Chocolate Bar Ratings Predictor

Team Candy
Bruce Jilek, Tahereh Javaheripour, Yan Kin Pang, Travis Loseke

Topic Selection

Predicting chocolate bar ratings from variable such as:

- **Chocolateur**
- **Bean type**
- **Bean Source**
- **Cocoa Percent**

Reason for Topic Selection

- **Examined several food and beverage topics**

- **Chocolate was universally loved by the team**

- **Dataset was expansive enough to cover final project requirement**

- **Team became genuinely interested in if cocoa bean types and their source drive chocolate quality.**

Data sources

- **Kaggle- Chocolate Bar ratings**
  - https://www.kaggle.com/rtatman/chocolate-bar-ratings
- **Kaggle - Countries and States Lat Long**
  - https://www.kaggle.com/paultimothymooney/latitude-and-longitude-for-every-country-and-state
- **Flavors of Cacao website**
  - http://flavorsofcacao.com/chocolate_database.html
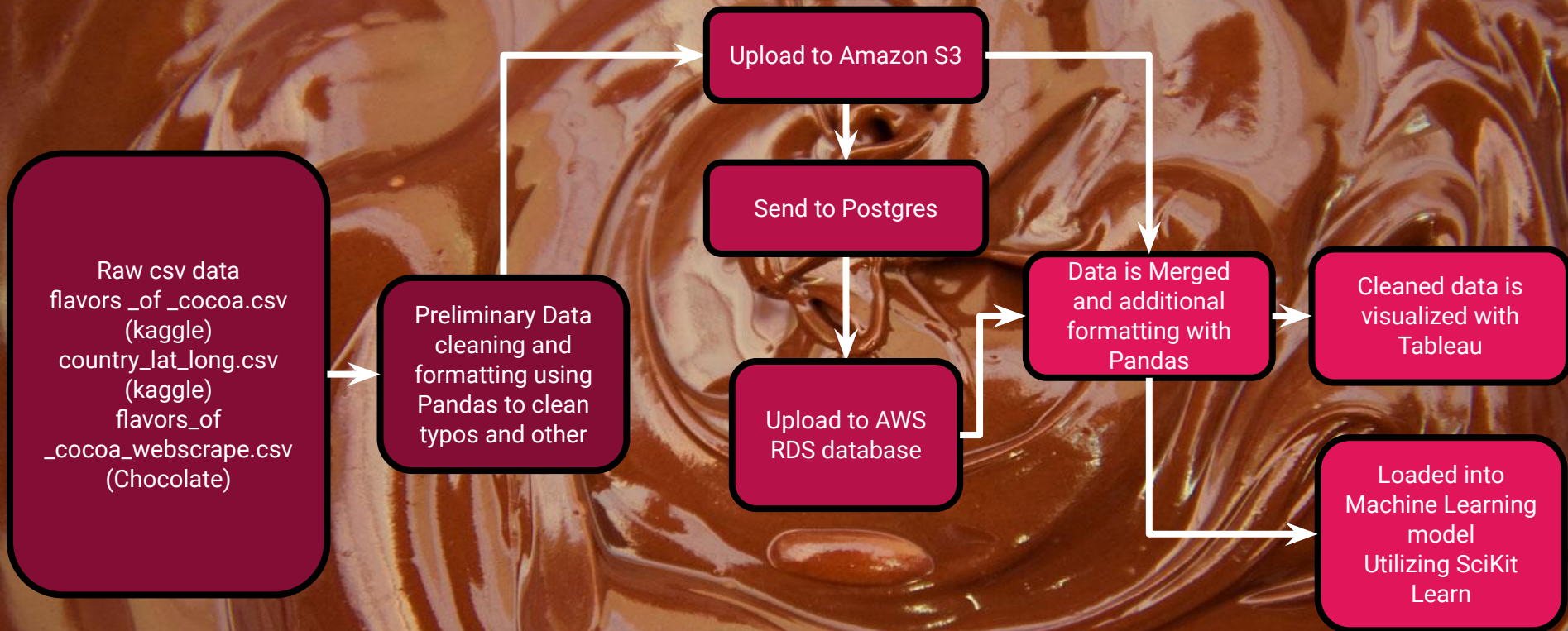
Four questions to answer

- **Where are the best cocoa beans grown?**

- **Which countries produce the highest-rated bars?**

- **What's the relationship between cocoa solids percentage and rating?**

- **Can we predict if a bar will be 4 stars?**

Data exploration phase

- **Pandas and Matplot lib were used to review, evaluate and clean datasets**

- **Tableau was used to make initial map plots to look at geographic distribution of cocoa bean sources**

-

Data processing and Evaluation Workflow

Raw csv data
flavors _of _cocoa.csv
(kaggle)
country_lat_long.csv
(kaggle)
flavors_of
_cocoa_webscrape.csv
(Chocolate)

Preliminary Data cleaning and formatting using Pandas to clean typos and other

Upload to Amazon S3

Send to Postgres

Upload to AWS RDS database

Data is Merged and additional formatting with Pandas

Cleaned data is visualized with Tableau

Loaded into Machine Learning model
Utilizing SciKit Learn

Data Preprocessing and Cleaning

- Pandas will be the primary avenue for cleaning and preparing the data
- Basic data editing of the csv will initially be performed
- Data will be uploaded to Amazon S3
- PostGres SQL with intake data and push to AWS RDS
- Python will merge data and format for use in Machine Learning and export out data to S3 for use in Tableau

Database Structure
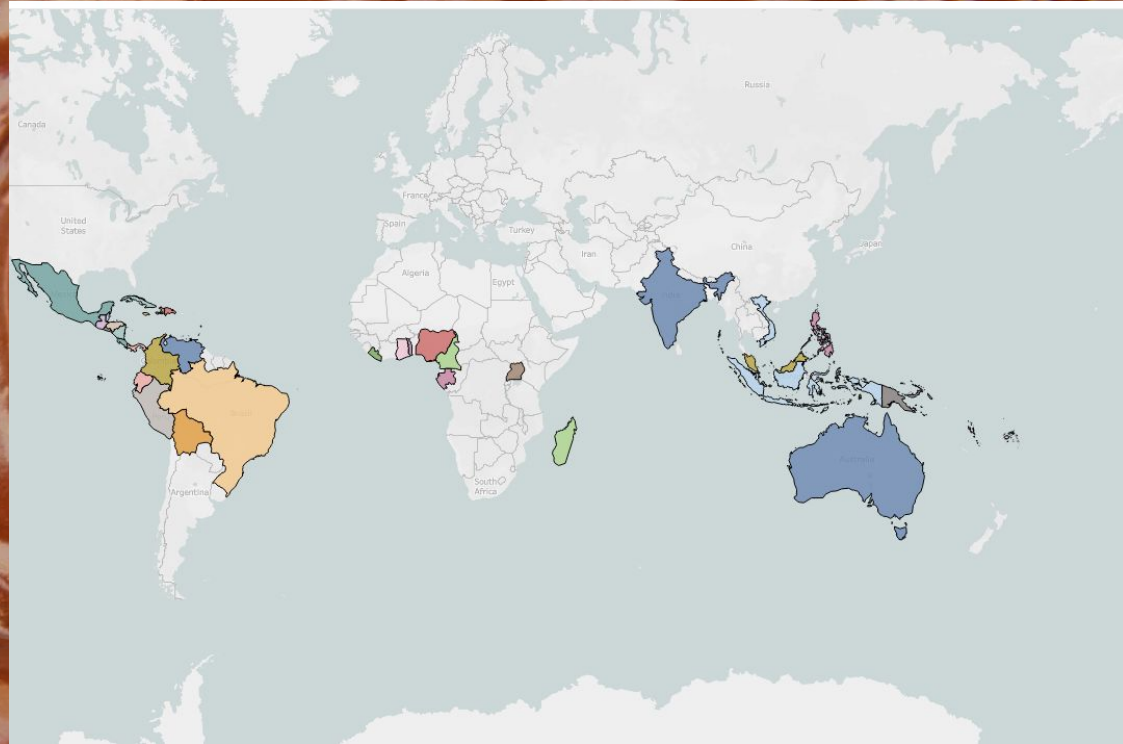
AWS Database- Postgres SQL Database

ERD:

Machine Learning Model

- Kera Neural Network
- Investigating others- Randon Forest and SVM

Dashboard

Tableau-

Global Chocolate production

Where are the best beans grown?

```
Broad_Bean_Origin
Venezuela                                                                      5.0
Madagascar                                                                     4.0
Dominican Republic,Madagascar                                                  4.0
Guatemala,Dominican Republic,Peru,Madagascar,Papua New Guinea                  4.0
Guatemala                                                                      4.0
Grenada,Papua New Guinea,Hawaii,Haiti,Madagascar                               4.0
South America                                                                  4.0
St Lucia                                                                       4.0
Peru                                                                           4.0
Ghana                                                                          4.0
Name: Rating, dtype: float64
```
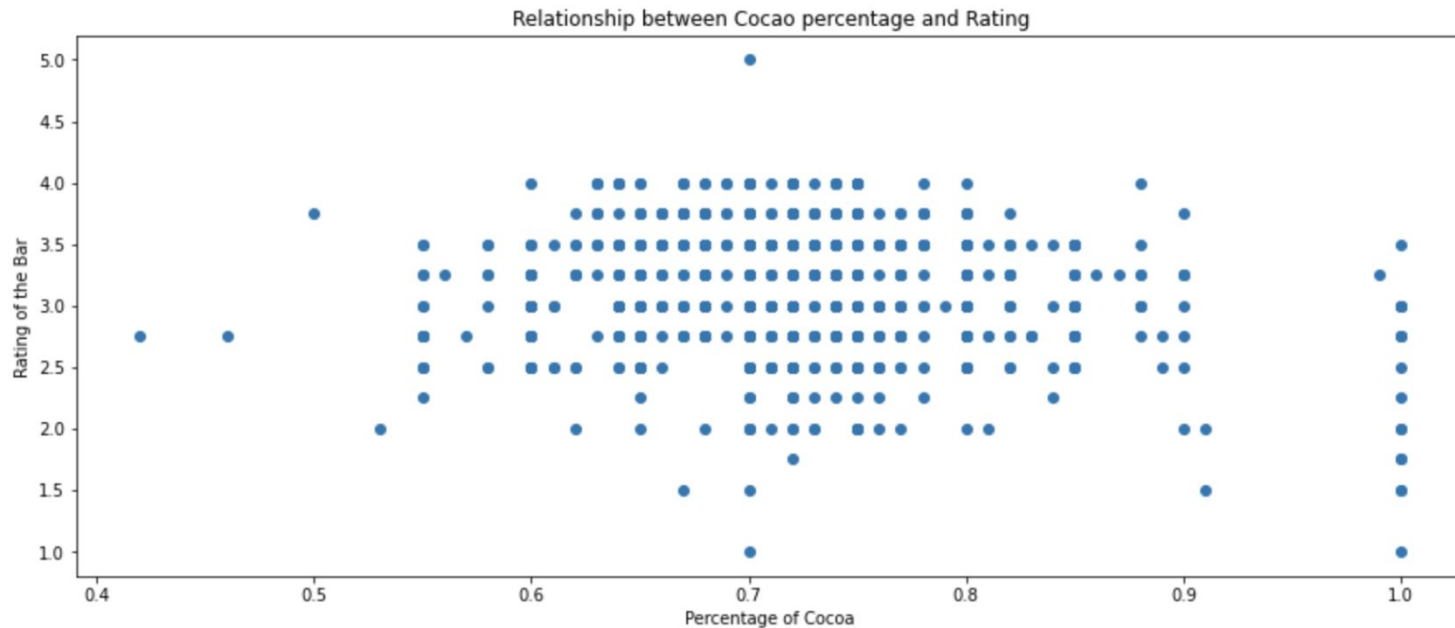
Does Cocoa % lead to better bar rating

Relationship between Cocao percentage and Rating

Best chocolate producing country

Top Chocolate Producing Countries in the World (Ratings above 4.0)