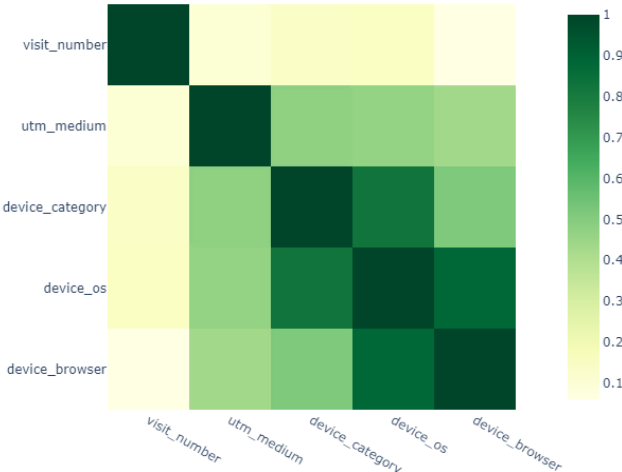
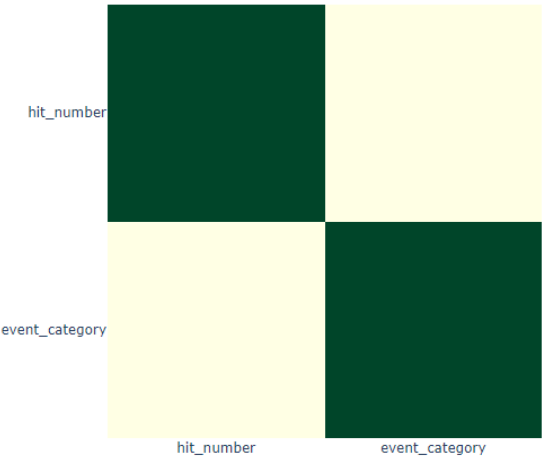


# Data Preparation

Sessions	Hits
Описание атрибутов	
<i>Одна строка = один визит на сайт.</i>	<i>Одна строка = одно событие в рамках одного визита на сайт</i>
session_id — ID визита; client_id — ID посетителя; visit_date — дата визита; visit_time — время визита; visit_number — порядковый номер визита клиента; utm_source — канал привлечения; utm_medium — тип привлечения; utm_campaign — рекламная кампания; utm_keyword — ключевое слово; device_category — тип устройства; device_os — ОС устройства; device_brand — марка устройства; device_model — модель устройства; device_screen_resolution — разрешение экрана; device_brand — марка устройства; device_model — модель устройства; device_screen_resolution — разрешение экрана; device_browser — браузер; geo_country — страна; geo_city — город.	session_id — ID визита; hit_date — дата события; hit_time — время события; hit_number — порядковый номер события в рамках сессии; hit_type — тип события; hit_referer — источник события; hit_page_path — страница события; event_category — тип действия; event_action — действие; event_label — тег действия; event_value — значение результата действия.
Размер выборки	
1860042 строк, 17 столбцов	15726470 строк, 11 столбцов
Дублирующихся строк нет.	Дублирующихся строк нет.
Данные приведены за время с 19 мая по 31 декабря 2021 года	Данные приведены за время с 22 сентября по 31 декабря 2021 года

<p>Типичный посетитель:</p> <p><b>Москвич</b>, заходящий на сайт</p> <p>с <b>мобильного телефона</b> марки <b>Apple</b></p> <p>через браузер <b>Chrome</b></p> <p>по нажатию на <b>баннер</b></p>	<p>Самая часто посещаемая страница - <b>podpiska.sberauto.com</b>,</p> <p>источник - <b>HboIMJUevbIAbkHCIEQa</b>,</p> <p>около половины действий происходят с <b>карточкой</b> (чаще всего - <b>просмотр</b>).</p> <p>Признаки <b>hit_type</b> и <b>event_value</b> не несут никакой информации.</p>
 <p><b>Операционная система</b> устройства сильно коррелирует с <b>категорией устройства и браузером</b>.</p> <p>Между <b>категорией устройства и браузером</b> - средняя корреляция</p>	 <p>Значимых корреляций не наблюдается</p>
Генерация признаков	
<p>month</p> <p>traffic_type</p> <p>is_from_social_network</p>	<p>brand</p> <p>model</p> <p>is_target</p>
Удаление столбцов	
<p>client_id</p> <p>visit_number</p> <p>utm_adcontent</p> <p>visit_date_time</p> <p>utm_keyword</p> <p>device_os</p> <p>device_model</p> <p>device_screen_resolution</p> <p>device_browser</p>	<p>hit_type</p> <p>hit_referer</p> <p>hit_page_path</p> <p>hit_date_time</p> <p>hit_number</p> <p>event_category</p> <p>event_action</p> <p>event_label</p> <p>event_value</p>

# Создание новых датафреймов

auto_brand	df
session_id brand model is_target	target_session utm_source utm_medium utm_campaign device_category device_brand geo_country geo_city week month traffic_type region is_from_social_network
Заполнение пропусков	
Пропуски в модели автомобиля -> (not set)	2344 строки, в которых отсутствуют данные по 10 признакам, удалены  Пропуски в device_brand и utm_campaign -> (not set)

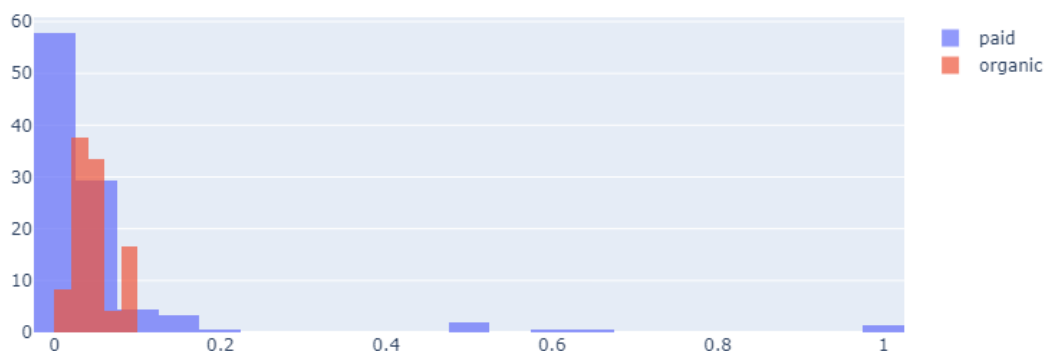
# Проверка гипотез

## Органический трафик не отличается от платного с точки зрения Conversion Rate в целевые события

Сводная таблица по типам трафика:

	target_sessions	all_sessions	cr
organic	20812	468920	0.044383
paid	29497	1263270	0.023350

Конверсия всего органического трафика с мая по декабрь 2021 года **выше** конверсии платного за тот же период на **2,1%**



### Тест Шапиро-Уилки

**H0:** выборки распределены нормально

Organic traffic: *statistic*=0.9308854937553406, *pvalue*=0.10203346610069275

Paid traffic: *statistic*=0.38410788774490356, *pvalue*=9.28321731388862e-23

Выборка платного трафика **не соответствует нормальному распределению**, поэтому используем непараметрический критерий

### Критерий Манна Уитни

**H0:** Органический трафик не отличается от платного с точки зрения CR в целевые события

**H1:** CR в целевые события от органического трафика выше, чем от платного

*statistic*=2678.0, *pvalue*=0.00017157920334444202

Данные позволяют отвергнуть нулевую гипотезу и сделать вывод:

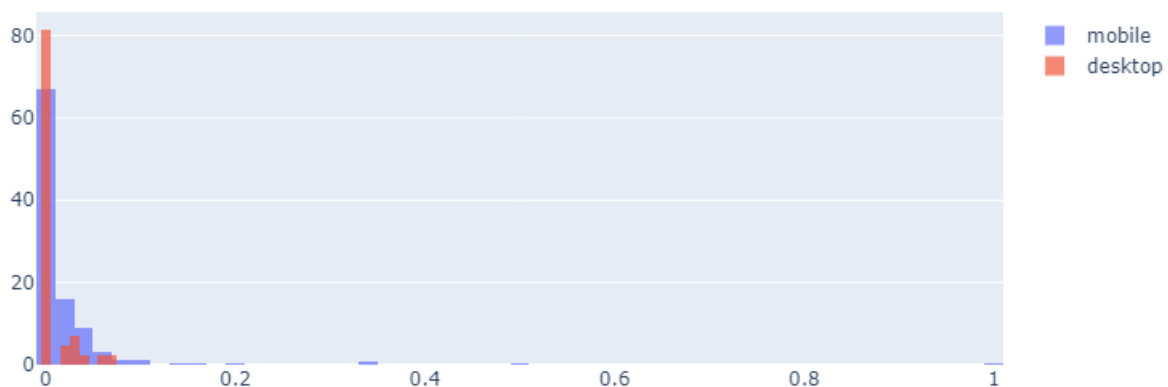
**Конверсия в целевые события от органического трафика выше, чем от платного**

# Трафик с мобильных устройств не отличается от трафика с десктопных устройств с точки зрения Conversion Rate в целевые события

Сводная таблица по типам устройств:

	target_sessions	all_sessions	cr
desktop	11510	346843	0.033185
mobile	38377	1368651	0.028040
tablet	422	16696	0.025276

Конверсия трафика с десктопных устройств с мая по декабрь 2021 года **выше** конверсии трафика с мобильных устройств за то же время



## Тест Шапиро-Уилки

**H0:** выборки распределены нормально

Mobile: statistic=0.24334490299224854, pvalue=0.0

Desktop: statistic=0.4796009659767151, pvalue=3.7743894815145396e-11

Выборки **не соответствуют нормальному распределению**, поэтому используем непараметрический критерий

## Критерий Манна Уитни

**H0:** Трафик с мобильных устройств не отличается от трафика с десктопных устройств с точки зрения CR в целевые события

**H1:** CR в целевые события от трафика с мобильных устройств выше, чем с десктопных  
statistic=21238.0, pvalue=0.019602618881964764

Данные позволяют отвергнуть нулевую гипотезу и сделать вывод:

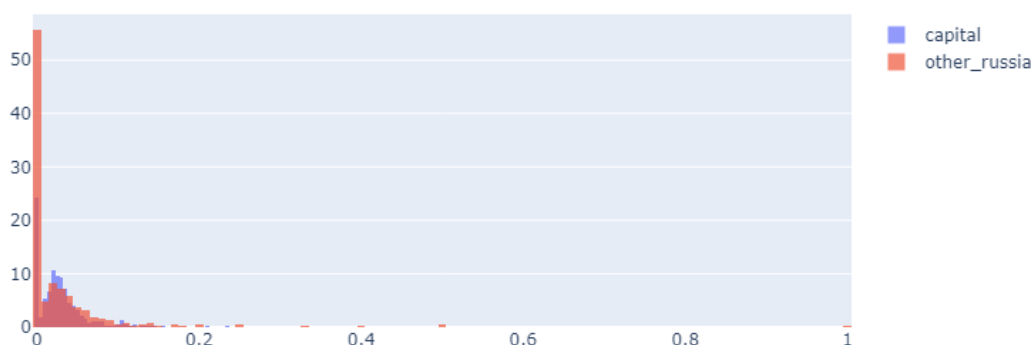
**Конверсия в целевые события от трафика с мобильных устройств выше, чем с десктопных**

# Трафик из городов присутствия (Москва и область, Санкт-Петербург) не отличается от трафика из иных регионов с точки зрения Conversion Rate в целевые события

Сводная таблица по регионам:

	target_sessions	all_sessions	cr
capital	33749	1127768	0.029925
other_russia	15402	554579	0.027772

Конверсия трафика из Мск, МО и СПб с мая по декабрь 2021 года **выше** конверсии трафика из других регионов России на **0,22%**



## Тест Шапиро-Уилки

**H0:** выборки распределены нормально

Capital: statistic=0.7998051643371582, pvalue=1.3265492870805587e-24

Other Russia: statistic=0.36756837368011475, pvalue=0.0

Выборки **не соответствуют нормальному распределению**, поэтому используем непараметрический критерий

## Критерий Манна Уитни

**H0:** Трафик из городов присутствия не отличается от трафика из иных регионов с точки зрения CR в целевые события

**H1:** CR в целевые события от трафика из городов присутствия выше, чем из иных регионов России  
statistic=919546.0, pvalue=1.3709502965637892e-21

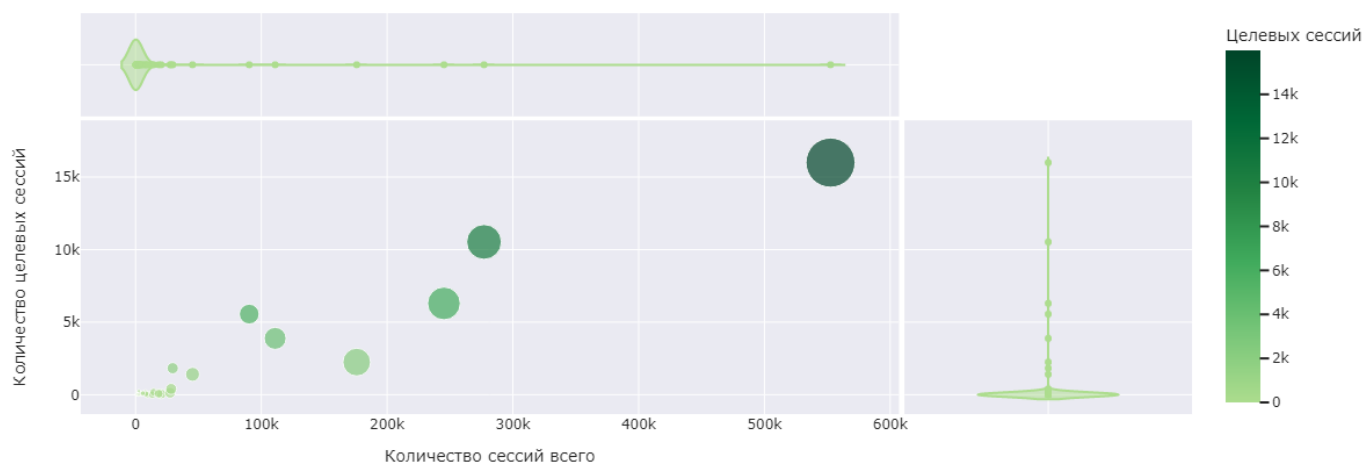
Данные позволяют отвергнуть нулевую гипотезу и сделать вывод:

**Конверсия в целевые события от трафика из городов присутствия выше, чем из иных регионов России**

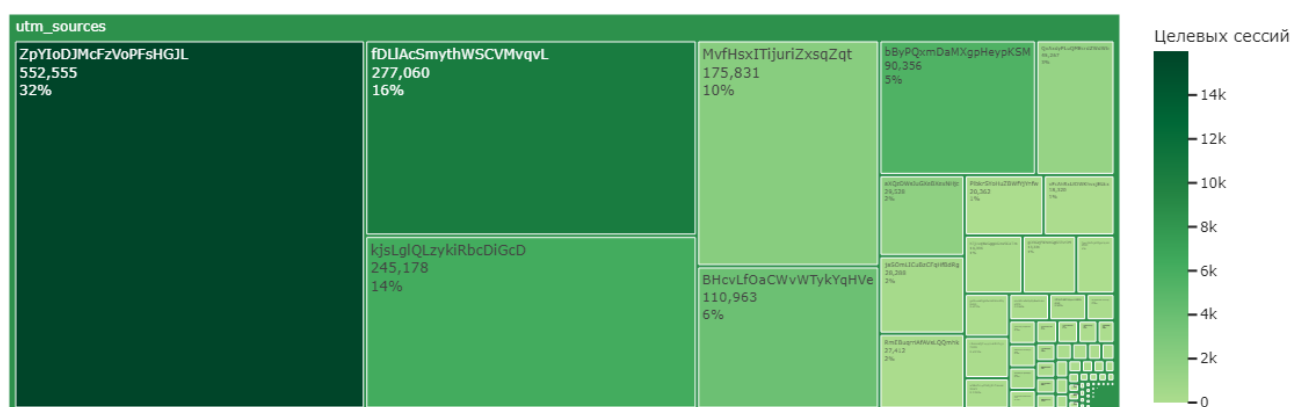
# Ответы на вопросы продуктовой команды

## Из каких источников идёт самый целевой трафик?

Распределение количественных показателей источников трафика.  
Размер - количество сессий. Цвет - количество целевых сессий.



Распределение трафика по источникам за все время.  
Размер - количество сессий всего. Цвет - количество целевых сессий.



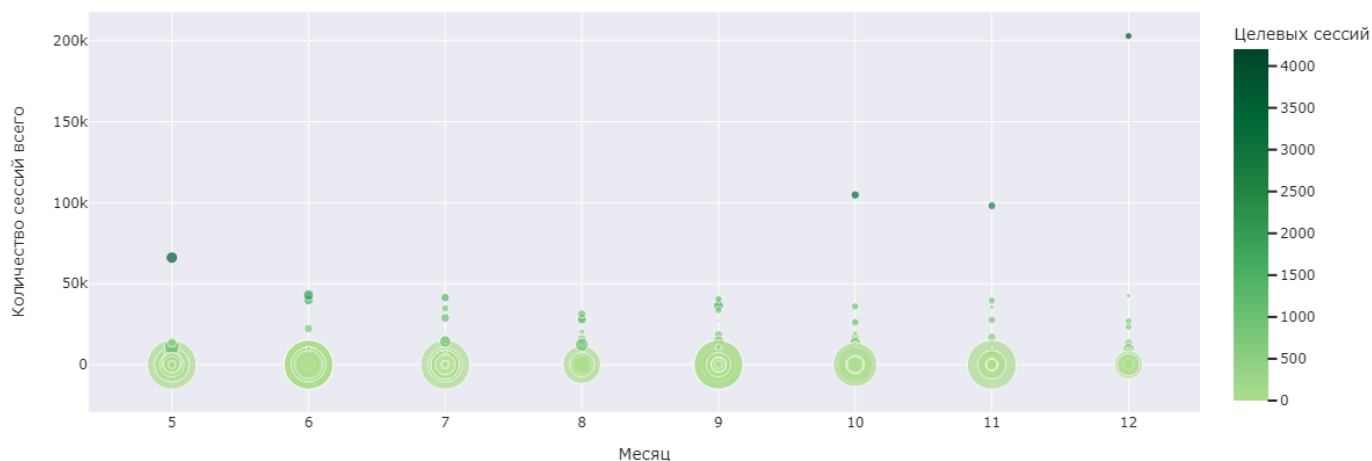
### Топ источников трафика (более 5%):

1. ZpYIoDJMcFzVoPFsHGJL - 32%
2. fDLIAcSmythWSCVMvqvL - 16%
3. kjsLglQLzykiRbcDiGcD - 14%
4. MvfHsxITijuriZxsqZqt - 10%
5. BHcvLfOaCWvWTykYqHVe - 6%
6. bByPQxmDaMXgpHeypKSM - 5%

### Топ источников целевого трафика (более 5%):

1. ZpYIoDJMcFzVoPFsHGJL - 32%
2. fDLIAcSmythWSCVMvqvL - 21%
3. kjsLglQLzykiRbcDiGcD - 13%
4. MvfHsxITijuriZxsqZqt - 11%
5. BHcvLfOaCWvWTykYqHVe - 8%

Распределение трафика по месяцам.  
Размер - конверсия. Цвет - количество целевых сессий.



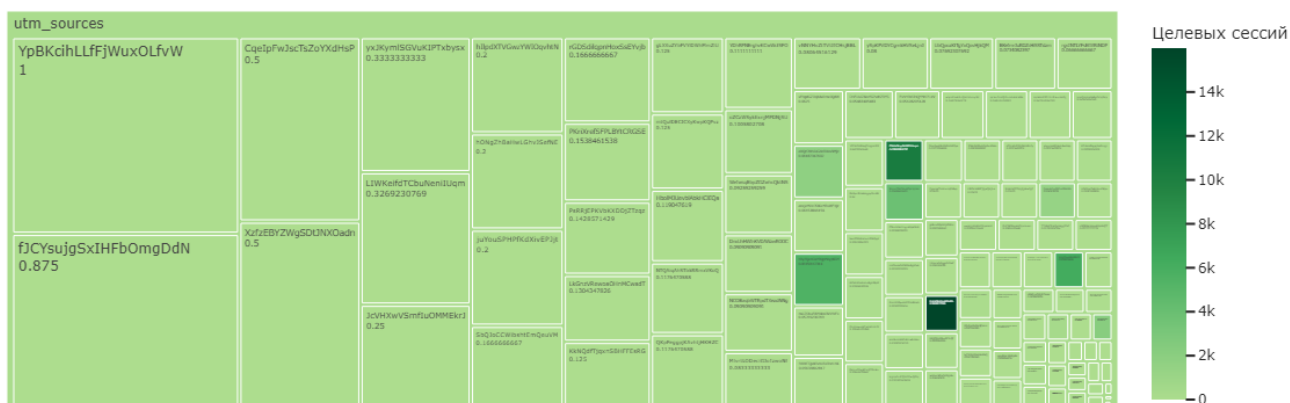
### Топ источников трафика по месяцам:

- ZpYloDJMcFzVoPFsHGJL - июль, октябрь - декабрь
- fDLIAcSmythWSCVMvqvL - май, июнь
- kjsLglQLzykiRbcDiGcD - август, сентябрь

### Топ источников целевого трафика по месяцам:

- ZpYloDJMcFzVoPFsHGJL - июль - декабрь
- fDLIAcSmythWSCVMvqvL - май, июнь

Распределение конверсии по источникам за все время.  
Размер - конверсия. Цвет - количество целевых сессий.



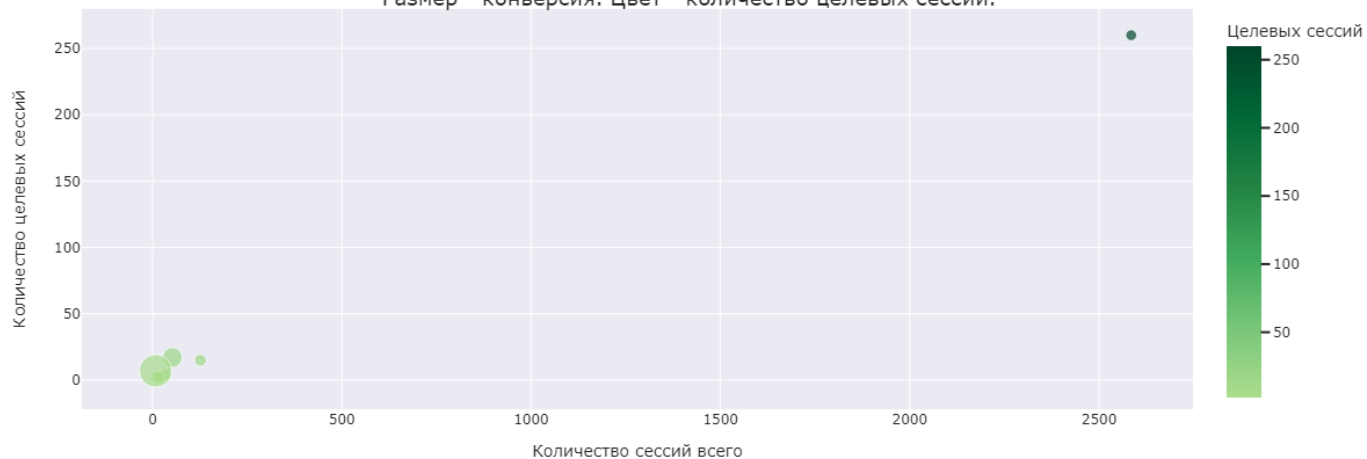
### Топ источников по конверсии (более 0,3):

1. YpBKcihLLfJWuxOLfvW - 1
2. fJCYSujgSxIHfBOmgDdN - 0,88
3. XzfzEBYZWgSDtJNXOadn - 0,5
4. CqelpFwJscTsZoYXdHsP - 0,5
5. yxJKymlSGVuKIPTxbyxs - 0,33
6. LIWKeifdTcBuNenilUqm - 0,33



Выборка источников, благодаря которым совершено более 1 целевого действия  
и с конверсией более 10%.

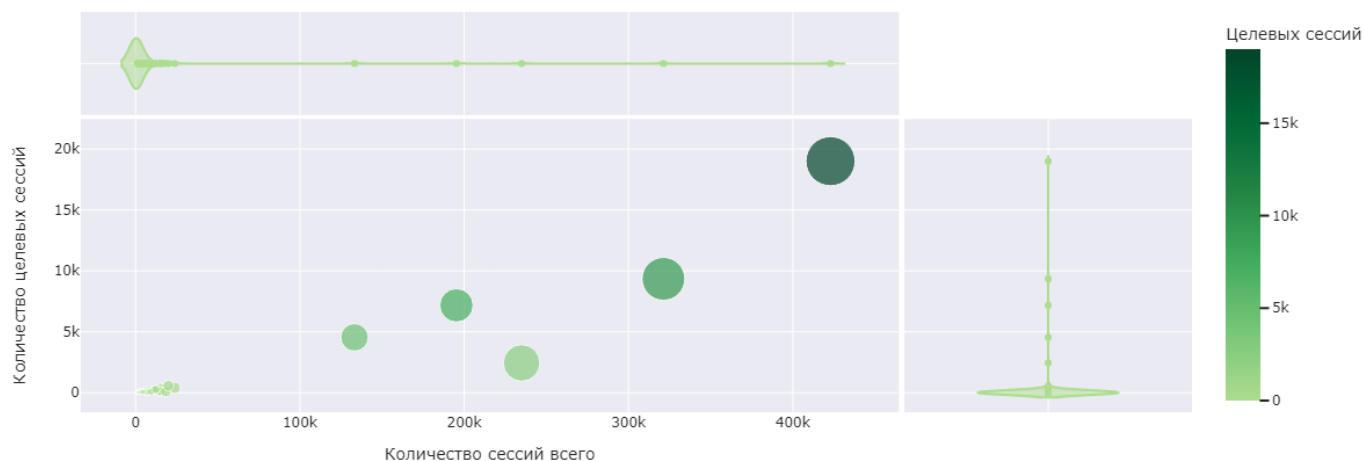
Размер - конверсия. Цвет - количество целевых сессий.



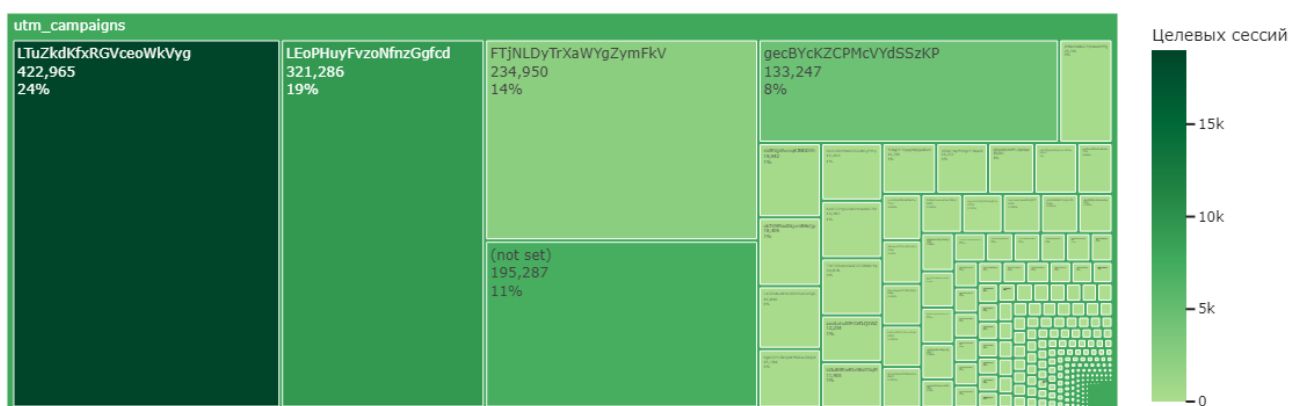
utm_source	target_sessions	all_sessions	cr
fJCYsujgSxIHfBOmgDdN	7	8	0.875000
LIWKeifdTCbuNenilUqm	17	52	0.326923
PKriXrefSFPLBYtCRGSE	2	13	0.153846
PsRRjEPKVbKXDDjZTzqz	2	14	0.142857
LkGnzVRewoaOHnMCwadT	3	23	0.130435
HbolMJUevblAbkHCIEQa	15	126	0.119048
NTQAiqAhSTbkRRmxVKoQ	4	34	0.117647
QKpFeggpjKAvhtjHKHZC	2	17	0.117647
oZCzWSykfixnjMPDNjSU	260	2585	0.100580

# Из каких кампаний идёт самый целевой трафик?

Распределение количественных показателей кампаний.  
Размер - количество сессий. Цвет - количество целевых сессий.



Распределение трафика по кампаниям за все время.  
Размер - количество сессий всего. Цвет - количество целевых сессий.



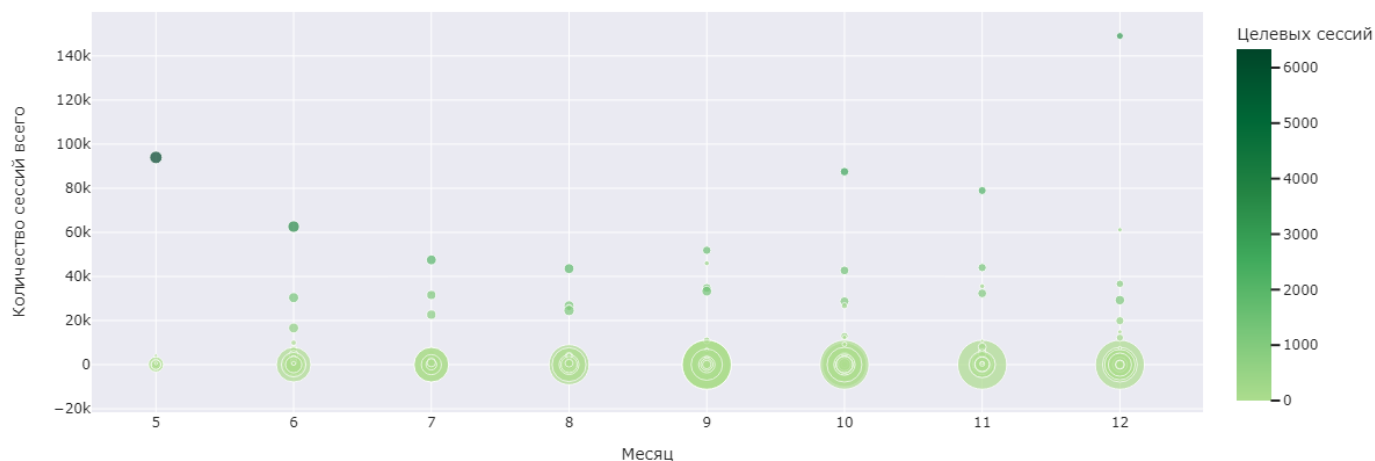
## Топ источников трафика (более 5%):

1. LTuZkdKfxRGVceoWkVyg - 24%
2. LEoPHuyFvzoNfnzGgfc - 19%
3. FTjNLDyTrXaWYgZymFkV - 14%
4. gecBYcKZCPMcVYdSSzKP - 8%

## Топ источников целевого трафика (не менее 5%):

1. LTuZkdKfxRGVceoWkVyg - 38%
2. LEoPHuyFvzoNfnzGgfc - 19%
3. gecBYcKZCPMcVYdSSzKP - 9%
4. FTjNLDyTrXaWYgZymFkV - 5%

Распределение трафика по месяцам.  
Размер - конверсия. Цвет - количество целевых сессий.



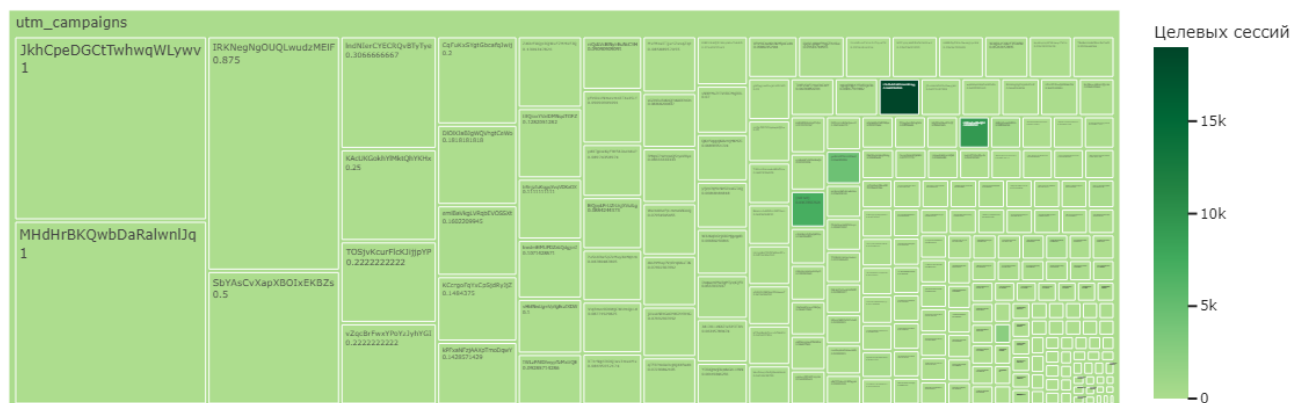
#### Топ источников трафика по месяцам:

- LTuZkdKfxRGVceoWkVyg - май - сентябрь
- LEOPHuyFvzoNfnzGgfc - ноябрь, декабрь
- gecBYcKZCPMcVYdSSzKP - октябрь

#### Топ источников целевого трафика по месяцам:

- LTuZkdKfxRGVceoWkVyg - май - сентябрь
- LEOPHuyFvzoNfnzGgfc - ноябрь, декабрь
- gecBYcKZCPMcVYdSSzKP - октябрь

Распределение конверсии по кампаниям за все время.  
Размер - конверсия. Цвет - количество целевых сессий.

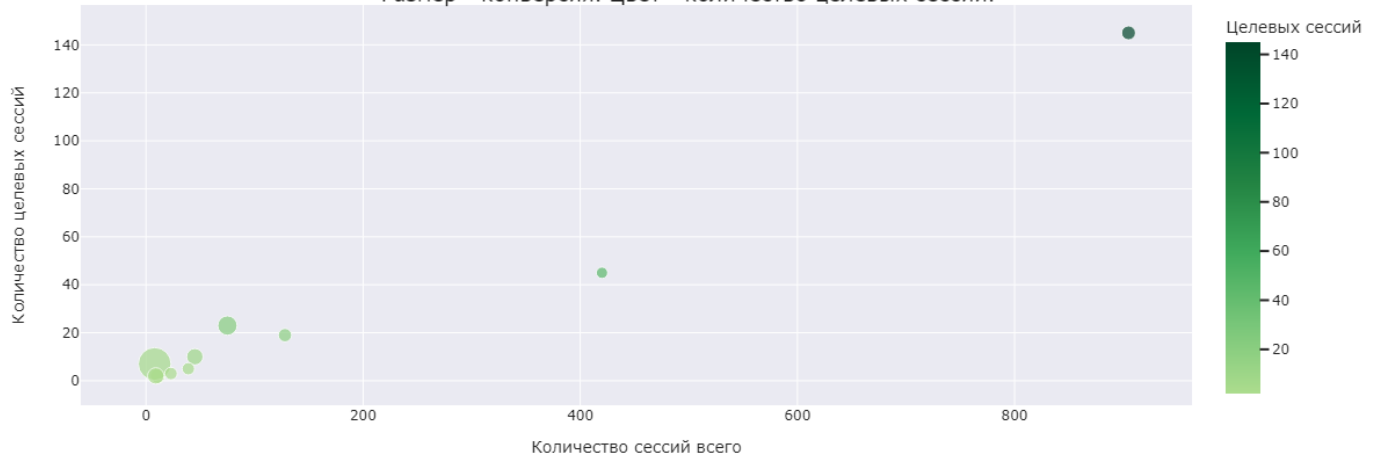


#### Топ кампаний по конверсии (более 0,3):

1. JkhCpeDGCTwhwqWLywv - 1
2. MHdHrBKQwbDaRalwnIJq - 1
3. IRKNegNgOUQLwudzMEIF - 0,88
4. SbYAsCvXapXBOIxEKBZs - 0,5
5. IndNlerCYECRQvBTyTye - 0,31

Выборка кампаний, благодаря которым совершено более 1 целевого действия  
и с конверсией более 10%.

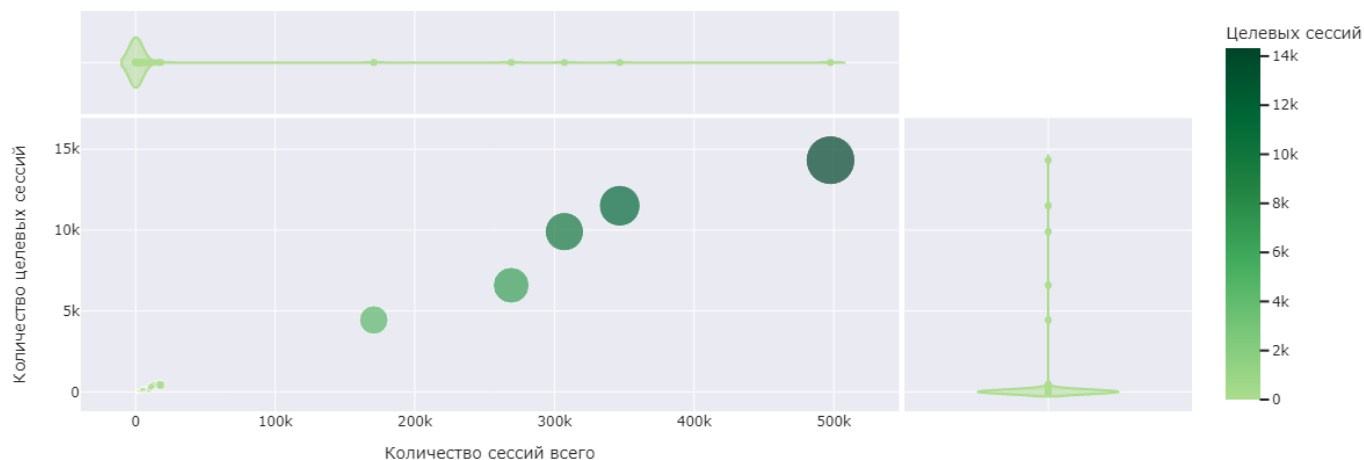
Размер - конверсия. Цвет - количество целевых сессий.



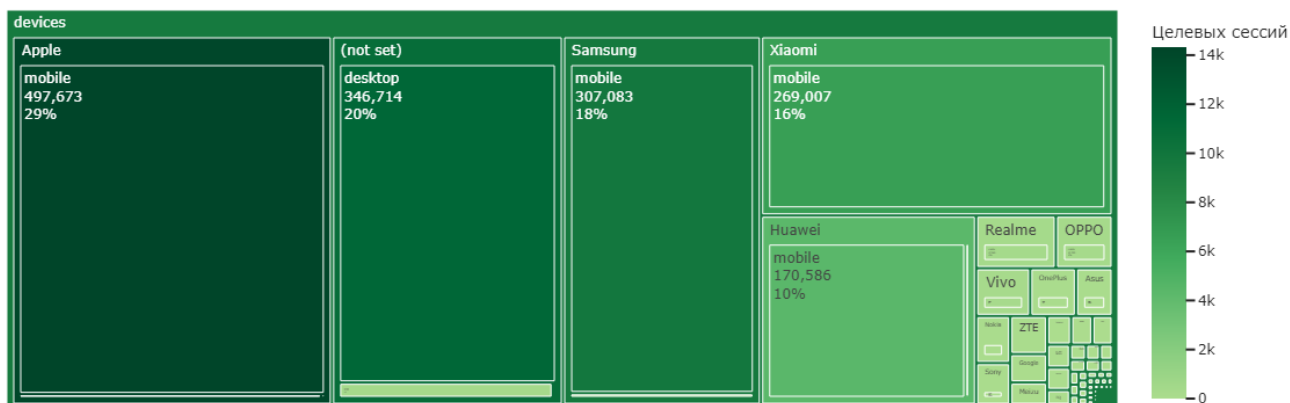
utm_campaign	target_sessions	all_sessions	cr
IRKNegNgOUQLwudzMEIF	7	8	0.875000
IndNlerCYECRQvBTyTye	23	75	0.306667
TOSjvKcurFlcKJijpYP	10	45	0.222222
vZqcBrFwxYPoYzJyhYGI	2	9	0.222222
DIOIXJaBJgWQVhgtCeWo	2	11	0.181818
emlBaVkgLVRqbEVOSSXt	145	905	0.160221
KCcrgoFqYxCpSjdRyJjZ	19	128	0.148438
ZdOcFAQpRQWuFZHHkEJQ	3	23	0.130435
tlfQioxYValDMNqdTGFZ	5	39	0.128205
kwdmEIMUPDZaLQdgjcsI	45	420	0.107143

# Из каких устройств идёт самый целевой трафик?

Распределение количественных показателей трафика с устройств.  
Размер - количество сессий. Цвет - количество целевых сессий.



Распределение трафика по устройствам за все время.  
Размер - количество сессий всего. Цвет - количество целевых сессий.



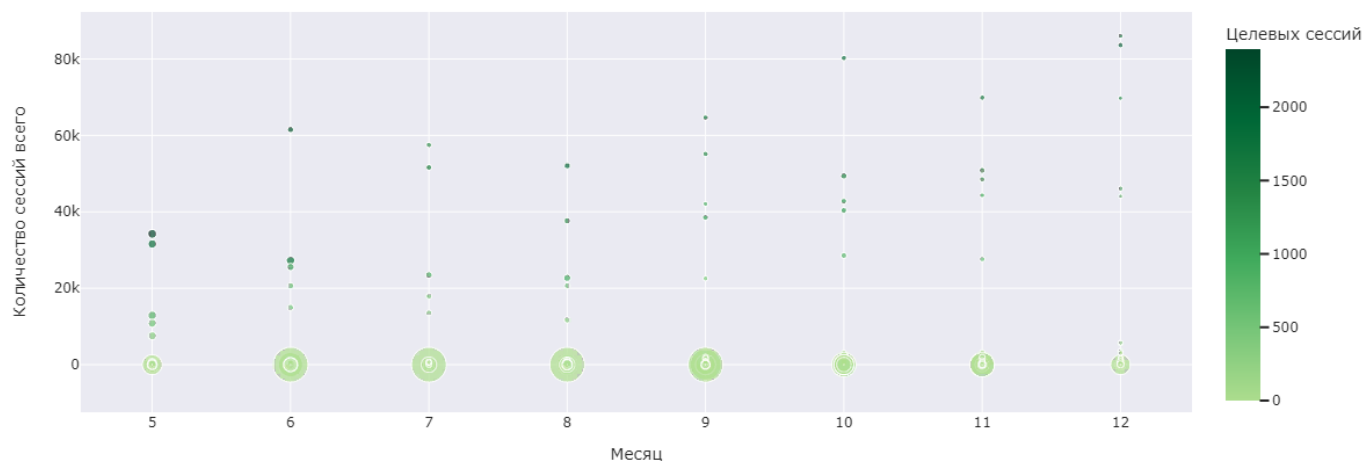
Топ источников трафика (более 5%):

1. Apple - 29%
2. desktop - 20%
3. Samsung - 18%
4. Xiaomi - 16%
5. Huawei - 10%

Топ источников целевого трафика (более 5%):

1. Apple - 28%
2. desktop - 23%
3. Samsung - 20%
4. Xiaomi - 13%
5. Huawei - 9%

Распределение трафика по месяцам.  
Размер - конверсия. Цвет - количество целевых сессий.



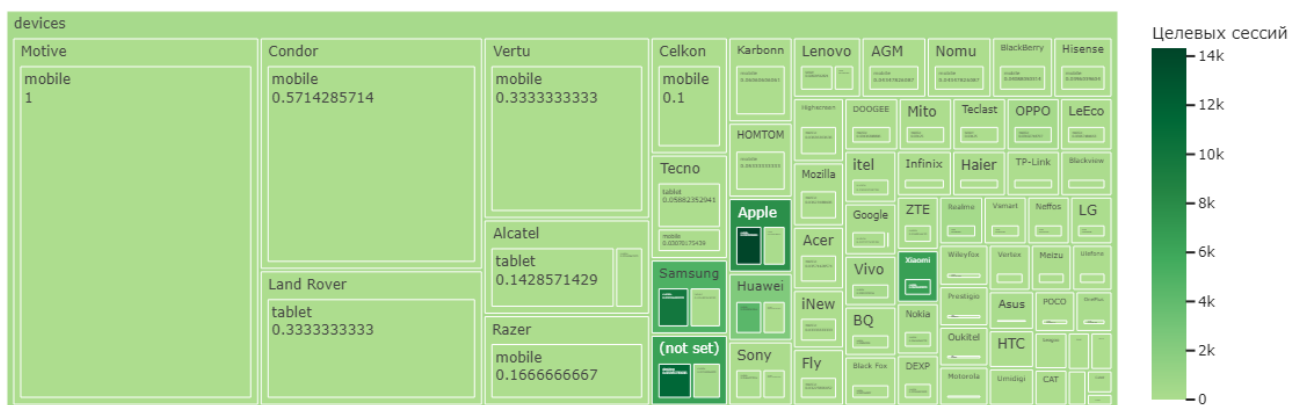
#### Топ источников трафика по месяцам:

- Apple - июнь, август - декабрь
- desktop - май, июль

#### Топ источников целевого трафика по месяцам:

- Apple - июнь - ноябрь
- Samsung - декабрь
- desktop - май

Распределение конверсии по устройствам за все время.  
Размер - конверсия. Цвет - количество целевых сессий.



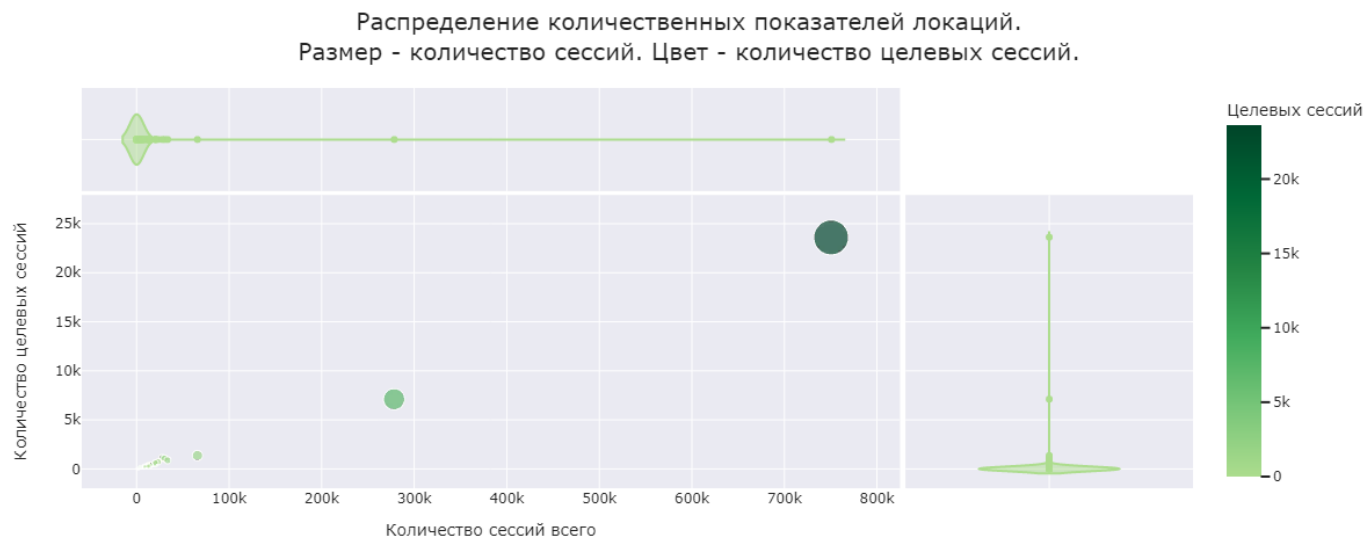
#### Топ устройств по конверсии (более 0,3):

1. mobile Motive - 1
2. mobile Condor - 0,57
3. tablet Land Rover - 0,33
4. mobile Vertu - 0,33

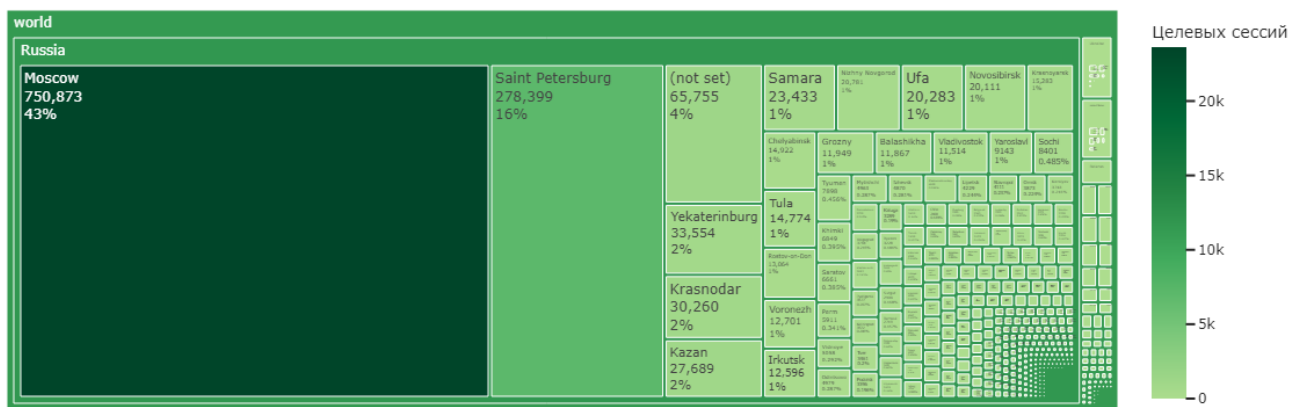


device_category	device_brand	target_sessions	all_sessions	cr
mobile	Condor	4	7	0.571429

# Из каких локаций идёт самый целевой трафик?



Распределение трафика по локациям за все время.  
Размер - количество сессий всего. Цвет - количество целевых сессий.



## Топ источников трафика (более 1%):

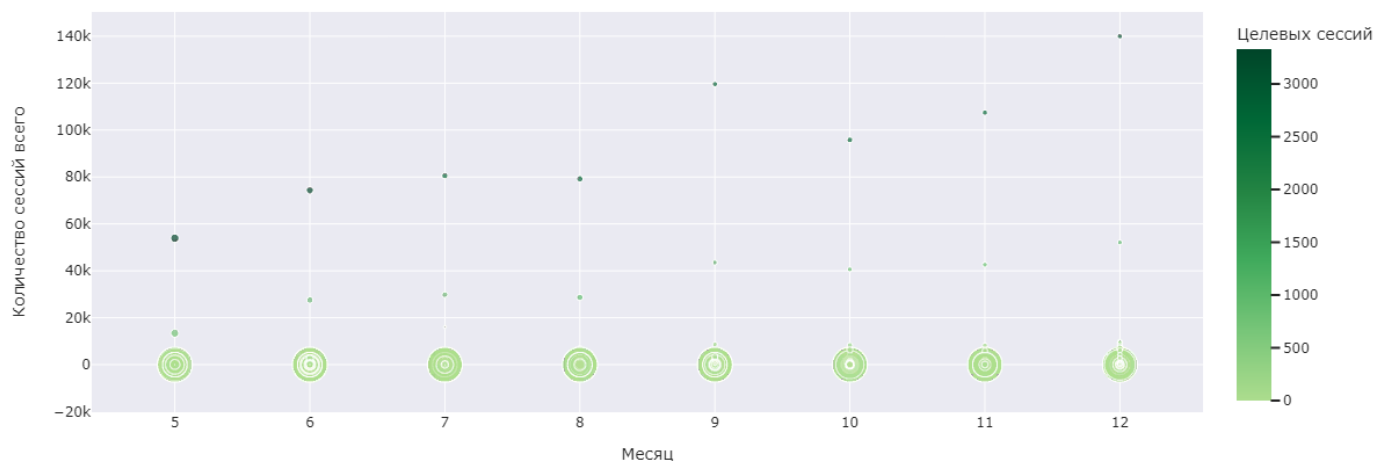
1. **Moscow** - 43%
2. **Saint Petersburg** - 16%
3. **Yekaterinburg** - 2%
4. **Krasnodar** - 2%
5. **Kazan** - 2%

## Топ источников целевого трафика (более 1%):

1. **Moscow** - 47%
2. **Saint Petersburg** - 14%
3. **Kazan** - 2%
4. **Krasnodar** - 2%
5. **Yekaterinburg** - 2%



Распределение трафика по месяцам.  
Размер - конверсия. Цвет - количество целевых сессий.



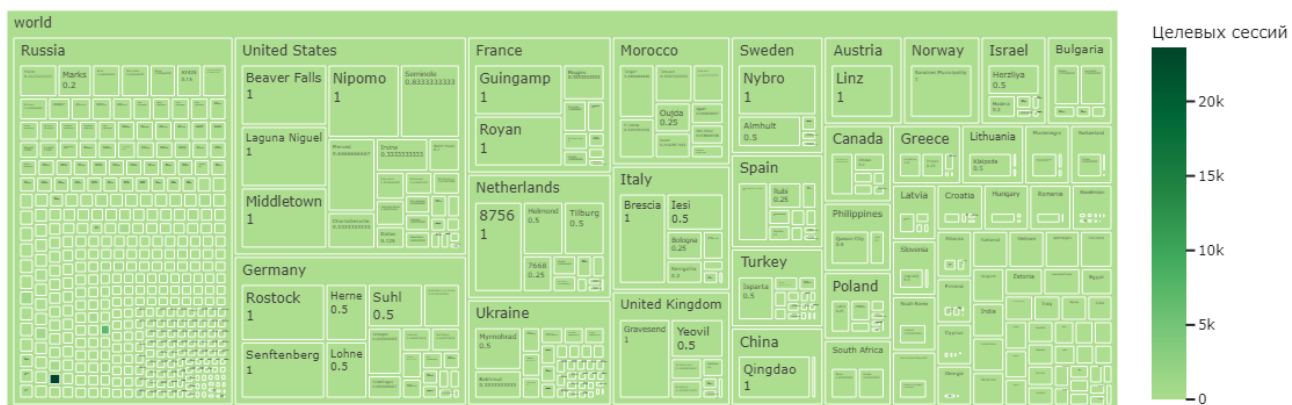
### Топ источников трафика по месяцам:

- Москва -- май - декабрь

### Топ источников целевого трафика по месяцам:

- Москва -- май - декабрь

Распределение конверсии по локациям за все время.  
Размер - конверсия. Цвет - количество целевых сессий.



Топ локаций по конверсии не выглядит показательным, т.к. локации с конверсией  $> 0.3$  (топ-50) дают в среднем не более 2 целевых сессий

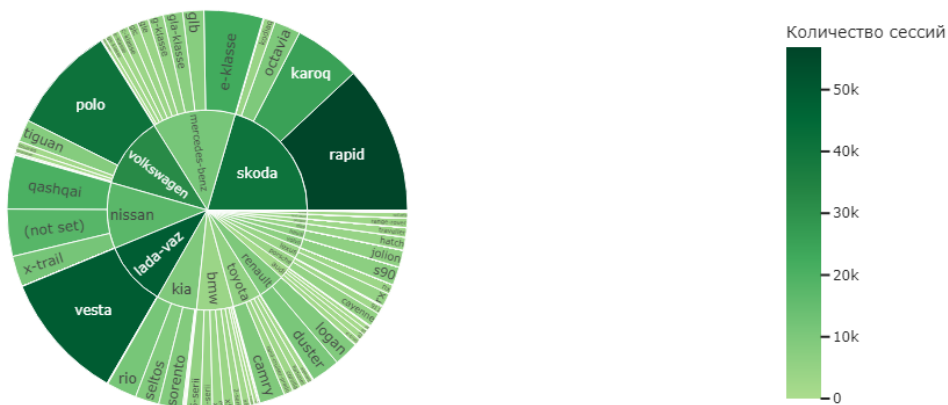


**Топ локаций по конверсии (с 2 и более целевыми сессиями):**

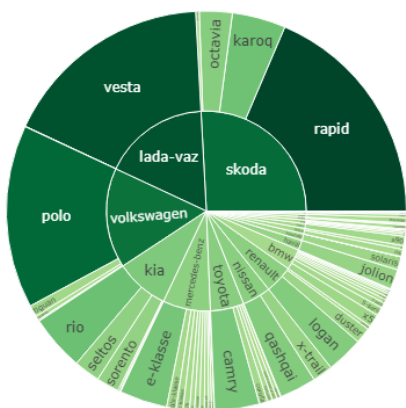
1. [Seminole](#) (United States) - 0,83
2. [Merced](#) (United States) - 0,67
3. [Helmond](#) (Netherlands) - 0,5
4. [Tangier](#) (Morocco) - 0,44
5. [\(Slovenia\)](#) - 0,3

## Какие авто пользуются наибольшим спросом?

Распределение марок и моделей авто  
(по количеству сессий)



Распределение марок и моделей авто  
(по количеству целевых сессий)

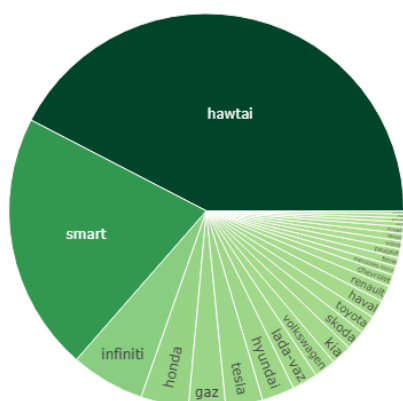


### Топ автомобилей по количеству целевых сессий:

1. skoda rapid - 19%
2. lada-vaz vesta - 17%
3. volkswagen polo - 15%
4. kia rio - 5%
5. skoda karoq - 4%
6. mercedes-benz e-klasse - 4%

# Какие авто пользуются наибольшим спросом?

Распределение брендов по CR



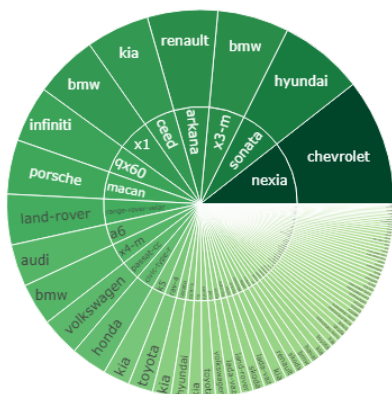
Топ брендов по конверсии с количеством целевых действий менее 10:

1. **hawaitai** - 1
2. **smart** - 0,5
3. **infiniti** - 0,14
4. **honda** - 0,09
5. **tesla** - 0,07
6. **gaz** - 0,07

Топ брендов с относительно высокой конверсией, но со значительно большим количеством целевых действий:

1. **hyundai** - 0,06
2. **lada-vaz** - 0,05
3. **volkswagen** - 0,04
4. **kia** - 0,04
5. **skoda** - 0,04
6. **toyota** - 0,04
7. **haval** - 0,04

Распределение марок по CR



Топ моделей по конверсии:

1. **chevrolet nexia** - 0,5
2. **hyundai sonata** - 0,33
3. **bmw x3-m** - 0,28
4. **renault arkana** - 0,28
5. **bmw x1** - 0,25
6. **kia ceed** - 0,25

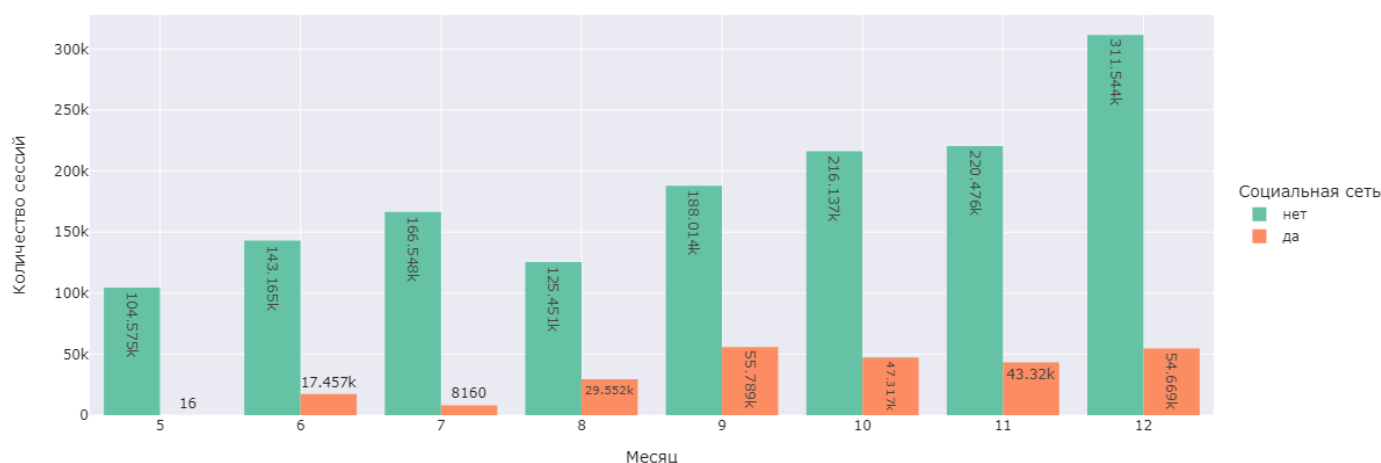
## Стоит ли увеличивать своё присутствие в соцсетях и давать там больше рекламы?

	target_sessions	all_sessions	cr
нет	46288	1475910	0.031362
да	4021	256280	0.015690

В среднем за все время **конверсия трафика из соцсетей ниже на 1,57%**, чем из остальных источников.

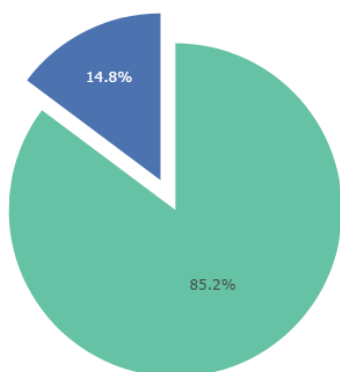
На первый взгляд, не стоит убирать рекламу из соцсетей, а попробовать оптимизировать ее, проанализировать эффективность каждого из источников.

Распределение трафика по месяцам (всего сессий)



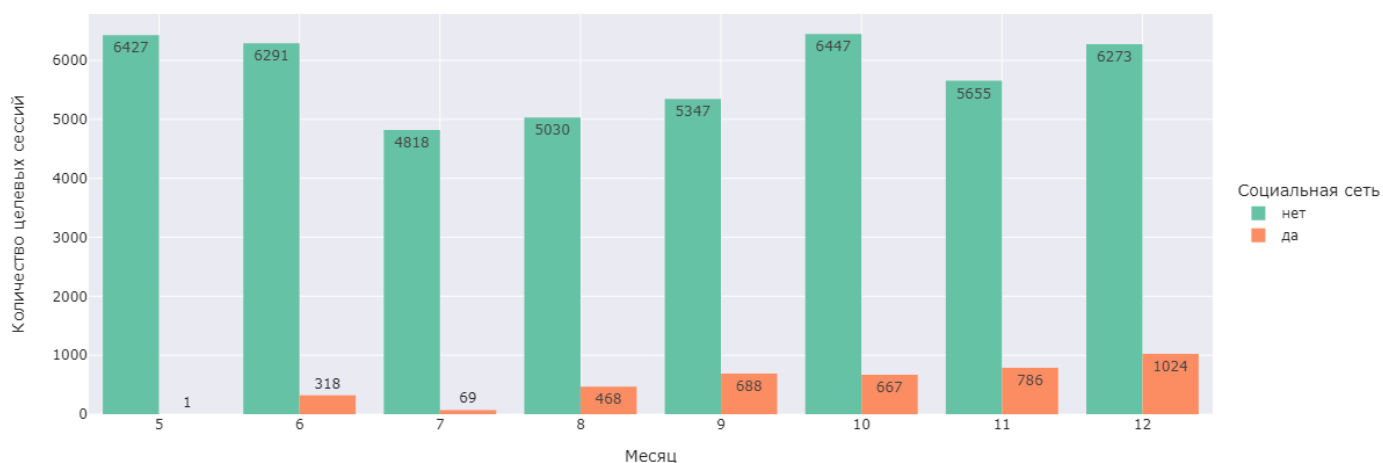
На диаграмме видно, что часть трафика из соцсетей (относительно всего трафика) выше в период с **августа по декабрь**

Распределение трафика за весь период (всего сессий)

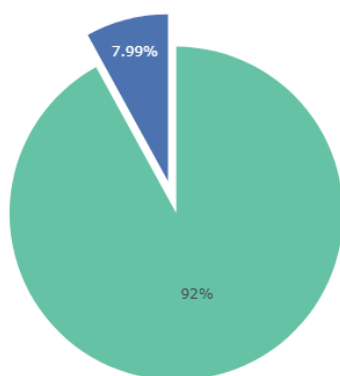


Трафик из соцсетей занимает хоть и небольшую часть всего трафика, но весомую - **около 15 %**

### Распределение целевого трафика по месяцам

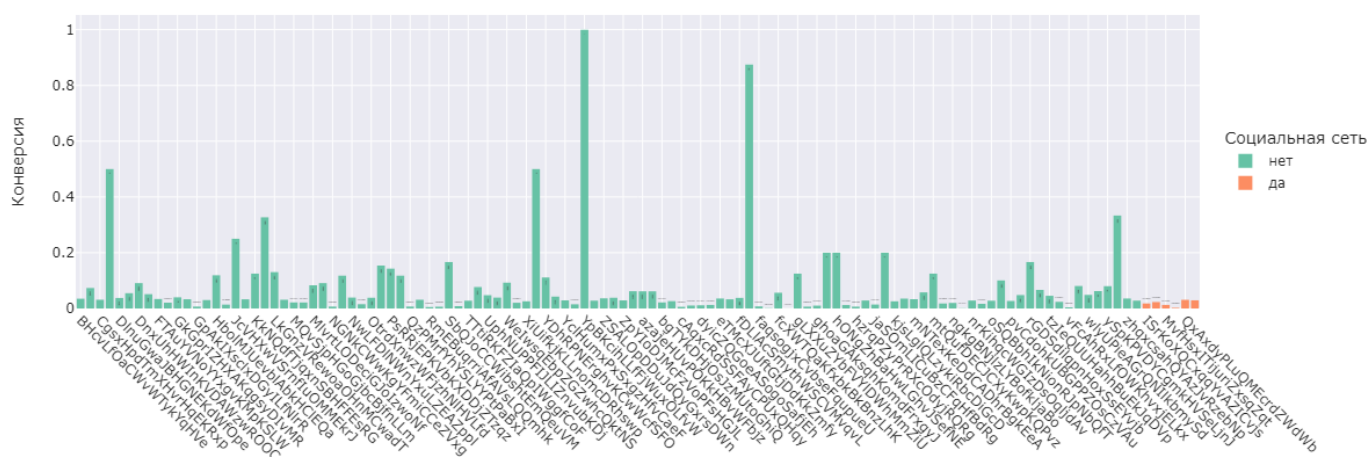


Распределение целевого трафика за весь период



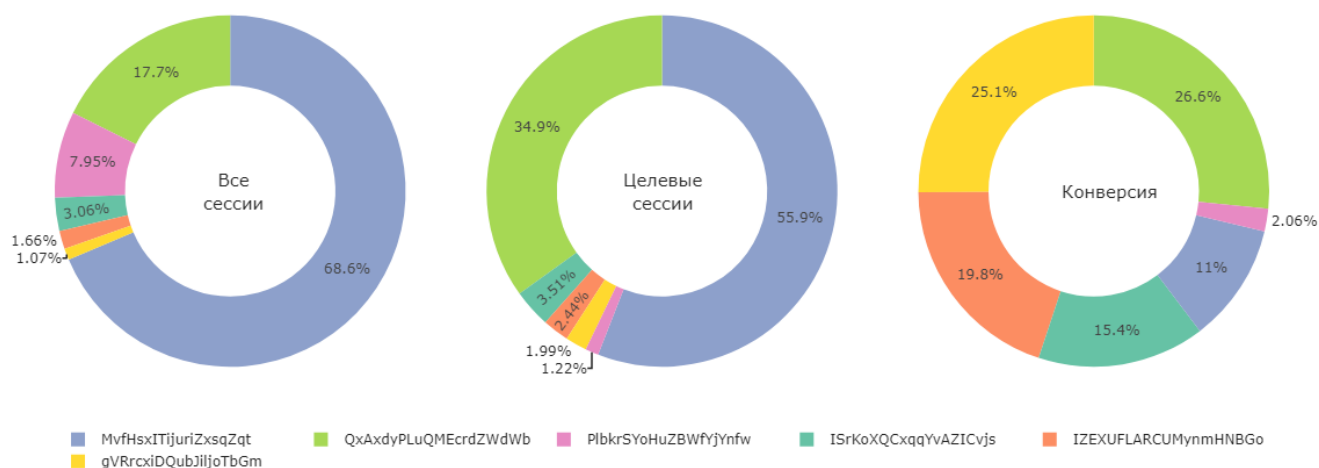
Доля целевого трафика из социальных сетей меньше - всего 8% от всего трафика. Но пока недостаточно аргументов, чтобы им пренебрегать.

### Распределение конверсии по источникам



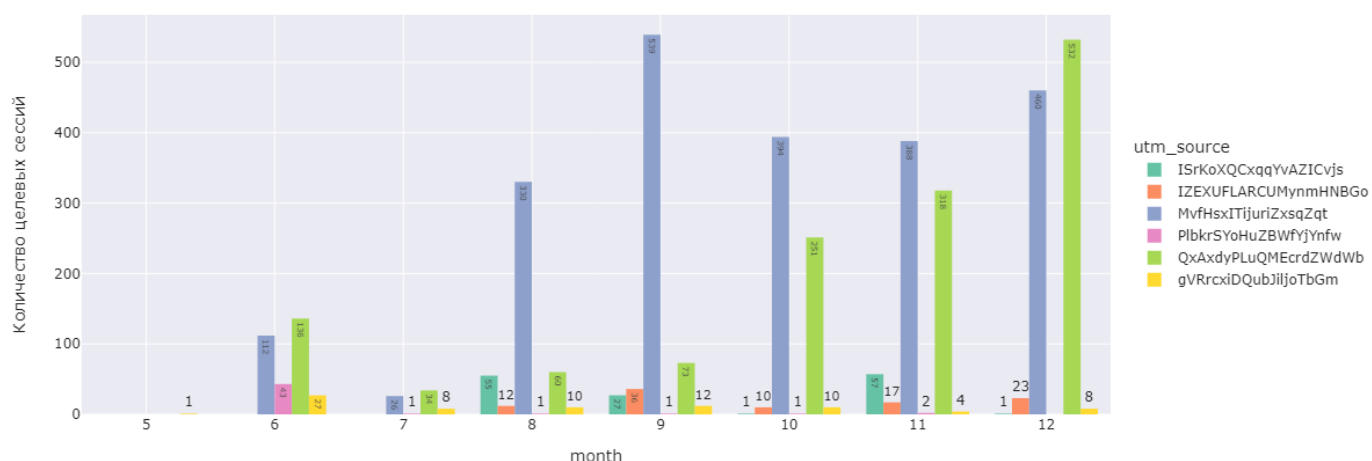
С точки зрения конверсии, трафик из соцсетей не выделяется, но все-таки заметен на диаграмме.

### Показатели источников, относящихся к социальным сетям



По круговым диаграммам можно сделать вывод, что лучшие показатели дают источники **MvfHsxITijuriZxsqZqt** и **QxAxdyPLuQMEcrdZWdWb**

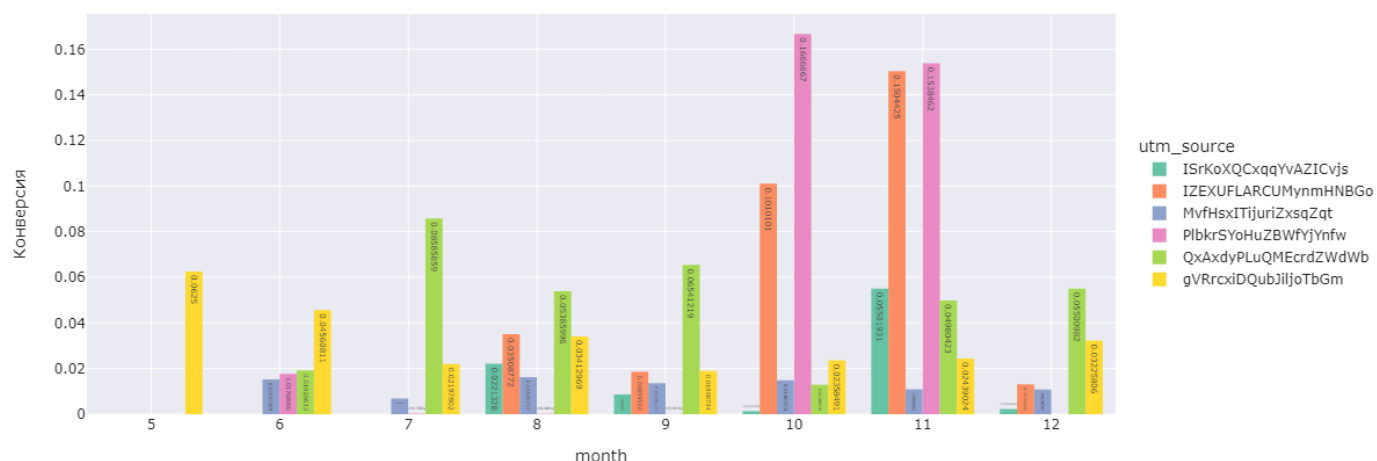
### Распределение целевого трафика по источникам из соцсетей (по месяцам)



Для источника **MvfHsxITijuriZxsqZqt** лучшим по показателям целевых сессий был период с **августа по декабрь 2021** года.

Для источника **QxAxdyPLuQMEcrdZWdWb** - период с **октября по декабрь 2021** года.

Конверсия источников из соцсетей (по месяцам)



Для источника **MvfHsxITijuriZxsqZqt** лучшими по показателям конверсии был периоды **с июля по сентябрь** и **с ноября по декабрь 2021** года

Источник **QxAxdyPLuQMEcrdZWdWb** держался на уровне 10-15% конверсии **в июне** и в период **с августа по декабрь 2021** года

#### Вывод:

Будет целесообразно оставить два источника трафика из соцсетей (или давать там больше рекламы)

- **MvfHsxITijuriZxsqZqt** - в августе, сентябре, ноябре и декабре
- **QxAxdyPLuQMEcrdZWdWb** - с октября по декабрь