**Does Coffee Stunt Growth?**

**Milan Sherman**

As an avid coffee drinker of modest stature, I am interested in the myth that coffee stunts your growth. This myth is often cited when attempting to discourage young people or teens from drinking coffee. Apparently, this myth started based on study that showed a correlation between people who drink coffee and low bone density (<https://theroasterie.com/myth-or-fact-coffee-stunts-your-growth/>, 2020). As low bone density is connected with osteoporosis, which in turn can cause people to actually shorten as they age, it was then concluded that coffee stunts growth. This idea was further perpetuated by a company that created a caffeine-free coffee alternative in the 1930s.

To refute this claim using the CRISP-DM process, I would begin by carefully considering the exact claim to be tested, as the rest of the process depends on that. For example, based on the above, are we testing the claim that coffee stunts growth, or that caffeine causes low bone density, or that coffee causes osteoporosis? This is the step of clarifying and understanding the business question, i.e. business understanding. Ultimately, the myth I am interested in is whether coffee stunts growth, so that is my focus. This question then drives a number of considerations with respect to the data needed to answer it: what should be the age(s) of participants (they should still be growing), is it ethical to incentivize participants to drink coffee, over what period of time should data be collected in order to detect differences in growth rates, how often does someone need to drink coffee to be considered a coffee drinker, how much data needs to be collected, how do we deal with attrition (participants who quit drinking coffee during the study), etc. Once these questions are answered, another important question is whether such data already exists or needs to be collected, as this will have a major impact on the resources required and the timeline of the project. This is all part of understanding the business understanding of the problem.

Subsequent steps are simpler once this first step is complete. In the data understanding step, we collect the data needed, either through identification of pre-existing datasets that may meet our criteria, or through conducting a study. Once this data is collected, we explore it in order to assess its quality in relation to the business question. This includes assessing the measures collected, e.g. participants’ ages vs. birth dates, looking for incomplete observations, outliers or bad data (such as typos), and summarizing and possibly visualizing each of these dimensions to gain a better understanding of the data. This process is completed for each dataset that is obtained.

In the data preparation step, we decide which dataset or datasets we will use based on the previous two steps, i.e., the business requirements and the data quality. In addition, we prepare and potentially merge the data, if there is more than one dataset. This may also include converting some of the dimensions to metrics more suitable for answering the question. For example, if the data has participants’ birth dates, and the dates that they began drinking coffee, these fields would likely need to be converted to ages and time periods, and if more than one dataset is used, it will be necessary to convert to common measures before merging.

The data modeling step for this example will depend on the data that is available. For example, if we have data with two samples of subjects, some of whom drink coffee and some that do not, then a simple t-test or other non-parametric test may suffice to test the claim. If our data only contains participants who drink coffee, then we might create a model that would predict their growth based on data prior to drinking coffee, and use this model to compare their actual growth after they began drinking coffee. Whatever models or statistical tests are used, they will need to be evaluated for accuracy and confidence, and potentially tweaked and iterated.

In the evaluation step, we consider the models generated and the degree to which they meet the criteria outlined in the first step related to business understanding. Strengths and weakness of each model or test are summarized, and either 1) one model/test is chosen, or 2) the results of more than one are integrated to provide an answer the question. In this example the final step, deployment, would likely consist of a write up of the results, describing and summarizing the entire process and providing an answer to the original question. An important part of each of the steps in the process is a summary report documenting what was done and summarizing the decisions and/or findings. In addition to being leveraged in the final report, these reports are important for communicating with stakeholders throughout the process and provide a basis for discussion of adjustments that may need to be made at each step.

References

Myth or fact: Does coffee stunt your growth? (2020, July 07). Retrieved March 29, 2021, from https://theroasterie.com/myth-or-fact-coffee-stunts-your-growth/