

**Maladies
Cardio-
vasculaire**

ANALYSE DE DONNÉES

PAR : ELYES KACEM

1 ING 1

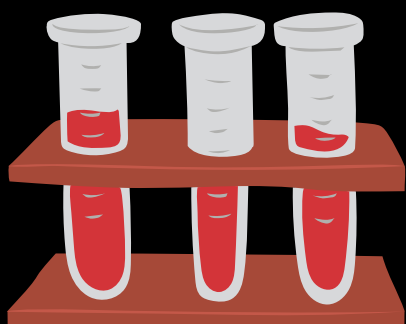
SOMMAIRE

DESCRIPTION DU PROJET

- 01** Titre
- 02** Lien Data set
- 03** Description des colonnes de data set
- 04** Objectifs et Questions de l'analyse

TRAVAIL RÉALISÉ

- 01** Description des méthodes d'analyse et d'apprentissage adoptées
- 02** bibliothèques Python importées
- 03** Les modules implémentés
- 04** Les résultats obtenus



DESCRIPTION DU PROJET

01

Titre : Analyse sur les maladies cardiovasculaire d'un échantillon de 69k personnes.

02

Lien du projet :

<https://colab.research.google.com/drive/10sJ9JXVQk2R4JePDILWMK8pzjF4PidPK?usp=sharing>

Lien data set :

<https://www.kaggle.com/sulianova/cardio-vascular-disease-dataset>

03

Description des colonnes de data set :

- **id** : l'identifiant du personne.
- **age** : l'age du personne
- **gender** : genre : 1 homme / 2: femme

- **height** : la taille du personne en CM
- **weight** : le poid du personne en KG
- **ap_hi** : la tension systolique : la tension artérielle systolique mesure la pression maximale exercée dans la phase de contraction du ventricule gauche.
- **ap_lo** : La tension artérielle diastolique : indique la pression résiduelle au moment de la phase de relâchement du cœur.
- **cholesterol** : 1 : taux moyenne, 2: un peu plus loin du normal, 3 : anormale
- **gluc** : 1 : taux moyenne, 2: un peu plus loin du normal, 3 : anormale
- **smoke** : 1: fumeur, 0 : non fumeur
- **alco** : 1: alchooliste, 0: non alchooliste
- **active** : 1 : active, 0 : non active
- **cardio** : 1 : malade, 0: non malade

04

Objectifs et Questions de l'analyse :

Préparer les données (elle était dans une mauvaise format), faire un model et choisir un model plus précis (J'ai utilisé la méthode Tree classification) et enregistrer les données bien organisées et le model dans drive, faire des étude sur l'influence de l'alchool, l'absence de l'activité , le poid, la taille , l'age et le genre sur la maladie.

TRAVAIL RÉALISÉ

01

Description des méthodes d'analyse et d'apprentissage adoptées :

Pour faire le model d'apprentissage, j'ai utilisé la méthode de Tree classification pour préviser si le personne est malade (1) ou non (0), et j'ai subdiviser les données en 2 groupes : données d'apprentissage (80%) et données de teste (20%). Cette méthode permet de faire une arbre de choix, en basant sur les données d'apprentissage. Mais ce type d'apprentissage fonctionne seulement sur les données réelles ou entier. Or j'ai des données de type chaine de caractère, donc j'ai essayé de coder ces données par un clé de type réelle. Après la construction de l'arbre, j'ai tester la précision de ce model en prévisionant les données de test et comparer les résultat obtenu par les valeurs réelles.

Pour l'analyse, j'ai essayer d'organiser les données dans un fichier CSV (car les données était sous une forme indéterminée), puis, j'ai éliminer les lignes conenants les valeurs NaN, et j'ai sauvegarder le nouveau fichier dans le drive.

Durant l'analyse, j'ai fais des graphes sur des différents informations en utilisant des différents bibliothèque et modules python que vous le trouverez dans les parties suivants.

02

Les modules implémentés :

- seaborn, plot
- matplotlib.pyplot
- pandas
- sklearn.tree
- sklearn.model_selection
- sklearn.metrics
- joblib
- google.colab
- accuracy_score

03

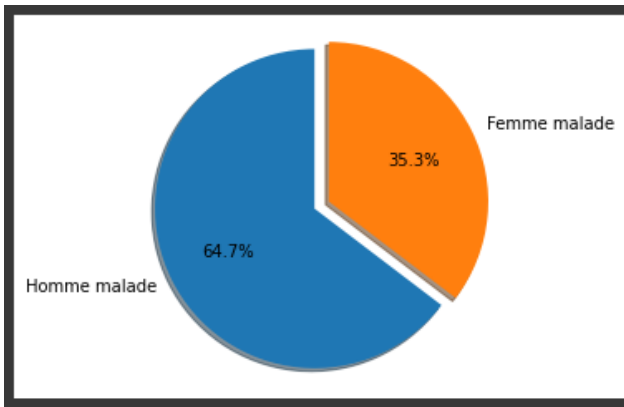
Les bibliothèque python importés :

- **drive** : pour l'interaction avec mon drive.
- **csv** : pour construire un fichier csv (contenant les données formater)
- **DecisionTreeClassifier** : pour construire le model d'apprentissage
- **train_test_split** : pour séparer les données d'apprentissage et les données du teste

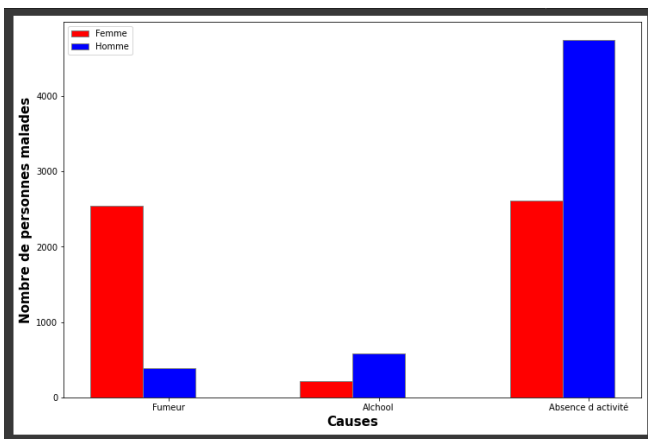
04

Les résultats obtenus :

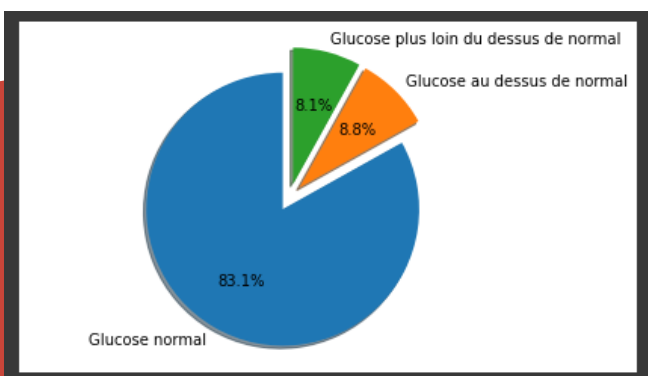
la pourcentage des malades chez les personnes selon le genre:



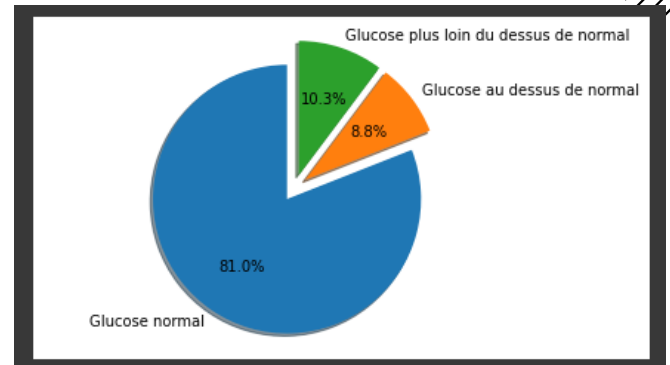
Nombre de personnes malades en fonction des causes, pour chaque genre :



Taux du glucose pour les femmes malades:

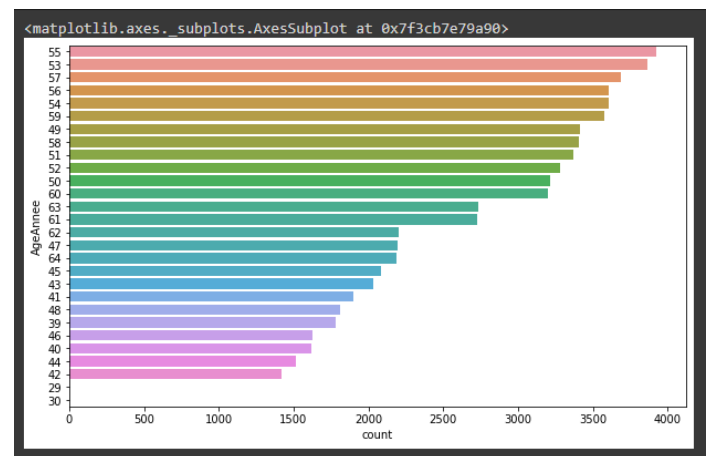


Taux du glucose pour les hommes malades:

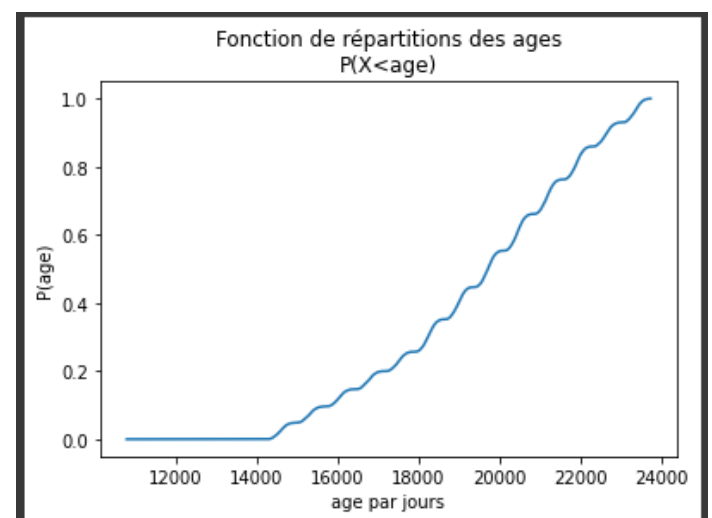


Concernant l'age des malades :

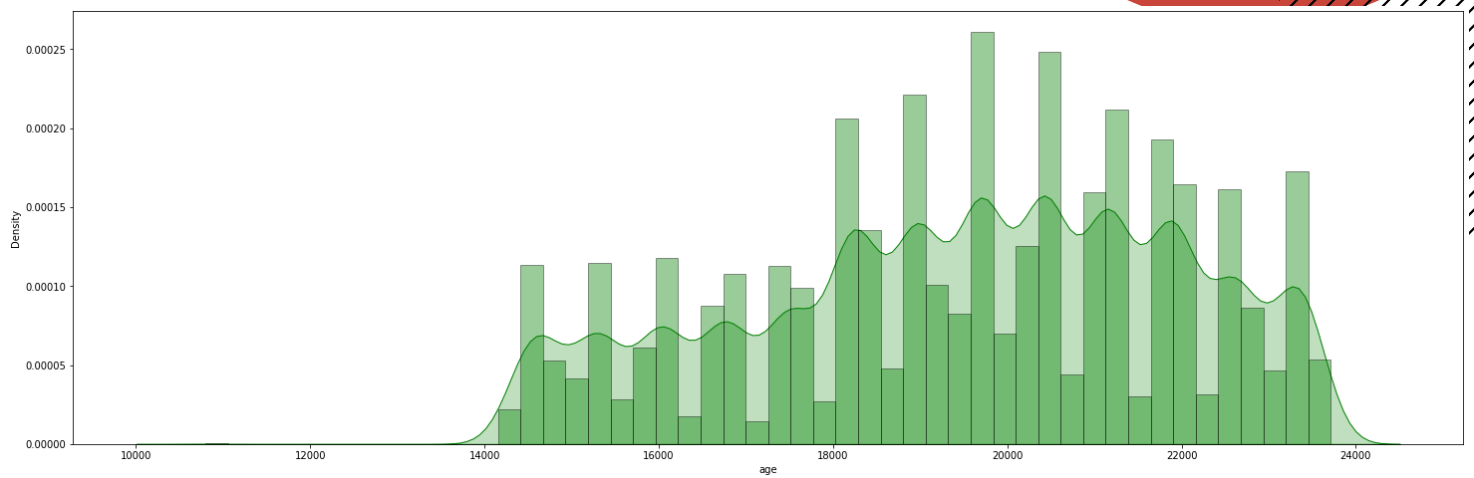
Nombre de personnes malade par age :



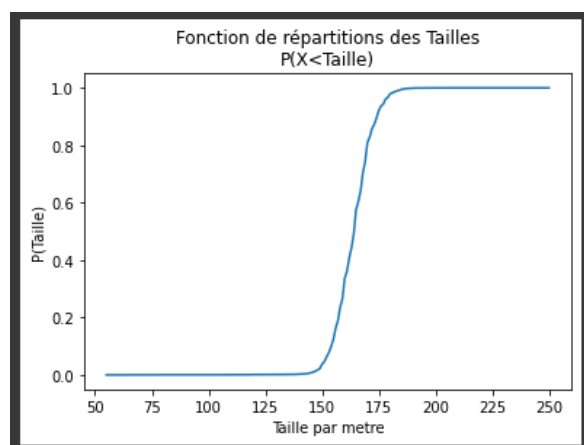
Fonction de répartition des ages :



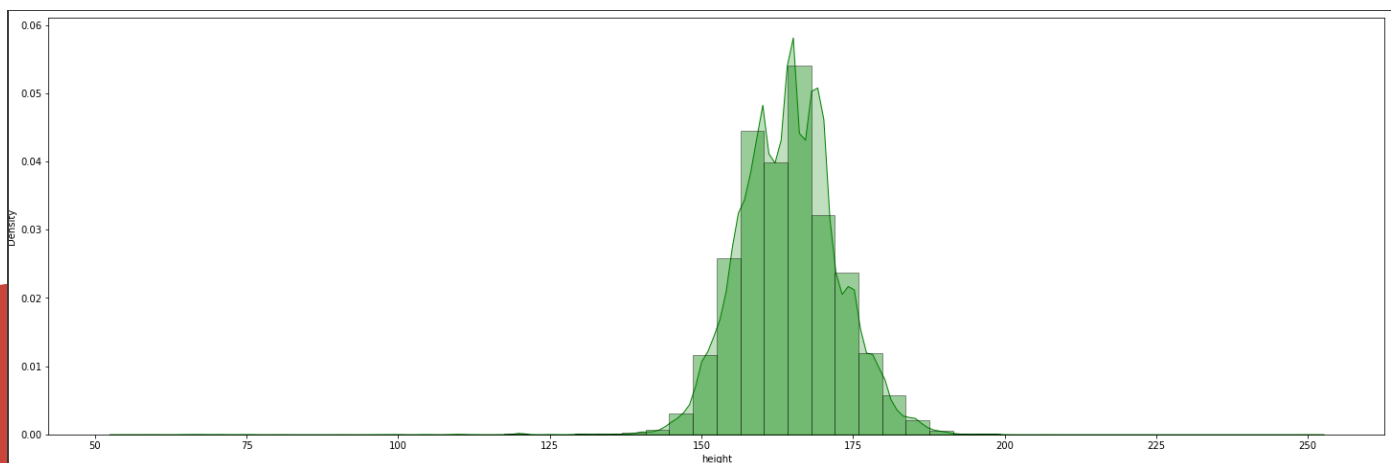
la densité en fonction d'age :



Concernant la taille des malades :

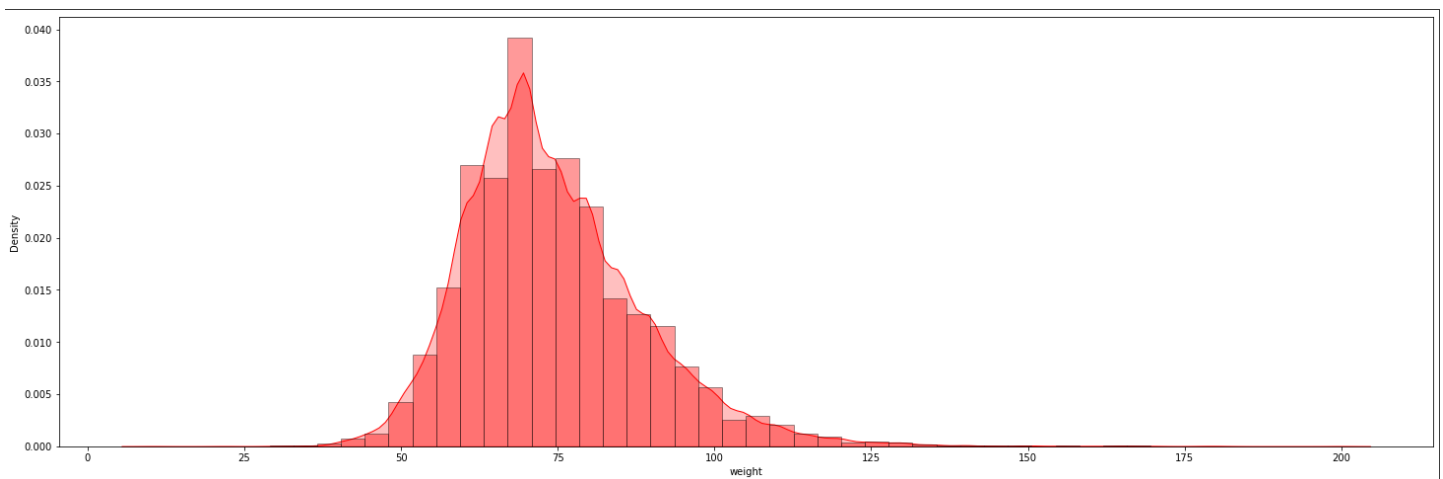


la densité en fonction de la taille en cm :

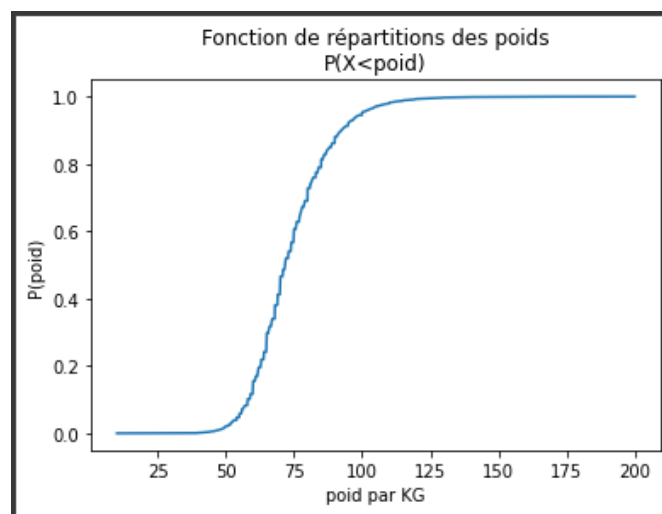


Concernant le poid des malades :

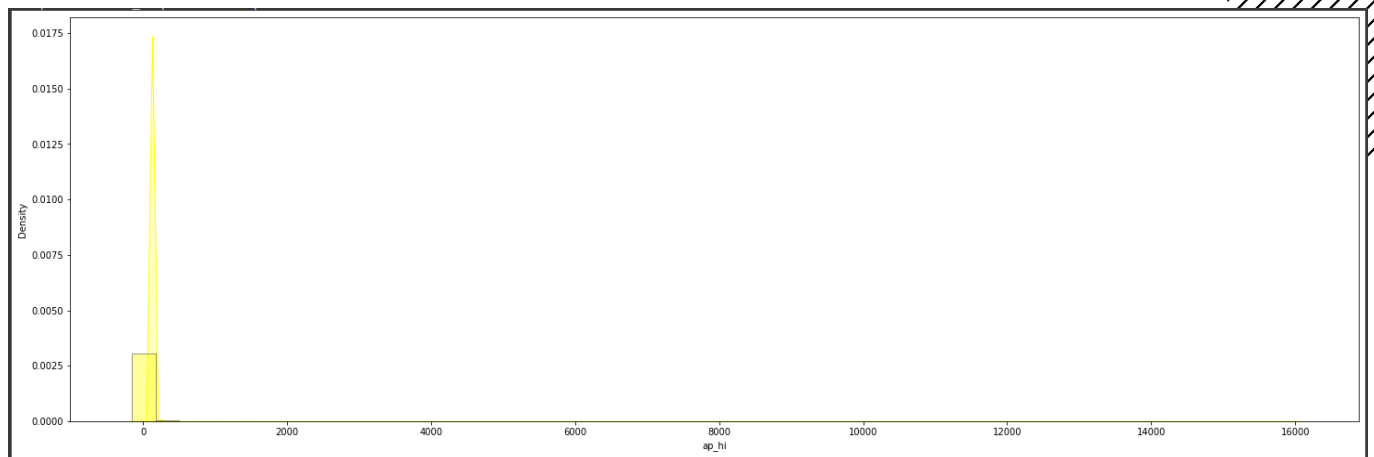
la densité en fonction du poid :



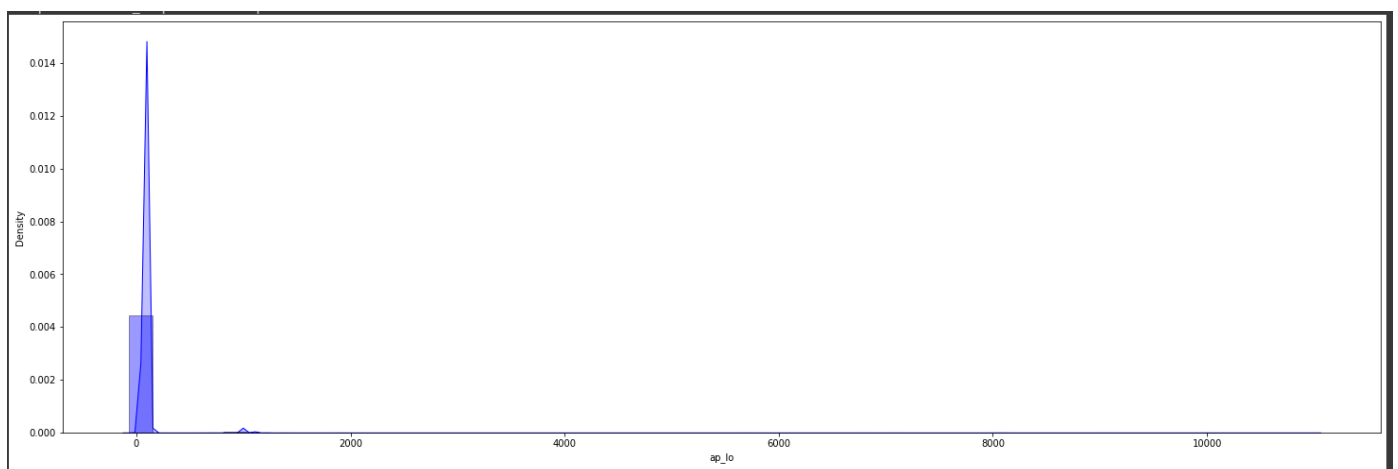
la fonction de répartition des poids chez
les malades :



la densité en fonction de la tension systolique :



la densité en fonction de la tension artérielle diastolique :



J'ai essayé de résumer mon projet dans ce fiche du projet. Pour savoir plus de détails sur le projets, veuillez ouvrir le lien du projet (j'ai travaillé sur Google Collab).
J'espère que vous avez aimez mon projet.

Cordialement,