Data of Patients (For Medical Field)[1]

Dữ liệu cho lĩnh vực y tế (Dữ liệu đã được làm sạch)

A. Giới thiệu về tập dữ liệu

Bộ dữ liệu bao gồm thông tin bệnh nhân, bao gồm các chi tiết nhân khẩu học như tuổi, giới tính và trạng thái, cùng với các yếu tố liên quan đến sức khỏe như sức khỏe tổng quát, BMI và tiền sử bệnh. Nó bao gồm các tình trạng như bệnh tim, đột quỵ, tiểu đường và hen suyễn, cũng như các lựa chọn về lối sống như hút thuốc, uống rượu và tình trạng tiêm chủng.

B. Thông tin về bệnh nhân

1.- *PatientID* : (int) Mã định danh duy nhất cho mỗi bệnh nhân.

2.- *State* : (string) Bang địa lý nơi cư trú. Gồm 54 bang.

3.- *Sex* : (string) Giới tính của bệnh nhân. Gồm 2 phái Male, Female.

4.- *GeneralHealth* : (string) Tự báo cáo tình trạng sức khỏe. Gồm 6 giá trị: Excellent, Very good, Good, Fair, Poor

5.- *AgeCategory* : (string) Nhóm tuổi được phân loại của bệnh nhân. Gồm 13 nhóm, khoảng tuổi của mỗi nhóm là 5. Riêng nhóm đầu là 7 (từ 18-24) và nhóm cuối (từ 80 trở lên)

6.- *HeightInMeters* : (float) Chiều cao của bệnh nhân (tính bằng mét).

7.- *WeightInKilograms*: (float) Cân nặng của bệnh nhân (tính bằng kilogam).

8.- *BMI* : (float) Chỉ số khối cơ thể, được tính từ chiều cao và cân nặng.

9.- *RaceEthnicityCategory*: (string) Chủng tộc hoặc sắc tộc của bệnh nhân. Gồm 5 giá trị: White only, Non-Hispanic; Black only, Non-Hispanic; Other race only, Non-Hispanic; Multiracial, Non-Hispanic; Hispanic

C. Tiền sử bệnh

10.- *HadHeartAttack* : (boolean: 0 và 1) Dấu hiệu cho biết bệnh nhân có bị đau tim hay không.

11.- *HadAngina* : (boolean: 0 và 1) Chỉ số cho biết bệnh nhân có bị đau thắt ngực hay không.

12.- *HadStroke* : (boolean: 0 và 1) Dấu hiệu cho biết bệnh nhân có bị đột quỵ hay không.

13.- *HadAsthma* : (boolean: 0 và 1) Dấu hiệu cho biết bệnh nhân có bị hen suyễn hay không.

14.- *HadSkinCancer* : (boolean: 0 và 1) Chỉ số cho biết bệnh nhân có bị ung thư da hay không.

15.- *HadCOPD* : (boolean: 0 và 1) Chỉ số cho biết bệnh nhân có mắc bệnh phổi tắc nghẽn mạn tính (COPD) hay không.

[1] Data of Patients (For Medical Field)

16.- *HadDepressiveDisorder*: (boolean: 0 và 1) Chỉ báo cho biết bệnh nhân có được chẩn đoán mắc chứng rối loạn trầm cảm hay không.

17.- *HadKidneyDisease* : (boolean: 0 và 1) Chỉ số cho biết bệnh nhân có bệnh thận hay không.

18.- *HadArthritis* : (boolean: 0 và 1) Chỉ số cho biết bệnh nhân có bị viêm khớp hay không.

19.- *HadDiabetes* : (string) Chỉ số cho biết bệnh nhân có mắc bệnh tiểu đường hay không. Gồm 4 giá trị: Yes; No; No, pre-diabetes or borderline diabetes; Yes, but only during pregnancy (female)

20.- *DeafOrHardOfHearing* : (boolean: 0 và 1) Biểu hiện tình trạng suy giảm thính lực.

21.- *BlindOrVisionDifficty*: (boolean: 0 và 1) Dấu hiệu suy giảm thị lực.

22.- *DifficultyConcentrating* : (boolean: 0 và 1) Chỉ số khó tập trung.

23.- *DifficultyWalking* : (boolean: 0 và 1) Chỉ số về những khó khăn trong việc đi lại.

24.- *DifficultyDressingBathing*: (boolean: 0 và 1) Biểu thị sự khó khăn trong việc mặc quần áo hoặc tắm rửa.

25.- *DifficultyErrands* : (boolean: 0 và 1) Biểu thị mức độ khó khăn khi thực hiện các công việc vặt.

D. Lối sống

26.- *SmokerStatus* : (string) Tình trạng bệnh nhân có hút thuốc hay không. Gồm 4 giá trị: Former smoker; Never smoked; Current smoker - now smokes every day; Current smoker - now smokes some days

27.- *EcigaretteUsage* : (string) Chỉ số sử dụng thuốc lá điện tử. Gồm 4 giá trị: Never used e-cigarettes in my entire life; Not at all (right now); Use them some days; Use them every day

28.- *AlcoholDrinkers* : (boolean: 0 và 1) Tình trạng bệnh nhân có uống rượu hay không.

E. Các xét nghiệm đã thực hiện

29.- *ChestScan* : (boolean: 0 và 1) Chỉ báo bệnh nhân đã được chụp ngực hay chưa.

30.- *HIVTesting* : (boolean: 0 và 1) Tình trạng bệnh nhân đã được xét nghiệm HIV hay chưa.

31.- *FluVaxLast12* : (boolean: 0 và 1) Tình trạng bệnh nhân có được tiêm vắc xin cúm trong 12 tháng qua hay không.

32.- *PneumoVaxEver* : (boolean: 0 và 1) Tình trạng bệnh nhân đã từng được chủng ngừa phế cầu khuẩn hay chưa.

33.- *TetanusLast10Tdap*: (string) Tình trạng bệnh nhân đã được tiêm vắc xin uốn ván trong 10 năm qua hay chưa. Gồm 4 giá trị: No, did not receive any tetanus shot in

the past 10 years; Yes, received Tdap; Yes, received tetanus shot but not sure what type; Yes, received tetanus shot, but not Tdap

34.- *CovidPos* : (boolean: 0 và 1) Tình trạng bệnh nhân có kết quả xét nghiệm dương tính với COVID-19 hay không.

F. Chỉ báo sức khỏe của năm trước

35.- *HighRiskLastYear* : (boolean: 0 và 1) Chỉ báo cho biết bệnh nhân có nguy cơ cao trong năm qua hay không.

G. Các yêu cầu thực hiện đối với bài tập

Trước khi thực hiện các yêu cầu cầu sau, SV cần đọc qua tất cả các yêu cầu. Như vậy, nếu thực hiện xong các yêu cầu sau đây, số lượng bảng phụ cần tạo ra là quá nhiều. SV hãy đề xuất và thực hiện gom gọn các table lại với nhau ngay khi thực hiện các yêu cầu dưới đây:

(1).- Phát sinh mã kiểu số nguyên (char hoặc tiny int) cho các bang để dữ liệu của thuộc tính State trong dữ liệu gốc trở thành kiểu số nguyên => cần lập bảng lưu lại kết quả mã để tiện dùng sau này.

(2).- Chuyển field "*Sex*" từ kiểu string thành kiểu bool với tên field mới là "*isFemale*". Do đó giá trị "*Female*" sẽ được chuyển thành *true* và "*male*" được chuyển thành *false*.

(3).- Chuyển field "*GeneralHealth*" từ kiểu string thành kiểu số nguyên (char hoặc tiny int) với giá trị từ 1 đến 5 (Excellent=1, Very good=2, Good=3, Fair=4, Poor=5). Do đó thuộc tính này trong dữ liệu gốc trở thành kiểu số nguyên => cần lập bảng lưu lại kết quả mã để tiện dùng sau này.

(4).- Chuyển field "*AgeCategory*" từ kiểu string thành kiểu số nguyên (char hoặc tiny int) với giá trị từ 1 đến 13. Do đó thuộc tính này trong dữ liệu gốc trở thành kiểu số nguyên => cần lập bảng lưu lại kết quả mã để tiện dùng sau này.

(5).- Do field BMI được tính từ 2 field *WeightInKilograms* và *HeightInMeters* (theo công thức BMI= *WeightInKilograms* / *HeightInMeters* ^{2}) nên cần xóa bỏ field này.

(6).- Chuyển field "*RaceEthnicityCategory*" từ kiểu string thành kiểu số nguyên (char hoặc tiny int) với giá trị từ 1 đến 5 (White only, Non-Hispanic=1; Black only, Non-Hispanic=2; Other race only, Non-Hispanic=3; Multiracial, Non-Hispanic=4; Hispanic=5). Do đó thuộc tính này trong dữ liệu gốc trở thành kiểu số nguyên => cần lập bảng lưu lại kết quả mã để tiện dùng sau này.

(7).- Chuyển field "*HadDiabetes*" từ kiểu string thành kiểu số nguyên (char hoặc tiny int) với giá trị từ 1 đến 4 (Yes=1; No=2; No, pre-diabetes or borderline diabetes=3; Yes, but only during pregnancy (female)=4). Do đó thuộc tính này trong dữ liệu gốc trở thành kiểu số nguyên => cần lập bảng lưu lại kết quả mã để tiện dùng sau này.

(8).- Nhóm các field kiểu bool thuộc mục "*Tiền sử bệnh*" (15 fields: từ field thứ 10 - *HadHeartAttack* - đến field thứ 25 - *DifficultyErrands* -, bỏ qua field *HadDiabetes*) thành 1 field duy nhất. Với tên field mới là "*Medical history*" => cần lập bảng lưu lại kết quả mã để tiện dùng sau này.

(9).- Nhóm các field kiểu bool thuộc mục "*Các xét nghiệm đã thực hiện*" (5 fields) thành 1 field duy nhất. Với tên field mới là "*Tests performed*" => cần lập bảng lưu lại kết quả mã để tiện dùng sau này.

(10).- Chuyển field "*SmokerStatus*" từ kiểu string thành kiểu số nguyên (char hoặc tiny int) với giá trị từ 1 đến 4 (Former smoker=1; Never smoked=2; Current smoker - now smokes every day=3; Current smoker - now smokes some days=4). Do đó thuộc tính

này trong dữ liệu gốc trở thành kiểu số nguyên => cần lập bảng lưu lại kết quả mã để tiện dùng sau này.

(11).- Chuyển field "*EcigaretteUsage*" từ kiểu string thành kiểu số nguyên (char hoặc tiny int) với giá trị từ 1 đến 4 (Never used e-cigarettes in my entire life=1; Not at all (right now)=2; Use them some days=3; Use them every day=4). Do đó thuộc tính này trong dữ liệu gốc trở thành kiểu số nguyên => cần lập bảng lưu lại kết quả mã để tiện dùng sau này.

(12).- Chuyển field "*TetanusLast10Tdap*" từ kiểu string thành kiểu số nguyên (char hoặc tiny int) với giá trị từ 1 đến 4 (No, did not receive any tetanus shot in the past 10 years=1; Yes, received Tdap=2; Yes, received tetanus shot but not sure what type=3; Yes, received tetanus shot, but not Tdap=4). Do đó thuộc tính này trong dữ liệu gốc trở thành kiểu số nguyên => cần lập bảng lưu lại kết quả mã để tiện dùng sau này.

- HẾT -