

# Анализ взаимодействия пользователей с карточками Яндекс.Дзен

Подготовила студент 38 когорты DA  
Зайцева Наталия

# Описание

- Яндекс.Дзен - платформа для создания и просмотра контента.
- Контент формируется в виде карточки (статьи), с которой могут взаимодействовать пользователи (кликать, просматривать и т.д.)
- Каждую карточку определяют её тема и источник, у которого тоже есть тема
- Пользователей системы характеризует возрастная категория

## Цель проекта

1. Автоматизировать процесс анализа взаимодействия пользователей с карточками Яндекс.Дзен - построить дашборд в Tableau.
2. Провести анализ взаимодействия пользователей с карточками Яндекс.Дзен на текущий момент.

# 1. Построение дашборда

# Техническое задание

- Основные **пользователи** дашборда: менеджеры по анализу данных
- **Частота пользования** дашбордом: не реже, чем раз в неделю
- **Состав данных:**
  - История событий по темам карточек (два графика - абсолютные числа и процентное соотношение);
  - Разбивка событий по темам источников;
  - Таблица соответствия тем источников темам карточек;
- **Группировка данных:**
  - Дата и время;
  - Тема карточки;
  - Тема источника;
  - Возрастная группа;
- **Обновление** данных: один раз в сутки, в полночь по UTC

# Макет дашборда

Название дашборда		Фильтр по дате и времени	
Описание дашборда		Фильтр по темам карточек	Фильтр по возрастным категориям
График истории взаимодействия «События по темам карточек» (абсолютные значения, stacked area chart)	График истории взаимодействия «% событий по темам карточек» (% от общего, stacked area chart)	График разбивки событий по темам источников «События по темам источников» (относительные значения, pie chart)	
Таблица соответствия тем карточек темам источников «Темы источников - темы карточек». В колонках должны быть указаны темы источников, в строках – темы карточек. На пересечении должно быть указано абсолютное количество событий. Ячейки таблицы должны иметь окраску в зависимости от числа взаимодействий (highlight table).			

# Исходные данные

- Предполагается, что дашборд будет работать на основе пайплайна, который будет брать сырые данные, обрабатывать их и агрегировать в таблицу.
- На данный момент имеется готовая агрегированная таблица `dash_visits.csv`, на основе которой строится дашборд

## Таблица dash\_visits.csv

- Время анализа: 24 сентября 2019 года с 18:28 по 19:00
- Всего уникальных записей за это время: 30 745
- Всего событий за это время: 310 207
- Столбцы:
  - id записи;
  - Тема карточки (25 уникальных тем);
  - Тема источника карточки (26 уникальных тем);
  - Возрастной сегмент пользователя;
  - Дата совершения событий;
  - Количество событий (просмотров, кликов и т.д.)



# Построение дашборда

- Поскольку тем карточек и тем источников слишком много, некоторые графики на дашборде будут выглядеть нечитаемыми
- Решено объединить их в **категории тем**
- Объединение по категориям будем производить на основании здравого смысла
- Корректнее было бы объединять на основе корреляции между просмотрами тем, но в данной работе этого делать не будем

## Категории тем карточек

<b>Семья и дети</b>	Семья Дети
<b>Отношения, психология</b>	Отношения Психология Женская психология Женщины Скандалы
<b>Красота и здоровье</b>	Красота Здоровье
<b>Общество, Россия, деньги</b>	Общество Россия Деньги
<b>Наука и искусство</b>	Искусство Культура Наука История
<b>Путешествия</b>	Путешествия Туризм
<b>Полезные советы</b>	Полезные советы Интересные факты
<b>Развлекательное</b>	Знаменитости Подборки Рассказы Шоу Юмор

## Категории тем источников

<b>Семья, дети, отношения</b>	Семья Дети Семейные отношения Психология
<b>Здоровый образ жизни</b>	Здоровье Спорт Еда
<b>Политика, Россия, деньги</b>	Политика Россия Деньги Финансы
<b>Искусство и история</b>	Искусство История Кино Музыка
<b>Ремонт, сад, дача</b>	Ремонт Сад и дача Строительство Технологии Интерьеры
<b>Путешествия, авто</b>	Путешествия Авто
<b>Мода и знаменитости</b>	Знаменитости Одежда
<b>Полезные советы</b>	Полезные советы Сделай сам

## Ссылка на дашборд

[https://public.tableau.com/app/profile/tata.zaitseva/viz/\\_16529714108670/sheet12\\_1?publish=yes](https://public.tableau.com/app/profile/tata.zaitseva/viz/_16529714108670/sheet12_1?publish=yes)

# Взаимодействие пользователей с карточками Яндекс.Дзен

Этот дашборд поможет анализировать взаимодействие пользователей с карточками Яндекс.Дзен.

Первые два графика (слева направо) показывают изменение количества взаимодействий во времени в разбивке по темам карточек в абсолютном (1) и относительном (2) представлении.

Третий график показывает соотношение количества событий по темам источников.

Четвертый график (внизу) показывает соответствие тем карточек темам источников (на пересечении - количество событий).

Дата и время

18:28:00

19:00:00

Возрастной сегмент

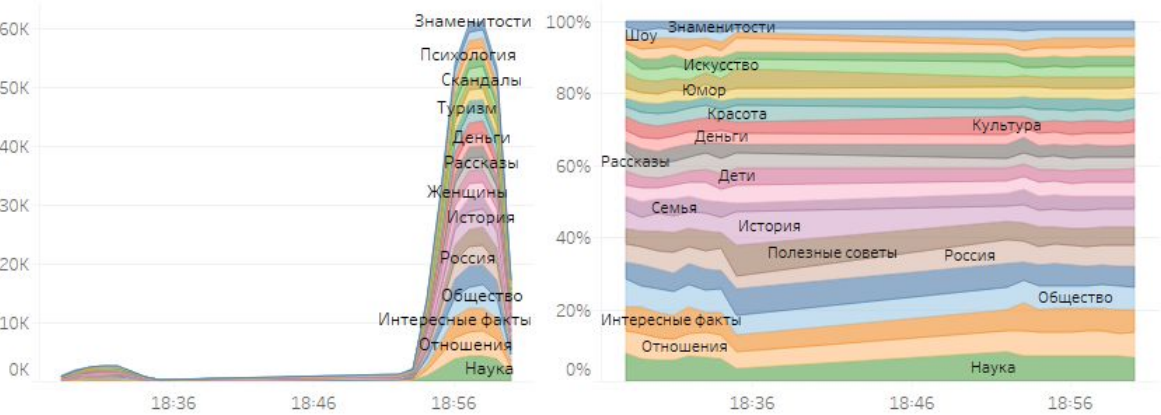
(All)

Тема карточки

(All)

Тема источника

(All)



Категория темы карточки	Категория источника									
	Семья, дети, отношения	Мода и знаменитости	Политика, Россия, деньги	Искусство и история	Полезные советы	Здоровый образ жизни	Ремонт, сад, дача	Путешествия, авто		
Отношения, психолог..	15 042	9 032	8 536	6 010	5 896	5 538	4 158	3 020		
Общество, Россия, деньги	11 375	3 728	11 568	6 220	3 543	3 079	2 276	5 108		
Наука и искусство	10 983	6 273	7 339	13 504	3 997	2 436	3 378	7 936		
Развлекательное	10 662	6 462	6 754	9 231	5 185	3 414	3 264	8 012		
Полезные советы	6 447	2 719	4 683	4 118	4 456	4 372	3 043	5 539		
Семья и дети	5 648	2 997	2 056	2 257	3 622	2 557	2 736	1 013		
Путешествия	4 052	1 408	2 419	3 015	2 424	1 404	2 422	1 628		
Красота и здоровье	3 969	3 221	1 602	1 227	4 576	2 176	2 007	1 435		

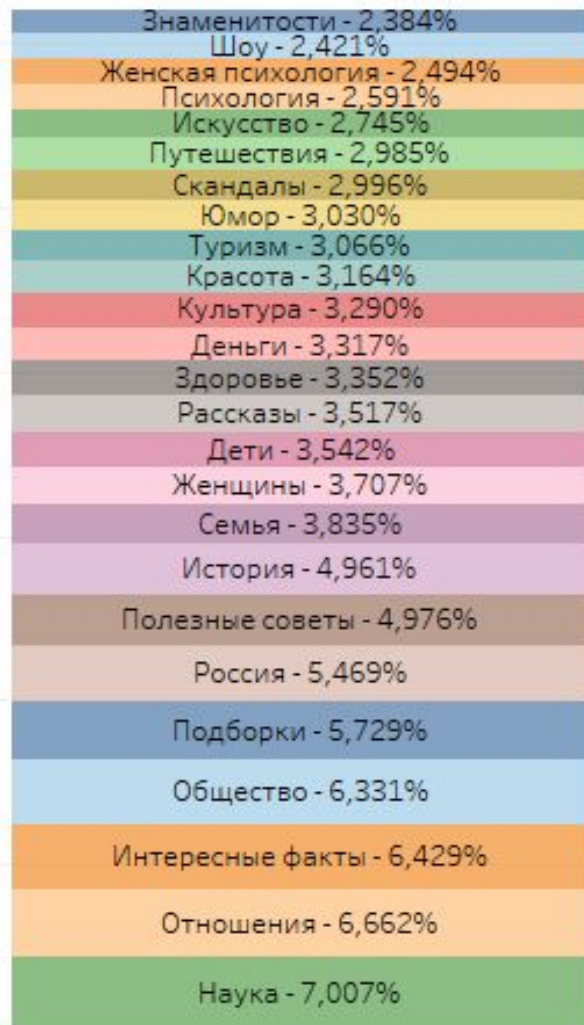
# Вывод

- Дашборд построен согласно заявленному макету
- Имеется группировка по дате, возрастной категории, темам карточек и темам источника
- При наведении на категорию темы на графике соответствия тем карточек и источников во всплывающем окне можно подробно рассмотреть информацию о темах, входящих в данную категорию

## 2. Анализ взаимодействия пользователей с карточками Яндекс.Дзен на текущий момент

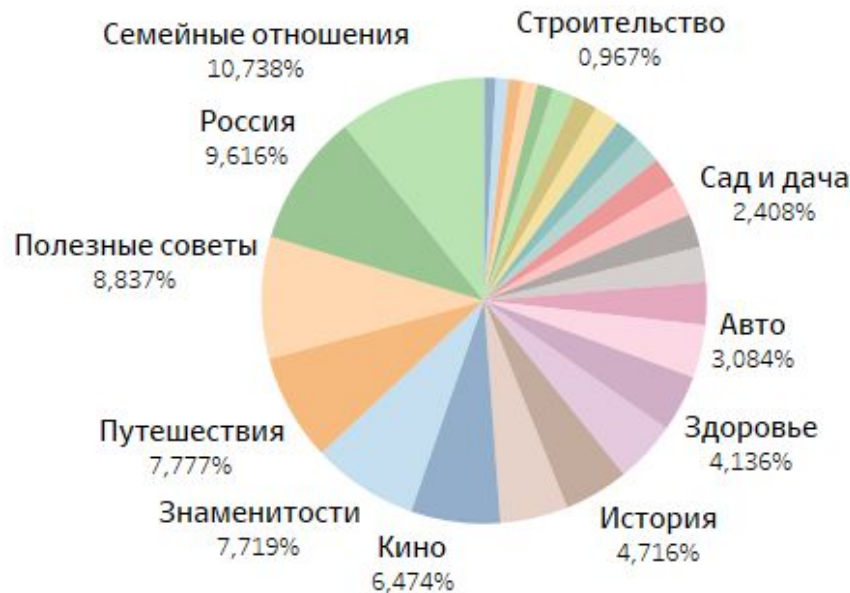
# Сколько взаимодействий пользователей с карточками происходит в системе с разбивкой по темам карточек?

- Всего взаимодействий за наблюдаемый период: **310 207**
- За **каждую минуту** происходит **до 60 тысяч** взаимодействий
- **Топ-5 популярных тем** - на них приходится почти **треть** всех взаимодействий:
  - Наука
  - Отношения
  - Интересные факты
  - Общество
  - Подборки
- **Наименее популярные темы:**
  - Знаменитости
  - Шоу
  - Женская психология



# Как много карточек генерируют источники с разными темами?

- **Больше всего** карточек генерируют источники с темами:
  - Семейные отношения
  - Россия
  - Полезные советы
  - Путешествия
  - Знаменитости
  - Кино
- На них приходится **более половины всех карточек**
- **Меньше всего** тем от источников:
  - Финансы
  - Музыка
  - Строительство
  - Технологии





# Как соотносятся темы карточек и темы источников?

## Темы карточек и темы источников соотносятся не очень хорошо:

- Например, источники с категорией “Семья, дети, отношения” генерируют много карточек, в том числе с темами про отношения. Но, в то же время, эти источники генерируют больше карточек на тему науки и искусства, чем на тему семья и дети.
- То же самое с другими категориями, например источники с категорией “Путешествия, авто” генерируют карточки с совершенно разными темами, не относящимися к путешествиям.
- Стоит учесть, что в данной таблице корректнее рассматривать количество уникальных карточек, а не событий по ним, но таких данных у нас нет

Категория темы карточки	Категория источника							
	Семья, дети, ..	Мода и знаменитости	Политика, Россия, д..	Искусство и история	Полезные советы	Здоровый образ жизни..	Ремонт, сад, дача	Путешествия, авто
Отношения, психология	15 042	9 032	8 536	6 010	5 896	5 538	4 158	3 020
Общество, Россия, деньги	11 375	3 728	11 568	6 220	3 543	3 079	2 276	5 108
Наука и искусство	10 983	6 273	7 339	13 504	3 997	2 436	3 378	7 936
Развлекательное	10 662	6 462	6 754	9 231	5 185	3 414	3 264	8 012
Полезные советы	6 447	2 719	4 683	4 118	4 456	4 372	3 043	5 539
Семья и дети	5 648	2 997	2 056	2 257	3 622	2 557	2 736	1 013
Путешествия	4 052	1 408	2 419	3 015	2 424	1 404	2 422	1 628
Красота и здоровье	3 969	3 221	1 602	1 227	4 576	2 176	2 007	1 435

# Вывод

- Таким образом, карточки с разными темами обладают разной популярностью. **Больше всего взаимодействий** с карточками тем: **наука, отношения, интересные факты.**
- Источники с разными темами генерируют разное количество карточек - **более всего продуктивны источники** с темами **“семейные отношения”, “Россия”, “полезные советы”, “путешествия”, “знаменитости” и “кино”** - на их долю приходится более половины всех карточек.
- **Тема источника не всегда связана с темой карточки** - источники генерируют карточки на абсолютно разные темы, часто не относящиеся к их собственной.