**Contexte**

Les maladies cardiaques font partie des maladies chroniques les plus répandues aux États-Unis. Elles touchent des millions d'Américains chaque année et représentent une charge financière importante pour l'économie. Rien qu'aux États-Unis, les maladies cardiaques font environ 647 000 victimes chaque année, ce qui en fait la première cause de décès. L'accumulation de plaques à l'intérieur des grosses artères coronaires, les changements moléculaires associés au vieillissement, l'inflammation chronique, l'hypertension artérielle et le diabète sont autant de causes et de facteurs de risque des maladies cardiaques.

Bien qu'il existe différents types de maladies coronariennes, la majorité des personnes n'apprennent qu'elles sont atteintes de la maladie qu'à la suite de symptômes tels que des douleurs thoraciques, une crise cardiaque ou un arrêt cardiaque soudain. Ce fait souligne l'importance des mesures préventives et des tests qui peuvent prédire avec précision les maladies cardiaques dans la population avant que des résultats négatifs comme les infarctus du myocarde (crises cardiaques) ne se produisent.

Les centres de contrôle et de prévention des maladies ont identifié l'hypertension artérielle, l'hypercholestérolémie et le tabagisme comme les trois principaux facteurs de risque des maladies cardiaques. Près de la moitié des Américains présentent au moins un de ces trois facteurs de risque. L'Institut national du cœur, des poumons et du sang (National Heart, Lung, and Blood Institute) met en évidence un plus large éventail de facteurs tels que l'âge, l'environnement et la profession, les antécédents familiaux et la génétique, les habitudes de vie, les autres conditions médicales, la race ou l'origine ethnique et le sexe, que les cliniciens peuvent utiliser pour diagnostiquer une maladie coronarienne. Le diagnostic tend à être guidé par une enquête initiale sur ces facteurs de risque communs, suivie d'analyses de sang et d'autres tests.

**Contenu**

Le Behavioral Risk Factor Surveillance System (BRFSS) est une enquête téléphonique sur la santé qui est menée chaque année par les CDC. Chaque année, l'enquête recueille les réponses de plus de 400 000 Américains sur les comportements à risque liés à la santé, les conditions de santé chroniques et l'utilisation de services préventifs. Elle a été menée chaque année depuis 1984. Pour ce projet, le jeu de données en fichier csv concerne l’enquête menée pour l'année 2015. Ce jeu de données original contient les réponses de 441 455 personnes et comporte 330 caractéristiques. Ces caractéristiques sont soit des questions directement posées aux participants, soit des variables calculées à partir des réponses individuelles des participants.

Cet ensemble de données contient 253 680 réponses à l'enquête BRFSS 2015 nettoyée, à utiliser principalement pour la classification binaire des maladies cardiaques. Il n'y a pas de fort déséquilibre entre les classes dans cet ensemble de données. 229 787 répondants n'ont pas eu de maladie cardiaque tandis que 23 893 ont eu une maladie cardiaque. Les questions à explorer sont les suivantes :

1. Dans quelle mesure les réponses à l'enquête du BRFSS peuvent-elles être utilisées pour prédire le risque de maladie cardiaque ?

et

2. Un sous-ensemble de questions de l'enquête BRFSS peut-il être utilisé pour le dépistage préventif de maladies comme les maladies cardiaques ?

**Description des features**

1. HeartDiseaseorAttack

Répondants ayant déjà déclaré avoir souffert d'une maladie coronarienne ou d'un infarctus du myocarde.

2. HighBP

Adultes à qui un médecin, une infirmière ou un autre professionnel de la santé a dit qu'ils souffraient d'hypertension.

3. HighChol

Un médecin, une infirmière ou un autre professionnel de la santé vous a-t-il JAMAIS dit que votre taux de cholestérol sanguin était élevé?

4. CholCheck

Contrôle du cholestérol au cours des cinq dernières années.

5. BMI

Indice de masse corporelle.

6. Smoker

Avez-vous fumé au moins 100 cigarettes au cours de votre vie ? (Note : 5 paquets = 100 cigarettes).

7. Stroke

(On vous a déjà dit) que vous aviez eu une attaque vasculaire cérébrale.

8. Diabetes

0 = pas de diabète, 1 = pré-diabète et 2 = diabète.

9. PhysActivity

Adultes ayant déclaré avoir pratiqué une activité physique ou un exercice au cours des 30 derniers jours en dehors de leur travail habituel.

10. Fruits

Consommez 1 ou plusiuers fruits 1 fois par jour.

11. Veggies

Consommer 1 ou plusieurs légumes 1 fois par jour.

12. HvyAlcoholConsump

Gros buveurs (hommes adultes ayant plus de 14 boissons par semaine et femmes adultes ayant plus de 7 boissons par semaine).

13. AnyHealthcare

Vous bénéficiez d'une quelconque couverture de soins de santé, y compris une assurance maladie, des plans prépayés?

14. NoDocbcCost

Au cours des 12 derniers mois, y a-t-il eu un moment où vous avez eu besoin de consulter un médecin mais n'avez pas pu le faire à cause du coût?

15. GenHlth

Diriez-vous qu'en général votre santé est :

16. MenHlth

Si vous pensez maintenant à votre santé mentale, qui comprend le stress, la dépression et les problèmes émotionnels, pendant combien de jours au cours des 30 derniers jours votre santé mentale n'était-elle pas bonne ?

17. PhsyHlth

En pensant maintenant à votre santé physique, qui comprend les maladies et les blessures physiques, pendant combien de jours au cours des 30 derniers jours votre santé physique n'était-elle pas bonne?

18. DiffWalk

Avez-vous de sérieuses difficultés à marcher ou à monter des escaliers?

19. Sex

Indiquez le sexe du patient.

20. Age

14 catégories d'âges.

21. Education

Quelle est la plus haute classe ou année d'études que vous avez achevée?

22. Income

Quel est le revenu annuel de votre famille, toutes sources confondues? (Si on refuse de répondre, codez "Refused").