

# Description of the "Smoking and Drinking" Data set

1. sex: Indicates the gender of the individual - male or female.
2. age: age was categorized into 5-year intervals; i.e. 20-24 years, 25-29 years, 85+ years. These intervals were then converted into numeric values by taking the lower bound of the interval.
2. height: Indicates the height of the individual (Rounded upto 5cm).
3. weight: Indicates the weight of the individual (Rounded upto 5kg).
4. waist: Indicates the circumference of the individual's waist.
5. sight\_left: represents the visual acuity of the individual's left eye. Visual acuity measures the ability of the eye to distinguish shapes and object details at a given distance. This ranges from 0.1-2.5, with values < 0.1 being shifted up to 0.1.
6. sight\_right: represents the visual acuity of the individual's right eye.
7. hear\_left: evaluates the person's left ear's hearing, assigning a 1 to any abnormal hearing and a 2 to normal hearing.
8. hear\_right: evaluates a person's right ear hearing using the same categorization scheme as hear\_left.
9. SBP: determines the person's maximum systolic blood pressure, expressed in millimeter-hours (mmHg). The pressure in the arteries during a heartbeat is measured by the systolic blood pressure.
10. DBP: determines the person's diastolic blood pressure, expressed in millimeter-hours (mmHg). The pressure in the arteries during the period between heartbeats is measured by diastolic blood pressure. The source mentioned above.

11. BLDS: measures the individual's fasting blood glucose, in mg/dL. This represents the concentration of glucose per 100ml of blood prior to eating a meal.

12. tot\_chole: measures the total concentration of (ester and non-ester) cholesterol in the individual, in mg/dL.

13. HDL\_chole: calculates the total cholesterol concentration in a person's HDL (high density lipoprotein) region, expressed in milligrams per deciliter. Good cholesterol, or HDL cholesterol, is a type of cholesterol that returns to the liver after being absorbed from the blood and eliminated. keep in mind that having higher HDL cholesterol can reduce your risk of heart disease.

14. LDL\_chole: quantifies the total cholesterol concentration in the LDL (low density lipoprotein) region in milligrams per deciliter. The majority of the cholesterol in the body is LDL cholesterol, sometimes referred to as bad cholesterol. Elevated levels of this may increase the risk of stroke and heart disease.

15. triglyceride: calculates the blood's total triglyceride content, expressed in milligrams per deciliter. One kind of lipid, or fat, that moves through our bloodstream is triglycerides, which are typically derived from the food we eat.

16. hemoglobin: measures the total concentration of hemoglobin in the individual's blood, in g/dL. Hemoglobin is a protein in our red blood cells that carries oxygen

17. urine\_protein: determines the person's urine's protein content. Elevated protein levels in the urine, also known as proteinuria, may indicate a number of health issues, including renal and heart failure. supply. The label encodings for this variate (1(-), 2(+/-), 3(+1), 4(+2), 5(+3), and 6(+4)) are not clear; however, we could surmise that they represent the number of standard deviations the value in the feature distribution resides at from the mean.

18. serum\_creatin: measures the concentration of creatinine in the individual's serum (which resides in their blood), in mg/dL. Creatinine is a waste product of creatine, which is produced to supply energy mainly to the muscles. Usually, this

is removed by your kidneys entirely; thus, if kidney function is abnormal, the concentration of creatinine will increase

19. SGOT\_AST: measures the SGOT (Glutamate-oxaloacetate transaminase) - AST (Aspartate transaminase) value in IU/L, which are values in blood tests that quantify liver, heart and other organs' performance. In particular, when these are damaged, the value of this goes up. Normal values are around 0-40 IU/L.

20. SGOT\_ALT: measures the SGOT (Glutamate-oxaloacetate transaminase) - ALT (Alanine transaminase) value in IU/L, which are values in blood tests that quantify liver performance. In particular, when this are damaged, the value of this goes up. Normal values are around 0-40 IU/L.

21. gamma\_GTP: determines the gamma-GTP (γ-glutamyl transpeptidase) value in IU/L, which is a bile duct value that indicates liver function. For men, the normal range is 11–63 IU/L, and for women, it is 8–35 IU/L.

22. SMK\_STAT\_TYPE\_CD: measures the individual's smoking state, where 1 corresponds to them never smoking, 2 corresponds to them having used to smoke but quit, and 3 corresponds to them still smoking.

23. DRK\_YN: is a flag that indicates whether the individual is a drinker or not.