

```
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)
```