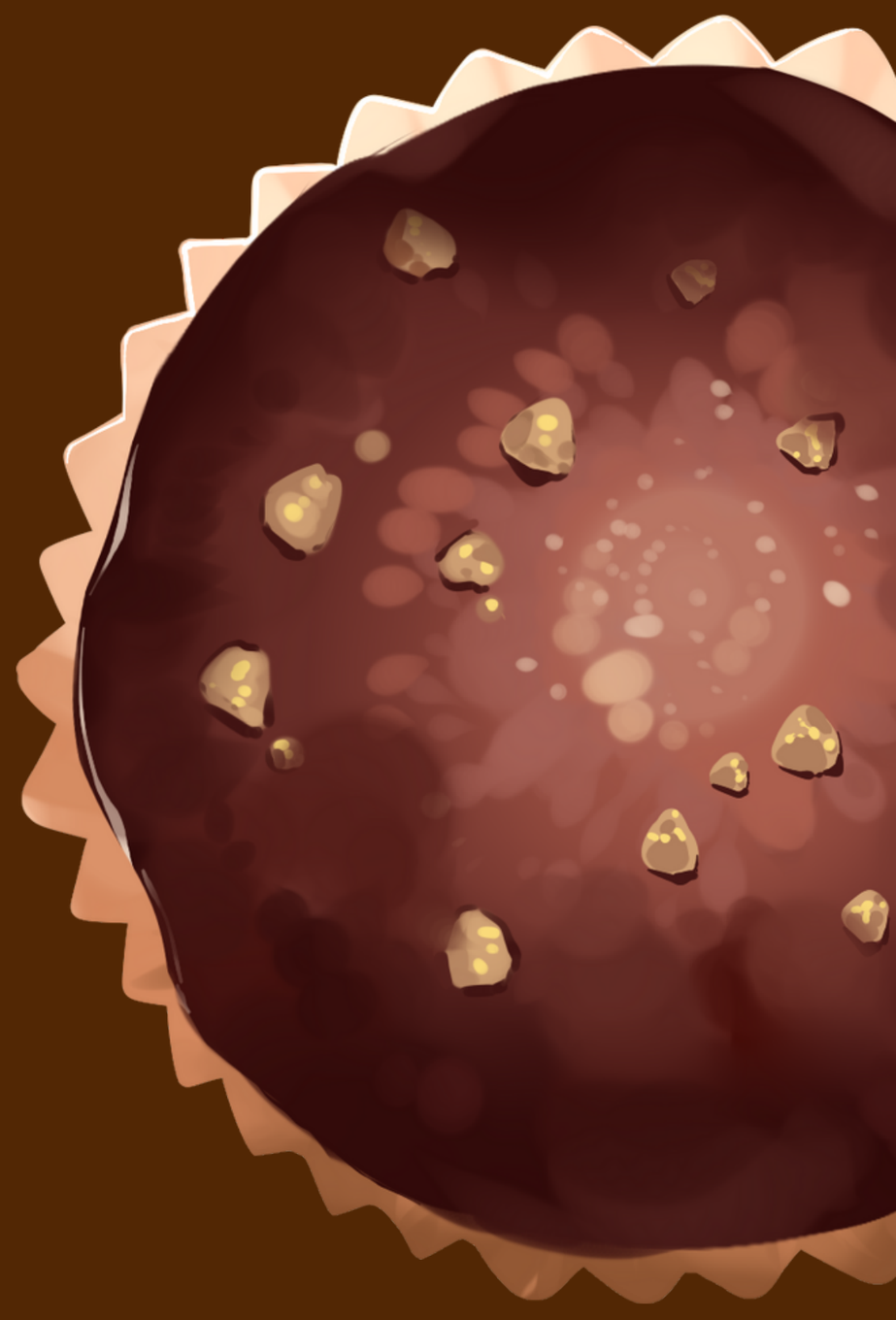
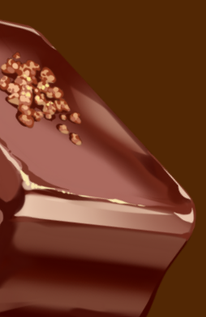
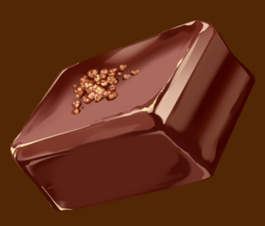
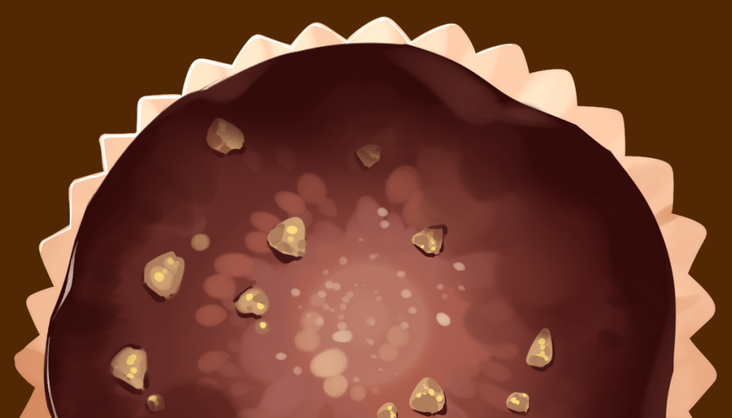
How do you like your chocolate?







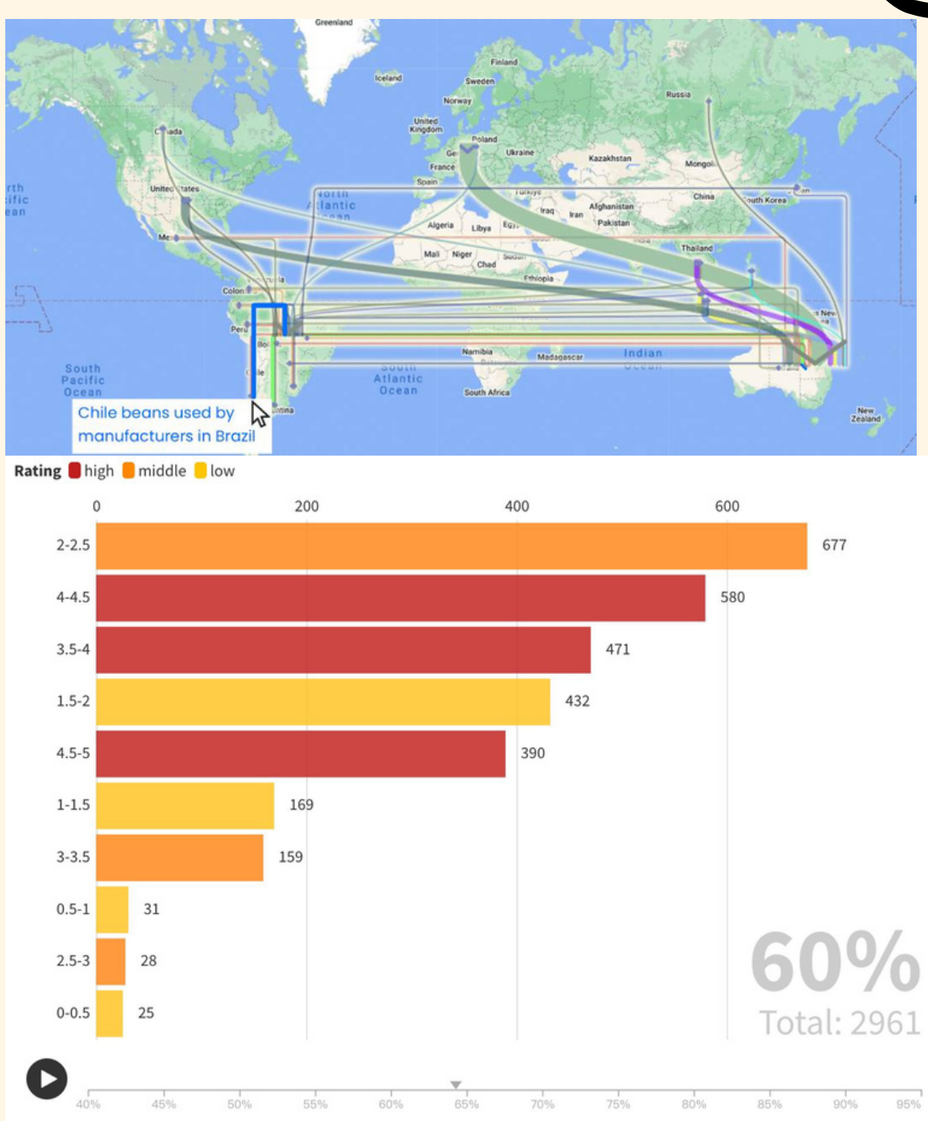


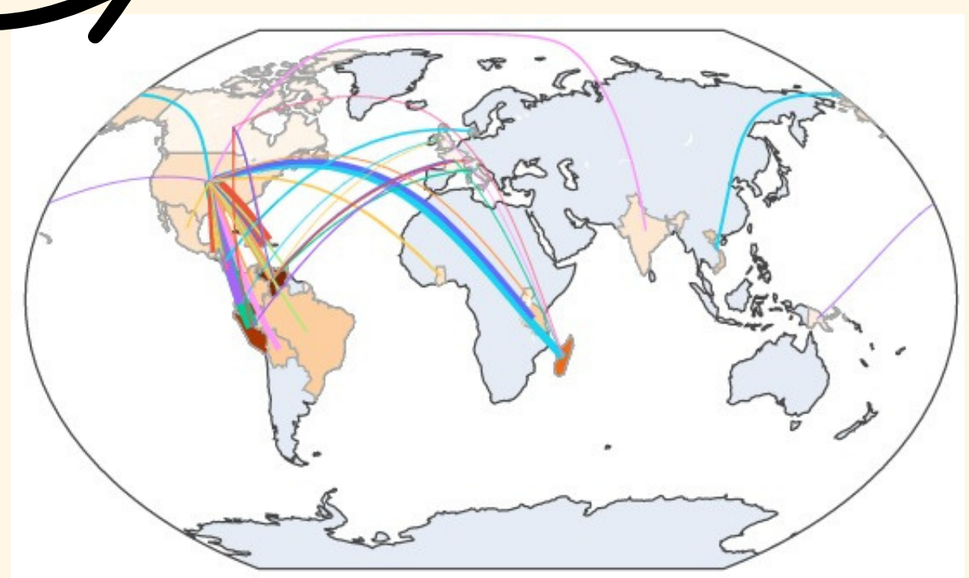
COM-480 Data Visualization
team: the-3-musketeers

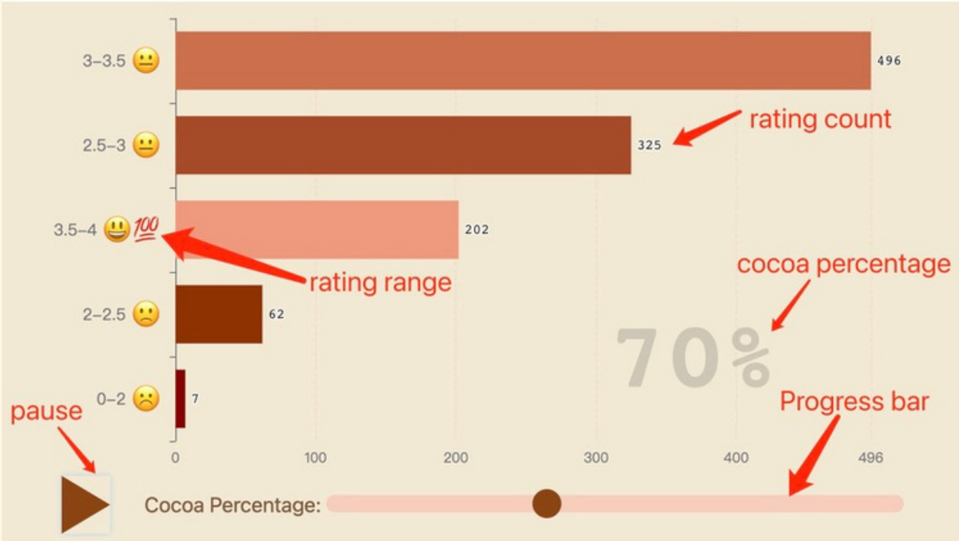
Introduction

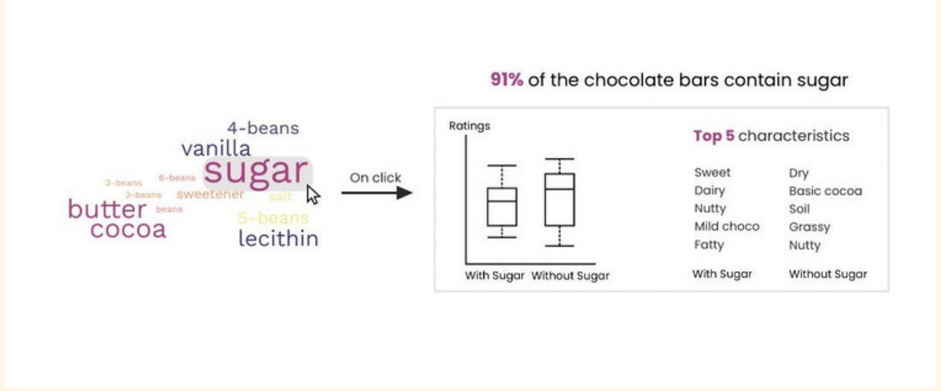
Our project revolves around exploring the intricate factors that influence chocolate quality, empowering producers to make well-informed decisions throughout the production process, and enlighting consumers about the fascinating world of the chocolate industry.

Our Journey









CHOCOLATE BAR

Path to obtain the final result

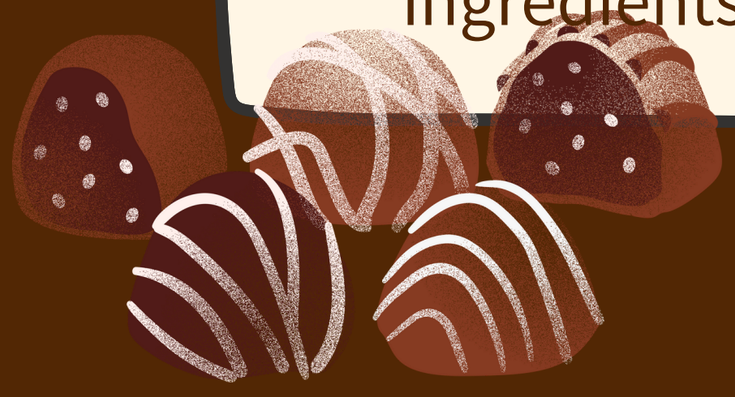
We first cleaned and preprocessed the dataset to ensure data quality, handle missing values, and remove any inconsistencies. Then we explored and analyzed the dataset to gain insights into its structure and variables. This provides a good basis for choosing the right visualization.

Considering our target users, we decided to focus on these areas:

- What affects the rating of chocolate
- Where do the chocolates on the market come from? How much?
- What flavors do consumers prefer in chocolate?

We came up with a number of ideas and in the end, taking into account feasibility and aesthetics, we kept the following three:

- **Flow map (MAP with SANKEY)** visualizes the flows of cocoa beans from certain origins to be used by manufacturers in certain locations.
- **Race bar chart** displaying rating changing over cocoa percentage in the form of an animated bar chart.
- **Word cloud** showing the comparison in the ratings and characteristics of chocolate bars based on specific ingredients and identifying patterns in how certain ingredients affect the overall rating of a chocolate bar.



Flow Maps

Interactivity

On both representations, users can hover over nodes (countries or sankey nodes) and links to obtain more information, such as number of links, weight of a link, etc.

Function

The world map representation allows us to see the geographical distribution of the main cacao bean flows. It presents the advantage of showcasing the dynamics not only between individual countries, but also between world regions. The sankey diagram allows us to better see the weights of the different flows, and which individual countries are the most involved in these flows, both as importer and exporter of beans.

Challenges encountered

The main challenge encountered is that the U.S.A. is by far more involved in these flows that any other country in the dataset. Thus, we needed to find a balance in the proportionality of links, and in the numbers of flows to showcase.

Changes since last milestone

Since the last milestone, we decided to add a sankey diagram along with the world map as a complement, since it provides us with additional information on each cacao bean flow in a readable manner. Moreover, we made the world map more minimalistic in order to better showcase the different flows, and to not visually overload the user.

Race Bar Chart

Interactivity

It's a dynamic animation. Users can pause the animation, or select a specific cocoa percentage then it will be static.

Function

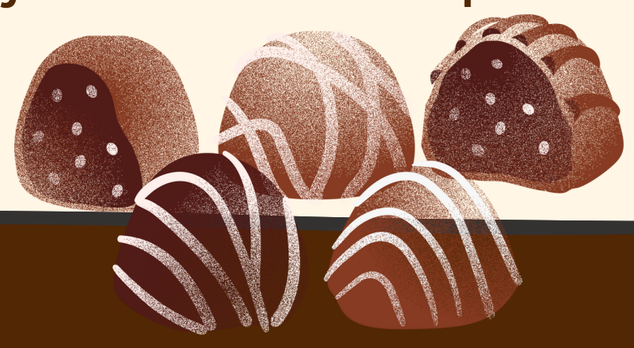
This graph gives chocolate producers a better understanding of the relationship between the rating score and the cocoa percentage. Helping them to make further production decisions about cocoa ratios.

Chanllenges encountered

When going deeper into our data, We found that the data was not so homogeneous, with only a very small amount of data in the 0-1 and 1-2 rating intervals, which would make our plot not so aesthetically pleasing, so we merged these two intervals without compromising understanding and interpretability. At the same time, chocolate products with less than 60% cocoa percentage were rarely present in the data set and their number was not statistically significant, so we excluded these samples.

Changes since last milestone

We have made some changes to the color scheme to harmonize the plot with the tone of our website. At the same time we give meaning to the brightness of the color: the lighter the color the higher its score. We've also added some emoji to make the plot more interesting.



Ingredients Word Cloud

Interactivity

The word cloud allows users to dive into their interested ingredients. By clicking on a word, a wealth of information regarding ratings and flavors associated with that ingredient will appear.

Function

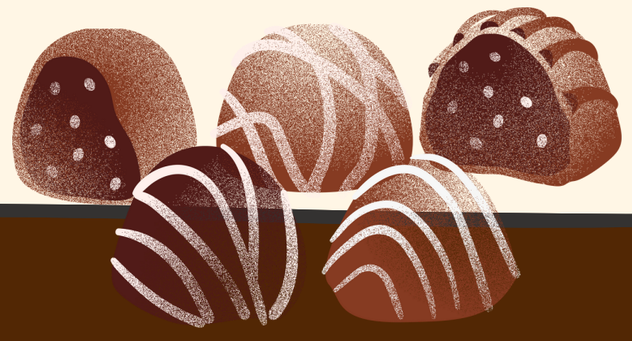
The word cloud visualizes an ingredient's frequency of being used. Once an ingredient catches your interest, we provide a box-plot showing the rating distribution, as well as the top 5 characteristics of chocolate bars both with and without that particular ingredient. This allows manufacturers to gain insights into how the chosen ingredient directly influences the flavors and overall ratings of chocolate bars.

Chanllenges encountered

Careful attention has been given to organize the display of divisions and sections in a visually pleasing manner. To enhance the user experience, font changes and transitions have been incorporated to improve the clicking animation.

Changes since last milestone

To explore the ingredients' roles, we have transformed the dataset into an ingredient-focused approach. There we construct for each ingredient an object with key, ratings statistics and top mentioned characteristics associated with it.



Designing the website

To design the website for our submission, we decided to use Github pages, a website hosting service offered by Github that uses the content of a repository to produce the content of the website itself. Github pages allows for easy customization using Jekyll themes.

After choosing a blog-style Jekyll theme called Quartz, we modified the relevant information regarding out project, like the titles, the footer, the copywrite and so on.

We then made the website revolve around the theme of chocolate. We did so by first changing the website colors scheme to a custom-made palette of brown chocolate tones. We then modified the website's favicon to show a chocolate bar.

With the form of the website all set, we went ahead and filled its content, first starting with an introduction explaining the goal of our project, then with three articles, each corresponding to a visualization.

Technologies we used

D3.js Echarts.js

Peer assessment

Dana Kalaaji:

- EDA
- Jekyll website build
- Responsible for the implementation of the Interactive World Map and the Sankey Diagram
- Work together with the reports of milestones 1 and 3

Dong Chu:

- Work together with the reports on each milestone
- Responsible for the implementation of the Word Cloud plot
- Video presentation

Zhan Li:

- Initial data processing and EDA
- Responsible for the implementation of the Race Bar Chart
- Work together with the reports on each milestone