

# Proposal Draft

Adam Dameron 2

September 24, 2022

## Abstract

This proposal is for a project which seeks to use machine learning and object oriented statistics to model international human trafficking behavior. The goal of this project is to show which aspects of trafficking are most correlated with changes in a nations development. This is for the purposes of informing law enforcement and policy makers of what characteristics to be searching for.

## 1 Introduction

Human trafficking occurs on a global scale, and has been recorded as an international concern since the year 1913 (Aromaa 2007). Many attempts have been made at defining it, yet these definitions remain unclear and vary greatly. The United Nations defines human trafficking as “recruitment, transportation, transfer, harboring or receipt of people through force, fraud or deception, with the aim of exploiting them for profit” (Raymond 2002). Human trafficking is strongly believed to occur in every country, and victimizes individuals of all ages, genders, and backgrounds (Jac-Kucharski 2012).

A major gap in the global understanding of international human trafficking, is in how human trafficking is effected by changes in a country’s economic and social development. The United Nations mentions this fact in its ”Introduction to Human Trafficking,” and describes trafficking as a ”multidimensional problem,” where a development is applicable (Kangaspunta 2008). However, there exists very little quantitative analysis into how trafficking and development are connected. This research seeks to provide more insight into what connections exist, if any.

## 2 Objective

The objective of this research is to apply advanced methods of statistics, machine learning, and geographic analysis to gain a better quantitative understanding of international human trafficking. More specifically, by using existing publicly available data, I hope to be able to create a model that can quantify

trends in human trafficking for countries of differing levels of development. This will aid law-enforcement and policy-makers in their efforts to detect and prevent human trafficking.

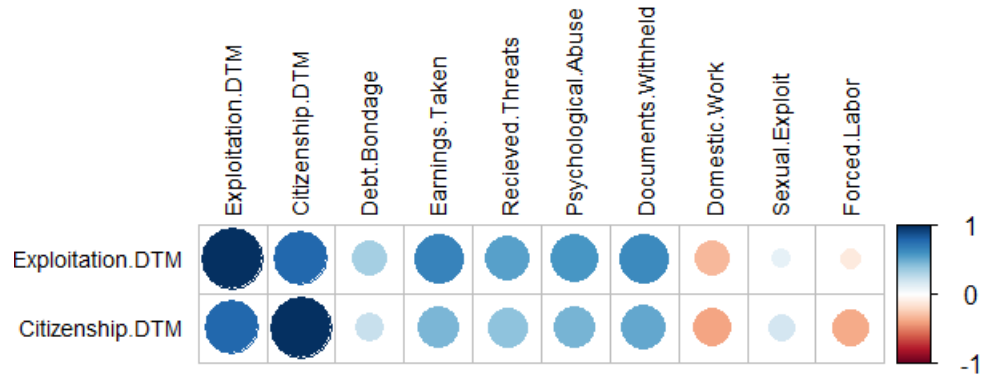
### 3 Preliminary Reasoning and Analysis

The Counter Trafficking Data Collaborative (CTDC) has a publicly available, anonymized dataset, which is a collection of over 400,000 entries for human trafficking victims (*The Global K-Anonymized Dataset* 2021). The database has plenty of useful information, such as country of citizenship, country of exploitation, and many variables expressing the type(s) of trafficking the victim experienced. This data set is what my research will be based on, as it is the largest publicly available database for instances of international human trafficking.

While the CTDC database is extremely large, it does not contain information needed for a quantitative analysis involving a country's development. For this, one needs a numerical way to categorize countries into "stages," as the goal is to make a generalized model, not one that pertains to specific countries. By far the simplest way to categorize countries would be by using the Demographic Transition Model (Bongaarts 2009). Using this model, a country can be categorized into one of 5 stages, with stage 1 being the lowest developed countries, and stage 5 being the most developed.

Unlike many other ways of classifying countries in terms of their development, this model uses birth, death, and migration rates to determine a country's stage (Bongaarts 2009). While these rates are not directly related to a country's economic growth or decline, the model has been shown to be very successful at expressing economic changes, while not being as strongly influenced by sudden, temporary economic setbacks or "booms" (Kirk 1996; Bar and Leukhina 2010; Galor and Weil 2000). The model has even been shown to help predict changes in the "shadow economy," in areas of illegal activity such as drug trafficking (Schneider 1994). It is for these reasons that I feel there is a way to efficiently model and predict changes in human trafficking as a country becomes more or less developed.

By looking at Complete entries in the dataset we are using, we can do a preliminary correlation comparison. We can take both the stage of the citizenship country, and exploitation country for a victim, and see if this correlates with any aspects of trafficking they experienced. The closer to a perfectly correlated two variables are, the more red or blue the circle will be. If a circle is red, that means the trait is seen more in low stage countries, and if it is blue, it is observed more often in high-stage countries.



All of the included variables have at least a moderate correlation with DTM, and it would be beneficial to further quantify and model what type of relationship does exist.

## 4 Methods of Research

- My Plan on how to address missing data
- Describe plans of cleaning data and how to determine which variables are best for modeling
- summarize how model will be created (methods for variable selection)
- Describe potential for bootstrapping to address missing data

## 5 Final Goals & Evaluation

- Explain how model will be tested (Train/test split) and evaluated for "goodness of fit"
- Explain significance of a good model being created for law-enforcement and lawmakers
- Explain limitations of data

## 6 Related Work

- Further analyze works existing on similar topics that validate statistical methods outlined
- Include works which inform me how to counteract bias due to missing data and how I will implement methods in my research

## 7 Technical and Time Constraints

- Timeline and feasibility for project
- Potential issues that may arise
- Lay out rough deadlines for parts of project

## References

- Aromaa, Kauko (2007). “Trafficking in Human Beings: Uniform Definitions for Better Measuring and for Effective Counter-Measures”. In: Springer, New York, NY.
- Bar, Michael and Oksana Leukhina (2010). “Demographic transition and industrial revolution: A macroeconomic investigation”. In: *Review of Economic Dynamics* 13.2, pp. 424–451. ISSN: 1094-2025. DOI: <https://doi.org/10.1016/j.red.2009.03.002>. URL: <https://www.sciencedirect.com/science/article/pii/S1094202509000210>.
- Bongaarts, John (2009). “Human population growth and the demographic transition”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 364.1532. DOI: 10.1098/rstb.2009.0137.
- Galor, Oded and David N. Weil (Sept. 2000). “Population, Technology, and Growth: From Malthusian Stagnation to the Demographic Transition and Beyond”. In: *American Economic Review* 90.4, pp. 806–828. DOI: 10.1257/aer.90.4.806. URL: <https://www.aeaweb.org/articles?id=10.1257/aer.90.4.806>.
- Jac-Kucharski, Alicja (2012). “The Determinants of Human Trafficking: A US Case Study”. In: *International Migration*.
- Kangaspunta, Kristiina (2008). *An introduction to human trafficking: Vulnerability, impact and action*. United Nations, p. 108.
- Kirk, Dudley (1996). “Demographic Transition Theory”. In: *Population Studies* 50.3, pp. 361–387. DOI: 10.1098/rstb.2009.0137.
- Raymond, Janice G. (2002). “The New UN Trafficking Protocol.” In: *Women’s Studies International Forum*.
- Schneider, Friedrich (1994). “Measuring the Size and Development of the Shadow Economy. Can the Causes Be Found and the Obstacles Be Overcome?” In: *Essays on Economic Psychology* 90.4, pp. 193–212. DOI: 10.1007/978-3-642-48621-0\_10.
- The Global K-Anonymized Dataset* (Dec. 14, 2021). URL: <https://www.ctdatacollaborative.org/global-k-anonymized-dataset>.