

Facebook Assignment Report

Joseph Yu
jyu425@gatech.edu

1 FACEBOOK ASSIGNMENT REPORT

1.1 Social Media Platform

I chose the social media platform, Facebook, and downloaded a JSON containing my last 3 years of data of all advertisers that use my information or my activity.

1.2 Number of Original Advertisers

The total number of original advertisers prior to sampling is 1557. This was found from the `advertisers_using_your_activity_or_information.json` file downloaded from my Facebook data.

1.3 Sample Number of Advertisers

I sampled 180 advertisers with 20 advertisers for each of the 9 categories I chose. The random generation code is attached at the end of the report PDF in Section 2. I first labeled all the advertisers within the last 3 years with a category and then performed the random sampling to ensure each of the categories was able to contain 20 advertisers.

1.4 Categories

I identified 9 total categories among all advertisers.

1. Technology (e.g. ScaleAI, Stripe)
2. Food (e.g. Doordash, Ubereats)
3. Travel (e.g. Delta Airlines, AAA)
4. Entertainment (e.g. Netflix, NBA)
5. Education (e.g. Georgia Tech)
6. Financial Services (e.g. Chase)
7. Fashion (e.g. Shein)
8. Health (e.g. Kaiser)
9. Retail (e.g. Amazon, Wish)

1.5 Data Buckets

I then classified each advertiser within each category into one of the following data buckets.

1. **Relevant** - Advertisers that I interact with regularly and am interested in their product.
2. **Somewhat Relevant** - Advertisers with which I do not interact regularly but have clicked on their ads or is tangentially related to my interests.
3. **Not relevant** - Advertisers that I have never heard of and are not related to my interests.

1.6 Sankeymatic.com Script

```
All Advertisers [20] Technology
All Advertisers [20] Food
All Advertisers [20] Retail
All Advertisers [20] Travel
All Advertisers [20] Entertainment
All Advertisers [20] Financial Services
All Advertisers [20] Fashion
All Advertisers [20] Health
All Advertisers [20] Education
```

```
Technology [8] Relevant #2ca02c
Technology [3] Somewhat Relevant #ff7f0e
Technology [9] Not Relevant #d62728
```

```
Food [7] Relevant #2ca02c
Food [4] Somewhat Relevant #ff7f0e
Food [9] Not Relevant #d62728
```

```
Retail [3] Relevant #2ca02c
Retail [3] Somewhat Relevant #ff7f0e
Retail [14] Not Relevant #d62728
```

```
Travel [6] Relevant #2ca02c
Travel [2] Somewhat Relevant #ff7f0e
```

Travel [12] Not Relevant #d62728

Entertainment [6] Relevant #2ca02c

Entertainment [2] Somewhat Relevant #ff7f0e

Entertainment [12] Not Relevant #d62728

Financial Services [4] Relevant #2ca02c

Financial Services [0] Somewhat Relevant #ff7f0e

Financial Services [16] Not Relevant #d62728

Fashion [4] Relevant #2ca02c

Fashion [3] Somewhat Relevant #ff7f0e

Fashion [13] Not Relevant #d62728

Health [3] Relevant #2ca02c

Health [2] Somewhat Relevant #ff7f0e

Health [15] Not Relevant #d62728

Education [4] Relevant #2ca02c

Education [2] Somewhat Relevant #ff7f0e

Education [14] Not Relevant #d62728

:All Advertisers #333

:Technology #1f77b4

:Food #17becf

:Retail #9467bd

:Travel #8c564b

:Entertainment #e377c2

:Financial Services #7f7f7f

:Fashion #bcdb22

:Health #2ca02c

:Education #ff7f0e

:Relevant #2ca02c

:Somewhat Relevant #ff7f0e

:Not Relevant #d62728

1.7 Dataflow Diagram

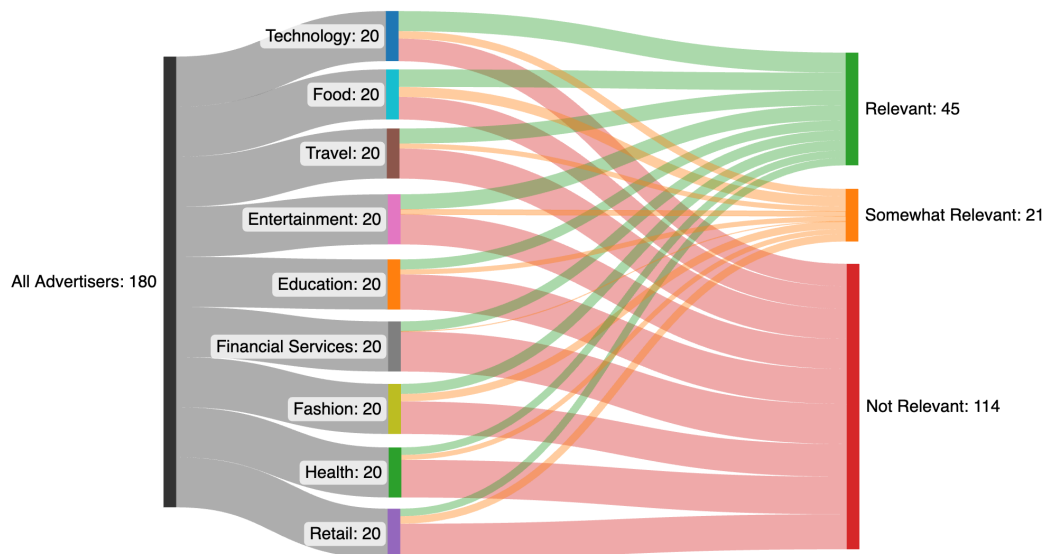


Figure 1—Bucketed Advertisers Sankey Diagram

1.8 Table: Summary Statistics

Table 1—Category Technology

Data Buckets	Count	Accuracy (% Relevant)	Rubbish (% Not Relevant)
Relevant	8		
Somewhat Relevant	3		
Not Relevant	9		
Total	20	40%	45%

Table 2—Category Food

Data Buckets	Count	Accuracy (% Relevant)	Rubbish (% Not Relevant)
Relevant	7		
Somewhat Relevant	4		
Not Relevant	9		
Total	20	35%	45%

Table 3—Category Travel

Data Buckets	Count	Accuracy (% Relevant)	Rubbish (% Not Relevant)
Relevant	6		
Somewhat Relevant	2		
Not Relevant	12		
Total	20	30%	60%

Table 4—Category Entertainment

Data Buckets	Count	Accuracy (% Relevant)	Rubbish (% Not Relevant)
Relevant	6		
Somewhat Relevant	2		
Not Relevant	12		
Total	20	30%	60%

Table 5—Category Fashion

Data Buckets	Count	Accuracy (% Relevant)	Rubbish (% Not Relevant)
Relevant	4		
Somewhat Relevant	3		
Not Relevant	13		
Total	20	20%	65%

Table 6—Category Education

Data Buckets	Count	Accuracy (% Relevant)	Rubbish (% Not Relevant)
Relevant	4		
Somewhat Relevant	2		
Not Relevant	14		
Total	20	20%	70%

Table 7—Category Financial Services

Data Buckets	Count	Accuracy (% Relevant)	Rubbish (% Not Relevant)
Relevant	4		
Somewhat Relevant	0		
Not Relevant	16		
Total	20	20%	80%

Table 8—Category Retail

Data Buckets	Count	Accuracy (% Relevant)	Rubbish (% Not Relevant)
Relevant	3		
Somewhat Relevant	3		
Not Relevant	14		
Total	20	15%	70%

Table 9—Category Health

Data Buckets	Count	Accuracy (% Relevant)	Rubbish (% Not Relevant)
Relevant	3		
Somewhat Relevant	2		
Not Relevant	15		
Total	20	15%	75%

1.9 Most and Least Accurate Categories

My most accurate category was **Technology**. My least accurate category was a tie between **Retail** and **Health**.

1.10 Table: Regulated Domain Information

Table 10—Regulated Domain Advertisers

Regulated Domain	Number of Advertisers	Advertiser Sample
Credit	6	Klarna, Chase
Education	20	University of Portland, Claremont Graduate University
Employment	1	Athlete Exchange
Housing & Public Accommodation	35	IHG Hotels Resorts, Airbnb

1.11 Reflection

- From the data collected I was able to see that the Technology and Food categories had the two highest relevant percentages at 40% and 35% respectively. In addition, the relevant and somewhat relevant data buckets for both of these categories made up of over half of the advertisers showing that for certain categories Facebook's targeted advertising is highly relevant. This was somewhat expected because my use of Facebook has mainly been interacting with posts about new restaurants or home-cooked meals as well as tech advertisements about blogs or newsletters for advancing my software knowledge. I was a bit surprised that Databricks and Stripe had appeared on my advertisers list since I had never encountered an ad by them on Facebook but they were two companies I had previously worked at. I was also not expecting the large number of advertisers in Retail and Health categories with both having a high percentage for irrelevance, 70% and 75% respectively. I rarely interact with retail or health related ads yet Facebook still has a non-negligible number of targeted advertisers in those categories.
- I was surprised by the amount of information collected by Facebook in the last 3 years. While my account has been active for a long time I rarely use it yet I still had over 1500 targeted advertisers over the last 3 years. In addition, I

was surprised to see companies that I had worked for listed in the targeted advertisers despite never having seen or interacted with their ads. It made me question what personal information Facebook was able to retrieve about me beyond just my use of their app. For example does my chat data in messenger or searches in Instagram also guide targeted advertisers on Facebook. I believe there are serious privacy concerns with one being the lack of transparency into what specific elements of my activity are being used by targeted advertisers. I also felt that I had little to no agency over my data within Facebook. While the help center allowed some control over my data I wish they gave me the option to opt-out of this tracking when I made my account. The help information additionally felt purposefully hidden. I also could not find a way to remove data shared with targeted advertisers or remove data Facebook has collected about me in order to build the list of targeted advertisers.

- This information definitely will change the way that I interact with social media. Noticing that Facebook is able to use my activity from other Meta applications such as Instagram and even external accounts including Adobe I plan to investigate into other social media accounts such as Reddit or LinkedIn and see if I can opt out of this form of tracking. In addition, most of my tracked activity data on Facebook cannot be cleared so I plan to continue to use it sparingly and be more vigilant around the advertisements and posts that I interact it. Finally, I intend to be more cautious around my usage of social media in general because while the advertiser data I retrieved from Facebook did not contain much personal information other sites may be able to extract even more personal information from my uses.

2 RANDOM SAMPLING CODE


```
In [40]: import json
        from collections import defaultdict
        import random
```

```
In [33]: with open('./data/targeted_advertisers_labeled.json', 'r', encoding='utf-8')
        advertisers_labeled = json.load(f)
```

```
In [34]: all_advertisers = advertisers_labeled['advertisers']
        category_to_advertiser = defaultdict(list)

        for advertiser in all_advertisers:
            category_to_advertiser[advertiser['category']].append(advertiser['name'])
```

```
In [41]: len(all_advertisers)
```

```
Out[41]: 1557
```

```
In [37]: category_sample_advertisers = defaultdict(list)
        for category, advertisers in category_to_advertiser.items():
            if category == "Miscellaneous":
                continue
            sample_size = min(20, len(advertisers))
            category_sample_advertisers[category] = random.sample(advertisers, sample_size)
```

```
In [38]: for category, sampled_advertisers in category_sample_advertisers.items():
        print(f"Category: {category}, Sampled Advertisers ({len(sampled_advertisers)})
```

Category: Technology, Sampled Advertisers (20): ['nanoleaf', 'norton', 'anthropic', 'razer', 'athlete exchange', 'meta for business', 'notion', 'merkle data partner', 'the rundown ai', 'corsair', 'acxiom', 'campaign adcloud', 'adparlor', 'zapier', 'angler ai', 'lexar', 'wiland data targeting', 'memoryos', 'waymo', 'wiland data strategies']

Category: Food, Sampled Advertisers (20): ['blaze pizza', 'handle the heat', 'hangukvegan', 'eatclean', 'guinness', 'factor', 'bush brothers company', 'hellofresh', 'gong cha', 'flemings prime steakhouse & wine bar', 'the hershey company', 'original kabob factory business manager', 'egghead sando cafe', 'grubhub', 'chamberlain coffee', 'savvy by shopmium', 'shef', 'dietz watson', 'dreamfarmplantlovers', 'papa johns pizza']

Category: Retail, Sampled Advertisers (20): ['dr squatch', 'living spaces', 'groupon inc', 'montreal weights', 'sculpd', 'il homestores', 'tshirt kingdom b55', 'ashley', 'toyota fan zone', 'haworth inc', 'guitar center', 'purple', 'our place', 'secretlab', 'sams club', 'wish', 'sally beauty', 'pg philippines', 'nectar sleep', 'bloomingdales']

Category: Travel, Sampled Advertisers (20): ['ski butlers', 'aaa', 'delta air lines', 'discover the palm beaches', 'drive toyota', 'ihg hotels resorts', 'drury hotels', 'tripadvisor', 'viator', 'autocamp', 'extended stay america', 'vrbo', 'the lenox hotel boston', 'lionsback resort', 'aaa western and central new york', 'aaa south jersey', 'jtb', 'airbnb', 'greenvillespartanburg international airport', 'aaa the auto club group']

Category: Entertainment, Sampled Advertisers (20): ['all day all night events', 'bleacher report', 'nba', 'the independent', 'mojo country club', 'espn', 'directv', 'reach ni', 'wmx', 'taketwo interactive', 'live nation sponsorship', 'thunderwolf media', 'walt disney world', 'beachlife festival', 'nixon america', 'flyquest', 'segerstrom center for the arts', 'contenidos el rey', 'the new york times', 'disney books']

Category: Financial Services, Sampled Advertisers (20): ['nibbles pet credit card', 'klarna', 'intuit turbotax', 'wealthy exile', 'lifevest insurance', 'credlize promotora', 'el show de andres gutierrez', 'changed app', 'visa', 'real estate investor couple', 'joy wallet finance', 'krost', 'axos bank', 'publiccom', 'chase', 'tenx', 'santander bank us', 'windfall data', 'amica insurance', 'ethos']

Category: Fashion, Sampled Advertisers (20): ['greats brand', 'protect socks', 'hockerty', 'brooks running', 'balance kimonos', 'nike mena', 'fitsyou', 'freya test account', 'saks fifth avenue', 'edikted', 'axel arigato', 'parks project', 'fear of god', 'marie nicole clothing', 'harajukustreetwear', 'flight club', 'brand collective Pty Ltd', 'stitch fix', 'tommy john', 'birdwell beach britches']

Category: Health, Sampled Advertisers (20): ['smiledirectclub', 'real chemistry', 'slimfast', 'relaxe', 'cardon', 'bsn', 'novartis clinical trials', 'cigna healthcare', 'houston methodist', 'les mills', 'muhammads boxing and martial arts center', 'dripdrop', 'st jude childrens research hospital', 'medscape', 'mycosan hong kong', 'huggies', 'endless pools', 'ymca of greater richmond', 'health union llc', 'boxing fitness']

Category: Education, Sampled Advertisers (20): ['canvas lms', 'gale media', 'codecademy', 'strayer university', 'claremont graduate university', 'tonyjeffries', 'mouratoglou tennis academy', 'aarp driver safety', 'university of portland', 'ruffalo noel levitz', 'masterclass', 'unidays', 'dk books', 'sounds true', 'coursera', 'deepstash business', 'jiu jitsu legacy', 'iep instituto europeo de posgrado', 'university of colorado boulder', 'life surge']

```
In [39]: with open('./data/category_sample_advertisers.json', 'w', encoding='utf-8')
         : json.dump(category_sample_advertisers, f, indent=4)
```