Инструкция по рецензированию

Часть 1. Работа с табличными данными.

- Во всех заданиях обязательно наличие кода решения. Письменные ответы на вопросы без сопутствующего кода оцениваются в 0 баллов.

Часть 2. Визуализация / Часть 3. Векторные вычисления

Все графики должны быть опрятные:

- Должно быть название (plt.title) графика;
- На графиках должны быть подписаны оси (plt.xlabel, plt.ylabel);
- Должны быть подписаны единицы измерения (если это возможно);
- Все названия должны быть понятны любому человеку, знакомому с терминологией, без заглядывания в код;
- Подписи тиков на осях не должны сливаться как на одной оси, так и между ними;
- Если изображено несколько сущностей на одном холсте (например несколько функций), то необходима поясняющая легенда (plt.legend);
- Все линии на графиках должны быть чётко видны (нет похожих цветов или цветов, сливающихся с фоном);
- Если отображена величина, имеющая очевидный диапазон значений (например, проценты могут быть от 0 до 100), то желательно масштабировать ось на весь диапазон значений (исключением является случай, когда вам необходимо показать малое отличие, которое незаметно в таких масштабах);
- Графики должны быть не супер-микро и не супер-макро по размерам, так, чтобы можно было увидеть все, что нужно.

При нарушении хотя бы 3 пунктов график считается НЕ опрятным, что ведет к снижению баллов по соответствующему критерию.

Файлы решения

numpy-pandas-matplotlib.ipynb (/solution_file/14726/numpy-pandas-matplotlib.ipynb)

Рецензии

Часть 1. Задание 1. Откройте файл с таблицей и выведите последние 5 строк

0.0 баллов
Пункт не выполнен / выполнен неправильно.

0.2 баллов
Выведены строки с индексами 3216-3220.

Рецензия №1 | Рецензия №2 | Рецензия №3

Часть 1. Задание 1. Посмотрите на данные и скажите, что они из себя представляют, сколько в таблице строк и столбцов?

0.0 баллов
Пункт не выполнен / выполнен неправильно.

0.2 баллов
Дан ответ, что данные представляют из себя таблицу с 3221 строкой и 15 столбцами.

Рецензия №1
Рецензия №2
Рецензия №3

Часть 1. Задание 2. Для скольких стран есть информация о видеороликах в таблице (включая None)? (колонка country)

0.0 баллов
Пункт не выполнен / выполнен неправильно.

0.2 баллов
Показано, что в таблице есть видеоролики для 20 уникальных стран.

Рецензия №1
Рецензия №2
Рецензия №3

Часть 1. Задание 2. Сколько есть видеороликов с разрешенными и запрещенными комментариями? (колонка comments_disabled). Под видеороликом считаем 1 запись в таблице.

0.0 баллов
Пункт не выполнен / выполнен неправильно.

0.2 баллов
Показано, что в таблице 3193 записей с разрешенными комментариями (comments_disabled = False) и 28 записей с запрещенными комментариями (comments_disabled = True).

Рецензия №1
Рецензия №2
Рецензия №3

Часть 1. Задание 2. Какое максимальное число просмотров у видеороликов, попавших в таблицу? (колонка view_count)

0.0 баллов
Пункт не выполнен / выполнен неправильно.

Рецензия №3

0.2 баллов
Показано, что максимальное число просмотров 123034293.

Рецензия №1
Рецензия №2

Часть 1. Задание 3. Есть ли в данных пропуски? В каких колонках?

0.0 баллов
Пункт не выполнен / выполнен неправильно.

0.2 баллов
Показано, что в таблице есть пропуски в колонках country и description.

Рецензия №1
Рецензия №2
Рецензия №3

Часть 1. Задание 3. Заполните пропуски пустой строкой для текстовых колонок, и удалите строки с пропусками для колонок с категориальными значениями.

0.0 баллов
Пункт не выполнен / выполнен неправильно.

0.3 баллов
В колонке description пропуски заполнены пустой строкой, а строки с пропущенными значениями в колонке country удалены.

Рецензия №1
Рецензия №2
Рецензия №3

Часть 1. Задание 4. Какого она типа?

0.0 баллов
Пункт не выполнен / выполнен неправильно.

0.1 баллов
Показано, что колонка publishedAt имеет тип object.

Рецензия №1 Рецензия №2
Рецензия №3

Часть 1. Задание 4. Измените тип колонки на datetime. Покажите, что после изменения колонка действительно имеет тип datetime.

0.0 баллов
Пункт не выполнен / выполнен неправильно.

0.2 баллов
Изменен тип колонки publishedAt на datetime.

Рецензия №1
Рецензия №2
Рецензия №3

Часть 1. Задание 4. В какой месяц было добавлено больше всего видеороликов?

0.0 баллов
Пункт не выполнен / выполнен неправильно.

0.3 баллов
Показано, что больше всего видеороликов было опубликовано в июле.

Рецензия №1 Рецензия №2
Рецензия №3

Часть 1. Задание 5. Переведите строки в нижний регистр и обрежьте ненужные пробелы справа и слева.

0.0 баллов
Пункт не выполнен / выполнен неправильно.

0.2 баллов
Значения в колонке description переведены в нижний регистр, пробельные символы справа и слева удалены.

Рецензия №1
Рецензия №2
Рецензия №3

Часть 1. Задание 5. Замените url-адреса в описаниях на специальный тэг <URL>.

0.0 баллов
Пункт не выполнен / выполнен неправильно.

0.3 баллов
Все url-адреса в колонке description заменены на <URL>.

Рецензия №1 Рецензия №2
Рецензия №3

Часть 1. Задание 5. Выведите полностью description для видеоролика с video_id = 'Oxexirlmdig'.

0.0 баллов
Пункт не выполнен / выполнен неправильно.

0.2 баллов
Показано, что значение description для видео с video_id 'Oxexirlmdig' равно: 'music line & lyrics : ghetto geasy & majkmusic produced : bigdeal records mix & master: bigdeal recordsavailable on:spotify: <URL> <URL> <URL>.

Рецензия №1
Рецензия №2
Рецензия №3

Часть 1. Задание 6. Какие 5 каналов в Италии (country='IT') опубликовали ролики с наибольшим числом комментариев (колонка comment_count)? .....

0.0 баллов
Пункт не выполнен / выполнен неправильно.

0.2 баллов
Выполнена сортировка по числу комментариев (колонка comment_count), в ответе присутствуют следующие 5 каналов: SMTOWN, Fleeting Films, FORMULA 1, TiniVEVO, PlayStation (не обязательно в таком порядке!). Вторая сортировка не выполнена

Рецензия №1
Рецензия №3

0.4 баллов
Выполнена сортировка по числу комментариев (колонка comment_count), в ответе присутствуют следующие 5 каналов: SMTOWN, Fleeting Films, FORMULA 1, TiniVEVO, PlayStation (не обязательно в таком порядке!). Выполнена вторая сортировка по кол-ву просмотров (колонка view_count) у видеороликов с выбранных каналов с самым большим числом комментариев / с самым большим числом просмотров.

Рецензия №2

Часть 1. Задание 7. Какой видеоролик был опубликован в наибольшем кол-ве стран? В качестве ответа приведите название видеоролика (колонка title).

0.0 баллов
Пункт не выполнен / выполнен неправильно.

0.4 баллов
Показано, что ролик с названием "Race Highlights | 2022 British Grand Prix" был опубликован в наибольшем числе стран.

Рецензия №1
Рецензия №2
Рецензия №3

Часть 1. Задание 8. Какое максимальное число просмотров (колонка view_count) у видеороликов, выпущенных во Франции или Италии (в колонке country значение FR или IT) в июле (колонка publishedAt)?

0.0 баллов
Пункт не выполнен / выполнен неправильно.

0.2 баллов
Показано, что максимальное число просмотров у видеороликов, выпущенных во Франции или Италии в июле, равно 82864422. Решение не в 1 строку

0.4 баллов
Показано, что максимальное число просмотров у видеороликов, выпущенных во Франции или Италии в июле, равно 82864422. Решение в 1 строку.

Рецензия №1
Рецензия №2
Рецензия №3

Часть 1. Задание 9. Посмотрите на названия каналов из 17 категории (колонка categoryId). Как вы думаете, что это за категория?

0.0 баллов
Пункт не выполнен / выполнен неправильно.

0.1 баллов
Дан ответ, что это категория связана со спортом (ответ аргументирован, например, выводом всех уникальных названий каналов из 17 категории).

Рецензия №1
Рецензия №2
Рецензия №3

Часть 1. Задание 9. Предположим, что была допущена ошибка, и все видеоролики из 17 категории, относятся к 16 категории. Исправьте нужные значения в колонке categoryId, после чего выведите все уникальные категории.

0.0 баллов
Пункт не выполнен / выполнен неправильно.

0.4 баллов
В строках, где значения 'categoryId' были равны 17, значения изменены на 16. После преобразований выведены все категории, среди которых нет значения 17, но есть 16.

Рецензия №1
Рецензия №2
Рецензия №3

Часть 1. Задание 10. Замените строки со значением '[none]' на пустую строку.

0.0 баллов
Пункт не выполнен / выполнен неправильно.

0.2 баллов
В строках, где значения 'tags' были равны '[none]', значения заменены на пустую строку.

Рецензия №1 | Рецензия №2
Рецензия №3

Часть 1. Задание 10. Какой самый популярный тэг? Обратите внимание, что у каждого видеоролика может быть более 1 тэга.

0.0 баллов
Пункт не выполнен / выполнен неправильно.

0.4 баллов
Дан ответ отличный от 'challenge', но ход решения, в целом, правильный. Значения в колонках tags были разбиты на уникальные значения (с помощью split по '|' и explode), но подсчёт выполняется по уникальным значениям колонки tags без предобработки (значения не приведены в нижний регистр, не удалены пробелы).

0.8 баллов
Дан ответ 'challenge'

Рецензия №1
Рецензия №2
Рецензия №3

Часть 2. Задание 1.

0.0 баллов
График не построен / неопрятный (не выполнены хотя бы 3 пункта из правил оформления).

0.5 баллов
Построен опрятный график.

Рецензия №1 | Рецензия №2
Рецензия №3

Часть 2. Задание 1.

0.0 баллов
График не построен / некорректный.

1.0 балл
Построен график типа line plot. График корректный: по оси x - дата публикации (publishedA), по оси y - кол-во просмотров у роликов, выпущенных в соответствующую дат. Допускаются решения, как с группировкой записей в таблице по video_id, так и без. При корректном построении пик приходится на начало июля.

Рецензия №1
Рецензия №2
Рецензия №3

Часть 2. Задание 2.

0.0 баллов
График не построен / неопрятный (не выполнены хотя бы 3 пункта из правил оформления).

0.5 баллов
Построен опрятный график.

Рецензия №1 | Рецензия №2
Рецензия №3

Часть 2. Задание 2.

0.0 баллов
График не построен / некорректный.

1.0 балл
Построен график типа bar plot. График корректный: по одной оси - 19 стран (country) (ИЛИ 20 с учетом None), по другой - кол-во опубликованных роликов. При корректном построении в стране ME наименьшее число опубликованных роликов.

Рецензия №1
Рецензия №2
Рецензия №3

Часть 2. Задание 3.

0.0 баллов
График не построен / неопрятный (не выполнены хотя бы 3 пункта из правил оформления).

0.5 баллов
Построен опрятный график.

Рецензия №1 | Рецензия №2
Рецензия №3

Часть 2. Задание 3.

0.0 баллов
График не построен / некорректный.

1.0 балл
Построен график типа scatter plot. График корректный: по одной оси - кол-во просмотров (view_count), по другой - кол-во комментариев (comments_count). Возможна логарифмическая шкала. Для данных есть фильтры country == 'FR' и comments_disabled == False.

Рецензия №1
Рецензия №2
Рецензия №3

Часть 3. Задача 7.

0.0 баллов
Графики не построены / неопрятные (не выполнены хотя бы 3 пункта из правил оформления хотя бы для одного графика).

0.5 баллов
Построены опрятные графики.

Рецензия №1
Рецензия №2
Рецензия №3

Часть 3. Задача 7.

0.0 баллов
Графики не построены / некорректные.

1.0 балл
Построены корректные графики сравнения для 1-3 функций: по оси x - размерность входных данных, по оси y - время работы функции, векторизованная и невекторизованная реализации сравниваются на одном графике, для них есть легенда.

Рецензия №3

1.5 балла
Построены корректные графики сравнения для 4-5 функций: по оси x - размерность входных данных, по оси y - время работы функции, векторизованная и невекторизованная реализации сравниваются на одном графике, для них есть легенда.

Рецензия №1

1.5 балла
Построены корректные графики сравнения для 4-5 функций: по оси x - размерность входных данных, по оси y - время работы функции, векторизованная и невекторизованная реализации сравниваются на разных графиках.

2.0 балла
Построены корректные графики сравнения для всех 6 функций: по оси x - размерность входных данных, по оси y - время работы функции, векторизованная и невекторизованная реализации сравниваются на одном графике, для них есть легенда.

Рецензия №2

Часть 3. Задача 7.

0.0 баллов
Нет выводов.

0.5 баллов
Есть выводы о том, что время работы векторизованных функций меньше, разница в скорости увеличивается с увеличением размерности входных данных.

Рецензия №1
Рецензия №2
Рецензия №3

Оцените, насколько понятным для вас был код автора

0.0 баллов
Все понятно.

Рецензия №1 Рецензия №2
Рецензия №3

0.0 баллов
В целом понятно, были неясные места.

0.0 баллов
Есть, куда расти.

Оцените, насколько понятным для вас было оформление ноутбука автора в целом

0.0 баллов Все понятно	Рецензия №1 Рецензия №2 Рецензия №3
0.0 баллов В целом понятно, были неясные места	---
0.0 баллов Есть, куда расти	---

Комментарии

Рецензия №1

Рецензия №2

Рецензия №3