Data Analytics

Chocolate Varieties though data analysist

Anh Huynh

December

Table of content

Introduction

Overview

Chocolate is paramount in our society, it is a food product made from cocoa beans, consumed as a solid, liquid, paste and use to make beverages in bakery products and other foods. Today chocolate industry estimated to be an USD 130.56 billion-dollar industry. The cacao tree was cultivated more than 3000 years ago by the Maya, Toltec and Aztec people. They use cocoa bean to prepare a beverage as a ceremonial drink, and also used the beans as a currency. The Maya considered chocolate to be the food of gods, held the cacao tree to be sacred and even buried dignitaries with bowls along with other items deemed useful in the afterlife. In fact, the word cacao (*ka-ka-w*) belongs to the Maya's phonetic manner of writing. In 1502, Christopher Columbus took cocoa beans to Spain after his fourth voyage. Spain was the earliest European country to incorporate chocolate into its cuisine. The first solid chocolate bar put into production was made by J. S. Fry & Sons of Bristol, England in 1847. Meanwhile, the making of chocolate spread overseas and grew in sophistication. Each chocolate is evaluated from a combination of both objective qualities and subjective interpretation. A rating represents an experience with one bar from different people with different perspectives. These are the basic ingredients in chocolate: Cacao beans, Cocoa Butter, Sugar, Milk Powder and some others ingredients (in milk and white chocolates). However, not all chocolate bars are created equal.

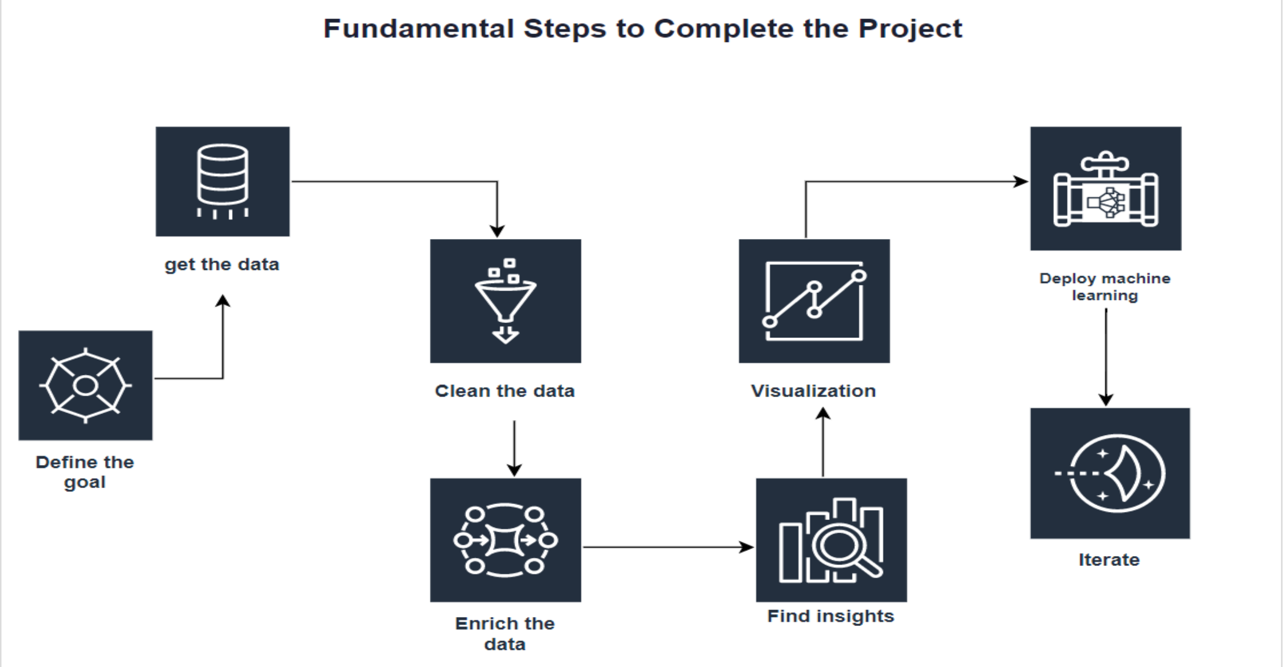
Project Main goals:

The main dataset that I have chosen is contains the ratings of over 2,500 chocolate bars from around the globe to apply in my project. The goal of this analyst is focused on rating on different chocolate bars with an aim of appreciating core different categories around chocolate type. Each chocolate is evaluated from a combination of both objective qualities and subjective interpretation. Overall opinion is really where the ratings reflect a subjective opinion. With this dataset, I want to be able to find out the correlation between displayed categories on rating, to explain the questions:

- Does the manufacture of producing chocolate make the high rating?
- Does the bean origin affect people rating?
- The ingredients make different chocolates effect on rating?
- Does cocoa percentage or ingredients affect the rating?
- What are the most memorable characteristics that made impressions on people and make them remember about each chocolate.

To better understand the topic of chocolate bar rating, I import two more datasets for showing more insights more explanations and to support for the main dataset. After retrieving the necessary information and raw data I need, I try to receive an overview of data I have imported, clean them and perform exploratory data analysis. Then, I create an entity-relationship model and explore some queries with MySQL in order to show the insights. Last, suitable machine learning model will be deployed on the main dataset to predict the future target.

Here are my summary steps to finish the project base on diagrammed model:



Data Descriptions

The main dataset name *'chocolate rating'* is analysed throughout this project has 2530 lines and 10 columns before cleaning. It bases on the rating of more than 2500 different types of chocolate from over the world. Belongs, others information is about the cacao origins, manufactures, cocoa percentage, ingredients of each chocolate, rating time, rating and comments from people. Rating system in this dataset is a range from 1 to 4, also the important element which go through all of the project including:

- 3.5 - 4 = Outstanding
- 3.0 - 3.49 = Recommended
- 2.0 - 2.9 = Disappointing
- 1.0 - 1.9 = Unpleasant

One of crucial elements is the ingredients that made each chocolate bar which are mentioned in one of the columns (Ingredients) are: B (Beans), S (Sugar), S* (Beet sugar), C (Cocoa Butter), (V) Vanilla, (L) L ecithin, Sa (Salt). Those characteristics are very important in giving some explanations and relations to rating results. In addition, the location of company and the cacao bean geography play a crucial part during this data analysis process.

On the other hand, to enrich the dataset and have more key information, I had found two more datasets to support for the main dataset and for my project processing. The second one 's name ***'**chocolate flavour'* which has not only had most of similar points with main dataset but also has another useful data to brings some new insights. To be more detail, I will explain more about how to apply the second dataset in later part about creation database and data importation with MySQL. The third data is about *'countries and continents'*, implemented to enrich main dataset, will be mentioned more detail in the part of data cleaning description.

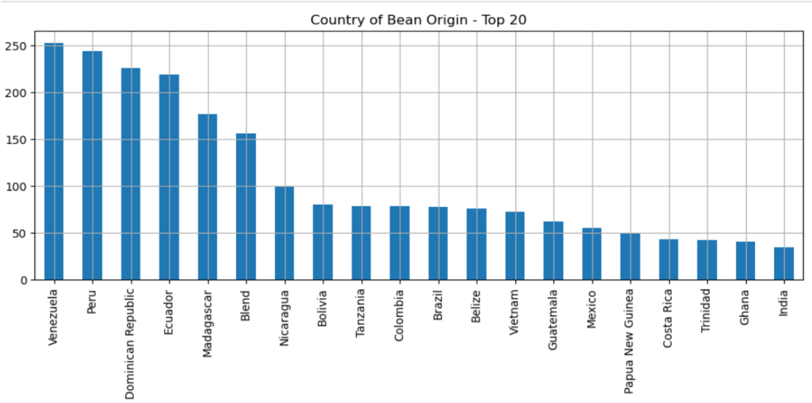
Data collection

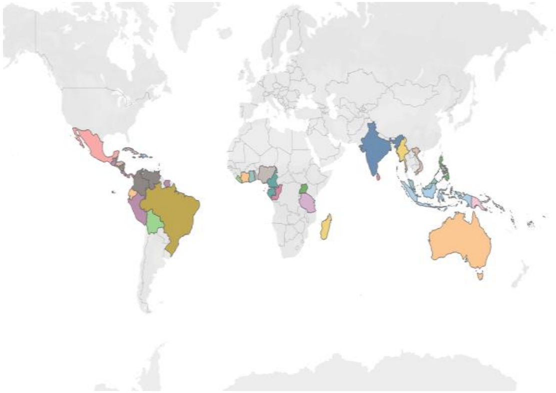
The source of data is collected from Kaggle, the world' largest data science community with a lot of useful resources and powerful tool for data science. The dataset described in the section before have been republished on Kaggle and I download them on my local machine. Then I imported them to Jupiter notebook using Pandas read csv function. From here I start to clean the data, enrich data, explore in visualization and do machine learning. Beside Kaggle, to understand correctly my data sets, I have read some articles about chocolate from www.britannica.com, www.chocolate.co.uk, https://damecacao.com, https://barandcocoa.com, to have an overview look on the ingredients in chocolate, the different varieties of cocoa beans and its history.

Data cleaning and Exploratory data analysis

Firstly, I imported all of the necessary libraries I need to work with on my Jupiter notebook such as: NumPy, pandas and matplotlib, seaborn for visualizations etc. Regarding the workflow of my data cleaning and my purpose of analyzing, I decided to focus strongly on the main dataset which I will use throughout the process. The first step after import dataset is always to have an idea on what the dataset show generally by using read_csv then use. head() function. Then reading data with .info(), checking the shape of dataframe, looking of different types of data in each columns and observing column names. Secondly, I use function .unique to check if the data have some problems with duplicate, uppercase, lowercase, unnecessary space and special characteristic then replace them in cleaned version. Next, find out the missing value with .isna().sum() and fill them with suitable values. In my dataset, there are 87 missing values in the column 'Ingredients'. To fill the missing value, I use function. mode() to fill on the most frequently values for the missing parts. In addition, there is some missing value that need to solve base on the column Cocoa percent who located in front of column 'Ingredients'. And for the second and third dataset, I also clean with a very similar way repeatedly at later process but less focus than the main data. I drop the columns that are not useful for my analyzing by function .drop('column', axis=1, inplace=True). Overall, the cleaning will not finish at the first part but it will go through all the analyzing depends on the process.

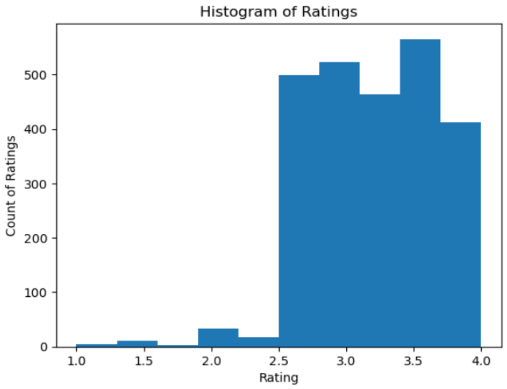
Secondly, to understand the indications of data, I have explored information on charts by visualizations use seaborn and matplotlib. First, I want to know the top 20 countries were well known as origin of cacao bean by using .value_counts() and I got the result is Venezuela is product the most cacao bean, then is Peru and dominican Republic go just after. Also, I import the data on tableau to see where are those bean country's locations. Here are results from matplotlib and tableau

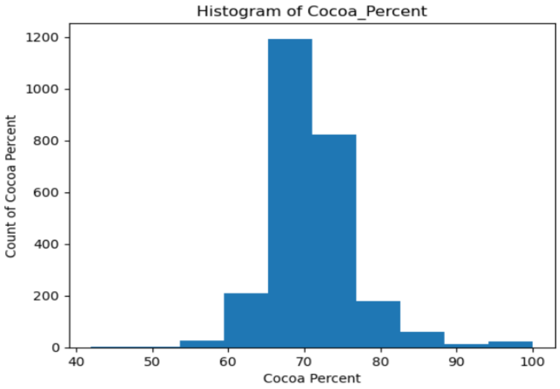




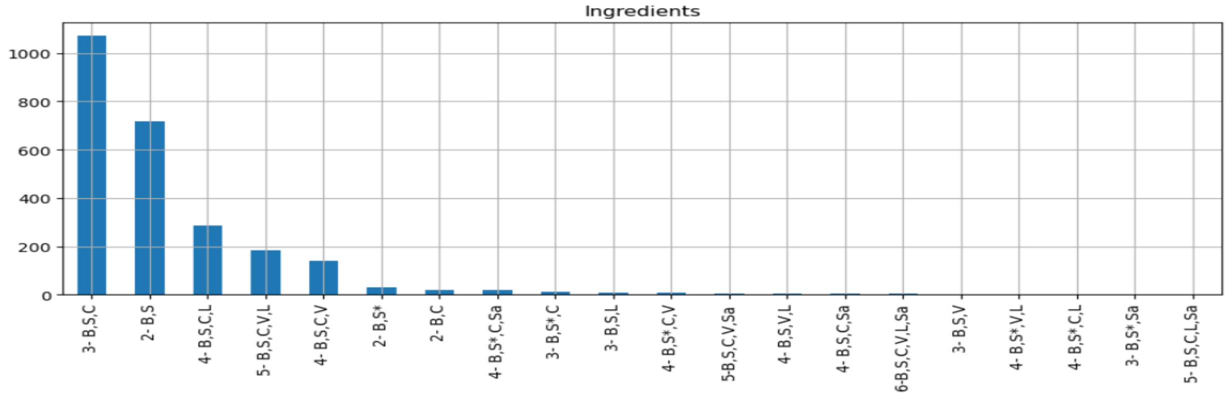
Next, I would like to know where have the top 10 companies who produce chocolate by the same function (list(.value_count()[:10].index), the result lead by USA, then Canada and France, follow-up are Italy, Belgium, Ecuador, Australia, Switzerland and Germany. Here we can see that the places that cocoa beans come from are not at all concern with where the chocolate are made and deliver to consumers. With the same function of code, list of top manufacturer are Soma, Fresco, A.morin, Dandelion, Pralus…Zotter.

Regarding to charts below, the rating of this dataset is mostly from 2.5 to 4 which is generally high. Still, some ratings are very low from 1 to 1.5 and 2 to 2.5 but they are minorities. Besides, the majority of chocolate bars have cocoa percent from 65 % to 75 % can take into account of the given data. It seems like the companies produce more chocolate bar that have cocoa percentage at 65 % to 75 % than others which have percentage less than 65 % and higher than 75 %. That is also an important insight for business cases to explain where is the high demand and may relate to companies' profit.

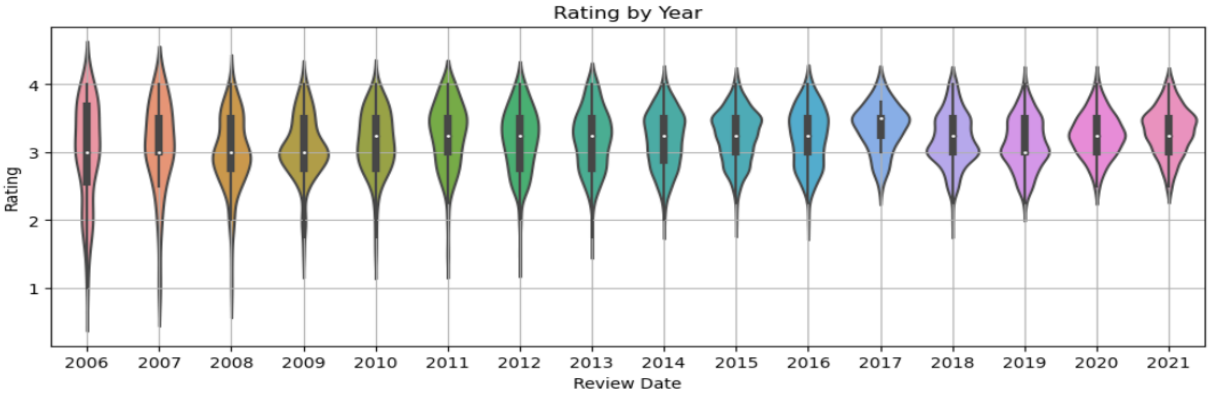




Now, let see what is ingredients which appear the most in the chocolate. From the dataset, below figure mostly focus on the ingredients with 3 categories: B,S,C (bean, sugar, cocoa butter). The second place is belonging to chocolate have 2-B, S (bean and sugar). The least is 5-B, S,C,L,Sa (bean, sugar, cocoa butter, lecithin and salt). Parallel with percentage of cocoa, the ingredients are also depended on how much makers add in each chocolate bar. In fact, we can tell that the highest grade of ingredients is most likely in range of 65% to 75% cocoa

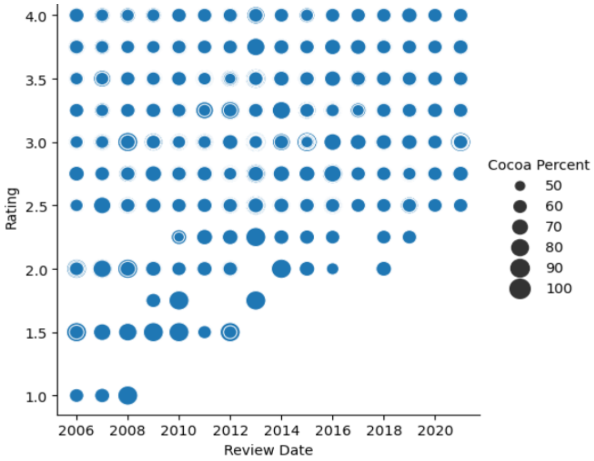


Thirly, I want to see what is the effect of each factor on the rating score. Here are several charts to indicate their correlation with rating. Rating by year from 2006-2021 give conclusion that over time the average rating isn't increasing but the instances of extreme ratings is decreasing, resulting in a smaller spread of ratings. The latest rating scores seem be the less different comparation of people thinking.



The country of manufacture and bean origin don't appear to have a large impact on rating, but both the maker and company can be seen to impact the average rating of a chocolate. This notebook showed that North America has a huge industry for the chocolate, while the Cacao beans are from Africa and South America countries. Since there are massive number of different origins and many origins contained more than one country, we will plot the rating and cocoa Bean Origins. We observed that Haiti has the highest quality cocoa beans. Furthermore, Venezuela, Madagascar and Brazil also produce high quantity and quality cocoa beans.

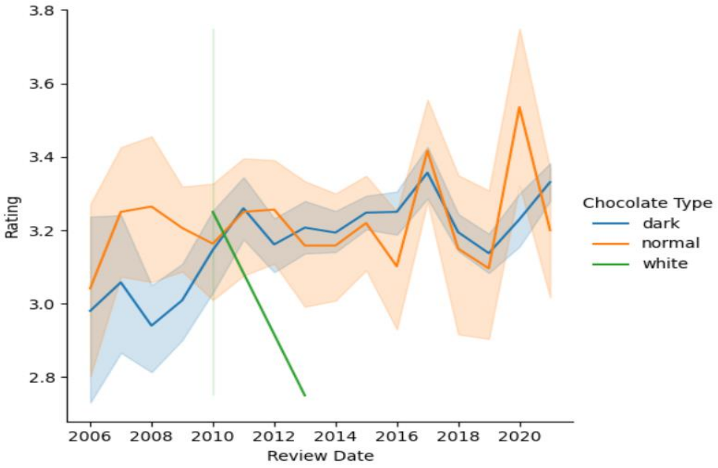
From this graph we can see when cocoa: 50% we still have the highest rating. However, the case is not persuasive enough, because the simple space size of 50% cocoa chocolate bars is not large enough. Consider the sample sizes into account, 70% cocoa bar still our best choice. Moreover, any chocolate bars from 65% - 75% percent has highest score. Furthermore, Rating base on ingredient has significant differents depends on what maker add in a chocolate bar.



As the visuallisation parts has showed in my Jupiter notebook, the top rating from 3-4 belongs to group of (B,S,C), (B,S) and (B,S,C,L). But it seems like the chocolate with ingredient B, C has the lowest rating which is black chocolate with include cocoa bean and cocoa butter. In conclusion, the high amount of chocolate type is produced are also the one has highest score in rating and small amount type of chocolate is also have very low rating chocolate.

Further exploratory data analysis

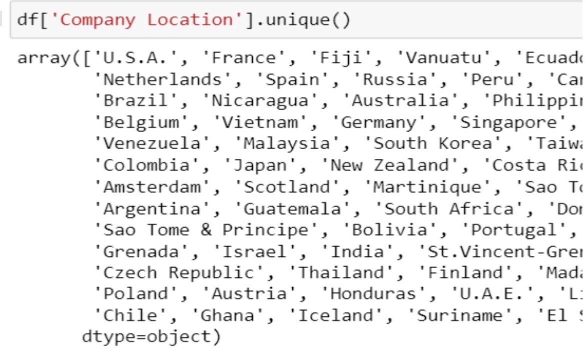
The previous part is to have general look all every elements stated in the main dataframe together with data cleaning and figures. Then, I like to enrich my data by create new columns base on existed database and to combine with another dataset to have more deeply point of view. From the column ‘Cocoa percent’ , I create a new column: ‘Chocolate type’ by using the funtion .apply(lambda) with condition statement if to separate into 3 different kinds of chocolate: ‘Dark, normal and white’.



I have set up condition: 'dark' if x>=70 else 'normal' if x < 70 and x > 50 else 'white', then I compare with the rating overtime(chart below). Here we can see, the liking for dark chocolates have increased year by year while normal chocolate it still increased in generally bu white chocolate seem to have less likely overtime.

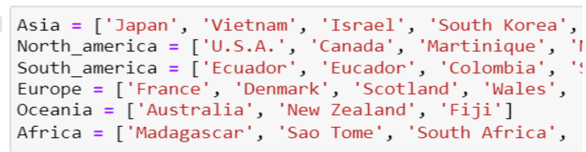
After, I want to create another column name: ‘Continents’ to apply in my main dataframe beside bean origin countries and company location. I imported another dataset about countries and continents which has 249 rows and 9 columns. My purpose here is to do left join the column of continents to my main database to have more useful information. The problem occurs when it also has some missing value and I need to fix it by searching more information on internet to find out and then use funtion. replace (np.nan,…) to fill the missing values. Then, second problem is some name of countries in 2 data sets are written differently so they gave some nan value because they are not matched. Then, I need to change the name in 1 of the table to have all of matching value.

Afterall, I found out another method than may easier instead of import another dataset. From the first dataset, I do function unique() for ‘Company Location', it will show me all of the countries and then I regroup them in six categories of continents: Asia, Africa, North America, South America, Europe and Oceania. Next step, I apply condition statement if and return function like the screenshot below. Then I has new column continents base on the countries. The goal of this work is to have simpler look on the dataset because there is a lot of information of locations that show in more than 2500 lines, also it is much easier to compare and divide for data preparation.

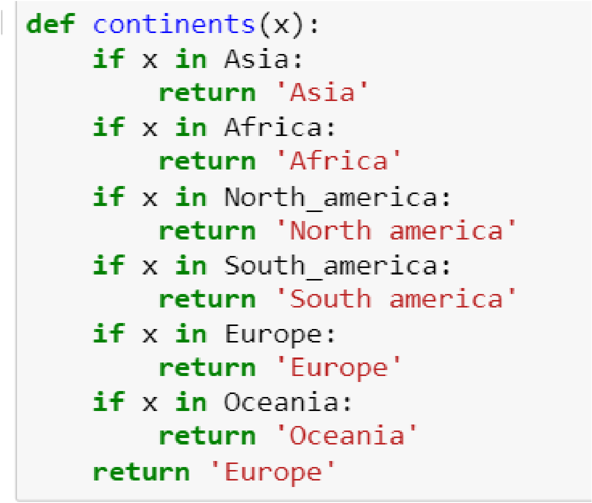


```
df['Company Location'].unique()

array(['U.S.A.', 'France', 'Fiji', 'Vanuatu', 'Ecuado
       'Netherlands', 'Spain', 'Russia', 'Peru', 'Ca
       'Brazil', 'Nicaragua', 'Australia', 'Philippi
       'Belgium', 'Vietnam', 'Germany', 'Singapore',
       'Venezuela', 'Malaysia', 'South Korea', 'Taiwa
       'Colombia', 'Japan', 'New Zealand', 'Costa Ri
       'Amsterdam', 'Scotland', 'Martinique', 'Sao T
       'Argentina', 'Guatemala', 'South Africa', 'Do
       'Sao Tome & Principe', 'Bolivia', 'Portugal',
       'Grenada', 'Israel', 'India', 'St.Vincent-Gre
       'Czech Republic', 'Thailand', 'Finland', 'Mad
       'Poland', 'Austria', 'Honduras', 'U.A.E.', 'L
       'Chile', 'Ghana', 'Iceland', 'Suriname', 'El
      dtype=object)
```

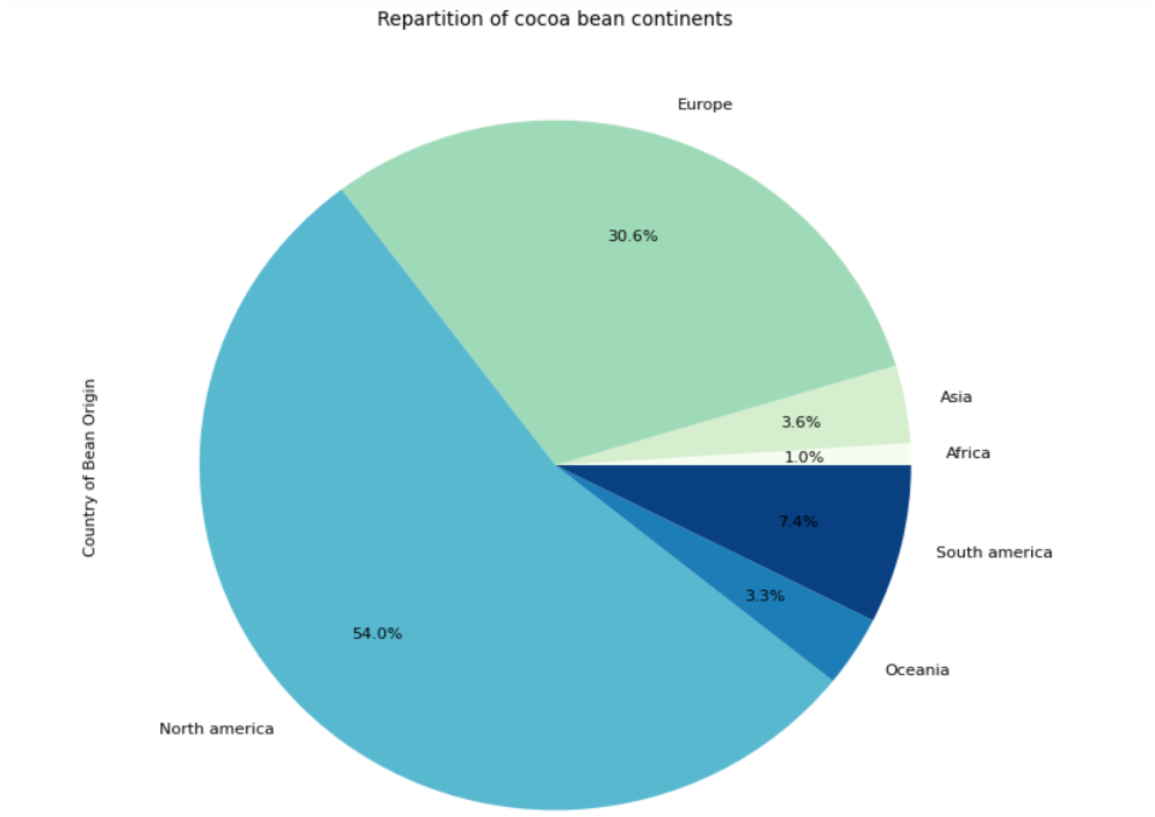


```
Asia = ['Japan', 'Vietnam', 'Israel', 'South Korea',
North_america = ['U.S.A.', 'Canada', 'Martinique', '
South_america = ['Ecuador', 'Eucador', 'Colombia', '
Europe = ['France', 'Denmark', 'Scotland', 'Wales',
Oceania = ['Australia', 'New Zealand', 'Fiji']
Africa = ['Madagascar', 'Sao Tome', 'South Africa',
```



```
def continents(x):
    if x in Asia:
        return 'Asia'
    if x in Africa:
        return 'Africa'
    if x in North_america:
        return 'North america'
    if x in South_america:
        return 'South america'
    if x in Europe:
        return 'Europe'
    if x in Oceania:
        return 'Oceania'
    return 'Europe'
```

After create new category for my data, I compare it with the people rating to what can show out for more understanding. It is interesting to see that half of repartition of cocoa bean origin from south America (54%), the second place is Europe, third is south America. If we see this graph, we can make out that this dataset is not very diversity and could be biased towards the European's and the North American's consumers.



Deciding on what type of database use

Follow my plan, the next step is creating my database, I decided to use SQL because this logistic allow me to import my cleaned and converted data frame in structured and I can create and inform my queries. The five critical different between SQL and NoSQL are:

SQL	NoSQL
- Relational - Use structured query language and predefined schema - Can work with smaller amount of data - Are table-based, multi-row transaction - OOP unfriendly (object-oriented programing)	- Non-relational - Have dynamic schemas for unstructured or semi structured data - Can work with big amount of data - Document key-value, graph, wide-column - OOP friendly

Furthermore, I work on different tables at the same time by using different queries like group by, order by, select, average, max, min, etc…Moreover, by using primary key and foreign key, it allows me to link and join quicky different data in various table just by making queries. In order to prepare my database, I create the following entity relationship model to clarify the link and the process of my work.

Entities. ERD

The following entities represent the application of my data processing. For the database I imported two data sets: first is my cleaned main data (chocolate rating), the second is new one (chocolate flavor) which have mostly similar categories. The second data is also need to clean before import on SQL because of some missing value. Then, I create some queries to get some insights in each dataset and between the two data. After create the database I create four tasks to explore more useful information that was not found in the previous process. I rename chocolate rating as: choco1 and chocolate flavor: as choco2 to have more simple name for 2 tables in SQL

The goal to in indicated quick answer with SQL:

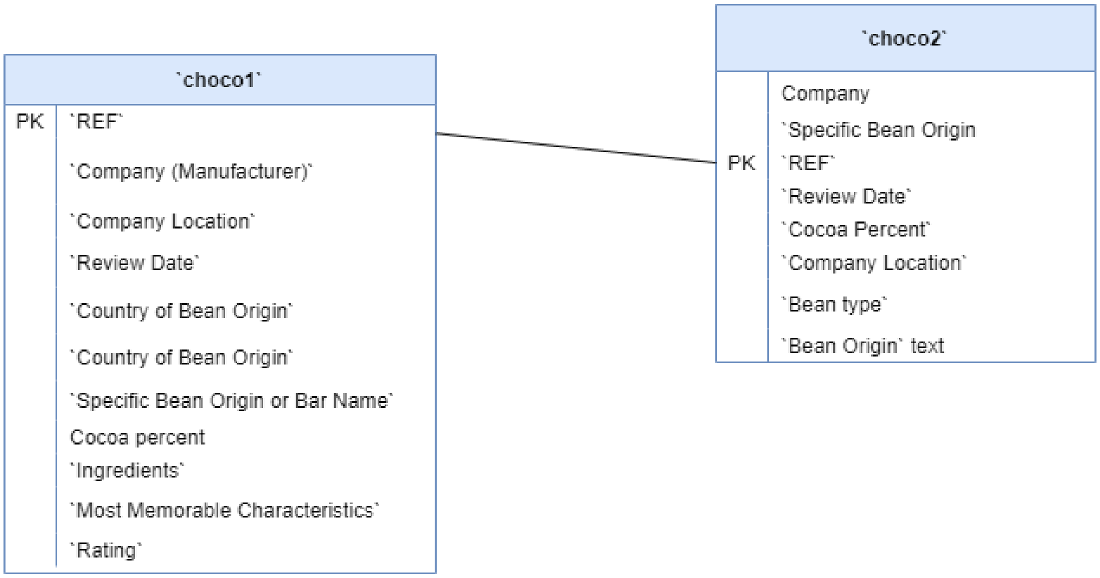
Task 1: Select the top 10 Country of Bean Origin, company base on Rating

Task 2: Find the highest, lowest and average rating base on Cocoa percent in choco1

Task 3: Add the column ingredients from choco1 to choco2

Task 4: Select company, cocoa percent and company location from choco2, calculate average rating from choco1 and join to choco2 include ingredients

Task 5: Description relation between country of bean origin(choco1) and bean type(choco2) base on rating(choco1)



	`choco1`
PK	`REF`
	`Company (Manufacturer)`
	`Company Location`
	`Review Date`
	`Country of Bean Origin`
	`Country of Bean Origin`
	`Specific Bean Origin or Bar Name`
	Cocoa percent
	`Ingredients`
	`Most Memorable Characteristics`
	`Rating`

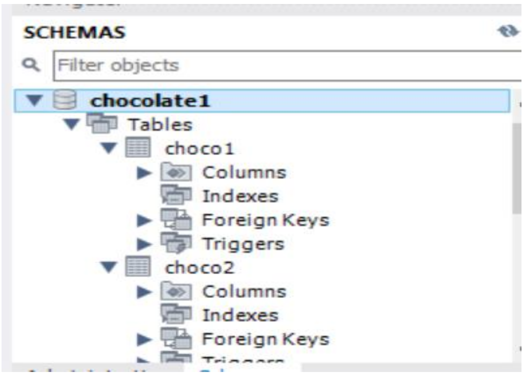
	`choco2`
	Company
	`Specific Bean Origin
PK	`REF`
	`Review Date`
	`Cocoa Percent`
	`Company Location`
	`Bean type`
	`Bean Origin` text

Database creation and data importation

After deciding on which type of database to use, I began creating my relational database on MySQL workbench with "create database if not exists":

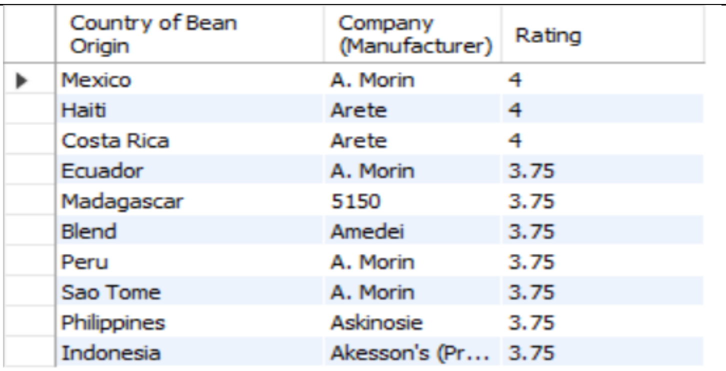
- `create database if not exists CHOCOLATE1;`
- `use CHOCOLATE1;`

Then I upload 2 table choco1 and choco2 by clicking on "*Table Data Import Wizard*". After import successfully, I start to make some queries to response for 5 tasks above in order to receive various insights from my data.



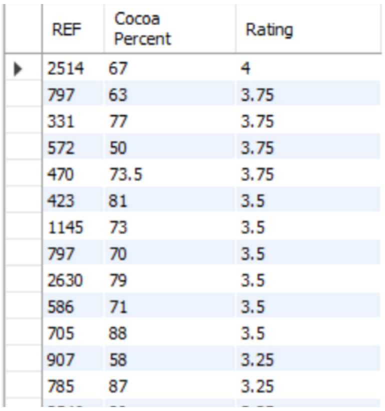
Insights

Task 1 is to find out top 10 Country of Bean Origin, company base on Rating in table choco1. As in the result, Mexico, Haiti and Costa Rica got the maximum rate at 4. The next are Ecuador, Madagascar, Blend, Peru, Sao Tome, Philippine and Indonesia. Hence, most of the bean origin have high rating come from south Americas and Asia. The brand is also playing an important role in rating score, A. Morin has the best score in the top 10 and repeat 3 times on the top of the list.

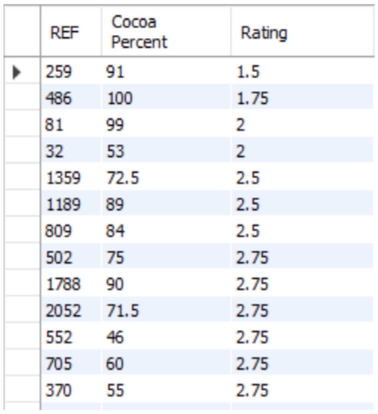


	Country of Bean Origin	Company (Manufacturer)	Rating
▶	Mexico	A. Morin	4
	Haiti	Arete	4
	Costa Rica	Arete	4
	Ecuador	A. Morin	3.75
	Madagascar	5150	3.75
	Blend	Amedei	3.75
	Peru	A. Morin	3.75
	Sao Tome	A. Morin	3.75
	Philippines	Askinosie	3.75
	Indonesia	Akesson's (Pr...	3.75

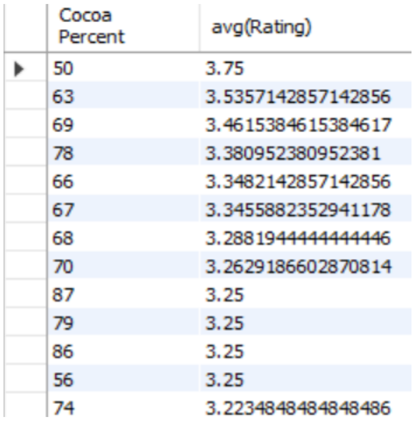
Task 2 is to find out the highest , lowest and average rating base on Cocoa percent in choco1. As we can see that the highest rating with 67 percent of cocoa with the maximum value, lowest rating belongs to 91 and 100 percent of cocoa. Then, I calculate the average of rating and range from 50 to 74 percentage of cocoa have stable and high rating.



	REF	Cocoa Percent	Rating
▶	2514	67	4
	797	63	3.75
	331	77	3.75
	572	50	3.75
	470	73.5	3.75
	423	81	3.5
	1145	73	3.5
	797	70	3.5
	2630	79	3.5
	586	71	3.5
	705	88	3.5
	907	58	3.25
	785	87	3.25

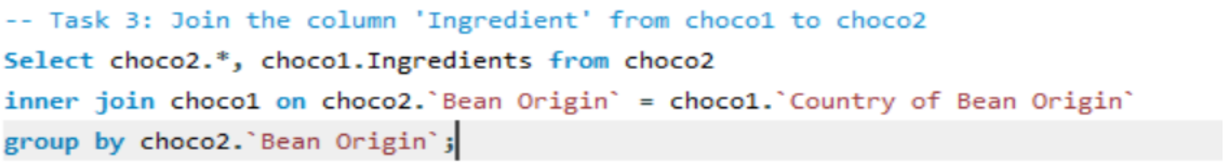


	REF	Cocoa Percent	Rating
▶	259	91	1.5
	486	100	1.75
	81	99	2
	32	53	2
	1359	72.5	2.5
	1189	89	2.5
	809	84	2.5
	502	75	2.75
	1788	90	2.75
	2052	71.5	2.75
	552	46	2.75
	705	60	2.75
	370	55	2.75



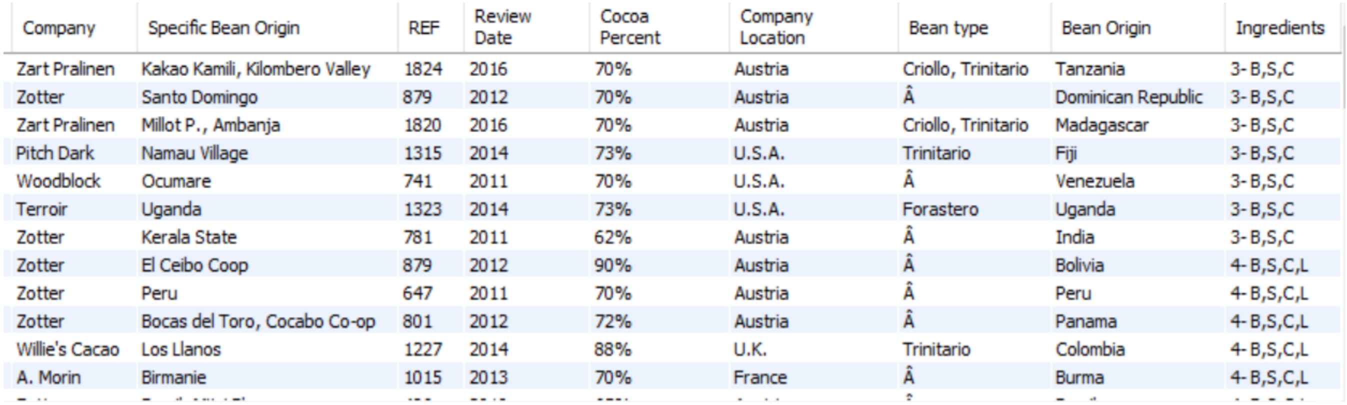
	Cocoa Percent	avg(Rating)
▶	50	3.75
	63	3.5357142857142856
	69	3.4615384615384617
	78	3.380952380952381
	66	3.3482142857142856
	67	3.3455882352941178
	68	3.2881944444444446
	70	3.2629186602870814
	87	3.25
	79	3.25
	86	3.25
	56	3.25
	74	3.2234848484848486

For the task 3, I would like to add ingredients from choco1 to choco2. In order to achieve the result, I use inner join to link the two table with key columns: 'Country of Bean Origin` from choco1 and "bean Origin" from choco2:



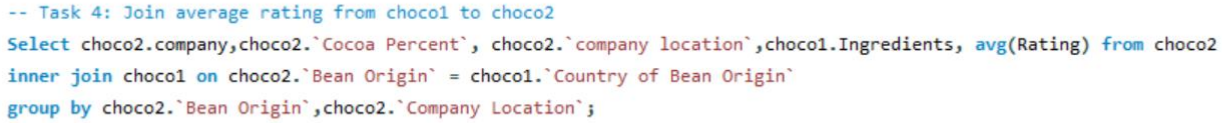
```
-- Task 3: Join the column 'Ingredient' from choco1 to choco2
Select choco2.*, choco1.Ingredients from choco2
inner join choco1 on choco2.`Bean Origin` = choco1.`Country of Bean Origin`
group by choco2.`Bean Origin`;
```

Then I got the table as below:



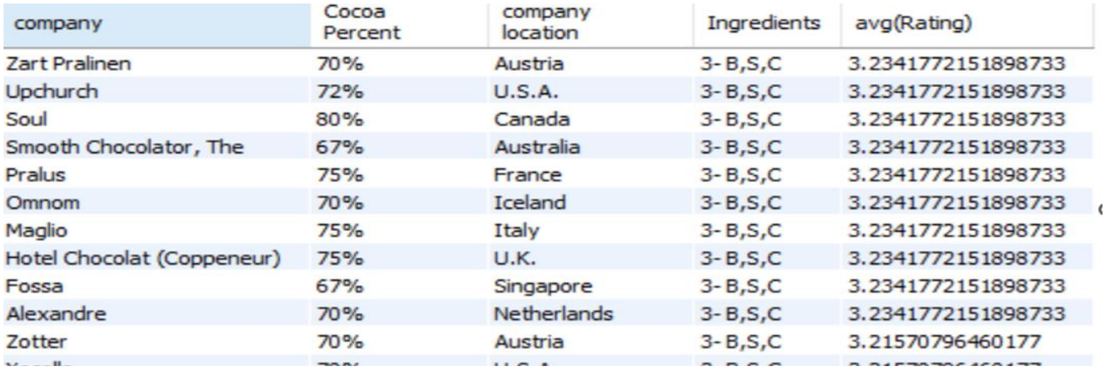
Company	Specific Bean Origin	REF	Review Date	Cocoa Percent	Company Location	Bean type	Bean Origin	Ingredients
Zart Pralinen	Kakao Kamili, Kilombero Valley	1824	2016	70%	Austria	Criollo, Trinitario	Tanzania	3- B,S,C
Zotter	Santo Domingo	879	2012	70%	Austria	Â	Dominican Republic	3- B,S,C
Zart Pralinen	Millot P., Ambanja	1820	2016	70%	Austria	Criollo, Trinitario	Madagascar	3- B,S,C
Pitch Dark	Namau Village	1315	2014	73%	U.S.A.	Trinitario	Fiji	3- B,S,C
Woodblock	Ocumare	741	2011	70%	U.S.A.	Â	Venezuela	3- B,S,C
Terroir	Uganda	1323	2014	73%	U.S.A.	Forastero	Uganda	3- B,S,C
Zotter	Kerala State	781	2011	62%	Austria	Â	India	3- B,S,C
Zotter	El Ceibo Coop	879	2012	90%	Austria	Â	Bolivia	4- B,S,C,L
Zotter	Peru	647	2011	70%	Austria	Â	Peru	4- B,S,C,L
Zotter	Bocas del Toro, Cocabo Co-op	801	2012	72%	Austria	Â	Panama	4- B,S,C,L
Willie's Cacao	Los Llanos	1227	2014	88%	U.K.	Trinitario	Colombia	4- B,S,C,L
A. Morin	Birmanie	1015	2013	70%	France	Â	Burma	4- B,S,C,L

With the task 4, I would like to select just columns company, cocoa percent and company location from choco2, calculate average rating from choco1 and join to choco2 include ingredients, then group by bean origin and company location. I create the query:



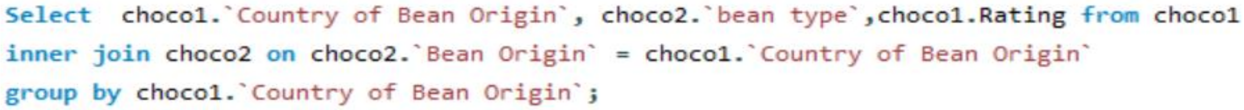
```
-- Task 4: Join average rating from choco1 to choco2
Select choco2.company,choco2.`Cocoa Percent`, choco2.`company location`,choco1.Ingredients, avg(Rating) from choco2
inner join choco1 on choco2.`Bean Origin` = choco1.`Country of Bean Origin`
group by choco2.`Bean Origin`,choco2.`Company Location`;
```

Here is the result:



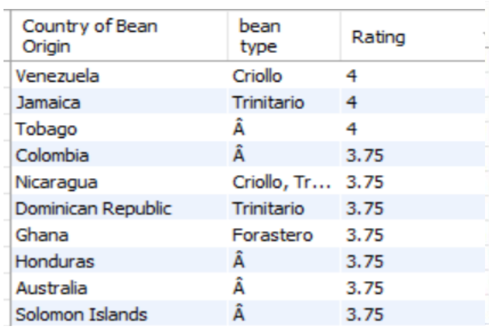
company	Cocoa Percent	company location	Ingredients	avg(Rating)
Zart Pralinen	70%	Austria	3- B,S,C	3.2341772151898733
Upchurch	72%	U.S.A.	3- B,S,C	3.2341772151898733
Soul	80%	Canada	3- B,S,C	3.2341772151898733
Smooth Chocolator, The	67%	Australia	3- B,S,C	3.2341772151898733
Pralus	75%	France	3- B,S,C	3.2341772151898733
Omnom	70%	Iceland	3- B,S,C	3.2341772151898733
Maglio	75%	Italy	3- B,S,C	3.2341772151898733
Hotel Chocolat (Coppeneur)	75%	U.K.	3- B,S,C	3.2341772151898733
Fossa	67%	Singapore	3- B,S,C	3.2341772151898733
Alexandre	70%	Netherlands	3- B,S,C	3.2341772151898733
Zotter	70%	Austria	3- B,S,C	3.21570796460177

On the last task, I want to know the relation between bean type(choco2) and country of bean origin (choco1) base on the rating score. I made a query with inner join:

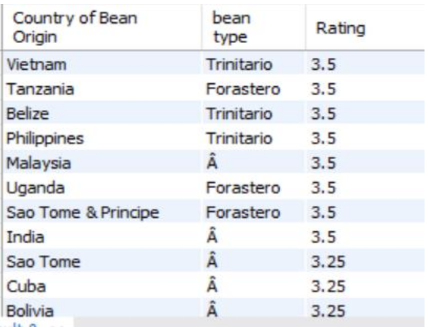


```
Select  choco1.`Country of Bean Origin`, choco2.`bean type`,choco1.Rating from choco1
inner join choco2 on choco2.`Bean Origin` = choco1.`Country of Bean Origin`
group by choco1.`Country of Bean Origin`;
```

From the table below we can see that the most popular kind of bean are Trinitario, Criollo Forastero, Amazon have 3.75 to 4



Country of Bean Origin	bean type	Rating
Venezuela	Criollo	4
Jamaica	Trinitario	4
Tobago	Â	4
Colombia	Â	3.75
Nicaragua	Criollo, Tr...	3.75
Dominican Republic	Trinitario	3.75
Ghana	Forastero	3.75
Honduras	Â	3.75
Australia	Â	3.75
Solomon Islands	Â	3.75



Country of Bean Origin	bean type	Rating
Vietnam	Trinitario	3.5
Tanzania	Forastero	3.5
Belize	Trinitario	3.5
Philippines	Trinitario	3.5
Malaysia	Â	3.5
Uganda	Forastero	3.5
Sao Tome & Principe	Forastero	3.5
India	Â	3.5
Sao Tome	Â	3.25
Cuba	Â	3.25
Bolivia	Â	3.25

Conclusion

In conclusion, this data can be very valuable to individuals in the chocolate industry. Consumer preference can be useful to determine how well a chocolate bar will sell. For instance, if it is not selling well, is it due to the amount of cocoa is in it? Over time the average rating isn't increasing but the instances of extreme ratings is decreasing, resulting in a smaller spread of ratings. As analysed in my work above, it can be stated that the country of manufacture doesn't appear to have a large impact on rating, but both the maker or the company brand can be seen to impact the average rating of a chocolate. In addition, there isn't a strong relationship between the cocoa percent and rating of the chocolate. We can see that most chocolate bars that have a cocoa percent from 65- 75% is much better rating than if they have more or less cocoa. Also, we can observe that cocoa beans that come from South American countries are rated much higher, we can then use this data to pitch ideas to companies where the most chocolate is made (USA, France, Canada, etc.). However, the broad bean origin of the bean doesn't seem to determine the quality, or rating, of the chocolate. Factors such as these help us get in touch with customer behavior and can help us predict how well a chocolate bar will do in the near future to see if we can increase sales, or perhaps customer preference of certain chocolate bars.

Sources links.

https://www.kaggle.com/datasets/rtatman/chocolate-bar-ratings

https://damecacao.com/ingredients-in-chocolate/

https://barandcocoa.com/pages/what-is-100-percent-chocolate#:~:text=A%20bar%20of%20100%25%20dark,of%2099.75%25%20cacao%20by%20volume.

https://www.chocolate.co.uk/blogs/news/the-different-varieties-of-cocoa-beans-criollo-trinitario-and-forastero#:~:text=world's%20cocoa%20production.-,Forastero,Sri%20Lanka%2C%20Malaysia%20and%20Indonesia.

https://gist.github.com/NEOissss/55c8637457f2a4e5b8d0ddec3ebb2e51

https://www.britannica.com/topic/chocolate