

# Rendu Machine learning

[Lien du Repo](#) (branche "test\_branch")

## Comment on a abordé le problème

Tout d'abord, nous avons fait un bilan des compétences de chacun. Ainsi, nous avons pu découvrir les affinités de chacun pour travailler efficacement.

Ensuite, nous avons analysé ensemble le jeu de données pour définir les bases et pour s'assurer qu'on parte tous sur la même longueur d'onde.

Enfin, nous avons chacun de notre côté essayé de manipuler le jeu de données avec Jupyter, Pandas, etc... car même si nous savions ce qu'on voulait réaliser, nous avons besoin de pratiquer pour au moins avoir les bases.

## Exploration du dataset

Forme : 319795 lignes et 18 colonnes

```
df.shape
✓ 0.4s
(319795, 18)
```

Types des différentes colonnes (float64 ou object)

Deux types de données :

- Variables de catégories (14)

	HeartDisease	Smoking	AlcoholDrinking	Stroke	DiffWalking	Sex	AgeCategory	Race	Diabetic	PhysicalActivity	GenHealth	Asthma	KidneyDisease	SkinCancer
0	No	Yes	No	No	No	Female	55-59	White	Yes	Yes	Very good	Yes	No	Yes
1	No	No	No	Yes	No	Female	80 or older	White	No	Yes	Very good	No	No	No

- Variables discrètes (4)

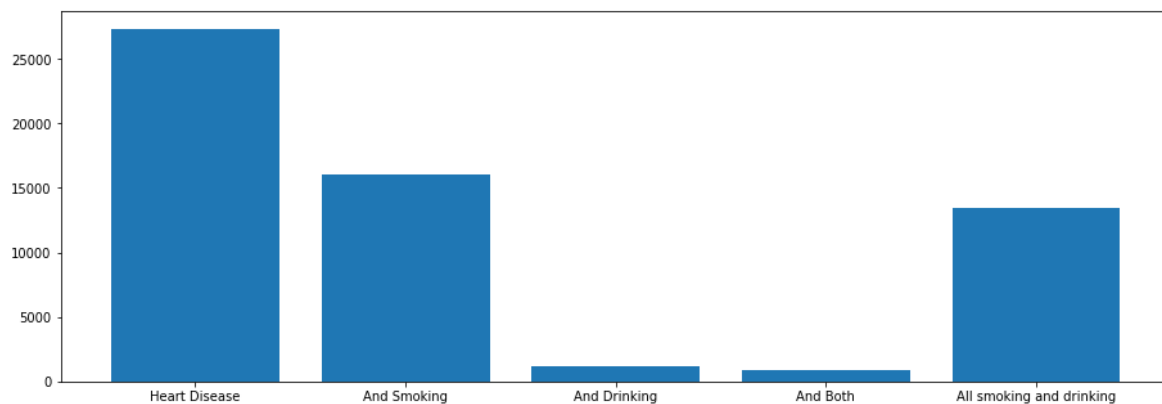
	BMI	PhysicalHealth	MentalHealth	SleepTime
0	16.60	3.0	30.0	5.0
1	20.34	0.0	0.0	7.0

Aucune valeur nulle.

Remarques supplémentaires :

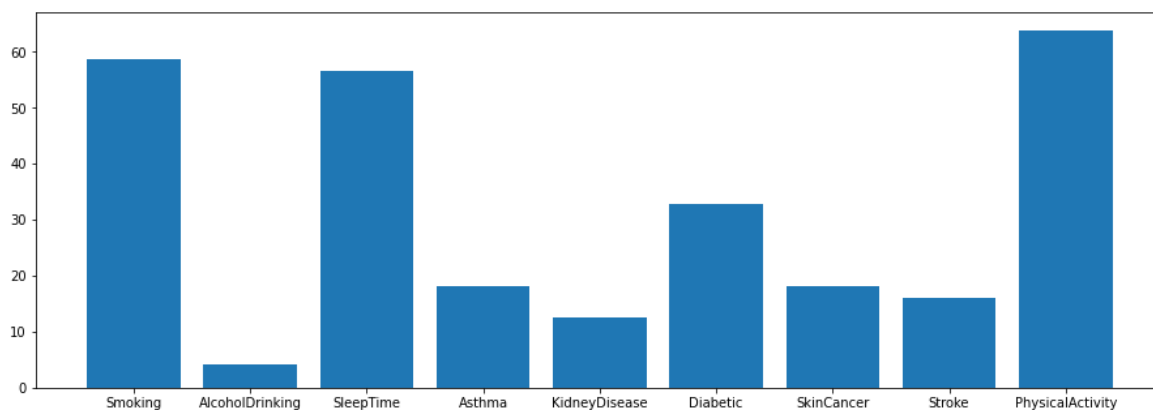
- 27373 personnes sur les 319795 qui souffrent de maladies cardiaques.
- Sur ces 27373 personnes, 16037 d'entre elles fument, 1141 boivent de l'alcool, 890 fument et boivent
- Aussi, sur les 27373 personnes, 15497 ont un temps de sommeil inférieur aux 8 heures recommandées

Heart Disease : 27373  
 And Smoking : 16037  
 And Drinking : 1141  
 And Both : 890  
 All Smoking and Drinking : 13415



Etude de chaque colonne (sur les personnes qui souffrent de maladies cardiaques) :

- la majorité des sujets sont entre 23 et 33 BMI
- 58.5 % des sujets fument, 41.5% ne fument pas
- 4.1% de sujets boivent de l'alcool
- 56.6% de sleeptime < 8 heures
- 18% qui font de l'asthme
- 12% qui ont une maladie des reins
- 32% qui ont le diabète
- 18% qui ont un cancer de la peau
- 16% qui ont déjà eu une crise cardiaque
- 63.9% font une activité physique régulière



Smoking : 58.586928725386336  
 AlcoholDrinking : 4.168341066013955  
 SleepTime : 56.614181858035295  
 Asthma : 18.021407956745698  
 KidneyDisease : 12.621926716107112  
 Diabetic : 32.722025353450476  
 SkinCancer : 18.19310999890403  
 Stroke : 16.034048149636504  
 PhysicalActivity : 63.89142585759691

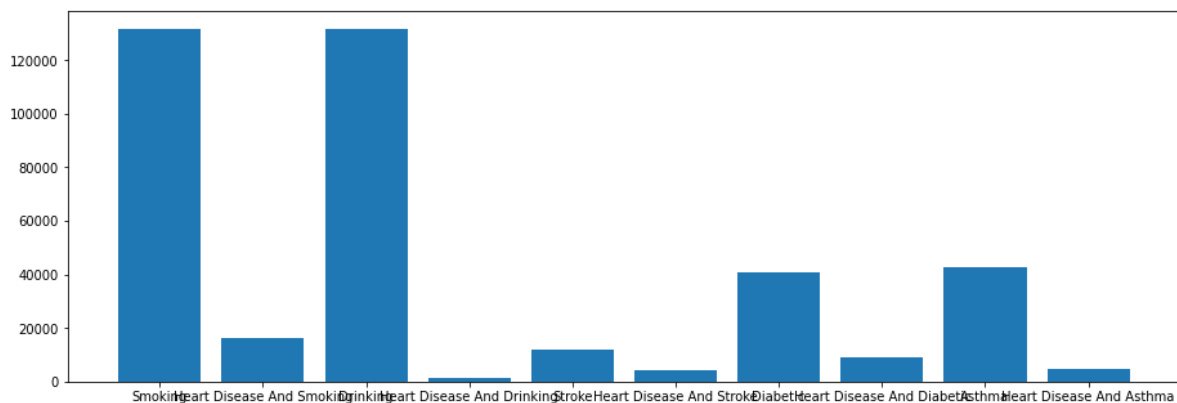
## Combinaisons de données

Après l'étude initiale, on essaie de comparer les sujets qui sont malades et ceux qui ne le sont pas à travers plusieurs différentes caractéristiques.

On commence par récupérer le nombre total de personnes qui ont "True" sur la caractéristique (par exemple Smoking) et on récupère le nombre de personnes qui souffrent de maladies cardiaques et de la caractéristique choisie aussi.

On compare les deux valeurs qu'on trouve :

```
Smoking : 131908
Heart Disease And Smoking : 16037, 12.157715983867543
Drinking : 131908
Heart Disease And Drinking : 1141, 0.8649968159626407
Stroke : 12069
Heart Disease And Stroke : 4389, 36.36589609743972
Diabetic : 40802
Heart Disease And Diabetic : 8957, 21.952355276702125
Asthma : 42872
Heart Disease And Asthma : 4933, 11.506344467251353
```



## Théories qu'on peut émettre

En observant les données qu'on a,

On peut voir que 36% des personnes qui souffrent de maladies cardiaques ont subi une crise cardiaque.

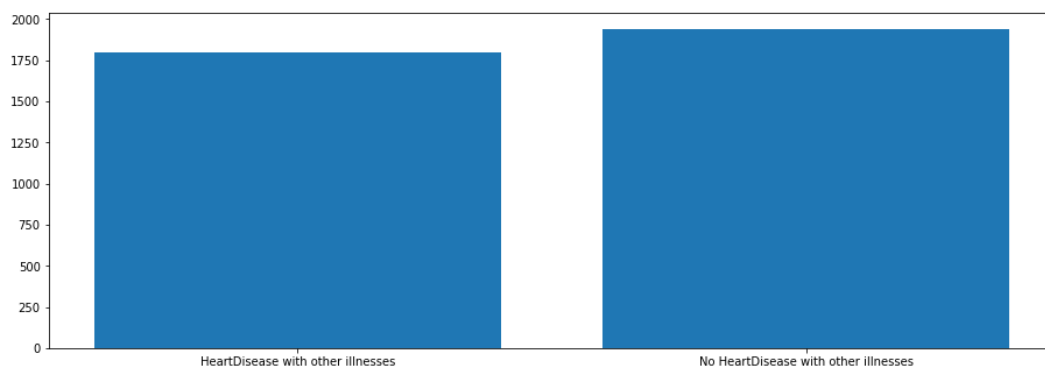
On peut aussi voir que 21% des personnes qui souffrent de maladies cardiaques sont diabétiques.

On peut alors combiner les deux pour voir combien de personnes qui sont diabétiques et qui ont subi une crise cardiaque souffrent d'une maladie du coeur.

HeartDisease with other illnesses: 1801

No HeartDisease with other illnesses: 1943

48.10363247863248% of the people that are diabetic and suffered a stroke have a heart disease.



Résultat: 48% sont malades d'une maladie du cœur.