

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ**

федеральное государственное автономное образовательное учреждение
высшего образования

**«Московский физико-технический институт
(национальный исследовательский университет)»
(МФТИ, Физтех)**

УТВЕРЖДАЮ

Ректор МФТИ

д-р физ.-мат. наук, профессор

Д. В. Ливанов

2023 г.



**Дополнительная общеобразовательная общеразвивающая программа
«Анализ данных на Python»**

Москва 2023

1. Общая характеристика программы

1.1 В обществе всё большее значение приобретает умение человека использовать компьютер не на пользовательском уровне, а на уровне начинающего программиста. В обязательном школьном курсе информатики программирование представлено на уровне, достаточном для прохождения экзамена, но не предполагает овладение практическими навыками применения языка. Следствием этого - формальное восприятие обучающимися основ современного программирования и неумение применять полученные знания на практике.

Программа «Анализ данных на Python» имеет техническую направленность, в её основу заложены принципы модульности и практической направленности, что обеспечит вариативность обучения. Содержание учебных модулей предполагает детальное изучение алгоритмизации, реализацию межпредметных связей, организацию проектной и исследовательской деятельности обучающихся.

Анализ данных применяется во многих областях науки и бизнеса для выявления важных закономерностей. Одним из наиболее эффективных инструментов для работы с данными являются языки программирования, в частности язык Python.

В рамках данного курса рассматриваются теоретические и практические аспекты, которые необходимо знать для решения задач первичного анализа данных. Пройдя этот курс, вы узнаете, какие среды разработки Python наиболее подходят для задачи анализа, поработаете с циклами и коллекциями, а также познакомитесь с некоторыми популярными библиотеками языка Python для обработки и визуализации данных.

Цель программы - помочь учащимся в освоении основных конструкций и библиотек анализа данных для языка Python, а также сформировать компетенции для проведения первичного анализа данных с помощью этих инструментов.

1.2. Категории слушателей, на обучение которых рассчитана программа дополнительного образования (далее – программа): учащиеся 8-11 классов.

1.3. Нормативный срок освоения программы – 144 академических часа.

1.4. Форма обучения: очная с применением дистанционных образовательных технологий.

Основной вид занятий – комбинированный, сочетающий в себе элементы теории и практики. Единицей учебного процесса является блок уроков (модуль). Каждый модуль охватывает отдельную информационную технологию или её часть. Внутри модулей разбивка по времени изучения производится педагогом самостоятельно, но с учётом рекомендованного календарно-тематического плана. Темп изучения отдельных разделов блока

определяется субъективными и объективными факторами и необходимостью повторения.

2. Планируемые результаты обучения

Личностные результаты

- формирование ответственного отношения к учению, готовности и способности, обучающихся к саморазвитию и самообразованию на основе мотивации к обучению и познанию, осознанному выбору и построению дальнейшей индивидуальной траектории образования на базе ориентировки в мире профессий и профессиональных предпочтений, с учётом устойчивых познавательных интересов;
- формирование целостного мировоззрения, соответствующего современному уровню развития науки и общественной практики, учитывающего социальное, культурное, языковое, духовное многообразие современного мира;
- формирование коммуникативной компетентности в общении и сотрудничестве со сверстниками, детьми старшего и младшего возраста, взрослыми в процессе образовательной, общественно полезной, учебно-исследовательской, творческой и других видов деятельности.
- повышение уровня самооценки благодаря реализованным проектам.

Метапредметные результаты

- умение самостоятельно определять цели своего обучения, ставить и формулировать для себя новые задачи в учёбе и познавательной деятельности, развивать мотивы и интересы своей познавательной деятельности;
- умение самостоятельно планировать пути достижения целей, в том числе альтернативные, осознанно выбирать наиболее эффективные способы решения учебных и познавательных задач;
- умение соотносить свои действия с планируемыми результатами, осуществлять контроль своей деятельности в процессе достижения результата, определять способы действий в рамках предложенных условий и требований, корректировать свои действия в соответствии с изменяющейся ситуацией;
- умение оценивать правильность выполнения учебной задачи, собственные возможности её решения;
- владение основами самоконтроля, самооценки, принятия решений и осуществления осознанного выбора в учебной и познавательной деятельности;
- умение определять понятия, создавать обобщения, устанавливать аналогии, классифицировать, самостоятельно выбирать основания и критерии для классификации, устанавливать причинно-следственные связи, строить логическое рассуждение, умозаключение (индуктивное, дедуктивное и по аналогии) и делать выводы;

- умение создавать, применять и преобразовывать знаки и символы, модели и схемы для решения учебных и познавательных задач;
- умение организовывать учебное сотрудничество и совместную деятельность с учителем и сверстниками; работать индивидуально и в группе: находить общее решение и разрешать конфликты на основе согласования позиций и учёта интересов; формулировать, аргументировать и отстаивать своё мнение;
- формирование и развитие компетентности в области использования информационно коммуникационных технологий.

Предметные результаты

- знание основ работы в среде разработки JupyterLab;
- знание основ программирования на языке Python;
- знание базовых классов, библиотек и функций Python для сбора, анализа и визуализации данных.
- умение применять инструменты библиотек NumPy и Pandas при работе с данными для вычисления простых метрик и анализа взаимосвязей;
- умение визуализировать данные при помощи инструментов Python и библиотеки Matplotlib;
- умение собирать данные из открытых источников при помощи инструментов библиотеки Beautiful Soup и сохранять их в виде файлов.

3. Структура программы Учебный план

№	Тема (модуль)	Кол-во часов	В том числе		
			Аудит. занятия	Самост. работа	Промежуточная аттестация
	Модуль 1. Базовые конструкции	36	18	14	4
1.1	Установка Python и JupyterLab. Понятие кода, интерпретатора, программы. Исполнение кода и отладка	5	3	2	
1.2	Переменные. Базовые типы данных. Основные операторы. Целочисленная арифметика	5	3	2	
1.3	Конструкция ветвление. Условный оператор. Каскадный условный оператор. Вложенные условия	5	3	2	
1.4	Контрольная работа	5	3	2	
1.5	Цикл for. Организация циклов	6	3	3	
1.6	Цикл while. Организация разветвлений в цикле	6	3	3	
	Промежуточная аттестация	4			4
	Модуль 2. Знакомство с коллекциями	36	18	14	4
2.1	Строки. Срезы. Методы строк	5	3	2	
2.2	Списки. Методы списков. Списочные выражения. Кортежи	5	3	2	
2.3	Функции. Области видимости переменных. Функции с переменным числом аргументов. Значения по умолчанию	5	3	2	
2.4	Контрольная работа	5	3	2	
2.5	Лямбда функции. Сортировка с параметром key	6	3	3	
2.6	Функции высшего порядка: map, filter, zip	6	3	3	
	Промежуточная аттестация	4			4
	Модуль 3. Работа с числовыми данными NumPy	36	18	14	4
3.1	Множества	5	3	2	
3.2	Словари	5	3	2	
3.3	Работа с файлами. Чтение и запись данных	5	3	2	
3.4	Особенные типы данных в NumPy. Работа с векторами и матрицами	5	3	2	
3.5	Вычисление главных статистических метрик с помощью NumPy (среднее, медиана, мода, дисперсия)	6	3	3	
3.6	Визуализация данных с Matplotlib	6	3	3	
	Промежуточная аттестация	4			4
	Модуль 4. Анализ данных с помощью библиотеки Pandas	36	18	14	4
4.1	Работа с табличными данными Pandas	5	3	2	

4.2	Сбор данных в сети Интернет. Библиотека Beautiful Soup	5	3	2	
4.3	Анализ взаимосвязей и описательная статистика	5	3	2	
4.4	Сводные таблицы	5	3	2	
4.5	Агрегирование данных и групповые операции	6	3	3	
4.6	Подготовка к итоговому проекту	6	3	3	4
Итого		144	72	56	16

4. Содержание программы

4.1. Содержание программы по модулям

Модуль 1. Базовые конструкции

Тема 1.1. Установка Python и JupyterLab. Понятие кода, интерпретатора, программы. Исполнение кода и отладка

Решение задач типа: Установите на свою рабочую станцию Python и JupyterLab подходящим для вашей системы пакетным менеджером. Выведите на экран сумму двух чисел для проверки работоспособности установленного интерпретатора.

Тема 1.2. Переменные. Базовые типы данных. Основные операторы. Целочисленная арифметика

Решение задач типа: Вычислите дробь $466/115$ с точностью до 6 знака после запятой. Вычислите остаток от деления числа 2^{2048} на 3.

Тема 1.3. Конструкция ветвление. Условный оператор. Каскадный условный оператор. Вложенные условия

Решение задач типа: Вычислите и выведите на экран евклидово расстояние между подаваемыми на вход координаты точек.

Тема 1.4. Контрольная работа

Решение задач типа: Реализуйте математическую функцию $\text{sign}(x)$. Напишите программу, которая принимает на вход номер месяца вашего рождения, и выводит строку «Вы родились в НАЗВАНИЕ_МЕСЯЦА».

Тема 1.5. Цикл for. Организация циклов

Решение задач типа: Получив на вход два числа, выведите на экран все простые числа, находящиеся между ними.

Тема 1.6. Цикл while. Организация разветвлений в цикле

Решение задач типа: Вычислите последовательность Фибоначчи для подаваемого на вход числа. Вычислите факториал подаваемого на вход числа. Выведите на экран сумму натуральных чисел от 1 до подаваемого на вход числа.

Модуль 2. Знакомство с коллекциями

Тема 2.1. Строки. Срезы. Методы строк

Решение задач типа:

Для строки '1234567890abcdefghij' выведите все символы с четными номерами.

В подаваемой на вход строке замените все цифры на точки.

Тема 2.2. Списки. Методы списков. Списочные выражения. Кортежи

Решение задач типа:

Дан числовой список элементов. Составьте из него новый список, в который попадут только элементы меньше 100.

Напишите программу, которая считывает список целых чисел и выводит на экран кортеж из уникальных элементов этого списка, упорядоченных по возрастанию.

Тема 2.3. Функции. Области видимости переменных. Функции с переменным числом аргументов. Значения по умолчанию

Решение задач типа:

Напишите функцию проверки четности числа.

Напишите функцию, удваивающую вхождение введенного символа в тексте.

Тема 2.4. Контрольная работа

Решение задач типа: Представьте в виде списка строк текст 'One;two;three;four;five', разделив его по символу ';' В переменной data хранится список из фиксированного набора описаний. Отобразите в новой строке все уникальные элементы data и их количество по убыванию.

Тема 2.5. Лямбда функции. Сортировка с параметром key

Решение задач типа:

С помощью лямбда-функции отсортируйте список чисел по последней цифре.

Тема 2.6. Функции высшего порядка: map, filter, zip

Решение задач типа:

Напишите лямбда-функцию для удвоения всех элементов списка.

Дан список точек из кортежей из двух чисел вида [(3, -2), (7, 1), (0, 4)]. С помощью лямбда-функции отсортируйте список по возрастанию расстояния от начала координат до точки.

Напишите лямбда-функцию для создания списка положительных чисел из списка вида [19, -8, 4, 0, -2, 15].

Модуль 3. Работа с числовыми данными NumPy

Тема 3.1. Множества

Решение задач типа:

На входе функция получает строку или список чисел. Преобразуйте их в множество и посчитайте его мощность.

Тема 3.2. Словари

Решение задач типа:

Создайте словарь, в котором ключами будут числа от 1 до 10, а значениями эти же числа, возведенные в куб.

Дан словарь, состоящий из пар слов, являющихся синонимами. Для подаваемого на вход слова определите его синоним.

Тема 3.3. Работа с файлами. Чтение и запись данных

Решение задач типа:

Считайте файл целиком при помощи метода `read()`. Выведите в обратном порядке содержимое всего файла.

Тема 3.4. Особенности типы данных в NumPy. Работа с векторами и матрицами

Решение задач типа:

На вход подается numpy массив `a` и целое `b`. Возвратите numpy массив, состоящий из индексов всех вхождений числа `b` в массив `a`.

Создайте трехмерный массив numpy $3 \times 3 \times 3$ со случайными значениями.

Тема 3.5. Вычисление главных статистических метрик с помощью NumPy (среднее, медиана, мода, дисперсия)

Решение задач типа:

Пусть `a` - непустой двумерный numpy массив. Найдите медиану и дисперсию по колонкам.

Считайте из файла массив. Вычислите среднее и среднеквадратичное отклонение элементов.

Тема 3.6. Визуализация данных с Matplotlib

Решение задач типа:

Постройте на общих осях графики $f(x) = x$ и $f(x) = 1/x$.

На вход подается имя файла с построчными координатами точек. Визуализируйте эти точки на графике, постепенно меняя с каждой точкой градиент цвета от красного к черному.

Модуль 4. Анализ данных с помощью библиотеки pandas

Тема 4.1. Работа с табличными данными Pandas

Решение задач типа:

Считайте из csv файла данные, положите их датафрейм Pandas и выведите на экран первые 10 строк.

Дан словарь, содержащий фамилии учеников и их оценку за экзамен. Преобразуйте словарь в датафрейм и посчитайте в нём значение среднего балла.

Тема 4.2. Сбор данных в сети Интернет. Библиотека BeautifulSoup

Решение задач типа:

Дана простая web страница с заголовком и несколькими параграфами. С помощью библиотеки BeautifulSoup выделите из страницы содержимое параграфов и запишите их в текстовый файл.

Тема 4.3. Анализ взаимосвязей и описательная статистика

Решение задач типа:

Дан фрейм данных с числовыми столбцами. С помощью функции `pd.plotting.scatter_matrix()` создайте матрицу точечных графиков для всех столбцов.

Тема 4.4. Сводные таблицы

Решение задач типа:

Дан csv файл вида «Врач,Диагноз,Продолжительность». С помощью сводной таблицы Pandas посчитайте, сколько раз какой врач ставил каждый из диагнозов.

Тема 4.5. Агрегирование данных и групповые операции

Решение задач типа:

Дан датафрейм со сводной таблицей животных в зоопарке, содержащей номер, вид и норму питьевой воды в сутки. С помощью методов агрегации посчитайте количество видов животных и общий объем требуемой воды.

Тема 4.6. Подготовка к итоговому проекту

Разбор содержания тем проектов:

Сбор отзывов с сайта доставки готовой еды и поиск зависимости оценки отзыва от сезона / времени дня.

Визуализация пиков заболеваемости Covid-19 на основе открытых датасетов.

Реализация Contact Book для хранения разнообразных контактных данных, на основе csv файла с поддержкой функции поиска по разным полям.

Реализация игры «Камень, Ножницы, Бумага» с хранением и выводом на экран статистики всех когда-либо сыгранных матчей.

Автоматизированный сбор данных о цене какого-либо товаре в течение месяца, визуализация динамики цены на графике и подсчет средних значений цены.

4.2. Список рекомендуемой литературы

Лутц М. Изучаем Python. Том 1 – Издательство «Диалектика» –2019 – 832с.

Рамальо Л. Python. К вершинам мастерства – Издательство «ДМК Пресс» – 2016 – 768с.

5. Материально-технические условия реализации программы

Требования к помещению:

- помещение для занятий, отвечающие требованиям СанПин для учреждений дополнительного образования;
- качественное освещение;
- столы, стулья по количеству обучающихся и 1 рабочим местом для педагога.

Информационное обеспечение:

- операционная система (желательно Windows);

- браузеры: Yandex Browser, Chrome, Chrome Mobile, Firefox, Opera, Safari, Mobile Safari, Edge, Python;
- среда Wing IDE 101 (версии 6 или выше);
- среда PyCharm Community Edition.

Требования к оборудованию

Таблица 4

Наименование специализированных аудиторий, кабинетов, лабораторий	Вид занятий	Наименование оборудования, программного обеспечения
Аудитория с доступом в Интернет	Аудиторные занятия	Компьютер, Visual Studio, проектор, видеочамера, доступ в Интернет
Аудитория с доступом в Интернет	Самостоятельная работа	Персональные компьютеры/ноутбуки, Visual Studio, доступ в Интернет

6. Оценка качества освоения программ

Оценка качества освоения программы осуществляется в процессе промежуточной аттестации.

Формы и методы промежуточного контроля представлены в таблице 5.

Таблица 5

Наименование модулей	Основные показатели оценки	Формы и методы контроля и оценки
Модуль 1. Базовые конструкции	Установленное количество выполненных заданий	Самостоятельное выполнение заданий
Модуль 2. Знакомство с коллекциями	Установленное количество выполненных заданий	Самостоятельное выполнение заданий
Модуль 3. Работа с числовыми данными NumPy (Numerical Python)	Установленное количество выполненных заданий	Самостоятельное выполнение заданий
Модуль 4. Анализ данных с помощью библиотеки pandas	Установленное количество выполненных заданий	Самостоятельное выполнение заданий

Примерные задания для проведения промежуточной аттестации по модулям

Модуль 1. Базовые конструкции

- Выведите на экран сумму двух чисел для проверки работоспособности установленного интерпретатора.
- Вычислите остаток от деления числа 2^{2048} на 3.
- Вычислите и выведите на экран евклидово расстояние между подаваемыми на вход координаты точек
- Выведите на экран сумму натуральных чисел от 1 до подаваемого на вход числа.

Модуль 2. Знакомство с коллекциями

- Для строки '1234567890abcdefghij' выведите все символы с четными номерами.
- Дан числовой список элементов. Составьте из него новый список, в который попадут только элементы меньше 100
- Напишите функцию проверки четности числа.
- Представьте в виде списка строк текст 'One;two;three;four;five', разделив его по символу ';'.
- Напишите лямбда-функцию для создания списка положительных чисел из списка вида [19, -8, 4, 0, -2, 15].

Модуль 3. Работа с числовыми данными NumPy (Numerical Python)

- На входе функция получает строку или список чисел. Преобразуйте их в множество и посчитайте его мощность.
- Считайте файл целиком при помощи метода read(). Выведите в обратном порядке содержимое всего файла.
- Считайте из файла массив. Вычислите среднее и среднеквадратичное отклонение элементов.
- Постройте на общих осях графики $f(x) = x$ и $f(x) = 1/x$.

Модуль 4. Анализ данных с помощью библиотеки pandas

- Считайте из csv файла данные, положите их датафрейм Pandas и выведите на экран первые 10 строк.
- Дан фрейм данных с числовыми столбцами. С помощью функции pd.plotting.scatter_matrix() создайте матрицу точечных графиков для всех столбцов.
- Дан датафрейм со сводной таблицей животных в зоопарке, содержащей номер, вид и норму питьевой воды в сутки. С помощью методов агрегации посчитайте количество видов животных и общий объем требуемой воды.

7. Составители программы

Малеев Алексей Викторович, директор Центра развития ИТ-образования МФТИ

Куклин Евгений Юрьевич, старший преподаватель Уральского Федерального университета, м.н.с. Института математики и механики УрО РАН

Созыкин Андрей Владимирович, к. т. н., заместитель директора Центра развития ИТ-образования МФТИ

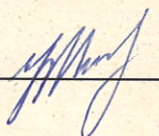
Мартемьянов Роман Юрьевич, заместитель директора Центра развития ИТ-образования МФТИ

Даревский Сергей Григорьевич, руководитель группы Центра развития ИТ-образования МФТИ

Сырцова Елена Леонидовна, PhD, к. пед. н., доцент, руководитель проектов Центра развития ИТ-образования МФТИ

Токмакова Ольга Викторовна, PhD, к. пед. н., доцент, специалист по учебно-методической работе Центра развития ИТ-образования МФТИ

Согласовано
Ведущий специалист отдела
сопровождения образовательных
программ

 Ж.И. Зубцова

Согласовано
Директор ЦРИТО

 А. В. Малеев

Федеральное государственное автономное образовательное учреждение
высшего образования
«Московский физико-технический институт
(национальный исследовательский университет)»

ВЫПИСКА ИЗ ПРОТОКОЛА № 8
заседания учебно-методического совета от 29 марта 2023 года.

ПОВЕСТКА:

Рассмотрение дополнительных общеобразовательных и профессиональных программ.

Проректор по учебной работе А. А. Воронов.

СЛУШАЛИ: заместителя директора Центра дополнительного, дополнительного профессионального и онлайн-образования «ПУСК» Рыбакову А.И. о представлении дополнительных общеобразовательных и профессиональных программ. (Центр «ПУСК» МФТИ).

ПОСТАНОВИЛИ:

Рекомендовать к утверждению в установленном порядке дополнительную общеобразовательную программу «Анализ данных на Python».

Решение принято единогласно.

Форма проведения заседания: заочная.

Председатель УМС МФТИ

Ученый секретарь УМС МФТИ



А.А. Воронов

М.В. Березникова