

Лабораторная работа №3

Задание 1.

Цель: Закрепление основных команд первичного анализа данных в Pandas.

Рекомендуется использовать библиотеку **pandas** и функции **merge()**, **explode()**, **groupby()**

Исследовательская группа GroupLens Research предлагает данные о рейтингах фильмов, предоставленных пользователями сайта MovieLens в конце 1990-х – начале 2000-х (ml-1m.zip). Данные распределены по трем таблицам: рейтинги фильмов, метаданные о фильмах (жанр и год выхода) и демографические данные о пользователях (возраст, почтовый индекс, пол и род занятий). Данные о возрасте и роде занятий кодируются целыми числами, а расшифровка приведена в прилагаемом к набору данных файлу **README**. Поскольку один фильм может относиться к нескольким жанрам, содержащиеся в нем значения разделены «|».

- 1) Требуется вычислить средние рейтинги для фильма с ID = № варианта в зависимости от пола;
- 2) Отобрать фильмы, получившие не менее 250 оценок, и для них определить средний рейтинг в зависимости от пола;

Результат вывести в виде DataFrame с колонками

```
['MovieID', 'Title', 'Rating_F', 'Rating_M']
```

- 3) Найти фильмы, оказавшиеся в топ 5 у зрителей-женщин и худшие 5 у зрителей-мужчин. Найти фильмы (топ 5), по которым мужчины и женщины сильнее всего разошлись в оценках;
- 4) Определить фильмы (топ 5), вызвавшие в среднем наибольшее разногласие у зрителей независимо от пола;
- 5) Определить средний рейтинг каждого жанра для различных возрастов. Установите, фильмы какого жанра в каждой категории возрастов имеют наибольший и наименьший рейтинги.

Если фильм относится к нескольким данным, то их надо разделить и сформировать несколько строчек с этим фильмом, но с различными жанрами, например:

UserID	MovieID	Rating	Timestamp	Gender	Age	Occupation	Zip-code	Title	Genres
1	661	3	978302109	F	1	10	48067	James and the Giant Peach (1996)	Animation Children's Musical

необходимо преобразовать в

UserID	MovieID	Rating	Timestamp	Gender	Age	Occupation	Zip-code	Title	Genres
1	661	3	978302109	F	1	10	48067	James and the Giant Peach (1996)	Animation
1	661	3	978302109	F	1	10	48067	James and the Giant Peach (1996)	Children's
1	661	3	978302109	F	1	10	48067	James and the Giant Peach (1996)	Musical

Задание 2.

Цель: Закрепление основных основных понятий теории вероятностей.

Рекомендуется использовать библиотеки **SciPy** и **matplotlib**

1) Случайная величина X имеет нормальное распределение, $M[X] = 10$ и $D[X] = 4$.

- Построить функцию и плотность распределения вероятностей для нормированной нормально распределенной случайной величины.
- Построить функцию и плотность распределения вероятностей для X .
- Вычислить вероятности $P(X < 11)$, $P(X > 5)$, $P(4 < X < 12)$.
- Отобразить, вероятности на графиках функции и плотности вероятностей.
- Проверить выполнение правила «3-сигм» для X , отобразить это правило на графике плотности вероятностей

2) Случайная величина Y имеет экспоненциальное распределение, $M[Y] = 2$.

- Построить для нее функцию и плотность распределения вероятностей.
- Вычислить вероятности $P(X < 3)$, $P(X > 2)$, $P(1 < X < 2.5)$.
- Отобразить, вероятности на графиках функции и плотности вероятностей.
- Проверить выполнение правила «3-сигм» для Y , отобразить это правило на графике плотности вероятностей.

Отчет устный по результатам выполненной работы.