

# Final Data Dictionary

## 1. Overview of the Data

**Dataset Name:** Heart Attack Risk Prediction Dataset

**Description:** This dataset is designed to analyze and predict heart attack risk based on various health, lifestyle, and demographic factors. It includes attributes such as age, cholesterol levels, blood pressure, smoking habits, exercise patterns, dietary preferences, and medical history. By leveraging predictive analytics and machine learning, this dataset can aid researchers and healthcare professionals in developing proactive strategies for heart disease prevention and management.

**We have two datasets for final data dictionary:**

**Data (Converted all binary variables)**

Total rows: 8763

Total columns: 26

**Data1 (Transformed “Data” categorical variables to dummy variables)**

Total rows: 8763

Total columns: 52

## 2. Variable Definitions

**Data**

Variable Name	Description	Data Type
Heart Attack Risk	Presence of heart attack risk (1: Yes, 0: No)	QL(Bin)
Age	Age of the patient (in years)	Q(Int)
Sex	Gender of the patient (Male/Female)	QL(2)
Cholesterol	Cholesterol levels of the patient (mg/dL)	Q(Int)
Heart Rate	Resting heart rate of the patient (bpm)	Q(Int)
Diabetes	Whether the patient has diabetes (Yes/No)	QL(Bin)

Family History	Family history of heart-related problems (1: Yes, 0: No)	QL(Bin)
Smoking	Smoking status of the patient (1: Smoker, 0: Non-smoker)	QL(Bin)
Obesity	Obesity status of the patient (1: Obese, 0: Not obese)	QL(Bin)
Alcohol Consumption	Level of alcohol consumption (None/Light/Moderate/Heavy)	QL(4)
Exercise Hours Per Week	Number of hours the patient exercises per week	Q(Real)
Diet	Dietary habits of the patient (Healthy/Average/Unhealthy)	QL(3)
Previous Heart Problems	Whether the patient had previous heart-related issues (1: Yes, 0: No)	QL(Bin)
Medication Use	Whether the patient takes medication (1: Yes, 0: No)	QL(Bin)
Stress Level	Self-reported stress level (Scale: 1-10)	Q(Int)
Sedentary Hours Per Day	Average number of sedentary hours per day	Q(Real)
Income	Income level of the patient	Q(Real)
BMI	Body Mass Index of the patient	Q(Real)
Triglycerides	Triglyceride levels of the patient (mg/dL)	Q(Int)
Physical Activity Days Per Week	Number of days the patient engages in physical activity per week	Q(Int)
Sleep Hours Per Day	Number of sleep hours per day	Q(Real)
Country	Country where the patient resides	QL(21)
Continent	Continent of residence	QL(7)
Northern.Hemisphere	Hemisphere of residence (Northern/Southern)	QL(Bin)
Systolic	Systolic blood pressure (mmHg)	Q(Int)
Diastolic	Diastolic blood pressure (mmHg)	Q(Int)

## Data1

Variable Name	Description	Data Type
Heart Attack Risk	Presence of heart attack risk (1: Yes, 0: No)	QL(Bin)
Age	Age of the patient (in years)	Q(Int)
Sex	Gender of the patient (Male/Female)	QL(2)
Cholesterol	Cholesterol levels of the patient (mg/dL)	Q(Int)
Heart Rate	Resting heart rate of the patient (bpm)	Q(Int)
Diabetes	Whether the patient has diabetes (Yes/No)	QL(Bin)
Family History	Family history of heart-related problems (1: Yes, 0: No)	QL(Bin)
Smoking	Smoking status of the patient (1: Smoker, 0: Non-smoker)	QL(Bin)
Obesity	Obesity status of the patient (1: Obese, 0: Not obese)	QL(Bin)
Alcohol Consumption	Level of alcohol consumption (None/Light/Moderate/Heavy)	QL(4)
Exercise Hours Per Week	Number of hours the patient exercises per week	Q(Real)
Previous Heart Problems	Whether the patient had previous heart-related issues (1: Yes, 0: No)	QL(Bin)
Medication Use	Whether the patient takes medication (1: Yes, 0: No)	QL(Bin)
Stress Level	Self-reported stress level (Scale: 1-10)	Q(Int)
Sedentary Hours Per Day	Average number of sedentary hours per day	Q(Real)
Income	Income level of the patient	Q(Real)
BMI	Body Mass Index of the patient	Q(Real)
Triglycerides	Triglyceride levels of the patient (mg/dL)	Q(Int)

Physical Activity Days Per Week	Number of days the patient engages in physical activity per week	Q(Int)
Sleep Hours Per Day	Number of sleep hours per day	Q(Real)
Northern.Hemisphere	Hemisphere of residence (Northern/Southern)	QL(Bin)
Systolic	Systolic blood pressure (mmHg)	Q(Int)
Diastolic	Diastolic blood pressure (mmHg)	Q(Int)
Diet_DietAverage	Patient follows an average diet (1: Yes, 0: No)	QL(Bin)
Diet_DietHealthy	Patient follows a healthy diet (1: Yes, 0: No)	QL(Bin)
Diet_DietUnhealthy	Patient follows an unhealthy diet (1: Yes, 0: No)	QL(Bin)
Country_CountryX (20 columns)	Dummy variables for country of residence	QL(Bin)
Continent_ContinentX (6 columns)	Dummy variables for continent of residence	QL(Bin)

**Note:** Q(Int): Quantitative (Integer);                      Q(Real): Quantitative (Real);

QL(Bin): Qualitative (Binary);                      QL(N): Qualitative (Categorical) with N levels.