AI, Ethics, and Society

Spring 2020

Readings:

- Intro & Chapter 1: Weapons of Math Destruction (What is a Model?)
- M. Kosinski, D. Stillwell, T. Graepel, "Digital records of behavior expose personal traits," 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 basic 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. We will focus on Facebook for this assignment (see the note below if you'd like to choose another social media platform).

- Step 1: Information on how to download a coy of your data can be found at: https://www.facebook.com/help/1701730696756992

- Step 2: Download your data associated with the "Select Ads" category (two formats are available: html and json).

- Step 3: Based on the data file *advertisers_who_uploaded_a_contact_list_with_your_information.html*, categorize advertisers into (no less than) 5 categories and (no more than) 10 categories.

- Step 4: Create a data flow graph that associates your advertiser categories with three types of data buckets: Relevant, Not Relevant, Way Off. Feel free to be creative in the naming of your buckets. A great tool to use is: http://sankeymatic.com/build/

- Step 5: Compute basic statistical measures on the data (per category): count, mean, accuracy (= %Relevant), and rubbish (%Way Off). Identify which category was the most accurate and which was the least.

- Step 6: Identify which advertisers are associated with a regulated domain in law (i.e. Credit, Education, Employment, Housing and 'Public Accommodation'). For each regulated domain, list how many fall within and the associated advertiser.

- Step 7: Turn in a report documenting your findings, including number of advertisers, categories identified, script (if using sankeymatic) and data flow graphic, statistical measures, regulated domain/advertiser list. As an example, here's the report associated with my data:

Prof. Ayanna Howard

Number of Advertisers: 1700

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. Applebee's Grill & Bar)

My script on sankeymatic.com:
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 [85] No
Shopping [85] U Got to be Kidding
Interest Groups [170] Yes
Interest Groups [170] No
Entertainment [85] Yes
Entertainment [85] No

My data flow graphic:

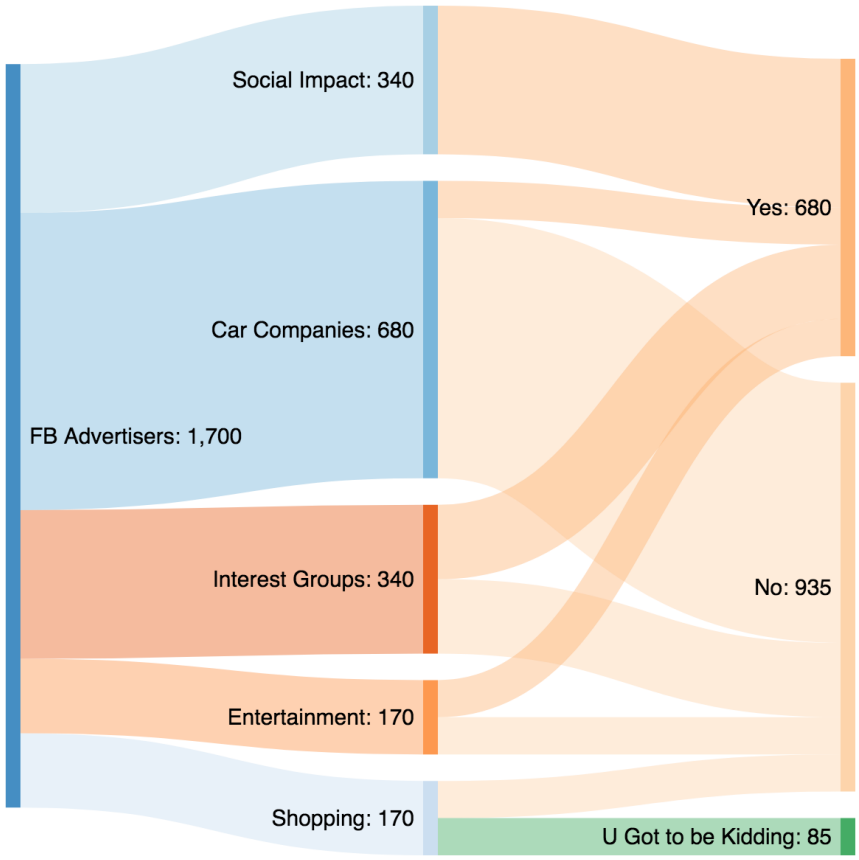

Table: Summary Statistics Count (*Partial Example for One Category*)

Variable	Ordinal Categories	Count	Mean	Accuracy	Rubbish Meter
Shopping	-1. U Got to be Kidding	85			
	0. No	85			
	1. Yes	0			
	Total	170	-0.5	0%	50%

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

Advertisers and Associated Regulated Domains in Law:
Credit (2):
Alliant Credit Union
Anchor Capital

Education (3):
Baylor College of Medicine
Daniels College of Business
Georgia State University

Employment (0)

Housing (2):
Ashton Woods Homes
Echo Fine Properties

Note: If choose to use another social media platform, you are allowed to do so for this assignment, but you must deliver similar information to the advertiser list provided by Facebook.