Zelda Mazur

Code Labs Data Science

September 23, 2024

# Class 1 Exercises

*[Describe your dataset, and describe what you hope to find from your dataset]*

**Example**: Today, I’m looking into a dataset about coffee costs in different regions and how that compares to its quality. I hope to find if there is truly a correlation between a coffee’s quality and its price. Perhaps there are correlations between different attributes and the data will tell us so.

### Descriptive Statistics

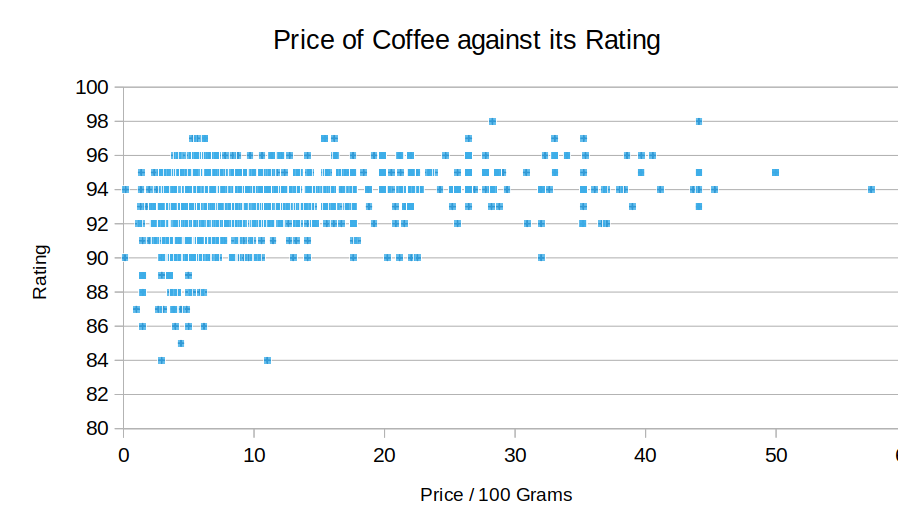
*[Talk about the descriptive statistics for quantitative data in the set: Means, medians, modes, ranges etc that stick out to you. You can build histograms in Excel to support your findings.]*

**Example**: In this set, the mean coffee price per 100 grams is $10USD, but I noticed the median is much lower. Given the range in price from $0.12 to $132, I suspect the large outliers is skewing the mean towards a much higher price.

### Correlations

*[Talk about correlations between two attributes in your dataset. Let’s also see a scatter plot. Did anything surprise you?]*

**Example**: When looking more deeply into my dataset, I found that there wasn’t much correlation between the price of the coffee and its quality: it only correlates 25.99% of the time. I believed at first there would be a strong correlation, that the highest-priced coffee would also be the highest quality. This simply isn’t the case. We can see that with the following scatter plot.

****