

**Спецификация конкурсных материалов для проведения *практического* этапа
Московского конкурса межпредметных навыков и знаний «Интеллектуальный
мегаполис. Потенциал» в номинации «ИТ-класс» по направлению *Большие данные и
технологии искусственного интеллекта***

1. Назначение конкурсных материалов

Материалы *практического* этапа Московского конкурса межпредметных навыков и знаний «Интеллектуальный мегаполис. Потенциал» (далее – Конкурс) предназначены для оценки уровня *практической* подготовки участников Конкурса.

2. Условия проведения

Практический этап Конкурса проводится в очной форме на базе вуза или очной дистанционной форме. При выполнении работы обеспечивается строгое соблюдение порядка организации и проведения Конкурса.

Если этап проводится в очном дистанционном формате с использованием технологии прокторинга. Участникам необходимо иметь компьютер (ПК или ноутбук; прохождение диагностики на мобильных устройствах - невозможно) с выходом в Интернет, веб-камерой и микрофоном, а также смартфон (или планшет) со стабильным интернетом и приложением для считывания QR-кодов. Требуется предварительная настройка оборудования:

https://im.mcko.ru/docs/Инструкция_для_участника_конкурса_Интеллектуальный_мегаполис_Потенциал.pdf. Браузер разрешается использовать только для прохождения заданий этапа и процедуры прокторинга.

Дополнительное ПО, разрешенное для прохождения: компьютеры. На компьютерах установлена версия Python 3.12 или выше, а также интегрированные среды разработки (IDE) PyCharm и Visual Studio Code. Участники имеют доступ к необходимым библиотекам Python, включая numpy, pandas и matplotlib, которые настроены и готовы к использованию. Для работы с базами данных на компьютерах также установлена СУБД PostgreSQL и визуальный интерфейс pgAdmin или DBeaver.

Чем пользоваться категорически нельзя (ведет к отклонению работы): веб-поиском, методическими рекомендациями по направлениям Конкурса.

3. Продолжительность выполнения

На выполнение заданий *практического* этапа Конкурса отводится 150 минут. Во время проведения мероприятия участник может выйти из зоны проведения мероприятия не более чем на 5 минут, предупредив *ответственного от вуза*. Мероприятие не продлевается на время отсутствия участника.

4. Содержание и структура

Индивидуальный вариант участника включает 8 заданий, базирующихся на содержании *элективных курсов Введение в ИТ-специальность и Информационные технологии*. Задания 5, 6 предусматривают решение одного задания из списка на выбор.

5. Система оценивания

Задание считается выполненным, если ответ участника совпал с эталоном. Максимальный балл за выполнение всех заданий – 60 баллов.

6. Приложения

1. План конкурсных материалов для проведения *практического* этапа Конкурса.
2. Демонстрационный вариант конкурсных заданий *практического* этапа Конкурса.
3. Пример решения задач

План конкурсных материалов для проведения *практического* этапа Конкурса

№ задания	Выбор задания для решения	Уровень сложности	Уникальные кодификаторы Конкурса	Контролируемые требования к проверяемым умениям	Балл
1.	-	базовый	Библиотека pandas.	Знать: - Основные структуры данных библиотеки pandas: Series, DataFrame - Методы чтения и записи данных (CSV, Excel, JSON) - Способы фильтрации, группировки и агрегации данных Уметь: - Создавать и изменять объекты Series и DataFrame - Загружать данные из внешних файлов и сохранять их - Применять группировку ('groupby') и агрегатные функции ('mean()', 'sum()' и др.)	5
2.	-	базовый	Библиотека numpy.	Знать: - Структуру и назначение массива ndarray - Основные операции с массивами: арифметика, трансформация формы, индексация - Функции генерации данных: arange, linspace, random Уметь:	5



				- Создавать одномерные и многомерные массивы - Выполнять векторные операции и логические фильтрации - Использовать встроенные функции numpy для генерации и обработки данных	
3.	-	базовый	Библиотека matplotlib.	Знать: - Основные элементы графика: фигура, оси, подписи, легенда - Команды для построения графиков (<code>plot`</code>, <code>bar</code>, <code>hist</code>, <code>scatter`</code>) - Параметры стилизации графиков: цвета, маркеры, линии Уметь: - Строить линейные графики, столбчатые диаграммы, гистограммы - Настраивать подписи, легенды, сетку и заголовки - Сохранять графики в файл в различных форматах (PNG, PDF и др.)	5
4.	-	базовый	Анализ и визуализация данных на Python.	Знать: - Понятие и цели одномерного анализа (распределение, мода, медиана, размах) - Назначение гистограммы и графика плотности - Типы распределений	5



				Уметь: - Строить графики функций и гистограммы по массиву данных - Интерпретировать распределения по графикам - Определять характер распределения	
5.	На выбор 5 или 6	повышенный	Введение в SQL. Примеры в Postgresql	Умение формировать и оптимизировать SQL-запросы, знание основ работы с базами данных. Практическое применение SQL в PostgreSQL.	24
6.	На выбор 1 или 2	повышенный	Библиотека numpy. Примеры. Библиотека pandas. Примеры Библиотека matplotlib. Примеры	Умение использовать библиотеку NumPy для создания и преобразования массивов, выполнения арифметических операций и генерации случайных чисел. Умение применять Pandas для работы с табличными данными: создание и модификация DataFrame, фильтрация, группировка, обработка пропущенных значений и объединение данных. Умение использовать Matplotlib для визуализации данных: построение графиков, настройка осей и	24



				оформления.	
7.	-	повышенный	Исчисление вероятностей и элементы комбинаторики	Умение применять знания основ исчисления вероятностей и элементов комбинаторики.	6
8.	-	повышенный	Одномерный анализ данных. График функции. Гистограммы. Распределения	Умение визуализировать данные с помощью графиков и гистограмм, интерпретировать полученные результаты. Знание основных типов графиков и их применения для анализа распределения данных. Понимание распределений данных (например, нормальное распределение, биномиальное распределение и т.п.)	10
Сумма баллов:					60

Демонстрационный вариант конкурсных заданий *практического* этапа Конкурса**№1****Пример состава задания базового уровня сложности *практического* этапа Конкурса.**

1.

Был создан датафрейм с оценками по предметам:

```
import pandas as pd
data = {'Математика': [4, 5, 3], 'Информатика': [5, 4, 4]}
df = pd.DataFrame(data)
```

Необходимо вычислить среднюю оценку по каждому предмету, в ответе укажите соответствующую команду.

Ответ: `df.mean()`

Критерии оценивания: если задание решено верно, участник получает максимальный балл, иначе 0 баллов.

Описание хода практической части в случае очной дистанционной формы проведения этапа Конкурса: *категорически нельзя (ведет к отклонению работы): веб-поиском, методическими рекомендациями по направлениям Конкурса*. Во время проведения мероприятия участник может выйти из зоны проведения мероприятия не более чем на 5 минут, предупредив *проктора на камеру*. Мероприятие не продлевается на время отсутствия участника.

2.

Необходимо создать массив из 5 одинаковых значений 7. В ответе напишите соответствующую команду. Способ решения должен быть корректным с помощью NumPy.

Ответ: `np.full(5, 7)`, либо любой другой корректный способ решения.

Критерии оценивания: если задание решено верно, участник получает максимальный балл, иначе 0 баллов.

Описание хода практической части в случае очной дистанционной формы проведения этапа Конкурса: *категорически нельзя (ведет к отклонению работы): веб-поиском, методическими рекомендациями по направлениям Конкурса*. Во время проведения мероприятия участник может выйти из зоны проведения мероприятия не более чем на 5 минут, предупредив *проктора на камеру*. Мероприятие не продлевается на время отсутствия участника.

3.

Необходимо построить график зависимости температуры от времени. Массивы `time`, `temperature` заданы. Напишите команду, учитывая, что она должна быть в формате `plt.__(__, __)`.

Ответ: `plt.plot(time, temperature)`

Критерии оценивания: если задание решено верно, участник получает максимальный балл, иначе 0 баллов.

Описание хода практической части в случае очной дистанционной формы проведения этапа Конкурса: *категорически нельзя (ведет к отклонению работы): веб-поиском, методическими рекомендациями по направлениям Конкурса*. Во время проведения мероприятия участник может выйти из зоны проведения мероприятия не более чем на 5 минут, предупредив *проктора на камеру*. Мероприятие не продлевается на время отсутствия участника.

4.

Необходимо загрузить CSV-файл `inf.csv` в `DataFrame` для анализа. Напишите команду, учитывая, что она должна быть в формате `pd.__(__)`

Ответ: `pd.read_csv('inf.csv')`

Критерии оценивания: если задание решено верно, участник получает максимальный балл, иначе 0 баллов.

Описание хода практической части в случае очной дистанционной формы проведения этапа Конкурса: *категорически нельзя (ведет к отклонению работы): веб-поиском, методическими рекомендациями по направлениям Конкурса*. Во время проведения мероприятия участник может выйти из зоны проведения мероприятия не более чем на 5 минут, предупредив *проктора на камеру*. Мероприятие не продлевается на время отсутствия участника.

Демонстрационный вариант конкурсных заданий *практического* этапа Конкурса**Пример состава задания повышенного уровня сложности *практического* этапа Конкурса.****Задача 5.**

Библиотека готовится к ежегодному фестивалю «Библиофест», где планируется наградить самых активных авторов и читателей, а также провести выставку популярных книг. Вам, как аналитику, поручили собрать ключевые данные из PostgreSQL-базы библиотеки. Таблицы базы данных:

- **authors** (id, name, country) — авторы.
- **books** (id, title, year, author_id, pages) — книги.
- **readers** (id, name, email, registration_date) — читатели.
- **book_loans** (id, book_id, reader_id, loan date, return_date) — записи о выдаче книг.

Для этого сформируйте необходимые запросы к базе данных для составления списков для следующих номинаций.

1. **Номинация «Мастер слова».** Для награждения авторов, чьи книги наиболее объемны и современны: вывести имя автора и общее количество страниц его книг, изданных после 2015 года. Сортировать по убыванию страниц.
2. **Номинация «Книжный фанат».** Читатели, претендующие на звание «Самый активный»: найти тех, кто брал книги чаще 2 раз в феврале 2023 года. Вывести их имя, email и количество выдач.

В качестве ответа необходимо предоставить два sql-запроса.

Запишите свое решение в файл: задача_1.txt.

Критерии оценивания:

Максимальная оценка за каждый запрос составляет 12 балла. Общее количество баллов за задание – 24.

Корректный синтаксис	1
Запрос выполняется без ошибок в PostgreSQL	1



Есть незначительные ошибки (например, пропущена запятая, неверный алиас), но запрос работает после исправления	0.5	
Синтаксические ошибки, из-за которых запрос не выполняется (например, неправильное использование ключевых слов, некорректные имена таблиц)	0	
Логика	7	
Результат полностью соответствует условию (учтены фильтры, сортировка, группировка, исключены дубли)	7	
Частичные ошибки в логике (например, не учтена сортировка по убыванию, пропущен год в фильтре)	3	
Логика нарушена (например, вывод книг, изданных до 2015 года вместо после)	0	
Оптимальность и дополнительные требования	4	
Использованы оптимальные методы (например, JOIN вместо подзапросов). Выполнены все дополнительные условия. Нет избыточных операций (например, SELECT * вместо конкретных колонок).	4	
Частично выполнено (например, NOT EXISTS заменен на LEFT JOIN, но результат верный)	2	
Грубые нарушения (например, запрос с CROSS JOIN, ведущим к декартову произведению)	0	
ИТОГО	12	

Задача 6.

Вы — аналитик данных в розничной сети. Вам предоставили файл `sales_data.csv` с историей продаж за 6 месяцев

Цель: выявить категории товаров с нестабильной динамикой и предложить меры по стабилизации продаж.

Файл `sales_data.csv`, содержащий ежемесячные данные о продажах. Формат таблицы:

- `product_id` — идентификационный номер продукта (целое число)
- `month` — месяц в формате YYYY-MM (строка или дата)
- `sales` — объем продаж (число, может быть ноль)
- `category` - категории продуктов (строка)

Требуется

1. Загрузить данные из файла `sales_data.csv` - файл должен находиться в том же проекте, что и программа.
 - Используйте `pd.to_datetime()`, `sort_values()`
 - Решение можно оформить так: загрузка → преобразование месяца → сортировка.

2. Для каждой товарной позиции (`product_id`) определить, изменились ли продажи по сравнению с предыдущим месяцем: Используйте `groupby()` и `diff()` для вычисления разницы.
 - Создайте новый столбец `sales_changed`, который принимает значение:
 - i. `True`, если продажи изменились
 - ii. `False`, если продажи не изменились
 - Для расчёта используйте `diff().ne(0)`.
3. Итоговая таблица должна содержать следующие столбцы (в указанном порядке):
 - `product_id`
 - `month`
 - `sales`
 - `category`
 - `sales_change`
 - Приведите дату обратно к формату 'YYYY-MM' через `.dt.strftime()`
 - *Проверьте, что порядок и формат соответствуют.*
4. Вывести первые 10 строк итоговой таблицы.
5. Для каждого сочетания `category` и `month` рассчитать медианные `sales`
6. Построить **линейный график**, где:
 - Ось X — `month`
 - Ось Y — медианные `sales`
 - Каждая линия — отдельная категория
7. Используйте `.median()`, `.reset_index()`, `matplotlib.pyplot` (`plot()`, `xticks()`, `legend()` и т. д.)
8. Сохранить изображение как `sales_lineplot.png`

Запишите свое решение в файл: `task_2.py`.

Документацию по приведенным методам можно получить с помощью функции **`help()`**.

Критерии оценивания:

Максимальная количество баллов за задание – 24.

Загрузка и предобработка данных		4
Данные загружены корректно Проведена проверка на пропуски (если требуется)	4	



Загрузка выполнена, но не учтены особенности данных (кодировка, разделители)	2	
Данные загружены с ошибками (например, неверный путь к файлу)	1	
Данные не загружены	0	
Расчет показателей	10	
Все метрики (динамика, средние, медианы и т.д.) рассчитаны верно Учтены краевые случаи (деление на ноль, отрицательные значения)	10	
Метрики есть, но есть ошибки в формулах или группировках	5	
Расчеты проведены, но не соответствуют условию задачи	1	
Показатели не рассчитаны	0	
Добавление новых столбцов	4	
Столбцы созданы через векторные операции (без циклов) Категоризация выполнена по условиям задачи	4	
Логика верна, но код не оптимален (например, использование apply)	2	
Столбцы добавлены, но с ошибками в условиях	1	
Новые столбцы отсутствуют	0	
Визуализация	5	
Построен график требуемого типа (heatmap, line, bar и т.д.). Наличие подписей осей, заголовка, легенды. График сохранен в файл	5	
График построен, но: Неверный тип визуализации, Отсутствуют подписи/легенда,	2	
Есть график, но не соответствует данным	1	
Визуализация отсутствует	0	
Оформление кода и вывод	1	
Код читаем, есть комментарии (где необходимо). Результаты выведены в требуемом формате (например, print или to_markdown).	1	
Код не структурирован, вывод отсутствует	0	
ИТОГО	24	

Задача 7.

В школе проводится лотерея для сбора средств на благотворительность. Участники покупают билеты, каждый из которых содержит уникальный номер из 5 цифр (включительно от 00000 до 99999). Призы разыгрываются по следующим правилам:

1. **Главный приз:** номер содержит ровно три одинаковые цифры (например, 12121, 33305).
2. **Дополнительный приз:** номер является «палиндромом» (читается одинаково слева направо и справа налево, например, 12321, 55055).

Какова вероятность, что случайно выбранный номер одновременно претендует на главный приз и является палиндромом?

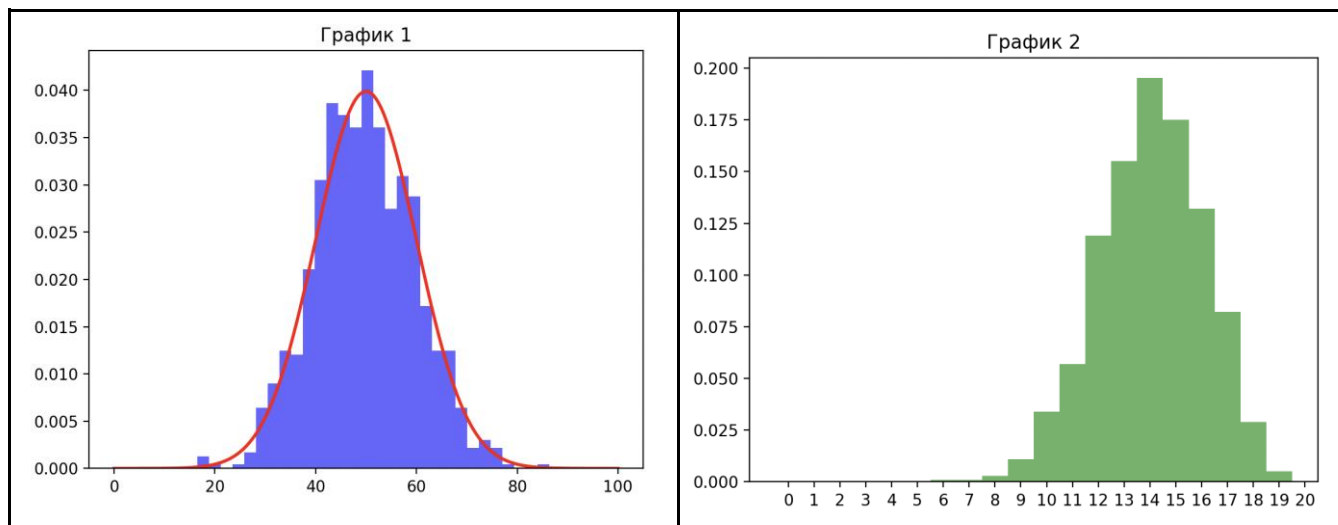
Запишите свое решение в файл: задача_3.txt. Ваше решение должно основываться на математических вычислениях - использование сторонних программ не допускается.

Критерии оценивания:

Максимальная количество баллов за задание – 6.

Логичность и отсутствие ошибок		3
Последовательность рассуждений, пояснение каждого шага	1	
Отсутствие арифметических, логических ошибок или противоречий	2	
Разделение на случаи (например, центральная цифра совпадает с первой/второй)		2
Выявление особенностей объектов, соответствующих условию (например, палиндромы, номера с повторяющимися цифрами)	1	
Правильное применение комбинаторных методов (правила умножения, сочетания, размещения)	1	
Предоставлен верный ответ		1
ИТОГО		6

Задача 8.



1. Определите тип распределения для каждого графика.
2. Назовите параметры распределений:
 - Для Графика 1: μ и σ .
 - Для Графика 2: число испытаний (n) и вероятность успеха (p).
3. Объясните, по каким визуальным признакам вы определили тип распределения.

Запишите свое решение в файл: задача_4.txt

Критерии оценивания:

Максимальная оценка за каждый график составляет 5 баллов. Общее количество баллов за задание – 10.

Определение типа распределения		1
Правильно указан тип распределения	1	
Тип распределения указан с неуверенностью или допущена несущественная ошибка (например, «похоже на нормальное» без точного названия)	0.5	
Тип распределения указан неверно или не указан вовсе	0	
Указание параметров распределения		2
Все параметры корректны и соответствуют графику	2	
Указан только один параметр правильно, или оба с частичной точностью	1	
Параметры не указаны или указаны с грубыми ошибками	0	



Обоснование выбора типа распределения		2
Обоснование подробное, логичное и основано на анализе формы графика (симметрия, пик, хвосты, размах и т.д.)	2	
Обоснование частично верное или не до конца раскрыто (общие фразы, без привязки к графику)	1	
Обоснование отсутствует или ошибочно	0	
ИТОГО		5



Пример решения задания повышенного уровня сложности практического этапа Конкурса.

Задача 5.

1.

```
SELECT a.name AS author_name, SUM(b.pages) AS total_pages
FROM authors a
JOIN books b ON a.id = b.author_id
WHERE b.year > 2015
GROUP BY a.id, a.name
ORDER BY total_pages DESC;
```

2.

```
SELECT r.name, r.email, COUNT(bl.id) AS loan_count
FROM readers r
JOIN book_loans bl ON r.id = bl.reader_id
WHERE bl.loan_date >= '2023-02-01' AND bl.loan_date < '2023-03-01'
GROUP BY r.id, r.name, r.email
HAVING COUNT(bl.id) > 2;
```

Задача 6.

```
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv('sales_data.csv')

df['month'] = pd.to_datetime(df['month'])
df = df.sort_values(['product_id', 'month'])
df['sales_change'] = df.groupby('product_id')['sales'].diff().ne(0)

df['month'] = df['month'].dt.strftime('%Y-%m')

df = df[['product_id', 'month', 'sales', 'category', 'sales_change']]

print(df.head(10))

median_sales = df.groupby(['category', 'month'])['sales'].median().reset_index()

plt.figure(figsize=(12, 6))
for category in median_sales['category'].unique():
    subset = median_sales[median_sales['category'] == category]
    plt.plot(subset['month'], subset['sales'], label=category)

plt.xticks(rotation=45)
plt.xlabel('Месяц')
plt.ylabel('Медианные продажи')
plt.title('Медианные продажи по категориям товаров')
plt.legend()
plt.tight_layout()

plt.savefig('sales_lineplot.png')
```

	product_id	month	sales	category	sales_change
0	1	2023-01	152	Электроника	True
1	2	2023-01	238	Одежда	True
2	3	2023-01	124	Продукты	True
3	4	2023-01	201	Мебель	True
4	5	2023-01	285	Книги	True
5	1	2023-02	229	Электроника	True
6	2	2023-02	70	Одежда	True
7	3	2023-02	252	Продукты	True
8	4	2023-02	180	Мебель	True
9	5	2023-02	207	Книги	True



Задача 7.

1. Общее количество номеров

Каждый номер — это 5-значное число от 00000 до 99999: $N_{total} = 10^5$

2. Структура пятизначного палиндрома.

Палиндром имеет вид: $abcba$, где a, b, c принадлежат множеству $\{0, 1, \dots, 9\}$.

В номере $abcba$ цифры a и b встречаются по два раза, c — один раз.

Чтобы было ровно три одинаковые цифры, возможны два случая:

Случай 1: $a = c, a \neq b$

Кол-во вариантов выбора a - 10

Кол-во вариантов выбора: b - 9

Кол-во возможных комбинаций: $10 \times 9 = 90$

Рассуждения для Случая 2: $b = c$, $a \neq b$ аналогичны

Итого всего количество билетов, удовлетворяющих условиям: $90 + 90 = 180$

3. Расчет вероятности: $\frac{180}{10^5} * 100\% = 0.18\%$

Ответ: 0.18%

Задача 8.

1. График 1 - нормальное распределение. График 2 - биномиальное распределение.

2. $\mu \approx 50$ и $\sigma \approx \frac{|40 - 60|}{2} \approx 10$.

$n = 20$, $p = \frac{\mu}{n} \approx \frac{14}{20} \approx 0.7$ (разрешается погрешность в ± 0.2)

Допустимые погрешности при учете ответа:

μ : $\pm 5\%$ от истинного значения (например, если $\mu = 50$, допустим ответ 48–52).

σ : ± 10 –15% (например, $\sigma = 10 \rightarrow$ ответ 9–11)

n : точное значение (например, $n = 20 \rightarrow$ только 20, так как это целое число).

p : ± 0.05 (например, $p = 0.7 \rightarrow$ ответ 0.65–0.75)

3. График 1: симметричная колоколообразная форма, соответствует нормальному распределению.

График 2: дискретные значения (целые числа от 0 до 20), пик около 14 ($n \cdot p = 14$)

**Демонстрационный вариант конкурсных заданий *практического* этапа Конкурса
№2**

Пример состава задания базового уровня сложности *практического* этапа Конкурса.

1.

Необходимо выбрать из DataFrame только те строки, где в столбце "Баллы" значение больше 80, используя метод query:

```
import pandas as pd
df = pd.DataFrame({'Имя': ['Аня', 'Борис', 'Вика'], 'Баллы': [75, 90, 85]})
```

В ответе укажите соответствующую команду.

Ответ: `df.query('Баллы > 80')`

Критерии оценивания: если задание решено верно, участник получает максимальный балл, иначе 0 баллов.

Описание хода практической части в случае очной дистанционной формы проведения этапа Конкурса: *категорически нельзя (ведет к отклонению работы): веб-поиском, методическими рекомендациями по направлениям Конкурса*. Во время проведения мероприятия участник может выйти из зоны проведения мероприятия не более чем на 5 минут, предупредив *проктора на камеру*. Мероприятие не продлевается на время отсутствия участника.

2.

Создайте массив из 100 случайных чисел от 0 до 10 с равномерным распределением и округлите до 2 знаков, учитывая, что ответ должен быть записан в виде `np.__(np.random.__(), __)`.

Ответ: `np.round(np.random.uniform(0, 10, 100), 2)`

Критерии оценивания: если задание решено верно, участник получает максимальный балл, иначе 0 баллов.

Описание хода практической части в случае очной дистанционной формы проведения этапа Конкурса: *категорически нельзя (ведет к отклонению работы): веб-поиском, методическими рекомендациями по направлениям Конкурса*. Во время проведения мероприятия участник может выйти из зоны проведения мероприятия не более чем на 5 минут, предупредив *проктора на камеру*. Мероприятие не продлевается на время отсутствия участника.

3.

Постройте гистограмму из массива data с 20 равными интервалами и отобразите сетку.

Ответ должен быть записан в виде `plt.__(__, __); plt.__()`.

Ответ: `plt.hist(data, bins=20); plt.grid()`

Критерии оценивания: если задание решено верно, участник получает максимальный балл, иначе 0 баллов.

Описание хода практической части в случае очной дистанционной формы проведения этапа Конкурса: *категорически нельзя (ведет к отклонению работы): веб-поиском, методическими рекомендациями по направлениям Конкурса*. Во время проведения мероприятия участник может выйти из зоны проведения мероприятия не более чем на 5 минут, предупредив *проктора на камеру*. Мероприятие не продлевается на время отсутствия участника.

4.

Постройте график плотности распределения (kernel density estimation) для столбца 'score' датафрейма df. В ответ впишите верную команду.

Ответ: `df['score'].plot.kde()`

Критерии оценивания: если задание решено верно, участник получает максимальный балл, иначе 0 баллов.

Описание хода практической части в случае очной дистанционной формы проведения этапа Конкурса: *категорически нельзя (ведет к отклонению работы): веб-поиском, методическими рекомендациями по направлениям Конкурса*. Во время проведения мероприятия участник может выйти из зоны проведения мероприятия не более чем на 5 минут, предупредив *проктора на камеру*. Мероприятие не продлевается на время отсутствия участника.

Демонстрационный вариант конкурсных заданий *практического* этапа Конкурса**Пример состава задания повышенного уровня сложности *практического* этапа Конкурса.****Задача 5.**

Образовательная онлайн-платформа анализирует активность пользователей и популярность курсов для улучшения качества обучения и обновления контента. Вам, как аналитику, поручили подготовить выборки из базы PostgreSQL.

Структура базы данных:

- instructors (id, name, expertise) — преподаватели.
- courses (id, title, year, instructor_id, duration_hours) — курсы.
- students (id, name, email, registration_date) — студенты.
- course_enrollments (id, course_id, student_id, enrollment_date, completion_date) — записи о прохождении курсов.

Подготовьте SQL-запросы к базе данных для следующих целей:

1. Анализ активности преподавателей

Вывести имена преподавателей и общее количество часов всех их курсов, выпущенных после 2020 года. Отсортировать по убыванию суммарной продолжительности.

2. Идентификация вовлеченных студентов

Найти студентов, которые записывались на курсы более 3 раз за март 2023 года. Вывести имя, email и общее количество записей в этом месяце.

В качестве ответа необходимо предоставить два sql-запроса.

Запишите свое решение в файл: задача_1.txt.

Критерии оценивания:

Максимальная оценка за каждый запрос составляет 12 балла. Общее количество баллов за задание – 24.

Корректный синтаксис		1
Запрос выполняется без ошибок в PostgreSQL	1	
Есть незначительные ошибки (например, пропущена запятая, неверный алиас), но запрос работает после исправления	0.5	
Синтаксические ошибки, из-за которых запрос не выполняется (например, неправильное использование ключевых слов, некорректные имена таблиц)	0	
Логика		7
Результат полностью соответствует условию (учтены фильтры, сортировка, группировка, исключены дубли)	7	
Частичные ошибки в логике (например, не учтена сортировка по убыванию, пропущен год в фильтре)	3	
Логика нарушена (например, вывод книг, изданных до 2015 года вместо после)	0	
Оптимальность и дополнительные требования		4
Использованы оптимальные методы (например, JOIN вместо подзапросов). Выполнены все дополнительные условия. Нет избыточных операций (например, SELECT * вместо конкретных колонок).	4	
Частично выполнено (например, NOT EXISTS заменен на LEFT JOIN, но результат верный)	2	
Грубые нарушения (например, запрос с CROSS JOIN, ведущим к декартову произведению)	0	
ИТОГО		12

Задача 6.

Вы — аналитик клиентского опыта в крупном интернет-магазине. Вам предоставили файл reviews_data.csv с историей отзывов клиентов за последние 12 месяцев.

Цель: оценить изменение клиентской удовлетворенности по продуктам и выявить группы с ухудшением восприятия.

Файл: reviews_data.csv, формат таблицы:

Колонка	Описание
product_id	Идентификатор продукта (целое число)
review_date	Дата отзыва (в формате YYYY-MM-DD)
rating	Оценка (целое число от 1 до 5)

sentiment	Категория тональности отзыва (positive/negative/neutral)
department	Отдел, к которому относится товар (строка)

Требуется:

1. **Загрузить данные из reviews_data.csv** — файл должен быть в том же проекте, что и программа.
2. **Преобразовать даты в формат месяца (YYYY-MM)** и сгруппировать данные по:
 - product_id
 - review_month (новый столбец, извлечённый из review_date)
 - Используйте pd.to_datetime(), .dt.to_period('M') или .dt.strftime('%Y-%m')
3. Для каждой группы рассчитать среднюю оценку (avg_rating) за месяц.
 - Используйте groupby(), agg() с mean
4. **Сформировать итоговую таблицу со столбцами (в указанном порядке):**
 - product_id
 - review_month
 - avg_rating
 - department
5. Отсортировать по review_month по возрастанию. Вывести первые 10 строк.
6. Построить **гистограмму распределения клиентских оценок** — rating
 - i. Ось X — rating
 - ii. Ось Y — количество отзывов, поставивших соответствующую оценку
 - Сохранить как rating_gist.png
 - Используйте matplotlib.pyplot (plot(), xticks(), legend(), hist() и т. д.)
 - Добавить подписи осей и заголовок графика.

Документацию по приведенным методам можно получить с помощью функции **help()**.

Критерии оценивания:

Максимальная количество баллов за задание – 24.



Данные загружены корректно Проведена проверка на пропуски (если требуется)	4	
Загрузка выполнена, но не учтены особенности данных (кодировка, разделители)	2	
Данные загружены с ошибками (например, неверный путь к файлу)	1	
Данные не загружены	0	
Расчет показателей		10
Все метрики (динамика, средние, медианы и т.д.) рассчитаны верно Учтены краевые случаи (деление на ноль, отрицательные значения)	10	
Метрики есть, но есть ошибки в формулах или группировках	5	
Расчеты проведены, но не соответствуют условию задачи	1	
Показатели не рассчитаны	0	
Добавление новых столбцов		4
Столбцы созданы через векторные операции (без циклов) Категоризация выполнена по условиям задачи	4	
Логика верна, но код не оптимален (например, использование apply)	2	
Столбцы добавлены, но с ошибками в условиях	1	
Новые столбцы отсутствуют	0	
Визуализация		5
Построен график требуемого типа (heatmap, line, bar и т.д.). Наличие подписей осей, заголовка, легенды. График сохранен в файл	5	
График построен, но: Неверный тип визуализации, Отсутствуют подписи/легенда,	2	
Есть график, но не соответствует данным	1	
Визуализация отсутствует	0	
Оформление кода и вывод		1
Код читаем, есть комментарии (где необходимо). Результаты выведены в требуемом формате (например, print или to_markdown).	1	
Код не структурирован, вывод отсутствует	0	
ИТОГО		24

Задача 7.

Школьная столовая планирует меню на неделю (5 учебных дней). В распоряжении поваров — 8 разных горячих блюд. Каждый день подают **одно** горячее блюдо, и каждое блюдо можно использовать **не более одного раза в неделю**.

Но чтобы не было скучно, вводятся два условия:

- Блюдо понедельника всегда – овощное рагу.
- В течение недели **ровно одно** блюдо должно начинаться с буквы “**К**” (например, «Котлета»).

Вот список блюд, которые есть в распоряжении поваров:

1. Котлета по-домашнему
2. Каша гречневая с подливкой
3. Плов с курицей
4. Макароны по-флотски
5. Овощное рагу
6. Рыбные тефтели
7. Запеканка картофельная
8. Суп гороховый

Сколько различных вариантов меню можно составить на неделю?

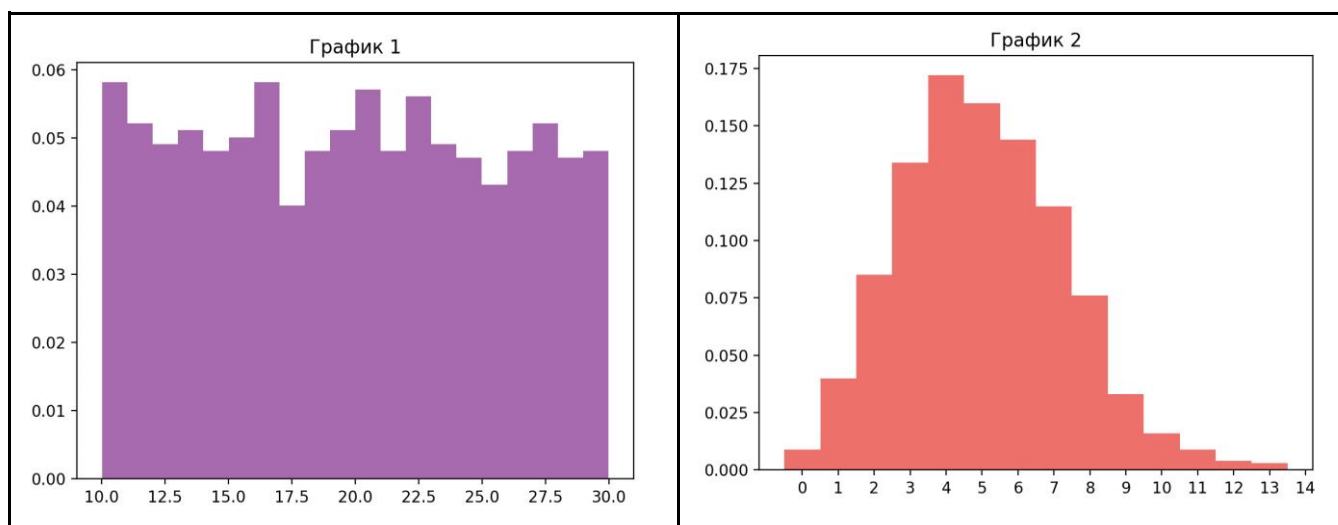
Запишите свое решение в файл: задача_2.txt. Ваше решение должно основываться на математических вычислениях - использование сторонних программ не допускается.

Критерии оценивания:

Максимальная количество баллов за задание – 6.

Логичность и отсутствие ошибок		3
Последовательность рассуждений, пояснение каждого шага	1	
Отсутствие арифметических, логических ошибок или противоречий	2	
Разделение на случаи (например, центральная цифра совпадает с первой/второй)		2
Выявление особенностей объектов, соответствующих условию (например, палиндромы, номера с повторяющимися цифрами)	1	
Правильное применение комбинаторных методов (правила умножения, сочетания, размещения)	1	
Предоставлен верный ответ		1
ИТОГО		6

Задача 8.



1. Определите тип распределения для каждого графика.
2. Оцените параметры распределений:
 - Для Графика 1: границы интервала (**a** и **b**).
 - Для Графика 2: параметр λ (лямбда).
3. Объясните, по каким визуальным признакам вы определили тип распределения.

Критерии оценивания:

Максимальная оценка за каждый график составляет 5 баллов. Общее количество баллов за задание – 10.

Определение типа распределения		1
Правильно указан тип распределения	1	
Тип распределения указан с неуверенностью или допущена несущественная ошибка (например, «похоже на нормальное» без точного названия)	0.5	
Тип распределения указан неверно или не указан вовсе	0	
Указание параметров распределения		2
Все параметры корректны и соответствуют графику	2	
Указан только один параметр правильно, или оба с частичной точностью	1	
Параметры не указаны или указаны с грубыми ошибками	0	
Обоснование выбора типа распределения		2
Обоснование подробное, логичное и основано на анализе формы графика (симметрия, пик, хвосты, размах и т.д.)	2	



Обоснование частично верное или не до конца раскрыто (общие фразы, без привязки к графику)	1	
Обоснование отсутствует или ошибочно	0	
ИТОГО		5



Пример решения задания повышенного уровня сложности практического этапа Конкурса.

Задача 5.

1.

```
SELECT
    i.name AS instructor_name,
    SUM(c.duration_hours) AS total_hours
FROM
    instructors i
JOIN
    courses c ON i.id = c.instructor_id
WHERE
    c.year > 2020
GROUP BY
    i.name
ORDER BY
    total_hours DESC;
```

2.

```
SELECT
    s.name,
    s.email,
    COUNT(*) AS enrollment_count
FROM
    students s
JOIN
    course_enrollments ce ON s.id = ce.student_id
WHERE
    ce.enrollment_date >= '2023-03-01' AND ce.enrollment_date < '2023-04-01'
GROUP BY
    s.name, s.email
HAVING
    COUNT(*) > 3;
```



Задача 6.

```
import pandas as pd
import matplotlib.pyplot as plt
```

```
df = pd.read_csv("reviews_data.csv")
```

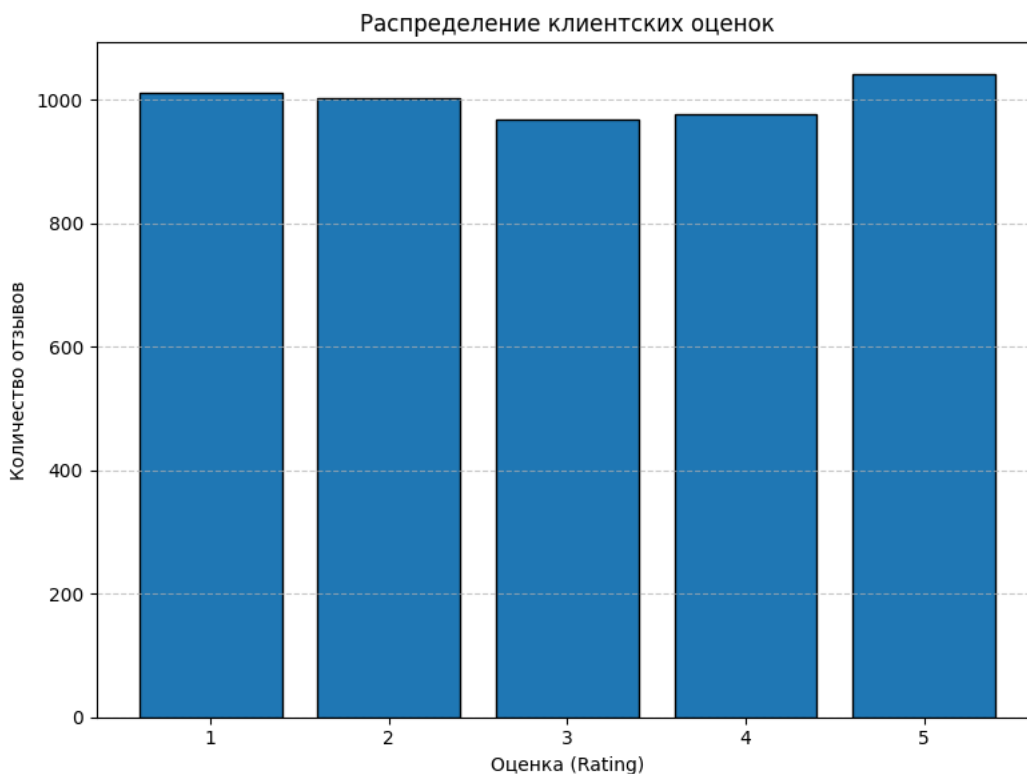
```
df['review_date'] = pd.to_datetime(df['review_date'])
df['review_month'] = df['review_date'].dt.to_period('M').astype(str) # Можно также .dt.strftime('%Y-%m')
```

```
grouped = df.groupby(['product_id', 'review_month', 'department']).agg(avg_rating=('rating', 'mean')).reset_index()
```

```
final_df = grouped[['product_id', 'review_month', 'avg_rating', 'department']]
```

```
final_df = final_df.sort_values(by='review_month')
print(final_df.head(10))
```

```
plt.figure(figsize=(8, 6))
plt.hist(df['rating'], bins=range(1, 7), align='left', edgecolor='black', rwidth=0.8)
plt.xlabel('Оценка (Rating)')
plt.ylabel('Количество отзывов')
plt.title('Распределение клиентских оценок')
plt.xticks(range(1, 6))
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.savefig('rating_gist.png')
plt.show()
```



	product_id	review_month	avg_rating	department
0	1000	2024-01	3.0	Beauty

430	1009	2024-01	4.0	Electronics
429	1009	2024-01	2.0	Clothing
428	1009	2024-01	1.5	Beauty
1725	1036	2024-01	3.0	Clothing
1726	1036	2024-01	1.0	Electronics
1727	1036	2024-01	3.0	Home
1728	1036	2024-01	4.0	Sports
1773	1037	2024-01	5.0	Clothing
1774	1037	2024-01	1.5	Electronics

Задача 7.

- Из 8 блюд два начинаются на "К" — выбираем одно: $C(2,1) = 2$.
 - Из 5 не-К-блюд (кроме Овощного рагу и выбранного "К") выбрать 3: $C(5, 3) = 10$.
 - Тогда кол-во способов выбрать 4 блюда: $2 * 10 = 20$
 - Эти 4 блюда (1 "К"-блюдо и 3 не-К) нужно **переставить по 4 дням** — Вт, Ср, Чт, Пт.
- Все блюда разные. Размещаем: $4! = 24$ способов
5. Общее число вариантов: $20 \times 24 = 480$.

Ответ: 480

Задача 8.

- График 1 — равномерное, График 2 — Пуассона
- График 1: **a ≈ 10, b ≈ 30** (значения равномерно распределены от 10 до 30).

График 2: **λ ≈ 5** (пик около 5, среднее = λ)

Допустимые погрешности:

a и b: ±2–3% от диапазона (например, данные от 10 до 30 → $a \approx 10 \pm 0.5$, $b \approx 30 \pm 0.5$).

λ: ±1–2 единицы (например, $\lambda = 5 \rightarrow$ ответ 4–6)

- График 1: гистограмма плоская, все значения в диапазоне [10, 30] встречаются примерно одинаково часто.

График 2: дискретные значения, пик около 5, асимметрия вправо (характерно для Пуассона).