

Данные для данного проекта были взяты с сайта [kaggle.com](https://www.kaggle.com)

Изначальный набор данных включает в себя 14952 аниме, проанализированным по 24 переменным.

Для достижения своих целей я использовал всего 5 переменных, так как считаю, что их вполне достаточно для анализа, а именно:

1. **“Name”**. Название аниме.
2. **“Score”**. Рейтинг аниме (учитывался тот рейтинг, который был указан в исходном наборе данных).
3. **“Source”**. Источник аниме (создано на основе манги, другого литературного жанра или создано с нуля).
4. **“Studios”**. Студия, выпустившая аниме.
5. **“Watching”**. Количество людей, просмотревших аниме.

В рамках своего исследования я не пытался выявить какие-либо закономерности между переменными, напротив, я занимался простой фиксацией данных. На мой взгляд, это тоже является полезным навыком для аналитика данных. Информация, полученная в ходе моего исследования, по моему мнению, может быть использована как основа для создания персонализированных подборок в различных онлайн-кинотеатрах, а также критерием, согласно которому различные государственные и негосударственные акторы могут осуществлять финансирование различных студий, компаний по производству кинопродукции, поощряя развитие приоритетных для них отраслей и жанров.

В рамках моего исследования были использованы:  
**Библиотека pandas** – для извлечения данных из файла CSV-формата, создание подходящих выборок на основе необходимых переменных.

**Библиотека NumPy** – для создания одномерных массивов, которые в дальнейшем успешно были визуализированы (после присвоения каждому массиву своего уникального значения).

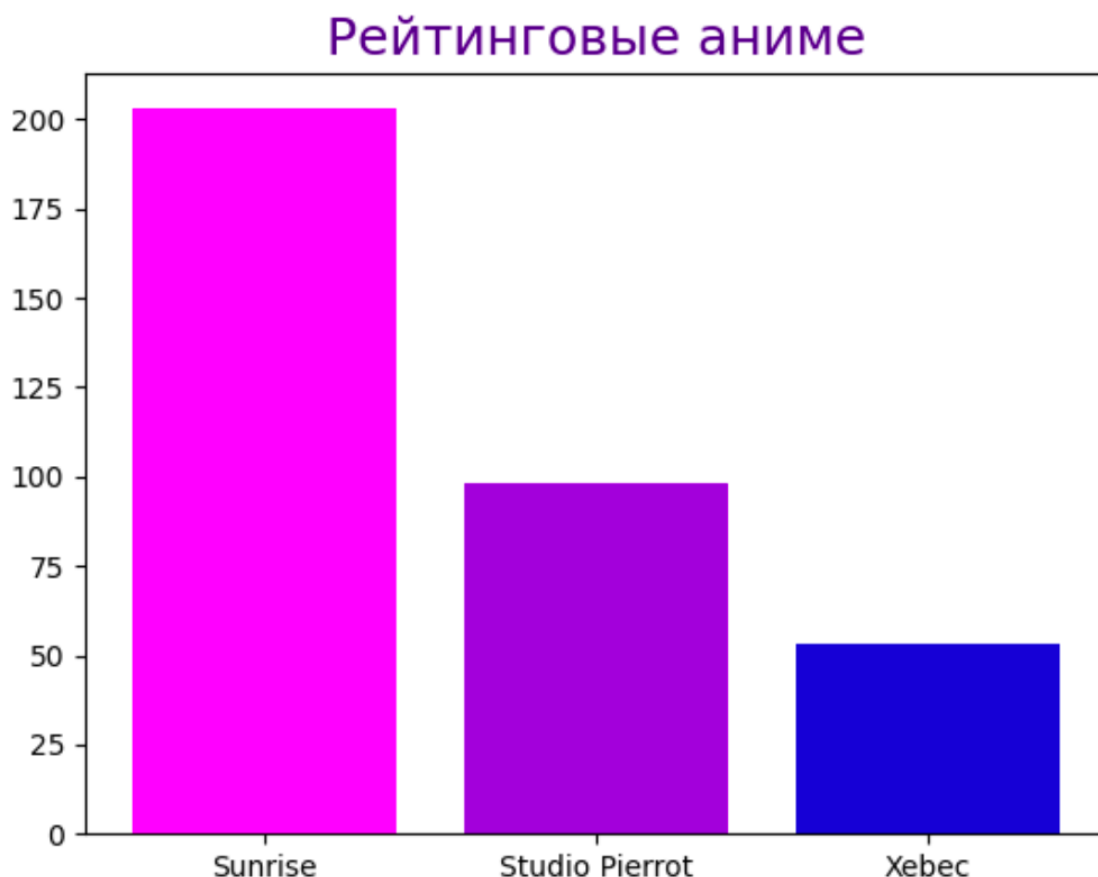
**Библиотека Matplotlib** – для визуализации полученных данных в виде столбчатой, круговой диаграммы, stem-графика.

Из всех пяти переменных, взятых для моего исследования, единственной неизменной может быть переменная “**Studios**”. Таким образом, в рамках моего исследования я анализировал аниме таких студий как **Sunrise**, **Studio Pierrot** и **Xebec**, так как эти студии при предварительном просмотре изначального набора данных чаще других встречались в таблице.

Наиболее интересным мне показалось проанализировать:

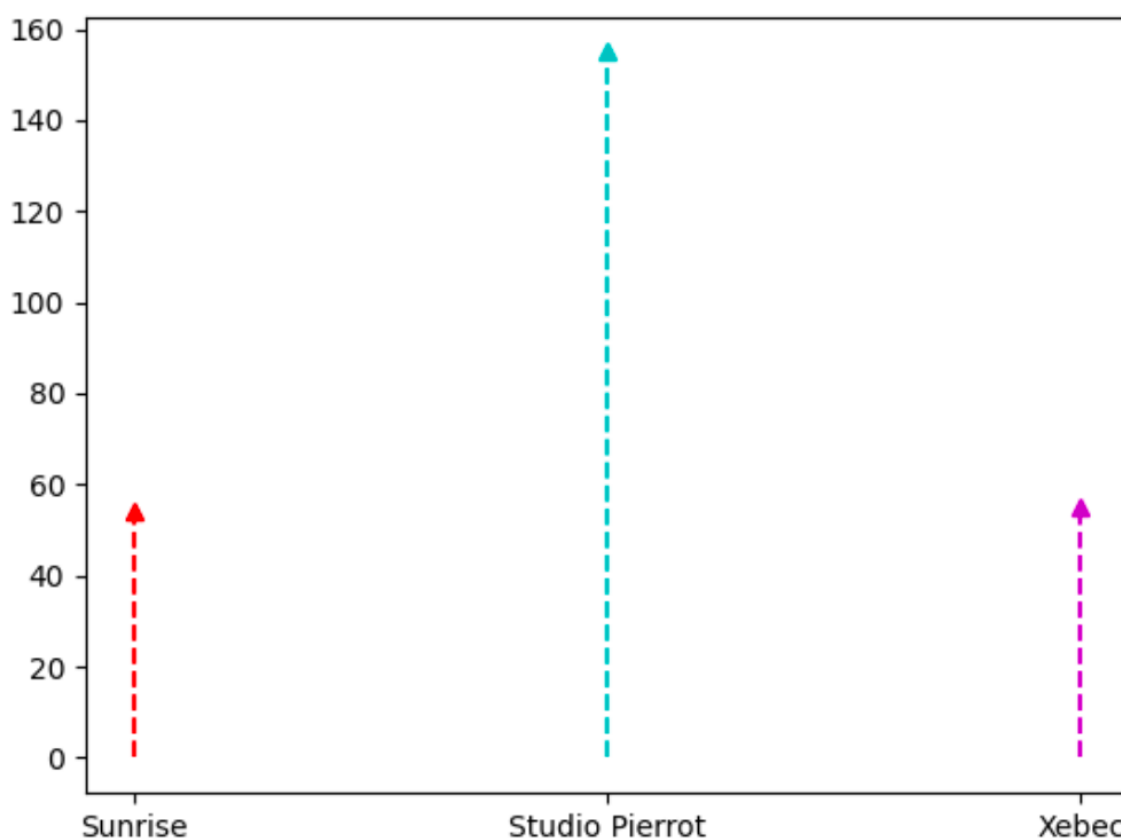
1. Сколько аниме с рейтингом  $\geq 7$  выпустили данные студии?
2. Сколько аниме, оригиналом которых служила манга, выпустили данные студии?
3. Сколько аниме, которые посмотрело  $\geq 10\,000$  человек, выпустили данные студии?

В результате анализа мне удалось узнать следующее:



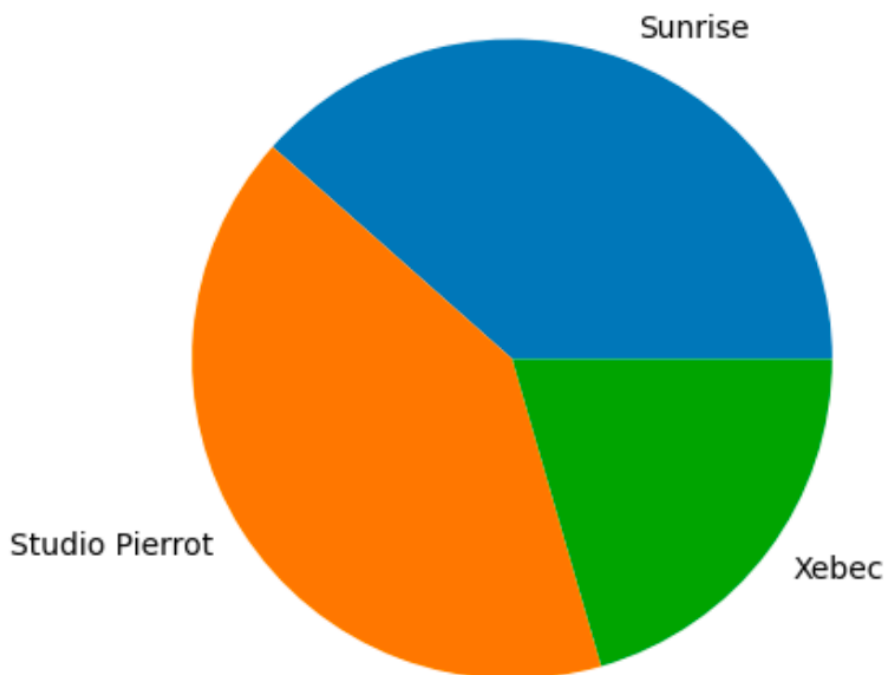
Наибольшее количество аниме с рейтингом  $\geq 7$  выпустила студия Sunrise. Следовательно, при составлении персональной подборки аниме в любом онлайн-кинотеатре аниме студии Sunrise могут занимать лидирующие позиции, поскольку государственные или негосударственные акторы могут материально поощрять студию, выпускающую большее количество рейтинговых аниме, чем другие. Таким образом, по данному параметру студия Sunrise значительно опережает своих конкурентов по выборке (больше, чем две другие студии вместе взятые).

## Манга - источник аниме



На данном графике по сравнению с предыдущим наблюдается совершенно иная картина. Здесь видно, что Studio Pierrot выпустила большее количество аниме, основанных на манге, чем другие студии из выборки вместе взятые. Данная информация может быть полезной для издательств и авторов различных манг. На мой взгляд, студия Studio Pierrot - наиболее удачный выбор для сотрудничества, поскольку именно она выпускает больше аниме по мангам, чем ее конкуренты.

## Количество просмотров аниме



Несмотря на то, что данная круговая диаграмма достаточно информативна, полученные данные не позволяют выявить явного преимущества у той или иной студии. Можно отметить лишь то, что Sunrise и Studio Pierrot выпустили больше аниме с количеством просмотров  $\geq 10\,000$  человек, чем студия Xebec. Данная информация может свидетельствовать не о необходимости поддержки двух данных студий, а может говорить о небольшом штате сотрудников студии Xebec, ограниченном финансировании и т.п. Таким образом, данный график лишь фиксирует преимущество двух вышеуказанных студий над третьей. Для того, чтобы понять, насколько высоко качество аниме у той или иной студии, необходимо собирать большее количество данных.

**ВЫВОД:** Данный проект показал, что библиотека pandas хорошо подходит не только для построения графиков зависимостей между различными переменными, но и просто для визуализации простых табличных данных, из которых без особого труда можно извлечь полезную информацию (библиотека NumPy была использована в данном проекте лишь для создания одномерного массива для упрощения процесса визуализации в библиотеке Matplotlib). Тем не менее, для более глубокого анализа и построения линии тренда нужны более сложные математические модели и более продвинутые возможности библиотек Python.