

# Présentation Projet LU3IN026 – Analyses des données Google Play Store

MISTRY Manishkumar, 3671086  
ERGUN Yavuz, 3670777

## Introduction

Les applications Google Play sont aujourd’hui populaire grâce à leur popularité, genre et leurs nouvelles mise à jours  
Dans ce projet, nous avons analyser différents donnés comme les Catégories, les Ratings (notes), les Install etc. Une fois les analyse faite nous sommes parvenu à avoir des problématiques pour les:

- classification supervisé
- classification non-supervisé.

## Classification Supervisée

## Problématique

Aujourd’hui, les notations (Ratings) sont un moyen naturel d’augmenter la découverte d’application et le nombre de téléchargements. En effet, une application est bien notée par de nombreuses évaluation sera mieux classée que les autres applications de la catégories.

Notre problématique sera, en quoi la note peut-elle convaincre l’utilisateur de choisir une applications?

## Modèle

Nous avons utilisé des modèles d’apprentissage supervisé comme le Perceptron, les Adalines Analytique. Nous avons utilisé ces modèle en Cross-validation (cross-validation strat) tout en séparant les donnés.

	Accuracy Test	Acurracy Apprentissage
<div>Perceptron</div>	82,15	87,00
Adaline	56,82	58,17
Adaline2	90,48	93,3

## Conclusion

Après les analyses, nous avons décidé de prendre les résultats du Perceptron qui nous donne un accuracy de 87%.  
D’après les analyse du poids:

- On constate que la majorité des applications sont gratuites.
- Ils sont les plus vues et meilleurs notés
- Les Applications les mieux notés sont donc ceux qui sont les plus choisi par les gens

Ainsi, nous pouvons en déduire que les applications avec des meilleurs notes ont un impact fort sur le nombre de son installations.

## Classification Non-Supervisée

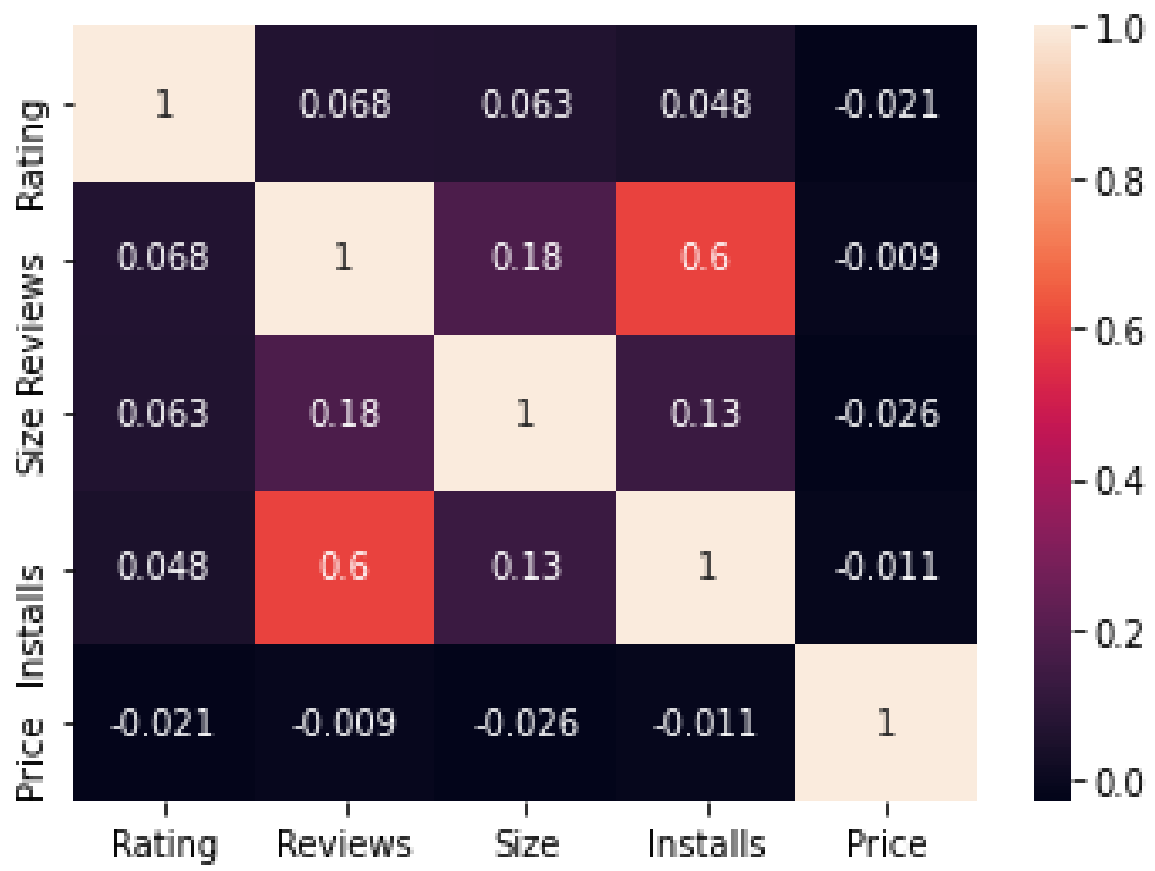
## Problématique

Dans le but de concevoir et de vendre une application sur le Play Store, nous avons décidé d’étudier la base de donnée pour isoler certaine caractéristique des application qui on très bien marché afin de répliquer leurs succès.

## Modèle

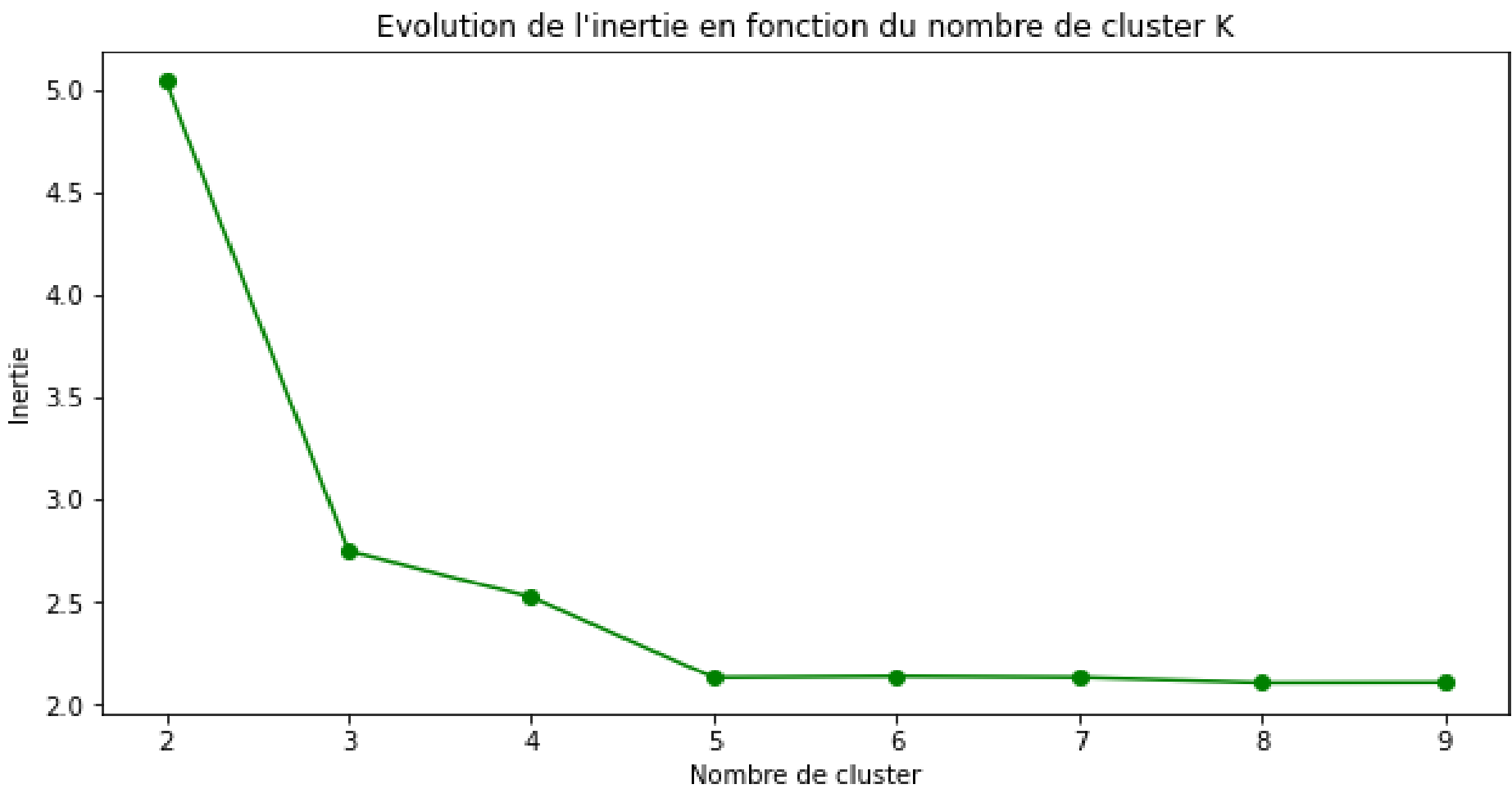
Pour ce faire nous avons utiliser le modèle d’apprentissage non supervisé des K-moyenne. Pour choisir notre K, nous avons utiliser la méthode du coude qui consiste a tester plusieurs k et de prendre celui ou l'inertie ne varie plus beaucoup après.

## Protocole



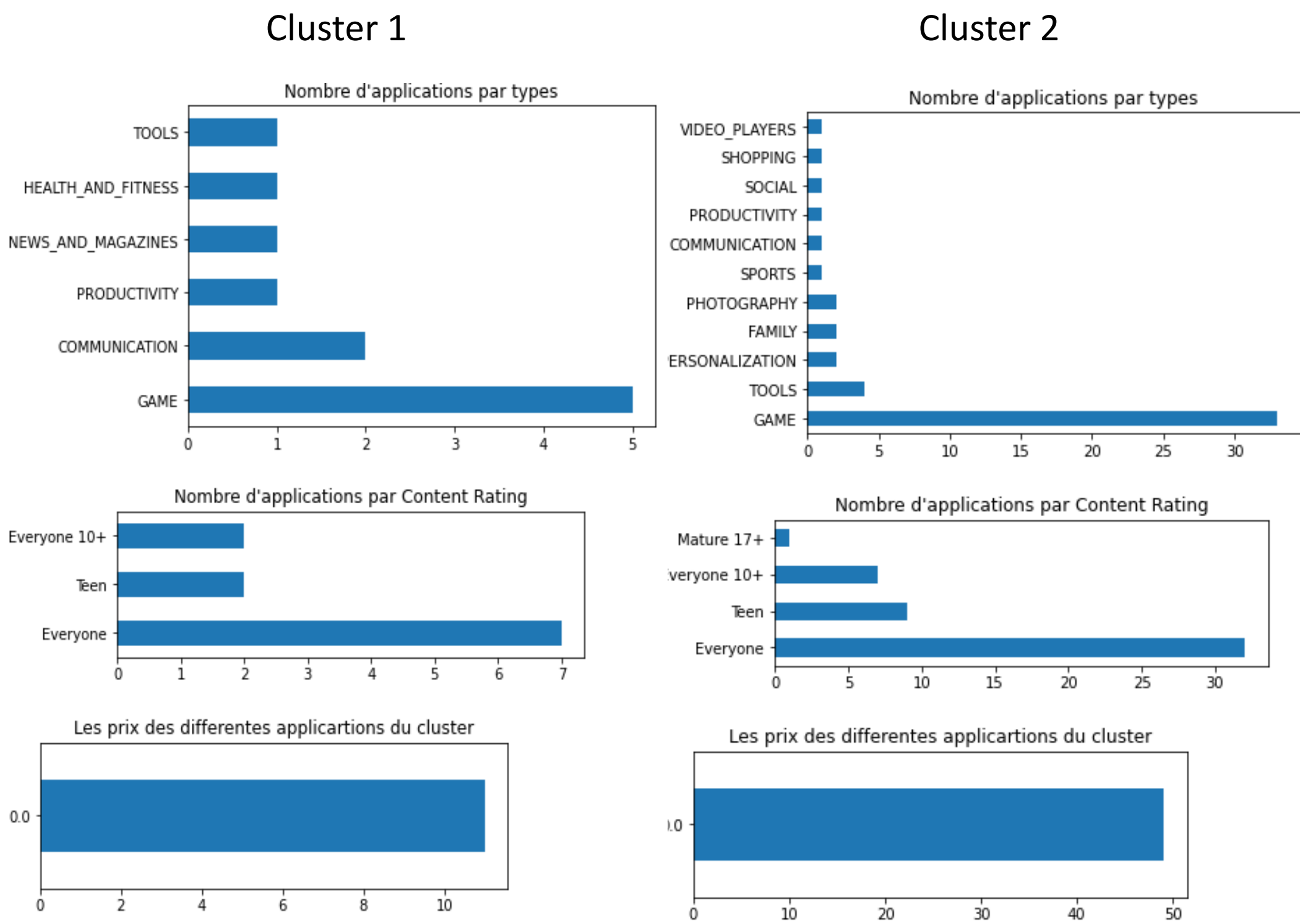
Premièrement nous avons chercher une certaine corrélation dans nos données afin de mieux ciblé, notre matrice de corrélation nous a clairement révéler une certaine corrélation entre les reviens et le nombre d’installation

Nous avons ensuite fixer notre nombre de cluster pour notre algorithme de K-moyenne. Pour ce faire nous avons utiliser la méthode du coude qui consiste a tester plusieurs k et de prendre celui ou l'inertie ne varie plus beaucoup après. Ce qui nous a donné 5 comme nombre de cluster optimale.



## Résultat

Après une visualisation de nos 5 clusters, seulement 2 nous intéresse, avec un qui sort particulièrement du lot avec sont nombre moyen d’installation très élever, ce sont les quelque applications exceptionnel qui on extrêmement bien marcher et l’autre cluster qui regroupe les applis ayant très bien marcher, les 3 autres cluster ne nous satisfaisait pas sur les critère d’installation et de Reviews donc nous les avons pas étudier plus que cela.



## Conclusion

En regardant attentivement ces deux clusters on se rend compte que les applications qui marche le mieux au niveau des Rating/Review/Install sont des jeux, cela nous donne une certaine réponse compte à la catégorie de notre application. Ensuite on observe qu'au niveau du genre des jeux les genres Action, Arcade, casual revienne le plus. Donc notre jeu sera d'un de ces 3 genres. Plus précisément notre jeu visera la globalité des utilisateurs et sera gratuit avec surement des achat in apps comme la majorité des jeux présent dans le cluster 1 et 2.