

Университет ИТМО

## Практическая работа №2

по дисциплине «Визуализация и моделирование»

**Автор:** Кривошапкина Айталиа Сергеевна

**Поток:** ВИМ 1.1

**Группа:** К3241

**Факультет:** ИКТ

**Преподаватель:** Чернышева А.В.

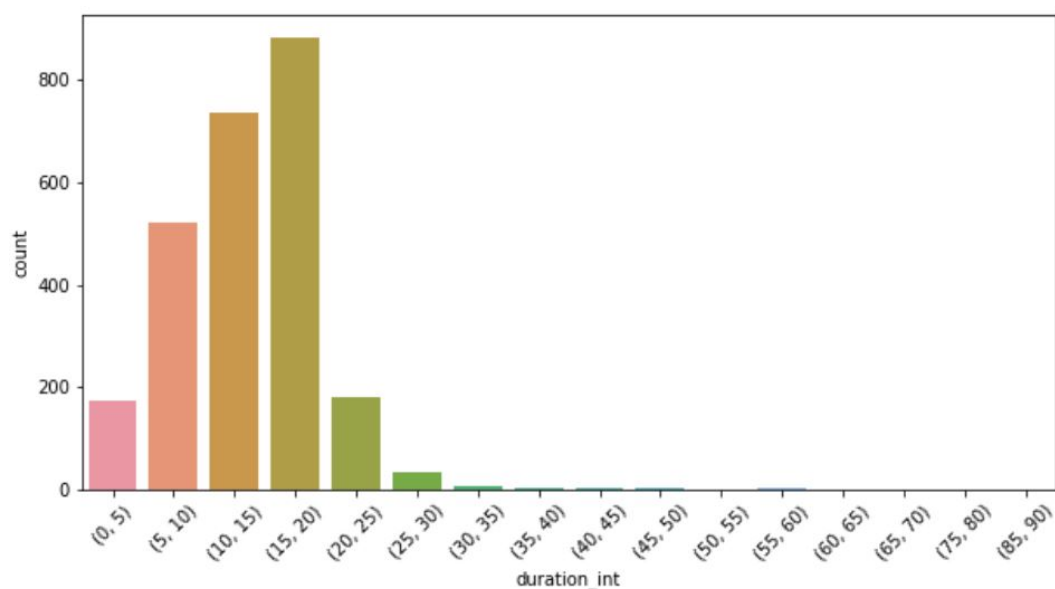
Санкт-Петербург, 2021 г.

Мною был выбран датасет "Ted Talks". Данный датасет содержит информацию обо всех аудио-видео записях выступлений TED, загруженный на официальный сайт ted.com до 21 сентября 2017 года. Основной набор данных содержит информацию обо всех выступлениях, включая количество просмотров, количество комментариев, описания, имена докладчиков и заголовки.

Таблица с описанием каждого заголовка:

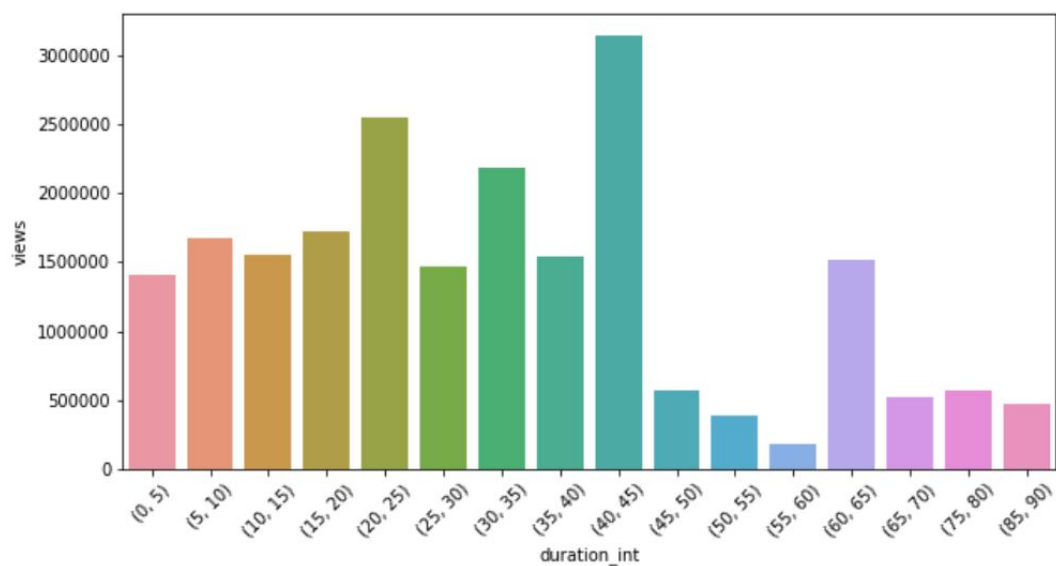
№	Название	Описание	Тип данных	Шкала
1	comments	Количество комментариев, сделанных во время выступления	int	относительная
2	description	Описание выступления	string	номинальная
3	duration	Продолжительность видео в секундах	int	интервальная
4	event	Место проведения выступления	string	номинальная
5	film_date	Продолжительность съемки в unix	int	интервальная
6	languages	Количество доступных языков	int	интервальная
7	main_speaker	Имя первого спикера	string	номинальная
8	name	Официальное название выступления, включающее в себя имя спикера	string	номинальная
9	num_speaker	Количество спикеров	int	относительная
10	published_date	Время публикации выступления в unix на ted.com	21.5	относительная
11	ratings	Слова, описывающие выступление	array	номинальная
12	related_talks	Список рекомендуемых словарей для последующего просмотра	array	номинальная
13	speaker_occupation	Специализация главного спикера	string	номинальная
14	tags	Темы, ассоциируемые с темой выступления	array	номинальная
15	title	Название выступления	string	
16	url	Ссылка на выступление	string	
17	views	Количество просмотров	int	относительная

1. Вопрос: Какой продолжительности видео выпускают чаще всего?



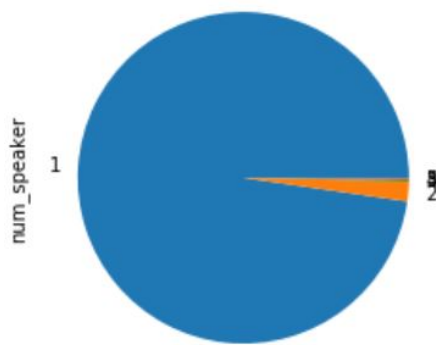
Ответ: в промежутке от 15 до 20 минут

2. Вопрос: Какой продолжительности видео смотрят чаще всего?

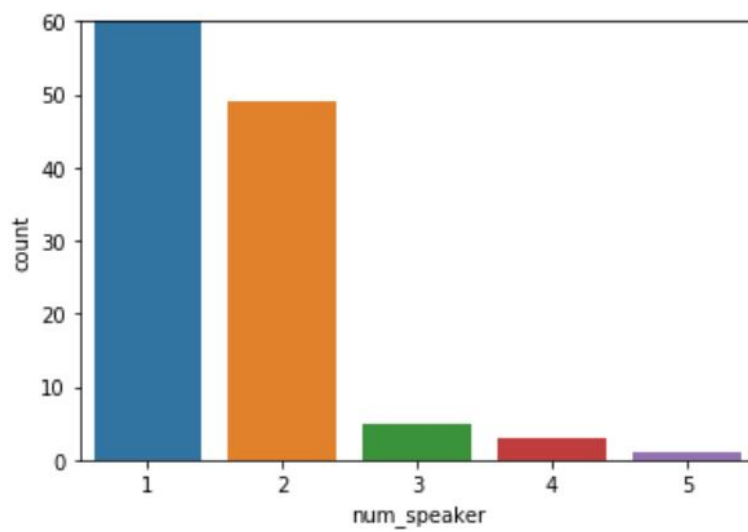


Ответ: в промежутке от 40 до 45 минут

3. Гипотеза: В основном в качестве спикера выступает только один человек

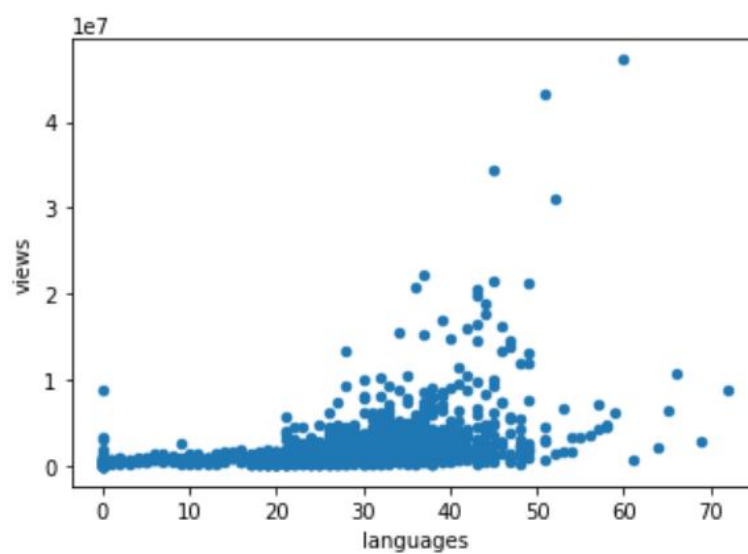


Рассмотрим лучше соотношение между меньшими долями на гистограмме:



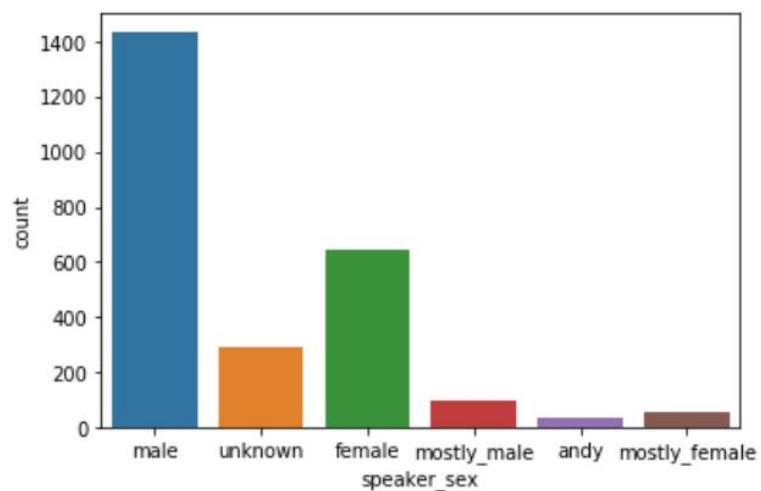
Ответ: да

4. Гипотеза: Чем больше доступных языков, тем больше просмотров у видео



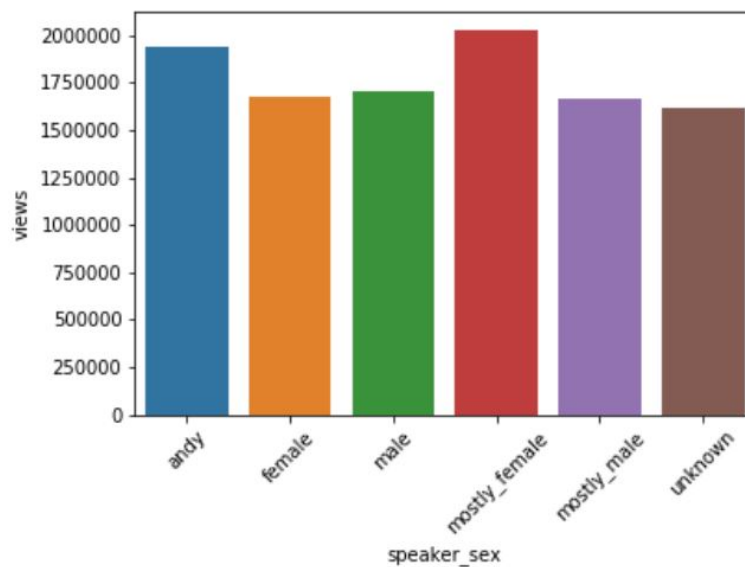
Ответ: да

5. Вопрос: Кто чаще выступает в качестве спикера: женщины или мужчины:  
Для этого воспользуемся библиотекой `gender_guesser`



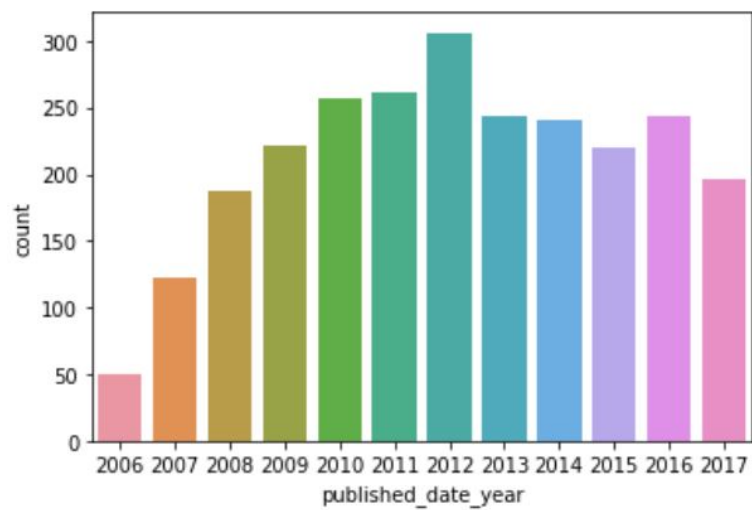
Ответ: мужчины

6. Вопрос: Кого смотрят больше: женщин или мужчин



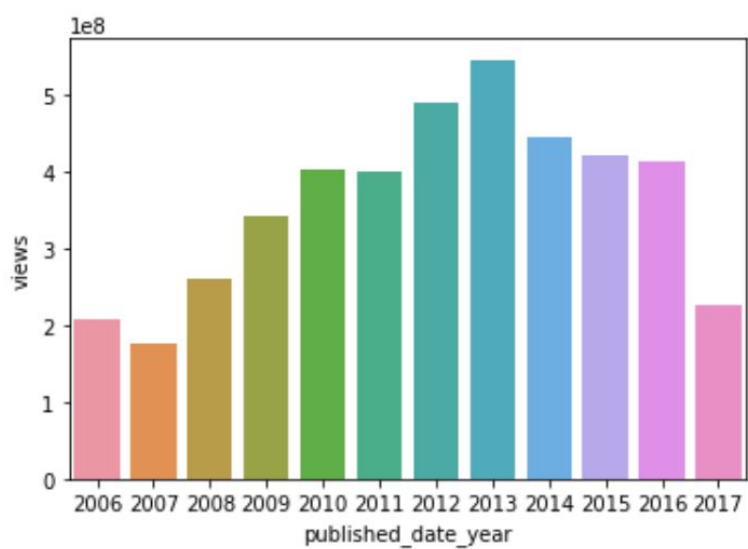
Ответ: женщин

7. Вопрос: В каком году вышло больше всего видео?

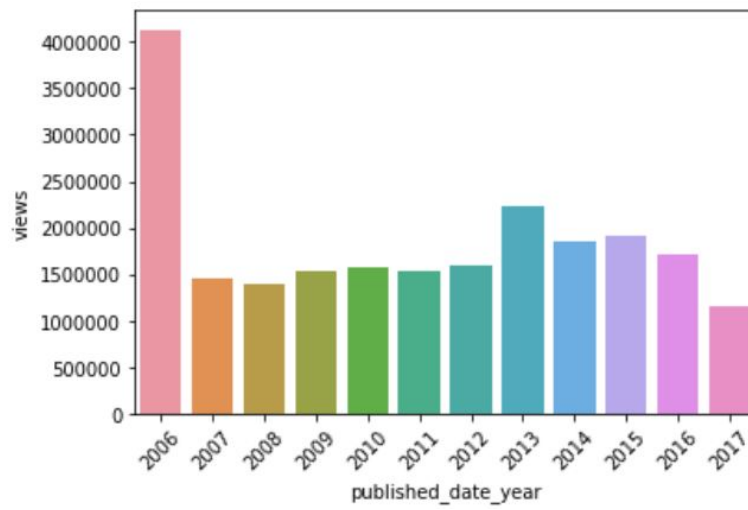


Ответ: 2012

8. Вопрос: Видео какого года имеет больше всего просмотров?

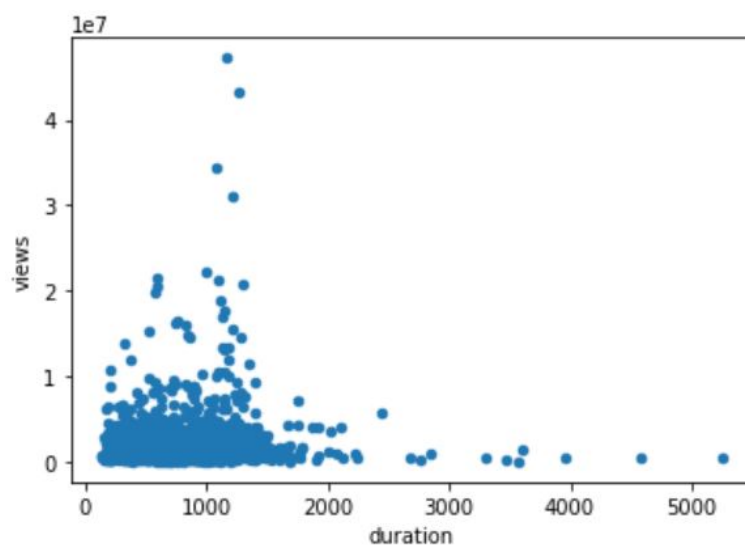


Видим, что в сумме просмотров больше в 2013 году. Но если взять среднее значение вместо суммы, то получим следующее:



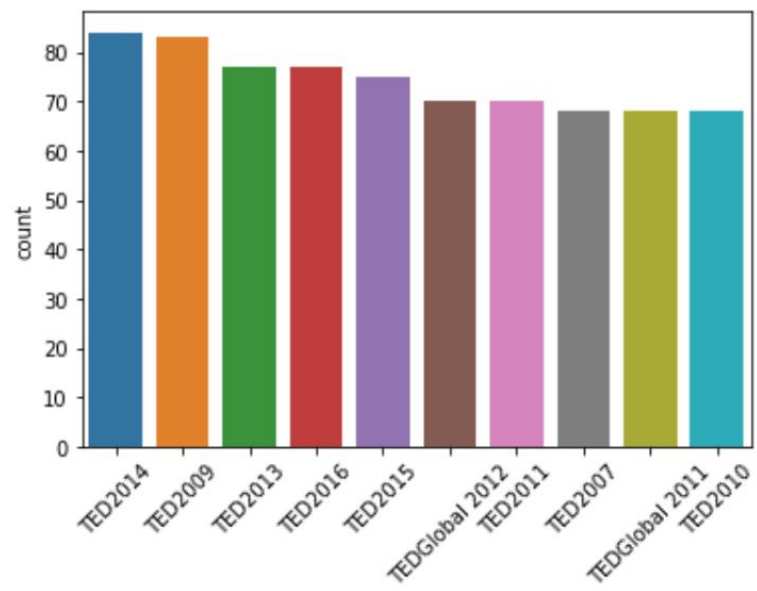
Ответ: 2006

9. Гипотеза: Чем дольше видео, тем меньше просмотров



Ответ: да

10. Вопрос: Какое мероприятие было самым масштабным?



Ответ: TED2014