

# Проект:

создание интерактивного дашборда анализа активности читательской аудитории Яндекс.Дзен – в зависимости от тем статей и источников, и возрастной группы аудитории

Доклад по статусу выполнения проекта

Аналитик данных Севастьянов А.А.

Материалы веб-встречи РГ по проекту от 30.09.2022

# План презентации:

1. Описание проекта. Состав участников и роли. Организационно-техническое обеспечение проекта
2. Ожидания стейкхолдеров проекта. Краткое техническое задание (ТЗ)
3. Утвержденный макет дашборда
4. Обзор данных, передаваемых в дашборд из автоматизированного пайплайна сбора данных
5. Финальный дашборд. Внешний вид и рабочая ссылка на сам дашборд
6. Обзор функционала дашборда и получение ответов на вопросы маркетинга напрямую из дашборда:
  - Сколько взаимодействий пользователей с карточками происходит в системе с разбивкой по темам карточек?
  - Как много карточек генерируют источники с разными темами?
  - Как соотносятся темы карточек и темы источников?
7. Общие выводы по проекту

# Структура проекта

Сроки выполнения: 19.09.2022 – 30.09.2022

Спонсор: отдел продуктового маркетинга

Рабочая группа проекта в составе:

специалист	роль
Проектный менеджер	Руководитель проекта и планирование
Менеджер продукта	Бизнес заказчик проекта и бюджет
Инженер данных	Архитектор автоматизации сбора данных и автор скриптов и требований к IT
Аналитик данных	Предварительный анализ собранных данных, подготовка для загрузки в специализированное ПО по построению дашбордов, разработка и публикация дашборда.
Специалист IT helpdesk	Обеспечение корпоративного IT
Специалист по ИБ	Информационная безопасность и ПД
Менеджер по закупкам ПО	Обеспечение лицензиями и услугами ТП на ПО 3-х сторон

Организационно-техническое обеспечение проекта:

- Доступ к журналам регистрации действий пользователей (лог-файлам)
- Управление планировщиками работ (автоматизация запусков скриптов)
- Доступ к базам данных
- Выделение ресурсов требуемой производительности в корпоративном IT облаке
- Обеспечение сетевой обвязки требуемой пропускной способности
- Создание пользователей систем и приложений с соответствующими полномочиями
- Защита персональных данных
- Активация лицензий ПО и ТП 3-х сторон

# Согласованное со стейкхолдерами ТЗ

Автор: аналитик данных

Согласовано: члены рабочей группы проекта

- Бизнес-задача: анализ взаимодействия пользователей с карточками Яндекс.Дзен.
- Насколько часто предполагается пользоваться дашбордом: не реже, чем раз в неделю.
- Кто будет основным пользователем дашборда: менеджеры по анализу контента.
- Состав данных для дашборда:
  - История событий по темам карточек (два графика - абсолютные числа и процентное соотношение);
  - Разбивка событий по темам источников;
  - Таблица соответствия тем источников темам карточек.
- По каким параметрам данные должны группироваться:
  - Дата и время;
  - Тема карточки;
  - Тема источника;
  - Возрастная группа.
- Характер данных:
  - История событий по темам карточек — абсолютные величины с разбивкой по минутам;
  - Разбивка событий по темам источников — относительные величины (% событий);
  - Соответствия тем источников темам карточек - абсолютные величины.
- Важность: все графики имеют равную важность.
- Источники данных для дашборда: сырые данные о событиях взаимодействия пользователей с карточками.
- База данных, в которой будут храниться агрегированные данные: дополнительные агрегированные таблицы в БД.
- Частота обновления данных: один раз в сутки, в полночь по UTC.
- Состав и порядок отображения графиков, а также элементы управления, - должны соответствовать утвержденному макету дашборда (приложение №1 к настоящему ТЗ).

# Утвержденный макет дашборда

Данный макет описывает поля дашборда и их наполнение. Дашборд интерактивен и позволяет выбирать критерии фильтрации визуализируемых данных. Фильтры применяются ко всем графикам и таблицам, где это уместно.

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).			

# Обзор данных передаваемых дашборд

Источник получения данных: таблица dash\_visits БД data-analyst-zen-project-db

Файл с финальными данными: dash\_visits.csv (загружен в Tableau Public)

Значимые поля записей из БД:

- item\_topics – одна из 25 тем карточек статей Яндекс.Дзен
- source\_topics – одна из 26 тем источников Яндекс.Дзен
- age\_segment – один из 6 диапазонов возрастов читателей
- dt – временная метка записи с точностью до минуты
- visits – агрегированное кол-во взаимодействий читателей

Всего 30745 записей без дубликатов и пропусков данных

Особенности исходных данных:

- item\_topics и source\_topics имеют 11 совпадающих тем и 14 и 15 несовпадающих соответственно
- dt имеет только 17 уникальных значений – все в 32-минутный период одного дня с шагом в 1 минуту (см. Рис.1 справа)
- visits содержит одно единственное значение для каждой уникальной комбинации других полей (см. Рис.2 справа)

В целом данные не требуют дополнительной предобработки

Рис.1: распределение исходных данных по времени dt

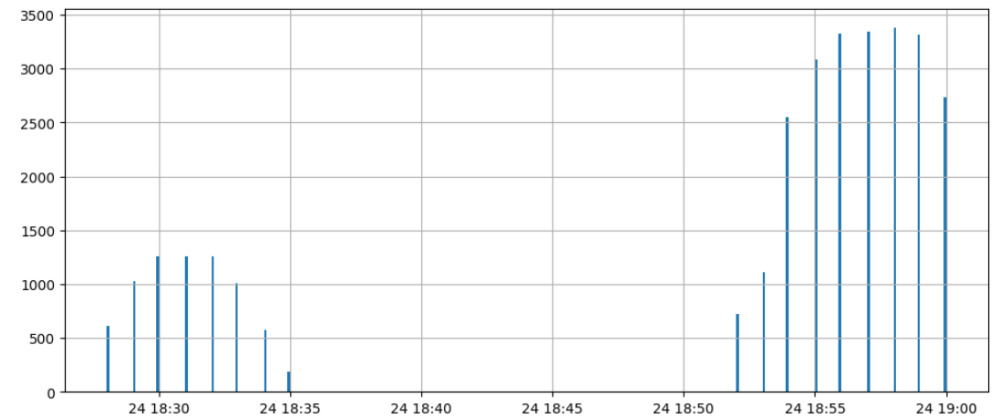
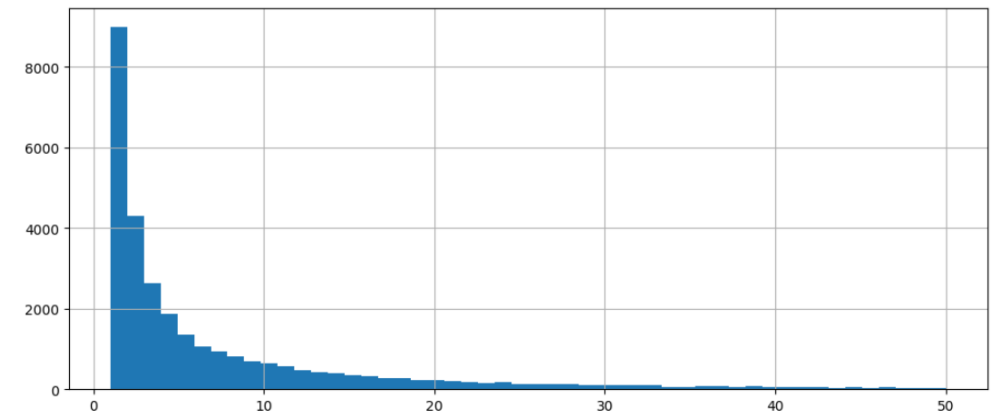


Рис.2: распределение 30745 записей по значениям visits





# Внешний вид финального дашборда

слайд 7

[https://public.tableau.com/views/DashboardDA49SevastyanovAA/DashboardDA49SevastyanovAA?:language=en-US&publish=yes&:display\\_count=n&:origin=viz\\_share\\_link](https://public.tableau.com/views/DashboardDA49SevastyanovAA/DashboardDA49SevastyanovAA?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link)

## Дашборд активности читательской аудитории Яндекс.Дзен по темам статей и источникам

фильтры данных:

Выберите период исходных данных:



Выберите темы статей:

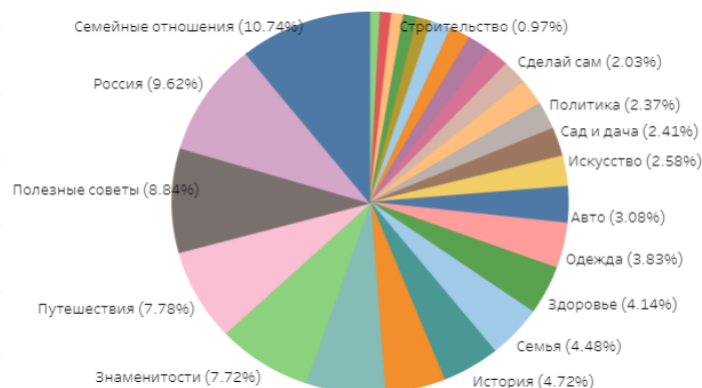
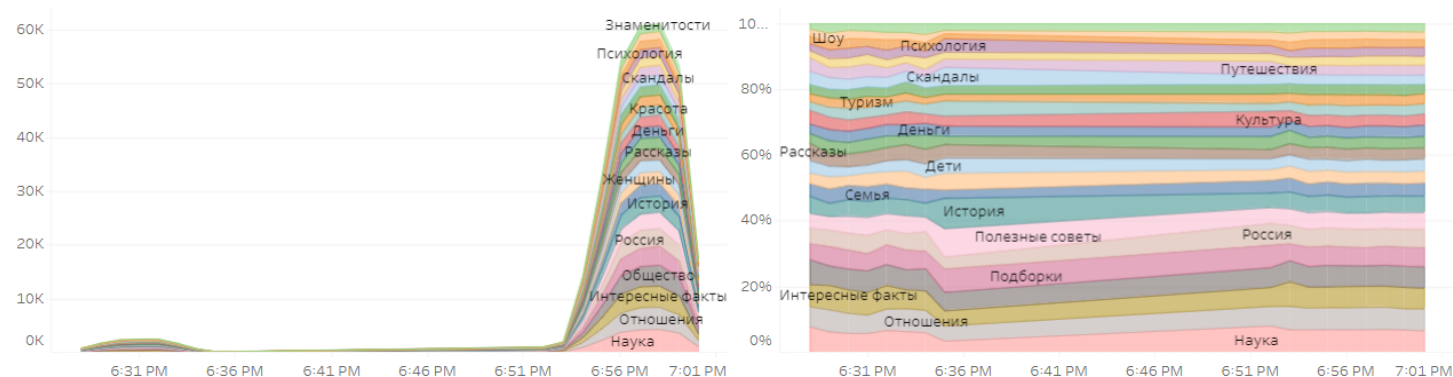
(All)

Выберите возрастную группу(ы):

(All)

Данный дашборд визуализирует:

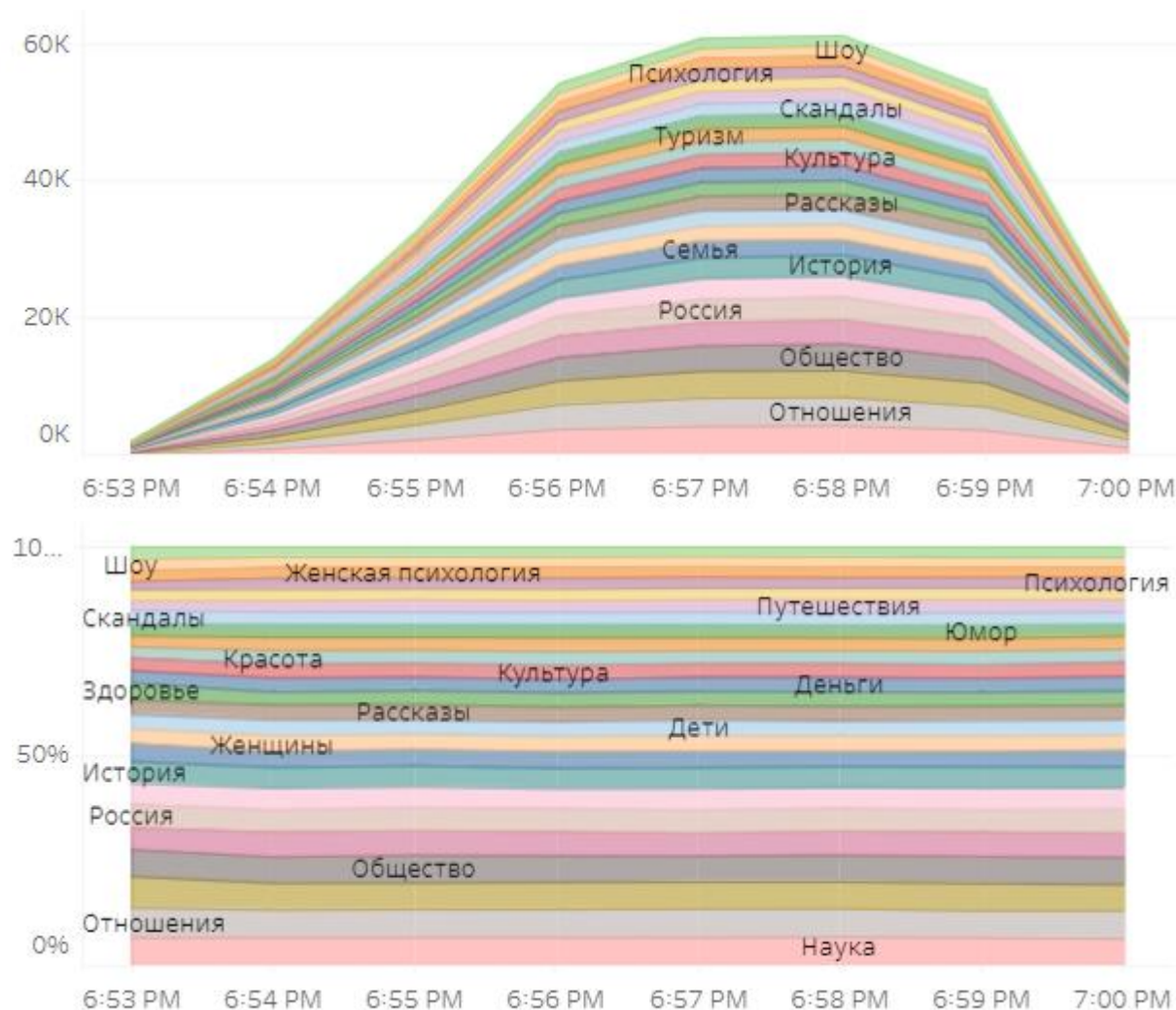
1. В **таблице внизу**: кол-во взаимодействий читателей с темами статей (строки) и темами источников (столбцы). Подсвечены большие значения.
  2. На **левом графике**: временную зависимость абсолютного кол-ва взаимодействий в разбивке по темам статей.
  3. На **графике в центре**: то же, что и на левом графике, но в относительных величинах - за 100% принимается кол-во взаимодействий в данную минуту.
  4. На **правом графике**: относительные доли взаимодействий с темами источников за весь выбранный период.
- Фильтры в правом верхнем углу позволяют выбирать а) временной период показываемых данных, б) подмножество тем статей Яндекс.Дзен, в) возрастную группу аудитории.



	Авто	Деньги	Дети	Еда	Здоровье	Знамени..	Интерье..	Искусст..	История	Кино	Музыка	Одежда	Полезны..	Политика	Психоло..	Путешес..	Ремонт	Россия	Сад и да..	Сделай с..	Семейн..	Семья	Спорт	Строите..	Техноло..	Финансы
Знаменитости	29	55	145	36	139	445	12	356	644	408	55	467	178	35	88	299	3	2,275	21	6	764	234	433	50	158	59
Интересные факты	1,254	145	327	391	2,090	720	54	697	1,273	1,033	325	837	2,023	340	449	1,448	175	2,567	825	341	1,323	611	186	234	219	55
Искусство	145	98	639	7	70	1,083	164	512	596	1,254	223	209	331	121	98	664	172	632	43	128	754	404	54	53	58	4
История	522	114	1,047	60	771	921	30	695	636	1,091	164	310	859	695	227	2,643	168	1,491	209	149	1,344	696	190	9	286	62
Красота	29	48	703	253	736	1,313	185	232	177	472	38	428	1,227	282	171	781	172	269	513	466	748	274	28	174	39	56
Культура	103	46	578	55	41	1,508	124	682	636	1,543	178	295	231	308	81	430	154	1,474	100	76	892	482	47	86	46	9
Наука	1,606	197	930	91	629	1,554	186	510	1,102	3,279	403	393	1,865	244	291	1,823	415	1,782	227	358	1,711	809	421	99	749	62
Общество	379	350	1,007	160	1,181	829	64	218	1,116	578	79	355	1,313	1,209	919	891	181	3,471	76	162	2,727	1,416	338	97	118	406
Отношения	478	281	242	573	1,173	2,040	862	374	679	672	139	1,428	2,716	830	197	1,269	510	2,278	944	517	1,030	194	312	513	119	296
Подборки	551	217	319	871	396	1,353	122	257	597	1,403	190	1,612	2,795	43	396	1,491	479	877	522	364	1,386	424	280	361	440	26
Полезные советы	749	916	707	223	1,346	271	427	83	437	143	127	891	1,413	51	438	2,088	485	403	253	679	1,781	811	136	77	294	206
Психология	19	86	1,233	62	583	582	46	306	286	219	19	183	862	224	141	261	34	936	86	45	855	656	125	88	23	76
Путешествия	343	363	509	357	187	509	26	555	1,018	289	33	382	888	60	126	641	451	588	499	127	709	306	88	85	44	77
Рассказы	78	454	286	28	248	615	47	291	568	357	36	30	358	63	233	4,587	123	421	201	92	875	539	222	74	1	82
Россия	1,885	229	493	23	282	1,579	134	355	1,104	919	324	276	730	232	38	1,079	345	2,847	85	114	1,962	608	703	179	316	125
Семья	50	458	361	1,236	275	689	321	288	266	486	41	544	1,293	78	176	340	447	519	548	434	1,489	1,131	4	351	34	38
Скандалы	14	35	844	48	43	1,992	93	91	410	707	132	281	54	374	18	284	13	1,560	24	56	952	386	597		273	13

# Обзор возможностей дашборда

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



Левый и центральный графики дашборда показывают временную динамику взаимодействий пользователей с карточками статей:

- Левый (относительно схемы дашборда) график показывает абсолютные значения взаимодействий читателей.
- Центральный (относительно схемы дашборда) график показывает относительные значения взаимодействий читателей – за каждую минуту агрегации данных.
- Показываемые на графиках темы карточек статей можно выбирать из выпадающего списка – в области фильтров дашборда (правый верхний угол дашборда). NB!: при этом относительные проценты вычисляются только по выбранному подмножеству тем статей.
- В анализ можно включать подмножество возрастных групп читательской аудитории, выбирая его из выпадающего списка – в области фильтров дашборда (правый верхний угол дашборда). NB!: при этом относительные проценты вычисляются только внутри выбранной возрастной группы читателей.
- Временной диапазон визуализации данных можно изменять в области фильтров дашборда (правый верхний угол дашборда).



# Обзор возможностей дашборда

Как много карточек генерируют источники с разными темами?

Правый (относительно схемы дашборда) график показывает доли взаимодействий пользователей с темами источников Дзена:

- Для увеличения наглядности (согласно техническому заданию) дашборд использует тип визуализации «пирог» (pie).
- Темы источников указываются по их названиям, за которыми в скобках приводится значение их доли в процентах – от общего «пирога». Куски «пирога» упорядочены по убыванию их долей против часовой стрелки.
- В анализ можно включать подмножество возрастных групп читательской аудитории, выбирая его из выпадающего списка – в области фильтров дашборда (правый верхний угол дашборда).
- Временной диапазон визуализации данных можно изменять в области фильтров дашборда (правый верхний угол дашборда).
- NB!: при этом относительные проценты вычисляются только внутри выбранных фильтрами срезов данных.
- Согласно техническому заданию и утвержденному макету дашборда, фильтры дашборда не позволяют выбирать подмножество тем источников – по умолчанию показаны все 26 тем источников.

## фильтры данных:

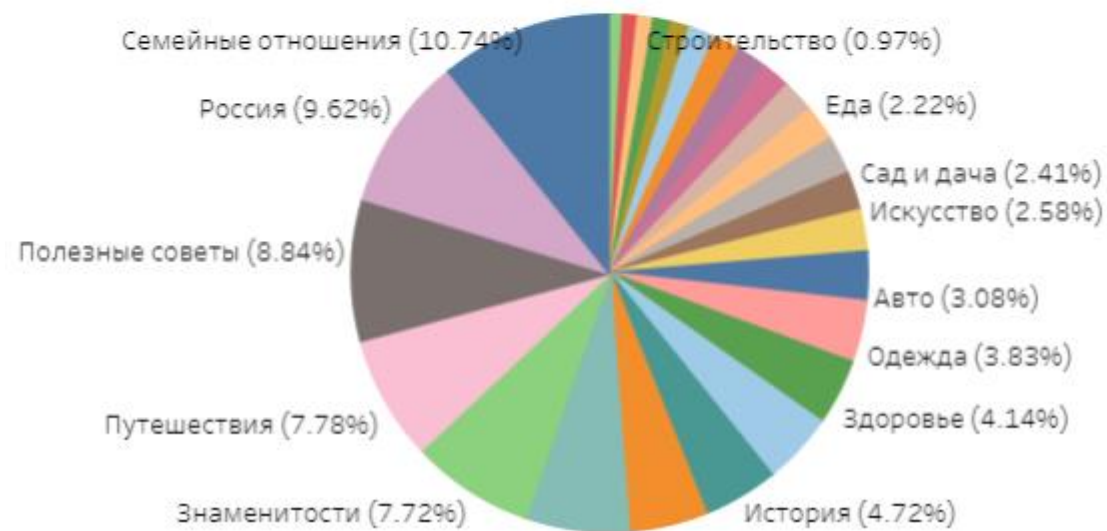
Выберите период исходных данных:

7:00:00 PM

темы статей:

Выберите возрастную группу(ы):

(All)



# Обзор возможностей дашборда

Как соотносятся темы карточек и темы источников?

Нижняя (относительно схемы дашборда) таблица дашборда показывают количество взаимодействий пользователей с карточками статей по темам источников: строки таблицы соответствуют темам (карточек) статей, а столбцы – темам источников Дзен.

- Показываемые на графиках темы карточек статей можно выбирать из выпадающего списка – в области фильтров дашборда (правый верхний угол дашборда).
- В анализ можно включать подмножество возрастных групп читательской аудитории, выбирая его из выпадающего списка – в области фильтров дашборда (правый верхний угол дашборда).
- Временной диапазон визуализации данных можно изменять в области фильтров дашборда (правый верхний угол дашборда).
- Для удобства визуального анализа таблица снабжена цветовой подсветкой – большие значения взаимодействий подсвечены более интенсивно.
- Хотя дашборд обладает автоматическим масштабированием, данная таблица слишком велика даже для больших экранов – используйте соответствующие полосы прокрутки для просмотра всей таблицы

	Авто	Деньги	Дети	Еда	Здоровье	Знамени..	Интерье..	Искусст..	История	Кино	Музыка	Одежда	Полезны..	Политика	Психоло..	Путешес..	Ремонт	Россия	Сад и да..	Сделай с..	Семейн..	Семья	Спорт	Строите..	Техноло..	Финанс
Общество	379	350	1,007	160	1,181	829	64	218	1,116	578	79	355	1,313	1,209	919	891	181	3,471	76	162	2,727	1,416	338	97	118	40
Отношения	478	281	242	573	1,173	2,040	862	374	679	672	139	1,428	2,716	830	197	1,269	510	2,278	944	517	1,030	194	312	513	119	29
Подборки	551	217	319	871	396	1,353	122	257	597	1,403	190	1,612	2,795	43	396	1,491	479	877	522	364	1,386	424	280	361	440	2
Полезные с..	749	916	707	223	1,346	271	427	83	437	143	127	891	1,413	51	438	2,088	485	403	253	679	1,781	811	136	77	294	20
Психология	19	86	1,233	62	583	582	46	306	286	219	19	183	862	224	141	261	34	936	86	45	855	656	125	88	23	7
Путешеств..	343	363	509	357	187	509	26	555	1,018	289	33	382	888	60	126	641	451	588	499	127	709	306	88	85	44	7
Рассказы	78	454	286	28	248	615	47	291	568	357	36	30	358	63	233	4,587	123	421	201	92	875	539	222	74	1	8
Россия	1,885	229	493	23	282	1,579	134	355	1,104	919	324	276	730	232	38	1,079	345	2,847	85	114	1,962	608	703	179	316	12
Семья	50	458	361	1,236	275	689	321	288	266	486	41	544	1,293	78	176	340	447	519	548	434	1,489	1,131	4	351	34	3
Скандалы	14	35	844	48	43	1,992	93	91	410	707	132	281	54	374	18	284	13	1,560	24	56	952	386	597		273	1
Туризм	161	74	406	382	311	269	279	173	618	252	77	248	721	307	133	483	403	867	523	688	1,580	283	79	74	38	8
Шоу	4	8	297	1	20	829	30	66	215	2,201	98	393	38	522	7	113	7	505			1,503	151	141	8		35
Юмор	224	79	404	194	124	334	17	205	715	502	67	384	1,150	17	140	636	201	621	321	204	1,887	584	281	55	11	4

# Общие выводы

## Результаты проекта:

- Автоматизирована отработка регулярных запросов продуктового маркетинга о статистике поведения пользователей (читателей) Яндекс.Дзен:
  - Сколько взаимодействий пользователей с карточками происходит в системе с разбивкой по темам карточек?
  - Как много карточек генерируют источники с разными темами?
  - Как соотносятся темы карточек и темы источников?
- Теперь продуктовый маркетинг может самостоятельно получить онлайн ответы на данный список вопросов в любое время, используя дашборд, доступный по линку:  
[https://public.tableau.com/views/DashboardDA49SevastyanovAA/DashboardDA49SevastyanovAA?:language=en-US&publish=yes&:display\\_count=n&:origin=viz\\_share\\_link](https://public.tableau.com/views/DashboardDA49SevastyanovAA/DashboardDA49SevastyanovAA?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link)
- Дашборд автоматически загружает данные из базы данных.
- Данные собирает и обрабатывает автоматизированный инженерами данных пайплайн сбора и обработки данных.

## Чего нет в результатах проекта:

- Предоставленный продуктовому маркетингу дашборд является инструментом анализа данных. Его разработка является финальным результатом проекта. Проект не включает никакого исследовательского анализа данных – ни с помощью дашборда, ни с помощью других средств анализа.
- В проекте не делается никаких выводов относительно характера поведения пользователей приложения Яндекс.Дзен – предполагается, что продуктовый маркетинг продолжает прежнюю практику определения продуктовых метрик по данным взаимодействия читателей, которые теперь поставляются автоматически в режиме онлайн пользователю дашборда в согласованной с бизнес заказчиком форме.
- На текущий момент реализован некоммерческий прототип решения на Tableau Public, поэтому загрузка данных из БД производилась вручную – согласно ограничениям Tableau Public. Для запуска в коммерческую эксплуатацию необходима реализация автоматической загрузки данных из БД в корпоративную систему Tableau.

СПАСИБО !