

Mini projet module Data mining

Analyse des caractères l'influence sur l'obésité

Réaliser par :Bouarfa abdelilah

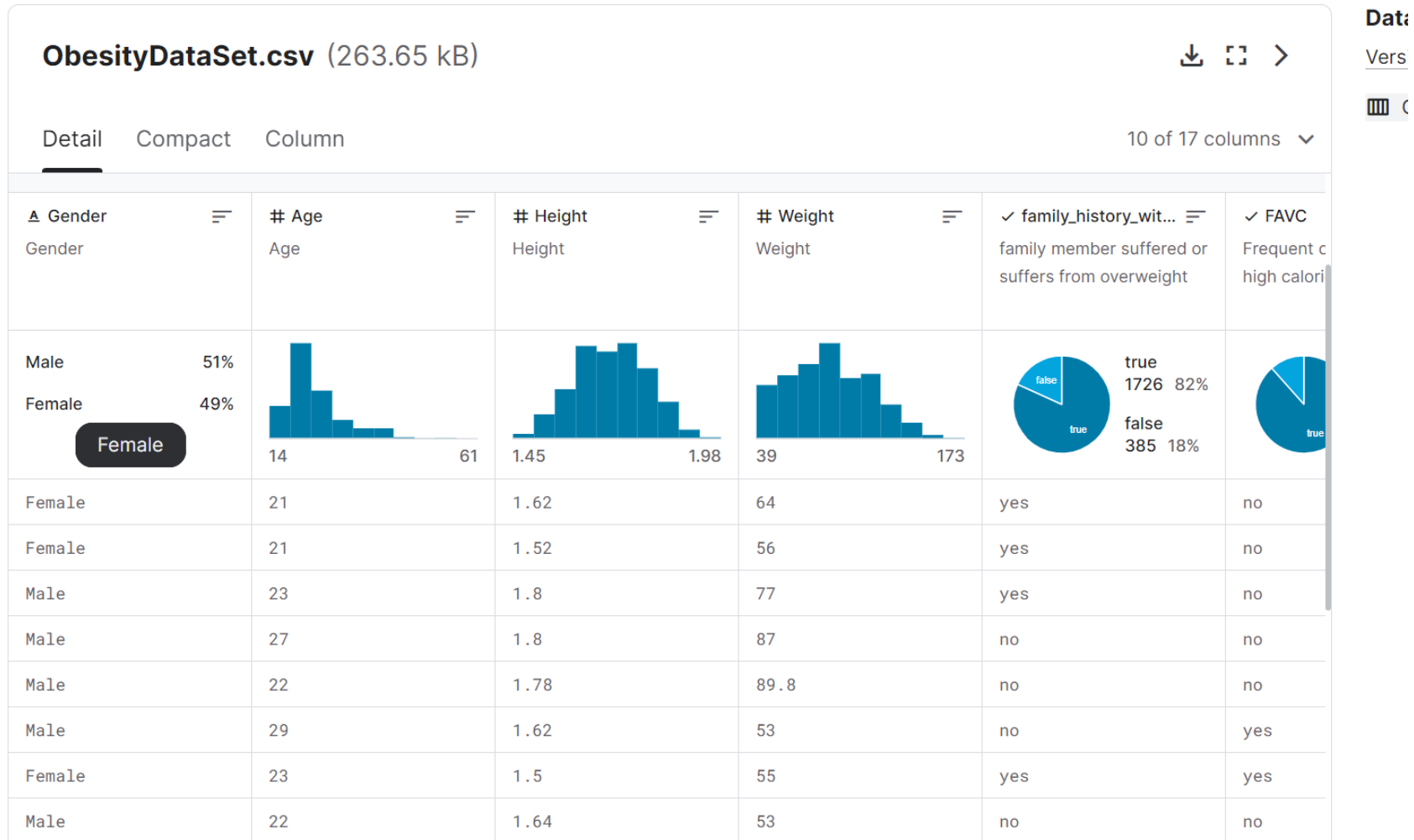
Encadre par: mohmed sabiri



Introduction

Bienvenue dans ce carnet de démarrage sur l'ensemble de données sur l'obésité. Les données consistent en l'estimation des niveaux d'obésité chez les personnes des pays du Mexique, du Pérou et de la Colombie, âgées de 14 à 61 ans et ayant des habitudes alimentaires et des conditions physiques diverses. Les données ont été collectées à l'aide d'une plateforme Web avec une enquête à laquelle des utilisateurs anonymes ont répondu. Chaque question, puis les informations ont été traitées pour obtenir 17 attributs et 2111 enregistrements..

Exploration de données



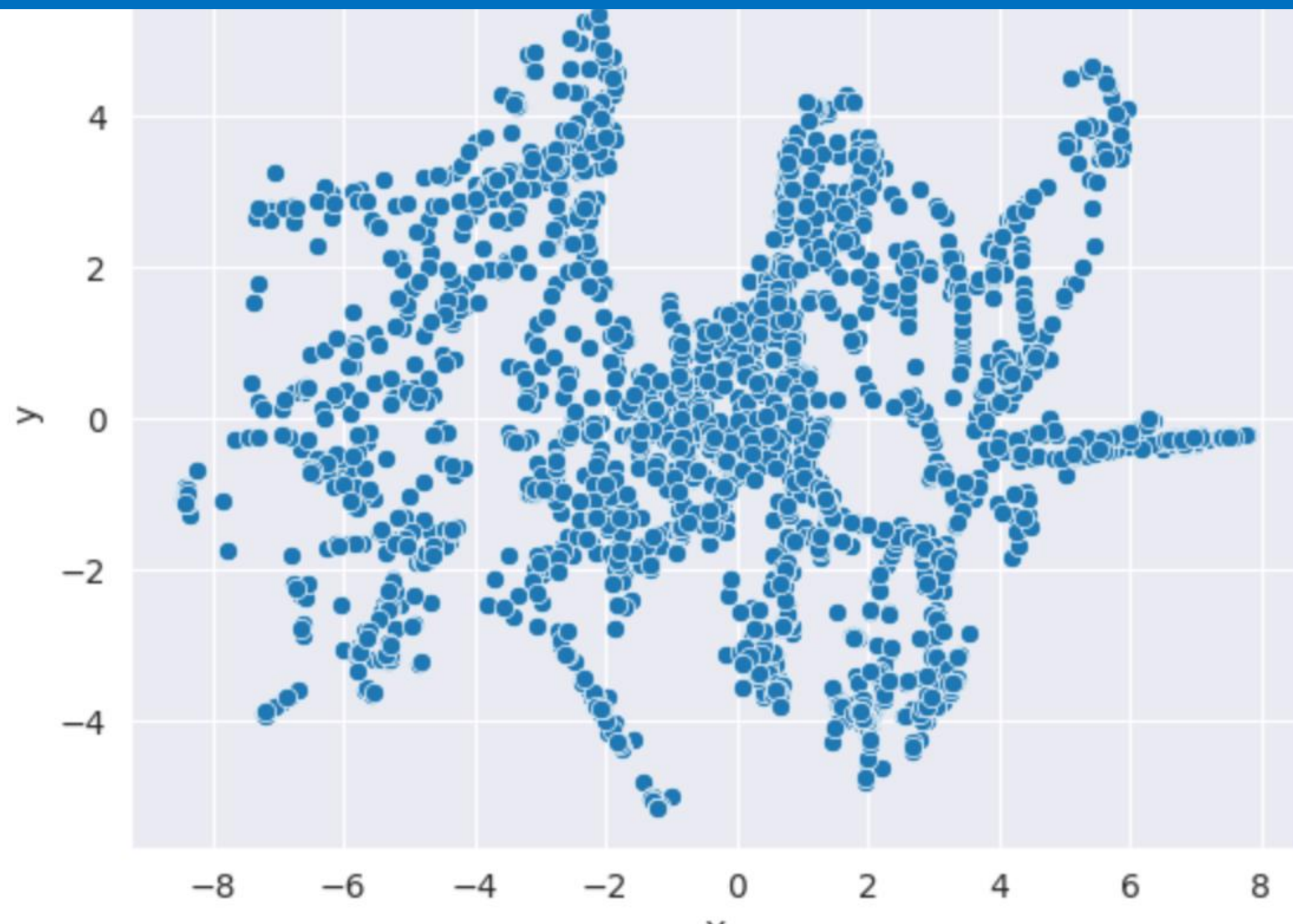
la correspondance de chaque attribut

- **FAVC** - Consommation fréquente d'aliments riches en calories
- **FCVC** - Fréquence de consommation de légumes
- **NCP** - Nombre de repas principaux
- **CAEC** - Consommation de nourriture entre les repas
- **CH2O** - Consommation d'eau quotidienne
- **CALC** - Consommation d'alcool
- **SCC** - Les attributs liés à la condition physique sont : Suivi de la consommation de calories
- **FAF** - Fréquence d'activité physique
- **TUE** - Temps passé à utiliser des appareils technologiques
- **MTRANS** - Transport utilisé

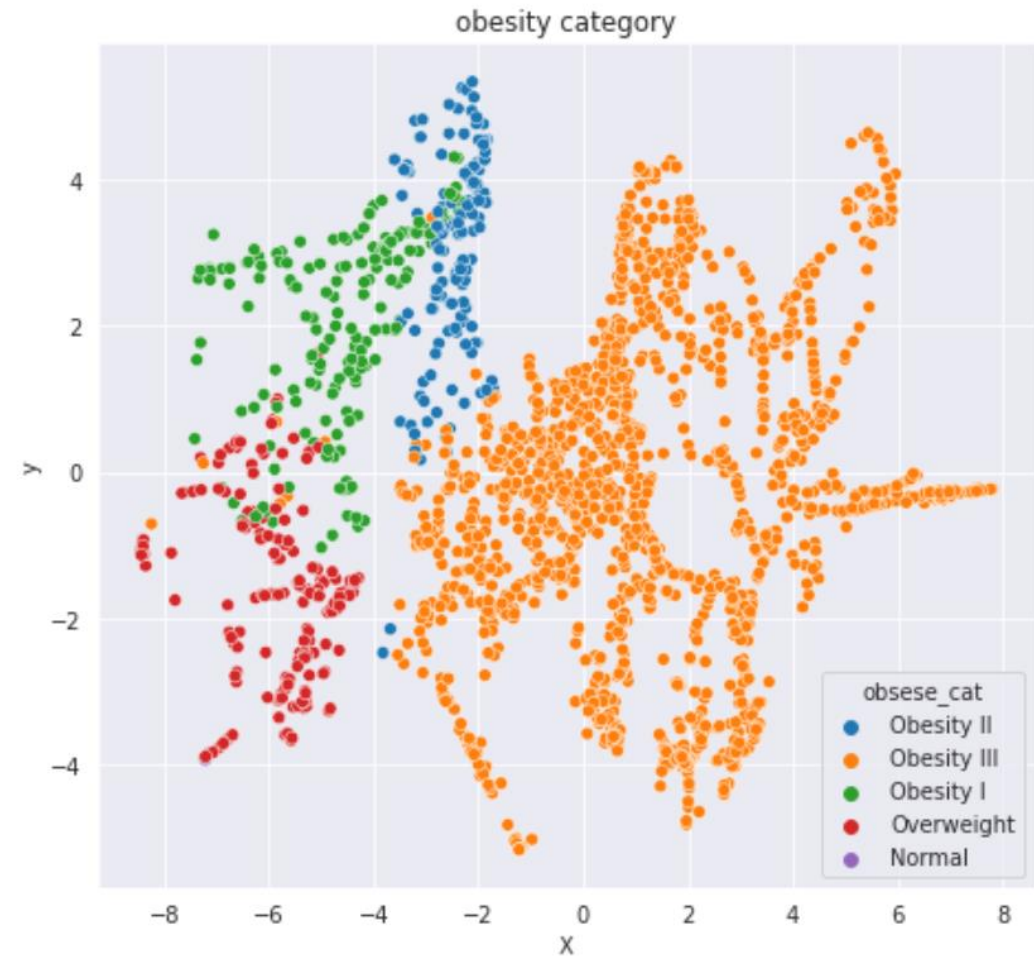
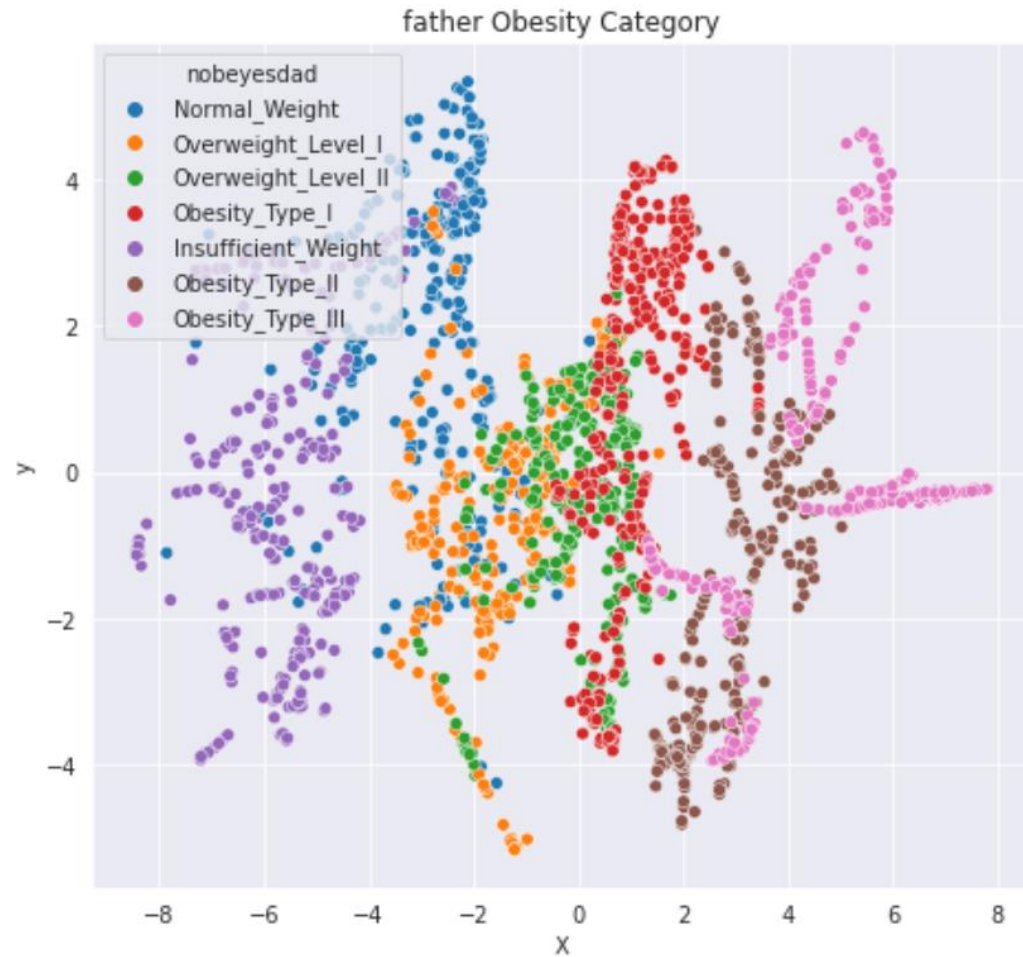
Description des données

:	Age	Height	Weight	FCVC	NCP	CH2O	FAF	TUE
count	2111.000000	2111.000000	2111.000000	2111.000000	2111.000000	2111.000000	2111.000000	2111.000000
mean	24.312600	1.701677	86.586058	2.419043	2.685628	2.008011	1.010298	0.657866
std	6.345968	0.093305	26.191172	0.533927	0.778039	0.612953	0.850592	0.608927
min	14.000000	1.450000	39.000000	1.000000	1.000000	1.000000	0.000000	0.000000
25%	19.947192	1.630000	65.473343	2.000000	2.658738	1.584812	0.124505	0.000000
50%	22.777890	1.700499	83.000000	2.385502	3.000000	2.000000	1.000000	0.625350
75%	26.000000	1.768464	107.430682	3.000000	3.000000	2.477420	1.666678	1.000000
max	61.000000	1.980000	173.000000	3.000000	4.000000	3.000000	3.000000	2.000000

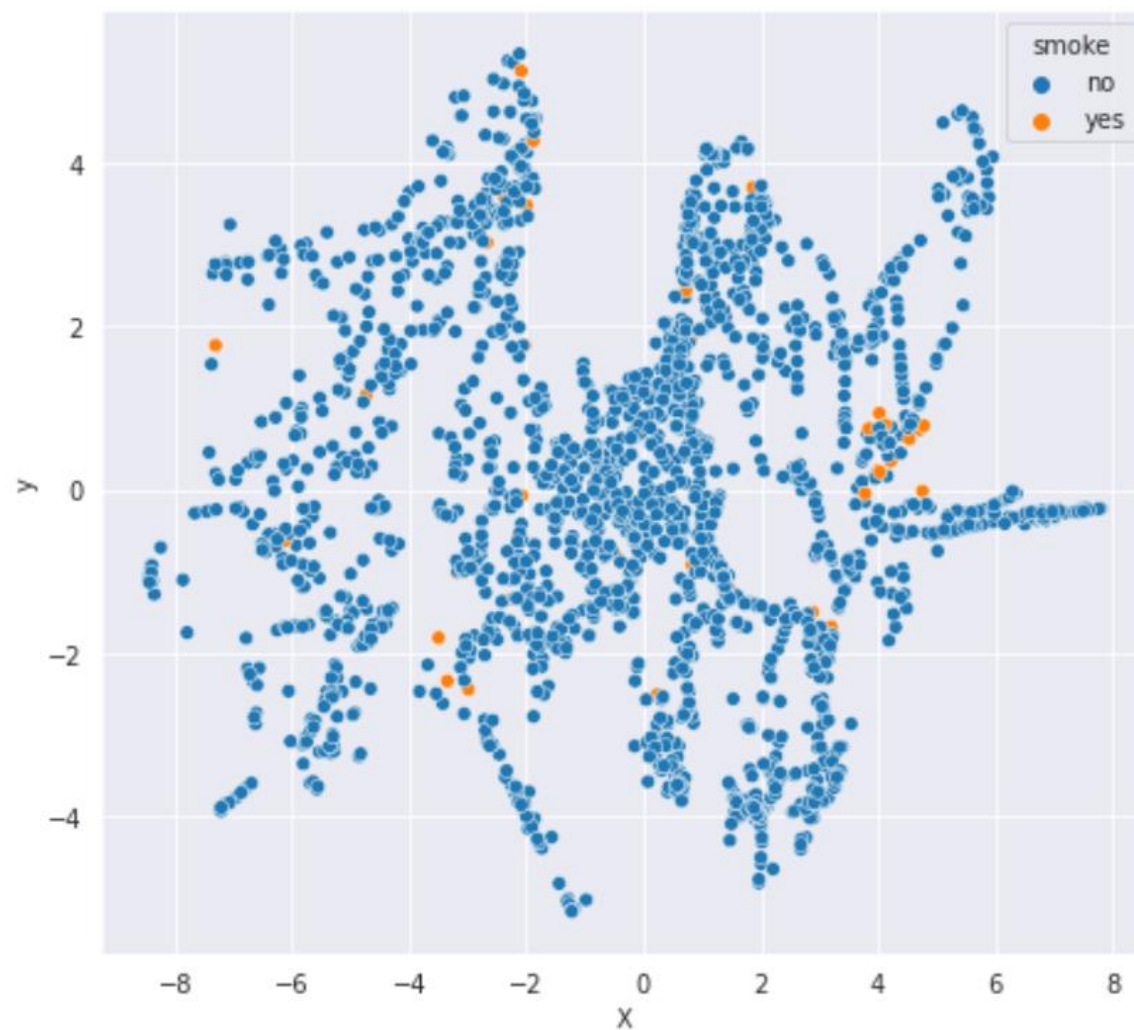
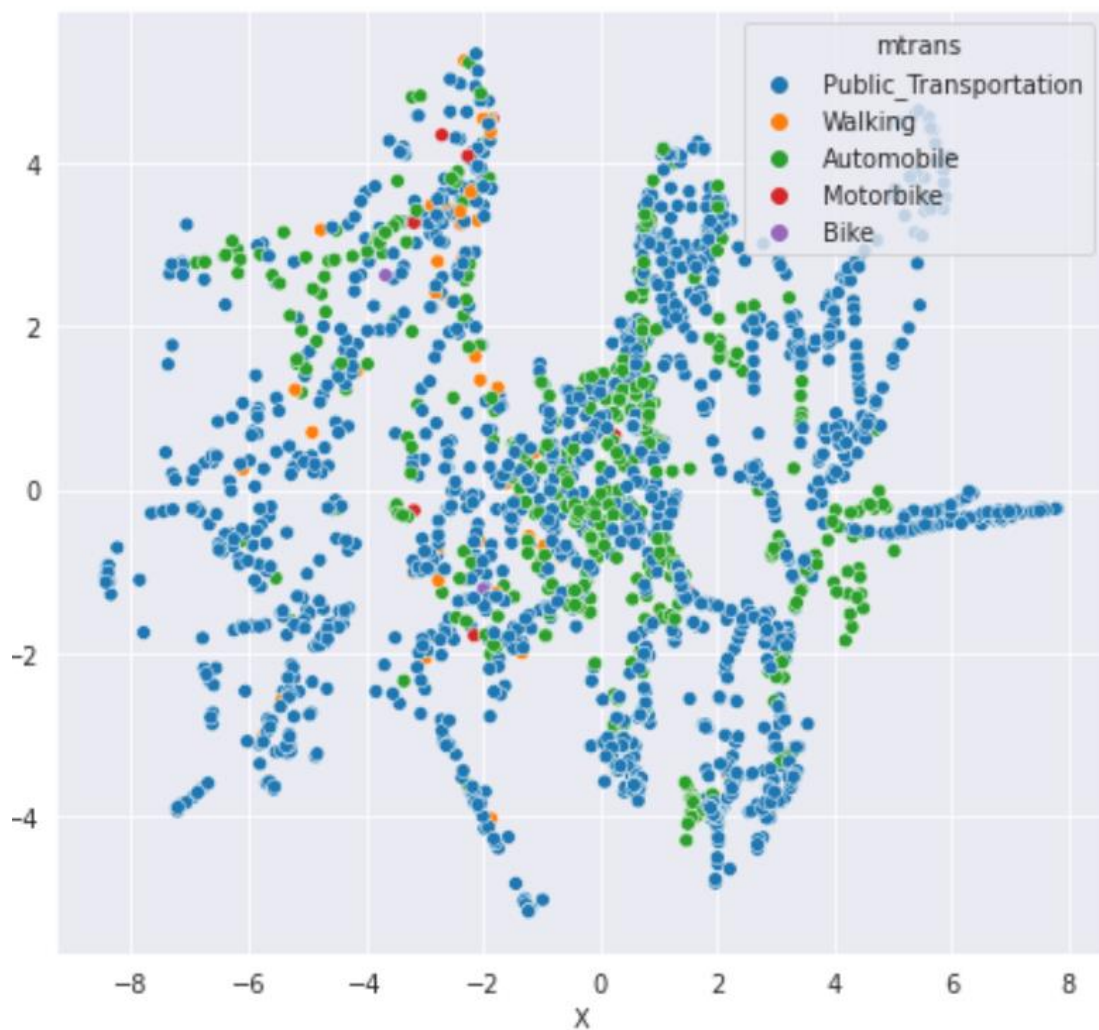
attribuer les fonctionnalités TNSE



l'obésité du père affecte la catégorie de poids de l'enfant.



le tabagisme a très peu d'impact sur l'obésité.



étapes

Seuil de variance

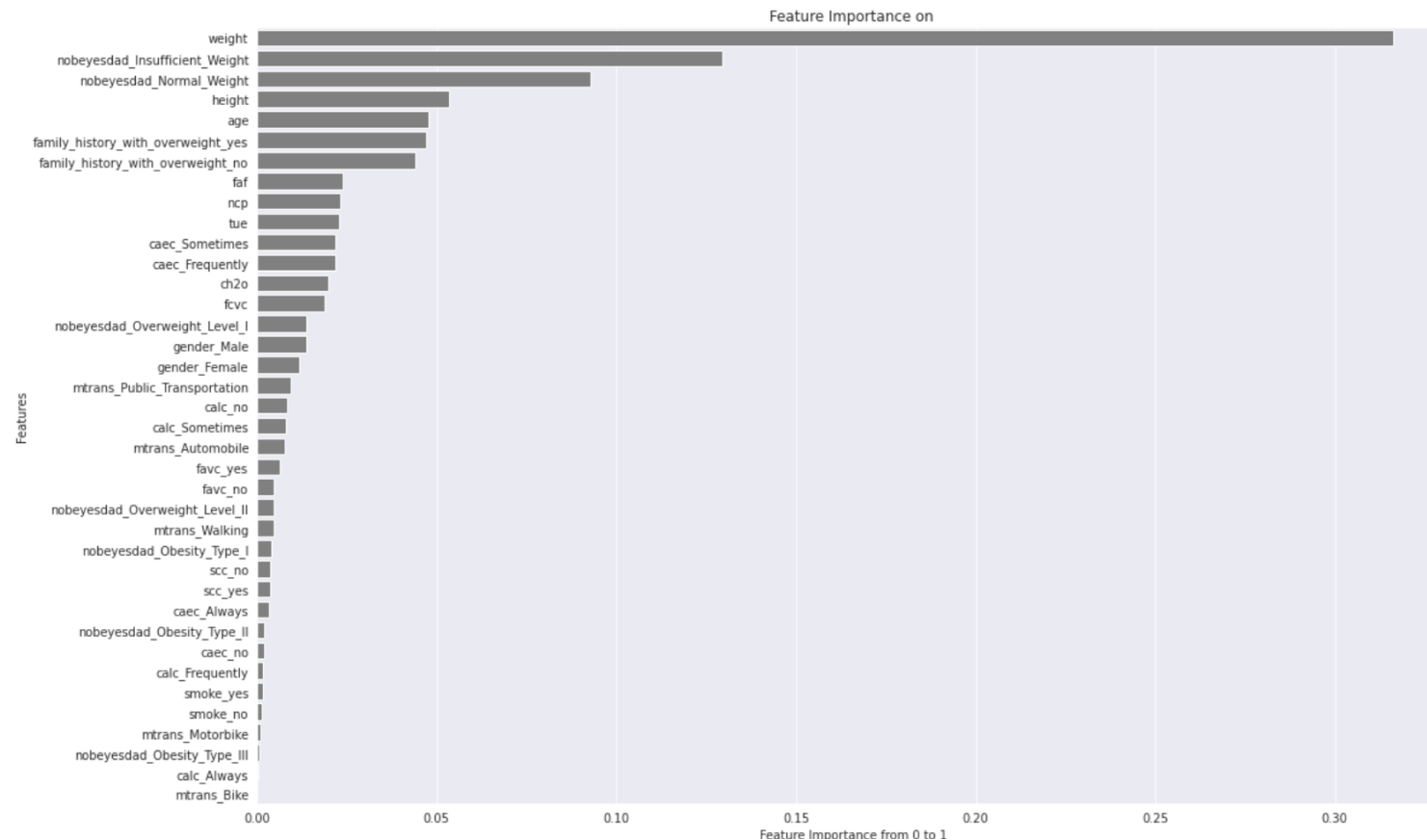
- Supprimer les données hautement corrélées
- Sélection d'une fonctionnalité pour les performances du modèle

Ingénierie des fonctionnalités

L'ingénierie des fonctionnalités fait référence à la manipulation (ajout, suppression, combinaison, mutation) de votre ensemble de données. pour améliorer la formation des modèles d'apprentissage automatique, conduisant à de meilleures performances et à une plus grande précision.

- **Utilisation du classificateur Random Forest**

Résultat



Le poids est le facteur majeur de l'obésité et le mode de transport vélo réduit les risques d'obésité