

Домашнее задание «Функции и работа с данными»



Задание

[Вопросы по заданию](#)

Преподаватель: Константин Башевой

****Домашнее задание к лекции.**

Задание 1**

Напишите функцию, которая классифицирует фильмы из материалов занятия по следующим правилам:

- оценка 2 и меньше - низкий рейтинг
- оценка 4 и меньше - средний рейтинг
- оценка 4.5 и 5 - высокий рейтинг

Результат классификации запишите в столбец class

Задание 2

Используем файл keywords.csv.

Необходимо написать гео-классификатор, который каждой строке сможет выставить географическую принадлежность определенному региону. Т. е. если поисковый запрос содержит название города региона, то в столбце 'region' пишется название этого региона. Если поисковый запрос не содержит названия города, то ставим 'undefined'.

Правила распределения по регионам Центр, Северо-Запад и Дальний Восток:

geo_data = {

```
1 'Центр': ['москва', 'тула', 'ярославль'],
2
3 'Северо-Запад': ['петербург', 'псков', 'мурманск'],
4
5 'Дальний Восток': ['владивосток', 'сахалин', 'хабаровск']
```

}

Результат классификации запишите в отдельный столбец region.

Задание 3 (бонусное)

Есть мнение, что “раньше снимали настоящее кино, не то что сейчас”. Ваша задача проверить это утверждение, используя файлы с рейтингами фильмов из прошлого домашнего занятия (файл movies.csv и ratings.csv из базы grouplens.org.../movielens). Т. е. проверить верно ли, что с ростом года выпуска фильма его средний рейтинг становится ниже.

При этом мы не будем затрагивать субъективные факторы выставления этих рейтингов, а пройдемся по следующему алгоритму:

1. В переменную years запишите список из всех годов с 1950 по 2010.
2. Напишите функцию production_year, которая каждой строке из названия фильма выставляет год выпуска. Не все названия фильмов содержат год выпуска в одинаковом формате, поэтому используйте следующий алгоритм:
 1. для каждой строки пройдите по всем годам списка years
 2. если номер года присутствует в названии фильма, то функция возвращает этот год как год выпуска
 3. если ни один из номеров года списка years не встретился в названии фильма, то возвращается 1900 год
3. Запишите год выпуска фильма по алгоритму пункта 2 в новый столбец 'year'
4. Посчитайте средний рейтинг всех фильмов для каждого значения столбца 'year' и отсортируйте результат по убыванию рейтинга

важно: при копировании юпитер ноутбука изменяет кавычки на `, что может способствовать ошибки в коде. В данной ситуации необходимо их просто заменить.`

Решение

Зачет

https://github.com/PeterM-lab/PYDA1/blob/main/10_apply_df/Untitled.ipynb 17 апр. 2021

LINK

Вы загрузили решение 17 апр. в 20:28

Константин Башевой поставил(а) зачет 18 апр. в 17:10

**Константин Башевой**

ПРЕПОДАВАТЕЛЬ

18 апреля 2021 17:10

Добрый вечер, спасибо за решения.
Круто сделано.



Вам понравилось?

Вопросы по заданию

Задайте вопрос — вам помогут одноклассники и эксперты

ПМ

На какой вопрос вы хотите получить ответ?

Добавьте более подробное описание вашего вопроса, если необходимо

Спросить

Все вопросы

Помочь с ответом

ИС Ирина Спасская

Верно ли указан файл в задании 2?

В файле keywords.csv всего 2 колонки:

keyword shows

0 вк 64292779

1 одноклассники 63810309

где искать нужный файл?

Написать ответ

**Алексей Нагорский**

Ответ студента

Все верно указано, нужно этот файл смотреть, я тоже не сразу понял :)

