**Read me**

To start this analysis, you will need to download this dataset as a csv: [Chocolate Bar Ratings](https://www.kaggle.com/datasets/rtatman/chocolate-bar-ratings).

You will then have to open the dataset into Excel to do some data preparation.

By searching Unique values in “Company Location” all countries can then be sorted into one of six regions: Asia, Europe, North America, South America, Oceania and Africa. This will be used for one of the bar charts in the blog.

In Excel, both Interactive maps can be made.

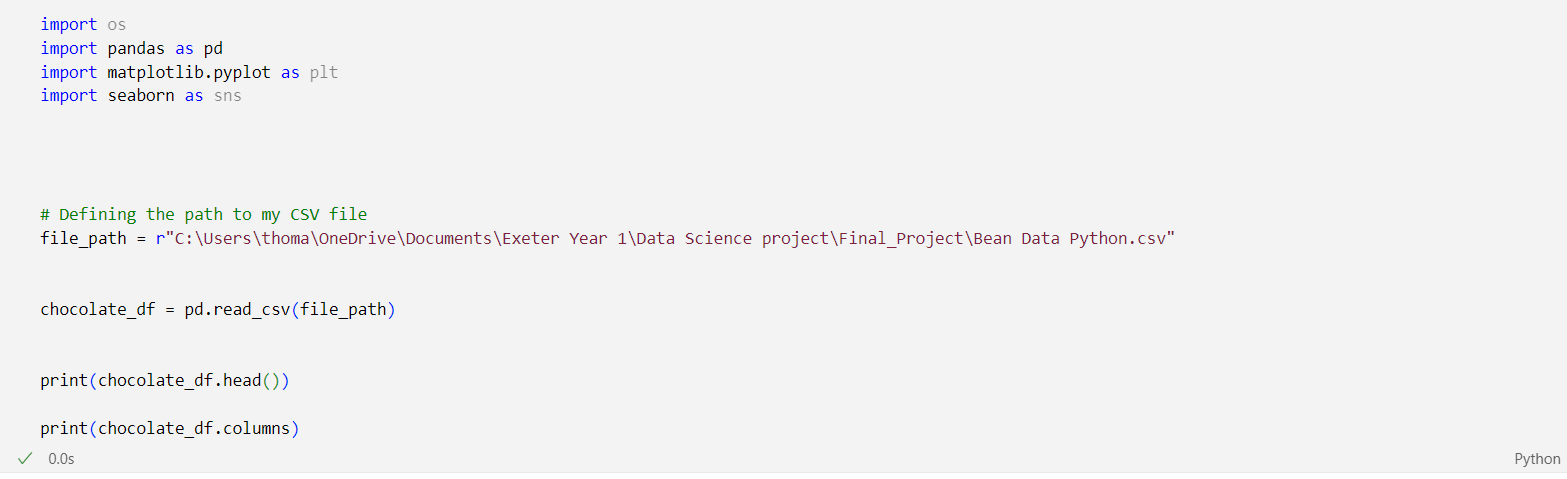
For the first map: Interactive map of average rating by country, we need to create a pivot table selecting Company Location as columns and Rating as Values. To get the Average of rating you will have to select averages by going into Value Field Settings. Once this is done the pivot table should appear. From here to create an interactive map you will need to copy the values into another sheet, pasting the values there. Then clicking on insert, you should see a Map icon which will create the map. You can then change the style to make it look like the one in my blog. Sadly, when extracting the map into VS code, it can no longer be interactive (when hovering over a country you see the average rating).

For the second map: We will need to do the same steps, however inserting Broad Bean Origin into the pivot table instead of Company Region. Following on from here, the same steps will need to be taken.

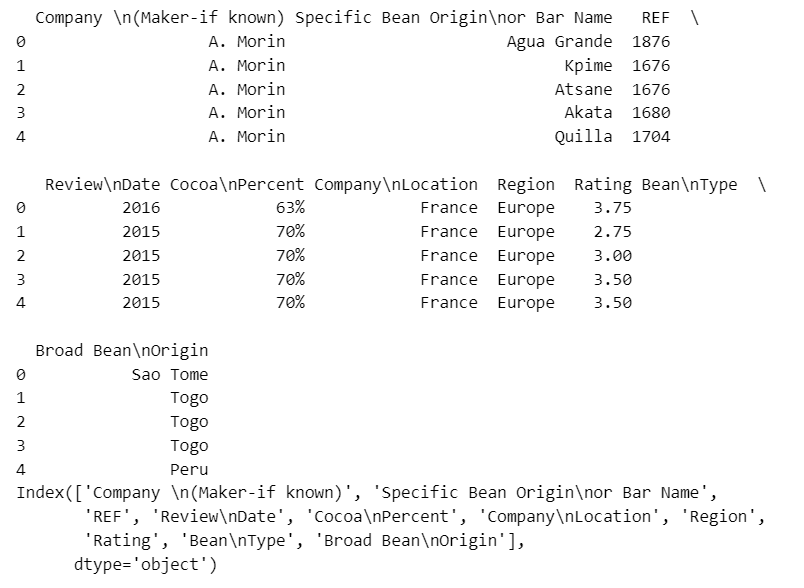
The highest rated chocolate bar can also easily be found by doing filter and find values. There is only one bar rated 5.

Loading the data into VS code:

You will need to run this code- inserting the adequate path to your csv file:



This should be the output:



Now our data is loaded into VSCode!

We can now create visualisations.

For the first plot, the distribution plot, you will need to execute this code:

A screenshot of a computer program

AI-generated content may be incorrect.

This should give this output:

A graph of chocolate bars

AI-generated content may be incorrect.

For the second bar chart, average bar rating by region, you will need to execute the following code:

A screenshot of a computer program

AI-generated content may be incorrect.

If the data preparation was done sufficiently well, you will not need lines 27 to 30. I however, named some South American countries as Latin America so had to combine them in VS Code. Using the groupby and mean functions, we can therefore create the average bar rating by region.

The average values all lay between 3.0 and 3.5, so on the y-axis I modified the range. This I understand can be misleading to the differences in regions average, however I believe it is more visually appealing and still gets the general point across as to which region has the highest average rated dark chocolate.

This should be the seen output:

A graph of a bar graph

AI-generated content may be incorrect.

To further the analysis and add some additional data on cacao bean prices, we will need to do some web scraping. The data I found is from this website https://www.indexmundi.com/commodities/?commodity=cocoa-beans&months=240

To successfully web scrape the data and load it into VS Code as a dataframe you will need to run this code:

A screenshot of a computer program

AI-generated content may be incorrect.

The output should be the following:

A screenshot of a computer screen

AI-generated content may be incorrect.

I then wanted to transfer this data into SQL to be able to manipulate it there and create clear tables for my blog.

To do so I used the following code:

A screenshot of a computer

AI-generated content may be incorrect.

You will need to modify the path to the correct one on your computer. Here, the double back slashes are important. Similarly, index=False is very important as when importing to SQL it automatically adds a column named index, this can easily block the code from working. You should now see the database in SQL like so:

A screenshot of a computer

AI-generated content may be incorrect.

In order to obtain the two tables in the blog you will need to run the following code in Execute SQL:

A screenshot of a computer program

AI-generated content may be incorrect.

The first section uses a substr to create a column with just the years, and then creates another column average\_price and orders them in descending from 2024 downwards. This should output the following:

A screenshot of a computer

AI-generated content may be incorrect.

The second part selects the month with the maximum percentage increase. This should output:

A screenshot of a computer

AI-generated content may be incorrect.

Lastly, I wanted to create a predictive model for future prices. To replicate this, you will need to recreate the following code in VS Code (it is split as it was too long):

A screenshot of a computer program

AI-generated content may be incorrect.

A screenshot of a computer program

AI-generated content may be incorrect.

A screen shot of a computer code

AI-generated content may be incorrect.

This should create the predictive random forest model, and plot the following graph:

A graph with blue and orange lines

AI-generated content may be incorrect.