

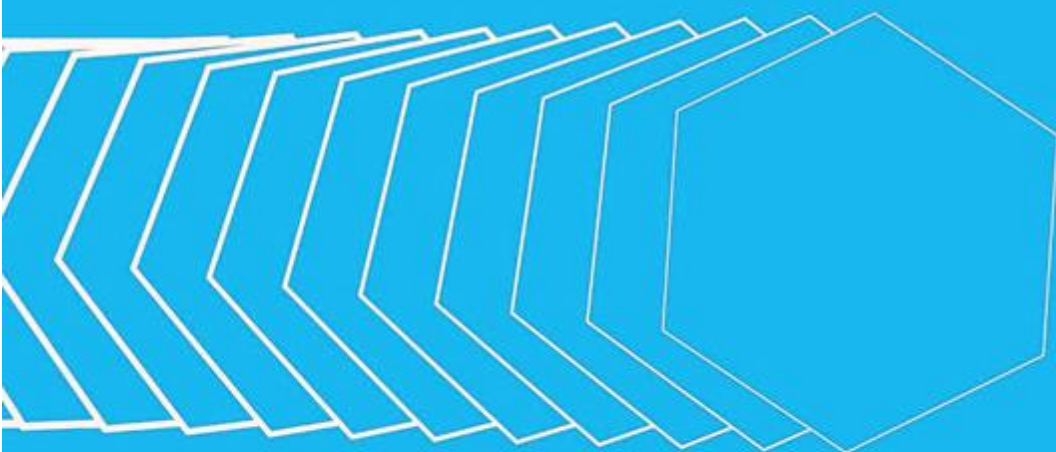


**ФГБОУ ДО
ФЦДО**

Методическая разработка

мастер-класса

«Визуализация данных в Plotly»



Направление:
IT-технологии

Автор: Сосулин Павел,
педагог дополнитель-
ного образования тех-
нической направлен-
ности

**2022
Москва**

Визуализации данных в Plotly

Цель мастер-класса:

Расширение представлений о технических возможностях инструментов визуализации данных языка программирования Python.

Задачи:

- знакомство с инструментами визуализации данных в виде библиотек "Matplotlib" и "Plotly".
- развитие навыков сравнительного анализ возможностей библиотек "Matplotlib" и "Plotly" на конкретном практическом примере в виде задачи оптимизации параметров процесса электронно-лучевой сварки тонкостенных конструкций аэрокосмического назначения.

Основная идея: познакомить с двумя инструментами визуализации данных языка программирования Python в виде библиотек "Matplotlib" и "Plotly" и сравнительным анализом полученных результатов при использовании каждой из них. В процессе знакомства и сравнительного анализа должно сформироваться положительное впечатление об использовании библиотеки "Plotly" в качестве инструмента визуализации данных.

ПАСПОРТ ПРАКТИКИ	УСЛОВИЯ РЕАЛИЗАЦИИ		
	Время	Форма	Методы
	2 встречи по 90 минут	Работа в малых группах	Репродуктивные, исследовательские
	КАТЕГОРИЯ ОБУЧАЮЩИХСЯ		
	Возраст	Особенности	Кол-во участников в рабочей группе
	13 – 17 лет	Владение основами языка программирования Python	До 36 человек
	РЕСУРСЫ		
	Оборудование и материалы	Базовые знания из других областей	Уровень сложности и Место в структуре курса
	Компьютер/ ноутбук с установленным ПО Anaconda3 2022.05 (Python 3.9.12)	Основы алгоритмизации и программирования	Средний уровень сложности.
	ОБРАЗОВАТЕЛЬНЫЕ РЕЗУЛЬТАТЫ		
	Артефакт (продукт)	Hard-skills, предметные навыки	Soft-Skills, метапредметные навыки
	Визуальные графические представления данных: графики, столбчатые диаграммы и т.д., с элементами интерактивности	<ul style="list-style-type: none"> – знания об основах программирования на языке Python – умение подключать и использовать библиотеку Plotly – основы анализа данных – знания о профессии «аналитик-данных» (Data Scientist) 	<ul style="list-style-type: none"> – навыки работы в малых группах, – развитие абстрактно-логического и аналитического видов мышления, – умение делать выводы на основе полученных результатов.
	ОПЫТ РЕАЛИЗАЦИИ		
	Инструкции для педагога	Открытая информация о результатах	Участие в конкурсах
	есть	нет	нет

ЗНАКОМСТВО С БИБЛИОТЕКОЙ PLOTLY

1 ВСТРЕЧА – 90 минут.

«Знакомство с библиотекой Plotly. Отличия Plotly от Matplotlib»

Планируемые результаты:

- умение использовать на практике язык программирования Python и библиотеку "Matplotlib" для визуализации данных
- умение использовать на практике язык программирования Python и библиотеку "Plotly" для визуализации данных
- знание отличий библиотек "Matplotlib" и "Plotly"

Оборудование для проведения мастер-класса:

- компьютер/ноутбук с установленным ПО Anaconda3 2022.05(Python 3.9.12)
- проекционное оборудование
- презентация и файл с программным кодом (Jupyter Notebook)

ЭТАП 1 ПОДГОТОВИТЕЛЬНЫЙ

Рекомендованное время: 10 минут

Педагог информирует участников о том, что сегодня в рамках этого мастер-класса они познакомятся с такими библиотеками представления данных в Python, как Plotly и Matplotlib и узнают на практическом примере их сходства и отличия друг от друга. Сравнение возможностей двух библиотек будет проведено на данных, полученных при решении задачи математического и алгоритмического обеспечения оптимизации процесса электронно-лучевой сварки тонкостенных конструкций аэрокосмического назначения. Полученные результаты очень сильно зависят от начальных параметров, которые варьируются в широком диапазоне, поэтому для анализа полученных данных целесообразно использовать графическое представление переменных и целевой функции. Программная реализация такой задачи разбита на два модуля:

- Модуль 1 содержит функции T_1 , T_2 и ψ
- Модуль 2 выполняет расчёт распределения температур ($T_1 + T_2$) по заданным параметрам с последующим выводом в отдельном окне с возможностью сохранения графика в файл.

Педагог уточняет, что для работы понадобится программа Jupyter Notebook, ссылка на процесс установки которой показана на слайде [презентации](#) (ПО Jupyter Notebook должно быть установлено на рабочие места школьников заранее. Слайд с ссылкой на процесс установки носит информативный характер (на случай, если учащиеся захотят повторить его у себя на домашних устройствах).

ЭТАП 2 ОСНОВНОЕ СОДЕРЖАНИЕ ДЕЯТЕЛЬНОСТИ

Рекомендованное время: 40 минут

ШАГ 1. Установка необходимых для работы библиотек

Для начала работы необходимо запустить Jupyter Notebook, создать в нём новый файл и добавить строки кода, указанные на слайде. Для задания точки с координатами X, Y, Z разработаем класс, код которого указан на этом слайде

Примечание: вместо последовательного набора программного кода за наставником учащимся можно раздать файл с готовым программным кодом мастер-класса, по которому они будут последовательно выполнять каждую ячейку и видеть результат выполнения.

ШАГ 2. Разработка вспомогательного класса для задания точки с координатами

[Plotly vs Matplotlib.ipynb](#) - файл с готовым программным кодом мастер-класса.

Дополнительно к файлу с программным кодом потребуются следующие файлы с исходными данными: [module 1 input.csv](#) – входные данные для Модуля 1 [module 2 input.csv](#) – входные данные для Модуля 2.

ШАГ 3. Создание вспомогательных функций

Педагог объясняет детям о необходимости разработки функции для импорта необходимых настроек из конфигурационного csv - файла. Функция принимает всего лишь один аргумент (указывается в скобках после названия функции) – это число, которое является номером модуля. Код функции представлен в презентации (приложение 1, п.1). Педагог обращает внимание детей на то, что если по каким-то причинам настройки из файла принять не удастся, то нужно обработать такую исключительную ситуацию и показать пользователю сообщение об ошибке (смотрите код презентации (приложение 1, п.1)).

Для анализа состояния температурного поля при воздействии быстро движущегося точечного источника излучения необходимо разработать функцию T_1, в которой реализовать вычисление по заранее известной формуле. Код функции представлен в презентации (приложение 1, п. 1).

Для анализа состояния температурного поля при воздействии быстро движущегося источника линейного излучения необходимо разработать функцию T_2, в которой нужно реализовать вычисление по заранее известной формуле. Код функции представлен в презентации (приложение 1, п. 1).

Нам потребуется специализированная функция PSI(Пси). Давайте разработаем и её. Код функции представлен на этом слайде.

Примечание: вместо последовательного набора программного кода за наставником учащимся можно раздать файл с готовым программным кодом мастер-класса, по которому они будут последовательно выполнять каждую ячейку и видеть результат выполнения.

[Plotly vs Matplotlib.ipynb](#) - файл с готовым программным кодом мастер-класса.

Дополнительно к файлу с программным кодом потребуются следующие файлы с исходными данными: [module 1 input.csv](#) – входные данные для Модуля 1, [module 2 input.csv](#) – входные данные для Модуля 2

ШАГ 4. Задание параметров

Педагог информирует обучающихся о переходе к Модулю 2.

Педагог объясняет о необходимости задать необходимые параметры. Код этих параметров представлен в презентации (приложение 1, п.1). Педагог обращает внимание детей на то, что считанные из csv-файла при помощи функции импорта настроек (код которой был разработан ранее) помещаются в виде массива строк в переменную `work_configs`. Далее строки из этой переменной подставляются в необходимые параметры.

ШАГ 5. Основная функция

Педагог информирует детей о финальном этапе мастер-класса, и что осталось разработать код основной функции `main()`, в которой будет располагаться основной код. Педагог сообщает детям о создании и заполнении матрицы основных координат точки в пространстве X, Y, Z. Педагог рекомендует детям о необходимости поместить в переменную `tmp` одну из 4-х представленных для анализа функций и далее добавлении результата из переменной `tmp` в качестве координаты Z. Код функции `main()` представлен в презентации (приложение 1 п.1).

Педагог объясняет детям о построении графиков того, что получилось при помощи использования библиотеки Matplotlib и Plotly. Для педагог с детьми продолжает разрабатывать код функции `main()`. Программный код представлен в презентации (приложение 1 п.1).

ЭТАП 3. АНАЛИЗ РЕЗУЛЬТАТОВ

Рекомендованное время: 40 минут

Педагог информирует обучающимся, что после выполнения кода в функции `main()` были получены два графика, один из которых построен при помощи библиотеки Plotly, другой - библиотеки Matplotlib. Вид полученных графиков представлен в презентации (приложение 1, п.1). В отличие от графика, построенного при помощи библиотеки Matplotlib, график, построенный при помощи библиотеки Plotly обладает интерактивностью, его можно вращать, сохранять в качестве файла с изображением, приближать или отдалять отдельные его части.

Педагог объясняет детям, что в итоге на практическом примере при использовании специализированного ПО для анализа данных Jupyter Notebook, языка программирования Python и ряда библиотек, среди которых есть библиотеки Plotly и Matplotlib были получены идентичные друг другу графики. Однако график, построенный при помощи библиотеки Plotly, обладает большей интерактивностью по сравнению с Matplotlib.

Примечание: вместо последовательного набора программного кода за наставником учащимся можно раздать файл с готовым программным кодом мастер-класса, по которому они будут последовательно выполнять каждую ячейку и видеть результат выполнения.

[Plotly vs Matplotlib.ipynb](#) (приложение 1 п.2) - файл с готовым программным кодом мастер-класса.

Дополнительно к файлу с программным кодом потребуются следующие файлы с исходными данными: [module 1 input.csv](#) (приложение 1 п.3) – входные данные для Модуля 1, [module 2 input.csv](#) (приложение 1 п.4) – входные данные для Модуля 2.

«Визуализация данных в Python. Библиотекой Plotly»

2 ВСТРЕЧА – 90 минут.

«Визуализация данных в Python. Библиотека Plotly»

Планируемые результаты:

- умение применять диаграммы и графики библиотеки “Plotly” для отображения наглядности наборов анализируемых данных
- умение логически мыслить и делать определённые выводы, основанные на практических результатах
- практический опыт анализа данных на конкретном практическом примере

Оборудование для проведения мастер-класса:

- компьютер/ноутбук с установленным ПО Anaconda3 2022.05(Python 3.9.12)
- проекционное оборудование
- презентация и файл с программным кодом (Jupyter Notebook)

ЭТАП 1 ПОДГОТОВИТЕЛЬНЫЙ

Рекомендованное время: 10 минут.

Педагог приветствует и знакомит обучающихся с целью и задачами этого мастер-класса.

Педагог акцентирует внимание обучающихся на том, что в первую очередь для работы им понадобится установить программу Jupyter Notebook, ссылка на процесс установки которой показан в презентации (приложение 1 п.1), а также может понадобиться ряд библиотек таких как: Plotly, Cufflinks, Chart studio, Pyarrow и Fastparquet.

Педагог сообщает обучающимся о том, что информацию о библиотеке Plotly они смогут найти на официальном сайте по ссылке, указанной в презентации (приложение 1, п.5).

Педагог сообщает обучающимся о том, что, как и у любой другой библиотеки, написанной на языке программирования Python, у библиотеки “Plotly” имеется документация с примерами программного кода. Просмотреть документацию можно по ссылке, указанной в презентации (приложение 1, п.5).

Также при помощи библиотеки “Plotly” можно строить различные диаграммы. Некоторые из них отображены в презентации (приложение 1, п.5).

Педагог сообщает обучающимся, что в начале необходимо произвести импорт необходимых библиотек. Для этого следует запустить программу Jupyter Notebook и создать в ней новый документ, в первую ячейку которого поместить программный код, указанный в презентации (приложение 1, п.5).

Источником данных для анализа будут являться данные платформы социальной журналистики “Medium”, внешний вид и ссылка на которую - представлены в презентации (приложение 1, п.5).

Педагог сообщает обучающимся, что интерес будут представлять реакции пользователей платформы в виде так называемых «хлопков в ладоши» (по-английски, claps). Это будет означать, что то или иное содержимое нравится пользователям. На слайде представлены реакции пользователей на две статьи авторов на платформе Medium в виде «хлопков в ладоши».

Примечание:

ПО Jupyter Notebook должно быть установлено на рабочие места школьников заранее. Слайд с ссылкой на процесс установки носит информативный характер (на случай, если учащиеся захотят повторить его у себя на домашних устройствах)

Можно попросить детей самостоятельно открыть официальный сайт библиотеки "Plotly" в интернет-браузере на своих рабочих компьютерах/ноутбуках/смартфонах.

Адрес сайта: <https://plotly.com>

Адрес сайта с документацией: <https://plotly.com/python>

Вместо последовательного набора программного кода за наставником учащимся можно раздать файл с готовым программным кодом мастер-класса, по которому они будут последовательно выполнять каждую ячейку и видеть результат выполнения.

[PlotlyDataVisualization.ipynb](#) (приложение 1, п.6) - файл с готовым программным кодом мастер-класса.

Примечание: учащиеся могут самостоятельно ознакомиться с платформой социальной журналистики "Medium", открыв её официальный сайт: <https://medim.com> на своих устройствах.

ЭТАП 2 ОСНОВНОЕ СОДЕРЖАНИЕ ДЕЯТЕЛЬНОСТИ

Рекомендованное время: 40 минут.

ШАГ 1. Подготовка данных

Педагог информирует обучающихся о необходимости воспользоваться уже готовым набором собранных данных, скачав его с сервиса GitHub, чтобы не утруждать себя сбором и накоплением таких данных.

Для этого в ячейке Jupyter Notebook необходимо набрать программный код на Python, указанный в презентации (приложение 1, п.1), и запустим ячейку с набранным кодом на выполнение. В результате будут выведены первые 5 строк набора данных, чтобы иметь представление о нём.

Примечание:

учащиеся могут самостоятельно ознакомиться с содержимым набора данных, открыв его в браузере на своём устройстве по ссылке:

https://github.com/WillKoehrsen/Data-Analysis/blob/master/plotly/data/medium_data_2019_01_06?raw=true, загрузив его на свои компьютеры, и открыв содержимое в текстовом редакторе, например, «Блокнот» в ОС Windows или «Notepad++». Вместо последовательного набора программного кода за наставником учащимся можно раздать файл с готовым программным кодом мастер-класса, по которому они будут последовательно выполнять каждую ячейку и видеть результат выполнения.

[PlotlyDataVisualization.ipynb](#) (приложение 1, п.6) - файл с готовым программным кодом мастер-класса.

ШАГ 2. Построение гистограмм

Педагог информирует обучающихся, что для анализа «хлопков в ладоши» пользователей (сколько пользователей и сколько раз показали свои реакции) по загруженным данным необходимо построить столбчатую гистограмму. Для этого

необходимо в новой ячейке с клавиатуры ввести следующий программный код, указанный в презентации (приложение 1, п.5)

Примечание:

вместо последовательного набора программного кода за наставником учащимся можно раздать файл с готовым программным кодом мастер-класса, по которому они будут последовательно выполнять каждую ячейку и видеть результат выполнения.

[PlotlyDataVisualization.ipynb](#) (приложение 1, п.6) - файл с готовым программным кодом мастер-класса.

Педагог акцентирует внимание обучающихся на результате в виде гистограммы, вид которой показан в презентации (приложение 1). Педагог информирует обучающихся о том, что гистограммы в Plotly интерактивные, и просит обучающихся выделить интересующие их столбцы гистограммы, чтобы увидеть детали как это показано в презентации (приложение 1).

Педагог сообщает учащимся о проверке разницы между временем начала создания статей на платформе Medium и временем их публикации. Для этого от обучающихся требуется построить гистограмму анализа времени с эффектом наложения, путём ввода с клавиатуры указанный в презентации (приложение 1, п.5) программный код.

В результате должна получиться гистограмма с эффектом наложения, вид которой представлен в презентации (приложение 1, п.5). Синим цветом столбцов на гистограмме указано время начала создания статей, а красным – время их публикации на платформе.

ШАГ 3. Построение столбчатых графиков

Педагог сообщает обучающимся о том, чтобы проанализировать ежемесячную частоту публикации статей, количество их просмотров и прочтений, усреднённые по дате публикации статьи, им необходимо построить столбчатый график. Для этого нужно ввести с клавиатуры код, представленный в презентации (приложение 1, п.5) в новой ячейке Jupyter Notebook.

В результате у обучающихся должен получиться столбчатый график, вид которого представлен в презентации (приложение 1, п.5). На этом графике красным цветом показано количество прочтений статьи, а синим – количество просмотров.

Педагог уточняет, что обучающиеся могут объединить возможности библиотек Pandas и Plotly и Cufflinks. В качестве примера обучающимся необходимо построить столбчатый график по всей истории публикаций. Чтобы это сделать, обучающимся потребуется использовать `pivot`, а затем вид графика "box" библиотеки PlotExpress. В новой ячейке Jupyter Notebook необходимо ввести с клавиатуры программный код, указанный в презентации (приложение 1, п.5).

Примечание:

вместо последовательного набора программного кода за наставником учащимся можно раздать файл с готовым программным кодом мастер-класса, по которому они будут последовательно выполнять каждую ячейку и видеть результат выполнения.

[PlotlyDataVisualization.ipynb](#) (приложение 1, п.6) - файл с готовым программным кодом мастер-класса.

В результате у обучающихся получится столбчатый график распределения подписчиков по всей истории публикаций. Вид графика представлен в презентации (приложение 1, п.5). График обладает интерактивностью, поэтому при желании углубиться в детали можно выделить его желаемую часть.

ШАГ 4. Построение диаграмм-рассеяния

Педагог информирует обучающихся, что помощью библиотеки Plotly можно строить так называемые диаграммы-рассеяния или по-английски, Scatterplots. Такие диаграммы используются очень часто для анализа эволюции данных во времени между двумя и более переменными. Пример таких диаграмм-рассеяния представлен в презентации (приложение 1, п.5)

Педагог предлагает обучающимся построить один из видов диаграммы-рассеяния – временные ряды (Time Series). Для этого обучающимся потребуется на диаграмме отобразить названия статей к науке о данных, в связи с чем будет необходимо создать соответствующий набор данных (дата фрейм). В новой ячейке Jupyter Notebook необходимо добавить программный код, указанный в презентации.

В результате у обучающихся должны получиться два временных ряда:

1 - количество реакций читателей автора за всю историю его публикаций (указан синим цветом)

2 – количество фанатов автора из числа его читателей за всю историю его публикаций (указан красным цветом).

Получившиеся временные ряды представлены в презентации. Интерактивность здесь также присутствует, поэтому для погружения в детали той или иной части временного ряда, её необходимо выделить мышью.

Примечание:

вместо последовательного набора программного кода за наставником учащимся можно раздать файл с готовым программным кодом мастер-класса, по которому они будут последовательно выполнять каждую ячейку и видеть результат выполнения.

[PlotlyDataVisualization.ipynb](#) (приложение 1, п.6) - файл с готовым программным кодом мастер-класса.

Педагог сообщает обучающимся о том, что для анализа реакции пользователей, публикаций, просмотров, частоте прочтения статей и количестве слов в них на платформе «Medium» им требуется построить ещё один вид диаграмм-рассеяния – матрицу рассеяния или Scatter Matrix. Перед построением обучающимся нужно собрать новый снимок данных (дата фрейм), куда включить требуемые столбцы из набора данных. Для этого обучающиеся добавляют в новую ячейку Jupyter Notebook программный код, указанный в презентации (приложение 1, п.5).

В результате у обучающихся должна получиться матрица рассеяния, вид которой представлен в презентации (приложение 1, п.5).

В легенде этой матрицы (столбец справа) представлены темы статей, по которым производился анализ. Педагог акцентирует внимание обучающихся на том, что им не следует забывать про интерактивность, если они захотят детально разобраться в полученных данных.

ШАГ 5. Построение тепловой карты корреляции

Педагог информирует обучающихся о том, что для визуализации корреляций между числовыми переменными, сначала эти корреляции необходимо вычислить тем или иным способом, а затем по этим корреляциям требуется создать тепловую аннотированную карту корреляции или Correlation Heatmap. Педагог предлагает обучающимся по имеющемуся в их распоряжении набору данных построить такую карту корреляции. В новой ячейке Jupyter Notebook обучающимся необходимо ввести с клавиатуры следующий код, указанный в презентации (приложение 1, п.5). Педагог обращает внимание обучающихся на то, что для тепловой карты корреляции набор данных будет необходимо загрузить снова.

Примечание:

вместо последовательного набора программного кода за наставником учащимся можно раздать файл с готовым программным кодом мастер-класса, по которому они будут последовательно выполнять каждую ячейку и видеть результат выполнения.

[PlotlyDataVisualization.ipynb](#) (приложение 1, п.6) - файл с готовым программным кодом мастер-класса.

В результате у нас должна получиться такая тепловая карта корреляции, какая показана на этом слайде.

ЭТАП 3. АНАЛИЗ РЕЗУЛЬТАТОВ

Рекомендованное время: 40 минут

Педагог информирует обучающихся о том, что библиотека Plotly позволяет визуализировать полученные наборы данных для их наглядного представления, в чём они должны были убедиться на практическом примере, рассмотренном в рамках проведения этого мастер-класса. Педагог акцентирует внимание обучающихся на том, что была рассмотрена только лишь некоторая часть из всего списка диаграмм, но при помощи неё были получены наглядные отображения требуемых данных. Основываясь на полученных результатах, по характеру этих данных можно сделать некоторые логические выводы и выдвинуть гипотезы. Педагог информирует обучающихся, что для каждой конкретной задачи анализа данных им могут потребоваться разные виды диаграмм из библиотеки Plotly. Вид таких диаграмм и их количество будут зависеть от цели, которую обучающиеся будут преследовать, анализируя данные, а также от самого исследуемого набора данных.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

Для педагога:

1. Официальный сайт библиотеки Plotly, электронный ресурс, режим доступа: <https://plotly.com>
2. Официальный сайт с документацией по библиотеке Plotly – электронный ресурс, режим доступа: <https://plotly.com/python>
3. Визуализация с Plotly.Express: подробное руководство, электронный ресурс, режим доступа: <https://ichi.pro/ru/vizualizacia-s-plotly-express-podrobnoe-rukovodstvo-258792796523351>

Для обучающихся:

1. Шпаргалка по визуализации данных в Python с помощью Plotly, электронный ресурс, режим доступа: <https://habr.com/ru/post/502958/>
2. Визуализация с Plotly.Express: подробное руководство, электронный ресурс, режим доступа: <https://ichi.pro/ru/vizualizacia-s-plotly-express-podrobnoe-rukovodstvo-258792796523351>

ПРИЛОЖЕНИЕ 1

1. «Знакомство с библиотекой Plotly. Отличия Plotly от Matplotlib». Презентация: <https://disk.yandex.ru/i/oD2gL0bP2DE8yg>
2. Дополнительный материал. Листинг программного кода по теме «Знакомство с библиотекой Plotly. Отличия Plotly от Matplotlib»:
<https://disk.yandex.ru/d/vWHQZ57UZsZTlw>
3. Дополнительный материал. Файл с входными данными модуля 1:
<https://disk.yandex.ru/d/eQVkfCmZSF6WEA>
4. Дополнительный материал. Файл с входными данными модуля 2:
<https://disk.yandex.ru/d/uh-0wggwZimaInA>
5. «Визуализация данных в Python. Библиотека Plotly». Презентация:
https://disk.yandex.ru/i/xlsHNcSh6D_XKQ
6. Дополнительный материал. Листинг программного кода по теме «Визуализация данных в Python. Библиотека Plotly»:
<https://disk.yandex.ru/d/vINoz6xJLEgLAW>