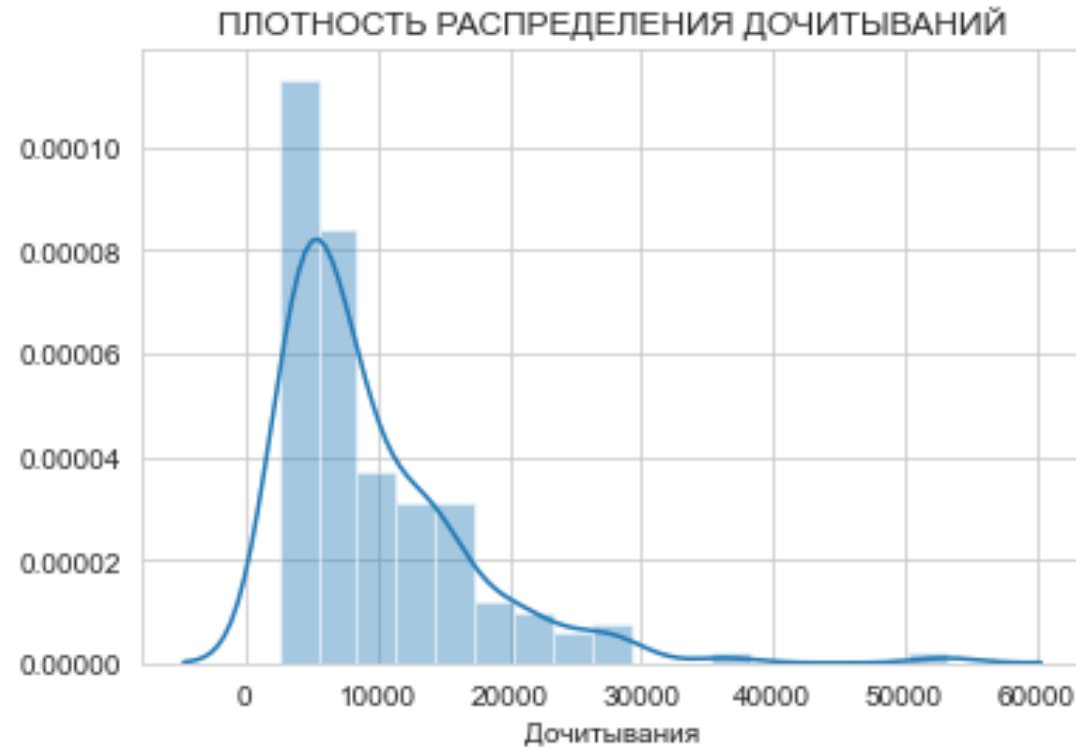


# Анализ больших массивов данных на примере параметров статей интернет-блога на платформе Яндекс.Дзен

Все графики созданы с помощью  
программного кода на языке Python

# График плотности распределения



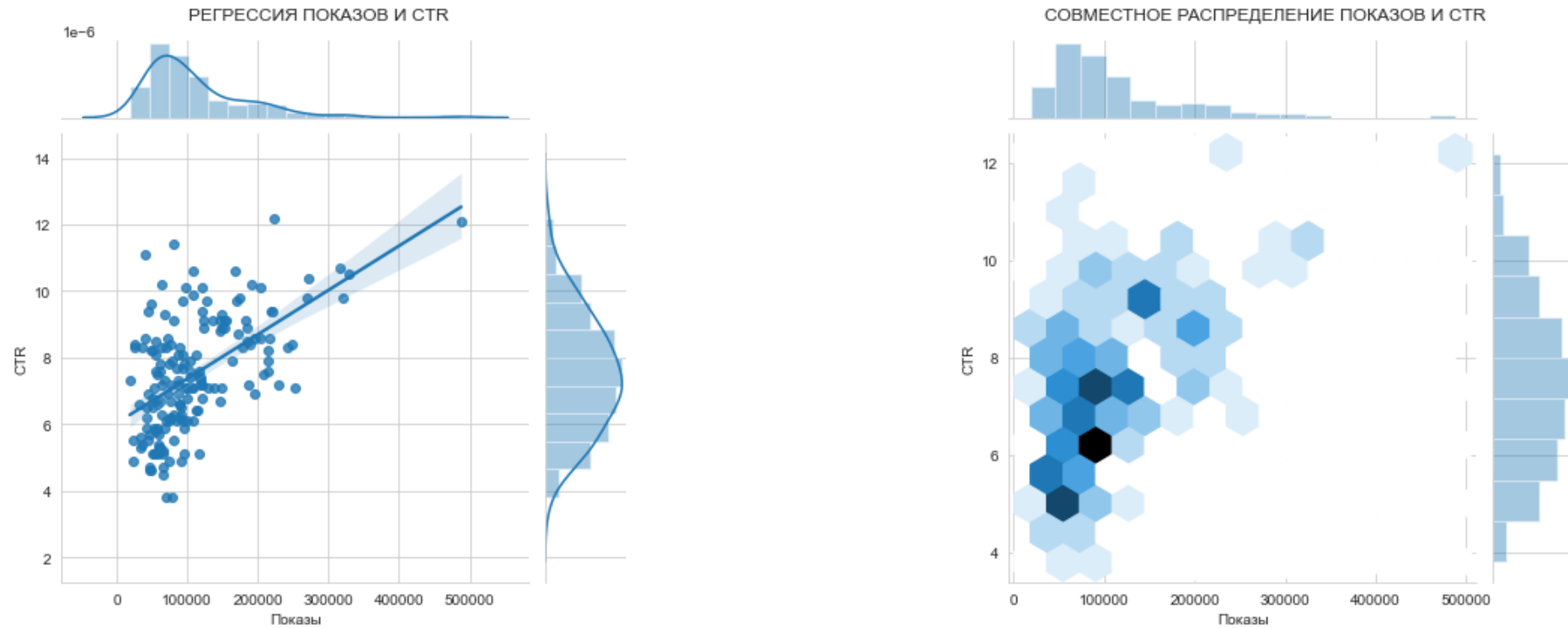
Из графика видно, что, в основном, статьи блога набирают до 10000 дочитываний (прочитаны полностью), более точно от 4 до 8 тыс.

# Тепловая карта зависимости параметров



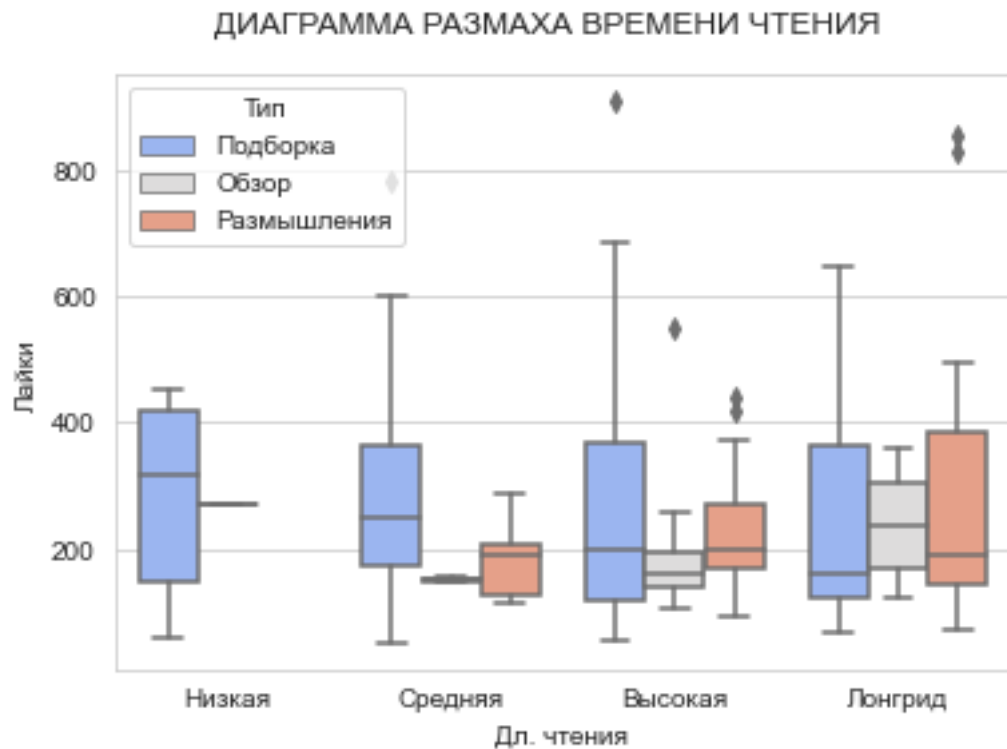
От «холодного» к «горячему» увеличивается зависимость одного параметра от другого. Так, мы видим, что дочитывания больше всего зависят от количества лайков на публикациях. А количество комментариев меньше всего зависит от уровня CTR

# Графики регрессии и распределения

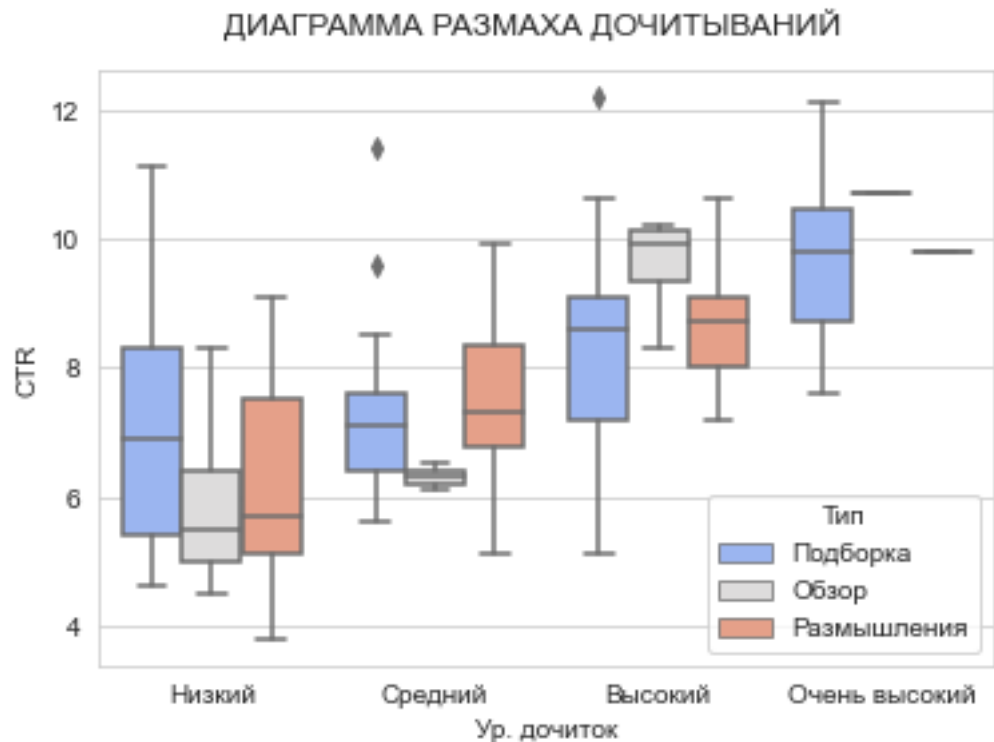


Из графика видна прямая зависимость показов от CTR. На графике слева представлена линия тренда, показывающая увеличение показов при повышении CTR. Основной объем статей набирает до 100 тыс. показов, имея CTR 6-8%. На графике справа представлена тепловая карта совместного распределения параметров

# Диаграммы размаха

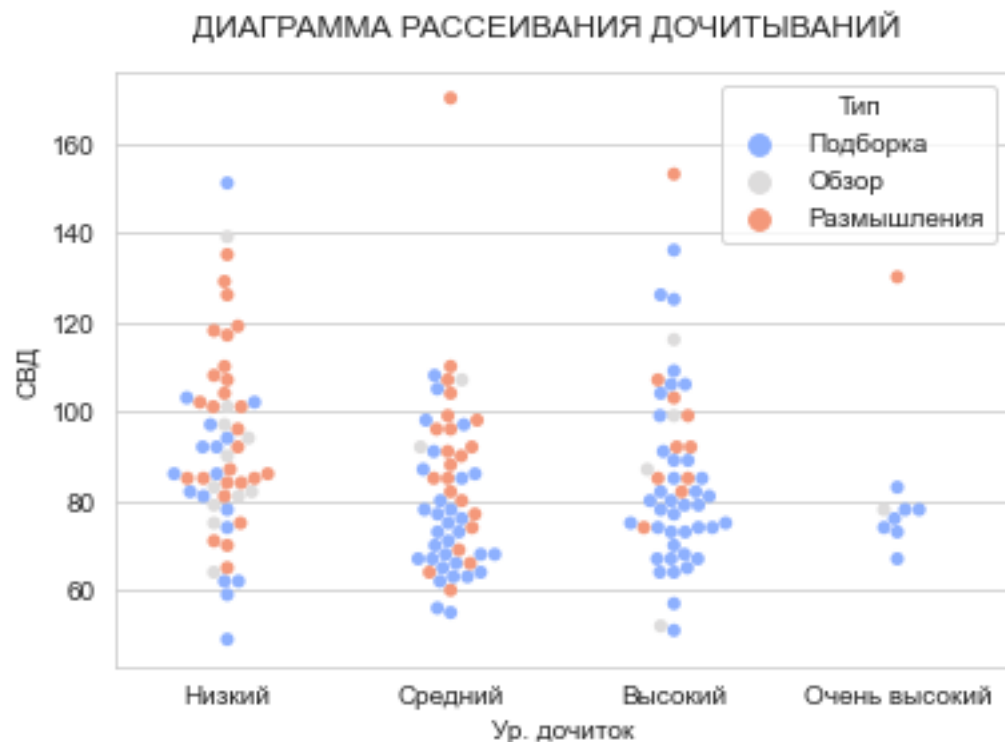


Можно отметить, что больше всего лайков набирают быстро читаемые подборки (из-за удобства потребления простого контента) и очень длинные статьи-размышления (из-за высокой вовлеченности и интереса к авторскому мнению)



Можно отметить, что размышления не получают высокий CTR (контент не для массового потребления). А самый большой CTR имеют подборки с очень высоким уровнем дочитываний

# Диаграммы рассеивания

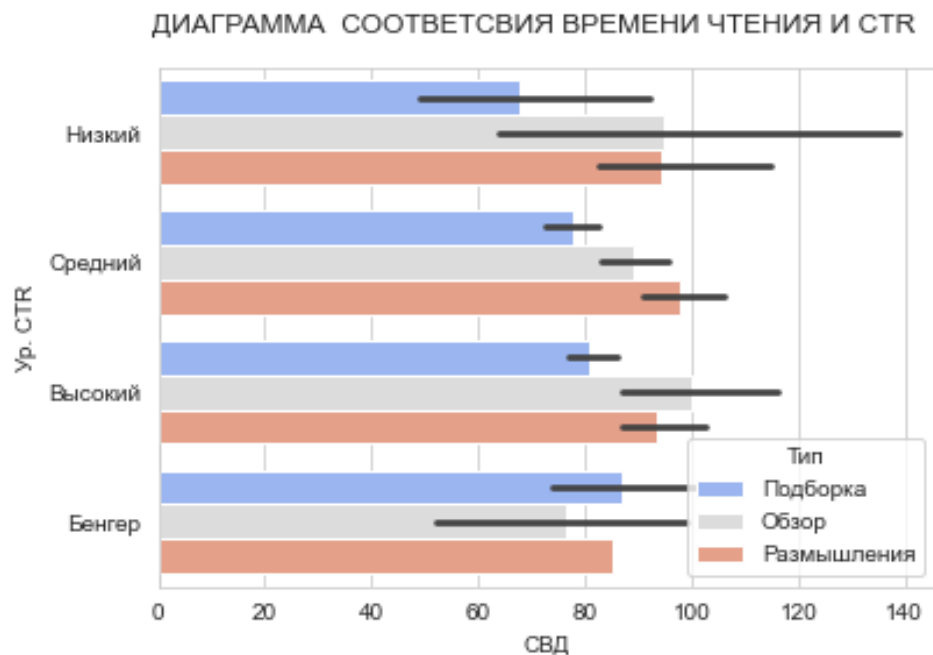


Так же из графика видно, что размышления имеют самое длительное среднее время дочитываний (СВД) и отсюда – наименьший уровень дочитываний. Подборки, в основном, на среднем и высоком уровне дочитываний при низком СВД



На данном графике видно, что у «лонгридов» явно хороший уровень вовлеченности относительно лайков. Также стоит отметить, что обзоры имеют, как правило, высокую и очень высокую длительности чтения

# Диаграммы соответствия



Из графика видно, что если для подборок СВД при увеличении CTR растет, то для размышлений, наоборот, снижается. Однако показатель СВД для подборок и размышлений в категории «Бенгер» (очень высокий CTR) примерно одинаковый. Это говорит о сильной зависимости CTR от длительности чтения статьи (пользователи должны дочитывать, для этого не стоит писать слишком большие текста

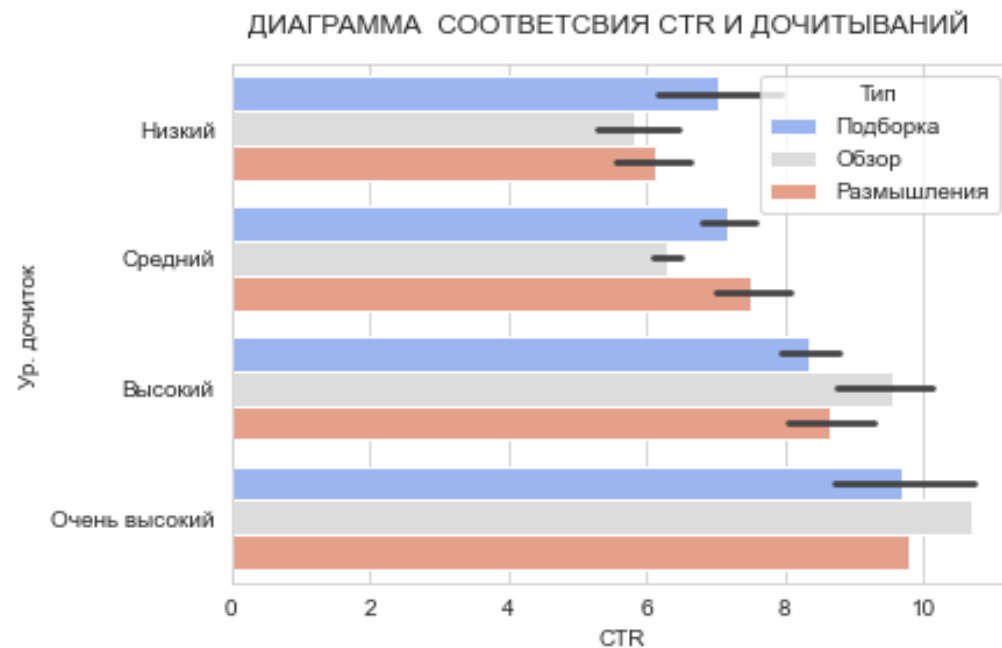
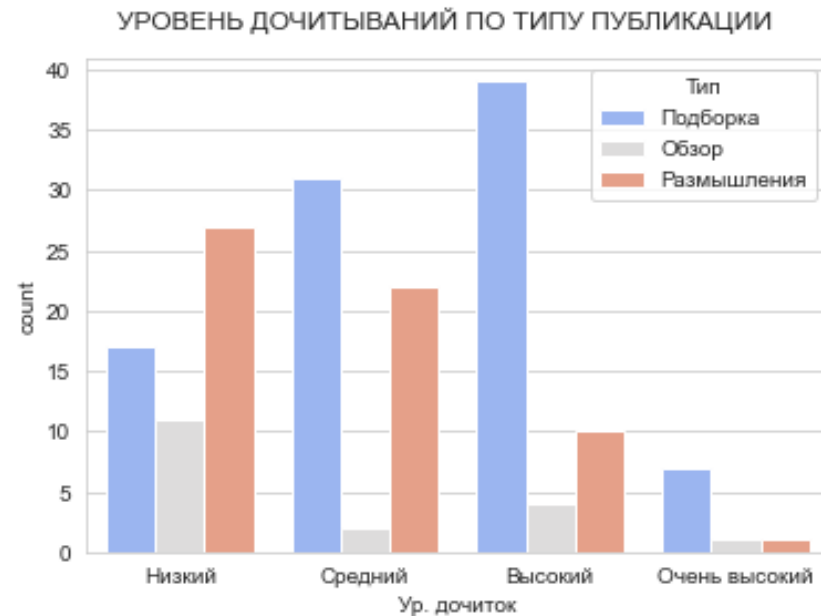
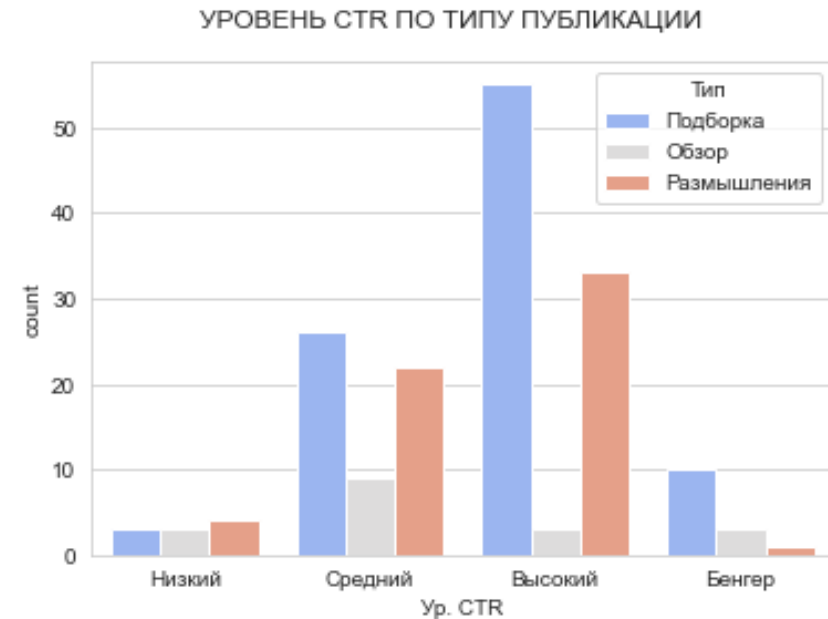


График четко демонстрирует прямую связь между уровнем дочитываний и CTR. Чем выше CTR, тем больше дочитываний, и чем больше дочитываний обычно набирают статьи, тем выше шанс, что следующие публикации будут хорошо ранжироваться алгоритмами платформы и получать большие охваты и больший CTR

# Графики количества статей по уровню показателя



Из графика можно отметить, что размышления дочитываются плохо. Большая часть имеют низкий или средний уровень дочиток. Отметим также, что в очень высоком уровне выделяются подборки как наиболее релевантный для больших охватов тип статей



Количество статей-размышлений несмотря на низкий уровень дочитываний демонстрирует положительную динамику CTR. Однако среди «Бенгеров» подобных статей практически нет.