**The Impact of Assumptions**

Bilal Kudaimi

Department of Data Science, Bellevue University

DSC 530: Exploratory Data Analysis

Dr. Matthew Metzger

August 29, 2020

Ever since I discovered the TV show *Mythbusters*, I have thought about the impacts that myths and urban legends have on society, as well as how many myths remain to be disproven. The show revolved around performing tests and data analyses to either prove or disprove urban legends such as “cellphones cause fires at gas stations.” Consequently, when I received news that I was to be analyzing an assumption using data, I was absolutely elated. While many myths remain untested, I will focus on the analysis of one such myth: Carrots improve your vision. I will be discussing the origins of the myth as well as how the CRISP-DM process can be applied to data from a research study to uncover the reality behind the myth. I believe that CRISP-DM is the key to extracting patterns from “clean” data.

Growing up, many of us have heard time and time again of the myth that carrots improve eyesight. The truth is that the myth was invented by the Allied forces of World War II to hide the existence of radar. Radar had recently been deployed to intercept German bombers before they reached Britain, but the British had to keep its existence a secret to avoid the possibility of a German counterattack. To this end, they started a propaganda campaign to cover it up by claiming that their pilots ate carrots – then known to be required for vision – to improve their eyesight and see the bombers from afar. The myth persists to this day due to the science *seeming* correct (if carrots are good for my eyes, then eating more *must* improve eyesight!).

So far, the origin might seem like just a governmental stretch of the truth, so how can we know if carrots trulyimprove eyesight? The known facts are 1) the human eye requires specific nutrients, and 2) carrots contain high amounts of said nutrients. Surely, this means eating more carrots improves eyesight, right? Not exactly; the eye needs these nutrients but presenting more will not improve vision. We can test this null hypothesis by using a controlled study, then using part of the CRISP-DM process to analyze our collected data and uncover the truth behind the orange vegetable. CRISP-DM is the Cross-Industry Standard Platform for Data Mining, which describes a process for data mining in six steps.

For the sake of simplicity, we will assume that the controlled research study contains one control and one experimental group. Both are sufficiently large to produce meaningful data and are of similar demographics. The experimental group eats a very large serving of carrots daily and the control eats no carrots. Eye exams will be administered regularly to check for sight improvement. With the acquired raw data of carrot consumption and eye exam results, we can apply CRISP-DM to mine the data for patterns.

Now that we have the question we want answered – whether carrots improve eyesight – then we would assess what data we would require to tackle our objective (the “Data Understanding” step), in this case the participants’ carrot consumption and eye exam results from the research study. We would then “clean” the data of outliers and get it into the form of a table displaying the aforementioned data (the “Data Preparation” step). This clean table will help us to better visualize the data. I believe scatterplots are the preferable visual medium here, as they better illustrate change over time.

The “Modeling” step is when we would apply algorithms to the data to discover any latent trends or a lack thereof. For example, if we are using scatterplots, we could build a regression model that would give us the Pearson correlation and the slope for vision test results over time and vision test results versus daily carrot consumption. The Pearson correlation shows the relationship, if any, between variables and it will help us arrive at a conclusion regarding carrots and eyesight (we fail to reject the null hypothesis; there should be no positive correlation between eating more carrots and eyesight). At this point, we will have answered our question by refuting any correlation between carrots and *improved* eyesight, so we can choose to conclude the CRISP-DM process here.

The use of data mining to investigate myths has made me realize that data analysis could be used in ways I have never even thought about, such as this myth “busting.” I’m left to ponder the question of if

References:

1. Azevedo, A. & Santos, M. (2008). *KDD, SEMMA, and CRISP-DM: An Overview.* IADIS European Conference for Data Mining. Retrieved August 30, 2020 from https://www.semanticscholar.org/paper/KDD%2C-SEMMA-and-CRISP-DM%3A-a-parallel-overview-Azevedo-Santos/6bc30ac3f23d43ffc2254b0be24ec4217cf8c845?p2df.
2. Kelleher, J. D., & Tierney, B. (2018). *Data Science.* The MIT Press.
3. Webb, D. (2017). *Do Carrots Improve Eyesight?* University of California, Berkeley Wellness. Retrieved on August 30, 2020 from https://www.berkeleywellness.com/healthy-eating/nutrition/article/do-carrots-really-improve-eyesight.