

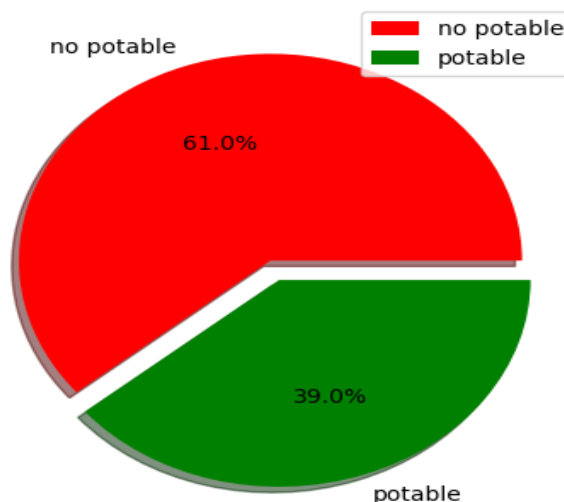
Analyse Exploratoire de Données

1) Analyse de Forme :

- Variables Target :
Potabilité
- Nombres de lignes et Nombres de colonnes :
 - lignes : 3276
 - colonnes : 10
- Nombres de types de chaque variable :
 - qualitatives : 1
 - quantitatives : 9
- Quantité des valeurs manquantes de chaque variable :
 - ph : 15%
 - Sulfate : 24%
 - Trihalomethanes : 49%
 - Aucune valeur manquante pour les autres variables

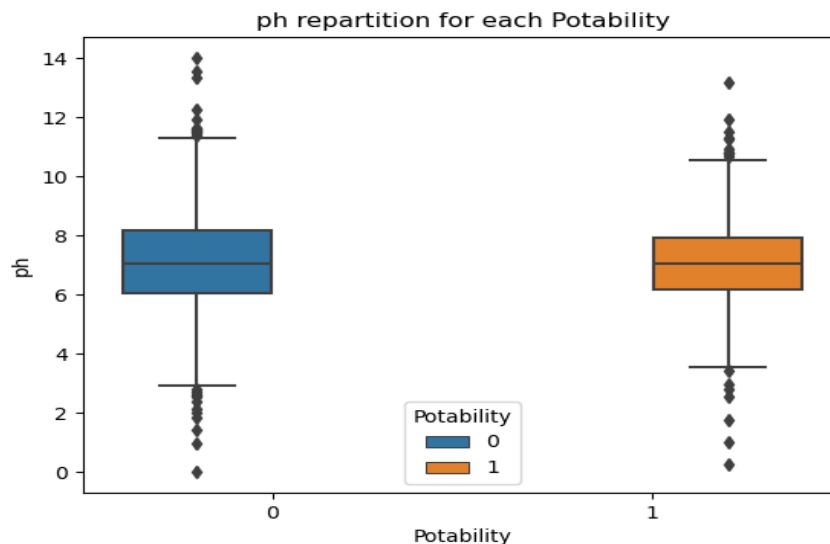
2) Analyse de fond et plus détaillée :

- Analyse de variables Target :
 - La classe de la variable Target n'est pas équilibré.
 - 61% de la variable Target prennent l'étiquette 'no potable', il est beaucoup plus grand par rapport à l'autre étiquette 'potable'.
 - 39% restant prennent l'étiquette 'potable'.

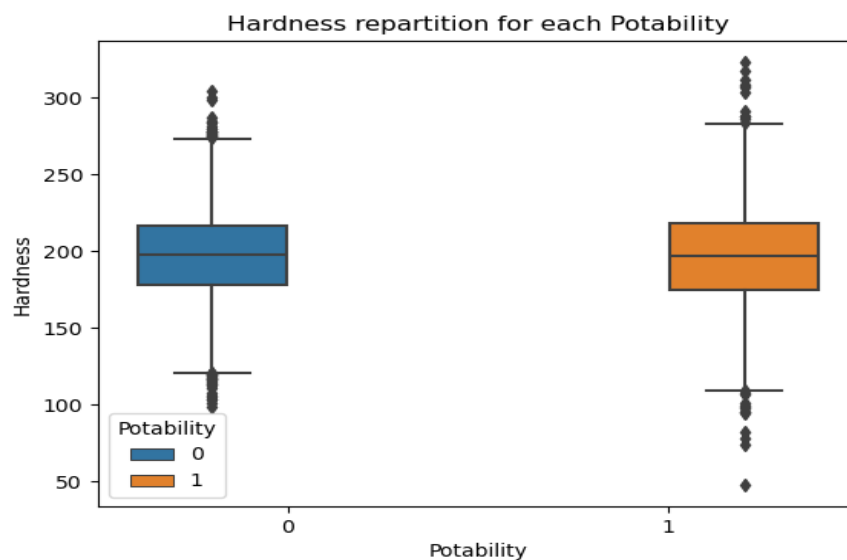


- **Analyse de relation entre les variables features et la variable Target :**

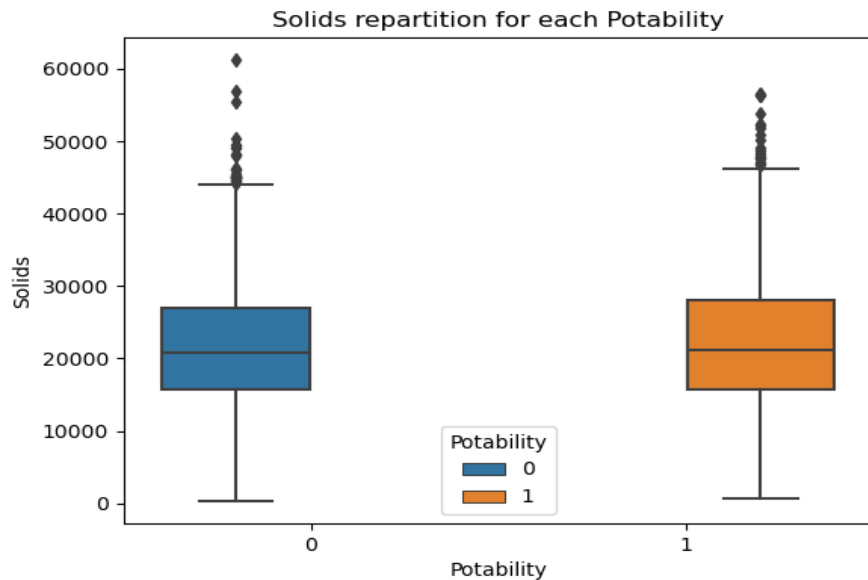
- **Target / Solides :** il semble que la répartition de la variable 'ph' sous la classe 0 est à peu près la même que celle de la classe 1. Le 'ph' de chaque classe est donc majoritairement entre 6 et 8 et il y a quelques valeurs extrêmes qui sont inférieurs à 3 voir jusqu'à 0 et supérieurs à 11 voir jusqu'à 14.



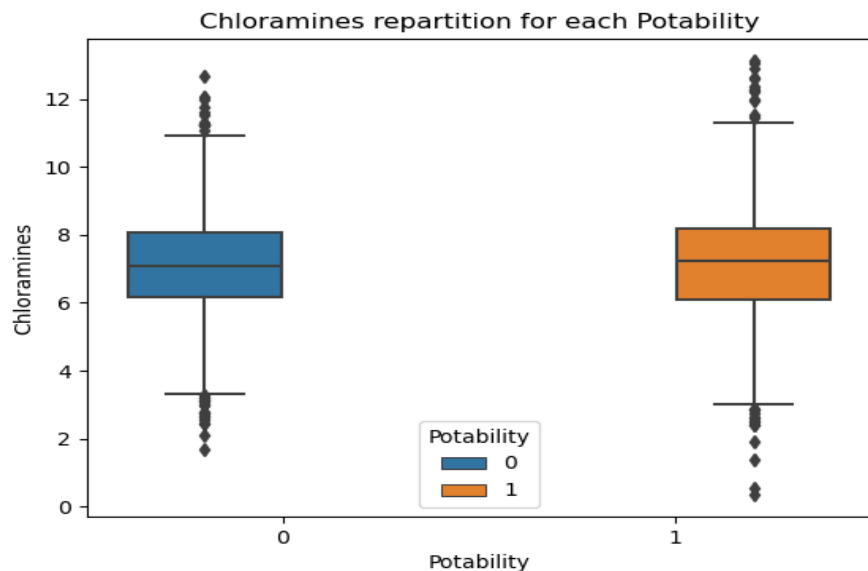
- **Target / Hardness :** il semble que la répartition de la variable dureté sous la classe 0 est à peu près la même que celle de la classe 1. La dureté de chaque classe est donc majoritairement entre 125 et 225 et il y a quelques valeurs extrêmes qui sont inférieurs à 125 voir jusqu'à 47 (qui est dans la classe 1) et supérieurs à 11 voir jusqu'à 304 (qui est dans la classe 1).



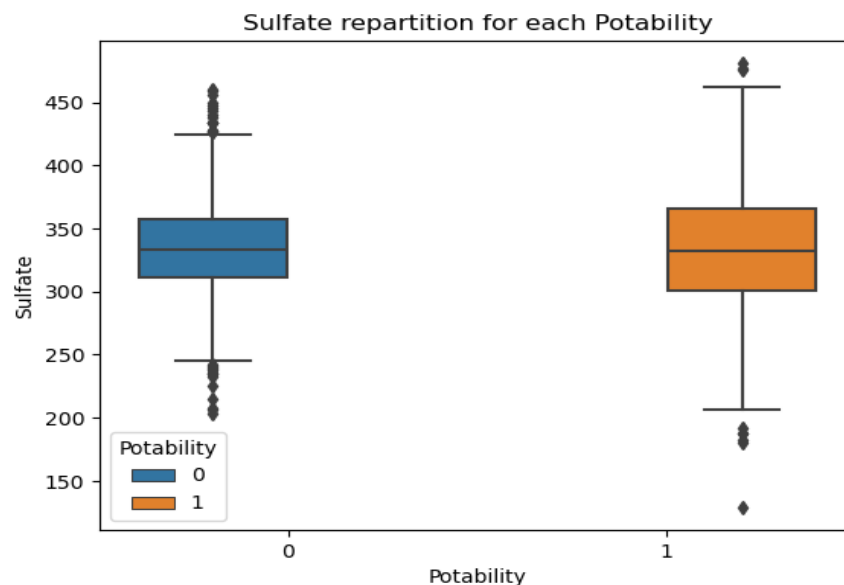
- **Target / Solides** : il semble que la répartition de la variable Solides sous la classe 0 est à peu près la même que celle de la classe 1. Les Solides de chaque classe sont donc majoritairement entre 15000 et 18000 et il y a quelques valeurs extrêmes qui sont supérieurs à 45000 voir jusqu'à 61000 (qui est dans la classe 0).



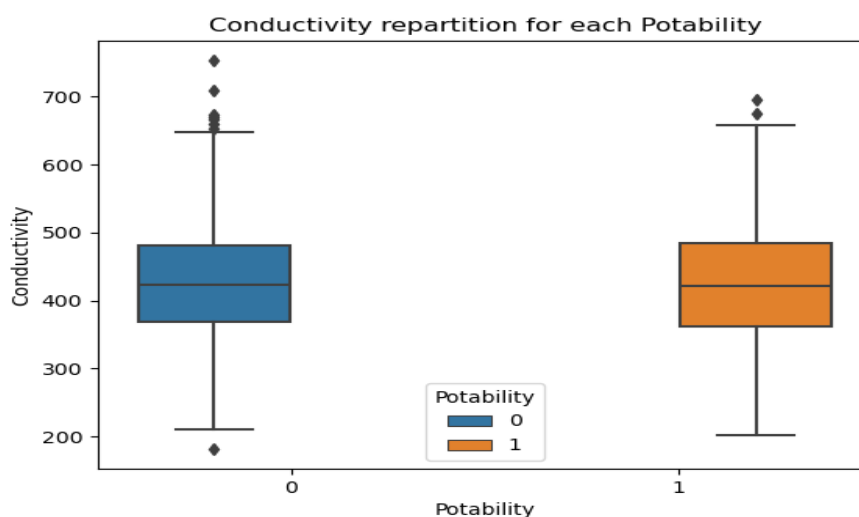
- **Target / Chloramines** : il semble que la répartition de la variable Chloramines sous la classe 0 est à peu près la même que celle de la classe 1. Les Chloramines de chaque classe sont donc majoritairement entre 6 et 8 et il y a quelques valeurs extrêmes qui sont inférieurs à 3 voir jusqu'à 0,35 (dans la classe 1) et supérieurs à 11 voir jusqu'à 13,12 (qui est dans la classe 1).



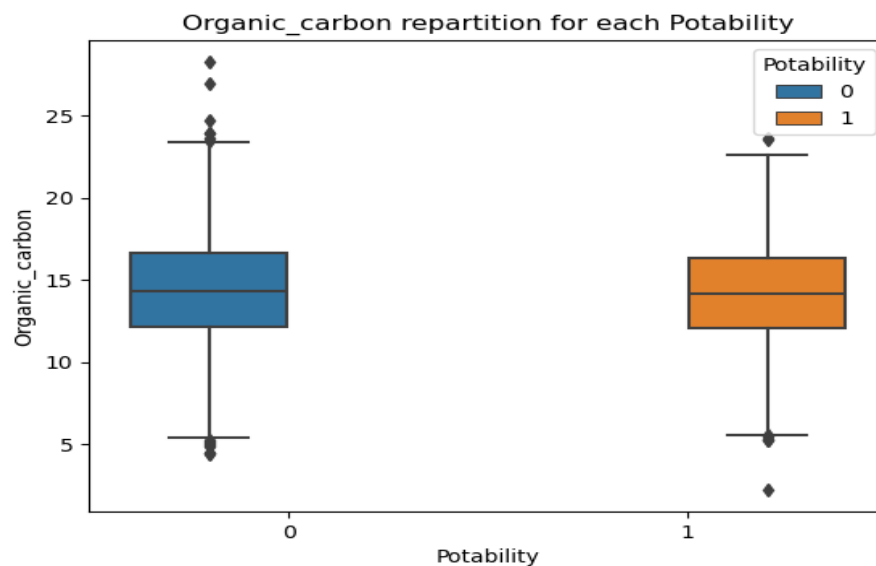
- **Target / Sulfate** : il semble que la répartition de la variable Sulfate sous la classe 0 est un peu différent que celle de la classe 1. La Sulfate dans la classe 0 est donc majoritairement entre 310 et 360 et il y a quelques valeurs extrêmes qui sont inférieurs à 250 voir jusqu'à 203 et supérieurs à 425 voir jusqu'à 460 tandis que la Sulfate dans la classe 1 est majoritairement entre 300 et 375 et il y a quelques valeurs extrêmes qui sont inférieurs à 280 voir jusqu'à 129 et supérieurs à 460 voir jusqu'à 481.



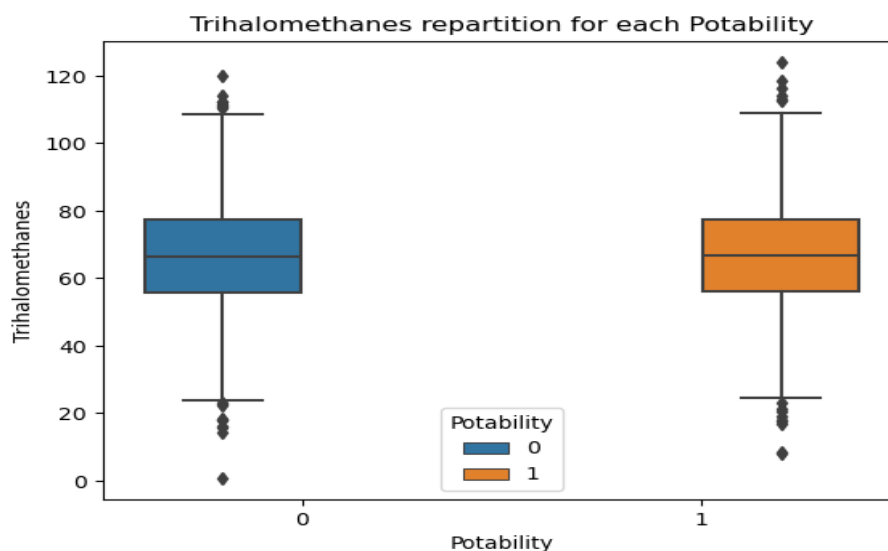
- **Target / Conductivité** : il semble que la répartition de la variable Conductivité sous la classe 0 est à peu près la même que celle de la classe 1. La Conductivité de chaque classe est donc majoritairement entre 370 et 480 et il y a quelques valeurs extrêmes qui sont supérieurs à 650 voir jusqu'à 753,34 (qui est dans la classe 0) et une seule valeur extrême minimale qui est dans la classe 0 et est égale à 181,48.



- **Target / Organic_carbon** : il semble que la répartition de la variable Organic_carbon sous la classe 0 est à peu près la même que celle de la classe 1. Le Organic_carbon de chaque classe est donc majoritairement entre 12,5 et 17,5 et il y a quelques valeurs extrêmes qui sont inférieures à 5 voir jusqu'à 2,2 (qui est dans la classe 1) et supérieurs à 23 voir jusqu'à 28,3 (qui est dans la classe 0).



- **Target / Trihalomethanes** : il semble que la répartition de la variable Trihalomethanes sous la classe 0 est à peu près la même que celle de la classe 1. Le Trihalomethanes de chaque classe est donc majoritairement entre 56 et 78 et il y a quelques valeurs extrêmes qui sont inférieures à 23 voir jusqu'à 0,73 (qui est dans la classe 0) et supérieurs à 110 voir jusqu'à 124 (qui est dans la classe 1).



- **Target / Turbidité** : il semble que la répartition de la variable Turbidité sous la classe 0 est à peu près la même que celle de la classe 1. La Turbidité de chaque classe est donc majoritairement entre 3,5 et 4,5 et il y a quelques valeurs extrêmes qui sont inférieurs à 2 voir jusqu'à 1,45 (qui est dans la classe 0) et supérieurs à 6 voir jusqu'à 6,73 (qui est aussi dans la classe 0).

