Data Cleaning Change Log

Alexander Brown

10-19-2023

**Age:** Split into two groups, adults 18-64, and older adults 65 and older, as per National Institutes of Health (2022).

**Sex:** Changed to binary for processing.

**Cholesterol:** Split into three categories. According to Robenson R (2021) less than 200 total is normal, 200-239 borderline high, 240 or greater is high.

**Blood Pressure:** Reclassified as hypertensive or not based on medical recommendations.

**Hypertensive:** Derived from blood pressure levels. If Diastolic is >90 or Systolic is >140 when Diastolic is <90 is 1, else 0. Blood pressure statistics we’re taken from Clinical Methods: The History, Physical, and Laboratory Examinations 3rd edition (Walker et al, 1990).

**Heart Rate:** According to Laskowski E (2022) a heart rate above 100 can be considered tachycardia and if the person is not a trained athlete a heart rate below 60 is considered bradycardia. For this instance, we will consider anything outside of these normal ranges as an issue, setting 1 as detection of a problem or 0 if within normal range. It is noted that more data would be required to determine if someone is a trained athlete.

**Diabetes:** Unchanged

**Family History:** Unchanged

**Smoking:** Unchanged

**Obesity:** Removed due to overlap with BMI

**Alcohol Consumption:** Unchanged

**Exercise Hours per Week:** Rounded to nearest whole number. Categorized by Inactive or Active, assumed to be moderate activity. Considered active if 2.5 hours or more per week, or inactive if less than 2.5 hours per week. Assumptions about activity were made based on information from Laskowski E (2023)

**Diet:** Changed to numeric, 0 for unhealthy, 1 for average, 2 for healthy

**Previous Heart Problems (Not Heart Attack):** Unchanged

**Medication Use:** Removed due to conflict of attribute. Does not provide useful information. For example, higher if on medication that increases risk of heart attack, or lower if on medication to control risk factors.

**Stress level:** According to an article by American Psychological Association (n.d.), a 3.6 on a 10-point scale is considered a healthy level of stress. While this is not an exact measure, it gives us a place to start. For this metric we will split them into two groups, a 4 or below will be considered healthy or nominal, represented by a 0, while a 5 or above will be considered high stress, represented by a 1.

**Sedentary Hours Per Day:** Removed. According to Park et al (2020) studies showed that meeting the minimum 150 minutes of exercise per week produced inconsistent results in increased mortality regardless of sedentary hours per day.

**Income:** Categorized by lower [<$28,007], lower middle [$28008-$55,000], middle [$55,001-$89,744], upper middle [$89,745-$149,131], or upper class [>$149,131]. Lower=0, Lower-Middle=1, Middle=2, Upper-Middle=3, Upper=4. Categorizations were made based on an article by Bieber C (2023).

**BMI:** Rounded to one decimal. Categorized as follows: Under 18.5 underweight, 18.5-24.9 healthy, 25-29.9 overweight, 30-39.9 obesity, 40 or over severe obesity. Stats provided by NHS Scotland (2023).

**Triglyceride:** According to National Institutes of Health (2023), Healthy levels for adults range below 150, borderline high from 150-199, and high 200-499.

**Physical Activity Days Per Week:** According to Cleveland Clinic (2023), it is recommended to split workouts over 5 days. Based on this, we will categorize 5 days per week or more as 1, and less than 5 days per week as 0.

**Sleep Hours per Night:** According to an article by the Center for Disease Control and Prevention (2022), adults should get at least 7 hours of sleep per night. We will classify enough sleep as 1, and suboptimal sleep as 0.

**Country:** Removed due to assumption that all patients are in United States.

**Continent:** Removed due to assumption that all patients are in United States.

**Hemisphere:** Removed due to assumption that all patients are in United States.

**Previous Heart Attack:** Changed from Heart Attack Risk, represented as 0 for no previous attack or 1 for any previous attack. This is the metric all others are weighed against.

**Heart Risk:** New addition, calculated as per Cleveland Clinic (2022) as follows: Low if <5%, Borderline if 5%-7.4%, Intermediate if 7.5-19.9%, or High if >=20%.

Sources:

National Institutes of Health (2022, September 2). *Age* https://www.nih.gov/nih-style-guide/age

Rosenson, Robert (2021, July 12). *Patient Education: High Cholesterol and Lipids (Beyond the Basics)* <https://www.uptodate.com/contents/high-cholesterol-and-lipids-beyond-the-basics/print>

Clinical Methods: The History, Physical, and Laboratory Examinations. 3rd edition.

Walker HK, Hall WD, Hurst JW, editors.

Boston: [Butterworths](http://www.elsevierdirect.com/imprint.jsp?iid=30); 1990.

<https://www.ncbi.nlm.nih.gov/books/NBK268/table/A562/?report=objectonly>

Laskowski, E (2022, October 8) *What’s a Normal Resting Heart Rate?* <https://www.mayoclinic.org/healthy-lifestyle/fitness/expert-answers/heart-rate/faq-20057979>

Laskowski, E (2023, July 26) *How Much Should the Average Adult Exercise Every Day?*

<https://www.mayoclinic.org/healthy-lifestyle/fitness/expert-answers/exercise/faq-20057916>

American Psychological Association (n.d) *Stress by Gender.*

<https://www.apa.org/news/press/releases/stress/2012/gender-report.pdf>

Park Et Al (2020, November 9) *Sedentary Lifestyle: Overview of Updated Evidence of Potential Health Risks*

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7700832/>

Bieber C (2023, September 5) *Does Your Income Make You Upper Class, Middle Class, or Lower Class?*

<https://www.fool.com/the-ascent/personal-finance/articles/does-your-income-make-you-upper-class-middle-class-or-lower-class/>

NHS Scotland (2023, January 4*) Body Mass Index (BMI)*

<https://www.nhsinform.scot/healthy-living/food-and-nutrition/healthy-eating-and-weight-loss/body-mass-index-bmi/>

National Institutes of Health (2023, April 19) *High Blood Triglycerides*

<https://www.nhlbi.nih.gov/health/high-blood-triglycerides>

Cleveland Clinic (2023, January 12) *How Often Should You Work Out?*

<https://health.clevelandclinic.org/how-often-you-should-work-out/>

Center for Disease Control and Prevention (2022, September 14)

<https://www.cdc.gov/sleep/about_sleep/how_much_sleep.html>

Cleveland Clinic (2022, July 1) *Cardiac Risk Calculator*

<https://my.clevelandclinic.org/health/articles/17085-heart-risk-factor-calculators>