gemiusAdReal Study

Methodology description

Updated: April 2024

Contents

INTRODUCTION

gemiusAdReal is both digital and traditional advertising research study that allows for monitoring and analysing of advertising activity and digital strategies of **all advertisers** on the market, based on their **display, text, video and audio ads** delivered to the users through **various media channels** (horizontal portals, social sites, search, video players, mobile applications, tv and radio stations).

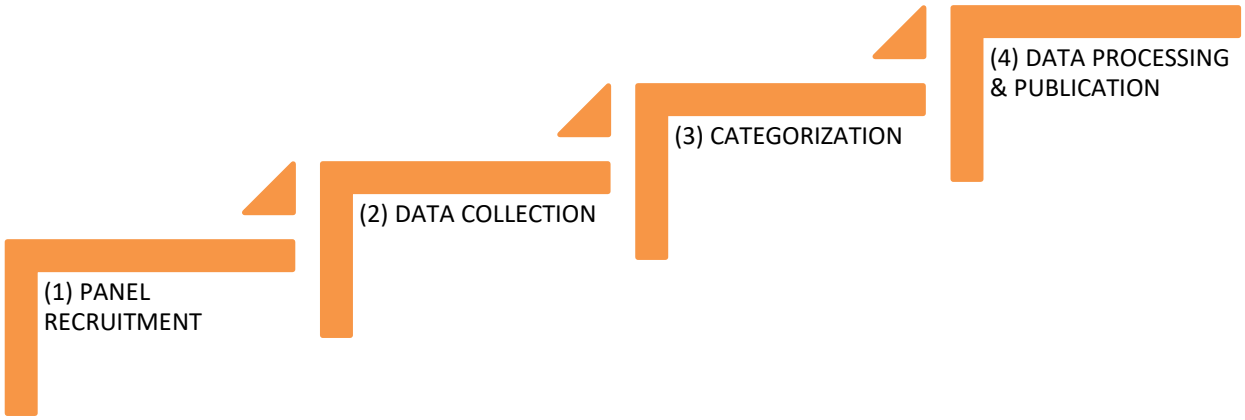
It is **a passive research based on a panel**, i.e. a representative group of the Internet users who have installed a **research software** inside their browsers. In the case of mobile panel (used also for TV and Radio study), software is pre-installed on devices provided by Gemius to the study participants. The software measures online behaviour of the panel members, focusing on their exposure to online advertising. Data collected from the observation of the participants of the study is categorised and processed in order to extrapolate an overall picture of the online and offline advertising market.

Currently, gemiusAdReal data is available for ads served on PC platform in France, Germany, Poland, Romania, Russia, Turkey and Ukraine,mobile in Poland, Germany and Turkey, TV in Poland and Germany and Radio in Poland.

The document hereby provides an overview of the gemiusAdReal methodology, as well as the information on the approach applied to measure viewability.

GEMIUSADREAL- STUDY REALIZATION PROCESS

gemiusAdReal data production process includes the following four main stages:



(1) PANEL RECRUITMENT

A user-centric approach, applied in the gemiusAdReal study, requires **recruiting an online panel** of the Internet users who agree to install **the research software on their devices**. The research software is a **browser extension** (available for the most popular web browsers) that passively collects data about the ads that are being displayed to the panellists while they use the Internet.

During recruiting a mobile panel, people who have expressed an interest in participating in the study receive from Gemius phones with a set of meters pre-installed at the operating system level.

At the recruitment process the panel members are asked basic **socio-demographic questions** that are used to ensure the representativeness of the panel structure with reference to the population of

the Internet users in a given country, as well as to analyse the study's results by selected socio-demographic characteristics.

The **panel maintenance** incorporates two essential activities: **supplementary recruitment** to replace the resigning panel members and regular **extension updates**. Both the research software and the approach used in panel recruitment are continuously being developed (e.g. inclusion of additional information on measured ads, new mobile applications in the study, new media types) and adjusted to the changing market reality (such as new ad formats or technologies used by publishers).

The size of the panel, as well as the recruitment method might differ by country.

In Poland, the study uses a panel of 12 000 panelists on PC devices and for the other media included in the study, i.e. mobile, tv and radio, is based on a sample of 2 500 panelists.

(2) DATA COLLECTION (RECOGNITION OF ADS)

gemiusAdReal research software installed in panellist browser **recognises the ads displayed on the screen** and verifies whether they were in-view for the user. At the moment gemiusAdReal software supports all of the major **internet browsers on PC** (i.e. Chrome, Firefox, Opera, as well as Yandex.Browser). Additionally the study is conducted on smartphones including advertising traffic from **mobile browsers** (i.e. Chrome, Chrome beta, Firefox, Samsung Browser, Yandex, Yandex Lite, Opera, Opera Touch, Dolphin, Edge, CM Browser) and **in mobile applications Facebook, YouTube, Instagram and TikTok**. Additionally, traffic from the Facebook In-App Browser is measured - it is a browser within the Facebook application that opens links to the external websites. These devices are also used to measure television and radio advertisements (in this case the meters installed on the smartphone's operating system level recognize the sounds of advertisements heard by panelists and save them as text files). Thanks to the use of a single-source panel in the gemiusAdReal study, it is possible to analyze advertising activity on smartphones (mobile browsers, Facebook, YouTube, Instagram and TikTok applications) and - thanks to the use of sound-matching technology - on TV and Radio.

The software recognises and collects data on all **types of ads** displayed to the Internet users, i.e.:

- **video ads (applies to PC, Mobile and TV),** including Video In-Display, Out-Stream, In-Stream, Social Video and TV AD, that were served to the panel members. In gemiusAdReal study each animation (including GIF), served through Facebook ecosystem, is counted as video ad. This is due to the lack of technical possibilities to distinguish between looped animation and video.
- **display and text ads (applies to PC, Mobile),** including non-video ads served to the Internet users in a form of display ads (image, animation) or text ads (e.g. sponsored links or context advertising). Although native ads are difficult to identify due to diverse market standards, data regarding standardised formats (such as: sponsored content on Facebook, special sections, properly labelled sponsored articles etc.) is also collected.
- **radio ads (applies to Radio),** including emissions of audio ads broadcasts during advertising breaks on radio stations included in the study.

The data is gathered from a **variety of sources**: traditional portals, social media sites (like Facebook or Twitter), advertising websites, e-commerce sites, auctions, pages with entertainment content (e.g.

YouTube), selected applications (Facebook, Instagram, YouTube, TikTok), selected TV stations (including TVP1, Polsat, TVN) and selected radio stations (including RMF FM, Radio ZET). The main criterion for including given a media channel to research is whether it has an advertising offer. If the page does not have any kind of paid advertising it is excluded from the study results. The list of these different sites varies from country to country, but usually includes websites and telecommunications companies, and some e-commerce sites that are limited by self-promotion campaigns.

Data collection in gemiusAdReal study includes the following two stages:

1. **identifying ad icontacts** on the media channels viewed by the panellist. Every second the extension checks if any new ads were served to a given panellist on the visited browsed web page, used application, watched TV channel or listened radio station.

    - **In the case of PC and Mobile research:** Ads are recognized thanks to predefined rules, partially adapted to individual markets.
    - **In the case of TV research:** smartphones distributed by Gemius to study participants recognize the sounds of advertisments heard by panellists and saves them as a text files. The reference base is the daily satellite signal provided by external provider, which contains tv ads and content broadcast emitted on each of 48 TV stations included in the study on a given day.

      List of tv stations is as follows (is a subject to change in the future):

      TVN, TVP 2, Polsat, TVP 1, TVN24, TV4, TV Puls, TVN 7, TTV, Stopklatka, TVP Seriale, TV6, Puls 2, Polsat News, Polsat 2, TVP INFO, Zoom TV, Fokus TV, Super Polsat, TVN Turbo, Nowa TV, Polsat Play, TVP 3, TVP HD, FX, WP TV, Polsat Film, Polsat Cafe, TVP Rozrywka, HGTV, TVP Historia, Metro TV, TVN Fabula, Kino Polska, National Geographic, Polsat Comedy Central Extra, TVN24 Biznes i Swiat, Polo TV, Polsat Sport, TVP ABC, TVP Kultura, EskaTV, Discovery, TVN Style, TVP Sport, 4FUN.TV, Eurosport 1, TV Trwam.

      The confidence level (accuracy of time participation) for the largest six TV stations is 99,92%, while for the other channels it is 99%.

    - **In the case of Radio research:** ads are recognized in a similar way TV study. The reference base is a radio reference signal provided by external providers, containing ads and radio auditions in audio format, broadcasted on a given day for the 6 radio stations included in the study.

      List of radio stations is as follows (is a subject to change in the future):

      Radio ZET, RMF FM, Trójka, Radio Złote Przeboje, Radio ESKA, Jedynka.

2. **downloading** page elements recognized as advertising creations and **sending** them to the Gemius server together with information about where and how long (if at all) the given ad was in the user's field of view. This stage applies only to PC and Mobile platforms. In the case of TV and Radio platforms, Gemius receives a daily program schedule of radio and tv programs in the form of video and audio files from the signal provider. By using sound-matching technology, the data received from the panel is combined with the reference base

of satellite and radio signals. In that way the files are created that contains information with which ads panelists have had contact with. The efficiency of audio recognition remains at ~98%. At the same time, it is worth to underline that throughout the whole process the signal of television and radio stations is additionally analyzed to identify advertising breaks. Since the online advertising ecosystem is constantly changing (new or custom ad formats appear or existing ways of presenting ads are modified), the ways of the ad-recognizing extension are verified and updated on an ongoing basis. In order to provide the highest quality data and minimize cases in which the extension does not register the ad contacts, daily, weekly and quarterly **quality tests** are carried out, including **manual and automatic checking of compliance of the most popular pages with the current version of the gemiusAdReal study software** (the number of checked pages increases with the length of the period covered testing, e.g., weekly tests cover more pages than daily tests). These tests ensure efficiency and ongoing quality verification at both stages of data collection: recognition of ads and providing information collected about them.

(3) CATEGORISATION OF ADS

During the process of categorization, individual advertising creatives (or groups of creatives having the same landing page) recognized and downloaded via research software, are given a detailed description (assigned automatically or manually) specifying the industry and the advertiser. When it comes to television and radio commercials, materials from advertising breaks go to the standard categorization process, in which they are assigned the right brands and industries. Due to the process automation, short advertising materials (e.g. sponsorship billboards) emitted outside of advertising breaks are not included in the study results. Further data processing is carried out in the same way as for online advertising (PC, Mobile).

- **Industry**
  - Sub-industry (in some cases)
- **Brand owner**
  - Brand
    - Product (in some cases).

This step of the process is crucial to the whole process, as – in order to deliver data of the best possible quality – the significant part of this procedure is **handled manually** by a highly qualified categorisation team, especially in case of ambiguous materials. This mainly applies to television and radio advertisements, which due to their nature are fully categorized manually. In addition video in-stream and out-stream ads are also categorised manually which means that these creatives needs to be watched by a member of the categorisation team. In case of non-video ads, many creatives undergo the same procedure as video (e.g. rich media). Remaining creatives are grouped and landing pages of these groups of creatives are categorised by a member of categorisation team. Since the number of ad creatives collected by the extension can amount monthly to several millions on a given market, this stage of the process is the most vulnerable to human error or negative impact of categorization team overload. Such situation manifests itself in either incorrect categorisation or in lack of categorization. In most cases lack of categorisation refers to a long tail campaigns with small reach, but occasionally can also refer to bigger campaigns with variety of landing pages assigned to a given ad creative. In order to minimise a risk of error or omission, a **set of procedures and precautions** was implemented, including:

- extended training program for categorisation team run by the Supervisor of the local market according to gemiusAdReal global procedures, with a set of training cases and selective re-categorisations;
- development of both, manual and automatic control systems that verify occurrences of common inaccuracies;
- additional feature in the clients' interface, enabling to report mistakes which are then immediately solved by the Supervisor.

(4) DATA PROCESSING AND PUBLICATION

In this stage of the process each panellist is being assigned an **individual weight** in a daily, weekly and monthly resolution, which allows to estimate and calculate respective indicators for the researched market, based on the data collected from the panel. Simultaneously, the domains visited by the panellists are being automatically matched with a pre-defined **media tree** in order to be presented in the following structure:

Media owner
Media channel
Section

Since February 1, 2019, the method of data processing has changed - from that moment a **virtual version of the panel** is created based on the current panel. The effect of this process is to assign **constant and equal weights** to all panelists and to eliminate the problem of rotation of study participants.

Constant Panel is a modelled, virtual panel, consisting of a very large number of panelists (having the same weight). The number of panelists or their weights do not change over time, hence the name "constant". Instead of changing the weight of a given panelist, we balance it by creating a sufficiently large group of constant panelists whose total weight will be the same as that of the panelist.

This reflects the situation from the actual population. Constant Panel can be defined as a replacement model, to which the activity of real panelists is assigned each time.

Thanks to the above approach, it is possible to calculate results for any custom periods (up to 12 months).

The implementation of the Constant Panel methodology also brings additional research improvements, such as the ability to analyze reach building, duplication between different media types and advertising activity in different types of media, both digital and traditional. Thanks to the use of Single Source technology, it is possible to analyze the results of campaigns conducted simultaneously on a PC, smartphones (in selected browsers and mobile applications), television and radio (selected TV and radio stations).

Since January 1, 2021, we updated the population and structural data. As a result, the results for radio and television have taken into account the entire population of users of these media, and not only Internet users. In addition, the total internet population has increased, including PC and mobile users. As a result of these changes, it is not possible to select the periods from the turn of December 2020 and January 2021, i.e. including both the data calculated with the current methodology and the new methodology.

At the end of the whole process, the **data is published** in the gemiusAdReal interface according to the following schedule:

- **daily** results are available by the next day;
- **weekly** results are available on Tuesdays;
- **monthly** results are available by the 7th day of the following month.

DATA PROVIDED BY GEMIUSADREAL STUDY

gemiusAdReal study provides detailed information about ads displayed to the Internet PC users, smartphones (selected web browsers and applications), television and radio (selected tv and radio stations). Availability of the information and statistics differs slightly by **ad type**, i.e. depends on whether the data involves results for video, display, text or audio ads.

The table below shows a summary of the creative attributes and indicators available for specific types of ads.

CREATIVE ATTRIBUTES	video ads	display and text ads	audio ads
CREATIVE PREVIEW	✓	✓	✓
CATEGORIZATION[1] by Advertiser (Brand) and Industry	✓	✓	✓
MEDIA CHANNEL (aggregated to Media group>Media channel>Section) where the ad was served	✓[2]	✓	✓
CREATIVE DIMENSIONS (size in pixels)	X	✓	X
POSITION ON WEBSITES (distance from the top of document measured in pixels)	X	✓	X
AUTOPROMOTION - metric informing if a given advertisement was a self-promotional advertisement (details are described in the following section)	✓	✓	✓
AD FORMAT SET - predefined most popular ad formats (e.g. Gigaboard, Half page)	X	✓	X
SOC-DEM – filtering by gender and age	✓	✓	✓
AD POSITION - information about the placements of and ad broken down into screens (current browser window)	X	✓	X
ADVERTISING TECHNOLOGY - dimension informing about the creative serving technology	X	✓	X

[1] The data reported in the gemiusAdReal study is based only on categorised ad creatives gathered from panellists. The categorisation level differs by market, with the target level of approx. 80% for video ads contacts gathered and 70% for display or text ad contacts gathered.

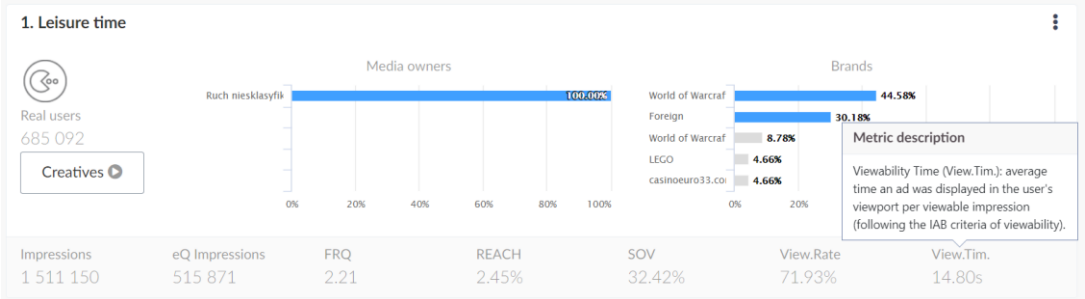
[2] Data regarding video ads for Media owners are presented by media channels on which the ad was displayed, not by the player that served the ad. For example, if the video was an embed from YouTube together with ad, the statistic will be presented for the website where it was displayed, not for YouTube.

METRICS (FOR BRAND OWNER, BRANDS, MEDIA OWNERS AND CHANNELS)	MEDIA TYPE - PC	MEDIA TYPE - Mobile	MEDIA TYPE – TV/Radio
REAL USERS - the number of people in the audience of the advertisement(s)	✓	✓	✓
AD CONTACTS – the total number of contacts with advertisement(s)	✓	✓	✓
VIEWABLE AD CONTACTS – the total number of contacts with viewable advertisement(s)	✓	✓	✓
AD CONTACT TIME – the average time of exposure to the viewable advertisement(s) per one viewable ad contact	✓	✓	✓
VIEWABILITY RATE – the share of viewable ad contacts in all ad contacts	✓	✓	✓
VIEWABLE REAL USERS 2+ - the number of users who have come into contact with the ad at least twice	✓[3]	✓[3]	✓[3]
VIEWABLE REAL USERS 3+ - the number of users who have come into contact with the ad at least three times	✓[3]	✓[3]	✓[3]
VIEWABLE REACH 2+ - reach in cross media for users who have come into contact with the ad at least twice	✓[3]	✓[3]	✓[3]
VIEWABLE REACH 3+ - reach in cross media for users who have come into contact with the ad at least three times	✓[3]	✓[3]	✓[3]
GRP - REACH*FRQ*100	✓	✓	✓
EQ GRP – quality equivalent of GRP	✓	✓	✓
REACH - reach in cross media population	✓	✓	✓
SOV – ad contact share in selected category	✓[4]	✓[4]	✓[4]
FRQ - the average number of ad contacts of the person	✓	✓	✓
EQ AD CONTACTS – quality equivalent of ad contacts	✓	✓	✓
EMISSIONS – the total number of emissions of the advertisement in tv or radio	X	X	✓[5]

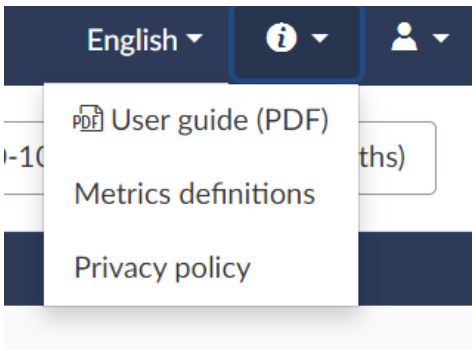
[3] Applies to Timeline view.
[4] Applies to Ranking view.
[5] Applies to List price view.

LIST PRICE – the total cost expressed in PLN of displaying the advertisement, according to the price lists	✓5	✓5	✓5
LIST PRICE (share) – the share of the costs of displaying the advertisement across the market or in a selected dimension	✓5	✓5	✓5

All definitions of the key metrics are available in **tooltips** (under mouse-over in the 'Ranking' section of the gemiusAdReal interface).



Additional detailed definitions are available from the menu in the upper-right corner of the interface:



Apart from analysing the data via the interface, it is also possible to receive additional reports (e.g. about display chains) or alerts about the ad appearing on media channels from brand safety list.

STATISTICAL ERROR

gemiusAdReal is based on data obtained from a carefully selected research panel. The data is processed and extrapolated to the entire population of internet users using PCs and smartphones, as well as traditional media users (TV and radio) who also use smartphones. The population and the way it is processed depends on the country - the above example applies to the Polish market. Its value is dependent mainly on the size of the research sample (i.e. panel size) and amounts to approximately +/-2[6] percentage points (with confidence level 95%) in countries where sample size exceeds 5000 or approx. +/-3 percentage points in countries where the sample size is around 3000. This error can be calculated directly only in reference to the number of Real Users (and the derived

[6] As compared to the standard statistical error values, the values provided hereby for gemiusAdReal data are adjusted additionally due to the usage of the weighting procedure.

metrics such as Reach), as this metric is connected with the population of the Internet users, whereas in case of ad contact there is no direct point of reference that can be treated as the population.

In order to indicate the sample size for particular results, a colour-code was implemented in the gemiusAdReal interface:

- **black** – 50 or more panellists;
- **dark grey** – 20 to 49 panellists;
- **light grey** – less than 20 panellists.

VIEWABILITY MEASUREMENT

Viewability for digital media (pc, mobile) is compliant with IAB's viewability standard - which refers to digital media only - and defines following criteria: an ad was visible at least in 50% in the browser viewport or 30% in case of "large canvas" ads (sized 242 500 pixels or more[7]) for at least 1 second for display and text ads and 2 seconds for video ads. Viewability for traditional media (tv, radio): in order to allow for cross-media analysis, criteria for digital media are adjusted and applied also to traditional media measurement (tv, radio): an ad was audible for at least 2 seconds. In addition in compliance with IAB guidelines, these measurements are executed in certain order, for instance time measurement occurs after fulfilling pixel criteria[8].

Viewability metrics are available for all ad formats (video, display and text) for digital media type starting from January 2017 (the availability of data for 2016 depends on the market) and from January 2020 for TV/Radio. In case of traditional media types (TV & Radio) and cross media analyses (including traditional media) both viewability rate and estimated viewable ad contact metrics do not function.

As is the case with other gemiusAdReal metrics, the **sample size** for viewability data is marked with different colours. Data in grey means that viewability was measured on a small sample that influences the estimation error. In viewability data presentation we take into consideration only the sample of panellists who have had the ad in the viewport of their browser (viewability panellist), so the sample for viewability metrics and other metrics may be different. It is worth to underline that in cases when there are none viewable ad contact measured in gemiusAdReal for a given ad creative, it does not necessarily mean that actually there were none. Especially in case of smaller campaigns or ads with low viewability such situation might result from a small sample size of panellists who had contact with the ad and – consequently – none that had it in viewable part of the browser.

VIEWABILITY DETECTION

To detect viewability for all ad formats (display, text and video) on digital media type, gemiusAdReal software identifies the **ad contact.** As mentioned earlier, every second the software inspects whether there are any new ad creatives loaded to the website visited by the panellist or whether there are any new viewable ad contacts.

The next step is to verify the **pixel requirements**, i.e. identify whether the ad contact:

- happened in the **active browser (and tab) or application window**;

[7] The 30% pixel threshold for large display ads (sized 242 500 pixels or more) was introduced in viewability measurement in gemiusAdReal in January 2018.

[8] Hereby we refer to the "MRC Viewable Ad Impression Measurement Guidelines. Prepared in collaboration with IAB Emerging Innovations Task Force Version 1.0 (Final) – June 30, 2014", available here: https://www.iab.com/wp-content/uploads/2015/06/MRC-Viewable-Ad-Impression-Measurement-Guideline.pdf.

- was in-view for at least **50%(or 30% for large format ads) of the ad pixels**.

The last step is related to verifying **time requirements** and it differs depending on the ad format and applies to each media type included in study:

- **display and text ads:** for all ad formats of display or text gemiusAdReal software detects for how long the display or text ad was in-view from the moment it has been fully loaded in an internet user's browser.
- **video ads:** the gemiusAdReal software detects viewability in two different ways:
    - for **in-stream video ads** gemiusAdReal software measures precisely the stream time (watched material in seconds). This viewability measurement supports the most popular html5 player (while flash and Silverlight players or – on some of the websites - video ads with blob URLs are not supported). Video in-stream ads creatives with a total duration time of less than two seconds are excluded from the study,
    - for **video-in-banner ads** the software measures viewability time of an ad creative inside the frame from the moment it has been fully loaded in the user's screen.

  Viewability metrics for video ads contain information measured by both methods. If we cannot measure viewability from all players for selected campaign (e.g. some ad contacts were running with flash player or blob and some with html5 player), in calculation of viewability rate we include only the measured ad contacts.

SELF-PROMOTION ADS

Information on the volume of self-promotional ads broadcasted by a given publisher is an extremely important element of many analyzes. Distinction of this advertising type is based on the advertiser-publisher relationship map, which is available in the gemiusAdReal interface. An ad is considered as self-promotion when both the given brand and the website on which they are displayed belong to one owner.

AD POSITION

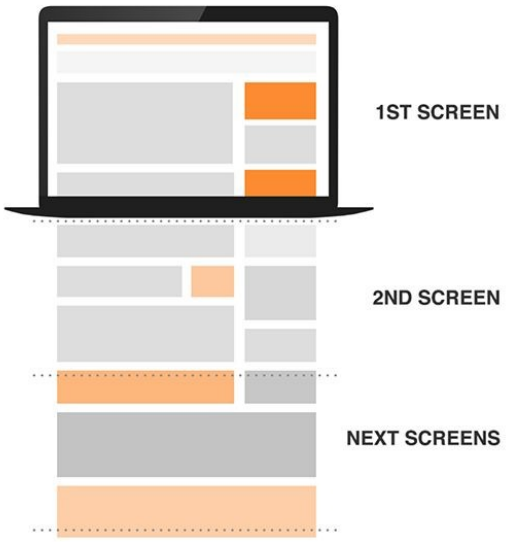
To provide users of gemiusAdReal study with as much information as possible about individual elements of the campaign and their impact on its effectiveness, the interface has a filter and the **Ad Position** dimension. They allow to check which creations are visible on the user's screen when the page is opened (first screen), and which ones require scrolling to be seen.

HOW AD POSITION IS DEFINED

The ad position is defined as a screen number from the top of the page, as in the following example:

Given user screen is based on his/hers current browser window size. Having that in mind the panelist which has active browser window resolution of **3840x2160,** during his page visit will have following screen aggregate values:

- **1st screen** – 0px-2160px,
- **2nd screen** – 2161px-4320px,
- **3rd screen** – 4321px-6480px,
- ….



On the other hand the user with active browser window size of **1920x1080** will have the following screen aggregate values:

- **1st screen** – 0px-1080px,
- **2nd screen** – 1081px-2160px,
- **3rd screen** – 2161px-3240px,
- ….

In other words, the screen is dynamic and always refers to the current size of the user's browser window. In gemiusAdReal interface, the screens are grouped as follows: 1. screen, screens 2-4, 5-7, 8-10, 11-13, 14-16, 17-19, 20+.

Ad position also allows the user to deepen the analysis of the effectiveness of an ad position by checking various metrics such as Ad contact, Real Users, Viewability Rate and others that are available in the study. Each of the metrics relates directly to the advertisement itself, not the screen, hence the calculation method has not changed.

ADVERTISING TECHNOLOGY

In the gemiusAdReal study, it is possible to analyze internet traffic broken down into advertising technologies which publishers use to serve their users ad creations. Currently, the **Advertising technology** dimension includes traffic from the most popular programmatic platforms: **AdSense** and **Criteo**.

HOW THE SOURCE OF ADVERTISING TRAFFIC IS DEFINED IN PROGRAMMATIC MODEL

Due to the nature of programmatic ads, it is extremely difficult to clearly define which of the creations are served using such technology. One of the factors hindering the detection of advertising platforms is the diverse approach of publishers to the implementation as well as the differences in the way these technologies work and serves creations. Having that in mind, the gemiusAdReal study distincts two of largest advertising technologies, whose standardization allows presenting consistent results between all publishers. The results are divided as follows:

- **Google AdSense / Criteo**: if one of the redirects of an ad matches the Google AdSense or Criteo technology pattern, then ad contacts are assigned to the given technology aggregate,
- **Other advertising technology**: if none of the redirects of an ad matches the Google AdSense or Criteo technology pattern, then ad contacts are assigned to the Other advertising technology aggregate (also includes publishers' own advertising technologies),
- **Unrecognized**: if the collected ad does not contain any information about redirects then ad contacts are assigned to the **Unrecognized** aggregate.

Gemius S.A.

Domaniewska 48 Street

02-672 Warsaw, Poland

Phone: + 48 22 390 90 90

+ 48 22 378 30 50

Fax: + 48 22 874 41 01

contact@gemius.com