**\*\*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 = {

​

   'Центр': ['москва', 'тула', 'ярославль'],

​

   'Северо-Запад': ['петербург', 'псков', 'мурманск'],

​

   '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 ratings.csv из базы https://grouplens.org/datasets/movielens). Т. е. проверить верно ли, что с ростом года выпуска фильма его средний рейтинг становится ниже.

​

При этом мы не будем затрагивать субьективные 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:

- для каждой строки пройдите по всем годам списка years

- если номер года присутствует в названии фильма, то функция возвращает этот год как год выпуска

- 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.\*\***

​