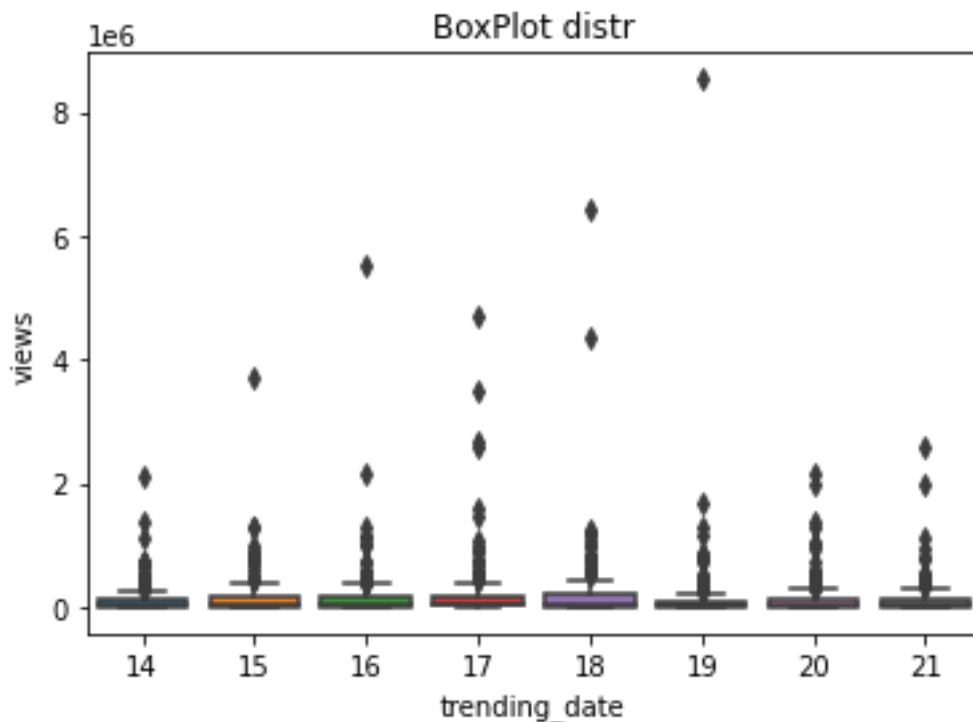


## Отчет по задаче “YouTube”

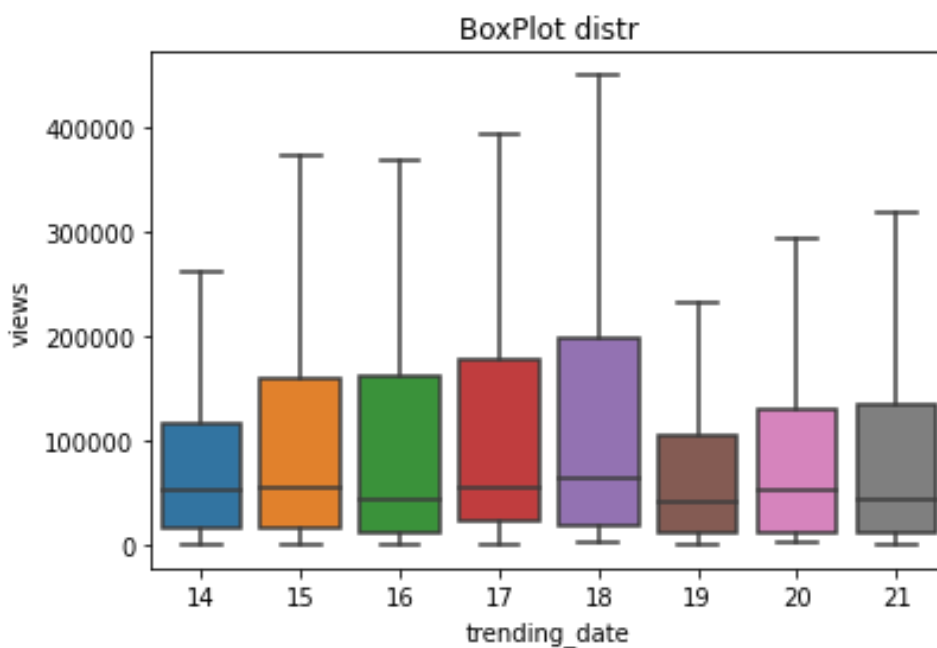
Для начала считываем данные и правильно распознаем дату.

Затем оставляем только нужные столбцы, поскольку в таблице присутствует много неинформативных столбцов. Оставляем только нужные, после по получившейся обрезанной таблице строим коробку с усами. В качестве оси x, выбираем день попадания в тренды, а в качестве оси y количество просмотров.



Гр. 1

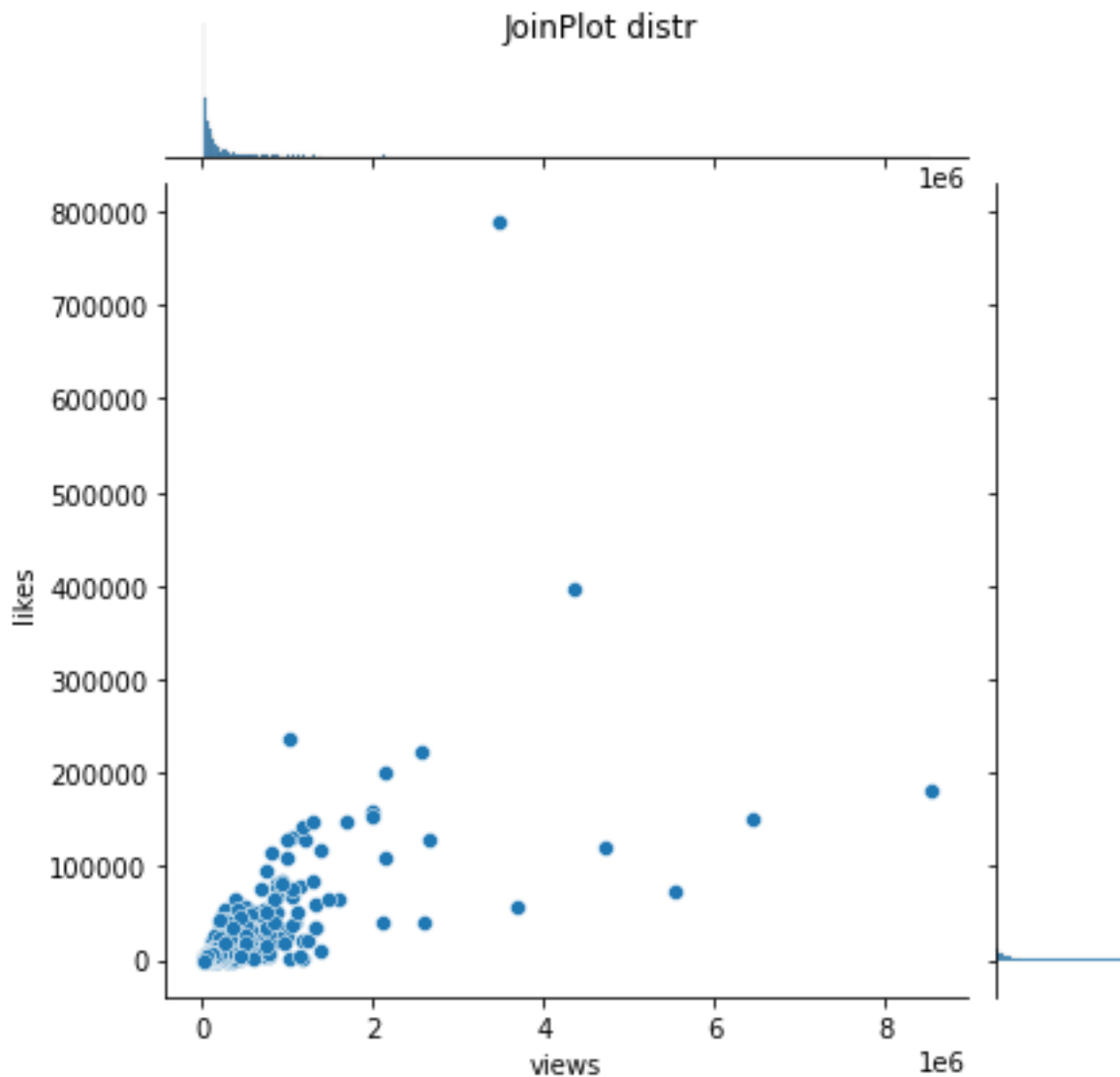
Как можем видеть, график получился малоинформативным. Это все из-за выбросов. Нужно от них избавиться. Удаляем выбросы, и получаем более информативный график.



Гр. 2

Избавившись от выбросов, можем увидеть более точный график(гр.2)

Теперь мы хотим аналогично построить график jointplot, только теперь с количеством просмотров по x, и лайками по y.

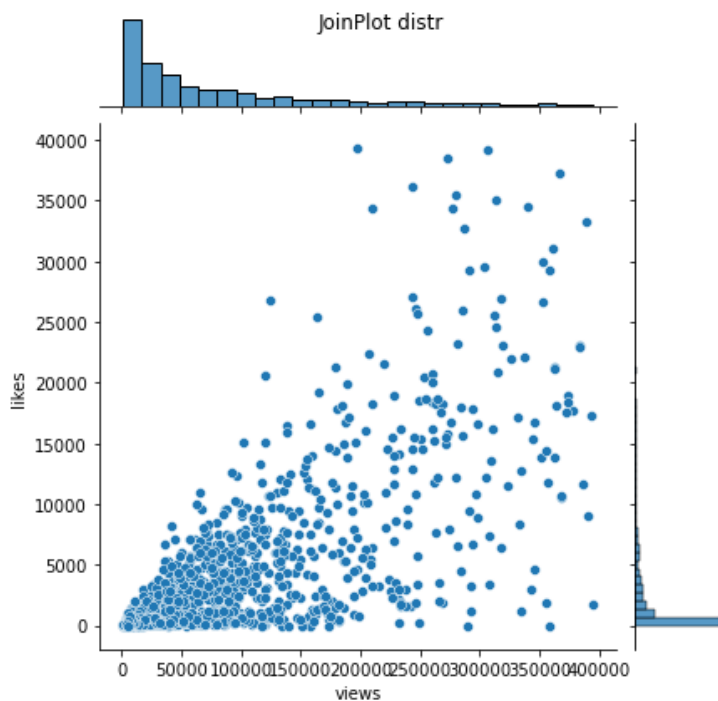


Гр.3

Исходя из Гр.3, у нас та же самая проблема – выбросы. Чтобы это исправить, ограничим данные, убрав тем самым выбросы.

После ограничения получаем следующее(Гр.4). График получился уже более информативным и полезным.

Выводы: из Гр.2 следует, что в среднем прирост количества просмотров, для видео которые находились в трендах, у всех примерно на одном уровне. Однако, некоторые видео имели стремительный рост, и затем его снизили. Их Гр.4 можно увидеть зависимость между Лайками и просмотрами. Видно, что их(лайков) на целый порядок меньше.



Гр.4

## Отчет по задаче “YouTube2”

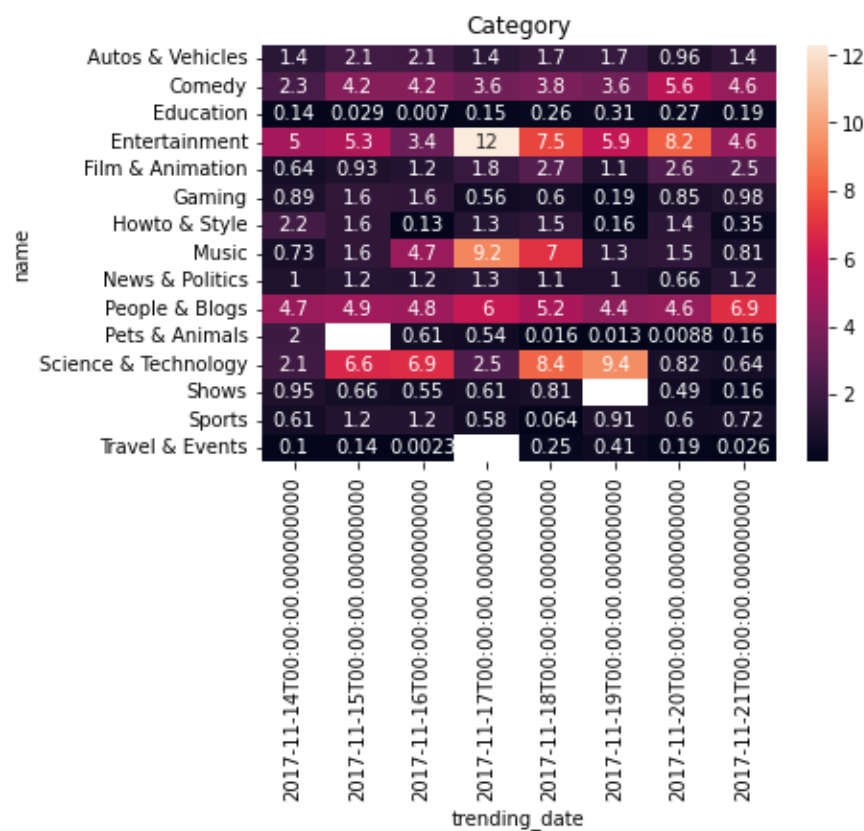
По аналогии с задачей YouTube сначала правильно читаем данные. Однако, теперь нам нужно прочитать еще одну табличку: категории. После успешного прочтения и дальнейшего слияние таблиц, по ключам id, получаем табличку в которой имеется категория видео в трендах.

Теперь нам нужно сводную таблицу о количестве просмотров по дням для каждой категории видео. Мы это делаем для того, чтобы в дальнейшем построить тепловую карту по сводной таблице(Гр.5). По тепловой таблице теперь, можно сказать о том какую категорию смотрят больше всего.

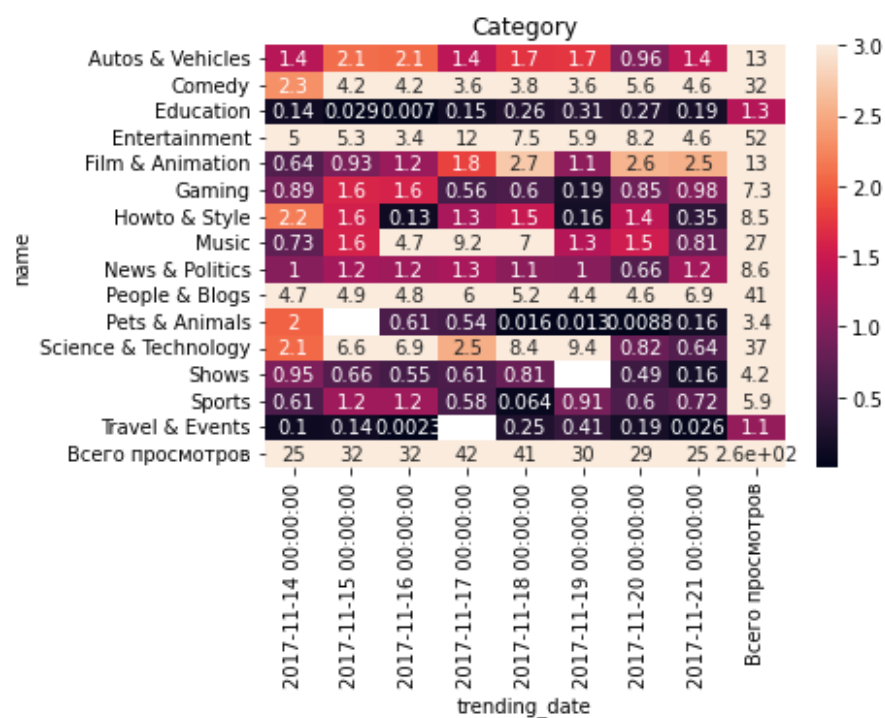
Теперь мы хотим получить общее количество просмотров. Для этого добавим отдельную колонку “Всего просмотров”, и будем наблюдать сколько каждая категория имеет всего просмотров за определенное время.

Однако, при таком решении могут возникать проблемы с визуализацией тепловой карты, поскольку она получается однотонной. Для того чтобы это исправить, ограничиваем максимальное значение просмотров. И получаем более информативную тепловую карту(Гр.6).

Выводы: из тепловой карты видно наибольшее количество просмотров имеет категория “Развлечение”. Меньше всего людей интересуют “Путешествия и Мероприятия”. Также как и в задаче YouTube, присутствует зависимость по дням, то есть распределение разное.



Гр.5



Гр.6