COSC 445: Project Proposal

I. OBJECTIVE

Analyze the rate of online piracy across various countries and compare it with income levels and availability of streaming services.

II. MOTIVATION

Despite it's questionable legality, piracy is an immensely popular means of consuming media. We aim to learn more about society's relationship with piracy. We will not only look at what types of media are pirated the most, but also dig into possible infrastructural and socio-economic factors that affect the overall rate of piracy in specific populations.

There are many different elements we believe affect not only overall piracy levels, but also the light in which a population views the act of piracy. We will look at many different data sets to get the best overall picture of society's relationship with piracy including the following: the availability of media in specific regions, DRM protected media such as region-locking, how income levels and local infrastructure affects piracy and media consumption, and how COVID affected the rates of piracy around the world, among others.

Piracy of media is no longer confined to basement servers and P2P file sharing. It is now much larger and more mainstream, complete with polished Netflix-like applications that bring an ease of use that was not present in the past. We believe the analysis of this kind of data will show us how piracy has evolved over time and shed light on the external factors that affect global media piracy.

III. DATA DISCUSSION

We want to compile a wide variety of data for various aspects of piracy that could play a role in our objective. First we want to look at wate of piracy across different countries to get a good idea of the specific places where it happens most frequently. Then we wanted to look at the availability of streaming services in each of those countries to see if a lack of accessibility was the main cause for piracy. However, we then realized that just looking at streaming services leaves out several other factors that might be key to our understanding of piracy. First, we also need to consider income levels inside the various countries we select, as a country with a large wage gap may have more piracy due to the cost of streaming services. On the other hand, piracy may go down in countries where internet access and personal computers are limited, so we need to gather data on that as well. If we want to expand past streaming services, we may look at other exclusive media like untranslated regional content and piracy versus movie going costs when compared to a country's average income. If the data is available, we'd also like to compare pre and post-covid piracy activity.

IV. MEMBER RESPONSIBILITIES:

Data collection with web API's (Kellen and Anthony)

Analyzing Data and compiling data with Numpy and Pandas data wrangling libraries (Andrew and Franklin)

GUI to display data (Riley)

V. MILESTONES

October 5 2021

Datasets that could be relevant to the project have been identified and marked down somewhere. Other sources of data for the project have been found and verified as viable for use.

October 10 2021

The chosen datasets have been read through, and ones deemed useful to the project are kept. Data from other sources has been gathered and made into datasets.

October 31 2021

Programs needed to parse the datasets have been created and are in a working state. Programs to take the parse data and create any graphical representations are also in a working state.

November 14 2021

A rough draft of the final project report has been completed. The draft is a comprehensive report of the outcome of the project, and includes data and graphics created from the user made programs.

November 21 2021

The draft of the final report has been improved on and the final draft of the report is ready to present at the end of the semester.

VI. EXPECTED OUTCOME

We would assume that countries with a lower average income would pirate the most. First, these places would have less access to TV and movies which results in a higher rate of piracy. Second, we would assume that countries with a lower income would not be able to access streaming sites which leads to more privacy. An outlier that could throw this assumption off is the availability of the internet in these countries.