**Capstone project for Springboard’s Data Science Workshop**

**By Ma. Eliza J. Villarino**

*What is the problem you want to solve?*

There is a general belief in the field of global development and humanitarian relief that having foreign language proficiency provides job seekers an edge in applying for full-time and consultancy assignments. My project aims to determine which foreign language will be most beneficial in building and sustaining a career in global development and humanitarian relief.

*Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn’t have otherwise?*

My client includes language learning centers and universities in English-speaking donor countries which offer courses focused on global development and humanitarian relief. Donor countries refer to high-income countries such as the United States, the United Kingdom, Canada, Australia and Ireland which have traditionally provided foreign aid. Foreign aid mainly funds global development and humanitarian relief positions, and these donor countries would often likely host organizations that hire people with foreign language skills, whether at headquarters or for their field projects in low-income countries.

This project provides basis for marketing: Should this project indicate that a particular language will open up a certain amount of opportunities for people, a language learning center for that particular language can use the project as a basis for promoting their language courses. In addition, this project could prompt universities in the covered English-speaking countries offering courses focused on global development and humanitarian relief to expand their curricula or improve enrollment in certain foreign language courses.

*What data are you going to use for this? How will you acquire this data?*

The job ads posted on Reliefweb will be used for this project. Reliefweb, unlike similar jobs boards that have a portion of their job ads restricted to their members, have made all their job ads publicly available.

The data will be acquired through an API search.

*In brief, outline your approach to solving this problem (knowing that this might change later).*

After extracting the data from Reliefweb, I will clean the data and create columns to indicate ads written in English and if that same ad mentions foreign languages often mentioned by published articles as to which foreign languages would be useful for English speakers in applying for global development and humanitarian relief jobs. English-written jobs will also include ads posted by organizations where the job location is in English-speaking donor countries.

As such, for each ad that mentions “English” or from an English-speaking donor country, a binary column would indicate “1.” If that same ad also mentions “French,” for instance, another binary column would indicate “1.” I will then sum up rows of these columns, with each row expected to have the sum 2 (1 for English and 1 for French. After getting the sum, divide the row sums by 2 to get the real number of jobs ads that require or prefer English speakers to have proficiency in both English and French. This number will then be divided by the total number of rows (i.e., job ads) to determine the proportion of ads that require or prefer English speakers to have proficiency in French. The same procedure will be used when determining the proportion of job ads that require or prefer English speakers that have proficiency in Arabic, Spanish, Chinese, Russian, Portuguese, Italian, local language and certain regional languages.

The results will be plotted using a histogram or a bar graph, using ggplot in R.

*What are your deliverables?*

My deliverables will include code, a paper and/or a slide deck.