
Project Motivation

Access to potable water is both a necessity for life and a basic human right; however, the vast majority of new and existing water sources are unfit for consumption. Determining which factors have the greatest effect on a water source's potability will provide people with concrete measures of the quality of their water. Additionally, understanding the importance of water characteristics as they relate to water quality will allow for a better understanding of how the water quality can be improved, which has significant health and development implications. For these reasons, along with my experience in researching the quantity and transport of water resources, I have chosen to examine water quality datasets for my data science capstone project.

Project Hypothesis

With my project motivation in mind, the specific questions I am seeking to answer through the data analysis are as follows:

1. What physical and chemical characteristics of water are most important to consider when determining water quality?
2. Do the physical and chemical characteristics exhibit any threshold values above which water is not considered potable?
3. Are any singular characteristics statistically significant in determining whether water is potable (safe for consumption)?

Using these questions, I form my null hypothesis, which is written below:
H_0 : There is no statistically significant relationship between a given physical or chemical characteristic of water and the water's potability.

Project Background Information

According to the National Science Foundation, water covers about 71% of the earth's surface, though 97% of the earth's water is found in the oceans, rendering it too salty for consumption and most agricultural or industrial uses. 3% of the earth's water, then, is considered fresh water, but 83% of this resource is unavailable as it is trapped in glaciers, ice caps, and the atmosphere, or is located too far below the earth's surface to be extracted. Thus, only 0.5% of earth's total water is available fresh water; however, even this small amount is not guaranteed to be potable - much of the world's available fresh water is contaminated.

Ample research has been conducted with regard to the different types and extent of water contamination as water is an essential resource and is uniquely vulnerable to pollution. Per the Environmental Protection Agency, the most common types of water pollution within the United States include agricultural pollution (fertilizer, waste, and pesticide runoff), wastewater pollution (stormwater runoff, sewage, and industrial byproducts/sludge), and oil pollution (spills, vehicle discharges, and plastics). Worldwide, water pollution types largely remain the same, with nutrient pollution from agricultural activities constituting about 60% of all freshwater contamination according to the World Health Organization.

Researchers have detailed the significant implications that consumption of contaminants has on human health as well as societal development. According to the United Nations, every year, approximately 297,000 children under five die from diseases linked to poor sanitation, poor hygiene, or unsafe drinking water (wastewater pollution). Beyond this, the Natural Resources Defense Council states that chemical pollutants, such as pesticides, fertilizers, and heavy metals can cause serious health problems if ingested, including cancer, hormone disruption, and damage to the immune, nervous, cardiovascular, lymphatic, and reproductive systems. A recent and very prominent (not to mention personal) example of the dangers of chemical pollutants is the Flint Water Crisis, in which residents of Flint, MI, ingested water that contained lofty levels of lead and other toxins, per Michigan's Environment, Great Lakes, and Energy agency. This immediately resulted in health problems like rashes and itchy skin; however, the long-term effects of this crisis are still to be observed in full.

While contamination may not be apparent from water's physical appearance, it does present itself in the physical and chemical characteristics of water. For instance, according to the Environmental Protection Agency, a body of water bearing a total dissolved solids measurement above a certain limit indicates a high proportion of minerals within the water. These minerals often result from nutrient runoff, which provides a good measure for agricultural pollution. Additionally, turbidity measurements are often used to gauge the light emitted from water resources (can indicate oil pollution), and trihalomethanes or organic carbon measurements can detect how much organic material is dissolved in the water (can indicate wastewater pollution). Examining these physical or chemical properties, along with several others, is useful for gauging whether or not a given water resource is contaminated, and thus may be useful for determining whether water is safe for consumption.