Состав датасета:

Файл `logs.csv` - логи просмотров за неделю. Содержат следующие поля:

1. event\_timestamp – таймстемп события

2. video\_id – уникальный ID видео

3. watchtime – время просмотра в секундах

4. region – регион расположения пользователя

5. city – город расположения пользователя

6. user\_id – уникальный ID пользователя

Файл `video\_stat.csv` - содержит информацию и статистику по каждому видео из логов:

1. video\_id – уникальный ID видео

2. v\_pub\_datetime – дата публикации видео на платформе. Момент, когда видео стало доступно на Rutube в публичном доступе

3. v\_total\_comments – количество комментариев под видео-если пользователь не прокоментировал видео у которого большая доля комментариев, то сильнее уменьшаем ему коэф того что ему понравится эта категория

4. v\_year\_views – полное число плеер стартов на видео

5. v\_month\_views – число плеер стартов на видео за последний месяц

6. v\_week\_views – число плеер стартов на видео за последнюю неделю

7. v\_day\_views – число плеер стартов на видео за последний день

8. v\_likes – число положительных эмоций на видео в данный момент времени

9. v\_dislikes – число отрицательных эмоций на видео в данный момент времени

10. v\_duration – длительность видео

11. v\_cr\_click\_like\_7\_days – конверсия плеер старта в положительную эмоцию за последние 7 дней

12. v\_cr\_click\_dislike\_7\_days – конверсия плеер старта в отрицательную эмоцию за последние 7 дней

13. v\_cr\_click\_vtop\_7\_days – конверсия плеер старта в нажатие «втоп» за последние 7 дней

14. v\_cr\_click\_long\_view\_7\_days – конверсия плеер старта в долгий просмотр за последние 7 дней

15. v\_cr\_click\_comment\_7\_days – конверсия плеер старта в комментарий за последние 7 дней

16. v\_cr\_click\_like\_30\_days – конверсия плеер старта в положительную эмоцию за последние 30 дней

17. v\_cr\_click\_dislike\_30\_days – конверсия плеер старта в отрицательную эмоцию за последние 30 дней

18. v\_cr\_click\_vtop\_30\_days – конверсия плеер старта в нажатие «втоп» за последние 30 дней

19. v\_cr\_click\_long\_view\_30\_days – конверсия плеер старта в долгий просмотр за последние 30 дней

20. v\_cr\_click\_comment\_30\_days – конверсия плеер старта в комментарий за последние 30 дней

21. v\_cr\_click\_like\_1\_days – конверсия плеер старта в положительную эмоцию за последний день

22. v\_cr\_click\_dislike\_1\_days – конверсия плеер старта в отрицательную эмоцию за последний день

23. v\_cr\_click\_vtop\_1\_days – конверсия плеер старта в нажатие «втоп» за последний день

24. v\_cr\_click\_long\_view\_1\_days – конверсия плеер старта в долгий просмотр за последний день

25. v\_cr\_click\_comment\_1\_days – конверсия плеер старта в комментарий за последний день

26. title – заголовок видео

27. description – описание видео

28. v\_avg\_watchtime\_1\_day – средний watchtime на видео за 1 день

29. v\_avg\_watchtime\_7\_day – средний watchtime на видео за 7 дней

30. v\_avg\_watchtime\_30\_day – средний watchtime на видео за 30 дней

31. v\_frac\_avg\_watchtime\_1\_day\_duration – средний watchtime на видео за 1 день в секундах делить на длительность видео в секундах

32. v\_frac\_avg\_watchtime\_7\_day\_duration – средний watchtime на видео за 7 дней делить на длительность видео в секундах

33. v\_frac\_avg\_watchtime\_30\_day\_duration – средний watchtime на видео за 30 дней делить на длительность видео в секундах

34. v\_category\_popularity\_percent\_7\_days – доля плеер стартов категории данного видео относительно всех плеер стартов за последние 7 дней

35. v\_category\_popularity\_percent\_30\_days – доля плеер стартов категории данного видео относительно всех плеер стартов за последние 30 дней

36. v\_long\_views\_1\_days – число плеер стартов с последующим "долгим" просмотром за 1 день

37. v\_long\_views\_7\_days – число плеер стартов с последующим "долгим" просмотром за 7 дней

38. v\_long\_views\_30\_days – число плеер стартов с последующим "долгим" просмотром за 30 дней

39. author\_id – ID автора видео

40. category\_id – ID категории видео

Определения:

1. Долгий просмотр

Если видео больше 5 минут, то долгим просмотром считается если watchtime > 0.25 \* duration

Если видео меньше 5 минут, то долгим просмотром считается watchtime > 30 секунд