

Университет ИТМО

## Практическая работа №1

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
1	comments	Количество комментариев, сделанных во время выступления	int
2	description	Описание выступления	string
3	duration	Продолжительность видео в секундах	int
4	event	Место проведения выступления	string
5	film_date	Продолжительность съемки в unix	int
6	languages	Количество доступных языков	int
7	main_speaker	Имя первого спикера	string
8	name	Официальное название выступления, включающее в себя имя спикера	string
9	num_speaker	Количество спикеров	int
10	published_date	Время публикации выступления в unix на ted.com	21.5
11	ratings	Слова, описывающие выступление	array
12	related_talks	Список рекомендуемых словарей для последующего просмотра	array
13	speaker_occupation	Специализация главного спикера	string
14	tags	Темы, ассоциируемые с темой выступления	array
15	title	Название выступления	string
16	url	Ссылка на выступление	string
17	views	Количество просмотров	int

На данном датасете можно решить следующие глобальные задачи:

- определить наиболее популярные темы слушателей
- определить темы, которые часто затрагиваются спикерами
- определить темы, которые потенциально могут быть интересны слушателям