

МИНИСТЕРСТВО ТРАНСПОРТА РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«РОССИЙСКИЙ УНИВЕРСИТЕТ ТРАНСПОРТА»
(РУТ (МИИТ))



Рабочая программа дисциплины (модуля),
как компонент образовательной программы
базового высшего образования
по направлению подготовки
09.03.01 Информатика и вычислительная техника,
утвержденной первым проректором ФГАОУ ВО
РУТ (МИИТ) Тимониным В.С.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Основы программирования на Python для анализа данных

Направление подготовки: 09.03.01 Информатика и вычислительная техника

Направленность (профиль): IT-сервисы и технологии обработки данных на транспорте

Форма обучения: Очная

Рабочая программа дисциплины (модуля) в виде
электронного документа выгружена из единой
корпоративной информационной системы управления
университетом и соответствует оригиналу

Простая электронная подпись, выданная ФГАОУ ВО РУТ
(МИИТ)

ID подписи: 937226

Подписал: директор центра Проневич Ольга Борисовна

Дата: 24.09.2026

1. Общие сведения о дисциплине (модуле).

Формирование у обучающихся практических навыков программирования на языке Python для решения задач анализа данных, обработки табличной информации, исследования пространственных данных и подготовки аналитических выводов с использованием современных библиотек Python.

В результате освоения дисциплины обучающийся:

- изучает основные инструменты Python для анализа данных;
- осваивает библиотеки Pandas и Polars для обработки больших массивов информации;
- приобретает навыки очистки, преобразования и объединения данных;
- изучает методы группировки, агрегирования и анализа данных;
- осваивает основы визуализации результатов анализа;
- знакомится с обработкой географических данных средствами GeoPandas;
- получает практический опыт выполнения полного цикла анализа данных от загрузки до подготовки аналитических выводов.

2. Планируемые результаты обучения по дисциплине (модулю).

Перечень формируемых результатов освоения образовательной программы (компетенций) в результате обучения по дисциплине (модулю):

МО-2 - Способен проектировать и внедрять системы автоматизации жизненного цикла машинного обучения (MLOps);

ППК-ДА2 - Способен подготавливать и обрабатывать данные для аналитических исследований;

ППК-Р1 - Способен разрабатывать и отлаживать программный код;

ППК-Р6 - Способен участвовать в промышленной разработке программного обеспечения.

Обучение по дисциплине (модулю) предполагает, что по его результатам обучающийся будет:

Знать:

- основные возможности языка Python для анализа данных и используемые библиотеки;
- методы обработки, преобразования, группировки и анализа табличных данных;

- принципы работы с пространственными (географическими) данными и средствами их визуализации.

Уметь:

- выполнять загрузку, очистку и подготовку данных из различных источников;
- использовать библиотеки Pandas, Polars и GeoPandas для решения аналитических задач;
- проводить исследовательский анализ данных и интерпретировать полученные результаты.

Владеть:

- навыками анализа данных с использованием современных библиотек Python;
- навыками построения аналитических отчетов и визуализации результатов;
- навыками обработки пространственных данных и выполнения базовой геоаналитики.

3. Объем дисциплины (модуля).

3.1. Общая трудоемкость дисциплины (модуля).

Общая трудоемкость дисциплины (модуля) составляет 2 з.е. (72 академических часа(ов)).

3.2. Объем дисциплины (модуля) в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении учебных занятий:

Тип учебных занятий	Количество часов	
	Всего	Семестр №1
Контактная работа при проведении учебных занятий (всего):	32	32
В том числе:		
Занятия семинарского типа	32	32

3.3. Объем дисциплины (модуля) в форме самостоятельной работы обучающихся, а также в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении промежуточной аттестации составляет 40 академических часа (ов).

3.4. При обучении по индивидуальному учебному плану, в том числе при ускоренном обучении, объем дисциплины (модуля) может быть реализован полностью в форме самостоятельной работы обучающихся, а также в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении промежуточной аттестации.

4. Содержание дисциплины (модуля).

4.1. Занятия лекционного типа.

Не предусмотрено учебным планом

4.2. Занятия семинарского типа.

Лабораторные работы

№ п/п	Наименование лабораторных работ / краткое содержание
1	Тема 1. Введение в Python для анализа данных. Рассматриваемые вопросы: - знакомство с Python как инструментом анализа данных; - установка и настройка среды разработки; - работа с Jupyter Notebook; - установка необходимых библиотек: pandas, polars; - обзор форматов данных для анализа: csv, tsv, parquet, json, geojson.
2	Тема 2. Основы работы с данными в Python. Рассматриваемые вопросы: - группировка данных, - работа с датой, - формирование временного ряда, - работа с текстом, re и библиотеки nltk.
3	Тема 3. Загрузка и первичный анализ данных из различных источников. Рассматриваемые вопросы: - чтение данных из CSV, Excel, JSON, Parquet; - работа с параметрами загрузки (кодировка, разделитель, пропуск строк); - первичная оценка объема данных и типов столбцов; - сохранение обработанных данных в различных форматах; - знакомство с реальными датасетами.
4	Тема 4. Структуры данных в Pandas: Series и DataFrame. Рассматриваемые вопросы: - создание Series и DataFrame из различных источников; - индексация, выборка и фильтрация строк и столбцов (loc, iloc, условиям); - сортировка и ранжирование данных; - работа с иерархической индексацией (MultiIndex); - применение пользовательских функций к данным (apply, map).
5	Тема 5. Очистка и подготовка данных для анализа. Рассматриваемые вопросы: - выявление и обработка пропущенных значений;

№ п/п	Наименование лабораторных работ / краткое содержание
	<ul style="list-style-type: none"> - удаление дубликатов и аномалий; - приведение типов данных; - переименование столбцов и работа с именами; - создание новых признаков на основе.
6	<p>Тема 6. Преобразование и объединение табличных данных.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - объединение данных из нескольких таблиц (merge, join, concat); - работа с разными типами соединений (INNER, LEFT, RIGHT, OUTER); - преобразование структуры таблиц (melt, pivot, stack/unstack); - работа со строковыми данными; - применение оконных функций.
7	<p>Тема 7. Группировка и агрегация данных.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - группировка данных по одному и нескольким признакам; - применение агрегирующих функций (sum, mean, count, min, max); - использование пользовательских агрегаций (agg, transform); - построение сводных таблиц; - многоуровневая группировка и анализ иерархических данных.
8	<p>Тема 8. Исследовательский анализ данных (EDA).</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - расчёт описательных статистик; - анализ распределений и выявление выбросов; - корреляционный анализ (corr, тепловые карты); - проверка статистических гипотез; - формулирование аналитических гипотез и выводов на основе данных.
9	<p>Тема 9. Визуализация данных в Matplotlib и Seaborn.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - построение базовых графиков (линейные, столбчатые, точечные); - визуализация распределений; - визуализация зависимостей; - оформление графиков (заголовки, легенды, подписи, сетка); - создание составных графиков.
10	<p>Тема 10. Интерактивная визуализация и дашборды.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - библиотека Plotly: построение интерактивных графиков; - создание панелей управления; - визуализация временных рядов с настройкой масштаба; - подготовка графиков для презентаций и отчетов; - практический кейс: построение дашборда по транспортным данным
11	<p>Тема 11. Работа с большими данными: библиотека Polars.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - особенности Polars: архитектура, работа с памятью, многопоточность; - создание DataFrame и чтение больших файлов; - фильтрация, выборка и преобразование данных в Polars; - работа с выражениями и контекстами; - сравнение синтаксиса Pandas и Polars.
12	<p>Тема 12. Оптимизация вычислений в Polars.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - работа с LazyFrame и ленивыми вычислениями; - агрегация и группировка в Polars;

№ п/п	Наименование лабораторных работ / краткое содержание
	<ul style="list-style-type: none"> - оптимизация запросов и план выполнения; - работа с большими объёмами данных (> 10 ГБ); - интеграция Polars с Pandas и другими библиотеками.
13	<p>Тема 13. Введение в геоаналитику: GeoPandas.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - понятие пространственных данных и геометрических объектов (точки, линии, полигоны); - чтение геоданных из OpenStreetMap; - работа с системами координат (CRS) и их преобразование; - визуализация геоданных, наложение на карту; - использование фоновых карт.
14	<p>Тема 14. Пространственный анализ данных.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - пространственная фильтрация; - пространственные соединения; - построение тематических карт; - агрегация геоданных; - практический кейс: анализ транспортной доступности или ДТП.
15	<p>Тема 15. Комплексный анализ данных (сквозной кейс).</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - интеграция работы с табличными, временными и пространственными данными; - загрузка, очистка и подготовка данных из нескольких источников; - исследовательский анализ и визуализация; - формулирование аналитических выводов и рекомендаций; - подготовка отчёта в Jupyter Notebook.
16	<p>Тема 16. Защита аналитического мини-проекта.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - постановка задачи на основе реальных данных (транспорт / городская среда); - самостоятельный выбор инструментов и методов анализа; - выполнение полного цикла обработки данных; - подготовка итогового отчёта и презентации; - публичное представление результатов с ответами на вопросы.

4.3. Самостоятельная работа обучающихся.

№ п/п	Вид самостоятельной работы
1	Работа с учебной литературой
2	Подготовка к практическим занятиям
3	Подготовка к промежуточной аттестации.
4	Подготовка к текущему контролю.
5	Подготовка к промежуточной аттестации.
6	Подготовка к текущему контролю.

5. Перечень изданий, которые рекомендуется использовать при освоении дисциплины (модуля).

№ п/п	Библиографическое описание	Место доступа
1	Сергеева, О. А. Программирование на Python : учебно-методическое пособие / О. А. Сергеева. — Кемерово : КемГУ, 2024. — 157 с. — ISBN 978-5-8353-3123-9	https://e.lanbook.com/book/420758
2	Янцев, В. В. Web-программирование на Python : учебное пособие для вузов / В. В. Янцев. — 3-е изд., перераб. — Санкт-Петербург : Лань, 2024. — 180 с. — ISBN 978-5-507-48364-8	https://e.lanbook.com/book/392993

6. Перечень современных профессиональных баз данных и информационных справочных систем, которые могут использоваться при освоении дисциплины (модуля).

<https://habr.com/ru> - база знаний в виде статей, обзоров

<http://library.miit.ru/> - электронно-библиотечная система научно-технической библиотеки МИИТ

<https://superset.apache.org/user-docs/intro/#documentation> – документация Apache Superset

<https://www.postgresql.org/> - официальная страница PostgreSQL

<https://yandex.cloud/ru/docs/datalens/> - официальная документация Yandex Datalens

<https://docs.docker.com/> - официальная документация Docker

<https://www.metabase.com/docs/latest/> - официальная документация Metabase

<https://tableau.pro/> - официальный сайт Tableau с лучшими практиками проектирования дашбордов

<https://revealthedata.com/blog/> - Блог о визуализации данных и BI-системах

7. Перечень лицензионного и свободно распространяемого программного обеспечения, в том числе отечественного производства, необходимого для освоения дисциплины (модуля).

MS Office (Excel, Word, PowerPoint)

Apache Superset

Yandex Datalens

Metabase

PostgreSQL

Docker

Apache Airflow

Браузер Chrome / Яндекс браузер
Draw.io

8. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю).

Компьютер преподавателя с доступом к сети Интернет

Компьютеры студентов с установленным Python-окружением и табличным процессором

Доступ к учебной базе событийных данных и экспериментальным выборкам

Доступ к PostgreSQL и BI-системе для визуализации результатов

Доступ к облачной виртуальной машине или Docker-контейнерам для воспроизводимого анализа

Проектор и экран для проектора

Маркерная доска

9. Форма промежуточной аттестации:

Экзамен в 1 семестре.

10. Оценочные материалы.

Оценочные материалы, применяемые при проведении промежуточной аттестации, разрабатываются в соответствии с локальным нормативным актом РУТ (МИИТ).

Авторы:

руководитель образовательной
программы

О.Б. Проневич

Согласовано:

Директор

Д.В. Паринов

Директор ЦС IT - программ ВИШ

О.Б. Проневич

Председатель учебно-методической
комиссии

Д.В. Паринов