Executive Summary for Product Quality Determination in Dark Chocolate Bars

**Problem**

Mershers’ Chocolates is a small chocolate company famous for its high-quality milk chocolate bars. The company is looking to expand its products by entering the dark chocolate bar market. The problem is that they are not sure which features of dark chocolate bars impact their quality rating. This project will improve the decision-making process for product development at Mershers by determining which variables and variables' values greatly impact a dark chocolate bar’s rating.

**Hypothesis**

The exploratory data analysis will discover at least three explanatory variables and their values that highly correlate to a high-quality dark chocolate bar.

**Summary of Data Analysis Process**

The data analysis process began by extracting the data from the database. I used Jupyter Notebooks and the Python library BeautifulSoup4 to extract and load the data into a CSV. From there, the CSV was loaded into Jupyter Notebooks in the form of a data frame. In Jupyter Notebooks I performed the initial observation, the data preparation (cleaning), and the Exploratory Data Analysis (EDA).

The initial observation included generating summaries of the whole dataset and individual variables including; data types, number of observations and columns, statistics, unique values, and simple distribution. The data preparation involved ridding the dataset of duplicate and null values, removing unnecessary columns, standardizing column names, verifying spelling and accuracy of data, and updating the columns to produce more accurate analysis.

After cleaning, I tested for normality in the quantitative data to determine if I needed to use parametric or non-parametric tests. The result was I could use parametric tests. I used ANOVA tests to prove statistical significance for categorical variables. I used Person’s R test to prove statistical significance for quantitative variables. The result was that the variables 'characteristic', 'cocoa\_percent', 'total\_ingredients','ingredient\_combination' were statistically significant. For each statistical signing variable, I performed a univariate analysis and a bivariate analysis against ‘rating’. I summarized interpretation and insights under each plot. Finally, I visualized the results of the EDA through a data story on Tableau.

**Findings**

The following are recommendations based on dark chocolate bars with ratings at or above 3.5. Order of recommendation with first being the top choice. The top 3 cocoa percentages and frequencies were 0.7 with 412, 0.75 with 95, and 0.72 with 91. The top 3 total ingredients and frequencies were 3 with 397, 2 with 267, and 4 with 149. The top 3 combinations of ingredients used and frequencies were BSC with 394, BS with 261, and BSCL with 110. The top 3 memorable characteristics and frequencies were nut with 135, cocoa with 123, and creamy with 119.

**Limitations of techniques and tools.**

The tools used were Python and Jupyter Notebooks. The technique used was Exploratory Data Analysis (EDA). The limitation of Python was that it can be slow and require a great deal of memory. The limitation of Jupyter Notebooks came from the lack of auto-completion for Python variables, autosave was variable and was not robust to handle anything beyond exploratory data analysis. The limitations of EDA was if it is not accurately performed, it could produce incorrect insights that can lead an analysis astray.

**Summary of Proposed Actions**

There are two proposed actions for this project. The first logical course of action would be to extend the analysis of dark chocolate bar rating to review by general consumers. Extends the analysis would provide ratings from a larger sample of ratings. Ratings from credible organization provide direction to the development, but in the end, it is the consumer purchases the drives the final decisions. The second logical course of action would be to create a predictive model based on the data research to predict the rating prototype dark chocolate bars formulated by the product development team. This would allow the product development team to test small batches of prototypes, decreasing time spent in field research and funding used by reducing the potential number of failed prototypes.

**Expected Benefits of Study**

The expected benefits of the study were to assist the product development team’s decision-making process by providing necessary and accurate information. Analyzing what variables of a product influence its quality rating gives the company a way to target product development to specific variables. This allows a company to not waste time and funding to produce a product where there is a great deal of uncertainty it will succeed.