

AI, Ethics, and Society

Homework Project #1

Readings:

- Intro & Chapter 1: Weapons of Math Destruction (What is a Model?)
- M. Kosinski, D. Stillwell, T. Graepel, “Private traits and attributes are predictable from digital records of human behavior,” Proceedings of the National Academy of Sciences Apr 2013, 110 (15) 5802-5805; DOI:10.1073/pnas.1218772110, <http://www.pnas.org/content/110/15/5802>

In this assignment, you'll begin the process of discovering how data from a user's social media profile is used by various organizations. You'll accomplish this task by examining your own data profile on social media. You are allowed to use any social media platform for this assignment, but you must be able to extract similar data (Steps 1-7) that companies can use to target you. Facebook is provided as an example in the steps below (but you can choose any other social media platform as long as you can complete the steps).

- Step 1: Research how to download a copy of your personal data from your selected social media platform
 - For Facebook: Information on how to download a copy of your data can be found at: <https://www.facebook.com/help/1701730696756992>
- Step 2: Download the stored data associated with your social media profile that can be used to target you. For the purpose of this assignment, targeted information is defined as any data that is collected and can be used to profile you. Targeted information can include, for example, data you've provided or is inferred by the social network (like your interests) or list of advertisers that may have already expressed interest in you (typically based on third-party data brokers, your profiled interests, or your browsing history).
 - For Facebook: This information is found in the categories listed below on Facebook's "Download Your Information" page (two formats are available: html and json).
 1. "Ads Information"
The downloaded data includes the files: *advertisers_you've_interacted_with.html* (ads that you have interacted with) and *advertisers_using_your_activity_or_information.html* (advertisers that are using your information)
 2. "Other Logged Information"
The downloaded data includes the files: *ads_interests.html* (keywords used to target you).
- Step 3: Based on the data associated with your targeted information, categorize the data into (no less than) 5 categories and (no more than) 10 categories.
 - For Facebook: I've selected the targeted advertiser list *advertisers_who_uploaded_a_contact_list_with_your_information.html* as an illustrative example in the next set of steps
- Step 4: Create a data flow graph (e.g. using <http://sankeymatic.com/build/>) that associates your categories with three types of data buckets: Relevant, Not Relevant, Way Off. Feel free to be creative in the naming and interpretation of your buckets, but you will need to define all three data buckets.

- Step 5: Compute basic statistical measures on the data (per each data bucket): **Count, Accuracy** (= %Relevant), **and Rubbish** (%Way Off). Identify **which category was the most accurate and which was the least.**
- Step 6: Identify which data items could be associated with a regulated domain in law as defined in the lectures (**Credit, Education, Employment, Housing and 'Public Accommodation'**). For each of these regulated domains, list how many fall within each and provide a sample of the associated data items.
- Step 7: Review the results of your report, such as the dataflow diagram (Step 4) and the summary statistics table (Step 5). Reflect on the results and answer the questions below:
 - **How do you feel** about the accuracy / inaccuracy of the data collected? **Explain.**
 - **Were you surprised** or not surprised by the amount of information collected by the social media platform? **Why?**
 - **Will this information prompt any changes** with how you interact with social media? **Why?**
- Step 8: Turn in a report documenting your findings, including social media platform, number of data items, number of categories, name of categories, data buckets identified, **script/code (to create data flow graphic)**, data flow graphic, statistical measures, regulated domain/data item list. The report should be submitted in JDF format. Reports that are not neat and well organized will receive up to a 10% deduction. The file name for submission is **GTuserName_Assignment_1**, for example, **Joyner03_Assignment_1**. Below is an example report not in JDF format associated with my advertiser data for reference:

Facebook Assignment Report

George P. Burdell

george.burdell@me.gatech.edu

1 FACEBOOK ASSIGNMENT REPORT

1.1 Social Media Platform

Facebook

1.2 Number of Advertisers

1700

1.3 Categories

- Number of categories identified: 5
- Car Companies (e.g. International Autos Mercedes Benz)

- Social Impact (e.g. The National Association for the Education of Young Children)
- Shopping (e.g. Tiffany & Co.)
- Interest Groups (e.g. AARP)
- Entertainment – (e.g. Applebees’ Grill & Bar)

1.4 Data Buckets

- Yes
- No
- U Got To Be Kidding

1.5 Sankeymatic.com Script

FB Advertisers [680] Car Companies

FB Advertisers [340] Social Impact

FB Advertisers [170] Shopping

FB Advertisers [340] Interest Groups

FB Advertisers [170] Entertainment

Car Companies [85] Yes

Car Companies [595] No

Social Impact [340] Yes

Shopping [84] No

Shopping [86] U Got to be Kidding

Interest Groups [170] Yes

Interest Groups [170] No

Entertainment [85] Yes

Entertainment [85] No

1.6 Dataflow Diagram

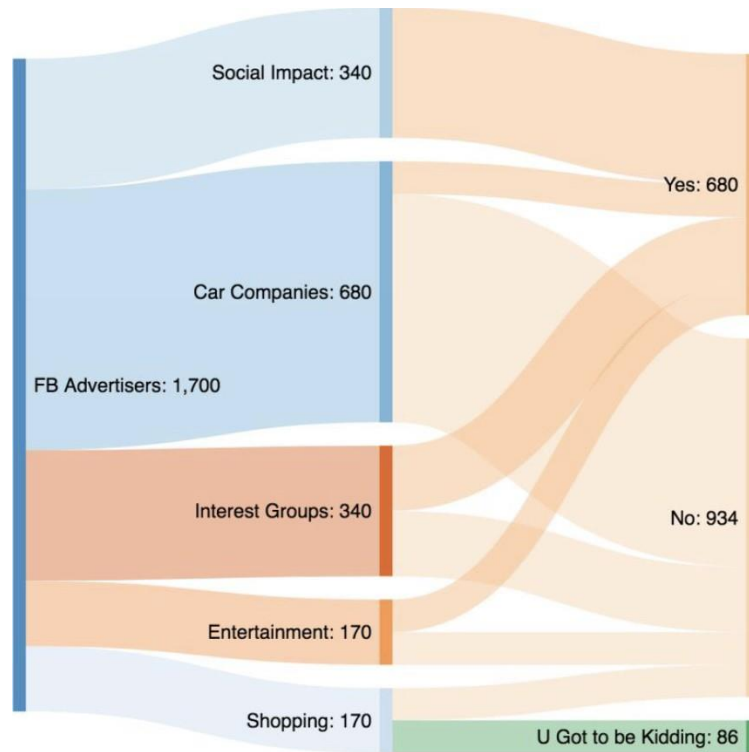


Figure 1 — Dataflow diagram

1.7 Table: Summary Statistics (Note: Partial Example for One Category)

Category	Data Bucket	Count	Accuracy	Rubbish
Shopping	U Got to be Kidding	86		
	No	84		
	Yes	0		
	Total Count	170	0%	51%

1.8 Most and Least Accurate Category

My most accurate category: Social Impact

My least accurate category (i.e. rubbish): Shopping

1.9 Table: Regulated Domain Information

Regulated Domain	Number of Items	Advertiser Sample
Credit	230	Alliant Credit Union Anchor Capital
Education	100	Baylor College of Medicine Daniels College of Business Georgia State University

Employment	0	
Housing & Public Accommodation	2	Ashton Woods Homes Echo Fine Properties

1.10 Reflection

My Facebook data had mixed accuracy results. For example, all the ads in the shopping category were rubbish or had no relevance to my interests. This was quite a relief to me, given the prevalence of targeted advertising. On the other hand, all the ads in the “social impact” category were relevant to me. I didn’t mind this since it brought awareness to many of the social impact organizations I wasn’t aware of before. I was most surprised by the amount of data Facebook had collected given my limited 5 months of use. Moving forward, I intend to limit my Facebook use, given the amount of data it collects.

2 REFERENCES