

МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Костромской государственный университет»
(КГУ)

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

АНАЛИЗ ДАННЫХ НА PYTHON

Направление подготовки 01.03.02 «Прикладная математика и информатика»

Направленность подготовки «Анализ данных»

Квалификация (степень) выпускника: бакалавр

Кострома
2025

Рабочая программа дисциплины «Анализ данных на python» разработана в соответствии с Федеральным государственным образовательным стандартом высшего образования по направлению подготовки 01.03.02 Прикладная математика и информатика (уровень подготовки бакалавриат), утвержденным приказом №9 от 10.01.2018 г.

Разработал: Ивков Владимир Анатольевич, доцент, к.э.н., доцент

Рецензент: Секованов Валерий Сергеевич, профессор КГУ

ПРОГРАММА УТВЕРЖДЕНА:

На заседании кафедры Прикладной математики и информационных технологий:

Протокол заседания кафедры №4 от 18.02.2025 г.

Заведующий кафедрой Прикладной математики и информационных технологий

Ивков Владимир Анатольевич, доц., к.э.н.

1. Цели и задачи освоения дисциплины

Цель дисциплины: изучение технологии анализа данных.

Задачи дисциплины:

- ознакомить с основными концепциями анализа данных их применением для решения информационных задач в прикладных областях;
- овладеть принципами программирования на python с использованием библиотек для анализа данных;
- усвоить технологии анализа и визуализации данных с помощью соответствующих библиотек.

2. Перечень планируемых результатов обучения по дисциплине

В результате освоения дисциплины обучающийся должен:

освоить компетенцию:

ОПК-1: (Способен понимать, анализировать и совершенствовать данные современных научных исследований).

Код и содержание индикаторов компетенции

ОПК-1.1. Знать: технологии разработки и анализа математических моделей в профессиональной деятельности

ОПК-1.2. Уметь: разрабатывать математические модели и проводить их анализ

ОПК-1.3. Иметь навыки: построения математических моделей в области профессиональной деятельности.

Знать:

- основные библиотеки python для анализа данных;
- основные методы статистической обработки данных;
- примеры задач анализа данных в прикладной области;

Уметь:

- обрабатывать табличные данные в pandas;
- составить и решить задачу оптимизации;
- использовать информационные технологии для анализа взаимосвязей и описательной статистики.

Владеть:

- методами проведения анализа данных при решении прикладных задач.

3. Место дисциплины в структуре ОП ВО

Дисциплина относится к обязательной части учебного плана. Изучается в пятом семестре обучения.

Изучение дисциплины основывается на ранее освоенных дисциплинах, полученных в рамках бакалавриата: «Математический анализ», «Линейная алгебра», «Теория вероятностей и математическая статистика», «Математика анализа данных».

Изучение дисциплины является основой для освоения последующих дисциплин, связанных с большими данными и интеллектуальным анализом данных.

Вместе с «Анализом данных на python» компетенцию ПК-1 формируют также дисциплины «Математика анализа данных», «Интеллектуальный анализ данных», «Технологии анализа больших данных».

4. Объём дисциплины «Анализ данных на python»

4.1. Объём дисциплины в зачётных единицах с указанием академических часов и виды учебной работы

Виды учебной работы,	Очная форма
Общая трудоемкость в зачетных единицах	3
Общая трудоемкость в часах	108
Аудиторные занятия в часах, в том числе:	28
Лекции	14
Практические занятия	
Лабораторные занятия	14
Самостоятельная работа в часах	80
Форма промежуточной аттестации	Зачет

4.2. Объём контактной работы на 1 обучающегося

Виды учебных занятий	Очная форма
Лекции	14
Практические занятия	
Лабораторные занятия	14
Консультации	8
Зачет/зачеты	
Экзамен/экзамены	
Курсовые работы	
Курсовые проекты	
Всего	28

5. Содержание дисциплины «Анализ данных на python», структурированное по темам (разделам), с указанием количества часов и видов занятий

5.1 Тематический план учебной дисциплины

№	Название раздела, темы	Всего з.е/час	Аудиторные занятия			Самостоятельная работа
			Лекц.	Практ.	Лаб.	
1	Основы python для анализа данных	0,4/15	2		2	11
2	NumPy. Вектора и матрицы	0,4/15	2		2	11
3	NumPy. Статистика	0,4/15	2		2	11
4	Визуализация данных с Matplotlib	0,4/15	2		2	11
5	Pandas. Табличные данные	0,4/15	2		2	11
6	Анализ взаимосвязей и описательная статистика	0,4/15	2		2	11
7	Анализ данных при помощи сводных таблиц	0,6/18	2		2	14
	Итого:	3/108	14		14	80

5.2. Содержание:

Тема 1. Основы python для анализа данных. Понимание основ языка, работа с библиотеками NumPy, Pandas, Matplotlib и Seaborn, а также знание статистических методов. Основы Python: Переменные и типы данных: Целые числа, числа с плавающей точкой, строки, списки, кортежи, словари, множества. Операторы: Арифметические, логические, сравнения, присваивания. Условные операторы: if, elif, else. Циклы: for, while. Функции: Определения, аргументы, возвращаемые значения. Библиотеки для анализа данных: NumPy: Основа для научных вычислений, работа с массивами и матрицами. Pandas: Работа с табличными данными, создание и обработка DataFrame. Matplotlib: Построение графиков и диаграмм. Seaborn: Расширенная библиотека для визуализации, основанная на Matplotlib. Анализ данных: Загрузка данных: Чтение данных из файлов (CSV, Excel, JSON и т.д.). Очистка данных: Обработка пропущенных значений, удаление дубликатов, преобразование типов. Анализ данных: Агрегирование, группировка, вычисление статистических показателей. Визуализация данных: Построение графиков, диаграмм, гистограмм для представления результатов. Дополнительные навыки: Работа с Jupyter Notebook: Интерактивная среда для написания и выполнения кода, визуализации результатов. Статистика: Основные понятия статистики (среднее, медиана, дисперсия и т.д.).

Тема 2. NumPy. Вектора и матрицы. В NumPy векторы и матрицы представлены объектами ndarray, которые являются многомерными массивами. Вектор - это одномерный массив, а матрица - двумерный. NumPy предоставляет широкий набор операций для работы с векторами и матрицами, таких как математические операции, индексация, срезы и т.д.

Тема 3. NumPy. Статистика. NumPy предоставляет широкий набор функций для статистических расчетов над массивами данных. Эти функции позволяют вычислять основные статистические показатели, такие как среднее, медиана, стандартное отклонение и другие. Они реализованы как методы объектов массивов NumPy, что делает их удобными для использования.

Тема 4. Визуализация данных с Matplotlib. Matplotlib – это библиотека Python для создания визуализаций данных. Она предоставляет широкий спектр возможностей для построения различных типов графиков и диаграмм, включая линейные графики, гистограммы, круговые диаграммы, и многое другое. Matplotlib позволяет как создавать простые графики, так и настраивать их внешний вид, что делает ее мощным инструментом для анализа данных.

Тема 5. Pandas. Табличные данные. Pandas - это библиотека Python, предназначенная для работы с табличными данными. Основная структура данных в Pandas - DataFrame, представляющая собой двумерную таблицу с возможностью хранения данных разных типов в столбцах. Pandas предоставляет мощные инструменты для фильтрации, сортировки, агрегации, преобразования и анализа данных. Он также поддерживает чтение и запись данных из различных форматов, включая CSV, Excel, SQL базы данных и JSON.

Тема 6. Анализ взаимосвязей и описательная статистика. Анализ взаимосвязей и описательная статистика - это два взаимодополняющих статистических подхода к изучению данных. Описательная статистика обобщает и представляет данные, используя числа, таблицы и графики. Анализ взаимосвязей, в свою очередь, исследует зависимости между переменными, выявляя, как изменение одной переменной влияет на другую. Вместе они предоставляют всестороннее понимание набора данных.

Тема 7. Анализ данных при помощи сводных таблиц. Сводные таблицы в Excel - мощный инструмент для анализа данных, позволяющий быстро и гибко обобщать, фильтровать, сортировать и анализировать большие объемы данных. Они незаменимы для выявления трендов, сравнения данных и создания наглядных отчетов.

6. Методические материалы для обучающихся по освоению дисциплины

«Анализ данных на python»

6.1. Самостоятельная работа обучающихся по дисциплине (модулю)

№ п/п	Раздел (тема) дисциплины	Задание	Часы	Методические рекомендации по	Форма контроля
-------	--------------------------	---------	------	------------------------------	----------------

				выполнению задания (при необходимости)	
1	Основы python для анализа данных	Выполнение лабораторной работы №1	11		Отчет по лабораторной работе
2	NumPy. Вектора и матрицы	Выполнение лабораторной работы №2	11		Отчет по лабораторной работе
3	NumPy. Статистика	Выполнение лабораторной работы №3	11		Отчет по лабораторной работе
4	Визуализация данных с Matplotlib	Выполнение лабораторной работы №4	11		Отчет по лабораторной работе
5	Pandas. Табличные данные	Выполнение лабораторной работы №5	11		Отчет по лабораторной работе
6	Анализ взаимосвязей и описательная статистика	Выполнение лабораторной работы №6	11		Отчет по лабораторной работе
7	Анализ данных при помощи сводных таблиц	Выполнение лабораторной работы №7	14		Отчет по лабораторной работе

6.2. Тематика и задания для практических занятий

Не предусмотрено.

6.3. Тематика и задания для лабораторных занятий

Тема 1. Основы python для анализа данных.

На вход программе подаются оценки студента (целые числа от 2 до 5), записанные в одну строку через пробел. Все эти оценки читаются и сохраняются в список с помощью следующей команды (прописать в программе): `marks = list (map(int, input().split()))`

Необходимо вычислить средний балл списка marks и вывести его на экран с точностью до десятых (один знак после запятой).

Тема 2. NumPy. Вектора и матрицы. Заполнить массив NumPy 16 случайными целыми числами в диапазоне от 0 до 100. Представить числа в виде двумерной матрицы. Вывести на экран транспонированную матрицу.

Тема 3. NumPy. Статистика. Имеются следующие данные: 52 33 10 22 28 34 39 29 21 27 31 12 28 40 46 51 44 32 16 11 29 31 38 44 31 24 9 17 32 41 47 31 42 15 21 29 50 55 37 19 57 32 7 28 23 20 45 18 29 25. Построить интервальный вариационный ряд. Для этого найти: а) размах варьирования признака по формуле $R = x_{\max} - x_{\min}$, где $x_{\min}$ – наименьшая, $x_{\max}$ – наибольшая варианты в данной выборочной совокупности; б) число интервалов вариационного ряда, пользуясь одним из приведенных ниже соотношений: $k = \sqrt[n]{n}$, $6 < k < 12$, $k = 1 + \log_2 n = 1 + 3,2 \lg n$, где n – объем выборки; в) длину h частичных интервалов по формуле $h = R / k$ и, если необходимо, округлить это значение до некоторого числа; г) записать полученный интервальный вариационный ряд. Построить дискретный вариационный ряд, взяв в качестве вариант середины интервалов непрерывного вариационного ряда, а в качестве частот – частоты непрерывного вариационного ряда.

Тема 4. Визуализация данных с Matplotlib. Построить круговую диаграмму с пустотой внутри по следующим данным: В группе английского языка занимаются 108 студентов, группе

немецкого языка – 55, группе французского языка – 32, и 20 студентов занимаются в группе китайского.

Тема 5. Pandas. Табличные данные. Предположим, у вас есть CSV-файл с данными о продажах. С помощью Python и библиотек, вы можете:

Загрузить данные в DataFrame с помощью Pandas.

Провести очистку данных (например, заполнить пропущенные значения).

Вычислить общую сумму продаж.

Построить гистограмму распределения продаж.

Сохранить результаты в отчет.

Тема 6. Анализ взаимосвязей и описательная статистика. Анализ взаимосвязей и описательная статистика - это два взаимодополняющих статистических подхода к изучению данных. Описательная статистика обобщает и представляет данные, используя числа, таблицы и графики. Анализ взаимосвязей, в свою очередь, исследует зависимости между переменными, выявляя, как изменение одной переменной влияет на другую. Вместе они предоставляют всестороннее понимание набора данных.

Тема 7. Анализ данных при помощи сводных таблиц. Сводные таблицы в Excel - мощный инструмент для анализа данных, позволяющий быстро и гибко обобщать, фильтровать, сортировать и анализировать большие объемы данных. Они незаменимы для выявления трендов, сравнения данных и создания наглядных отчетов.

7. Перечень основной и дополнительной литературы, необходимой для освоения дисциплины

а) основная:

1. Анализ данных: учебник для вузов / Под ред. Мхитаряна В.С. – М.: Издательство Юрайт, 2025. — 448 с.

2. Косников С. Н. и др. Основы анализа данных и интеллектуальные системы. Учебное пособие. – М.: Лань, 2025. — 176 с.

3. Баланов А. Н. Анализ данных на компьютере. Учеб. пособие. – М.: Лань, 2024. — 80 с.

б) дополнительная:

4. Алексеев Д. С., Щекочихин Д.В. Технологии интеллектуального анализа данных. – М.: Лань, 2024. — 176 с.

5. Миркин Б.Г. Базовые методы анализа данных: Учебник и практикум для вузов. - М.: Юрайт, 2025. — 297 с.

8. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

Информационно-образовательные ресурсы:

1. Библиотека ГОСТов. Все ГОСТы, [Электронный ресурс], URL: <http://vsegost.com/>

2. Национальный открытый университет <http://intuit.ru/>

Электронные библиотечные системы:

1. ЭБС Университетская библиотека онлайн - <http://biblioclub.ru>

2. ЭБС «Лань» <https://e.lanbook.com>

3. ЭБС «ZNANIUM.COM» <http://znanium.com>

9. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

Лекционные занятия проводятся в аудиториях с требуемым числом посадочных мест, оборудованные мультимедиа.

Лабораторные работы проводятся в компьютерных классах.

Свободно распространяемое программное обеспечение:

- рсuагm;
- офисный пакет.