**Project Proposal**

**Project Description**

Our team will collect and analyze recent Backpage and Craigslist postings in efforts to identify potential patterns of online sex trafficking. Backpage and Craigslist are classified advertisements websites where individuals can sell products or services to others, or alternatively, make personal connections. Unfortunately, these websites have also been used to traffic unwilling sex workers. We are particularly interested in analyzing potential patterns of known coded words, emojis, descriptions, ages, ethnicities, and locations, among other possible insights, within our clusters of interest.

**Data Source**

There is no provided API for Backpage and Craigslist. We will apply alternative web scraping methods to collect all posts within each of the websites’ Women seeking Men section for major cities in California (San Francisco, Los Angeles, San Diego, etc).

**Analysis**

We will perform a clustering analysis on the collected documents in attempts to differentiate between possible trafficking of victims/young victims versus likely non-trafficking cases. Within the trafficking cluster(s) we will perform additional analyses to assess the distributions of the cluster(s) regarding age/age range, race, location, and any additional insights that can be gained from this effort.

**Challenges**

We anticipate we will need to overcome the following challenges as part of our project:

1. Performing standard stop word removal, stemming, and phrase clustering on short posts that may not be grammatically correct, that may not use correct spelling, that contain numerous abbreviations, etc.
2. Differentiating posts that contain known coded words used in a potential trafficking context versus posts that contain the same words used in a different context.

**Data Source Justification**

In spite of the fact that Craigslist ceased its adult services section in 2010 and Backpage shut down its own adult ads section in early 2017 as a result of mounting political pressure, use of these platforms for the specific purpose of trafficking victims remains strong across other sections of the websites (e.g. Women seeking Men). Hundreds of adult ads are posted daily in nearly every single city where Backpage and Craigslist operate. Therefore, we feel confident we can collect the raw data needed to support our goals and analysis for this project.

**Deliverables**

We will provide the following deliverables at the end of our project:

1. A dataset containing posts scraped from each website’s sections of interest
2. A frequency distribution analysis by demographics of interest
3. A short in-class presentation of our findings, discussions of their meaning, and general "lessons learned" from our project