


МИНОБРНАУКИ РОССИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
(ФГБОУ ВО «ВГУ»)

УТВЕРЖДАЮ

Заведующий кафедрой
цифровых технологий

 / Кургалин С.Д.

03.05.2023 г.

РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ

Б1.О.09 СОВРЕМЕННЫЕ ТЕХНОЛОГИИ АНАЛИЗА ДАННЫХ

1. Код и наименование направления подготовки:

02.04.01 Математика и компьютерные науки

2. Профиль подготовки:

компьютерное моделирование и искусственный интеллект

3. Квалификация выпускника: магистр

4. Форма обучения: очная

5. Кафедра, отвечающая за реализацию дисциплины: цифровых технологий

6. Составители программы:

Крыловецкий Александр Абрамович, к. ф.-м. н., доцент

7. Рекомендована: НМС ФКН (протокол № 4 от 03.05.2024)

8. Учебный год: 2024-2025 **Семестр:** 4

9. Цели и задачи учебной дисциплины

Целями освоения учебной дисциплины являются: овладение обучаемыми методами и моделями анализа данных.

Задачи учебной дисциплины:

- изучение и повторение математических оснований анализа данных;
- изучение основных методов и алгоритмов анализа данных;
- освоение технологий разработки ПО для анализа данных.

10. Место учебной дисциплины в структуре ООП

Дисциплина относится к обязательной части учебного плана (блок Б1).

11. Планируемые результаты обучения по дисциплине/модулю (знания, умения, навыки), соотнесенные с планируемыми результатами освоения образовательной программы (компетенциями) и индикаторами их достижения:

Код	Название компетенции	Код(ы)	Индикатор(ы)	Планируемые результаты обучения
ОПК-1	Способен находить, формулировать и решать актуальные и значимые проблемы прикладной и компьютерной математики.	ОПК-1.1	Обладает фундаментальными знаниями и практическим опытом в формулировке и решении актуальных и значимых проблем прикладной и компьютерной математики.	Знать: фундаментальные математические основы анализа данных, методы проектирования и создания программных продуктов для задач анализа данных
		ОПК-1.2	Умеет использовать их в профессиональной деятельности.	Уметь: применять математические методы при решении задач анализа данных, использовать различные инструментальные средства разработки ПО для анализа данных.
		ОПК-1.3	Имеет навыки решения актуальных и значимых проблем прикладной и компьютерной математики.	Владеть: методами прикладной математики для анализа данных, технологиями создания программных продуктов для реализации методов анализа данных в прикладных задачах.

12. Объем дисциплины в зачетных единицах/час — 2/72.

Форма промежуточной аттестации: зачет с оценкой.

13. Трудоемкость по видам учебной работы

Вид учебной работы	Трудоемкость	
	Всего	По семестрам
		4 семестр
Аудиторные занятия	28	28
в том числе:	лекции	
	практические	14
	лабораторные	14

Самостоятельная работа	44	44
в том числе: курсовая работа (проект)		
Форма промежуточной аттестации Зачёт с оценкой		
Итого:	72	72

13.1. Содержание дисциплины

п/п	Наименование раздела дисциплины	Содержание раздела дисциплины	Реализация раздела дисциплины с помощью онлайн-курса, ЭУМК *
1. Практические занятия			
1.1	Библиотека Pandas и анализ данных	Знакомство с библиотекой Pandas Обработка данных, анализ, наука и библиотека Pandas Обработка данных Анализ данных Наука о данных Предназначение библиотеки Pandas Процесс анализа данных Процесс Взаимосвязь между книгой и процессом анализа данных Понятия «данные» и «анализ» в контексте нашего знакомства с библиотекой Pandas Типы данных Переменные Временные ряды Общие понятия анализа и статистики Другие библиотеки Python, работающие вместе с библиотекой Pandas Численные и научные вычисления – NumPy и SciPy Статистический анализ – StatsModels Машинное обучение – scikit-learn PyMC – стохастическое байесовское моделирование Визуализация данных – matplotlib и seaborn	https://edu.vsu.ru/course/view.php?id=4018
1.2	Представление одномерных данных с помощью объекта Series	Настройка библиотеки Pandas Создание объекта Series Создание объекта Series с помощью питоновских списков и словарей Создание объекта Series с помощью функций NumPy Создание объекта Series с помощью скалярного значения Свойства index и values Размер и форма объекта Series Установка индекса во время создания объекта Series Использование методов head(), tail() и take() для вывода значений Получение значений в объекте Series по метке или позиции Поиск по метке с помощью оператора [] и свойства ix[] Явный поиск по позиции с помощью свойства iloc[] Явный поиск по меткам с помощью свойства loc[] Создание срезов объекта Series Выравнивание данных по меткам индекса Выполнение логического отбора Переиндексация объекта Series Модификация объекта Series на месте	https://edu.vsu.ru/course/view.php?id=4018
1.3	Представление табличных и многомерных данных с помощью объекта DataFrame	Настройка библиотеки Pandas Создание объектов DataFrame Создание объекта DataFrame на основе результатов функций NumPy Создание объекта DataFrame с помощью питоновского словаря и объектов Series Создание объекта DataFrame на основе CSV-файла Доступ к данным внутри объекта DataFrame Отбор столбцов в объекте DataFrame Отбор строк в объекте DataFrame Поиск скалярного значения по метке и позиции с помощью at[] и iat[] Создание среза датафрейма с помощью оператора [] Логический отбор строк Одновременный отбор строк и столбцов	https://edu.vsu.ru/course/view.php?id=4018

1.4	Выполнение операций над объектом DataFrame и его содержимым	Переименование столбцов Добавление новых столбцов с помощью оператора [] и метода insert() Добавление столбцов за счет расширения датафрейма Добавление столбцов с помощью конкатенации Переупорядочивание столбцов Замена содержимого столбца Удаление столбцов Присоединение новых строк Конкатенация строк Добавление и замена строк за счет расширения датафрейма Удаление строк с помощью метода drop() Удаление строк с помощью логического отбора Удаление строк с помощью среза	https://edu.vsu.ru/course/view.php?id=4018
1.5	Индексация данных	Важность применения индексов Типы индексов библиотеки Pandas Основной тип Index Индексы IntIndex и RangeIndex, в качестве меток используются целые числа Индекс FloatIndex, в качестве меток используются числа с плавающей точкой Представление дискретных интервалов с использованием IntervalIndex Категории в качестве индекса – CategoricalIndex Индексация по датам и времени с помощью DatetimeIndex Индексация периодов времени с помощью PeriodIndex Работа с индексами Создание и использование индекса в объекте Series или объекте DataFrame Отбор значений с помощью индекса Преобразование данных в индекс и получение данных из индекса Переиндексация объекта библиотеки Pandas Иерархическая индексация	https://edu.vsu.ru/course/view.php?id=4018
1.6	Категориальные данные	Создание категориальных переменных Переименование категорий Добавление категорий Удаление категорий Удаление неиспользуемых категорий Установка категорий Вычисление описательных статистик для категориальной переменной Обработка школьных оценок	https://edu.vsu.ru/course/view.php?id=4018
1.7	Численные и статистические методы	Применение численных методов к объектам библиотеки Pandas Выполнение арифметических операций над объектами DataFrame или Series Вычисление количества значений Определение уникальных значений (и их встречаемости) Вычисление минимума и максимума Вычисление n наименьших значений и n наибольших значений Вычисление накопленных значений Выполнение статистических операций с объектами библиотеки Pandas Получение итоговых описательных статистик Измерение центральной тенденции: среднее, медиана и мода Вычисление дисперсии и стандартного отклонения Вычисление ковариации и корреляции Дискретизация и квантилизация данных Вычисление ранга значений Вычисление процентного изменения для каждого наблюдения серии Выполнение операций со скользящим окном Создание случайной выборки данных	https://edu.vsu.ru/course/view.php?id=4018
1.8	Загрузка данных, приведение данных в порядок	Работа с CSV-файлами и текстовыми/табличными данными Исследование CSV-файла Чтение CSV-файла в датафрейм Указание индекса столбца при чтении CSV-файла Вывод и спецификация типа данных Указание имен столбцов Указание конкретных столбцов для загрузки Сохранение датафрейма в CSV-файл Работа с данными, в которых используются разделители полей Обра-	https://edu.vsu.ru/course/view.php?id=4018

		<p>ботка загрязненных данных, в которых используются разделители полей Чтение и запись данных в формате Excel Чтение и запись JSON-файлов Чтение HTML-файлов из Интернета Чтение и запись HDF-файлов Загрузка CSV-файлов из Интернета Чтение из базы данных SQL и запись в базу данных SQL Загрузка данных с удаленных сервисов Загрузка Базы данных по экономической статистике Федерального резервного банка Сент-Луиса Загрузка данных Кеннета Френча Загрузка данных Всемирного банка</p> <p>Настройка библиотеки Pandas Что такое приведение данных в порядок? Как работать с пропущенными данными Поиск значений NaN в объектах библиотеки Pandas Удаление пропущенных данных Обработка значений NaN в ходе арифметических операций Заполнение пропущенных данных Прямое и обратное заполнение пропущенных значений Заполнение с помощью меток индекса Выполнение интерполяции пропущенных значений Обработка дублирующихся данных Преобразование данных Сопоставление значений другим значениям Замена значений Применение функций для преобразования данных</p>	
1.9	Агрегирование данных	<p>Обзор схемы «разделение – применение - объединение» Данные для примеров Разделение данных Группировка по значениям отдельного столбца Просмотр результатов группировки Группировка по нескольким столбцам Группировка по уровням индекса Применение агрегирующих функций, преобразований и фильтров Применение агрегирующих функций к группам Преобразование групп данных Общий процесс преобразования Заполнение пропущенных значений групповым средним Вычисление нормализованных z-значений с помощью преобразования Исключение групп из процедуры агрегирования</p>	<p>https://edu.vsu.ru/course/view.php?id=4018</p>
1.10	Анализ временных рядов	<p>Представление дат, времени и интервалов Объекты datetime, day и time Создание временной метки с помощью объекта Timestamp Использование объекта Timedelta для представления временного интервала Введение во временные ряды Индексация с помощью объекта DatetimeIndex Создание временного ряда с определенной частотой Вычисление новых дат с помощью смещений Представление временных интервалов с помощью смещений дат Привязанные смещения Представление промежутков времени с помощью объектов Period Создание временного интервала с помощью объекта Period Индексация с помощью объекта PeriodIndex Обработка праздников с помощью календарей Нормализация временных меток с помощью часовых поясов Операции с временными рядами Оперережение и запаздывание Преобразование частоты временного ряда Увеличение или уменьшение шага дискретизации временного ряда Применение к временному ряду операций на основе скользящего окна</p>	<p>https://edu.vsu.ru/course/view.php?id=4018</p>
2. Лабораторные занятия			
2.1	Библиотека Pandas и анализ данных	<p>Знакомство с библиотекой Pandas Обработка данных, анализ, наука и библиотека Pandas Об-</p>	<p>https://edu.vsu.ru/course/view.php?id=4018</p>

		<p>работка данных Анализ данных Наука о данных Предназначение библиотеки Pandas Процесс анализа данных Процесс Взаимосвязь между книгой и процессом анализа данных Понятия «данные» и «анализ» в контексте нашего знакомства с библиотекой Pandas Типы данных Переменные Временные ряды Общие понятия анализа и статистики Другие библиотеки Python, работающие вместе с библиотекой Pandas Численные и научные вычисления – NumPy и SciPy Статистический анализ – StatsModels Машинное обучение – scikit-learn PyMC – стохастическое байесовское моделирование Визуализация данных – matplotlib и seaborn</p>	<p>w.php?id=4018</p>
2.2	<p>Представление одномерных данных с помощью объекта Series</p>	<p>Настройка библиотеки Pandas Создание объекта Series Создание объекта Series с помощью питоновских списков и словарей Создание объекта Series с помощью функций NumPy Создание объекта Series с помощью скалярного значения Свойства index и values Размер и форма объекта Series Установка индекса во время создания объекта Series Использование методов head(), tail() и take() для вывода значений Получение значений в объекте Series по метке или позиции Поиск по метке с помощью оператора [] и свойства ix[] Явный поиск по позиции с помощью свойства iloc[] Явный поиск по меткам с помощью свойства loc[] Создание срезов объекта Series Выравнивание данных по меткам индекса Выполнение логического отбора Переиндексация объекта Series Модификация объекта Series на месте</p>	<p>https://edu.vsu.ru/course/view.php?id=4018</p>
2.3	<p>Представление табличных и многомерных данных с помощью объекта DataFrame</p>	<p>Настройка библиотеки Pandas Создание объектов DataFrame Создание объекта DataFrame на основе результатов функций NumPy Создание объекта DataFrame с помощью питоновского словаря и объектов Series Создание объекта DataFrame на основе CSV-файла Доступ к данным внутри объекта DataFrame Отбор столбцов в объекте DataFrame Отбор строк в объекте DataFrame Поиск скалярного значения по метке и позиции с помощью at[] и iat[] Создание среза датафрейма с помощью оператора [] Логический отбор строк Одновременный отбор строк и столбцов</p>	<p>https://edu.vsu.ru/course/view.php?id=4018</p>
2.4	<p>Выполнение операций над объектом DataFrame и его содержимым</p>	<p>Переименование столбцов Добавление новых столбцов с помощью оператора [] и метода insert() Добавление столбцов за счет расширения датафрейма Добавление столбцов с помощью конкатенации Переупорядочивание столбцов Замена содержимого столбца Удаление столбцов Присоединение новых строк Конкатенация строк Добавление и замена строк за счет расширения датафрейма Удаление строк с помощью метода drop() Удаление строк с помощью логического отбора Удаление строк с помощью среза</p>	<p>https://edu.vsu.ru/course/view.php?id=4018</p>
2.5	<p>Индексация данных</p>	<p>Важность применения индексов Типы индексов библиотеки Pandas Основной тип Index Индексы IntIndex и RangeIndex, в качестве меток используются целые числа Индекс FloatIndex, в качестве меток используются числа с плавающей точкой Представление дискретных интервалов с использованием IntervalIndex Категории в качестве индекса – CategoricalIndex Индексация по датам и времени с помощью DatetimeIndex Индексация периодов времени с помощью PeriodIndex Работа</p>	<p>https://edu.vsu.ru/course/view.php?id=4018</p>

		с индексами Создание и использование индекса в объекте Series или объекте DataFrame Отбор значений с помощью индекса Преобразование данных в индекс и получение данных из индекса Переиндексация объекта библиотеки Pandas Иерархическая индексация	
2.6	Категориальные данные	Создание категориальных переменных Переименование категорий Добавление категорий Удаление категорий Удаление неиспользуемых категорий Установка категорий Вычисление описательных статистик для категориальной переменной Обработка школьных оценок	https://edu.vsu.ru/course/view.php?id=4018
2.7	Численные и статистические методы	Применение численных методов к объектам библиотеки Pandas Выполнение арифметических операций над объектами DataFrame или Series Вычисление количества значений Определение уникальных значений (и их встречаемости) Вычисление минимума и максимума Вычисление n наименьших значений и n наибольших значений Вычисление накопленных значений Выполнение статистических операций с объектами библиотеки Pandas Получение итоговых описательных статистик Измерение центральной тенденции: среднее, медиана и мода Вычисление дисперсии и стандартного отклонения Вычисление ковариации и корреляции Дискретизация и квантилизация данных Вычисление ранга значений Вычисление процентного изменения для каждого наблюдения серии Выполнение операций со скользящим окном Создание случайной выборки данных	https://edu.vsu.ru/course/view.php?id=4018
2.8	Загрузка данных, приведение данных в порядок	Работа с CSV-файлами и текстовыми/табличными данными Исследование CSV-файла Чтение CSV-файла в датафрейм Указание индекса столбца при чтении CSV-файла Вывод и спецификация типа данных Указание имен столбцов Указание конкретных столбцов для загрузки Сохранение датафрейма в CSV-файл Работа с данными, в которых используются разделители полей Обработка загрязненных данных, в которых используются разделители полей Чтение и запись данных в формате Excel Чтение и запись JSON-файлов Чтение HTML-файлов из Интернета Чтение и запись HDF-файлов Загрузка CSV-файлов из Интернета Чтение из базы данных SQL и запись в базу данных SQL Загрузка данных с удаленных сервисов Загрузка Базы данных по экономической статистике Федерального резервного банка Сент-Луиса Загрузка данных Кеннета Френча Загрузка данных Всемирного банка Настройка библиотеки Pandas Что такое приведение данных в порядок? Как работать с пропущенными данными Поиск значений NaN в объектах библиотеки Pandas Удаление пропущенных данных Обработка значений NaN в ходе арифметических операций Заполнение пропущенных данных Прямое и обратное заполнение пропущенных значений Заполнение с помощью меток индекса Выполнение интерполяции пропущенных значений Обработка дублирующихся данных Преобразование данных Сопоставление значений другим значениям Замена значений Применение функций для преобразования данных	https://edu.vsu.ru/course/view.php?id=4018
2.9	Агрегирование данных	Обзор схемы «разделение – применение - объединение» Данные для примеров Разделение дан-	https://edu.vsu.ru/course/view.php?id=4018

		ных Группировка по значениям отдельного столбца Просмотр результатов группировки Группировка по нескольким столбцам Группировка по уровням индекса Применение агрегирующих функций, преобразований и фильтров Применение агрегирующих функций к группам Преобразование групп данных Общий процесс преобразования Заполнение пропущенных значений групповым средним Вычисление нормализованных z-значений с помощью преобразования Исключение групп из процедуры агрегирования	w.php?id=4018
2.10	Анализ временных рядов	Представление дат, времени и интервалов Объекты datetime, day и time Создание временной метки с помощью объекта Timestamp Использование объекта Timedelta для представления временного интервала Введение во временные ряды Индексация с помощью объекта DatetimeIndex Создание временного ряда с определенной частотой Вычисление новых дат с помощью смещений Представление временных интервалов с помощью смещений дат Привязанные смещения Представление промежутков времени с помощью объектов Period Создание временного интервала с помощью объекта Period Индексация с помощью объекта PeriodIndex Обработка праздников с помощью календарей Нормализация временных меток с помощью часовых поясов Операции с временными рядами Опережение и запаздывание Преобразование частоты временного ряда Увеличение или уменьшение шага дискретизации временного ряда Применение к временному ряду операций на основе скользящего окна	https://edu.vsu.ru/course/view.php?id=4018

13.2. Темы (разделы) дисциплины и виды занятий

№ п/п	Наименование темы (раздела) дисциплины	Виды занятий (количество часов)				
		Лекции	Практические	Лабораторные	Самостоятельная работа	Всего
1	Библиотека Pandas и анализ данных		2	2	4	8
2	Представление одномерных данных с помощью объекта Series		1	1	4	6
3	Представление табличных и многомерных данных с помощью объекта DataFrame		1	1	4	6
4	Выполнение операций над объектом DataFrame и его содержимым		2	2	4	8
5	Индексация данных		2	2	4	8
6	Категориальные данные		1	1	4	6
7	Численные и статистические методы		1	1	4	6
8	Загрузка данных, приведение данных в порядок		1	1	4	6
9	Агрегирование данных		1	1	6	8

10	Анализ временных рядов		2	2	6	10
	Итого:		14	14	44	72

14. Методические указания для обучающихся по освоению дисциплины

Освоение дисциплины складывается из аудиторной работы (учебной деятельности, выполняемой под руководством преподавателя) и внеаудиторной работы (учебной деятельности, реализуемой обучающимся самостоятельно).

Аудиторная работа состоит из работы на лекциях и выполнения практических (или лабораторных) заданий в объёме, предусмотренном учебным планом. Лекция представляет собой последовательное и систематическое изложение учебного материала, направленное на знакомство обучающихся с основными понятиями и теоретическими положениями изучаемой дисциплины. Лекционные занятия формируют базу для практических (или лабораторных) занятий, на которых полученные теоретические знания применяются для решения конкретных практических задач. Обучающимся для успешного освоения дисциплины рекомендуется вести конспект лекций и практических (лабораторных) занятий.

Самостоятельная работа предполагает углублённое изучение отдельных разделов дисциплины с использованием литературы, рекомендованной преподавателем, а также конспектов лекций, презентационным материалом (при наличии) и конспектов практических (лабораторных) занятий. В качестве плана для самостоятельной работы может быть использован раздел 13.1 настоящей рабочей программы, в котором зафиксированы разделы дисциплины и их содержание. В разделе 13.2 рабочей программы определяется количество часов, отводимое на самостоятельную работу по каждому разделу дисциплины. Большее количество часов на самостоятельную работу отводится на наиболее трудные разделы дисциплины. Для самостоятельного изучения отдельных разделов дисциплины используется перечень литературы и других ресурсов, перечисленных в пунктах 15 и 16 настоящей рабочей программы.

Успешность освоения дисциплины определяется систематичностью и глубиной аудиторной и внеаудиторной работы обучающегося.

Требуется регулярное посещение лекций и лабораторных занятий, выполнение всех заданий, работа над самостоятельным проектом.

15. Перечень основной и дополнительной литературы, ресурсов интернет, необходимых для освоения дисциплины

а) основная литература:

№ п/п	Источник
1	Кузьмин, В. И. Методы анализа данных : учебное пособие / В. И. Кузьмин, А. Ф. Гадзаов. — 2-е изд., перераб. и доп. — Москва : РТУ МИРЭА, 2020. — 155 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/171433
2	Сандуляк, Д. А. Анализ и обработка экспериментальных данных : учебно-методическое пособие / Д. А. Сандуляк, М. Н. Полисмакова. — Москва : РТУ МИРЭА, 2023. — 64 с. — ISBN 978-5-7339-1902-7. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/382541

б) дополнительная литература:

№ п/п	Источник
1	Алексеев, Д. С. Технологии интеллектуального анализа данных / Д. С. Алексеев, О. В. Щечко-чихин. — 2-е изд., стер. — Санкт-Петербург : Лань, 2024. — 176 с. — ISBN 978-5-507-48763-9. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/362915
2	Вольфсон, М. Б. Анализ данных : учебно-методическое пособие / М. Б. Вольфсон. — Санкт-Петербург : СПбГУТ им. М.А. Бонч-Бруевича, 2023. — 69 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/381533

в) информационные электронно-образовательные ресурсы (официальные ресурсы интернет):

№ п/п	Ресурс
1	Электронная библиотека ВГУ https://lib.vsu.ru
2	ЭБС «Лань» https://e.lanbook.com/
3	«Университетская библиотека online» https://biblioclub.ru/
4	Сайт, посвящённый машинному обучению: http://www.machinelearning.ru/

16. Перечень учебно-методического обеспечения для самостоятельной работы

№ п/п	Источник
1	Кузьмин, В. И. Методы анализа данных : учебное пособие / В. И. Кузьмин, А. Ф. Гадзаов. — 2-е изд., перераб. и доп. — Москва : РТУ МИРЭА, 2020. — 155 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/171433

17. Образовательные технологии, используемые при реализации учебной дисциплины, включая дистанционные образовательные технологии (ДОТ), электронное обучение (ЭО), смешанное обучение):

В процессе изучения дисциплины используются традиционные технологии, формы и методы обучения: лекционные и практические занятия, контрольные работы, устные опросы, консультации. Углублённое освоение учебного материала происходит во время самостоятельной работы с учебной литературой.

18. Материально-техническое обеспечение дисциплины:

Компьютерный класс: специализированная мебель, персональные компьютеры на базе i3-9100-3,6ГГц, мониторы ЖК 19» (30 шт.), мультимедийный проектор, экран.

ПО: ОС Windows v.7, 8, 10, Набор утилит интерпретатор языка CPython, интерпретатор языка Anaconda, IDE PyCharm, редактор Jupiter.

19. Оценочные средства для проведения текущей и промежуточной аттестаций

Порядок оценки освоения обучающимися учебного материала определяется содержанием следующих разделов дисциплины:

№ п/п	Наименование раздела дисциплины (модуля)	Компетенция(и)	Индикатор(ы) достижения компетенции	Оценочные средства
1	Библиотека Pandas и анализ данных	ОПК-1	ОПК-1.1, ОПК-1.2, ОПК-1.3	Индивидуальный проект
2	Представление одномерных данных с помощью объекта Series	ОПК-1	ОПК-1.1, ОПК-1.2, ОПК-1.3	Индивидуальный проект
3	Представление табличных и многомерных данных с помощью объекта DataFrame	ОПК-1	ОПК-1.1, ОПК-1.2, ОПК-1.3	Индивидуальный проект
4	Выполнение операций над объектом DataFrame и его содержимым	ОПК-1	ОПК-1.1, ОПК-1.2, ОПК-1.3	Индивидуальный проект
5	Индексация данных	ОПК-1	ОПК-1.1, ОПК-1.2, ОПК-1.3	Индивидуальный проект
6	Категориальные данные	ОПК-1	ОПК-1.1, ОПК-1.2, ОПК-1.3	Индивидуальный проект
7	Численные и статистические методы	ОПК-1	ОПК-1.1, ОПК-1.2, ОПК-1.3	Индивидуальный проект

№ п/п	Наименование раздела дисциплины (модуля)	Компетенция(и)	Индикатор(ы) достижения компетенции	Оценочные средства
8	Загрузка данных, приведение данных в порядок	ОПК-1	ОПК-1.1, ОПК-1.2, ОПК-1.3	Индивидуальный проект
9	Агрегирование данных	ОПК-1	ОПК-1.1, ОПК-1.2, ОПК-1.3	Индивидуальный проект
10	Анализ временных рядов	ОПК-1	ОПК-1.1, ОПК-1.2, ОПК-1.3	Индивидуальный проект
Промежуточная аттестация форма контроля – зачёт с оценкой				Перечень вопросов к зачёту

20 Типовые оценочные средства и методические материалы, определяющие процедуры оценивания

20.1 Текущий контроль успеваемости

Контроль успеваемости по дисциплине осуществляется с помощью следующих оценочных средств:

– презентация индивидуального проекта.

Примерный перечень индивидуальных проектов.

1. Численные и научные вычисления – NumPy и SciPy.
2. Статистический анализ – StatsModels.
3. Машинное обучение – scikit-learn.
4. PyMC – стохастическое байесовское моделирование.
5. Визуализация данных – matplotlib и seaborn.
6. Обработка школьных оценок.
7. Применение численных методов к объектам библиотеки Pandas.
8. Выполнение арифметических операций над объектами DataFrame или Series.
9. Вычисление количества значений.
10. Определение уникальных значений (и их встречаемости).
11. Вычисление минимума и максимума.
12. Вычисление n наименьших значений и n наибольших значений.
13. Вычисление накопленных значений.
14. Выполнение статистических операций с объектами библиотеки Pandas.
15. Получение итоговых описательных статистик.
16. Выполнение операций со скользящим окном.
17. Создание случайной выборки данных.
18. Работа с CSV-файлами и текстовыми/табличными данными.
19. Загрузка Базы данных по экономической статистике Федерального резервного банка Сент-Луиса.
20. Загрузка данных Кеннета Френча.
21. Загрузка данных Всемирного банка.
22. Применение функций для преобразования данных.
23. Обзор схемы «разделение – применение - объединение».
24. Создание временного ряда с определенной частотой.
25. Вычисление новых дат с помощью смещений.
26. Представление временных интервалов с помощью смещений дат.
27. Обработка праздников с помощью календарей.
28. Нормализация временных меток с помощью часовых поясов.

29. Операции с временными рядами.
30. Опережение и запаздывание.
31. Преобразование частоты временного ряда.
32. Применение к временному ряду операций на основе скользящего окна.

20.2 Промежуточная аттестация

Промежуточная аттестация по дисциплине осуществляется с помощью следующих оценочных средств: перечень вопросов к зачёту.

Перечень вопросов к зачёту

1. Библиотека Pandas и анализ данных.
2. Представление одномерных данных с помощью объекта Series.
3. Представление табличных и многомерных данных с помощью объекта DataFrame.
4. Выполнение операций над объектом DataFrame и его содержимым.
5. Индексация данных.
6. Категориальные данные.
7. Численные и статистические методы.
8. Загрузка данных, приведение данных в порядок.
9. Агрегирование данных.
10. Анализ временных рядов.

Для оценивания результатов обучения на экзамене используется 4-балльная шкала: «отлично», «хорошо», «удовлетворительно», «неудовлетворительно».

Критерии оценивания компетенций	Уровень сформированности компетенций	Шкала оценок
Дан полный, развёрнутый ответ на поставленный вопрос (вопросы), обучающийся свободно оперирует основными понятиями дисциплины, ориентируется в предметной области. Изложение материала не содержит ошибок, отличается последовательностью, грамотностью, логической стройностью.	Повышенный уровень	Отлично
Дан развёрнутый ответ на поставленный вопрос (вопросы), обучающийся свободно оперирует основными понятиями дисциплины, ориентируется в предметной области. Материал изложен в целом последовательно и грамотно, отсутствуют грубые ошибки, однако имеются отдельные неточности в определениях, вычислениях, доказательствах, изложениях положений теории.	Базовый уровень	Хорошо
Ответ на поставленный вопрос (вопросы) содержит изложение только базового теоретического материала, имеются ошибки в определениях, вычислениях, доказательствах, формулировках положений теории. Нарушена логическая последовательность в изложении материала.	Пороговый уровень	Удовлетворительно
Ответ на поставленный вопрос (вопросы) отсутствует, либо содержит грубые ошибки в определениях, вычислениях, доказательствах, формулировках положений теории. Обучающийся не владеет основными понятиями дисциплины. Отсутствует логическая последовательность в изложении материала.	–	Неудовлетворительно