

1 Dataset Movies

Наборы содержат рейтинги фильмов, метаданные о фильмах (жанр и год выхода) и демографические данные о пользователях (возраст, почтовый индекс, пол и род занятий). Такие данные часто представляют интерес для разработки систем рекомендаций, основанных на алгоритмах машинного обучения.

Набор MovieLens 1M содержит 1 миллион рейтингов 4000 фильмов, предоставленных 6000 пользователей. Данные распределены по трем таблицам: рейтинги, информация о пользователях и информация о фильмах.

Задачи по обработке датасета:

1. Загрузить каждую таблицу в отдельный объект `DataFrame` с помощью метода **`pd.read_table`**:
2. Объединить все данные в одну таблицу, применяя функцию **`pd.merge`** из библиотеки `pandas`. Сначала объединим `ratings` с `users`, а затем результат объединим с `movies`. `Pandas` определяет, по каким столбцам объединять, ориентируясь на совпадение имен.
3. Получить средние рейтинги всех фильмов при группировке по полу. Получить средние рейтинги всех фильмов при группировке по возрасту (использовать метод **`pd.pivot_table`**).

4. Отсортировать полученные средние рейтинги при группировке по полу в порядке убывания популярности у женской аудитории. Отсортировать полученные средние рейтинги при группировке по возрасту в порядке убывания популярности у самой возрастной аудитории.
5. С помощью метода **groupby()** получить серию, содержащую количество оценок для каждого наименования фильма. Вывести наименования, получившие не менее 250 оценок.
6. Получить данные фильмов, по которым мужская и женская аудитории сильнее всего разошлись в оценках (для этого можно добавить столбец, содержащий разность средних в датафрейм, а затем отсортировать по нему). Вывести первые 15 фильмов, которым женская аудитория поставила высокие оценки, а мужская - низкие, и наоборот.