визуализация данных.

Виды графиков и диаграмм: линейные, точечные, столбчатые, гистограммы, коробчатые и секторные. Основные элементы диаграммы: области (Figure, Axes), оси, сетка, легенда, аннотации. Создание и настройка диаграммы. Форматы изображений: PNG, SVG, PDF; экспорт и настройка DPI.

Рабочая программа темы 5. Прикладная визуализация данных в Matplotlib. Практические кейсы.

Построение графика функции. Создание гистограммы распределения. Настройка стилей и подписей.

Рабочая программа темы 6. Библиотека Pandas. Статистика и анализ данных.

Объект Series: индекс, значения, метаданные. Объект DataFrame: создание, просмотр, выборка строк и столбцов. Группировка и агрегирование данных. Описательные статистики и сводные таблицы.

Рабочая программа темы 7. Структурирование данных в Pandas. Практические кейсы.

Первичная обработка данных: переименование, фильтрация, сортировка. Загрузка и выгрузка данных (CSV, Excel, JSON). Организация колонок и строк. Работа с пропусками и повторами.

Рабочая программа темы 8. Статистика и временные ряды.

Типы данных для представления времени: Timestamp, DatetimeIndex, Period. Объект Period и периодические индексы. Основные операции статистики

Рабочая программа темы 9. Анализ статистики и агрегация данных. Практические кейсы.

Статистический анализ итогового набора. Объединение данных из разных источников. Сцепление и наложение (concatenating, stacking). Слияние (merging, joining).

Директор Института развития профессиональных компетенций и квалификаций



Т.А. Болтенко