SEVEN SEAS

An Analysis of TV Torrenting

Andrew Banks, Riley Barklow, Anthony Sutherland, Franklin Xie

What we set out to learn.

◦ Originally we wanted to see how the limited availability of streaming services effected piracy in different countries.

◦ We abandoned that after realizing there was no good way to obtain country information on torrenting.

◦ Our goal then changed to analyzing the torrenting of streaming service exclusives.

  ◦ How much does each streaming service suffer from torrenting?

  ◦ How does torrenting compare to streaming services that are available internationally to ones with limited availability?

  ◦ How has torrenting been affected by covid-19 and the lockdown?

What We Analyzed

100 streaming service exclusives

- Streaming services we used were
  - Netflix
  - Hulu
  - Disney +
  - Amazon Prime Video
  - HBO Max
- Limited availability
  - HBO Max
  - Amazon Prime Video
  - Disney + (questionable)

20 non-exclusives (control group)

- 20 shows that were available either through multiple streaming services, or through regular TV were selected as a control group.
- These non-exclusives were limited to shows that had come out within the last ten years.
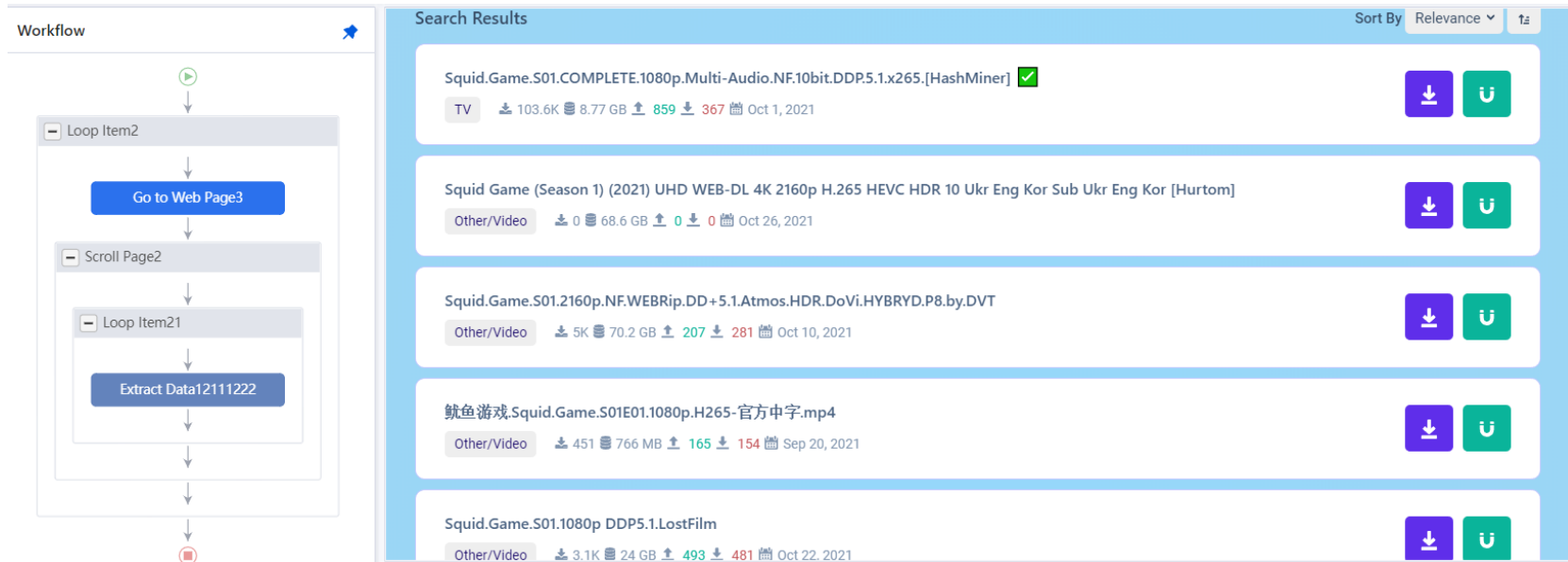
SOLID TORRENTS

OCTOPARSE

PYTHON PARSER

DATA CLEANUP

Tools We Used

Solid Torrents is a torrenting search engine that keeps track of the download numbers for each torrent.

Solid Torrents searched through the internet for torrents of each of the 120 shows while data was collected on each.

Data Collection

Octoparse is a special program that assists users in designing web scraping routines.

This was used to scrape every result produced by solid torrents.

Info collected:

- Number of Torrents
- Date
- File name (for cleaning)

Parsing and Cleaning Data

◦ We constructed our own parser using python which:
   ◦ Created a data set for each show.
      ◦ Total number of torrents
      ◦ Date of oldest torrent
   ◦ Categorized them by streaming service.
   ◦ Tracked the dates for each individual torrent (For covid numbers).

◦ The cleanup process was split between Octoparse and our own parser.
   ◦ Octoparse ignored useless lines of data like those that contained zero downloads.

   ◦ We built a library of words for shows with generic names or ones part of a franchise.
   ◦ If those words were in the file name for a torrent for that show, the line of data was Skipped.
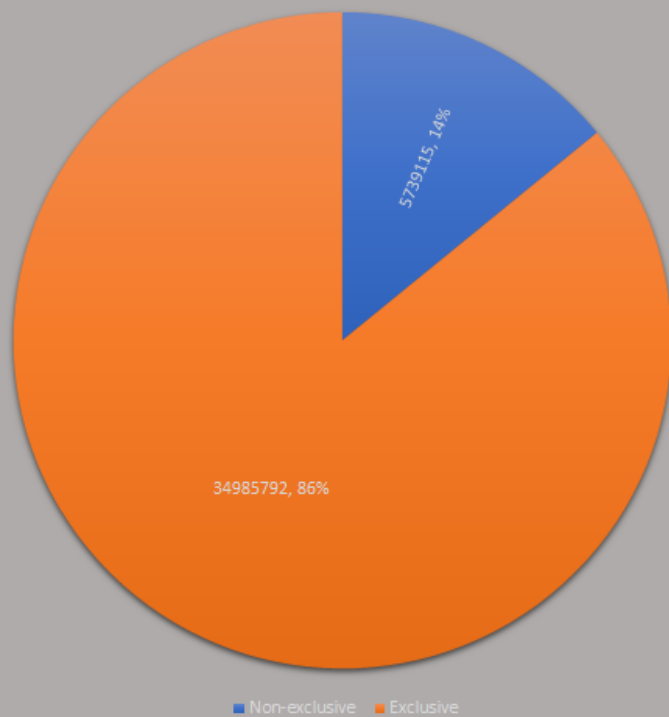
Example:
The Witcher ⬅ Netflix show
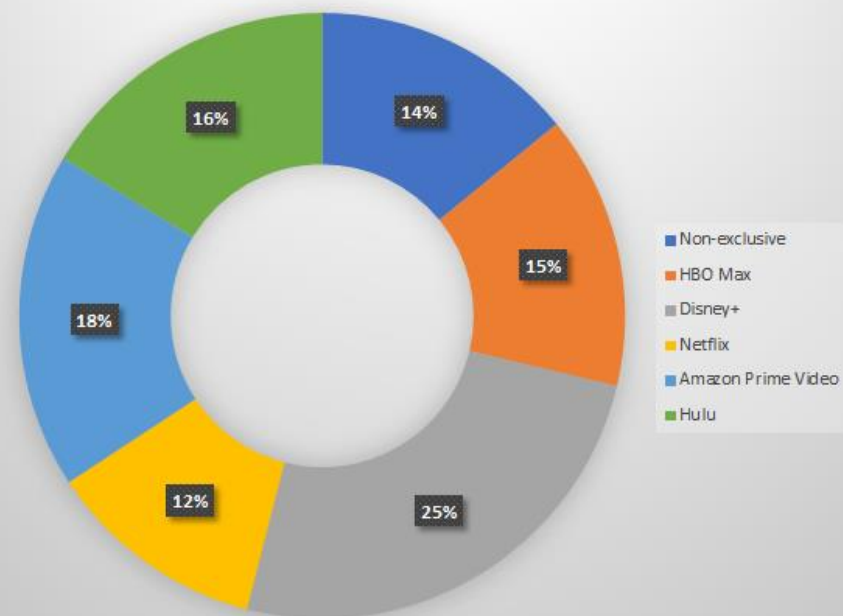
The Witcher 3: ⬅ Video Game
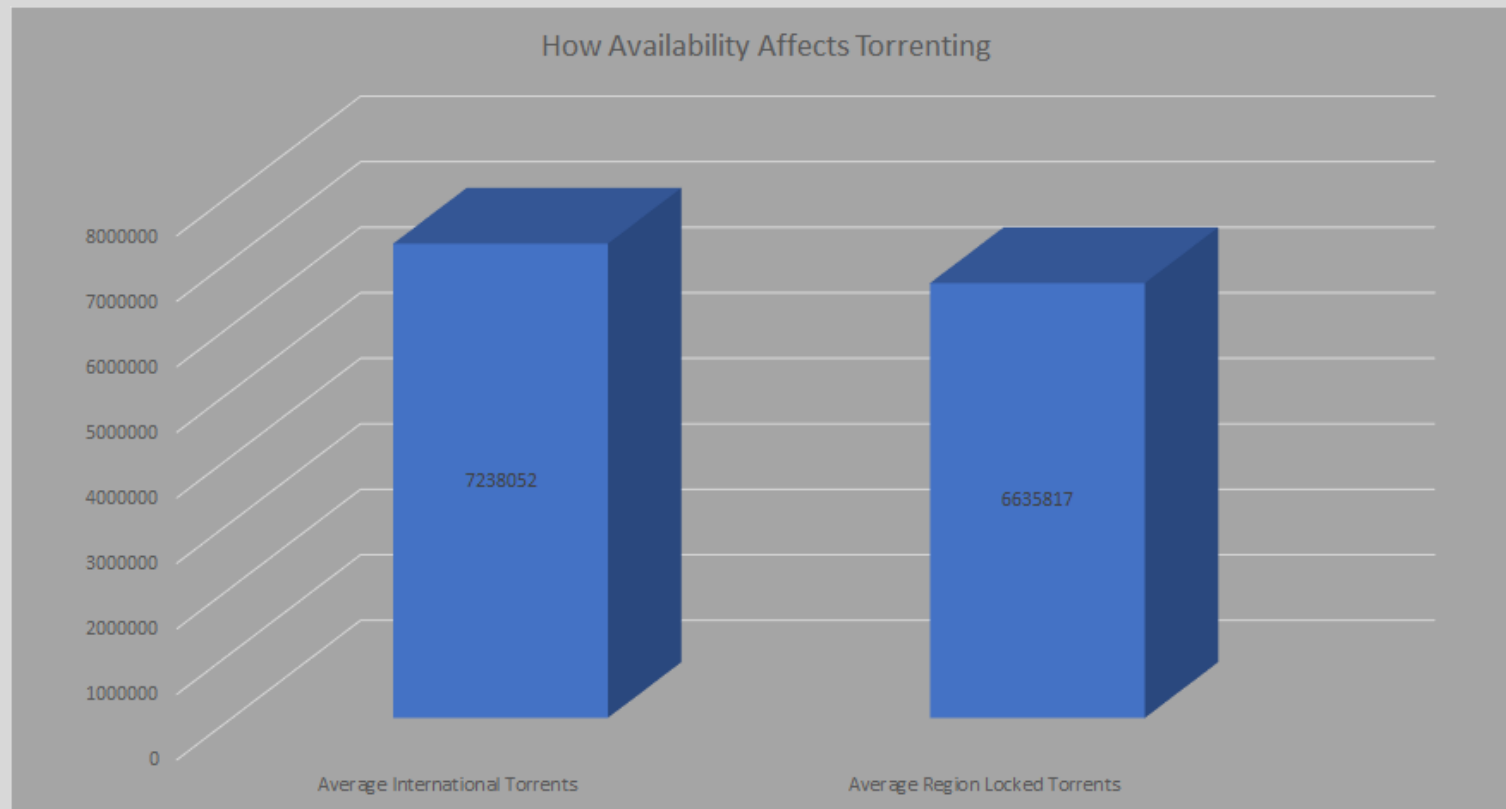Wild hunt

Termination words: Wild, Hunt

Our results

Total exclusive vs Non-exclusives

5739115, 14%
34985792, 86%

Non-exclusive   Exclusive

Total Torrent Downloads

16% 14%
18% 15%
12% 25%

Non-exclusive
HBO Max
Disney+
Netflix
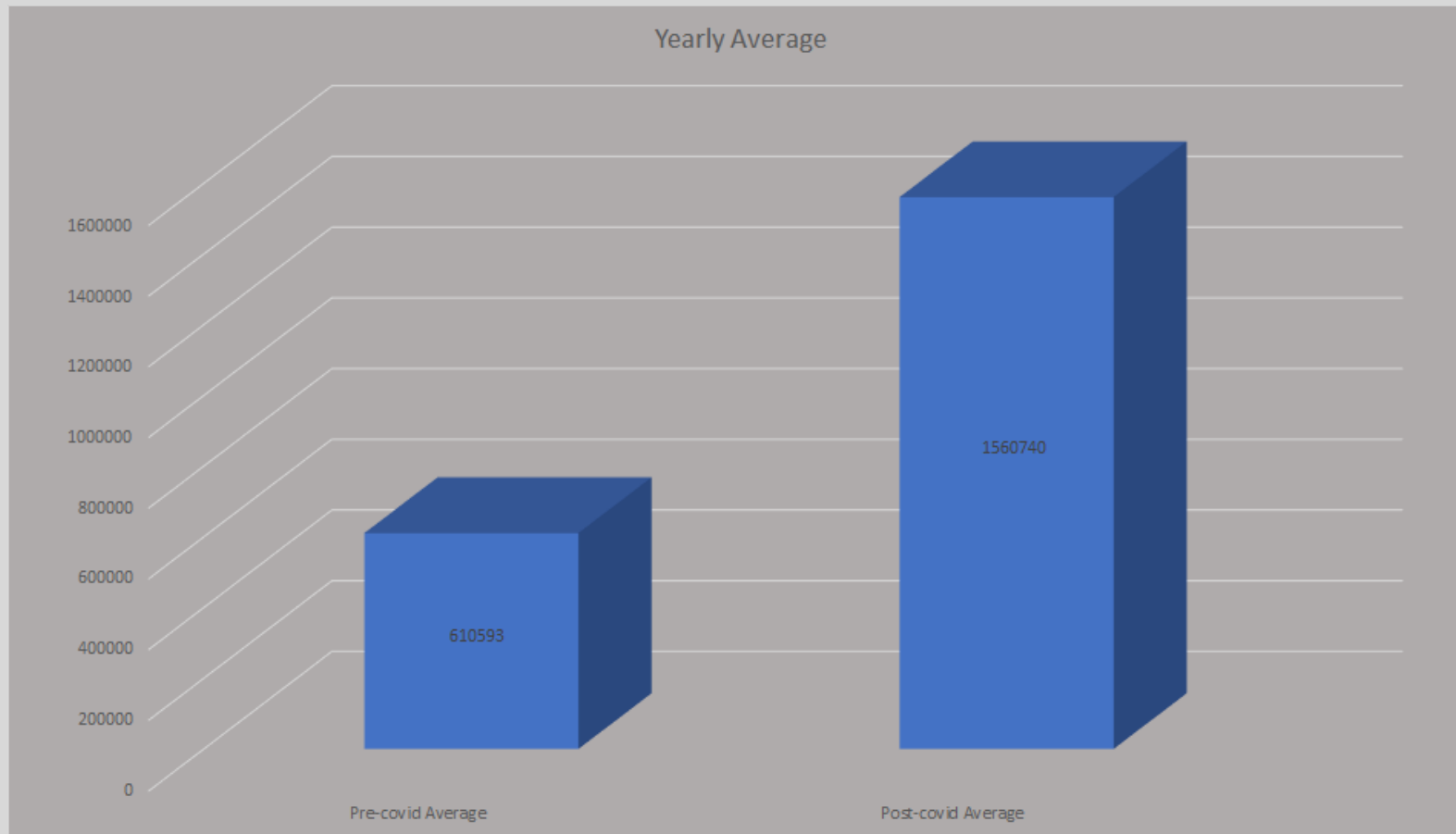Amazon Prime Video
Hulu

Our results

How Availability Affects Torrenting

Our results

Conclusions

◦ Despite being one of the newest streaming services, Disney + seems to suffer the most from streaming,

◦ Netflix suffers the least despite being one of the longest running and most popular services.

   ◦ This may be due to the fact that Netflix was more widely available for a longer amount of time than Disney +
   ◦ Disney+ came out only a few months before covid started and torrenting nearly tripled.
   ◦ Disney+ exclusives include Marvel and Star Wars shows.

◦ International availability was hard to determine due to uncertainty in how to categorize Disney +. More in depth research may be required.

◦ Covid-19 and the subsequent lockdown correlated heavily with an increase in torrenting. This could also be due to the recency of several streaming services.

Q&A