

# Using SAS Viya Textual Analytics to Identify and Predict Human Trafficking Attractors and Movement

Lauren Agrigento, Louisiana State University



Author: Lauren Agrigento

**Abstract:** Human trafficking networks commonly move from city to city in an effort to both avoid the police and to take advantage of certain events which may increase their business. In this study, unsupervised and unstructured textual data scraped from advertisements on Backpage.com is analyzed through SAS Viya Textual Machine Learning Analytics techniques to ultimately identify the types of appeals that human traffickers utilize in recruiting ads. Custom Concepts in SAS Viya allows us to classify indicators of the ad being the reason for the advertisement's appearance, such as certain words or phrases which may indicate if the trafficking network moved to the area in order to benefit from the recruitment's increased demand.

## Introduction

The human trafficking epidemic is one of the most prevalent, pressing issues faced by America today. According to DoSomething.org, human trafficking is the third largest international crime industry and, in turn, generates billions of dollars. Trafficking involves exploitation in many forms: forcing victims into prostitution, subjecting victims to slavery or involuntary servitude and compelling victims to commit sex acts for the purpose of creating pornography. Women and girls make up the bulk of the over 4 million victims being coerced into human trafficking with the false promises of things such as economic opportunity and a life of independence. One of the major avenues that human traffickers use to recruit victims is online advertising. Using advertisements on websites such as Backpage.com, an innocent victim is easily connected to a predator. Using contextual analysis, characteristics of profiles thought to be associated with human trafficking are constructed. Then, a new ad is compared to each of the profiles and assigned a score representing how likely the ad is to be associated with the human trafficking profile. These profiles can be used by law enforcement to minimize the set of advertisements that need to be investigated, reducing the amount of time officers spend sifting through advertisements.

## Data Preparation

WebHarvy.com offers this explanation of web scraping:

*Data displayed by most websites can only be viewed using a web browser. Web Scraping is the technique of automating this process, so that instead of manually copying the data from websites, the Web Scraping software will perform the same task within a fraction of the time.*

When information from the advertisements is extracted from Backpage.com, the important information is surrounded by code that formats how it is displayed in an internet browser. The format code is removed, leaving blocks of text that needed to be analyzed. This raw data is then further refined to extract useful information such as phone numbers, text, dates, and locations. In order to format the data in an easily interpretable way, variables such as Ad Category, Ad Content, Posting Body, Location, Date, and Phone Number were created. The parsing program was able to identify and extract this information from nearly 90% of the advertisements, even when the individuals posting the ads attempted to evade text mining software. For example, letters and symbols were inserted into phone numbers; a 10 digit phone number could be represented as 7eight5 0l3 259b. The parsing program used the SASPy framework for Python to translate the deceptive information into a usable phone number. Finally, in order to investigate the geographic location in which the advertisement originated, the City variable was parsed from the Location variable, then merged with a list of all existing latitudes and longitudes by city.



## Methods

This study used advertisements scraped from Backpage.com, which included 531,128 observations and 23 variables. These observations were collected by our team and span in time from 2013 to late 2018. A second dataset, produced with help of State Police, contains 1,685 records of verified human trafficking ads. The second dataset will be used to create a profile in which all other ads will be compared to.

# Using SAS Viya Textual Analytics to Identify and Predict Human Trafficking Attractors and Movement

Lauren Agrigento, Louisiana State University

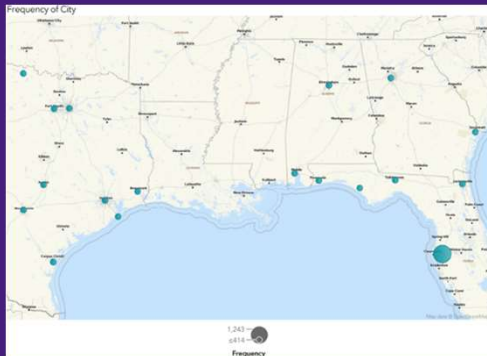


Author: Lauren Agrigento

## Profile of Recruiting Advertisements

Recruiting ads are a small category that take up 0.3% of all ads. It is the highest priority for analysis, because the contact information in these ads may provide direct links to the human traffickers and pimps (i.e., phone numbers, email addresses, and more). An example of a recruiting ad is below:

*"DAILY CASH PAYOUTS \$500-1200 daily Safe professional clients -- WE SCREEN ALL. Upscale locations All bookings and advertising provided A decade of experience Innovative and organized management WHO WE EXPECT TO WORK WITH: Attractive females 27-45 YRS Sexy and open-minded and personality Must be reliable and trustworthy Must have a reliable car Must be NOT OVERWEIGHT Must always remember that CLIENT SATISFACTION IS NUMBER ONE Priority SEND 1 PHOTO AND CONTACT INFO WHEN REPLYING!! WE WILL CONTACT YOU IF INTERESTED WITHIN 24 HRS."*



Some of the most notable information that describes the recruiting ads is the geographic range. Based on feedback from victims who have escaped, it appears that recruiters often remove the victims from their usual environment and quickly move them into a new, unfamiliar place, often across state lines. In the map below, we see that there are ads in many different cities and states. Tampa, specifically, has the highest concentration of recruiting ads with 1,243 observations.

Phone	Frequency ▼
7272174110	230
7276234810	230
7278272903	225
7279548887	220
7272534995	220
8135404094	66
8779766889	53
3342032829	36
8053083934	30
2016869604	30

Above is a list of the top 10 most used phone numbers in recruiting ads. It is interesting to note the number of ads associated with each phone number. The most used phone number has 230 uses, which is strongly indicative of a human trafficking ring or organization.

Topics (26)			
<input type="checkbox"/>	Topic	Created by	Documents ▼
<input type="checkbox"/>	+interest, modeling, +career, +model, modeling	User	414
<input type="checkbox"/>	+model, modeling, +career, +interest, number	System	378

Topic modelling can be described as a method for finding a group of words (i.e topic) from a collection of documents that best represents the information in the collection. It can also be thought of as a form of text mining a way to obtain recurring patterns of words in textual material. When the text mining pipeline was run, the results indicated that most ads fall under a topic involving modeling and careers, with over 400 ads saying generally the same thing. As stated earlier, women are being victimized by the prospect of modeling and careers they never would have had otherwise

Start Modeling Career with NEW Lingerie Magazine! - 34 Our NEW Victoria Secret style lingerie magazine is looking for some sexy models that are comfortable with their body and are interested in being featured in this new upcoming magazine. MODELS MUST BE WILLING TO TRAVEL! Models must be 18 or older. BEGINNER MODELS WELCOME! NATIONAL EXPOSURE! HOUSING INCLUDED! If you are interested in starting a REAL MODELING CAREER send PHOTOS and PHONE NUMBER to this email. We will develop you! We will build website and build your professional modeling portfolio! Also feel free to call us at 503 268 2446. Thank you. (PHOTOS AND PHONE NUMBER)

Relevancy:  
0.970

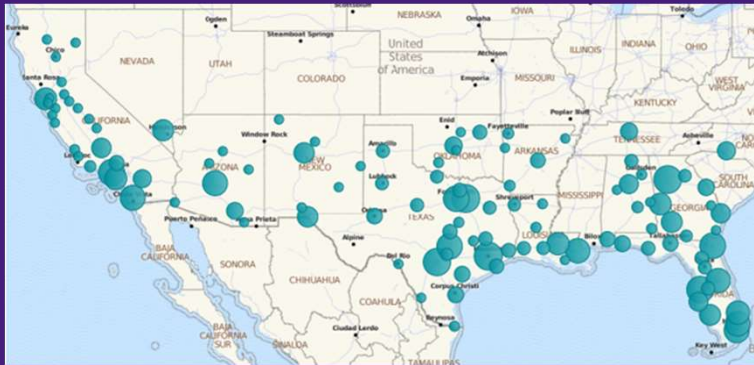
Sentiment:  
Positive

When this topic is examined, all matches associated with it are assigned a relevancy score, indicating how well a document satisfies a rule or model. The best match has a score of 1 and reflects a perfect (100%) match. Our model's top classification is the match above.



Author: Lauren Agrigento

## Classifying Other Advertisements



Using the geographical range, phone numbers, and topics from the already known recruiting ads, we can classify ads from our second, larger data set. Upon further review of the other advertisements, we can see the even larger geographical range from the increase in ads.

A further look into the phone numbers is even more revealing. Out of the top phone numbers, 19 out of 20 used in the recruiting ads were used in other, completely different ads. For example, the phone number (813) 540 4094 has the following two ads associated with it:

*"Adult Entertainers: Make Today! We're looking for new faces. attractive Females escorts Wanted for up-Scale Adult Establishment. Please tell us about yourself and send pface and body pics to e-mail: thelingeriestudiotampa@gmail.com"*

*"Now Hiring Lingerie Models for 6PM-2AM shifts and weekends. Start Today! Text 813- 540-4094 - 29 Remain independent and make Daily Clean Up-Scale Environment On Facebook: <https://www.facebook.com/thestudiotampa/> ATM available inside In-call Only Were currently Hiring please inquire within. The Studio 1053 West Busch Blvd Tampa FL 33612 New Phone # 813-933-4411 Hours: Sun-Tue 10 AM-2AM Wed-Sat: open 24 Hours Look For The Red Lights Brandon clearwater st Pete New Tampa tampa"*

Although completely different syntax, their general nature is the same. We can, in turn, conclude that they came from the same source and for the same purpose.

The most compelling piece of evidence, however, comes from the use of the best fitting topic created for the recruiting advertisements. Using the concept node in the SAS® Viya text mining software, the team created a custom concept that addressed the same key terms as did the topic. Using this, we can see hundreds of other advertisements that likely have the same motive. Below is an example of a promise of fame and fortune that any unsuspecting woman could easily fall victim to.:

*"Ladies ONLY, Looking to become the next BIG ADULT SUPER STAR? Come join our ELITE talent management company! Have fun and be safe STARRING in NON DEGRADING HIGH END Adult Film Magazines ie. Penthouse Hustler Cheri etc. \$800+/per shoot. STEADY, LONG TERM, QUICK, AND EASY MONEY! PLUS FREE HOUSING!! SAVE \$200/WK THAT THE OTHER AGENTS CHARGE AND YOU WON'T EVEN GET YOUR OWN ROOM OR BATHROOM. WE CAN ACCOMODATE THAT TOO! WE HAVE 2 OTHER MODEL HOUSES THAT YOU WILL NOT HAVE TO SHARE YOUR ROOM OR BATHROOM FOR THE SAME FEE. PLENTY OF REFERENCES, WE ARE HAPPY TO PUT YOU ON THE PHONE WITH OUR CURRENT MODELS. Just text some smiling selfies and bikini pics, or give a call to Company owner Bella. If email is easier for you, you can just email pictures to her"*



Author: Lauren Agrigento

## Conclusions

Human trafficking ads are posted many times every day, all across the country. When profiles are created based on the geographical range, phone numbers, and topics within the text, newer advertisements can be classified into the dangerous recruiting category. With this information, state and federal police can pinpoint rings, predators, and individual advertisements at risk for potential human trafficking violations. By doing that, potential victims are spared, and lives are saved.

## Recommendations

Ads related to human trafficking are constantly updated, so we suggest an ongoing web scrape to help to detect current activity. For future studies, we suggest a focus on commonly used phone numbers to discover phone numbers associated with specific advertisements. With a focus on frequent phone numbers, movement patterns of both individuals and organizations can be tracked.

## References

- "11 Facts About Human Trafficking." DoSomething.org | Volunteer for Social Change, [www.dosomething.org/us/facts/11\\_facts\\_about\\_human\\_trafficking](http://www.dosomething.org/us/facts/11_facts_about_human_trafficking).
- "7 Facts You Didn't Know about Human Trafficking." USC Suzanne Dworak Peck School of Social Work, [dworakpeck.usc.edu/news/7\\_facts\\_you\\_didnt\\_know\\_about\\_human\\_trafficking%3Futm\\_source%3Dtest](http://dworakpeck.usc.edu/news/7_facts_you_didnt_know_about_human_trafficking%3Futm_source%3Dtest).
- "Facts on Human Trafficking." Sisters of Mercy, [www.sistersofmercy.org/resources/facts\\_on\\_human\\_trafficking/](http://www.sistersofmercy.org/resources/facts_on_human_trafficking/).
- "Human Trafficking Statistics & Facts." Safe Horizon, [www.safehorizon.org/get\\_informed/human\\_trafficking\\_statistics\\_facts/#description/](http://www.safehorizon.org/get_informed/human_trafficking_statistics_facts/#description/).
- "KDnuggets." KDnuggets Analytics Big Data Data Mining SysNucleus. "How Tos." Web Scraping Explained, [www.webharvy.com/articles/howto.html](http://www.webharvy.com/articles/howto.html).
- "KDnuggets." KDnuggets Analytics Big Data Data Mining and Data Science, [www.kdnuggets.com/2016/07/text\\_mining\\_101\\_topic\\_modeling.html](http://www.kdnuggets.com/2016/07/text_mining_101_topic_modeling.html).
- SysNucleus. "How Tos." Web Scraping Explained, [www.webharvy.com/articles/howto.html](http://www.webharvy.com/articles/howto.html).