

# Chocolate Bar Ratings

**08.30 Cohort- Brenda: Madeline Ye**

# What's in the data set?

- 1,700 plain dark chocolate bars
- Regional Origin information
- Percentage of cocoa
- Variety of chocolate bean used
- Where the beans were grown
- Expert rating of the bar



These ratings were compiled by Brady Brelinski, Founding Member of the Manhattan Chocolate Society

<http://flavorsofcacao.com/index.html>

Kaggle: <https://www.kaggle.com/ratatman/chocolate-bar-ratings>

# Cleaning Steps



- Column name formatting
- Two null values:
  - Bean origin and Bean type

```
#Checking the specific on these null values  
df[df.isnull().any(axis=1)]
```

	Company(Maker)	Specific_Bean_Origin_Bar_Name	REF	Review_Date	Cocoa_percent	Company_location	Rating	Bean_type	Broad_Bean_Origin
1072	Mast Brothers	Madagascar	999	2012	72%	U.S.A.	2.5	Trinitario	NaN
1544	Soma	Three Amigos(Chuao, Wild Bolivia, D.R.)	676	2011	70%	Canada	4.0	NaN	Ven, Bolivia, D.R.

```
#Deciding to drop the null values as they are not a simple fix of imputing or backfilling and it is only 2 data points  
df = df.dropna()
```

# Cleaning Continued...

- Dropping Review Date Column, irrelevant
- Changing cocoa percentage to float
- Empty column value for Bean type

```
In [27]: df['Bean_type'].value_counts()
```

```
#We see that there are a lot of blank values for bean type. Might be worth filling in this with "Unknown"
```

```
Out[27]:
```

	887
--	-----

Trinitario	418
------------	-----

Criollo	153
---------	-----

Forastero	87
-----------	----

Forastero (Nacional)	52
----------------------	----

Blend	41
-------	----

Criollo, Trinitario	39
---------------------	----

Forastero (Arriba)	37
--------------------	----

Criollo (Porcelana)	10
---------------------	----

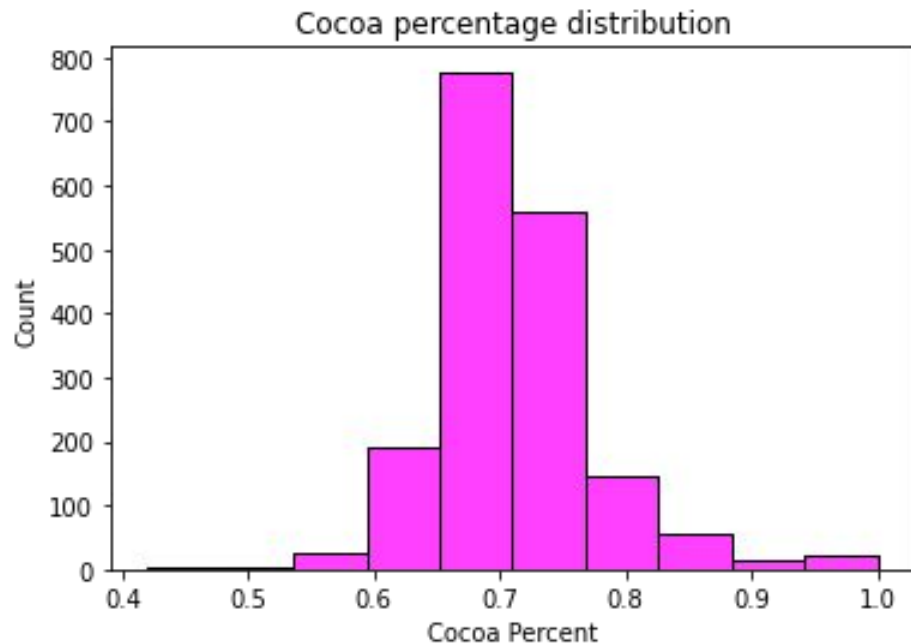
Trinitario, Criollo	9
---------------------	---

Forastero (Parazinho)	8
-----------------------	---

Forastero (Arriba) ASS	6
------------------------	---

# Exploratory Visuals: Cocoa percentage

- Most bars are in the high 60s, low 70s



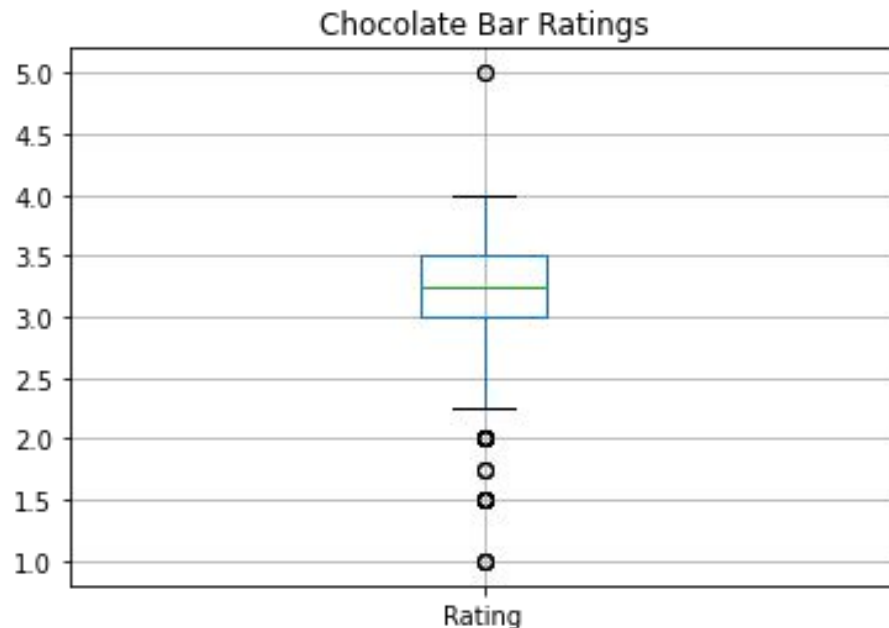
# Exploratory Visuals: Rating

---

- Rated on:
  - Flavor
  - Texture
  - Aftermelt
  - Overall

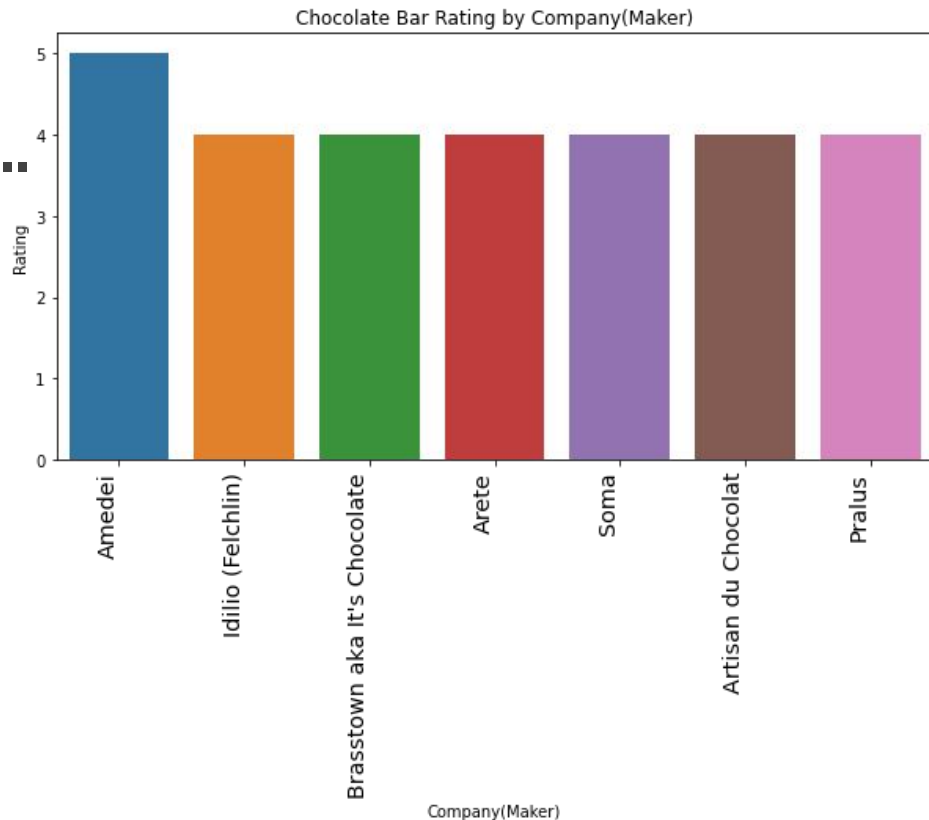
## Flavors of Cacao Rating System:

- 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)



# Exploratory Visuals Top Brands...

- Top 7 rated chocolate bars
- No other rating comes close to the top



Company(Maker)	Specific_Bean_Origin_Bar_Name	REF	Cocoa_percent	Company_location	Rating	Bean_type	Broad_Bean_Origin
Amedei	Toscano Black	40	0.7	Italy	5.0	Blend	
Amedei	Chuo	111	0.7	Italy	5.0	Trinitario	Venezuela

# Thank you!

— — —