Alicia Ramkissoon

2.21.2019

ETL Analysis

For this ETL project, I decided to use three sources from Kaggle, which hones in on suicide rates, gender suicide, and poverty levels. To do this, I used Pandas to read and clean my data. First, I merged two of the CSV files, then eventually merged the third. I was able clean up the data by merging the data using a common column. To make columns common, I renamed the column in Pandas. I was able to drop columns I did not need and dropped any null values. For aggregation, I wanted to find the mean of the poverty values, male: female ratio, and the total suicide rate for each country by year; this made it easier to read and to see any patterns, if any.

By doing so, I was able to see that most countries around the world have male to female suicide ratios that are above 1, which means that most countries have higher rates of women committing suicide. By reading the data, it isn’t so evident that poverty is a cause of suicide. If very well could be, but the data in this ETL project doesn’t show strong evidence for or against.