**DOCUMENTATION**

**DATA PREPARATION**

What I understood as the business task is to identify the countries where the best cocoa beans are grown based on rating. Also, to see if there’s a correlation (if any and what kind) that exists between cocoa % and rating.

**Observations/bias**

I observed blanks under the ‘bean type’ and ‘broad bean origin’ columns. The data is incomplete.

**DATA CLEANING**

* Re-aligned the headers
* Checked for and removed duplicates
* Applied filter and checked for misspellings on all columns and removed blanks under ‘broad bean origin’ column.
* Dropped about 4 columns because they were not important for the analysis

**ANALYSIS**

VENEZUELA had the highest rated chocolate and it is where the best cocoa bean is grown. I compared the broad regions against the rating column on a pivot table. I limited the countries that produce the highest rated bars to 5.

The higher cocoa % had the lowest ratings. I believe ratings were determined also considering the broad bean origin of the cocoa bean considering the fact that venezuela did not produce the cocoa bean with the highest cocoa % but still had the highest rating.

My conclusion is that varying factors asides cocoa % such as bean type and country of origin determine the rating of a cocoa bean.