Tony Nguyen

Dr. Gina Sprint

CPSC 222 01

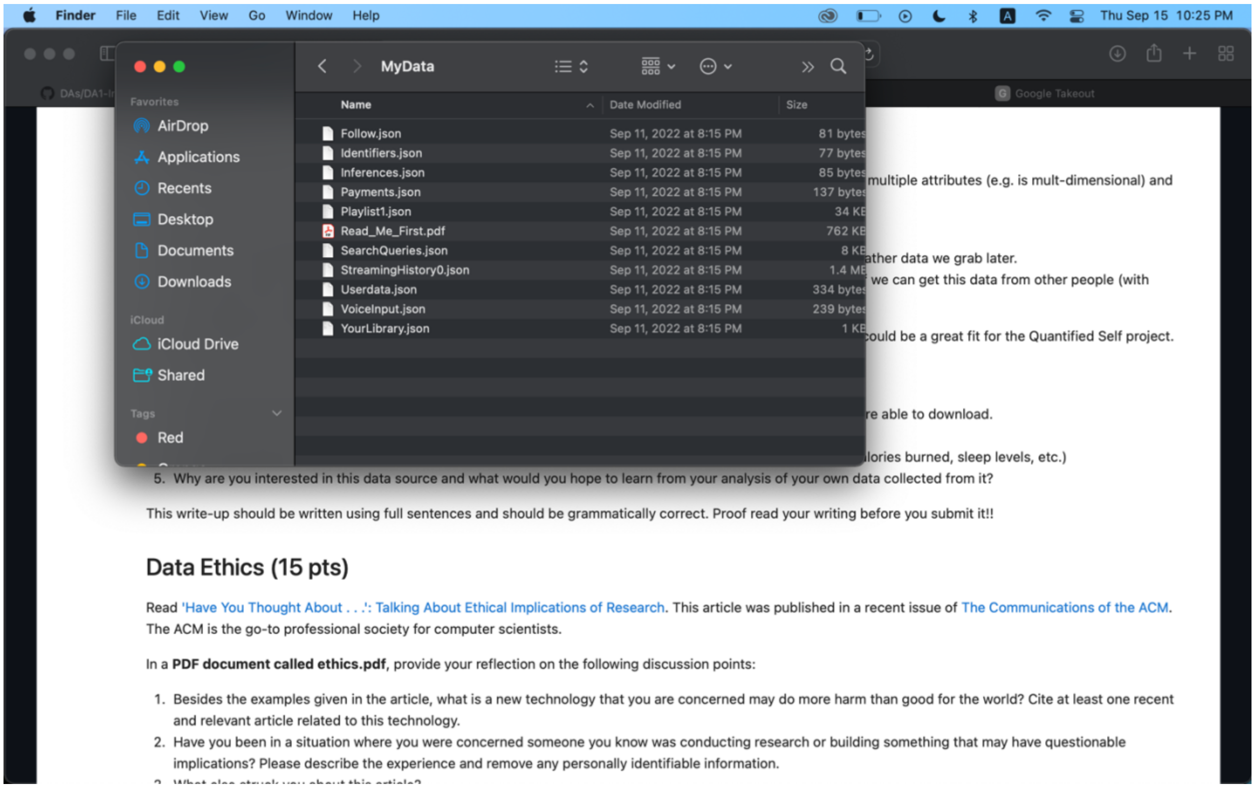
16 September 2022

Project Part 1

For this project, I chose three data sources: Spotify, Netflix, and YouTube history.

1. Spotify

   I am most excited about Spotify since this is probably my most frequently used application. My data for Spotify is an automatically generated one, which I requested from its website. Spotify collects data on various groups, from my payment information, daily mix 1, search queries, and the most notable, streaming history. For each group of data, Spotify has different intervals for collecting the information. With payment information, it notes the date I added my current credit card for monthly payments to Spotify. Speaking of daily mix 1, it includes every song that has been added to the mix, ranging from 2019 to 2022. For search queries, it collects my search history as well as the search keywords for three months since the day I downloaded the file. And for streaming history, it contains the song name, artist name, and the time stamp for each stream over one year, from September 2021 until today. I think this data topic will be interesting since it gives me a look over my genres over time as well as my listening habit. I can know more about when I usually listen to music, whether during the day or at night. It is also worth noting that Spotify uses JSON file format for its data report.



Name	Date Modified	Size
Follow.json	Sep 11, 2022 at 8:15 PM	81 bytes
Identifiers.json	Sep 11, 2022 at 8:15 PM	77 bytes
Inferences.json	Sep 11, 2022 at 8:15 PM	85 bytes
Payments.json	Sep 11, 2022 at 8:15 PM	137 bytes
Playlist1.json	Sep 11, 2022 at 8:15 PM	34 KB
Read_Me_First.pdf	Sep 11, 2022 at 8:15 PM	762 KB
SearchQueries.json	Sep 11, 2022 at 8:15 PM	8 KB
StreamingHistory0.json	Sep 11, 2022 at 8:15 PM	1.4 MB
Userdata.json	Sep 11, 2022 at 8:15 PM	334 bytes
VoiceInput.json	Sep 11, 2022 at 8:15 PM	239 bytes
YourLibrary.json	Sep 11, 2022 at 8:15 PM	1 KB

multiple attributes (e.g. is mult-dimensional) and

ather data we grab later.

we can get this data from other people (with

ould be a great fit for the Quantified Self project.

re able to download.

lories burned, sleep levels, etc.)

5. Why are you interested in this data source and what would you hope to learn from your analysis of your own data collected from it?

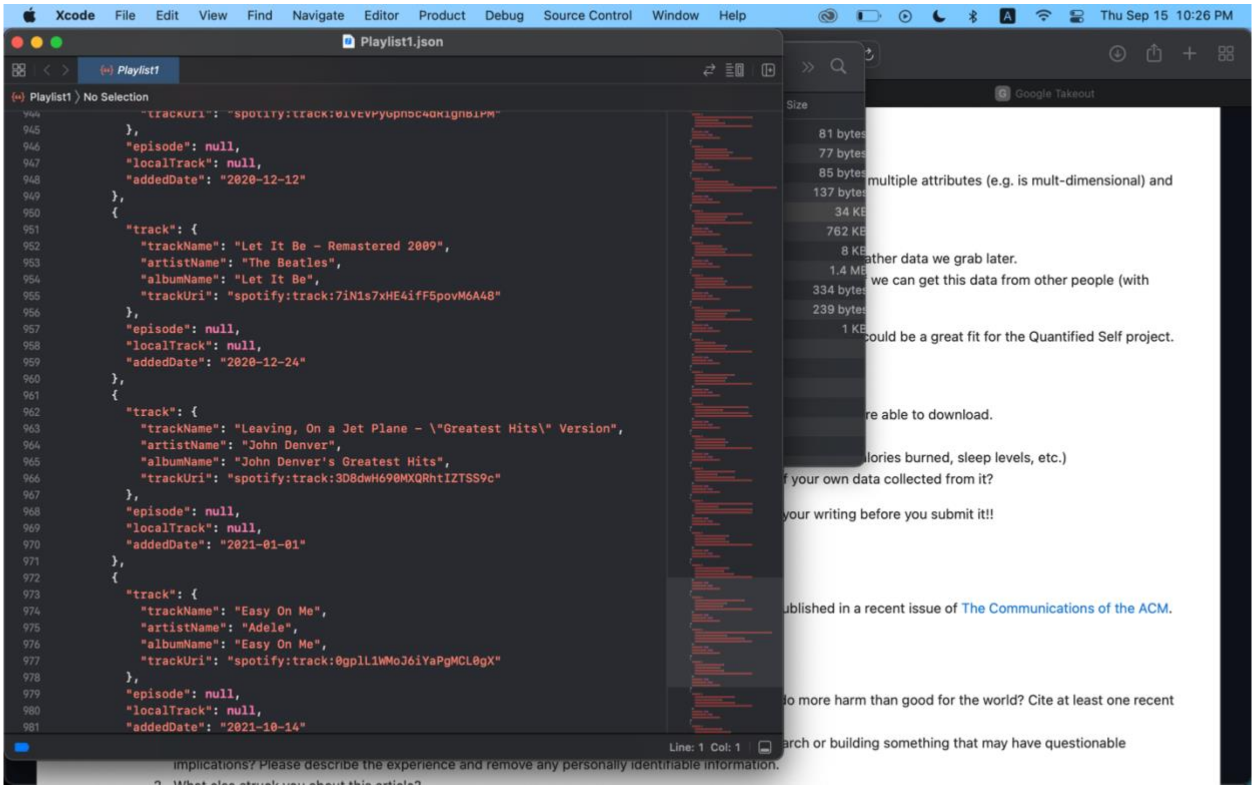
This write-up should be written using full sentences and should be grammatically correct. Proof read your writing before you submit it!!

Data Ethics (15 pts)

Read 'Have You Thought About . . .': Talking About Ethical Implications of Research. This article was published in a recent issue of The Communications of the ACM. The ACM is the go-to professional society for computer scientists.

In a **PDF document called ethics.pdf**, provide your reflection on the following discussion points:

1. Besides the examples given in the article, what is a new technology that you are concerned may do more harm than good for the world? Cite at least one recent and relevant article related to this technology.
2. Have you been in a situation where you were concerned someone you know was conducting research or building something that may have questionable implications? Please describe the experience and remove any personally identifiable information.
3. What else struck you about this article?



```
              "trackUri": "spotify:track:01VEvPyGpn5c4dRignBlPM"
            },
            "episode": null,
            "localTrack": null,
            "addedDate": "2020-12-12"
          },
          {
            "track": {
              "trackName": "Let It Be - Remastered 2009",
              "artistName": "The Beatles",
              "albumName": "Let It Be",
              "trackUri": "spotify:track:7iN1s7xHE4ifF5povM6A48"
            },
            "episode": null,
            "localTrack": null,
            "addedDate": "2020-12-24"
          },
          {
            "track": {
              "trackName": "Leaving, On a Jet Plane - \"Greatest Hits\" Version",
              "artistName": "John Denver",
              "albumName": "John Denver's Greatest Hits",
              "trackUri": "spotify:track:3D8dwH690MXQRhtIZTSS9c"
            },
            "episode": null,
            "localTrack": null,
            "addedDate": "2021-01-01"
          },
          {
            "track": {
              "trackName": "Easy On Me",
              "artistName": "Adele",
              "albumName": "Easy On Me",
              "trackUri": "spotify:track:0gplL1WMoJ6iYaPgMCL0gX"
            },
            "episode": null,
            "localTrack": null,
            "addedDate": "2021-10-14"
```

81 bytes
77 bytes
85 bytes
137 bytes
34 KB
762 KB
8 KB
1.4 MB
334 bytes
239 bytes
1 KB

multiple attributes (e.g. is mult-dimensional) and

ather data we grab later.

we can get this data from other people (with

ould be a great fit for the Quantified Self project.

re able to download.

lories burned, sleep levels, etc.)

f your own data collected from it?

your writing before you submit it!!

ublished in a recent issue of The Communications of the ACM.

o more harm than good for the world? Cite at least one recent

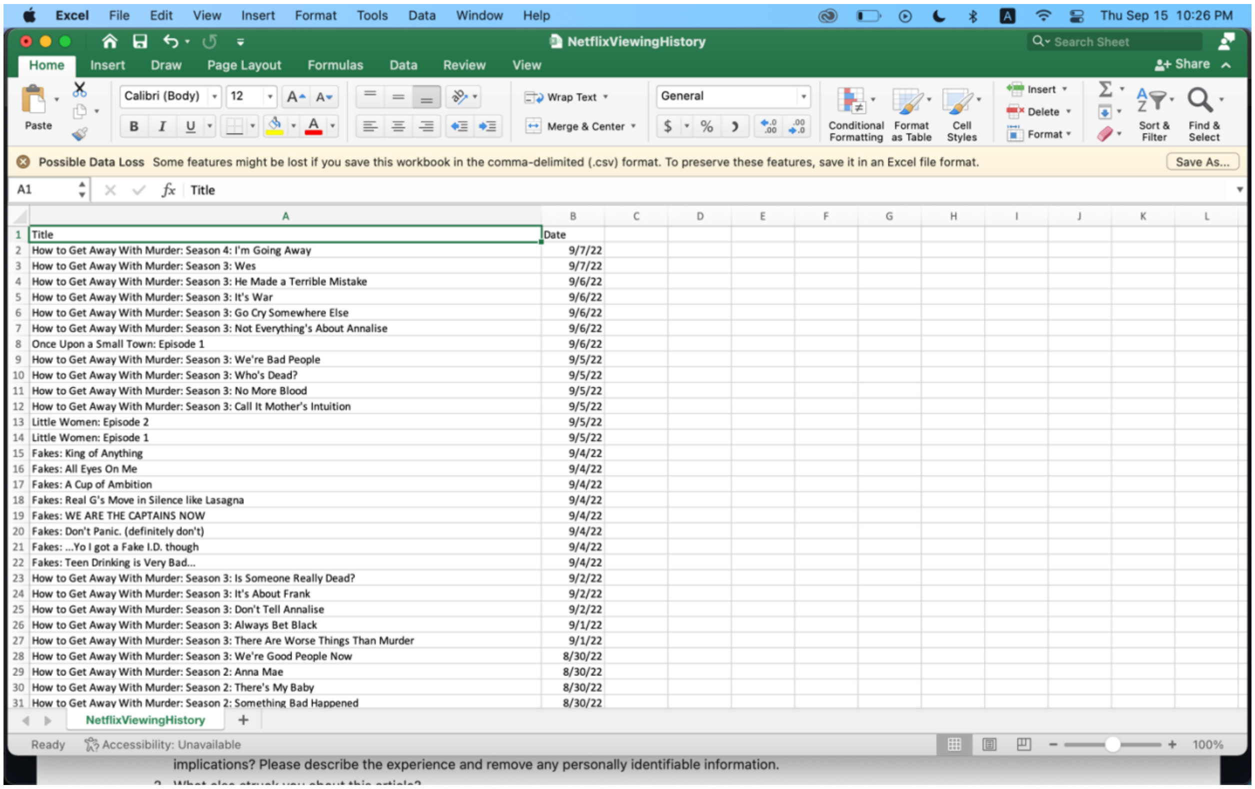
arch or building something that may have questionable

implications? Please describe the experience and remove any personally identifiable information.

3. What else struck you about this article?

2. Netflix

Netflix has a more straightforward dataset than Spotify. The takeout is in CSV format, which is accessible compared to other sources. Netflix is also an automatic-generated file that notes every time I use the service. The file has very few attributes since it only has the name of the movie or episodes of a series and the viewing date since I first used it in January 2020. I would hope that Netflix can provide more information, for example, the length of each episode and which type of movie I enjoy watching most. I know they can possibly do this since their algorithm for recommending movies is pretty accurate, which is the main reason I stick with Netflix but not other services. Regardless, I still think Netflix's data is worth looking at since I can see which movies I have watched over time and how many times I have rewatched some of the series.



Possible Data Loss Some features might be lost if you save this workbook in the comma-delimited (.csv) format. To preserve these features, save it in an Excel file format.

	A	B
1	Title	Date
2	How to Get Away With Murder: Season 4: I'm Going Away	9/7/22
3	How to Get Away With Murder: Season 3: Wes	9/7/22
4	How to Get Away With Murder: Season 3: He Made a Terrible Mistake	9/6/22
5	How to Get Away With Murder: Season 3: It's War	9/6/22
6	How to Get Away With Murder: Season 3: Go Cry Somewhere Else	9/6/22
7	How to Get Away With Murder: Season 3: Not Everything's About Annalise	9/6/22
8	Once Upon a Small Town: Episode 1	9/6/22
9	How to Get Away With Murder: Season 3: We're Bad People	9/5/22
10	How to Get Away With Murder: Season 3: Who's Dead?	9/5/22
11	How to Get Away With Murder: Season 3: No More Blood	9/5/22
12	How to Get Away With Murder: Season 3: Call It Mother's Intuition	9/5/22
13	Little Women: Episode 2	9/5/22
14	Little Women: Episode 1	9/5/22
15	Fakes: King of Anything	9/4/22
16	Fakes: All Eyes On Me	9/4/22
17	Fakes: A Cup of Ambition	9/4/22
18	Fakes: Real G's Move in Silence like Lasagna	9/4/22
19	Fakes: WE ARE THE CAPTAINS NOW	9/4/22
20	Fakes: Don't Panic. (definitely don't)	9/4/22
21	Fakes: ...Yo I got a Fake I.D. though	9/4/22
22	Fakes: Teen Drinking is Very Bad...	9/4/22
23	How to Get Away With Murder: Season 3: Is Someone Really Dead?	9/2/22
24	How to Get Away With Murder: Season 3: It's About Frank	9/2/22
25	How to Get Away With Murder: Season 3: Don't Tell Annalise	9/2/22
26	How to Get Away With Murder: Season 3: Always Bet Black	9/1/22
27	How to Get Away With Murder: Season 3: There Are Worse Things Than Murder	9/1/22
28	How to Get Away With Murder: Season 3: We're Good People Now	8/30/22
29	How to Get Away With Murder: Season 2: Anna Mae	8/30/22
30	How to Get Away With Murder: Season 2: There's My Baby	8/30/22
31	How to Get Away With Murder: Season 2: Something Bad Happened	8/30/22

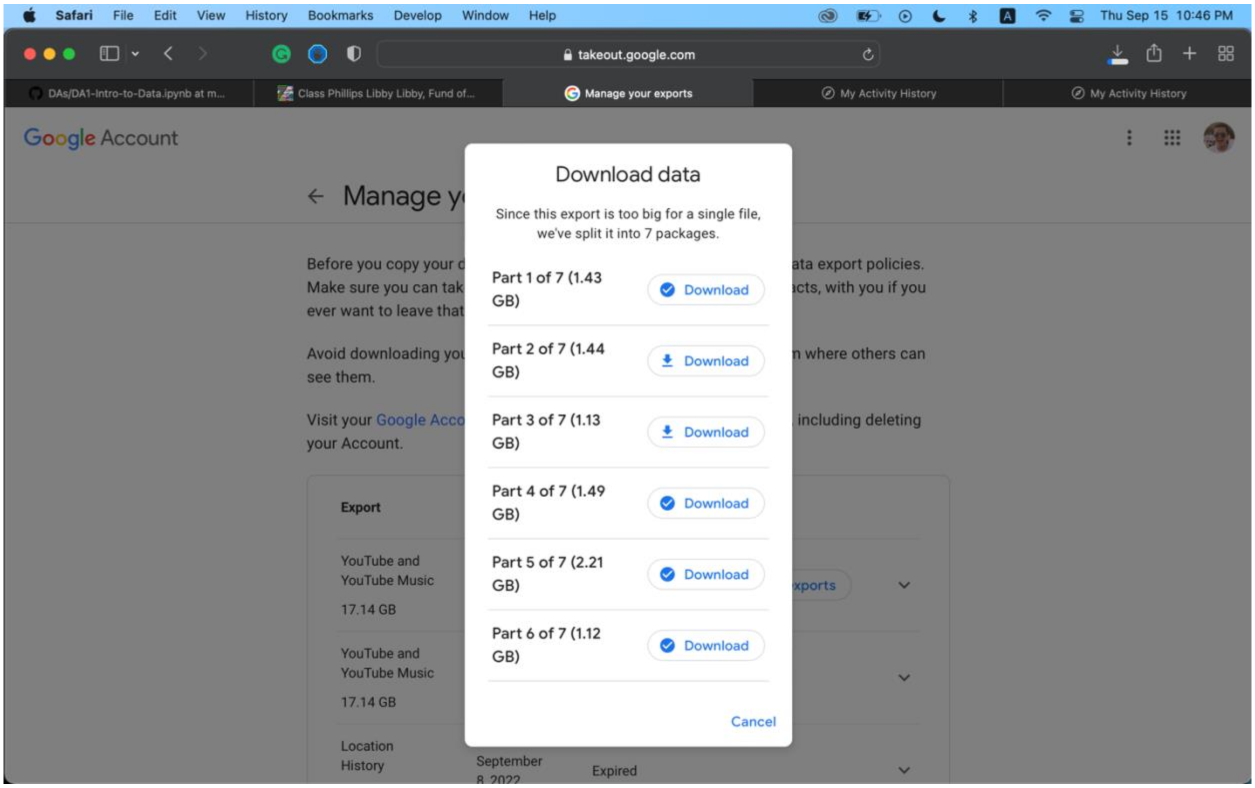
implications? Please describe the experience and remove any personally identifiable information.

2. What else struck you about this article?

(How to Get Away With Murder is really good!)

3. YouTube and YouTube Music

YouTube's data is the largest compared to Netflix and Spotify, with 17.14GB split into seven smaller compressed files. I was surprised to see this number, knowing that much data is a lot. However, over the seven files, the majority of storage consisted of my previously uploaded video to my channel. YouTube deems those videos a part of my takeout, which makes sense to some extent, but it would be nice for them to let me know beforehand so I can decide whether to download them or not to save more time. YouTube data is automatically collected, including my streaming history, downloaded videos, subscription list, and more. The only interval in the dataset appears in the streaming history when YouTube remembers every video I have watched since October 2012 and its timestamp, which is nearly ten years of data collection. I think having such big data from YouTube would be nice since it provides a flashback to my childhood and sees what I have watched over my journey of growing up. However, I do not have many ideas about what kind of analysis I can make with this information since watching preferences change over time, and individual videos cannot reflect it to the furthest extent. While other items in the dataset are written in CSV files, YouTube's streaming history is presented in HTML format and conveniently accessible using any internet browser. However, I notice a problem with the files is that it does not support Vietnamese characters. I don't know if this is due to the HTML format or other technical issues from YouTube. Still, it converts some Vietnamese video titles to unreadable characters and affects the dataset's quality.


Safari File Edit View History Bookmarks Develop Window Help
Thu Sep 15 10:46 PM
takeout.google.com
DAs/DA1-Intro-to-Data.ipynb at m...
Class Phillips Libby Libby, Fund of...
Manage your exports
My Activity History
My Activity History
Google Account
Manage y

Before you copy your d… ata export policies. Make sure you can tak… acts, with you if you ever want to leave that…

Avoid downloading you… n where others can see them.

Visit your Google Acco… including deleting your Account.

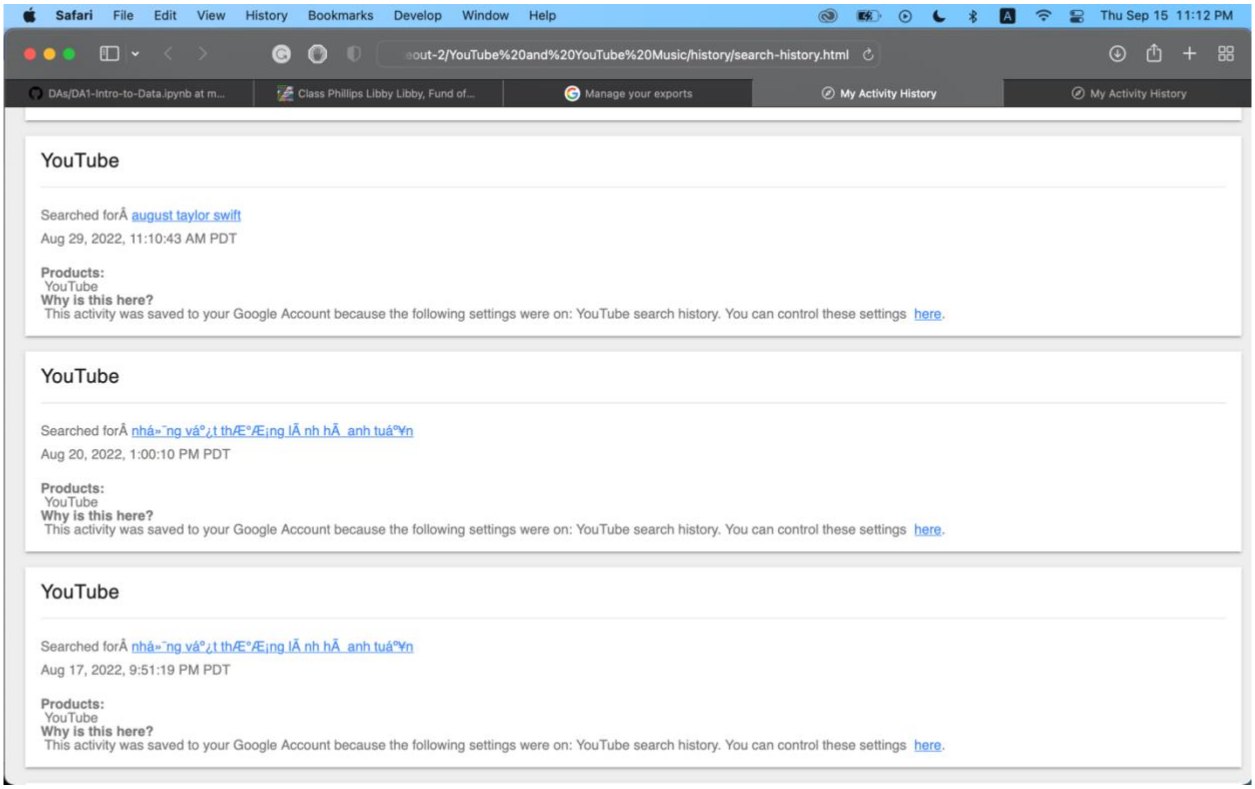
Export			
YouTube and YouTube Music 17.14 GB		xports	
YouTube and YouTube Music 17.14 GB			
Location History	September 8, 2022	Expired	

Download data

Since this export is too big for a single file, we've split it into 7 packages.

Part 1 of 7 (1.43 GB)	Download
Part 2 of 7 (1.44 GB)	Download
Part 3 of 7 (1.13 GB)	Download
Part 4 of 7 (1.49 GB)	Download
Part 5 of 7 (2.21 GB)	Download
Part 6 of 7 (1.12 GB)	Download

Cancel


Safari File Edit View History Bookmarks Develop Window Help
Thu Sep 15 11:12 PM
out-2/YouTube%20and%20YouTube%20Music/history/search-history.html
DAs/DA1-Intro-to-Data.ipynb at m...
Class Phillips Libby Libby, Fund of...
Manage your exports
My Activity History
My Activity History

YouTube

Searched forÂ august taylor swift
Aug 29, 2022, 11:10:43 AM PDT

Products:
YouTube
Why is this here?
This activity was saved to your Google Account because the following settings were on: YouTube search history. You can control these settings here.

YouTube

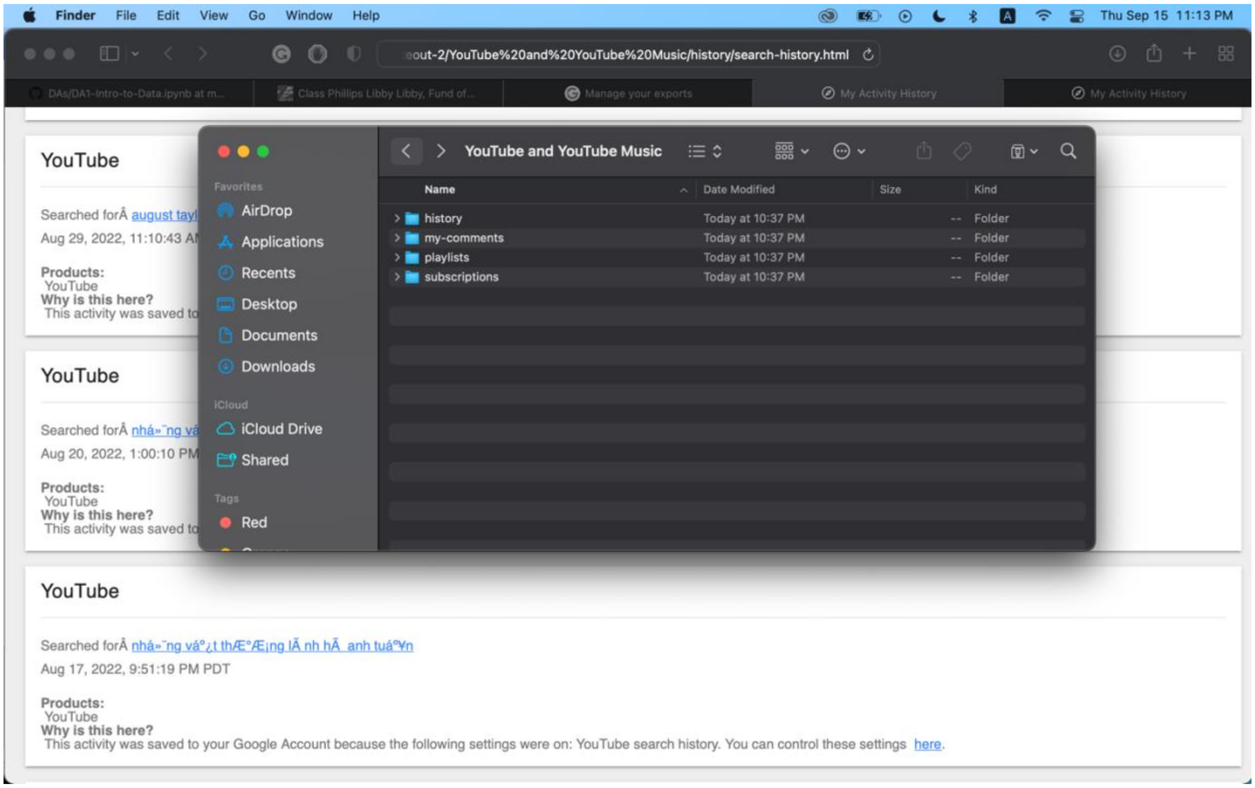
Searched forÂ nhá»¯ng vá°¿t thÆ°Æ¡ng lÃ nh hÃ anh tuá°¥n
Aug 20, 2022, 1:00:10 PM PDT

Products:
YouTube
Why is this here?
This activity was saved to your Google Account because the following settings were on: YouTube search history. You can control these settings here.

YouTube

Searched forÂ nhá»¯ng vá°¿t thÆ°Æ¡ng lÃ nh hÃ anh tuá°¥n
Aug 17, 2022, 9:51:19 PM PDT

Products:
YouTube
Why is this here?
This activity was saved to your Google Account because the following settings were on: YouTube search history. You can control these settings here.



…out-2/YouTube%20and%20YouTube%20Music/history/search-history.html

YouTube and YouTube Music

Name	Date Modified	Size	Kind
history	Today at 10:37 PM	--	Folder
my-comments	Today at 10:37 PM	--	Folder
playlists	Today at 10:37 PM	--	Folder
subscriptions	Today at 10:37 PM	--	Folder

YouTube

Searched forÂ august tayl

Aug 29, 2022, 11:10:43 AM

Products:
YouTube
Why is this here?
This activity was saved to

YouTube

Searched forÂ nhá»¯ng vá

Aug 20, 2022, 1:00:10 PM

Products:
YouTube
Why is this here?
This activity was saved to

YouTube

Searched forÂ nhá»¯ng vá°¿t thÆ°Æ¡ng lÃ nh hÃ anh tuá°¥n

Aug 17, 2022, 9:51:19 PM PDT

Products:
YouTube
Why is this here?
This activity was saved to your Google Account because the following settings were on: YouTube search history. You can control these settings here.