

**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.	<p>Знать:</p> <ul style="list-style-type: none">- Основные структуры данных библиотеки pandas: Series, DataFrame- Методы чтения и записи данных (CSV, Excel, JSON)- Способы фильтрации, группировки и агрегации данных <p>Уметь:</p> <ul style="list-style-type: none">- 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.	<p>Знать:</p> <ul style="list-style-type: none">- Структуру и назначение массива ndarray- Основные операции с массивами: арифметика, трансформация формы, индексация- Функции генерации данных: arange, linspace, random <p>Уметь:</p>	5



				<ul style="list-style-type: none">- Создавать одномерные и многомерные массивы- Выполнять векторные операции и логические фильтрации- Использовать встроенные функции numpy для генерации и обработки данных	
3.	-	базовый	Библиотека matplotlib.	<p>Знать:</p> <ul style="list-style-type: none">- 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 <p>Уметь:</p> <ul style="list-style-type: none">- 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.	<p>Знать:</p> <ul style="list-style-type: none">- Понятие и цели одномерного анализа (распределение, мода, медиана, размах)- Назначение гистограммы и графика плотности- Типы распределений	5



				<p>Уметь:</p> <ul style="list-style-type: none">- 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` — объём 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.

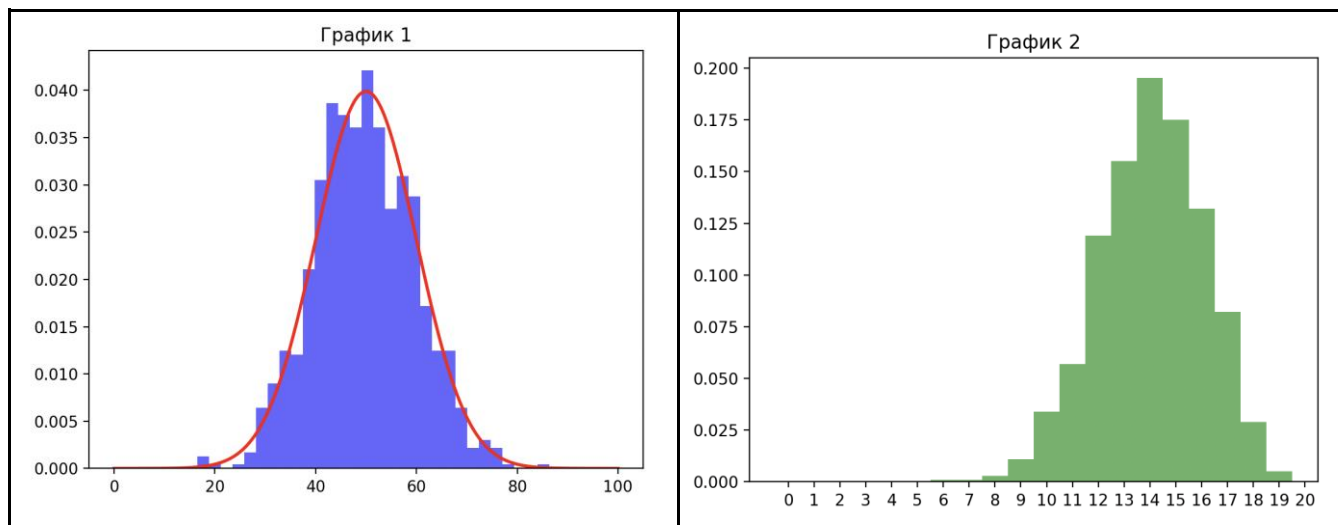
Критерии оценивания:

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