**Using CRISP-DM to Determine if Coffee Stunts One’s Growth**

\*\*\*\* \*\*\*\*\*\*\* \*\*\*\*\*

College of Science and Technology, Bellevue University

DSC500-T301: Introduction to Data Science (2227)

Sashidhar Bezawada

June 17, 2022

**Belief that Coffee Stunts Growth**

Growing up, many of us were cautioned against drinking coffee. This is due to a commonly held belief that coffee stunts your growth. Researchers believe that the “fact” originated from a perceived link between coffee and osteoporosis. Additionally, science shows that caffeine consumption can lead to an increased elimination of calcium. However, the link between coffee and osteoporosis with height loss has not been confirmed. The previous studies seemed to determine correlation, not causation because the beverage choices of coffee drinkers may feature less calcium-rich drinks (“Can coffee really stunt your growth,” 2020).

**CRISP-DM Methodology Applications**

To investigate the claim that coffee stunts one’s growth, questions and analysis must be performed. CRISP-DM is the recommended methodology to be followed to refute this claim. The primary steps to be followed include business understanding, data understanding, data presentation, modeling, evaluation, and deployment (“Crisp dm methodology,” 2020).

**Business Understanding**

To begin, an objective must be set. In this case, the objective is to determine the answer to the question “Does regular coffee consumption stunt one’s growth?” With the data, we may be able to refute this commonly held belief. Therefore, the business success criteria would be proving that drinking coffee does not stunt one’s growth.

An assessment of the situation shows that many people believe that drinking coffee, especially at a young age impacts one’s height. There are not existing extensive studies on this topic; however, some studies and theories are available. The medical data may be limited to aggregate numbers due to patient confidentiality laws and regulations. Through data mining, the findings could be used to prove that there is not a correlation between height and coffee consumption.

There must be enough data available to determine if there is a definitive link and that causation is clear. The project plan centers on utilizing readily available public resources such as Google and scholarly articles as tools. There may be significant limitations on the data availability, which could impact final project delivery timing.

**Data Understanding**

To collect the initial data, we will look for existing studies and scholarly articles to review any past methods and conclusions. We must seek out publicly available coffee consumption and height data. Scientific journals may help provide the medical information to determine the age at which the main stages of the body’s growth occur. This age data can then be compared with age data on when most users begin drinking coffee. Research on the genetics side of growth impact will also help direct the types of health data required.

Once the data is collected, the surface properties of the data can then be described and explored through simple analysis. A data exploration report will be drafted. Then the data quality must be evaluated. Gaps in the data should be noted and verification for the identification of errors must occur.

**Data Preparation**

Next, the data will be selected and readied for data mining. This data may include the number of individuals suffering from osteoporosis and the number of patients that lose height because of fractures. Additionally, data on the typical amount of milk and other calcium rich drinks that coffee drinkers consume may be utilized. Once the data is selected, the data is then cleaned, constructed, integrated, and formatted.

**Modeling**

Once the data is ready, the modeling technique will be selected, and any modeling assumptions must be captured. Then we will generate the test design and build the model. During this process, the parameters will be recorded as well as the model description. A possible model may trend the heights of regular coffee drinkers compared to the national average heights. After the model is created, the team must assess the model. At this time, the parameter settings should be fine-tuned.

**Evaluation**

As the CRISP-DM processes reaches final stages, the results must be evaluated to determine whether the business objective has been met by the model. Did we establish causation and correlation between drinking coffee and lessened growth or increased likelihood of osteoporosis? If we could not, does our model show substantial backing to refute the claim? The process should be reviewed and any steps that need to be repeated should be noted. After which, next steps may be determined.

**Deployment**

Finally, we reach the stage of deployment. At this stage, the strategy, steps, and instructions are documented. An outline of future plan monitoring and maintenance should be laid out. Lastly, the final report is produced and the results are presented. As a review of the process, the team should collaborate to capture any failures, hints, and report.

**Conclusion**

In summation, the CRISP-DM methodology can be successfully utilized to prove or disprove commonly held beliefs that individuals regard as “facts.”

**Reference**

*Can coffee really stunt your growth?* Harvard Health. (2020, January 7). https://www.health.harvard.edu/staying-healthy/can-coffee-really-stunt-your-growth.

*Crisp dm methodology*. Smart Vision Europe. (2020, June 17). https://www.sv-europe.com/crisp-dm-methodology/#one.