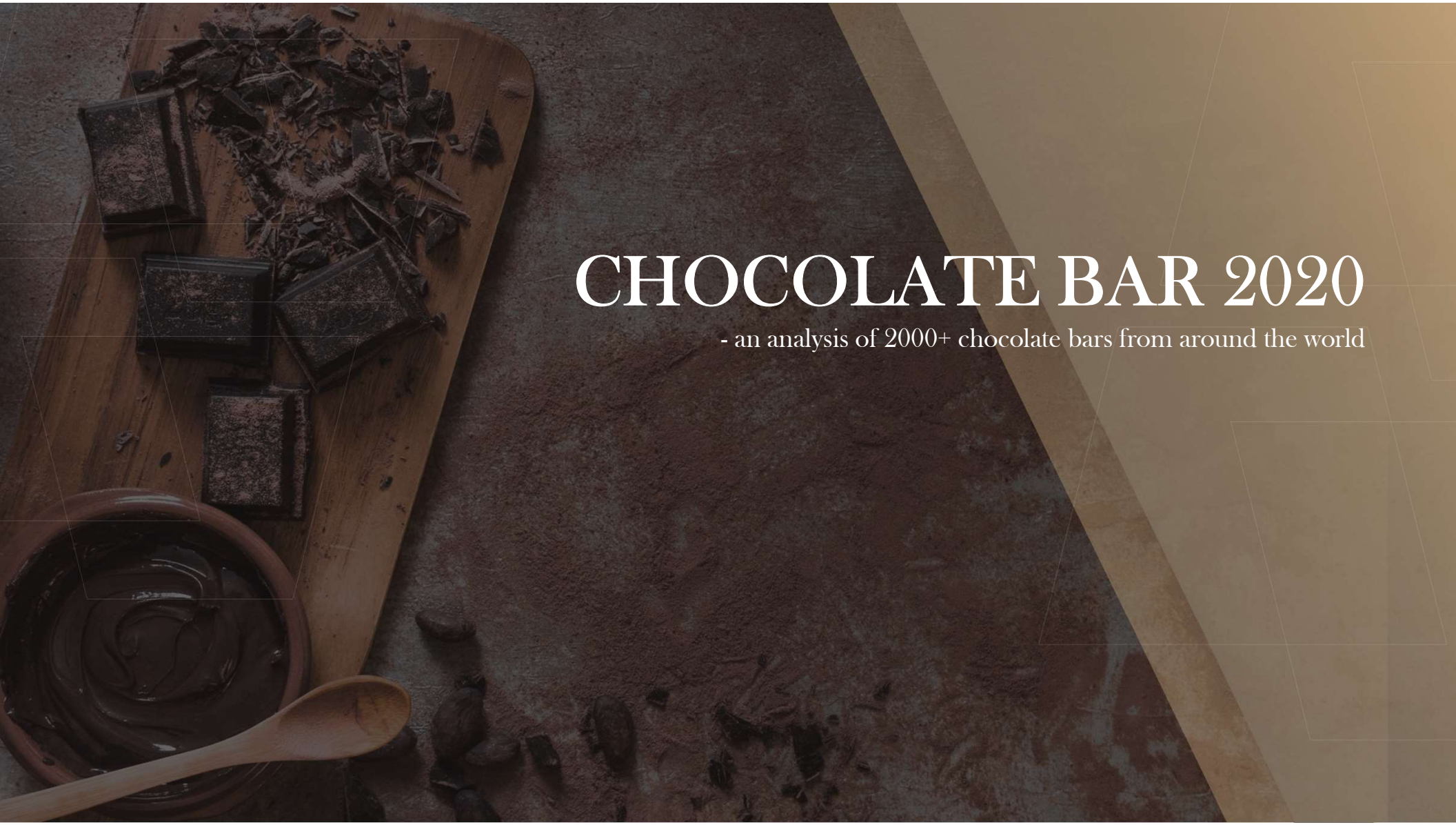


CHOCOLATE BAR 2020

- an analysis of 2000+ chocolate bars from around the world

WORLD OF CHOCOLATE

FLAVOR

600+ flavor compounds while red wine only has 200.

TEXTURE

Impacts flavor – a good way to evaluate the maker's vision

AFTERMELT

Chocolate is the only edible substance to melt around 32°C (around human body) - the last impression from chocolate influences the overall experience

OVERALL EXPERIENCE

The smell of chocolate increases theta brain waves which triggers relaxation – the most prominent impressions will last forever.

DATA DESCRIPTION

Two tables: chocolate and chocolate_taste_dataset

Chocolate dataset contains 2000+ expert ratings in 66 countries along with key variables like country_of_bean_origin, cocoa_percent, and counts_of_ingredients

Chocolate_taste_dataset has different tastes of chocolate bars

DATA CLEANING & CHALLENGES

CLEAN DATA

We only keep data with review_count> 10

Group by counts_of_ingredients
Having review_count > 10

***Observations*:**

“vanilla” column but string values are “have_not_vanila” or “have_vanila”

DATA DUPLICATION & NULL VALUES

There is not duplicated record in the dataset.

There is no NULL value found for company, company_location, country_of_bean_origin

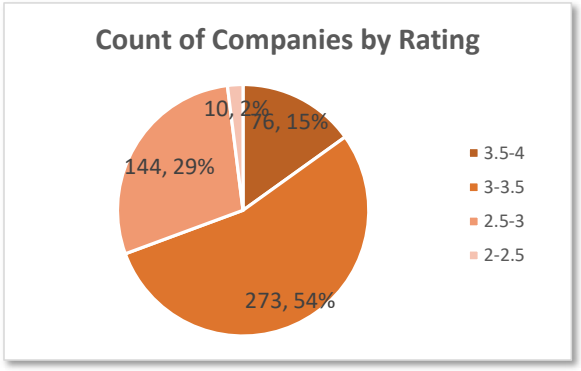
There are NULL values found for second_taste, third_taste, and fourth_taste

CHANGE INGREDIENTS COLUMN INTO BINARY

beans	cocoa_butter	vanilla	lecithin	salt	sugar	sweetener_without_sugar
have_bean	have_cocoa_butter	have_not_vanila	have_not_lecithin	have_not_salt	have_sugar	have_not_sweetener_without_sugar

beans_binary	cocoa_butter_binary	vanilla_binary	lecithin_binary	salt_binary	sugar_binary	sweetener_without_sugar_binary
1	1	0	0	0	1	0

WHAT AFFECTS THE RATING - LIKES?



ZOKOKO
Artisan Chocolate

OCELOT

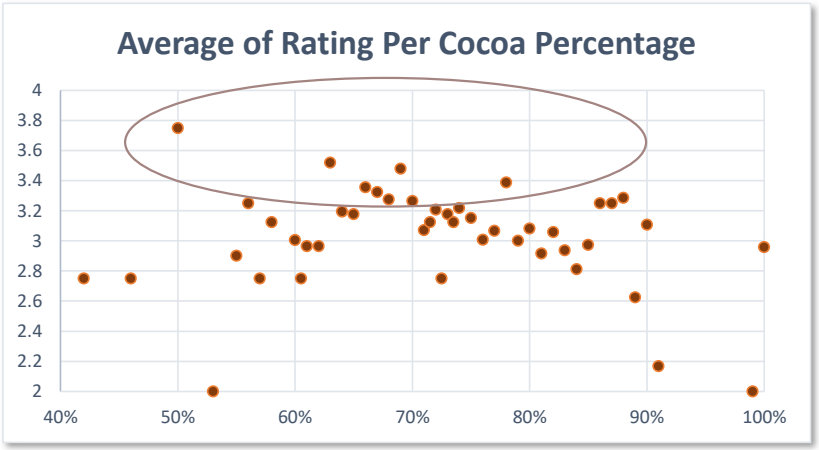
Idilio Origins
Premium Swiss Chocolate

Patric CHOCOLATE

Matale CHOCOLATE

Company	AVG_RATING
Zokoko	3.88
Ocelot	3.88
Matale	3.81
Patric	3.79
Idilio (Felchlin)	3.78

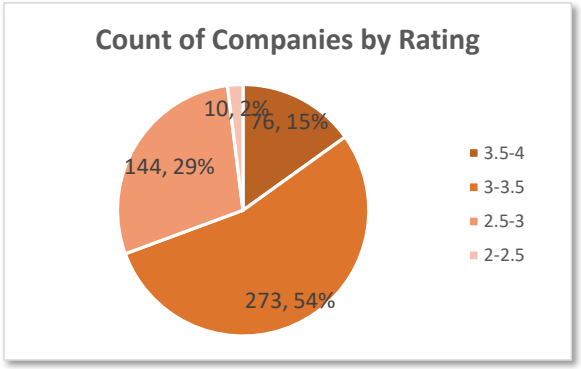
- Have sugar - NO salts
- Average cocoa percentage is 71%



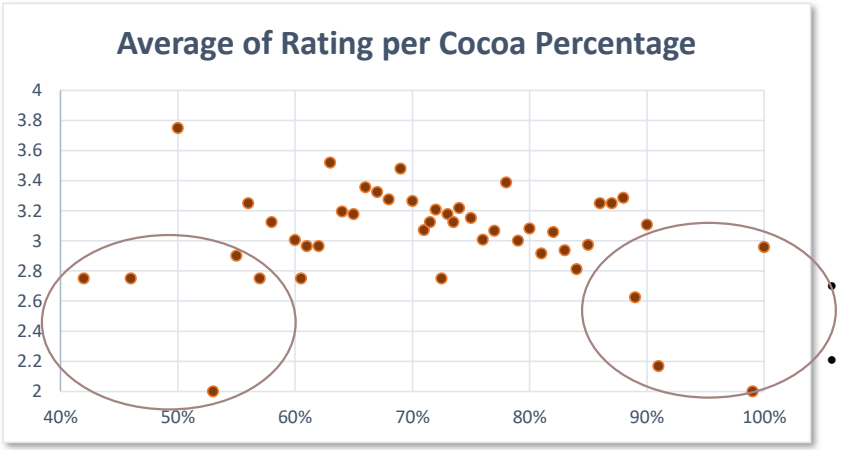
Cocoa Percentage	AVG_RATING
0.50%	3.75
0.63%	3.52
0.69%	3.48
0.78%	3.39
0.66%	3.36

- A higher cocoa percentage does NOT mean a higher rating. Cocoa Percentages around 50% to 80% have higher rating

WHAT AFFECTS THE RATING - DISLIKES?



Company	AVG_RATING
Ki' Xocolatl	2.00
Jacque Torres	2.00
Casa	2.00
Majani	2.00
Love Bar	2.00

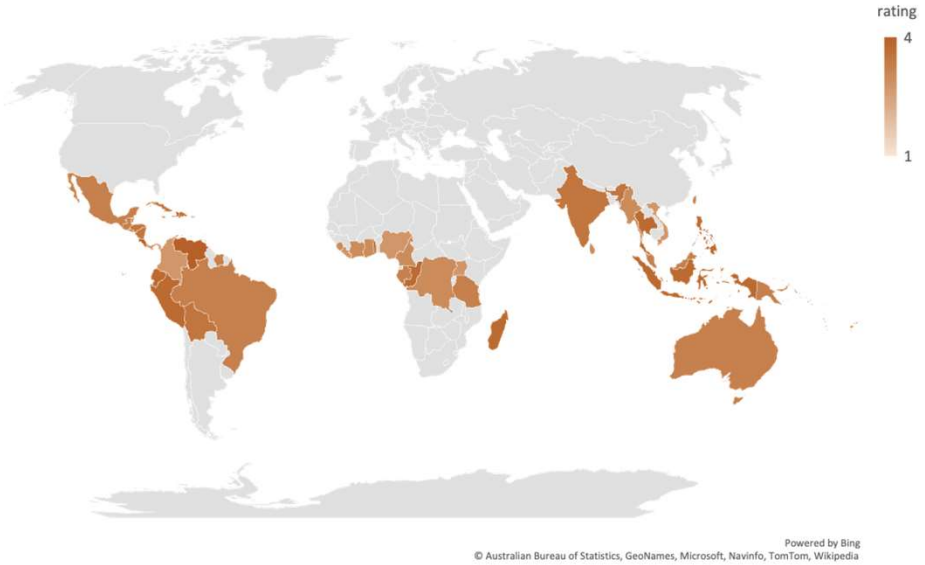


Cocoa Percentage	AVG_RATING
0.53%	2.00
0.99%	2.00
0.91%	2.21
0.89%	2.67
0.42%	2.75

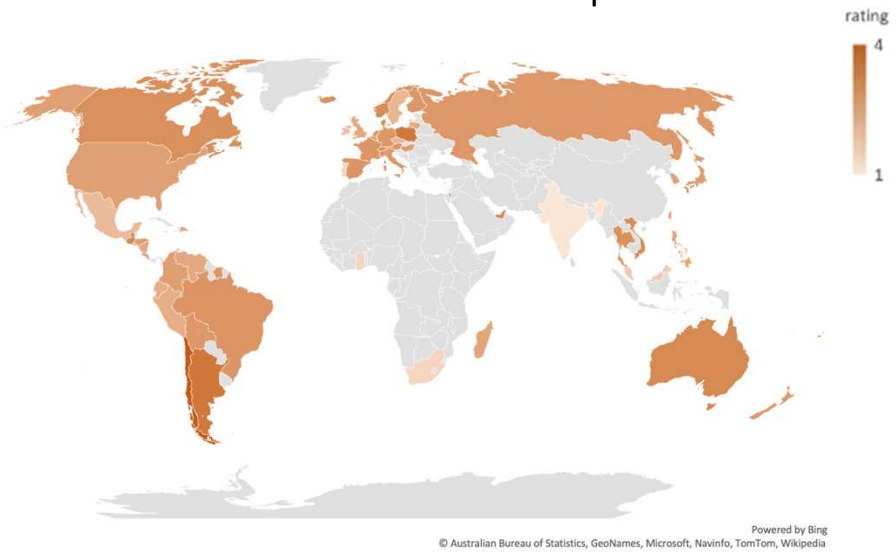
- Companies who have a rating below or equal to 2.5 are among the bottom 10%
- People don’t like extremes in cocoa percentage

WORLD CHOCOLATE MAP

Where are the best cocoa beans grown?



Where are the best chocolate companies located?



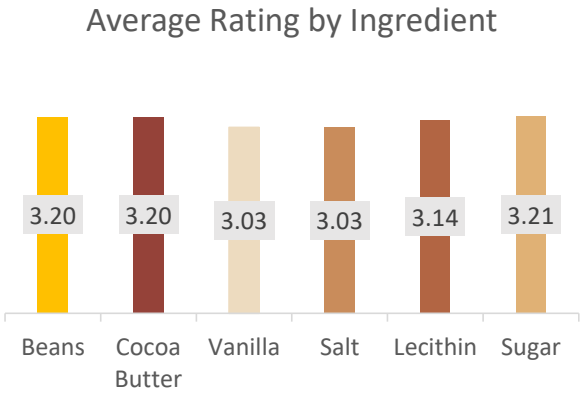
Key Takeaways :

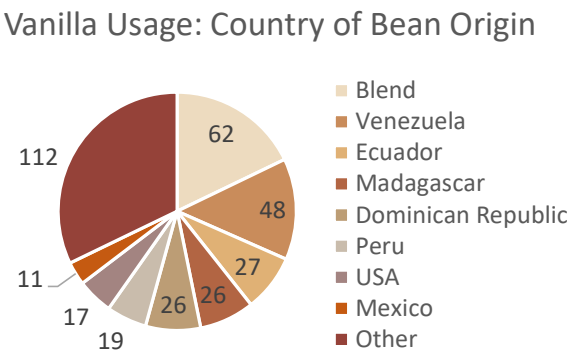
The highest average ratings were for chocolate bars that used 3 of the above ingredients (3.26/4)

Salt is the least common ingredient used (37/2221 chocolate bars)

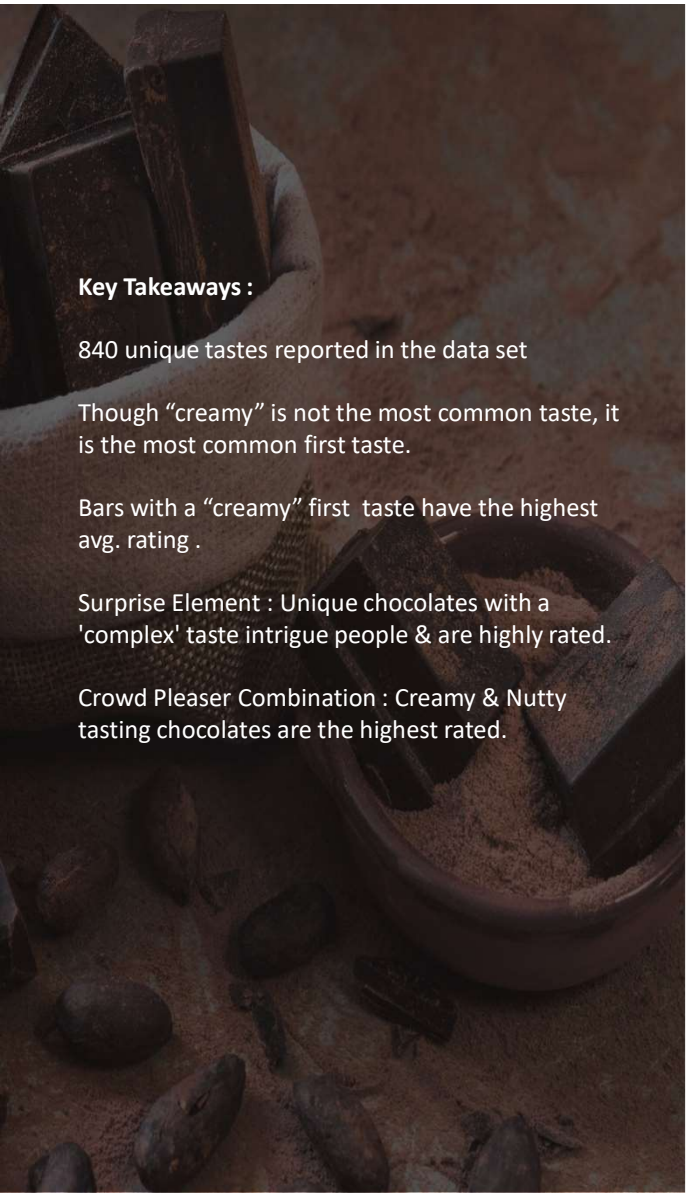
8 chocolate bars out of 2221 used neither sugar nor sweetener

A closer look at INGREDIENTS





Ingredient Combination	Avg. Rating	Review Count
Cocoa Butter, Sugar	3.27	882
Sugar	3.21	633
Cocoa Butter, Lecithin, Sweetener	3.20	272
Cocoa Butter, Salt, Sweetener	3.11	20
Cocoa Butter, Vanilla, Lecithin, Sugar	3.09	184
Cocoa Butter, Vanilla, Sugar	2.97	136
Sweetener	2.96	31
Cocoa Butter, Sweetener	2.94	12



Key Takeaways :

840 unique tastes reported in the data set

Though “creamy” is not the most common taste, it is the most common first taste.

Bars with a “creamy” first taste have the highest avg. rating .

Surprise Element : Unique chocolates with a 'complex' taste intrigue people & are highly rated.

Crowd Pleaser Combination : Creamy & Nutty tasting chocolates are the highest rated.

Let's TASTE

Most common taste

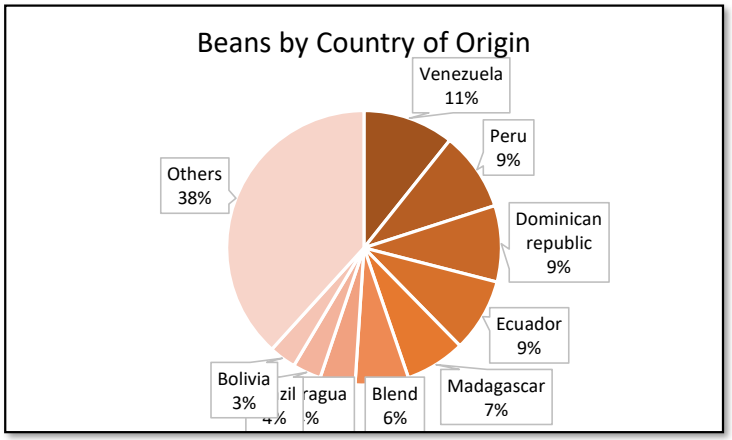
Understanding the Human Palate

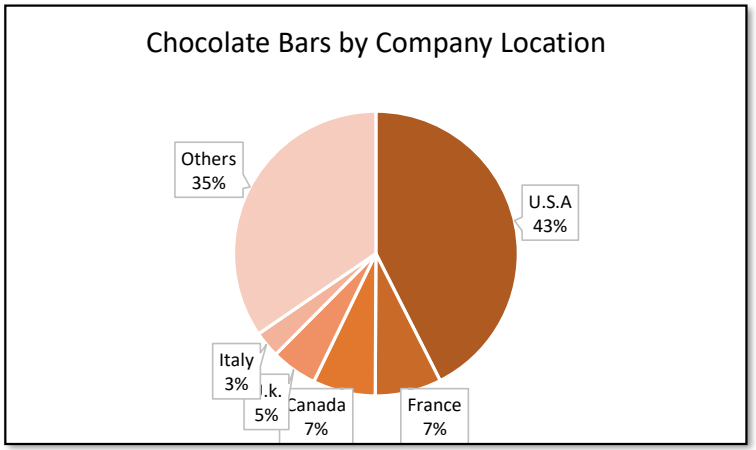
First Taste -> Second Taste -> Third Taste

Highest Rated First Taste		
First Taste	Review Count	Avg. Rating
Complex	26	3.54
Tart	15	3.52
Rich Cocoa	23	3.50
Creamy	159	3.48
Smooth	20	3.48
Rich	12	3.44
Spice	11	3.41
Dried Fruit	20	3.40
Spicy	47	3.34
Cocoa	29	3.33

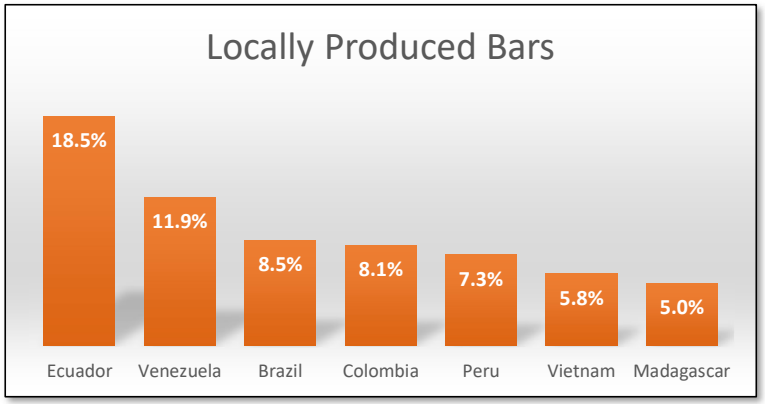
Highest Rated Taste Pairing			
First Taste	Second Taste	No. of Reviews	Avg. Rating
Creamy	Nutty	16	3.6
Sandy	Sweet	34	3.1
Gritty	Sweet	11	2.8

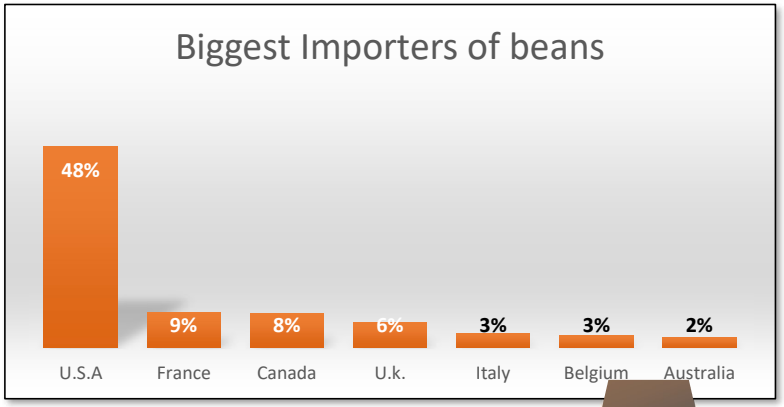
Beans & Bars by Geography

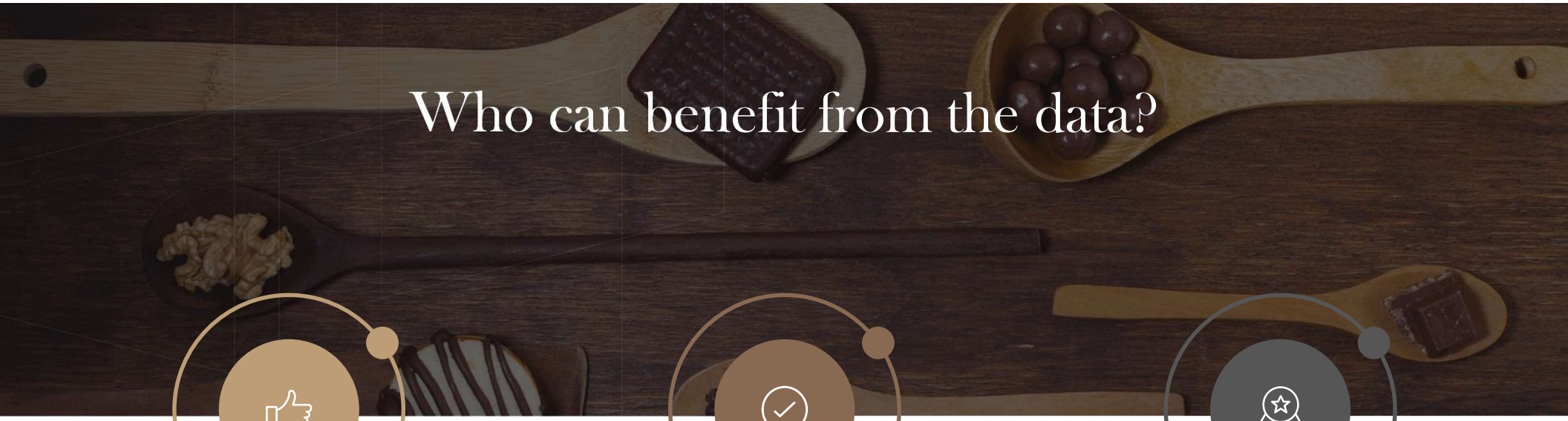




Out of the total Chocolate Bars produced only 12% are made locally, rest of the beans are exported to other Countries.







Who can benefit from the data?

01

Chocolate Makers

From details like where different cocoa beans originate and what type of beans are used to make what type of chocolate.

02

Cocoa Exporters

This gives exporters a good idea of where they could import what type of beans

03

Chocolate Enthusiasts

From chocolate lovers to chocoholics, chocolate enthusiasts have so many names

CHOCOLATE – MORE IS LESS

(Other data points that could have made the experience better)

Reviewer Demography

The data-set could have provided information about the reviews and their location as well. This would help with establishing a proper target segment for any kind of analysis.

Multiple Factors for Ratings

The ratings do not reflect health benefits, social missions, or organic status.

Categories and Labelling

Proper categories for taste and ingredients would make it easier to analyze the data set.

Price Points

Adding the price of the different bars or adding the price of extracting the Cocoa in different countries would also be of great importance.


THANK
YOU

Appendix- Data Cleaning

Check for duplicates:

```
ALTER TABLE chocolate

ADD COLUMN ChocoID VARCHAR(255) AFTER country_of_bean_origin;

UPDATE chocolateSET ChocoID = CONCAT(ref,company,"-
",specific_bean_origin_or_bar_name, cocoa_percent,rating);

SELECT ChocoID, count(*) as RowCount, company, country_of_bean_origin,
cocoa_percent, company_location,rating,beans, review_datefrom
chocolate group by ChocoIDorder by rowcount desc;
```

Check for NULL values:

```
select * from chocolate where company is NULL or company_location is
NULL or country_of_bean_origin is NULL;
```

Variables:

Ref: unique number for company

Company: company name

Company_location: company location:

Review_date: data review for chocolate bar

Country_of_bean_origin: country of chocolate bean

specific_bean_origin_or_bar_name: province of chocolate bean

cocoa_percent: percent of chocolate bar

Rating: chocolate bar rating

counts_of_ingredients: number of ingredients

Appendix – Likes and dislikes

Find the average ratings for each cocoa percent:

```
Select cocoa_percentage, AVG(Rating) as
AverageRatingPerCocoaPercent

FROM chocolate

GROUP BY cocoa_percentage

ORDER BY AverageRatingPerCocoaPercent DESC;
```

Find 5 cocoa percentage with the highest and the lowest rating

```
Select cocoa_percentage, AVG(Rating) as
AverageRatingPerCocoaPercent

FROM chocolate

GROUP BY cocoa_percentage

ORDER BY AverageRatingPerCocoaPercent DESC

Limit 5;
```

```
Select cocoa_percentage, AVG(Rating) as
AverageRatingPerCocoaPercent

FROM chocolate

GROUP BY cocoa_percentage

ORDER BY AverageRatingPerCocoaPercent ASC

Limit 5
```

Calculate count of companies in each rating range.

```
SELECT AverageRatingPerCompany AS RatingRange, COUNT(*) AS
 CompanyCount FROM
( SELECT
 CASE WHEN RatingRange between 2 and 2.5 then ' 2-2.5'
when RatingRange between 2.5 and 3 then '2.5-3'
when RatingRange between 3 and 3.5 then '3-3.5'
else '3.5-4' end as range RatingRange from RatingbyCompany)
group by AverageRatingPerCompany;
```

Find 5 companies with the highest and the lowest rating

```
Select companies, AVG(Rating) as AverageRatingPerCompany

FROM Chocolate

GROUP BY company

Order BY AverageRatingPerCompany DESC

Limit 5;
```

```
Select companies, AVG(Rating) as AverageRatingPerCompany

FROM Chocolate

GROUP BY company

Order BY AverageRatingPerCompany ASC

Limit 5;
```

Appendix- Taste Analysis Code

1. Most popular taste pairing

```
select distinct first_taste,second_taste, count(*) as review_count,
avg(rating) as avg_rating
from chocolate
group by first_taste,second_taste
having review_count>5
order by avg_rating desc;
```

2. Highest Rated Taste

```
select first_taste, count(*) as ReviewCount, avg(rating)
from chocolate
group by first_taste
having ReviewCount>10
order by avg(rating) desc
limit 10;
```

3. Most Common First Taste - Top 5

```
select first_taste, count(*) as TasteCount
from chocolate
group by first_taste
order by TasteCount desc
limit 5;
```

Appendix- Ingredient Analysis Code

```
create table chocolate_binary select *, case when beans="have_bean" then 1 else 0 end as beans_binary, case when cocoa_butter="have_cocoa_butter" then 1 else 0 end as cocoa_butter_binary, case when vanilla="have_vanila" then 1 else 0 end as vanilla_binary, case when lecithin="have_lecithin" then 1 else 0 end as lecithin_binary, case when salt="have_salt" then 1 else 0 end as salt_binary, case when sugar="have_sugar" then 1 else 0 end as sugar_binary, case when sweetener_without_sugar="have_sweetener_without_sugar" then 1 else 0 end as sweetener_without_sugar_binary from chocolate;

select beans_binary, count(*) as review_count, avg(rating) as avg_rating, count(distinct ref) as bar_count from chocolate_binary group by beans_binary;

select cocoa_butter_binary, count(*) as review_count, avg(rating) as avg_rating, count(distinct ref) as bar_count from chocolate_binary group by cocoa_butter_binary;

select vanilla_binary, count(*) as review_count, avg(rating) as avg_rating, count(distinct ref) as bar_count from chocolate_binary group by vanilla_binary;

select lecithin_binary, count(*) as review_count, avg(rating) as avg_rating, count(distinct ref) as bar_count from chocolate_binary group by lecithin_binary;

select salt_binary, count(*) as review_count, avg(rating) as avg_rating, count(distinct ref) as bar_count from chocolate_binary group by salt_binary;

select sugar_binary, count(*) as review_count, avg(rating) as avg_rating, count(distinct ref) as bar_count from chocolate_binary group by sugar_binary;

select sweetener_without_sugar_binary, count(*) as review_count, avg(rating) as avg_rating, count(distinct ref) as bar_count from chocolate_binary group by sweetener_without_sugar_binary;

select country_of_bean_origin, count(*) as review_count from chocolate_binarywhere vanilla_binary=1 group by country_of_bean_originhaving review_count>10 order by review_count desc;

select distinct counts_of_ingredients, count(*) as review_count, avg(rating) as avg_ratingfrom chocolate_binary group by counts_of_ingredientshaving review_count>10;

select distinct cocoa_butter_binary, vanilla_binary, lecithin_binary, salt_binary,  sugar_binary,sweetener_without_sugar_binary, count(*) as review_count, avg(rating) as avg_rating from chocolate_binary group by cocoa_butter_binary, vanilla_binary, lecithin_binary, salt_binary, sugar_binary, sweetener_without_sugar_binary having review_count>10 order by avg_rating desc;
```

Appendix: Other Insights

1) select country_of_bean_origin, count(*) as CountryCountfrom chocolategroup by country_of_bean_originorder by CountryCount desc;

2) select company_location, count(*) as CountryCountfrom chocolategroup by company_locationorder by CountryCount desc;

3) select company_location, count(company_location)

from chocolate

where company_location = country_of_bean_origin

group by company_location

order by count(company_location) desc;

select company_location, count(company_location)

from chocolate

where company_location != country_of_bean_origin

group by company_location

order by count(company_location) desc;