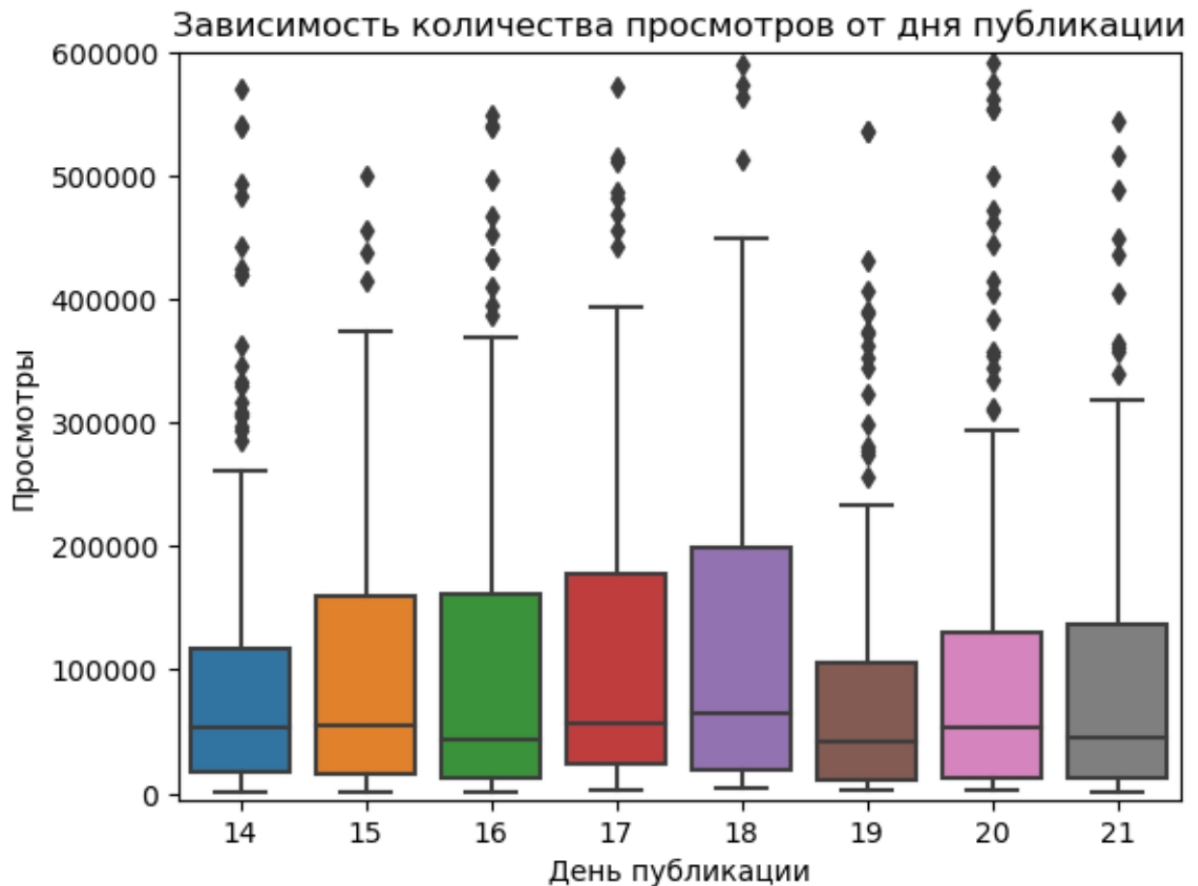


Задача 4 из ДЗ 2 (Комлев Роман)

Задача 2, пункт 4: построить и проанализировать ящики с усами на каждый день с 14 ноября по 21 ноября 2017 года по количеству просмотров в российском сегменте Youtube.



Видим, что среднее количество просмотров у видео примерно равно 50000, особо не меняется в течение недели. Достаточно много видео ежедневно набирали от 20000 до 150000 просмотров.

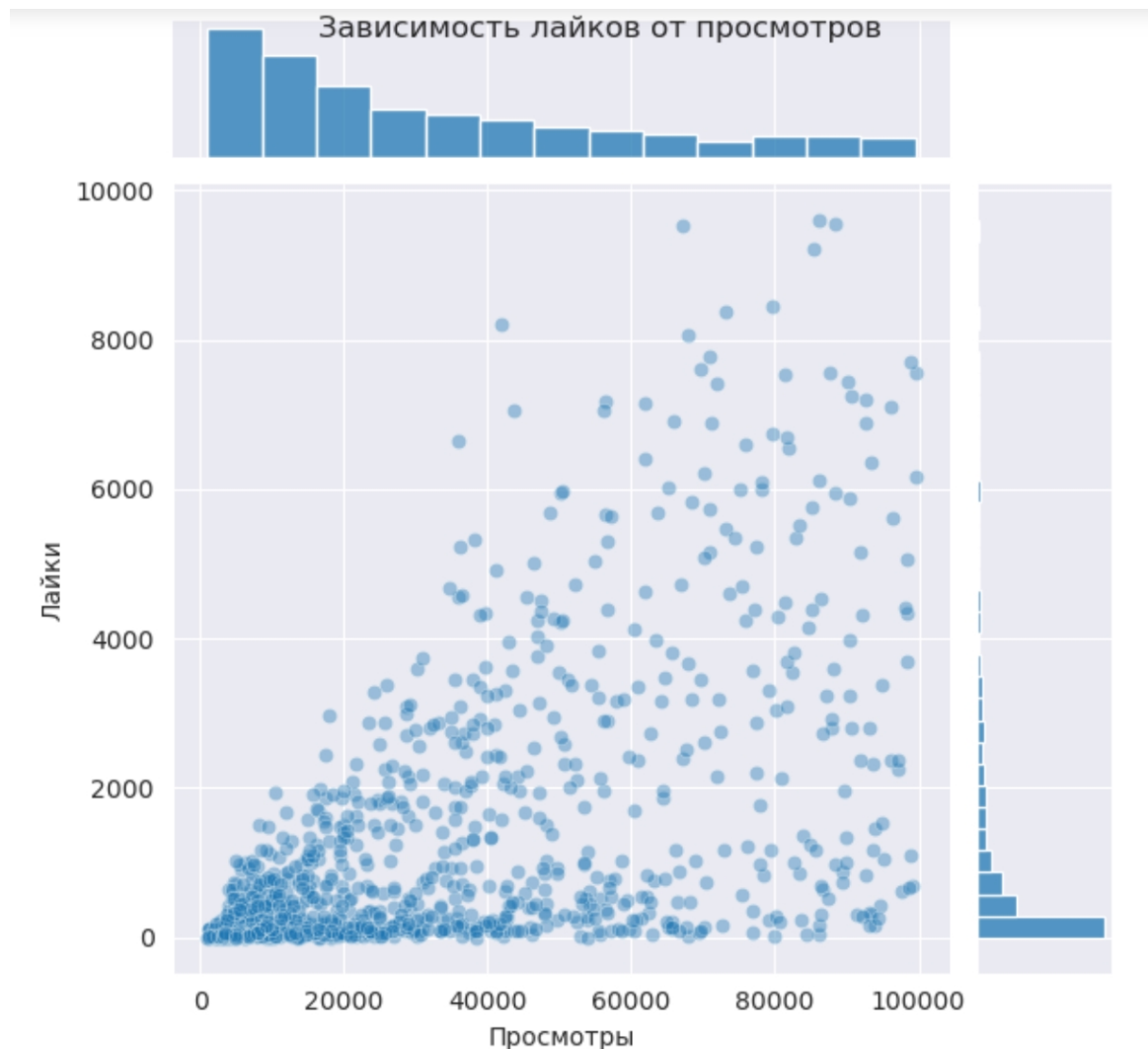
18 числа несколько видео набрали очень много просмотров, поэтому среднее значение в этот день является наибольшим за неделю.

19 числа почти все видео не набрали более 400000 просмотров. Среднее значение просмотров в 19 день является самым маленьким из всех рассмотренных дней.

Если не учитывать видео, собравшие более 300000 просмотров, то боксплоты по просмотрам для каждого из дней будут выглядеть почти одинаково.

Задача 2, пункт 6: построить и проанализировать jointplot по данным о видео в российском сегменте YouTube с 14 ноября по 21 ноября 2017 года для

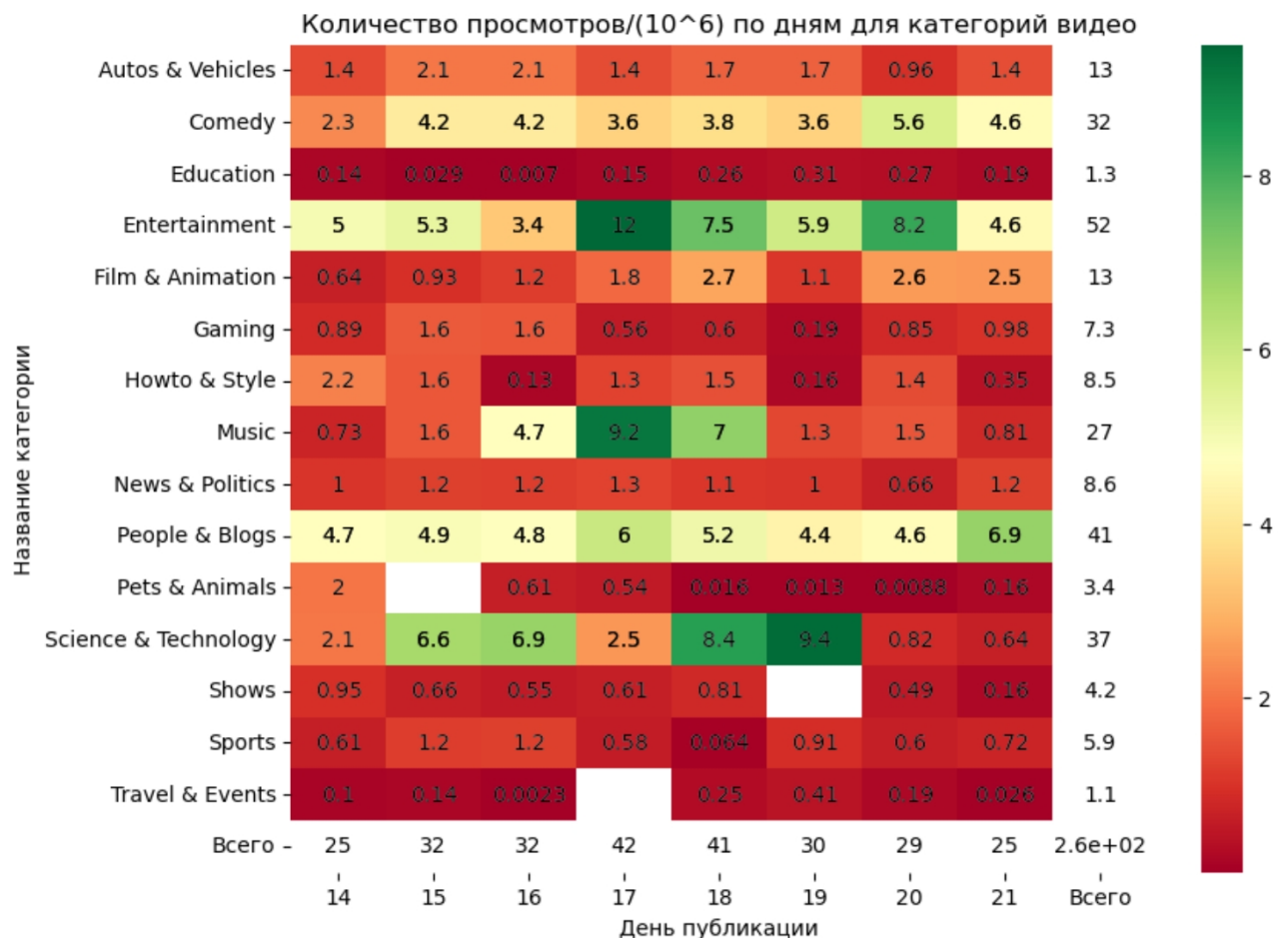
количества просмотров по горизонтальной оси и количества лайков по вертикальной.



При построении графика не были отображены данные о видео, которые набрали более 100000 просмотров или более 10000 лайков. Зависимость для исключенных видео выглядела аналогично, или же точка была отброшена по причине отсутствия других соседей в большом радиусе (выброс).

Из графика видно, что существует большое количество видео, которые набрали много просмотров, но имеют 0 лайков или совсем незначительное их количество. Чаще всего количество лайков на видео не превосходит 10% от количества просмотров. Большинство проанализированных видео имеют до 20000 просмотров и до 1000 лайков.

Задача 3, пункт 5: визуализировать и проанализировать таблицу по данным о видео в российском сегменте YouTube с 14 ноября по 21 ноября 2017 года с помощью heatmap. В таблице должна содержаться информация о количестве просмотров по дням для каждой категории видео, с добавлением суммы по столбцам и строкам.



При построении таблицы с помощью heatmap для информативности делим все числа на 10^6 . Значения в добавленных строке и столбце с суммой просмотров намного больше, чем все остальные значения, что негативно влияет на цветовую раскраску таблицы. Поэтому не красим строки "Всего", цвет чисел изменим на черный. Также меняем цветовую карту, чтобы черные числа были хорошо видны по всей таблице. Вручную задаем значения для привязки цветовой карты. Теперь таблица heatmap стала достаточно информативной.

Из таблицы видно, что категории Travel & Events и Education смотрели наименьшее количество раз из всех категорий. Больше всего просмотров за неделю было в категории Entertainment.

Меньше всего просмотров у видео было 14 и 21 ноября 2017 года, оба этих дня были вторниками. Больше всего видео смотрели 17 и 18 ноября, в пятницу и субботу.

Всего по данным в российском сегменте YouTube с 14 ноября по 21 ноября 2017 года видео смотрели примерно 260 млн раз.