

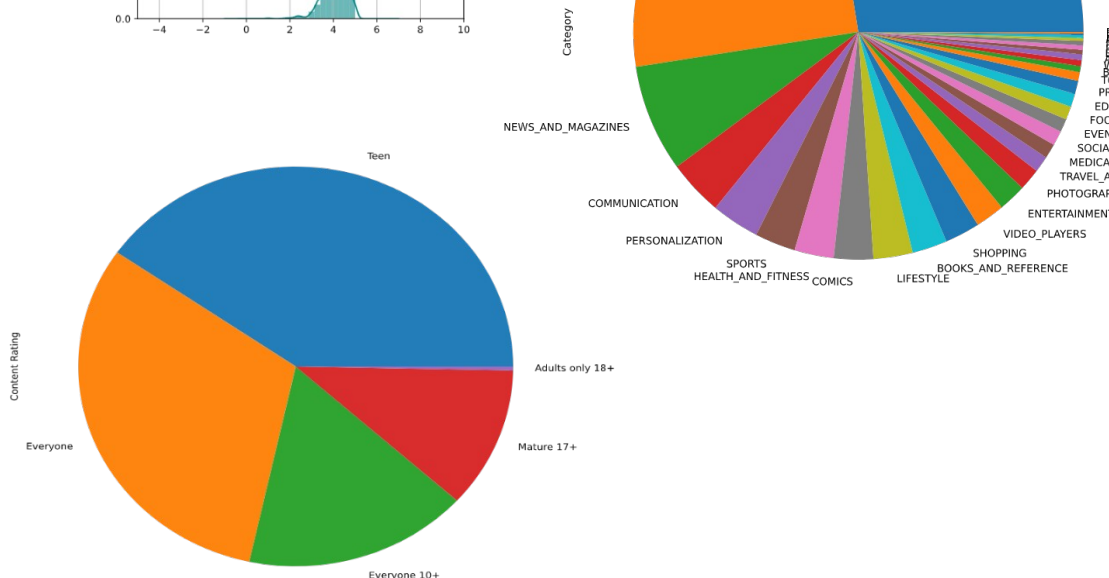
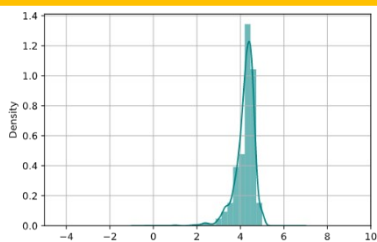
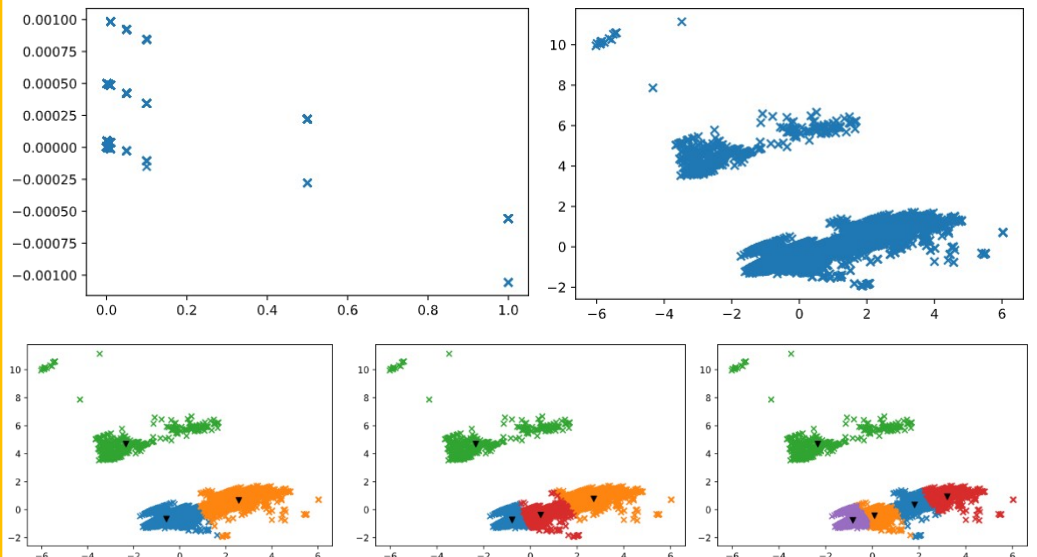
Problème 2 : Apprentissage non supervisé

-Trouver la catégorie et le Content Rating d'applications ayant le plus haut nombre d'installations ainsi qu'une bonne note, grâce à l'utilisation de l'algorithme des k-moyennes .

Étape 1:

Visualisation des données

- 1-Filtrage, One hot Encoding.
- 2-Changer la dimension de nos données de N dimensions à 2 dimensions.
- 3-Trouver le bon K, avec l'algorithme des k-moyennes.



Étape 2 :

Analyse des clusters:

1-Appliquer l'algorithme des K-moyennes sur les données pour $k=5$.

2-Afficher le Rating, Content Rating et les catégories de chaque cluster .

3-Calculer la moyenne et la variance sur le nombre d'installations de chaque cluster.

Étape 3 :

Résultats et Comparaison des clusters

1-Comparaison de la moyenne et de la variance du nombre d'installations de chaque cluster.

2-Déterminer le cluster avec le plus grand nombre d'installations et représenter ses catégories et son Content Rating.

```
cluster n° 0
moy : 275304.8487804878
var : 16740005731992.092
cluster n° 1
moy : 2641076.4518664046
var : 92594285284434.19
cluster n° 2
moy : 53279252.9676071
var : 3.0833157453792404e+16
cluster n° 3
moy : 24637643.420752566
var : 6125334585173725.0
cluster n° 4
moy : 41123153.20261438
var : 2.189179201342625e+16
```