Context

The data set CHDdata.csv contains cases of coronary heart disease (CHD) and variables associated with the patient's condition: systolic blood pressure, yearly tobacco use (in kg), low density lipoprotein (Idl), adiposity, family history (0 or 1), type A personality score (typea), obesity (body mass index), alcohol use, age, and the diagnosis of CHD (0 or 1).

1. **sex**: This column represents the gender of the individuals (female- male).
2. **age**: This column represents the age of the individuals in the dataset. Age is a crucial factor in assessing the risk of coronary heart disease.
3. **education**: This column represents the level of education of the individuals. It could be coded using categorical values indicating different levels of education attainment.
4. **smokingStatus**: This column likely represents the smoking status of the individuals, indicating whether they are smokers(yes), non-smokers(no).
5. **cigsPerDay**: If an individual is a smoker, this column represents the number of cigarettes smoked per day.
6. **BPMeds**: This column indicates whether the individual is taking blood pressure medications (binary: 0 for not taking, 1 for taking).
7. **prevalentStroke**: This column indicates whether an individual has had a stroke prior to the study (binary: 0 for no, 1 for yes).
8. **prevalentHyp**: This column indicates whether an individual has hypertension (binary: 0 for no, 1 for yes).
9. **diabetes**: This column indicates whether an individual has diabetes (binary: 0 for no, 1 for yes).
10. **totChol**: This column represents the total cholesterol level of the individuals.
11. **sysBP**: This column represents the systolic blood pressure of the individuals.
12. **diaBP**: This column represents the diastolic blood pressure of the individuals.
13. **BMI**: This column represents the Body Mass Index (BMI) of the individuals, which is a measure of body fat based on height and weight.
14. **heartRate**: This column represents the resting heart rate of the individuals.
15. **glucose**: This column represents the fasting blood glucose level of the individuals.
16. **CHDRisk**: This column likely represents the Ten-Year Coronary Heart Disease (CHD) Risk for each individual, which is the target variable that you may want to predict or analyze.