Measuring the Facebook Advertising Ecosystem

Athanasios Andreou
EURECOM, France

Marcio Silva
UFMG, Brasil

Fabricio Benevenuto
UFMG, Brasil

Oana Goga
Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, France

Patrick Loiseau
Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP, LIG, France &
MPI-SWS, Germany

Alan Mislove
Northeastern University, USA

ABSTRACT

The Facebook advertising platform has been the source of a number of controversies in the past years regarding privacy violation, lack of transparency, as well as its capacity to be used by dishonest actors for discrimination or propaganda.

In this study we aim to shed light into the Facebook advertising ecosystem by analyzing data from more than 400 real-world users to answer two big questions: (i) *Who are the advertisers?*; and (ii) *How are the advertisers using the platform?*. Our analysis is based on the data we collected using AdAnalyst, a browser extension that collects the ads real users receive when they browse their Facebook timeline.

Our results reveal that users are targeted by a wide range of advertisers from popular to niche ones whose trustworthiness is hard to assess; that a large fraction of advertisers are part of potentially sensitive categories such as news, politics, health and even religion; and that the targeting strategies employed by advertisers are either very invasive or opaque. Overall, our work emphasizes the need for better mechanisms to audit ads and advertisers in social media.

1. INTRODUCTION

The Facebook advertising platform has been the source of a number of controversies in recent years regarding privacy violations [25, 32], lack of transparency in the ways Facebook provides information about the ads users see [17], and lately, Facebook's ability to be used by dishonest actors for discriminatory advertising [11, 4, 31] or ad-driven propaganda to influence elections [13]. As example, Propublica demonstrated that Facebook allowed advertisers to reach hate groups such as 'Jew Haters' [4], and also allowed advertisers to exclude people from ads about employment based on their age [11].

This situation has led many governments and privacy advocates to push Facebook to make its platform more transparent and more accountable for the ads that circulate on it [9]. However, providing transparency can be tricky with such a complex system. For example, Andreou et al. [17] recently showed that current transparency mechanisms provided by Facebook, which explain why a user has received a particular ad, are incomplete and sometimes misleading. In addition, a new report from Upturn[12] (supported by many privacy advocates) also finds that Facebook's ad transparency efforts are far from sufficient as

> *Facebook's ad transparency tools do not include an effective way for the public to make sense of the millions of ads running on its platform at any given time ...* [We recommend to] *provide a strong baseline of access to* all *ads, not just those identified as political in nature ...* [and] *disclose data about ads' reach, type, and audience—especially for ads that implicate important rights and public policies.*

So, despite the recent efforts from Facebook to provide transparency, little is still known about the ads inside Facebook and how the platform is used by advertisers.

There are two main features that allow dishonest actors to misuse the platform. First, everyone with a Facebook account can be an advertiser in a matter of 2 minutes and 5 clicks; there is no verification required to place ads, no need to provide a scan of ab identity card or proof that of a legitimate registered business. Second, the platform exposes advertisers to a wide range of ways to target users. For example, advertisers are able to target users that satisfy a precise list of characteristics such as "interested in tennis and having very liberal convictions" that they can choose from a list of over 200,000 attributes provided by Facebook [2]; or they can target specific users if they know their email address or phone number (see Section 2 for more details). We do not aim to debate whether such targeting strategies should be allowed in the first place, but we do believe that it is necessary to understand how they are being used by advertisers.

In this paper we aim to shed light into the Facebook advertising ecosystem by answering two major questions: (i) *Who are the advertisers?* (Section 4); and (ii) *How are the advertisers using the platform?* (Section 5). We investigate questions such as what are the most common targeting strategies advertisers use, who are the users advertisers targeted the most and how do advertisers tailor their ads to specific users. To investigate these questions we analyze data from 481 real-wold Facebook users from two datasets. We collect this data from two versions of AdAnalyst [1], a browser extension that we developed to collect the ads users receive when they browse their Facebook timeline and the corresponding explanations Facebook provides about the reasons they have been targeted with that particular ad. In total we analyze data about 73K/57K ads and 20K/15K advertisers, (Section 3). While our data is unique, difficult to collect, and provides a new perspective on the Facebook advertising ecosystem, it does have biases due to the way we disseminate AdAnalyst and limitations due to the incompleteness of ad explanations provided by Facebook. We provide precise descriptions of how these limitations impact the results and findings throughout the paper.

Our study differs from previous works in three significant ways. First, it is the first-of-its-kind study of ads and advertisers in Facebook. While there have been many studies about online ad targeting [20, 21, 27, 26, 29, 28, 33], none focused on social media advertising or Facebook. Second, it is the first study to analyze the ad targeting strategies of a large number of advertisers. While in our previous work we did focus on explanations of ad targeting, we only performed controlled experiments to evaluate the transparency mechanisms Facebook provides. Third, it is the first study that analyzes ads collected from real-world users (inside or outside Facebook). While there have been several studies that analyze online ads [20], the traditional techniques for collecting ads is to create fake personas and visit a predefined set of websites to collect the corresponding ads.

Our analysis of *who are the advertisers* that use Facebook reveals that the ecosystem is broad and complex. There exist advertisers that are well known, popular (i.e., having more than 100K likes, covering 31% of all advertisers) and trustworthy (more than 73% of popular advertisers have a verified badge) but also many advertisers that are niche (i.e., have less than 1K likes, coverings 18% of all advertisers) and whose trustworthiness is difficult to manually/visually assess (less than 6% of them have a verified badge). We also see that a significant fraction of advertisers (15%) are part of potentially sensitive categories such as news, politics, health and even religion.

Some of the highlights of our analysis on *how the advertisers are using the platform?* reveal that:

(i) *Targeting strategies advertisers use:* A significant fraction of targeting strategies (20%) are either invasive (e.g., make use of Personally Identifiable Information (PII) and attributes from data brokers collected in the offline world to target users) or opaque (e.g., use the "lookalike audiences" feature that lets Facebook decide to whom to send the ad based on a proprietary algorithm). This represents a shift from more traditional targeting strategies such as location-based, behavioral, and re-targeting. Finally, most advertisers (66%) target users with one single ad, and only a small fraction (3%) target users persistently over long periods of time.

(ii) *Which users do advertisers target:* A significant fraction of advertisers (23%) use multiple attributes to target users going to as many as 65 attributes. While in most cases the targeting attributes are in accordance with the business of the advertiser, we do find cases of questionable targeting even from large companies, which emphasized the need for more visibility and accountability in what users advertisers target.

(iii) *How do advertisers tailor their ads to specific users:* A surprisingly large number of advertisers change the content of their ads either across users (74%[1]), across targeting attributes (84%[1]), or across time (84%[1]). While this practice is not inherently malicious, we found several cases where the tailoring of the content could be problematic.

Overall, our analysis points to the fact that users receive ads that can potentially affect their world view, deal with sensitive information, and whose quality is difficult to assess, allowing for potential manipulation of users or even scams. Overall, our work emphasizes the need for better mechanisms to audit ads and advertisers in social media to protect users from dishonest practices that are not only focused on political ads but also concern ads from all kinds of advertisers. As a step forward, AdAnalyst provides users with aggregate statistics about who are the advertisers that target them and what are the properties of other users that were targeted by the same advertisers, which we hope will help users protect themselves from dishonest practices.

2. BACKGROUND

In this section we take a quick look at how one can advertise on Facebook and what are the transparency mechanisms Facebook provides to users.

2.1 Advertising in Facebook

Becoming a Facebook advertiser is trivial, one needs a (personal) Facebook account and than in a few clicks he can become an advertiser by simply clicking on the menu "Create Ads" from the upper left dropdown menu. To send an ad, advertisers need to create a *targeting audience* where they specify the properties of users they want to target, choose some optimization criteria, and place a bid.

To create a *targeting audience* Facebook exposes prospective advertisers to a plethora of options. First, advertisers can target users based on *age, gender, location* and *languages* they speak. Additionally, advertisers can choose to send their ads to a *custom audience* or a *lookalike audience*. Custom audience is a list of specific users advertisers can target directly. Advertisers can use various types of data to create a custom audience list ranging from specifying the emails, phone numbers or physical addresses of people they want to reach, to users that have visited their website, installed their

[1] Out of the relevant set of advertisers.

mobile application, or liked their Facebook Page. Lookalike audiences allow user to let Facebook choose to whom to sends their ads based on previous campaigns. Finally, advertisers can choose from a long list of *targeting attributes* the characteristics they want users who receive their ads to have (e.g., users interested in pingpong and pizza). Advertisers can choose one or multiple such attributes (and even create targeting formulas) and can combine the different targeting options that are provided by Facebook.

2.2 Facebook's transparency mechanisms

Facebook provides explanations to users on why they have received a specific ad, namely *ad explanations*. To obtain such explanations users need to click on the "Why am I seeing this?" button that is in the upper right corner of every ad. Ad explanations look like:

> *One reason you're seeing this ad is that* BMW Karriere *wants to reach people interested in* Software*, based on activity such as liking Pages or clicking on ads.*
>
> *There may be other reasons you're seeing this ad, including that* BMW Karriere *wants to reach people ages 18 and older who live in Germany. This is information based on your Facebook profile and where you've connected to the internet.*

Some explanations are more informative than others. For example, explanations about data broker attributes do not present the attribute that was used in the targeting, while interest explanations are more specific, although limited as well. A comprehensive analysis and evaluation of the ad explanations can be found in reference [17].

3. DATASET

In this paper we use the dataset collected with the help of AdAnalyst [1]. AdAnalyst is a browser extension (available for Chrome and Firefox) that collects two main types of information when users log in their Facebook accounts: (1) the *ads* users receive when they browse their Facebook timeline; and (2) the *ad explanations* provided by Facebook of why they receive a particular ad. We deployed AdAnalyst in two different instances; one for broader worldwide audiences, and one with a focus on Brazilian users. The Brazilian instance was disseminated as part of a project[2] to provide transparency about political campaigns in the 2018 Brazilian elections.

In this study, we will look at data collected from both versions of AdAnalyst. We call the version for broader audiences ADANALYST-WORLWIDE, and the version focused on Brazilian users ADANALYST-BRAZIL. When we do not mention results from ADANALYST-BRAZIL or combined results explicitly, we will be referring to results from ADANALYST-WORLWIDE. We only use data from users that installed AdAnalyst for more than one day. In total we have 99 users in the ADANALYST-WORLWIDE and 382 in ADANALYST-BRAZIL. Next, we provide more details about the data we collect and how we collect it.

[2] www.eleicoes-sem-fake.dcc.ufmg.br

3.1 Data collection

Ads: In order to capture the ads that users receive on Facebook, we look at the DOM of Facebook's HTML code for the tag *"Sponsored"*. This tag is used by Facebook to help users distinguish sponsored content from the rest of the posts in their Facebook feed. The captured frame contains the media content of the ad (either a video or an image), the text of the ad, and a link to the advertiser Facebook page. AdAnalyst does not collect video-ads that appear when a user is watching a video on Facebook. Ads are accompanied by an ad id, which we can use to identify unique ads. In total, we have collected 73.3K unique ads in ADANALYST-WORLWIDE and 56.9K in ADANALYST-BRAZIL. The average number of unique ads per day for a user can range from 1 to 68 and the median is 11.

Ad explanations: By simulating the click on the "Why am I seeing this" button of each ad, we collect the explanation that the user can see regarding the respective ad. Facebook imposes very strict rate limits with respect to the maximum number of explanations we can retrieve. Thus, we developed a scheduling mechanism where we collect all the http requests that can be used to retrieve explanations and get only 10 explanations per hour. Additionally, we do not collect an explanation for an ad if we have already collected the explanation for the same ad for the same user, the last two days before the ad reappeared. In total we collected 68.8K unique ads with their explanations (49.5K for ADANALYST-BRAZIL). We did not manage to collect explanations for 4.4K ads (7.3K for ADANALYST-BRAZIL).

We parse these explanations to retrieve the targeting attributes that are mentioned. For each targeting attribute we also retrieve from the Facebook Advertising Interface [3] its reach (e.g., the number of users that satisfy the attribute).

Advertisers: From all the ads we collected in our dataset we extracted 20K unique advertisers (15K for ADANALYST-BRAZIL). In order to be able to advertise on Facebook, advertisers currently need to create a Facebook Page, while that was not the case in the past. We managed to retrive the Facebook Page of 98.3% of advertisers in both datasets.

The Facebook Pages can provide lots of information about advertisers, we parse: the categories that the advertiser belongs to, the webpage the advertiser has provided, the number of people who have liked the page, or checked-in there and the verification badge (if the advertiser is verified). We do not always have all these informations because they are not mandatory, the only two mandatory informations are the name and the category of the Page/advertiser. The average number of unique advertisers that target users each week, varies from 1 to 194 with a median of 24.

Table 1: Geographical distribution of the datasets.

Location	WORLDWIDE Users	Ads	Adv.	BRAZIL Users	Ads	Adv.
Europe	75	62K	17K	4	2K	789
South America	1	297	131	373	53K	13K
North America	12	4K	2K	5	2K	1K
Rest	11	7K	2K	0	0	0
France	42	18K	7K	1	43	36
Germany	16	42K	11K	1	1K	514
Brazil	1	297	131	372	53K	13K
United States	12	4K	2K	3	2K	993
Total	99	73K	20K	382	57K	15K

3.2 Data limitations

There are two sources of biases and limitations in our dataset, one that comes from users that installed AdAnalyst and one that comes from the way Facebook provides ad explanations.

Representativeness and bias: Representativeness is an important but challenging issue in any empirical study, as ours. We designed a methodology to gather Facebook ads that is as thorough as possible, given our practical constraints. We used two different strategies to disseminate AdAnalyst. The first consisted of disseminating it in our social and family circles as well as in the conferences we attended. For this version, users had to set their Facebook language to English or French. The second dissemination strategy consisted of providing Adnalyst as part of a system focused on bring transparency to an election, in a version that also work in Portuguese. We acknowledge that both strategies are biased towards specific populations, but we hope the setup of two different dissemination strategies may provide hints on the extent to which these biases have influenced our findings. The geographical distributions of our datasets, across continents as well as in some selected countries, is depicted in Table 1.

Limitations on ad explanations: Andreou et al. [17] showed that ad explanations are *incomplete*. This means that in each explanation, at most one targeting attribute appears (plus age/gender/location information), regardless of how many attributes the advertisers use. This means that explanations might reveal only part of the targeting attributes that were used, providing us–and the users–with an incomplete picture of the attributes that advertisers were using. However, in the same study, authors performed a number of controlled experiments that suggest–but not inconclusively prove–that there is a logic behind which attributes appear in an explanation and which not. Specifically, they uncovered the following precedence: Demographics & Age/Gender/Location $>$ Interests $>$ PII-based $>$ Behaviors. When these targeting types are combined by an advertiser, the type with the highest precedence appears in the explanation. Additionally, when attributes of the same type are used in the targeting (i.e., two Interests), the one that appears is the one with the highest estimated audience size. These observations allow us estimate whether our results about a specific targeting type are underestimated or not. We will detail how this limitation impacts the results throughout the paper.

3.3 Ethical considerations

It is important to mention that the code of our developed browser plugins are open source as they can be viewed in the client's machine, as any Chrome and Firefox extension. We only collect information about the ads and clearly stated what we collect to the volunteers who install the extensions and accept our terms. All data collection that we present in this paper was reviewed and approved by the Ethical Review Board of the University of Saarland and by the Institutional Review Board of Northeastern University. Due to IRB restrictions, and in order to minimize any risk of exposure of users' sensitive information, we will not share our data or make them publicly available.

4. WHO ARE THE ADVERTISERS?

In order to analyze the advertisers that target users on Facebook, we proceed with a set of questions and hypothesis that we answer or verify.

4.1 Advertiser's popularity

How popular are advertisers?.
Facebook offers a platform where anyone with a Facebook account can be an advertiser without going through any verification process. This means that the platform is open to both popular and well known advertisers such as Coca Cola as well as more niche advertisers such as the tattoo shop around the corner. We consider the number of likes advertisers got on their Facebook Pages as a measure of their popularity and bin advertisers in three different categories: (i) **niche**, with $1K$ likes or less, (ii) **ordinary**, with likes between $1K$ and $100K$ likes, and (iii) **popular**, which have more than $100K$ likes. Niche advertisers constitute 18% of the Facebook advertisers in our dataset, ordinary 51%, and popular 31% (12%; 58%; 30% for ADANALYST-BRAZIL).

Are there more worldwide or city-wise advertisers?.
Facebook allows location micro-targeting. This possibly attracts advertisers that are local such as restaurants. We identify worldwide advertisers by detecting advertisers that maintain pages in different countries. These advertisers have a global brand root id which corresponds to the advertiser, and then different local ids per country [7]. To identify city-wise advertisers we select the advertisers that have an address in their Facebook Page. While these two proxies are not perfect we find that 8% of the advertisers in our dataset are worldwide, 45% city-wise, and 47% are neither worldwide not city-wise (4%; 45%; 51% for ADANALYST-BRAZIL).

Which advertisers draw the highest fraction of ads?. While there are more ordinary advertisers than popular, popular advertisers contribute to a larger number of ads: 62% of all unique ads we collected come from popular, 32% from ordinary and 6% from niche advertisers (62%; 34%; 4% for ADANALYST-BRAZIL). Additionally, 21% of ads come from worldwide, while 36% come from city-wise and 43% come from the rest (14%; 39%; 47% for ADANALYST-BRAZIL). Hence while a large fraction of Facebook's revenue comes from worldwide and popular advertisers, a non-negligible fraction comes from city-wise and niche advertisers as well.

4.2 Advertiser's categories

When advertisers on Facebook create a Page, they can self-report one or more categories that correspond to their business. Advertisers can either choose from a predefined list of $1,543$ different categories (organized in a hierarchical tree with a max. depth of 6) or input a free-text category.

We observe 932 unique categories in our dataset (824 in ADANALYST-BRAZIL). Figure 1 presents the 20 most common categories among advertisers (they appear in 51.5% of advertisers in our dataset).

Many advertisers only report a general category such as `Website`, `Company`, or `Product/Service` which are not very informative about the sector in which the advertiser works, while others report very fine-grained categories such as `Evangelical Church`, or `Aquarium`, or `Opera House`. To be able to analyze which sectors advertisers come from and to have more homogeneous categories for all, we map advertisers in our dataset to categories in the Interactive Advertising Bureau (IAB) taxonomy [10]. This taxonomy provides categories for advertising purposes and is the standard in advertising. It is composed of 29 Tier-1 categories such as `News and Politics` or `Education`. For the Facebook category `Public Figure` there is no suitable existing IAB category, so we create a new category. For advertisers with only coarse-grained categories such as `Company` or `Website` we do not assign to them any IAB category. In total we manage to map 77% advertisers to a IAB category.

Tables 2 and 3 present the top 10 IAB categories and the respective percentage of advertisers and ads that appear in our datasets. The tables also shows (in the bottom) categories that we consider as possibly sensitive, such as `Medical Health` and `Legal` and are not part of the top 10.[3] The tables show that the top 10 IAB categories are the same in the two datasets with the exception of one category: `Travel` that only appears for ADANALYST-WORLWIDE and `Public Figure` that only appears for ADANALYST-BRAZIL. Besides, there is a significant number of advertisers and ads that come from potentially sensitive categories such as `News and Politics` or `Education`.

4.3 Advertisers's trustworthiness

[3] We consider a category sensitive if, intuitively, we think it can have significant negative consequences on users; we admit that our definition is arbitrary.

(a) ADANALYST-WORLWIDE (b) ADANALYST-BRAZIL

Figure 1: Most popular Facebook advertiser categories.

Table 2: Popular and sensitive (in bold) IAB advertiser categories for ADANALYST-WORLWIDE.

IAB Tier-1 category	Advertisers	Ads
Food and Drink	9.2%	6.2%
Style and Fashion	8.4%	5.9%
News and Politics	7.1%	9.8%
Shopping	6.6%	5.1%
Community Organization	6.4%	3.9%
Technology and Computing	6.4%	7.8%
Travel	4.5%	3.0%
Education	4.3%	5.5%
Healthy Living	4.1%	2.5%
Music and Audio	3.1%	1.2%
Business and Finance	2.0%	2.2%
Medical Health	0.9%	0.5%
Legal	0.2%	0.1%
Religion and Spirituality	0.1%	0.1%

Advertisers can verify their Facebook Page and acquire a badge as proof [14]. There exist two types of badges. Blue badges are for profiles of public interest, and require a copy of an official government-issued photo identification such as a passport. Gray badges are for businesses and require a publicly listed phone number, or a document such as a telephone bill that is associated with the business.

Table 4 shows the fraction of verified advertisers for worldwide, city-wise, niche, ordinary and popular advertisers. In both datasets niche advertisers tend to be less frequently verified (0.2% for blue and 5.5% for gray verification) compared to ordinary (9.8% and 12.7%) and popular advertisers (67% and 59%). In total only 26% of advertisers have a blue badge and 9% a gray one. Our data shows that the majority (54%) of ads come from advertisers that are not verified. Since the advertising platform offers a direct channel to users for (potentially malicious) advertisers, it is essential to be able to estimate the trustworthiness of such advertisers and make them accountable.

4.4 Takeaways

The ecosystem of advertisers in Facebook is broad and complex. There exist advertisers that are global, popular and trustworthy. On the other side, there exist many niche advertisers for which it is difficult to assess the trustworthiness without manual effort. We see that users can be targeted by advertisers that belong to categories dealing with sensitive information such as politics, health, or religion. We also see

Table 3: Popular and sensitive (in bold) IAB advertiser categories for ADANALYST-BRAZIL.

IAB Tier-1 category	Advertisers	Ads
Education	8.8%	10.2%
Food and Drink	7.6%	6.2%
News and Politics	7.3%	8.6%
Music and Audio	6.5%	3.1%
Shopping	6.3%	6.4%
Technology and Computing	5.4%	6.6%
Style & Fashion	5.2%	4.4%
Public Figure	4.6%	4.2%
Community Organization	4.4%	3.2%
Healthy Living	2.7%	1.8%
Medical Health	1.7%	0.8%
Business and Finance	1.6%	2.4%
Legal	0.3%	0.2%
Religion and Spirituality	0.2%	0.1%

Table 4: Fractions of advertisers that are verified (B = blue badge, G = gray badge).

Dataset	World.	Loc.	Niche	Ordinary	Popular
WORLDWIDE	B:82.3%	B:17.4%	B:0.2%	B: 9.8%	B:67.0%
	G:2.9%	G:0.0%	G:5.5%	G:12.7%	G:5.9%
BRAZIL	B:86.6%	B:13.0%	B:0.0%	B: 5.1%	B:53.3%
	G:1.1%	G:0.0%	G:2.3%	G:11.3%	G:11.0%

that many advertisers deal with news delivery, implying that news and media companies focus a lot on Facebook advertising. In total, our analysis points to the fact that users receive ads that can affect their world view, deal with sensitive information, and whose quality is difficult to assess, allowing for potential manipulation of users or even scams.

5. HOW ARE THE ADVERTISERS TARGETING USERS?

For the different types of advertisers identified in the previous section we analyze (i) how they target users; (ii) which users they target; and (iii) how they customize their ads.

5.1 Analysis of targeting strategies

Breakdown of targeting types.

Advertisers on Facebook can choose from a wide range of ways to reach users – see Section 2 for more details. To analyze the different ways advertisers reach people we mine the ad explanations provided by Facebook in the "why am I seeing this?" feature. As mentioned in Section 3.2, explanations are incomplete and only reveal part of the targeting. This means that we do not have a full picture of the targeting strategies used. For each result in this section we describe how this limitation impacts the interpretation of our results.

By looking at the patterns of ad explanations as well as information in the Facebook Advertising Interface, we have identified several broad *targeting types*:

Age/Gender/Location – when advertisers only target users based on their age, gender and location.

Attribute-based – when advertisers target users that satisfy a precise list of attributes. We split this in 5 subcategories based on the source of data: *Behaviors*, *Demographics* and *Interests* – which corresponds to attributes inferred by Facebook from the user's activities on the platform; *Data brokers* – when the targeting is based on attributes inferred by data brokers offline and not by Facebook; and *Profile data* – when attributes correspond to information users provided in their Facebook profiles such as martial status, employer or degree and university attended.

PII-based – where advertisers create a targeting audience based on a list of emails, phone numbers or physical addresses of users they possess.

Retargeting – when advertisers target users that already interacted with their business such as users that visited their page, liked the advertiser's page, responded to an event, or used their mobile app.

Lookalike audiences – where advertisers let Facebook choose their audience based on past results and the characteristics of previous audiences.

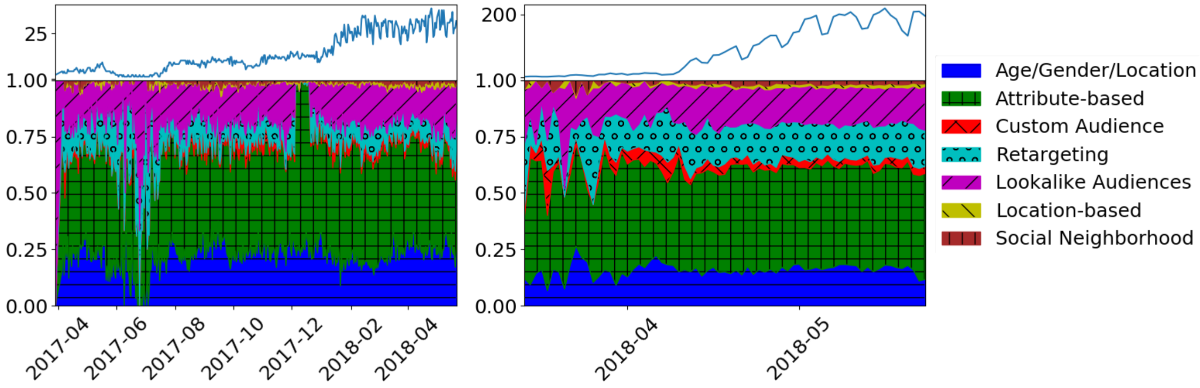
Location-based targeting – when advertiser target users that were or passed by a precise GPS location.

Social neighborhood – when advertisers targets users whose friends liked their Facebook page.

Figures 2a and 2b present a timeline of daily frequency of each targeting type wrt the total number of ads we collected each day. There are not many fluctuations, and in general, the proportion of each targeting type does not change over time or over dataset.[4] Table 5 shows the overall frequency of each targeting type wrt number of ads that have been targeted and fraction of advertisers that have used these targeting types, as well as the fraction of users that have been targeted with these types.

Impact of biases and limitations in the dataset: In the fifth column, Table 5, shows the precedence of each targeting types according to Andreou et al. [17]. In case of multi-type/multi-attribute targeting (e.g., advertisers that use both PII-based and attribute-based targeting at the same time), Facebook only shows one reason in the corresponding explanations (see Section 3). The way Facebook selects the reason shown impacts the frequencies reported in the table. According to [17] the multi-type targeting precedence is: *Demographics & Age/Gender/Location* $>$ *Interests* $>$ *PII-based* $>$ *Behaviors*. All targeting types with a precedence higher than 1 are possibly underestimated, but since we do not know how often advertisers are using multi-type targeting, we cannot estimate the degree of underestimation. On the other hand, there is no overestimation in the results.

[4]The big increase for *Attribute-based* around December and January 2018 can be attributed to a possible bug from Facebook, where many explanations from different advertisers showed the same demographic attribute, namely *Member of a Family-based household*.



(a) ADANALYST-WORLWIDE. (b) ADANALYST-BRAZIL. (c) Labels.

Figure 2: Breakdown of targeting types across time wrt number of ads (across all users). Above: daily number of active users.

Table 5: Breakdown of targeting types with the respective fraction of ads, advertisers, and users who were targeted. Last column presents the attribute precedence (1 – highest precedence; 5 – lowest precedence; nk – not known).

	Ads	**Advertisers**	**Users**	**Precedence**
Age/Gender/Location	23%	35%	95%	1
Behaviors	2%	2%	38%	4
Demographics	2%	3%	33%	1
Interests	37%	48%	94%	2
Profile Data	7%	8%	89%	nk
Data Brokers	1%	1%	30%	nk
PII-based	2%	1%	70%	3
Retargeting	8%	7%	81%	nk
Lookalike Audiences	17%	18%	92%	nk
Location-based	1%	3%	73%	nk
Social Neighborhood	1%	3%	60%	nk

The fact that there are no big fluctuations, and in general the proportion of each category does not change over time or over dataset (even in the beginning of the timeline when the data comes from a smaller number of users) gives us confidence that the numbers we see in this section are not overly biased by the population in our datasets.

From Table 5 we can see that:

(1) *Age/Gender/Location* with 23% of ads and *Attribute-based* with an aggregate of 40% of ads (*Interests* taking the biggest share 37%) are the most prevalent targeting types. These targeting types can be seen as the two more traditional ways of targeting users online.

(2) A large fraction of ads are targeted using *Lookalike audiences*, 17%. This is a newer targeting strategy employed by social media advertising platforms that allows advertisers to ask Facebook to choose who to send the ad to based on previous ad campaigns. This targeting mechanism is problematic because the algorithm behind lookalike audiences is unknown to the public and users have no way of knowing why they received such an ad. On top of this it has been shown that lookalike audiences are vulnerable to deceptive advertisers that can use the mechanism to increase the discrimination in their targeting [31].

(3) A fair share of ads – 8% – are part of *Retargeting*.

(4) While a small share of ads – 2% – are part of $PII-based$ targeting (note that this targeting type has one of the lowest precedence and it is underestimated), a large number of users (70%) have been targeted with at least one $PII-based$ ad – i.e., there exists at least one advertiser that knows the email or the phone number of the user. To date there is no verification process of where advertisers gathered such information and list of phone numbers and emails can be easily bought online [5]. It is important to give special attention to this targeting mechanism especially because it has been shown that it can be used for discriminatory advertising [31] and can be exploited to find more PIIs of users [32].

(5) Surprisingly, *Social neighborhood* targeting accounts only for a very small fraction of ads 1%, which is unexpected on a social media where other posts are based on neighborhood.

Table 6 presents the frequency of each targeting type in terms of ads, advertisers and users in Europe, North America, Brazil, and the rest of the world.[5] We can see that:

(1) *Data brokers* and *PII-based* targeting types are much more frequent in North American, reaching 2% and 5% of the ads respectively (compared to 1% and 2% in Europe). *PII-based* targeting types are much more prominent among users as well – 83% of the North American monitored users have received such ads, while there are only 67% Europeans. This might reflect the differences regarding privacy laws and handling of personal data in general [6].

(2) European advertisers use less retargeting, less lookalike audiences and more age/gender/location targeting. This raises

[5]Note that we assume that the precedence we observe in explanations is consistent across countries.

the question of whether current privacy discussions and laws [6] have an impact on European advertisers' strategies.

Country-specific vs. worldwide targeting.

We now investigate whether advertisers target users in one country or across borders by mining the second part of explanations that specifies age, gender, and location targeting criteria (see Section 2). The overwhelming majority (93%) of advertisers in our datasets target users in only one country.

Impact of biases and limitations in the dataset: Since we have data from users only across a small set of countries, the fraction of worldwide advertisers will be underestimated.

Andreou et al. [17] showed that ad explanations are sometimes misleading: if the advertiser did not specify any location then the current location of the user will appear in the ad explanation, however, if the advertiser did specify a location it will appear as such in the explanation. For advertisers that are country-specific this limitation does not impact the results as users will receive the same explanation no matter where they are traveling. For the worldwide advertisers this can impact the results in both ways: if the users targeted do not cross the border then it will lead to an underestimation; if the users targeted cross the border it will reduce the underestimation due to the country-bias of our dataset.

Table 7 shows the top 10 advertisers wrt the number of countries they target. As we see, while most of them are known companies, they belong to a wide range of IAB categories. Table 8 presents the most popular advertisers that appeared only in one country (where popularity is measured as number of users who received an ad from them).

Persistent vs. one-shot targeting.

We define a *persistent advertiser* as an advertiser that has targeted at least one user for more than two weeks and with more than five ads; and *one-shot advertiser* as an advertiser that targeted all users only once.

Impact of biases and limitations in the dataset: In order not to overestimate the fraction of one-shot advertisers we report results on only advertisers for users for which we have more than 2 weeks of data. We also looked at one-shot advertisers for users for which we have more than 4 and 6 weeks of data and the results are similar so we omit them.

Our results show that the large majority of advertisers 66% $(12,850)$ are one-shot and only a small minority 3% (596) persistently target users (59%; 2% for ADANALYST-BRAZIL). 88% of persistent advertisers have targeted persistently only one or two users, however, some of them have targeted persistently up to 15 users (e.g., `Facebook`, `Data Camp`, `Google`). Table 9 compares the characteristics of the two types of advertisers. We can see the following:

Popularity: in general persistent advertisers are more popular and trustworthy but there exist also niche persistent advertisers (e.g., `SEMY Awards`, an organization that gives industry awards; `Vianex-Fast-Remit` a money transfer company with only 54 likes).

Attribute targeting types: For persistent advertisers, we observe that they use more *PII-based* and *Retargeting* and less *Age/Gender/Location* targeting types (compared to Table 5). For one-shot advertisers, we can see that they use more *Age/Gender/Lo* and *Attribute-based* and less *Lookalike*, *PII-based* and *Retargeting* targeting types. Surprisingly a large fraction (7%) of targeting types for one-shot advertisers are *Location-based* and *Social Neighborhood* (compared to 2% in Table 5).

Adverisers' IAB categories: 13% of persistent advertisers are part of the *News and Politics* IAB category (e.g., `PokerGO` a Facebook page that covers news in Poker; `JB Pritzker` an American politician; the `European parliament`); while only 6% of one-shot advertisers are part of this category. Regarding more sensitive categories, there exist two *Medical Health* persistent advertisers `THINX` related to women's health, and `Merck Group`, pharmaceutical company.

We will discuss in the next section how the text of the ads change across time when a user receives multiple ads from the same advertiser.

Takeaways.

Marketing strategies are changing and it is important to check their vulnerabilities. They are more invasive (custom audiences, data brokers) and opaques (lookalike audiences). The data used from targeting comes from a multitude of sources: advertisers (custom audiences), ad platform (interests), offline (data brokers). There are differences in targeting strategies across countries: more users are targeted with custom audiences and data brokers in the US than Europe and the rest of the world. Most advertisers target users in only one country with one single ad, only a small fraction target users persistently over long periods of time.

5.2 Analysis of targeting attributes

In this section we study the precise attributes advertisers specify to create their targeting audiences. There are four types of attributes according to the Facebook Advertiser API: *Interest (I), Behaviors (B), Demographics (D)* and *Profile data (PD)*. We analyze in this section data on 10K advertisers which have targeted users with 31K ads that have used 2,560 attributes (7K; 21K; 2,838 for ADANALYST-BRAZIL).

Impact of biases and limitations in the dataset: For the analysis in this section we mine the attributes present in the ad explanations provided by Facebook. Andreou et al. [17] showed that if the advertiser uses multiple attributes to create his targeting audiences, only the attribute with the highest reach will appear in the explanation. Thus all the results in the section are biased towards the popular attributes.

What is the reach of targeted attributes?.

The median reach for interests-based attributes is $25.3M$ while the maximum is $1.6B$ and the minimum is $6.7K$ ($17.3M$;

Table 6: Breakdown of targeting types split geographically; with the respective fraction of ads, advertisers and user targeted.

	Europe (75 users)			North America (12 users)			Brazil (372 users)			Rest (11 users)		
	Ads	**Advertisers**	**Users**	**Ads**	**Advertisers**	**Users**	**Ads**	**Advertisers**	**Users**	**Ads**	**Advertisers**	**Users**
Age/Gender/Location	24%	36%	97%	21%	27%	100%	17%	26%	93%	17%	27%	73%
Behaviors	2%	2%	36%	1%	2%	42%	0%	0%	0%	2%	3%	45%
Demographics	2%	3%	31%	1%	2%	42%	0%	0%	0%	2%	4%	36%
Interests	37%	47%	95%	26%	38%	92%	41%	55%	95%	41%	48%	91%
Profile Data	7%	8%	88%	4%	6%	100%	4%	4%	74%	10%	11%	82%
Data Brokers	1%	1%	32%	2%	3%	50%	1%	2%	42%	0%	0%	0%
PII-based	2%	1%	68%	5%	4%	83%	3%	3%	68%	2%	2%	64%
Retargeting	8%	7%	77%	12%	10%	92%	15%	12%	91%	9%	10%	91%
Lookalike Audiences	16%	17%	92%	25%	29%	100%	16%	15%	91%	16%	19%	82%
Location-based	1%	3%	76%	1%	2%	67%	2%	4%	48%	1%	2%	55%
Social Neighborhood	1%	3%	55%	1%	2%	83%	3%	7%	53%	1%	4%	64%

Table 7: Advertisers that advertise users across countries (✓for verified advertisers, **L** – number of likes, **C** – number of countries, **U** – number of users).

Name	**L**	**IAB Categories**	**C**	**U**
Google ✓	$26M$	Tech. and Comp.	12	84
Netflix ✓	$47M$	News and Politics	11	164
Airbnb ✓	$13M$	Travel	11	104
Udemy ✓	$4M$	Education	10	160
Crossover	$634K$	Other	10	42
Booking.com ✓	$11M$	Travel	9	124
The Economist ✓	$9M$	News and Politics	9	52
Toggl	$25K$	Other	9	45
Must-see Kickstarter projects	$41K$	Other	9	23
DigitalOcean ✓	$121K$	Tech. and Comp.	9	16

$1.6B$; $9.7K$ for ADANALYST-BRAZIL). The median reach for behavior and demographic-based attributes are $80.5M$ and $17.1M$ respectively. Note that this is an upper bound of the actual reach of attributes used in targeting.

Do advertisers tend to use more predefined attributes or free-text ones?.

Interest-based attributes can either be predefined, where the advertisers can browse in a tree structure of attributes; or free-text, where the advertisers can type something they believe is related with their desired targeting and get exposed to related attributes (which usually correspond to users that have engaged with a particular Facebook Page). In our dataset, 71% of the ads contain free-text attributes while only 29% contain predefined ones (72%; 28% for ADANALYST-BRAZIL). We did not expect to observe such a high fraction of free-text attributes and this percentage is likely underestimated given they have a smaller reach than predefined ones. Free-text attributes can be used as a proxy to discriminate against people [31] and can sometimes be more sensitive.

Do advertisers use multiple attributes?.

Since Facebook offers so many options for targeting, we expect that some advertisers might use of many different attributes. In our dataset 23% of advertisers have used more than one attribute in their targeting with some using more than 15 different attributes.[6] Table 13 in [16][7] shows the advertisers that have used the largest number of attributes. We can see advertisers such as `Google`, `Adidas` or `Forbes`. While many of the attributes used seem relevant to their business, some of them are more questionable. For example `Google` has used attributes such as `Married`, `Family`, `Women's rights`, `Politics and social issues` and `US politics (very liberal)` to target users. We will investigate in the next section how the ads of an advertisers vary with the targeting attributes he uses.

What are the most and least used attributes?.

Figure 3 shows the top 10 attributes that appear most frequently in ad explanations (3a), were used by the largest fraction of advertisers (3b) and were seen by the largest number of users in their ad explanations(3c). We can see that most attributes are either languages, or broad interest-based attributes such as `Travel` and `Entertainment`. Besides, 39% of attributes appear in only one ad (Table 10 presents a random sample); 50% have been used by only one advertiser; and 65% have been seen by only one user (44%; 55%; and 56% for ADANALYST-BRAZIL).

Who targets what?.

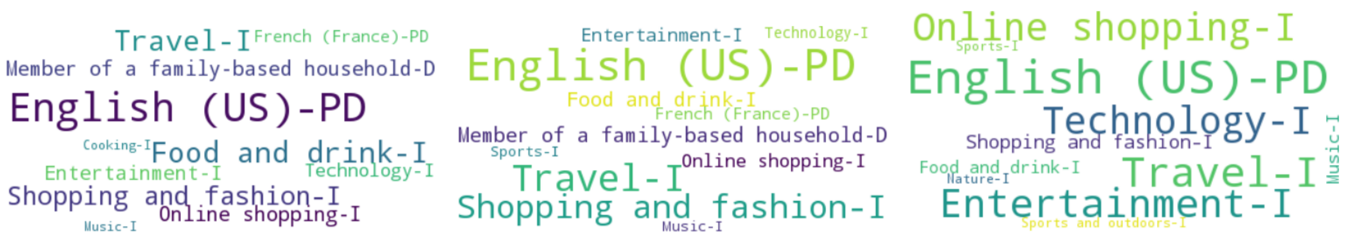
Table 14 in [16] shows the 10 most frequent attributes per IAB category of advertisers (Table 15 in [16] for ADANALYST-BRAZIL). Generally, the most frequent attributes used for each category are in concordance with the type of the advertiser. However, there are attributes that are used by some of the advertisers that are questionable. For example, a real estate advertisers target people that are `Engaged`, have an `iPhone`; a politician target people that are interested in `LGBT community`, `Homosexuality`, and `Anti-fascism` and a medical insurance company targets users with interests in `Fitness and wellness`, `Politics and social issues` and education statuses such as `Master's Degree`.

[6] Remember that we can only observe one attribute for each ad campaign (even if the advertiser used multiple), here we check the number of different observed attributes across multiple ad campaigns.

[7] Due to space constraints, Tables 13–15 are presented only in our technical report [16].

Table 8: Most popular advertisers that have appeared only in one country (**U** – number of users who received an ad from them).

France			Germany			Brazil			US		
Name	**U**	**IAB**	**Name**	**U**	**IAB**	**Name**	**U**	**IAB**	**Name**	**U**	**IAB**
Les Echos	20	News and Polit.	Telekom Shop	15	Tech. and Computing	TIM Brasil	139	Tech. and Computing	International Rescue Committee	11	Community Organization
McDonald's France	18	Food and Drink	Saturn Deutschland	14	Other	Mercado Livre	129	Other	AT&T	9	Other
Sosh	15	Other	germantaxes.de	14	Business and Fin.	TAG - Experiłncias Literrias	118	Other	Amazon.com	9	Shopping
Amazon.fr	12	Shopping	REWE	13	Food and Drink	Santander Brasil	107	Business and Fin.	Verizon	8	Other
Renfe-SNCF	11	Other	EDEKA	12	Food and Drink		101	Other	Airtable	6	Other
Just Eat France	11	Food and Drink	Sky Ticket	12	Television	Renner	97	Style and Fashion	Brandless	6	Shopping
Heineken FR	11	Food and Drink	Amazon.de	11	Shopping	Cheetos Brasil	95	Food and Drink	Xfinity	6	Other
Monoprix	11	Other	Vodafone Deutschland	11	Other	Kanui	88	Other	Starbucks	6	Food and Drink
VICE France	11	News and Polit.	Domino's Deutschland	11	Food and Drink	StartSe	83	News and Polit.	WIRED	6	News and Polit.
Dacia France	10	Automotive	ING-DiBa	10	Business and Fin.	NET	80	Tech. and Computing	GEICO	6	Business and Fin.



(a) wrt the fraction of ads with the attribute (ADANALYST-WORLWIDE).

(b) wrt the fraction of advertisers that use the attribute (ADANALYST-WORLWIDE).

(c) wrt the fraction of user that were targeted with the attribute (ADANALYST-WORLWIDE).

Figure 3: Top targeting attributes (I for Interests, B for behaviors, D for demographics, PD for profile data).

Takeaways.

Most interest-based attributes used in targeting are free-text ones and not predefined, free-text attributes are vulnerable to a wider range of privacy attacks. They can be used to discriminate against people and are sometime more privacy sensitive. A significant fraction of advertisers use multiple attributes to target users going to as many as 65 attributes. While in most cases the targeting attributes are in accordance with the business of the advertiser, we do find cases of questionable targeting even from big companies, which emphasizes the need for more visibility and accountability in what users advertisers target.

5.3 Analysis of targeted ads

For marketing reasons advertisers could tweak the content of their ads to get better engagement. In this section we analyze whether (and how) advertisers tailor their ads across three dimensions: (1) *over time* for the same user, (2) *across users*, and (3) *across targeting attributes*. While these practice are not necessarily evil, they might be problematic in some cases such as political advertising.

Ads that change over time for the same user.

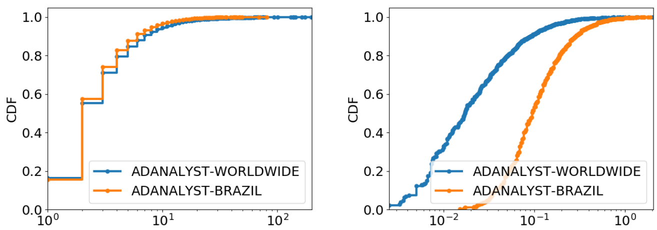
To measure what percentage of advertisers change the content of their ads over time for a specific user we look user-advertisers pairs. Out of the 33K user-advertisers pairs we have in our dataset, in 35% of them the advertiser send two or more ads to a user; which we consider in this analysis.

To identify advertisers that change the content of their ads we count the number of ads with different texts for each user-advertiser. Figure 4a shows the CDF of the number of ads with different texts for each user-advertiser pair. The figure shows that 84% of user-advertiser pairs have two or more ads with different texts (and this corresponds to 83% of the advertisers we consider). Furthermore, 5.4% of user-advertiser pairs have more than 10 ads with different texts.

To study the properties of advertiser that change the most often their text we need to normalize the number of texts in each user-advertiser pair by the number of days in which we

Table 9: Characteristics of persistent and one-shot advertisers.

	Persistent	One-shot
Verified	67%	26%
Popular/Ordinary/Niche	74%/24%/2%	22%/56%/22%
Top targeting types	Attr-based 44% A/G/L 19% Lookalike 16% Retargeting 15% PII 4% Social N. 1% Location 0%	Attr-based 51% A/G/L 29% Lookalike 10% Retargeting 3% PII 0$ Social N. 4% Location 3%
Top IAB categories	News & Pol. 15% Style & F. 14% Food & Dr. 11% Tech. & Comp. 11% Shopping 9%	Style & F. 14% Food & Dr. 12% Education 7% Shopping 7% News & Pol. 6%



(a) raw (b) normalized

Figure 4: CDF of number of different texts in ads for each user-advertiser pair.

have collected ads for the user as for some users we collected data for longer periods of time than others. We analyze next advertisers corresponding to the top 10% user-advertisers pairs with most text changes in their ads (normalized). This corresponds to 706 advertisers that have targeted 90 users (543; 297 for ADANALYST-BRAZIL). Table 11 shows the most frequent IAB categories of these advertisers. We can see that advertisers that change the most often the content of their ads are part of `News and Politics` (12.6%).

To understand how these advertisers are changing the content of their ads Table 12 presents a sample of advertisers and the text of their ads from `News and Politics`.

Ads that change over users.

To analyze the advertisers that change the content of their ads across users we consider two measures: (1) *all-disjoint* – where given an advertiser, each user has been targeted with a different ad, i.e., there is no overlap in the ads received by *any* of the users; and (2) *one-disjoint* – where given an advertiser, there exist at least one user that received ads that are different than the rest of the users targeted by the advertiser, i.e., there exist a user with an empty overlap between his ads and the ads received by the rest of the users.

We consider that two ads are different if the text that appears is different. To account for the fact that the text that appears in two ads is different just because it is in two different languages we only consider ads that are in English. We also repeat the analysis for only ads that are in Portuguese, French and German. In order to detect the language of a text, we use the Google Translate API [8]. For the analysis we also consider only advertisers that targeted more than three users.

Out of the 678 advertisers in our dataset that have sent ads in English and have targeted more than three users, 73.9% are one-disjoint and 10.6% are all-disjoint. For Portuguese, French and German ads the percentage of all-disjoint advertisers are 2.7%, 11.3% and 7.8%.

We analyze next the all-disjoint advertisers with English ads. Table 11 presents the fraction of these advertisers that belong to the different IAB categories. Again `News and Politics` is the top category.

Table 12 presents a sample of advertisers and the text of their ads for different users from the `News and Politics` and `Medical` categories.

Ads that change over targeting attributes.

To analyze the advertisers that change the content of their ads over targeting attributes we consider again the two previously introduced measures: (1) *all-disjoint* – where given an advertiser, there is no overlap in the text of the ads targeted to different attributes; and (2) *one-disjoint* – where given an advertiser, there exist a targeted attribute with an empty overlap between his ads and the ads targeted with other attributes.

We are going to filter out advertisers that have targeted with only one attribute. Out of the 2.436 advertisers we considered, 83.09% are one-disjoint and 61.8% are all-disjoint (1.826; 73.27%; 48.5% for ADANALYST-BRAZIL). Table 11 presents the fraction of all-disjoint advertisers that belong to the different IAB categories. Again `News and Politics` is the top category. Table 12 presents a sample of advertisers and the text of their ads for different targeting attributes from the `News and Politics`.

Takeaways.

A surprisingly large number of advertisers change the content of their ads either across users, across targeting attributes or across time and the largest fraction of them are the News and Politics category. While this practice is not necessarily evil, we found examples of several cases where the tailoring of the content might be problematic.

6. RELATED WORK

Facebook is a multibillionaire social network in which its main revenue comes from ads. Not surprisingly, its ads platform has showed to be quite effective in many marketing segments[8]. However, the many behavioral, demographic, and interest options that Facebook provides to advertisers

[8]https://www.wordstream.com/blog/ws/2017/02/28/facebook-advertising-benchmarks

Table 10: Random sample of attributes that have appeared in just one ad explanation.

Attribute Type	**Attributes**
Interests	Country music, Animal rescue group, Fundraising, Fallout (series), IOS, Urdu, Andreea Raicu, Clique, Chocolate cake, Tattoo removal
Behaviors	Uses a mobile device (18-24 months), Primary Browser: Safari, Anniversary in 61-90 Days, Smartphone Owners, Nexus 5, Expats (Italy), Returned from trip 2 weeks ago, HTC, Primary OS Windows 7, Expats (Colombia)
Demographics	Anniversary within 30 Days, Close Friends of Women with a Birthday in 7-30 days , Upcoming birthday, Birthday in 01 January
Profile Data	Universitatea BABE - BOLYAI, Student, CTO, UCLA, Saarland University, Croatian, IIT Kharagpur, Professor

Table 11: Fraction of advertisers that belongs to different IAB categories and change the content of their ads across time, users and attributes.

	WORLDWIDE			BRAZIL		
IAB category	**Time**	**Users**	**Attr.**	**Time**	**Users**	**Attr.**
Food & Drink	10.0%	4.2%	13.6%*	15.3%*	7.3%	8.9%
Style & F.	12.6%*	16.9%*	10.6%	8.4%	2.4%	6.2%
News & Pol.	12.6%*	21.1%*	12.2%*	15.3%*	12.2%*	10.8%*
Shopping	7.4%	7.0%	9.2%	7.4%	12.2%	9.2%
Community O..	4.2%	1.4%	3.6%	2.7%	0.0%	4.4%
Tech. & Comp.	11.6%	15.5%	8.2%	7.6%	4.9%	5.1%
Travel	8.2%	11.3%	9.2%	3.8%	2.4%	4.6%
Education	6.2%	5.6%	6.0%	10.7%	26.8%*	16.3%*
Healthy Living	3.5%	1.4%	3.8%	2.9%	0.0%	1.9%
Music & Audio	1.4%	0.0%	1.5%	1.7%	4.9%	5.6%
Business & F.	4.3%	0.0%	2.9%	2.5%	2.4%	3.2%
Medical H.	0.5%	1.4%	0.9%	0.6%	0.0%	0.7%
Legal	0.2%	0.0%	0.1%	0.2%	0.0%	0.2%
Religion & S.	0.2%	0.0%	0.0%	0.0%	0.0%	0.1%

have been raising concerns about its use for political campaigns [13] as well as a form to create discriminative ads. Particularly, Speicher *et al.* [31] investigated the different targeting options provided by Facebook and their ability to be potentially abused by malicious advertisers to target users based on gender and race attributes. Similarly, Korolova *et al.* [25] provides a detailed discussion about how the design of the Facebook ads platform could be exploited to violate the users privacy. Security issues on the Facebook ads platform have also been investigated by Venkatadri *et al.* [32], where authors demonstrate forms of attacks that allowed an adversary to exploit the interface to infer users' PII as well as to infer their activity. Andreou *et al.* [17] investigated the level of transparency of Facebook explanations and showed that the Facebook ad explanations are often incomplete and sometimes misleading, while data explanations are often incomplete and vague. Complementarily, Eslami *et al.*[23] provides a better understanding on how communicating aspects of the algorithmic ad curation process affects user's perception of their ad experience. There has also been a growing number of recent efforts that exploit the Facebook Ads API to extract behavioral and demographic patterns from user populations. This approach has showed to be useful for many different applications, including monitoring lifestyle diseases [18], study worldwide gender inequality [24], to study the movement of migrants [35], and to infer the political leaning of news outlets in large scale [30].

There are other efforts that attempt to understand how ads are displayed in other systems. Wills *et al.*[34] investigated what Google Ad do with the information they know about users. Authors studied the ads shown to users during controlled browsing as well as examine the inferred demographics and interests shown in Ad Preference Managers provided by advertisers. Their findings suggest that the Google Ad Network provides contextual, behavioral, location-based ads, and, in some cases, behavioral aspects of users like sexual orientation, health and financial matters. In the same line, Barford [19] provided an in-depth understanding about the features, mechanisms and dynamics of display advertising on the web. Particularly, they show when targeting is used, the specific types of ads delivered generally correspond with the details of user profiles, and also with users' patterns of visit. Another set of efforts concentrate on identifying how trackers are used to gather users data. In this line, Acar [15] studied the mechanism of maintaining persistent cookies even though the user removes browser cookies. They show that there are other mechanisms of user's tracking even savvy users cannot remove. More recently, Englehardt [22] created a tool to help users to discover how intrusive are the online trackers.

Our work effort is complementarily to the above studies, as we gather the ads using a browser plugin, a methodology that provides a unique perspective about the ads shared in Faceebook. Our approach opens this black-box ads ecosystem from Facebook, providing a unique understanding on how advertisers are using this particular system.

7. CONCLUSION

In this study, we shed some light into the Facebook advertising ecosystem by collecting and analyzing data from more than 400 real-world users. We tackle two main questions: (i) *Who are the advertisers?*; and (ii) *How are the advertisers using the platform?*. Our results reveal for instance that users are targeted by a wide range of advertisers from popular to niche ones whose trustworthiness is hard to assess; that a large fraction of advertisers are part of potentially sensitive categories such as news, politics, health and religion; and that the targeting strategies employed by advertisers can be invasive or opaque.

Overall, our work shows that there is a range of poten-

Table 12: Examples of advertisers that change the content of their ads across targeting attributes, users and time.

Name	Att/Usr/Time	Text of ads
VICE News	The New York Times	As of September 1, U.S. citizens can no longer travel to North Korea. We went to the Hermi Kingdom with one of the last tourists to go. ** As North Korea celebrated its founder's 105th birthday, VICE returned to the Hermit Kingdom to see how its citizens are reacting to the growing crisis. ** There's a giant inflatable Trump Chicken on the south lawn of The White House . ** It was supposed to be a press conference about infrastructure, but then it took a turn. ** Donald Trump always seems to say what Donald Trump won't say. **
	PC Magazine	A self-driving, flying taxi could soon be a reality **
	US politics (very liberal)	BuzzFeed News' plan to fight a lawsuit related to the infamous ?pee tape? dossier: prove some of the allegations against Donald Trump are true. ** One of the reasons it's hard for Trump to navigate the guns issue after Parkland is that the gun rights community itself is still trying to figure out what change is acceptable. **
	Finance	A $10 billion lawsuit could finally unmask the creator of bitcoin news.vice.com ?Dave was found dead in his home. The scene of Kleiman?s death was gruesome. His body was... **
	Democratic Party	Mr. Trump and Mr. Cohen have a lot of explaining to do. ** VICE News had exclusive access from the front-lines of Charlottesville, and you can watch the full episode now. ** VICE News: We're possibly the only media organization to be certified as "fake news incorporated" by Sebastian Gorka. **
	I fucking love science	But can they get it delivered to the International Space Station in 30 minutes or less? **
En Marche	Europe	Europe, structuration territoriale, engagement citoyen... Vous avez rat la confrence de presse de rentre de Christophe Castaner vendredi? Allez... on vous la rsume en 2 minutes. Top chrono ! ? **
	Fair trade	C?tait une promesse de campagne d'Emmanuel Macron, aujourd?hui a t lance la # FrenchImpact : un acclrateur pour permettre le dveloppement de l'ESS et faire en sorte que les initiatives locales qui fonctionnent deviennent des solutions nationales ! ?? **
	Emmanuel Macron	LIVE — Suivez notre confrence sur la biothique en prsence de Didier Sicard, Monique Canto-Sperber, Irne Thry et Alain Fontanel. ** ??? Connaissez-vous le RGPD ? Non ? Et pourtant, c?est une petite rvolution. **
Merck Group	Master's degree	How our smart innovations are driving the future of personal mobility. # alwayscurious **
	Healthcare and Medical	Escape the desk: create an environment where curiosity thrives. # catchcurious ** Does your business model empower curiosity? # catchcurious ** Optimizing curiosity curiosity.merckgroup.com Escape the desk: create an environment where curiosity thrives. #catchcurious ** Merck Group curiosity.merckgroup.com ** Can curiosity take higher education further? # catchcurious ** Curiosity as a means of survival? Find out more: www.curiosity.merckgroup.com/stories/curiosity-and-brain # catchcurious ** Into the unknown curiosity.merckgroup.com The will of discovery ? curious for more! #catchcurious ** Survival through curiosity curiosity.merckgroup.com Curiosity as a means of survival? Find out more: www.curiosity.merckgroup.com/stories/curi... **
Durex	User 1	Happy Holi ;) Buy now http://bit.ly/2un0NBQ ** This Rose Day, # CutTheCliches with Durex. Buy now: http://bit.ly/2un0NBQ *
	User 2	Is Kate your perfect girl? Every hour, at least one person in Ireland* is diagnosed with an STI**. (*ROI only **Based on 2016 HPSC data) Date with Durex. **
	User 3	IIf you were hating condoms, you'll love Durex AiR, so thin, it's like it's not even there. What are you waiting for? Shop now: http://amzn.to/2oTWcTm \# HateCondomsLoveDurexAiR ** "Durex Feel Thin, gets closer than ever before". Buy Now : http://amzn.to/2ryWiDa ** What's happening India, we want to know why? **
Bloomberg	User 1	A doctor told him to go home to die.
	User 2	Even though Ma "had no business plan."
	User 3	Just look at Cape Town. ** The world is more complex than ever, which makes big risks more dangerous.
	User 4	Offshore oil rigs have a $38 million problem. ** Only 3-5% of oil and gas equipment is currently connected to the cloud.
	User 5	Your petabytes can help you prepare. ** What IoT developers can learn from Apple. ** This sector is predicted to surge... ten times over. ** It will be bigger than the smartphone market. ** Is your company ready to shop for its next digital merger? ** Elon Musk thinks AI poses the biggest threat to humanity. **
New York Times		When polarizing ideas dominate the discussion, inform your opinions with The New York Times. ** Subscribe to The New York Times, and trade final exams for fitness with the Well blog. ** In Moscow. In Bolivia. In Beijing. In Damascus. Everywhere the story is taking place. ** This is your day to get a year of The New York Times. Subscribe now. Cancel anytime. ** How will the U.S. view Macron? Subscribe to The New York Times and find out. ** This is a chance to hold power to account. ** Investigative reporting has never been so important. ** The news you need. The journalism you deserve. ** Nationalism. Centrism. Socialism. Journalism. ** Find bold opinions and fresh perspectives, daily. Save on The New York Times. ** "I'm not sure it's possible to justify my liaisons with married men, but what I learned from having them warrants discussion." ** So ... you're saying we shouldn't adopt one? ** Following the world?s most important stories wherever they lead. Subscribe now. Cancel anytime. ** Facts. We seek them out. We check them. We help you make sense of them. The New York Times. ** See how President Macron will shape the E.U. and the world. Subscribe to The New York Times. ** No. 1: Wear comfortable underwear ** A victory for Merkel. But also for the far-right. ** More photojournalists on staff than any other newsroom. ** ?I?m hoping for a crib death,? wrote one user. ?Deport the scum immediately,? read another online comment. ** "I have never understood why some guys seem to think flattery is the key to a bedroom they?ve already been welcomed into." ** The truth is vital. Our journalists won?t rest until they find it. The New York Times. ** Journalism that shakes the status quo. ** Does a ?do not resuscitate? tattoo speak for itself? ** Where will Macron lead France? Subscribe to The New York Times and get the facts. ** This is your day to get a year of The New York Times. ** Need a study break? Subscribe to The New York Times and explore award-winning Times Video. ** The most innovative newsroom in journalism. And reporters who still knock on doors. ** ?Something that started decades ago and was applauded and inoffensive is now politically incorrect. What can you do?? Lisa Simpson says. ** Get The New York Times for as low as $1 a week. ** Reporting for those who want to know more. ** Find your perfect post-exams podcast. Subscribe to The New York Times. ** France?s next chapter, page one. ** Our journalists investigate the stories that matter.

tially questionable uses of the platform, which calls for better mechanisms to audit ads and advertisers in social media.

Our analysis is based on the data we collected using AdAnalyst, a browser extension that collects the ads real users receive when they browse their Facebook timeline. In addition to collecting data, AdAnalyst provides societal benefit as it helps users better understand the data that the platform has about them and how it is being used.

8. REFERENCES

[1] Adanalyst. `https://adanalyst.mpi-sws.org`. Accessed: 2018-05-25.

[2] The definitive list of what everyone likes on facebook. `https://bit.ly/2kqmIlu`. Accessed: 2018-05-25.

[3] Facebook ads. `https://www.facebook.com/business/products/ads`. Accessed: 2018-05-25.

[4] Facebook enabled advertisers to reach jew haters. `https://bit.ly/2h62JGP`. Accessed: 2018-05-25.

[5] Facebook plans crackdown on ad targeting by email without consent. `https://techcrunch.com/2018/03/31/custom-audiences-certification`. Accessed: 2018-05-25.

[6] General data protection regulation. `https://gdpr-info.eu/`. Accessed: 2018-05-25.

[7] Global pages. `https://www.facebook.com/business/help/905034079579176`. Accessed: 2018-05-25.

[8] Google translate-detect language. `https://cloud.google.com/translate/docs/detecting-language`. Accessed: 2018-05-25.

[9] Honest ads. `https://cbsn.ws/2Lsl1QK`. Accessed: 2018-05-25.

[10] Interactive advertising bureau (iab). `https://www.iab.com/`. Accessed: 2018-05-25.

[11] Is it age discrimination if you don't know youre being discriminated against? `https://bit.ly/2ncNgsC`. Accessed: 2018-05-25.

[12] Leveling the platform: Real transparency for paid messages on facebook. `https://www.teamupturn.org/reports/2018/facebook-ads/`. Accessed: 2018-05-25.

[13] These are the ads russia bought on facebook in 2016. `https://www.nytimes.com/2017/11/01/us/politics/russia-2016-election-facebook.html`. Accessed: 2018-05-25.

[14] What is a verified page or profile? `https://www.facebook.com/help/196050490547892`. Accessed: 2018-05-25.

[15] G. Acar, C. Eubank, S. Englehardt, M. Juarez, A. Narayanan, and C. Diaz. The web never forgets: Persistent tracking mechanisms in the wild. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, CCS '14, pages 674–689, New York, NY, USA, 2014. ACM.

[16] A. Andreou, M. Silva, F. Benevenuto, O. Goga, P. Loiseau, and A. Mislove. Measuring the facebook advertising ecosystem, 2018. Technical Report available at https://lig-membres.imag.fr/gogao/IMC2018-long.pdf.

[17] A. Andreou, G. Venkatadri, O. Goga, K. P. Gummadi, P. Loiseau, and A. Mislove. Investigating ad transparency mechanisms in social media: A case study of facebooks explanations. In *The Network and Distributed System Security Symposium (NDSS)*, 2018.

[18] M. Araujo, Y. Mejova, I. Weber, and F. Benevenuto. Using facebook ads audiences for global lifestyle disease surveillance: Promises and limitations. In *Proceedings of the 9th ACM Conference on Web Science (WebSci)*, 2017.

[19] P. Barford, I. Canadi, D. Krushevskaja, Q. Ma, and S. Muthukrishnan. Adscape: Harvesting and analyzing online display ads. In *Proceedings of the 23rd International Conference on World Wide Web*, WWW '14, pages 597–608, New York, NY, USA, 2014. ACM.

[20] J. M. Carrascosa, J. Mikians, R. Cuevas, V. Erramilli, and N. Laoutaris. I always feel like somebody's watching me: Measuring online behavioural advertising. In *Proceedings of the 11th ACM Conference on Emerging Networking Experiments and Technologies*, CoNEXT '15, pages 13:1–13:13, New York, NY, USA, 2015. ACM.

[21] A. Datta, M. C. Tschantz, and A. Datta. Automated experiments on ad privacy settings. *PETS*, 2015.

[22] S. Englehardt and A. Narayanan. Online tracking: A 1-million-site measurement and analysis. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, CCS '16, pages 1388–1401, New York, NY, USA, 2016. ACM.

[23] M. Eslami, S. R. Krishna Kumaran, C. Sandvig, and K. Karahalios. Communicating algorithmic process in online behavioral advertising. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 432:1–432:13, New York, NY, USA, 2018. ACM.

[24] D. Garcia, Y. M. Kassa, A. Cuevas, M. Cebrian, E. Moro, I. Rahwan, and R. Cuevas. Facebook's gender divide. *arXiv preprint arXiv:1710.03705*, 2017.

[25] A. Korolova. Privacy violations using microtargeted ads: A case study. In *2010 IEEE International Conference on Data Mining Workshops*, pages 474–482, Dec 2010.

[26] M. Lécuyer, G. Ducoffe, F. Lan, A. Papancea, T. Petsios, R. Spahn, A. Chaintreau, and R. Geambasu. Xray: Enhancing the web's transparency with differential correlation. In *USENIX Security*, 2014.

[27] M. Lecuyer, R. Spahn, Y. Spiliopolous, A. Chaintreau, R. Geambasu, and D. Hsu. Sunlight: Fine-grained targeting detection at scale with statistical confidence. In *ACM CCS*, 2015.

[28] B. Liu, A. Sheth, U. Weinsberg, J. Chandrashekar, and R. Govindan. Adreveal: improving transparency into online targeted advertising. In *ACM HotNets*, 2013.

[29] J. Parra-Arnau, J. P. Achara, and C. Castelluccia. Myadchoices: Bringing transparency and control to online advertising. *ACM Trans. Web*, 2017.

[30] F. Ribeiro, L. Henrique, F. Benevenuto, A. Chakraborty, J. Kulshrestha, M. Babaei, and K. P. Gummadi. Media bias monitor: Quantifying biases of social media news outlets at large-scale. In

Proceedings of the International AAAI Conference on Web and Social Media, ICWSM'18, Stanford, USA, June 2018.
[31] T. Speicher, M. Ali, G. Venkatadri, F. N. Ribeiro, G. Arvanitakis, F. Benevenuto, K. P. Gummadi, P. Loiseau, and A. Mislove. Potential for discrimination in online targeted advertising. In S. A. Friedler and C. Wilson, editors, *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, volume 81 of *Proceedings of Machine Learning Research*, pages 5–19, New York, NY, USA, 23–24 Feb 2018. PMLR.
[32] G. Venkatadri, A. Andreou, Y. Liu, A. Mislove, K. P. Gummadi, P. Loiseau, and O. Goga. Privacy risks with Facebook's PII-based targeting: Auditing a data broker's advertising interface. In *S&P 2018, IEEE Symposium on Security and Privacy*, 2018.
[33] C. E. Wills and C. Tatar. Understanding what they do with what they know. In *ACM WPES*, 2012.
[34] C. E. Wills and C. Tatar. Understanding what they do with what they know. In *Proceedings of the 2012 ACM Workshop on Privacy in the Electronic Society*, WPES '12, pages 13–18, New York, NY, USA, 2012. ACM.
[35] E. Zagheni, I. Weber, and K. Gummadi. Leveraging facebook's advertising platform to monitor stocks of migrants. *Population and Development Review*, 2017.

Table 13: Top 20 advertisers who micro-target.

Name	Nb Attr.	Attributes
Google	85	Harvard Business Review-I, Graduation-I, Online shopping-I, Employment-I, Data analysis-I, TechCrunch-I, Mashable-I, Education-I, Cover letter-I, Televisions-I, Kickstarter-I, Master's degree-PD, Digital media-I, Squarespace-I, Greenpeace-I, Business school-I, Tourism-I, Mobile phones-I, Google-I, Business and industry-I, Indeed.com-I, Motorola-B, WordPress-I, Educational technology-I, Politics and social issues-I, Recruitment-I, Zalando-I, Bachelor's degree-PD, Society-I, Instagram-I, Hacker News-I, Sales-I, English (US)-PD, Startup company-I, Drop shipping-I, IPhone-I, Technology early adopters-B, Employment website-I, Entrepreneurship-I, Charitable organisation-I, Facebook Page Admins-B, Android (operating system)-I, PayPal-I, Magento-I, Smartphones-I, Rsum-I, Start-Up-I, High school-I, Udemy-I, Job-I, Postgraduate education-I, Digital data-I, Food and drink-I, Aviva (empresa)-I, Job hunting-I, US politics (very liberal)-D, Business Insider-I, Personal development-I, Facebook-I, Higher education-I, Learning-I, Software engineering-I, Forbes-I, Technology late adopters-B, Socit nationale des chemins de fer franais-I, Advertising-I, Software development-I, Student-I, Women's rights-I, Power Editor-I, Airbnb-I, Kiehl's-I, Computer science-I, Public university-I, Glassdoor-I, Online advertising-I, Marketing-I, Udacity-I, Job interview-I, Married-PD, Family-I, Codecademy-I, Coursera-I, Shopify-I, AliExpress-I
Indiatimes	34	Physical exercise-I, Friendship-I, Running-I, Mobile phones-I, Bollywood movies-I, Stand-up comedy-I, Films-I, TV reality shows-I, Dance-I, Sports and outdoors-I, Entrepreneurship-I, Selfie-I, Viral video-I, BuzzFeed-I, Movies-I, Photography-I, Facebook-I, India-I, Humour-I, YouTube-I, Aishwarya Rai Bachchan-I, Sports-I, Community and Social Services-D, Technology-I, Fitness and wellness-I, Motivation-I, Entertainment-I, Finance-I, Parents (All)-D, Married-PD, Family-I, Love-I, Travel-I, Bollywood films-I
Udemy	30	Web development-I, Python (programming language)-I, Machine learning-I, Public speaking-I, C++-I, Spirituality-I, Reiki-I, Adobe Illustrator-I, English (UK)-PD, English (US)-PD, Analytics-I, Writing-I, Web design-I, Buddhism-I, Software-I, Adobe After Effects-I, Drawing-I, Personal development-I, Tennis-I, Computer programming-I, Music-I, Data science-I, Digital marketing-I, Big data-I, Online advertising-I, Mindfulness-I, Guitar-I, Udacity-I, Software developer-I, Programming language-I
Great Big Story	29	The Nightmare Before Christmas-I, Thai cuisine-I, Italy-I, Asia-I, Tourism-I, Science fiction movies-I, Frequent International Travelers-B, WhatsApp-I, I fucking love science-I, Giant panda-I, Member of a family-based household-D, Slate-I, Nature-I, Adventure travel-I, The New York Times-I, Food & Wine-I, Architecture-I, Documentary movies-I, Recipes-I, Airbnb-I, Food Network-I, Cheese-I, Star Wars-I, Humans of New York-I, Family-I, Adventure-I, Travel-I, Baking-I, TED (conference)-I
Quartz	28	Happiness-I, Emotion-I, Personal finance-I, English (US)-PD, Startup company-I, Cloud computing-I, Member of a family-based household-D, History-I, Entrepreneurship-I, Dating-I, Information technology-I, Culinary art-I, Friends-I, Psychology-I, Family and relationships-I, Knowledge-I, Delta Air Lines-I, Food and drink-I, Computers-I, Leadership-I, YouTube-I, Business-I, Computer science-I, Healthcare and Medical-D, Career-I, Steve Jobs-I, Love-I, quora-I
Udacity India	26	Web development-I, Artificial neural network-I, Data analysis-I, Artificial intelligence-I, Kickstarter-I, Machine learning-I, Training-I, English (US)-PD, Member of a family-based household-D, Linux-I, App Store (iOS)-I, Android (operating system)-I, Social media-I, Mercedes-Benz-I, Udemy-I, Software engineering-I, Computer programming-I, Data science-I, Software development-I, Business-I, Computer science-I, Statistics-I, Big data-I, Technology-I, Software developer-I, Programming language-I
Humble Bundle	25	Video games-I, Smartphones and tablets-B, Machine learning-I, Humble Bundle-I, TripAdvisor-I, Steam-I, Mass Effect-I, C++-I, English (UK)-PD, English (US)-PD, Cloud computing-I, Action games-I, Puzzle video games-I, First-person shooter games-I, Software-I, Fallout 2-I, Computer programming-I, Left 4 Dead 2-I, Gaming computer-I, Computer science-I, Game of Thrones-I, Shooter games-I, Horror movies-I, Robot-I, Bitcoin-I
Kialo	23	Amnesty International-I, The New Yorker-I, TechCrunch-I, Environmentally friendly-I, The Economist-I, Sustainability-I, English (US)-PD, Startup company-I, Tech News-I, NASA-I, Philosophy-I, North Korea-I, Feminism-I, 20th-century philosophy-I, Conselho da Europa-I, Geek-I, The New York Times-I, UNICEF-I, Politics-I, Organic food-I, Atheism-I, quora-I, Religion-I
Samsung	23	Online shopping-I, Mountains-I, Games-I, Samsung-I, Mobile phones-I, Motorola-B, (A) Affinity for High Value Goods - India-B, Netflix-I, Watch-I, Technology early adopters-B, Sports and outdoors-I, Smartphones-I, All iOS devices-B, Comedy movies-I, Fashion design-I, FC Bayern Munich-I, Cycling-I, Sports-I, Technology-I, Parents (All)-D, Pets-I, Family-I, All Android devices-B
adidas	23	Online shopping-I, Gareth Bale-I, Physical exercise-I, uber-I, Google Play-I, Adidas-I, Sports and outdoors-I, Shopping and fashion-I, Complex (magazine)-I, Association football (Soccer)-I, Paris Saint-Germain Handball-I, Physical fitness-I, Pop music-I, UEFA Champions League-I, School-I, Tennis-I, Music-I, Sports-I, Basketball-I, Marathons-I, Fitness and wellness-I, Swimming-I, Reebok-I
Intel Developer Zone	22	Video games-I, Python (programming language)-I, Artificial intelligence-I, Machine learning-I, C++-I, Android (operating system)-I, Analytics-I, Electronics-I, Intel-I, Software-I, Computers-I, Software engineering-I, Computer programming-I, Software development-I, Computer science-I, Statistics-I, Technology-I, Application software-I, Healthcare and Medical-D, Software developer-I, Robot-I, Stack Overflow-I
MensXP	21	Engineering-I, Online shopping-I, Physical exercise-I, Virat Kohli-I, Star Plus-I, Bollywood movies-I, Cricket-I, Shoes-I, Member of a family-based household-D, Android (operating system)-I, BuzzFeed-I, Movies-I, Live events-I, Marvel Comics-I, Deepika Padukone-I, Student-I, Indian Premier League-I, Entertainment-I, Love-I, India national cricket team-I, Single-PD
Forbes	20	Harvard Business Review-I, Fortune (magazine)-I, TripAdvisor-I, Air travel-I, President of the United States-I, Business and industry-I, Indeed.com-I, Power (social and political)-I, Politics and social issues-I, Ecotourism-I, English (US)-PD, Culture-I, Real estate-I, Higher education-I, The New York Times-I, Computer science-I, Politics-I, Technology-I, Cornell University-I, Travel-I
Crossover	19	Python (programming language)-I, Cascading Style Sheets-I, Database-I, English (UK)-PD, Bachelor's degree-PD, Sales-I, English (US)-PD, Cloud computing-I, Member of a family-based household-D, Architecture and Engineering-D, JavaScript-I, Software engineering-I, Java (programming language)-I, Computer science-I, Big data-I, English language-I, Salesforce.com-I, Finance-I, Software developer-I
PayPal	19	Harvard Business Review-I, Online shopping-I, TechCrunch-I, (A+B) Affinity for Mid-High Value Goods - India-B, Fossil-I, Paytm-I, TripAdvisor-I, Facebook for Business-I, Entrepreneurship-I, PayPal-I, Tablet computers-I, Movies-I, Udemy-I, Amazon.com-I, Business Insider-I, Business-I, Personal computer-I, All frequent travelers-B, quora-I
Hotstar	19	Luxury goods-I, Bengali language-I, Animated movies-I, Kolkata Knight Riders-I, all india bakchod-I, Virat Kohli-I, Cricket-I, TV reality shows-I, Delhi Daredevils-I, YourStory-I, Marvel Cinematic Universe-I, Royal Challengers Bangalore-I, Iron Man-I, Academy Awards-I, Community and Social Services-D, Technology-I, Kolkata-I, Expats (India)-B, Telugu-PD
Google AdWords	19	Kickstarter-I, Digital media-I, Web search engine-I, Facebook for Business-I, Instagram-I, Startup company-I, Entrepreneurship-I, Information technology-I, Facebook Page Admins-B, Search engine optimization-I, Asana-I, Restaurants-I, Digital marketing-I, Power Editor-I, YouTube-I, Life, Physical, and Social Science-D, Online advertising-I, Marketing-I, Coursera-I
BARMER	18	Rock music-I, Tea-I, Literature-I, Meme-I, Bachelor's degree-PD, English (US)-PD, Member of a family-based household-D, 9GAG-I, Association football (Soccer)-I, Culture-I, Ozzy Osbourne-I, Computer programming-I, Music-I, Student-I, Recipes-I, Sports-I, Writer-I, Family-I
BookMad	18	Reading-I, Online shopping-I, United States-I, Poetry-I, Christmas-I, Literature-I, Books-I, English (US)-PD, History-I, Philosophy-I, College-I, Higher education-I, Student-I, Sports-I, Technology-I, Hiking-I, Family-I, God-I
Airbnb	18	Design-I, TripAdvisor-I, uber-I, Portuguese (Brazil)-PD, Literature-I, French (France)-PD, English (UK)-PD, English (US)-PD, CNN-I, Culture-I, Outdoor recreation-I, Nature-I, German-PD, Airbnb-I, Married-PD, All frequent travelers-B, Spanish (Spain)-PD, Travel-I

Table 14: IAB categories and 10 attributes that are used by most advertisers with the respective percentage of advertisers (Interests-I, Behaviors-B, Demographics-D, Profile Data-PD)

IAB Category	Nb. Adv.	Nb. Attr.	Attributes
News and Politics	799	609	English (US)-PD (9.89%), Entertainment-I (4.88%), Travel-I (4.76%), French (France)-PD (4.63%), Technology-I (4.26%), Member of a family-based household-D (3.88%), Food and drink-I (3.38%), Music-I (2.75%), Food-I (2.13%), Netflix-I (2.13%)
Technology and Computing	727	453	English (US)-PD (9.35%), Technology-I (6.74%), Member of a family-based household-D (6.19%), Software engineering-I (5.23%), Video games-I (3.03%), Computer science-I (2.75%), Software developer-I (2.34%), Online shopping-I (2.34%), Software-I (2.20%), Computers-I (2.20%)
Shopping	695	378	Online shopping-I (8.06%), Shopping and fashion-I (7.63%), English (US)-PD (4.75%), Food and drink-I (4.32%), Travel-I (3.60%), Member of a family-based household-D (3.60%), French (France)-PD (3.45%), Cooking-I (3.45%), Video games-I (2.88%), Technology-I (2.73%)
Community Organization	541	360	English (US)-PD (8.32%), French (France)-PD (3.33%), Member of a family-based household-D (3.33%), Online shopping-I (2.96%), Travel-I (2.96%), Kickstarter-I (2.77%), Cooking-I (2.77%), Music-I (2.03%), Entrepreneurship-I (1.66%), Business-I (1.29%)
Style & Fashion	807	360	Shopping and fashion-I (24.16%), Online shopping-I (8.67%), English (US)-PD (5.95%), Member of a family-based household-D (5.45%), Travel-I (4.83%), French (France)-PD (3.84%), Entertainment-I (3.35%), Beauty-I (3.35%), Fitness and wellness-I (2.73%), Shopping-I (2.73%)
Food and Drink	983	345	Food and drink-I (18.51%), Cooking-I (5.70%), Food-I (4.27%), Recipes-I (3.97%), Entertainment-I (3.66%), Beer-I (2.95%), Alcoholic beverages-I (2.85%), Coffee-I (2.64%), Chocolate-I (2.64%), Veganism-I (2.54%)
Education	485	314	English (US)-PD (7.63%), Education-I (5.77%), Higher education-I (4.12%), Business-I (3.92%), Bachelor's degree-PD (3.92%), Member of a family-based household-D (3.51%), French (France)-PD (3.51%), Software engineering-I (2.68%), Marketing-I (2.68%), Big data-I (2.47%)
Travel	567	244	Travel-I (34.57%), All frequent travelers-B (10.93%), English (US)-PD (6.70%), Food and drink-I (4.94%), French (France)-PD (4.94%), Frequent International Travelers-B (4.41%), Sports and outdoors-I (3.88%), Nature-I (3.53%), Tourism-I (3.35%), Member of a family-based household-D (3.00%)
Business and Finance	207	211	English (US)-PD (8.21%), Finance-I (5.31%), Member of a family-based household-D (4.83%), Bachelor's degree-PD (3.86%), Online shopping-I (3.86%), French (France)-PD (3.86%), Entrepreneurship-I (3.86%), Expats (India)-B (3.38%), Sports-I (3.38%), German-PD (3.38%)
Television	172	197	Game of Thrones-I (7.56%), English (US)-PD (6.98%), Entertainment-I (6.40%), Travel-I (5.81%), Food and drink-I (5.23%), Netflix-I (4.65%), Association football (Soccer)-I (4.07%), CNN-I (2.33%), BuzzFeed-I (2.33%), Video games-I (2.33%)
Healthy Living	387	196	Fitness and wellness-I (15.25%), Beauty-I (10.08%), English (US)-PD (8.27%), Sports-I (5.68%), Online shopping-I (5.43%), Shopping and fashion-I (5.17%), French (France)-PD (3.88%), Physical exercise-I (3.36%), Yoga-I (2.84%), Sports and outdoors-I (2.84%)
Home & Garden	227	163	Member of a family-based household-D (7.05%), Design-I (7.05%), Interior design-I (7.05%), Home and garden-I (5.73%), Online shopping-I (5.29%), French (France)-PD (4.85%), Food and drink-I (4.41%), English (US)-PD (4.41%), Technology-I (3.96%), Cooking-I (3.52%)
Events and Attractions	173	160	Music-I (6.94%), Entertainment-I (6.94%), Electronic music-I (4.05%), Resident Advisor-I (3.47%), Technology-I (2.89%), Arts and music-I (2.89%), Food and drink-I (2.89%), Sports and outdoors-I (2.31%), Shopping and fashion-I (2.31%), Hip hop music-I (2.31%)
Books and Literature	130	156	English (US)-PD (7.69%), Food and drink-I (6.15%), French (France)-PD (6.15%), Entertainment-I (4.62%), Shopping and fashion-I (3.85%), Hollywood-I (3.08%), Reading-I (3.08%), Sports-I (3.08%), Books-I (3.08%), Travel-I (3.08%)
Public Figure	226	146	Entertainment-I (6.19%), Member of a family-based household-D (5.31%), English (US)-PD (5.31%), Travel-I (3.54%), Entrepreneurship-I (3.54%), Sports-I (3.10%), French (France)-PD (3.10%), Online advertising-I (2.21%), Fitness and wellness-I (1.77%), Technology-I (1.77%)
Music and Audio	235	142	Music-I (21.28%), Entertainment-I (8.94%), Rock music-I (3.83%), Arts and music-I (2.98%), Spotify-I (2.13%), Electronic music-I (2.13%), French (France)-PD (2.13%), Pop music-I (1.70%), Member of a family-based household-D (1.70%), Concerts-I (1.70%)
Hobbies & Interests	186	139	Photography-I (8.60%), Travel-I (5.91%), Family-I (5.38%), English (US)-PD (5.38%), Video games-I (4.84%), Member of a family-based household-D (4.30%), Online shopping-I (4.30%), French (France)-PD (4.30%), Games-I (3.76%), Music-I (3.23%)
Movies	175	134	Movies-I (21.71%), Comedy movies-I (13.71%), Entertainment-I (10.86%), Action movies-I (5.71%), Cannes Film Festival-I (3.43%), Netflix-I (3.43%), Video games-I (2.86%), Independent film-I (2.86%), Thriller movies-I (2.29%), 9GAG-I (2.29%)
Sports	287	129	Sports-I (12.20%), Association football (Soccer)-I (8.01%), Sports and outdoors-I (6.97%), Skiing-I (4.53%), Fitness and wellness-I (3.83%), CrossFit-I (3.48%), Tennis-I (3.14%), Travel-I (3.14%), Marathons-I (2.44%), Physical fitness-I (2.44%)
Non-Business Places	136	128	Travel-I (14.71%), Food and drink-I (4.41%), Business-I (3.68%), Music-I (3.68%), Sports and outdoors-I (3.68%), Bachelor's degree-PD (2.94%), Mountains-I (2.94%), Finance-I (2.21%), Skiing-I (2.21%), Education-I (2.21%)
Video Gaming	108	108	Video games-I (39.81%), Games-I (7.41%), First-person shooter games-I (6.48%), League of Legends-I (6.48%), French (France)-PD (5.56%), English (US)-PD (4.63%), Association football (Soccer)-I (3.70%), Action games-I (3.70%), Technology-I (2.78%), Online games-I (2.78%)
Automotive	134	101	Automobiles-I (25.37%), Vehicles-I (9.70%), Sports-I (8.21%), Technology-I (7.46%), Travel-I (5.97%), Music-I (4.48%), Family-I (3.73%), Auto racing-I (3.73%), Shopping and fashion-I (3.73%), All frequent travelers-B (3.73%)
Fine Art	146	98	Music-I (9.59%), Entertainment-I (9.59%), Arts and music-I (8.22%), English (US)-PD (6.85%), Photography-I (3.42%), Artist-I (2.74%), Travel-I (2.74%), Dance-I (2.05%), Classical music-I (2.05%), Movies-I (2.05%)
Real Estate	107	77	Real estate-I (26.17%), Business-I (6.54%), English (US)-PD (4.67%), Co-Founder/CEO-PD (3.74%), Apartment-I (3.74%), Cornell University-I (2.80%), Engaged-PD (2.80%), Travel-I (2.80%), Investment-I (2.80%), IPhone-I (2.80%)
Other Media	75	73	Entertainment-I (5.33%), English (US)-PD (5.33%), Travel-I (5.33%), Music-I (4.00%), Hip hop music-I (4.00%), Rock music-I (4.00%), Association football (Soccer)-I (2.67%), Resident Advisor-I (2.67%), Live events-I (2.67%), Food and drink-I (2.67%)
Medical Health	88	65	Healthcare and Medical-D (11.36%), English (US)-PD (9.09%), French (France)-PD (7.95%), Fitness and wellness-I (6.82%), Married-PD (4.55%), Beauty-I (4.55%), Sports-I (3.41%), Shopping and fashion-I (3.41%), Family-I (2.27%), Sports and outdoors-I (2.27%)
Legal	17	21	Business and industry-I (11.76%), Automobiles-I (5.88%), Law-I (5.88%), Returned from trip 2 weeks ago-B (5.88%), Software developer-I (5.88%), University of Maryland, College Park-PD (5.88%), Audi-I (5.88%), Employment-I (5.88%), Justice-I (5.88%), Tourism-I (5.88%)
Pets	37	18	Dogs-I (35.14%), Pets-I (18.92%), Cats-I (13.51%), Member of a family-based household-D (8.11%), English (US)-PD (5.41%), French (France)-PD (5.41%), Automobiles-I (2.70%), Clothing-I (2.70%), Gardening-I (2.70%), Online shopping-I (2.70%)
Religion and Spirituality	13	17	Christianity-I (15.38%), English (US)-PD (15.38%), Dudeism-I (7.69%), Scientist-I (7.69%), Israel-I (7.69%), Hillsong Worship-I (7.69%), Entertainment-I (7.69%), Evangelist Daniel Kolenda-I (7.69%), The Big Lebowski-I (7.69%), Buddhism-I (7.69%)
Career	4	5	Python (programming language)-I (25.00%), Higher education-I (25.00%), Doctor of Philosophy-I (25.00%), Indian Institute of Technology Joint Entrance Examination-I (25.00%), Graduate school-I (25.00%)

Table 15: IAB categories and 10 attributes that are used by most advertisers with the respective percentage of advertisers (Interests-I, Behaviors-B, Demographics-D, Profile Data-PD) for ADANALYST-BRAZIL.

IAB Category	**Nb Advertisers**	**Nb Attributes**	**Attributes**
News and Politics	560	612	Entertainment-I (5.00%), Business-I (3.75%), Portuguese (Brazil)-PD (3.57%), Reading-I (3.21%), Education-I (2.86%), Travel-I (2.68%), Association football (Soccer)-I (2.68%), Books-I (2.68%), Technology-I (2.50%), Politics and social matters-I (2.50%)
Education	662	490	Education-I (13.14%), Higher education-I (6.04%), Technology-I (5.44%), Business-I (4.38%), English language-I (3.63%), Portuguese (Brazil)-PD (2.87%), Bachelor's degree-PD (2.42%), Exame Nacional do Ensino Mdio-I (2.11%), Entrepreneurship-I (2.11%), English (US)-PD (1.96%)
Technology and Computing	424	408	Technology-I (10.14%), Portuguese (Brazil)-PD (4.95%), English (US)-PD (4.48%), Business-I (4.01%), Entrepreneurship-I (3.07%), Software engineering-I (2.59%), Information technology-I (2.59%), Cloud computing-I (2.36%), Games-I (2.12%), Digital marketing-I (2.12%)
Shopping	474	350	Shopping and fashion-I (9.49%), Online shopping-I (7.59%), Entertainment-I (4.85%), Shopping-I (4.01%), Food and drink-I (3.80%), Sports and outdoors-I (3.80%), Technology-I (2.53%), Games-I (2.53%), Portuguese (Brazil)-PD (2.11%), Sports-I (2.11%)
Food and Drink	533	316	Food and drink-I (10.69%), Entertainment-I (9.76%), Music-I (6.57%), Beer-I (6.19%), Food-I (4.32%), Rock music-I (3.56%), Restaurants-I (3.56%), Chocolate-I (3.19%), Association football (Soccer)-I (2.81%), Fast food-I (2.63%)
Community Organization	310	292	English (US)-PD (4.84%), Portuguese (Brazil)-PD (4.52%), Online shopping-I (2.90%), Education-I (2.58%), Entertainment-I (2.26%), Sustainability-I (1.94%), Music-I (1.94%), Kickstarter-I (1.94%), Travel-I (1.61%), Home and garden-I (1.61%)
Music and Audio	462	289	Music-I (12.12%), Entertainment-I (6.49%), Rock music-I (5.84%), Electronic music-I (3.25%), Spotify-I (3.03%), Heavy metal music-I (2.60%), Msica popular brasileira-I (2.38%), alok-I (2.38%), Arts and music-I (1.73%), Blues music-I (1.73%)
Style & Fashion	364	271	Shopping and fashion-I (21.70%), Online shopping-I (10.99%), Sports and outdoors-I (3.02%), Beauty-I (3.02%), Shopping-I (2.75%), Shoes-I (2.47%), Portuguese (Brazil)-PD (2.47%), Fashion accessories-I (2.20%), Association football (Soccer)-I (1.92%), Netflix-I (1.92%)
Public Figure	267	239	Business-I (4.49%), Education-I (4.12%), Portuguese (Brazil)-PD (3.37%), Entertainment-I (3.37%), Entrepreneurship-I (3.37%), Digital marketing-I (3.00%), Stand-up comedy-I (2.62%), Politics-I (2.62%), Reading-I (1.87%), Humour-I (1.87%)
Television	102	206	Entertainment-I (9.80%), Netflix-I (7.84%), Sports-I (4.90%), Association football (Soccer)-I (3.92%), Game of Thrones-I (3.92%), Music-I (3.92%), Family Guy-I (3.92%), porta dos fundos-I (2.94%), Action movies-I (2.94%), HBO-I (2.94%)
Movies	115	178	Movies-I (31.30%), Entertainment-I (16.52%), Action movies-I (12.17%), Comedy movies-I (8.70%), Netflix-I (6.96%), Marvel Comics-I (4.35%), Star Wars-I (4.35%), Film festival-I (3.48%), Filmmaking-I (3.48%), Animated movies-I (3.48%)
Fine Art	161	162	Entertainment-I (6.83%), Music-I (4.97%), Photography-I (4.35%), Arts and music-I (3.73%), Culture-I (3.11%), English (US)-PD (3.11%), Portuguese (Brazil)-PD (2.48%), Theatre-I (2.48%), alok-I (2.48%), Live events-I (1.86%)
Business and Finance	123	155	Business-I (8.94%), Finance-I (8.13%), Portuguese (Brazil)-PD (7.32%), Investment-I (7.32%), Vehicles-I (4.88%), Bitcoin-I (4.88%), Personal finance-I (4.88%), English (US)-PD (4.88%), Travel-I (4.07%), Online shopping-I (3.25%)
Travel	187	147	Travel-I (34.22%), Tourism-I (11.23%), Portuguese (Brazil)-PD (7.49%), Nature-I (5.88%), Entertainment-I (3.74%), Air travel-I (2.67%), Online shopping-I (2.14%), Food and drink-I (2.14%), German-PD (2.14%), English (US)-PD (2.14%)
Events and Attractions	142	138	Music-I (11.27%), Entertainment-I (9.86%), Rock music-I (7.04%), Pop music-I (4.23%), Photography-I (4.23%), Technology-I (2.82%), Arts and music-I (2.82%), Dance-I (2.11%), Video games-I (2.11%), Live events-I (2.11%)
Sports	154	137	Association football (Soccer)-I (9.74%), Sports-I (9.09%), Sports and outdoors-I (9.09%), Auto racing-I (5.19%), Video games-I (3.90%), League of Legends-I (3.25%), Tennis-I (3.25%), Physical exercise-I (2.60%), Martial arts-I (2.60%), Portuguese (Brazil)-PD (2.60%)
Healthy Living	169	130	Beauty-I (9.47%), Cosmetics-I (7.69%), Health and wellness-I (4.73%), Physical exercise-I (4.73%), Online shopping-I (4.73%), Sports and outdoors-I (4.73%), Shopping and fashion-I (4.73%), Sports-I (4.14%), Fitness and wellness-I (2.96%), Aesthetics-I (2.96%)
Books and Literature	81	125	Reading-I (17.28%), Books-I (11.11%), Literature-I (7.41%), Music-I (6.17%), Education-I (4.94%), Technology-I (3.70%), Arts and music-I (3.70%), Online shopping-I (3.70%), Entertainment-I (3.70%), English (US)-PD (3.70%)
Non-Business Places	113	123	Entertainment-I (6.19%), Education-I (5.31%), Business-I (5.31%), Higher education-I (5.31%), Technology-I (3.54%), Food-I (3.54%), House-I (3.54%), Shopping and fashion-I (3.54%), Travel-I (3.54%), Online shopping-I (2.65%)
Home & Garden	137	120	Home and garden-I (9.49%), Online shopping-I (5.11%), Interior design-I (5.11%), Design-I (4.38%), Family-I (3.65%), House-I (3.65%), Architecture-I (2.92%), English (US)-PD (2.92%), Married-PD (2.92%), Luxury goods-I (2.19%)
Video Gaming	131	114	Portuguese (Brazil)-PD (20.61%), Video games-I (12.98%), Games-I (11.45%), League of Legends-I (8.40%), PlayStation 4-I (7.63%), First-person shooter games-I (6.87%), Steam (software)-I (5.34%), Online games-I (3.82%), Game consoles-I (3.82%), Gamer-I (3.82%)
Hobbies & Interests	148	113	Photography-I (9.46%), Role-playing games-I (5.41%), Video games-I (4.73%), Music-I (4.73%), Portuguese (Brazil)-PD (4.05%), Games-I (4.05%), Online shopping-I (3.38%), English (US)-PD (3.38%), Technology-I (2.70%), Massively multiplayer online role-playing games-I (2.70%)
Automotive	107	100	Automobiles-I (22.43%), Luxury goods-I (8.41%), Auto racing-I (8.41%), Travel-I (8.41%), Vehicles-I (7.48%), Technology-I (5.61%), Motorcycles-I (5.61%), Automotive industry-I (5.61%), Cars (film)-I (5.61%), Sports-I (4.67%)
Medical Health	95	74	Family-I (8.42%), Medicine-I (7.37%), Health and wellness-I (6.32%), Psychology-I (6.32%), Entertainment-I (5.26%), Beauty-I (5.26%), Physician-I (4.21%), Medical school-I (3.16%), Happiness-I (3.16%), Motherhood-I (3.16%)
Real Estate	99	48	Real estate-I (23.23%), House-I (15.15%), Business-I (10.10%), Family-I (8.08%), Luxury goods-I (6.06%), Apartment-I (6.06%), Married-PD (4.04%), Shopping and fashion-I (3.03%), Master's degree-PD (2.02%), Single-PD (2.02%)
Legal	20	19	Lawyer-I (20.00%), Business-I (10.00%), Conselho Federal da OAB-I (10.00%), Portuguese (Brazil)-PD (10.00%), Education-I (5.00%), Natural environment-I (5.00%), Politics and social matters-I (5.00%), Technology-I (5.00%), Justia-I (5.00%), Sports and outdoors-I (5.00%)
Pets	28	13	Pets-I (50.00%), Dogs-I (25.00%), Online shopping-I (10.71%), Cats-I (10.71%), Pet store-I (7.14%), SK Gaming-I (3.57%), Golden Retriever-I (3.57%), English (US)-PD (3.57%), Royal Canin-I (3.57%), Fnatic-I (3.57%)
Religion and Spirituality	10	11	Bible-I (10.00%), Travel-I (10.00%), Education-I (10.00%), Umbanda-I (10.00%), Books-I (10.00%), Toys-I (10.00%), Igreja Catlica-I (10.00%), Portuguese (Brazil)-PD (10.00%), Meditation-I (10.00%), Catechism-I (10.00%)
Career	1	1	Internship-I (100.00%)