Lydia Graveline
CART451
November 17, 2023

Project Description

My project involves analyzing and visualizing my personal data collected from my social platforms of choice. During the prototyping phase, my focus was primarily on delving into and comprehending the data structured, and conducting experiments with visualizations.

When you request your data from these platforms, they compile and provide you with a downloadable folder that contains a variety of information associated with your account.

- Profile Information: Details provided during account creation, such as name, email address, and phone number.
- Posts and Media: My shared photos, videos, and other content.
- Interactions: Information about my engagements on the platform, including likes, comments, and messages.
- Ad Interactions: Data pertaining to advertisements you've engaged with on the platform and custom audiences.
- Login and Account Activity: Insights into when and where you've logged into your account.
- Settings: Your account preferences and settings.

Notably, these platforms don't necessarily *want* to provide you with your data. The motivation is likely rooted in new privacy regulations and the growing user awareness about data collection practices. Regulations like the General Data Protection Regulation (GDPR) in Europe and the California Consumer Privacy Act (CCPA) in the United States have compelled companies to be more transparent about the data they collect.

This project may have it's roots in this unfinished endeavor I crafted in 2019, where scans of all the items I carried that day were transferred onto watercolor paper. Drawing further inspiration from notable works like Tracey Emin's "Everyone I Have Ever Slept With, 1963 - 1995," my interest lies in the vulnerable showcase of personal collections, defining an individual through their accumulated objects. My analysis of personal data functions as self-portraiture, and emotional self-discovery. Given the intimate nature of personal data, my project aims to uncover the ways in which I am defined through my accumulated interactions, ultimately fostering a better understanding of my (online) self.

Presenting the prototype to peers felt exposing, I had to fight the urge to say "Oh, but I'm not *actually* interested in that topic!" The data's accuracy of course is not absolute; certain topics may not align with my actual interests. Yet the fear of potential judgments based on revealed interests or advertisers is where meaning can be derived. Given the sheer amount of data I've provided these platforms over the years, perhaps they know me better than I do.

Self Portrait from 2019:

Current Project Stage

At this point I have requested just about all of the data I can. Response times varied, with some platforms providing insights within hours, while others adhered to a timeframe of up to 30 days. This phase has also involved understanding and formatting the content I've received so far.

The prototype served as a testing ground for experimenting with the acquired data, using p5.js for the visualizations and MongoDB for database storage. From the requested data, I focused on "advertisers" and "topics." Advertisers, representing companies targeting the user, were sourced from Instagram, Facebook, and Twitter. Topics, representing inferred made about my interests, were derived from Instagram, Pinterest, and Twitter.

The process also involved manual formatting decisions to ensure data compatibility. Although the interaction methods require further refinement, and there's a need to develop a compelling narrative for the project, the prototype has provided valuable insights into the potential directions and outcomes of the analysis. This iterative approach allows for ongoing adjustments and improvements, ensuring meaningful exploration of the obtained data.

Data Formatting

Because all the data was formatted differently, I had to find a way to make the data easier to work with. Ensuring uniformity across varied data formats consumed a significant portion of my time thus far. For the prototype's initial phase, only Instagram and Facebook data were utilized, given their compatibility in terms of completeness and formatting. The structure remained unchanged when implemented into mongoDB, with the only manual addition being the information on the data's source.

Instagram "advertisers" data:


```
"ig_custom_audiences_all_types": [
  {
    "advertiser_name": "Epsilon Audience Data Provider",
    "has_data_file_custom_audience": true,
    "has_remarketing_custom_audience": false,
    "has_in_person_store_visit": false
  },
```

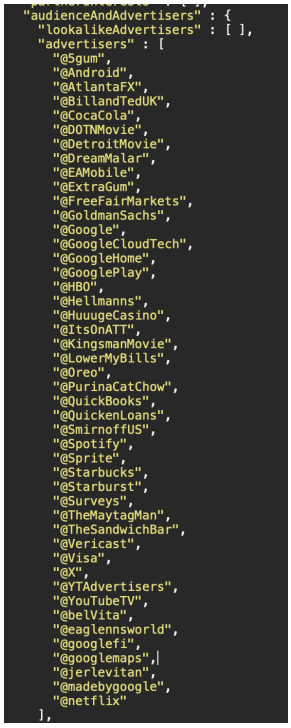

Facebook "advertisers" example:


```
{  "custom_audiences_all_types_v2": [
   {
    "advertiser_name": "Aeroplan",
    "has_data_file_custom_audience": true,
    "has_remarketing_custom_audience": true,
    "has_in_person_store_visit": false
   },
```


Advertiser data formatted in Mongo:


```
▾ source: Object
    site: "instagram"
    account: "lydiagraveline"
▾ custom_audiences: Array (283)
  ▾ 0: Object
      advertiser_name: "Epsilon Audience Data Provider"
      has_data_file_custom_audience: true
      has_remarketing_custom_audience: false
      has_in_person_store_visit: false
```


Midway through development, my Twitter data arrived, which included a list of "audienceAndAdvertisers." I wanted to include this in my prototype alongside the facebook and instagram advertisers, despite it having a different structure. Using ChatGPT, I executed a series of prompts to modify the structure to match that of Facebook and Instagram advertisers, such as "remove the "@" from each item," and "bracket each one in { }." I was able to copy and paste this into MongoDB, where I manually added the "source" information.



The “topics” from each platform had similar variations. Recognizing the need for uniformity for multiple platforms beyond Instagram and Facebook, I devised a program specifically for Instagram's data package format to automatically extract and format relevant information. The idea was that if this worked, I could create extend the program to work across all the provided data formats. The instagram topics were embedded in several layers “object>string_man_data > name > value,” the progrsam could extract the topics from a “your_topics.json” file, and output the topics as an array of strings. I copied and pasted this array and manually added it to mongo, and added the source information. While this approach was successful for Instagram, which I had three accounts to request data from, it proved to be too platform-specific and file-specific, and extending it would have been too time consuming for now.

My program and the outputted array of strings:

Upload a JSON file

Choose File your_topics.json Upload and Display JSON


```
["South Asian Food","Pea","Foods","Meat & Seafood","Fashion Products","Legumes & Beans","East
Asian Food","Baked Goods","Pasta & Noodle Recipes","Mangoes","Body Modification","Electronic
& Dance Music","Fashion","Cookbooks","Foods of India","Chia Seeds","Rock Music","Non-
Alcoholic Beverages","Drinks","Regional Cuisines","Types of Meat & Seafood","Sushi &
Sashimi","Fashion Styles & Trends","Beauty","Music by Genre","Plants & Trees","Visual
Arts","Makeup","Indian Food","Cheese","Rice","Beauty Product
Types","Produce","Yogurt","Coffee Drinks","Japanese Food","Greek
Yogurt","Comedy","Seafood","Rice, Grains & Noodles","Beauty Products","Horticulture &
Gardening","Avocados","Stone Fruits","Mammals","Dogs","Gardening &
Landscaping","Dairy","Jewelry","Nails & Nail Care","Noodles & Pastas","Asian Food","Seafood
Recipes","Faces & Face Care","Food in Media","Plants","Hair Care","Animals","Body Art","Nuts
& Seeds","Fruits","Literary Forms","Recipes","Memes","Clothing & Accessories","Desserts"]
```

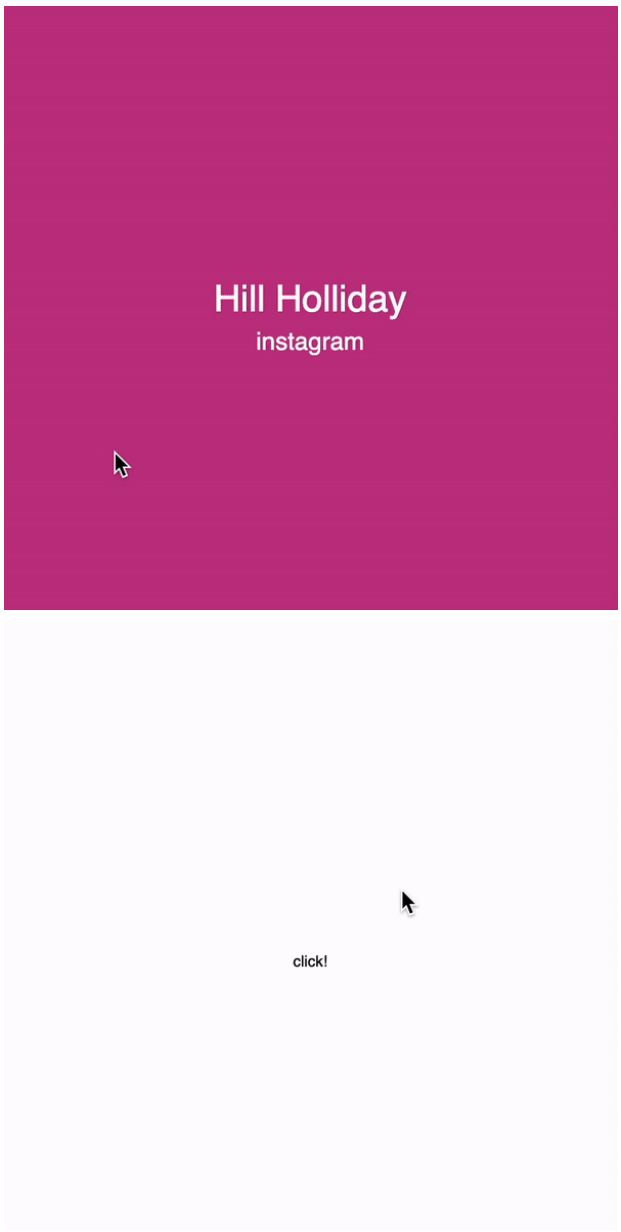

Instagram topics formatted in MongoDB:

Pinterest's data, received as an HTML file, was straightforward to integrate due to the simplicity of my Topics collection. For the Twitter topics data, I did the same process as the advertisers, utilizing ChatGPT to reformat it into an array.

Prototype Visualization

With the formatted data in place, I was ready to start prototyping.For each data type—advertisers and topics—I developed two visualizations using p5.js. The first visualization mapped data items on the y-axis, with the x-axis corresponding to the source of the data (social media platform and my account). The arrays are quite long, so every inch of the canvas tells you something new. This approach provided a straightforward yet informative interaction, allowing one to navigate through advertisers or topics seamlessly.

The second visualization felt a bit more playful. Users could click, triggering the appearance of a circle with either a topic or advertiser name at its center. The data within these circles was intentionally randomized, drawing from any source and appearing in any order. While initially unintentional, the circles expand indefinitely, obscuring the previously displayed data. I liked this way of interacting; as curiosity grows and the user clicks more and more, the information becomes progressively more obscured.



Next Steps

Moving forward, my focus will be on refining interaction methods, selecting and formatting data, and developing a coherent narrative based on the insights gained from the prototype. The final iteration of this project will take the form as an interactive website inspired by the hypertext storytelling, akin to "My Boyfriend Came Back From The War." Here, the audience will act as both web users and voyeurs, delving into my personal world of data.

To achieve this, I will need to develop a compelling structure that unfolds as users explore my personal data. Users will navigate through various data points and insights, revealing layers of information akin to a digital diary, a storyline, or a self-portrait. Each click or interaction will serve as a narrative choice, influencing the direction and tone of the user's digital journey.

Throughout the website, different data will be presented with varying interaction methods, building on the prototype while introducing new methods, like multimedia elements such as images, videos, and sound that could be integrated to evoke emotions, memories, or reflections associated with specific data points. For example, a photo shared on Instagram or a conversation snippet from Facebook could serve as visual or auditory triggers.

Incorporating interactive elements, such as decision points, surveys, or feedback mechanisms, will prompt users to actively engage with the unfolding story, creating a dialogue between the user and the data. I will carefully curate what data to include in the final version, addressing any remaining formatting challenges and ensuring the narrative's coherence.

Beyond introspection, I want my final version to serve as a dynamic digital archive or diary, cataloging my digital persona at the time of its creation. This archival aspect will address challenges in contemporary cultural archives, including issues such as censorship, data loss, and obsolescence.

Given the challenges of conveying sincerity on social media, and the fine line between self-expression and self-presentation, the narrative will explore the complexities of the performance of self in the digital realm. Inspired by the discomfort and vulnerability felt this project has generated, I want to further delve into hermeneutic violence, where social media interactions are overly scrutinized, and the immense desire for authenticity amid the evolving landscape of self-expression. The final version will be a reflection of my digital self in the midst of contemporary societal and technological pressures.