Data Processes second assignment (second part)

Assignment details

Heart attacks are among the most prevalent chronic diseases in the world, impacting millions of people each year and exerting a significant financial burden on the economy. The buildup of plaques inside larger coronary arteries, molecular changes associated with aging, chronic inflammation, high blood pressure, and diabetes are all causes of and risk factors for heart disease.

A national healthcare system has done a telephone survey asking people some life habits and health conditions. The results of the survey can be found in the dataset whose features are described in the annex. The goal is to reduce the costs associated to heart attacks in the national healthcare system by a 20%.

We know that we can reduce the risk of suffering a heart attack in a person who is prone to it by a 75% if we settle up a specific lifestyle modification plan for that person. The cost of the treating a person who has suffered a heart attack is of €50,000 in average, while the customized plan for a person costs €1,000. The estimation is that 85% of the people who is offered the plan would accept it, but not all of them will adhere to the plan properly.

Then, you need to do a cost analysis to check the percentage of adherence (from those people who has already accepted the prevention plan) we need to obtain to achieve our goal.

A **ZIP file** that contains a PDF file with the explanations about the treatment given to the dataset and the cost analysis and the Python file code used to make the analysis must be uploaded to the Moodle task for the assignment. The deadline to upload this assignment is **January 10th 2023, 23:55**.

The assignment must be done in **groups of 4 people**.

Annex

The variables included in the dataset “heart_disease.xlsx” are the following:

- HeartAttack: binary variable indicating whether the patient has suffered a heart attack or not.
- HighBP: binary variable indicating whether the patient has high blood pressure or not.
- HighChol: binary variable indicating whether the patient has cholesterol levels in blood higher than the standards or not.
- CholCheck: binary variable indicating whether the patient controls their level of cholesterol in blood periodically or not.
- BMI: body mass index of the patient.
- Smoker: binary variable indicating whether the patient smokes or not.
- Stroke: HeartAttack: binary variable indicating whether the patient has previously suffered a brain stroke or not.
- Diabetes: binary variable indicating whether the patient has diabetes or not.
- PhysActivity: binary variable indicating whether the patient does physical activity regularly or not.
- Fruits: binary variable indicating whether the patient eats fruit every day or not.
- Veggies: binary variable indicating whether the patient eats vegetables every day or not.
- HvyAlcoholConsump: binary variable indicating whether the patient drinks alcohol regularly or not.
- GenHlth: grade given by the patient to their general health status from 1 to 5.
- MentHlth: grade given by the patient to their mental health status from 1 to 5.
- PhysHlth: grade given by the patient to their physical health status from 1 to 5.
- DiffWalk: binary variable indicating whether the patient has difficulty to walk or not.
- Sex: sex of the patient.
- Age: age of the patient.
- Education: grade given to the educational level of the patient from 1 to 6.
- Income: grade given to the annual income of the patient from 1 to 8.