CS 6603: AI, Ethics, and Society
Homework Project #1

Cleo Zhang

yzhang3761@gatech.edu

- ***Abstract***—this homework starts discovering how organizations use users' data from their social media. I will examine my Facebook profile and Ads Information as the "target information" to achieve this assignment's learning goal. As required, the topics will cover the total number of data items, the number and name of categories, identified data buckets, data flow graphic and the script used, basic statistical measures, and regulated domain with data item list.

1 DATA FLOW GRAPH

In this section, we look into the data from *advertisers_using_your_activity_or_information.json.* The total number of data items is 141 over the last 3 years, and the below table lists the distribution for 8 categories.

Table 1 — Number of data items for each category

Category	Number
Online Shopping	61
SaaS platforms	23
Food	13
Education	12
Finance	10
Social Media	10
Health and Fitness	7
Real Estates	5

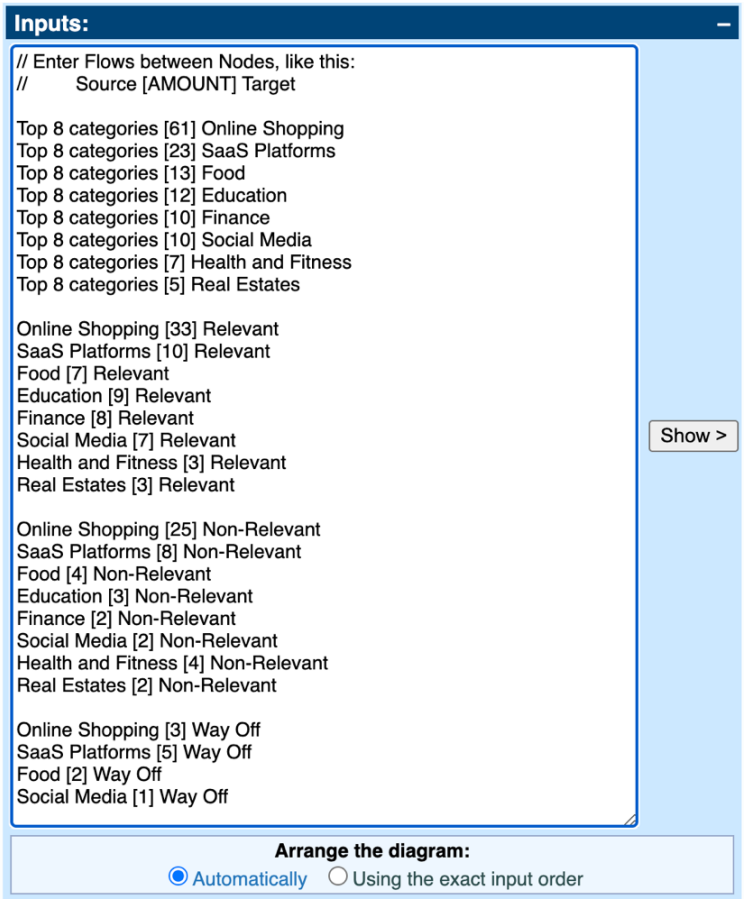

Figure 1 — Scripts for the data flow graph.

I have divided the data into 3 "data brackets" – Relevant, Non-relevant and Way-off according to my online experience and interactions. The data flow graph is generated from `http://sankeymatic.com/build/` and shown in *Figure 2*, using the script written in *Figure 1*.

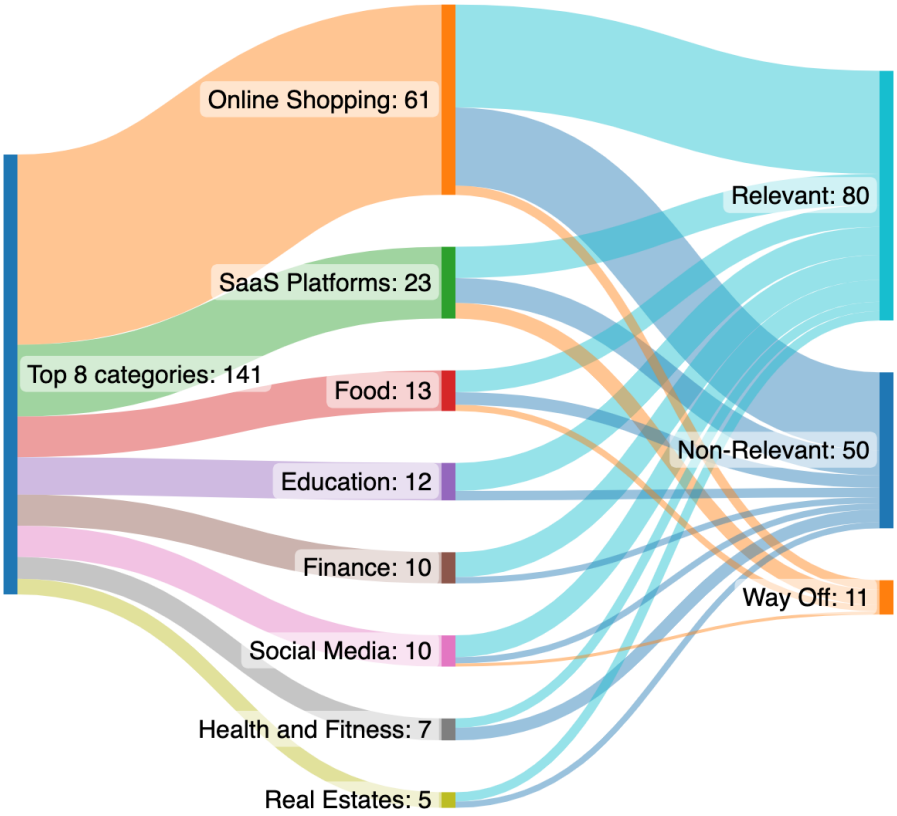

Figure 2 — Data flows from the 8 categories to relevant, non-relevant and way-off data brackets.

2 BASIC STATISTICAL MEASURES FOR EACH CATEGORY

Table 2 shows the calculation of Accuracy (= %Relevant), Inaccuracy (= %Non-relevant) and Rubbish (=% Way-off) for each category in *Table 1*. For the categories with less than 10 data items, the percentage seems too high (Accuracy for Finance is 80%) or too low (Rubbish for Real Estate is 0%). As the sample size decreases, measurement reliability needs further validation.

According to *Table 2*, the Finance category is the most accurate, while Health and Fitness have the least accuracy.

Table 2 — Basic Measurements for each category

	Count	**Accuracy**	**Inaccuracy**	**Rubbish**
Online Shopping	*61*	*54.1%*	*41%*	*4.9%*
SaaS platforms	*23*	*43.5%*	*34.8%*	*21.7%*
Food	*13*	*53.8%*	*30.8%*	*15.4%*
Education	*12*	*75%*	*25%*	*0%*
Finance	*10*	*80%*	*20%*	*0%*
Social Media	*10*	*70%*	*20%*	*10%*
Health and Fitness	*7*	*42.9%*	*57%*	*0%*
Real Estates	*5*	*60%*	*40%*	*0%*

3 IDENTIFICATION OF REGULATED DOMAIN

Data items that are associated with regulated domains defined in lectures are listed below:

- Credit: 8 items in total
    - Paypal
    - Canadian Tire Bank
    - CIBC
    - BMO Financial Group
    - TD
    - Tangerine
    - Scotiabank
    - Wealthsimple
    - RBC

- Education: 8 items in total
    - Udemy
    - Khan Academy
    - Coursera
    - Brain Station
    - The Lighthouse Labs
    - Triplebyte
    - DataCamp
    - Codeacademy
- Employment: 3 items in total
    - Toptal
    - Fiverr
    - Linkedin
- Housing and Public Accommodation: 4 items in total
    - BlueSky by Bosa Properties
    - Concord Pacific – Canada
    - Livrent
    - Airbnb