

АННОТАЦИЯ РАБОЧЕЙ ПРОГРАММЫ ДИСЦИПЛИНЫ

Введение в анализ данных с Python

Автор: Шилин К.Ю.

Код и наименование направления подготовки, профиля: 38.04.01 Экономика («Экономика и финансы»)

Квалификация (степень) выпускника: Магистр

Форма обучения: очная

Цель освоения дисциплины:

Углубить компетенции в сфере аналитической работы в области экономики и финансов, а также научно-исследовательской работы.

План курса:

Тема 1. Основы языка программирования Python

Дистрибутив Anaconda. Jupyter Notebooks. Основы языка программирования. Структуры данных, функции, файлы.

Тема 2. Основы NumPy. Вычисления с векторами и матрицами (массивами)

Объекты NumPy ndarray. Основные функции. Векторное программирование. Математические и статистические методы в NumPy. Булевы операции. Сортировка. Поиск уникальных имен. Линейная алгебра. Генератор «случайных» чисел.

Тема 3. Введение в Pandas.

Структурированные данные: Series, DataFrame. Индексация. Основные функции. Расчет статистик таблиц. Импорт текстовых данных, таблиц csv, таблиц Excel.

Тема 4. Очистка и подготовка данных

Обработка отсутствующих данных. Преобразование данных: поиск дубликатов, преобразование по условию, замена значений по условию. Обнаружение и фильтрация выбросов. Дамми. Манипуляции со столбцами. Векторизация строковых данных.

Тема 5. Объединение, комбинация и изменение размера данных

Иерархическое индексирование. Переконфигурирование данных с сортировкой. Вычисление статистик. Объединение наборов данных. Изменение размерности данных, сводные таблицы.

Тема 6. Визуализация

Визуализация matplotlib. Визуализация seaborn.

Иерархическое индексирование. Переконфигурирование данных с сортировкой. Вычисление статистик. Объединение наборов данных. Изменение размерности данных, сводные таблицы.

Тема 7. Агрегирование и группировки

Иерархическое индексирование. Переконфигурирование данных с сортировкой. Вычисление статистик. Объединение наборов данных. Изменение размерности данных, сводные таблицы.

Тема 8. Временные ряды

Типы данных: время, даты. Конвертация времени и дат. Временной ряд. Диапазоны дат, частотность и смещение. Периоды, расчеты с периодами. Смена размерности и частотности временного ряда. Подвижное окно выбора.

Тема 9. Примеры анализа данных

Анализ имен новорожденных. Анализ результатов выборов.

Контактные часы: 48

Формы текущего контроля и промежуточной аттестации: домашнее задание, зачет с оценкой

Основная литература:

1. Wes MCKinney. Python for Data Analysis. 2nd edition. 2018