

Датасет находится в файле: **anime.csv**.

Набор данных содержит информацию о 1992 различных аниме.

1. Считайте файл
2. Выведите первые несколько строк (например, 10) датасета
3. Выведите информацию о колонках датасета: название, тип значений (целочисленный, вещественный, строковый и т.д.). Также для каждой колонки укажите, она содержит числовые или категориальные значения.  
Зам: Про разные типы значений можно прочитать по ссылке:  
<https://medium.com/nuances-of-programming/статистические-типы-данных-используемые-в-машинном-обучении-b8891039b09>
4. Приведите названия колонок к единому виду: все символы в названии колонок должны быть в нижнем регистре, между словами в названии должен быть символ «\_» (изменения делать с помощью `dataframe.columns`)
5. Для каждой колонки, содержащей числовые значения, построить статистики: min, max, mean, median, std, 25, 75, 90 перцентиль (`describe`)
6. Для каждой из колонок, содержащих категориальные признаки, вывести все возможные значения с количеством повторений этих значений.
7. Проанализировать значения во всех столбцах. Проверить, есть ли пропуски в данных. Если есть, то предложите один из способов устранения пропусков. Объясните, почему заполнили пропуски определённым образом или почему не стали этого делать.
8. Визуализируйте простые зависимости
  - a. Построить столбчатую диаграмму, описывающую зависимость между названием компаний (*Production Company*) и количеством выпущенных этой компанией аниме. В левой части диаграммы отобразите компании с наименьшим количеством аниме, в правой - с наибольшим. Какая компания выпустила больше всех аниме?
  - b. Каждое аниме состоит из нескольких эпизодов (*Number of Episodes*). Определите количество аниме с одинаковым количеством эпизодов и визуализируйте полученные данные в виде столбчатой диаграммы. Какое количество эпизодов чаще всего встречается?
  - c. Определите топ-3 популярных источника (*Source*) для аниме (пример: *Manga*, *Light Novel*, *Game* и т.д.) и визуализируйте полученные данные в виде столбчатой диаграммы. Какой источник наиболее популярный?
  - d. У каждого аниме есть тема (*Theme*). Определите количество аниме по каждой теме, отсортируйте в порядке возрастания и визуализируйте в виде столбчатой диаграммы. Какая тема наиболее популярная?
  - e. У каждого аниме есть дата выхода (*Airdate*). Определите количество аниме, вышедших в каждый год, постройте столбчатую диаграмму или линейный график. Какой год был самым богатым на выпущенные аниме?
9. Для каждой компании (*Production Company*) определите средний рейтинг (*Rating*) выпущенных данной компанией аниме (*mean*), отсортируйте все компании по возрастанию среднего рейтинга, постройте столбчатую диаграмму. Какие 3 компании выпускают самые популярные аниме?
10. У каждого аниме есть рейтинг (*Rating*). Рассмотрим интервалы рейтингов: [0, 1), [1, 2), [2, 3), [3, 4), [4, 5), [5, 6), [6, 7), [7, 8), [8, 9), [9, 10). Для каждого интервала

определите количество аниме с рейтингом из данного интервала. Полученные данные визуализируйте в виде столбчатой диаграммы или линейного графика. Какой интервал рейтинга оказался самым большим?

11. У каждого аниме есть жанр (*Genre*), который состоит из нескольких тегов (например: *Action, Adventure, Fantasy*) и тема (*Theme*), которая тоже состоит из нескольких тегов (например: *Historical, Military*). Определить, есть ли зависимость рейтинга от жанра, от темы аниме? Для определения зависимости рейтинга от жанра, можно, например, воспользоваться разбиением из п.10 и определить, какие теги есть только у самых популярных аниме, возможно именно поэтому они стали популярными. Для определения зависимости рейтинга от темы можете поступить так же, как с жанром. Вы также можете предложить свои способы определения зависимости рейтинга от жанра и темы.
12. Будем считать, что количество людей, оценивших аниме (*Voter*) – это количество зрителей. Постройте линейный график зависимости рейтинга аниме от количество зрителей. Есть ли корреляция между этими величинами?
13. Постройте 2 гипотезы. Обоснуйте или опровергните предложенные гипотезы.
14. Напишите общие выводы.