

# Ticketland ML Contest

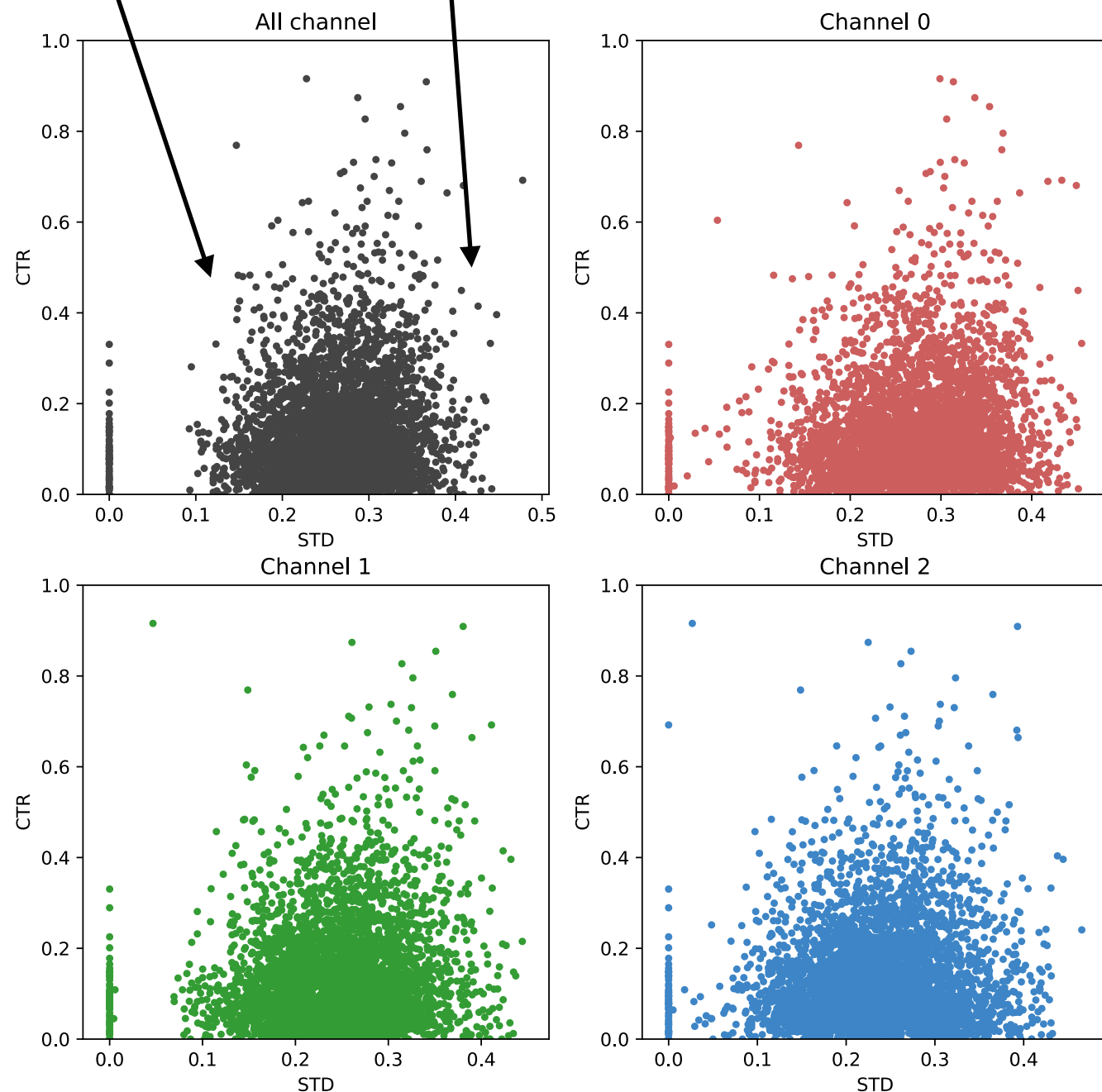
Каюмов Эмиль, 517  
ММП ВМК МГУ

Курс «Прикладные задачи анализа данных»

# Посмотрим на изображения

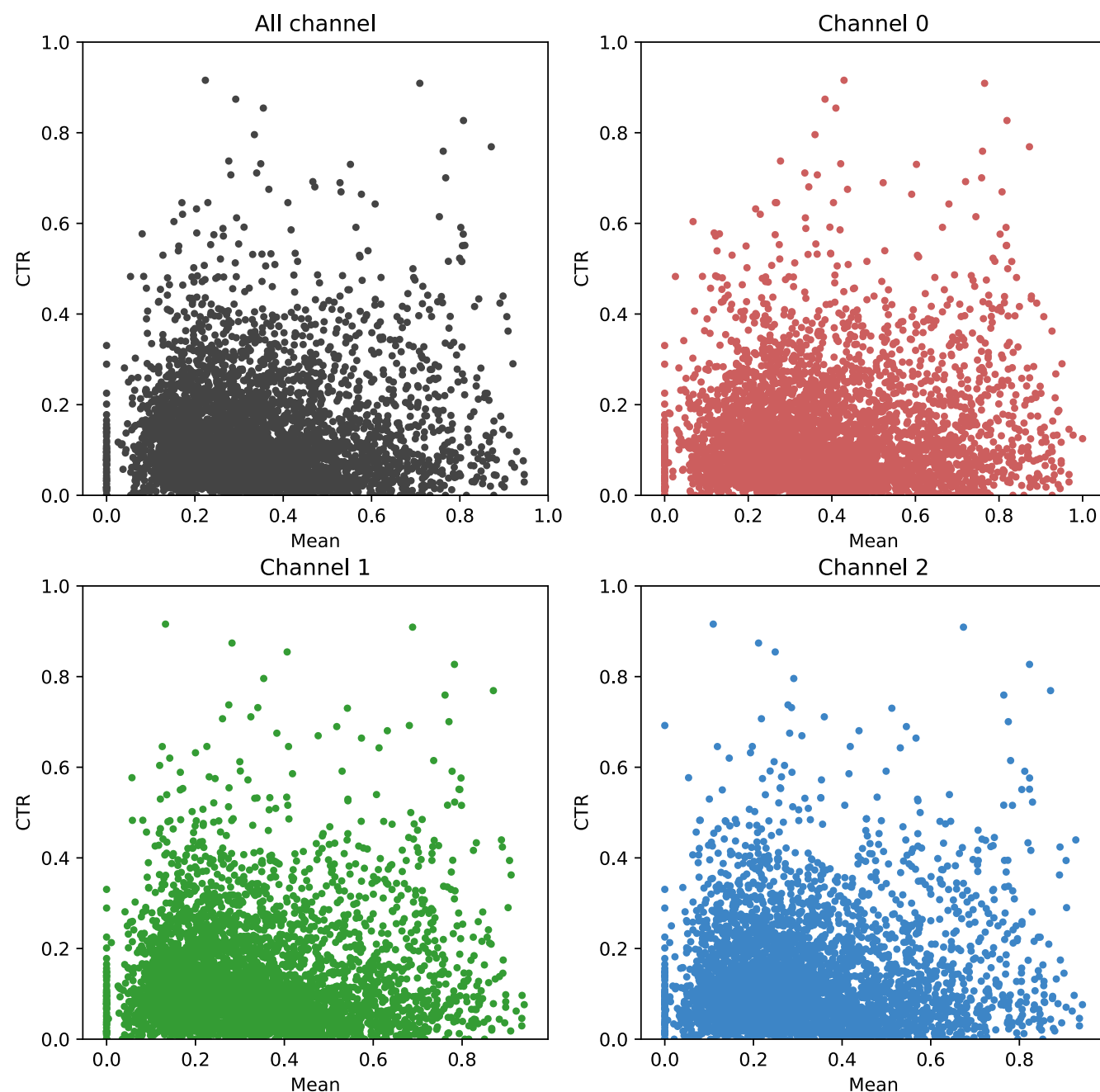
уныло

ЗАШКВАР



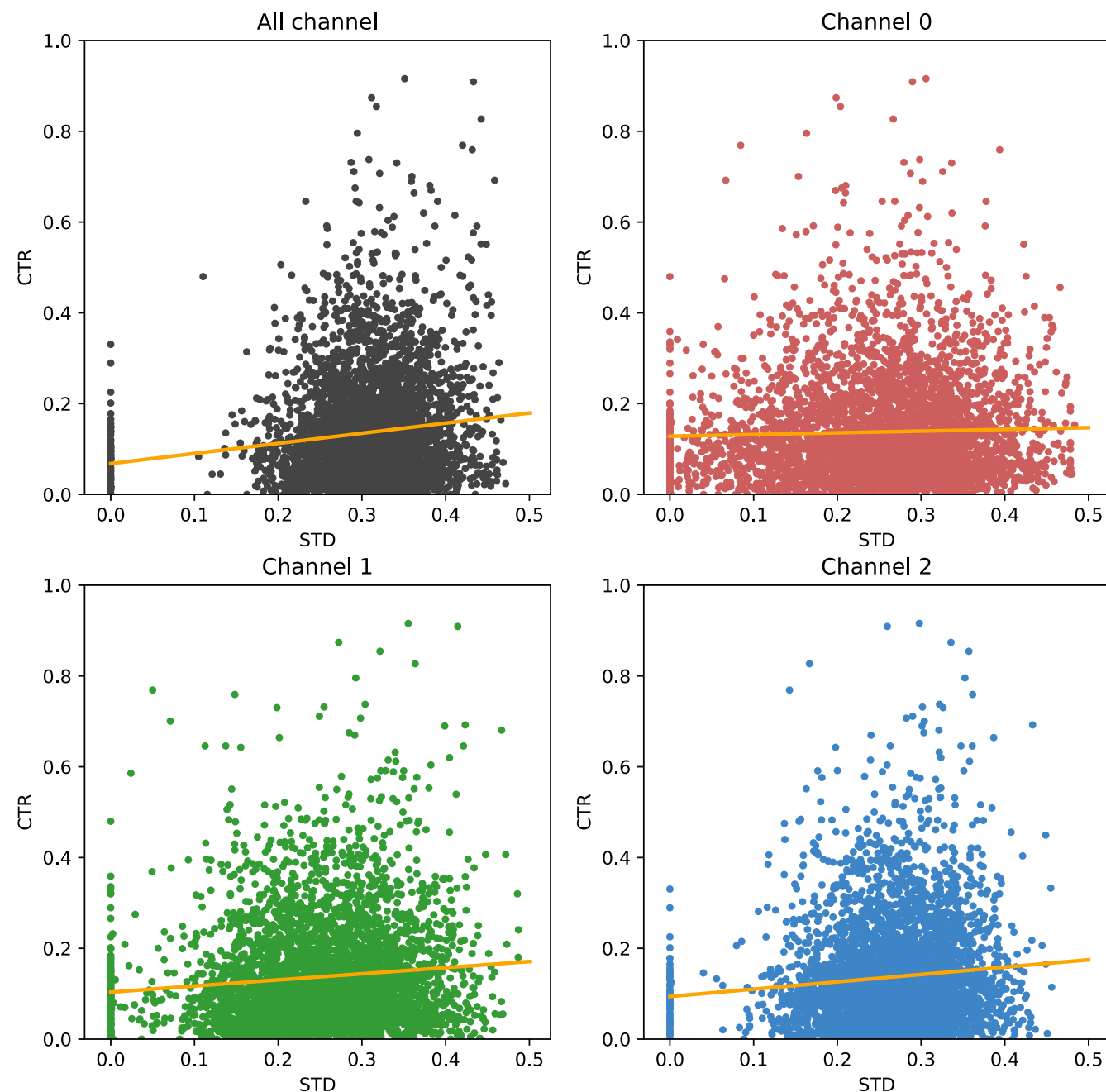
- STD по каналам (+общий)
- Столбик в нуле — нет изображений
- Но никакой явной зависимости нет

# Посмотрим на изображения (2)



- Mean по каналам (+общий)
- Столбик в нуле — нет изображений
- Но никакой явной зависимости опять нет

# Может, не RGB...



- STD по каналам HSV
- Линия без учёта столбика!
- Уже что-то...

# Дубли

id_show	image_id
25553	98220
25971	98222
16311	98217
6120	98219
16698	98221



98217.png



98219.png



98220.png



98221.png



98222.png

- Встречаются одинаковые изображения у разных шоу — можно объединить и считать общие статистики
- (Найдено случайно, а не специально лучший пример)

# Best overfit model

- На самом деле все ненулевые
- CTR специально переобученные (если и без того переобучение...)
- **USERxSHOWxDT**
- (Конечно, на логарифм смотреть лучше)

