At the end of the week, your team will submit a Final Report that describes the following:

* **E**xtract: your original data sources and how the data was formatted (CSV, JSON, pgAdmin 4, etc).
* **T**ransform: what data cleaning or transformation was required.
* **L**oad: the final database, tables/collections, and why this was chosen.

Please upload the report to Github and submit a link to Bootcampspot.

Team Coocoo4Cocoa worked on looking at data concerning chocolate. All of the data we found were comma delimited. We found a csv file with data about companies with farming locations around the world which produce chocolate. From that data set we chose to take the information on the company, country in which they produce cocoa from, the percentage of cocoa within each bar of chocolate, and the rating that was given to each bar by industry experts. To pair with this data, we have a set of data with a few countries that use child labor in order to farm cocoa. The countries that were not listed were put as countries that do not utilize children as workers. The third set of data that we are combining is the weight of chocolate consumption per capita to certain countries.

In our flavors of cacao file, which held the ratings of cocoa and percentages in certain chocolates, we had to drop several columns that we unnecessary. The specific bean origin location, chocolate reference number, review date, company’s main location, and bean type were all dropped as columns with information that was not relevant. From the list of goods excel file we were able to get countries and the goods they produce along with a marker on whether or not they use child labor. We sorted through the data to pick out only cocoa producing countries and dropped anything else. With that list, we could find the countries with child labor and cocoa producers. Data from the global chocolate consumption contained two columns, both which were used. However, since the weight of chocolate was given in kilograms, and seeing as we are unique Americans with no sense of the metric system, we converted the kilograms to pounds by multiplying the amount in the columns by 2.205, the average weight of one kilogram to pounds, resulting in each listed country’s chocolate consumption per capita.

For loading the data, we used a relational database to contain all the data. We used the country names as a primary key because each database held country as a similar column; each of them containing the corresponding information concerning the child labor and chocolate consumption data. Because our data was in csv format we used SQL to create our database.

