# **Final Data Science Project: Projecting the Migration of Unaccompanied Minors from Latin America**

## **Executive Summary**

To complete our Data Science Program, we are required to create a project presentation analyzing data using what we’ve learned start to finish. We will have six weeks to choose a project and complete it. Princess, Christina, and I formed a team called “Unaccompanied Adults Supporting Unaccompanied Minors”.

This document will explain the purpose and the scope of the project.

## **Business Objectives**

To demonstrate the mastery of our data science skills we have learned through the Data Science Program. We will use R, Python, Excel, Tableau, and other programs to wrangle, analyze, and present our findings related to datasets from the Inter-University Consortium for Political and Social Research titled, “Addressing the Surge in Unaccompanied Migrant Children” and data from the US Border and Customs Department titled, “Southwest Land Border Encounters”.

Our end goal is to visually present one aspect of the story concerning migrant unaccompanied minors and to project possible surges. We will share this story with our peers, instructors, potential employers, and other interested parties.

## **Background**

This project serves as a culmination of skills learned and practiced through the Data Science Program.

We have chosen this topic as it is of great interest to all of us. This subject is highly politicized and the population very vulnerable. It is of the utmost importance we understand the trends of migration in order to protect these children by being proactive.

## **Scope**

Christina, Princess, and Suzy will use software we’ve become familiar with in the program. We will use software of the most interest and that is the most appropriate to the task.

In relation to the actual project content, our scope is to be able to project when influxes of unaccompanied minors from Latin America migrate into the United States. This includes ages, timeframes, country of origin, and how they are designated.

## **Functional requirements**

*Data Wrangling*: We will need to combine datasets and create a sub-set with the relevant data needed. The datasets will need to be narrowed by removing the unusable columns. We will need to make sure it is formatted best for the suited analysis.

*Data Analysis*: We will familiar ourselves with the topic, datasets, and each column of given information. With this understanding, we will brainstorm questions to ask of this data in order to best tell its story and help narrow our focus to know which functions to use to create models and predictions.

*Data Visualization:* Once we have been able to properly analyze our datasets, we will need to create visuals using Tableau or other graphing programs to help present our findings in our PowerPoint Presentation.

*Presentation:* Christina, Princess, and Suzy will schedule a time to present their findings to their instructor, Margaret Martinez via Zoom. We will be dressed in a professional manner and communicate clearly in presenting our evaluation of the data. We will keep our presentation about 20 minutes.

## **Personnel requirements**

Our team (Christina, Princess, and I) will be the developers in this project. We will work to communicate regularly through text, Slack, and Zoom throughout the process in order to succeed. As conflicts, questions, concerns come, we will touch base with each other and work through them with understanding. We all have great strengths and have taken care to work through our challenges.

Once a week we will meet with our instructor to seek guidance on continuing the process.

We may also need to reach out to our other mentors if other guidance is needed.

## **Delivery schedule**

*Week 1:* We will get to know our team members and find some common times and ways we can communicate. A topic will need to be chosen and researched with appropriate datasets. We will get our Github setup.

*Week 2:* We will get set up in R and Python to study the dataset and ask questions, find possible correlations, look at distributions, and figure out which predictive models to generate and examine graphs.

*Week 3: M*odeling/Optimization (Combined Stepwise - Forward and Backward Selection) and Machine Learning (Random Forest.)

*Week 4:* Review and validate findings from the previous week, and draw insights/conclusions. Double check everything. After this week, NO MORE ANALYSIS for the sake of time and completion of the project.

*Week 5:* Compile findings into a PowerPoint presentation, review presentation with our instructor, and practice our presentation. Work on timing of the presentation and flow along with the style and layout of the presentation.

*Week 6:* Final edits to the Power Point presentation and review again..We will practice presenting at least a couple times and at least once with their instructor.

## **Other requirements**

Any programs used will be free or free to user. Considerations might be made for Tableau.

## **Assumptions**

Programs and platforms used for this project need to be updated, available, and usable. All team members will work together to complete the project.

## **Limitations**

Unforeseen circumstances during our project timeline could delay the project. If we reach a challenge in our processing, we may also be delayed.

## **Risks**

Possible risks that may occur could be severe storms causing power outages and damages, or any possible family emergencies.