

# **Exploratory Data Analysis**

## **Workshop**

# Motivation

# Motivation

 **Understanding the Data Distribution**





# Motivation

-  **Understanding the Data Distribution**
-  **Identifying Patterns and Relationships**

# Motivation

- 🍫 **Understanding the Data Distribution**
- 🍫 **Identifying Patterns and Relationships**
- 🍫 **Data Cleaning and Preprocessing**

# Motivation

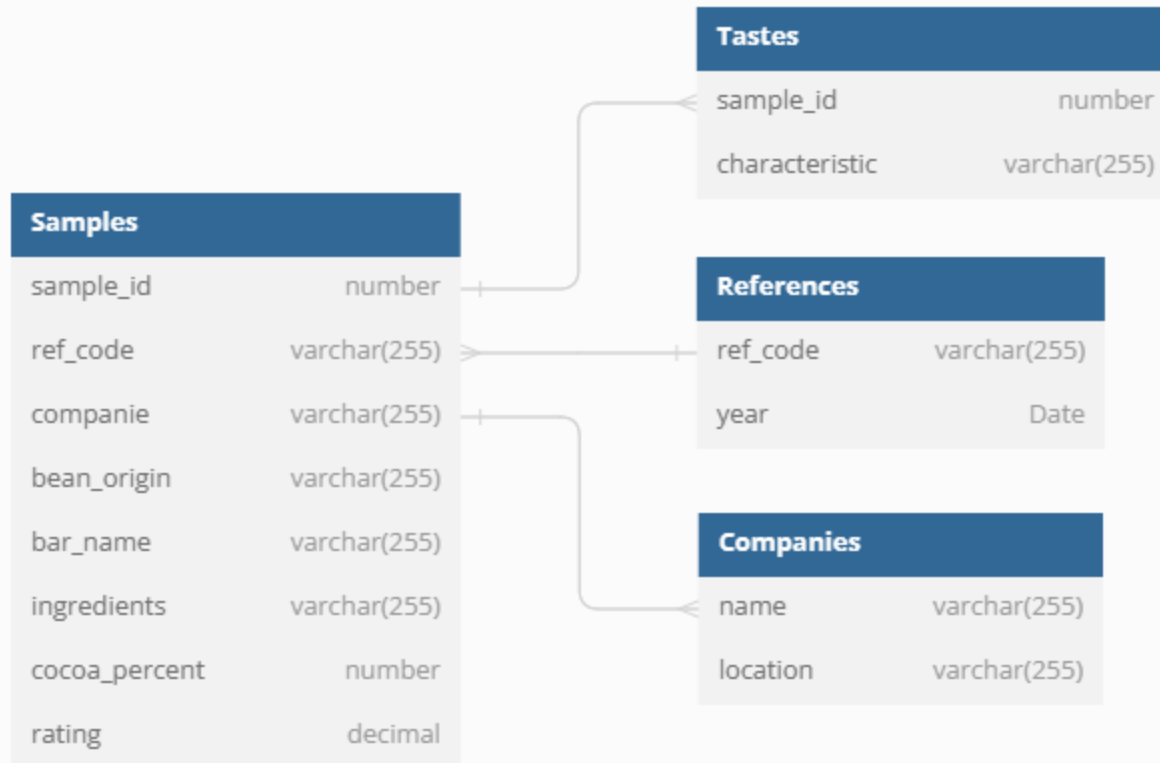
-  **Understanding the Data Distribution**
-  **Identifying Patterns and Relationships**
-  **Data Cleaning and Preprocessing**
-  **Hypothesis Generation**

# Dataset: Flavors Of Cacao

[https://flavorsofcacao.com/chocolate\\_database.html](https://flavorsofcacao.com/chocolate_database.html)



# Dataset: Flavors Of Cacao





# Cacao origins

## **Best**

Tobago

China

Sao Tome & Principe

Solomon Islands

Congo

## **Worst**

Bali

Puerto Rico

France

Principe

Martinique

# Cacao Flavors

## **Best**

---

rounded

robust cocoa

unripened fruit

tart citrus

homey

## **Worst**

---

this is not chocolate

strong off flavor

very bitter

klingy

unpleasant

# Cacao Ingredients

ingredient	t_statistic	p_value	significant
beans			FALSE
sugar	4.38	3.23E-05	FALSE
sweetener	-4.02	1.29E-04	FALSE
cocoa butter	-0.64	5.24E-01	TRUE
vanilla	-6.80	3.50E-11	FALSE
lecithin	-2.43	1.55E-02	TRUE
salt	-2.22	3.29E-02	TRUE

# Bye

**Ben Sivan**  
Data Scientist

---

+972-546925522

[ben.sivan@celleste-bio.com](mailto:ben.sivan@celleste-bio.com)