**Conceptualization of research problem:**

According to food science, eating chocolate not only tastes good but it can also elevate your emotional state. Chocolate contains chemicals like tryptophan which the brain uses to produce serotonin, the neurotransmitter and hormone responsible for feelings of happiness and joy. So other than the obvious chemical effect in our bodies, what exactly makes a chocolate taste so good?

Chocolate is a billion dollar industry with many competitors offering their own take on the perfect chocolate recipe. With heavy competition in the market it is important for companies to analyze what characteristics make a chocolate delicious and contribute to creating repeat customers. Our team’s conclusion will help identify the major characteristics of good chocolate.

**Extent of data exploration:**

So, in the following figures shown below we have used count plots for analysis between the target variable (rating) and all the other features to find the highly correlated features.Chart, histogram

Description automatically generated

**Figure 1.1 Cocoa\_percent vs. Rating**

Figure 1.1 shows that the highest distribution of good chocolate ratings (greater than 2.5) are around 70% cocoa. Furthermore, this suggests that 70% is the optimal level of cacao in highly rated chocolate.

**Figure 1.2 Nutty vs. Ratings**Chart, bar chart, histogram

Description automatically generated

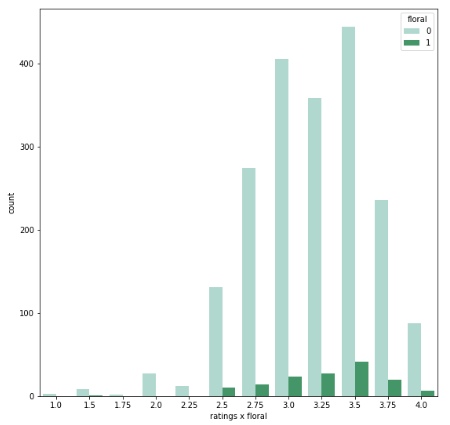
This figure shows that good chocolate ratings (greater than 2.5) include nutty flavor being present in the chocolate.

Chart, bar chart, histogram

Description automatically generated

**Figure 1.3 Roasty vs. Ratings**

This figure shows that good chocolate ratings (greater than 2.5) can include roasty flavor being present in the chocolate.



**Figure 1.4 Floral vs. Ratings**

This figure shows that good chocolate ratings (greater than 2.5) can include floral flavor being present in the chocolate.

**Figure 1.5 Spicy vs. Ratings**Chart, histogram

Description automatically generated

This figure shows that good chocolate ratings (greater than 2.5) can include spicy flavor being present in the chocolate.

Chart, bar chart

Description automatically generated

**Figure 1.6 Fruit vs. Ratings**

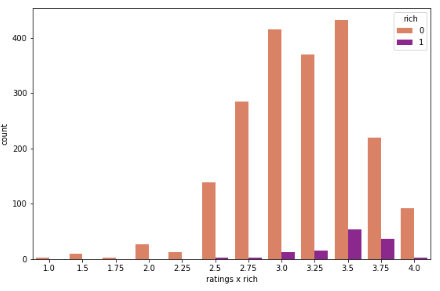
This figure shows that good chocolate ratings (greater than 2.5) can include fruit flavor being present in the chocolate.

Chart, histogram

Description automatically generated

**Figure 1.7 Creamy vs. Ratings**

This figure shows that good chocolate ratings (greater than 2.5) should include creamy flavor being present in the chocolate.



**Figure 1.7 Creamy vs. Ratings**

This figure shows that good chocolate ratings (greater than 2.5) should include rich flavor being present in the chocolate.

As you can see, count plot graphs show the bar for 0 is significantly bigger compared to the bar of 1, this is because there are very few entries for chocolate with these features present. So, we can see that the distribution for these features is mostly in the ratio 80/20 (absence vs presence). For this project we only worked on Dataiku and not on any other python notebook. So, it was a little too late for us when we got to know that we cannot perform oversampling on features, and that Dataiku only supports under-samplings.

**Models used:**

We used the following models to find out what characteristics contribute towards the likeability of a chocolate.

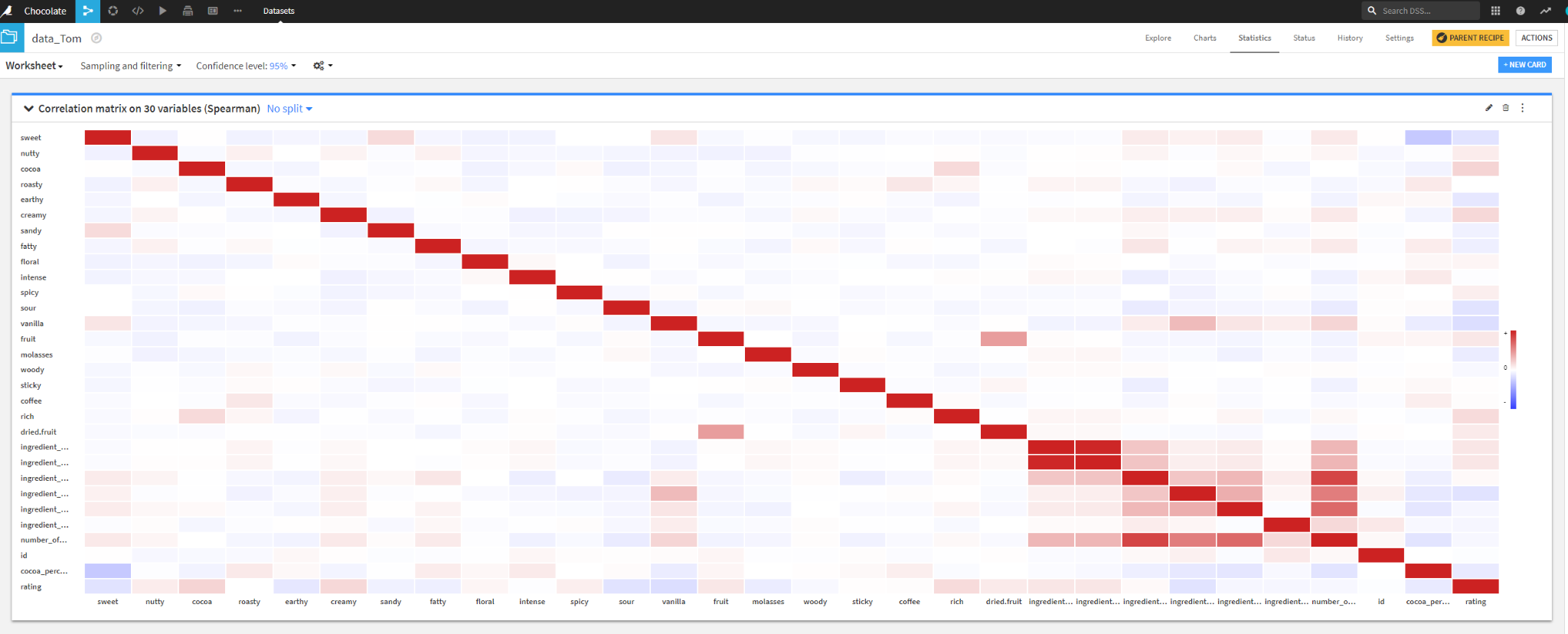
*Random forest, Gardient Boosted Trees, Ordinary Least Square, Ridge(L2), Lasso (L1), LightGBM, XGBoost, Decision Tree, SVM, SGD, Extra Tree, Artificial Neural Network and LASSO-LARS.*

Table

Description automatically generated

**Results:**

We found a high correlation between cocoa\_percent and rating. We also noticed a high correlation between ingredient\_Sugar, ingredient\_Beans and ingredient\_cocoa\_butter with the ratings. The chocolate characteristics such as Fruit,Cocoa, Creamy, Nutty and Rich contributed highly towards the rating. The RMSE value that we got for our machine learning model was 0.325.



**Conclusions:**

From the analysis we came to the conclusion that, the factors/characteristics that makes a chocolate likable according to the data are, that it should contain 70% of cocoa percentage and the preferred ingredients are Sugar, Beans and Cocoa-butter. We also got to know the important characteristics of a chocolate that make it likable according to our analysis were the creaminess, cocoa, nuttiness, richness, and fruitiness. Therefore, If the company wants to make a profit, they can increase these ingredients/characteristics in the chocolate to increase their chocolate rating.