**Stop Human Trafficking: An Analytical Approach**
**MGT 6203 - Team 064**
**Lucas Chen, Amit Dommeti, Roy Garcia, Garrison Winter** TA: David Firrincieli

**Background:**
Human trafficking is typically considered a form of modern-day slavery. Human Trafficking, as defined by Mariam Webster is "organized criminal activity in which human beings are treated as possessions to be controlled and exploited (as by being forced into prostitution or involuntary labor)." It has an everlasting impact on affected individuals, leading to lifelong damage to the survivor's mental and physical health. As depicted by Love Justice, there are "an estimated 50 million people trapped in slavery around the world" as a result of human trafficking. There is an increasing amount of public attention that is being driven toward the problem of human trafficking. There is a shortage of analytics solutions that can assess decisions related to preventing human trafficking.

**Problem Statement:**
Trafficking has been a world-wide problem, and using 130K+ records of human trafficking data, our team aims to predict how human trafficking patterns will change over the next few years, so that the right number of resources can be allocated to help mitigate this issue.

**Research Question:**

- What is the most used tactic for abducting individuals?
- How has human trafficking trended over the last 2 decades?
- Has human trafficking leaned or shifted towards a specific group of people? (specific gender, age group, and others…)

**Initial Approach:**
Using the dataset, our team will perform a time series analysis to understand trafficking trends as well as segment out certain groups that may be at highest risk of being targeted. The team is conducting analysis through Python. We will also perform and considerate other approaches.

**Unexpected Challenges:**
We had to change the original dataset proposed to a new one due to this one having more datapoints in certain years.
This new dataset also has a holistic and better representation of global human trafficking while showing a better protection to the privacy of the victims among other benefits such as:

- To resolve the challenges in previous datasets, the team finalized on a new dataset that has 130K+ records globally. The distribution of human trafficking by year is a lot more evenly distributed than previous dataset. Dataset now also contains more detail on relationship between victim and perpetrator, region of citizenship, etc
- The data contains k-anonymity ("safety in numbers") for group privacy and differential privacy ("safety in noise") for individual privacy – working with sensitive data, we must consider the ethics behind it

- Dataset created in collaboration with Microsoft Research and the CTDC to create an accurate Global Synthetic dataset that contains data from 156,000 victims and trafficking survivors across 189 countries and territories
    - Allows for complete data that protects the privacy of individuals so their information/status is not publicly exposed
    - Preserves all statistical properties and relationships in original data

**Data model:**
Our data model is a complex one which contains 39 columns with the Year of Registration, Gender, Age Range, Majority Status, Traffick Months, Citizenship, among many other columns that provide great insights such as Means of Control (e.g. Physical Abuse, Psychological Abuse, False Promises,…), type of labor (if is Forced Labor or Sexual), type of labor (Agriculture, Construction,…), type of sex trafficking( pornography, prostitution,..), and the recruiter relationship with the victim (partner, friend, family,…), each scenario is in a separate column and we have a lot of data to analyze and find the best analysis scope and model to be analyzed.

**Data cleaning process:**
Most of the data is already cleaned but will refer to the CTDC Codebook for all data cleansing methods. On top of those we performed some extra steps such as the creation of an Index Column, replacing all values that are empty or –99 to null values, replacing the citizenship with country of exploitation, filtering accordingly to the necessary data for each model tested/proposed.

**Initial Findings:**
Working with global data due to the numerous locations drives into a lot of problems so we decided to focus first on a global scale and evaluate later if it makes sense to go down to a country level.
We did an exploratory analysis of the top countries of trafficking volume compared to the bottom countries of trafficking volume. One thing to note is that the bottom countries had a lot less data points, so we pulled a lot more countries from the bottom portion and used the top 5 countries in the top portion. First, we broke down the countries based on country of citizenship. In terms of age, both the top countries and bottom countries exhibited similar trends with the 9-17 and 30-38 age groups having the highest volume. However, the top countries had the highest volume in the 30-38 age group while the bottom countries had the highest volume in the 9-17 age group. With the gender breakdown, both sets of countries had about twice as many females getting trafficked vs male. The top countries exhibit a shorter amount of overall time trafficked with both countries having the highest volume in 0-1 year but the second highest volume in top countries was for 1-2 years while for the bottom countries it was 2-5 years.

Next, we broke down the countries by the country of exploitation. The top five countries shifted with USA remaining in the top 5 for both categories. The age breakdown reversed from the citizenship analysis with the 9-17 age group being the highest volume in the top countries while the 30-38 age group being the bottom countries. The gender ratio for both sets of countries remained about the same with females being trafficked at twice the rate of males. Lastly, both top and bottom countries exhibit the same trend as citizenship with top countries trending towards shorter trafficking times than bottom countries.

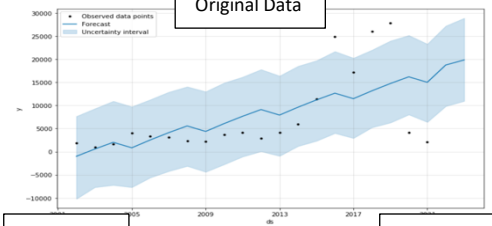
**Research:**
After the initial findings we decided to conduct even more research to start understanding what the experts are saying about Human Trafficking:

- Research has shown that human trafficking is one of the fastest growing forms of transnational crime. The clandestine nature of trafficking makes it much harder to get evidence of actual trends happening worldwide. This project will attempt to create trends based on the dataset we discovered.
- Traffickers take advantage of the unequal status of women and children in disadvantaged countries and capitalize on the cheap labor and the promotion of sex labor in some countries.
- About half of all humans trafficked are under the age of 18.
- Other big populations that are at risk of being trafficked are runaways and homeless youth as these groups don't have a strong sense of family in their normal lives.
- Within the next 10 years human trafficking is expected to surpass drug and arms trafficking in its incidence.
- The annual revenue exchanged from human trafficking is projected to total to 32 billion US dollars.

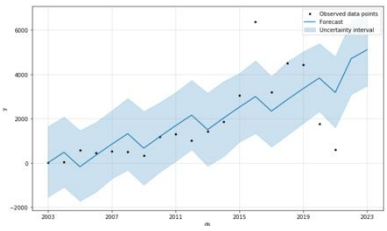
This research proves the importance of this problem in the current time and why there should be a greater focus on it.

**Prophet Model:**
Based on our initial research, our primary focus was to create a time series analysis via Prophet. The output is not favorable for extensive research to be conducted due to the date attribute only being present at the year dimension. We were able to notice this with how inputting of the data only yielded trend lines relatively and not true seasonality changes over each year. This would have been more interesting if the data points spanned across more than just ~20 years/data points. We were able to dive deeper into the model with specific trend lines (ie gender trend lines for both Male and Female categories), yet again the output was rudimentary. Going forward we do believe the data is still useful, just not in the form of a time series model. A time series model would be a lot more fruitful if the date attribute went down to at least the month level since the size of the dataset is reasonable.

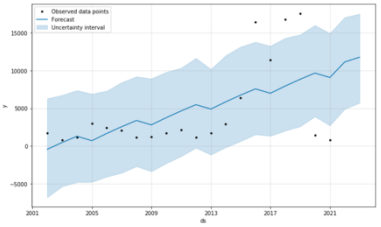
Original Data



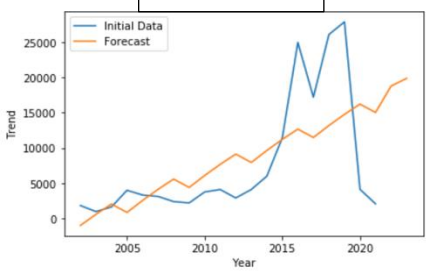

Male Data




Female Data




Forecasted Data



Plotly Interactive Capability

| country | pop | continent | lifeExp |
|---|---|---|---|
| Filter column... | Filter column... | Filter column... | Filter column... |
| Afghanistan | 31889923 | Asia | 43.828 |
| Albania | 3600523 | Europe | 76.423 |
| Algeria | 33333216 | Africa | 72.301 |
| Angola | 12420476 | Africa | 42.731 |
| Argentina | 40301927 | Americas | 75.32 |
| Australia | 20434176 | Oceania | 81.235 |
| Austria | 8199783 | Europe | 79.829 |
| Bahrain | 708573 | Asia | 75.635 |
| Bangladesh | 150448339 | Asia | 64.062 |
| Belgium | 10392226 | Europe | 79.441 |

« ‹ 1 / 15 › »

**Clustering Model:**

With this model we have not started research yet due to our early findings within the time series model. We do plan to find more insight in this model since we should be able to venture down clustering of the target groups through variables such as exploited country, gender, method of exploitation, and exploitation service. Our clustering method will most likely be KMeans instead of DBScan since we can cluster the data based on each distinct value within the attribute under examination. With this approach we believe the ability to drill down into each cluster would be very beneficial to support our initial hypothesis.
This should be capable of testing via the Data Table Interactivity function within Plotly. With this function we can filter the data and select the attribute to cluster to format for a specific environment that is wanting to be explored. The biggest hurdle for us will be the formatting of the data to fit into a clustering model with drill down capability. If the drill down capacity is not present, we still should be able to find insight from a proper grouping of the data.

**Logit Model for Global Estimates of diverse types of slavery:**

Despite diligent data collection efforts, there is a lack of dependable information on the entire population affected by modern slavery. Surveys are insufficient in estimating certain populations, particularly children in modern slavery and victims of forced commercial sexual exploitation, due to ethical concerns and stigma. While survey data is reliable for measuring adult forced labor exploitation and forced marriage, it falls short for forced commercial sexual exploitation and child labor. To complement survey data, odds ratio will be used to estimate the prevalence of these hidden populations from the CTDC dataset. A logit model will be employed, examining the correlation between the type of exploitation and factors like gender and majority status.

**Data preparation for logit model**

For this analysis we had to prepare the data to match the data available in the surveys. Surveys only contain data for male and female victims and for adults and minor.
Our dataset contains several types of genders and majority status. We limited the scope to male and female and to minor and adult to match the data available in the surveys.

For the dependent variable we transformed the null values (NaN) in the isSexualExploit and isForcedLabour columns to 0 and 1 to be able to fit the Logistic Regression correctly.

**Logit Model:**

See the model being proposed to the right:

$$\ln\left(\frac{p}{1-p}\right) = b0 + b1 * gender + b2 * majorityStatusAtExploit + b3 * gender * majorityStatusAtExploit$$

After fitting the model we obtained the next output:
For gender we have two levels: female and male. For this model female is the base case. For Majority Status we also have two levels: Adult and minor/ For this model Adult will be the base case.

| | | | | | | (b0+b1+b2+b3) | p/1-p |
|---|---|---|---|---|---|---|---|
| **Genre** | **Majority Status at Exploit** | **b0** | **b1** | **b2** | **b3** | **ln(p/1-p)** | **Odds ratio** |
| Female | Minor | 0.19016 | | 1.89411 | | 2.08427 | 8.038721 |
| Female | Adult | 0.19016 | | | | 0.19016 | 1.209443 |
| Male | Minor | 0.19016 | -4.68387 | 1.89411 | 3.19678 | 0.59718 | 1.816988 |
| Male | Adult | 0.19016 | -4.68387 | | | -4.49371 | 0.011179 |

In the last column we are reporting the obtained Odds Ratio, if the ratio is greater than 1 means that a victim matching that genre and majority status is more likely to fall to forced commercial sexual exploitation than forced labour exploitation. These are minor females, adult females, and Minor males. The only ones that are more likely to fall in forced labour exploitation are Adult males. These ratios will help us to estimate the forced commercial sexual exploitation of e.g. adult males and females by using their corresponding odds ratio. From a survey shared by the International Labour Organization (ILO) and the International Organization for Migration (IOM) in 2022 they state that Men (Adult, male) forces to labour in Thousands are equal to 10,656. By applying their odds ratio to fall in forced commercial sexual exploitation (0.011179) we get that approximately 120 men will be forced to do so.

Other great analysis that can be obtained from this logit model is how male stack vs female to be sexually exploit. By applying the next calculation **(exp(model1$coefficients[-1])-1) * 100** to each coefficient obtained in the model we can obtained the odds of being sexually commercialized. This means that the male's odds of being sexually commercialized are 99.07 smaller than female's odds of being sexually commercialized.

**Project Timeline/Planning:**

We are on track with the timeline proposed in the Project Proposal. In the upcoming weeks we will be working on further analysis and meeting with David to gather more insights into what could be done in a better way and what could be improved.

Also, the team is ahead of schedule and on a great pace to get the final model, and results by the end of next week. The power point template for the final report has been created already, which gives us an advantage to finish the project ahead of time.

**References**


- Weitzer, R. (2014). New Directions in Research on Human Trafficking. *The ANNALS of the American Academy of Political and Social Science*, *653*(1), 6–24. https://doi.org/10.1177/0002716214521562
- Clawson, Heather J., et al. "Human trafficking into and within the United States: A review of the literature." *Washington, DC: Office of the Assistant Secretary for Planning and Evaluation, US Department of Human and Heath Services. Retrieved December* 25 (2009): 2009.
- Wheaton, Elizabeth M., et al. "Economics of Human Trafficking." *International Migration*, vol. 48, no. 4, 2010, pp. 114–141, https://doi.org/10.1111/j.1468-2435.2009.00592.x.