

Онлайн-кинотеатр KION — часть продуктовой экосистемы МТС. Включает более 200 ТВ-каналов, тысячи фильмов, сериалов, мультфильмов и документальных проектов.

В датасете представлены данные, собранные на основе анализа поведения пользователей сервиса с 13 марта 2021 года по 22 августа 2022 года.

Датасет включает информацию о взаимодействии пользователей с контентом, демографическую информация о пользователях и мета-информацию о фильмах.

users.csv

- `user_id` — ID пользователя
- `age` — возрастная группа пользователя, строка вида “M_N”
- данный признак — результат работы модели
- `18_24` — от 18 до 24 лет включительно
- `25_34` — от 25 до 34 лет включительно
- `35_44` — от 35 до 44 лет включительно
- `45_54` — от 45 до 54 лет включительно
- `55_64` — от 55 до 64 лет включительно
- `65_inf` — от 65 и старше
- `sex` — пол пользователя
- данный признак — результат работы модели

- M — мужчина
- Ж — женщина
- income — доход пользователя, строка вида "M_N"
- данный признак — результат работы модели
- income_0_20
- income_20_40
- income_40_60
- income_60_90
- income_90_150
- income_150_inf
- kids_flg — флаг «наличие ребенка»
- данный признак — результат работы модели

items.csv

В данном файле содержится информация об объектах (фильмы/сериалы):

- item_id — ID контента
- content_type — Тип контента (фильм, сериал)
- title — Название на русском
- title_orig — Название оригинальное
- genres — Жанры из источника (онлайн-кинотеатры)
- countries — страны
- for_kids — флаг «контент для детей»
- age_rating — возрастной рейтинг

- studios — студии
- directors — директора
- actors — актеры
- keywords — ключевые слова
- description — описание
- valid_from_dttm — дата, с которой контент доступен на KION
- rating_kp — рейтинг на Кинопоиске

interactions.csv

В данном файле содержится информация о взаимодействиях пользователей с контентом:

- user_id — ID пользователя
- item_id — ID контента
- last_watch_dt — Дата последнего просмотра
- total_dur — Общая продолжительность всех просмотров данного контента в секундах
- content_type — Тип контента (фильм, сериал)