**Project 1 – Written Report**

**Kim’s Hypothesis**

There are a total of four flavors that were used by Team 2 from a dataset found on Kaggle. Team 2 compared these four flavors to their respective rating, effects, location, and breeder. After cleaning the data, Team 2 realized that flavor four only had 43 datapoints out of 2351 entries. The team decided to drop this flavor four dataset due to it being so small compared to the other flavors.

A research question the team posed was that the sweet flavor cannabis would have the highest ratings. Typically, when looking at flavors one would think that the population would lean towards favoring sweet, flavored products as seen with popular vape flavors. Due to this assumption, the team decided that sweet flavor cannabis would be the most popular flavor and have the highest rating.

The raw data was cleaned to breakout the flavors into flavor one, two, and three by taking the first listed flavor in the entire “Flavor” column listed in the original data frame (reference figure 1 and 2). Those flavors made up the flavor one profile. The same steps were repeated to make the flavor two and three profiles. This data is what was used for statistical analysis..

A screenshot of a computer

Description automatically generated

Figure 1: Original data frame

A screenshot of a computer

Description automatically generated

Figure 2: Flavors broken out to separate to columns from “Flavor” column in original data frame

A screenshot of a computer

Description automatically generated

Figure 3: Combined broken out flavors into original data frame

Flavor profile one resulted in its ratings spanning across a range of 4.1 to 4.9. There were 47 flavors that made up this profile. The minimum rating went to a flavor called blue and the highest rating was for Tea. The mean for all the ratings in flavor profile one was 4.43 and the mode was 4.40.

The flavor profiles were large and the cleanest way to visually depict the datapoints per flavor would be to use a box plot chart. At a quick glance, the boxplot shows the minimum, first quartile, median, third quartile, and maximum values. Analyzing the graph for flavor profile one, the median lines line up horizontally across the board and there were a limited number of outliers. The y-axis shows the rating, and the x-axis lists out each flavor.

A chart with different colored boxes

Description automatically generated

Figure 4: Box plot of flavor one profile

Flavor profile two had its ratings span across a range of ratings from 4.2 to 4.6. It encompassed 46 separate flavors. The minimum rating went to a flavor called apricot and the highest rating was for violet. The mean for all the ratings in flavor profile one was 4.45 and the mode was 4.50. The box plot of flavor profile two has a similar resemblance to flavor profile one.

A chart with different colored squares

Description automatically generated

Figure 5: Box plot of flavor two profile

Flavor profile three ratings spanned across a range of 4.2 to 4.6. It was comprised of 48 flavors. The minimum rating went to a flavor called blue and the highest rating was for coffee. The mean for all the ratings in flavor profile one was 4.45 and the mode was 4.40. The box plot of flavor profile two has a similar resemblance to flavor profile one.

A chart with different colored boxes

Description automatically generated

Figure 6: Box plot of flavor two profile

Comparing all three flavor profiles, the datapoint values are extremely similar with tight ranges, similar modes and means which inclines Team 2 to trust comparing values across all the flavors. Each flavor profile our roughly equal, giving a high confidence level that the data samples are similar enough to compare for statistical analysis.

In conclusion, hypothesis number 2 was incorrect with assuming sweet as having the highest rating. The team was surprised to find that tea had the highest rating with coffee closely behind it. One would not naturally associate cannabis flavors being popular for mimicking beverage flavors, but coffee and tea are drank nationwide so maybe consumers are accustomed to those flavors. The team would want to explore if this result would be seen globally as well since coffee and tea can be found worldwide. Future research of this dataset would include grouping the flavors into larger groups for analysis. Grouping flavors by the five different types of taste on human tongue receptors such as salt, sweet, sour, bitter, and umami would allow for more general analysis on what flavors are preferred by the population.