

Synthesis Activity: Data collection brainstorm

Due Thursday, 04/14/2022 at 11:59 pm (10 points)

Material borrowed from Dr. Matthew Burton

Activity overview

The goal of these activities is to learn about what and how companies are collecting data about you. Recently, many companies have started disclosing their data collection procedures and in some cases share the data they have collected with you. For this activity, you will create a Google Sheet of all the different types of data being collected about you from different sources. First you will examine private corporations; then you'll examine what University of Pittsburgh.

For this activity you'll be submitting a spreadsheet. You can use any spreadsheet editing software you like, although I strongly suggest using Google Sheets to make it easier to collaborate.

You may complete this activity alone, or in a group of up to 4.

Learning objectives

- Explore representations, structure, function, and classification/categorization of data

Part 1: Corporate data collection

Form together into small groups. The goal of this exercise is to figure out what data various corporations are collecting about you and how they represent you.

First, use your internet sleuthing skills to find out, to the best of your abilities, what data the major companies are collecting about you.

Second, create a set of categories for organizing the information you have gathered. Make a spreadsheet with companies as rows and the data point, category, and how it is collected as columns. Each row should be a datapoint. See the example below:

Company	Data Point	Category	How is it collected?
---------	------------	----------	----------------------

Stahbucks	Coffee preference	Consumer behavior	Order history
Stahbucks	Work schedule	Consumer behavior	Order time
Evil Corp	DNA	Biometric	Toothbrush sensor
Evil Corp	Fingerprint	Biometric	Touch ID Harvesting
Hogwarts	House Name	Student Identity	Sorting Hat

Using the following list as a starting point, try to find out what information each of these companies has about you. There are some links provided for some of the companies below, but you should try and find additional information if it is available. Here is a list of potential companies, but feel free to explore others!

- Google
- Facebook
- Amazon
- Netflix
- Apple
- Instagram
- Twitter
- Snapchat
- Fortnite
- Steam
- What other platforms do you use?
- For each company think about the following questions:
 - What data are they collecting about you? (Location information, topics of interest, spending habits)
 - How are they collecting this information? (Mobile app, website, ad tracking, etc.)
 - What are they doing with this information? (Selling ads, improving services, controlling your mind?)

Here are some links to a couple companies data collection information to help you get you started:

Google

Here is a link to your location history on Google: www.google.com/maps/timeline

- Are there any locations that surprise you?
- What have you learned about travel habits based on your location history?

Facebook

Here is a link to some data collection info on

Facebook: <https://www.facebook.com/help/930396167085762>

Instagram

Visit this URL for general information about their data collection: <https://help.instagram.com/181231772500920>

Here is a link specifically about your “Instagram Ad Interests”: https://www.instagram.com/accounts/access_tool/ads_interests

- Do these “Ad Interests” reflect your actual interests?
- Compare Ad Interests with someone else in your group, are they the same or different?
- Talk about how you use the platform and how the similarities or differences would emerge from your user behavioral data.

Twitter

Look at your Twitter advertising data: https://twitter.com/settings/your_twitter_data/ads

- Do these interests describe you? Yes/No?
- Does anything surprise you?

Your spreadsheet should have at least 10 companies.

Part 2: Pitt Data Collection

Get together in your group and start thinking about all the different kinds of data Pitt may be collecting about you. Start making a list of all the data Pitt could theoretically collect about you.

Use the following questions as a guide:

- What are the various kinds of data that Pitt has about you?
- How do you think Pitt collects that data?
- What do you think Pitt should do with these data?
- What do you think Pitt is currently doing with these data?
- How would you feel if Pitt was collecting all these data about you?

Create a new tab for the spreadsheet you used for part 1. Enumerate the data points, put them into some categories, and think about the following questions:

- Who is collecting the data? (Registrar, Instructors, Admissions, IT, etc. If you don’t know, provide a guess!)
- How is it collected? (It is automated or does someone create it?)
- How could it be used? (For good or bad?)

You should have at least 10 data points.

Data Point	Category	Who is collecting?	How is it collected?	How could it be used?
------------	----------	--------------------	----------------------	-----------------------

GPA	Student Assessment	Registrar?	Generated from course grades	Employers use it to evaluate applications Financial aid decisions Parent judgement
-----	--------------------	------------	------------------------------	--

Part 3: Pitt data discussion question

Data are likely gathered by many different entities on the Pitt campus (e.g. financial aid, the registrar, individual course instructors) and used separately. What are some possible outcomes if these data were combined? Put your answer into a write up. Your answer should be at least two paragraphs long.

Submitting your assignment

Submit your writeup and your spreadsheet on Canvas. If you used Google Sheets to make your spreadsheet, download your work as an xlsx file to submit it. You only need to submit one writeup and spreadsheet for each group. You will use Canvas's group submission feature to do this. First, you and your teammates should join the same group. You can find instructions for how to do that [here](#). Make sure you join one of the "Data collection brainstorm activity" groups. Then, one (and only one) of you needs to submit the writeup and spreadsheet, and your entire group will get credit.

Grading rubric

Corporate data collection: 4 points

Pitt data collection: 4 points

Implications of combining Pitt data: 2 points