**Задание:**

**Преподаватель:**Константин Башевой

**Домашнее задание к лекции**

**Задание 1**  
Напишите функцию, которая классифицирует фильмы из материалов занятия по правилам:

* оценка 2 и ниже — низкий рейтинг;
* оценка 4 и ниже — средний рейтинг;
* оценка 4.5 и 5 — высокий рейтинг.

Результат классификации запишите в столбец class.

**Задание 2**  
Используйте файл keywords.csv.

Нужно написать гео-классификатор, который каждой строке сможет выставить географическую принадлежность определённому 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 = {

12345'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 movies.csv и ratings.csv [из базы](https://grouplens.org/datasets/movielens). Нужно проверить, верно ли, что с ростом года выпуска фильма его средний рейтинг становится ниже.

Вы не будете затрагивать субьективные факторы выставления этих рейтингов, а пройдётесь по алгоритму:

1. В переменную years запишите список из всех годов с 1950 по 2010 года.
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 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.

Все файлы находятся в архиве «Дополнительные файлы к лекции и дз».

**Важно! При копировании юпитер ноутбука изменяет кавычки ’ на ` , что может вызывать ошибку в коде. В этом случае их нужно просто заменить.**

**Решение:**

<https://colab.research.google.com/drive/1P8_Qe9DmwA_9j5JPZFXwP4GCdNJX6DoX?usp=sharing>

Изображение выглядит как текст

Автоматически созданное описание