Chocolate Bar Quality Analysis

By: Shalmali Ambike

ASK

Problem to solve:

- Where are the best cocoa beans grown?
- Which countries produce the highest-rated bars?
- What's the relationship between cocoa solids percentage and rating?

Business Task: Identify which factors contribute to a higher rating of a chocolate bar by experts.

PREPARE

– The data can be found [here](#).

– The data has been made available by Rachael Tatman on Kaggle and complied by Brady Brelinski, Founding Member of the Manhattan Chocolate Society.

ABOUT THE DATASET

– This .csv dataset contains expert ratings of over 1,700 individual chocolate bars, along with information on their regional origin, percentage of cocoa, the variety of chocolate bean used and where the beans were grown.

– The Rating System:

- 5= Elite (Transcending beyond the ordinary limits)
- 4= Premium (Superior flavor development, character and style)
- 3= Satisfactory(3.0) to praiseworthy(3.75) (well made with special qualities)
- 2= Disappointing (Passable but contains at least one significant flaw)
- 1= Unpleasant (mostly unpalatable)

*The database is narrowly focused on plain dark chocolate with an aim of appreciating the flavors of the cacao when made into chocolate. The ratings do not reflect health benefits, social missions, or organic status.
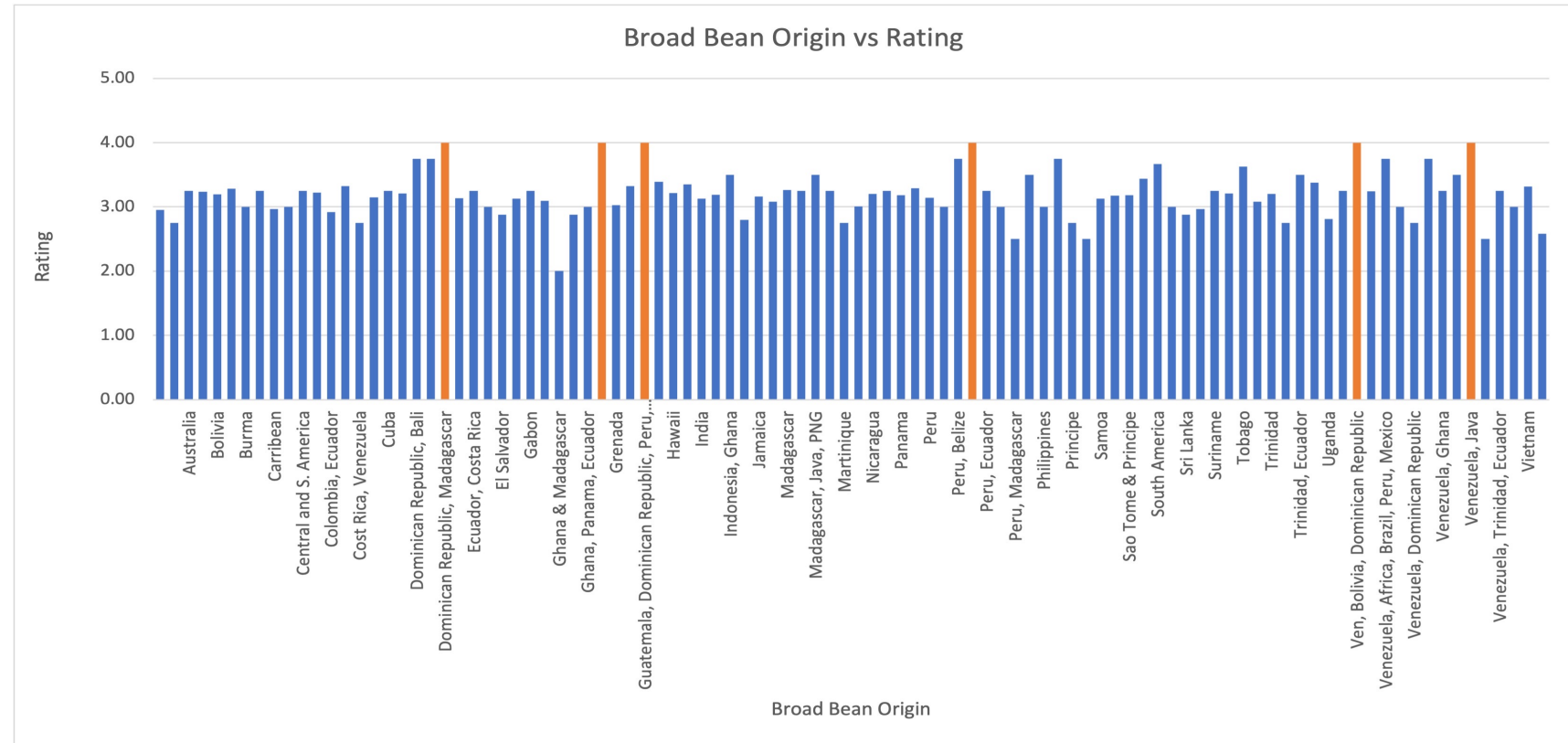
PROCESS

– The .csv file was imported into Excel. Then I explored the data checked for duplicates. I did not find any duplicates in the dataset.

– Next, I wanted to make sure that there were no leading or trailing spaces for the Cocoa Percent, Company Location, Rating and Broad Bean Origin columns. To ensure this I used the TRIM function on all the values of these columns. Since TRIM only returns text values, I had to re-convert values of Cocoa Percent and Rating into Percentage and Number respectively.

– Next, I filtered the data to check for misspellings in any of the text columns and to check for blank rows. I did not find any misspellings entire or blank rows. The data did, however, have blank entries for Broad Bean Origin column which I counted using COUNTBLANK function.
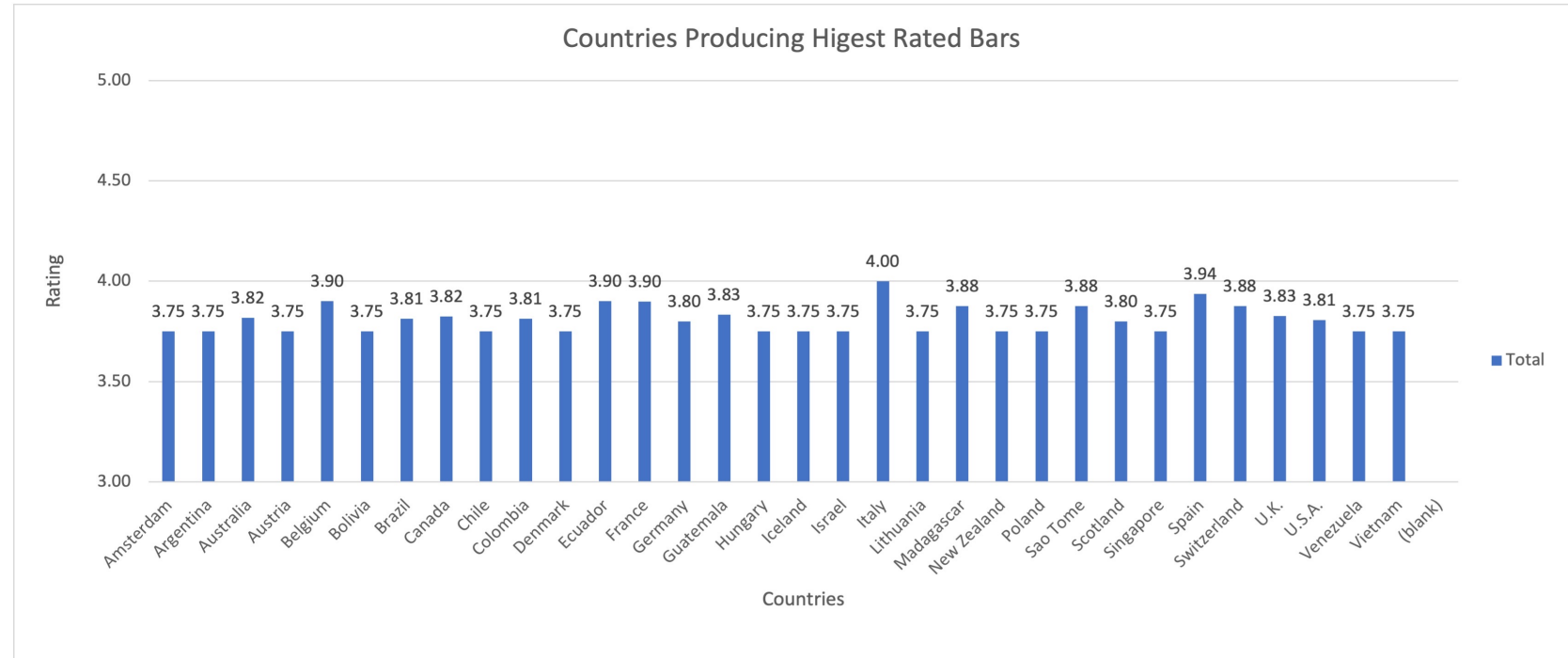
PROCESS cont.

– Now, I wanted to use the Broad Bean Origin column in my analysis but upon further inspection, I found out that the entries in that column were not standardized. For example, the country of Dominican Republic had entries by the name of D.R, DR, Dom Rep etc.

– To standardize the values in this column, I used the excel function UNIQUE to see how many distinct values were there in the column.

– After that, I used the Find and Replace function in excel to find different names of same value and standardized those names so that they could be used for analysis.

– Lastly, I made sure all the columns were assigned proper datatypes according to their values.

- Where are the best cocoa beans grown?
  - In order to determine the geographical location of the best cocoa beans, I made a pivot table with the column broad bean origin which is the broad geo region for the bean origin and calculated the average bar rating for each of those locations.
  - After plotting them on a bar chart there are 6 locations where we can find highest rated bars which are highlighted in the chart.
  - From the data, the best cocoa beans are grown in Dominican Republic, Madagascar, Guatemala, Peru, Papua New Guinea, Bolivia, Venezuela, Belize and Java.
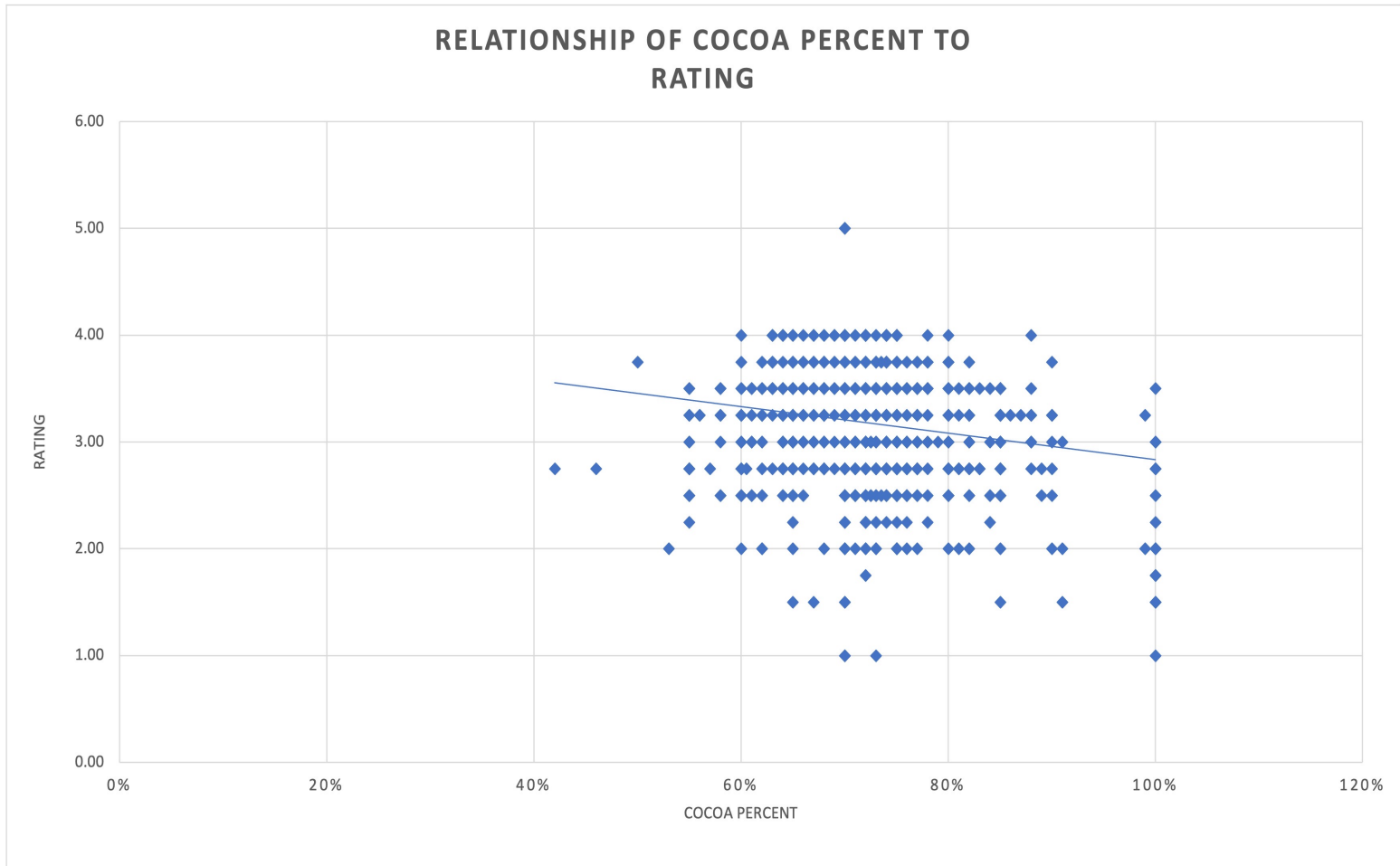

Broad Bean Origin vs Rating

– Which countries produce the highest rated bars?

   – To find this I filtered out all the countries with bar ratings greater than 3.75.

   – From this data, I constructed a pivot table where I calculated average of the highly rated bars from those countries.

   – We can see from the data that Italy followed by Spain, Belgium, Ecuador and France are the countries with highest rated bars.

– What is the relationship between cocoa solids percent and rating?

  – I plotted cocoa percent vs rating as a scatter plot to see the relationship between the two variables.

  – I also calculated the correlation coefficient for the two variables using the CORREL function which returned a value of -0.16.

  – This indicates that there is negligible relationship between the two variables.

  – We can hence conclude that cocoa percent has no bearing on the rating of the bar.

RELATIONSHIP OF COCOA PERCENT TO RATING

CONCLUSION

– In conclusion,

   – The best cocoa beans are grown in Dominican Republic, Madagascar, Guatemala, Peru, Papua New Guinea, Bolivia, Venezuela, Belize and Java.

   – Italy and Spain are the countries that produce highest rated bars.

   – Cocoa Percent does not affect the rating of a chocolate bar.