

Карточки и топики Яндекс Дзен

Статистика за 24 Сентября 2019 года
Подготовил Хоружевский Георгий

Цели и задачи.

Цели:

Составление дашборда для менеджеров
Яндекс Дзен, который отвечал бы на
вопросы о темах карточек

Задачи:

Ответить на вопросы:

1. Сколько взаимодействий пользователей с карточками происходит в системе с разбивкой по темам карточек?
2. Как много карточек генерируют источники с разными темами?
3. Как соотносятся темы карточек и темы источников?

Данные

дата-инженеры подготовили агрегирующую таблицу `dash_visits`.

Таблица хранится в базе данных `zen`.

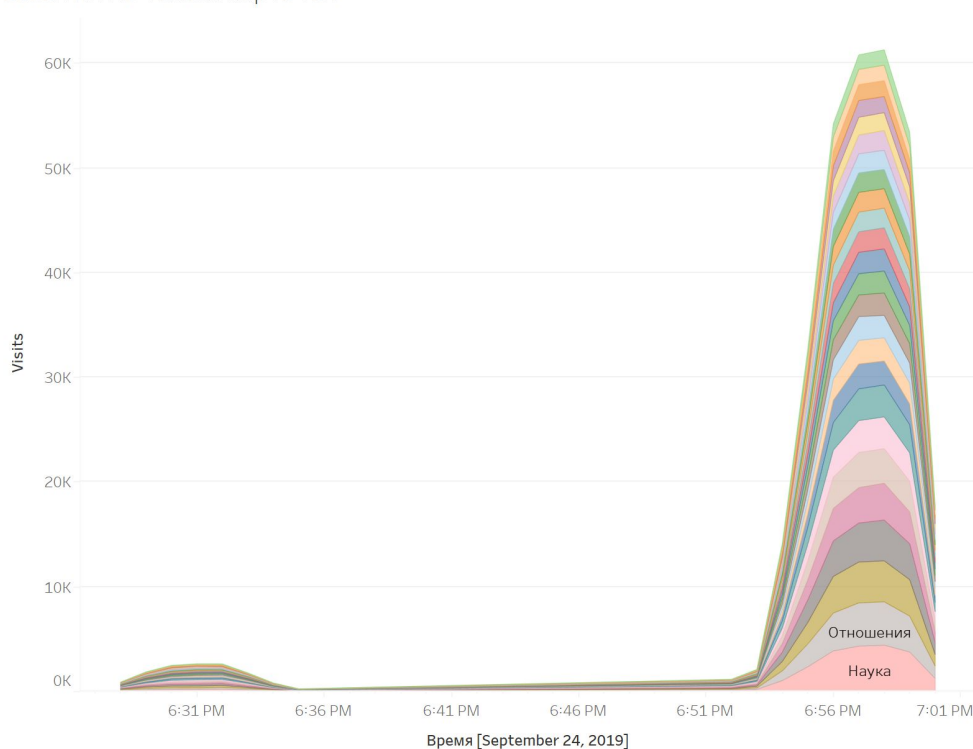
Структура таблицы:

`record_id` — первичный ключ,
`item_topic` — тема карточки,
`source_topic` — тема источника,
`age_segment` — возрастной сегмент,
`dt` — дата и время,
`visits` — количество событий.

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

1/2

События по темам карточек

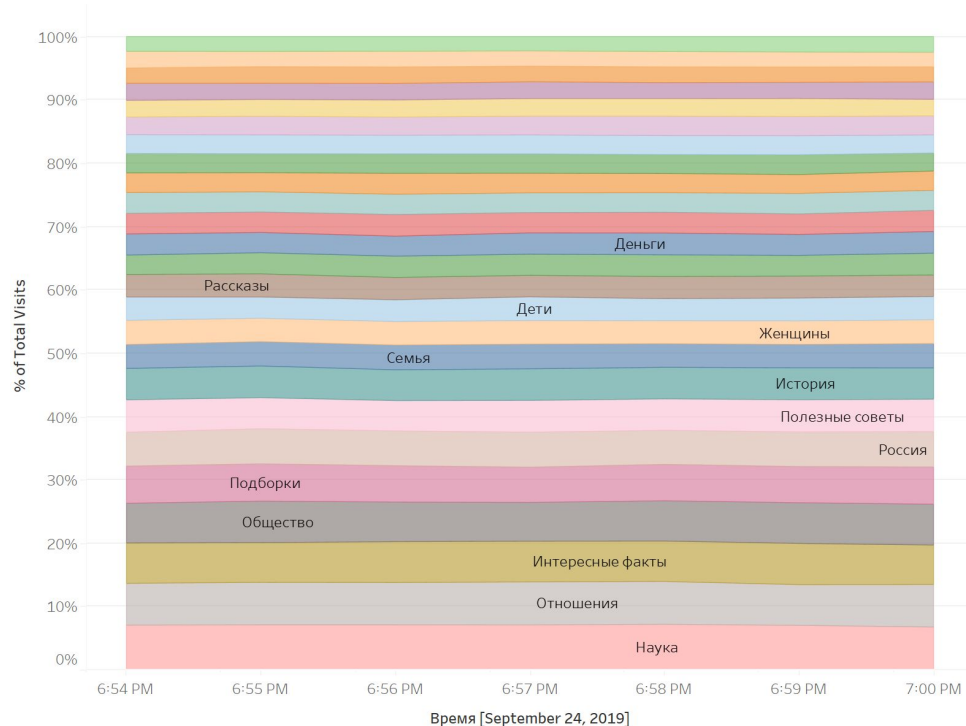


Количество взаимодействий критически зависит от времени (возможно, в силу особенностей сбора данных), но на пике в наиболее популярных категориях Наука и Отношения в минуту происходит более 4000 взаимодействий.

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

2/2

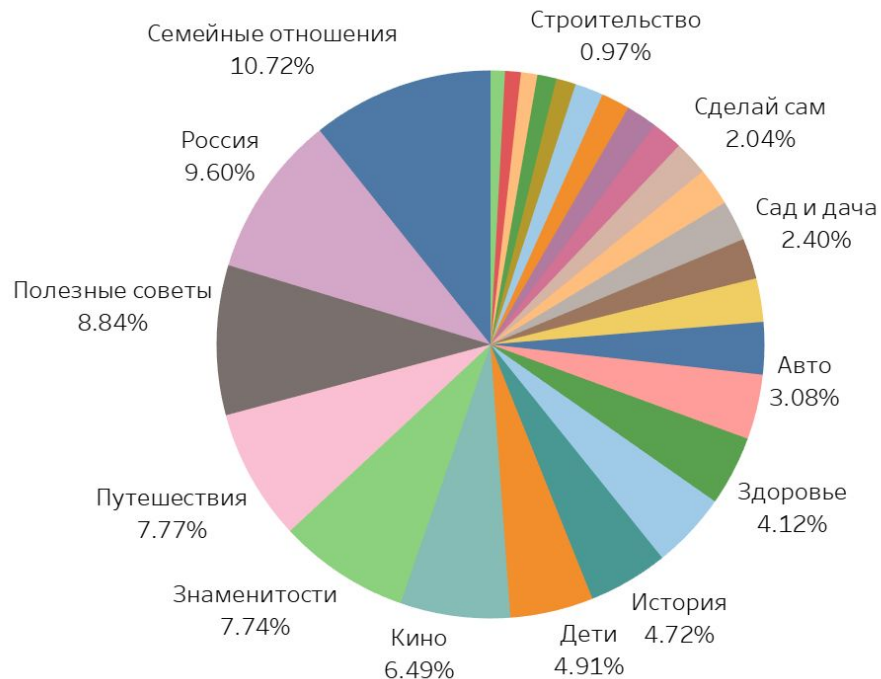
% событий по темам карточек



В то же время, в процентном соотношении взаимодействия по карточкам стабильны во времени.

Наука и Отношения занимают примерно по 7%,
Интересные факты, Общество и Подборки - 6%
Россия и Полезные советы - 5%

Как много карточек генерируют источники с разными темами?



Семейные отношения указаны в источниках в 10.72% случаев
Россия - 9.6%
Полезные советы - 8.84% и т.д.

Наименее популярны Финансы, Музыка и Строительство - менее процента.
Скорее всего, причина в том, что это нишевые темы, в то время как лидируют наиболее общие топики

Как соотносятся темы карточек и темы источников?

1/3

Темы источников - темы карточек

Item Topic	Source Topic														
	Авто	Деньги	Дети	Еда	Здоровь..	Знамен..	Интерь..	Искусст..	История	Кино	Музыка	Одежда	Полезны е советы	Полити..	Пси
Деньги	39	33	41	31	23	39	23	37	39	41	25	30	42	42	
Дети	34	38	42	41	37	42	30	37	40	40	6	37	42	35	
Женская пс..	14	36	41	19	37	39		29	33	38	3	41	36	16	
Женщины	35	28	41	39	42	41	37	39	39	40	22	39	41	26	
Здоровье	36	32	40	40	39	42	23	25	25	33	14	42	42	31	
Знаменито..	17	28	31	16	33	41	10	39	41	40	17	39	30	20	
Интересны..	42	33	42	38	42	41	25	41	41	40	37	41	42	38	
Искусство	32	33	40	7	25	42	32	40	40	42	37	39	38	30	
История	42	35	41	23	42	40	17	41	40	42	33	37	42	40	
Красота	16	20	41	37	42	42	34	37	36	39	22	40	41	35	
Культура	28	21	41	20	23	42	30	41	41	42	36	38	35	35	
Наука	42	36	40	25	41	42	32	42	41	42	40	40	42	37	
Общество	40	40	40	33	42	42	23	36	41	38	20	39	42	41	
Отношения	42	34	37	39	42	42	41	40	41	39	33	42	42	41	
Подборки	41	34	39	42	39	41	32	37	42	41	37	42	42	21	
Полезные с..	42	40	40	33	42	38	40	27	40	34	33	42	42	20	
Психология	14	29	41	23	40	40	21	36	35	37	11	35	41	36	
Путешеств..	39	41	40	37	36	40	16	42	42	38	18	41	42	25	
Рассказы	27	38	36	18	37	40	24	38	40	40	15	19	39	29	
Россия	42	35	40	11	36	42	34	39	41	42	39	39	41	38	
Семья	25	42	39	42	37	40	39	38	36	37	19	41	42	27	
Скандалы	10	22	40	21	20	42	26	28	40	42	28	40	19	39	
Туризм	36	24	40	39	39	39	38	36	41	37	28	38	41	37	
Шоу	4	8	39	1	12	42	14	24	34	42	30	39	18	39	
Юмор	37	26	38	30	30	39	14	36	40	41	25	38	42	11	

Интересно, что некоторые темы источников полностью отсутствуют среди топиков. Например, тема Авто присутствует только в источниках.

Как соотносятся темы карточек и темы источников?

2/3

Темы источников - темы карточек

Source Topic														
Item Topic	Авто	Деньги	Дети	Еда	Здоровь..	Знамен..	Интерь..	Искусст..	История	Кино	Музыка	Одежда	Полезны е советы	Полити..
Россия	42	35	40	11	36	42	34	39	41	42	39	41	38	
Полезные с..	42	40	40	33	42	38	40	27	40	34	33	42	42	20
Отношения	42	34	37	39	42	42	41	40	41	39	33	42	42	41
Наука	42	36	40	25	41	42	32	42	41	42	40	40	42	37
История	42	35	41	23	42	40	17	41	40	42	33	37	42	40
Интересны..	42	33	42	38	42	41	25	41	41	40	37	41	42	38
Подборки	41	34	39	42	39	41	32	37	42	41	37	42	42	21
Общество	40	40	40	33	42	42	23	36	41	38	20	39	42	41
Путешеств..	39	41	40	37	36	40	16	42	42	38	18	41	42	25
Деньги	39	33	41	31	23	39	23	37	39	41	25	30	42	42
Юмор	37	26	38	30	30	39	14	36	40	41	25	38	42	11
Туризм	36	24	40	39	39	39	38	36	41	37	28	38	41	37
Здоровье	36	32	40	40	39	42	23	25	25	33	14	42	42	31
Женщины	35	28	41	39	42	41	37	39	39	40	22	39	41	26
Дети	34	38	42	41	37	42	30	37	40	40	6	37	42	35
Искусство	32	33	40	7	25	42	32	40	40	42	37	39	38	30
Культура	28	21	41	20	23	42	30	41	41	42	36	38	35	35
Рассказы	27	38	36	18	37	40	24	38	40	40	15	19	39	29
Семья	25	42	39	42	37	40	39	38	36	37	19	41	42	27
Знаменито..	17	28	31	16	33	41	10	39	41	40	17	39	30	20
Красота	16	20	41	37	42	42	34	37	36	39	22	40	41	35
Психология	14	29	41	23	40	40	21	36	35	37	11	35	41	36
Женская пс..	14	36	41	19	37	39		29	33	38	3	41	36	16
Скандалы	10	22	40	21	20	42	26	28	40	42	28	40	19	39
Шоу	4	8	39	1	12	42	14	24	34	42	30	39	18	39

При этом, значения в таблице не превышают 42

Источник Интерьеры не упоминается ни разу для карточек о Женской психологии.

Как соотносятся темы карточек и темы источников?

3/3

Темы источников - темы карточек

		Source Topic														
		Семейны		Полезны												
Item Topic	Сем..	Росси	е отно..	е советы	Дети	Знамен..	Здоров..	Кино	Полити..	Искусст..	Путеше..	Психол..	Одежда	История	Спорт	
Психолог..	42	42	41	41	41	40	40	37	36	36	35	35	35	35		
Полезные с..	42	40	42	42	40	38	42	34	20	27	42	40	42	40		
Общество	42	42	42	42	40	42	42	38	41	36	42	42	39	41		
Наука	42	42	42	42	40	42	41	42	37	42	42	39	40	41		
История	42	42	42	42	41	40	42	42	40	41	42	36	37	40		
Семья	41	37	41	42	39	40	37	37	27	38	37	37	41	36		
Женщины	41	41	42	41	41	41	42	40	26	39	36	39	39	39		
Женская пс..	41	41	42	36	41	39	37	38	16	29	36	40	41	33		
Дети	41	42	42	42	42	42	37	40	35	37	42	32	37	40		
Рассказы	40	39	41	39	36	40	37	40	29	38	42	36	19	40		
Юмор	39	39	42	42	38	39	30	41	11	36	41	33	38	40		
Подборки	39	42	42	42	39	41	39	41	21	37	42	39	42	42		
Культура	39	42	40	35	41	42	23	42	35	41	41	29	38	41		
Красота	39	39	41	41	41	42	42	39	35	37	41	33	40	36		
Искусство	39	40	42	38	40	42	25	42	30	40	40	28	39	40		
Интересны..	39	42	42	42	42	41	42	40	38	41	41	39	41	41		
Знаменито..	39	42	40	30	31	41	33	40	20	39	40	28	39	41		
Скандалы	38	42	40	19	40	42	20	42	39	28	37	13	40	40		
Россия	38	42	42	41	40	42	36	42	38	39	42	22	39	41		
Путешеств..	36	41	41	42	40	40	36	38	25	42	42	32	41	42		
Деньги	36	41	42	42	41	39	23	41	42	37	40	39	30	39		
Отношения	35	42	42	42	37	42	42	39	41	40	42	36	42	41		
Шоу	34	37	42	18	39	42	12	42	39	24	31	3	39	34		
Туризм	34	41	42	41	40	39	39	37	37	36	41	32	38	41		
Здоровье	33	40	41	42	40	42	39	33	31	25	39	36	42	25		

Можно отметить, что сочетания наиболее общих тем дают наибольший результат. Например, в теме Общество с источником Россия 42 взаимодействия - рекордное количество.

В то же время, тема Рассказы имеет источник Технологии лишь 1 раз - логично, тема и источник являются нишевыми и мало связанными друг с другом.

Выводы.

1. Взаимодействие с карточками не стабильно в количественном отношении во времени, но стабильно в процентном соотношении.
Темы Наука и Отношения занимают по 7% взаимодействий.
Наименее популярны нишевые топики, такие как Шоу и Женская психология.
2. Семейные отношения и Россия - наиболее мощные источники топики, они генерируют примерно по 10% карточек. Меньше всего приходится на источники Финансы, Музыка и Строительство - менее процента.
3. Соотношение источников и тем карточек сложно описать единой формулой. Можно лишь отметить, что наиболее популярными являются пересечения наиболее общих тем и источников.